International Journal of Medical Informatics (2005) 74, 125—132

International Journal of Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

A novel view on information content of concepts in a large ontology and a view on the structure and the quality of the ontology

Carl Van Buggenhout[a,*], Werner Ceusters[b,1]

[a] *Second Cycle in Electrical Engineering, Option Photonics at the University of Ghent, Belgium*
[b] *European Centre for Ontological Research, University of Saarland, Saarbrücken, Germany*

Received 31 October 2003; received in revised form 13 March 2004; accepted 17 March 2004

KEYWORDS
Semantic distance;
Information content

Summary Semantic distance and semantic similarity are two important information retrieval measures used in word sense disambiguation as well as for the assessment of how relevant concepts are with respect to the documents in which they are found. A variety of calculation methods have been proposed in the literature, whereby methods taking into account the information content of an individual concept outperform those that do not. In this paper, we present a novel recursive approach to calculate a concept's information content based on the information content of the concepts to which it relates. The method is applicable to extremely large ontologies containing several million concepts and relationships amongst them. It is shown that a concept's information content as calculated by this method provides additional information with respect to an ontology that cannot be approximated by hierarchical edge-counting or human insight. In addition, it is suggested that the method can be used for quality control within large ontologies and that it can give you an impression on the structure and the quality of the ontology.

1. Introduction

Semantic distance and semantic similarity are important, though loosely defined measures in concept-based information retrieval; definitions, if given at all, depending on the approach used to calculate these measures. They have been used with varying degrees of success in applications such as medical coding [1], semantic indexing [2], word sense disambiguation [3,4] and image caption retrieval [5] to mention only a few. Whereas semantic distance measures how closely two concepts are topologically related in a semantic network, semantic similarity captures to what extent two concepts might represent the same thing. Obviously, the two notions are closely related, although not the same.

* Corresponding author.
E-mail addresses: carl.vanbuggenhout@ugent.be (C.V. Buggenhout), werner.ceusters@ecor.uni-saarland.de (W. Ceusters).
URL: http://www.ecor.uni-saarland.de.
[1] Tel. +32 475 486 587.

1386-5056/$ — see front matter
doi:10.1016/j.ijmedinf.2004.03.009

A concept such as ''fractured arm'' should have a very short semantic distance towards ''arm fracture'', whereas the semantic similarity should be small: a ''fracture'' cannot stand for an ''arm'' or the other way round. But, to complicate matters, any decent system should be able to compute the semantic distance of post-coordinated concepts such as ''patient''-WITH-''arm fracture'' and ''patient''-WITH-''fractured arm'' as being minimal, and the semantic similarity as being maximal.

Various approaches to calculate both values have been proposed. They tend to fall in two categories. *Edge-based methods* exploit mainly the idea of path-length in a network with or without additional weights according to the type of link traversed, whereas *node-based methods* also take into account the probability to find each concept in a large corpus [6]. The idea behind them is that the ''information content'' (i.e. a loosely defined measure expressing how much information is conveyed in a word, phrase, sentence or entire text) of concepts occurring often in a corpus is lower than of concepts that occur rarely, and that these information-low concepts tend to appear higher in an ontology. Recently, a similar idea is introduced in methods that are intrinsically edge-based. It tends to capture the feeling that the semantic difference between upper level concepts in an ontology is bigger than between lower level concepts [7]. The implementation in [7] is entirely based on the hierarchical ISA-relationship.

In this paper, we expand this idea by taking also into account the associative relationships amongst concepts. We argue that, although semantic distance, semantic similarity and information content are very different notions, they are related to each other. We propose a method for information content calculation that preserves this relationship.

2. Material and methods

LinKBase® is a large scale medical ontology developed and maintained by the modeling team of Language and Computing nv. LinKBase® contains currently over one million language-independent medical and general-purpose concepts, linked to natural language terms in several languages, including English [8,9]. These concepts are linked together into a semantic network like structure using approximately 450 different link types for expressing formal relationships. These relationships are based on logics dealing with issues such as mereology and topology [10,11], time and causality [12] and models for semantics driven natural language understanding [13,14]. Link types form a multi-parented hierarchy on their own, such as the partitive hierarchy (from narrower to broader) ''is-tangential-part-of'' > ''is-proper-part-of'' > ''is-part-of''.

It is very important to note that in LinkBase® the formal subsumption relationship covers only about 15% of the total number of relationships amongst concepts. As such, LinkBase® is a much richer structure than terminological systems in which term-relationships are expressed as strictly ''narrower'' or ''broader''. Important to note also is that LinkBase® is a ''living'' ontology, in which data are changed on a daily basis, and at a rate of 2000—4000 modifications a day. Moreover, it is not required for concepts added to be perfectly modeled from the very beginning [15].

We defined the initial information content (IC_0) of a concept in LinkBase® as:

$$\forall k : IC_0(C_k) = 1 + \sum_i (LW(L_i) \times IC_0(C_i)) \qquad (1)$$

where C_k is the source concept, C_i the target concept, $LW(L_i)$ the link weight of the link between C_k and C_i, and the sum is going over all the outgoing concepts of C_k.

We defined the initial link weight (LW_0) of a link type as:

$$\forall k : LW_0(L_k) = 1 + \sum_i (LW_0(\mathrm{ParentLink}_i(L_k))) \qquad (2)$$

where the sum goes over all the parents of link type L_k.

From these formulas it follows that the IC of individual concepts can only be found by setting up and subsequently solving a (huge!) system of equations, i.e. one per concept in the ontology. Once all values computed, they were normalised using a straight line in the range of 0—1. We used the following function for the ICs:

$$IC := (C, a) \rightarrow \frac{IC_0(C)}{a} \quad \text{where } a = \max(IC_0(C_k)), \forall k \qquad (3)$$

and we used this function to normalize the LWs:

$$LW := (L, a) \rightarrow \frac{LW_0(L)}{a}$$
$$\text{where } a = \max(LW_0(L_k)), \forall k \qquad (4)$$

An algorithm was designed to compute the ICs in real time, taking advantage of the network-structure of the ontology.

A first analysis of the results was carried out by comparing the information content ranking of the algorithm with those of four human judges: all concepts containing the substring ''tachycard'' as part of their knowledge name were extracted, ranked alphabetically, and given to the ontology modelers (medical doctors with at least one year experience in ontology building). Their task was to rank the concepts according to their subjective information content, a notion which through their daily work is clear to them, but without explaining them how the algorithm works. The results of the evaluators were then compared with those of the algorithm. The null-hypothesis for this analysis is that there would be no difference between human judges as a group, and the automatic ranking based on IC.

A second analysis involved a comparison of the depth of a concept in the hierarchy with respect to its computed IC value. For the IC to fulfill its purpose, there should be no complete correlation.

3. Results

The program ran on an Intel Pentium 2.4 GHz processor and used approximately 600 MB of RAM memory. It took about 15 min to calculate the 450 link-type weights and the 1 million ICs, and also some 10 min to store all the concept values in a file, sorted on IC.

3.1. Evaluation of the ranking experiment

Fig. 1 captures the results of the ranking performed by the algorithm (diagonal line with diamonds) and the human evaluators (irregular line with squares, based on the overall mean of the different human ratings). The concept for instance that was ranked by the algorithm on the 43rd place, was set by the evaluators on average on position 100. The third, unmarked line, is the result of a trend analysis performed on the human ratings. From this it follows that the upward trend of the ranking produced by the algorithm is followed by the human raters, but at a considerable lower slope.

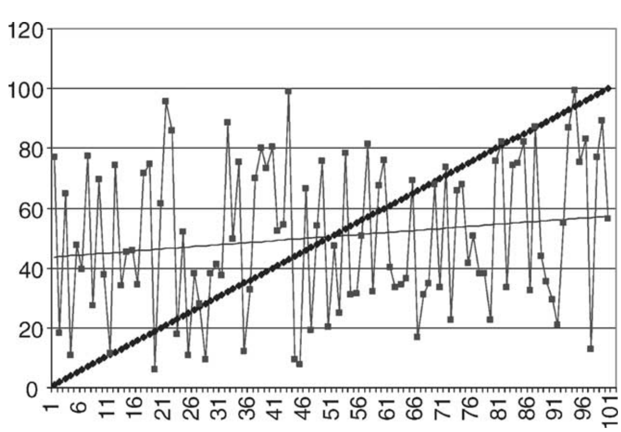

Fig. 1 IC-ranking by algorithm and human evaluators.

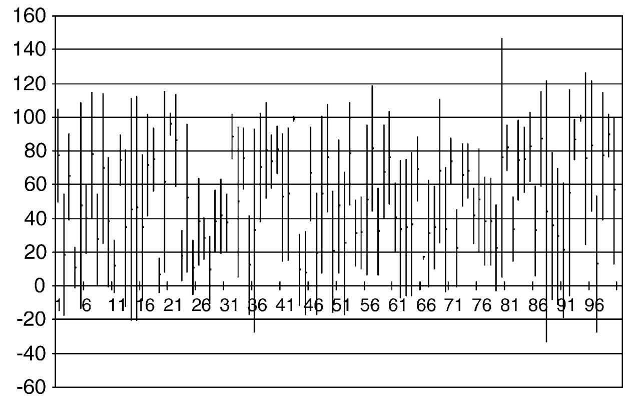

Fig. 2 Inter-rater agreement amongst human evaluators.

Fig. 2 shows the mean rankings of the human raters within a 2 standard deviation confidence interval. From this it follows that the human raters made quite different assessments of the information content. The evaluators together think that the concept that was placed by the algorithm on the fist position, must be placed somewhere between positions 50 and 100. We can also see that the concept placed by the algorithm on position 88, is placed by the evaluators somewhere between -30 and 120. This is clearly impossible, but it does show that the evaluators were not certain about the real position to put that concept.

3.2. Correlation between the hierarchical depth of a concept in the ontology and its IC

Fig. 3 shows a scatter plot of the normalized ICs of the ''tachycardia''-related concepts, versus the depth according to the ISA-hierarchy. Because the normalised ICs are very small, they were re-scaled using Formula (5), for pure visualization purposes.

$$\text{Rescaled IC} = \frac{-100}{\log(\text{IC})} \tag{5}$$

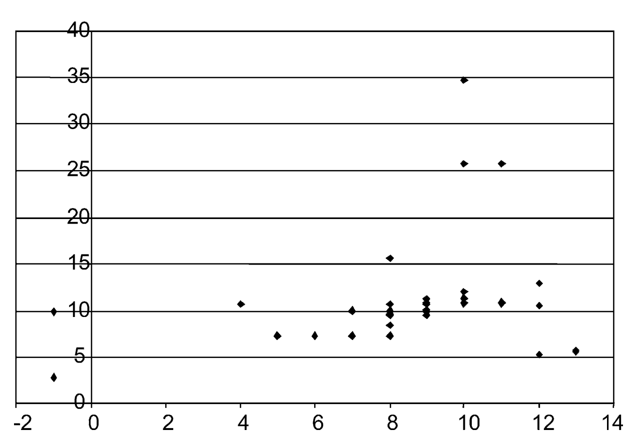

Fig. 3 IC/depth plot.

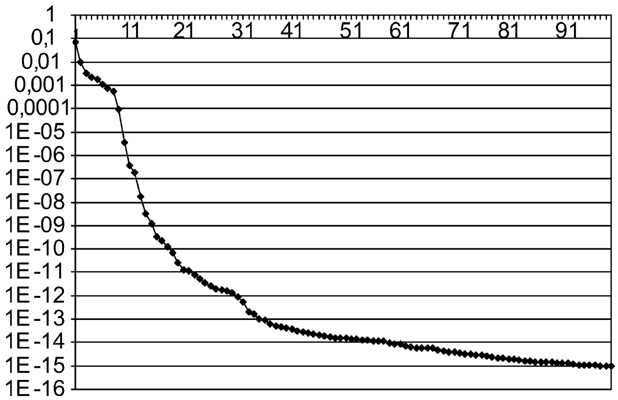

Fig. 4 IC distribution.

As can be seen, the hierarchical depth of a concept in the ontology is not the most important factor that contributes to the information content.

Fig. 4 gives an impression concerning the distribution of the normalized ICs on a logarithmic scale. Only the first 100 concepts with the highest IC are shown.

4. Discussion

4.1. Algorithm design

To assess how much information a concept contains about other concepts, we had to set up a system of equations. This system was set up by assuming that the information content of a concept relates to the sum of the ICs of the concepts used to describe it, weighted by a link type specific value (Formula (1)). These link type specific values were calculated in exactly the same way, but without a weighting factor (Formula (2)).

A naïve approach to set up the system of equations would be to generate the equation for each concept individually and then to use a matrix formalism to calculate the IC for each concept. However, given the huge size of the ontology, this would not be sensible.

We developed a novel algorithm exploiting the fact that generation and partially solving the equations could be guided by the structure of the ontology.

There are two steps in the algorithm. In the first step LWs are calculated that are used in the second step to compute the ICs of the concepts.

We start the work with a list of all the link types. For every link type L_1 that we did not visit yet, we calculate its LW by following all the parent link types of this link type, and using their LWs to compute the value of L_1 according to Formula (2). If the LW of one of the outbound link types (say L_2) is not computed up till now, we calculate the value for link type L_2 by recursively applying the same procedure that we used for calculating the LW of L_1, that is, we follow all the parent link types of link type L_2, and use their LWs to compute the value of L_2, until we reach a link type L_i that has no parents. The value for this link type L_i is 1. After we calculated all the LWs, we normalize them using Formula (4).

We then work with a list of all the concepts. For every concept C_1 that we did not visit yet, we calculate its IC by following all the outbound links of this concept, and using the concept values of these outbound concepts to compute the value of C_1 according to Formula (1). If the concept value of one of the outbound concepts (say C_2) is not computed up till now, we calculate the value for concept C_2 by recursively applying the same procedure as for calculating the concept values of C_1, that is, we follow all the outbound links of concept C_2, and use their ICs to compute the value of C_2. This process comes to an end if we reach a concept C_i that has no parents. Its value equals 1. After we calculated all the ICs, we normalize them using Formula (3).

There is one condition that must be taken into account: the algorithm we used iterates until it reaches a value that has already been calculated, so this value has not to be computed for a second time. However when there are cycles — i.e. paths from a concept towards the same concept — in the ontology, we must stop the iteration process for instance after *n* iterations. We are sure that the value of such a concept converges to its ''real'' value if the system of equations is not false. In that case the more iterations the program does, the more precise these converging values get.

But what if the system is false? With values the six links as in Fig. 5, the system of equations is not solvable. This is because the determinant of the coefficient-matrix of the system equals zero. Therefore we cannot use algorithms that make use of the inverse of that matrix because the inverse does not exist. Such algorithms are for instance Gauss-Seidel (GS) or SOR (successive over relaxation) [16]. Both algorithms assume that the

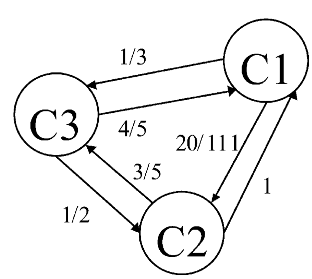

Fig. 5 Unsolvable configuration.

coefficient-matrix is diagonally. The more diagonally the matrix, the faster the solution converges. But they fail on situations as in Fig. 5. Our modified algorithm, as explained in the section below, handles this situation by calculating the best approximation. This is not a mathematical approximation because one cannot make an approximation of the ICs because they will diverge. Instead it will be an approximation of the ICs meaning that the values obtained by our algorithm will be sufficient to declare differences between ICs as being significant enough. This is a result of the structure of the modified algorithm when only some iterations are being taken in the calculation of the IC values. This effect will be further examined in the next paragraph.

4.2. Modification of the algorithm

The algorithm as it is implemented by us differs from the algorithm described above. This has some very important consequences. First this means that we always obtain positive IC-values, but on the other hand, these values are no real solutions of the system of linear equations. However the values do reflect the amount of information they contain about other concepts. These ICs tell us to what extent the concepts are modeled. This means that a concept with a high value is more modeled and has more properties than a concept with a lower value. The following example will clarify all this.

Say we have got an ontology with five concepts C_1—C_5 linked together with their corresponding weights like this (Fig. 6):

If we want to calculate the IC-values for the concepts C_1—C_5, we take the following steps:

Step 1. We write down the system of equations as explained by Formula (1). In this case the system turns out to be (with the shorter notation C_i for $IC_0(C_i)$):

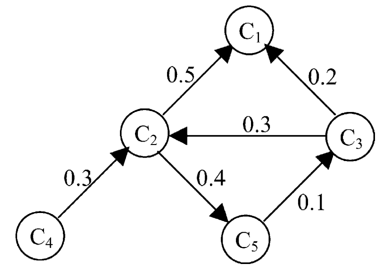

Fig. 6 Example ontology.

$C_1 = 1$

$C_2 = 1 + 0.5 \times C_1 + 0.4 \times C_5$

$C_3 = 1 + 0.2 \times C_1 + 0.3 \times C_2$

$C_4 = 1 + 0.3 \times C_2$

$C_5 = 1 + 0.1 \times C_3$

Step 2. We initiate all the values to 1: $(C_1, \ldots, C_5) = (1, 1, 1, 1, 1)$.

Step 3. For each concept C_i we visit all the outbound concepts of that concept C_i that we did not visit yet from concept C_i. This means that a concept can be traversed several times, but only once by traversing a link from the concept C_i.

Say we start by visiting concept C_2.

The table below shows the progress of the algorithm. In the right column the instant values of the concepts are shown. The bold values represent values that will not be changed any longer.

$C_2 = 1 + 0.5 \times C_1 + 0.4 \times C_5 \rightarrow$ visit C_1	$(C_1, \ldots, C_5) = (1, 1, 1, 1, 1)$
$C_1 = 1$	$(C_1, \ldots, C_5) = (\mathbf{1}, 1, 1, 1, 1)$
$C_2 = 1 + 0.5 \times 1 + 0.4 \times C_5 \rightarrow$ visit C_5	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1, 1, 1)$
$C_5 = 1 + 0.1 \times C_3 \rightarrow$ visit C_3	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1, 1, 1)$
$C_3 = 1 + 0.2 \times C_1 + 0.3 \times C_2 \rightarrow$ visit C_1	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1, 1, 1)$
$C_1 = 1$	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1, 1, 1)$
$C_3 = 1 + 0.2 \times 1 + 0.3 \times C_2 \rightarrow$ visit C_2	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1.2, 1, 1)$
$C_2 = 1 + 0.5 \times 1$	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, 1.2, 1, 1)$
$C_3 = 1 + 0.2 \times 1 + 0.3 \times 1.5$	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, \mathbf{1.65}, 1, 1)$
$C_5 = 1 + 0.1 \times 1.65$	$(C_1, \ldots, C_5) = (\mathbf{1}, 1.5, \mathbf{1.65}, 1, \mathbf{1.165})$
$C_2 = 1 + 0.5 \times 1 + 0.4 \times 1.165$	$(C_1, \ldots, C_5) = (\mathbf{1}, \mathbf{1.966}, \mathbf{1.65}, 1, \mathbf{1.165})$
$C_4 = 1 + 0.3 \times 1.966$	$(C_1, \ldots, C_5) = (\mathbf{1}, \mathbf{1.966}, \mathbf{1.65}, \mathbf{1.5898}, \mathbf{1.165})$

The real mathematical solution of this system equals:

$$\{C_1 = 1, \quad C_2 = 1.97, \quad C_3 = 1.791,$$
$$C_4 = 1.592, \quad C_5 = 1.179\}$$

When using this algorithm we calculate values based on the structure of the ontology and based on the links between the different concepts. As a result the values obtained by this calculation are good results for the amount of information that a concept contains about all the other concepts in the ontology. There will be a small difference in the resulting values when we start calculating for instance with C_5 instead of starting with concpet C_2 but this difference will not be significant for our purpose.

4.3. Structure of the ontology

When applying the algorithm to LinkBase, we can see there is a trend that very much concepts in the ontology have relatively low IC-values. This can also be seen in Fig. 4. This may be unexpected, because the IC-value of a concept is a sum for all the outbound concepts of the concept whose IC we are calculating. An important question is to what extent this result reflects the structure of the ontology we are investigating.

Hence let us take a closer look at the example ontology used to describe the modified algorithm (Fig. 6).

First we can see that concept C_1 has no parents. Its IC-value will not depend on the values of C_2 until C_5. Second we notice that C_4 is connected to the ontology only by a link towards concept C_2. For this reason the IC-value of C_4 will be influenced by the value of C_2. The values of concepts C_2, C_3 and C_5 will influence each other.

This makes it possible to divide the ontology into three parts as follows.

The circle in the middle in Fig. 7 represents an ontology in which three concepts are linked together, but where no concepts appear that have only outbound links like concept C_4 and where no concepts are present that have only inbound links like concept C_1. All the concepts in this sub-ontology have at least one outbound and at least one inbound link that connect the concepts by paths in the ontology. It is allowed for concept C_1 to have one or more outbound links towards other concepts, but these links may not connect C_1 by a path to the sub-ontology $\{C_2, C_3, C_5\}$.

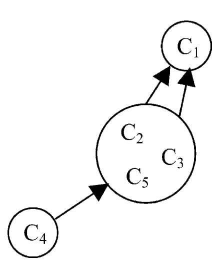

Fig. 7 Structure of the ontology shown in Fig. 6.

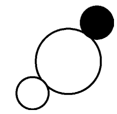

Fig. 8 Abstraction of Fig. 7.

Fig. 8 shows a shorter notation for an ontology like the one in Fig. 7. The little white circle is a concept that has no influence on the sub-ontology in the middle, while the value of the black concept is determinant for the mean IC-value of the sub-ontology. Because all the concepts are linked together the variation of the IC-values of the concepts in the sub-ontology to the mean value will be small. This mean value will depend on the value of the black concept. It will also depend on the values of the links in the sub-ontology but even more on the link between the sub-ontology and the black concept. If for instance the value of the black concept equals 10E30 and the range of the link type weights is between 10E-4 and 1, then the mean IC-value of the sub-ontology can be somewhere in the interval 10E30 $\times$ [10E-4, 1]. The mean value of the sub-ontology can be larger than 10E30, but the probability is small because in that case the values of the link type weights in the sub-ontology must equal almost 1; otherwise the values will decrease because of the normalized link type weights.

One way to obtain high values in a sub-ontology is to have many black concepts the sub-ontology is linked towards as shown in Fig. 9(a). Another way is that many sub-ontologies are linked together as in Fig. 9(b). The effect in Fig. 9(b) is that the mean values of the sub-ontologies will slightly increase, while the mean value of the sub-ontology in Fig. 9(a) can be much higher than the values of the black concepts.

Another important result is that two sub-ontologies like those in Fig. 10 are independent of each other. Their mean values will not be influenced by each other. This is a consequence of the fact that the black concepts do not have outbound links at all.

Now we are ready to make use of these sub-structures to determine the global structure of the ontology by use of the distribution of the calculated

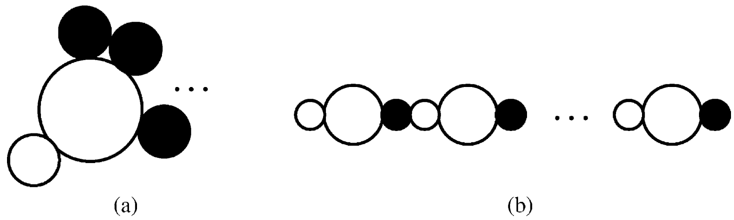

Fig. 9 Some possible sub-ontologies.

IC-values of all the concepts. We are not able to determine two sub-structures that are independent as a consequence of the fact that their values will not influence each other, but the difference made in Fig. 9 is subject of our interest. These different structures can be determined by investigating the calculated IC-values.

Fig. 11 shows a possible representation of the analysis for LinkBase. There are many concepts that do not have parents, which makes their values equal 1 (the black concepts on the right). The other part of LinkBase consists of sets of sub-structures merged together like in Fig. 9(b), where all the concepts are well modeled and linked together. This makes LinkBase a rich structure containing a lot of information.

4.4. Interpretation of the results

It was no surprise to us that the modelers' rankings differed considerably amongst each other, as well as with respect to the algorithm's ranking. Differences amongst modelers could be explained most often by inaccurate estimations of the IC of additional criteria associated to the concept that appeared to be more central. The IC of the concept ''chronic tachycardia'' was by all modelers correctly judged lower than the IC of ''fetal tachycardia'', which on its turn was judged lower than for ''fetal tachycardia affecting management of mother''. But typically, the IC-differences for ''chronic'', ''fetal'', and ''affecting management of mother'' were seriously underestimated. When informed about these differences, some modelers accepted this view without critique, while others judged the differences as real but irrelevant, a situation similar as in [17] where only 2 out of 19 hierarchic relationships generated by a description logic classifier (hence mathematically correct) were judged ''accurate'' enough by human reviewers to be taken into account.

Fig. 10 Two independent sub-ontologies.

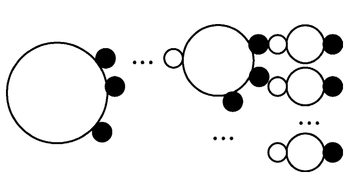

Fig. 11 A possible representation of the structure of LinKBase.

It was also no surprise that the IC of a concept is relatively (though not complete of course) independent from its place in the hierarchy. For a given hierarchical depth, the range of ICs is typically large. This is a result of the fact that the number of outbound links of concepts for a given depth is not constant. There can be concepts on a given depth that have many outgoing links and as result have higher IC values than other concepts on the same hierarchical depth that have less outbound links. This explains why the IC of a concept not only depends on the place of the concept in the hierarchy. However, when standard statistical techniques for outlier detection were used, the majority of outliers turned out to be the result of inappropriate modeling. As such, this method might be useful for quality control.

5. Conclusion

We have been able to design a novel algorithm to calculate the information content of concepts in extremely large ontologies. The method adds another dimension to the notions of semantic distance and semantic similarity as the calculated ICs are relatively independent from the hierarchical depth within an ontology. Because information content has been shown to be an important parameter for accurate information retrieval [5,7], our method might give an important contribution in that field. In addition, we have indications that the method can also be used for quality control. The distribution of the IC-values can also give us an impression on the quality of the ontology, i.e. how the ontology is structured and how well the ontology is modeled.

References

[1] Department of Health, Information Strategy and Development Documentation. Available at: C. Bousquet, M. Jaulent, G. Chatellier, P. Degoulet, Using Semantic Distance for the Efficient Coding of Medical Concepts, in: Proceedings of the AMIA 2000. http://medicine.ucsd.edu/F2000/D200437.htm.

[2] B. Jackson, W. Ceusters, A novel approach to semantic indexing combining ontology-based semantic weights and in-document concept co-occurrences, in: R. Baud, P. Ruch (Eds.), Proceedings of the EFMI Workshop on Natural Language Processing in Biomedical Applications, Cyprus, March 8—9, 2002, pp. 75—80.

[3] M. Sussna, word sense disambiguation for free-text indexing using a massive semantic network, in: Proceedings of the Second International Conference on Information and Knowledge Management, Arlington, VA, USA, 1993, pp. 67—74.

[4] E. Agirre, G. Rigau, Word sense disambiguation using conceptual density, in: Proceedings of the 16th International Conference on Computational Linguistics, Copenhagen, 1996, pp. 16—22.

[5] A.F. Smeaton, G. Quigley, Experiments on using semantic distances between words in image caption retrieval, in: Proceedings of the 19th International Conference on Research and Development in Information Retrieval (SIGIR96), Zurich, Switzerland, 1996, pp. 174—180.

[6] G. Polčicová, P. Návrat, Semantic similarity in content-based filtering, in: Y. Manolopoulos, P. Návrat (Eds.), Proceedings of the ADBI2002 Advances in Databases and Information Systems, Springer LNCS 2435, 2002, pp. 80—85.

[7] J. Zhong, H. Zhu, J. Li, Y. Yu, Conceptual graph matching for semantic search, in: U. Priss, D. Corbett, G. Angelova (Eds.), Conceptual Structures: Integration and Interfaces (ICCS2002), 2002, pp. 92—106.

[8] W. Ceusters, P. Martens, C. Dhaen, B. Terzic, LinkFactory: an advanced formal ontology management system, in: Proceedings of the Interactive Tools for Knowledge Capture Workshop, Victoria, BC, Canada, KCAP-2001, October 20, 2001 (http://sern.ucalgary.ca/ksi/K-CAP/K-CAP2001/).

[9] F. Montyne, The importance of formal ontologies: a case study in occupational health, in: Proceedings of the OES-SEO2001 International Workshop on Open Enterprise Solutions: Systems, Experiences, and Organizations, Rome, September 14—15, 2001 (http://cersi.luiss.it/oesseo2001/papers/28.pdf).

[10] B. Smith, Mereotopology: a theory of parts and boundaries, Data Knowledge Eng. 20 (1996) 287—301.

[11] B. Smith, A.C. Varzi, Fiat and bona fide boundaries, in: Proceedings of the COSIT-97, Springer-Verlag, 1997, pp. 103—119.

[12] F. Buekens, W. Ceusters, G. De Moor, The explanatory role of events in causal and temporal reasoning in medicine, Met. Inform. Med. 32 (1993) 274—278.

[13] W. Ceusters, F. Buekens, G. De Moor, A. Waagmeester, The distinction between linguistic and conceptual semantics in medical terminology and its implications for NLP-based knowledge acquisition, Met. Inform. Med. 37 (4/5) (1998) 327—333.

[14] J.A. Bateman, Ontology construction and natural language, in: Proceedings of the International Workshop on Formal Ontology, Padua, Italy, 1993, pp. 83—93.

[15] A. Flett, M. Casella dos Santos, W. Ceusters, Some ontology engineering processes and their supporting technologies, in: A. Gomez-Perez, V.R. Benjamins (Eds.), Ontologies and the Semantic Web, EKAW2002, Springer 2002, pp. 154—165.

[16] W. Hackbusch, Iterative solution of large sparse systems of equations, in: Applied Mathematical Sciences, Springer-Verlag, New York, USA, 1993.

[17] N.R. Hardiker, A.L. Rector, Structural validation of nursing terminologies, J. Am. Med. Inform. Assoc. 8 (3) (2001) 212—221.