Dissertation submitted to the Graz University of Technology,
Faculty of Computer Science, for the attainment of the degree of
Doctor of Engineering Sciences (Dr. techn.)

by

Nicolas Weber

March 06, 2012

A User-centered Approach to Information Quality Assessment in Social Software Systems

Thesis supervisors

First Advisor
Prof. Dr. Stefanie Lindstaedt
Knowledge Management Institute
Graz University of Technology

Second Advisor
Prof. Dr. Tobias Ley
Institute of Informatics
Tallinn University

Acknowledgements

Though only my name appears on the cover of this dissertation, many people have contributed. I owe my gratitude to all those people who have made this dissertation possible.

I would like to thank Prof. Dr. Stefanie Lindstaedt, head of the Knowledge Management Institute and supervisor of my thesis and Prof. Dr. Tobias Ley as second reader of my thesis for their scientific expertise and support over the last years.

Further, I want to thank my colleagues from the Knowledge Management Institute and the Knowledge Evolution Group of the Know-Center for inspiration and feedback, especially Dr. Barbara Kump for her encouragement and practical advice. I am grateful for the fruitful cooperation with the people in the MATURE project, especially Dieter Theiler, Tobias Nelkner (University of Paderborn), Simone Braun (FZI) and Karin Schöfegger.

I thank my lovely wife for her support. We went together through the ups and downs of writing a dissertation. Eventually, I thank my family for their patience and understanding.

This work has been partially funded by the European Commission as part of the MATURE IP (grant no. 216346) within the 7th Framework Programme of IST

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am ……………………………   ………………………………………………..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……………………………   ………………………………………………..
date   (signature)

To design for better quality,
it is necessary first to understand WHAT QUALITY MEANS and HOW IT IS MEASURED.

Wand, Y., & Wang, R. (1996)

Abstract

Over the past decade, the enormous increase in user generated content on the Web has provided an inexhaustible source of information, so much so that being able to find or locate a required piece of information has become more and more like finding a needle in a haystack. The problem is not so much the vast amount of information available as the varying quality of information contributed by non-experts. After all, searching for information online is as much about finding the relevant information on a particular topic as finding information of quality - at least as far as the user is concerned. It means therefore that perception of quality cannot be assessed merely in general terms but always in relation to the context of the user.

The aim of this work is to increase if not create awareness of information quality in systems dealing with user generated content. With such an aim, understanding the notion and meaning of information quality is therefore critical. Building on the theoretical foundation of information quality research, I present a model for assessing the quality of information in social software systems. In the light of this model, creating quality awareness necessarily involves two steps. One the one hand, aspects of quality ought to be measured in order to track the qualitative status of available resources. On the other hand, the user requirements regarding quality aspects also have to be elicited.

For the first step in creating quality awareness, measuring quality aspects, I propose a metric framework facilitating the assessment of the current quality status of resources. In a field study I tested the metric framework with real world data of a classroom setting. The study shows that these metrics facilitate the assessment of different factors of information quality.

For the second step in creating quality awareness, collecting user requirements, I propose mechanisms that implicitly or explicitly elicit the users´ needs. Furthermore, in order to be able to find resources that fit the users' needs, I

propose an algorithm for calculating similarities in quality patterns. I evaluated the feedback mechanisms and the similarity algorithm by comparing the perceived quality of the users with the calculated quality status. The study shows that the collected requirements correspond to the perceived quality of resources.

To conclude, this work proposes an approach that enables social software systems to provide information corresponding to the user's requirements, not only in terms of information content but even more importantly in terms of information quality.

Kurzfassung

In den letzten Jahren hat die starke Zunahme nutzergenerierter Inhalte das Web zu einer fast unerschöpflichen Quelle von Informationen gemacht. Dadurch gestaltet sich die Suche nach der richtigen Information ähnlich der Suche nach einer Nadel im Heuhaufen. Das Problem dabei ist jedoch nicht unbedingt die Menge der Informationen, sondern vielmehr die schwankende Qualität der oft von Laien erstellten Inhalte. Bei der Suche nach Informationen ist es daher wichtig, dass diese nicht nur inhaltlichen Kriterien, sondern vor allem auch den Qualitätsansprüchen der Suchenden entsprechen. Qualität kann in diesem Zusammenhang aber nicht allgemein bewertet werden, sondern ist immer von der Situation, in der sich die Benutzer/-innen befinden, abhängig. Obwohl die Qualität von Information ein wichtiger Aspekt ist, wird diese von derzeitigen Informationssystemen im Web kaum berücksichtigt.

Ziel dieser Arbeit ist es daher, qualitätssensitive Mechanismen zur Bereitstellung von nutzergenerierten Inhalten in sozialen Softwaresystemen zu untersuchen und bereitzustellen. Dabei wird zunächst das Verständnis von Informationsqualität für diese Arbeit herausgestellt. Aufbauend auf diesen theoretischen Betrachtungen wird ein Modell zur Bewertung von Informationsqualität vorgeschlagen. Dieses umfasst zwei wesentliche Schritte: Einerseits muss die Qualität einer Information im Bezug auf bestimmte Kriterien beurteilt werden. Die Evaluation des Metrik-Frameworks in einer Wiki-Studie zeigt, dass die gewählten Metriken als Indikatoren für die Informationsqualität geeignet sind.

Andererseits müssen jedoch auch die Ansprüche der Benutzer/-innen an die Informationsqualität ermittelt werden. Hierfür werden Mechanismen vorgestellt, die die Ermittlung sowohl von implizit als auch von explizit ausgedrückten Qualitätsanforderungen unterstützen. Ein Algorithmus zur Berechnung der Ähnlichkeit von Qualitätsmustern ermöglicht das Auffinden von Inhalten entsprechend der Qualitätsanforderungen. Die Evaluation dieses Ansatzes zeigt, dass dieser zur Steigerung der Qualität in Informationssystemen beiträgt und die

Mechanismen es ermöglichen Inhalte, die den Qualitätsanforderungen der Benutzer entsprechen, zu identifizieren.

Damit stellt diese Arbeit einen Ansatz vor, wie Informationssysteme, die Informationen aus nutzergenerierten Inhalten anbieten, den Erwartungen der Benutzer/-innen entsprechen können – und zwar nicht nur Inhaltlich sondern auch im Bezug auf Informationsqualität.

Table of Contents

1 INTRODUCTION

In recent years, the Web has gone through a metamorphosis, from a more or less static source of information to a network of users who are active contributors. A new consciousness of web usage and new technologies has enabled the user to share knowledge on the web. Systems that allow for the user to be author and consumer at the same time are rapidly evolving. Discovering a lack of quality is the bottleneck in many social information systems because the provision of high-quality data is essential for system acceptance. Therefore, it has become increasingly important to check and to ensure the quality of information in social information systems.

Information provided by social software systems like Wikis is usually created by a community of users. In contrast to traditional systems where content is produced by dedicated authors, no additional quality assurance is provided in these systems. Both the creation of content as well as assurance of a certain level of quality are carried out by the community. In some cases, this works astonishing well while in other cases acceptance of systems may be low due to a lack of content quality. Delone and McLean (2003) investigated the influence of information quality on information system acceptance. Their *Model of Information System Success* describes information quality as one of the vital prerequisites for system acceptance.

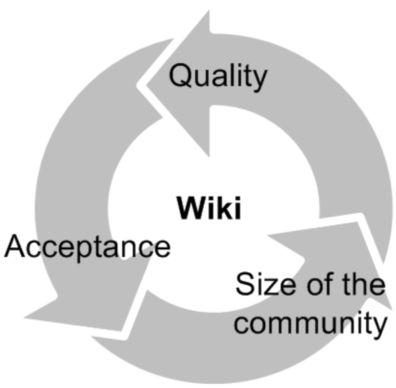


Figure 1-1 Cause-and-effect chain in Wiki systems

One factor that influences content quality in information systems is the size of the community. This effect is often denoted as *the principle of many pairs of eyes*. The idea is that having many pairs of eyes is one strategy for weeding out errors in system content. In this way, with millions of users, Wikipedia attains an outstanding quality in its articles. In many cases, articles created collaboratively exceed traditional encyclopedias in terms of content quality (Giles, 2005). Hence, the vicious circle in information system quality is that acceptance of a system depends on the quality of system content that on its part is dependent on the size of the community which in turn depends on system acceptance (Figure 1-1). This means that ensuring the quality of information in information systems is a vital factor for the acceptance and success of the entire system.

But not all systems are suffering from insufficient quality of information. There are also small- and middle-sized systems with a moderate number of users that provide good quality of information. These examples show that there must be other influential factors beyond the mere number of users. Especially in communities of practice, amongst others, motivation, group dynamics and consciousness of responsibility are crucial factors when it comes to quality. Measuring these factors can also be used as an indicator of quality.

Small- and middle-sized systems provide information about a certain topic or product and are hosted by companies or communities of interest. Compared to the few big systems, there is a vast quantity of them on the internet, even if these are not the biggest systems regarding the amount of content. In contrast to huge social systems, which number less than 50 on the web, there are thousands of social systems dealing with a specific topic; corporate Wikis for hardware and software products, forums and Wikis operated by communities of interest. Due to their specific content, the community of users is smaller and so there are less *pairs of eyes* observing the quality of content. This effect is often denoted as *The Long Tail* which means that although the popular social systems have millions of articles and millions of users, the sum of content and users in small- to middle-sized systems is even more. In these systems, it is often difficult to maintain a certain level of quality.

For the reasons mentioned above, the approach described in this work focuses on middle-sized systems in an organizational or community environment. Therefore, I propose a user centered approach for quality assessment in social systems. Three questions are addressed: First, how can I **detect and represent** the quality needs of the user? Second, how can I **measure** the qualitative status of a resource? Third, how can I map the resource quality status to the user **quality requirements** in order to provide resources that comply with the users' quality needs? These questions led me to the main research questions of this work which are presented and discussed in the next section.

1.1 RESEARCH QUESTIONS AND SCIENTIFIC CONTRIBUTION

The introduction has already provided an insight into the problem statement and its relevance to this field. The research questions concretize the focus of this work and help finding the appropriate answers to the right question. The main research question of this work is defined as follows:

How can quality aware system behavior in a social software system be achieved?

It follows therefore that three sub-questions will have to be answered:

1. ***How can the qualitative status of resources be measured and represented?***

2. ***How can the user requirements regarding quality be identified and represented?***

3. ***How can resources that correspond to the user requirements be found?***

This work brings together various well researched approaches from different areas. The strength of this approach is to build on those existing ones in order to attain the goal of this research which is quality aware system behavior.

Research on information quality has a long tradition in science, and it only stands to reason to reuse an existing quality model instead of beginning from scratch. Previous approaches similarly used these models to assess quality in information systems. The new aspect in this approach is to assess quality not only in terms of the object itself but also in terms of the requirements of the user. Similar to other recommendation systems, addressing the users' needs is also the main goal of this approach. Most approaches (try to) reach this goal by identifying objects whose content fits best with the users' requirements. Usually, documents, pictures, websites, as well as multimedia are recommended based on their content. This approach differs in that it identifies artifacts only on the basis of their qualitative status. The semantics of content is not relevant for this process. However, in order to have both topic-based recommendation and quality-based recommendation, hybrid approaches seem to be the most promising.

The identification process is based on pattern similarity. Algorithms that calculate pattern similarity are utilized for various purposes such as analyzing stock markets, face detection, and analyzing buying behavior in e-commerce. The algorithm reused in this approach was originally intended for identifying similarities in gene samples. This means that the algorithm was designed to handle huge datasets whereof this approach benefits in scalability and performance. In the course of this work, the algorithm has been modified, applied and evaluated in the context of detecting similar quality profiles. The aim of my approach is to find mechanisms for identifying meaningful resources. The use of these mechanisms, e.g. for proactive recommendation, is outside the scope of this work.

The abstract representation of users and resources in agile systems has characterized many potentially promising approaches. This work proposes a semantic model for representing the qualitative status of a resource as well as enriching the user profile with information about quality requirements.

1.2 STRUCTURE OF THE THESIS

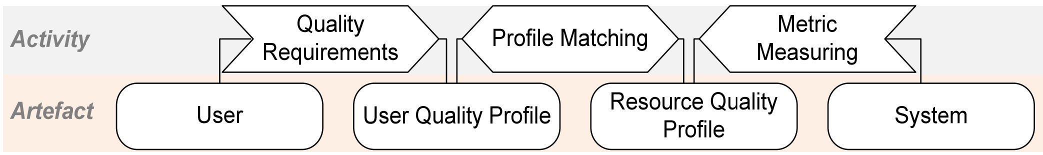


Figure 1-2 Schematic description of approach and structure of thesis

Figure 1-2 shows a graphical representation of steps and artifacts presented in this thesis. The various sections of this work are arranged according to this figure. The two-layered graph depicts entities comprising this approach on the lower layer and activities that transform those entities into other ones on the higher layer. The fundamental entities in this process are the *User* on one side and the *System* on the other. For both of these entities, an abstraction is created: the *User Quality Profile* and the *Resource Quality Profile*. The *User Quality Profile* is created by determining the quality requirements of the user. The *Resource Quality Profile* is created by measuring the system using metrics. The process of comparing *User Quality Profile* and *Resource Quality Profile* for ascertaining resources from the *System* that fit in with the *User's* needs is denoted as *Profile Matching*.

In the section *Social Server - Semantic Enhancement of the Social system*, I describe the technical system that I implemented so as to be able to assess quality in social software systems. The *Social Server* acts as an additional layer on top of the social system. The main task of the server is creating and maintaining the semantics, content and usage representation of the underlying system as well as providing a metric framework for a flexible and easy administration and adaption of quality measures.

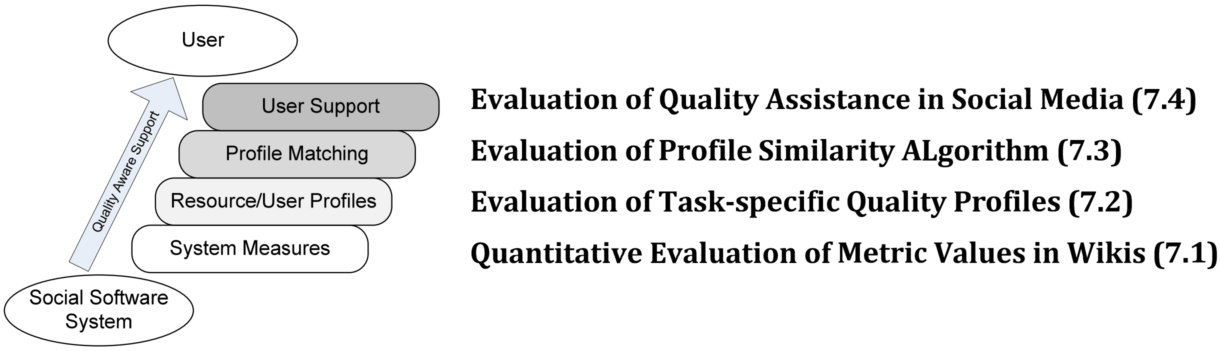


Figure 1-3 Multi-layered evaluation approach - the figure shows layers on top of the social software systems that are needed to provide quality aware user support

The evaluation of my approach is divided into several studies that describe the different layers required to provide quality aware user support. I have chosen the multi-layered evaluation approach in order to be able to utilize different evaluation paradigms depending on the layer of system abstraction.

Four studies evaluate the particular aspects of my approach (cf. Section 8). The section on *Quantitative Evaluation of metric values in Wikis* (8.1) presents a study aimed at measuring quality in a Wiki using quantitative metrics. The study on *Evaluation of task-specific quality profiles* (8.2) evaluates the pattern similarity algorithm used for pattern matching. The study on *Evaluation of profile similarity algorithm* (8.3) presents an empirical approach towards elicitation of user quality requirements as basis for creation of user quality profile. Additionally, the functionality described in this approach has been integrated into two prototypes for evaluating such an approach in a real world setting. The first prototype is focused rather on system representation and measuring quality using metrics. The second prototype focuses on the user quality requirements and recommendation of resources. Both are described in the study on *Evaluation of Quality Assistance in Social Media* (8.4).

2 Quality in Social systems

The salient points of this section have been published in
Weber, N., & Lindstaedt, S. N. (2011). A User Centred Approach for Quality Assessment in Social Systems. *Proceedings of the International Conference on Knowledge Management and Information Sharing 2011,* Paris, France

As an instance of social software systems, Wikis are prime examples of tools that allow for a collective construction of knowledge in a community setting (Cress & Kimmerle, 2008). There are certainly good examples of Wikis being used as tools for creating a collective online encyclopedia, for teaching and learning purposes, and for organizational knowledge management (Jaksch, Kepp, & Womser-Hacker, 2008; Majchrzak, Wagner, & Yates, 2006; Reinhold, 2006). In my opinion, Wikis are very well suited for analyzing content quality, especially because of the ease of editing content and the policy that everyone can edit anything. In particular, this feature often leads to information quality problems due to several reasons. Therefore, in the course of this thesis, I would often use Wikis as an exemplary instance of a social software system.

2.1 Quality perception in Wikis

The purpose of the social software systems like Wikis is to store and provide information. Usually, a web browser is used to interface the system. Quality problems may occur every time a user opens a Wiki page and the content does not correspond to his/her quality requirements. This can happen for various reasons: the article addresses another audience, meaning that the article was intended for readers from a different community, with a different educational background or for users conducting a different task. Quality problems can also be caused by a simple lack of expertise on the part of the content author or volitional quality problems in the case of vandalism.

Qualitative characteristics of an artifact that must be met in order to fulfill the user's quality needs are denoted as quality requirements. These requirements can be observed at different levels (cf. Figure 2-1). There are, for example, very common quality requirements that are generally valid. For instance, *freedom from error, accuracy* and *completeness* are important quality requirements that are relevant in almost all cases regardless of the current task or user (cf. evaluation in Section 8.2). On the other hand, there are also very individual quality requirements; for instance, quality requirements depending on the user's background knowledge or the user's experience in a certain area. An expert on a topic would assess the quality of an article differently from someone who is new to the same topic. Similarly, a child may have different quality requirements from an adult.

2.1.1 Common understanding of quality in communities

Besides these individual and general factors, quality can also be defined for and by a group of people. Communities also have, besides common goals and interests (Wenger, 1998), a common understanding of quality. In terms of communities in an organizational environment, quality requirements are driven by organizational quality standards if they are used internally. If the artifacts are intended for customers, customer requirements might be considered as well. Only in the rarest of cases are community quality standards explicitly represented and externalized in a guideline. Quality requirements can be seen rather as the common understanding of quality in a group which arises from the group's tasks and activities. In this case, quality perception can be framed by intentional aspects (Pipino, Lee & Wang, 2002).

Figure 2-1 Quality requirements in different contexts

A community (of practice) can be a division or project team in a company but also a community of interest on the web. In both cases, the quality of artifacts is prescribed by the quasi standard within the community. The difference between the community within an organization and the web community is the motivation to meet the quality standard. On web communities, a common interest for the most part encourages members to meet a quality standard while contributing (Zhang, 2006). Members of communities in an organizational environment try to (must) meet the organizational quality standard as part of their membership commitment.

2.1.2 QUALITY AWARE SYSTEM BEHAVIOR

Thus far, for interacting with social software systems, there is no context aware behavior. That means that the system reacts similarly to user input regardless of the user and his/her current context (Rath et al., 2008). The fact that different users or users in different contexts have varying requirements stays out of consideration. Traditional systems only decide on the user input. The objective of this work is to make information systems aware of quality. This means that the system adapts its behavior according to the quality requirements of the user or community. To attain this goal, this system has to process the particular quality requirements as input parameters for its algorithms. But the question is: at which level should the system meet the user’s quality requirements - individual, community or general level? At the general level, quality would be evaluated, independent of the user or community because general requirements are universally valid. Using these requirements would be contradictory to the idea of quality awareness since all users would get similar results. In contrast, focusing on individual quality requirements would mean fully supporting the individual user. The problem here is that the causative aspects of individual quality requirements such as competence, experience, environmental context, etc., are not easy to determine, especially in online systems. There are approaches to detect and model contextual information (cf. Lindstaedt & Christl, 2011; or Rath et al., 2008) but to adapt these approaches to the application area of this effort would go far beyond the scope of this work.

My approach presents a compromise between the general and individual perspective. People working together with the same interest have similar quality requirements for artifacts used in this context (cf. Evaluation). In organizations, teams work together on one task in small- and/or middle-sized groups. These can be seen as a community of practice with similar quality requirements. My approach is to provide quality awareness in social systems according to the quality requirements of a community.

2.2 MODEL FOR QUALITY ASSESSMENT IN SOCIAL SOFTWARE SYSTEMS

The quality of information is a very general and unspecified term. Juran (1992) defined information quality as data that is *fit for use in their* (users') *tasks*. This definition is unspecific but it suggests that quality is strongly connected to the user and her/his requirements (Francalanci & Pernici, 2004). If we assume that information quality can be measured by looking at the performance of a system which is based on that information (Ivanov, 1972), we still must acknowledge that for social media the performance can differ depending on the target groups: different users may assess the quality of one and the same Wiki article quite differently depending on their situation and current tasks. This means that since the objective is to assess the quality of social content, one can never assess quality without having information about the consumer of the social system data.

Apart from different factors influencing the perception of quality, the concept of quality itself, among most researchers, is understood as a concept with multiple dimensions (Pipino, Lee, & Wang, 2002; Francalanci, & Pernici, 2004). In the literature, the approach to define quality by regarding it as a set of quality dimensions is very popular. Knight and Burn (2005) provide a comparison of knowledge dimensions among various approaches. The importance of these dimensions for assessing the quality of a given resource is preset by the individual weighting of the quality dimensions. About 40 years ago, Ivanov (1972) proposed a set of intuitively selected quality dimensions The problem here is with the selection of quality dimensions that describe the best quality for the intended goal.

The proposed dimensions are:

Accuracy	Goodness
Value	Availability
Validity	Privacy
Dependability	Trueness
Integrity	Relevance
Correctness	Reasonableness
Precision	Pertinence
Timeliness	Acceptability
Freedom from Error	Refinement
Exactness	Approximation
Quality	Currency
Secrecy	Rightness
Usefulness	Accessibility
Confidentiality	Security
Consistency	Coverage
Authenticity	Completeness
Reliability	
Degree of Detail Recency	
Controllability	

Table 2-1 Quality attributes (Ivanov, 1972)

Quality and the related quality dimensions depend on several factors. One problem that seems to be a gap in quality assessment is the selection of these dimensions. How can we elicit dimensions that fit the users' current requirements?

Several approaches have been made in recent years. Wang and Strong (1996) categorized these attempts in three groups: **Intuitive approaches, theoretical approaches, and empirical approaches**. The next section discusses these three approaches and shows why in this context the empirical approach seems to be the most promising.

The **intuitive approach** includes quality dimensions that are intuitively chosen for quality assessment. For example, correctness is considered intuitively as a quality dimension given that the interconnection between quality and correctness seems obvious. The problem with this approach is that the selection of quality dimensions is based on an assumption of user requirements. The risk with this approach is to evaluate quality facets that have no high priority for the user, further leaving out facets that are vital. This approach is used very often in literature. For instance, Hu et al. (2007) propose an approach to measure the quality of articles in Wikipedia based on this intuitive assumption: “the higher the authority of the authors is, the better the article quality is” and “the higher the reviewer authority is, the higher the article quality is”. In Agichtein et al. (2008), these authors evaluate the quality of entries in answering portals such as Yahoo! Answers. They thus assume that an answer can be assessed by using the following dimensions: compositionality, style accuracy and soundness. In order to test the hypothesis, they compare the results of calculated metrics with human judgments. In another approach, which is also based on answering portals, Jeon et al. (2006) propose that “good answers tend to be relevant, informative, objective, sincere and readable”.

As seen in these examples, intuitively selecting quality dimensions results in various attributes which overlap in some cases or are completely different. But the most important fact is that these dimensions do not inevitably correspond to the user’s perception of quality. They are selected based on the intuition of authors’ quality. The evaluation of intuitively selected quality attributes presented in the literature often addresses only one facet of quality.

The second approach for selection of quality dimensions is the **theoretical approach**. In this approach, measures are derived from theoretical considerations. Often, models are used as the basis for describing the information system. Based on these, the model measures can be inferred. Usually, information that acts as the basis for an information system is measured on the results derived from using the system. Pammer and Lindstaedt (2009) describe an approach where new information, gained by inference over an ontology, is evaluated. The result of the evaluation enables the user to assess the quality of the underlying ontology. That is, the quality of an ontology is measured by its success in a real world scenario.

The **empirical approach** applies a user-centered elicitation of quality dimensions. Therefore, the quality perception of the end-user is evaluated using empiric methods in order to find out what aspects of quality are relevant for the community that uses the system. Wang and Strong (1996) present an empirical approach for elicitation of quality dimensions in an information system. Because perception of quality can be described at different levels of granularity, they propose a stepwise procedure. In a first survey, they collected 179 quality attributes directly from the end-users of the information system. The quality attributes represent a description of quality at a very low level of granularity. In the next step, the quality attributes provide the basis for 20 attribute clusters, denoted as quality dimensions. This step is required because the quality attributes are very user-specific with many overlaps between the attributes. Some of them lack discriminatory power and semantic ambiguity compared to other attributes.

The quality dimensions derived by clustering the attributes provide a more general perspective on quality. Wang and Strong (1996) present a further abstraction from quality dimensions to quality categories. In this way, the quality of a system can be described in only four quality categories.

The quality dimensions proposed by Wang and Strong (1996) provide the foundation for the quality modeling approach described in Section 4.

2.3 User support by providing quality awareness

In the course of trying to avoid if not prevent information quality problems, quality assessment can be used for various purposes. It can be the basis for recommendation of resources. Furthermore, it can be used for knowledge gardening (Weller & Peters, 2008) by detecting resources that do not match a certain quality standard. This section describes requirements providing the motivation for this work on the one hand, and possible solutions by means of creating quality awareness on the other.

2.3.1 Use Cases

The creation of use cases based on empirical data was part of the research and development activities in the EU FP7 project MATURE [1]. Several studies were conducted in this context (MATURE Consortium, 2009a, 2010). Amongst others, these studies resulted in use cases describing the desired system behavior or situation. One aspect in analyzing and supporting maturing processes (Schmidt, 2005) is the awareness and improvement of quality in several dimensions of organizational knowledge. Hence, several of the use cases focus on quality of artifacts. These use cases represent the motivation for this approach and at the same time the goal this work is aiming at. Considering the fact that the studies involved real end users, it emphasizes the relevance of the use cases in providing the motivational foundation of this work.

2.3.1.1 Use Case 1: Improve the quality of content artifacts

During the process of writing an article, the user's context is constantly analyzed using various indicators. These measure the resource status in several dimensions, e.g. readability for the content dimension or the semantic markup on the structure dimension. The results are shown to the user and appropriate recommendations are made by the system to improve the quality of the text. In case the user

[1] http://www.mature-ip.eu

experiences problems during writing, for example, if he/she is quite new to the field he/she is working on, the system offers the possibility of sending the newly created article to experts and to collect feedbacks from colleagues.

Problem situation, motivational aspects

A person is interested in improving the quality of articles, documents, reports, etc. This happens due to the user lacking enough experience or does not have enough knowledge on certain topics or is interested in improving his/her own work. Additionally, the user might want to contribute to articles or documents written by others.

Added Value

Supporting features that help users improve the quality of knowledge artifacts facilitate the knowledge maturing process. Quality improvement can happen, for example, by adding additional semantic mark up or improving the provided readability scores. Additionally, in case the user has certain questions, he/she is provided with a collaborative initiation service such that interested or more experienced colleagues may be asked to help out.

2.3.1.2 Use Case 2: Rate and assess the quality of artifacts

This use case describes how the system supports users in rating artifacts related to their quality and/or usefulness in certain situations. A user comes across artifacts of diverse quality during her work. The system helps her to rate and comment on these artifacts by considering the following aspects:

- Ratings relate to the quality and/or usefulness of the artifact
- Usefulness is influenced by the work context, e.g. the case a user is working on. Quality may also include features such as topicality or presumed general usefulness

- The usefulness of an artifact largely depends on the (work) context in which it is used. Therefore, the context in which the rating is given (e.g. a case the user is working on) should be preserved so that other users can judge whether the rating(s) really apply to their own current context.
- In addition, users should be enabled to comment on their ratings, as a basis for potential later discussions
- To overcome the cold-start problem of ratings, some artifacts can receive automatic initial ratings (e.g. from text mining)

Problem situation, motivational aspects

Good ratings prevent wasting time due to redundancy (including duplication), uncertainty of information value, incompleteness, incorrectness, un-conciseness. That is, users want to rate artifacts for their own interest by marking high quality artifacts from bad ones.

Added Value

Quality control is an important reseeding activity that prevents resources from being allocated to less effective maturing processes. It helps put the focus on the important ones. This "importance" is defined by the perceived usefulness of peers.

2.3.1.3 Use Case 3: Garden shared knowledge spaces

The use case describes the process of a person or group of people taking ownership of gardening shared knowledge spaces, e.g. in a folksonomy (Weller & Peters, 2008). This can also be for a certain (sub-) topic that is of relevance to the organization as a whole as well as initiating an organizational maturing process by refining collective knowledge and artifacts within this topic. When having loosely collected ideas and existing artifacts brought together, it is necessary to initiate refinement processes for the artifacts and their context including the surrounding semantic structures, to add missing links between topics and between artifacts, to merge similar ones and smooth inconsistencies, and to delete unused, outdated or incorrect artifacts.

Problem situation, motivational aspects

Problems in accessing knowledge/information/data are due to:

- missing links between artifacts
- distributed information over multiple artifacts, e.g. multiple Wiki pages for the same topic
- inconsistencies/contradictions between artifacts
- different structure, labeling, etc., deviating from users´ expectation

Added Value

A major barrier to wider dissemination is the lack of shared structures. But such shared structures - if they are useful - represent negotiated meaning, which usually grows in an evolutionary way. But this growth also needs reseeding activities to prune and improve the structure, and to reflect on interrelations.

2.3.2 EXEMPLARY SOLUTIONS IN QUALITY AWARE SYSTEMS

The three use cases mentioned above illustrate different situations in which quality awareness is necessary. In the subsequent section, examples of quality aware system functionality for supporting these use cases are provided. This section shows the possible impact of mechanisms provided by this approach without going too much into detail.

2.3.2.1 Information Search

The main purpose of a Wiki system is to act as knowledge base for a particular domain. The possibility of accessing information that corresponds to the knowledge of a whole community makes it an important information source on the internet. Especially in organizational contexts, Wikis are often used to make information available to employees and customers. In this case, it is important not only to provide information, but the ability of content creation for all users is also another important feature of the Wiki. However, at a certain point in time when the Wiki gets more and more articles, it gets more and more difficult to find information on the Wiki. Search functions included in the Wiki do not always find the appropriate information. Usually, Wiki-systems come with two included search mechanisms.

The **full text search** is included in most web systems. The characteristic of this type of search is that the user types in one or more words and gets back all articles or websites which contain such word(s). The drawback of this approach is that the user is only able to find content which s/he remembers to contain a particular word or s/he has to guess which words could be in the article.

The second way to find information on the Wiki is to **browse** through the content by following the links in the article. This is probably the more frequently used search method in small- and middle-sized Wikis. In many cases, the main page is used as navigation page that links to other content. The problem of this type of information search is that the result depends on the quality of the Wiki's link structure. Articles that have no incoming links from other articles cannot be found. If a user starts at the main page, only articles that are reachable by a path of links from that page can be found. Since the author of an article cannot directly influence the incoming links of the article, s/he cannot guarantee that the content would be easy to find.

Both approaches to information search on Wikis have two things in common: first, as mentioned, both do not always provide satisfying results. Second, the result of information search is the same for all users. In other words, for one given keyword the full text search provides the same results regardless of who is sitting in front of the computer and what the person is going to do. Since the perception of quality depends on several aspects, the same search results can be useful for one user and useless for another (cf. Section 2.1). Aspects that influence the perception of quality can be experience, the current task, and the user's competence level. In order to provide adequate search results, these aspects should influence the search so as to provide search results that correspond to the user's quality requirements.

One approach described in this work is to enhance the existing search mechanisms. In this, the two search mechanisms described above are enhanced with quality awareness. Therefore, the quality profile of the user acts in addition to the search term as an input parameter. Based on the quality profile of the articles, the system can filter and/or rank the search results according to the user's needs. That means, from the set of resources that result from a full text search, resources that match the user's quality requirements are ranked on top.

In the case of filtering, those results which contain the search term but do not correspond to the user's needs are deleted from the result set. So, the system would only present appropriate results in terms of the search term and quality requirements.

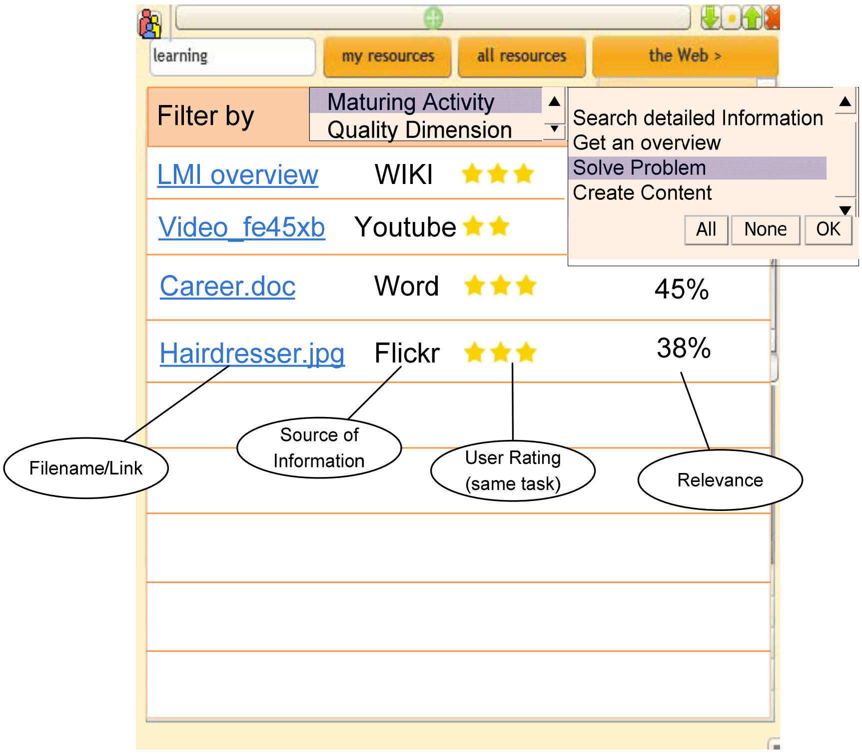


Figure 2-2 Quality aware ranking mechanism – quality dimensions as well as user's current task within a community, influencing the ranking of results

Just as well, the second search mechanism can be enhanced with quality aware search support. This exemplary solution is adapted from *adaptive hypermedia* (Brusilovsky, 1996). In this case, the user opens an article and checks the links on the page currently opened. This view corresponds to the presentation of search results because the user has to choose one of the links to browse further through the Wiki. Similar to the search results of a full text search, the result-set can be enhanced with information about how the quality of articles behind links relates to the quality expectations (De Bra & Calvi, 1998). As such, the user is aware of the quality of the linked articles before s/he opens the link. One possible way of reaching this goal would be to change the appearance of the link itself.

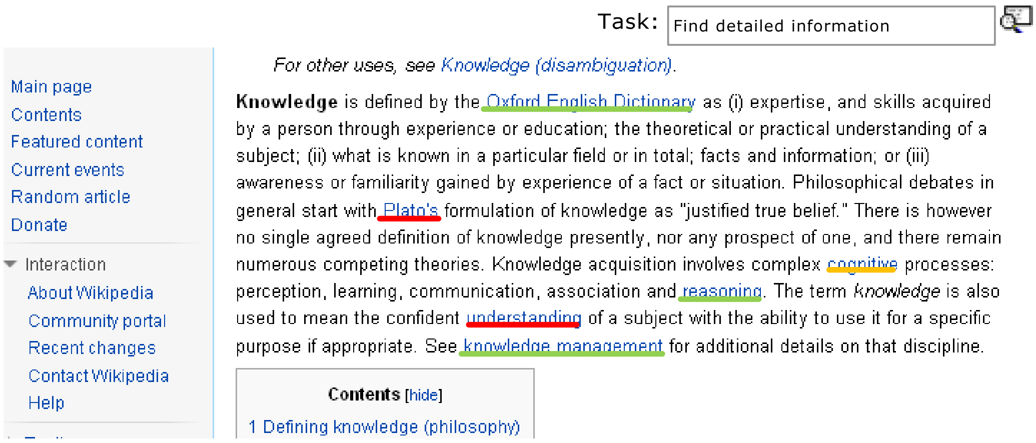


Figure 2-3 Link color based on resource quality status – how well the article that is linked fits in with quality requirements is indicated by the color of the link

Figure 2-3 shows how the links of a Wiki article are modified in order to provide information about the linked article's quality status in relation to the quality requirements.

2.3.2.2 Content creation support

Usually, the creation of content in Wiki systems happens in editors that are provided by the system. I assume that the author strives to create quality content – although other behavior is also known as well (Potthast, 2010). As mentioned in section 2.2, quality dimensions describe different facets of quality. In order to achieve a good quality article, these dimensions can act as indicator of the quality status. The objective is to create awareness of the current article's quality status by showing the status of the quality dimensions. The user is able to track the quality status of the content s/he is currently editing. So, s/he can assess the quality of the article.

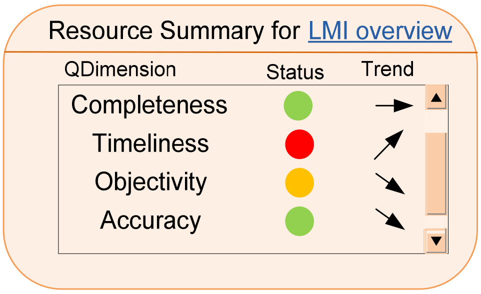


Figure 2-4 Resource quality status indicator

The information about the current quality status is provided by the resource quality model. Therefore, the approach requires that the model is updated in real-time during article editing.

Considering the role of the user quality profile, I distinguish two cases. First, the editor of the article writes for no special audience. In this case, the user quality profile is not considered. In the second case, the editor has a certain intended audience in mind; for instance, people searching for detailed information in order to write a scientific work. In this case, there is a particular quality requirement: the quality requirement of the community for which the article is intended. If it is known which quality dimensions are relevant, the user could use the quality status as guidance to write an article that suits the intended communities' quality requirements. This approach has been implemented and evaluated in *Evaluation of Quality Assistance in Social Media* (Section 8.4).

2.3.2.3 Gardening activity support

Wikis are often used as a base of common knowledge in companies and communities of interest. The Wiki is used to provide information about a particular product or topic of interest. Thereby, one characteristic of the Wiki systems is that (almost) all users are able create, edit and delete content. Due to this openness, errors occur in the system by reason of collaborative editing. This affects the content, but also the link structure and vocabulary. Usually, following the Web 2.0 idea, the community is responsible for correcting errors in the Wiki.

But there are some cases where an administrative task of correcting errors may be required. In the case of syntactic errors and typos within the articles, the community is able to correct them. But some situations require a system-wide prospect which is normally given only to administrators. For instance, if one category is given to an article, the community can decide whether it suits the article or not but the community cannot identify other categories with similar semantics that should be combined in order to avoid any ambiguity among the categories. So the task of gardening (Peters, Weller, 2008; Udell, 2010; Weller & Peters, 2008) is to conduct administrative tasks as a community member. Some systems have special functions to support gardening and to give community members the overview required for doing gardening.

Quality awareness in social systems provides a significant contribution to identifying gardening needs. Since syntactic errors, inconsistencies in the vocabulary, inadequate linking and typos all have an impact on the quality profile of the resources, quality awareness supports finding gardening needs. Based on the quality profile of resources, there can only be given hints as to where gardening is needed and where the gardening task itself must be done by the user. In this context, two scenarios are possible; the user triggered search for gardening needs and the proactive gardening recommendation. In the first case, the user decides to weed out errors in the system and asks for articles that have a low assessment in a particular quality dimension in order to improve it. In the second case, the system requests the user to correct errors when it detects them.

3 CONCEPTUAL FOUNDATION AND RELATED APPROACHES

This section gives an overview of existing approaches. These approaches either provided a conceptual/technical foundation for this work or discussed alternative approaches. Figure 3-1 depicts the composition of this section.

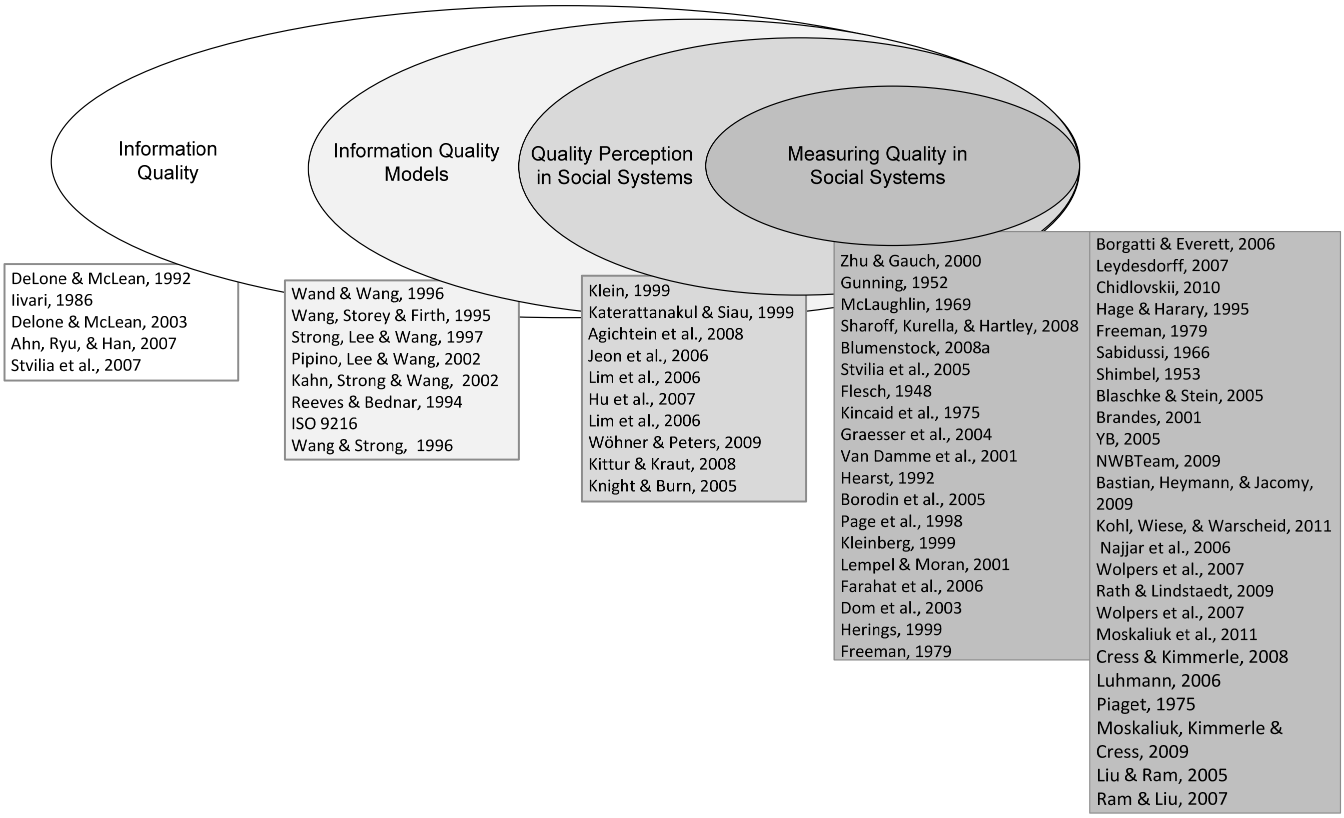


Figure 3-1 Specialization of related work topics

The first part addresses information quality in general as a conceptual frame of this work. The second part narrows the field down to models of information quality. Based on these models, the third part focuses on how quality is perceived in social systems. The fourth path focuses on approaches to measuring quality according to the perception of the user.

This section gives an overview on related research and approaches upon which this work is based. The definition of knowledge, information and data varies among disciplines. Since work from various disciplines is cited in the subsequent section, some address *quality of data*, others *quality of information*. If there is no explicit remark, *information* and *data* are interchangeable and mean the same in this context. Within this work, I use the concept of *information quality.*

3.1 INFORMATION QUALITY

3.1.1 RELATED WORK

Quality in information systems is often denoted as a factor for the success of the system. DeLone and McLean (1992), in *Information Systems Success: The Quest for the Dependent Variable*, present a model that identifies activators and inevitable prerequisites for system success. The authors differentiate between several dimensions related to information system success: S*ystem Quality, Information Quality, Information Use, User Satisfaction, Individual Impact* and *Organizational Impact*. In this way, *System Quality* is defined as quality on a technical level and *Information Quality* as quality on a semantic level. In the case of *Information Quality* and *System Quality*, the authors differentiate between measuring the output of a system (information) and measuring the system itself.

The proposed dimensions are the result of an empirical analysis of 180 articles. The basis for the dimension of *Information Quality* is adopted from the *Iivari-Koskelka satisfaction measure* (Iivari, 1986) which includes three measures for information quality: *informativeness (relevance, comprehensiveness, recentness, accuracy, credibility), accessibility (convenience, timeliness interoperability),* and *adaptability.* Furthermore, the authors provide an extensive overview of empirical measures for information quality. The updated IS success model (Delone & McLean, 2003) adds another factor for system success: *service quality.* Due to the fact that nowadays more and more organizations shift their field of action from information providers to service providers, another possible output of information systems has been added to the model: services.

As the updated model of Delone and McLean shows, the model was applied and adapted in several cases. System success is particularly important in e-commerce. Ahn, Ryu, and Han (2007), in *The impact of Web quality and playfulness on user acceptance of online retailing,* describes the interconnection between information quality and acceptance. The foundation of this work builds up the quality categorization into system quality, information quality and service quality. These three concepts of quality are described as follows: **System quality** as engineering-oriented performance characteristics (rating interface design, functionality, response time, etc.), **Information quality** with both engineering and operational characteristics (rating data format, completeness, timeliness, etc.), and **Service quality** which refers to the availability of communication mechanisms for accepting consumer complaints and their timely resolution with responsiveness, assurance, and follow-up services (i.e., how well a delivered service level matches customer expectations). In order to evaluate the relationship between quality and user acceptance, a study of 942 users has been conducted. During this study, the following quality facets have been evaluated:

- System quality
- Information quality
- Service quality
- Playfulness
- Perceived ease of use
- Perceived usefulness
- Attitude toward use
- Behavioral intention to use

The results showed that customers' intention to visit an online retailing site depends on its playfulness as well as two beliefs: perceived ease of use and usefulness.

In *A framework for information quality assessment* (Stvilia et al., 2007), the authors describe prerequisites that are required in order to assess the quality of information. As one interesting aspect, the authors pursue the question of what leads to information quality problems. They name five different sources for variance in information quality: *mapping, changes to the information entity, changes to the underlying entity* or *condition,* and *context changes.* Thereby *mapping* means unclean, ambiguous or wrong relations in and between information entities, whereas the other sources address the domain and the ever-changing context of information and thus require permanent update. Understanding the sources of information quality problems helps identify the underlying reason in a particular case. In addition, the authors propose a model for quality by proposing quality dimensions and the corresponding metrics.

The next section discusses different approaches to quality models and shows some differences but more notably their similarities.

3.1.2 Relevance for this work

The study of Delone & McLean (1992) provided substantial motivation for this work. The result of this empirical study shows the importance of quality improvement for all types of information systems. However, dividing quality as a whole into information quality, system quality, etc., could not be retraced due to the strong interferences between these dimensions. Assessment of quality in the context of my work means assessing the output of a socio-technical system as an information system.

However, the scope of this thesis only covers information quality's related aspects but the paper shows that these are crucial factors for system success. The quality dimensions providing the foundation for this work do not distinguish between information quality, system quality and service quality. My approach, as presented in this thesis, is rather user-driven since it assesses the quality perception of the user which is always a combination of experiencing both system and information.

3.2 QUALITY MODELS

The question what quality of information is and how it can be measured has a long tradition. Defining quality as a concept is very complex, hence in most of the papers and also in this work, quality is considered as a multidimensional concept. Dividing quality into several facets makes it easier to understand and assess the quality of information.

One problem in identifying facets is to decide whether they are relevant and respectively if a discovered set is complete. Missing quality facets may lead to ignoring important aspects of quality assessment.

3.2.1 RELATED WORK

In Wand and Wang (1996), *Anchoring data quality dimensions in ontological foundations,* the information system is defined as a representation of the real world. The approach towards identifying relevant dimensions describing information quality is based on a number of citations in the literature. When measuring quality, authors often intuitively chose dimensions that are indicators of quality. This effects a specialization of the quality concept and makes quality easier to measure/evaluate. The authors present a set of dimensions based on the literature review (Wang, Storey & Firth, 1995). This work can be seen as a starting point in identifying a complete set of quality dimensions.

Dimension	# cited	Dimension	# cited	Dimension	# cited
Accuracy	25	Format	4	Comparability	2
Reliability	22	Interpretability	4	Conciseness	2
Timeliness	19	Content	3	Freedom from bias	2
Relevance	16	Efficiency	3	Informativeness	2
Completeness	15	Importance	3	Level of detail	2
Currency	9	Sufficiency	3	Quantitativeness	2
Consistency	8	Usableness	3	Scope	2
Flexibility	5	Usefulness	3	Understandability	2
Precision	5	Clarity	2		

Table 3-1 Quality dimensions used in the literature (Wang et al., 1995)

Another approach to identifying quality facets is described in *Beyond accuracy: What data quality means to data consumers* (Wang & Strong, 1996). This work provides the theoretical foundation for this work regarding the facets of quality. Wang and Strong (1996) argue that quality can be described at different levels of granularity. At the lowest level is quality attributes. Their fine granular nature allows for mapping them directly to metric values. Quality attributes are clustered to quality dimensions which in turn can be grouped into quality categories. These categories provide the four main aspects of quality. The quality attributes and their generalizations are gained from two studies with a total of 135 participants. The result is the division of quality into the categories of *Intrinsic Data Quality, Contextual Data Quality, Representational Data Quality* and *Accessibility Data Quality.* These categories are depicted in the following empirically corroborated model:

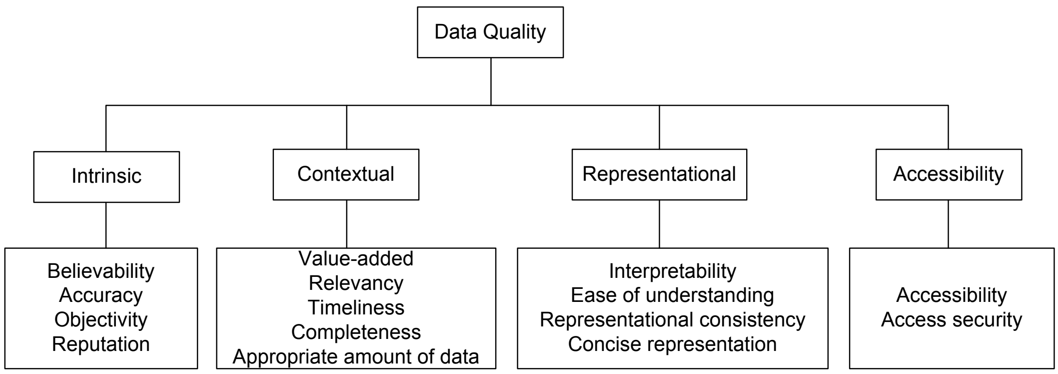


Figure 3-2 Conceptualization of quality (Wang & Strong, 1996)

The preceding papers focus mainly on quality in general. But in an organizational context in particular, lacking information quality is associated with high costs. In Strong, Lee and Wang (1997), *Data quality in context,* the authors approach information quality from an organizational perspective. The main focus of their paper is to develop strategies for improving information quality. Therefore, 42 projects addressing quality improvement in different organizations have been studied by means of interviews. For assessment of the qualitative status of information, the quality dimensions in Wang and Strong (1996) have been adopted.

For effectively measuring quality, a tool for measurement is required. Pipino, Lee and Wang (2002) in *Data Quality Assessment,* discuss the importance of selecting metrics that fit the application scenario. In this approach, data quality is assessed from two perspectives, the subjective assessment and the objective assessment, whereby objective means measuring quality using metrics. Depending on the data and the context in which data quality is assessed, the development of metrics plays an important role. The paper shows how metrics can be evaluated and mapped according to quality dimensions. On the other hand, the process of subjective quality assessment and mapping between subjective and objective assessment in order to identify discrepancies is described. The normalization and interpretation of metric values, described in section 5.5, is based on the content of this paper.

Organizations strive to provide good quality information to their customers. Therefore, one prerequisite is awareness of the qualitative status of information, similar to other goods that are produced and benchmarked. Kahn, Strong and Wang (2002), in *Information Quality Benchmarks: Product and Service Performance,* propose a model for benchmarking information quality. The model is referred to as product and service performance model for information quality (PSP/IQ). In this context, the authors define good information quality as conformance to specification and exceeding customer expectations. In this definition, the two additional views on quality – excellence and value – mentioned in Reeves and Bednar (1994) are ignored by the authors due to their non-measurability. Similar to the Model of Information System Success (Delone & McLean, 2003), quality is divided into different areas. Whereas the IN Success model divides quality into 8 areas, the PSP/IQ model distinguishes 2 areas: Product Quality and Service Quality. According to the definition used in this context two conditions must be met: information should conform to the specification, and exceed (or at least meet) the expectations of the user. This results in a two-dimensional model. In order to assess the status of each field in this model, the quality dimensions presented in Wang and Strong (1996) have been adopted.

	Conforms to Specifications	Meets or Exceeds Consumer Expectations
Product Quality	Sound Information • Free-of-Error • Concise Representation • Completeness • Consistent Representation	Useful Information • Appropriate Amount • Relevancy • Understandability • *Interpretability* • *Objectivity*
Service Quality	Dependable Information • Timeliness • Security	Usable Information • Believability • Accessibility • Ease of Manipulation • Reputation • Value-Added

Figure 3-3 Specified conceptualization of quality (Reeves & Bednar, 1994)

A study has been conducted to test if the model can be used to evaluate the quality of information. Thus, 75 employees in three organizations filled in a questionnaire assessing the different quality dimensions on a Likert scale. The result of the study is that the model is useful for comparing the qualitative status of information across organizations.

Apparently, information quality is also a crucial topic for standardization. For this reason, there is an international standard for information quality (ISO 9216). The ISO quality model has three sub-models of software product quality (internal quality, external quality, and quality in use), 10 quality characteristics, 24 sub-characteristics, and more than 250 measures proposed to quantify these quality characteristics and sub-characteristics. Internal and external quality have the same characteristics and sub-characteristics. The difference is in the "quality" measures. Quality in use has no sub-characteristics. The characteristics, proposed in the standard, are:

Characteristics	DATA QUALITY	
	Inherent	System dependent
Accuracy	X	
Completeness	X	
Consistency	X	
Credibility	X	
Currentness	X	
Accessibility	X	X
Compliance	X	X
Confidentiality	X	X
Efficiency	X	X
Precision	X	X
Traceability	X	X
Understandability	X	X
Availability		X
Portability		X
Recoverability		X

Figure 3-4 ISO standard for information quality (ISO 9216)

3.2.2 *RELEVANCE FOR THIS WORK*

The previous section impressively shows the similarity between the existing models. The perception of quality as a multifaceted concept is widely accepted. This insight led me to the model of quality layers described in Section 4. However, considering the different models, the question arises as to which one is the most relevant. The main difference between the models is the number and granularity of quality dimensions. Hence, all models are similarly relevant as long as the dimensions are selective and have no big overlaps. In my approach, the evaluation of the used quality dimensions' selectivity has been evaluated in Section 8.1. This work is based on the quality model of Wang and Strong (1996) due to its empirical establishment and its wide acceptance in information quality research.

3.3 Information Quality in Social Media

The previous section approaches information quality from a rather traditional standpoint. Information systems are considered as systems that are used in organizations to manage and provide knowledge. In recent years, another type of information systems has emerged: systems that are accessed online and maintained by communities. Information in these systems is often denoted as social media. Due to the nature of user-contributed content, the need for mechanisms to assess information quality has increased. The work presented in this section addresses the assessment of quality in social media.

3.3.1 Related Work

Since information about almost every topic is available on the web, user-generated content is widely used for information search. More than traditional text sources, the user often has to deal with defective content. Information from the web is only useful if the reader knows the author's level of expertise. Klein (1999), in *Perceptions of Information Quality: A Study of Internet and Traditional Text Sources,* evaluates how quality from different sources is perceived.

Most content on the web is represented on websites. The quality of a website is critical for their success. Katerattanakul and Siau (1999), in *Measuring information quality of web sites: development of an instrument,* evaluate which aspects of websites affect quality. In case of websites, not only does the quality of the provided information play an important role but also the presentation of the information as well as related technical aspects. Based on the model of Wang and Strong (1996), they identified important aspects for website quality in each of the four quality categories: Intrinsic quality, contextual quality, representational quality, and accessible quality. The following model is derived from these considerations:

Intrinsic	• Accuracy and errors of the content • Accurate, workable, and relevant hyperlinks
Contextual	• Provision of author's information
Representational	• Organization, visual settings, typographical features, and consistency • Vividness and attractiveness • Confusion of the content
Accessibility	• Navigational tools provided

Figure 3-5 Interpretation of quality categories (Katerattanakul & Siau, 1999)

The model was tested in an individual website context using a questionnaire. The reliability tests indicated that it is a highly reliable instrument for quality assessment. The factor analysis reveals that the developed instrument is fairly consistent with the research framework. Two of the four factors (contextual and accessibility), emerging from the factor analysis, are consistent with two information quality categories in the research framework. The results, however, suggest that the other two factors (intrinsic and representation) may be merged when used in individual or personal Web context.

The paper by Agichtein et al. (2008), *Finding high-quality content in social media,* describes an approach for assessing information quality on the web. The system on which the author focuses is the question and answering portal YAHOO! Answers. The approach of the paper is to evaluate the quality of the answers provided in this system. Therefore, three research questions ought to be dealt with:

1. What are the elements of social media that can be used to facilitate automated discovery of high-quality content? In addition to the content itself, there is a wide array of non-content information available, from links between items to explicit and implicit quality rating from members of the community. What is the utility of each source of information to the task of estimating quality?

2. How are these different factors related? Is content alone enough for identifying high-quality items?

3. Can community feedback approximate judgments of specialists?

In order to assess quality and to answer these questions, quality is measured in several dimensions. The first dimension is the intrinsic content quality. Thus, the content of questions and answers has been analyzed. The second dimension is the social network of users. Here, the authority of the answering person gives information about the answer’s quality. The third dimension is usage of the system. Click statistics and dwelling times are used to assess quality in the system. In order to obtain information about the question and answering process, the interactions within the systems were represented in activity models. The model shows which actions can be done in the system and how one can derive qualitative information from these resources. Based on this model, the authors developed a framework for classifying questions and answers in two groups: high quality and normal/low quality. An evaluation of the classifier based on a YAHOO dataset shows the accuracy of the classification. Another comparable approach is Jeon, Croft, Lee and Park, (2006) in *A framework to predict the quality of answers with non-textual features*. Similar to the previous approach, the authors develop a framework for qualitative assessment of question-and-answer pairs in QA systems. The difference in this approach is to infer the qualitative status from textual features. In the light of which, 13 features have been identified: Answerer's Acceptance Ratio, Answer Length, Questioner's Self-Evaluation, Answerer's Activity Level, Answerer's Category Specialty, Print Counts, Copy Counts, Users' Recommendation, Editor's Recommendation, Sponsor's Answer, Click Counts, Number of Answers, Users' Dis-recommendation. As we can see here, the selection of possible features depends on the data captured and provided by the system. The more metadata (e.g. explicit user ratings) is provided, the more features can be evaluated. The features were evaluated on 6.8 million QA pairs retrieved from the Korean Q&A service, Naver. In addition, 125 randomly selected search queries were collected. The objective of the evaluation part is to demonstrate that an existing retrieval mechanism can be enhanced by using additional features. The result shows that non-textual features improve search quality. Results from these two studies assessing QA quality can hardly be compared to this work since they do not proffer any definition of quality. Both are not oriented toward one of the quality models (Agichtein et al. 2008) that define quality in what Wikipedia users tag as *high quality*. Jeon et al. (2006) reckon that *good answers tend to be relevant, informative, objective, sincere and readable.*

Another approach that is also based on the authority of users provides a model for assessing the quality of user generated content. In Lim et al. (2006), *Measuring qualities of articles contributed by online communities,* user-created articles in systems like Wikis are analyzed. The authors' approach for assessing quality is based on the *mutual reinforcement principle.* Two assumptions are at the foundation of this principle:

- Quality: An article has high quality if it is contributed by a high authority author
- Authority: A contributor has high authority if s/he contributes to high quality articles

The authors define two models: the basic model and the peer review model. The basic model assumes that the higher the authority of the authors, the higher is the quality of the resulting content. The authority of contributors increases with the amount of created content. The second model, the peer review model, takes into consideration that editors improve only the parts of an article that is of low quality. This implies that editors perceive the quality of the residual content as sufficient. And so the authority of the editing user can be applied to the whole content. The two models have been evaluated on a dataset of 77 Wikipedia articles containing information about different countries. The reason for this selection is the high number of contributors to these articles. Some of them have up to 1000 contributors. This facilitates tracking changes in article quality based on editing behavior.

Hu et al. (2007), in *Measuring Article Quality in Wikipedia: Models and Evaluation,* present models similar to Lim et al. (2006) but they add one more aspect. In addition to the Basic model and the Peer Review model, they introduce the ProbReview model. In contrast to the Peer Review model, the ProbReview model assumes that an editing user is not aware of the qualitative status of the residual content. The mentioned use case concerns an editor who changes only part(s) of a document without knowing the content of the whole article. In this model, the authors introduce a probability for the case that the contributor knows the document content. So ProbReview acts as a refinement of the Peer Review model. The evaluation of the model shows that this refinement reflects the real editing behavior better than the Peer Review model.

Another interesting approach is the evaluation of usage during the lifecycle of an article to assess quality. In Wöhner and Peters (2009), *Assessing the quality of Wikipedia articles with lifecycle based metrics,* low quality articles and high quality articles are identified based on their usage patterns. Therefore, the authors introduce two new metrics. These are based on tracking the changes in the editing intensity throughout the entire existence of an article; the authors refer this as the lifecycle of the article. The idea of new metrics is to measure the *persistent contribution* and the *transient contribution* in order to assess quality. The transient contribution refers to the number of words which were changed and reversed in the same given period of time. The persistent contribution refers to all effective edits which remain in the article beyond the period. To evaluate their metrics, the authors use the user-given quality categories of a Wikipedia dataset. These are: featured article, good article, and article for deletion. The evaluation ascertained a high level of effectiveness for quality measurement of the lifecycle-based metrics, in particular of metrics related to persistent contribution. For this work, this observation has led to the tracking of quality value development over a longer period in order to identify quality problems based on usage pattern.

The preceding papers show that usage patterns can be an indicator for quality. Some assumptions imply that the number of editors has an immediate effect on article quality. The following paper evaluates the group dynamic editing behavior in detail. Kittur and Kraut (2008), in *Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination,* state that the size of the group of editors is not exclusively decisive for article quality. The crucial factor is the coordination in the group. Wikipedia and most other Wiki systems provide, besides their collaborative editing functionality, the ability to discuss the contributions. The authors of the mentioned papers assume that these discussions foster the qualitative improvement of articles created by groups of users. This can happen explicitly in the already mentioned discussion pages but also implicitly within the article itself by giving structure or force a particular direction.

Possibly, the most exhaustive compendium of quality models is provided by Knight and Burn (2005) in *Developing a Framework for Assessing Information Quality on the World Wide Web– The Big Picture - What Is Information Quality?* The authors compare information quality frameworks beginning in 1996 until 2006. The outstanding approaches are abstracted above. As to the approaches listed in the paper, an astonishing fact is that all resulting frameworks are quite similar. That means there is a common understanding for considering quality as a multidimensional concept. Even the nature of the dimensions is widely accepted. But it is not only the finding that quality is multidimensional; another fact is the multi-granularity of the concept quality. Quality can be described at high granularity levels – like quality categories – but also at low levels like quality attributes.

3.3.2 Relevance for this work

The previous section describes approaches for measuring information quality in social systems (socio-technical systems). All these approaches understand quality as a multifaceted concept and try to assess one or more of these dimensions. For this reason, the approaches are comparable to my work. The main difference is that most of these approaches focus on one particular system. They measure Wikipedia, Yahoo! Answers, and other popular systems. For my work, these approaches are the basis for a more general approach with less focus on a particular system. Some concepts, like considering hubs and authorities for assessing the relevance of a node or involving the negotiation process into quality evaluation, have been directly adopted in my approach since they are system-independent.

3.4 Metrics

The previous sections have addressed the rather higher levels of quality. The following literature describes how quality can be measured at a low level of granularity. Metrics can directly be applied onto system representations to measure features of a system that are descriptive for the system, respectively its information quality. One important question regarding metrics is their interpretation. In other words, how can a mapping between low granularity quality facets and metric values be established. One option is to evaluate empirically the correlation between the metric value und the user perception of the quality facet. The following literature review gives an overview of which metrics are implemented and evaluated in research as a tool for measuring quality.

3.4.1 Related Work

One example for a mapping between metrics and quality dimensions is given in Zhu and Gauch (2000) in *Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web.* The authors provide an approach for assessing the quality of websites. They identify six quality aspects and their metrics to measure them.

The following list shows the assignment between quality dimensions and metrics:

Dimension	Metric Description
Currency	time stamp of the last modification of the document.
Availability	number of broken links on a page divided by the total number of links it contains
Information-to-Noise Ratio	total length of the tokens after preprocessing divided by the size of the document
Popularity	The number of links pointing to a Web page was used to measure the popularity of the Web page. The information about how many links that point to a particular Web page was obtained from the AltaVista (1999) site
Cohesiveness	determined by how closely related the major topics in the Web page are
Authority	The authority of a Web page was based on the Yahoo Internet Life (YIL) reviews, which assigns a score ranging from 2 to 4 to a reviewed site. If a Web page was from a site that was not even reviewed by YIL, its authority was assumed to be 0.

Figure 3-6 Mapping between quality dimensions and metrics

As this example shows, there are multiple dimensions to which metrics can be applied. The ones that are most prominent and also serve as the basis for my approach are: content, structure and usage. The following literature review follows these three dimensions as well.

3.4.1.1 Content

The objective of qualitative approaches for assessing content quality is to measure qualitative conditions like average number of sentences to make a qualitative statement.

Very early on, Gunning (1952) discovered the interconnection between quantitative measures for text and readability. The result of this observation was the formulation of the Gunning FOG Readability Formula. Since then, many other approaches have been published but all of them have one thing in common: the formulas are based on quantitative text measures like sentence length, number of words, average number of syllables, etc. For instance, the Gunning Fog Index is calculated as follows:

$$\mathrm{GFI} = \left(\frac{\mathrm{W}}{\mathrm{S}} + \mathrm{D}\right) \cdot 0{,}4$$

where W is number of words, S the number of sentences, and D the number of words with more than three syllables. Another more sophisticated approach is presented and evaluated by McLaughlin (1969) in *SMOG grading: A new readability formula*. The grading value is calculated as follows:

1. Count 10 consecutive sentences near the beginning of the text to be assessed, 10 in the middle and 10 near the end. Count as a sentence any string of words ending with a period, question mark or exclamation point.

2. In the 30 selected sentences, count every word of three or more syllables. Any string of letters or numerals beginning and ending with a space or punctuation mark should be counted if you can distinguish at least three syllables when you read it aloud in context. If a polysyllabic word is repeated, count each repetition.

3. Estimate the square root of the number of polysyllabic words counted. This is done by taking the square root of the nearest perfect square. For example, if the count is 95, the nearest perfect square is 100, which yields a square root of 10. If the count lies roughly between two perfect squares, choose the lower number. For instance, if the count is 110, take the square root of 100 rather than that of 121.

4. Add 3 to the approximate square root. This gives the SMOG Grade, which is the reading grade that a person must have reached if he is to understand fully the text to be assessed.

Although this reading score is rather difficult to calculate manually, there are already algorithms that automatically calculate the value for a given text. Hence, readability is a frequently used tool for assessing the quality of text in information systems. In Sharoff, Kurella, & Hartley (2008), *Seeking needles in the Web haystack: Finding texts suitable for language learners,* several metrics are evaluated in order to assess text difficulty. At this, the authors evaluate not only features relevant for text assessment but also the peculiarity of different languages regarding these features. One result of this paper is that features of low complexity often perform best. This observation is also made in Blumenstock (2008), *Size Matters – Word Count as a Measure of Quality in Wikipedia*. In this contribution, the easily measurable metric *word count* is applied to assess the quality of Wikipedia articles. Similar to the previously described approaches, the author uses the user-given assessment for Wikipedia (featured/not featured) as quality standard. The approach resulted in an astonishing accuracy of 96.31% just by classifying articles with more than 2000 words as featured articles. The accuracy of this result is comparable to complex classification algorithms like k-nearest neighbor (96,94%), logit model (96,74%) or random-forest (95,8%). Stvilia et al. (2005), in *Information Quality Discussions in Wikipedia,* presents an approach based on reading scores. Which reading scores are used exactly is not clear in the paper because the authors denote the metrics as *Flesch* and *Kincaid Reading Score* but reference Gunning (1952). I assume that the authors mean Flesch (1948) and Kincaid et al. (1975). However, the approach shows that the quality of Wikipedia articles can accurately be classified based on readability measures.

Graesser et al. (2004), in *Coh-metrix: analysis of text on cohesion and language,* proposes a framework of more than 200 content metrics. The objective is to provide tools for measuring text cohesion which can be understood as one quality dimension. The framework contains simple metrics like word count, sentence length, but also rather complex metrics like *Proportion of content words that overlap between adjacent sentences* and the reading scores *Flesch Reading Ease*

Score and *Flesch-Kincaid Grade Level.* Additionally, the authors present 18 metric classes for content metrics (a detailed description can be found in the paper):

- LSA space: latent semantic analysis as statistical representation of word and text meaning
- Word information: particular characteristics of the words in text
- Word frequency: frequency of particular words occurring in a certain language compared to the frequency of words in text
- Part of Speech: identification of particular word types: nouns, lexical verbs, adjectives, adverbs, prepositions, determiners, and pronouns
- Density Scores: incidence, ratio or proportion of particular word classes or constituents in the text
- Logical Operators: frequency of logical operators like *or, and, not, and if–then*
- Connectives: finding keywords indicating connectives such as: (1) clarifying connectives such *as in other words* and *that is*; (2) additive connectives such as *also* and *moreover*; (3) temporal connectives such as *after*, *before*, and *when*; and (4) causal connectives such as *because, so,* and *consequently*
- Type: Token Ratio, i.e. each unique word in a text is a word type. Each instance of a particular word is a token.
- Polysemy and Hypernym: polysemy is measured as the number of senses of a word (ambiguity). The hypernym count is defined as the number of levels in a conceptual taxonomic hierarchy that is above a word.
- Syntactic Complexity: syntactic complexity involves a number of metrics that assess how difficult it is to analyze the syntactic composition of sentences.
- Readability: assessing text difficulty by means of readability formulas.

3.4.1.2 Structure

The previous section describes metrics based on content. Most of them measure features based on textual content. One important piece of information is disregarded by these metrics, namely, how the content objects are related to other artifacts. There are several possible types of relations an object can have. Mostly investigated in research are two types of relational links between websites in hyperlinked environments and connections between people in social networks. The subsequent literature review presents papers describing structure metrics and papers showing the importance of these metrics for quality considerations.

In link graphs, mostly hyperlink structures (as websites and their links as relations) are considered. This can be a Wiki page but also an item in a Folksonomy (Damme & Coenen, 2001) or a text pattern (Hearst, 1992). Since the links are directed, a node can have two types of relations: incoming and outgoing relations. In addition, the node from which or to which a link is made plays an important role. In Borodin et al. (2005), *Link analysis ranking: algorithms, theory, and experiments,* metrics that deal with these features are extensively analyzed. Particularly, the group of metrics that considers the importance of linking notes (hubs and authorities) plays a crucial role in today's web (Google).

The authors of the paper analyze and compare existing metrics and identify potential weaknesses of the algorithms. The algorithms analyzed in this work are: InDegree (Borodin et al., 2005), PageRank (Page, Brin, Motwani, & Winograd, 1999), HITS (Kleinberg, 1999), SALSA (Lempel & Moran, 2001). All of these algorithms are based on the link graphs and thus are assigned to the structure representation. *PageRank, HITS* and *inDegree* are also used and evaluated within this thesis. Furthermore, the authors propose some new algorithms; some of them variations of hubs/authorities algorithms, some based on a Bayesian approach. In the evaluation section, experiments are described comparing rankings of the different measures with several queries. The authors conclude thus: "*We observed that some of the theoretically predicted properties (for example, the TKC effect for the HITS algorithm) were indeed prominent in our experiments. We were* *surprised to discover that some of the "simpler" algorithms, such as INDEGREE and BFS, appear to perform better than more sophisticated algorithms, such as PAGERANK and HITS".* Another contribution that deals with the same algorithms is Farahat et al. (2006), *Authority Rankings from HITS, PageRank, and SALSA: Existence, Uniqueness, and Effect of Initialization.* The authors analyze the behavior of the four ranking algorithms and recommend modifications to avoid unexpected behavior. Furthermore, they characterized the situations on which some algorithms fail.

Dom et al. (2003), in *Graph-based ranking algorithms for e-mail expertise analysis,* show the practical application of these algorithms. In this work, the graph represents the email correspondence between people. People as senders and receivers are nodes, while edges represent email communication from one person to another. The objective of this research is to determine the expertise of people based on this social network. Therefore, the performance of different graph-based ranking algorithms is compared. In order to find experts, emails have to be divided in two groups: emails requesting information, and emails providing information. People that provide information often are defined as experts. The compared algorithms are:

- Affinity: naïve approach that counts the absolute number of edges for each node
- Successor: based on the assumption that relations always lead from higher expertise to lower expertise, while all successors of a node are ranked lower than this node regarding the expertise
- PageRank: (cf. Page et al. 1998)
- Positional Power: (cf. Herings 1999)
- HITS: (cf. Kleinberg 1999)

The evaluation shows that PageRank and Successor perform best and HITS was significantly worse than these two. In terms of quality, expertise in this work is seen as one aspect of personal quality.

Besides algorithms that use indegree, outdegree, hubs and authorities for measuring graphs, there is a class of algorithms that consider the position of a node in the graph. Centrality measures were originally used to determine the importance of people in social networks. Freeman (1979), in *Centrality in social networks conceptual clarification,* presents an extensive overview of the importance and calculation of centrality measures in social network analysis. But centrality is not only applicable to social networks but also to content networks. Borgatti and Everett (2006), in *A Graph-theoretic perspective on centrality,* extensively describe the characteristics of different centrality measures in terms of graph metrics. The authors identify three groups of centrality measures:

- Degree-like Measures
  This group of measures focuses on incoming and outgoing edges in a graph. A prominent degree-like measure in social software systems is *inDegree* which indicate the number of pages that link to a given one.
- Closeness-like Measures
  Closeness-like measures calculate the distance of a given node to another node, a group of nodes or every node in the graph. This measure is often used in social network analysis since it describes, for instance, whether a person can be reached easily.
- Betweenness-like Measures
  The group of betweenness-like measures always regards the graph as a whole. *Betweenness Centrality* calculates the shortest paths in a graph and indicates the number of paths in which a particular node is located. This measure is also an important measure in social network analysis since it indicates which people are hubs in a communication network.

The subsequent papers describe approaches wherein different types of centrality calculations are used to assess features of nodes. Leydesdorff (2007), in *Betweenness centrality as an indicator of the interdisciplinarity of scientific journals,* presents an approach which uses centrality scores in a network of publications. Centrality measures have the characteristic to rate nodes that represent connectors between two or more networks. In the case of social networks, this means that if nodes in two different networks communicate, all communication goes through the connecting node. This shows the importance of central nodes. In the case of publications, the different clusters in the network represent different disciplines. The connecting nodes between these clusters are interdisciplinary since they are members of two clusters. The authors of the paper show that centrality measures perform well when assessing the interdisciplinarity of journal papers. In Chidlovskii (2010), *Multi-label Wikipedia classification with textual and link features,* the authors classify a given set of Wikipedia articles according to their graph structure. Therefore, several graph metrics are applied, amongst others, to different types of centrality scores: mean centrality, degree centrality (Hage & Harary, 1995), betweenness centrality (Freeman, 1979), closeness centrality (Sabidussi, 1966), and stress centrality (Shimbel, 1953). The result shows significantly the performance of these features for article classification.

Blaschke and Stein (2005), in *Methods and Measures for the Analysis of Corporate Wikis: A Case Study,* present a study conducted in an organization. Within this study, the authors analyze a corporate Wiki using graph measures which are usually applied in social network analysis. This approach is not only limited to link graphs but also to communication graphs. These are created from analyzing which users contributed to the same article. The metrics used in the paper are *betweenness, closeness, indegree,* and *outdegree.*

Performance in terms of time and space consumption is always a critical factor regarding graph metrics. Considerations regarding these issues have been made by Brandes (2001) in *A faster algorithm for betweenness centrality.* The authors analyze the runtime complexity of centrality algorithms and propose modification for betweenness centrality in terms of runtime improvement. Due to the sensitivity regarding runtime for the approach described in this thesis, this work provides the basis for centrality related calculations.

Reviewing literature, the metrics mentioned above are the most relevant for graph analysis. But graph metrics are not limited to those. Since implementing metrics is not part of this work, I resort to the existing frameworks. Therefore, the succeeding review of graph analysis frameworks should give an overview of existing metrics. Because metrics must be adapted to the system architecture often, modifications to the interfaces are required. This task is easier to accomplish by having the source codes. Hence, this review is limited to open source frameworks.

Madhain et al. (2005), in *Analysis and visualization of network data using JUNG,* presents the **JUNG** framework (Java Universal Network/Graph Framework). Jung provides a large set of graph representation, manipulation and analysis functionalities. As an open source project, it is freely available and is thus permanently improved by the community. Hence, JUNG is well accepted in the community and used in many projects and experiments. Madhain et al. (2005) describes its major features as:

• Support for a variety of representations of entities and their relations, including directed and undirected graphs, multi-modal graphs (graphs which contain more than one type of vertex or edge), graphs with parallel edges (also known as multigraphs), and hypergraphs (which contain hyperedges, each of which may connect any number of vertices).

• Mechanisms for annotating graphs, entities, and relations with metadata. These capabilities facilitate the creation of analytic tools for complex data sets that can examine the relations between entities, as well as the metadata attached to each entity and relation.

• Implementations of a number of algorithms from graph theory, exploratory data analysis, social network analysis, to machine learning. These include routines for clustering, decomposition, optimization, random graph generation, statistical analysis, and calculation of network distances, flows, and ranking measures (centrality, PageRank, HITS, etc.).

• A visualization framework that makes it easy to construct tools for the interactive exploration of network data. Users can choose among the provided layout and rendering algorithms, or use the framework to create their own custom algorithms.

• Filtering mechanisms which extract subsets of a network; this allows users to focus their attention, or their algorithms, on specific portions of a network.

The Network Workbench Tool (NWBTeam, 2009) is an open source project for providing a large-scale network analysis, modeling and visualization toolkit for biomedical, social science and physics research. The framework comprises a large set of integrated metrics for graph analysis and clustering. The use of standardized input formats facilitates usage of the framework. Formats like XGMML, TreeML and GraphML, which are also used in this implementation, can be used as input. Furthermore, the Network Workbench Tool provides a graphical user interface for visualizing and analyzing graph structures.

A tool which is rather focused on graph visualization though also containing analysis capabilities is **Gephi**: *An open source software for exploring and manipulating networks* (Bastian, Heymann, & Jacomy, 2009). The main idea of the open-source project Gephi is to enable non-expert users to conduct exploratory data analysis. Therefore, it provides a graphical user interface for opening many different file formats like GraphML, Pajek and CSV. The maximal number of nodes in a graph is specified as 500,000, what would correspond to the link graph of a large Wiki but would not be enough to visualize Wikipedia or Flickr. Furthermore, Gephi comes with a set of integrated metrics in different categories and in addition to it, the system provides a plug-in infrastructure for adding new components.

Another very interesting open source graph analysis framework is **Cytoscape** (Kohl, Wiese, & Warscheid, 2011). The framework is more often used in a biomedical context but it brings interesting analysis functionalities which are also applicable to this context. The developers describe Cytoscape thus: "*Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization*". The project benefits from a very active community of developers implementing a large set of plug-ins for enhancing functionality. The flexibility of the framework makes the application of the analysis functionality complex, thus the application in this approach would go beyond the scope of this work. However, applying the enhanced analysis capabilities would be an interesting possibility for further research in this area.

3.4.1.3 Usage

As described in the previous sections, quality of resources can be determined based on the content of a resource and how it is related to another resource. This section reviews papers that focus on a third dimension: **usage**.

Here, we have to distinguish between two underlying assumptions. First, the qualitative status of a resource influences the way people interact with the resource. That means, usage patterns differ depending on the resource status. Second, the type of user activity is decisive for the qualitative status. That means, activities representing interactions with a resource can be grouped into more or less quality fostering activities. Considering the activities contained in the history of a resource, the qualitative status can be inferred from past activities.

One challenge in evaluating system usage is the tracking and representation of user interactions. Some approaches are not limited to one system but rather collect as much information about the user as possible. Najjar et al. (2006), in *Attention metadata: Collection and management,* proposes techniques for observation and representation of user interactions. Information about the interaction is stored in a data model which is described in the Context Attention Metadata Schema (Wolpers et al., 2007):

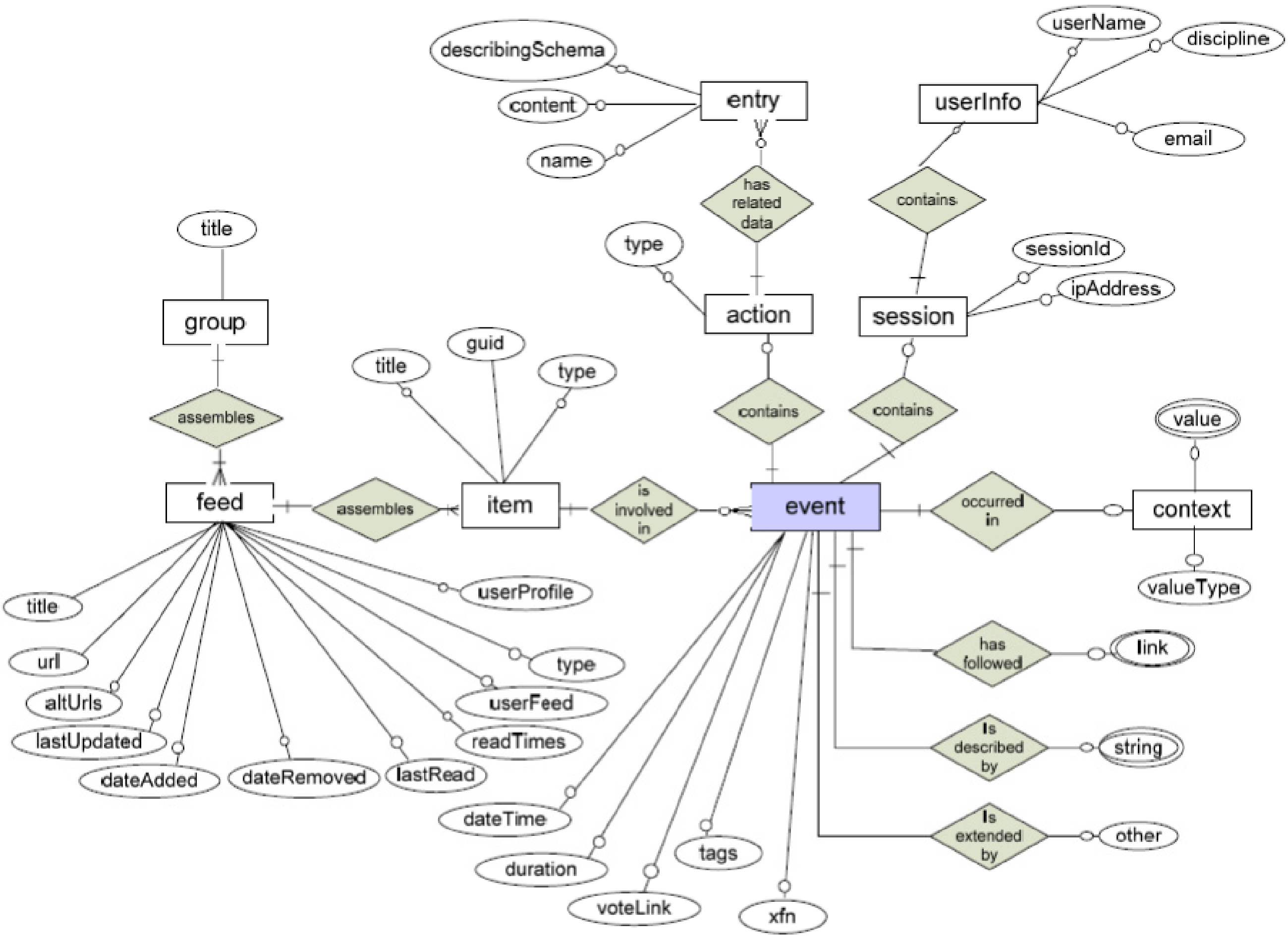


Figure 3-7 CAM metadata schema (Wolpers et al., 2007)

The central concept in this schema is a user event. Events are described by attributes and their relations to other concepts. These concepts are independent from operating system and application what allows for populating the model in various application environments.

A similar approach is proposed by Rath and Lindstaedt (2009) in *UICO: An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop.* The context ontology focuses rather on representing application specific metadata from Microsoft Windows and Office. Different to the approach described in Wolpers et al. (2007), the central concepts are less activity-based with more focus on the application metadata. The model is divided into three logical parts: the user description, the application metadata, and the interactions between user and application.

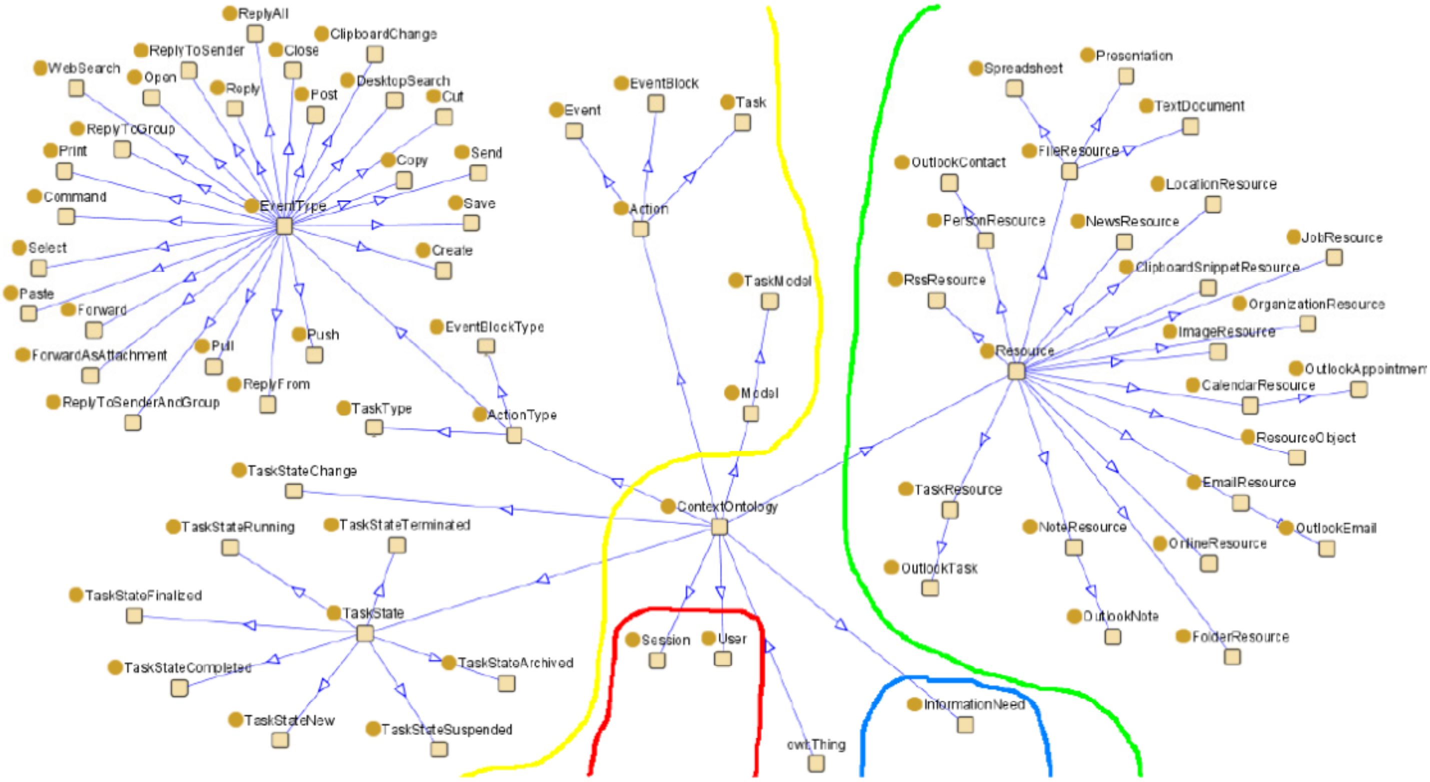


Figure 3-8 UICO model (Rath & Lindstaedt, 2009)

The previous section describes how a qualitative improvement can be measured by observing user tasks. Usually, these tasks consist of very fine-granular user events up to events that consist only of keystrokes and mouse movements. All these approaches are only to a limited extent transferable to this approach because in the context of social systems web technology is mostly used. This only allows for tracking user activity on a higher level of granularity. Generally, tracking user activities on the server side of the web is limited to tracking the URLs that are requested by the user. Based on this information, dwell times can be inferred (but without knowing whether the user is actually looking at the page). In addition, system-specific logs can be used to track usage activity. Usually, all web-based systems log the activities of their users. Thereby, the activities are system-specific as well. For instance, Wiki activities like creating an article, commenting on an article, editing and deleting an article are logged. The next paper shows an approach toward using these logs for determining quality in Wiki systems.

In Liu and Ram (2005), *Who Does What: Collaboration Patterns in the Wikipedia and Their Impact on Data Quality,* the focus is on rather fine granular system-specific user activities from which the qualitative status can be inferred.

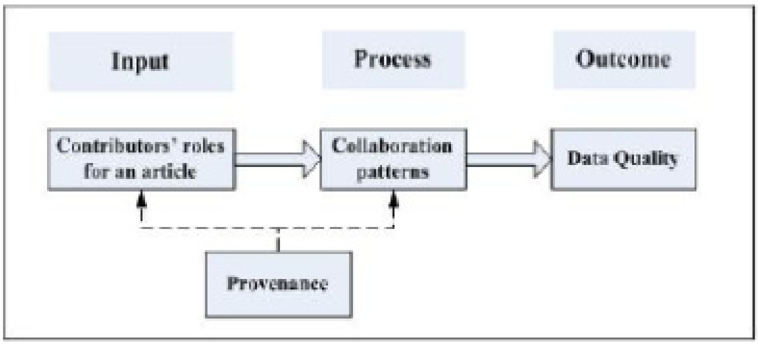


Figure 3-9 Process chain of collaboration patterns (Liu & Ram, 2009)

The underlying assumption of the author is that contributors to a Wiki article can play several roles. In so doing, a role consists of a set of expected and enacted behaviors from a contributor. A collaboration pattern represents the interaction between different roles. The roles of the contributors as well as their interaction patterns can be identified by considering the provenance of a Wiki article. Provenance, in the case of Wiki articles, means the history of activities that led to the current status. In Ram and Liu (2007), *Understanding the Semantics of Data Provenance to Support Active Conceptual Modeling,* the authors present an event-model for a semantic description of resource provenance. The following list of events is based on this model:

Entity	**Action**
Sentence	• creation
	• modification
	• deletion
Link	• creation
	• modification

	• deletion
Reference	• creation
	• modification
	• deletion
Article	• revert

Figure 3-10 User activities in a Wiki (Ram & Liu, 2007)

Based on these activities (see Figure 3-10), different roles have been identified as in the following:

- *All-round editors (Performing actions more frequently than an average contributor)*
- *Watchdogs (Focusing on reverts. Performing actions more frequently than an average contributor)*
- *Starters (Focusing on sentence creations and seldom engaging in other actions. Performing actions less frequently)*
- *Content Justifiers (Focusing on three types of actions: sentence creations, link creations and reference creations. Performing actions less frequently)*
- *Copy Editors (Focusing on sentence modifications)*
- *Cleaners (Focusing on removing sentences, references and links)*

As Figure 3-10 shows, based on article provenance, the roles of authors and their interaction patterns can be identified. The assumption of the authors is that interaction patterns influence article quality or the other way around, namely, that article quality can be assessed through evaluation of interaction patterns. The authors tested this hypothesis in an experiment with Wikipedia articles. They evaluated the correlation of quality assessment based on interaction pattern and the user-given quality categories in Wikipedia.

The previous approach shows how system specific activity logs can be used to find out more about the user. In terms of granularity, the activities are still on an article level. That means that the events describe what is happening with an article but not which part of an article is edited. This information can possibly be inferred from comparing the current version with a past version, but to detect which part of a text the user reads (in other words, which part of the text is visible by scrolling up/down) is not possible. A similar problem is to detect which part of an article a comment refers to since the article is commented upon as a whole. To detect these usage activities, more sophisticated approaches are required. The two subsequent papers present approaches for acquiring fine granular usage data on web based systems. Both approaches use additional tools for tracking user activity.

In Dekel (2007), *A Framework for Studying the Use of Wikis in Knowledge Work Using Client-Side Access Data,* the author describes an approach for tracking fine granular user activities by modifying a MediaWiki. The modification in the Wiki changes the content of the page that is sent to the user. The HTML page includes an AJAX script that is executed on the client side and thus allows for tracking fine granular events like keystrokes, scrolling or mouse movements.

A similar approach is presented in Atterer, Wnuk and Schmidt (2006), *Knowing the User's Every Move – User Activity Tracking for Website Usability Evaluation and Implicit Interaction.*

Here, the authors interpose a proxy server between the web-server and the client. When the client sends a page request, the response from the web-server is sent over the proxy-server and is modified there. Similar to the former approach, a script is injected into the response code. This script, in this case a JavaScript, is executed on the client-side and thus allows for tracking user activity. The difference in terms of the former approach is that this one is not bound to a particular system. The usage tracking works for all pages that are sent through the proxy. The granularity of recoded usage data is similar to the previous approach. Figure 3-11 shows the recorded mouse trails on a website.

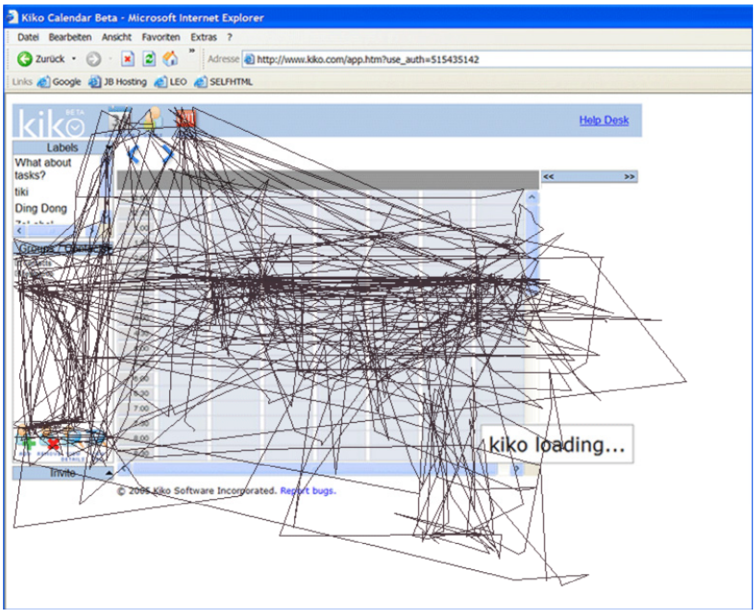


Figure 3-11 Trace of mouse movements on a webpage – mouse trails can be used to evaluate how websites are used and how the user navigates (Atterer et al., 2006)

3.4.2 Relevance for this work

The related work, presented in the former section, constitutes a very important part of this work. Metrics play a crucial role in measuring aspects of a system. The scope of this work does not include designing and implementing new metrics. I would rather reuse existing and evaluated metrics for measuring certain parameters. Papers in the related work section describe various metrics and their evaluation in a particular context. At the end of the previous section, I presented some frameworks which provide a plentitude of metrics. I reused for this work several metrics from different categories. For measuring quality in structures, I mainly used PageRank, different types of centrality measures, and HITS. For the content representation, readability measures like Gunning Fog and Flesch Kincaid Reading Ease Score were utilized. The usage metrics are conceptually reused since the activities which provide the basis for these metrics are highly system-dependent. For evaluating the usage of Wiki systems, I use the system specific events from the activity log. In the context of my work, I consider the approaches for tracking fine granular usage data by using a proxy, as not applicable. Injecting scripts using a proxy server is a big security risk. The user cannot appraise the effects of the injected code on his/her machine. Furthermore, this technique does not consider the user's privacy. All data from the user is sent over a proxy that would be able to track visited sites, private data, passwords, etc. This cannot easily be switched off. For my approach, I decided not to use such methods in order to avoid a lack of user acceptance due to security risks and unsolved privacy issues. However, from a technical point of view, these approaches are interesting and definitely usable in a controlled lab experiment.

4 QUALITY MODELING

One challenge in designing quality aware systems is how to describe and represent quality. Two interacting entities dealing with quality have to be modeled: the resource and the user.

User models usually represent different characteristics for the user. This can include user context (Rath et al., 2008) but also the users' tasks (Lindstaedt et al., 2009) and competencies (Ley, Kump & Albert, 2010). This approach introduces modeling a new dimension – the quality dimension. The quality models contain the representation of user requirements regarding quality.

The quality resource model provides information about the current qualitative status of a resource. Resources in the context of this work are mostly Wiki articles since the Wiki is exemplarily used as an instantiation of a social software system. This section describes how quality is modeled in these entities. For both user quality profile and resource quality profile, I describe the overall quality by means of quality dimensions. The fundamental model which provides the basis for understanding quality in this work is the quality model by Wang and Strong (1996).

Intrinsic Quality	Contextual Quality	Representational Quality	Accessibility Quality	Quality Categories
Believability Accuracy Objectivity Reputation	Value-added Relevancy Timeliness Completeness Appropriate amount of data	Interpretability Ease of understanding Representational consistency Concise representation	Accessibility Access security	Quality Dimensions

Table 4-1 Categories and dimensions of quality proposed by Wang and Strong (1996)

Table 4-1 shows the resulting quality dimensions taken from the quality model. These are dimensions used in the model to describe users and resources as well. Within the models, quality dimensions are attributes of the resource concept. For both user and resource model, the representation is similar in terms of used quality dimensions. The main difference lies in the interpretation; the user quality model represents the user quality preferences for a particular task, whereas the resource model contains the resource quality status.

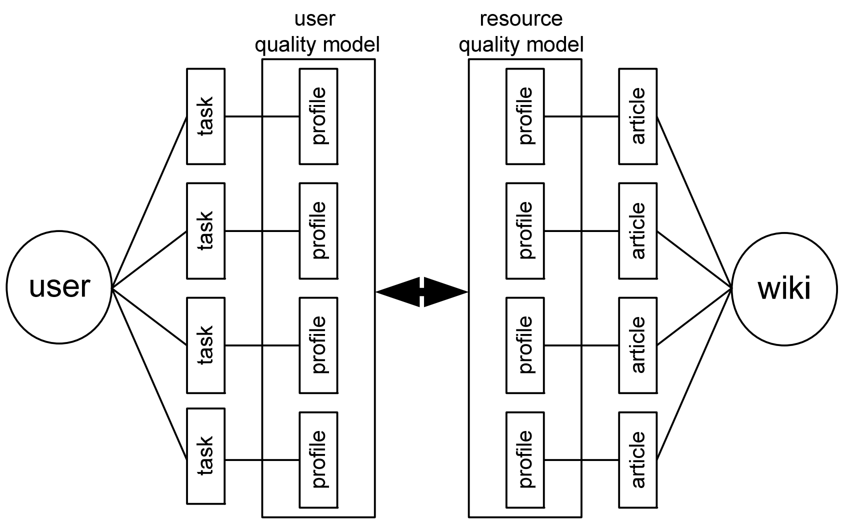
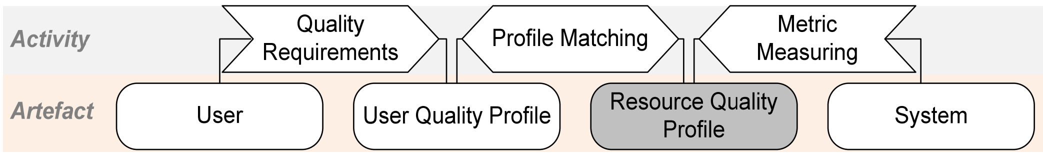


Figure 4-1 Abstraction of user and system in quality profiles

Figure 4-1 shows the user on the one side and the system on the other. Since user requirements may not be similar for different tasks, the requirements of the user are modeled in user profiles that are mapped to a particular task. On the system side the quality status of each article is represented in a resource quality model.

4.1 RESOURCE QUALITY MODEL



The aim of this work is to enable the Wiki to make decisions also regarding the qualitative status of resources. This means the qualitative status must be represented in a machine-readable format. The schematic specification of the representation of resource quality is denoted as **Resource Quality Model**. Each single resource has information about its qualitative status stored in the metadata. This metadata is an instantiation of the resource quality model and is denoted as **Resource Quality Profile**.

The resource quality profile contains information about the current quality status of a resource. In the case of Wiki, content is represented in articles. Therefore, a quality profile is always related to an article. The representation of an article regarding its quality is stored in a semantic model that describes the current quality status of the article. Not only is the current status important for the assessment of an article, but the maturing process can be an important quality indicator as well. That means, the current status is important for assessing the quality at a certain point in time but the progression of quality can be important as well. Hence, the profile contains in addition information about the qualitative change in the development of the resource. The use of both the user and the resource quality model leads to a quality-aware system. The additional information provided by the models can be used to make quality-aware recommendations, to enhance search with qualitative filtering, and to allow for supporting the task of gardening in the Wiki.

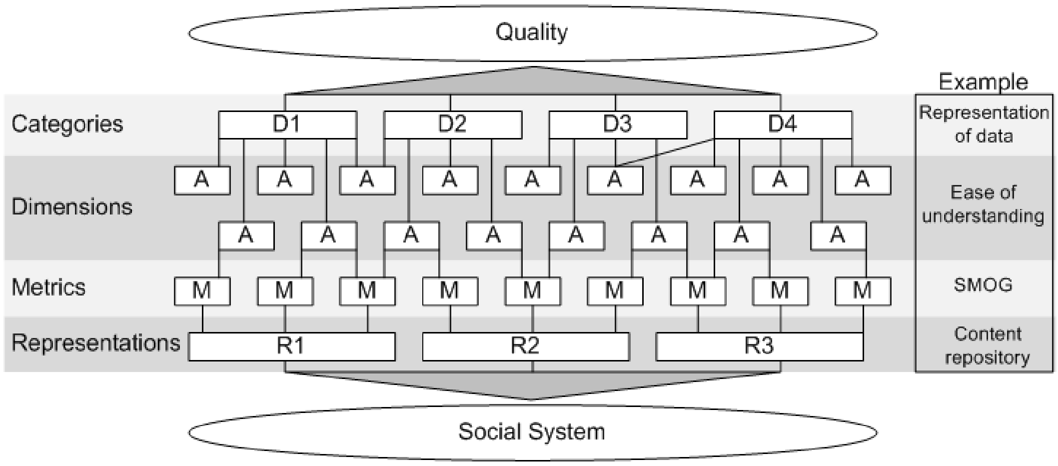
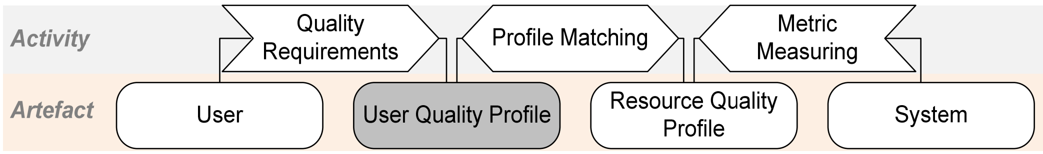


Figure 4-2 Multilayered model for quality assessment

Figure 4-2 shows the generalization process for obtaining the status of a social system. Therefore, we have to pass through different layers of abstraction. The quality model of Wang and Strong (1996) provides the theoretical foundation for this process. Thereby, quality is described at different levels. The category level and the dimensions level are taken from the existing model. In order to get information about the dimensions, we need metrics to measure the system. Metrics can directly be mapped to the dimensions. That means one dimension is measured using one or more metrics. For measuring the system, a formal representation of the system is needed. Depending on the type of metric, different representations are required. The system representation is the lowest layer of abstraction from the system, and so the lowest layer in our process model.

4.2 USER QUALITY MODEL



In order to derive a representation of users' quality requirements, several steps are necessary. One prerequisite is knowledge about the users' requirements regarding quality. These requirements are represented by the multiple facets of the user context. In this work, only task-related parts of the user context attributes are considered. Aspects like education or competence are omitted.

In general, the objective of this work is the elicitation of users' requirements regarding quality. There are three ways to reach this goal. The first is to ask the user directly – the explicit method. The advantage of this method is that one can assume that the user knows best his/her own preferences. The drawback is that sometimes users are not aware of their requirements, and so they may not be able to express them. The second way is a combination of implicit and explicit requirement elicitations. In this case, the user explicitly rates the overall quality of articles. By rating the user implicitly, it gives information about his/her quality preferences. The third way is to evaluate exclusively implicit information. In this case, the system tracks the system usage. From the analysis of these events, articles that the user perceives useful due to their article quality can be determined. These can be used to infer user quality requirements. The next required step towards a quality aware system is the representation of the requirements derived in the previous steps. This representation acts as a model for user requirements and is used as the basis for further processing with the inclusion of the user's needs. The user quality model is populated with implicit or explicit feedback from the user. That means the more information can be collected in the model about the user, the more precise the user quality model will be. Information about the user is stored as a machine-readable semantic model (see Section 7.3.2.2). The user quality model is directly associated with a particular user. That means the model can only belong to one user. On the other hand, one user can have more than one quality profile within the model depending on his/her current context. As stated in Section 2.1, this part of the work is to elaborate the notion that quality requirements are highly dependent on executed tasks. So a user can have several quality profiles, one for each of his/her tasks.

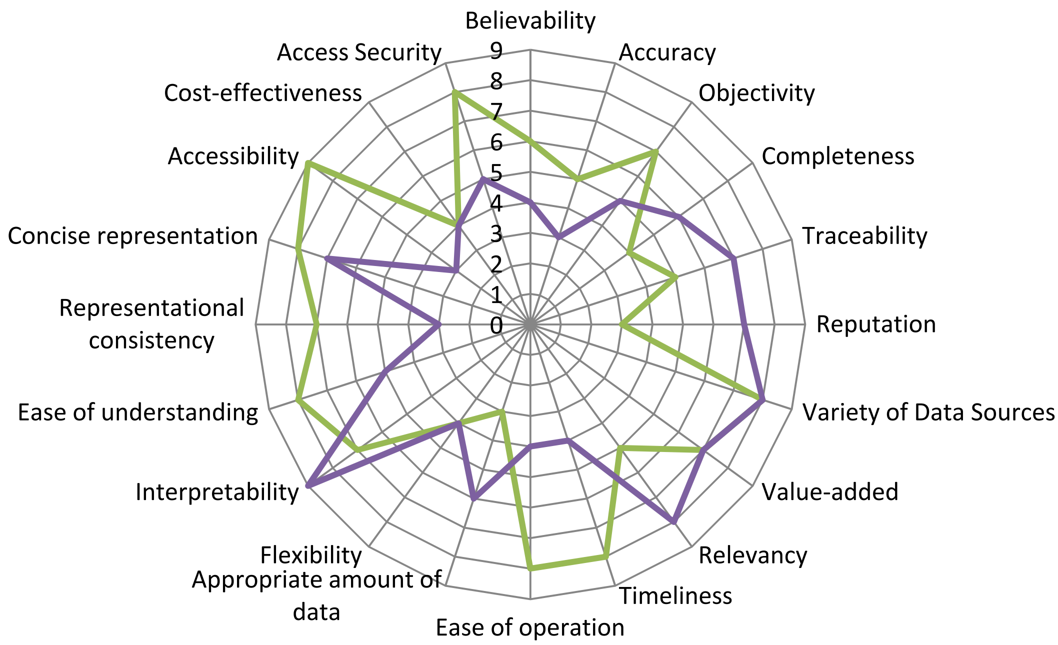
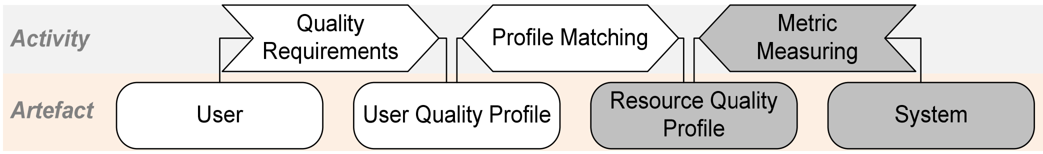


Figure 4-3 Quality preferences in a quality profile

Figure 4-3 shows a quality profile containing quality preferences for two different tasks on a scale from 0 to 9 (0 = not important, 9 = vital). Each line represents the quality preferences for a particular task. High values in one dimension mean that this facet of quality is important for fulfilling such a task.

5 WIKI METRICS



"*Measuring is defined as the process of mapping the attribute-level distributions of real-world entities to numbers or symbols in an objective and systematic way. Accordingly, a measure is defined as a relation associating the attribute-level distributions of real-world entities or processes with numbers or symbols.*" (Stvilia et al., 2007)

Models are used as a formal representation of an entity or situation. Often, only a certain aspect of an entity is represented by a model. In the majority of cases, the model leads to simplification and reduces complexity. A machine readable model enables algorithms to measure features of the model and consequently make a statement about the entity that is represented by the model. This statement can be quantitative or qualitative depending on the interpretation of values.

Since the quality of a resource depends on many factors, including factors that cannot be objectively measured, a qualitative assessment is often subjective. Things like experience or emotions play an important role in quality assessment. Thus, a user may assess the quality of a text to be good but cannot express why. Quality is a very general concept; many aspects play a role in quality perception. A text can have a good quality because it is well structured or because it is enriched with pictures and tables. On the other hand, it can be of high quality because its content describes an entity very well. Additionally, the user itself is a relevant factor for quality perception; especially the context in which the text is used. A particular text can be more or less useful in different situations. Last but not least, the user context and experience are crucial factors for quality assessment. A professor assesses the quality of a document differently from a student.

The subjectivity of qualitative statements on resources makes it difficult to assess quality automatically. In contrast to humans, machines cannot make qualitative statements. In order to make a quantitative statement, machines use metrics which provide values acting as a quantitative measure. In order to be able to make qualitative statements anyway, one option is to find measures that provide values which correlate with the human qualitative assessment. In this way, one can make qualitative statements about a resource using a quantitative metric.

For example, for specific texts, readability formulas correlate with the quality perception among users (Farahat et al., 2006), meaning that if the readability formula results in a high value for a particular text, the user assesses the text as high qualitative as well. Needless to say, such a readability formula can only state the quality regarding understandability and not structuredness or syntactic correctness. Quality is complex and can only be measured by using a set of complementing metrics. The next section describes how these metrics can be categorized.

5.1 System Dimensions as Basis for Metrics

Metrics serve as a quantitative measure for resources. Based on formal machine readable representations of the resource, they provide a value that can be interpreted to make a quantitative statement. Metrics can be categorized on the type and origin of data they use as the basis for calculation. This section proposes a categorization schema of the metrics used in this work. The categorization schema is based on different dimensions in the representation of a social system.

As described in Section 3.4, metrics need a particular basis for calculating measures. The feature that should be measured must be represented in this data basis. For instance, metrics measuring usage-related indicators need a data basis representing the users' activities, e.g. a system event log. Similarly, structure-based metrics need a special representation containing structural information of the system represented in a graphic structure.

5.1.1 System Representations based on the Knowledge Ecosystem

Since there are various system representations and the number of possible representations is not limited, one question here concerns which representations that would be required for assessing quality. Lindstaedt et al. (2008) describes the knowledge ecosystem as a model showing interacting entities in systems supporting the knowledge of workers (cf. MATURE Consortium, 2009b). Since social software systems can be seen as one special type of such a system, the model of interacting entities is also valid for my approach. As shown in Figure 4-1, the three interacting entities are: **people**, **content** and s**tructure** (cf. MATURE Consortium, 2009b).

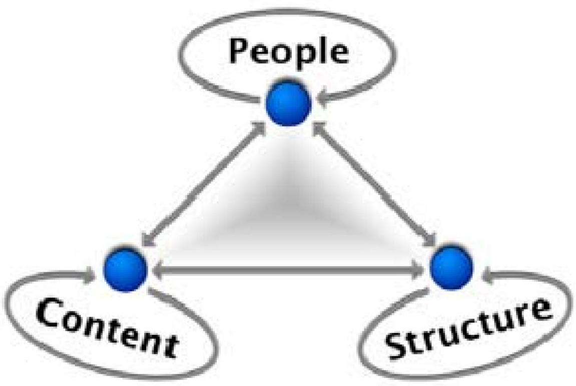


Figure 5-1 Interacting entities in a knowledge ecosystem

Content is the essential part of almost all social software systems. Content can be represented by texts, images, tables, and in the meantime also as video and audio streams. More and more, content is shifted from traditional documents to web-based platforms. The Wiki, as a common knowledge base and one example for such a social software system, gets more and more popular; e.g. in 2006 IBM used internally 13,313 Wikis (Arazy, Croitoru, & Jang, 2009).

Semantic and **Temporal Structures**. This type of knowledge entity is probably the least visible within organizations. Semantics connect the different entities and support the individual learning processes by providing the basis for mutual understanding. Without semantic integration, grassroots approaches encouraging people to contribute their individual views, experiences and insights would get stuck in misinterpretations and lengthy negotiation processes. Semantic structures can be represented by tag clouds and emerging folksonomies, folder structures, competence models, local or global enterprise ontologies, social networks, user representations, etc.

People use social software systems as source of information but also for contributing information in a community. They interact with multi-media content articles, structures and other users. This interaction is reflected in the system as **usage** data.

All these entities represent an artifact of knowledge, an externalization of the authors' knowledge. Content-based metrics facilitate measuring certain aspects of content objects and so they allow for an evaluation of underlying knowledge. In addition to the representation of content, social systems provide structure which mostly means that relational information between parts of the content is expressed in digraphs. Structure-based metrics measure the status of this graph. The interpretation of the values given by the metrics depends on the semantics of the graph. For instance, a centrality measure can be calculated for a graph representing the link structure between articles or a graph representing relations between article authors. The interpretation of almost every metric depends on the semantics of the underlying model. The third knowledge entity of *people* is, due to its complexity, hard to represent in a model. The knowledge of people is expressed in content as well as in structured models. Additionally, people can be characterized by the way they interact with the system. The representation of the interaction of people with the social system is denoted as the usage model which provides the basis for usage-based metrics.

The three dimensions used in my approach are derived from the three knowledge entities of the knowledge ecosystem: people, content, and structure. The relevance of the categories depends on the application case and the interpretation of the resulting values. This approach is not limited to these three categories, but they are considered to be relevant and provide a sufficient basis for metric-based quality assessment. For each of the metric dimensions, various metrics can be applied.

5.1.2 System Representations Based on Used Metrics

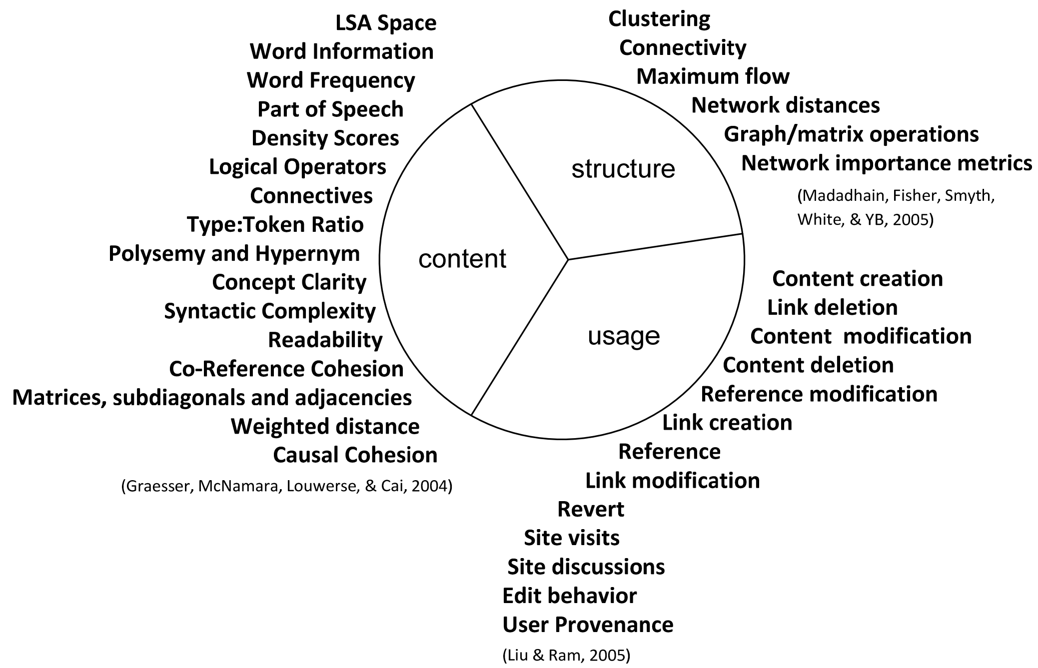


Figure 5-2 Metric categories based on system representations

Figure 5-2 shows categories to which metrics already used to measure information quality can be mapped. The metrics in this figure use three different representations – content, structure, and usage. The previous section describes a top-down approach by selecting system representations reflecting the entities interacting in the social software system. A rather pragmatic bottom-up approach is to select the metrics first and then provide system representations that are required for the used metrics. Therefore, the set of metrics to be used must be fixed before using the system. This hampers adding metrics at runtime since the metric may not fit the available system representations.

5.2 GRAPH-BASED METRICS

A very common representation is the structure based on representation in a directed graph. Many approaches use graphs that represent relations between objects. One example is the social network analysis (SNA) where a social system is represented by nodes (people in the system) and edges (relations/activities/events) to make a statement about the whole system or individuals in the system (Dom et al., 2003; Zhang, Ackerman, & Adamic, 2007; Brandes & Wagner, 2003). Graph representations are used not just for the analysis of social networks; Jeon et al. (2006) applied metrics to a graph structure to assess the quality of answers in answering services where people ask and other people give answers. Hotho et al. (2006) present an approach for analysis of folksonomies based on their graph representation. Madhain (2005) proposes the following classification of graph-based metrics:

- **Clustering**

finding groups of similar entities regarding a particular feature (e.g. edge betweenness, weak components)

- **Connectivity**

measures robustness of a network based on the graph structure (e.g. K-neighborhood extraction)

- **Maximum flow**

analyzes graphs in terms of the capacity of their edges, indentifies weak edges in a graph structure

- **Network distances**

evaluation of graphs based on paths through the graphs along the edges (e.g. Betweenness Centrality, Eigenvector Centrality)

- **Graph/matrix operations**

(squaring, mean first passage)

- **Network importance metrics**

calculations considering the importance of nodes from incoming and outgoing edges (Page Rank, social network analysis centrality)

5.3 Content based metrics

Parts of this section have been published as:

Moskaliuk, J., Weber, N., Stern, H., Kimmerle, J., Cress, U., & Lindstaedt, S. N. (2011). Evaluation of social media collaboration using task-detection methods. In C. Delgado Kloos, D. Gillet, R. M. Crespo García, F. Wild, & M. Wolpersm(Eds.), *Towards ubiquitous learning: Proceedings of the 6th European Conference on Technology Enhanced Learning 2011* (pp. 248-259). Berlin, Heidelberg, Dordrecht, London, New York: Springer.

Another representation of information stored in the Wiki is the content-based representation. This representation consists of only the textual (and multimedia) content of the system. Content-based metrics assess the status of texts, video, audio and pictures. Since Wikis still almost solely consist of text, for this work only text-based content metrics are considered. One very common approach to assess the quality of text is by means of reading scores. Examples of reading score-based approaches can be found in Gunning Fog Reading Ease Score, Flesh-Kincaid Readability Formula (Agichtein et al., 2008) and the SMOG Reading Score (McLaughlin, 1969). But metrics for content-based quality assessment are not only limited to reading scores. Graesser et al. (2004) propose, for instance, text coherence as one indicator relevant for text quality. In addition, they present a framework consisting of more than 200 metrics for text assessment. Graesser et al. (2004) propose the following schema for classification of content-based metrics.

- LSA space: ‘latent semantic analysis’ as statistical representation of word and text meaning
- Word information: particular characteristics of the words in the text
- Word frequency: frequency with which particular words occur in a certain language compared to the frequency of words in the text
- Parts of Speech: identification of particular word types: nouns, lexical verbs, adjectives, adverbs, prepositions, determiners, and pronouns
- Density Scores: incidence, ratio, or proportion of particular word classes or constituents in the text

- Logical Operators: frequency of logical operators like *or, and, not, and if–then*
- Connectives: finding keywords indicating connectives like: (1) clarifying connectives such as *in other words* and *that is*; (2) additive connectives such as *also* and *moreover*; (3) temporal connectives such as *after*, *before*, and *when*; and (4) causal connectives such as *because, so,* and *consequently*
- Type: Token Ratio: each unique word in a text is a word type. Each instance of a particular word is a token.
- Polysemy and Hypernym: Polysemy is measured as the number of senses of a word (ambiguity). The hypernym count is defined as the number of levels in a conceptual taxonomic hierarchy that is above a word.
- Syntactic Complexity: syntactic complexity involves a number of metrics that assess how difficult it is to analyze the syntactic composition of sentences.
- Readability: assessing text difficulty by means of readability formulas.

5.4 USAGE-BASED METRICS

In addition to the already presented metrics, resources can be assessed based on the way they are used in the system. The usage of resources denotes any interaction between users and resources in the system. The assumption behind the application of usage metrics is that, if the quality of a resource changes the interaction patterns change as well. This means users interact differently with an article if it is of high quality than with one of low quality. Lih (2004) shows that there is a direct correlation between the quality of an article and the number of edits in a particular time span, respectively the number of unique authors. Moskaliuk et al. (2011) show that there are interaction patterns that lead to the qualitative improvement of an article and some that do not influence the quality. So it is both; interactions can influence the quality of an article and on the other hand interactions can be used as indicator of article quality. In most systems, some attributes that would be relevant for usage-based article assessment are hard to detect, for instance, the expertise of an editing user. Liu and Ram (2005) identified the following interaction patterns for categorizing usage-based indicators. The indicators concern how often the activity occurs:

- Content: modification; changes in the content of a Wiki article
- Reference: adding references to a Wiki article
- Content deletion: deleting whole articles or parts of it
- Reference modification: modifying a reference within an article
- Link creation: adding a link to a Wiki article
- Link modification: modifying an existing link within an article
- Revert: reverting edits in the history of a Wiki article
- Site visits: a user visits a Wiki article
- Site discussions: a discussion is added to an article
- Edit behavior: specific pattern of user activities
- User Provenance: provenance of the users activities in a Wiki article
- On-Site actions: overall activities on one Wiki page

5.5 MAPPING METRICS TO DIMENSIONS

Figure 5-3 shows the different layers required for assessing quality in a social system. Previous sections describe the creation of system representations and the application of metrics to these representations. Furthermore, the mapping of quality categories to quality dimensions has been elaborated. What is now missing is the mapping between metrics and dimensions. The problem here is to find a metric that reveals the status of the current resource regarding a certain dimension. In other words, we need metrics that correlate with the users' perception of the qualitative status in a particular quality dimension. For example, in considering the quality dimension *Ease of understanding*, how can we find a metric that corresponds to the user perception of this dimension? Basically, one option is to empirically analyze if a given metric correlates to the user perception. In the case of *Ease of understanding,* one would test if SMOG (McLaughlin, 1969) is a metric feasible to predict the perception of the user.

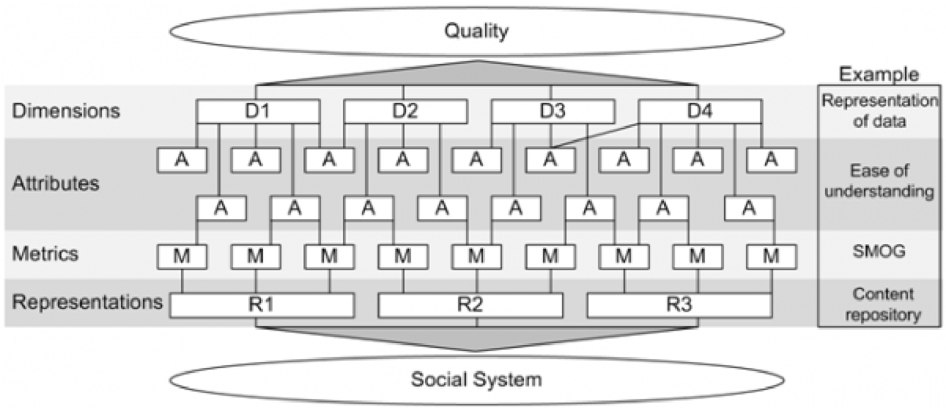
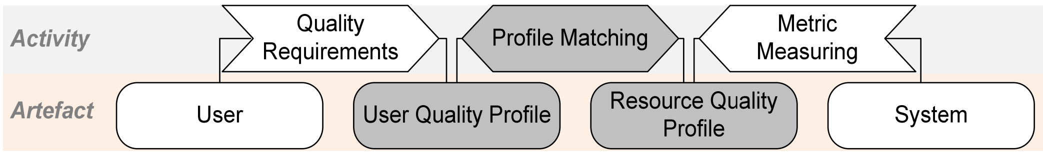


Figure 5-3 Multi layered quality assessment model

The approach that I have chosen for this work is to reuse existing empirical studies so as to find evidence for mappings between metrics and quality dimensions. Since this approach is not limited to a given set of metrics, one task in adopting this approach would be to find suitable metrics that fit the context of the system. The succeeding table shows the metric mapping in two application cases and references in which the metrics are applied and evaluated.

6 QUALITY ASSESSMENT BASED ON PATTERN SIMILARITY



Quality evaluation and quality-based user support are always based on quality perceptions of the users. Hence, the process of quality assessment is always a matching process between quality profile and resource footprint (resource profile). The quality profile represents the quality requirements of a user for a particular task. The quality footprint describes the current qualitative status of a resource. Both of them can be either represented in combinatory patterns (strings, value series, system logs) or spatial patterns (discrete structures) (Hagedoorn et al., 2000). Since the values in the user and resource profile are available as a vector of values and not as discrete functions, graph-based similarity measures as well as combinatory measures can be used. As mentioned in Section 2.1.2, several types of quality-aware behavior are possible. These functionalities should be examined here when considering the different types of pattern similarity.

1. **Resource profile – Resource profile comparison**
   For this method, the resource quality profiles are compared with each other. The result can be used for several application areas. It can be used to find resources of similar quality and to group them according to their quality into a cluster. These article clusters can only be used if the quality, not the content, of an article is relevant. In the case of a Wiki gardening, one use case could happen by querying all articles that should be improved in some aspect of quality. Another use case could be the pre-filtering of search results. Only a subset of articles is used as a basis for search. This guarantees that search results do not contain only the information that is

requested, but also conforms to the quality requirements of the user. This method can also be used to recommend to the user a resource that has the same qualitative status as the one the user is currently working on.

2. **User profile – Resource profile comparison**

   This method compares the requirements of a user with the status of a resource. That means the algorithm compares the similarity of user requirements with a set of resource profiles. The objective of this comparison is to find resources that correspond to the users' requirements. The results can be used for proactive search mechanisms as well as for ranking search results.

3. **User profile – User profile comparison**

   This type of profile comparison provides people with similar quality interests. Results can be used to find hubs for additional resources that fit in with the user's needs. Identifying a person with similar quality requirements implies that resources created and used by this person correspond to his/her quality requirements which could also be useful for users with similar requirements.

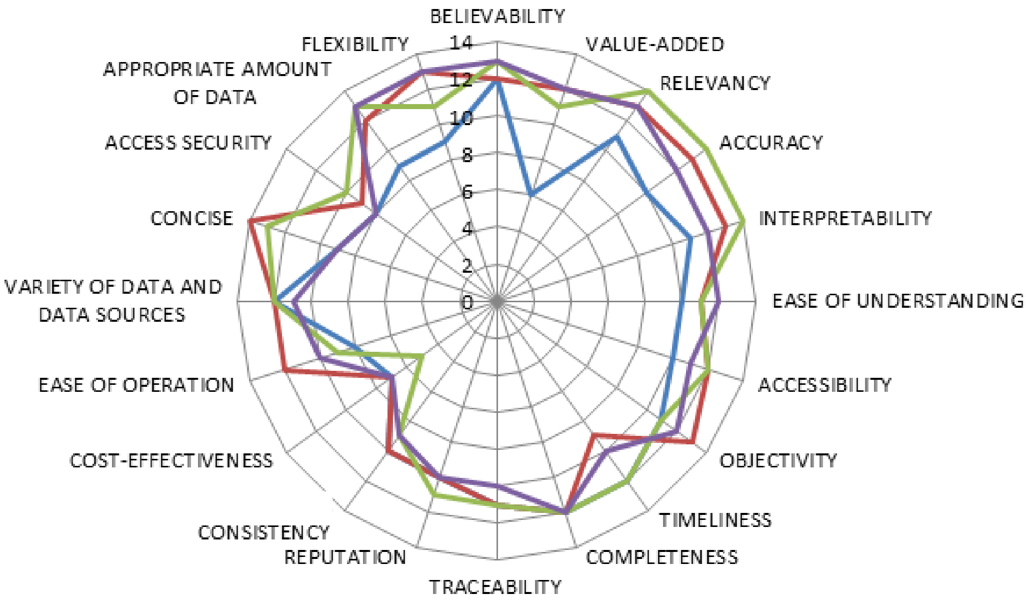


Figure 6-1 Graphic representation of user profile

In general, for analyzing pattern similarity, two graphs are compared to each other. The result is a quantitative indicator for similarity.

Figure 6-1 shows the graphic representation of an excerpt from a user profile for one task with three different persons. The figure shows that requirements regarding quality dimensions are different. The goal of this approach is to provide for each of the profiles the best fitting resources. To achieve this goal, I use pattern analysis to find similar patterns in quality profiles. Identification of similar patterns is a common problem in research. Examples from other research areas where pattern analysis is employed are the analysis of DNA samples or comparing two fingerprints. These examples conjure up the problem in pattern analysis: if a person leaves two fingerprints (from one finger), the fingerprints are not identical. Due to the noise in the scanning process, different surfaces or just dust on the finger, the resulting patterns will very likely not be identical. So the task of identifying fingerprints that are identical from a given set is (relatively) easy but to find fingerprints that similar or where they can be assumed to have come from the same finger is not so simple. The same goal should be achieved in this work. Resource profiles that are identical to the user requirements are probably rare; this approach aims therefore at finding resources whose quality profile is similar enough to be perceived as matching the user's needs. In the subsequent section, I will present some frequently used pattern analysis algorithms and discuss their applicability to the problem statement.

The intuitive approach toward evaluating the similarity of a value vector is to calculate the absolute value of the distance as follows.

$$G_{sim} = \sum_{i=1}^{n} |V_{metric}[i]|$$

Let n be the number of metric values/ quality dimensions and V is their value. G denotes the similarity of two value vectors based on their absolute distance. A problem with this algorithm is that a positive deviation followed by a negative deviation neutralizes the result.

Another frequently used similarity measure is *Cosine Similarity* (Qian, Sural, Gu, & Pramanik, 2004). On this, every metric value is represented as a point in a multidimensional space. Each of the points is part of a vector beginning at the origin. The similarity of vectors is defined as the angle between the vectors. The smaller the angle, the more similar are the values. One benefit of this algorithm is that one value cannot neutralize another value because every value is represented in its own dimension.

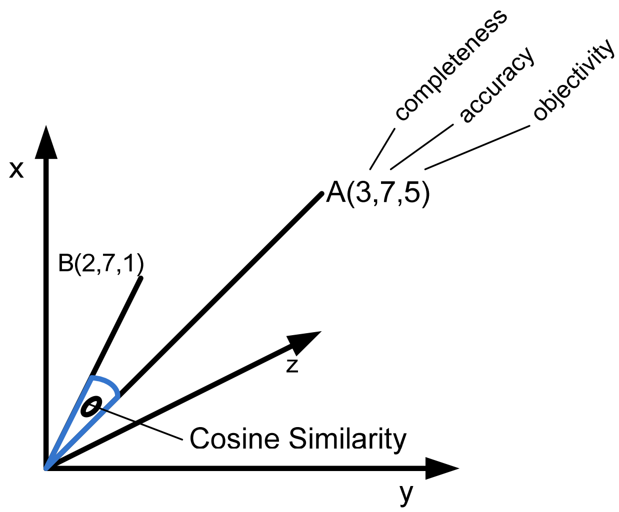


Figure 6-2 Cosine similarity used on quality profile data

$$\textit{Cosine Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

A and B are vectors consisting of metered values. In this context, the vector consists of all values from the metrics. As shown in Figure 6-2, each value stands for a particular quality attribute/dimension. Figure 6-2 exemplarily shows a vector with three values, thus, the points are aligned in a three-dimensional space. Using *n* metrics, the points would be arranged in an n-dimensional space. This similarity measure is often used for analyzing similarity of text documents (Klieber et al., 2009). Here, every value in the vector corresponds to the number of occurrence of a word in the text.

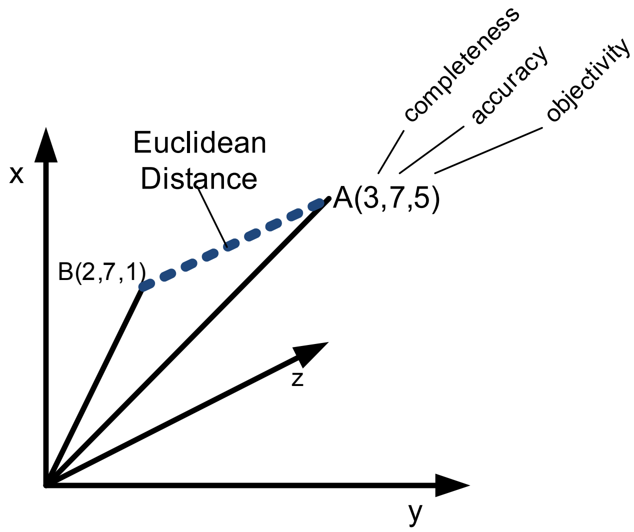


Figure 6-3 Euclidean distance used on quality profile data

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}.$$

The *Euclidean Distance* (Suebsing & Hiransakolwong, 2009) is based on *Cosine Similarity.* Similar to *Cosine Similarity,* the points are arranged as points in an n-dimensional space. In contrast to *Cosine Similarity,* the angle is not calculated but the absolute distance of the points. The closer the vectors, the more similar are the objects represented by the point.

6.1.1 *Problems - Pattern Similarity*

When analyzing the pattern similarity of quality pattern using the algorithms presented above, three effects lead to inadequate results. These effects are known as shifting, scaling, and subset identification. The next section explains these effects and provides arguments as to why the algorithm used in this approach must avoid these effects.

6.1.1.1 Shifting

The problem known as shifting occurs if two patterns are similar but at different levels. This means their values are similar if one adds an offset to each value.

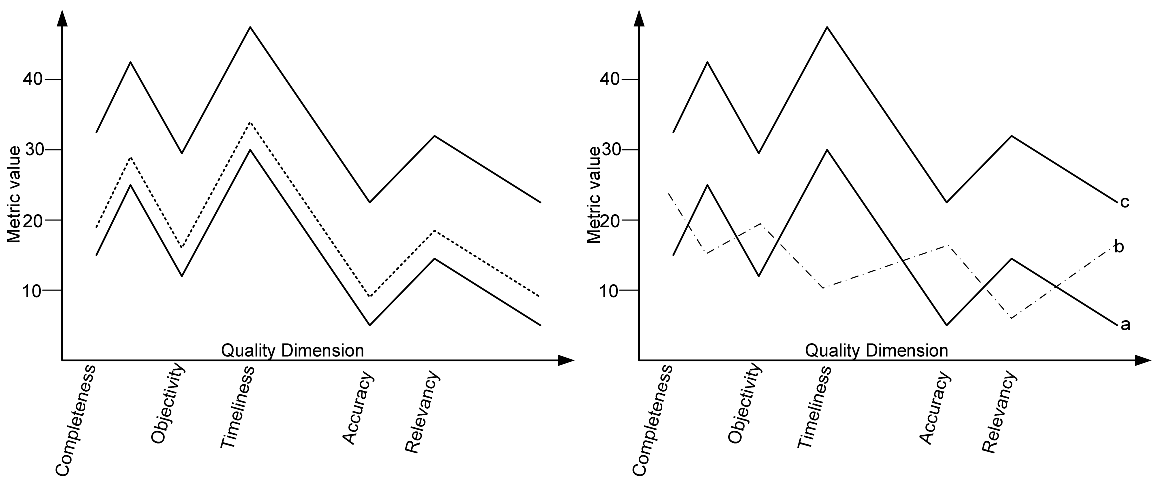


Figure 6-4 Shifting problem

Figure 6-4 shows the quality profiles of three resources represented by graphs. The graphical representation is identical but at different levels. When considering quality pattern, shifting can occur if users explicitly declare their quality preferences and interpret the scale differently. A user could, for instance, rate quality dimensions that are relevant for him/her as always *important* (4 of 5), while another may rate them as always vital (5 of 5) even though they mean the same. Since in this case the focus is on detection of similar patterns and not on absolute values, an algorithm should be tolerant against shifting effects. Because the algorithms described above are based on distances (except cosine similarity), they would result in a greater similarity for a and b than for a and c.

6.1.1.2 Scaling

This is another effect that occurs if the patterns are similar but the values are differently scaled. That means the trend is the same but the values are multiples of the others.

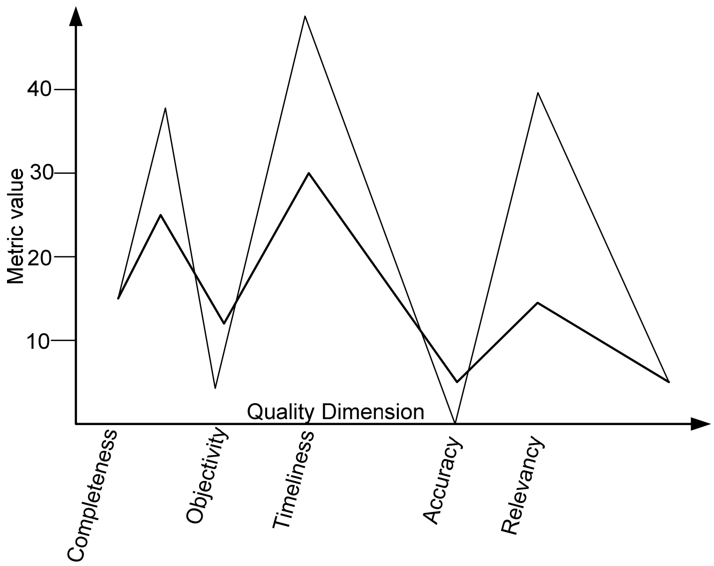


Figure 6-5 Scaling problem

Figure 6-5 shows two patterns that have similar trends for each metric but with different values. The peaks are in the same place but at different levels. In this approach, this can occur due to different rating behavior among users or different scaling behavior by the used metrics (even if I use normalized values). Some users tend to give extreme values in rating, while some others always rate around average. In case of metrics, since the metrics are based on different inputs, some are more sensitive and result in higher values than others. Considering the access count of an article as quality indicator, the absolute number of accesses is not as relevant as the trend. Therefore, an algorithm that finds similarities in quality patterns must be tolerant against shifting.

6.1.1.3 Subset identification

So far, I have assumed that either the whole pattern is similar to another pattern or it is not. Another rather likely case is that only a subset of attributes is similar to the same subset of another pattern. For the algorithms described above, it would mean that the result is falsified due to the fact that the matching part is not involved in the assessment.

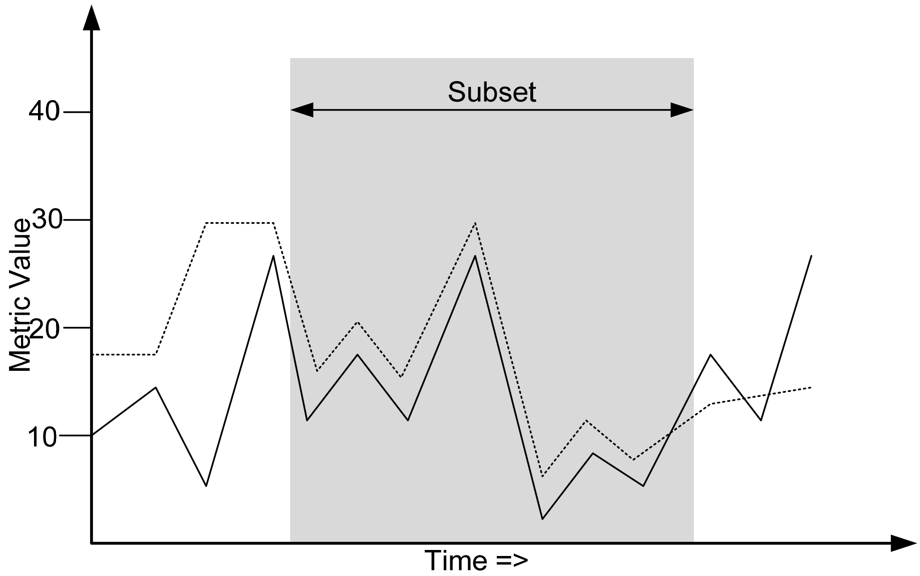


Figure 6-6 Subset clustering example

Figure 6-6 shows two quality patterns with a subset of values that are almost identical. The remaining values have no similarity. An algorithm that is not able to identify subsets might provide an inadequate result.

6.2 APPROACH – MODIFIED PCLUSTER

The algorithms mentioned above fail more or less in terms of some of the problem statements because all of them calculate directly or indirectly the distance between graphs. As shown in Figure 6-4, Figure 6-5 and Figure 6-6, the absolute distance is not always relevant to the similarity of quality pattern. One algorithm whose intention is to solve these problems is the pCluster algorithm presented in Wang et al. (2002). This approach calculates the similarity of values relative to other values. Thus, the patterns are related to each other and not to their absolute value. Shifting effects are avoided since the absolute position of a vector in the n-dimensional space is not relevant. The following formula calculates the pScore for a 2-dimesional quality profile:

$$pScore\left(\begin{bmatrix} d_{xa} & d_{xb} \\ d_{ya} & d_{yb} \end{bmatrix}\right) = \left|(d_{xa} - d_{xb}) - \left(d_{ya} - d_{yb}\right)\right|$$

Considering the two quality profiles *X* and *Y,* each of which having the attributes a and b, then let d_{xa} be the value of the quality profile *X* for the attribute a. The algorithm evaluates the relative difference between two values. In order to avoid also scaling effects, a threshold for the relative difference is introduced as the *pScore-threshold* δ. A δ-pCluster is a 2 x 2 submatrix X if *pScore(X)* < δ. The user defines the threshold on which the pScore is valid. Thus, the absolute value of the attribute is not relevant as long as it is above the threshold. The formula above shows the algorithm for two quality patterns, each having two attributes. Wang et al. (2002) show that every N x M pattern (N pattern, M attributes) can be fragmented in a 2 x 2 submatrix, and so the algorithm is applicable for an arbitrarily large set of patterns. However, the limiting factor in the size of the pattern set is the runtime. The author specifies the runtime as $O(M^2 N \log N + N^2 M \log M)$. The quadratic runtime estimation can lead to unacceptable long runtimes for huge datasets. The performance analysis shows that for 3000 objects (what corresponds to the average number of articles in a corporate Wiki) and 30 attributes per objects (we propose 20), the runtime is between 50 and 400 seconds depending on the chosen threshold δ. Since the complete cluster analysis does not take place at runtime but rather in fixed intervals, the algorithm provides a reasonable runtime in this context. However, due to its runtime behavior, this algorithm would not be applicable for a Wiki consisting of millions of articles.

6.2.1 *Clustering Algorithm Modifications*

The pCluster algorithm provides the information required to identify patterns that are *near* to a given pattern. In our context, *near* means that the patterns are similar to a subset of dimensions. In order to adapt the algorithm to the application context, some modifications have been made. The modifications are aimed on the one hand at increasing the performance of the algorithm, while on the other a special functionality is required to fulfill the intended task. Clustering algorithms like pCluster usually provide the function to form groups of similar objects. The objects are of the same kind, or at least can be treated similarly, and the similarity is calculated on several features. In this approach, the application scenario of the clustering algorithm is different. First, I do not use objects of the same kind. Depending on the similarity measure, either user profiles or resource profiles are clustered. In this way, clusters contain not necessarily only one type of object. The first modification enhances the clustering algorithm to deal with different types of resources. This modification is needed to improve the performance of the algorithm.

The second modification adapts the algorithm to the application case. Given the primal intention of finding users and resources that match a particular quality profile, two options seem to be possible: clustering and classification. I have chosen the classification approach due to better performance and not requiring supervision. But the clustering approach carries further problems. Since the approach intends to find entities that are in the same cluster as qualitative similar entities, the question is what happens if the resource is not in a cluster? The second modification describes how the algorithm is adapted to find entities at different levels of similarity.

6.2.1.1 Clustering different entities

Usually, clustering algorithms are designed to cluster entities of the same kind into groups. One prerequisite for clustering is that the relevant features exist for all entities. In the same way, the pCluster algorithm does not differentiate between various entities. Given the representation of the two entities (user profile and resource profile), both are represented identically. So, the algorithm can be applied to both. Since the algorithm is not optimal when it comes to its runtime-behavior, I have tried to keep the input set as small as possible. Depending on the aimed result, either the user profiles or the resource profiles should be the output of the algorithm. For identifying resources that match a given user profile or finding resources with similar quality profiles, only resource profiles should be contained in the result set. For finding users with similar quality needs, only user profiles should be in the result set. So the idea of this modification is to make the algorithm aware of the different entity types so as to avoid processing entities that can be excluded from the result set. The modification of the pCluster algorithm consists of introducing different entity types in the processing of the features. The entity types that should be contained in the result set can be determined via an additional parameter. That way, other entity types are ignored during the processing steps. This plays an important role if the amount of users and number of resources are evenly distributed; for then the processing of many objects can be avoided. With regard to a future application of the algorithm for clustering resources, the modification is not limited to two-entity types. A conceivable scenario would be to incorporate different types of multimedia objects. The corresponding application could thus be to find video and audio material that meets a particular quality standard.

6.2.1.2 Clustering with different levels of similarity

Measuring pattern similarity is always a tradeoff between accuracy and the number of similar patterns found. The endeavor in this approach is to find patterns that are as similar as possible. The ideal case would be to find identical patterns. As already mentioned, particularly in the context of quality pattern identification, finding identical patterns is not very likely. Not to be able to provide results in terms of similar resources is unsatisfactory for the end user. So the tradeoff is to identify an adequate number of resources with a sufficient level of similarity. This problem is addressed through the modification of the pCluster algorithm described in this section. The parameters are an influencing factor for the resulting clusters, hence processing data with a given set of parameters results in an unpredictable set of clusters. To decide beforehand whether the algorithm with the given parameters provides adequate results is not possible. The idea of this approach is to iteratively adjust the parameters at runtime to overcome the tradeoff between finding enough entities and providing adequate precision in terms of the results. The two parameters used for adjusting the algorithm are the **size of subset** and the **variance of feature values** within one cluster.

The size of the subset determines how many quality dimensions must be similar for all objects within one cluster. Resource profiles as well as user profiles match in a certain number of quality dimensions. The parameter **size of subset** acts as a condition over how many different quality dimension objects in one cluster must match. The lower the value of the **size of subset** parameter, the more objects are found in one cluster. On the other hand, the higher the number of matching dimensions in the subset, the more similar are the objects in the resulting cluster. So the tradeoff for this parameter is to adjust the **size of subset** parameter to a level where enough resources are found and where the resources are still similar enough to be useful for the user.

The **variance of feature values** for documents within one cluster defines how the maximum deviance from one feature to a given feature may be recognized as a similar feature. This means that if variance of features is set to 1% and a subset of ten similar features is identified, then all these features have a maximum deviance of one percent to the features of the other objects in the same cluster. The **variance** parameter is used to determine the similarity of objects in one cluster. The lower the variance, the more similar the objects are. A variance of zero percent would mean that the objects are identical (in the subset). On the other hand, the lower the variance, the less objects are found.

So, as well as the **size of subset**, the **variance of feature values** is a tradeoff between similarity and the number of objects found within a cluster. The selection of the values for the two parameters depends on several factors. One of these is the constitution of the data set. Large datasets allow for finding more objects with a higher level of similarity. Another factor is the distribution of feature values. If the values are distributed around one value with a low variance, more objects may be found than if the values are equally distributed in the whole spectrum.

The approach toward a modified pCluster algorithm is to iteratively apply such an algorithm with changing parameters. This leads to a shorter runtime of the algorithm because the processing stops when a sufficient amount of resources with an adequate quality is found. Therefore, the algorithm processes the input data with different parameters. The range of these parameters depends on the characteristic of the dataset.

Experiment: Determining expedient parameter values in the MATURE Wiki

To get an idea of possible ranges of input parameters in the context of quality pattern analysis, I tested the behavior of the algorithm on a typical dataset. The data is derived from a Wiki system that is used as a project documentation tool within MATURE[2]. The Wiki contains about 800 pages maintained by 50 editors. For each article, 18 features have been calculated. By applying the pCluster algorithm with changing parameters, it should be possible to determine the range for expedient parameter values. The independent variables in this experiment are the **size of the subset** and **variation of features**, while the dependant variable is the number of clusters found. One assumption is that if more objects are contained in the clusters, it is more likely that similar objects be found.

[2] http://www.mature-ip.eu

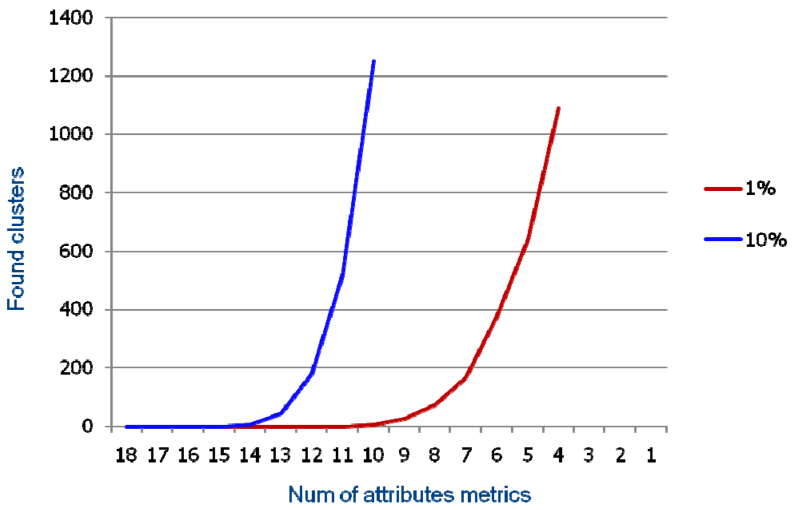


Figure 6-7 Evaluation of parameter values

Figure 6-7 shows the results for different input parameters. The x-axis shows the values for the *size of subset* parameter. For each of these values, the algorithm has been processed with a *variance of features* of 1% and 10%. The blue graph shows the number of clusters found for variance of 10%. That means that two features are evaluated as similar if their relative pCluster values vary less than 10%. The red graph shows the number of clusters found for variance of 1%. The graphs show clearly the interrelation between the size of the subset and clusters found on the one hand and the variance of feature values and number of clusters on the other. For this dataset, no clusters are found on subsets greater than 10 features if I were to assume a precision of 1% variation. In the case of 10% variation, clusters are found on a feature subset of 15. The graphs show that for smaller subsets and higher variance, more clusters can be identified. So for this dataset it would not make sense to run the algorithm with the **size of subset** parameters higher than 15 in case of 10% variance or higher than 10 in case of 1% variance since the algorithm would then provide an empty result. A lower bound for the size of subset would be at 10 for 10% variance and 4 in case of 1% variance because all articles are contained in clusters for these values.

One aspect which is not considered here is how the user perception of quality correlates with the parameter values. In particular, the question over what the threshold for the parameters is in order for users to perceive two objects as qualitatively similar should be further elaborated. In the case of subset size and variation, an open question is the minimal size of the subset for which the results are perceived as similar and may thus be provided as useful recommendations.

6.3 Pattern recognition

In opposition to finding similar patterns in quality profiles, pattern recognition is aimed at finding a particular pattern in the chronological sequence of a quality profile (Bishop, 2006). The previous section describes the analysis of a quality profile at a certain point in time. That means one quality profile is compared to another profile. For the pattern recognition approach, I take the temporal development of metric attributes of one quality profile into account. Assuming that not only is the current status of metric attributes relevant but also the development of values, I then try to identify a recurrent pattern. For that reason, I analyze metric values recorded over a period of time.

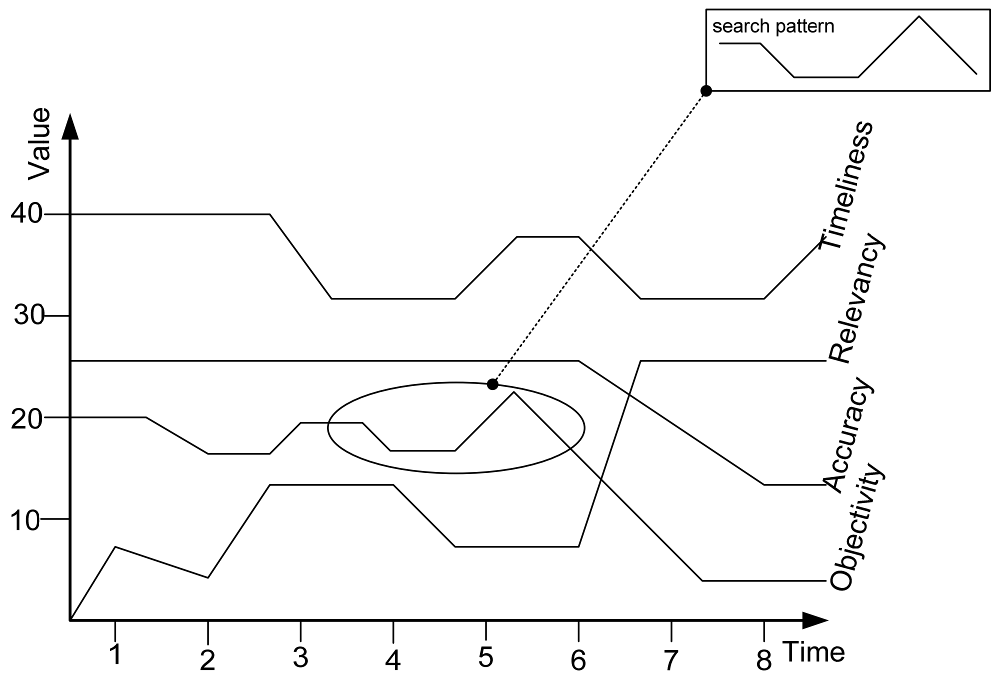
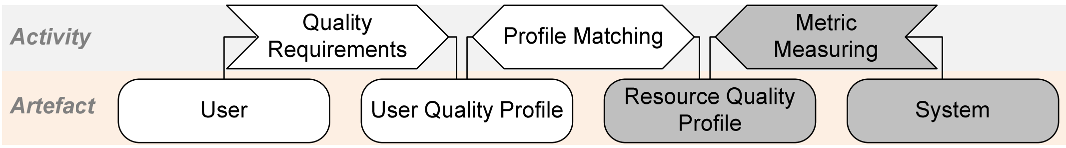


Figure 6-8 Pattern recognition in quality values

Figure 6-8 shows the development of metric values. In contrast to the former approach where the attributes were on the x-axis, this representation of metric values shows the progression of time on the x-axis. Pattern recognition is aimed at identifying patterns that are relevant for the qualitative status of a resource. One example for such a pattern could be the identification of activities for quality improvement (see section 3.4.1.3). Figure 6-8 shows the development of four quality dimensions over eight days.

Pattern recognition approaches have two different application areas. The first approach is to find patterns that are recurring and characteristic for a particular task. Here, the algorithm observes the values and finds *pattern models.* This approach is, for instance, used to identify task patterns in event logs (Dongen et al., 2005) but also for explorative analysis of patterns (e.g. gene analysis). The second approach is based on a given set of patterns. During pattern recognition, the algorithm identifies these patterns in the temporal development of values. The two vary in the input of the algorithm. The first approach is not supervised, the second is. The second approach is used to identify system conditions which must be recognized to trigger events. The pattern for triggering user notification can be defined. One example is the detection of vandalism in the Wiki (Potthast, Stein, & Gerling, 2008; Smets, Goethals, & Verdonk, 2007). When vandalism occurs, qualitative indicators do change over a longer period of time and so it can be detected by pattern recognition. First, the content-based metrics will change due to the edited content. After that, structure as well as usage indicators will change too since vandalized content is used and linked differently. Other conceivable scenarios are to detect activity patterns that correspond to qualitative development or to identify content which is often used but has a low quality especially as a basis for gardening activities.

7 SOCIAL SERVER - SEMANTIC ENHANCEMENT OF THE SOCIAL SYSTEM



One focus of this work is to evaluate the perception of quality in the context of community. In this sense, quality of artifacts can only be assessed by considering their interconnection with other artifacts or community members. This means that from a technical point of view, contextual metadata must be attached to the resources and users. When using existing socio-technical systems as a source of information, there are two problems. First, contextual information about the artifacts represented in the system is not available (or at least not sufficiently). Second, even if this information can be deduced from other sources, these systems have not the ability to store complex contextual information. As an example, a Wiki system does not provide information about which other users are interested in the same topics as me. But even if it is possible to extract this information from the Wiki (i.e. by searching authors that edited the same articles as me), there is no possibility to represent this relational information among users in the system.

Another problem is that even if complex semantic relations can be stored in the system (Völkel et al., 2006), i.e. users who are interested in the same topics, this information could be valid and useful to other systems as well. The problem in this case is the lack of exchangeability in semantic information between systems.

This section describes how data from a socio-technical system can be consolidated with a system-boundary-crossing semantic layer. This layer acts as a store for complex semantic information as well as a common repository for exchanging data between systems. In the course of this thesis and as part of the

MATURE [3] I developed a system that provide a common semantic layer interconnecting various entities from different systems – the Social Server. Subsequent sections will give a conceptual and technical insight into this development. Some aspects are not directly related to quality assessment but focus rather on supporting social interaction in a community as a basis for quality assessment.

7.1 LOGICAL PERSPECTIVE

The purpose of the Social Server is to establish a network of resources, the so called *Content Network*. This network connects all types of resources through a great variety of semantic relations to one common network. This allows for representing relations between users (social network), relations between digital resources (similarity, links, clusters), and relations between users and digital resources (authorship, expertise, recommendations). All knowledge containing artifacts with a unique identifier can be added as a digital resource to the *Content Network*. The purpose of this network is to provide a common semantic model as the basis for analysis and further processing. Since the knowledge entities, people, structure and content (cf. Section 5.1) and their interactions belong to the system, these have to be covered by the model. Semantic models that represent people and their interactions and relations are very common. The research area of social network analysis (SNA) proposes lots of modeling techniques for these models. In most cases, the models are represented in directed graphs where nodes are people and the edges express one particular relation or interaction (Carrington, Scott, & Wasserman, 2005). But there are also sophisticated ontology-based semantic models describing complex networks. Many of these relations can also be found in the Wiki. Examples of these relations are two authors writing an article collaboratively, two authors writing articles in the same category, people interested in the same topic, etc. Identical to people, resources in the Wiki also have relations among themselves. For instance, articles can be related because

[3] http://www.mature-ip.eu

they are hyperlinked or found in the same topical cluster. Articles can be related by tags or categories or because they have the same author. All these relations are implicitly enclosed in the Wiki database and the goal of the semantic model is to make them explicit.

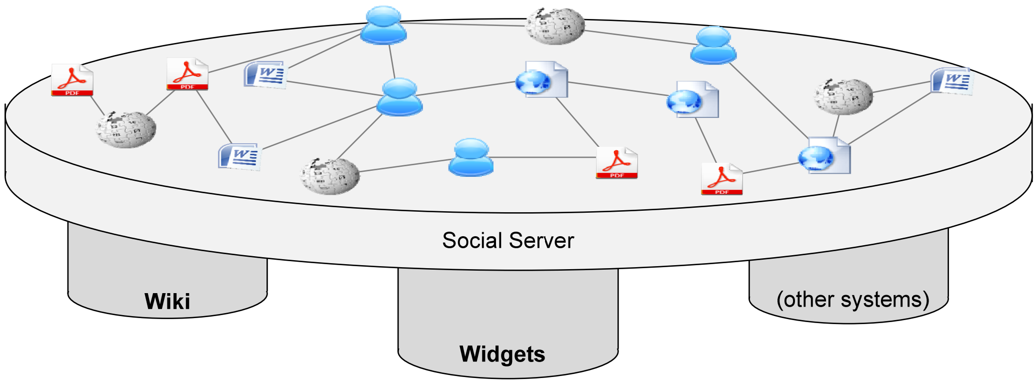


Figure 7-1 Content network – semantic representation of relations between artifacts like web resources, users and documents

As Figure 7-1 shows, from a logical standpoint, the Content Network can be seen as a layer on top of the existing system. The ability to represent complex semantic relations between various types of resources enriches the semantic expressivity of the underlying systems. The *Social Server* populates the content network with data from the underlying systems by analyzing system data. The *Content Network* can be accessed from any connected systems. Information provided by the network can either flow back into the source system(s) or can also be presented or further processed by other connected systems. In this way, the exchange of complex semantic information over system boarders is achieved.

7.2 MODEL PERSPECTIVE

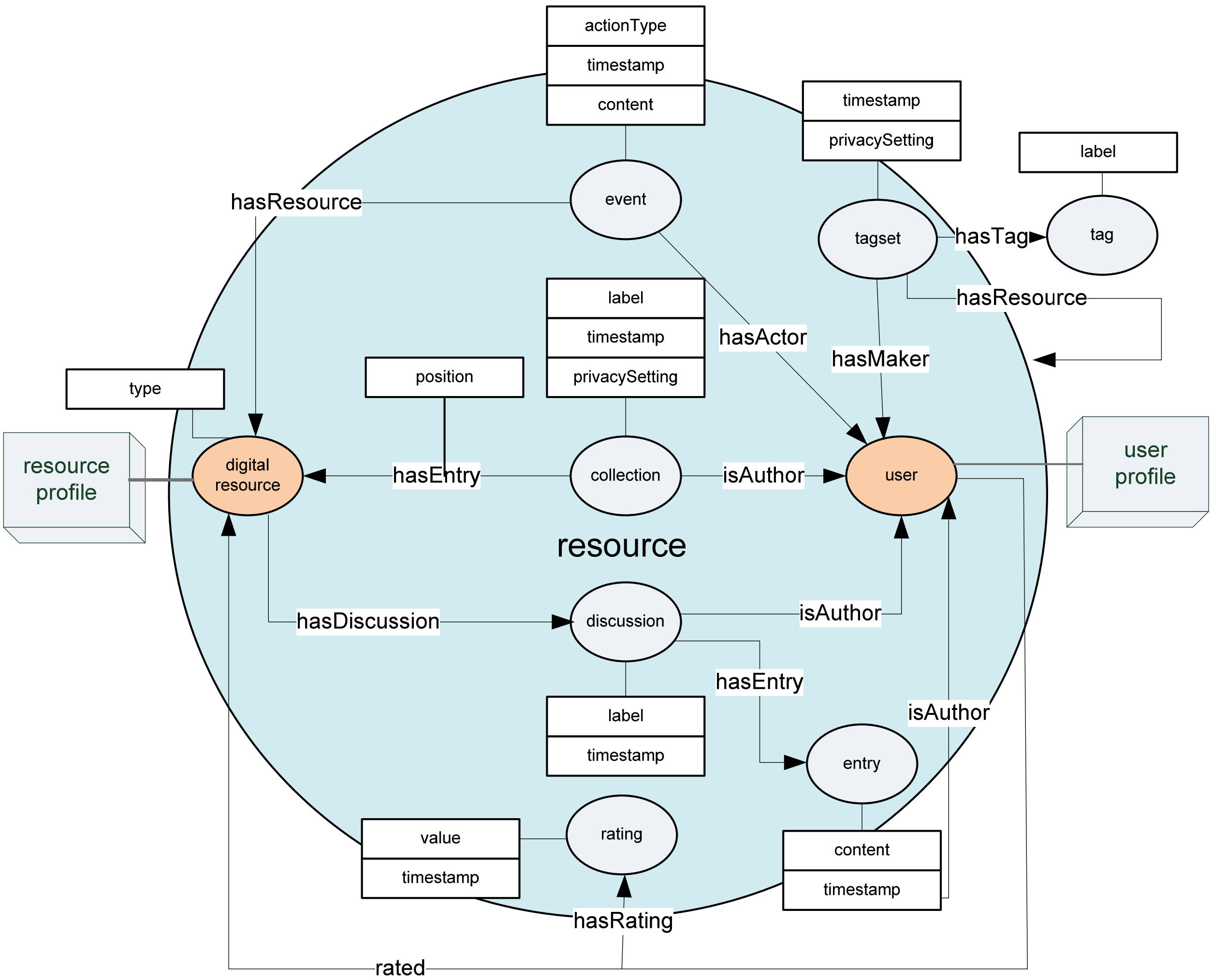


Figure 7-2 Schematic view of the semantic model including its concepts, attributes and semantic relations

Figure 7-2 shows the schema/ontology of the semantic model. The focus of the model is to represent the interactions and relations between users and resources, users among each other, and resources among themselves. Hence, the main entities are the **user** and the **digital resource** (red) which are interconnected by semantic relations. Through its relation to other entities, the digital resource is structured in many ways. By adding the resource to a collection, belonging to a certain cluster of resources is indicated. The discussion enables the user to track the negotiation process thereby leading to the current status of a resource. Thus, discussions are always related to resources. The rating represents an explicit assessment of the resource. The *isAuthor* relation specifies a user as actor for all activities. *Collections, Discussions* and *Ratings* always represent a component of information structuring between the actor and resource. Another instrument for structuring is used slightly different. Similar to collections, tagging relates tags to resources, and thus users to resources (through tags), and resources to resources (through tags). The difference here is that the tagged object is the resource itself enclosing all digital resources as are all other entities except tags (tagging tags leads to inconsistencies in the model). Concretely, this means that users can tag not only digital resources but also collections, discussions, ratings and other users. The resource profile and the user profile provide condensed information about the user and resource. By tracking the user interaction with the resources, these models are created and maintained. Since the profiles describe changes in the base model, these can be seen as metamodels, thereby representing a further level of abstraction.

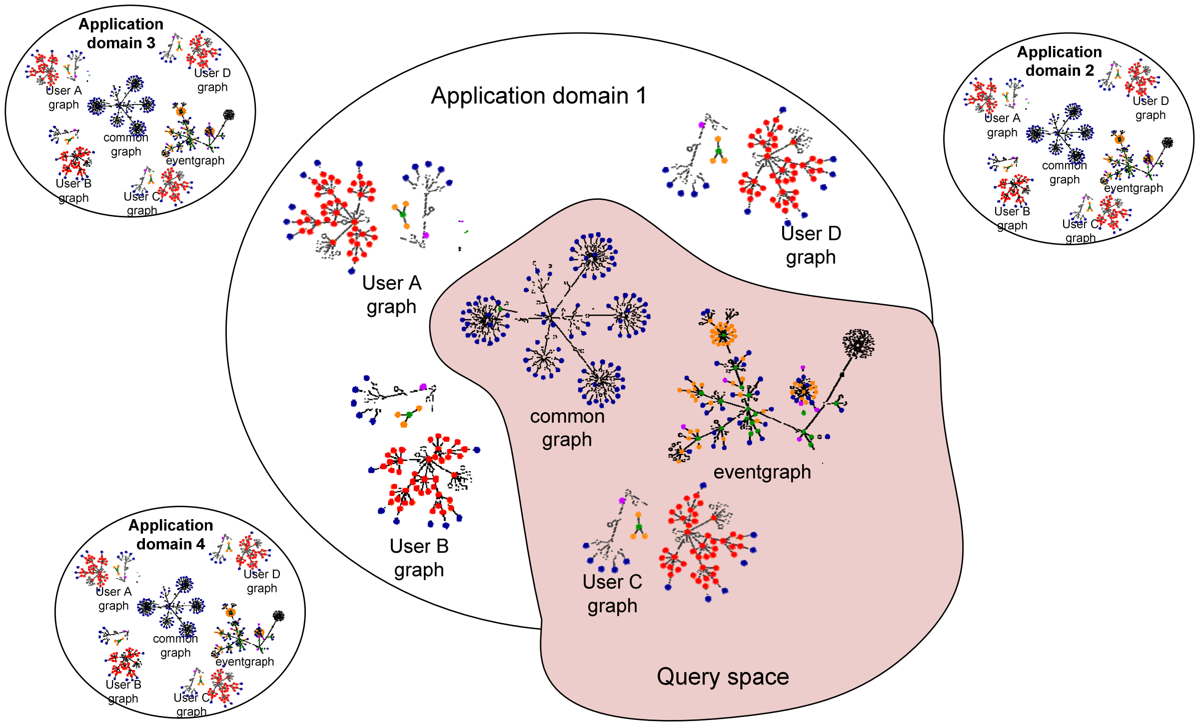


Figure 7-3 Multi-graph architecture – facilitating the definition of information space at query time

Additional requirements for the design and development of the modeling layer were meant to guarantee privacy on the one hand and to provide fast query responses on the other. An intuitive approach toward such a model would be to establish and maintain one graph containing all information which can be queried at runtime. For my approach, this turned out to have multiple drawbacks. First, the graph contains all information, including users' private information. The query engine cannot guarantee that a result would contain private data. Second, depending on the number of users and the amount of stored information, the graph grows continuously and reaches multimillions of triples. Due to the representation of graph structures and query performance being dependent on graph size, queries and modifications have (unacceptable) long response times. The solution for both problems is to split the graph into many independent smaller graphs. In our model, the graph is cut into logical pieces. Each user has its own graph where private information is stored; one graph contains all information that is generally visible. The specific feature of this approach is that the graphs are merged only at query time involving only graphs that are relevant. This guarantees privacy since graphs containing data that must remain hidden from the user are not part of the queried graphs thus keeping the queried graphs small to avoid a long waiting time for query responses. Figure 7-3 shows the different graphs in an application domain. The query space of a particular query assembles three graphs that contain possibly information relevant for the user. The *common graph* contains freely visible information; in addition, the graph of the current user is added because only the current (querying) user is allowed to see his/her private data. One special case is the *eventgraph*. For the reason that every interaction is logged in fine granular events contained in the *eventgraph*, this graph grows very fast and contains in large part the triples of the whole system. So, due to performance reasons this graph is only added to the query space if the evaluation of the event data is absolutely required (i.e. for the creation and maintenance of the user and resource models). In addition to the separated graphs, there are also separated spaces. A space is a group of graphs belonging to a particular application scenario. In contrast to graphs, space cannot be merged during query time. The idea behind this is to clearly border datasets from different contexts, e.g. if the server is used for two companies, each of them has its own space while accessing another space other than one's own is not possible.

7.3 Technical Perspective

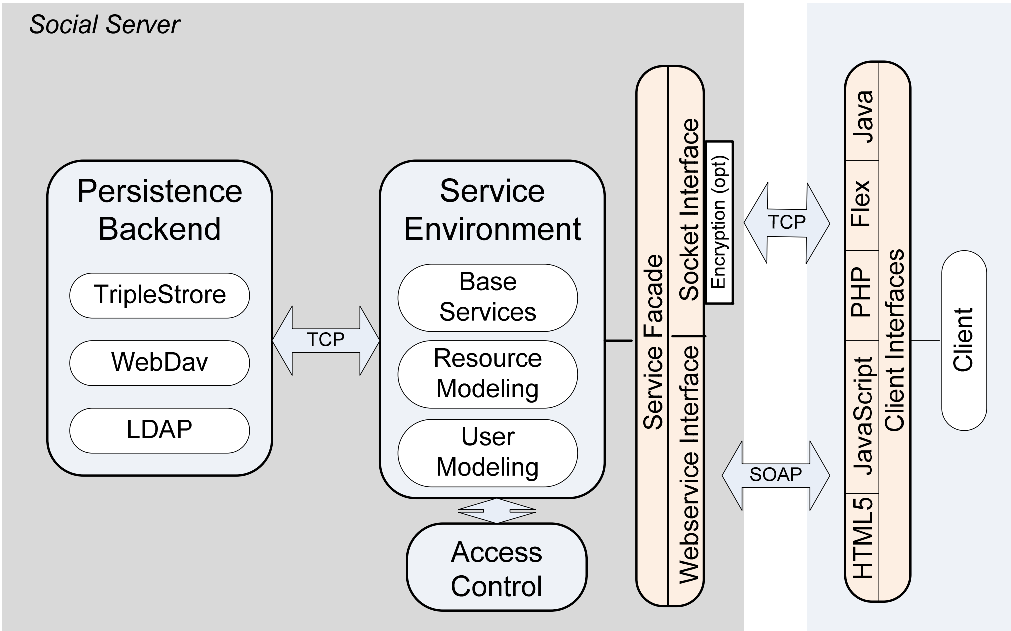


Figure 7-4 Architectural overview and technical infrastructure of the Social Server

In order to be able to access the semantic model, the socio-technical system connects through the client interface to the *Social Server*. One focus in the server development is to provide an interface that allows as many different technologies as possible. For interfacing the server, I implemented two different protocols. At the application level (HTTP), a SOAP based web service interface is provided. Due to the weakness of web services for this application case (immense overhead and procrastination in case of short payload) especially for time critical applications, I implemented a second interface channel. An evaluation of different protocols has shown that socket connections transferring data serialized in JSON turned out to be an approach with high scalability and the best performance. One problem in contrast to web services (WS-Security) is a lack in the capability to provide security functions. I compensated for this by adding access control functionalities in an extra module as well as an optional encryption function.

Figure 7-4 shows the programming languages of the already implemented clients. Furthermore, the chosen interface technology allows for accessing the client in almost every programming language.

The Service Environment consists of different sets of services. The *Social Server* provides a flexible and intelligent environment for services including some basic services. That means the services can easily be added to improve functionality. The services are grouped according to their level of abstraction from the system (cf. Figure 7-5).

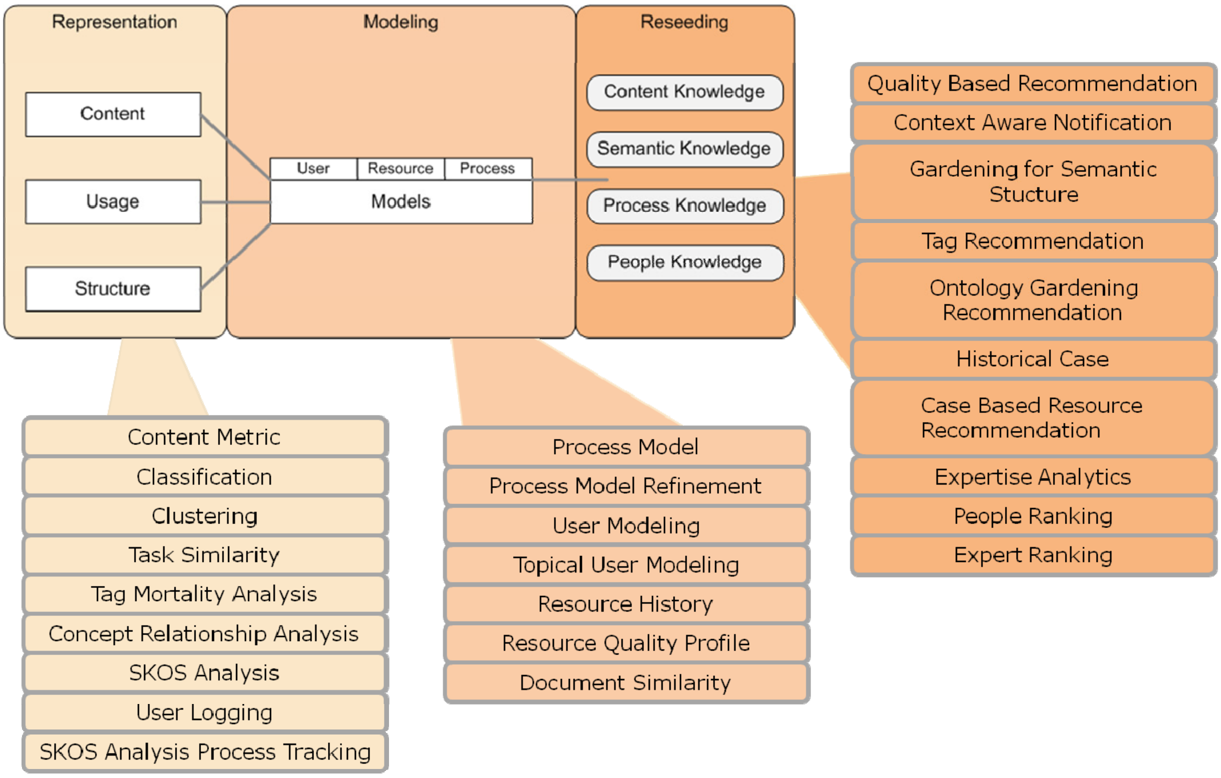


Figure 7-5 Implemented services included in the service environment

Figure 7-5 shows a conceptual overview of the service categorization relating to the *Social Server* service environment. The service categories are explained in more detail in the subsequent section. Moreover, the figure shows services already implemented for each of the categories. These services have been developed by different partners in line with the mature project. As contribution to the project and in course of this thesis I have designed, developed and implemented all information quality related services. These are:

- the Social Server as service environment and infrastructure
- the persistence backend
- the base services to provide a basis-functionality
- the server interfaces
- access-control and encryption
- client applications for MediaWiki
- quality related parts of the Widget interface (resource quality summary)
- all quality related services:
    - content, usage and structured metric services
    - resource quality profile service
    - quality-based recommendation service

7.3.1 Representation Services

Services at the *Representation* level create formal representations of the underlying system. These representations are the basis for other services. According to the metric representation model, these services are grouped into *content, usage* and *structure.*

7.3.1.1 Representation of Structure

The representation of the system structure covers all entities and the relations upon which entities are linked within the system. Particularly in Web-based systems, one important relation is the hyperlink structure. Nodes in this graph represent web pages whereas edges represent hyperlinks between web pages.

Figure 7-6 MediaWiki link graph visualisation

The structure representation is stored in a Java object link structure with export function for GraphML. I have chosen GraphML (Brandes et al., 2002) because of its XML-based input format which is widely used in many graph analytic frameworks. Due to its flexibility, it can be used in many different contexts and has the power to model even complex facts. The following snippet shows the definition of a simple directed graph.[4]

```
<graph id="G" edgedefault="directed">
  <node id="n0"/>
  <node id="n1"/>
  ...
  <node id="n10"/>
  <edge source="n0" target="n2"/>
  <edge source="n1" target="n2"/>
  ...
  <edge source="n8" target="n10"/>
</graph>
```

Figure 7-7 GraphML example

Furthermore, the features of GraphML are the following:

- **Attributes:** to store additional data on nodes and edges. These are used in this context to store semantic metadata about nodes and edges, i.e. edge weight, node type, etc.
- **Parse Info**: since GraphML is not used for dynamic modeling due to performance reasons, some information is already calculated in the system graph. In order to avoid recalculating this information, it is stored as parse info and can be reconstructed during parsing.
- **Nested Graphs:** to represent different granularity levels of a webpage, graphs can be nested in nodes. This means graphs can be handled as nodes so that a network of networks can be established. In the context of hyperlink structures, a node can represent a web page but the web page itself can also be a network (of content objects).
- **Hyperedges:** edges can connect more than two nodes. The endpoint of an edge is an arbitrary number of edges.

[4] http://graphml.graphdrawing.org/primer/graphml-primer.html

- **Ports**: an edge is outgoing from a certain node and incoming to a certain node. Ports allow for defining a logical location for the outgoing and incoming end of the edge. Two edges can have the same source and target nodes but at different ports.

7.3.1.2 Usage Representation

The interactions of users with the system are represented in an event-based format. Similar to Wolpers et al. (2007), events are the central concept in our representation model. Since I focus in this implementation only on one system, viz. the Wiki, I have chosen a less general but more lightweight approach. The approach can be seen as a combination of Wolpers et al. (2007), where I got the event-based representation, with Rath and Lindstaedt (2009) for the application of specific action types. The result is a lightweight, easy to process and flexible event format which therefore forfeits somewhat on expressivity. Previous approaches have shown that particularly in working with fine granular events, performance and time consumption are critical factors. The lightweight representation approach chosen for this work is intended to overcome this barrier.

Basically, events are a representation of users' interaction with the system. Therefore, our usage representation only consists of three concepts: user, event, and resource.

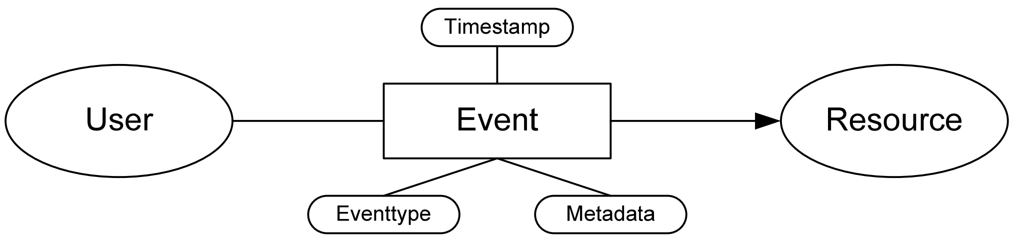


Figure 7-8 Lightweight event format schema

Figure 7-8 shows the semantic event structure. The event always represents an interaction of a user with a resource (the resource can be empty if no resource is involved). The concept event is annotated with a timestamp, the type of event and event-specific metadata. For instance, if a user tags a resource, the metadata would contain information about the tag.

7.3.1.3 Content Representation

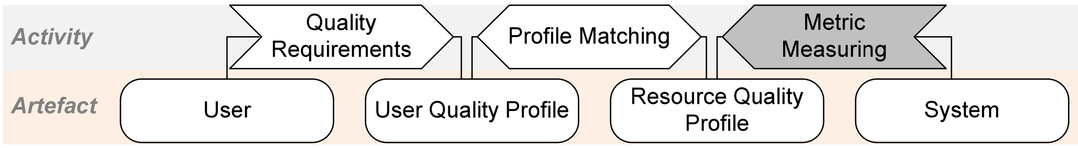
The content representation of the system consists of a collection of textual content. In this implementation the content consists of four parts:

- **Content id**: the unique identifier for content object. This identifier is used in all system representations to identify the same content object. In this way, information about content, structure and usage of one artifact can be interlinked
- **Textual content**: the current content of an article at the time the representation is created or last updated
- **Content revisions**: the array of content objects that represents the different revisions of one article
- **Formatting** and **structuring**: the information about the format of the text. Depending on the system, this information is stored directly within the content (e.g. MediaWiki content); in other systems formatting information is stored in metadata (e.g. HTML)

7.3.2 *MODEL SERVICES*

The objective of model services is to create an abstraction from the representation layer. The two services that are most relevant for this work will be described in the following.

7.3.2.1 Creation of Resource Quality Profile



The objective of the metric measuring process is to obtain values describing the attributes of the system. These values will be used as a basis for creating the profile of the resources. Basically, the intention of this process is to measure several aspects of the information system. Similar to a metering rule that is used to measure one aspect of an object (e.g. size), in this context metrics are used as tools to gather data about an article. As there are many aspects that can be measured, various metrics are needed for evaluating different aspects of an article. From the standpoint of implementation, the question here is: which metrics should be used for the determination of article quality? One requirement for metrics is that they are to be commonly used in order to get comparable results. According to the metering rule example, one would use a standardized unit of measuring (e.g. meters) instead of one's own in order to be able to compare the height of a piece of furniture to the height of a room. Another advantage of using quasi standardized metrics is to benefit from existing evaluations. All metrics presented here are used in other research works as well and in particular their influence on quality is already evaluated and shown. So, if I use reading ease scores, there is no need to evaluate if there is a correlation between the perception of understandability in texts and the value of the metric. The same applies to graph metrics like *Betweenness Centrality* whose significance has been clarified in many research papers. Another important aspect is that the set of metrics cannot be fixed. As described in Section 2.1, there is only an indirect relation between quality perception and metric values. That means different metrics can be used to give information about one quality attribute. Which metrics will be needed depends on the user quality requirements and the mapping between metrics and quality attributes. Therefore, to make metrics exchangeable and to allow for adapting the metrics to the domain, the implementation is equipped with a plug-in infrastructure for metrics.

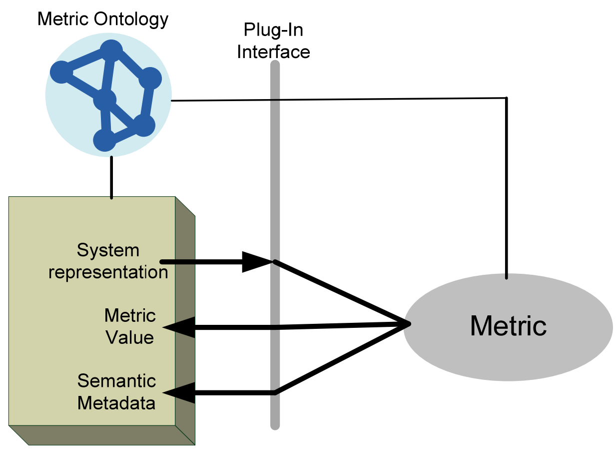


Figure 7-9 Plug-in architecture for flexible inclusion of various metrics

By using the metric interface, additional metrics can be added at runtime. The basis for calculation is always one of the system representations. Depending on the type of metric and representation of the content, the article's relation or usage is provided by the interface. The result of the metric calculation is always a double precision value. In order to be able to process this value, additional semantic information is provided at the interface. The following metadata must be provided by the metric:

- Metric id – for comparison of metric values among several articles
- Metric mapping – mapping between quality attribute and metric for interpretation of values
- Metric range – for normalization of values

The resource profile represents an instance of the resource model. The resource model contains concepts and attributes relevant to the qualitative status of a resource. I used a two-layered model to represent the quality attributes on the one hand and the metric values on the other. Because of this approach, I am able to adapt/adjust the semantic model at runtime. This is required particularly when changing used metrics. The upper layer is static and provides the connection to the profile matching algorithms. In this way, I am able to ensure that required information is available in the resource models. The concepts from the upper layer are related to the metric-specific concepts on the lower layer. These concepts are provided by the metric ontology. The metric ontology also provides the relation to the upper level concepts. This means that if a new metric is added, such a metric is equipped with its own (part of) ontology describing the metric as a concept through its attributes and establishing the interconnection to the static part of the model.

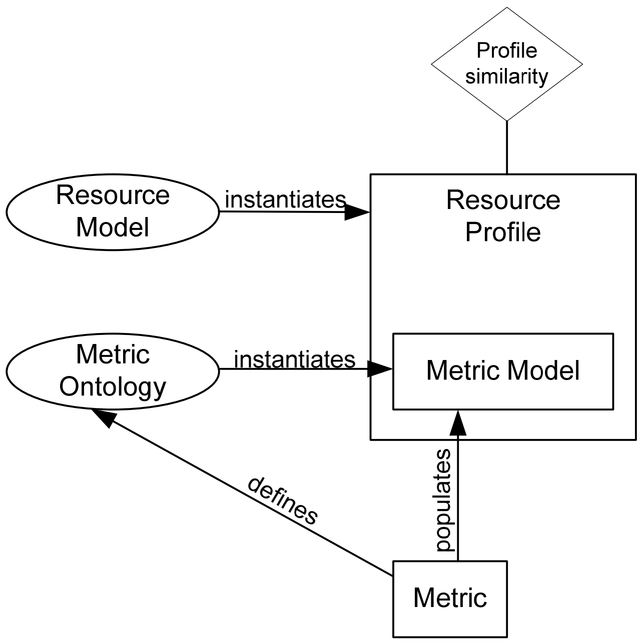
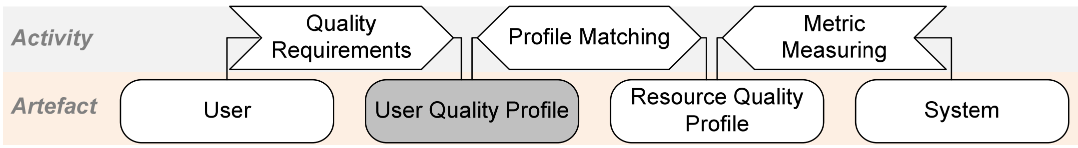


Figure 7-10 Semantic schema for quality metadata

The metric model is one part of the resource profile and contains the status of the resource regarding a certain metric. Due to the different system representations on which the metrics are working and the different implementations of the metrics, they are very diverse. But this affects only the input side of the metric. The output is characterized by the metric model and, apart from different attributes, similar for all metrics. Thus, the description in the metric model is required for all metrics in order to fit the resource model. Hence, the concept is flexible and metrics with new features can be integrated at runtime by adapting the metric ontology.

Another important aspect is the temporal component. As already mentioned, not only the current status is important, but also the temporal component. Therefore, the resource model tracks the history of metric values by storing snapshots of the metric values at several points in time. This facilitates tracking the development of a resource as well as searching for patterns in the resource history.

7.3.2.2 Representation of user quality profile



The user quality profile defines the quality requirements of the user. These are represented in a way similar to the resource quality profile in quality dimensions. Different to the resource quality profile, the meaning of the dimension values does not represent the current status but rather the weight of this dimension's importance. This means that if a user attaches importance to good readability of texts, this would imply that resources which are recommended to the user are good in terms of readability. Therefore, the value for readability would be high in the user profile.

One observation in the evaluation (see Section 8.3) was that quality requirements are not directly tied to the user. In fact, users can have different roles in which quality requirements are completely different. Even within these roles, different tasks with different quality requirements can be fulfilled.

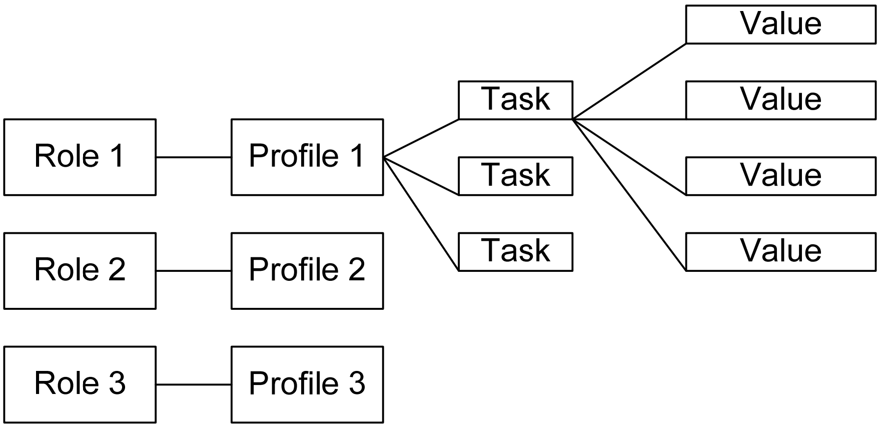
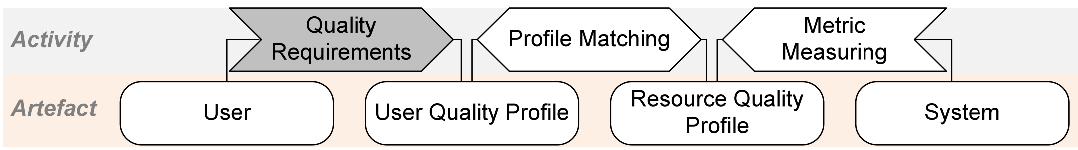


Figure 7-11 User quality profile schema

Therefore, the implementation of the user profile does not represent a user but a particular role of the user. Each profile consists of several task descriptions (at least one) for each task that is fulfilled by a person in a particular role. Thus, the quality dimensions are unambiguously defined depending on their mapping to a task, role or user.

7.3.2.3 Gathering user quality requirements



Wiki questionnaire

The process of gathering requirements provides information about user quality profile to the general public. The objective of this step is to gather information on how quality dimensions are weighted from the user perspective. Therefore, different approaches have been tested. The first is the explicit approach. This approach is characterized by asking the user directly about the importance of the quality dimensions (see Section 7.3.2.3). This can be done, for instance, by using a questionnaire. Figure 7-12 shows the integrated questionnaire which was evaluated in a study (see Section 8.3).

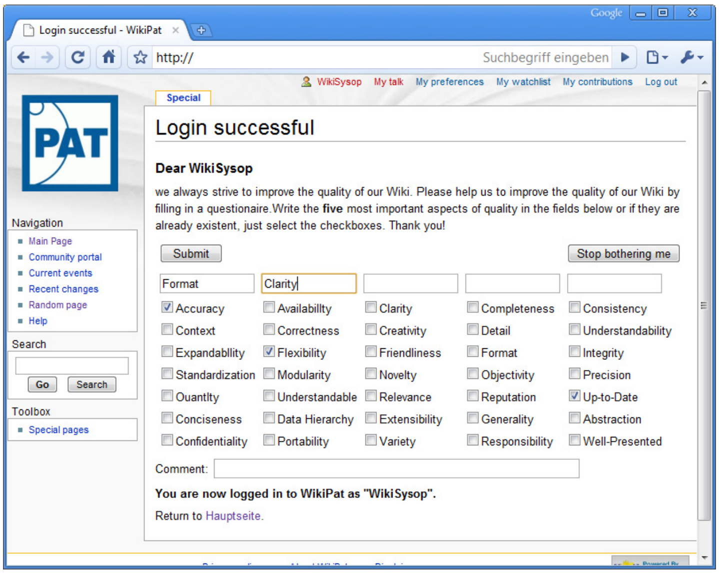


Figure 7-12 Wiki questionnaire plug-in

The questionnaire (see Figure 7-12) is integrated in a Wiki and is presented after login. One simplification is to refrain from exact weightings (e.g. a Likert scale) in order to make the questionnaire easier to use and lower the barrier for filling in the questionnaire. The result of the questionnaire provides an explicit description of users' quality needs.

Wiki rating

The implicit approach collects information about the users' quality requirements during system usage. Particular weightings of dimensions can be derived from user behavior within the system. The ability to derive user quality requirements from system usage depends on the interface/functionality of the system as well as the detail level/accessibility of the system login (in case of evaluating usage behavior from system logs). In the case of the Wiki system, various information about user activities are available in the articles. The article history provides information about their development process. This includes all types of editing like adding and deleting content, but also discussing the article. One functionality that was missing is rating which makes it hard to get information about the users' quality perception regarding an article. Therefore, I have implemented an easy-to-use rating functionality for Wiki articles.

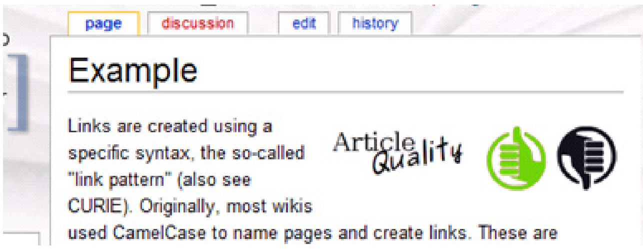


Figure 7-13 Wiki rating plug-in

The rating buttons are integrated in each article; the user can give his/her rating by clicking either the green (good quality) or black (bad quality) button. In this way, the user makes an explicit statement regarding the quality of an article and also selects implicitly quality attributes. Therefore, after each rating, all available metrics calculate the values for the page that was rated. If the rating is positive, the system searches for metrics which show high values for the given text. Since each metric in the system is connected to a quality attribute, this method implicitly provides candidates for quality attributes. For assessing in a particular case whether the value of a metric is high/low, the deviance from the median of the Wiki text corpus is calculated. The following formula shows how metrics are implicitly selected based on user rating. *M* represents the Metrics, P the article, *i* as id (from 0 to n), *Mcurrent* is the current metric, with T as the threshold for a metric.

$$\left| M_{current} - \frac{\sum M_x(P_i)}{n} \right| > T_{metric}$$

Example: A user rates ten articles as *good quality* (green thumb). For all articles, the values for the RES metric (Laughlin et al., 1969) and the interaction metric are very high. The RES metric is connected to the *Readability* attribute because it correlates with the readability of the text. The interaction metric shows that the article is updated very often; it is connected to the quality attribute *Up-To-Date*. Since the user apparently perceives articles that are easy to read and up-to-date as high quality articles, these attributes are stored in the user profile and can be used for future recommendations.

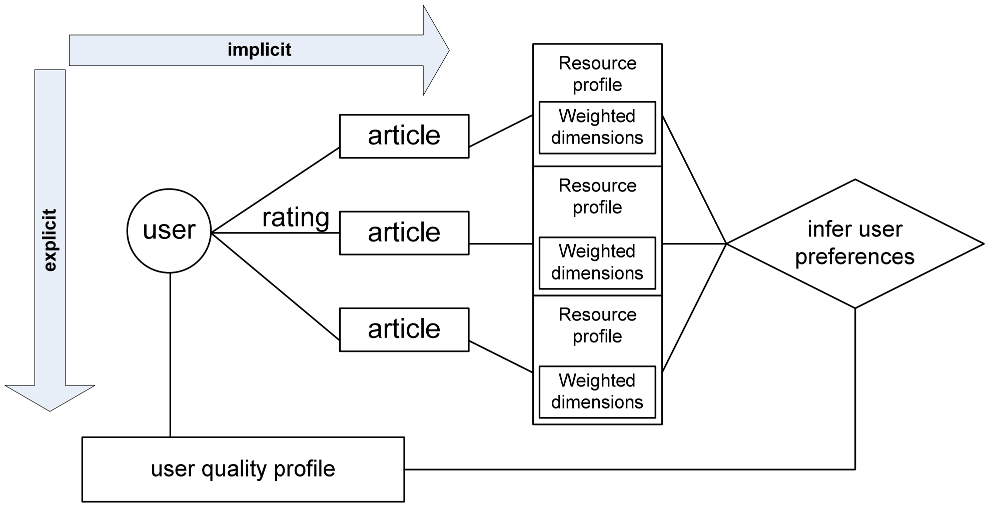


Figure 7-14 Implicit and explicit gathering of profile data

Figure 7-14 shows the implicit and explicit gathering of user quality requirements. Implicit gathering is done by rating articles and inferring the user preferences from the resource profile, whereas explicit gathering provides requirements directly via a Wiki questionnaire.

Wiki learning paths

Although rating implicitly provides quality requirements, to a certain extent the rating itself is explicit. In Weber et al. (2011), I presented an approach for creating learning paths out of Wiki content. The creator of a learning path can easily add existing articles by dragging and dropping them into a learning collection. The selection of an article as learning content can also be seen as quality statement. Since almost all quality dimensions are relevant for learning content, I thus infer that the user considers the selected content as *high quality* (of course besides the correctness with regard to the content).

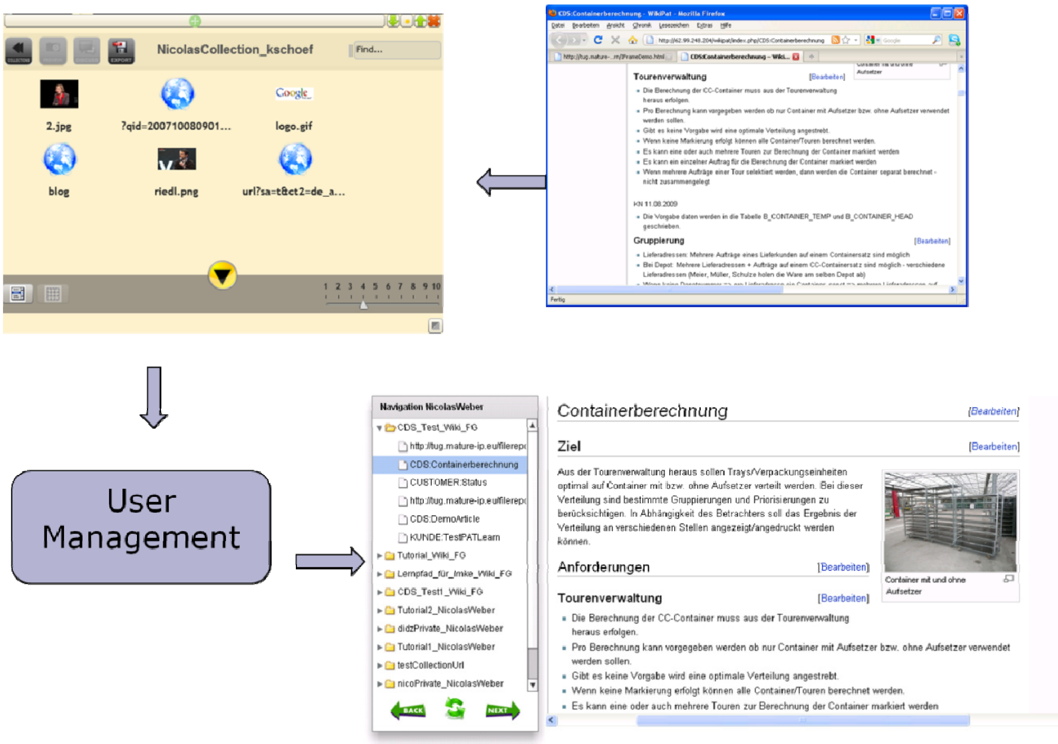


Figure 7-15 Learning collection authoring environment

Figure 7-15 shows the authoring environment for learning paths. Content from the Wiki can simply be added by *drag n´ drop* to a learning collection. At the same time, the author implicitly rates the quality of the content and so defines his/her quality preferences.

7.3.2.4 Implementation Details

The Wiki questionnaire and the rating plug-in are implanted as MediaWiki[5] extensions. In order to contribute to MediaWiki development, both extensions are available as open-source. The extensions are written in PHP and use MediaWiki hooks to get notices about activities in the Wiki. The rating function is embedded in the HTML compilation and adds the button to every article page. The questionnaire is integrated into the login process and is shown after every login.

http://www.mediawiki.org[5]

The information given by the user, either by rating or by filling in the questionnaire, is transferred to a central server where profile data is maintained and evaluated independently from the Wiki. In this way, information from different Wikis can be evaluated in one single system.

Widget Personal Learning Environment (Widget PLE)

In the case of the widget that was developed in MATURE, no explicit definition of user quality requirements is intended. Populating the user quality profile is done in conjunction with the user profile. Therefore, several mechanisms have been implemented in order to be able to track the users' needs (MATURE Consortium, 2010).

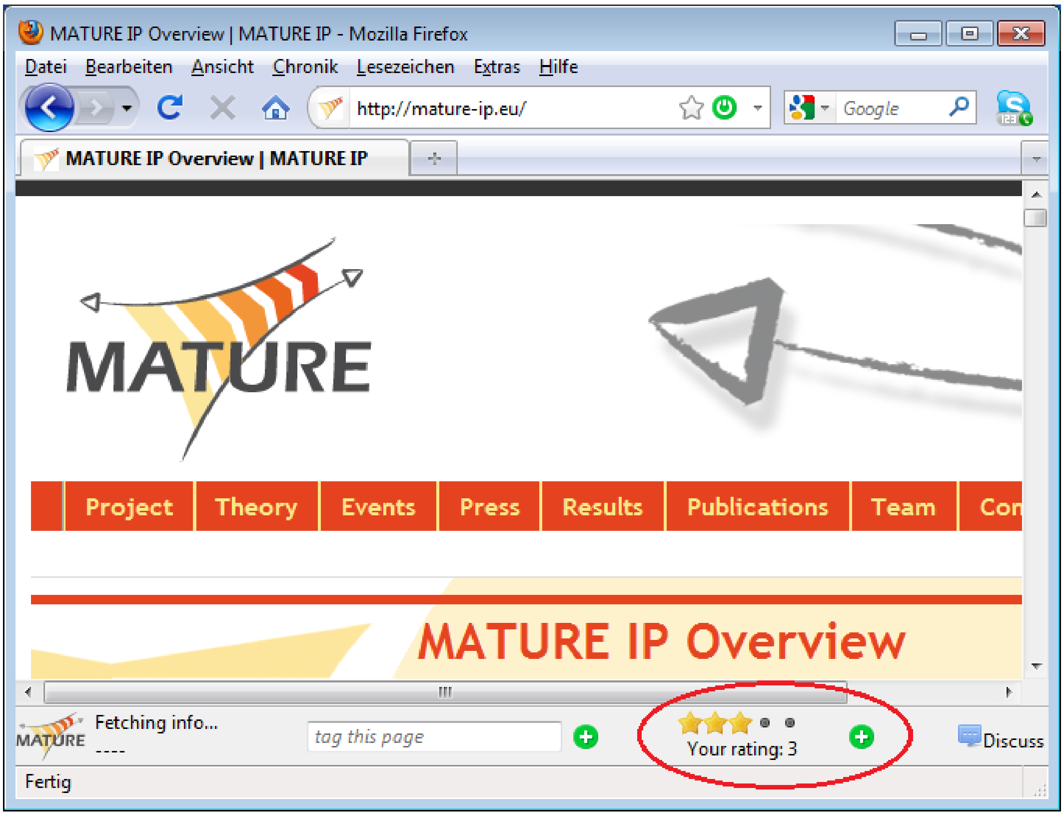
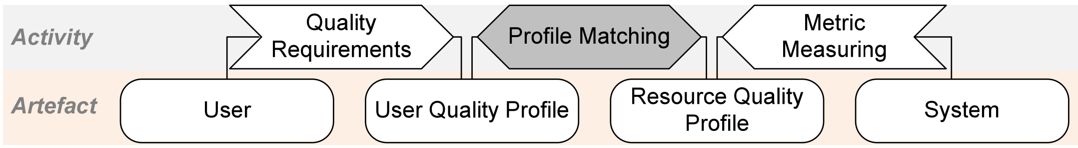


Figure 7-16 Firefox rating extension

Figure 7-17 Widget rating extension

Amongst others, the Widget PLE contains rating functionalities as well. In this approach, the rating functionality is not limited to web resources. As shown in Figure 7-16 and Figure 7-17, two different rating interfaces are provided. The browser-based plug-in focuses on web resources, like websites, social networks, blogs and Wikis. The simple rating function within the browser improves the acceptance and provides quality assessment from different sources. The rating widget provides the same functionality for desktop resources. Since in an organizational environment, most content is still stored in desktop resources like word or pdf documents, this widget facilitates the rating of local and intranet documents.

7.3.3 *RESEEDING SERVICES - QUALITY PROFILE MATCHING*



The profile matching process consists of several steps. To some extent, it is always about measuring similarities between profiles. Depending on the recommendation task, different similarity measures can be calculated depending on the entity represented by the profile. The user profile is compared to other user profiles in order to identify users with similar interests. The resource profile is compared to other resource profiles in order to identify group resources according to their qualitative status. The resource profile is compared to the user profile to find resources matching the users' qualitative needs. The following scenario shows the application in context of an organizational Wiki. Initially, the system has to establish the different profiles. This happens in two areas. In the resource area, representations of the systems are created. In the case of Wiki, three representations are created and maintained over the runtime of the system.

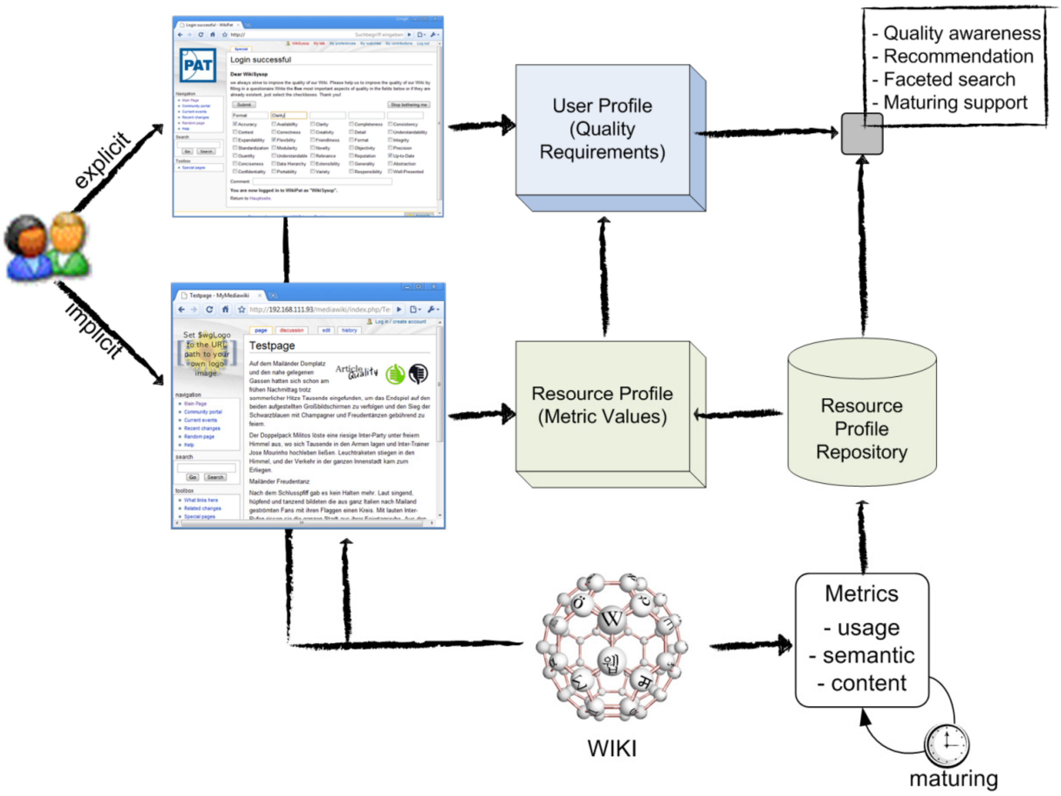


Figure 7-18 Wiki example for quality profile matching

In the Wiki (similar to most other information systems), two main entities are differentiated: agents (users) and artifacts (articles). For both of them, their qualitative statuses/requirements are represented in quality profiles. The implementation provides two semantic repositories for both of them. These contain semantic representations of the profiles and also relations describing the similarity between profiles.

For calculating the similarity between profiles, the modified pCluster algorithm described in Section 6 is applied. Due to performance improvement, I have set out two different ways of using the algorithm: periodical calculation, and event triggered calculation. The periodical calculation is triggered by a timer and updates information about similarities in both repositories. Therefore, the whole dataset is clustered into quality clusters and the information whose entities are in the same cluster is stored in the similarity relations. In order to avoid slowing down the system due to the amount of semantic relations (*n to n* : n^2), only clear similarities over a particular threshold (and a particular size of the subset) are added at this time. Since for some metrics the calculation is only possible on the whole dataset (e.g. PageRank) and metric values are a prerequisite for profile updates, the clustering process happens in combination with the updating of metric values. Usually, this update mechanism runs at night when the system load is low.

The second type of calculation is event-triggered. That means there are certain events defined in the system which trigger a similarity calculation. Mostly, these calculations use only a subset of entities as input. Triggers can be resources that are newly added to the system and for which similar resources should be found or users whose profile changes thereby resulting in a recalculation of quality mappings. Basically, this type of calculation is applied either when resources change or if special results are needed. *Special* means in this case that a resource with a higher or lower similarity should be returned. In some cases, it is required that the entities are almost identical regarding qualitative statuses/needs. In other cases, it is more important to provide a large set of resources. In these cases, parameters of the clustering algorithms are changed to meet the requirements. Due to the fact that these calculations happen on a subset of data, results are mostly available within seconds, whereas calculations for updating the whole database may take up to five minutes (~1000 articles, 50 users).

8 EVALUATION

This section conducts an evaluation of my approach to quality awareness in social software systems. Due to the many facets of the approach, conducting one overall evaluation turned out to be not appropriate. The evaluation described in this section is divided into several levels of system abstraction. Abstraction in this context means, for instance, that system measures do not contain the content of the system; they are rather an abstract representation of the system. Based on the measures, the content of the system cannot be restored but the measures can be seen as abstract representation. Similarly, profiles are created out of the measures, inferring information from the measures but losing the values of the measures. So every step represents an abstraction from the system.

The advantage of splitting the evaluation into several parts is to be able to analyze different aspects of my approach and to apply different evaluation paradigms according to the level of abstraction.

Furthermore, the different parts of the evaluation also reflect the design and development process of this work.

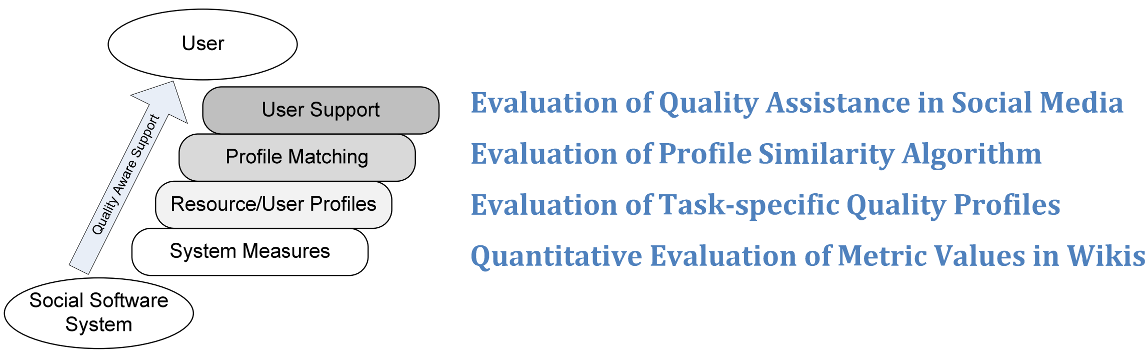


Figure 8-1 Parts of evaluation – Evaluation of quality aware user support in the social software system

Figure 8-1 shows the different levels of system abstraction and the corresponding evaluation sections. Basically, the aim of the evaluation is to analyze the benefit of information quality awareness in social software from the perspective of a user. Thus, all parts of the evaluation analyze aspects of the user-system interaction at different levels. At the lowest level of system abstraction, metrics providing system measures are evaluated. A combination of explorative data analysis and descriptive statistical methods seem here to be an adequate evaluation paradigm. The approach and results can be found in Section 8.1

From the low-level system measures, I conducted a step for consolidating and enriching the system data in user/resource profiles. This can also be seen as a level of abstraction from system measurement. The qualitative status of a resource and the quality needs of a user can be best assessed by the user itself. Thus, the evaluation paradigm I chose is the empirical approach. Here, the user rates the relevance of profile features by filling in a questionnaire. The results and the analysis of these results by means of descriptive statistics can be found in Section 8.2.

At the next level of system abstraction after the user and resource profiling, my approach describes how to determine similarities in these profiles. In this step, the profiles are further enriched with information about the similarity of profiles. The aim of the study was to compare the user's quality perception with the automatically calculate quality status of a resource. Therefore, explicit user rating has been compared to quality profiles. The detailed description of the approach and results can be found in Section 6.

At the highest level is the user interacting with the system. In order to evaluate the user experience while using the social software, two systems have been equipped with quality-sensitive user support. To gather information about their usefulness, users were observed while working and then interviewed after working. The results of this study can be found in Section 8.4.

8.1 Quantitative Evaluation of metric values in Wikis

The first of the four studies described in my approach deals with results from metrics directly applied to the system representation. The aim is to understand the expressivity of metrics as an indicator of quality in social software systems. The results provide a basis for the second part of the evaluation, namely, the creation of quality profiles out of metric values.

8.1.1 Objective

This study examines the characteristics of metrics applied in Wiki systems. The objective is to analyze and understand indicators for qualitative improvement of resources. In order to be able to automatically identify and recommend resources, meaningful indicators for particular quality dimensions are necessary. Therefore, representative metrics for three system representations – usage, structure, and content – were evaluated over a longer period of time. The statistical evaluation of the metric development provides information about several interconnections and characteristics of the metrics which are important for the mapping between metrics and quality dimensions. This study shows how qualitative changes in Wiki systems are reflected in different dimensions of representation. The study is aimed at answering the questions: How can one apply metrics in a Wiki system? Do the metrics reflect the editing process in a Wiki? Which metrics should be combined for measuring certain quality dimensions?

8.1.2 Study Context

8.1.2.1 Data Collection

The data set used to evaluate the metric values was provided by several Wiki systems that were used as knowledge base in a lecture about Knowledge Management at the Graz University of Technology. In order to get comparable results, the students were grouped into 6 groups of 50 students.

As stated in Section 2.1.1, communities have a common understanding of quality. Therefore, an important prerequisite for this study was the membership of the participants in one community. Given the group of student-participants in the study, this prerequisite is fulfilled according to the definition of communities of practice (Wenger, 1998). Since all students were from the same study course, visiting the same lecture and executing the same tasks, the common interest as well as the common goal is a given for this group. One difference between this group of participants and users on the web is that one cannot expect altruistic behavior as is often observed in social software systems on the web. But for this study, this effect is only of minor importance.

The task assignment was similar for all groups: The students should write articles according to the content of the lecture. Each concept from the lecture should be one article in the Wiki. The result should be a representation of both the textual description of the topics and the interconnection between the topics. The creation of articles was done collaboratively within each student group. Thereby, students documented the concepts of the current lecture in the Wiki. The Wiki was edited over a period of 3 months. At the end of the study, each of the six Wikis contained between 500 and 850 articles. Besides the comparability of the results, another reason for splitting the user group into six groups editing one Wiki was to be able to identify anomalies in the Wiki usage. These anomalies, e.g. abnormal editing behavior, would distort the metric results. The overall number of articles analyzed in this study is approximately 3,200.

In order to track the maturing process of the articles and to analyze (qualitative) changes in the Wiki, snapshots of the Wiki systems were taken. Therefore, the current state of the Wiki was saved every day. This resulted in 120 snapshots for each Wiki (720 snapshots in total) in the usage period of 3 months. Within the snapshots, the Wiki is represented by its structure, content and usage model and that way, suitable representations for all metrics are available.

Figure 8-2 shows the creation of the dataset for this study. During the Wiki usage period, I took snapshots every day for each Wiki. These snapshots contain content, structural information and usage data. In the course of dataset creation for each snapshot, I calculated the metrics from a set of measures. This results in a dataset with values in a timeline for each of the metrics.

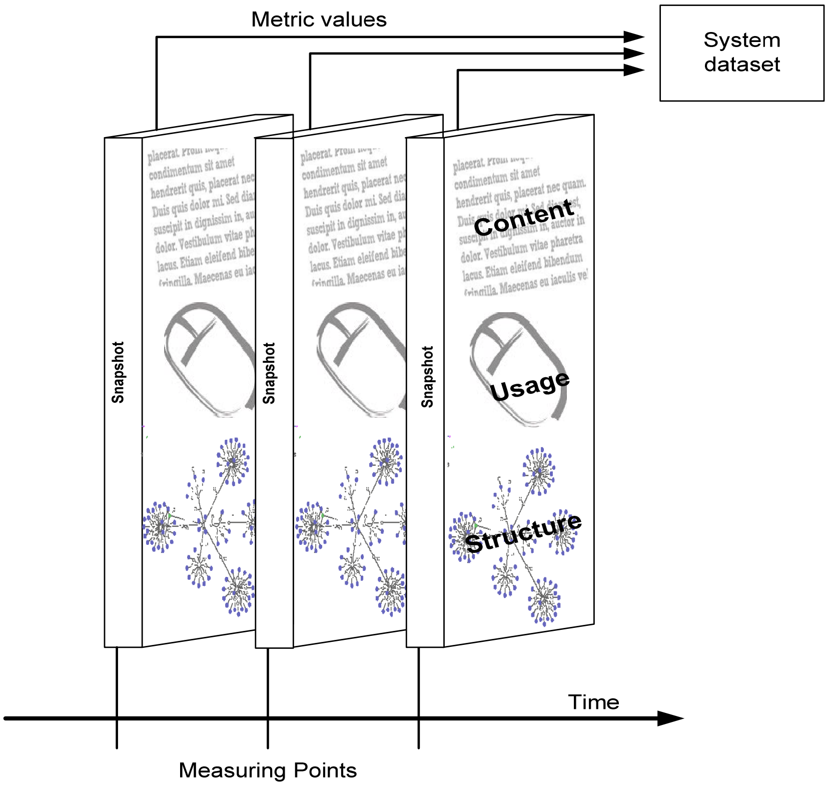


Figure 8-2 Tracking change by means of Wiki snapshots – the change of Wiki content is captured in snapshots. The development of metric values can be observed by arranging the metric values of the snapshots in a timeline

8.1.2.2 Characteristics of the Data

All articles in the study have a unique identifier that allows for identifying one article in each of the representations – content, usage, and structure. The content representation simply consists of a mapping between the text of the article and the unique identifier. The usage dimension contains a temporally ordered list containing all events that are related to an article. This can be editing activities like writing/deleting text or creating an article but also reading activities and page accesses. The structure representation consists of a graph representation containing three types of nodes – articles, users, and discussions. Relations represented in the graph are:

- article – article: One article hyperlinks to another, directed
- user – article: A user is author or one of the editors of the articles, directed
- article – discussion: Discussion about an article, directed (not evaluated)
- user – user: Co-authorship, users are editors of the same article, undirected (not evaluated)

Figure 8-3 shows the graph representation of the link structure for one Wiki at day 75 containing more than 400 articles. Nodes in this picture represent articles in a Wiki, edges between the nodes represent links between articles.

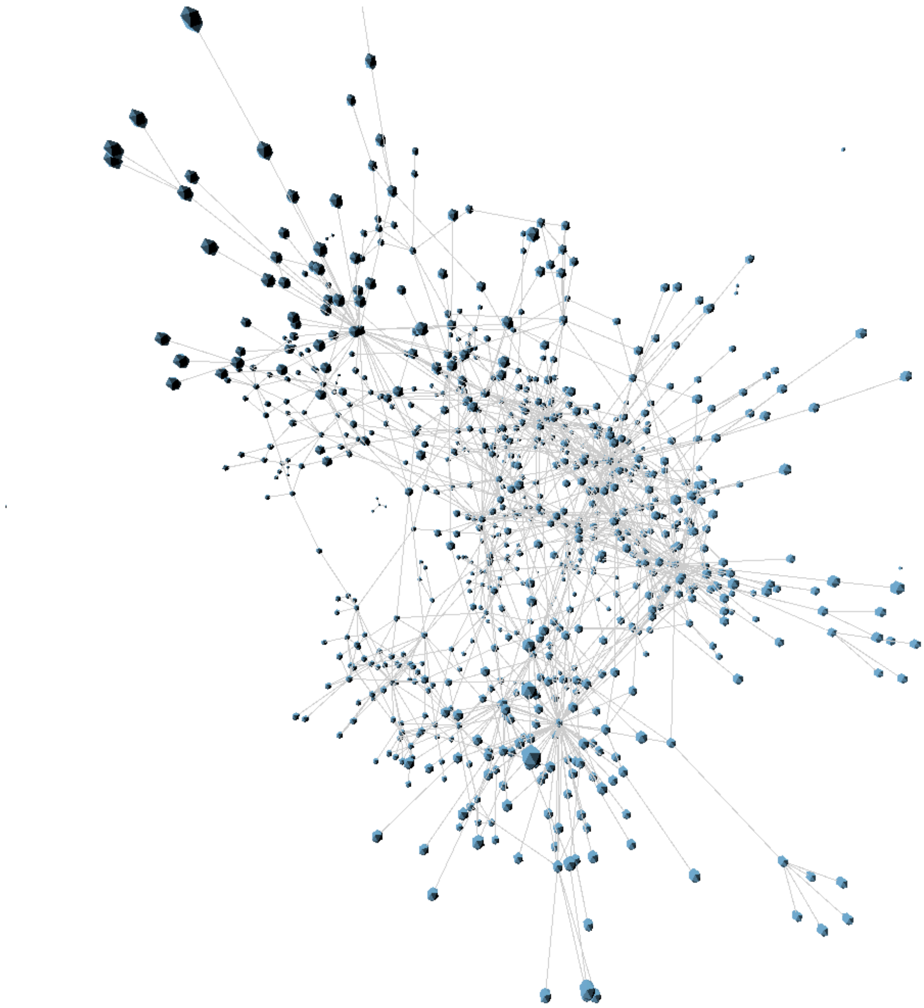


Figure 8-3 Graphical representation of Wiki link graph – the graph represents the articles (nodes) and the links between the articles (edges). The graph shows a Wiki containing ~400 articles

8.1.2.3 Descriptive Statistics of the Data

As stated above, the data for this study have been collected from several Wikis in order to avoid inaccurate results caused by anomalies in one system. However, to obtain comparable results, several factors must be controlled to make sure that changes in the measured system dimensions are not influenced by changes in these factors. These are average user activity, number of users, number of created pages, observed time span, and usage pattern. Table 8-1 shows the characteristics of each Wiki.

#	Users	Articles	Avg. Articles/Day	Snapshots
Wiki 1	50	672	5.6	120
Wiki 2	50	568	4.7	120
Wiki 3	50	831	6.9	120
Wiki 4	50	557	4.6	120
Wiki 5	50	599	4.9	120
Wiki 6	50	854	7.1	120
Total	**300**	**3227**	**5.6**	**720**

Table 8-1 characteristics of the Wiki instantiations – the table shows for each of the Wiki instantiations (1-6) the number of contributing users, the number of articles created in the test period, the average number of articles per day calculated at the end of the test period, and the number of snapshots taken

The number of average articles per day in Table 7-1 implies that usage patterns in Wikis were different. In order to evaluate how the Wikis were used, the development in the number of articles has been plotted in a graph. I assume that the number of articles in a Wiki reflects the usage of the Wiki, meaning that activities like editing, deleting, reading, etc., happen in much the same amount as the creation of articles. Voss (2005) found this interrelation for Wikipedia as well. Figure 8-4 shows the development in the number of pages for each Wiki. The graph shows clearly that the development is not linear and not on the same level for all Wikis but the relative change is similar for all Wikis.

Figure 8-5 illustrates this observation. In this graph, the relative change in the number of articles is compared to the previous day. The values are calculated as follows: for value X at time t, $X._{rel}(t)=X._{abs}(t)-X._{abs}(t-1)$. The graph shows that activities are not equally distributed over time but activity patterns nevertheless are similar for all Wikis.

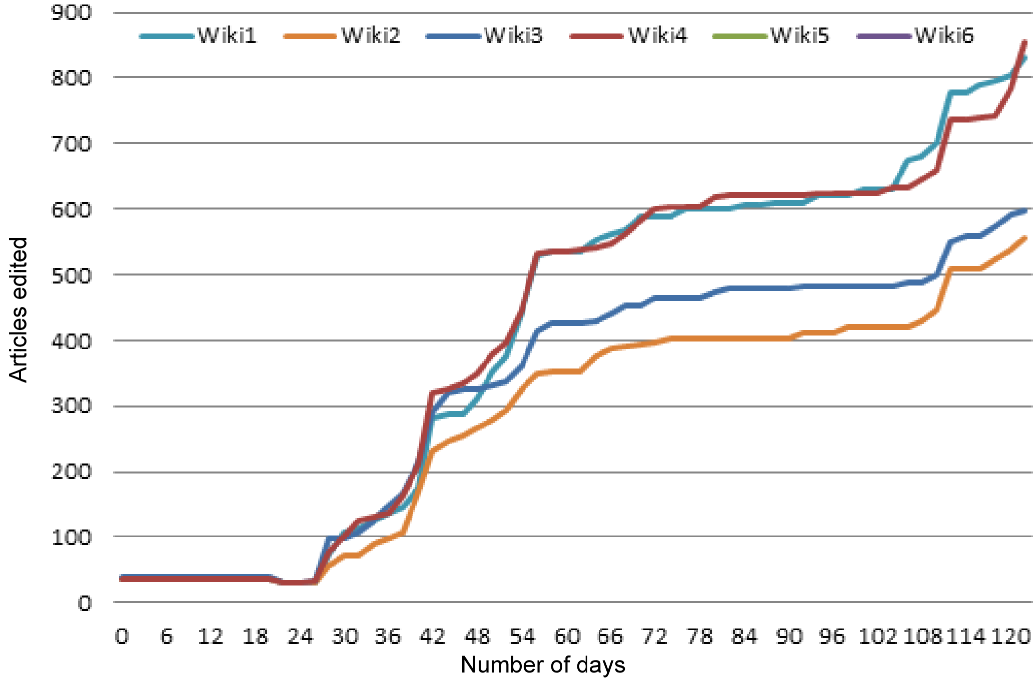


Figure 8-4 Wiki activity graph – absolute number of articles for the test period of 120 days

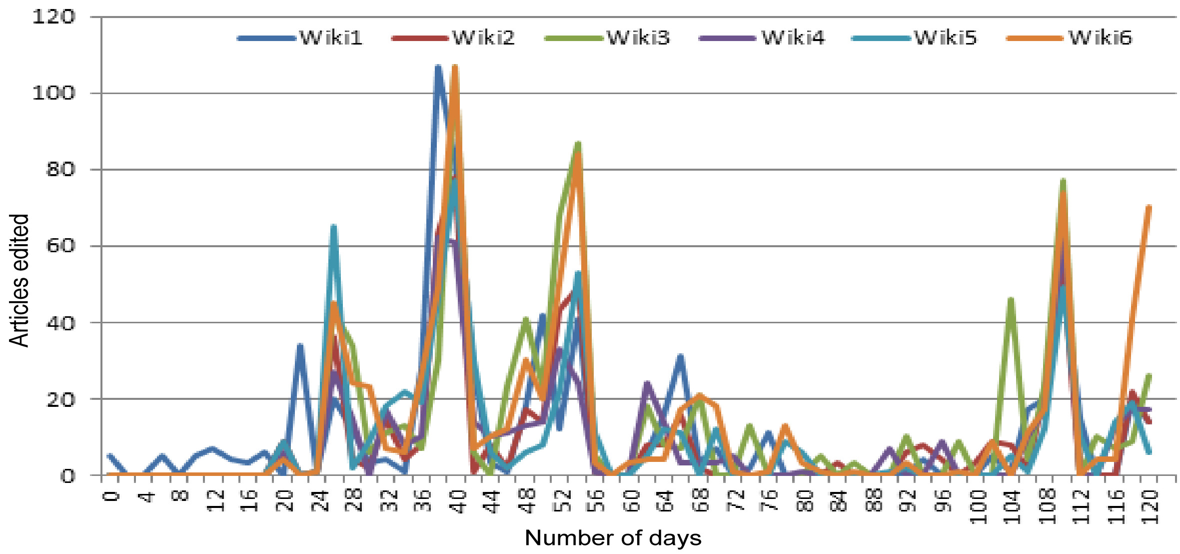


Figure 8-5 Wiki activity graph – relative number of articles

8.1.2.4 Metrics to be analyzed in the study

For each metric category – usage, content, and structure – a set of metrics that ought to be evaluated has been selected. The selection of these metrics is based on previous approaches. Concretely, this means all of the metrics have separately been used to assess the quality of information (see Figure 8-6). Some metrics, which would be just as relevant, were not incorporated due to their predictable correlation with other metrics. It means that some of the metric values are calculated on the same parameters and so a correlation is obvious. Particularly, content-based metrics are often based on parameters like sentence length or number of syllables. If two metrics are influenced by the same parameter and differ only by a constant factor, a correlation is obvious. Since I am considering here only correlations instead of absolute values, these metrics are redundant.

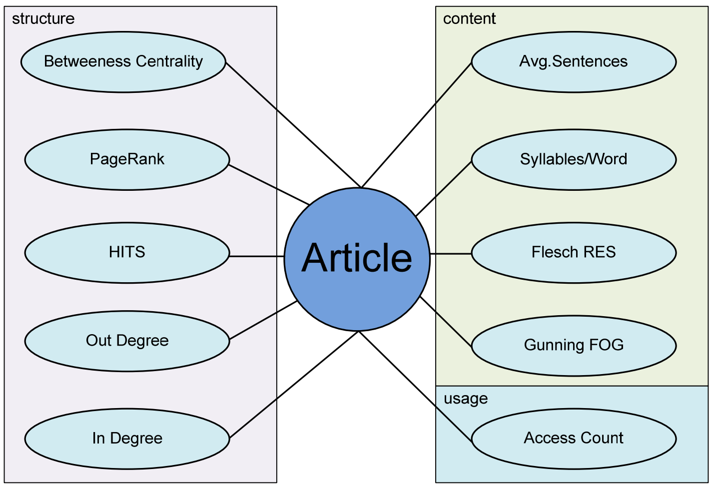


Figure 8-6 Metrics and their system representations used in this study

The selection resulted in 10 metrics with three categories. The parameters on which the metrics are calculated are mathematically independent; hence, the metric values should also be independent from each other. The metrics evaluated in this study are fleshed out in the following.

8.1.2.4.1 *Betweenness Centrality (Brandes, 2001; Leydesdorff, 2007)*

Betweenness Centrality is one of the most important metrics in social network analysis. It calculates for a node the number of shortest paths in a graph that go through this node. In social network analysis, communication between users (nodes) is usually represented in edges. So, Betweenness Centrality is used to determine the importance of a person for communication in a social network. The application in my approach is similar. The difference here is that nodes in my graph are articles while edges are links between articles. Hence, the metric gives information on how many shortest link paths an article is located. This is a very important indicator of how easy it is to find an article in the system. Betweenness Centrality is calculated as follows:

Betweenness centrality of a node $v$is the sum of the fraction of all-pairs shortest paths that pass through v:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where $V$is the set of nodes, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ is the number of those paths passing through some node v other than s, t. If $s = t$, $\sigma(s,t) = 1$, and if $v \in s, t$, $\sigma(s,t|v) = 0$

8.1.2.4.2 PageRank (Page et al., 1999) and *HITS* (Farahat et al., 2006)

Originally, PageRank was intended to rank search results in web searches. Its characteristic is that it simulates the surf behavior of the user group. Therefore, the relevance of a website is defined by its number of incoming links. But not only is this number to be considered but also the relevance of the website the link comes to. The assumption behind this algorithm is that linking to a website implies that this website is relevant since the author points people to this other website. This assumption can easily be transferred to our approach because there are links in social software systems as well, and linking has the same meaning as on the web. Furthermore, the algorithm is applicable here since social software is a subset of the web. PageRank is calculated as follows:

$$PR_i = \frac{1-d}{N} + d \sum_{\forall j \in \{(j,i)\}} \frac{PR_j}{C_j}$$

Where N is the number of pages and d is an attenuation coefficient.

From this idea, the approach of *HITS* is similar to *PageRank* with the difference that the algorithm differentiates between *Hubs* and *Authorities*. Every node has a hub-value and an authority-value. The hub-value is the sum of all authority-values of sites the node links to. The authority-value is the sum of all hub-values of sites that link to the node. The values are calculated as follows:

$$h_i = \delta \sum_{j=1}^{n} A_{ij} a_j$$

$$a_i = \lambda \sum_{k=1}^{n} A^T_{ik} h_k$$

Where a is the authority-value, h is the hub-value, and A is the linkage matrix.

Despite the conceptual similarity to *PageRank*, in this approach *HITS* turned out to provide different measures which, as the subsequent section shows, often differ from the values of *PageRank.* Both algorithms have proven to be useful in the context of quality indicators.

8.1.2.4.3 Flesch Reading Ease Score (FLESCH, 1948) and Gunning Fog (Gunning, 1952)

There is a long tradition in finding indicators for the understandability of texts. The objective of these indicators is to automatically assess the quality of texts in terms of their understandability. Therefore, the indicators should provide information about the educational level for which a given text is appropriate in terms of readability. One can assume that semantics of words plays a crucial role in readability. The reading scores mentioned here only use quantitative measures of the text (sentence length, number of words, and average number of syllables) even though they reach an astonishing correlation to the users' perception of readability. Whether a text is considered good in terms of readability will also depend on the characteristics of the language. Thus, particular reading scores are intended for only one language. For my approach, I used different reading scores for the languages English and German, depending on the target language. Both reading scores mentioned here can only be used with English texts. Flesh Reading Ease Score (FRE) uses a scale from 0 to 100, with 0 equivalent to the 12th grade and 100 equivalent to the 4th grade, and is thus calculated as follows:

$$\mathrm{FRE} = 206{,}835 - (1{,}015 \cdot \mathrm{ASL}) - (84{,}6 \cdot \mathrm{ASW})$$

Where ASL is average sentence length and ASW is the average number of syllables per word.

The calculation of the Gunning Fog score (GFI) is:

$$\mathrm{GFI} = \left(\frac{\mathrm{W}}{\mathrm{S}} + \mathrm{D}\right) \cdot 0{,}4$$

Where W is the number of words in the sample and S is the absolute number of sentences while D is the number of words with more than 3 syllables. These words are considered as *complex words*.

8.1.2.4.4 Access Count (cf. Najjar et al., 2006)

Access count in a social system is how often or the number of times an artifact is retrieved by the user. In the case of Wiki articles, it refers to how often a page is read. On the web as well as in social software systems, the main goal is to make the content of a page popular. Thus, access count is often used to determine how often a page has been opened, e.g. accounting for advertisements on websites. The access count measure used in my approach is the article access count provided by MediaWiki.

8.1.2.4.5 InDegree and OutDegree (Kamps & Koolen, 2008)

The number of incoming and outgoing links on a webpage is an important indicator for both the quality of content and connectedness in the link graph. Even though the number of incoming links and the number of outgoing links is not directly related to the content, it gives indirectly information on the qualitative status of the content. In the case of outDegree, an indicator of quality is linking to other content which means that an article links to further information in other articles. But that may not be a good quality page since, for instance, navigation pages contain lots of links but may provide only little information. What seems clearer is the interconnection between inDegree and quality since people from other pages consciously link to a page, directing people there to find information. Both indicators, inDegree and outDegree, are also the basis for PageRank and HITS.

Another metric that is calculated but not considered here is *Number of Revisions*. Several publications reveal a strong correlation between quality and this particular metric. The reason why I omitted this metric is the manner of selecting articles. The articles are selected based on the number of activities in each article. This is strongly correlated with the *Number of Revisions*. Using this dependent metric would distort the results. The selection of relevant articles is described in the next section.

8.1.3 Data Analysis

8.1.3.1 Selection of relevant articles

Some articles have not been read or edited over the whole test period and therefore there is no value change. Obviously, values that are static correlate with other static values. Further, it is important to have normally distributed values as a prerequisite for the statistical analysis. This is not given for values that do not change. Thus, the set of evaluated articles has been narrowed down so that only articles that had changes in metric values on at least 50 of the 120 days of the testing period have been selected. This resulted in a set of 912 articles for all Wikis. These articles have been ranked according to their number of value changes. The top ranked 100, which represent all other articles in all Wiki instances, have been evaluated within this study

8.1.3.2 Explorative Analysis of Metric Values

The first part of the study is an explorative analysis of metric values. The purpose of this part is to identify patterns and interconnections in the temporal development of metric values. Following a visual analytic approach, I first plotted the timelines of the 50 most edited Wiki articles for each Wiki.

This allowed for comparing the change of values at every point in time. Figure 8-7 shows the metric value development of an article for a period of 20 days. Some of the values stay steady over the test period, some change drastically. Moreover, one can see that some metric values change at the same time. This behavior is observable in almost all article plots. This led me to the assumption, that multiple metrics measure the same aspect of quality. From a more theoretical standpoint, this would mean that they load on one common factor. One example, depicted in Figure 8-7, is the similar pattern in metric values for inDegree (purple) and AccessCount (green).

One reason for correlating values is that they are based on mutual variables or similar factors. Particularly, the example above reveals that the values (inDegree and AccessCount) have no mutual variables. The *inDegree* value represents the amount of articles that link to a certain article. *AccessCount* is the number of users that access (read, edit) an article within a particular time span. Obviously, these metrics have no common components.

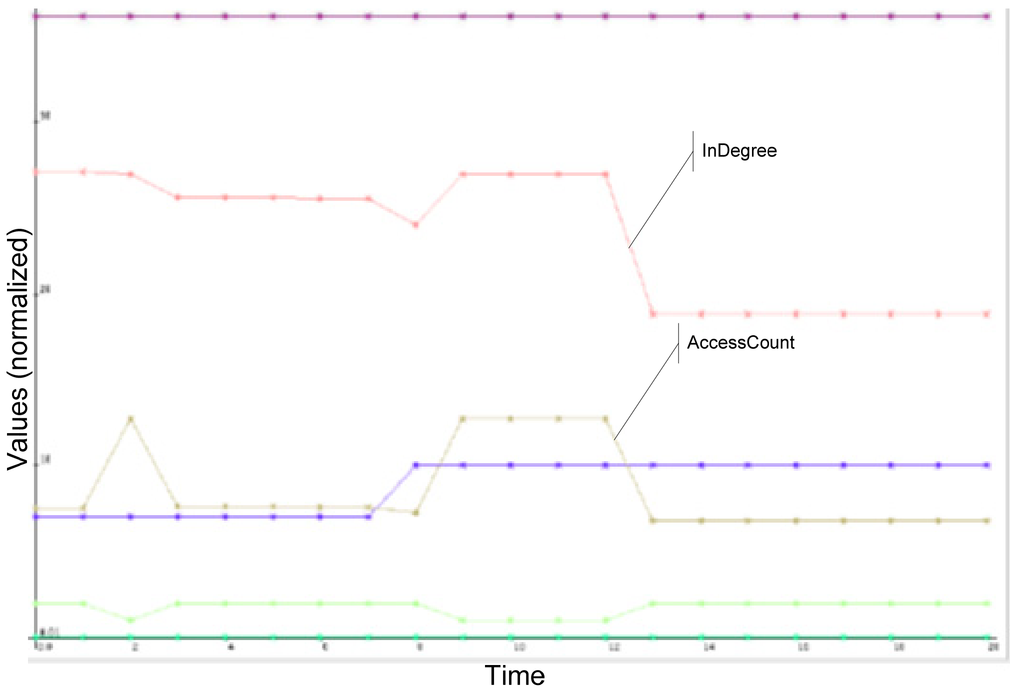


Figure 8-7 Metric development – the plot gives an impression of the temporal development of several metric values. In addition to the tagged metrics, measures for readability (green) and structure (blue) are also shown

In order to analyze the progression of the metric values, the absolute number plays only a minor role. Hence, in this phase, the transitions between the snapshots have been calculated as relative difference between the current value and the last value. It means that if there is no change in the values, the relative difference is always independent of the absolute value and the used metric. Figure 8-8 shows the relative difference of Gunning Fog (readability) and Betweenness Centrality (reachability) metric values. Some of the peak values are shifted to the left or to the right. I assume that this is due to the time that passes by before the consequential effect happens. For instance, improving the readability of an article will not instantly lead to a value change in the access count; this effect may take a while to manifest itself in the data. Figure 8-8 and Figure 8-9 show the metric development of the articles *Semantic Web* and *Web 2.0*, respectively. The difference here to the previous plot is that the graphs show the relative change in values with regard to the previous point in time. Both graphs strongly suggest a correlation. Figure 8-8 shows the metric values for *Gunning Fog,* a readability indicator created from the content representation; the other one is the structure representation-based metric *Betweenness Centrality*. Figure 8-9 shows the *Flesch Reading Ease Score* and *Betweenness Centrality*. The difference here is that the correlation seems to be negative.

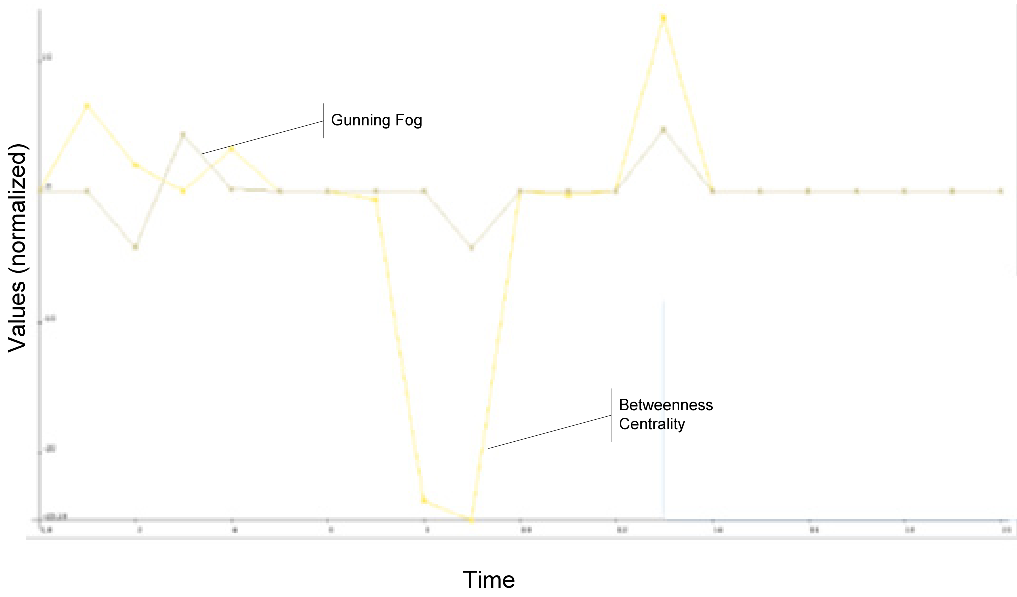


Figure 8-8 Metric development for the article *Semantic Web* over a period of 20 days

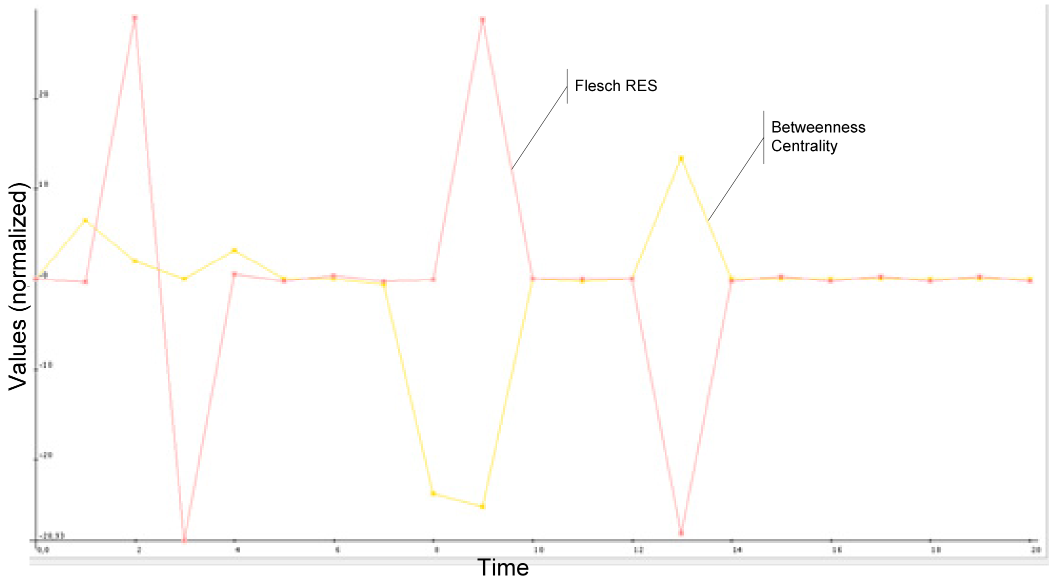


Figure 8-9 Metric development for the article *Web 2.0* over a period of 20 days

This visual presentation of metric values suggests a correlation between them. This substantiates the assumption that several metrics are together capable of predicting a particular aspect of quality.

8.1.3.3 Factor Analysis

As a next step, based on the observation that some metrics seem to be associated among them, a factor analysis was carried out on the data. The aim of the factor analysis was to strengthen the hypothesis that some metrics are correlated due to underlying quality characteristics. That means, metrics should be identified that measure a similar aspect of quality. It could also mean obtaining 'bundles' of metrics for measuring a particular quality aspect more reliably than with other (separate) metrics.

For the factor analysis, I used a *varimax* factor rotation. Table 8-3 shows the results of the factor analysis. The result shows that in all three cases, content metrics (reading scores) and structure metrics (HITS) load on a common factor.

Similarly, sage (access count) and structure (centrality) metrics load on a common factor as well. These results will be verified in the correlation study.

	Factor1 (Readability)	Factor2 (Objectivity, Accuracy)	Factor3 (Accessibility)
Access.Count	0.460	0.805	
Centrality	0.110	0.852	0.426
FleschGL	0.924	0.377	
FleschRES	-0.937	-0.334	
FOG	-0.889	-0.448	
inDegree	0.540	0.811	
HITS	-0.802	-0.135	0.380
outDegree	0.387	0.853	0.128
PageRank.1000	-0.175	-0.760	
Syllables.per.Word	0.950	0.252	0.174
Words.per.Sentence	0.892	0.447	

	Factor1	Factor2	Factor3
SS loadings	5.529	3.526	0.971
Proportion	0.503	0.321	0.088
Cumulative	0.503	0.823	0.911

Table 8-2 Results of factor analysis, green fields show values higher than 0.6 (self-defined threshold)

The factor analysis revealed 3 independent factors that shall be looked at more closely now. Factor 1 in all three groups has its main focus on content quality. This factor mainly concerns reading scores and text-based metrics (Flesch Grade Level, Flesch Reading Ease Score, Gunning Fog) but also a structure based metric (HITS) in all three cases. This can be interpreted along the sense that *text quality* *can be measured using reading ease scores but the result can be verified when the structure based metric HITS is added to the calculation.* The second factor somewhat focuses on the structure-related quality dimensions of *Objectivity, Accuracy* and *Relevancy.* The decisive metrics here are mainly based on the link structure. The third factor has not really a relevant metric but *Betweenness Centrality* has medium high values for all Wiki systems. Testing the real meaning of the identified factors and their relation to the concept of quality goes beyond the scope of this work.

8.1.3.4 Correlation Matrix

The correlation study is aimed at testing the hypothesis derived from the visual metric analysis and the factor analysis. In a Wiki-based dataset, correlating metrics in different system representations may be found. Thus, this part of the evaluation focuses on the following questions. Are user interactions in the system observable in different representations? More concretely, is there an interrelation between structure/usage and the quality of content? This could help to answer future questions like: do people interact differently on high quality pages in relation to low quality pages? Is the number of people that read an article interconnected to content quality and/or the centrality in the link graph?

The aim of correlating metric values is to answer these questions. Therefore, from a set of articles the values of metrics (for a particular time-span) are transformed into vectors. The vectors for two metrics (of one article) are then correlated. According to the results of the visual analysis and the factor analysis, each of the three graph-based metrics – *Betweenness Centrality, Page Rank* and *KStep* - was evaluated in terms of correlation with each of the other metrics.

In this part of the study, a Pearson correlation (Bobko, 2001; Wasson, 2008) was calculated. The dataset consisted of merged data from all 6 Wikis. Subsequently, the metric values of one article for all Wikis and for all points in time were considered for calculation. The Pearson correlation coefficient results in values between 1 and -1; with 1 meaning absolute correlation and -1 absolute inverse correlation. 0 means no correlation at all.

According to Cohen (1988), I assumed the following effect-sizes:

Size of effect	ρ	% variance
small	.1	1
medium	.3	9
large	.5	25

Table 8-2 shows an excerpt from the result of the correlation test. The excerpt consists of the set of articles most frequently edited or accessed. The coloring serves only for making the results more readable and easier to interpret (green = strong correlation; white = weak correlation).

	F/K RES Betweenness Centrality	Gunning FOG Betweenness Centrality	AccessCount Betweenness Centrality	F/K RES PageRank	FOG PageRank	AccessCount PageRank	F/K RES HITS	FOG HITS	AccessCount HITS
Social_Web	0,60	0,50	0,63	0,49	0,65	0,55	0,21	0,23	-0,04
Hypertext_Markup_Language	0,55	0,48	0,72	0,52	0,60	0,11	0,24	0,36	-0,20
Technorati	0,53	0,49	0,63	0,67	0,75	0,61	0,78	0,85	0,79
Uniform_Resource_Identifier	0,68	0,48	0,63	0,16	0,46	0,14	-0,01	0,43	0,48
Social_Software	0,60	0,51	0,75	0,71	0,77	0,45	0,81	0,83	0,75
Cloud_Computing	0,65	0,59	0,78	0,67	0,72	0,49	0,83	0,84	0,76
Java	0,56	0,52	0,60	0,63	0,63	0,55	0,50	0,50	0,43
Folksonomy	0,44	0,62	0,71	0,41	0,53	0,55	0,40	0,76	0,82
Web_application	0,55	0,51	0,60	0,78	0,79	0,80	0,89	0,88	0,96
Flickr	-0,27	-0,33	0,76	0,85	0,88	0,51	0,59	0,61	0,78
Markup_Language	0,60	0,51	0,62	0,40	0,45	0,11	0,56	0,66	0,38
Linked_Data	0,65	0,55	0,60	0,50	0,60	0,55	0,13	0,16	0,16
Wikipedia	-0,02	0,70	0,72	0,90	0,73	0,71	0,85	0,68	0,73
Instant_Messaging	-0,34	0,41	0,65	0,68	0,88	0,68	0,04	0,76	0,84
Web_Ontology_Language	0,50	0,53	0,62	0,33	0,19	-0,05	0,23	0,12	-0,13
Dynamic_Web_Pages	0,73	0,72	0,81	0,66	0,67	0,55	0,45	0,45	0,39
Information_Retrieval	-0,44	0,22	0,88	-0,27	0,47	0,77	-0,06	0,60	0,73
Social_network	0,74	0,53	0,69	0,58	0,74	0,83	0,79	0,86	0,94
Web_2,0_Technology	0,85	0,67	0,75	0,65	0,88	0,81	0,78	0,91	0,90
Tim_Berners_Lee	0,29	0,62	0,78	0,62	0,43	0,07	0,80	0,79	0,84
Vimeo	0,56	0,87	0,94	0,92	0,75	0,73	0,77	0,86	0,90
Semantic_Web_Applications	0,33	0,56	0,65	0,32	0,29	0,12	0,13	0,03	-0,23
World_Wide_Web_Consortium	0,07	0,53	0,63	0,74	0,58	0,66	0,58	0,69	0,81
BibSonomy	0,37	0,58	0,71	0,44	0,56	0,28	0,53	0,70	0,83
GeoNames	0,63	0,64	0,65	0,62	0,57	0,57	0,22	0,18	0,18
Social_Tagging_Systems	0,65	0,69	0,75	0,78	0,74	0,72	0,85	0,84	0,86
Extensible_Markup_Language	0,14	0,55	0,67	0,62	0,49	0,25	0,78	0,35	0,04
Internet	0,71	0,51	0,76	0,70	0,88	0,64	0,90	0,97	0,89
Social_Networking_Platform	0,37	0,50	0,63	0,88	0,81	0,56	0,91	0,87	0,89
Data_Mining	0,42	0,49	0,74	0,94	0,95	0,85	0,82	0,83	0,86
Query_Language	-0,19	0,63	0,74	0,82	0,49	0,21	0,21	0,67	0,73

Figure 8-10 Table of correlating metrics in Wiki articles – the columns show in each case the correlation of two measures, e.g. Flesch Kinkaid RES (F/K RES) and Betweenness Centrality. The rows show the correlation for a specific article in all value pairs of all Wiki instantiations at all points in time.

This result shows that some of the metric combinations have a strong correlation for many articles. This indicates that there can be a dependency between the metric values which implies that the quality attributes influence each other. Particularly noticeable are the numerous strong correlations between *Access Count* and *Betweenness Centrality*. *Access Count* is a usage-based metric and denotes the number of edits and views of an article within a certain time-span. *Betweenness Centrality* is a graph-based metric that is often used in social network analysis. It denotes the number of shortest paths on which the node is located. In the link structure graph, this means that the article is often traversed while browsing through the Wiki, since it is located on the paths that are most often used. The reason for the correlation could be that an article is accessed more often when it is easier to find. Or the other way around, the more people read and edit an article, the more frequent they link to this article from other articles, what results in a higher *Betweenness Centrality*. As an example, this correlation could be used to increase the number of page accesses by improving *Betweenness Centrality,* for instance, by creating links to the article. Another correlation that is noticeable is between *Page Rank* and *Gunning Fog*. Since *Page Rank* is a graph-based metric which is often used for ranking search results and *Gunning Fog* is a readability score, it is obvious that these metrics are calculated based on different information. Still, their correlation suggests that both of them are influenced by the article quality.

However, the Pearson correlation provides information only about similar tendencies in value development. It makes no statement about causal dependencies. To infer that, for instance, better readability is a result of a high access count is not feasible. Hence, one can only show that user interactions have implications for all system representations.

8.1.4 Conclusion

The aim of this study was to answer the question of whether changes in the social software system can be observed in several system representations. The visual data analysis, facilitated by using plots of the metric value timelines, led me to the discovery that there seems to be a correlation between some of the metrics in different system representations. This observation is interesting because metrics in different representations are mathematically independent so that a correlation must be caused by measuring similar factors in the system. In order to strengthen this hypothesis, I further conducted a factor analysis. The result of the factor analysis strengthened the assumption about correlating measures. In a correlation study, I tested the metrics which indicated a correlation behavior in the former tests. The result shows that there is a correlation in some of the metrics. This result does not allow for concluding that there is a cause–effect relation. Thus, finding out which activity causes change in the value ought to be part and parcel of any further research.

8.2 Evaluation of task-specific quality profiles

Parts of this section have been published in

Weber, N., & Lindstaedt, S. N. (2011). A User Centred Approach for Quality Assessment in Social Systems. *Proceedings of the International Conference on Knowledge Management and Information Sharing 2011,* Paris, France

The evaluation of metric values in the previous section had indicated that there seem to be several facets of quality that can be measured using the metrics as quality indicators. The underlying assumption for the next study is that these quality facets differ in their relevance depending on the common understanding of task requirements in a community.

8.2.1 Objective of the Study

Section **2.1** describes the influence of various factors on the user's perception of quality. The common understanding of quality in communities is considered here as a decisive factor. But also within communities, quality of a resource can be perceived differently in various contexts. The knowledge of which quality is required for artifacts involved in a task is also part of the common understanding of quality. The objective of this study is to evaluate the relevance of different quality dimensions for tasks achieved in a community. The result plays an important role in the recommendation of useful resources based on the user's current task.

8.2.2 Study Context

This study was conducted in the course of the MATURE demonstrator evaluation. The objective of this part of the project was to support quality assurance in the career guidance sector (Weber et al., 2009). The participants in the study are domain experts from career guidance working in different areas and on different management levels. This facilitates on the one side capturing the current status from different perspectives yet ensures on the other remaining within the selected domain. All participants are working in the same domain (career guidance), thus, they can be considered as one community with similar interests and goals.

Within this study, the users tested a Wiki-based information system which was designed and developed for assuring quality in the context of career guidance. A more detailed description of the system can be found in Section 7.

8.2.3 APPROACH

The users performed several tasks in order to test the system in different situations. These tasks were selected according to familiar work tasks of the end-users. So, the first phase of the study was gathering relevant tasks from the end-users by interviewing them. The tasks were divided into two task groups: the first group was about acquiring information, e.g. by searching. The second group was about providing information as in writing articles. The following tasks resulted from this phase:

1. TASK 1 - Searching for information around higher education.
   Task description: The objective of this task is to provide a broad overview of a particular area in the field of higher education. This task is part of the work of a personal adviser (PA) in career guidance, and relates to activities like finding general information on higher education, the subjects that can be studied, where the particular subjects can be studied, things to consider when deciding when and where to go, and what graduates do. The goal is to find relevant information for the different steps a PA has to consider when giving information, advice and guidance to a young person by taking into account their post-compulsory education options.
   In considering this task, the test users were asked to rate the resources and information they found during their search with regard to the different quality aspects listed below (cf. Figure 7 11).

2. TASK 2: Providing and scaffolding information on higher education.
   Task description: The goal is to make sense of information on higher education and create a summary of information for young people.
   Again, the users were asked to rate the summary they have created for a young person with regard to the different quality aspects listed below (cf. Figure 8-11).

3. TASK 3: Information search for studying medicine at higher education level.
   This task comprises focused activities for finding information on choosing and studying medicine at a higher educational level. The objective of this task is to provide detailed information on a specific topic.
   Here, the users were asked to rate the importance of the required information with regard to the different quality aspects listed below (cf. Figure 8-11).

4. TASK 4: Providing information for studying medicine at higher education level (detailed information).
   This task is about providing information over a particular area on a specific topic. Information must be very detailed and focused.
   Again, the users should rate the content they have created with regard to the different quality aspects listed below (cf. Figure 8-11).

In the second phase, the users assessed the relevance of quality dimensions for each of the tasks. The quality dimensions proposed to the users were taken from Wang and Strong (1996). For each task and quality dimension, the users could select from a 5-point Likert scale beginning from *not relevant* to *vital*. In order to have a common understanding of the meaning of quality dimensions, these have been briefly explained at the bottom of the questionnaire.

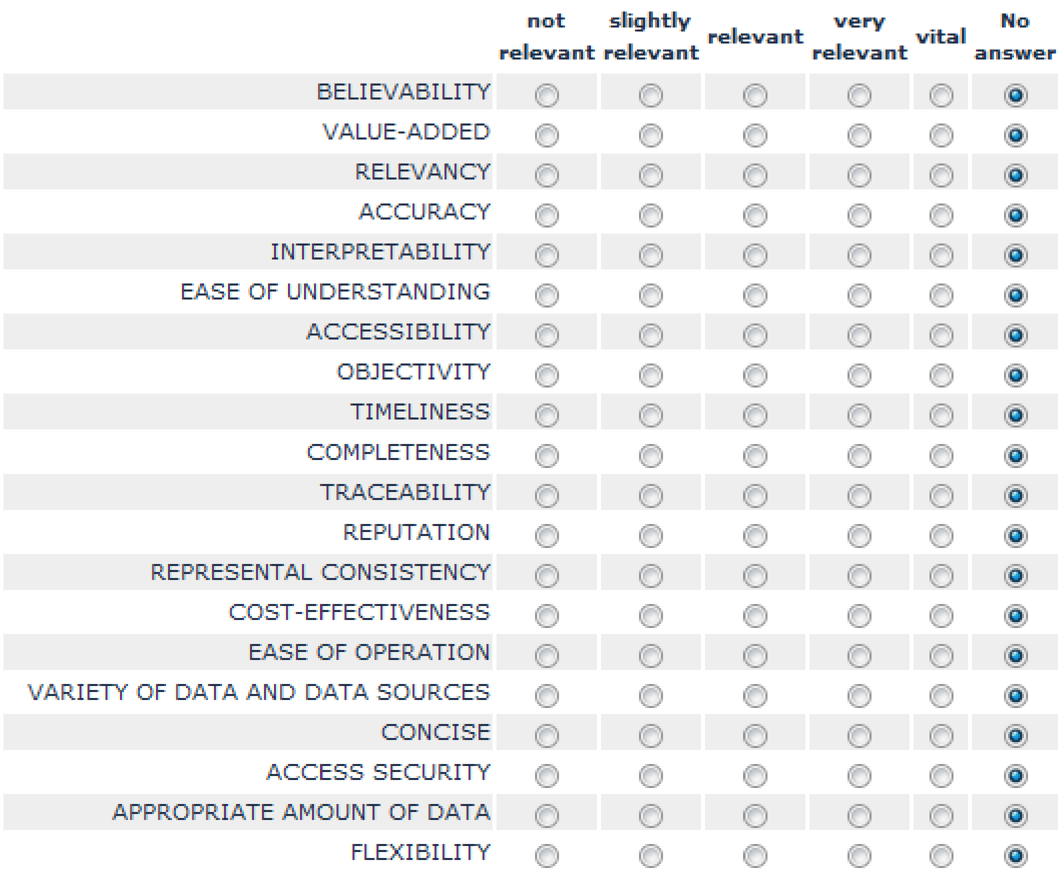


Figure 8-11 Questionnaire example – participants had to assess each of the quality dimensions (rows) on a five-point Likert scale (columns); additionally, 'no answer' could be chosen

The quality dimensions were explained as follows:

- BELIEVABILITY: The extent to which data are accepted or regarded as true, real, and credible.
- VALUE-ADDED: The extent to which data are beneficial and provide advantages from their use.
- RELEVANCY: The extent to which data are applicable and helpful for the task at hand.
- ACCURACY: The extent to which data are correct, reliable, and certified free of error.
- INTERPRETABILITY: The extent to which data are rendered in appropriate language and units and the data definitions are clear.
- EASE OF UNDERSTANDING: The extent to which data are clear without ambiguity and easily comprehended.
- ACCESSIBILITY: The extent to which data are available or easily retrievable.
- OBJECTIVITY: The extent to which data are unbiased (unprejudiced) and impartial.
- TIMELINESS: The extent to which the age of the data is appropriate for the task at hand.
- COMPLETENESS: The extent to which data are of sufficient breadth, depth, and scope for the task at hand.
- TRACEABILITY: The extent to which data are well documented, verifiable, and easily attributed to a source.
- REPUTATION: The extent to which data are trusted or highly regarded in terms of their source or content.
- REPRESENTATIONAL CONSISTENCY: The extent to which data are always presented in the same format and are compatible with previous data.
- COST-EFFECTIVENESS: The extent to which the cost of collecting appropriate data is reasonable.
- EASE OF OPERATION: The extent to which data are easily managed and manipulated (i.e., updated, moved, aggregated, reproduced, customized).
- VARIETY OF DATA AND DATA SOURCES: The extent to which data are available from several different data sources.
- CONCISE: The extent to which data are compactly represented without being overwhelming (i.e., brief in presentation, yet complete and to the point).
- ACCESS SECURITY: The extent to which access to data can be restricted and hence kept secure.
- APPROPRIATE AMOUNT OF DATA: The extent to which the quantity or volume of available data is appropriate.
- FLEXIBILITY: The extent to which data are expandable, adaptable, and easily applied to other needs.

8.2.4 *RESULTS*

The questionnaire was filled in by 4 experts acting as proxy for a particular domain. The experts had background knowledge in the area and the required overview to assess the relevance of the quality dimensions. Experts from four domains assessed the relevance of quality dimension representing 24 personal advisors.

Due to the small number of participants, a first question was whether the participants assess the importance of dimensions in a similar manner. Based on my assumption that participants from one community have a common understanding of quality, the rating of the users must be similar. In order to test this assumption, I calculated the correlation of the answers from each user – user pair. Since the dataset fulfills all preconditions, I then calculated a Pearson Correlation.

	u1	u2	u3	u4
u1	1	0,302	0,502	0,568
u2		1	0,192	0,443
u3			1	0,703
u4				1

Tabelle 8-1 correlation matrix – the matrix shows the Pearson correlation for the assessments of four users (u1 – u4)

The results show that there is a strong correlation between users 3 and 4. This correlation shows that the ratings of these users are almost identical. The combinations u1-u2, u1-u3, u1-u4 and u2-u4 still have a medium correlation. A low correlation can be observed with users u2 and u3. Considering these results, similarity in the rating of all participants can be assumed. In addition, this result also reveals that there is only a small standard deviation from the ratings. Based on this observation, further calculations were executed using the mean value of all participants.

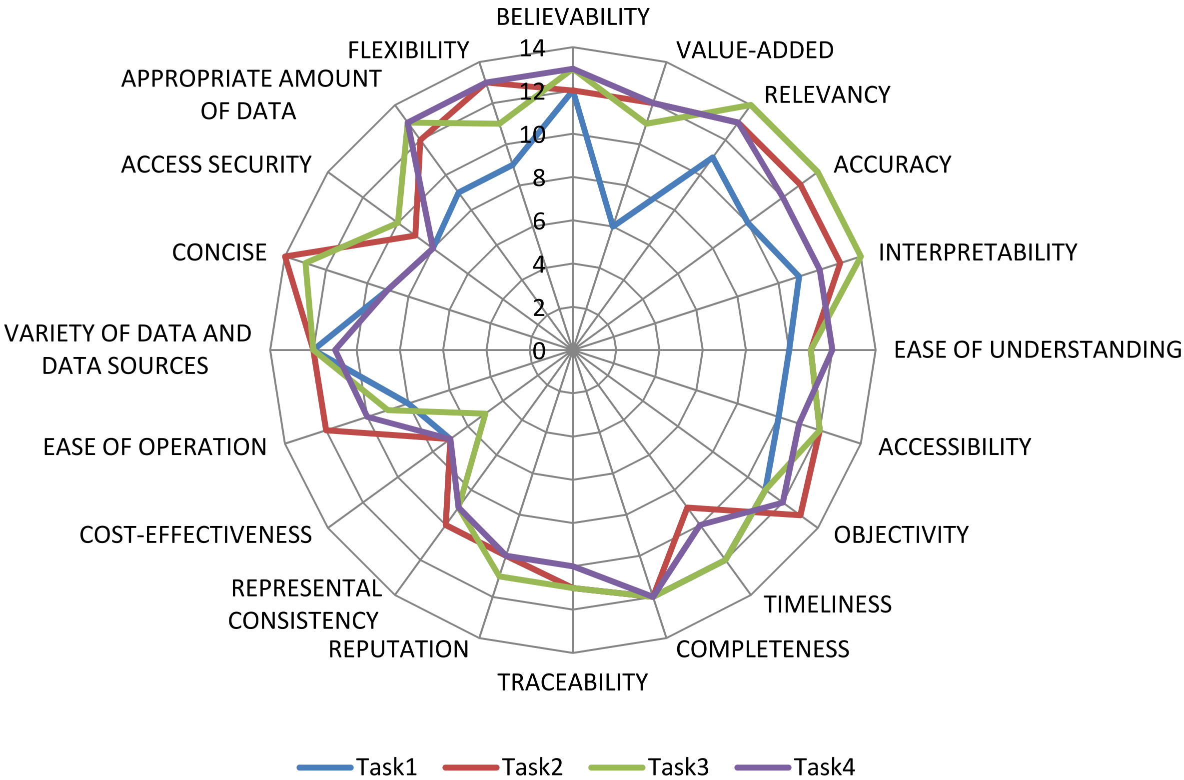


Figure 8-12 Quality profiles for tasks 1 to 4 – the graph shows the mean values of all users rating quality dimensions for each task

The summary of the results shows that the relevance of quality dimensions is weighted differently for various tasks. Figure 8-12 shows the accumulated results for the four tasks. One noticeable fact is that some of the quality dimensions are more dependent on the task than others which are similar for all tasks. Two examples, shown in Figure 8-12, are *Completeness* (which seems to be important, independent of the task) and *Cost Effectiveness* (which seems to be rather marginally relevant for the selected domain. In contrast, some values are obviously dependant on the task. In the case of *Concise*, the relevance for the second and third task is high, whereas it is low for the first and fourth task. The dimension *Timeliness* is assessed higher for tasks 1 and 3 than for tasks 2 and 4.

Regarding the fact that tasks 1 and 3 address the quality of resources that are presented to the user and tasks 2 and 4 are where the user provides information by himself/herself, the different weights of the quality dimensions make sense.

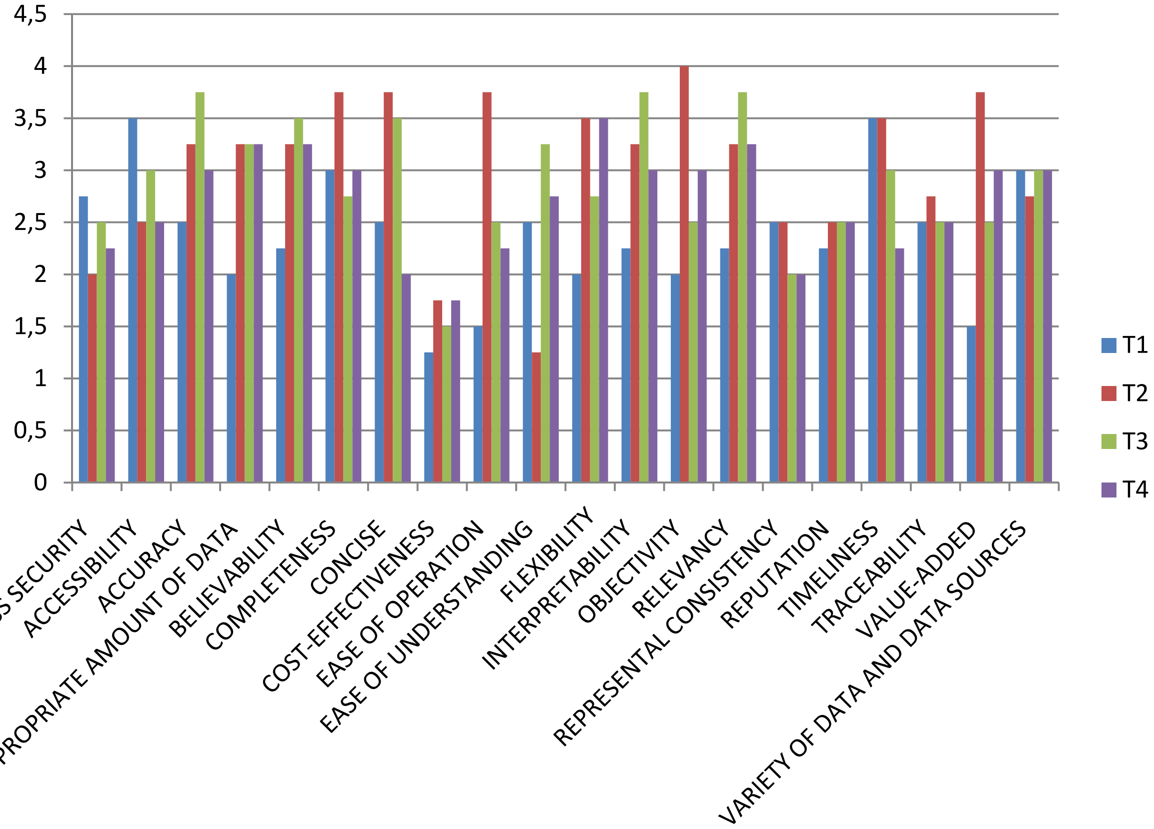


Figure 8-13 mean values for each task (T1-T4) and each quality dimension

In order to verify this result and to evaluate the discrimination of tasks, I calculated the standard deviation of the ratings for each quality dimension. Therefore, the mean value of all task ratings was first calculated for each quality dimensions. As a result, I got the assessment of quality independent of the task. Figure 8-13 shows the mean values for each quality dimension. Even though these values are not very meaningful because they do not reflect the deviation of the individual tasks, they show that some dimensions are generally not very important. For example, *Cost Effectiveness* is rated as not relevant independent from the task. This result has also been confirmed in interviews with the participants. On the other hand, other dimensions have been assessed as very relevant. In my opinion, these are the dimensions that one would intuitively choose as very relevant, viz.: *Accuracy, Completeness, Relevancy* and *Timeliness.*

As already mentioned, the mean values are only meaningful if the deviations from the mean values are considered as well. The values in Figure 8-14 show the standard deviation for each quality dimension. The higher the value, the more do the individual task ratings differ from the mean value. This could be interpreted as the following: the higher the standard deviation, the more diverge are the ratings for the tasks and thus the more task-sensitive is the quality dimension. As an example, *Reputation* has a low standard deviation. This means that the ratings for the tasks are almost identical to what can also be observed in Figure 8-14. Other task-independent dimensions are: *Representational Consistency, Traceability* and *Variety of Data Sources.* On the other hand, there are some dimensions that seem to be highly task-dependent such as *Ease of Operation, Ease of Understanding* and *Objectivity.*

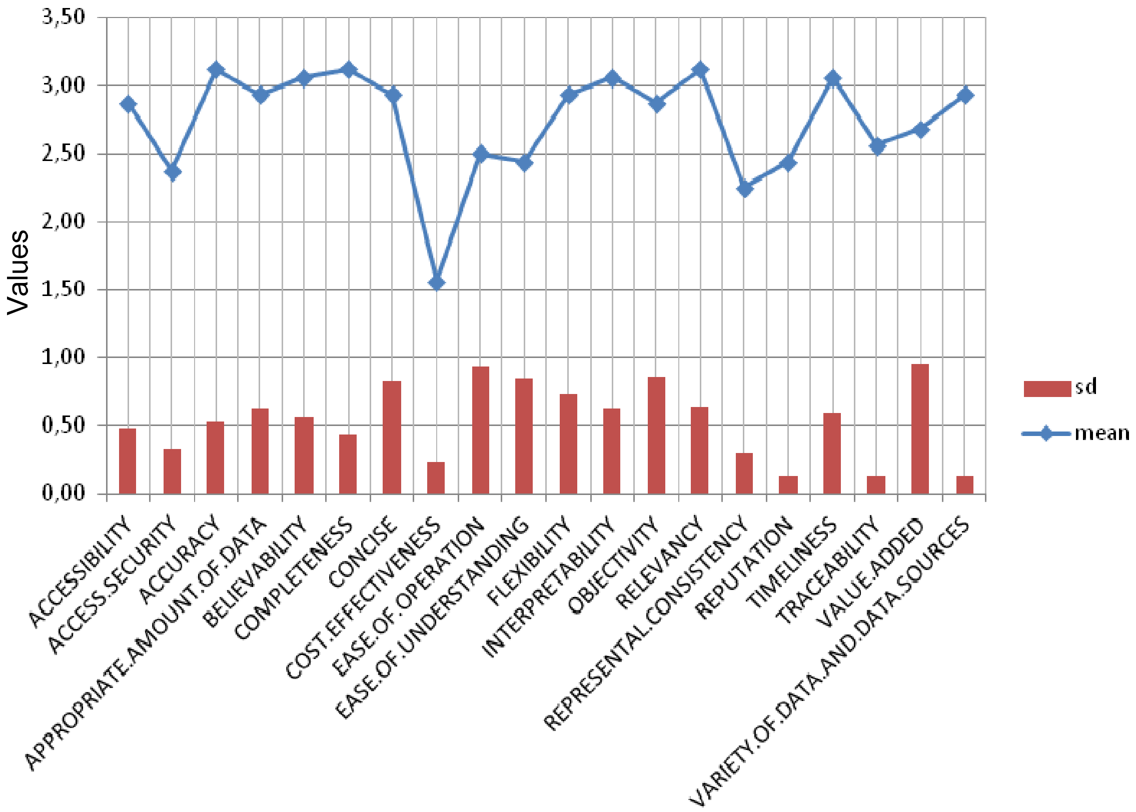


Figure 8-14 shows the mean value of all tasks together for each of the quality dimensions (red bars); the blue line shows the standard deviation of the four tasks on each of the quality dimensions. A high standard deviation means that the ratings on one quality dimension was similar for all tasks

The result of the study shows that the requirements for resources are not always the same. The requirements are mainly dependent on the user's current task, regardless of the provision of content and the consumption of information. Both are important for tracking the qualitative status of the resources and for knowing what the user wants to do with the resources.

8.2.5 Conclusion

We assume that perception of quality depends on the user's context. Amongst others, the user's current task is an important part of the context. This study aims to show the dependency between quality perception and user tasks. Therefore, four characteristic tasks of participants were rated regarding their quality requirements. The results show that some quality dimensions highly depend on the task, while others are independent of it. The results of the study have influenced this work in such a way that user profiles are subdivided into user-task profiles, representing the user requirements for each of the user tasks.

8.3 EVALUATION OF PROFILE SIMILARITY ALGORITHM

Parts of this section have been published in

Supporting the knowledge maturing process in organizations using Web 2.0
(Unterstützung des Wissensreifungsprozesses durch Einsatz von Web 2.0 in Unternehmen)
Weber, Nicolas, Frühstück, G. und Ley, Tobias
In Proceedings: 6th Conference on Professional Knowledge Management - From Knowledge to Action. 21. - 23. February 2011, Innsbruck, Austria, ISBN 978-3-88579-276-5 , pp. 231 - 240 (2011 /02)

The previous study indicated that the relevance of quality facets can be perceived differently depending on the common understanding of quality in a community. One goal of my approach is to find methods for providing resources according to the quality requirements of a community. Therefore, the quality profile of a community member has to be compared to the qualitative profile of resources. This section presents the evaluation of an algorithm for comparing these quality profiles.

8.3.1 INTRODUCTION

The study is aimed at evaluating the performance of the pattern similarity measuring approach. In particular, how accurately do the recommended artifacts correspond to the users' quality requirements? Therefore, the explicitly given quality dimension weighting is compared to the user model created by rating articles. The result shows that the pattern clustering algorithm would have recommended the same articles that are considered *good* by the user.

Although providing recommendation mechanisms is not within the scope of my approach, calculating the similarity between quality profiles is one prerequisite for providing quality-aware functionalities (cf. Section 2.1.2). The performance of the profile similarity algorithm is crucial for the accuracy of any quality-aware system behavior. Therefore, a recommendation situation is simulated in order to be able to evaluate the performance of the profile similarity algorithm.

The modified pCluster approach, described in Section 6, is intended to provide recommendations for useful artifacts. Useful, in this context, means that the artifacts correspond to the users' quality requirements. In this study, the user rating is compared to the results of the clustering algorithms. The focus is on the question of whether the user would rate as good an article which is recommended by the system.

8.3.2 STUDY CONTEXT

The study is conducted in an organizational Wiki operated by a software company. The small-sized software company (~25 employees) develops ERP applications for different purposes and operates in several locations across Europe. The software company aims to improve the quality of the Wiki content and expects to get information about the quality requirements of its customers and employees.

The Wiki acts as a common knowledge base for horticultural ERP software. The Wiki is thus used for different purposes; as documentation tool in the software development and planning process, as internal knowledge base for employees, but also as software handbook and support portal for customers. The system currently contains about 2,800 articles in different categories. The average number of page views per month is 1,750.

Based on the observation of the system usage, four different tasks were identified: finding information about a certain topic, getting support for a problem, writing an article (internal), and writing an article (handbook). These are also the tasks which could be selected in the questionnaire (Figure 8-16). For every task, the user could determine a quality dimension rating. The quality dimensions rated in the questionnaire were taken from Wang and Strong (1996). The results presented in this study have been gathered over a period of 3 months during which both extensions were used in the system.

8.3.3 Approach

The clustering algorithm is able to evaluate the similarity between two given patterns. In case of quality recommendation, these patterns are either resource profiles, representing the current quality status of a resource, or user profiles, containing information about the quality requirements of a user. Clustering these profiles means to identify patterns that are similar to a given one. Similar resource profiles mean that the resources are similar in terms of their quality dimensions. A user profile that is similar to a resource profile means that the resource complies with the users' quality needs and can be recommended to the user. Two similar user profiles imply similar requirements in particular quality dimensions. Here, it is important to mention that patterns only represent information about quality dimensions and do not consider the topic or domain of a resource. Thus, recommendations can meet the user's needs regarding quality but may be useless due to the wrong content domain or topic.

This approach compares two user-given profiles. One is explicitly given by filling in a questionnaire. The second is implicitly given by rating articles in a Wiki.

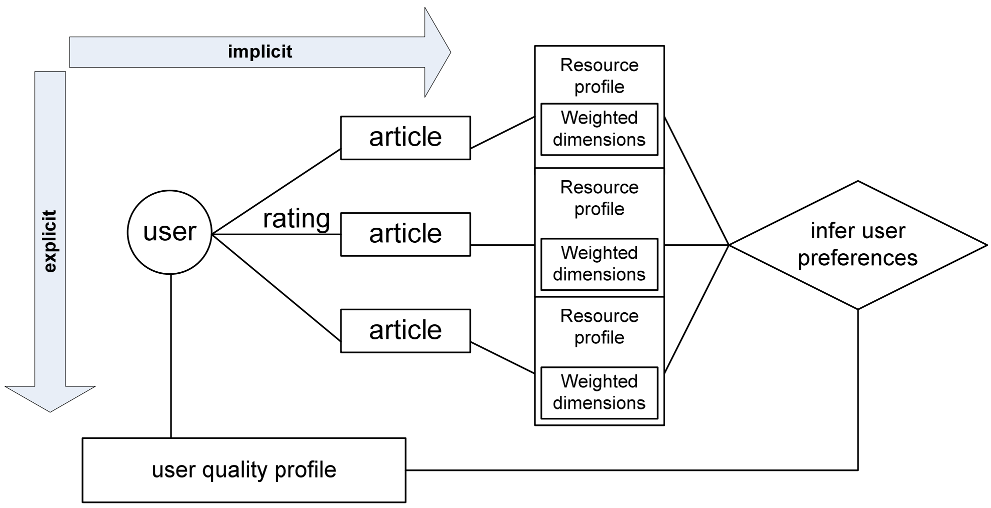


Figure 8-15 Explicit and implicit quality profile creation – the figure shows the two ways (implicit and explicit) of deriving the user quality profile. In the explicit way, the user directly states his/her preferences; in the implicit way, the user rates articles and from the features of the articles the user preferences are directly inferred

Figure 8-15 shows both processes for establishing quality profiles. In the explicit way, the user rated each quality dimension according to his/her own weighting. The implicit way uses article ratings for elicitation of the user's weighting. Based on the qualitative status of the rated resource, one can infer the quality recommendations of the user. If, for instance, the user rates articles in which the weighted dimension for *ease of understanding* is high, then one can infer that the user prefers articles with relative *ease of understanding*. So, all articles rated by one user are analyzed to identify similarities in the weighted quality dimensions. The result is a user quality profile implicitly created by rating. Assuming that the user is able to fairly express her quality requirements in the questionnaire, both the implicit and explicit user profile should be in one cluster. The aim of this study is to evaluate the accuracy with which the cluster algorithm clusters the profiles. Another approach might be to test if the clusters also contain documents that have been rated as good. The problem here is that some articles are rated as good due to reasons other than quality, for instance, the topic.

The functionalities for facilitating implicit and explicit rating can be integrated directly into a Wiki. Explicit rating takes place by means of a Wiki-integrated questionnaire (Figure 8-16). This questionnaire is presented in the Wiki after successful user login. Within the questionnaire, the user rates the quality dimensions; these dimensions are always related to a particular task. By reason of simplification, only two tasks are differentiated in this study: providing information, and finding information. Furthermore, these tasks are bound to the roles of the users. This means, based on the account data of the user, I decide whether the user fulfills the task of searching information or providing information. The assumption is that employees would rather provide information while customers would rather search for information.

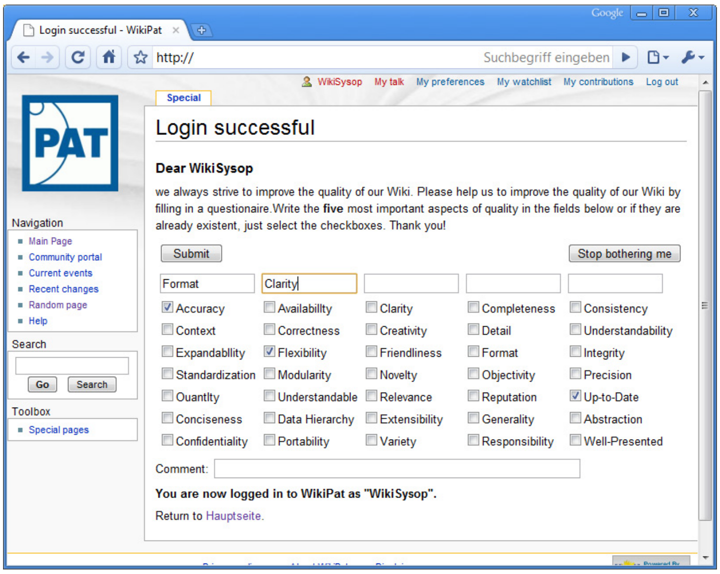


Figure 8-16 Wiki questionnaire plug-in (version 1)

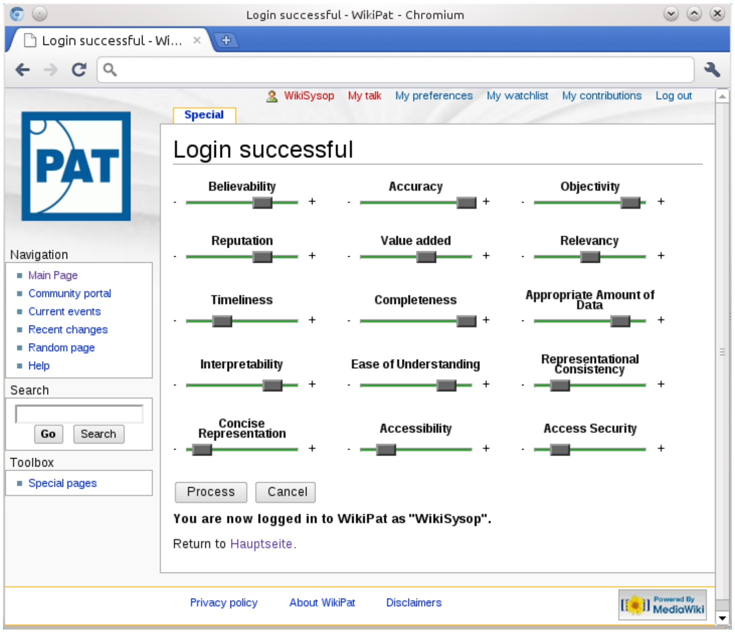


Figure 8-17 Wiki questionnaire plug-in (version 2)

Up to now, there is only the option of marking one dimension as important. Unmarked dimensions are ignored. This decision has been made in order to improve the acceptance of the questionnaire in the Wiki. Simply marking the dimensions is less time-consuming and thus increases acceptance. The rating given by the user is stored in the user quality profile. This profile contains different ratings for several tasks. Once the questionnaire is filled in, it is not presented anymore. The reason for this decision was not to bother the user; however, the drawback is that the user is not able to change his/her settings for an already rated task. But since the benefit for the user is not immediately noticeable and the users represent real customers and employees, the focus was to avoid annoying the users. Figure 8-17 shows the improved version providing a scale for setting the importance of quality dimensions.

The second Wiki extension facilitates implicit elicitation of user preferences by enabling the user to rate articles. This technique is well-known from many Web 2.0 portals. For expressing feedback, users are able to click rating buttons directly within the content. In this study, rating is limited to positive (good) and negative (bad) ratings. A good rating means that the article quality corresponds to the user's quality requirements; bad ratings mean the article quality must be improved.

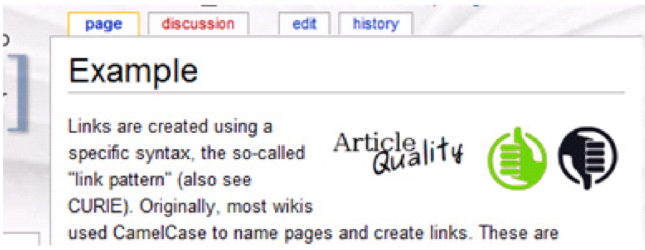


Figure 8-18 Wiki rating plug-in

Figure 8-18 shows the buttons that are integrated in every article. All feedback of the users is collected in a result table. This table contains, besides the username, the name of the article, rating, and a timestamp. The timestamp is important for recovering the quality status at the rating time because the content and thus the quality changes over the time.

After rating, the user quality profile can be inferred from the given user rating. Therefore, the rated articles are evaluated. The set of positively (good) rated articles is analyzed in order to find a pattern which best describes the user quality requirements. Such a pattern could be a particular combination of quality dimensions that has high ratings in all articles. This would mean that for a certain user, it is important for this combination of quality dimensions to be high and that the recommendation system should recommend only articles where these dimensions have a high status. The approach of this study is to analyze how the cluster algorithm must be configured to recognize the implicit and explicit profiles as one pattern.

8.3.4 IMPLEMENTATION

Both features are implemented as extensions for the prevalent open-source Wiki system MediaWiki. The extensions are written in PHP and available as open source. Using MediaWiki hooks, the questionnaire's rating buttons are integrated into the presented content. The data entered in the extensions is sent via web services to a central server. In this central place, the rating results as well as the user profile are stored, aggregated and maintained. By now, the information stored in the user profiles does not flow back into the system.

8.3.5 RESULTS

The aim of the study was to compare the calculated quality status of a resource and the user quality perception of this resource. For that reason, I used the Wiki plug-in for explicitly weighting the quality dimensions. In this way, I created a quality profile that represents which quality dimensions are notably important. For this experiment, I assumed that the context of the user is static, which means the tasks are always the same. The second step was to use the rating buttons to collect the quality rating of Wiki pages given by users. Thereby, I gathered information about which articles correspond to the users' quality needs. The last step was to calculate the quality status based on metric measuring. The objective was to understand the user's perception of good quality, i.e. which articles he/she rates as good quality and which articles the system recommends that correspond to the user's needs. The evaluation considers whether the system measures correspond to the user's rating. During the test period, 78 ratings were given by 18 users. 66 ratings were positive, 12 negative. 4 users (from the group of rating people) filled in the questionnaire and provide enough ratings to establish the user profile. The dependent variable in this experiment is the number of dimensions that are similar in both the resource status and the user profile. The independent variable is the threshold which defines similarity. Figure 8-19 shows the results for users A, B, C and D. The similarity threshold defines the threshold for which two dimension values are considered equal. To recall, the dimension values compared in this study are the explicit user rating on the one hand and the automatically calculated dimensions of the resource quality status on the other. These ratings would have to be similar if the user had rated a particular resource as “good quality”.

The threshold is an input parameter for the similarity algorithm and is defined as a percentage from the maximal absolute value. For instance, if a dimension value has a range from 0 to 100 with a threshold of 0%, only identical values would be considered as similar. With a threshold of 5%, a value of 60 and another value of 65 would be considered as similar.

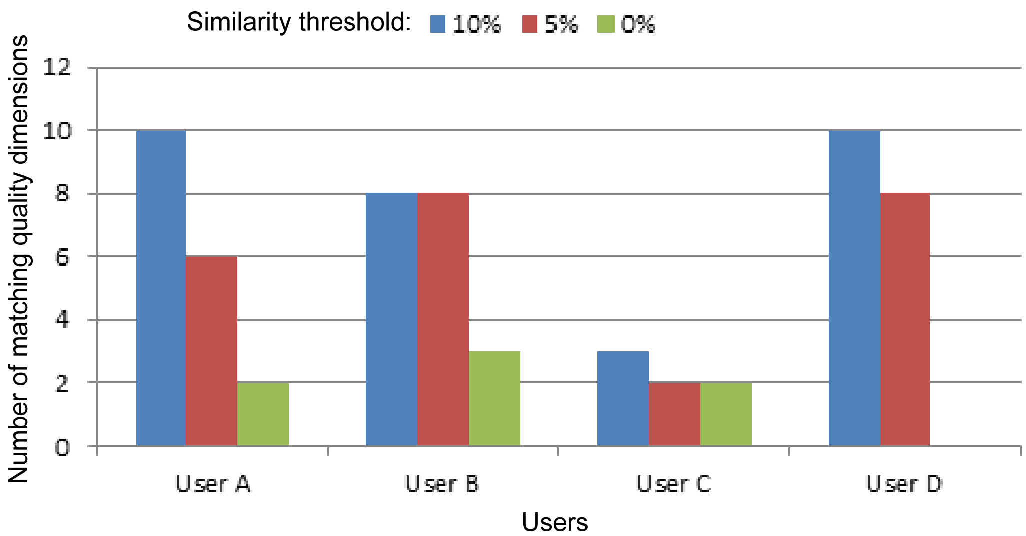


Figure 8-19 Graphical representation of results

Figure 8-19 shows the results for similarity thresholds of 0%, 5% and 10%. 12 quality dimensions were evaluated in this study. On a threshold of 10%, it would mean that if the values differ less than or equal to 10%, they are considered as similar. For users A and D, there is an overlap of 10 dimensions. User B has still 8 and user C only 3. The mean value of the 10% threshold is 7.75. With a threshold of 5%, the number of similar dimensions is between 6 and 8 except for user C with only 2. The mean value here is 6. For the 0% threshold, the number of similar dimensions goes dramatically down; A and C have 2, B has 3, and D has 0 common categories.

8.3.6 *CONCLUSION*

The objective of the study was to analyze the performance of the proposed algorithm in terms of finding similar quality profiles. The study is based on data gathered from a corporate Wiki. By means of additional rating functionality, the users assessed the importance of quality dimensions of articles in the Wiki. In addition, they rated the quality of the existing data. The question in this study was: for which parameters would the algorithm recommend the articles that have been rated positively by the users? The result has shown that there is a tradeoff between the threshold and the number of similar dimensions. What is missing here is the evaluation of how parameters can be changed so that the user perceives the recommendation as still useful. Both parameters influence the user’s perception of the recommendation’s usefulness. The smaller the subset (number of similar dimensions), the less will be the resource that corresponds to the user’s requirements. On the other side, increasing the similarity threshold leads to results that are less similar and thus correspond less to the user’s requirements. The study showed that by choosing a large similarity threshold, the subset is larger than with a small subset. A threshold of 10% seems to provide a subset that is large enough to be able to say that the profiles are qualitatively similar. A subject of further research would be to examine whether a threshold of 10% is still sufficient for the user’s quality requirements. A limitation of this study is the assumption that all users in the system are one community and that there exist only two tasks.

8.4 Evaluation of Quality Assistance in Social Media

Parts of this section have been published in:

Weber, N., Schoefegger, K., Bimrose, J., Ley, T., Lindstaedt, S., Brown, A., Barnes,S.A.,. (2009). Knowledge Maturing in the Semantic MediaWiki: A Design Study in Career Guidance. *In Proceedings of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines (EC-TEL '09), Ulrike Cress, Vania Dimitrova, and Marcus Specht (Eds.). Springer-Verlag, Berlin, Heidelberg*, 700-705. DOI=10.1007/978-3-642-04636-0_

Weber, N., Nelkner, T., Schoefegger, K., Lindstaedt, S. N. (2010) SIMPLE - a social interactive mashup PLE *In Proceedings: CEUR Workshop proceedings series, Proceedings of the Third International Workshop on Mashup Personal Learning Environments* (MUPPLE09), in conjunction with the 5th European Conference on Technology Enhanced Learning (EC-TEL2010) (2010 /10)

The previous studies mainly focus on algorithms processing data within a system. One important factor has yet to be considered – the user. This part of the evaluation focuses on how quality awareness in social software systems is perceived by the end users.

8.4.1 Introduction

This part of the evaluation shows the benefit of quality-aware systems from the user's perspective. The focus here is evaluating the interaction of users with interfaces that are enriched with quality indicators. In this part of the evaluation, I present several systems and interfaces which result from the iterative design and development process. The first phase in this process describes a design study that was conducted to derive requirements for the second phase, or the widget-based personal learning environment (Widget PLE).

8.4.2 Background

The evaluation describes the process of designing and developing a social software system in order to evaluate the benefit of quality awareness. The development and evaluation of user interfaces were conducted within the framework of the ongoing EU-funded project MATURE[6]. The goal of the project is to understand the maturing process and provide maturing support for knowledge workers in a collaborative environment.

From the project's perspective, the main goal of the study is to examine quality improvement within the context of knowledge maturing in a career guidance setting (MATURE Consortium, 2009c). In this setting, the task of career advisors is to personally consult individuals (such as students, graduates or their parents) on their job prospects, and to advise them on potential careers based on their interests and the general job situation in their region. In doing so, they make use of a large body of formally documented knowledge artifacts; for instance, statistics and reports on job opportunities or labor market development in certain employment sectors and regions. Additionally, they draw on a considerable amount of informal knowledge derived from their experiences with concrete cases. This knowledge in use is more or less systematically applied to their job, as well as being more or less systematically shared among practitioners. The aim of this part of the project is to systematically support personal advisers by providing tools for improvement of information quality.

The study is divided into two parts: the **first part**, presented in this section, was conducted in the form of a design study. Based on a rapid prototyping approach, design studies represent focused user-centered techniques to deepen the understanding of user requirements, and to analyze their feasibility from the perspective of technology. This approach takes models and algorithms into the field by developing quick prototypes that can be evaluated by prospective end-users, and thus providing early feedback on the viability of the approach.

[6] http://www.mature-ip.eu

The **second part** was designed and implemented as demonstrator - a prototype that is already usable over a longer period of time by the end-user. It describes the design and development of the widget PLE. Within this part of the evaluation, I want to explore ways on how to support the collaborative creation of high quality and reliable content-based knowledge artifacts in a self-organized way. The main focus of this demonstrator is on the first four phases of the knowledge maturing model (Schmidt, 2005). I try to analyze how to support content creation from an early stage up to high quality artifacts.

8.4.3 Design

In order to align our research closely to the real needs of our application partners, the consortium of the MATURE project decided to apply a user-centered approach (specifically, a participatory design approach) to our design, development, and evaluation process. In a participatory design (Bødker, Kensing, & Simonsen, 2004; Muller, 2003), end-users are invited to cooperate with researchers and developers during the innovation process. In fact, end-users participate during several stages of the process: during the initial exploration to help define the problem, during the development to help focus on the most appropriate solutions, and during the evaluation to provide feedback on the proposed prototype. The approach consists in developing successive feasible prototype versions of the final product. Rapid prototyping is characterized by agile, iterative, and incremental development cycles with integral testing, frequent, use-case-centric, and adaptive requirement analysis.

As a starting point, our application partners from career guidance described a scenario in which learning and knowledge maturing are supported. This scenario represented the main functional source for the development of our prototype. On the other hand, the main technological source was provided by our analysis of existing state-of-the-art technical approaches. I focused on investigating existing tools in order to avoid developing completely new functionalities, but also to find innovative ways of realizing the pragmatic functionalities elicited by the scenario.

At the beginning of the design phase, together with our application partners and other project members, I organized a workshop for evaluating the prototype. Participants were divided into two smaller groups to facilitate discussions and contributions from all participants. The prototype was presented to each group separately. We, the workshop organizers, encouraged them to discuss the efficacy of the prototype for supporting the previous developed scenario, its potential application by individuals in an organizational context, as well as future perspectives and new possibilities. The key points of each group were then presented to, and discussed with, the other group. All participants were also asked to fill in a feedback questionnaire, aimed at gathering specific requirements and additional design ideas. From the analysis of the gathered data, I collated the following information quality-related requirements (Weber et al., 2009):

- adaptation of the system according to the learning style and user preferences

  This requirement addresses the quality-aware adaption of the system. This means that system functionalities like search and recommendation ought to adapt to user preferences and learning styles. In that way, it provides results according to the user quality requirements.

- quick access to relevant information

  This addresses support in terms of scaffolding during content creation in order to facilitate the creation of high quality content. This includes providing article summaries and community ratings in search results and recommendations.

- community-driven quality assurance

Community support is an important factor for assuring quality since in social software systems the content is mainly maintained by the community. The community is responsible for keeping the content accurate. Functionalities supporting these activities must be incorporated in the system. In addition, community activities also facilitate keeping one up-to-date on a specific topic.

- integration with existing systems and datasets

  Particularly in an organizational environment, the integration of legacy systems in the social software system is a vital requirement for the acceptance of the system.

- awareness of collaboration

  Especially in social software systems, the aim of the users is not only to find and contribute information but also to share/spread information within the community. Sharing content and networking within the community facilitates the qualitative improvement of content.

Because of the bottom-up and user-centered nature of our design approach, I opted to ground the development of our prototype on concrete Use Cases, in addition to the rather abstract requirements. Our Use Cases are written in a high-level concrete language, independently from any technical implementation, and describe the system from the actors' viewpoint. They are based on the requirements derived from the design study, which realize their functional goals. They are also based on the *Personas* derived from the ethnographic studies (MATURE Consortium, 2009a), thus prolonging the real human aspect of our requirement specification process. Another key aspect of our Use Cases is their testability, in the sense that their post-conditions can be judged as satisfied or not. They constitute the key representational scheme of our development process, and can evolve during the stages of the developmental lifecycle. A detailed description (including involved Personas, underlying activities, organizational aspects, relation to knowledge maturing, and limitations) of our Use Cases and of the five classes used to group them can be found in MATURE Consortium, (2009a). Based on these Use Cases, the preliminary scenario developed by our application partner was refined.

8.4.4 SMW Design Study

The design study, as part of the requirements for elicitation and analysis, is aimed at identifying requirements for a future system that supports the maturing process of knowledge objects. The basic idea of the study is to stimulate knowledge maturing by providing support for improving the quality of resources. For that purpose, the three dimensions of content, semantics, and processes (cf. Section 5.1) were taken into account.

Furthermore, the objective of our design study was to explore initial key ideas in order to have them validated prior to a full-scale development process, during which important issues would have been more difficult to address. Another aim of this study was to further analyze the users' requirements, in synchronization with the ethnographic study (MATURE Consortium, 2009a), which was performed in parallel. Figure 8-20 illustrates the detailed work plan which shows the iterative design, development and user-feedback cycles, as well as the parallel conducted ethnographic study that provides additional results/insights.

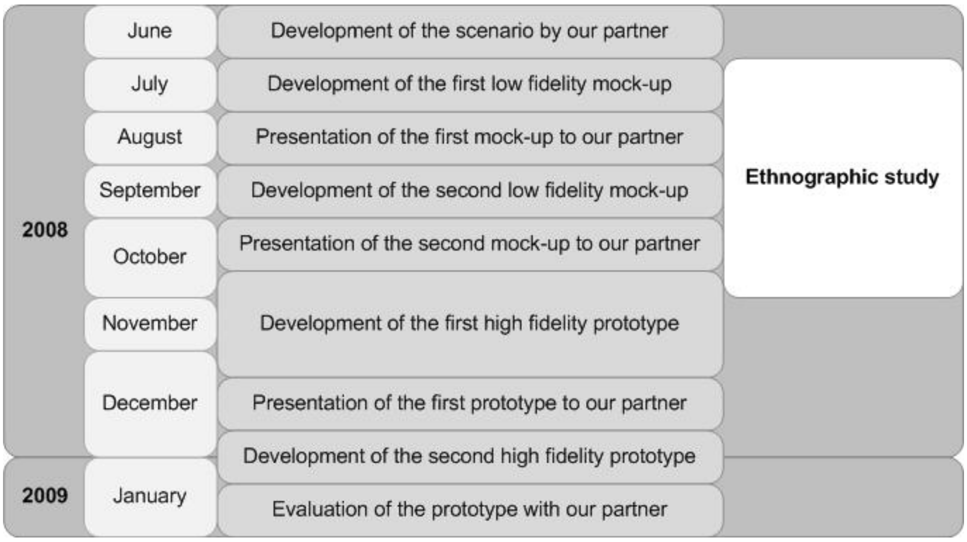


Figure 8-20 Development phases – the development phases show the participatory design approach by iteratively involving the end-users in the design process

The main focus of this approach is to make the user aware of the current text quality. From the perspective of a knowledge worker, Wikis are very well suited for enabling the maturing of artifacts, especially due to the ease of editing content and the policy that everyone can edit anything. Additionally, they allow for tracing the collective construction process (utilizing the Wiki's history functionality) and discussion processes around artifacts. The career guidance sector is heavily content-dependent, thus a Semantic MediaWiki was chosen as the basis for a prototype supporting knowledge maturing in this sector. Therefore, I developed services for content and structure analysis and integrated them into a Semantic MediaWiki[7]. The objective of analyzing the article content is to facilitate the assessment of the quality of a document and enable the user to improve article quality. The current quality level makes it possible to decide whether the quality of a certain document should be improved by supporting the user in creating or editing content. That is why I developed mechanisms for the automatic assessment and semi-automatic improvement of article quality. The bottleneck in assessing quality of text is the selection of qualified attributes that

[7] http://semantic-mediawiki.org/

reflect the quality of the content. Because of this, I adopted readability formulas as well as quantitative measures for semantic markup.

These functionalities have been developed to enrich the Semantic MediaWiki in terms of searching, collaborating, adding semantic mark-up and visualization. Some of these services for knowledge work support will be described in detail in the following section.

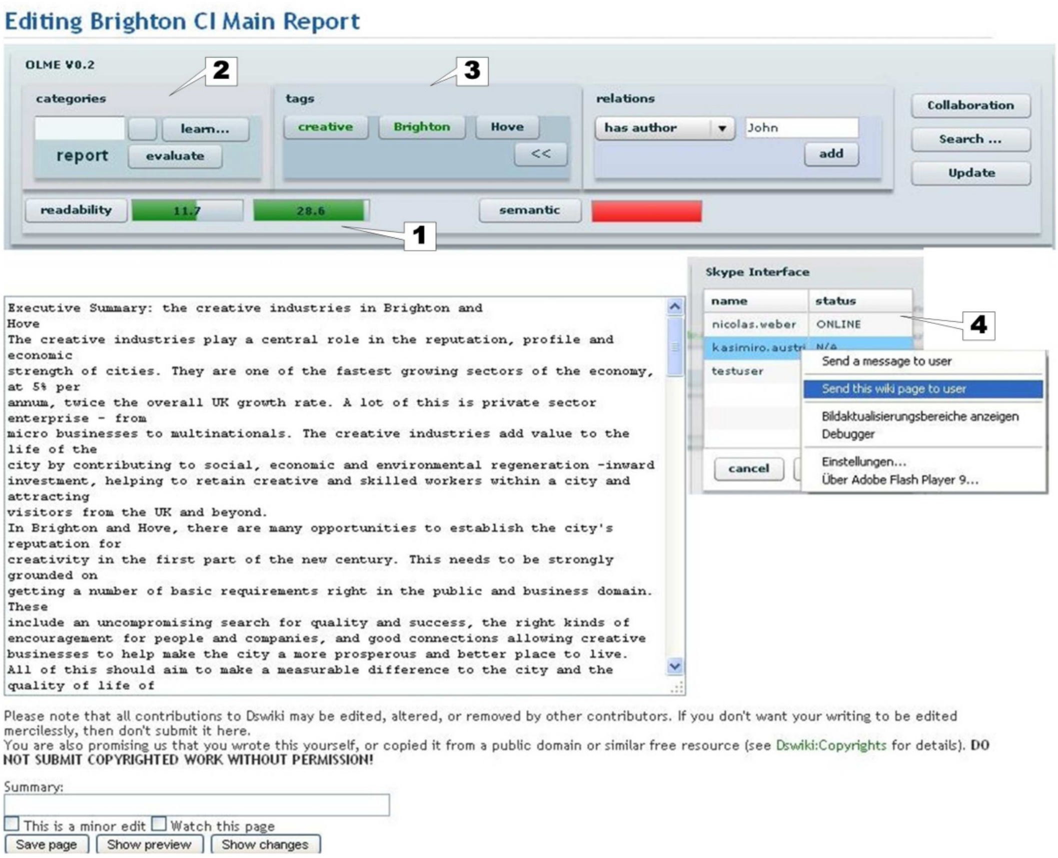


Figure 8-21 Semantic MediaWiki extension – editing page enhanced with top-bar

8.4.4.1 Text Quality Indicator Services (1)

The objective of analyzing content is to facilitate the assessment of the document's quality status. This allows for deciding whether the quality of a certain document should be improved by supporting the user in creating or editing a knowledge artifact. The bottleneck of assessing quality of text is the selection of qualified attributes that reflect the quality of content. Assuming that readability and quality have a strong correlation (see Braun & Schmidt, 2007), I tested within the design study two metrics for readability scores.

Both are based on reading scores (Stvilia et al., 2005) calculated from quantitative metrics like sentence length, number of syllables or number of words. The indicator for content maturity is calculated in real time during the editing process of a knowledge artifact. The result of the content analysis is then used to display the current status within the user interface. So the user is urged to improve the text quality. In addition, the quality status can be the basis for the recommendation of documents with similar quality. The user can derive patterns from these documents in order to improve the quality of his/her documents. Furthermore, the indicators can be used to trigger actions depending on a threshold; thus, if the system identifies a low quality status, it can provide tools and resources for quality improvement.

Figure 8-22 Readability indicator

In the **Flesch Reading Ease** test, higher scores indicate material that is easier to read; lower numbers mark sections that are more difficult to read. The formula for the Flesch Reading Ease Score (FRES) test (Si & Callan, 2001) is:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

Scores can be interpreted as shown in the table below.

Score	Notes
90.0–100.0	easily understandable by an average 11-year old
60.0–70.0	easily understandable by 13- to 15-year old students
0.0–30.0	best understood by college graduates

Table 8-3 Interpretation of FRES

′

The **Gunning fog index** measures the *readability* of a sample of English texts. The resulting number is an indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading.

The complete formula is as follows:

$$0.4 * \left(\left(\frac{\text{words}}{\text{sentence}}\right) + 100\left(\frac{\text{complex words}}{\text{words}}\right)\right)$$

These indicators are based on the content of a Wiki article; the semantic indicator provides a quantitative measure for the semantic annotation. Since semantic mark-up is a very important factor for identifying relevant information in the Wiki, this indicator should enable the user to assess the amount of semantic mark-up of his/her article and additionally to motivate him/her to add mark-up until the bar switches its color from red to yellow and then green.

Figure 8-23 Semantic indicator

The **semantic indicator** is calculated as follows:

Given that #semantic attributes > 0, the index is calculated with the following formula, otherwise I_{sem} is set to 0.

$$I_{sem} = \frac{\#words}{\#semanticattributes * 100}$$

The resulting score is classified into three categories (see the following table) so as to present the indicator in an easier way to the user.

0.0 – 0.69	Red
0.7 – 1.39	yellow
1.4 - ∞	Green

8.4.4.2 Categorization Service (2)

The categorization service automatically provides a category which best suits an article. The service is based on machine learning techniques. Depending on the content of an article, the system analyzes the words that are used and their frequencies to recommend the most used keywords as tags for the article. In order to categorize articles, the system suggests already existing categories which best correspond to the newly created content. Additionally, the user can add a certain category which seems to be appropriate and can train the service with this category such that the system will suggest this category in the future for appropriate and related articles.

Figure 8-24 Category recommendation

The recommended category can be added to the article by clicking the add button. The user is enabled to improve the precision of the classifier by training the classifier using the learn function. The service facilitates structuring content by recommending a suitable category from a given set of categories.

8.4.4.3 Mark-up Recommendation Service (3)

One problematic aspect of Wikis is their inability to deal with more formal content or structures. In a standard Wiki, it may seem that any artifact is constructed from scratch in a community setting, and that there is no end to this construction process. This is an unrealistic observation in most settings and especially in an organizational setting where knowledge generation uses artifacts that fluctuate between the informal and formal pole.

The Semantic Media Wiki is enhanced with Semantic Media Wiki markup so that it is easy to use and users do not have to be familiar with formal markup for writing articles with improved quality. Semantic mark-up supports the author and, most importantly, other users, to refine articles concerning a certain topic of interest. The mark-up recommendation bar is divided into two areas: tag recommendation and relation selector.

The markup recommendation services strive for two goals. First, they aim to lower the barrier for creating mark-up which replaces the complex Semantic MediaWiki syntax; and second, they try to improve the quality of structure by recommending meaningful, pre-consolidated mark-up. Creating semantic mark-up contributes to the enrichment of Wiki content. The additional annotation of articles enables the user to browse through the Wiki and facilitates the retrieval of knowledge based on semantic mark-up. In addition, the mark-up is used as the basis for the recommendation of useful resources and visualization of emergent content structures.

Depending on the content of an article, the system analyzes used words and their frequencies to recommend the keywords most often used as tags for the article, see Figure 8-21 (marker 3).

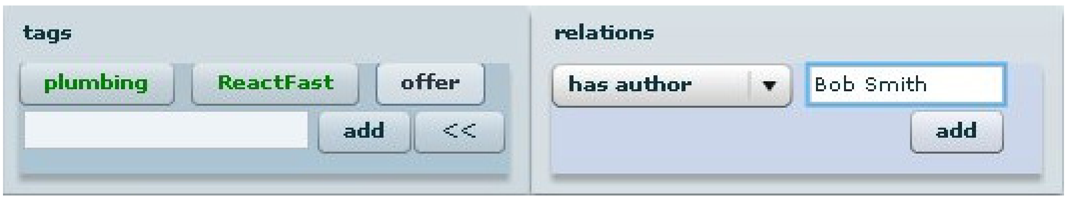
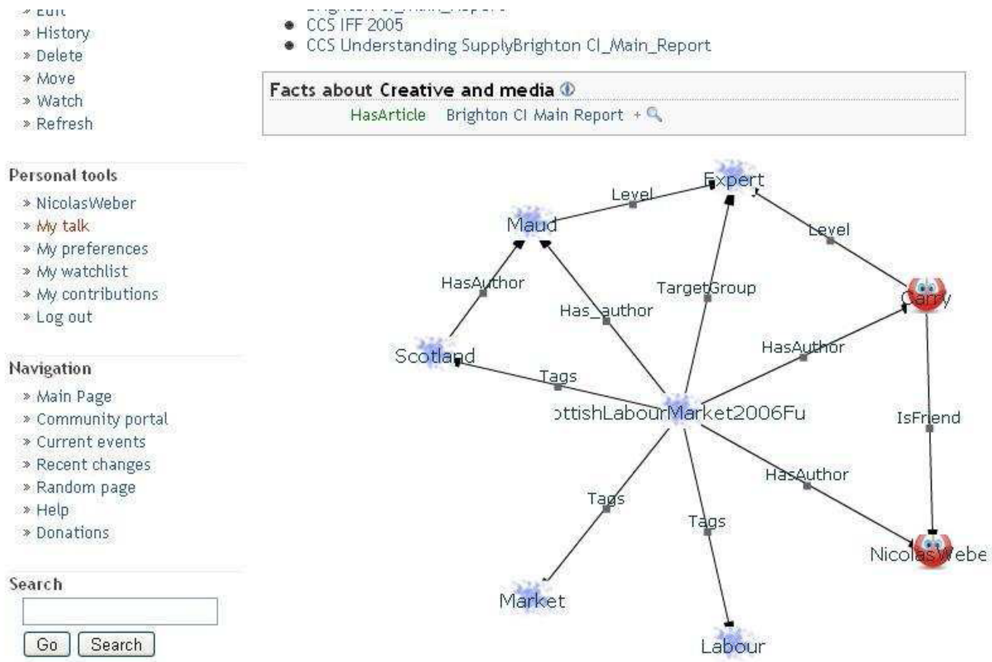


Figure 8-25 Markup recommendation plug-in

The tag recommendation is based on the Information Extraction Service described in Klieber et al. (2007). This Wiki extension aims to make tagging of resources as easy as possible. The system recommends a pre-consolidated set of tags, based on the result of content analysis. Adding a tag to the article requires just one click on the recommended tag.

8.4.4.4 Service for Visualization of Semantic Structures

This service provides visualization for the content of the Semantic Media Wiki. Each node in the graph represents either an article in the Wiki or a registered user. Directed edges represent the relations; for instance, an article might have an assigned category, author, tag or linked article. A user might have written one or more articles, or a category might contain one or more sections, articles, tags, etc. Depending on the choice of the maximum path-length shown, the user can define how many levels (and nodes) of the network are shown in the visualization, as well as the type of representing graph (e.g. hierarchical, cyclic). By clicking on a node in the graph, the visualization is updated and its connected nodes are shown; this enables the user to browse easily through the content of the Wiki within the graph. Additionally, new nodes (users or articles) can be created; articles corresponding to a certain node in the graph can be opened and edited in a new browser window; and users corresponding to nodes can be contacted by using the Collaboration Initiation Service. This service supports the daily work of users by enabling visual browsing through Wiki content from article to related articles or users. Thus, the service provides an overview of related topics and experts and offers easy negotiation by embedding a collaboration service (see next Section).



8-26 Service for Visualization of Semantic Structures

8.4.4.5 Collaboration Initiation Service (4)

This service offers the facility to easily initiate collaboration with authors of articles or interested persons via Skype (see Figure 8-21 (marker 4)) by not having to switch to another tool since it is embedded into the Wiki. The user can send messages or web-links to Wiki articles in order to support negotiation and consolidation of artifacts. Additionally, within the visualization of the Wiki network, every author related to an article in the Wiki can be contacted by clicking on the author's node.

8.4.5 *WIDGET PLE*

In this section, I present the design and development of a mash-up learning and maturing system. The approach is based on widget technology which is seamlessly integrated into the familiar work environment of the end user. The widgets provide functionalities for creating high quality content on the one hand and access to indicators for quality assurance on the other.

The application scenario is also related to the work process of a personal adviser within career guidance organizations. This prototype is thought to be a continuation of the first study and responds to feedback gathered from the previous study. Thus, this demonstrator is also based on a MediaWiki for content creation and sharing but provides a different, widget-based interface. Furthermore, it allows for different spaces, taking into account privacy and security restrictions (see Section 7 on Social Server).

The design and development process was again split into two phases. These phases focused on different aspects of supporting quality improvement. The first phase is characterized by the creation of semantic structures in a content network which provides the basis for the second phase (cf. Weber, Nelkner, & Schoefegger, 2010). The second phase focuses rather on developing mechanisms for quality-aware search support and recommendation. Results from this work have mainly been integrated into the second phase of the demonstrator, whereas results from the first phase influenced the development of this approach.

8.4.5.1 Widget Frontend – Phase 1

As already mentioned, this demonstrator is based on a light-weight widget-based technology; Figure 8-27 shows a personal desktop with a set of widgets opened. In the following, different kinds of services, which this demonstrator offers as part of the widgets, are described in detail.

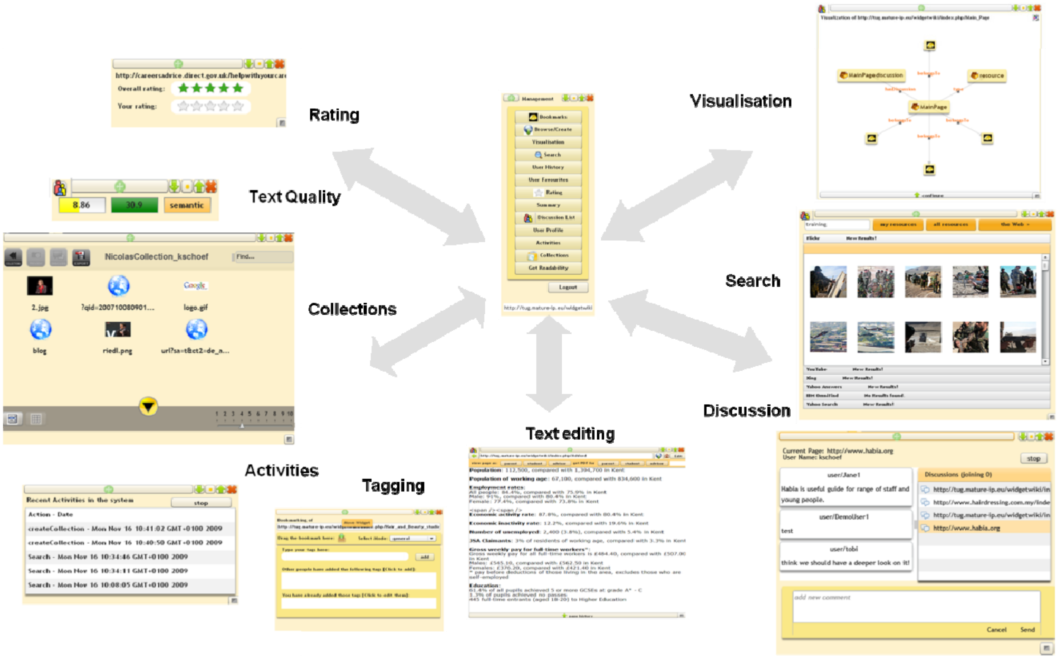


Figure 8-27. Desktop showing various widgets of demonstrator 1

The maturing activities within the work process of personal advisors described in Attwell (2007) include access to and search for information, aggregation of information, manipulation of documents, analysis and reflection of previous work, presentation and representation of created knowledge objects, sharing of information, and networking with other people. According to the knowledge maturing activities in MATURE Consortium (2010), the functionality of the widgets cover the following activities that are mostly relevant in the context of career guidance:

1. Find relevant digital resources

The search widget provides tools for searching various types of resources (websites, pictures, video, documents, experts) in different sources. The information quality of created material is improved by providing various information sources as a seed.

2. Embed information at individual or organizational level

All information can be stored either in a private repository, not visible to other users, or in a common repository in order to share information and collaborate. The user can decide whether the quality of the created artifact corresponds to the quality requirements of the community.

3. Keep oneself up-to-date with organization-related knowledge

Resource summaries provide information about latest changes and activities of other users in a resource. This helps to stay up-to-date with the documents' content and ensures timeliness of information.

4. Familiarize oneself with new information

By means of tags, discussions and collections, information is always presented with contextual metadata. This function helps provide an overview of new/existing information.

5. Reorganize information at individual or organizational level

By using tags and collections, information can be easily reorganized either in the private space or in a shared space. The quality of artifacts is improved by adding meaningful metadata.

6. Reflect on and refine work practices or processes

Supporting functionalities for reflecting on own activities in the system are planned but not yet implemented.

7. Collaborate with creation and co-development of digital resources

Digital resources can be shared with other users within the system. In this way, collaboration is facilitated. For collaborative creation and editing of resources, a Wiki-system is connected and can be used together with the widgets.

8. Share and release digital resources

The widgets allow easy access for other users by using public collections for sharing digital resources.

9. Restrict access and protect digital resources

Private repositories allow each user to hide information and protect his/her resources. All private information can be published by the user at any time.

10. Find people with particular knowledge or expertise

The user profile generation, implemented in the Social Server, evaluates relevant topics for each user. This information will in future be used to recommend experts depending on the current user context.

11. Communicate with people

Discussion functionalities, integrated in the discussion widget, allow for communicating directly within the widgets.

12. Assess, verify and rate information

Rating functions for all types of resources allows for explicit feedback on relevance and quality of the resources.

8.4.5.2 Widget Frontend – Phase 2

In phase 2, the focus of development was on refinement of maturing support and quality awareness functionality based on the end-user feedback from phase 1. According to the results of formative evaluations, usability issues were addressed. Especially, Search, User & Resource Profiling, and Awareness Provider were either newly developed or have been adapted according to the requirements of the integrated instantiations. More specifically, the following building blocks were developed.

Resource Search

Resource Search is fully based on the integration activities. Our application partner uses this Building Block as Search Widget. Specifically, it provides two integration options:

- Facets: Users are provided with a faceted search. The facets are adopted from SOBOLEO spaces (Braun et al., 2008). One space is collaboratively managed by all users and another space is a system space, which allows for predefining facets for the search. In addition, quality indicators can be used as facets to narrow down search results. This functionality goes beyond showing indicators how to create quality awareness since the user can actively arrange search results with regard to their quality status.

- Person Search: As the common vocabulary is used for tagging digital resources and persons, the search mechanism also uses search terms for retrieving persons. Thus, users can find experts more easily and more context-dependent. Moreover, as people are provided with digital resources, both can be brought together more easily. As in resource search, users can also be arranged according to quality aspects. But in the case of users, quality rather addresses the quality of resources the user is in touch with or his/her position in a social network.

- Usability: According to results of the formative evaluation, usability was improved. Users can now see the most important information of a search result at a glance. Moreover, digital resource search results have a link to the following activities: Tagging, Add to Collection, Show Details, and Discuss.

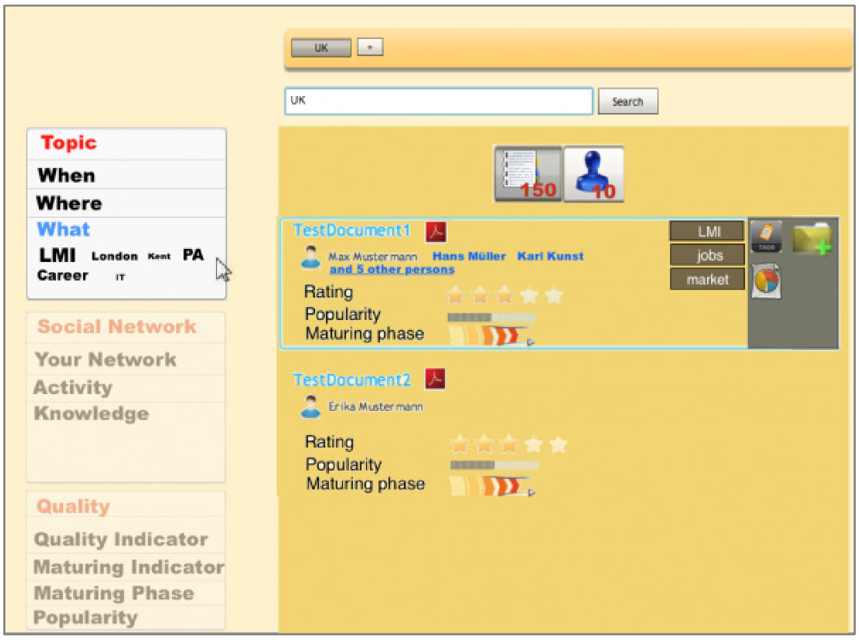


Figure 8-28 Faceted search based on user-quality requirements

Resource profiling provides details about digital resources. Figure 8-28 shows the details page with related resources, related persons, quality indicators, recent activities on the resource, quality statistics, and a tag cloud. In a first step, related resources are only collections which contain the resource and all discussions about that resource. In further steps, this could also encompass digital resources that show a certain degree of similarity. Related persons are those who have added the resource, discussed it, used it in a collection, and so forth. Furthermore, maturing indicators which fit the current status of the resource are displayed. It can be a subset of all defined maturing indicators (MATURE Consortium, 2010); for example, *'The resource was just created'*. This can provide a maturing status of the resource and may help improve it. Recent activities encompass a list of actions carried out with this resource, e.g. "X has been tagged." This usually includes more statements than the list of KMIs. The quality statistics present different metrics of text quality, e.g. readability. The overall list of metrics and the corresponding maturing service is described in D4.3. Finally, a tag cloud is presented.

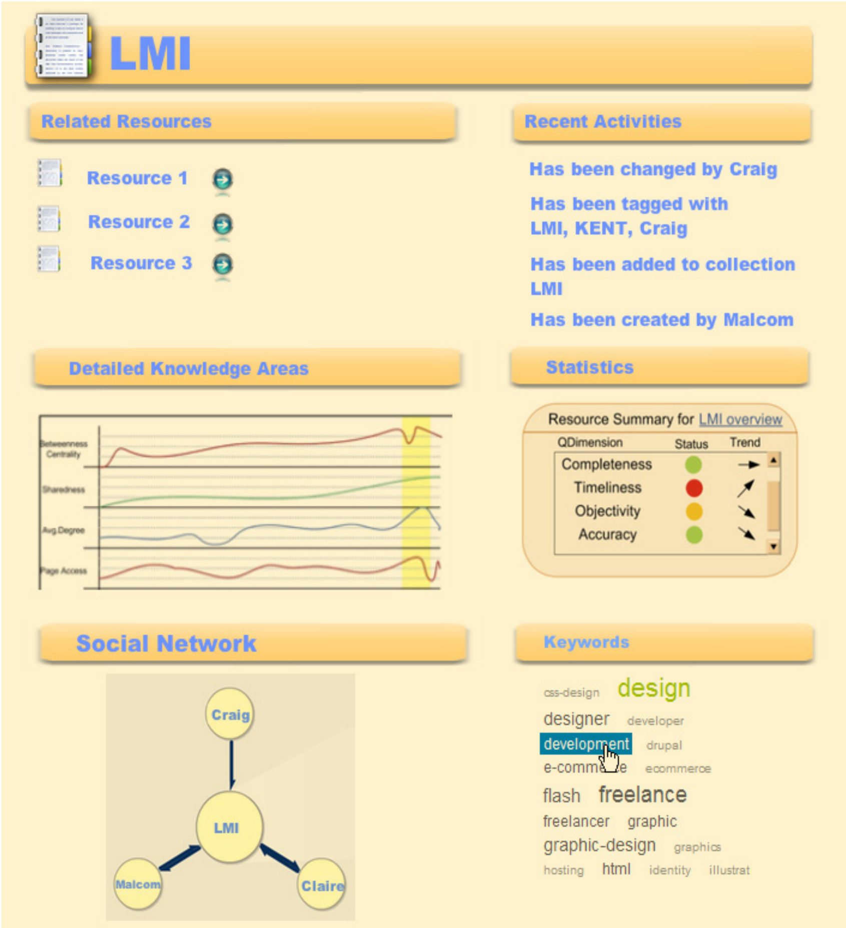


Figure 8-29 Resource summary containing qualitative development and status

8.4.6 *Results*

In order to gain more insights and new ideas about how a system could support improving content quality, prototypes have been developed and evaluated in a real world context of career guidance organizations. The implementation of the prototype for this design study was done using rapid prototyping which involves iterative design phases using mock-ups and development phases. This was combined with regular input to generate feedback over the viability of our approach when it comes to supporting knowledge maturing in the context of career guidance. The services were evaluated within a workshop. By means of hands-on sessions, the end users got an impression of the look and feel of the system. For the elicitation of feedback, we (the design and development team) recorded and evaluated the workshop discussions. In order to assess special aspects/functions, we used questionnaires and face-to-face interviews. The following section specifies the general results of the workshop.

- **Visual Appearance**
  Visual adaptation of the system would be necessary depending on individual preferences and learning styles. The easier a user can adapt the system to his/her (quality) needs, the more likely that his/her motivation for using the system in everyday work will grow.
- **Easy access to relevant information**
  Users might not have all the time in the world to research information, so easy access to relevant articles should be provided. To support this, each article could have a summary which is shown when articles are listed as a search result or at the top of a page. Additionally, this summary could be shown within the visualization of the Wiki content when an employer moves the mouse over a node representing this article. This result reflects the aim of my work very well since *relevant information* can be interpreted as meeting the user's quality needs.

- **Accuracy control**
  Accuracy control concerning time and content is necessary to make sure that the data is accurate, up-to-date and relevant. Long articles are unlikely to be read let alone being too time-consuming for the user when searching for any specific information. Instead of a moderator, the idea of automatic date flags could be used to remind authors and editors to update a certain knowledge artifact. As to my approach, accuracy, timeliness and relevance are features that can be tracked in quality aware systems.
- **Awareness of collaboration**
  Collaboration within organizations encourages employees to discuss new ideas and provide help when questions arise or problems are encountered. The user should be able to see immediately who is online and who can be asked for help or discussion.

In order to evaluate the effectiveness of our approach, we (the MATURE consortium) are pursuing a long-term field evaluation within the work settings of our application partner. I focus here on the end-user evaluation rather than the technical evaluation. The high-level key questions guiding us are:

- Is the prototype usable?
- Does the prototype support knowledge maturing and thus quality improvement?
- How useful and acceptable is the prototype in an organizational context?

This section provides an excerpt from the results of the demonstrator's formative evaluation in the MATURE project. Results that mainly focus on content quality are presented here. For a comprehensive overview of the results, see MATURE Consortium (2010b).

The first evaluation phase consisted of a structured scenario-based walkthrough, coupled with an interview. It aimed at assessing usability, as well as gathering functional and non-functional requirements. As a preparatory work, together with the end-users we devised scenarios (realistic tasks with their goals and constraints, based on the results of the ethnographic studies and the Use Cases) that were used for walking through the prototype features. For the walkthrough, two groups were formed, each one with two end-users, one developer providing guidance and technical assistance, and one facilitator audio-recording or documenting the end-users' interactions and reactions. The two end-users collaboratively used the prototype and were asked to think aloud, even though some questions were also triggered by the facilitator. After the walkthrough, the end-users individually provided feedback by filling in a questionnaire.

The evaluation of the demonstrator was conducted iteratively in several workshops. During these workshops, feedback was collected by means of questionnaires, interviews, audio and video protocols. The resulting artifacts consisted of 291 statements (items) extracted from all types of media. The 192 items of Phase 1 were clustered into the 3 major categories of *non-functional requirements*, *use cases* and *functional requirements*, with the emphasis on the final category covering 9 subcategories and 123 statements. While in Phase 1, 64% of all items were instantiations of functional requirements, in Phase 2 this category only comprised 35% of the textual data corpus. Among the *functional requirements,* 23.6% of the items were allocated to the category *improvement of existing widgets*, such as the extension of the search widget by filter-functions and the option to refine and broaden search results; as well as *information about certain usage data* (9%), as in the functionality to log and point out recent modifications of articles and indicate who has edited the articles. Two use cases were conceived by the P.A.s: the *creation of an organizational vocabulary* and a *community-driven quality assurance*.

Figure 8-30 shows statements assigned to demonstrator functionality clusters. The graph shows that quality indicators, cooperation support and tagging were perceived as useful, whereas gardening and recommendation need further improvement.

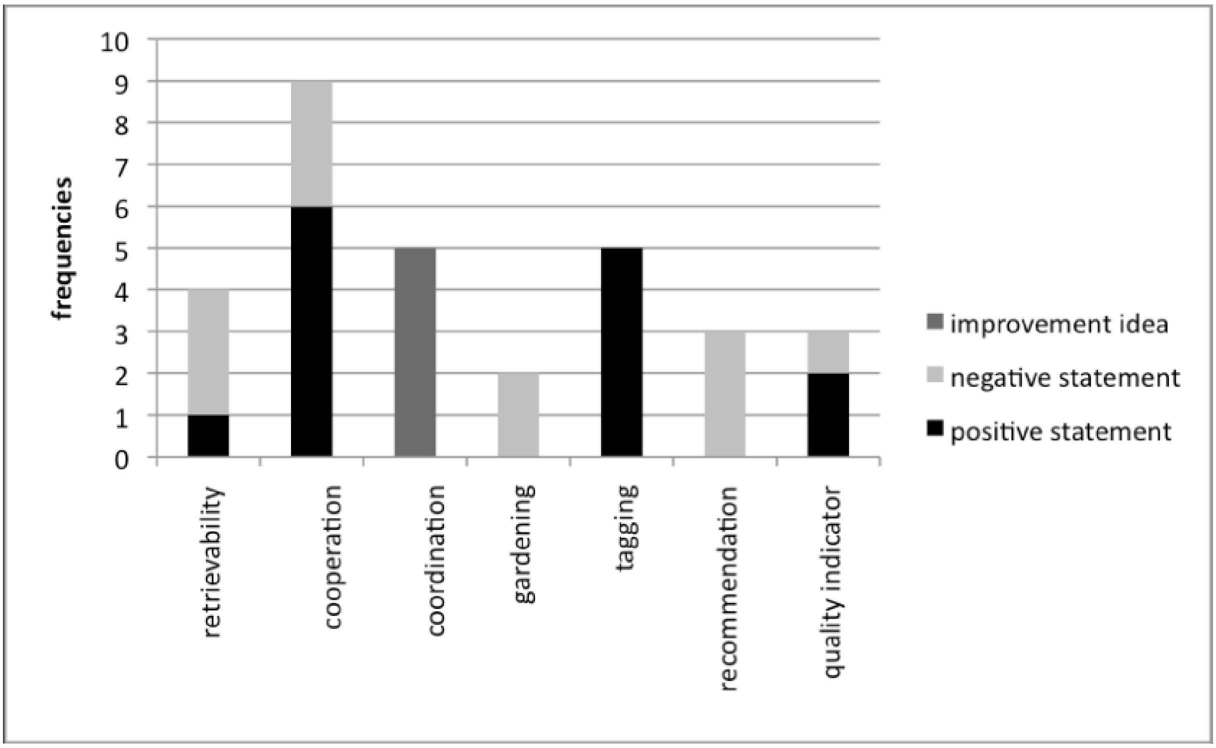


Figure 8-30 Results of evaluation

Figure 8-31 provides a more consolidated version of the formative evaluation results. The topic map shows concepts that were identified for improvement and their relations and interconnections. In terms of quality awareness, the graph names two important concepts: quality indicators for information quality improvement, and interface quality for usability improvement. Furthermore, several topics indirectly influence quality. In particular, semantic structures are referred to as a prerequisite for irretrievability, what represents an important factor for quality. Furthermore, information quality was identified as a prerequisite for one of the overall goals of the demonstrator: a community-driven quality assurance.

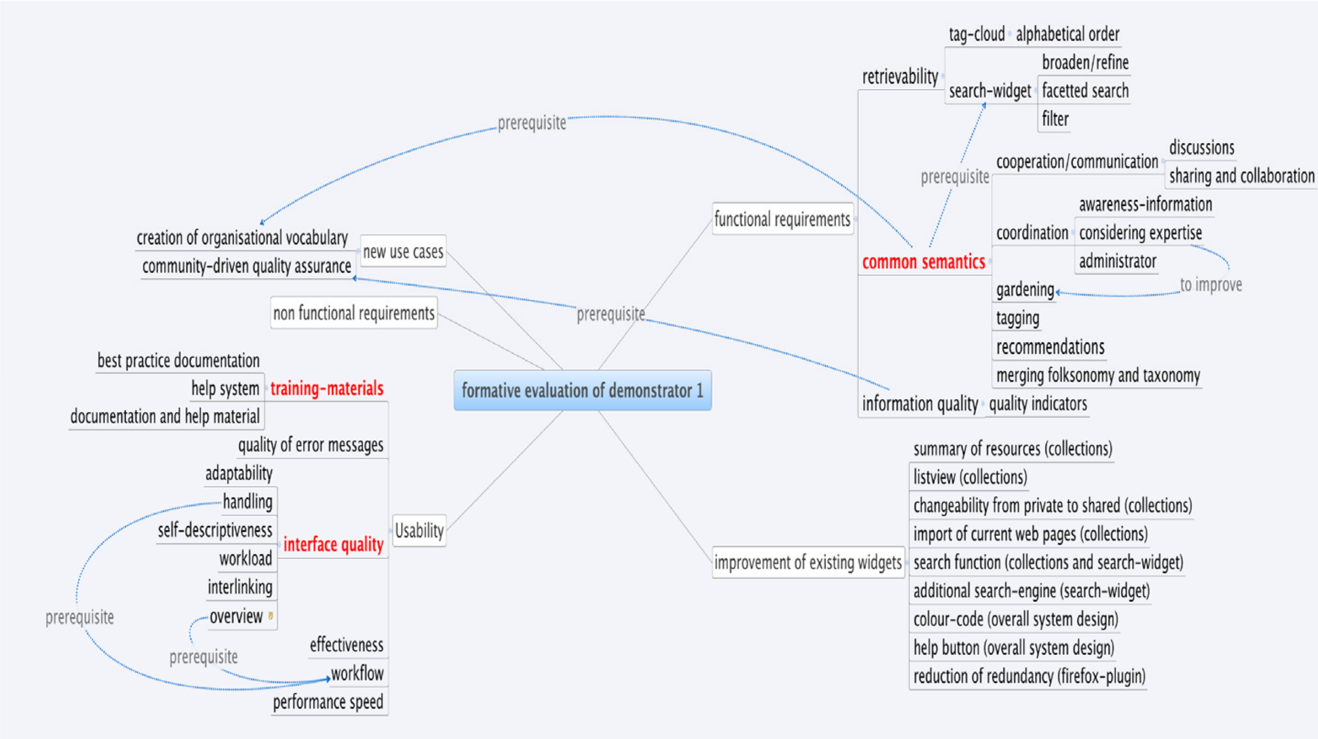


Figure 8-31 Interconnection between entities

Finally, this evaluation has identified two further issues to be considered in future design development. Both concern the provision of information to improve usability on the one hand, and facilitating quality assurance on the other:

1. Supplying information in the form of training materials should improve usability. The interaction with the system will become easier if the best practice documentation about the purpose and potential use of the system clearly explains the handling of widgets and buttons.

2. Providing information in the form of quality indicators should support quality assurance. System efficiency should be enhanced by presenting reliable indicators of information quality, such as rating or readability scores.

8.4.7 *Conclusion*

From this project's perspective, the main purpose of this work was to gain an insight into the knowledge maturing process in the real-world-context of organizations from the career guidance sector by developing a tool that supports quality improvement. The potential of the system in this context was explored, with the research firmly focused on how the utility of this system could be further enhanced. Several quality indicators and services have been designed in an attempt to bridge the gaps in the maturing process. Furthermore, an evaluation of the prototype in a real-world-context helped obtain a deeper insight into the features relevant for supporting quality improvement in career-guiding organizations. The main aspects of their requirements could easily be adopted for a system supporting knowledge maturing in terms of any knowledge worker in other contexts.

One result of the Semantic MediaWiki design study is that the willingness to provide high quality content depends for the most part on the time-consumption for the user. Since the user contributes to organizational knowledge and benefits only indirectly, the creation of mark-up has to be easy and fast. *Markup Recommendation Services* are connected to both the content on which the mark-up is based and the semantic structure which describes the content. Thus, the recommendation service is able to compute relations and attributes based on the content and eases the creation of mark-up by accessing the semantic model directly. A bottleneck in the creation of semantic annotations (such as relations, attributes, and tags) is an inconsistent vocabulary. Ambiguity, misspelling, similarity in term semantics hamper the creation of meaningful annotation and necessitate the correction and consolidation of the semantic structure. Recommendation services use NLP techniques for the discovery of semantic relations and attributes. By mapping the identified mark-up to a common vocabulary, the service avoids inconsistencies within the semantic model. A design principle in developing recommendation services is to make sure that the user is always in control. The objective of these services is not to create mark-up automatically but rather to provide support by recommending annotations. Using quantitative metrics as an indicator for content quality has proved to be useful. By presenting the resulting values to the user, awareness of information quality problems is raised.

The two phases of this approach represent the two steps required for creating quality-aware system behavior. One important prerequisite is to provide functionalities for establishing and maintaining content metadata and structure. Functionalities, such as tagging, rating and annotating resources, facilitate *gardening* of emerging knowledge structures. These structures are required for creating entity models as in user models (user quality profile) and resource models (resource quality profile). Based on these models, high level services, like recommendation, awareness- or context-sensitive services, can be provided. The provision of these services is covered by phase two of the demonstrator's development. The evaluation has shown that these services (including gardening) facilitate a community-driven quality assurance as well as the creation of an organizational vocabulary, what all along was the intended and principal goal of the demonstrator.

9 Conclusion

The idea behind the here proposed approach is to determine both the quality requirements of a user (or a group) and the qualitative status of resources, and to have them both represented in terms of quality dimensions. This allows involving an additional feature to the decision-making processes in social software systems – information quality. The main result following this approach is that the system does not only provide information based on the textual content, the category or other metadata, but is also based on the users' quality requirements. This behavior is denoted as 'quality aware system behavior'.

Thus, the objective of this approach is to design, implement and evaluate functionalities fostering quality awareness in social systems. Therefore, the main research question of this work has been expressed as follows (cf. Section 1.1):

How can quality aware system behavior in social software systems be achieved?

In the context of this work, quality awareness means that decisions in the system are made by taking information quality into account. This goal is attained in this approach by making quality explicit. Representations of artifacts and users are enriched with metadata about their qualitative status and quality requirements. Quality aware system behavior is facilitated by proposing algorithms that include quality metadata in their calculations. The result is a system-user interaction influenced by the qualitative status of the resources and the quality requirements of the user. To highlight the impact of this work, in the following, three application scenarios are sketched, emphasizing the strengths of the proposed approach:

Ranking of search results according to their qualitative status: Usually, search results are ranked according to their similarity to a particular search term. More sophisticated approaches are able to sort resources according to the user's topic of interest or current task. The here suggested approach adds another dimension by considering the user's requirements regarding quality. This enables social software systems to provide search results that correspond to the users' requirements, not only regarding information content but also in terms of information quality.

Giving an overview of the qualitative status of resources: In many situations, users can choose a resource from a set of given resources. Attributes that support the decision which one is the required resource can be (file-) names, user ratings, tags, etc. The proposed approach overcomes the problem that resources first have to be read/viewed in order to assess the qualitative status. It facilitates finding adequate resources by showing their qualitative status.

Recommendation based on the qualitative status: Recommender systems provide proactive support to the user by providing resources that are relevant for him/her. In this context, relevance can be defined at different levels; resources that topically fit to the current domain of the user, resources that help users to fulfill a current task, as well as experts that can be contacted to provide support. Quality awareness in this context means that the qualitative status is incorporated as well. This work proposes functionalities that facilitate quality-based recommendation(s) of resources and experts, and in this way provides proactively information that fits to the users´ quality needs at the time the user needs it.

My approach for creating quality awareness in social systems comprises three main pillars of which each is reflected in one of the three research questions:

1. How can the qualitative status of resources be measured and represented?

My extensive literature review of quality models has shown that considering quality as a multidimensional concept is the most popular and adequate solution. This perception is also adopted in this work by integrating and extending the quality model of Wang and Strong (1996). Based on this model, I have developed a multilayered quality assessment model for social systems (cf. Figure 4-2).

This multilayered approach is also reflected in the semantic representation of the resource quality. This approach proposes a model for representing the current status of the resource quality by describing several quality dimensions of a resource. The resource quality model extends existing approaches in terms of expressiveness and flexibility. Two studies have been conducted to answer the first research question.

In the first study (*Quantitative Evaluation of Metric Values*, cf. Section 8.1), I analyzed several instances of a Wiki regarding article quality. I therefore created system representations in all three categories. Based on these representations, I calculated the metric values over a longer period of time. I evaluated the development of the metric values using statistical methods. The result of this study shows, that qualitative changes are reflected in different system representations. That means, using the metric model proposed in this approach, various aspects of information quality in Wiki articles can be measured.

The second somewhat application-oriented study (*Evaluation of Quality Assistance in Social Media,* cf Section 8.4) evaluated how users can be supported by providing indicators based on quality metrics. Therefore metric values for *readability* and *structuredness* have been presented to the user in a MediaWiki interface. The question was whether these indicators support improving quality. The result of this study shows that quality indicators stimulate qualitative improvement of information in Wiki articles. The fact, that the metric values were perceived as useful, implies that they are a real indicator for quality.

To answer the first research question, the metric framework developed in course of this work, is applicable to obtain measures for the quality of information. The conducted studies showed that these measures are an indicator for quality. Presenting the metric values to the user brings a real benefit for quality improvement. The resource model adopted from the theoretical foundation of information quality research facilitates representing the qualitative status of a resource.

2. How can the user requirements regarding quality be identified and represented?

One characteristic of this approach is that the definition of quality depends on user preferences. In other words, *if you want to support the user by providing quality-aware services, you have to know the user*. To achieve adaptivity regarding user preferences, the user must be modeled in the system. This work describes an approach how quality requirements for different tasks can be represented by a semantic model. This model includes the weighting of all quality dimensions. For populating the user-quality model, three approaches are described in this work. In the explicit approach, users are explicitly asked for their preferences regarding the quality of resources. For the implicit approach, user activities and ratings are evaluated in order to infer their preferences regarding quality. The third approach, the hybrid approach, is a combination of the two other approaches. The hybrid approach was not evaluated within this work. To investigate how user requirements can be identified and represented, the following two studies were conducted.

The study *Evaluation of task-specific quality profiles* (cf. Section 8.2) focused on capturing community-specific quality requirements. In this empiric approach, questionnaires were used to create quality profiles. The aim of the study was to identify similarities and differences in community-specific quality requirements. The result shows that there are similar quality requirements within a community of participants. This supports the assumption that there is a common understanding of information quality within a certain community. Furthermore, the result shows differences when it comes to the task dependence of quality dimensions. The relevance assessment indicated dimensions that are task-sensitive (*Representational*

Consistency, Traceability, *Variety of Data Sources*) and others that are (almost) independent from the task (*Reputation, Completeness*).

The study *Evaluation of profile similarity algorithm*, described in Section 8.3, focuses on the evaluation of recommendation mechanisms based on user quality profiles. One first step in this study was to determine the user requirements regarding quality. As part of this, I applied two ways for determining user quality requirements in a Wiki: explicitly by providing a questionnaire for giving a weight to each quality dimension, and implicitly by inferring quality requirements from article ratings. The result of the study regarding the user quality profile shows that the explicit and the implicit were almost similar. On the one hand this shows that the user is able to define own quality requirements explicitly. On the other hand it shows that the less intrusive way also leads to useful results. A drawback of this approach is that the social software system must be enhanced with additional functionality in both ways.

In order to answer the second research question, the goal was to find ways to determine and represent user quality requirements. A quality profile has been created and evaluated in both studies. The first study shows that quality requirements for different tasks can be defined in terms of quality dimensions. These dimensions are the basis for representing quality requirements in the user quality profile. The second study presents mechanisms (implicit/explicit) for identifying user quality requirements.

3. How can resources that correspond to the user requirements be found?

One important part of this work is the design, implementation and evaluation of mechanisms for measuring similarity between quality profiles. As addressed in the previous sub-questions, modeling quality of resources and quality requirements of individuals is an essential part of this work and a vital basis for any quality-aware system behavior. However, to provide quality adaptive functionality, services that incorporate quality models into their calculations are needed. An important functionality that provides the basis for quality adaptive services is measuring similarity between quality profiles. This work proposes an algorithm for recognizing similar quality patterns. Based on an existing approach, modifications for adapting the algorithm to a specific application case have been designed and implemented. The main focus of the design and implementation was applicability to real world environments. Concretely, this means being able to deal with the special characteristics of quality profiles, scalability to handle large datasets, and customizability regarding special user needs and application scenarios. The proposed algorithm is applicable to different application scenarios and can be used to identify:

- resources that are qualitatively similar to a given one
- resources that match certain quality criteria
- resources that have to be qualitatively improved
- resources that fit the task/requirements of the user
- users with similar quality requirements to any given user

The study *Evaluation of profile similarity algorithm*, described in Section 8.3 evaluates the algorithm intended for this purpose. In this Wiki-based study, quality profiles considered as similar by the users are tested on similarity by the algorithm. That means, the study evaluates whether the perceived similar information quality corresponds to the result of the algorithm. The result shows that the algorithm can identify clusters of similar profiles with sufficient accuracy, provided that the parameters are adapted to the application case. Based on these clusters containing resources of similar qualitative status, meaningful recommendations of recourses containing high quality information can be made.

The study *Evaluation of Quality Assistance in Social Media* (cf Section 8.4) evaluated the acceptance and perceived usefulness of quality-aware functionalities in social software systems. In a preparatory design study, the usage of metrics as quality indicators and the provision of quality aware recommendation mechanisms have been evaluated. Within the study mockups and design studies of quality-aware search mechanisms for providing quality-aware search result ranking were presented and evaluated. The qualitative evaluation showed that these functionalities, combined with existing search mechanisms, are beneficial to system users.

With regard to the third research question, one prerequisite is required for providing quality aware system behavior: an algorithm that is able to identify clusters of similar quality patterns. This algorithm is presented in my approach. The evaluation does not only highlight the ability to identify resources that correspond to the users' needs in terms of quality but also that quality aware resource recommendation provides a real benefit for the users.

DISCUSSION AND STARTING POINTS FOR FURTHER RESEARCH

This section discusses limitations of my approach and gives starting points for further research. According to the three main pillars of my work, this section is divided into these subsections as well

Measuring and representing the qualitative status of resources

I proposed an approach for measuring the qualitative status of resources using metrics for creating quality profiles as an abstraction from system representations. Although I strived to provide general solutions that are applicable to many different social software systems, the nature of system representations is such that it is system-specific. In this work, I presented the creation of system representations for MediaWiki systems. But creating this representation for another system may be completely different. Even though the step for creating the system representation is different, the result can be similar. The aim of my approach is to be able to apply metrics to system representations independent of the underlying system. Thus standardized system representations are required. Results have shown that reusing existing metrics leads to satisfying results. Due to the large amount of existing measures provided by different disciplines and areas of research, no development of additional metrics was required in this area. The selection of metrics cannot be generally defined as well. Metrics must be selected based on the application case, the expressivity of the system representation, and the favored quality dimensions. Exemplary approaches for the selection of metrics have been presented in this work but a general statement cannot yet be given.

Eliciting and representing the users' quality requirements

Two different methods for eliciting user quality requirements are presented in this approach: implicit and explicit. While the implicit method processes usage activities from the system in order to infer quality requirements, the explicit method requires the user to express his/her quality requirements explicitly within the system. That has the advantage of allowing the user to state his/her requirements with regard to a particular system. The drawback of this method is that the system must be modified with additional functionality to enable the user to state such requirements. This is usually associated with a major effort of system modification. Another point is that in the course of creating the user quality profile, lots of contextual information about the user must be detected, stored and processed. In other words, *the better the system gets to know the user, the better will be the results of quality aware functions*. But especially, collecting user data and user profiling often evokes concerns regarding privacy. Although I consider privacy as an important topic, addressing it in this context goes beyond the scope of my work but could well be a starting point for further research.

Identifying matches between the qualitative status of resources and users' quality requirements.

Calculating similarities between quality profiles plays a crucial role in my approach. I proposed an algorithm that provided sufficient results when calculating these similarities. Experiments have shown that the performance of the algorithm strongly depends on the set parameters. Which parameters perform best will depend on the characteristics of the dataset, e.g. the number of measured quality dimensions, variety of quality attributes, size of subset, etc. Basically, the parameterization is a tradeoff between the number of results and similarity between them. That means, the more results are returned, the lower is the similarity, and thus the recommendation is perceived to be less useful. Since the selection of parameters depends on the dataset, system, and end-users, no general statement can be made about what the best fitting parameters might be. They have to be adjusted according to the contextual characteristics.

Even if the aim of this work was to provide a generally applicable approach, the previous points show that some points are highly dependent on given factors. My approach gives case examples on these points, hence adjustments are necessary for other systems and application cases. The approach described in this work focuses mainly on Wikis as one (very common) social system. Algorithms, designs and evaluations address Wiki systems. A suggestion for further research may be to generalize the ideas of this work for other social systems like folksonomies or social networks. Furthermore, one aspect that supposedly influences the performance of this approach significantly is the selection of metrics. The focus of this work is not to design or develop metrics. But the use here of more sophisticated metrics with more complex data structures seems promising. Solutions presented in this work are mainly focused on corporate Wikis. Thereby, information about quality is assessed always in one single system. Usually, users use several social systems on the web. Through standardization of the information metadata, qualitative information could be exchanged and maintained across several systems.

LIST OF FIGURES

LIST OF TABLES

REFERENCES

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the international conference on Web search and web data mining - WSDM '08*, 183. New York, New York, USA: ACM Press. doi:10.1145/1341531.1341557

Ahn, T., Ryu, S., & Han, I. (2007). The impact of Web quality and playfulness on user acceptance of online retailing. *Information & Management*, *44*(3), 263-275. doi:10.1016/j.im.2006.12.008

Arazy, O., Croitoru, A., 2010. The sustainability of corporate wikis: A time-series analysis of activity patterns. *ACM Trans. Manage. Inf. Syst.* 1, 1, Article 6 (December 2010), 24 pages. DOI=10.1145/1877725.1877731

Atterer, R., Wnuk, M., & Schmidt, A. (2006). Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. *Proceedings of the 15th international conference on World Wide Web* (pp. 203–212). ACM.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi : An Open Source Software for Exploring and Manipulating Networks. *American Journal of Sociology*, 361-362. AAAI.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. (M. Jordan, J. Kleinberg, & B. Scholkopf, Eds.)*Pattern Recognition* (Vol. 4, p. 738). Springer. doi:10.1117/1.2819119

Blaschke, S., & Stein, K. (2008). Methods and measures for the analysis of corporate wikis: A case study. *Proceedings of the 58th Annual Conference of the International Communication ICA* (pp. 1-24).

Blumenstock, J. E. (2008). Size matters: word count as a measure of quality on wikipedia. *Proceeding of the 17th international conference on World Wide Web* (pp. 1095–1096). ACM.

Bobko, P. (2001). Applications of pearson correlation to measurement theory. *Correlation and regression Applications for industrial organization* (pp. 67-96). Sage Publications. ISBN: 9780761923039

Borgatti, S. P., & Everett, M. G. (2006). A Graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466-484. Elsevier.

Borodin, A., Roberts, G. O., Rosenthal, J. S., & Tsaparas, P. (2005). Link analysis ranking: algorithms, theory, and experiments. *ACM Transactions on Internet Technology*, *5*(1), 231-297. doi:10.1145/1052934.1052942

Bra, P. D., & Calvi, L. (1998). AHA: a Generic Adaptive Hypermedia System. *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia* (pp. 5-12), Pittsburgh, 1998

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(2), 163–177. Taylor & Francis.

Brandes, U., & Wagner, D. (2004). Analysis and visualization of social networks. *Graph drawing software* (pp. 321–340). Citeseer.

Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., & Marshall, M. S. (2002). GraphML Progress Report Structural Layer Proposal. *Lecture Notes in Computer Science*, *2265*, 501-512. Springer-Verlag. doi:10.1007/3-540-45848-4_59

Braun, S, & Schmidt, A. (2007). Wikis as a Technology Fostering Knowledge Maturing: What we can learn from Wikipedia. In K. Tochtermann & H. Maurer (Eds.), *7th International Conference on Knowledge* (pp. 321-329). Springer.

Braun, S., Zacharias, V., & Happel, H.-J. (2008). Social Semantic Bookmarking. *Practical Aspects of Knowledge Management 7th International Conference PAKM2008 Yokohama Japan*, *5345 LNAI*, 62-73.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, *6*(2-3), 87-129. doi:10.1007/BF00143964

Bødker, K., Kensing, F., & Simonsen, J. (2004). Participatory IT Design - Designing for Business and Workplace Realities. *The MIT Press, Cambridge, Massachusetts*.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). Models and methods in social network analysis. (P. J. Carrington, J. Scott, & S. Wasserman, Eds.) Cambridge Univ Pr, Year 2005.

Chidlovskii, B. (2010). Multi-label Wikipedia classification with textual and link features. *INEX. editor(s) Geva, Shlomo and Kamps, Jaap and Trotman, Andrew. Lecture Notes in Computer Science, (6203) 387-396, Springer, Year 2009.*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (L. Erlbaum, Ed.)*Statistical Power Analysis for the Behavioral Sciences* (Vol. 2, p. 567). Erlbaum.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *The International Journal of Computer-Supported Collaborative Learning*, *3*(2), 105-122. Springer New York. doi:10.1007/s11412-007-9035-z

Damme, C. V., & Coenen, T. (2001). Deriving a Lightweight Corporate Ontology form a Folksonomy: Methodology and its Possible Applications. *In Scalable Computing: Practice and Experience - Scientific International Journal for Parallel and Distributed Computing, editor(s) Flejter, Dominik and Kaczmarek, Tomasz and Kowalkiewicz, Marek. (9) 4: 293-301, Year 2008.*

Dekel, U., (2007). A framework for studying the use of wikis in knowledge work using client-side access data. In *Proceedings of the 2007 international symposium on Wikis* (WikiSym '07). ACM, New York, NY, USA, 25-30. DOI=10.1145/1296951.1296954

DeLone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60-95. INFORMS: Institute for Operations Research. doi:10.1287/isre.3.1.60

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of management information systems*, *19*(4), 9-30.

Dom, B., Eiron, I., Cozzi, A., & Zhang, Y. (2003). Graph-based ranking algorithms for e-mail expertise analysis. *In Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery (DMKD '03). ACM, New York, NY, USA, 42-48. DOI=10.1145/882082.882093*

Dongen, B. F. V., Medeiros, A. K. A. D., Verbeek, H. M. W., Weijters, A. J. M. M., & Van Der Aalst, W. M. P. (2005). The ProM framework: A new era in process mining tool support. (G. Ciardo & P. Darondeau, Eds.)*Applications and Theory of Petri Nets 2005*, *3536*(i), 444-454. Springer. doi:10.1007/11494744_25

FLESCH, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233. Elsevier.

Farahat, A., LoFaro, T., Miller, J. C., Rae, G., & Ward, L. a. (2006). Authority Rankings from HITS, PageRank, and SALSA: Existence, Uniqueness, and Effect of Initialization. *SIAM Journal on Scientific Computing*, *27*(4), 1181. doi:10.1137/S1064827502412875

Francalanci, C., & Pernici, B. (2004). Data quality assessment from the user's perspective. *Proceedings of the 2004 international workshop on Information quality in information systems* (pp. 68–73). ACM.

Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215-239. doi:10.1016/0378-8733(78)90021-7

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. Nature Publishing Group.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, *36*(2), 193-202.

Gunning, R. (1952). *The technique of clear writing*. New York, NY: McGraw-Hill International Book Co.

Hage, P., & Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks*, *17*(1), 57-63. doi:10.1016/0378-8733(94)00248-9

Hagedoorn, M., (2000). *Pattern matching using similarity measures*. PhD thesis, Utrecht University, the Netherlands, 2000. 2.

Hearst, M. a. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th conference on Computational linguistics* -, 539. Morristown, NJ, USA: Association for Computational Linguistics. doi:10.3115/992133.992154

Herings, P.J.J. & Laan, G. van der & Talman, A.J.J., (2001).Measuring the Power of Nodes in Digraphs, Discussion Paper 2001-72, Tilburg University, Center for Economic Research.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Information retrieval in folksonomies: Search and ranking. The Semantic Web: Research and Applications. editor(s) Sure, York and Domingue, John. Lecture Notes in Computer Science, (4011) 411-426, Springer, Heidelberg, Year 2006.

Hu, M., Lim, E. P., Sun, A., Lauw, H. W., & Vuong, B. Q. (2007). Measuring article quality in wikipedia: models and evaluation. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 243–252). ACM.

Iivari, J. (1986). An innovation research perspective on information system implementation. *International Journal of Information Management*, *6*(3), 123-144. doi:10.1016/0268-4012(86)90001-0

Ivanov, K. (1972). Quality-control of information: On the concept of accuracy of information in data banks and in management information systems. The University of Stockholm and The Royal Institute of Technology. Doctoral dissertation.

Jaksch, B., Kepp, S.-J., & Womser-Hacker, C. (2008). Integration of a Wiki for Collaborative Knowledge Development in an E-Learning Context for University Teaching. (A. Holzinger, Ed.)*Education*, *LNCS 5298*, 77-96. Springer.

Jeon, J., Croft, W. B., Lee, J. H., & Park, S. (2006). A framework to predict the quality of answers with non-textual features. *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 228--235, ACM Press, New York, NY, USA, Year 2006.*

Juran, J. M. (1992). *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services* (p. VI + 538). The Free Press.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM*, *45*(4), 184–192. ACM.

Kamps, J., & Koolen, M. (2008). The importance of link evidence in Wikipedia. ECIR. editor(s) Macdonald, Craig and Ounis, Iadh and Plachouras, Vassilis and Ruthven, Ian and White, Ryen W.. Lecture Notes in Computer Science, (4956) 270-282, Springer, Year 2008.

Katerattanakul, P., & Siau, K. (1999). Measuring information quality of web sites: development of an instrument. *Proceedings of the 20th international conference on Information Systems* (Vol. pp, pp. 279–285). Association for Information Systems.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. *Proceedings of the ACM 2008 conference on Computer supported cooperative work* (pp. 37–46). ACM.

Klein, B.D. (1999) Perceptions of Information Quality: A Study of Internet and Traditional Text Sources,*Proceedings of the Fifth Americas Conference on Information Systems*, 1999, 618-620.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604–632. ACM New York, NY, USA.

Klieber, W., Sabol, V., Muhr, M., Kern, R., Öttl, G., & Granitzer, M. (2009). Knowledge discovery using the KnowMiner framework. In: IADIS*, Vol. SingleIADIS International Conference Information Systems 2009* (2009) , p. 307--314.

Knight, S.-ann, & Burn, J. (2005). Developing a Framework for Assessing Information Quality on the World Wide Web Introduction – The Big Picture What Is Information Quality ? *Informing Science Journal*, 8 (2005), pp. 160–172

Kohl, M., Wiese, S., & Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Methods In Molecular Biology Clifton Nj*, *696*, 291-303. Humana Press.

Lempel, R., & Moran, S. (2001). SALSA: the stochastic approach for link-structure analysis. *ACM Transactions on Information Systems*, *19*(2), 131-160. doi:10.1145/382979.383041

Ley, T., Kump, B., & Albert, D. (2010). A methodology for eliciting, modelling, and evaluating expert knowledge for an adaptive work-integrated learning system. *International Journal of Human-Computer Studies*, *68*(4), 185-208. Elsevier. doi:10.1016/j.ijhcs.2009.12.001

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*(9), 1303-1319. doi:10.1002/asi.20614

Lih, A. (2004). Wikipedia as Participatory Journalism : Reliable Sources ? Metrics for evaluating collaborative media as a news resource. *In Proceedings of the 5th International Symposium on Online Journalism*. 16--17, Year 2004.

Lim, E. P., Vuong, B. Q., Lauw, H. W., & Sun, A. (2006). Measuring qualities of articles contributed by online communities. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 81–87). IEEE Computer Society.

Lindstaedt, S N, Scheir, P., Lokaiczyk, R., Kump, B., Beham, G., & Pammer, V. (2008). Knowledge Services for Work-integrated Learning. (P. Dillenbourg & M. Specht, Eds.)*Times of Convergence Technologies Across Learning Contexts*, 234-244. Springer.

Lindstaedt, S N, & Christl, C. (2011). APOSDLE – learn@work: Firsthand Experiences and Lessons Learned. In P. Keleher, K. Klinger, L. Johnston, J. Gamon, N. Pronio, M. Vracarich, & N. Newcomer (Eds.), *WorkIntegrated Learning in Engineering Built Environment and Technology Diversity of Practice in Practice* (pp. 52-71). IGI-Global.

Lindstaedt, S N, Beham, G., Kump, B., & Ley, T. (2009). Getting to Know Your User–Unobtrusive User Model Maintenance within Work-Integrated Learning Environments. EC-TEL. editor(s) Cress, Ulrike and Dimitrova, Vania and Specht, Marcus. Lecture Notes in Computer Science, (5794) 73-87, Springer, Year 2009.

Liu, J., & Ram, S. (2009). Who does what: Collaboration patterns in the wikipedia and their impact on data quality. *Proceedings of nineteenth Annual Workshop on Information Technologies and Systems (WITS 2009)*. Phoenix, Arizona, USA, Year 2009.

MATURE Consortium. (2009a). D1.1 Results of the Ethnographic Study and Conceptual Knowledge Maturing Model. Retrieved from http://mature-ip.eu/files/deliverables/D1.1_Ethnographic_Studies_Knowledge_Maturing_Model.pdf

MATURE Consortium. (2009b). D4.1 Maturing Services Definition. Retrieved from http://mature-ip.eu/files/deliverables/D4.1_Maturing_Services.pdf

MATURE Consortium. (2009c). D3.1 Model of organizational requirements and of supporting services of the OLME. Retrieved from http://mature-ip.eu/files/deliverables/D3.1_OLME_Requirements_Services.pdf

MATURE Consortium. (2010). D1.2 Results of the representative study and refined conceptual knowledge maturing model. Retrieved from http://mature-ip.eu/files/deliverables/D1.2_Representative_Study_KMM.pdf

Madadhain, J., Fisher, D., Smyth, P., White, S., & YB. (2005). Analysis and visualization of network data using JUNG. In Journal of Statistical Software, (10) 1--35, Citeseer, Year 2005.

Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate wiki users: results of a survey. *Proceedings of the 2006 international symposium on Wikis*, *104*(September 2006), 99-104. ACM. doi:10.1145/1149453.1149472

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of reading*, *12*(8), 639–646. JSTOR.

Moskaliuk, J., Rath, A., Weber, N., Lindstaedt, S., & Cress, U. (2011). Automatic detection of accommodation steps as an indicator of knowledge maturing. *Interacting with Computers*. Year 2011.

Muller, M. J. (2003). Participatory design: the third space in HCI. In A. Sears & J. A. Jacko (Eds.), *Communications of the ACM* (Vol. 4235, pp. 1-32). Lawrence Erlbaum Associates, Inc. doi:10.1145/153571.255960

NWBTeam. (2009). Network Workbench Tool. *Social Science*, 1-77. Indiana University, Northeastern University, and University of Michigan.

Najjar, J., Duval, E., & Wolpers, M. (2006). Attention metadata: Collection and management. *WWW2006 workshop on logging traces of web activity: the mechanics of data collection* (pp. 1–4).

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Library Technologies Project*. Stanford InfoLab.

Pammer, V., & Lindstaedt, S N, (2009). Ontology evaluation through assessment of inferred statements: Study of a prototypical implementation of an ontology questionnaire for owl dl ontologies. *Knowledge Science, Engineering and Management*, 394–405. Springer.

Peters, I., & Weller, K. (2008). Tag gardening for folksonomy enrichment and maintenance. *Webology*, *5*(3), 1-23.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. ACM.

Potthast, M. (2010). Crowdsourcing a Wikipedia Vandalism Corpus. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '10). ACM, New York, NY, USA, 789-790. DOI=10.1145/1835449.1835617

Potthast, M., Stein, B., & Gerling, R. (2008). Automatic vandalism detection in Wikipedia. *Proceedings of the IR research, 30th European conference on Advances in information retrieval* (Vol. 4956, pp. 663–668). Springer-Verlag.

Qian, G., Sural, S., Gu, Y., & Pramanik, S. (2004). Similarity between Euclidean and cosine angle distance for nearest neighbor queries. *Proceedings of the 2004 ACM symposium on Applied computing - SAC '04*, 1232. New York, New York, USA: ACM Press. doi:10.1145/967900.968151

Ram, S., & Liu, J. (2007). Understanding the semantics of data provenance to support active conceptual modeling. In *Active conceptual modeling of learning*, Peter P. Chen and Leah Y. Wong (Eds.). Lecture Notes In Computer Science, Vol. 4512. Springer-Verlag, Berlin, Heidelberg 17-29.

Rath, A. S, Weber, N., Kröll, M., Granitzer, M., Dietzel, O., & Lindstaedt, S. N. (2008). Context-aware knowledge services. *Proceedings of Computer Human Interaction (CHI 2008), Workshop on Personal Information Management: PIM 2008*. Year 2008

Rath, A.S., Devaurs, D., & Lindstaedt, S. N. (2009). UICO: an ontology-based user interaction context model for automatic task detection on the computer desktop. *Proceedings of the 1st Workshop on Context, Information and Ontologies* (p. 8). ACM.

Reeves, C. A., & Bednar, D. A. (1994). Defining Quality: Alternatives and Implications. *Academy of Management Review*, *19*(3), 419. doi:10.2307/258934

Reinhold, S. (2006). WikiTrails: augmenting Wiki structure for collaborative, interdisciplinary learning. In D. Riehle & J. Noble (Eds.), *WikiSym 06 Proceedings of the 2006 international symposium on Wikis* (Vol. 2006, pp. 47-58). ACM Press. doi:10.1145/1149453.1149467

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581-603. Springer New York. doi:10.1007/BF02289527

Schmidt, A., (2005). Knowledge maturing and the continuity of context as a unifying concept for knowledge management and e-learning. *Proceedings of I-KNOW 05*, Graz, Austria. Year 2005

Sharoff, S., Kurella, S., & Hartley, A. (2008). Seeking needles in the web haystack: Finding texts suitable for language learners. *Proceedings of the 8th Teaching and Language Corpora Conference (TaLC-8). Lisbon, Portugal*.

Shimbel, A. (1953). Structural Parameters of Communication Networks. *Bulletin of Mathematical Biology*, *15*(1), 501-507.

Si, L., & Callan, J. (2001). A statistical model for scientific readability. *CIKM* (Vol. 1, pp. 574-576). ACM Press. doi:10.1145/502585.502695

Smets, K., & Goethals, B. (2008). Automatic vandalism detection in Wikipedia: Towards a machine learning approach. *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy* (WikiAI08). 43--48, AAAI Press, Year 2008.

Strong, D. M., & Lee, Y. W. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103–110. ACM.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, *58*(12), 1720-1733. doi:10.1002/asi.20652

Stvilia, B., Twidale, M. B., & Gasser, L. (2005). Information quality discussions in Wikipedia. *Knowledge Management: Nurturing Culture, Innovation, and Technology - Proceedings of the 2005 International Conference on Knowledge Management*. editor(s) Hawamdeh, S.. 101-113, Year 2005

Suebsing, A., & Hiransakolwong, N. (2009). Feature Selection Using Euclidean Distance and Cosine Similarity for Intrusion Detection Model. *2009 First Asian Conference on Intelligent Information and Database Systems*, 86-91. Ieee. doi:10.1109/ACIIDS.2009.23

Udell, B. J. (2010). Collaborative knowledge gardening. *Source*, 1-2.

Voss, J. Measuring Wikipedia, 2005. In International Conference of the International Society for Scientometrics and Informetrics : 10th,Stockholm (Sweden),24-28 July 2005.

Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R. (2006). Semantic Wikipedia. Proceedings of the 15th international conference on World Wide Web WWW 06 5 (January 2001) p. 585

Wand Y., &Wang R. Y., 1996. Anchoring data quality dimensions in ontological foundations. *Commun. ACM* 39, 11 (November 1996), 86-95. DOI=10.1145/240455.240479

Wang, H., Wang, W., Yang, J., & Yu, P. S. (2002). Clustering by pattern similarity in large data sets. *Proceedings of the 2002 ACM SIGMOD international conference on Management of data - SIGMOD '02*, *2*, 394. New York, New York, USA: ACM Press. doi:10.1145/564691.564737

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), 33. ME Sharpe, Inc.

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, *7*(4), 623-640. doi:10.1109/69.404034

Wasson, J. (2008). Pearson Product Moment Correlation Coefficient. Minnesota State University.

Weber, N., Nelkner, T., Schoefegger, K., Lindstaedt, S. N.. SIMPLE - a social interactive mashup PLE . *In CEUR Workshop Proceedings, Proceedings of the Third International Workshop on Mashup Personal Learning Environments* (MUPPLE09), in conjunction with the 5th European Conference on Technology Enhanced Learning (EC-TEL2010). editor(s) Wild, Fridolin and Kalz, Marco and Palmér, Matthias and Müller, Daniel. Barcelona, Spain, Year 2010..

Weber, N., Schoefegger, K., Bimrose, J., Ley, T., Lindstaedt, S., Brown, A., & Barnes, S. A. (2009). Knowledge maturing in the semantic mediawiki: A design study in career guidance. *Learning in the Synergy of Multiple Disciplines*, 700–705. Springer.

Weber, N., Frühstück, G., & Ley, T. (2011). Unterstützung des Wissensreifungsprozesses durch Einsatz von Web 2.0 in Unternehmen. *Tagungsband zum Workshop E20Success: Enterprise 2.0 – Mehr Erfolg mit Web 2.0 im Unternehmen*. Conference on Professional Knowledge Management 2011, Innsbruck, Austria

Weller, K., & Peters, I. (2008). Seeding, weeding, fertilizing. different tag gardening activities for folksonomy maintenance and enrichment. (S. Auer, S. Schaffert, & T. Pellegrini, Eds.) *Proceedings of I-Semantics '08, International Conference on Semantic Systems*. Graz, Austria, September 3-5, S. 110-117

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. (J. S. Brown & J. Hawkins, Eds.)*Learning in doing* (p. 318). Cambridge University Press.

Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking Actual Usage : the Attention Metadata Approach. *Educational Technology & Society*, *10*, 106-121.

Wöhner, T., & Peters, R. (2009). Assessing the quality of Wikipedia articles with lifecycle based metrics. *Proceedings of the 5th International Symposium on Wikis and Open Collaboration - WikiSym '09*, 1. New York, New York, USA: ACM Press. doi:10.1145/1641309.1641333

Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: structure and algorithms. *Proceedings of the 16th international conference on World Wide Web* (Vol. pages, p. 230). ACM.

Zhang, X. (2006). Intrinsic motivation of open content contributors: The case of wikipedia. *Workshop on Information Systems and Economics (WISE)*, Dec. 2006, Chicago, IL.

Zhu, X., & Gauch, S. (2000). Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '00* (pp. 288-295). New York, New York, USA: ACM Press. doi:10.1145/345508.345602