G^{con}: A Graph-Based Technique for Resolving Ambiguity in Query Translation Candidates

Dong Zhou
University of Nottingham
School of Computer Science
Nottingham, NG8 1BB, UK
(+44) 0115-8466591
dxz@cs.nott.ac.uk

Mark Truran
University of Teesside
School of Computing
Middlesbrough, TS1 3BA, UK
(+44) 01642-342667
m.a.truran@tees.ac.uk

Tim Brailsford
University of Nottingham
School of Computer Science
Nottingham, NG8 1BB, UK
(+44) 0115-9514231
tjb@cs.nott.ac.uk

Helen Ashman
University of South Australia
School of Computer & IS
Crawley, WA 6009, Australia
(+61) 8830-25335
helen.ashman@unisa.edu.au

James Goulding
University of Nottingham
School of Computer Science
Nottingham, NG8 1BB, UK
(+44) 0115-8466591
jog@cs.nott.ac.uk

ABSTRACT

In the field of cross-language information retrieval (CLIR), the resolution of lexical ambiguity is a key challenge. Common mechanisms for the translation of query terms from one language to another typically produce a set of possible translation candidates, rather than some authoritative result. Correctly reducing a list of possible candidates down to a single translation is an enduring problem. Thus far, solutions have concentrated upon the use of the use of term co-occurrence information to guide the process of resolving translation-based ambiguity. In this paper we introduce a new disambiguation strategy which employs a graph-based analysis of generated co-occurrence data to determine the most appropriate translation for a given term.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; E.1 [**Date Structures**]: Graphs and networks.

General Terms

Algorithms, Design, Experimentation, Performance.

Keywords

Cross Language Information Retrieval, Disambiguation, Co-Occurrence Measure, Graph Analysis.

1 INTRODUCTION

Ambiguity is an inherent problem when translating a multi term query using a bilingual dictionary. This problem stems from the choice of possible translations. A typical bilingual dictionary will provide a set of alternative translations for each term within any given query. Choosing the correct translation of each term is a difficult task, and one that can seriously impact upon the efficiency of any related retrieval functions. Research into this problem has repeatedly suggested that co-occurrence information, extracted from a representative document collection, may point the way towards the resolution of translation ambiguity [4, 5, 6, 8, 9, 14]. The approach described here involves measuring the degree with which all of the possible translation candidates co-occur with each other within the confines of a particular corpus. This operation typically yields a calculation of the most likely translation of each term in the particular query. In this paper we extend the scope of this basic methodology by considering co-occurrence information as a network susceptible to graph based analysis (G^{con}*: Graph* over *Co-Occurrence Network*).

The remainder of the paper is organized as follows: section 2 introduces the proposed graph-based technique for resolving translation ambiguities. In section 3, we describe an evaluative experiment using data provided by the NTCIR[1] and CLEF[2] organization committees. Section 4 briefly examines related work and its relationship to our algorithm. Finally, in section 5, we present our conclusions and speculate on further research opportunities.

2 TERM CO-OCCURENCE DATA

The hypothesis behind the use of co-occurrence data to resolve translation ambiguities is fairly simple. For any query (containing multiple terms) which must be translated, the correct translations of individual query terms will tend to co-occur as part of a given sub-language, while the incorrect translations of individual query terms will not. Therefore, by examining the pattern of co-occurrence statistics for each pair of possible translations of individual query terms within some representative document collection, it is possible to infer the most likely translation of the query as a whole. Various techniques can, and have, been used to calculate the significance of word co-occurrence for this very

SAC'08, March 16-20, 2008, Fortaleza, Ceará, Brazil.

[1] http://research.nii.ac.jp/ntcir/

[2] http://www.clef-campaign.org/

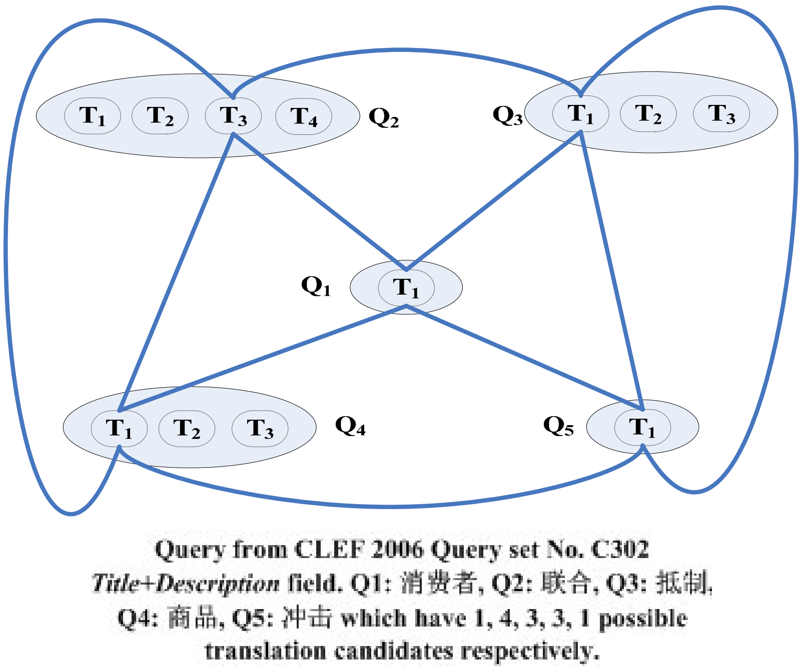

Figure 1. Co-occurrence Network Viewed as a Graph

purpose, including the Sorenson-Dice similarity co-efficient, the Chi-squared test and measures based on mutual information (MI) [6, 9].

This general technique is particularly effective when query translation is effected using a machine readable bilingual dictionary. However, one inherent drawback of to this general approach has been identified and studied extensively. This problem relates to the mutually dependent relationship between each term within a multiple term query. Ideally, for each query term under consideration, we would like to choose the best translation that is *consistent* with the translations selected for all remaining query terms. However, this process of inter-term optimization has proved computationally complex for even the shortest of queries [4]. A common workaround, used by several researchers working on this particular problem [4, 8], involves use of the following greedy algorithm:

1. Given a query *C* in source language (say, Chinese) which contains several terms $\{c_1, c_2, \cdots, c_n\}$, for each term $c_i, i \in [1, n]$ in *C*, obtain the translation candidates $T(c_i) = \{t_{i,1}, t_{i,2}, \cdots, t_{i,m}\}$

2. for each set of candidates $T(c_i)$

   (1) for each $t_{i,j} \in T(c_i), j \in [1, m]$, define the similarity score between the term $t_{i,j}$ and the other set $T(c_k), k \neq i$ as the sum of the similarities between $t_{i,j}$ and each term in set $T(c_k)$. As follows

$$sim(t_{i,j}, T(c_k)) = \sum_{t_{k,l} \in T(c_k)} sim(t_{i,j}, t_{k,l})$$

   (2) compute the cohesion score for $t_{i,j}$ as

$$co(t_{i,j}) = \sum_{\forall i' \neq i} sim(t_{i,j}, T(c_{i'}))$$

   (3) select the term $t \in T(c_i)$ with the highest cohesion score

$$t = \arg\max co(t_{i,j})$$

The similarity measurement is usually calculated using the mutual information between two terms as shown below:

$$sim(x, y) = MI(x, y) = P(x, y) \times \log \frac{P(x, y)}{P(x) \times P(y)}$$

Where $P(x)$ is the unigram probability for word *x*, and $P(x, y)$ is the joint probabilities for word *x* and *y* to co-occur in a pre-defined text window. Both can be acquired by simply counting the term frequencies.

2.1 Viewing Term Co-occurrence as a Graph

The co-occurrence of possible translation terms within a given corpus can be viewed as a *graph*. In such a graph each translation candidate of a source query term is represented by a single *node*. Edges drawn between these nodes are weighted according to a particular co-occurrence measurement. Figure 1 illustrates a graph view of the possible translation candidates for a sample five term query. In this diagram, translation candidates for the same query term have been grouped together, and only the candidates with the highest co-occurrence scorings have been connected.

Viewing the co-occurrence of possible translation terms within a given corpus in this way has one particular advantage. It converts simple concurrence statistics into a pattern susceptible to various analytical techniques pioneered in the field of information retrieval. In the next section we will explain how we have adapted popular graph based algorithms such as HITS [7] and Page Rank [1] to enable the efficient resolution of translation ambiguities.

2.2 Graph Analysis Algorithm

The G^{con} algorithm for determining the most appropriate translation for a given query term, using a term co-occurrence graph, is as follows:

1. Given a query *C* in a source language contains several terms $\{c_1, c_2, \cdots, c_n\}$, for each term $c_i, i \in [1, n]$ in *C*, obtain the translation candidates $T(c_i) = \{t_{i,1}, t_{i,2}, \cdots, t_{i,m}\}$

2. A*ll* possible translation candidates of the query terms are generated, to form a undirected weighted graph: *G=<F, W>*, where *F* is the set of vertices representing one translation candidate $t_{i,j}$ to the query term c_i, and *W* is a complete set of *weighting functions*. Hence, every possible pairing of translation candidates sets has a non-negative weight attribute, *w*, which indicates the probable strength of any link potential between them. The set of weights as whole can be described as:

$$W : F \times F \rightarrow \{w \in R : w \geq 0\}$$

An individual weighting between two translation candidates $t_{i,j}$ and $t_{k,l}$ is given by the function:

$$w(t_{i,j} \leftrightarrow t_{k,l})$$

(n.b. In G^{con} each of these weights are calculated using co-occurrence statistics, but within this remit there are still two possible approaches, as discussed in Section 2.4)

3. For each translation candidate, $t_{i,j}$, compute the *Centrality Score* and in order *to* determine: $Cen(t_{i,j})$ for every single translation candidate in the graph (this process is discussed in more detail in section 2.4).

4. The translation of a query term is then determined by selecting the translation candidate, $t_{i,j}$, which produces the max *Centrality Score* in the correspondent set of translation candidates:

$$t(c_i) = \underset{t_{i,j} \in T(c_i)}{Max}(t_{i,j})$$

2.3 Significance of the Centrality Score

Graph-based analysis is essentially a technique for deciding the importance of a single node using global information recursively drawn from the entire graph. In our formulation of this approach, which is tailored to candidate translation selection, the importance of a node within the graph is described as its 'centrality'. When constructing the graph, edges are drawn between two nodes if and only if the co-occurrence scores for the two terms represented by those nodes are non-zero. The more edges connecting a single node to the rest of the graph, the greater the likelihood that the term that particular node represents is a 'good' translation of a given query term.

This can be stated formally as follows: for the given directed graph $G' = <V, E>$, where *V* denotes set of nodes and *E* denotes set of edges, for a given V_i let $\{V_i\}_{IN}$ be a set of nodes that point to it and let $\{V_i\}_{OUT}$ be a set of nodes that V_i points to, then, the centrality score of V_i is defined as follows:

$$Cen(V_i) = (1-d)/N + d \times \sum_{j \in \{V_i\}_{IN}} \frac{w_{i,j}}{\sum_{V_k \in \{V_j\}_{OUT}} w_{j,k}} Cen(V_j)$$

Where *d* is a dampening factor which integrates the probability of jumping from one node to another at random[3] (normally set to 0.85) and *N* is the total number of nodes in the graph. Starting with an arbitrary value assigned to every node in the graph, this algorithm is guaranteed to iterative until convergence below a certain threshold[4].

This is, perhaps, a relatively new area for graph based analysis. This class of technique was initially developed for use with *directed graphs*, most notably those graphs representing the associative structure of a hyperlinked corpus. A graph of co-occurrence phenomena is of course an *undirected graph* – similarity scores, which are signified by connecting edges, have no inherent direction and apply to the two connecting nodes equally. However, with some minor modifications to the basic algorithm, this lack of directionality poses no serious obstacle.

2.4 Matrix Notation

By using the matrix notation and viewing the whole process as a Markov chain, we can re-formulate these definitions. Let **A** be the similarity matrix where rows and columns represent the translation candidates and each entry denotes the similarity score between them. Note that the similarity scores between translation candidates for the same query term are initially set to zero. Let **z**

---

[3] This dampening factor is an adapted version of the "random surfer model" employed by Brin and Page [1] in the context of the PageRank algorithm.

[4] This is guaranteed because the Markov chain here is irreducible and aperiodic.

Table 1: Overview of Test Collections

CLEF 2006 (English, UTF-8)		CLEF 2007 (English, UTF-8)		NTCIR 5 (Chinese, Big5)	
Source	Num of Docs	Source	Num of Docs	Source	Num of Docs
Los Angeles Times 1994	113,005	Los Angeles Times 2002	135,153	United Daily News 2001-02	466,564
Glasgow Herald 1995	56,472			United Express 2001-02	92,296
				Ming Hseng News 2001-02	169,739
				Economic Daily News 2001-02	172,847
Total	169,477	Total	135,153	Total	901,446

be the centrality vector that corresponds to the stationary distribution of **A**, and let **U** be a square matrix with all elements being equal to *1/N*. The equation above can be written in the matrix form as

$$z = [(1-d)U + dA]^T z$$

The transition kernel $[(1-d)U + dA]$ of the resulting Markov chain is a mixture of two kernels **U** and **A**. A random walker on this Markov chain chooses one of the adjacent states of the current state with probability *d*, or jumps to any state in the graph, including the current state, with probability *1-d*.

2.5 The Weighting Function

In the graph based analysis of hypermedia structure, it is common to assign nodes in the graph a weighting. Thus far, the algorithm has assumed an unweighted graph. However, it is useful to indicate and incorporate into the model the "strength" of the link between two nodes into the co-occurrence graph. Therefore, a weight *w* is added to each edge that connects a pair of nodes.

Two variations of a weighting function have been developed: $w(t_{i,j} \leftrightarrow t_{k,l})$, which are called *StrengthWeighting* (*SW*) and *FixedWeighting* (*FW*) respectively. The *SW* function can be reviewed as an undirected weighted graph calculation while *FW* function is the normal unweighted version:

StrengthWeighting: If the similarity score between two terms is more than zero, the weights between the two terms are set as the similarity score, or else the weights are set to zero.

$$w(t_{i,j} \leftrightarrow t_{k,l}) = \langle \begin{array}{ll} sim(t_{i,j}, t_{k,l}) & sim(t_{i,j}, t_{k,l}) > 0 \\ 0 & otherwise \end{array}$$

FixedWeighting: If the similarity score between two terms is more than zero the weights between the two terms are set as one, otherwise they are set as zero.

$$w(t_{i,j} \leftrightarrow t_{k,l}) = \langle \begin{array}{ll} 1 & sim(t_{i,j}, t_{k,l}) > 0 \\ 0 & otherwise \end{array}$$

3 EXPERIMENTAL DESIGN

The purpose of the experiment described below is to evaluate the G^{con} algorithm. Our evaluation will focus on the following two questions:

- *Is the G^{con} method effective when used in the context of cross language information retrieval?* To answer this question, we examine the performance of the proposed algorithm as part of a Chinese-English /English-Chinese cross language retrieval experiment.
- *Is the G^{con} approach a significant improvement over simpler techniques designed to resolve translation ambiguities?* To answer this question we conduct a variety of experiments using the NTCIR and CLEF data sets comparing the G^{con} algorithm to less sophisticated approaches which exploit co-occurrence data.

3.1 Text Collection

The text corpus used in our experiment was made up from elements of the CLEF-2006, CLEF-2007 collections (Chinese queries against English documents) and the NTCIR-5 collection (English queries against Chinese documents). These collections are described in greater detail in Table 1.

All of the documents in the experiment were indexed using the Lemur toolkit[5]. Prior to indexing, Porter's stemmer and a stopwords list[6] were used for the English documents, and the Chinese documents were processed using a segmentation tool[7]. No further text processing was carried out.

3.2 Queries

We also used the CLEF-2006, CLEF-2007 and NTCIR-5 query sets. The query sets consist of 50 topics in Chinese, Chinese and English respectively, all of which were used in the experiment. Each topic is composed of several parts such as: *Title, Description, Narrative*. We chose to conduct *Title* runs as short queries and *Title+Description* runs as long queries. The English language text parsed from these fields parsed from these fields was stemmed and all stop words were removed. The Chinese language text was

5 http://www.lemurproject.org

6 ftp://ftp.cs.cornell.edu/pub/smart/

7 Available from http://www.madarintools.com

segmented. The relevance judgments for the original queries are used as the relevance judgments for their translations. The NTCIR relevance judgments are divided into two categories: *rigid* (strictly relevant) and *relaxed* (not strictly relevant). We adopted both of them in our evaluation.

3.3 Testing Environments

The Chinese-English and English-Chinese bilingual dictionaries used in our experiments were provided by the Linguistic Data Consortium[8]. We ran queries against the document collections using the Lemur Toolkit.

3.4 Description of Experimental Runs

Several experimental retrieval runs were conducted as follows:

MONO (***monolingual***): This part of the experiment involved retrieving documents from the three different test collections using manually translated queries. These translations were provided by the NTCIR-5 and CLEF-2006/2007 workshop committees respectively.

ALLTRANS (***all translations***): This part of the experiment involved retrieving documents from the three different test collections with queries translated using a bilingual dictionary. In this case, *all* of the translations offered by the dictionary were used.

FIRSTONE (***first translations***): This part of the experiment involved retrieving documents from the three different test collections with queries translated using a bilingual dictionary. In this case, only the *first* entry in the dictionary was used. This typically equates to the most popular translation for that term.

COM (***co-occurrence translation***): This part of the experiment involved retrieving documents from the three different test collections with queries translated using a bilingual dictionary in combination with the co-occurrence algorithm described in section S.2. The target document collection was used to calculate the co-occurrence scorings needed for this operation.

GCONW (*weighted graph analysis*): This part of the experiment involved retrieving documents from the three different test collections with queries translated using a bilingual dictionary in combination with our analysis of a *weighted* co-occurrence graph (i.e. we used the *SW* weighting function).

GCONUW (*unweighted graph analysis*): As above, we retrieved documents from the collections using query translations suggested by our analysis of the co-occurrence graph, only this time we used an *unweighted graph* (i.e. we used the *FW* weighting function).

A discussion of the results obtained can be found below.

3.5 Results and Discussion

As illustrated by the results given in tables 2 and 3, document retrieval with no disambiguation of the candidate translations (ALLTRANS) was consistently the lowest performer in terms of mean average precision. With just one exception, using the first translation offered by a bilingual dictionary (FIRSTONE) always led to an improvement in retrieval effectiveness when compared to the ALLTRANS scorings. However, the magnitude of this improvement was uneven across the three test collections, an anomaly which may be related to the bilingual dictionaries that were used. The English-Chinese dictionary provided by LDC contains more translation alternatives than the Chinese-English counterpart, thereby creating greater scope for ambiguity.

When the translation for each query term was selected using a basic co-occurrence model (COM), retrieval effectiveness always exceeded ALLTRANS and FIRSTONE. Interestingly, this result is inconsistent with earlier work published by [8] observing the opposite trend in the context of a TREC retrieval experiment. With the CLEF collection, a greater improvement was observed for longer queries when compared to shorter queries. However, the exact opposite was found to be the case when using the NTCIR collection. Again, this phenomenon seems to be related to the different characteristics of the bilingual dictionaries being used in the experiment.

The empirical results obtained using G^{con} technique are very promising. A weighted co-occurrence graph (GCONW) outperformed the basic co-occurrence model in five retrieval runs, exceeding the COM baseline by between 0.15% to 6.24%. However, the unweighted co-occurrence graph proved less satisfactory, recording results inferior to the basic co-occurrence model.

4 RELATED WORK

The resolution of ambiguities arising from the process of translation represents a classic problem for cross language information retrieval. Many of the successful solutions proposed to date have centred upon the use of co-occurrence statistics extracted from external resources to resolve ambiguities encountered during translation [4, 5, 6, 8, 9, 14]. In this vein, Gao *et al*. provide an excellent description of the basic co-occurrence model within a discussion of the use of decaying factors for mutual information measures [5]. Readers are also referred to the work of Liu *et al*. [8], who provide a further variation on a theme.

The G^{con} technique extends this approach by considering the network of term co-occurrence as a graph susceptible to recursive analysis. This approach was inspired by the current popularity of techniques based on graphs induced by implicit relationships between documents or other linguistic items [3, 10, 14, 16]. The work of Erkan *et al*. [3], who used graph based analysis to improve the salience of text summarizations, was particularly influential, as was the key word extraction algorithms of Michalcea [10]. The work of Monz & Dorr's [14], which describes the use of an iterative machine learning disambiguation strategy based on co-occurrence statistics, also provided further instruction.

This work also shares strong commonalities with research investigating graph based algorithms for large scale word sense disambiguation (WSD) [11, 12]. In this particular field, each node within the graph structure represents one possible sense of a given word sequence, while the edges connecting the nodes correspond to semantic dependencies between those interpretations (e.g. antonymy, synomy). The process of determining the correct sense for a given word sequence exploits these semantic dependencies to calculate the most 'important' node, thereby resolving the ambiguous sequence[9]. The algorithm described in this paper, which selects the most appropriate translation for a given term using recursive graph based analysis, could be viewed as a logical

[8] http://www.ldc.upenn.edu/

[9] See [2, 13] for alternative graph based approaches which determine word sense using similarity based algorithms dependent on context.

extension to this technique, with equal application to any process requiring "broad-coverage language understanding" [15].

5 CONCLUSIONS

In this paper we have introduced G^{con}, a computationally inexpensive graph-based model for the resolution of translation ambiguity. We evaluated this model in the context of a cross language retrieval experiment. The weighted G^{con} model fared well in our experiments, outperforming a baseline system reliant upon basic co-occurrence measures.

Use of the NTCIR and CLEF document collections in tandem led to some interesting observations. There seems to be a distinct difference between these two collections and the TREC alternatives commonly used by researchers in this field (*TREC 5/6*). Historically, the use of co-occurrence information to aid disambiguation has led to disappointing results on TREC retrieval runs [8]. This was certainly not the case for the NTCIR and CLEF document collections. A suggested avenue for future work may involve a side by side examination of all three document collections to identify those characteristics indicating an affinity for co-occurrence based resolution techniques.

6 ACKNOWLEDGEMENTS

We are most grateful to members of the University of Nottingham Web Technology Group for many useful discussions. This work was supported by a scholarship from the University of Nottingham and the Hunan University.

7 REFERENCES

[1] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the Seventh international Conference on World Wide Web 7 (Brisbane, Australia). P. H. Enslow and A. Ellis, Eds. Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 107-117.

[2] Brody, S., Navigli, R., and Lapata, M. 2006. Ensemble methods for unsupervised WSD. In Proceedings of the ACL/COLING, Sydney, Australia.

[3] Erkan, G. and Radev, D.R. 2004. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. Journal of Artificial Intelligence. 457-479.

[4] Gao, J. and Nie, J. 2006. A study of statistical models for query translation: finding a good unit of translation. In Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Seattle, Washington, USA, August 06 - 11, 2006). SIGIR '06. ACM Press, New York, NY, 194-201.

[5] Gao, J., Zhou, M., Nie, J., He, H., and Chen, W. 2002. Resolving query translation ambiguity using a decaying co-occurrence model and syntactic dependence relations. In Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Tampere, Finland, August 11 - 15, 2002). SIGIR '02. ACM Press, New York, NY, 183-190.

[6] Jang, M., Myaeng, S. H., and Park, S. Y. 1999. Using mutual information to resolve query translation ambiguities and query term weighting. In Proceedings of the 37th Annual Meeting of the Association For Computational Linguistics on Computational Linguistics (College Park, Maryland, June 20 - 26, 1999). Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, 223-229.

[7] Kleinberg, J.M. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM. 46(5): 604-632.

[8] Liu, Y., Jin, R., and Chai, J. Y. 2005. A maximum coherence model for dictionary-based cross-language information retrieval. In Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Salvador, Brazil, August 15 - 19, 2005). SIGIR '05. ACM Press, New York, NY, 536-543.

[9] Maeda, A., Sadat, F., Yoshikawa, M., and Uemura, S. 2000. Query term disambiguation for Web cross-language information retrieval using a search engine. In Proceedings of the Fifth international Workshop on on information Retrieval with Asian Languages (Hong Kong, China, September 30 - October 01, 2000). IRAL '00. ACM Press, New York, NY, 25-32.

[10] Mihalcea, R. and Tarau, P. 2004.. Textrank: Bringing order into texts. In L. Dekang and W. Dekai, editors, Proceedings of EMNLP 2004, Barcelona, Spain, July 2004. Association for Computational Linguistics. 404-411.

[11] Mihalcea, R. 2005. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing Association for Computational Linguistics. 411-418.

[12] Mihalcea, R., Tarau, P., and Figa, E. 2004. PageRank on semantic networks, with application to word sense disambiguation. In Proceedings of the 20th international Conference on Computational Linguistics Association for Computational Linguistics, 1126.

[13] Mohammad, S. and Hirst, G. 2006. Determining word sense dominance using a thesaurus. In Proceedings of the 11th EACL, Trento, Italy. 121-128.

[14] Monz, C. and Dorr, B. J. 2005. Iterative translation disambiguation for cross-language information retrieval. In Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Salvador, Brazil, August 15 - 19, 2005). SIGIR '05. ACM Press, New York, NY, 520-527.

[15] Navigli, R. and Lapata, M. 2007. Graph Connectivity Measures for Unsupervised Word Sense Disambiguation. Proceedings of the 20th International Joint Conference on Artificial Intelligence, 1683-1688.

[16] Zhou, D., Goulding, J., Truran, M., and Brailsford, T. 2007. LLAMA: automatic hypertext generation utilizing language models. Proceedings of the 18th conference on Hypertext and hypermedia, Manchester, UK. 77-80.

Table 2. Experimental Results for the CLEF test collection
Measured in Mean Average Precision, R-prec and Precision at 10.

	MAP	R-Prec	P@10	% of MONO	Improvement over ALLTRANS	Improvement over FIRSTONE	Improvement over COM
Short Query for CLEF 2006							
MONO	0.3468	0.3474	0.386	N/A	N/A	N/A	N/A
ALLTRANS	0.2017	0.2	0.232	58.16%	N/A	N/A	N/A
FIRSTONE	0.214	0.2116	0.226	61.71%	6.10%	N/A	N/A
COM	0.2166	0.2163	0.238	62.46%	7.39%	1.21%	N/A
GCONW	0.2166	0.2163	0.238	62.46%	7.39%	1.21%	0.00%
GCONUW	0.2249	0.2218	0.242	64.85%	11.50%	5.09%	3.83%
Long Query for CLEF 2006							
MONO	0.3403	0.3384	0.398	N/A	N/A	N/A	N/A
ALLTRANS	0.2368	0.2511	0.262	69.59%	N/A	N/A	N/A
FIRSTONE	0.2603	0.2656	0.276	76.49%	9.92%	N/A	N/A
COM	0.2692	0.2729	0.282	79.11%	13.68%	3.42%	N/A
GCONW	0.2707	0.2762	0.282	79.55%	14.32%	4.00%	0.56%
GCONUW	0.2619	0.2659	0.282	76.96%	10.60%	0.61%	-2.71%
Short Query for CLEF 2007							
MONO	0.4078	0.4019	0.486	N/A	N/A	N/A	N/A
ALLTRANS	0.2567	0.2558	0.304	62.95%	N/A	N/A	N/A
FIRSTONE	0.2638	0.2555	0.284	64.69%	2.77%	N/A	N/A
COM	0.2645	0.2617	0.306	64.86%	3.04%	0.27%	N/A
GCONW	0.2645	0.2617	0.306	64.86%	3.04%	0.27%	0.00%
GCONUW	0.2711	0.2619	0.294	66.48%	5.61%	2.77%	2.50%
Long Query for CLEF 2007							
MONO	0.3753	0.3806	0.43	N/A	N/A	N/A	N/A
ALLTRANS	0.2671	0.2778	0.346	71.17%	N/A	N/A	N/A
FIRSTONE	0.2516	0.2595	0.286	67.04%	-5.80%	N/A	N/A
COM	0.2748	0.2784	0.322	73.22%	2.88%	9.22%	N/A
GCONW	0.2748	0.2784	0.322	73.22%	2.88%	9.22%	0.00%
GCONUW	0.2606	0.2714	0.286	69.44%	-2.43%	3.58%	-5.17%

Table 3. Experimental Results for the NTCIR test collection Measured in Mean Average Precision, R-prec and Precision at 10.

	MAP	R-Prec	P@10	% of MONO	Improvement over ALLTRANS	Improvement over FIRSTONE	Improvement over COM
Short Query for NTCIR5-rigid							
MONO	0.3692	0.3601	0.416	N/A	N/A	N/A	N/A
ALLTRANS	0.0704	0.0828	0.09	19.07%	N/A	N/A	N/A
FIRSTONE	0.1004	0.1012	0.126	27.19%	42.61%	N/A	N/A
COM	0.1302	0.1333	0.158	35.27%	84.94%	29.68%	N/A
GCONW	0.1308	0.1332	0.156	35.43%	85.80%	30.28%	0.46%
GCONUW	0.1132	0.1133	0.132	30.66%	60.80%	12.75%	-13.06%
Long Query for NTCIR5-rigid							
MONO	0.3494	0.3523	0.416	N/A	N/A	N/A	N/A
ALLTRANS	0.0697	0.0736	0.088	19.95%	N/A	N/A	N/A
FIRSTONE	0.119	0.1239	0.132	34.06%	70.73%	N/A	N/A
COM	0.1428	0.1509	0.166	40.87%	104.88%	20.00%	N/A
GCONW	0.1483	0.153	0.172	42.44%	112.77%	24.62%	3.85%
GCONUW	0.1099	0.1204	0.138	31.45%	57.68%	-7.65%	-23.04%
Short Query for NTCIR5-relax							
MONO	0.4294	0.4	0.534	N/A	N/A	N/A	N/A
ALLTRANS	0.0787	0.0907	0.124	18.33%	N/A	N/A	N/A
FIRSTONE	0.1174	0.1203	0.166	27.34%	49.17%	N/A	N/A
COM	0.1298	0.1375	0.178	30.23%	64.93%	10.56%	N/A
GCONW	0.13	0.1369	0.176	30.27%	65.18%	10.73%	0.15%
GCONUW	0.1311	0.1341	0.168	30.53%	66.58%	11.67%	1.00%
Long Query for NTCIR5-relax							
MONO	0.4032	0.3921	0.546	N/A	N/A	N/A	N/A
ALLTRANS	0.0618	0.0706	0.098	15.33%	N/A	N/A	N/A
FIRSTONE	0.1377	0.144	0.186	34.15%	122.82%	N/A	N/A
COM	0.1507	0.1635	0.204	37.38%	143.85%	9.44%	N/A
GCONW	0.1601	0.1737	0.228	39.71%	159.06%	16.27%	6.24%
GCONUW	0.13	0.1474	0.186	32.24%	110.36%	-5.59%	-13.74%