Carsten Keßler

# CONTEXT-AWARE SEMANTICS-BASED INFORMATION RETRIEVAL

Mai 2010

Geoinformatik

<span style="color:red">CONTEXT-AWARE SEMANTICS-BASED
INFORMATION RETRIEVAL</span>

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
im Fachbereich Geowissenschaften
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Carsten Keßler
aus Georgsmarienhütte, Deutschland
im Mai 2010

## ABSTRACT

Information retrieval can benefit from contextual information to adapt the results to a user's current situation and personal preferences. In this respect, *semantics-based* information retrieval is especially challenging because a change in context may require modifications to the knowledge base at hand, such as updates to or reclassifications of individuals. This thesis introduces a novel approach for context-aware semantics-based information retrieval that covers two aspects.

First, context-aware system design requires an identification of *relevant* contextual information. For information retrieval, the impact of a contextual aspect on the query results determines its relevance. Performing the same query in different contexts often leads to different result rankings. The comparison of such rankings can provide insights into the effects of context changes on the information retrieval results. While numerous methods exist for assessing the result relevance with respect to a query, the question how *different* two result rankings are has not been tackled yet. The first part of this thesis is therefore concerned with the definition of a cognitively plausible dissimilarity measure for information retrieval results (DIR). It is based solely on the results and thus applicable independent of the retrieval method. The DIR measure supports cognitive engineering tasks, such as work flow and user interface design: Using DIR, developers can identify which contextual aspects strongly influence the outcome of the retrieval task and should therefore be in the user's focus. DIR's purpose is to reflect how human users quantify the changes in information retrieval result rankings. Its cognitive plausibility has been evaluated in two human participants tests, which show a strong correlation with user judgments.

Second, the relevant contextual aspects have to be modeled in a way that supports interaction with semantics-based knowledge bases. The Semantic Web is based on nominal data and it is therefore inherently difficult to integrate information from the Sensor Web, which is an increasingly important source of contextual information. The second part of this thesis introduces an approach based on semantic rules that bridge these two worlds to enable context-aware information retrieval from the Semantic Web. It demonstrates how user preferences can be modeled in the Semantic Web Rule Language (SWRL). SWRL's support for rules with free variables allows for reasoning on the individuals in an ontology – in the running scenario, the current conditions at surf spots in California are compared against a user model and ranked on the basis of their deviation from a user's preferences. Moreover, novel SWRL built-ins are introduced to dynamically read observations from the Sensor Web during rule execution, and to perform queries by example on individuals' data type values. This approach allows for a strict separation of static knowledge about individuals in an ontology and any dynamic information through an explicit link to sensors.

## PUBLICATIONS

This thesis is based on ideas, fragments and figures that have appeared previously in the following publications:

C. Keßler. Similarity Measurement in Context. In B. Kokinov, D. Richardson, T. Roth-Berghofer, and L. Vieu, editors, Modeling and Using Context. 6th International and Interdisciplinary Conference, CONTEXT 2007, Roskilde, Denmark, August 20–24, 2007, Proceedings, volume 4635 of Lecture Notes in Artificial Intelligence, pages 277–290. Springer, 2007.

C. Keßler, M. Raubal, and K. Janowicz. The Effect of Context on Semantic Similarity Measurement. In R. Meersman, Z. Tari, and P. Herrero, editors, On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, OTM Academy Doctoral Consortium, MONET, OnToContent, ORM, PerSys, PPN, RDDS, SSWS, and SWWS 2007, Proceedings, Part II, volume 4806 of Lecture Notes in Computer Science, pages 1274–1284, Vilamoura, Portuga, November 2007. Springer.

K. Janowicz, C. Keßler, M. Schwarz, M. Wilkes, I. Panov, M. Espeter, and B. Bäumer. Algorithm, Implementation and Application of the SIM-DL Similarity Server. In F. T. Fonseca, A. Rodríguez, and S. Levashkin, editors, Second International Conference on GeoSpatial Semantics (GeoS 2007), volume 4853 of Lecture Notes in Computer Science, pages 128–145, Mexico City, Mexico, November 2007. Springer.

K. Janowicz, C. Keßler, I. Panov, M. Wilkes, M. Espeter, and M. Schwarz. A Study on the Cognitive Plausibility of SIM-DL Similarity Rankings for Geographic Feature Types. In L. Bernard, A. Friis-Christensen, and H. Pundt, editors, 11th AGILE International Conference on Geographic Information Science (AGILE 2008), Lecture Notes in Geoinformation and Cartography, pages 115–133, Girona, Spain, 5–8. May 2008. Springer.

C. Keßler, M. Raubal, and C. Wosniok. Semantic Rules for Context-aware Geographical Information Retrieval. In P. Barnaghi, editor, 4th European Conference on Smart Sensing and Context, EuroSSC 2009, University of Surrey. Volume 5741 of Lecture Notes in Computer Science, pages 77–92, Springer, 2009.

K. Janowicz, S. Schade, A. Bröring, C. Keßler, P. Maué, and C. Stasch. Semantic Enablement for Spatial Data Infrastructures. Transactions in GIS, 14(2):111–129, 2010.

C. Keßler. What's the Difference? A Cognitive Dissimilarity Measure for Information Retrieval Result Sets. Knowledge and Information Systems, accepted subject to minor revisions.

# ACKNOWLEDGMENTS

# CONTENTS

xv

## LIST OF TABLES

# ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AI | Artificial Intelligence |
| AJAX | Asynchronous JavaScript and XML |
| DIR | Dissimilarity Measure for Information Retrieval Results |
| DL | Description Logics |
| IR | Information Retrieval |
| OAT | One At a Time |
| OGC | Open Geospatial Consortium |
| OWL | Web Ontology Language |
| RDF | Resource Description Framework |
| REST | Representational State Transfer |
| SOS | Sensor Observation Service |
| SWE | Sensor Web Enablement |
| SWRL | Semantic Web Rule Language |
| UML | Unified Modeling Language |
| WPS | Web Processing Service |
| WRS | Web Reasoning Service |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |
| XSD | XML Schema Definition |

# INTRODUCTION AND MOTIVATION

1

The first chapter provides the background for this thesis and motivates the idea of context-aware semantics-based information retrieval. The hypothesis as well as the methodology for testing it are introduced. A use case for context-aware geographic information retrieval serving as a practical example throughout this thesis is outlined.

## 1.1 INFORMATION RETRIEVAL AS A CORNERSTONE OF THE INFORMATION SOCIETY

The ever-growing amount of information available poses significant challenges when it comes to *finding* specific pieces of information. Before the invention of digital mass storage, index card systems were sufficient as a search tool for the most extensive information repositories of that time – libraries. Since the dawn of the digital age, hard disk capacities have been rising continuously, approximately doubling every two years[1]. The development of the World Wide Web has caused another vast increase in the amount of accessible information, allowing users to get hold of information distributed over countless servers located around the globe. Evidently, such an extensive amount of information requires search methods that go beyond index cards. The economic success of Web search companies such as Google shows that information retrieval is no longer just a scientific field that is merely of interest to specialists in the area – it is shaping the way the information society interacts with its most important good. As Boisot and Canals put it, "information has now become the main focus of economic transactions, and not merely a support for them" [22, p. 64].

*Information retrieval*

The largest portion of information available online has been made available for human users, mostly in the form of hypertext augmented with images and other kinds of multimedia. This hypertext Web is crawled and statistically analyzed by Web search engines in order to be able to present their users rankings of the most relevant Web pages for a specific search term. The idea of a *Semantic Web* was brought forward by Berners-Lee [16] to allow machines to understand content on the Web and to enable meaningful human-machine as well as machine-machine communication. Almost ten years later, the vision that the Semantic Web would allow users to seamlessly integrate any data (and eventually also services) must be regarded as overly ambitious and could not be realized yet. Nonetheless, the technologies that it builds upon (see Figure 1) have been successfully applied in numerous cases to integrate data across different services. This new emphasis of the

*Semantic Web*

---

1 According to *Moore's Law* [131], the number of transistors per integrated circuit doubles roughly every two years. This law has also been shown to apply to other properties of digital electronic devices such as memory.
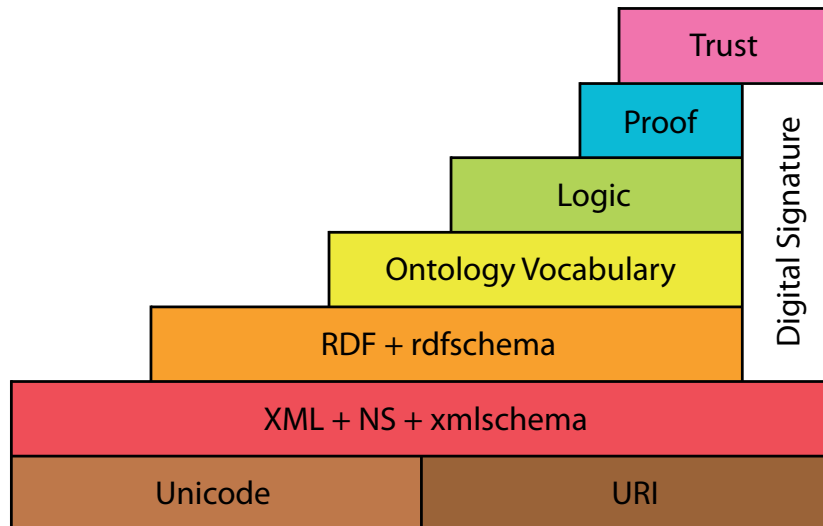
Figure 1: The Semantic Web Stack illustrates the different technologies that enable the Semantic Web, showing how they build upon each other. Adopted from Berners-Lee [15, slide 10].

Semantic Web on building distributed data repositories was recently coined *Linked Data* [20]. It also reflects the approach taken by specialized communities that adopt Semantic Web technologies to create structured, unambiguous specifications of their vocabulary for data annotation. Examples include the field of biology that has developed specifications for proteins, diseases and genomes[2], the cultural heritage community with its CIDOC CRM[3], as well as the field of geographic information science that will be in the focus of this thesis.

The main motivation for a community to make an effort and specify a domain ontology [65] (see Section 2.1) is the need for a formal (i.e., unambiguous) specification of the core terms of a domain and the relationships between them. Once a domain ontology has been developed, it does not only facilitate communication within the domain, but it also helps to overcome the lack of support from statistics-based information retrieval methods for small, specialized fields with their own vocabularies and comparatively small amounts of data online.

Standard information retrieval methods aim at "finding material [...] of an unstructured nature [...] from within large collections" [122, p.1]. In contrast, representations in logic-based languages such as the Web Ontology Language (OWL) [189] provide a structure that supports information retrieval based on logical reasoning. They enable queries such as "which geographic feature types are inland water bodies and carry salt water?". Novel approaches such as similarity-based retrieval

---

2 Further examples can be found in online repositories such as http://obofoundry.org or http://bioportal.bioontology.org.

3 CIDOC is the International Committee for Museum Documentation, which has developed a conceptual reference model (CRM) for cultural heritage documentation, see http://cidoc.ics.forth.gr.

[79, 83] build on these reasoning capabilities to produce intuitive results for queries such as "what is similar to a salt lake?".

The usefulness of the results of such queries, however, often depends on whether and how contextual information is taken into account – independent of the applied information retrieval method. The answer to the query for things similar to salt lakes, for example, depends on whether the context is *salt production* or *health resorts*.

The human ability to interpret such queries depending on the current context has been mostly neglected in inference-based approaches to information retrieval from semantically annotated information so far. They focus on the actual retrieval method and treat the query context as a mere add-on. The aim of this research is to stress the importance of context for cognitive information retrieval [177] and to demonstrate how contextual information can improve the retrieved results. The strong spatio-temporal component in almost any kind of contextual information makes context-aware retrieval a highly relevant topic for geographical information science. Typical kinds of contextual information such as available resources, nearby friends or colleagues, or the weather are inseparably tied to the current location. Combined with personal information about preferences and social network [47], such information can be used to provide highly personalized services. Information about a user's future movements allows a context-aware system to adapt information retrieval results to suit her travel plans[4]. As such, context-aware retrieval covers a number of the core aspects of geographical information science as defined by Goodchild [59], including data capture, measurement, algorithms and processes, display, uncertainty, as well as ethical issues.

The outcome of this thesis will facilitate the development of context-aware tools for semantics-based information retrieval by providing a measure for the impact of context changes as well as a new approach for modeling context based on rules. Both approaches are evaluated within the surf spot finder, a tool for the context-aware and personalized recommendation of surf spots at California's central coast. This use case has been chosen because it is a prime example of a retrieval task that depends heavily both on personal preferences and external influences such as weather and wave conditions. Finding a suitable surf spot requires both components that are characteristic of a context-aware application, a user model and the inclusion of external information collected from sensors. The latter is specifically challenging, as recent research has shown how sensor networks can benefit from semantic technologies [171, 36, 71, 178]. Hence, the surf spot finder covers multiple aspects that are frequently found in context-aware information retrieval.

---

4 Services such as `http://dopplr.com` are social networks for travelers that enable retrieval of accommodation opportunities, sights, etc., based on a user's contacts' recommendations.

Web search is the most common form of information retrieval today. Searching for a term such as "digital camera", for example, shows that the results will be a collection of results from different – albeit potentially overlapping – contexts. Results for such a query will likely include links to shopping sites, product reviews by customers and experts, discussion forums for photographers, pictures of (and taken with) different cameras, and more. A user, however, is most often only interested in results for one of these contexts and has to subsequently refine her query with additional keywords, such as "digital camera reviews". Different approaches have been developed to automate this disambiguation task [106, 107, 97, 77]. However, an increasing number of information retrieval tasks cannot be solved satisfactorily with such disambiguation strategies. Queries such as "where can I go for lunch today?" depend on multiple factors such as personal preferences or abilities, time of the year and weather, as well as the user's location.

Location-dependent queries are a prime example of context-aware information retrieval. Consider the screen shot of a search for *Pizza* on Google Maps™ shown in Figure 2. The top two screen shots show a map-based search in Münster, Germany, near central station, but with a slightly different map extent. The corresponding top results shown on the left have five matches in common, but in different order. The remaining matches are different in the two rankings. For the bottom screen shot, the spatial context has been moved to the area around Alexanderplatz in Berlin. Evidently, the results in the top two screen shots are very much alike, whereas the list in the bottom one is completely different from the other two[5]. This points to the problem that it is often hard to identify which contextual factors influence the outcome for a query, and to what degree.

Taking context into account is crucial when solving many information retrieval tasks in order to produce intuitive results and ultimately *Cognitive information retrieval* enable cognitive information retrieval [177]. While existing approches to information retrieval reach high precision and recall rates, they fail to adapt their results to user preferences and external influences. Hence, the goal of this research is to achieve high recall and precision for a *specific* user in a *specific* situation. Moreover, the question how context parameters that significantly influence the outcome of a given task can be identified has not been tackled so far. Accordingly, the objective of this thesis is to enable information retrieval to adapt to contextual influences as well as the users' preferences; it is *not* the goal to develop another retrieval algorithm. While the impact of location on the search results is relatively easy to assess for the pizza example, a use case of a surf spot finder application, which is the running scenario of this thesis, has a complex set of input parameters such as weather, water, as well

---

5 Some of the changes in the result rankings appear unintuitive (such as the drop of *La Cucina* from position 1 to 4 in the top two screen shots). Since Google keeps the criteria for these rankings undisclosed, we can only speculate that customer ratings and the page ranks [140] of the listed restaurants' Web pages probably play a role.

Figure 2: The spatial context for the *Pizza* query in the top two screen shots is very similar, leading to almost identical results. In the bottom screen shot, both the spatial context and the corresponding results differ completely from the top two.

as social aspects. It will be shown why such applications require tools to analyze the influence of different context parameters, and how such tools can be implemented.

*Pervasive computing*

With the growing computational power of mobile phones, research activities on context-aware computing (also referred to as pervasive or ubiquitous computing) have strongly increased. They aim at the development of software and devices that adapt automatically to the user's current context – her location, activities, users and devices in the vicinity, personal preferences, etc. To enable the development of such tools, research has come up with a number of definitions of context [40, 11], approaches for context modeling [181, 70], management [103, 199] and reasoning [28, 34] as well as development of context-aware software [41, 6]. In order to improve the context awareness in formal approaches to information retrieval, this thesis adapts ideas from context-aware computing and shows how they can be applied in information retrieval. Semantics-based information retrieval presents a special case of this problem. The technologies of the Semantic Web are all based on different flavors of formal logics. They have been built to provide support for different reasoning approches [66]. This comes (among other problems) at the cost of a lacking support for processing of numerical values, such as a pair of GPS coordinates. The second major aspect covered in this thesis will hence be the question how to model context using existing Semantic Web technologies so that processing of real-time information provided by sensors is enabled.

## 1.3 HYPOTHESIS AND RESEARCH QUESTIONS

Existing approaches for context-aware information retrieval often suffer from the lack of a clear definition of context. In order to clarify what context in information retrieval means, we will refer to the following definition throughout this thesis (see Section 2.4 for other context definitions and Section 3.1 for a detailed motivation of this one):

> DEFINITION 1: An information retrieval query's context is any information whose change significantly modifies the query results.

Note that this definition is itself context-dependent, as it is up to the specific application what a *significant modification* of the results is. In some applications, any slight changes might play a role (such as in medical applications), while for other applications, only substantial modifications are of importance (such as in Web search, where often only changes on the first result page are relevant). To make Definition 1 useful in practice, a measure for the impact of context changes on a information retrieval task's results is required. This will enable developers to assess which contextual information plays a role for their respective applications.

The according measurement method developed in this thesis is the DIR measure. It is a <u>D</u>issimilarity measure for <u>I</u>nformation retrieval <u>R</u>esult sets. DIR is based solely on the changes of position of a specific

match within the results (see the pizza example in Figure 2) and thus applicable independent of the retrieval method. In DIR, these *shifts* are weighted to highlight changes at the top ranks, as well as new or missing matches. If a result appears in both rankings under consideration, the difference in rank is weighted according to the higher of the two ranks. Results that appear in only one of the rankings are stressed with the maximum weight. DIR is the sum of all weighted shifts, normalized to the interval $[0, 1]$, where $0$ indicates identical input result rankings, and $1$ means completely different results that share no common entries. In order to contribute to cognitive information retrieval, it is crucial that the DIR measure really reflects how people perceive changes in result sets, i.e., that it bears *cognitively plausible* results that correlate strongly with human dissimilarity judgments. In this research we will compare dissimilarity measures obtained from DIR to human dissimilarity judgments of the perceived degree of change in a result set observed in human participants tests. DIR will be called cognitively plausible if its results correlate strongly with the results from the human participants tests. The statistical evaluation of the tests is therefore based on the null hypothesis ($H_0$) that there is no correlation between human dissimilarity judgments and the normalized sum of pairwise rank differences, each weighted by the higher of the two ranks. If we can falsify $H_0$, we have to accept the alternative hypothesis [143, ch.4]:

> HYPOTHESIS $H_A$: The normalized sum of pairwise rank differences, each weighted by the higher of the two ranks, correlates strongly with human judgments of dissimilarity between two rankings.

While a cognitively plausible measure for the impact of context changes helps to better understand how human users perceive such changes in information retrieval processes, it does not solve the problem of context representation. Current approaches towards context-aware information retrieval are limited to *internal* representations of context. The underlying assumption is that any contextual information that might play a role in the information retrieval process is already present in the knowledge base at hand. The context for a specific query is then defined by either limiting the query to a subset of the knowledge base [159, 79] or by assigning weights to the parts of the knowledge base [186, 86]. This does not only exclude other potentially important contextual information from the information retrieval process, it also contradicts the basic understanding of context in ubiquitous and pervasive computing research. In these fields, context refers to external information that is used to better solve a given task (see also the discussion of different definitions of context in Section 2.4). To overcome this limited understanding of context in semantics-based information retrieval, a rule-based approach is introduced that allows for the "injection" of external contextual information into existing knowledge bases.

From these two central aspects – context representation and measurement of the impact of context changes – arise the following research questions that are dealt with in the course of this thesis:

1. *How can relevant contextual aspects be identified?* In order to develop useful and performant context-aware applications, the identification of relevant context aspects is crucial. Since semantics-based information retrieval is the focus of this thesis, a measure is required to quantify the impact of context changes on the information retrieval task's results. The approach of the DIR measure introduced in this thesis circumvents the pitfall of manually selecting seemingly important contextual information and gives developers a tool to prove which of the available information (available from sensors, for example) must be taken into account.

2. *Which properties should a context impact measure have?* This part of the research includes a review of existing statistical correlation measures and an analysis of their characteristics to check whether existing methods can be employed to measure the impact of context changes. The goal is to specify DIR's behavior in terms of focus, weighting and normalization.

3. *How can the cognitive plausibility of the context impact measure be shown?* The DIR measure is supposed to reflect how users judge changes in information retrieval result rankings under context changes. This cognitive part of the thesis comprises designing, carrying out and statistically evaluating two human participants tests to show that the results obtained from DIR correlate with human judgments (i.e., to falsify the hypothesis).

4. *How can context be modeled so that it can be used to modify existing knowledge bases?* The functionality that is required to make a context representation useful for semantics-based information retrieval has to be analyzed. This includes a review of existing approaches towards context modeling and an argumentation for a rule-based approach (as opposed to a statistical approach, for example).

5. *How can the context model be linked to real-time information?* In order to make an information retrieval system context-aware, it must have access to contextual information. Such information can come in different flavors, such as user profiles or real-time information obtained from sensors. In order to realize these features, the rule-based model will be extended with sensor queries based on standard Semantic Web and Sensor Web technologies.

Figure 3 gives an overview of these research questions and shows in which chapters they will be dealt with.

## 1.4 EXPECTED RESULTS AND CONTRIBUTION

In order to develop context-aware semantics-based information retrieval into a useful tool, this thesis deals with two perspectives on context.

The first part of the thesis is concerned with the identification of relevant contextual information. Up to now, the decision which contextual aspects to cover in an application is up to the developer, who
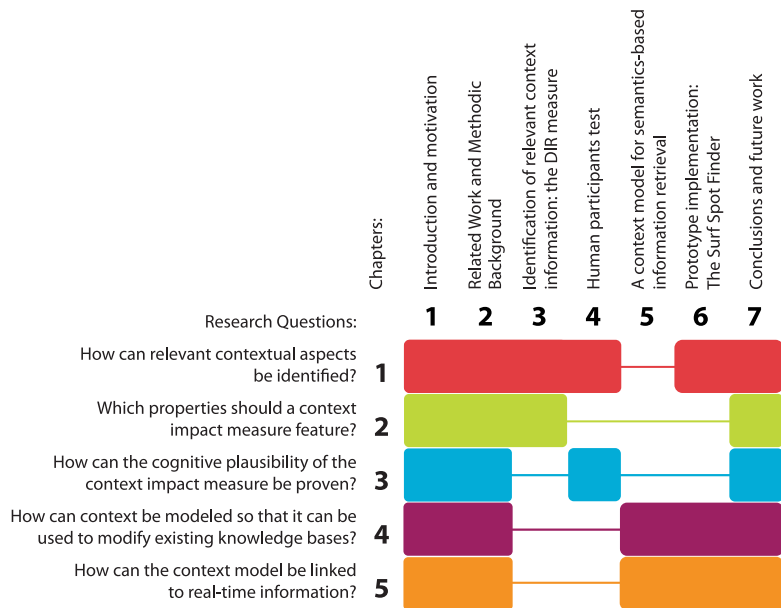
Figure 3: Overview of coverage of research questions in the chapters of the thesis.

decides based on personal experience or impression which aspects are relevant. Existing context definitions fail to guide the developer in choosing relevant context aspects for a specific application. To solve this problem, a dissimilarity measure for information retrieval results is introduced. This quantitative measure reflects how strongly users perceive changes in result sets. It thus allows for the assessment of the relevance of contextual aspects based on their influence on the results. This approach therefore allows for a definition of context based on relevance that also supports the developer in selecting contextual aspects. The change in results that are caused by modifications of the context should be measured in a way that reflects users' judgments of these changes. The measure is therefore tested with human participants to evaluate its cognitive plausibility and to investigate whether the result visualization influences user judgments (see Figure 4).

*Context identification*

*Cognitive plausibility*

Second, a new qualitative approach for the representation of context based on rules is developed. It allows for the inclusion of any contextual information into the information retrieval process and thus overcomes the limitation of existing approaches which rely solely on contextual information from the knowledge base at hand. Unlike the DIR measure, which is independent of the actual information retrieval method and knowledge representation, context rules are tied to Semantic Web technologies for execution. The rule-based approach allows for context-based temporal enrichment of knowledge bases with external information that is not possible with existing semantics-based retrieval methods and thus overcomes the static notion of information on the Semantic Web. From a modeling perspective, this approach forces a strict

*Context representation*

Figure 4: While most Web search engines, such as Yahoo!, use flat result lists (a), other tools such as the SIM-DL plugin [83] also specify the results' relevance via values (b) or by visualizing the results as tag clouds (c), where the font size indicates the quality of the results.

separation between the static part of an ontology, and those aspects that depend on the current context. A generic application design for the approach based on semantic rules is introduced and evaluated in a mobile prototype for personalized geographic information retrieval.

These two novel approaches developed during the research for this thesis form a framework that facilitates the development of context-aware information retrieval systems.

## 1.5 APPLICATION SCENARIO

Finding suitable spots (or routes) for activities such as hiking, diving or climbing [195] is a highly context-dependent task that does not reduce context to the user's location, as in some other cases [167]. In addition to the actual location of a specific spot, weather (and possibly water) conditions play a specific role. Especially the user's preferences and skills have a crucial impact on the selection of an appropriate spot for such activities: seemingly good spots may be boring, too challenging or even dangerous for a specific user.

In this thesis, the task of finding an appropriate surf spot will serve as a scenario [95, 198]. While the location plays an important role in the choice of a suitable spot (in terms of its distance from the current
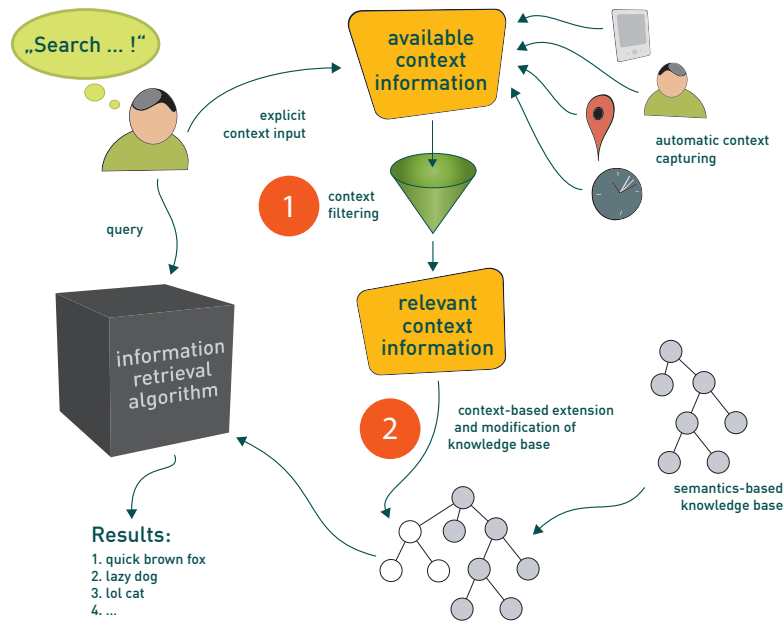
Figure 5: Context-aware semantics-based information retrieval workflow: Available context information is filtered to identify relevant information (1). This relevant information is used to extend and modify the knowledge base (2), which is then used by the IR algorithm to answer the user's query. This approach is independent of the applied retrieval method, which is treated as a black box here. Adapted from [92].

user), wave hight and frequency, currents and seasonal conditions have a big influence, too. These factors have to match the user's skills and personal preferences. Moreover, the social aspect plays an important role in the selection of a surf spot. Users may want to meet with friends for surfing and most surfers are looking for beaches that are not too crowded. Existing websites such as `http://magicseaweed.com` or `http://wannasurf.com` provide generic information on the current conditions, but without sophisticated functionality for personalized search.

This scenario has been chosen because it is a prime example of a geographic information retrieval [87] task that can only be solved satisfactorily in consideration of the query context. Any generic recommendations that do not take the user preferences and the current conditions into account can be seen as wild guesses that may or may not fit the current user. The surf spot finder will be a mobile application used to demonstrate how relevant context aspects can be identified and a rule-based system can be implemented that reflects different user preferences. It will also demonstrate how queries are adapted to the current conditions making use of the Semantic Sensor Web [171]. Once the client for this new surf spot recommendation system is finished, it will provide a query tool for spatial information that better matches human reasoning as outlined in Egenhofer's vision of the Semantic Geospatial Web [46].

*Geographic information retrieval*

Figure 5 shows the two steps within the context-aware information retrieval process this thesis deals with. First, relevant contextual information needs to be identified among the potentially available information that is captured from sensors or directly provided by the user. Since the goal of this research is to develop a method that completes this task in a cognitively plausible way, human participants testing is indispensable. Performing a human participants test includes test design [130] and careful sample size determination [114]. The goal is to make sure that the results obtained from the statistical evaluation of the collected data are significant. In the case of this thesis, significance of the evaluation is strongly tied to the notion of *cognitive plausibility* discussed in Section 2.5.

*Human participants testing*

Second, the contextual information that has been identified as relevant in step one must be integrated with the knowledge base at hand. The context model to be designed can be divided into two parts: user model and sensor model. These two aspects need to be designed in a way that is compatible and implementable with the chosen representation language, which consists of the standard languages of the Semantic Web (RDF, OWL, SWRL) in this case. These languages and their theoretical foundations are, in this respect, the method to put the presented approach into practice.

A context model for information retrieval on the Semantic Web must be based on a thorough analysis of the requirements in terms of functionality and flexibility to make sure that it fulfills its purpose. This analysis is especially important to identify technological restrictions. The standard approach for requirements analysis has been introduced by Maciaszek [120]. This strategy is driven by the Unified Modeling Language (UML) and focuses on commercial software development, covering aspects such as subcontractors, stakeholders, and project management. The overall process includes the requirements determination, requirements specification (development of use cases and class diagrams), architecture design, implementation, integration of modules, and maintenance. We will follow this well-established methodology for requirements determination and specification of the context model in this thesis, which acts as a transparent proxy between the IR method and the knowledge base. However, not all steps of the standard process will be applicable, since the UML-driven approach is targeted towards object oriented software, which is not the case for the context model in this thesis (at least in the narrower sense of object orientation[6]).

*Requirements analysis*

The remainder of this thesis is organized as follows (see Figure 3): the next chapter gives an overview of relevant related work from the areas of semantics and ontology, information retrieval, the Semantic Web, context modeling and context-aware applications, cognitive engineering as well as statistical correlation measures. Chapter 3 motivates

---

6 In this thesis, the term *object orientation* is used in the sense of object oriented programming, a software development paradigm that is based on object-like data structures with fields and methods [23].

and formalizes the DIR approach, followed by an evaluation of DIR's cognitive plausibility in two human participants tests in Chapter 4. Chapter 5 introduces a generic development approach for context-aware semantics-based information retrieval. Based on a requirements analysis, approaches for user modeling and linking to sensor networks are introduced. This generic approach is demonstrated in the Surf Spot Finder application presented in Chapter 6. This chapter also demonstrates the use of the DIR measure for the contextual aspects to include in a real application. Chapter 7 summarizes the results of this thesis and gives directions for future research.

RELATED WORK AND METHODIC BACKGROUND

This chapter points to relevant related work from the fields of semantics and ontology research (Section 2.1), information retrieval (Section 2.2), Semantic Web (Section 2.3), context modeling (Section 2.4), cognitive engineering (Section 2.5) and statistics (Section 2.6). The methods for semantics-based information retrieval, rule-based modifications of ontologies and statistical evaluation of human participants tests that will be applied in the remainder of the thesis are introduced.

## 2.1 SEMANTICS AND ONTOLOGY

Research on semantics deals with meaning in communication. While the view on semantics differs depending on the field of study, information systems research is mostly concerned with the question how the meaning of pieces of information and ways to interact with them can be unambiguously described. Any information system ultimately goes back to a conceptualization – the objects in the respective domain of interest, also referred to as the *universe of discourse*, their functions, and the relations between them [53]. This abstraction from the complexity of the real world is made explicit in order to achieve a shared and common understanding of a particular domain of interest:

*Universe of discourse*

> DEFINITION 2: An ontology is an explicit specification of a conceptualization [63].

A distinction is drawn between three kinds of ontologies. *Foundational ontologies*, also referred to as *upper-level ontologies* [146], are "axiomatic theories of domain-independent top-level notions such as object, attribute, event, parthood, dependence, and spatio-temporal connection" [168, p. 91]. Examples include the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [126] and OpenCyc [153]. While foundational ontologies provide generic specifications on an abstract level, *domain ontologies* specify the shared conceptualization of a specific community, such as the AGROVOC Ontology currently under development by the Food and Arrgriculture Organization of the United Nations[1]. *Application ontologies* are the most specific family of ontologies, as they specify the conceptualizations underlying specific applications [146]. In an ideal case, these three groups of ontologies are used in a layered fashion: an application ontology should be aligned to a corresponding domain ontology, which should in turn commit to a foundational ontology. Ultimately, *semantic interoperability* [19] between all parts of this ontological hierarchy is enabled.

*Semantic interoperability*

As mentioned above, ontologies are based on conceptualisations. The objects forming the universe of discourse underlying such a concep-

---

1 See http://www.fao.org/aims/agrovoccs.jsp.

tualisation are *concepts* – mental entities in people's heads. The idea of concepts goes back to Aristotelian philosophy, which defines them as "complex mental representations whose structure generally encodes a specification of necessary and sufficient conditions for their own application" [112, p.191]. This view on concepts is also known as the Classical Theory. Due to its inability to explain phenomena such as typicality (for example, although both dogs and gold fish are pets, dogs are more typical of the concept pet), alternative theories of concept have been proposed in the past decades, such as Prototype Theory [161] and Dual Theory [139]. An overview and discussion of the different theories of concept is given by Laurence and Margolis [112]. The above-mentioned theories are all epistemological views on concepts, which are based on knowledge (or justified believe) about apparent properties. This is also the general perspective on concepts in information systems, as opposed to metaphysical theories which focus on strict physical properties of concepts, sometimes down to an atomic level [175].

Despite the well-known shortcomings, research on ontology for information systems is still largely inspired by the Classical Theory. Within this thesis, the notion of concept will be used as follows (adapted from [56]):

> DEFINITION 3: A concept is a mental representation[2] of a set of entities that are grouped together.

Note that we will use the term concept interchangeably both for the mental entity as well as for its ontological representation; if this distinction is necessary, it will be noted. As Barsalou et al. [10] point out, mental concepts are inherently dynamic. They are subject to change when we are learning. For example, your concept of a PhD thesis may have changed since you started reading, depending on how many theses you have read before. Concepts may also change based on sensory input, as can be seen from the change of the concept *terrorist* since 9/11. Recent research tries to cover this dynamics of concepts in formal representations [151]. For this thesis, it is particularly important to note that concepts are context-dependent [10] – this phenomenon is formalized in the rule-based approach presented in Chapter 5.

Concepts – both as mental entities and their ontological representations – always abstract from the real world, since it is virtually impossible to know (and specify) all facts that apply to, for example, buildings. An ontological specification of such a concept can be *instantiated*:

> DEFINITION 4: An ontological instance of a specification of a concept is a concrete individual sharing its characteristics.

Within this thesis, instances are occurrences of ontological concepts. In many cases, instances refer to real-world entities: a domain ontology for transportation networks might define the concept *road*, and an instance of this concept would be a representation of a road in the real world (an entity). An ontology consisting of concepts and instances

---

2  The question how concepts are represented in people's minds and how well these transfer to external (for example, ontological) representations is a very interesting one, but out of scope for this research.

is also referred to as a populated ontology or knowledge base. In the remainder of this thesis, we will deal with information retrieval from such knowledge bases.

## 2.2 INFORMATION RETRIEVAL

The research field of information retrieval (IR) emerged in the 1950s out of the need for automatic ways to search large amounts of stored information[3]. While the field matured considerably over the decades, it was not until the advent of the World Wide Web that IR attracted attention beyond a comparatively small group of specialists. It turned out that IR methods are indispensable for searching the ever-growing host of information online and thus making it accessible in the first place. Although some definitions (such as the one by Manning et al. [122, p. 1]) limit IR to the retrieval of *unstructured* information, the term is generally also used for the retrieval of information from structured sources such as XML documents, databases or ontologies:

> DEFINITION 5: Information retrieval (IR) is a measurement $m$ for the relevance $\Re$ of an object $O$ for a user's information need. The information need consists of a query $Q$ and implicit information $I$, as well as information inferred ($\vdash$) from $I$: $IR = m[\Re(O, (Q, \langle I, \vdash \rangle))]$. [43, p. 10]

Note that this definition already contains a notion of context in the form of the implicit information $\langle I, \vdash \rangle$ to be taken into account. When the distinction is drawn, retrieval from structured sources is sometimes referred to as data retrieval. Strictly speaking, this distinction contradicts the common differentiation between data, information and knowledge [22]: information can only be extracted from data if the recipient knows how to interpret the data, and information ultimately becomes knowledge when it is *justified true belief*[4].

*Data, information and knowledge*

In many cases, a combination of techniques from structured and unstructured sources is required, for example when searching (structured) metadata that contain free-text descriptions. This combination also applies to semantics-based information retrieval and geographic information retrieval as targeted in this research. Classic information retrieval methods such as boolean, vector space or probabilistic-based methods, however, have been developed for unstructured information sources. These models were iteratively refined and combined with other methods such as artificial neural networks or natural language processing. With the advent of the Web, strategies were developed for semi-structured sources that take the connections between Web pages into account by analyzing links [140]. *Recall* and *precision* have been

*Recall and precision*

---

3 See, for example, Luhn [119] as a representative of the first influential papers written at that time.

4 This classical definition goes back to Plato; however, there is still no consensus among philosophers on the correct definition of knowledge.

defined as measures of quality, independent of the kind of information source and retrieval method [91]:

$$recall = \frac{|\{relevant\ docs\} \cap \{retrieved\ docs\}|}{\{relevant\ docs\}} \qquad (2.1)$$

$$precision = \frac{|\{relevant\ docs\} \cap \{retrieved\ docs\}|}{\{retrieved\ docs\}} \qquad (2.2)$$

Recall measures how many of the relevant documents are among the retrieved documents (Eq. 2.1). It is easy to reach a recall of 100% by simply returning *all* documents; these will necessarily also contain all relevant documents, along with all irrelevant documents. Thus, recall is only meaningful in combination with precision, which measures how many of the retrieved documents are actually relevant (Eq. 2.2). Differently weighted combinations of recall and precision are referred to as the F-measure [188]; the traditional $F_1$ measure refers to the harmonic mean of precision and recall. A perfect IR method would reach 100% recall and 100% precision ($F_1 = 1$), i.e., the retrieved results contain *all* relevant documents, but *no* irrelevant documents. Measurement of recall and precision usually requires the involvement of experts who can reasonably make the distinction between relevant and irrelevant documents for a specific query.

In practice, retrieving relevant information for a given query often entails finding contents *similar* to the query. This insight lead to the investigation of *similarity measurement* as a novel IR method. Similarity measurement has been a research topic for over half a century in psychology [57]. The objective of the research in this field is to understand the cognitive processes that underly human similarity judgments. While the actual process humans perform during similarity measurement could not be clarified completely yet, scientists have come up with a number of different approaches to modeling it, including feature-based [186], network-based [148], geometric [52] and alignment models [54].

*Similarity measurement*

The artificial intelligence (AI) community started investigating the topic from a different perspective, focusing on how human similarity ratings can be mimicked by algorithmic approaches [155]. The objective is to achieve similarity ratings that correspond to those of humans and can thus be applied to information retrieval use cases. AI research thus aims at the outcome rather than the processes leading to those results. Note that AI research looks at similarity measurements among computable *representations* of human conceptualizations as formalized in ontologies or other kinds of knowledge bases. In this respect, the main challenge with *semantic* similarity measurement is the comparison of meanings as opposed to purely structural (such as lexical or statistical) comparison. To enable such measurements, a representation language and an according similarity measure that is able to handle the expressivity of this language are required (see Section 2.3).

After successful application in areas such as case based reasoning [88], similarity measurement has also gained attention as a tool for geographic information retrieval [170]. The matching-distance similarity measure (MDSM) is a feature-based approach that has been the first similarity measure developed specifically for this domain [158, 159]. Other approaches [150] employ conceptual spaces [52], determine similarity based on affordances [81], or provide similarity measurements for spatial scenes [116, 134]. Several measures have been developed to quantify similarity within (geo-)ontologies [25, 38, 169, 83].

Semantic similarity measurement is especially attractive as a tool for *cognitive* information retrieval due to its potential to simulate human similarity judgments. A key factor for the cognitive plausibility (see Definition 8) of a similarity measure is the treatment of context. Various human participants tests have demonstrated that the perceived similarity of two objects varies considerably depending on the current context [186, 108, 127, 17, 58]. As Murphy and Medin put it, "the relative weighting of a feature [. . . ] varies with the stimulus context and task, so that there is no unique answer to the question of how similar one object is to another" [132, p.292]. In that respect, context-free similarity measurements can be compared to topographic (i.e. multi-purpose) maps, whereas context-dependent similarity measurements are tailored to specific tasks and situations, comparable to thematic maps.

Context is a crucial aspect to incorporate when applying similarity measurement as a tool for cognitive information retrieval [177]. The treatment of context in existing semantic similarity measures, however, is mostly limited to a restriction of the set of potential results to a subset of the knowledge base under consideration [159, 79]. The context model developed in this thesis enables context-aware similarity measurement that allows other modifications to the knowledge base, independent of the actual similarity measurement. It is evaluated in a prototype application that evaluates the similarity of specific properties of a surf spot and the corresponding user preferences. Moreover, the evaluation of the relevance of specific context parameters for an application is based on a *dissimilarity* measure for information retrieval result sets. This approach follows the assumption that the less similar the results for a query posed in two different contexts are, the bigger the influence of the context is.

The potential of similarity measurement as a tool for IR relies on the availability of structured data as provided on the Semantic Web. Independent of the underlying kinds of knowledge representation that IR algorithms work on, recent research efforts strive for a more user-centered approach in information retrieval. Context-awareness is only one part that contributes to cognitive information retrieval [177], amongst interactive IR, feedback and explanations. With similar objectives, the term *human-computer information retrieval* (HCIR) was coined for a research field that combines information retrieval with human-computer interaction to open up new opportunities towards more intuitive, interactive, and goal-directed IR [162]. Marchionini [124] suggests "that integrating the human and system interaction is

the main design challenge to realizing these opportunities and that what is required is design that recognizes a kind of syminforosis – people as organic information processors continuously engaged with information in the emerging cyberinfrastructure."

## 2.3 SEMANTIC WEB

*Semantics-based information retrieval*

Classical information retrieval methods make assumptions about the relevance of documents for specific queries based on statistical measures. Due to the unstructured nature of the documents under consideration, it is not possible to *logically infer* whether and how a document is related to a query. This lack of reasoning capabilities was one of the main motivations for the Semantic Web. The Resource Description Framework (RDF) [123] and the Web Ontology Language (OWL) [189] are the most prominent approaches that have been developed under this umbrella term in order to enable semantics-based information retrieval. RDF was developed by the World Wide Web Consortium (W3C) as a metadata model based on triples consisting of subject, predicate and object. The most widely used implementation of the abstract RDF model is realized in XML, as shown in the following extract which specifies the title and publisher for the Wikipedia page about Tim Berners-Lee:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
    rdf:about="http://en.wikipedia.org/wiki/Tim_Berners-Lee">
    <dc:title>Tim Berners-Lee</dc:title>
    <dc:publisher>Wikipedia</dc:publisher>
  </rdf:Description>
</rdf:RDF>
```

While the information in this example is certainly also easily extractable for classical statistics-based IR methods, RDF triples can also be used to express more complex facts such as relations between people[5] or representation of the WordNet lexical database[6]. Since RDF only allows for the specification of relations between instances, RDF Schema [30] has been introduced to enable the specification of concepts and thus add categorization capabilities to RDF. The SPARQL Query Language for RDF [147] allows searching RDF repositories based on predefined criteria, comparable to SQL for relational databases. RDF and SPARQL are also central technologies for the ongoing efforts to provide Linked Data [20], an initiative that aims at the connection of annotated online datasets as an intermediate step towards the Semantic Web.

OWL is built on top of RDF and adds additional expressivity, which varies depending on which of the three different sublanguages OWL-Lite, OWL-DL or OWL-Full is used. The three sublanguages correspond

---

5 The Friend of a Friend (FOAF) project is a popular application of RDF; see http://foaf-project.org.

6 See http://w3.org/TR/wordnet-rdf.

to different languages from the family of Description Logics (DLs) [7].
DLs are used to model concepts and entities in a knowledge base. Such
a knowledge base consists of a TBox containing the terminology, i.e. the
vocabulary describing a given domain, and an ABox storing assertions
about instances. Description logics distinguish between logical and
non-logical symbols. The former have a pre-defined meaning grounded
in set theory, while the latter are domain specific. Logical symbols are
either constructors ($\sqcap, \sqcup, \exists, \forall, \leqslant, \geqslant$) used to compose complex concepts
out of primitive ones or connectives such as equality ($\equiv$) or inclusion
($\sqsubseteq$). OWL-Lite corresponds to the $\mathcal{SHIF}$ description logic, while OWL-
DL is based on $\mathcal{SHOIN(D)}$[7]. OWL-Full, however, is based on different
semantics than the former two and is not decidable, i.e., it does not
necessarily support reasoning services such as subsumption reasoning.
Note that the three OWL languages build on each other: OWL-Full
contains OWL-DL, which in turn contains OWL-Lite. The following
extract from an OWL document introduces the restriction that the
concept specified in node A42 (specifying the concept *IrrigationCanal*,
not shown here) has the property *constructedFor* and that this property
can only take values from *Supply*; in other words, this specifies that
irrigation canals are always constructed for a certain kind of supply:

```
<rdf:Description rdf:nodeID="A42">
  <owl:someValuesFrom rdf:resource="#Supply"/>
  <owl:onProperty rdf:resource="#constructedFor"/>
  <rdf:type
    rdf:resource="http://www.w3.org/2002/07/owl#Restriction"/>
</rdf:Description>
```

The specification of OWL ontologies enables a host of new IR meth-
ods that exploit the structured information constructed from such OWL
fragments. These IR methods are all based on combinations of the
reasoning mechanisms OWL (i.e., the respective underlying DL) offers.
Reasoning services include satisfiability (consistency checking), sub-
sumption (finding subconcepts) and equivalence reasoning. Ontology
editors such as Protégé enable communication with reasoners such as
Pellet [172] or FaCT++ [183] via the DIG interface [12]. The recently
introduced new version of the Web Ontology Language, OWL 2 [190],
provides novel profiles targeted at different reasoning capabilities as
supported by the OWL API[8] [74].

The dedication of OWL Lite and OWL DL to the decidability of
key inference problems leads (among other restrictions) to a lack of
expressivity for specific description tasks. Examples include ternary
relations such as the common *uncle* example ("if A is B's brother, and B
is C's parent, then A is C's uncle") which cannot be expressed in OWL.
In a Semantic Web environment, this limitation to binary relationships
becomes relevant for specifications of pre and post conditions of Web
services. Various approaches have been developed to overcome these
restrictions by combining ontology languages with rules (see Antoniou

---

7 The description logic complexity navigator gives an overview of the expressivity of
different DL languages: `http://www.cs.man.ac.uk/~ezolin/dl/`.
8 See also `http://owlapi.sourceforge.net`.

et al. [5] for an overview). The Semantic Web Rule Language (SWRL) [76] adds rule capabilities to OWL ontologies. SWRL rules are Horn-like clauses [73], consisting of an antecedent (body) and a consequent (head), each containing a set of atoms. For example, the rule

```
hasBrother(?a,?b) ∧
hasParent(?c,?b)
→ hasUncle(?c,?a)
```

formalises the "uncle" example given above in SWRL prefix notation. The SWRL submission to W3C introduces a concrete XML syntax for use with OWL, which is achieved by combination with RuleML[9] statements. SWRL has been successfully applied for geographic information for the specification of category membership rules, where the categories are formalized in OWL-DL [99, 100]. Moreover, a rule-based approach closely related to SWRL has been implemented to deduce implicit spatial semantics and check spatial integrity [174, 173].

SWRL can be extended by additional reasoning mechanisms via built-ins. Among the built-ins that have been implemented so far are libraries for mathematics, string handling and date support, enabling rules such as

```
Person(?p) ∧
hasAge(?p,?a) ∧
swrlb:lessThan(?a,18)
→ Minor(?p).
```

Built-ins thus play a central role for the mapping from numeric sensor inputs to ontologies via SWRL. Chapter 5 demonstrates how SWRL can be used to develop user models for the Semantic Web that allow for modifications of knowledge bases according to a user's personal preferences. Moreover, SWRL's built-in mechanism is used to include contextual information from sensors in the information retrieval process.

## 2.4 CONTEXT

Any definition of context is heavily dependent on the field of application, as shown by the analysis of 150 different definitions by Bazire and Brézillon [11]. Looking at definitions within the field of computer science, the literature mostly falls back on enumerations of examples. In other cases, the definitions are too specific to be transferable to other application areas [141]. A widely adopted definition of context for ubiquitous computing – which is also concerned with context-aware information retrieval – defines context as follows:

> DEFINITION 6: Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves [40, p.5].

---

9 See http://www.ruleml.org.

22

The central aspects in this definition are identity (user), activity (interaction with an application), location and time (as the temporal constraints of a certain situation). This list does not claim completeness, nor do all of the aspects always play a role. In the same paper, Dey defines context awareness:

> DEFINITION 7: A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task [40, p.5].

Dey's definition of context (Definition 6) points to potential sources of contextual information on a generic level. His definition of context awareness (Definition 7), however, is directly related to the view on context for information retrieval taken in this research, wich defines an information retrieval task's context as *any information whose change significantly modifies the task's outcome* (see Definition 1). The goal of this research is therefore to develop an approach for semantics-based information retrieval that enables a system to provide relevant information to the user.

While the idea of context in many cases remains at least slightly ambiguous, research activities on context-aware information retrieval have increased remarkably in recent years. The ubiquitous and pervasive computing communities have developed numerous approaches to automatically provide users with information and services based on their current situation – see [118] for an overview. Existing context-aware applications range from smart spaces [193] over mobile tour guides [35] to generic prototyping platforms [149]. Context has been successfully employed in a number of ways for personalization [98, 95] and to improve retrieval from the Web [49, 107], from email archives [194], from RDF graphs [3], from Wikipedia [187], as well as for ontology-based alert services [115], intention recognition [96], and for the disambiguation of historical place names [89]. Different approaches based on vector spaces have been proposed to enable context-aware information retrieval [52, 150, 128]. Maeda identifies context as one of his *laws of simplicity*, stating that "what lies in the periphery of simplicity is definitely not peripheral" [121, p.53] – transferred to information retrieval, this statement is a clear hint that the context of a retrieval task needs to be taken into account to make completing the task as simple as possible for the user. Lawrence [113] provides an overview of the different strategies to make Web search context-aware. Yahoo introduced context-aware tools [107] that automatically extend the user's query by text from the Web site the user is currently visiting or from the file she is working on. Google provides a custom search extension for Wikipedia[10] that allows a user to search Wikipedia pages related to the article the user is currently viewing. The context is in this case restricted to a given topic or to pages the current article refers to. Dumitrescu and Santini [44] present a context-aware information retrieval method for documents. Contexts are traces of documents the user has been working on, which they represent as self-organizing maps [101]. While the authors point

*Context-aware applications*

---

10 See http://en.wikipedia.org/wiki/Wikipedia:Google_Custom_Search.

out that the advantage of this approach is that context does not need to be represented by logical means, it also prevents logical reasoning on the contextual information. Semantic annotations are therefore an inevitable requirement to reason about contextual information.

Most of the research outlined above focused on solutions for specific applications or tasks and the proposed solutions are thus not easy to generalize or adapt to different tasks. The Semantic Web community has investigated different approaches towards semantic models of context and contextualizing ontologies, since the static notion of concepts in OWL does not meet the requirements of context-aware applications. While some approaches propose to use OWL ontologies to model context [64], the more generic approaches try to modify existing Semantic Web technologies with extensions for context representation and inclusion. CTXML [27] is an XML dialect that enables local, contextualized models within ontologies. It specifies an XML-encoding for contexts as concept hierarchies that can be assigned information on the owner, the groups the context is specified for, and on security and history. Mappings between different contexts based on distributed description logics (DDL) [24] allow for the identification of different specifications that refer to the same concepts. CTXML eventually merged into C-OWL [29], which is an extension to OWL that enables contextual ontologies via bridge rules that map between the local ontologies. C-OWL thus adds the capability to handle potentially inconsistent knowledge. *MultiContext Ontologies* (MuCO) [13] have been designed with a similar goal. They are based on an extended version of description logics that enables different interpretations of concept terms depending on context. Comparable to C-OWL, MuCO builds on predefined contexts built into the ontology; for example,

*Student* $= (\exists$ *StudentNumber.Number*$)$[s1] $\sqcup$ $(\exists$ *Takes.Class*$)$[s2]

specifies that in context s1, the concept *Student* is restricted to the entities that have a student number assigned, whereas in context s2, it is restricted to entities that take a class.

Korpipää and Mäntyjärvi [102] propose an RDF-based context ontology. It specifies context types (e.g. Environment:Light:Intensity) that can be filled with values (e.g. *Dark*). However, they do not describe how the mapping from the raw sensor data onto the context values (*Dark* in the above example) works.

The use of SWRL for reasoning about context information has already been proposed within the A-MUSE project [200]. However, the utilization of SWRL within the project was limited to *reasoning* about context information present in the ontology. For example, SWRL was used to find out whether a person is currently attending a meeting in a certain room based on the property values of the individuals in the knowledge base. We extend this approach by using SWRL to *represent* context (instead of only reasoning about it) and by querying context information directly from a novel SWRL built-in.

*Contextualized ontologies*

The idea of *cognitive engineering* as an interdisciplinary approach to improve the usability of machines and user interfaces goes back to Norman [137, 138]. He identified a *gulf of execution and evaluation* which results from discrepancies between a system's variables and the cognitive variables of a user. Cognitive engineering applies methods from the cognitive sciences to identify mismatches between the user's expectations (which may differ from one user to another) and the functionality offered by the system. The user-centered system design put forward by Norman has since been applied in numerous application areas ranging from unser interface design for information systems [136] and geographic information science applications such as location based services [152] to the design of complex working environments such as a passenger plane's cockpit [55].

While the classical view of cognitive engineering is focused on user interfaces, information engineering is another field that benefits from more intuitive representations. Kuhn [110] presents a *semantic engineering* perspective, arguing that the mental concept to be modeled can be treated as a black box for engineering purposes, as the engineer has no access to it whatsoever. Instead, he proposes to focus on the specifications of the symbols and their use in the communication of (spatial) information. The symbols are specified by constraints, which should ultimately be grounded in reality. Analogous to spatial reference systems, he proposes the use of *semantic reference systems* to anchor the symbol's specifications in the real world [145, 165, 166].

*Semantic engineering*

In a less engineering-driven approach, Gärdenfors [52] claims that information representation should be grounded in human perception and cognition. His theory of *conceptual spaces* builds on quality dimensions with a geometrical or topological structure for one or more domains. The basic dimensions are grounded in the human perception of color, for example, with the dimensions *hue*, *saturation* and *brightness*. Concepts cover multiple domains and are modeled as n-dimensional regions. Every object or member of the corresponding category is represented as a point in the conceptual space. This allows for expressing the similarity between two objects as the spatial distance between their points. Moreover, conceptual spaces support the paradigm that concepts are dynamical systems [9]. Dietze et al. [42] demonstrate how to utilize conceptual spaces for context-aware information retrieval. They introduce a meta model for *conceptual situation spaces* and apply it successfully to retrieve personalized learning material, based on a user's preferred learning style. Adams and Raubal [1] introduce the Conceptual Space Markup Language (CSML), an XML dialect to enable online sharing of and reasoning with conceptual spaces. CSML is the first attempt to turn conceptual spaces into a functioning computational environment that overcomes the limitations of the logic-based standard Semantic Web languages. While CSML is a promising approach towards more intuitive reasoning on the Semantic Web [51], this research builds on standard Semantic Web technologies to ensure applicabil-

*Conceptual spaces*

ity of the developed approaches in existing infrastructures. Moreover, the motivation is to test the flexibility of existing standards for novel applications.

While Gärdenfors does not claim that conceptual spaces model how human conceptualization works, his theory certainly appears more intuitive than the Semantic Web technologies and their foundations in logics (see Section 2.3). Strube [182] has first addressed this issue for knowledge engineering and defined different degrees of cognitive adequacy. *Weak cognitive adequacy* is reached when a "system has been carefully built with the human user in mind" [182, pp.165–166]. *Strong cognitive adequacy*, in contrast, implies the systems apply "the very same principles of cognitive functioning as human experts do" [182, p.167]. He further divides strongly adequate systems into *absolutely* adequate systems, which model "the human expert's knowledge and way of reasoning in every aspect", and *relatively* adequate systems, where the "kind of knowledge representation and reasoning as used in the system can be found in human experts, too" [182, p.167].

We will not discuss here on whether research in psychology has yet revealed enough details about the human cognitive apparatus to build a model of it that holds up to the criterion of strong cognitive adequacy. This discussion is clearly out of scope for this research. While some engineering disciplines such as robotics may claim this goal, we are concerned with user-centered information retrieval. The goal is therefore not to build a system that performs information retrieval in the way a human agent would perform this task. Instead, we focus on the system output, which should reflect the output of human agents performing this task. Our goal is thus between Strube's definitions of weak and strong cognitive adequacy: we want the system to *plausibly* mimic the output human users produce. This is a higher requirement than building a system "with the human user in mind", yet we do not strive for a system that models "the human experts way of reasoning in every aspect". While the latter is hard to verify, the notion of cognitive plausibility used in this research can be checked by evaluation against the output of human users:

> DEFINITION 8: A system is called *cognitively plausible* if the output it produces in performing its task correlates strongly with human output when performing the same task.

This definition of cognitive plausibility will be of importance for the specification (see Chapter 3) and testing (see Chapter 4) of the DIR measure. We are thus not interested in *how* human agents compare sets of information retrieval results, but the measure is supposed to reflect human dissimilarity judgments.

## 2.6 STATISTICAL CORRELATION MEASURES

Correlation is a well known problem with established statistical solutions. Correlation measures such as Pearson's $r$ [157], Spearman's $\rho$ [176] and Kendall's $\tau$ [90] go back to the 1930s and have proven useful for

countless studies across different disciplines. They have been developed for different kinds of data to analyze. Pearson's product-momentum correlation coefficient r can be applied to data if they do not deviate too far from the normal distribution. If this is not given, distribution free methods have to be applied. Among these are rank correlation coefficients which are not based on the actual data values, but on their order. Spearman's rank correlation coefficient ρ requires equidistant ranks; if this is not given, Kendall's rank correlation efficient τ can still be applied. The coefficients functions all map to the interval $[-1, 1]$, where 1 indicates perfect positive correlation (i.e., perfect agreement), 0 indicates no correlation (i.e., independent data) and $-1$ indicates perfect negative correlation (i.e., perfect disagreement). While these three are the most prominent correlation coefficients, numerous variants and tests based on them have been developed for specific kinds of data, especially in the behavioral sciences [62]. In the setting of this research, statistical correlation measures play a twofold role.

First, the comparison of information retrieval results can be regarded as a specific correlation problem. However, none of the standard correlation measures are tailored to the specific requirements of this task. This is due to the fact that those correlation coefficients have been developed to calculate correlation between sets of result pairs where every individual in one ranking has its corresponding counterpart in the other ranking. In information retrieval, however, cases where individuals (single results in this particular case) appear in one ranking, but not in the other, are very common. Thus, the two rankings to be compared often consist of different numbers of ranks. Finally, each rank may hold more than one result of equal relevance[11]. While this concept is hardly reflected in user interfaces for reasons of simplicity (see [83] for an example), IR algorithms mostly deal with such shared ranks internally. These ranks contain results that have been assigned the same relevance rating with respect to the current query. The DIR measure developed in this thesis must be able to cope with all these kinds of input. While these peculiarities can be dealt with in principle [135], the output of statistical correlation measures ranges from $-1$ for perfect negative correlation to 1 for perfect positive correlation. In the case of IR results, negative correlation is both hardly to expect and very difficult to explain. A mapping interval of $[0, 1]$ is therefore more appropriate for the desired dissimilarity measure, where 0 applies for two identical result rankings and 1 applies for any pair of rankings that share no common entries.

From a cognitive perspective, statistical correlation coefficients present a purely mathematical view on the rankings at hand. While they can be employed to compare result rankings in principle (such as Kendall's τ in [45]), they do not reflect the human perspective on the results at hand, including the focus on top results [2]. This property, however, should be in the centre of a measure that supports cognitive IR. Comparable to the semantic similarity measures building on similarity research in

---

11 We use the term *relevance* here to refer to the calculated relevance of a relevance based on the applied IR method, not for relevance feedback collected from users.

psychology, the DIR measure needs to reflect the user's view on IR results. The cognitive aspect of this research is therefore concerned with the question how to map from two given result rankings to a value between 0 and 1, so that the order of the resulting values corresponds to human judgments.

The second aspect in which correlation measures play a role in this research is the evaluation of the correspondance to human judgments. The cognitive plausibility of the DIR measure discussed in Section 2.5 will be evaluated by a human participants test in Chapter 4. As mentioned above, correlation coefficients have their roots in a the analysis of the data generated in such tests. Accordingly, the test data collected during the human participants tests will also be analyzed by these standard methods. The choice of the respective method depends on the characteristics of the data at hand and will be discussed in Sections 4.1 and 4.2.

# IDENTIFICATION OF RELEVANT CONTEXT INFORMATION: THE DIR MEASURE

This chapter introduces the cognitively plausible dissimilarity measure for information retrieval results (DIR). The need for such a measure is motivated in Section 3.1. The characteristics of the measure are discussed using examples from geographic information retrieval (Section 3.2) and the measure's formalization is introduced (Section 3.3).

## 3.1 MOTIVATION

The question which contextual aspects to include in a context-aware application is crucial from the users' as well as from a computational perspective. Users want the application to account for any *relevant* context changes, but processing power can be a bottle neck, especially on mobile devices. Moreover, the collection of context information can be costly when additional sensors are required. Furthermore, the computing of too much sensor input can slow down the application and hamper usability. Accordingly, it is important for application designers to be able to assess the impact of context changes to decide which aspects to include in their applications, and which to ignore. Finally, being able to measure the impact of context changes is also interesting from a theoretical perspective, since it enables distinguishing between noise and intended context[1] [80]: intended contextual information must have an impact that goes beyond a threshold value δ, otherwise it is considered noise. The goal for user interfaces is therefore to clean them from contextual noise and reduce the interaction options for the user to intended contextual information.

*Identifying intended context*

Research on user interaction with information retrieval results has been of growing importance since the broad breakthrough of Web search engines. Understanding how users work with search interfaces is important whenever large collections of text and multimedia are made available. Providing users with an intuitive interface for retrieval is crucial for businesses selling and organizing information. Such research that concentrates on the presentation of results to the users and their interaction with them [8, Chapter 10] has been coined human-computer information retrieval (HCIR) [124]. Developing HCIR applications is especially challenging when it comes to context-aware retrieval that adapts the results to personal interests and preferences, the user's

*Human-computer information retrieval*

---

1 As kinds of context for similarity measurement, Janowicz [80] further distinguishes user context (personal background, environment, motivation), application context (application-specific parameters), discourse context (domain of application), representation context (context-dependent changes of the knowledge base; see Chapter 5) and interpretation context (interpretation of one result with respect to other results). These kinds of context also apply for the general case of information retrieval, except for the discourse context which is specific to similarity measurement.
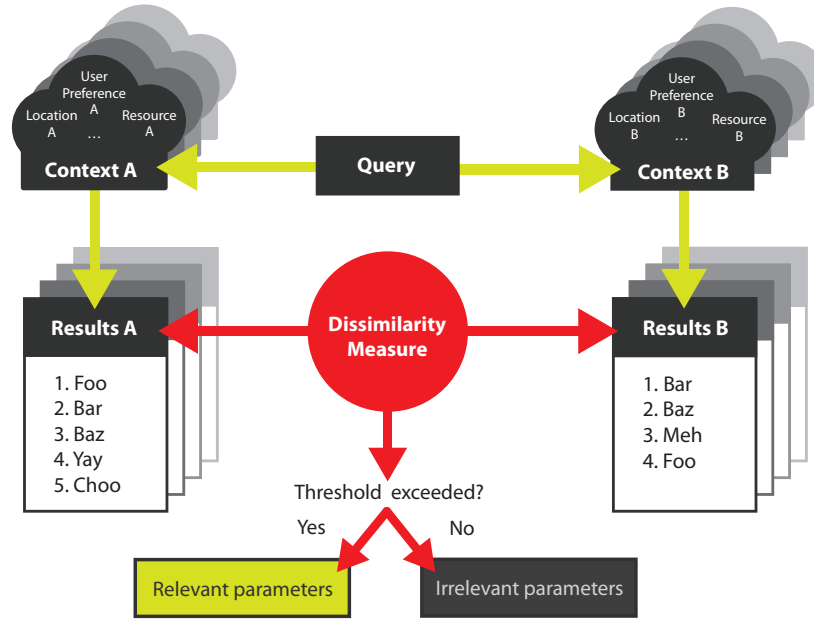
Figure 6: Identification of relevant context parameters using the DIR measure. Adapted from [93].

current location and other context information relevant to the given task (see Definition 1 and [32]). As discussed before, changes of context can cause an adaptation of the result ranking for a given query, so that the same query does not necessarily always lead to the same result [77].

Comparisons of result rankings stemming from the same query posed in different contexts can provide useful insights into the effects of context changes on the IR results. While the context for the pizza example (see Figure 2) consists only of the selected map extent, applications such as a surf spot finder (see Section 1.5 and [95, 198]) require more detailed context models: In order to allow a user to retrieve a personalized selection of surf spots based on the current conditions, the context model must cover wind and water conditions, as well as the user's skills and preferences. In such a complex setting, developers have to focus on the context aspects that can cause the biggest turbulences in the results when they change. An analytical tool that supports this assessment task helps to reduce the complexity of the application by removing context aspects that have a negligible influence on the outcome. A measure for the *quantification* of the users' perception of the difference between two rankings would support the reduction of user interfaces as well as computational complexity of context aware applications, as shown in Figure 6: Relevant context parameters can be identified if their change causes a modification of the results for a given query that exceeds an application-specific threshold. For this purpose, the same query is posed repeatedly in different contexts, which differ in certain aspects. If the corresponding result rankings are modified beyond the given threshold, they have to be taken into account for
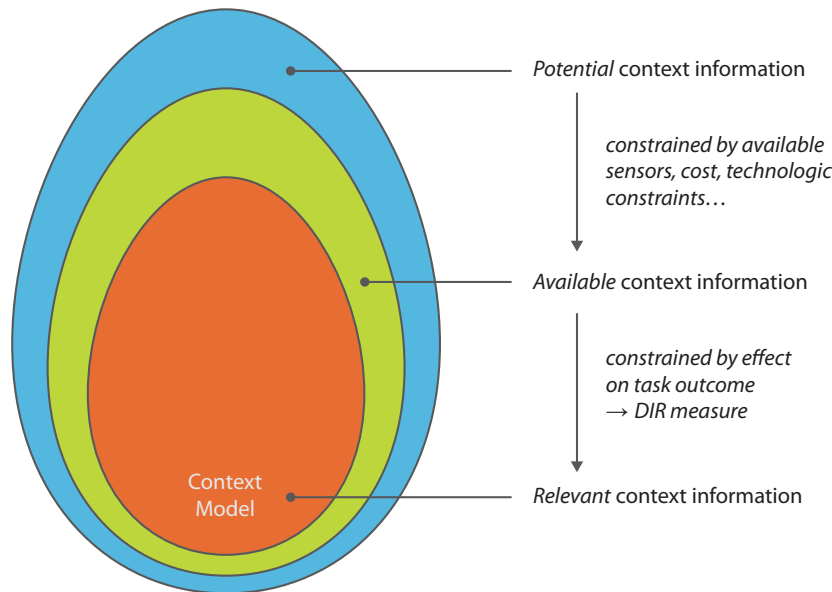
Figure 7: The specification of context information relevant for a given application is a three step process, consisting of listing any potentially important information (1, blue), determining available context information (2, green), and then identifying relevant context information using the DIR measure (3, orange).

the application. Otherwise, they can be neglected. This process can be automated for large numbers of queries (based on standard sensitivity analysis procedures; see below) and different constellations of context parameters to gain a detailed understanding of the influence of the different aspects of the context. In this way, such a measure would also support the development of simpler user interfaces for cognitive IR [177] by reducing the information shown to the user to relevant aspects of the task. This selection of contextual information to cover in a given application can be regarded as a three step process (see Figure 7). First, *potential* context information is determined by the developer, who needs a general idea of what contextual information may play a role for the application. Second, this potential context information is reduced to *available* context information. This pick is based on contraints imposed by available sensors, costs, the technology chosen for implementation, and more. Third, the selection of context information chosen for the implementation is further reduced to *relevant* information using the DIR measure.

At first glance, assessing how IR results respond to context changes looks like an application scenario for sensitivity analysis [164]. Sensitivity analysis comprises a number of techniques to trace back changes in the output of a mathematical model to variations in different parameters of the input. Typical application areas are business and environmental models. Methods from sensitivity analysis cannot be directly applied to the case of IR results, however, since they depend on a numeric model,

*Sensitivity analysis*

which is not given here since we only take the results and their order into account. Sensitivity analysis and DIR thus work on different scales of measurement [180]. Assigning a numeric value to these changes is what the measure introduced in this chapter is about. Hence, the *output* of the DIR measure can be used as input to a sensitivity analysis in a second step. Chapter 6 demonstrates how DIR can be used for a straightforward sensitivity analysis. Sensitivity analysis, however, is not per se concerned with cognitive plausibility. While one could imagine sensitivity analyses that inspect the effects of parameter changes on the cognitive plausibility of a model, this is not the prime motivation for a sensitivity analysis. In particular, the analysis itself is not meant to be cognitively plausible, which is a core aspect of this work.

The cognitively plausible Dissimilarity measure for Information Retrieval Results (DIR) introduced in this chapter analyzes the individual results in two rankings and compares them for overlap, focussing on the top results using a weighting mechanism to emphasize differences at the top ranks of the given result rankings. DIR returns values between 0 and 1, so that the calculated values are on an interval scale. A value of 0 indicates that two rankings are completely identical, and 1 indicates that they are completely different, i.e., they share no common results. Working solely on the results, the DIR measure can be applied independent of the actual retrieval method or search algorithm[2]. Rankings can be generated, for example, using the probability ranking principle [156] or novel approaches such as mean-variance analysis [192] or similarity-based retrieval [83].

## 3.2 DIR APPROACH

*The* Imp *measure*

The definition of the DIR measure is purely based on result rankings, independent of the applied knowledge representation and retrieval method, respectively. In contrast, the context impact measure Imp introduced in [94] aggregates the changes in ontological concepts under consideration during similarity measurement. The Imp measure can only be applied to logic-based knowledge representations, where context is formalized as rules modifying the knowledge base. Imp is defined for a knowledge base K and a specific query with search and target concepts ($c_s$ and $c_t$). The calculated value represents the share of the search- and target concepts' superconcepts (affected concepts $c_a$) that are changed by the modifying concepts $c_m$ which add ($+$) or remove ($-$) parts of the definitions in a specific context:

$$\text{Imp}(K, c_s, c_t) = \sum \frac{\pm c_m}{|\{c_a | c_a \sqsubseteq c_s \sqcup c_t\}|} \tag{3.1}$$

Besides its restriction to a specific kind of knowledge representation, such a result pair based measure is also of limited use since it does not take a relevance value's interpretation context [80] into account, i.e., its

---

2 In fact, the retrieval and ranking algorithm may even be unknown, as it is the case for the pizza example shown in Figure 2.

relativeness regarding other values: A relevance value of 0.8 may result in a top rank, but depending on the other results, it can also be topped by a large number of other results with relevance values $> 0.8$, leading to a lower rank. Moreover, while a single relevance value (or even all values in a ranking) may change, the order within the ranking may still remain the same. Finally, as discussed in [94], it stands to reason how the *Imp* measure translates to the results of a specific query, as the changes to a knowledge base may have different effects on the results depending on the applied retrieval method.

Accordingly, we propose an approach that looks at result rankings instead of modifications on the input. This ranking-driven approach also adheres to the cognitive aspect of IR, as the top of a ranking, presenting the best results for a specific query, is in the users' focus [2]. For DIR, this means that changes at the top of a ranking need to be emphasized by a higher weight, as opposed to changes that affect the lower ranks. The same applies to concepts that only appear in one of the two rankings at hand: they change the result rankings and disappear (and pop up, respectively) when going from one context to another. These cases are also weighted higher in DIR as they change the overall set of results. Both aspects are handled by the weighting mechanism introduced in Section 3.3.

*Ranking-driven approach*

DIR provides a tool that is applicable independent of the actual retrieval method, since the measure is based on the outcome of an information retrieval task in different contexts. Note that the measure is not intended to model – let alone explain – how human users compare two given rankings. This question touches upon a number of issues discussed in the cognitive sciences and is clearly out of scope for this research. Instead, the intention behind DIR is to reflect how human users *quantify* the differences between them. As discussed in Section 2.5, the focus of this research is on cognitive *plausibility* (see Definition 8) rather than cognitive *adequacy*.

Statistical rank correlation measures are not applicable in information retrieval, as discussed in Section 3.1. They map to the interval $[-1, 1]$, where a value of 1 indicates identical rankings, $-1$ indicates perfectly inverse rankings, and 0 indicates no statistical correlation. DIR takes a different approach, mapping to the interval from 0 for equal rankings to 1 for the extreme case where no result appears in both rankings. The normalization to the interval $[0, 1]$ allows for a comparison of DIR values independent of the rankings' lengths and their properties. DIR does not produce negative values, since inverse rankings are both hard to interpret and very unlikely to occur in the comparison of IR results. The most important aspect, however, is the lack of a weighting mechanism to stress differences at the top ranks. Consider the following example of three different rankings that demonstrate this issue:

```
1.  apple      apple      orange
2.  mouse      mouse      mouse
3.  tree       tree       tree
4.  boat       boat       boat
5.  goat       ape        ape
```

Both the left and the right ranking differ in only one result from the ranking in the middle, which leads to the same rank correlation of 0.7 based on Spearman's ρ for both pairs of rankings. From a cognitive IR perspective, however, the left two rankings are much more alike than the right two rankings: while the former two only differ in the last result, the latter differ in the top result which is in the focus of the IR task.

The presentation of the results to the user is an important aspect in human-computer information retrieval. Flat, list-style visualizations of IR results do not provide any information about the actual relevance of the shown results. These lists are the current standard, especially for Web search. Novel visualization techniques, such as tag clouds, give the user an impression of how good the results are. In particular, they allow the user to assess how much better result 1 is than result 2, for example. Figure 8 shows two visualizations of a result ranking from geographic information retrieval: on the left, the results are presented without any indication of the relevance of the results. The tag cloud on the right indicates the relevance of the different results and the differences between them by font size: the better the result, the bigger the font used. Other visualization techniques also fall into these two categories, such as lists with relevance values or groupings of results into predefined categories [83]. An assumption that suggests itself is that the visualization of information about the relevance of results also influences user judgments of changes in result rankings. In order to investigate whether this assumption holds, we have defined two different versions of the DIR measure: $\mathrm{DIR}_{rank}$ is solely based on the position of the results in a ranking (i.e., the order within the ranking) and is applied to list-style visualizations. In contrast, $\mathrm{DIR}_{rel}$ takes the relevance values of the individual results into account and is applied to result visualizations that allow the user to assess the single results' respective relevance.

*Evaluation based on result visualization*

1. Port
2. IrrigationCanal
3. Aqueduct, Canal
4. Dam
5. Sluice

# Port IrrigationCanal
## Aqueduct Canal Dam <sub>Sluice</sub>

Figure 8: Visualization of the same result ranking as a ranked list without indication of the relevance values (left) and as a tag cloud, where the results' relevance determines the font size (right).

Both variants of DIR are symmetric measures. They are therefore not sensitive to whether a query is first posed in context A and then in context B, or the other way around. In contrast, most semantic similarity measures [108, 86, 159, 83] are asymmetric. Tversky [186] has shown in a series of human participants tests that similarity assessments are strongly influenced by the prominence of the concepts to be compared. The less prominent concept was shown to be rated more similar to the more prominent one than vice versa; for example, North Korea was constantly rated more similar to China than vice versa. While

34

such effects of prominence or relative salience are also to be expected when comparing IR result rankings, it is not possible to take them into account without knowledge about the query and the result contents. We can therefore not take them into account on a generic level and specify DIR as a symmetric measure. The mathematical definitions of the two variants of DIR will be introduced in the following subsection.

## 3.3 FORMALIZATION

DIR is based on the comparison of two result rankings, where a ranking R consists of an ordered set of ranks. Each rank consists of a relevance value $v \in [0, 1]$ and a non-empty set of results $k$, where $v$ represents the relevance of the results $k$ (i.e., a rank groups all results of the same relevance). We assume that the ranks are in descending order with respect to the relevance values. Each rank is assigned an ascending rank number $n$, such that

$$
\begin{aligned}
R = \langle \{1, v_1, (k_i, \ldots, k_j)\}, \{2, v_2, (k_l, \ldots, k_p)\}, \ldots, \\
\{n, v_n, (k_q, \ldots, k_r)\} \rangle, \quad \text{where } v_1 > v_2 > \cdots > v_n.
\end{aligned}
\tag{3.2}
$$

DIR is based on the *shift* that every concept $k$ undergoes when a query is posed in two different contexts. This shift is calculated for every result that is part of either of the two rankings (or both). Based on the position in either ranking, individual weights are assigned to all shift values. The sum of all weighted shifts is then divided by the maximum possible dissimilarity $md$ for normalization. By slight abuse of notation, we denote the appearance of a result $k$ in a ranking $M$ as $k \in M$. For a pair of result rankings $M, N$, we define DIR as

$$
DIR(M, N) = \frac{\sum shift(k) * weight(k)}{md(M, N)}, k \in \{N \cup M\}.
\tag{3.3}
$$

As mentioned above, DIR is symmetric, i.e., DIR(M,N) = DIR(N,M). In the following, we define the $shift$, $weight$, and $md$ functions for the two variants $DIR_{rank}$ and $DIR_{rel}$.

THE SHIFT FUNCTION. The basic formalization of DIR shown in Equation 3.3 can be filled with different functions depending on whether relevance values should be taken into account or not. The *shift* function for a purely rank-based DIR is the difference in rank number between the two rankings. A special case occurs when the result appears in only one of the two rankings: in this case, the shift is the distance between its current position and the position "behind" the last rank of the longer ranking (see Figure 9). We therefore treat the result as if it drops out of

*Rank-based and relevance-based shift*
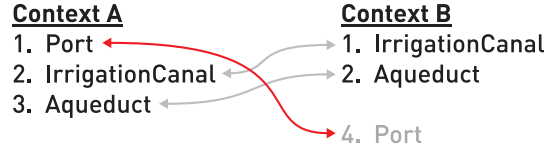
Figure 9: When a result appears in one ranking, but not in the other (such as *Port* in this example), it is treated by $shift_{rank}$ like it dropped out of the longer of the two rankings. The calculated values for this example are $shift_{rank}$ (*Port*) = 3, $shift_{rank}$ (*IrrigationCanal*) = 1, and $shift_{rank}$ (*Aqueduct*) = 1.

the longer ranking[3]. Let |M| be the number of ranks (not results) in M (and |N| the number of ranks in N), we define the rank-based *shift* as

$$shift_{rank}(k) = \begin{cases} |rank_k(M) - rank_k(N)| & \text{if } k \in M \wedge k \in N, \\ |M| - rank_k(M) + 1 & \text{if } k \notin N, \\ |M| - rank_k(N) + 1 & \text{if } k \notin M. \end{cases} \tag{3.4}$$

The relevance-based *shift* function takes the same approach, though based on relevance values instead of ranks. It takes the relevance differences into account. The special case where a result appears in only one of the two rankings does not require special handling: results that do not appear in a ranking have a relevance value $v$ of 0, so that the following equation covers all possible cases:

$$shift_{rel}(k) = |v_k(M) - v_k(N)| \tag{3.5}$$

THE WEIGHTING FUNCTION. The purpose of the weighting function is to stress shifts that affect the top of either ranking. DIR's weighting function, which is used both for $DIR_{rank}$ and $DIR_{rel}$, employs the rank number to determine the weight. One could think of the weights as inverse rank numbers, i.e. in a ranking M, rank no. 1 is weighted |M|, rank no. 2 is weighted $|M| - 1$, and so forth. The last rank is weighted 1. The maximum possible weight is determined by the longer of the two rankings. As for $shift_{rank}$, two different cases have to be distinguished: If a result appears in both rankings, the higher of the two calculated weights (the "inverse ranks") is used. This stresses changes at the top of the rankings and puts the weight in relation to the maximum possible weight for the given configuration (i.e., the weight for the results at rank 1 in the longer ranking). If the result appears in only one ranking, the total number of ranks of the longer ranking is used as the weight, so that the maximum weight applies. This stresses results that pop up or

---

3 Shifting a result "behind" its own ranking's last rank does not work here. For rankings with a big difference in length, this can produce situations where the shift of a result from the top of a short ranking to the end of a long ranking would exceed the $shift_{rank}$ value for dropping out, which is supposed to be the maximum possible shift.

| Context A | Context B | Weights | |
|-----------|-----------|---------|---|
| 1. Port | 1. IrrigationCanal | Port | 3 |
| 2. IrrigationCanal | 2. Aqueduct | IrrigationCanal | 3 |
| 3. Aqueduct | | Aqueduct | 2 |

Figure 10: Weights for the example shown in Figure 9.

disappear during a context change, as explained in Section 3.2. Figure 10 shows two sample rankings and the resulting weights. Assuming that M is the longer (or equal) ranking, i.e. $|M| \geqslant |N|$, the weighting function is defined as

$$
weight(k) = \begin{cases} 1 + max(|M| - rank_k(M), |M| - rank_k(N)) \\ \qquad\qquad\qquad \text{if } k \in M \wedge k \in N, \\ \\ |M| \qquad\qquad\qquad \text{if } k \notin M \vee k \notin N. \end{cases} \tag{3.6}
$$

DETERMINING THE MAXIMUM DISSIMILARITY. According to Equation 3.3, the sum of all single impact values, $\sum shift(k) * weight(k)$, is then divided by the maximum possible dissimilarity value for normalization. The maximum dissimilarity $md$ is reached when the two rankings are completely different. To calculate this value, we assume that this is the case for the given constellation of rankings with their specific relevance values and corresponding ranks. This maximum dissimilarity value depends on the shift function, therefore, the calculation of $md$ is different for the rank-based and the relevance-based versions of DIR. For the rank-based version, this means that every concept is shifted out of its ranking and weighted according to the number of ranks in the longer ranking. Correspondingly, just like the shift function, the maximum possible impact depends on the lengths of the two rankings, and on the number of concepts at every rank. Let $k \in r_m$ be the concepts at rank $m$, $|k \in r_m|$ the number of concepts at rank $m$, and given $|M| \geqslant |N|$, the maximum impact for the rank-based DIR is

*Rank-based and relevance-based maximum dissimilarity*

$$
md_{rank}(M, N) = |M| * ( \sum_{m=1}^{|M|} (|M| + 1 - m) * |k \in r_m| + \\ \sum_{n=1}^{|N|} (|M| + 1 - n) * |k \in r_n|). \tag{3.7}
$$

For the relevance-based DIR, this straight-forward calculation based on rank numbers is not possible. Instead, the relevance value at a given rank must be taken into account. Accordingly, $md_{rel}$ additionally depends on the exact relevance values associated with the ranks. Let

$v_m$ be the relevance value at rank $m$, the maximum impact for the relevance-based DIR is

$$md_{rel}(M, N) = |M| * \left( \sum_{m=1}^{|M|} v_m * |k \in r_m| + \sum_{n=1}^{|N|} v_n * |k \in r_n| \right). \quad (3.8)$$

The difference between $md_{rank}$ and $md_{rel}$ is therefore in the calculation of the maximum shift per result: while $md_{rank}$ pushes every result out of the longer ranking ($|M| + 1 - m$), $md_{rel}$ drops the relevance values of all results to 0, so that the shift is $v_m$ for every result. Both versions of $md$ provide the maximum possible impact value for a given configuration of two relevance rankings. It is thus made sure that whenever two completely different result rankings are passed to DIR, it will return the value 1.

# 4

## HUMAN PARTICIPANTS TEST

To make sure that DIR measurements reflect the users' impression of changes in IR result rankings, the measure's cognitive plausibility has been evaluated in two human participants tests. This chapter is thus concerned with testing the hypothesis phrased in Section 1.3. In the first test (Section 4.1), the participants were to compare two randomly generated result rankings to a reference ranking and rate which of them differed more from the reference ranking. In the second test (Section 4.2), the task was to rate the difference between two randomly generated rankings by positioning a slider between the two extremes *indistinguishable* and *no commonalities*. For both tests, the participants' judgments were statistically evaluated[1]. The results are discussed in Section 4.3.

### 4.1 TEST ONE: JUDGMENT BY ORDER

The first human participants test (referred to as test one in the following) consisted in the task to compare two result rankings to a third reference ranking. The participants were students at the University of California, Santa Barbara.

TEST SETUP.     Test one was a computer based test. The screen presented to the participants is shown in Figure 11. The participants had to rate which of the bottom two rankings differs more from the reference ranking at the top. Three options were given to choose from: one could either select the left or right ranking to be more different from the reference ranking, or alternatively choose "cannot decide". The first test therefore focussed on the *order* of the different DIR values and its correlation with the order of the participant judgments [163, 84].

Though designed for completion on the Web, the test was not publicly accessible[2]. Login information was only provided for registered participants, who were shown an introductory page with an example to clarify the task. In some cases, the test was completed in a university lab; however, no additional instructions were provided. It was thus made sure that every participant received the same information. Moreover, written instructions have been shown to be preferred by participants over spoken instructions [67].

Both the list style and the tag cloud visualization techniques were tested and every participant was shown all tasks using only one of them. The result rankings contained existing English words. For every

---

1 The data collected during the two human participants tests are available for download at `http://ifgi.uni-muenster.de/~kessler/phd/hpt-data.zip`

2 After the study, the first human participants test has been made publicly available at `http://v-simcat.uni-muenster.de:8080/UCSB-Test/`.
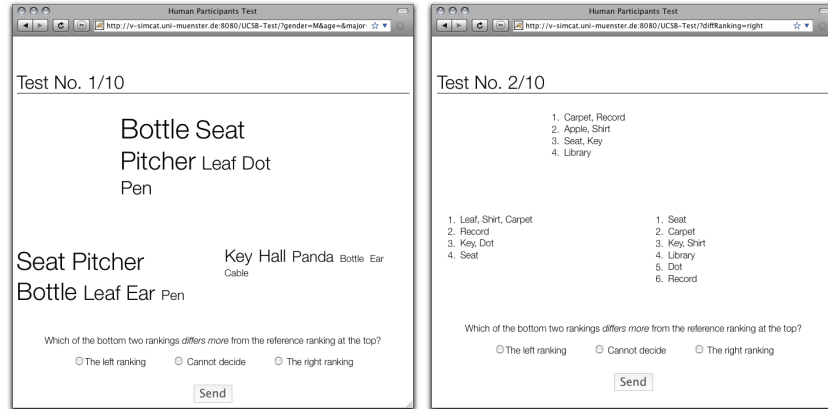
Figure 11: Screen shot of single comparison tasks in test one for the relevance-based (left) and rank-based (right) DIR variant.

participant, ten comparison tasks were randomly generated so that the ten difference[3] values (*diff*) were equally distributed between 0 and 1: for every randomly generated triplet, $DIR_L$ =DIR(Left, Reference) and $DIR_R$ =DIR(Right, Reference) were calculated, so that *diff* = $|DIR_L - DIR_R|$. The *diff* values for the 10 tests shown to every participant were distributed randomly between 0 and 1. The assumption for the test was that the bigger the *diff* value, the easier it should be for a participant to decide, reflected in higher numbers of judgments in line with DIR. Moreover, it was assumed that tasks with a bigger *diff* value are solved in less time than those with smaller *diff* values, i.e., that there is a negative correlation between the time required per task and the *diff* value for the task. The time required to complete a single comparison task was logged to evaluate whether this assumption holds.

EVALUATION.    The participants in test one consisted of a group of 52 undergraduate students of the Geography 5 class at the University of California, Santa Barbara, who received credit points for participation. The participants were between age 17 and 27, with a mean age of 18.5. The group consisted of 37 female and 14 male[4] participants majoring in different fields, most of them still undecided about their major.

The evaluation of test one was carried out on the complete collection of all 52 participant datasets. Figure 12 gives an overview of the collected data and shows that the cases where the participant judgements coincide with the calculated *diff* value based on DIR form the large majority. Nonetheless, some participants largely disagree with the calculated values, such as participant 02 in the list visualization, or participant 05 in the tag cloud visualization. Out of the 520 comparison tasks completed by the participants, 341 (~ 66%) were judged in com-

---

3 The DIR values for the evaluations of the two tests were calculated using a Java implementation of the measure which is available from `http://ifgi.uni-muenster.de/~kessler/phd/DIR.zip`.

4 One participant did not provide any personal information.
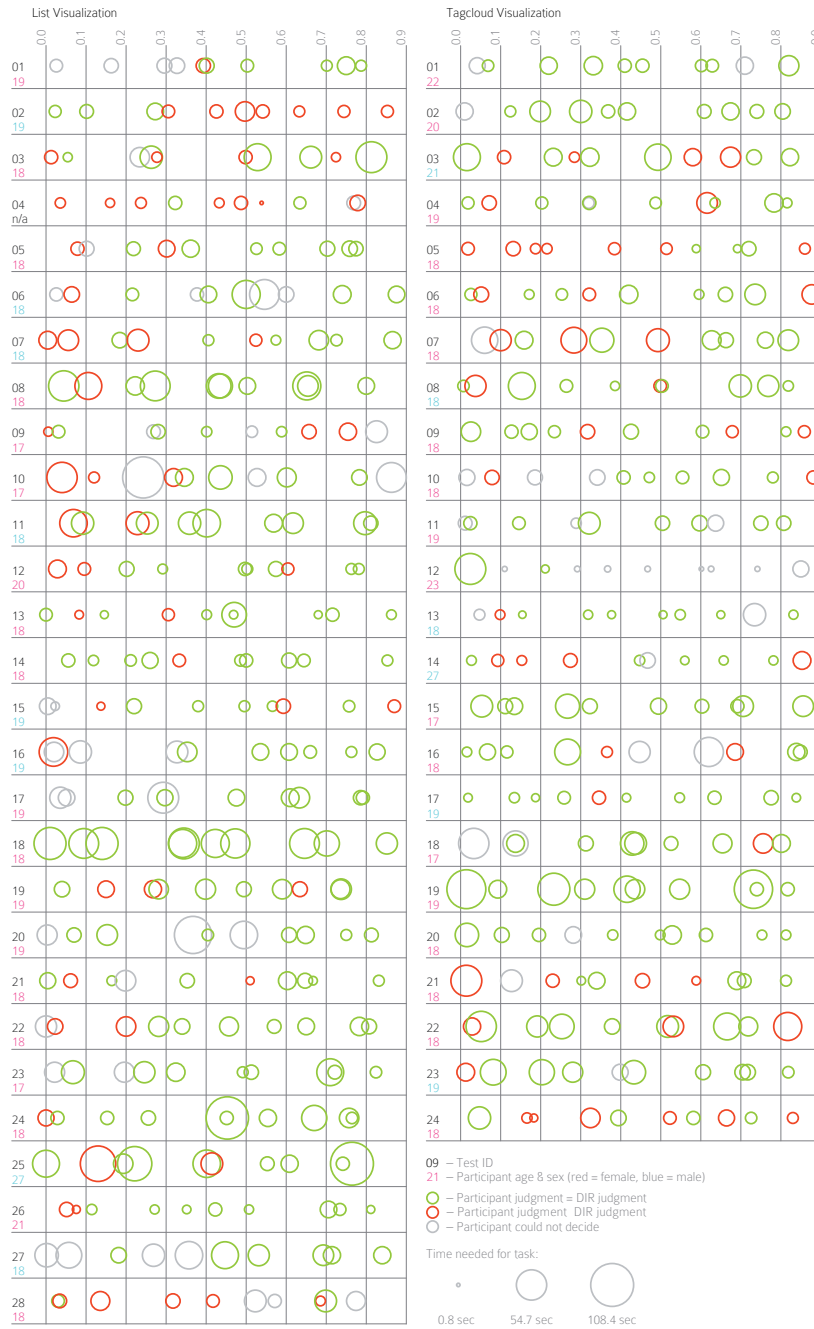
Figure 12: Results for the 52 participants in test one. Every row in the two columns represents one participant, with one circle per comparison task. The circle's color indicates whether the user judgment matches DIR, the circle area reflects the time the participant needed to answer. The circles' positions in the row reflect the respective *diff* values [93].

pliance with DIR, 111 (~ 21%) were judged inconsistent with DIR, and in 68 cases (~ 13%), the participants could not decide.

Figure 13 shows the distribution of "correctly" (i.e., in line with DIR), "falsely" (contradicting DIR) and undecided tasks per user. The bottom histogram shows that only 5 participants (~ 10%) judged more than 4 tasks contradicting DIR; for comparison, 22 participants (~ 42%) judged only 1 task contradicting DIR. The top left histogram shows that 37 participants (~ 71%) judged more than half of the tasks according to DIR. These numbers indicate a strong agreement of the participants with the DIR measure, which is supported by the results of a test for correlation. For this test, all 520 tasks completed by the participants were split into 20 groups of 26 tasks each. For each group, the mean *diff* value and the number of participant judgments that were in line with DIR were computed. Since the resulting ranks are not equidistant, the correlation between mean *diff* value and number of 'correctly' solved tasks was tested using Kendall's τ. The test yields a positive correlation of τ = 0.602. The result is significant at the 1% level (p= 0.000029). Therefore, the null hypothesis of no correlation can be falsified, and we can accept the alternative hypothesis that higher differences between the DIR values lead to a higher recognition rate among the participants.

The second hypothesis to test was that a higher *diff* value in a specific task also allows the participants to solve it faster. Looking for a negative linear correlation between the respective *diff* values and logged completion times, Pearson's product-moment correlation was calculated for the 520 single tasks. The test yields a slight negative correlation of ρ = −0.11, which is significant at the 5% level (p= 0.0123). Accordingly, the *diff* values have a comparably small (yet statistically significant) effect on the time required to solve a task.

The development of the two different versions of DIR followed the assumption that user judgements depend on whether the result visualization includes hints on inter-rank distances or not. If this is the

*Relevance-based versus rank-based DIR*

case, the correct choice between $DIR_{rel}$ and $DIR_{rank}$ with respect to the presentation of the results should increase the correlation between user judgements and the DIR calculations. To test this hypothesis, the two correlation values were also calculated with a dataset where the *diff* values for all tasks carried out by the participants were computed using $DIR_{rank}$. The DIR values were thus only based on the order of the results, independent of whether the participants were shown tag clouds containing information on inter-rank distance. In this setting, Kendall's τ yields a positive correlation of τ = 0.613, significant at the 1% level (p= 0.00020). The correlation is thus even slightly stronger when DIR is calculated solely based on ranks, which suggests that $DIR_{rank}$ is a plausible measure for perceived changes in result rankings, independent of result visualization. Accordingly, inter-rank distance seems to play a secondary role for users, who seem to focus on the order of the results. The correlation between *diff* values and completion times remains largely unchanged by the limitation to $DIR_{rank}$.
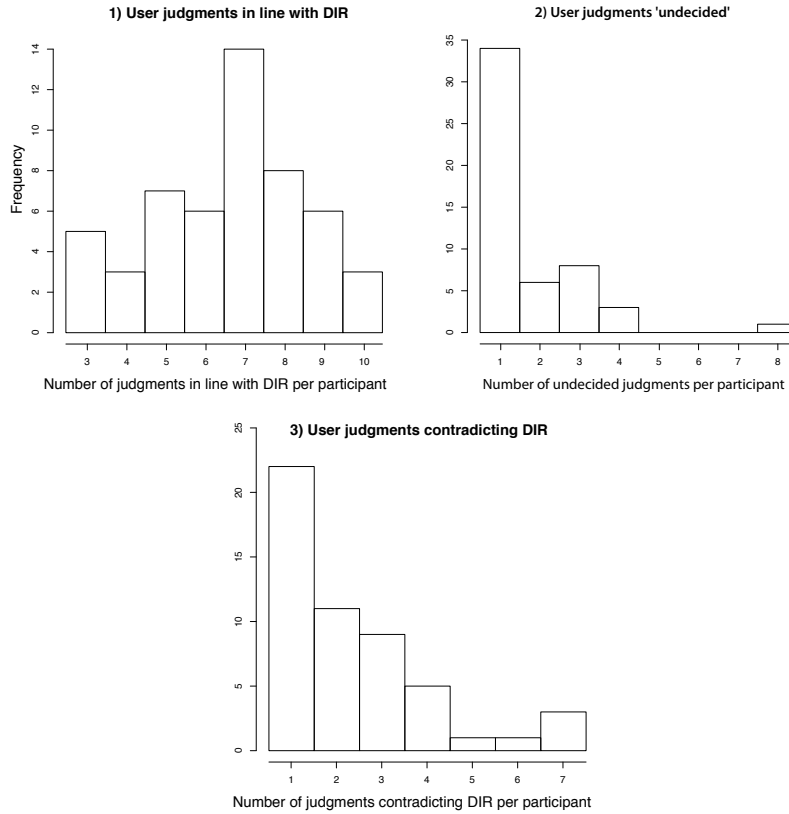
Figure 13: Histograms of the number of tasks solved in line with DIR (1), unde-
cided tasks (2) and tasks solved contradicting DIR (3) per participant.
In the left histogram, for example, the rightmost bar shows that 3
participants had solved all 10 tasks in line with DIR [93].

## 4.2 TEST TWO: JUDGMENT BY NUMBERS

The second test (referred to as test two below) focused on a direct
comparison of DIR values. It was run as an open online test where
everyone could participate.

TEST SETUP. Test two was also designed as a Web-based test[5], yet
with an open call. As in test one, participants were shown randomly
generated result rankings visualized either as tag clouds or lists (see
Figure 14). The objective of this test was to investigate whether the
DIR measure does not only correlate with participant judgments in
terms of order, but also in terms of the actual values. The participants'
task consisted in comparing pairs of result rankings and answering the
question "how much does the left result set differ from the one on the
right in your opinion?". To answer this question, the participants had to
move a slider ranging from "indistinguishable" to "no commonalities".

---

5 The test is available online at http://v-simcat.uni-muenster.de:8080/OnlineCIMTest/
UserTest?language=en.
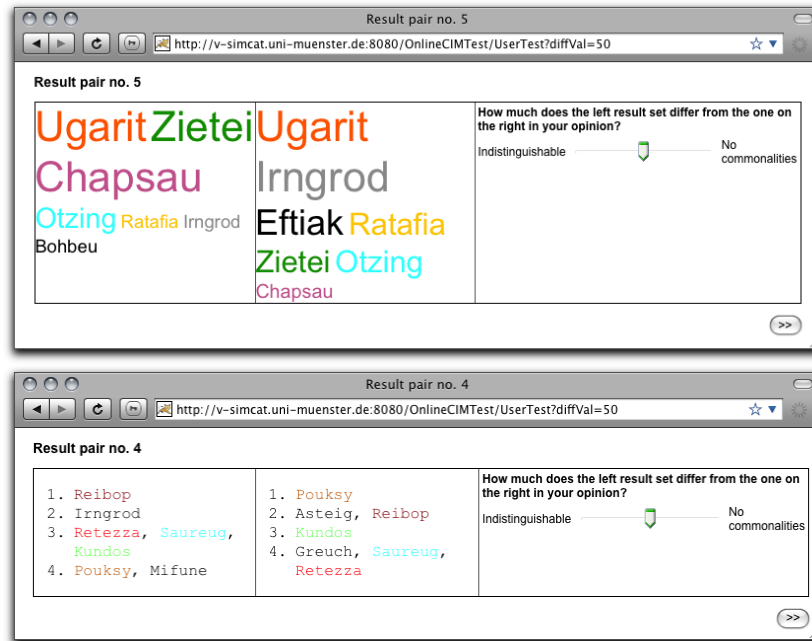
Figure 14: Screen shots of single comparison tasks in test one for the relevance-based (top) and the rank-based (bottom) DIR variants [93].

The position of the slider was internally mapped to values in the interval $[0, 1]$.

The test was available in English and German to increase the number of potential participants. It was explained with an example and some background information on an introductory Web page. The participants were specifically advised that there are no right or wrong answers in the test to make sure that they adjust the slider according to their personal opinion. In addition to the slider position, the time every participant spent on every pair of result rankings until he or she committed the judgment was logged. Each ranking of test data was stored along with the computer's IP address to make sure that every participant takes the test only once.

As in test one, every participant was presented with all comparison tasks using the same visualization technique – either list-based or as a tag cloud – to avoid confusion. The result rankings were limited to seven entries on either side to prevent cognitive overload, as a high number of results in the two rankings would have made it very hard for the participants to thoroughly compare the rankings based on the single results. As opposed to test one, the results in the rankings consisted of made-up words, so that the participants could not perform any undesirable reasoning on the result rankings, such as "*result ranking A contains 'tree' and result ranking B contains 'apple', therefore...*". Moreover, these made up words guaranteed that the German and the English speaking participants could be shown the same test data. The entries were also color-coded to facilitate recognition of identical entries in the

*Made-up words and color-coding*

44

two result rankings. Entries appearing in both rankings were displayed in the same color, whereas entries shown only either left or right were displayed in black.

The seven pairs of result rankings were arranged according to pre-defined ranges using DIR. Every participant was shown two extreme cases, one with completely identical rankings (DIR = 0) and one with completely different rankings (no common entries, DIR = 1). These two cases were included to make sure that every participant had read and understood the instructions. The remaining five were generated so that the corresponding DIR values were within five equal intervals between 0 and 1 (i.e., $[0.0, 0.2[, [0.2, 0.4[ \ldots [0.8, 1.0])$. The seven tests were randomly generated for every new participant and presented in random order.

EVALUATION. The 81 participants who took the test were anonymous volunteers recruited by distributing an announcement on several mailing lists. As no compensation was granted for participation, the test was intentionally kept very short to avoid participants from getting bored and breaking off in the middle of the test. No personal data such as age, sex, or occupation were collected, as the reliability of such data in an open, Web-based human participants test is per se very limited. It is very likely that the majority are adults with a college education background, given that the call was distributed on several geography, geoinformatics and computer science mailing lists.

Out of the 81 participants who took test two, 42 were shown the rank-based test and 39 were shown the relevance-based test. Only tests where the participants had correctly understood the instructions were taken into account for the statistical analysis: if one of the two extreme cases contained in each test (DIR = 0 or DIR = 1) were judged more than 0.25 "off" (user judgment $> 0.25$ and $< 0.75$, respectively), these tests were not used any further. After this filtering, 31 rank-based and 25 relevance-based tests were used for evaluation. For this collection of 56 tests, the participants spent a mean time of 4 min, 16 sec on the completion (including instructions) and a mean time of 2 min, 14 sec on reading the instructions. It it thus safe to assume that the tests used for evaluation were completed thoroughly and that the according participants had read the instructions and understood the task.

Figure 15 shows bubble plots of DIR values against the values selected by the participants using the slider. The bubble size reflects the time spent on the respective task. The distribution of the bubbles, each representing the result of one task completed by a participant, already gives an impression of the high variance of the user judgments compared to the calculated DIR values (see also Figure 16). Nonetheless, the regression line in Figure 15 shows that the calculated DIR values generally follow the same trend as the participant values.

The correlation between the user judgments and computed DIR values for test two can be calculated directly on the values entered by the users; it is not necessary to group the tasks, as in the evaluation of test one. When both visualization types are analyzed with the respective
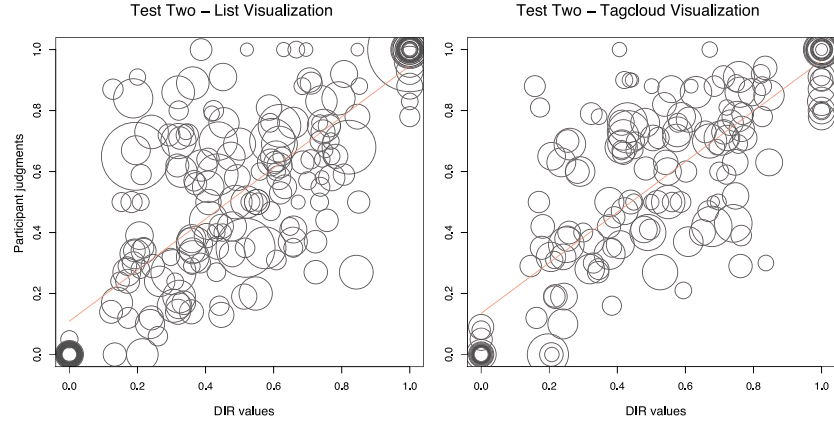
Figure 15: Bubble plots with regression lines of the calculated DIR values against the participant judgments for list visualization (left, with $\text{DIR}_{rank}$) and tag cloud visualization (right, with $\text{DIR}_{rel}$) in test two. The bubble size indicates the logged completion time for the respective task [93].
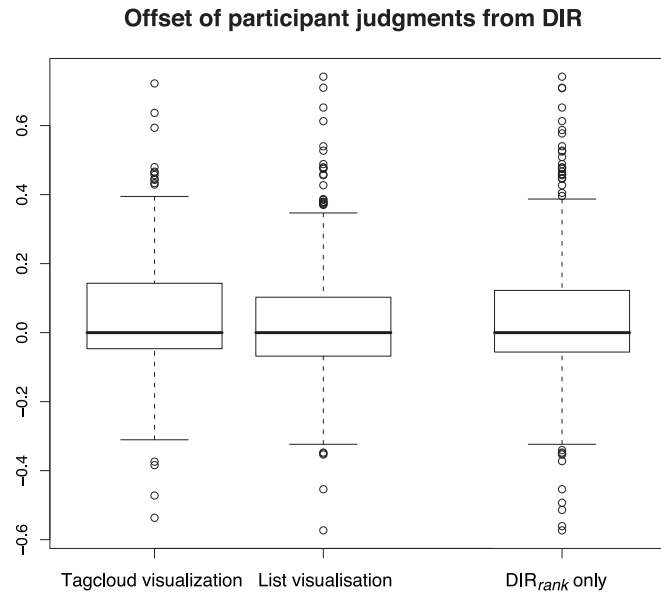


Figure 16: Box plots of the two variants of test two, showing the deviation of the participant judgments from the computed DIR values. The left box shows both visualization types with their corresponding DIR variants, the right box shows all 392 tasks solved by the participants compared to $\text{DIR}_{rank}$ [93].

DIR variant, Pearson's product-moment correlation yields a positive correlation of 0.805, significant at the 1% level.

In this constellation, the logged completion time correlates slightly with the respective DIR values at 0.106, significant at the 5% level. Interestingly, the users thus needed slightly longer, the more different (according to DIR) the two shown rankings were.

For test two, we also analyzed the effect of the choice of DIR method on the correlation. If all results are recalculated using only $DIR_{rank}$, the correlation between user judgments and DIR values decreases slightly to 0.803 (significant at the 1% level). The correlation between DIR values and completion times does not change notably either, dropping to 0.102.

## 4.3  RESULTS & DISCUSSION

Two human participants tests were carried out to test the hypothesis that dissimilarity measurements for information retrieval results calculated using the DIR method correlate strongly with human dissimilarity judgments (see Section 1.3). The evaluation of the test data leads to the result that this hypothesis cannot be falsified. In both cases, the human judgments correlate with the DIR values, significant at the 1% level.

The most striking outcome of the evaluation is the fact that the purely rank-based variant of DIR correlates equally strong with the participant judgments. Apparently, users do not take the relevance values of the results into account when they compare two rankings. While the information that, for example, result 1 is twice as good as result 2, is certainly valuable from an objective point of view, users do not seem to distinguish this from a case where result 1 is only slightly better than result 2. The outcome of the two human participants tests thus confirms the approach taken by all popular Web search engines in presenting their results to the users as flat lists. At the same time, it renders $DIR_{rel}$ superflous, at least for the cases studied in this research. This may change when applying the measure to real IR use cases, where users typically only skim through the results and do not take such a close look as the participants in these two studies.

*Ignorance of relevance information*

The evaluation of both human participants tests revealed a strong correlation between the calculated DIR values and the respective participant judgments. Completion time, however, does not seem to be influenced by how different (according to DIR) the presented result rankings are. A slight negative correlation could be observed in test one, showing that the more different two pairs of result rankings were, the faster the tasks were solved. The evaluation of test two did not support this finding, though. Since in both cases, the correlation was very weak, we have to assume that there is no connection between DIR and processing times – at least at the level of number of results presented in this study. A stronger correlation may turn out with higher numbers of results presented to the participants. Previous research indicates that the small number of results presented to the participants were still within a range that can easily be processed [129]. It thus remains to be shown whether this finding also holds when the participants are

shown large numbers of results. Moreover, this outcome may have been influenced by the setting of the study, where all participants looked thoroughly at the tasks before submitting their judgments; to further investigate this correlation, observing users in solving actual information retrieval tasks seems to be a promising approach.

For both tests presented in this chapter, the participants can be assumed to be experienced Web users who are familiar with IR systems, at least in the form of Web search engines. Test one was even completed by a group of participants consisting only of *digital natives* who have grown up with the Web. We cannot deduce from the test data whether these findings transfer to other user groups (and to what degree). However, with the growing number of Internet users, it is very likely that these results apply to an increasing fraction of the population that can be described as experienced Web users. Due to sample sizes and participant selection, it was not possible to test any effects of the participants' age or gender. Whether these play a role remains as an open question for future research.

Chapter 3 has introduced a novel method to compute the dissimilarity of information retrieval result rankings. Its cognitive plausibility was verified in two human participants tests in this chapter. While this method can be applied to *identify* relevant contextual aspects that influence the outcome of an IR task, it does not solve the problem of *modeling* and managing context in an IR application. In the following chapter, we introduce a novel method for context representation based on Semantic Web standards. Its usefulness in application development is shown in Chapter 6, along with a demonstration of DIR as a selector for relevant context information.

# A CONTEXT MODEL FOR SEMANTICS-BASED INFORMATION RETRIEVAL

This chapter describes an approach for a generic implementation of a context model for information retrieval from semantically annotated information using the Semantic Web Rule Language. Based on the requirements analysis in Section 5.1, we develop a user model (Section 5.2) that allows personalized retrieval. In Section 5.3, we show how a novel SWRL buit-in allows the retrieval process to be adapted to the current conditions including information from real-world sensors. Theoretical and practical implications of the approach are discussed in Section 5.4.

## 5.1 REQUIREMENTS ANALYSIS: CHARACTERISTICS OF A CONTEXT MODEL

The process of requirements analysis is the standard approach to generate a solid foundation for large-scale commercial software products. According to Maciaszek [120] – whose methodology we will follow in this section – requirements analysis is specifically critical for information systems. He distinguishes three levels of management that information systems are developed for: operational, tactical and strategic. Knowledge processing systems, which also include the ontologies and the corresponding management tools discussed in this thesis, fall into the category of information systems for the strategic management layer. In many cases, such systems require detailed representation of business processes. Developing knowledge processing systems is especially demanding because of the potentially large number of users that may have very different expectations concerning the system's functionality. The success of a software development project inherently depends on whether the developed system works the way users expect it to work. This aspect also concerns the cognitive engineering background of this work (see Section 2.5).

*Knowledge processing systems*

While the full software development lifecycle covers the phases of analysis, design, implementation, integration, deployment, operation and maintenance [120, p.26], we will restrict the considerations here to the analysis phase and touch upon design and implementation, as we limit ourselves to a high-level description of a generic implementation approach in this chapter. In the following, we will go through the phases of requirements determination, requirements specification, as well as design and architecture. Throughout this chapter, we will describe the general process of developing a context-aware semantics-based tool for information retrieval, thus treating the context model as a setup consisting of several software components. Existing standards and open source solutions will be used where possible.

REQUIREMENTS DETERMINATION. The goal of the requirements determination phase is to analyze and document the hierarchy of (business) processes underlying the application to be developed. Figure 17 shows an overview of the context-aware IR process from ontologies in the Business Process Modeling Notation[1] (BPMN). From a user's perspective, the retrieval process splits down to four subprocesses. Management of the knowledge base is a general prerequisite which makes sure that a stable conceptualization and information annotated with this conceptualization is at hand. This process may or may not be part of the functionality accessible to the user; an implementation for end users should try to hide most of this complexity from the user. Management of context dependencies based on input from sensors and the user herself enables context-aware retrieval based on the aspects selected using the DIR measure. This process requires user input, at least at the level of creating a user profile. Query formalisation and result interpretation (including visualization) are not further analyzed here, since they fall into the category of user interface issues that are too application specific to be handled at a generic level[2]. In the following, we detail the system services that are required to build a software representation of the processes underlying context-aware, semantics-based information retrieval. Moreover, we discuss constraints imposed on the implementation.
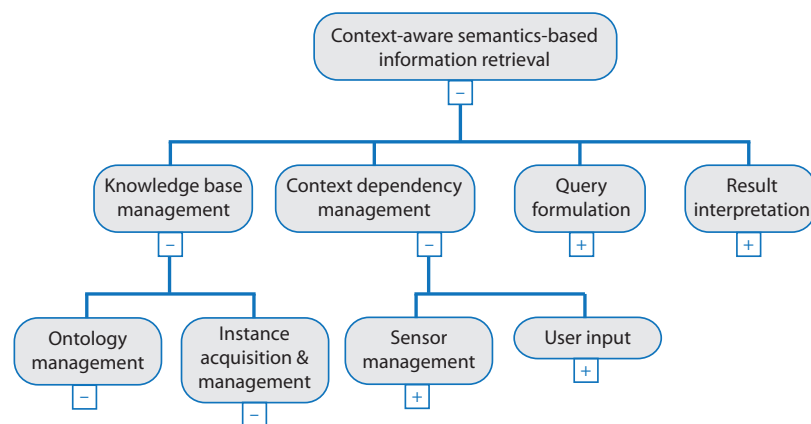
Figure 17: Process hierarchy for the retrieval task. Processes marked with a − indicate atomic processes, while processes whose subprocesses are not shown are marked with a +.

*System services*    The core functionality of the system is to allow the user to pose queries to a knowledge base (i.e., a populated ontology) that take the current context into account when they are processed. In particular, the system must be able to (a) adapt the results to the user's preferences and (b) adapt the results to external contextual information provided by sensors. Moreover, the system allows the user to change her profile and thus the way the system reacts to context changes. While the

---

1 See http://omg.org/spec/BPMN/1.2/.

2 The user interface of the surf spot finder will be discussed in Section 6.2.

system adheres to existing standards of the Semantic Web to maximize interoperability and the number of potential input sources, the applied retrieval algorithm must be easy to replace. The component of the application that enables the context-awareness must therefore act as a transparent proxy between the IR method and the knowledge base. Figure 18 shows the relationship between this context model, the knowledge base, and the retrieval method.
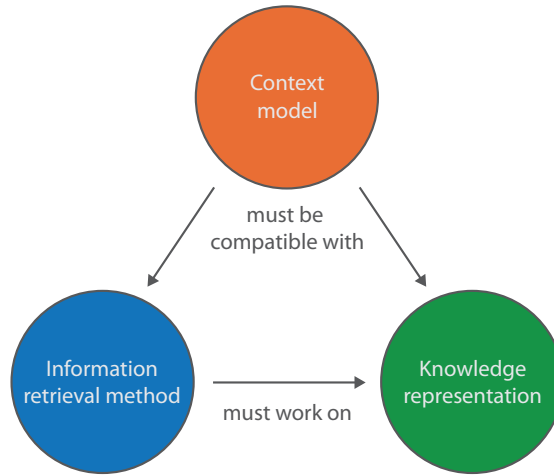


Figure 18: The information retrieval method must always be applicable to the language chosen for knowledge representation. Likewise, the context model must be able to interoperate with the knowledge representation and make its information available for the retrieval method. In this thesis, we stick to the languages of the Semantic Web to ensure this interoperability between the three components.

The constraints for the system implementation are imposed by the Semantic Web standards it should adhere to. The Resource Description Framework (RDF) [123] and the Web Ontology Language (OWL) [189] are the most important standards to mention here (see also Section 2.3). The system component handling context awareness must therefore be compatible with knowledge bases expressed in OWL. The acquisition of information from sensors is the second important area where existing standards need to be considered. The Open Geospatial Consortium[3] (OGC) has developed a number of standards under the umbrella of the Sensor Web Enablement (SWE) [26] initiative that are implemented for a rapidly increasing number of sensors around the world. In order to maximize the number of potential input sensors for the system, the implementation must also be compliant with the OGC SWE standards.

*System constraints*

REQUIREMENTS SPECIFICATION. The aim of the requirements specification process is to determine system states and changes between them, as well as the system's behavior. Figure 19 shows the five use

---

3 See http://opengeospatial.org.
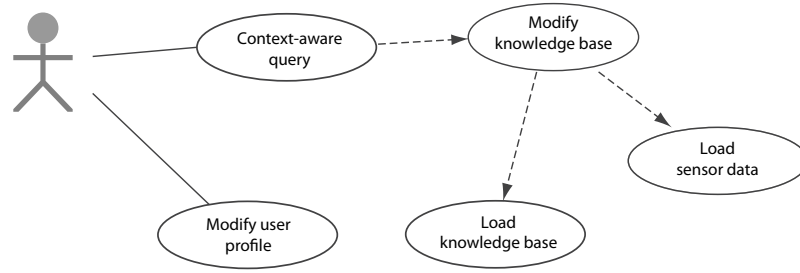
Figure 19: Use cases within the context-aware semantics-based IR system.

cases covered by the system, where the *context aware query* and *modify profile* use cases are triggered by the user:

1. *Context aware query*. The main use case from the user's perspective is posing a context aware query, where the results returned by the system are adapted to her user profile and the current conditions measured by sensors. To complete this task, the system internally triggers the *Modify knowledge base* use case.

2. *Modify profile*. In order to make sure that the user's profile always reflects her preferences (which may change over time), the system must allow her to edit the profile.

3. *Modify knowledge base*. This use case is performed by the system to adapt the knowledge base according to the user profile. The approach will be detailed in Section 5.2. It invokes the *load knowledge base* and *load sensor data* use cases.

4. *Load knowledge base*. Before the knowledge base can be adapted according to the user preferences, it must be loaded into the system.

5. *Load sensor data*. The adaptation of the knowledge base is based on a combination of the user preferences and the current real-world conditions provided by sensors. Access to the sensor data is detailed below.

The different use cases have been transformed into a sequence diagram in Figure 20 that shows the order in which the different steps are completed, and which components are involved at which step of the task.

DESIGN AND ARCHITECTURE. The requirements determination and specification phases provided an abstract view on the system to be developed. They concentrated on the identification of (business) processes the system needs to represent, and one of the components that are required to realize this system. The development of design and architecture, in contrast, provides the foundation for the actual implementation of the system. The deployment diagram in Figure 21 shows an overview of the components of the system and their interaction.
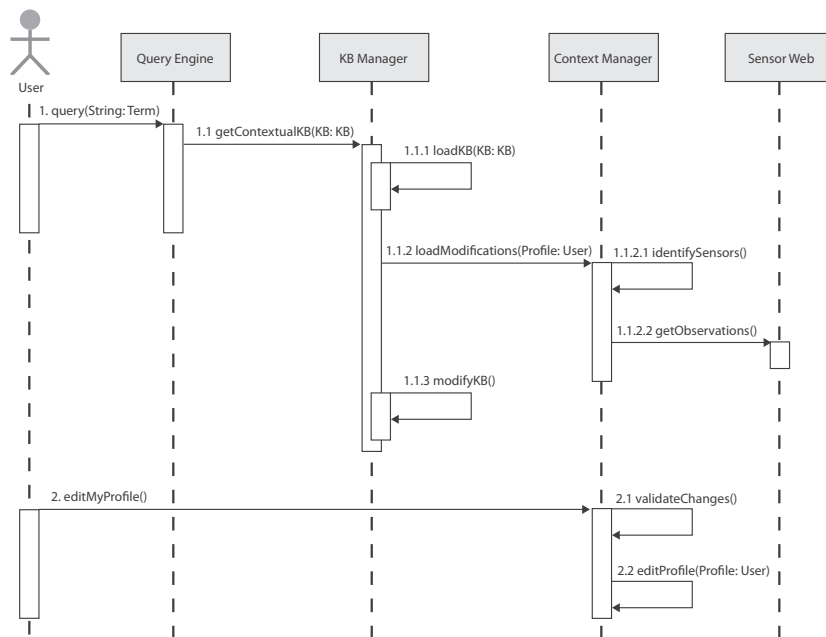
Figure 20: Sequence diagram for the two interaction options the user has with the system: pose context aware queries (1) and edit her profile (2).
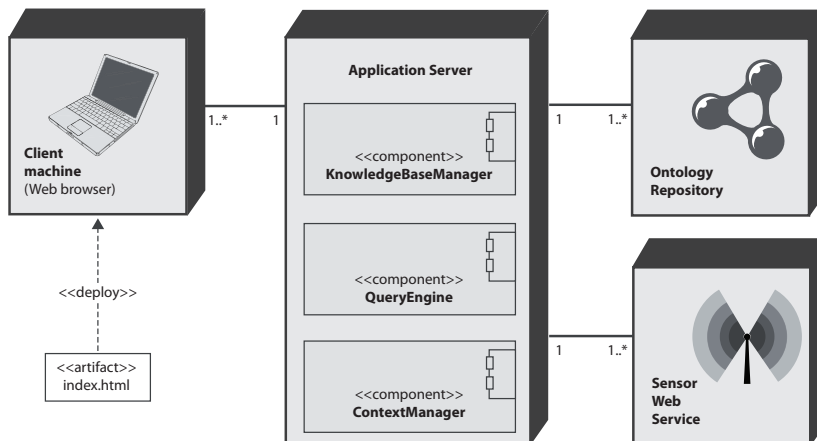


Figure 21: Deployment diagram for the system, making use of existing ontology repositories and sensor web services.

We assume that the client system runs in a Web browser and is therefore independent of any specific prerequisites on the user side concerning a specific operating system or software that needs to be installed. The main components to be developed are found on the application server side. As mentioned before, the goal is to reuse existing software wherever possible. We will use the Protégé ontology editor[4] as the main component for knowledge base management. Protégé is an open-source knowledge base framework that supports the development of OWL ontologies with a graphical user interface. Moreover, the community has developed variants of Protégé which support collaborative editing of an ontology. This is realized either by connecting the Protégé application to the Protégé server [184], or via the Web Protégé [185] interface that runs in a Web browser. Besides these different variants focusing on different user interfaces, Protégé's API can also be accessed directly from third party code, which is the approach we take here for the application server.

*Protégé*

We have picked Protégé as the component for knowledge base management because of its wide user base and the broad range of available plugins. One such extension to the core Protégé functionality is the integration of plugins for editing and executing rules defined in the Semantic Web Rule Language. In the following, we will demonstrate how SWRL – and particularly its support for custom built-ins – can be used to implement a system for context-aware semantics-based IR as specified above.

### 5.2 USER MODEL

Context-aware information retrieval requires an interpretation of the processed information with respect to the user. It depends on the user's personal preferences whether a shop has interesting offers, a blog article is worth reading or a surf spot is adequate with respect to the user's expectations and abilities. OWL-based ontologies on the Semantic Web, however, lack the processing and mapping capabilities that are required to express such dependencies on the user profile. It is not possible to specify that a user finds surf spots with rocky bottoms too dangerous, or that another one finds waves below 6ft unattractive. This problem becomes even more evident when the application incorporates Linked Data [20] from other sources around the Web, which cannot (and should not) be changed based on a user's personal preferences. We are thus looking for a workflow for the context management component (see Figure 21) that enables *temporary* changes to a personalized copy of the ontology based on the users profile. The problem at hand is thus to find a mechanism that allows for personalized mappings between the numeric sensor world and information stored in ontologies, as shown in Figure 22. In the following, we describe how semantic rules can be applied for this task. A similar – yet more high level – approach based on rules has been put forward by Roman et al. [160], who introduced

---

4 See `http://protege.stanford.edu`.
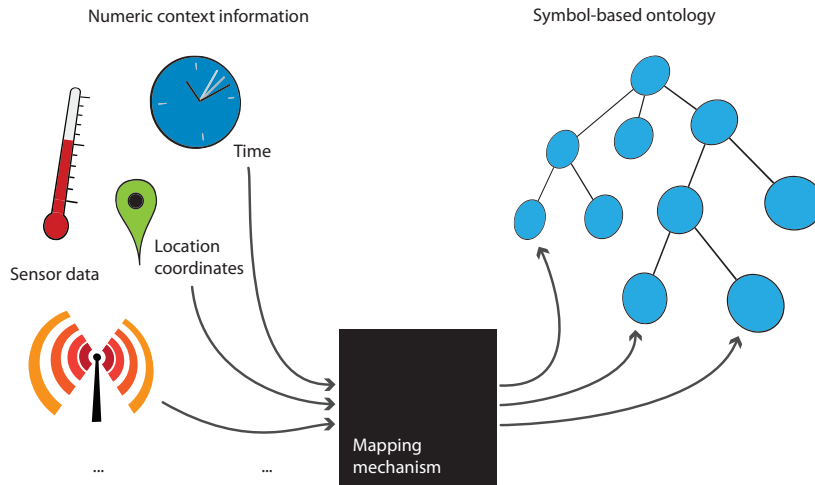
Figure 22: Context-aware geographical information retrieval on the Semantic Web requires a mapping from the context model (left) to the static ontological information (right) [95].

a formal context model where the rules specify the different context variables' behavior.

We have selected the Semantic Web Rule Language (SWRL) [76] for the implementation. As discussed in Section 2.3, SWRL extends OWL by Horn-like clauses [73] that enable n-ary relationships. Existing built-ins as well as the possibility to develop custom built-ins for specific reasoning tasks are the main reason to pick SWRL. To the best of our knowledge, there exists no standardized rule language that extends OWL with additional inference mechanisms. At the time of writing of this thesis, SWRL has the status of a W3C submission and is therefore designed for integration with existing W3C Semantic Web standards, though not a standard by itself yet.

SWRL rules are part of the ontology. They are instances of the `swrl:Imp` class, which defines implication rules in the SWRL ontology[5]. Accordingly, the knowledge base management and context management components of the implementation (see Figure 21) have to be tightly integrated. Two of the plugins available for the Protégé ontology editor are the SWRL Tab and the Jess Tab[6]. While the former is an editor for SWRL rules with basic syntax checking, the latter allows the user to transfer the knowledge base and the rules to the Jess rule engine[7]. The rules can then be executed and inferred knowledge can be transferred back into the knowledge base, such as reclassifications of individuals. While Jess itself is not open source software, it is the only rule engine

*Rule engines*

---

5 See `http://w3.org/Submission/SWRL/swrl.owl`.
6 See `http://protege.cim3.net/cgi-bin/wiki.pl?SWRLTab` and
  `http://protege.cim3.net/cgi-bin/wiki.pl?SWRLJessTab`.
7 See `http://jessrules.com`.

that has been integrated into Protégé with full support for custom built-ins (discussed below). We thus picked the combination of Protégé and Jess, using one of the free licenses that can be obtained for academic use. It can be expected that more (and open source) rule engines will be integrated with Protégé in the future to provide a broader selection for the execution of SWRL rules. This setup needs to be wrapped as an HTTP server to be made accessible for the browser-based user interface; details on a specific implementation of this approach will be provided in Chapter 6. The following section demonstrates how SWRL rules using built-ins can be utilized to represent a user model as shown in Figure 23.
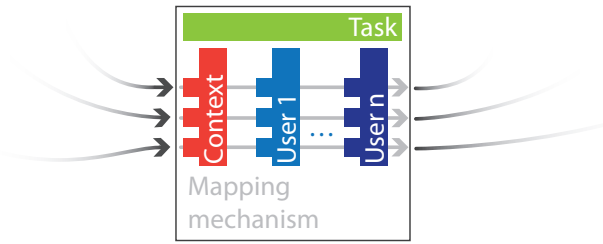


Figure 23: The contents of the mapping mechanism black box shown in Figure 22: The mapping consists of task model (green, not further specified), context model (red) and user models (blue). The context model provides access to dynamically changing information from sensors as described in Section 5.3, which is then fed into the user models. The user models, each consisting of a SWRL rule, filter the incoming context information based to the users' preferences [95].

USER MODEL FORMALIZATION IN SWRL. A user model can be represented as a set of preferences specified by the user (or learned by an adaptive user interface, which is out of scope for this research). We will first discuss the straight forward case where the user specifies concrete constraints in her profile, and how these can be implemented in the antecedent of a SWRL rule. This body of a rule can be used to represent a conjunction of the different aspects of a user's preferences. We illustrate this approach with an example from the Surf Spot Finder application outlined in Section 1.5. Consider the following rule which specifies that a certain `SurfSpot` is reclassified as an `AppropriateSurfSpot` if it has a `SandyBottom` (for users afraid of potential dangers):

```
01    SurfSpot(?spot) ∧
02    hasBottom(?spot, ?bottom) ∧
03    SandyBottom(?bottom)
04    → AppropriateSurfSpot(?spot)
```

*Reasoning about classes and relations*

When executing this rule, the rule engine reasons about the individuals in the ontology and first picks all individuals of the class `SurfSpot` (line one). Those are then restricted to individuals with the relation `hasBottom` (line two), which is filled with an individual from the class `SandyBottom` (line three). For all individuals where lines one to three

are evaluated as true, the head of the rule applies which reclassifies them as individuals of the `AppropriateSurfSpot` class (a subclass of `SurfSpot` in the ontology). Evidently, this approach requires that the classes `SurfSpot`, `SandyBottom` and `AppropriateSurfSpot`, as well as the relation `hasBottom` are defined in the ontology.

While the first example is restricted to reasoning about the individuals' types and relations between them, the standard built-ins for comparisons and mathematical calculations[8] can be applied to reason about the individual's datatype property values. The following example specifies the user preference for wave heights over 2.5 meters (the question where the information on the current wave height comes from is discussed in Section 5.3):

```
01    SurfSpot(?spot) ∧
02    hasWaveHeight(?spot, ?height) ∧
03    swrlb:greaterThan(?height, 2.5) ∧
04    → AppropriateSurfSpot(?spot)
```

As these two examples show, user preferences can be easily represented by a combination of restrictions about the individuals, types, relations, and datatype property values. The rules can be further extended by SWRL's date and time built-ins to enable spatio-temporal reasoning, for example when looking for shops within a certain range that are currently open. Combinations of such rules – where each rule represents the set of one user's preferences – can be applied to enable SWRL-based group decision support [78]. The individuals that match the combination of all users' preferences would therefore also accommodate the preferences of the whole group. Obviously, this would require the collection of rules to be free of conflicts. The combination of the previous wave height example with the atom `swrlb:lessThan(?height,2)` would obviously not bear any results.

While we have simply reclassified matching individuals in the examples above, one could also imagine a selection of those individuals via SQWRL[9], a library of built-ins that extend SWRL with SQL-like querying mechanisms. Both approaches have the drawback that they do not provide ranked results: individuals in the knowledge base either match the given criteria, or they do not. In order to use SWRL in a way that is closer to other information retrieval approaches, we propose a novel built-in that allows querying by example for the individuals' datatype property values. For example, it would be more intuitive for users if they could state that their preferred price for lunch is 6$, and the retrieval component would return a ranked list of lunch options, ordered by ascending price difference. This functionality is provided by the `ir:rank` built-in, which takes a datatype property value, compares it against the values of the individuals under considerations, and calculates a ranked list (see Section 3.3) of these individuals. Merging such a ranked list back into the knowledge base does not make sense (and would be hard to implement), since it merely represents the query

---

8 See `http://www.daml.org/2004/04/swrl/builtins.html`.
9 See `http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL`.

results for a single user's preferences. The outcome must therefore be passed to the query engine component (see Figure 21), which may include it in further processing, such as combinations with other result lists.

The implementation of custom built-ins is realized via Protégé's SWRL Built-in Bridge[10]. From the ontological perspective, new built-ins must be individuals of the class `BuiltIn`, which is defined in the SWRL ontology. This ontology has to be imported to enable the definition of rules. The processing functionality that evaluates the single built-ins must be written in Java as a class called `SWRLBuiltInLibraryImpl`. It must override some standard methods required for built-in processing and is dynamically loaded by the bridge.

A concrete example with user profiles represented in SWRL and an application using the `ir:rank` built-in will be introduced in Chapter 6. In the following, we will show how the built-in mechanism can be used to make retrieval from ontologies context aware by including measurements from the Sensor Web.

## 5.3 LINKING TO REAL-WORLD SENSORS

While we have demonstrated how SWRL rules can be employed as user models, there is still a missing link to measurements from sensors out in the real world. So far, dynamic values still have to be managed and updated manually for the individuals in a knowledge base. This is not only cumbersome, it also contradicts the idea of context-aware computing. In this section, we outline an approach to this problem that enables querying of sensors from within SWRL rules.

SWRL built-ins bear the potential to link OWL ontologies to external information sources, such as the SQWRL built-ins mentioned above. We adopt this approach for (geo)sensors [179]. In the following, we describe how to collect observations from the Sensor Web from a SWRL built-in. The presented approach allows for dynamic data type values for specific instances in an otherwise static ontology. It can be regarded as a more light-weight variant of the Semantic Sensor Web [171, 71], which builds on fully annotated sensor web services. Other related visions are the Semantic Reality [68] and the Semantic Geospatial Web [46].

The Open Geospatial Consortium has introduced a number of service specifications to enable a fully interoperable Sensor Web. One of these standards is the Sensor Observation Service (SOS) [133], which allows clients to request observations for a specific sensor via the *Sensor Web built-in* `GetObservation` request. Following the idea of "SQWRL for sensors", we have developed the `sos:getLastObservation` built-in[11]. To keep the approach as lightweight as possible, we only model the SOS' core properties:

---

10 See `http://protege.cim3.net/cgi-bin/wiki.pl?SWRLBuiltInBridge`.

11 The SWRL built-ins developed for this thesis are available from the 52° North Semantics Community at `http://52north.org/semantics`.

1. `ServiceURL`: The HTTP address under which the SOS can be found online.

2. `ObservedProperty`: This property describes the phenomenon that is observed. The corresponding definitions are provided by phenomenon dictionaries that assign unique URNs, such as `urn:ogc:def:phenomenon:OGC:1.0.30:temperature`.

3. `ObservationOffering`: A grouping of observations that is assigned a keyword for identification. The observations are often grouped by phenomenon at a more general level, such as `WeatherBerlin` or `WaterQualityRhine`, or by region or time interval.

4. `FeatureOfInterest`: An abstract feature in the sense of OGC's Geography Markup Language (GML) [144]. This can be a named feature such as `FOI_LakeConstance`, specified by the services' capabilities document, and it must have the `ObservedProperty` as its property.

We use the feature of interest in our built-in as the link between the basic sensor ontology[12] and the concepts in the corresponding application ontologies. For example, for the surf spot finder, the features of interest of the Sensor Observation Services modeled in the ontology are instances of the class `SurfSpot`. The corresponding workflow of a SWRL-enabled OWL-ontology using the `sos:getLastObservation` built-in is shown in Figure 24.

Using `sos:getLastObservation` thus requires the ontology to contain an individual representing the SOS to be queried. This instance is an individual of the class that can be generated from the SOS XML schema[13], both in its current basic form and in future more comprehensive versions. The usage pattern for the built-in binds the value observed for the given `?foi`, `?offering` and property `?urn` of an `?sos` individual to `?obsValue`:

`sos:getLastObservation(?obsValue, ?sos, ?foi, ?urn, ?offering)`

With this built-in, it is possible to remove any dynamically changing information from the ontology. It then only contains static information about the individuals as well as the sensors measuring the dynamic aspects, which are linked to the central concepts of the application ontology as features of interest. The dynamic context information is collected on demand, i.e., when the rules representing the user model are executed. The actual communication with the Sensor Observation Services is internally handled by 52 North's OX Framework[14]. This *OX Framework* Open Web Service Access Framework is a collection of open source Java libraries for accessing different kinds of OGC Web Services that integrates well with the Protégé- and Jess-based solution proposed in this chapter.

---

12 The ontologies mentioned in this chapter are available from `http://sim-dl.sf.net/downloads/`.

13 See `http://schemas.opengis.net/sos/1.0.0/sosGetCapabilities.xsd`.

14 See `http://52north.org/oxf`.
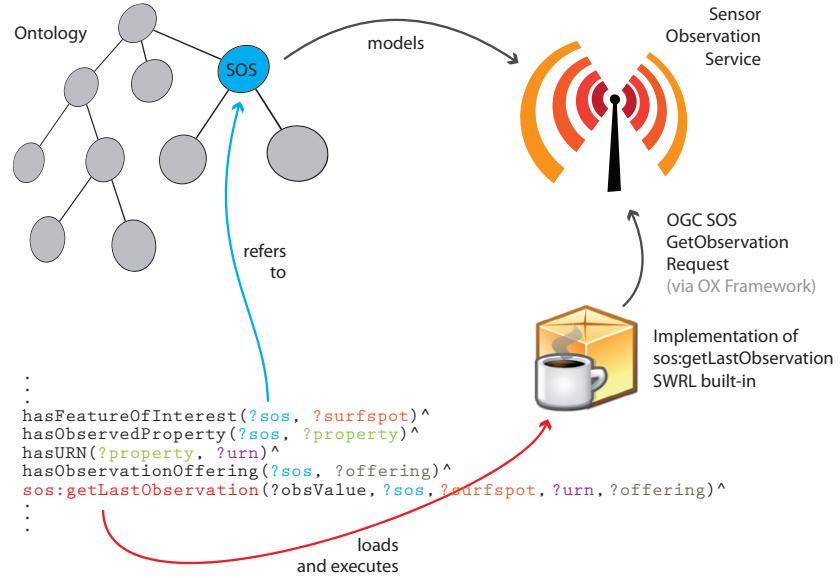
Figure 24: Workflow of the `sos:getLastObservation` built-in: An SOS has to be modeled in the ontology, so that it can be referred to from SWRL. During rule execution, the implementation of the built-in is dynamically loaded and executed. It queries the SOS and returns the value to the rule engine for further processing (adapted from [95]).

## 5.4 DISCUSSION

As mentioned before, widespread Semantic Web technologies lack support for context-aware mappings required for a threefold model consisting of task, user and context as shown in Figure 23. More specifically, the implementation of such a model needs (1) rules (*if...then...*) and (2) free variables, as required for mathematical processing capabilities. This applies both to RDF/OWL (including C-OWL; see Section 2.4) as well as to alternative approaches such as the ISO Topic Map standard [18]. Although these languages have not been developed to support such functionality initially, these restrictions place constraints on their applicability to real-world problems. Their underlying realist view on semantics [52] ignores the fact that a large number of applications in information retrieval, linguistics or location-based services, for example, can only produce useful results if they take the user's current context into account. Even the context extension to OWL, C-OWL, is limited to fixed contexts (i.e., local micro theories) and mappings between them. It can thus not live up to the requirements of the notion of *dynamic context awareness* [125] from AI, which heavily relies on sensor input. The same applies to Topic Maps, which allow users to define *scopes* specifying in which context assertions are valid [142]. Scopes form a mechanism that is very useful when stating assertions such as the topic name *München* is only valid in the scope of the German language or the topic name *Jaguar* appears in different scopes (animal, car, operating system). However, same as for C-OWL, scopes neither support rules

nor free variables. Overall, these limitations result in the rare use of semantic technologies in context-aware software in AI, which is still largely based on statistics-driven approaches.

As demonstrated in this chapter, SWRL helps to overcome these limitations by specifying user models as rules that contextualize OWL ontologies in terms of these models. This new functionality comes with a number of implications, though. First, modeling in OWL and SWRL requires a thoughtful separation between the notions modeled in OWL and those modeled in SWRL. Depending on the sensor measurements imported via `sos:getLastObservation`, the same reasoning steps may lead to different results. Hence, monotonicity is no longer guaranteed. The ontology engineer must already have a clear idea of the rules that will interact with the application ontology at the very beginning of the design phase. Positively speaking, the presented approach forces clean ontology engineering, as the static part (put into the OWL ontology) has to be strictly separated from the dynamic part (formalized in the SWRL rules). This separation of static and dynamic content becomes especially important when decidability plays a role: since the integration of SWRL renders OWL undecidable [75], it may be desirable to retain the static part of the ontology for reasoning purposes.

*Monotonicity and decidability*

Concerning SWRL, the number of tools for rule engineering and execution is still very limited. In practice, the rules quickly become verbose, even with the prefix syntax used in this thesis. The XML-syntax used for storage is even more extensive, so that neither of the two formats is adequate for interaction of non-experts. These end users, such as the surfers in our scenario, cannot be expected to edit rules by hand. Instead, user interfaces that hide the complexity of the rules from the user are required. This problem will be addressed in the next chapter.

# 6

PROTOTYPE IMPLEMENTATION: THE SURF SPOT FINDER

This chapter demonstrates the approaches for identification of relevant contextual information and SWRL-based context modeling using the surf spot finder application outlined in Section 1.5. Section 6.1 describes how the DIR measure was applied to distinguish relevant from non-relevant contextual information (see Chapters 3 and 4). Section 6.2 describes the user interface, and Section 6.3 outlines how the contextual aspects identified in Section 6.1 were implemented in the server component.

## 6.1 IDENTIFICATION OF RELEVANT CONTEXT INFORMATION

In order to demonstrate the feasibility of the DIR approach for the development of context-aware applications, we have applied it for the selection of the contextual information covered in the surf spot finder prototype. As mentioned in Section 1.5, it will be a mobile application – more specifically, a Web application optimized for the iPhone, since it is one of the first devices that enable access to the user's position via GPS from a Web browser[1]. Besides access to GPS, the implementation as a Web application running in the browser is especially attractive since it does not require the user to install software. Moreover, optimization for other devices is easy to realize. Due to the limited screen size on mobile devices (480 by 320 pixels in case of the iPhone), the focus on the most relevant context parameters becomes even more important to provide an easy-to use, uncluttered user interface. The screen size for the planned application offers sufficient space for five drop-down menues, where the user can select her preferences (see Section 6.2). In the following, we discuss how DIR has been applied to identify the five most important contextual aspects to be included in the Web application.

*Web applications*

The surfing website `http://magicseaweed.com` offers an extensive archive of swell- and wind data for numerous spots around the world. These archives have been used to enable a realistic simulation of the variability of the different environmental aspects that influence the conditions at a surf spot. The service offers comprehensive overviews of daily average minimum and maximum swell heights and periods, as well as the monthly wind direction and speed distribution. The average highs, lows, and the distributions have been accumulated from several years of data. Moreover, information on the wave breaking direction and predominant kind of bottom (sandy, rocky or reef) were collected from `http://wannasurf.com` for every spot. Since neither of the two

*Surf spot data*

---

1 This functionality is described in the draft specification of the W3C Geolocation API [191].

surfing websites offer any archive data on air and water temperature, monthly average lows and highs for both temperatures were used for Santa Barbara[2] and the Santa Barbara Channel[3], respectively. Those data were selected because of the central location of Santa Barbara with respect to the surf spots selected for the study. Table 1 shows sample data for one of the twenty spots along the central and southern California coast that were taken into account.

| Property | Value/range |
| --- | --- |
| Location | lat: 34.4, lon: −119.84 |
| Bottom | Rock |
| Wave breaking direction | Right |
| Swell height | 3–7ft |
| Swell period | 5–16sec |
| Wind direction | NNW-NWW |
| Wind speed | 10–20mph |
| Water temperature | 13–21°C |
| Air temperature | 15–25°C |

Table 1: Sample data for the surf spot Campus Point. The weather data are based on a snapshot for July 1st. The data on wind speed and direction actually consist of detailed monthly distributions; the table shows only the predominant speed and direction for the given month.

Several services on the Web offer access to the current weather- and water conditions. The DIR-based simulations as described in Chapter 3, however, require a complete overview of the conditions to expect throughout the whole year. Otherwise, it is not possible to realistically simulate the variance of the different contextual parameters that are available for the application. The archive data described above thus offer a solid foundation for a realistic simulation with DIR.

*User profile properties*

In order to evaluate the effect of the different context parameters on the outcome of a retrieval task, a set of typical values for every aspect of a user profile has been created. The aspects of the user profiles correspond to the different properties of the surf spots, representing the temporal, spatial and thematic components that characterize geographic information [60]. Table 2 shows an overview of these values. The third column indicates knock out criteria: if any of these criteria are part of the user profile, only spots that exactly match this preference will be part of the results. This is the case for aspects where a ranking is not feasible because the variable's scale is nominal [180]. Moreover, the maximum distance the user is willing to travel is a knock out criterion. All spots within this range are ranked by distance from the user's location[4]. For all other parameters, the values specify a preference, and the spots

---

2 Taken from `http://en.wikipedia.org/wiki/Santa_Barbara,_California`.
3 Taken from `http://www.santabarbarachannelswim.org/46053_sea_temperature.pdf`.
4 Future versions of the surf spot finder, however, could also take account of spots outside this range if they match the user preferences well in most other aspects.

are ranked by deviation from this preferred value. For the ranking returned to the user, the mean rank is calculated for every surf spot, based on the ranks for the single parameters. More sophisticated user interfaces could also allow the user to assign weights to the different aspects as proposed by Rinner and Raubal [154] for location-based services; however, this functionality has been omitted for the prototype implementation to keep the user interface as simple as possible[5]. Any of the parameters of the user profile can also be empty, indicating that the user does not have a preference for this aspect. The only compulsory parameter is the user location: if no further parameters are present, the spots are ranked by distance.

| Context variable | Values | K.o. |
|---|---|---|
| User location | Santa Barbara, Santa Cruz, Pacifica, San Luis Obispo, Oxnard | |
| Max. distance | 50, 100, 200 miles | √ |
| Water temperature | 16, 18, 20, 22°C | |
| Air temperature | 14, 19, 22, 25, 28°C | |
| Swell height | 5, 10, 15 ft. | |
| Swell period | 3, 5, 10, 20 sec. | |
| Wind direction | N, E, S, W | |
| Wind speed | 10, 20, 30 mph | |
| Breaking direction | Left, right, both | √ |
| Bottom | Sand, rock, reef | √ |

Table 2: Different values used for the sample user profiles. The different locations were represented by the respective lat/lon values in the simulation.

For the simulation, 1000 random combinations of the values shown in Table 2 were generated. For each of these profiles, 12 random dates were picked (one per month on average) to reflect the weather conditions at different times of the year. Based on each of these daily weather data snapshots, a ranking was generated for each user profile. First, every aspect of the user profile was compared against the corresponding value of the weather data, leading to a ranking of the spots for this specific aspect. The resulting collection of rankings was then merged into a single ranking based on the mean ranks for every spot. The user profile was stored along with this ranking. In the next step, all $12,000^2$ combinations of two of the rankings were compared using DIR. However, only such combinations were taken into account where the user profile differed in only one aspect[6]. This way, it is possible to

*Simulation approach*

---

5 Implementing the weighting functionality would also require an extension of the DIR-based simulation approach, so that it includes different weights for the contextual aspects.
6 This approach known as *One At a Time* (OAT) in sensitivity analysis may be problematic; see the corresponding discussion in Section 7.1. Nonetheless, we chose to follow this approach here to demonstrate the use of DIR, since the focus of this research is not on sensitivity analysis.

analyze how changes in one of the contextual aspects affect the results, while all other aspects remain constant. Figure 25 shows box plots of the set of DIR values that were produced for every context parameter. Based on the findings from the human participants tests in Chapter 4, only $DIR_{rank}$ was applied.
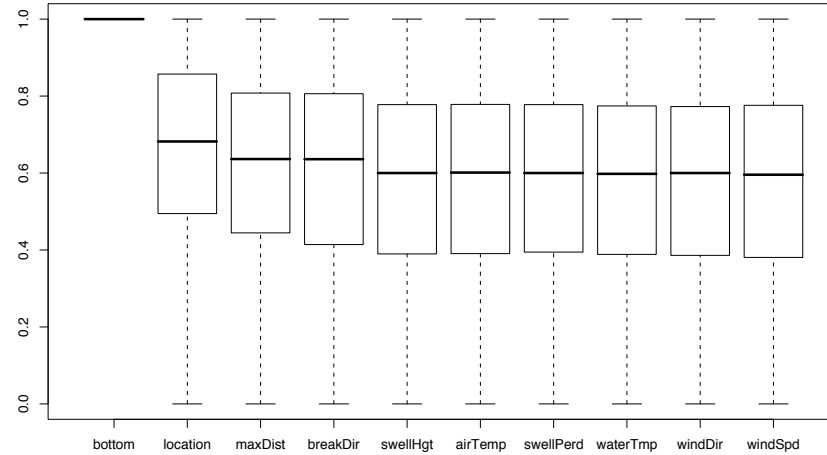


Figure 25: Box plots of the $DIR_{rank}$ values per context parameter.

According to this analysis, the knock out criterion *bottom* has the largest influence on the outcome. This is straightforward, since a change of preference in this aspect automatically excludes all surf spots that matched the previous preferences. The DIR value is therefore always 1 when the user changes the preference for this aspect. For the second and third most influential context aspects, DIR has identified the *location* and the *maximum distance* the user is willing to travel. These two aspects require only one input field for the maximum distance in the user interface, since the location is automatically set via GPS. According to DIR, the fourth most influential context parameter is the *breaking direction*. While the same outcome as for *bottom* (DIR= 1 for all cases) could be expected here, spots that are characterized by waves breaking in *both* directions are also shown to users who selected *left* or *right* as their preference. Vice versa, if the user selects *both*, spots with waves breaking *left* or *right* are also included in the results. Hence, there is still an overlap in the results, even when this knock out criterion changes.

For the remaining six context parameters, there are only marginal differences in the mean DIR values, which range around 0.57. Additional information is therefore required to pick the remaining two context aspects for the user interface. We have based this selection on the study by Wosniok [198], who evaluated 150 questionnaires filled in by occasional surfers. We picked the two aspects that have been rated most important by the participants of the study. We have left out aspects that have already been selected for the prototype (*distance*, *dangers* – which refers to the kind of bottom) and aspects that we do not have any data for (*accessibility*, *crowd size*, *friends going there*, *rideable wave length*). This

*Combination with questionnaire-based methods*

leaves the *swell height* (called *wave height* in Wosniok's study, ranked no. 1) and *wind speed* (ranked no. 7) for the prototype [198, p.27].

Based on the analysis in this section, the contextual aspects to cover in the prototype include *bottom*, *maximum distance* the user is willing to travel, *wave breaking direction*, *swell height* and *wind speed*. Additionally, the user's current location is set via GPS. While DIR was useful in selecting the most influential parameters, the evaluation of the measure in a development process has also shown that DIR alone is not always sufficient to fully evaluate the contextual aspects to cover in an application. In case of the surf spot finder, six out of the ten contextual aspects had an equally large influence on the results. We could therefore only identify three[7] of the five most important aspects (that the user interface offers space for) based on DIR only. The remaining two aspects had to be identified by other means, for which an existing questionnaire-based study was picked [198]. In many cases, it will therefore still be required to evaluate the user needs – which is a crucial step in professional application development in any case [120]. DIR thus adds a new perspective to the requirements analysis that is based on the contextual variances influencing an application. This data-driven view needs to be combined with other analysis methods to get a comprehensive overview of the requirements. This section has shown how DIR can be used together with a questionnaire-based analysis method to develop a solution that suits both the users' needs and is tailored to the specific contextual dependencies of an information retrieval task. In the following, we discuss the implementation based on this selection of contextual aspects, both for the client and the server side.

## 6.2 USER INTERFACE

The implementation of the user interface of the surf spot finder Web application is based on jQTouch, a plugin for the popular jQuery[8] JavaScript framework. The plugin adds support for multitouch gestures and allows for the straight forward development of Web applications that mimic the iPhone OS user interface elements. Users can therefore be presented with familiar interface elements and interaction patterns.

*jQuery and jQTouch*

Figure 26 gives an overview of the prototype's user interface and shows the work flow. After agreeing to let the Web application access the user's current position, she can set her preferences for the five different aspects. In case she has used the Surf Spot Finder before, her previously selected preferences are automatically loaded based on a cookie stored in her browser. Any change in one of the parameters automatically updates the spot map, which shows all spots that match the user's preferences, along with her current location. The user can hence switch back and forth between the two views (preferences and spot map) to see how her changes affect the set of matching surf spots.

---

7 One of the four most influential aspects was the location, which is not covered in the user interface since it is measured via GPS.

8 Both jQuery and the jQTouch plugin are free and open source; see `http://jqtouch.com/` and `http://jquery.com/`.
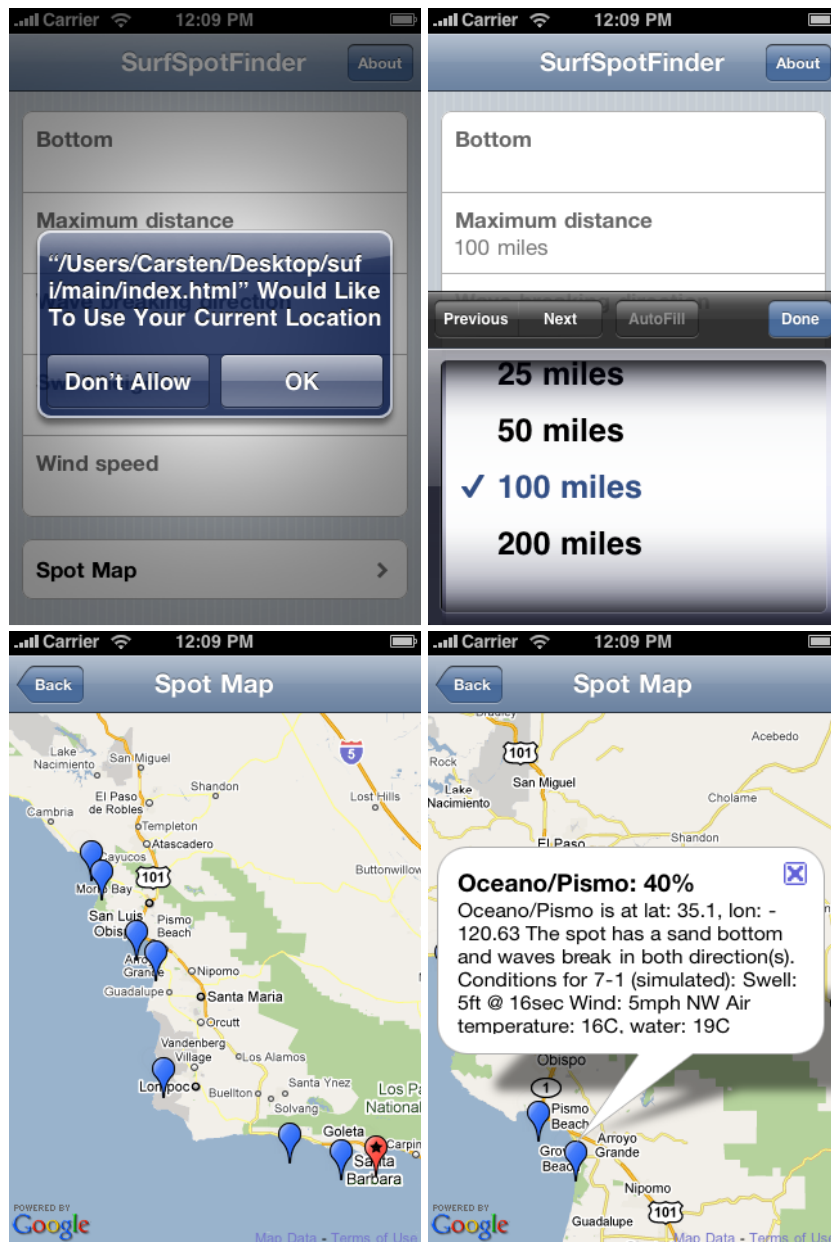
Figure 26: Screen shots of the prototype user interface. When starting the Web application, the user is asked whether she allows it to use her current location (top left). After selecting values for the user profile (top right), the user proceeds to the spot map (bottom left), which shows all spots matching her preferences with detailed information on the current conditions (bottom right).

Communication with the server component is realized using Asynchronous JavaScript and XML (AJAX), which allows the Web application to update the map without reloading the whole application. The AJAX functionality is provided by jQuery and straightforward to use: whenever the user changes one of the parameters in her profile, this change is sent to the server and the updated set of matching spots is returned as a KML[9] file and shown on the map. The server side implementation of the prototype is discussed in the following.

## 6.3 SERVER-SIDE IMPLEMENTATION

The implementation of the prototype's server component follows the generic approach described in Section 5.1. The architecture for the interaction between client and server follows the Representational State Transfer (REST) approach [48], where clients interact with resources on the server. In HTTP-based architectures that follow the REST principles, resources are made available via descriptive URIs. The interaction with these resources is realized based on the different HTTP methods, among which the prototype implementation makes use of the GET, PUT and DELETE operations. Table 3 shows an overview of the resources offered by the RESTful server implementation.

*RESTful Web services*

| Resource URI | HTTP Method | Description |
| --- | --- | --- |
| /users | GET | Returns a list of all users, represented as SWRL rules |
| /users/{id} | GET | Returns the SWRL rule representing a specific user |
| /users/{id} | DELETE | Deletes the specified user profile |
| /users/{id}/spots | GET | Returns the spots matching the user profile as a KML file |
| /users/{id}/bottom | GET | Returns the information on the type of bottom preferred by the user |
| /users/{id}/bottom | PUT | Writes the information on the type of bottom preferred by the user to the profile and returns the updated matching spots as a KML file |
| /users/{id}/bottom | DELETE | Deletes the type of bottom preferred by the user from the profile and returns the updated matching spots as a KML file |

Table 3: Resources offered by the RESTful server implementation. The three operations on the bottom resource are also available correspondingly for all other aspects of the user profiles.

---

9 KML was the XML-based file format originally developed for annotations and visualizations in Google Earth; it is now an OGC standard [197].

For the prototype implementation, the architecture proposed in Figure 21 has been wrapped as a RESTful Web service. This wrapping has been implemented as a Java Servlet that maps the different resources to (atoms within) the SWRL rules that model the user preferences as discussed in Section 5.2. In order to maintain the connection of a user to her specific rule, the ID assigned to every user is stored in a cookie on the client side. On the server side, a new resource is created for every user, based on the ID (see Table 3). This maintains the stateless nature of REST, since no session data have to be stored on the server. Moreover, this approach also enables the user to start with her previous preferences when she has already used the service before.
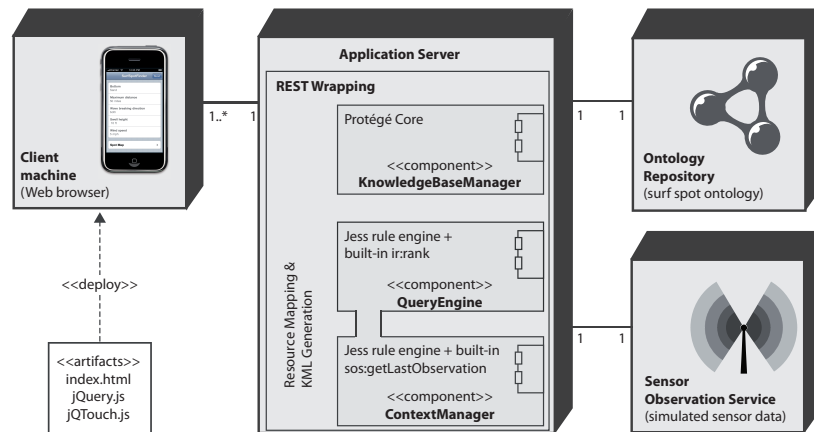


Figure 27: Extended deployment diagram for the prototype implementation based on Figure 21.

For all resources that require a computation of the matching spots, the user's SWRL rule is executed, using the ir:rank (see Section 5.2) and sos:getLastObservation (see Section 5.3) built-ins. The output of the ir:rank built-in, which computes a ranking based on deviance from the user preference as discussed in Section 6.1, is handled by the Java wrapper, which combines the different rankings into a single one: Each call to the ir:rank built-in produces a ranking of surf spots for one contextual aspect. These single rankings are then merged by the wrapper based on the mean rank of every spot. In the next step, the wrapper creates a KML file based on this ranking and returns it to the client. The KML files are generated using the Java API for KML[10].

The Sensor Web component shown in Figure 21 consists of a single Sensor Observation Service in case of the prototype implementation. Due to a lack of online sensors at the surf spots, this service simulates realistic conditions at the spots in the way discussed in Section 6.1. Besides the lack of access to real sensors, the simulation approach is also useful to test the Web application and evaluate its behavior when the conditions change. With real sensors, this evaluation would be a lengthy process, since weather conditions most often only change very

---

10 See http://labs.micromata.de/display/jak/.

70

slowly. Figure 27 shows the extended architecture of the prototype's server implementation.

This chapter has successfully applied the DIR measure introduced in Chapter 3 in the development of the Surf Spot Finder application. A simulation approach based on the DIR measure has been suggested. The results of the simulation were analyzed statistically to identify the most influential context parameters, taking the results of a questionnaire-based approach into account. Moreover, the generic approach for context modeling presented in Chapter 5 has been implemented in a working application. Finally, a REST-based wrapping for the user model management component has been introduced that can be reused and extended for other context-aware applications.

# CONCLUSIONS AND FUTURE WORK

7

This chapter summarizes the results obtained during the research for this thesis. Section 7.1 discusses the outcome with respect to the hypothesis and research questions phrased in Chapter 1. Potential application areas are discussed in Section 7.2, followed by an outlook on future work in Section 7.3.

## 7.1 SUMMARY AND DISCUSSION

The goal of this thesis was to develop a novel approach for context-aware information retrieval from semantics-based sources. The proposed method covers two aspects.

The first part of the thesis has dealt with the identification of contextual aspects that influence the outcome of context-aware information retrieval tasks. This identification process is required for the development of context-aware applications to make sure that important contextual information is taken into account. At the same time, it reveals irrelevant information that can be ignored to save computing power (especially on mobile devices) and reduce the complexity of user interfaces. In this thesis, we have introduced the cognitively plausible dissimilarity measure for information retrieval results (DIR) to support this task. Based on the changes in IR results caused by context changes, DIR accumulates the shifts of single results between two rankings. It weighs them according to their position within the ranking, so that the results at the top are emphasized, and normalizes the sum to the interval $[0, 1]$. Since the purpose of the measure is to reflect how people quantify the differences between result rankings, its cognitive plausibility was evaluated in two human participants tests. The tests' statistical evaluations indicate a strong correlation between the participant judgments and those calculated by the DIR measure. Based on these findings, the null hypothesis of no correlation has been falsified, and we have to accept the alternative hypothesis (see page 7). According to these results, DIR is a cognitively plausible measure for human judgments of IR result ranking dissimilarity, as it is based on the properties described in the hypothesis.

In the second part of this work, we have dealt with the fact that the Semantic Web is inherently static and it is therefore difficult to integrate information from the Sensor Web – a crucial requirement to make the retrieval process really context-aware. We have demonstrated how semantic rules can be employed to bridge these two worlds to enable context-aware geographical information retrieval from the Semantic Web. Using the Semantic Web Rule Language, we have shown how a model of user preferences can be fully represented in a SWRL rule. SWRL's support for rules with free variables allows for reasoning on the

*Identifying context*

*Modeling context*

73

individuals in an ontology – in our example, surf spots were compared against a user model and ranked based on deviation from the given user's preferences. Moreover, we have introduced two novel SWRL built-ins: one to dynamically read observation values from the Sensor Web during execution of a rule based on SWRL built-ins, and another one for performing queries by example on individuals' data type values. This approach allows for a strict separation of static knowledge about the individuals in an ontology and any dynamic information represented by an explicit link to sensors. Following this rule-based approach, we have introduced a generic architecture for context-aware information retrieval tools.

These two novel approaches for the identification and modeling of contextual information have been successfully applied in the development of the surf spot finder application, which served as a use case for this research. In the following, we will recall the research questions phrased on page 8 and discuss the results:

*How can relevant contextual aspects be identified?* Research question 1 has been answered by the DIR-based simulation approach introduced in Chapter 3 and demonstrated in the prototype implementation (see Chapter 6). Starting from the set of available context information (determined by available sensors, for example), different combinations of parameter values are fed into the system. The output is compared using the DIR measure: if it exceeds an application-specific value, the context parameter under consideration will be taken into account. If this is not the case, we can assume that changes of this parameter have a negligible effect on the results and can therefore be ignored. As such, the DIR measure can be used as a component in sensitivity analysis [164], which requires numeric input. DIR provides this numeric input by assigning values from the interval $[0, 1]$ to pairs of result rankings. The sensitivity analysis then consists in systematically changing context parameters that affect the information retrieval process to investigate the effect of these changes on the degree of change in the results. The approach proposed here to change single parameters – known as *One At a Time* (OAT) in sensitivity analysis – can be problematic, though. Visually speaking, the collection of all available context parameters spans a context parameter space. With an increasing number of available context parameters, the fraction of the subspace that is covered by changing parameters following the OAT approach is getting smaller. To solve this problem and cover the whole context parameter space, the complete set of all parameters must be taken into consideration, particularly to reveal interaction effects between different context factors. However, this approach makes it significantly harder to assign major changes in the results to a particular context parameter.

*Which properties should a context impact measure have?* Research question 2 has been discussed in Chapter 3, where we motivated and introduced the properties of the measure. The properties of DIR are based on four assumptions about users' interaction with information retrieval results. First, we have assumed that the user's impression of changes between two result rankings can be quantified only based

*DIR-based simulation*

on the position of the results within the ranking. DIR is thus based on what we called *shifts* – the difference in rank (and relevance value, respectively) between two rankings. This approach ignores other factors that potentially influence the users' judgments, such as system response times or the general layout of the system's user interface[1]. Second, we have assumed that the users' impression of the difference between two result rankings depends on whether the result visualization includes information about the results' relevance or not. Intuitively, one would think that users make a difference between a case where the first result is only slightly better than the second one, and a case where it is twice as good, for example. We have therefore introduced two different variants of the DIR measure, one for rank-based visualizations, and one for visualizations that include information on the results' relevance for the given search term (this distinction does not seem to be made by users; see the discussion of research question 3 below). Third, previous research has shown that users focus on the top results when they interact with information retrieval results [2]. We have modeled this fact by a weighting approach that applies the inverse ranks as weights to the shifts. Fourth, we assume that there is a common ground in this behavior for all users, i.e., that the users' perception of changes in result rankings is sufficiently uniform so that it makes sense to define one measure for it. This last assumption is fundamental for the definition of a generic measure and verified in research question 3.

*How can the cognitive plausibility of the context impact measure be shown?* Research question 3 has been evaluated in the two human participants tests described in Chapter 4. Both tests have shown a strong correlation between the participant judgments and the values calculated by the DIR measure, so that the hypothesis $H_A$ of this research can be accepted. Nonetheless, not all assumptions made prior to the tests were borne out by the results. The most striking result was that the users do not seem to take further information on the single results' relevance into account when they assess the dissimilarity of two result rankings. They seem to be focusing only on the order of the results. In both tests, this was indicated by the fact that the purely rank-based $DIR_{rank}$ showed an equally high correlation with the participant judgments as $DIR_{rel}$ – independent of whether the participants were given information about the results' relevance or not. We can only speculate about the reason for this ignorance of information that is very useful from a subjective point of view. One reason for these user judgments may be the high cognitive load of processing detailed information on the results' relevance. Taking these relevance values into account during dissimilarity judgment is a more complex task than simply looking at the results' order. While we have to leave the study of this question for future research, this finding confirms the approach taken by most Web search engines,

---

1 In fact, it has been shown that the user interface layout does influence the users' interaction with the results. Google's vice president Marissa Mayer presented some findings on how the white space around the results influences user behavior at the IO conference 2007 in San Francisco; the experiment is also briefly mentioned in an interview with her published at `http://searchengineland.com/qa-with-marissa-mayer-google-vp-search-products-user-experience-10370`.

which present their results as flat lists without any information on the relevance.

The second striking result of the evaluation of the first human participants test was the absence of a correlation between how (dis)similar two result rankings were, and the time the participants needed to solve the task. This analysis was done on the data from test one where the users were to decide which of two rankings is more different from a result ranking (see Section 4.1). The analysis was based on *diff* values – the difference in the DIR values between the two rankings and the reference ranking. Beforehand, we had assumed that tests with a larger *diff* value should be easier – and thus faster – to decide than cases where the *Processing time* difference in DIR is very small. Based on the test data, we can say that this is not the case. The only plausible explanation for this unexpected result is the limited number of results in each ranking presented during the test. It has been shown that the numbers of result were still within the range that can be reasonably processed by humans at the same time [129]. We therefore assume that this result does not transfer to settings with bigger numbers of results; however, this question has to be left for future research.

*How can context be modeled so that it can be used to modify existing knowledge bases?* Research question 4 was aimed at the problem of modeling context in a Semantic Web environment without limiting the approach to predefined contexts as previous approaches do. We proposed to *SWRL-based user* use the Semantic Web Rule Language to model user preferences and *model* include contextual information provided by sensors in the information retrieval process (see the discussion of research question 5 below). The approach has been demonstrated for the surf spot finder prototype in Chapter 6. There are two main advantages in using SWRL for this task. First, SWRL is part of the W3C's efforts towards establishing standards for the Semantic Web and is therefore compatible with all important standards in this field, particularly OWL, which it extends with rules expressed in RuleML. Second, the use of SWRL is not bound to a particular retrieval method. While we have built the retrieval functionality into the rules for this thesis, it is also possible to execute the rules before the actual retrieval. Other retrieval approaches may be based on standard reasoning or novel approaches such as similarity-based reasoning [83, 82]. While the import of the SWRL ontology can render OWL undecidable [75], which may be problematic depending on the application, the approach forces a clean distinction between the static parts – the actual ontology, modeled in OWL – and the dynamic parts that change depending on user preferences or context input, which are modeled as rules.

Writing SWRL rules can be cumbersome, especially for lay users. Applications building on SWRL therefore need to hide this complexity from the user with intuitive user interfaces. One approach for this has been introduced in Chapter 6, where the user picks combinations of preferences that are then translated into SWRL. The translation has been implemented as a two-step process in the prototype Web application's user interface: The user enters her preferences on a simple HTML form,

where each change in the form triggers an AJAX call of a REST resource on the server. In the second step, this resource is mapped to a SWRL atom, and the rule representing the user is modified. This encapsulation of the reasoner and rule engine expose their functionality to a wider choice of clients which do not necessarily need to know about the internal reasoning. It is sufficient for them to be able to interact with a RESTful service interface.

*How can the context model be linked to real-time information?* Research question 5 covers was concerned with making information retrieval on the Semantic Web really context-aware. The notion of context-awareness in AI is inherently connected to the idea that applications adjust their behavior to the current conditions in the real world, for example, the conditions that define the situation of the current user. As for the user model, we also used SWRL for this approach, particularly its support for custom built-ins. We have shown how the open source OX framework can be wrapped in a novel built-in that enables instances in the ontology to be updated with the current values measured by OGC compliant sensor networks. For example, it is no longer necessary to manually update information on the swell height for the surf spot instances in our application ontology; the `sos:getLastObservation` built-in takes care of this process now. The built-in integrates the retrieval process with the Sensor Web. The minimum requirement to enable this functionality is that the OGC Sensor Observation Services accessed by the built-in are created as instances in the ontology, with basic information on their service capabilities. These representations of Sensor Observation Services then have to be linked to the actual instances under consideration in the application; in the case of our scenario, the surf spot instances.

*SWRL built-in for the Sensor Web*

This approach renders all reasoning on the instances of the ontology that are linked to sensor observations non-monotonic. While this may seem obvious and constitutes the basic idea of context-aware computing, the implication that the same reasoning task may lead to different outcomes – depending on the current context defined by the sensor observations – must be kept in mind when offering services and data on the Semantic Web. While we have only discussed the case where a compatible user interface wraps the functionality of the service, one could also imagine a reasoning service that can be embedded in semantically annotated service chains [85]. In this case, the non-monotonic behavior of the service must be included in the service descriptions.

As this discussion of the research questions phrased in Chapter 1 shows, the questions have all been addressed in the course of this research. The hypothesis $H_A$ could be accepted, i.e., human judgments of the dissimilarity of information retrieval result rankings correlate strongly with the normalized sum of pairwise rank differences, each weighted by the higher of the two ranks. In the following, we will discuss potential application areas, both for the DIR measure and for the SWRL-based retrieval approach.

We have focused on particular applications for the DIR measure and SWRL-based context models throughout this research. This section discusses other potential application areas for the developments of this thesis.

APPLICATIONS FOR THE DIR MEASURE. In this research, the DIR measure has been used to assess the influence of different context parameters on the results of an information retrieval task. However, since DIR is based solely on result rankings, it can be applied for a number of analyses in information retrieval. For example, DIR could be used *Search engines* to analyse the effects of search engine index updates on the results. The same applies for updates to the ranking algorithms. From a search engine provider's point of view, it may be undesirable to have such an update cause too many changes in the results, as users might get the impression that the results are unstable, unreliable, or inconsistent. DIR can be used to prevent such unwanted effects. Users most often do not know about the aspects that go into the retrieval and ranking algorithms. From their perspective, DIR can be used to assess the semantic relatedness[2] of different search terms. The underlying assumption is that closely related search terms should lead to very similar results, reflected in a DIR value close to 0. Whether this assumption holds and whether it is also possible to specify the kind of relationship (meronymy or functional relationships, for example) between two different search terms remains an open research question.

Further potential uses of the DIR measure in context-aware information retrieval are in the comparison of different approaches for including context. Yahoo has investigated different ways to make their Web search engine context aware (see Section 2.4). Their comparison of the different approaches focused on required changes to their search engine and computational costs, which are major considerations for such a large-scale service [107]. However, it would also be interesting to see to what degree the different solutions affect the results of specific queries. The DIR measure enables search engine developers to specify how sensitive to context changes their approach really is.

Outside of information retrieval, ranked lists also play a role for music and book sales charts, popularity rankings (of politicians based *Charts analyses* on pre-election polls, for example), league tables in sports, and more. In most of these cases[3], DIR could allow interesting insights on the consequences of certain actions on these lists. For example, one could analyze the degree of change in politicians' popularity rankings to statements made between two polls. Changes in book or music sales charts could be connected to influential reviews. As these examples show, there are a number of applications for the DIR measure outside

---

2 Semantic relatedness is a concept from natural language processing that includes different kinds of relationships between entities, such as meronymy, antonymy or functional relationships. Semantic similarity is thus a special case of semantic relatedness [33, 69].

3 League tables in sports are an exception, since the maximum possible degree of change is determined by the respective scoring system.

of information retrieval that potentially affect business or political-strategical decisions.

APPLICATIONS FOR SWRL-BASED CONTEXT MODELS.  User modeling in SWRL as proposed in this thesis is directly adoptable for use cases similar to the surf spot finder, such as retrieval of climbing or hiking routes [195], or selecting areas for activities like scuba diving or hang gliding. The selection of an appropriate spot or route for any of these activities is highly dependent on personal skills and preferences, as well as the current outside conditions. Hence, this approach is especially useful where personalized retrieval is required, and where the contextual information has a strong spatio-temporal component. *Spatio-temporally dependent tasks* While the mentioned examples all consist of non-critical retrieval tasks around recreational activities, one can also think of applications in more professional settings. In the disaster relief domain, for example, initial choices often have to be made about where to place personnel and equipment, where to set up camps and hospitals, etc. These decision processes could be supported by rule-based approaches. Instead of user preferences, the rules could represent specific staff members' skills and abilities to assure an optimal use and distribution of resources. As mentioned in Chapter 5, the rule-based user models can be easily transferred to user groups by combination of the single users' preferences. This approach only works as long as the rules do not contain any contradictions. Future research in this area should therefore extend this approach with trade-offs to enable real group decision support [47].

SWRL as a means to include sensor observations in the reasoning process via the `sos:getLastObservation` built-in is useful in any kind of inference process that relies on real-time data. Sheth and colleagues propose the identification of weather situations such as *freezing rain* based on rules [171, 71]. Since a growing number of sensors for all kinds of observations is made available as OGC compliant Sensor Observation Services, one can imagine a number of different context-aware reasoning systems. Examples range from traffic control systems *SWRL-based evaluation of sensor data* based on street network and air quality sensors, over systems based on marine sensors that monitor the conditions for dangerous increases in the number of algae, to crop production control systems. While some of these applications already exist – mostly based on statistical modeling approaches –, a semantics-based solution would allow for an easy integration of different Linked Data [20] sources and eventually enable meaningful sensor plug and play [31].

Putting this vision into practice, however, depends largely on the number of semantically annotated data sources and sensors available online. While their number is growing and annotation is becoming a straightforward process using tools such as the Semantic Annotations Proxy[4], there are a number of open issues to be addressed in future research.

---

4 See `http://semantic-proxy.appspot.com`.

This section rounds the thesis off with directions for future research.

APPLICATION DEVELOPMENT.    In this research, the DIR measure has been introduced as a method to identify relevant contextual information for an application under development. While its usefulness has been demonstrated in the surf spot finder prototype, the DIR measure has to be better integrated into the design and development of context-aware applications to tap its full potential. The influence of a contextual parameter on the outcome of an information retrieval task is only one aspect that should guide the choice of contextual information to cover in an application. User requirements, cost (especially when sensors have to be deployed), as well as potential stakeholder requirements also have to be taken into consideration. When these different views *Integrating DIR in* on an application conflict, the developers have to balance them. Exist-*the development* ing approaches for system analysis and design [14, 120] only include *process* guidelines for negotiating conflicting user requirements, but not for divergent outcomes of the analysis process.

In this respect, the focus on the outcome of the DIR measure during the selection of contextual information for the prototype in this thesis cannot be seen as an all-embracing approach. A combination of DIR with traditional questionnaire-based methods such as done by Wosniok [198] has already been outlined in Chapter 6. Moreover, the information retrieval process – such as selecting a surf spot whose current conditions match the user's preferences – is often only one functional aspect of an application. For the surf spot finder, profile management is another crucial functionality. Other aspects such as social networking capabilities might be added later. Depending on their impact on the information retrieval process, such changes to the application may require a reevaluation of the contextual aspects that play a role. This also applies when new sensors become available. In terms of application development, future research should therefore target the integration of DIR into the analysis and design process, and explore its applicability in system redesign and maintenance.

EVALUATION OF THE DIR MEASURE.    The human participants tests conducted for this research have shown that the outcome of the DIR measure correlates strongly with human dissimilarity judgments of information retrieval result rankings. While the results were positive for the specific participant group under consideration, future research should evaluate whether these results are replicable for a broader participant group. Specifically, it should be checked whether the results hold for other age groups, since the participants in test one (see Section 4.1) were all undergraduate students[5]. Other factors that potentially influenced the results of the study are the participants' educational level,

---

5 Since no personal data were collected in test two for a lack of reliability, we can only assume that the participant group in this test was more diverse; however, it is still likely that most of the participants of test two had a college education background, given the channels chosen for advertising the online study.

the fact that most of them can be assumed to have grown up with the Web (*digital natives*) and consequently their age. To confirm the results of the human participants tests, future studies should therefore look at other user groups with a broader range in age, education, and level of experience in the use of computers and the Web. Moreover, the participant groups in these two studies were too small to identify any gender effects. Potential future studies should therefore also investigate whether there is a difference in the perception of dissimilarities of result rankings between male and female users.

The independence of decision times from the difference in dissimilarity between two pairs of result rankings is an effect that could not be fully explained in the studies conducted in this research. In test one, the difference between two DIR values DIR(*a*,*b*) and DIR(*a*,*c*) did not influence how long the participants needed to decided whether *a* and *b* or *a* and *c* were more dissimilar. Apparently, the comparably small number of results used in the studies led to this effect. To evaluate this assumption, future studies should also work with higher numbers of results in the rankings, which are beyond the level that can be easily processed simultaneously by human users ($7 \pm 2$; see [129]).

While the goal of DIR is not to model or explain how human users reason about the dissimilarity of result rankings, insights into this process would be interesting both from a cognitive science perspective and for the further development and refinement of the DIR measure. Explanatory studies could shed light on this aspect of the dissimilarity of result rankings, where the participants are asked to comment on how they came to a certain decision in a given comparison task. Since interaction with IR systems is a very artificial process, it should be investigated whether there is a common approach at all, and how it relates to studies on similarity measurement from psychology [54, 57, 61, 105, 111]. Future research should specifically investigate whether effects based on salience and direction of comparison can also be observed for abstract entities such as IR result rankings.

Finally, DIR should also be evaluated within real applications, for example, when users are interacting with Web search engines. In this case, DIR can be applied to measure the effects of iterative refinement of the search phrase, which is a common strategy. A non-obstructive evaluation that does not interrupt the users' workflow, however, is challenging. One promising approach is to observe a user's query refinement process, for example, from `pizza` over `pizza recipe` to `pizza recipe Chicago style`. In order to not interrupt her workflow, one could ask the user about her impressions of the result changes between the different queries *after* she has found what she was looking for. This retrospective approach seems to be the least interruptive strategy, since it is not feasible to collect information on a user's impression of the degree of change within the results after a query refinement without directly asking.

SEMANTIC SENSOR WEB AND SDI.    The term *Semantic Sensor Web* has been coined by Sheth et al. [171], who propose the semantic anno-

tation of sensor data. The querying of OGC compliant Sensor Observation Services from a SWRL built-in proposed in this thesis extends this idea since it allows to link any semantically annotated information items to sensors providing real-time information. Typically, these information items represent real-world objects, such as surf spots in our scenario, which are the features of interest for the Sensor Observation Services under consideration. The light-weight method proposed in this research is only based on four properties described in a Sensor Observation Service's capabilities: `ServiceURL`, `ObservedProperty`, `ObservationOffering` and `FeatureOfInterest`. Creating an instance in the ontology that represents a specific SOS and linking it to the corresponding feature of interest is thus straightforward.

While this light-weight approach demonstrates the general feasibility of the mapping between OGC schemas and OWL ontologies, more complex applications require a full ontological representation of the OGC service specifications. Such a representation would allow for an on-the-fly annotation of Sensor Observation Services for use in semantics-based applications. One could imagine SWRL-based reasoning that involves different aspects of the SOS specifications such as sensor types and observation process specifications, combined with spatio-temporal reasoning about the features of interest. The scenario of a gas plume dispersion and the required analysis steps has been discussed by Janowicz et al. [85] for a semantically enabled Spatial Data Infrastructure. The annotation of the transferred documents can be done via the semantic annotations proxy mentioned above. This would enable an automatic and complete extraction of instances for an ontology based on the services' capabilities documents. Creating an OWL encoding based on the OGC service schemas is not straightforward, though. Several proposals have been made for translating XML schema (XSD) to OWL[6] [21, 4]. However, these strategies only cover the basic constructs in XSD and do not provide sufficient support for the resolution of included and imported schemas or the handling of selection lists, for example. The Sensor Observation Service specification, however, imports a variety of schemas, ranging from the Geographic Markup Language to Sensor Markup Language and digital rights management XSDs. A useful ontological representation of all these aspects can only be built iteratively, starting from the central concepts and with the help of the (natural language) specifications provided by OGC and ISO. One starting point for this ontological framework has been provided by Henson et al. [71], who developed on OWL encoding of a subset of the OGC Observations and Measurements XML Schema.

We have discussed the tight integration of semantics-based reasoning services into Spatial Data Infrastructures in [85]. The integration approach put forward in the paper consists in the establishment of a Web Reasoning Service (WRS) profile of the OGC Web Processing Service (WPS). While we have discussed the encapsulation of the SIM-DL similarity server in the paper, one could imagine the same wrapping

---

6 See also `http://xml2owl.sourceforge.net` and `http://rhizomik.net/html/redefer/xsd2owl`.

strategy for a SWRL-based reasoning service. Figure 28 shows a sample workflow for such a SWRL-based Web Reasoning Service. In this case, the conditions for precipitation would be formalized as SWRL rules that access Sensor Observation Services via the `sos:getLastObservation` built-in. For the Web Feature Service and the Web Processing Service that interact with the SWRL-based WRS, it acts as a standard WPS. While the chance for precipitation could also be offered directly by an SOS, the SWRL-based solution has the advantage that the service can automatically combine information from different Sensor Observation Services to calculate the chance for precipitation. The different SOS must offer the same feature of interest to enable the combination.
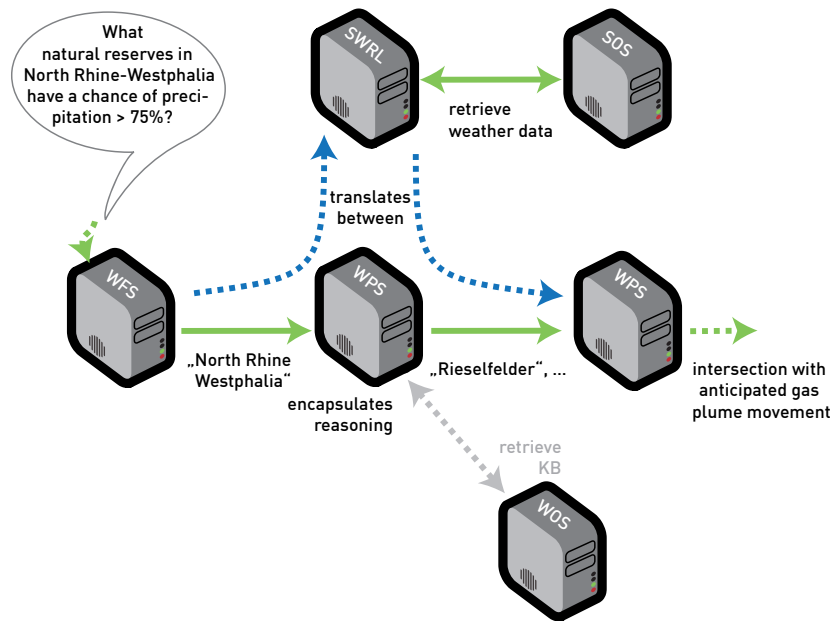
Figure 28: Workflow of a SWRL-based Web Reasoning Service, encapsulated as a Web Processing Service within the SDI (adapted from [85]).

Both on the Semantic Sensor Web and in Spatial Data Infrastructures, the combination of similarity reasoning [83, 82] and semantic rules is promising to combine reasoning about concepts (via SIM-DL) with rule-based inference about individuals. One could think of a reclassification of the ontology via rules before measuring similarity, which would put the ontology in the context of the current similarity query. Vice versa, subsets of the ontology which are most similar with respect to a given query could serve as input to rules, which could then adapt these preliminary results to the current context. In both cases, this would require a consistency check on the ontology after any reclassification. Exploring the combination of SIM-DL with rule-based representations of context is part of the SimCat II project[7].

---

7 SimCat II is the second phase of the SimCat project funded by the German Research Foundation (DFG Ja1709/2-2); see `http://sim-dl.sourceforge.net`.

SCALABILITY AND PERFORMANCE. The `sos:getLastObservation` built-in for SWRL introduced in Chapter 5 queries the Sensor Observation Services every time a rule using this built-in is executed by the rule engine. Moreover, this process also repeats itself for every individual that is the feature of interest of an SOS instance – even if it is always the same SOS. While this approach was sufficient for a proof-of-concept, execution time is a crucial aspect for real applications. Any rule making use of the built-in will have to wait for the Sensor Observation Service(s) to deliver the observation values in order to execute the rest of the rule. The more SOS instances are queried, the longer the user will have to wait for the system to response to her query.

*Caching*   To speed up processing, caching mechanisms are required. In a first step, the observations collected during one run of the rule engine should be cached, so that every SOS is only contacted once per rule engine run. The caching can then be extended based on pre-defined time schedules, especially for observed values that do not change very fast. Once caching has been implemented, offline querying of sensors would allow for another significant improvement in performance. Instead of querying the sensors when the rule is executed, they can be queried based on a fixed schedule. The SWRL built-in then accesses the cached data, instead of the costly operation on an SOS. For phenomena such as wave height, hourly observation is certainly sufficient (at least for the surfing scenario; a tsunami warning system, in contrast, would require frequent updates). Since the update schedules depend largely on the observation offerings, the SWRL built-in would need a configuration option for the update schedule. Alternatively, one could *Self-configuration* imagine an automatic configuration based on the Sensor Observation Services' archived data, which most of them also provide. The built-in could identify the average interval where the change in observed value exceeds a threshold $\delta$. The implementation of the SWRL-based user modeling approach in large-scale systems also raises the question of scalability. Future research should thus also investigate whether existing knowledge base management tools and rule engines can cope with massive user numbers and incoming sensor data, and explore potential optimization strategies.

UNCERTAINTY. When processing sensor data, information on the reliability of the data is crucial for many applications. In case of a toxic gas plume, for example, the area to evacuate must be calculated. In this process, it is critical to quantify the degree of uncertainty in the input data, as well as the degree of uncertainty added during further calculations based on these data. Errors occur during measurement and processing, and the data can be corrupted during (wireless) communication and storage. In the OGC Sensor Web specifications, the communication of uncertainty has not been considered, except for the (largely unspecified) `resultQuality` element in the Observations and Measurements standard [37].

*UncertML*   In order to address these issues, UncertML has been developed as "an XML encoding for the transport and storage of information about

uncertain quantities, with emphasis on quantitative representations based on probability theory"[196, p.5]. Since handling uncertainty is an issue that is currently being addressed in the Sensor Web community, the developed approaches also need to be evaluated for context aware computing applications. While we have assumed that there is a sharp distinction between entities (such as surf spots) that fit the user preferences, and those which do not, this line cannot always be drawn that clearly. Uncertainties in the input data (both from sensors and from user input), as well as in the processing models have to be taken into account in order to enable a probabilistic view on contextual information. This would also move retrieval from semantics-based sources that is currently purely binary closer to classical information retrieval. In order to enable the processing of contextual information with metadata on uncertainty in an ontology-based setting as discussed in this thesis, a mapping of the central concepts of UncertML to OWL would be required. A sound ontological specification of uncertainty also requires a clean separation of (observation) processes and events, *Processes and events* which has recently been addressed by Galton and Mizoguchi [50] and Devaraju and Kuhn [39], among others. While it does not yet account for uncertainty, the functional ontology for observations and measurements proposed by Kuhn [109] provides a solid foundation for further research in this direction.

CONTEXT MODELING.    This thesis has proposed a context modeling approach based on Semantic Web technologies that integrates observations from the Sensor Web. The development of context-aware applications has thus been shifted away from the implementation and deployment of system-specific sensors to the use of standardized Sensor Web services. These services can potentially serve numerous different applications with observation data and hence reduce the development and deployment cost for context-aware applications. The context model introduced in Chapter 5, though, was developed in a traditional way, starting from the application to be developed and selecting the contextual aspects to cover in this application by hand (and filtered using the DIR measure). In contrast to this top-down approach, the growing *Top-down vs.* number of sensor data made available through standardized services *bottom-up context* bears potential for a bottom-up approach: starting from the sensor *integration* data that are available, one could identify data that are potentially useful for specific applications. This would require a semantic annotation of the process model developed during requirements analysis (see Figure 17). Ontology-based business process models are subject to ongoing research [72, 117, 104]. Augmenting such semantic business process models with links to the sensor (or observation) types that are required to complete context-dependent tasks would allow for an automatic allocation of Sensor Observation Services that offer the required data. Similarity-based reasoning [83, 85] allows for the identification of potentially useful sensors if the exact required sensor type is not available.

When a large enough collection of semantically annotated Sensor Observation Services and business processes is available, it would even be possible to derive the set of supported processes from the sensor data at hand. Starting from this collection of processes, developers could follow a bottom-up approach that is based on the potential combinations of processes ad sensor data. The developer could then combine them in "process mash-ups", comparable to mash-ups that were enabled by Web 2.0 application programming interfaces. From a context modeling perspective, the combination of Semantic Sensor Web services and semantic business process specifications would provide developers with a toolbox that supports rapid prototyping of context-aware applications. Different context models for specific applications could be easily implemented and tested. Besides the required annotation of sensors and processes, however, this approach requires further research on the notion of context awareness: While the DIR measure supports a definition of context based on its impact on the output produced by an application, the impact of context-aware *processes* built into an application will be significantly harder to analyze. A process-driven understanding of context awareness hence requires research to understand how processes interact and how these interactions respond to changes in the sensor data that they handle.

On a more generic level, there is need for a modeling language for contextual information. Existing research – including this thesis – starts with the preconditions given in a specific environment determined by available data, implementations and infrastructures. In this respect, context awareness is always added as extra functionality on top of an existing infrastructure. A modeling language that supports the specific requirements of the design of context aware applications could put the context into the focus of application development. This will require an identification of the generic requirements of context-aware applications. Future research therefore needs to investigate whether (and how) an identification of such requirements is achievable. Moreover, the practicality of existing programming languages for such applications is up for debate, as turning such context-centric models into running systems may require novel software development paradigms.

PRIVACY CONCERNS.    Context-aware computing relies largely on information on the user and her current environment. The surf spot finder, for example, depends on a user's profile specifying her preferences, and her current location. The more detailed a user profile is, the better are the chances that the system can make an educated guess about which spot the user would like best. However, some of this information – especially her current location – can be sensitive and must therefore be properly protected from illegitimate access. Moreover, the
*Proportionality*    system must follow the principle of proportionality of data collection. It should only ask the user for information that it needs for its operation. For example, it is not necessary for the surf spot finder to know the user's real name to provide the functionality described in this thesis. While many online services collect such information to learn about

their users and market advertising space, context-aware applications should economize when it comes to sensitive personal information. The DIR measure can make a contribution to a spare data collection by reducing the set of potentially collectable information to information that has a significant impact for the given application. There is hope that users will become more sensitive in the future when they provide private data for online services, as the recent discussions about privacy issues in social networks have shown. A sparse data collection strategy that makes it transparent why certain data are required for a service to operate could thus become a distinctive feature for context-aware applications.

BIBLIOGRAPHY

[1] B. Adams and M. Raubal. Conceptual Space Markup Language (CSML): Towards the Cognitive Semantic Web. In *International Conference on Semantic Computing*, pages 253–260, Los Alamitos, CA, USA, 2009. IEEE Computer Society. (Cited on page 25.)

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 3–10. ACM New York, NY, USA, 2006. (Cited on pages 27, 33, and 75.)

[3] B. A. Aleman-Meza, C. Halaschek, B. I. Arpinar, and A. Sheth. Context-aware semantic association ranking. In *Semantic Web and Databases Workshop Proceedings*, pages 33–50, Berlin, Germany, September 2003. (Cited on page 23.)

[4] N. Anicic, N. Ivezic, and Z. Marjanovic. Mapping XML Schema to OWL. In G. Doumeingts, J. Müller, G. Morel, and B. Vallespir, editors, *Enterprise Interoperability*, chapter 23, pages 243–252. Springer London, London, 2007. (Cited on page 82.)

[5] G. Antoniou, C. V. Damásio, B. Grosof, I. Horrocks, M. Kifer, J. Ma, and P. F. Patel-Schneider. Combining Rules and Ontologies. A survey. REWERSE Deliverable I3-D3. Technical report, Institutionen för Datavetenskap, Linköpings Universitetet, 2005. (Cited on page 22.)

[6] D. Ayed, D. Delanote, and Y. Berbers. MDD Approach for the Development of Context-Aware Applications. In B. Kokinov, D. C. Richardson, T. R. Roth-Berghofer, and L. Vieu, editors, *Modeling and Using Context. 6th International and Interdisciplinary Conference, CONTEXT'07, Roskilde, Denmark*, Lecture Notes in Computer Science 4635, pages 15–28. Springer-Verlag Berlin Heidelberg, 2007. (Cited on page 6.)

[7] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003. (Cited on page 21.)

[8] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999. (Cited on page 29.)

[9] L. Barsalou. Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 5:513–562, 2003. (Cited on page 25.)

[10] L. Barsalou, W. Yeh, B. Luka, K. Olseth, K. Mix, and L. Wu. Concepts and meaning. In K. Beals, G. Cooke, K. Kathman, K. McCullough, S. Kita, and D. Testen, editors, *Papers from the parasession on conceptual representations*, volume 29 of *Chicago Linguistics Society*, pages 23–61. University of Chicago: Chicago Linguistics Society, 1993. (Cited on page 16.)

[11] M. Bazire and P. Brézillon. Understanding Context Before Using It. In A. K. Dey, B. N. Kokinov, D. B. Leake, and R. M. Turner, editors, *Modeling and Using Context, 5th International and Interdisciplinary Conference (CONTEXT 2005)*, volume 3554 of *Lecture Notes in Computer Science*, pages 29–40. Springer, July 2005. (Cited on pages 6 and 22.)

[12] S. Bechhofer. The DIG Description Logic Interface. In *International Workshop on Description Logics (DL2003)*, Rome, Italy, September 2003. (Cited on page 21.)

[13] D. Benslimane, A. Arara, G. Falquet, Z. Maamar, P. Thiran, and F. Gargouri. Contextual ontologies. In T. Yakhno and E. J. Neuhold, editors, *Advances in Information Systems*, volume 4243 of *Lecture Notes in Computer Science*, pages 168–176. Springer Berlin Heidelberg, 2006. (Cited on page 24.)

[14] B. Berenbach, D. Paulish, J. Kazmeier, and A. Rudorfer. *Software & Systems Requirements Engineering: In Practice*. McGraw-Hill Osborne Media, 2009. (Cited on page 80.)

[15] T. Berners-Lee. Semantic Web on XML. Slides from XML 2000 conference, Washington DC. Available from http://w3.org/2000/Talks/1206-xml2k-tbl/, December 2000. (Cited on pages xii and 2.)

[16] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001. (Cited on page 1.)

[17] Y. Biberman. A context similarity measure. In *Machine Learning: ECML-94*, pages 49–63, 1994. (Cited on page 19.)

[18] M. Biezunski, M. Bryan, and S. R. Newcomb. ISO/IEC 13250:2000 Topic Maps: Information Technology – Document Description and Markup Languages, December 1999. (Cited on page 60.)

[19] Y. Bishr. Overcoming the Semantic and Other Barriers to GIS Interoperability. *International Journal of Geographical Information Science*, 12(4):299–314, 1998. (Cited on page 15.)

[20] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009. (Cited on pages 2, 20, 54, and 79.)

[21] H. Bohring and S. Auer. Mapping XML to OWL Ontologies. In *Leipziger Informatik-Tage*, volume 72 of *Lecture Notes in Informatics*, pages 147–156. Springer, 2005. (Cited on page 82.)

[22] M. Boisot and A. Canals. Data, information and knowledge: have we got it right? *Journal of Evolutionary Economics*, 14(1):43–67, January 2004. (Cited on pages 1 and 17.)

[23] G. Booch. *Object-Oriented Analysis and Design with Applications*. Addison-Wesley, 1997. (Cited on page 12.)

[24] A. Borgida and L. Serafini. Distributed description logics: Assimilating information from peer sources. *Journal on Data Semantics*, 1:153–184, 2003. (Cited on page 24.)

[25] A. Borgida, T. Walsh, and H. Hirsh. Towards Measuring Similarity in Description Logics. In *International Workshop on Description Logics (DL2005)*, volume 147 of *CEUR Workshop Proceedings*. CEUR, Edinburgh, Scotland, UK, July 2005. (Cited on page 19.)

[26] M. Botts, G. Percivall, C. Reed, and J. Davidson. OGC Sensor Web Enablement: Overview And High Level Architecture. Technical report, OGC, 2007. (Cited on page 51.)

[27] P. Bouquet, A. Donà, L. Serafini, and S. Zanobini. ConTeXtualized local ontologies specification via CTXML. In *AAAI-02 workshop on Meaning Negotiation, Edmonton, Canada, July 2002*. AAAI Press, 2002. (Cited on page 24.)

[28] P. Bouquet, C. Ghidini, F. Giunchiglia, and E. Blanzieri. Theories and uses of context in knowledge representation and reasoning. *Journal of Pragmatics*, 35(3), 2003. (Cited on page 6.)

[29] P. Bouquet, F. Giunchiglia, F. Harmelen, L. Serafini, and H. Stuckenschmidt. C-OWL: Contextualizing ontologies. In *The Semantic Web – ISWC 2003*, volume 2870 of *Lecture Notes in Computer Science*, pages 164–179. Springer, 2003. (Cited on page 24.)

[30] D. Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. W3C recommendation, available from http://w3.org/TR/rdf-schema, February 2004. (Cited on page 20.)

[31] A. Bröring, K. Janowicz, C. Stasch, , and W. Kuhn. Semantic challenges for sensor plug and play. In *Web and Wireless Geographical Information Systems, 9th International Symposium, W2GIS 2009, Maynooth, Ireland, December 7-8, 2009*, volume 5886 of *Lecture Notes in Computer Science*, pages 72–86. Springer, 2009. (Cited on page 79.)

[32] P. J. Brown and G. J. F. Jones. Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering. *Personal and Ubiquitous Computing*, 5:253–263, 2001. (Cited on page 30.)

[33] A. Budanitsky and G. Hirst. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1): 13–47, March 2006. (Cited on page 78.)

[34] H. Chen, T. Finin, and A. Joshi. An Ontology for Context-Aware Pervasive Computing Environments. *Knowledge Engineering Review*, 18(3):197–207, 2003. (Cited on page 6.)

[35] S. Chou, W. Hsieh, F. Gandon, and N. Sadeh. Semantic Web technologies for context-aware museum tour guide applications. In *Advanced Information Networking and Applications, 2005. AINA 2005. 19th International Conference on*, volume 2, 2005. (Cited on page 23.)

[36] M. Compton, C. Henson, H. Neuhaus, L. Lefort, and A. Sheth. A survey of the semantic specification of sensors. In *2nd International Workshop on Semantic Sensor Networks, at 8th International Semantic Web Conference*, 2009. (Cited on page 3.)

[37] S. Cox, M. Botts, A. Robin, J. Davidson, S. Falke, A. Woolf, N. Ardlie, J. Greenwood, H. Shoebach, I. Simonis, G. Jatteau, and G. Creager. OGC Implementation Standard 07-022r1: Observations and Measurements – Part 1 – Observation schema, 2007. (Cited on page 84.)

[38] C. d'Amato, N. Fanizzi, and F. Esposito. A dissimilarity measure for $\mathcal{ALC}$ concept descriptions. In *Proceedings of the 2006 ACM Symposium on Applied Computing (SAC2006)*, pages 1695–1699, Dijon, France, April 2006. ACM. (Cited on page 19.)

[39] A. Devaraju and W. Kuhn. A Process-Centric Ontological Approach for Integrating Geo-Sensor Data. In *6th International Conference on Formal Ontology in Information Systems (FOIS 2010), Toronto, Canada, May 11-14, 2010*. (Cited on page 85.)

[40] A. K. Dey. Understanding and Using Context. *Personal Ubiquitous Computing*, 5(1):4–7, 2001. (Cited on pages 6, 22, and 23.)

[41] A. K. Dey, D. Salber, and G. Abowd. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction*, 16:97–166, 2001. (Cited on page 6.)

[42] S. Dietze, A. Gugliotta, and J. Domingue. Fuzzy context adaptation through conceptual situation spaces. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008 (IEEE World Congress on Computational Intelligence)*, pages 461–466, 2008. (Cited on page 25.)

[43] S. Dominich. *The Modern Algebra of Information Retrieval (The Information Retrieval Series)*. Springer, 2008. (Cited on page 17.)

[44] A. Dumitrescu and S. Santini. Think locally, search globally; context based information retrieval. In *International Conference on Semantic Computing*, pages 396–401, Los Alamitos, CA, USA, 2009. IEEE Computer Society. (Cited on page 23.)

[45] M. Efron. Using multiple query aspects to build test collections without human relevance judgments. In M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, editors, *ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, volume 5478 of *Lecture Notes in Computer Science*, pages 276–287. Springer-Verlag, 2009. (Cited on page 27.)

[46] M. J. Egenhofer. Toward the semantic geospatial web. In *GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, pages 1–4, New York, NY, USA, 2002. ACM Press. (Cited on pages 11 and 58.)

[47] M. Espeter and M. Raubal. Location-based decision support for user groups. *Journal of Location Based Services*, 3(3):165–187, 2009. (Cited on pages 3 and 79.)

[48] R. Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine, USA, 2000. (Cited on page 69.)

[49] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppin. Placing search in context: the concept revisited. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 406–414, New York, NY, USA, 2001. ACM Press. (Cited on page 23.)

[50] A. Galton and R. Mizoguchi. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied Ontololgy*, 4(2):71–107, 2009. (Cited on page 85.)

[51] P. Gärdenfors. How to Make the Semantic Web More semantic. pages 17–34. (Cited on page 25.)

[52] P. Gärdenfors. *Conceptual Spaces – The Geometry of Thought*. Bradford Books, MIT Press, Cambridge, MA, 2000. (Cited on pages 18, 19, 23, 25, and 60.)

[53] M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, 1987. (Cited on page 15.)

[54] D. Gentner and A. B. Markman. Structure mapping in analogy and similarity. *American Psychologist*, 52(1):45–56, 1997. (Cited on pages 18 and 81.)

[55] J. R. Gersh, J. A. McKneely, and R. W. Remington. Cognitive engineering: Understanding human interaction with complex systems. *Johns Hopkins APL Technical Digest*, 26(4):377–382, 2005. (Cited on page 25.)

[56] R. L. Goldstone and A. Kersten. Concepts and categorization. In A. Healy and R. Proctor, editors, *Comprehensive Handbook of Psychology*, volume 4, pages 599–621. 2003. (Cited on page 16.)

[57] R. L. Goldstone and J. Son. Similarity. In K. Holyoak and R. Morrison, editors, *Cambridge Handbook of Thinking and Reasoning*, pages 13–36. Cambridge University Press, Cambridge, 2005. (Cited on pages 18 and 81.)

[58] R. L. Goldstone, D. L. Medin, and J. Halberstadt. Similarity in context. *Memory & Cognition*, 25:237–255, 1997. (Cited on page 19.)

[59] M. F. Goodchild. Geographical information science. *International journal of geographical information systems*, 6(1):31–45, 1992. (Cited on page 3.)

[60] M. F. Goodchild, M. J. Egenhofer, K. K. Kemp, D. M. Mark, and E. Sheppard. Introduction to the Varenius project. *International Journal of Geographical Information Science*, 13(8):731–745, 1999. (Cited on page 64.)

[61] N. Goodman. Seven strictures on similarity. In *Problems and projects*, pages 437–447. Bobbs-Merrill, New York, 1972. (Cited on page 81.)

[62] F. J. Gravetter and L. B. Wallnau. *Statistics for the Behavioral Sciences*. Wadsworth Publishing, 8$^{th}$ edition, 2008. (Cited on page 27.)

[63] T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. (Cited on page 15.)

[64] T. Gu, X. Wang, H. Pung, and D. Zhang. An Ontology-based Context Model in Intelligent Environments. In *Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2004. (Cited on page 24.)

[65] N. Guarino. Formal ontology and information systems. In N. Guarino, editor, *International Conference on Formal Ontology in Information Systems (FOIS1998)*, pages 3–15. IOS Press, Trento, Italy, 1998. (Cited on page 2.)

[66] N. Guarino. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In A. T. Borgida, V. K. Chaudhri, P. Giorgini, and E. S. Yu, editors, *Conceptual Modeling: Foundations and Applications*, chapter 4, pages 52–67. Springer, 2009. (Cited on page 6.)

[67] S. M. Harrison. A comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 82–89, Denver, Colorado, USA, May 1995. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA. (Cited on page 39.)

[68] M. Hauswirth and S. Decker. Semantic Reality – Connecting the Real and the Virtual World. In *SemGrail Workshop, Redmond, WA, June 21–22, 2007*. Microsoft, 2007. (Cited on page 58.)

[69] B. Hecht and M. Raubal. GeoSR: Geographically explore semantic relations in world knowledge. In L. Bernard, A. Friis-Christensen, and H. Pundt, editors, *The European Information Society – Taking Geoinformation Science One Step Further. 11th AGILE International Conference on Geographic Information Science, Girona, Spain*, Lecture Notes in Geoinformation and Cartography, pages 95–114. Springer-Verlag, 2008. (Cited on page 78.)

[70] K. Henricksen, J. Indulska, and A. Rakotonirainy. Modeling context information in pervasive computing systems. In F. Mattern and M. Naghshineh, editors, *Pervasive '02: Proceedings of the First International Conference on Pervasive Computing*, pages 79–117. Springer-Verlag Berlin Heidelberg, 2002. (Cited on page 6.)

[71] C. A. Henson, J. K. Pschorr, A. P. Sheth, and K. Thirunarayan. SemSOS: Semantic Sensor Observation Service. In *Proceedings of the 2009 International Symposium on Collaborative Technologies and Systems (CTS 2009), Baltimore, MD*, 2009. (Cited on pages 3, 58, 79, and 82.)

[72] M. Hepp and D. Roman. An ontology framework for semantic business process management. *Proceedings of Wirtschaftsinformatik 2007*, 2007. (Cited on page 85.)

[73] A. Horn. On sentences which are true of direct unions of algebras. *Journal of Symbolic Logic*, 16:14–21, 1951. (Cited on pages 22 and 55.)

[74] M. Horridge and S. Bechhofer. The OWL API: A Java API for Working with OWL 2 Ontologies. In *OWLED 2009, 6th OWL Experienced and Directions Workshop*, Chantilly, Virginia, October 2009. (Cited on page 21.)

[75] I. Horrocks. OWL rules, OK? In *W3C Workshop on Rule Languages for Interoperability*, 2005. (Cited on pages 61 and 76.)

[76] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, 21 May 2004, available from http://w3.org/Submission/2004/SUBM-SWRL-20040521/, 2004. (Cited on pages 22 and 55.)

[77] J. Hu and P. Chan. Personalized Web Search by Using Learned User Profiles in Re-ranking. In *Workshop on Knowledge Discovery on the Web, KDD conference*, pages 84–97, 2008. (Cited on pages 4 and 30.)

[78] C. Hwang and M. Lin. *Group Decision Making under Multiple Criteria: Methods and Applications*. Springer, 1987. (Cited on page 57.)

[79] K. Janowicz. Sim-DL: Towards a semantic similarity measurement theory for the description logic $\mathcal{ALCNR}$ in geographic information retrieval. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006*

*Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, COMINF, IS, KSinBIT, MIOS-CIAO, MONET, OnToContent, ORM, PerSys, OTM Academy Doctoral Consortium, RDDS, SWWS, and SeBGIS. Proceedings, Part II*, volume 4278 of *Lecture Notes in Computer Science*, pages 1681–1692. Springer, October 2006. (Cited on pages 3, 7, and 19.)

[80] K. Janowicz. Kinds of Contexts and their Impact on Semantic Similarity Measurement. In *5th IEEE Workshop on Context Modeling and Reasoning (CoMoRea'08) at the 6th IEEE International Conference on Pervasive Computing and Communication (PerCom'08)*, Hong Kong, March 2008. IEEE Computer Society. (Cited on pages 29 and 32.)

[81] K. Janowicz and M. Raubal. Affordance-based similarity measurement for entity types. In S. Winter, M. Duckham, L. Kulik, and B. Kuipers, editors, *Spatial Information Theory, 8th International Conference (COSIT 2007)*, volume 4736 of *Lecture Notes in Computer Science*, pages 133–151, Melbourne, Australia, September 2007. Springer. (Cited on page 19.)

[82] K. Janowicz and M. Wilkes. SIM-DL$_A$: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-Concept to Inter-Instance Similarity. In *The 6th Annual European Semantic Web Conference (ESWC2009)*, volume 5554 of *Lecture Notes in Computer Science*, pages 353–367. Springer, 2009. (Cited on pages 76 and 83.)

[83] K. Janowicz, C. Keßler, M. Schwarz, M. Wilkes, I. Panov, M. Espeter, and B. Bäumer. Algorithm, implementation and application of the SIM-DL similarity server. In F. T. Fonseca, A. Rodríguez, and S. Levashkin, editors, *Second International Conference on GeoSpatial Semantics (GeoS 2007)*, number 4853 in Lecture Notes in Computer Science, pages 128–145, Mexico City, Mexico, November 2007. Springer. (Cited on pages xii, 3, 10, 19, 27, 32, 34, 76, 83, and 85.)

[84] K. Janowicz, C. Keßler, I. Panov, M. Wilkes, M. Espeter, and M. Schwarz. A study on the cognitive plausibility of SIM-DL similarity rankings for geographic feature types. In L. Bernard, A. Friis-Christensen, and H. Pundt, editors, *11th AGILE International Conference on Geographic Information Science (AGILE 2008)*, Lecture Notes in Geoinformation and Cartography, pages 115–133, Girona, Spain, 5-8. May 2008. Springer. (Cited on page 39.)

[85] K. Janowicz, S. Schade, A. Bröring, C. Keßler, P. Maué, and C. Stasch. Semantic Enablement for Spatial Data Infrastructures. *Transactions in GIS*, 14(2):111–129, 2010. (Cited on pages xv, 77, 82, 83, and 85.)

[86] M. Johannesson. Modelling asymmetric similarity with prominence. *British Journal of Mathematical and Statistical Psychology*, 53: 121–39, 2000. (Cited on pages 7 and 34.)

[87] C. B. Jones and R. S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219–228, 2008. (Cited on page 11.)

[88] I. Jurisica. Technical Report DKBS-TR-94-5: Context-based similarity applied to retrieval of relevant cases. Technical report, University of Toronto, Department of Computer Science, Toronto, 1994. (Cited on page 19.)

[89] T. Kauppinen. *Methods for Creating and Using Geospatio-Temporal Semantic Web*. PhD thesis, Aalto University School of Science and Technology, Faculty of Information and Natural Sciences, Department of Media Technology, Helsinki University Print, Helsinki, April 2010. (Cited on page 23.)

[90] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30 (12), 1938. (Cited on page 26.)

[91] A. Kent, M. M. Berry, J. Fred U. Luehrs, and J. W. Perry. Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American Documentation*, 6(2):93–101, 1955. (Cited on page 18.)

[92] C. Keßler. Similarity Measurement in Context. In B. Kokinov, D. Richardson, T. Roth-Berghofer, and L. Vieu, editors, *Modeling and Using Context. 6th International and Interdisciplinary Conference, CONTEXT 2007, Roskilde, Denmark, August 20–24, 2007, Proceedings*, volume 4635 of *Lecture Notes in Artificial Intelligence*, pages 277–290. Springer, 2007. (Cited on pages xii and 11.)

[93] C. Keßler. What's the Difference? A Cognitive Dissimilarity Measure for Information Retrieval Result Sets. *Knowledge and Information Systems*, accepted subject to minor revisions. (Cited on pages xii, xiii, 30, 41, 43, 44, and 46.)

[94] C. Keßler, M. Raubal, and K. Janowicz. The effect of context on semantic similarity measurement. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, OTM Academy Doctoral Consortium, MONET, OnToContent, ORM, PerSys, PPN, RDDS, SSWS, and SWWS 2007, Proceedings, Part II*, volume 4806 of *Lecture Notes in Computer Science*, pages 1274–1284, Vilamoura, Portuga, November 2007. Springer. (Cited on pages 32 and 33.)

[95] C. Keßler, M. Raubal, and C. Wosniok. Semantic Rules for Context-Aware Geographical Information Retrieval. In P. Barnaghi, editor, *4th European Conference on Smart Sensing and Context, EuroSSC 2009*, volume 5741 of *Lecture Notes in Computer Science*, pages 77–92. University of Surrey, Springer, 2009. (Cited on pages xiv, 10, 23, 30, 55, 56, and 60.)

[96] P. Kiefer and C. Schlieder. Exploring context-sensitivity in spatial intention recognition. In *In Proceedings of the Workshop on Behaviour Monitoring and Interpretation (BMI'07)*, pages 102–116, 2007. (Cited on page 23.)

[97] H. R. Kim and P. Chan. Personalized search results with user interest hierarchies learnt from bookmarks. In O. Nasraoui, O. Zaïane, M. Spiliopoulou, B. Mobasher, B. Masand, and P. S. Yu, editors, *Advances in Web Mining and Web Usage Analysis*, volume 4198 of *Lecture Notes in Computer Science*, chapter 9, pages 158–176. Springer, 2006. (Cited on page 4.)

[98] H. R. Kim and P. K. Chan. Learning implicit user interest hierarchy for context in personalization. *Applied Intelligence*, 28(2): 153–166, 2008. (Cited on page 23.)

[99] E. Klien. A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS, Special Issue on the Geospatial Semantic Web*, 11(3):437–452, 2007. (Cited on page 22.)

[100] E. Klien. *Semantic Annotation of Geographic Information*. PhD thesis, Institute for Geoinformatics, University of Münster, Germany, 2008. (Cited on page 22.)

[101] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2000. (Cited on page 23.)

[102] P. Korpipää and J. Mäntyjärvi. An Ontology for Mobile Device Sensor-Based Context Awareness. In P. Blackburn, C. Ghidini, R. M. Turner, and F. Giunchiglia, editors, *Modeling and Using Context*, volume 2680 of *Lecture Notes in Artificial Intelligence*, pages 451–458. Springer, 2003. (Cited on page 24.)

[103] P. Korpipää, J. Mäntyjärvi, J. Kela, H. Keranen, and E. J. Malm. Managing context information in mobile devices. *IEEE Pervasive Computing*, 2(3):42–51, 2003. (Cited on page 6.)

[104] A. Koschmider and A. Oberweis. Ontology based business process description. In *Proceedings of the CAiSE*, volume 5, pages 321–333, 2005. (Cited on page 85.)

[105] L. Kotovsky and D. Gentner. Comparison and categorization in the development of relational similarity. *Child Development*, 67: 2797–2822, 1996. (Cited on page 81.)

[106] R. Kraft, F. Maghoul, and C. C. Chang. Y!Q: contextual search at the point of inspiration. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 816–823, New York, NY, USA, 2005. ACM. (Cited on page 4.)

[107] R. Kraft, C. C. Chang, F. Maghoul, and R. Kumar. Searching with context. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 477–486, New York, NY, USA, 2006. ACM Press. (Cited on pages 4, 23, and 78.)

[108] C. L. Krumhansl. Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review*, 85:445–463, 1978. (Cited on pages 19 and 34.)

[109] W. Kuhn. A functional ontology of observation and measurement. In K. Janowicz, M. Raubal, and S. Levashkin, editors, *GeoSpatial Semantics. Third International Conference (GeoS 2009), Mexico City, 3–4 December 2009*, volume 5892 of *Lecture Notes in Computer Science*, pages 26–43. Springer, 2009. (Cited on page 85.)

[110] W. Kuhn. Semantic engineering. In G. Navratil, editor, *Research Trends in Geographic Information Science*, Lecture Notes in Geoinformation and Cartography, pages 63–76. Springer Berlin Heidelberg, 2009. (Cited on page 25.)

[111] L. B. Larkey and A. B. Markman. Processes of similarity judgment. *Cognitive Science*, 29(6):1061–1076, 2005. (Cited on page 81.)

[112] S. Laurence and E. Margolis. Concepts. In S. P. Stich and T. A. Warfield, editors, *Blackwell Guide to Philosophy of Mind*, pages 190–213. Wiley-Blackwell, 2003. (Cited on page 16.)

[113] S. Lawrence. Context in Web Search. *IEEE Data Engineering Bulletin*, 23(3):25–32, 2000. (Cited on page 23.)

[114] R. V. Lenth. Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician*, 55:187–193. (Cited on page 12.)

[115] A. Leonidis, G. Baryannis, X. Fafoutis, M. Korozi, N. Gazoni, M. Dimitriou, M. Koutsogiannaki, A. Boutsika, M. Papadakis, H. Papagiannakis, G. Tesseris, E. Voskakis, A. Bikakis, and G. Antoniou. AlertMe: A Semantics-based Context-Aware Notification System. In *33rd Annual IEEE International Computer Software and Applications Conference*, pages 200–205. IEEE, 2009. (Cited on page 23.)

[116] B. Li and F. Fonseca. TDD – a comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation*, 6(1):31–62, 2006. (Cited on page 19.)

[117] L. Liao and H. K. N. Leung. An ontology-based business process modeling methodology. In *ACST'07: Proceedings of the third conference on IASTED International Conference, Phuket, Thailand*, pages 114–119. ACTA Press, 2007. (Cited on page 85.)

[118] S. Loke. *Context-Aware Pervasive Systems: Architectures for a New Breed of Applications*. Auerbach Publications, 2006. (Cited on page 23.)

[119] H. P. Luhn. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4):309–317, 1957. (Cited on page 17.)

[120] L. A. Maciaszek. *Requirements analysis and system design*. Pearson Education, Essex, UK, 3rd edition, 2007. (Cited on pages 12, 49, 67, and 80.)

[121] J. Maeda. *The Laws of Simplicity (Simplicity: Design, Technology, Business, Life)*. The MIT Press, August 2006. (Cited on page 23.)

[122] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. (Cited on pages 2 and 17.)

[123] F. Manola and E. Miller. RDF Primer. W3C recommendation, available from http://w3.org/TR/rdf-primer, February 2004. (Cited on pages 20 and 51.)

[124] G. Marchionini. Toward Human-Computer Information Retrieval. June/July 2006 Bulletin of the American Society for Information Science and Technology, available from http://asis.org/Bulletin/Jun-06/marchionini.html, 2006. (Cited on pages 19 and 29.)

[125] C. Mascolo, L. Capra, and W. Emmerich. Mobile computing middleware. In E. Gregori, G. Anastasi, and S. Basagni, editors, *Advanced lectures on networking – NETWORKING 2002 Tutorials*, pages 20–58. Springer-Verlag New York, 2002. (Cited on page 60.)

[126] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology. Technical report, ISTC-CNR, 2002. (Cited on page 15.)

[127] D. Medin, R. Goldstone, and D. Gentner. Respects for similarity. *Psychological Review*, 100(2):254–278, 1993. (Cited on page 19.)

[128] M. Melucci. A basis for information retrieval in context. *ACM Transactions on Information Systems*, 26(3):1–41, 2008. (Cited on page 23.)

[129] G. A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63:81–97, 1956. (Cited on pages 47, 76, and 81.)

[130] D. R. Montello and P. C. Sutton. *An Introduction to Scientific Research Methods in Geography*. Sage Publications Ltd., 2006. (Cited on page 12.)

[131] G. E. Moore. Cramming More Components onto Integrated Circuits. *Electronics*, 38(8):114–117, April 1965. (Cited on page 1.)

[132] G. L. Murphy and D. L. Medin. The role of theories in conceptual coherence. *Psychological Review*, 92(3):289–316, 1985. (Cited on page 19.)

[133] A. Na and M. Priest. OGC Implementation Specification 06-009r6: OpenGIS Sensor Observation Service (SOS), 2007. (Cited on page 58.)

[134] K. Nedas and M. Egenhofer. Spatial similarity queries with logical operators. In T. Hadzilacos, Y. Manolopoulos, J. Roddick, and Y. Theodoridis, editors, *SSTD '03 – Eighth International Symposium on Spatial and Temporal Databases*, volume 2750 of *Lecture Notes in Computer Science*, pages 430–448. Springer, Santorini, Greece, 2003. (Cited on page 19.)

[135] K. A. Nedas and M. J. Egenhofer. Integral vs. separable attributes in spatial similarity assessments. In C. Freksa, N. S. Newcombe, P. Gärdenfors, and S. Wölf, editors, *Proceedings of the international conference on Spatial Cognition VI*, volume 5248 of *Lecture Notes in Artificial Intelligence*, pages 295–310. Springer, 2008. (Cited on page 27.)

[136] J. Nielsen. *Usability Engineering (Interactive Technologies)*. Morgan Kaufmann, September 1993. (Cited on page 25.)

[137] D. A. Norman. Steps toward a cognitive engineering: Design rules based on analyses of human error. In *Proceedings of the 1982 conference on Human factors in computing systems*, pages 378–382, New York, NY, USA, 1982. ACM. (Cited on page 25.)

[138] D. A. Norman. Cognitive Engineering. In D. A. Norman and S. Draper, editors, *User Centered System Design*, pages 31–61. Lawrence Erlbaum Assoc., Hillsdale, 1986. (Cited on page 25.)

[139] D. N. Osherson and E. E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58, 1981. (Cited on page 16.)

[140] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998. (Cited on pages 4 and 17.)

[141] J. Pascoe. Adding generic contextual capabilities to wearable computers. In *Second International Symposium on Wearable Computers (ISWC'98)*, pages 92–99, 1998. (Cited on page 22.)

[142] S. Pepper and G. O. Grønmo. Towards a general theory of scope. In *Extreme Markup Languages 2001, August 14–17, Montréal, Québec*, 2001. (Cited on page 60.)

[143] K. Popper. *The Logic of Scientific Discovery*. Routledge, new edition, 2002. (Cited on page 7.)

[144] C. Portele. OGC Implementation Specification 07-036: OpenGIS Geography Markup Language (GML) Encoding Standard. OpenGIS Standard, Open Geospatial Consortium, 2007. (Cited on page 59.)

[145] F. Probst. Observations, measurements and semantic reference spaces. *Applied Ontology*, 3:63–89. (Cited on page 25.)

[146] F. Probst. *Semantic Reference Systems for Observations and Measurements*. PhD thesis, Institute for Geoinformatics, University of Münster, 2007. (Cited on page 15.)

[147] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. W3C Recommendation, available from http://w3.org/TR/rdf-sparql-query, January 2008. (Cited on page 20.)

[148] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man and Cybernetics*, 19:17–30, 1989. (Cited on page 18.)

[149] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. Context-Phone: A Prototyping Platform for Context-Aware Mobile Applications. *IEEE Pervasive Computing*, 4(2):51–59, 2005. (Cited on page 23.)

[150] M. Raubal. Formalizing Conceptual Spaces. In A. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004)*, volume 114 of *Frontiers in Artificial Intelligence and Applications*, pages 153–164. IOS Press, Torino, Italy, November 2004. (Cited on pages 19 and 23.)

[151] M. Raubal. Representing Concepts in Time. In *Proceedings of the international conference on Spatial Cognition VI*, pages 328–343, Berlin, Heidelberg, 2008. Springer-Verlag. (Cited on page 16.)

[152] M. Raubal. Cognitive Engineering for Geographic Information Science. *Geography Compass*, 3(3):1087–1104, 2009. (Cited on page 25.)

[153] S. Reed and D. Lenat. Mapping Ontologies into Cyc. In *AAAI 2002 Conference Workshop on Ontologies For The Semantic Web*, July 2002. (Cited on page 15.)

[154] C. Rinner and M. Raubal. Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. *Journal of Geographic Information Sciences*, 10:149–156, 2004. (Cited on page 65.)

[155] E. L. Rissland. AI and Similarity. *IEEE Intelligent Systems*, 21(3): 39–49, 2006. (Cited on page 18.)

[156] S. E. Robertson. The probability ranking principle in IR. pages 281–286, 1997. (Cited on page 32.)

[157] J. L. Rodgers and W. A. Nicewanderer. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, 1988. (Cited on page 26.)

[158] A. Rodríguez. *Semantic Similarity Among Spatial Entity Classes.* Ph.D. Thesis, Department of Spatial Infromation Science and Engineering University of Maine, 2000. (Cited on page 19.)

[159] A. Rodríguez and M. Egenhofer. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*, 18(3):229–256, 2004. (Cited on pages 7, 19, and 34.)

[160] G.-C. Roman, C. Julien, and J. Payton. A Formal Treatment of Context-Awareness. In M. Wermelinger and T. Margaria, editors, *Fundamental Approaches to Software Engineering, 7th International Conference, FASE 2004, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004 Barcelona, Spain, March 29 – April 2, 2004, Proceedings*, volume 2984 of *Lecture Notes in Computer Science*, pages 12–36. Springer, 2004. (Cited on page 54.)

[161] E. Rosch. Principles of categorization. In E. Rosch and B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1978. (Cited on page 16.)

[162] D. E. Rose and D. Levinson. Understanding user goals in web search. In S. Feldman and M. Uretsky, editors, *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 13–19. ACM Press, 2004. (Cited on page 19.)

[163] S. Rosset, C. Perlich, and B. Zadrozny. Ranking-based evaluation of regression models. *Knowledge and Information Systems*, 12(3): 331–353, 2007. (Cited on page 39.)

[164] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. John Wiley Sons, 2008. (Cited on pages 31 and 74.)

[165] S. Scheider, K. Janowicz, and W. Kuhn. Grounding Geographic Categories in the Meaningful Environment. In K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, editors, *Spatial Information Theory*, Lecture Notes in Computer Science, pages 69–87. Springer. (Cited on page 25.)

[166] S. Scheider, A. Devaraju, K. Janowicz, P. Maué, S. Schade, C. Keßler, J. Ortmann, M. Bishr, T. Fincke, T. Weigel, and W. Kuhn. Grounding Geographic Information. In *The European Information Society - Taking Geoinformation Science One Step Further (Poster Proceedings of the 12th AGILE International Conference on Geographic Information Science, June, Hannover, Germany)*, 2009. (Cited on page 25.)

[167] A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers and Graphics*, 23(6):893–901, 1999. (Cited on page 10.)

[168] L. Schneider. Designing Foundational Ontologies. In I.-Y. Song, S. W. Liddle, T. W. Ling, and P. Scheuermann, editors, *Conceptual Modeling - ER 2003. 22nd International Conference on Conceptual Modeling. Chicago, IL, USA, October 13-16, 2003*, volume 2813 of *Lecture Notes in Computer Science*, pages 91–104. Springer, 2003. (Cited on page 15.)

[169] A. Schwering. *Semantic Similarity Measurement including Spatial Relations for Semantic Information Retrieval of Geo-Spatial Data*. PhD thesis, Institute for Geoinformatics, University of Münster, Germany, 2006. (Cited on page 19.)

[170] A. Schwering. Approaches to Semantic Similarity Measurement for Geo-Spatial Data – A Survey. *Transactions in GIS*, 12(1):5–29, 2008. (Cited on page 19.)

[171] A. Sheth, C. Henson, and S. S. Sahoo. Semantic Sensor Web. *Internet Computing, IEEE*, 12(4):78–83, 2008. (Cited on pages 3, 11, 58, 79, and 81.)

[172] E. Sirin and B. Parsia. Pellet: An OWL DL reasoner. In V. Haarslev and R. Möller, editors, *Proceedings of the 2004 International Workshop on Description Logics (DL2004)*, volume 104 of *CEUR Workshop Proceedings*, Whistler, British Columbia, Canada, June 2004. (Cited on page 21.)

[173] P. D. Smart. *A Semantic Web Rule Language for Geospatial Domains*. PhD thesis, Cardiff University, School of Computer Science, 2009. (Cited on page 22.)

[174] P. D. Smart, A. Abdelmoty, B. El-Geresy, and C. Jones. A Framework for Combining Rules and Geo-ontologies. In M. Marchiori, J. Z. Pan, and C. Marie, editors, *Web Reasoning and Rule Systems*, volume 4524 of *Lecture Notes in Computer Science*, chapter 10, pages 133–147. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. (Cited on page 22.)

[175] E. E. Smith. Three Distinctions about Concepts and Categorization. *Mind & Language*, 4(1–2):57–61, 1989. (Cited on page 16.)

[176] C. Spearman. "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904. (Cited on page 26.)

[177] A. Spink and C. Cole, editors. *New Directions in Cognitive Information Retrieval*. Springer, 2005. (Cited on pages 3, 4, 19, and 31.)

[178] C. Stasch, K. Janowicz, A. Bröring, I. Reis, and W. Kuhn. A Stimulus-Centric Algebraic Approach to Sensors and Observations. In N. Trigoni, A. Markham, and S. Nawaz, editors, *GeoSensor Networks. Third International Conference, GSN 2009, July 13–14, 2009. Proceedings*, volume 5659 of *Lecture Notes in Computer Science*, pages 169–179. Springer, 2009. (Cited on page 3.)

[179] A. Stefanidis and S. Nittel, editors. *GeoSensors Networks*. CRC Press, 2005. (Cited on page 58.)

[180] S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946. (Cited on pages 32 and 64.)

[181] T. Strang and C. Linnhoff-Popien. A Context Modeling Survey, 2004. (Cited on page 6.)

[182] G. Strube. The role of cognitive science in knowledge engineering. *Proceedings of the First Joint Workshop on Contemporary Knowledge Engineering and Cognition*, 622:161–174, 1991. (Cited on page 26.)

[183] D. Tsarkov and I. Horrocks. FaCT++ Description Logic Reasoner: System Description. In *Proceedings of the International Joint Conference on Automated Reasoning (IJCAR 2006)*, volume 4130 of *Lecture Notes in Artificial Intelligence*, pages 292–297. Springer, 2006. (Cited on page 21.)

[184] T. Tudorache, N. Noy, S. Tu, and M. Musen. Supporting Collaborative Ontology Development in Protégé. In A. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, and K. Thirunarayan, editors, *The Semantic Web – ISWC 2008*, volume 5318 of *Lecture Notes in Computer Science*, pages 17–32, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. (Cited on page 54.)

[185] T. Tudorache, J. Vendetti, and N. F. Noy. Web-Protégé: A Lightweight OWL Ontology Editor for the Web. In *OWL: Experiences and Directions (OWLED 2008). Fifth International Workshop, Karlsruhe, Germany, October 26-27, 2008 co-located with ISWC 2008*, 2008. (Cited on page 54.)

[186] A. Tversky. Features of similarity. *Psychological Review*, 84(4): 327–352, 1977. (Cited on pages 7, 18, 19, and 34.)

[187] A. Ukkonen, C. Castillo, D. Donato, and A. Gionis. Searching the Wikipedia with Contextual Information. In *CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge mining*, pages 1351–1352, New York, NY, USA, 2008. ACM. (Cited on page 23.)

[188] C. J. van Rijsbergen. *Information Retrieval*. Butterworth, 2 edition, 1979. (Cited on page 18.)

[189] W3C. OWL Web Ontology Language Reference. W3C recommendation, available from `http://www.w3.org/TR/owl-ref/`, February 2004. (Cited on pages 2, 20, and 51.)

[190] W3C. OWL 2 Web Ontology Language. W3C recommendation, available from `http://www.w3.org/TR/2009/REC-owl2-overview-20091027/`, October 2009. (Cited on page 21.)

[191] W3C. Geolocation API Specification. W3C editor's draft, available from `http://dev.w3.org/geo/api/spec-source.html`, February 2010. (Cited on page 63.)

[192] J. Wang. Mean-Variance Analysis: A New Document Ranking Theory in Information Retrieval. In M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, editors, *ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, volume 5478 of *Lecture Notes in Computer Science*, pages 4–16. Springer-Verlag, 2009. (Cited on page 32.)

[193] X. Wang, J. S. Dong, C. Chin, S. Hettiarachchi, and D. Zhang. Semantic Space: An Infrastructure for Smart Spaces. *IEEE Pervasive Computing*, 3(3):32–39, July 2004. (Cited on page 23.)

[194] W. Weerkamp, K. Balog, and M. de Rijke. Using Contextual Information to Improve Search in Email Archives. In *Advances in Information Retrieval. 31st European Conference on Information Retrieval Conference (ECIR 2009)*, pages 400–411. 2009. (Cited on page 23.)

[195] M. Wilkes. A graph-based alignment approach to context-sensitive similarity between climbing routes. Diploma thesis, Institute for Geoinformatics, University of Münster, Germany, 2008. (Cited on pages 10 and 79.)

[196] M. Williams, D. Cornford, L. Bastin, and E. Pebesma. Uncertainty Markup Language (UncertML). OGC Discussion Paper 08-122r1, available from `http://www.uncertml.org/documents/UncertML.pdf`. (Cited on page 85.)

[197] T. Wilson. OGC KML. OGC standard, Open Geospatial Consortium, 2008. (Cited on page 69.)

[198] C. Wosniok. A Context-Aware Surf Spot Search Engine Based on Conceptual Spaces. Diploma thesis, Institute for Geoinformatics, University of Münster, Germany, 2010. (Cited on pages 10, 30, 66, 67, and 80.)

[199] A. Zimmermann, M. Specht, and A. Lorenz. Personalization and context management. *User Modeling and User-Adapted Interaction*, 15(3–4):275–302, 2005. (Cited on page 6.)

[200] H. Zwaal, M. Hutschemaekers, and M. Verheijen. Manipulating context information with SWRL (A-MUSE Deliverable D3.12). Technical report, 2006. (Cited on page 24.)

**Carsten Keßler**

Geboren am 8. Februar 1979 in Georgsmarienhütte.
Familienstand: Ledig.

*Eltern:*
Hans-Jürgen Keßler,
Ursula Keßler, geb. Sandkämper.

*Schulbildung:*
Grundschule:          August 1985 – Juli 1989, Hagen a.T.W.
Orientierungsstufe:   August 1989 – Juli 1991, Hagen a.T.W.
Gymnasium:            August 1991 – Juni 1998, Georgsmarienhütte.

*Hochschulreife (Abitur):*
Am 23. Juni 1998 in Georgsmarienhütte.

*Zivildienst:*
Juli 1998 – Juli 1999, Alten- u. Pflegeheim St. Marien, Georgsmarienhütte.

*Studium:*
Diplom Geoinformatik   Oktober 1999 – Dezember 2004,
                       Westfälische Wilhelms-Universität Münster.

*Prüfungen:*
Diplom im Fach Geoinformatik am 10. Dezember 2004,
                       Westfälische Wilhelms-Universität Münster.

*Promotionsstudium:*
Geoinformatik          Mai 2007 – Mai 2010,
                       Westfälische Wilhelms-Universität Münster.

*Tätigkeiten:*

| | |
|---|---|
| Studentische Hilfskraft | Oktober 2000 – Juni 2002, |
| | Datenverarbeitungszentrale, |
| | Fachhochschule Münster. |
| Studentische Hilfskraft | August 2002 – Januar 2005, |
| | Institut für Geoinformatik, |
| | Westfälische Wilhelms-Universität Münster. |
| Wissenschaftliche Hilfskraft | Februar 2005 – Oktober 2006, |
| | Institut für Geoinformatik, |
| | Westfälische Wilhelms-Universität Münster. |
| Wissenschaftlicher Mitarbeiter | Seit November 2006, |
| | Institut für Geoinformatik, |
| | Westfälische Wilhelms-Universität Münster. |

*Beginn der Dissertation:*

Oktober 2006 bei Prof. Dr. Martin Raubal am Institut für Geoinformatik,
Westfälische Wilhelms-Universität Münster.

Münster im Mai 2010

.............................................
*Unterschrift*