Similarity Measurement in Context

Carsten Keßler

Institute for Geoinformatics, University of Münster, Germany
carsten.kessler@uni-muenster.de

Abstract. Context plays a crucial role when measuring the similarity of two concepts. Nonetheless, the modelling of context has been mostly neglected in existing similarity measurement theories. In this paper, we explore the influence of context in existing similarity measurement approaches for the geospatial domain, focussing on whether and how these approaches account for it. Based on these observations, the processing of context during similarity measurement is analysed, and general implementation issues, especially ease of integration into existing reasoning systems and computability, are discussed. The results of the different analyses are then combined into a generic set of characteristics of context for similarity measurement, with regard to the geospatial domain.

Keywords: Similarity measurement, context, geospatial concepts.

1 Introduction and Motivation

The importance of context for similarity measurement has long been observed and is beyond dispute. In fact, context is required for similarity measures to make sense in the first place. 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" [18], p.292. As an example, imagine being asked to compare two buildings in New York City: the Chrysler Building and the Radio City Music Hall. The answer depends on whether you are currently talking about functional aspects, which makes both very dissimilar – or whether you are talking about architectural styles, which results in a high similarity of the two Art Deco buildings. To that effect, measuring similarity without taking context into consideration is in most cases useless [10].

Even so, the actual modeling and incorporation of context into similarity measurement has mostly been neglected or appears as future work in the literature. Existing similarity theories [9] produce satisfying results in psychological experiments. However, it must be noted that these experiments are carefully designed such that the subject's similarity ratings are not biased due to environmental – i.e. contextual – influences [23]. Such an isolated perspective on similarity has two drawbacks: on the one hand, it is based on unrealistic preconditions, as people's similarity ratings in everyday situations are always influenced by their current context; on the other hand, such theories are missing the chance to utilize contextual information to make similarity measurements more accurate and tailored to the situation of an individual.

B. Kokinov et al. (Eds.): CONTEXT 2007, LNAI 4635, pp. 277–290, 2007.

The motivation of this paper is to improve similarity measurements by explicitly integrating context. Such an integrated model would allow for more precise queries, not only retrieving the *generic* similarity of two concepts or individuals, but directly referring to the respects which need to be taken into consideration. Concerning applications using similarity measurement, contextual information can be useful to clarify ambiguous situations, e.g. when searching knowledge bases by query concepts. In such search scenarios, the knowledge base typically contains a lot of information that is insignificant for a comparison. The context can specify what information needs to be considered, and what is out of scope for the current task.

To develop a useful context model, "we must attain a better understanding of what context is" [4], p.2. We are thus interested in a definition of context that is application-driven, i.e. that allows us to figure out what context parameters are important for a particular comparison task. The specific aim of this paper is hence a notion that helps putting context for similarity measurement into computational practice. The long-term objective is the development of a tool which supports developers in assessing the influence of the available context parameters on the overall similarity measurement.

This paper focuses on the geospatial domain because there is a big interest in context for similarity measurement within this research area. On the one hand, similarity measurement has been an important field of research within the community during the last years, e.g. to enhance retrieval of geographic information, or to integrate heterogeneous spatial data sources [12, 20]. On the other hand, location is an important aspect of context and plays a crucial role in different applications such as location based services (LBS) or mobile decision support systems [22, 25]. An improved understanding of context for similarity measurement in the geospatial domain can thus contribute to further developments in various branches of this research field. Nonetheless, the anticipated results are expected to be largely transferable to other application areas for similarity measurement.

The remaining part of the paper is organized as follows: We first present relevant related work from the fields of similarity measurement and context. Three applications of similarity measurement from the geospatial domain are then analyzed regarding their incorporation of context. Finally, a definition of context for similarity measurement, and formal characteristics of context are presented, followed by conclusions and open research questions.

2 Related Work

This section presents relevant related work from the fields of similarity measurement and from other research areas with an interest in contextual information. A generic definition of context is presented as a starting point for a notion of context for similarity measurement.

2.1 Similarity Measurement

Similarity measurement theories stem from research on the human ability to intuitively determine how similar two objects are, and to put those similarity ratings in relation (e.g. "computer science is more similar to mathematics than geography"). There are two main interests within this research area: on the one hand, psychologists aim at understanding and modeling how humans rate similarity; on the other hand, the artificial intelligence (AI) community is interested in designing formal – and thus computable – methods for ambiguous reasoning tasks; however, integrated approaches that take both perspectives into account are rare. Although the basic idea of similarity measurements is to reflect human ratings, the design of cognitively adequate algorithms that reproduce the human similarity rating process is difficult in practice. This is not only because of a lack of understanding concerning the underlying cognitive processes, but also because existing knowledge representations such as ontologies focus on formalizing knowledge, rather than matching the mental concept representations of human agents. Hence, the focus of this paper is on the AI perspective of similarity measurement, striving for cognitively plausible results which match human similarity ratings; the discussion whether the applied methods that lead to those results correspond to human cognitive processes is thus secondary.

From the psychological perspective, there are different approaches to similarity measurement, relying on different ways of representing concepts. Similarity in feature-based approaches can be computed in such representations following different strategies, for example counting (and possibly weighting) common features, and integrating structural similarity [27]. Geometry-based approaches, in contrast, assign dimensions with a geometric or topologic structure to the concepts which represent their properties [8]. All concepts are thus placed in an n-dimensional space, which allows for similarity measurement based on the semantic distance between two concepts. Network models put the stress on the edges in the network, and are mostly used to reproduce similarity ratings from human subject tests. Independent of the approach chosen for concept representation, similarity values are usually normalized to values between 0 (completely dissimilar) and 1 (identical). Although this list is not complete[1], it is sufficient to show what different preconditions a generic notion of context must be able to adapt to.

2.2 Defining 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 [2]. 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 similarity measurement [19]. A generic definition of context for ubiquitous computing is presented in [4], pp.3-4:

"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."

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) [4]. This list does not claim completeness, nor do all of the aspects always play a role, as

[1] For a comprehensive list of similarity theories, see [9].

will be shown. The definition will serve as a starting point for this paper, since it is from a related field of research, yet still generic enough to be transferred to similarity measurement for the geospatial domain. We will develop a more specific definition for similarity measurement throughout this paper.

3 Similarity and Context in the Geospatial Domain

In this section, we analyze three different approaches for similarity measurement in the geospatial domain. The chosen scenarios stem from research publications and represent a broad range of applications for similarity measurement in this scientific field, both in terms of concept representation method and kind of application. The objective of this review is to show which aspects of context play a role in the presented use cases, and to point out which of them have been considered in the corresponding approaches. For this purpose, the categories identity, activity, location and time from the definition in section 2.2 are used for reference. Moreover, this review demonstrates the need for context in similarity measurement, as none of the presented tasks can be completed satisfyingly without taking context into consideration.

3.1 Comparing Geospatial Features

Rodríguez and Egenhofer (2004) introduce the Matching-Distance Similarity Measure (MDSM) [24], an approach for the comparison of geospatial features in ontologies. MDSM is a weighted sum of the similarities of two concepts' parts, functions and attributes, extending Tversky's ratio model [27]. It allows for asymmetric similarity measurement, as the perceived similarity of a to b is not always the same as the similarity of b to a. This fact is either based on the varying prominence of the instances at hand (e.g. the Kaufmann Concert Hall is more similar to the Radio City Music Hall than vice versa) [13, 27], or on the comparison of sub- and super-concepts (e.g. Concert Halls are more similar to Buildings than vice versa) [5].

MDSM explicitly includes context, modeling it as a set of tuples consisting of operations and their arguments. This information is processed in two manners: First, the domain of application is determined by selecting all features that are ontologically related to the operations' arguments. Second, weights for mereological, functional and attributional similarity are derived from the context. These weights can be calculated based on variability (focusing on a feature's informativeness) or on commonality (focusing on how characteristic a feature is for the application domain).

The notion of context included in MDSM is based on the assumption that all relevant contextual information is immanent in the task the user wants to perform – the activity, using the terminology of our current context definition. However, referring to the other aspects of the definition, spatial and temporal constraints are not supported by this context view. This limitation is based on the structure of the underlying ontology, which lacks spatial and temporal information. The user preferences are represented through the operations selected for the context. Looking at the examples given in the paper, such as "the user's intention is to play a sport", location and time provide important contextual information: a system that considers a more detailed form of context could reduce the domain of application to locations in the user's vicinity, and opening times could be considered. A user model would even allow for a weighting by the user's preferences, e.g. higher weighting of swimming pools than soccer fields. It must, however, be noted that the ontology used in the paper does not contain individuals; consequently, instance-specific information (location, opening times) cannot be considered in the measurements. Although the notion of context in MDSM could be refined by further information, it introduced the first inclusion of context for similarity measurement in the geospatial domain.

3.2 Geographic Information Retrieval

Janowicz (2006) introduces Sim-DL, a similarity theory for concepts specified in the $\mathcal{ALCNR}$ description logic (DL) [11]. The development of Sim-DL aims at closing the gap between similarity theories from psychological research and formal knowledge representations used in the AI community. Similarity in Sim-DL is asymmetric and calculated as the normalized weighted sum of the similarities of all descriptions of two concepts. The similarity of the single parts is the overlap of their concept descriptions in normal form. Comparable to the approach presented in the previous section, contextual information is used in Sim-DL to specify the domain of application. Moreover, weights are used to express the impact of a part on the overall similarity. The method for determining weights is not specified within Sim-DL.

Context is explicitly stated together with the search concept when starting a similarity query. The author uses the example of "botels" in Amsterdam to illustrate the SIM-DL approach. When measuring whether botels are conceptually closer to hotels or to boat houses, the user explicitly states that the context for this comparison should be, for example, housing. Accordingly, all concepts within the knowledge base related to housing are used for the similarity measurement. Concepts which are related to the query concept, but not related to housing (such as tub or water taxi), are not taken into consideration. Regarding our current definition of context, the main question is how to model identity, location and time, which cannot be represented in $\mathcal{ALCNR}$ (activity is represented through the choice of the domain of application). As the author points out, $\mathcal{ALCNR}$ is not expressive enough, e.g. to explicitly state geographic locations (which requires concrete domains), but only topological relationships. Likewise, temporal relations can be expressed, but no specific points in time. This lack of expressiveness limits what can be said about instances. Moreover, reasoning in description logics is expensive concerning computation time, even on simple knowledge bases with only a few concepts. To improve this, more efficient reasoning algorithms for DL are required. Consequently, Sim-DL is an approach to similarity measurement with a limited notion of context which is compatible with AI knowledge representations, but which still has limitations in practice.

3.3 Landmark Selection for Pedestrian Navigation

Raubal (2004) [21] presents a formalization of Gärdenfors' conceptual spaces [8], a theory from cognitive semantics accounting for the fact that different people may have different understandings of the same expressions. Conceptual spaces are sets of quality dimensions with a geometric or topologic structure. Concepts are represented as regions in such a space (instances as points, respectively), allowing for similarity measurements based on semantic distance. Raubal formalizes Gärdenfors' model as conceptual vector spaces, employing z-transformations to standardize the values for dimensions of the space. As opposed to the approaches presented above, similarity is calculated at the instance level in this case.

The approach is demonstrated by using a pedestrian navigation scenario, where user and system[2] have different conceptualizations of landmarks. For example, the system's conceptualization may include information about buildings' historical importance, which is irrelevant to most users. Such semantic gaps are closed via transformations and projections between the corresponding vector space representations.

The conceptual spaces approach includes various aspects of context. Conceptual spaces are centered on the user, so that there is a detailed user model at hand, i.e. the user's conceptual space. Moreover, the paper introduces methods to match this user profile with external conceptualizations, which can be utilized to extend existing systems with user profiles, and to match between different systems. The choice of a landmark at every decision point during the navigation task is context-dependent: Among the landmarks available at a decision point, the most distinct one is chosen, i.e. the landmark with the largest semantic distance to the landmark prototype. The prototype is an imaginary landmark instance, calculated as the combination of the mean values for each dimension. Beyond user and location, the author discusses temporal aspects of context. For the scenario, the time of day is crucial. If a landmark sticks out because of its color, it is salient during daytime, but not at night [28]. The different contexts are represented by weightings on the dimensions of the conceptual space, for example the color is assigned a high weight for the daytime context, and a low one for the nighttime context.

Comparing the presented approach to our current definition of context, only activity is not explicitly modeled, whereas user, location and time are already included. However, this is mostly due to the use case chosen, which includes a fixed activity (pedestrian navigation). Contextual aspects depending on the task could easily be included by adapting the weights. The limitations of this approach are based on the requirement for every quality to be at least ordered in some way; data on the nominal scale cannot be represented properly[3] [14]. Moreover, conceptual spaces have only been used for small numbers of dimensions so far, and the scenarios were mostly of limited complexity. Further research is required to demonstrate how this approach can be applied in more complex situations.

3.4 Summary

The three approaches presented in this section embark on different strategies for concept representation and similarity measurement, and also for the inclusion of context. Nonetheless, they share the idea of assigning weights to the single factors that go into a similarity measurement to reflect a specific context. Accordingly, these

---

[2] More precisely, the system reflects the system designer's conceptualization.

[3] It is possible to integrate nominal values by creating a Boolean dimension for every one, but this easily leads to a large number of dimensions, rendering the whole approach impracticable.

weights have a big impact on the overall result of the measurement. Rodríguez & Egenhofer have analyzed two different strategies – commonality and variability – for the different scenarios in their paper. Although the change of strategy did not alter the overall ranking drastically, the commonality approach (which puts the stress on common features) produces more variation in the results. From a cognitive perspective, this approach seems to be the more plausible one compared to the variability strategy, since psychological research has found that people appear to focus on commonalties, also referred to as the *max effect* [17].

Context is also used to determine the domain of application. This is either done by automatic extraction of concepts from the user's query [24], or by explicit statement of context concepts [11]. In both cases, these concepts are used to select all related concepts from the knowledge base as the domain of application. It is remarkable that none of the presented approaches allow for the inclusion of additional contextual information that is not already present in the knowledge base, because the essential idea of context in other fields of research is mostly to add supplementary information to what is already known. In some cases, context is even regarded as completely external to the knowledge base [6].

Concerning the similarity measurement itself, all approaches assume the existence of a common understanding of the basic terms of the knowledge base, usually defined in a shared vocabulary such as a top level ontology. As the presented strategies only select context from within the knowledge base, this applies also to the context. In conclusion, it must be pointed out that all of the presented approaches were focusing on the similarity theory itself, and that context was only a part of the theory. The notions of context engaged within the theories are thus not complete, but show how context can generally be integrated in the similarity measurement process.

4 Context for Similarity Measurement Applications

In this section, we present the requirements and constraints for context in similarity measurement applications. Following from those theoretical and practical prerequisites, a definition of context for similarity measurement is given, and a set of generic properties for this notion of context is formalized.

4.1 General Requirements

The observations from section 3 have shown that it is not possible to come up with a fixed context model for similarity measurement in the geospatial domain. The context parameters that play a role depend to a great extent on the application. Although the categories identity, activity, location and time have been used for the analysis, these categories are not of great help. On the one hand, they are too generic, since *every* contextual parameter can be squeezed into one of those categories. On the other hand, they do not say anything about the relevance of these categories for a comparison task.

Apart from the relevance of concepts, the question of how to obtain data for those parameters plays a big role in practice. Aspects of context that are not available in the knowledge base must either be captured automatically, or, if this is not possible, explicitly provided by the user. Collection of context information by user input is a usability issue and must be balanced in every case. The formalization of the knowledge base is also important for the context model [2]. As we have seen in section 3, the context must in any case be in the same form of representation as the knowledge base; otherwise, it is not possible to integrate both. For example, providing additional contextual information, formalized in the Web Ontology Language (OWL) [16], for an application built on conceptual spaces would be hard to utilize, since both are based on very different kinds of concept representation. Moreover, the context must refer to the same shared vocabulary as the knowledge base to enable integration, where the knowledge base can also serve as the shared vocabulary. Such integration also allows for comparison of different contexts. Intuitively, a similarity measurement should produce similar results under similar contexts. This behavior could also be observed in MSDM: changes in the strategy for selection of weights, resulting in slight changes to the context, caused only small changes in the outcome of the similarity measurement.

Research on similarity measurement has lead to the development of models that produce reliable results. Accordingly, context should be established as an add-on to existing similarity theories – instead of inventing yet another similarity theory. Specific context models are heavily depending on different aspects of the application, even within the specific field of similarity measurement; nonetheless, it is still possible to make generic statements about context for similarity measurement. The next sections will give an overview of the typical environment for contextual information, and then define context for similarity measurement on a generic level, which provides the basic conditions for specific context models built for applications.

4.2 The Context Processing Chain

Applications that make use of contextual information generally follow a certain process chain when completing a task for the user. For a context-aware similarity application, this chain starts when the user defines the kind of problem he wants to solve. These problems are composed of comparisons of concept pairs at the lowest level. Such a query may be augmented with an explicit context statement, but parts of the context can also be automatically extracted[4] and then interactively refined by the user. Time and location, for example, are contextual aspects which can easily be captured automatically.

After this first initialization step, the user query and the context information have to be aligned with the knowledge base, i.e. it must be checked whether the knowledge base already contains all context information provided with the query. If this is not the case, the additional information must be “injected” to the knowledge base, relying on a shared vocabulary for alignment. The domain of application is then a subset of this extended knowledge base, consisting of those parts of the knowledge base that are conceptually related to the context. Within the domain of application, weights are assigned to the concept in the next step. The steps completed so far can be regarded as

[4] Techniques for automatic context extraction are beyond the scope of this paper. First solutions, which can partly be transferred to similarity measurement applications, can be found in [15] for context-aware web search engines, and in [3] for ubiquitous computing.

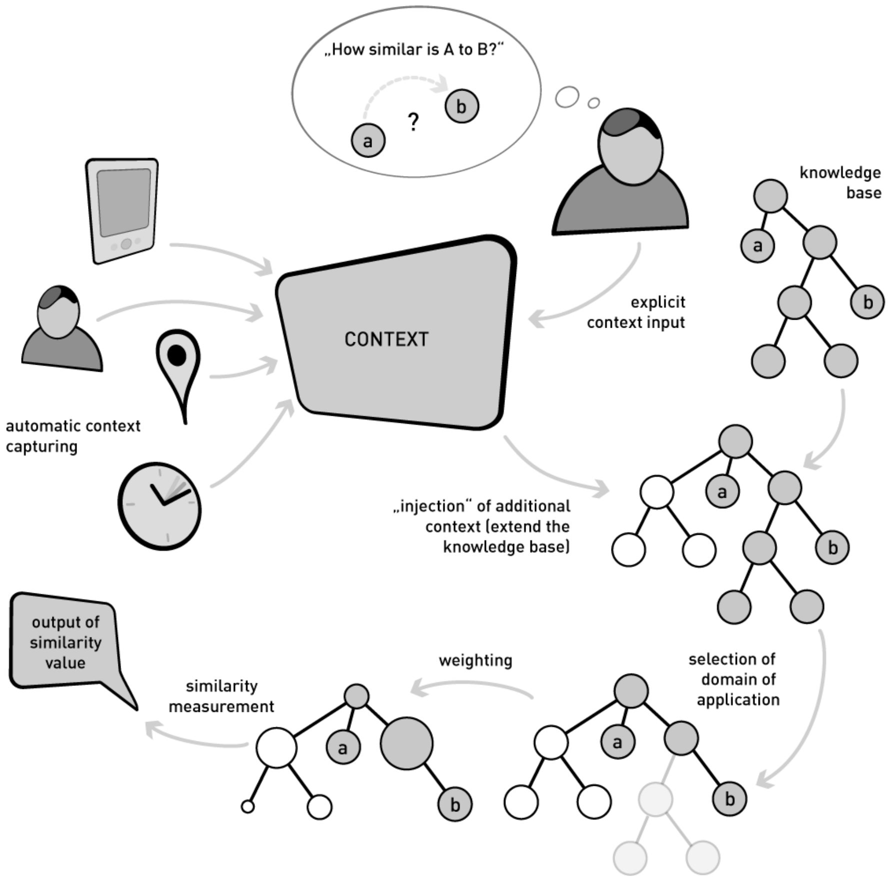

Fig. 1. The overview of the context processing chain uses a tree-structured knowledge base for reasons of simplicity. The general sequence of the process remains the same for other kinds of knowledge base representation.

a preparation for the actual similarity measurement, which is then carried out on the weighted domain of application. The method applied for similarity measurement again depends on the kind of knowledge representation. For complex queries, several iterations of this process may be required, until the results are finally presented to the user. Such results may, for example, consist of a ranked list of the most similar concepts compared to the query concept. Figure 1 shows on overview of the context processing.

4.3 Definition of Context for Similarity Measurement

A generic definition of context was given in section 2.2, focusing on the four elements identity, activity, location and time. As explained in the previous section, this definition is not useful for similarity measurement, as it does not support the choice and weighting of context parameters. Consequently, we define context for similarity measurement as follows:

A similarity measurement's context is any information that helps to specify the similarity of two entities more precisely concerning the current situation. This information must be represented in the same way as the knowledge base under consideration, and it must be capturable at maintainable cost.

With the help of this definition, developers can check parameters that come into question for an application-specific context model for the following properties:

- *Impact*: does this parameter render the similarity measurement more precisely?
- *Representation*: can this parameter be represented in the knowledge base?
- *Capturing*: can this parameter be captured at maintainable cost?

The example of a mobile sight recommendation system for tourists, offered for rent by a tourist information office, shows the usefulness of the approach. Assuming that the sights are represented in a conceptual space, different contextual parameters are taken into consideration: both location and time have a high impact on the results, since users want recommendations of nearby sights, which should be open to the public at query time. These parameters can be represented in a conceptual space and they can be captured automatically. The history of the user's previously visited sights also has a high impact on the results, as it shows the user's interests. It can also be represented in a conceptual space, but it is hard to capture due to the fact that most tourists only use the system once, and manual input is not feasible. Manual input of keywords of interest, however, might still be acceptable, but cannot be represented in a conceptual space. As a final example, information on maintenance costs for a sight could be available from the municipality and easily represented in a conceptual space, but it does not affect the recommendations for tourists. This list of candidate context parameters is not complete, but it shows how developers can check candidates based on the criteria of impact, representation and capturing.

4.4 Generic Characteristics of Context for Similarity Measurement

Although we do not propose a specific formalized context model here, as this would have to be tailored both to the application and to the model of concept representation used in the knowledge base, it is still possible to formalize a set of generic characteristics of context. This affects especially the relationships between context, knowledge base and domain of application. Those characteristics will be shown in the following, referring to a similarity task with query concept a and target concept b, as this is the underlying operation for all complex similarity measurement tasks.

The following statements assume the existence of a similarity theory based on a language L with symbols and grammar, which allows for the construction of complex concepts and relationships among instances. Both the knowledge base K and the context C are expressed in L, and are assumed to be consistent in the following. As illustrated in section 4.2, it cannot always be assumed that all contextual information is already present in the knowledge base. Accordingly, we define an extended knowledge base K_E as the union of K with context C (note that K_E may be equal to K, if the context is already completely covered by the knowledge base):

$$K_E = C \bigcup K \tag{1}$$

K_E is also assumed to be consistent. As mentioned in section 4.1, a shared vocabulary is required to make sure that a connection between context and knowledge base can be established. Accordingly, the existence of at least one concept which is part of both the knowledge base and the context is required:

$$C \bigcap K \neq \varnothing \tag{2}$$

Going back to the definition of context for similarity measurement given in section 4.3, the impact of a potential context parameter (i.e. a concept c) for the overall similarity is crucial for the decision whether to include it in a context model for a specific application. The minimum impact is represented by an application-dependent constant δ, so that all potential context parameters can be checked against this threshold value. The final context then includes all parameters with an impact greater than δ, where the impact is defined as the mean difference between a similarity measurement in a context *with* the parameter compared to one *without* the parameter:

$$C = \{c \mid imp(c) > \delta\} \tag{3}$$

$$imp(c_n) = \frac{\sum \mid sim_{(c_n \in C)}(a,b) - sim_{(c_n \notin C)}(a,b) \mid}{\mid C \mid} \tag{4}$$

Following the process illustrated in section 4.2, the extended knowledge base K_E is then used to determine the domain of application D from the extended knowledge base. The domain of application then consists of all concepts c from K_E that are used to define either a or b. To enable this step, the language L must support subsumption:

$$D = \{c \in K_E \mid c \sqsupseteq a \sqcup c \sqsupseteq b\} \tag{5}$$

Besides the sets of concepts introduced so far, there exists a function w which assigns weights to the concepts c in the domain of application, reflecting the importance of every concept in a given context. The sum of all weights is 1:

$$w : D \times D \rightarrow [0,1], \sum w = 1 \tag{6}$$

Similarity is then a function that computes a similarity value between 0 and 1 for a pair of query and target concepts from the weighted domain of application D_W:

$$sim(q,t) : D_w \times D_w \rightarrow [0,1] \tag{7}$$

As the context is itself represented in the same form as the knowledge base, different contexts can be compared using a context-free comparison, where the domain of application comprises the whole context (without any reduction or addition), and the weights are all equal. This can be used to formalize the statement from section 4.1: the more similar two contexts are, the less a similarity measurement should change under those two contexts. In other words, the difference between the results of a similarity measurement in two different contexts converges to 0 with a growing similarity of the two contexts:

$$\lim_{sim(C_1,C_2)\to 1} sim_{C_1}(a,b) - sim_{C_2}(a,b) = 0 \tag{8}$$

The characteristics presented above are independent of the actual knowledge representation; however, subsumption has been taken for granted here, which cannot be assumed in general, but is supported by common languages for knowledge representation such as OWL. Together with the definition of context presented in section 4.3, they provide a generic foundation for the design of application-specific context models for similarity measurement.

5 Conclusions and Future Work

A proper incorporation of context into similarity measurement has mostly been neglected so far, missing the chance to disambiguate similarity measurements and to tailor them to specific situations. In this paper, we have analyzed three approaches to similarity measurement in the geospatial domain and discussed the influence of context on the corresponding use cases. Accordingly, the context models included in the similarity theories at hand were analyzed.

Based on the broad range of models for concept representation and corresponding methods for similarity measurement, a definition of context for similarity measurement was presented. The definition provides application developers with a notion of context that supports the selection of context parameters for similarity measurement applications, based on impact of the parameters, compatibility with the knowledge base (representation), and practicability (capturing). This is in line with the analysis of general requirements for a context model (consistency and compatibility with knowledge base) and the way context is processed when a similarity measurement is completed. The definition of context was finally complemented with a set of formal characteristics of context on an abstract level.

Future research should focus on context model implementations which put the generic findings of this paper into practice, to enable research on specific problems at the application level. Specifically, robust methods for the assignment of weights to the parts of the domain of application must be developed, depending on the current context. Newly developed methods must then be verified in human subject tests to evaluate whether the results correspond to user ratings. Sensitivity analyses are required to show which context parameters have the biggest influence to the overall similarity. Options for combination with other strategies for context parameter selection, for example based on granularity [26] or on cognitive processes [7], need to be investigated. More research is also required concerning the integration with existing knowledge base and reasoning systems. This is especially crucial as it is unlikely that existing knowledge bases will be converted to new formats required for similarity measurement, causing additional work for developers. Instead, such new functionality must be compatible with widely used representation languages such as OWL. Concerning the research on context for similarity measurement in general, the differences between similarity among instances, concepts and whole knowledge bases requires further research, as context comes in different flavors depending on what is compared [1].

The next steps within this research will be the development of a context-enabled similarity server as part of the SimCat project (http://sim-dl.sourceforge.net). The server will then be used for different use cases for context and similarity measurement. The first scenario planned for implementation is a portal for cyclists, allowing for context-dependent comparison of bike routes.

Acknowledgements

This research is funded by the German Research Foundation (DFG) under the project "Semantic Similarity Measurement for Role-Governed Geospatial Categories". Special thanks go to Martin Raubal and Krzysztof Janowicz for helpful comments.

References

1. Albertoni, R., De Martino, M.: Semantic Similarity of Ontology Instances Tailored on the Application Context. In: Meersman, R., Tari, Z. (eds.) On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE. LNCS, vol. 4275, Springer, Heidelberg (2006)
2. Bazire, M., Brézillon, P.: Understanding Context Before Using It. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
3. Chen, G., Kotz, D.: Context Aggregation and Dissemination in Ubiquitous Computing Systems. In: Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications, IEEE Computer Society Press, Los Alamitos (2002)
4. Dey, A.K.,, Abowd, G.D.: Towards a Better Understanding of Context and Context-Awareness. In: CHI 2000 Workshop on the What, Who, Where, When, Why and How of Context-Awareness. The Hague, The Netherlands (2000)
5. Egenhofer, M.J., Mark, D.M.: Naive Geography. In: Kuhn, W., Frank, A.U. (eds.) COSIT 1995. LNCS, vol. 988, Springer, Heidelberg (1995)
6. Ehrig, M., et al.: Similarity for Ontologies - A Comprehensive Framework. In: ECIS, Regensburg, Germany (2005)
7. Freksa, C., Klippel, A., Winter, S.: Invisible Geography-A Cognitive Perspective on Spatial Context. In: Spatial Cognition: Specialization and Integration 2007, Schloss Dagstuhl, Germany (2007)
8. Gärdenfors, P.: Conceptual Spaces - The Geometry of Thought. MIT Press, Cambridge (2000)
9. Goldstone, R.L., Son, J.: Similarity. In: Holyoak, K., Morrison, R. (eds.) Cambridge Handbook of Thinking and Reasoning, pp. 13–36. Cambridge University Press, Cambridge (2005)
10. Goodman, N.: Seven strictures on similarity. In: Goodman, N. (ed.) Problems and projects, pp. 437–447. Bobbs-Merrill, New York (1972)
11. Janowicz, K.: Sim-DL: Towards a Semantic Similarity Measurement Theory for the Description Logic ALCNR in Geographic Information Retrieval. In: Meersman, R., Tari, Z., Herrero, P. (eds.) On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. LNCS, vol. 4277, Springer, Heidelberg (2006)

12. Janowicz, K.: Similarity-Based Retrieval for Geospatial Semantic Web Services Specified using the Web Service Modeling Language (WSML-Core). In: The Geospatial Web - How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society, Springer, Heidelberg (2007) (forthcoming)
13. Johannesson, M.: Modelling Asymmetric Similarity with Prominence. British Journal of Mathematical and Statistical Psychology 53, 121–139 (2000)
14. Keßler, C.: Conceptual Spaces for Data Descriptions. In: The Cognitive Approach to Modeling Environments (CAME), Workshop at GIScience 2006, Münster, Germany: SFB/TR 8 Report No. 009-08/2006 (2006)
15. Kraft, R., Maghoul, F., Chang, Y.Q.C.C.: Contextual Search at the Point of Inspiration. In: Proceedings of the 14th Conference on Information and Knowledge Management (CIKM), Bremen, Germany, ACM Press, New York (2005)
16. McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language Overview. In: W3C Recommendation, pp. 2003–2004 (2004)
17. Medin, D., Goldstone, R., Gentner, D.: Respects for similarity. Psychological Review 100(2), 254–278 (1993)
18. Murphy, G.L., Medin, D.L.: The role of theories in conceptual coherence. Psychological Review 92(3), 289–316 (1985)
19. Pascoe, J.: Adding generic contextual capabilities to wearable computers. In: Second International Symposium on Wearable Computers (ISWC'98) (1998)
20. Quix, C., et al.: Matching schemas for geographical information systems using semantic information. In: Meersman, R., Tari, Z., Herrero, P. (eds.) On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. LNCS, vol. 4277/4278, Springer, Heidelberg (2006)
21. Raubal, M.: Formalizing Conceptual Spaces. In: Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004), IOS Press, Amsterdam (2004)
22. Rinner, C., Raubal, M.: Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. Journal of Geographic Information Sciences 10(2), 149–156 (2004)
23. Rissland, E.L.: AI and Similarity. IEEE Intelligent Systems 21(3), 39–49 (2006)
24. Rodríguez, A., Egenhofer, M.: Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure. International Journal of Geographical Information Science 18(3), 229–256 (2004)
25. Schiller, J., Voisard, A.: Location Based Services. Morgan Kaufmann Publishers Inc., San Francisco (2004)
26. Schmidtke, H.R.: Granularity as a Parameter of Context. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
27. Tversky, A.: Features of Similarity. Psychological Review 84(4), 327–352 (1977)
28. Winter, S., Raubal, M., Nothegger, C.: Focalizing Measures of Salience for Wayfinding, in Map-based Mobile Services - Theories. In: Meng, L., Zipf, A., Reichenbacher, T. (eds.) Methods and Implementations, pp. 127–142. Springer, Heidelberg (2005)