Semantic Similarity between Queries in QA System using a Domain-specific Taxonomy

Hilda Kosorus[1], Andreas Bögl[2] and Josef Küng[1]
[1]*Institute of Applied Knowledge Processing, Johannes Kepler University, Altenbergerstraße 69, Linz, Austria*
[2]*MEOVI, Hagenberg, Austria*

Keywords: Query recommendation, Semantic similarity, Short text similarity, Taxonomy.

Abstract: Semantic similarity has been extensively studied in the past decades and has become a rapidly growing field of research. Sentence or short text similarity measures play an important role in text-based applications, such as text mining, information retrieval and question answering systems. In this paper we consider the problem of semantic similarity between queries in a question answering system with the purpose of query recommendation. Our approach is based on an existing domain-specific taxonomy. We define innovative three-layered semantic similarity measures between queries using existing similarity measures between ontology concepts combined with various set-based distance measures. We then analyse and evaluate our approach against human intuition using a data set of 90 questions. Further on, we argue that these measures are taxonomy-dependent and are influenced by various factors: taxonomy structure, keyword mappings, keyword weights, query-keyword mappings and the chosen concept similarity measure.

1 INTRODUCTION

Current implementations of QA systems that incorporate a recommendation mechanism are based on (i) methods using external sources, like user profiles, (ii) methods based on expectations (e.g. query patterns, models) or (iii) methods using query logs (Marcel and Negre, 2011). These methods do not take into account the semantic meaning of queries. In the past two decades researchers have been studying semantic similarity in order to improve information retrieval and develop intelligent semantic systems.

A semantic sentence similarity measure can have an important role in the development of a query recommender system. Nevertheless, such measures can be successfully used in other directions, like query clustering for discovering "hot topics" or to find the query that best represents a cluster, pattern recognition for identifying user groups or in web page retrieval to calculate page title similarities.

Studies of semantic similarity in the past decades has been focusing on two extremes: either measuring the similarity between single words or concepts or between documents. However, there is a growing need for an effective method to compute short text similarity. Web search technologies incorporate tasks, such as query reformulation, query recommendation, sponsored search and image retrieval, that rely on accurately computing similarity between two very short segments of text. Unfortunately, traditional techniques for detecting similarity between documents and queries fail when directly applied to these tasks. Such methods rely on analysing shared words or the co-occurence of terms in both the query and the document.

In this paper we define innovative three-layered semantic similarity measures between queries using existing similarity measures between ontology concepts combined with various set-based distance measures. We then analyze and evaluate our approach against human intuition using a dataset of 90 questions. The goal of this paper is to present semantic query similarity measures that can be successfully integrated into query recommender systems and to evaluate and compare them in terms of human judgement.

The rest of the paper is structured as follows. In section 2 we review related work in the area of semantic similarity measures between concepts, between sets of concepts and the area of short text similarity. In section 3 we present and define the domain-specific taxonomy on which our semantic similarity measures are based. In section 4 we introduce similarity measures between queries as a combination of topic similarity and keyword similarity using the defined ta-

Kosorus H., Bögl A. and Küng J..
Semantic Similarity between Queries in QA System using a Domain-specific Taxonomy.
DOI: 10.5220/0003965902410246
In *Proceedings of the 14th International Conference on Enterprise Information Systems* (ICEIS-2012), pages 241-246
ISBN: 978-989-8565-10-5

xonomy. In section 5 we analyze and evaluate these similarity measures. Finally, in section 6 we summarize the contents of this paper, drawing some important conclusions and present our future work.

2 RELATED WORK

The problem of similarity is a heavily researched subject in particular in information retrieval, but also in general in computer science, artificial intelligence, philosophy and natural language processing. Measuring similarity between documents has a long tradition in information retrieval, but these approaches compare only vectors of document features (Burgess et al., 1998; Landauer et al., 1998a; Landauer et al., 1998b), usually single words or word stems, by counting their occurrence in the document.

There is extensive literature on measuring similarity between concepts within a taxonomy (Rada et al., 1989; Lee et al., 1993; Wu and Palmer, 1994; Resnik, 1995; Jiang and Conrath, 1997; Leacock and Chodorow, 1998; Lin, 1998; Resnik, 1999; Li et al., 2003; Bouquet et al., 2004; Haase et al., 2004; Cordì et al., 2005; Al-Mubaid and Nguyen, 2006; Wang et al., 2006; Lee et al., 2008; Dong et al., 2009; Bin et al., 2009), while there are few publications that cover the area of short text semantic similarity (Li et al., 2006; O'Shea et al., 2010; Oliva et al., 2011) and some related to semantic similarity between sets of concepts (Bouquet et al., 2004; Haase et al., 2004; Cordì et al., 2005). In (Li et al., 2006) it is argued that existing long text similarity measures have some limitations and drawbacks and their performance is unsatisfactory when applied to short sentences.

In the following we will briefly present the related research in the domain of semantic similarity between concepts and between sets of concepts.

2.1 Semantic Similarity between Concepts using Taxonomies

There are basically two ways of using an ontology or taxonomy to determine the semantic similarity between concepts: the *edge-based approach* and the *information content-based approach* (Resnik, 1995; Resnik, 1999; Lin, 1998). In the following we will make a short overview of the edge-based approaches.

Intuitively, the similarity of different concepts in an ontology is measured by computing the distance within the ontology. Namely, if two concepts reside closer in the ontology, then we can conclude that they are more similar. When computing the ontology distance we actually use the specialization graph of objects and we define it as being the shortest path between the two concepts (Rada et al., 1989).

Rada, Mili, Bicknell and Blettner (1989) defined the conceptual distance as

$$sim(c_1,c_2) = minimum \quad number \quad of \quad edges \quad separating \quad c_1 \quad and \quad c_2,$$

where c_1 and c_2 are the node representation of the two concepts in the ontology. Wu and Palmer (2004) redefined the edge-based similarity measure taking into account the depth of the nodes in the hierarchical graph:

$$sim(c_1,c_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3}, \qquad (1)$$

where N_1 and N_2 are the number of nodes from c_1 and c_2, respectively, to c_3, the *least common superconcept* (LCS) of c_1 and c_2, and N_3 is the number of nodes on the path from c_3 to the root node.

Li et al. (2003) defined the similarity between two concepts as:

$$sim(c_1,c_2) = \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & , \text{if } c_1 \neq c_2 \\ 1 & , \text{otherwise} \end{cases} \qquad (2)$$

where, similarly, the parameters α and β scale the contribution of the two values $l = N_1 + N_2$ and $h = N_3$. Based on the benchmark data set, they obtained the optimal parameters $\alpha = 0.2$ and $\beta = 0.6$.

2.2 Semantic Similarity between Sets of Concepts

Defining a semantic similarity measure between sets of concepts was the next step in computing semantic similarity mainly for information retrieval purposes.

In (Bouquet et al., 2004) the ontological distance between sets of concepts is computed by summing up the distances between every pair (c_1,c_2), where $c_1 \in C_1$ and $c_2 \in C_2$. Haase et al. (2004) used the edge-based similarity measure between concepts defined by Li et al. (2006) (see 2) to introduce the similarity between sets of concepts as:

$$Sim(C_1,C_2) = \frac{1}{|C_1|} \cdot \sum_{c_1 \in C_1} \max_{c_2 \in C_2} sim(c_1,c_2), \qquad (3)$$

which computes an average of distances between $c_1 \in C_1$ and the most similar concept in C_2.

In (Cordì et al., 2005) a new similarity measure between sets of concepts was introduced, which gives more weight to keyword pairs with a higher similarity, but still allowing lower values to contribute to the final outcome.

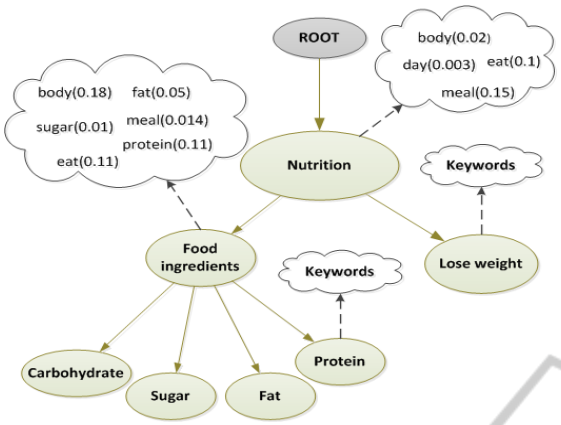

Figure 1: Snapshot of the topic-tree with keywords and their weights.

3 THE DOMAIN-SPECIFIC TAXONOMY

Before introducing our proposed semantic query similarities, it is important to understand the structure of the underlying domain-specific taxonomy. While most of the previously described similarity measures make use of the english lexical taxonomy WordNet[1], our similarity measures are based on a new domain-specific (nutrition) taxonomy with a tree-like structure, where the links between nodes represent IS-A relationships. In the following we will refer to this structure as "topic-tree".

Our topic-tree is composed of a set of *topics*:

$$\mathcal{T} = \{t_1, t_2, ..., t_n\},$$

an IS-A relationship between topics:

$$\mathcal{L} \subset \mathcal{T} \times \mathcal{T}, (t_p, t_q) \in \mathcal{L} \iff t_p \quad parent \quad of \quad t_q,$$

a set of *keywords*:

$$\mathcal{K} = \{k_1, k_2, ..., k_m\},$$

a *mapping* relationship between topics and keywords:

$$\mathcal{M} \subset \mathcal{T} \times \mathcal{K}, (t_p, k_q) \in \mathcal{M} \iff k_q \quad mapped \quad to \quad t_p,$$

and the corresponding mapping *weights*:

$$w : M \to (0, 1],$$

where the value $w(t_p, k_q)$ represents how relevant is keyword k_q for topic t_p.

Figure 1 shows a partial snapshot of the above defined taxonomy. The *topics* represent selected categories and sub-categories in the specified domain (i.e. nutrition), the mapped keywords are frequent relevant words occuring within these topics which were obtained by crawling related websites and/or documents. The corresponding weights were calculated using the TF-IDF method (Salton and Buckley, 1988).

[1] http://wordnet.princeton.edu/

4 PROPOSED SEMANTIC SIMILARITY MEASURES

Let $Q = \{q_1, q_2, ..., q_N\}$ be a set of queries in the nutrition domain. We want to define a semantic similarity measure $sim_q : Q \times Q \to [0, 1]$ between these queries using the topic-tree defined in section 3. We assume that to each query $q \in Q$ we can assign a set of keywords $\mathcal{S}_q \subset \mathcal{K}$, where $\mathcal{S}_q$ was extracted from q using some natural language processing methods (HaCohen-Kerner et al., 2005; Turney, 2000; Hulth, 2003). For example, for

$q =$ *"What type of food can I eat and at what time in order to lose weight?"*

$$\mathcal{S}_q = \{food, eat, time, lose\ weight\}.$$

In the following we will define the semantic query similarity sim_q using three other similarity measures: between topics, between keywords and between sets of keywords, each incorporating the one before.

4.1 Semantic Similarity between Topics

Let $sim_t : \mathcal{T} \times \mathcal{T} \to [0, 1]$ be the *topic similarity* function where $sim_t(t_p, t_q)$ represents the semantic similarity between two topics $t_p, t_q \in \mathcal{T}$ using the structure of the topic-tree. For our experiments, we defined sim_t using the similarity measures (1) and (2).

4.2 Semantic Similarity between Keywords

Let $sim_k : \mathcal{K} \times \mathcal{K} \to [0, 1]$ be the *keyword similarity* function where $sim_k(k_p, k_q)$ represents the semantic similarity between two keywords $k_p, k_q \in \mathcal{K}$. We define sim_k in the following way:

$$sim_k(k_p, k_q) = \frac{w_p + w_q}{2} sim_t(t_p, t_q) \qquad (4)$$

where

$$w_i = \max_{(t, k_i) \in \mathcal{M}} w(t, k_i), \quad i \in \{p, q\}$$

and

$$t_i = arg \max_{(t, k_i) \in \mathcal{M}} w(t, k_i), \quad i \in \{p, q\}.$$

4.3 Semantic Similarity between Sets of Keywords

Let $sim_{ks} : \mathcal{P}(\mathcal{K}) \times \mathcal{P}(\mathcal{K}) \to [0, 1]$ be the *keyword-set similarity* function where $sim_{ks}(\mathcal{S}_p, \mathcal{S}_q)$ represents the semantic similarity between two sets of keywords $\mathcal{S}_p, \mathcal{S}_q \subset \mathcal{K}$ and $\mathcal{P}(\mathcal{K})$ contains all subsets of $\mathcal{K}$. In the following we will introduce several possible definitions of sim_{ks} using well-known set distance measures from the literature.

4.3.1 The Sum of Maximum Similarities

The sum of minimum distances measure was originally defined by Niiniluoto (1987) to measure truthlikeness in belief revision theory. We apply the same concept to define the similarity measure sim_{ks} between sets of keywords (the *sum of maximum similarities*):

$$sim_{ks}(\mathcal{S}_p, \mathcal{S}_q) = \frac{1}{2}\left(\frac{1}{|\mathcal{S}_p|}\sum_{k_p \in \mathcal{S}_p} Sim(k_p, \mathcal{S}_q) + \frac{1}{|\mathcal{S}_q|}\sum_{k_q \in \mathcal{S}_q} Sim(k_q, \mathcal{S}_p)\right) \quad (5)$$

where

$$Sim : \mathcal{K} \times \mathcal{P}(\mathcal{K}) \rightarrow [0,1], \;\; Sim(k, \mathcal{S}) = \max_{k_s \in \mathcal{S}} sim_k(k, k_s).$$

is the semantic similarity between a keyword $k \in \mathcal{K}$ and a set of keywords $\mathcal{S} \subset \mathcal{K}$.

4.3.2 The Surjection Measure

The surjection measure was introduced by Oddie (1979), who suggested defining the distance between two sets by considering surjections that map the larger set to the smaller one. We applied this concept to measure similarity between sets of keywords, and defined *surjection similarity measure*, sim_{ks}, as

$$sim_{ks}(\mathcal{S}_p, \mathcal{S}_q) = \max_{\eta} \frac{1}{|\eta|} \sum_{(k_p, k_q) \in \eta} sim_k(k_p, k_q). \quad (6)$$

where the maximum is taken over all surjections η that maps the larger set to the smaller one.

4.3.3 The Maximum Link Similarity Measure

The minimum link distance measure was proposed in (Eiter and Mannila, 1997) as an alternative to the previously mentioned distance measures between point sets. First, let us define the *linking* between $\mathcal{S}_p$ and $\mathcal{S}_q$ as a relation $\mathcal{R} \subseteq \mathcal{S}_p \times \mathcal{S}_q$ satisfying

(a) for all $k_p \in \mathcal{S}_p$ there exists $k_q \in \mathcal{S}_q$ such that $(k_p, k_q) \in \mathcal{R}$

and

(b) for all $k_q \in \mathcal{S}_q$ there exists $k_p \in \mathcal{S}_p$ such that $(k_p, k_q) \in \mathcal{R}$.

We now apply this concept to define the *maximum link similarity* between sets of keywords as

$$sim_{ks}(\mathcal{S}_p, \mathcal{S}_q) = \max_{\mathcal{R}} \frac{1}{|\mathcal{R}|} \sum_{(k_p, k_q) \in \mathcal{R}} sim_k(k_p, k_q), \quad (7)$$

taking the maximum over all relations $\mathcal{R}$.

4.4 Semantic Similarity between Queries

Finally, we define the *query similarity* measure $sim_q : \mathcal{Q} \times \mathcal{Q} \rightarrow [0,1]$ as

$$sim_q(q_a, q_b) = sim_{ks}(\mathcal{S}_{q_a}, \mathcal{S}_{q_b}) \quad (8)$$

where $\mathcal{S}_{q_a}, \mathcal{S}_{q_b} \subset \mathcal{K}$ are the corresponding set of keywords extracted from q_a and q_b, respectively.

5 COMPARISON AND EVALUATION

In order to evaluate these similarity measures we conducted a survey with 15 persons, men and women, age between 25 and 60. We randomly sampled 50 pairs from a dataset of 90 different questions in the nutrition domain and asked the survey participants to compare and measure the relatedness of each pair by ranking them with a value between 0 and 4 (0=not related at all, 1=somehow related, 2=related, 3=very related, 4=similar).

Finally, we compared the participants' ranking against six different semantic similarity measures: the one defined by Haase et al. (3), the sum of all similarities (Bouquet et al., 2004), the one introduced by Cordì (2005), the cosine similarity (Li et al., 2003), the sum of maximum similarities (5), the surjection similarity (6) and the maximum link similarity (7).

While some question pairs were ranked almost the same by all participants (low variance), there were some cases where participants answered very differently (high variance). This reflects how *diversely* is the "relatedness" of two questions perceived by humans. Table 1 contains the mean, maximum and minimum variances calculated by question pairs rankings.

Table 2 contains the correlation values of each semantic similarity method with the average participant ranking values.

Table 1: Survey results - Variances calculated by question pair rankings.

Mean variance	0.93
Maximum variance	2.14
Minimum variance	0

Based on our experiments and the above results we make the following observations:

- The semantic similarity measures depend on the structure of the taxonomy (Bernstein et al., 2005). In our case, the topic hierarchy, the keyword-topic mappings and the assigned keyword weights affect the computed similarity.

Table 2: Correlation between survey results and the semantic similarity measures.

Method	Correlation
Haase	0.605
Sum of All	0.597
Cordì	0.563
Cosine	0.563
Sum of Maximum	0.617
Surjection	**0.634**
Maximum Link	0.626

- The similarity measure between sets of keywords, and therefore between queries, depends on the chosen topic similarity (edge-based or information content-based) and on the keyword similarity. In our experiments we used the edge-based similarity measures defined by Wu and Palmer (1994) and Li et al. (2003).

Table 3: Types of question pairs based on ranking variance and difference between average survey ranking and semantic similarity values.

Type	Var.	Diff.	Percentage
A	low	low	48%
B	high	low	20%
C	low	high	12%
D	high	high	20%

- Although the correlation between the participants' ranking and the evaluated measures are rather low (see table 2), this can be explained by the following factors:
  - the queries are selected from a specific and narrow domain (nutrition),
  - the concepts that appear in the queries are rather complex,
  - the participants' ranking for some question pairs was very diverse,
  - the participants tend to understand the ranking values or the question pair "relatedness" differently.
- The correlation results (between 0.563 and 0.634) do not contradict the fact that the semantic similarity measures reflect on some level the human perception. Most of the question pairs were evaluated by the participants and the semantic similarity measures almost the same. In our evaluation, compared to the surjection measure, 48% of the question pairs were of type A and 20% of type B (see table 3).

6 CONCLUSIONS AND FUTURE WORK

In this paper we introduced innovative three-layered semantic similarity measures between queries using a domain-specific taxonomy. We evaluated our measures by conducting an on-line survey and comparing them and other four existing semantic similarity measures against the participants' intuition. The results show that our similarity measures have a higher correlation with the average survey ranking than the other four measures. We believe that measuring semantic similarity between concepts using taxonomies can improve significantly the results retrieved by recommender systems. We also argue that these measures depend on the structure of the underlying taxonomy (hierarchy, keyword-topic mappings, keyword weights, etc.) and on the chosen concept-to-concept similarity measure. In the future, we plan to analyze the aspects that alter the behavior of the semantic similarity measures.

In this context, we distinguish two types of recommendations. The first type can be directly obtained by using the semantic similarity measure and retrieving the queries with the highest similarity to the user's last query. These recommendations will be rather "general" and maybe "too similar" to the last query (i.e. predictions with low probability). The second type of recommendations requires a much elaborate analysis (extracting patterns, clustering) of all users' history and then comparing the learned query patterns to the current user's history. With this type of recommendations we can predict the user's next set of questions (with a high probability) and, on the long run, his interests and goals. In the future we intend to focus on the second type of recommendations. We also plan to test the goodness of the semantic recommendations by analyzing users' feedback.

ACKNOWLEDGEMENTS

The authors would like to thank MEOVI[2] for the financial support during their research that lead to the findings presented in this paper.

REFERENCES

Al-Mubaid, H. and Nguyen, H. A. (2006). A cluster-based approach for semantic similarity in the biomedical domain. In *Proceedings of the 28th IEEE EMBS*

[2]www.meovi.com

Annual International Conference, pages 2713–2717, New York, USA.

Bernstein, A., Kaufmann, E., Bürki, C., and Klein, M. (2005). How similar is it? Towards personalized similarity measures in ontologies. In *7. Internationale Tagung Wirtschaftsinformatik*, pages 1347–1366.

Bin, S., Liying, F., Jianzhuo, Y., Pu, W., and Zhongcheng, Z. (2009). Ontology-based measure of semantic similarity between concepts. In *World Congress on Software Engineering*, volume 2, pages 109–112.

Bouquet, P., Kuper, G., Scoz, M., and Zanobini, S. (2004). Asking and answering semantic queries. In *Proceedings of Meaning Coordination and Negotiation Workshop (MCNW-04) in conjunction with International Semantic Web Conference*.

Burgess, C., Livesay, K., and Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, 25(2-3):211–257.

Cordì, V., Lombardi, P., Martelli, M., and Mascardi, V. (2005). An ontology-based similarity between sets of concepts. In *6th Joint Workshop "From Objects to Agents": Simulation and Formal Analysis of Complex Systems*, pages 16–21, Camerino,Italy.

Dong, H., Hussain, F. H., and Chang, E. (2009). A hybrid concept similarity measure model for ontology environment. In *Proceedings of the Confederated International Workshops and Posters on the Move to Meaningful Internet Systems*, pages 848–857.

Eiter, T. and Mannila, H. (1997). Distance measures for point sets and their computation. *Journal Acta Informatica*, 34:103–133.

Haase, P., Siebes, R., and Harmelen, F. V. (2004). Peer selection in peer-to-peer networks with semantic topologies. In *International Conference on Semantics of a Networked World: Semantics for Grid Databases*.

HaCohen-Kerner, Y., , Gross, Z., and Masa, A. (2005). Automatic extraction and learning of keyphrases from scientific articles. In Gelbukh, A., editor, *Computational Linguistics and Intelligent Text Processing*, volume 3406 of *Lecture Notes in Computer Science*, pages 657–669. Springer Berlin / Heidelberg.

Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 216–223, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jiang, J. and Conrath, W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference Research on Computational Linguistics*, pages 19–33, Taiwan.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998a). Introduction to latent semantic analysis. *Discourse Processes*, 25(2-3):259–284.

Landauer, T. K., Laham, D., and Foltz, P. (1998b). Learning human-like knowledge by singular value decomposition: A progress report. In *Advances in Neural Information Processing Systems 10*, pages 45–51. MIT Press.

Leacock, C. and Chodorow, M. (1998). *Combining Local Context and WordNet Similarity for Word Sense Identification*, pages 305–332. In C. Fellbaum (Ed.), MIT Press.

Lee, J. H., Kim, M. H., and Lee", Y. J. (1993). Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*, 49(2):188–207.

Lee, W. N., Shah, N., Sundlass, K., and Musen, M. (2008). Comparison of ontology-based semantic-similarity measures. In *AMIA Annual Symposium Proceedings*, pages 384–388.

Li, Y., Bandar, Z. A., and McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4).

Li, Y., McLean, D., Bandar, Z. A., O'Shea, J. D., and Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1138–1150.

Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304.

Marcel, P. and Negre, E. (2011). A survey of query recommendation techniques for data warehouse exploration. *7èmes Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA)*, B-7.

Oliva, J., Serrano, J. I., del Castillo, M. D., and Iglesias, A. (2011). Sysmss: A syntax-based measure for short-text semantic similarity. *Data and Knowledge Engineering*, 70:390–405.

O'Shea, J., Bandar, Z., Crockett, K., and McLean, D. (2010). Benchmarking short text semantic similarity. *International Journal of Intelligent Information and Database Systems*, 4(2):103 – 120.

Rada, R., Mili, H., Bicknell, E., and Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(1):17–30.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of IJCAI-95*, pages 448–453, Montreal, Canada.

Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal or Artificial Intelligence Research*, 11:95–130.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. In *Information Processing and Management*, pages 513–523.

Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information Retrieval*, 2(4):303–336.

Wang, G. H., Wang, Y. D., and Guo, M. Z. (2006). An ontology-based method for similarity calculation of concepts in the semantic web. In *Proceedings of the 5th International Conference on Machine Learning and Cybernetics*, pages 1538–1542, Dalian, China.

Wu, Z. and Palmer, M. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138.