On ontology-based querying

Troels Andreasen and **Henrik Bulskov** and **Rasmus Knappe**
Department of Computer Science,
Roskilde University,
P.O. Box 260, DK-4000 Roskilde, Denmark
{troels,bulskov,knappe}@ruc.dk

Abstract

In this paper we introduce an approach to exploit knowledge represented in an ontology in answers to queries to an information base. We assume that the ontology is embedded in a knowledge base covering the domain of the information base. The ontology is first of all to influence ranking of objects in answers to queries as measured by similarity to the query. We consider a generative framework where an ontology in combination with a concept language defines a set of well-formed concepts. Well-formed concepts is assumed to be the basis for an indexing of the information base in the sense that these concepts appear as descriptors attached to objects in the base. Concepts are thus applied to obtain a means for descriptions that generalizes simple word-based information base indexing. In effect query evaluation is generalized to be a matter of comparison at the level of concepts rather than words.

1 Introduction

The approach presented here concerns ontology-based querying. For the information base targeted for querying we assume an ontology (probably embedded in a knowledge base) covering the domain of the information base.

The aim is to utilize knowledge from a domain-specific ontology to obtain better and closer answers on a semantical basis comparing concepts rather than words. Better answers are primarily better ranked information base objects which in turn is a matter of better means for computing the similarity between a query and an object from the base.

The ontology plays its role behind the scenes – it defines and relates the concepts that are the basis for comparing queries and answers. However, even though it may for other reasons be relevant, it is not essential that the ontology and the concepts and relations it encloses are revealed to users. For this reason issues on editing, browsing and visualization of the ontology become subordinate and the problem of representation of ontology can be dealt with in a different perspective.

Our claim is that when the ontology is no longer the primary base in focus, more restrictive language with less expressive power is more suited in the present context. The main argument for this is that we can do with an incremental volume of knowledge represented in the ontology. Even very small fragments from a domain, such as a few related concepts, makes sense as an ontology if only there are queries with answers that can be improved from this. There is no need at all to insist on completeness on the coverage of a domain or a subdomain.

We consider a generative framework where an ontology in combination with a concept language defines a set of well-formed concepts. Well-formed concepts is assumed to be the basis for an indexing of the information base in the sense that these concepts appear as descriptors attached to objects in the base. Concepts are thus applied to obtain a means for descriptions that generalizes simple word-based information base indexing. In effect query evaluation is generalized to be a matter of comparison at the level of concepts capturing fragments of meaning rather than words.

The goal is thus a semantic basis for querying in text retrieval environments. In this context, one of the major problems is to determine the similarity between the semantic elements. It is no longer only simple match of keywords in the text objects, but also the meaning of them, we have to take into consideration when we calculate the similarity between queries and objects in the base.

The foundation of this paper is our previous work[Bulskov *et al.*, 2002] and our affiliation to the interdisciplinary research project ONTOQUERY(Ontology-based Querying)[Andreasen *et al.*, 2000; 2002b; OntoQuery, 2002].

An Environment for Ontology-based Querying

As introduced in the following section we consider a generative ontology that defines a set of well-formed concepts from a basis ontology. This basis ontology defines a vocabulary of concepts and situates these in a concept inclusion lattice (a taxonomy). We assume an environment where queries as well as objects from the base are attached descriptions formed from descriptors which basically are well-formed concepts. Query evaluation is then a matter of comparison of descriptions.

The environment for this type of querying may be a system that automatically can produce conceptual descriptions (conceptual indexing) of text objects and support textual/word list queries by initial transformation into descriptions.

2 A generative ontology

The purpose of the ontology is to define and relate concepts that can be used in descriptions. The ontology framework is generative in the following sense. A basis ontology defines a set of atomic concepts and situates these in a concept inclusion lattice, which basically is a taxonomy over single or multi-word concepts that are treated as atomic in the modelling of the domain. In combination with a given basis ontology, a concept language (description language) defines a set of well-formed concepts.

The concept language in focus here, ONTOLOG[Nilsson, 2001], defines a set of semantic relations which can be used for "attribution" (feature-attachment) to form compound concepts. The suitable number of available relations may vary with different domains, but among the more important relations that probably will be present in most domain modellings are WRT (With-respect-to), CHR (Characterized-by), CBY (Caused-by), TMP (Temporal), LOC (Location).

Expressions in ONTOLOG are descriptions of concepts situated in an ontology formed by an algebraic lattice with concept inclusion (ISA) as the ordering relation.

Attribution of concepts – combining atomic concepts into compound concepts by attaching attributes – can be written as a feature structures. Simple attribution of a concept c_1 with relation r and a concept c_2 is denoted $c_1[r\colon c_2]$.

We assume a set of atomic concepts **A** and a set of semantic relations **R**, as indicated with **R**$=\{\text{WRT}, \text{CHR}, \text{CBY}, \text{TMP}, \text{LOC}, \ldots\}$. Then the set of well-formed terms **L** of the ONTOLOG language is recursively defined as follows.

- if $x \in \mathbf{A}$ then $x \in \mathbf{L}$
- if $x \in \mathbf{L}$, $r_i \in \mathbf{R}$ and $y_i \in \mathbf{L}, i = 1, \ldots, n$ then $x[r_1\colon y_1, \ldots, r_n\colon y_n] \in \mathbf{L}$

It appears that compound terms can be built from nesting, for instance $c_1[r_1\colon c_2[r_2\colon c_3]]$ and from multiple attribution as in $c_1[r_1\colon c_2, r_2\colon c_3]$. The attributes of a multiple attributed term $T = x[r_1\colon y_1, \ldots, r_n\colon y_n]$ is considered as a set, thus we can rewrite T with any permutation of $r_1\colon y_1, \ldots, r_n\colon y_n$.

The basis for the ontology is a simple taxonomic concept inclusion relation ISA_{KB}, which is atomic in the sense that it defines a relation over the atomic concepts **A**. It is considered as domain or world knowledge and may for instance express the view of a domain expert. We distinguish this (knowledge base) relation ISA_{KB} because concepts are assumed to be related by specific knowledge over the domain. For that reason we cannot expect the relation to be transitively closed. The relation ISA is the transitive closure of ISA_{KB}, while the relation $\text{ISA}_{\text{REDUC}}$ is the transitive reduction of ISA_{KB}.

Based on ISA, the transitive closure of ISA_{KB}, we can generalize into a relation over all well-formed terms of the language **L** by the following.

- if x ISA y then $x \leq y$
- if $x[\ldots] \leq y[\ldots]$ then also
  $x[\ldots, r\colon z] \leq y[\ldots]$, and
  $x[\ldots, r\colon z] \leq y[\ldots, r\colon z]$,
- if $x \leq y$ then also

$$z[\ldots, r\colon x] \leq z[\ldots, r\colon y]$$

where repeated ... in each inequality denotes identical lists of zero or more attributes of the form $r_i\colon w_i$ The purpose of the language introduced above is to describe fragments of meaning in text at a more thoroughly way than what can by obtained from simple keywords, while still refraining from full meaning representations which is obviously not realistic in general search applications (with a huge database).

Take as an example the sentence: "*the black dog is making noise*" which can be translated into this semantic expression *noise[*CBY*: dog[*CHR*: black]]*.

Descriptions of text expressed in this language goes beyond simple keyword descriptions partly due to formation of compound terms and to the reference to the ontology. A key question in the framework of querying is of course the definitions of similarity or nearness of terms, now that we no longer can rely on simple matching of keywords.

3 From Ontology to Similarity

In building a query evaluation principle that draws on an ontology, a key issue is of course how the ontology influence the matching of values, that is, how the different relations of the ontology may contribute to similarity.

We have to decide for each relation to what extent related values are similar and we must build similarity functions, mapping values into similarities, that reflect these decisions.

We discuss below how to introduce similarity upon an ontology. We introduce firstly a shortest-path approach to similarity based on the key ordering relation in the ontology, ISA. Based on a definition for atomic concepts of the basis ontology we discuss how to extend the notion of similarity to cover general compound concepts as expressions in the language ONTOLOG. Secondly we introduce an alternative approach for devising a similarity measure based on the notion of shared nodes corresponding to lattice join in the lattice of the arguments of the similarity function. This approach can be considered as taking into account not only the shortest path but in principle all possible paths connecting two concepts.

3.1 Shortest-path similarity on atomic concepts

The concept inclusion relation plays a central role as the ordering relation that bind the ontology in a lattice. Concept inclusion intuitively imply strong similarity in the opposite direction of the inclusion (specialization), but also the direction of the inclusion (generalization) must contribute with some degree of similarity. Take as an example the small fraction of an ontology in figure 1. With reference to this ontology the atomic concept *dog* can be directly expanded to cover also *poodle* and *alsatian*.

This expansion respects the ontology in the sense that every instance of the extension of the expanded concept *dog* (that is, every element in the union of the extensions of *dog*, *poodle* and *alsatian*) by definition bear the relation ISA to *dog*. The intuition is that to a query on *dog* an answer including instances *poodle* is satisfactory (a specific answer to a general query). Since the hyponymy relation obviously is transitive we can by the same argument expand to further specializations e.g. to include *poodle* in the extension of *animal*.

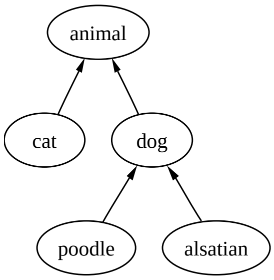

Figure 1: Inclusion relation (ISA_{KB}) with upwards reading, e.g. *dog* ISA_{KB} *animal*.

However similarity exploiting the lattice should also reflect 'distance' in the relation. Intuitively greater distance (longer path in the relation graph) corresponds to smaller similarity.

Further also generalization should contribute to similarity. Of course it is not strictly correct in an ontological sense to expand the extension of *dog* with instances of *animal*, but because all *dogs* are *animals*, *animals* are to some degree similar to *dogs*. This substantiates that also a property of generalization similarity should be exploited and, for similar reasons as in the case of specializations, that also transitive generalizations should contribute with decreasing degree of similarity.

A concept inclusion relation can be mapped into a similarity function in accordance with the described intuition as follows.

Assume an ontology given as a domain knowledge relation ISA_{KB}. Figure 1 shows an example. The corresponding transitive closure relation ISA includes for instance also *poodle* ISA *animal*. To make "distance" influence similarity we need to consider the transitively reduced relation $\text{ISA}_{\text{REDUC}}$. Similarity reflecting distance can then be measured from path-length in the graph corresponding to the $\text{ISA}_{\text{REDUC}}$ relation. A similarity function sim based on distance in $\text{ISA}_{\text{REDUC}}$ $dist(X, Y)$ should have the properties:

1. sim: $U \times U \rightarrow [0, 1]$, where U is the universe of concepts
2. $sim(x, y) = 1$ only if $x = y$
3. $sim(x, y) < sim(x, z)$ if $dist(x, y) > dist(x, z)$

By parameterizing with two factors δ and γ expressing similarity of immediate specialization and generalization respectively, we can define a simple similarity function: If there is a path from nodes (concepts) x and y in the hyponomy relation then it has the form

$$P = (P_1, \cdots, P_n)$$

where

$$P_i \text{ISA}_{\text{REDUC}} P_{i+1} \text{ or } P_{i+1} \text{ISA}_{\text{REDUC}} \ P_i$$

for each i with $X = P_1$ and $Y = P_n$.

Given a path $P = (P_1, \cdots, P_n)$, set $s(P)$ and $g(P)$ to the numbers of specializations and generalizations respectively along the path P thus:

$$s(P) = |\{i | P_i \text{ISA}_{\text{REDUC}} P_{i+1}\}|$$

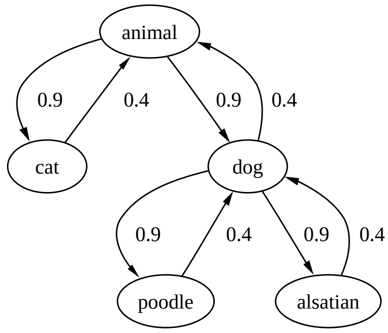

Figure 2: The ontology transformed into a directed weighted graph, with the immediate specialization and generalization similarity being σ = 0.9 and γ = 0.4 respectively as weights. Similarity is derived as maximal (multiplicative) weighted path length, thus $sim(poodle, alsatian) = 0.4 * 0.9 = 0.36$.

and

$$g(P) = |\{i | P_{i+1} \text{ISA}_{\text{REDUC}} P_i\}|$$

If $P^1, \cdots, P^m$ are all paths connecting X and Y then the degree to which Y is similar to X can be defined as

$$sim(X, Y) = \max_{j=1,\ldots,m} \left\{ \sigma^{s(P^j)} \gamma^{g(P^j)} \right\} \quad (1)$$

This similarity can be considered as derived from the ontology by transforming the ontology into a directional weighted graph, with σ as downwards and γ as upwards weights and with similarity derived as the product of the weights on the paths. Figure 2 shows the graph corresponding to the ontology in figure 1. An atomic concept T can then be expanded to a fuzzy set, including T and similar values $T_1, T_2, \ldots, T_n$ as in:

$$T+ = 1/T + sim(T, T_1)/T_1 + \cdots + sim(T, T_n)/T_n \quad (2)$$

Thus for instance with σ = 0.9 and γ = 0.4 the expansion of the concepts *dog*, *animal* and *poodle* into sets of similar values would be:

$dog+ = 1/dog + 0.9/poodle + 0.9/alsatian + 0.4/animal$

$poodle+ = 1/poodle + 0.4/dog + 0.36/alsatian + 0.16/animal + 0.144/cat$

$animal+ = 1/animal + 0.9/cat + 0.9/dog + 0.81/poodle + 0.81/alsatian$

3.2 General shortest-path similarity

The semantic relations, used in forming concepts in the ontology, indirectly contribute to similarity through subsumption. For instance *noise*[CBY: *dog* [CHR: *black*]] is subsumed by - and thus extensionally included in - each of the more general concepts *noise*[CBY: *dog*] and *noise*. Thus with a definition of similarity covering atomic concepts, and in some sense reflecting the ordering relation (concept inclusion), we can extend to similarity on compound concepts by a relaxation, which takes subsumed concepts into account when comparing descriptions.

The principle can be considered to be a matter of subsumption expansion. Any compound concept is expanded (or relaxed) into the set of subsuming concepts, thus

$$noise[\text{CBY}: dog[\text{CHR}: black]]$$

is expanded to the set

$$\{noise, \quad noise[\text{CBY}: \quad dog], \; noise[\text{CBY}: dog[\text{CHR}: black]]\}$$

One approach to query-answering in this direction is to expand the description of the query along the ontology and the potential answer objects along subsumption.

For instance a query on dog could be expanded to a query on similar values like:

$$dog+ = 1/dog + \ldots + 0.4/animal + \ldots$$

and a potential answer object like $noise[\text{CBY}: dog[\text{WRT}: black]]$ would then be expanded as exemplified above.

While not the key issue here, we should point out the importance of applying an appropriate averaging aggregation when comparing descriptions. It is essential that similarity based on subsumption expansion, exploits that for instance the degree to which $c[r_1 : c_1]$ is matching $c[r_1 : c_1[r_2 : c_2]]$ is higher than the degree for c with no attributes is matching $c[r_1 : c_1[r_2 : c_2]]$. Approaches to aggregation that can be tailored to obtain these properties, based on order weighted averaging[Yager, 1988] and capturing nested structuring[Yager, 2000], are described in [Andreasen, 2002a; 2002b].

An alternative to the above described subsumption expansion is to include edges corresponding to semantic relations in the computation of shortest path similarity as a generalization of the principle of aggregating weights by multiplying cost factors described in the previous subsection. While the similarity between c and $c[r_1 : c_1]$ can be claimed to be justified by the ontology formalism (subsumption) or simply by the fact that $c[r_1 : c_1]$ ISA c, it is not strictly correct in an ontological sense to claim similarity likewise between c_1 and $c[r_1 : c_1]$.

For instance $noise[\text{CBY}: dog]$ is conceptually not some kind of a dog. On the other hand it would be reasonable to claim that $noise[\text{CBY}: dog]$ in a broad sense has something to do with (and thus has similarities to) dog (simply supported by the fact that concept $noise[\text{CBY}: dog]$ is present in the base). Most examples tend to reveal the same characteristics and this phenomenon is one good explanation for the comparative success of conventional word-based querying approaches. Basically the (incorrect) assumption of no correlation between words in NL phrases, which is underlying any strictly word-based approach, does not lead to serious failure because the correlation that appears is not dominating.

This could of course be an argument for not looking at compound concepts at all, but rather these considerations points in the direction of redrawing some of the importance of correlation in NL phrases when developing similarity measures.

Consider figure 3. The solid edges are ISA references and the broken are references by other semantic relations – in this example CBY and CHR are in use. Each compound concept has broken edges to its attribution concept. Strictly the spelling out of the compound concept expression as the label of a node is redundant since the concept expression can be derived from the connecting edges.

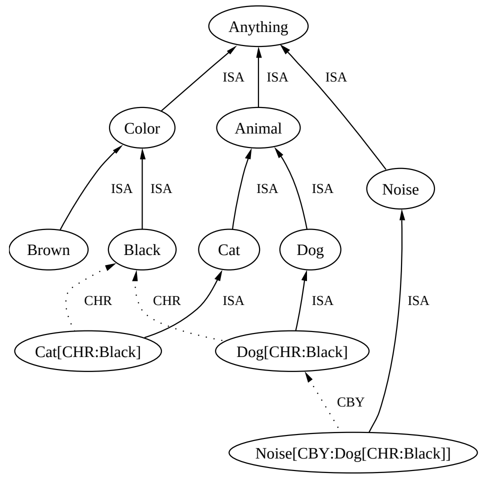

Figure 3: An ontology where attribution with semantic relations is shown as dotted edges.

The principle of weighted path similarity can be generalized by introducing similarity factors for the semantic relations. The extensional arguments used to argue for differentiated weights depending on direction does not apply to semantic relations and seemingly there is no obvious way to differentiate based on direction at all. Thus one approach in the generalization is simply to introduce a single similarity factor and to transform to bidirectional edges.

Assume that we have k different semantic relations $R^1, \ldots, R^k$ and let $\rho_1, \cdots, \rho_k$ be the attached similarity factors. Given a path $P = (P_1, \cdots, P_n)$, set $r^j(P)$ to the number of R^j edges along the path P thus:

$$r^j(P) = \left|\left\{ i \mid \; P_i \quad R^j \quad P_{i+1} \; \right\}\right| \tag{3}$$

If $P^1, \cdots, P^m$ are all paths connecting c_1 and c_2 then the degree to which Y is similar to X can be defined as

$$sim(X,Y) = \max_{j=1,\ldots,m} \left\{ \sigma^{s(P^j)} \gamma^{g(P^j)} {\rho_1}^{r^1(P^j)} \cdots {\rho_k}^{r^k(P^j)} \right\} \tag{4}$$

The result of transforming the ontology in figure 3 is shown in 4 . Here two semantic relations CHR and CBY are in use. The corresponding edge count functions are r^{WRT} and r^{CBY} and the attached similarity factors are denoted ρ_{WRT} and ρ_{CBY}. The figure shows the graph with the attached similarity factors as weights. Again the degree to which a concept c_1 is similar to a concept c_2 is based on shortest path (and derived as the maximum of the products of edge weights over the set of paths connecting c_1 and c_2).

For instance we can derive from figure 4 that $sim(cat, dog) = 0.9 * 0.4 = 0.36$ and $sim(cat[\text{CHR}: black], color) = 0.3 * 0.4 = 0.12$.

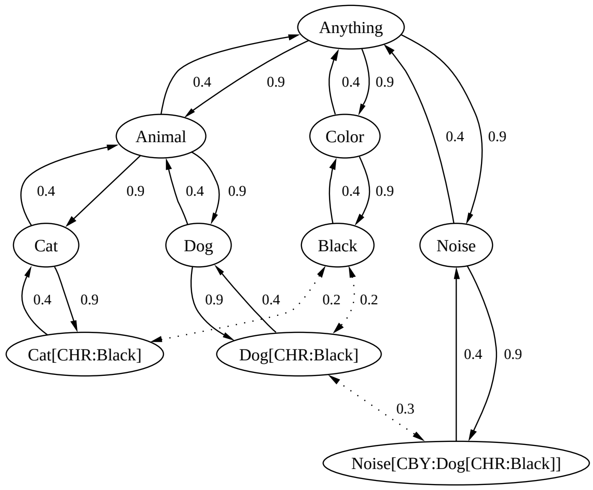

Figure 4: The ontology of figure 3transformed into a directional weighted graph with the similarity factors for specialization: σ = 0.9, for generalization: γ = 0.4, for CBY: ρ_{CBY} = 0.3 and for CHR: ρ_{WRT} = 0.2.

The weights in the example are assigned in a rather ad hoc manner. Such assignment in practice needs a careful effort by domain experts. Furthermore the similarity principle in general needs to be verified empirically.

3.3 Shared Nodes Similarity

The shortest path approach described above is straightforward and does not entail computational problems. However one aspect that must be assumed to contribute to similarity is ignored. When two concepts are connected by multiple paths only the shortest contribute to similarity. Considering the ontology in figure 4 the similarities between $cat[\text{CHR} : black, \text{CHR} : brown]$ and $dog[\text{CHR} : black, \text{CHR} : brown]$ will not be greater than the similarity between cat and dog. This example shows that other connections than the shortest path in some cases should contribute and it also indicates that similarity should be proportional to the number of possible paths connecting two concepts. Obviously a similarity measure that takes into account all possible paths will impose increased computational complexity and calls for considerations on possible optimization approaches.

In this direction we suggest a similarity measure that utilizes a well-defined subset of all possible paths. The goal then is to encircle a basis in the form of a subontology for measuring similarity.

In broad terms our simplified "all-possible-paths" approach is a "shared nodes" approach, where shared nodes between two concepts are nodes that are "upwards reachable" from both concepts and where the similarity is proportional to the number of shared nodes.

To this end we define first the term-decomposition $\tau(c)$ and the upwards expansion $\omega(c)$ of a concept term c. The term-decomposition is defined as the set of all subterms of c, which thus includes all concepts subsuming c and all attributes of subsuming concepts for c. The term-decomposition is defined as follows:

$$\tau(c) = \{x | c \leq x \vee c \leq y[r\colon x], x \in \mathbf{L}, y \in \mathbf{L}, r \in \mathbf{R}\}$$

As an example the term $noise[\text{CBY}\colon dog[\text{CHR}\colon black]]$ decomposes to resulting in the set containing the following concepts:

$$\tau(noise[\text{CBY}\colon dog[\text{CHR}\colon black]]) = \{noise[\text{CBY}\colon dog[\text{CHR}\colon black]], noise[\text{CBY}\colon dog]], noise, dog[\text{CHR}\colon black], dog, black\}$$

The upwards expansion $\omega(C)$ of a set of terms C is the transitive closure of C with respect to ISA_{KB}.

$$\omega(C) = \{x | x \in C \vee y \in C, y \text{ ISA } x\}$$

This expansion thus only adds atoms to C.

We define further the upwards spanning subgraph (subontology) $\gamma(C)$ for a set of concepts $C = \{c_1, \ldots, c_n\}$ as the graph that appears when decomposing C and connecting the resulting set of terms with edges corresponding to the ISA_{KB}relation and to the semantic relations used in attribution of elements in C. We define the triple (x, y, r) as the edge of type r from concept x to concept y.

$$\gamma(C) = \begin{array}{l} \{(x, y, \text{ISA}) | x, y \in \omega(\tau(C)), x \text{ISA}_{\text{REDUC}}\ y\} \\ \cup \\ \{(x, y, r) | x, y \in \omega(\tau(C)), r \in \mathbf{R}, x[r\colon y] \in \tau(C)\} \end{array}$$

Figure 5 shows an example of such an subontology spanned by the two terms.

Now a shared node between concepts c_1 and c_2 is a node that is reachable from both c_1 and c_2. With the example in figure 5 both $Animal$ and $Black$ are shared nodes for $Cat[\text{CHR} : Black]$ and $Dog[\text{CHR} : Black]$.

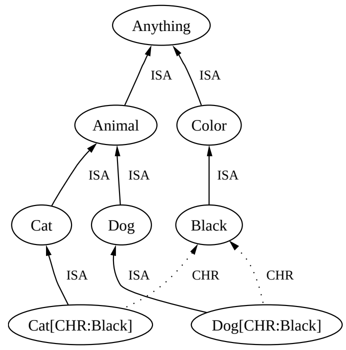

Figure 5: An example of an upwards spanning subgraph for the concepts $Cat[\text{CHR} : Black]$ and $Dog[\text{CHR} : Black]$ where $Animal$ and $Black$ (and $Anything$ and $Color$) are shared nodes.

If we for a concept c defines $\alpha(c)$ to be the set of nodes reachable from c, that is, $\alpha(c) = \omega(\tau(c))$, then $\alpha(c_1) \bigcap \alpha(c_2))$ is the set of shared nodes for two concepts c_1 and c_2.

A very simple similarity measure based on shared nodes can be defined as

$$sim(x, y) = \frac{|\alpha(x) \cap \alpha(y)|}{|\alpha(y)|}$$

It appears that this measure has good properties in sustaining to the intuition behind the ontology. First of all we can see that the similarity $sim(cat[\text{CHR} : black, \text{CHR} : brown], dog[\text{CHR} : black, \text{CHR} : brown])$ is increased as compared to the similarity $sim(cat, dog)$. We have that when $c_1 \leq c_2$ then $sim(c_1, c_2)$ is smaller than $sim(c_2, c_1)$ (a general concept is not as good as replacement for a specific as vise-versa). Furthermore it follows that steps along edges become more expensive when the edges are closer to the top of the ontology.

However it appears that $sim(x, y)$ is independent of nestings of x. For instance consider the following example where the similarity between $dog[\text{CHR} : white, \text{LOC} : tarmac[\text{CHR} : black]]$ and $cat[\text{CHR} : black]$ is equal to the similarity between $dog[\text{CHR} : white, \text{LOC} : tarmac[\text{CHR} : black]]$. Obviously there is a need for refinement of the similarity measure, that takes into consideration the nesting of x in $sim(x, y)$. This is subject for further investigation.

The inclusion of non-minimal paths in the computation of the similarity measure has as a consequence, that the locations of the concepts in the ontology, influences the measure. This is due to the fact that concepts in the upper part of the ontology have fewer potential paths than concepts in the lower part. One could argue, from a pragmatic point of view, that concepts with longer common paths to the top of the ontology are stronger connected, which substantiate the intuition of increased similarity.

4 Conclusion

We have described different principles for measuring similarity between both atomic and compound concepts, all of which incorporate meta knowledge.

- Similarity between atomic concepts based on distance in the ordering relation of the ontology, concept inclusion ($\text{ISA}_{\text{REDUC}}$).
- Similarity between general compound concepts based on subsumption expansion.
- Similarity between both atomic and general compound concepts based on shared nodes.

The notion of measuring similarity as distance, either in the ordering relation or in combination with the semantic relations, seems to indicate a usable theoretical foundation for design of similarity measures.

The purpose of similarity measures in connection with querying is of course to look for similar rather than for exactly matching values, that is, to introduce soft rather than crisp evaluation. As indicated through examples above one approach to introduce similar values is to expand crisp values into fuzzy sets including also similar values. Expansion of this kind, applying similarity based on knowledge in the knowledge base, is a simplification replacing direct reasoning over the knowledge base during query evaluation. The graded similarity is the obvious means to make expansion a useful - by using simple threshold values for similarity the size of the answer can be fully controlled.

Acknowledgments

The work described in this paper is part of the OntoQuery[1] project supported by the Danish Technical Research Council and the Danish IT University.

References

[Bulskov *et al.*, 2002] Bulskov, H., Knappe, R. and Andreasen, T.: On Measuring Similarity for Conceptual Querying, LNAI 2522, pp. 100-111 in T. Andreasen, A. Motro, H. Christiansen, H.L. Larsen (Eds.): Flexible Query Answering Systems 5th International Conference, FQAS 2002. Copenhagen, Denmark, October 27-29, 2002. Proceedings

[Andreasen, 2002a] Andreasen, T.: On knowledge-guided fuzzy aggregation. In *IPMU'2002, 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 1-5 July 2002, Annecy, France

[Andreasen, 2002b] Andreasen, T.: Query evaluation based on domain-specific ontologies. In *NAFIPS'2001, 20th IFSA / NAFIPS International Conference Fuzziness and Soft Computing*, pp. 1844-1849, Vancouver, Canada, 2001.

[Andreasen *et al.*, 2000] Andreasen, T., Nilsson, J. Fischer & Thomsen, H. Erdman: Ontology-based Querying, in Larsen, H.L. *et al.* (eds.) Flexible Query Answering Systems, *Flexible Query Answering Systems, Recent Advances*, Physica-Verlag, Springer, 2000. pp. 15-26.

[Andreasen *et al.*, 2002a] Andreasen, T., Jensen, P. Anker, Nilsson, J. Fischer, Paggio, P., Pedersen, B. Sandford & Thomsen, H. Erdman: *OntoQuery:* Ontology-based Querying of Texts, AAAI 2002 Spring Symposium, Stanford, California, 2002.

[Andreasen *et al.*, 2002b] Andreasen, T., Jensen, P. Anker, Nilsson, J. Fischer, Paggio, P., Pedersen, B. Sandford & Thomsen, H. Erdman: Ontological Extraction of Content for Text Querying, to appear in NLDB 2002, Stockholm, Sweden, 2002.

[Jensen *et al.*, 2001] Jensen, P. Anker & Skadhauge, P. (eds.): Proceedings of the First International OntoQuery Workshop *Ontology-based interpretation of NP's*, Department of Business Communication and Information Science, University of Southern Denmark, Kolding, 2001.

[Meadow *et al.*, 2000] Meadow, C.T., Boyce, B.R., Kraft, D.H.: Text information retrieval systems, second edition, Academic Press, 2000.

[Nistrup Madsen *et al.*, 2001] Nistrup, B. Madsen and B. Sandford Pedersen and H. Erdman Thomsen: Semantic Relations in Content-based Querying Systems: a

[1] The project has the following participating institutions: Centre for Language Technology, The Technical University of Denmark, Copenhagen Business School, Roskilde University, and the University of Southern Denmark.

Research Presentation from the OntoQuery Project, in [Jensen *et al.*, 2001].

[Nilsson, 2001] Nilsson, J. Fischer: A Logico-algebraic Framework for Ontologies ONTOLOG, in [Jensen *et al.*, 2001].

[OntoQuery, 2002] ONTOQUERY project net site: www.ontoquery.dk

[Yager, 1988] Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making, in IEEE Transactions on Systems, Man and Cybernetics, vol 18, 1988.

[Yager, 2000] Yager, R.R.: A hierarchical document retrieval language, in Information Retrieval vol 3, Issue 4, Kluwer Academic Publishers pp. 357-377, 2000.