JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 2 (2011), pp. 29–57
doi:10.5311/JOSIS.2011.2.3

RESEARCH ARTICLE

The semantics of similarity in geographic information retrieval

Krzysztof Janowicz[1], Martin Raubal[2], and Werner Kuhn[3]

[1]Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA
[2]Department of Geography, University of California, Santa Barbara, CA 93106, USA
[3]Institute for Geoinformatics, University of Münster, D-48151 Münster, Germany

Received: May 24, 2010; returned: July 7, 2010; revised: November 8, 2010; accepted: December 21, 2010.

Abstract: Similarity measures have a long tradition in fields such as information retrieval, artificial intelligence, and cognitive science. Within the last years, these measures have been extended and reused to measure semantic similarity; i.e., for comparing meanings rather than syntactic differences. Various measures for spatial applications have been developed, but a solid foundation for answering what they measure; how they are best applied in information retrieval; which role contextual information plays; and how similarity values or rankings should be interpreted is still missing. It is therefore difficult to decide which measure should be used for a particular application or to compare results from different similarity theories. Based on a review of existing similarity measures, we introduce a framework to specify the semantics of similarity. We discuss similarity-based information retrieval paradigms as well as their implementation in web-based user interfaces for geographic information retrieval to demonstrate the applicability of the framework. Finally, we formulate open challenges for similarity research.

Keywords: semantic similarity, geographic information retrieval, ontology, similarity measure, context, relevance, description logic, user interface

1 Introduction and motivation

Similarity measures belong to the classical approaches to information retrieval and have been successfully applied for many years, increasingly also in the domain of spatial information [82]. While they have been working previously in the background of search engines, similarity measures are nowadays becoming more visible and are integrated into user interfaces of modern search engines. A majority of these measures are purely syntactical, rely on statistical measures or linguistic models, and are restricted to unstructured data such as text documents. Lately, the role of similarity measures in searching and browsing multimedia content, such as images or videos has been growing [59]. Similarity measures have also been studied intensively in cognitive science and artificial intelligence [80] for more than 40 years. In contrast to information retrieval, these domains investigate similarity to learn about human cognition, reasoning, and categorization [31] from studying differences and commonalities in human conceptualizations. Similarity measures have also become popular in the Semantic (geospatial) Web [20]. They are being applied to compare concepts, to improve searching and browsing through ontologies, as well as for matching and aligning ontologies [84]. In GIScience, similarity measures play a core role in understanding and handling semantic heterogeneity and, hence, in enabling interoperability between services and data repositories on the Web. In his classic book *Gödel, Escher, Bach—An Eternal Golden Braid*, Hofstadter named among other facts the abilities "to find similarities between situations despite differences which may separate them [and] to draw distinctions between situations despite similarities which may link them" as major characteristics of (human) intelligence [38, p.26].

Modern similarity measures are neither restricted to purely structural approaches nor to simple network measures within a subsumption hierarchy. They compute the conceptual overlap between arbitrary concepts and relations, and, hence, narrow the gap between similarity and analogy. To emphasize this difference, they are often referred to as *semantic* similarity measures. Similar to syntactic measures, they are increasingly integrated into front-ends such as semantically enabled gazetteer interfaces [44]. In contrast to subsumption-based approaches, similarity reasoning is more flexible in supporting users during information retrieval. Most applications that handle fuzzy or ambiguous input—either from human beings or from software agents—potentially benefit from similarity reasoning.

However, the interpretation of similarity values is not trivial. While the number of measures and applications is increasing, there is no appropriate theoretical underpinning to explain what they measure, how they can be compared, and which of them should be chosen to solve a particular task. In a nutshell, the challenge is to make the semantics of similarity explicit. Abstracting from various existing theories, we propose a generic framework for similarity measures, supporting the study of these and related questions. In our work and review we focus on inter-concept similarity and particularly on comparing classes in ontologies. While the methods to measure inter-concept and inter-instance similarity overlap, the former is more challenging. This is mainly for two reasons. First, in contrast to data on individuals, ontologies describe multiple potential interpretations. For instance, there is no single graph describing a concept in an OWL-based ontology [37]. Secondly, an interpretation may have an infinite number of elements and, hence, may describe an infinite graph.

The remainder of this article is structured as follows. First we introduce related work on geographic information retrieval and semantic similarity measurement. Next, we propose a generic framework and elucidate the introduced steps by examples from popular similarity theories. While we focus on inter-concept similarity, the framework has also been successfully adapted to inter-instance measures [90], and, moreover, can be generalized to the comparison of spatial scenes [60,72]. We then discuss the role of similarity in semantics-based information retrieval and show its integration into user interfaces. We conclude by pointing to open research questions.

2 Related work

This section introduces geographic information retrieval, similarity measurement, and points to related work.

2.1 Geographic information retrieval

Information retrieval (IR) is a broad and interdisciplinary research field including information indexing, relevance rankings, search engines, evaluation measures such as recall and precision, as well as robust information carriers and efficient storage. In its broadest definition, information retrieval is concerned with finding relevant information based on a user's query [18]. Here, we focus on the relevance relationship and leave other aspects such as indexing aside. Following Dominich [18], information retrieval can be formalized as:

$$IR = m[\Re(O, (Q, \langle I, \mapsto \rangle))] \quad (1)$$

where

- $\Re$ is the relevance relationship,
- O is a set of objects,
- Q is the user's query,
- I is implicit information,
- $\mapsto$ is inferred information, and
- m is the degree (or certainty) of relevance.

Accordingly, information retrieval is about computing the degree of relevance between a set of objects, such as web pages, and the search parameters, e.g., keywords, specified by the user. Besides defining suitable relevance measures, the main challenge for information retrieval is that "we are asking the computer to supply the information we want, instead of the information we asked for. In short, users are asking the computer to reason intuitively" [10, p.1]. Not all information relevant to a search can be entered into the retrieval system. For instance, classical search engines offer a single text field to enter keywords or phrases. Implicit information, such as the user's age, cultural background, or the task motivating the search are not part of the query. Some of this implicit information can be inferred and used for the relevance rankings. In case of search engines for the web, the language settings of the browser or the IP-address reveal additional information about the user.

Geographic information retrieval (GIR) adds space and sometimes time as dimensions to the classical retrieval problem. For instance, a query for "pubs in the historic center of Münster" requires a thematic and a spatial matching between the data and the user's query. According to Jones and Purves [50], GIR considers the following steps. First, the geographic references have to be recognized and extracted from the user's query or a document using methods such as named entity recognition and geo-parsing. Second, place names are not unique and the GIR system has to decide which interpretation is intended by the user. Third, geographic references are often vague; typical examples are vernacular names ("historic center") and fuzzy geographic footprints. In case of the pub query, the GIR system has to select the correct boundaries of the historic center [71]. Fourth, and in contrast to classical IR, documents also have to be indexed according to particular geographic regions. Finally, geographic relevance rankings extend existing relevance measures with a spatial component. The ranking of instances does not only depend on thematic aspects, e.g., the pubs, but also on their location, e.g., their distance to the historic center of Münster.

2.2 Semantic similarity measurement

Research on similarity investigates commonalities and differences between individuals or classes. Most similarity measures originated in psychology and were established to determine why and how individuals are grouped into categories, and why some categories are comparable to each other while others are not [31, 69]. The following approaches to semantic similarity measurement can be distinguished: feature-based, alignment-based, network-based, transformational, geometric, and information theoretic (see [31] for details).

These similarity measures are either syntax- or semantics-based. Classical examples for syntactic similarity measures are those which compare literals, such as edit-distance; but there are also more complex theories. The main challenge for *semantic* similarity measures is the comparison of meaning as opposed to structure. Lacking direct access to individuals or categories in the world, any computation of similarity rests on terms expressing concepts. Semantic similarity measures use specifications of these concepts taken from ontologies [34]. These may involve (unstructured) bags of features, regions in a multidimensional space, algebras, or logical predicates (e.g., in description logics, which are popular among Semantic Web ontologies). Consequently, similarity measures do not only differ in their expressivity but also in the degree and kind of formality applied to represent concepts, which makes them difficult to compare. Besides the question of representation, context and its integration is another major challenge for similarity measures [40, 52]. Meaningful notions of similarity cannot be determined without defining (or at least controlling) the context in which similarity is measured [23,32,69]. While research from many domains including psychology, neurobiology, and GIScience argues for a situated nature of conceptualization and reasoning [8, 9, 12, 58, 67, 91], the concept representations used by most similarity theories from information science are static and de-contextualized. An alternative approach was recently presented by Raubal [79] arguing for a time-indexed representation of concepts.

Similarity has been widely applied within GIScience. Based on Tversky's feature model [88], Rodríguez and Egenhofer [81] developed the matching distance similarity measure (MDSM) which supports a basic context theory, automatically determined weights, and a symmetric as well as a non-symmetric mode. Ahlqvist, Raubal, and Schwering [2, 77, 83] used conceptual spaces [26] for models based on geometric distance. Sunna and Cruz [14, 86] applied network-based similarity measures for ontology alignment. Several measures [4, 5, 11, 15, 16, 39, 48] have been developed to close the gap between ontologies specified in description logics and classical similarity theories which had not been able to handle the expressivity of these logics so far. Other theories [60, 73] have been established to determine the similarity between spatial scenes, handle uncertainty in the definition of geographic categories [25], or to compute inter-user similarity for geographic recommender systems [68]. Similarity has also been applied as a quality indicator in geographic ontology engineering [45]. The ConceptVISTA [24] ontology management and visualization toolkit uses similarity for knowledge retrieval and organization. Klippel [54, 55] provided first insights into measuring similarity between geographic events and the dynamic conceptualization of topological relations.

3 Semantics of similarity

Similarity has been applied to various tasks in many domains. One consequence is that there is no precise and application-independent description of how and what a similarity theory measures [32, 69]. Even for semantics-based information retrieval, several similarity measures have been proposed. This makes the selection of an appropriate measure for a particular application a challenging task. It also raises the question of how to compare existing theories. By examining several of these measures from different domains we found generic patterns which jointly form a framework for describing how similarity is computed [44, 48]. The framework consists of the following seven steps:

1. definition of application area and intended audience;
2. selection of context and search (query) and target concepts;
3. transformation of concepts to canonical form;
4. definition of an alignment matrix for concept descriptors;
5. application of constructor specific similarity functions;
6. determination of standardized overall similarity; and
7. interpretation of the resulting similarity values.

The implementation of these steps depends on the similarity measure as well as the used representation language. Steps which may be of major importance for a particular theory, may play only a marginal role for others. The key motivation underlying the framework is to establish a systematic approach to describe how a similarity theory works by defining in which ways it implements the seven steps. By doing so, the theory fixes the semantics of the computed similarity values as well as important characteristics, such as whether the measure is symmetric, transitive, reflexive, strict, or minimal [6, 13, 31]. Moreover, the framework also supports a separation between the process of computing similarity (i.e., *what* is measured) and the applied similarity functions (i.e., *how* it is measured). Note that we distinguish between similarity functions and similarity measures (or theories). A similarity measure is an application of the proposed framework, while similarity functions are specific algorithms used in step 5. For instance, a particular similarity theory may foresee the use of different similarity functions depending on the tasks or users. This difference is discussed in more detail below. While the framework has been developed for inter-concept similarity measures, it can be reused and modified to understand inter-instance similarity as well. The reason for focusing on inter-concept similarity lies in their complex nature which makes understanding particular steps and design decisions necessary.

In the following, a description of each step is given; examples from geometric, feature-based, alignment, network, and transformational similarity measures demonstrate the generalizability of the framework.

3.1 Application area and intended audience

Which functions should be selected to measure similarity depends on the application area. Theories established for (geographical) information retrieval and in the cognitive sciences tend to use non-symmetric similarity functions to mimic human similarity reasoning [31], which is also influenced by language, age, and cultural background [40, 63, 69]. The ability to adjust similarity measures also plays a crucial role in human-computer interaction. In contrast, similarity theories for ontology matching and alignment tend to utilize symmetric functions as none of the compared ontologies plays a preferred role. In some cases, the choice of a representation language influences which parameters have to be taken into account before measuring similarity. For instance, for logical disjunctions among predicates one needs to choose between computing the maximum, minimum [16], or average similarity [44]. With respect to the introduced information retrieval definition, this step is responsible for adjusting the similarity theory using inferable implicit information.

3.2 Context, search, and target concepts

Before similarity is measured, concepts have to be selected for comparison. Depending on the application scenario and theory, the search concept C_s can be part of the ontology or built from a shared vocabulary; in the latter case the term query concept C_q may be more appropriate [39, 44, 62]. The target concepts $C_{t_1}, ..., C_{t_i}$ form the so-called context of discourse $\mathcal{C}_d$ [40] (called domain of application in case of the MDSM [81]) and are selected by hand or automatically determined by specifying a context concept C_c. In the latter case, the target concepts are those concepts subsumed by C_c. Equation 2 shows how to derive the context of discourse for similarity theories using description logics as representation language.

$$\mathcal{C}_d = \{C_t \mid C_t \sqsubseteq C_c\} \tag{2}$$

In case of the matching distance similarity measure, the context (C) is defined as a set of tuples over operations (op_i) associated with their respective nouns (e_j, equation 3). These nouns express types, while the operations correspond to verbs associated with the functions defined for these types (see [81] for details). For instance, a context such as $C = \langle(play, \{\})\rangle$ restricts the domain of application to those types which share the functional feature *play*.

$$C = \langle(op_i, \{e_1, ..., e_m\}), ..., (op_n, \{e_1, ..., e_l\})\rangle \tag{3}$$

Other knowledge representation paradigms such as conceptual spaces require their own definitions, e.g., by computing relations between regions in a multi-dimensional space.

The distinction between search and target concept is especially important for non-symmetric similarity. As will be discussed in the similarity functions step, the selection of a particular context concept does not only define which concepts are compared but also directly affects the measured similarity. The following list shows some exemplary similarity queries from the domain of hydrology, defined using search, target, and context concept:

- How similar is *Canal* (C_s) to *River* (C_t)?
- Which kind of *Waterbody* (C_c) is most similar to *Canal* (C_s)?
- What is most similar to *Waterbody* $\wedge$ *Artificial* (C_q)?
- What is more similar to *Canal* (C_s), *River* (C_t) or *Lake* (C_t)?
- What are the two most similar *Waterbodies* (C_c) in the examined ontology?

In the first case, *Canal* is compared to *River*, and in the second case to all subconcepts of *Waterbody* (e.g., *River*, *Lake*, *Reservoir*). In contrast, the third case shows a query over the whole ontology. All concepts are compared for similarity to the query concept formed by the conjunction of *Waterbody* and *Artificial*. Note that the query and context concepts are not necessarily part of the ontology, but can be defined by the user. The fourth query is an extended version of the first, with two target concepts selected by hand. Symmetric similarity measures can be defined without an explicit search and target concept, though this is difficult to argue from a cognitive point of view as direction is implicitly contained in many retrieval tasks.

3.3 Canonical normal form

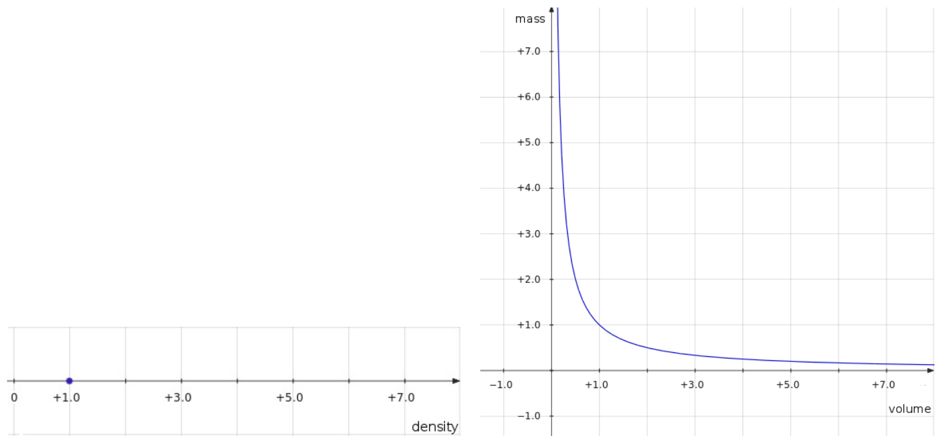
Semantic similarity measures should only be influenced by what is said about concepts, not by how it is said (syntactic differences). If two concept descriptions denote the same referents using different language elements, they need to be rewritten in a common form to eliminate unintended syntactic influences. This step mainly depends on the underlying representation language and is most important for structural similarity measures. Two simple examples for description logics are:

1. **Condition** $(\leq nR.C)$ and $n \leq 0$ **Rewrite** $(\leq nR.C)$ to $\perp$
2. **Condition** $\forall R.C \sqcap \forall R.C'$ **Rewrite** $\forall R.C \sqcap \forall R.C'$ to $\forall R.(C \sqcap C')$

One may also think of canonizations for conceptual spaces. For instance, if the dimensions *density*, *mass*, and *volume* are part of a knowledge base: the category of all entities with a density value 1ρ can be either expressed as a point on the density axis or as a curve in the space with dimensions mass and volume. Per definition, the denoted category contains the same entities, but the similarity value would be 0 using classical geometry-based similarity measures (see Figure 1). In such a case, a rewriting rule has to map one representation to the other. Of course, this example requires that the semantics of the involved dimensions is known. A first approach to handle these difficulties was presented by Raubal, introducing projection and transformation rules for conceptual spaces [78]. However, from a perspective of human cognition canonization may not always be possible.

Similar examples can be constructed for so-called transformational measures [35]. They define semantic similarity as a function over a set of transformation rules to derive a representation from another one. Among others, transformation rules include deletion, mirroring or shifting. Canonization may be required on two levels. First, it has to be ensured that the same set of transformations is used and that no transformation can be constructed out of others (as this would increase the transformation distance and, hence, decrease similarity). Second, the same representation has to be used. For instance X2OX3OX3OX may be a condensed representation of the stimulus XXOXXXOXXXOX [31] and, hence, has to be unfolded before comparison to ensure that a shift of the first O towards the second counts 3 instead of 2 steps.

In general, canonization is a complex and expensive task and should be reduced to a minimum. For instance, SIM-DL_A uses the same similarity functions as our previous SIM-DL theory [44] but reduces the need of canonization and syntactic influence by breaking down the problem of inter-concept similarity to the less complex problem of inter-instance similarity [48]. This is achieved by comparing potential interpretations for overlap instead of a structural comparison of the formal specifications. In doing so, SIM-DL_A addresses some of the challenges discussed in to introduction, namely how to deal with the multitude of potential graph representations. This is especially important for concepts specified using expressive description logics.



(a) 1ρ on the density dimension

(b) 1ρ; $\rho = m/v$ on the mass(m) and volume(v) dimensions

Figure 1: The category of all entities with the density of 1ρ specified using one dimension (a) or two dimensions (b)

3.4 Alignment matrix

While the second step of the framework selects concepts for comparison, the alignment matrix specifies which concept descriptors (e.g., dimensions, features) are compared and how. We use the term "alignment" in a slightly different sense, but based on research in psychology that investigates how structure and correspondence influence similarity judgments [22,27,64,66,69]. The term "matrix" points to the fact that the selection of comparable tuples of descriptors requires a matrix $C_s^D \times C_t^D$ (where C_s^D and C_t^D are the sets of descriptors forming C_s and C_t, respectively).

Alignment-based approaches were developed as a reaction to classical feature-based and geometric models, which do not establish relations between features and dimensions. This also affects relations to other concepts or to instances. For example, in feature-based and geometric models it is not possible to state that two concepts are similar, because their instances stand in a certain relation to instances of another concept. As depicted in Figure 2, the topological relation *above(circle, triangle)* [31] does not describe the same fact as *above(triangle, circle)*. During a similarity assessment participants may judge *above(circle, triangle)* more similar to *above(circle, rectangle)* than to *above(triangle, circle)* because of the same role (namely being above something else) that the circle plays within the first examples (see also [65]).

The motivation behind alignment-based models is that relations between concepts and their instances are of fundamental importance to determine similarity [28, 29, 66]. If instances of two compared concepts share the same color, but the colored parts are not related to each other, then the common feature of having the same color may not influence the similarity assessments. This means that subjects tend to focus more on structures and relations than on disconnected features. Hence, alignment-based models claim that similarity cannot be reduced to matching features, but one must determine how these features *align with* others [31].

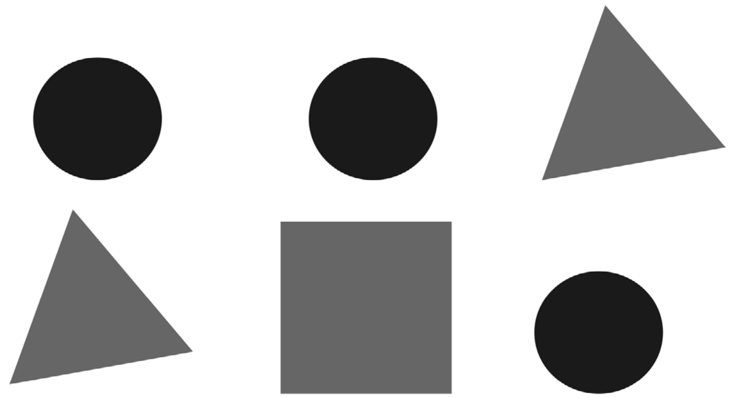


Figure 2: Being above something else as common feature used for similarity reasoning (see [31] for details)

From a set of available concept descriptors, humans tend to select those for comparison which correspond in a meaningful way [22, 27, 64, 66, 69]. The literature distinguishes between alignable commonalities, alignable differences, and non-alignable differences. In the first case, entities and relations match. For instance, in *above(circle,triangle)*, *above(circle,triangle)*, *above(circle,rectangle)*, and *smaller(circle,triangle)*, the first two assertions are alignable because both specify an above relation, and common because of the related entities. In contrast, the second and third assertion form an alignable difference. While the assertions can be compared for similarity, the related entities do not match (but could still be similar). Non-alignable differences cannot be compared for similarity in a meaningful way. For instance, no meaningful notion of similarity can be established between *above* and *smaller*. While this example relates individuals within spatial scenes, the same argumentation holds for the concept level. The fact, for instance, that rivers are connected to other water bodies can be compared to the connectedness of roads. For this reason, both can be abstracted as being parts of transportation infrastructures. (At the same time, this example also demonstrates the vague boundaries between similarity and analogy-based reasoning.) In contrast, this connectedness cannot be compared to a has-depth relation of another water body as they form a non-alignable difference.

In the proposed similarity framework the alignment matrix tackles the following questions: in most similarity theories each concept descriptor from (C_s) is compared to exactly one descriptor from (C_t)—how are these tuples selected? If the compared concepts are specified by a different number of descriptors, how are surplus descriptors to be treated [78]? Does it make a difference whether the remaining descriptors belong to the search or target concept? Are there specific weights for certain tuples or are all tuples of equal importance? How similar are concepts to their super-concepts and vice versa? Does the similarity measure depend on the search direction?

While the distinction between search and target concept was introduced in step 1, the question of how the search direction influences similarity also depends on the alignment. In theory, the following four settings can be distinguished:

A user is searching for a concept that exactly matches the search concept (C_s) ...

- and every divergence reduces similarity.
- or is more specific.
- or is more general.
- or at least overlaps with C_s.

In the first case, similarity is 1 if $C_s \equiv C_t$ and decreases with every descriptor from C_s or C_t that is not part of both specifications. Similarity reaches 0 if the compared concepts have no common descriptor. Asymmetry is not mandatory in this setting, but can be introduced by weighting distinct features differently depending on whether they are descriptors of C_s or C_t. In the second scenario, similarity is 1 if $C_s \equiv C_t$ or if C_t is a sub-type of C_s; else, similarity is 0. Such a notion of similarity is not symmetric. If C_t is a sub-concept of C_s, the similarity $sim(C_s, C_t)$ is 1, while $sim(C_t, C_s) = 0$. The third case works the other way around, similarity is 1 if $C_s \equiv C_t$ or if C_s is a sub-type of C_t. In the last scenario, similarity is always 1, except for the case when C_s and C_t do not share a single descriptor.

In contrast to the first setting, the remaining cases can be reduced to subsumption-based information retrieval, as described by Lutz and Klien [62]. These settings only distinguish values between 1 and 0. In the second and third case, the search (query) concept is injected into the examined ontology. After reclassification, all sub- or super-concepts of C_s are part of the result set [49, 62]. The last scenario can be solved accordingly by searching for a common super-concept of C_s and C_t.

Consequently, a similarity theory should be based on the first case or a combination of the first and second, or first and third case. Such combinations necessarily lead to non-symmetric similarity measures. For instance, SIM-DL is a combination of setting one and two. (To be more precise, SIM-DL allows to choose between a symmetric and non-symmetric mode.) The similarity between two concepts decreases with a decreasing overlap of descriptors, while the similarity between a type and its sub-types is always 1. The geometric similarity measure defined by Schwering and Raubal [83] applies the following rules to handle (non-)symmetry: 1. The greater the overlap and the less the non-overlapping parts, the higher the similarity between compared concepts; 2. Distance values from subconcepts to their superconcept are zero; 3. Distance values from superconcept to subconcepts are always greater than zero, but not necessarily 1.

It is important to keep in mind that these design decisions are driven by the application and not by a generic law of similarity [32, 33, 75, 85].

3.5 Similarity functions

After selecting the compared concepts and aligning their descriptors, the similarity for each selected tuple is measured. Depending on the representation language and application, different similarity functions have to be applied. In most cases, each similarity function itself takes care of standardization (to values between 0 and 1).

In case of the matching distance similarity measure (MDSM) [81], the features are distinguished into different types during the alignment process: parts, attributes, and functions. Although a contextual weighting is computed for each of these types, the same similarity function is applied to all of them.

$$S_t(c_1, c_2) = \frac{|C_1 \cap C_2|}{|C_1 \cap C_2| + \alpha(c_1, c_2) * |C_1 \backslash C_2| + (1 - \alpha(c_1, c_2)) * |C_2 \backslash C_1|} \tag{4}$$

Equation 4 describes the non-symmetric similarity function for each of the feature types. $S_t(c_1, c_2)$ is defined as the similarity for the feature type t between the entity classes c_1 and c_2. C_1 and C_2 are the sets of features of type t for c_1 and c_2, while $|C_1 \cap C_2|$ is the cardinality of the set intersection and $|C_1 \backslash C_2|$ is the cardinality of the set difference. The relative importance α (equation 5) of the different features of type t is defined in terms of the distance d between c_1 and c_2 within a hierarchy that takes taxonomic and partonomic relations into account. Lub denotes the least upper bound, i.e., the immediate common superclass of c_1 and c_2 [81]. The distance is defined as $d(c_1, c_2) = d(c_1, lub) + d(c_2, lub)$.

$$\alpha(c_1, c_2) = \begin{cases} \frac{d(c_1,lub)}{d(c_1,c_2)}, d(c_1, lub) \leq d(c_2, lub) \\ 1 - \frac{d(c_1,lub)}{d(c_1,c_2)}, d(c_1, lub) > d(c_2, lub) \end{cases} \tag{5}$$

MDSM accounts for context by introducing weights for the different types of features. While the integration of these weights (ω_t in equation 13) plays a role for the overall similarity, the two weighting functions are introduced here. The relevance of each feature type is defined either by the variability P_t^v (equation 6) or commonality P_t^c function (equation 7) and then normalized with respect to the remaining feature types so that the sum of $\omega_p + \omega_f + \omega_a$ is always 1.

$$P_t^v = 1 - \sum_{i=1}^{l} \frac{o_i}{n * l} \tag{6}$$

The variability describes how *diagnostic* [30, 88] or characteristic a feature *t* is within a certain application. A certain feature of type *t* has low relevance if it appears in many classes and high relevance if it is not common to the classes within the domain. P_t^v is the sum of the diagnosticity of all features of the type *t* in the domain and therefore 0 when all features are shared by all entity classes (P_t^v=1-1=0), and close to 1 if each feature is unique (o_i is the number of occurrences of the feature within the domain) and the number of features *l* and classes *n* in the domain is high.

$$P_t^c = \sum_{i=1}^{l} \frac{o_i}{n * l} = 1 - P_t^v \tag{7}$$

Commonality is defined as the opposite of variability ($P_t^c = 1 - P_t^v$) and assumes that by defining a domain of application the user implicitly states what features are relevant [81].

In contrast to MDSM, SIM-DL and SIM-DL$_\text{A}$ distinguish between several similarity functions for roles and their fillers, e.g., functions for conceptual neighborhoods, role hierarchies, or co-occurrence of primitives. Primitives (also called base symbols) occur only on the right-hand side of definitions. To measure their similarity (sim_p, see equation 8), an adapted version of the Jaccard similarity coefficient is used. It measures the degree of overlap between two sets S_1 and S_2 as the ratio of the cardinality of shared members (e.g., features) from $S_1 \wedge S_2$ to the cardinality retrieved from $S_1 \vee S_2$. In SIM-DL, the coefficient is applied to compute the context-aware co-occurrence of primitives within the definitions of other (non-primitive) concepts [44]. Two primitives are the more similar, the more complex concepts are defined by both (and not only one) of them. If $sim_p(A, B) = 1$, both primitives always co-occur in complex concepts and cannot be distinguished. As similarity depends on the context of discourse [40], only those concepts C_i are considered which are subconcepts of C_c (see step two of the similarity framework).

$$sim_p(A,B) = \frac{|\{C \mid (C \sqsubseteq C_c) \wedge (C \sqsubset A) \wedge (C \sqsubset B)\}|}{|\{C \mid (C \sqsubseteq C_c) \wedge ((C \sqsubset A) \vee (C \sqsubset B))\}|} \tag{8}$$

SIM-DL uses a modified network-based approach [76] to compute the similarity between roles (R and S) within a hierarchy. Similarity (sim_r, see equation 9) is defined as the ratio between the shortest path from R to S and the maximum path within the graph representation of the role hierarchy; where the universal role U ($U \equiv \triangle^{\mathcal{I}} \times \triangle^{\mathcal{I}}$) forms the graph's root. Compared to sim_p, similarity between roles is defined without reference to the context. This would require to take only such roles into account which are used within quantifications or restrictions of concepts within the context. The standardization in equation 9 is depth-dependent to indicate that the distance from node to node decreases with increasing depth level of R and S within the hierarchy. In other words, the weights of the edges used to determine the path between R and S decrease with increasing depth of the graph. If a path between two roles crosses U, similarity is 0. The $lcs(R,S)$ is the least common subsumer, in this case the first common super role of R and S.

$$sim_r(R,S) = \frac{depth(lcs(R,S))}{depth(lcs(R,S)) + edge\_distance(R,S)} \tag{9}$$

Similarity between topological or temporal relations (sim_n, see equation 10) equals their normalized distance within the graph representation of their conceptual neighborhood. In contrast to sim_r, the normalization is not depth-dependent but based on the longest path within the neighborhood graph.

$$sim_n(R,S) = \frac{max\_distance_n - edge\_distance(R,S)}{max\_distance_n} \tag{10}$$

The similarity between role filler pairs (sim_{rf}, see equation 11) is defined by the similarity of the involved roles R and S times the overall similarity of the fillers C and D which can again be complex concepts.

$$sim_{rf}(R(C), S(D)) = sim_r(R,S) * sim_o(C,D) \tag{11}$$

Some similarity measures define role-filler similarity as the weighted average of the role and filler similarities, but the multiplicative approach has proven to be cognitively plausible [43] and allows for simple approximation and optimization techniques not discussed here in detail.

In the case of geometric approaches to similarity, the spatial distance in the conceptual (vector) space is interpreted as the semantic distance d. Consequently, similarity increases with decreasing spatial distance. A classical function for geometry-based similarity measures is given by the Minkowski metric (see equation 12). The parameter r is used to switch between different distances, such as the Manhattan distance ($r = 1$) and the Euclidean distance ($r = 2$) [31]. A more detailed discussion with regard to a metric conceptual space algebra including weights is given by Adams and Raubal [1].

$$d(c,d) = \left[\sum_{i=1}^{n} |\, c_i - d_i\,|^r\right]^{\frac{1}{r}} \tag{12}$$

Note that, while we focus on inter-concept similarity here, certain similarity functions can also take knowledge about instances into account to derive information about concept similarity [15–17].

3.6 Overall similarity

In the sixth step of the framework, the single similarity values derived from applying the similarity functions to all selected tuples of compared concepts are combined to an overall similarity value. In most theories this step is a standardized (to values between 0 and 1) weighted sum.

For MDSM, the overall similarity is the weighted sum of the similarities determined between functions, parts, and attributes of the compared entity classes c_1 and c_2. The weights indicate the relative importance of each feature type using either the commonality or variability model introduced before (equation 13). At the same time, the weights act as standardization factors ($\sum \omega = 1$) [81].

$$S(c_1, c_2) = \omega_p * S_p(c_1, c_2) + \omega_f * S_f(c_1, c_2) + \omega_a * S_a(c_1, c_2) \tag{13}$$

In case of SIM-DL, each similarity function takes care of its standardization using the number of compared tuples or the graph depth. Each similarity function returns a standardized value to the higher-level function by which it was called. Hence, overall similarity is simply the (standardized) sum of the single similarity values.

For geometric approaches, the overall similarity is given by the z-transformed sum of compared values [77], in order to account for different dimensional units. Each z_i score is computed according to equation 14 where x_i is the i-th value of the quality dimension X, $\overline{x}$ is the mean of all X_i of X, and s_x is the standard deviation of these x_i.

$$z_i = \frac{x_i - \overline{x}}{s_x} \tag{14}$$

The overall similarity is then defined using the Minkowski metric (see equation 12) where n is the number of quality dimensions and c and d are the z-transformed values for the compared concepts (per dimension).

3.7 Interpretation of similarity values

All of the introduced measures map two compared concepts to a real number. They do not explain their results or point to descriptors for which the concepts differ. Such a single value (e.g., 0.7) is difficult to interpret. For instance, 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 remains unclear. If the least similar target concept in an ontology has a similarity value of 0.65 to the source concept and the most similar concept yields 0.9, a similarity value of 0.7 is not necessarily a good match. It is difficult to argue why a single similarity value is cognitively plausible without reference to other results [51]. Moreover, the threshold value above which compared concepts are considered *similar* depends on the specific application and context.

Therefore, measures such as MDSM or SIM-DL rely on similarity rankings. They compare a search concept to all target concepts from the domain of discourse and return the results as an ordered list of descending similarity values. Consequently, one would not argue that a particular similarity value is cognitively plausible, but that a ranking correlates with human estimations [43]. Such a ranking puts a single similarity value in context by delivering additional information about the distribution of similarity values and their range. We call this context the interpretation context ($\mathcal{C}_i$, see [40] for more details on different kinds of contexts and their impact on similarity measures).

$$\mathcal{C}_i : (C_s, C_t, simV) \in \Delta_{sim} \times \mathcal{C}_a \rightarrow \Psi(C_s, C_t) \in \Delta_\Psi \tag{15}$$

The interpretation context (see equation 15) maps the triple search concept (C_s), target concept (C_t), similarity value ($simV$) from the set of measured similarities between the search concept and each target concepts $\in \mathcal{C}_d$ (Δ_{sim}) and the restrictions specified by the application context ($\mathcal{C}_a$) to an interpretation value ($\Psi(C_s, C_t)$) from the domain of interpretations (Δ_Ψ). The application context [40] describes the settings by which a similarity measure can be adapted to the user's needs, e.g., whether the commonality or variability weightings in MDSM should be selected.

The simplest domain of interpretation can be formed by $\Delta_\Psi = \{t, f\}$. Depending on the remaining pairs of compared concepts from Δ_{sim} as well as the application area, each triple is either mapped to true or false. Therefore, the question of whether concepts are similar is answered by yes or no. For graphical user interfaces, similarity values can also be mapped to font sizes using a logarithmic tag cloud algorithm (see Figure 3). Note that as $\mathcal{C}_i$ depends on Δ_{sim}, it does not simply map an isolated similarity value to yet another domain. For example, the maximum font size will always be assigned to the target concept with the highest similarity to the search concept, independent of the specific value.

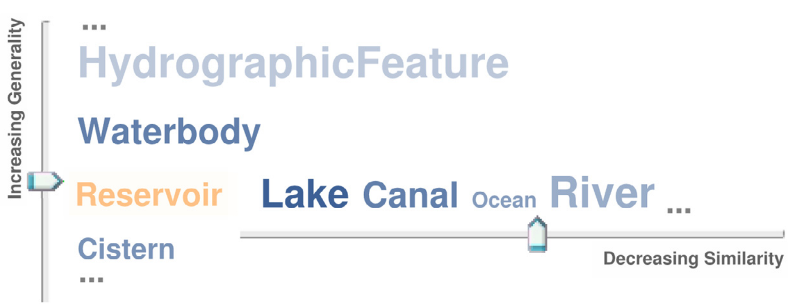


Figure 3: Font size scaling for similarity values, based on [47]

3.8 Properties of similarity measures

The proposed framework helps to understand how similarity theories work and what they measure. This is essential for choosing the optimal measure for a specific application, to compare similarity measures, and to interpret similarity values and rankings. The framework also unveils basic properties of a particular measure, e.g., whether it is reflexive, symmetric, transitive, strict, minimal, etc. (see [6, 13, 31, 75] for a detailed discussion from the perspectives of computer science and psychology). As an example, the following paragraphs discuss strictness and symmetry for the SIM-DL/SIM-DL$_\mathrm{A}$ theory, as well as the relation between similarity and dissimilarity. The triangle inequality is discussed as an important property of geometric approaches.

Strictness is often referred to as an important property of similarity [87]. Formally, strictness states that the maximum similarity value is only assigned to equal stimuli (e.g., concepts): $sim(C, D) = 1$ if and only if $C \equiv D$. This is related to the minimality property, which claims that two different stimuli are less (or equally) similar than the stimulus is to itself: $sim(C, D) \leq sim(C, C)$ [6, 31]. In the literature, minimality is defined for dissimilarity: $dis(C, D) \geq dis(C, C)$. In SIM-DL, the similarity value 1 is interpreted as *equal or not distinguishable (within a given context)*. This is for two reasons: co-occurrence between primitives and non-symmetry. The comparison of two primitives yields 1 if they cannot be differentiated, i.e., if they always appear jointly within concept definitions (see equation 8). As SIM-DL focuses on information retrieval, a target concept satisfies the user's needs ($sim(C_s, C_t) = 1$) if it is a sub-concept of the search concept (step 4 of the framework). Consequently, similarity in SIM-DL is not strict.

Symmetry is one of the most controversial properties of similarity. While several theories from computer science argue that similarity is essentially a symmetric relation [61], research from cognitive science favors non-symmetric similarity measures [56, 69, 74, 88]. As argued in the previous sections, SIM-DL allows the user to switch between a symmetric and a non-symmetric mode. From Tversky's [88] point of view, one may argue that this is nothing more than indecision. However, the understanding of symmetry underlying SIM-DL is driven by Nosofsky's notion of a biased measure [74]. Symmetry is not a characteristic of similarity as such, but of the process of measuring similarity. This process is driven (biased) by a certain task—namely information retrieval. Whether the comparison of two concepts is symmetric or not depends on the application area and task (and therefore on the alignment process), but not on the measure as such. This again reflects the need for a separation between the alignment and the application of concrete similarity functions.

Dissimilarity and similarity are often used interchangeably assuming that dissimilarity is simply the counterpart of similarity: $dis(C, D) = 1 - sim(C, D)$. While this may be true for certain cases, it is not a valid assumption in general [31]. As argued by Tversky [88], Nosofsky [74], and Dubois and Prade [19], similarity and dissimilarity are different views on stimuli comparison. SIM-DL, for instance, stresses the alignment of descriptors. If the task is to find dissimilarities between compared concepts, other tuples might be selected for comparison and alignment. One can demonstrate that the assumption $dis(C, D) = 1 - sim(C, D)$ is oversimplified and counter-intuitive using SIM-DL's maximum similarity function for concepts formed by logical disjunction. For simplification, consider the concepts $C \equiv A \sqcap B$ and $D \equiv C \sqcup E$ where A, B, and E are primitives. To measure the similarity $sim(C, D)$, SIM-DL unfolds their definitions and creates the following alignment tuples: (A, A), (A, B), (A, E), (B, A), (B, B), and (B, E). Out of this set, the tuples (A, A) and (B, B) are chosen for further computation and finally, $sim(C, D)$ returns 1. Consequently, the resulting dissimilarity $dis(C, D)$ should be 0. This is true, if one still applies the maximum similarity function. Instead, when searching for dissimilarities between compared concepts, one would rather use a minimum similarity function and thus take E into account for comparison to A or B. In both cases, $dis(C, D)$ can be greater than 0.

Triangle Inequality describes the metric property according to which the distance between two points cannot be greater than the distance between these points reached via an additional third point. Surprisingly, it turns out that even such fundamental properties of geometry cannot be taken for granted. Instead, Tversky and Gati demonstrated that the triangle inequality does not necessarily hold for cognitive measures of similarity [89].

4 Similarity in semantics-based information retrieval

While the proposed framework defines how similarity is measured, this section demonstrates its role in semantics-based geographic information retrieval and its integration into user interfaces.

4.1 Retrieval paradigms

Previously, we defined information retrieval by the degree of relevance $m[\Re(O,(Q,\langle I,\mapsto\rangle))]$ without stating how to measure this relevance. Based on this definition and without going into any details about query rewriting and expansion, we explain the role of similarity by restricting the definition such that:

- O is a set of target concepts (C_t) in an ontology,
- Q is a particular concept phrased or selected for the search (C_s),
- I and $\mapsto$ are additional contextual information at execution time (C_c),
- $\Re$ is the similarity relationship between pairs of concepts, and
- and m is the degree of similarity between pairs of concepts.

In contrast to purely syntactic approaches, semantics-based information retrieval takes the underlying conceptualizations into account to compute relevance and hence improves searching and browsing through structured data. In general, one can distinguish between two approaches for concept retrieval: those based on classical subsumption reasoning and those that rely on semantic similarity measures [49]. Simplifying, subsumption reasoning can be applied to vertical search, while similarity works best for horizontal search, i.e., similarity values are difficult to interpret when comparing sub- and super-types.

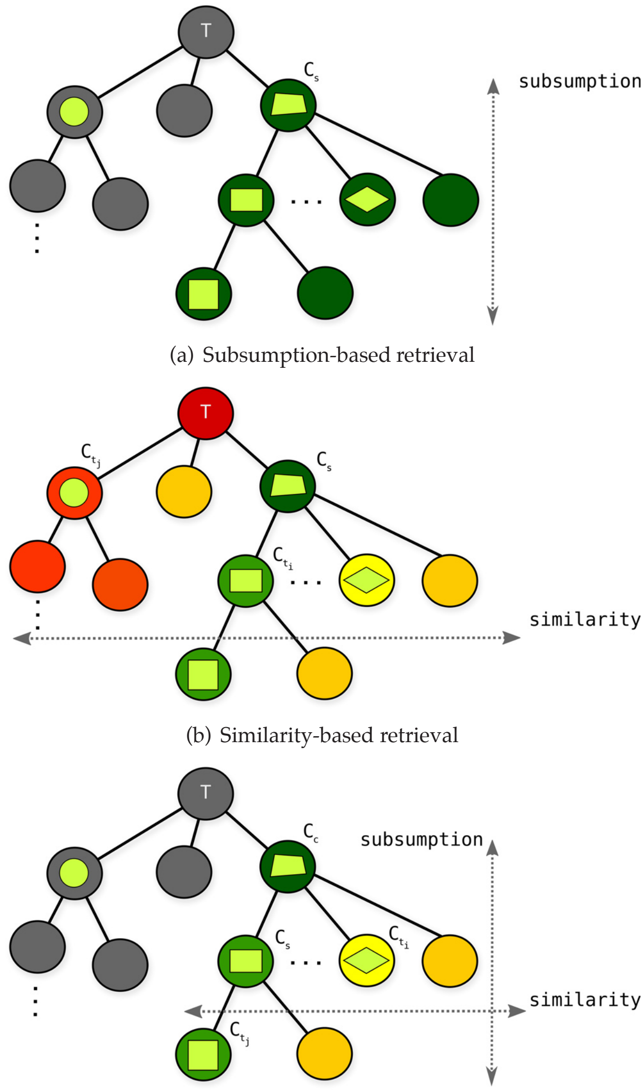
Formally, the result set for a subsumption-based query is defined as $\mathbb{RS} = \{C \mid C \in O \wedge C \sqsubseteq Q\}$. As each concept in $\mathbb{RS}$ is a subsumee of the search/query concept, it meets the user's search criteria (see Figure 4a). Consequently, there is no degree of of relevance m; or, to put it in other words, it is always 1. The missing relevance information and rigidity of subsumption make selecting an appropriate search concept the major challenge for subsumption-based retrieval. In many cases, the search concept will be an artificial construct and not necessarily the *searched* concept (see [49] for details). If it is too generic (i.e., too close to the top of the hierarchy) the user will get a large part of the queried ontology back as an unsorted result set; if the search concept is too narrow, the result set will only contain a few or even no concepts.

For similarity-based retrieval as depicted in Figure 4b, the result set is defined as $\mathbb{RS} = \{C \mid C \in O \wedge sim(Q,C) > t\}$; where t is a threshold defined by the user or application [44, 49]. In contrast to subsumption-based retrieval, the search concept is the concept the user is really *searching* for, no matter whether it is part of the queried ontology or not. As similarity computes the overlap between concept definitions (or their extensions [16, 48]) it is more flexible than a purely subsumption-based approach. Moreover, the results are ranked—returned as an ordered list with descending similarity values representing the relevance m. This makes it easier for the user to select an appropriate concept from the results. However, it is not guaranteed that the returned concepts match all of the user's search criteria. Consequently, the benefits similarity offers during the retrieval phase, namely to deliver a flexible degree of (conceptual) overlap with a searched concept, stands against shortcomings during the selection phase, because the results do not necessarily match all of the user's requirements.

To overcome these shortcomings, similarity theories such as SIM-DL and MDSM combine subsumption and similarity reasoning by introducing contexts to reduce the set of potential target concepts (see equations 2 and 3). As depicted in Figure 4c, only those concepts are compared for similarity that are subconcepts of the context concept C_c. This way, the user can specify some minimal characteristics all target concepts need to share. Typically, user interfaces and search engines will be designed in a way to infer or at least approximate C_c from additional, implicit contextual information $(I, \mapsto)$. Consequently, for the combined retrieval paradigm the result set is defined as $\mathbb{RS} = \{C \mid C \in O \wedge C \sqsubseteq C_c \wedge sim(Q, C) > t\}$.

Figure 4 shows an ontology of geometric figures as a simplified example to illustrate the differences between the introduced paradigms. Note that some quadrilaterals and relations between them have been left out to increase readability. We assume that a user is searching for quadrilaterals with specific characteristics. In the subsumption only case, the result set contains types such as *Rectangle*, *Rhombus*, *Square*, and so forth without additional information about their degree of relevance. In the similarity only case, the result set contains additional relevance information for these types but also geometric figures such as *Circle* which do not satisfy all the requirements specified by the user. Note however that they would appear at the end of the relevance list due to their low similarity (indicated by the shift from green over yellow to red in Figure 4b). In case of the combined paradigm a user could prefer quadrilaterals with right angles by specifying *Rectangle* as search concept and *Quadrilateral* as context concept. In contrast to the similarity only case, the result set does not contain *Circle* but still delivers information about the degree of relevance.

Before going into details about the integration of the combined approach into user interfaces, we briefly need to discuss two questions which have remained unanswered so far. First, one could argue that combining subsumption and similarity reasoning by introducing the context concept as a least upper bound only shifts the query formulation problem from the search concept to the context concept. If the user chooses a context concept that is too narrow, then this has the same effects as in the subsumption only case. While this is true in general, we will demonstrate in the next section that the context concept can be derived as inferred information from the query, which is not the case for the search concept. Moreover, the combined approach still delivers ranked results instead of an unstructured set. Second, so far we have restricted our concept retrieval cases to queries based on the notion of a search or query concept and therefore to intensional retrieval paradigms. Nevertheless, there are also extensional paradigms for retrieval, e.g., based on non-standard inference techniques such as computing the least common subsumer (*lcs*) or most specific concept (*msc*) [49, 57, 70]. We will discuss these approaches using a *query-by-example* interface in which reference individuals are selected for searching.



(a) Subsumption-based retrieval

(b) Similarity-based retrieval

(c) Subsumption and similarity-based retrieval

Figure 4: Semantics-based retrieval in a simplified ontology of geometric figures

4.2 Application

This section introduces two web-based user interfaces implementing similarity and subsumption-based retrieval. The interfaces have been implemented, evaluated [43, 47], and are available as free and open source software[1]. Their integration into spatial data infrastructures was recently discussed by Janowicz et al. [46] and is left aside here.

[1]http://sim-dl.sourceforge.net/applications/

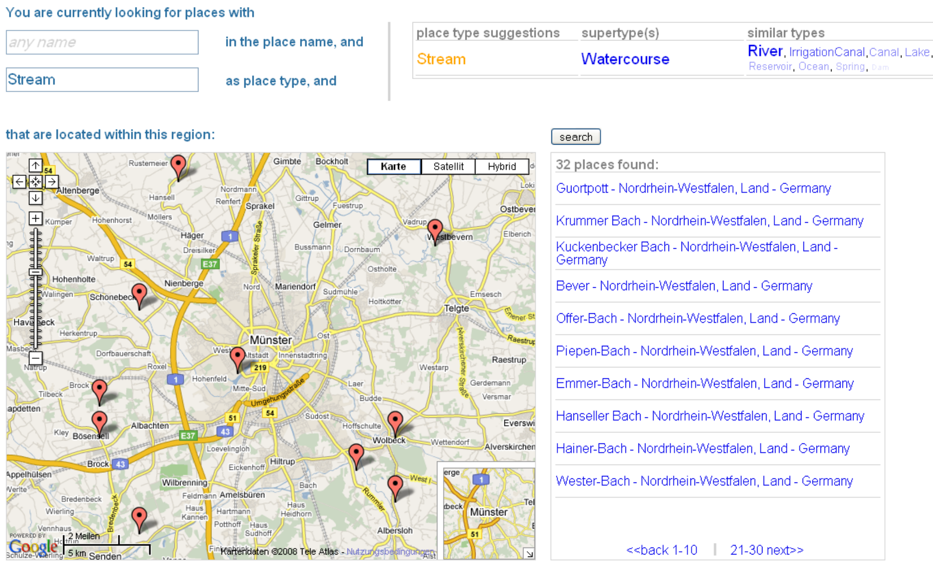


Figure 5: A subsumption and similarity-based user interface for Web gazetteers [47]

4.2.1 Selecting a search concept

Figure 5 shows a semantics-based user interface for the Alexandria Digital Library Gazetteer. The interface implements the intensional retrieval paradigm based on a combination of similarity and subsumption reasoning. A user can enter a search concept using a *search-while-you-type* AJAX-based text field. To improve the navigation between geographic feature types, the interface displays the immediate super-type as well as a list of similar types [42, 47]. Based on the question of interpretation discussed in Section 3.7, a decreasing font size indicates decreasing similarity between the search concept and the proposed target concepts. In the example query, the type *Stream* is selected for comparison and the interface displays *Watercourse* as super type to broaden the search. *River* is the most similar concept followed by other hydrographic feature types. By clicking on a super- or similar type it gets selected as search concept for a new query. The map is used to restrict the search to a specific area. The interface displays features on the right side and on the map. It does not support the selection of a context concept by the user. This would overload the interface and the underlying idea of a context concept may be difficult to explain to ordinary users. Nevertheless, the context concept can be inferred from implicit information, e.g., using the map component. The context concept can be derived by computing the least common subsumer of all feature types which have features in the map extent. Yet, this approach only works well for particular zoom levels and will become meaningless if the user searches a larger area.

4.2.2 Query-by-example

Figure 6 shows a user interface implementing an extensional (example based) paradigm using similarity and non-standard inference. It overcomes two shortcomings of the pre-

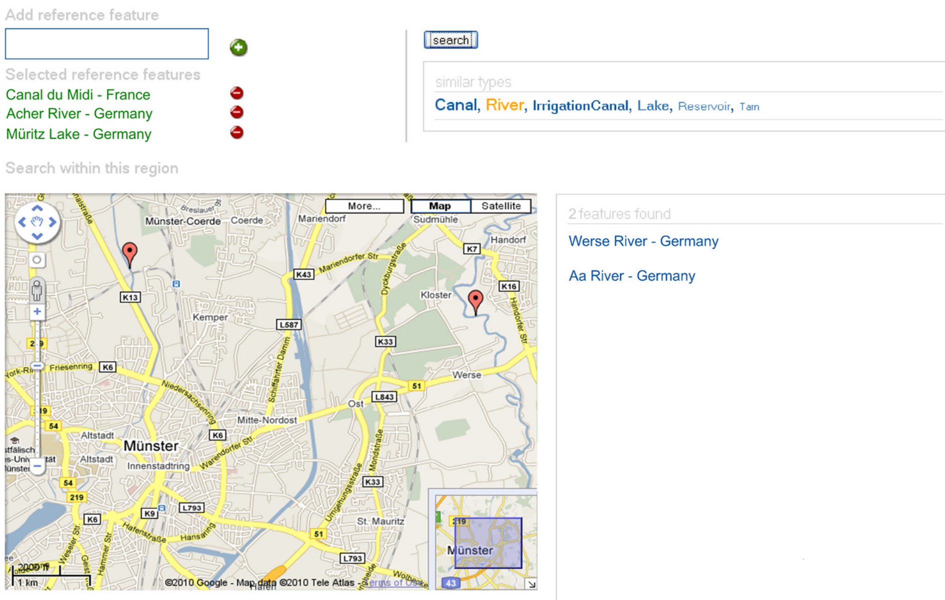


Figure 6: A conceptual design of a query-by-example based Web interface for recommender services (see [90] for an implementation of such an interface for climbing routes using the SIM-DL server)

vious interface. First, some users may be unfamiliar with using feature types for search and navigation; second, the previous interface does not offer a convincing way to infer the context concept with a minimum of user interaction. The query-by-example interface allows the user to select particular reference features instead of types. The most specific concept [57] is computed for each of these types. Based on these concepts, the least common subsumer [57] can be determined and used as context concept to deliver an inter-concept similarity ranking [90]. In the example query, three different water bodies are selected as reference features and *Canal* is computed to be the most similar concept to the least common subsumer of those concepts instantiated by the selected features. While the first interface is typical for web gazetteers, the second interface focuses on decision support and recommender services. For instance, if the user is searching for interesting canoing spots for her next vacation, the selected water bodies may be picked from previous canoing trips at different locations [49].

5 Conclusions and further work

In this article we introduced a generic framework for semantic similarity measurement. The framework consists of seven sequential steps used to explain what and how a particular theory measures. The framework clearly separates the process of measuring similarity and finding alignable descriptors from the concrete functions used to compute similarity values for selected tuples of these descriptors. It also discusses the role of context, additional application-specific parameters, and the interpretation of similarity values. We do not try to squeeze all existing similarity measures into our framework, but argue that by applying this framework—in describing the realization of the proposed steps—a measure defines the semantics of similarity. This, however, is a prerequisite for comparing existing measures and selecting them for specific applications. A similar argumentation was proposed before by Hayes for the notion of *context* [36]. Besides offering new insights into similarity theories used in GIScience and beyond, the article also discusses the role of these measures in semantics-based geographic information retrieval, introduces paradigms, and shows their implementations and limitations for real user interfaces.

Further work should focus on the following issues. First, while progress has been made on developing similarity theories for more expressive description logics [4, 17, 48], the approximation and explanation of similarity values is still at an early stage. Both topics are crucial for the adaptation of similarity-based information retrieval paradigms into more complex applications. Approximation techniques aim at reducing the computational costs for similarity measurements. While the theories reviewed here can compare dozens of concepts within a reasonable time frame, they do not scale well. In general, two directions for future work seem reasonable. On the one hand, one could try to improve the selection and alignment process to reduce the number of comparable concepts and tuples in the first place. On the other hand, one could approximate the similarity values and only compute exact values for candidates that are above a certain threshold. In SIM-DL, for instance, the role-filler similarity is defined by multiplying role and filler similarities. The computation of role similarities is realized by a simple network-based distance. Hence, if the resulting value is below the defined threshold the more complex filler similarity does not need to be computed.

The downside of using more expressive description logics and approximation techniques is that similarity values become even harder to interpret. In the long term, it will be necessary to assist the user by providing explanations in addition to plain numerical values or rankings. Future reasoners could list which descriptors were taken into account and visualize their impact on overall similarity. While this is important for information retrieval, it would be even more relevant for ontology engineering and negotiation [45]. This way, similarity reasoning could be used to establish bridges between communities across cultures and ages. So far, there has been no work on explaining similarity values but an adaptation of recent work on axiom pinpointing [7] may be a promising starting point.

Next, evaluation methods to compare computational similarity measures to human similarity rankings are still restricted. An interesting research direction towards semantic precision and recall was recently proposed by Euzenat [21], while Keßler [52] investigates whether and how one can go beyond simple correlation measures to evaluate the cognitive plausibility of similarity theories. Another approach to adjust similarity values to the user's needs would be to compute weights out of partial knowledge gained from user feedback [41].

Additionally, similarity depends on context in many ways. Most existing measures, however, reduce context to the selection or similarity functions steps of the framework. Advanced theories should take contextual information into account to alter these functions, the alignment of descriptors, and the computational representations of the compared entities and concepts [40, 53]. One promising direction for future research is to investigate whether and to what degree context can be modeled by changing the alignment process—this would also lead to interesting insights about the graded structure of ad-hoc categories [8, 30].

Moreover, the application of similarity measures is not restricted to information retrieval. Using them for complex data mining, clustering, handling of uncertainty in ontology engineering, and so forth requires more work on visualization methods as well as integration with spatial analysis tools. Semantic variograms [3], parallel coordinate plots, or radar charts may be interesting starting points in this respect.

Finally, while we provided a framework for understanding the semantics of similarity and for articulating the differences between existing measures, a formal apparatus to quantify these differences and translate between similarity values obtained by existing theories is missing. While work on category theory may be a promising direction for further research, the key problem that remains concerns the heterogeneity of the used approaches, application areas, and the difference between idealized measures and human cognition (the triangle inequality discussed in Section 3.8 is just one example). For the same reason, we cannot argue that our framework is necessary and sufficient for all potential similarity measures.

Acknowledgments

We are thankful to our colleagues from the Münster Semantic Interoperability Lab (MUSIL), Benjamin Adams, and the three anonymous reviewers for their input to improve the quality and clarity of this article.

References

[1] ADAMS, B., AND RAUBAL, M. A metric conceptual space algebra. In *Conference on Spatial Information Theory (COSIT)* (2009), K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, Eds., vol. 5756 of *Lecture Notes in Computer Science*, Springer, pp. 51–68. doi:10.1007/978-3-642-03832-7_4.

[2] AHLQVIST, O. Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science 19*, 7 (2005), 831–857. doi:10.1080/13658810500106729.

[3] AHLQVIST, O., AND SHORTRIDGE, A. Characterizing land cover structure with semantic variograms. In *Progress in Spatial Data Handling, 12th International Symposium on Spatial Data Handling* (2006), A. Riedl, W. Kainz, and G. Elmes, Eds., Springer, pp. 401–415. doi:10.1007/3-540-35589-8_26.

[4] ARAÚJO, R., AND PINTO, H. S. Semilarity: Towards a model-driven approach to similarity. In *International Workshop on Description Logics (DL)* (2007), vol. 20, Bolzano University Press, pp. 155–162. doi:10.1.1.142.7321.

[5] ARAÚJO, R., AND PINTO, H. S. Towards semantics-based ontology similarity. In *Proc. Workshop on Ontology Matching (OM), International Semantic Web Conference (ISWC)* (2007), P. Shvaiko, J. Euzenat, F. Giunchiglia, and B. He, Eds. doi:10.1.1.143.1541.

[6] ASHBY, F. G., AND PERRIN, N. A. Toward a unified theory of similarity and recognition. *Psychological Review 95* (1988), 124–150. doi:10.1037/0033-295X.95.1.124.

[7] BAADER, F., AND PENALOZA, R. Axiom pinpointing in general tableaux. In *Proc. 16th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods TABLEAUX* (2007), N. Olivetti, Ed., vol. 4548 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 11–27. doi:10.1007/978-3-540-73099-6_4.

[8] BARSALOU, L. Ad hoc categories. *Memory and Cognition 11* (1983), 211–227.

[9] BARSALOU, L. Situated simulation in the human conceptual system. *Language and Cognitive Processes 5*, 6 (2003), 513–562. doi:10.1080/01690960344000026.

[10] BERRY, M., AND BROWNE, M. *Understanding Search Engines: Mathematical Modeling and Text Retrieval*, 2nd ed. SIAM, 2005.

[11] BORGIDA, A., WALSH, T., AND HIRSH, H. Towards measuring similarity in description logics. In *International Workshop on Description Logics (DL2005)*, vol. 147 of *CEUR Workshop Proceedings*. CEUR, 2005.

[12] BRODARIC, B., AND GAHEGAN, M. Experiments to Examine the Situated Nature of Geoscientific Concepts. *Spatial Cognition and Computation 7*, 1 (2007), 61–95. doi:10.1080/13875860701337934.

[13] CROSS, V., AND SUDKAMP, T. *Similarity and Computability in Fuzzy Set Theory: Assessments and Applications*, vol. 93 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, 2002.

[14] CRUZ, I., AND SUNNA, W. Structural alignment methods with applications to geospatial ontologies. *Transactions in GIS 12*, 6 (2008), 683–711. doi:10.1111/j.1467-9671.2008.01126.x.

[15] D'AMATO, C., FANIZZI, N., AND ESPOSITO, F. A semantic similarity measure for expressive description logics. In *Convegno Italiano di Logica Computazionale (CILC)* (2005).

[16] D'AMATO, C., FANIZZI, N., AND ESPOSITO, F. A dissimilarity measure for $\mathcal{ALC}$ concept descriptions. In *Proc. ACM Symposium on Applied Computing (SAC)* (2006), ACM, pp. 1695–1699. doi:10.1145/1141277.1141677.

[17] D'AMATO, C., FANIZZI, N., AND ESPOSITO, F. Query answering and ontology population: An inductive approach. In *Proc. 5th European Semantic Web Conference (ESWC)* (2008), S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds., vol. 5021 of *Lecture Notes in Computer Science*, Springer, pp. 288–302. doi:10.1007/978-3-540-68234-9_23.

[18] DOMINICH, S. *The Modern Algebra of Information Retrieval*, 1st ed. Springer, 2008. doi:10.1007/978-3-540-77659-8.

[19] DUBOIS, D., AND PRADE, H. A unifying view of comparison indices in a fuzzy set-theoretic framework. In *Recent Development in Fuzzy Set and Possibility Theory*, R. Yager, Ed. Pergamon Press, 1982, pp. 3–13.

[20] EGENHOFER, M. Toward the semantic geospatial web. In *Proc. 10th ACM International Symposium on Advances in Geographic Information Systems* (2002), ACM, pp. 1–4. doi:10.1145/585147.585148.

[21] EUZENAT, J. Semantic precision and recall for ontology alignment evaluation. In *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI)* (2007), pp. 348–353.

[22] FALKENHAINER, B., FORBUS, K., AND GENTNER, D. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence 41* (1989), 1–63. doi:10.1016/0004-3702(89)90077-5.

[23] FRANK, A. U. Similarity measures for semantics: What is observed? In *COSIT'07 Workshop on Semantic Similarity Measurement and Geospatial Applications* (2007).

[24] GAHEGAN, M., AGRAWAL, R., JAISWAL, A. R., LUO, J., AND SOON, K.-H. A platform for visualizing and experimenting with measures of semantic similarity in ontologies and concept maps. *Transactions in GIS 12*, 6 (2008), 713–732. doi:10.1111/j.1467-9671.2008.01124.x.

[25] GAHEGAN, M., AND BRODARIC, B. Examining uncertainty in the definition and meaning of geographical categories. In *Proc. 5th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (2002), G. J. Hunter and K. Lowell, Eds. doi:10.1.1.61.9168.

[26] GÄRDENFORS, P. *Conceptual Spaces—The Geometry of Thought*. Bradford Books, MIT Press, 2000.

[27] GENTNER, D., AND FORBUS, K. D. MAC/FAC: A model of similarity-based retrieval. In *Proc. 13th Annual Conference of the Cognitive Science Society* (1991), Erlbaum, pp. 504–509. doi:10.1207/s15516709cog1902_1.

[28] GOLDSTONE, R. L. Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition 20* (1994), 3–28. doi:10.1037/0278-7393.20.1.3.

[29] GOLDSTONE, R. L., AND MEDIN, D. Similarity, interactive activation, and mapping: An overview. In *Analogical Connections: Advances in Connectionist and Neural Computation Theory*, K. Holyoak and J. Barnden, Eds., vol. 2. Ablex, 1994, pp. 321–362.

[30] GOLDSTONE, R. L., MEDIN, D. L., AND HALBERSTADT, J. Similarity in context. *Memory and Cognition 25* (1997), 237–255.

[31] GOLDSTONE, R. L., AND SON, J. Similarity. In *Cambridge Handbook of Thinking and Reasoning*, K. Holyoak and R. Morrison, Eds. Cambridge University Press, 2005, pp. 13–36. doi:10.2277/0521531012.

[32] GOODMAN, N. Seven strictures on similarity. In *Problems and projects*. Bobbs-Merrill, 1972, pp. 437–447.

[33] GREGSON, R. *Psychometrics of similarity*. Academic Press, 1975.

[34] GRUBER, T. A translation approach to portable ontology specifications. *Knowledge Acquisition 5*, 2 (1993), 199–220. doi:10.1006/knac.1993.1008.

[35] HAHN, U., CHATER, N., AND RICHARDSON, L. B. Similarity as transformation. *Cognition 87* (2003), 1–32. doi:10.1016/S0010-0277(02)00184-1.

[36] HAYES, P. Contexts in context. In *Context in Knowledge Representation and Natural Language, AAAI Fall Symposium* (1997), AAAI Press.

[37] HITZLER, P., KRÖTZSCH, M., AND RUDOLPH, S. *Foundations of Semantic Web Technologies*. Textbooks in Computing, Chapman and Hall/CRC Press, 2010.

[38] HOFSTADTER, D. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, 1999.

[39] JANOWICZ, K. Sim-DL: Towards a semantic similarity measurement theory for the description logic $\mathcal{ALCNR}$ in geographic information retrieval. In *On the Move to Meaningful Internet Systems, Proc. OTM, Part II*, R. Meersman, Z. Tari, and P. Herrero, Eds., vol. 4278 of *Lecture Notes in Computer Science*. Springer, 2006, pp. 1681–1692. doi:10.1007/11915072_74.

[40] JANOWICZ, K. Kinds of contexts and their impact on semantic similarity measurement. In *Proc. 5th IEEE Workshop on Context Modeling and Reasoning (CoMoRea), 6th IEEE International Conference on Pervasive Computing and Communication (PerCom)* (2008), IEEE Computer Society. doi:10.1109/PERCOM.2008.35.

[41] JANOWICZ, K., ADAMS, B., AND RAUBAL, M. Semantic referencing—determining context weights for similarity measurement. In *Proc. 6th International Conference Geographic Information Science (GIScience)* (2010), S. I. Fabrikant, T. Reichenbacher, M. J. van Kreveld, and C. Schlieder, Eds., vol. 6292 of *Lecture Notes in Computer Science*, Sp, pp. 70–84. doi:10.1007/978-3-642-15300-6_6.

[42] JANOWICZ, K., AND KESSLER, C. The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science 10*, 22 (2008), 1129–1157. doi:10.1080/13658810701851461.

[43] JANOWICZ, K., KESSLER, C., PANOV, I., WILKES, M., ESPETER, M., AND SCHWARZ, M. A study on the cognitive plausibility of SIM-DL similarity rankings for geographic feature types. In *Proc. 11th AGILE International Conference on Geographic Information Science (AGILE)* (2008), L. Bernard, A. Friis-Christensen, and H. Pundt, Eds., Lecture Notes in Geoinformation and Cartography, Springer, pp. 115–133. doi:10.1007/978-3-540-78946-8_7.

[44] JANOWICZ, K., KESSLER, C., SCHWARZ, M., WILKES, M., PANOV, I., ESPETER, M., AND BAEUMER, B. Algorithm, implementation and application of the SIM-DL similarity server. In *Proc. Second International Conference on GeoSpatial Semantics (GeoS)* (2007), F. T. Fonseca, A. Rodriguez, and S. Levashkin, Eds., no. 4853 in Lecture Notes in Computer Science, Springer, pp. 128–145. doi:10.1007/978-3-540-76876-0_9.

[45] JANOWICZ, K., MAUÉ, P., WILKES, M., BRAUN, M., SCHADE, S., DUPKE, S., AND KUHN, W. Similarity as a quality indicator in ontology engineering. In *Proc. 5th International Conference on Formal Ontology in Information Systems (FOIS)* (2008), C. Eschenbach and M. Grüninger, Eds., vol. 183, IOS Pres, pp. 92–105.

[46] JANOWICZ, K., SCHADE, S., BRÖRING, A., KESSLER, C., MAUE, P., AND STASCH, C. Semantic enablement for spatial data infrastructures. *Transactions in GIS 14*, 2 (2010), 111–129. doi:10.1111/j.1467-9671.2010.01186.x.

[47] JANOWICZ, K., SCHWARZ, M., AND WILKES, M. Implementation and evaluation of a semantics-based user interface for web gazetteers. In *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW)* (2009).

[48] JANOWICZ, K., AND WILKES, M. SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity. In *Proc. 6th Annual European Semantic Web Conference (ESWC)* (2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvoenen, R. Mizoguchi, E. Oren, M. Sabou, and E. P. B. Simperl, Eds., vol. 5554 of *Lecture Notes in Computer Science*, Springer, pp. 353–367. doi:10.1007/978-3-642-02121-3_28.

[49] JANOWICZ, K., WILKES, M., AND LUTZ, M. Similarity-based information retrieval and its role within spatial data infrastructures. In *Proc. 5th International Conference on Geographic Information Science (GIScience)* (2008), Springer, pp. 151–167. doi:10.1007/978-3-540-87473-7_10.

[50] JONES, C. B., AND PURVES, R. S. Geographical information retrieval. *International Journal of Geographical Information Science 22*, 3 (2008), 219–228. doi:10.1080/13658810701626343.

[51] JURISICA, I. Dkbs-tr-94-5: Context-based similarity applied to retrieval of relevant cases. Tech. rep., University of Toronto, Department of Computer Science, Toronto, 1994.

[52] KESSLER, C. What's the difference? a cognitive dissimilarity measure for information retrieval result sets. *Knowledge and Information Systems* (2011; accepted for publication).

[53] KESSLER, C., RAUBAL, M., AND JANOWICZ, K. The effect of context on semantic similarity measurement. In *On the Move to Meaningful Internet Systems, Proc. OTM Part II* (2007), R. Meersman, Z. Tari, and P. Herrero, Eds., no. 4806 in Lecture Notes in Computer Science, Springer, pp. 1274–1284. doi:10.1007/978-3-540-76890-6_55.

[54] KLIPPEL, A., LI, R., HARDISTY, F., AND WEAVER, C. Cognitive invariants of geographic event conceptualization: What matters and what refines. In *Proc. 6th International Conference on Geographic Information Science (GIScience)* (2010), S. I. Fabrikant, T. Reichenbacher, M. van Krefeld, and C. Schlieder, Eds., LNCS, Springer, pp. 130–144. doi:10.1007/978-3-642-15300-6_10.

[55] KLIPPEL, A., WORBOYS, M., AND DUCKHAM, M. Identifying factors of geographic event conceptualisation. *International Journal of Geographical Information Science*, 22(2) (2008), 183–204. doi:10.1080/13658810701405607.

[56] KRUMHANSL, C. L. Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review 85* (1978), 445–463. doi:10.1037/0033-295X.85.5.445.

[57] KÜSTERS, R. *Non-Standard Inferences in Description Logics*, vol. 2100 of *Lecture Notes in Artificial Intelligence*. Springer, 2001. doi:10.1007/3-540-44613-3.

[58] LARKEY, L., AND MARKMAN, A. Processes of similarity judgment. *Cognitive Science 29*, 6 (2005), 1061–1076. doi:10.1207/s15516709cog0000_30.

[59] LEW, M., SEBE, N., DJERABA, C., AND JAIN, R. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications 2*, 1 (2006), 1–19. doi:10.1145/1126004.1126005.

[60] LI, B., AND FONSECA, F. Tdd—a comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation 6*, 1 (2006), 31–62. doi:10.1207/s15427633scc0601_2.

[61] LIN, D. An information-theoretic definition of similarity. In *Proc. 15th International Conference on Machine Learning* (1998), Morgan Kaufmann, pp. 296–304.

[62] LUTZ, M., AND KLIEN, E. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science 20*, 3 (2006), 233–260. doi:10.1080/13658810500287107.

[63] MARK, D., TURK, A., AND STEA, D. Does the semantic similarity of geospatial entity types vary across languages and cultures? In *Workshop on Semantic Similarity Measurement and Geospatial Applications, COSIT 2007* (2007).

[64] MARKMAN, A. B. *Similarity and Categorization*. Oxford University Press., 2001, ch. Structural alignment, similarity, and the internal structure of category representations., pp. 109–130.

[65] MARKMAN, A. B., AND GENTNER, D. Structural alignment during similarity comparisons. *Cognitive Psychology 25*, 4 (1993), 431–467. doi:10.1006/cogp.1993.1011.

[66] MARKMAN, A. B., AND GENTNER, D. Structure mapping in the comparison process. *American Journal of Psychology 113* (2000), 501–538. doi:10.2307/1423470.

[67] MARKMAN, A. B., AND STILWELL, C. Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence 13*, 4 (2001), 329–358. doi:10.1080/09528130110100252.

[68] MATYAS, C., AND SCHLIEDER, C. A spatial user similarity measure for geographic recommender systems. In *Proc. Third International Conference on GeoSpatial Semantics (GeoS)* (2009; forthcoming), K. Janowicz, M. Raubal, and S. Levashkin, Eds., vol. 5892 of *Lecture Notes in Computer Science*, Springer. doi:10.1007/978-3-642-10436-7_8.

[69] MEDIN, D., GOLDSTONE, R., AND GENTNER, D. Respects for similarity. *Psychological Review 100*, 2 (1993), 254–278. doi:10.1037/0033-295X.100.2.254.

[70] MÖLLER, R., HAARSLEV, V., AND NEUMANN, B. Semantics-Based Information Retrieval. In *Proc. International Conference on Information Technology and Knowledge Systems (IT&KNOWS-98)* (1998), pp. 49–56.

[71] MONTELLO, D., GOODCHILD, M., GOTTSEGEN, J., AND FOHL, P. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation 3*, 2 (2003), 185–204. doi:10.1207/S15427633SCC032&3_06.

[72] NEDAS, K., AND EGENHOFER, M. Spatial similarity queries with logical operators. In *Proc. Eighth International Symposium on Spatial and Temporal Databases*, T. Hadzilacos, Y. Manolopoulos, J. Roddick, and Y. Theodoridis, Eds., vol. 2750 of *Lecture Notes in Computer Science*. 2003, pp. 430–448. doi:10.1007/978-3-540-45072-6_25.

[73] NEDAS, K., AND EGENHOFER, M. Spatial-scene similarity queries. *Transactions in GIS 12*, 6 (2008), 661–681. doi:10.1111/j.1467-9671.2008.01127.x.

[74] NOSOFSKY, R. M. Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology 23*, 1 (1991), 94–140. doi:10.1016/0010-0285(91)90004-8.

[75] OSGOOD, C. E., SUCI, G. J., AND TANNENBAUM, P. H. *The Measurement of Meaning*. University of Illnois press, 1967.

[76] RADA, R., MILI, H., BICKNELL, E., AND BLETTNER, M. Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics 19* (1989), 17–30. doi:10.1109/21.24528.

[77] RAUBAL, M. Formalizing conceptual spaces. In *Proc. Third International Conference Formal Ontology in Information Systems (FOIS)*, A. Varzi and L. Vieu, Eds., vol. 114 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2004, pp. 153–164.

[78] RAUBAL, M. Mappings for cognitive semantic interoperability. In *Proc. 8th AGILE Conference on Geographic Information Science (AGILE)* (2005), F. Toppen and M. Painho, Eds., pp. 291–296.

[79] RAUBAL, M. Representing concepts in time. In *Spatial Cognition* (2008), C. Freksa, N. S. Newcombe, P. Gärdenfors, and S. Wölfl, Eds., vol. 5248 of *Lecture Notes in Computer Science*, Springer, pp. 328–343. doi:10.1007/978-3-540-87601-4_24.

[80] RISSLAND, E. L. Ai and similarity. *IEEE Intelligent Systems 21*, 3 (2006), 39–49. doi:10.1109/MIS.2006.38.

[81] RODRÍGUEZ, A., AND EGENHOFER, M. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science 18*, 3 (2004), 229–256. doi:10.1080/13658810310001629592.

[82] SCHWERING, A. Approaches to semantic similarity measurement for geo-spatial data—a survey. *Transactions in GIS 12*, 1 (2008), 5–29. doi:10.1111/j.1467-9671.2008.01084.x.

[83] SCHWERING, A., AND RAUBAL, M. Spatial relations for semantic similarity measurement. In *Perspectives in Conceptual Modeling: ER 2005 Workshops CAOIS, BP-UML, CoMoGIS, eCOMO, and QoIS.*, J. Akoka, S. Liddle, I.-Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.-J. vanden Heuvel, M. Kolp, J. Trujillo, C. Kop, and H. Mayr, Eds., vol. 3770 of *Lecture Notes in Computer Science*. Springer, 2005, pp. 259–269. doi:10.1007/11568346_28.

[84] SHVAIKO, P., AND EUZENAT, J. Ten challenges for ontology matching. In *Proc. On the Move to Meaningful Internet Systems (OTM)* (2008), R. Meersman and Z. Tari, Eds., vol. 5332 of *Lecture Notes in Computer Science*, Springer, pp. 1164–1182. doi:10.1007/978-3-540-88873-4_18.

[85] SMITH, L. B. *Similarity and analogy*. Cambridge University Press, 1989, ch. From global similarities to kinds of similarities: The construction of dimensions in development, pp. 146–178.

[86] SUNNA, W., AND CRUZ, I. Using the agreementmaker to align ontologies for the oaei campaign 2007. In *Proc. Second International Workshop on Ontology Matching, 6th International Semantic Web Conference (ISWC)* (2007).

[87] TAN, P.-N., STEINBACH, M., AND KUMAR, V. *Introduction to Data Mining*. Addison Wesley, 2005.

[88] TVERSKY, A. Features of similarity. *Psychological Review 84*, 4 (1977), 327–352. doi:10.1037/0033-295X.84.4.327.

[89] TVERSKY, A., AND GATI, I. Similarity, separability, and the triangle inequality. *Psychological Review 89(2)* (1982), 123–154. doi:10.1037/0033-295X.89.2.123.

[90] WILKES, M., AND JANOWICZ, K. A graph-based alignment approach to similarity between climbing routes. In *Proc. First International Workshop on Information Semantics and its Implications for Geographic Analysis (ISGA)* (2008).

[91] YEH, W., AND BARSALOU, L. The situated nature of concepts. *American Journal of Psychology 119* (2006), 349–384. doi:10.2307/20445349.