IADIS International Journal on Computer Science and Information Systems
Vol. 6, No.1, pp. 45-57
ISSN: 1646-3692

A SEMANTIC CENTRALITY MEASURE FOR FINDING THE MOST TRUSTWORTHY ACCOUNT

Myriam C. Traub. *Faculty of Digital Media, Furtwangen University, Germany.*
myriam.traub@googlemail.com

Maarten H. Lamers. *Leiden Institute of Advanced Computer Science (LIACS), Leiden University, The Netherlands.*
lamers@liacs.nl

Wilhelm Walter. *Faculty of Digital Media, Furtwangen University, Germany.*
wilhelm.walter@hs-furtwangen.de

ABSTRACT

We propose an algorithmic approach for ranking of differing textual descriptions (accounts) of the same event or story according to their likeliness to best describe the source. Application domains include the ranking of eyewitness reports, historical accounts, and news reports. For this, we developed the concept of "semantic centrality" as a measure of how central a text is among a collection of texts, in terms of its semantic overlap or similarity with all other texts. This measure is based on natural language processing theory, as well as graph theory.
Using three different collections of humanly generated texts (gathered through a recall task, "Chinese Whispers" task, and real-world news reports), we evaluated the proposed method for algorithmic ranking of textual accounts by their trustworthiness to describe source events. In one experiment algorithmic ranking is compared to human ranking. Results indicate that semantic centrality as a measure for trustworthiness of textual accounts is promising and deserves further research attention.

KEYWORDS

Natural language processing, semantic similarity, graphs, eyewitness reports

1. INTRODUCTION

Evaluating differing accounts of an event or story is central in many professional tasks. When investigating crimes, investigators often rely on the accounts of eyewitnesses, which may show dissimilarity concerning minor or even major details, from which they attempt to distill the events as they occurred. Historians deal with similar problems when comparing differing accounts of one event. While criminal investigators still have a good chance to find the most reliable witness by consulting other sources, historians often lack this possibility and an objective way to classify these accounts.

Neuropsychology formulated reasons for distortions of (witness) accounts in the interference theory (Bjork 2009). According to this, the distortions are caused by proactive interference (items memorized before interfere with newly memorized items), retroactive interference (newly memorized items are distorted by items learned afterwards) and output interference (the act of retrieval interferes with the retention of memories). Naturally, reasons for purposefully misrepresenting events in one's account can be found particularly in criminal investigations.

Another type of report that is difficult to evaluate for a general public are online news items. Numerous online news portals offer texts about the latest happenings worldwide, typically condensed into few hundred words. Although the primary source of information is very often the same (e.g. a press conference or news agency), the articles commonly diverge with respect to details.

This raises the question whether it is possible to find in a systematic way from a collection of accounts (or stories) the one that represents best the actual happenings or source story, based on nothing more than the contents of the accounts themselves. Such a systematic or algorithmic approach is obviously not able to reconstruct the absolute truth, since it may not even be (fully) described by the accounts. However, it may be able to generate a prioritization of accounts, based on criteria that would otherwise have remained unseen by a human judge of the accounts. Such a prioritization could then aid the forming of a human judgment of the relative validity of the differing accounts.

The aim of our research is to algorithmically find the most "trustworthy" account from a collection of distorted accounts *T* of the same source event or story *S*. Since there is no practical way to determine if the facts that make up accounts *T* are objectively true, the following approach was taken.

We assume that deviations from the source story *S* vary among accounts in *T*. In other words, multiple accounts in *T* will likely differ from the source *S* on varying aspects, not all on the same aspect. As a result, we are interested in the shared or similar elements among the different texts that make up *T*, because they are most likely to reflect the source *S*. Naturally, when speaking about similarity, we intend the semantic similarity of elements of accounts in *T*, not necessarily the syntactic similarity. Since it is practically unfeasible to reconstruct a hypothetical account *A* of source *S* from the overlapping semantic elements in *T*, we aim to answer the following question:

> *Given a source story or event* S *and a collection of derived (and therefore distorted) texts* T *describing* S*, how can we select from* T *the text that best describes* S*?*

As with most other fields of natural language processing, the measuring of semantic similarity "lacks unified, objective, accepted standards on measurement of similarity, which makes the comparability and comprehensibility of similarity not good" (Rong and Wen 2008). Still, measuring the semantic similarity of texts seems to be an appropriate tool for our purposes as it has the faculty to tackle at least approximately the "two main problems of natural language processing" (Gabrilovich and Markovitch 2009): polysemy and synonymy.

2. RELATED WORK

Research on the concept of “similarity” is not only done on the basis of texts in natural language processing, but also for example in cognitive psychology where it is opposed with the concept of “rules” (Pothos 2005). While rules are used to make a clear statement about an object (“if there are no geometrical shapes, this cannot be a site diagram” (Withrow 2009)), “similarity” describes objects in a more imprecise and associative way (“this looks like a laptop computer, so it goes in my ‘laptop’ category” (Withrow 2009)). The results of this branch of psychological research allow conclusions on cognitive processes that deal with categorization, which itself is related to our research question.

Another field related to cognitive psychology is case-based reasoning. Here, a problem is solved using an adapted approach taken from a known set of more or less similar problems (Aha 2001). An approach that goes further in the direction of the previously mentioned psychological similarity research is similarity-based reasoning. A research work that combines natural language processing (NLP) and crime analysis data was done by Richard Bache et al. (Bache et al 2007). They compare police reports of unsolved cases to reports of solved cases. Assuming that serial offenders commit their crimes usually in a fairly similar way, they argue that same-offender police reports contain many similar terms and therefore the similarity of the texts is very high. Bache and his team demonstrated that it is possible to prioritize known criminals by the likelihood of them being the suspect in an unsolved case, through NLP based comparison of crime records.

At first glance, this approach resembles the previously mentioned case-based reasoning. However, there is a substantial difference. In case-based reasoning, information about prior solutions to similar cases is used to adapt the previous solution to suit the current case. Bache et al., on the other hand, use "old" cases as part of the solution. The solutions that were used to solve the old crime cases are not part of Bache’s approach.

3. MEASURING SEMANTIC SIMILARITY

Our approach relies strongly on the capability of determining semantic similarity between texts. Semantic analysis is a major topic of natural language processing research, a broad and highly challenging field of research that embraces different human-computer interaction problems that deal with generation or understanding of spoken or written natural language. Problem domains of NLP include speech segmentation, part-of-speech tagging and word-sense disambiguation. Explaining current research of all subfields of natural language processing and especially of information retrieval that are related to our work would exceed the scope of this paper. However, some interesting and for this work influential ideas are mentioned.

Since in the problem domain of semantic analysis it is often not possible to deduce required contextual information only from a text that is being processed, Gabrilovich and Makovitch (Gabrilovich and Markovitch 2009) suggest to enrich input texts with information gained from a source that would as well serve humans who lack certain knowledge to fully understand a topic. Their Explicit Semantic Analysis (ESA) accepts different types of texts as input data to set them in a relation to the knowledge gained from Wikipedia. In their own words, the central aspect of their work “is representing the meaning of text as a weighted combination of all Wikipedia concepts" (Gabrilovich and Markovitch 2009). The resulting high-dimensional representation of a texts semantics can then be used for further processing or classification.

A straightforward approach to text similarity measuring is lexical matching, as implemented in the Text::Similarity::Overlaps package for Perl written by Michelizzi and Pedersen (Michelezzi and Pedersen 2009). The algorithm is able to find pairs of identical words in two input texts. It only matches literally overlapping strings ("cat" and "cats" would not match). The more identical words or word sequences the two texts contain, the higher is their resulting similarity score.

This simple approach implies that words with similar meanings, like "building" and "house" for example are not recognized as semantically related. Therefore, lexical matching can be quite imprecise for texts with different vocabularies. Likewise, semantically unrelated texts that use similar vocabulary (for example about the financial institution and the riverside "bank") would be falsely classified as related.

Another straightforward semantic similarity measure is the Google Similarity Distance (Cilibrasi and Vitanyi 2007). It measures the semantic similarity for two words based on the results of corresponding Google web search. Words that co-occur very often in documents indexed by Google are considered to have related meaning. This, however, results in the fact that words with contrary meanings (like "true" and "false") are rated as semantically similar because they often co-occur in the same documents (Cilibrasi and Vitanyi 2007). While studying the Google Similarity Distance, the idea arose to use Google.com the same way as in the Google Similarity Distance project, but for entire texts. Unfortunately, Google processes only the first 32 words of a given query (in October 2009), so it is not possible to use texts as input for a search query.

4. SEMANTIC CENTRALITY

Our approach to find from a collection of texts *T* the text that best describes the source *S*, is based on identifying the text that has the largest semantic overlap with all other texts in *T*. One might consider this the most "semantically central" of all texts in a collection *T*. Our hypothesis is that it best describes the source *S* of all derived texts in *T*. This hypothesis is tested through experiments described further in this paper.

To find this semantically central text from collection *T*, we construct a fully connected, directed and weighted graph in which the vertices correspond to the texts in *T*. Every arc between vertices for texts and is weighted with a measure of the semantic similarity between these texts.

4.1 Constructing a Semantic Graph

The semantic similarity measure we apply for our experiments is based on an approach proposed by Courtney Corley and Rada Mihalcea (Corley and Mihalcea 2005). Their work uses research on word-to-word similarity metrics and applies it to semantic comparison of texts. By lack of a ready-to-use implementation of the Corley and Mihalcea algorithm, we had to rebuild it step-by-step. Figure 1 offers an overview of the modules used in our implementation.

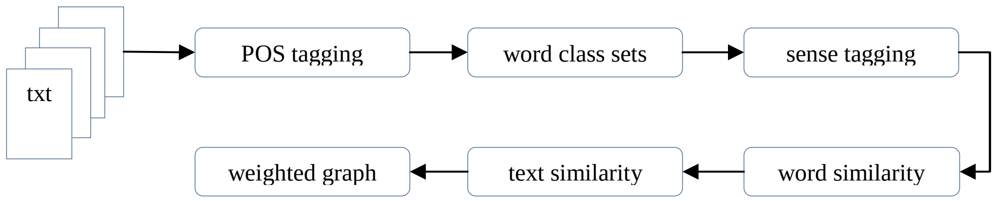

Figure 1. Flow chart for the sequence of steps taken to obtain a weighted graph representing all semantic relations between the texts

We start with a collection of different texts that describe a common story or event. To illustrate how we obtain a semantic similarity measure for all the texts in reference to each other, we use the example of one sentence:

"In the experiment, a cat is put in an airtight box, together with a bomb."

Each text is processed by the part-of-speech (POS) tagger developed at Stanford University (Toutanova 2009). A POS tagger assigns each word a tag with the respective word category, for example "/NN" for a noun or "/DT" for determiner. The output is a text with tagged words, e. g.:

"In/IN the/DT experiment/NN a/DT cat/NN is/VBZ put/VBN in/IN an/DT airtight/JJ box,/NN together/RB with/IN a/DT bomb./NN".

This text is further processed and all verbs, nouns, adjectives and adverbs are put in separate word class sets. Words that are irrelevant for the semantic of a text are filtered out (like "and", "the", "a", ...). The result appears as:

verbs = {is, put}
nouns = {experiment, cat, box, bomb}
adjectives = {airtight}
adverbs = {}.

The word class sets are processed further by adding a sense tag to each word, using a Perl module entitled WordNet::SenseRelate::AllWords developed by Ted Pedersen and colleagues at the University of Minnesota Duluth. It is based on WordNet (Princeton University 2009), a lexical database organized in sets of synonyms (synsets). For the word "cat" WordNet would return a list of the different word classes and senses the word can have, starting with "feline mammal usually having thick soft fur and no ability to roar" upto "a method of examining body organs by scanning them with X-rays and using a computer to construct a series of cross-sectional scans along a single axis" (Princeton University 2009).

A tag is post-fixed to each word containing its wordclass (e.g. "#n" for a noun) and a number for the sense. In our example sentence the cat is more probable to be a feline mammal than a medical apparatus, so it is tagged with "#1" (for the first entry in the list of possible word senses).

verbs = {is#v#1 put#v#1}

nouns = {experiment#n#1 cat#n#1 box#n#1 bomb#n#1}
adjectives = {airtight#a#2}
adverbs = {}

After processing the text as described, it is possible to calculate a measure of semantic relatedness between words by using another module developed by Pedersen and colleagues, entitled the WordNet::Similarity::Path. It works from the assumption that words are strongly semantically related if their vertices in a tree-like representation of WordNet are separated via only a short path.

A program developed by us finds the highest similarity value for the words of text T_i by semantically comparing each word of text T_i with each word of text T_j.

The algorithm proposed by Corley and Mihalcea (Corley and Mihalcea 2005) requires computation of an “inverse document frequency” for each word *w*. It is defined as the logarithm of the total number of documents in a reference collection divided by the number of these documents that contains word *w*. For the reference document collection we use a corpus developed by Microsoft Research Laboratories (Microsoft Research 2009), that contains manually collected 2-sentence documents about random topics.

We can then calculate the similarity measure between a pair of texts T_i and T_j, as proposed by Corley and Mihalcea (Corley and Mihalcea 2005). It is these values that serve as weights in the directed graph connecting all texts *T*.

We chose a directed graph because the similarity measures we obtain are one-directional. This means that comparing a text T_i to a text T_j may return a different value than a comparison of text T_j to text T_i. For texts that are identical, the semantic similarity is 100%, expressed in a value of 1.0. This is represented in the arcs that point back on their source vertex (loops).

When we compare a text T_i to text T_j and T_i is a paragraph taken from T_j, we find semantically strongly connected words for each word of text in text T_i. The resulting similarity is therefore 1.0. If we compare text T_j with text T_i, however, we find only suitable matches for some of the words in text. The resulting similarity value is consequently lower than 1.0, although we still compare parts of the same text.

Corley and Mihalcea (Corley and Mihalcea 2005) go one step further in their work and combine the one-directional semantic similarity values into one bi-directional value. For our purposes, this step would mean a loss of precision as we explicitly look for the semantic similarity of one text towards all other texts.

After processing the texts within a collection *T* with the described algorithms we obtain for *T* a fully connected and directed graph that represents all semantic relations among its texts.

4.2 Finding the Most Central Text

Our next step is to find the most “semantically central” text represented as the most central vertex in our graph. In graph theory, many different concepts of centrality are known. According to Brandes and Erlebach (Brandes and Erlebach 2005) “almost everybody introduced his or her centrality without giving a strict definition for centrality in general”. Discussing all of them would go beyond the scope of the paper, therefore we confine ourselves to the most common centrality measures for a graph $K_n = (V;E;w)$:

• **Degree centrality**. In an undirected graph, degree centrality is defined as the degree $d(v)$ of a vertex *v*. In directed graphs it is defined as the in-degree centrality and out-degree centrality. Since we have a fully connected digraph, all vertices have the same degree $|V|$. Degree centrality is thus not applicable to our graphs representing semantic relatedness.

• **Shortest paths centrality**. In graph theory, there are several centrality measures based on shortest paths, like stress centrality or shortest path betweenness centrality. This measure indicates for example if the respective vertex is part of many shortest paths within the graph. In our fully connected digraph, the (topologically) shortest paths are by definition direct connections. When calculating the shortest path in metric space, the result probably looks very much alike. Moreover, the paths in our graph are not relevant for the problem we attempt to solve.

• **Closeness centrality**. This approach defines the centrality of a vertex *v* as the sum of its distances to all other vertices *u* in the graph, as in Equation 1. This can be calculated for both, metric and topological space.

Equation 1. Calculation of the closeness centrality measure for vertex *v* in a graph with vertices *V*

$$\sum_{u \in V} \mathrm{dist}(v, u)$$

• **Eigenvector centrality**. Eigenvector centrality uses the adjacency matrix of a graph to weigh not only direct connections as with most other similarity measures, but also indirect connections of any length. Again, in our fully connected graph, this measure is not applicable.

The concept of centrality that applies best to our work is closeness centrality. In combination with the semantic similarity measure that defines the weights in a graph for texts in collection *T*, closeness centrality yields a measure of semantic centrality for each text in *T*. We can then rank the texts according to their semantic centrality within the whole collection of texts *T*.

5. EXPERIMENTS

This section introduces the texts we used for our experiments. To show the broad applicability we tried to gain the text collections from as different sources as possible.

5.1 Schroedinger's Cat

Data for this experiment was gathered during a lecture at Leiden University's Media Technology M.Sc. program. As part of the lecture, a short video (163 seconds) was shown from the video-podcast “Nanotechnology” (published by Oxford University, 2009) about the thought experiment known as “Schroedinger's Cat” by the physicist Erwin Schroedinger. After the lecture, the students were asked to retell in written form what they had seen in the video. A set of 18 texts written by the students were collected, varying in size between 35 and 180 words, to which we added the transcript of the speaker in the video (about 450 words).

The un-edited texts describing the Schroedinger's Cat video were given to 23 international engineering students from Furtwangen University. Each student received a subset of five randomly chosen texts, and was asked to rate them from “most trustworthy” to “least trustworthy" in describing the events in the video. Each text was rated by five students. Since most students in the test group were unfamiliar with the topic of quantum mechanics, the thought experiment by Schroedinger was unknown to them. Moreover, they were not shown the video. As a result, they could only judge the texts based on their contents. The results from this human judgment were compared to those from our algorithmic approach.

5.2 Chinese Whispers: King M

This experiment is named after the popular "Chinese Whispers" children's game, in which children successively whisper words or sentences to each other and thereby often unwittingly change their content. According to Claude Shannon, a pioneer of information theory, most communication channels are perturbed by a noise source (Shannon 1998). The Chinese Whispers game is a good example for such a "noise" changing the content of human communication.

The source text was a short story written by Crispin Oduobuk, entitled "King M" (393 words). Subjects in this experiment were all students from different engineering faculties of Furtwangen University. The story was given to four test subjects with the instructions to read it, to put it away and then to retell it. Transcripts of the retold stories (T_1 ... T_4) were given to another four test subjects, whom repeated the procedure to retell them again (T_5 ... T_8). This happened three times yielding 12 "distorted" texts (T_1 ... T_{12}) plus the original source story *S*.

With the texts gained from the Chinese Whispers experiment we hope to have a data set in which noise increases by repetitious retelling. The results of our algorithmic approach can be viewed in light of the resulting decrease of retelling accuracy.

5.3 News

In June 2009 an Air France Airbus crashed into the Atlantic Ocean near the Brazilian Fernando de Noronha islands killing all passengers and crew members. Only about three weeks later, the first victims of the crash could be identified. For our experiment, nine articles about the identification of victims of the AirFrance crash were collected on June 26, 2009 from different international news websites such as CNN, BBC and CBCnews. The number of words in the texts ranges from 87 to 552.

6. ANALYZING SEMANTIC CENTRALITY

In this section we analyze the data from our experiments with the concept of semantic centrality, different texts, and human ranking of texts.

6.1 Schroedinger's Cat

The analysis for this experiment is split in two parts. First, we analyze the graph obtained from the algorithmic semantic analyses. Secondly, we analyze the correlation between the semantic centrality ranking and human ranking of all texts.

Table 1. Semantic centrality values and ranking order, and human ranking order of the Schroedinger's Cat texts, including the source transcript *S*. Ranking order 1 indicates the most centrally / trustworthy ranked representation of source *S*

Text	Word count (w)	Semantic centrality value (x)	Semantic centrality ranking order (y)	Human ranking order (z)
T_1	67	9.99	6	10
T_2	74	8.58	12	8/9
T_3	75	9.71	8	15/16
T_4	33	9.78	7	18
T_5	115	7.37	14	2
T_6	143	8.66	11	12
T_7	109	12.19	1	4
T_8	93	6.51	17	8/9
T_9	116	10.37	3	3
T_{10}	103	10.31	4	1
T_{11}	110	5.48	19	13
T_{12}	115	7.09	16	5
T_{13}	115	8.51	13	11
T_{14}	57	10.08	5	14
T_{15}	39	6.24	18	17
T_{16}	72	10.60	2	15/16
T_{17}	101	9.69	9	6
T_{18}	182	9.62	10	7
S	505	7.20	15	-

Correlation (w,x) = -0.21, p (one-sided) = 0.19, p (two-sided) = 0.38
SROC (y,z) = 0.14, p = 0.59

6.1.1 Semantic Centrality Graph

The semantic centrality values we obtain for the Schroedinger's Cat collection of texts vary between 5.48 for the least central vertex (T_{11}) to 12.19 for the most central vertex (T_7) as shown in Table 1. Since the source text *S* was also processed, one would reasonably assume it to be at least in the upper third of the ranking if not on the first position. The fact that it is found at rank order 15 (out of 19) may convey the impression that our semantic centrality measure, failed.

However, the implications of these results are more complex than it may seem at first sight. The source text *S* is not only by far the longest text in the collection, but also the most detailed one. Most of the texts in collection *T* contain less detail than *S*. Facts of *S* that do not show up in *T* are (according to our hypothesis) not considered representative. Consequently the semantic centrality value for *S* is rather low.

6.1.2 Correlation with Human Ranking

Table 1 shows the rank orders for semantic centrality and human ranking. The lower its human ranking value, the better a text is deemed to represent the unknown source story *S*; e.g. a ranking value of 1 means it is assumed to best represent the source story. Text T_2 and T_8, and texts T_3 and T_{16} respectively were ranked equally trustworthy by human participants.

Since the source story *S* itself was not ranked by human judges, it could not be included in a rank order comparison. The corrected Spearman Rank Order Correlation (SROC) between the semantic centrality and human ranking orders of texts T_1 ... T_{18} is 0.14 (p=0.59), indicating weak positive, non-significant correlation between both rankings. The correlation between word count and semantic centrality value of all texts (including *S*) is -0.21 (Table 1) with *p*-values of 0.19 (one-sided) and 0.38 (two-sided).

6.2 Chinese Whispers

As in the Schroedinger's Cat experiment, the original "King M" story *S* is also ranked near-last by semantic centrality (see results in Table 2). The reasons are the same as in the Schroedinger's Cat experiment: with 393 words *S* is by far the longest and most detailed story in the collection.

The corrected SROC between the semantic centrality ranking and the sequence ranking is -0.29 (p=0.33).

The sequence ranking order represents the number of distortion steps that the source text went through to generate each of the texts. It was defined such that it can be applied in corrected Spearman Rank Order Correlation calculations.

Table 2. Chinese Whispers data set results. Sequence ranking order represents the different stages of distortion of the original source story *S*

Text	Word count (w)	Number of distortion steps	Semantic centrality value (x)	Semantic centrality ranking order (y)	Sequence ranking order (z)
T_1	126	1	7.57	12	3.5
T_2	114	1	9.67	1	3.5
T_3	103	1	9.07	4	3.5
T_4	217	1	8.47	10	3.5
T_5	80	2	7.95	11	7.5
T_6	114	2	9.57	2	7.5
T_7	118	2	8.88	6	7.5
T_8	223	2	8.64	8	7.5
T_9	104	3	8.66	7	11.5
T_{10}	91	3	8.47	5	11.5
T_{11}	120	3	9.07	3	11.5
T_{12}	168	3	8.51	9	11.5
S	393	0	6.94	13	1
Correlation (w,x) = -0.59, p (one-sided) = 0.02, p (two-sided) = 0.04					
SROC (y,z) = -0.29, p = 0.33					

6.3 News

The graph depicting the semantic similarity between all news items in the collection shows that the vertex representing news item T_3 taken from skynews.com.au (SkyNews 2009) is by far the most semantically central vertex in the graph (Table 3). This news item is with 87 words the shortest one in the collection and gives only a very condensed overview of the events.

Table 3. Algorithm-based ranking of the nine news item texts

Text	Word count (w)	Semantic centrality value (x)	Semantic centrality ranking order (y)
T_1	552	3.76	9
T_2	306	4.70	7
T_3	351	4.90	3
T_4	137	4.79	5
T_5	87	6.01	1
T_6	235	5.17	2
T_7	526	4.41	8
T_8	238	4.75	6
T_9	410	4.81	4
Correlation (w,x) = -0.81, p (one-sided) = 0.00, p (two-sided) = 0.01			

The least central vertex in this graph represents a news item taken from cnn.com (T_1). It consists of 552 words and gives information that is not only related to the identification of the victims, but also to other aspects of the air crash like the location of the wreckage. For example, no other news item mentions the difficulties the authorities had when they tried to find relatives of all 228 victims. This aspect leads to low semantic centrality among the collection of news items.

7. DISCUSSION

With this project we propose an algorithmic approach to ranking the texts within a collection according to their likeliness to best describe their source story or event. For this, we developed the concept of semantic centrality as a measure of how central a text is among a collection of texts, in terms of semantic overlap or similarity. In constructing this measure of semantic centrality we make use of extensive natural language processing theory, as well as graph theory. Using different collections of humanly generated texts, we then experimented with our semantic centrality measure, evaluating its application in ranking texts.

For the Schroedinger's Cat collection of texts, we were able to compare semantic centrality rankings to human rankings. The semantic centrality rankings showed only a very low positive correlation (p>0.1) with the human ranking of trustworthiness. This may be due to the fact that the criteria for the human and algorithmic ranking significantly diverge, and that the concept of semantic centrality in itself may not correspond to perceived trustworthiness. It would be interesting to know what influences the reader of a collection of texts, to rate one text more trustworthy than another.

When reflecting on the setup and realization of this experiment we would suggest a few changes. First, we would choose a topic that our human test subjects can relate to. Although most test subjects a-priori stated to have basic knowledge about Schroedinger's thought experiment, it became clear a-posteriori that most of them had indeed no such knowledge.

When designing the experiment, a topic was selected that is not well-known. As a result, when ranking the texts trustworthiness, the readers cannot rely on contextual knowledge but only on information within the texts. Thereby we ensure that the ranking is not disturbed by external influences. The lack of this prior knowledge in combination with the large variations concerning the quality of the texts (in terms of grammar, writing style, and content), however, may have made it difficult for subjects to rank the texts. This issue may be avoided by choosing texts with simpler content, for example about a topic from popular sciences.

Surprisingly, the correlation between word count and semantic centrality was much lower (and far less statistically significant) for the Schroedinger's Cat texts than for the other two collections. The correlation is statistically significant strong negative for both the Chinese Whispers and news items collections.

Altogether we did succeed in finding a plausible measure for semantic centrality within collections of texts, the semantic centrality. The answer to the question, if this measure is actually viable for expressing trustworthiness to describe source events, however, requires further research.

REFERENCES

Aha, D. W. et al, 2001. *Conversational Case-Based Reasoning.* Applied Intelligence 14, pp. 932.

Bache, R. et al, 2007. *Application of Language Models to Suspect Prioritisation and Suspect Likelihood in Serial Crimes*, Third International Symposium on Information Assurance and Security.

Bjork, R.A, 2009. *Interference and Forgetting*, In J. H. Byrne (Ed.), Encyclopedia of learning and memory, 2nd ed., pp. 268-273 (accessed: 21/09/2009).

Brandes, U. and Erlebach, T. (Eds.), 2005. *Network Analysis - Methodological Foundations*, Springer Verlag.

Cilibrasi, R. and Vitanyi, P., 2007. *The Google Similarity Distance*, IEEE Transactions On Knowledge And Data Engineering, Vol. 19, No. 3.

Corley, C. and Mihalcea, R., 2005. *Measuring the Semantic Similarity of Texts*, Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, pp. 13-18.

Forbus, K. and Gentner, D., 1991. *Similarity-based Cognitive Architecture,* SIGART Bulletin, Vol. 2, No. 4.

Gabrilovich, E. and Markovitch, S., 2009. *Wikipedia-based Semantic Interpretation for Natural Language Processing*. Journal of Artificial Intelligence Research 34, 2009

Klebanov, B. et al., 2004. *Text Simplification for Information-Seeking Applications*, On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE, Meersman, R. and Tari, Z. (eds), LNCS Vol. 3290, pp. 735-747.

Michelizzi, J. and Pedersen, T., 2009. *Text::Similarity::Overlaps*., Perl module, University of Minnesota, Duluth (accessed: 26/08/2009).

Microsoft Research, 2009. Microsoft *Research Paraphrase Corpus*, http://research.microsoft.com/ (accessed: 17/10/2009).

Pothos, E., 2005. *The rules versus similarity distinction*. Behavioral and Brain Sciences, No. 28, pp. 1 - 49.

Rong, F. and Wen, X. 2008. *Study on Text Semantic Similarity in Information Retrieval*. Proceedings of the 2008 IEEE International Conference on Information and Automation.

Shannon, C., 1998, *Communication in the Presence of Noise*. Classic paper, Proceedings of the IEEE, Vol. 86, No. 2.

SkyNews.com.au Australia. *11 identified from Air France flight* (accessed: 26/06/2009).

Toutanova, K., 2009. *Stanford Log-linear Part-Of-Speech Tagger*. The Stanford Natural Language Processing Group (accessed: 17/10/2009).

Wells, G. and Olson, E., 2003. *Eyewitness Testimony*. Annual Review of Psychology, No. 54, pp. 277-295.

Withrow, J., 2009, *Cognitive Psychology & IA: From Theory to Practice*.

http://www.boxesandarrows.com/view/cognitive_psychology_ia_from_theory_to_practice (accessed: 27/09/2009).

Princeton University, Cognitive Science Laboratory. *WordNet Search 3.0*., http://wordnetweb.princeton.edu/perl/webwn (accessed: 26/08/2009).