International Journal of Semantic Computing

BUILDING SEMANTIC NETWORKS FROM PLAIN TEXT AND WIKIPEDIA WITH APPLICATION TO SEMANTIC RELATEDNESS AND NOUN COMPOUND PARAPHRASING

PIA-RAMONA WOJTINNEK and STEPHEN PULMAN
Department of Computer Science, Oxford University
Wolfson Building, Parks Road, Oxford, OX1 3QD, United Kingdom
{*pia-ramona.wojtinnek,stephen.pulman*}*@cs.ox.ac.uk*
http://www.cs.ox.ac.uk/activities/compling

JOHANNA VÖLKER
KR&KM Research Group, University of Mannheim
B6 26, 68159 Mannheim, Germany
voelker@informatik.uni-mannheim.de

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

The construction of suitable and scalable representations of semantic knowledge is a core challenge in Semantic Computing. Manually created resources such as WordNet have been shown to be useful for many AI and NLP tasks, but they are inherently restricted in their coverage and scalability. In addition, they have been challenged by simple distributional models on very large corpora, questioning the advantage of structured knowledge representations.

We present a framework for building large-scale semantic networks automatically from plain text and Wikipedia articles using only linguistic analysis tools. Our constructed resources cover up to 2 million concepts and were built in less than 6 days. Using the task of measuring semantic relatedness, we show that we achieve results comparable to the best WordNet based methods as well as the best distributional methods while using a corpus of a size several magnitudes smaller. In addition, we show that we can outperform both types of methods by combining the results of our two network variants. Initial experiments on noun compound paraphrasing show similar results, underlining the quality as well as the flexibility of our constructed resources.

Keywords: Semantic Networks; Semantic Relatedness; Unsupervised Noun Compound Paraphrasing; Wikipedia

1. Introduction

Finding suitable representations of knowledge for Semantic Computing has been a challenge since its early beginnings. Early designs of semantic networks such as those in [1] have led to today widely used lexical knowledge resources such as WordNet [2] or logic-based knowledge representation frameworks known as ontolo-

gies. However, while the constructed resources have been proven to be useful for many AI and Natural Language Processing (NLP) tasks, their major drawback is their costly manual acquisition. WordNet, for example, has been successfully used for measuring semantic relatedness, disambiguating word senses and recognising textual entailment. However, building the resource has been an ongoing project for more than 10 years. It currently covers 117,000 concepts, linked by a limited set of relations such as hyponymy and meronymy. In comparison, the encyclopedia Wikipedia has about 3.5 million entries, indicating the vast number of concepts a general knowledge representation should cover. WordNet, and with it any method based on it, is restricted to some core general purpose concepts and cannot scale to cover specific domains. Furthermore, some WordNet based approaches to NLP tasks have been challenged or even outperformed by simple distributional models on very large corpora, posing the question whether structured representations of semantic knowledge are necessary.

In this article, we build large-scale semantic networks directly from plain text as well as from Wikipedia articles (i.e. text with Wikipedia markup) using only state-of-the-art linguistic analysis tools. Our work presents a further development of the semantic networks introduced by [3]. Our largest plain text network covers approximately 870,000 concepts while the Wikipedia based network contains more than 2 million concepts. The underlying corpus size is comparable for both networks. However, due to a more fine-grained distinction between concepts in the Wikipedia network, it has broader coverage, is less dense and takes less time to build. Including all linguistic preprocessing, constructing these two networks took about 6 and 4.5 days, respectively. By tackling the semantic relatedness task, we show that our automatically created resources achieve results comparable to the best WordNet based methods. In addition, our results are comparable to the best distributional models while our corpus is several magnitudes smaller. This shows that by using a network representation of the concepts, relations and attributes occurring in a corpus, we gain more information from the underlying text than models that use the text directly. Furthermore, by combining the results of the plain text and the Wikipedia networks, we outperform both the best WordNet based methods and the best distributional methods. This gives us a powerful framework to rapidly enhance the performance of our general-purpose networks with domain specific text as necessary for the application at hand. Initial experiments on noun compound paraphrasing show similar results, underlining the quality of our constructed networks as well as their applicability to a variety of tasks.

2. Building Semantic Networks

We build our large-scale semantic networks directly from text using state-of-the-art linguistic analysis tools. The network provides a structured representation of the concepts, relations and attributes occurring in the text and can be thought of as an approximation of a knowledge representation covering these concepts. Our frame-

work can build two slightly different types of networks - networks from plain text and networks from Wikipedia text. The plain text networks have the advantage of availability for any domain while the Wikipedia based ones are more precise and can more easily be integrated with other existing knowledge resources. The basic building process is similar for both networks. They are built by translating every sentence in the text into a network fragment based on semantic analysis and then merging these networks into a large network by mapping all occurrences of the same concept onto one node. Figure 1 contains a sample text snippet and the network derived from it. In the example, concepts such as STUDENT and DISSERTATION and relations such as *write* are identified and syntactic surface structures are translated to their semantic content. Multiple occurrences of these concepts in different sentences are integrated. In this way, concepts are connected across sentences and documents, resulting in a high-level view of the information contained.

2.1. *Plain Text Networks*

Our first semantic network variant is built from plain text and has been previously described in [4]. The concepts are derived from the occurring nouns and the resulting network represents those concepts and their extracted relationships. The core advantage of this network is that the type of corpus it is based on (i.e. plain text) is available for any domain and usually in large quantities. It therefore presents an easy and quick way to approximate a domain-specific knowledge representation. The network is built incrementally by parsing every sentence, translating it into a small network fragment and then mapping that fragment onto the main network generated from all previous sentences. Our translation of sentences from text to network is based on ASKNet [3]. It makes use of two NLP tools, the Clark and Curran parser [5] and the semantic analysis tool Boxer [6], both of which are part of the

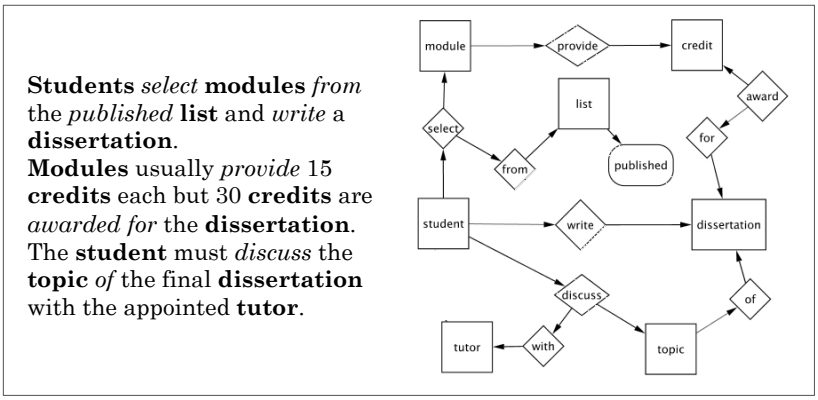

Fig. 1: Sample text snippet and according network representation. For explanation of node shapes see Section 2.1

C&C Toolkit[a]. The parser uses Combinatory Categorial Grammar (CCG) and has been trained on 40,000 manually annotated sentences of the Wall Street Journal. It is both efficient and robust. Boxer is then designed to convert the CCG parsed text into a logical representation based on Discourse Representation Theory (DRT). This intermediate logical form representation provides an abstraction from different syntactic surface forms to their semantic core information. For example, the syntactical forms *progress of student* and *student's progress* have the same Boxer representation and so do *the student who attends the lecture* and *the student attending the lecture*. In addition, Boxer provides some elementary co-reference resolution.

The translation from the Boxer output into a network is straightforward and an example is given in Figure 2. The network structure distinguishes between concept nodes (rectangular), relational nodes (diamonds) and attributes (rounded rectangles) and different types of links such as subject, object and attribute links. As the different link types currently have no impact on our application, we omit further elaboration here. Details can be found in [7]. The large unified network is then built by merging every occurrence of a concept into one node, thus accumulating the information on this concept. In this example, the *lecture* node would be merged with occurrences of *lecture* in other sentences. Figure 3 gives a subset of a network based on a few paragraphs from Oxford student handbooks. Multiple occurrences of the same path between two object nodes are drawn as overlapping.

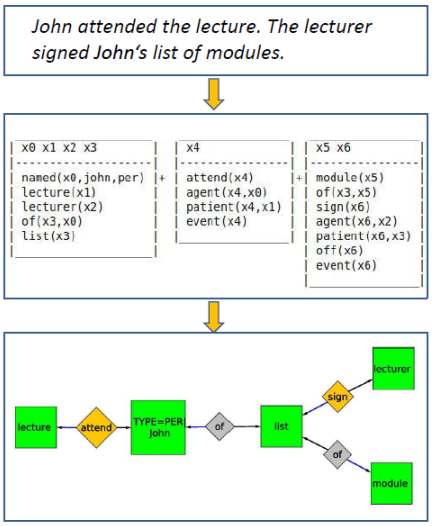

Fig. 2: Example of translation from text to network over Boxer semantic analysis.

[a]http://svn.ask.it.usyd.edu.au/trac/candc

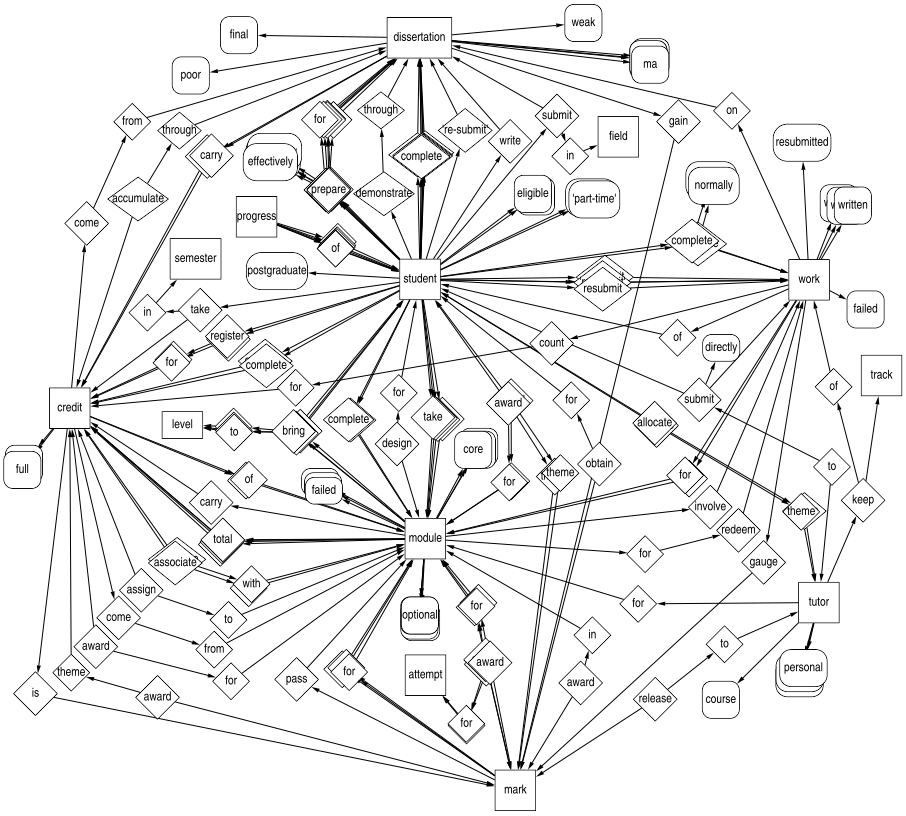

Fig. 3: Subgraph displaying selected concepts (rectangular nodes), relations (diamond nodes) and attributes (rounded rectangular nodes) from sample network. The figure is taken from [4].

2.2. *Wikipedia Networks*

Our second network variant is entirely constructed from Wikipedia text and makes use of the available hyperlinks to uniquely identify concepts. It is based on the observation that we can think of an article as referring to a concept. Then a hyperlink to this article in Wikipedia text unambiguously provides information on the corresponding concept. For example, in the first sentence of the "Roberts bank" article *'Roberts Bank' is an [[undersea bank‖bank_(geography)]] on the south side of the estuary of the [[Fraser River]]*, "undersea bank" is linked to the article BANK_(GEOGRAPHY), which disambiguates the word "bank" and establishes a relation between the two concepts ROBERTS BANK and BANK_(GEOGRAPHY). The network fragment of this sentence would contain three concept nodes ROBERTS BANK, BANK_(GEOGRAPHY) and FRASER RIVER, and relations extracted using the linguistic tools as before. The concept nodes in the Wikipedia network are ex-

actly those that have been identified with an article and we will call them *article concept nodes* for clarity.

The Wikipedia-based networks have several advantages over plain text networks. Concepts are disambiguated and the correspondence with Wikipedia articles allows for integration of the network resource with other Wikipedia-based resources such as DBpedia[b]. In addition, as the text is encyclopedic, the extracted relations are more characteristic of the concept they apply to. However, Wikipedia is a limited resource and despite its broad coverage of currently 3.6 million articles, not all domains will be equally well covered.

The network is built in four steps shown in Figure 4[c]. The Corpus Extractor and Preprocessor component extracts the relevant text and produces a plain text version as well as a version preserving the hyperlinks to other articles. Depending on the objective, the extracted text can be first sentences, first paragraphs or whole articles and it can either take into account all Wikipedia articles or only a subset for a specific domain. Each plain text sentence is then translated into a network fragment using ASKNet in the same way as for the plain text network. In the following Wikipedia Tag component, the hyperlink annotations are integrated into the network fragment. This often means that two or more original nodes turn into one newly created article node, such as in the case of "academic degree". Article concept nodes are identified by their article name, but the original plain text tokens are also kept. For example, the article concept node ACADEMIC DEGREE will keep track of the token "academic degrees" and in the sample sentence above, the node BANK_(GEOGRAPHY) will have the token "undersea bank". Finally, the merging component incrementally integrates all sentence network fragments to form one unified semantic network by merging all article concept nodes with the same article name tag into one. In this way, the OXFORD UNIVERSITY article node from the first sentence is merged with the OXFORD UNIVERSITY article node in the second sentence, while both tokens "Oxford University" and "University of Oxford" are kept.

2.3. *Constructed Network Resources and their Statistics*

In this paper, we built general-purpose, broad-coverage networks of both variants. As the corpus for the plain text networks, we chose the British National Corpus (BNC)[d]. It is one of the largest standardised English corpora and contains approximately 5.9 million sentences. The diversity of the corpus ensures good coverage of concepts as well as realistic overall connectedness. For the Wikipedia-based network, we extracted the first sentence of each of the 3.6m articles in our Wikipedia snapshot of 16 Jan 2010. Not all Wikipedia articles start with a full first sentence,

[b] `http://dbpedia.org/`

[c] "Oxford University Computing Laboratory" renamed to "Department of Computer Science" in June 2011.

[d] http://www.natcorp.ox.ac.uk/

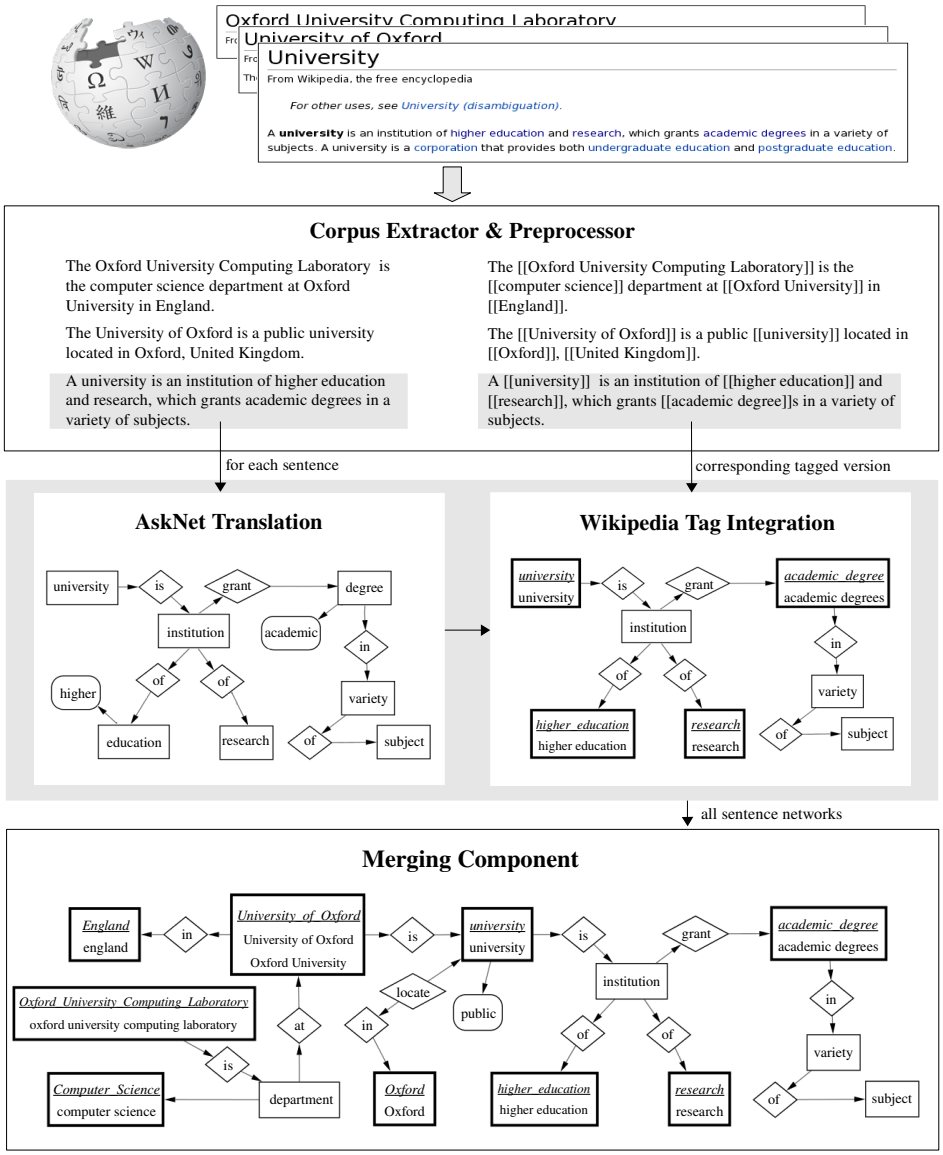

Fig. 4: Illustration of framework for building Semantic Wikipedia Networks.

in particular, many disambiguation and list pages do not. The resulting corpus contains just under 3m sentences (see Table 1). The first sentences provide a good base for the initial resource presented in this paper, because they tend to contain more crucial information and more links than sentences in later paragraphs of the article. However, we can easily extend the network to include all sentences in Wikipedia that contain a link or the primary entity. In order to be able to compare the two types of networks, we build the plain network on a 3m sentence subcorpus of the

Table 1: Corpus sizes for (a) BNC (b) Wikipedia.

BNC Corpus Statistics			
	1m sentences	**2m sentences**	**3m sentences**
Number of sentences	1,140,712	2,016,681	3,221,798
Successfully parsed	1,095,067	1,937,587	3,084,758
	(96.0%)	(96.1%)	(95.7%)

Wikipedia Corpus Statistics	
Total number of Wikipedia pages	3,650,225
(of which disambiguation or list pages)	(174,798)
Extracted first sentences	2,946,409
After filtering	2,823,195
Successfully parsed sentences (C&C + Boxer)	**2,775,369** (98.3%)

BNC and in addition networks from 1m and 2m sentences to demonstrate the effect of corpus size on the quality of representation as well as network structure and building time.

2.3.1. *Node Statistics and Building Times*

The resulting networks contain a total of about 15m to 43m nodes for the BNC networks and about 34m for the Wikipedia network (cf. Table 2). The BNC networks cover approximately 500,000 to 870,000 concepts while the Wikipedia network covers over 2 million. The primary reason for this difference is the detailed annotation provided by the Wikipedia links. In other words, the distinction into concepts is more fine-grained in the Wikipedia network. Concept nodes or article concept nodes only make up for a small fraction of the network. Most of the nodes are part of relations between the concepts and attributes. As to be expected, we can see that the fraction of concept nodes gets smaller as the network gets bigger. More concepts are repeated in the text and their occurrences merged. Relations between concepts are collected.

Table 2: Node statistics of all networks.

Node Statistics				
	BNC Networks			**Wikipedia**
	1m sent.	**2m sent.**	**3m sent.**	**network**
Total number of nodes	15,704,437	27,446,743	43,282,918	34,168,598
Concept/article concept nodes	498,920 (3.3%)	635,008 (2,3%)	869,309 (2,0%)	2,123,097 (6,8%)
Other nodes	15,205,517	26,811,735	42,413,609	32,045,501

The complete construction time including the linguistic preprocessing for the BNC networks was between just under 2 days for the 1 million sentences network and approximately 6 days for the 3 million network (cf. Table 3). The Wikipedia network took 4.5 days to construct. The largest chunk of construction time is taken up by the linguistic preprocessing including the syntactic and semantic parsing, accounting for 76%-91% of construction time. As the parsers take one sentence at a time, the linguistic processing time is linear in the number of sentences. However, this can easily be sped up by parallelisation. The average parsing time per Wikipedia sentence is slightly lower than the one for the BNC sentences. This is probably due to the first sentences of Wikipedia articles being less complex than sentences in the BNC and on average more than a word shorter (20.74 words per sentence in BNC, 19.44 in the Wikipedia corpus). This is reflected by the lower percentage of successfully parsed sentences in the BNC (95.93%) compared to Wikipedia (98.3%). The actual network building time lies between a bit more than 4 hours for the 1m BNC network and just under 35 hours for the 3m network and around 14 hours for the Wikipedia one. Table 3 also lists the running time excluding time spent in garbage collection. As we get closer to the memory capacity of the machine used, garbage collection is called more often. This distorts the runtime growth analysis. Using the garbage collection corrected times, we can clearly see the average building time per node grows linearly in the number of nodes and the total building time grows quadratically.

Table 3: Complete construction times of the networks. We present both the network building time with and without garbage collection (GC).

Network Construction Times				
	BNC Networks			**Wikipedia**
	1m sent.	**2m sent.**	**3m sent.**	**network**
Linguistic Parsing				
Total	$42h\ 43min$	$75h\ 33min$	$112h\ 42min$	$93h\ 36min$
Av. per node	$0.13s$	$0.13s$	$0.13s$	$0.11s$
Network Building				
Total (with GC)	$4h\ 15min$	$13h\ 13min$	$34h\ 39min$	$14h\ 11min$
Total (without GC)	$2h\ 21min$	$7h\ 27min$	$18h\ 56min$	$1h\ 05min$
Av. per node (without GC)	$0.54ms$	$0.98ms$	$1.58ms$	$0.11ms$
Total				
Parsing & Building (with GC)	$\mathbf{46h\ 58min}$	$\mathbf{88h\ 46min}$	$\mathbf{147h\ 21min}$	$\mathbf{107h\ 47min}$

2.3.2. *Network Structures and Concept Coverage*

Structurally, the BNC networks and the Wikipedia network have common features, but also differences. Both types of networks are *scale-free* [8] as their degree distribution follows the power law (cf. Figure 5): most nodes have few links, but a small

percentage of nodes, called "hubs", have many links. This can also clearly be seen from the link statistics in Table 4, which gives an overview of the number of links of the concept or article concept nodes.[e] In the 3m BNC network, 89% of the concept nodes have up to 10 links, 57% have even only up to 3 links. On the other hand, there are nodes with up to 211,594 links. The Wikipedia network is more sparse with almost 95% of article concept nodes only having up to 10 links and 84.5% up to 3 links, while the hubs go up to 145,620 links. This is reflected in the average number of links, which is considerably higher in the 3m BNC network (35.13 per node) than in the Wikipedia network (6.72 per node). We can conclude two points from this observation. Firstly, the BNC network contains more accumulated information per concept than the Wikipedia network. Indeed, with more than 61.41% of article concept nodes only having one link in the Wikipedia network, information is very sparse for most concepts. Secondly, the BNC network is much more interconnected, meaning that it is easier to reach one concept from another.

The list of top five hubs in the two networks in Table 5 shows the difference in concepts covered and their relative prominence in the network. In the BNC network, the top most linked concepts are very general such as *time*, *year*, *people* or *part*, while those in the Wikipedia network are primarily countries such as United_States and France. This result is not surprising. The most linked concept nodes in the BNC network are those derived from frequent nouns and therefore rather generic. However, the top concepts in the Wikipedia network are those that occurred frequently and at the same time were often considered to be worth linking to an article. The Wikipedia guidelines advise contributors to avoid "overlinking",

Table 4: Link statistics of the concept/article concept nodes.

Link Statistics (concept nodes only)				
	BNC Networks			**Wikipedia**
	1m sent.	**2m sent.**	**3m sent.**	**network**
max. num. of links per node	75,195	136,628	211,594	145,620
avg. num. of links per node	27.18	30.63	**35.13**	**6.72**
% of nodes with $\leq$ 10 links	88.70%	89.17%	**89.08%**	**94.90%**
% of nodes with $\leq$ 3 links	58.11%	58.02%	**57.19%**	**84.50%**
% of nodes with = 1 link	23.31%	23.34%	**22.70%**	**61.43**%

i.e. to avoid linking plain English words unless particularly relevant to the topic or the article and to add links that are of value to the reader.[f] Therefore, in the sentence "New York is a city in the United States", *United States* is likely to be

[e]We only consider concept nodes here as these provide the backbone of the network and determine the degree of connectedness.

[f]see http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style_(linking)#Overlinking

Table 5: Top 5 most linked concept/article concept nodes.

Most linked concepts				
	3m BNC		**Wikipedia**	
Rank	**number of links**	**concept**	**number of links**	**concept**
1	211,574	time	145,620	United_States
2	173,411	year	110.796	Association_football
3	156,557	way	92,910	France
4	156,200	people	86,479	Village
5	123,385	man	74,145	Departments_of_France

linked to the state's article, but linking *year* in "Anne Hathaway received the Oscar this year" to the elaborate description of the word year and its definition would be against the guidelines. The hubs are in some way complementary in the two types of networks, with Wikipedia accumulating more information on concrete entities and concepts and not providing much information on concepts known well to every reader. The top BNC concept *time* is only found at rank 3001 in Wikipedia, the top Wikipedia concept UNITED_STATES is placed at rank 1220 in the BNC network.

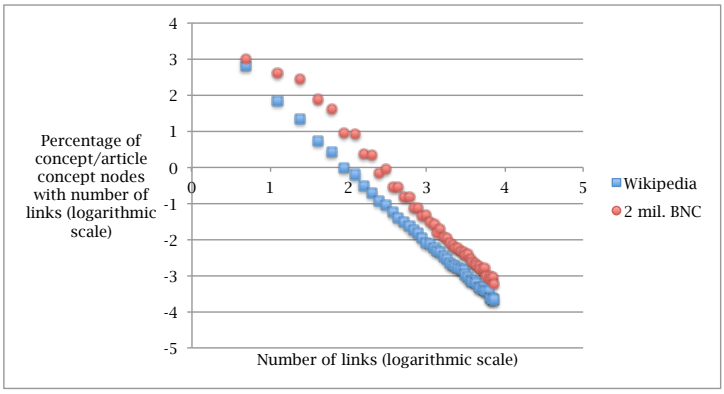

Fig. 5: Degree distribution in the range of of 2 to 35 links for both networks. Both axis are in logarithmic scale. The fact that the plot follows a straight line with negative slope on the log-log scale shows the Power Law relation.

2.4. *Standardised outputs and integration*

In order to be able to use external visualisation and query tools as well as to facilitate the integration with other knowledge resources, we represent the networks using W3C standards for data exchange and integration. Our primary output for-

mat is the Resource Description Framework (RDF), a standard data representation language originally developed as a basis for the Semantic Web, and well suitable for modeling the semantics of the various types of nodes and links in our networks. More specifically, we decided to use the N-Triples[g] syntax of RDF, because it enables a particularly simple representation of the networks that can be easily split and distributed among multiple files – an important prerequisite for efficient processing of large-scale networks. The Wikipedia network representation contains about 90 million triples. Like any other RDF syntax N-Triples is supported by numerous tools and applications, such as Cytoscape[h], an open source platform for complex, large-scale networks, which we use for visualizing and browsing the data. However, despite RDF being suitable to express many core aspects of the network, some could not be translated at all (or only in a very unintuitive way), for example weights and nested reifications. [9] therefore suggest to design a semantic network markup language covering some of these and more aspects as an interface to RDF.

3. Measuring Semantic Relatedness

We evaluate the quality of our semantic networks as a structured representation of knowledge using the task of quantifying semantic similarity and relatedness of concepts. Humans show good agreement on judging that the concepts *baby* and *mother* are more related than *dollar* and *loss* and that *drink* and *ear* are only remotely related while *king* and *cabbage* are unrelated. Within Natural Language Processing applications, this task has shown to be important for word sense disambiguation, text summarisation and information retrieval [10]. However, due to the complex background knowledge and intuition used by humans to judge relatedness scores, this task poses a challenge for automatic systems.

Most approaches to measuring semantic relatedness fall into one of two categories. They either make use of pre-existing knowledge resources such as WordNet [11] or look at distributional properties based on corpora [12, 13]. The resource based approaches achieve good results, but they are inherently restricted in coverage and domain adaptation due to their reliance on costly manual acquisition of the resource. In addition, hierarchical, taxonomically structured resources are generally better suited for measuring semantic similarity than relatedness [10]. In order to take advantage of the strengths of different methods, a recent trend towards supervised combinations of resource-based and distributional approaches can be seen [13, 14]. The semantic relatedness task allows us to compare the quality of our automatically built semantic network resources in comparison to manually created resources. We can also investigate the effect of building a structured representation of underlying text compared to using the text directly.

The most common evaluation setting for Semantic Relatedness is the

[g] `http://www.w3.org/2001/sw/RDFCore/ntriples/`
[h] `http://www.cytoscape.org/`

WordSimilarity-353 data set [12], which provides average human judgments scores of the degree of relatedness of 352 word pairs.[i] The collection contains classically similar word pairs such as *tiger* and *jaguar* as well as topically related pairs such as *movie* and *popcorn*. However, no distinction was made while judging and the instruction was to rate the general degree of semantic relatedness. In addition, in cases of ambiguous words, judges were asked to consider the two words related if they were related in at least one of their senses. For example, *minister-party* would be judged on the basis of *party* referring to a political party. We follow the common practice and use this dataset for our evaluation.

3.1. *Approach*

We measure the semantic relatedness of two concepts by measuring the similarity of the surroundings of their corresponding nodes in the network. It is based on the assumption that semantically related nodes are connected to a similar set of nodes. In other words, we use the context of a node in the network as a representation of its meaning. The approach consists of two steps. First, we use spreading activation to retrieve the network context of a specific node and determine the level of significance of each node in the context. Then, we derive a vector representation of the contexts and measure their similarity using cosine similarity. We restrict the context to only include concept nodes in the BNC networks and article concept nodes in the Wikipedia network. The spreading activation algorithm is based on [7], while we set parameters as appropriate for our model. A certain amount of initial activation is given to a target node and when it is fired, the activation is split evenly between all links of the target node (incoming as well as outgoing) and spreads over the links to the neighbouring nodes. These receive the activation and in turn fire if their activation level exceeds a certain threshold, which varies depending on the node type. We modified the algorithm so that activation does not spread back through a link it just came from by introducing "blocked links" for a node. The spreading activation process stops when no node can fire anymore. We can attenuate the amount of activation spread to ensure that it weakens with distance and that the algorithm reaches a stable state. Spreading activation is a parallel process in which a node fires as soon as its threshold is met, however, we use a sequential implementation. For further implementational details, refer to [7]. In our model, the target node is given the initial activation $\text{iniAct}(x) = \text{numberOfLinks}(x)$. This allows us to ensure that the structure of the context retrieved will be comparable for all words regardless of their number of links. We use attenuation of the activation only on concept/article concept nodes, passing on 90% of it, and leave all other nodes to pass on all of their activation. For all types of nodes, the activation is evenly split between all non-blocked links. For the firing thresholds, we also distinguish between concept

[i]The dataset provided under http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353 contains a duplicate word pair *money-cash*, giving 353 pairs in total

nodes and other nodes. Concept nodes are allowed to fire if their activation level exceeds $\text{thr}(cn) = 0.05{*}\text{numberOfLinks}(cn)$.[j] Taking the attenuation into account, this means that a concept node can only fire if it can still send at least 0.045 units per link. This setting ensures that a node only fires if it has been significantly activated, taking into account the node's level of connectedness. It also prevents tiny amounts of activation from being passed on. This significantly reduces the noise in the system. All other types of nodes have a threshold of 0, passing on any activation they receive. For the context representation, we use the total amount of activation a node has received during the process. The different levels of node activation reflect their significance within the context of the target node.

Table 6: Overview over the settings of the spreading activation algorithm. $\text{nl}(i) = \text{numberOfLinks}(i)$; $\text{nbl}(i) = \text{numberOfBlockedLinks}(i)$

Spreading Activation Settings		
Initial activation of target node x		$\text{iniAct}(x) = \text{nl}(x)$
Firing threshold $\text{thr}(i)$	$0.05{\times}\text{nl}(i)$	if $\text{type}(i) = \text{conceptNode}$
	0	else
Activation received by node j when node i fires at step p of the algorithm $\text{act}_i^p(j)$	0	if $\text{link}_{i\to j} \in \text{blockedLinks}(i)$
	else $\begin{cases} 0.9 \times \frac{\text{act}^p(i)}{\text{nl}(i)-\text{nbl}(i)} \\ \frac{\text{act}^p(i)}{\text{nl}(i)-\text{nbl}(i)} \end{cases}$	if $\text{type}(i) =$conceptNode else

We further weight the retrieved activation level of a context node using an inverse weighting function *iwf* over its number of neighbours (i.e. concept nodes that are directly linked by a path) in order to reduce the impact of highly linked nodes. The more links a context node has, the more likely it is to be activated by another node and the less helpful it is to describe the meaning of a node. In a similar fashion, we smooth the activation levels using a function *sf* to smooth out differences in spreading activation caused by sparseness or imbalance of information. We test two smoothing functions, linear and square root. The vector representation of $\vec{v}(x)$ of the network context of x then has the entries

$$v_i(x) = \frac{\text{sf}(\text{totalAct}(n_i))}{\text{iwf}((n_i))} \tag{1}$$

where the nodes n_i are concept/article concept nodes.[k] We use cosine similarity to compare the context vectors.

$$\text{sim\_rel}(x, y) = cos(\vec{v}(x), \vec{v}(y)) = \frac{\vec{v}(x) \cdot \vec{v}(y)}{\|\vec{v}(x)\| \ \|\vec{v}(y)\|} \tag{2}$$

[j]The parameter was picked without optimisation on the dataset to avoid overfitting.
[k]The implementation contains a canonical ordering of the nodes.

As spreading activation takes several factors into account, such as number of paths, length of paths, level of density and number of connections, this method leverages the full interconnected structure of the network.

3.1.1. *Disambiguation*

In order to use our relatedness measure on the Wikipedia network, we first need to map the words in the word pairs to Wikipedia articles. This mapping is not necessarily straightforward. On the one hand, disambiguation pages often offer a wide range of possible articles. For example, the disambiguation page for KING has 65 entries including MONARCH, KING_(CHESS) and KING_(BAND). On the other hand, some senses may not have a corresponding article such as *round* as part of a fight or the most general meaning of *delay* or *arrival*. This is in particular true for words that refer to abstract concepts such as *importance*. This mismatch can be primarily attributed to the encyclopedic nature of Wikipedia and current lack of coverage. The task of mapping each word in a word pair to a Wikipedia article for semantic relatedness then adds the difficulty of picking the right sense with regard to the respective other word. For example, for the pair *jaguar-cat* we want to pick the articles JAGUAR-CAT, but for *jaguar-car* the articles JAGUAR_CARS-AUTOMOBILE would be appropriate.

In order to separate the performance of our semantic relatedness measure from mapping issues, we created an agreed mapping to Wikipedia articles. We asked two annotators to provide a mapping of all WordSim-353 pairs. Additionally, three annotators mapped a subset of 92 pairs that the previous annotators disagreed on or that at least one of the annotators marked as a low confidence mapping. The annotator agreement rate reflected the difficulties described above, with 25 pairs not achieving an agreement at all (less than 3 annotators agreeing on at least one of the words in the pair) and an additional 34 pairs with low agreement (exactly 3 agreeing annotators). Out of the 327 agreed pairs, 40 contained at least one word without an appropriate Wikipedia article. Overall, we have a set of 280 word pairs with an agreed mapping to article concepts nodes in the network.

We provide the results of the semantic relatedness approach on this set of 280 pairs as well as on an extended supervised mapping of all 352 pairs. To create the extended set, we give each pair for which one of the words had no mapping the median similarity score of the existing pairs. This follows the rationale that if we cannot get a judgement of relatedness from the network, the best bet is to assign a neutral score. For those pairs for which the annotators could not agree, we pick the article which the word was more frequently linked to in the text out of the candidate articles given by the annotators. In other words, we take the "most frequent sense" of the word. In order to investigate the effect of an appropriate mapping, we compare the results of these two sets to a naive mapping of all 352 pairs, taking either the article with the exact name (allowing for redirects) or, if a disambiguation page is returned, the first sense. We will refer to the three sets as 280 Pairs agreed (AG),

352 pairs supervised extension (SE) and 352 pairs naive mapping (NM).

3.2. *Results*

We present our results in four steps. First, we look at the performance of the individual networks, the effect of network size and the difference between the BNC and the Wikipedia networks. Then we combine the two network variants, showing that due to their complementary nature, we can increase performance by taking both into account. In the third part, the results are further refined with regards to measuring semantic similarity versus semantic relatedness using a split of the dataset. Finally, we compare our results to previous approaches.

3.2.1. *Individual networks*

The results on all three sizes of BNC networks are given in Table 7. The different options for inverse weighting, no weighting and linear weighting ($iwf(n_i) = 1$ and $iwf(n_i) = numNeighb(n_i)$) are combined with no and square root smoothing of the spreading activation values ($sf(act(n_i)) = act(n_i)$ and $sf(act(n_i)) = \sqrt{act(n_i)}$. As a network-based baseline, we use all direct neighbours (i.e. concept/article concept nodes connected to the target node by a direct path) as a context, giving each of them equal value. Comparing the results on the 1m network with those of the

Table 7: Spearman correlation results for the three sizes of BNC networks.

BNC results			
		$iwf(n_i) = 1$	$iwf(n_i) = numNeighb(n_i)$
3 million	$act(n_i)$	0.28	0.40
	Baseline	*0.054*	
2 million	$act(n_i)$	0.38	**0.42**
	$\sqrt{act(n_i))}$	0.09	0.31
	Baseline	*0.050*	
1 million	$act(n_i)$	0.20	0.32
	Baseline	*0.044*	
2 million	$act(n_i)$	0.32	0.42
(280 pairs)	Baseline	*0.11*	

2m network, we can see that performance increases with the size of the network. In addition, the substantial amount of improvement over the baseline grows as the network gets bigger, from 0.32 versus 0.04 on the 1m network to 0.42 versus 0.05 on the 2m one. This in accordance with the intuition that more information allows for better scoring and for more accurate distinction of context nodes by spreading activation and link weighting. However, the results drop from the 2m to the 3m network while the baseline still improves. This indicates that the spreading

activation algorithm needs to be adjusted to account for increased density as the networks grow. In the further sections, we will base the combined and detailed results on the 2m network. In all three cases, linear inverse weighting improves the result while no smoothing clearly outperforms square root smoothing (rows omitted for 1m and 3m). We also give the results on the 280 pairs AG subset to provide a direct comparison with the Wikipedia results on this network.

Table 8 shows the results on the Wikipedia network on the two manually disambiguated datasets 280 Pairs (AG) and 352 Pairs (SE) as well as on the naively mapped 352 Pairs (NM). The relatedness measurement gives considerably higher results on the Wikipedia network than on the BNC network (0.65 versus 0.42), confirming the higher quality of the representation due to the use of the hyperlinks and the encyclopedic content. In particular, the simple baseline on the 280 Pairs (AG) already achieves a score of 0.66, which increases to 0.68 by using the spreading activation method. In contrast, the baseline of the BNC network on the same set is 0.11, increased to 0.42. The large difference in baseline performance can be explained by differences in the quality of the relations and concept distinctions in the networks. The fact that the spreading activation method yields a larger improvement over the baseline for the BNC network than for the Wikipedia one can be attributed to the higher density of the first (cf. Table 4). The denser and more interconnected the network, the larger distinctions generated by the spreading activation.

Table 8: Results for the Wikipedia network for all combinations two inverse weighting functions and two smoothing functions.

Wikipedia results			
		$iwf(n_i) = 1$	$iwf(n_i) = numNeighb(n_i)$
280 Pairs (AG)	$act(n_i)$	0.56	0.61
	$\sqrt{act(n_i))}$	**0.68**	0.65
	Baseline	0.66	
352 Pairs (SE)	$act(n_i)$	0.54	0.62
	$\sqrt{act(n_i))}$	**0.65**	0.62
	Baseline	0.62	
352 Pairs (NM)	$act(n_i)$	0.50	0.50
	$\sqrt{act(n_i))}$	**0.53**	0.50
	Baseline	0.52	

The differently mapped datasets illustrate the importance of an appropriate mapping to Wikipedia articles. The results on the subset of pairs with agreed mapping (AG) are the highest at 0.68. On the whole dataset for which the agreed set is extended in a supervised way (SE), the results drop only slightly to 0.65 while they go down to 0.46 with a naive mapping. In contrast to the BNC networks,

the best results are achieved by not using an inverse weighting function but the square root smoothing function on the spreading activation values. Again, this can be attributed to the difference in density of the two network types. For a less linked network, the differences in the number of paths between two nodes are less significant, hence the spreading activation values need to be smoothed. In conclusion, the Wikipedia network outperforms the plain text BNC network of comparable size on the semantic relatedness task, showing higher quality of the network built from hyperlinked text.

3.2.2. *Combination of BNC and Wikipedia networks*

As described in Section 2.3.2, the two types of networks are to some extent complementary in the types of concepts primarily covered and differ in the amount of information collected on individual concepts. Therefore, although the BNC network has a lower overall performance on the semantic relatedness task, we find that for 108 out of 280 and 144 out of 352 pairs the rank given by the BNC network is better than the one by Wikipedia. In order to make use of these respective strengths and create a better overall ranking of the pairs, we combine the two rankings by assigning each pair a weighted average of the its BNC and Wikipedia rank as a score and then ranking the pairs according to this score. The results are given in Table 12. We found the ideal weighting to be 0.75 for the Wikipedia rank and 0.25 for the BNC rank. However, the results are robust towards changes in the exact weighting. The SE score is larger than 0.67 for any Wikipedia weight between 0.65-0.85 and according BNC weight of 0.35-0.15.

Table 9: Results of combing the 2m BNC and the Wikipedia network ranks.

BNC & Wikipedia combination				
	BNC (2 million)	Wikipedia	automatic combination	combination ceiling
280 Pairs (AG)	0.42	0.68	**0.70**	0.83
352 Pairs (SE)	0.42	0.65	**0.68**	0.81

As can be seen from the table, the results from the combination are better than those from the individual networks, with 0.68 versus 0.42 and 0.65 on all 352 pairs. However, a global weighting, i.e. one that is the same for each word pair, is not ideal. The combination ceiling uses an oracle to decide which of the two ranks is better and uses this as a score to show the maximal potential of the combination of the two networks. For the 352 pairs, this ideal combination gives an impressive score of 0.81. In future work, we will therefore investigate under which conditions one or the other gives better results in order to decide the weighting for each pair individually. The factors we have identified so far are the number of links of the target nodes, with sparsely linked target nodes being less adequately described by their context

than well linked nodes, as well as the difference in number of links between the two target nodes of a pairs, with a large imbalance leading to less accurate relatedness scores.

3.2.3. *Relatedness versus Similarity*

When the WordSim353 dataset was established, annotators judged the relatedness on the pairs without making a distinction between similar and related pairs [12]. In order to have a basis for investigating whether approaches perform better on similar pairs versus related pairs, [13] split the WordSim353 dataset into a *relatedness* subset (union of related and unrelated pairs) and a *similarity* subset (union of similar pairs and unrelated pairs). Our results on these subsets are given in Table 10. Both networks individually as well as their combination achieve considerably stronger results on the similarity than on the relatedness subset.

Relatedness versus Similarity Results								
	Relatedness Subset				**Similarity Subset**			
	252/352 or 192/280 Pairs				203/352 or 159/280 Pairs			
	BNC (2m)	Wiki.	autom. comb.	ceiling	BNC (2m)	Wiki.	autom. comb.	ceiling
280 Pairs (AG)	0.43	0.64	0.68	0.82	0.52	0.76	0.78	0.86
352 Pairs (SE)	0.40	0.60	**0.63**	0.81	0.50	0.73	**0.76**	0.87

Table 10: Results on the relatedness and similarity subsets as introduced in [13]

3.2.4. *Comparison with other approaches*

Previous approaches to semantic relatedness fall into two broad categories, those that use a single resource or method and those that combine different types of resources or methods in order to make use of their respective strengths. The combination approaches are supervised, i.e. they are trained on the gold standard using a Support Vector Machine to provide the ideal combination, while the single method approaches have no element of supervision. The latter set of approaches can further be divided into WordNet-based, distributional or corpus-based, Wikipedia-based and semantic network based approaches. The currently best performing single approach, [15], is based on Wikipedia. It is in its core a distributional approach that takes advantage of the fact that the corpus split into Wikipedia articles. The best supervised combination system, [14], integrates the [15] approach with a WordNet-based as well as a distributional method.

Our Wikipedia network by itself performs just under the best WordNet-based approach, [13], with a Spearman coefficient of 0.65 to 0.66. WordNet is a manually created, expert-engineered resource that links concepts (called *synsets*) using a fixed set of relations such as hypernymy, meronymy or causality. In addition, it provides

Comparison with previous approaches		
Single resource or method		
WordNet-based		
Hughes and Ramage [11]	WordNet Graph	0.55
Agirre et al. [13]	WordNet Graph	0.56
Agirre et al. [13]	WordNet Graph incl. disamb. glosses	0.66
Distributional, corpus-based		
Finkelstein et al. [12]	Web corpus	0.56
Agirre et al. [13]	Web corpus	0.66
Wikipedia-based		
Strube and Ponzetto [16]	Wikipedia Category Structure	0.19-0.48
Yeh et al. [17]	Wikipedia Link Structure	0.49
Milne and Witten [18]	Wikipedia Link Structure	0.69
Gabrilovich and Markovitch [15]	Wikipedia articles	$(0.709^{h}$-$)0.75$
Semantic Network		
Harrington [19]	Autom. built semantic network	0.62
Multiple resources or methods (supervised combination)		
Agirre et al. [13]	WordNet(+glosses)+Webcorpus	0.78
Haralambous and Klyuev [14]	ESA+WordNet+GoogleBook corpus	0.86
Our networks		
2m BNC network		0.42
Wikipedia network (SE)		**0.65**
2m BNC & Wikipedia (SE) autom. comb.		**0.68**

Table 11: Comparison of our results to previous approaches on the WordSim-353 dataset.

glosses for the concepts, in which the words were manually linked to their appropriate *synset*. The approach makes use of both of these to create a WordNet graph. In contrast, the relations in our network are automatically and purely linguistically derived from text and the set of concepts as well as the annotations (hyperlinks) in the text are crowd-sourced, sparse and inconsistent. We have therefore demonstrated that the quality of our automatically generated resource is comparable to WordNet on the semantic relatedness task. In addition, by combining the Wikipedia resource with a network built in the same way using only plain text, we can outperform WordNet, yielding a score of 0.68. This is in particular important for specific domains that are not covered by WordNet.

The Wikipedia network also performs just under the best distributional approach, [13], with 0.65 to 0.66, while the combination of the BNC and Wikipedia network outperforms the distributional approaches. In addition, the web corpus the result is based on is several magnitudes larger than our BNC and Wikipedia corpus combined. [13] report major drops in the performance of their distributional mod-

els when they are run on smaller corpora. For example, their bag of words method decreases from 0.64 to 0.52 when restricted to a corpus still one magnitude larger than ours. This is a good indicator that our approach of translating text into a structured network representation first and using annotations when possible makes better use of the corpus text.

In the group of Wikipedia-based approaches, our Wikipedia-based network falls behind [18] by 0.03 and [15] by 0.06-0.10. The first is more comparable to our structure as it uses in- and outgoing hyperlinks of an article to describe the meaning of the underlying concept. It currently has three possible advantages over our approach. First, it uses outgoing links in the whole of the concept's article while we take only those in the first sentence. Secondly, it uses a combination of two different relatedness measurements over the link structure, which yields a higher score than each measure individually. Thirdly, the mapping of words to Wikipedia articles is done by a sophisticated automatic disambiguation that uses the same algorithm later used to measure semantic relatedness to pick the mapping that maximises the relatedness score of the word pair. They show that this automatic disambiguation leads to a substantial improvement of the result over manual disambiguation, revealing difficulties of human annotators in doing the mapping. In future research we aim to integrate these aspects into our approach and investigate whether they lead to an improvement in our case, too.

[15] use an inherently different method. For a word, they build a vector containing the Wikipedia articles it appeared in, filtered and weighted appropriately. For a word pair, these vectors are compared by cosine similarity. The method an be described as distributional while taking advantage of the underlying corpus being structured into topics (ie articles). It is not restricted to concepts that have a corresponding Wikipedia article. Therefore, it does not face the difficulty of zero scores due to non-availability as ours and the other Wikipedia-based approaches do. In addition, it uses the whole of Wikipedia text in contrast to just first sentences in our case. The approaches using a supervised combination of two or more different types of resources are inherently difficult to compete with using one (type of) resource and no supervision. Our approach falls back behind these optimised ones.

The only previous results on the similarity versus relatedness split of the dataset are by [13] themselves. On both subsets, our automatically built Wikipedia network by itself either outperforms or performs equally well as the presented methods on the manually engineered WordNet. The simple weighted combination of the BNC and Wikipedia network then outperforms the stronger distributional bag of words model on the relatedness subset and achieves a result just under the distributional context windows model on the similarity subset. Both distributional models are based on a corpus several magnitudes larger than ours. This result indicates the gain achieved by structuring text in a semantic network instead of taking the surface

[h]Reimplementations in [17] and [14] on Wikipedia snapshots of the respective time of publication scored 0.709 and 0.7394, respectively

form directly. By using semantic analysis and a structured representation, we are able to achieve results comparable to those using simple distributional models but very large corpora.

Relatedness versus Similarity Results						
	Relatedness subset			**Similarity subset**		
Agirre et al. [13]	WN	WNg	BoW	WN	WNg	CW
	0.38	0.56	0.62	0.73	0.72	0.77
	Wikipedia. network		gl.weight.. comb.	Wikipedia network		gl.weight. comb.
352 Pairs (SE)	**0.60**		**0.63**	**0.73**		**0.76**

Table 12: Comparison of results on similarity and relatedness subsets with [13]. WN = WordNet, WNg = WordNet & glosses, BoW = Bag of Words, CW = Context Windows.

4. Noun Compound Interpretation

As a second application we look at noun compound interpretation. In contrast to the semantic relatedness approach, which primarily evaluated the structure of the networks, it allows us to investigate the quality of the relations and paths as descriptions of the relations between concepts. The objective of noun compound interpretation is to find a paraphrase of the relation between the two heads of the compound. For example, an *apple cake* is a *'cake that is made from apples'* or a *'cake that is baked with apples'*. A standardised evaluation for this task was set up as part of SemEval-2010 [20]. The dataset contains 250 training and 388 test noun compounds. For each of them an average number of 71 human annotators were asked to provide paraphrases consisting of a verb or a verb plus prepositions. The resulting paraphrases were ranked according to the number of times annotators came up with them. The aim for systems is to recreate the ranking. The final score is the average Spearman coefficient over all compounds.

In our initial simple network approach, we retrieve paths between the nodes corresponding to the compound parts and match the paraphrases. We then rank the paraphrases according to the number of unique paths they occurred on. Our hypothesis is that due to the network structure, we can make better use of the underlying corpus by reducing sparseness. For example, in the example in Figure 6, the direct connection between *apple* and *cake* only gives the paraphrase *contain*. However, by taking into account paths over the concepts *fruit* and *banana*, we also find *baked with* and *made from*. In addition, we can use structural information such as length of the path or the number of links of the nodes on the path to weight the paths. In this way we can lower the influence of long paths or paths over hubs. These weightings will be part of future work. In the next section we present the results of the simple frequency approach on a small test set.

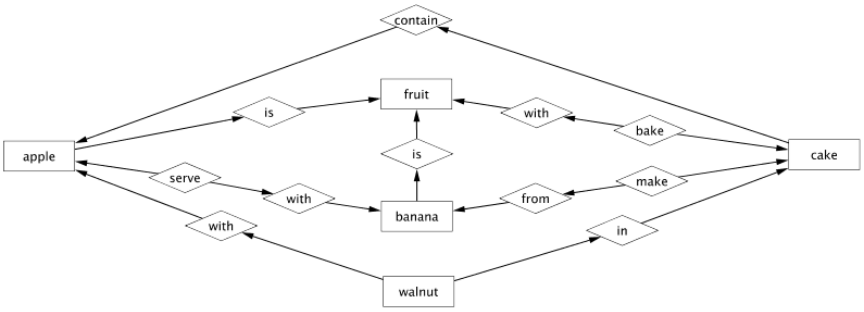

Fig. 6: Illustration of some paths relevant to noun compound paraphrasing of *apple cake*.

4.1. *Preliminary results*

For our initial experiment, we randomly picked 23 noun compounds out of the SemEval2010 Task 9 dataset. We retrieved only paths of maximum concept depth 2 (i.e. paths containing at maximum one concept/article concept node between the source and target node) and allowed for a fallback to depth 3 paths for the Wikipedia network, if there were none of depth 2. As the BNC network is denser than the Wikipedia network, more paths were retrieved on the BNC network. This is reflected in the average recall of 38% of the paraphrases on the BNC against 16% on the Wikipedia network. Table 13 lists our preliminary results in comparison with previous unsupervised approaches.[1] Interestingly, in this case the BNC network performs better than the Wikipedia network, probably due to its higher degree of connectedness. The results indicate that our approach has the potential to outperform previous Wikipedia graph based methods as well as statistical models on large corpora. On the test set, our BNC network performs better than the UCAM model, which is based on the whole of the BNC, again indicating the gain of a structured representation over direct corpus usage. In future work, we will scale our approach to the whole dataset and include weighting of the paths for more sophisticated rankings.

5. Conclusion

In this article, we presented a framework for automatically building large-scale semantic networks from plain text and Wikipedia text. Within a matter of a few days, we were able to build plain text networks covering 870,000 concepts and a Wikipedia network covering more than 2 million concepts. Our spreading activation based semantic relatedness measure on the Wikipedia network achieved comparable results to the best method based on the manually constructed WordNet on the

[1]We exclude the unsupervised system UCD-PN found in [20] from the list as it is somewhat incomparable to our approach. It is based on the dataset itself and scores the probability of a paraphrase appearing in the same set as other paraphrases.

Comparison with previous approaches		
Unsupervised approaches		
NC-Interp	Model using verb-argument frequencies from parsed Web snippets and WordNet smoothing	0.186
Miklosch [21]	Bag of words model on first sentences of Wikipedia articles on selected paths in Wikipedia graph	0.214
UCAM	Model using verb-argument frequencies from the BNC	0.267
UCD-Google-I [22]	Prob. model using pattern frequencies estimated from Google-N-Gram corpus	0.380
Our networks (subset of 23 noun compounds)		
Wikipedia network		**0.247**
2m BNC network		**0.274**

Table 13: Comparison of our preliminary results to previous unsupervised and comparable approaches, based on [20].

WordSim-353 dataset. Furthermore, by combination of the Wikipedia network with the plain text BNC network, we were able to outperform the WordNet based measures. The same combination outperformed the best distributional method while being based on a corpus several magnitudes smaller. This indicates that by using a structured representation of the concepts, relations and attributes occurring the corpus, we can gain more information from the underlying text. We also presented encouraging initial results on noun compound interpretation using our networks that underline the above results and illustrate the flexibility of the constructed resources. In future work, we plan to improve the spreading activation algorithm used in the semantic relatedness measure to better make use of larger network sizes. In addition, we will investigate further strategies of combining the network variants in order to make full use of the shown potential of this combination.

Acknowledgements Pia-Ramona Wojtinnek is funded by the EPSRC and the Oxford University Department of Computer Science. Johanna Völker is financed by a Margarete-von-Wrangell scholarship of the European Social Fund (ESF) and the Ministry of Science, Research and the Arts Baden-Württemberg.

References

[1] M. Ross Quillian. The Teachable Language Comprehender: A Simulation Program and Theory of Language. *Communications of the ACM*, 12(8), 1969.

[2] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.

[3] Brian Harrington and Stephen Clark. ASKNet: Creating and Evaluating Large Scale Integrated Semantic Networks. *International Journal of Semantic Computing (IJSC)*, 2(3), 2008.

[4] Pia-Ramona Wojtinnek and Stephen Pulman. Semantic Relatedness from Automatically Generated Semantic Networks. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '11. Association for Computational Linguistics, 2011.

[5] Stephen Clark and James R. Curran. Parsing the WSJ using CCG and Log-Linear Models. In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 2004.

[6] Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. Wide-Coverage Semantic Representations from a CCG Parser. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, 2004.

[7] Brian Harrington. *ASKNet: Automatically Creating Semantic Knowledge Networks from Natural Language Text.* PhD thesis, University of Oxford, 2009.

[8] Albert-Lásló Barabási and Réla Albert. Emergence of Scaling in Random Networks. *Science*, 286, 1999.

[9] Brian Harrington and Pia-Ramona Wojtinnek. Creating a Standardized Markup Language for Semantic Networks. In *Proceedings of the 5th IEEE International Conference on Semantic Computing*, 2011.

[10] Alexander Budanitsky and Graeme Hirst. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1), 2006.

[11] Thad Hughes and Daniel Ramage. Lexical Semantic Relatedness with Random Graph Walks. In *EMNLP-CoNLL'07: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007.

[12] Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. Placing Search in Context: The Concept Revisited. *ACM Trans. Inf. Syst.*, 20(1), 2002.

[13] Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Paşca, and Aitor Soroa. A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. In *NAACL '09: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2009.

[14] Yannis Haralambous and Vitaly Klyuev. A Semantic Relatedness Measure Based on Combined Encyclopedic, Ontological and Collocational Knowledge. *CoRR*, abs/1107.4723, 2011.

[15] Evgeniy Gabrilovich and Shaul Markovitch. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artifical intelligence*, 2007.

[16] Michael Strube and Simone Paolo Ponzetto. WikiRelate! Computing Semantic Relatedness Using Wikipedia. In *AAAI'06: proceedings of the 21st national conference on Artificial intelligence*, 2006.

[17] Eric Yeh, Daniel Ramage, Christopher D. Manning, Eneko Agirre, and Aitor Soroa. WikiWalk: Random Walks on Wikipedia for Semantic Relatedness. In *TextGraphs-4: Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, Morristown, NJ, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-54-1.

[18] David Milne and Ian H. Witten. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links. In *In Proceedings of AAAI 2008*, 2008.

[19] Brian Harrington. A Semantic Network Approach to Measuring Semantic Relatedness. In *COLING'10: Proceedings of the 23rd International Conference on Computational Linguistics*, 2010.

[20] Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. SemEval-2 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S10-1007`.

[21] Yvonne Miklosch. Wikipedia-based Analysis of Nominal Compounds (Wikipedia-basierte Analyse Nominaler Komposita). Diplomarbeit, Universität Mannheim, 2011.

[22] Guofu Li, Alejandra Lopez-Fernandez, and Tony Veale. UCD-Goggle: A Hybrid System for Noun Compound Paraphrasing. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S10-1051`.