Concept-Analyzer: A tool for analyzing fuzzy concepts*

Sergio Guadarrama†
European Center for Soft Computing
sergio.guadarrama@softcomputing.es

Marta Garrido
Universidad Politécnica de Madrid
martagarridoarribas@hotmail.com

Abstract

In this work we want to test the feasibility of using ConceptNet to analyze concepts expressed by words, list of words or sentences. That is, we want to explore the capability of Concept-Net and the theory of similarity provided by fuzzy logic to compare concepts expressed by words, list of words and sentences in a conceptual and relational way (see [12], [8]).

Keywords: ConceptNet, Similarity, Fuzzy Concepts, Synonyms, Antonyms.

1 Introduction

Concept-Net [5] is a freely available commonsense [4] knowledgebase and natural-language-processing toolkit. Whereas similar large-scale semantic knowledgebases like Cyc [3] and WordNet [7] are carefully handcrafted, ConceptNet is generated automatically from the 700,000 sentences of the Open Mind Common Sense Project [10] a World Wide Web based collaboration with over 14,000 authors. ConceptNet is a structured resource as is WordNet but with a general scope as Cyc, although it is important to remark that they pursue different goals. While ConceptNet is devoted to extract commonsense knowledge from web users, WordNet is devoted to organize and categorize concepts by a group of experts, and Cyc is to devoted to collect assertions or facts related to general knowledge by a company.

ConceptNet is a semantic network of commonsense knowledge that at present contains 1.6 million edges connecting more than 300,000 nodes. Nodes are semi-structured English fragments, interrelated by an ontology of twenty semantic relations. A partial snapshot of actual knowledge in ConceptNet is given in figure 1. Although we talk about ontology, ConceptNet has not a formal ontology since it is not intended to be complete or totally sound, but fairly approximate. It includes a natural language analyzer MontyLingua (also developed by Hugo Liu at MIT) that we have used to parse sentence into sets of concepts.

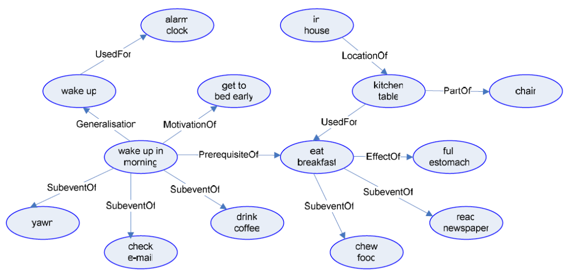


Figure 1: Extract of ConceptNet

ConceptNet's has twenty relation-types that are grouped by their thematic:

- K-Lines: 'ConceptuallyRelatedTo', 'ThematicKLine' and 'SuperThematicKLine'.
- Things: 'IsA', 'PropertyOf', 'PartOf', 'MadeOf' and 'DefinedAs'.

*This paper has been partially supported by MEC (Spain) under project TIN2005-08943-C02-01
† Corresponding author

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1084–1089
Torremolinos (Málaga), June 22–27, 2008

- Agents: 'CapableOf'.
- Events: 'PrerequisiteOf', 'First-SubeventOf', 'SubeventOf' and 'Last-SubeventOf'.
- Spatial: 'LocationOf'.
- Causal: 'EffectOf' and 'DesirousEffectOf'.
- Functional: 'UsedFor' and 'CapableOfReceivingAction'.
- Affective: 'MotivationOf' and 'DesireOf'.

2 Concept Analyzer

Concepts can be described by words, but there are many different ways of doing it. So we need a way to compare different ways and see how similar they are, taking into account that words are context dependent and therefore their choice and combination has influence in the results. In this work we want to be able to compare concepts from a relational point of view to establish a similarity relation. Using the idea that a word can be understood as an imprecise concept (and therefore, as a concept gradually related to other concepts) we have built "Concept-Analyzer". It is a software developed to manage the language interface and to analyze the relations between concepts and sentences. It allows us to take into account the context and the point of view in order to do a proper analysis.

Concept-Analyzer has two main objectives, first to help us to analyze the ideas expressed by words and sentences in a conceptual and relational way using for that purpose, the knowledge collected in Concept-Net and the theory of similarity and difference provided by fuzzy logic [11]. And, secondly, to use in the future this system to correlate user queries expressed by words or list of words and the answers returned by search engine, like Google or Yahoo (see [6], [2], [8]).

We have used Concept-Net due to the fact that it is the biggest and free available common sense knowledge base. It includes a set of tools to manage the net of concepts and a natural language syntactic processor. The inclusion of the context allows us to focus on a specific meaning of a word. Also the possibility of defining a point of view (i.e. general, affective, taxonomic, detailed) has allowed us to capture the concrete use of a word in the given context. And all of this has facilitated the comparison of concepts as one can see in figure 4.

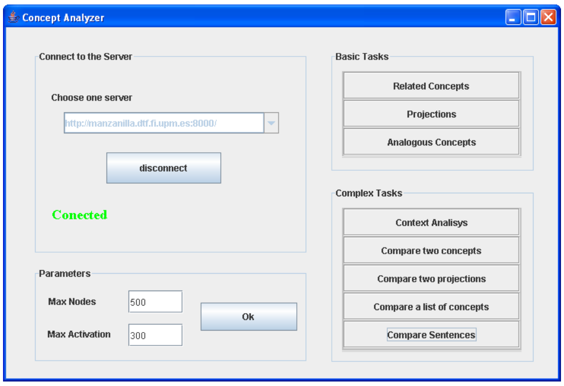


Figure 2: Concept Analyzer

3 Similarities

In order to compare different projections of concepts we have tested many different fuzzy similarities functions taken from [1] [9], which can be divided in two groups:

3.1 Set based

It uses fuzzy sets operations to calculate their similarity. For example, given an universe two fuzzy sets $A(x) = a_1/x_1 + \cdots + a_n/x_n$ and $B(x) = b_1/x_1 + \cdots + b_n/x_n$, and taking as fuzzy cardinality measure $|A| = \sum_{i=1}^{n} a_i$, we can define:

- M-Similarity as:

$$M_{A,B} = \frac{|A \cap B|}{|A \cup B|}$$

- T-Similarity as:

$$T_{A,B} = \sup_{x \in X}(A \cap B)$$

- $Dice$-Similarity as:

$$Dice_{A,B} = \frac{2|A \cap B|}{|A| + |B|}$$

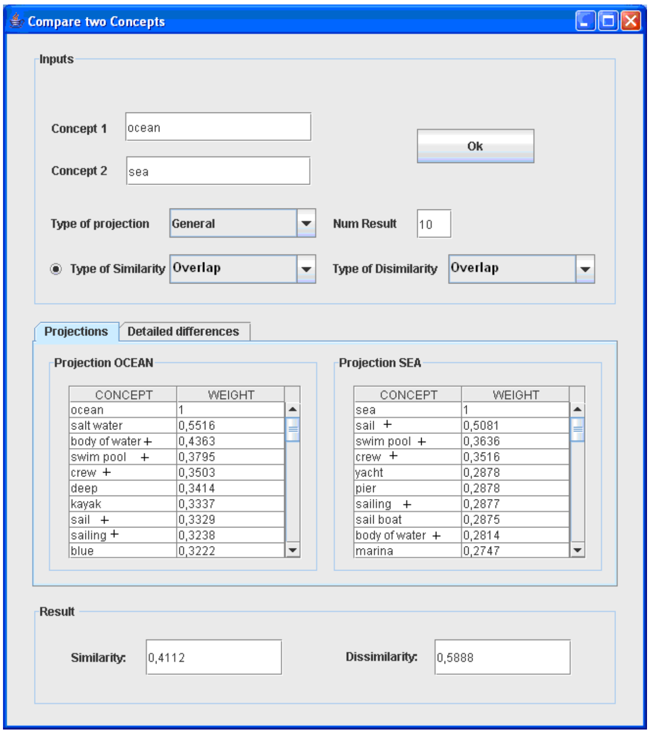


Figure 3: Compare Ocean and Sea

- *Cosine*-Similarity as:

$$Cosine_{A,B} = \frac{|A \cap B|}{\sqrt{|A|} + \sqrt{|B|}}$$

- *Overlap*-Similarity as:

$$Overlap_{A,B} = \frac{|A \cap B|}{min(|A|, |B|)}$$

- *Mutual*-Similarity as:

$$Mutual_{A,B} = \frac{\frac{|A \cap B|}{|A|} + \frac{|A \cap B|}{|B|}}{2}$$

3.2 Distance based

It assumes that fuzzy sets are points in a N-dimensional space and by using a distance function calculates their similarity. For example, given an universe $X = x_1, \ldots, x_n$ and two fuzzy sets $A(x) = a_1/x_1 + \cdots + a_n/x_n$ and $B(x) = b_1/x_1 + \cdots + b_n/x_n$, we can define:

- *L*-Similarity as:

$$L_{A,B} = 1 - max(|a_i - b_i|)$$

- *S*-Similarity as:

$$S_{A,B} = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{\sum_{i=1}^{n} (a_i + b_i)}$$

- *W*-Similarity as:

$$W_{A,B} = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{n}$$

- *Euclid*-Similarity as:

$$Euclid_{A,B} = 1 - \sqrt{\frac{\sum_{i=1}^{n} (a_i - b_i)^2}{n}}$$

4 Related concepts

Given a concept or list of concepts we can use ConceptNet to get a list of related concepts with their relatedness degree.

We have two different ways of obtaining these related concepts, one is by using the default point of view, that is, by using the general projection, and second is by specifying one point of view by means of a type of projection. We have defined 11 types of projections to test the validity of ConceptNet by changing the weights assigned to each type of relation:

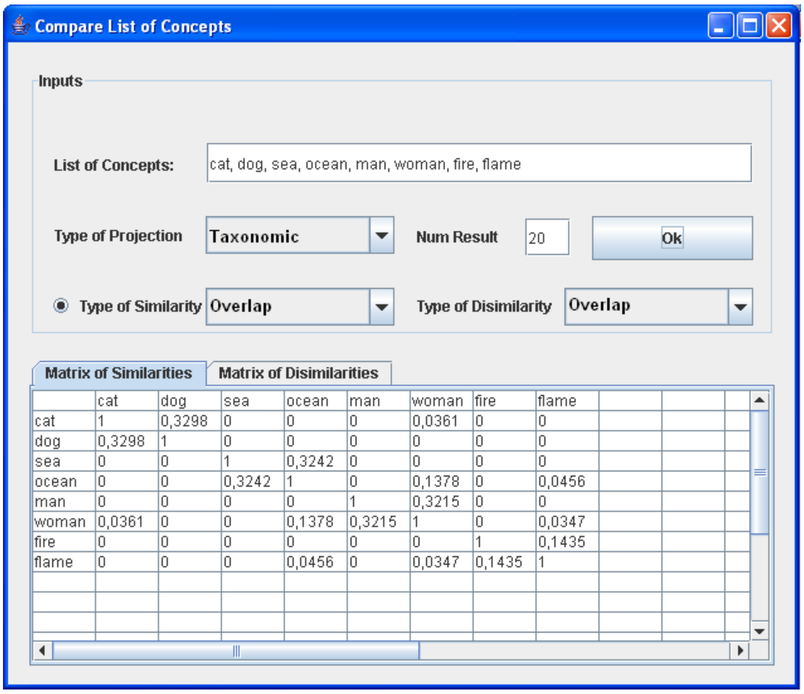


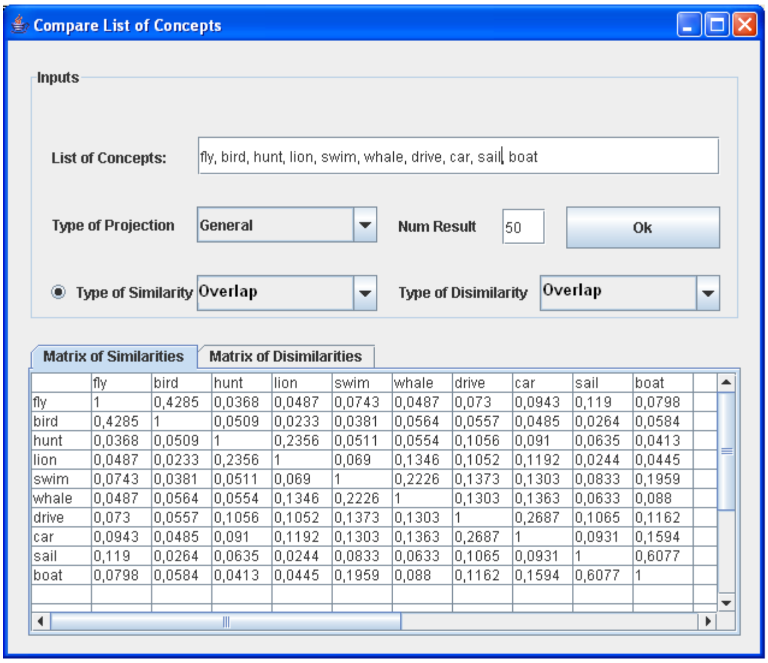
Figure 4: Similarities between entities

1. General: it assigns different weights to all direct relations trying to obtain a global view (for example see figure 3).
2. Details: it obtains concepts related with properties, attributes, parts, ...
3. Affective: it obtains concepts related with sentiments, sensations, desires, motivations, ...
4. Consequences: it obtains concepts related with verbs, actions and states.
5. Spatial: it obtains concepts related with places, situations, states, ...
6. Inferred: it obtains concepts which relation where inferred by ConceptNet.
7. Taxonomic: it obtains concepts related with properties, classes, attributes,... (for example see figure 4)
8. Abstraction: it obtains concepts related inverse of properties, classes, attributes,...
9. All ones: it assigns 1 to weights of all direct and inverse relations to obtain the maximum propagation.
10. Direct ones: it assigns 1 to weights of all direct relations to obtain the maximum forward propagation.
11. Inverse ones: it assigns 1 to weights of all inverse relations to obtain the maximum backward propagation.

4.1 Compare two concepts

Given two concepts we can use the lists of related concepts with their relatedness degree and different similarities functions to obtain their similarity and dissimilarity degrees. For example, if we compare 'ocean' and 'sea' we obtain a similarity degree of 0.41 using a general point of view (see figure 3) and a similarity degree of 0.32 using a taxonomic point of view (see figure 4).

Let us explain in a little detail how it works for this case, first we look for concepts related to 'ocean' and to 'sea' using a general point of view, then find the concepts that they have in common (marked with '+' in figure 3), and depending of the similarity chose we calculate the degree of similarity. In the case of the overlapping similarity we obtain the following



	fly	bird	hunt	lion	swim	whale	drive	car	sail	boat
fly	1	0,4285	0,0368	0,0487	0,0743	0,0487	0,073	0,0943	0,119	0,0798
bird	0,4285	1	0,0509	0,0233	0,0381	0,0564	0,0557	0,0485	0,0264	0,0584
hunt	0,0368	0,0509	1	0,2356	0,0511	0,0554	0,1056	0,091	0,0635	0,0413
lion	0,0487	0,0233	0,2356	1	0,069	0,1346	0,1052	0,1192	0,0244	0,0445
swim	0,0743	0,0381	0,0511	0,069	1	0,2226	0,1373	0,1303	0,0833	0,1959
whale	0,0487	0,0564	0,0554	0,1346	0,2226	1	0,1303	0,1363	0,0633	0,088
drive	0,073	0,0557	0,1056	0,1052	0,1373	0,1303	1	0,2687	0,1065	0,1162
car	0,0943	0,0485	0,091	0,1192	0,1303	0,1363	0,2687	1	0,0931	0,1594
sail	0,119	0,0264	0,0635	0,0244	0,0833	0,0633	0,1065	0,0931	1	0,6077
boat	0,0798	0,0584	0,0413	0,0445	0,1959	0,088	0,1162	0,1594	0,6077	1

Figure 5: Similarity between verbs and entities related

result:

$\mu_{ocean} = ocean/1 + saltwater/0.55 + \cdots$
With $|\mu_{ocean}| = 4.37$
$\mu_{sea} = sea/1 + sail/0.50 + \cdots$
With $|\mu_{sea}| = 3.92$
$\mu_{ocean} \wedge \mu_{sea} = swimpool/0.36 + crew/0.35 +$
$sail/0.33 + sailling/0.29 + bodyofwater/0.28$
With $|\mu_{ocean} \wedge \mu_{sea}| = 1,61$
$Over = \frac{|\mu_{ocean} \wedge \mu_{sea}|}{min(|\mu_{ocean}|,|\mu_{sea}|)} = \frac{1.61}{min(4.37,3.92)} = 0.41$

4.2 Compare a list of concepts

In this case from a list of concepts we obtain a matrix of similarities and of dissimilarities, that allows us to obtain a better comparison between concepts, and it is something that later on could be used for clustering. For example we introduce the list [‘cat’, ‘dog’, ‘sea’, ‘ocean’, ‘man’, ‘woman’, ‘fire’, ‘flame’] and chose a taxonomic point of view obtaining the following significant similarities (see figure 4):

- Between ’cat’ and ’dog’ is 0.33.
- Between ’sea’ and ’ocean’ is 0.32.
- Between ’man’ and ’woman’ is 0.32.
- Between ’fire’ and ’flame’ is 0.14.
- Between ’woman’ and ’ocean’ is 0.14

In general, these results are very coherent with our expectations with only the surprise that ’woman’ and ’ocean’ are related to some degree and the degree of similarity between ’fire’ and ’flame’ is a little bit low.

Finally to test the capabilities of this approach we have compared the similarity between verbs and entities related. That is, we wanted to see if we can discriminate between different actions when fixing some entities. We compare the list of entities [‘bird’, ‘lion’, ‘whale’, ‘car’, ‘boat’] with the list of actions [‘fly’, ‘hunt’, ‘swim’, ‘drive’, ‘sail’] using a general point of view and obtaining the following most related pairs with their associated degree: (fly, bird, 0.43), (hunt, lion, 0.23), (swim, whale, 0.22), (drive, car, 0.27) and (sail, boat, 0.61), as we expected (see figure 5). Although some degrees are not as high as I would like they are at least much higher that the others pairs, as for example (fly, car, 0.09), (drive, boat, 0.11) and (swing, car, 0.13), while there are some possible interesting pairs as (swing, boat, 0.19) or (car, boat, 0.16).

5 Conclusions

We have used Concept-Net due to the fact that it is the biggest and free available common sense knowledge base. It includes a set of tools to manage the net of concepts and a natural language syntactic processor.

We can say that the inclusion of words as context allows us to precise a specific meaning and obtain concepts more related and with greater degree of relatedness. It could allow us to discriminate among different meanings of polysemy words.

When obtaining concepts the variation the type of projections allow us to obtain different results for the same concept from different points of view. That is, the point of view have a great influence on determining the related concepts.

Regarding the variety of similarity functions tested in this work we can conclude that some of then have a coherent behavior, M-Similarity, P-Similarity, S-Similarity, $Dice$-Similarity, $Overlap$-Similarity, $Mutual$-Similarity and $Cosine$-Similarity but others have not a coherent behavior with our expectations, for example $Euclid$-Similarity, T-Similarity, L-Similarity or W-Similarity, but others have a Among all of them the best is the $Overlap$-Similarity.

References

[1] S.M. Chen, M.S. Yeh, and P.Y. Hsiao. A comparison of similarity measures of fuzzy values. *Fuzzy Sets and Systems*, 72(1):79–89, 1995.

[2] C.S. Lee, Z.W. Jian, and L.K. Huang. A fuzzy ontology and its application to news summarization. *Systems, Man and Cybernetics, Part B, IEEE Transactions on*, 35(5):859–880, 2005.

[3] D.B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[4] H. Liu and P. Singh. Commonsense reasoning in and over natural language. Springer, 2004.

[5] H. Liu and P. Singh. ConceptNetA Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*, 22(4):211–226, 2004.

[6] C. Loer, H. Singh, A. Cheung, S. Guadarrama, and M. Nikravesh. Evaluating Ontology Based Search Strategies. In *Soft Computing For Information Processing And Analysis*, volume 164 of *Studies in fuzziness and soft computing*, pages 189–202. Springer, 2005.

[7] G.A. Miller. WordNet: A Lexical Database for English. *COMMUNICATIONS OF THE ACM*, 38(11):39, 1995.

[8] J.A. Olivas, P.J. Garcés, and F.P. Romero. An application of the FIS-CRM model to the FISS metasearcher. *International Journal of Approximate Reasoning*, 34(2-3):201–219, 2003.

[9] C.P. Pappis and N.I. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56(2):171–174, 1993.

[10] P. Singh and et al. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proc. of the 1st Int. Conf. on Ontologies, Databases, and Applications of Semantics for Large Scale I.S*, volume 2519. Springer, 2002.

[11] E. Trillas, C. Moraga, S. Guadarrama, S. Cubillo, and E. Castiñeira. Computing with Antonyms. In *Forging New Frontiers: Fuzzy Pioneers I*, volume 217 of *Studies in fuzziness and soft computing*, pages 133–153. Springer, 2007.

[12] E. Trillas, E. Renedo, and S. Guadarrama. Fuzzy Sets vs Language. In B. Reusch, editor, *Computational Intelligence: Theory and Practice*, volume 164 of *Studies in fuzziness and soft computing*, pages 353–366. Springer, 2006.