MODELING AND ACQUIRING KNOWLEDGE ABOUT HUMAN GOALS

Dissertation in fulfilment of the requirements for the academic degree
Doctor of Technical Sciences (Dr.techn.) in Computer Science

at the

Graz University of Technology

submitted by

Mark Kröll

Knowledge Management Institute

Graz, Austria, October 2012

First Reader:	Univ.-Doz. Dr. Markus Strohmaier
Second Reader:	Univ.-Prof. Dr. Michael Granitzer

Graz University of Technology

Abstract

Realizing the vision of machines behaving intelligently requires the acquisition of real-world knowledge including, but not limited to, knowledge about human goals. Understanding human goals can help (i) to answer "why" questions about user behavior and user interactions, (ii) to reason about people's goals or (iii) to generate action sequences that implement goals (plan generation). Structuring and organizing (human) goal knowledge has recently been gaining traction in several areas including Commonsense Knowledge Acquisition, Linguistics or Commonsense Psychology. Yet, each research area appears to have developed its own understanding of (human) goal knowledge which is reflected in their definitions and models. A consequence of this specialization is that it becomes increasingly difficult to analyze, evaluate or compare goal knowledge across research prototypes or projects.

The theoretical part of this work introduces an intentional framework which structures and describes human goal knowledge. This framework seeks to create a common perspective on human goal knowledge. The framework's development is based on and synthesizes existing work on modeling goal knowledge from a number of distinct, yet interrelated research domains including Goal-Oriented Requirements Engineering, Commonsense Knowledge Acquisition and Information Retrieval. The framework illustrates and elaborates on the engineering process as a whole including modeling, extracting and representing knowledge about human goals.

In the practical part of this work, three case studies are conducted which instantiate different parts of the framework and thus cover different aspects of goal knowledge. Case studies as well as application scenarios address the automatic acquisition of human goal knowledge and demonstrate its value in practical problem settings such as (i) analysing textual resources from an intentional perspective or (ii) complementing commonsense knowledge bases. This PhD thesis' theoretical and practical contributions will serve as a starting point for knowledge engineers interested in modeling and in acquiring human goal knowledge and will hopefully enable them to build systems that exhibit higher levels of awareness and representation of human goals.

Kurzfassung

Um die Vision von sich intelligent verhaltenden Maschinen zu realisieren, bedarf es der Akquise von Wissen über die reale Welt - inklusive Wissen über die Ziele der Menschen. Das Verständnis menschlicher Ziele kann dabei behilflich sein (i) "Warum"-Fragen über Benutzerverhalten und Benutzerinteraktionen zu beantworten, (ii) Schlußfolgerungen über menschliche Ziele zu ziehen, oder (iii) Sequenzen von Aktionen zu generieren, die der Zielerreichung dienen (Planerzeugung). Menschliches Zielwissen zu strukturieren und zu organisieren gewinnt zunehmend an Bedeutung in mehreren Forschungsgebieten wie z.B. der Akquise von Allgemeinwissen, der Linguistik oder der Alltagspsychologie. Jedes Forschungsgebiet scheint jedoch ein eigenes Verständnis von menschlichem Zielwissen entwickelt zu haben, was sich in deren Modellen und Definitionen wiederspiegelt. Diese Spezialisierung erschwert es, Zielwissen über Forschungsprototypen oder Forschungsgruppen hinweg zu analysieren, zu evaluieren oder zu vergleichen.

Im theoretischen Teil dieser Arbeit wird ein generelles Framework entwickelt, das menschliches Zielwissen strukturiert und abbildet. Dieses Framework versucht eine gemeinsame Perspektive über menschliches Zielwissen zu schaffen. Die Entwicklung des Frameworks basiert einerseits auf existierenden Arbeiten, Zielwissen zu modellieren, und anderseits synthetisiert es diese. Diese Arbeiten umfassen eine Reihe unterschiedlicher, jedoch miteinander in Beziehung stehender Forschungsbereiche wie z.B. zielorientierte Anforderungserhebung, Akquise von Alltagswissen oder Informationswiedergewinnung. Das vorgestellte Framework beinhaltet zudem eine detaillierte Beschreibung des gesamten Akquiseprozesses inklusive der Modellierung, Extraktion und Repräsentation von Wissen über menschliche Ziele.

Im praktischen Teil dieser Arbeit werden drei Fallstudien durchgeführt, die unterschiedliche Teile des Frameworks instantiieren und somit unterschiedliche Aspekte von Zielwissen behandeln. Im Zuge dieser Fallstudien wird die automatische Akquise von menschlichem Zielwissen addressiert sowie dessen Wert in praktischen Anwendungsszenarien demonstriert wie z.B. (i) dem Analysieren von textuellen Ressourcen aus einer intentionalen Perspektive sowie (ii) dem Erweitern von bereits vorhandenen Wissensbasen. Theoretische und praktische Beiträge dieser Dissertation werden Wissensingenieuren behilflich dabei sein, menschliches Zielwissen zu modellieren und zu akquirieren. Es ist zu hoffen, dass dieses Wissen sie in die Lage versetzt, Systeme zu bauen, die sich in höherem Ausmaß menschlicher Ziele bewußt sind.

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)

Dedication

To my parents for their never-ending love and support.

To Agnes who never stopped believing in me.

To my fellow PhD students and Know-Center colleagues for hours of honest and supportive discussions (in alphabetical order): Albert Bifet, Roman Kern, Christian Körner, Johannes Liegl, Markus Muhr, Peter Prettenhofer, Andreas S. Rath, Vedran Sabol, Monika Schubert, Christin Seifert and Nicolas Weber.

To my co-authors (in alphabetical order): Yusuke Fukazawa, Fleur Jeanquartier, Christian Körner, Jun Ota, Peter Prettenhofer and Markus Strohmaier.

To the Know-Center for their support during the last years.

To Andrew S. Gordon for an unforgettable stay and for showing me the lightness of research.

To Michael Granitzer for his valuable advice and for his continuous support.

And to Markus Strohmaier, my supervisor, for teaching me the art of research, for guiding me through the process of writing a thesis and for providing me with valuable feedback to continuously improve this thesis.

Thank you.

Mark Kröll
Graz, Austria, October 2012

Contents

List of Figures

List of Tables

PART I: INTRODUCTION

Chapter 1

Introduction

1.1 Motivation

In [110], Marvin Minsky lays out a vision of machines that are capable of behaving intelligently: "*...if we want our computers to understand us, we will need to equip them with adequate knowledge. Only then can they become truly concerned with our human affairs. To help us work, they must know what our jobs are. To entertain us they'll need to know what their audiences like or need.*" Realizing this vision requires the acquisition of real-world knowledge including, but not limited to, knowledge about human goals. Goal knowledge has been acknowledged as an asset in various research and application domains, for instance, to quote Henry Lieberman [101]: "*In our research on Commonsense reasoning, we have found that an especially important kind of knowledge is knowledge about human goals. Especially when applying Commonsense reasoning to interface agents, we need to recognize goals from user actions (plan recognition), and generate sequences of actions that implement goals (planning).*"

A better understanding of humans and their goals can help (i) to predict user behavior and user interactions, e.g. answering why questions (cf. [157]), (ii) to reason about people's goals or (iii) to generate action sequences that implement goals (planning). Figure 1.1 illustrates a toy example of knowledge about the human goal "`lose weight`" and serves to familiarize the reader with (i) this work's understanding of goal knowledge and (ii) the kind of knowledge it seeks to model and acquire. As Figure 1.1 indicates, human goal knowledge covers a wide spectrum not only consisting of human goals but of other intentional components as well including means or resources. Various types of relations connect goal components with each other thereby organizing and structuring them. Reasoning and planning mechanisms build upon this structural information to process goal knowledge.

Figure 1.2 illustrates (i) top-down or (ii) bottom-up reasoning by a simple toy example where knowledge on the human goal "`lose weight`" has been modeled. Means-Ends relations indicate either a positive (helps) or a negative (hurts) contribution. Means in this example also affect other goals such as "`achieve quick success`" and "`increase body awareness`". In top-down reasoning, the impact of certain means on other goals such as "`achieve quick success`" or "`increase`

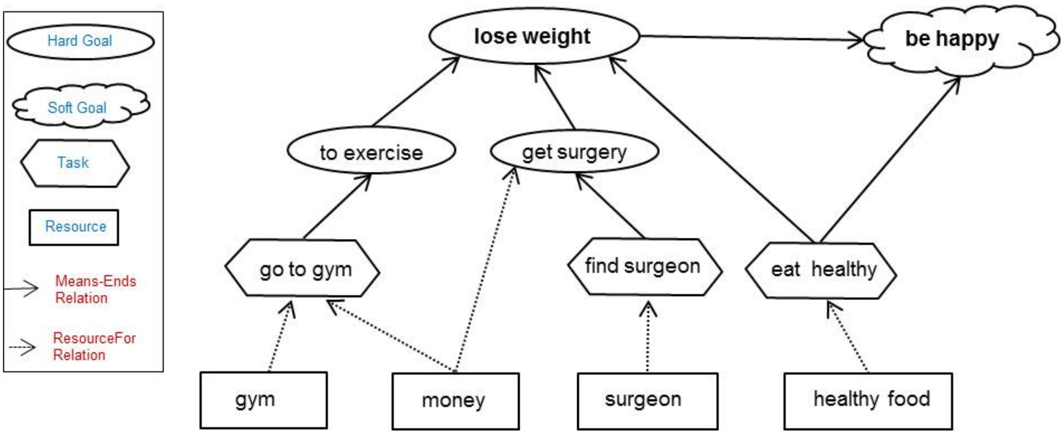


Figure 1.1: An example of a toy goal graph encoding knowledge about the goal "`lose weight`" represented in i* notation [188].

`body awareness`" is taken into account during reasoning, i.e. only those paths are selected which positively contribute to simultaneous (soft) goals the person might have. In contrast, bottom-up reasoning starts from leaf levels, e.g. "`get surgery`", traverses the hierarchy upwards and examines the impact of alternatives on soft goals, e.g. "`increase body awareness`".

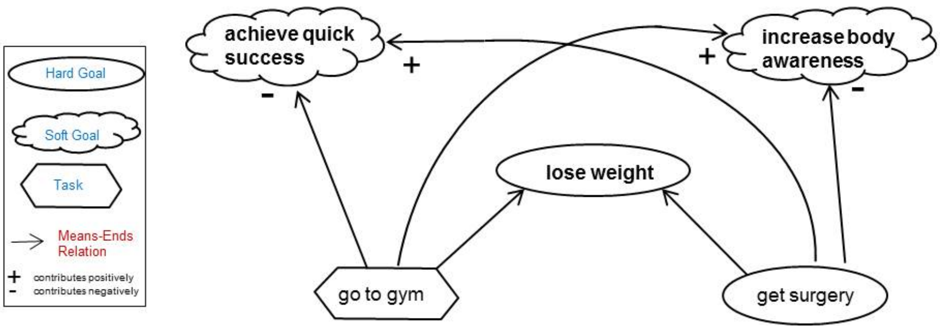


Figure 1.2: A simple toy example which encodes knowledge about the human goal "`lose weight`" (represented in i* notation [188]). Means-Ends relations indicate either a positive (helps) or a negative (hurts) contribution. This information can be utilized in (i) top-down or (ii) bottom-up reasoning.

A major rationale behind engineering human goal knowledge thus is to make goal knowledge utilizable and accessible by machines. Consequently, a lot of research has been dedicated to model and acquire knowledge about human goals across many domains including commonsense knowledge acquisition (cf. [101], [92]), context or situation awareness (cf. [77]), information retrieval (cf. [21], [146]), semantic task retrieval (cf. [120]), linguistics (cf. [149]), commonsense psychology (cf. [68]) or psychology (cf. [30]). Requirements on goal knowledge vary from area to area so that it appears natural that these research areas have developed their own understanding of human goal knowledge. This understanding concerns the richness of goal models or the extent to which goal knowledge has been made explicit.

At the same time, a common understanding (ontology) of human goal knowledge certainly would be beneficial for analyzing, evaluating or comparing human goal knowledge across research prototypes, research groups or research domains. Figure 1.3 sketches the general process of engineering knowledge taken from related work (cf. [105], [165] or [166]). This PhD thesis analyzes and specifies this process with respect to engineering human goal knowledge. Required steps to explicate knowledge about human goals include getting familiar with the domain's understanding of goal knowledge in the elicitation step and keeping goal knowledge operable in the operationalization step.

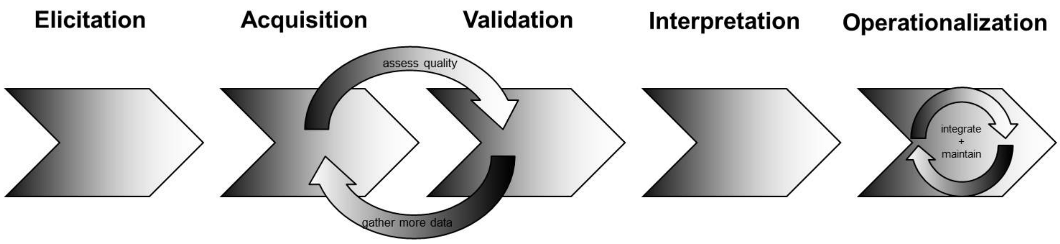


Figure 1.3: The general process on engineering knowledge. (based on Sure's [166] Knowledge Meta Process)

The traditional way of engineering knowledge was based on manual participation. In the elicitation step, for example, knowledge engineers conducted interviews with domain experts. The resulting protocols then served as data for the subsequent acquisition step. Yet, continuous growth of available data made a manual approach tedious and time-consuming. With more data at hand in all engineering steps, the focus shifted towards approaches exhibiting a reduced amount of human participation (cf. [39], [105]). Engineering knowledge about human goals faces similar challenges. With the advent of the social web, unprecedented opportunities opened up for constructing broad knowledge bases containing human goal knowledge. According to Lenhart [94] people tend to share information about their lives including their goals on the social web, making it a viable source of data. The social web facilitates accessing and gathering decent amounts of resources about human goal knowledge. Exemplary websites include “43Things.com”, “jig.com” or “goalmigo.com”. Publishing platforms such as weblogs or twitter have become an accepted means of communication for people exchanging their ideas, thoughts and experiences. Acceptance of social web applications therefore supports the creation of large repositories. Faced with these large amounts of data, automating the goal knowledge engineering process appears to be a promising problem to address.

1.2 Research Objectives

The previous section motivated human goal knowledge, i.e. its value, its applicability and thus its acquisition. Although this thesis is meant to serve as a starting point for knowledge engineers to model, acquire and represent knowledge about human goals, a detailed analysis of the entire engineering process is out of scope. This work addresses selected aspects related to the modeling and acquisition of human goal knowledge. Respective rationals are presented in the following two subsections.

1.2.1 Ontological Aspects of Human Goal Knowledge

When reviewing related literature, it can be observed that knowledge about (human) goals is omnipresent in several research areas including commonsense knowledge acquisition, linguistics or information retrieval. Yet, each research area appears to have developed its own understanding of human goal knowledge which is reflected in their definitions and models. It is often the case that these definitions and models are implicit and that they were devised to comply with specific research questions and particular applications in their respective area. A consequence of this diversity is that it becomes increasingly difficult to analyze, evaluate or compare goal knowledge across research prototypes or research groups. A unified view on human goal knowledge would contribute to standardize the engineering process which is sketched in Figure 3. To bring research closer to this unified view, the first core contribution of this PhD thesis is stated as follows:

> *This thesis develops a framework that is general enough to provide a common perspective on knowledge of human goals.*

To explicate and formalize knowledge about human goals, this PhD thesis proposes a framework which consists of (i) a data model which organizes intentional components and intentional relations and (ii) a meta-process which describes required steps to make human goal knowledge operable and thus useful. In brief, the framework specifies the encoding of human goal knowledge and seeks to represent a common perspective on human goal knowledge. To that end, this work synthesizes and bases its development on existing work on modeling goal knowledge from a number of distinct, yet interrelated research domains including goal-oriented requirements engineering (cf. [187], [41] or [119]), intelligent agent theory (cf. [18], [19] or [174]) or commonsense knowledge acquisition (cf. [101]). Particularly, in the requirements engineering (RE) domain, goals and goal-related components are utilized to inform the process of designing systems. To formalize goal knowledge, this work thus taps into RE's rich corpus of techniques.

1.2.2 Acquisitional Aspects of Human Goal Knowledge

To demonstrate the value and potential of human goal knowledge in practical problem settings, three case studies are conducted. In course of these case studies, techniques are explored and developed to automate the acquisition of human goal knowledge from textual resources. Efforts on automating the acquisition process appear necessary given the continuously growing amount of available data (cf. [39]). To advance these efforts, the second core contribution of this PhD thesis is stated as follows:

> *This thesis implements various aspects of the proposed framework in practical problem settings and thus operationalizes knowledge about human goals.*

Automating the process addresses (i) the goal acquisition problem (or bottleneck), which refers to the costs associated with knowledge acquisition [101] and (ii) the goal coverage problem, which refers to the difficulty of capturing the tremendous variety and range in the set of human goals [43].

In addition, the case studies contribute to illustrating the versatility of goal knowledge (i) across research domains such as commonsense knowledge acquisition or textual analysis, and (ii) across textual resources ranging from resources exhibiting poor grammatical structures, e.g. search queries, to resources with rich grammatical structures such as political speeches or blog posts.

1.2.3 Structure of this Thesis

This PhD thesis is composed of 9 chapters and is divided into four parts - Part I: Introduction, Part II: Framework, Part III: Case Studies & Applications and Part IV: Conclusion. Figure 1.4 illustrates this thesis' structure and chapter interrelationsships.

Chapter 1 introduces the concept of human goal knowledge by sketching its value and its potentials. The chapter thereby sets the frame for this thesis' objectives and contributions.

Chapter 2 surveys and synthesizes related literature including modeling of (human) goal knowledge and acquiring (human) goal knowledge.

Chapter 3 introduces the proposed framework on human goal knowledge, i.e. (i) the data model which encodes knowledge about human goals and (ii) the meta-process which characterizes the engineering process of goal knowledge. The data model's development is based on and synthesizes existing work from other research areas. The meta-process is designed as a special case of the general knowledge engineering process, yet focusing on aspects which are distinctive for human goal knowledge. Discussing potential operations on goal knowledge concludes this chapter.

Chapters 4 to 6 represent three case studies which seek to operationalize knowledge about human goals in various problem settings. To do that, each case study instantiates and covers selected parts of the proposed framework.

Chapter 4 presents a case study on acquiring human goal knowledge from search query logs. In this study, a classification algorithm is devised which achieves useful precision/recall values in extracting human goal instances from two large search query logs, recorded by AOL and Microsoft Research in 2006. To learn more about characteristics and distribution of these goal instances, qualitative, quantitative and comparative analyses are conducted.

Chapter 5 presents a case study on analyzing textual resources from a goal-oriented perspective. Political speeches given by the two American presidential candidates in 2008 are categorized into a human goal taxonomy. Emerging goal profiles then allow a comparison of these speeches and thus a comparison of the presidential candidates' aspirations with respect to intentional aspects.

Chapter 6 seeks to automatically construct a hierarchy of goal concepts, i.e. infering hierarchical structures by applying unsupervised machine learning techniques. To analyze and compare resulting concept hierarchies, a ground truth is generated which allows calculating taxonomic overlaps. An algorithmic approach is presented to automatically complement goal concept

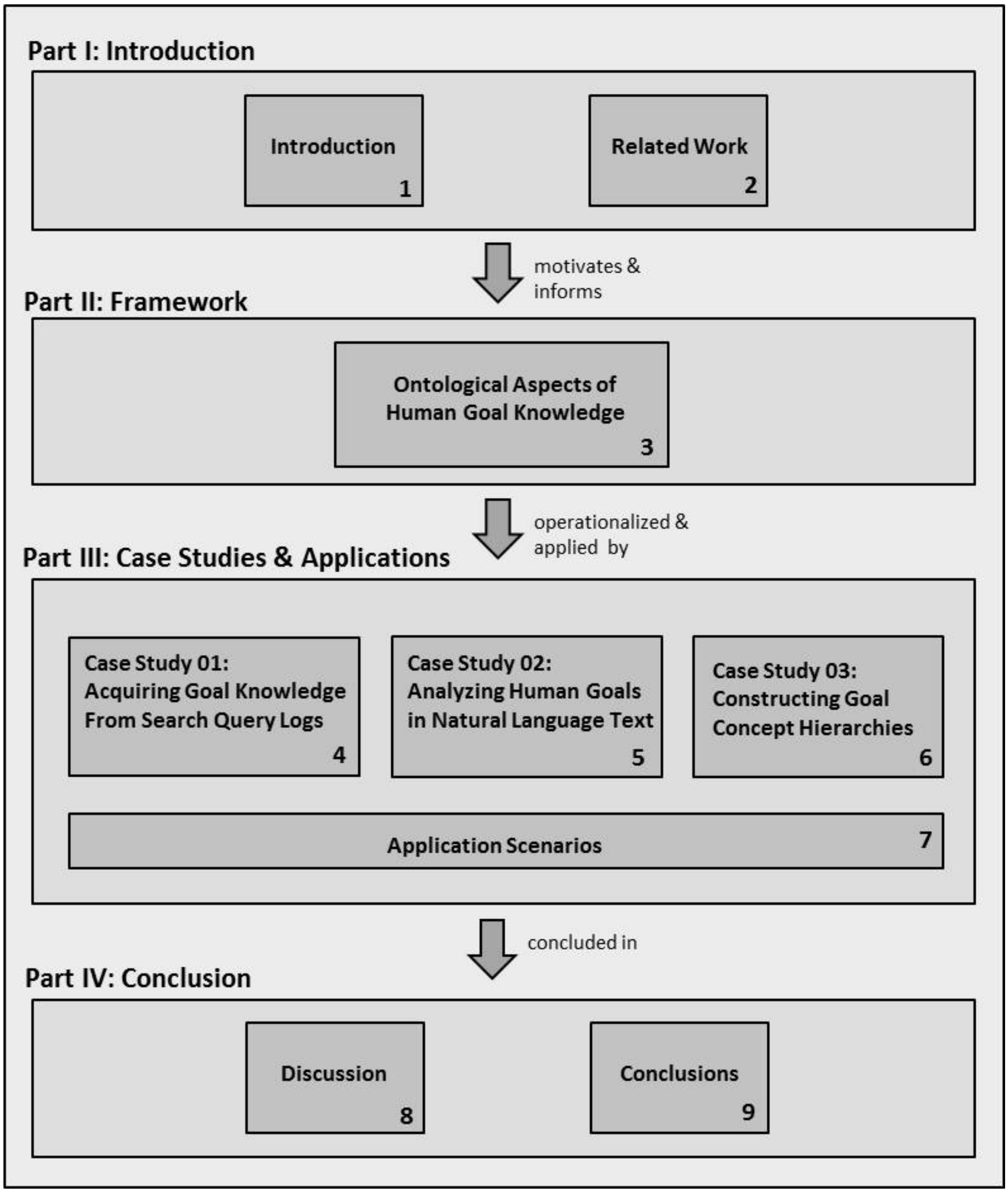


Figure 1.4: A structural overview of this PhD thesis' 9 chapters.

hierarchies by Means-Ends relations, i.e. relating goals to tasks which potentially contribute to their accomplishment.

Chapter 7 encompasses various application scenarios which demonstrate the practical applicability and usefulness of human goal knowledge. The scenarios present theoretical as well practical approaches illustrating benefits of formalizing human goal knowledge. One scenario, for instance, explores methods to complement commonsense knowledge bases, in this case ConceptNet [104]. Another scenario introduces intentional query suggestion, which encourages users to make their search goal more explicit (e.g. "`buy a car`") rather than formulating their query in a rather artificial manner (e.g. "`car dealership`").

Chapter 8 summarizes and discusses this thesis' contributions by relating the core research objectives to the research results and Chapter 9 concludes this work by pointing to open challenges and future directions.

1.3 Scientific Contributions

This PhD thesis covers the results of four years of research work (2008-2011) in the field of knowledge management and knowledge acquisition. This thesis focuses on and contributes to modelling, acquiring and applying knowledge about human goals.

Overall, this PhD thesis makes the following two contributions to theoretical and practical aspects of modeling and acquiring human goal knowledge.

1. It presents a general framework on human goal knowledge which (i) synthesizes and (ii) makes an important step towards unifying existing research on modeling and acquiring human goal knowledge.
2. It operationalizes human goal knowledge in three case studies and demonstrates its value and its potential in several practical problem settings.

Additional contributions of this PhD thesis are described in the following.

◇ It advances the explication and thus the understanding of human goal knowledge.

◇ It explores and adapts existing techniques to automate the acquisition of goal knowledge.

◇ It examines different types of textual resources with respect to their value for acquiring human goal knowledge.

◇ It presents a codification of the data model in OWL, the Web Ontology Language, to be useable and accessible for semantic web technologies.

Parts of this PhD thesis have been published in various conference proceedings and journals:

- Strohmaier, M., Kröll, M. and Körner, C. (2009). Intentional query suggestion: Making user goals more explicit during search. In *Proceedings of the 2009 Workshop on Web Search Click Data (WSCD'09) held in conjuction with WSDM'09.*
- Strohmaier, M. and Kröll, M. (2009). Studying databases of intentions: Do search query logs capture knowledge about common human goals? In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09).*

- Kröll, M. and Strohmaier, M. (2009). Analyzing human intentions in natural language text. In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09).*
- Jeanquartier, F., Kröll, M. and Strohmaier, M. (2009). Intent Tag Clouds: An intentional approach to visual text analysis. In *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe'09).*
- Kröll, M. and Strohmaier, M. (2009). Extracting Human Goals from Weblogs. In *Workshop "Knowledge Discovery, Data Mining and Machine Learning 2009" held in conjunction with LWA'09.*
- Kröll, M., Körner, C. and Strohmaier, M. (2010). iTag: Automatically annotating textual resources with human intentions. In *Emerging Technologies in Web Intelligence.*
- Strohmaier, M. and Kröll, M. (2011). Acquiring knowledge about human goals from search query logs. In *Information Processing and Management*, Elsevier.
- Kröll, M., Fukazawa, Y., Ota, J. and Strohmaier, M. (2011). Automatically constructing concept hierarchies of health-related human goals. In *Proceedings of the 5th International Conference on Knowledge Science, Engineering and Management (KSEM'11).*

Lastly, it is important to state that theoretical and practical results of this work seek to contribute to and advance the research area of Goal Knowledge Acquisition, an important problem in Commonsense Knowledge Acquisition and Representation. This work refrains from studying, discussing or generating cognitive models of human goals which is the focus of work in psychology or philosophy. While this work is motivated and informed by other areas' approaches, definitions and models, the proposed framework is a product of synthesizing existing work and is meant to further explicate knowledge about human goals.

1.4 Support

This work has in part been funded by the FWF Austrian Science Fund Grant P20269 TransAgere.

Chapter 2

Related Work

The related work chapter is divided into two main parts which correspond to ontological and acquisitional aspects of human goal knowledge, i.e. this work's core contributions. Section 2.1 reviews work on how goal knowledge is specified and modeled in other areas including goal-oriented requirement engineering, intelligent agent theory or commonsense knowledge acquisition. Examing other areas' understanding of goal knowledge informed the development of the proposed framework. Section 2.2 gives an overview of information extraction and knowledge acquisition techniques which are necessary to automate the acquisition of human goal knowledge from textual resources.

2.1 Human Goal Knowledge

Knowledge about human goals has been found to be an important kind of knowledge for a range of challenging research problems. These challenges include goal recognition from people's actions, reasoning about people's goals or the generation of action sequences that implement goals (planning) (cf. [149] or [24]). Explicitly representing and modeling goals has been acknowledged as an asset in various research and application domains. This work has been inspired by other domains' understanding on goal knowledge. Goal knowledge is not only composed of goals (intentional concepts) but also of intentional relations. Modeling goal knowledge has been addressed in a number of research areas including (i) Requirements Engineering (RE), (ii) Intelligent Agent Theory, (iii) Commonsense Knowledge Acquisition, (iv) Information Retrieval and (v) Human Language Technology.

2.1.1 Models on Goal Knowledge

2.1.1.1 Goal Modeling in Requirements Engineering:

Goal-directed or goal-oriented requirements engineering (RE) reflects the importance of using the concept of goals in various stages of the RE process. These stages include requirements elicitation, requirements negotiation, requirements specification and requirements validation. The explication of goal concepts during modelling requirements provides several positive effects: making goals explicit positively affects the relevance of requirements, i.e. their pertinence (cf. [189]), it allows for estimating whether requirements specification is sufficiently complete (cf. [189]) or it offers rationales for requirements (cf. [90], [112]). By incorporating goal concepts into the design process, emerging "why" questions can be addressed leading to an increased understanding of motivations and interests of involved stakeholders. In van Lamsweerde [172], further motivating aspects are presented which speak in favour for including goal concepts in RE. Complementary, goals may help to detect conflicts amongst requirements and they may be of help to resolve these conflicts eventually (cf. [144], [173]). Last but not least, identifying goals often proves helpful during the acquisition of requirements (cf. [41]).

RE literature (cf. [138], [173]) discerns the following goal types which often are regarded as instances of goal concepts or categories: "Achieve", "Cease", "Maintain", "Avoid", "Optimize" goals. These goal types are often denoted as goal patterns referring to the pattern of temporal behavior they exhibit. Table 2.1 gives an overview of used goal types in RE, their corresponding temporal pattern and behavior.

Goal Type	***Temporal Pattern***	***Temporal Behavior***
Achieve	$G \Rightarrow \diamond Q$	Q holds in some future state.
Cease	$G \Rightarrow \diamond \neg Q$	Q will not hold at some future state.
Maintain	$G \Rightarrow \Box Q$	Q always holds in the future.
Avoid	$G \Rightarrow \Box \neg Q$	Q will never hold in the future.
Optimize	$G \Rightarrow \diamond Q_{(max/min)}$	Q will hold in some future state with the additional constraint that an objective function related to Q is maximized or minimized.

Table 2.1: An overview of goal types in requirements engineering and their corresponding temporal behavior. (cf. [81])

In this case, temporal behavior is expressed by linear temporal logic, where G and Q denote conditions and where $\diamond$ (a condition holds in some future state) and $\Box$ (a condition holds in all future states) are temporal operators. Explicating temporal behavior by means of temporal operators appears essential within system specifications not only to charactize goal instances but to be able to better design system procedures.

Besides categorizing goals with regard to their temporal behavior, other high-level categorization schemes for goal concepts are reported in RE literature (cf. [81], [172] or [118]). These schemes were introduced to account for different aspects of goal instances including range, ownership or functionality. In the following, established schemas are reviewed which discern goals along several dimensions.

(i) System vs. Individual Goals (cf. [172]): An individual or private goal is typically pursued by a single actor in a system. In contrast, a system goal represents a high-level goal whose achievement requires the orchestration of many components in the system. System goals are further specialized into "SatisfactionGoals", "InformationGoals", "RobustnessGoals", "ConsistencyGoals", "SafetyGoals" and "PrivacyGoals". Table 2.2 gives an overview of their purpose as well as of their temporal behavior, i.e. whether a system goal is to achieve or avoid a certain state of affairs or whether it is to uphold a desired state.

Specialized System Goal	**Description**	**Pattern**
Satisfaction Goal	satisfy an agent's request	achieve
Information Goal	keep an agent informed about certain states	maintain
Robustness Goal	recover from human or automated agent's mistakes	achieve
Consistency Goal	maintain consistency between parts of the system	maintain
Safety / Privacy Goal	maintain agents in states which are safe and observable	avoid

Table 2.2: Various kinds of system goals, a corresponding description and its pattern which reflects the temporal behavior. (cf. [172])

(ii) Functional vs. Non-functional Goals: van Lamsweerde [172] describes functional goals to "*underly services that the system is expected to deliver [...]*" and non-functional goals to "*refer to expected system qualities such as security, safety, performance [...]*" The distinction functional vs. non-functional goals relate to the distinction hard vs. soft goals. Softgoals have been used to analyze non-functional requirements (goals) since softgoals do not adhere to objective success criteria; the assessment whether a softgoals has been achieved or not is subjective in nature and varies from individual to individual. In further consequence, softgoals are utilized to analyze non-functional requirements such as safety, usability or flexibility (cf. [82]).

The importance of non-functional requirements has long been neglected [118]. Since non-functional requirements refer to global qualities of system, they have long been handled in an informal way. Yet, modeling and thereby acknowledging non-functional requirements has been tremendously beneficial for the RE process, e.g. by detecting conflicts at an early design stage.

2.1.1.2 Goal Modelling in Planning

A plan is a sequence of operator instances, e.g. actions, such that executing them in the initial state will change the world into a state which satisfies the goal state description. The first action of a planning agent is to ask for or generate a goal state. Thus, the representation of goals or rather goal states proves essential in the planning domain. Situation Calculus (SC) was one of the first attempts to formalize states and actions by logically describing situations. SC is a version of first order logic augmented to allow reasoning about actions in time. A planning problem in its simplest form can be described by three components, i.e. initial state, goal state and operators represented by logical sentences. In SC, an exemplary initial /goal state pair (from [125]) could be:

Initial State:		At(home, S_0) AND $\neg$ Have(milk, S_0)
Goal State:	Exist s	At(home, s) AND Have(milk, s)

Restrictions of situation calculus led to the development of the STRIPS language [48] which describes actions in terms of preconditions and their effects. The former planning example reads:

Initial State:	At(home) AND ¬ Have(milk)
Goal State:	At(home) AND Have(milk)

where states are represented by conjunctions of function-free ground literals; predicates are applied to constant symbols. Operators in STRIPS consist of three components (i) description of an action, (ii) precondition and (iii) effects. Extensions of STRIPS are successfully applied to real-world scenarios nowadays, for example, surveying and controlling space missions.

Subareas of planning such as goal/ intent recognition (cf. [150], [180]) or plan recognition (cf. [24]) continuously enriched these classical approaches by incorporating additional components, e.g. various goal types or relations amongt plan constituents. Richer models and formalizations emerged which were most welcomed in the face of real-world planning applications requiring expressive plan representations. In [52], Yolanda Gil argues in favor of description logics [6] to represent and reason about plans. She discusses description logics based approaches to represent plan knowledge and groups them into four categories: (i) object taxonomies, (ii) action taxonomies, (iii) plan taxonomies and (iv) goal taxonomies.

Approaches not using description logics include Kautz's formal theory of plan recognition (cf. [80]). In his framework, goals and plans an agent might have are stored in an event hierarchy which represents, e.g. behavioral aspects of the target domain. Event nodes can represent actions, goals or plans. Among these event types, no formal distinction is made. The advancement of Kautz's work in the area of plan recognition was that he showed how conclusions about people's unobserved actions or goals could be deduced by using logical entailment. A common criticism addresses the fact that his approach does not contain preference mechanisms, e.g. to prefer one goal or plan over another.

Probabilistic reasoning techniques were then applied to determine which goals or plans were likely. Approaches included baysesian believe nets for plan inference (cf. [26], [70]) and application of the Dempster-Shafer theory (cf. [23], [14]). Besides known challenges of probabilistic approaches including requirement of a posteriori probabilities as well as costly calculation and processing of probabilities, they also have to face up face up to scaling issues because of the correlation between the network's and the domain's size.

2.1.1.3 Goal Modeling in Intelligent Agent Theory:

In the context of Goal-Based Intelligent Agents [125], the belief-desire-intention (BDI) model [18] attempts to model an agent's mental state and is one of the first models of practical reasoning agents; the BDI model is a software model developed for programming intelligent agents. Follow up work was conducted by [36] who explored the rationale balance among agents' beliefs, goals, actions and intentions. In his work on BDI agents, Bratman argues that rational behavior cannot just be analyzed in terms of beliefs and desires, yet that a third mental state is necessary: intentions. Intentions are desires to which the agent has to some extent committed itself to: in practice, this means that the agent has begun executing a series of actions. In BDI notation, a goal is understood as a desire which is actively pursued by the agent, i.e. a future state which the agent wants to achieve. Existing approaches to goals typically focus on one or several types of goals (cf. [174]).

Goals in Intelligent Agent Theory often are distinguished along the following dimensions:

- **Procedural vs. Declarative Goals** (cf. [181]): A procedural goal is satisfied when a process is launched, e.g. the execution of actions. Procedural goals thereby differ from declarative goals which aim to reach a state of affairs. So to say, declarative goals relate to the result of an action whereas procedural goals focus on the action per se. “AchievementGoals”, ”MaintenanceGoals”, “PerformGoals, “TestGoals” or ”QueryGoals” (cf. [19]) represent instances of procedural and declarative goals. A perform goal is the goal to execute actions and is the sole instance of procedural goals. The remaining goals seek to either achieve (AchievementGoal) or maintain (MaintenanceGoal) a state of affairs. QueryGoals or TestGoals coincide as their common desire is to have a certain piece of information. Since QueryGoals and TestGoals also attempt to reach a state, i.e. a state where the respective piece of information is available, these goals are also classified as declarative goals.

- **System vs. Individual Goals** (cf.[172]): This distinction was adopted from the area of goal-oriented requirements engineering (GORE). An individual goal is typically pursued by a single actor in a system. In contrast, a system goal represents a high-level goal whose achievement requires the orchestration of many components in the system.

- **Hard vs. Soft Goals** (cf.[19]): Intelligent agent theory subscribes to the distinction between hard and softgoals researched and introduced in the area of requirements engineering (cf. functional vs. non-functional goals by [118]).

Intelligent agents’ goals can range from simple to complex goals depending on the application domain. To give an example, simple goals rather take place in constraint environments such as the well-known Wumpus world example [125] where the agent pursues goals such as “`find gold`” and “`survive while doing so`”. On the web, intelligent agents [65] are likely confronted with complex goals related to real-world scenarios such as “`organize my trip to Vienna`”. Making the representation of goals explicit has been gaining importance in the past years (cf. [181], [19]). An explicit representation of goals is advantageous, e.g. to apply reasoning techniques to agent goals. BDI agent technologies including Prometheus [126] or Tropos [55] have subscribed to these ideas.

2.1.2 Intentional Concepts

The concept “Goal” belongs to a broader family of concepts, i.e. “Intentional Concepts” which originated in Intelligent Agent Theory literature (cf. [36], [169]). These concepts did not only include goals but also wider aspects such as commitment, belief or ability. Eric Yu adopted intentional concepts in his strategic rationale model [187] to allow for an intentional perspective in the process of engineering requirements. Yu deemed a formalization of goals and goal components valuable for modeling explanations and rationales for requirements. Explanatory aspects such as intentional concepts were thus included into the strategic rationale model. Besides focusing on the “what” and “how” in the engineering process, these aspects contributed to answering the “why”. For this work’s data model development, Yu’s work is of particular value because of its analytical capabilities. However, goals in RE often are dedicated to describe a desired state of a system (see previous paragraphs) and thus slightly differ from the notion of a human goal. To agree on a definition of human goals, this work seeks to develop a common understanding by surveying relevant literature. In the following goal definitions from several research areas are discussed with respect to modeling and acquiring human goals. Reviewed research areas include (i) Commonsense Knowledge Acquisition, (ii) Information Retrieval, (iii) Human Language Technology, (iv) Requirements Engineering and (v) Intelligent Agent Theory.

Commonsense Knowledge spans a broad spectrum of human experiences such as "a lemon is sour" or "if you forget your anniversary, someone might be unhappy with you". Commonsense comprises fact-based knowledge as well as knowledge about other aspects including commonsense goals [101]. Examples of commonsense goals include "buy a house", "paint the bathroom" or "get out of debt" (from [157]). These goals would intuitively be categorized as human goals as well. [101] give an informal definition of goals: "*Goals often answer the why questions about human behavior, and provide good clues as to the when, how, and other considerations. They are therefore fundamental to explanation.*"

In Information Retrieval (IR) research, the focus is on approximating a user's underlying search goal to inform and improve retrieval processes, e.g. retrieving more relevant resources. For this purpose, Broder et al. [21] introduced a high level taxonomy of search intent by classifying search queries into three categories: navigational, informational and transactional. Two years later, Rose et al. [146] elaborated on this distinction by introducing a more detailed categorization of search intent. They repeatedly refined their goal categories based on empirical evidence. Their efforts resulted in a search intent hierarchy where high-level categories resembled Broder's taxonomy. Follow up research led to evolutions of Broder's work which included collapsing categories, adding categories [7] and/or focusing on subsets only [91]. In summary, two observations can be made: First, high-level definitions of search intent, e.g. Broder's, do not require a query's search goal to be made explicit. Second, the adopted abstraction level often is too high so that information about individual human goals, e.g. "how to sell my car", is lost. From a knowledge acquisition perspective, IR definitions of (search) intent thus appear to be impractical.

In the area of Human Language Technology, Marta Tatu [167] defined goals as "*expressions of a particular action that shall take place in the future, in which the speaker is some sort of agent.*" This definition is based on the concept of a "private state" [137] which refers to a state not open to objective observation or verification. A private state thus represents mental or inner states covering opinions, beliefs, thoughts, feelings, emotions or goals. In her work, Tatu seeks to identify goals in textual resources, yet not for acquisition purposes but as means to inform and improve the task of Question Answering (Q&A).

According to Regev [138], Requirements Engineering (RE) literature reports three established definitions of goals depending on the respective approach. In their KAOS approach, Dardenne et al. [41] define a goal as "*a nonoperational objective to be achieved by the composite system*". A goal is refined until it becomes an objective which is satisfiable by the prospective system, e.g.

through state transitions. Anton [5] defines goals to be "*targets for achievement which provide a framework for the desired system. Goals are high level objectives of the business, organization, or system. They express the rationale for proposed systems and guide decisions at various levels within the enterprise.*" This definition explicitly contains a motivation for goals for they serve as rationales in the RE process. The goal "maximize corporate profits" complies with this definition and represents a high-level enterprise goal. In the goal-oriented requirements language [72], a goal is defined as "*a condition or state of affairs in the world that the stakeholders would like to achieve. How the goal is to be achieved is not specified, allowing alternatives to be considered.*" This definition explicates that in RE a clear distinction is made between the "why" something is done and the "how" something is done. Which steps are taken to accomplish a goal, is usually specified by other goal knowledge components such as tasks. Although these definitions appear to be rather system-centric definitions, they apply to individuals as well specifying a "condition" or a "state of affairs" which is to be reached.

In Intelligent Agent Theory as well as in Planning, goals can range from simple to complex goals depending on the application domain. To give an example, simple goals rather take place in constraint environments such as the well-known *Wumpus* world example [125] where the planning agent pursues goals such as "`find gold`" and "`survive while doing so`". On the web, intelligent agents [65] are likely confronted with complex goals relating to real-world scenarios such as "`organize my trip to Vienna`". Making the representation of goals explicit has been gaining importance in the past years (cf. [181], [19]). An explicit representation of goals is advantageous, e.g. to apply reasoning techniques to agents' goals. BDI agent technologies like Prometheus [126] or Tropos [55] have subscribed to these ideas.

Research Area	**Purpose**
Commonsense Knowledge Acquisition (CSKA)	- to provide context-sensitive help and debugging facilities (cf. [99]) - to advance intelligent user interfaces (cf. [110]) - to apply commonsense reasoning to interface agents (cf. [101])
Information Retrieval (IR)	- to inform result ranking (cf. [21], [184]) - to trigger vertical search engines (cf. [69], [96]) - for query recommendation/suggestion (cf. [109]) - to inform interface design (cf. [145])
Human Language Technology (HLT)	- to support the task of Question Answering (Q&A) tasks (cf.[167])
Requirements Engineering (RE)	- as rationales for requirements, e.g, answering why questions (cf. [90], [112], [187]) - to provide a criterion (i) for sufficient completeness of requirements specification and (ii) for requirements pertinence (cf. [189]) - to detect conflicts (cf. [144], [173])
Planning / Intelligent Agent Theory (PIAT)	- to select optimal actions in accordance with given goals (cf. [125], [18])

Table 2.3: Various purposes human goals serve depending on the research domain.

In previous paragraphs, (human) goal definitions from related literature have been surveyed. Table 2.3 provides an overview of main purposes goals and goal constructs are supposed to answer in their respective areas. It summarizes different perspectives on human goal instances and illustrates the diversity of purposes reflecting the underlying motivation to introduce goals. Goals sometimes appear to be mere auxiliary structures - definitively a valid assumption for IR. Table 2.3's statement corroborates and motivates the development of a general data model since each research area utilizes and thus defines goal knowledge to comply with specific research questions and particular applications.

Research Area	**Goal Instances**	**References**
Information Retrieval(IR)	"aloha airlines" (navigational goal), "why are metals shiny" (informational goal), "download songs" (transactional goal)	[21], [146]
Commonsense Knowledge Acquisition (CSKA)	"lose 50 pounds", "beat depression", "be more confident", "be happy with myself"	[101], [157]
Human Language Technology (HLT)	"leave the country", "end a pitching slump", "give up drink", "go for a walk"	[167]
Requirements Engineering (RE)	"meeting scheduled", "low effort", "minimize costs", "serve more passengers", "user friendly", "notify instructor", "mark assignment", "change password"	[171], [188], [124]
Planning / Intelligent Agent Theory (PIAT)	"stay alive", "find gold", "organize trip to Vienna"	[125], [18]
Goal Knowledge Acquisition	"avoid wrinkles", "look young", "get a degree", "pass a drug test", "sell my car", "play online poker", "get pregnant"	[161], [86]

Table 2.4: Representative goal instances from five research areas are presented including (i) Information Retrieval, (ii) Commonsense Knowledge Acquisition, (iii) Human Language Technology, (iv) Requirements Engineering and (v) Goal Knowledge Acquisition.

Table 2.4 presents representative, exemplary goal instances according to their research area. It thus compares research areas with respect to representative goal instances. Reviewing these instances reveals that instances from commonsense knowledge acquisition, human language technologies, requirements engineering and planning/intelligent agent theory more or less ressemble each other as they are mostly formulated in form of verb phrases describing a state of affairs to reach. Only some goal instances in RE are formulated differently and are more system-centric, yet still expressing a desired goal state. An outlier in this table are goal instances from IR which appears to be natural since IR has no interest in acquiring goal instances, yet merely recognizing the intent behind a query be it implicit or explicit. The purpose of utilizing goals appears to differ from area to area (see Table 2.3), nevertheless research areas share some common grounds (i) on how goals are expressed and (ii) on their nature. For the purpose of comparison, the last row of Table 2.4 contains exemplary goal instances of Goal Knowledge Acquisition, this work's focus.

2.1.3 Intentional Relations

To be of value, goal knowledge components need to be related to each other. Only a network or hierarchy of relations enables more sophisticated operations on knowledge such as reasoning or planning. "Intentional Relations" therefore are introduced as relations who connect goal knowledge components with each other, e.g. a goal to a subgoal or a goal to a task. So far these relations fell into the category of semantic relations. In the context of goal knowledge, it appears beneficial

Relation Type	Relation Instance / Lexico-Syntactic Pattern
Means-Ends	“IsAchievedBy” [51] “InOrderTo”, “CrucialFor”, “NecessaryFor”, “EssentialFor”, “HelpsTo” [87] “Means-Ends” [187] “Goal”, “Means” [141]
Causal	“Causes/ IsCausedBy”, “LeadsTo” [116] “ResultsIn”,” EnablesTo”, “ForThePurposeOf” [87] “EffectOf”, “DesiredEffectOf” [104], [155] “Cause” [141], [54] “Reciprocal” [54]
Affective	“MotivationOf”, “DesireOf”, “IsMotivatedByGoal” [104], [155]
Functional	“UsedFor”, “CapableOfReceivingAction” [104], [155]

Table 2.5: This table gives an overview of other areas’ relation types which exhibit goal-oriented characteristics.

to add another type of relation to indicate their contribution to goal-oriented behavior. In related literature, many types of relations appear to partly exhibit goal-oriented, intentional characteristics such as the “Cause” relation type. To give an example: “She married him because he made her laugh” can be transformed into “make a woman laugh” (means) so that “she marries you” (end) (example taken from [54]). Other examples do not need to be transformed such as the triple (“eat breakfast”, “EffectOf”, “full stomach”) (example taken from [104]). Table 2.5 agglomerates a non-exhaustive list of relation types which have already been introduced into research literature. These relation types all exhibit some sort of goal-oriented behavior and can thus accordingly be denotated as intentional relation.

2.1.4 Goal Knowledge Representation

In [110], Marvin Minsky elaborates on the challenges of knowledge representation and starts his discussion with following statement: *“There is no best way to represent knowledge.”* In summary, each form of representation has its advantages and disadvantages; logic might be too inflexible, neural networks too rigid and semantic networks might be too flexible. Instead of recommending a particular way to represent goal knowledge, this section seeks to provide knowledge engineers with sufficient information for an educated decision. For that purpose, existing representation approaches are reviewed ranging from formal to informal ones. Each representation’s characteristics are pointed out and discussed, e.g. to what degree reasoning is affected. Approaches to represent knowledge about human goals can be roughly divided in two categories: (i) formal, logical and (ii) informal, semi-structured natural language approaches.

2.1.4.1 Formal Representation of Goal Knowledge

Artificial intelligence research has been investigating possibilities for processing knowledge, i.e. represent it in such a way to make it utilizable by machines, e.g. intelligent systems. These systems were meant to act intelligently, i.e. making decisions or taking actions which were logically correct. For that purpose, respective knowledge representation languages should be expressive, concise, unambiguous and independent of context. To satisfy these requirements, formal languages have been employed or advanced such as first-order logic, description logic [183], situation calculus [108], event calculus [115], fluent calculus [168] or episodic logic [153]. These languages have thus been used to represent commonsense knowledge including human goals (cf. [56], [68]). Gordon et al. [56] discuss the need for formalizations of commonsense psychology and present a new methodology for constructing formal theories in commonsense knowledge domains. They highlight and describe $\sim$30 representational areas including “Goals” and “Plans” with regard to planning knowledge. Hobbs and Gordon [68] provide a formal theory of goals which is authored in first order logic. Their logical constructs adhere to intelligent agent theory, i.e. supporting intelligent agents to execute plans. Amongst these are standard constructs such as (i) introducing a hierarchy of plans, goals and subgoals, (ii) types of goals or (iii) failing and succeeding mechanisms. In addition, they also define “thriving” which describes the agent’s process of achieving various subgoals in order to achieve the top level goal.

Research projects having (partly) devoted themselves to a formal representation include Cyc [92], ThoughtTreasure [114] or KNEXT [152]. Cyc and ThoughtTreasure are both projects which aim to acquire commonsense knowledge whereas KNEXT focuses on knowledge about the general world. Cyc [92] developed an internal formal language, CycL, to represent commonsense knowledge. CycL’s syntax was inspired by first-order predicate calculus and from the programming language Lisp. CycL’s vocabulary consists of terms which can be divided into constants, non-atomic terms, variables, and a few other types of objects. Meaningful CycL expressions are generated by combining terms. These expressions are used to make assertions in the Cyc knowledge base. While ThoughtTreasure was inspired by Cyc, the project utilizes other representations but logic as well including scripts [149] or grids, both for stereotypical settings. KNEXT [152] denotes an ongoing research project dedicated to acquire knowledge about the general world including statements such as “doors can be opened” or “airplanes can fly”. The boundaries between commonsense and general world knowledge have not yet been strictly determined and are therefore rather vague. In KNEXT, knowledge is represented as form of episodic logic that facilitates the mapping of natural language to logic and vice versa. Knowledge in KNEXT therefore aims at a greater expressiveness than other forms of logical representations.

2.1.4.2 Informal Representation of Goal Knowledge

From the previous section, we know that knowledge representation languages should be expressive, concise, unambiguous and independent of context to be of any value, e.g. in inference processes. Informal representation languages such as (a structured form of) natural language do not satisfy these requirements. The reason, why informal languages are still in use, lies in the nature of the knowledge to represent: (i) Human goal knowledge is inexact in a logical sense, i.e. everybody considers goals and ways to achieve them individualistically. (ii) Goal knowledge is defeasible, i.e it is true but not always and might contain exceptions to the rule. (iii) Approaches to accomplish a goal may differ from person to person depending on, e.g. their cultural or financial background. Goal knowledge thus appears to be sensitive to the context.

The motivation to use informal representations is that by forcing goal knowledge into a formal representation much of its characteristics are lost. Human behavior including goals and activities

	Ambiguity	Reasoning	Transformation steps required	Knowledge Acquisition	Handling Uncertainty	Resembles Human Reasoning
Formal Representation	no	deductive	yes	by experts	no	no
Informal Representation	yes	statistical	only to a certain degree	by non-experts, e.g. volunteers	yes	yes

Table 2.6: Characteristics of formal and informal representations are summarized with respect to goal knowledge.

of everyday life is too diverse to be precisely described by logical definitions or deductions. Formal representations are not designed for handling inconsistent, controversial or ambiguous knowledge. Informal representations can cope with these shortcomings, e.g. by employing methods from statistics.

Research projects having (partly) devoted themselves to an informal representation include Open Mind Common Sense (OMCS) [155], ConceptNet [104] or DART [34]. OMCS aimed at capturing commonsense knowledge by a volunteer-based approach. It later evolved into ConceptNet which uses semi-structured natural language [61] to represent commonsense concepts which are organized in a semantic network. ConceptNet was designed to make practical context-based inferences over real-world texts, i.e. inferences are not logical deductions as e.g. in Cyc, but rather are based on graph reasoning methods like spreading activation or network traversal. Peter Clark introduced DART [34], a knowledge base which contains general world knowledge. He also decided on an informal representation approach by normalizing statements in canonical form. The sentence "The camouflaged helicopter landed" would produce following tuples (AN "camouflaged" "helicopter") and (S "helicopter" "land") where AN stands for adjective-noun and S for sentential tuple (example from [34]).

2.1.4.3 Contrasting Formal and Informal Representations

This section contrasts formal and informal representations with respect to represent knowledge about human goals. In general, it does not appear reasonable to speak of strengths and weaknesses but rather to point out each representation's characteristics (summarized in Table 2.6). Whether these turn out to be positive or negative depends on the use case's requirements.

- Ambiguity: While formal representations such as logic prohibit ambiguous assertions, informal representations are less restrictive. Knowledge representations in natural language, for instance, allow for lots of ways to express and thereby represent the same knowledge. To reduce ambiguity, informal approaches also seek to introduce more structure. As an example, real-world knowledge is conceptualized in ConceptNet (cf. [61]): concepts represent real-life aspects or situations expressed in natural language. A normalization process merges many related natural language phrases to a concept, i.e. a concept then encompasses all phrases that normalize to the same text.

  In brief, the less formalization the higher, the degree of ambiguity. Ambiguity per se is not a bad property. Context is key, i.e. whether the domain or problem setting benefit from a more ambiguous representation. Advantages of ambiguity include better adaption to new situations or potential exploration of new possibilities as in case of human DNA.

- Reasoning: Reasoning or inference in its traditional form is a subfield of logics and describes the process of reaching a conclusion which is based on already existing data. Formal approaches allow for deductive reasoning, i.e. from more general knowledge infer more specific knowledge. Due to characteristics such as ambiguity or uncertainty, informal representations cannot perform deductive reasoning. They have to resort to statistical reasoning which leads to plausible or practical inferences. Representations such as semantic networks support reasoning which is based on spreading activation or network traversal. Graph based reasoning is associative and thus not as expressive, exact or certain as logical inferences, but it is much more straightforward to perform and useful for reasoning practically over text.

- Transformation Steps: It requires far more effort to transform real-world facts, e.g. people's goals and their actions, into a formal representation. Let's take for example reasoning over natural language text. If a formal framework such as Cyc were employed, reasoning over text could become quite complex: Text, which is inherently ambiguous, must first be mapped into Cyc's unambiguous logic, which often turns out to be a cumbersome and time-consuming task. By using an informal representation, the requirement for transformation can be reduced.

- Acquisition: The knowledge representation's approach has determining influence on acquisition strategies. For formal representations, human experts are required which are familiar with the representation language and thus can conduct the transformation steps. Informal approaches, however, can tap into a rich acquisition source, i.e. volunteers providing knowledge. Since informal representations are close to natural language representations, (i) volunteers do not need to be trained to provide knowledge and (ii) acquisition strategies can be kept simple, e.g. by posing questions via a user interface. In addition, informal approaches appear to be more amenable for automating the process of knowledge acquisition than formal approaches, simply because less complex transformation steps are required.

- Uncertainty: Knowledge about human goals is complex and dynamic, i.e. it is unlikely to know all the facts - the entire context. Yet, knowledge engineers still want to operate on this knowledge, make decisions under uncertainty as well. While formal approaches were not designed to take uncertain knowledge into account, informal approaches can handle uncertainty, e.g. by using probability theory or statistics. On the border between logic and statistics are fuzzy logic and fuzzy set theory [190], which were designed for reasoning about phenomena with a high degree of uncertainty.

- Human Reasoning: There appears to be a divergence between formal approaches, e.g. logic, to reasoning and what is known about how people reason (cf. [103]). Human reasoning has a number of properties that distinguish it from traditional logical reasoning, e.g. it is based on induction, abduction and empirical evidence. The ability to perform inductive reasoning, i.e. generalizations from known, single experiences, has proven valuable in mankind's evolution. Humans are successful inductive reasoners since they acquire a wealth of information about their environment from which then patterns can emerge. In contrast, formal representations appear to have difficulties handling the imprecise way that humans categorize and compare things.

By contrasting formal to informal knowledge representation approaches, each approach's characteristical features have been emphasized. Each representation might be favorable in particular problem settings. Yet, to adequately represent real-world knowledge including human goals, a multiple representation approach is considered to be most beneficial as experiences already illustrate [114]. In [110], Marvin Minsky puts similar thoughts forward to discussion by introducing several levels of representation which reflect formal as well as informal representations such as logics, neural networks or natural language. In Minsky's opinion, a combined approach will compensate for their disadvantages.

2.2 Information Extraction & Knowledge Acquisition

Information Extraction (IE) can be considered a key technology for knowledge acquisition and refers to the automatic extraction of structured information from unstructured textual resources. It also refers to the process of making information explicit and thereby useable both by humans and machines; information that otherwise remains hidden in vast amounts of digital text documents, e.g. newspaper archives or social media platforms. Extracted information is often structured in relational databases where the information becomes accessible, e.g. to querying mechanisms. To better understand the functions IE systems should perform, consider following sentence "Vienna is the stylish capital of Austria". An IE system takes this sentence as input and is expected to output a mapping to a relational tuple, for instance, ("Vienna", "CapitalOf", "Austria"). Accurately extracting tuples often requires existing knowledge, i.e. a certain degree of human involvement. This condition has been termed knowledge engineering bottleneck. The advent of the World Wide Web triggered a paradigm shift in IE due to changing requirements: lesser human involvement due to the sheer amount of digital resources. Traditional IE approaches had to evolve towards automatic procedures. In the following, this evolution will be briefly surveyed; the chronological grouping was partly adopted from Etzioni et al. [44].

2.2.1 Traditional Information Extraction

Traditional Information Extraction (IE) systems focused on locating instances of pre-specified relations such as time and place of events, from small, homogeneous, domain-specific corpora. The spectrum of relation types was continually enlarged to support, for instance, the construction of lexical databases such as WordNet [63]. Donald Hindle [67] and Marti Hearst [62] were among the first who pioneered hand-crafted, textual patterns in the early 1990's. For their analyses, they took into account the surrounding context including syntactical and grammatical characteristics. While Hindle focused on predicate-argument structures, Hearst developed lexico-syntactic patterns, i.e. part-of-speech enriched regular expressions, to extract hyponymy ("is-a") relations from textual resources. Being hand-crafted, only instances of predefined relation types could be identified and extracted. In addition, Traditional Information Extraction was disadvantageous (i) for being too domain-dependent allowing no portation of successful patterns and (ii) for being heavily time consuming.

2.2.2 Automated Information Extraction

The objective of Automated Information Extraction (IE) is to continuously reduce human involvement in the IE process thereby seeking (i) to increase processable data volumns and (ii) to broaden the spectrum of extractable relation and entity types. Machine learning was among the preferably used methods and ideally complemented the previously knowledge-based approaches. Human involvement was still key in order to provide learning algorithms with annotated training examples. Yet, instead of crafting patterns by hand, researchers attempted to automatically learn these patterns to reduce human efforts. First work in this direction included Soderland's *CRYSTAL* system [158], Kim's *PALKA* system [83], and Riloff's *AutoSlog-TS* system [142]. In contrast to *CRYSTAL* and *PALKA*, the *AutoSlog-TS* system represented an unsupervised approach to IE by generating extraction patterns using untagged text. The process of compiling large amounts of training data proved to be a major bottleneck to highly scalable IE systems. To minimize human involvement, Agichtein and Gravano [2] presented the IE system *Snowball* which extracted structured data from plaintext documents. The idea is to provide few but frequent training examples in combination with a regular expression the examples have to match. *Snowball* uses iteration cycles to repeatedly check the quality of the extracted instances. These cyclic quality checks reduce error propagation and therefore represent the main advancement compared to Brin's Dual Iterative Pattern Expansion (DIPRE) algorithm [20].

Self-supervised IE systems can be regarded as a subcategory of unsupervised methods. Yet, unlike classic unsupervised methods, self-supervised IE systems find and annotate examples on their own to train a classifier. Representatives include *KnowItAll* [45], a domain-independent system that automatically extracts information from the web. *KnowItAll* is seeded with an extensible ontology and a small number of generic rule templates from which it creates text extraction rules for each class and relation in its ontology. The authors also provided detailed information on lessons-learnt towards the development of an Open Information Extraction (Open IE) system. The term "Open IE" was coined by Michele Banko (cf. [9]) and represents a novel extraction paradigm; the paradigm is meant to address challenges of extracting information from web-scale corpora: Open IE does not require (i) domain specific training data, (ii) in advance specification of relations to extract, (iii) but does require linear scalability due to massive data amounts. *TextRunner* [9] represents a fully implemented Open IE system which features these requirements. When compared to its predecessor *KnowItAll* [45], *TextRunner*'s average error rate is significantly lower while identifying an almost identical number of correct extractions. In addition, *TextRunner* extracts information from all relations at once thereby drastically reducing processing time.

The concept of "never-ending" [25] or "life-long" [10] learning developed simultaneously. It refers to an open ended effort to continuously extract information from the web. In comparison to Open IE, the emphasis lies on constructing a comprehensive reflection of the web's factual content;

predicates to learn are given in advance. Carlson et al. [25] presented *NELL*, an implementation of a never-ending language learning system. The system consists of four subsystem components which utilize semi-supervised learning methods thereby independently attempting to extract candidate facts. A component called knowledge integrator is then responsible for upgrading highconfidence candidates to the status of beliefs. First evaluations of *NELL* yielded promising precision results while constantly accumulating knowledge.

2.2.3 Acquiring Goal Knowledge from Text

The task of acquiring goals from textual resources is refered to as *Goal Mining*. This area covers a broad range of interesting aspects including the acquisition of goals from scientific articles [71], organizational policies [136], organizational guidelines and procedures [98], interviews/diaries [134], search query logs [164] and others. In the area of understanding natural language text, knowledge about human goals gains significance as a novel dimension to be considered. Passonnneau et al. [133] theoretically analyzed whether the task of text segmentation can be informed by employing knowledge about peoples' goals. Marta Tatu [167] analyzed human goals in natural language text to improve the task of question answering. Extracting expressions of human goals to complement social media monitoring tools has been recently explored by [87]. In this previous work political speeches were studied from a goal-oriented perspective and classified into a human goal taxonomy [30]. Political speeches could be compared not only by traditional topic category distributions but also by human goal category distributions. Knowledge about human goals has been found to play a fundamental role in explanation, justification, and rationalization as well. Understanding peoples' goals can help to answer "why" questions about user behavior and user interactions (cf. [46], [101] or [157]). In commonsense enabled applications [99], explicit representations of goal knowledge are crucial for plan recognition and planning. In addition, they are an enabler for intelligent user interfaces which exhibit traits of commonsense understanding such as goal-oriented search [102] or goal-oriented event planning [156].

This PhD work seeks to demonstrate the value of human goal knowledge in practical problem settings. To that end, state-of-the-art information extraction and knowledge acquisition techniques are applied and adapted to operationalize and instantiate knowledge about human goals. In that sense this work is inspired by existing methodological concepts, e.g. for devising extraction patterns to automatically identify human goal knowledge.

PART II: FRAMEWORK

Chapter 3

Ontological Aspects of Human Goal Knowledge

Artificial intelligence research has been driven by the desire to narrow the gap between human intelligence and machine intelligence. To mimick intelligent behavior, it is necessary to equip machines with sufficient knowledge to make reasonable decisions (based on given facts). Intelligent behavior is strongly connected with explicitly knowing "what has to be done" and "how to achieve it". Intelligent agents, for instance, are constantly confronted with critical situations to be handled with care in order to survive. A building block along this road represents knowledge about (human) goals which encompasses useful components to describe, to understand and to reason about human goals. To explicate knowledge and thereby making it accessible and usable, ontological constructs have become popular.

To encode and thus formalize knowledge about human goals, this PhD thesis proposes a data model to inform modelling and acquisition processes. The data model's development has been guided by the following requirements:

- Generality: The data model is supposed to represent a common perspective on human goal knowledge across domains. Since goal knowledge may vary from individual to indivual, from organization to organization, a formal model has to lay the foundations for modelling a wide variety of goal characteristics in different settings. To provide generality, the development is based on and synthesizes existing work on modeling goal knowledge from a number of distinct, yet interrelated research domains including goal-oriented requirements engineering (cf. [187], [41] or [119]), intelligent agent theory (cf. [18], [19] or [174]) or commonsense knowledge acquisition (cf. [101]).
- Clarity: The data model's representation has to be clear and unambiguous to be of use for researchers and knowledge engineers. To provide clarity, the Unified Modelling Language (UML)[1] [17] has been selected for modeling aspects .

[1] UML is a general-purpose modeling language used to describe the classes of a system and their relationships to each other.

- Accessability: The data model is to be accessible by researchers and knowledge engineers across domains and research areas, which is in part realized by using UML as modelling language. To further facilitate accessability and adoption, e.g. through the Semantic Web, this work provides an OWL codification of the data model. OWL stands for Web Ontology Language[2], a logic-based language characterized by well-defined semantics.

- Consistency: The data model must not be inconsistent, i.e. unsatisfiable. Inconsistency might result from (i) over-constraint class descriptions so that a class cannot have any instances or from (ii) class descriptions which allow two classes to have the same set of instances.

- Instantiability: The data model is supposed to serve as a template and guide for acquisition processes. A theoretical validation of this quality appears non-trivial. Yet, a practical evaluation has been conducted by instantiating various components in three case studies (see Chapter 4 to Chapter 6).

This chapter is divided into four sections. Section 3.1 seeks to familiarize the reader with this work's understanding of human goal knowledge. Main modelling components are introduced including intentional concepts and intentional relations. Section 3.2 then presents the data model designed in UML and provides details on all model components. Section 3.3 discusses selected operations on goal knowledge such as reasoning or planning. Each operation is set into relation with the data model, i.e. which components need to be instantiated to enable the respective operation. An elaboration of the meta-process on engineering knowledge about human goals concludes this chapter in Section 3.4.

3.1 Modeling Knowledge about Human Goals

This thesis understands knowledge as a world model which comprises structural and functional properties of the real world. Thus, knowledge about human goals encompasses goal instances as well as additional components such as resources, agents (intentional actors) and different goal types which are necessary to describe and represent the world of human goals. Knowledge about human goals includes contributing components such as actions and therefore complies with Becerra-Fernandez's definition of knowledge to guide actions and to inform decisions [15].

This chapter introduces and characterizes main components of the data model such as intentional concepts including goals, subgoals or tasks as well as intentional relations connecting them. These modelling features are based on Erik Yu's i* framework for modelling and reasoning about organizational environments in early-phase requirements engineering [187]. The i* framework consists of two models, the Strategic Dependency (SD) model and the Strategic Rationale (SR) model. The SD model describes dependency relationships between actors in an organizational context. The SR model, on the other hand, allows modelling interests inside an actor. Both models use intentional elements such as goals, soft goals, tasks and resources to model intentionality in organizational environments. The proposed data model builds upon i* and extends it by providing a more detailed

[2]Dean M, et al. Web Ontology Language (OWL) reference version 1.0. 2002. http://www.w3.org/tr/owl-guide/

characterization of (human) goal knowledge, e.g. adding attributes to existing features, adding types of intentional relations and adding intentional concepts such as goal categories. While not all extensions will prove useful in the context of early-phase requirements engineering, some of them might inspire adaptions to the i* framework.

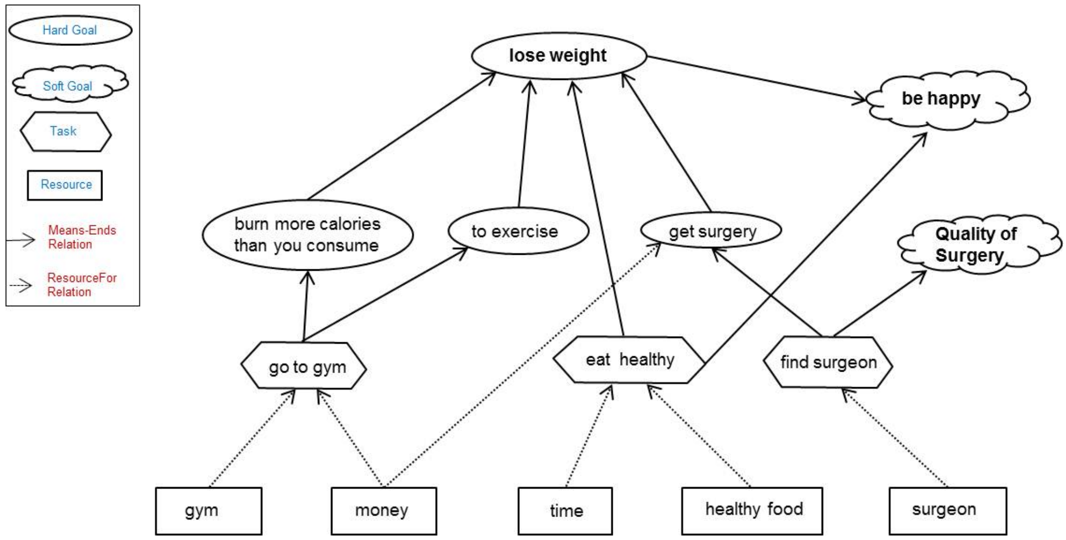


Figure 3.1: Manually generated toy example which models knowledge about the human goal "`lose weight`" represented in i* notation [188]. The model's components are defined in Section 3.1.1

Figure 3.1 illustrates a manually generated toy example which explicates knowledge about the human goal "`lose weight`", thus focusing on intentional concepts and intentional relations while neglecting other aspects. Goal knowledge components such as goals, subgoals, tasks, resources are represented in i* notation throughout this chapter. In this notation, Means-Ends relations connect means to an end. A means can represent a task which specifies a particular way of doing something. In Figure 3.1, the human goal "`lose weight`" expresses a person's goal to achieve a state of lesser weight. Several Means-Ends relations offer a range of possibilities to reach this condition, for instance, you can "`exercise`" and "`go to the gym`" in order to "`lose weight`". Resources such as "`money`", "`healthy food`" or "`time`" represent resources which contribute to the goal's accomplishment.

3.1.1 Modeling Features

This section characterizes the main components of the proposed data model including goals, tasks, resources and intentional relations.

3.1.1.1 Human Goal

Many formal as well as informal definitions of (human) goals exist throughout related literature. Agreeing on a clear definition is vital; it represents a prerequisite for modelling and acquisition efforts. The definition should thus be (i) stringent enough to comply with related literature and (ii) practical enough to be applicable to automatically extract goal instances from textual resources. This PhD thesis therefore uses following definition:

> *A textual statement is regarded as explicit human goal whenever the statement 1) contains at least one verb and 2) describes a plausible future state of affairs that a person may want to achieve, avoid or maintain (cf. [138]) 3) in a recognizable way [164].*

This definition is based on the following rationals: The first part of this definition addresses the crucial role verbs play in explicating human goal instances in textual resources. To extract human goal instances, requiring an explicit goal representation saves us from having to deal with ambiguous expressions. The concept of a "state" is recognizable in definitions from human language technology as well as requirements engineering. Its purpose is to specify a condition or situation in the future which differs from the current one. A goal expresses a person's concrete attempt to achieve, avoid or maintain a plausible, future state. "Plausible" refers to an external observer's assessment whether the human goal could likely represent the goal of a person who formulates the given text phrase. "Recognizable", in the third part, refers to what David Kirsh [84] defines as "trivial to identify" by a subject within a given attention span. According to Kirsh, "trivial to identify" represents the ability to make a decision in constant time. This requirement relates to the definition's purpose to be practical, e.g. enabling human annotations during the process of knowledge acquisition. Results from the first case study (see Chapter 4) corroborate the practicability of this definition.

3.1.1.2 Hard/Soft Goal

Hard and soft goals differ with respect to their success criteria. A hard goal's success can be objectively measured whereas success criteria for soft goals are rather subjective in nature and may vary from individual to individual. Following examples should clarify this distinction. The goal instance "`lose 10 pounds`" is refered to as a hard goal, since a weight loss of 10 pounds can be objectively measured, e.g. by means of a pair of scales. In contrast, there are no clear-cut success criteria for the goal instance "`be happy`", thus this work refers to this instance as soft goal.

3.1.1.3 Goal Decomposition

Complex human goals can be decomposed into simpler (sub) goals. Reducing complexity typically facilitates finding solutions, i.e. the finding of means which contribute to a goal's accomplishment. In Figure 3.1, the high-level goal "`lose weight`" can be approached by choosing amongst a series of subgoals, e.g. "`burn more calories than you consume`", which all entail different sets of tasks and resources. In this process, two decomposition types need to be considered: AND-decompositions and OR-decompositions. A goal can be decomposed in a set of subgoals (i) where each one of them is well suited to contribute to the higher-level goal, i.e. OR-decomposition, or (where) all the goals in the set of subgoals need to be accomplished to contribute to the higher level goal, i.e. AND-decomposition. To clarify, an AND-decomposed goal requires that all its subgoals are addressed and accomplished while an OR-decomposed goal can be accomplished by only one of the alternative subgoals. The concept of AND/OR decompositions in a graph structure has been adopted from Nilsson [123] to reduce complex problems to simpler ones for which it is more likely to find solutions. Figure 3.2 illustrates both types of goal decompositions exemplified by real-world examples.

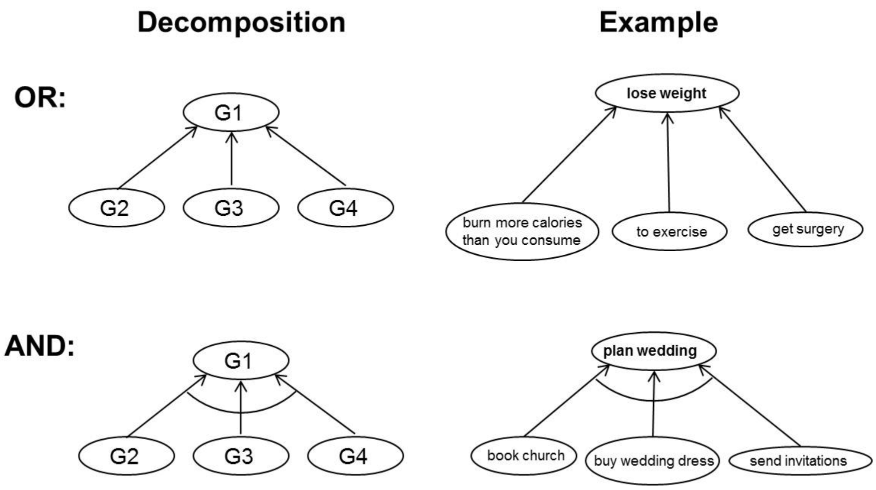


Figure 3.2: The concept of AND/OR decompositions of goals in agreement with [123]. Decomposition notations are exemplified by real-world situations.

In agreement with [123], a restructuring of populated goal models is proposed if sets of subgoals contain both decomposition types. In this case, restructuring means to introduce another level of intermediate goals which are OR-related or where each one is decomposed in a set of AND goals.

3.1.1.4 Subgoal

Through decomposition, goal hierarchies naturally emerge where goals become more complex the higher up the hierarchy they are positioned. A human goal can be decomposed into a set of subgoals and/or a set of tasks.

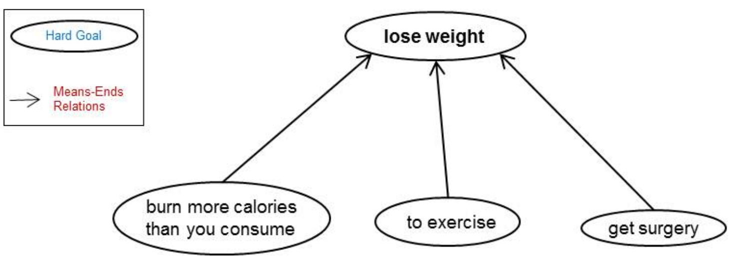


Figure 3.3: Goal - Subgoal relations are illustrated - in this case a decomposition of the high-level goal "`lose weight`" into three subgoals representing three alternatives to achieve the high-level goal. Goal-Subgoal relations are representatives for Means-Ends relations where subgoals represent means to achieve an end. (represented in i* notation [188])

Figure 3.3 shows some goal – subgoal decomposition examples, e.g. "`burn more calories than you need`" would be a subgoal of "`lose weight`".All three subgoals represent valid alternatives, i.e. OR decompositions, for accomplishing the corresponding higher level goal "`lose weight`". A goal is connected to a subgoal by a Means-Ends relation since in this case a subgoal represents kind of a building block to accomplish the higher level goal, i.e. a means to reach an end.

3.1.1.5 Task

A task describes an atomic action or activity which cannot be reasonably decomposed further positioning them at the bottom of goal hierarchies. This complies with other definitions from related areas such as personal information management (cf. [12]) where tasks refer to well-defined actions within more complex structures, e.g. within business processes. Compared to goals, tasks (i) describe concrete actions and (ii) rather specify how something is done. This work's definition deviates from Yu's strategic rationale model where a task can be decomposed further into subgoals and subtasks. This adaptation is considered necessary with respect to automating the acquisition process of goal knowledge.

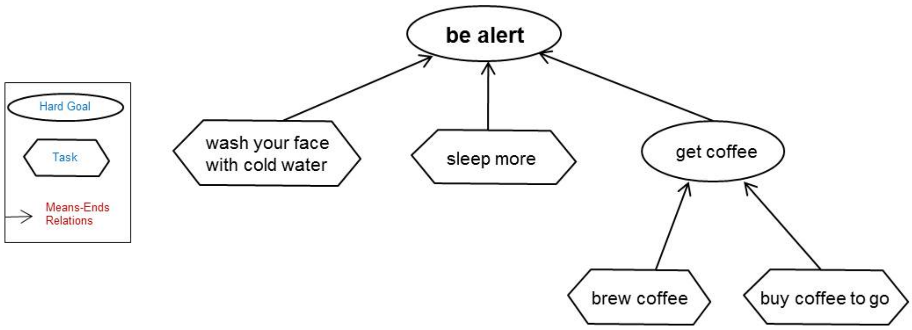


Figure 3.4: Means-Ends relations are exemplified by demonstrating Goal - Task relations as well as Subgoal - Task relations. (represented in i* notation [188])

Figure 3.4 shows a goal hierarchy excerpt of the human goal "`be alert`". As illustrated, tasks can be directly connected to goals as well as to subgoals. The task "`wash your face with cold water`" represents one alternative for achieving the higher-level goal. Similarly, two task alternatives are offered to accomplish the subgoal "`get coffee`".

3.1.1.6 Resource

Knowledge about human goals does not only encompass goals, subgoals or tasks. To accomplish goals or tasks, resources play an important role. A resource is defined to be an entity (either physical or non-physical) which is in one way or the other relevant for satisficing goals. Figure 3.5 illustrates the role of resources in the context of knowledge about the goal "`be alert`".

The high level goal, in this example "`be alert`", can be achieved by "`drinking coffee`". Two tasks offer alternatives on how to get some coffee to drink, i.e. "`buy coffee to go`" or "`brew coffee`". "`Coffee`" represents the critical resource in this example. The other resources can be regarded as optional, i.e. they might have little effect on the hard goal "`be alert`", but they might indirectly affect the soft goal "`enjoy good coffee`". Information about availability and criticality thus appears valuable in the context of goals and their resources.

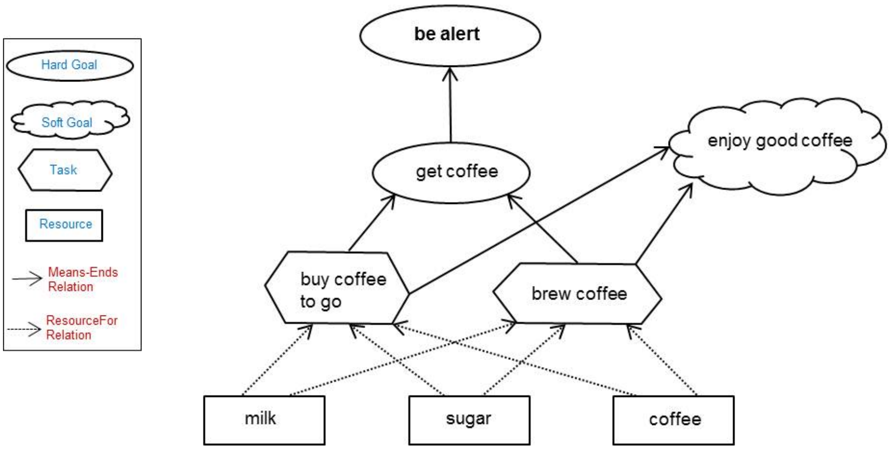


Figure 3.5: ResourceFor relations in the context of the high-level goal "`be alert`". In this example the goal state can be reached by "getting coffee". For "getting coffee", some resources are relevant, one of them critical – the coffee – the other more or less optional. The availability of these optional resources might have a positive effect on the soft goal "`enjoy good coffee`". This example is represented in i* notation [188].

3.1.1.7 Goal Category

A goal category groups together a set of goal instances which are are similar to each other. It thus represents a high-level goal, e.g. "`Physical Health`" and is characterized by a description which clearly defines the boundaries to other categories. Goal categories can form goal schemas which are mental constructs to organize goals, e.g. by arranging them in hierarchical structures. Goal schemas can be adopted from other domains such as psychology (cf. [30], [107] or [139]), web search (cf. [21], [146]) or requirements engineering (cf. [173]).

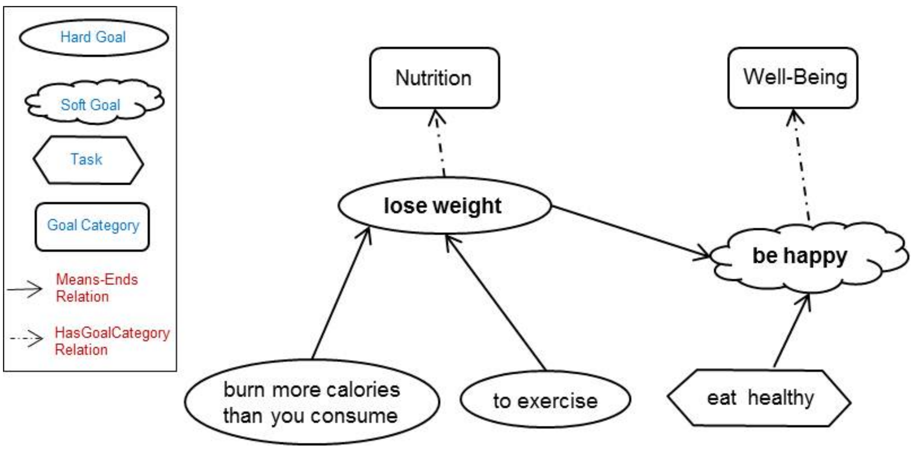


Figure 3.6: HasGoalCategory relations are exemplified by connecting the hard goal "`lose weight`" and the soft goal "`be happy`" to goal categories (taken from [30]). This example is partly represented in i* notation [188].

In the context of goal knowledge acquisition, theoretical work from psychology appears useful, for instance, the social-psychological theoretical framework by Ada Chulef [30]. Chulef's framework organizes the motivations and high-level goals of people including "`Family`", "`Religion`", or "`Physical Health`". In Figure 3.6, the hard goal "`lose weight`" is connected to the goal category "`Nutrition`" and the soft goal "`be happy`" to the goal category "`Well-Being`". Both categories are taken from Chulef's taxonomy of human goals. For knowledge acquisition purposes, research adapted machine learning techniques to automate the process of classifying human goals into goal categories. These methods were successfully applied in many areas including web search (cf. [78], [91] or [73]) or text analysis (cf. [87]).

3.1.1.8 Agent

An agent is defined as an intentional actor, e.g. the person who pursues the goal. From a knowledge acquisition viewpoint, the task is to automatically identify the goal's pursuer. In this context, a research direction called semantic role labelling (cf. [53]) appears to be contributing. Semantic role labelling refers to the identification of roles and their characteristics in textual resources. Approaches typically analyze the grammatical structure ranging from simply identifying a sentence's parts-of-speech to constructing complex parse trees of a text passage.

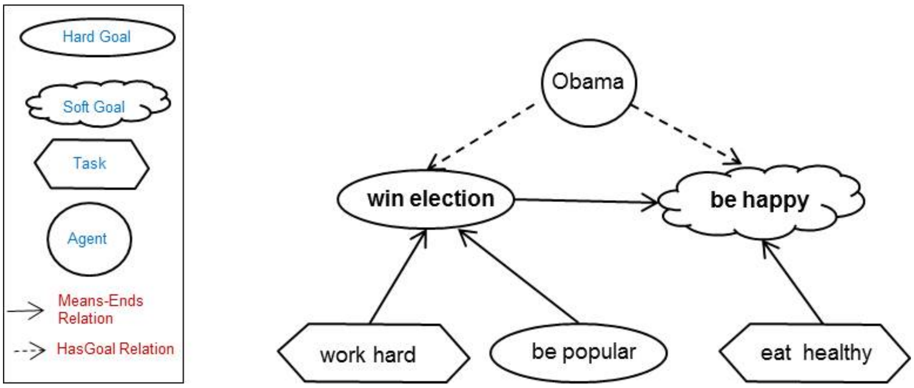


Figure 3.7: This toy example illustrates two goals an intentional actor, in this case Barack Obama, might have as well as means to accomplish them. (represented in i* notation [188])

The toy example in Figure 3.7 shows two goals an intentional actor, in this case Barack Obama, might pursue. The hard goal "`win election`" and the soft goal "`be happy`" are connected to "`Obama`" via HasGoal relations. Several means are presented which might contribute to accomplishing these goals. This example also demonstrates that explicating relationships between goal knowledge components is useful to recognize dependencies amongst an agent's goals, e.g. the accomplishment of one goal directly contributes to the accomplishment of another one.

3.1.1.9 Intentional Relation

Intentional Relations are introduced as relations that exhibit a clear goal-oriented focus. Figure 3.8 provides an overview of this work's intentional relation types, i.e. Means-Ends, HasGoal, ResourceFor, Conflict and HasGoalCategory. Relation types are exemplified by domain examples. By characterizing various steps in the process of reaching a goal state, all these relation types

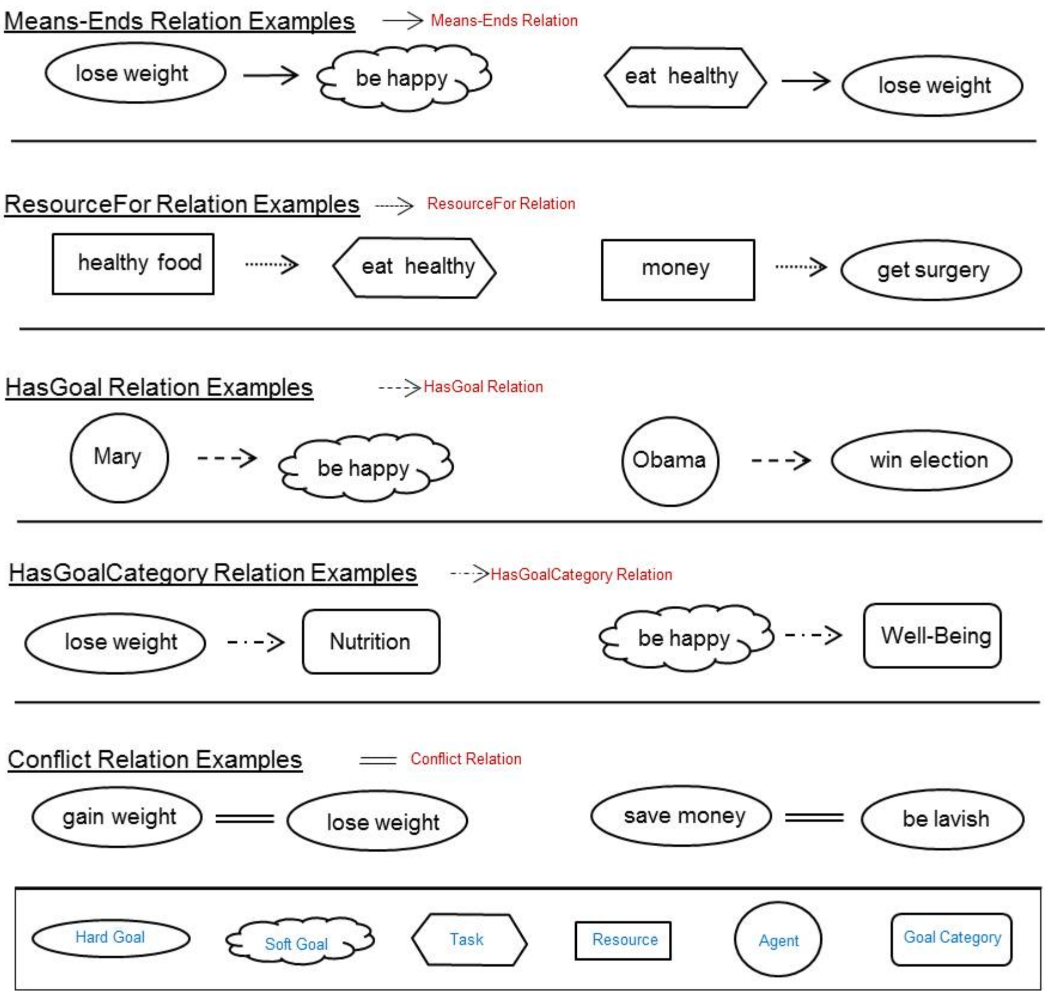


Figure 3.8: Types of intentional relations and domain examples represented in i* notation [188]. Goals have been assigned to goal categories according to Chulef's taxonomy on human goals [30].

contribute to describing and thus modeling knowledge about human goals. So far these relation types have been known and categorized as semantic relations. To better explicate the goal-oriented focus of these relation types, a novel characterization amongst relations appears justified which indicates their contribution to intentional knowledge. An intentional relation is thus defined

1. as a binary relation $R \subseteq A \times B$
2. to describe an intentional context and
3. to contribute to goal-oriented behaviour, i.e. relating goal components to each other.

Set A and set B contain varying intentional concepts depending on the intentional relation:

Intentional Relation	Set A	Set B
MeansEnds Relation	{Task, Goal}	{Goal}
ResourceFor Relation	{Resource}	{Task, Goal}
HasGoal Relation	{Agent}	{Goal}
HasGoalCategory Relation	{Goal}	{GoalCategory}
Conflict Relation	{Goal}	{Goal}

Except for the Conflict relation, the relation types can be further characterized to be direct, asymmetric and non reflexive binary relations. Each relation type features different characteristics and describes a different behaviour for modelling human goal knowledge. Related work on intentional relations is described in more detail in Section 2.1.3.

3.2 A Data Model for Human Goal Knowledge

Researchers have been attempting to continuously optimize the process of modeling and acquiring goal knowledge with respect to their domain. These separate attempts led to a multitude of perspectives on modeling and representing (human) goal knowledge. It appears that these models were often customized to serve specific research problems or to satisfy particular application requirements. As a consequence, it has become difficult, e.g. to analyze, evaluate or compare (human) goal knowledge across research prototypes or research groups. To alleviate this situation, this PhD thesis proposes a general data model which specifies the encoding of human goal knowledge. The data model is meant to provide a unifying view on modeling knowledge about human goals. To justify and inform the development process, existing modeling approaches from other research domains have been reviewed including goal-oriented requirements engineering (cf. [172], [41]) or intelligent agent theory (cf. [19], [174]). Informed by existing work, goal knowledge components have been adopted, adapted and complemented according to the introduced requirements. The resulting formal construct can be regarded as a template which has the potential to inform acquisition processes.

The Unified Modelling Language (UML) [17] is chosen to model knowledge about human goals. UML represents a general-purpose modeling language which can be used to describe the classes of a system and their relationships to each other. It has been already utilized to model knowledge (cf. [1]) and is accepted as standard in many areas, e.g. in software engineering. Other common techniques to model knowledge include CommonKADS [151], Multi-perspective Modelling (MpM) [27] or TELOS [117] which is mainly used in requirements engineering.

This work refrains from arguing completeness of the proposed data model with respect to modeling human goal knowledge. While the proposed data model remains open for extensions, the proposed version's usefulness has to a certain extent been validated in a series of case studies and application scenarios. Respective results indicate that the data model is reasonably complete and thus valuable for certain practical application scenarios.
Figure 3.9 shows an excerpt of the proposed data model on human goal knowledge modeled in UML. The remaining section provides an overview of the data model's classes and attributes.

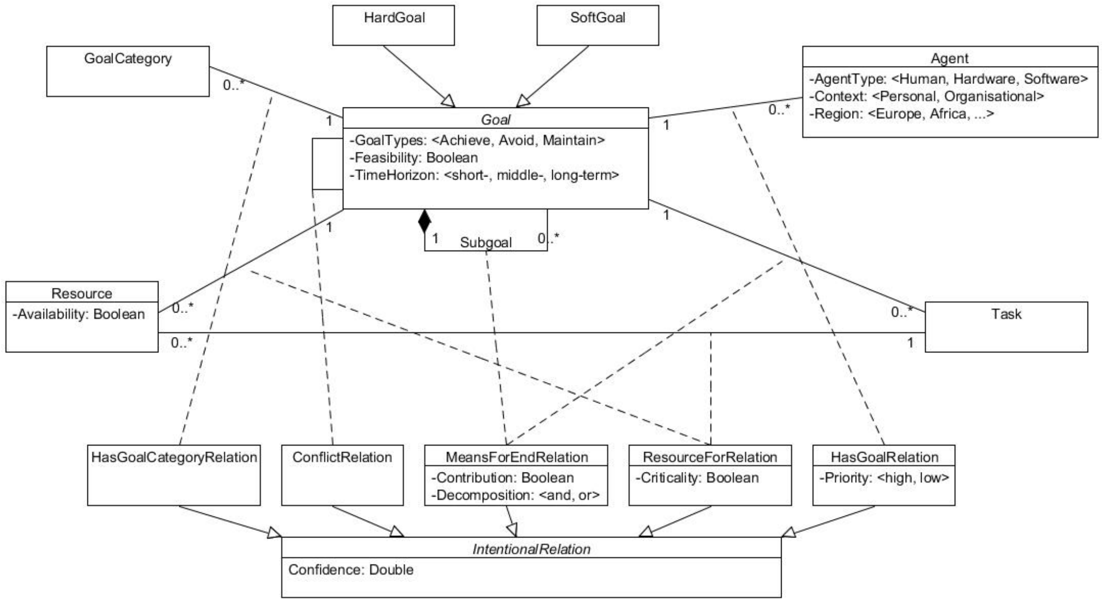


Figure 3.9: Structure of the proposed data model in UML notation. The class diagram contains and explicates essential components to model knowledge about human goals.

The **Goal** class represents one of the main components in the proposed data model. It denotes an abstract class implementable by following two classes: **HardGoal** and **SoftGoal** (cf. [187]). Figure 3.10 provides a data model excerpt of the abstract *Goal* class and its implementations.

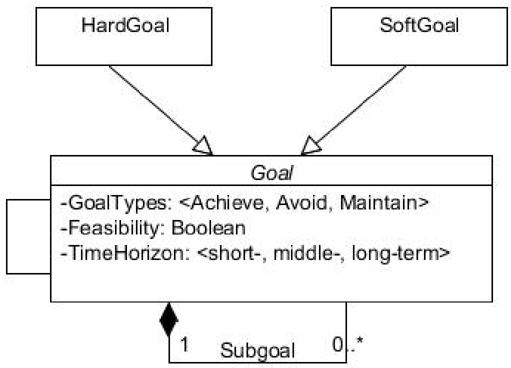


Figure 3.10: Data model excerpt of the abstract *Goal* class and its implementations.

Subgoals are the result of a decomposition process and are linked the *Goal* class via the *MeansForEndRelation* class. Decomposition structures represent a prerequisite for operations on goal knowledge such as planning or reasoning. In accordance with Nilsson [123], the *MeansForEndRelation* class specifies two decomposition types, i.e. "AND" or "OR".

The goal class is characterized by following attributes: (1) *GoalType*, (2) *Feasibility* and (3) *Time-Horizon.*

1. *GoalType* specifies whether a human goal is a goal to (a) achieve, (b) avoid or (c) maintain. Maintain goals represent states which an agent wants to preserve over a period of time, while achieve and avoid goals refer to a certain point in time where the corresponding goal state is achieved or avoided respectively. These goal classes express a temporal behavior which linear temporal logic describes as follows:

Goal Type	***Temporal Pattern***	***Temporal Behavior***
Achieve	$G \Rightarrow \diamond Q$	Q holds in some future state.
Cease	$G \Rightarrow \diamond \neg Q$	Q will not hold at some future state.
Maintain	$G \Rightarrow \Box Q$	Q always holds in the future.
Avoid	$G \Rightarrow \Box \neg Q$	Q will never hold in the future.
Optimize	$G \Rightarrow \diamond Q_{(max/min)}$	Q will hold in some future state with the additional constraint that an objective function related to Q is maximized or minimized.

Table 3.1: An overview of goal types in Requirements Engineering and their corresponding temporal behavior. (cf. [81])

where G and Q denote conditions and where $\diamond$ (a condition holds in some future state) and $\Box$ (a condition holds in all future states) are temporal operators in Table 3.1.

An AchieveGoal represents a goal an agent wants to achieve, i.e. it wants to reach a certain state. Goal instances include "`buy car`".

An AvoidGoal represents a goal an agent wants to avoid, i.e. it wants to avoid a certain state. Goal instances include "`fail exam`" or "`lose job`".

A MaintainGoal represents a goal an agent wants to maintain, i.e. the agent wants to uphold a certain state. Exemplary goal instances include "`be happy`" or "`be in a relationship`".

2. *Feasibility* states whether a goal can be accomplished or not from a present day perspective. Feasible goal instances include "`lose weight`" or "`buy car`". Infeasible goal instances include "`live forever`" or "`live on Mars`". The feasibility attribute does not take into account the agent's context, e.g. financial aspects necessary to accomplish a goal.

3. *TimeHorizon* informs within which time interval or at which time point the goal is in principle accomplishable: (a) short-term, (b) middle-term and (c) long-term. A short-term goal can be accomplished within several minutes to several days, e.g. "`feed dog`". A middle-term goal can be accomplished within several weeks, e.g. "`buy car`". A long-term goal can be accomplished within several months to several years or within a lifetime, such as "`start a family`" or "`find enlightenment`". This attribute does not take into account an agent's motivation to pursue the goal. It is remarked that without proper commitment, even the accomplishment of a short-term goal could take years.

The **Task** class is connected to the *Goal* class by the *MeansForEndRelation* class and to the *Resource* class by the *ResourceForRelation* class. A task is defined as an atomic action or activity which cannot be reasonably decomposed further. In a goal hierarchy, tasks therefore represent concrete actions at the bottom. Task instances include "`eat healthy`" or "`do cardio`".

The **Resource** class is connected to the *Task* and *Goal* class by the *ResourceForRelation* class. A resource is defined as either a physical or non-physical entity. Resources are characterized by following attribute (1) *Availability*.

1. *Availability* refers to whether a resource is available or not in the process of accomplishing a goal.

The **GoalCategory** class is a representative for a high-level goal such as "`Physical Health`" thus forming a set of similar goal instances. Each goal category is characterized by a description which clearly defines the boundaries to other categories. Goal categories can be organized by a schema which structures these categories and thereafter human goals. The *Goal* class is linked to the *GoalCategory* class via the *HasGoalCategoryRelation* class.

The **Agent Class** represents the intentional actor, e.g. the human agent who pursues a goal. The *Agent* class is linked to the *Goal* class via the *HasGoalRelation* class.
The *Agent* class is characterized by following attributes: (1) *AgentType*, (2) *Context* and (3) *Region*.

1. *AgentType* distinguishes different types of agents, e.g. whether the agent is human or artificial. The attribute denotes an agent to be a human agent, a hardware agent or a software agent agent.

2. *Context* refers to an agent's setting, e.g. whether the agent is motivated by personal, group or by organizational interests.

3. *Region* takes into account varying cultural backgrounds which can have effects, e.g. on an agent's preferences. To give an example, funeral ceremonies differ with respect to the culture. In Asia, the color of grief is white whereas in Europe and Northern America it is traditionally black. The cultural background thus might affect an agent's choice of means or its goals. The Region attribute currently encompasses seven categories, i.e. Africa, Asia, Europe, Australia, NorthAmerica, SouthAmerica and Other. If required, this categorization can be changed or be further split down, e.g. with respect to countries.

The data model defines an abstract **IntentionalRelation** class which can be implemented by following five classes: *MeansForEndRelation*, *ResourceForRelation*, *HasGoalRelation*, *HasGoalCategoryRelation* and *ConflictRelation*. These relation classes connect intentional concepts with each other including goals, subgoals or tasks. The abstract class is characterized by following attribute: (1) *Confidence*.

1. *Confidence* refers to a quality value which expresses relation's accuracy. The attribute can be regarded as normalized weight value with a range of [0,1].

The **MeansForEndRelation** class connects subgoals or tasks (means) to a goal (end) and is thus highly relevant in planning, reasoning or inference procedures.
It is characterized by following attributes: (1) *Contribution* and (2) *Decomposition.*

1. *Contribution* refers to a binary attribute, either positive or negative, characterizing the relationship between a means and an end. *Contribution* information is considered valuable, e.g. to inform reasoning processes (see Section 3.3.3 for an example).

2. *Decomposition* refers to whether a relation expresses an AND or an OR decomposition. Many factors or components can be involved in the process of accomplishing a goal. In this process, two types are considered to decompose a high-level goal: AND-decompositions and OR-decompositions. In case of OR decompositions, there exist several means which satisfy a respective goal. With reference to Figure 3.2, there are several means to accomplish the goal instance "`lose weight`" including "`get surgery`" or "`burn more calories than you consume`". AND decompositions, on the other hand, require all means to satisfy a respective goal. As an example, consider the goal "`plan wedding`" where means such as "`buy wedding dress`", "`send invitations`" or "`book church`" all equally contribute to accomplish the goal. The concept of AND/OR decompositions in a hierarchy structure was adopted from Nilsson [123].

The **ResourceForRelation** class connects the *Resource* class to the *Goal* class and to the *Task* class. This relation provides intentional concepts such as goals or tasks with resources. These resources might be essential in the overall process of reaching goal states. The relation's characterization thus suggests that a resource contributes to the accomplishment of a human goal.
It is characterized by following attribute: (1) *Criticality.*

1. *Criticality* refers to whether a resource is critical to a goal's accomplishment or not. Without these resources the goal cannot be accomplished, e.g. a surgeon is required in the process of "`get surgery`".

The **HasGoalRelation** class establishes a relation between the *Goal* class and the *Agent* class. It thus links a goal to an agent that intends to pursue the respective goal instance.
It is characterized by the following attribute: (1) *Priority.*

1. *Priority* describes whether the accomplishment of a goal is of (a) high or (b) low priority. An agent might want to quickly accomplish the goal "`find job`", i.e. assigning it a high-priority. In contrast, the goal instance "`clean up basement`" might not be time critical, thus having a low priority. The priority decision lies with the agent, independent of the respective goal's characteristics. In other domains such as requirements engineering the priority attribute is also known as urgency attribute.

The **HasGoalCategoryRelation** class connects the *GoalCategory* class to the *Goal* class. Goal instances can thus be assigned to goal categories.

The **ConflictRelation** class defines a reflexive relation of the *Goal* class. If two goal instances are in conflict with each other, they exert antagonistic effects on each other. Accomplishing both of them simultaneously is not possible. To give an example, the two goal instances "`lose weight`" and "`gain weight`" cannot be accomplished at the same time and are thus in conflict with each other.

3.2.1 An OWL Version of the Data Model

In addition to the proposed data model, this PhD thesis provides an OWL model to ontologize human goal knowledge. By providing a machine-readable OWL model, web services can access and utilize a shared conceptualization of human goal knowledge, e.g. for immediate usage in OWL-based reasoning systems. Semantic Web applications thus might benefit from the development of a goal ontology fostering cooperation between humans and computers.

To construct the OWL model, the Protégé ontology editor[3] is used, which is an open source ontology editor and a knowledge-base framework. OWL represents a family of knowledge representation languages and offers three sub-languages: a lite, a description logic-based (DL) and a full version to allow for different degrees of expressiveness, i.e. richness of semantic information. This work chooses the DL sub-language which guarantees computable conclusions and decidability.

OWL does not provide constructs to characterize classes by attributes. So called Value Partitions[4] need to be used as a surrogate for modeling class attributes. The following example explains the process of converting the class attribute *Goal.GoalType* which encompasses three options, i.e. Achieve, Avoid and Maintain, into a possible OWL construct. (i) Create a ValuePartition class *GoalType*. (ii) Create subclasses of the ValuePartition to represent possible options for the ValuePartition. In case of the *GoalType* class, *Achieve*, *Avoid* and *Maintain* classes are created. These option classes are disjoint from each other so that an individual cannot be a member of more than one of them. (iii) The list of value types has to be made exhaustive by introducing a covering axiom. A covering axiom can be introduced by filling the *GoalTypes'* equivalent classes with the respective options, i.e. *Achieve*, *Avoid* and *Maintain* classes. (iv) An object property for the *GoalType* class has to be created, in this case *hasGoalType*, to link the *Goal* class to its class attribute *GoalType*. (v) The domain of the property has to be set to the *Goal* class and the range to the *GoalType* class.

Object properties are in general used to connect two OWL classes to each other. These properties cannot be characterized further by attributes in OWL. In the proposed data model, intentional relation classes such as the *MeansForEndRelation* class are further described by attributes, i.e. Decomposition and Contribution. Since OWL does not support this kind of characterization,

[3] The Protégé project. http://protege.stanford.edu.
[4] ValuePartitions are not part of OWL, they are a "design pattern".

Data Model Classes	Ontology Classes
Goal	Goal
Task	Task
Resource	Resource
Agent	Agent
GoalCategory	GoalCategory

Table 3.2: Correspondance overview of data model classes and ontology classes.

attributes of intentional relation classes are not are not integrated into the provided OWL model. The following paragraphs elaborate on the ontology's classes and object properties which correspond to the data model's classes and attributes. Table 3.2 shows the one-to-one mapping of data model classes and ontology classes with respect to *Agent*class, *GoalCategory* class, *Goal* class, *Task* class and *Resource* class. Corresponding data model attributes are modeled as ValuePartition classes.

The **Goal** class represents one of the main components in the ontology. It denotes an abstract class implementable by following two classes: **HardGoal** and **SoftGoal** (cf. [187]). Subgoals are the result of a decomposition process and are linked to the *Goal* class via the *isMeansForEnd* object property class. Corresponding partition value classes are *GoalType*, *Feasibility,* and *TimeHorizon*.

- The *Goal* class is linked to its goal type attribute via the *hasGoalType* property.
- The *Goal* class is linked to its feasibility attribute via the *hasFeasibility* property.
- The *Goal* class is linked to its horizon attribute via the *hasTimeHorizon* property.

Subclasses of *GoalType* are *Achieve*, *Avoid* and *Maintain*. Subclasses of *Feasibility* are *Feasible* and *NotFeasible*. Subclasses of *TimeHorizon* are *ShortTerm*, *MiddleTerm* and *LongTerm*.

The **Task** class is connected to the *Goal* class by the *isMeansForEnd* property class.

The **Resource** class is connected to the *Task* and *Goal* class by the *isResourceFor* property class. The corresponding partition value class is *Availability*.

- The *Resource* class is linked to its availability attribute via the *hasAvailability* property.

Subclasses of *Availability* are *Available* and *NotAvailable*.

The **Agent** class represents the intentional actor, e.g. the person who pursues a goal. The *Agent* class is linked to the *Goal* class via the *hasGoal* property class. Corresponding partition value classes are *AgentType, Context* and *Region*.

- The *Agent* class is linked to its context attribute via the *hasContext* property.
- The *Agent* class is linked to its agent type attribute via the *hasAgentType* property.
- The *Agent* class is linked to its region attribute via the *hasRegion* property.

Subclasses of *AgentType* are *HumanAgent*, *SoftwareAgent* and *HardwareAgent*. Subclasses of *Context* are *PersonalContext*, *GroupContext* and *OrganizationContext*. Subclasses of *Region* are *RegionAfrica*, *RegionEurope*, *RegionAsia*, *RegionNorthAmerica*, *RegionSouthAmerica*, *RegionAustralia* and *RegionOther*.

The **GoalCategory** class is a representative for a high-level human goal and is a container for related human goals. The *Goal* class is linked to the *GoalCategory* class via the *hasGoalCategory* property class.

In contrast to other object-oriented knowledge representation languages, OWL uses object properties to describe relationships between two individuals. The ontology specifies an **isIntentionalRelation** object property which acts as a super class for following object properties: *isMeansForEnd*, *hasGoal*, *hasGoalCategory*, *isResourceFor* and *isInConflictWith*.

Table 3.3 gives an overview of corresponding data model relation classes and ontology object properties.

Data Model Relation Classes	Ontology Object Properties
IntentionalRelation	isIntentionalRelation
MeansForEndRelation	isMeansForEnd
HasGoalRelation	hasGoal
ResourceForRelation	hasResourceFor
HasGoalCategoryRelation	hasGoalCategory
ConflictRelation	isInConflictWith

Table 3.3: Correspondance overview of data model relation classes and ontology object properties.

The **isMeansForEnd** object property connects intentional concepts with each other including goals, subgoals or tasks. Typically, it connects a means, e.g. a subgoal or a task, to an end, e.g. a goal.

The **isResourceFor** object property connects resources with intentional concepts including goals, subgoals or tasks.

The **isInConflictWith** object property connects two conflicting goal instances, i.e. instances which exert antagonistic effects on each other.

The **hasGoal** object property establishes a relation between a goal and an agent pursuing the goal.

The **hasGoalCategory** object property establishes a relation between a goal and a corresponding goal category.

3.3 Operations on Human Goal Knowledge

The value of modeling and acquiring human goal knowledge does not only lie in the explication of real-world knowledge alone. Additional value lies in the application of operations on human goal knowledge to support decision making processes, e.g. of intelligent agents or of intelligent systems. Acquiring sufficient amounts of goal knowledge paves the way for knowledge engineers to benefit from this type of knowledge.

Describing the full range of applicable operations is beyond this work's scope. The focus is on operations (i) which are known to be standard operations in the area of artificial intelligence and (ii) which have already been used to implement goal-oriented behavior. This section thus elaborates on a selective set of operations, i.e. searching, reasoning and planning, and discusses them in the context of human goal knowledge, i.e. which components are necessary or useful in terms of an operation's success.

3.3.1 Search

The concept of "search" often is associated with the retrieval of information. Search engines act as gateways to access and retrieve high-quality information, e.g. from the web. This section, however, discusses search processes and search strategies from an artificial intelligence viewpoint, not from an information retrieval one. According to Norvig and Russell [125], "*Search and planning are the subfields to AI devoted to finding action sequences that do achieve the agent's goals*". To do that, it is helpful to think of the search process as constructing a search tree. How the search tree is traversed depends on the selected search strategy. There exist several search strategies such as breadth-first or depth-first which can be characterized by their completeness, optimality, time and space complexity. Each search strategy has its own strengths and weaknesses with respect to the use case. To give an example, depth-first search is impractical in search trees of large depths; breadth-first requires a lot of (often too much) memory. Bidirectional search strategies can greatly reduce the time complexity. These strategies simultaneously progress forward from the initial state and backward from the goal state to output an action sequence. The process stops when a branch of one search process meets a branch of the other.

To implement AI's perspective on search in its simplest way, the data model's *Goal, Task* and *MeansForEndRelation* classes are relevant. Finding action sequences then translates to moving along Means-Ends relations from instantiated tasks and subgoals to accomplish a given goal. The data model, however, encodes additional characteristics of human goal knowledge which can inform the search process. Accomplishing a goal might require certain resources indicated by the data model's *Resource* and *ResourceForRelation* classes. In case of a multitude of goals to accomplish, data model components such as the *Resource.Availability* attribute, the *HasGoalRelation.Priority* attribute or the *ConflictRelation* class can contribute to the search process' decisions.

3.3.2 Planning

Planning describes the process of generating action sequences to achieve a predefined goal state (cf. [125]). These action sequences can be regarded as "how to" instructions partly specifying also the required ordering of actions to be performed. Planning has a long tradition in artificial intelligence and was partly inspired by research in robotics; the STRIPS program, for instance, was designed to control Shakey, a little robot, roaming the premises of the SRI research institute in the 1970s. According to Norvig and Russell [125], "*[...] planning comes down to an exercise in finding a language which is just expressive enough for the problems you want to solve, but still admits a reasonably efficient algorithm. [...]*" Traditionally, planning algorithms require formal representations of states to be in and actions to execute. Situation calculus was one of the first logical representations which planning algorithms were capable of processing. Yet, restrictions of situation calculus led to the development of the STRIPS language [48] which describes actions in terms of preconditions and their effects. Extensions of STRIPS are successfully applied to real-world scenarios nowadays, for example, in surveying and controlling space missions.

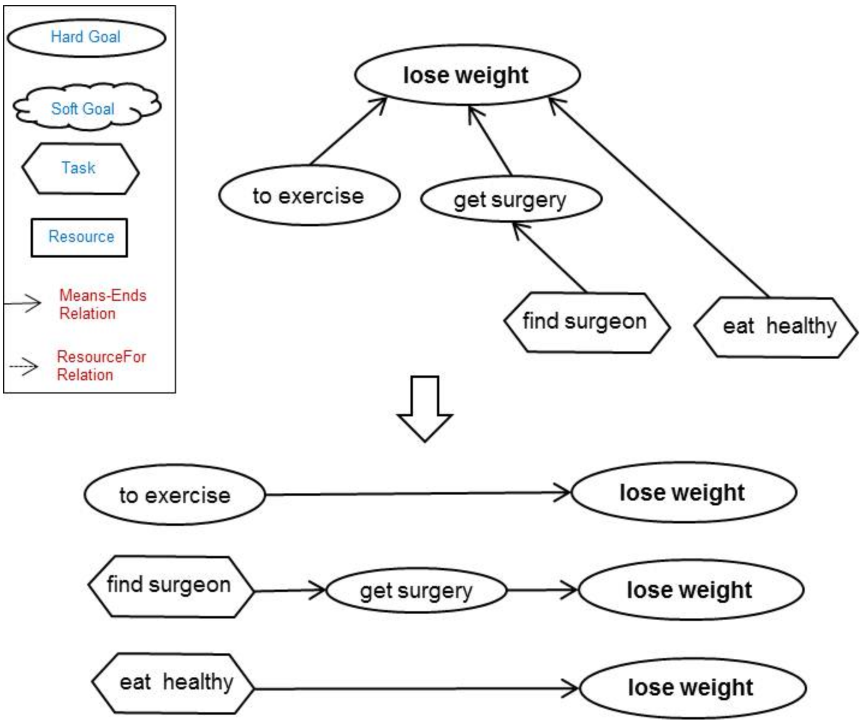


Figure 3.11: A simple toy example representing knowledge about the human goal "`lose weight`" and a respective linearization which results in alternative paths through a potential problem space. (represented in i* notation [188]).

Forward and backward planning refer to the method of Means-Ends analysis [140]. This analysis identifies differences between the current state and the goal state. From a repertoire of possible actions the agent needs to select one to reduce this difference. Forward planning refers to an agent which examines all currently possible actions in an effort to walk forward through the problem space to the goal state. Yet, if there are lots of actions, searching for a solution starting from the initial state appears to be hopeless. Backward planning could alleviate this situation. It refers to an agent which is already at the goal state and examines possible actions to go backward through the problem space. The decision to utilize backward or forward planning can be made dynamically by a weighted counting of the number of actions relevant to the current state and the same weighted counting of the number of actions that will result in the goal.

Figure 3.11 illustrates a linearization of alternative paths to accomplish the human goal "`lose weight`". These alternative paths can also be considered as paths through a problem space where the planner would have to make appropriate decisions. Forward planning would correspond to a situation where a planning algorithm chooses between alternative paths, in this case choosing amongst the "exercise path", the "surgery path" or the "diet path". Backward planning would correspond to a situation where a planner starts at the goal state and examines incoming paths.

The proposed data model provides additional characteristics on goal knowledge components. Its classes and their attributes such as the *Goal* class, the *Task* class, the *Resource* class and the *MeansForEndRelation* class can further inform planning algorithms in their decision making processes. The *MeansForEndRelation.Contribution* attribute provides information on whether a means will positively or negatively contribute to an end. If several goals are involved in the planning process, the *ConflictRelation* class, the *Resource.Availability* attribute or the *ResourceForRelation.Criticality* attribute can indicate potential obstacles and bottlenecks. A resulting plan can thus be fine-tuned to fit the respective situation.

With respect to goals and goal knowledge, planning examples from other domains include hierarchical task networks and commonsense knowledge. The hierarchical task network (HTN) is an approach to automated planning and has been used by a variety of planners, e.g. SHOP2 [121]. An HTN represents a network of tasks which are organized according to their dependencies in a hierarchy. At the bottom of the hierarchy, primitive tasks, i.e. simple actions, are located whereas complex goal tasks are rather located at the top. Complex and primitive tasks have their equivalents in STRIPS, i.e. goals and actions, yet HTN are capable of expressing compound tasks as well. An agent that has access to such task structures can generate and execute plans more effectively since the HTN knowledge specifies how to decompose complex tasks into simpler ones. Nejati et al. [122] present DLIGHT, a goal-directed incremental algorithm, to learn successful solution strategies for the *Escape* domain. DLIGHT's contribution is to produce more balanced solutions, i.e. solutions less specific than explanation based learners and less general than LIGHT's, DLIGHT's predecessor. While HTNs provide an expressive, formal language, human goal knowledge often is represented informally, i.e. by semi-structured natural language as in ConceptNet [104]. This form of representation allows only for a statistical approach to planning. Statistical planning algorithms operate on a set of associations between sequences of actions and statements of goals. In Lieberman and Espinosa [100], statistical planning mechanisms are applied to facilitate human computer interaction in the area of consumer electronics. For a given user goal, for instance, "`listen to a CD`", their system provides a set of partially ordered actions such as first "`turn the recorder on`", second "`open the CD player door`" and so on.

3.3.3 Reasoning

Reasoning or inference in its traditional form is a subfield of logics and describes the process of reaching a conclusion which is based on already existing data (cf. [125]). The reasoning quality depends to a certain extent on the type of knowledge representation; the more formal the representation, the stronger drawn inferences, i.e. the reasoning quality. While formal representations, e.g. first order logic, allow for deductive reasoning, more informal representations such as semi-structured natural language allow for a weaker form of reasoning including graph reasoning methods like spreading activation [37] or network traversal. To give an example, ConceptNet [103] provides three implementations of flexible inference: *Context Finding*, *Inference Chaining* and *Conceptual Analogy*. After mapping a respective query into ConceptNet's structure, the network is traversed; in case of *Context Finding*, all nodes which are one hop away are returned regardless of their relation type. In case of *Inference Chaining*, the network is traversed to identify a path between two concepts such as "`buy food`" and "`fall asleep`".

Norvig [125] discerns two types of reaching a conclusion. The first type is denoted as data-driven, forward chaining or forward reasoning. This type starts with all known data and progresses toward possible conclusions. The second type is denoted as goal driven, backward chaining, or backward reasoning. In this case, the reasoning mechanism selects a possible conclusion, i.e. a conclusion candidate, and attempts to prove its validity by looking for supporting evidence. The second reasoning type is best known for solving diagnostic problems where the number of possible conclusions is small. In the context of goal knowledge, reasoning can also be divided into (i) top-down and (ii) bottom-up reasoning. Figure 3.12 outlines these two modes of operation by means of the human goal "`lose weight`". This example also demonstrates the potential of goal components – in this case *SoftGoal* – to act as preferences within reasoning processes, i.e. to inform (i) top-down or (ii) bottom-up reasoning. In top-down reasoning, the contribution on soft goals is taken into account, i.e. paths are favoured which positively contribute to soft goals.

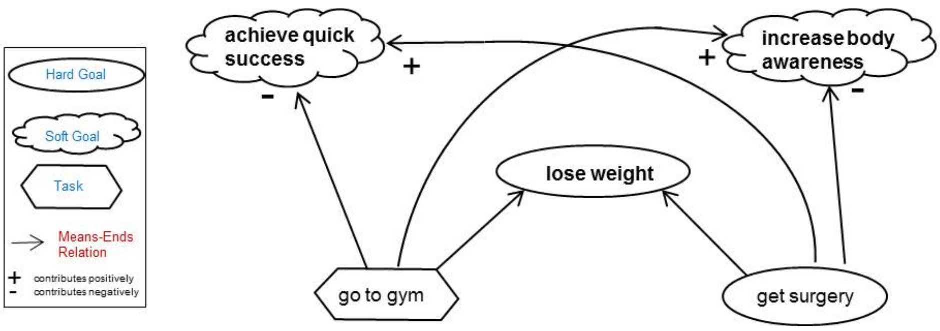


Figure 3.12: A simple toy example representing knowledge on the human goal "`lose weight`" represented in i* notation [188]. Means-Ends relations indicate either a positive (helps) or a negative (hurts) contribution. This information can be utilized in (i) top-down or (ii) bottom-up reasoning.

Let's assume a person's goal is to "`lose weight`" while "`increasing her body awareness`". In this example, "`get surgery`" does not contribute positively to the soft goal "`increase body` `awareness`", so the reasoning mechanism would refrain from selecting this path in favor of the "`go to gym`" path. In contrast, bottom-up reasoning starts from low hierarchy levels (tasks in the proposed data model). This kind of reasoning traverses the hierarchy upwards and examines the impact on soft goals if particular pathways were selected. In the example, this would mean to select a starting point, e.g. "`get surgery`" and examine the impact on possible hard and soft goals. According to Figure 3.12, this choice would negatively contribute to the soft goal "`increase body awareness`", yet it would positively contribute to the soft goal "`achieve quick success`".

3.3.3.1 Prediction as a Form of Reasoning

Prediction can be defined as making a statement about something that has yet to happen. Time series analysis [177] is one of the first fields devoting their research on "understanding the past to forecast the future". By analyzing data characteristics at hand, time series analysis seeks to extract repeating patterns to make more or less accurate predictions. Time series analysis has been found useful for many applications including economic forecasting, stock market analysis or utility studies.

In the context of human goal knowledge, "prediction" can be understood as predicting a person's goal. The rationale for predicted goals can be based (i) on a person's behavior, e.g. on a series of performed actions, or (ii) on a set of resources a person utilized. Thus, the rationale for the decision is based on already observed data upon that a conclusion is reached. Goal predication can thus be regarded is special form of forward reasoning about goal knowledge.

3.4 The Meta-Process of Engineering Human Goal Knowledge

This section elaborates on the process of engineering knowledge about human goals. The process starts with an Elicitation step, i.e. familiarizing with the domain and ends with an Operationalization step, i.e maintaining and updating acquired knowledge. This process is partitioned into five distinct processing steps (i) Elicitation, (ii) Acquisition, (iii) Validation, (iv) Interpretation and (v) Operationalization based on related work on knowledge engineering.

The introduced meta-process on goal knowledge deviates from standard knowledge engineering processes (cf. [105], [165] and [166]) by focusing on engineering human goal knowledge, i.e. specifying respective inputs and outputs. In addition, each processing step discusses the respective role of goal knowledge as well as potential ways of dealing with its characteristics.

The remaining section describes these steps which are summarized and briefly explained in Figure 3.13. The meta-process starts with the *Elicitation* step to get familiar with the target domain. This includes developing an understanding of how goal knowledge is defined and expressed. This understanding is encoded by the proposed data model on human goal knowledge. The *Acquisition* step encompasses all common approaches to explicate goal knowledge according to the data model. It outputs unchallenged goal knowledge which has to be qualitatively assessed

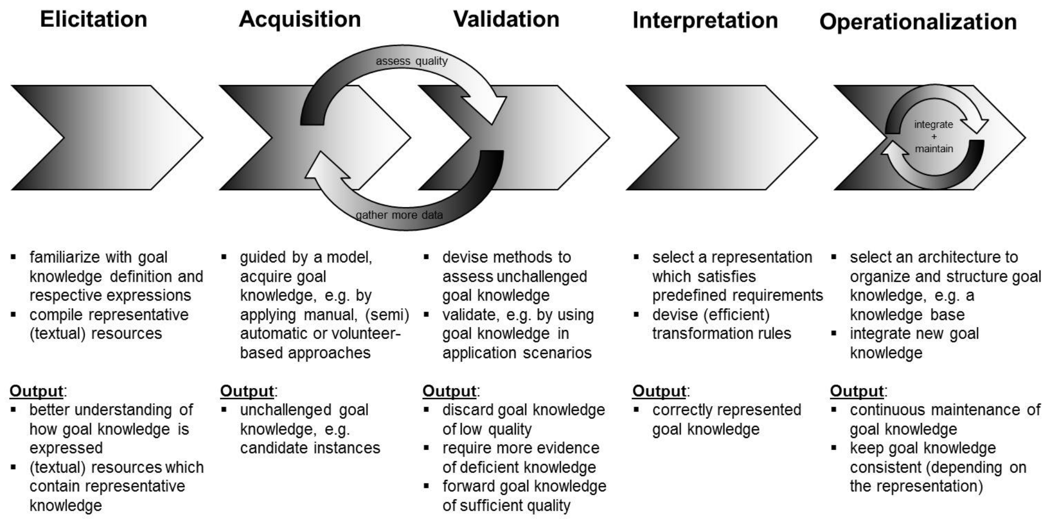
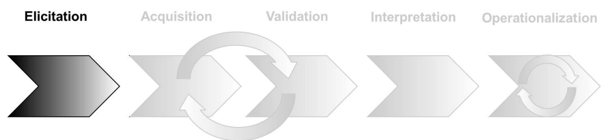


Figure 3.13: An overview of the meta-process on goal knowledge. It consists of five processing steps: 1.) Elicitation, 2.) Acquisition, 3.) Validation, 4.) Interpretation and 5.) Operationalization. (based on Sure's Knowledge Meta Process [166])

in the *Validation* step. The *Validation* step discards or forwards goal knowledge. In case of deficient instances, this step requires additional evidence with respect to the instance's quality. In the *Interpretation* step, high quality goal knowledge is translated into the representation language selected by the knowledge engineer. The *Operationalization* step is in charge of keeping the knowledge operable, i.e. systems should be able to access, process and utilize captured goal knowledge. This comprises the organization and structuring as well as maintenance of goal knowledge.

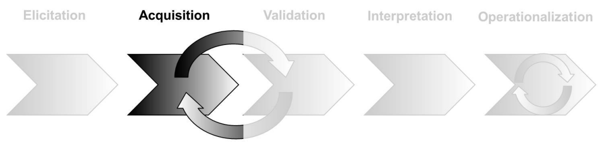


3.4.1 Elicitation

Elicitation serves the knowledge engineer to familiarize herself with the domain from which goal knowledge shall be acquired or captured. In [165], the objective of the elicitation step is to *"acquire informal descriptions of the knowledge about the specific domain"*. In early knowledge engineering attempts, this was done by conducting structured interviews resulting in knowledge protocols which contained domain-relevant knowledge. These interviews primarily served to get a better understanding of domain-relevant definitions and expressions of domain knowledge. In addition, they also represented textual resources describing the domain for subsequent acquisition purposes.

Human goal knowledge is different since it potentially covers a wide spectrum of domains. So, instead of getting familiar with a single domain knowledge engineers need to get familiar with a community's understanding of human goal knowledge. To gain this familiarity, alternative options exist which can be divided into manual and automated. In manual approaches, knowledge engineers interview large numbers of people about their personal goals and ways they think to accomplish them. In addition, personal writings such as letters or dairy entries are gathered and then analyzed. Yet, manual approaches are known to be time consuming and tedious. The advent of the social web allows automating the process of gathering personal information and decent amounts of resources about human goal knowledge, e.g. from websites such as "43Things.com", "jig.com" or "goalmigo.com". According to Lenhart [94], people tend to share information about their lives including their goals on the social web. Social media applications emerge replacing traditional media such as dairies with digital ones such as weblogs.

The result of the *Elicitation* step is a better understanding of how goal knowledge is defined and expressed. Rather than focusing on a domain, this understanding could focus on a particular subpopulation or on a particular type of medium such as social networks.

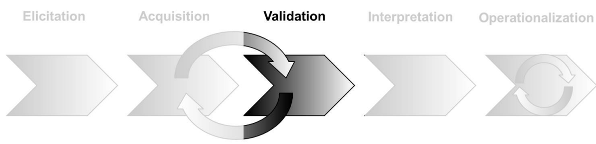


3.4.2 Acquisition

With a deepened understanding on human goal knowledge, this step comprises the actual acquisition, i.e. the population or instantiation of an abstract model. In this case, the data model informs us by specifying how goal knowledge is to be encoded but leaves the how to the knowledge engineers. In the context of goal knowledge, different strategies can be applied including human knowledge engineering (cf. [92]), volunteer-based (cf. [155]), game-based ([101] or [89]) or semi-automatic approaches (cf. [43]). All these approaches have in common that first a common understanding needed to be established on how the respected goal knowledge is to be expressed. In case of game-based approaches, this understanding transitions into the design of the game, e.g. what kind of questions to ask, or in case of semi-automatic approaches into the structure of extraction patterns.

To be of value, the amount of goal knowledge available needs to be substantial. Similar to [39] or [105], this work proposes to automate the acquisition process, i.e. reducing human participation in the process. Other research areas have been developing and successfully applying automated techniques to acquire goal knowledge. These research areas include (i) situation awareness (cf. [148], [77]), (ii) semantic task retrieval (cf. [51]), (iii) requirements engineering (cf. [124]), (iv) human language technologies (cf. [167]), (v) commonsense knowledge acquisition (cf. [154], [22], [57]) or (vi) information retrieval (cf. [91], [186]). Automated approaches apply a wealth of techniques including machine learning, information extraction and natural language processing. Yet, even with automated approaches human participation can only be minimized to a certain degree. Machine learning techniques largely require annotated samples to construct classification or clustering models. Similarly, information extraction techniques often apply hand-crafted extraction patterns. To further advance the automation of knowledge acquisition, a lot of work has been dedicated recently including (i) self-supervised IE, (ii) Open IE and (iii) “life-long” or “never-ending” learning. Especially the concept of “never-ending” [25] or “life-long” [10] learning appears promising in this context since it refers to an open-ended effort to extract information from the web.

In the end, the knowledge engineer has to decide on a suitable approach. This decision might be affected by several factors including familiarity with techniques, quantity or quality of (textual) resources. The *Acquisition* step outputs goal knowledge, e.g. a list candidate instances, whose quality is evaluated in the next step.



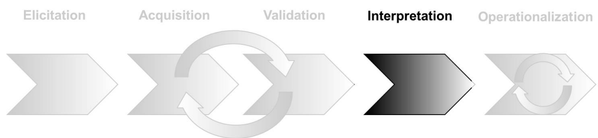
3.4.3 Validation

The *Validation* step’s task is to assess the quality of acquired goal knowledge. Each research community applies their own procedures and metrics to assess knowledge validity. Common metrics are metrics such as precision, recall and F-Score (cf. [8]), accuracy, inter-rater agreement κ [35] or taxonomic overlap [106], which all represent standard metrics for measuring quality in domains such as information retrieval, machine learning, information extraction or natural language processing.

Strategies to evaluate or validate knowledge range from manual to automatic approaches. Manual approaches are characterized by high quality and by a fairly low throughput which appears disadvantageous in a web-scale context. Automatic strategies require less human participation, i.e. higher throughput often at the expense of quality. A method that has recently gained popularity uses the so called “wisdom of crowds”, i.e. the accumulated knowledge of many people. The evaluation is then based on statistics which encode and represent the crowd’s wisdom. Such statistics can be generated, e.g. by analyzing search engine results (cf. [33]) where the search engine provides access to the crowd’s wisdom. Human computation (cf. [176]) and crowdsourcing approaches form another category in between manual and automated strategies. The underlying idea is to utilize human resources for quality assessments, i.e. to motivate them to actively participate in the assessment process. Motivations can range from entertainment aspects, e.g.

playing a game, to financial aspects, e.g. small amount of money for every assessed instance; a business model pursued by *MechanicalTurk*. A representative for this approach is Common Consensus [101] which is a web based game to acquire commonsense knowledge. It is based on the idea of "Games with a Purpose" (cf. [176]) where users contribute knowledge online by playing a game. Since the interface mimics the popular TV game Family Feud, goal knowledge is acquired in a question-answer manner. Each question addresses a different aspect, e.g. synonymy, a Means-Ends relation or locality.

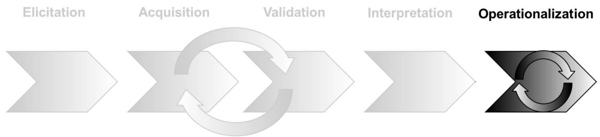
The *Validation* step decides (i) whether goal knowledge is discarded or (ii) whether additional evidence is required to make a final decision. Validated goal knowledge meets all quality criteria and is forwarded to the next step.



3.4.4 Interpretation

In this step, goal knowledge is transformed into the target representation language which can range from formal to informal approaches. Each approach has its strengths and weaknesses which should be taken into account (see Section 2.1.4 for a detailed comparison). The respective goal knowledge then has to be transformed according to the representation language's specifications. The complexity of the transformation procedure depends on the representation. The degree of human participation is therefore directly related to the representation language. Formal approaches such as logic are rather complex and require a careful transformation. Automating this transformation turns out to be complicated, yet there is ongoing research to tackle this issue (cf. [154]). Informal approaches do not demand such a tight formal corset and are therefore more amenable to automation. To give an example, [61] uses a semi-structured, natural language representation of concepts. This representation can be easily achieved by applying a series of natural language processing techniques such as stop word removal or stemming.

The *Interpretation* step outputs a correct representation of human goal knowledge with respect to the chosen representation language's specification, e.g. description logic.



3.4.5 Operationalization

The last step of the meta-process comprises techniques to make the acquired goal knowledge operable. Making and keeping goal knowledge operable allows (intelligent) systems to access, to process and to utilize the stored goal knowledge. Two strategies exist to store knowledge: (i) the symbolic and (ii) the connectionist approach which allow different perspectives on knowledge. The connectionist approach resembles the way humans memorize knowledge, i.e. a network of neurons where the actual knowledge is represented by connection strengths. Yet, the connectionist approach bears some weaknesses when it comes to digitally update, search or maintain knowledge. Thus, this section focuses on symbolic approaches which include storing knowledge, e.g. in knowledge bases. In a knowledge base, goal knowledge components are put into relation to each other and thereby structured, e.g. into hierarchies. These relations and structures then allow intelligent systems to operate on goal knowledge, e.g. to reason about human goals. A knowledge base is an umbrella term, i.e. a knowledge base can contain knowledge expressed in various representation languages. Erik Mueller's Thought Treasure [113] is an example for a commonsense knowledge base which uses logic, finite automata, grids, and scripts to represent knowledge. The representation language influences updating and maintaining operations. To give an example, formal approaches such as monotonic logics require the knowledge base to remain consistent, i.e. a statement to be added must not be in conflict with existing statements and reasonings.

PART III: CASE STUDIES & APPLICATIONS

Three case studies are conducted to operationalize knowledge about human goals as well as to instantiate and to use selected components of the proposed data model. The case studies also serve to illustrate and validate the introduced framework, i.e. the data model and the meta-process on engineering human goal knowledge.

Chapter 4's case study explores real-world characteristics of human goal instances. For that purpose, large amounts of goal instances are extracted from search query logs thereby instantiating the data model's *Goal* class. These instances are quantitatively and qualitatively analyzed to learn more about their nature. In addition, a comparative analysis is conducted to study query log goals with respect to commonsense characteristics. Chapter 4 is based on (Strohmaier and Kröll [161]) and (Strohmaier et al. [162]).

Chapter 5's case study seeks to better understand human goal knowledge in natural language text. It therefore analyzes textual resources from a goal-oriented perspective, i.e. relating textual passages to goal categories. The analysis process uses two mappings (i) from people's tasks to people's goals (the data model's *Task* and *Goal* Class) and (ii) from people's goals to people's goal categories (the data model's *Goal* and *GoalCategory* Class). The quality of the automated mapping process is evaluated by analyzing and comparing political speeches from an intentional perspective. Chapter 5 is based on (Kröll et al. [86]) and (Kröll and Strohmaier [87]).

Chapter 6's case study focuses on extracting Means-Ends relations thereby instantiating the data model's *MeansForEndRelation* class. In the first part of the case study hierarchical relations between goal concepts are automatically inferred. To construct a hierarchy of goal concepts, well-established clustering techniques are applied. In the second part this case study presents an automated method to complement goal hierarchies by Means-Ends relations, i.e. relating goals to tasks which potentially contribute to their accomplishment. Chapter 6 is based on (Kröll et al. [85]).

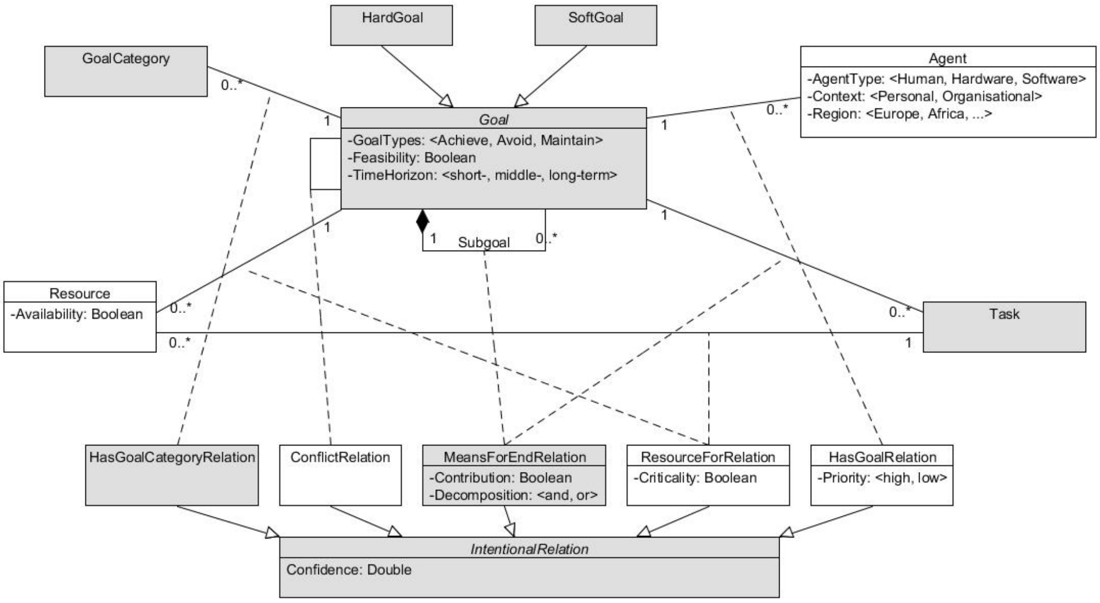


Figure 3.14: Overview of instantiated data model components. Shaded parts are covered by this work's case studies.

The presented case studies explore approaches to automatically instantiate selected parts of the data model (see Figure 3.14), thus operationalizing knowledge about human goals. By doing so, these case studies contribute to illustrating the versatility of goal knowledge (i) across research domains such as commonsense knowledge acquisition or textual analysis, and (ii) across textual resources ranging from resources exhibiting poor grammatical structures, e.g. search queries, to resources with rich grammatical structures such as political speeches or blog posts.

Chapter 7 demonstrates the value of acquired human goal knowledge in three practical problem settings. (i) In case of commonsense knowledge acquisition: exploring methods to complement ConceptNet [104]. (ii) In case of information retrieval: exploring methods to make a person's search query more explicit. (iii) In case of visual analytics: exploring possibilities to visually evaluate text from a goal-oriented perspective. Chapter 7 is based on (Strohmaier and Kröll [161]), (Kröll et al. [86]), (Jeanquartier et al. [75]) and (Strohmaier et al. [162]).

The case studies and the application scenarios thus demonstrate the framework's value and potential in various problem settings thereby indicating its generality.

Chapter 4

Case Study 01: Acquiring Goal Knowledge from Search Query Logs

4.1 Contributions to PhD Objectives

The first case study explores the feasibility of automatically instantiating the *Goal* class, the main component of the proposed data model on human goal knowledge as illustrated in Figure 4.1.

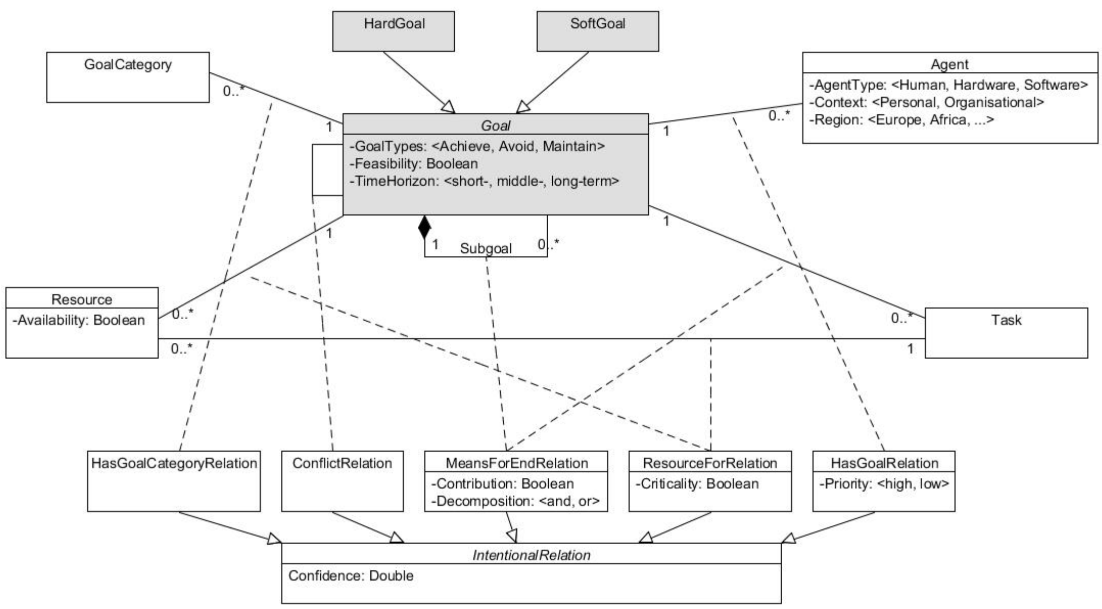


Figure 4.1: Components of the proposed data model covered by this case study are highlighted.

Covered Components	**Description**
Goal, HardGoal, SoftGoal	Most of the identified goal instances can be classified as HardGoal instances since objective success criteria can be applied. HardGoal instances include, for instance, "`buy house`" or "`get pregnant`". SoftGoal instances include "`be happy`" or "`be funny`".

This case study further validates the framework by investigating the agreeability and practicability of this PhD thesis' definition of a "human goal". It thus contributes to operationalizing human goal knowledge. A sufficiently practicable definition is a prerequisite for devising algorithms to automatically recognize and eventually extract human goal instances. This case study explores search query logs as a candidate resource for harvesting large amounts of human goal instances since every search query can be considered to be an expression of a person's (search) goal. The characterization of query logs from an intentional perspective might motivate further research in other areas including search intent (cf. [21], [58], [69]), semantic retrieval (cf. [120], [51]) or goal-oriented search interfaces (cf. [157], [104]).

This case study makes following contributions to the understanding of human goal knowledge which eventually benefits its formalization process.

1.) The case study examines the agreeability and thus practicability of the definition on human goals which is a prerequisite for devising algorithmic approaches to acquire human goal knowledge. For that purpose, a human subject study is conducted where human annotators are required to manually classify 3000 search queries into two classes: queries containing an explicit goal and queries which do not. The resulting inter-rater agreement Kappa [35] between human subjects is high which is indicative for agreeability and thus also for the practicability of the definition. In addition, the manual classification task informs about the class' characteristics and boundaries (cf. [21] or [146]).

2.) The case study provides an implementation of a classification algorithm thus contributing to the automation of acquiring human goal knowledge. Inspired by related work from search intent classification (cf. [78], [91] or [40]), a wide spectrum of feature types is taken into account including click-through as well as part-of-speech information. Feature selection methods are applied to decide on a final set of feature types that appear promising for the classification task. Having identified discriminative features, two established classification models are applied, i.e. Naive Bayes (cf. [50]) and Support Vector Machines (cf. [42], [175]). To evaluate these models, three-fold cross validation is performed and standard metrics such as precision, recall and F1-measure are calculated.

3.) The case study fosters our understanding of real-world goal knowledge by providing qualitative and quantitative information. Quantitative as well as qualitative analyses are conducted to learn more about the nature of human goals acquired from search query logs. It is refered to work from query log analysis (cf. [132], [74] or [76]) which analyzes different aspects of queries, e.g. temporal aspects. While evaluation strategies appear to be similar, the intention behind query log analysis often is to improve retrieval performance. To learn more about the diversity of human goals, verbs in human goals are analyzed by classifying them into selected Levin's verb classes [95]. In addition, characteristics of commonsense goals are studied by generating verb class histograms of selected Levin's verb classes. By learning more about common and uncommon features, a better understanding is developed of how human goals from search query logs might contribute to complementing commonsense knowledge.

4.2 Motivation

Two ongoing research projects, Cyc [92] and ConceptNet/Openmind (cf. [104], [155]), have been capturing commonsense knowledge, including knowledge about human goals, over the past years aiming to continuously refine, improve and extend their commonsense knowledge with different strategies. While Cyc partly relies on human experts to develop and build their knowledge base, ConceptNet aggregates and processes contributions made by volunteers all over the world. These existing attempts illustrate two main problems in the process of constructing a knowledge base about human goals: (1) the goal acquisition problem (or bottleneck), which refers to the costs associated with knowledge acquisition [101] and (2) the goal coverage problem, which refers to the difficulty of capturing the tremendous variety and range in the set of human goals [43]. These problems have hindered progress in capturing broad knowledge about human goals. To address the goal acquisition problem, this case study explores search query logs referred by [13] to as *Databases of Intentions* – as source for extracting human goal instances.

Search query logs appear to represent an appropriate candidate resource for the goal acquisition problem since every query can be considered to be an expression of a person's (search) goal. In most cases, a query submitted to a search engine expresses some user's underlying goal or motivation. While some goals contained in search queries might be very explicit, other queries might contain more implicit goals, which would mean that they are more difficult to recognize by, for example, an external observer. To give an example: in terms of intentional explicitness, the query "car Miami" differs from the query "buy a car in Miami" [163]. This observation suggests that it is useful to distinguish between at least two classes of queries: (1) queries that contain explicit goals and (2) queries which do not. Table 4.1 contrasts queries of both classes. Results from a larger human subject study corroborate the existence of these two classes and furthermore hint towards a theoretical separability [164].

Queries containing explicit goals	***Queries not containing explicit goals***
"sell my car"	"Mazda dealership"
"play online poker"	"online games"
"find home to rent in Florida"	"Miami beach houses"
"passing a drug test"	"drug test"
"raising your credit score"	"credit cards"

Table 4.1: Exemplary queries from two introduced query classes are contrasted: Queries containing explicit goals and queries which do not contain explicit goals. The presented exemplary queries were obtained from the AOL search query log. (taken from [161])

With respect to this case study's contributions, three research questions are addressed:

RQ 01: Is this work's definition on a "human goal" agreeable and thus practical enough to be used in automated mechanisms to extract human goal instances from search query logs? (see Section 4.4.1)

RQ 02: To what degree is it feasible to automate the process of extracting queries containing explicit goals? (see Section 4.4.2)

RQ 03: What are characteristics of queries containing explicit goals automatically extracted from search query logs? Do search query logs contain commonsense goals, i.e. goals that are found in ConceptNet, a commonsense knowledge base? If they do, what is the nature of human goals shared by ConceptNet and search query logs and how do they differ? (see Section 4.4.3)

4.3 Experimental Setup

4.3.1 Search Query Logs from AOL & MS Research

Two large search query logs are used which were recorded by AOL and Microsoft Research in 2006. These two search query logs are combined to (i) increase the number of queries as well as (ii) to decrease potential domain and population bias that is introduced by using only one search query log. To give an example, queries such as "`cancel AOL account`" or "`how to delete the msn account`" reflect a certain degree of domain bias. The first query log, the MSN search query log[1] excerpt, contains ∼15 million queries (from US users) that were sampled over one month in May, 2006. The second log, the AOL search query log [132], contains ∼20 million queries (from US users) recorded between March 1, 2006 and May 31, 2006. Search queries from both logs were extracted using the same method, and underwent several sanitization and pre-processing steps in order to reduce noise to an acceptable level:

1. Empty Queries: Blank queries and queries containing just a minus character were removed.
2. Short Queries: Only queries with at least three tokens (n>2) were taken into account for the following two reasons: (i) inherent ambiguity of short queries and (ii) the lack of syntactical structure to express human goals. This restriction resulted in a removal of ∼65% of the queries contained in the original datasets.
3. URL queries: Queries containing URLs or fragments of URLs using regular expressions were removed.
4. Queries containing lyrics or movie titles: In preliminary experiments it was observed that queries for music lyrics (I need love lyrics) often contained a verb, but refered to songs rather than actual human goals. This bears the risk of confusing the classification approach

[1] The MSN Search Asset Data Spring 2006 represents a data set which was provided by Microsoft for selected paper proposals at the Workshop on Web Search Click Data 2009 in conjunction with WSDM'09 (http://research.microsoft.com/en-us/um/people/nickcr/wscd09/) accessed Feb 13th, 2012.

that is in part based on syntactic features. Such queries can be identified, since they often contain keywords such as "lyrics" or result in click-through to lyrics or movie related websites (e.g. "http://www.seeklyrics.com"). Limited term and website blacklisting was performed to heuristically reduce the number of such queries in the datasets.

5. Syntax check: Queries containing tokens were removed, which are not numbers or sequences of letters. This filter was used to eliminate corrupted character encodings.

6. Removed misspellings: Misspelled queries were discarded. Whether or not a consecutive query represents a spelling correction was determined by the Levenshtein distance between two consecutive query strings. A query was removed if the Levenshtein distance between the query and its successor is <= 2 and the first query has no click-through event attached.

By applying these filtering steps, >95% of all queries are discarded.

4.3.2 Commonsense Knowledge obtained from ConceptNet

This study chooses ConceptNet [104] as commonsense knowledge base because of its open availability, its natural language knowledge representation and its considerable size. Moreover, knowledge in ConceptNet is partly represented in free-form text which facilitates the comparison with search queries. Knowledge which is contained within ConceptNet is regarded as commonsense knowledge. Commonsense goals [101] in ConceptNet are identified by querying concepts (ConceptNet nodes) which are connected by intentional relations such as "`MotivatedByGoal`", "`UsedFor`" and "`CapableOf`". A subset of entries from ConceptNet was compiled that consists of commonsense goals and imposed the following restrictions on all entries: Commonsense goals had to contain at least one verb and at least one noun. To enforce this restriction, corresponding part-of-speech tags[2] were examined. For the experiments, a set of ∼68.000 commonsense goals from ConceptNet was obtained.

4.4 Results

4.4.1 Practicability of Human Goal Definition

With respect to goal instances in search query logs, the proposed definition on human goals (see Section 3.1.1.1) has been adapted to comply with queries containing explicit goals:

> A search query is regarded to contain an explicit goal whenever the query 1) contains at least one verb and 2) describes a plausible state of affairs that the user may want to achieve, avoid or maintain (cf. [138]) 3) in a recognizable way (cf. [164]).

To evaluate the practicability of the proposed definition and the feasibility of an automatic approach, a human subject study was conducted in which 4 judges (Computer Science graduate students) were instructed to annotate a small query sample. In this task, the

[2] Stanford Part-Of-Speech Tagger version 1.6 available from http://nlp.stanford.edu/software/tagger.shtml accessed Feb 13th, 2012.

judges conducted a question answering task; they were required to independently answer a single question for each of 3000 queries randomly obtained from the AOL search query log[3]. The question for each query followed this schema: Given a query X, do you think that Y (with Y being the first verb in X, plus the remainder of X) is a plausible goal of a searcher who is performing the query X? Two examples should illustrate the process:

> Given query: "how to increase virtual memory"
> Question: Do you think that "increase virtual memory" is a plausible goal of a searcher who is performing the query "how to increase virtual memory"?
> Potential Answer: Yes
>
> Given query: "boys kissing girls"
> Question: Do you think that "kissing girls" is a plausible goal of a searcher who is performing the query "boys kissing girls"
> Potential Answer: No

After the question answering task, answers for each query to the corresponding categories were assigned in the following way: each answer "Yes" resulted in classifying the query as a "query containing an explicit goal"; each answer "No" resulted in classifying the query as a "query not containing an explicit goal". The chart in Figure 4.2 shows that 243 queries out of 3000 have been labeled as containing an explicit goal by all 4 subjects (8.1%, right most bar), and 134 queries have been labelled as containing an explicit goal by 3 out of 4 subjects. This shows that 1) Search query logs contain human goals and 2) the number of queries containing human goals is expected to be small[4]. Noteworthy is the dichotomous characteristic of the agreement distribution in Figure 4.2, which provides preliminary evidence for (i) the agreeability of construct and (ii) the potential for an automatic classification approach. To further explore agreeability, the inter-rater agreement Kappa [35] was calculated between all pairs of human subjects A, B, C and D. The κ values in the human subject study range from 0.65 to 0.76 (see Figure 4.2). The average inter-rater agreement κ yields $\sim$0.72 which hints towards a principal (yet not optimal) agreeability of the proposed definition.

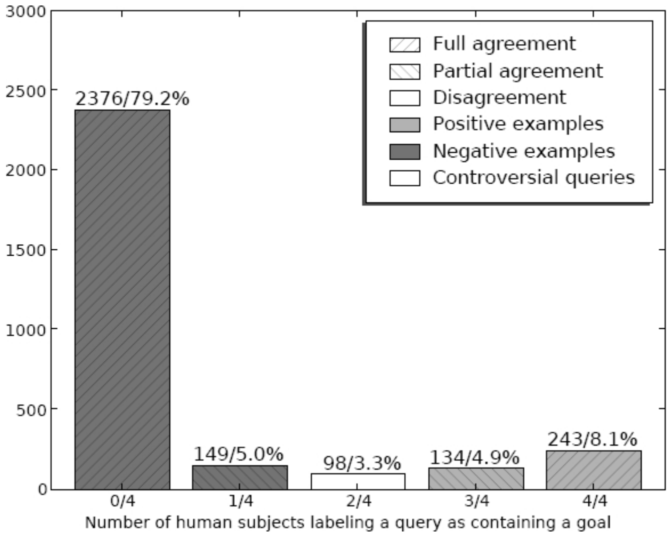
4.4.2 Feasibility of an Automated Approach

To devise an algorithm to automatically identify queries containing explicit goals, a training data set was compiled based on a majority vote among the participants of the human subject study presented in the previous subsection. Out of the 3000 labeled queries, the negative examples were defined by the two bars on the left hand side of Figure 4.2 (2525 total), and the positive examples were defined by the two bars on the right hand side (377). The bar in the middle represents controversial queries[5] which were removed. Altogether, the training set for the classification task comprised 2902 queries. Several feature types were considered for the automatic classification

[3] The MSN query log was not available at the time the agreeability study was conducted.

[4] Note that a query subset was used in the human subject study. If filtered queries were taken into account as well, the fraction of queries containing human goals would be even smaller.

[5] Controversial queries are queries where the majority of judges do not agree whether the query contains an explicit goal or not. Controversial queries could include ambiguous as well as unambiguous queries which do not contain explicit human goals.



Pair	κ
A-B	0.70
A-C	0.69
A-D	0.65
B-C	0.76
B-D	0.75
C-D	0.76

$$\kappa = \frac{P(O)-P(C)}{1-P(C)}$$

Figure 4.2: (left) The percentage of human subjects who labeled a given query as containing an explicit goal. The inter-rater agreement κ was calculated between all pairs of human subjects A, B, C and D to further explore agreeability (see table on the right). An average inter-rater agreement κ of 0.72 indicates good (but not optimal) agreeability of the proposed definition. The formula for calculating κ is denoted in the right lower corner where P(O) denotes the relative observed agreement among raters and P(C) denotes the hypothetical probability of chance agreement, i.e. if the two raters were totally independent. (taken from [161])

approach including "Plain Text", "Part-of-Speech Trigrams", "Query Length", "Click- Through" and "Language Modeling". Preliminary evaluation cycles showed that only plain text and part-of-speech trigrams exhibited sufficient discriminative power for the classification task:

- Plain Text: Queries are represented as binary word vectors or "Set of Words" (SoW). The Porter stemming algorithm [135] was used for word conflation and removing stop words.
- Part-of-Speech Trigrams: Each query is translated from a sequence of tokens into a sequence of part-of-speech (POS) tags. The part-of-speech tagging used a Maximum Entropy Tagger[6] that had been trained on sections 0 to 18 of the Wall Street Journal part of the Penn Treebank corpus. Trigrams were generated by moving a fixed sized window of length 3 over the POS sequence. The sequence boundaries were expanded by introducing a single marker ($) at the beginning and at the end allowing for length two POS features. The query "buying/VBG new/JJ car/NN" would yield the following trigrams:

  $ VBG JJ; VBG JJ NN; JJ NN $

  The intuition behind introducing trigrams was to exploit the grammatical structure of explicit goal queries, i.e. putting emphasis on verb phrases.

Throughout the experiments WEKA [182] was used as data mining toolkit for feature pre-processing, feature selection, classification and evaluation of classification models. Experiments were conducted using several feature types such as word n-grams, part-of-speech n-grams or query length. By ranking these features according to the results of a chi-square feature

[6] http://homepages.inf.ed.ac.uk/lzhang10/maxent toolkit.html accessed Feb 28th, 2011.

selection most discriminative features were determined which eventually led to the decision to use only word unigrams and part-of-speech trigrams. Table 4.2 lists the 20 most discriminative features together with example queries for each feature and the number of occurrences of the feature in the positive class (#). Word features such as "how" and "where" were expected to be amongst highranking features for identifying queries containing human goals. A (probably too strict) stop word removal by the Porter stemmer was suspected to be responsible for this absence. However, the value of indicators like "how" and "where" to the classification task is preserved in some highly ranked partof-speech trigrams, i.e. "$ WRB TO" or "WRB TO VB". Moreover, it can be observed that only a fifth of the features in Table 4.2 are unigram features, notably all of them verbs. Thus, it appears that the most discriminative features for identifying queries containing explicit goals are POS features complemented by verbs.

Nr.	#	Feature	Example Matching Query	Nr.	#	Feature	Example Matching Query
1	126	$ WRB TO	[*$ where to*] find shrooms in Georgia	11	12	TO VB PRP	how [*to copyright your*] photos
2	130	WRB TO VB	[*how to live*] jewishly	12	14	WRB VBP PRP	my hair turned orange [*how do I*] fix it
3	83	TO VB NN	drink milk [*to lose weight*]	13	26	TO VB NNS	what [*to pay Mexicans*]
4	41	buy	*buy* acoustic guitar	14	28	VB NN NNS	[*make business cards*]
5	58	VB NN NN	[*find property values*] calculator	15	19	TO VB DT	teach yourself [*to play the*] piano
6	20	find	*find* an old friend for free	16	9	VB PRP JJ	how to [*get yourself sick*]
7	36	TO VB JJ	I want [*to download instant*] messenger	17	45	TO VB IN	places [*to stay in*] Gatlinburg
8	27	make	*make* your own parable	18	8	install	*install* Microsoft windows 2000
9	52	$ VB NN	[*$ find lawyer*] in Georgia to form llc	19	14	$ VB PRP	[*$ customize your*] aol buddy icon
10	29	VB NN IN	[*borrow money from*] Donald Trump	20	22	VB PRP NN	how to [*obtain us passport*]

Table 4.2: The top 20 most discriminative features are illustrated resulting from applying WEKA's chi-square feature selection. The part of the search query that matches the respective part-of-speech feature is enclosed by brackets ([ ]). To aid readability, descriptions of selective part-of-speech tags are provided according to the Penn Treebank Tag Set: WRB represents a Wh- adverb, VB represents the base form of a verb, PRP represents a personal pronoun, NN represents a noun in singular form, IN represents a preposition and DT represents a determiner. (taken from [161])

After having identified a set of discriminative features, two common classification models were applied for handling textual data (i) a Naive Bayes (NB) classifier [50] and (ii) a linear Support Vector Machine (SVM) (cf. [42], [175]). Similar to prior work on query classification [97], the F1 measure, i.e. the harmonic mean of precision and recall, was chosen for evaluation. Since this study is mainly interested in achieving high values for the positive class, i.e. queries containing explicit goals, only precision, recall and F1 values for the positive class are reported. In conducting experiments with regard to the F1 score, this study aims to identify configurations that balance precision and recall in a way that is useful for acquiring expressions of human goals. The selected linear classification models were evaluated with regard to varying feature set sizes (see Figure 4.3). Feature sets are generated by applying WEKA's chi-square feature selection and keeping the top N features. For each classifier/feature set size combination, 10 trials of three- fold cross-validation are carried out. The resulting scores for each combination are averaged over all trials. Figure 4.3 presents two resulting learning curves, i.e. the F1 scores of different feature set sizes and classification models.

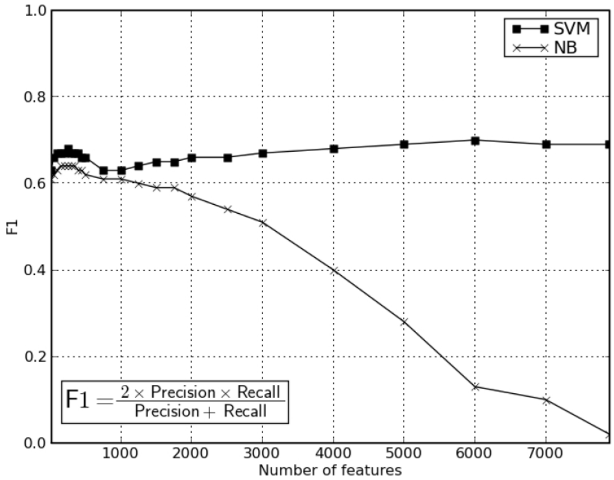


Figure 4.3: Learning curves for two classification models, i.e. the linear Support Vector Machine and the Naive Bayes classifier are shown. F1-scores indicate the respective classifier's performance with varying feature numbers. In this case, the F1-scores refer to the positive class, i.e. queries containing explicit goals. (taken from [161])

These results indicate that the SVM appears to be better suited for the classification task, in particular with increasing number of features. The results also illustrate the NB classifier's dependence on accurate feature selection prior to training and classification. For the NB classifier, the performance significantly deteriorates when more features are used. Informed by these results, the linear SVM as the classification model is selected for subsequent experiments.

Precision	**Recall**	**F1 - Score**
0.77	0.63	0.69

Table 4.3: Following averaged Precision, Recall and F1-scores are achieved on the manually labeled data set. A linear SVM takes into account all features to conduct the classification task. Precision, recall and F1 scores refer to the positive class, i.e. queries containing explicit goals. (taken from [161])

Table 4.3 shows the precision, recall and F1 scores for the positive class, i.e. queries containing explicit goals. The values result from averaging 10 trials of three-fold cross-validation keeping all features. A precision of 77% means that in 77% where the classification model believed the query contained a goal, the majority of human subjects agreed. This form of evaluation allows statements about the generalization capabilities of the algorithm: A precision of 77% is the quality to expect when applying this algorithm to search query logs.

Table 4.4 shows classification results from the automated method in form of a confusion matrix. It provides an overview of the query distribution regarding true positive (TP), false positive (FP), false negative (FN) and true negative (TN) information. In addition to frequency values, Table 4.4 provides corresponding query examples. Examining exemplary queries categorized as FP or FN can

Classified as → Annotated as ↓	Containing an Explicit Goal	Not Containing an Explicit Goal
Containing an Explicit Goal	TP: # 239 / 8.2% Query Examples: how to write a resume, make money from home, obtaining a passport	FN: # 138 / 4.8% Query Examples: flyfishing around park city Utah, cruising down the free way
Not Containing an Explicit Goal	FP: # 73 / 2.5% Query Examples: dancing with the stars, stem cell research, living room furniture	TN: # 2452 / 84.5 % Query Examples: online games, national car rental, vegas ride shuttle, shoppers guide

Table 4.4: Classification results are shown in form of a confusion matrix, i.e. reflecting true positive (TP), false positive (FP), false negative (FN) and true negative (TN) information. The table provides an overview of the query distribution as well as corresponding query examples. Examining exemplary queries categorized as FP or FN can be beneficial to better understand the algorithm's behavior and to improve its performance. (taken from [161])

be beneficial to better understand the algorithm's behavior as well as to improve its performance. Incorrectly classified entries are mainly due to incorrect part-of-speech tagging.

A simple baseline approach would guess that a query containing a verb always contains an explicit goal. Yet, such a baseline would perform significantly worse: While the baseline would excel on recall (= yielding a recall of 1.0 due to the definition of explicit goals requiring a query to contain a verb), it would perform worse with regard to the extraction task due to low precision. In experiments, the baseline achieved a precision of 0.13 and a F1 score of 0.23.

4.4.3 Characteristics of Queries Containing Explicit Goals

To gain more insights into the nature of human goals, quantitative as well as qualitative analyses were conducted. The automatic classification method was applied to ∼35 million queries[7], i.e. the AOL and the MSN search query log combined. The set of queries this system classified as containing human goals, which is called the result set, comprises ∼142,000 queries, 110,000 of which are unique. With a precision of 77%, this means an estimated 109,000 queries in the result set actually do contain goals. 109,000 queries might appear small in the light of ∼35 million queries contained in the original search query logs. Yet, in case of the AOL search query log, 20 million queries reportedly represent only 0.33% of the total number of queries served during that time. Considering the large numbers of queries served every day, the approach would be able to continuously extract human goal expressions on larger datasets. Stop word removal and stemming were applied to the result set to obtain a more accurate frequency ranking. Similar entries such as "`buy a new car`" and "`buying new cars`" are then merged into one entry "`buy car`" with higher frequency values. To enhance readability, all queries in subsequent result figures and tables are manually post-processed: stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning.

[7]After initial filtering, ∼1.7 million queries remained meaning that >95% of all queries are discarded.

4.4.3.1 Quantitative Analysis

The 40 most frequent queries from the result set are presented in Table 4.5. Each example is accompanied by rank and frequency information. Queries containing the token "`http`" are filtered out and those queries containing expletives or sex-related content are replaced by "`deleted`".

Nr.	Query	Freq.	Nr.	Query	Freq.	Nr.	Query	Freq.	Nr.	Query	Freq.
1	enterprise rent car	311	11	dollar rent car	142	21	buy buy baby	98	31	"`deleted`"	78
2	build bear	212	12	rent car	142	22	find people free	97	32	hertz rent car	73
3	pimp ride	195	13	find person*	139	23	find grave	96	33	lose weight*	73
4	rent center	192	14	find email address	138	24	listen free music	84	34	trick truck	70
5	listen to music*	190	15	tie ties	111	25	make money home	83	35	buy house*	67
6	find phone number*	185	16	meaning of name	109	26	start own business	83	36	work home*	66
7	assist sell	173	17	change password	107	27	write resume	83	37	lose weight fast*	64
8	pimp space	167	18	find address*	103	28	flash rack	81	38	make own website	63
9	budget rent car	166	19	pimp myspace	102	29	cancel aol account	78	39	play guitar*	63
10	find zip code	154	20	"`deleted`"	102	30	get pregnant*	78	40	gain weight*	62

Table 4.5: The 40 most frequent queries containing goals in the result set are listed. Queries marked with a "*" represent queries that are contained in ConceptNet indicating the existence of commonsense goals in search query logs. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning. (taken from [161])

The information in this table reflects – to some extent – the needs and goals of the North-American web population, i.e. users of AOL/MS search during the period of the dataset recordings. Some of the most frequent queries containing human goals relate to commonsense goals such as "`lose weight`", "`get pregnant`" or "`listen to music`" (marked with a "*" in Table 4.5). These goals are refered to as commonsense goal because of their relation to ConceptNet, a commonsense knowledge base. "`lose weight`" represents a ConceptNet node which is connected to other nodes by intentional relations such as "`MotivatedByGoal`". The existence of commonsense goals among the most frequent queries in the result set provides some evidence that search query logs are suited for the task of commonsense knowledge acquisition. Combining the AOL and MSN search query logs allowed to partly decrease bias that would be introduced by using just one dataset. Yet, the remaining bias introduced by the corpus itself (search queries) and the population (i.e. AOL and MSN users) deserves attention: A fraction of frequent queries deals with web-related or AOL/MSN specific issues, such as the queries "`find e-mail address`" or "`cancel aol`" account. Entries such as "`meaning of name`", and "`buy buy baby`" likely represent false positives, revealing two kinds of shortcomings of this approach: First, the automatic classification approach relies on linguistic patterns generated by part-of-speech (POS) tagging. In case of the query meaning of name, the POS tagger mistakenly tagged meaning as a verb (VBG) which yields an incorrect decision. Other examples include "`enterprise rent car`" and "`hertz rent car`" where the word rent has again been mistakenly tagged as a verb. A part-of-speech tagger that is trained on a more suitable corpus might help alleviating such problems in the future. Second, certain queries containing explicit goals resemble book titles, TV shows or music themes such as "`buy buy baby`". This problem could be addressed by including domain knowledge (for instance "imdb.com") in the classification task or inspecting and analyzing click-through data and anchor text, which can be expected to improve the overall performance of the approach.

	home (2541)	**car** (2227)	**account** (1729)	**card** (1537)	**house** (1365)	**music** (1322)	**money** (1276)	**phone** (1109)	**name** (1109)	**window** (1086)
make (12600)	294*	55*	24	318*	23*	97*	769*	31*	85	59*
buy (11905)	426*	457*	15	139	339*	84	74	154	17	57
find (11836)	231*	127*	26	33	133*	65*	72*	560*	260	41
get (7260)	102*	51*	35*	75*	84*	36	66*	38*	42	33
do (4767)	52	22*	25*	40	21	19	33	21	22	42
listen (2864)	14	0	0	0	2	702*	0	6	1	7
use (2490)	26	145*	24	22*	9	7	15*	52*	14	33
clean (2485)	29*	17*	2	1	72*	0	3	0	0	28*
build (2410)	108*	46*	15	3	128*	2	5	0	0	7
write (1899)	3	3	0	38*	0	14*	10	1*	16*	3

Table 4.6: The 10 most frequent verbs and the 10 most frequent nouns in the result set are ilustrated as well as frequency numbers of corresponding co-occurrences. Frequency values marked with a "*" represent queries that are contained in ConceptNet. (taken from [161])

To better understand the nature of identified human goals, a term analysis was conducted by identifying the 10 most frequent nouns and verbs in the result set and analyzing verb/noun co-occurrences. The most popular verb/noun co-occurrences in Table 4.6 seem to be indicative of typical human goals on the web, such as "`make money`", "`listen music`" or "`find phone`". Preliminary evaluations of the top verb/noun correlations reveal that many of these human goals are also contained in the ConceptNet commonsense knowledge base (marked with a "*"). This can be understood as a further indicator of the usefulness of search query logs for acquiring knowledge about human goals. It also suggests that search query logs might be useful to automatically complement knowledge contained in existing commonsense knowledge bases, which has been attempted before [43]. If search query logs would be utilized for such a purpose, a relevant question to ask is: How diverse is the set of human goals contained in search query logs? The diversity of goals would ultimately constrain the utility of a given dataset for complementing existing knowledge bases. In order to explore this question, explicit goal queries from the long tail [3] were classified into selected Levin's verb classes [95]. The histogram in Figure 4.4 reflects the classification result. This distribution provides evidence that the majority of queries containing explicit goals are diverse in nature, i.e. covering a broad spectrum of goals.

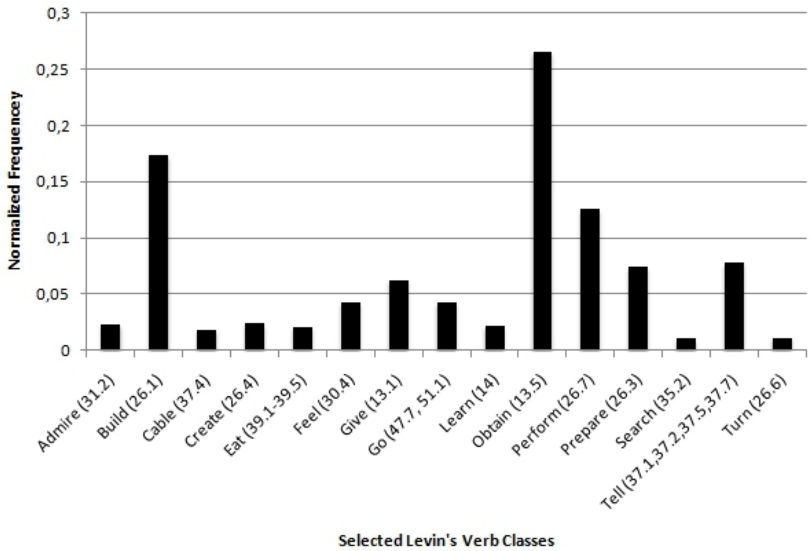


Figure 4.4: This histogram illustrates the distribution of explicit goal queries from the long tail over selected Levin's verb classes. The distribution is diverse with regard to all 15 verb classes (except predominance of verb classes "Build" and "Obtain") and thus provides evidence that the majority of queries containing explicit goals are diverse in nature. Frequency values are normalized to allow better comparison. Levin's verb classes along with corresponding indices are denoted in brackets. (taken from [161])

4.4.3.2 Qualitative Analysis

While the analyses conducted so far provide statistical insights into the nature of human goals contained in search queries, it is difficult to infer information about their quality. To address this issue, limited qualitative analyses was performed through inspection. Four verbs and four nouns were manually selected to represent a range of important activities and topics typically addressed by web search users and inspect explicit goal queries which contain them. In Table 4.7, the 10 most frequent goals in the result set are listed, which contain either the verbs "get", "make", "change" or "be". Frequency refers to the occurrence in the result set. The human goals listed in Table 4.7 are the result of identifying the first verb in a query containing a goal, and truncating any tokens prior to this verb. Goals marked with a "`*`" represent goals that are contained in ConceptNet. Many entries in Table 4.7 are related to existing commonsense goals, such as "`be pregnant`", "`be rich`" or "`be funny`".

To gain further insights, interesting nouns were selected which belong to Information Extraction classes: "`money`" [economy], "`birth`" [event], "`home`" [location] and "`people`" [person]. The corresponding top 10 most frequent human goal expressions are depicted in Table 4.8.

Table 4.7 and Table 4.8 make an interesting case for using search query logs to complement existing commonsense knowledge bases as is demonstrated in Section 7.1.

Nr.	Verb: get	Verb: make	Verb: change	Verb: be
1	get pregnant* (78)	make money home (83)	change password (107)	be anorexic* (27)
2	get rid of ants (45)	make your own website (63)	change screen name (47)	be funny* (11)
3	get passport* (43)	make money online (56)	change name* (31)	be bulimic (11)
4	get rid of love handles (23)	make wish foundation (53)	change AOL password (28)	be cool* (8)
5	get out of debt* (22)	make money* (52)	change profile (20)	be loved (8)
6	get rid of stretch marks (20)	make the band 3 (49)	change AOL screen name (17)	be millionaire* (8)
7	get myspace school(20)	make money fast (42)	change e-mail address (15)	be sexy* (6)
8	get rid of moles (19)	make new screen name (33)	change home page (14)	be emo (6)
9	get rid of belly fat (16)	make crossword puzzle (31)	change welcome screen (11)	be romantic* (6)
10	get rich* (15)	make ethanol (30)	change life* (7)	be happy* (5)

Table 4.7: This table provides an overview of the 10 most frequent human goals from search query logs containing the verbs "`get`", "`make`", "`change`" or "`be`". Human goals marked with a "*" indicate goals that are also contained in ConceptNet. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning. (taken from [161])

Nr.	Noun: money	Noun: birth	Noun: home	Noun: people
1	make money from/at home (83)	find birth mother (6)	make money from/at home (83)	find people for free (97)
2	make money online (56)	get birth certificate (6)	work at home*(66)	find people (16)
3	make money* (52)	find birth parents (6)	buy home* (54)	find peoples phone number (12)
4	make money fast (42)	buy birth control pills (5)	sell home (35)	find missing people (9)
5	save money* (23)	get birth control (3)	build own home (31)	find peoples address (9)
6	make money on ebay (18)	buy birth control online (3)	design own home (24)	find people online (8)
7	ways to make money (17)	buy birth control (3)	sell own home (21)	win friends and influence people (8)
8	make money on internet (16)	use birth control pill (3)	find home* (18)	deal with difficult people (8)
9	find lost money (10)	obtain birth certificate (2)	make msn home page (16)	loop up people (6)
10	invest money* (9)	look up birth parents (2)	organize home* (16)	search for people (6)

Table 4.8: In this table the 10 most frequent human goals from search query logs containing the nouns "`money`", "`birth`", "`home`", or "`people`" are presented. Human goals marked with a "*" indicate goals that are also contained in ConceptNet. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning. (taken from [161])

4.4.3.3 Comparative Analysis

The qualitative analysis indicated the existence of commonsense goals in search query logs. To elaborate on that issue, human goals acquired from search query logs were compared to commonsense goals (cf. [101]) from ConceptNet. Commonsense goals in ConceptNet were identified by querying concepts (ConceptNet nodes) connected by relations such as `UsedFor`, `CapableOf` and `MotivatedByGoal`. A subset of entries from ConceptNet was compiled that consists of commonsense goals and imposed the following restrictions on all candidates to increase quality: Commonsense goal candidates had to contain at least one verb and at least one noun. This approach obtained an overall number of ∼68,000 commonsense goals which were compared to the ∼110,000

unique human goals acquired from search query logs.
To learn whether search query logs contained commonsense goals, the intersection between search query log and ConceptNet goal sets was calculated with the following intuition: An adequate number of shared entries would indicate the presence of commonsense goals in search query logs. To calculate the intersection, a simple goal matching algorithm (similar to [101]) was devised to identify matching pairs of human goals from the two goal sets. All entries were pre-processed: stop words were removed and all remaining tokens were stemmed using the Porter stemmer [135]. Following table shows examples of commonsense goal examples from ConceptNet before and after applying pre-processing steps.

Before Pre-Processing	**After Pre-Processing**
to protect your family	['protect', 'famili']
have something to do during breakfast	['have', 'do', 'breakfast']
how to tell kids about suicide	['tell', 'kid', 'suicid']

For two goals to match, they had to contain an equal number of identical stems. The matching algorithm focused on lexical characteristics only; semantic similarities were not taken into account. This idea is similar to Liu's process of normalization to identify similar instances in ConceptNet [60]. Table 4.9 illustrates some examples of matching and non-matching entries from ConceptNet and search query logs.

ConceptNet Goals	**Search Query Log Goals**	**Match**
<u>make paper airplanes</u>	how to <u>make paper airplanes</u>	yes
<u>get</u> in <u>shape</u>	<u>get</u>ting into <u>shape</u>	yes
we <u>buy</u> <u>houses</u>	<u>buy</u> a <u>house</u>	yes
we buy houses	purchase a house	no
<u>make money</u>	<u>make</u> more <u>money</u>	yes
make money	make online money	no

Table 4.9: Examples for matching and non-matching ConceptNet and search query log entries. Matching sequences are underlined. (taken from [161])

Eventually, ∼2,300 ConceptNet goals and ∼3,100 search query log goals (occurrences) were obtained that produced positive matches. While the number of shared goals appears small, these findings provide first evidence of the existence of commonsense goals in search query logs. To learn more about the nature of commonsense goals in search query logs, the set of ∼3,100 commonsense search query log goals was categorized into a subset of Levin's verb class taxonomy [95]. From this taxonomy, 15 verb classes were selected that were deemed relevant for reflecting human activities, e.g. "`eat`" or "`learn`". Figure 4.5 shows the resulting verb class histogram of commonsense goals from search query logs. Three dominant verb classes build, obtain and perform can be identified from Figure 4.5. Their dominance might be explained by occurrences of verbs in frequently stated commonsense goals such as "`make money`", "`buy food`" or "`play an instrument`". Class `build` (verb "`make`"), class `obtain` (verb "`buy`") and class `perform` (verb "`play`") represent corresponding Levin's verb classes. An interesting observation is that the verb class search is not dominant in the dataset. This class might be underrepresented due to the fact that search engines already represent a means for searching the web, i.e. goals in search queries do not need to contain verbs expressing the goal to search itself.

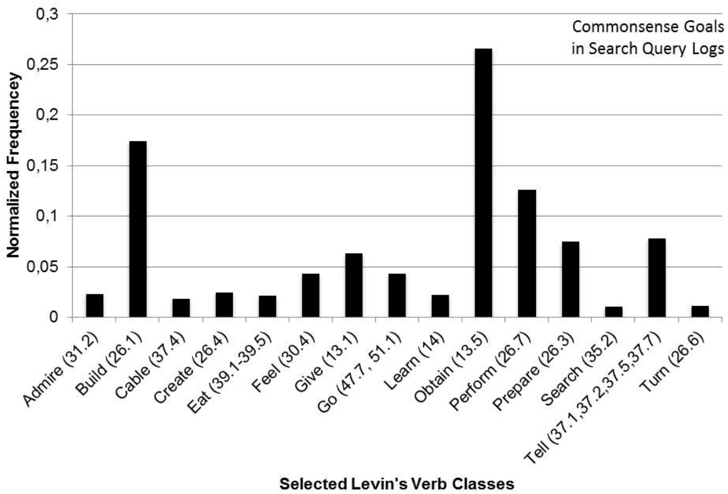


Figure 4.5: This histogram illustrates how commonsense goals in search query logs are distributed over selected Levin's verb classes. Frequency values are normalized to allow better comparison. Levin's verb classes along with corresponding indices are denoted in brackets. (taken from [161])

The remaining section studies the extent to which commonsense goals in search query logs differ to those in ConceptNet. To illustrate different characteristics between goals from search query logs (QL) and ConceptNet (CN), a verb class histogram for the complementary sets of goals was generated. In set-theoretic terms, this reads as follows: CN $-$ QL and QL $-$ CN. The initial intuition was that ConceptNet's commonsense goals would be biased towards everyday situations and human characteristics such as eating, feeling and living. The results confirm this intuition: The verb histogram in Figure 4.6 shows that verb classes "`eat`", "`gorge`", "`touch`" and "`feel`" are more prominent in the ConceptNet set. Similarly, we expected classes such as "`obtain`" to be the dominating search query log goals. This can be observed in our results: Levin's verb class '`obtain`' dominates the human goals acquired from search query logs, which contain frequently occurring verb instances such as "`get`", "`buy`" and "`find`".

In addition, Figure 4.6 reveals that each dataset, i.e. CN $-$ QL and QL $-$ CN, favors a different range of human activities. In fact, this suggests that query logs could actually "*help increase coverage of commonsense knowledge bases*", for example by focusing on types of goals that are more prevalent in search query logs, such as "`obtain`" and "`build`". The distribution also suggests that search query logs are not suited to contribute to commonsense goals from verb classes such as "`eat`" or "`feel`".

Following three statements conclude this section: 1.) Search query logs appear to be a potential source for commonsense goals. 2.) Search query logs and ConceptNet each emphasize different goal classes. 3.) Search query logs might represent a useful resource to complement existing commonsense knowledge bases such as ConceptNet.

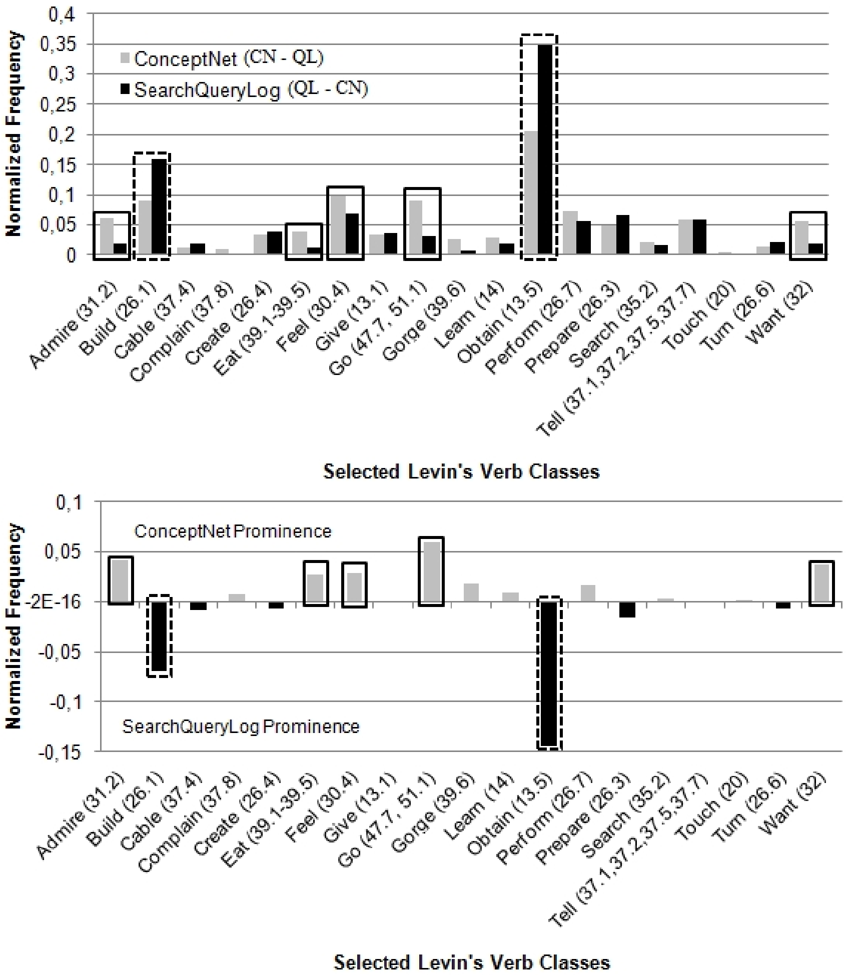


Figure 4.6: The top figure compares verb class distributions from two complementary sets, i.e. CN − QL and QL − CN. This comparison illustrates for which classes search query logs can potentially contribute to increasing coverage of ConceptNet. The bottom figure represents a different presentation of the upper figure's frequency values. It shows relative differences in frequency values to provide an overview of verb class prominences. Frequency values are normalized to allow better comparisons. Levin's verb classes along with corresponding indices are denoted in brackets.(taken from [161])

4.5 Discussion

This case study validates the framework's applicability by providing evidence that the proposed definition on human goals is useful with respect to practicability and agreeability in a particular domain, the search domain. Future work might explore the definition's usefulness in other domains. The case study also contributes to the development of the data model by advancing the understanding of real-world human goal knowledge, i.e. "what goals do people have?", "what are their attributes and characteristics?" or "how are they distributed?" Qualitative analyses inform us about usable goal attributes such as types of goals (achieve vs. avoid) and different goal horizons. Quantitative analyses inform us about goal distributions and reveal a great variety of human goals. The comparative analysis reveals the existence of commonsense goals in search query logs and thus establishes a connection between human goal instances from two domains, the commonsense and the search domain. Overall, this gained understanding eventually benefits the formalization process of human goal knowledge, i.e. the development of a data model encoding human goal knowledge.

With respect to the meta-process of engineering human goal knowledge, this case study addresses the *Acquisition* step and the *Validation* step. It allows first estimates on the extent to which acquiring knowledge about human goals from search query logs can be automated. Human participation is required to compile training and test sets as well as to engineer adequate features. Once this process is completed, the classification algorithm is capable of extracting qualitative human goal instances on its own. Results indicate that the presented automatic approach is a viable alternative to manual or semi-automatic approaches which are often costly and time-consuming. This work is thus also relevant for knowledge engineers who are interested in mining human goal instances from the web.

Finally, this case study introduces search query logs as a viable source for extracting human goal instances. And since search query logs are a natural by-product of human activity on the web, they represent a largely untapped, renewable resource for the goal knowledge acquisition task, i.e. the *Acquisition* step.

Chapter 5

Case Study 02: Analyzing Human Goals in Natural Language Text

5.1 Contribution to PhD Objectives

The case study addresses the categorization of textual resources into human goal categories. It thus benefits and supports the development of the proposed data model by indicating components which are necessary to model human goal knowledge, e.g. the *GoalCategory* class, the *Task* class and the *MeansForEndRelation* class as illustrated in Figure 5.1.

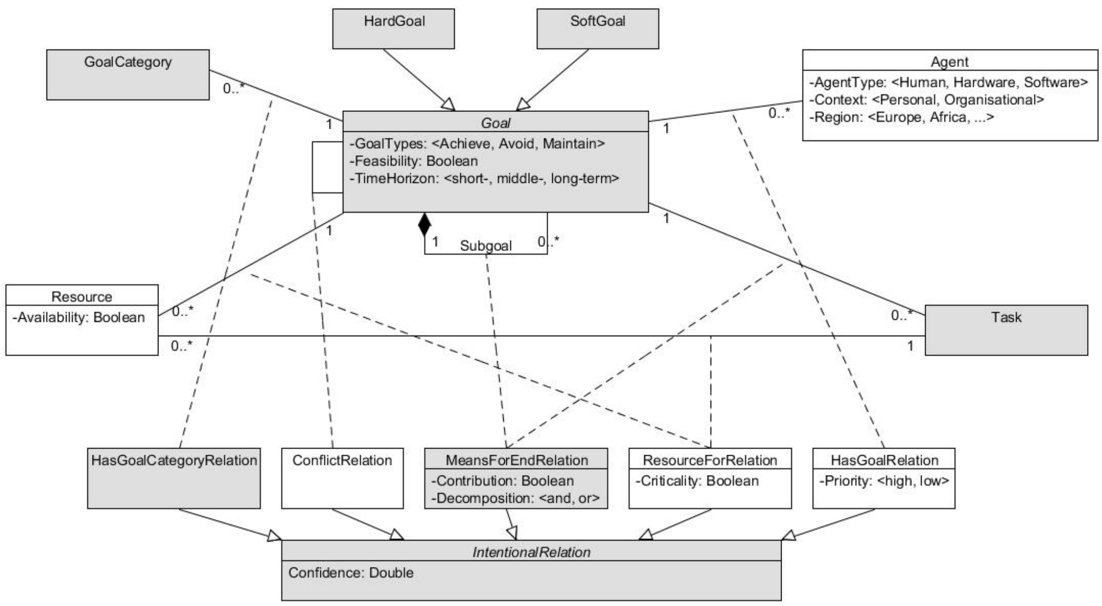


Figure 5.1: Components of the proposed data model on human goal knowledge covered by this case study are highlighted.

This case study seeks to understand how human goal knowledge is expressed in natural language text and thus informs the automation of the meta-process' *Elicitation* step. To approach this understanding, explorative questions are posed such as "How do people express their goals?", "To what extent can textual resources be (automatically) profiled from a goal-oriented perspective?" and "Do goal categories evolve over time?" This understanding (i) supports efforts to devise algorithmic approaches, which are part of the meta-process' *Acquisition* step, and (ii) moreover informs the development of the proposed data model. This case study introduces a new domain, i.e. political speeches, to be analyzed from an intentional perspective. Political speeches can be expected to contain a broad variety of human goals and are chosen over other resources, e.g. blog posts, because they typically have a clearer focus on discussing, conveying or achieving goals.

Covered Components	**Description**
Goal	Human goal instances in this case study are generated manually to describe and characterize corresponding goal categories. This characterization is necessary to construct a mapping between people's actions (tasks) and goal categories. Goal instances thus serve as bridge between people's actions (tasks) and the goal categories.
GoalCategory	This case study uses existing goal categories from a socio-psychological taxonomy of 135 goal categories [30]. This taxonomy reflects a possible spectrum of human goals by organizing high-level human goals into a hierarchical construct. Categories include "`A good marriage`", "`Getting an education`" and "`Taking care of family`". Goal categories are exchangeable to shift the focus to other, e.g. more domain-specific categories such as the search domain (cf. [146]).
Task	People often rather write about what they are going to do, i.e. their actions (tasks), to accomplish their goals rather than write about their goals in an explicit manner. It appears that people use tasks as surrogate for their goals in textual resources. This case study explores and uses this aspect for assigning textual resources to goal categories.
IntentionalRelation, MeansForEndRelation, HasGoalCategoryRelation	The extraction of intentional relations is a byproduct of creating mappings between the (*GoalCategory*/*Goal*) class and the (*Goal*/*Task*) class.

With respect to the core objectives of this PhD thesis, this case study makes following contributions:

1.) The case study studies the extent to which textual resources can be assigned to goal categories in an automated manner. This categorization can, for instance, be used to generate goal profiles which allow the analysis and comparison of textual resources from an intentional perspective. For that purpose, the case study provides an algorithmic approach thus contributing to automate the acquisition of human goal knowledge from natural language text. The approach uses two mappings (i) from people's tasks to people's goals and (ii) from people's goals to people's goal categories. In the process of generating these mappings, goal knowledge is made operational by instantiating and using following components of the goal data model: the *Goal* class, the *GoalCategory* class, the *Task* class and the *MeansForEndRelation* class. For evaluation purposes, an exploratory study is conducted that focuses on analyzing transcripts of political speeches given by US presidential candidates in 2008.

2.) This case study contributes to advancing our understanding of human goal knowledge in natural language text, in particular transcripts of political speeches (monologues). The study examines and demonstrates the usefulness of relations between goals, tasks, agents and goal categories which then have been integrated into the proposed data model. In preliminary experiments textual resources are analyzed with respect to how people express their goals. This gained understanding is beneficial for both (i) developing the data model with respect to incorporating real-world structural information and (ii) characterizing the meta-process' *Acquisition* step with respect to algorithmic approaches and techniques.

5.2 Motivation

Traditional topic categorization attempts to classify a document according to its predominant subject matter (what the page is about, e.g. sports or politics). Besides this traditional topical classification of text, orthogonal categories have been applied and explored such as sentiments, opinions or genres [49]. Sentiment Analysis, also known as opinion mining [127], seeks to classify textual resources according to their contained emotions and opinions. Sentiment classification has been a challenging topic in Natural Language Processing [179] and can be defined as a binary classification task to assign a sentence either positive or negative polarity [128].

This work takes a different approach and attempts to classify documents according to the human goals described within them (what goals a page is about, e.g. *Achieve Happiness* or *Maintain Good Health*). Similar to sentiment analysis and opinion mining, an intentional analysis represents an orthogonal view on topic categorization and aims at creating goal profiles of textual documents. Instead of analyzing text according to the conveyed sentiments or opinions, this form of analysis attempts to answer *which human goals* (i.e. *future* states of affairs that some agent wants to achieve) *are referenced in a given document.* This analysis deals with a different temporal focus than sentiment analysis, where a *present* (emotional) state is approximated. Using a sample of web documents it could be observed that people rarely state their goals explicitly in text. As an example, consider the human religious intention to `Achieve Salvation` (taken from [30]). Although this is an activity pursued by many, it is extremely rare to find someone who states their plan on how to accomplish this goal. However, people are quite prolific in writing about the actions (tasks) they participate in on a daily basis, such as "`convert to Christianity`", which indirectly contribute to "`Achieve Salvation`".

This work therefore explores the use of such *actions* as a proxy for inferring intentions from textual content. This intentional analysis now can be understood as identifying a corresponding goal category for every *action* indicative of a goal in a given text. The analysis task is to approximate the unknown function *f:* $S \times C^G \rightarrow \{True, False\}$, where $C^G = \{c^G{}_1, c^G{}_2 \ldots c^G{}_n\}$ is a set of predefined goal categories, D is a domain of text documents and each document d_i contains a sequence of sentences $S = \{s_1, s_2 \ldots s_{|S|}\}$.

With respect to this case study's contributions, two research questions are addressed:

> RQ 01: To what extent is it feasible to automatically annotate natural language text with goal categories? (see Section 5.4.1)
>
> RQ 02: What are characteristics of human goal knowledge acquired from transcripts of political speeches (monologues)? (see Section 5.4.2)

5.3 Experimental Setup

5.3.1 Enriching a Taxonomy of Human Goals

The social-psychological theoretical framework [30] was employed that organizes high-level goals of people into 135 goal categories including ”`A good marriage`”, ”`Getting an education`“ and ”`Taking care of family`”. A useful property of taxonomies in general is that categories are hierarchically grouped into high-level categories, in this case top level categories such as “`Family`”, “`Religion`” and “`Money`” (not depicted in Figure 5.2).

Abbreviation	Full label
Achieving salvation	Achieving salvation
Arts	Appreciating the arts
Aspirations	Achieving my aspirations
Attracting sexually	Being able to attract, please, sexually excite a sexual partner
Avoiding failure	Avoiding failure
Avoiding guilt	Avoiding feelings of guilt
Avoiding rejection	Avoiding rejection by others
Avoiding stress	Avoiding stress
Being able to fantasize	Being able to fantasize, imagine
Being affectionate	Being affectionate toward others
Being ambitious	Being ambitious, hard-working
Being better than others	Being better than others, beating others
Being carefree	Being lighthearted, carefree, enjoying life
Being clean	Being clean, neat (personal care)
Being conventional	Maintaining conventional views, avoiding innovation
Being creative	Being creative (e.g., artistically, scientifically, intellectually)
Being curious	Being curious, inspecting, learning
Being disciplined	Being disciplined, able to follow-through with projects I start, following my intentions with behavior
Being free	Having freedom (being a free person)
Being good looking	Being good looking

Figure 5.2: An excerpt of Chulef's taxonomy of human goals [30]. The left part lists the first 20 goal categories and the right part provides additional information to each category.

While the taxonomy of human goals provides abbreviations and full labels for each goal category, further category descriptions are not available. In order to semantically enrich these category descriptions, it was attempted to find corresponding descriptive phrases, i.e. human goals, for each category. To give an example: Descriptive phrases for the category “`Achieve Salvation`” included “`to reach spiritual enlightenment`” or “`to get into heaven`”. The manual process of enriching the taxonomy with descriptive phrases was iterative. Together with Dr. Read, one of the co-authors of Chulef et al. [30], these mappings were evaluated. Moreover, Dr. Read provided assistance to better understand goal category distinctions during the evaluation phase.

5.3.2 Knowledge Base Construction

In a first step a knowledge base was generated consisting of actions that indicate relevance for one of 135 categories. A large set of indicative actions was acquired by searching for sentences on the web (cf. [33]) that contained both (i) one of the descriptive phrases for the category, and (ii) an action-based intentional relation. To achieve that, a series of query strings was constructed by concatenating each descriptive phrase with each of the following five intentional relation phrases: "`in order to`", "`for the purpose of`", "`essential for`", "`necessary for`" and "`critical for`". Then, exact phrase searches were issued to the web using the Yahoo! BOSS API for all constructed query strings. The textual content of the first 500 result pages was retrieved, parsed and sentence delimited. Sentences that contained query phrases were stored in the knowledge base which was implemented via an Apache Lucene index. Table 5.1 shows sample phrase queries and retrieved sentences with the respective indicative actions underlined.

Query String intentional relation + *descr. Phrase*	**Retrieved Sentences (Yahoo)** indicative actions
"in order to *look young*"	In order to *look young* and beautiful, you need to take care of your skin.
"for the purpose of *looking young*"	While we know that fitness is one of the keys to remaining healthy, *we* also exercise for the purpose of *looking young and sexy.*
"in order to *look youthful*"	It was in the context of people drinking a lot of water in order to *look youthful.*
"in order to *avoid wrinkles*"	You need to moisturize inside and out, in order to *avoid wrinkles.*

Table 5.1: Exemplary query strings for the category "`Looking Young`" and retrieved sentences containing indicative actions. (taken from [86])

5.3.3 Sentences to Goal Categories Matching

To automatically generate intentional annotations for a given textual resource, the document was first segmented into a set of sentences for subsequent analysis. Then, each sentence in the document was issued as a query to the knowledge base using Lucene's default similarity measure. This allowed identifying the most similar sentence in the knowledge base containing indicative actions. A similarity greater than 0.5 (1.0 equals an exact match) was required as a quality criterion of the retrieved sentences. Then the goal category associated with the knowledge base entry is assigned as the intentional annotation in a "winner takes it all" approach. Intentional annotations for entire documents are produced by aggregating annotations of all sentences.

5.4 Results

Section 5.4.1 explores the feasibility to automatically annotate textual resources with goal categories, e.g. for generating goal profiles which allow the analysis and comparison of textual resources from an intentional perspective. For evaluation purposes, an exploratory study is conducted that focuses on analyzing transcripts of political speeches given by US presidential candidates in 2008.

Section 5.4.2 examines selected characteristics of human goal knowledge, i.e. knowledge base characteristics and temporal evolution of goal categories.

5.4.1 Annotating Textual Resources with Goal Categories

Due to the exploratory nature of this case study, political speeches are used because (i) political speeches typically have a clear focus on discussing, conveying or achieving goals (ii) transcripts of political speeches are less affected by noise compared to other resources, and (iii) political speeches can be expected to contain a broad variety of goals. These factors facilitate evaluation and make political speeches particularly suitable to explore the prospects of intentional annotations in a simplified setting.

Textual resources of 44 transcripts of political speeches were retrieved and preprocessed. The speeches were given in April and June 2008 by the two leading American presidential candidates, John McCain and Barack Obama. After data cleansing and sentence delimitation, every sentence was treated as a query for the knowledge base. Figure 5.3 illustrates 21 speeches and their relation to 135 goal categories. Each cell contains a weight describing the relative importance of a given goal category for a particular speech. This figure indicates that certain categories dominate throughout all speeches analyzed in this study such as the goal category “`Helping Others`” or “`Charity`”, while other categories exhibit temporal bursts, for example “`Being Better Than Others`”.

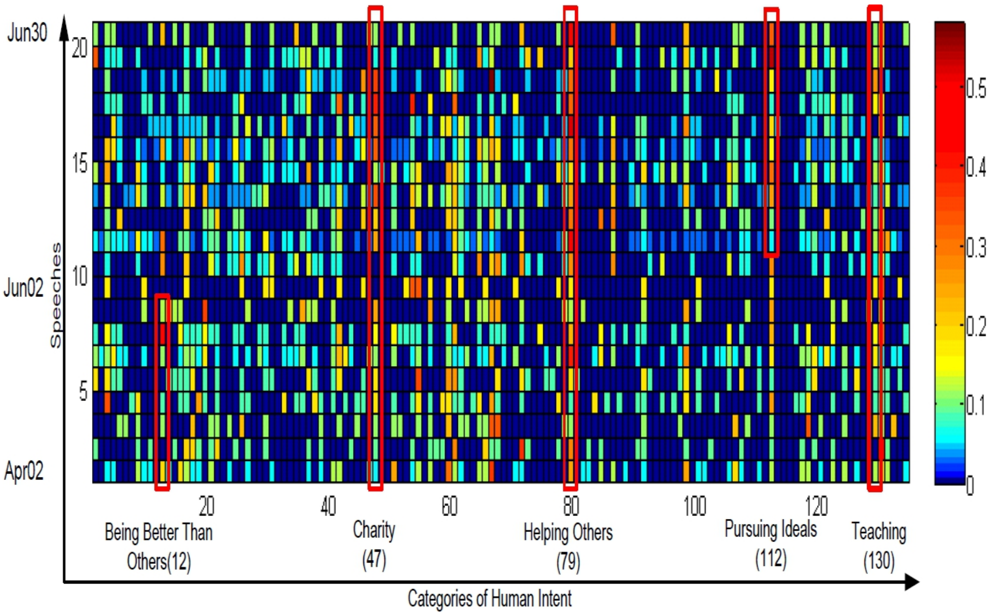


Figure 5.3: Overview of intentional annotations for 21 speeches given by Barack Obama in April and June 2008. Selected categories which are predominant over a certain period of time are highlighted. (taken from [86])

The data can be analyzed from a number of perspectives. Figure 5.4 illustrates goal profiles for an intentional comparison of Barack Obama’s and John McCain’s speeches. At a first glance, similarities and differences between the two candidates can easily be recognized, providing some sort of intentional summary of the speech contents. Both candidates conveyed messages to their audience that were often assigned to high-level goal categories such as "`Leader`", "`Helping Others`" and "`Charity`".

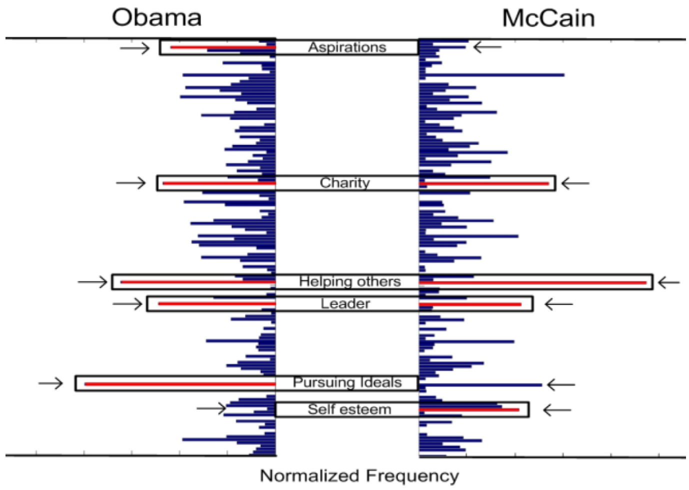


Figure 5.4: Goal profiles to compare Obama's and McCain's speeches. Results are averaged over 44 (21+23) speeches (April and June 2008). Predominant categories such as "`Charity`" are highlighted. (taken from [86])

Figure 5.4 also reveals goal categories that are stronger associated with one of the two candidates: categories such as "`Self Esteem`" have a higher weight for McCain's speeches whereas Obama's speeches seem to emphasize other categories such as "`Pursuing Ideals`" and "`Aspirations`".

5.4.1.1 Automated vs. Manual Categorization

To evaluate the quality of the presented implementation, automated categorizations were compared to categorizations produced in a human subject study. All three annotators were Computer Science graduate students. Two human subjects annotated Obama's speeches and assigned goal categories to sentences that the subject believed would contain indicative actions.

The annotators judged 3,722 sentences from 21 speeches and agreed upon 3,382 sentences to either assign no or at least one category to the same sentence. The corresponding Kappa κ =0.82 reflects useful agreement amongst the raters. In case of McCain's speeches, the third human subject annotated McCain's 23 speeches, altogether 2,677 sentences. Manual annotations were used to produce a ranking of goal categories for each candidate. In case of Obama's speeches, the union of annotations was taken produced by the two human subjects to mitigate data sparsity. Table 5.2 and Table 5.3 present the most frequent goal categories assigned either automatically or manually based on the aggregation of 21 speeches by Barack Obama and 23 speeches by John McCain. Out of the top 25 automatically assigned goal categories, the manually assigned goal categories agreed in 10 cases (40%). Agreement for McCain's speeches was similar, with 11 (44%) goal categories shared by automated and manual category ranking.

Automated Categorization (21 Speeches by Barack Obama)		Rank	**Manual Annotation** (21 Speeches by Barack Obama)	
Goal Category	**Weight**		**Goal Category**	**Weight**
Pursuing ideals	0.1991	1	Helping others	0.1798
Helping others	0.1615	2	Contribution	0.0944
Leader	0.1223	3	Difficult things	0.0831
Charity	0.1177	4	Bills	0.0742
Aspirations	0.1094	5	Job	0.0607
Being free	0.0993	6	Seeking equality	0.0494
Teaching	0.0968	7	Charity	0.0449
Being better than others	0.0965	8	Education	0.0404
Control over others	0.0956	9	Feeling safe	0.0404
Being creative	0.0940	10	Being better than others	0.0382
Education	0.0886	11	Seeking fairness	0.0382
Exercising	0.0874	12	Being responsible	0.0270
Ethical	0.0807	13	Being ambitious	0.0247
Exploring	0.0797	14	Money	0.0247
Feeling safe	0.0771	15	Being innovative	0.0202
Being likeable	0.0771	16	Control over others	0.0157
Content with myself	0.0762	17	Seeking justice	0.0157
Money	0.0721	18	Avoiding failure	0.0157
Attracting sexually	0.0709	19	Overcoming failure	0.0157
Knowing many others	0.0645	20	Teaching	0.0112
Easy life	0.0644	21	Providing family	0.0112
Being curious	0.0576	22	Own guidelines	0.0112
Avoiding stress	0.0550	23	Close children	0.0090
Sexual experiences	0.0540	24	Leader	0.0007
Being self-sufficient	0.0530	25	Being free	0.0045

Table 5.2: Comparison of the top 25 automatically assigned goal categories and human annotators for Obama's speeches. Weights represent normalized frequency values. Highlighted entries represent entries that were assigned by both automated and manual categorization. (taken from [86])

Highlighted entries in Table 5.2 and Table 5.3 represent entries that were assigned by both automated as well as manual categorization approaches. In order to gauge the quality of the automatically inferred goal annotations, the top 25 manual annotations were used as relevant annotations (right columns in Table 5.2 and Table 5.3) and the remaining manual annotations were judged to be irrelevant. Using the manual annotations as "ground truth", Figure 5.5 shows the performance of automatically inferred annotations in comparison to a simple baseline approach. The baseline approach ranks goal categories in a random manner.

Figure 5.5 shows that the presented approach outperforms the simple baseline approach (up to 70% with respect to recall levels). The results illustrate that for up to 40% recall (10 relevant annotations) this approach achieves a precision of 50% and above. While there is room for improvement, the results demonstrate the principle feasibility of automatically categorizing textual resources into human goal categories and represent a first step towards more sophisticated approaches.

Limitations: The process of automatically generating intentional annotations faces a number of challenges illustrated by following example:

Automated Categorization (23 Speeches by John McCain)		Rank	**Manual Annotation** (23 Speeches by John McCain)	
Goal Category	**Weight**		**Goal Category**	**Weight**
Helping others	0.2368	1	Avoiding failure	0.0958
Being better than others	0.1513	2	Aspirations	0.0949
Charity	0.1350	3	Standing up for beliefs	0.0873
Pursuing ideals	0.1278	4	Helping others	0.0863
Leader	0.1058	5	Being respected	0.0852
Self esteem	0.1039	6	Pursuing ideals	0.0586
Ethical	0.1030	7	Being recognized	0.0543
Money	0.0990	8	Persuading others	0.0383
Being socially attractive	0.0919	9	Being responsible	0.0362
Seeking justice	0.0862	10	Overcoming failure	0.0319
Seeking fairness	0.0811	11	Novel ideas	0.0309
Being intelligent	0.0805	12	Own guidelines	0.0277
Easy life	0.0773	13	Leader	0.0266
Belonging	0.0747	14	Support from others	0.0266
Career	0.0738	15	Being better than others	0.0191
Peace of mind	0.0673	16	Control over others	0.0181
Being honest	0.0653	17	Teaching	0.0170
Teaching	0.0651	18	Others' trust	0.0170
Feeling safe	0.0643	19	Seeking fairness	0.0170
Being respected	0.0616	20	Being honest	0.0160
Being creative	0.0590	21	Seeking justice	0.0150
Good parent	0.0567	22	Freedom of choice	0.0128
Personal growth	0.0543	23	Career	0.0128
Content with myself	0.0529	24	Seeking equality	0.0110
Being responsible	0.0525	25	Taking care of family	0.0110

Table 5.3: Comparison of the top 25 automatically assigned goal categories and human annotators for McCain's speeches. Weights represent normalized frequency values. Highlighted entries represent entries that were assigned by both automated and human categorization. (taken from [86])

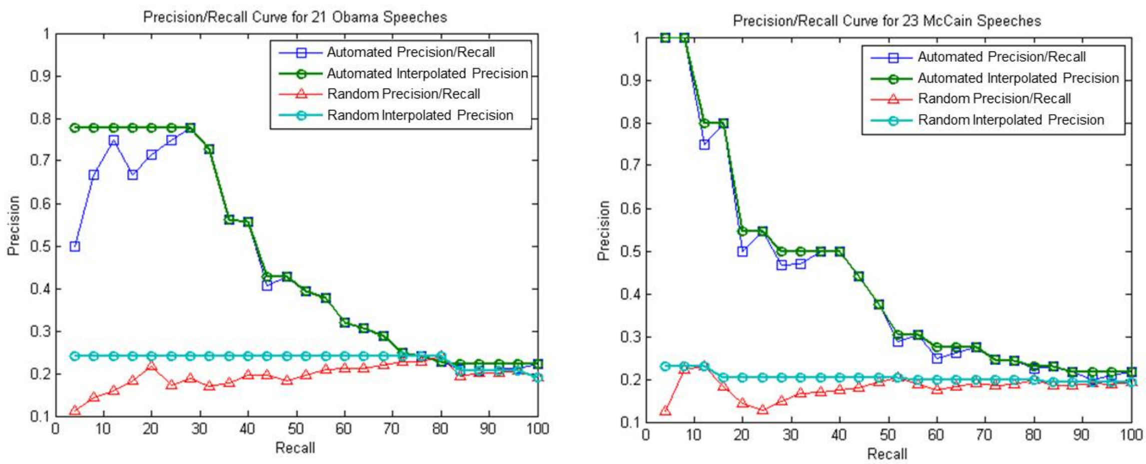


Figure 5.5: Comparison of the automated vs. random approaches for Obama's (left) and McCain's (right) speeches in terms of precision and recall. (taken from [86])

Consider the search query: *"in order to age well"*, which corresponds to the goal category "`Looking Young`". Among other results, this query could produce the following problematic search result:

> "Cork has been used for over 400 years, and many winemakers today still believe that *in order to age well*, wine <u>needs gradual exposure to oxygen.</u>"

Such problems cause sentences being misclassified, and negatively influence results. However, because the automated approach is based on aggregating evidence and taking a "winner takes it all" approach, it is tolerant against occasional misclassification of sentences, and the evaluation of knowledge base entries revealed that a majority of indicative actions represent suitable proxies for the automated intentional annotation task. However, an option to reduce this problem in future work could be to employ parsing to alleviate the semantic problem (cf. [178]) and/or using machine learning techniques to distinguish between sentences that should be assigned goal categories and those that should not.

5.4.2 Characteristics of Human Goal Knowledge

5.4.2.1 Knowledge Base Characteristics

On a general level, the usefulness of a knowledge base can suffer from a single or a combination of the following issues: (i) the knowledge base entry does not contain an indicative action, (ii) the entry contains an action but it is unrelated to the corresponding goal category and (iii) the entries for a given category only represent a minor fraction of possible actions. Combined, these factors have the potential to introduce noise and bias to the knowledge base. The following paragraphs aim to estimate the usefulness of the knowledge base by investigating qualitative and quantitative aspects.

The minimum number of knowledge base entries per category was 12 (Category: "`Firm Values`"), the maximum number was 4,497 (Category: "`Helping Others`") and the average number was 752. The final number of sentences in the knowledge base totaled 101,490. The distribution of knowledge base entries is skewed as depicted in Figure 5.6 yet only a minor fraction of categories received less than 100 entries.

In order to evaluate the quality of knowledge base entries, a random sample of 674 entries was drawn from the knowledge base. The sample was judged by a linguistics undergraduate student with regard to 1) whether the entry contains indicative actions and 2) whether the entry is relevant for the corresponding category.

57% of the entries in the sample contained actions indicative of the corresponding goal category, which says that while there is a certain level of false positives, the majority of entries are useful. To evaluate relevance of knowledge base entries, comparative analyses of different intentional relations were conducted. Table 5.4 shows success rates for every intentional relation where success rate is defined as #correct entries divided by #all entries regarding a particular intentional relation. Two relations, i.e. "`in order to`" and "`for the purpose of`", exhibit success rates beyond 50% suggesting better quantities and higher-quality knowledge better entries than others, such as "`essential for`" and "`necessary for`". For the results reported in this study, only the intentional relations "`in order to`" and "`for the purpose of`" were used, information acquired through other relations was discarded as a result of this evaluation. This choice of intentional relations does not cover all potential goal – action pairs. Enforcing this restriction will result in

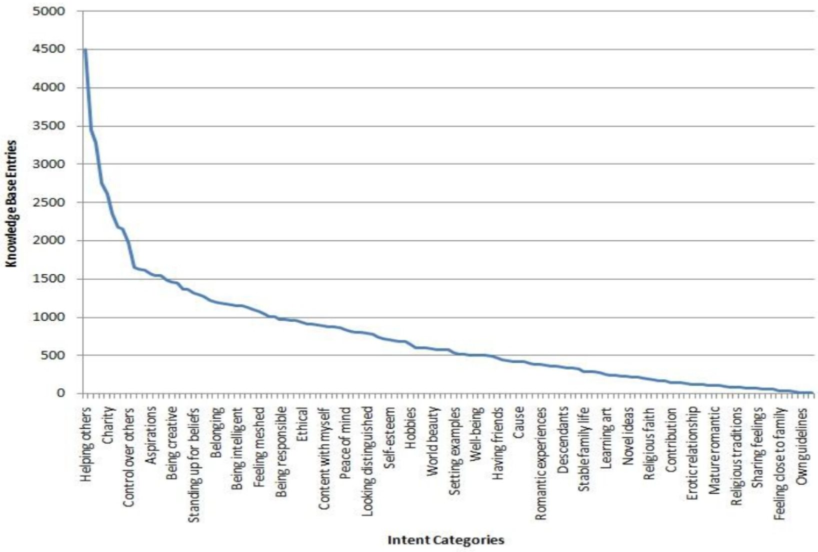


Figure 5.6: Distribution of knowledge base entries per category. (taken from [86])

neglecting many pairs. Yet, this work does not aim to achieve an optimal coverage of goal – action pairs for every category but a reasonably sufficient one.

	in order to	essential for	necessary for	critical for	for the purpose of
Success Rate	59.2 %	32%	35.5%	16.7%	59.8%

Table 5.4: Quality of the five used intentional relations. Only those exhibiting a success rate beyond 50% were taken into account for further processing steps. (taken from [86])

In addition to this evaluation, people's agreement on sentences was investigated that contain indicative actions. A Cohen's Kappa coefficient [35] of $\kappa = 0.79$ was obtained. This indicates that human annotators can largely agree on what constitutes suitable entries in the knowledge base.

Limitations: While the knowledge base was helpful to produce intentional annotations that achieve a useful level of agreement with human annotators, it suffers from (i) a skewed distribution of #entries per category, (ii) a certain amount of false-positive indicative actions, (iii) and noise. Some of these concerns were addressed in this work (e.g. by eliminating intentional relations that tend to produce false positives), but there are several opportunities for improvement. To study whether the skewed distribution of entries in the knowledge base biases the intentional analysis of textual resources, additional analyses were conducted. Spearman's rank correlation was calculated between the ranked list of knowledge base categories (where the category with the highest number of entries ranks first) and the automatically assigned goal categories for the speeches by Obama and McCain. The results of this calculation reveal that there is a weak correlation between the ranked goal categories and the automatically assigned annotations, i.e. Obama = 0.38 and McCain = 0.42. The assignment of categories thus appears to be weakly biased towards the number of entries per category in the knowledge base.

5.4.2.2 Temporal Evolution

Temporal information about the date of the speeches is available allowing a number of interesting additional analyses. For example, Figure 5.7 illustrates the temporal evolution of goal categories over 21 speeches given by Barack Obama in April and June 2008.

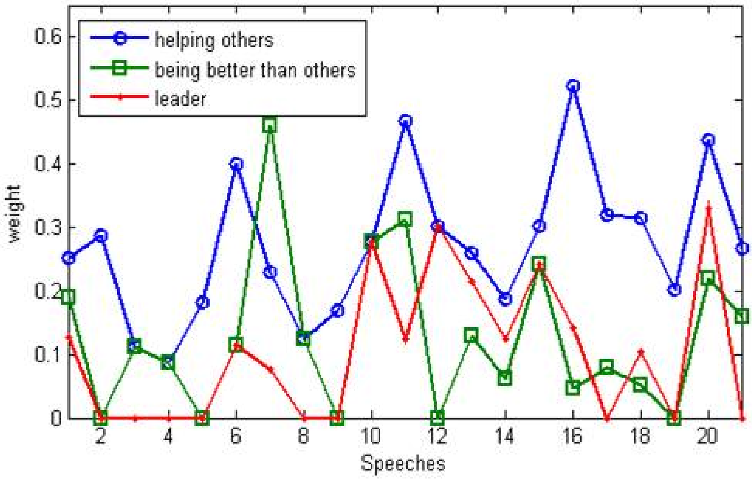


Figure 5.7: Temporal evolution chart of three selected Goal Categories “`Helping Others`”, ”`Being Better Than Others`“ and ”`Leader`“ over 21 speeches given by B. Obama. (taken from [86])

Several observations can be made when focusing the comparison on a few selected goal categories. For example, the chart in Figure 5.7 shows that certain categories such as “`Helping Others`” are prominent over the entire observed period. Peaks where individual goal categories dominate can be easily detected, such as “`Being better than others`” in speech No. 7 or “`Leader`” in speech No. 20. It is conceivable that applying this type of analysis to other resources, such as search query logs, blog posts or discussion forums, could open up new opportunities to interlink textual resources on the web or to monitor social media activities.

5.5 Discussion

This case study fosters our understanding of human goal knowledge in natural language text, in particular transcripts of political speeches (monologues). It reveals how people express their goals, i.e. that they often use actions (tasks) as surrogates for goal statements and that these surrogates can in further consequence be used to assign goal categories. This observation (i) suggests and corroborates the integration of additional components into the proposed data model, e.g. the *GoalCategory* class, and (ii) supports knowledge engineers in their decision on which corpus to use for mining large amounts of human goal instances. In addition, this study shows that identifying the intentional actor, i.e. instantiating the data model's *Agent* class, is a non-trivial task and will require additional information, e.g. from semantic role labeling literature. Overall, this case study benefits the data model's development by providing insights into relational structures between goals, tasks, agents and goal categories.

This gained understanding relates to the *Elicitation* step in the goal knowledge engineering process which focuses on resource compilation and familiarization with the respective community. In addition, this case study addresses the *Acquisition* step and the *Validation* step, in particular the automatic acquisition of *Task* class instances. The automatic acquisition is based on existing (i) search and retrieval mechanisms to compile a candidate set of task instances, (ii) information extraction techniques to process the task instances and (iii) machine learning principles to assign goal categories. In course of this case study human goal knowledge has been made applicable and accessible by instantiating and using following components of the data model: the *Goal* class, the *GoalCategory* class, the *Task* class and the *MeansForEndRelation* class.

Chapter 6

Case Study 03: Constructing Goal Concept Hierarchies

6.1 Contribution to PhD Objectives

This case study's emphasis lies on the automatic inference and extraction of relations between intentional concepts such as goals, subgoals and tasks. As illustrated in Figure 6.1, the following components of the proposed data model are instantiated: the *Goal* class, the *Task* class and the *MeansForEndRelation* class. For that purpose, the case study explores methods (i) to automatically infer hierarchical goal structures, i.e. relating goals to subgoals, and (ii) to automatically complement goal structures by reasonable actions (tasks) which potentially contribute to their accomplishment, i.e. relating goals to tasks. The corresponding problem setting encompasses the construction of concept hierarchies with respect to domain-specific human goals in this case from the health domain.

Covered Components	**Description**
Goal	The case study focuses on human goal instances from the health domain. To compile a set of health-related goal instances, previous work was used to extract instances from search query logs. The majority of these health-related goals are hard goals, i.e. objective success criteria can be applied. In the experiments only instances are considered whose goal type attribute is either achieve, e.g. "`buy healthy food`" or maintain, e.g. "`keep emotional health`". Though avoid goals are neglected, they still provide interesting insights into people's fears and concerns. These instances include, for example, "`lose mind`", "`lose hair`" or "`catch aids`".
Task	Complementing goal concept hierarchies with tasks widens the range of operations which can be applied, e.g. generating action sequences to support planning procedures. Task candidates are automatically retrieved and evaluated by a "wisdom of crowds" approach.
IntentionalRelation, MeansForEndRelation	This case study's emphasis lies on the automatic inference and extraction of relations between intentional concepts such as goals, subgoals or tasks. In the first part, hierarchical relations amongst goals are inferred. The second part seeks to automatically relate goals in the hierarchy to action (tasks) which contribute to their accomplishment. Relations in both parts can be classified as instances of the *MeansForEndRelation* class.

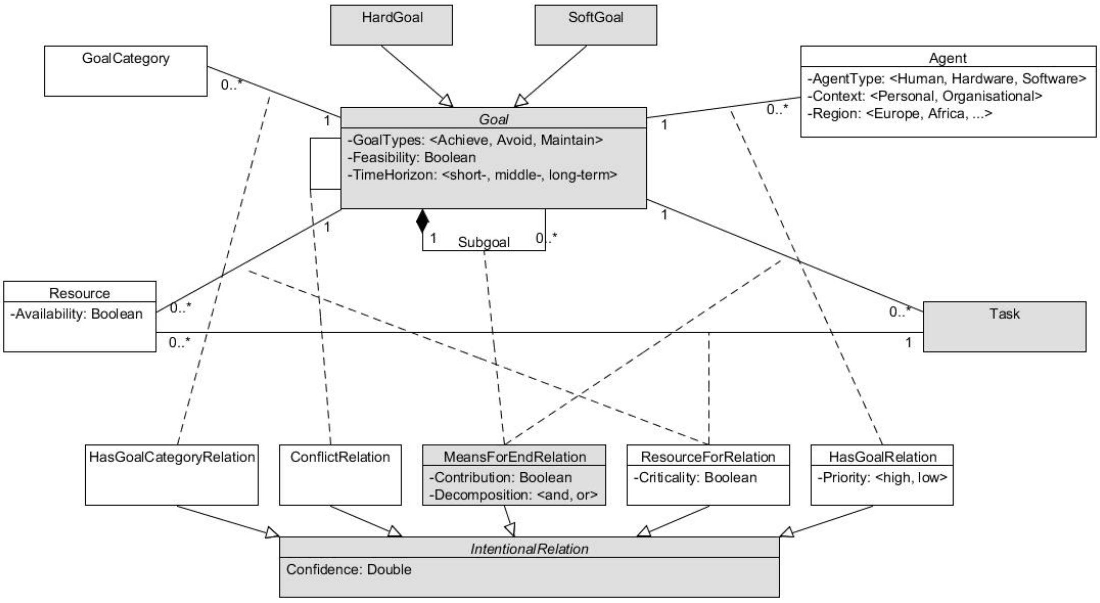


Figure 6.1: Components of the proposed data model on human goal knowledge covered by this case study are highlighted.

This case study seeks to automatically construct a hierarchy of goal concepts which reflects goal knowledge from the health domain, i.e. hierarchically relating health-related human goal concepts with each other. Existing techniques from unsupervised machine learning, i.e. clustering algorithms, are applied and adapted for the construction process.

With respect to the core objectives of this PhD thesis, this case study makes following two contributions:

1.) This case study explores methods to automatically infer hierarchical structures amongst goals, i.e. introducing Means-Ends relations between goals and subgoals. To generate goal concept hierarchies, established clustering techniques such as Bi-Section K-Means, Formal Concept Analysis and Hierarchical Agglomerative Clustering are investigated. These techniques were successfully applied in the past to construct concept hierarchies (cf. [32]). To compare resulting concept hierarchies, this work constructs a ground truth and calculates taxonomic overlaps (cf. [106]).

2.) The case study seeks to complement resulting goal concept hierarchies by tasks, i.e. introducing Means-Ends relations between goals and actions which potentially contribute to their accomplishment. To give an example, the action "`use condoms`" potentially contributes to accomplishing the health-related goal "`prevent aids`". Enriching goal concept hierarchies by Means-Ends relations enables more complex operations on goal knowledge, e.g. the generation of action sequences in planning procedures (cf. [24]).

Queries containing explicit goal instances	Queries not containing explicit goal instances
“lose weight fast”	“weight loss”
“writing medical case studies”	“case study research”
“passing a drug test”	“drug test”

Table 6.1: Examples of health-related queries which do or do not contain human goal instances. (taken from [85])

6.2 Motivation

To enable planning or reasoning, knowledge about human goals needs to be structured and organized, e.g. by arranging it in hierarchical structures. In this context, hierarchies of goal concepts have proven valuable in several research areas including (i) web search ([21], [186]), (ii) intelligent user interfaces ([157], [102]) or (iii) semantic task retrieval ([51], [120]). Concept hierarchies are meant to mimic mental constructs, thus reflecting a domain's abstract representation. (i) In web search, goal concept hierarchies organize users' underlying search goals to inform and improve search engines' retrieval performances. (ii) By utilizing structured human goal knowledge, intelligent user interfaces are capable of better understanding relationships between people's goals and their actions. (iii) Finding appropriate web services is facilitated by a better understanding of which tasks or actions are required to accomplish people's goals. Fukazawa et al. [51] have already experimented with modelling goal structures and connecting them to web services.

With respect to this case study's contributions, two research questions are addressed:

> RQ 01: How accurately can concept hierarchies of health-related human goals be automatically inferred? (see Section 6.4.1)
>
> RQ 02: To what extent is it feasible to automatically complement goal concept hierarchies with Means-Ends relations, i.e. relating goals to actions which potentially contribute to their accomplishment? (see Section 6.4.2)

6.3 Experimental Setup

This section describes how search query logs are utilized as a source for extracting (health-related) human goal instances. These instances are then transformed into concepts and manually organized in a hierarchy. This hand-crafted hierarchy will serve as ground truth in the experiments.

6.3.1 Health Related Goals from Search Query Logs

Search query logs have been shown to be a particularly valuable source for extracting human goal instances (cf. [161]). Each submitted query represents a person's search goal which can be expressed explicitly or implicitly. Table 6.1 presents some examples of health-related queries which do or do not contain instances of human goals (obtained from [132]):
A previously developed approach [164] is applied to identify those queries which contain explicit goal instances. The algorithm automatically extracts ∼90,000 queries of which 77 out of 100

queries actually contained explicit goal instances (77% precision). The ~90,000 queries were then filtered with respect to (i) health-related keywords such as "healthy" or "disease" and (ii) health-related URLs such as "http://www.camh.net/" or "http://www.healthandage.com/". Keywords as well as URLs were compiled and refined manually, i.e. by means of brainstorming sessions, manual inspection and using the open directory project[1]. The keyword-based approach identified fewer health-related goal queries than the URL-based approach, yet with a higher accuracy, i.e. 73.2% over 44.8%. To gather a useful set of health-related goal queries, both filtering approaches were combined. Duplicates were removed as well as false positives such as "`follow your heart`", "`donate your car`" or "`find healthy dog food`".

To reduce ambiguity, health-related goal instances were conceptualized, i.e. converted into goal concepts. Forming concepts was done by normalizing individual instances of human goals, i.e. by removing stop words and punctuations, by lemmatizing verbs, i.e. reducing them to their base form and by transforming nouns to their singular form like similar work by [60]. A concept then encompasses all instances that normalize to the same text as illustrated in the following:

Human Goal Concepts	**Corresponding Human Goal Instances**
"increase health"	"increasing health", "increase your health"
"lose weight"	"lose some weights", "losing a lot of weight"

As a consequence, goal concepts have a correspondence to verb phrases. Yet, goal concepts are not literal strings of text but stand for mental artifacts. A concept can represent related or synonymous instances of human goals. For this case study's experiments, the set of health-related goal concepts totals 489. This set size is considered to be sufficient since this work focuses on exploring means to automatically construct goal concept hierarchies.

6.3.2 Ground Truth

To the best of my knowledge, there is no publicly available ground truth which hierarchically relates health-related goal concepts. To that end, it was decided to handcraft a ground truth which will enable a comparison of algorithms and thus an assessment of the quality of resulting concept hierarchies. The generation of the ground truth consisted of three steps: (1) to group ~500 health-related goal concepts by similar topics, e.g. pregnancy-related goal concepts, (2) to hierarchically relate these topic groups, e.g. to connect plastic surgery-related and skin-related goals by introducing the higher level cluster beauty-related goals, and eventually (3) to hierarchically relate goal concepts within each group. Table 6.2 shows the top-level structure of the ground truth after the first two steps.

The third step introduces hierarchical structures within each topic group. For each group, general goal concepts are selected as higher level node candidates. General goal concepts tend to consist of a verb and only few objects. Afterwards, the focus is on specializations of general goal concepts. Two kinds of specializations are considered: 1) object specialization, e.g. "`buy cheap diet pill`" is an object specialization of "`buy diet pill`" and 2) method specialization, e.g., "`prepare weight loss diet`" is a method specialization of "`lose weight`". These two steps are

[1] *http : //www.dmoz.org/*

1^{st} **Level Concepts**	2^{nd} **Level Concepts**
Sex & Baby	Sex, Pregnancy, Baby
Health & Beauty	Daily Health-Care, Healthy Diet, Weight-Control, Fitness, Beauty, Face-Care, Mental-Health
Disease	Body-Disease, Brain-Disease
Other	Authority, Drugs

Table 6.2: The ground truth's 1st and 2nd level structures. (taken from [85])

repeated until no other concepts remain in the respective group. Figure 6.2 visualizes a small excerpt from the ground truth.

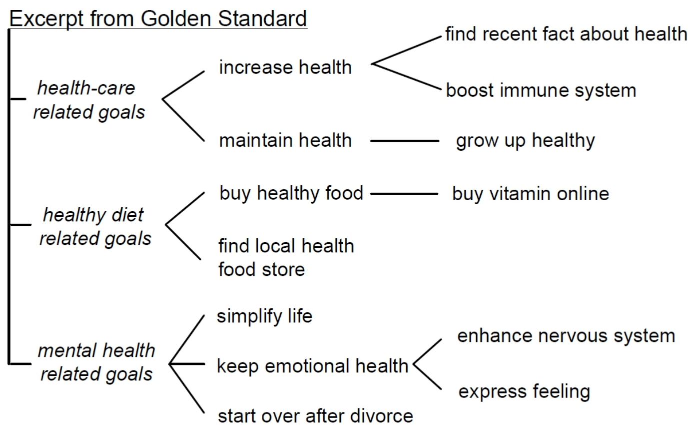


Figure 6.2: A small excerpt from the ground truth. Higher-level categories are inserted by hand when no appropriate goal concept was present (in italics). (taken from [85])

This section is concluded by analyzing the resulting structure of the ground truth and its components. This analysis might give some indication of health-related concerns as well as e-health trends (at that time). The ground truth's four top level groups partly reflect people's health conditions, their situations in life as well as what their fears and problems are. The first group "Sex & Baby" comprises all goal concepts related to sexual activities as well as consequences, e.g. pregnancy, baby care, yet also dealing with AIDS. The second topic group "Health & Beauty", on the one hand, reflects people's longing for an immaculate appearance, e.g. by having plastic surgery. On the other hand, it reveals a trend towards a healthier lifestyle including diet, fitness and weight issues. The third group "Disease" contains goal concepts which are related to physical dysfunctions, e.g. "`eliminate large kidney stone`" or "`go off seizure medication`". The fourth group "Other" comprises all remaining health-related goal concepts which were further divided into authority-related and drugs-related goals. In addition, there were goals present which people did not want to achieve but rather avoid (indicated by the keyword "not"). To give some examples, selected avoid goals include: "`keep getting sick`", "`have nocturnal seizure`", "`have dark circle under eye`", "`lose mind`", "`lose hair`" or "`catch aids`". These goals were not included in the ground truth, yet these instances are considered interesting for they provide insights about people's fears and concerns.

6.3.3 Algorithmic Approach

Three algorithms are examined which are introduced in [31] as viable practices to automatically construct concept hierarchies: Formal Concept Analysis (FCA), Hierarchical Agglomerative Clustering (HAC), and Bi-Section K-Means (BSKM). Three feature types are experimented with as representation mode: token-based (T), neighborhood-based (N) and click-through-based (C).

Feature Type	Description
token-based (T)	For every goal concept, all corresponding goal instances are tokenized and sanitized, e.g. stop words are removed. Then a characteristic token vector is formed to represent the goal concept.
neighborhood-based (N)	It is assumed that neighboring queries are suitable features to represent goal queries. This neighborhood information is obtained from search query logs. The set of neighboring queries encompasses queries issued by the same user before and after the goal query. To give an example, the query "lose weight fast" possesses a number of neighboring queries including "weight loss supplements", "types of diet pills" or "Lipo6". After tokenization and sanitization steps (as in the token-based approach), a characteristic term vector is generated using a set of neighboring queries without including the goal query.
click-through-based (C)	The search query log also contains click-through information, i.e. given a query which resulting URLs were clicked. The search query log is traversed for each human goal query and all corresponding clicked URLs are collected. To be added to the feature vector, each URL must have been clicked at least twice.

FCA, HAC and BSKM were implemented and applied using standard settings without optimization steps:

Formal Concept Analysis (FCA) originated as data analysis technique before it was successfully applied to construct concept hierarchies [32]. In the experiments an existing java library was used which provides an efficient implementation.
Hierarchical Agglomerative Clustering (HAC) is a similarity-based bottom-up clustering algorithm. In the experiments, HAC algorithm implementation according to [31] was used. The cosine measure was used as similarity measure and as linkage metric we chose single linkage since it possesses the lowest computational complexity, i.e. $O(n^2)$, compared to the other linkage metrics.
Bi-Section K-Means (BSKM) represents a clustering technique that repeatedly applies the traditional K-Means algorithm. The Bi-Section K-Means algorithm was implemented according to [31]. As cluster centers, two data points were chosen randomly and as similarity measure, the cosine measure was used.

As evaluation metric, taxonomic overlap was used pioneered by [106] as one of the first metrics to compare two concept hierarchies with each other. The metric allows a comparison not only on a lexical level but also on a conceptual one. The principal idea behind this metric is that two concept hierarchies are similar (i) if they have a lot of concepts in common and (ii) if these common concepts share many super/sub concepts. A high overlap between an automatically constructed concept hierarchy and the hand-crafted ground truth would thus indicate a good quality. To calculate the taxonomic overlap (TO) in respect to one common concept (c), following formula was used:

$$TO(c, O_1, O_2) = \max_{c' \in C_2} \frac{|SC(c,O_1) \cap SC(c',O_2)|}{|SC(c,O_1) \cup SC(c',O_2)|}$$

where O_1 represents the ground truth, O_2 the automatically constructed concept hierarchy, C_2 the set of O_2's concepts and SC stands for *semantic cotopy*, i.e. the set of concept c's super and sub concepts (cf. [106]). The calculation of the SC does not take into account concepts without name which have been created during the clustering process. By repeating this calculation for all concepts, an averaged $\overline{TO}$ is obtained which reflects the similarity between two concept hierarchies O_1 and O_2:

$$\overline{TO}(O_1, O_2) = \frac{1}{|C_1|} \sum_{c \in C_1} TO(c, O_1, O_2)$$

where $|C_1|$stands for the number of concepts in the ground truth O_1. The taxonomic overlap can be calculated in both directions, i.e. $\overline{TO}(O_1,O_2)$ equaling precision and $\overline{TO}(O_2,O_1)$ equaling recall. Since this work seeks to explore the potential of three algorithms to hierarchically structure a fixed set of goal concepts, all reported taxonomic overlaps represent precision results.

6.4 Results

6.4.1 Infering Hierarchical Goal Structures

This section explores the potential of three algorithmic approaches to automatically infer concept hierarchies of health-related human goals. To compare these approaches, taxonomic overlaps are calculated between resulting concept hierarchies and the ground truth. Table 6.3 summarizes taxonomic overlap results for all (algorithm, feature type) combinations. Each overlap value reflects the degree of matching between an automatically constructed goal concept hierarchy and the ground truth.

	Token-Based (T)	Neighborhood (N)	Click-Through (C)
BSKM	48.06 %	45.04 %	39.69 %
FCA	41.52 %	40.91 %	39.39 %
HAC	*50.82 %*	47.55 %	46.11 %

Table 6.3: Taxonomic overlaps for (algorithm, feature type) combinations. (taken from [85])

The combination (T, HAC) yields the highest taxonomic overlap value 50.82% and (C, FCA) the lowest value 39.39%. HAC and BSKM appear to be equally well suited for the construction process. All clustering algorithms achieve highest results when using the token-based features. In addition, experiments with combinations of feature types have been conducted which did not yield better results. To further investigate the applicability of the clustering algorithms, taxonomic overlaps for 14 subdomains are calculated which differ, e.g. in their number of concepts. The subdomains are formed by taking second level categories (see Table 6.2) as roots.

Results in Table 6.4 show that the HAC algorithm yields highest taxonomic overlap values in most cases, e.g. 59.67% for the authority subdomain. These results indicate that the degree of overlap does not depend on the number of concepts, i.e. high overlaps were achieved for

Health Sub Domain	#Concepts	Taxonomic Overlap		
		FCA	BSKM	HAC
Body-Disease	79	32.35%	44.46%	*44.64%*
Daily Health-Care	78	36.26%	41.91%	*50.92%*
Sex	58	35.82%	*47.81%*	46.09%
Beauty	54	38.65%	46.48%	*49.93%*
Weight control	52	32.00%	46.45%	*46.89%*
Pregnancy	51	38.74%	47.16%	*51.36%*
Mental Health	43	40.03%	47.13%	*50.44%*
Face-Care	40	42.98%	48.28%	*49.56%*
Healthy diet	39	39.24%	49.62%	*55.23%*
Fitness	28	41.30%	37.04%	*42.14%*
Baby	25	42.39%	*53.63%*	52.08%
Authority	23	51.13%	54.30%	*59.67%*
Brain-Disease	18	52.43%	49.8%	*54.59%*
Drug	10	58.00%	*58.72%*	57.91%

Table 6.4: Individual taxonomic overlaps for 14 subdomains (token-based features). (taken from [85])

subdomains with few and many concepts. To learn more about what causes low taxonomic overlaps, the Fitness health subdomain is examined which yields a taxonomic overlap of only 42.14%. Figure 6.3 compares excerpts from the clustering result (T, HAC) to the ground truth.

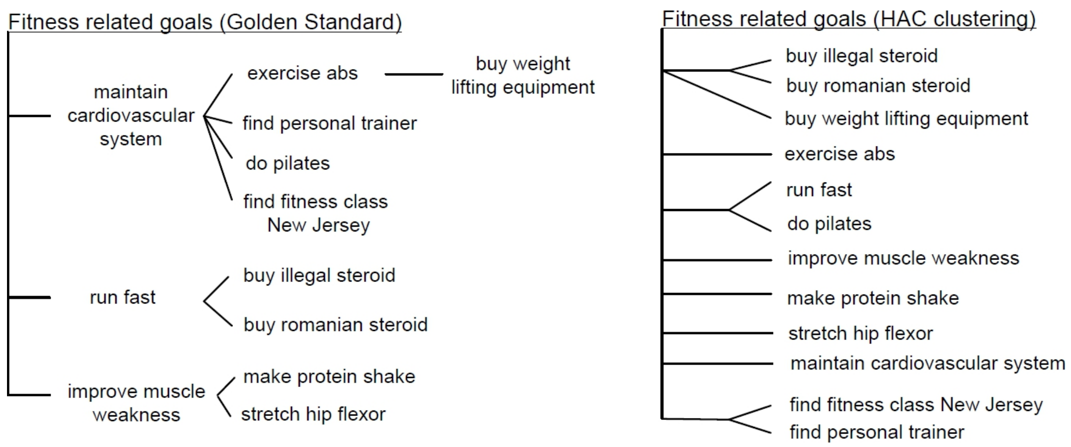


Figure 6.3: An excerpt of the ground truth's structure is compared to HAC's respective clustering result. The excerpts show goal concepts from the health subdomain "Fitness". (taken from [85])

From visual inspection, a primary reason for the low overlap appears to be the flattening of the hierarchical structure. This might be a consequence of the used feature representations not capturing the required information for generating a hierarchy according to the ground truth. This work thus proposes to use context information[2] to represent goal concepts. Context information can, for instance, be acquired from social media corpora such as weblogs (cf. [88]). This kind of representation might also better comply with FCA which did not seem to reach its full potential with the used features types. Lastly, an implication of token-based representations can be observed:

[2]According to Harris' distributional hypothesis [59].

human goal concepts are clustered together based on equal verbs, e.g. "buy" or "find", although they do not belong together from a semantic point of view. Separating verb from noun semantics and treating, e.g. weighting, them differently might alleviate this issue.

6.4.2 Complementing Concept Hierarchies with Means-Ends Relations

This section provides a prototypical implementation to automatically complement goal concept hierarchies with Means-Ends relations, i.e. relating goals to actions which potentially contribute to their accomplishment. Complementing goal concept hierarchies by Means-Ends relations widens the range of operations which can be applied, e.g. generating action sequences to support planning procedures. The advent of social media platforms allows automating the process of extracting Means-Ends relations from the web.

Yahoo!Answers (Y!A), a social media platform, is utilized as resource to extract candidate actions which contribute to health-related goals. Y!A mediates the process of people posing questions and of people answering questions. Question/answer pairs are of particular value since a question often represents a person's goal and its answers potentially contain means for accomplishment. This idea is illustrated by reviewing following question/answer pair from Y!A:

Question:	"How can I lose weight easily?"
Answer:	"The best way to lose weight or maintain a healthy weight is through changing your eating habits and exercising regularly."

This example indicates that actions such as "`change your eating habits`" or "`exercise regularly`" might contribute to the goal "`lose weight easily`". The Y!A's API is utilized to programmatically retrieve answers to submitted goal queries. Goals are translated into questions, e.g. by prefixing "how to", and are sent to Y!A. Answers are collected and prepared for (i) pre-processing, (ii) pattern-based action extraction and (iii) post-processing described in detail in the following.

Pre-Processing	It is assumed that clauses are natural boundaries for singular actions. During the pre-processing step, the task is to identify all clauses in the Y!A answers. A rule based approach is pursued by compiling a list of clause delimiters such as "if", "and" or "\n", as well as punctuations. Eventually, the clause's constituents are part-of-speech tagged using the Stanford log-linear part-of-speech tagger.
Pattern-Based Action Extraction	Extraction patterns are used to identify candidate actions. Extraction patterns are a combination of an indicator phrase and a verb phrase. By examining a small set of answers, a list of heuristic indicator phrases is manually compiled including phrases such as "you should", "you have to", "try", "start" or "you can". If an indicator phrase precedes a verb phrase (determined by part-of-speech tag information), the verb phrase is added to the list of candidate actions.
Post-Processing	Sanitization steps are applied: (i) candidate actions with less than three tokens are removed, (ii) duplicate entries are removed and (iii) too general actions and actions without proper object are removed by blacklisting; the manually compiled blacklist contains entries such as "do it", "make sure" or "know how".

To automate the ranking of candidate actions, a "wisdom of crowds" approach is used which is based on (web) statistics. If a candidate action often co-occurs with a human goal, it is assumed that it contributes to the goal's accomplishment and thus a high rank is assigned. To access web statistics, phrase searches are issued to the web using the Yahoo!BOSS API[3]. Query strings for goal/action pairs are constructed where manual processing is applied to increase the likelihood for hits. This manual processing includes correcting misspellings or adding personal pronouns at correct positions, e.g. "`clean out my body`". As ranking metric, an adapted version of point-wise mutual information [170] is used:

$$Score(goal, action) = \frac{Hits(action + "to" + goal)}{Hits(action) * Hits(goal)}$$

where the term Hits(goal) can be eliminated from the denominator as it is common to all candidate actions for a particular goal. This implementation selects the top three ranked candidate actions[4]. Actions are conceptualized as it is illustrated with goal concepts. Figure 6.4 shows an excerpt of the ground truth which is complemented with action concepts.

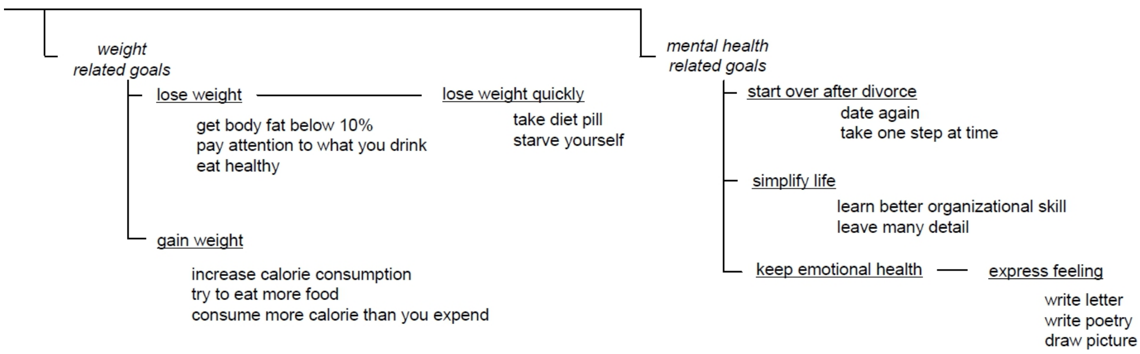


Figure 6.4: An excerpt of the ground truth is shown, i.e. weight and mental health related goals. Underlined goal concepts are complemented with action concepts. (taken from [85])

The presented approach is capable of complementing goal concept hierarchies with reasonable action concepts (cf. Figure 6.4). However, manual assessment of extracted and highly ranked action candidates reveals that the approach is vulnerable to false positives (due to the lack of human judgment). Moreover, as Y!A is used, it is assume that goals (as part of a question) have already been asked and answered. Hence, actions could be extracted only for a third of all goals. To find more goal/action pairs, other social web sources such as how-to sites appear to be promising. Lastly, goals without extracted actions were examined more closely. While some goals are simply not contained in Y!A such as "`recover from open reduction ankle surgery`" or "`make sure skin heals after cryosurgery`", it is hypothesized that Question/Answer systems are not suitable for "find" goals, e.g. "`find vascular surgeon`", and "buy" goals, e.g. "`buy weight watchers food online`". Specialized services on the web, e.g. identified by semantic retrieval [120], appear to be more suitable means for these goals.

[3]http://developer.yahoo.com/search/boss/
[4]Other strategies are conceivable, e.g. introducing a threshold.

6.5 Discussion

This case study benefits the development of the proposed data model by gaining a better understanding of relational structures amongst intentional concepts such as goals, subgoals and tasks in a particular domain (health). Automating the instantiation process of components such as the *Goal* class, the *Task* class and the *MeansForEndRelation* class, corroborates and validates the proposed data model. The presented algorithmic approaches relate to the meta-process' *Acquisition* step and *Validation* step. In the *Acquisition* step, existing techniques from (i) natural language processing, (ii) information retrieval and (iii) machine learning are adopted to construct hierarchical goal structures. The *Validation* step verifies the correctness of human goal knowledge, i.e. automating the ranking of candidate actions by a "wisdom of crowds" approach which is based on (web) statistics. In course of automating the process of engineering goal knowledge, the study shows that human participation is still necessary to obtain qualitative results. With respect to constructing goal concept hierarchies, human participation is required for pre- and post-processing, e.g. to compile gazetteer lists or to devise extraction patterns.

This case study provides algorithmic approaches to automate the inference and extraction of Means-Ends relations which relate to existing work to automatically extracting semantic relations (cf. [129], [147] or [11]). The practical application addresses the construction of health-related goal concept hierarchies which are supposed to be mental constructs to organize knowledge. Concept hierarchies are considered vital for knowledge-based systems since they allow for a concise and abstract representation of a domain. Hierarchies of goal concepts have already been proven valuable in several research areas including (i) web search (cf. [186]), (ii) intelligent user interfaces (cf. [157]) or (iii) semantic task retrieval (cf. [120]).

Chapter 7

Application Scenarios

In a series of application scenarios the proposed framework's applicability, relevance and usefulness are demonstrated and to a certain extent validated by applying human goal knowledge to various research areas including commonsense knowledge acquisition, information retrieval, text understanding and goal-oriented user interfaces. Prototypical implementations and results from these application scenarios will prove helpful for (i) conducting intelligent, goal-oriented analyses and for (ii) building and designing goal-oriented, intelligent systems.

Section 7.1 and Section 7.2 build upon theoretical considerations from Chapter 4 and seek to implement them. Section 7.1 explores two methods to complement Concept Net's [104] commonsense knowledge: (i) refinement and (ii) extension. The refinement process seeks to find more specialized versions of a given goal, e.g. "`write an informative paper`" or "`write an argumentative paper`" to the given goal "`write a paper`". With respect to the framework's data model, new instances of the *Goal* class are generated. The extension process seeks to link goal instance from query logs to ConceptNet actions via Means-Ends relations. Thus, this application scenario also generates new instances of *MeansForEndRelation* class, *Goal* class and the *Task* class. Section 7.2 presents an alternative approach to query suggestion, i.e. it explores methods to make a person's search goal more explicit. The suggestion mechanism builds upon goal instances acquired from query logs. Given a short, ambiguous query such as "house", the algorithm returns queries such as "`repair house`" or "`insure house`" which potentially explicate the person's search goal. In this case, no data model components are newly instantiated, yet the presented approach can be used to introduce a certain degree of similarity between suggested goal instances.

Section 7.3 implements ideas from Chapter 5 and presents iTag, a prototype for automatically annotating textual resources with human goal categories and explores possibilities to visually evaluate text from a goal-oriented perspective. By introducing and applying formalized human goal knowledge, this application scenario thus opens up further research opportunities in areas such as (i) automatic tag generation and (ii) visual analytics.

7.1 Complementing Commonsense Knowledge Bases

This section investigates whether human goals acquired from search query logs have the potential to contribute to commonsense knowledge. Two methods are explored for complementing ConceptNet: (i) refinement and (ii) extension. The internal structure of ConceptNet facilitates refinements and extensions simply by adding novel triples consisting of two concepts connected by a semantic relation. First, refining existing nodes is proposed by making them more specific, e.g. by adding adjectives or adverbs. The ConceptNet goal finding friends can thus be refined by following selected goals "`find old friends`", "`find a lost friend`" or "`find free military friends`" (obtained from search query logs). Table 7.1 provides potential refinements for a set of commonsense goals from ConceptNet.

ConceptNet Goal	List of Refinements From Query Logs
now buy this car	buy new car (20), buy a cheap car (2), buying rental cars (2), buy electric car (2), buy a used car (19), buy old cars (3), buying wise car (1)
finding friends	to find old friends (4), find high school friends(1), find lost friend (11), find best friends (1), find elementary school friend (1), find free online friends (1), find past military friends (1)
writing a paper	write an argumentative paper (1), write an informative paper (1), write an autobiographical paper (1), write a narrative paper (1)
cutting your hair	cutting my own hair (1), cut short hair (3), cut black hair (1), cutting long hair (1)
feeding the baby	feeding a newborn baby (1)
find a partner	finding sexual partners (1)
train a dog	train an abused dog (1), train a deaf dog (1)
making coffee	make perfect coffee (1), make a flaming coffee (1)

Table 7.1: This table presents exemplary refinement candidates for selected ConceptNet goals. Frequencies from search query logs are denoted in brackets and provide a first indicator towards quality of the candidates. (taken from [161])

In addition, frequency information is denoted in brackets for each potential refinement in Table 7.1. A quality criterion for refinement candidates could, for instance, be to introduce a frequency threshold.

Second, a possible approach is demonstrated to extend commonsense knowledge by human goals (acquired from search query logs). The presented approach starts from ConceptNet triples which contain the MotivatedByGoal relation such as ["`wait tables`", "`MotivatedByGoal`", "`make money`"]. The left concept "`wait tables`" represents a potential action to perform in order to attain the corresponding goal make money in the right concept. For each ConceptNet goal, a list of actions from ConceptNet was extracted. In case of the goal "`make money`", the list of actions includes "`wait tables`", "`go to work`" or "`apply for a job`". To extend the knowledge base with human goals from search query logs, a list of candidate human goals was compiled which were similar to the ConceptNet goal. A simple bag-of-words metric was employed to identify similar goals: If the search query log goal contained all tokens of the ConceptNet goal, it was added to the candidate set of the corresponding ConceptNet goal. To give an example, "`make money quickly`" represent a candidate for the ConceptNet goal "`make money`". This metric considers only refinements as candidates and consequently as potential new ConceptNet entries. The following example should clarify the procedure:

	ConceptNet Action	Relation Type	Similar Goals
ConceptNet Entry:	wait tables	MotivatedByGoal	make money
Potential New ConceptNet Entry:	wait tables	MotivatedByGoal	make money quickly

Table 7.2 shows potential new combinations of human goals from search query logs and corresponding actions that were already contained in ConceptNet. The examples in Table 7.2 illustrate selected combinations that were rated positively by human judges. These combinations could be integrated into ConceptNet by using the MotivatedByGoal relation.

List of ConceptNet Actions (Left Concept)	**Semantic Relation**	**Human Goal from Search Query Logs (Right Concept)**
wait tables, go to work, work the box office, serve customers, tell a story, get a contract, buy a house, apply for a job, pass a course	MotivatedByGoal	make some money quickly
meet interesting people, meet people	MotivatedByGoal	make new friends in your area
surf the net, surf the web, use a computer	MotivatedByGoal	find credit information
eat ice cream	MotivatedByGoal	ways to gain weight
go jogging, eat healthily, release your energy, go for a run, play sports, get exercise, get some physical activity, eat vegetables	MotivatedByGoal	lose maximum weight fast

Table 7.2: This table shows selected combinations of human goals from search query logs and corresponding actions from ConceptNet. These combinations were positively rated by human annotators and could be integrated into ConceptNet using the MotivatedByGoal relation. (taken from [161])

A human subject study was conducted to evaluate the compiled ("search query goal" / "ConceptNet action") pairs which were annotated by two annotators. On the whole, 528 decisions had to be made. The human subjects were given a list of goals and corresponding actions. For every goal/action pair (G/A), they had to answer following question with yes or no: "Do you think that a person's goal could be G when performing action A?" A softer variant of the known precision metric was introduced, i.e. a human goal from search query logs was considered a potential goal if at least one ConceptNet action had been positively annotated. The human subject study achieved an average soft precision of 64%, meaning that 77 out of 120 goals from search query logs were regarded reasonable goals for the given actions. These findings suggest that human goals from search query logs can contribute to complement ConceptNet, a commonsense knowledge base.

7.2 Intentional Query Suggestion

This section experiments with the idea of intentional query suggestion, i.e. suggesting queries to users to make the intent more explicit during search. While traditional query suggestion often aims to make a query resemble more closely the documents a user is expected to retrieve (which might be unknown to the user), an alternative approach is explored: expanding queries to make searchers' goals more explicit.

Figure 7.1: A screenshot of the prototypical implementation which realizes the idea of intentional query suggestion. Triggered by the short query "poker", the suggested queries are offered on the screen's right upper corner.

To give an example: In traditional query suggestion, a query "car" might receive the following suggestions: "car rental", "car insurance", "enterprise car rental", "car games" (actual suggestions produced by Yahoo.com on Nov 27^{th} 2008). Intentional query suggestions include, for instance, "`buy a car`", "`rent a car`", "`sell your car`", "`repair your car`". Figure 7.1 shows a screenshot of the prototypical implementation which realizes the idea of intentional query suggestion. Triggered by the short query "poker", the suggested queries are offered on the screen's right upper corner.

Table 7.3 contrasts traditional and intentional query suggestions to illustrate the differences between both approaches. Intuitively, the better a search engine could understand a person's search goal the better the level of service that can be provided.

The parametric algorithm for intentional query suggestion approximates the person's search goal by combining text-based and neighborhood-based approaches. For the text-based approach, the tokens of input queries are textually compared to all tokens of explicit goal queries. Several text-based similarity measures including Cosine Similarity, Dice Similarity, Jaccard Similarity and Overlap Similarity were taken into account. Because the similarity measures did not exhibit signif-

Initial Query	Semantic Query Suggestion (from Yahoo!)	Semantic Query Suggestion (from MSN)	Intentional Query Suggestion
car	car rental, car insurance, enterprise car rental, car games	used cars, new cars, 2007 new cars, used cars for sale, cars for sale, fast cars, classic cars, car games	buy a car, rent a car, sell your car, repair your car
poker	online poker, poker games, world series of poker, party poker, free poker	free online poker, full tilt poker, free poker games, free poker, poker rules, absolute poker, online poker, poker hands	cheating at poker, learn to play poker, buy poker table, design your own poker chips
house	house plans, white house, house of fraser, columbia house, house of blues, full house	house TV show, houses for sale, houses for rent, house plans, house MD, house fox, haunted houses, Hugh Laurie	insure my house, sell your house, make offer on house, buy house online, build my own house

Table 7.3: Traditional (provided by Yahoo! and MSN) and intentional query suggestions are contrasted to illustrate the idea of making user goals more explicit during search. (taken from [162])

icant differences, it was agreed on using Jaccard Similarity throughout the experiments for reasons of simplicity. For the neighborhood-based approach query logs are conceptualized as consisting of two types of nodes (a bipartite graph), where nodes of one type correspond to explicit intentional queries and nodes of the other type correspond to implicit intentional queries. A bipartite graph is constructed based on session proximity between these two types of nodes. Thereby, neighboring queries are used to further describe and characterize explicit intentional queries, building characteristic term vectors for explicit intentional queries. Figure 7.2 shows an excerpt of the resulting bipartite graph (right side) and corresponding query log entries (left side). Details on constructing the bipartite graph are provided in [162].

Type	Query	Date
$q_{a,1}$	types of diet pills	2006-05-24 13:34:16
$q_{a,2}$	Lipo6	2006-05-24 13:36:24
$q_{a,3}$	lose 20 pounds in 8 weeks	2006-05-24 13:37:23
$q_{e,1}$	lose weight fast	2006-05-24 13:38:42
$q_{a,4}$	lose weight fast	2006-05-24 13:39:06
$q_{a,5}$	weight loss upplements	2006-05-24 13:39:51
$q_{a,6}$	weight loss supplements	2006-05-24 13:39:56

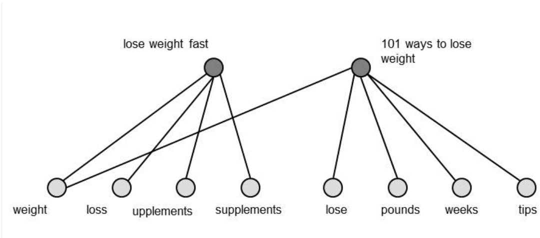


Figure 7.2: Excerpt of the resulting bipartite graph (right side) and corresponding query log entries (left side). (taken from [162])

The graph illustrates relations between explicit intentional queries and meaningful terms in the session neighborhood, representing characteristic term vectors for explicit intentional queries. The example also shows that the neighborhood-based approach is agnostic to misspellings. The bipartite graph is useful in two ways: (i) bottom-up, it can help to produce intentional query suggestions based on cooccurrence (e.g. “upplements” → “lose weight fast”). (ii) top-down, the graph can help to transform explicit intentional queries into implicit ones (which is not further pursued in this example). When input queries are processed by the algorithmic approach, both similarity measures are calculated. In this approach, a linear combination determines the overall similarity between an input query and every explicit intentional query in the dataset yielding a ranked list of potential user intentions.

7.2.1 Quality of Suggested Queries

A human subject study was conducted to learn more about the quality of suggested intentional queries. Annotators were asked to categorize the 10 top-ranked suggested explicit intentional queries for 30 queries into one of the following two relevance classes.

Relevance Classes:

1. Potential User Goal: The suggested query represents a plausible user goal behind a short query.

Initial Query	Top-ranked Intentional Query Suggestions
"anime"	"draw anime", "draw manga"
"playground mat"	"buy playground equipment", "build a swing set"

or the suggested query represents an unlikely yet still related user goal as illustrated by following examples:

Initial Query	Top-ranked Intentional Query Suggestions
"Boston herald"	"getting around Boston", "sightseeing in Boston"
"ginseng coffee"	"moving coffee stains", "fix my keyboard"

2. Clear Misinterpretation: The suggested query has no relation with the initial query. Suggestions that do not conform to this definition are assigned this category as well.

Initial Query	Top-ranked Intentional Query Suggestions
"Boston herald"	"care for Boston fern", "flying to Nantucket"
"playground mat"	"raise money for our playground", "weave a basket fifth grade project"

30 queries of length 1 or 2 were randomly drawn from the MSN search query log. The prospective queries were filtered with regard to (i) reasonableness, i.e. discarding queries such as "wiseco" or "drinkingmate" and to (ii) non American raters, i.e. discarding queries such as "target" or "espn".

In order to evaluate intentional query suggestions that are provided by the algorithmic approach, the percentage of correct suggestions was calculated, i.e. query suggestions that were assigned to relevance class 1. Achieved precision values are illustrated in Table 7.4.

	X	Y	Z
Precision	0.61	0.73	0.8

Table 7.4: Precision values of the algorithm's suggestions as rated by three human annotators (X, Y and Z). (taken from [162])

The average precision amounts to 0.71, i.e. in seven out of ten cases the algorithm returns a potential user intention. In addition, the inter-rater agreement κ [35] between all pairs of human subjects X, Y, and Z was calculated. Cohen's κ measures the average pair-wise agreement corrected for chance agreement when classifying N items into C mutually exclusive categories. Cohen's κ formula reads:

$$\kappa = \frac{P(O)-P(C)}{1-P(C)}$$

where P(O) is the proportion of times that a hypothesis agrees with a standard (or another rater), and P(C) is the proportion of times that a hypothesis and a standard would be expected to agree by chance. The κ value is constrained to the interval [-1,1]. A κ-value of 1 indicates total agreement, 0 indicates agreement by chance and -1 indicates total disagreement. Table 7.5 shows the achieved κ-values in the human subject study.

	X-Y	X-Z	Y-Z
Cohen's Kappa (κ)	0.64	0.51	0.67

Table 7.5: Kappa values amongst three annotators (X, Y and Z) for the two relevance classes. (taken from [162])

The κ-values (see Table 7.5) range from 0.51 to 0.67 (0.61 on average) containing two values above 0.6 indicating substantial agreement.

7.2.2 Implications of Intentional Query Suggestion

The following two sections discuss potential implications of intentional query suggestion for web search:

First, diversity of search results has recently gained importance in web search [38]. For example in informational queries, web search results should not provide monolithic search result sets but rather cover as many different aspects (topics) as possible. This work is thus interested in exploring the influence of explicit intentional queries on the diversity of search result sets. If result sets of explicit intentional queries would be more diverse, intentional query suggestion could help to better focus and guide searchers in exploratory searches.

Second, click through rates have been frequently used as a proxy for measuring relevance in large document collections. This work is thus interested in studying whether explicit intentional queries would yield other/better click-through rates than implicit intentional queries. If explicit intentional queries would yield higher click-through rates, making search goals more explicit would represent an interesting new mechanism to improve search engine performance.

7.2.2.1 Influence on Diversity of Search Results

The diversity within search results is examined by calculating the intersection size between different URL result sets produced by different/same query suggestion mechanisms. Two experiments were conducted, seeking to answer the following questions:

(i) *Intersection between different Query Suggestion Mechanisms:* How many URLs (top level domains only) intersect between URL result sets retrieved by 1) the original queries, 2) the corresponding Yahoo! expanded queries and 3) the corresponding intentional query suggestions?

(ii) *Intersection within same Query Suggestion Mechanisms:* How many URLs (top level domains only) intersect between result sets that were retrieved by the same query suggestion mechanism regarding one original query?

400 queries of length 1 or 2 were randomly drawn from the MSR search query log. Following constraints were made: original queries (i) should yield at least 10 suggestions by the algorithmic approach, (ii) should not contain misspellings and (iii) must not be 'adult' phrases. For each selected query, the top 10 suggestions were produced by using the Yahoo! API and by the intentional query suggestion algorithm. The top 50 result URLs for each suggestion were processed. Searches were conducted by applying the Yahoo! BOSS API. In order to compare the original query results with both expanded results sets, 500 resulting URLs are retrieved for every original query. For each query, it was calculated how many URLs were shared on average between the URL result sets taking into account only unique URLs as well as only top level domains of the resulting set. The intersection metric equals the Jaccard metric that has already been introduced. The averaged results over all candidate queries are shown in Table 7.6.

Compared URL result sets	**Average Intersection**
Original Queries vs. Yahoo! Suggestions	0.191
Original Queries vs. Intentional Suggestions	0.047
Yahoo! Suggestions vs. Intentional Suggestions	0.051

Table 7.6: Average intersection sizes for URL sets of original queries and their corresponding suggestions. (taken from [162])

The results in Table 7.6 imply that original query results share more URLs with results yielded by Yahoo! expanded queries than with results yielded by queries that reflect potential search goals. This suggests that if queries are expanded by a person's search goal more diverse result sets can be achieved. In addition, the inner intersection size of the result sets was calculated, i.e. the overlap between different result sets produced by the same suggestion mechanism. The results were again averaged over all queries and are shown in Table 7.7.

Compared URL result sets	**Average Intersection**
Yahoo! Suggestions	0.103
Intentional Suggestions	0.026

Table 7.7: Average intersection sizes for URL sets of either Yahoo! expanded queries or expanded by the person's search goal. (taken from [162])

The results in Table 7.7 suggest that queries expanded by Yahoo! yield more overlapping URLs than queries expanded by a person's search goal. It is speculated that queries that express a specific intention lead to more diverse results than queries that attempt to approximate the expected document content to retrieve. Considering the presented results, it can be hypothesized that search processes could be made more focused if the person's search goal is included in the search process. It appears that intentional query suggestions diversify search results to a greater extent.

Moreover, these results suggest that intentional query suggestions cover a wider range of topics than suggestions by traditional approaches.

7.2.2.2 Influence on Click-Through

To study the influence of explicit intentional queries on click through, the number click-through events for different token lengths was analyzed. The click-through numbers for different token lengths in the MSR query dataset were obtained and following token length bins were created: one token queries, two token queries, three to four token queries, five token queries, six to ten token queries and queries consisting of more than ten tokens. In addition, the number of click through events was analyzed for explicit intentional queries. For each category, a random sample of 5,000 queries was drawn and the corresponding click-through events were registered and counted. Explicit intentional queries were only considered in the respective category and filtered out otherwise. Table 7.8 shows the number of click through events for each bin (#click-through), and for the set of explicit intentional queries. For example, Table 7.8 shows that for 5,000 random queries of length five, 5,559 click-through events were registered. Note that the click-through data was collapsed for this analysis.

	Implicit Intentional Queries						**Explicit Intent. Queries**
Query Length	*1*	*2*	*3-4*	*5*	*6-10*	*>10*	*5.33*
#click-through	855,649	358,327	64,313	5,559	2,728	960	7,236

Table 7.8: Click-through distribution for different query lengths. An average query length of 5.33 represents the number of click-through for explicit intentional queries. (taken from [162])

It can be observed that explicit intentional queries appear to have a ∼30% higher number of click through events (#click-through = 7,236) than implicit intentional queries of comparable length (length 5, #click-through = 5,559). The higher click-through numbers of explicit intentional queries suggest that such queries retrieve more relevant results, which seems to represent an interesting finding and first evidence for the potential utility of intentional query suggestion.

7.3 iTag: Human Intent Annotation

This application scenario explores the potential of human goal knowledge to contribute to the area of automatic tag generation and thereby to the analysis of textual resources. While existing tagging algorithms largely focus on automatically describing the general topics covered by a resource (such as “career”, “education”), focusing on a different tagging dimension is suggested: automatically annotating resources with human goals.

Annotations have become an increasingly popular means for organizing, categorizing and finding resources on the social web. Yet, only a minor fraction of resources on the web are annotated [66]. This has led to the development of automatic tag generation algorithms aiming to augment and approximate human tagging behavior. Recent attempts include TagAssist, an approach to automatically suggest appropriate topic tags for blog posts (such as “politics” or “news”) [159] or P-Tag, an algorithm to automatically produce personalized tags for web pages [28]. Results reported by these early attempts are encouraging and demonstrate that for selected tagging dimensions useful approximations can be produced. While certain dimensions of tags dominate folksonomies in many applications such as search [16], a particularly interesting yet currently not very well understood dimension of annotations is human intent.

In contrast to topic or quality annotations, intent annotations focus on future states of affairs that some agent wants to achieve, and describe which human goals are relevant in the context of a given textual resource. To give an example: While a particular blog post might focus on the topics “cars” and “automobiles”, the underlying intention of the author might be to “`Achieve mobility`” or to “`Reduce ecological footprint`”. Human goals can be assumed to play a fundamental role in user interactions on the web, including interpreting and understanding resources. Intent annotations could be useful, for example, to quickly grasp the main aspirations implicitly addressed by resources or to enable goal-oriented navigation of resources, such as blogs, on the web (cf. for example, [160]). Figure 7.3 shows an example tag cloud of intent annotations.

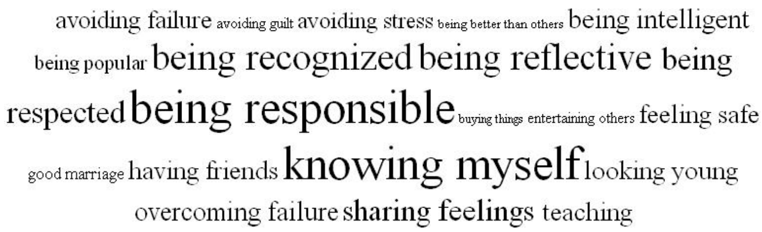


Figure 7.3: An example tag cloud of intent annotations. (taken from [86])

Figure 7.3 aims to illustrate the notion of intent annotations by giving an example of a tag cloud revealing information about goals and intentions referenced in a textual resource. Without knowing the underlying resource, a range of interesting analyses becomes possible. From knowing authors’ goals and interests, one might be able to infer their opinions, their relationship with other people or their attitude towards life. However, existing folksonomy-based systems do not support or encourage users in assigning intent tags and as a result - this type of annotation is hardly used “in the wild”.

Existing automatic tag suggestion approaches largely focus on annotating a document according to its predominant subject matter (what a resource is about, e.g. “sports” or “politics”). This application scenario aims to annotate resources according to the human goals described within them (what goals a resource is about, e.g. “`Achieve Happiness`” or “`Maintain Good Health`”). This type of annotation can be expected to introduce a new and interesting perspective on textual resources. Intent annotations thereby represent an orthogonal view on topic annotations by attempting to answer which human goals are referenced in a given textual resource. Intent annotations thus deal with future states of affairs that some agent wants to achieve (goals), as opposed to topic, sentiment, or opinion tags where typically a present state is approximated. In addition, goals are frequently represented by compound tokens consisting of at least one verb and one or more other tokens ("looking young" as opposed to "youth").

7.3.1 Automatic Intent Annotation with iTAG

There are a number of alternative datasets that could be used as a basis for intent annotation, such as goals acquired from resources themselves, goals acquired from other resources (such as 43things.com, search query logs, etc) or human goals modeled in theoretical frameworks. The iTAG automated annotation approach is based on the latter - an existing socio-psychological taxonomy of 135 categories of human goals [30]. This has two advantages: First, the theoretical framework was compiled by psychologists, and can be considered to be exhaustive to a certain extent by covering a broad range of different aspects of human goals. Second, the limited set of goal categories facilitates evaluation by transforming the large set of potential human goal instances into a manageable number of goal categories. To automatically generate intent annotations, the iTAG approach utilizes the implementation to analyze natural language text from a goal-oriented perspective (see Section 5.3). Textual resources of 44 transcripts of political speeches were retrieved and preprocessed. The speeches were given in April and June 2008 by the two leading American presidential candidates, John McCain and Barack Obama. After data cleansing and sentence delimitation, every sentence was treated as a query for the knowledge base.

iTAG yields a ranked list of intent annotations based on the 25 most dominant goal categories identified for a given textual resource. Figure 7.4 and Figure 7.5 present tag clouds of intent annotations for speeches given by Obama and McCain. The text size of intent tags is based on the weight of annotations assigned to Barack Obama's and John McCain's speeches. Text size in these clouds scales linear; to visualize the clouds existing online services were used. In both cases the top 25 tags were retained. While the tag clouds depicted in Figure 7.4 and Figure 7.5 aggregate intent annotations for a number of speeches given by the candidates, iTAG could be applied on an individual speech and/or passage level as well, assuming the presence of a sufficient number of sentences containing indicative actions.

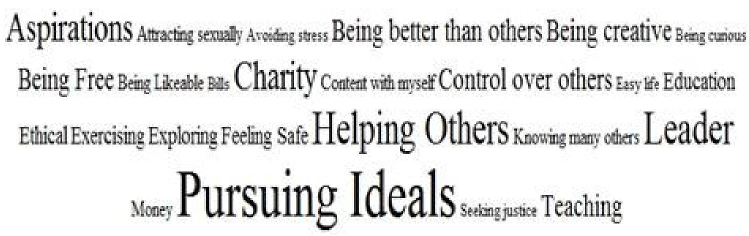


Figure 7.4: A tag cloud of intent annotations for 21 speeches given by Barack Obama. (taken from [86])

The two tag clouds presented in Figure 7.4 and Figure 7.5 reveal further interesting differences between the goals pursued by the two presidential candidates. While McCain's most dominant goal categories are "`Helping Others`" and "`Being better than others`", "`Pursuing ideals`" and "`Helping Others`" represent the highest-weighted annotations for Obama, according to iTAG.

This is an interesting, yet anecdotal, result concurring with a popular media characterization of Obama's political motivations as driven by and aspiring to ideals.

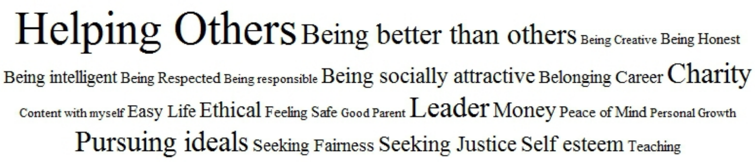


Figure 7.5: A tag cloud of intent annotations for 23 speeches given by John McCain. (taken from [86])

These tag clouds indicate intent annotations' potential as a quick, visual evaluation of text from an intentional perspective.

7.3.2 Intent vs. Content Annotations

To visually illustrate the differences between intent and content annotations, different tag clouds were compared – Intent Tag Clouds vs. Traditional Tag Clouds. Intent tag clouds were produced by iTAG, while "traditional" tag clouds were produced by counting word occurrences in the text, and eliminating words based on a list of stop words[1]. Figure 7.6 illustrates an excerpt of the tag clouds produced. On the right hand side, traditional tag clouds of McCain's and Obama's speeches are presented, while on the right hand side of Figure 7.6, intent tag clouds show a different perspective on the same data. It can be observed that while traditional tag clouds provide a rough overview of the vocabulary used by the two candidates, intent tag clouds highlight the goal categories that are most important to them.

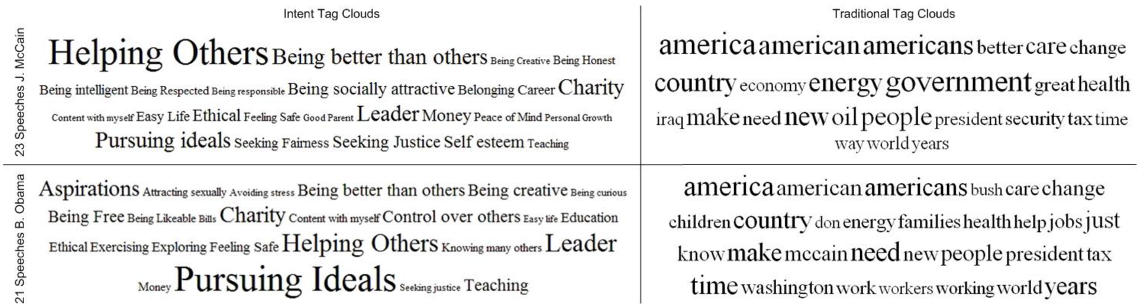


Figure 7.6: Visual comparison between Intent- and Traditional Tag Clouds based on all speeches given by Barack Obama and John McCain. The figure should illustrate that there is no rivalry between intent and traditional tags. They rather complement each other by providing two different perspectives onto political speeches. (taken from [86])

Jeanquartier et al. [75] explored the usability and the acceptance of intent tag clouds to support users in visually analyzing textual content. Tag clouds per se can have positive effects on basic visual tasks, e.g. getting a quick impression of the resource's content [143]. These effects can be ascribed (i) to their compact layout, (ii) to their ability to easily visualize several feature

[1] http://www.dcs.gla.ac.uk/idom/ir_resources/linguistic_utils/stop_words, last accessed February 15th 2009.

dimensions, e.g. size or color of tag cloud elements, or (iii) to their potential to point out important information, e.g. most frequent terms (cf. [64]). Tag clouds are often preferred over other visualization techniques because of their simple usage and applicability, e.g. for prototyping (cf. [79]).

A formative test [4] was conducted to learn more about usability and information quality of intent tag clouds. Four human subjects participated in this study assessing traditional and intent tag clouds. The generation of tag clouds was based on Martin Luther King's famous speech "I have a dream". After completing the visual comparison, human subjects were asked to fill out a questionnaire to convey their impressions. The questionnaire contained a number of questions to cover different aspects such as readability, text understanding, acceptance or (perceived) usefulness. For more a more detailed description of the formative test and the respective questionnaire, it is refered to [75]. The results indicate a positive attitude towards (intent) tag clouds as assistance tool. The participants believed that intent tag clouds provided a novel perspective on textual resources, complementing the traditional tag clouds and thus contributing to a better understanding of the textual contents. This was reflected in their answers that they thought traditional tag clouds were more suitable to give a quick topic overview whereas intent tag clouds facilitated recognizing human goals in textual content.

In summary, intent annotations add interesting information to textual resources, which is difficult to extract from the resource itself. In the past, automatic tag generation approaches demonstrated their usefulness in a broad range of different applications, including tag suggestion, resource clustering, resource enrichment or tag-based navigation. Thus, this application scenario adds a novel dimension to the set of tag dimensions identified in the literature. Moreover, intent tag clouds appear to be useful to complement visual analytics tools.

PART IV: CONCLUSION

Chapter 8

Discussion

This work offers novel perspectives on modeling, acquiring and applying knowledge about human goals. A main motivation of this work is the desire to better understand how goal knowledge can be modeled semantically to make it operable, utilizable and applicable in interactive systems. The value of goal knowledge has been already acknowledged and demonstrated across many research domains including Requirements Engineering, Intelligent Agent Technology or Commonsense Knowledge Acquisition. However, what has been missing so far is a common view on engineering human goal knowledge across research domains.

This work is meant as a first step towards unifying existing research on modeling human goal knowledge by proposing a general framework which is based on and synthesizes existing work from other fields. The framework serves as common denominator between research domains which in further consequence allows research domains to relate their work to each other and thus to benefit from each other's experiences and experimental results. Moreover, it allows researchers to compare, analyze and evaluate human goal knowledge across domains, e.g. to identify synergetic potentials. By operationalizing knowledge about human goals in three case studies, this work advances the explication and thus the understanding of human goal knowledge in natural language text and thus serves as a starting point for knowledge engineers who are interested in integrating human goal knowedge into their systems.

This work demonstrates the value and the potential of applying human goal knowledge in various problem settings and application scenarios. Prototypical implementations will prove helpful for (i) conducting goal-oriented analyses and for (ii) building and designing intelligent systems. Preliminary experiments indicate that using human goal knowledge is valuable and inspires further research opportunities in following research domains including

- Commonsense Knowledge Acquisition by complementing existing commonsense knowledge bases with human goal knowledge.

- Information Retrieval by offering alternative ways to suggest queries, i.e. making users' search goals more explicit during the search process.

- Text Analysis by analysing natural language text based on referenced human goals in a given textual resource. This form of analysis deals with a different temporal focus than orthogonal analysis techniques, e.g. sentiment analysis where a present, emotional state is approximated. This analysis therefore represents a novel addition to the repertoire of textual data analysis techniques and contributes to expanding the knowledge that can be inferred from textual resources.

- Visual Analytics by visually comparing textual resources from a goal-oriented perspective instead of, for instance, a topical one.

8.1 Limitations

The author is aware that this PhD work can only be a first step towards a unifying view on engineering knowledge about human goals. This work is limited in theoretical and practical aspects which may inform future research efforts. Selected limitations include

- Large parts of this work are theoretical in nature, i.e. informed by experiences drawn from case studies, discussions with other researchers and reviews of related literature. While generality and practicability of the proposed framework are corroborated by three case studies and application scenarios, an evaluation with respect to completeness is beyond the scope of this thesis.

- While this work outlines instructions to acquire large amounts of human goal knowledge, it does not construct broad knowledge bases. Large-scale analyses of human goal knowledge have thus not been conducted.

- This work only discusses selected operations on human goal knowledge such as reasoning or planning. The implementation of these operations is left to future work.

- The practical part of this work instantiates and applies a subset of data model components including the *GoalCategory* class, the *Goal* class, the *Task* class, the *HasGoalCategoryRelation* class and the *MeansForEndRelation* class. Other components need to be instantiated in future work to construct broad knowledge bases of human goal knowledge. Section 8.2 outlines possible approaches to further instantiate the data model by relating to existing work from other areas.

8.2 Instantiation of Data Model Components

This section outlines possible starting points to automate the acquisition and the instantiation of selected data model components which have not been covered in this work.

<u>Goals and Tasks:</u> The case studies presented methods to extract, to structure and to categorize goal instances. Extracting information about a goal's attributes ressembles – to a certain extent – a template-filling task. To approach this task, it is recommended to examine a goal's contextual

environment [62] to identify characteristic features for extracting attribute information. To give an example, a statistical analysis of temporal keywords might be beneficial for the *Goal.Horizon* attribute.

Agents and Resources: To automatically identify the intentional actor, semantic role labelling (cf. [53]) appears to be a good starting point. The objective of semantic role labelling is to assign roles to sentence parts to identify the grammatical agent or the patient. The agent, from Latin "agens" the one doing, exerts control over the verb whereas the patient is being acted upon. Thus, whence the goal instance has been recognized, identifying the grammatical agent might be a candidate for the intentional actor, and the grammatical patient, a candidate for a goal's resource. The *HasGoalRelation.Priority* attribute describes whether the accomplishment of a goal is of (a) high or (b) low priority. An agent might want to quickly accomplish the goal "`find job`", i.e. assigning it a high-priority. Potential approach: (i) gather a lot of data containing the respective goal instance, (ii) examine the contextual environment and (iii) scan for temporal indicators such as "as soon as possible", "asap" or "quickly". Other possible features include, e.g. the writing style to infer whether the intentional actor is tense or under pressure.

Intentional Relations: Automatically extracting semantic relations has been gaining track in recent years (cf. [129], [147] or [11]). Although intentional relations are not explicitly addressed, these approaches appear to be a good starting point for devising algorithms. In Pantel and Pennacchiotti [129] for instance, the authors introduce *Espresso*, a weakly-supervised, general-purpose algorithm for automatically extracting semantic relations from textual resources. The algorithm's advantage is that it requires only few seed examples per relation type. In Pantel and Pennacchiotti [130], their work is continued by studying techniques to ontologize semantic relations, i.e. to automatically link them into existing semantic repositories. The *ConflictRelation* class represents an exceptional case since it connects two conflicting human goals with each other, i.e. goals which exert antagonistic effects on each other. Analyzing the contextual environment might not be successful in this case, simply because of rare co-occurrences. This task might be best approached by utilizing external resources such as WordNet [47] or VerbOcean [29] which allow access to semantic or commonsense knowledge. WordNet, for instance, can be used to connect related noun parts such as "vehicle" and "car". VerbOcean on the other hand contains antonymy relations, i.e. it explicates contrasting meanings between verb pairs such as "lose" and "gain".

Chapter 9

Conclusions

This PhD thesis is motivated by a vision of machines that are capable of behaving intelligently. Intelligent behavior requires that machines understand humans and their goals. The encoding of human goal knowledge and the construction of broad knowledge bases thus represent important steps along the way to realize this vision. This work's core contributions add to these efforts in the following way.

- A general framework has been introduced which (i) synthesizes and (ii) makes a significant step towards unifying existing research on engineering human goal knowledge. The generality of the framework has been partly validated by instantiating and applying its data model's components in a series of theoretical and practical problem settings.
- Three case studies have explored mechanisms to automatically acquire human goal knowledge from textual resources thus instantiating selected parts of the data model components. The framework's applicability and practicability have been demonstrated in three different domains.

Additional contributions of this work encompass the following aspects.

- ◇ Advancing the explication and thus the understanding of human goal knowledge.
- ◇ The adaption of existing techniques to further automate the acquisition of goal knowledge in three case studies and several application scenarios.
- ◇ The examination of different types of textual resources with respect to their value for acquiring human goal knowledge.
- ◇ The codification of the framework's data model in the Web Ontology Language (OWL) to be useable and accessible for semantic web technologies.

This PhD thesis offers a novel perspective on knowledge about human goals and demonstrates its value and potential for advancing the understanding and the interaction between humans and machines. This PhD thesis' theoretical and practical contributions are relevant for knowledge engineers interested in modeling and in acquiring human goal knowledge from textual resources and will hopefully enable them to build systems that exhibit higher levels of awareness and representation of human goal knowledge.

9.1 Open Challenges & Future Directions

This work is considered as a first step towards unifying human goal knowledge across research domains. To step further into this direction, human goal knowledge needs to be integrated into real-world applications by building intelligent systems capable of understanding goal knowledge. At the same time, a number of future research directions open up such as (i) integrating ideas from the field of Open Information Extraction into the acquisition process to generate broad goal knowledge bases or (ii) enhancing the interaction between humans and machines by equipping intelligent systems with goal knowledge.

Construction of Comprehensive Human Goal Knowledge Bases

To be usable in real-world applications, large amounts of human goal knowledge have to be acquired. While this work's case studies provide techniques to acquire knowledge about human goals, they represent separate attempts to instantiate various parts of the framework's data model. Since human goal knowledge covers a wide spectrum of domains, the construction of comprehensive goal knowledge bases – probably in an open-ended, continuous process – is valuable for the web community. Human goal knowledge would then complement other types of knowledge such as commonsense knowledge (cf. Cyc [92], ConceptNet [104], KNEXT [152] or MindNet [141]). Creating a link between different types of knowledge (and knowledge bases) could also pave the way for broader and richer knowledge about human goals.

Enhance Human-Computer-Interaction

To provide machines with means to better understand and even predict human behavior, user interfaces should be equipped with knowledge about human goals so that they can provide better support and guidance. Existing user interfaces benefitting from some form of human goal knowledge include goal-oriented search [102], intentional query suggestion (cf. [111], [162]), goal-oriented web browsing [46], intelligent ambient assistants [131], semantic search assistants [93] or semantic task retrieval [120]. This PhD thesis lays the groundwork for creating comprehensive goal knowledge bases accessible by intelligent user interfaces to better recognize users' goals based upon their actions.

Ontology-Based Real-World Applications

The OWL encoding of the data model provides a formal specification which can contribute to a unifying way of describing human goal knowledge across the Semantic Web. A unified view enables knowledge sharing and comparing as well as applications of web-wide retrieval or large-scale statistics. At the moment it is difficult to find aspects of human goal knowledge using contemporary search engines since simple text based indexing is less accurate and less flexible than what would be achieved with an ontology-based formalization process. Deploying the goal ontology in real-world applications also supports reasoning over human goal knowledge as is, for instance, necessary to answer (i) standard data retrieval queries as well as (ii) conceptual queries about the structure of goal knowledge. Reasoning functionality allows statements on correctness and reliability of an ontology-based system. These are important quality indicators, e.g. for safety-critical systems such as medicine or traffic control. Implementing goal ontology-based applications thus contributes to the growth of the Semantic Web and to the extended usage of human goal knowledge on the Semantic Web.

Complement Reasoners and Planners with Human Goal Knowledge

Reasoning or inference describes the process of reaching a conclusion which is based on already existing data. The proposed framework provides a rich characterization of human goal knowledge including the specialization into hard and soft goals. Soft goals, for instance, already inform reasoning procedures by serving as preferences in goal-oriented requirements engineering (cf. [118]). Other components such as the *Resource.Availability* attribute or the *HasGoalRelation.Priority* attribute appear equally well suited to inform (i) top-down or (ii) bottom-up reasoning procedures.

Planning describes the process of generating action sequences to reach a predefined goal state. Planning processes might benefit from additional information about the current goal, e.g. whether critical resources are available or not or whether required goals or subgoals are in conflict with each other. The framework outlines a rich model on human goal knowledge whose components can be used to inform the planning process. The *Goal.Horizon* attribute, for instance, informs about how long a goal's accomplishment is likely to take. Future work could aim for developing prototypical implementations to integrate human goal knowledge into reasoning and planning processes.

Appendix A

Research Methodology

This thesis uses methods from case study research [185] and related literature to inform the development of (i) a data model which organizes required components and relations and (ii) a meta-process which describes necessary steps to make human goal knowledge operable and thus useful. Although the structure of this thesis is linear, it was an iterative process of conducting case studies and of revising the framework to take newly gained insights and experiences into account.

Case study research is not only adequate for exploratory phases of an investigation but also for explanatory and descriptive phases. Yin defines a case study as "*. . . empirical inquiry that investigates a contemporary phenomenon within its real life context, especially when the boundaries between context and phenomenon are not clearly evident*". Case studies therefore represent a suitable research strategy for understanding complex phenomena which cannot be easily studied separated from their context without losing information; they seek to gain an integral understanding of a phenomenon. In summary, case study research can be used as research strategy given following four characteristics: (i) "how" or "why" questions are being asked, (ii) the phenomenon (set of events) is not separable from the context (iii) the phenomenon is contemporary (iv) the researcher has no control over the phenomenon (in contrast to experiments). Since these characteristics comply with studying theoretical and practical aspects of engineering knowledge about human goals, this thesis demonstrates the proposed framework's value and potential by means of case studies. Moreover, to make robust statements about "engineering human goal knowledge", this thesis' central research design is based on a multiple case study approach.

In this thesis, three case studies are conducted as a preliminary evaluation approach to illustrate and validate the introduced framework:

1. The first case study addresses the acquisition of goal knowledge from search query logs. It explores the agreeability of the definition of a human goal and the feasibility to automatically extract human goal instances from search query logs. This case study's results inform us about (i) characteristics of goal instances, e.g. which goal types or attributes appear

reasonable in a data model on human goal knowledge and (ii) their statistical distribution.

2. The second case study addresses the annotation of political speeches with goal categories. This case study is explorative as well as descriptive. It explores the feasibility of categorizing textual resources into a human goal taxonomy. Moreover, the case study fosters our understanding of human goal knowledge in natural language text, e.g. how people express their goals. It has thus a descriptive character as well. Based on this understanding algorithmic approaches are devised to analyze textual resources from a goal-oriented perspective.

3. The third case study addresses the inference and extraction of intentional relations between intentional concepts such as goals, subgoals and tasks. For that purpose, the case study explores methods (i) to generate goal concept hierarchies, and (ii) to complement these hierarchies with reasonable actions (tasks) which potentially contribute to their accomplishment.

To ensure the results are of scientific value, this work considers Yin's threats to validity. (i) Construct Validity: To validate this thesis' definitions and results, human subject studies have been conducted which all yielded reasonable scores for inter-rater agreements [35] and reasonable distributions of human annotators' judgements. (ii) External Validity: All case studies refer and relate to existing work from literature. Experiments are evaluated by using standardized measures in the community such as precision, recall or taxonomic overlap. (iii) Reliability: All case studies use existing toolkits, e.g. for natural language processing (NLTK[1]) or machine learning (WEKA) [182], so that reproducing experimental results is possible.

[1] http://www.nltk.org/

Appendix B

List of Abbreviations

AOL	American Online
API	Application Programming Interface
BDI	Belief, Desire, Intention
BOSS	Build your Own Search Service
BSKM	BiSection K-Means
CSKA	Commonsense Knowledge Acquisition
FCA	Formal Concept Analysis
GOOSE	GOal-Oriented Search Engine
HAC	Hierarchical Agglomerative Clustering
HLT	Human Language Technology
HTN	Hierarchical Task Networks
IAT	Intelligent Agent Technology
IE	Information Extraction
IR	Information Retrieval
i*	iStar
κ	Kappa
KNEXT	Knowledge Extraction from Text
ML	Machine Learning
MS	Microsoft
NB	Naïve Bayes
NLP	Natural Language Processing
OMCS	Open Mind Common Sense
OWL	Web Ontology Language
POS	Part-Of-Speech
RE	Requirements Engineering
SC	Situation Calculus
SoW	Set of Words
SQL	Search Query Log
SVM	Support Vector Machine
TO	Taxonomic Overlap
URL	Unified Resource Locator
US	United States

Bibliography

[1] Abdullah, M., Paige, R., Benest, I., and Kimble, C. (2005). Unified modeling language for knowledge modelling. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet*.

[2] Agichtein, E. and Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In *Proceedings of the 5th ACM Conference on Digital Libraries (DL'00)*.

[3] Anderson, C. (2004). The long tail. *Wired Magazine*, 12(10).

[4] Andrews, K. (2006). Evaluating information visualisations. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors (BELIV'06)*.

[5] Anton, A. (1997). *Goal identification and refinement in the specification of software-based information systems*. Doctoral Dissertation, Georgia Institute of Technology, Atlanta GA.

[6] Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P. (2003). *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.

[7] Baeza-Yates, R., Calderon-Benavides, L., and Gonzalez-Caro, C. (2006). The intention behind web queries. In *String Processing and Information Retrieval (SPIRE'06)*.

[8] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley.

[9] Banko, M., Cafarella, M., Soderland, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI'07)*.

[10] Banko, M. and Etzioni, O. (2007). Strategies for lifelong knowledge extraction from the web. In *Proceedings of the 4th International Conference on Knowledge Capture (K-CAP'07)*.

[11] Banko, M. and Etzioni, O. (2008). The tradeoffs between open and traditional relation extraction. In *Proceedings of the 46th Annual Meeting on Association for Computational Linguistics (ACL'08)*, pages 28–36, Columbus, Ohio. Association for Computational Linguistics.

[12] Bannon, L., Cypher, A., Greenspan, S., and Monty, M. (1983). Evaluation and analysis of users' activity organization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'83)*.

[13] Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Penguin Group.

[14] Bauer, M. (1995). A dempster-shafer approach to modeling agent preferences for plan recognition. *User Modeling and User-Adapted Interaction*, 5:317–348.

[15] Becerra-Fernandez, I. (2004). *Knowledge Management: Challenges, Solutions, and Technologies*. Prentice Hall.

[16] Bischoff, K., Firan, C., Nejdl, W., and Paiu, R. (2008). Can all tags be used for search? In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08)*.

[17] Booch, G., Rumbaugh, J., and Jacobson, I. (1999). *The Unified Modelling Language User Guide*.

[18] Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.

[19] Braubach, L., Pokahr, A., Moldt, D., and W., L. (2004). Goal representation for BDI agent systems. In *PROMAS*.

[20] Brin, S. (1999). Extracting patterns and relations from the world wide web. In *Selected Papers from the International Workshop on World Wide Web and Databases (WebDB'99)*.

[21] Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2).

[22] Cankaya, H. and Moldovan, D. (2009). Method for extracting commonsense knowledge. In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09)*.

[23] Carberry, S. (1990). Incorporating default inferences into plan recognition. In *Proceedings of the 8th National Conference on Artificial Intelligence - Volume 1 (AAAI'90)*.

[24] Carberry, S. (2001). Techniques for plan recognition. *Journal of User Modeling and User-Adapted Interaction*, 11.

[25] Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr., E., and Mitchell, T. (2010). Toward an architecture for never-ending language learning. In *Proceedings of the 27th National Conference on Artificial Intelligence (AAAI'10)*.

[26] Charniak, E. and Goldman, R. (1993). A bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79.

[27] Chen-Burger, J. (2001). *Knowledge sharing and inconsistency checking on multiple enterprise models*. AIAI, University of Edinburgh, Edinburgh.

[28] Chirita, P., Costache, S., Nejdl, W., and Handschuh, S. (2007). P-Tag: Large scale automatic generation of personalized annotation tags for the web. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*.

[29] Chklovski, T. and Pantel, P. (2004). VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*.

[30] Chulef, A., Read, S., and Walsh, D. (2001). A hierarchical taxonomy of human goals. *Motivation and Emotion*, 25(3).

[31] Cimiano, P. (2006). *Ontology learning and population from text*.

[32] Cimiano, P., Hotho, A., and Staab, S. (2005). Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 24.

[33] Cimiano, P. and Staab, S. (2004). Learning by googling. In *SIGKDD Explorations Newsletter*.

[34] Clark, P. and Harrison, P. (2009). Large-scale extraction and use of knowledge from text. In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09)*.

[35] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:1–37.

[36] Cohen, P. and Levesque, H. (1990). Intention is choice with commitment. *Artificial Intelligence*, 42.

[37] Collins, A. and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82:407–428.

[38] Crabtree, D., Andreae, P., and Gao, X. (2007). Exploiting underrepresented query aspects for automatic query expansion. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*.

[39] Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., and Slattery, S. (2000). Learning to construct knowledge bases from the world wide web. *Artificial Intelligence*, 118(1-2):69–113.

[40] Dai, H., Zhao, L., Nie, Z., Wen, J., Wang, L., and Li, Y. (2006). Detecting online commercial intention. In *Proceedings of the 15th International Conference on World Wide Web (WWW'06)*.

[41] Dardenne, A., van Lamsweerde, A., and Fickas, S. (1993). Goal-directed requirements acquisition. *Science of Computer Programming*, 20.

[42] Dumais, S., Platt, J., Heckerman, D., and Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM'98)*.

[43] Eslick, I. (2006). *Searching for commonsense*. Master's thesis at the Massachusetts Institute of Technology.

[44] Etzioni, O., Banko, M., Soderland, S., and Weld, D. (2008). Open information extraction from the web. *Communications of the ACM*, 51(12):68–74.

[45] Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D., and Yates, A. (2004). Web-scale information extraction in knowitall: (preliminary results). In *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*, pages 100–110.

[46] Faaborg, A. and Lieberman, H. (2006). A goal-oriented web browser. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'06)*.

[47] Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

[48] Fikes, R. and Nilsson, N. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208.

[49] Finn, A. and Kushmerick, N. (2003). Learning to classify documents according to genre. In *Workshop on Computational Approaches to Style Analysis and Synthesis (IJCAI'03)*.

[50] Friedman, J. (1997). On bias, variance, 0/1 loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*.

[51] Fukazawa, Y. and Ota, J. (2010). Automatic modeling of user's real world activities from the web for semantic IR. In *Proceedings of the 3rd International Semantic Search Workshop*.

[52] Gil, Y. (2005). Description logics and planning. *AI Magazine*, 26.

[53] Gildea, D. and Jurafsky, D. (2002). Automatic labelling of semantic roles. *Computational Linguistics*, 23:245–288.

[54] Girju, R. (2010). Toward social causality: An analysis of interpersonal relationships in online blogs and forums. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10)*.

[55] Giunchiglia, F., Mylopoulos, J., and Perini, A. (2002). The tropos software development methodology: processes, models and diagrams. In *Proceedings of the 3rd International Conference on Agent-Oriented Software Engineering (AOSE'02)*.

[56] Gordon, A. and Hobbs, J. (2004). Formalizations of commonsense psychology. *AI Magazine*, 25.

[57] Gordon, J., Van Durme, B., and Schubert, L. (2010). Learning from the web: extracting general world knowledge from noisy text. In *Workshops at the 24th AAAI Conference on Artificial Intelligence*.

[58] Guo, Q. and Agichtein, E. (2010). Ready to buy or just browsing? Detecting web searcher goals from interaction data. In *Proceedings of the 33rd International ACM Conference on Research and Development in Information Retrieval (SIGIR'10)*.

[59] Harris, Z. (1968). *Mathematical Structures of Language*. Wiley, New York.

[60] Havasi, C. (2007). ConceptNet 3: A flexible, multilingual semantic network for common sense knowledge. In *Proceedings of the 22nd Conference on Artificial Intelligence*.

[61] Havasi, C., Speer, R., and Alonso, J. (2007). ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In *Advances in Natural Language Processing*.

[62] Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics*.

[63] Hearst, M. (1998). Automated discovery of WordNet relations. In Fellbaum, C., editor, *WordNet: An Electronic Lexical Database and Some of its Applications*. MIT Press.

[64] Hearst, M. and Rosner, D. (2008). Tag clouds: Data analysis tool or social signaller? In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS'08)*.

[65] Hendler, J. (2007). Where are all the intelligent agents? *IEEE Intelligent Systems*, 22:2–3.

[66] Heymann, P., Koutrika, G., and Garcia-Molina, H. (2008). Can social bookmarking improve web search? In *Proceedings of the International Conference on Web Search and Web Data Mining (WSDM'08)*.

[67] Hindle, D. (1990). Noun classification from predicate-argument structures. In *Proceedings of the 28th Annual Meeting on Association for Computational Linguistics (ACL'90)*.

[68] Hobbs, J. and Gordon, A. (2010). Goals in a formal theory of commonsense psychology. In *Proceedings of the 2010 Conference on Formal Ontology in Information Systems (FOIS'10)*. IOS Press.

[69] Hu, J., Wang, G., Lochovsky, F., Sun, J., and Chen, Z. (2009). Understanding user's query intent with wikipedia. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*.

[70] Huber, M., Durfee, E., and Wellman, M. (1994). The automated mapping of plans for plan recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1460–1460.

[71] Hui, B. and Yu, E. (2005). Extracting conceptual relationships from specialized documents. *Data and Knowledge Engineering*, 54(1):29–55.

[72] ITU (2001). - Telecommunication standardization sector draft specification of the goal-oriented requirement language (z.151).

[73] Jansen, B., Booth, D., and Spink, A. (2008). Determining the informational, navigational, and transactional intent of web queries. *Information Processing and Management*, 44(3):1251–1266.

[74] Jansen, B., Spink, A., and Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36:207–227.

[75] Jeanquartier, F., Kröll, M., and Strohmaier, M. (2009). Intent tag clouds: An intentional approach to visual text analysis. In *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe'09)*.

[76] Jones, R. and Diaz, F. (2007). Temporal profiles of queries. *ACM Transactions on Information Systems*, 25.

[77] Jung, Y., Ryu, J., Kim, K., and Myaeng, S. (2010). Automatic construction of a large-scale situation ontology by mining how-to instructions from the web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8:110–124.

[78] Kang, I. and Kim, G. (2003). Query type classification for web document retrieval. In *Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR'03)*.

[79] Kaser, O. and Lemire, D. (2007). *Tag-cloud drawing: Algorithms for cloud visualization*.

[80] Kautz, H. (1991). A formal theory of plan recognition and its implementation. In Brachman, R. J., Allen, J. F., Kautz, H. A., Pelavin, R. N., and Tenenberg, J. D., editors, *Reasoning about plans*, pages 69–124. Morgan Kaufmann Publishers Inc.

[81] Kavakli, E. and Loucopoulos, P. (2004). Goal modeling in requirements engineering: Analysis and critique. In John Krogstie, T. H. and Siau, K., editors, *Information Modeling Methods and Methodologies (Adv. Topics of Database Research)*, pages 102 –124. IDEA Group.

[82] Keller, S., Kahn, L., and Panara, R. (1990). Specifying software quality requirements with metrics. *System and Software Requirements Enginering*, pages 145–163.

[83] Kim, J. and Moldovan, D. (1995). Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transactions on Knowledge and Data Engineering*, 7(5):713–724.

[84] Kirsh, D. (1990). When is information explicitly represented? *Information, Language and Cognition - The Vancouver Studies in Cognitive Science*, pages 340–365.

[85] Kröll, M., Fukazawa, Y., Ota, J., and Strohmaier, M. (2011). Automatically constructing concept hierarchies of health-related human goals. *The 5th International Conference on Knowledge Science, Engineering and Management (KSEM'11)*.

[86] Kröll, M., Körner, C., and Strohmaier, M. (2010). iTag: Automatically annotating textual resources with human intentions. *Emerging Technologies in Web Intelligence*, 2(4):333–342.

[87] Kröll, M. and Strohmaier, M. (2009a). Analyzing human intentions in natural language text. In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09)*.

[88] Kröll, M. and Strohmaier, M. (2009b). Extracting human goals from weblogs. In *Workshop "Knowledge Discovery, Data Mining and Machine Learning 2009" in conjunction with LWA'09*.

[89] Law, E., Mityagin, A., and Chickering, M. (2009). Intentions: A game for classifying search query intent. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI'09)*.

[90] Lee, J. (1991). Extending the potts and bruns model for recording design rationale. In *Proceedings of the 13th International Conference on Software Engineering (ICSE'91)*.

[91] Lee, U., Liu, Z., and Cho, J. (2005). Automatic identification of user goals in web search. In *Proceedings of the 14th International Conference on World Wide Web (WWW'05)*.

[92] Lenat, D. (1995). CyC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

[93] Lenat, D., Witbrock, M., Baxter, D., Blackstone, D., Deaton, C., Schneider, D., Scott, J., and Shepard, B. (2010). Harnessing CyC to answer clinical researchers' ad hoc queries. *AI Magazine*, 31:13–32.

[94] Lenhart, A. and Fox, S. (2006). Bloggers: A portrait of the internet's new storytellers. In *Pew Interent American Life Project*.

[95] Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

[96] Li, X., Wang, Y., and Acero, A. (2008). Learning query intent from regularized click graphs. In *Proceedings of the 31st International ACM Conference on Research and Development in Information Retrieval (SIGIR'08)*.

[97] Li, Y., Zheng, Z., and Dai, H. (2005). KDD Cup-2005 report: Facing a great challenge. *ACM SIGKDD Explorations Newsletter*, 7(2):91–99.

[98] Liaskos, S., Lapouchnian, A., Yu, Y., Yu, E., and Mylopoulos, J. (2006). On goal-based variability acquisition and analysis. In *Proceedings of the 14th IEEE International Requirements Engineering Conference (RE'06)*.

[99] Lieberman, H. (2008). Usable AI requires commonsense knowledge. In *Workshops and Courses: Usable Artificial Intelligence held in conjunction with CHI'08*.

[100] Lieberman, H. and Espinosa, J. (2007). A goal-oriented interface to consumer electronics using planning and commonsense reasoning. *Knowledge-Based Systems*, 20.

[101] Lieberman, H., Smith, D., and Teeters, A. (2007). Common consensus: A web-based game for collecting commonsense goals. In *Proceedings of the Workshop on Common Sense and Intelligent User Interfaces held in conjunction with IUI'07*.

[102] Liu, H., Lieberman, H., and Selker, T. (2002). GOOSE: A goal-oriented search engine with commonsense. In *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*.

[103] Liu, H. and Singh, P. (2004a). Commonsense reasoning in and over natural language. In *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES'04)*.

[104] Liu, H. and Singh, P. (2004b). ConceptNet - A practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226.

[105] Maedche, A. and Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16.

[106] Maedche, A. and Staab, S. (2002). Measuring similarity between ontologies. In *Proceedings of the European Conference on Knowledge Engineering and Management*.

[107] Maslow, A. (1943). A theory of human motivation. *Psychological Review*, 50:370–396.

[108] McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502.

[109] Mei, Q., Zhou, D., and Church, K. (2008). Query suggestion using hitting time. In *Proceedings of the 17th International Conference on Information and Knowledge Management (CIKM'08)*.

[110] Minsky, M. (2000). Commonsense-based interfaces. *Communications of the ACM*, 43(8).

[111] Mostert, J. and Hollink, V. (2008). Effects of goal-oriented search suggestions. In *Proceedings of the 20th Belgian-Netherlands Conference on Artificial Intelligence*.

[112] Mostow, J. (1985). Towards better models of the design process. *AI Magazine*, 6.

[113] Mueller, E. (1998). *Natural language processing with ThoughtTreasure*. Signiform.

[114] Mueller, E. (2003). Thoughttreasure: A natural language commonsense platform. Online: http://xenia.media.mit.edu/ mueller/papers/tt.html.

[115] Mueller, R. (2009). Automating commonsense reasoning using the event calculus. *Communications of the ACM*, 52:113–117.

[116] Mulkar-Mehta, R., Hobbs, J., Liu, C., and Zhou, X. (2009). Discovering causal and temporal relations in biomedical texts. In *AAAI Spring Symposium*. AAAI Press.

[117] Mylopoulos, J., Borgida, A., Jarke, M., and Koubarakis, M. (1990). Telos: Representing knowledge about information systems. *ACM Transactions of Information Systems*, 8:325–362.

[118] Mylopoulos, J., Chung, L., and Nixon, B. (1992). Representing and using nonfunctional requirements: A process-oriented approach. *IEEE Transactions on Software Engineering*, 18:483–497.

[119] Mylopoulos, J., Chung, L., and Yu, E. (1999). From object-oriented to goal-oriented requirements analysis. *Communications of the ACM*, 42(1):31–37.

[120] Naganuma, T. and Kurakake, S. (2005). Task knowledge-based retrieval for service relevant to mobile user's activity. In *Proceedings of the International Semantic Web Conference*.

[121] Nau, D., Au, T., Ilghami, O., Kuter, U., Murdock, W., Wu, D., and Yaman, F. (2003). Shop2: An HTN planning system. *JAIR*, 20:379–404.

[122] Nejati, N., Könik, T., and Kuter, U. (2009). A goal- and dependency-directed algorithm for learning hierarchical task networks. In *Proceedings of the 5h International Conference on Knowledge capture (K-CAP'09)*.

[123] Nilsson, N. (1971). *Problem-solving methods in artificial intelligence*. New York: McGraw-Hill.

[124] Niu, N. and Easterbrook, S. (2008). Extracting and modeling product line functional requirements. In *Proceedings of the 16th IEEE International Requirements Engineering Conference (RE'08)*.

[125] Norvig, P. and Russell, S. (1995). *Artificial Intelligence: A Modern Approach*. Prentice-Hall, NJ, USA.

[126] Padgham, L. and Winikoff, M. (2002). Prometheus: A methodology for developing intelligent agents. In *Proceedings of AOSE*.

[127] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2.

[128] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*.

[129] Pantel, P. and Pennacchiotti, M. (2006). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (ACL'06)*, pages 113–120.

[130] Pantel, P. and Pennacchiotti, M. (2008). Automatically harvesting and ontologizing semantic relations. In *Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*.

[131] Panton, K., Matuszek, C., Lenat, D., Schneider, D., Witbrock, M., Siegel, N., and Shepard, B. (2006). Common sense reasoning from CyC to intelligent assistant. *Ambient Intelligence in Everyday Life*, pages 1–31.

[132] Pass, G., Chowdhury, A., and Torgeson, C. (2006). A picture of search. In *Proceedings of the 1st International Conference on Scalable Information Systems*.

[133] Passonneau, R. and Litman, D. (1993). Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics (ACL'93)*.

[134] Piao, S., Bental, D., Whittle, J., Aylett, R., Makri, S., and Sun, X. (2012). A pilot study: Deriving a users' goal framework from a corpus of interviews and diaries. In *Workshop on Semantic Relations II. Enhancing Resources and Applications*.

[135] Porter, M. (1997). An algorithm for suffix stripping. In *Readings in Information Retrieval*.

[136] Potts, C., Takahashi, K., and Anton, A. (1994). Inquiry-based requirements analysis. *Software, IEEE*, 11(2):21 –32.

[137] Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York, Longman.

[138] Regev, G. and Wegmann, A. (2005). Where do goals come from: The underlying principles of goal-oriented requirements engineering. In *Proceedings of the 13th International Conference on Requirements Engineering (RE'05)*.

[139] Reiss, S. (1994). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, 8(3):179 – 193.

[140] Rich, E. and Knight, K. (1990). *Artificial intelligence*. McGraw-Hill.

[141] Richardson, S., Dolan, W., and Vanderwende, L. (1998). Mindnet: acquiring and structuring semantic information from text. In *Proceedings of the 17th International Conference on Computational Linguistics*.

[142] Riloff, E. (1996). Automatically generating extraction patterns from untagged text. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI'96)*.

[143] Rivadeneira, A., Gruen, D., Muller, M., and Millen, D. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*.

[144] Robinson, W. (1989). Integrating multiple specifications using domain goals. In *Proceedings of the 5th International Workshop on Software Specification and Design*.

[145] Rose, D. (2006). Reconciling information-seeking behavior with search user interfaces for the web. *Journal of the American Society for Information Science and Technology*, 57:797–799.

[146] Rose, D. and Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*.

[147] Roxana Girju, R., Badulescu, A., and Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.

[148] Ryo, J., Jung, Y., Kim, K., and Myaeng, H. (2010). Automatic extraction of human activity knowledge from method-describing web articles. In *Proceedings of the 1st Workshop on Automated Knowledge Base Construction*.

[149] Schank, R. and Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.

[150] Schmidt, C., Sridharan, N., and Goodson, J. (1978). The plan recognition problem: an intersection of psychology and artificial intelligence. *Artificial Intelligence*, 11:45–83.

[151] Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., de Velde, W., and Wielinga, B. (1999). *Knowledge Engineering and Management: The CommonKADS Methodology*. Massachusetts, MIT Press.

[152] Schubert, L. and Tong, M. (2003). Extracting and evaluating general world knowledge from the brown corpus. In *Proceedings of the HLT-NAACL 2003 Workshop on Text meaning, Association for Computational Linguistics*.

[153] Schubert, L., Van Durme, B., and Bazrafshan, M. (2010). Entailment inference in a natural logic-like general reasoner. In *AAAI Fall Symposium on Commonsense Knowledge (CSK'10)*.

[154] Shah, P., Schneider, D., Matuszek, C., Kahlert, R., Aldag, A., Baxter, D., Cabral, J., Witbrock, M., and Curtis, J. (2006). Automated population of CyC: extracting information about named-entities from the web. In *Proceedings of the 19th International FLAIRS Conference*.

[155] Singh, P., Lin, T., Mueller, E., Lim, G., Perkins, T., and Zhu, W. (2002). Open mind common sense: Knowledge acquisition from the general public. In *Proceedings of the 1st International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*.

[156] Smith, D. (2007). *EventMinder: A personal calendar assistant that understands events*. Master's thesis at the Massachusetts Institute of Technology.

[157] Smith, D. and Lieberman, H. (2010). The why UI: Using goal networks to improve user interfaces. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI'10)*.

[158] Soderland, S., Fisher, D., Aseltine, J., and Lehnert, W. (1995). CRYSTAL: Inducing a conceptual dictionary. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*.

[159] Sood, S., Owsley, S., Hammond, K., and Birnbaum, L. (2007). Tagassist: Automatic tag suggestion for blog posts. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'07)*.

[160] Strohmaier, M. (2008). Purpose tagging - capturing user intent to assist goal-oriented social search. In *Workshop on Search in Social Media (SSM'08) held in conjunction with CIKM'08*.

[161] Strohmaier, M. and Kröll, M. (2011). Acquiring knowledge about human goals from search query logs. *Information Processing and Management*.

[162] Strohmaier, M., Kröll, M., and Körner, C. (2009). Intentional query suggestion: Making user goals more explicit during search. In *Proceedings of the 2009 Workshop on Web Search Click Data (WSCD'09)*.

[163] Strohmaier, M., Lux, M., Granitzer, M., Scheir, P., Liaskos, S., and Yu, E. (2007). How do users express goals on the web? - An exploration of intentional structures in web search. In *International Workshop on Collaborative Knowledge Management for Web Information Systems*.

[164] Strohmaier, M., Prettenhofer, P., and Kröll, M. (2008). Acquiring explicit user goals from search query logs. In *International Workshop on Agents and Data Mining Interaction (ADMI'08) held in conjunction with WI'08*.

[165] Studer, R., Benjamins, V., and Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data Knowledge Engineering*, 25:161–197.

[166] Sure, Y., Staab, S., and Studer, R. (2002). Methodology for development and employment of ontology based knowledge management applications. *SIGMOD Rec.*, 31:18–23.

[167] Tatu, M. (2005). Automatic discovery of intentions in text and its application to question answering. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.

[168] Thielscher, M. (1997). Ramification and causality. *Artificial Intelligence*, 89:317–364.

[169] Thomas, B., Shoham, Y., Schwartz, A., and Kraus, S. (1991). Preliminary thoughts on an agent description language. *International Journal of Intelligent Systems*, 6:498–508.

[170] Turney, P. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning*.

[171] van Lamsweerde, A. (2000). Requirements engineering in the year 00: a research perspective. In *Proceedings of the 22nd International Conference on Software Engineering (ICSE'00)*.

[172] van Lamsweerde, A. (2001). Goal-oriented requirements engineering: A guided tour. In *Proceedings of the 5th IEEE International Symposium on Requirements Engineering (RE'01)*.

[173] van Lamsweerde, A., Letier, E., and Darimont, R. (1998). Managing conflicts in goal-driven requirements engineering. *IEEE Transactions on Software Engineering*, 24:908–926.

[174] van Riemsdijk, M., Dastani, M., and Winikoff, M. (2008). Goals in agent systems: A unifying framework. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08)*. International Foundation for Autonomous Agents and Multiagent Systems.

[175] Vapnik, V. (1998). *Statistical Learning Theory*. New York, NY, USA Wiley-Interscience.

[176] Von Ahn, L. (2006). Games with a purpose. *Computer*, 39:92–94.

[177] Weigend, A. and Gershenfeld, N. (1994). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison-Wesley.

[178] Wiebe, J. and Mihalcea, R. (2006). Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics (ACL'06)*.

[179] Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39:165–210.

[180] Wilensky, R. (1983). *Planning and understanding*. Addison Wesley.

[181] Winikoff, M., Padgham, L., Harland, J., and Thangarajah, J. (2002). Declarative and procedural goals in intelligent agent systems. In *Proceedings of Knowledge Representation (KR'02)*.

[182] Witten, I. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 2nd ed.

[183] Wu, X., Zeng, G., Han, F., and Wang, R. (2007). Goal representation and reasoning based on description logics (dls). In *Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering (ISKE'07)*.

[184] Yi, X., Raghavan, H., and Leggetter, C. (2009). Discovering users' specific geo intention in web search. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*.

[185] Yin, R. (2002). *Case study research: Design and methods (Applied Social Research Methods)*. SAGE Publications.

[186] Yin, X. and Shah, S. (2010). Building taxonomy of web search intents for name entity queries. In *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*.

[187] Yu, E. (1996). *Modelling strategic relationships for process reengineering*. Doctoral Thesis. UMI Order Number: UMI Order No. GAXNN-02887 (Canadian Dissertation), University of Toronto.

[188] Yu, E. (1997). Towards modeling and reasoning support for early-phase requirements engineering. In *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE'97)*.

[189] Yue, K. (1987). What does it mean to say that a specification is complete? In *Proceedings of the 4th International Workshop on Software Specification and Design (IWSSD'87)*.

[190] Zadeh, L. (1974). Fuzzy logic and its application to approximate reasoning. In *Proceedings of the International Federation for Information Processing (IFIP) Congress*.