# Building Scalable and Smart Multimedia Applications on the Semantic Web

Michael Hausenblas

# Building Scalable and Smart Multimedia Applications on the Semantic Web

Doctoral Thesis

at

Graz University of Technology

submitted by

**Michael Hausenblas**

Institute for Software Technology (IST),
Graz University of Technology
8010 Graz, Austria

and

Institute of Information Systems & Information Management (IIS),
JOANNEUM RESEARCH Forschungsges. mbH
8010 Graz, Austria

2008

© Copyright 2004—2008 by Michael Hausenblas

First Reader: Univ.-Prof. DI Dr. Wolfgang Slany
Second Reader: Univ.-Prof. Dr.rer.nat. Klaus Tochtermann

TUG Graz University of Technology
Erzherzog-Johann-University

JOANNEUM RESEARCH

# Abstract

The Semantic Web has become reality over the past couple of years. While certain practical topics—such as interoperability, etc.—have at least partially been addressed, scalability and expressivity issues regarding the utilisation of multimedia metadata on the Semantic Web are still widely neglected. However, existing Web (2.0) applications handling millions of multimedia assets are starting to take advantage of Semantic Web technologies.

This work contributes to design decisions regarding scalable and smart multimedia applications on the Semantic Web. Based on an analysis of practical issues—stemming from diverse projects and activities the author has participated in over the past four years—three areas have been identified, namely (i) performance and scalability issues on the data access level, (ii) the effective and efficient representation of multimedia content descriptions, and (iii) the deployment of multimedia metadata on the Semantic Web. The three research areas have as its common base the trade-off between expressivity and scalability.

We present our findings regarding scalable, yet expressive Semantic Web multimedia applications in a number of practical settings and discuss future directions, such as "interlinking multimedia".

# Kurzfassung


Im Laufe der letzten Jahre wurde das Semantic Web Realität. Obgleich einige praktische Fragen, wie beispielsweise Interoperabilität, schon teilweise behandelt wurden, sind die Themen Skalierbarkeit und Expressivität in Bezug auf die Ausnützung von Multimedia-Metadaten im Semantic Web bislang vernachlässigt worden. Bestehende Web (2.0) Anwendungen, die Millionen von multimedialen Inhalten handhaben, beginnen von Semantic Web Technologien zu profitieren.

Die vorliegende Arbeit unterstützt Designentscheidungen beim Bau von multimedialen Semantic Web Applikationen. Dabei wurden, ausgehend von einer umfassenden Analyse praxisnaher Probleme (basierend auf Projekten bei denen der Verfasser dieser Arbeit beteiligt war) drei Bereiche identifiziert: Erstens, Performanz und Skalierbarkeitsfragen auf der Datenzugriffsebene, zweitens, effiziente und effektive Repräsentation von Beschreibungen multimedialer Inhalte, und schließlich der Gebrauch von Multimedia-Metadaten am Semantic Web. Gemein ist den oben genannten Forschungsbereichen die Kompromissfindung in Bezug auf Expressivität vs. Skalierbarkeit.

In der Arbeit werden die Erkenntnisse bezüglich skalierbarer und dennoch ausdrucksstarker Semantic Web Applikationen im Multimediabereich im Rahmen einer Reihe realitätsnaher Aufgabenstellungen dargestellt. Schließlich werden zukünftige Entwicklungen (wie “interlinking multimedia”) diskutiert.

*I hereby certify that the work presented in this thesis is my own and that work performed by others is appropriately cited.*

Michael Hausenblas, June 2008.

# Acknowledgements

My uttermost thanks and thoughts to my family: Saphira, Ranya, Iannis and Anneliese. Without your support I could not possibly have done this work. Especially in the darker moments you have motivated me and pointed me towards the light a the end of the tunnel. Thank you.

I especially wish to thank my supervisor, Wolfgang Slany, for his support and for letting me benefit from his rich experience. It was most enjoyable to publish with you, and I learned a lot from you. Both regarding scientific work and the practical side as well; I am deeply thankful.

To my colleagues at JOANNEUM RESEARCH I would like to say thanks for your time and permanent will to discuss—without any special order these colleagues are: Werner Bailer, Werner Haas, Harald Mayer, Wolfgang Halb, Herwig Rehatschek, Martin Umgeher, Rudi Schlatte, Rene Kaiser, Martin Höffernig, Robert Sorschag, Lena Lauber, Hannes Bauer, Georg Mittendorfer, Herwig Zeiner, Selver Softic, Georg Thallinger, Gert Kienast, and Wolfgang Weiss. I further wish so express my gratitudes to my colleagues from the research projects NM2, K-Space, and SALERO: Doug Williams, Ian Kegel and Tim Stevens (all British Telecom), Marian Ursu (Goldsmith College), Maureen Thomas (Cambridge University), further Lynda Hardman, Jacco van Ossenbruggen, Raphaël Troncy, Frank Nack, Zeljko Obrenovic, and Lloyd Rutledge (CWI) as well as Tobias Bürger (STI Innsbruck) and Yves Raimond (Centre for Digital Music, Queen Mary, University of London).

From my work within W3C I learned a lot regarding standardisation processes, and administrative as well as community issues. I like to express my deep thankfulness to my colleagues at the Multimedia Semantics Incubator Group (especially Jeff Z. Pan, Raphaël Troncy, Vassilis Tzouvaras, Susanne Boll, Tobias Bürger, and Oscar Celma) and the RDFa Task Force of the Semantic Web Deployment Working Group (Ben Adida, Ralph Swick, Mark Birbeck, Steven Pemberton, Shane McCarron, and Manu Sporny) as well as other people within W3C who have strongly influenced my view: Ivan Herman, Dan Connolly, Guus Schreiber, and Fabien Gandon. Thanks for your patience. Finally, kudos to the genuine Linking Open Data chaps, Chris Bizer, Tom Heath, Richard Cyganiak, Kingsley Idehen, Yves Raimond—you shaped my understanding regarding the practical side of the Web of Data quite a lot. Thanks for your brilliant and visionary ideas and for sharing them with me. Regarding the practitioner's point of view I wish to thank Danny Ayers and Keith Alexander (both Talis), Benjamin Nowack (for ARC2), Dan Brickley (FOAF, uF, etc.), Lee Feigenbaum (regarding scovo/SPARQL) and David Peterson (our ISWC07 discussions); working with you certainly helped me to better understand and attack real-world issues.

Last but not least I want to thank my parents Gerhard and Gertrude, as well as my sister Monika and her family (Herbert, Larissa, and Elena) for their patience and support over the past couple of years.

Michael G. Hausenblas
Graz, Austria, 2008

# Credits

This thesis is based on Keith Andrews' wonderful LaTeX template. An up-to-date version can be obtained by visiting the FTP site of the Institute for Information Processing and Computer Supported New Media (http://ftp.iicm.tugraz.at/pub/keith/thesis/).

The *fancyhdr* package was used for the chapter headings. Kudos for this nice work to Piet van Oostrum. Note that the package is available via the CTAN network (http://www.ctan.org/tex-archive/macros/latex/contrib/fancyhdr/).

Some of the content and metadata stems from the FP6 EU project "New Media for a New Millennium" (NM2) (http://www.ist-nm2.org/). The author expresses deepest gratitude to the creative people in the production teams. The X3D example used in Fig. 3.3 on page 36 and the Fig. 3.4 on page 50 has been created within the NM2 project.

By courtesy of http://www.x-smiles.org the SMIL example in Fig. 3.2 on page 34 has been used.

The Fig. 4.1 on page 57 originates from *The Description Logic Handbook* [13].

The W3C Semantic Web stack used in Section 4.3 originally stems from W3C's Semantic Web activity homepage (http://www.w3.org/2001/sw/); especially credits go to the W3C Semantic Web activity lead, Ivan Herman.

# Contents

# List of Figures

# List of Tables

# List of Listings

# Part I

# Scope and Foundations

# Chapter 1

# Introduction

*"Vade mecum."*

(Latin phrase)

When "a message from Chad and Steve"[1] reached the YouTube community in early October 2006, people would ask: Why is Google going to put 1.65 billion dollar[2] on the counter? Without being in the board of Google it is hard to tell, though the core of the story is obvious: *it is the multimedia, stupid*!

Two fundamental types of resources are at odds on the Web: textual resources, and multimedia resources—or more specific, audio-visual content—such as a PNG still image, a MP3 music clip, or an AVI video clip. While for textual resources an array of research [85; 285] and tools are available[3], multimedia issues w.r.t. the Semantic Web have not yet been widely addressed. In this work we focus on multimedia resources, or—to be a bit more precisely—their description, and the respective usage of the descriptions.

## 1.1 Motivation

The demand for real-world applications on the Semantic Web is steadily increasing. Simultaneously, existing Web applications handling millions of multimedia assets are starting to take advantage of Semantic Web technologies [237]. Although in the past five to ten years an increase of research activities in the media semantics area can be noticed, several core problems are still not satisfactory solved. Effectively and efficiently accessing distributed data sources, dealing with the **Semantic Gap** in multimedia content descriptions, and deploying media asset descriptions on a Web-scale; these and other related issues stemming from real-world requirements may be one of the reasons for the—still widely academic minted—reputation of Semantic Web (multimedia) applications.

Different parameters may influence the performance, and the functionality of a Semantic Web multimedia application (SWMA). Attempting to build such scaleable and smart

[1] http://www.youtube.com/watch?v=QCVxQ_3Ejkg
[2] http://www.google.com/press/pressrel/google_youtube.html
[3] http://www.txtkit.sw.ofcd.com/

applications, one has to research manifold aspects of multimedia metadata generation, representation, and consumption. A multidimensional analysis is necessary to identify the requirements for a successful utilisation of media assets on the Semantic Web.

Regarding accessing distributed data sources, it can be noted that the *RDFising* process has not yet been widely researched. Some practical work has been reported, such as [264]. However, performance and scalability issues were neglected by and larger so far.

Furthermore, automated understanding of multimedia content is an issue in Semantic Web multimedia applications; often referred to as the "Semantic Gap", which is, following Smeulders et.al. [271]

> ... the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

Although substantial research efforts have been undertaken, a generic, domain-independent solution to the problem is not at hand. Understanding from a set of low-level features, such as colour, shape, etc. that these actually stand for (that is "mean") a certain entity in a domain—for example "tree"—is a non-trivial task.

Most of the activities or projects addressing the Semantic Gap are seldom more than research prototypes, using toy data sets. While the focus is often put on the expressivity of the description, aspects as performance and scalability, extensibility, and interoperability still have not been widely addressed. Studer et. al. [284] recently claimed:

> Another challenge is to manage the expressivity-scalability trade-off of reasoning over declarative knowledge, enabling reasoning over large-scale distributed knowledge bases for suitably expressive knowledge representations. Automated knowledge acquisition will typically yield knowledge that's uncertain—for example, fuzzy or probabilistic. Such knowledge must be represented and reasoned with in an adequate and scalable way. As knowledge from distributed knowledge bases is aggregated, a deeper semantics can emerge, letting intelligent agents discover patterns across people, roles, and tasks.

This work aims at addressing the *expressivity-scalability tradeoff* in the realm of multimedia applications operating on the Semantic Web. The following example illustrates, how easy one may run into troubles, when dealing with a detailed description of audio-visual content.

***Example 1.1 (Low-level feature description of a media asset with RDF).***
A video clip with a duration of **one hour** is described with MPEG-7. Several visual low-level features ($\mathfrak{F}$) as colour, shape, texture, etc. are extracted for a number of spatial segments ($S$) per key frame ($K$). A multimedia ontology is then used to represent the MPEG-7 descriptors formally (on basis of RDF); an average number of RDF triples is assumed for each descriptor ($T_D$). An estimation of the resulting RDF graph size then is $\mathfrak{F} \cdot K \cdot S \cdot T_D$. Let us assume that we want to capture 10 features, some 1000 key frames may exist, 10 spatial segments are marked up, and finally 10 triples are required per descriptor. This yields a total RDF graph size of **1 million triples**—just for describing the low-level features of an hour of video footage. □

Finally, even if the above mentioned issue were resolved, another open issue exists: The deployment of multimedia metadata along with the content in the context of the Semantic Web. To the best of our knowledge no proposal exists that addresses performance and scalability, as well as enabling the formal descriptions of the multimedia resources.

## 1.2 Problem Definition

Several issues arise when building Semantic Web multimedia applications; based on a thorough analysis (cf. Chapter 6) we identify three issues to be most significant regarding scalability and expressivity:

- performance and scalability issue in distributed multimedia metadata sources,
- efficient and effective representation of multimedia vocabularies and instances, and
- scaleable multimedia metadata deployment on the Semantic Web.

The following sections describe each of the above listed areas of research in greater detail, and formulate according research questions. The reader is invited to note that although the three selected research areas are not strongly interrelated, they have a recurrent theme: they all focus on both effectiveness and efficiency, hence the name of this thesis—Building scaleable and smart multimedia applications on the Semantic Web. While the three areas may be seen as orthogonal, they address different aspects in the design and implementation of a Semantic Web multimedia applications.

### 1.2.1 Performance and Scalability Issues in Distributed Metadata Sources

A Semantic Web multimedia application needs to process RDF-based metadata stemming from a range of sources. When accessing and processing distributed metadata sources on the RDF-level, the application has to deal with real-world limitations as bandwidth, downtimes, etc.

While from the point of view of a Semantic Web agent it might not be of interest where the triples come from, it may—for the human user who has instructed the agent to carry out a task—well be of interest how long a certain operation takes.

**Research Questions**. What are the characteristics of (multimedia) data sources available on the (Semantic) Web. How can these efficiently be RDFised? Which are practical performance and scalability indicators?

**Scope**. For this problem, we assume that we deal with global descriptions of multimedia assets. For the performance and scalability indicator a static, a simple setup is assumed; it should be evaluate it in an multimedia Web application.

### 1.2.2 Efficient and Effective Representation of Multimedia Metadata

When building Semantic Web multimedia applications, the content being dealt with has to be described appropriately. In order to describe the content appropriately, a language has to meet a range of requirements. It has to be expressive to represent objects, events, and relations. There must be ways to assign descriptions to temporal and spatial segments. The granularity of the content description has to be adjustable. The language has to deal with concrete data types in all its forms (scalars, vectors, and matrices).

To find a tradeoff between expressivity and scalability, several aspects should be taken into account[4]:

[4]For a detailed discussion on these aspects, the reader is invited to refer to Chapter 6.

- The **granularity of the description** usually has an impact on the size of the resulting description. The discriminator here is that of scope: A audio clip might be described global in terms of genre (*this MP3 file is a Jazz clip*) or there might be a detailed description of the wave shape, energy, etc. for a certain time period (*from time code X to time code Y the following parameters have been extracted*: `vector of signal parameters`).
- The required **inferential capabilities** of the system influence the choice of the representation. If no or only simple queries are expected (*return all documents that are mono and less than 1min playtime*), a simple metadata format (such as ID3 for music) might be sufficient. When advanced, and even domain specific retrieval operations are on target (*find me contemplative scenes with at least two people in it*), usually formal-grounded languages as Description Logics or rule-based languages are a good choice.
- The **usage of the content** that may further be differentiated into:
  - number of users (limited group vs. Web-scale)
  - content delivery (streaming, interactive, off-line)
  - metadata deployment (embedded vs. referenced)
  - access mode (broadcast vs. point-to-point)
  - read-only vs. read/write
  - personalisation of content

The reader is invited to note that no single language currently covers all the above mentioned aspects. Where, e.g., MPEG-7 is a good choice for representing low-level features, it fails supporting the engineer in the modelling of high-level semantics. On the other hand, for example OWL is quite expressive but lacks built-ins for complex concrete data types (as matrices), temporal descriptions, and support for multimedia description in general[5].

**Research Questions**. How can (formal) descriptions of multimedia content be represented effectively and efficiently? Is there a trade-off between scalability and expressivity and if yes, where?

**Scope**. A closed-world scenario is assumed; we focus on spatio-temporal descriptions of audio-visual material. Common multimedia metadata formats such as MPEG-7 should be taken into account.

### 1.2.3 Scaleable Multimedia Metadata Deployment on the Semantic Web

Many multimedia metadata formats, such as Exif or MPEG-7 are available to describe what a multimedia asset is about, who has produced it, etc. With the advent of User Generated Content—be it blogs, Wikis, etc.—a need for deploying these M3 formats in (X)HTML pages can be identified. Another motivation stems from the professional content realm. There, detailed descriptions of cross-media content is on target, along with rights-management.

Again, in the context of building Semantic Web multimedia applications, one key question regarding the deployment of the metadata is how to enable existing multimedia metadata formats to enter the Semantic Web in order to make them accessible to Semantic Web agents capable of handling RDF-based metadata.

[5] We note that the ongoing work regarding OWL 2 have not been taken into consideration; see also http://tw.rpi.edu/weblog/2008/04/16/towards-rdfs-30-or-owl-2-r-full

**Research Questions**. How can existing multimedia metadata formats be deployed effectively and efficiently on the Semantic Web? What are the use cases?

**Scope**. It is assumed that reusability of exiting material should be maximised. We assume a prototypical implementation as sufficient as a proof of concept. The deployment description should be available as an vocabulary.

## 1.3 Reader's Guide

The thesis at hand is roughly structured into five parts as shown in Fig. 1.1.

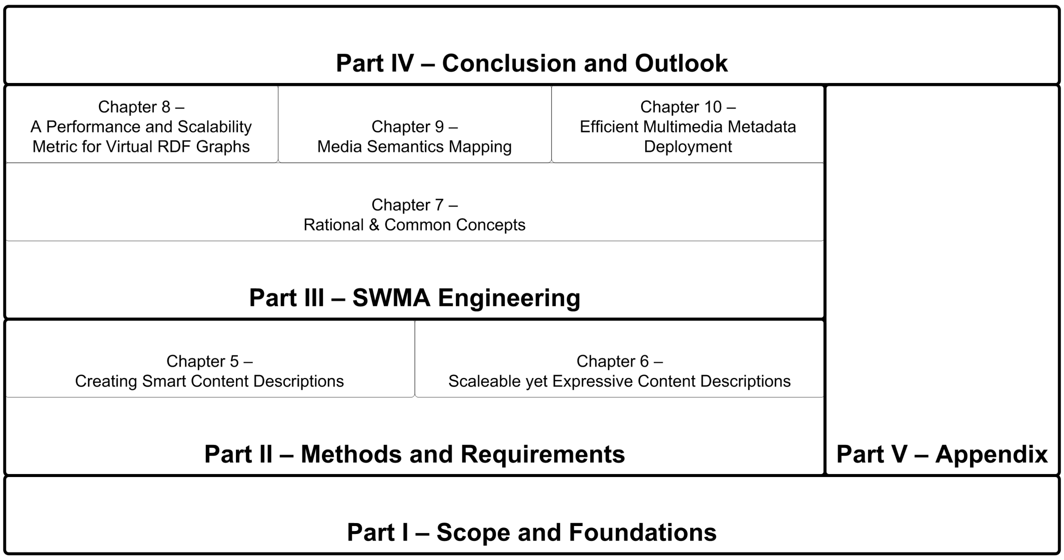



Figure 1.1: A visual guide through the thesis.

- Part I introduces the foundations and lists existing and related work;
- Part II discusses methods and requirements regarding scaleable, yet expressive multimedia content descriptions;
- Part III addresses the three core issues of engineering Semantic Web multimedia applications as of the problem definition;
- Part IV contains conclusions and contemplates about future directions regardingw.r.t. Semantic Web multimedia applications;
- Part V (Appendix) gathers sources and the author's contributions.

Readers familiar with both multimedia metadata and equipped with knowledge of the Semantic Web (technologies) may choose to skip Part I and directly start with Part II. The core of the thesis is in Part III, as it addresses the research question given earlier in this chapter (cf. section 1.2).

Note that a detailed explanation of the research this thesis is built upon is given later in section B.1 (Appendix B). This work was accompanied by the author's activities within

W3C. The author was active in the first Multimedia Semantics Incubator Group (MMSEM-XG) in 2006/2007. Further, the author has been active in the Semantic Web Deployment Working Group (ongoing) focusing on the RDFa specification[6], more specially on the use cases, test cases and the implementation report. Finally, the author has been active in the LinkingOpenData[7] project, realising the formalisation and interlinking of statistical data [125] and proposing a new interlinking method [144; 143].

In the following a detailed reader's guide—on the chapter level—is given:

**Part I–Scope and Foundations** Introduces the foundations and sums up existing work. The goal is to make the reader familiar with the problem domain.

- **Introduction** Gives a motivation and defines the research questions.
- **Related and Existing Work** Related work is discussed and critically reviewed.
- **Multimedia Metadata** Foundations of multimedial metadata (M3) are explained.
- **Semantic Web** Semantic Web basics are explained.

**Part II–Methods and Requirements** Constitutes theoretical elaborations on scaleable yet expressive multimedia content descriptions.

- **Creating Smart Content Descriptions** Elaborates on how multimedia content descriptions are created (from extraction to ontology engineering).
- **Scaleable yet Expressive Content Descriptions** Introduces requirements for scaleable yet expressive multimedia content descriptions.

**Part III–SWMA Engineering** Addresses three core issues of engineering Semantic Web Multimedia applications.

- **Rational & Common Concepts** Lists basic design principles and defines common concepts.
- **A Performance and Scalability Metric for Virtual RDF Graphs** Addresses issues w.r.t. the access of distributed metadata sources.
- **Media Semantics Mapping** Addresses issues regarding the Semantic Gap in media descriptions.
- **Efficient Multimedia Metadata Deployment** Addresses multimedia metadata deployment issues.

**Part IV–Conclusion and Outlook** Discusses lessons learned and future directions.

- **Concluding Remarks** The work is reviewed and discussed.
- **Outlook** A number of possible developments regarding SWMA is presented.

**Part V–Appendix** Gathers sources and author's contributions.

- **Sources** Lists sources of RDF graphs and applications in the context of this work.
- **Author's Contribution** Lists the author's contributions in the realm of this thesis.
- **Reference Material** Offers a collection of good practice material for SWMA.
- **Glossary** Gives a short explanation of terms used in this work.

[6] http://www.w3.org/TR/xhtml-rdfa-primer/

[7] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

## 1.4 What this work is NOT about

This thesis does *not attempt to define a solution for a semantic description of multimedia content*. Defining such a formal specification, i.e., an ontology, would contradict with the genuine idea behind **ontologies**: to be based on an agreement of domain experts. The reader is invited to note the plural form; it is rather due to the fact that ontologies are based on a *shared* understanding of a domain than to the circumstance that the author is not able to or willing to perform such a task.

Not in the scope of this thesis are multimedia *content* issues, such as compression, codecs, etc. Further, issues as data access and delivery (caching, streaming, broadband, etc.) or access control-issues, such as ACLs, etc., are not in the primary scope of the thesis. However, we refer to such issues if they have a significant impact on the issues discussed earlier, i.e., issues that **are** in the scope of the work.

Chapter 2

# Related and Existing Work

*"In statu nascendi."*

(Latin phrase)

The title of this work—Building Scaleable and Smart Multimedia Applications on the Semantic Web—contains terms, which have to be clarified, and put into context before one is able to go into greater detail. For quite a lot of these terms, no general definition is available. Were appropriate, such a definition in the context of the thesis is given. Hence, this chapter discusses the following terms, along with their interpretation in the context of the work at hand:

- "Semantic Web Applications", cf. section 2.1
- "Multimedia Applications", cf. section 2.2
- "Scalability and Expressivity", cf. section 2.3

For each of the phrases an explanation is given, relevant existing and related work is discussed. Where applicable, research projects are listed exemplary—some of them the author has participated in.

As an aside, it is worth noting that each technology undergoes certain phases ranging from foundational academic research to practical exploitation. Semantic Web technologies are—as time of writing of this thesis—according to Gartner's Hype Cycle[1] in the so called "Technology Trigger" phase. This first phase of a Hype Cycle is the breakthrough, product launch or other event that generates significant press and interest. While from the infrastructural point of view a lot work has already been done (annotations, languages, services, etc.), practical aspects as for example scalability of metadata have not been widely addressed. However, a range of activities can be noticed in this field, be it grass-root-like or educational and outreach activities.

[1] *A Hype Cycle is a graphic representation of the maturity, adoption and business application of specific technologies.*, http://www.gartner.com

## 2.1 Semantic Web Applications

In 2007, the Semantic Web Challenge[2] (SWC) is being held the fifth time in a row. The SWC—an event for demonstrating practical progress towards achieving the vision of the Semantic Web—is organized in conjunction with the *International Semantic Web Conference* (ISWC). Several purposes are served, namely (i) the SWC enables to illustrate to society what the Semantic Web can provide, (ii) gives researchers an opportunity to showcase their work and compare it to others, and (iii) stimulates current research to a higher final goal by showing the state-of-the-art every year.

To ensure a certain level of comparability, the SWC has listed a number of minimal requirements, a Semantic Web applications must meet in order to be able to participate in the challenge. These criteria are outlined and discussed in the following.

1. The **meaning of data** has to play a central role.
   - Meaning must be represented using formal descriptions,
   - Data must be manipulated/processed in interesting ways to derive useful information, and
   - This semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all.
2. The **information sources** ...
   - Should have diverse ownerships (i.e. there is no control of evolution),
   - Should be heterogeneous (syntactically, structurally, and semantically), and
   - Should contain real world data, i.e. are more than toy examples.
3. It is required that all applications **assume an open world**, i.e. assume that the information is never complete.

Discussing the above listed criteria, we note the following w.r.t. the scope of this work.

Regarding the "meaning of data": Though **formal** is formulated quite liberal, in the context of the Semantic Web the languages of choice are somehow limited to being RDF-based, such as OWL and the like.

Regarding the "information sources": Firstly, the requirement that the sources need to have diverse ownerships is obviously needed to be able to demonstrate the **Web** characteristic; cf. Definition 2.2. Secondly, asking for real world data rather than for constructed, limited toy examples supports the very issue of this thesis.

Regarding the "open world assumption": Due to the Web-scale reasoning process, this is a non-trivial issue; recently Fensel and van Harmelen [98] elaborated on that issue.

We subscribe to the above stated view on the requirements for Semantic Web applications, and additionally point out that a Semantic Web application is a **Web application**, after all. The lessons learned in this area should be taken into account, as well. Well-known infrastructure, processes, and methodologies [3] for handling content and metadata should be utilised. Consequently, before we give a definition of what is to be understood by a Semantic Web application, we define Web application as follows[4].

[2] http://challenge.semanticweb.org/

[3] e.g. http://java.sun.com/blueprints/guidelines

[4] See also http://www.whatwg.org/specs/web-apps/current-work/ for characteristics of Web applications.

**Definition 2.1 (Web Application).**
A Web Application is a software program that meets following minimal requirements:

- It is based on the HyperText Transfer Protocol **HTTP** [169] and Uniform Resource Identifiers **URI** [33; 60][5];
- For human agents, the primary presentation format is the Hypertext Markup Language (X)HTML[6] [330];
- For software agents, the primary interface is REST-compliant [100] or may be based on Web services[7]—cf. **SOAP** [276], **WSDL** [55], and **UDDI** [301];
- The application operates on the Internet;
- The number of (concurrent) users is undetermined. ❖

Note that where *primary* is used in Definition 2.1, it is possible and likely that other rendering formats (such as PDF[8]) or protocols (for example XMPP[9]) may as well be offered by a Web application in addition to the ones mentioned. Note as well that the last characteristic both effects the scalability and performance of a Web application.

In the next step we give—based on Definition 2.1 and motivated by the requirements of the Semantic Web Challenge—a definition of a Semantic Web application.

**Definition 2.2 (Semantic Web Application).**
A Semantic Web Application is a Web application that additionally to the requirements listed in Definition 2.1, meets the following minimal requirements:

- The metadata (metadata sets) used in the Web application must be machine readable and machine interpretable[10], i.e, it is based on the Resource Description Framework **RDF** [203][11];
- A set of formal vocabularies—potentially based on **OWL** [239]—is used to capture the domain of discourse[12]; at least one of the utilised vocabularies and/or metadata sets has to be proven **not** to be under (full) control of the Semantic Web application maintainer;
- SPARQL [253] should be used for querying, and RIF[260] may be utilised for exchanging rules. ❖

The restriction that a (Semantic) Web application is expected to operate on the Internet is to ensure that Intranet—or for the sake of correctness: Intraweb—applications *utilising* (Semantic) Web technologies are **not** understood as (Semantic) Web applications in the narrower sense per se. The reader is invited to note that this requirement is a matter of the control over the data and the schemas rather than a question of the sheer size of the deployed application.

---

[5] See section 4.3.1 for details
[6] http://www.w3.org/html/wg/
[7] http://www.w3.org/2002/ws/
[8] http://www.adobe.com/devnet/pdf/pdf_reference.html
[9] http://www.xmpp.org/rfcs/
[10] For a discussion on this issue the reader is invited to refer to [312, Section 1.1]
[11] Section 4.3.3
[12] See section 4.3.4 for details

Another very important aspect of Semantic Web Applications was paraphrased by Ora Lassila in his keynote at the Scandinavian Conference on AI (SCAI) 2006 [193]:

> Any specific problem (typically) has a specific solution that does not require Semantic Web technologies.
>
> Q: Why then is the Semantic Web so attractive?
> A: For future-proofing. Semantic Web can be a solution to those problems and situations that we are yet to define.

It was also Lassila who coined the term "serendipity"; serendipity in **interoperability** (how to interoperate with systems we knew nothing about at design time?), serendipity in information **reuse** (accessible semantics of the information), and serendipity in information **integration** (can information from independent sources be combined?).

As an exemplary Semantic Web application[13], we discuss `mle`, the mailing list explorer [148] in the following[14].

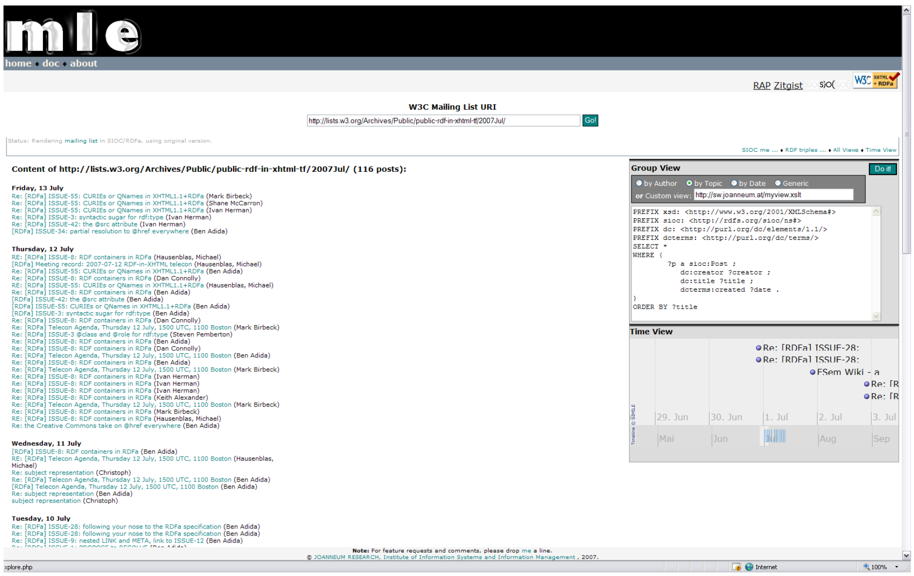



Figure 2.1: An examplary Semantic Web application.

Following and understanding discussions on mailing lists is a prevalent task for executives and policy makers in order to get an impression of one's company image. However, existing solutions providing a Web-based archive require substantial manual effort to search for or filter certain information. With `mle` (cf. Fig. 2.1) we propose a new way to automatically process mailing list archives. The tool is realised based on two Semantic Web technologies: Firstly, SIOC (Section 4.4.2) is utilised as the primary vocabulary for describing posts, people, and topics; secondly the RDF metadata is deployed by means of embedding it in the Web page encoded in XHTML+RDFa (see section 4.6.2).

[13] Demonstrated at the Semantic Web Challenge 2007

[14] The application is available at `http://sw.joanneum.at/mle/xplore.php`

### 2.1.1 Projects & Activities

In the following, projects and activities dealing with how to build, enhance, or utilise Semantic Web applications are discussed. The discipline of Semantic Web application building is a quite new one[15], hence selecting appropriate tools is an elaborate task[16].

A good starting point for applications of ontologies is the *Handbook on Ontologies* [279]. Fensel et. al. [97] describe areas for application of the Semantic Web, focusing on knowledge management and electronic commerce.

At the time of writing, the W3C *Semantic Web Education and Outreach Interest Group*[17] seeks to develop strategies and materials to increase awareness among the Web community of the need and benefit for the Semantic Web, and educate the Web community regarding related solutions and technologies.

The following is a short overview on prominent (research) projects operating in the Semantic Web application domain:

- **SWAD-Europe** - **S**emantic **W**eb **A**dvanced **D**evelopment in Europe,
  Projects goals from http://www.w3.org/2001/sw/Europe/:

  > The SWAD-Europe project aims to support W3C's Semantic Web initiative in Europe, providing targeted research, demonstrations and outreach to ensure Semantic Web technologies move into the mainstream of networked computing. The project aims to support the development and deployment of W3C Semantic Web specifications through implementation, research and testing activities. Semantic Web Advanced Development for Europe (SWAD-Europe) aims to play a key role in the evolution of the Semantic Web, through education and outreach to developers, organisations and content creators; through Open Source implementation and testing, and through pre-consensus technology development to drive and inform the creation of new Semantic Web standards.

- **Knowledge Web**
  Mission from http://knowledgeweb.semanticweb.org/:

  > The mission of Knowledge Web is to strengthen the European industry and service providers in one of the most important areas of current computer technology: Semantic Web enabled E-work and E-commerce. The project concentrates its efforts around the outreach of this technology to industry. Naturally, this includes education and research efforts to ensure the durability of impact and support of industry.

- **SIMILE** - **S**emantic **I**nteroperability of **M**etadata and **I**nformation in un**L**ike **E**nvironments,
  Due to http://simile.mit.edu/, SIMILE ...

[15] To coordinate so called Semantic Web engineers, the author of this thesis founded a social site dedicated to the exchange of this issue. This social network is open for subscription to everyone interested in this area; it can be found at http://semanticwebengineers.crowdvine.com/

[16] The interested read is referred to a repository at W3C, giving an overview on tools and environments: http://esw.w3.org/topic/SemanticWebTools

[17] http://esw.w3.org/topic/SweoIG

> ... seeks to enhance inter-operability among digital assets, schemata / vocabularies / ontologies, metadata, and services. A key challenge is that the collections which must inter-operate are often distributed across individual, community, and institutional stores. We seek to be able to provide end-user services by drawing upon the assets, schemata/vocabularies/ontologies, and metadata held in such stores.
>
> [...] The project also aims to implement a digital asset dissemination architecture based upon web standards. The dissemination architecture will provide a mechanism to add useful ”views” to a particular digital artifact (i.e. asset, schema, or metadata instance), and bind those views to consuming services.

- **CAS** - **C**S **AK**Tive **S**pace,
  http://www.aktors.org/technologies/csaktivespace/ states that:

  > CAS is an integrated Semantic Web application which provides a way to explore the UK Computer Science Research domain across multiple dimensions for multiple stakeholders, from funding agencies to individual researchers.

- **SIOC** - **S**emantically-**I**nterlinked **O**nline **C**ommunities
  The homepage http://sioc-project.org/ states that SIOC

  > provides methods for interconnecting discussion methods such as blogs, forums and mailing lists to each other. It consists of the SIOC ontology, an open-standard machine readable format for expressing the information contained both explicitly and implicitly in internet discussion methods, of SIOC metadata producers for a number of popular blogging platforms and content management systems, and of storage and browsing/searching systems for leveraging this SIOC data.

  The reader is invited to note that SIOC has been submitted to W3C for standardisation and—given its widespread use—is likely to become a recommendation, soon.

- **Semantic Web Search Engines**
  Although the available amount of RDF-based data on the Web is rather limited compared to the overall size of the Web, already dedicated search engines and indexer are available[18]. Typically, they operate on the triple level, i.e., indexing triple along with their provenance information. Among those SE, the following are the ones widely used:

  - Sindice (DERI): http://sindice.com/—a semantic indexer;
  - Falcon (ISW China): http://iws.seu.edu.cn/services/falcons/—a Semantic Web search engine;
  - Swoogle (UMBC): http://swoogle.umbc.edu/—a mature, hybrid SE with some 6 million URIs indexed;
  - Zitgist LLC's PTSW: http://pingthesemanticweb.com/—a web service archiving the location of recently updated RDF documents on the Web.

[18] http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines

## 2.2 Multimedia Applications

In this section we first explain the term *smart multimedia content* in the context of this work. Then, multimedia metadata deployment issue are discussed. Further, the notion of a *Multimedia Application on the Semantic Web* is defined. Finally, this section concludes with a review of recent projects and activities illustrating the current status of *Semantic Multimedia Applications*[177; 152].

### 2.2.1 Smart Multimedia Content

In the area of smart multimedia content handling [289], the past few years of research have produced a notable output. This section summarizes known approaches to bridge the *Semantic Gap*, and highlights activities in this area. For a comprehensive discussion on the state-of-the-art, and a proposal for a research agenda including open research issues in the development of the Semantic Web from the perspective of hypermedia research, the reader is referred to [311].

An array of research is available [129; 84; 336; 116; 334] dealing with the Semantic Gap. However, a generic, domain-independent solution to the problem is not at hand. An example instantiation for the Semantic Gap is depicted in Fig. 2.2 and described in the following example, respectively.

***Example 2.1 (Semantic Gap Example).***
Take for example some visual content source—such as a video clip—depicting a soccer ball. Further, assume we have low-level features, say, colour and shape, which describe the content throughout. This setup is shown in Fig. 2.2.

The question now is if and how it is possible to map the two available low-level features `shape=circular` and `colour={black, white}`, occurring in a certain region, to the (logical) concept *soccer ball*. If it is possible, then the question is under which circumstances it can be realised. The ultimate goal would be to handle the generic case, i.e., without any further knowledge about the domain. □

Starting with a critical review of both MPEG-7 and OWL—as well as an analysis of their respective interoperability—in [310] and [229] we can state that there is a certain degree of freedom in choosing some markup for a certain task. From a methodological point of view there exist a number of approaches that may be taken to realise smart media content descriptions[19], reviewed in the following.

- The *purist approach*, where either a metadata format as MPEG-7 or logic-based approach (e.g. an description based on OWL) is assumed to fulfil the task. An example for a logic language extended for multimedia retrieval can be found in [209];

- The *integration approach*, that tries to embed or translate (parts of) one language into the other—a prominent exponent is [171]; a recent example can be found in [107];

- The *layer* approach, also known as the "principle of subsidiarity", where each vocabulary is used in the appropriate realm. For example [290] is a promising research work that represents this approach. Related work can also be found in [299];

[19]The reader is invited to refer to Chapter 5 for a detailed discussion on this issue.

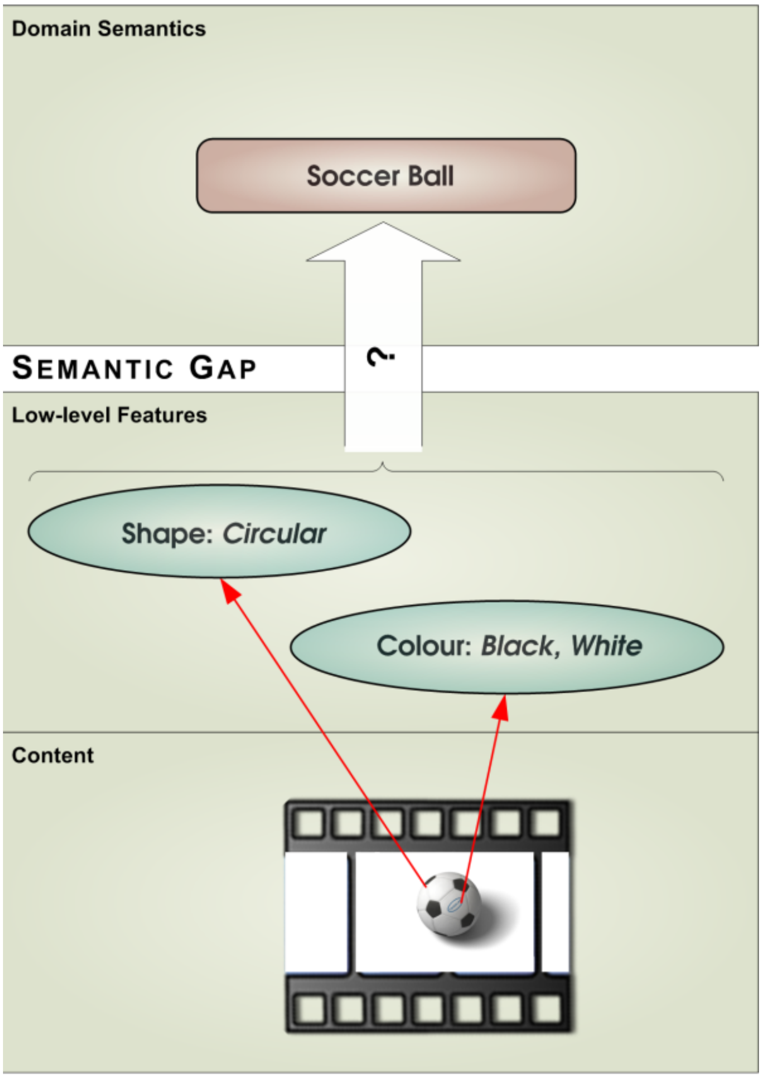



Figure 2.2: The Semantic Gap in Multimedia Content Description.

- Finally it is possible to *invent a new vocabulary*—such as proposed in [170], and independently in [12]—which is not advisable in terms of interoperability[20].

**Standardisation**. Since 2005 considerable work has been carried out in the World Wide Web Consortium (W3C) [317] on multimedia content description and understanding The Multimedia Task Force of the Semantic Web Best Practices and Deployment Working Group[21] has elaborated on multimedia markup for a while. In early 2006 another W3C activity has been launched,the Multimedia Semantics Incubator Group [315]. Its mission was to ...

> ... show metadata interoperability can be achieved by using the Semantic Web technologies to integrate existing multimedia metadata standards. Thus, the goal of the XG is NOT to invent new multimedia metadata formats, but to leverage and combine existing approaches

The author of this thesis has been active since and contributed to various deliverables, such as the Incubator Group report "Multimedia Vocabularies on the Semantic Web" [141].

The scope of the Moving Picture Experts Group (MPEG)[22] has been extended recently from only signal coding to multimedia metadata, processes and applications. The "Multimedia Content Description Interface" (MPEG-7) standard [222; 220; 221; 218] specifies the description of multimedia content, integrating content structure (e.g. shots of video, regions of image), low-level visual and audio features and high-level descriptions (e.g. production information). The high-level descriptors allow linking external thesauri or knowledge bases and thus the integration of media oriented content descriptions with Semantic Web technologies. MPEG-7 profiles have been proposed as subsets for certain application areas to reduce the interoperability problem caused by the comprehensiveness and generality of the MPEG-7 standards. The Detailed Audiovisual Profile (DAVP) [19], the first MPEG 7 profile with formal semantics of the description elements (in order to solve the interoperability problem), has been developed with contributions from the author of this thesis[23].

While standardisation efforts are emerging and already produce first substantial results, the term "smart multimedia content" is still not uniformly defined. Even in marketing slang the term has been (mis)used. To avoid confusion what is meant by smart media content, we define it in the context of this work as follows:

**Definition 2.3 (Smart Multimedia Content—SMC).**
Smart Multimedia Content is multimedia content along with metadata enabling interoperable and advanced operations across systems. Typically SMC has two characteristics:

- The metadata can be available both in terms of low-level features, as well as formal domain descriptions;
- It is self-descriptive (see also the "The Self-Describing Web"[24]). ❖

The reader is invited to note that the definition of SMC is deliberately kept quite vague. This is due to the nature of SMC. Many forms of SMC may exist, and many technologies may be utilised to realise SMC. Hence, the above definition can be seen a least common denominator. Notable efforts regarding SMC have been reported from the mobile devices area and from ubiquitous computing.

[20] See also http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability/
[21] http://www.w3.org/2001/sw/BestPractices/MM
[22] http://www.mpeg.org/
[23] http://mpeg-7.joanneum.at
[24] http://www.w3.org/2001/tag/doc/selfDescribingDocuments.html

### 2.2.2 Multimedia Metadata Deployment

To the best of our knowledge research regarding multimedia metadata deployment has not been widely performed. Current approaches are either not specific to multimedia or do not scale to the size of the Web. We will discuss available proposals and highlight issues with them in the following.

The W3C's Protocol for Web Description Resources (POWDER) Working Group currently works on a very powerful, but rather generic standard[25] to facilitate the publication of descriptions of multiple resources such as all those available from a Web site.

Adobe's Extensible Metadata Platform (XMP)[26], which is primarily used for PDF documents (but also usable with other formats, such as JPEG, PNG, etc.), shares some of our objectives. The Open Archives Initiative (OAI) has published the Compound Information Objects draft [191], a specification dealing with the publication of aggregations of distinct information units.

The Upper Mapping and Binding Exchange Layer (UMBEL) specification[27] is a high-level subject layer for mapping various ontologies with simple binding mechanisms for any structured formalism. Simply stated, UMBEL is both a high-level reference "bag of subjects" and light-weight mechanisms for binding to Web ontologies via proxies for those subjects. For linked datasets, the semantic sitemaps extension [65] is available, providing basic deployment descriptions on the data access level. We are currently working on a proposal labelled voiD—"Vocabulary of Interlinked Datasets"[28] allowing the description of linked datasets on the content-level.

However, the main issue with multimedia metadata deployment is the so called "last mile".

**The Last Mile**. For communications provider, the last mile is the final leg of delivering connectivity to a customer. Equally, in business the last mile is used to describe the process of getting any deliverable to the final consumer. In the case of multimedia metadata deployment, the last mile is the delivery of multimedia metadata to the end-user, i.e., a Semantic Web agent able to "understand" RDF. In this work, multimedia metadata deployment has been assigned a precise meaning:

**Definition 2.4 (Multimedia Metadata Deployment).**
Multimedia Metadata Deployment is the packaging and the delivery of the metadata along with an multimedia asset. Regarding the Semantic Web at least two requirements need to be fulfilled:

- The data model of the deployment has to be RDF;
- While typical container formats such as SMIL, SVG, and PDF may be used, there has to been at least one deployment path that works with (X)HTML. ❖

In Fig. 2.3 the basic setup regarding multimedia deployment on the Semantic Web is given. Typically a couple of distinct players is involved, namely (i) human users consuming a Web page and its embedded media objects, (ii) Semantic Web agents consuming the RDF-based metadata describing the media objects, further (iii) Semantic Web languages and ontologies, and finally (iv) multimedia metadata formats.

[25] http://www.w3.org/TR/powder-grouping/
[26] http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf
[27] http://www.umbel.org/proposal.xhtml
[28] http://community.linkeddata.org/MediaWiki/index.php?VoiD

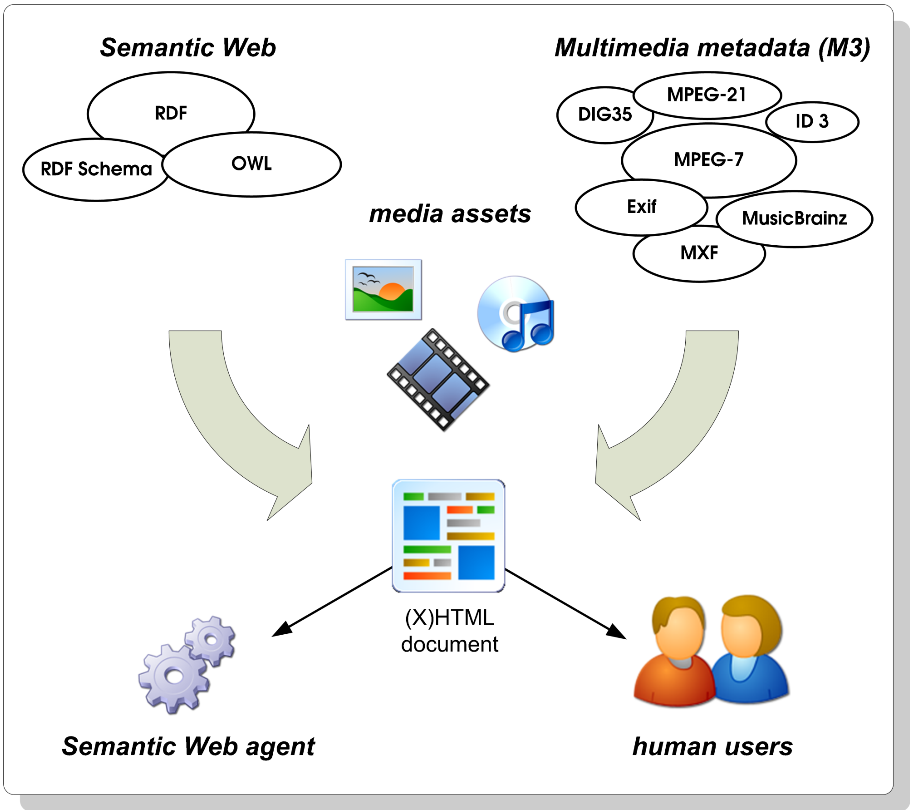



Figure 2.3: Deploying multimedia formats on the Semantic Web.

### 2.2.3 Semantic Web Multimedia Applications—SWMA

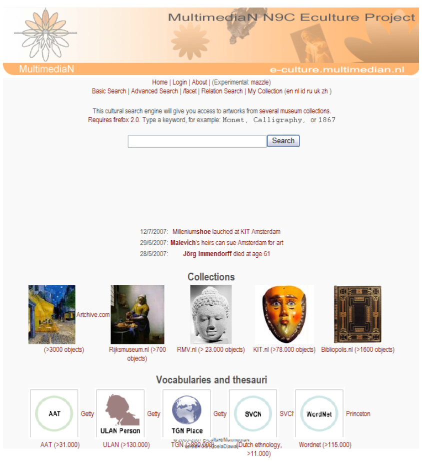
Two kinds of applications can roughly be distinguished, namely (i) Web applications and (ii) multimedia applications. As stated above, Semantic Web applications are a subset of Web applications. In the intersection of Semantic Web Applications, and Multimedia Applications we finally find the so called "Semantic Web Multimedia Applications" (SWMA), as depicted in Fig. 2.5.

Based on the definitions 2.2 (cf. page 13) and 2.3 (cf. page 19), it is now possible to define:

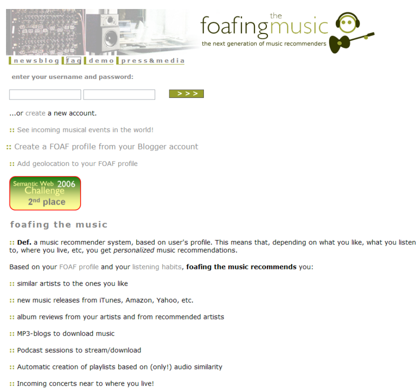
**Definition 2.5 (Semantic Web Multimedia Application—SWMA).**
A Semantic Web Multimedia Application (SWMA) is a Semantic Web Application (cf. Def. 2.2) dealing with smart multimedia content (cf. Def. 2.3). The SWMA may support sharing, creating, manipulating, or delivering of the multimedia content; at least one of the following characteristics applies:

- The application deals with spatio-temporal issues w.r.t. the content description;
- The application deals with the *Semantic Gap*. ❖

State-of-the-Art examples of Semantic Web multimedia applications are DBTune.org[29], the Podcast Pinpointer[30], the MultimediaN E-Culture demo[31], and FOAFing-the-Music[32]. The latter two are depicted in the Fig. 2.4.




**(a)** MultimediaN E-Culture Demo.




**(b)** FOAFing-the-Music.

Figure 2.4: Examples of Semantic Web Multimedia Applications.

While the MultimediaN E-Culture Demo (Fig. 2.4(a)) focuses on artworks supporting a range of vocabularies (AAT, ULAN, WordNet, etc.), FOAFing-the-Music (Fig. 2.4(b)) is a semantic music recommender system, based on a user's FOAF profile. Both SWMA's have successfully taken part in the 2006 Semantic Web challenge (first and second prize).

[29] http://dbtune.org/
[30] http://citeseer.ist.psu.edu/hogan05podcast.html
[31] http://e-culture.multimedian.nl/demo/search
[32] http://foafing-the-music.iua.upf.edu/

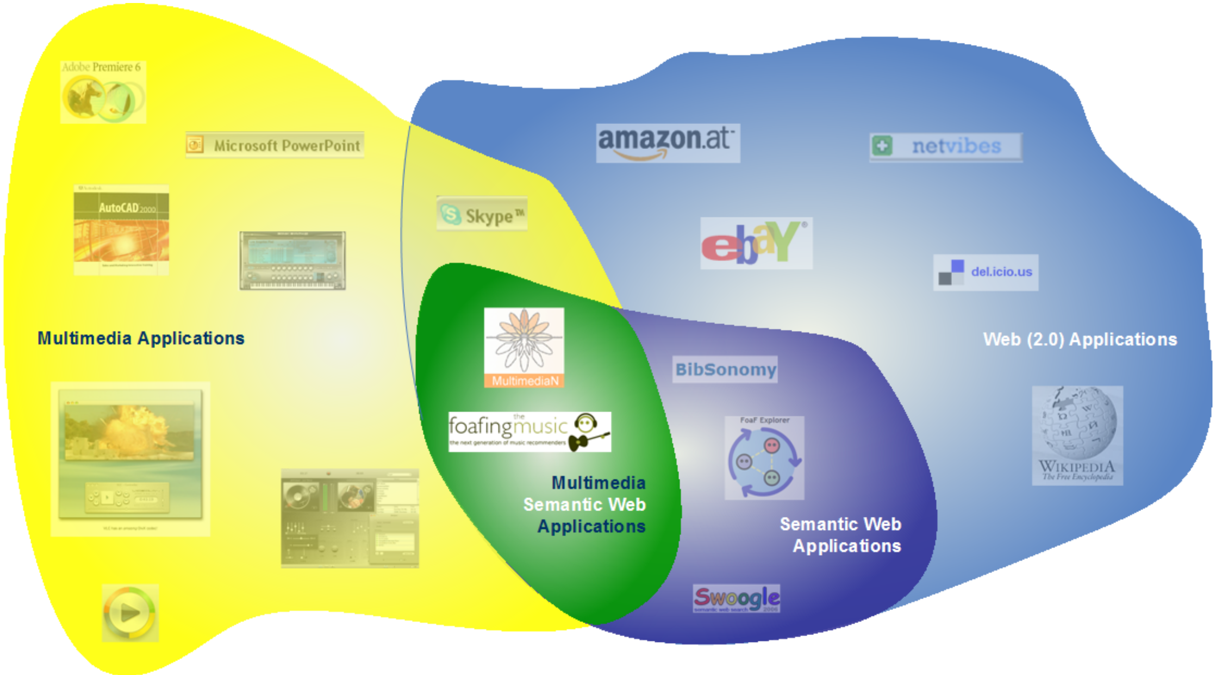



Figure 2.5: Semantic Web Multimedia Applications (SWMA).

### 2.2.4 Projects & Activities

A range of projects (EU-funded research projects, national programmes, and international project) focuses on smart media; a selective overview is given in the following. **Note**: In projects marked with ▷ the author of this work has been or still is active.

- **aceMedia** - http://www.acemedia.org/aceMedia
  Citing [189]:

  > [...] an approach for knowledge and context-assisted content analysis and reasoning based on a multimedia ontology infrastructure is presented. [...] In aceMedia, ontologies will be extended and enriched to include lowlevel audiovisual features, descriptors and behavioural models in order to support automatic content annotation. This approach is part of an integrated framework consisting of: user-oriented design, knowledge-driven content processing and distributed system architecture. The overall objective of aceMedia is the implementation of a novel concept for unified media representation: the **Autonomous Content Entity** (ACE), which has three layers: content, its associated metadata, and an intelligence layer.

- ▷ **K-Space** **K**nowledge **Space** of Semantic inference for automatic annotation and retrieval of multimedia content,
  From http://kspace.qmul.net/ we learn that ...

  > K-Space integrates leading European research teams to create a Network of Excellence in semantic inference for semi-automatic annotation and retrieval of multimedia content. The aim is to narrow the gap between content descriptors that can be computed automatically by current machines and algorithms, and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media: *The Semantic Gap*.

- **MUSCLE** **M**ultimedia **U**nderstanding through **S**emantics, **C**omputation and **Le**arning, http://www.muscle-noe.org/ describes the goals as follows:

  > MUSCLE aims at creating and supporting a pan-European Network of Excellence to foster close collaboration between research groups in multimedia datamining on the one hand, and machine learning on the other [...]

- ▷ **NM2** **N**ew **M**illennium, **N**ew **M**edia
  The project homepage, http://www.ist-nm2.org/, says:

  > NM2 unites leading media and technology experts from across Europe to develop compelling new media genres, which utilise the unique characteristics of broadband networks. The project is creating new production tools for the media industry that allow the easy production of interactive non-linear broadband media genres, which can be personalised to suit the preferences of the individual viewer. Viewers are able to interact directly with the medium and influence what they see and hear according to their personal choices and tastes.

- ▷ **SALERO** - **S**emantic **A**udiovisua**L** **E**ntertainment **R**eusable **O**bjects
  At http://www.salero.info/ it is claimed that ...

> SALERO aims at making cross media-production for games, movies and broadcast faster, better and cheaper by combining computer graphics, language technology, semantic web technologies as well as content based search and retrieval.

- **REVEAL THIS**[33] - **RE**trieval of **V**id**E**o **A**nd **L**anguage for **T**he **H**ome user in an **I**nformation **S**ociety
The project homepage—http://www.reveal-this.org/—explains the scope as follows:

> REVEAL THIS addresses a basic need underlying content organisation, filtering, consumption and enjoyment by developing content processing systems that will help European citizens keep up with the explosion of digital content scattered over different platforms (radio, TV, World Wide Web, etc), different media (speech, text, image, video) and different languages. People should be spending most of their leisure time enjoying the content, not searching for it. REVEAL THIS aims at developing content processing technology able to capture, semantically index, categorise and cross-link multiplatform, multimedia and multilingual digital content, as well as provide the system user with semantic search, retrieval, summarisation and translation functionalities.

- **FilmEd**
[266] and http://metadata.net/filmed/ state:

> The FilmEd project's original aim was to provide the tertiary education sector with broadband access to high quality and unique film and video content stored within Australian moving image archives to enhance curriculum based programs concerned with screen literacy, film and media studies, journalism and Australian culture and history. A prototype called Vannotea has been developed which enables the collaborative indexing, annotation and discussion of audiovisual content over high bandwidth networks. It enables geographically distributed groups connected across broadband networks (GrangeNet) to perform real time collaborative sharing indexing, discussion and annotation of high quality digital film/video and images (and shortly 3D objects).

- **MAENAD** - **M**ultimedia **A**ccess across **E**nterprise **N**etworks **a**nd **D**omains
Quoting http://www.dstc.edu.au/Research/maenad-ov.html:

> The objectives of this project are to develop an underlying data model, metadata mapping schemas (RDF, XML), metadata generators, metadata repositories, query languages, search interfaces and search engines which can provide solutions to the problems of resource discovery, preservation, delivery and management. Resource Discovery of single-medium atomic digital objects has advanced in the past 5 years due to the development of metadata standards such as Dublin Core which provides semantic interoperability for textual documents and MPEG-7 which will provide the same

[33] In 2006, a workshop in the framework of the LREC2006 conference was organized by REVEAL THIS in Genoa, Italy. With the authors contribution, [256] was presented at the Crossing Media for Improved Information Access workshop.

> for audio, video and audiovisual documents. However the future will lead to many more compound multimedia documents on the web which combine text, image, audio and video in rich complex structured documents in which temporal, spatial, structural and semantic relationships exist between the components. The problems of indexing, archiving, searching, browsing, retrieving and managing these kinds of structured dynamic documents are infinitely more complex than the resource discovery of simple atomic textual documents.

## 2.3 Scalability and Expressivity

A good starting point for discussing scalability and expressivity issues is the *OWL Use Cases and Requirements* document [242], which states:

> Expressivity determines what can be said in the language, and thus determines its inferential power and what reasoning capabilities should be expected in systems that fully implement it. An expressive language contains a rich set of primitives that allow a wide variety of knowledge to be formalized. A language with too little expressivity will provide too few reasoning opportunities to be of much use and may not provide any contribution over existing languages.

**Expressivity**. A good place to start the discussion on expressivity is the work of Levesque and Brachman [198]. They examine computational limits on automated reasoning and its effect on knowledge representation. The conclusion they draw is that there exists a tradeoff between the expressiveness of a representational language and its computational tractability: When one limits what can be in a knowledge base its implication are more manageable computationally. Restricting the logical form of a knowledge base can lead to very specialized forms of inference. Well-known and practical relevant forms are: the relational (database) form [258], Description Logics (cf. section 4.1.2), and the logic-program form (cf. section 4.1.3),

Dixon et. al. [79] have discussed issues associated with systems evolution in decentralised organisations. They have proposed a five layer model of information expressivity that provides a theoretical framework for classifying the system variants.

James Hendler sketched the main motivation for expressivity in [153]:

> However, I argue that semantic web techniques can, and must, go much further. The first use of ontologies on the web for this purpose is pretty straightforward—by creating the service advertisements in an ontological language, tools could use the hierarchy (and property restrictions) to find matches via the class/subclass properties or other semantic links. For example, someone looking to buy roses might find florists (who sell flowers) even if there were no exact match that served the purpose. Using, for example, description logic (or other inferential means), the user could even find categorizations that werent explicit. So, for example, specifying a search for animals that were of ”size = small” and ”type = friendly,” the user could end up finding the Pet Shop Mary is working for, which happens to specialize in hamsters and gerbils.

[200] has investigated the expressive power and parsing complexity of the a formalism originally designed for displaying formal propositions and proofs in natural language, the

so called Grammatical Framework. Recently a survey of the usage of ontology languages and their expressivity has been performed [321].

**Scalability**. While expressivity is defined relatively sharp, the term scalability is used to address a range of issues in different domains. A generic definition is not available, hence does not make sense when comparing the differing interpretations in various domains. For example, Hill [154] stated in 1990 in the context of microprocessor systems that

> [...] I first examine formal definitions of scalability, but I fail to find a useful, rigorous definition of it. I then question whether scalability is useful and conclude by challenging the technical community to either (1) rigorously define scalability or (2) stop using it to describe systems.

The following Table 2.1 lists some examples for the usage of scalability and issues connected with.

| **Domain** | **Issues** |
|---|---|
| software systems | requirements, users [86] |
| parallel processing | number of nodes, memory [180] |
| grid computing | services, distributed implementations [102] |
| web applications | sessions, granularity of content [54] |
| information retrieval systems | capacity, performance, large corpora [212] |

Table 2.1: Scalability in selected domains.

In the literature some research already has been performed regarding scalability. Most importantly, Bondi [44] recently elaborated on scalability issues on a generic level; there he considers four types of scalability: *load scalability*, *space scalability*, *space-time scalability*, and *structural scalability*:

- **Load scalability**. If a system has the ability to function gracefully, i.e., without undue delay and without unproductive resource consumption or resource contention at light, moderate, or heavy loads while making good use of available resources;
- **Space scalability**. If its memory requirements do not grow to intolerable levels as the number of items it supports increases.
- **Space-time scalability**. If a system continues to function gracefully as the number of objects it encompasses increases by orders of magnitude.
- **Structural scalability**. If its implementation or standards do not impede the growth of the number of objects it encompasses, or at least will not do so within a chosen time frame.

The above given attributes form the basis of the analysis given in chapter 6.

### 2.3.1 Infrastructure Level

On the infrastructure level—as RDF stores, reasoning facilities, and the like—a number of research activities can be listed. The outcome here mostly are benchmarks, evaluations, and guides how to implement the infrastructure in an optimal way.

**RDF stores**. Some practical research w.r.t. triple stores has been reported from the SIMILE[34] project [196]. A quite complete survey on RDF storage systems with special attention on scalability is available as a deliverable of the Semantic Web Advanced Development for Europe (SWAD-Europe) project [25]. Wielemaker et.al. [325] outline an Prolog-based infrastructure for loading and saving RDF triples, elementary reasoning with triples and visualization. A predecessor of the infrastructure described there has been used in applications for ontology-based annotation of multimedia objects. The library aims at fast parsing, fast access and scalability for fairly large but not unbounded applications up to 40 million triples. In [120] Guo et.al. present an evaluation of four knowledge base systems w.r.t. to use in large OWL applications. The datasets used range from 15 OWL files totalling 8MB to 999 files totalling 583MB. They evaluated two memory-based systems and two systems with persistent storage. The conclusion of the work is that existing systems need to place a greater emphasis on scalability. For a criticism of benchmarks we invite the reader to refer to [324].

**Reasoning**. Wache et.al. have published some related research on *Scalability Techniques for Reasoning with Ontologies* [318]. In [214] some practical related results w.r.t representing and reasoning about incomplete information are presented. In [282]—based on contexts—a theoretical approach and implementation of Contextual Reasoning in a Semantic Web KB and the associated testing results are presented. Another work on reasoning worth mentioning in the realm of ubiquitous computing is [232]. Heflin [151] proposes in his PhD thesis "... to use reasoning methods that are not sound and complete", which is sensible due to the Open World Assumptions, and further ...

> ... as an alternative to description logics is to use Horn logic. It has been shown that although Horn-logic and the most common description logics can express things the other cannot, neither is more expressive than the other.

**Software & Data Engineering**. The development of Semantic Web applications from an object-oriented programmer's point of view is discussed in [188]. Alba et.al. [7] report on IBMs Semantic Super Computing platform that has been designed to ingest, augment, store, index and support queries on billions of documents. They describe the challenges and lessons learned in the areas of solution design, hardware, operations, middleware, algorithms, and testing.

### 2.3.2 Application Level

The application level covers the *What?* rather than the *How?*. This level is mainly in the scope of this thesis.

**Multimedia/Hypermedia**. For the multimedia realm however, there exists little research efforts. One of the few to mention is [190] that deals with the data rather with the metadata level of scalability. Another work is [9] that focuses on scalability "in both the data and application domains" with an industrial hypermedia system as the testbed.

**Semantic Web Applications**. In [132], Hartmann and Sure describe their contribution to the 2003 Semantic Web Grand Challenge that realizes semantic-based search and access facilities to information represented by semantic portals. Such portals typically provide knowledge about a specific domain and rely on ontologies to structure and exchange this

[34] http://simile.mit.edu

knowledge. They claim that their approach has the following benefits: (i) Significant reduction of content maintenance overhead, (ii) knowledge accessibility for both human and machine agents, atop existing information sources, and (iii) suitability for productive environments.

### 2.3.3 Projects & Activities

In the recent years a number of project implicitly and explicitly addressing scalability issues can be noticed. In the following we discuss some prominent examples of research projects in this area.

- **REOL** - **R**easoning for **E**xpressive **O**ntology **L**anguages,
  http://reol.cs.manchester.ac.uk/ states the project goals as follows:

  > The primary goal of this project is to develop techniques that address the expressivity requirements of various applications. This is to be achieved by a synergy between two previously disjoint techniques. Currently, tableaux algorithms are the state-of-the-art for reasoning with DL ontologies. However, in recent years, great progress has been made in designing resolution-based algorithms for reasoning with ontologies. These two types of algorithms seem to enjoy two complementary properties: tableaux are model-building calculi that seem to perform well on satisfiable problems, whereas resolution is a refutation calculus that seems to perform well on unsatisfiable problems. The main idea of this project is to extend both calculi and provide a common framework for integrating them.

- **REWERSE** - **R**easoning on the **We**b with **R**ules and **S**emantics,
  The homepage of the project, http://rewerse.net/, gives following overview:

  > The community networked and structured by REWERSE will (i) develop a coherent and complete, yet minimal, collection of inter-operable reasoning languages for advanced Web systems and applications; (ii) test these languages on context-adaptive Web systems and Web-based decision support systems selected as test-beds for proof-of-concept purposes; (iii) bring the proposed languages to the level of open pre-standards amenable to submissions to standardisation bodies such as the W3C.

- **MOSES** - **MO**dular and **S**calable **E**nvironment for the **S**emantic Web,
  Projects goals from http://www.hum.ku.dk/moses/:

  > The only way to make the Semantic Web a success is a bottom up approach, enabling it to emerge from the aggregation of locally organized knowledge fragments of varying size. To demonstrate this approach MOSES will create a small but scalable ontology-based Knowledge Management System and an ontology-based search engine that will accept queries and produce answers in natural language.
  >
  > The main goals of the scalable environment will be to demonstrate that it is possible to upgrade ontological systems from new contents, either creating new knowledge domains or incorporating new knowledge into a pre-existing domain. This requires extracting structured knowledge from plain content.

- ▷ **SCALEX** - **Sca**lable **Ex**hibition Server,
  Projects goals from http://www.scalex.info/:

  > SCALEX is an easy to use toolbox for museums and companies that deal with the creation of digital content. With SCALEX it is possible to combine digital content, as for example texts, images, videos and audios, with real exhibition objects. In addition to that SCALEX also supports the creation of purely virtual exhibitions. The presentation of the digital media is directly coupled to the interests of the specific visitor. Exhibitions that are enhanced with digital media open up new interaction possibilities and thereby offer the visitors a completely new experience during exhibition visits.

- **DIET** - **D**ecentralised **I**nformation **E**cosystems **T**echnology,
  http://www.dfki.uni-kl.de/IVS/IVSEnglisch/Projects/diet.html states:

  > The project will involve the theoretical study, implementation and validation of a novel information management framework which will use ecosystem metaphors to turn the global information infrastructure into an open, adaptive, scalable and stable environment for service provision. Initially this will involve the design of an overall framework in which infohabitants - entities which can process information - can interact and coexist in societies set in an information environment.

**Note**: In projects marked with ▷, the author of this work has been or still is active.

While generic scalability and expressivity issues have been researched in an array of projects, practical issues often are neglected. The Semantic Web, as an extension of the Web is per definition a "scaling entity". Real success in terms of user loyalty, community, and market is only truly possible if one is able to offer solutions that scale to the size of the.

Chapter 3

# Multimedia Metadata

*"Tabula rasa."*

(Latin phrase)

This chapter gives an overview on multimedia metadata formats relevant in the context of the Web; it addresses still-images, audio, audio-visual, and multimedia container formats. Based on the work performed by the author and colleagues in the realm of the W3C Multimedia Semantics Incubator Group (MMSEM-XG) [315] diverse aspects of the multimedia metadata formats are discussed herein.

We will, however, not discuss the basic multimedia content formats—as for example the PNG format[1] for still-images, or the Ogg format[2] for videos—used on the (Semantic) Web to represent, and deliver audio-visual content. For an overview and a sound discussion on these basic multimedia formats, and their capabilities w.r.t. metadata, the interested reader is invited to refer to [105], and [63].

## 3.1 Multimedia Container Formats

In contrast to the basic multimedia content formats mentioned above, **container formats** are able to host a range of media, typically including text. Multimedia container formats may differ regarding their capabilities of arranging, and presenting the content. This representational aspect can further be differentiated into (i) a spatial dimension (2D, 3D, layout, etc.), and (ii) a temporal dimension (synchronisation, parallel playout, etc.). Another feature relevant in the context of container formats is the support for manipulating the content (or at least its components) dynamically. Last but not least the support for metadata handling is of great importance, which is valid not only in the realm of the work at hand.

The formats discussed in the following have been selected due to their alignment with the Semantic Web, viz. they meet some minimal requirements[3], as being based on XML, utilising URIs, etc.

[1] http://www.w3.org/Graphics/PNG/
[2] http://www.xiph.org/ogg/
[3] The reader is invited to refer to [128] for further discussions on this topic.

### 3.1.1 eXtensible HyperText Markup Language–(X)HTML

(X)HTML is the family name for the group of languages that form the *lingua franca* of the World Wide Web[4]. While HTML 4.01 [168] is the latest revision of the SGML-based branch of hypertext language dialects, the Web is based on, XHTML [330; 331] now takes over the role of the Web workhorse. Figure 3.1 depicts a HTML page conceived as a simple multimedia container. It demonstrates, how multimedia assets, as still-images, etc. can be embedded into a page and—in conjunction with links—forms a simple hypermedia document.


Figure 3.1: Example HTML page as a multimedia container.

Though (X)HTML is in wide spread use, it bears some serious limitations. It only allows for a spatial arrangement of multimedia assets. Temporal issues can only be handled using scripting functionality in combination with the (X)HTML document object model (DOM).

#### (X)HTML and Metadata

Whilst earlier versions of (X)HTML took a global point of view regarding metadata[5], the newest generation—XHTML 2.0 [331]—heads after a sound basis for integrating metadata. The author of this thesis participates in this effort [5], known as XHTML+RDFa;for further details see section 3.2.3, below.

[4]Cf. http://www.w3.org/MarkUp/Activity

[5]See for example http://www.w3.org/TR/html401/struct/global.html

### 3.1.2 Scalable Vector Graphics–SVG

SVG [314] is a modularized language for describing two-dimensional vector and mixed vector/raster graphics in XML. It allows for describing scenes with vector shapes (e.g. paths consisting of straight lines, curves), text, and multimedia (e.g. still images, video, audio). These objects can be grouped, transformed, styled and composited into previously rendered objects.

SVG files are compact and provide high-quality graphics on the Web, in print, and on resource-limited handheld devices. In addition, SVG supports scripting and animation, so SVG is ideal for interactive, data-driven, personalized graphics. SVG is based on the download-and-play concept. SVG has also a mobile specification, SVG Tiny, which is a subset of SVG.

```
<?xml version="1.0"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN"
  "http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">

<svg xmlns="http://www.w3.org/2000/svg"
 width="467" height="462">
  <rect x="80" y="60" width="250" height="250" rx="20"
      style="fill:#ff0000; stroke:#000000;stroke-width:2px;" />

  <rect x="140" y="120" width="250" height="250" rx="40"
      style="fill:#0000ff; stroke:#000000; stroke-width:2px;
      fill-opacity:0.7;" />
</svg>
```

Listing 3.1: A sample SVG markup.

A sample SVG's document code[6] is depicted in listing 3.1. Note, that though not primarily intended, it is possible to use SVG as a general purpose media container format[7].

Metadata which is included with SVG content is specified within the metadata elements[8], with contents from other XML namespaces such as Dublin Core or RDF. The specification states

> Individual industries or individual content creators are free to define their own metadata schema but are encouraged to follow existing metadata standards and use standard metadata schema wherever possible to promote interchange and interoperability. If a particular standard metadata schema does not meet your needs, then it is usually better to define an additional metadata schema in an existing framework such as RDF and to use custom metadata schema in combination with standard metadata schema, rather than totally ignore the standard schema.

When looking at the deployment of SVG+RDF, the results are rather disillusioning. For example a Google-search in the form of `filetype:svg rdf` in early September 2007 yielded only some 500 hits.

---

[6] From http://en.wikipedia.org/wiki/Scalable_Vector_Graphics

[7] For an example see, e.g., http://www.bluishcoder.co.nz/2007/08/svg-video-demo.html

[8] See section 21 of the SVG Recomendation at http://www.w3.org/TR/SVG11/metadata.html

### 3.1.3 Synchronized Multimedia Integration Language–SMIL

SMIL[9] is a XML-based W3C recommendation [272] for describing interactive multimedia presentations. The language allows for describing the temporal behaviour of a multimedia presentation, associate hyperlinks with media objects and describe the layout of the presentation on a screen.

Components based on SMIL are used for integrating timing into XHTML [330] and into SVG [314]. SMIL components may have different media types, such as audio, video, image or text. The begin and the end time of different components are specified according to events in other media components. For example, in a slide show, a particular slide is displayed when the narrator in the audio starts talking about it. Hyperlinks embedded in the presentation allow for random navigation through the presentation.


Figure 3.2: Example SMIL document as a multimedia container.

Microsoft and others proposed a SMIL-based variant of HTML with the *Timed Interactive Multimedia Extensions for HTML* (HTML+TIME)[10]; the dissemination of SMIL still is somehow limited. This might as well be rooted in the complex issues w.r.t. multimedia presentations. An exemplary SMIL markup[11] is depicted in listing 3.2 on page 35.

#### SMIL and Metadata

With the metainformation module[12], SMIL now supports—additionally to the use of the `meta` element (from SMIL 1.0)—the description of metadata using the Resource Description Framework (RDF) model and Syntax (cf. section 4.3.3).

[9] pronounced as “smile”
[10] http://www.w3.org/TR/NOTE-HTMLplusTIME
[11] From http://www.streaming-media.biz/cnt330.html
[12] http://www.w3.org/TR/SMIL2/metadata.html

```xml
<?xml version="1.0"?>

 <smil xmlns="http://www.w3.org/2001/SMIL20/Language"

  <head>

   <layout>
    <root-layout width="320" height="240" background-color="black" />
     <region id="r1" left="0" top="0" width="320" height="240" fit="fill"
         z-index="1"/>
   </layout>

  </head>

  <body>

    <seq>
     <video src="/media/realvideo8_sure.ram" region="r1" dur="60s" />
     <video src="/media/realvideo8_sure1.ram" region="r1" begin="30s"
         dur="60s"/>
     <video src="/media/realvideo8_sure2.ram" region="r1" begin="60s"
         dur="60s"/>
    </seq>

  </body>

 </smil>
```

Listing 3.2: A sample SMIL markup.

### 3.1.4 eXtensible 3D—X3D

Extensible 3D (X3D) [329] is a ISO standard for defining interactive web- and broadcast-based 3D content integrated with multimedia. X3D is intended for use on a variety of hardware devices and in a broad range of application areas such as engineering and scientific visualization, multimedia presentations, entertainment and educational titles, web pages, and shared virtual worlds. X3D is also intended to be a universal interchange format for integrated 3D graphics and multimedia.

As the successor to the Virtual Reality Modeling Language (VRML), X3D improves upon VRML with new features, advanced APIs, additional data encoding formats, and a componentised architecture allowing for a modular approach to supporting the standard. X3D has a rich set of features[13] including 2D and 3D graphics, animation, spatialised audio and video, user interaction, and scripting capabilities. For a discussion on the current state of the standardisation of Virtual Reality the reader is referred to [122].

The Fig. 3.3 (page 36) depicts a sample X3D document; an excerpt of the markup of this sample is shown in listing 3.3 on page 37 . The X3D in this setup is used as a 3D interface for a video summary browser along with logical annotations.

[13]For a comprehensive list, cf. http://www.web3d.org/x3d/specifications/

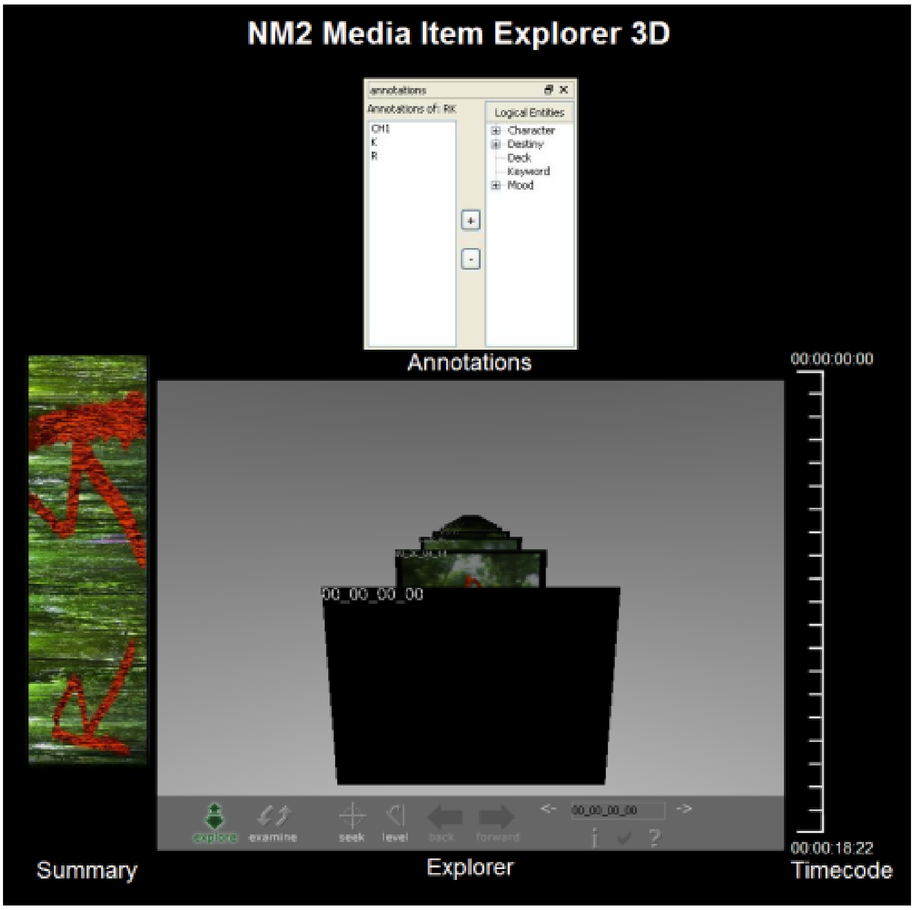



Figure 3.3: Example X3D document as a multimedia container.

### X3D and Metadata

*Part 1: Architecture and base components* of the X3D specification [329] generically talks about metadata in the sense:

> Metadata is information that is associated with the objects of the X3D world but is not a direct part of the world representation.

It further defines an abstract interface `X3DMetadataObject` identifying a node as containing metadata and metadata nodes that specify metadata values in various data types. Three basic representation types are specified: strings, floating point values, and integers. Each piece of metadata has two additional strings that describe (i) the metadata standard (if any) from which the metadata specification emanates, and (ii) the identification for the particular piece of metadata being provided.

In [249] an approach for associating semantic information to 3D worlds based on the integration of the X3D language and Semantic Web languages is proposed. The approach is characterized also by the definition of scene-independent ontologies and by the definition of semantic zones that complement the role of semantic objects for giving a complete description of the environment.

```xml
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE X3D PUBLIC "ISO//Web3D//DTD X3D 3.0//EN"
    "http://www.web3d.org/specifications/x3d-3.0.dtd">
<X3D profile="Immersive">
   <Scene>
      <NavigationInfo avatarSize="3 5 3" visibilityLimit="0" speed="2"
          headlight="true" type="'FLY'" />

      <Fog DEF="Fog" fogType="EXPONENTIAL" color="0 0 0"
          visibilityRange="15.0" />

      <Transform translation="-10 0 20">
         <Background groundAngle="0 3" groundColor="0 0 0 1 1 1 0 0 0" />
      </Transform>

      <Transform>
         <Group>
            <Group>
               <Transform translation="0 0 16" rotation="0 1 0 1.570796">
                  <Shape>
                     <Appearance>
                        <ImageTexture url="keyframes/00_00_00_00.jpg" />
                     </Appearance>

                     <Box size="0.01 1.2 1.6" solid="true" />
                  </Shape>
               </Transform>

               <Transform translation="-0.8 0.635 16.01" rotation="0 0 1
                   0">
                  <Shape>
                     <Appearance>
                        <Material ambientIntensity="0.6" shininess="0.1"
                            diffuseColor="1 1 1" specularColor="0 0 0" />
                     </Appearance>

                     <Text string="00_00_00_00">
                        <FontStyle family="Arial" size="0.1" />
                     </Text>
                  </Shape>
               </Transform>
            </Group>

      ...

      </Transform>

      <Viewpoint description="00_00_00_00" jump="true" fieldOfView="1.0"
          position="0 1.0 18" orientation="1 0 0 -0.2" />

      ...

      <Viewpoint description="00_00_18_12" jump="true" fieldOfView="1.0"
          position="0 1.0 -18" orientation="1 0 0 -0.2" />

      <Viewpoint description="00_00_18_22" jump="true" fieldOfView="1.0"
          position="0 1.0 -20" orientation="1 0 0 -0.2" />
   </Scene>
</X3D>
```

Listing 3.3: A sample X3D markup.

## 3.2 Aspects of Multimedia Metadata

The term metadata has become particularly common with the popularity of the Web; the underlying concepts have been around for at least 20 years. Till the early 1990s, metadata was a term most prevalently used by communities involved with the management of library systems, public institutions, and the like.

For these communities, metadata referred to a suite of disciplinary standards as well as additional internal and external documentation and other data necessary for the identification, representation, interoperability, technical management, performance, and use of data contained in an information system [110].

### 3.2.1 An Attempt of A Definition

Trying to find a common definition for metadata yields a plethora of options. The most widely used definitions (or explanations alike) of metadata are gathered in the following.

**Dictionary-oriented** The prefix *meta-* is defined[14] as meaning *after* or *beyond*. According to [53] this means that the concept of metadata inherently asserts a certain relationship between the metadata and the information resource which it describes.

**Data-oriented** Here, metadata describes the content, quality, condition, and other characteristics of data. Metadata answers who, what, when, where, why, and how questions about every facet of the data that are being documented in a consistent and precise format. This allows a potential user to decide whether the data is appropriate. That means that metadata can relieve potential users of having to have full advance knowledge of the existence of a dataset and characteristics; cf. [328].

**Application-oriented** Metadata literally is *data about data*; it is an ubiquitous term that is understood in different ways by diverse professional communities that use information systems and resources. For example in the realm of GIS-applications cf. [213].

**Web-oriented** From an architectural point of view [30], metadata consists of assertions about data. Such assertions typically take the form of a name or type of assertion and a set of parameters. The metadata can be represented as a set of independent assertions.

This model implies that in general, two assertions about the same resource can stand alone and independently. When they are grouped together in one place, the combined assertion is simply the sum of the independent ones. Due to this commutative characteristic, collections of assertions are essentially unordered sets. In [192] introduces a new way of using metadata on the Web: The Resource Description Framework.

It can be noted that for each use case the metadata—and its relation to the actual data it is referring to—has a different significance. Not only the focus might vary depending on the application environment, but also the way metadata is actually is utilised.

In the following, we give a definition of metadata in our understanding, as used throughout this work. We do not share the attempts of most authors to define metadata as a rigid property; rather **metadata** is viewed as a **role** data plays in a certain context.

[14] by Merriam-Webster, http://www.m-w.com/

**Definition 3.1 (Metadata).**
**Metadata** is a functional aspect (a role) of data with the following characteristics:

1. Its purpose is to support a task (as query, ranking, etc.);
2. It refers to some other data. ❖

The first requirement actually embodies the rational for introducing metadata; metadata is not an end in itself, but rather enables, supports, simplifies, etc. a certain task. Regarding the second requirement it can be noted that the actual bit metadata may refer to can be data in the simple case, or even (other) metadata—sometimes referred to as meta-metadata.

### 3.2.2 Types of Metadata

In [52], metadata is differentiated into the following categories:

1. **Syntactic Metadata**, which describes non-contextual information about the content (as the document size) with little or no contextual relevance.
2. **Structural Metadata** that provides information regarding the structure of the content; it depends on the type of the document.
3. **Semantic Metadata**, cf. also [6]—Adds relationships, rules and constraints to syntactic and structural metadata. It describes contextually relevant or domain-specific information about the content; uses domain specific metadata model or ontology. For further discussion see Chapter 4.

Although the above given categorisation is not a functional one, we subscribe to this view.

In the following we give a categorisation that reflects basic types of metadata functionality based on work suggested in [236], [53], and [30].

**Administrative**
Metadata used in managing and administering information resources. This type of metadata can be used for acquisition information, rights tracking, location information, and version control.

**Descriptive**
Metadata used to describe information resources. This type of metadata can be used for: cataloging records, finding aids and specialized indexes, relationships between resources, annotations by users and the like and can further be differentiated into:

- *Low-level Descriptive*. Descriptive metadata with informal semantics and/or without domain-specific value.
- *High-level Descriptive*. Descriptive metadata with formal grounding and usually based on domain knowledge.

**Contextual**
Metadata used to put information into a context. It can be utilised for Human Computer Interaction (HCI), for retrieval purposes, and for machine learning techniques.

### 3.2.3 Scope of Metadata

The scope of metadata may be discriminated along several dimensions. In the following we examine three dimensions widely being addressed in multimedia metadata formats, namely

- The level of formality (regarding the semantics);
- Granularity of content descriptions;
- Spatio-temporal invariance.

#### Level of Formality Regarding Semantics

The following table 3.1 lists the scope of metadata along dimension of semantics. The discriminator in this setting is the level of formality w.r.t. semantics effecting the type of query one is able to perform.

| Functional Type | Example | Sample Formats |
|---|---|---|
| Administrative | resolution of an image, e.g., 200x300 pixel | Exif, IPTC/XMP |
| Low-level Descriptive | colour and shape of an object | MPEG-7 |
| High-level Descriptive | a certain region depicts a frog | RDF-S/OWL, WSML |
| Contextual | $hasColour(x, GREEN) \leftarrow type(x, Frog), state(x, Living)$ | Prolog, Jena rules, RIF |

Table 3.1: Scope of Metadata regarding their Functional Type.

#### Granularity of Content Descriptions

The scope of metadata can be distinguished along the discriminator of content description granularity. A global description refers to an entire multimedia asset, such as the dimension of an image or its author whereas a local description only addresses a certain spatio-temporal segment of a multimedia asset. For example the dominant colour of a certain region of an image has a local character. With fine granular descriptions available, the question of how to properly address them arises. We have started work on this already; see our position paper [294] at the W3C Video on the Web Workshop[15].

#### Invariance

Further, the scope of metadata can be understood in terms of invariance. Static features are independent of time and space, hence spatio-temporally invariant. When taking a picture with a camera, one piece of static metadata is the information about the person that shot the picture. However, most features have dependencies w.r.t. both time and space.

[15] http://www.w3.org/2007/08/video/

## 3.3 Multimedia Metadata Formats

In the following sections an overview on multimedia metadata (M3) formats, based on a deliverable [141] of the W3C Multimedia Semantics Incubator Group [315] is presented. The reader is invited to note that the author of this thesis has been mainly responsible for editing this technical report. The presentation of the M3 formats in the following is organised based on [273].

### 3.3.1 Metadata for Still Images

In the following section, M3 formats are listed dealing with the description of still image content.

#### Visual Resource Association–VRA

The Visual Resource Association (VRA)[16] is an organisation consisting of over 600 active members, including many American Universities, galleries and art institutes. These often maintain large collections of (annotated) slides, images and other representations of works of art. The VRA has defined the VRA Core Categories to describe such collections. The VRA Core is a set of metadata elements used to describe works of visual culture as well as the images that represent them.

#### Exchangeable image file format–Exif

One of today's commonly used image format and metadata standard is the Exchangeable Image File Format (Exif)[17]. The standard defines the format for both images and sound captured using digital still cameras; additionally, the Exif format provides a standard specification for storing metadata regarding images and sound.

The metadata tags which the Exif standard provides covers metadata related to the capture of the image and the context situation of the capturing. This includes metadata related to the image data structure (e.g., height, width, orientation), capturing information (e.g., rotation, exposure time, flash), recording offset (e.g., image data location, bytes per compressed strip), image data characteristics (e.g., transfer function, colour space transformation), as well as general tags (e.g., image title, copyright holder, manufacturer). In these days new camera also write GPS information into the header. Lastly, we point out that metadata elements pertaining to the image are stored in the image file header and are marked identified by unique tags serving as an element identifier.

#### NISO Z39.87

The NISO Z39.87 standard[18] defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files. Tags cover a wide spectrum of metadata: basic image parameters, image creation, imaging performance assessment, history. This standard is intended to facilitate the development of applications

[16] http://www.vraweb.org/vracore3.htm
[17] http://www.digicamsoft.com/exif22/exif22/html/exif22_1.htm
[18] http://www.niso.org/standards/resources/Z39-87-2006.pdf

to validate, manage, migrate, and otherwise process images of enduring value. Such applications are viewed to be essential components of large-scale digital repositories and digital asset management systems.

The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management and continuing access to digital image collections.

#### DIG35

The DIG35 specification[19] includes a standard set of metadata for digital images, which promotes interoperability and extensibility, as well as a uniform underlying construct to support interoperability of metadata between various digital imaging devices.

The metadata properties are encoded within an XML Schema and cover:

- Basic Image Parameter (a general-purpose metadata standard);
- Image Creation (e.g. the camera and lens information);
- Content Description (who, what, when and where);
- History (partial information about how the image got to the present state);
- Intellectual Property Rights;
- Fundamental Metadata Types and Fields (define the format of the field defined in all metadata block).

### 3.3.2 Metadata for describing Audio Content

This section contains metadata for audio content, be it related to music, or speech.

#### ID3

ID3[20] is a metadata container used and embedded with an MP3 audio file format. It allows to state information about the title, artist, album, etc. about a song. The ID3 specification aims to address a broad spectrum of metadata (represented in so called 'frames') ranging from encryption, over involved people list, lyrics, band, relative volume adjustment to ownership, artist, and recording dates. Additionally user can define own properties. A list of 79 genres is defined (from Blues to Hard Rock).

#### MusicBrainz Metadata Initiative 2.1

MusicBrainz defines a (RDF-S based) vocabulary[21]: three namespaces are defined. The core set is capable of expressing basic music related metadata (as artist, album, track, etc.). Instances in RDF are being made available via a query language. The third namespace is reserved for future use in expressing extended music related metadata (as contributors, roles, lyrics, etc.).

[19] http://xml.coverpages.org/FU-Berlin-DIG35-v10-Sept00.pdf
[20] http://www.id3.org/Developer_Information
[21] http://musicbrainz.org/MM/

#### MusicXML

Recordare has developed MusicXML[22] technology to create an Internet-friendly method of publishing musical scores, enabling musicians and music fans to get more out of their online music. MusicXML is a universal translator for common Western musical notation from the $17^{th}$ century onwards. It is designed as an interchange format for notation, analysis, and retrieval for music notation nd digital sheet music applications. The MusicXML format is open for use by anyone under a royalty-free license, and is supported by over 75 applications.

### 3.3.3 Metadata for describing Audio-Visual Content

In this section, multimedia metadata formats for describing audio-visual content in general are described.

#### Advanced Authoring Format–AAF

The Advanced Authoring Format (AAF) [1] is a cross-platform file format that allows the interchange of data between multimedia authoring tools. AAF supports the encapsulation of both metadata and essence, but its primary purpose involves the description of authoring information. The object-oriented AAF object model allows for extensive timeline-based modeling of compositions (i.e. motion picture montages), including transitions between clips and the application of effects (e.g. dissolves, wipes, flipping). Hence, the application domain of AAF is within the post production phase of an audiovisual product and it can be employed in specialized video work centers. Among the structural metadata contained for clips and compositions, AAF also supports storing event-related information (e.g. time-based user annotations and remarks) or specific authoring instructions.

AAF files are fully agnostic as to how essence is coded and serve as a wrapper for any kind of essence coding specification. In addition to describe the current location and characteristics of essence clips, AAF also supports descriptions of the entire derivation chain for a piece of essence, from its current state to the original storage medium, possibly a tape (identified by tape number and time code), or a film (identified by an edge code for example).

The AAF data model and essence are independent of the specificities of how AAF files are stored on disk. The most common storage specification used for AAF files is the Microsoft Structured Storage format, but other storage formats (e.g. XML) can be used.

The AAF metadata specifications and object model are fully extensible (e.g. subclassing existing objects) and the extensions are fully contained in a metadata dictionary, stored in the AAF file. In order in order to achieve predictable interoperability between implementations created by different developers, due to the format's flexibility and use of proprietary extensions, the Edit Protocol was established. The Edit Protocol combines a number of best practices and constraints as to how an Edit Protocol-compatible AAF implementation must function and which subset of the AAF specification can be used in Edit Protocol-compliant AAF files.

[22] http://www.recordare.com/xml.html

#### Material Exchange Format–MXF

The Material Exchange Format (MXF) [275] is a streamable file format optimized for the interchange of material for the content creation industries. MXF is a wrapper/container format intended to encapsulate and accurately describe one or more 'clips' of audiovisual essence (video, sound, pictures, etc.). This file format is essence-agnostic, which means it should be independent of the underlying audio and video coding specifications in the file. In order to process such a file, its header contains data about the essence. An MXF file contains enough structural header information to allow applications to interchange essence without any a priori information. The MXF metadata allows applications to know the duration of the file, what essence codecs are required, what timeline complexity is involved and other key points to allow interchange.

MXF uses KLV coding throughout the file structure. This KLV is a data interchange format defined by the simple data construct: Key-Length-Value, where the Key identifies the data meaning, the Length gives the data length, and the Value is the data itself. This principle allows a decoder to identify each component by its key and skip any component it cannot recognize using the length value to continue decoding data types with recognized key values. KLV coding allows any kind of information to be coded. It is essentially a machine-friendly coding construct that is datacentric and is not dependent on human language. Additionally, the KLV structure of MXF allows this file format to be streamable.

Structural Metadata is the way in which MXF describes different essence types and their relationship along a timeline. The structural metadata defines the synchronization of different tracks along a timeline. It also defines picture size, picture rate, aspect ratio, audio sampling, and other essence description parameters. The MXF structural metadata is derived from the AAF data model. Next to the structural metadata described above, MXF files may contain descriptive and dark metadata.

MXF descriptive metadata comprises information in addition to the structure of the MXF file. Descriptive metadata is metadata created during production or planning of production. Possible information can be about the production, the clip (e.g. which type of camera was used) or a scene (e.g. the actors in it). DMS-1 (Descriptive Metadata Scheme 1) [274] is an attempt to standardize such information within the MXF format. Furthermore DMS-1 is able to interwork as far as practical with other metadata schemes, such as MPEG-7, TV-Anytime, etc. and Dublin Core. The SMPTE Metadata Dictionary is a thematically structured list of metadata elements, defined by a key, the size of the value and its semantics.

Dark Metadata is the term given to metadata that is unknown by an application. This metadata may be privately defined and generated, it may be new properties added or it may be standard MXF metadata not relevant to the application processing this MXF file. There are rules in the MXF standard on the use of dark metadata to prevent numerical or namespace clashes when private metadata is added to a file already containing dark metadata.

#### EBU P_Meta

P_Meta [250] is a standardised metadata exchange scheme which offers a way of sharing the meaning of electronic information necessary or useful for the business-to-business exchange of content.

The P_Meta Scheme is basically a set of definitions which provide a semantic framework for the information which is typically exchanged along with audio-visual material. It includes the identification of concepts (simple or complex) that are referenced by P_Meta names and P_Meta Identifiers. P_Meta refers to existing standards and uses existing reference data when applicable.

As a key principle, the data analysis has been set at the lowest level to identify concepts or subjects which cannot be further divided, thus giving maximum precision in meaning, with maximum flexibility in the use and re-use of basic elements.

The P_Meta Scheme is intended for use in a business-to-business scenario where the participating organisations may retain their internal data structures, workflows, and concepts. The P/Meta definition uses a three-layer model. The standard specifies the definition layer (i.e. the semantic of the description). The technology layer defines the encoding used for exchange; currently KLV (key, length, value) and XML representations are specified. The lowest layer, the data interchange layer, is out of scope of the specification. P/Meta consists of a number of attributes (some of them with a controlled list of values), which are organized into sets. The standard covers the following types of metadata:

- Identification
- Technical metadata
- Programme description and classification
- Creation and production information
- Rights and contract information
- Publication information

It is worth mentioning that EBU is working on replacing P_Meta by NewsML-G2, cf. section 3.3.5.

### 3.3.4 Multimedia Content Description Interface–MPEG-7

The MPEG-7 standard [226; 227; 206; 222], formally named "Multimedia Content Description" aims to be an overall for describing any multimedia content. The objective of the group that developed this standard is to enable efficient search, filtering and browsing of the multimedia content.Standardization has started in 1996 and is partly still on-going[23].

Possible applications are in the areas of digital audiovisual libraries, electronic news media and interactive TV. MPEG-7 provides standardized description schemes that allow creating descriptions of material that are directly linked with the essence to support efficient retrieval. The audiovisual information can be represented in various forms of media, such as pictures, 2D/3D models, audio, speech, and video. Nowadays, there are an increasing number of cases where the audiovisual information is created, exchanged, retrieved, and re-used by computational systems. Because MPEG7 was developed to be global, it is independent of how the content is coded or stored. There exist several usage scenarios:

- Image acquisition (digital cameras, etc.);
- Media conversion (speech to text, etc.);
- Information retrieval, pull media filtering: quickly and efficiently searching for various types of multimedia documents of interest to the user;

[23] See the MPEG-7 section at http://www.itscj.ipsj.or.jp/sc29/29w42911.htm

- Filtering in a stream of audiovisual content description, push media filtering: e.g. from a variety of digital TV channels only those matching the user's preferences are presented.

MPEG-7 standardizes so-called ”description tools” for multimedia content: Descriptors (Ds), Description Schemes (DSs) and the relationships between them. Descriptors are used to represent specific features of the content, generally low-level features such as visual (e.g. texture, camera motion) or audio (e.g. melody), while description schemes refer to more abstract description entities (usually a set of related descriptors). These description tools as well as their relationships are represented using the Description Definition Language (DDL), a core part of the language. The W3C XML Schema recommendation has been adopted as the most appropriate schema for the MPEG-7 DDL, adding a few extensions (array and matrix datatypes) in order to satisfy specific MPEG-7 requirements. MPEG-7 descriptions can be serialized as XML or in a binary format defined in the standard.

This standard is subdivided into nine parts:

**Part 1** Systems: specifies the tools (in the sense of description schemas) for preparing descriptions for efficient transport and storage, compressing descriptions, and allowing synchronization between content and descriptions;

**Part 2** Description Definition Language (DDL): specifies the language for defining the standard set of description tools (Description Schemes, Descriptors, and Data Types);

**Part 3** Visual: specifies the description tools pertaining to visual content—cf. [220];

**Part 4** Audio: specifies the description tools pertaining to audio content—cf. [221];

**Part 5** Multimedia Description Schemes (MDS): specifies the generic description tools pertaining to multimedia including audio and visual content—cf. [218];

**Part 6** Reference Software: a software implementation of relevant parts of the standard with normative status;

**Part 7** Conformance testing: guidelines and procedures for testing conformance of implementations to the standard;

**Part 8** Extraction and use of descriptions: informative material (in the form of a technical report) about the extraction and the use of some of the Description tools;

**Part 9** Profiles and levels: relevant parts of the standard grouped for interoperability—cf. [219].

Because MPEG7 defines a set of descriptors that express different viewpoints of the description of the audio-visual content, it is designed to take into account all the viewpoints under consideration by other leading standards such as Dublin Core [155], SMPTE Metadata dictionary etc. [309].

The comprehensiveness results from the fact that the standard has been designed for a broad range of applications and thus employs very general and widely applicable concepts. The standard contains a large set of tools for diverse types of annotations on different semantic levels (the set of MPEG-7 XML Schemas define 1182 elements, 417 attributes and 377 complex types). The flexibility is very much based on the structuring tools and allows the description to be modular and on different levels of abstraction. MPEG-7 supports fine grained description, and it provides the possibility to attach descriptors to arbitrary segments on any level of detail of the description. The possibility to extend MPEG-7 according to the conformance guidelines defined in part 7 provides further flexibility.

Two main problems arise in the practical use of MPEG 7 from its flexibility and comprehensiveness: complexity and limited interoperability. The complexity is a result of the use of generic concepts, which allow deep hierarchical structures, the high number of different descriptors and description schemes and their flexible inner structure, i.e. the variability concerning types of descriptors and their cardinalities. This causes sometimes hesitance in using the standard. The interoperability problem is a result of the ambiguities that exist because of the flexible definition of many elements in the standard (e.g. the generic structuring tools). There can be several options to structure and organize descriptions which are similar or even identical in terms of content, and they result in conformant, yet incompatible descriptions. The description tools are defined using DDL. Their semantics is described textually in the standard documents. Due to the wide application are, the semantics of the description tools are often very general. Several works have already pointed out the lack of formal semantics of the standard that could extend the traditional text descriptions into machine understandable ones. These attempts that aim to bridge the gap between the multimedia community and the Semantic Web, either for the whole standard, or just one of its part, are detailed below.

**Profiles in MPEG-7**

Profiles and levels have been proposed as a means to reduce the complexity of MPEG-7 descriptions (cf. *Definition of MPEG-7 Description Profiling*, ISO/IEC 15938-9). Like in other MPEG standards, profiles are subsets of the standard that cover certain functionalities, while levels are flavours of profiles with different complexity. In MPEG-7, profiles are subsets of description tools for certain application areas, levels have not yet been used.

Several profiles have been under consideration for standardization and three profiles have been standardized (they constitute part 9 of the standard, with their XML schemas being defined in part 11):

- **Simple Metadata Profile** (SMP) allows describing single instances of multimedia content or simple collections. The profile contains tools for global metadata in textual form only. The proposed Simple Bibliographic Profile is a subset of SMP. Mappings from ID3, 3GPP and EXIF to SMP have been defined;

- **User Description Profile** (UDP) consists of tools for describing user preferences and usage history for the personalization of multimedia content delivery;

- **Core Description Profile** (CDP) enables describing image, audio, video and audiovisual content as well as collections of multimedia content. Tools for the description of relationships between content, media information, creation information, usage information and semantic information are included. The profile does not include the visual and audio description tools defined in parts 3 and 4.

The adopted profiles will not be sufficient for a number of applications. If an application requires additional description tools, a new profile must be specified. It will thus be necessary to define further profiles for specific application areas. For interoperability it is crucial, that the definitions of these profiles are published, to check conformance to a certain profile and define mappings between the profiles. It has to be noted, that all of the adopted profiles just define the subset of description tools to be included and some tool constraints; none of the profile definitions includes constraints on the semantics of the tools that clarify how they are to be used in the profile.

Apart from the standardized ones, a profile for the detailed description of single audiovisual content entities called Detailed Audiovisual Profile (DAVP) has been proposed. The profile includes many of the MDS tools, such as a wide range of structuring tools, as well as tools for the description of media, creation and production information and textual and semantic annotation, and for summarisation. In contrast to the adopted profiles, DAVP includes the tools for audio and visual feature description, which was one motivation for the definition of the profile. The other motivation was to define a profile the supports interoperability between systems using MPEG-7 by avoiding possible ambiguities and clarifying the use of the description tools in the profile. The DAVP definition thus includes a set of semantic constraints, which play a crucial role in the profile definition. Due to the lack of formal semantics in DDL, these constraints are only described in textual form in the profile definition [18]. In our work on formalising MPEG-7 profiles [291; 19] we have shown possible further directions on how to handle this issue.

#### Controlled vocabularies in MPEG-7

Annotation of content often contains references to semantic entities such as objects, events, states, places, and times. In order to ensure consistent descriptions (e.g. make sure that persons are always referenced with the same name) some kind of controlled vocabulary should be used in these cases. MPEG-7 provides a generic mechanism for referencing terms defined in controlled vocabularies. The only requirement is that the controlled vocabulary is identified by a URI, so that a specific term in a specific controlled vocabulary can be referenced unambiguously. In the simplest case, the controlled vocabulary is just a list of possible values of a property in the content description, without any structure. The list of values can be defined in a file accessed by the application or can be taken from some external source, for example the list of countries defined in ISO 3166. The mechanism can also be used to reference terms from other external vocabularies, such as thesauri or ontologies.

**Classification schemes** (CSs) are a MPEG-7 description tool that allows to describe a set of terms using MPEG-7 description schemes and descriptors. It allows to define hierarchies of terms and simple relations between them, and allows the term names and definitions to be multilingual. Part 5 of the standard already defines a number of classification schemes, and new ones can be added. The CSs defined in the standard are for those description tools, which require or encourage the use of controlled vocabularies, such as

- Technical media information: encoding, physical media types, file formats, defects;
- Content classification: genre, format, rating;
- Other: affection, role of creator, dissemination format.

Fig. 3.4 on page 50 depicts a video clip[24]—including extracted key frames and a stripe image—being described by an MPEG-7 document as shown in listing 3.4 (cf. page 49).

[24]Taken by author at the EuroITV 2007 conference

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
       xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
<Description xsi:type="ContentEntityType">
 <MultimediaContent xsi:type="AudioVisualType">
  <AudioVisual id="AVID_1">
   ...
   <MediaSourceDecomposition criteria="modalities">
    <VideoSegment id="VSID_1">
     <StructuralUnit
        href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005/vis.programme">
      <Name>Programme</Name>
     </StructuralUnit>
     <TemporalDecomposition criteria="visual shots">
      <VideoSegment id="TRID_1" xsi:type="ShotType">
       <StructuralUnit
          href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.shot">
        <Name>Shot</Name>
       </StructuralUnit>
       <MediaTime>
        <MediaTimePoint>T00:00:00:0F25</MediaTimePoint>
        <MediaDuration>P0DT0H0M17S21N25F</MediaDuration>
       </MediaTime>
       <VisualDescriptor xsi:type="CameraMotionType">
        &lt;Segment xsi:type="MixtureCameraMotionSegmentType">
         <MediaTime>
          <MediaTimePoint>T00:00:07:0F25</MediaTimePoint>
          <MediaDuration>P0DT0H0M0S14N25F</MediaDuration>
         </MediaTime>
         <AmountOfMotion>
          <PanRight>10</PanRight>
          <TiltDown>117</TiltDown>
         </AmountOfMotion>
        &lt;/Segment&gt;
       </VisualDescriptor>
       <TemporalDecomposition>
        <VideoSegment id="KFID_1">
          <MediaLocator>
           <MediaUri>00_00_00_00.jpg</MediaUri>
          </MediaLocator>
         <StructuralUnit
            href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.keyframe">
          <Name>Keyframe</Name>
         </StructuralUnit>
         <MediaTime>
          <MediaTimePoint>T00:00:00:0F25</MediaTimePoint>
          <MediaDuration>P0DT0H0M0S1N25F</MediaDuration>
         </MediaTime>
        </VideoSegment>
       </TemporalDecomposition>
      </VideoSegment>
     </TemporalDecomposition>
    </VideoSegment>
   </MediaSourceDecomposition>
  </AudioVisual>
 </MultimediaContent>
</Description>
</Mpeg7>
```

Listing 3.4: Excerpt of an exemplary MPEG-7 document.


Figure 3.4: A sample video clip described using MPEG-7.

### 3.3.5 Formats For Describing Specific Domains Or Workflows

**MPEG-21**

The aim for MPEG-21 is to describe how these various elements fit together. The result is an open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain. This open framework thus provides content creators and service providers with equal opportunities in the MPEG-21 enabled open market. This will also be to the benefit of the content consumer providing them access to a large variety of content in an interoperable manner.

The vision for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. The standard is currently[25] (after the March 2004 MPEG meeting) divided into 16 parts, most of them still under development. The number of parts may still increase.

MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction (the Digital Item) and the concept of Users. A Digital Item is a structured digital object with a standard representation, identification and standard representation, identification and metadata. It can be considered the *what* of the Multimedia Framework (e.g., a video collection, a music album).

The Users can be considered the *who* of the Multimedia Framework. From a technical perspective, all parties that have a requirement within MPEG-21 to interact are categorised equally as Users. A User is any entity that interacts in the MPEG-21 environment or makes use of a Digital Item. Each User will assume specific rights and responsibilities according to their interaction with other users.

Users include for example individuals, organisations, corporations, communities, consortia, governments and other standards bodies and can assume roles like creators, consumers, rights holders, content providers, distributors, etc.

In its "'Vision, Technologies and Strategy"' technical report (ISO/IEC TR 21000-1, [224]) MPEG has identified seven key architectural elements for the MPEG-21 multimedia framework that are needed to support the multimedia delivery chain, and is in the process of

[25]See the MPEG-21 section at http://www.itscj.ipsj.or.jp/sc29/29w42911.htm

defining the relationships between and the operations supported by them. MPEG will elaborate the elements by defining the syntax and semantics of their characteristics, such as interfaces to the elements. MPEG-21 in particular will address the necessary framework functionality, such as the protocols associated with the interfaces, and mechanisms to provide a repository, composition, conformance, etc.

The seven identified key elements are:

1. Digital Item Declaration (a uniform and flexible abstraction and interoperable schema for declaring Digital Items)

2. Digital Item Identification and Description (a framework for identification and description of any entity regardless of its nature, type or granularity)

3. Content Handling and Usage (provide interfaces and protocols that enable creation, manipulation, search, access, storage, delivery, and (re)use of content across the content distribution and consumption value chain)

4. Intellectual Property Management and Protection (the means to enable Digital Items and their rights to be persistently and reliably managed and protected across a wide range of networks and devices)

5. Terminals and Networks (the ability to provide interoperable and transparent access to content across networks and terminals)

6. Content Representation (how the media resources are represented

7. Event Reporting (the metrics and interfaces that enable Users to understand precisely the performance of all reportable events within the framework)

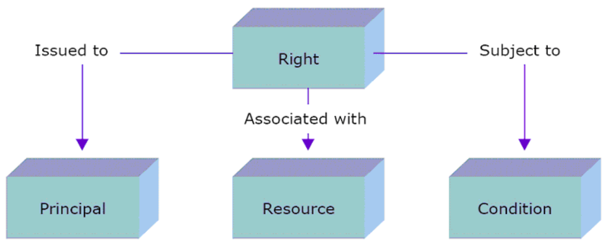



Figure 3.5: The MPEG-21 REL.

Regarding DRM issues the MPEG REL adopts a simple and extensible core data model for its key concepts and elements. The basic MPEG-21 REL element is the license. A license (Fig. 3.5) may contain as of [223] one or more grants. This basic relationship structurally consists of the following parts:

- The principal to whom the grant is issued
- The right that the grant specifies
- The resource to which the right in the grant applies
- The condition that must be met before the right can be exercised

The author of this work has—together with partners from the Know Center[26]—researched on applying MPEG-21 w.r.t. DRM in the realm of the eLARM project[27]. There, we investigated approaches how to semantically describe and integrate various DRM systems [202].

```xml
<?xml version="1.0" encoding="UTF-8"?>
<license xmlns="urn:mpeg:mpeg21:2003:01-REL-R-NS"
 xmlns:sx="urn:mpeg:mpeg21:2003:01-REL-SX-NS"
 xmlns:mx="urn:mpeg:mpeg21:2003:01-REL-MX-NS"
 xmlns:dsig="http://www.w3.org/2000/09/xmldsig#"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xml:base="http://sw.joanneum.at/drm#"
 xsi:schemaLocation="urn:mpeg:mpeg21:2003:01-REL-MX-NS
 http://RightsExpress.ContentGuard.Com/schemas/rel-mx.xsd">
<grant licensePartId="grant_0">
 <forAll varName="principal_0">
  <propertyPossessor>
   <sx:propertyUri definition="http://sw-app.org/mic.xhtml#i"/>
  </propertyPossessor>
 </forAll>
 <principal varRef="principal_0" />
 <mx:play />
 <mx:diReference>
  <mx:identifier>
   http://sw.joanneum.at/NM2/testbed/example.avi
  </mx:identifier>
 </mx:diReference>
 <allConditions>
  <validityInterval>
   <notBefore>2004-09-01T00:00:00Z</notBefore>
   <notAfter>2007-09-01T00:00:00Z</notAfter>
  </validityInterval>
  <sx:territory>
         <sx:location>
          <sx:country
              xmlns:iso="urn:mpeg:mpeg21:2003:01-REL-SX-NS:country">
                iso:AT
          </sx:country>
   </sx:location>
  </sx:territory>
 </allConditions>
</grant>
</license>
```

Listing 3.5: An exemplary MPEG-21 license.

In the listing 3.5 a sample MPEG-21 license is rendered. It basically expresses that a user with the identity `http://sw-app.org/mic.xhtml#i` is allowed to `play` (i.e. view) a certain asset (`http://sw.joanneum.at/NM2/testbed/example.avi`) during a given time period—that is between 1. September 2004 and 1. September 2007.

[26] http://www.know-center.tugraz.at/

[27] http://www.know-center.tugraz.at/forschung/wissenserschliessung/projekte_wissenserschliessung/projekte_2004

### NewsML-G2

For easing the exchange of news, the International Press Telecommunication Council (IPTC) has developed the News Architecture for G2-Standards[28] whose goal is to provide a single generic model for exchanging all kinds of newsworthy information, thus providing a framework for a future family of IPTC news exchange standards. This family includes NewsML-G2, SportsML-G2, EventsML-G2, ProgramGuideML-G2 or a future WeatherML. All are XML-based languages used for describing not only the news content (traditional metadata), but also their management, packaging, or related to the exchange itself (transportation, routing).

### TV-Anytime

The TV-Anytime[29] Forum is an association of organizations which seeks to develop specifications to provide value-added interactive services, such as the electronic program guide, in the context of TV digital broadcasting. The forum identified the metadata as one of the key technologies enabling their vision and have adopted MPEG-7 as the description language. They have extended the MPEG-7 vocabulary with higher-level descriptors, such as, for example, the intended audience of a program or its broadcast conditions.

Within the TV-Anytime environment, the most visible parts of metadata are the attractors/descriptors used e.g. in Electronic Program Guides (EPG), or in Web pages to describe content. This is the information that the consumer, or intelligent agents, will use to search and select content available from a variety of internal and external sources. Another important set of metadata consists of describing user preferences, representing user consumption habits, and defining other information (e.g. demographics models) for targeting a specific audience.

## 3.3.6 Interoperability

Multimedia applications on the Web typically deal with media assets of mixed media types, which are indexed on the basis of strongly divergent metadata standards. This severely hampers the inter-operation of such systems. In [300] we have started to work on an multimedia metadata interoperability framework. We suggest that machine understanding of metadata coming from different applications is a basic requirement for the inter-operation of distributed multimedia systems. In [300], we present how interoperability among metadata, vocabularies and services is enhanced using Semantic Web technologies. In addition, it provides guidelines for semantic interoperability, illustrated by use cases.

Related approaches have been suggested in [174] as well as in [172], where the ABC model is proposed as a solution for multimedia metadata interoperability.

[28] http://www.iptc.org/NAR/
[29] http://www.tv-anytime.org/workinggroups/wg-md.html#docs

Chapter 4

# Semantic Web

*"Sui generis."*

(Latin phrase)

On xml.com, Dan Zambonini triggered quite a discussion in early December 2006 when he attempted to "Explaining the Semantic Web in 10 seconds"[1]

- *Web 1.0 was about connecting-up documents*
- *Web 2.0 is about connecting-up people.*
- *The Semantic Web is about connecting-up data.*

A week later I responded, stating that:

- *The Web (be it 2.0 or the "old" one :) is all about curiosity. It requires people to explore (=click on a link) the available space.*
- *The Semantic Web really is about laziness. You want some clever piece of software to take care of boring or otherwise non-exciting tasks for you.*

It is very likely that the truth is somewhere in the middle. However, as a matter of fact the Semantic Web has started to lift off, meanwhile.

This chapter explains the foundations of the Semantic Web ranging from its logical foundations to practical issues. We discuss the logic underpinning the Semantic Web Vision and the (in)famous Semantic Web stack. Further, we give a brief overview on accessing Semantic Web resources by means of querying and elaborate on Semantic Web vocabularies. Finally, we discuss recent developments, such as the recently so called "Web 3.0".

## 4.1 Logic and the Semantic Web

Bearing [225] in mind, one can state that "... the basic problem of knowledge representation (KR) [262] is the development of a sufficiently precise notation for representing knowledge. [...] we shall refer to any such notation as a *(knowledge) representation scheme*. Using such a scheme, one can specify a *knowledge base* consisting of *facts*."

[1] http://www.oreillynet.com/xml/blog/2006/12/explaining_the_semantic_web_in.html

### 4.1.1 Knowledge Representation

A logical representation scheme employs the notions of constant, variable, function, predicate, logical connective and quantifier to represent facts as logical formulas in some logic, as First Order Logic (FOL) or Higher Order Logic, Multivalued, Modal, or Fuzzy Logic. A **knowledge base** (KB) in this view is a collection of logical formulas which provides a partial description of a world. Levesque and Brachman [198] state that there are two major properties that KB have to satisfy. First the KB structures have to be expressions in language that has a *truth theory*. Secondly, the symbolic structures in the KB must play a causal role in the behaviour of the system; this should align with the understanding of them as propositions representing knowledge. The role of a KB is therefore to determine whether the truth of a sentence $\alpha$ is implicit in the KB, viz. answering the following question denoted as $KB \models \alpha$:

> Assuming the world is such that what is believed is true, is $\alpha$ also true?

Further, following Levesque and Brachman [198] it is possible to limit the expressive power of a KB. Taking FOL as a starting point and restricting the kind of information that can be represented yields a number of practical useful languages. Examples of such semantically restricted FOL forms are the database (relational) form, the logic-programming form, the semantic-network, or the frame-description form. In the following sections semantically restricted FOL, hence logic languages are examined. Two prevalent logic languages constitute the foundation of the Semantic Web: *Description Logics* and *Logic Programming*.

### 4.1.2 Description Logics (DL)

This section gives an overview on Description Logics (DL) following [15; 158; 167; 244]. It discusses the main features of DL and demonstrates how it can be used in a practical context. DL is the most recent name for a family of KR formalisms that represent the knowledge of an application domain by first defining the relevant concepts of the domain (its terminology), and then using these concepts to specify properties of objects and individuals occurring in the domain (the world description).

DL are descended from so-called "structured inheritance networks", which were introduced to overcome the ambiguities of early semantic networks and frames. The following three ideas, first put forward in Brachmans work on structured inheritance networks, have largely shaped the subsequent development of DLs:

- The basic syntactic building blocks are atomic concepts (unary predicates), atomic roles (binary predicates), and individuals (constants).
- The expressive power of the language is restricted in that it uses a rather small set of (epistemologically adequate) constructors for building complex concepts and roles.
- Implicit knowledge about concepts and individuals can be inferred automatically with the help of inference procedures. In particular, subsumption relationships between concepts and instance relationships between individuals and concepts play an important role: unlike *IS-A* links in Semantic Networks, which are explicitly introduced by the user, subsumption relationships and instance relationships are inferred from the definition of the concepts and the properties of the individuals.

Figure 4.1[2] sketches a knowledge representation system based on Description Logics. While the TBox introduces the vocabulary of a domain, the ABox contains assertions about individuals in terms of this vocabulary. The vocabulary consists of concepts that denote sets of individuals and roles, which denote binary relationships between individuals. Elementary descriptions are atomic concepts and atomic roles. Complex descriptions can be built from them inductively with concept constructors. Description Logics are distinguished by the constructors they provide. The language $\mathcal{AL}$ (= attributive language) has been introduced in [265] as a minimal language that is of practical interest; other DL languages are extensions of $\mathcal{AL}$.

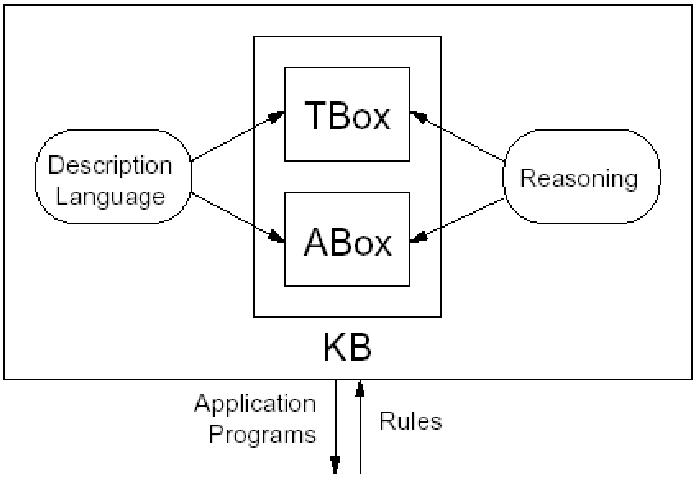



Figure 4.1: Architecture of a KR system based on Description Logics.

Next, we give a more formal account of a DL Knowledge Base, including a definition of TBox and ABox.

**Definition 4.1 (Description Logics (DL) Knowledge Base).**
A DL knowledge base $\Sigma$ is a tuple $\langle \mathcal{T}, \mathcal{A} \rangle$.

$\mathcal{T}$ denotes the Terminological Box or **TBox**. $\mathcal{T}$ is a finite set of statements called terminological axioms of the form

- $C \sqsubseteq D$ (general inclusion axiom) or $C \equiv D$ (general equivalence axiom), which is short for $C \sqsubseteq D$ and $D \sqsubseteq C$, where $C, D$ are concepts, and
- $R \sqsubseteq S$ (rule inclusion axiom) or $R \equiv S$ (rule equivalence axiom), which is short for $R \sqsubseteq S$ and $S \sqsubseteq R$, where $R, S$ are roles.

$\mathcal{A}$ being the Assertional Box or **ABox**. $\mathcal{A}$ is a finite set of statements—individual axioms—of the form $a : C$, called concept assertions or $\langle a, b \rangle : R$, called role assertions.

An interpretation $\mathcal{I}$ is said to satisfy $\Sigma$, viz. $\mathcal{I} \models \Sigma$, iff it satisfies $\mathcal{T}$ and $\mathcal{A}$. $\mathcal{I}$ consist of a non-empty set $\Delta^{\mathcal{I}}$—the domain of the interpretation—and an interpretation function, which assigns to every atomic concept A a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to every atomic role R a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

❖

[2]Credits: The Description Logic Handbook [13].

Note: The interpretation function is extended to concept descriptions by the inductive definitions listed in Table 4.1.

**Syntax and Semantic**. Following [13], Description Logics are usually given a Tarski-style[3] declarative semantics, which allows them to be seen as sub-languages of predicate logic. They are considered an important formalism unifying and giving a logical basis to the well known traditions of frame-based systems, semantic networks and KL-ONE-like languages, object-oriented representations, semantic data models, and type systems.

The basic building blocks are **concepts**, **roles** and individuals. Concepts describe the common properties of a collection of individuals and can be considered as unary predicates which are interpreted as sets of objects. Roles are interpreted as binary relations between objects. Each description logic defines also a number of language constructs (such as intersection, union, etc.) that can be used to define new concepts and roles.

The main reasoning tasks in DL are classification and satisfiability, subsumption and instance checking. Subsumption represents the is-a relation. Classification is the computation of a concept hierarchy based on subsumption.

In the following we give an demonstration of the expressivity of DL in an exemplary Knowledge Base.

```
1  Female ; atomic concept
2
3  Male ; atomic concept
4
5  Male ⊑ ¬Female ; females and males are disjoint
6
7  Human ⊑ Male ⊔ Female ; there are male and female humans
8
9  Man ≡ Human ⊓ Male ; men are male humans
10
11 Woman ≡ Human ⊓ Female ; women are female humans
12
13 hasMother ≡ isMotherOf⁻ ; bi-directional relationships
14
15 ⊤ ⊑≤ 1hasMother ; one can only have a single mother
16
17 hasMother ⊑ hasParent ; motherhood is a special kind of parenthood
18
19 mary : Woman ; mary is a woman
20
21 tim : Man ⊓ ∃hasMother.{mary} ; tim is a man and his mother is mary
```

Listing 4.1: A sample DL knowledge base.

***Example 4.1 (An Example DL Knowledge Base).***
An exemplary family is modelled in a DL knowledge base, called $\Sigma^{Fam}$. This is depicted in Listing 4.1 and Fig. 4.2 (page 59). The TBox of $\Sigma^{Fam}$—lines 1 to 17—states that there are two atomic concepts ($Female$, and $Male$), and some defined sublcasses. Further a role describing the motherhood relationship ($hasMother$) is defined in line 13 to 17. $\Sigma^{Fam}$'s ABox—lines 19 to 21—lists two individuals, namely $tim$ and his mother $mary$, along with their relation. Based on this DL knowledge base, we can now for example infer that $tim$ has a parent $mary$ and that $mary$ is a $Female$, etc. □

[3] http://plato.stanford.edu/entries/tarski-truth/

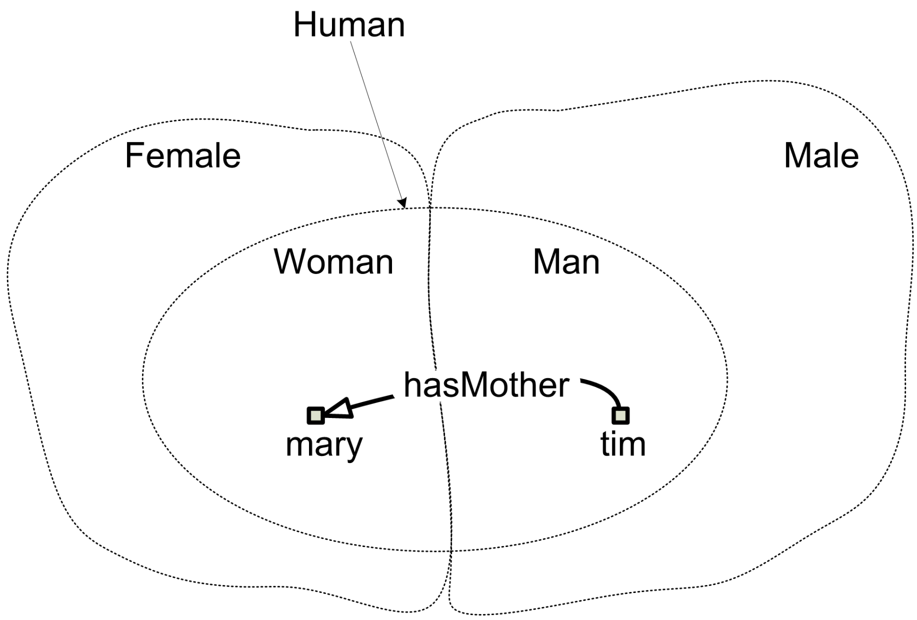



Figure 4.2: A sample DL knowledge base.

Some basics characteristics of DL-based languages—with special attention to their application on the Semantic Web—are listed, below.

*Unique Naming Assumption* (UNA). Again, based on [13], in DL we assume that distinct individual names denote distinct objects. Therefore, if $a, b \in \mathbf{I}$ are distinct individual names, it implies that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$. In contrast, according to [216] the semantics of, e.g., $\mathcal{ALC}$ does not require different symbols to be interpreted as different objects. However, the reader is invited to note that OWL [239, Section 4.2][4] does not use a UNA, per se. In the context of the Semantic Web this seems pretty straight-forward and natural. There may be different labels (URIs) for the same "thing" or for different purposes, still referring or representing the same entity. Note, that such a statement may well be valid only for a certain period of time. For example the two following URIs

- http://en.wikipedia.org/wiki/Kofi_Annan, and
- http://en.wikipedia.org/wiki/United_Nations_Secretary-General

certainly *did* refer to the same resource, namely the human being Kofi Atta Annan in the time frame from January 1, 1997 to January 1, 2007.

[4]For further examinations on OWL see also section 4.3.4

*Open vs. Closed World Assumption*. For a basic discussion on the Closed World Assumption (CWA) we refer the reader to [257]. DL Knowledge Bases have open-world semantics. If—as stated in [14]—we cannot deduce from an DL knowledge base that an individual $i$ is an instance of the concept $C$, we do not assume that $i$ belongs to $\neg C$. Put in the context of the Semantic Web, we note that OWL [239, Section 2]—as a DL—makes an *Open World Assumption* as well:

> That is, descriptions of resources are not confined to a single file or scope. While class `C1` may be defined originally in ontology `O1`, it can be extended in other ontologies. The consequences of these additional propositions about `C1` are monotonic. New information cannot retract previous information. New information can be contradictory, but facts and entailments can only be added, never deleted.

**Application**. Practical applications of DL have been suggested in [81] and [80]. A whole family of knowledge representation systems have been built using these languages and for most of them complexity results for the main reasoning tasks are known. Description logic systems have been used for building a variety of applications including conceptual modelling, information integration, query mechanisms, view maintenance, software management systems, planning systems, configuration systems, and natural language understanding. For a more detailed discussion on applications of DL-based languages see the following sections on the Semantic Web and its use cases.

Table 4.1[5] on page 61 summarises the syntax, the semantics, and applications of a sample DL language: $\mathcal{SHOIN}(D)$. As a matter of fact the Semantic Web is based on OWL [239], which is well known to be corresponding to a variant of $\mathcal{SHOIN}(D)$.

[5] Based on the *Description Logic Handbook* [13, Chapter 2], the *OWL Web Ontology Language Semantics and Abstract Syntax* specification [240], [117], and [165].

| Name | Syntax | Interpretation | Example | Alternative Name |
|---|---|---|---|---|
| | | *Constructors* | | |
| atomic concept | $A$ | $A^{\mathcal{I}}$ | $Human$ | class |
| atomic role | $R$ | $R^{\mathcal{I}}$ | $hasChild$ | property |
| individual | $o$ | $o \in \Delta^{\mathcal{I}}$ | $tim$ | instance |
| universal concept | $\top$ | $\Delta^{\mathcal{I}}$ | everything | top |
| bottom concept | $\bot$ | $\emptyset$ | the empty set | - |
| intersection | $C \sqcap D$ | $C^{\mathcal{I}} \cap D^{\mathcal{I}}$ | $Human \sqcap Male$ | - |
| union | $C \sqcup D$ | $C^{\mathcal{I}} \cup D^{\mathcal{I}}$ | $Doctor \sqcup Lawyer$ | - |
| complement | $\neg C$ | $\Delta^{\mathcal{I}} \setminus C$ | $\neg Male$ | negation |
| one of | $\{o_1\ o_2\}$ | $\{o_1^{\mathcal{I}},\ o_2^{\mathcal{I}}\}$ | $\{tim\ mary\}$ | nominals |
| exists | $\exists R.C$ | $\{a \in \Delta^{\mathcal{I}} \mid \exists b.(a,\ b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}$ | $\exists hasChild.Lawyer$ | existential quant. |
| for all | $\forall R.C$ | $\{a \in \Delta^{\mathcal{I}} \mid \forall b.(a,\ b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$ | $\forall hasChild.Doctor$ | value restriction |
| min. cardinality | $\geq n\ R.C$ | $\left\{a \in \Delta^{\mathcal{I}} \,\middle|\, \lvert\{b \in \Delta^{\mathcal{I}} \lvert (a,\ b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert \geq n\right\}$ | $\geq 2\ hasChild.Male$ | qual. number restriction |
| max. cardinality | $\leq n\ R.C$ | $\left\{a \in \Delta^{\mathcal{I}} \,\middle|\, \lvert\{b \in \Delta^{\mathcal{I}} \lvert (a,\ b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}\rvert \leq n\right\}$ | $\leq 4\ hasSister.Lawyer$ | qual. number restriction |
| has value | $\exists R.\{o\}$ | $\{a \in \Delta^{\mathcal{I}} \mid (a,\ o^{\mathcal{I}}) \in R^{\mathcal{I}}\}$ | $\exists hasMother.\{mary\}$ | role filler |
| | | *Axioms* | | |
| subclass | $C \sqsubseteq D$ | $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ | $Human \sqsubseteq Animal$ | - |
| same class | $C \equiv D$ | $C^{\mathcal{I}} = D^{\mathcal{I}}$ | $Man \equiv Human \sqcap Male$ | - |
| concept assertion | $C(o)$ | $o \in C^{\mathcal{I}}$ | $Man(tim)$ | - |
| disjoint | $C \sqsubseteq \neg D$ | $C^{\mathcal{I}} \cap D^{\mathcal{I}} = \emptyset$ | $Male \sqsubseteq \neg Female$ | - |
| same individual | $\{o_1\} = \{o_2\}$ | $\{o_1^{\mathcal{I}}\} = \{o_2^{\mathcal{I}}\}$ | $\{mary\} = \{mom\}$ | - |
| subproperty | $R \sqsubseteq P$ | $R^{\mathcal{I}} \subseteq P^{\mathcal{I}}$ | $hasDaughter \sqsubseteq hasChild$ | - |
| same property | $R \equiv P$ | $R^{\mathcal{I}} = P^{\mathcal{I}}$ | $hasMother \equiv hasMom$ | - |
| role assertion | $R(o_1,\ o_2)$ | $(o_1^{\mathcal{I}},\ o_2^{\mathcal{I}}) \in R^{\mathcal{I}}$ | $hasMother(tim,\ mary)$ | - |
| symmetric | $R \equiv R^{-}$ | $R^{\mathcal{I}} = \{(a,\ b) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (b,\ a) \in R^{\mathcal{I}}\}$ | $hasSibling \equiv hasSibling^{-}$ | - |
| inverse | $R \equiv P^{-}$ | $R^{\mathcal{I}} = \{(a,\ b) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (b,\ a) \in P^{\mathcal{I}}\}$ | $hasChild \equiv hasParent^{-}$ | - |
| transitive | $R^{+} \sqsubseteq R$ | $\{(a,\ b) \in R^{\mathcal{I}} \wedge (b,\ c) \in R^{\mathcal{I}} \rightarrow (a,\ c) \in R^{\mathcal{I}}\}$ | $ancestor^{+} \sqsubseteq ancestor$ | - |
| functional | $\top \sqsubseteq \leq 1\ R$ | $\{(a,\ b) \in R^{\mathcal{I}} \wedge (a,\ c) \in R^{\mathcal{I}} \rightarrow b = c\}$ | $\top \sqsubseteq \leq 1\ hasMother$ | - |
| inverse functional | $\top \sqsubseteq \leq 1\ R^{-}$ | $\{(b,\ a) \in R^{\mathcal{I}} \wedge (c,\ a) \in R^{\mathcal{I}} \rightarrow b = c\}$ | $\top \sqsubseteq \leq 1\ isMotherOf^{-}$ | - |

Table 4.1: Description Logics Axioms.

### 4.1.3 Logic Programming (LP)

The following section is based on [20] and gives an overview on what Logic Programming (LP) is and how it can be used in a practical context.

Declarative logic programs (LP) is the KR whose semantics underlies in a large part four families of rule systems: SQL relational databases, OPS5-heritage production rules, Prolog, and Event-Condition-Action rules. Recently, proposals for rules in the context of the Semantic Web have been made. As described by Boley and Kifer in the context of the RIF [43]:

> There are several equivalent ways to define first-order semantic structures. The one we adopted has the advantage that it generalizes to rule sets with negation as failure (NAF) and to logics for dealing with uncertainty and inconsistency. The difficulty is that some popular theories for NAF, such as the well-founded semantics, are based on three-valued semantic structures. Some popular ways to handle uncertain or inconsistent information (which is certainly important in the Web environment) rely on four-valued and other multi-valued logics. Therefore, following M. Fitting, Fixpoint Semantics for Logic Programming A Survey [101], we build our definitions to be compatible with future RIF dialects, which will be based multivalued logics.

The commonly used expressiveness of full LP includes features, notably negation-as-failure/priorities and procedural attachments, that are not expressible in First-Order-Logic (FOL). An ordinary—in [20] called general—logic program is a set of rules each having the form:

$$H \leftarrow B_1 \wedge \ldots \wedge B_m \wedge \sim B_{m+1} \wedge \ldots \wedge \sim B_n$$

where

- $H$ and $B_i$ are *atomic formulae*,
- $\sim$ is a logical connective called *negation as failure*[6] (cf. [58]),
- $\leftarrow$ is to be read as *if*, so that the overall rule should be read as "[head] if [body]",
- and $n \geq m \geq 0$.

The left-hand side of the rule is called the rule's head (or conclusion/consequent); the right-hand side is called the rules body (or premise/antecedent). Note that no restriction is placed on the arity of the predicates appearing in these atoms. Logical variables, and logical functions (with any arity), may appear unrestrictedly in these atoms.

A definite LP is an ordinary LP in which negation-as-failure does not appear, i.e., a set of rules each having the form:

$$H \leftarrow B_1 \wedge \ldots \wedge B_m$$

where $m \geq 0$.

[6] a logically non-monotonic form of negation whose semantics differs from the semantics of classical negation

Definite LP is closely related syntactically and semantically to the Horn fragment of FOL[7]. A clause in FOL has the form:

$$L_1 \vee \ldots \wedge L_k$$

where each $L_i$ is a classical literal.

A clause is said to be Horn when at most one of its literals is positive. A Horn clause is said to be definite when exactly one of its literals is positive. A definite Horn clause is also known as a Horn rule which thus can be written in the form:

$$H \leftarrow B_1 \wedge \ldots \wedge B_m$$

This Horn rule corresponds to the definite LP rule that has the same syntactic form, and vice versa. Horn-logic and Datalog[8] was studied in the area of deductive databases and logic programming and a number of efficient evaluation strategies are available. An important property of Datalog programs is that it is decidable whether a given query is logically entailed by a Datalog program. As described in [254], Horn clauses form the basis of Prolog. In Datalog, the (deductive) database consists of two parts, namely the extensional database (EDB), consisting of a set of facts, and the intensional database (IDB) consisting of a set of rules. Assuming that the sets of extensional and intensional predicates are disjoint means that no extensional predicate is allowed to occur in the head of a rule in IDB (cf. [302]).

Motivated by [312] the Table 4.2 below summaries different types of Logic Programming clauses.

| **Clause** | **Name** | **Type** |
|---|---|---|
| $H \leftarrow$ | (positive) unit clause | **fact** |
| $H \leftarrow B_1 \wedge \ldots \wedge B_n$ | definite program clause | **rule** |
| $\leftarrow B_1 \wedge \ldots \wedge B_n$ | definite goal clause | **query** |

Table 4.2: Terminology for LP clause types.

**Semantics**. There are two basic styles for defining the semantics of logic programs. The first is the model-theoretic semantics, in which the semantics of a program *P* is given by the minimal Herbrand model *MP*. The other style of semantics is the computational semantics, in which the semantics of a program *P* is given by the least fixpoint of the direct-consequence operator *TP*. A single application of *TP* yields all atoms that can be derived by a single application of some rule in *P* given the atoms in the interpretation *I*.

**Application**. Gelfond and Leone [108] give an overview on latest development in the application of Logic Programming and Knowledge Representation. Prolog has a long-standing and successful record on application fields and a widely deployed base, as for example XSB[9], SWI-Prolog[10], or even extensions to higher-order dimensions as HiLog[11]. From a practical point of view, LP-based systems are seen quite mature and scaleable; see for example [325].

[7] or Horn-clause logic

[8] Horn-logic only with 0-ary function symbols and additionally, no negation and only "safe" rules, i.e. all variables in the head also occur in the body.

[9] http://en.wikipedia.org/wiki/XSB

[10] http://www.swi-prolog.org/

[11] http://flora.sourceforge.net/aboutHiLog.php

### 4.1.4 Integrating DL and LP

Based on [83] and [117] we discuss how LP and DL can be aligned in the context of the Semantic Web. Due to the limitations of the expressive power of the OWL language family [166; 161] the need for a richer set of descriptions regarding properties emerges. Both OWL-DL and function-free Horn rules are decidable fragments of FOL with orthogonal expressive power. Combinations of OWL-DL and rules are desirable for the Semantic Web, however may lead to undecidability. In [217] a decidable combination is presented where rules are required to be DL-safe: each variable in the rule is required to occur in a non-DL-atom in the rule body.

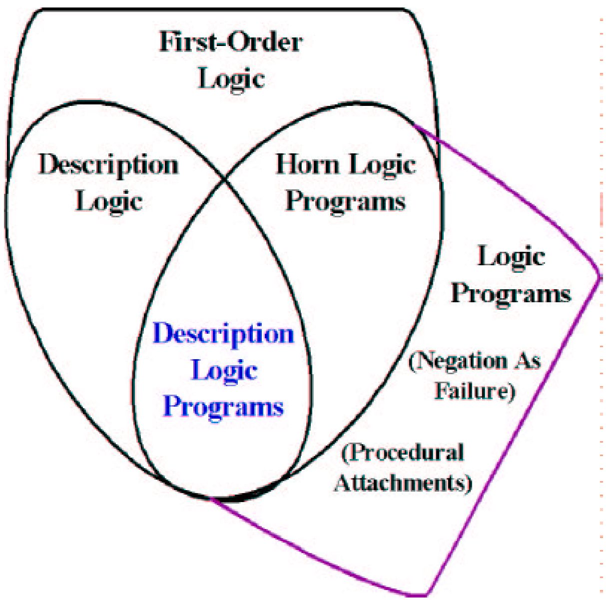



Figure 4.3: DL and LP overlapping.

Figure 4.3[12] illustrates the relationship between various knowledge representation languages and their expressive classes. DL and Horn are strict (decidable) subsets of FOL. LP, on the other hand, intersects with FOL but neither includes nor is fully included by FOL. According to [117], Description Logic Programs (DLP) is contained within the intersection of DL and LP. Full LP, including non-monotonicity and procedural attachments, can thus be viewed as including an "ontology sub-language", namely the DLP subset of DL.

Horrocks et.al. [161] recently suggested SWRL (Semantic Web Rules Language), a Horn clause rules extension to a specific DL dialect. A logical framework for handling rules and DL-systems together was presented in [103]. Reasoning support for SWRL is enabled through e.g. the translation of OWL-DL into the *Thousands of Problems for Theorem Provers* (TPTP) Problem Library [286] implemented in the Hoolet system[13], using a first order prover such as Vampire [259] to reason with the resulting first order theory. Note that recently a performance comparison between FO-prover and DL-Reasoner has been reported in [297]. In [288] more recently an extension of SWRL to handle function-free unary and binary first-order logic was proposed.

[12] By courtesy of [117]
[13] http://owl.man.ac.uk/hoolet/

## 4.2 Semantic Web Vision

In this section, we shortly review the history of the Semantic Web, further the current state of the Semantic Web and examine future directions. The original—and meanwhile adapted—intention of the Semantic Web is discussed, and relations to other fields are described.

### 4.2.1 Synopsis

The history of the Semantic Web really was starting with multimedia some 60 years ago. When Vannevar Bush [50] sketched the "Memex" and its functionality, he was actually describing a hypermedia system, though it took some 20 or more years to see tangible results (cf. Engelbart [90] and Nelson [231]). It was finally Sir Tim Berners-Lee who set the ball rolling when he implemented a very simple version of a hypertext system at CERN in the late 1980's, now known as the WorldWideWeb (WWW). In [32, p82], where Berners-Lee et al. summarise and introduce the original work, the future developments are expected to include

> [...] Evolution of objects from being principally human-readable documents to contain more machine-oriented semantic information, allowing more sophisticated processing.

While in a first wave between 1999 an 2001 the foundations were rolled out (such as RDF, community activities, etc.), a major revision has been done in 2004 (RDF Semantics, OWL, etc.). The World Wide Web Consortium (W3C) has originally[14] defined **Semantic Web** in its "Semantic Web Activity Statement" [287] in late 2001 as following.

> The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is the idea of having data on the Web defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications. The Web can reach its full potential if it becomes a place where data can be shared and processed by automated tools as well as by people.

### 4.2.2 Current State

The past four years (i.e., 2004–2008) have in many respects been utterly interesting regarding the Semantic Web. Firstly, important standards have been finalised: GRDDL (see page 90) and SPARQL (cf. page 73) became recommendations and XHTML+RDFa (page 91) is a W3C Candidate Recommendation at the time of writing. Secondly, many worthwhile activities, such as the ones from the Semantic Web Education and Outreach (SWEO) Interest Group, bring the Semantic Web closer to the people, hence closer to reality. In early 2007, the headline at the W3C Semantic Web Activity [287] reads:

> The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).

[14] http://www.w3.org/2001/sw/EO/points

The need for a Semantic Web has been motivated by the fact that "the mix of content on the web has been shifting from exclusively human-oriented content to more and more data content". The Semantic Web provides a solution by having data defined and linked in a way such that these techniques can be provided on top. According to Berners-Lee [31] the following requirements must be fulfilled by a Semantic Web:

- providing a common syntax for machine understandable statements,
- establishing common vocabularies,
- agreeing on a logical language,
- using the language for exchanging proofs.

At W3C there is currently (2007) a range of groups working in the Semantic Web Activity[15]; most importantly:

- *RDF Data Access Working Group*, to evaluate the requirements for an query language and network protocol for RDF and defined formal specifications and test cases for supporting such requirements.
- *Rules Interchange Working Group*, to produce a core rule language plus extensions which together allow rules to be translated between rule languages and thus transferred between rule systems.
- *Web Ontology Language (OWL) Working Group*, to produce a W3C Recommendation that refines and extends the 2004 version of OWL. The proposed extensions are a small set that: have been identified by users as widely needed, and have been identified by tool implementers as reasonable and feasible extensions to current tools.
- *Gleaning Resource Descriptions from Dialects of Languages Working Group*, to complement the concrete RDF/XML syntax with a mechanism to relate other XML syntaxes (especially XHTML dialects or "microformats") to the RDF abstract syntax via transformations identified by URIs.
- *Semantic Web Deployment Working Group*, to provide guidance in the form of W3C Technical Reports on issues of practical RDF development and deployment practices in the areas of publishing vocabularies, OWL usage, and integrating RDF with HTML documents.
- *Semantic Web Health Care and Life Sciences Interest Group*, to improve collaboration, research and development, and innovation adoption in the health care and life science industries. Aiding decision-making in clinical research, Semantic Web technologies will bridge many forms of biological and medical information across institutions.
- *Semantic Web Education and Outreach Interest Group*, collect proof-of-concept business cases, demonstration prototypes, etc, based on successful implementations of Semantic Web technologies, collect user experiences, develop and facilitate community outreach strategies, training and educational resources.

Currently, a shift from the academic-dominated to a more practical-rooted Semantic Web can be noticed. Triggered by the Web 2.0-success-stories, the Semantic Web community has become interested into solving real-world problems, hence in finding questions to the answers it has been given, so far.

[15] http://www.w3.org/2001/sw

### 4.2.3 Future

Ora Lassila wrote a follow-up to a blog entry[16] in late 2006 w.r.t. the ongoing Web 3.0 debate where he states:

> Reading the aforementioned blog entry[17], I feel it misunderstands the Semantic Web w.r.t. the term "Strong AI" [...] This misunderstanding may come from the first sentence of the definition of the term "Strong AI" in Wikipedia[18] which claims that Strong AI is the belief that "some forms of AI can truly reason and solve problems". [...]"Weak AI", on the other hand, treats AI as a field of computer science, with AI techniques as useful additions to the computer scientists arsenal of techniques that can be used when building software. If the Semantic Web is about AI [...], it is about Weak AI—even the original SciAm article does not make claims about Strong AI.

Very likely the next five to ten years will decide if and in which form the Semantic Web will be realised. It seems that grass-rooted approaches in metadata handling[19] gaining more and more popularity will be the first places where Semantic Web technologies can demonstrate their advantages. However, it remains unclear if the more complex parts of the Semantic Web (such as ontologies and rules) will ever be dominating the Semantic Web outside the researcher's labs. The interested reader is invited to note that we discuss the so called "Web 3.0" in greater detail below; cf. section 4.6.

### 4.2.4 Related Fields

**Semantic Web vs. Artificial Intelligence**. Indeed, the Semantic Web utilises Knowledge Representation techniques and applies them. The Semantic Web can be seen as a representative of the weak AI branch as Lassila and Hendler [194] put it:

> From one viewpoint, the Semantic Web is the symbiosis of Web technologies and knowledge representation (KR), which is a subfield of artificial intelligence (AI) concerned with constructing and maintaining (potentially complex) models of the world that enable reasoning about themselves and their associated information. As such, we can understand the Semantic Web through the lessons learned from the Webs development and adoption, as well as (perhaps somewhat painfully) from the deployment of AI technologies.

**Semantic Web vs. Relational Data**. One of the cornerstones of the Semantic Web—Description Logics—have already been investigated in the light of relational data. A logic view on relational data has been reported in [258]; a common view on logical languages and their representation (incl. the relational view) has been given by [198].

A more down-to-earth approach recently came up: To enable RDMBS to enter the Semantic Web by using SPARQL, so called SPARQL-endpoints[20].

[16] http://www.lassila.org/blog/archive/2006/11/microformats_we.html
[17] http://microformatique.com/?p=57
[18] http://en.wikipedia.org/wiki/Strong_ai
[19] such as http://machinetags.org/
[20] http://esw.w3.org/topic/SparqlEndpoints and http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/

## 4.3 Semantic Web Stack

This section presents the proposed architecture of the Semantic Web [34] in terms of the Semantic Web Stack. Since the Semantic Web is still in genesis, the current State-of-the-Art and possible directions are discussed herein.

Semiotics, following [267] is the general science of signs and how their meaning is transmitted and understood. A sign is generally defined as something that stands for something else. Semiotics is composed of three fundamental components:

- **Syntax**. It deals with the formal or structural relations between signs and the production of new ones when exchanging data.
- **Semantic**. The study of relations between the system of signs and their meanings:
  - Implicit semantics. The kind that is implicit in data and that is not represented explicitly in any machine processable syntax.
  - Formal semantics. Semantics that are represented in some well-formed syntactic form.
  - Powerful (soft) semantics. The ability to represent and utilize knowledge that is imprecise, uncertain, partially true, and approximate is lacking, at least in the base/standard models.
- **Pragmatic**. According to [121], the "relation of signs to (human) interpreters". If the effect is different from the senders initial intention a pragmatic disturbance occurs. Pragmatics relate to some extend to usability.

To explain the Semantic Web, its component and interdependencies, very often a stack or a layered model[21] is utilised. The Fig. 4.4 on page 69 depicts the layering of Semantic Web standards as seen by the W3C[22] in early 2007. Note, however that the proposed layering is not free of any difficulties; we discuss some of them at the end of this section, in 4.3.6.

The Semantic Web architecture grounds itself on available standards for referring to entities, viz. Uniform resource identifiers (URIs), and encoding of character symbols, i.e. Unicode (The Unicode Consortium, 2003). It reuses existing Web technologies, such as the eXtensible Markup Language (XML) for syntactic purposes. At the core of the stack is an integrative data model that scales well on the Web-level. On top of it are vocabularies and rules to be found; the top layers providing Proof and Trust are starting to be addressed by research, nowadays. Consequently, we cannot give an explicit account of these layers but only describe their intention. According to [31], the ability to check the validity of statements made in the (Semantic) Web is important. Therefore the creators of statements should be able to provide a proof of correctness of the statement which is verifiable by a machine. At this level, it is not required that the machine that reads the statements finds the proof itself, it "just" has to check whether the proof provided by the creator is feasible enough to trust the provided statements. In the following sections, a closer look at each of the layers is made and the according technologies are discussed.

[21] analogous to the OSI network reference model
[22] http://www.w3.org/2007/03/layerCake.png

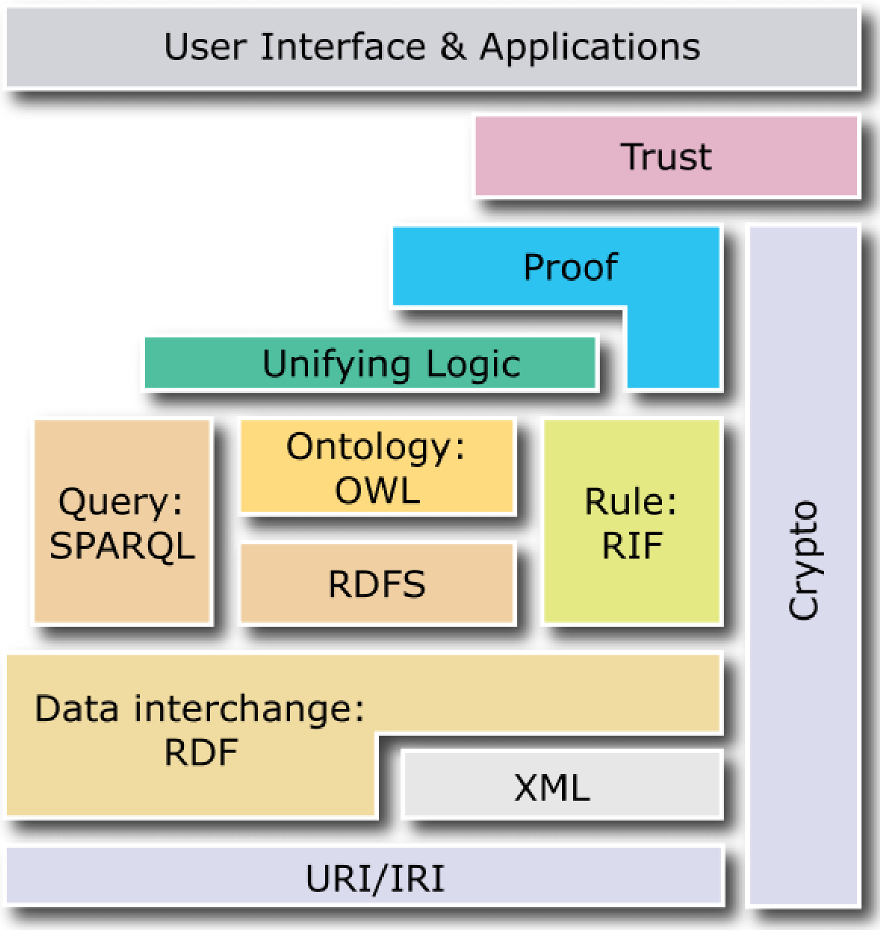



Figure 4.4: The W3C Semantic Web Stack.

### 4.3.1 Encoding & Addressing

Computers usually process binary sequences (the Bits) as `10110010`, while humans don't. Human beings work with symbols, e.g. "A" or "5". To support both worlds thus enabling humans to work with low-level computer representations, Unicode[23] is utilised. Unicode provides a set of mappings between numbers and assigned symbols (the what) that—together with the Unicode Transformation Format (UTF) schemes—to allow symbols from all of the writing systems of the world to be consistently represented.

When talking about things (in special resources), one has to name them. Addressing[24] is another word for naming things uniquely [176]. A Uniform Resource Identifier (URI) [33] plays the role of naming things on the Web[25]. Berners-Lee, in [32], states beneath other things that URIs are strings used as addresses of objects on the Web. The introduction of URIs[26] actually was a key factor for the success of the Web; it allowed for a uniform addressing, hence treatment of different protocols (FTP, NNTP, but also the back then rivaling WAIS or Gopher system).

A joint W3C/IETF note on URIs[27] stated that during the early years of discussion of Web identifiers (early to mid 90s), people assumed that an identifier type would be cast into one of two or possibly more classes. An identifier might specify the location of a resource—a URL [35]) or its name—a URN—independent of the location. There was discussion about generalizing this by addition of a discrete number of additional classes; for example, a URI might point to metadata rather than the resource itself, in which case the URI would be a Uniform Resource Citation (URC)[28]. URI space was thus viewed as partitioned into subspaces: URL and URN, and additional subspaces, to be defined. An excellent overview on the relation of URIs, URLs, and URNs has recently been written by Dan Connolly (cf. "Untangle URIs, URLs, and URNs"[29]).

A simplified depiction of the relations between URIs, URLs and URNs is shown in Fig. 4.5 on page 71.

Internationalized Resource Identifiers (IRIs) [87]—a recent IETF RFC—extend the syntax of URIs to a much wider repertoire of characters. Further, IRIs define "internationalized" versions corresponding to other constructs from [33], such as URI references.

Regarding the role of URIs in the Semantic Web the interested reader is invited to follow the ongoing standardisation activities of the W3C Semantic Web Deployment Working Group (SWD WG). At the time of writing of this thesis the SWD-WG[30] is working on "Best Practice Recipes for Publishing RDF Vocabularies"[31], which elaborates on dereferencing issues. Further, the W3C Technical Architecture Group (TAG) has discussed the so called *http-range-14* [118] issue; this TAG finding describes the nature of HTTP URIs. A document addressing these issues in greater detail is available elsewhere [263].

---

[23] `http://www.unicode.org/`
[24] `http://www.w3.org/Addressing/`
[25] The reader is referred to design aspects of URIs `http://www.w3.org/Provider/Style/URI`
[26] See also [32] for a discussion of URIs and the URLs.
[27] `http://www.w3.org/TR/uri-clarification/`
[28] `http://www.ukoln.ac.uk/metadata/desire/overview/rev_22.htm`
[29] `http://www-128.ibm.com/developerworks/xml/library/x-urlni.html`
[30] The author of this thesis is a member of the SWD and has also contributed to URI issues by reviewing, etc.
[31] `http://www.w3.org/TR/swbp-vocab-pub/`

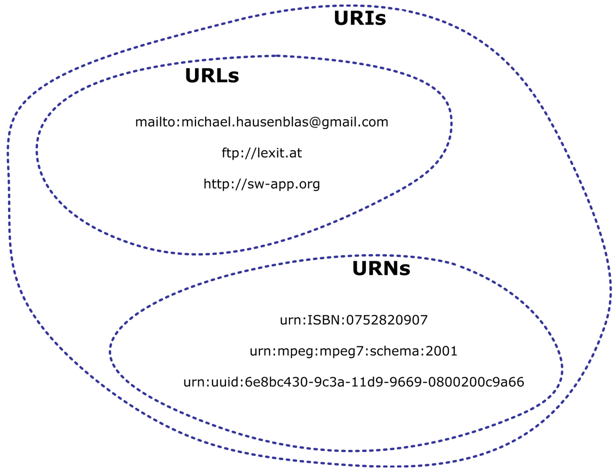



Figure 4.5: URIs, URLs and URNs.

### 4.3.2 Data Structure and Exchange

Based on the explanation of the basic encoding and addressing used on the Semantic Web, we now proceed to the next layer. This data structure and exchange layer at its core is a simple tree-based model enabling the interoperable exchange of structured data. The proposed standard for exchanging information on the Semantic Web is the *eXtensible Markup Language* (XML) [332]. XML intends to specify a syntactic encoding of documents by defining a set of rules to parse textual documents as labelled trees. Documents that conform to the syntactic rules of XML are called *well-formed*. In section 4.3.6 we discuss the potential problems with XML, especially the layering issue regarding RDF.

The XML Namespace specification [45] has been devised to distinguish the names of different applications. Each name is prefixed with a namespace, which must be declared as an attribute in one of the ancestor elements; typically this is done in the root element. The namespace declaration assigns an URI fragment to a prefix expanding with the namespace URI when the XML document is parsed.

XML itself is based on the simple idea of representing documents as trees. However, many applications can only make use of special trees, viz. documents that are valid by conforming[32] to a predefined structure and vocabulary. Tree grammars are used to restrict the structure of a document.

To define tree grammars, there exist several ways. The Document Type Definition (DTD)[33] is a mature but limited way of defining the grammar. XML Schema [333] is more powerful than DTD since it supports a kind of "subclassing"—by resctricting or by extending a base type—and provides a set of atomic data types, such as integers, strings, dates, etc. in addition to character data. An alternative to XML Schema is Relax NG [56], a more lightweight but equally powerful XML constraints language.

[32] For a discussion on XML schema languages cf. http://www.daimi.au.dk/~fagidiot/thesis/.

[33] http://www.isgmlug.org/sgmlhelp/g-index.htm

### 4.3.3 Data Model

While the Web is composed of documents that are interlinked, the Semantic Web is actually an interlinked data Web. In the following we describe how the data is represented formally, and even more important, how the interlinking takes place; the explanation is based on an example from [3].

**Statements**. The basic unit of information is called a statement. To enable automation the *Resource Description Framework* (RDF) [203] defines a structure for these statements. A statement is formally called a triple, meaning that it is made up of three components; the subject of the triple what we are making our statements about, the second part of a triple is called the property (of the subject), and the final part of a triple is called the object. To illustrate this concept the following statements are assumed.

```
Albert was born on March 14, 1879, in Germany.
A picture of him at is available at the URL
http://en.wikipedia.org/wiki/Image:Albert_Einstein_Head.jpg.
```

When the above statements are rearranged and rewritten, an RDF graph as depicted in the listing 4.2 may be the result.

```
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf:  <http://xmlns.com/foaf/0.1/> .
@prefix dbpp: <http://dbpedia.org/property/> .

<http://dbpedia.org/resource/Albert_Einstein>
 foaf:name "Albert Einstein" ;
 dbpp:dateOfBirth "March 14, 1879" ;
 dbpp:birthPlace <http://dbpedia.org/resource/Germany> ;
 foaf:depiction
     <http://en.wikipedia.org/wiki/Image:Albert_Einstein_Head.jpg> .
```

Listing 4.2: Example RDF statements.

RDF has a simple data model that is easy for applications to process and manipulate [186]. The data model is independent of any specific serialization syntax[34] [27]; “model” used herein has a completely different sense to its use in the term “model theory”; cf. section 4.3.4.

Before going into details of defining what an RDF graph is, we elaborate on some important terms, needed for a sound understanding of the actual definition.

**URI reference (URIref)**. Technically, a URIref in an RDF graph is a Unicode string that (i) does not contain any control characters[35], and (ii) would produce a valid URI [33] character sequence representing an absolute URI[36] with optional fragment identifier. URIrefs are used for disambiguation purposes; this may happen on the schema level (see section 4.3.4, below) or to uniquely identify objects in metadata statements. For example, in listing 4.2 the statement `Albert was born in Germany` was rewritten using URIrefs. To uniquely state that we are talking about *the* Albert Einstein, rather than using the string “`Albert Einstein`”, we use the URIref `<http://dbpedia.org/resource/Albert_Einstein>`, etc.

[34] For an excellent synopsis on RDF serialisation syntax, the read is referred to [104, Appendix A].

[35] i.e., it does not contain #x00 - #x1F, #x7F-#x9F

[36] For a discussion on URI equivalence, the interested reader is referred to http://www.w3.org/2001/tag/issues.html#URIEquivalence-15.

**Blank node (bNode)**. The bNodes[37] in an RDF graph are drawn from an infinite set. This set of blank nodes, the set of all URIrefs and the set of all literals (see below) are pairwise disjoint; otherwise the set of blank nodes is arbitrary. The concept of bNodes is useful when talking about things that need no (global) identity. However, there are good reasons to assign URIs to the parts of a graph that are intended to be publicly available[38].

**Literals**. Literals may contain one or two named components. The *lexical form*—being a Unicode string[39]—is common to all literals, where **plain literals** may have an optional language tag, normalized to lowercase, and **typed literals** a datatype URI (a URIref). For a detailed discussion on datatypes, the reader is referred to section 4.3.6. With the above discussion, we can now define a RDF Graph formally, as follows.

**Definition 4.2 (RDF Graph).**
An *RDF graph* is a set of triples, where each triple contains three components:

- a subject, which is an URIref or a bNode,
- an object, which is an URIref, a literal or a bNode,
- a predicate that denotes a relationship[40], which is an URIref. ❖

The assertion of an RDF triple says that some relationship, indicated by the predicate, holds between the things denoted by subject and object of the triple. The assertion of an RDF graph amounts to asserting all the triples in it, so the meaning of an RDF graph is the conjunction (logical AND) of the statements corresponding to all the triples it contains. A formal account of the meaning of RDF graphs has been given in [150].

### 4.3.4 Ontologies, Rules & Query

One of the first well-known work on ontology engineering was reported by Uschold and Gruninger [307]. In this paper they discuss a methodology for developing and evaluating ontologies, discussing both informal and formal techniques. Note that the relation between a DL Knowledge Base (cf. section 4.1.2 on page 56) and ontologies can be stated as follows: *Every DL-knowledge base is representing an ontology, but not every ontology is necessarily represented using DL*—alternative ways to represent an ontology may be F-Logic, etc.

Fensel discusses the importance of ontologies in [96] where he states that "ontologies provide a shared and common understanding of a domain that can be communicated between people and application systems".

**RDF-S**. A simple way of expressing a formal vocabulary is RDF Schema (RDF-S) [255]. RDF provides a way to express simple statements about resources. However there exists a need to define the vocabularies, which can be used in those statements. Following the RDF Primer [203, Section 5], RDF-S provides the facilities needed to describe classes and properties, and to indicate which classes and properties are expected to be used together, that is RDF-S provides a type system for RDF. The RDF-S type system is similar in some

[37] RDF makes no reference to any internal structure of bNodes, but given two bNodes, it is must be possible to determine whether or not they are the same.

[38] For example through Semantic Web indexer such as sindice.com.

[39] In Normal Form C; cf. http://www.unicode.org/unicode/reports/tr15/

[40] The direction of the arc is significant; it always points from the subject to the object.

respects to the type systems of object-oriented programming languages such as Java. However, RDF classes and properties are in some respects very different from programming language types. RDF class and property descriptions do not create a straightjacket into which information must be forced, but instead provide additional information about the RDF resources they describe.

The RDF-S terms are themselves provided in the form of an (predefined) RDF vocabulary; the resources in the RDF-S vocabulary have URIrefs with the prefix `http://www.w3.org/2000/01/rdf-schema#`

**OWL**. The Web Ontology Language (OWL) [240; 241; 243; 239] is the W3C recommendation, viz. the standard for representing ontologies on the Semantic Web. In [159] the relation between OWL (DL and Lite) ontology languages to certain description logics is explained in detail.

In the Fig. 4.6 we show how a very simple multimedia ontology can be represented using RDF-S and OWL.

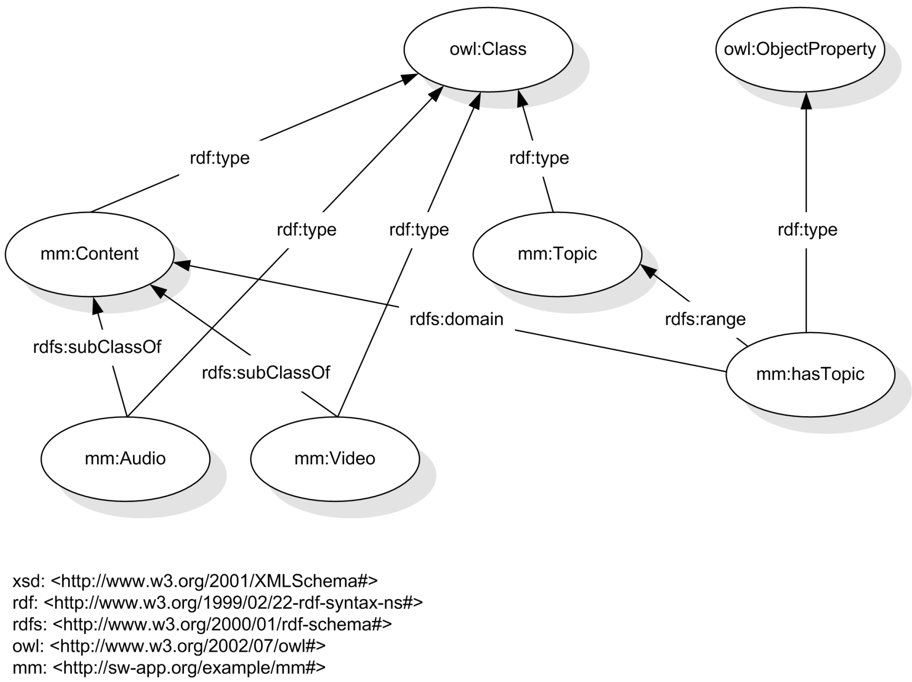



Figure 4.6: A simple multimedia ontology in RDF-S/OWL.

**Rules**. Nowadays research usually considers the ontology and the logic levels together, as any semantic specification like ontologies has to be grounded in logic. As we have seen in the previous section 4.1, languages such as OWL do not only specify a vocabulary—hence constrain the use of that vocabulary by restrictions—but also provide axioms allowing to deduce new information from explicit information.

However, OWL does currently not allow the definition of general rules over properties. For example, one cannot express property chaining in OWL. Since there is currently no consensus on how a rule layer could look like, W3C decided not to standardise rules itself, but rather their exchange—cf. the Rule Interchange Format (RIF) [260].

In the following, earlier proposals regarding the design of the rule layer are listed.

- *XML-based approaches*, such as RuleML[41] relying on specialized XML vocabularies to create various types of logic programs and their underlying Knowledge Bases;
- *RDF-based approaches*. Triple and N3 are the most prominent examples of this type. The latter one is for example used in CWM[42]. Triple reasons with RDF data in a frame-based syntax, which has been inspired by F-Logic [183]. Triple programs are based on Horn logic and are compiled into Prolog programs. Unlike F-Logic, Triple does not have a fixed semantics for object-oriented features like classes and inheritance.
- *OWL-based approaches*. In this approach, DL-based systems and rule systems are used together; cf. section 4.1.4.

**Query**. Without doubt it is necessary to have sound and interoperable representations of the information on the Semantic Web. However, after reviewing the KR-based issues in the previous sections, one might be tempted to ask: *How to make use of it?*

This is a very valid and obvious question; indeed all the RDF-based metadata, all the ontologies and rules are of little use if they are not accessible through the Network. Therefore, standardised protocols are needed that allow for accessing and query RDF-based data sets. The W3C RDF Data Access Working Group noted in their *RDF Data Access Use Cases and Requirements* [57]:

> Despite the lack of standards, developers in commercial and in open source projects have created many query languages for RDF data[43]. But these languages lack both a common syntax and a common semantics. In fact, the extant query languages cover a significant semantic range: from declarative, SQL-like languages, to path languages, to rule or production-like systems. The existing languages also exhibit a range of extensibility features and built-in capabilities, including inferencing and distributed query.

On the Semantic Web documents enriched by annotations about the documents, as well as machine interpretable statements capturing some of the meaning of the documents content are very likely to be available. In [208] a framework for integrating search and inference in this setting has been proposed. It supports both retrieval-driven and inference-driven processing, using text and markup as indexing terms, and exploiting Web search engines, thus tightly binding retrieval to inference.

The SPARQL Protocol and RDF Query Language (SPARQL) [253] is the W3C standard for querying RDF. The specification defines the syntax and semantics of the SPARQL query language for RDF. SPARQL can be used to express queries across diverse data sources; it contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be (XML or JSON[44]) results sets or RDF graphs. In the following example the usage of SPARQL is demonstrated (see also Fig. 7.2 on page 128): It asks for videos with a certain topic (news).

[41] http://www.ruleml.org/
[42] http://www.w3.org/2000/10/swap/doc/
[43] http://www.w3.org/2001/11/13-RDF-Query-Rules/
[44] JavaScript Object Notation, see http://www.json.org/

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
PREFIX mm:  <http://www.sw-app.org/examples/multimedia#> .
PREFIX mex:  <http://www.sw-app.org/examples/multimedia/abox1#> .

SELECT
 *
FROM
 <http://sw.joanneum.at/sandbox/m3-sample-onto.rdf>
WHERE {
 ?video rdf:type mm:Video ;
        mm:hasTopic mex:News .
}
```

Listing 4.3: Example SPARQL query.

Executing the SPARQL Query from listing 4.3 yields the following binding

`?video` = `<http://www.sw-app.org/examples/multimedia/abox1#v2>`.

### 4.3.5 Trust & Data Provenance

The notion of context is widely studied in different areas of artificial intelligence (AI). It became a popular issue only in the late 1980s, when J. McCarthy proposed to formalise context as a possible solution to the problem of generality:

> When we take the logic approach to AI, lack of generality shows up in that the axioms we devise to express common sense knowledge are too restricted in their applicability for a general common sense database ... Whenever we write an axiom, a critic can say that the axiom is true only in a certain context. With a little ingenuity the critic can usually devise a more general context in which the precise form of the axiom doesn't hold.

When discussing context in an interdependent, interconnected environment as the Semantic Web [38], two important aspects immediately arise: data provenance and trust. Deciding which among many possibly inconsistent sources is most reliable is a challenging task. In [78] an approach has been proposed to agent knowledge outsourcing inspired by the use trust in human society. Two important practical issues are discussed: learning trust and justifying trust. An agent can learn trust relationships by reasoning about its direct interactions with other agents and about public or private reputation information, i.e., the aggregate trust evaluations of other agents.

In [234] common strategies of trust are highlighted and their costs and benefits w.r.t. implementation are discussed. On the one hand the focus is on technology-driven areas (such as digital signatures, certificates, etc.), on the other hand the human factor (social networking). In [111] an approach for integrating the two to build a Web of trust in a more social respect is described.

As the Semantic Web is an open-world system, everybody is free to make statements about everything. Potential difficulties of data abuse are likely. Hence, *data provenance*[45] is an important issue on the Semantic Web.

[45] http://www.ibm.com/developerworks/xml/library/x-rdfprov.html

Dan Connolloy recently took "A look at emerging Web security architectures from a Semantic Web perspective"[46]. First trust-based implementations—mostly based on FOAF, OpenID[47], etc.—are emerging, such as the *Beatnik Address Book*[48].

### 4.3.6 Semantic Web Issues

The Semantic Web stack as introduced above has some serious flaws. Early discussions can be traced back to the standardisation of OWL [245; 166]:

> OWL was not designed in a vacuum. [...] As OWL is an effort in W3Cs Semantic Web activity, it had to fit into the Semantic Web vision of a stack of languages including XML and RDF. As OWL is supposed to be an ontology language, it had to be able to represent a useful group of ontology features.
>
> [...]
>
> The multiple influences on OWL resulted in some difficult trade-offs. Also, and somewhat surprisingly, considerable technical work had to be performed to devise OWL in such a way that it could be shown to have various desirable features, while still retaining sufficient compatibility with its roots.

In the following we take a closer look at the two important issues nowadays, namely (i) layering issues, and (ii) data-type issues.

#### Layering Issues

Criticism on how to layer logic on top of RDF have been discussed in the past years, see for examples [160]. Pan has proposed a layering [244] of OWL onto a DL-ised version of RDFS—called RDFS(FA)—addressing most of the issues, such as the built-in semantics of RDF triples, non-existing restrictions on how to use the built-in vocabulary, etc.

A radical new approach to the layering-problem is performed within the WSMO (Web Service Modeling Ontology) project [327], initiated and driven by DERI, the Digital Enterprise Research Institute. The Web Service Modeling Language (WSML)—due to a position statement[49]—provides a framework of different language variants to describe semantic Web services. WSML is a frame based language with an intuitive human readable syntax and XML and RDF exchange syntaxes, as well as a mapping to OWL. It provides different variants, allowing for open and closed world modelling; it is a fully-fledged ontology and rule language with defined variants grounded in well known formalisms, namely Datalog, Description Logic and Frame Logic. Taking the key aspects of WSML as a starting point, we rationalize the design decisions which we consider relevant in designing a proper layering of ontology and rule languages for the Semantic Web and semantic Web services.

#### Data-type Issues

In **XML Schema**, datatypes[50] are defined as 3-tuples, consisting of (i) a set of distinct values (*value space*), (ii) a set of lexical representations (*lexical space*), and (iii) a set of facets that

[46] http://www.w3.org/2006/03dc-aus-lga/swauth
[47] http://openid.net/
[48] http://blogs.sun.com/bblfish/entry/beatnik_change_your_mind
[49] http://www.w3.org/2004/12/rules-ws/paper/44/
[50] http://www.w3.org/TR/xmlschema-2/

characterize properties of the value space, individual values or lexical items. Datatypes are either built-in datatypes or user-derived datatypes.

**RDF(S)** defines datatypes[51] and data values, viz. allows for using an external type system (as XML Schema): A datatype consists of a *lexical space* (a set of Unicodes[52] strings), a *value space* and a *lexical-to-value mapping* (a total mapping from the lexical to the value space). Further *literals*[53] are used to identify values such as numbers and dates by means of a lexical representation; a *plain literal* is a self-denoting string with an optional language tag, whereas a *typed literal* is a string combined with a datatype URI. It denotes the member of the identified datatype's value space obtained by applying the lexical-to-value mapping to the literal string.

Finally **OWL**[54] reuses many of the built-in XML Schema datatypes and is aligned with the RDF(S) datatypes concepts, though a critical difference between OWL and RDF(S) datatyping concerns the relation between datatypes and classes. In OWL-DL, object and datatype domains are disjoint (cf. also [165]).

Wang et al. [320] recently gave an analysis the way that natural languages handle continuous quantities. They propose a general semantics based on metric spaces, and describe how to treat semantic values computationally.

[51] http://www.w3.org/TR/rdf-concepts#section-Datatypes
[52] http://www.w3.org/TR/rdf-concepts/#ref-unicode
[53] http://www.w3.org/TR/rdf-concepts/#section-Literals
[54] http://www.w3.org/TR/owl-semantics/direct.html#3.1

## 4.4 Semantic Web Vocabularies

In this section predominant vocabularies on the Semantic Web are discussed. They do not necessary focus on multimedia issues, although most of them are used as well in the context of a SWMA. We roughly differentiate into (i) generic vocabularies, (ii) social vocabularies, (iii) spatio-temporal vocabularies, and (iv) other vocabularies in the following overview.

### 4.4.1 Generic Vocabularies

In the following, vocabularies are discussed that do not focus on a specific domain, but rather provide for a generic vocabulary framework.

#### Simple Knowledge Organisation Systems–SKOS

The *Simple Knowledge Organisation System* (SKOS) [210; 211] is a model for expressing the basic structure and content of concept schemes, such as thesauri, classification schemes, subject heading lists, taxonomies, "folksonomies", and other types of controlled vocabulary.

The SKOS Core Vocabulary is an RDF application. Using RDF allows data to be linked to and/or merged with other RDF data by Semantic Web applications. In practice, this means that data sources can be distributed across the web in a decentralised way, but still be meaningfully composed and integrated by applications, often in novel and unanticipated ways. The SKOS Core Vocabulary is a set of RDF properties and RDF-S classes, that can be used to express the content and structure of a concept scheme as an RDF graph.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.my.com/#canals">
    <skos:definition>A feature type category for places such as the Erie
        Canal</skos:definition>
    <skos:prefLabel>canals</skos:prefLabel>
    <skos:altLabel>canal bends</skos:altLabel>
    <skos:altLabel>canalized streams</skos:altLabel>
    <skos:altLabel>ditch mouths</skos:altLabel>
    <skos:altLabel>ditches</skos:altLabel>
    <skos:altLabel>drainage canals</skos:altLabel>
    <skos:altLabel>drainage ditches</skos:altLabel>
    <skos:broader
        rdf:resource="http://www.my.com/#hydrographic\%20structures"/>
    <skos:related rdf:resource="http://www.my.com/#channels"/>
    <skos:related rdf:resource="http://www.my.com/#locks"/>
    <skos:related
        rdf:resource="http://www.my.com/#transportation\%20features"/>
    <skos:related rdf:resource="http://www.my.com/#tunnels"/>
    <skos:scopeNote>Manmade waterway used by watercraft or for drainage,
        irrigation, mining, or water power</skos:scopeNote>
  </skos:Concept>

</rdf:RDF>
```

Listing 4.4: A sample SKOS document.

A simple application of SKOS is given in the listing 4.4[55].

[55] Based on http://www.xml.com/pub/a/2005/06/22/skos.html

| Content | Intellectual Property | Instantiation |
|---|---|---|
| Coverage | Contributor | Name |
| Description | Creator | Date |
| Type | Publisher | Format |
| Relation | Rights | Identifier |
| Source | | Language |
| Subject | | |

Table 4.3: Dublin Core Elements.

### Dublin Core–DC

The Dublin Core Metadata scheme [70] is a set of 15 elements and each element has a limited set of attributes. It was initially created as a library scheme, favouring therefore documents as objects but it has since then been used in other applications as well. The idea of Dublin Core is to be kept as small and simple as possible to allow a non-specialist to create simple descriptive records. Some of the Dublin Core metadata content is controlled by defined vocabularies to improve search results. In other cases, ontologies have been used to solve the interoperability problem. In table 4.3 the 15 elements are listed:

The Dublin Core Initiative has also issued a list of recommended Qualifiers that can be divided into two groups: element refinement and encoding scheme qualifiers. The first make the element meaning narrower or more specific and the latter include controlled vocabularies and formal notations from where element values are being selected.

### Extensible Metadata Platform–XMP

The main goals of the *Extensible Metadata Platform* (XMP)[56] are to attach more powerful metadata to media assets in order to enable better management of multimedia content, to allow better ways to search and retrieve content and thus to improve consumption of assets. Furthermore XMP aims to enhance reuse and repurposing of content and to improve interoperability between different vendors and systems. The Adobe XMP specification standardizes the definition, creation, and processing of metadata by providing a data model, storage model (serialization of the metadata as a stream of XML), and formal schema definitions (predefined sets of metadata property definitions that are relevant for a wide range of applications). XMP makes use of RDF in order to represent the metadata properties associated with a document.

With XMP, Adobe provides a method and format for expressing and embedding metadata in various multimedia file formats. It provides a basic data model as well as metadata schemas for storing metadata in RDF, and provides storage mechanism and a basic set of schemas for managing multimedia content like versioning support etc.

The most important components of the specification are the data model and the predefined (and extensible) schemas.

- **XMP Data Model**. The data model is derived from RDF and is a subset of the RDF data model. It provides support for: metadata properties to attach metadata to a resource. Properties have property values, which can be structured (structured proper-

[56] `http://partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf`

ties) or simple types or arrays. Properties also may have properties (property qualifiers) which my provide additional information about the property value.

- **XMP Schemas**. Schemas consist of predefined sets of metadata property definitions. Schemas are essentially collections of statements about resources which are expressed using RDF. It is possible to define new external schemas, extend the existing ones or add some if necessary. There are some predefined schemas included in the specification like a Dublin Core Schema , a basic rights schema or a media management schema.

There is a growing number of commercial applications that already support XMP[57]. The International Press and Telecommunications Council (IPTC) has integrated XMP in its Image Metadata specifications and almost every Adobe application like Photoshop or In-Design supports XMP.

### 4.4.2 Social Vocabularies

This section gathers vocabularies that are centred around human beings and things they do, such as communicate, share, etc.; in the Fig. 4.7 a sample orchestration[58] of some of these vocabularies discussed in the following is depicted.

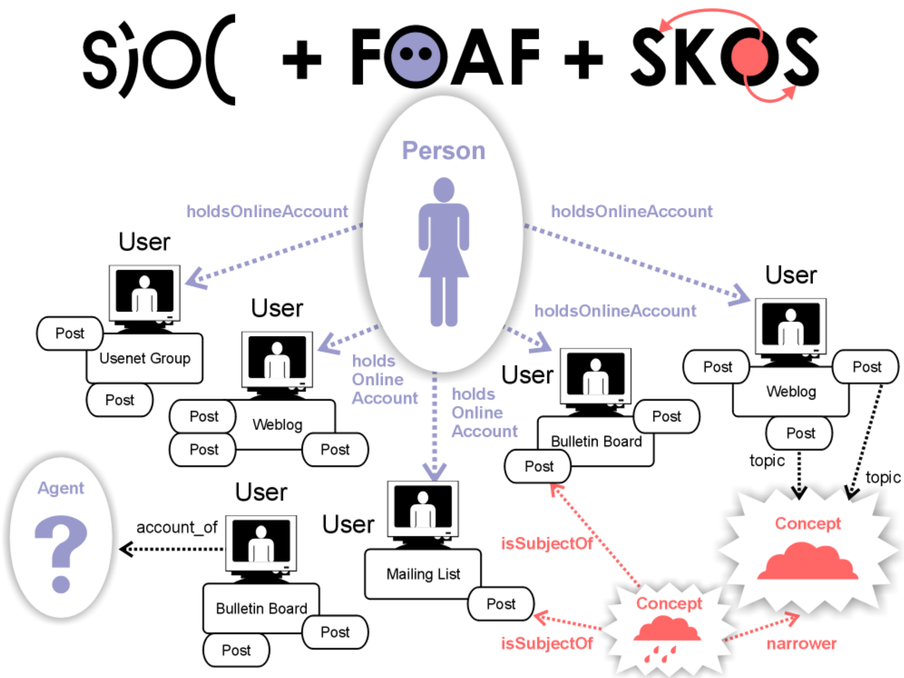



Figure 4.7: Social Vocabularies Orchestration.

[57] http://www.adobe.com/products/xmp/partners.html

[58] With kudos to Jon Breslin (DERI Galway), who put together this beautiful illustration; cf.http://www.johnbreslin.com/blog/2006/09/27/sioc-foaf-skos/

### Friend-Of-A-Friend–FOAF

The Friend of a Friend (FOAF) [46] project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do. FOAF is about humans places in the Web, and the Web's place in a humans world. FOAF is a simple technology that makes it easier to share and use information about people and their activities (e.g., photos, calendars, weblogs), to transfer information between Web sites, and to automatically extend, merge and re-use it on-line.

The FOAF project is based around the use of machine readable Web homepages for people, groups, companies and other kinds of thing. To achieve this we use the "FOAF vocabulary" to provide a collection of basic terms that can be used in these Web pages. At the heart of the FOAF project is a set of definitions designed to serve as a dictionary of terms that can be used to express claims about the world. The initial focus of FOAF has been on the description of people, since people are the things that link together most of the other kinds of things we describe in the Web: they make documents, attend meetings, are depicted in photos, and so on.

The FOAF Vocabulary definitions presented here are written using a computer language (RDF/OWL) that makes it easy for software to process some basic facts about the terms in the FOAF vocabulary, and consequently about the things described in FOAF documents. A FOAF document, unlike a traditional Web page, can be combined with other FOAF documents to create a unified database of information. An example FOAF document is depicted in Fig. 4.5 in N3 syntax.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix : <#> .

:mic a foaf:Person; # define profile of person
       foaf:based_near  [
           geo:lat "47.064";
           geo:long "15.453" ];
       foaf:depiction <http://sw-app.org/img/mic_2006_03.jpg>;
       foaf:homepage <http://sw-app.org/about.html>;
       foaf:interest <http://en.wikipedia.org/wiki/MPEG-7>,
                     <http://en.wikipedia.org/wiki/Semantic_Web>;
       foaf:mbox <mailto:michael.hausenblas@sw-app.org>;
       foaf:name "Michael G. Hausenblas"^^xsd:string;
       foaf:workplaceHomepage <http://www.joanneum.at/iis/> .
```

Listing 4.5: A sample FOAF document.

### Semantically-Interlinked Online Communities Project–SIOC

SIOC[59] is a vocabulary[42] to describe interconnected discussions in various so-called containers, such as blogs, forums and mailing lists etc. It partially builds upon and extends FOAF; recently, SIOC was submitted for W3C standardisation[60], hence a wide-spread and

[59] http://sioc-project.org/
[60] http://www.w3.org/Submission/2007/02/

uniform adoption is very likely. For a comprehensive list of SIOC applications and implementations, the reader is referred to [41].

#### Access and Rights Management

Access Control Lists (ACL) are designed to express access rules in a logical, unambiguous, machine-accessible format. A client with ACL data should be able to prove access privileges. For example the W3C site takes advantage of this flexibility with a dynamic, file-level access control system. In W3C ACL Schema [252] the ACL storage and query mechanisms used by W3C is described, as well as the availability and use of this data on the Semantic Web.

#### News Feeds

Atom RFC4287 [233] defines a feed format for representing and a protocol for editing Web resources such as Weblogs, online journals, Wikis, and similar content. The feed format enables syndication; that is, provision of a channel of information by representing multiple resources in a single document. The editing protocol enables agents to interact with resources by nominating a way of using existing Web standards in a pattern.

A re-formulation of the RFC4287 semantics on a formal basis (using OWL) has been proposed recently, called AtomOWL.[61]

### 4.4.3 Spatio-temporal Vocabularies

On the Semantic Web an array of vocabularies exists, which deal with dimensions, such as time or space. This section takes a closer look on some prominent and widely deployed, so called *spatio-temporal* vocabularies.

#### Geospatial Information

The W3C has proposed a basic RDF vocabulary[62] providing the Semantic Web community with a namespace for representing lat(itude), long(itude) and other information about spatially-located things, using WGS84 as a reference datum.

Useful information on geo-spatial research is provided by Harry Chen's blog, available at http://www.geospatialsemanticweb.com/.

#### Temporal Issues

OWL-Time[63] defines temporal concepts, for describing the temporal content of Web pages and the temporal properties of Web services. The ontology provides a vocabulary for expressing facts about topological relations among instants and intervals, together with information about durations, and about date-time information.

[61] http://bblfish.net/work/atom-owl/2006-06-06/AtomOwl.html
[62] http://www.w3.org/2003/01/geo/
[63] http://www.w3.org/TR/owl-time/

### 4.4.4 Other Vocabularies

In this last section of the Semantic Web vocabularies, the remainder of the vocabularies not fitting into any of the above categories are listed.

#### Description Of A Project–DOAP

Description Of A Project (DOAP)[64] is an attempt to make an RDF schema and XML vocabulary to describe open-source projects. The format was created and initially developed by Edd Dumbill[65] to convey semantic information associated with open-source software projects, such as:

- Global project characteristics (programming language, license, homepage, etc.);
- Maintainer characteristics (human represented in FOAF);
- Development detail characteristics (release, repository, etc.).

In the listing 4.6 below an excerpt of the Apache Tomcat DOAP document[66] is shown.

```
@prefix : <http://usefulinc.com/ns/doap#> .
@prefix asfext: <http://projects.apache.org/ns/asfext#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://tomcat.apache.org/> a :Project;
   :category <http://projects.apache.org/category/network-server>;
   :created "2006-01-27";
   :download-page <http://tomcat.apache.org>;
   :homepage <http://tomcat.apache.org>;
   :license <http://usefulinc.com/doap/licenses/asl20>;
   :mailing-list <http://tomcat.apache.org/lists.html>;
   :maintainer  [
       a foaf:Person;
       foaf:mbox <mailto:dev@tomcat.apache.org>;
       foaf:name "Tomcat PMC" ];
   :name "Apache Tomcat";
   :programming-language "Java";
   :release  [
       a :Version;
       :created "2006-04-14";
       :name "Latest Stable Release";
       :revision "5.5.17" ];
   :repository  [
       a :SVNRepository;
       :browse <http://svn.apache.org/repos/asf/tomcat/>;
       :location <http://svn.apache.org/repos/asf/tomcat> ] .
```

Listing 4.6: A sample DOAP document.

Latest developments resulted in a set of generators and validators, as well as viewers to enable more projects to be able to be included in the Semantic Web. DOAP repositories[67] are available as well.

[64] http://usefulinc.com/doap
[65] http://times.usefulinc.com/
[66] http://svn.apache.org/repos/asf/tomcat/site/trunk/docs/doap_Tomcat.rdf
[67] http://doapspace.org/

## 4.5 Linked Data

As we have summarised in [145] the basic idea of linked data was outlined by Sir Tim Berners-Lee[68]. In his note, a set of rules is being provided:

1. All items should be identified using *URI references* (URIrefs)[69];
2. All URIrefs should be *dereferenceable*; using HTTP URIs allows looking up the items identified through URIrefs;
3. When looking up an URIref (i.e. an RDF-property is interpreted as a hyperlink) it leads to more data;
4. Links to other URIrefs should be included in order to enable the discovery of more data [36].

The Linking Open Data (LOD) project is a collaborative effort aiming at bootstrapping the Semantic Web by publishing datasets in RDF on the Web and creating large numbers of links between these datasets [37]. As of time of writing roughly two billion triples and three mil-

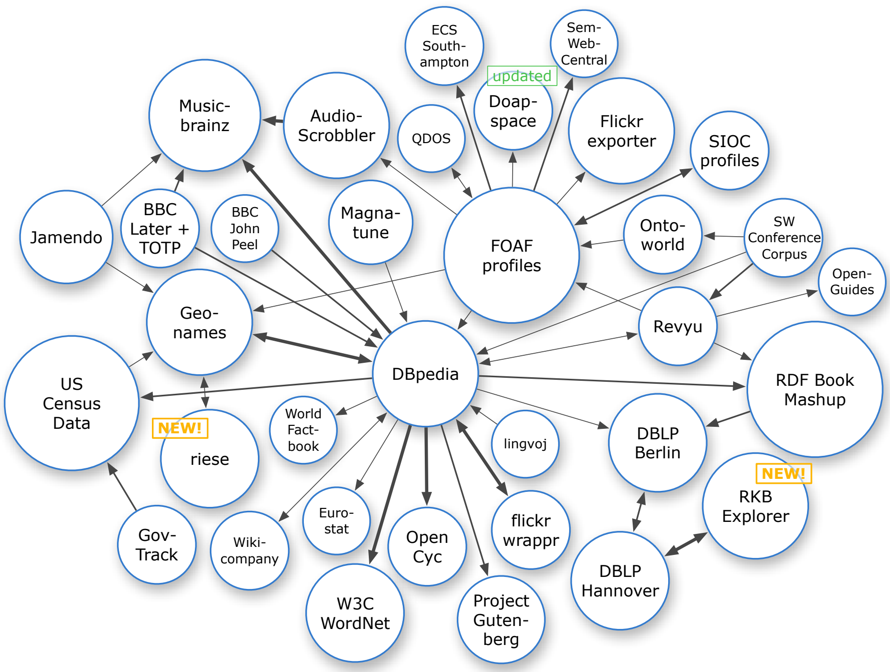



Figure 4.8: The Linking Open Data dataset at time of writing.

lion interlinks have been reported (cf. Fig. 4.8, by courtesy of Richard Cyganiak[70]), ranging from rather centralised ones to those that are very distributed. The LOD dataset can roughly be partitioned into two distinct types of datasets, namely (i) single-point-of-access datasets, such as DBpedia or Geonames, and (ii) distributed datasets (e.g. the FOAF-o-sphere). This distinction is significant regarding the access of the data in terms of performance and scalability.

[68] http://www.w3.org/DesignIssues/LinkedData.html
[69] http://www.w3.org/TR/rdf-concepts/#section-Graph-URIref
[70] http://richard.cyganiak.de/2007/10/lod/

## 4.6 Web 3.0

This section examines current trends and development in the realm of the Semantic Web and puts them into context. While a lot of buzzwords are floating around, some fundamental changes are taking place during the completion of this work. One of the most notable is the so called *Web 3.0*—the ultimate fusion of Web 2.0 and the Semantic Web.

### 4.6.1 Web 2.0: Ajax & Mashups

Asynchron JavaScript and XML[71] is a set of technologies and APIs regularly utilised in Web 2.0 applications. It allows for new kind of Web-client implementations, freeing the Server from a certain load. Using Ajax allows for more responsive and smarter Web-Clients, and enables Desktop-like GUIs.

With the raise of Ajax, the second layer of the Semantic Web stack—the data structure and exchange layer—has been de facto extended. *JavaScript Object Notation* (JSON)[72]—a lightweight data-interchange format based on a subset of the JavaScript language—started to compete XML; not only in the context of Web 2.0, but increasingly in the Web Service domain. There are a number of applications supporting JSON as input or output format alike.

```
"items" : [
{
        "id" :          "{AMSM}",
        "booktitle" : "1$^{st}$ International Conference on New Media
            Technology (iMedia07) ",
        "pub-type" :  "inproceedings",
        "author" :    "Hausenblas,-M.",
        "type" :      "Publication",
        "year" :      "2007",
        "label" :     "{Applying Media Semantics Mapping in a Non-linear,
            Interactive Movie Production Environment}",
        "address" :   "Graz, Austria",
        "key" :       "Hausenblas:IMEDIA07"
}
],
"types" :       {
        "Publication" : {
        "label" :       "Publication",
        "uri" :
            "http://simile.mit.edu/2006/11/bibtex#Publication",
        "pluralLabel" : "Publications"
        }
},
"properties" : {
        "journal" :         {
                "uri" : "http://simile.mit.edu/2006/11/bibtex#journal"
        }
}
```

Listing 4.7: An excerpt of a sample JSON document.

[71] http://en.wikipedia.org/wiki/Ajax_(programming)
[72] www.json.org

In the listing 4.7 below an excerpt of the JSON file listing the author's publications[73] is shown. Using Exhibit[74]—a light-weight Ajax publication framework—a view on the bibliographic data allowing for faceted browsing can be created (cf. Fig. 4.9).


Figure 4.9: An sample Exhibit document using JSON.

JSON is not only important w.r.t. XML. Recently a resource-centric serialisation of RDF in JSON[75] was proposed. Many Semantic Web APIs (such as SPARQL) offer JSON as an alternative format to XML. The rational for this is to lower the barrier for developers unfamiliar with RDF or RDF/XML.

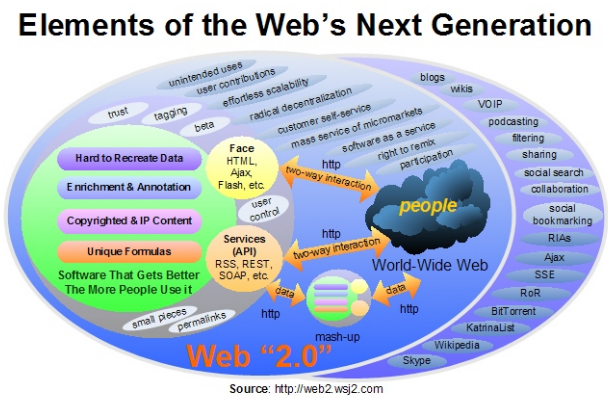



Figure 4.10: Web 2.0: Ajax, Mashups and more ...

**Mashups**. In the Fig. 4.10, the current state of the Web 2.0 is depicted[76]. One important part are the so called mashups, Web applications that combine data from more than one source into a single integrated site. Note that an overview on over 500 APIs is available,[77] listing possible combinations into mashups.

[73] http://sw-app.org/pub/exhibit_mic_pubs.html
[74] http://simile.mit.edu/exhibit/
[75] http://n2.talis.com/wiki/RDF_JSON_Specification
[76] Source: http://web2.wsj2.com/
[77] http://www.programmableweb.com/matrix

### 4.6.2 Metadata in HTML

In the following we shed some light on issues around metadata and HTML, sometimes misleadingly referred to as semantic HTML. First, we have a closer look on a grass-rooted approach coined as *microformats*. We then examine an approach how to embed metadata based on the RDF-model, which is conforming to the Semantic Web. For a comparison on the approaches described below, the reader is invited to visit a service (supplied by Benjamin Novack) allowing to compare them side-by-side[78].

#### Microformats

Microformats[79] are a set of simple, open data formats allowing to embed structured data into Web pages, cf. Fig. 4.11. The syntax of microformats varies from one application domain to another; this makes it rather hard to treat them uniformly and causes problems when mixing them[80].

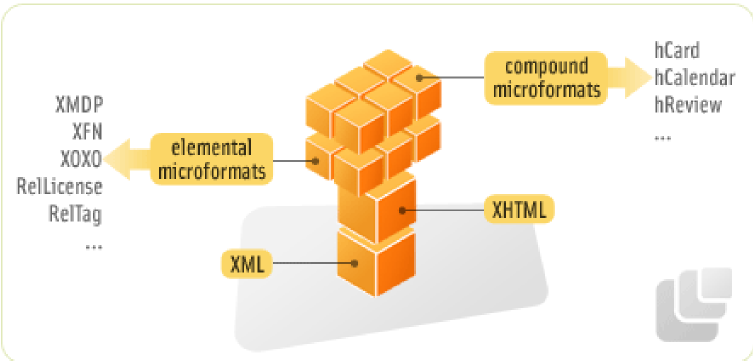



Figure 4.11: An overview on microformats.

Although microformats bear certain limitations, they are quite useful for a range of use cases; common microformats[81] are listed in the following. We roughly distinguish (i) elemental microformats, which are minimal solutions to a single problem, built from standard XHTML elements; acting as a building block for larger microformats, and (ii) compound microformats (see below).

Defined *elemental microformats* so far are:

- *rel-nofollow*. By adding `rel="nofollow"` to a hyperlink, a page indicates that the destination of that hyperlink should not be afforded any additional weight by user agents such as search engines;
- *rel-tag*. Used to indicate that the destination of that hyperlink is an author-designated "tag"for the current page;
- *rel-license*. Usable for indicating content licenses;

[78] http://bnode.org/blog/2007/02/12/comparison-of-microformats-erdf-and-rdfa
[79] http://microformats.org/
[80] http://internet-apps.blogspot.com/2007/02/jon-udell-on-microformats-and-rdftime.html
[81] http://microformats.org/wiki/Main_Page#Specifications

- *VoteLinks*. Suggests three new values for the `rev` attribute of the `<a>`(hyperlink) tag in HTML, namely (i) ”vote-for” (ii) ”vote-abstain”, and (iii) ”vote-against”, which are mutually exclusive, and represent agreement, abstention or indifference, and disagreement respectively;
- *XHTML Friends Network* (XFN). Is a simple way to represent human relationships using hyperlinks;
- *Extensible Open XHTML Outlines* (XOXO). Serves as the basis for XHTML friendly outlines for processing by XML engines and for easy interactive rendering by browsers.

*Compound microformats* are built from elemental microformats; they may be a mapping of an existing standardised schema describing a compound data type:

- *hCalendar*. A calendaring and events format, based on the iCalendar standard (RFC 2445 [68]);
- *hCard*. A format for representing people, companies, organizations, and places; based on vCard (RFC 2426 [67]) properties and values;
- *hReview*. Suitable for embedding reviews of products, services, businesses, events, etc.
- Additionally, an array of draft for microformats exists, including provisions for example to mark up geographic coordinates, for publishing resumes and CVs, etc.

The success of microformats—some 450 million instances of microformatted content on the Web has been reported recently[82]—has influenced other efforts to embed metadata into HTML, as discussed below.

### Folksonomies

Folksonomy[83] is the practice and method of collaboratively creating and managing tags to annotate and categorize Content. The tagging area began when *del.icio.us*[84] introduced tags for roughly describing the content of a bookmark some couples of years ago (cf. Fig. 4.12 on page 90).

In a recent keynote[119], Tom Gruber stated:

> Ontologies are enabling technology for the Semantic Web. They are a means for people to state what they mean by formal terms used in data that they might generate or consume. Folksonomies are an emergent phenomenon of the social web. They are created as people associate terms with content that they generate or consume. Recently the two ideas have been put into opposition, as if they were right and left poles of a political spectrum. This piece is an attempt to shed some cool light on the subject, and to pre-view some new work that applies the two ideas together to enable an Internet ecology for folksonomies.

The approach taken by Gruber is a promising one that could help to bridge another gap: The one between emerging Web 2.0 and Folksonomy community on the one side, and the academic-driven Semantic Web development on the other.

[82] http://rbach.priv.at/Microformats-IRC/2007-06-20#T235145
[83] http://www.atomiq.org/archives/2004/08/folksonomy_social_classification.html
[84] http://del.icio.us/

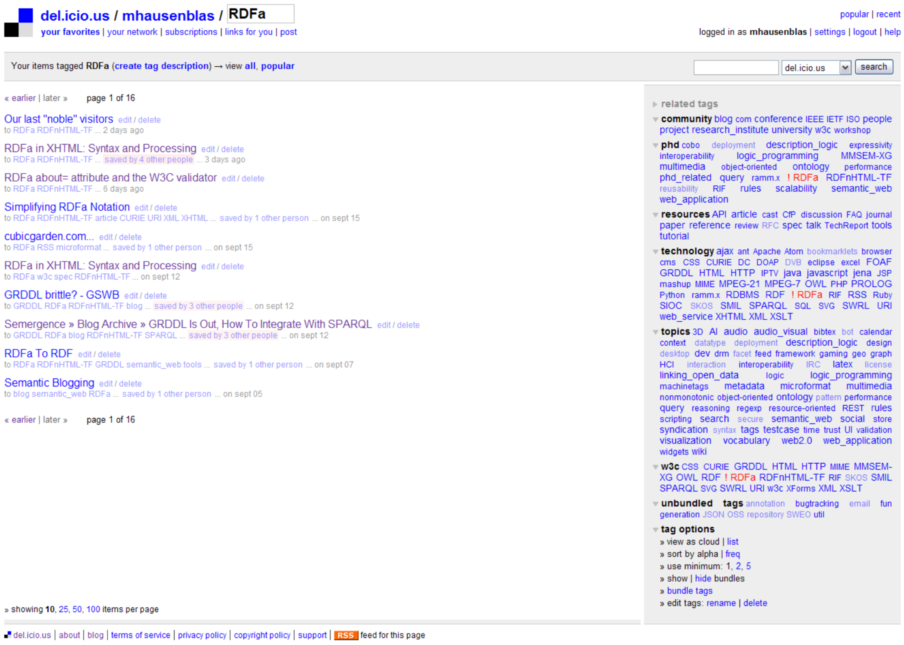



Figure 4.12: An exemplary tag-based environment (del.icio.us).

The reader is invited to note that the author of this thesis has as well participated in an pre-standardisation activity regarding the utilisation of Semantic Web technologies in tagging systems[85].

### Gleaning Resource Descriptions from Dialects of Languages–GRDDL

While adding metadata explicit to an HTML page is an option especially for new or evolving systems; current Web sites (including microformats) can be easily turned into Semantic Web sites when using Gleaning Resource Descriptions from Dialects of Languages (GRDDL) [61]. The GRDDL specification proposes mechanisms for declaring that an XML document includes data that is compatible with RDF. Further GRDDL defines the linking to algorithms—which are typically represented in XSLT—for extracting this data from the document. In contrast to harvester or wrapper generators such as [23] that are defined and applied at the consumer's side, GRDDL enables this transformation on the author's side.

[85] http://www.w3.org/2005/Incubator/mmsem/wiki/Tagging_Use_Case_Review

### RDF in HTML

As of [2], current Web pages contain inherent structured data: calendar events, contact information, photo captions, song titles, copyright licensing information, etc. When authors and publishers can express this data precisely, and when tools can read it robustly, a new world of user functionality becomes available, letting users transfer structured data between applications and web sites.

The W3C has—motivated by the success of the microformats—initiated a standardisation effort to embed RDF in (X)HTML. RDFa (RDF in attributes) [5; 4] is a syntax that allows for embedding an RDF graph into an (X)HTML document via attributes. RDFa lets XHTML authors express this structured data using extra XHTML attributes.

The following XHTML attributes are relevant in the scope of RDFa [4]:

- `@rel`, list of CURIEs, used for expressing relationships between two resources;
- `@rev`, list of CURIEs, used for expressing reverse relationships between two resources;
- `@href`, a URI for expressing the partner resource of a relationship;
- `@src`, a URI for expressing the partner resource of a relationship when the resource is embedded.

However, in contrast to alternative proposals to embed RDF into HTML, such as eRDF[86], the RDFa specification introduces new, RDFa-specific attributes allowing to represent an arbitrary RDF graph:

- `@about`, a URI or a CURIE, used for stating what the data is about
- `@property`, a list of CURIEs, used for expressing relationships between the subject and some literal text;
- `@resource`, a URI or a CURIE for expressing the partner resource of a relationship (not intended to be 'clickable')
- `@datatype`, a CURIE representing a datatype, to express the datatype of a literal
- `@content`, a string, for supplying alternative, machine-readable content for a literal.
- `@instanceof` a list of CURIEs that indicate the RDF type(s) to associate with the subject.

**CURIE**. The RDFa Syntax document [4] defines not only how RDFa must be processed, but also introduces the notion of *Compact URIs* (CURIE). A CURIE is comprised of two components, a prefix which maps to a URI, and a reference. The prefix is separated from the reference by a colon. The main reason for supporting CURIEs is that QNames [45, Sec. 4] have certain restrictions[87], resulting in a lack of support for certain use cases. For a detailed discussion on the CURIE vs. QName discussion, the reader is referred to a blog post from Mark Birbeck[88], one of the editors of the RDFa Syntax.

[86] http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml
[87] http://lists.w3.org/Archives/Public/public-rdf-in-xhtml-tf/2007Aug/0086.html
[88] http://internet-apps.blogspot.com/2005/10/curies-compact-uri-syntax-semantic.html

In the Fig. 4.13, an exemplary serialisation of a FOAF document (cf. section 4.4.2) using XHTML+RDFa is depicted. The numbers in the figure represent the components and their dependencies as follows:

1. Is the view on the Web page as a human user consumes it; it is the plain HTML rendering;

2. Depicts the XHTML+RDFa source code. There, RDFa is used to embed the RDF-based FOAF information (the RDF graph) into the XHTML document;

3. When a RDFa-aware agent processes the page, it is able to extract the encoded RDF graph. It is then possible to, e.g., execute a SPARQL query on it.

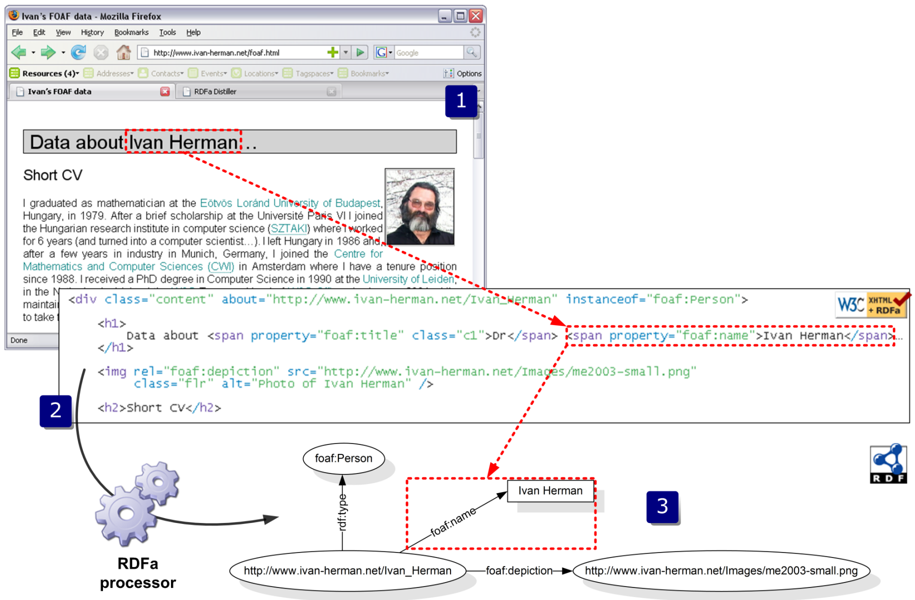



Figure 4.13: An exemplary XHTML+RDFa version of a FOAF document.

Note that at the time of writing the standardisation of RDFa is being finalised[89]. As a member of the RDFa Task Force, the author of this thesis is responsible for the RDFa Test Suite[90]. Further, a list of applications utilising RDFa[91] is maintained by the author.

### 4.6.3 Web 2.0 + Semantic Web = Web 3.0?

Although Web 2.0 seems to some as too much of a buzz word, we are about to approach Web 3.0, already [194]. Motivated by the ongoing discussions[92] regarding Web 2.0 and the Semantic Web, in the following a more generic point of view on the issue is taken. We examine this very issues in terms of a communications model capturing basic communication paths between human and computers. Note that this model is roughly based on the Maurer-Tochtermann model [207].

[89] http://rdfa.info/2007/09/22/one-step-closer-to-bridging-the-clickable-and-semantic-webs
[90] http://www.w3.org/2006/07/SWD/RDFa/testsuite/
[91] http://rdfa.info/rdfa-in-the-wild/
[92] http://www.oreillynet.com/xml/blog/2005/09/sparql_web_20_meet_the_semanti.html

In the Fig. 4.14 the basic communication paths between humans and computers are depicted.

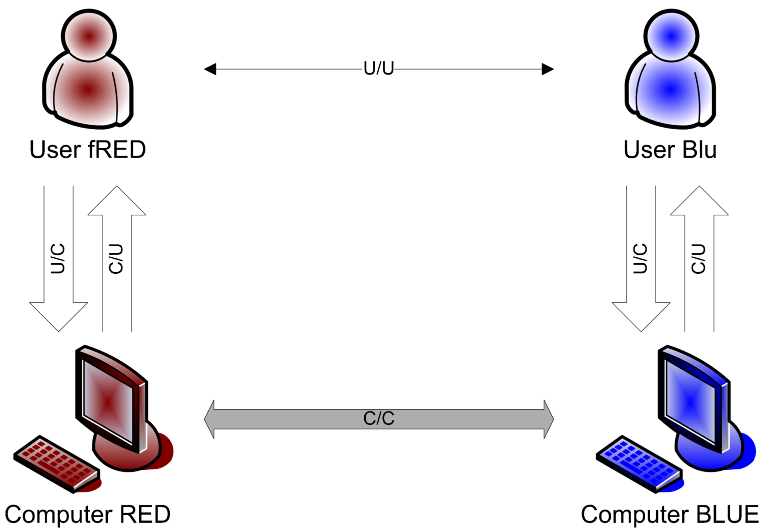



Figure 4.14: Human-Computer Communication Model in the Web 3.0 context.

In this setup, we have two users (fRED and Blu), and their computers RED and BLUE, respectively. So the following potential communication patterns can be identified:

1. **User to User** (U/U). That is, either face to face, or via phone, IRC, etc.;
2. **User to Computer** (U/C). Thus, a user tells the computer to perform an operation ("load web page", "connect to 123 ...", etc.);
3. **Computer to User** (C/U). A computer reports about the output of an operation (display web page, "can't find host", etc.);
4. **Computer to Computer** (C/C). The (fully) automatic exchange of information between computers, based on defined formats and protocols.

Based on the above stated, the following observations can be made:

*Regarding U/U*. The focus is obviously on the human communication, therefore, even if computers are involved (Blog, ICQ, etc.), they are merely tools to support human users to communicate. Nevertheless, computers are involved in the whole communication (except face-to-face), therefore all follwing patterns are included.

*Regarding U/C and C/U*. The field of Human Computer Interaction (HCI)[93] is broad. A lot of ways how humans may interact with computers have been proposed and researched. In our context, not the way of interaction is of interest, but the following basic facts:

- Human users have a defined set of commands to tell a computer to perform an operation, and to specify the expected result, respectively;
- Human users have to learn these commands and abstract them to fulfil a certain task;

[93] http://www.hcibib.org/

- Computers "understand" this set of commands in terms of that they execute a corresponding piece of code using the current users input and come up with the results.

Note that the author's view regarding Semantic Web and HCI issues has been published,[94] already.

*Regarding C/C.* The focus of this communication pattern is on automatic exchange of information, i.e., without involving users directly. The main statements are that (i) Computers need unambiguous, formal defined formats and protocols to perform a communication, and that (ii) normally, a C/C communication is triggered by a user command (directly or indirectly), therefore (user defined) parameters are available that control the communication.

Semantic Web is all about C/C communication, hence to provide infrastructure that can be used by machines to perform tasks on behalf of human users. Web 2.0 is all about social networking, focusing on human users that share information and communicate using interconnected computers. Finally Web 2.0 may use SW-infrastructure to support human users in using Web 2.0 functionality. Concluding, the merge of the two worlds may be called *Web 3.0*.

Recently a Web 3.0 architecture was proposed[95], depicted in Fig. 4.15. There, the Web 3.0 architecture is described as follows.

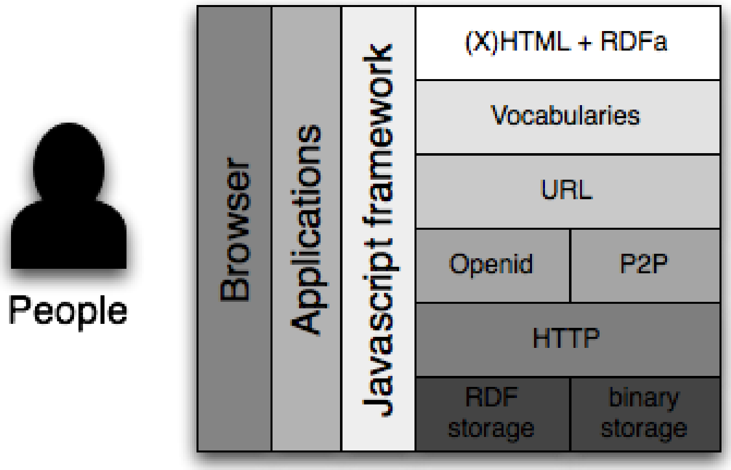



Figure 4.15: A Proposed Web 3.0 Architecture.

> Web3.0 is about taking the web into the browser. [...] Web3.0 applications are distributed and run seamlessly on client and servers. The JavaScript framework implements the model–view–controller design pattern, provides necessary libraries to read, write and handle graph data in both local RDF stores or as (X)HTML+RDFa web pages. These pages actually carry over HTTP the state of the applications. OpenID provides the framework with a distributed identity mechanism, and with a URL for everyone. [...] The vocabularies are agreements made by humans and used to describe human or machine semantics, see social tagging as a human semantics and RDF vocabularies as machine semantics.

[94] In the realm of the Semantic Web User Interaction Workshop 2007 (SWUI 07), see also http://swui.semanticweb.org/swui2007/

[95] http://blog.cedricmesnage.org/articles/2007/07/24/the-web3-0-architecture

## 4.7 Conclusion

Although the Semantic Web itself is still in its early stages there is already an array of applications utilising Semantic Web technologies. The W3C hosts a list of Use Cases[96] as well as commercial products using RDF as a base[97]. For Semantic Web developers there exists a range of sources they may use to gather information and look up appropriate tools. Examples are listed in the following:

- The W3C Semantic Web FAQ: http://www.w3.org/2001/sw/SW-FAQ
- A list of Semantic Web tools: http://esw.w3.org/topic/SemanticWebTools
- A comprehensive Semantic Web applications list maintained by Michael Bergman: http://www.mkbergman.com/?page_id=325

[96] http://www.w3.org/2001/sw/sweo/public/UseCases/
[97] http://esw.w3.org/topic/CommercialProducts

# Part II

# Methods and Requirements

Chapter 5

# Creating Smart Content Descriptions

*"Modus operandi."*

(Latin phrase)

To gain a true understanding of the problems in the context of multimedia applications on the Semantic Web, this chapter elaborates on the creation and on the usage of content descriptions. This is done both on the low-level feature level as w.r.t. domain semantics.

In [123] we argue that standard methodologies to describe and search specific content, e.g. an image, video or piece of music are mostly utilizing "piggy-back" text technologies based working on metadata. Text and metadata may be manually entered, gained from optical character recognition (OCR) or from automatic speech recognition (ASR). Content Based Indexing & Retrieval (CBIR) methods are extracting meaning directly from multimedia objects.

While this is relatively easy for low level features like colour, texture, pitch or volume, it is extremely difficult to extract objects or genres, to name just a few real world concepts. At this point we have to resort to knowledge. Semantic Web technologies are offering a way to formalise the knowledge and help us in describing (and later on in finding) the content in a much more user-oriented way. Even more, single content objects knowing about their meaning will on the long run be able to combine and reconfigure themselves on the fly into meaningful sequences, following established drama rules.

## 5.1 Information Flow and Media Semantic Web Stack

The information flow in a multimedia application on the Semantic Web is depicted in Fig. 5.1. There, on the left side the **content creators** produce multimedia data and metadata using some application. This application usually has direct access to the data and uses a stack of protocols and languages to handle the metadata. On the other side, the **content consumers** make use of an application to view and interact with the content by utilising metadata to find, browse, or otherwise manage it.

To span the space of possibilities, we initially take a look at two edge cases. The Semantic Web approach with RDF and OWL on the one hand (cf. Section 4.3), and the MPEG-7-based approach (cf. Section 3.3.4) on the other. The justification for the choice of MPEG-7 is

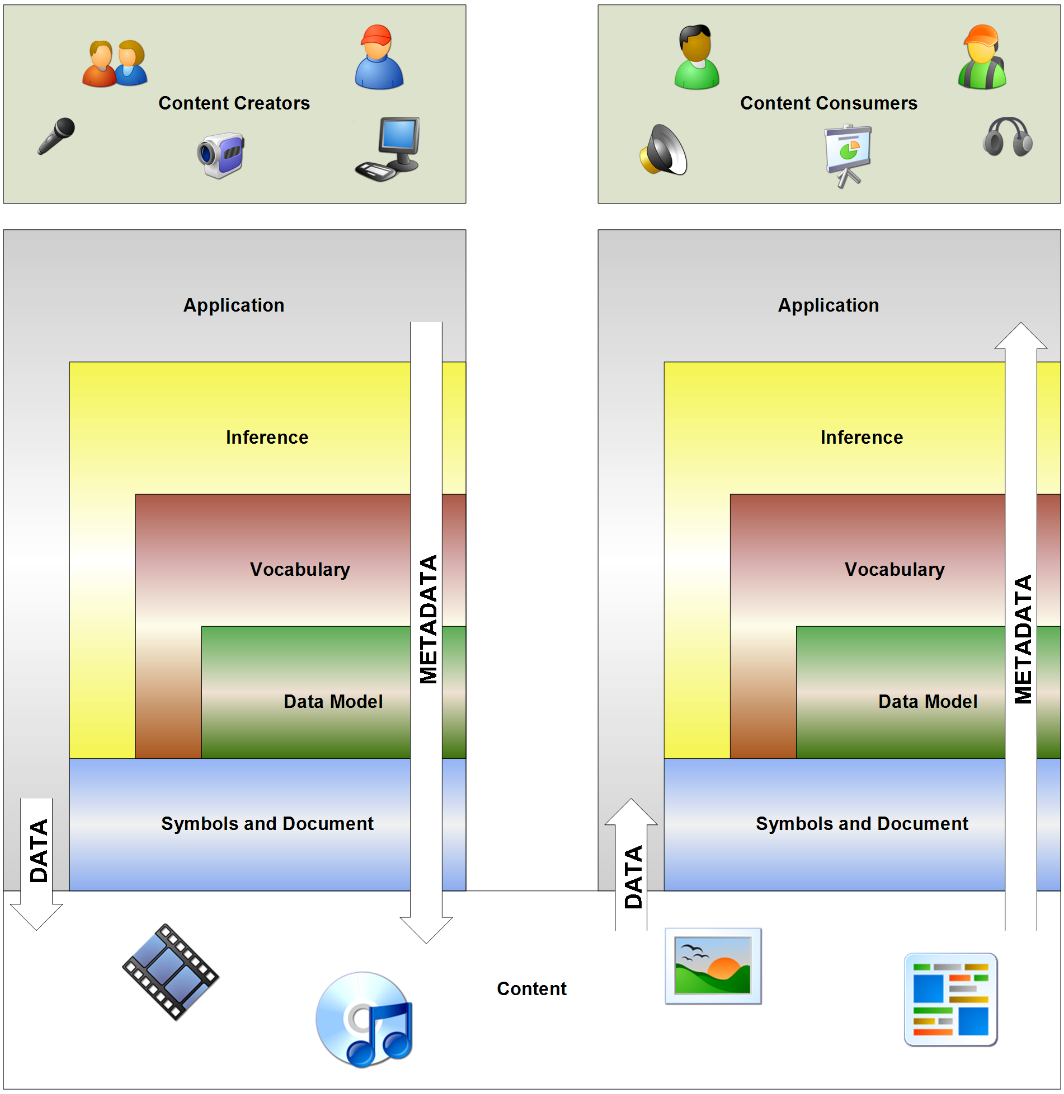



Figure 5.1: Flow of Information in Multimedia Applications on the Semantic Web.

that this is the most complete multimedia metadata standard regarding modalities and the granularity of the content description. The justification for choosing OWL is obvious: it is the ontology language at the very heart of the Semantic Web. In Fig. 5.2 we put the Semantic

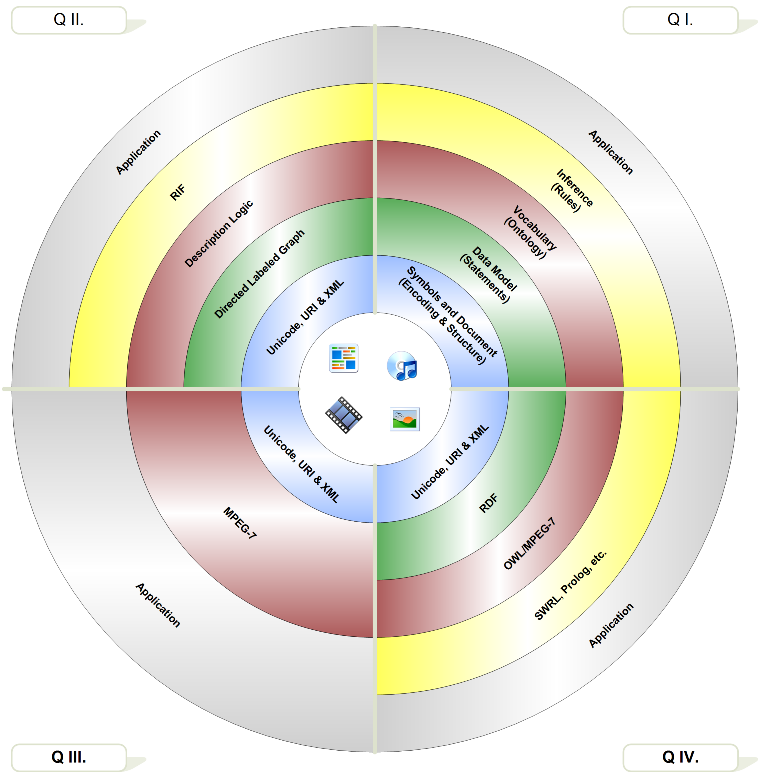



Figure 5.2: The Semantic Web Stack in the Realm of Multimedia Applications.

Web Stack in the context of multimedia applications. In the centre of the diagram, the actual media assets are depicted, such as an audio or video file. The farther away from the content in the centre, the higher is the abstraction level used in the figure. The four quadrants in Fig. 5.2 denote some possible layering of the metadata ranging from pure symbol representation up to inferential tasks, as discussed in the following.

**Q I.** The *generic layers*:

1. The lowest layer is responsible for representing symbols and documents including unambiguous addressing of parts of the description.
2. The second layer provides a data model that allows for making statements.
3. On top of the second layer a vocabulary layer is found that captures domain semantics.
4. The inference layer supports the task to make implicit information explicit.

**Q II.** The *Semantic Web* (W3C stack).
Unicode, URI, and XML represent the innermost layer. The data model is provided through the RDF model. Description Logics (OWL) form the basis for controlled vocabularies and finally RIF (together with OWL) is able to fill the inference layer.

**Q III.** An *MPEG-7 Application*.
The lowest layer is identical with the Semantic Web stack, hence here exists a certain interoperability cf. [310] and [229]. The MPEG-7 provides both a model and a vocabulary (Descriptors, Description Schemes, etc.). Due to the missing formal base, the inference is regularly done (implicit) on the application layer.

**Q IV.** A typical *multimedia application on the Semantic Web*.
Here, the two bottom layers are used according to the pure Semantic Web approach, then a layer follows that somehow combines MPEG-7 and OWL. For example certain MPEG-7 tools (cf. 3.3.4) are modeled in OWL; additionally the application domain is typically formalised using OWL.

## 5.2 Extraction vs. Annotation

In this work *extraction* is understood as the process of automated generation of metadata, whereas *annotation* is the human-led, i.e., manual process of creating metadata. While extraction is fast, cheap, and can be used for a high volume of content, the results are of low quality and almost always need some kind of human post-editing or supervision. Annotation on the other hand yields high-quality output, but is resource consuming.

### 5.2.1 Extraction

A key goal of content analysis is to describe the media being analysed in an efficient and exchangeable way. As low-level features are typically contained by many or all pixels of an image (e.g. every pixel has a colour), it is necessary to derive more compact representations of these features. Such a representation is called descriptor, and represents a certain feature of an image, a set of images (e.g. an image sequence), or a region of an image. The definition of a descriptor shall consist of three parts:

- Representation: which data is contained in the descriptor and what is meant by it
- Extraction: how to create descriptor data from a given visual media
- Comparison: how to determine the distance or similarity between two descriptors of the same type

Extraction is the process of creating a descriptor from a given visual media item. The extraction algorithm typically has to reduce the complexity of the data and to extract representative components. Thus, extraction algorithms often include clustering, component analysis or transform domain approaches.

### 5.2.2 Annotation

In [126; 40], a framework—CREAM—is discusses that allows for creation of metadata.In the annotation mode the system allows to create metadata for existing web pages, whereas the authoring mode lets authors create metadata while putting together the content of a page. As a particularity of the framework it allows to create relational metadata, i.e., metadata instantiating interrelated definitions of classes in a domain ontology rather than a comparatively rigid template-like schema as Dublin Core. They discuss requirements one has to meet when developing such an ontology-based framework, e.g. the integration of a metadata crawler, inference services, document management and a meta-ontology, and describe its implementation, viz. OntoMat, a component-based, ontology-driven Web page authoring and annotation tool—cf. also [40] for a more recent discussion on OntoMat.

Fig. 5.3 (on page 104) depicts an example of an multimedia retrieval application. In this example a combination of extraction and annotation is utilised to produce high volume and high quality multimedia metadata.

## 5.3 How To Deal with the Semantic Gap

In the following we discuss known approaches allowing to bridge the Semantic Gap.

### 5.3.1 Low-level Feature Based Approach

Throughout the 1980s and 1990s low-level based approaches have dominated the work on content-based image retrieval (CBIR). Using the query-by-example (QBE) paradigm, similarity between multimedia documents is defined in term of low-level features that can be directly derived from the multimedia data, such as colour, texture and shape in the visual domain, or pitch and frequency spectrum in the audio domain. A huge number of feature descriptors have been developed throughout this period, along with efficient matching and indexing approaches, some of these feature descriptors have been standardized in the MPEG-7 standard [222]. The advantage of this approach is that the problem of interpreting the multimedia data is avoided, with the drawback that queries can only be formulated by presenting a signal representation of the query example.

### 5.3.2 Model-based Approach

This class of approaches emerged from the application of computer vision and image understanding research to multimedia indexing and retrieval. Generally speaking, the concepts (objects, events, etc.) to be detected in the content are modelled and connected to their low-level feature representations by training classifiers using supervised learning approaches. If the domain of the multimedia content is known, these approaches yield satisfying results in practical applications (e.g. content-based description of sports video). However, in general applications the quality of the results depends crucially of the grounding of the concept in the low-level features. As can be seen from concept detection benchmarks (e.g. TRECVID[1]), the performance for abstract concepts is rather poor.

### 5.3.3 Semantic Web Approach

Grounded on the Semantic Web vision, the approach is to use RDF and Description Logic-based languages to model audio-visual features and domain semantics. To overcome the problem that comes along with multimedia metadata standards, viz. the missing formal basis [310; 229], this seems to be a good idea. This purist Semantic Web approach solves interoperability issues and allows for sound retrieval operations. Though, it solves some of the problems, it introduces new ones: the lack of support for basic multimedia requirements as time-based descriptions, weak data typing, and scalability issues, just to mention a few.

[1] http://www-nlpir.nist.gov/projects/trecvid

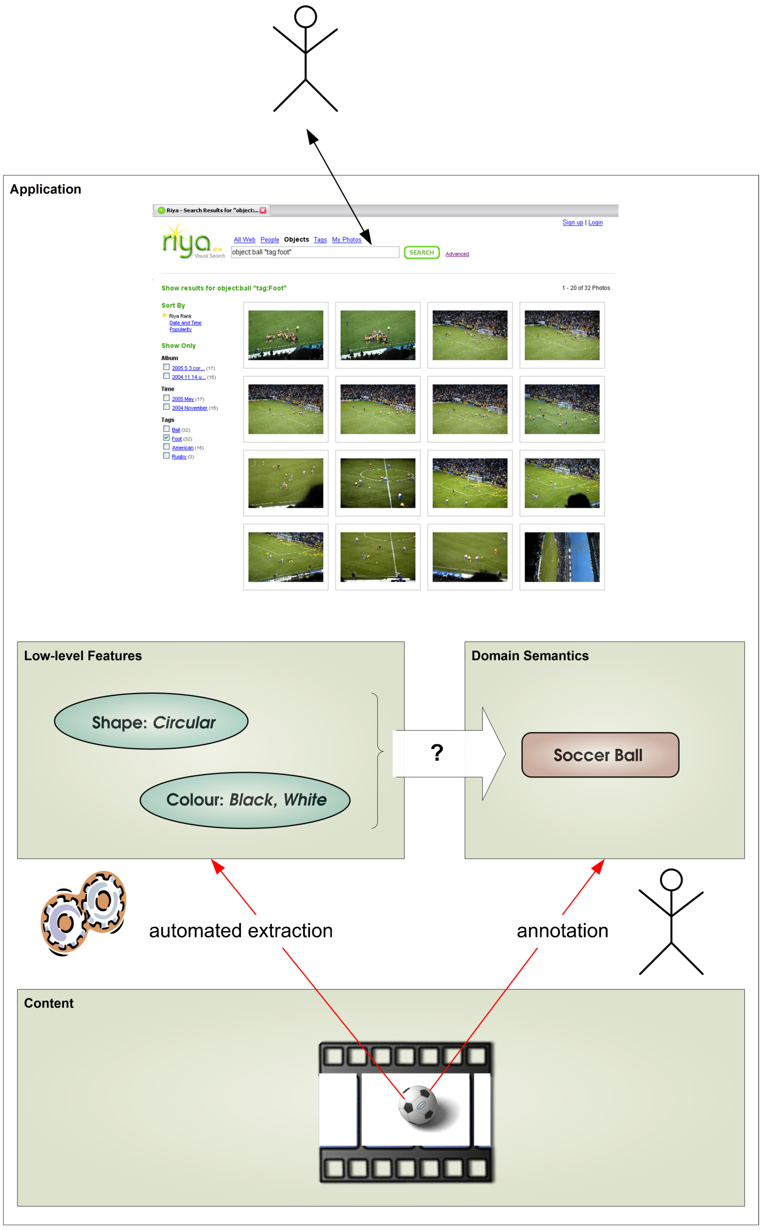



Figure 5.3: Extraction & Annotation Yielding High Quality Metadata.

### 5.3.4 Hybrid Approach

Instead of either using a bottom-up (multimedia data based) or top-down (knowledge based) approach, hybrid approaches aim at integrating both worlds, as both can mutually benefit from one another. How to do this integration practically is still a very active research topic and different approaches have been proposed recently, differing in the metadata representations and technologies being used. As this approach brings together before separated communities, it is a very promising, yet not mature one.

Hare et.al. [129] report on their work in the "Bridging the Semantic Gap in Image Retrieval" project. They have been investigating how the use of test-bed ontologies can meet the needs of real image searchers in limited domains. In order to investigate the potential of ontology-driven search, a thesaurus of the image metadata was created, and modelled this using SKOS (cf. section 4.4.1).

## 5.4 Multimedia Ontology Engineering

When creating an ontology-based application, one of the major question is how to create the ontology it is based upon. Although the process is similar to developing a relational database schema, there exist differences. In the process itself, two basic methods may be applied, namely

- manual creation by utilising for example dedicated ontology editors, such as Protégé [187];
- (semi)automated creation by utilising statistical or machine-learning techniques.

The process of creation is only the first step in the chain. After the ontology has been integrated into the application, it is usually tested and evaluated [106]. Very likely updates occur in an ontology's life-time leading to versioning issues [59].

In the following we take a closer look on ontology engineering methodologies and tools, and finally review existing multimedia ontologies. For further reading on the ontology engineering issue the reader is referred to Simperl and Tempich [269].

### 5.4.1 Methodologies

An excellent starting point for ontology engineering is the *Handbook on Ontologies* [279]. It provides for systems, tools, and applications. Numerous proposals exist how to create and populate ontologies.

In [179], the *Conceptual Data Modelling* for ontology engineering is suggested. They propose an ontology engineering-framework that enables reusing conceptual modelling approaches in modelling and representing ontologies. In the approach presented, they prevent application-specific knowledge to enter or to be mixed with domain knowledge.

The *CYC methodology* [197] CYC methodology consists of the following steps: first, you have to extract, by hand, common sense knowledge that is implicit in different sources. Next, once you have enough knowledge in your ontology, new common sense knowledge can be acquired either using natural language or machine learning tools.

The *CommonKADS/KACTUS* methodology [29] suggests to built the ontology on the basis of an application knowledge base, by means of a process of abstraction, viz. following a bottom-up strategy. The more applications are built, the more general the ontology becomes.

In other words, they propose to start building a KB for a specific application. Later, when a new knowledge base in a similar domain is needed, they propose to generalize the first KB into an ontology and adapt it for both applications. Applying this method recursively, the ontology would represent the consensual knowledge needed in all the applications.

*DILIGENT* [248] is a ontology engineering methodology for distributed, loosely controlled and evolving engineering of ontologies. Engineering a shared ontology is a social process that likely will take place in a distributed way. DILIGENT does not assume that the ontology developed with the methodology covers all aspects of the domain from the beginning (completeness). It rather assumes that the ontology will evolve over time and adapt to the user needs.

A methodology for analysing ontologies based on formal, domain-independent properties of classes (the meta-properties) is *OntoClean* [279, p 151–172]. It was the first attempt to formalise notions of ontological analysis for information systems. The idea was to justify the kinds of decision that experienced ontology builders make, and explain the common mistakes of the inexperienced. The notions OntoClean focuses on are drawn from philosophical ontology. The basis of OntoClean are the domain-independent properties of classes, the OntoClean meta-properties: identity, unity, rigidity, and dependence. Recent work has added two more meta-properties: permanence and actuality.

The *Ontolingua* ontology development environment [92] provides a suite of ontology authoring tools and a library of modular, reusable ontologies. The environment is available as a Web service and has a substantial user community. The tools in Ontolingua are oriented toward the authoring of ontologies by assembling and extending ontologies obtained from a library. In their work, they describe Ontolingua's formalism for combining the axioms, definitions, and words (non-logical symbols) of multiple ontologies. Further, they describe Ontolingua's facilities that enable renaming of words non-logical symbols from multiple component ontologies and that provide unambiguous mapping between words and text strings disambiguate symbol references during input and output.

*On-To-Knowledge Methodology* [280] applies ontologies to electronically available information to improve the quality of knowledge management in large and distributed organizations. The methodology provides guidelines for introducing knowledge management concepts and tools into enterprises, helping knowledge providers and seekers to present knowledge efficiently and effectively. The methodology includes the identification of goals that should be achieved by knowledge management tools and is based on an analysis of usage scenarios and different roles knowledge workers and other stakeholders play in organisations.

The methodology proposed by Uschold [306] includes some general steps to develop ontologies, which are (1) to identify the purpose, (2) to capture the concepts and the relationships between these concepts, and the terms used to refers to these concepts and relationships, and (3) to codify the ontology. The ontology has to be documented and evaluated. Other ontologies can be used to build the new one. They also outline requirements for a comprehensive methodology for building ontologies.

### 5.4.2 Ontology Engineering Tools

In [74] recently a survey on ontology tools was presented. The article compared the tools regarding a range of dimensions. Along dimensions, such as base language, import/export functionality, visualisation capabilities, etc. nearly 100 ontology tools were examined. Further, a comprehensive list of ontology engineering tools is available through [39].

A tool for supporting the ontology engineering process is [93]. For an overview on multimedia ontology annotation and engineering tools the interested reader is referred to Appendix C.2.

### 5.4.3 Review of Existing Multimedia Ontologies

The term *multimedia ontology* itself is inherently ambiguous. There exist communities that interpret any (formal) representation of multimedia metadata as multimedia ontology. Others argue that multimedia ontologies have at their core some modelling of media primitives—cf. also MPEG-7 basic descriptors in section 3.3.4 on page 45.

Historically, Hunter [171] provided the first attempt to model parts of MPEG-7 into RDF(S), covering the upper part of MPEG-7. Tsinaraki et al. [298] took the Hunter-ontology as a starting point and extend it to cover the full Multimedia Description Scheme (MDS) part of MPEG-7, ending up in an OWL-DL ontology.

In [107] Garcia and Celma proposed to automatically generate a multimedia ontology based on mappings from XSD to OWL producing a OWL-Full ontology. Common to all the above mentioned approaches is a fairly straight-forward translation of MPEG-7 descriptors, and descriptor schemes into OWL concepts, and roles: The interoperability problems—cf. Section 3.3.6—stay more or less the same.

Fundamental discussions on requirements for designing multimedia ontologies have been reported in the literature [293; 292; 189; 109]; we will further examine on this issue in the next chapter, 6. Regarding video metadata representation Hunter & Armstrong [173] paper compare the capabilities of the RDF(S), XML-Schema, and SOX for supporting and validating hierarchical video descriptions based on Dublin Core, MPEG-7 and a specific hierarchical structure. They propose a hybrid schema based on features from each of these schemas.

An important activity regarding multimedia ontologies has been performed in the aceMedia project[2]; a list of requirements for has been assembled and discussed in great depth. Eleftherohorinou et.al. [89] discuss this list based on input from a range of (EU) projects[3]. In particular [295] specifies the requirements as follows. A MMO should be interoperable with multimedia content description formats (as MPEG-7), should provide formal meaning for low-level features, and should be compliant with Semantic Web standards; further it should allow for complex data type representations, and uncertain information. Finally, a multimedia ontology should support the distinction between annotations addressing a digital artefact and the physical object depicted by a digital artefact.

In the realm of the EU FP6 K-Space project (the author is participating in) very recently a powerful multimedia ontology has been proposed: COMM—*A Core Ontology for Multimedia* [10]. COMM is based on both the MPEG-7 standard and the DOLCE[4] foundational ontology. A Java API for COMM is available[5], allowing to process and create media assets descriptions.

[2] See also section 2.2.4

[3] With input of the author of this thesis.

[4] http://www.loa-cnr.it/DOLCE.html

[5] http://comm.semanticweb.org/

Chapter 6

# Scaleable yet Expressive Content Descriptions

*"Via media."*

(Latin phrase)

To describe the content that a Semantic Web multimedia application deals with appropriately a language has to address a range of requirements. These requirements, where they stem from, and how they can be met is discussed in the following.

We note that this chapter is largely based on a paper written together with Tobias Bürger (Why Real-World Multimedia Assets Fail to Enter the Semantic Web [48]) and presented by the author at the Semantic Authoring, Annotation and Knowledge Markup (SAAKM07)[1] Workshop in Whistler, Canada in late 2007.

## 6.1 Introduction

Making multimedia assets on the one hand first-class objects on the Semantic Web, while keeping them on the other hand conforming to existing multimedia standards is a non-trivial task. Most proprietary media asset formats are binary, optimized for streaming or storage. However, the semantics carried by the media assets are not accessible directly. In addition, multimedia description standards lack the expressiveness to gain a semantic understanding of the media assets. There exists an array of requirements regarding media assets and the Semantic Web, already. Based on a critical review of these requirements we investigate how ontology languages fit into the picture. We finally analyse the usefulness of formal accounts to describe spatio-temporal aspects of multimedia assets in a practical context.

## 6.2 Motivation and Scenarios

Today a huge explosion of content can be experienced on the Web generated by and for home users [204]: An increasing number of people produce media assets (as photos, video

[1] http://saakm2007.semanticauthoring.org/index.html

clips, etc.), and share them on popular sites such as Flickr[2], and YouTube[3].

More recently, the popular attraction was guided away from image sharing to richer content sharing of videos. This can be seen by the launch of video portals like iFilm.com, Ziddio.com or the dozen of other portals that appeared recently to compete with YouTube[4].

Unsurprisingly there is already a portal called VideoRonk[5] trying to combine other portals by providing a MetaSearch interface, which is quite of an help as one does not want to search on ten or more different sites. However, what is missing is the link between the contents of all these sites, enabling distributed recommendations, cross-linking, etc.

Still, for example a cross-site search on the semantic level is close to impossible. The most obvious reason is due to a lack of metadata coming along with all the content. The power of providing metadata along with content on the Web can be seen at prospering mashups that not just combine APIs—provided by parties as Google[6]— but also trying to mashup things on a semantic level. This can be observed for example at Joost [268]. Having metadata about everything, as video content, blog posts, news feeds and the users of the system makes this new experience of watching TV through the Internet possible. To take this even one step further: Would every stream or video available on the Internet be described more detailed even content on the Internet could be matched with user profiles from applications like Joost and could be offered to watch.

As pointed out in [310; 229], high-quality metadata is essential for multimedia applications. Our recent work within initiatives [315] and research projects[7] has shown, there is a need for going beyond current metadata standards to annotate media assets. Current XML-based standards [141] are diverse, often proprietary and not ad hoc interoperable; see also [308].

In SALERO, for example, we are facing the problem to offer a semantic search facility over a diverse set of multimedia assets, e.g., image, videos, 3D objects or character animations. The same is true for the Austrian project GRISINO[8] where we aim to realize a semantic search facility for cultural heritage collections. Automating the handling of metadata for these collections and automating linkage between parts of these collections is hard as the vocabularies to describe them are mostly diverse and do not offer facilities to attach formal descriptions.

**A Motivating Scenario**. Imagine a person that wants to watch the recent clips similar to the ones of his favourite experimental artist. Tons of clips are potentially distributed on the Web, which makes searching for them sometimes time consuming and laborious. Thus a central facility to search for and negotiate content is needed. This facility should allow to formulate a search goal, including the characteristics, the subject matter, a maximum price, and the preferred encoding and file format of the clip. In a next step, all portal offerings will be scanned in order to retrieve and negotiate content that matches the users' intention. Note that also parts of a video may match his intention which means that videos need to be fine granular and sufficiently well enough described.

In order for this scenario to work, the descriptions of (1) the goal formulation, (2) the description of the media content by all content owners and (3) the negotiation semantics

[2] http://www.flickr.org
[3] http://www.youtube.com
[4] http://www.youtube.com
[5] http://www.videoronk.com
[6] http://code.google.com/apis/
[7] as, e.g., EU project SALERO, http://www.salero.info
[8] http://www.grisino.at

have to be compatible. Three important focal points of these semantic descriptions are:

- *Expressivity for high level semantic descriptions of content* as typical users are not thinking in terms of colour histograms and spatial / temporal constructs. The characteristics of the media should be described detailed enough.
- The *need for rules*: To effectively identify the part of the content that matches the users' intention, rules are needed to map high level semantic concepts to spatial and temporal segments of the video (eg., because ratings and classifications could only apply to parts of the content, ie., a scene including crime is only suitable for adults)
- *Fine grain semantic descriptions* as of bandwidth, user effort, or cost reason to transfer the whole content is not possible. Thus parts of the content should be described detailed enough.

To reach out, we want to provide answers to the question: Why do we need rich semantic descriptions of media assets on the Web, and (why) is there a need to bundle these descriptions together with the multimedia assets? Simultaneous, we want to provide answers to the questions: How can descriptions be provided? Why are the metadata features of multimedia standards not enough?

## 6.3 Requirements for the Description of Multimedia Assets

Requirements for multimedia content descriptions have been researched in a number of papers [109; 310; 229; 49] before and investigations of the combination of multimedia descriptions with features of the Semantic Web are yet numerous [171; 11; 290; 298; 10]. In the following, we give a summarisation of the proposed requirements and add two additional ones (Authoring & Consumption and Performance & Scalability).

**Representational Issues** A basic prerequisite is the formal grounding and neutral representation of the format used to describe multimedia assets.

- *Neutral Representation:* The ideal multimedia metadata format has a platform and application independent representation, and is both human and machine processable;
- *Formal Grounding:* Knowledge about media assets must be represented in formal languages, as it must be interpretable by machines to allow for automation.

**Extensibility & Reusability** It is requested that the format at hand is extensible, e.g., via an extension mechanism as found in MPEG-7. It should be possible to integrate or reference existing vocabularies [141].

**Multimedia Characteristics and Linking** The format should reflect the characteristics of media assets, hence allow linking between data and annotations:

- *Description Structures*. The format should support description structures at various levels of detail, including a rich set of structural, cardinality, and multimedia datatyping constraints;

- *Granularity*. The language has to support the definition of the various spatial, temporal, and conceptual relationships between media assets in a commonly agreed-upon format;

- *Linking*. It has to facilitate a diverse set of linking mechanisms between the annotations and the data being described, including a way to segment temporal media.

**Authoring & Consumption** A major drawback of existing metadata approaches is its lacking support for authors in creating annotations along with the lacking benefits of generated annotations.

- *Engineering support*. Appropriate tools are a prerequisite for uptake of new vocabularies. There is the need for at least authoring and consumption environments making use of the vocabularies to demonstrate their usefulness.

- *Deployment*. Multimedia Assets need to be exchangeable, and there must be ways to deploy descriptions along with the assets.

**Performance & Scalability** The language should yield descriptions that can be stored, processed, exchanged and queried effectively and efficiently.

## 6.4 Environment Analysis: The Semantic Web

A good starting point for the analysis of our targeted hosting environment—the Semantic Web—is the Architecture of the World Wide Web [176], in which its three main building blocks are discussed: *identification*, *interaction*, and *data formats*. The Semantic Web, as an extension of the well-known Web roughly has the following characteristics:

- It is a *highly distributed system*. *Identification of resources* is based on URIs—for both data and services;

- There is no single, central "registry", viz. *authorities are decentralised*; data and metadata are under control of a lot of distinct individuals (companies, standardisation bodies, private, etc.)

- Alike in the Web a fundamental building block are *relations* between data, whereas the relations in the Semantic Web are *named*, may be of *any granularity* and allow the automatic interchange of data;

- *Contribuser*[9] inhabit it; each participant may play different roles at once: consuming content and contributing via comments, links, etc.

- Finally, there exists a number of standards. As, RDF allowing formal definitions of the intended meaning, SPARQL for querying, RDF(S), OWL or SKOS to classify content and OWL, WSML, or RIF for describing logical relationships.

[9] a portmanteau word; contributor and user

Any multimedia metadata format that is after the successful application on the Semantic Web has to be in-line with the above listed characteristics. While some requirements, as formats (e.g. XML) are rather easy to meet, other can pose serious problems regarding the integration into the Semantic Web. For example, the requirement to address resource using URIs often is an issues; not all known multimedia formats are able to handle URIs, at least not in a complete way.

## 6.5 Multimedia Assets on the Semantic Web

Firstly, addressing the environmental requirements together with an efficient layering of the semantic descriptions on top of the existing metadata (sub-symbolic level - symbolic level - semantic level) is a necessary prerequisite for multimedia assets to enter the Semantic Web successfully. Secondly, from the requirements gathered in section 6.3 and the environmental analysis done in section 6.4 we deduce the following characteristics for multimedia assets on the Semantic Web:

**Formality of Descriptions** Formal descriptions are the basic building blocks of the Semantic Web. To enable automatic handling like retrieval, and negotiation of multimedia assets formality of descriptions is a pre-requisite.
Three different (semantic) levels of multimedia metadata can be identified [109]: (1) At the *subsymbolic layer* covering the raw multimedia information typically binary formats are used which are optimized for storage or streaming and which mostly do not provide metadata. (2) The *symbolical layer* provides an additional structural layer for the binary essence stream. For this level standards like MPEG-7, Dublin Core or MPEG-21 can be used. The semantics of the information encoded with these standards are only specified within each standards framework. (3) Therefore the *semantic and logical layer* is needed to provide the semantics for the symbolical layer. This layer should be formally described.

**Efficient layering and referencing of descriptions** It is necessary to support different levels of meaning attached to multimedia assets, i.e., meaning at the bit-level, *traditional metadata* and semantic (high-level) information. As there are already widely adopted standards available for the description of multimedia assets, the semantic layer must be efficiently put upon those traditional description layers and should not aim to replace it. Furthermore *semantic* descriptions from these traditional layers shall be re-used. As content, parts of content, and traditional and semantic descriptions may be distributed efficient referencing mechanisms for multimedia content must be present.

Based on recent discussions[10] we now give a summarisation of the approaches and conclude on the advantages and disadvantages.

The setup is as follows: the multimedia asset is denoted with `A`, for the multimedia metadata format (such as MPEG-7) we write `M`, the ontology language is `O`, and finally an external reference mechanism[11] is labelled with `R`. The linking is depicted with $\hookrightarrow$. The following combinations might be taken into account:

- `M` $\hookrightarrow$ `A`. That is, the assets content is referenced from the multimedia metadata format; the ontology layer has to deal with it, separately.

[10] `http://lists.w3.org/Archives/Public/public-xg-mmsem/2007Apr/0002.html`
[11] such as `http://www.annodex.net/TR/URI_fragments.html`

- `M` $\hookrightarrow$ `O`. The multimedia metadata format references the ontology layer.
- `O` $\hookrightarrow$ `M`. The ontology layer references the multimedia metadata format.
- `O` $\hookrightarrow$ `A`. The ontology layer references the asset content directly.
- `O`, `M` $\hookrightarrow^R$ `A`. The ontology layer and the multimedia metadata format use a common reference mechanism to link to the content of a multimedia asset.

**Interoperability among descriptions** Many formats used in various communities cause interoperability problems when dealing with multimedia content. To overcome this, an RDF based semantic layer should be added on top of these numerous formats to ease their semantic and syntactic integration. However, there are some open problems resp. the integration of existing annotation standards and semantic approaches [310; 229]: The stack of Semantic Web languages and technologies provided by the W3C is well suited to the formal, semantic descriptions of the terms in a multimedia document's annotation. But, as also pointed out in [281], the Semantic Web based languages lack the structural advantages of the XML-based approaches. Additionally, there is a huge amount of work already done on multimedia document annotation within the framework of other standards. This is why a combination of the existing standards is the most promising path for multimedia document description in the near future.

**Subjectivity and granularity of descriptions** Opinions and views of content differ among users because of their personal background, culture or previous experiences. As many users are potential contributors to descriptions of assets, opinions may differ. Many of these opinions sometimes do not serve to a unique whole opinion. This is why it should be possible to separately attach these opinions to multimedia assets and keep them separate.

**Trust and IPR issues** The Web consists of decentralized authorities and a huge number of contribusers. As descriptions of content—especially in the new changing Web 2.0 environment—are subject to vandalism, there need to be ways to guarantee the validity of the descriptions and to secure descriptions that are just read-only for a user group. Popular portals like Flickr or YouTube show that there is no need to own content in order to annotate it. Furthermore copyright is critical when dealing with multimedia content.

**Functional Descriptions** Sometimes the fact that metadata is created to support some specific function is forgotten when summarizing the requirements for a metadata schema. For the metadata creator it should be clear beforehand for what purpose the metadata will be used and what benefits he gains from it [215], ie., using this part of the metadata scheme enhances retrieval, raises social attention or helps you protect your assets.
This in turn also applies to the consumer of the metadata, functional descriptions of what type of information can be inferred from the attached metadata or what type of actions can be performed on the content are essential: this is especially true for information that is obfuscated prior to a possible negotiation phase of the content.

**Engineering Support** The presence of metadata is a prerequisite to make multimedia assets accessible, and deployable on the Semantic Web, hence to enable their automated processing. From a developers perspective, there must be tools and standards enabling an integrated authoring, testing, and deployment of multimedia assets along with their associated metadata. In the following the most important areas of engineering support are listed:

- *Edit & Visualise*. To aid the engineer in handling the annotations, editor tools, and IDEs[12] are needed. These may include validator services[13], converter or mapper, and visualisation modules.
- *Libraries & Applications*. When developing applications, the availability of APIs is a core requirement. In special for Semantic Web applications, interface and mapping issues are of importance [113].
- *Deployment* Multimedia containers as HTML, SMIL, etc. require the metadata either being referenced from within the media assets, or being embedded into it. As the data model needs to be RDF—in contrast to existing, flat (tags, etc.) technologies—upcoming approaches as RDFa [136; 2; 5] need to be utilised thoroughly.

## 6.6 Formal Descriptions of Multimedia Assets

In this part ontology languages which are thought to be used for the advanced requirements which were identified in the sections before. In its core it comprises a a comparison of two families of ontology languages against the requirements postulated in section 6.5.

The reader is invited to note that not all of the existing languages have the same expressiveness and not all have the same inferential capabilities. Further, the underlying knowledge representation paradigms (such as Description Logics, Logic Programming, etc.) can differ. Corcho and Gomez-Perez [114] present a framework that allows for analysing and comparing the expressiveness and reasoning capabilities of ontology languages, which can be used in the decision process. The process of choosing and selecting the appropriate ontology language includes questions like:

- What expressiveness does an ontology language have?
- What are the inference mechanisms (reasoning capabilities) of it?
- Are there any supporting tools for that language?
- Is the language appropriate for exchanging ontologies between applications?
- Are there translators that transform the ontology implemented in a source into a target language (to enhance reusability, exchangeability or interoperability)?

We are going to take these questions into consideration and simultaneously verify if the languages meet the requirements discussed in section 6.5.

### 6.6.1 Ontology Languages

A number of logical languages have been used for the description of different kinds of knowledge (i.e., ontologies and rules) on the Semantic Web: First Order Logic, Description Logics, Logic Programming and Frame-based Logics. Each of which allow the description of

[12]as for example http://www.topbraidcomposer.com/
[13]http://phoebus.cs.man.ac.uk:9999/OWL/Validator

different statements and each imply different complexity results for certain reasoning tasks with these languages.

In this section we want to introduce two of the most promising ontology language families, i.e., the OWL- and the WSML- family of languages. The OWL family of languages is a standardisation effort of the W3C and the WSML family of languages is an effort of the WSMO working group, whereas WSML is a formal language for the description of ontologies and Semantic Web Services.

### Web Ontology Language (OWL) Family

The Web Ontology Language (OWL) family was designed in a W3C standardisation process because of the need for an ontology language that can be used to formally describe the meaning of terminology used in Web documents, thus, making it easier for machines to automatically process and integrate information available on the Web. This language should be layered on top of XML and RDF (W3C's Resource Description Framework[14]) in order to build on XML's ability to define customized tagging schemes and RDF's approach to representing data.

OWL has some well-known limitations regarding the composition of properties [162; 166]. On the other hand OWL 2[15] is under development; it extends OWL in several ways: the underlying DL now is is $\mathcal{SROIQ}$, which provides increased expressive power with respect to properties and cardinality restrictions. Further, OWL 1.1 has user-defined datatypes and restrictions involving datatype predicates, and a weak form of meta-modelling known as punning.

The usage of rules in combination with DL has been investigated for some time [83; 117]—in the Semantic Web stack, it is expected that a rule language will complement the ontology layer.

### The WSML family of languages

The activities of the WSMO Working group[16] have yielded proposals of new ontology languages, namely WSML (WSML-Core, WSML-DL, WSML-Flight, WSML-Rule, WSML-Full), OWL- ("OWL minus") [71] and OWL Flight [73]. OWL- is a well-founded reduction of OWL that combines efficient reasoning with a high degree of expressiveness. In [99] unique key

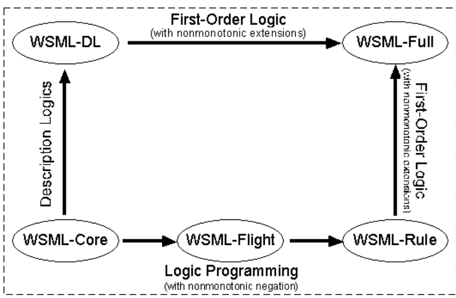



Figure 6.1: WSML Variants

[14] http://www.w3.org/TR/rdfprimer/
[15] http://www.w3.org/2007/OWL/wiki/OWL_Working_Group
[16] http://www.wsmo.org

features of WSML in comparison of other language proposals are presented. Compared to OWL key features include (1) WSML offers one syntactic framework for a set of layered languages, and (2) it separates between conceptual and logical modelling.

- One syntactic framework for a set of layered languages
- Normative, human readable syntax
- Separation of conceptual and logical modelling
- Semantics based on well-known formalisms
- WWW language
- Frame based syntax

The different variants of the WSML framework, all having different expressiveness [195], are:

- **WSML-Core** corresponds with the intersection of DL and Horn Logic (without function symbols and without equality), extended with datatype support in order to be useful in practical applications. WSML-Core is fully compliant with a subset of OWL.
- **WSML-DL** extends WSML-Core to an expressive Description Logic, namely, SHIQ, thereby covering that part of OWL which is efficiently implementable.
- **WSML-Flight** extends WSML-Core in the direction of Logic Programming. WSML-Flight has a rich set of modelling primitives for different aspects of attributes, such as value and integrity constraints. Furthermore, WSML-Flight incorporates a rule language, while still allowing efficient decidable reasoning. More precisely, WSML-Flight allows any Datalog rule, extended with inequality and (locally) stratified negation.
- **WSML-Rule** extends WSML-Flight to a fully-fledged Logic Programming language, by allowing function symbols and unsafe rules.
- **WSML-Full** unifies all WSML variants under a common First-Order umbrella with non-monotonic extensions which allow to capture non-monotonic negation of WSML-Rule.

### The Relation of WSML to OWL

The relation of WSML to OWL is presented in [72]: WSML-Core is a semantic subset of OWL-Lite and WSML-DL is semantically equivalent to OWL-DL. A major difference between ontology modelling in WSML and ontology modelling in OWL is that WSML separates between conceptual modelling for the non-expert users and logical modelling for the expert user as it—unlike OWL—uses an epistemology which abstracts from the underlying logical language. WSML-Flight and WSML-Rule are based on the Logic Programming paradigm, rather than the Description Logic paradigm. Thus, their expressiveness is quite different from OWL. On the one hand, WSML-Flight/Rule allow chaining over predicates and non-monotonic negation, but do not allow classical negation and full disjunction and existential quantification. With WSML Logic Programming and the Description Logics paradigm are captured in one coherent framework whereas interaction between the paradigms is achieved through a common subset, WSML-Core.

### 6.6.2 Rules

Due to the manifold availability of rule systems, harmonisation efforts have not been successful so far. A relatively new W3C initiative, the Rule Interchange Format Working Group, is now after defining a core rules language for exchanging rules. This Rule Interchange Format Core[17] (RIF Core) language aims at achieving maximum interoperability while preserving rule semantics; from a theoretical perspective, RIF Core corresponds to the language of definite Horn rules. As standardisation is still in its infancy, we will not go further into detail regarding rules, but one has to note that the careful integration of ontology languages is an issue to be addressed; for example the usage of DL concepts in a rule has to be well-defined.

### 6.6.3 Comparing Formal Descriptions Regarding the Requirements

In the following a high-level comparison of formal description paradigms for multimedia assets is performed. We chose OWL+RIF on the one side, and WSML/OWL-Flight on the other to achieve a somehow realistic scenario; the result can be found in Table 6.1[18].

| **Requirement** | **OWL 1.1 + RIF** | **WSML/OWL-Flight** |
|---|---|---|
| Formal Description | ++ | ++ |
| Layering of Descriptions | + | + |
| Interoperability | ++ | + |
| Granularity | + | + |
| Trust & IPR issues | - | - |
| Functional Descriptions | - | + |
| Engineering Support | ++ | + |
| Datatype Support | + | ++ |

Table 6.1: Comparison of Formal Descriptions for Media Assets.

In the following, we elaborate in detail on each of the items in Table 6.1, and argue therefore our findings regarding the comparison of OWL 1.1 + RIF vs. WSML/OWL-Flight.

#### Formal Description

Both OWL and WSML provide a framework for the formal (machine-processable) descriptions of ontologies. An ontology in WSML consists of the elements concept, relation, instance, relationInstance and axiom. The primary elements of an OWL ontology concern classes and their instances, properties, and relationships between these instances. The formality of the descriptions is based on logics that allow machines to reason on the information. Whereas OWL is based on Description Logics, the WSML family members are based on different logic languages (i.e., Description Logics, Logic Programming or First Order Logic).

Despite the fact, that OWL is more widely adopted and used we believe that WSML with its layered framework is conceptually superior to OWL. A major difference between ontology modelling in WSML and ontology modelling in OWL is that WSML separates conceptual modelling for the non-expert users, and logical modelling for the expert user as

[17] http://www.w3.org/TR/rif-core/

[18] ++ . . . good support, + . . . available , - . . . not supported

it—unlike OWL—uses an epistemology, which abstracts from the underlying logical language making the *surface syntax* nicer. Even if an application later requires OWL, one is able to use WSML tools to convert ontologies that reside in popular logic/language fragments automatically into equivalent OWL ontologies. Furthermore the WSML family framework enables one to choose exactly which language with the needed expressiveness is intend to be used, and later allows an easy switch to another family member because of its common grounding. WSML Rule and WSML Flight also include rule-support. Thus, unlike with OWL, no additional rule language is needed.

#### Layering of Descriptions

An array of existing multimedia metadata formats have been used for years in diverse application areas. However, when one aims at using these formats (as MPEG-7, ID3, etc.) in the context of the Semantic Web, the options are limited. Hence, to enable an efficient layering of RDF-based vocabularies on top of existing multimedia metadata, one may use hybrid techniques.

As a result of our works in the media semantics area, we have recently proposed the RDFa-deployed Multimedia Metadata (ramm.x) specification [138]. ramm.x is a light-weight framework allowing existing multimedia metadata to hook into the Semantic Web using RDFa [2]; see also chapter 10.

A different but as well Web compatible approach is described in [184]. There, the authors propose the concept of semantic documents; semantic documents include any information regarding the document and its relationships to other documents. The concept is realised by including XMP descriptions in PDF documents which can be rendered in any browser with available plugins. XMP is a format for embedding metadata in documents using RDF.

#### Interoperability

To adhere to the architecture of the Web, OWL uses (1) URIs for naming, (2) RDF to provide extensible descriptions, (3) builds on RDF Schema and adds additional vocabulary for describing properties and classes, and (4) the datatype support for OWL is grounded on XML Schema.

WSML has a number of features which allow to integrate it seamlessly in the Web. It (1) uses IRIs[19] [88] for the identification of resources, (2) adopts the namespace mechanism of XML, and WSML and XML Schema datatypes are compatible, (3) has an XML-based and an RDF-based syntax for exchange over the Web. To reach compatibility between WSML and OWL, WSML has a set of defined translators between OWL and WSML [75; 76].

#### Granularity

As stated above, when referring to granularity, we understand the support of the definition of various spatial, temporal, and conceptual relationships regarding annotations. In this sense, OWL and WSML meet the minimal requirements, but do not explicitly address this issue. Depending on the granularity, obviously scalability and performance issues come along. In this respect, again, OWL and WSML can be perceived comparable.

[19] IRIs are the successors of URIs

#### Trust and IPR

In an interdependent, interconnected environment as the Semantic Web, two important aspects immediately arise: data provenance and trust [38]. Requirements regarding trust issues gathered from [235; 112] contain costs and benefits w.r.t. implementation, technology-driven vs. social networking, etc.

Both WSML and OWL do not have explicit provisions for handling trust and IPR issues, respectively.
However, as WSML also is a language capable of describing different aspects of a Web service, it is capable of describing so called "non-functional properties". Non-functional properties typically are used to constrain functional—i.e. the formal specification of what a service can do—and behavioural aspects, namely how the functionality can be achieved in terms of interaction of a Web services; they may also be utilised to specify trust and IPR properties [238].

#### Functional Descriptions

WSML is a language for the specification of ontologies and different aspects of Web services. As such it not only provides means for modeling and description of ontologies but also functional (service) descriptions, i.e. the description of a service capability by means of precondition, assumptions, postconditions and effects [182]. In contrast, OWL does not support such kind of descriptions.

#### Engineering Support

Tool support for WSML and especially OWL is constantly growing. However, the amount of tools available for OWL [316] and WSML [77] can drastically not be compared. As OWL is a W3C Recommendation, the support for it is huge.

#### Data Type Support

The reader is invited to note that both OWL and WSML ground their datatype support on XML Schema. In WSML, XML Schema primitive datatypes, simple types and XML Schema derived datatypes are supported [261]; OWL adopts the RDF(S) specification of datatypes [244], though some XML Schema built-ins are problematic.

## 6.7 Conclusions

The first question we kept open is "What are real-world multimedia assets"? Real-world multimedia assets are multimedia objects which can be currently found embedded in HTML pages on the Web, such as images, videos, etc. We see two main reasons why media assets (currently) fail to enter the Semantic Web.

**No critical mass of annotated content on the Web** This is mainly due to the large scale automation of (semantic) visual analysis has not gone that far. This is why the user is the central person in the process in order to provide manual annotations. Motivating user to attach complex annotations to content is not easy to achieve.

**Shallow approaches to multimedia annotation are not useful to achieve the goals of the Semantic Web** Recalling, the most important aspects that the Semantic Web intends to solve are:

- Annotation, i.e., how to associate metadata with a resource;
- Information integration, i.e., how to integrate information about resources;
- Inference, i.e., reasoning over known facts to unleash hidden facts.

Existing multimedia metadata standards such as MPEG-7 can be used to annotate but keep a certain amount of ambiguity amongst these annotations. As it is a standard it allows easy integration based on it, but inference is not possible with the information attachable to an MPEG-7 document. The problem with tagging is manifold: There are a lot of open issues regarding tagging: (a) how can you guarantee consistency among tags of different users? (b) how do you reconcile tags? (c) how do you associate tags with parts of the tagged content? This huge amount of uncertainty will not allow reliable information integration, nor allow to reason on it.

However, recent research such as OntoGames [270]) make us believe that it is possible to combine the participative nature of Web 2.0 and semantic technologies by bringing in an incentive that ordinary users create high-level annotations needed to enable multimedia assets to finally enter the Semantic Web.

# Part III

# SWMA Engineering

Chapter 7

# Rational & Common Concepts

*"Ora et labora."*

(Latin phrase)

After an in-depth introduction of the foundations of both multimedia metadata and Semantic Web in part I and a discussion of requirements regarding Semantic Web multimedia applications in part II this third part presents the genuine idea of the work.

In this chapter we explain our view on the Semantic Web stack regarding Semantic Web multimedia applications and outline basic design principles. Further, common concepts used throughout in the subsequent chapters are introduced. An analysis and according definitions for both expressivity and scalability in the context of this work is given. The reader is invited to note that this chapter forms the base for the following three chapters 8, 9, and 10.

## 7.1 The Semantic Web Stack regarding SWMA

In section 5.1 we have initially discussed possible views on the Semantic Web stack. However, in the realm of this work, we subscribe to the view as depicted in Fig. 7.1.

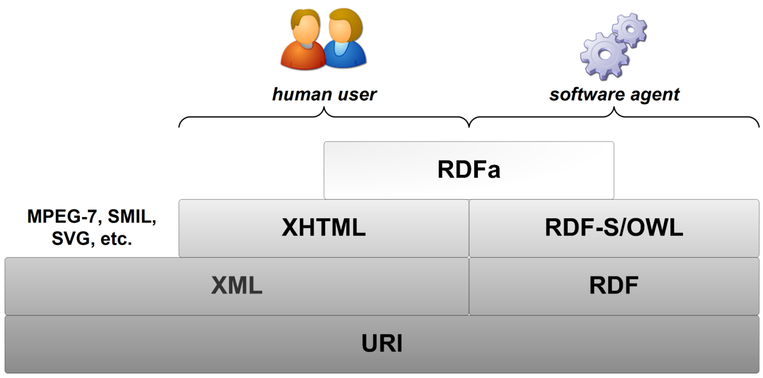



Figure 7.1: The Semantic Web Stack in the context of this work.

While most representations of the Semantic Web stack do not explicitly take the duality of the Semantic Web into account, we differentiate information entities as follows:

- Information primarily targeting at human users. This is the content, e.g., a video on a Web page that a person consumes, edits, or somehow processes;

- Information primarily targeting at software (or, more specially, Semantic Web agents). This information, commonly called metadata is *about* the content, for example key frames of the video represented in MPEG-7, or usage information, e.g., in RDF.

## 7.2 Design Principles and Common Concepts

### 7.2.1 Occam's Razor

*Occam's razor*[1] is a principle stating that the explanation of any phenomenon should make as few assumptions as possible, eliminating those that make no difference in the observable predictions of the explanatory hypothesis or theory. Alternatively this principle is known as the *KISS principle*, an acronym of the phrase "Keep It Simple, Stupid", or as Albert Einstein puts it: "everything should be made as simple as possible, but no simpler".

We note that when applying this principle in the design of a Semantic Web application, this has the consequence that one is likely to use expressive, yet scaleable technologies. For a deeper discussion on this topic the reader is referred to the discussion on "The Rule of Least Power"[2].

### 7.2.2 Follow-your-nose

While the KISS principle is a quite generic design principle, "follow-your-nose" (FYN)[3] is a cornerstone of the Web architecture [176]. It describes the general practice of performing Web retrieval on URIs in a knowledge base to obtain more knowledge. In the context of messaging, follow-your-nose is about walking through a message, outside-in, and interpreting semantics via the successive application of unambiguously associated specifications.

There is indeed a deep and ongoing discussion on various Semantic Web channels about this principle; see for example the Linking Open Data mailing list archive[4]. The RDF-based metadata world as well as the microformats community—with GRDDL as its ultimate bridge—acknowledge the FYN principle. The reader is referred to a talk Dan Connolly and Harry Halpin have given in early 2007: "Deploying Web-scale Mash-ups by Linking Microformats and the Semantic Web"[5].

Further, recently a promising article about FYN on the Web of data has been published[6] and the W3C TAG has an according draft finding[7] in its pipeline.

[1] http://en.wikipedia.org/wiki/Occam's_Razor
[2] http://www.w3.org/2001/tag/doc/leastPower
[3] http://esw.w3.org/topic/FollowLinksForMoreInformation
[4] http://simile.mit.edu/mail/SummarizeList?listId=14
[5] http://www.w3.org/2001/sw/grddl-wg/tut7/gtut2.html
[6] http://inkdroid.org/journal/2008/01/04/following-your-nose-to-the-web-of-data/
[7] http://www.w3.org/2001/tag/doc/selfDescribingDocuments.html

### 7.2.3 Reuse & Layering

Reusability is fostered by the wide deployment of existing vocabularies (such as FOAF and SIOC). Further, good practice on using and extending vocabularies is gathered by the community in diverse groups (for example the RDF Schema Development discussion group[8]) and the development is supported by upcoming tools such as Neologism[9]. The increasing usage—along with the implicit agreement on the semantics of the vocabulary—of simple vocabularies in turn enables richer vocabularies (for examples domain ontologies such as the Gene Ontology[10]) to settle.

Currently, (X)HTML and RDF exist side by side. Though there are (non-standard) ways to discover Semantic Web content, GRDDL and RDFa are both approaches enabling a *real* layering of the Web of Data on the current Web of Documents.

## 7.3 Expressivity on the Semantic Web

Based on the analysis of related and existing work in Section 2.3 we give a more specialised explanation of *Expressivity on the Semantic Web*, as we understand it in this work.

Following Felleisen [95], who developed a formal notion of expressivity regarding programming languages, and the Description Logic Handbook [13] we understand expressivity as follows.

**Definition 7.1 (Expressivity on the Semantic Web).**
Let $\mathbb{L}$ be a knowledge representation language with formal grounding. Expressivity is a measure for the computational complexity of the reasoning algorithms of $\mathbb{L}$.

❖

The following simple usage of RDF(S) illustrates the above definition in a practical context. In example 7.2 a light-weight multimedia schema is defined, along with a couple of instances.

***Example 7.1 (Expressivity in RDF Graphs).***
The RDF graph[11] in Fig. 7.2 (page 128) represents a knowledge base consisting of a T-Box and a A-Box capturing some basic multimedia properties. From this representation it can be inferred, for example, that `v1` is of type `Content` and is "about" `Economy`. □

When RDF-S inference is performed on the RDF graph from example 7.2 a reasoner would effectively added edges (that is, RDF properties) based on the RDF-S entailment rules[12]. Consequently, the graph density increases.

#### Open World Reasoning

In his PhD [151], Heflin argued that "... to scale to the size of the ever growing Web, we must either restrict the expressivity of our representation language or use incomplete reasoning algorithms." Recently this issue was even further exaggerated [98]:

[8] http://groups.google.com/group/rdf-schema-dev
[9] http://neologism.deri.ie/
[10] http://www.geneontology.org/
[11] For the source of the RDF graph see Appendix A.1 on page 199.
[12] http://www.w3.org/TR/rdf-mt/#RDFSRules

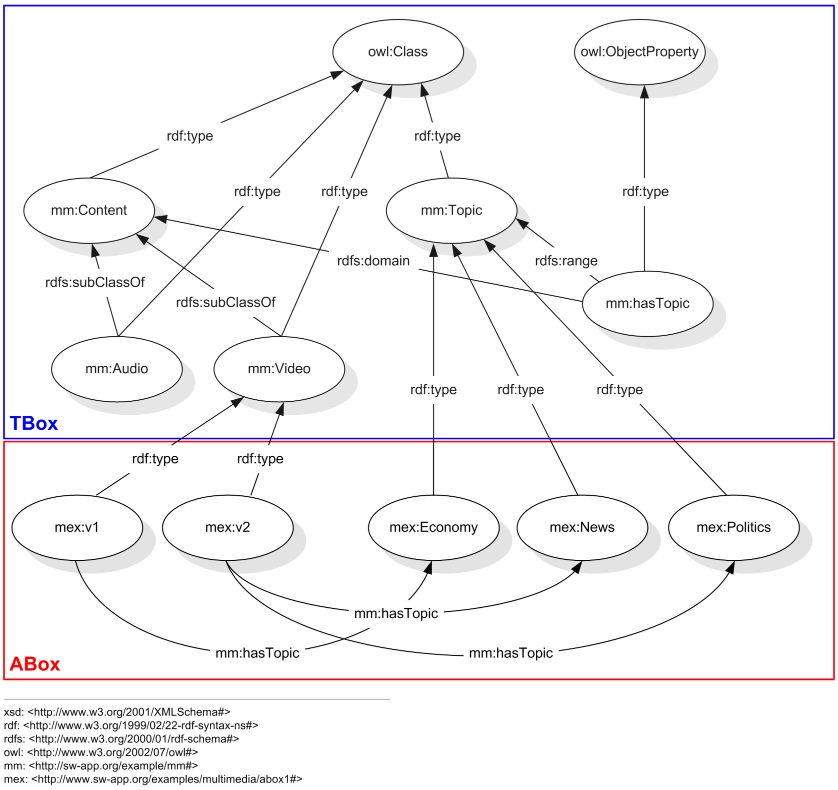



Figure 7.2: A sample RDF graph and basic RDF-S inference.

> Spoken cynically, current reasoning engines have inherited clumpy syntax from the Web (XML, RDF, and URIs), and in return, the Web has received toy engines that neither meet its requirements nor scale to its size. Basically, both sides have been aligned at a level too superficial to generate something useful. The basic underlying assumptions of pure logical reasoning dont seem to match the reality the Web provides.

Recently, research efforts have started that cover this issue, notably the EU Large-Scale Integrating Project LarKC (Large Knowledge Collider)[13]. LarKC aims at building a platform for massive distributed incomplete reasoning, targeting at the removal of the scalability barriers of currently existing reasoning systems for the Semantic Web.

## 7.4 Scalability on the Semantic Web

Referring to the analysis done in Section 2.3 we explain our understanding of *Scalability on the Semantic Web* in the realm of the work at hand.

A Semantic Web application—cf. definition 2.2 on page 13—has, per definition (i) a potential huge number of users, and (ii) the schemas and instances used in the Semantic Web application are under distributed control. Therefore, scalability is a strong requirement to meet—this is true per design and should not be conceived as a later add-on.

[13] http://www.larkc.eu/

A recent blog post[14] highlights this:

> Scalability has little to do with performance; moreover, a scalable solution is one that does not need to change when the problem size increases. Vertical scaling (which is buying bigger machines to crunch the problem faster) is inherently limited by the hardware available on today's market; and while it is always getting faster, it is likely that your cool new idea will manage to exceed the capacity of one of these monsters. Horizontal scaling (adding more machines, not faster ones) is what "scalability" is all about. And clearly this has nothing to do with performance.

Regarding SWMA, the reasons why the amount of multimedia metadata may grow are obvious; several possible reasons can be identified:

- The **number of users** (content consumer and/or producer) may increase. As this is not in the primary scope of this work, this aspect will not further be investigated;
- The **amount of data** can increase. In the same manner the amount of metadata may increase, either directly due to an (automatic) extraction task performed on the content or indirectly through (manual) annotation;
- Even when neither the user base increases, nor the content size grows, the amount of metadata may increase due to a **raise of detail of the content description**, for example, a transition from a global to a fine-grained description;
- Finally, the **application itself can grow** in terms of implemented features or interfaces to external systems.

In the following scalability issues with basic Semantic Web technologies—URI, XML, RDF, and OWL—are discussed.

**URI Scalability** In [32] Tim Berners-Lee stated that

> URIs are central to the W3 architecture. The fact that it is easy to address an object anywhere on the Internet is essential for the system to scale, and for the information space to be independent of the network and server topology.

The reason why URIs are scaleable is simple—they use a hierarchical structure, allowing for independency between it components.

**XML Scalability** Performance and scalability issues of XML have been discussed, recently. For example in [319] typical requirements for data retrieval from XML documents are examined. This work provides an in-depth analysis of the existing XML data retrieval strategies and why they are not optimal.

However, there exists a number of reasons why XML is believed not to be scaleable in our context.

- XML is verbose by nature[15];

[14] http://www.lethargy.org/~jesus/archives/91-Scalability-vs.-Performance-it-isnt-a-battle.html

[15] http://msdn2.microsoft.com/en-us/library/ms998559.aspx

- Due to the tree-based structure of XML it is cumbersome to represent certain data structures;
- Especially regarding the layering of RDF, there exists a number of issues [26] preventing an adoption on the Web-scale;
- Along with XML, the XML Schema is not scaleable per design. It does not allow for distributed extensions and overloading.

While in closed domain applications—for example in a desktop environment—certain of the above mentioned issues may not effect the overall scalability behaviour, their impact on Semantic Web applications can not be disregarded.

**RDF Scalability** The RDF model is scaleable by design. However, there are issues regarding the concrete serialisation of RDF in its various forms[16], as RDF/XML [26] or embedded formats alike. Note that the latter aspect will be investigated in greater detail in chapter 8.

**RDF-S & OWL Scalability** As RDF-S and OWL both are building on top of RDF, they are believed to be scaleable, but with certain limitations[17]. In particular when is comes to reasoning with certain vocabularies their exist fundamental issues to be resolved [98].

To summarise the above given analysis, we give the following definition of scalability in the context of the Semantic Web.

**Definition 7.2 (Scalability on the Semantic Web).**
Scalability w.r.t. the Semantic Web is a system-intrinsic property mainly depending on two factors:

- *Decentralisation*. The degree and type of format, to which the metadata is distributed over the Web. This factor influences the time that is needed to gather, convert, etc. the metadata into a central point of processing (for example to execute a SPARQL query);
- *Inference*. The type and breadth of inference, i.e., making implicit facts explicit, a system is expected to carry out. It heavily depends on the formalisation used, but is also dependent on the decentralisation factor, above. ❖

For example, in the design of voiD (the "Vocabulary of Interlinked Datasets"[18]), which allows the description of linked datasets on the content-level, we have decided to let the linked dataset providers keep their annotations locally. That is, each maintainer of a linked dataset (such as DBpedia, Geonames, riese, etc.) offers voiD-descriptions at his or her site. In a second step a Semantic Web indexer (such as `http://sindice.com`) can gather the descriptions and offer a single point of access for discovering the linked data sets; this is a scaleable, yet expressive solution per design.

[16] See Appendix A of [104] for a detailed history of RDF Serialization formats
[17] `http://code.google.com/p/owl1-1/wiki/P2NotesOwled2007`
[18] `http://community.linkeddata.org/MediaWiki/index.php?VoiD`

## 7.5 Conclusion

In this chapter we have discussed general design principles regarding Semantic Web multimedia applications and presented our view on the Semantic Web stack. In order to address the issues raised above, this thesis contributes to the research as described in the following:

- On the **RDF-level**, we propose a performance and scalability metric and evaluate it. This work is described in chapter 8;
- Regarding the **formal representation of multimedia vocabularies**, a scalable multimedia ontology is proposed as described in chapter 9;
- Finally, to address an effective and efficient **multimedia metadata deployment**, a new framework is proposed and its application is discussed in chapter 10.

Chapter 8

# A Performance and Scalability Metric for Virtual RDF Graphs

*"Cum grano salis."*

(Latin phrase)

This chapter is largely based on a paper [149] the author of this thesis has published together with Wolfgang Slany and Danny Ayers. The author has presented the paper at the $3^{rd}$ Workshop on Scripting for the Semantic Web (SFSW07), in Innsbruck, Austria, in June 2007. The original research paper has been extended and the discussions at the workshop have been incorporated.

In this chapter we focus on the data model layer (the second layer from the bottom), as depicted in Fig. 8.1. We note that mainly access issues are discussed, as a Semantic Web multimedia application must be able to efficiently and effectively access structured data on the Web.

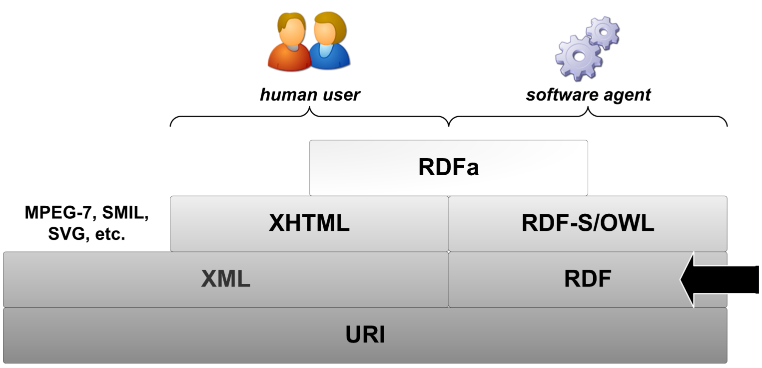



Figure 8.1: The Semantic Web Stack: Focus of chapter 8.

From a theoretical point of view, the Semantic Web is understood in terms of a stack with RDF being one of its layers. A Semantic Web application operates on the common data model expressed in RDF. Reality is a bit different, though. As legacy data has to be processed in order to realise the Semantic Web, a number of questions arise when one is after processing RDF graphs on the Semantic Web. This work addresses performance and scalability issues (PSI), viz. proposing a metric for *virtual RDF graphs on the Semantic Web*—in contrast to a local RDF repository, or distributed, but native RDF stores.

## 8.1 Motivation

The Semantic Web is—slowly—starting to take-off. As it seems this is mainly due to a certain pressure stemming from the Web 2.0 success stories. From a theoretical point of view the Semantic Web is understood in terms of a stack. The Resource Description Framework (RDF) [186] is one of the layers in this stack, representing the common data model of the Semantic Web.

However, practice teaches that this is not the case, in general. In the perception of the Semantic Web there exists a tremendous amount of legacy data: that is, HTML pages with or without microformats[1] in it, relational databases (RDBMS), various XML applications as Scalable Vector Graphics (SVG)[2], and the like. These formats are now being absorbed into the Semantic Web by approaches as Gleaning Resource Descriptions from Dialects of Languages (GRDDL) [61], RDF-in-HTML [136], etc.

Take Fig. 8.2 on page 135 as an example for a real-world setup of a Semantic Web application. There, a Semantic Web agent operates on an RDF graph with triples that actually originate from a number of sources, non-RDF or "native" RDF alike.

The Semantic Web agent operates on its local RDF graph in terms of performing for example a SPARQL Protocol And RDF Query Language [253] query to accomplish a certain task. While from the point of view of the SPARQL engine it might not be of interest where triples come from and how they found their way into the local RDF graph, the Semantic Web agent—and finally the human user who has instructed the agent to carry out a task—may well be interested in how long a certain operation takes.

But how do the triples arrive in the local RDF graph? Further, how do the following issues influence the performance and the scalability of the operations on the local RDF graph:

- The number of sources that are in use;
- The types of sources, as RDF/XML, RDF in HTML, RDBMS, etc.;
- Characteristics of the sources: a fixed number of triples vs. dynamic, as potentially in case of a SPARQL end point.

This chapter attempts to answers these questions. We first give a short overview of related and existing work, then we discuss and define virtual RDF graphs, their types, and characteristics. How to RDF-ize the flickr Web API[3], based on the recently introduced *machine tags*[4] feature, serves as a showcase for the proposed metric. Finally we conclude on the current work and sketch directions for further investigations.

[1] http://microformats.org/
[2] http://www.w3.org/Graphics/SVG/
[3] http://www.flickr.com/services/api/
[4] http://www.flickr.com/groups/api/discuss/72157594497877875/

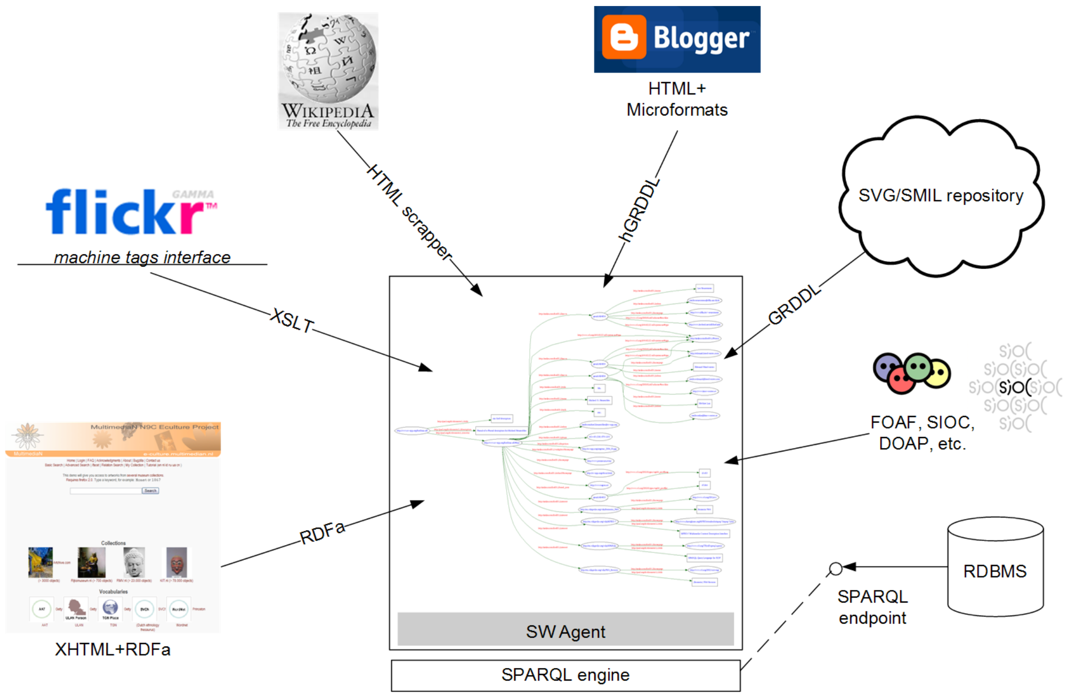



Figure 8.2: A real-world setup for a Semantic Web application.

## 8.2 Related and Existing Work

Practice-oriented research regarding RDF stores has been reported in the SIMILE[5] project [196], and in a joint W3C-EU project, Semantic Web Advanced Development for Europe [25]. In the Advanced Knowledge Technologies (AKT) project, 3store [131]—a scalable RDF store based on Redland[6]—was used. A framework for testing graph search algorithms where there is a need for storage that can execute fast graph search algorithms on big RDF data is described in [178].

At W3C, RDF scalability and performance is an issue[7], though the scope is often limited to local RDF stores. There are also academic events that address the scalability issue, such as the International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS)[8] which took place the second time in 2006.

Some research already exists regarding distributed RDF stores. Though the focus is on distributed RDF repositories, it is always assumed that one is dealing with 'native' RDF sources. An excellent example is the work of Stuckenschmidt et. al. [283]. They present an architecture for optimizing querying in distributed RDF repositories by extending an existing RDF store (Sesame). Cai and Frank [51] report on a scalable distributed RDF repository based on a peer-to-peer network—they propose RDFPeers that stores each triple at three places in a multi-attribute addressable network by applying globally known hash functions. From the Gnowsis project, a work has been reported that comes closest to ours: Sauermann and Schwarz [264] propose an adapter framework that allows for integrating data sources as PDFs, RDBMS, and even Microsoft Outlook.

To the best of our knowledge no research exists that addresses the issue of this chapter, i.e. performance and scalability issues of virtual RDF graphs on the Semantic Web. Regarding the distributed aspect, the authors of [283] listed two strong arguments, namely (i) the *freshness*, viz. in not using a local copy of a remote source frees one from the need of managing changes, and (ii) the gained *flexibility*—keeping different sources separate from each other provides a greater flexibility concerning the addition and removal of sources. We subscribe to this view and add that there are a number of real-world use cases which can only be addressed properly when taking distributed sources into account. These use cases are to be found in the news domain, stock exchange information, etc.

## 8.3 Virtual RDF Graphs

To establish a common understanding of the terms used in this work, we first give some basic definitions. Based on the definition of a Semantic Web application we have a closer look at virtual RDF graphs. The nature of virtual RDF graphs, and their intrinsic properties are discussed in detail. Definition 2.2 on page 13 requires that a Semantic Web application operates on the RDF data model, which leads us to the virtual RDF graph, defined as follows.

[5] http://simile.mit.edu

[6] http://librdf.org/

[7] http://esw.w3.org/topic/TripleStoreScalability, and http://esw.w3.org/topic/LargeTripleStores

[8] http://www.cs.vu.nl/~holger/ssws2006/

**Definition 8.1 (Virtual RDF Graph).**
A virtual RDF graph (vRDF graph) is an RDF graph local to a Semantic Web application that contains triples from potentially differing, non-local sources. The primary function of the vRDF graph is that of enabling CRUD[9] operations on top of it. The following is trivially true for a vRDF graph:

1. it comprises actual source RDF graphs (henceforth sources), with $N_{src}$ being the number of sources;
2. each source $S^i_{src}$ contributes a number of triples $T^i_{src}$ to a vRDF graph, with $0 < i \leq N_{src}$;
3. The vRDF graph contains $\sum T^i_{src}$ triples. ❖

### 8.3.1 Types Of Sources

Triples may stem from sources that utilise various representations. In Fig. 8.3 (on page 138) the representational properties of the sources are depicted, ranging from the RDF model[10] to non model-compliant sources.

The two middle layers of the pyramid denote representations that are RDF model-compliant and have a serialisation, hence may be called *native RDF*. Representations that do have a serialisation, but are not RDF model-compliant, may be referred to as *non-RDF sources*. We therefore differentiate:

**Standalone, RDF model-compliant Representations.** These type of sources, for example RDF/XML, can be stored, transmitted, and processed on their own. For example an in-memory Document Object Model (DOM) representation of an RDF/XML document can be built by utilising a SAX parser.

**Embedded, RDF model-compliant Representations.** Sources of this type, as XHTML+RDFa [136] or eRDF[11] need a host to exist. Their representation is only defined in the context of this host. Here, the triples are produced by applying a transformation.

**Representations non-compliant to the RDF model.** The majority of the data sources on the Web, standalone or embedded, is of this type:

- GRDDLable resources. GRDDL [61] is utilised to generate RDF from, for example, XHTML. The same applies to microformats which can be RDFised using hGRDDL[12];
- An RDBMS that provides for a SPARQL end point [253] can be used to contribute triples[13];
- Feed formats such as Atom [233];
- From an HTML page without explicit metadata, triples may be gathered through screen scrapers (for example [23]).

[9] **c**reate, **r**ead, **u**pdate and **d**elete—the four basic functions of persistent storage
[10] http://www.w3.org/TR/rdf-concepts/
[11] http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml
[12] http://www.w3.org/2006/07/SWD/wiki/hGRDDL_Example
[13] Though, this source may also be considered as being native in terms of the interface.

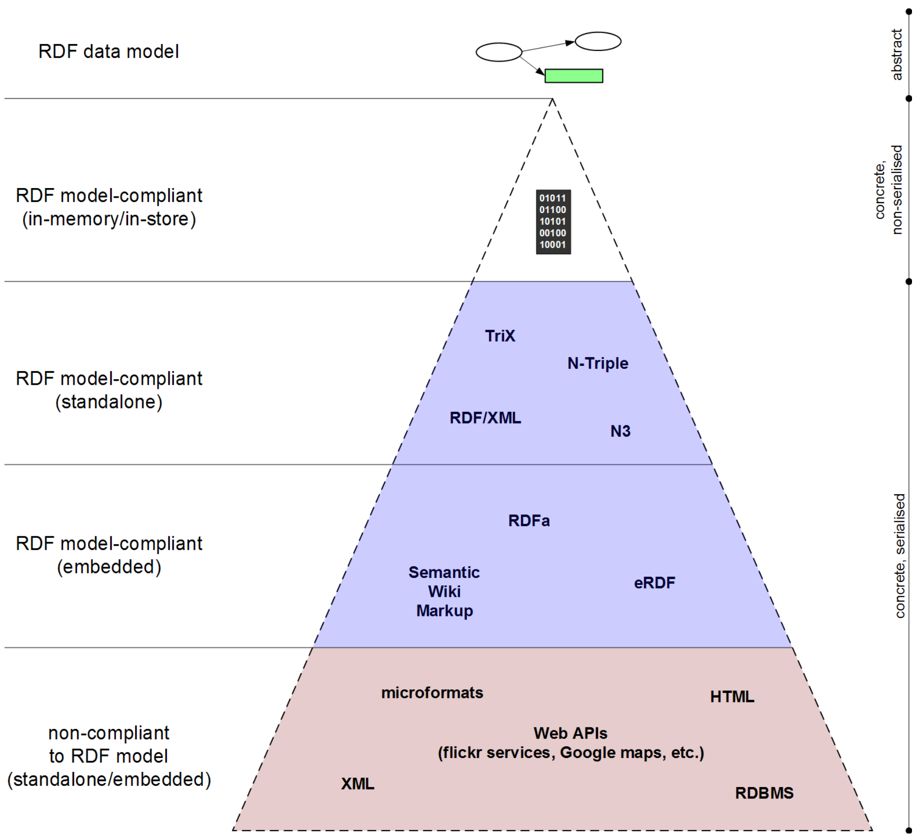



Figure 8.3: The RDF representation pyramid.

In order to be processed, the serialisation is required to be "converted" from a representation with a concrete syntax into an in-memory representation. This conversion may occur through applying a transformation[14], or by parsing the specified syntax.

### 8.3.2 Characteristics Of Sources

Besides the type of the source, a further distinction w.r.t. the number of triples from a source can be made. We distinguish between fixed sized sources and undetermined—or dynamic—sized source.

Take for example a Wiki site that serves as a source for a vRDF graph. Let us assume an HTML scraper is used to generate triples from selected Wiki pages, for example based on a category. The number of resulting triples then is in many cases stable and can be assessed in advanced. In contrast to this, imagine an RDBMS that provides for a SPARQL end point—the D2R Server[15] is a prominent example for this—as an example for a dynamic source. Based on the query, the number of triples varies. A border case are social media sites, as blogs. They are less dynamic than data provided by a SPARQL end point but constantly changing and growing as more comments come in.

## 8.4 A Metric for Virtual RDF Graphs

In this section we describe a performance and scalability metric that helps a Semantic Web application developer to assess her vRDF graph. A showcase for a non-native RDF source is then used to illustrate the application of the metric.

The execution time of an operation on a vRDF graph is influenced by a number of factors, including the number of sources in a vRDF graph $N_{src}$, the overall number of triples $\sum T^i_{src}$, and the type of the operation. The metric proposed in Definition 8.2 can be used to assess the performance and scalability of an vRDF graph.

**Definition 8.2 (Execution Metric).**
The overall execution time for performing a CRUD function (such as inserting a triple, performing a SPARQL ASK query, etc.) is denoted as $t_P$; the time for converting a non-RDF source representation into an RDF graph is referred to as $t_{2RDF}$. The total time delays due to the network (Internet) transfer are summed up as $t_D$; the time for the actual operation performed locally is denoted as $t_O$. Obviously,

$$t_P = t_O + t_{2RDF} + t_D$$

The "conversion time vs. the overall execution time"-ratio is defined as

$$coR = \frac{t_{2RDF}}{t_P}$$

❖

[14] see http://esw.w3.org/topic/ConverterToRdf, and also http://esw.w3.org/topic/CustomRdfDialects

[15] http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/

To illustrate the above introduced metric, a showcase has been set up, which is described in the following.

**A Showcase For Non-Native RDF Sources: PSIMeter** The showcase[16] demonstrates the application of the metric by RDFising the flickr API. Three different methods have been implemented; the non-native RDF Source used in the PSIMeter showcase (cf. Fig. 8.4) is the information present in the machine tags.

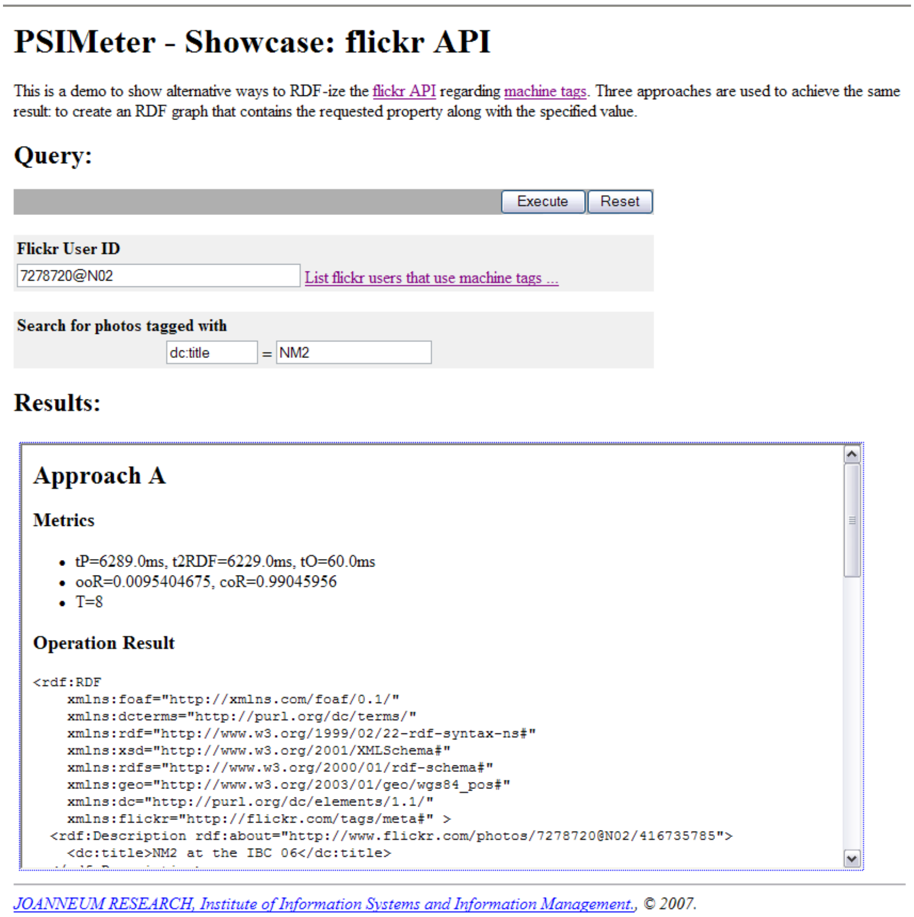



Figure 8.4: PSIMeter Showcase Screenshot.

The goal for each of the three methods is to allow a Semantic Web agent to perform a SPARQL construct statement, as for example:

```
CONSTRUCT { ?photoURI dc:subject ?subject }
WHERE     { ?photoURI dc:subject ?subject.
            FILTER regex(?subject, "XXX", "i") }
```

The experiments to compare the three approaches were run on a test-bed that comprised up to 100 photos from a single user, along with annotations in the form of machine tags, up to 60 in total. Machine tags were selected as the source due to their straightforward mapping to the RDF model.

[16] http://sw.joanneum.at:8080/psimeter/

The three methods for constructing the vRDF graph work as follows:

1. *Approach A* uses the search functionality of the flickr API[17] in a first step to retrieve the IDs of photos tagged with certain machine tags. In a second step the flickr API is used to retrieve the available metadata for each photo. Finally the result of the two previous steps is converted into an RDF representation, locally;
2. *Approach B* uses the flickr API to retrieve all public photos[18] of a certain user. It then uses a local XSL transformation to generate the RDF graph;
3. *Approach C* retrieves all public photos, as in Approach B. Then, for each photo an external service[19] is invoked to generate the RDF graph.

Firstly, for a *fixed query*, `dc:subject=marian`, the overall execution time $t_P$ has been measured depending on the number of photos (Fig. 8.6(a) on page 142). In Fig. 8.6(b), the size of the vRDF graph in relation to the number of annotations, with a fixed number of photos, is depicted.

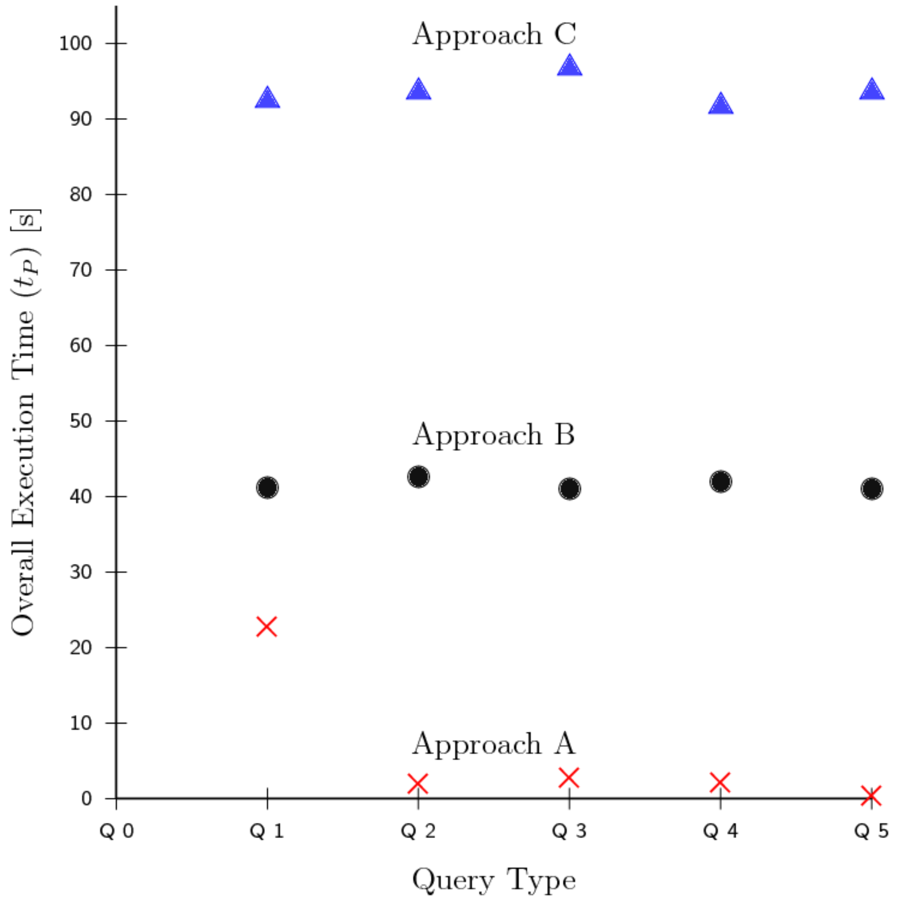
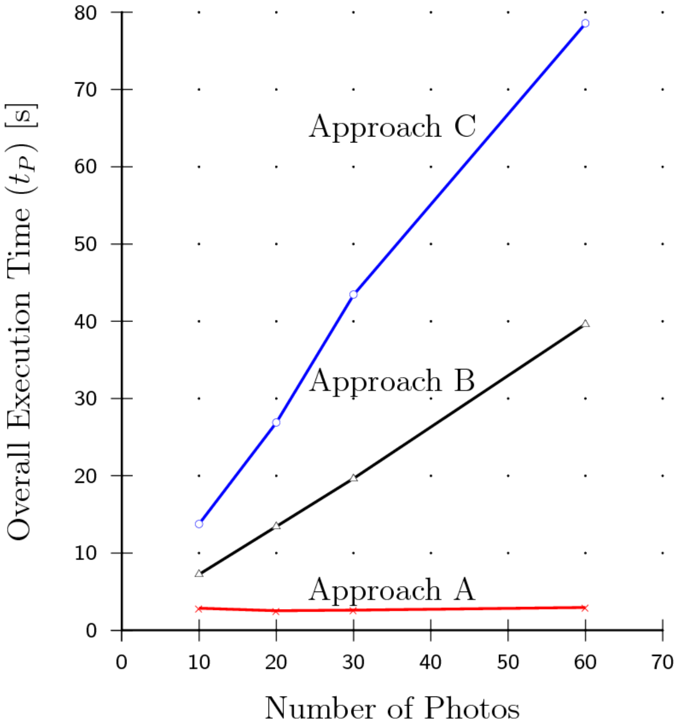
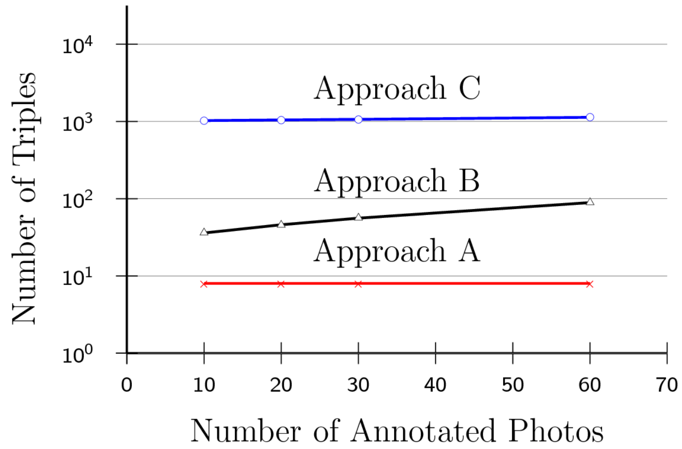



Figure 8.5: PSIMeter: Metric in dependency on Query Type.

[17] http://www.flickr.com/services/api/flickr.photos.search.htm
[18] http://www.flickr.com/services/api/flickr.people.getPublicPhotos.html
[19] http://www.kanzaki.com/works/2005/imgdsc/flickr2rdf




**(a)** Overall execution time dependency on number of photos.




**(b)** vRDF graph size dependency on number of annotated photos.

Figure 8.6: PSIMeter: Metric for a fixed query.

The second experiment focused on the impact of the *query type* on the overall execution time. The results of the second experiment are depicted in Fig. 8.5, with the according queries listed in Table 8.1 on page 143.

Note, that more than 80% of the photos were tagged with `dc:subject=dummy`, hence Q1 exhibits an exception. Another finding of the experiment, was that in all evaluation runs, $coR$ tended towards 1 (ranging from 0.95 to 0.99), i.e. most of the time the system was busy converting the data to RDF and only a small fraction was dedicated to the local execution of the SPARQL construct statement.

| Reference | Query |
|---|---|
| Q1 | `dc:subject=dummy` |
| Q2 | `dc:title=NM2` |
| Q3 | `dc:title=` |
| Q4 | `geo:location=athens` |
| Q5 | `dc:dummy=test` (empty match) |

Table 8.1: Query Types

## 8.5 Conclusion

When building Semantic Web applications, it is not only important to operate on distributed RDF graphs by means of virtue, but also to question how the triples in a vRDF graph where produced. We have looked at variables that influence the performance and scalability of a Semantic Web application, and proposed a metric for vRDF graphs. As all types of sources must be converted into an in-memory representation in order to be processed, the selection of the type of sources is crucial. The experiments highlight the importance to use existing search infrastructure, such as the flickr search API in our case, as far as possible.

Another generic hint is to avoid conversion cascades. As long as there exists a direct way to create an in-memory representation, this issue does not play a vital role. Though, regarding performance this is of importance in case an intermediate is used to create the in-memory representation, such as with, e.g., the hGRDDL[20] approach.

Finally, the incorporation of dynamic sized sources is a challenge one has to carefully implement. This is a potential area for further research in this field.

## 8.6 Acknowledgements

The research reported in this chapter has been carried out in two projects: the "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content" (K-Space) project[21], partially funded under the 6th Framework Programme of the European Commission, and the "Understanding Advertising" (UAd) project[22], funded by the Austrian FIT-IT Programme.

[20] http://lists.w3.org/Archives/Public/public-rdf-in-xhtml-tf/2006Apr/0069
[21] http://kspace.qmul.net/
[22] http://www.sembase.at/index.php/UAd

Chapter 9

# Media Semantics Mapping

*"Experimentum crucis."*

(Latin phrase)

This chapter is based on two research works, (i) a full paper at the $1^{st}$ International Conference on New Media Technology (iMedia07) [134], and (ii) an article in a special issue of the ACM Multimedia Systems Journal [135] (to appear 2008).

A Semantic Web multimedia application needs to not only handle low-level metadata, but deal with domain semantics in order to allow rich queries (likely combined with inference). Hence, we focus in this chapter on the third layer from the bottom—the ontology or vocabulary layer—as depicted in Fig. 9.1.

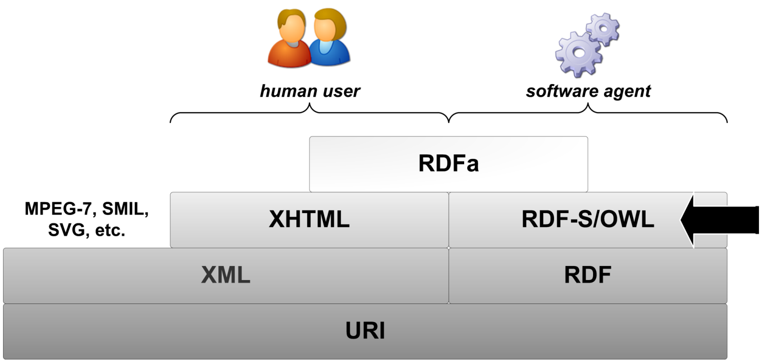



Figure 9.1: The Semantic Web Stack: Focus of chapter 9.

In the following we propose how to deal with the Semantic Gap in closed domains. That is, we propose to bridge the Semantic Gap by means of mapping well-known low-level feature patterns found in MPEG-7 descriptions to formal concepts. The key contributions of the proposed approach are (i) the utilisation of ontologies, and rules to enhance the retrieval capabilities (effectiveness), and (ii) the realisation of the feature matching process being carried out on the structural level through indexed MPEG-7 descriptions (efficiency). We discuss advantages and shortcomings of our approach, and illustrate its application in the realm of a non-linear, interactive movie production environment. Finally, we show how our approach can be understood in terms of the canonical model of media production [127].

## 9.1 Environment

The "New Media for a New Millennium" (NM2)[1] is an Integrated Project under the European Unions 6th Framework Programme in the thematic priority of Information Society Technology. The vision of NM2 is

> [...] to create a new media genre by taking into consideration the facilities of modern broadband communication and interactive terminals.

NM2 aims at developing tools for the media industry that enable the efficient production of non-linear, interactive broadband media. Additionally to the production values and aesthetic pleasures of television and cinema, productions based on NM2 technologies are influenced through the interaction of the engager according to their personal preferences. To provide non-linearity, NM2 productions are not final edited pieces of media, rather they consist of a pool of small media units to be recombined at run-time.

The NM2 project [256]—an Integrated Project of the EU 6th Framework Programme (2004—2007) with 13 partners from eight European countries—focused on the creation of technologies for non-linear, interactive narrative-based movie production.

The tools for personalised, reconfigurable media productions are elaborated in six audio-visual productions that range from news reporting and documentaries through a quality drama serial to an experimental television production. Targeted end-user devices are Windows Media Centre-PCs, game consoles, and mobile phones. For a detailed overview on the project objectives, system capabilities and the productions, the reader is referred to [326].

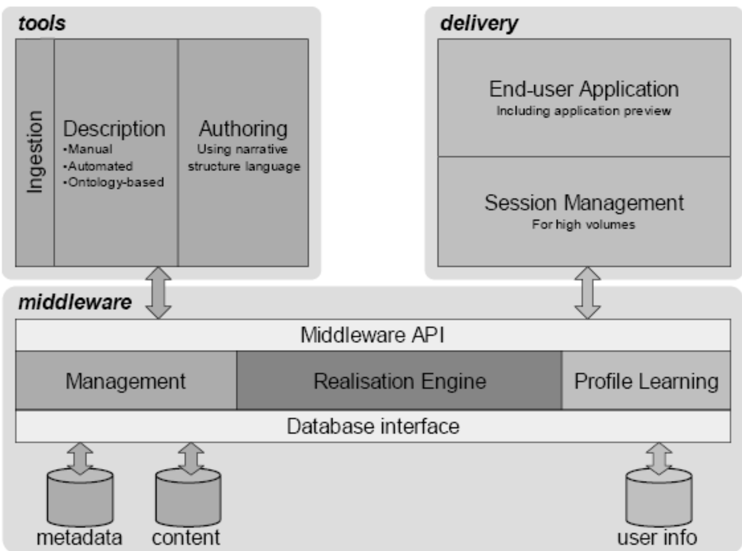



Figure 9.2: The NM2 system architecture.

The overall NM2 system architecture is depicted in Figure 9.2. The NM2 system consists of the NM2 Tools, the NM2 Delivery System and the NM2 Middleware as described in the following.

[1] http://www.ist-nm2.org/

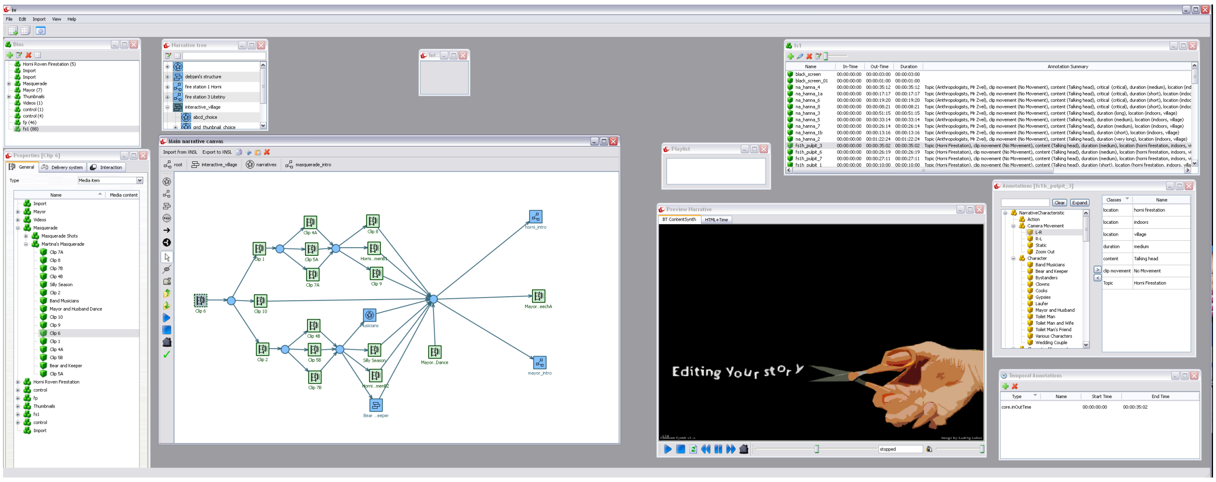



Figure 9.3: The NM2 Toolkit v3.1 (July 2007).

**NM2 Tools** The NM2 Tools (depicted in Fig. 9.3) support the authors of NM2 productions to produce non-linear story lines. The NM2 Tools comprise:

- The *Ingestion Tool* enables producers to import the essence (media files, as video-footage, audio clips, etc.) and metadata (scripts, shot-logs, EDLs, etc.) into the NM2 system.
- The *Media Semantics Mapping Utility* (pronounced: Monsoon) enables the definition of logical entities that may occur in the essence (cf. section 9.4). The output is a knowledge base that is used in the Description Tool to (semi)-automatically markup media items.
- The *Description Tool* allows creating, modifying, and deleting media items. Both low-level features as well as ontology-based metadata can be expressed in this tool, including automatic content analysis (performed on MPEG-7/low-level features).
- The *Authoring Tool* is used by the producers to create interactive narratives. Narrative objects can be placed and interconnected onto a workspace, media items from the Description Tool can be added to the workspace, and specific rules and heuristics can be entered to lay out the logic of the respective narrative.
- *Simulation and Test* provides functions which enable an interactive narrative to be tested and validated from within the NM2 Tools application environment.

**NM2 Delivery System** Engagers interface with the Delivery System that presents the output to them and manages interaction. The delivery system is set up in a client-server model that is already supported in many popular domestic devices such as PCs, advanced set-top boxes and games consoles.

**NM2 Middleware** The Middleware mediates between the Production Tools and the Delivery System by managing and interpreting the metadata and content. The Middleware handles all data management tasks, and the automatic assembly of media essence. It includes the Realisation Engine, which is responsible for dynamically creating a user-specific story, based on a given story world and the interaction of a particular engager.

Within NM2, the central units when talking about multimedia content description are media items. A media item refers to some multimedia essence[2], and is to provide a machine processable description of the essence to make its semantics explicit. Such descriptions form the basis for the retrieval tasks and have to enable semantic queries in narrative objects. The media items semantics is made explicit in two ways: first, by generating and attaching an MPEG-7 description of the essence. Second, through the annotation with logical entities defined in an ontology that represents a specific domain. The major task lies in bridging the Semantic Gap, that is to map media intrinsic information (captured within MPEG-7 descriptions) to logical entities.

One key problem of multimedia content understanding—bridging the **Semantic Gap**—still is not satisfactory solved. In the realm of non-linear, interactive movie productions, one major challenge is the dynamic matching of appropriate clips based on a formal expression describing the desired content. In our setup, a movie is—based on the users interaction—assembled on-the-fly, requiring the retrieval of the audio-visual content to be performed in near-real-time. For example, in one point of the narration, there could be a query for some material that *is about soccer*, *has an interview* in it and *starts with a PAN LEFT* camera motion. The challenge to find matching media assets in the realm of the NM2 project is depicted in Fig. 9.4 on page 149.

## 9.2 Related Work

Thoughts on how to conceptually bridge the Semantic Gap are probably as old as the multimedia content itself [115]. Most of the work in the realm of multimedia content representation focuses on the integration of multimedia metadata—as MPEG-7 [222]—with logic-based ontology languages, typically OWL [241].

The constitutive work of Hunter et.al. [171] has led to numerous related efforts [290; 107] that all share the translational approach of mapping MPEG-7 to OWL. For the field of ontology-based video retrieval, for example [298] reports a methodology to support interoperability of OWL with MPEG-7.

Media Streams—developed by Davis [66] in his PhD thesis— is a system for annotating, retrieving, repurposing, and automatically assembling digital video. It uses a stream-based, semantic representation of video content with an iconic visual language interface of hierarchically structured, composable, and searchable primitives. Nack and Putz presented the Authoring System for Syntactic, Semantic and Semiotic Modelling (A4SM) framework [228] that includes the creation, and retrieval of media material. The project goal was to have a framework at hand that would allow for semi-automated annotation of audiovisual objects, and to demonstrate the applicability in a news production environment. Both the Media Streams system and the A4SM can be understood as precursor to our proposed architecture.

Further related work can be found in [11]. In this paper, a knowledge-assisted analysis (KAA) platform is described. The interaction between the analysis algorithms and the knowledge is continuous and tightly integrated, instead of being just a pre- or post-processing step in the overall architecture. A matching process queries the knowledge base

[2]which could be a video clip, or an audio clip, or even a still-image

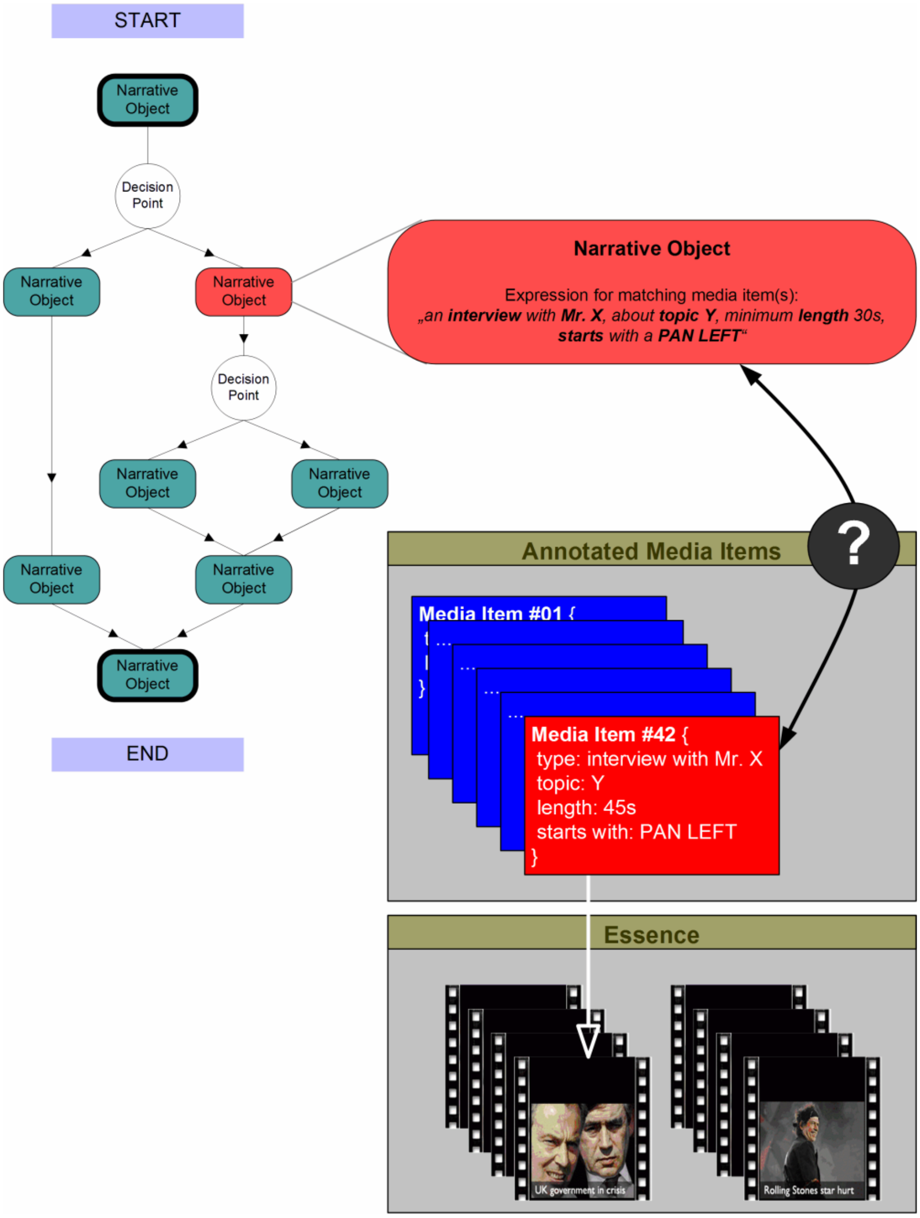



Figure 9.4: Challenge of finding matching media assets in NM2.

and assigns each region with a list of possible concepts along with a degree of relevance.

Other attempts closely related to ours are [133] that use rules to construct semantic descriptions for LOMs and [205] which proposes a similar approach regarding abstraction levels. [230] reports an approach to express behavior with rules in combination with ontologies and MPEG-7.

Audio content analysis and description has been a comparably active research field in the last 20 to 30 years. MPEG-7 Audio, in conjunction with the Multimedia Description Schemes part of the standard, provides structures for describing audio content. These structures are a set of low-level descriptors, for audio features across many applications (spectral, parametric, and temporal features of a signal) as well as high-level description tools that are more specific to specific application domains. Those high-level tools include general sound recognition and indexing, instrumental timbre, spoken content, audio signature, and melodic tools to facilitate query-by-humming. Audio analysis and retrieval applications are usually based on Independent Component Analysis of spectral components (ICA) and classifiers based on continuous Hidden Markov Models. Related work regarding the task of classification in the audio domain can be found in [21; 22; 185].

Motivated by the promising work reported in [199] and [157] the proposal presented inhere is based on our experiences with MPEG-7 annotation and retrieval [19].

## 9.3 Media Semantics Mapping

In this section, we discuss the *Media Semantics Mapping* (MSM) foundations, the terminology used, and give an oiverview on the utilised metadata. We further introduce the MSM more formally, and show how to apply it in the realm of the NM2 project.

### 9.3.1 Data and Metadata

The Narrative Structure Language (NSL) [305], developed within NM2, is a language for expressing non-linear narratives. We distinguish between specific narratives and global narratives. A specific narrative is a set of representations of media items arranged into a playlist that is delivered to an NM2 engager. A global narrative contains the same references to media items, but instead of fixed sequences, it specifies rules that are used to create a specific narrative on-the-fly based on context information. A specific narrative can insofar be regarded as an instantiation of a global narrative. The software that interprets a global narrative, producing a specific narrative, is referred to as the Realisation Engine.

OWL-DL [241] is used to formalise global production characteristics, and the domain of each specific production. Furthermore OWL-DL is used to add contextual information and to interface to the NSL.

*Modality* in our understanding is a path of communication between the human and the computer. Major modalities are vision and audition (others are tactition, olfaction, etc.).

In the following, audio-visual data is referred to as *essence*. Essence is the actual piece of data that resides e.g. in the file system. The essence can be document-based or stream-based and can be "played", using one or more modalities. The effect of "playing" depends on the modality being used. Example: an MP3 file using the audio modality or a GIF file using the visual modality.

With a *Media Item* (MI) we mean a proxy for the essence that has a pointer to the essence, enables the attachment of metadata (MPEG-7 description), and finally serves as the pivot for semantic annotations.

In NM2, MPEG-7 [222] is utilised for capturing intrinsic low-level features of the essence (colour descriptors, etc.), extracting as much as possible automatically from the essence to produce sound MPEG-7 descriptions using technologies from our Multimedia Mining Toolbox [19, Section 6]. In the visual domain we use the Dominant Color Descriptor and the Color Layout Descriptor to capture colour features. To describe textures, we make use of the Edge Histogram Descriptor. Shapes can be recognized via the Contour-Based Shape Descriptor. The Camera Motion Descriptor is utilised to describe camera movements (pan, tilt, zoom, etc.). Although a representation of low-level features on the ontological level would be possible, we do not lift MPEG-7 descriptions and description schemes onto the logical level, rather MPEG-7 fragments are referenced from within the ontology.

We want to stress the fact that due to the wide spectrum of domains that the six productions are about, a flexible and generic approach is needed to handle data and metadata. This is in contrast to e.g. [199], where only a single domain is formalised.

### 9.3.2 Media Semantics

*What are media semantics*? According to [130], any language definition comprises syntax, semantic domain, and a semantic mapping from the syntactic elements to the semantic domain. When talking about media semantics inhere, we subscribe to this point of view. In our understanding the essence itself does not "have" semantics. A piece of essence may be consumed or manipulated, nevertheless, essence "carries" the semantics and it is up to the consumer of the essence to interpret what she understands from it. Hence we do not try to define what in the general case an object "looks like" or "sounds like". We therefore understand that the ontological constructs in combination with the rules are our syntactical framework, further the semantic domain is conceived as being the domain of the LE that can occur in the essence, and finally define the semantic mapping as described below.

### 9.3.3 Spaces of Abstraction

We allow for two orthogonal conceptual paradigms to model media semantics: spaces and the well-known class/instance pattern. A *space* represents a certain level of abstraction, ranging from low-level, as colour or shape to abstract entities such as human feelings. Classes and instances are used to define the actual LE. Therefore "the soccer ball" instance in the context of a soccer game is defined to be black, white, and round but this does not mean that "a ball" in general—referring to the class level—has these properties.

Fig. 9.5[3] depicts the spaces available in our approach, with V denoting the modality vision, and A the modality audition. Following [115] we introduce relationships that can hold between an essence and some logical entities. The spaces are defined—listed by increasing level of abstraction—as follows.

1. The Feature Space—**F-Space**. Contains LE that represent low-level features. A low-level feature (LLF) is a single aspect of a certain (spatio-temporal) part of a media item.

[3] Note that this figure has already been presented at the ESWC 2006 tutorial "What you Mean is What you Watch: Multimedia and the Semantic Web", cf. `http://gate.ac.uk/conferences/eswc2006/multimedia-tutorial/`

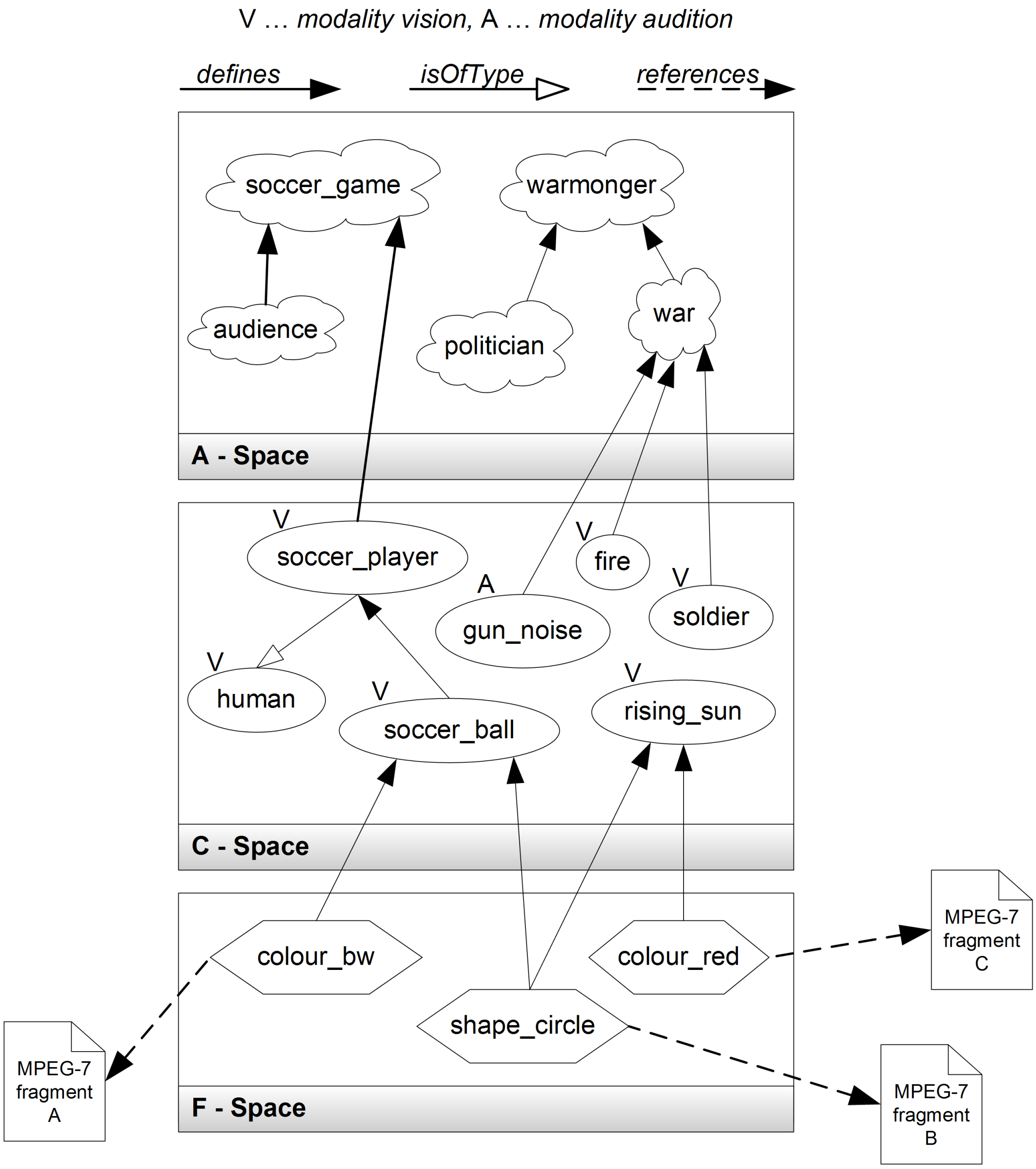



Figure 9.5: The Media Semantics Modelling Spaces.

For example the dominant colour of a spatial region (black and white) is represented as a LLF.

2. The Concrete Logical Entity Space—**C-Space**. Contains LE that can directly be recognized in the essence. A concrete logical entity (CLE) is a distinct object being defined by a combination of low-level features and their respective values (simple CLE) or using other concrete logical entities (composite CLE). For example, in the soccer domain, the CLE soccer ball may be defined by the LLF dominant colour black and white, and circular shape, i.e., a simple CLE. A table could be defined consisting of a CLE tabletop and four CLE table legs, resulting in a composite CLE.

3. The Abstract Logical Entity Space—**A-Space**. Contains LE that are not directly observable. An abstract logical entity (ALE) can be defined by a combination of CLE (simple ALE) or other ALEs (composite ALE). For example the ALE soccer game may be defined by the simultaneous presence of an ALE audience and a number of CLE soccer player.

Within each space, the class-instance modelling can be used to add further semantics, as taxonomies, object relations, etc. In each of the six NM2 productions, a domain-specific ontology is defined covering concepts and instances. This might be 'church', or 'painting' in the case of the documentary production about England's Golden Age in the $16^{th}$ century, or certain actors, moods, and keywords, as found in the drama Accidental Lovers [147].

Details about the NM2 core ontology can be found in Appendix A.3.

### 9.3.4 Built-in rules

Due to the well-known limitations of DL-ontology languages [163] we utilise rules in addition to DL-ontologies (see also [278]) to define the semantics of a logical entity in the context of a production. However, using rules can lead to serious problems w.r.t. organisational issues. We therefore only provide a minimalistic set of so called built-in rules, and automatically generate the actual rules as described below.

An overview of the available built-in rules is given in Table 9.1.

| BIR | Informal Semantics |
|---|---|
| `defines`, `contains` | If a LE is defined via a set of LLF, and these occur in a MI, the LE also occurs in that MI. |
| `transitions` | If a MI contains a certain LLF at the beginning/end, the MI starts/ends with it. |
| `modalities` | If a LE has a certain modality, the hosting MI has this modality as well. |
| `temporal` | If a LE occurs in a (temporal) part of a MI, the MI as a whole contains it as well (part-whole inference) |

Table 9.1: Overview on the Media Semantics Mapping built-in rules.

Two properties, defined in the NM2 core ontology (cf. also Appendix A.3.1) , enable the incorporation of rules, hence assisting to define the semantics of a logical entity.

The property `defines` allows a combination of `ConcreteLogicalEntity` instances to define either another `ConcreteLogicalEntity` instance (which forms a composite pattern) or an `AbstractLogicalEntity` instance, hence an inter-space mapping. For each (partial) `defines`-property in the ABox of the ontology appropriate atoms are added to the corresponding rule.

A media item `contains` a number of `LogicalEntity` instances along with `LLFeature` instances representing an occurrence of a logical entity in a media item. Equally as above, for each occurrence atoms are added accordingly.

An exemplary built-in rule defining the mapping from the F-Space to the C-Space is shown below, in listing 9.1.

```
1 contains(mi, cle) ← defines(llf_1, cle) ∧ ... ∧ defines(llf_i, cle)∧
2                     contains(mi, llf_1) ∧ ... ∧ contains(mi, llf_i)
```

Listing 9.1: The F/C-Space mapping rule.

Given that a set of low-level features $\{llf_1 \dots llf_i\}$ `defines` a certain logical entity $cle$ (line 1), and it is known that a certain media item $mi$ `contains` this set of low-level features (line 2), it can be inferred that $mi$ `contains` $cle$.

However, to ensure the correctness of the definition, some constraints must be put on the variables: $\forall llf_i \in LLFeature$, further $cle \in ConcreteLogicalEntity$, and $mi \in MediaItem$, which highlights the connection to the NM2 core ontology that defines each of the concepts.

**Transitions** Although the assembly of the essence is handled on the level of the NSL, the need for providing some information on how a media item starts or ends is obvious. In the visual domain this kind of information is captured for all camera motions. In the audio domain this is done in the general case because the cut in the audio modality almost always is critical w.r.t. sound intensity and phase. Similar to [246] a set of temporal semantics regarding audio-visual content is defined. Currently our temporal granularity is that of key-frames.

The information how a media item ends (respectively starts) can be made explicit (Fig. 9.6).

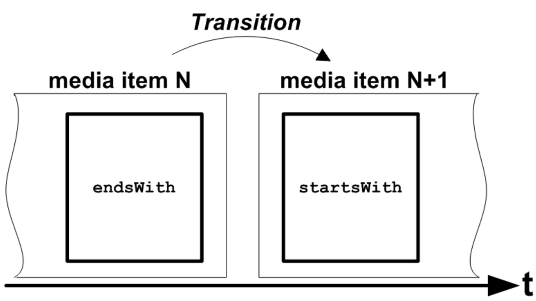



Figure 9.6: Transitions in A/V-essence

This information can then be used to support the creation of a specific narrative embodying the directors style, e.g. “never let a pan left directly follow a pan right”.

The explicit transition information can immediately be checked from the NSL by utilising `endsWith(X, Y) - startsWith(X, Y)` properties on consecutive media items.

Utilising temporal semantics is demonstrated in the following: for each low-level feature $\Phi$) that is of type camera motion (pan, tilt, zoom, etc.) or any supported audio feature, the meta-rule from listing 9.2 is applied.

$$endsWith(m, \Phi) \leftarrow containsAtEnd(mi, l) \wedge \Phi(l)$$

Listing 9.2: Examplary transition rule.

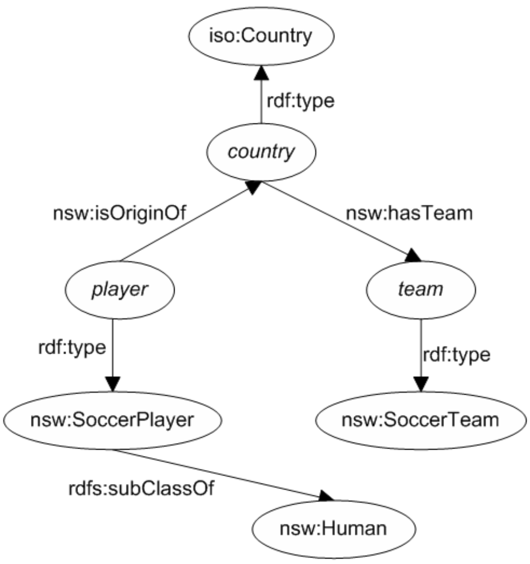
### 9.3.5 User-defined rules

To enhance the domain-specific ontologies further, so called user-defined rules can be manually defined. In listing 9.3 an example of an user-defined rule[4] is depicted. Applying this rule on the production ontology (Fig. 9.7(a)) yields the result depicted in Fig. 9.7(b).

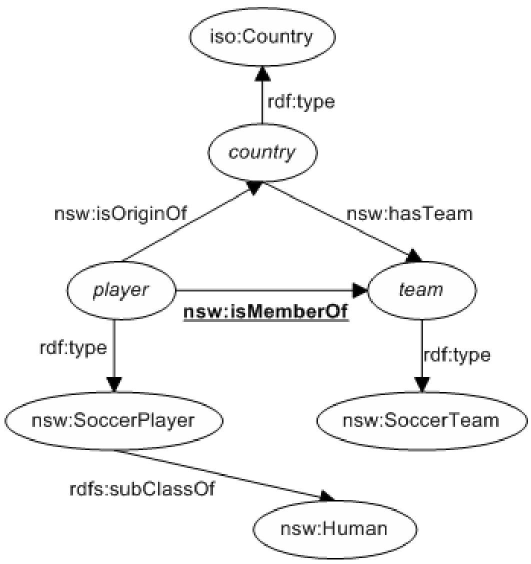
$$isMemberOf(p, t) \leftarrow isOriginOf(p, c) \wedge hasTeam(c, t) \wedge$$
$$SoccerPlayer(p) \wedge SoccerTeam(t) \wedge Country(c)\,.$$

Listing 9.3: An exemplary user-defined rule.

In Fig. 9.7(b) the inferred property `nsw:isMemberOf` is depicted. This also shows the usage of a subclass (`nsw:SoccerPlayer`) and the incorporation of an existing, external ontology (`iso:Country`).




**(a)** Production ontology base.




**(b)** Production ontology after applying user defined rule.

Figure 9.7: Applying a user-defined rule on a production ontology.

[4]Taken from the NM2 MyNewsSportsMyWay production.

### 9.3.6 The MSM Knowledge Base

The ontology and the rules together form the knowledge base $\mathcal{KB}_{MSM}$, defined as follows:

**Definition 9.1 (The MSM Knowledge Base ($\mathcal{KB}_{MSM}$)).**
The knowledge base $\mathcal{KB}_{MSM}$ is a tuple $(\mathcal{O}_D, \mathcal{R})$.

- with $\mathcal{O}_D$ being an ontology that consists of an A-Box and a T-Box, and
- with $\mathcal{R}$ being a rule-base including built-in rules and user defined rules. ❖

$\mathcal{KB}_{MSM}$ is further used to annotate the essence automatically. The rule-base $\mathcal{R}$ is represented using the Semantic Web Rule Language (SWRL) [164] to ensure a homogeneous format w.r.t. the ontology $\mathcal{O}_D$. For the purpose of applying $\mathcal{R}$ onto $\mathcal{O}_D$, the SWRL representation encoded in RDF/XML is converted into a number of Prolog rules. The T-Box and the A-Box of $\mathcal{O}_D$ represented in OWL-DL and encoded in RDF/XML are converted into a number of facts in Prolog. The outcome of the inference process is reflected in an update of the A-Box of $\mathcal{O}_D$.

## 9.4 Applying the Media Semantics Mapping

The **M**edia **S**emantics **M**apping Utility (MSM-Utility) is used to define instances based on the built-in rules, described above to generate $\mathcal{KB}_{MSM}$. For managing MPEG-7 documents we use our MPEG-7 Document Server [19, Section 5.2], which provides access to MPEG-7 documents for a number of clients and allows the exchange of whole documents or fragments thereof utilising XPath. Access to parts of documents is crucial for the efficiency of the system, as MPEG-7 documents of larger media items tend to have considerable size. The MPEG-7 documents used in the system are compliant with the Detailed Audiovisual Profile (DAVP) [19].

For processing the ontological information, we use a high-performing RDF-library, the Redland RDF library[5], wrapped up in an Object-Oriented-API (C++) that enables manipulation and query on the ontological level. Applying $\mathcal{R}$ onto $\mathcal{O}_D$ is done utilising Prolog. $\mathcal{O}_D$ represented in OWL-DL is converted into a number of SWI-Prolog[6] facts.

Typically, users of the NM2 toolkit lay out their production-specific ontologies by means of creating concepts and instances. Through $\mathcal{KB}_{MSM}$ the system is then able to automatically tag the essence in two subsequent steps. Firstly, the low-level features are extracted automatically on the MPEG-7 level. Secondly, $\mathcal{KB}_{MSM}$ is used to match against the generated description of the essence, triggering an update of the ABox of $\mathcal{O}_D$.

There are three distinct cases in which the incorporation of somehow existing work is desired. Although they are quite different in their actual intention, the can be enabled through a single functionality (import):

- Migration. Internal changes to the core ontology can largely be handled through the import of the previous version of the core ontology.

[5] http://librdf.org/
[6] http://www.swi-prolog.org/

- External Ontologies. In many productions the need for incorporating existing ontologies (ISO/Countries, Time, DC, etc.). This is best supported by displaying a list of importable classes and properties to the user and let him decide whereto import to (e.g. as a concept, etc.).

- Default Ontology. As defined in [62, Section 5.2], there exists a number of classes that apply to all productions, i.e. generic narration terms as Topic, Subject, Action, Location, Time, Prop, and Genre. If a production wishes to use this predefined classes it can import it in a single step (fixed mapping to core classes).

## 9.5 Mapping the NM2 Workflow to the Canonical Model

In the following we summarise, based on [135] our findings with non-linear, interactive movie productions. We carry out this discussion based on experiences gathered during the development and evaluation of an authoring suite that enables the creation and testing of non-linear narratives. The paper describes our approach and shows how this dynamic workflow can be understood in terms of the canonical process of media production [127].

### 9.5.1 The NM2 Workflow

In Fig. 9.8 on page 158 the overall NM2 workflow is depicted. In the upper half of the production workflow is shown. People with different roles (such as editors annotating material or authors creating a story) work in a collaborative fashion with the NM2 Tools in order to create and manage the repositories. While the generic story itself is constructed on the logical level using a graph-based tool, it references the media items either directly or via a logical expression.

In the lower half of Fig. 9.8 the actual run-time setup of the NM2 system is illustrated. The repositories produced during earlier production stages are used by the NM2 Delivery System to create an actual output based on the logical description of the story world and the interaction of the end-users through the end-user devices.

For a detailed overview of the NM2 project objectives, system capabilities and the productions, the reader is referred to [326; 256].

### 9.5.2 Authoring Of Non-linear Stories

Following Bulterman and Hardman [47], the NM2 Tools support the author in creating narratives with a combination of the graph-, structure- and script-based paradigms. We take a declarative approach to modelling interactive programmes, where the narratives are represented in a formal language called the Narrative Structure Language (NSL) [304; 303], made of building blocks called Narrative Objects (NO). The smallest building block that can be used to specify a narrative is the media item, pointing to a media asset, such as a video clip. Note that it is assumed that standard non-linear editing tools, such as Avid or Final Cut pro are used to edit media assets.

An Atomic Narrative Object refers to a media item (i) by directly referencing it, or (ii) using an expression. For example, a Narrative Object could query for some material that "is about soccer", "has an interview" in it and "starts with a pan left" camera motion. To realise the match, a set of existing technologies is utilised. MPEG-7 (cf. section 3.3.4) is utilised

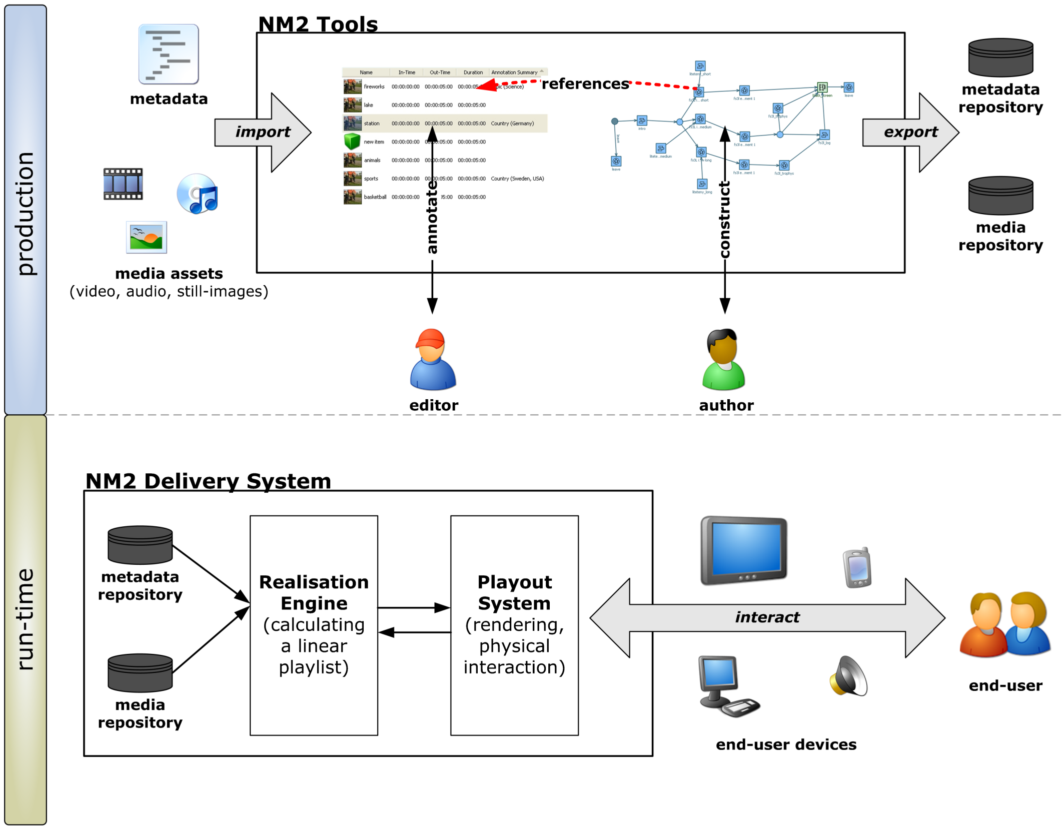



Figure 9.8: The NM2 workflow.

for representing low-level features of the media items, such as colour descriptors. OWL-DL (see section 4.3.4) is used to formalise production specific entities.

A non-atomic Narrative Object (NO) is called a Structured Narrative Object and is a grouping of other NOs. The NSL consists of a set of recursive primitive narrative structures through which higher level narrative structures, such as "3-Acts", are expressed [322]:

- A **link structure** is a directed graph, possibly with cycles. Each node is a NO, each edge specifies a potential path a narrative could take from the origin to the target node. Each edge has an enabling (boolean) condition referring to the metadata of narrative objects, input from the engager and context information (such as the play-list compiled so far).

- A **layer structure** has a number of layers, each consisting of a narrative object. Reaching a layer structure leads to the media asset referred to by each layer being added in parallel to the play-list. This means that they are played concurrently, starting at the same time. Layer structures can be used, for example, to associate a background-audio with video.

- A **selection group** has content, selection criteria and constraints. The content is a collection of narrative objects. The content may be specified by directly referencing narrative objects, or using an expression, as in the case of an atomic NO. Each selection groups has an optional termination condition which may be used to loop its interpretation.

### 9.5.3 Example NM2 Productions

The NM2 Tools have been used and evaluated in seven media productions[7]. These media productions were chosen to reflect a range of content genres (including drama, fiction, news, and a documentary). Three example productions are discussed in the following.

*Accidental Lovers* (AL) [147] is a participatory romantic black comedy, for television, mobile phone and Internet. The end-user can affect in real-time the unfolding drama of the unlikely romantic couple, Juulia in her sixties and Roope in his thirties. The outcome of the drama is shaped by sending text messages to a system that triggers story events based on keyword recognition. The interaction grammar is based on keyword-recognition: each submitted SMS is scanned for keywords selecting a single clip for either one of the main characters.

*My News & Sports My Way* (MNSMW), a digital interactive archive that allows end-users to discover, select and recombine news and sports items into stories which meet their individual tastes. The MNSMW production, produced in collaboration with the Swedish public service television (SVT), aims to show how the production tools of the NM2 system may improve the possibilities for producers of news and sports to create a reconfigurable media output that can be offered to end-users in different situations and in different formats. The idea is that as a finished product, the output will reach the end-users through broadband and will be consumed with the help of Windows Media Center.

*Gormenghast Explore* (GX) [201] is an experimental, spatially-organised drama for interactive delivery over broadband, made from the original footage of BBC TV's 4-part adaptation of Mervyn Peake's "Gormenghast" novels. The visitor enters the virtual Octagonal Gallery, and approaches any of the eight "living portraits" around it, guided by the direction of the lighting as well as by sound. The portraits are video loops, each offering a taster of the character whose story can be found behind the portrait panel. In GX, viewers can choose to follow the story of a particular character, or explore the rooms of the castle to find what happens there.

In section 9.5.5 the three above presented productions will be used to exemplify parts of concrete NM2 workflows.

### 9.5.4 Lessons Learned

In NM2 the productions are made with audio-visual material, such as video clips, audio recordings and graphics. In the case of fiction or scripted factual stories, the pool of media items reflects the pool of script items (such as scenes in the action treatment or voice over commentary). The media items are automatically assembled at viewing-time according to the rules determined by creators into stories reflecting both the narrative patterns devised by the creators, and the choices and preferences of end-users.

Due to production-specific requirements, the NM2 tools have to support a range of import functions (metadata, media), while providing a generic interface for the narrative design. Depending on the genre (fiction, news) it is possible to identify production assets that influence the workflow:

- Script items are available, hence can be used to assist in the post-production phase in certain genres, such as fiction or drama.

[7] For an overview the reader is referred to an online video available at http://www.ist-nm2.org/media_productions/CU_intro_video/CU-intro.htm

- The audio-visual material itself can be available in advance, or just-in-time, as is the case with news.

- Multimedia metadata [141] for the audio-visual material can be extracted and/or annotated manually.

One of the findings in the NM2 project is that handling the above mentioned issues—independently of the type of production—is hard to realise. This is due to the actual production requirements: For example, in the news domain, certain events (such as the eruption of a volcano) demand for an immediate change of a programme; the annotations in this case are only available after a while (e.g. for archiving purposes), hence may not be available for shaping the narrative.

Another challenge is the need for iteration based testing of the narrative. Due to the potential numbers of ways a story might evolve, a preview of parts of the story has to be offered. As the end-user devices along with the physical interaction can take many forms, a generic interface is required to render the preview of the narratives. This interface simulates the logically choices an end-user is able to make, hence drives the actual construction of the specific story. This functionality is offered by the NM2 system, called the Narrative Preview (Fig. 9.9). In using the Narrative Preview, a non-linear media editor is able to verify how, and if the story works, both on the logical level and from the artistic point of view. Through the Interaction Window, the editor is able to test all branches; a challenging task for complex and deep-structured narratives.

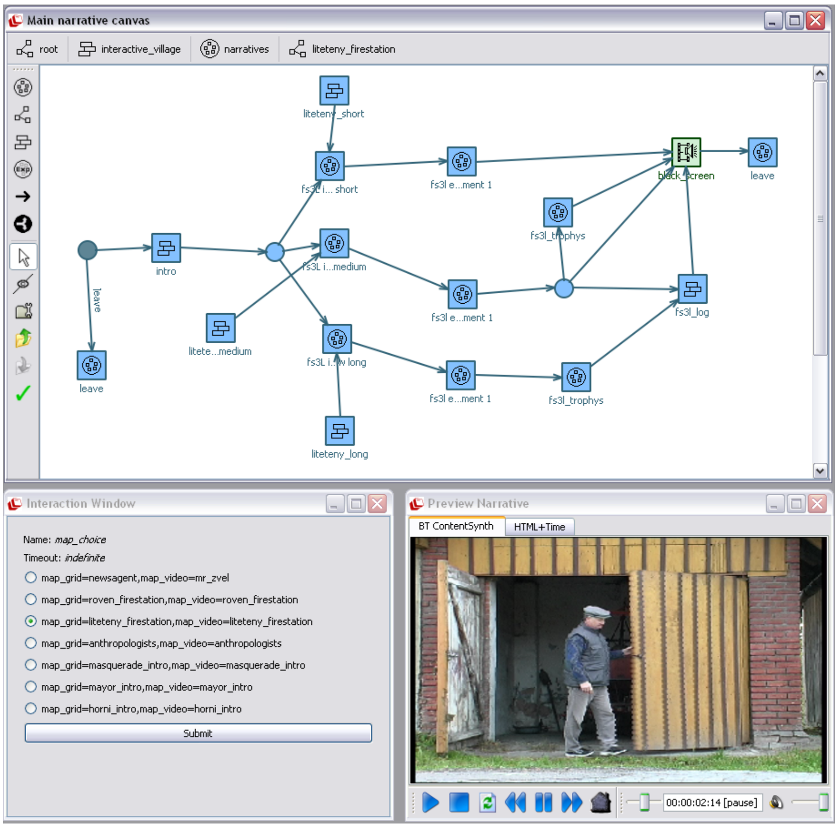



Figure 9.9: Preview a narrative of a non-linear media production.

Although validating a story on the syntactic level is possible and available, it turns out that "debugging" a story on the semantic level is a laborious and time-consuming task. Whilst for a human being it is rather straightforward to tell if a story is consistent and compelling, for a computer it is not. To this end we found that enforcing artistic aspects, a human in the loop is still required.

### 9.5.5 The NM2 Workflow in Terms of Canonical Processes

In this section we describe the NM2 workflow (depicted in Fig. 9.8 on page 158) and suggest a mapping to the Canonical Processes [127]. Note that no single NM2 production precisely instantiates this workflow. The actual workflow depends on, e.g. the genre of the production, or the availability of existing production assets (such as scripts). In the following, the three NM2 productions introduced in section 9.5.3 are used to highlight concrete NM2 workflows.

#### Preproduction

In NM2 the embodiment of the canonical *Premeditate* process heavily depends on the genre of the actual production. Ranging from a news production that directly incorporates existing metadata structures over a drama that utilises existing scripts, each production has its own requirements and ways to express `what` and `why` to capture. For example, in the My News & Sports My Way production, local events may be more interesting to the end-user, compared to global, unspecific ones. On the other hand, the story-world of Accidental Lovers was carefully designed long before the actual shooting.

Typically in the NM2 pre-production phase a production ontology is defined. It extends the NM2 core ontology in order to represent production specific entities, such as objects, actors, places. In the My News & Sports My Way, for example, the production ontology was initially derived from a taxonomy provided by the SVT; in a second step the production ontology was extended with other concepts, such as location and people. On the other hand, in Gormenghast Explore, the ontology was created from scratch, based on Aristotle's categories, such as time, place and action.

Another aspect influencing this process are decisions regarding the logical interactions, i.e. the way the end-user can influence the shaping of the material. The author has to decide where branches are introduced and how and when the branches join again. The NM2 Tools can be used in this phase to experiment with narrative structures. This is occasionally valuable even without the presence of actual media assets to, e.g. test if and how a narrative works. Again, take for example the Accidental Lovers narrative. The global story with its basically four options (Juulia falls in love with Roope, Roope falls in love with Juulia, both fall in love with each other, or both do not) as well as the micro-stories within each branch could be examined—prior to the (expensive) production.

#### Production

In NM2 we assume standard-NLE tools to be used to actually edit clips. This is due to the focus of the NM2 project rather than due to technical limitations. Although media assets have actually been created in the realm of NM2 productions (for example in Accidental Lovers), it happened outside of the scope of the NM2 Tools. We therefore understand that the *Create Media Asset* process of the canonical model is not present in the NM2 workflow.

The NM2 Tools allow for a number of ways to import media assets and metadata. After manual import, or a bulk import into the NM2 Tools, the single piece of essence is a "media item". The author is supported during the import through filters (such as media type or date) allowing to select certain media assets (*Query* canonical process). Data structures called "bins" are used to logically group media items. In case of a manual import (e.g. in Accidental Lovers), the editor selects a clip (and may choose to trim it). The editor can import the media item into an existing bin, or a create a new bin to hold he imported media asset; this task can be understood as the *Package* canonical process.

### Description

The canonical *Annotate* process is present in manifold ways in the NM2 Tools' workflow. Most commonly used are semantic annotations, based on the production ontology.

The editor is supported in describing media items in two modes: On the one hand automatic content analysis yields low-level features that are represented in MPEG-7. For example shot boundaries have been extracted in several productions, allowing the definition of smooth transitions. An author can, e.g. state that a pan-left shot must not follow a pan-right shot; this can be achieved based on the automatically extracted shot boundaries. The extracted low-level features can further be used to generate high-level descriptions rooted in the production ontology [134]. On the other hand, an editor may—again based on the production ontology—manually annotate the media items. This can be done globally, for example stating "this media item is an interview" in My News & Sports My Way, or using temporal annotations, such as "this media item starts with a pan-left followed by a two people sitting next to each other". For example, in the My News & Sports My Way production the media assets are directly exported from the SVT archive; the metadata (available as Dublin Core attributes) is mapped to instances of the production ontology. The editor then extends the description based on a taxonomy provided by SVT to capture a category (e.g. economy) a media item belongs to.

In some of the NM2 productions, logging tools, such as a shot logger, have been utilised to record information on the usefulness of the rushes. This information can be imported as well.

In order to find media items to manually annotate them, the author uses a *Query* process of the canonical model.

### Authoring

The construction of the story world can be performed bottom-up, top-down or middle-out. This phase in the NM2 workflow (upper half of Fig. 9.8) can be mapped to the *Package*, *Query*, *Construct Message*, and *Organise* canonical processes of [127]. An author might start off with a set of micro-stories and use narrative constructs of the NSL (cf. section 9.5.2) to generate bigger blocks of the story. Another authoring style that has been used in NM2 productions is to start with a rather rough, global set of narratives, and iteratively—as the material is available—refine them (My News & Sports My Way).

Within the narrative objects the NM2 system allows for two ways to select media items: by directly referencing a media item, or via an expression (*Query* and *Construct Message*), using the vocabulary from the production ontology. Expressions in narrative objects define the desired content, e.g., "I need a media item here that is maximally 45s long, starts with an interview, and followed by a close-up onto a painting".

The actual story is created dynamically depending on the editing rules, and on the usage of by reference/by expression in narrative objects. The actual shape of the story is defined only when user interaction takes place. This process of actually calculating a concrete path in the story world can be mapped to the canonical *Organise* process.

**Narrative preview**. The narrative preview is an integral part of NM2 authoring phase. However, testing a story—as discussed in section 9.5.4—is a non-trivial task. During the NM2 authoring phase one is very likely interested to see and interact with the outcome in the NM2 Tools; this is achieved via the so called narrative preview functionality. Note that authoring and narrative preview are actually interdependent and represent two sides of the same coin. The narrative preview renders the logical decision points of a NM2 narrative in a generic way, hence allowing the emulation of the production-specific front-end; it can be mapped to the canonical *Organise* and *Publish* processes.

#### Delivery

In the NM2 Delivery system on the one hand the Realisation Engine calculates dynamically a linear play-list—based on the narrative and the interaction of a particular user—and on the other hand the Playout System takes care for the actual rendering on a client device, including the physical interaction. Regarding the first part, i.e. the Realisation Engine, this is identically implemented as within the narrative preview. The actual playout is always production-dependent and represented through the NM2 Playout system. For example, Accidental Lovers was broadcasted on the TV, using mobile phones (SMS) as the back-channel; a majority-vote of all the incoming messages would influence the story. In contrast, Gormenghast Explore was realised using the Windows Media Centre along with an remote control.

Taken together, several canonical process can be identified. Driven by the end-users interaction, the global story is populated with media assets (*Query*), converted into a linear path (*Organise*), and rendered on the client device, where the physical interaction takes place (*Publish* and *Distribute*).

## 9.6 Discussion

We have shown in this chapter how to map low-level features extracted from multimedia essence to logical entities. This enables an effective and efficient retrieval of the essence. Another source for the entity definition process are scripts, shot-logs, etc., which are incorporated through the ingestion process. We also plan to include the support for *guided definitions*. This means to extract MPEG-7 features from a reference image or audio-clip, display the extracted values and let the user select a combination of the extracted values for definition purposes, quite similar to [199].

To allow for queries as "find me all MI with an interview as establishing shot, **followed** by a ZOOM_IN onto a painting", we currently work on the integration of so called *temporal annotations* to be used within a media item, based on [8].

Further, we have shown how the NM2 workflow can be understood as an instantiation of one or more of the identified canonical processes [127], or indeed a complex process combining several canonical processes. Non-linear movie productions have special needs and characteristics, such as a highly iterative workflow, strong emphasis on previewing and testing the story.

The content in a non-linear movie is assembled on-the-fly, triggered by the end-user's interaction with the content. Our finding was that mapping the canonical process is straightforward in the case of linear productions. For the domain of non-linear, interactive movie productions it might be handy to specialise the canonical process allowing to capture non-linear workflows, especially delivery and assembly of the media assets. Concluding, a formal representation of the core canonical processes would allow an automated mapping and enable translation between heterogeneous media workflows, hence fostering reusability.

# Chapter 10

# Efficient Multimedia Metadata Deployment

*"Coram publico."*

(Latin phrase)

Parts of this chapter have been presented at the First International Workshop on Cultural Heritage on the Semantic Web, collocated with the 6$^{th}$ International Semantic Web Conference (ISWC) 2007 and are available in its proceedings [140]. Further, we have published on the work reported in this chapter [16] at the First Workshop on Semantic Interoperability in the European Digital Library (SIEDL), collocated with the European Semantic Web Conference (ESWC) 2008.

We focus in this chapter on the top layer of the proposed Semantic Web stack. This layer—the deployment or integration layer–as depicted in Fig. 10.1 is important for Semantic Web multimedia applications, enabling an effective and efficient delivery of the multimedia content along with the metadata, serving both humans and machines.

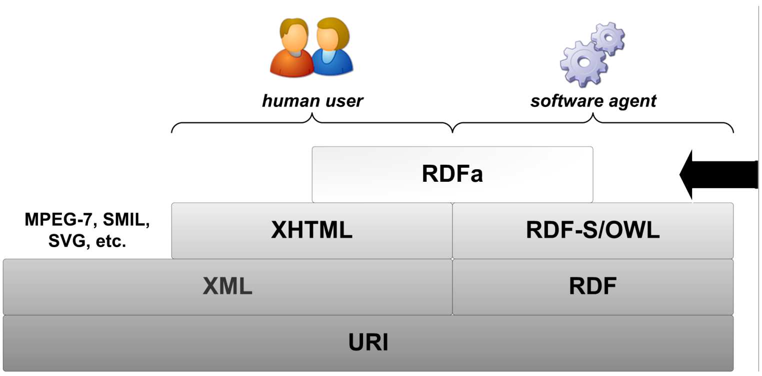



Figure 10.1: The Semantic Web Stack: Focus of chapter 10.

As a reaction to the shortcomings in deploying multimedia metadata (M3) formats on the Semantic Web, we have proposed a solution allowing to deploy multimedia metadata formats for Semantic Web agents in a scaleable and non-disruptive way. The author of this doctoral thesis has—together with other multimedia metadata researchers—launched the RDFa-deployed Multimedia Metadata (ramm.x)[1] initiative [138] to address this issue.

Further, this work has been accompanied by the author's activities in the W3C Multimedia Semantics Incubator Group [315] (2006/2007), and the W3C Semantic Web Deployment Working Group, as well as the RDFa Taskforce [175] (2006-ongoing).

## 10.1 Motivation

Many multimedia metadata formats, such as ID3, Exif, or MPEG-7 are available to describe—on various levels—what a multimedia asset is about, who has produced it, etc.. With the dissemination of User Generated Content, a need for deploying these metadata in (conventional) HTML pages found in blogs, Wikis, etc. arises. To enable the deployment of multimedia metadata on the (Semantic) Web (2.0), one valid approach is to use the RDF data model for a generic deployment (description) of an arbitrary multimedia metadata format. The step of RDFising is called "formalisation", in our context. To actually deliver the metadata along with the content being served, a new W3C Semantic Web standard—-RDFa—is utilised together with a light-weight vocabulary. This allows a Semantic Web agent to determine the formalisation steps in order to, for example perform a validation, or carry out inference.

### 10.1.1 Last Mile of Multimedia Metadata Deployment

For communications provider, the last mile is the final leg of delivering connectivity to a customer. Equally, in business the last mile is used to describe the process of getting any deliverable to the final consumer. In our case, the last mile is the delivery of multimedia metadata to the end-user, i.e. to a Semantic Web agent. The following observations highlight the basic characteristics we are faced with:

1. Quite an amount of multimedia metadata in formats such as ID3 or Exif is available on the Web [141];
2. RDF [186] is the common data model on the Semantic Web;
3. Formalisations (both on the schema level as on the service level) are available.

With ramm.x we aim at enabling existing multimedia metadata formats to enter the Semantic Web. The goal is to provide self-descriptive media asset descriptions allowing to apply the follow-your-nose principle. We focus on the deployment of multimedia metadata; pretty much as communications provider, we focus on the so called last mile—in our case the consumption and the processing of metadata.

### 10.1.2 Related Work

A generic examination of related work has already been given in Section 2.2.2. We focus on the specific parts in the following.

[1] `http://sw.joanneum.at/rammx/`

To the best of our knowledge no comparable proposal to ramm.x exists. However, we build on existing standards and works. The two main works worth mentioning are the *Multimedia Vocabularies on the Semantic Web* XGR [141] and *Embedding RDF in XHTML* RDFa [2; 5]—ramm.x's deployment and main serialisation syntax. Another work that partially influenced the ramm.x use cases is *Image Annotation on the Semantic Web* XGR [296].

A somehow related approach to ramm.x was proposed by Pfeiffer et.al. [247] back in 2005: The Continuous Media Markup Language (CMML). CMML specifies XML based markup for time-continuous data to allow it to become an integral part of the WWW. The specification allows to attach free-text annotations, metadata, captions and other textual information to clips of time-continuous data, thus enabling a timed textual representation of the data, which can be indexed by Web search engines. Further, CMML also allows to attach hyperlinks to clips of time-continuous data, enabling Web search engines to crawl the content. This also enables users to surf seamlessly between time-continuous data and other Web resources, integrating clips of media into the browsing history of a Web browser. Although the proposal seems very promising, a wide adoption may be not have been taken place due to the complexity and the disruptive model CMML is based on. This is in contrast to ramm.x, which builds on existing standards and only defines a very light-weight vocabulary.

### 10.1.3 Design Principles

The ramm.x specification and its accompanying documents have three core objectives:

- Allow the deployment of any multimedia metadata format in Web documents (such as XHTML, SMIL, SVG, etc.), enabling it to be part of the Semantic Web;
- Utilise and reuse existing multimedia metadata formats (cf. [141, Section 3]), rather than introduce new description formats;
- Offer self-descriptive media asset descriptions allowing to apply the 'follow-your-nose' principle.

In order to realise our objectives, we have decided to utilise RDFa (see section 4.6.2) to deploy the metadata. RDFa allows for a uniform deployment of both the content and the metadata. Further, we have agreed on using formalisations of multimedia metadata formats such as the ones discussed in [141, Section 4]. Finally, we will introduce a light-weight vocabulary enabling tools to process the RDF-based metadata in an ramm.x container.

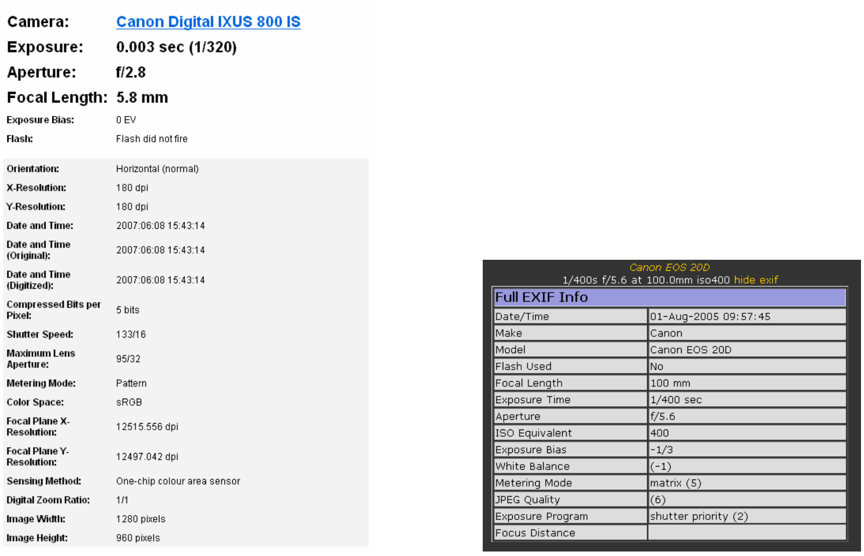
## 10.2 Use Cases

In the following, we list a number of ramm.x use cases. The use cases below have in common that:

1. One or more media assets are published on the Web;
2. Along with the media asset the multimedia metadata (such as Exif) is published;
3. The metadata itself is neither global (free-text) nor natively represented in an RDF-based vocabulary (such as Music Ontology, etc.);
4. A Semantic Web agent seeks to access the multimedia metadata.

### 10.2.1 Use Case: Annotate and Share Photos Online

The XG report [296] presents five uses cases for image annotation using semantic-based technologies. In the solutions presented for the use cases, the metadata descriptions are RDF documents. Deploying them on the Web would often require to embed the complete description into a HTML document using RDFa. In some of the use cases, the source metadata come from Dublin Core, Exif or TV Anytime descriptions. In these cases, ramm.x could be used as an alternative deployment strategy, which avoids to embed the complete RDF document into the HTML page.




**(a)** flickr.com rendering of Exif. **(b)** PBase rendering of Exif.

Figure 10.2: Exif metadata visualizations.

Photo sharing is without any doubt one of the most popular Web 2.0 applications. A number of photo sharing services exist and all of them allow to add some kind of metadata to the images uploaded. However, there are two shortcomings that are addressable by ramm.x, (i) exchanging annotations across services, and (ii) unsupported metadata at hosting service.

**Exchanging and using annotations across services** A number of useful metadata elements for images are for example readily available in the images' Exif (cf. Section 3.3.1) information, for example date and time and in some cases the GPS coordinates of the location. Some photo sharing services, such as Flickr[2] or PBase[3] allow to organize photos using some of these metadata—cf. Fig. 10.2.

However, it is difficult to collect all photos of a certain event (shot around the same geographical position and in the same time range) from different photo sharing sites, or to combine this information with videos (hosted again on another site), blog entries, news articles etc. Embedding the Exif information using ramm.x would make the metadata accessible for other Semantic Web applications.

[2] flickr.com
[3] http://www.pbase.com/

**Metadata not supported by hosting service** Most users use some kind of software for organizing their personal photo collection (e.g. Apple iPhoto, Google Picasa or Adobe Photoshop Album). These applications allow to add various kind of metadata to the photos, from simple things like title, to ratings and identification of the people appearing on the images. The metadata are often stored in standard or de-facto standard metadata formats (an example is XMP). When publishing images to photo sharing services, some of these metadata are lost (as they are not supported by the hosting site) or it is at least tedious to transfer them. Provided that a conversion service from the source format to RDF exists, ramm.x allows to embed or link the complete metadata description, even if the hosting site does not support some of the metadata elements, and makes them also accessible to other applications on the Web.

### 10.2.2 Use Case: Purchasing Music Online

Music is increasingly sold online, both by ordering traditional media like CDs online and by downloading files. Music stores, such as the iTunes[4] music store depicted in Fig. 10.3, provide at least the most basic metadata with their content. As there are a number of common formats for this kind of metadata, using ramm.x would allow to link the metadata of music stores with other Sematic Web resources. This enables applications that automatically link items on the store's site with artist information, reviews, information from fan sites, etc.

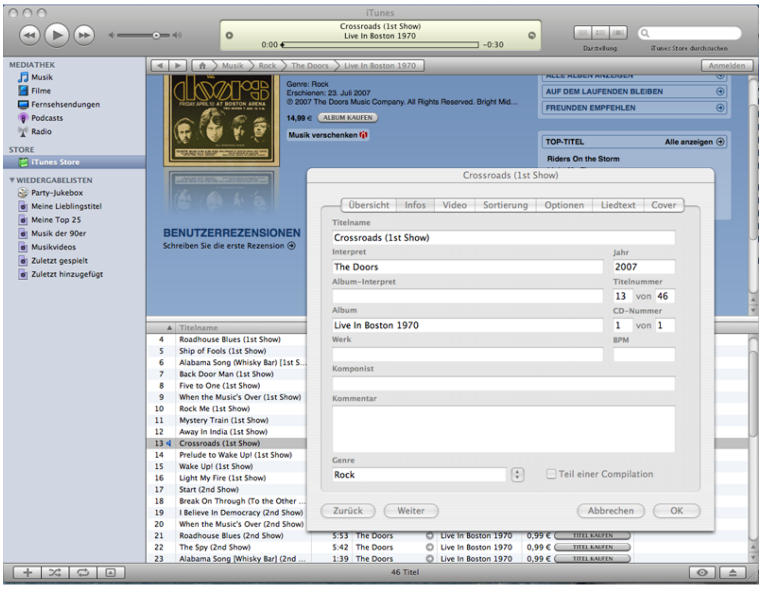



Figure 10.3: Metadata in the iTunes music store.

### 10.2.3 Use Case: Describing the Structure of a Video

While global metadata can be described in a number of simple formats, the metadata related to a certain temporal or spatial range of the content requires the use of more advanced metadata formats. A typical example is the description of the structure of a video, i.e. its

[4] http://www.apple.com/itunes/

scenes, sequences etc. This structure does not only serve as container for metadata valid only for a certain segment, but can also be used for navigation and abstraction of the content.

Assume that Ann creates a video of her last vacation, along with a description of the segmentation of the video into segments that correspond to the different places she has visited. She publishes the video on YouTube together with a ramm.x description of the video structure. Other Web applications can for example use this description as follows:

- A video search engine can provide a key frame or video skim based summary of Ann's video in the result display by selecting frames or clips from the segments. This allows a user to quickly judge the relevance of the result.
- A travel portal site could enhance an article about a certain place by using the corresponding clip from Ann's video.
- A smart video player could not only visualize the video structure like the chapters of a DVD, but also use the annotations of each of the segments to search the Web for related information and display it in context.

### 10.2.4 Use Case: Publishing Professional Content with Metadata

More and more professional content providers offer high quality content on the Web; an example is the BBC Motion Gallery[5], cf. Fig. 10.4 on page 171. In contrast to user generated content, detailed and accurate metadata are available for this kind of content. Currently the metadata are published only in part, and just as text on the Web site. The application of ramm.x does not only allow to publish the metadata in Semantic Web compliant way, but also to directly link the description in a format that is used by the content provider for business to business exchange (such as EBU P_Meta in the broadcast domain), provided that a service for conversion to RDF is available.

### 10.2.5 Use Case: Expressing and Using Complex Rights Information

When media assets are published, it is also important to make the related rights information accessible. This information is for example interesting for multimedia agencies if they want to retrieve images from the Web automatically that they can re-use in advertisings, catalogs, etc. If the rights information only consists of the reference to a certain license (e.g. a specific Creative Commons license) this is trivial. If there is, however, more complex rights metadata (cf. Fig. 10.5 on page 172), for example expressed in MPEG-21 REL, then ramm.x can be used to deploy this metadata.

### 10.2.6 Use Case: Detailed Description of Large Media Assets

Imageine a Web application that allows to create highlight and summary videos of NBA basketball games. Besides its presence on TV, footage of NBA games is available on the Web, even entire games are broadcast via broadband. Basketball content is both spectacular and multifaceted, and therefore well suited for interactive consumption [181].

Almost every aspect of an NBA basketball game is covered by exhaustive statistics. This includes statistics about teams, players (averages, career bests), and games (all game

[5] http://www.bbcmotiongallery.com/

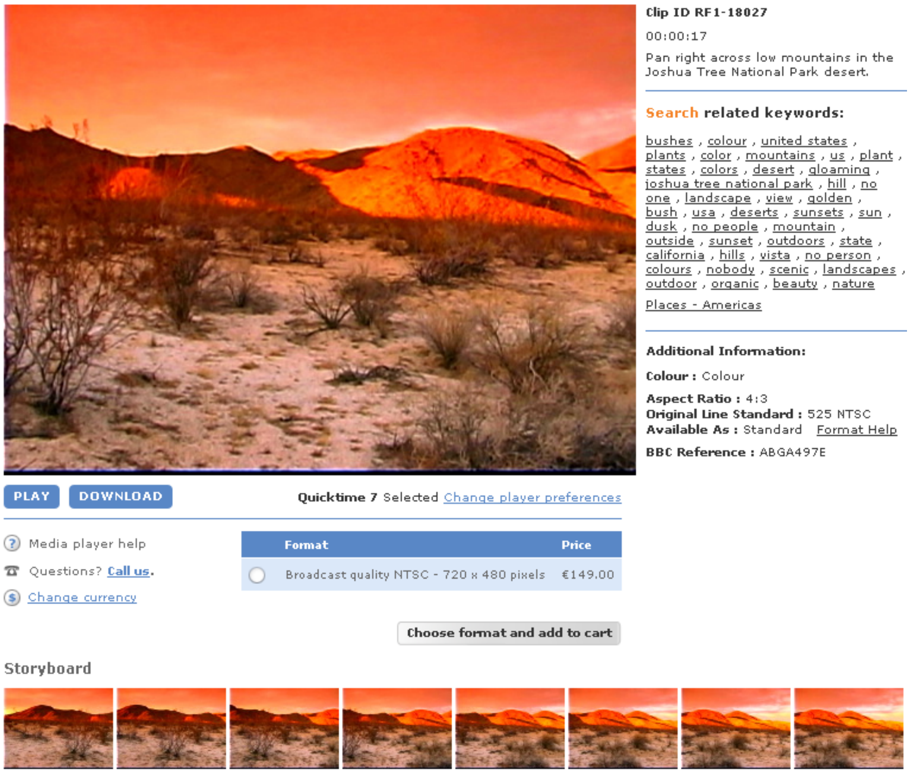



Figure 10.4: Asset offered at BBC Motion Gallery and metadata.

events by exact time, involved players, and action, e.g., free throws or turnovers) including extensive game logs. Comprehensive statistics—both official NBA statistics and further analysis—are publicly available on different web sites. In addition, metadata could be automatically extracted by content analysis approaches, e.g. segmenting and tracking players and describing their trajectories. This means that a huge amount of metadata in different formats is available for one video (cf. Fig. 10.6 on page 173).

When a user watches parts of the videos of a game, e.g. highlights selected based on his personal preferences, a Semantic Web application could use the ramm.x deployed metadata of the basketball game to gather related information and present it to the user. However, the complete description of the game is large and it is time consuming to process all of it, if only the description of a small segment is needed.

### 10.2.7 Use Case: Cultural Heritage

Managing and using multimedia metadata (M3) to facilitate access to cultural objects has always been of particular importance for memory institutions. With the advances of the available technologies, such as digitisation, it became possible to provide access to catalogues, and often the digitised objects itself are made available on the Web. These Web-based applications provide for several kinds of browsing and searching facilities.

Imagine the case of an archive collecting historical newspapers (cf. Fig. 10.7 on page 174), which are scanned per page. Optical character recognition (OCR) can be applied to retrieve the text in the articles and to make it searchable by full text. Other elements of the

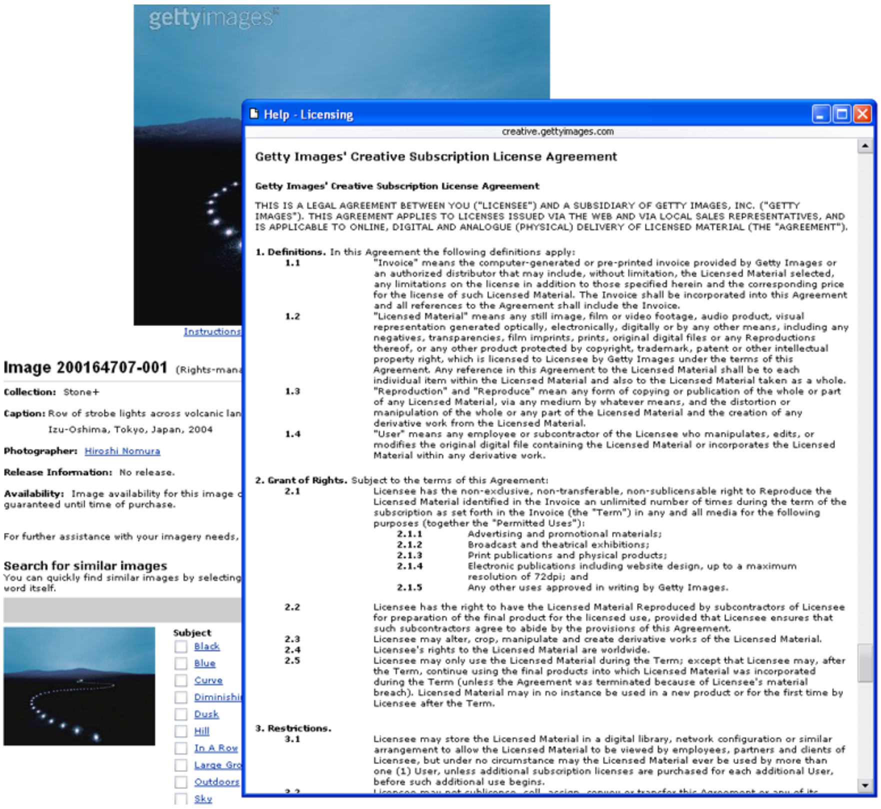



Figure 10.5: Textual rights information for an image offered at Getty Images.

pages, such as illustrations, photographs, advertisements, etc. can be located and extracted during the digitisation process, but are not self-descriptive. In addition, a number of metadata about the asset exists, for example descriptive and administrative metadata.

These metadata are commonly represented using the METS standard [94]; in our example the suggested historical newspaper profile [64] would be appropriate. Another type of metadata is information about the digitisation process (e.g. device, resolution, date/time), which is usually stored as Exif data [91] embedded in the digital image.

The archive in our example decides to make its collection available on the Web. It publishes the original scanned images, the text transcript and the extracted non-text elements. The most relevant of the available metadata elements are put into the asset description on the HTML page. This is very useful for a human viewer of the page.

Let us now assume that a TV journalist wants to edit a documentary on the Hungarian Revolution of 1956. He uses a Semantic Web agent to gather video and image material on this event. Clearly the image on the frontpage of the newsletter depicted in Fig. 10.7 would be relevant in this context, but how could it be linked to other information on the Semantic Web? Some simple descriptive metadata could be represented using the Dublin Core vocabulary [69]. As the Semantic Web agent understands RDF, we could use the RDF representation of Dublin Core [28], either in a separate document or preferably embedded in the HTML page using RDFa [2]. *But what about the information contained in the Exif and METS descriptions?*

For a second case, consider a broadcast archive offering video footage from its historical

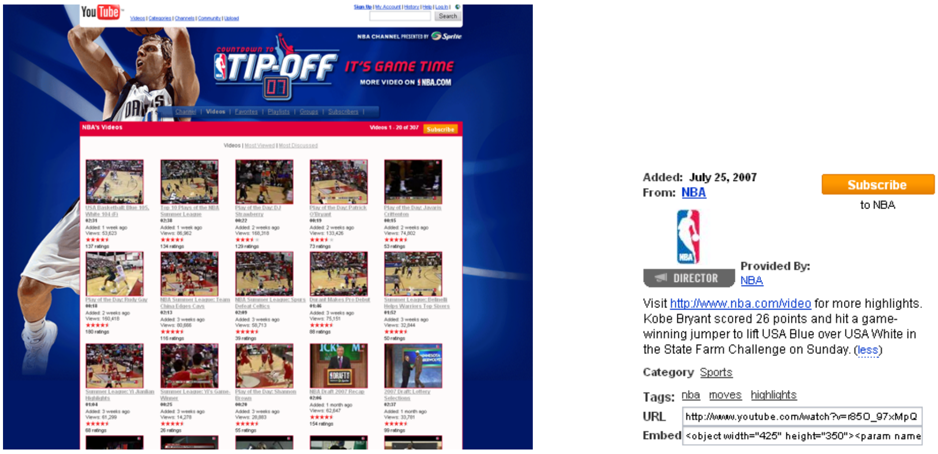



Figure 10.6: NBA content on YouTube and metadata published with it.

news collection for sale, such as the BBC Motion Gallery[6]. Broadcast archives have often comprehensive metadata for their content (especially in the case of news), represented in broadcast industry data formats/models such as EBU P_Meta [251], MXF DMS-1 [82] or SMEF [24]. Especially for time-based media such as video the use of these specific and more comprehensive metadata formats provides a big advantage, as they are for example able to describe the temporal structure of the video and thus allow to reference only the relevant clip of a longer video. But the problem is similar as above: *How could the journalist's Semantic Web agent access these metadata if they are in their native format or rendered into HTML for a human user?*

### 10.2.8 Derived Requirements from the Use Cases

To allow an efficient and effective deployment of multimedia metadata a format has to meet a range of requirements.

From the use cases outlined in Section 10.2, we can derive a set of requirements for an M3 deployment format usable on the Semantic Web (cf. also Fig. 2.3 on page 21):

- Embed references to existing M3 formats in Web container documents, such as (X)HTML. It should be possible to describe also certain parts of a page rather, and the resulting description must be interpretable an Semantic Web agent operating on the RDF-model;
- Provide references to services capable of mapping between a specific multimedia metadata format and RDF (formalisation);
- Several descriptions may be available for a media asset (e.g. in different formats, covering different aspects) and there may be several ways to formalise the description available in a certain format;
- Due to amount of metadata, it may be necessary to "stream" the metadata in pieces related to spatio-temporal segments.

[6] http://www.bbcmotiongallery.com

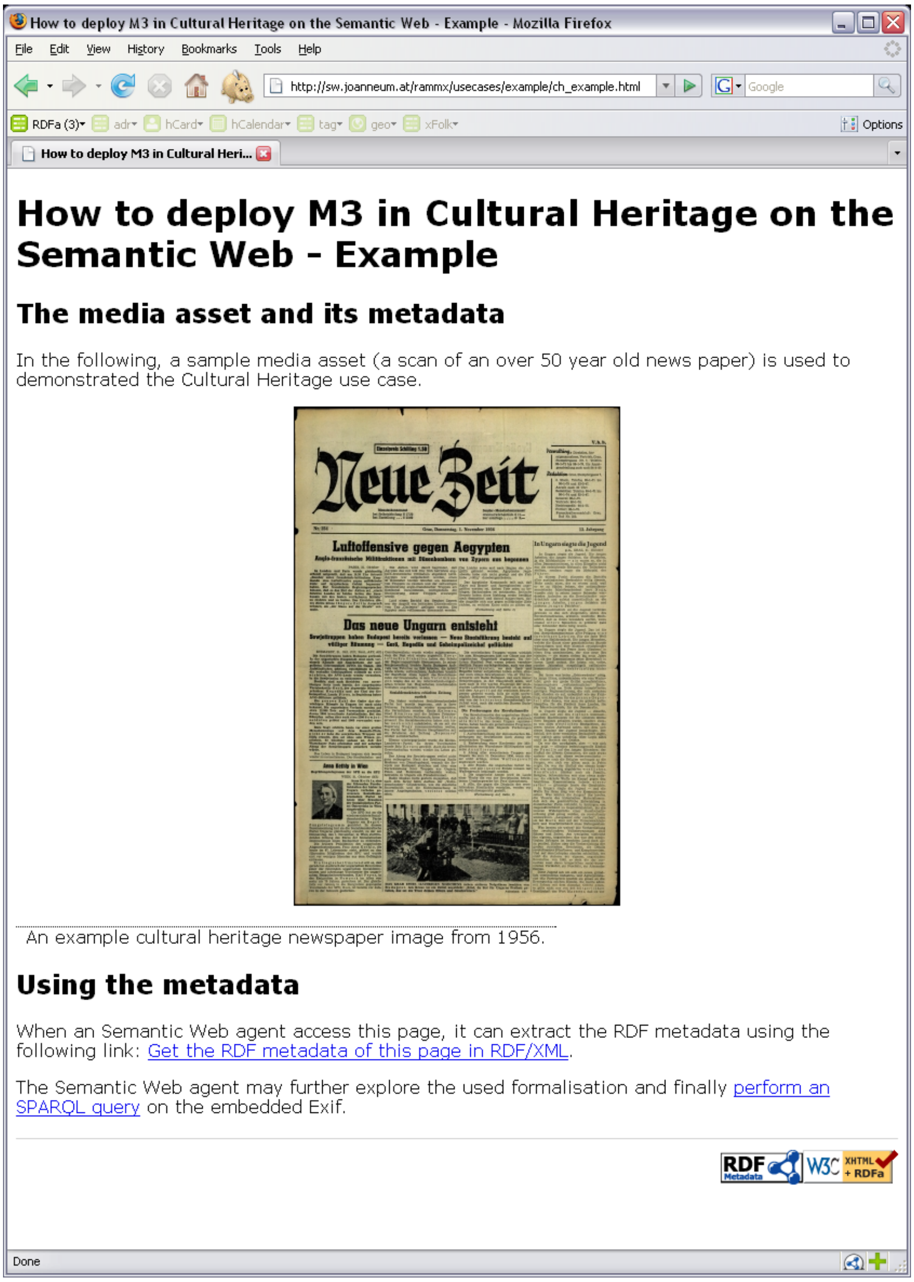



Figure 10.7: A cultural heritage newspaper scan in an XHTML container document.

## 10.3 RDFa-deployed Multimedia Metadata

A vast array of multimedia metadata (M3) formats [141], as Exif, MPEG-7, ID3, etc. is available to describe what an multimedia asset is about. To enable an effective and efficient deployment of this M3 along with the content in an Web environment, the RDFa-deployed Multimedia Metadata (ramm.x) specification [138] was proposed. The ramm.x specification builds upon RDFa [2; 5]—a concrete serialisation syntax of the core Semantic Web data model RDF [186] as the main deployment format. Further, ramm.x utilises formalisations of multimedia metadata formats[141, Sec. 4], this is to say RDF-based schemas of the formats for the actual descriptions. The ramm.x specification allows a Semantic Web agent to determine the formalisation steps in order to perform a validation, or carry out inference.

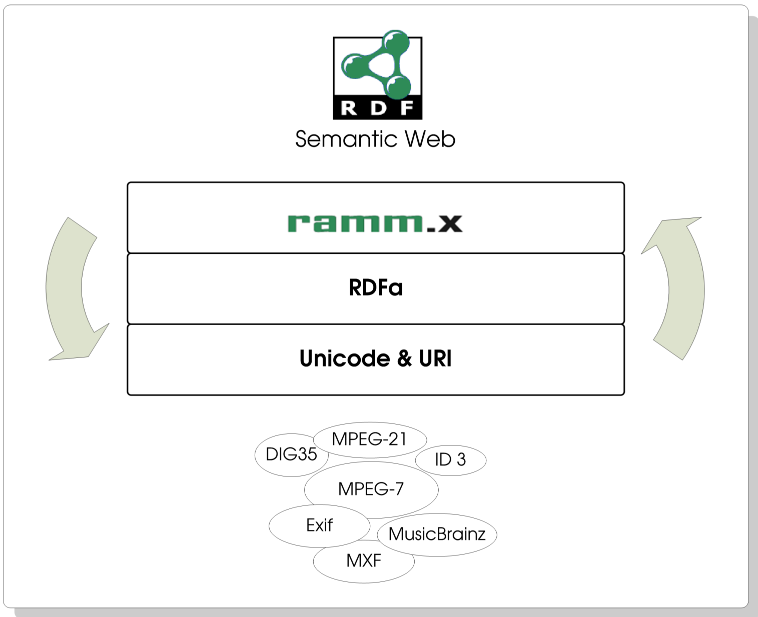



Figure 10.8: Multimedia metadata deployment on the Web.

With ramm.x we aim at enabling existing multimedia metadata formats to enter the Semantic Web—see Fig. 10.8, below. ramm.x targets at self-descriptive media asset descriptions allowing to apply the follow-your-nose principle. With ramm.x, we focus on the deployment of multimedia metadata. Pretty much as communications provider, we focus on the so called last mile—in our case the consumption and the processing of metadata.

### 10.3.1 ramm.x Vocabulary

In the ramm.x core vocabulary four modelling primitives (classes) are defined: `MediaAsset`, `MediaAssetDescription`, `Formalisation`, and `Container`. In the following these terms, and their relations are described. An overview of the ramm.x core vocabulary is given in Fig. 10.9, for a more detailed explanation, the reader is referred to the ramm.x Specification [138], available at http://sw.joanneum.at/rammx/spec/.

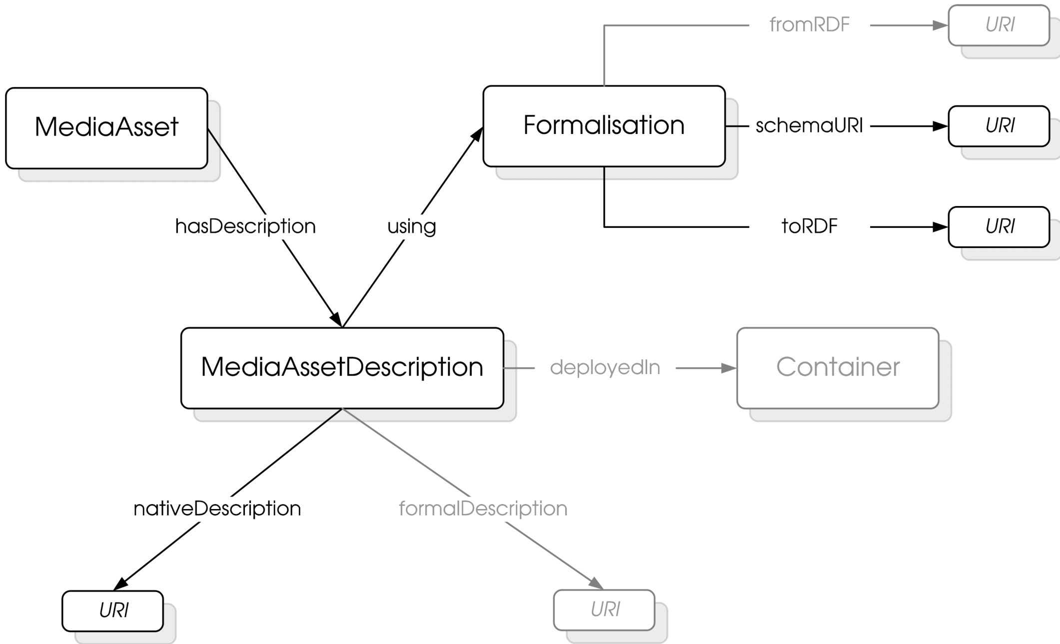



Figure 10.9: The ramm.x core vocabulary at a glance.

#### MediaAsset

A **media asset** (`MediaAsset`) is an information resource as of [118], hence we use[7] the URL of a media asset to identify the media asset itself.

#### MediaAssetDescription (MAD)

**Media asset descriptions** (`MediaAssetDescription`—MAD), are the pivot element for bridging the gap between a multimedia metadata format in its native representation and the formalised version usable on the Semantic Web. Each MAD points exactly to one native description (URL of the native metadata document), and uses exactly one formalisation. Optional there might be a formalised version of the description available as well.

#### Formalisations

A **formalisation** (`Formalisation`) is an explicit representation of the formalisation (or RDFising) process regarding a schema (in RDF-S, OWL, etc.). It provides for URLs pointing to converter services from and to a RDF-based representation, along with the respective schema.

#### Container

A **container** (`Container`), is an—optional—logical unit used to bundle MAD, which has been introduced for a number of reasons:

- One might want to assign certain properties to an array of MAD. Take a copyright note for example: Using a container, a `cc:license` is attached and hence all contained MAD inherit this property;
- Quite a lot of hosting sites allow only for partial control of a page's content (as in a Blog, Wiki, etc. ). With a container it is possible to control the deployment to meet the individual requirements;
- Due to granularity reasons. With a container it is possible to have the whole page as the actual physical host, some parts of it (e.g., a `<div>` element), or even an external location.

Although a container is a handy thing to have, it is an optional element. A fallback rule exists, which states that in case no container is present explicitly, the page in which the MAD is embedded is the container.

### 10.3.2 ramm.x extensions

To keep the ramm.x core specification as light-weight, hence implementable as possible, so called extensions can be defined. We have discussed extensions [139] regarding the following identified issues:

[7] For a discussion on URI declaration vs. usage see also http://dbooth.org/2007/uri-decl/

- **Extensions of the vocabulary**.
  Further classes and/or properties are added to support more complex use cases, such as use case 6. In order to support this use case, the so called "STreaming Extension INterface" (ramm.stein) is proposed;

- **Templates**.
  To foster reusability parts of the specification might be provided as templates. Take for example a formalisation such as an MPEG-7 mapping; for known formalisations a template collection can be provided;

- When a media asset is taken out of its original context (container), one would certainly expect to not lose the associated metadata. This can be achieved by the use of watermarking techniques carrying a pointer to the original container URI.

### 10.3.3 Processing ramm.x Descriptions

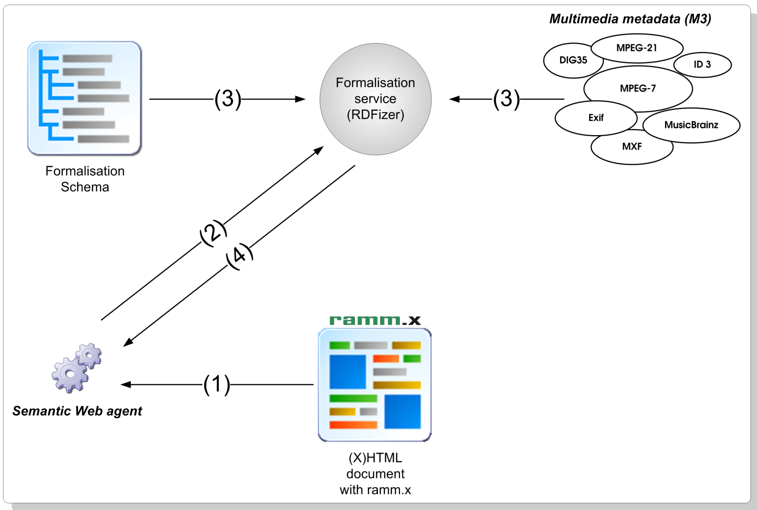



Figure 10.10: Processing ramm.x descriptions.

To actually use the ramm.x deployed metadata, a Semantic Web agent needs to process the container document. The following steps (cf. in Fig. 10.10) are necessary to access the formalised M3:

1. The Semantic Web agent fetches the ramm.x container (typically an XHTML document) with the embedded ramm.x and extracts the RDF using an RDFa extractor (onboard or as an online service—see the RDFa Implementation page[8] for further details);

2. The formalisation service is detected and invoked—this might be a invoking a RESTfull Web service, or a preparation of a SOAP message, etc.

3. Using the `toRDF` property of the `Formalisation` along with the defined schema (from `schemaURI`), the original M3 format (denoted by the value of `nativeDescription`) is converted to its RDF representation by the formalisation service;

[8] http://esw.w3.org/topic/RDFa#Implementations

4. The resulting RDF graph (i.e. the formal representation of the M3 document) is transferred to the Semantic Web agent—ready to be used in a query, or to perform an inference on top of it.

## 10.4 Examples

### 10.4.1 Deploying a Still Image along with Exif Metadata

In [139] we have given a simple application of ramm.x regarding the deployment of a still image along with Exif metadata (cf. section 3.3.1).

Figure 10.11: A sample still image with embedded Exif metadata.

The example[9] in Fig. 10.11 contains an JPEG still image with embedded Exif metadata. The goal is to allow a Semantic Web agent, such as an indexer or a syndication service expecting RDF-based metadata, to make use out of the Exif metadata in the still image.

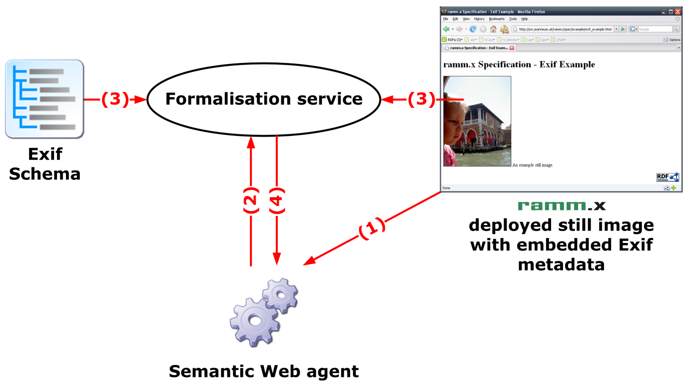



Figure 10.12: Processing ramm.x on a still image with Exif metadata.

In Fig. 10.12 the processing steps are shown a Semantic Web agent typically needs to undertake in order to use the ramm.x deployed metadata:

[9]Available at http://sw.joanneum.at/rammx/spec/example/exif_example.html

1. The Semantic Web agent fetches the ramm.x container (i.e. the XHTML document) with the embedded metadata and extracts the RDF using an RDFa extractor [137];
2. The formalisation service[10] is detected;
3. Using the `toRDF` property of the `Formalisation` along with the defined schema (from `schemaURI`), the Exif metadata is converted to an RDF representation by the formalisation service;
4. The resulting RDF graph is transferred back to the Semantic Web agent and can be further used in, for example, a query.

### 10.4.2 An Example from Cultural Heritage

We now revisit the use case introduced in Section 10.2 regarding a Cultural Heritage application dealing with newspapers. In the following we discuss, how to use ramm.x to deploy a scan of a newspaper's frontpage along with its embedded Exif metadata. When using an XHTML document as the container for both the media asset (the still image of the scanned paper) and the ramm.x MAD, it may yield a result as depicted in Fig. 10.6 on page 173.

The interesting part of the XHTML document is the RDFa-embedded ramm.x MAD along with the formalisation; an excerpt of an exemplary document[11] might look as depicted in listing 10.1.

```
<div about="#exif_formal" typeof="ramm:Formalisation">
  <span rel="ramm:schemaURI" href="http://www.kanzaki.com/ns/exif" />
  <span rel="ramm:toRDF" href="http://www.kanzaki.com/test/exif2rdf" />
</div>
<div about="#sample_mad" typeof="ramm:MediaAssetDescription">
  <div about=
       "http://sw.joanneum.at/rammx/usecases/example/ch_example.jpg"
       typeof="ramm:MediaAsset" >
    <img src="ch_example.jpg"
         alt="Example cultural heritage newspaper image from 1956."/>
      <span rel="ramm:hasDescription" href="#sample_mad"></span>
      <p property="dc:title" datatype="xsd:string">
        An example cultural heritage newspaper image from 1956.
      </p>
  </div>
  <span property="dcterms:created" content="2007-08-24"
        datatype="xsd:date" />
  <span rel="ramm:nativeDescription" href=
        "http://sw.joanneum.at/rammx/usecases/example/ch_example.jpg" />
  <span rel="ramm:using" href="#exif_formal" />
</div>
```

Listing 10.1: XHTML source code excerpt of the deployed media asset.

Put in simple words, the code snippet in 10.1 tells us that there is a media asset (`ch_example.jpg`), which we identify by its URI, hence is both a information and a non-information resource.

[10] Such as http://www.kanzaki.com/test/exif2rdf

[11] Available at http://sw.joanneum.at/rammx/usecases/example/ch_example.html

The media asset has a description `#sample_mad` referring to the native Exif metadata embedded in the media asset. Using the formalisation `#exif_formal`, the full formal description can be obtained. The formal description conforms to the schema proposed by

`http://www.kanzaki.com/ns/exif`

and can be generated using

`http://www.kanzaki.com/test/exif2rdf`.

A Semantic Web agent may access the page[12] and use an RDFa extractor[13] to retrieve the embedded RDF in this page (shown in Fig. 10.2). The RDF N3 notation is used in this example for readability.

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix h: <http://www.w3.org/1999/xhtml> .
@prefix ramm: <http://sw.joanneum.at/rammx/ns#> .
@prefix : <http://sw.joanneum.at/rammx/usecases/example/ch_example.html#>

 :exif_formal a ramm:Formalisation;
               ramm:schemaURI <http://www.kanzaki.com/ns/exif>;
               ramm:toRDF <http://www.kanzaki.com/test/exif2rdf> .

 :sample_mad a ramm:MediaAssetDescription;
             dcterms:created "2007-08-24"^^<xsd:date>;
             ramm:nativeDescription
             <http://sw.joanneum.at/rammx/usecases/example/ch_example.jpg>;
             ramm:using :exif_formal .

 <http://sw.joanneum.at/rammx/usecases/example/ch_example.jpg>
  a ramm:MediaAsset;
  ramm:hasDescription :sample_mad;
  dc:title "Example newspaper image from 1956."^^<xsd:#string> .
```

Listing 10.2: Extracted RDF from an historical newspaper page.

Further, the Semantic Web agent may want to perform an SPARQL query as shown in Fig. 10.3 on page 182. The result of the SPARQL query, hence the answer to the question: *When was the scanned newspaper image digitised?* is as follows: executing the SPARQL query on the resulting RDF graph yields:

`http://sw.joanneum.at/rammx/usecases/example/ch_example.jpg`

was digitised on

`2006-11-13T14:47:0`.

[12] http://sw.joanneum.at/rammx/usecases/example/ch_example.html

[13] Such as http://www.w3.org/2007/08/pyRdfa/

```
prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix k: <http://www.kanzaki.com/ns/exif#>

SELECT ?img ?digiValue

FROM <http://www.kanzaki.com/test/exif2rdf?u=
      http\%3A\%2F\%2Fsw.joanneum.at\%2F
      rammx\%2Fusecases\%2Fexample\%2Fch_example.jpg>

WHERE {
 ?img rdf:type foaf:Image ;
      k:exifdata ?exifData .
 ?exifData k:exif_IFD_Pointer ?ePointer .
 ?ePointer k:dateTimeDigitized ?digiValue.
}
```

Listing 10.3: Querying the embedded RDF metadata of the newspaper scan.

## 10.5 Conclusion and Future Work

Archives and libraries increasingly publish parts of their collection on the Web, including not only textual information but also multimedia assets. In contrast to text documents multimedia assets can only be indexed and interpreted if published together with metadata describing them. We have identified a shortcoming in deploying multimedia metadata (M3) on the Semantic Web and have proposed a solution which allows to deploy M3 vocabularies so that they can be used by Semantic Web agents understanding RDF. We have described two related use cases motivated by the search for image and video motivation, one based on a newspaper archive and the other on a broadcast archive. Using the newspaper example, we have outlined how ramm.x can be employed for deploying metadata so that it is accessible for a Semantic Web agent.

The RDFa-deployed Multimedia Metadata (ramm.x) framework allows to deploy references to media asset descriptions represented in any multimedia metadata vocabulary in a Web document. It uses a small RDF vocabulary consisting of four classes which is embedded into the host document using RDFa. It has to be emphasized that ramm.x is not an RDF-based multimedia description vocabulary or a multimedia ontology, nor does it replace existing technologies. It is complementary to multimedia metadata vocabularies, their formalisation and Web technologies in that it allows the deployment of non-RDF media asset descriptions on the Semantic Web, linking the host document of the media asset, the media description and a service capable of providing a formalisation (i.e. a RDF representation) of the media assset description.

ramm.x is useful if the media asset description is not natively available in RDF, otherwise RDFa would be sufficient. Using specific M3 vocabularies is necessary the more detailed the media description is, e.g. for describing the (spatio-)temporal structure of a video, with annotations on all of the segments. Also in many archives and libaries, the description in a native M3 format is readily available and thus ramm.x simplifies deployment. Of course the successful use of ramm.x depends on the availability of a service capable of formalising the native description format.

From the exemplary use of ramm.x in the cultural heritage domain two issues have been identified that should be addressed by future extensions. One concerns collecting formalisations of different M3 vocabularies as well as the services capable of doing the formalisation, which facilitates the practical use of ramm.x. The other issue comes from the broadcast archive use case described in Section 10.2 and concerns the deployment of larger media asset descriptions, such as a detailed annotation of a one hour documentary. Clearly ramm.x is more beneficial in such a case than when just describing a few global metadata elements. But as the media asset description can be very large in that case, a kind “streaming mode”, that allows to access the part of the description that is relevant to the current segment of the media asset would be useful.

# Part IV

# Conclusion and Outlook

# Chapter 11

# Conclusions

*"Acta est fabula, plaudite!"*

(Ending phrase of ancient Roman comedies)

The cornerstones of using multimedia metadata on the Semantic Web have been extensively researched in the recent years [310; 229; 141]. However, while certain issues, such as interoperability, etc. have been addressed, the practical utilisation of multimedia metadata on the Semantic Web regarding scalability and expressivity is still widely neglected. However, existing Web applications handling millions of multimedia assets are starting to take advantage of Semantic Web technologies [237].

The research of the doctoral thesis at hand contributes to building scaleable and smart Semantic Web multimedia applications by focusing on selected real-world issues. We have discussed related efforts in this work, and have described the state-of-the art both regarding multimedia metadata and Semantic Web tehcnologies in the first part of this work. Further, we have elaborated on methods and requirements regarding the design of scaleable and smart Semantic Web multimedia applications in the second part. Based on an analysis of practical issues stemming from diverse projects and activities the author has participated in over the past four years the following areas have been identified:

- On the data access level we have proposed a performance and scalability metric
- Regarding the effective and efficient representation of multimedia content descriptions a scalable approach has been proposed;
- Finally, to address an effective and efficient multimedia metadata deployment, a new deployment framework has been proposed and its application has been discussed.

**Processing Metadata Sources** In chapter 8) we have proposed a performance and scalability metric [149], along with a demonstration in a social media site for image-sharing. Further, we have recently shown how to apply good practices in building a Semantic Web application [148], and have demonstrated the system at the Semantic Web Challenge 2007.

**Formal Representation of Multimedia Content Descriptions** Today's multimedia metadata formats (such MPEG-7) are typically not grounded in formal languages. Semantic Web languages based on RDF are regularly utilised to resolve interoperability issues and to formalise the domain of interest, but typically do not offer special multimedia content description facilities.

To address this issue we have shown in chapter 9 we have proposed a scalable multimedia ontology [134]. Further, this work has been applied [147] in the 'New Millennium, New Media' (NM2) project. Finally, we have shown how the workflow of the exemplary non-linear media production environment in NM2 can be understood as an instantiation of one or more of the identified canonical processes [127].

**Multimedia Metadata Deployment on the Semantic Web** . To address the multimedia metadata deployment issue, the author has—together with Werner Bailer, Tobias Bürger and Raphael Troncy—initiated the RDFa-deployed Multimedia Metadata (ramm.x) initiative [138]. In chater 10 we have outlined the ramm.x framework and discussed potential use cases. We have also shown how to apply ramm.x and highlighted several extensions. The work on ramm.x was accompanied by the author's activities in the W3C Multimedia Semantics Incubator Group (2006/2007) and the RDFa Task Force of the Semantic Web Deployment Working Group (ongoing).

Concluding we note that scalability is a critical success factor regarding Semantic Web applications. A lot of research works only use toy-datasets in their setup. However, to realise the Semantic Web one needs sound and Web-scale datasets including emerging semantics. We have learned that it is possible to achieve scalability and expressivity on a Web-scale when taking practical limitations into account (see also chapter 7).

Recent activities such as the Linking Open Data community project[1], where the author of this thesis has been active since mid 2007, are trailblazing examples of "real-world" semantic content. However, also in this area future work regarding multimedia assets is urgently needed as up to now only global descriptions of textual resources have been RDFised and interlinked.

Based on the work presented herein we can formulate the following three generic advises for building scalable, yet smart Semantic Web multimedia applications:

- Start with a simple approach (KISS principle) and add complexity as required by the functionality (such as inferential power, etc.);
- Build on a decentralised architecture—it is at the heart of the Web architecture to be distributed, hence any application running on it should follow this approach;
- Use both human and computational power: there are several problems a human is better in solving it; this should be exploited.

We conclude this work with a reference to a possible solution that we hereby entitle "interlinking multimedia" (IM); for further details on IM the reader is referred to section 12.2.2.

[1] http://linkeddata.org/

# Chapter 12

# Outlook

*"Quid nunc?"*

(Latin phrase)

This last chapter aims to present an outlook regarding Semantic Web applications. Especially multimedia content is an emerging topic influencing already broad audiences. Most of the involved parties are still passive, though the situation is changing: from the consumer to the contribuser[1]. First we will discuss open issues regarding the work presented herein. Possible extensions are highlighted and future work is sketched. Then, the chapter highlights some general trends in the realm of multimedia applications on the Semantic Web and points out possible directions.

## 12.1 Semantic Web multimedia applications now and in 10 years time

This section discusses how currently available multimedia Web applications might migrate to the Semantic Web by means of utilising Semantic Web technologies, or at least some structured metadata technologies allowing to gently entering the Semantic Web.

### 12.1.1 Emerging Metadata

This sections contemplates on ongoing efforts regarding metadata in multimedia systems on the Web. Most of the approaches are grass-root approaches usually focusing on practicability, scalability and performance.

**hAudio** Manu Sporny (Digital Bazaar) and others proposed *hAudio*[2] to capture global music related metadata and make it interoperable on the Semantic Web. hAudio is a simple, open, distributed format, suitable for embedding information about audio recordings in (X)HTML, Atom, RSS, and arbitrary XML. hAudio is one of several microformats open standards. They defined a mapping of hAudio to RDFa.

[1] a portmanteau word; from **contrib**utor and **user**

[2] http://wiki.digitalbazaar.com/en/HAudio_RDFa

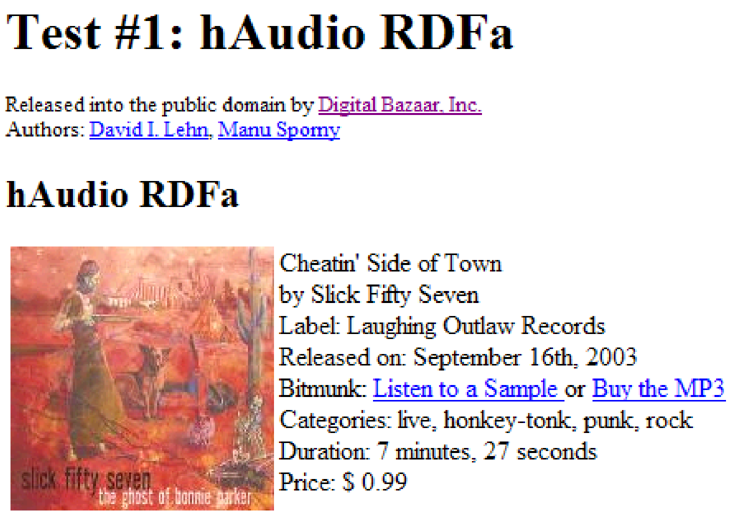



Figure 12.1: hAudio example.

An exemplary usage[3] of hAudio is depicted in Fig. 12.1. The resulting triples[4] are listed below, in Fig. 12.1.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix hcommerce: <http://www.microformats.org/2007/12/hcommerce/> .
@prefix hmedia: <http://www.microformats.org/2007/11/hmedia/> .
@prefix money: <http://www.microformats.org/2007/10/money/> .
@prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
@prefix : <http://sw.joanneum.at/rdfa/hAudio/haudio_rdfa_test.html#> .

:cheatin_side_of_town_rdfa a hmedia:Audio;
    hmedia:image-summary _:r1189671498r28118r1;
    dc:title "Cheatin' Side of Town";
    dc:creator "Slick Fifty Seven";
    dc:contributor [
        vcard:role "Label";
        vcard:org "Laughing Outlaw Records";
        vcard:fn "Laughing Outlaw Records" .
    ];
    dc:date "2003-09-16"^^xsd:date;
    hmedia:sample <http://www.bitmunk.com/sample/6011101>;
    hcommerce:payment <http://www.bitmunk.com/purchase/webbuy/6011101>;
    dc:type "live", "honkey-tonk", "punk", "rock";
    hmedia:duration "447S";
    money:price [
        money:currency "USD";
        money:amount "0.99" .
    ] .
```

Listing 12.1: Resulting triples from the hAudio example.

[3] http://sw.joanneum.at/rdfa/hAudio/haudio_rdfa_test.html

[4] The base URI is assumed implicitly.

**Machine tags** Flickr[5] rolled out a new feature called *machine tags* in early 2007. Machine tags[6] allow users to more precisely tag and search their photos; they are tags using a special syntax to define extra information about a tag. Machine tags have a namespace, a predicate and a value. The namespace defines a class or a facet that a tag belongs to ("geo", "flickr", etc.), and the predicate is the name of the property for a namespace ("latitude", "user", etc). For example, `dc:title='Message in the Bottle'`, could denote the title of a well-known Police song.

**RDFa in multimedia metadata**. Existing multimedia Web applications start to utilise RDF for different purposes. For example, Joost[7] is about to adopt RDFa. Further, flickr already utilises RDFa[8] for global metadata (see Fig. 12.2).

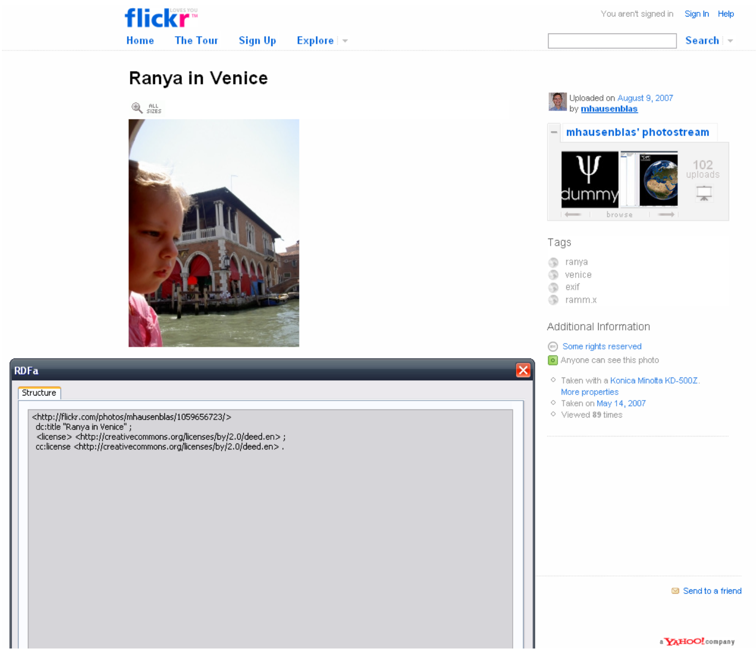



Figure 12.2: Usage of RDFa in flickr.

### 12.1.2 Advanced Annotation Techniques

For tasks computers currently are too limited, human computational power can be exploited. This approach leverages differences in abilities and alternative costs between humans and computer agents to achieve symbiotic human-computer interaction. In traditional computation, a human employs a computer to solve a problem: a human provides a formalized problem description to a computer, and receives a solution to interpret. In human-based computation, the roles are often reversed: the computer asks a person or a large number of people to solve a problem, then collects, interprets, and integrates their solutions.

Recently, Luis von Ahn [313] proposed a possible application of human computation for annotating media assets. This technology is now being used in the Google Image Labeler,

[5] flickr.com
[6] http://www.flickr.com/groups/api/discuss/72157594497877875/
[7] http://www.joost.com
[8] http://flickr.com/photos/mhausenblas/1059656723/


Figure 12.3: Google's Image Labeler.

as depicted in Fig. 12.3.

Related approaches—often labelled artificial artificial intelligence—to utilise the human computational power are already in use. For example Amazon started to offer a service called *Mechanical Turk*[9] a couple of years ago that allows people to contribute in solving simple tasks on a micropayment base. A similar service is ChaCha[10], a human-powered search engine, realised through employing people to help users sift through search results using a chat interface.

However, there are also scalability issues with these approaches. Firstly, by and large all human computation approaches assume a rather huge use group being available, which is certainly not always feasible. Secondly, it has to be questioned, if the approaches is generally applicable as it assumes the availability of public available data. Promising research [270]) is ongoing in this area, however.

### 12.1.3 Interactive Media

The desire to interact with digital media is steadily increasing. In [147] we have discussed the future of interactive media and concluded that ...

> ... the real challenge will be, however, to provide these tools not only for environments based on the desktop paradigm (graphic interface, keyboard, mouse), but rather for interfaces allowing an even more direct interactive way of manipulating media, such as authoring tools on mobile phones (a production unit with camera and voice control), which certainly will be more accessible to those who are used to talk rather than write but are highly trained in visual communication, as large populations in Asia.

[9] http://www.mturk.com/
[10] http://www.chacha.com/

A trailblazing example of a social media tool is the *International Remixer*. Visitors at the San Francisco International Film Festival used the International Remixer (Fig. 12.4), a Web-based editing suite, to remix multimedia selections from 19 films[11]. We note that we have

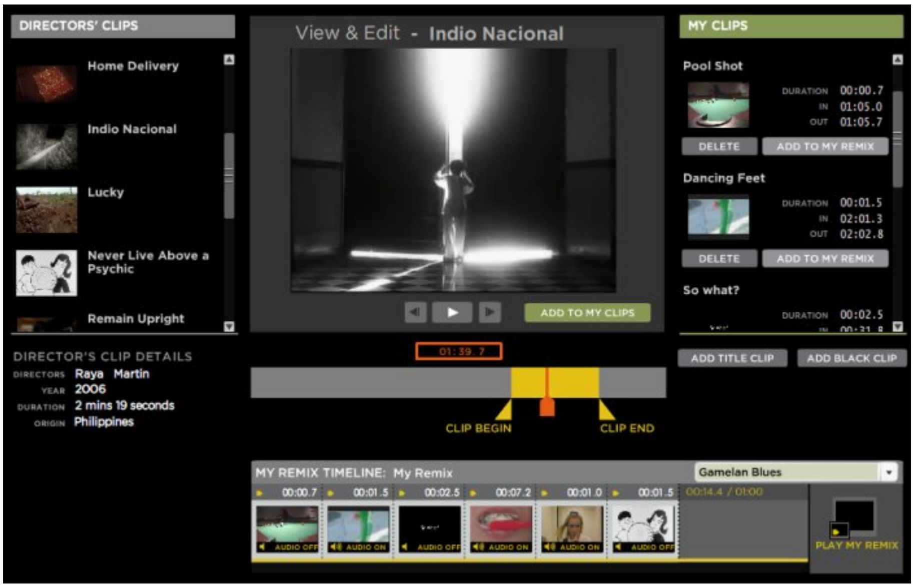



Figure 12.4: International Remix.

recently started to investigate the possibilities in this realm [181].

## 12.2 Future Work

### 12.2.1 Meshups and More

Meshups might well turn out to be the next big step towards a true Web 3.0. As Kingsley Idehen, CEO of OpenLink software, puts it[12]:

> A defining characteristic of the Data Web (Context Oriented Web 3.0) is that it facilitates Meshups rather than Mashups.
>
> - Quick Definitions
>   - Mashups—Brute force joining of disparate Web Data
>   - Meshups—Natural joining of disparate Web Data
> - Reasons for the distinction:
>   - Mashups are Data Model oblivious
>   - Meshups are Data Model driven
> - Examples
>   - Mashups are based on RSS 2.0 most of the time (RSS 2.0 is at best a Tree Structure that contains untyped or meaning challenged links.

[11] See http://timetags.research.yahoo.com/ for example results.
[12] http://www.openlinksw.com/blog/~kidehen/?date=2007-03-22

– Meshups are RDF based and the data is selfdescribing since the links are typed (posses inherent meaning thereby providing context).

In order to support the discovery and usage of linked datasets, hence creating Meshups, we have recently started to develop "vocabulary of interlinked datasets" (voiD)[13] allowing to describe the content of linked datasets in a formal way.

### 12.2.2 Multimedia and the Web of Data

In late 2006 we, that is, some members of the Multimedia Metadata Incubator Group, have organised a panel entitled "The Role of Multimedia Metadata Standards in a (Semantic) Web 3.0"[14] for the WWW07. Though some issues have been heavily discussed, the base vibe was a positive one. However, one of the panellists, Mor Naaman from Yahoo! Research Berkeley, declared quite provocatively "the Semantic Web dead"[15].

Another very promising path is the W3C "Video in the Web" activity[16]. We have participated in a workshop that kicked-off this activity in late 2007 and presented there our position regarding spatio-temporal fragment identifiers for audio-visual content [294]. As one result, the "Media Fragments Working Group"[17] has been established; the author of this thesis is a member of the Woking Group.

With "Catch Me If You Can" (CaMiCatzee) [142] we have only recently started to demonstrate the capabilities and opportunities of "interlinking multimedia" (IM).

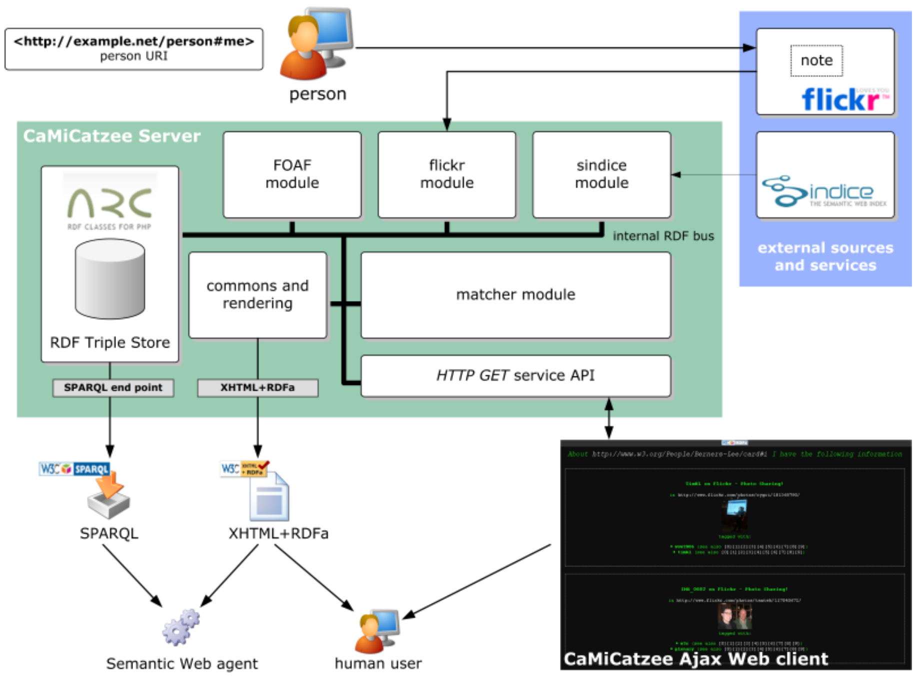



Figure 12.5: CaMiCatzee's system architecture.

[13] http://community.linkeddata.org/MediaWiki/index.php?VoiD
[14] http://www2007.org/panel4.php
[15] http://tinyurl.com/26dyx6
[16] http://www.w3.org/2008/WebVideo/
[17] http://www.w3.org/2008/WebVideo/Fragments/

CaMiCatzee[18] (Fig. 12.5) is an IM concept demonstrator which main goal is to show how regions in still images (from flickr) can be interlinked on the Web of Data. In CaMiCatzee, the User Contributed Interlinking (UCI) [143] principle is applied to multimedia assets the first time. The primary domain for annotations in CaMiCatzee is the depiction of people, hence the usage of FOAF profiles.

In the realm of IM we have also initiated according discussions[19] which we hope to extend in order to advance this field.

[18] http://sw.joanneum.at/CaMiCatzee/

[19] http://community.linkeddata.org/MediaWiki/index.php?InterlinkingMultimedia

# Part V

# Addendum

Appendix A

# Sources

In this chapter diverse sources are listed which have been used throughout in this work. This may well be data formats (such as RDF source code) or program source code.

## A.1 RDF Source Codes

This section gathers source codes of RDF graphs used in the doctoral thesis at hand.

### A.1.1 Minimalistic Media Ontology Example

The following are details about an exemplary RDF graph used in section 7.3. It demonstrates how an ontology can be established utilising RDF, and shows some minimal possible inference.

**T-Box**

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix mm: <http://sw-app.org/example/mm#> .

mm:Content a owl:Class .

mm:Video a owl:Class;
         rdfs:subClassOf mm:Content .

mm:Audio a owl:Class;
         rdfs:subClassOf mm:Content .

mm:Topic a owl:Class .

mm:hasTopic a owl:ObjectProperty;
            rdfs:domain mm:Content;
            rdfs:range mm:Topic .
```

Listing A.1: RDF source code of the minimal media ontology (T-Box).

Listing A.1 contains the source code of a simple, exemplary RDF graph. It represent a minimalistic multimedia ontology comprising content and topic classes.

In the following, instances utilising this T-Box are shown.

**A-Box**

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix mm: <http://www.sw-app.org/examples/multimedia#> .
@prefix mex: <http://www.sw-app.org/examples/multimedia/abox1#> .

<> a owl:Ontology;
   owl:imports <http://www.sw-app.org/examples/multimedia> .

mex:Economy a mm:Topic .

mex:News a mm:Topic .

mex:Politics a mm:Topic .

mex:v1 a mm:Video;
       mm:hasTopic mex:Economy .

mex:v2 a mm:Video;
       mm:hasTopic mex:News,
                   mex:Politics .
```

Listing A.2: RDF source code of the minimal media ontology (A-Box).

In listing A.2 a sample RDF graph utilising the multimedia ontology (T-Box) from listing A.1 is depicted. It basically states that there are several topics (`News`, etc.) and two videos *about* these topics.

## A.2 Program Source Code

### A.2.1 Performance and Scalability Metric Showcase

As described in chapter 8, a showcase for non-native RDF sources (PSIMeter[1]) has been developed in the realm of this work. The showcase demonstrates the application of the metric by RDFising the flickr API.

In the following the Java source code constituting the core of the PSIMeter application is listed. It contains the implementation of the three approaches to RDFise the Flickr-API, namely (i) Approach A uses the search functionality of the flickr API, (ii) Approach B uses the flickr API to retrieve all public photos and then uses a local XSL transformation to generate the RDF graph, and (iii) Approach C retrieves all public photos, as in Approach B and then, for each photo an external service is invoked to generate the RDF graph.

The complete source code including the Eclipse development files as well as the runtime environment settings (a Tomcat Web application) can be found at http://sw.joanneum.at/psimeter/.

[1] http://sw.joanneum.at:8080/psimeter/

```java
/**
 *
 */
package at.joanneum.sw.psimeter;

import java.io.BufferedReader;
import java.io.File;
import java.io.IOException;

import java.io.ByteArrayInputStream;
import java.io.InputStreamReader;
import java.io.StringReader;
import java.io.StringWriter;
import java.io.UnsupportedEncodingException;
import java.net.MalformedURLException;
import java.net.URL;
import java.net.URLEncoder;
import java.util.Date;
import java.util.HashMap;
import java.util.Iterator;
import java.util.Vector;

import javax.xml.parsers.DocumentBuilder;
import javax.xml.parsers.DocumentBuilderFactory;
import javax.xml.parsers.ParserConfigurationException;
import javax.xml.transform.Transformer;
import javax.xml.transform.TransformerConfigurationException;
import javax.xml.transform.TransformerException;
import javax.xml.transform.TransformerFactory;
import javax.xml.transform.TransformerFactoryConfigurationError;
import javax.xml.transform.dom.DOMSource;
import javax.xml.transform.stream.StreamResult;
import javax.xml.transform.stream.StreamSource;

import org.xml.sax.SAXException;

import com.aetrion.flickr.Flickr;
import com.aetrion.flickr.FlickrException;
import com.aetrion.flickr.people.PeopleInterface;
import com.aetrion.flickr.people.User;
import com.aetrion.flickr.photos.Photo;
import com.aetrion.flickr.photos.PhotoList;
import com.aetrion.flickr.photos.PhotoUtils;
import com.aetrion.flickr.photos.PhotosInterface;
import com.aetrion.flickr.photos.SearchParameters;
import com.aetrion.flickr.tags.Tag;
import com.aetrion.flickr.tags.TagsInterface;
import com.hp.hpl.jena.query.Query;
import com.hp.hpl.jena.query.QueryExecution;
import com.hp.hpl.jena.query.QueryExecutionFactory;
import com.hp.hpl.jena.query.QueryFactory;
import com.hp.hpl.jena.query.QuerySolution;
import com.hp.hpl.jena.query.ResultSet;
import com.hp.hpl.jena.query.ResultSetFormatter;
import com.hp.hpl.jena.rdf.model.Model;
import com.hp.hpl.jena.rdf.model.ModelFactory;
import com.hp.hpl.jena.util.FileManager;
import com.hp.hpl.jena.vocabulary.DC;

/**
```

```
 * @author Michael Hausenblas
 *
 */
public class PSIMeter {

public static boolean doDebug = false;
public static boolean doInfo = false;
public static String graphOutFormat = "N3";

public static final String CONSTRUCT_RESULT_PHOTO_DIRECT =
" prefix address: <http://example.org/address#> " +
" prefix cell: <http://machinetags.org/wiki/Cell/> " +
" prefix filtr: <http://example.org/filtr#> " +
" prefix flickr: <http://flickr.com/tags/meta#>  " +
" prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> " +
" prefix location: <http://example.org/location#> " +
" prefix people: <http://example.org/people#> " +
" prefix ph: <http://example.org/ph#> " +
" prefix sxsw: <http://sxsw.com/> " +
" prefix upcoming: <http://upcoming.org/> " +
" prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> " +
" prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> " +
" prefix dc: <http://purl.org/dc/elements/1.1/> " +
" prefix dcterms: <http://purl.org/dc/terms/> " +
" prefix foaf: <http://xmlns.com/foaf/0.1/> " +
" prefix xsd: <http://www.w3.org/2001/XMLSchema#> " +
" construct { ?pURL dc:title ?value  }" +
" where { " +
"  ?pURL dc:title ?value .  " +
"  FILTER regex(?value, \"nm2\", \"i\") " +
" } ";

public static final String SELECT_PHOTO_ID =
"prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> " +
"prefix dc: <http://purl.org/dc/elements/1.1/> " +
"prefix foaf: <http://xmlns.com/foaf/0.1/>  " +
"prefix flickr: <http://flickr.com/tags/meta#> " +
" select ?pID" +
" where { " +
"  ?p rdf:type flickr:Photo ;" +
"     flickr:photoID ?pID ." +
" } ";

public static final String SELECT_USER_ID =
"prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> " +
"prefix dc: <http://purl.org/dc/elements/1.1/> " +
"prefix foaf: <http://xmlns.com/foaf/0.1/>  " +
"prefix flickr: <http://flickr.com/tags/meta#> " +
" select distinct ?uID" +
" where { " +
"  ?p rdf:type flickr:Photo ;" +
"     flickr:ownerID ?uID ." +
" } ";


private FlickrProxy fp;
private String baseURL;

private HashMap mMap;

```

```java

public PSIMeter(String serviceBaseURL) {
        fp = new FlickrProxy();
        this.baseURL = serviceBaseURL;
        this.mMap = new HashMap<String,Metrics>();
}

/**
 * @param args
 */
public static void main(String[] args) {
        String ham = "michael.hausenblas@joanneum.at";
        String danny = "danny.ayers@gmail.com";
        String userEMailAddress = ham;
        String uID;
        File resultFile = new File("results/"+ (new Date()).hashCode() +
            ".metrics");

        // for 'list user that use machine tags in namespace'
        String inNamespace = "geo:";

        // handle parametrised query
        String userProp = "dc:subject";
        String userValue = "Marian";
        String userQuery = PSIMeter.CONSTRUCT_RESULT_PHOTO_DIRECT;
        userQuery = userQuery.replaceAll("dc:title", userProp);
        userQuery = userQuery.replaceAll("nm2", userValue);

        String[] tags = {"\"machine_tags\" => \""+ userProp + "=\\\""+
            userValue +  "\\\""};

        PSIMeter pm = new PSIMeter(""); // empty base -> run local

        uID = pm.userEMail2ID(userEMailAddress);

        PSIMeter.doDebug = true;
        PSIMeter.doInfo = true;

        pm.executeApproachA(uID, tags, userQuery);
        System.out.println(pm.getMetric("A"));

        pm.executeApproachB(uID, userQuery);
        System.out.println(pm.getMetric("B"));

        pm.executeApproachC(uID, userQuery);
        System.out.println(pm.getMetric("C"));

        // Serialise all metrics
        MetricsUtil.asXML(pm.getMetricsMap(), resultFile,
            pm.getNumOfPublicPhotosOfUser(userEMailAddress));
        try {
                String fURL = resultFile.toURL().toString();
                System.out.println(fURL);
                pm.dumpContentFromURL(fURL);
        } catch (MalformedURLException e) {
                e.printStackTrace();
        }
}


```

```java

/**
 * Uses the flickr API to retrieve photos of a user tagged with a certain
     machine tag,
 * and adds the tags to each photo, converting the result into RDF using
     a local XSLT.
 *
 * @param userIDs
 * @param tags
 * @param sparqlQuery
 */
public String executeApproachA(String userID, String[] tags, String
    sparqlQuery) {
        Vector pIDList = new Vector();
        String result = "";
        String fRetrievePhotosMTaggedFromUserURL = "";
        Model uGraph = ModelFactory.createDefaultModel();
        Metrics mA = new Metrics("Approach A", sparqlQuery, uGraph);

        mA.startOverallOperation();
        mA.startConversion();

        // create URL for flickr API to search in machine tags
        fRetrievePhotosMTaggedFromUserURL =
            fp.createFlickrPhotosSearchURLOfUserWithID(userID, tags[0]);

        if(PSIMeter.doDebug) System.out.println("[CMD] " +
            fRetrievePhotosMTaggedFromUserURL);

        // add matching photos (photo URL and owner ID only) to graph
        uGraph.add(constructRDFGraphFromStr(
                   applyTransformation(fRetrievePhotosMTaggedFromUserURL,
                   this.baseURL + "transform/flickrphoto2rdf.xslt")));

        // extract list of photo IDs ...
        pIDList  = createPhotoIDList(uGraph);

        // ... to add tags for each photo to the graph
        for (int i = 0; i < pIDList.size(); i++) {
                this.addTagsToPhotoInGraph(uGraph,
                    (String)pIDList.get(i));
        }
        mA.endConversion();
        mA.startOperation();
        // the target operation to execute
        result = dumpRDFGraphToStr(doConstructQuery(sparqlQuery, uGraph),
            PSIMeter.graphOutFormat);
        mA.endOperation();
        mA.endOverallOperation();

        if(PSIMeter.doInfo) {
                System.out.println(mA);
                System.out.println(dumpRDFGraphToStr(uGraph,
                    PSIMeter.graphOutFormat)); // dump merged RDF graph
                System.out.println(result);
        }

        this.mMap.put("A", mA);
        return result;
}
```

```
/**
 * Using flickr API to retrieve all public photos of user, and construct
 * RDF using local XSLT.
 *
 * @param userID
 * @param sparqlQuery
 */
public String executeApproachB(String userID, String sparqlQuery) {
        PhotoList pl = null;
        Model uGraph = ModelFactory.createDefaultModel();
        String result = "";
        Metrics m = new Metrics("Approach B", sparqlQuery, uGraph);

        m.startOverallOperation();
        m.startConversion();

        // create list of public photos of a user using the flickrj API
        pl = fp.listPublicPhotosOfUserWithID(userID);

        // add all public photos (including tags) of user to the graph
        for (Iterator iter = pl.iterator(); iter.hasNext();) {
                Photo p = (Photo) iter.next();
                this.addTagsToPhotoInGraph(uGraph, p.getId());
        }
        m.endConversion();
        m.startOperation();
        // the target operation to execute
        result = dumpRDFGraphToStr(doConstructQuery(sparqlQuery, uGraph),
            PSIMeter.graphOutFormat);
        m.endOperation();
        m.endOverallOperation();

        if(PSIMeter.doInfo) {
                System.out.println(m);
                System.out.println(dumpRDFGraphToStr(uGraph,
                    PSIMeter.graphOutFormat)); // dump merged RDF graph
                System.out.println(result);
        }
        this.mMap.put("B", m);
        return result;
}


/**
 * Uses the flickr API to retrieve all public photos of user, then, for
    each photo
 * an external
    service(http://www.kanzaki.com/works/2005/imgdsc/flickr2rdf) is
    invoked to
 * generate the RDF graph (incl. the addition of the tags to each photo,
    localy).
 *
 * @param userID
 * @param sparqlQuery
 */
public String executeApproachC(String userID, String sparqlQuery) {
        PhotoList pl = null;
        Model uGraph = ModelFactory.createDefaultModel();
```

```
        String result = "";
        Metrics m = new Metrics("Approach C", sparqlQuery, uGraph);

        m.startOverallOperation();
        m.startConversion();

        // create list of public photos of a user using the flickrj API
        pl = fp.listPublicPhotosOfUserWithID(userID);

        // add all public photos (including tags) of user to the graph
        for (Iterator iter = pl.iterator(); iter.hasNext();) {
                Photo p = (Photo) iter.next();

                // construct graph by involving an external service
                Model tmpGraph =
                    constructRDFGraph(fp.createFlickr2RDFURL(p.getUrl()));

                // remove dc:title props that have been added by external
                    service
                tmpGraph.removeAll(null, DC.title, null);

                // add tags for each photo to the graph
                this.addTagsToPhotoInGraph(tmpGraph, p.getId());

                uGraph.add(tmpGraph);
        }
        m.endConversion();
        m.startOperation();
        // the target operation to execute
        result = dumpRDFGraphToStr(doConstructQuery(sparqlQuery, uGraph),
            PSIMeter.graphOutFormat);
        m.endOperation();
        m.endOverallOperation();

        if(PSIMeter.doInfo) {
                System.out.println(m);
                System.out.println(dumpRDFGraphToStr(uGraph,
                    PSIMeter.graphOutFormat)); // dump merged RDF graph
                System.out.println(result);
        }

        this.mMap.put("C", m);
        return result;
}



/**
 * Adds the tags of the photo with <code>photoID</code> to the RDF graph.
 * @param rdfGraph
 * @param p
 */
public void addTagsToPhotoInGraph(Model rdfGraph, String photoID) {
        String tmpGraphStr;

        if(PSIMeter.doDebug)
         System.out.println(dumpContentFromURL(
          fp.createFlickrPhotosGetInfoURL(photoID))); // dump flickr XML
              response

```

```
        tmpGraphStr = applyTransformation(fp.createFlickrPhotosGetInfoURL(photoID), this.baseURL + "transform/flickrphototags2rdf.xslt");

        if(PSIMeter.doDebug) System.out.println(tmpGraphStr); // dump tmp RDF graph

        rdfGraph.add(constructRDFGraphFromStr(tmpGraphStr)); // add tmp RDF graph
}



/**
 * Lists all users that use a specified machine tags.
 *
 * @param mtag
 * @return
 */
public Vector listUserWithMTags(String mtag) {
        Vector uList = new Vector();
        Vector result = new Vector();
        String fRetrievePhotosMTagged = "";

        Model uGraph = ModelFactory.createDefaultModel();

        fRetrievePhotosMTagged = fp.createFlickrPhotosSearchURL(mtag);

        if(PSIMeter.doDebug) System.out.println("[CMD] " + fRetrievePhotosMTagged);

        // matched photos (URL and user only!)
        uGraph.add(constructRDFGraphFromStr(
                    applyTransformation(fRetrievePhotosMTagged,
                    this.baseURL + "transform/flickrphoto2rdf.xslt")));

        if(PSIMeter.doDebug) System.out.println(dumpRDFGraphToStr(uGraph, PSIMeter.graphOutFormat));

        uList = createUserList(uGraph);

        // create user info from list of user objects
        for (int i = 0; i < uList.size(); i++) {
                User u = (User)uList.get(i);
                String uName = u.getRealName();
                if(uName == null){
                        result.add("<a href=\"http://www.flickr.com/photos/" + u.getId() + "\">" + u.getId() + "</a>");
                }
                else{
                        result.add("<a href=\"http://www.flickr.com/photos/" + u.getId() + "\">" + u.getId()  + "</a> (" + uName + ")");
                }
        }

        return result;
```

```java
}



/**
 * Creates a list of flickr user objects from the RDF graph.
 *
 * @param rdfGraph
 * @return A list of <code>com.aetrion.flickr.people.User</code> objects.
 */
public Vector createUserList(Model rdfGraph) {
        Vector userList = new Vector();
        Query query = QueryFactory.create(PSIMeter.SELECT_USER_ID);
        QueryExecution qe = QueryExecutionFactory.create(query, rdfGraph);
        ResultSet result = qe.execSelect();

        for(Iterator qIter = result; qIter.hasNext();) {
                QuerySolution res = (QuerySolution)qIter.next();
                String uID = res.getLiteral("uID").getLexicalForm();
                if(!userList.contains(uID)){
                        userList.add(this.fp.getUserWithID(uID));
                }
        }
        qe.close();

        return userList;
}

public Vector createPhotoIDList(Model rdfGraph) {
        Vector pIDs = new Vector();
        Query query = QueryFactory.create(PSIMeter.SELECT_PHOTO_ID);
        QueryExecution qe = QueryExecutionFactory.create(query, rdfGraph);
        ResultSet result = qe.execSelect();

        for(Iterator qIter = result; qIter.hasNext();) {
                QuerySolution res = (QuerySolution)qIter.next();
                pIDs.add(res.getLiteral("pID").getLexicalForm());
        }
        qe.close();

        return pIDs;
}

public String dumpContentFromURL(String URLStr) {
        String eURL;
        URL srcUrl = null;
        BufferedReader in = null;
        String inputLine;
        StringWriter sw = new StringWriter();
        try {
                eURL = URLEncoder.encode(URLStr,"UTF-8");
                System.out.println("Trying to read from: " + eURL);
                srcUrl = new URL(URLStr);
                in = new BufferedReader(new
                    InputStreamReader(srcUrl.openStream()));
                while ((inputLine = in.readLine()) != null){
                    System.out.println(inputLine);
                    sw.write(inputLine);
                }
                in.close();
```

```
                sw.close();

        } catch (MalformedURLException murle) {
                murle.printStackTrace();
        } catch (IOException ioe) {
                ioe.printStackTrace();
        }
        return sw.toString();
}

private Model doConstructQuery(String queryString, Model rdfGraph){
        Query query = QueryFactory.create(queryString);

        // execute the query and obtain results
        QueryExecution qe = QueryExecutionFactory.create(query, rdfGraph);
        Model resultGraph = qe.execConstruct();

        qe.close();

        return resultGraph;
}


private String applyTransformation(String xmlDocURL, String xsltFileName){
        DocumentBuilder builder;
        Transformer transformer;
        StringWriter sw = new StringWriter();
        try {
   builder = DocumentBuilderFactory.newInstance().newDocumentBuilder();
   transformer = TransformerFactory.newInstance().newTransformer(
                 new StreamSource(new File(xsltFileName)));
         transformer.transform(new DOMSource(builder.parse(xmlDocURL)),
                               new StreamResult(sw));
        }
        catch (TransformerConfigurationException tce) {
                tce.printStackTrace();
        }
        catch (TransformerFactoryConfigurationError tfe) {
                tfe.printStackTrace();
        }
        catch (ParserConfigurationException pce) {
                pce.printStackTrace();
        }
        catch (TransformerException te) {
                te.printStackTrace();
        }
        catch (SAXException saxe) {
                saxe.printStackTrace();
        }
        catch (IOException ioe) {
                ioe.printStackTrace();
        }
        return sw.toString();
}


public Model constructRDFGraph(String rdfGraphURL){
         return FileManager.get().loadModel(rdfGraphURL);
}

```

```java
public Model constructRDFGraphFromStr(String rdfGraphStr) {
        Model rdfGraph = ModelFactory.createDefaultModel();
    StringReader sr = new StringReader(rdfGraphStr);
    ByteArrayInputStream bais = new
        ByteArrayInputStream(rdfGraphStr.getBytes());


    rdfGraph.read(bais, "");
    sr.close();
    return rdfGraph;
}

public String dumpRDFGraphToStr(Model rdfGraph, String outFormat){
        StringWriter sw = new StringWriter();
        rdfGraph.setNsPrefix("rdf",
            "http://www.w3.org/1999/02/22-rdf-syntax-ns#");
        rdfGraph.setNsPrefix("rdfs",
            "http://www.w3.org/2000/01/rdf-schema#");
        rdfGraph.setNsPrefix("dc", "http://purl.org/dc/elements/1.1/");
        rdfGraph.setNsPrefix("dcterms", "http://purl.org/dc/terms/");
        rdfGraph.setNsPrefix("foaf", "http://xmlns.com/foaf/0.1/");
        rdfGraph.setNsPrefix("geo",
            "http://www.w3.org/2003/01/geo/wgs84_pos#");
        rdfGraph.setNsPrefix("xsd", "http://www.w3.org/2001/XMLSchema#");
        rdfGraph.setNsPrefix("flickr", "http://flickr.com/tags/meta#");
        rdfGraph.write(sw, outFormat);
    return sw.toString();
}

public String userEMail2ID(String userEMailAddress){
        return fp.getUserWithEMail(userEMailAddress).getId();
}

public int getNumOfPublicPhotosOfUser(String userEMailAddress){
        return fp.listPublicPhotosOfUserWithID(
                  fp.getUserWithEMail(userEMailAddress).getId()).size();
}

public Metrics getMetric(String mID){
        return (Metrics) this.mMap.get(mID);
}

public HashMap<String,Metrics> getMetricsMap(){
        return this.mMap;
}

}
```

Listing A.3: Java source code of the PSIMeter application.

## A.3 Diagrams

### A.3.1 Media Semantics Mapping

On the following two pages, details about the NM2 core ontology produced as an output of the Media Semantics Mapping, as discussed in chapter 9 are depicted.

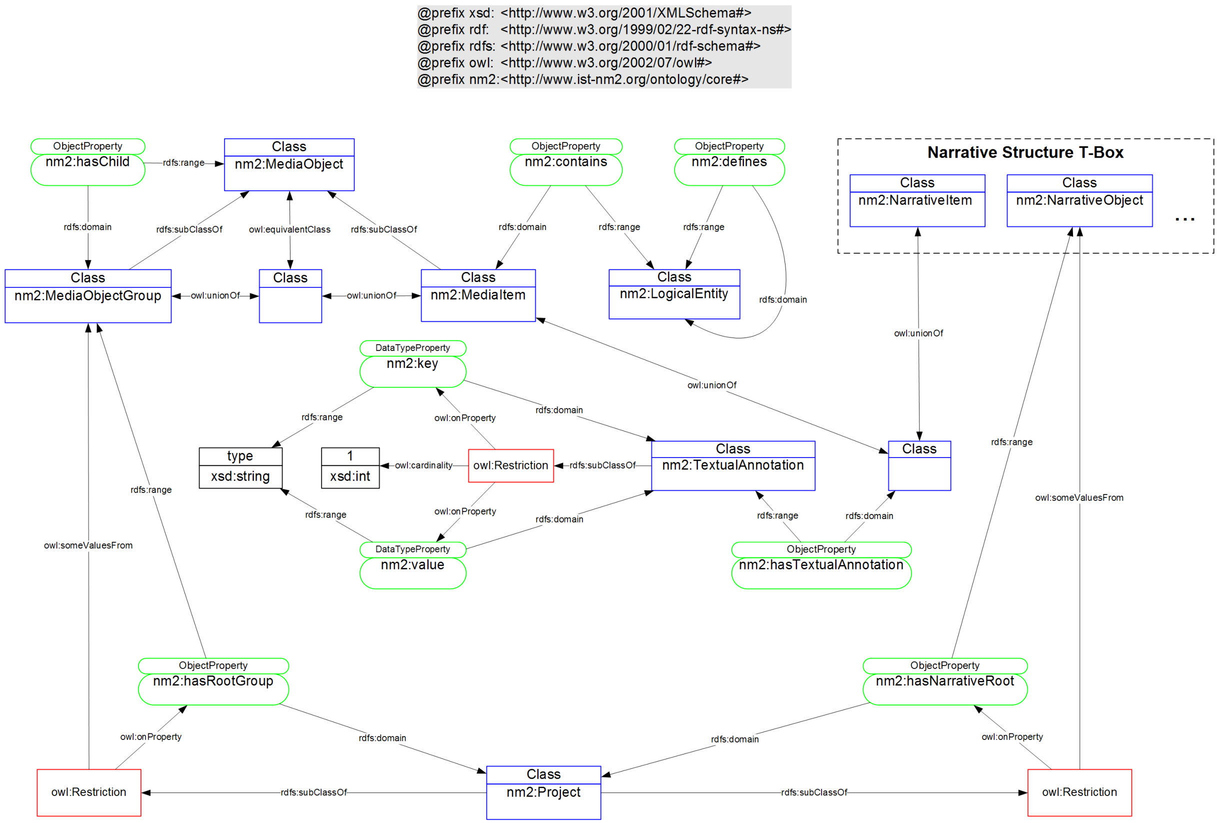



Figure A.1: NM2 core ontology.

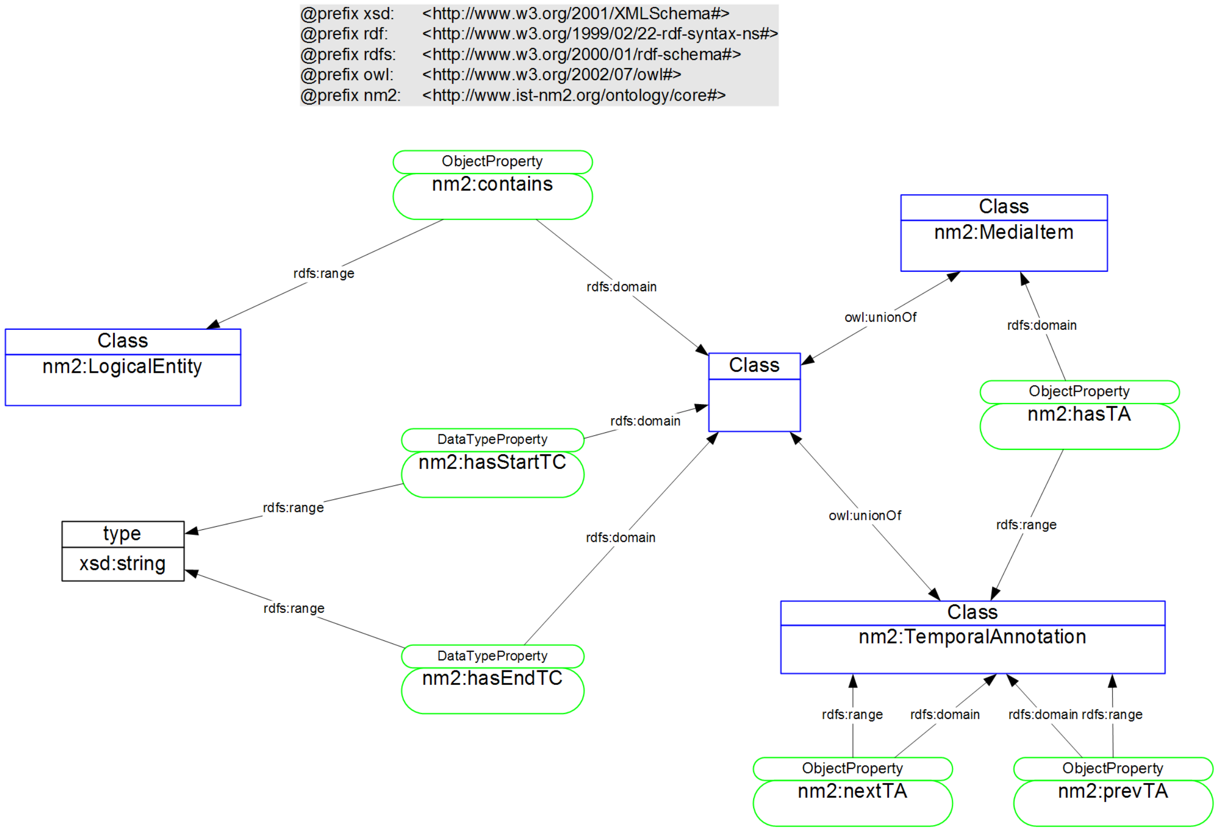



Figure A.2: Temporal Annotations in the NM2 core ontology.

Appendix B

# Author's Contribution

## B.1 Publications

This work is by and large based on research papers that have been published at diverse conferences, workshops, and journals between 2006 and 2008. Beside this academic publishing, the author of this thesis has been active in various projects and standardisations activities in the realm of media semantics. The core of this doctoral thesis is on six peer-reviewed papers:

- Chapter 6 is based on a paper the author has written together with Tobias Bürger: "Why Real-World Multimedia Assets Fail to Enter the Semantic Web" [48]; the author has presented the paper at the Semantic Authoring, Annotation and Knowledge Markup (SAAKM07) Workshop in Whistler, Canada in October 2007;

- Chapter 8 is based on a paper entitled "A Performance and Scalability Metric for Virtual RDF Graphs" [149] written together with Wolfgang Slany and Danny Ayers. The author has presented the paper at the $3^{rd}$ Workshop on Scripting for the Semantic Web (SFSW07) in Innsbruck, Austria, in June 2007;

- In chapter 9 the paper "Applying Media Semantics Mapping in a Non-linear, Interactive Movie Production Environment" [134] of the author has been used as a base. It was presented at the $1^{st}$ International Conference on New Media Technology (I-Media07) in Graz, Austria, in September 2007. Further, an article in a special issue of the ACM Multimedia Systems Journal [135] (to appear 2008) has been incorporated;

- Chapter 10 is based on "Deploying Multimedia Metadata in Cultural Heritage on the Semantic Web" [140], a paper written with Werner Bailer and Harald Mayer, presented at the First International Workshop on Cultural Heritage on the Semantic Web, collocated with the $6^{th}$ International Semantic Web Conference (ISWC07) in November 2007. Further, our short paper [16] at the First Workshop on Semantic Interoperability in the European Digital Library (SIEDL), collocated with the European Semantic Web Conference (ESWC) 2008 has been integrated.

Further works, such as a journal article, and technical reports have been incorporated into this thesis. These are:

- The section 3.3 is taken from the W3C Incubator Group Report (XGR) "Multimedia Vocabularies on the Semantic Web" [141], an XGR with contributions from and main editing responsibility by the author of this thesis;
- Certain parts of chapter 5 stem from an invited feature article with Werner Haas and Werner Bailer: "Media Semantics" [123];
- In chapter 9, further aspects of the non-linear, interactive movie production environment (NM2) have been discussed in [147], an IEEE MultiMedia article published together with Frank Nack.

## B.2 Projects

The research leading to these publications was carried out by and large in three projects. Without any special order these projects are: The "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content" (K-Space) project[1], partially funded under the 6th Framework Programme of the European Commission, the "New Media for a New Millennium" (NM2) project[2] (an Integrated Project under the European Unions 6th Framework Programme in the thematic priority of Information Society Technology), and the "Understanding Advertising" (UAd) project[3], a national project funded by the Austrian FIT-IT Programme. In the following, the above projects are discussed in greater detail.

### B.2.1 Media Production

The "New media for a New millennium" (NM2) project, an EU-FP6 Integrated Project running from 2004 to 2007 focused on creating tools for non-linear, interactive story authoring. The author of this thesis was heavily involved in ontology engineering design issues, as well as in the management of this project at JOANNEUM RESEARCH[4]. Aspects of the non-linear, interactive movie production environment NM2 with author's contribution were discussed in [256; 147; 304; 134].

We have further elaborated on future applications of the NM2 approach in [181]; there, we have discussed the metadata-based adaptive assembly of video clips on the Web; this publications was accepted in the main track of the $2^{nd}$ International Workshop on Semantic Media Adaptation and Personalization (SMAP 07)[5].

### B.2.2 Media Analysis

In early 2007 we started, based on results from the MediaCampaign project[6], to develop the foundations of the Understanding Advertising (UAd) project[7]. In the UAd project the focus is put on the so called semantic media analysis framework (SEMAF). The SEMAF is capable

[1] http://kspace.qmul.net/
[2] http://www.ist-nm2.org/
[3] http://www.sembase.at/index.php/UAd
[4] http://www.joanneum.at/iis
[5] http://www.smap2007.org/
[6] http://www.media-campaign.eu/
[7] http://www.sembase.at/index.php/UAd

of taking as input low level feature metadata received from visual analysis of advertisements and public knowledge. The public knowledge may stem from diverse sources such as a company's website, stock market information, statistical data or discussion forums. SEMAF will be a generic framework of modules with open interfaces which can be tailored easily by instantiation to specific domains of customer queries, e.g., trends for advertising expenditure in automotive sector as domain. Within the UAd application area specific ontologies and rules will be developed and applied together with the public knowledge in order to relate and homogenize information from heterogeneous sources and generate an answer to the user query. First results regarding the mining of opinions from discussion forums on the Web have been reported at the Social Data on the Web Workshop at the International Semantic Web Conference 2008 [277].

### B.2.3 Other Activities

An overview of the current state-of-the-art and some approaches how to deal with the Semantic Gap has been given in [123]. A state-of-the-art survey of content analysis tools (for video, audio and speech) has been performed, and has been made available in the realm of the PrestoSpace project [17].

In the "E-Learning And Rights Managements" eLARM project [202] we proposed a way to handle the access to learning objects by utilising an MPEG-21 based DRM ontology. Applying Semantic Web technologies in the realm of Geographical Information Systems (GIS) to enable innovative LBS has been discussed in [335].

Recently, we have elaborated on RDF-graph visualisation and debugging techniques [323]. The author has been at the Semantic Web Service Workshop [146] in 2005 to represent the position of JOANNEUM RESEARCH.

## B.3 Academic Activities

The author of this thesis has been active in the academic area as well. The following is a chronological list of Programme Committee memberships and reviewing activities:

- 1st Social Data on the Web workshop SDoW2008 (2008)
- Doctoral consortium at International Semantic Web Conference DC ISWC08 (2008)
- First International Workshop on Story-Telling and Educational Games STEG08 (2008)
- Semantic Authoring, Annotation and Knowledge Markup Workshop SAAKM08 (2008)
- 3rd International Conference on Semantic and Digital Media Technologies SAMT (2008)
- 4th Workshop on Scripting for the Semantic Web SFSW (2008)
- Linked Data on the Web Workshop LDOW (2008)
- European Interactive TV Conference EuroITV (2008)
- Semantic Web User Interaction at CHI SWUI (2008)
- Workshop on Multimedia Metadata Management & Retrieval MR3 (2008)

- 1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies MAReSO (2007)
- Knowledge Acquisition from Multimedia Content Workshop KAMC (2007)
- Multimedia Metadata Applications M3A (2007)

Further, the author has been active as a reviewer for journals:

- International Journal on Semantic Web and Information Systems
- IEEE Multimedia
- Multimedia Tools and Applications Journal—Special Issue on "Semantic Multimedia"

Finally, the author has two tutorials at the International Semantic Web Conference (ISWC) 2008[8], one concerning RDFa and one on linked data.

## B.4 Activities

This section lists activities, the author of this thesis is or has been active in; most if not all of them directly influenced the final shape of this work.

### B.4.1 W3C participation

#### Multimedia Semantics Incubator Group (MMSEM-XG)

The mission of the Multimedia Semantics Incubator Group (MMSEM-XG) [315]—in which the author has been active—is as follows:

- Show how metadata interoperability can be achieved by using the Semantic Web technologies to integrate existing multimedia metadata standards. Thus, the goal of the XG is NOT to invent new multimedia metadata formats, but to leverage and combine existing approaches.
- Show, in addition to the interoperability advantages, the added value of the formal semantics provided by the Semantic Web. The XG will describe practical applications and services that provide extra functionality by using, for example, subsumption reasoning or rule-based approaches. These applications could be multi-platform, i.e. adapted to any device that accesses the web.
- Provide best practices for annotating and using multimedia content on the Web, based on practical use cases that identify the users, the type of content and the type of metadata that they want to provide.

The author of this thesis has acted as the editor of the "Multimedia Vocabularies on the Semantic Web" Incubator Report [141]. Further, an overview of formal representations of MPEG-7 [141][9] has been given.

[8] http://iswc2008.semanticweb.org/program/tutorials/

[9] cf. http://www.w3.org/2005/Incubator/mmsem/wiki/Vocabularies#f_MPEG-7, with contributions of the author of this thesis

#### RDF in XHTML Taskforce (RDFa-TF)

The RDF in XHTML Taskforce [175] is chartered to:

- State requirements for representing metadata in RDF within an XHTML document
- Evaluate proposed solutions against those requirements.
- Recommend to the HTML WG and the SWBPD WG how to proceed to achieve a common, widely accepted way of representing RDF metadata within an XHMTL document.

In the realm of metadata deployment we have sketched the use cases of RDFa [5]. The author of this work is responsible for the RDFa Test Cases, along with the RDFa Test Suite[10]. Finally, the author takes care of the RDFa Implementation Report [137].

#### SWEO-IG: Linking Open Data

In the realm of the SWEO project "Linking Open Data" we have RDFised and interlinked the EuroStat data set[11], resulting in some 4.000.000.000 RDF triples. Further, we have proposed a new way of interlinking datasets, the so called User Contributed Interlinking [125; 144]. Finally, we currently work on describing linked dataset, resulting in the voiD (Vocabulary of Interlinked Datasets) vocabulary[12].

### B.4.2 ramm.x initiative

The author of this doctoral thesis has launched the RDFa-deployed Multimedia Metadata (ramm.x) initiative[13] together with Werner Bailer (as well JOANNEUM RESEARCH). This initiative can be seen as an outcome of the experience gathered in the media semantics project of the past years. The main goal is to produce the ramm.x specification [138], including a reference implementation and use cases.

In a perfect world, where everyone uses RDF to represent metadata on the Web, ramm.x would certainly be of no big use. As a matter of fact, we—at least from the Semantic Web point-of-view—do not live in a perfect world. There exists an array of existing multimedia metadata formats that have been used for years in diverse applications. However, when one is after using these formats in the context of the Semantic Web, the options are limited. What is basically missing is a framework that allows existing multimedia metadata hooking into the Semantic Web. With ramm.x we aim at enabling existing multimedia metadata formats to enter the Semantic Web. ramm.x targets at self-descriptive media asset descriptions allowing to apply the follow-your-nose principle. With ramm.x, we focus on the deployment of multimedia metadata. Pretty much as communications provider, we focus on the so called last mile—in our case the consumption and the processing of metadata.

[10] http://www.w3.org/2006/07/SWD/RDFa/testsuite/
[11] http://riese.joanneum.at
[12] http://community.linkeddata.org/MediaWiki/index.php?VoiD
[13] http://sw.joanneum.at/rammx

### B.4.3 Related to MPEG-7

In the realm of the K-Space project[14] our work on semantic validation of MPEG-7 profiles has been presented at the SAMT 2006[15] [291]. Further the "Semantics of Temporal Media Content Descriptions" [156] have been reported at the M3A workshop in 2007. In [19] we presented a case study of establishing a description infrastructure for an audiovisual content-analysis and retrieval system. The description infrastructure consists of an internal (MPEG-7-based) metadata model and access tool for using it.

[14] http://www.k-space.eu/

[15] http://www.samt2006.org/

Appendix C

# Reference Material

## C.1 Multimedia Ontologies

Based on [141; 89], an overview on existing multimedia ontologies is given.

### C.1.1 aceMedia Visual Descriptor Ontology

| | |
|---|---|
| **Responsibility** | http://www.acemedia.org/ |
| **Description** | [40] |

The Visual Descriptor Ontology (VDO) developed within the aceMedia project for semantic multimedia content analysis and reasoning, contains representations of MPEG-7 visual descriptors and models Concepts and Properties that describe visual characteristics of objects. The term descriptor refers to a specific representation of a visual feature (color, shape, texture etc) that defines the syntax and the semantics of a specific aspect of the feature. For example, the dominant color descriptor specifies among others, the number and value of dominant colors that are present in a region of interest and the percentage of pixels that each associated color value has. Although the construction of the VDO is tightly coupled with the specification of the MPEG-7 Visual Part, several modifications were carried out in order to adapt to the XML Schema provided by MPEG-7 to an ontology and the data type representations available in RDF Schema.

### C.1.2 Mindswap Image Region Ontology

| | |
|---|---|
| **Responsibility** | http://www.mindswap.org/ |
| **Description** | [124] |

### C.1.3 Music Ontology Specification

| | |
|---|---|
| **Responsibility** | Frederick Giasson and Yves Raimond |
| **Description** | http://musicontology.com/ |

The Music Ontology Specification provides main concepts and properties for describing music (i.e., artists, albums and tracks) on the Semantic Web. It was based on the MusicBrainz editorial metadata; cf. Section 3.3.2.

### C.1.4 Kanzaki Audio Ontology

| | |
|---|---|
| **Responsibility** | http://www.kanzaki.com/ |
| **Description** | http://www.kanzaki.com/ns/music |

A vocabulary to describe classical music and performances. Classes (categories) for musical work, event, instrument and performers, as well as related properties are defined.

### C.1.5 Core Ontology for Multimedia (COMM)

| | |
|---|---|
| **Responsibility** | http://multimedia.semanticweb.org/COMM/ |
| **Description** | [10] |

The Core Ontology for Multimedia (COMM) is based on both the MPEG-7 standard and the DOLCE[1] foundational ontology. COMM is an OWL DL ontology. It is composed of multimedia patterns specializing the DOLCE design patterns for Descriptions & Situations and Information Objects. The ontology covers a very large part of the MPEG-7 standard. The explicit representation of algorithms in the multimedia patterns allows also to describe the multimedia analysis steps, something that is not possible in MPEG-7.

## C.2 Multimedia Annotation Tools

Table C.1 on page 221 gives an overview on existing and available multimedia annotation tools.

[1] http://www.loa-cnr.it/DOLCE.html

| **Name** (Institution) | **Source** |
|---|---|
| PhotoStuff (Mindswap) | http://www.mindswap.org/2003/PhotoStuff/ |
| M-OntoMat-Annotizer (CERTH-ITI) | http://mkg.iti.gr/publications/eswc06.pdf |
| CONFOTO (private) | http://www.confoto.org/home |
| VideoAnnex (IBM) | http://www.alphawoks.ibm.com/tech/videoannex |
| VIDETO (ZGDV) | http://www.zgdv.de/zgdv/zgdv/departments/zr4/Produkte/videto/ |
| Vannotea (Univ. of Queensland) | http://www.itee.uq.edu.au/~eresearch/projects/vannotea/ |
| ViPER-GT (Univ. of Maryland) | http://viper-toolkit.sourceforge.net/docs/gt/ |
| Ricoh MovieTool (Ricoh) | http://www.ricoh.co.jp/src/multimedia/MovieTool/ |
| Vizard (JOANNEUM RESEARCH) | http://www.joanneum.at/index.php?id=376&L=1 |
| AKTive Media (Univ. of Sheffield) | http://www.dcs.shef.ac.uk/~ajay/html/cresearch.html |
| MuViNo (Univ. of Klagenfurt) | http://vitooki.sourceforge.net/components/muvino/code/ |
| ProjectPad2 (NorthWestern University) | http://projectpad.northwestern.edu/ppad2/ |
| Ontolog (NTNU) | http://www.idi.ntnu.no/~heggland/ontolog/ |
| Elan (Max Planck Institute) | http://www.mpi.nl/tools/elan.html |
| Transcriber | http://trans.sourceforge.net/en/presentation.php |
| Mdefi (INRIA) | http://opera.inrialpes.fr/VideoMadeus.html |
| OPALES (INA) | http://www.ina.fr/recherche/projets/encours/opales.fr.html |
| CMWeb (CSIRO) | http://www.ict.csiro.au/page.php?did=71 |
| Advene (LIRIS, Univ. of Lyon) | http://liris.cnrs.fr/advene/ |
| Schema rs (Schema Project) | http://media.iti.gr/SchemaRS |
| Photocopain (Univ. of Southampton) | http://eprints.ecs.soton.ac.uk/12004/ |
| LabelMe (MIT) | http://labelme.csail.mit.edu/ |

Table C.1: An overview of Multimedia Annotation Tools.

# Bibliography


[1] AAF Association. http://www.aafassociation.org, 2005.

[2] B. Adida and M. Birbeck. RDFa Primer 1.0—Embedding RDF in XHTML. Editor's draft, W3C Semantic Web Deployment Working Group, 2007.

[3] B. Adida, M. Birbeck, S. McCarron, and S. Pemberton. RDFa in XHTML: Syntax. W3c editor's draft 21 september 2007, W3C Semantic Web Deployment WG and XHTML 2 WG, 2007.

[4] B. Adida, M. Birbeck, S. McCarron, and S. Pemberton. RDFa in XHTML: Syntax and Processing. W3C Working Draft 18 October 2007, W3C Semantic Web Deployment Working Group, 2007.

[5] B. Adida and M. Hausenblas. RDFa Use Cases: Scenarios for Embedding RDF in HTML. W3C Working Draft, W3C Semantic Web Deployment Working Group, 2007.

[6] H. S. Al-Khalifa and H. C. Davis. The evolution of metadata from standards to semantics in e-learning applications. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 69–72, New York, NY, USA, 2006. ACM Press. ISBN 1-59593-417-0.

[7] A. Alba, V. Bhagwan, M. Ching, A. Cozzi, R. Desai, D. Gruhl, K. Haas, L. Kato, J. Kusnitz, B. Langston, F. Nagy, L. Nguyen, J. Pieper, S. Srinivasan, A. Stuart, and R. Tang. A funny thing happened on the way to a billion ... In *Bulletin of the Technical Committee on Data Engineering*, pages 27–36. IEEE Computer Society, 2006.

[8] J. F. Allen and G. Ferguson. Actions and Events in Interval Temporal Logic. Technical Report TR521, University of Rochester, 1994.

[9] K. M. Anderson. Data scalability in open hypermedia systems. In *HYPERTEXT '99: Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots*, pages 27–36. ACM Press, 1999. ISBN 1-58113-064-3.

[10] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *Proc. of the $6^{th}$ International Semantic Web Conference (ISWC'2007), Busan, Korea, November 11-15, 2007*, 2007.

[11] T. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis, and Y. Kompatsiaris. Using a Multimedia Ontology Infrastructure for Semantic

Annotation of Multimedia Content. In *Proc. of $5^{th}$ International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05), Galway, Ireland, November 2005*, 2005.

[12] O. Aubert, P.-A. Champin, and Y. Prie. Integration of Semantic Web technology in an annotation-based Hypervideo System. Technical Report RR-LIRIS-2006-004, LIRIS Research Report, 2006.

[13] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003. ISBN 0-521-78176-0.

[14] F. Baader, B. Ganter, B. Sertkaya, and U. Sattler. Completing Description Logic Knowledge Bases Using Formal Concept Analysis. In *Proc. of the Third International Workshop OWL: Experiences and Directions (OWLED 2007)*, pages 230–235, 2007.

[15] F. Baader and W. Nutt. Basic description logics. In F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors, *The Description Logic Handbook*, chapter 2, pages 47–100. Cambridge University Press, 2003.

[16] W. Bailer, M. Hausenblas, and W. Haas. Enabling Audiovisual Metadata Interoperability with the European Digital Library. In *First Workshop on Semantic Interoperability in the European Digital Library (SIEDL 08)*, Tenerife, Spain, 2008.

[17] W. Bailer, F. Hoeller, A. Messina, D. Airola, P. Schallauer, and M. Hausenblas. State of the Art of Content Analysis Tools for Video, Audio and Speech. Eu-project deliverable, JOANNEUM RESEARCH Forschungsgesellschaft gmbH, 2005.

[18] W. Bailer and P. Schallauer. The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems. In *Proceedings of $12^{th}$ International Multi-Media Modeling Conference*, Beijing, China, 2006.

[19] W. Bailer, P. Schallauer, M. Hausenblas, and G. Thallinger. MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System. In *Proceedings of SPIE - Storage and Retrieval Methods and Applications for Multimedia*, volume 5682, pages 284–295, 2005. ISBN 0-7695-2084-7.

[20] C. Baral and M. Gelfond. Logic Programming and Knowledge Representation. *Journal of Logic-Programming*, 19–20:73–148, 1994.

[21] E. Batlle, H. Neuschmied, P. Uray, and G. Ackermann. Recognition and analysis of audio for copyright protection: The RAA project. *JASIST*, 55(12):1084–1091, 2004.

[22] S. Baumann and A. Kluter. Super-convenience for non-musicians: Querying MP3 and the semantic web. In *Proceedings of the Third International Conference on Music Information Retrieval: ISMIR 2002*, pages 297–298, Paris, France, 2002. IRCAM - Centre Pompidou.

[23] R. Baumgartner, G. Gottlob, M. Herzog, and W. Slany. Annotating the Legacy Web with Lixto. In *Proc. $3^{rd}$ International Semantic Web Conference, ISWC2004, Hiroshima, Japan*, 2004.

[24] BBC. Standard Media Exchange Framework (SMEF) Data Model 1.5. http://www.bbc.co.uk/guidelines/smef, 2000.

[25] D. Beckett. Deliverable 10.1: Scalability and storage: Survey of free software/open source rdf storage systems. http://www.w3.org/2001/sw/Europe/reports/rdf_scalable_storage_report/, 2002.

[26] D. Beckett. Modernising Semantic Web Markup. In *Proc. of the XML Europe 2004*, page N.A., 2004.

[27] D. Beckett and B. McBride. RDF/XML Syntax Specification (Revised). W3c recommendation, World Wide Web Consortium, 2004.

[28] D. Beckett, E. Miller, and D. Brickley. Expressing Simple Dublin Core in RDF/XML (DCMI Recommendation). http://dublincore.org/documents/dcmes-xml, 2002.

[29] A. Bernaras, I. Laresgoiti, and J. M. Corera. Building and reusing ontologies for electrical network applications. In *$12^t h$ European Conference on Artificial Intelligence, Budapest, Hungary, August 11-16, 1996, Proceedings*, pages 298–302. John Wiley and Sons, Chichester, 1996.

[30] T. Berners-Lee. Metadata Architecture. http://www.w3.org/DesignIssues/Metadata.html, 1997.

[31] T. Berners-Lee. Semantic Web roadmap. http://www.w3.org/DesignIssues/Semantic.html, 1998.

[32] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret. The World-Wide Web. *Communications of the ACM*, 37(8):76–82, 1994.

[33] T. Berners-Lee, R. Fielding, U. Irvine, and L. Masinter. Uniform Resource Identifiers (URI): Generic Syntax. RFC 2396, IETF—Network Working Group, 1998.

[34] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web, 2001. Available at http://www.sciam.com/2001/0501issue/0501berners-lee.html.

[35] T. Berners-Lee, L. Masinter, and M. McCahill. Uniform Resource Locators (URL). RFC 1738, IETF—Network Working Group, 1994.

[36] C. Bizer, R. Cyganiak, and T. Heath. How to Publish Linked Data on the Web. http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/, 2007.

[37] C. Bizer, T. Heath, D. Ayers, and Y. Raimond. Interlinking Open Data on the Web (Poster). In *4th European Semantic Web Conference (ESWC2007)*, pages 802–815, 2007.

[38] C. Bizer and R. Oldakowski. Using context- and content-based trust policies on the Semantic Web. In *Proceedings of the $13^{th}$ international World Wide Web conference on Alternate track papers & posters*, pages 228–229. ACM Press, 2004.

[39] C. Bizer and D. Westphal. Developers Guide to Semantic Web Toolkits for different Programming Languages. http://www.wiwiss.fu-berlin.de/suhl/bizer/toolkits/, 2005.

[40] S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, S. Handschuh, Y. Kompatsiaris, S. Staab, and M. G. Strintzis. Semantic Annotation of Images and Videos for Multimedia Analysis. In *The Semantic Web: Research and Applications: Proceedings of the Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29-June 1, 2005*, volume 3532 of *Lecture Notes in Computer Science*, pages 592–607. Springer, 2005.

[41] U. Bojars, J. Breslin, and A. Passant. SIOC Ontology: Applications and Implementation Status. W3C Member Submission 12 June 2007, W3C Member Submission, 2007.

[42] U. Bojars and J. G. Breslin. SIOC Core Ontology Specification. http://rdfs.org/sioc/spec/, 2007.

[43] H. Boley and M. Kifer. RIF Core Design. W3c working draft 30 march 2007, W3C Rule Interchange Format (RIF) Working Group, 2007.

[44] A. B. Bondi. Characteristics of scalability and their impact on performance. In *WOSP '00: Proceedings of the 2$^{nd}$ international workshop on Software and performance*, pages 195–203, New York, NY, USA, 2000. ACM Press.

[45] T. Bray, D. Hollander, A. Layman, and R. Tobin. Namespaces in XML 1.1 (Second Edition). W3c recommendation 16 august 2006, World Wide Web Consortium, 2006.

[46] D. Brickley and L. Miller. FOAF Vocabulary Specification. http://xmlns.com/foaf/0.1/, 2004.

[47] D. Bulterman and L. Hardman. Structured multimedia authoring. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 1(1):89–109, 2005.

[48] T. Bürger and M. Hausenblas. Why Real-World Multimedia Assets Fail to Enter the Semantic Web. In *International Workshop on Semantic Authoring, Annotation and Knowledge Markup (SAAKM07)*, Whistler, Canada, 2007.

[49] T. Bürger and R. Westenthaler. Mind the Gap - Requirements for the Combination of Content and Knowledge. In *Poster Proceedings of the SAMT 2006 Conference, Athens, Greece*, 2006.

[50] V. Bush. As We May Think. *The Atlantic Monthly*, 176(1):101–108, 1945.

[51] M. Cai and M. Frank. RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network. In *WWW '04: Proc. of the 13$^{th}$ International Conference on World Wide Web*, pages 650–657, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-844-X.

[52] J. Cardoso and A. P. Sheth. *Semantic Web Services, Processes and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387302395.

[53] W. Cathro. Metadata: an overview. http://www.nla.gov.au/nla/staffpaper/cathro3.html, 1997.

[54] W. Chiu. Design for scalability - an update. http://www-128.ibm.com/developerworks/websphere/library/techarticles/hipods/scalability.html, 2001.

[55] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web Services Description Language (WSDL) 1.1. http://www.w3.org/TR/wsdl, March 2001.

[56] J. Clark. RELAX NG Specification. Committee Specification 3 December 2001, OASIS, 2001.

[57] K. G. Clark. RDF Data Access Use Cases and Requirements. W3c working draft 25 march 2005, W3C RDF Data Access Working Group, 2007.

[58] K. L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Databases*, pages 293–322. Plenum Press, 1978.

[59] E. Compatangelo, W. Vasconcelos, and B. Scharlau. The Ontology Versioning Manifold at its genesis: a distributed blackboard architecture for reasoning with and about ontology versions. Technical Report Technical Report AUCS/TR0404, Dept. of Computing Science, Univ. of Aberdeen, UK, 2004.

[60] D. Connolly. Web naming and addressing overview (URIs, URLs, ...). http://www.w3.org/Addressing/, 2006.

[61] D. Connolly. Gleaning Resource Descriptions from Dialects of Languages (GRDDL). http://www.w3.org/TR/2007/REC-grddl-20070911/, 2007.

[62] J. Cook, M. Hausenblas, G. Mittendorfer, V. Solachidis, M. Thomas, M. Ursu, and L. Xu. D5.1: Detailed Specifications of Understanding Visual Content. Deliverable to EC (permission required), NM2 consortium, 2004.

[63] J. Crowcroft, M. Handley, and I. Wakeman. *Internetworking Multimedia*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999. ISBN 1558605843.

[64] M. V. Cundiff. METS Profile for Historical Newspapers (Draft). http://www.loc.gov/standards/mets/test/ndnp/00000010.html, 2006.

[65] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, and G. Tummarello. Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web. In *The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain*, pages 690–704, 2008.

[66] M. Davis. *Media streams: representing video for retrieval and repurposing*. PhD thesis, Massachusetts Institute of Technology, 1995.

[67] F. Dawson and T. Howes. vCard MIME Directory Profile. RFC 2426, IETF—Network Working Group, 1998.

[68] F. Dawson and D. Stenerson. Internet Calendaring and Scheduling Core Object Specification (iCalendar). RFC 2445, IETF—Network Working Group, 1998.

[69] DCMI. Information and documentation—The Dublin Core metadata element set. ISO 15836, 2003.

[70] Dublin Core Metadata Initiative. http://purl.org/dc, 1999.

[71] J. de Bruijn and A. P. (eds.). OWL$^-$. WSML Deliverable D20.1v0.2 WSML Working Draft 05-15-2005, http://www.wsmo.org/TR/d20/d20.1/v0.2/, 2005.

[72] J. de Bruijn, H. Lausen, A. Polleres, and D. Fensel. The Web Service Modeling Language WSML: An Overview. In *ESWC*, pages 590–604, 2006.

[73] J. de Bruijn (ed.). OWL Flight. D20.3v0.1 OWL Flight WSML Working Draft 23-08-2004, http://www.wsmo.org/2004/d20/d20.3/v0.1/, 2004.

[74] M. Denny. Ontology Tools Survey, Revisited. http://www.xml.com/pub/a/2004/07/14/onto.html, 2004.

[75] DERI. OWL - WSML Translator v1.0. http://tools.deri.org/wsml/owl2wsml-translator/v0.1/, 2007.

[76] DERI. WSML - OWL Translator v1.0. http://tools.deri.org/wsml/wsml2owl-translator/v0.1/, 2007.

[77] DERI. WSML Tools. http://tools.deri.org/wsml/, 2007.

[78] L. Ding, L. Zhou, and T. Finin. Trust based knowledge outsourcing for semantic web agents. In *WI '03: Proceedings of the IEEE/WIC International Conference on Web Intelligence*, pages 379–387. IEEE Computer Society, 2003.

[79] M. Dixon, J. Kalmus, and K. Jeffery. A 5 layer information expressivity model applied to semantic heterogeneity in a decentralised organisation. http://citeseer.ist.psu.edu/dixon95layer.html, 1995.

[80] Implemented description logic-based systems, 2003. Available at http://www.research.att.com/sw/tools/classic/imp-systems.html.

[81] Description Logics Web Page, 2003. Available at http://dl.kr.org.

[82] DMS-1. Material Exchange Format (MXF) – Descriptive Metadata Scheme-1. SMPTE 380M, 2004.

[83] F. M. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. AL-log: Integrating Datalog and Description Logics. *Journal of Intelligent Information Systems*, 10(3):227–252, 1998.

[84] C. Dorai and S. Venkatesh. Bridging the Semantic Gap with Computational Media Aesthetics. *IEEE MultiMedia*, 10(2):15–17, 2003.

[85] J. Dörre, P. Gerstl, and R. Seiffert. Text mining: finding nuggets in mountains of textual data. In *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 398–401. ACM Press, 1999.

[86] L. Duboc, D. S. Rosenblum, and T. Wicks. A framework for modelling and analysis of software systems scalability. In *ICSE '06: Proceeding of the 28$^{th}$ international conference on Software engineering*, pages 949–952, New York, NY, USA, 2006. ACM Press.

[87] M. Duerst and M. Suignard. Internationalized Resource Identifiers (IRIs). RFC 3987, IETF—Network Working Group, 2005.

[88] M. Duerst and M. Suignard. Internationalized Resource Identifiers (IRIs). IETF RFC 3987, 2005. http://www.ietf.org/rfc/rfc3987.txt.

[89] H. Eleftherohorinou, V. Zervaki, A. Gounaris, V. Papastathis, Y. Kompatsiaris, and P. Hobson. Towards a Common Multimedia Ontology Framework. http://www.acemedia.org/aceMedia/files/multimedia_ontology/cfr/MM-Ontologies-Reqs-v1.3.pdf, 2006.

[90] D. C. Engelbart. Augmenting human intellect: A conceptual framework. Technical report, SRI, 1962.

[91] Exif. Digital Still Camera Image File Format Standard (Exchangeable image file format for Digital Still Camera: Exif). JEIDA-49, version 2.1, 1998.

[92] A. Farquhar, R. Fikes, and J. Rice. The Ontolingua server: A tool for collaborative ontology construction. Technical report, Stanford KSL, 1996.

[93] A. Farquhar, R. Fikes, and J. Rice. The ontolingua server: A tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, pages 707–728, 1997.

[94] D. L. Federation. METS – Metadata Encoding and Transmission Standard. http://www.loc.gov/standards/mets, 2006.

[95] M. Felleisen. On the expressive power of programming languages. In *ESOP '90 3rd European Symposium on Programming, Copenhagen, Denmark*, volume 432, pages 134–151, New York, N.Y., 1990. Springer-Verlag.

[96] D. Fensel. Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. http://citeseer.ist.psu.edu/413498.html, 2000.

[97] D. Fensel, C. Bussler, Y. Ding, V. Kartseva, M. Klein, M. Korotkiy, B. Omelayenko, and R. Siebes. Semantic Web Application Areas. http://citeseer.ist.psu.edu/fensel02semantic.html, 2002.

[98] D. Fensel and F. V. Harmelen. Unifying Reasoning and Search to Web Scale. *IEEE Internet Computing*, 11(2):94–96, March/April 2007.

[99] D. Fensel, H. Lausen, A. Polleres, J. de Bruijn, M. Stollberg, D. Roman, and J. Domingue. *Enabling Semantic Web Services: The Web Service Modeling Ontology*. Springer, 11 2006. ISBN 3540345191.

[100] R. T. Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine, 2000.

[101] M. Fitting. Fixpoint semantics for logic programming a survey. *Theoretical Computer Science*, 278(1-2):25–51, 2002.

[102] I. Foster, C. Kesselman, and S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *The International Journal of High Performance Computing Applications*, 15(3):200–222, 2001.

[103] E. Franconi and S. Tessaris. Rules and Queries with Ontologies: A Unified Logical Framework. In *Principles and Practice of Semantic Web Reasoning, Second International Workshop, PPSWR 2004, St. Malo, France, Proceedings*, pages 50–60, 2004.

[104] T. Furche, F. Bry, S. Schaffert, R. Orsini, I. Horrocks, M. Krauss, and O. Bolzer. Survey over Existing Query and Transformation Languages. Rewerse deliverable, Institute for Informatics, University of Munich, 2004.

[105] B. Furht. Multimedia Systems: An Overview. *IEEE MultiMedia*, 1(1):47–59, 1994.

[106] A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann. A theoretical framework for ontology evaluation and validation. In *Proc. of SWAP 2005, the 2nd Italian Semantic Web Workshop, Trento, Italy, December 14-16, 2005, CEUR Workshop Proceedings*, 2005.

[107] R. Garcia and O. Celma. Semantic Integration and Retrieval of Multimedia Metadata. In *5th International Workshop on Knowledge Markup and Semantic Annotation*, Galway, Ireland, 2005.

[108] M. Gelfond and N. Leone. Logic programming and knowledge representation — the a-prolog perspective. *Artificial Intelligence*, 138(1–2):3–38, 2002.

[109] J. Geurts, J. van Ossenbrugen, and L. Hardman. Requirements for practical multimedia annotation. In *Proc. of the Workshop on Multimedia and the Semantic Web, 2nd European Semantic Web Conference (ESWC)*, pages 4–11, Heraklion, Crete, 2005.

[110] A. Gilliland-Swetland. Setting the Stage. In: Murtha Baca ed.: Introduction to Metadata, Pathway to Digital Information. Getty Information Institute. http://www.getty.edu/research/institute/standards/intrometadata/, 2000.

[111] J. Golbeck, , B. Parsia, and J. Hendler. Trust networks on the semantic web. In *Proceedings of Cooperative Intelligent Agents 2003*, 2003.

[112] J. Golbeck, B. Parsia, and J. Hendler. Trust Networks on the Semantic Web. In *Proceedings of Cooperative Intelligent Agents 2003*, 2003.

[113] N. M. Goldman. Ontology-Oriented Programming: Static Typing for the Inconsistent Programmer. In *Proceedings of the Second International Semantic Web Conference - ISWC 2003*, pages 850–865, 2003.

[114] A. Gomez-Perez, M. Fernandez-Lopez, and O. Corcho-Garcia. *Ontological Engineering*. Springer, Berlin, 2007.

[115] W. I. Grosky. Multimedia Information Systems. *IEEE MultiMedia*, 1(1):12–24, 1994.

[116] W. I. Grosky and R. Zhao. Negotiating the Semantic Gap: From Feature Maps to Semantic Landscapes. *Lecture Notes in Computer Science*, 2234:33–52, 2001.

[117] B. Grosof, I. Horrocks, R. Volz, and S. Decker. Description Logic Programs: Combining Logic Programs with Description Logics. In *Proc. of the 12th International World Wide Web Conference (WWW'03)*, Budapest, Hungary, 2003.

[118] W. T. A. Group. httpRange-14: What is the range of the HTTP dereference function? http://www.w3.org/2001/tag/issues.html?type=1#httpRange-14, 2005.

[119] T. Gruber. Ontology of Folksonomy:A Mash-up of Apples and Oranges. *First on-Line conference on Metadata and Semantics Research (MTSR'05)*, 2005.

[120] Y. Guo, Z. Pan, and J. Heflin. An Evaluation of Knowledge Base Systems for Large OWL Datasets. In *Proc. of the* $3^{rd}$ *International Semantic Web Conference (ISWC)*, pages 274–288, Hiroshima, Japan, 2004.

[121] C. A. Gurr. Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *Journal of Visual Languages and Computing*, 10(4):317–342, 1999.

[122] M. A. Gutierrez, F. Vexo, and D. Thalmann. Standardized Virtual Reality, Are We There Yet? In *CW '06: Proceedings of the 2006 International Conference on Cyberworlds*, pages 191–197, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2671-3.

[123] W. Haas, W. Bailer, and M. Hausenblas. Media Semantics. *ÖGAI Journal*, 26(1):24–29, 2007.

[124] C. Halaschek-Wiener, A. Schain, J. Golbeck, M. Grove, B. Parsia, and J. Hendler. A Flexible Approach for Managing Digital Images on the Semantic Web. In Proc. of the $5^{th}$ International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005), Galway, Ireland, 2005.

[125] W. Halb, Y. Raimond, and M. Hausenblas. Building Linked Data For Both Humans and Machines. In *WWW 2008 Workshop: Linked Data on the Web (LDOW2008)*, Beijing, China, 2008.

[126] S. Handschuh and S. Staab. Cream - creating metadata for the semantic web. *Computer Networks*, 42:579–598, 2003. Elsevier.

[127] L. Hardman, Z. Obrenovic, F. Nack, B. Kerherve, and K. Piersol. Canonical Processes of Semantically Annotated Media Production. In *Special Issue of Multimedia Systems Journal on Canonical Process of Media Production*. ACM Press, 2008.

[128] L. Hardman and J. van Ossenbruggen. Multimedia and the Web Workshop Report. http://homepages.cwi.nl/~lynda/www9/report.html, 2000.

[129] J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. B. Enser, and C. J. Sandom. Bridging the Semantic Gap in Multimedia Information Retrieval: Top-down and Bottom-up approaches. In *Mastering the Gap: From Information Extraction to Semantic Representation,* $3^{rd}$ *European Semantic Web Conference (ESWC)*, Budva, Montenegro, 2006.

[130] D. Harel and B. Rumpe. Meaningful Modeling: What's the Semantics of "Semantics"? *Computer*, 37(10):64–72, 2004.

[131] S. Harris and N. Gibbins. 3store: Efficient bulk RDF storage. In *Proc.* $1^{st}$ *International Workshop on Practical and Scalable Semantic Systems (PSSS'03), Sanibel Island*, pages 1–15, 2003.

[132] J. Hartmann and Y. Sure. An Infrastructure for Scalable, Reliable Semantic Portals. *IEEE Intelligent Systems*, 19(3):58–65, 2004.

[133] M. Hatala and G. Richards. Value-Added Metatagging: Ontology and Rule Based Methods for Smarter Metadata. In *Rules and Rule Markup Languages for the Semantic Web,* $2^{nd}$ *International Workshop (RuleML 2003)*, volume 2876 of *Lecture Notes in Computer Science*, pages 65–80, Sanibel Island, Florida, USA, 2003.

[134] M. Hausenblas. Applying Media Semantics Mapping in a Non-linear, Interactive Movie Production Environment. In *$1^{st}$ International Conference on New Media Technology (I-Media 07)*, pages 33–40, Graz, Austria, 2007.

[135] M. Hausenblas. Non-linear Interactive Media Productions. *Special Issue of ACM Multimedia Systems Journal on Canonical Processes of Media Production*, N.N.(N.N.):to appear, 2008.

[136] M. Hausenblas and B. Adida. RDFa in HTML Overview. Editor's Draft, W3C Semantic Web Deployment Working Group, 2007.

[137] M. Hausenblas and B. Adida. RDFa Implementation Report. Editor's Draft, W3C Semantic Web Deployment Working Group, 2008.

[138] M. Hausenblas and W. Bailer. Deploying Multimedia Metadata on the Semantic Web - RDFa-deployed Multimedia Metadata (ramm.x). Specification, ramm.x Working Group, 2007.

[139] M. Hausenblas, W. Bailer, T. Bürger, and R. Troncy. Deploying Multimedia Metadata on the Semantic Web (Poster). In *2nd International Conference on Semantics And digital Media Technologies (SAMT 07)*, 2007.

[140] M. Hausenblas, W. Bailer, and H. Mayer. Deploying Multimedia Metadata in Cultural Heritage on the Semantic Web. In *First International Workshop on Cultural Heritage on the Semantic Web, collocated with the $6^{th}$ International Semantic Web Conference (ISWC07)*, Busan, South Korea, 2007.

[141] M. Hausenblas, S. Boll, T. Bürger, O. Celma, C. Halaschek-Wiener, E. Mannens, and R. Troncy. Multimedia Vocabularies on the Semantic Web. W3C Incubator Group Report, W3C Multimedia Semantics Incubator Group, 2007.

[142] M. Hausenblas and W. Halb. Interlinking Multimedia Data. In *Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics08)*, 2008.

[143] M. Hausenblas and W. Halb. Interlinking of Resources with Semantics (Poster). In *5th European Semantic Web Conference (ESWC2008)*, 2008.

[144] M. Hausenblas, W. Halb, and Y. Raimond. Scripting User Contributed Interlinking. In *4th Workshop on Scripting for the Semantic Web (SFSW08)*, Tenerife, Spain, 2008.

[145] M. Hausenblas, W. Halb, Y. Raimond, and T. Heath. What is the Size of the Semantic Web? In *I-Semantics 2008: International Conference on Semantic Systems*, Graz, Austria, 2008.

[146] M. Hausenblas, H. Mayer, and G. Thallinger. Position Paper for W3C Workshop on Frameworks for Semantics in Web Services. Position paper, JOANNEUM RESEARCH Forschungsgesellschaft gmbH, 2005.

[147] M. Hausenblas and F. Nack. Interactivity = Reflective Expressiveness. *IEEE MultiMedia*, 14(2):1–7, 2007.

[148] M. Hausenblas and H. Rehatschek. mle: Enhancing the Exploration of Mailing List Archives Through Making Semantics Explicit. In *Semantic Web Challenge 2007 at the $6^{th}$ International Semantic Web Conference (ISWC07)*, Busan, South Korea, 2007.

[149] M. Hausenblas, W. Slany, and D. Ayers. A Performance and Scalability Metric for Virtual RDF Graphs. In *3rd Workshop on Scripting for the Semantic Web (SFSW07)*, Innsbruck, Austria, 2007.

[150] P. Hayes and B. McBride. RDF Semantics. W3c recommendation, World Wide Web Consortium, 2004.

[151] J. Heflin. *Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment*. PhD thesis, Faculty of the Graduate School of the University of Maryland, College Park, 2001.

[152] R. S. Heller and C. D. Martin. A Media Taxonomy. *IEEE MultiMedia*, 2(4):36–45, 1995.

[153] J. Hendler. Agents and the semantic web. *IEEE Intelligent Systems*, 16(2):30–37, 2001.

[154] M. D. Hill. What is scalability? *SIGARCH Comput. Archit. News*, 18(4):18–21, 1990.

[155] D. Hillman. Using Dublin Core. http://dublincore.org/documents/usageguide, 2001.

[156] M. Höffernig, M. Hausenblas, and W. Bailer. Semantics of Temporal Media Content Descriptions. In *7th Workshop of Multimedia Metadata Applications (M3A)*, Graz, Austria, 2007.

[157] L. Hollink, S. Little, and J. Hunter. Evaluating the application of semantic inferencing rules to image annotation. In *3rd International Conference on Knowledge Capture (K-CAP 2005)*, pages 91–98, Banff, Alberta, Canada, 2005. ACM.

[158] I. Horrocks. Description Logics - Basics, Applications, and More. Tutorial given at ECAI-2002, 2002. Available at http://www.cs.man.ac.uk/~horrocks/Slides/ecai-handout.pdf.

[159] I. Horrocks and P. F. Patel-Schneider. Reducing OWL Entailment to Description Logic Satisfiability. In *Proc. of the 2003 Description Logic Workshop (DL 2003)*, volume 81, pages 1–8, 2003.

[160] I. Horrocks and P. F. Patel-Schneider. Three Theses of Representation in the Semantic Web. In *Proc. of the Twelfth International World Wide Web Conference (WWW 2003)*, pages 39–47. ACM, 2003.

[161] I. Horrocks, P. F. Patel-Schneider, S. Bechhofer, and D. Tsarkov. OWL Rules: A Proposal and Prototype Implementation. *J. of Web Semantics*, 3(1):23–40, 2005.

[162] I. Horrocks, P. F. Patel-Schneider, S. Bechhofer, and D. Tsarkov. OWL rules: A proposal and prototype implementation. *Journal of Web Semantics*, 3(1):23–40, 2005.

[163] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. http://www.w3.org/Submission/SWRL/, 2004.

[164] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Technical report, World Wide Web Consortium, May 2004. http://www.w3.org/Submission/SWRL/.

[165] I. Horrocks, P. F. Patel-Schneider, and F. van Harmelen. Reviewing the Design of DAML+OIL: An Ontology Language for the Semantic Web. In *Proc. of the 18th Nat. Conf. on Artificial Intelligence (AAAI 2002)*, pages 792–797. AAAI Press, 2002.

[166] I. Horrocks, P. F. Patel-Schneider, and F. van Harmelen. From $\mathcal{SHIQ}$ and RDF to OWL: The Making of a Web Ontology Language. *J. of Web Semantics*, 1(1):7–26, 2003.

[167] HP. Amateur intro to Description Logics, 2001. Available at http://www.hpl.hp.com/semweb/download/DescriptionLogicsIntro.pdf.

[168] HTML 4.01 Specification. http://www.w3.org/TR/html4/, 1999.

[169] Hypertext Transfer Protocol – HTTP/1.1 – W3C. http://www.w3.org/Protocols/rfc2616/rfc2616.html, 1999.

[170] M. J. Hu and Y. Jian. Multimedia description framework (MDF) for content description of audio/video documents. In *DL '99: Proceedings of the fourth ACM conference on Digital libraries*, pages 67–75. ACM Press, 1999. ISBN 1-58113-145-3.

[171] J. Hunter. Adding Multimedia to the Semantic Web: Building an MPEG-7 ontology. In *Proceedings of the First Semantic Web Working Symposium*, pages 261–281, 2001.

[172] J. Hunter. Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):49–58, 2003.

[173] J. Hunter and L. Armstrong. A Comparison of Schemas for Video Metadata Representation. *Computer Networks*, 31(11-16):1431–1451, 1999.

[174] J. Hunter and C. Lagoze. Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles. In *$10^{th}$ International World Wide Web Conference (WWW'01)*, pages 457–466, Hong Kong, 2001.

[175] W. R. in HTML TF. RDF in XHTML Taskforce. http://www.w3.org/2001/sw/BestPractices/HTML, 2007.

[176] I. Jacobs and N. Walsh. Architecture of the World Wide Web, Volume One. http://www.w3.org/TR/webarch/, 2004.

[177] R. Jain. Semantics in Multimedia Systems. *IEEE MultiMedia*, 1(2):3–4, 1994.

[178] M. Janik and K. Kochut. BRAHMS: A WorkBench RDF Store and High Performance Memory System for Semantic Association Discovery. In *Proc. $4^{th}$ International Semantic Web Conference, ISWC 2005, Galway, Ireland*, pages 431–445, 2005.

[179] M. Jarrar, J. Demey, and R. Meersman. On Using Conceptual Data Modeling for Ontology Engineering. *J. Data Semantics*, 1:185–207, 2003.

[180] P. Jogalekar and M. Woodside. Evaluating the scalability of distributed systems. *IEEE Trans. Parallel Distrib. Syst.*, 11(6):589–603, 2000.

[181] R. Kaiser, M. Umgeher, and M. Hausenblas. Metadata-based Adaptive Assembling of Video Clips on the Web. In *$2^{nd}$ International Workshop on Semantic Media Adaptation and Personalization (SMAP07)*, London, UK, 2007.

[182] U. Keller, H. Lausen, and M. Stollberg. On the Semantics of Functional Descriptions of Web Services. In *The Semantic Web: Research and Applications (Proceedings of ESWC 2006)*, pages 605–619, 2006.

[183] M. Kifer, G. Lausen, and J.Wu. Logical foundations of object-oriented and frame-based languages. *J. ACM*, 42(4):741–843, 1995.

[184] H. Kim, H. Kim, J. H. Choi, and S. Decker. Translating Documents into Semantic Documents using Semantic Web and Web 2.0. In *Proceedings of the 1$^{st}$ Semantic Authoring and Annotation Workshop (SAAW2006)*, 2006.

[185] H. Kim, N. Moreau, and T. Sikora. Audio classification based on MPEG-7 spectral basis representations. *IEEE Transactions on Circuits and Systems for Video Technology 7, Special Issue on Audio and Video Analysis for Multimedia Interactive Services*, 14(5): 716–725, 2004.

[186] G. Klyne, J. J. Carroll, and B. McBride. RDF/XML Syntax Specification (Revised). W3c recommendation, World Wide Web Consortium, 2004.

[187] H. Knublauch, R. W. Fergerson, N. F. Noy, and M. A. Musen. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In *Proc. 3$^{rd}$ International Semantic Web Conference, ISWC2004, Hiroshima, Japan*, volume 3298 of *Lecture Notes in Computer Science*, pages 229–243, 2004.

[188] H. Knublauch, D. Oberle, P. Tetlow, and E. Wallace. A semantic web primer for object-oriented software developers. W3c working group note, W3C, 2006.

[189] I. Kompatsiaris, Y. Avrithis, P. Hobson, and M. G. Strintzis. Integrating Knowledge, Semantics and Content for User-Centred Intelligent Media Services: The aceMedia Project. http://citeseer.ist.psu.edu/651411.html, 2004.

[190] M. Korkea-aho. Scalability in distributed multimedia systems. http://citeseer.ist.psu.edu/korkea-aho95scalability.html, 1995.

[191] C. Lagoze and H. V. de Sompel. Compound Information Objects: The OAI-ORE Perspective. Discussion Document, Open Archives Initiative, 2007.

[192] O. Lassila. Web metadata: A matter of semantics. *IEEE Internet Computing*, 2(4), 1998.

[193] O. Lassila. Semantic Web, quo vadis? Keynote address at the Scandinavian Conference on AI and the Finnish AI Symposium. http://www.lassila.org/publications/2006/SCAI-2006-keynote.pdf, 2006.

[194] O. Lassila and J. Hendler. Embracing ”Web 3.0”. *IEEE Internet Computing*, 11(3): 90–93, 2007.

[195] H. Lausen, J. de Bruijn, A. Polleres, and D. Fensel. WSML - a Language Framework for Semantic Web Services. In *Proceedings of the W3C Workshop on Rule Languages for Interoperability*, 2005.

[196] R. Lee. Scalability report on triple store applications. Technical report, MIT, 2004.

[197] D. B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[198] H. J. Levesque and R. J. Brachman. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93, 1987.

[199] S. Little and J. Hunter. Rules-By-Example - A Novel Approach to Semantic Indexing and Querying of Images. In *3$^{rd}$ International Semantic Web Conference (ISWC'04)*, volume 3298 of *Lecture Notes in Computer Science*, pages 534–548, Hiroshima, Japan, 2004.

[200] P. Ljunglöf. *Expressivity and Complexity of the Grammatical Framework*. PhD thesis, Göteborg University and Chalmers University of Technology, Gothenburg, Sweden., 2004.

[201] L. Lohse, J. Sussner, and M. Thomas. Gormenghast Explore: iTV Drama. In *Interactive TV: A Shared Personal Experience. TICSP Adjunct Proceedings of EuroITV 2007*, pages 181–182, 2007.

[202] M. Lux, M. Granitzer, W. Klieber, M. Hausenblas, and H. Mayer. Digital Rights Management for Distributed Multimedia E-Learning Content. In *Interactive Computer Aided Learning Conference*, September 2005. Available at http://sw-app.org/pub/ICL2005_DRM.pdf.

[203] F. Manola and E. Miller. Resource Description Framework RDF Primer. W3c recommendation, World Wide Web Consortium, 2004.

[204] J. Markoff. Web content by and for the masses. New York Times Online, June 2005.

[205] O. Marques and N. Barman. Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques. In 2$^{nd}$ *International Semantic Web Conference (ISWC'03)*, volume 2870 of *Lecture Notes in Computer Science*, pages 550–565, Sanibel Island, Florida, USA, 2003.

[206] J. M. Martinez. MPEG-7 Overview (version 9); ISO/IEC JTC1/SC29/WG11N5525. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm, 2003.

[207] H. A. Maurer and K. Tochtermann. On a New Powerful Model for Knowledge Management and its Applications. *J. UCS*, 8(1):85–96, 2002.

[208] J. Mayfield and T. Finin. Information retrieval on the Semantic Web: Integrating inference and retrieval. In *Proceedings of the SIGIR Workshop on the Semantic Web*, August 2003.

[209] C. Meghini, F. Sebastiani, and U. Straccia. Mirlog: A logic for multimedia information retrieval. In *Logic and Uncertainty in Information Retrieval: Advanced models for the representation and retrieval of information*. Kluwer Academic Publishing, Dordrecht, NL, 1998. Forthcoming., 1998.

[210] A. Miles and D. Brickley. SKOS Core Guide. W3c working draft, World Wide Web Consortium, 2005.

[211] A. Miles and D. Brickley. SKOS Core Vocabulary Specification. W3c working draft, World Wide Web Consortium, 2005.

[212] E. L. Miller, D. Shen, J. Liu, and C. Nicholas. Performance and scalability of a large-scale N-gram based information retrieval system. *Journal of Digital Information (online refereed journal)*, 2000.

[213] T. Minter. GIS MetaData Dictionary. http://gis.ashevillenc.gov/PublicAccess/metadata/mdd.htm, 2000.

[214] R. Möller, V. Haarslev, and M. Wessel. On the scalability of description logic instance retrieval. In *29. Deutsche Jahrestagung für Künstliche Intelligenz*, Lecture Notes in Artificial Intelligence. Springer Verlag, 2006.

[215] A. Morgan and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM Press. ISBN 9781595935939.

[216] B. Motik. *Reasoning in Description Logics using Resolution and Deductive Databases*. PhD thesis, Universität Karlsruhe (TH), 2006.

[217] B. Motik, U. Sattler, and R. Studer. Query answering for OWL-DL with rules. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 3(1):41–60, 2005.

[218] MPEG-7. Information Technology—Multimedia Content Description Interface - Part 5: Multimedia Description Schemes. ISO/IEC. 15938-5, 2001.

[219] MPEG-7. Information Technology—Multimedia Content Description Interface - Part 9: Profiles and levels. ISO/IEC. 15938-9, 2001.

[220] MPEG-7. Information Technology—Multimedia Content Description Interface, Part 3: Visual. ISO/IEC. 15938-3, 2001.

[221] MPEG-7. Information Technology—Multimedia Content Description Interface, Part 4: Audio. ISO/IEC. 15938-4, 2001.

[222] MPEG-7. Multimedia Content Description Interface. Standard No. ISO/IEC n15938, 2001.

[223] RightsCom: The MPEG-21 Rights Expression Language. A White Paper. http://www.rightscom.com/Portals/0/whitepaper_MPEG21-RELCB.pdf, 2003.

[224] ISO/IEC. TR 21000-1:2001(E) Multimedia framework (MPEG-21) Part 1: Vision, Technologies and Strategy, 2001.

[225] J. Mylopoulos. An overview of Knowledge Representation. In *Proceedings of the 1980 workshop on Data abstraction, databases and conceptual modeling*, pages 5–12. ACM Press, 1980. ISBN 0-89791-031-1.

[226] F. Nack and A. T. Lindsay. Everything you wanted to know about MPEG-7 (Part I). *IEEE Multimedia*, 6(3):65–77, 1999.

[227] F. Nack and A. T. Lindsay. Everything you wanted to know about MPEG-7 (Part II). *IEEE Multimedia*, 6(4):64–73, 1999.

[228] F. Nack and W. Putz. Designing annotation before it's needed. In *ACM Multimedia*, pages 251–260, 2001.

[229] F. Nack, J. van Ossenbruggen, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web, Part 2. *IEEE MultiMedia*, 12(1):54–63, 2005.

[230] Y. Naudet, D. Mathevon, A. Vagner, S. Renault, and J.-S. Brunner. MPEG-7 and the Semantic Web: An Hybrid Approach for Semantic and Behavioral Descriptions in Interactive Object-Oriented Multimedia Applications. *axmedis*, 0:213–216, 2005.

[231] T. Nelson. *Literary Machines*. Eastgate Systems, 1982.

[232] R. F. B. Neto and M. da Graça Pimentel. Performance evaluation of inference services for ubiquitous computing. In *WebMedia 2006: Proceedings of the 12$^{th}$ Brazilian symposium on Multimedia and the Web*, pages 27–34. ACM Press, 2006.

[233] M. Nottingham and R. Sayre. The Atom Syndication Format. RFC 4287, IETF—Network Working Group, 2005.

[234] K. O'Hara, H. Alani, Y. Kalfoglou, and N. Shadbolt. Trust strategies for the semantic web. In *ISWC Workshop on Trust, Security, and Reputation on the Semantic Web*, 2004.

[235] K. O'Hara, H. Alani, Y. Kalfoglou, and N. Shadbolt. Trust Strategies for the Semantic Web. In *ISWC Workshop on Trust, Security, and Reputation on the Semantic Web*, 2004.

[236] J. C. on Metadata (BAER/DLC). Definition of types of metadata. digital library committee. http://staffweb.library.northwestern.edu/dl/metadata/standardsinventory/definition.html, 2002.

[237] S. Ortiz. Searching the Visual Web. *Computer*, 40(6):12–14, 2007.

[238] J. O'Sullivan, D. Edmond, and A. H. ter Hofstede. Formal description of non-functional service properties. technical report. Technical report, Queensland University of Technology, Brisbane., 2005.

[239] OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide, 2004.

[240] OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/, 2004.

[241] OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref, 2004.

[242] OWL Web Ontology Language Use Cases and Requirements. http://www.w3.org/TR/webont-req/, 2004.

[243] OWL Web Ontology Language Semantics and Abstract Syntax. http://www.w3.org/TR/owl-semantics, 2004.

[244] J. Z. Pan. *Description Logics: Reasoning Support for the Semantic Web*. PhD thesis, School of Computer Science, The University of Manchester, 2004.

[245] P. F. Patel-Schneider and D. Fensel. Layering the semantic web: Problems and directions. In *Proc. of the First International Semantic Web Conference (ISWC2002), Sardinia, Italy*, 2002.

[246] M. Petkovic and W. Jonker. Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events. In *IEEE Workshop on Detection and Recognition of Events in Video*, pages 75–82, 2001.

[247] S. Pfeiffer, C. Parker, and A. Pang. The Continuous Media Markup Language (CMML), Version 2.0. Internet-Draft, The Internet Engineering Task Force, 2005.

[248] H. S. Pinto, C. Tempich, S. Staab, and Y. Sure. Distributed engineering of ontologies (diligent). In S. Stuckenschmidt and S. Staab, editors, *Semantic Web and Peer-to-Peer*, pages 301–320. Springer, 2005.

[249] F. Pittarello and A. D. Faveri. Semantic description of 3D environments: a proposal based on web standards. In *Web3D '06: Proceedings of the eleventh international conference on 3D web technology*, pages 85–95, New York, NY, USA, 2006. ACM Press. ISBN 1-59593-336-0.

[250] EBU Project Group P/Meta. http://www.ebu.ch/trev_290-hopper.pdf, 2002.

[251] P_Meta. The EBU Metadata Exchange Scheme. EBU Tech 3295, version 1.2, 2005.

[252] E. Prud'hommeaux. W3c acl system. http://www.w3.org/2001/04/20-ACLs, 2001.

[253] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. W3c candidate recommendation 14 june 2007, W3C RDF Data Access Working Group, 2007.

[254] K. Ramamohanarao and J. Harland. An Introduction to Deductive Database Languages and Systems. *The VLDB Journal*, 3(2):107–122, 1994.

[255] Vocabulary Description Language 1.0: RDF Schema – W3C. http://www.w3.org/TR/rdf-schema, 2004.

[256] H. Rehatschek, M. Hausenblas, G. Thallinger, and W. Haas. Cross Media Aspects in the Areas of Media Monitoring and Content Production. In *5th International Conference on Language Resources and Evaluation (LREC) 2006, cross-media indexing workshop*, pages 25–31, Genoa, Italy, 2006.

[257] R. Reiter. On closed world data bases. In *Logic and Data Bases*, pages 55–76, 1977. ISBN 0-306-40629-2.

[258] R. Reiter. Data bases: A logical perspective. In *Proceedings of the 1980 workshop on Data abstraction, databases and conceptual modeling*, pages 174–176, New York, NY, USA, 1980. ACM Press. ISBN 0-89791-031-1.

[259] A. Riazanov and A. Voronkov. The design and implementation of VAMPIRE. *AI Commun.*, 15(2-3):91–110, 2002.

[260] RIF. Rule Interchange Format. http://www.w3.org/2005/rules/, 2007.

[261] D. Roman, H. Lausen, and U. Keller. Web Service Modeling Ontology (WSMO), WSMO Deliverable D2v1.0., WSMO Working Draft 20 September 2004, September 2004.

[262] S. Russel and P. Norvig. *Artificial Intelligence, A Modern Approach*. Prentice Hall, Inc., 1995. ISBN 0-13-360124-2.

[263] L. Sauermann and R. Cyganiak. Cool URIs for the Semantic Web. W3C Interest Group Note, W3C Semantic Web Education and Outreach Interest Group., 2008.

[264] L. Sauermann and S. Schwarz. Gnowsis Adapter Framework: Treating Structured Data Sources as Virtual RDF Graphs. In *Proc. $4^{th}$ International Semantic Web Conference, ISWC 2005, Galway, Ireland*, 2005.

[265] M. Schmidt-Schauß and G. Smolka. Attributive Concept Descriptions with Complements. *Artificial Intelligence*, 48(1):1–26, 1991.

[266] R. Schroeter, J. Hunter, and D. Kosovic. FilmEd - Collaborative Video Indexing, Annotation and Discussion Tools Over Broadband Networks. In Y.-P. P. Chen, editor, *Proceedings of the $10^{th}$ International Multimedia Modeling Conference (MMM 2004), 5-7 January 2004, Brisbane, Australia*, pages 346–353. IEEE Computer Society, 2004. ISBN 0-7695-2084-7.

[267] A. Sheth, C. Ramakrishnan, and C. Thomas. Semantics for the Semantic Web: The Implicit, the Formal and the Powerful. *International Journal on Semantic Web & Information Systems*, 1(1), 2005.

[268] L. Simons. RDF at the Venice Project. http://www.leosimons.com/2006/rdf-at-the-venice-project.html, 2006. Blog Post.

[269] E. P. B. Simperl and C. Tempich. Ontolgy Engineering: A Reality Check. In *Proc. of the $5^{th}$ International Conference on Ontologies, Data Bases, and Applications of Semantics (ODBASE2006)*, volume 4275 of *Lecture Notes in Computer Science*, pages 836–854. Springer, 2006.

[270] K. Siorpaes and M. Hepp. OntoGame: Weaving the Semantic Web by Online Games. In *European Semantic Web Conference (ESWC 2008)*, pages 751–766. Springer, 2008.

[271] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[272] SMIL. Synchronized multimedia integration language. http://www.w3.org/TR/SMIL2/, 2006.

[273] J. R. Smith and P. Schirling. Metadata Standards Roundup. *IEEE MultiMedia*, 13(2): 84–88, 2006.

[274] SMPTE. Material exchange format (mxf) - descriptive metadata scheme-1 (standard) smpte 380m. http://www.pro-mpeg.org/publicdocs/mxf.html, 2004.

[275] SMPTE. Material exchange format (mxf) - file format specification (standard) smpte 377m. http://www.pro-mpeg.org/publicdocs/mxf.html, 2004.

[276] Simple Object Access Protocol. http://www.w3.org/TR/SOAP/, May 2000.

[277] S. Softic and M. Hausenblas. Towards Opinion Mining Through Tracing Discussions on the Web. In *Social Data on the Web (SDoW 2008) Workshop at the 7$^{th}$ International Semantic Web Conference*, Karlsruhe, Germany, 2008.

[278] S. Staab, J. Angele, S. Decker, B. Grosof, I. Horrocks, M. Kifer, and G. Wagner. Where are the rules? *IEEE Intelligent Systems, Trends & Controversies*, 18(5):76–83, 2003.

[279] S. Staab and R. Studer, editors. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer, 2004. ISBN 3-540-40834-7.

[280] S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure. Knowledge Processes and Ontologies. *IEEE Intelligent Systems*, 16(1):26–34, 2001.

[281] G. Stamou, J. van Ossenbruggen, J. Z. Pan, and G. Schreiber. Multimedia annotations on the semantic web. *IEEE MultiMedia*, 13(1):86–90, 2006.

[282] H. Stoermer, I. Palmisano, and D. Redavid. Achieving scalability and expressivity in an rdf knowledge base by implementing contexts. In *In Proceedings of SWAP 2006, the 3$^{rd}$ Italian Semantic Web Workshop*. CEUR Workshop Proceedings, 2006.

[283] H. Stuckenschmidt, R. Vdovjak, G.-J. Houben, and J. Broekstra. Index Structures and Algorithms for Querying Distributed RDF Repositories. In *WWW '04: Proc. of the 13$^{th}$ International Conference on World Wide Web*, pages 631–639, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-844-X.

[284] R. Studer, A. Ankolekar, P. Hitzler, and Y. Sure. A semantic future for ai. *IEEE Intelligent Systems*, 21(4):8–9, 2006.

[285] G. Stumme, B. Berendt, and A. Hotho. Usage Mining for and on the Semantic Web. In *Next Generation Data Mining. Proc. NSF Workshop, Baltimore, Nov. 2002*, pages 77–86, 2002.

[286] G. Sutcliffe and C. Suttner. The TPTP problem library: CNF release v1.2.1. *Journal of Automated Reasoning*, 21(2):177–203, 1998. Available at http://www.cs.miami.edu/~tptp/.

[287] Semantic Web Activity – W3C. http://www.w3.org/2001/sw/, 2001.

[288] A Proposal for a SWRL Extension towards First-Order Logic. http://www.w3.org/Submission/SWRL-FOL/, 2005.

[289] J. Tasic, U. Burnik, M. Ansorge, and J. Cid-Sueiro. Content Understanding for Smart Multimedia Systems. In *The International Conference on Computer as a Tool (EUROCON 2005)*, pages 175–178, Belgrade, Serbia, 2005.

[290] R. Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *Proc. of the 2$^{nd}$ International Semantic Web Conference (ISWC'03)*, volume LNCS 2870, pages 566–581, 2003.

[291] R. Troncy, W. Bailer, M. Hausenblas, P. Hofmair, and R. Schlatte. Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. In *1$^{st}$ International Conference on Semantics And digital Media Technology (SAMT'06)*, pages 41–55, Athens, Greece, 2006.

[292] R. Troncy and J. Carrive. A Reduced Yet Extensible Audio-Visual Description Language: How to Escape From the MPEG-7 Bottleneck. In *$4^{th}$ ACM Symposium on Document Engineering*, Milwaukee, Wisconsin, USA, 2004.

[293] R. Troncy, O. Celma, S. Little, R. Garcia, and C. Tsinaraki. MPEG-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue? In *1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies Workshop (MAReSO07)*, Genova, Italy, 2007.

[294] R. Troncy, L. Hardman, J. van Ossenbruggen, and M. Hausenblas. Identifying Spatial and Temporal Media Fragments on the Web. Position Paper, W3C Video on the Web Workshop, 2007.

[295] R. Troncy, J. van Ossenbruggen, C. Halaschek-Wiener, G. Stamou, and V. Tzouvaras. Multi-Perspective Requirements for a Common Multimedia Ontology Framework. Call for requirements for a multimedia ontology framework, 2006.

[296] R. Troncy, J. van Ossenbruggen, J. Z. Pan, and G. Stamou. Image Annotation on the Semantic Web. Xgr, W3C Multimedia Semantics Incubator Group, 2007.

[297] D. Tsarkov and I. Horrocks. DL reasoner vs. first-order prover. In *Proc. of the 2003 Description Logic Workshop (DL 2003)*, volume 81, pages 152–159, 2003.

[298] C. Tsinaraki, P. Polydoros, N. Moumoutzis, and S. Christodoulakis. Integration of OWL ontologies in MPEG-7 and TV-Anytime compliant Semantic Indexing. In *$16^{th}$ International Conference on Advanced Information Systems Engineering*, Riga, Latvia, 2004.

[299] G. Tummarello, C. Morbidoni, P. Puliti, A. F. Dragoni, and F. Piazza. From Multimedia to the Semantic Web using MPEG-7 and Computational Intelligence. In *4th International Conference on WEB Delivering of Music (WEDELMUSIC 2004), Barcelona, Spain*, pages 52–59. IEEE Computer Society, 2004.

[300] V. Tzouvaras, R. Troncy, and J. Pan. Multimedia Annotation Interoperability Framework. W3c incubator group editor's draft 14 august 2007, W3C Multimedia Semantics Incubator Group, 2007.

[301] UDDI Technical White Paper. http://www.uddi.org/.

[302] J. D. Ullman. *Principles of Database and Knowledge-Based Systems. Volume I: Classical Database Systems*. Computer Science Press, 1988.

[303] M. Ursu and J. Cook. D5.3: Languages for the representation of visual narratives. Deliverable to EC (restricted), NM2 consortium, 2005.

[304] M. Ursu, J. Cook, V. Zsombori, I. Kegel, D. Williams, M. Hausenblas, and M. Tuomola. ShapeShifting Screen Media: A Declarative Computational Model for Interactive Reconfigurable Moving Image Narratives. In *Poster at the Artificial Intelligence and Interactive Digital Entertainment Conference*, 2007.

[305] M. F. Ursu and J. Cook. D5.3: Languages for the representation of visual narratives. Deliverable to EC (permission required), NM2 consortium, 2005.

[306] M. Uschold. Towards a Methodology for Building Ontologies. *Towards a Methodology for Building Ontologies Workshop on Basic Ontological Issues in Knowledge Sharing, held in conduction with IJCAI-95*, 1995.

[307] M. Uschold and M. Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 1996.

[308] V. Tzouvaras (ed.). Multimedia Annotation Interoperability Framework; MMSEM XG Report. http://www.w3.org/2005/Incubator/mmsem/wiki/Semantic_Interoperability, 2007.

[309] P. van Beek, J. Smith, T. Ebrahimi, T. Suzuki, and J. Askelof. Metadata driven multimedia access. *IEEE Signal Processing magazine*, pages 40–52, 2003.

[310] J. van Ossenbruggen, F. Nack, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web, Part 1. *IEEE MultiMedia*, 11(4):38–48, 2004.

[311] J. R. van Ossenbruggen, L. Hardman, and L. W. Rutledge. Hypermedia and the semantic web: A research agenda. Technical report, CWI, Amsterdam, The Netherlands, 2001.

[312] R. Volz. *Web Ontology Reasoning with Logic Databases*. PhD thesis, Universität Karlsruhe, Fak. f. Wirtschaftswissenschaften, 2004.

[313] L. von Ahn. Games with a Purpose. *Computer*, 39(6):92–94, 2006.

[314] W3C. Scalable Vector Graphics (SVG) 1.1 Specification. http://www.w3.org/TR/SVG11/, 2003.

[315] W3C. Multimedia Semantics Incubator Group (MMSEM XG). http://www.w3.org/2005/Incubator/mmsem, 2007.

[316] W3C. Semantic Web Development Tools. http://esw.w3.org/topic/SemanticWebTools, 2007.

[317] W3C. The World Wide Web Consortium – W3C. http://www.w3.org, 2007.

[318] H. Wache, L. Serafini, and R. García-Castro. Survey of scalability techniques for reasoning with ontologies. Technical report, KnowledgeWeb Deliverable D2.1.1, 2004.

[319] J. Wang, M. Scardina, and K. Zhou. High-Performance XML Data Retrieval. In *Proc. of the XML Europe 2004*, page N.A., 2004.

[320] S. Wang, D. E. Rydeheard, and J. Z. Pan. The semantic processing of continuous quantities for discrete terms in ontologies. *Journal of Logic and Computation*, 2007.

[321] T. D. Wang, B. Parsia, and J. Hendler. A survey of the web ontology landscape. In *The Semantic Web - ISWC 2006, 5$^{th}$ International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, pages 682–694, Athens, GA, USA, 2006. Springer.

[322] K. Ward, M. Ursu, and V. Zsombori. D9.2 NM2 Training Package (User Manual). Deliverable to EC (public), NM2 consortium, 2005.

[323] W. Weiss and M. Hausenblas. Visual Exploration, Query, and Debugging of RDF Graphs. In *Semantic Web User Interaction Workshop (SWUI 08) at CHI 2008*, 2008.

[324] T. Weithöner, T. Liebig, M. Luther, and S. Böhm. What's Wrong with OWL Benchmarks? In *Proc. of the Second Int. Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006)*, pages 101–114, Athens, GA, USA, 2006.

[325] J. Wielemaker, G. Schreiber, and B. J. Wielinga. Prolog-Based Infrastructure for RDF: Scalability and Performance. In D. Fensel, K. P. Sycara, and J. Mylopoulos, editors, *International Semantic Web Conference*, volume 2870 of *Lecture Notes in Computer Science*, pages 644–658. Springer, 2003. ISBN 3-540-20362-1.

[326] D. Williams, M. Ursu, J. Cook, V. Zsombori, M. Engler, and I. Kegel. ShapeShifted TV – Enabling Multi-Sequential Narrative Productions for Delivery over Broadband. In *The 2$^{nd}$ IET Multimedia Conference, 29-30 November 2006*. ACM Press, 2006.

[327] Web Service Modeling Language (WSML). http://www.w3.org/Submission/WSML/, 2005.

[328] Metadata 101—a two minute introduction to metadata. The Wyoming Geographic Information Sciences Center. http://www.wygisc.uwyo.edu/clearinghouse/metainfo.html, 2002.

[329] X3D. Extensible 3D (X3D). ISO/IEC 19775:2004, 2004.

[330] XHTML 1.0 The Extensible HyperText Markup Language. http://www.w3.org/TR/xhtml1/, 2002.

[331] XHTML 2.0. http://www.w3.org/TR/xhtml2/, 2006.

[332] Extensible Markup Language (XML) 1.0 (Fourth Edition) – W3C. http://www.w3.org/TR/xml/, 2006.

[333] XML Schema – W3C. http://www.w3.org/XML/Schema, 2006.

[334] A. Yoshitaka and T. Ichikawa. A Survey on Content-Based Retrieval for Multimedia Databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999.

[335] H. Zeiner, G. Kienast, M. Hausenblas, C. Derler, and W. Haas. Video assisted geographical information systems. In *Cross-Media Service Delivery*, pages 205–215, 2003.

[336] R. Zhao and W. I. Grosky. Bridging the Semantic Gap in Image Retrieval. *Distributed multimedia databases: techniques and applications*, pages 14–36, 2002.

# Glossary

**Agent** A player in the ↪ SEMANTIC WEB . Uses the Semantic Web infrastructure to fulfil a (human) users instruction, i.e. is equipped with a goal and the users preferences to achieve a certain task.

**DC** **D**ublin **C**ore iss s bibliographic metadata format. Defines a number of generic terms (e.g. TITLE, SUBJECT, etc.) for use on the markup of (Web) documents. See also http://dublincore.org/

**DL** Description Logics. See also http://dl.kr.org/

**DLG** **D**irected **L**abled **G**raph. A graph is a set of objects called vertices connected by links called arcs which are-in the case of a DLG-directed.

**Expressivity** As of our definition 7.1 on page 127, expressivity is a measure for the computational complexity of the reasoning algorithms of a knowledge representation language.

**Linked Data** A term used to describe openly accessible and interlinked data about "things" on the Web. Sir Tim Berners-Lee outlined four principles of Linked Data, namely (i) to use URIs to identify things that you expose to the Web as resources, (ii) to use HTTP URIs so that people can locate and look up (dereference) these things, (iii) to provide useful information about the resource when its URI is dereferenced, and (iv) to include links to other, related URIs in the exposed data as a means of improving information discovery on the Web.
See also http://www.w3.org/DesignIssues/LinkedData.html

**Semantic Web multimedia application** As of our definition 2.5 on page 22, a piece of software dealing with multimedia entities, accessible on the Semantic Web;

**Metadata** *data about data.*

**microformats** Fixed set of structured metadata in HTML using attributes.
See also http://microformats.org/.

**MPEG-7** Formally the "Multimedia Content Description"—a universal multimedia metadata standard from ISO.

**Multimedia Asset** Multimedia content (such as a video clip) possibly along with certain metadata.

**Ontology** An ontology is a formal specification of a conceptionalisation of some domains concepts and relations.

**OWL** The Web Ontology Language. A Description Logic-based ontology language defined by W3C. Builts on RDF(S)and XML as graph and serialisation format. The basis for the formal definition of vocabularies to be used in the ↪ SEMANTIC WEB . See also http://www.w3.org/TR/owl-ref/

**ramm.x** A light-weight RDFa-based mechanism for deploying multimedia metadata on the Semantic Web (cf. chapter 10).

**RDFa** A concrete serialisation syntax for the ↪ RDF data model in (X)HTML using attributes. See also http://www.w3.org/TR/xhtml-rdfa-primer/

**RDF** **R**esource **D**escription **F**ramework is a DLG-based model for representing knowledge on the Web. RDF forms the basis of ↪ OWL and ↪ RDF SCHEMA .
See also http://www.w3.org/RDF/.

**RDF Schema** A schema language for defining basic ontology constraints.
See http://www.w3.org/TR/rdf-schema/.

**Semantic Web** The extension of the current Web with (formal, machine processable) semantics using RDF(S), OWL, etc.

**Semantic Web application** As of our definition 2.2 on page 13, a piece of software accessible on the Semantic Web.

**Semantic Web Agent** See ↪ AGENT .

**Scalability** As of our definition 7.2 on page 130, scalability is a system-intrinsic property mainly depending on two factors, namely (i) decentralisation, i.e. the way metadata is distributed over the Web, and (ii) the type of inference, i.e., making implicit facts explicit, a system is expected to carry out.

**URI** **U**niform **R**esource **I**dentifier. See also http://www.ietf.org/rfc/rfc2396.txt

**Vocabulary** A light-weight form of an ↪ ONTOLOGY .

**voiD** Vocabulary of interlinked datasets. A light-weight vocabulary to describe the content of linked data.

**Web application** As of our definition 2.1 on page 13, a piece of software accessible on the Web.

**W3C** **W**orld **W**ide **W**eb **C**onsortium. See also http://www.w3c.org/

**XML** e**X**tensible **M**arkup **L**anguage. Is a tree-based text format for representing structured documents (defined by ↪ W3C ).

**XML Schema** A schema language for XML allowing the definition of constraints (order, inheritance, range, etc.). See also http://www.w3.org/XML/Schema

**XPath** A (tree)pattern language for selecting of sub-trees in an XML tree.
See also http://www.w3.org/TR/xpath

**XSL** e**X**tensible **S**tyle **L**anguage. Is a XML-application, for transforming XML-trees.

# Index