The Similarity Jury: Combining expert judgements on geographic concepts

Andrea Ballatore,[1] David C. Wilson,[2] and Michela Bertolotto[1]

[1] School of Computer Science and Informatics
University College Dublin, Ireland
{andrea.ballatore,michela.bertolotto}@ucd.ie
[2] Department of Software and Information Systems
University of North Carolina, USA
davils@uncc.edu

Abstract. A cognitively plausible measure of semantic similarity between geographic concepts is valuable across several areas, including geographic information retrieval, data mining, and ontology alignment. Semantic similarity measures are not intrinsically right or wrong, but obtain a certain degree of cognitive plausibility in the context of a given application. A similarity measure can therefore be seen as a domain expert summoned to judge the similarity of a pair of concepts according to her subjective set of beliefs, perceptions, hypotheses, and epistemic biases. Following this analogy, we first define the *similarity jury* as a panel of experts having to reach a decision on the semantic similarity of a set of geographic concepts. Second, we have conducted an evaluation of 8 WordNet-based semantic similarity measures on a subset of OpenStreetMap geographic concepts. This empirical evidence indicates that a jury tends to perform better than individual experts, but the best expert often outperforms the jury. In some cases, the jury obtains higher cognitive plausibility than its best expert.

Keywords: Lexical similarity; Semantic similarity; Geo-semantics; Expert judgement; WordNet

1 Introduction

Since 2005, the landscape of geo-information has been experiencing rapid and dramatic changes. The concurrent explosion of Web 2.0 and web mapping has resulted in a complex nexus of phenomena, including geo-crowdsourcing, location-based services, and collaborative mapping. Traditional expert-generated geographic information has witnessed the advent of *produsers*, i.e. users engaged in production as well as consumption of spatial data. This resurgence of interest for maps among non-experts online users has been defined Volunteered Geographic Information (VGI) [11]. OpenStreetMap (OSM), a user-generated world map, is a particularly representative instance of these trends.[3]

[3] http://www.openstreetmap.org (acc. August 4, 2012)

As diverse communities generate increasingly large geo-datasets, semantics play an essential role to ground the meaning of the spatial objects being defined. In his vision of a Semantic Geospatial Web, Egenhofer stressed that semantic geo-technologies would enable higher interoperability, integration, and effective information retrieval [7]. When dealing with multiple sources of data, common tasks are those of information integration, and ontology alignment. For example, it might be necessary to retrieve the objects representing mountains from two datasets, one labelling them *mountain*, and the other one *peak*. If not supervised by a human, this semantic mapping is very challenging, because of the intrinsic ambiguity and fuzziness of geographic terms.

To identify automatically similar concepts in different datasets or within the same dataset, measures of semantic similarity are needed. Research in semantic similarity has produced a wide variety of approaches, classifiable as knowledge-based (structural similarity is computed in expert-authored ontologies), corpus-based (similarity is extracted from statistical patterns in large text corpora), or hybrid (combining knowledge and corpus-based approaches) [20, 23]. In the area of Geographic Information Science (GIScience), similarity techniques have been tailored on specific formalisms [25, 24, 12].

Typically, geographic concepts in geospatial datasets are described by a short lexical definition. For example, on the OSM Wiki website, the concept of a wetland is described as an 'area subject to inundation by water, or where waterlogged ground may be present.'[4] These definitions are used by data consumers to interpret the meaning of a feature, and by contributors to create appropriate meta-data for new features. As the OSM semantic model does not specify fine-grained ontological aspects of the concepts, a suitable approach to compute the semantic similarity of two concepts relies exclusively on their lexical definition. Lexical semantic similarity is an active research area in natural language processing, and several approaches have been proposed [20, 19]. The lexical database WordNet has turned out to be a key resource to develop knowledge-based measures [8].

In general, a judgement on lexical semantic similarity is not simply right or wrong, but rather shows a certain cognitive plausibility, i.e. a correlation with general human behaviour. For this reason, selecting the most appropriate measure for our domain is not trivial, and represents in itself a challenging task. A semantic similarity measure bears resemblance with a human expert being summoned to give her opinion on a complex semantic problem. When facing critical choices in domains such as medicine and economic policy, experts often disagree [17].

Instead of identifying the supposedly 'best' expert in a domain, a possibility is to rely on a jury of experts, extracting a representative average from their diverging opinions [3]. In this study we apply this strategy to the problem of lexical similarity for the domain of OSM geographic concepts, restricting the scope to a set of general-purpose WordNet-based measures. Rather than developing a new

[4] http://wiki.openstreetmap.org/wiki/Wetland (acc. August 4, 2012)

measure for geo-semantic similarity, we aim at exploring the idea of combining existing ones into a *similarity jury*.

The remainder of this paper is organised as follows. Section 2 reviews relevant related work in the areas of lexical semantic similarity, and WordNet-based similarity measures. The similarity jury is outlined in Section 3, while Section 4 presents and discuss an empirical evaluation. Finally, Section 5 draws conclusions about the jury, and indicates directions for future work.

2 Related work

The ability to assess similarity between concepts is considered a central characteristic of human beings [25]. Hence, it should not come as a surprise that semantic similarity is widely discussed in areas as diverse as philosophy, psychology, artificial intelligence, linguistics, and cognitive science.

Geographic information science is no exception, and over the past 10 years a scientific literature on similarity for geospatial concepts has been generated [13]. Schwering surveyed and classified semantic similarity techniques for geographic concepts, including network-based, set-theoretical, and geometric approaches [25]. Notably, Rodríguez and Egenhofer have developed the Matching-Distance Similarity Measure (MDSM) by extending Tversky's set-theoretical similarity for geographic concepts [24]. In the area of Semantic Web, SIM-DL is a semantic similarity measure for spatial concepts expressed in description logic (DL) [12].

WordNet is a well-known resource for natural language processing [8]. The usage of WordNet in the context of semantic similarity has fostered the development of numerous knowledge-based approaches, exploiting its deep taxonomic structure for nouns and verbs [15, 23, 16, 26, 1]. Table 1 summarises popular WordNet-based measures [2]. Some measures rely on shortest path between concepts, some include the information content of concepts, and others rely on the WordNet *glosses*, i.e. definition of concepts.

Whilst geospatial measures such as MDSM and SIM-DL can compute context-sensitive similarity in specific ontological formalisms, they cannot be applied directly to the OSM semantic model, in which geo-concepts are loosely described by natural language definitions. By contrast, general-purpose WordNet-based measures are easily applicable to the OSM concept lexical definitions. Spatial-geometric properties of the features – area, shape, topological relations, etc – have a role at the *instance* level, but are beyond the scope of this study, which focuses on abstract geographic *classes*. To the best of our knowledge, WordNet-based measures have not been applied to the geographic domain and, given their high cognitive plausibility in other domains, are worth exploring.

In this paper, we identify an analogy between computable semantic similarity measures and the combination of expert judgements, a problem relevant to several areas. Indeed, expert disagreement is not an exceptional state of affairs, but rather the norm in human activities characterised by uncertainty, complexity, and trade-offs between multiple criteria [17]. As Mumpower and Stewart put it,

Name	Authors	Description	SPath	Gloss	InfoC
path	Rada et al. [21]	Edge count in the semantic network	$\surd$		
lch	Leacock and Chodorow [15]	Edge count scaled by depth	$\surd$		
res	Resnik [23]	Information content of *lcs*	$\surd$		$\surd$
jcn	Jiang and Conrath [14]	Information content of *lcs* and terms	$\surd$		$\surd$
lin	Lin [16]	Ratio of information content of *lcs* and terms	$\surd$		$\surd$
wup	Wu and Palmer [26]	Edge count between *lcs* and terms	$\surd$		
lesk	Banerjee and Pedersen [1]	Extended gloss overlap		$\surd$	
vector	Patwardhan and Pedersen [19]	Second order co-occurrence vectors		$\surd$	

Table 1. WordNet-based similarity measures. *SPath*: shortest path; *Gloss*: lexical definitions (glosses); *InfoC*: information content; *lcs*: least common subsumer.

the "character and fallibilities of the human judgement process itself lead to persistent disagreements even among competent, honest, and disinterested experts" [18, p. 191].

Because of the high uncertainty in complex systems, experts often disagree on risk assessment, infrastructure management, and policy analysis [17, 5]. Mathematical and behavioural models have been devised to elicit judgements from experts for risk analysis, suggesting that simple mathematical methods perform quite well [4]. From a psychological perspective, in cases of high uncertainty and risk (e.g. choosing medical treatments, long term investments, etc), decision makers consult multiple experts, and try to obtain a representative average of divergent expert judgements [3].

To date, we are not aware of studies that address the possibility of combining lexical similarity measures in the context of geographic concepts. This corpus of diverse research areas informs our approach to addressing the problem.

3 The similarity jury

A computable measure of semantic similarity can be seen as a human domain expert summoned to rank pairs of concepts, according to her subjective set of beliefs, perceptions, hypotheses, and epistemic biases. When the performance of an expert can be compared against a gold standard, it is a reasonable policy to trust the expert showing the best performance. Unfortunately, such gold standards are difficult to construct and validate, and the choice of most appropriate expert remains highly problematic in many contexts.

To overcome this issue, we propose the analogy of the *similarity jury*, seen as a panel of experts having to reach a decision about a complex case, i.e. ranking the semantic similarity of a set of concepts. In this jury, experts are not human beings, but computable measures of similarity. Formally, the similarity function sim quantifies the semantic similarity of a pair of geographic concepts c_a and c_b ($sim(c_a, c_b) \in [0, 1]$). Set P contains all concept pairs whose similarity needs to be assessed, while set S contains all the existing semantic similarity measures.

Function sim enables the ranking of a set P of concept pairs, from the most similar (e.g. *mountain* and *peak*) to the least similar (*mountain* and *wetland*). These rankings $rank_{sim}(P)$ are used to assess the cognitive plausibility of sim against the human-generated ranks $rank_{hum}(P)$. The cognitive plausibility of sim is therefore the Spearman's correlation $\rho \in [-1, 1]$ between $rank_{hum}(P)$ and $rank_{sim}(P)$. If ρ is close to 1 or -1, sim is highly plausible, while if ρ is close to 0, sim shows no correlation with human behaviour.

A similarity jury J is defined as a set of lexical similarity measures $J = \{sim_1, sim_2 \ \ldots \ sim_n\}$, where all $sim \in S$. For example, considering the 8 measures in Table 1, jury a has 2 members ($J_a = \{jcn, lesk\}$), while jury b has 3 members ($J_b = \{jcn, res, wup\}$).

Several techniques have been discussed to aggregate rankings, using either unsupervised or supervised methods [4]. However, Clemen and Winkler stated that simple mathematical methods, such as the average, tend to perform quite well to combine expert judgements in risk assessment [4]. Thus for this initial exploration, we define the rankings of jury J as the *mean of the rankings* computed by each of its individual measures $sim \in J$. For example, if three measures rank five concept pairs as $\{1, 2, 3, 4, 5\}$, $\{2, 1, 4, 3, 5\}$ and $\{1, 2, 5, 3, 4\}$, the means are $\{1.3, 1.7, 4, 3.3, 4.7\}$, resulting in the new ranking $\{1, 2, 4, 3, 5\}$.

Furthermore, we define ρ_{sim} as the correlation of an individual measure sim (i.e. a jury of size 1), and ρ_J the correlation of the judgement obtained from a jury J. If $\forall sim \in J : \rho_J > \rho_{sim}$, the jury has *succeeded* in giving a more cognitively plausible similarity judgement. On the other hand, when $\exists sim \in J : \rho_J < \rho_{sim}$, the jury has *failed*, being less plausible than its constituent measure sim. A jury J enjoys a *partial success* against sim if $\rho_J > \rho_{sim}$, where $sim \in J$. Similarly, a jury obtains a *total success* if it outperforms all of its members, $\forall sim \in J : \rho_J > \rho_{sim}$.

4 Evaluation

In this section we evaluate the similarity jury, by comparing 154 juries with 8 individual measures, through an experiment on lexical similarity on OSM concepts.

Experiment setup. In order to study the similarity jury, we selected an existing dataset of human-generated similarity rankings on 54 pairs of geographic concepts, collected by Rodríguez and Egenhofer from 72 human subjects [24]. This dataset represents a high-quality sample of human judgements on geospatial similarity, covering large natural entities (e.g. *mountain*, *forest*) and man-made

	$\|J\|$	jcn	lch	lesk	lin	path	res	vector	wup	mean
		Jury containing *sim* (%)								
Partial success $\rho_J > \rho_{sim}$	2	69.3	62.9	84.6	55.0	60.4	79.6	55.0	66.4	66.6
	3	80.1	68.5	84.4	60.5	58.8	86.5	61.8	72.6	71.6
	4	84.4	73.1	83.7	60.4	61.9	87.2	65.6	73.9	73.8
	all	81.3	70.4	84.0	59.8	60.7	86.2	63.2	72.7	72.3
Total success $\forall sim \in J : \rho_J > \rho_{sim}$	2	46.1	42.5	35.7	43.9	42.1	34.6	35.0	42.1	40.2
	3	43.9	37.3	34.9	40.4	31.0	32.0	33.1	36.4	36.1
	4	39.7	32.7	33.9	35.0	28.6	29.5	30.9	33.1	32.9
	all	41.8	35.3	34.4	37.8	30.9	30.9	32.1	35.2	34.8
Plausibility	ρ	.72	.68	.45	.56	.66	.69	.56	.64	.62

Table 2. Results of the evaluation of the lexical similarity on 154 juries. For example, juries of cardinality 2 containing *jcn* obtain a partial success in the 69.3% of the cases.

features (e.g. *bridge*, *house*). The concepts of the human-generated dataset were manually mapped onto the corresponding concepts in the OSM, based on their lexical definitions.

To explore the performance of a similarity jury versus individual measures, we have selected a set of 8 *sim* term-to-term WordNet-based measures, $S = \{jcn, lch, lesk, lin, path, res, vector, wup\}$ (see Table 1). The open source project *WordNet::Similarity*[5] implements all of these measures, and was used to compute the similarity scores [20]. As the focus in this study is on the comparison of short segments of text, rather than individual words, the word similarity scores are combined using the technique developed by Corley and Mihalcea [6]. Since the OSM Wiki website holds about 1,900 concept definitions, the complete, symmetric similarity matrix for OSM concepts would contain about 1.8 million rankings.

In the context of risk assessment, large panels with more than 5 experts do not seem to outperform smaller ones [9]. Therefore, we consider the range of jury sizes $|J| \in [2, 4]$ to be appropriate for this study. All the subsets of S of cardinality two, three, and four were computed, resulting respectively in 28, 56, and 70 juries, for a total of 154 juries. The experiment was carried out through the following steps:

1. Compute $rank_{sim}(P)$ for the 8 measures on the OSM definitions.
2. Combine the individual measures into 154 jury $rank_J(P)$ by averaging the $rank_{sim}(P)$ of their members.
3. Compute cognitive plausibility against human-generated rankings for the 8 individual measures (ρ_{sim}) and the 154 juries (ρ_J).
4. Compute partial and total success ratio for juries containing a given *sim*.

Experiment results. Table 2 summarises the results of this experiment, showing the success and total success ratio of the juries containing a given *sim*, and

[5] http://wn-similarity.sourceforge.net (acc. August 4, 2012)

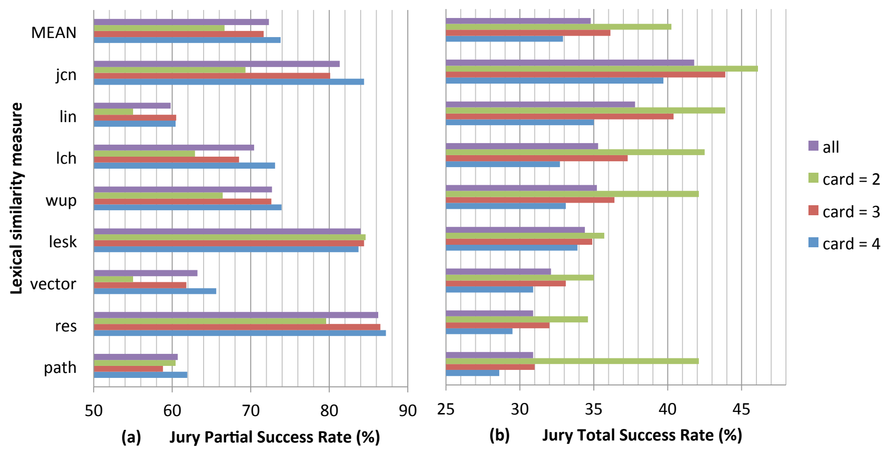

Fig. 1. Results of the lexical jury experiment: (a) partial success of the jury versus an individual measure; (b) total success of the jury versus all its member measures. *MEAN*: mean of success rates; *card*: cardinality of jury J.

the total success for each measure. The table shows the cognitive plausibility ρ for each measure *sim*, computed against the human rankings. It is possible to note that measures *jcn*, *res*, and *lch* have the highest cognitive plausibility. The jury results are grouped by jury cardinality (2, 3, and 4), and overall results (*all*). The results of the experiment are also displayed in Figure 1, which shows the success ratio of the juries grouped by their cardinality. For example, 80.1% of all juries of cardinality 3 containing the measure *jcn* are better than *jcn* in isolation. These results show a clear pattern: most juries enjoy a partial success over a given *sim* ($> 59.8\%$), while a minority of the juries obtain total success on all of their members ($< 41.8\%$). It is interesting to note that, in the experimental results, the plausibility of a jury is never inferior to that of all of its members, $\exists sim \in J : \rho_J < \rho_{sim}$.

The jury size has a clear impact on the success rate. Small juries of cardinality 2 tend to have a lower partial success ($mean = 66.6\%$), than those with 3 and 4 members (respectively 71.6% and 73.8%). Therefore larger juries have higher chances to obtain partial success over an individual measure. On the other hand, an opposite trend can be observed in the total success of a jury over all of its member measures. Juries of cardinality 2 tend to have a higher total success rate ($mean = 40.2\%$), compared with larger juries ($mean = 36.1\%$ for cardinality 3, and 32.9% for cardinality 4). As larger juries include more measures, it is more likely that one member outperforms the jury.

This empirical evidence shows that in 93.2% of the cases, the jury performs better than the average of the cognitive plausibility of its members, which would be by definition always lower than the plausibility of the best member: if the jury were simply returning the mean plausibility, its total success rate would always be 0%. By averaging the rankings, the jury reduces the weight of individual bias, converging towards a shared judgement. Such shared judgement is not necessarily the best fit in absolute terms, but tends to be more reliable than most individual judgements.

Given that we are measuring the cognitive plausibility of these similarity measures by the correlation with human rankings, the relationship between ρ of *sim* and the jury success ratio needs to be discussed. Interestingly, the cognitive plausibility ρ_{sim} shows no correlation with the jury partial and total success ratios (Spearman's $\rho \approx .1$). This suggests that even measures with high plausibility (such as *jcn* and *res*) still benefit from being combined with other measures. For example, the most plausible measure is *jcn* ($\rho = .72$), so it would be reasonable to expect a low success ratio, given that the measure is the most qualified expert in the panel. This expectation is not met: *jcn* shows a high partial and total success ratio (respectively 81.3% and 41.8%). The juries not only outperform individual measures in most cases, but can also obtain higher cognitive plausibility than its best member.

5 Conclusions and future work

In this paper we have proposed the analogy of the *similarity jury*, a combination of semantic similarity measures. The idea of jury was then evaluated in the context of lexical similarity for OSM geographic concepts, using 8 WordNet-based semantic similarity measures. Based on empirical results, the following conclusions can be drawn:

- In the context of the lexical similarity of geographic concepts, a similarity jury J is generally more cognitively plausible than its individual measures *sim* (partial success ratio mean $\approx$ 72%).
- A jury J is generally less cognitively plausible than the best of its members, i.e. $max(\rho_{sim}) > \rho_J$ (total success ratio mean $\approx$ 35%).
- In a context of limited information in which the optimal measure *sim* is not known, it is reasonable to rely on a jury J rather than on an arbitrary measure. The jury often outperforms even the most plausible measures.
- The similarity jury is consistent with the fact that, as Cooke and Goossens pointed out, "a group of experts tends to perform better than the average solitary expert, but the best individual in the group often outperforms the group as a whole" [5, p. 644].

In this initial study we have investigated the general behaviour of the similarity jury, by combining term-to-term WordNet-based similarity measures *sim*, in the context of geographic concepts of OSM. Our findings are consistent with those in the area of expert judgement combination for risk assessment [4, 5]. This indicates that the analogy of the jury is sound in the context of semantic similarity measures. However, in order to generalise these results, more work would be needed.

We have adopted a simple technique to combine rankings, i.e. a simple mean. More sophisticated techniques to combine rankings could be explored [22]. Furthermore, the empirical evidence presented in this paper was collected in a specific context, i.e. the lexical similarity of the geographic concepts defined in OSM. General-purpose similarity datasets, such as that by Finkelstein et al. [10], could be used to conduct experiments across other semantic domains.

The importance of semantic similarity measures in information retrieval, natural language processing, and data mining can hardly be underestimated [25]. A scientific contribution can be given not only by devising new similarity measures, but also by identifying effective ways to combine existing measures. In this sense, we believe that the similarity jury represents a promising direction worth investigating further, given its potential to enhance the cognitive plausibility of computational measures of semantic similarity.

Acknowledgements. The research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

References

[1] Banerjee, S., Pedersen, T.: An adapted Lesk algorithm for word sense disambiguation using WordNet. Computational Linguistics and Intelligent Text Processing pp. 117–171 (2002)
[2] Budanitsky, A., Hirst, G.: Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics 32(1), 13–47 (2006)
[3] Budescu, D., Rantilla, A.: Confidence in aggregation of expert opinions. Acta Psychologica 104(3), 371–398 (2000)
[4] Clemen, R., Winkler, R.: Combining probability distributions from experts in risk analysis. Risk Analysis 19(2), 187–203 (1999)
[5] Cooke, R., Goossens, L.: Expert judgement elicitation for risk assessments of critical infrastructures. Journal of Risk Research 7(6), 643–656 (2004)
[6] Corley, C., Mihalcea, R.: Measuring the semantic similarity of texts. In: Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment. pp. 13–18. Association for Computational Linguistics (2005)
[7] Egenhofer, M.: Toward the Semantic Geospatial Web. In: Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems. pp. 1–4. ACM (2002)
[8] Fellbaum, C.: WordNet: An electronic lexical database. The MIT press (1998)
[9] Ferrell, W.: Combining individual judgments. Behavioral decision making pp. 111–145 (1985)
[10] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E.: Placing Search in Context: The Concept Revisited. ACM Transactions on Information Systems 20(1), 116–131 (2002)

[11] Goodchild, M.: Citizens as Sensors: the world of volunteered geography. GeoJournal 69(4), 211–221 (2007)
[12] Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., Bäumer, B.: Algorithm, implementation and application of the SIM-DL similarity server. In: GeoSpatial Semantics: Second International Conference, GeoS 2007. Lecture Notes in Computer Science, vol. 4853, pp. 128–145. Springer (2007)
[13] Janowicz, K., Raubal, M., Kuhn, W.: The semantics of similarity in geographic information retrieval. Journal of Spatial Information Science (2), 29–57 (2011)
[14] Jiang, J., Conrath, D.: Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In: Proceedings of International Conference on Research in Computational Linguistics (ROCLING X). vol. 1, pp. 19–33 (1997)
[15] Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. WordNet: An electronic lexical database 49(2), 265–283 (1998)
[16] Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning. vol. 1, pp. 296–304. San Francisco (1998)
[17] Morgan, M., Henrion, M.: Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press (1992)
[18] Mumpower, J., Stewart, T.: Expert judgement and expert disagreement. Thinking & Reasoning 2(2-3), 191–212 (1996)
[19] Patwardhan, S., Pedersen, T.: Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In: Proceedings of the EACL 2006 Workshop 'Making Sense of Sense' – Bringing Computational Linguistics and Psycholinguistics Together. vol. 1501, pp. 1–8 (2006)
[20] Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet::Similarity: measuring the relatedness of concepts. In: Demonstration Papers at HLT-NAACL 2004 on XX. pp. 38–41. Association for Computational Linguistics (2004)
[21] Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics 19(1), 17–30 (1989)
[22] Renda, M., Straccia, U.: Web metasearch: rank vs. score based rank aggregation methods. In: Proceedings of the 2003 ACM symposium on Applied computing. pp. 841–846. ACM (2003)
[23] Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence. vol. 1, pp. 448–453. Morgan Kaufmann (1995)
[24] Rodríguez, M., Egenhofer, M.: Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure. International Journal of Geographical Information Science 18(3), 229–256 (2004)
[25] Schwering, A.: Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. Transactions in GIS 12(1), 5–29 (2008)
[26] Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on Association for Computational Linguistics. pp. 133–138. Association for Computational Linguistics (1994)