A Study on the Cognitive Plausibility of SIM-DL Similarity Rankings for Geographic Feature Types

Krzysztof Janowicz, Carsten Keßler, Ilija Panov, Marc Wilkes, Martin Espeter, and Mirco Schwarz

Institute for Geoinformatics, University of Münster, Germany
janowicz|carsten.kessler|i.panov|marc.wilkes|m.espeter|mirco.schwarz
@uni-muenster.de

Abstract. The SIM-DL theory has been developed to enable similarity measurement between concept specifications using description logics. It thus closes the gap between similarity theories from psychology and formal representation languages from the AI community, such as the Web Ontology Language (OWL). In this paper, we present the results of a human participants test which investigates the cognitive plausibility of SIM-DL, that is, how well the rankings computed by the similarity theory match human similarity judgments. For this purpose, a questionnaire on the similarity between geographic feature types from the hydrographic domain was handed out to a group of participants. We discuss the set up and the results of this test, as well as the development of the according hydrographic feature type ontology and user interface. Finally, we give an outlook on the future development of SIM-DL and further potential application areas.

1 Introduction and Motivation

Human judgments of similarity have been subject to research in psychology for more than fifty years now [1]. Different approaches to modeling similarity have been developed, including feature-based, network-based, and geometric approaches. More recently, the artificial intelligence (AI) community started investigations on computational similarity models as a new method for information retrieval [2]. The Matching Distance Similarity Measure (MDSM) [3] was the first similarity-based model that has been developed specifically for the geospatial domain. The description logic based SIM-DL similarity theory [4, 5] discussed in this paper has been introduced to overcome the gap between models from psychology and formal knowledge representation languages used in the AI community.

Numerous geospatial applications carry potential for the implementation of similarity-based information retrieval techniques. Geoportals could supplement result pages with matches that do not exactly fit the user's query, but share certain characteristics with the matches. Location-based services could derive points of interest in the user's current vicinity from similar, previously visited places. From a developer's point of view, similarity measurements bear a great potential to simplify and accelerate processes that still require tedious manual configuration, such as data integration or ontology alignment. Those are just a few examples; in general, every tool that has to deal with fuzzy or ambiguous input—either from users or from other systems—is a candidate for the application of similarity.

For this paper, an application scenario from gazetteer research has been chosen. Current activities in this research area aim at the development of a distributed gazetteer infrastructure to replace existing isolated gazetteers [6], as shown in figure 1. In this envisioned infrastructure, similarity measurement is applied to enable interoperability among different gazetteers. Moreover, similarity allows for novel user interfaces which allow for imprecise input and do not require the user to know about the internal organization of the gazetteer any longer (see [6] for details).

However, to enable such new techniques, a formal representation of the geographic feature types organized in gazetteers is required. As a starting point for first tests, an ontology specifying hydrographic feature types has been created. This ontology is used for the human participants test presented in this paper. The purpose of this test is to show that the similarity rankings calculated by SIM-DL correspond to human similarity judgments. The participants were asked to rate the similarity of a number of concepts such as *Lake*, *Ocean*, or *River* to the search concept *Canal*, solely based on the given concept definitions.

The remainder of the paper is organized as follows: section 2 presents previous work on similarity measurement and gazetteers. Section 3 introduces the SIM-DL theory and presents its implementation within the SIM-DL similarity server, as well as a novel gazetteer user interface that communicates with this server. Section 4 presents the human participants test and discusses the results, followed by conclusions in section 5.

2 Related Work

This section points to previous work on similarity and gazetteer research with a special focus on similarity theories applied within GIScience.

2.1 Semantic Similarity Measurement

The theory of similarity has its origin in cognitive science and was established to determine why and how entities are grouped into categories, and why some categories are comparable to each other while others are not [1, 7]. The main challenge with respect to *semantic* similarity measurement is the comparison of meanings as opposed to purely structural comparison. A language has to be specified to express the nature of entities and a measurement theory needs to be established to determine how (conceptually) close compared entities are. While entities can be expressed in terms of attributes, the representation of

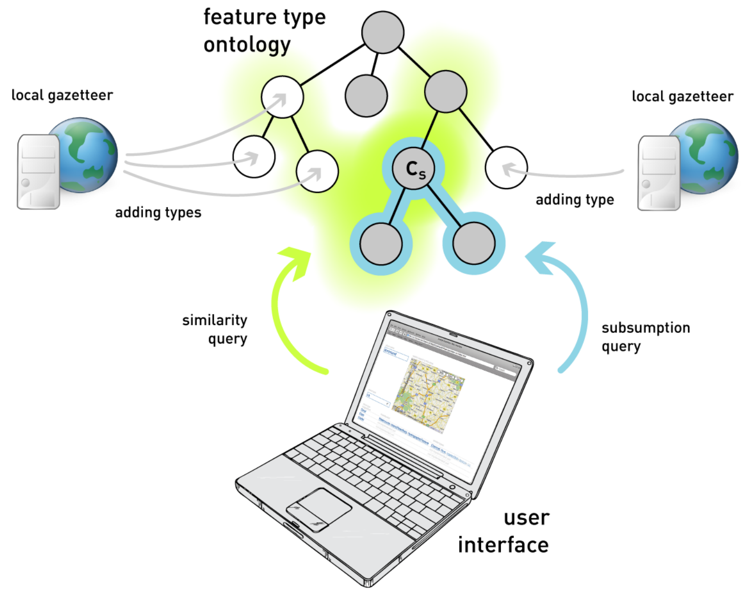

Fig. 1. Subsumption and similarity based information retrieval within the proposed gazetteer infrastructure [6].

entity types (concepts) is more complex. Depending on (computational) characteristics of the representation language, types are specified as sets of features, dimensions in a multidimensional space, or formal restrictions specified on sets using various kinds of description logics. While some representation languages have an underlying formal semantics (e.g., model theory), the grounding of several representation languages remains on the level of an informal description. As the compared types are representations of concepts in human minds, similarity depends on what is said (in terms of computational representation) about these types. This again is connected to the chosen language, leading to the fact that most similarity theories cannot be compared. Beside the question of representation, context is another major challenge for similarity assessments. In many cases, meaningful notions of similarity cannot be determined without defining in respect to what similarity is measured [7–10].

Similarity has been widely applied within GIScience over the past few years. Based on Tversky's feature model [11], Rodríguez and Egenhofer [3] developed the Matching Distance Similarity Measure that supports a basic context theory, automatically determined weights, and asymmetry. Raubal and Schwering [12, 13] used so-called conceptual spaces to implement models based on distance measures within geometric space, while Sunna and Cruz [14] applied a network based similarity measure for ontology alignment. Several measures [4, 15, 16] were developed to close the gap between (geo-)ontologies described by various kinds of description logics, and similarity theories that had not been able to handle the expressivity of such languages. Other similarity theories [17, 18] have been established to determine the similarity between spatial scenes. The ConceptVISTA [19] ontology management and visualization toolkit uses similarity for knowledge retrieval and organization.

2.2 Gazetteer Research

Gazetteers are knowledge organization systems for spatial information. They deliver feature types and geographic footprints for searched place names [20]. The categorization into feature types is crucial for geographic information retrieval, as it enables concept-based queries and reasoning in the first place (e.g., when looking for *villages in Catalunya*). However, current gazetteers such as the Alexandra Digital Library (ADL) Gazetteer[1] are based on semi-formal feature type thesauri with limited support for formal reasoning methods.

The human participants test described in this paper has been carried out on a subset of a geographic feature type *ontology* (FTO). The FTO is currently being developed based on the ADL feature type thesaurus to demonstrate the benefits of subsumption and similarity based reasoning for Gazetteers [6]. Such an ontology provides support for innovative user interfaces as discussed in section 3.3. Subsumption and similarity reasoning allow the user to intuitively *browse* the gazetteer, continuously being provided with all relevant information for the current query. Moreover, as outlined in section 1, they allow for new approaches towards gazetteer interoperability.

To ensure that such a user interface actually returns similar results as expected by the user, the cognitive plausibility of the similarity theory must be approved. In the following, we will outline the basic characteristics of the similarity theory SIM-DL, and discuss the human participants test and its results.

3 SIM-DL Theory, Implementation and Application

This section gives a brief insight into the SIM-DL similarity measurement theory, its implementation within the SIM-DL server, and the gazetteer web interface.

3.1 SIM-DL Theory

SIM-DL [4, 5, 21] is an asymmetric and context aware similarity measurement theory used for information retrieval and organization. It compares a search concept C_s with target concepts $\{C_t\}$ from an ontology (or several ontologies using the same shared vocabulary). The concepts themselves can be specified using various kinds of expressive description logics [22].

Within SIM-DL, similarity between concepts in canonical form [4, 23] is measured by comparing their definitions for overlap, where a high level of overlap indicates high similarity and vice versa. Description logics concepts are specified based on primitive concepts and roles using language constructors such as

[1] http://www.alexandria.ucsb.edu/gazetteer/

intersection, union, and existential quantification. Hence, similarity is defined as a polymorphic, binary, and real-valued function $C_s \times C_t \rightarrow$ R[0,1] providing implementations for all language constructs offered by the used description logics. The overall similarity between concepts is the normalized (and weighted) sum of the single similarities calculated for all parts (i.e., superconcepts) of the concept definitions. A similarity value of 1 indicates that the compared concepts cannot be differentiated, whereas 0 implies total dissimilarity. As most feature and geometric approaches, SIM-DL is an asymmetric measure, i.e., the similarity $sim(C_s, C_t)$ is not necessarily equal to $sim(C_t, C_s)$. Therefore, the comparison of two concepts does not only depend on their descriptors, but also on the direction in which both are compared.

A single similarity value (e.g., 0.67) computed between two concepts hides most of the important information. It does not answer the question whether there are more or less similar target concepts in the examined ontology. It is not sufficient to know that possible similarity values range from 0 to 1 as long as their distribution is unclear. Imagine an ontology where the least similar target concept has a value of 0.6 (compared to the source concept), while the comparison to the most similar concept yields 0.9. In this case, a similarity value of 0.67 is not high at all. Beside these interpretation problems, isolated comparison puts too much stress on the concrete similarity value. It is hard to argue that and why the result is (cognitively) plausible without other reference values [24].

Accordingly, SIM-DL focuses on similarity rankings. The search concept is compared to all target concepts derived from the measurement context [4, 9, 25, 26], i.e., a subset of the ontology, also referred to as the domain of application. The result is an ordered list with descending similarity values. Consequently, in the following we do not argue that single similarity values are cognitively plausible, but that the computed order correlates with human ranking judgments.

3.2 SIM-DL Server

The SIM-DL similarity server and a client plug-in for the Protégé Ontology Editor[2] are available as an open-source cross-platform project at SourceForge.net. The current beta version[3] supports basic reasoning services (e.g., subsumption reasoning) and similarity measurement up to $\mathcal{ALCHQ}$[4] [5, 22].

The reasoning component implements a tableaux algorithm to determine TBox subsumption based on ABox satisfiability, while the similarity component is based on the SIM-DL framework and theory. Each similarity request involves interaction with the reasoning component to determine all target concepts in the context. Furthermore, the reasoner is required for several similarity functions and optimization.

---

[2] http://protege.stanford.edu/

[3] The current release can be downloaded at http://sim-dl.sourceforge.net/.

[4] See http://www.cs.man.ac.uk/~ezolin/dl/ and [5] for more details about the used description logic and its computational characteristics.

The SIM-DL server interprets incoming requests and starts the similarity and reasoning engines. The requests conform to the DIG 1.1 specification [27] which provides a standardized XML-based interface for reasoning services. In our previous work, the DIG interface has been extended in order to support similarity measurement between concepts [5]. Within this paper, the server is used to compute the similarity values for the compared concepts and to interact with the new gazetteer web interface.

3.3 Application Scenario: Gazetteer Web Interface

The Gazetteer Web Interface connects the SIM-DL Server with the Alexandria Digital Library (ADL) gazetteer offering an enhanced search mechanism for geographic features. It is realized as a mashup combining an auto-suggest input field for feature types, an input field for feature names, and Google Maps™. The map is used to restrict a query's spatial extent and to display matching features retrieved by the gazetteer.

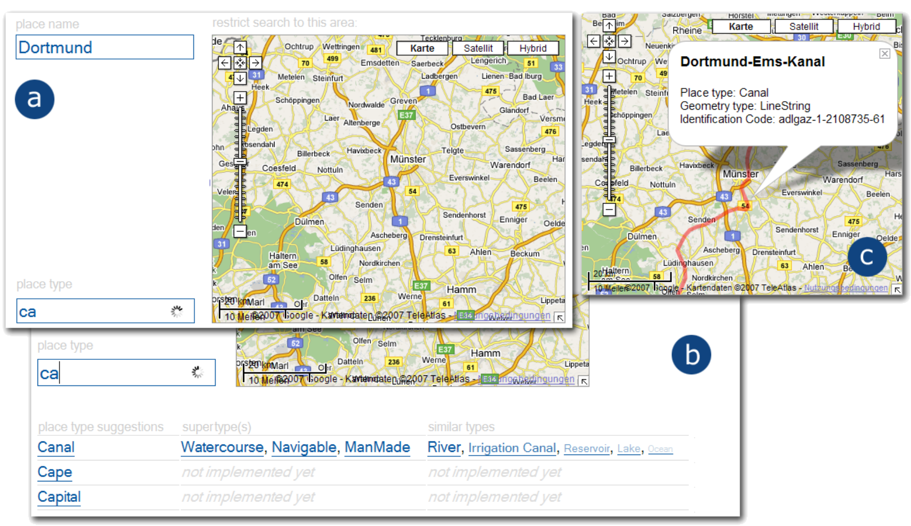

Fig. 2. The similarity enabled gazetteer web interface [6].

The intention with regard to the development of this interface is to optimize the gazetteer request procedure using similarity reasoning. In contrast to the standard ADL front-end, it does not require knowledge of the gazetteer's internal feature type thesaurus (FTT) hierarchy. The SIM-DL server uses an ontology (FTO) extending the FTT hierarchy that provides the concepts and relations being utilized within the SIM-DL similarity measurement process. The auto-suggest text field used for searching feature types is based on *Asynchronous Javascript and XML* (AJAX) technology: as the user enters the name of the requested feature type, feature types matching the letters entered so far are automatically retrieved and displayed. The suggestions returned by the SIM-DL server consist of the suggested type itself, its super types, and similar types. The similar types are presented in different font sizes, reflecting their similarity to the suggested concept. Comparable to *tag clouds*, the bigger a concept is displayed, the more similar it is. All suggestions are hyperlinked and are shifted into the input field when clicked.

Beside the selection of a feature type and the definition of an area of interest, users can also search by place names, such as the *Dortmund-Ems Canal*. The results are displayed on the selected map extent as shown in figure 2.

4 Evaluation

This section presents the human participants test that has been performed to prove the cognitive plausibility of SIM-DL similarity rankings. The test results are evaluated and discussed.

4.1 Motivation

SIM-DL is intended to measure similarity between computational representations of concepts. The motivation is to improve the accessibility of tasks such as information retrieval and organization for human users. This can only be achieved if there is a high correlation between the similarity rankings calculated by SIM-DL and human similarity judgments. The SIM-DL measurement process has been developed based on findings from cognitive science. It takes aspects such as asymmetry, alignment, and context into account which are known to play an important role for human similarity ratings. SIM-DL tries to approximate aspects from the human process of reasoning about similarity to achieve meaningful results. Nevertheless, it is a computational theory for description logics rather than a framework towards understanding cognitive processes. Consequently, we neither claim that SIM-DL models (or even explains) the process of human similarity judgments nor that humans represent concepts (if they do) in any kind of logic based serialization.

Figure 3 illustrates the relation between a similarity reasoning service such as the SIM-DL server and human reasoning about similarity. The box at the top represents the cognitive process (marked as dotted line) of deriving similarity judgments. Without discussing whether there is something such as representation and cognition (or only perception) [28, 29], up to now no direct mapping to computational representations is possible. Similarity theories developed in cognitive science model (i.e., approximate) this process by partitioning it into observable units. The effect of each unit is studied by changing its settings, while all other units remain stable[5]. Such units include Context, Alignment, Asymmetry, and the Max-Effect [7]. Each of them is depicted as a box on the

[5] or by studying patients with lesions.

dotted process line to indicate that they are fragments of the whole process. Most theories from cognitive science focus on the explanation of human similarity reasoning rather than the development of executable services[6]. Consequently, the chosen representation is more on the informal side. In contrast, information science is interested in computational representations to provide a basis for executable theories. While these theories try to approximate cognitive theories, their goal is not explanatory. Instead, they adopt such units that can be computed with appropriate resources. From this point of view, computational models form a subset of theories established in cognitive science. Typical application areas include human computer interaction and information retrieval.

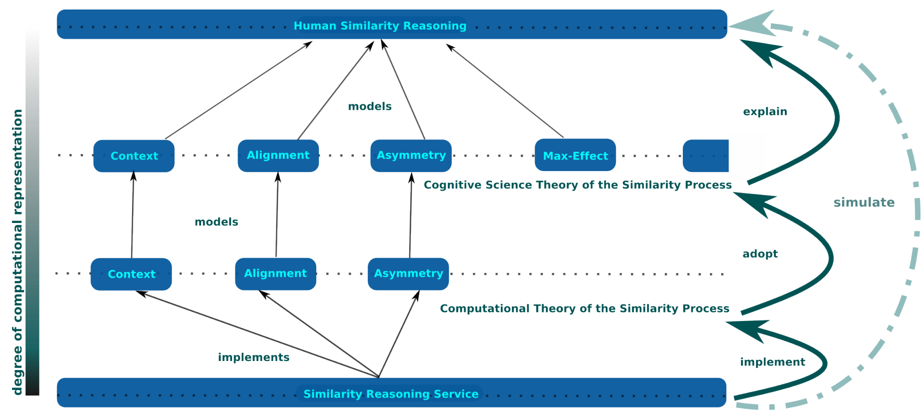

Fig. 3. From human similarity reasoning to similarity services such as the SIM-DL similarity server.

The box at the bottom of figure 3 represents concrete similarity reasoning services such as the SIM-DL similarity server. These services implement the computational theories as standalone applications or as parts of a knowledge infrastructure such as the ConceptVISTA[7] toolbox. The motivation for developing similarity-aware applications is to simulate human similarity judgement, thus making tasks such as information retrieval more accessible to the user. It is important to note that not the cognitive process is simulated, but the final similarity ranking, i.e., the reasoning results. The dashed arrow indicates that there is no direct link between the similarity service and human similarity judgments. Computational similarity ratings depend on how compared entities and concepts are represented and which units (parts) of the human similarity process are modeled within the implemented computational theories.

The term *cognitively plausible* will be used, if the similarity rankings produced using SIM-DL correlate with human similarity rankings. In contrast, *cognitively*

[6] For some exceptions, see SME and MAC/FAC [30, 31].

[7] http://www.geovista.psu.edu/ConceptVISTA/

adequate would require a comparison of the underlying processes (their units) and is out of scope of this work.

4.2 Test Setting

A human participants test has been performed to prove that the results calculated by the SIM-DL theory introduced in section 3 correlate with human similarity judgments. Very few objective tests have been carried out so far concerning the usage of similarity measurement in practice. Due to the fact that the participants were all native German speakers, the test was in German, too. In the following, we will translate the German parts of the test into the best-fitting English expressions where it is necessary for the understanding.

28 participants were recruited for the human participants test. The group of participants consisted of 16 males and 12 females. The mean age of the 28 participants was 27.3 with a range from 22 to 31 years. The mean female age was 26.4 and the mean age of the males was 27.8. The questionnaire[8] was distributed randomly among the participants [32].

The first step for every participant was to read the introduction, consisting of a brief motivation for the test, and instructions on how to fill it in. According to [33], written instructions are preferred by participants over spoken instructions. Next, every participant was asked to read the concept descriptions of the given feature types: the named search concept *Canal* (ger.: Kanal) and a set of anonymous target concepts (figure 4). Every participant was requested to assess the similarity between the description of the search concept and every description of the target concepts by placing a mark between a line ranging from minimum to maximum similarity. Additionally, the participants made a statement how confident they felt when placing the mark using a discrete scale with five classes from *not sure* (ger.: nicht sicher) to *sure* (ger.: sicher). It is assumed that a continuous scale for assessing the concept similarity is reasonable due to the provided granularity[9] which is not required for the confidence assessments. The range for the continuous scale went from minimum similarity (ger.: minimale Ähnlichkeit) to maximum similarity (ger.: maximale Ähnlichkeit). The reason for omitting the names of the target concepts was to ensure that the similarity judgments only depend on the concept descriptions and are not biased by the participants individual conceptualizations.

In the final step, the participants were asked to assign a given list of (concept) names to the anonymous concept descriptions. This final step was introduced to check whether the presented concept descriptions corresponds to the participants conceptualization. Moreover, wrong assignments of the concept names are a strong hint that the test was filled in randomly and thus useless for the evaluation; this check was considered necessary as there was no financial compensation for the participants' effort.

[8] The questionnaire is available at http://sim-dl.sourceforge.net/downloads/.

[9] For example, to allow statements such as "the similarity between *Canal* and concept A is almost equal to the similarity between *Canal* and concept B, but the former seems to be a bit higher".

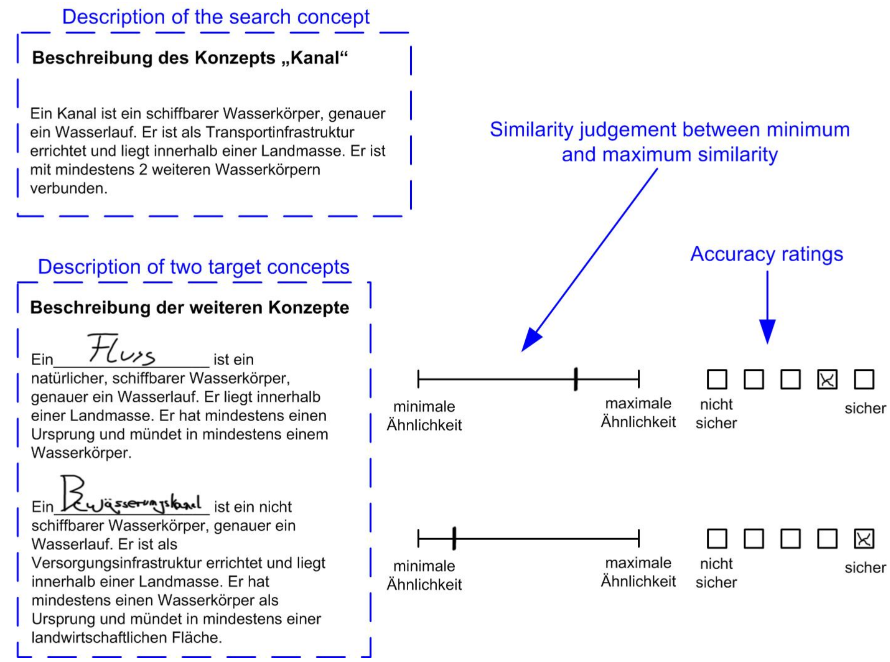

Fig. 4. Section of the questionnaire, showing the search concept *Kanal* (eng. Canal: “A canal is a navigable body of water, namely a watercourse. It is constructed as transport-infrastructure, that is inside landmass. It is connected to at least two other bodies of water”) and two of the six target concepts, *Fluss* (eng. River: “A river is a natural, navigable body of water, namely a watercourse. It is inside landmass. It is connected to at least one spring as origin and at least one body of water as destination”) and *Bewässerungskanal* (eng. Irrigation Canal: “An irrigation canal is a non-navigable body of water, namely a watercourse. It is constructed for supply and infrastructure and is inside landmass. It is connected to at least one body of water as origin and at least one agricultural area as destination.”)

While a detailed insight into the underlying feature type ontology used for similarity reasoning is out of scope here, the following example demonstrates how the concepts were specified.

$$
\begin{aligned}
Canal \sqsubseteq\, & WaterBody \sqcap Watercourse \sqcap Navigable \sqcap (\exists inside.Landmass) \\
& \sqcap (\exists constructedAs.Transportation) \\
& \sqcap (\geq 2\, connectedTo.Waterbody)
\end{aligned}
$$

While some concepts used to describe *Canal* are primitives (e.g., *Navigable*), other are defined within the ontology. For instance, *WaterBody* is a subconcept of *HydrographicFeature*. Note that for reasons of simplification, and to keep the cognitive load low, the plain text descriptions presented to the participants hide some details. Beside their role in transportation, canals can also be constructed for additional proposes[10].

4.3 Results

Out of the 28 questionnaires, 26 were taken for further processing. First, it was checked whether the concept names were properly assigned to the descriptions. All 26 questionnaires satisfy this requirement, however, several participants made updates (changed the names) while performing the test. Next, the similarity values and confidence assessments were transformed to values and weights,respectively, between 0 and 1. Each confidence box corresponds to a weighting step of 0.2. The first box was transformed to 0.2, the second to 0.4, and so on. Table 1 shows the absolute similarity values obtained using the SIM-DL similarity server, the arithmetic mean of the human similarity judgments, and the weighted mean using the confidence assessments.

Table 1. Mean (absolute) similarity judgments by test subjects, compared to SIM-DL.

Concept	Fluss (River)	Bewässerungskanal (Irrigation Canal)	Stausee (Reservoir)	See (Lake)	Ozean (Ocean)	Förderplattform (Offshore Platform)
SIM-DL server	0.75	0.67	0.58	0.5	0.38	0.08
Arithm. mean	0.7	0.53	0.59	0.43	0.33	0.14
Weighted mean	0.72	0.55	0.6	0.43	0.32	0.13

In a next step, the absolute similarity values from each questionnaire were transformed to ordinal scale, i.e., into a descending similarity ranking. The most similar concept (with respect to *Canal*) was ranked 6, while the least similar got the rank 1. If two or more concepts had the same absolute similarity values, a mean rank (tie) was chosen (e.g., 4.5). The weights have no influence on the ranking position. Figure 5 shows the resulting box plot for the 26 questionnaires. It depicts the lowest non-outlier ranking, the lower quartile (25%), the median, upper quartile (75%), and highest non-outlier ranking per target concept. The stars and dots represent mild and extreme outliers. *River*, *Reservoir*, *Lake*, and *Ocean* have a comparable interquartile range, while the boxplot for the *Offshore Platform* is collapsed. In contrast, the *Irrigation Canal* boxplot shows a high distribution among test subjects.

As depicted in table 2, the individual ranking data from each questionnaire was used to compute the median and mode for each target concept. In both cases, the resulting order corresponds to the computed similarity ranking except that *River* and *Irrigation Canal* share the same rank. In terms of frequencies,

[10] This does not influence similarity, as the same mapping was performed for all concepts used for the test and the participants had to compare the descriptions (not knowing which concepts were actually described).

Table 2. Median and mode similarity ranks for the target concepts based on the test results.

		Fluss (River)	Bewässerungskanal (Irrigation Canal)	Stausee (Reservoir)	See (Lake)	Ozean (Ocean)	Förderplattform (Offshore Platform)
N	Valid	26	26	26	26	26	26
	Missing	0	0	0	0	0	0
Median		5.0000	5.0000	4.0000	3.0000	2.0000	1.0000
Mode		5.00^a	6.00	4.00	3.00	2.00	1.00
frequency (#)							
6th rank		**12**	**8**	6	0	0	0
5th rank		**12**	6	4	1	1	0
4.5th rank[b]		1	-	1	-	-	-
4th rank		1	2	**12**	10	1	0
3rd rank		0	3	1	**14**	7	2
2nd rank		0	3	2	1	**16**	3
1st rank		0	4	0	0	1	**21**

a: Multiple modes exist (5 and 6). The smallest value is shown.
b: This rank is caused by the normalized ranking process of SPSS.

this means that the majority of test subjects has chosen the same rank as SIM-DL for *Reservoir*, *Lake*, *Ocean*, and *Offshore Platform*. In case of *River*, the same number of participants had chosen the 6^{th} and 5^{th} rank (12 times), while SIM-DL ranks *River* as most similar concept to *Canal* (6^{th} rank). The remaining two participants selected the 4^{th} rank. While the median for *Irrigation Canal* corresponds to the computed 5^{th} rank, the mode is 6. This is caused by the high dispersion for this concept. The human rankings range from the first (4 times) up to the sixth rank (8 times).

A correlation analysis between the median human similarity ranking and the ranking computed by SIM-DL yields $r_s = 0.986$ ($p < 0.01$) using Spearman's ρ. As depicted in figure 5, the data is not normally distributed, i.e., skewed. In addition, we cannot assume equi-distance between the ranks. Hence, the correlation was also determined using Kendall's τ and yields 0.966 ($p < 0.01$).

To measure the consensus among participants with respect to the chosen rank, Kendall's coefficient of concordance W was used. To determine whether an obtained W value is significant, chi-square was computed for given degrees of freedom and compared to significance tables for probability. The analysis (taking the ties from the ranking process into account) yields a value of 0.632 for W with a *Chisq(5)* of 82.1. If we hypothesize that the participant's ranks are associated, this corresponds to a proability of $p < 0.001$ that we accept the hypothesis while it is false. Consequently, and with respect to the high number of participants, the results are significant.

4.4 Discussion

The test shows a strong and significant correlation between human similarity rankings and those obtained using the SIM-DL similarity server. Based on our previous definition, the computed similarity judgments can be called cognitively plausible. The correspondence between the absolute similarity values is difficult

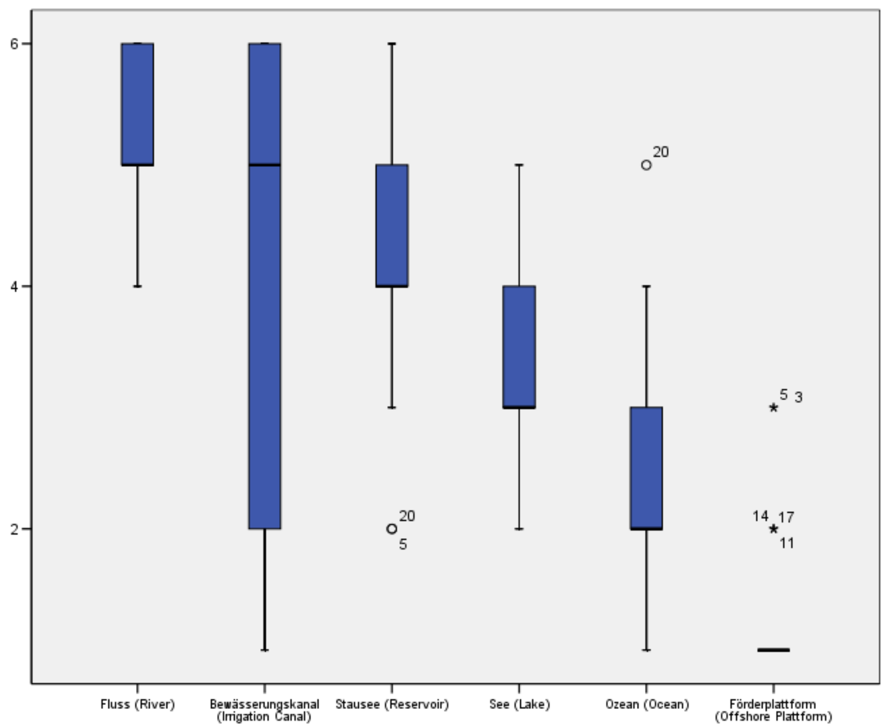

Fig. 5. Boxplot showing the human similarity rankings and their dispersion.

to interpret. Each participant has its own (cognitive) similarity scale and distribution, i.e., the similarity value for the most and least similar concept differs between participants. For instance, the absolute values for the concept *River* range from 0.93 to 0.73 for participants that had chosen *River* to be the most similar concept to *Canal*. Overall, SIM-DL values are close to the (weighted) mean similarity judgments, but tend to overestimate.

While these results look promising, the interquartile ranges raise some questions. This becomes especially apparent in case of *Irrigation Canal* and partly also for *Reservoir*. In the first case, while most participants had chosen a high similarity (5^{th} or 6^{th} rank), several subjects ranked *Irrigation Canal* as very dissimilar. There may be two potential explanations for these results. Out of all compared concept descriptions, *Irrigation Canal* is the only one specified as a *non-navigable body of water*, while all others (except *Offshore Platform*) are *navigable*. When subjects compare *Irrigation Canal* to *Canal*, they use the previously made similarity judgments as points of reference. While *Offshore Platform* is too different to serve as a reference, all other concepts share a feature that is missing for *Irrigation Canal*. In this case *navigable* becomes the characteristic feature of the set of compared concepts and gets a high weighting. This explanation corresponds well to the variability context weighting[11] proposed by Rodrígez and Egenhofer [3] as well as to Tversky's notion of diagnosticity [11]. Tversky argues that features which are diagnostic for a particular classification have a disproportionate influence on similarity judgments.

A second explanation could be based on different kinds of information processing and extraction. One has to keep in mind that while the similarity server and the participants share the same information about the presented concepts, their representation is different (plaintext versus description logics). The similarity ranking task involves some deductive reasoning steps. For instance, canals were defined as entities which are connected to at least two bodies of water, while rivers have at least one waterbody as origin and one waterbody as destination. The underlying ontology represents this using the three relations *connectedTo* and its sub-relations *hasOrigin* and *hasDestination*. When searching for entities connected to waterbodies, an entity with a waterbody as origin satisfies this requirement and should be similar. Participants seem to perform this kind of reasoning and therefore assign a high rank to *River*. In contrast, irrigation canals have at least one waterbody as origin and one agricultural area as destination. Instead of judging the origin and destination separately, participants may summarize both to a non matching feature [11].

5 Conclusions and Further Work

Based on the performed human participants test, the SIM-DL theory returns cognitively plausible similarity rankings. To ensure that both the human and the computer similarity judgments are based on the descriptions of the concepts, i.e., their representations, the concept names were left blank during the first step of the test. Accordingly, the participants had to assess the similarity of the concept *descriptions*, instead of their own conceptualizations. The complexity of the ontology used for the test was limited to a small number of concepts that were specified only by their most important characteristics to avoid a cognitive overload for the participants. Future work needs to investigate how more complex ontologies can be presented during a human participants test.

While the test shows that the similarity rankings correlate, it does not answer the question whether their integration and visualization within the proposed gazetteer web interface improves usability. Strictly speaking, one should also not argue that the rankings delivered by the Gazetteer Web Interface correspond to human similarity judgments. The motivation for using a gazetteer might put the focus on other parts of the concept description and hence influence similarity. Consequently, the next step has to be an evaluation of the web interface. Moreover, the feature type ontology needs to be extended by more concepts and more detailed specifications to demonstrate that the developed methods are able to cope with larger information bases.

[11] Up to now, SIM-DL only supports a context weighting comparable to the commonality weighting in MDSM.

Parts of the SIM-DL theory have been used within other projects such as a web service for identity assumptions for historical places. As SIM-DL has no own visualization module, an integration within the ConceptVISTA toolkit might be a promising next step.

Acknowledgments

This work is funded by the project *Semantic Similarity Measurement for Role-Governed Geospatial Categories* (*SimCat*) granted by the German Research Foundation (DFG Ra1062/2-1).

References

1. Goldstone, R., Son, J.: Similarity. In Holyoak, K., Morrison, R., eds.: Cambridge Handbook of Thinking and Reasoning. Cambridge University Press (2005)
2. Rissland, E.L.: Ai and similarity. IEEE Intelligent Systems **21**(3) (2006) 39–49
3. Rodríguez, A., Egenhofer, M.: Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. International Journal of Geographical Information Science **18**(3) (2004) 229–256
4. Janowicz, K.: Sim-dl: Towards a semantic similarity measurement theory for the description logic $\mathcal{ALCNR}$ in geographic information retrieval. In Meersman, R., Tari, Z., Herrero, P., al., e., eds.: SeBGIS 2006, OTM Workshops 2006. Volume 4278 of Lecture Notes in Computer Science. Springer, Berlin (2006) 1681 – 1692
5. Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., Bäumer, B.: Algorithm, implementation and application of the sim-dl similarity server. In Fonseca, F., Rodríguez, M.A., eds.: Second International Conference on GeoSpatial Semantics, GeoS 2007. Lecture Notes in Computer Science 4853, Springer-Verlag Berlin Heidelberg (2007) 128–145
6. Janowicz, K., Keßler, C.: The role of ontology in improving gazetteer interaction. International Journal of Geographical Information Science (accepted for publication 2008)
7. Medin, D., Goldstone, R., Gentner, D.: Respects for similarity. Psychological Review **100**(2) (1993) 254–278
8. Goodman, N.: Seven strictures on similarity. In Goodman, N., ed.: Problems and projects. Bobbs-Merrill, New York (1972) 437–447
9. Keßler, C.: Similarity measurement in context. In Kokinov, B., Richardson, D.C., Roth-Berghofer, T.R., Vieu, L., eds.: 6th International and Interdisciplinary Conference, CONTEXT 2007. Lecture Notes in Artificial Intelligence 4635, Roskilde, Denmark, Springer-Verlag Berlin Heidelberg (2007) 277–290
10. Frank, A.U.: Similarity measures for semantics: What is observed? In: COSIT'07 Workshop on Semantic Similarity Measurement and Geospatial Applications, Melbourne, Australia (September 2007)
11. Tversky, A.: Features of similarity. Psychological Review **84**(4) (1977) 327–352
12. Raubal, M.: Formalizing conceptual spaces. In Varzi, A., Vieu, L., eds.: Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Volume 114 of Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam, NL (2004) 153–164

13. Schwering, A., Raubal, M.: Spatial relations for semantic similarity measurement. In Akoka, J., Liddle, S., Song, I.Y., Bertolotto, M., Comyn-Wattiau, I., vanden Heuvel, W.J., Kolp, M., Trujillo, J., Kop, C., Mayr, H., eds.: Perspectives in Conceptual Modeling: ER 2005 CoMoGIS Workshop, Klagenfurt, Austria. Volume 3770 of Lecture Notes in Computer Science. Springer, Berlin (2005) 259–269
14. Sunna, W., Cruz, I.: Using the agreementmaker to align ontologies for the oaei campaign 2007. In: The Second International Workshop on Ontology Matching, collocated with the 6th International Semantic Web Conference ISWC, Busan, Korea. (November 2007)
15. d'Amato, C., Fanizzi, N., Esposito, F.: A dissimilarity measure for $\mathcal{ALC}$ concept descriptions. In: Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France (2006) 1695–1699
16. Borgida, A., Walsh, T., Hirsh, H.: Towards measuring similarity in description logics. In: Proceedings of the 2005 International Workshop on Description Logics (DL2005). Volume 147 of CEUR Workshop Proceedings. CEUR, Edinburgh, Scotland, UK (2005)
17. Li, B., Fonseca, F.: Tdd - a comprehensive model for qualitative spatial similarity assessment. Spatial Cognition and Computation **6**(1) (2006) 31–62
18. Nedas, K., Egenhofer, M.: Spatial similarity queries with logical operators. In Hadzilacos, T., Manolopoulos, Y., Roddick, J., Theodoridis, Y., eds.: SSTD '03 - Eighth International Symposium on Spatial and Temporal Databases, Santorini, Greece. Volume 2750 of Lecture Notes in Computer Science. (2003) 430–448
19. Gahegan, M., Agrawal, R., Banchuen, T., DiBiase, D.: Building rich, semantic descriptions of learning activities to facilitate reuse in digital libraries. International Journal on Digital Libraries **7**(1) (2007) 81–97
20. Hill, L.L.: Georeferencing: The Geographic Associations of Information (Digital Libraries and Electronic Publishing). The MIT Press (2006)
21. Janowicz, K.: Similarity-based retrieval for geospatial semantic web services specified using the web service modeling language (wsml-core). In Scharl, A., Tochtermann, K., eds.: The Geospatial Web - How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society. Lecture Notes in Computer Science. Springer, Berlin (2007)
22. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
23. Horrocks, I.: Implementation and optimization techniques. In: The description logic handbook: theory, implementation, and applications. Cambridge University Press, New York, NY, USA (2003) 306–346
24. Jurisica, I.: 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)
25. Keßler, C., Raubal, M., Janowicz, K.: The effect of context on semantic similarity measurement. In: Proceedings of the 3rd International IFIP Workshop On Semantic Web & Web Semantics (SWWS '07). Lecture Notes in Computer Science, Vilamoura, Portuga, Springer Verlag (2007; forthcoming)
26. Janowicz, K.: Kinds of contexts and their impact on semantic similarity measurement. In: 5th IEEE Workshop on Context Modeling and Reasoning (CoMoRea) at the 6th IEEE International Conference on Pervasive Computing and Communication (PerCom'08), Hong Kong, IEEE Computer Society (March 2008; forthcoming)
27. Bechhofer, S.: The dig description logic interface: Dig/1.1. In: DL2003 Workshop, Rome, Italy (2003)

28. Gibson, J.: The theory of affordances. In Shaw, R., Bransford, J., eds.: Perceiving, Acting, and Knowing - Toward an Ecological Psychology. Lawrence Erlbaum Ass., Hillsdale, New Jersey (1977) 67–82
29. Markman, A. B. & Dietrich, E.: In defense of representation. Cognitive Psychology **40** (2000) 138–171
30. Falkenhainer, B., Forbus, K., Gentner, D.: The structure-mapping engine: Algorithm and examples. Artificial Intelligence **41** (1989) 1–63
31. Gentner, D., Forbus, K.D.: MAC/FAC: a model of similarity-based retrieval. In: Proceedings of the 13th Cognitive Science Conference, Chicago, Erlbaum, Hillsdale (1991) 504–509
32. Montello, D., Sutton, P.: An Introduction to Scientific Research Methods in Geography. Sage Publications Ltd. (2006)
33. Harrison, S.: 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, ACM Press/Addison-Wesley Publishing Co. New York, NY, USA (1995) 82–89