Proceedings of HLT-NAACL 2003
Main Papers , pp. 48-55
Edmonton, May-June 2003

An Analysis of Clarification Dialogue for Question Answering

Marco De Boni

School of Computing
Leeds Metropolitan University
Leeds LS6 3QS, UK
Department of Computer Science
University of York
York Y010 5DD, UK
mdeboni@cs.york.ac.uk

Suresh Manandhar

Department of Computer Science
University of York
York Y010 5DD, UK
suresh@cs.york.ac.uk

Abstract

We examine clarification dialogue, a mechanism for refining user questions with follow-up questions, in the context of open domain Question Answering systems. We develop an algorithm for clarification dialogue recognition through the analysis of collected data on clarification dialogues and examine the importance of clarification dialogue recognition for question answering. The algorithm is evaluated and shown to successfully recognize the occurrence of clarification dialogue in the majority of cases and to simplify the task of answer retrieval.

1 Clarification dialogues in Question Answering

Question Answering Systems aim to determine an answer to a question by searching for a response in a collection of documents (see Voorhees 2002 for an overview of current systems). In order to achieve this (see for example Harabagiu et al. 2002), systems narrow down the search by using information retrieval techniques to select a subset of documents, or paragraphs within documents, containing keywords from the question and a concept which corresponds to the correct question type (e.g. a question starting with the word "Who?" would require an answer containing a person). The exact answer sentence is then sought by either attempting to unify the answer semantically with the question, through some kind of logical transformation (e.g. Moldovan and Rus 2001) or by some form of pattern matching (e.g. Soubbotin 2002; Harabagiu et al. 1999).

Often, though, a single question is not enough to meet user's goals and an elaboration or clarification dialogue is required, i.e. a dialogue with the user which would enable the answering system to refine its understanding of the questioner's needs (for reasons of space we shall not investigate here the difference between elaboration dialogues, clarification dialogues and coherent topical subdialogues and we shall hence refer to this type of dialogue simply as "clarification dialogue", noting that this may not be entirely satisfactory from a theoretical linguistic point of view). While a number of researchers have looked at clarification dialogue from a theoretical point of view (e.g. Ginzburg 1998; Ginzburg and Sag 2000; van Beek at al. 1993), or from the point of view of task oriented dialogue within a narrow domain (e.g. Ardissono and Sestero 1996), we are not aware of any work on clarification dialogue for open domain question answering systems such as the ones presented at the TREC workshops, apart from the experiments carried out for the (subsequently abandoned) "context" task in the TREC-10 QA workshop (Voorhees 2002; Harabagiu et al. 2002). Here we seek to partially address this problem by looking at some particular aspect of clarification dialogues in the context of open domain question answering. In particular, we examine the problem of recognizing that a clarification dialogue is occurring, i.e. how to recognize that the current question under consideration is part of a previous series (i.e. clarifying previous questions) or the start of a new series; we then show how the recognition that a clarification dialogue is occurring can simplify the problem of answer retrieval.

2 The TREC Context Experiments

The TREC-2001 QA track included a "context" task which aimed at testing systems' ability to track context through a series of questions (Voorhees 2002). In other words, systems were required to respond correctly to a kind of clarification dialogue in which a full understanding of questions depended on an understanding of previous questions. In order to test the ability to answer such questions correctly, a total of 42 questions were prepared by NIST staff, divided into 10 series of related question sentences which therefore constituted a type of clarification dialogue; the sentences varied in length between 3 and 8 questions, with an average of 4 questions per dialogue. These clarification dialogues were however presented to the question answering systems already classified and hence systems did not need to recognize that clarification was actually taking place. Consequently systems that simply looked for an answer in the subset of documents retrieved for the first question in a series performed well without any understanding of the fact that the questions constituted a coherent series.

In a more realistic approach, systems would not be informed in advance of the start and end of a series of clarification questions and would not be able to use this information to limit the subset of documents in which an answer is to be sought.

3 Analysis of the TREC context questions

We manually analysed the TREC context question collection in order to determine what features could be used to determine the start and end of a question series, with the following conclusions:

- Pronouns and possessive adjectives: questions such as "*When was it born?*", which followed "*What was the first transgenic mammal?*", were referring to some previously mentioned object through a pronoun ("it"). The use of personal pronouns ("he", "it", …) and possessive adjectives ("his", "her",…) which did not have any referent in the question under consideration was therefore considered an indication of a clarification question..
- Absence of verbs: questions such as *"On what body of water?"* clearly referred to some previous question or answer.
- Repetition of proper nouns: the question series starting with "*What type of vessel was the modern Varyag?*" had a follow-up question "*How long was the Varyag?*", where the repetition of the proper noun indicates that the same subject matter is under investigation.
- Importance of semantic relations: the first question series started with the question "*Which museum in Florence was damaged by a major bomb explosion?*"; follow-up questions included "*How many people were killed?*" and "*How much explosive was used?*", where there is a clear semantic relation between the "explosion" of the initial question and the "killing" and "explosive" of the following questions. Questions belonging to a series were "about" the same subject, and this aboutness could be seen in the use of semantically related words.

4 Experiments in Clarification Dialogue Recognition

It was therefore speculated that an algorithm which made use of these features would successfully recognize the occurrence of clarification dialogue. Given that the only available data was the collection of "context" questions used in TREC-10, it was felt necessary to collect further data in order to test our algorithm rigorously. This was necessary both because of the small number of questions in the TREC data and the fact that there was no guarantee that an algorithm built for this dataset would perform well on "real" user questions. A collection of 253 questions was therefore put together by asking potential users to seek information on a particular topic by asking a prototype question answering system a series of questions, with "cue" questions derived from the TREC question collection given as starting points for the dialogues. These questions made up 24 clarification dialogues, varying in length from 3 questions to 23, with an average length of 12 questions (the data is available from the main author upon request).

The differences between the TREC "context" collection and the new collection are summarized in the following table:

	Groups	Qs	Av. len	Max	Min
TREC	10	41	4	8	4
New	24	253	12	23	3

The questions were recorded and manually tagged to recognize the occurrence of clarification dialogue.

The questions thus collected were then fed into a system implementing the algorithm, with no indication as to where a clarification dialogue occurred. The system then attempted to recognize the occurrence of a clarification dialogue. Finally the results given by the system were compared to the manually recognized clarification dialogue tags. In particular the algorithm was evaluated for its capacity to:

- recognize a new series of questions (i.e. to tell that the current question is not a clarification of any previous question) (indicated by New in the results table)

- recognize that the current question is clarifying a previous question (indicated by Clarification in the table)

5 Clarification Recognition Algorithm

Our approach to clarification dialogue recognition looks at certain features of the question currently under consideration (e.g. pronouns and proper nouns) and compares the meaning of the current question with the meanings of previous questions to determine whether they are "about" the same matter.

Given a question q_0 and *n* previously asked questions $q_{-1}..q_{-n}$ we have a function Clarification_Question which is true if a question is considered a clarification of a previously asked question. In the light of empirical work such as (Ginzburg 1998), which indicates that questioners do not usually refer back to questions which are very distant, we only considered the set of the previously mentioned 10 questions.

A question is deemed to be a clarification of a previous question if:

1. There are direct references to nouns mentioned in the previous *n* questions through the use of pronouns (he, she, it, …) or possessive adjectives (his, her, its…) which have no references in the current question.
2. The question does not contain any verbs
3. There are explicit references to proper and common nouns mentioned in the previous *n* questions, i.e. repetitions which refer to an identical object; or there is a strong sentence similarity between the current question and the previously asked questions.

In other words:

Clarification_Question (q_n,$q_{-1}..q_{-n}$)
is true if
1. q_0 has pronoun and possessive adjective references to $q_{-1}..q_{-n}$
2. q_0 does not contain any verbs
3. q_0 has repetition of common or proper nouns in $q_{-1}..q_{-n}$ or q_0 has a strong semantic similarity to some $q \in q_{-1}..q_{-n}$

6 Sentence Similarity Metric

A major part of our clarification dialogue recognition algorithm is the sentence similarity metric which looks at the similarity in meaning between the current question and previous questions. WordNet (Miller 1999; Fellbaum 1998), a lexical database which organizes words into synsets, sets of synonymous words, and specifies a number of relationships such as hypernym, synonym, meronym which can exist between the synsets in the lexicon, has been shown to be fruitful in the calculation of semantic similarity. One approach has been to determine similarity by calculating the length of the path or relations connecting the words which constitute sentences (see for example Green 1997 and Hirst and St-Onge 1998); different approaches have been proposed (for an evaluation see (Budanitsky and Hirst 2001)), either using all WordNet relations (Budanitsky and Hirst 2001) or only is-a relations (Resnik 1995; Jiang and Conrath 1997; Mihalcea and Moldvoan 1999). Miller (1999), Harabagiu et al. (2002) and De Boni and Manandhar (2002) found WordNet glosses, considered as micro-contexts, to be useful in determining conceptual similarity. (Lee et al. 2002) have applied conceptual similarity to the Question Answering task, giving an answer *A* a score dependent on the number of matching terms in *A* and the question. Our sentence similarity measure followed on these ideas, adding to the use of WordNet relations, part-of-speech information, compound noun and word frequency information.

In particular, sentence similarity was considered as a function which took as arguments a sentence s_1 and a second sentence s_2 and returned a value representing the semantic relevance of s_1 in respect of s_2 in the context of knowledge *B*, i.e.

$$\text{semantic-relevance}(s_1, s_2, B) = n \in \mathbb{R}$$

semantic-relevance(s_1,*s*,*B*) < *semantic-relevance*(s_2,*s*, *B*) represents the fact that sentence s_1 is less relevant than s_2 in respect to the sentence *s* and the context *B*. In our experiments, *B* was taken to be the set of semantic relations given by WordNet. Clearly, the use of a different knowledge base would give different results, depending on its completeness and correctness.

In order to calculate the semantic similarity between a sentence s_1 and another sentence s_2, s_1 and s_2 were considered as sets *P* and *Q* of word stems. The similarity between each word in the question and each word in the answer was then calculated and the sum of the closest matches gave the overall similarity. In other words, given two sets *Q* and *P*, where $Q=\{qw_1,qw_2,\ldots,qw_n\}$ and $P=\{pw_1,pw_2,\ldots,pw_m\}$, the similarity between *Q* and *P* is given by

$$\sum_{1<p<n} \text{Argmax}_m \text{ similarity}(qw_p, pw_m)$$

The function *similarity*(w_1, w_2) maps the stems of the two words w_1 and w_2 to a similarity measure m representing how semantically related the two words are; *similarity*(w_i, w_j)< *similarity*(w_i, w_k) represents the fact that the word w_j is less semantically related than w_k in respect to the word w_i. In particular *similarity*=0 if two words are not at all semantically related and *similarity*=1 if the words are the same.

$$similarity(w_1, w_2) = h \in \mathbb{R}$$

where $0 \leq h \leq 1$. In particular, *similarity*(w_1, w_2) = 0 if $w_1 \in ST \vee w_2 \in ST$, where ST is a set containing a number of stop-words (e.g. "the", "a", "to") which are too common to be able to be usefully employed to estimate semantic similarity. In all other cases, h is calculated as follows: the words w_1 and w_2 are compared using all the available WordNet relationships (is-a, satellite, similar, pertains, meronym, entails, etc.), with the additional relationship, "same-as", which indicated that two words were identical. Each relationship is given a weighting indicating how related two words are, with a "same as" relationship indicating the closest relationship, followed by synonym relationships, hypernym, hyponym, then satellite, meronym, pertains, entails.

So, for example, given the question "Who went to the mountains yesterday?" and the second question "Did Fred walk to the big mountain and then to mount Pleasant?", Q would be the set {who, go, to, the, mountain, yesterday} and P would be the set {Did, Fred, walk, to, the, big, mountain, and, then, to, mount, Pleasant}.

In order to calculate similarity the algorithm would consider each word in turn. "Who" would be ignored as it is a common word and hence part of the list of stop-words. "Go" would be related to "walk" in a is-a relationship and receive a score h_1. "To" and "the" would be found in the list of stop-words and ignored. "Mountain" would be considered most similar to "mountain" (same-as relationship) and receive a score h_2: "mount" would be in a synonym relationship with "mountain" and give a lower score, so it is ignored. "Yesterday" would receive a score of 0 as there are no semantically related words in Q. The similarity measure of Q in respect to P would therefore be given by $h_1 + h_2$.

In order to improve performance of the similarity measure, additional information was considered in addition to simple word matching (see De Boni and Manandhar 2003 for a complete discussion):

- *Compound noun information*. The motivation behind is similar to the reason for using chunking information, i.e. the fact that the word "United" in "United States" should not be considered similar to "United" as in "Manchester United". As opposed to when using chunking information, however, when using noun compound information, the compound is considered a single word, as opposed to a group of words: chunking and compound noun information may therefore be combined as in "[the [United States] official team]".
- *Proper noun information*. The intuition behind this is that titles (of books, films, etc.) should not be confused with the "normal" use of the same words: "blue lagoon" as in the sentence "the film Blue Lagoon was rather strange" should not be considered as similar to the same words in the sentence "they swan in the blue lagoon" as they are to the sentence "I enjoyed Blue Lagoon when I was younger".
- *Word frequency information*. This is a step beyond the use of stop-words, following the intuition that the more a word is common the less it is useful in determining similarity between sentence. So, given the sentences "metatheoretical reasoning is common in philosophy" and "metatheoretical arguments are common in philosophy", the word "metatheoretical" should be considered more important in determining relevance than the words "common", "philosophy" and "is" as it is much more rare and therefore less probably found in irrelevant sentences. Word frequency data was taken from the Given that the questions examined were generic queries which did not necessarily refer to a specific set of documents, the word frequency for individual words was taken to be the word frequency given in the British National Corpus (see BNCFreq 2003). The top 100 words, making up 43% of the English Language, were then used as stop-words and were not used in calculating semantic similarity.

7 Results

An implementation of the algorithm was evaluated on the TREC context questions used to develop the algorithm and then on the collection of 500 new clarification dialogue questions. The results on the TREC data, which was used to develop the algorithm, were as follows (see below for discussion and an explanation of each method):

TREC	Meth.0	Meth.1	Meth.2	Meth.3a	Meth.3b
New	90	90	90	60	80
Clarif.	47	53	59	78	72

Where “New” indicates the ability to recognize whether the current question is the first in a new series of clarification questions and “Clarif.” (for “Clarification”) indicates the ability to recognize whether the current question is a clarification question.

The results for the same experiments conducted on the collected data were as follows:

Collected	Meth.0	Meth.1	Meth.2	Meth.3a	Meth.3b
New	100	100	100	67	83
Clarif.	64	62	66	91	89

Method 0. This method did not use any linguistic information and simply took a question to be a clarification question if it had any words in common with the previous *n* questions, else took the question to be the beginning of a new series. 64% of questions in the new collection could be recognized with this simple algorithm, which did not misclassify any "new" questions.

Method 1. This method employed point 1 of the algorithm described in section 5: 62% of questions in the new collection could be recognized as clarification questions simply by looking for "reference" keywords such as he, she, this, so, etc. which clearly referred to previous questions. Interestingly this did not misclassify any "new" questions.

Method 2. This method employed points 1 and 2 of the algorithm described in section 5: 5% of questions in the new collection could be recognized simply by looking for the absence of verbs, which, combined with keyword lookup (Method 1), improved performance to 66%. Again this did not misclassify any "new" questions.

Method 3a. This method employed the full algorithm described in section 5 (point 3 is the similarity measure algorithm described in section 6): clarification recognition rose to 91% of the new collection by looking at the similarity between nouns in the current question and nouns in the previous questions, in addition to reference words and the absence of verbs. Misclassification was a serious problem, however with correctly classified "new" questions falling to 67%.

Method 3b. This was the same as method 3a, but specified a similarity threshold when employing the similarity measure described in section 6: this required the nouns in the current question to be similar to nouns in the previous question beyond a specified similarity threshold. This brought clarification question recognition down to 89% of the new collection, but misclassification of "new" questions was reduced significantly, with "new" questions being correctly classified 83% of the time.

Problems noted were:

- False positives: questions following a similar but unrelated question series. E.g. "Are they all Muslim countries?" (talking about religion, but in the context of a general conversation about Saudi Arabia) followed by "What is the chief religion in Peru?" (also about religion, but in a totally unrelated context).
- Questions referring to answers, not previous questions (e.g. clarifying the meaning of a word contained in the answer, or building upon a concept defined in the answer: e.g. "What did Antonio Carlos Tobim play?" following "Which famous musicians did he play with?" in the context of a series of questions about Fank Sinatra: Antonio Carlos Tobim was referred to in the answer to the previous question, and nowhere else in the exchange. These made up 3% of the missed clarifications.
- Absence of relationships in WordNet, e.g. between "NASDAQ" and "index" (as in share index). Absence of verb-noun relationships in WordNet, e.g. between to die and death, between "battle" and "win" (i.e. after a battle one side generally wins and another side loses), "airport" and "visit" (i.e. people who are visiting another country use an airport to get there)

As can be seen from the tables above, the same experiments conducted on the TREC context questions yielded worse results; it was difficult to say, however, whether this was due to the small size of the TREC data or the nature of the data itself, which perhaps did not fully reflect “real” dialogues.

As regards the recognition of question in a series (the recognition that a clarification I taking place), the number of sentences recognized by keyword alone was smaller in the TREC data (53% compared to 62%), while the number of questions not containing verbs was roughly similar (about 6%). The improvement given by computing noun similarity between successive questions gave worse results on the TREC data: using method 3a resulted in an improvement to the overall correctness of 19 percentage points, or a 32% increase (compared to an improvement of 25 percentage points, or a 38% increase on the collected data); using method 3b resulted in an improvement of 13 percentage points, or a 22% increase (compared to an improvement of 23 percentage points or a 35% increase on the collected data), perhaps indicating that in "real" conversation speakers tend to use simpler semantic relationships than what was observed in the TREC data.

8 Usefulness of Clarification Dialogue Recognition

Recognizing that a clarification dialogue is occurring only makes sense if this information can then be used to improve answer retrieval performance.

We therefore hypothesized that noting that a questioner is trying to clarify previously asked questions is important in order to determine the context in which an answer is to be sought: in other words, the answers to certain questions are constrained by the context in which they have been uttered. The question "What does attenuate mean?", for example, may require a generic answer outlining all the possible meanings of "attenuate" if asked in isolation, or a particular meaning if asked after the word has been seen in an answer (i.e. in a definite context which constrains its meaning). In other cases, questions do not make sense at all out of a context. For example, no answer could be given to the question "where?" asked on its own, while following a question such as "Does Sean have a house anywhere apart from Scotland?" it becomes an easily intelligible query.

The usual way in which Question Answering systems constrain possible answers is by restricting the number of documents in which an answer is sought by filtering the total number of available documents through the use of an information retrieval engine. The information retrieval engine selects a subset of the available documents based on a number of keywords derived from the question at hand. In the simplest case, it is necessary to note that some words in the current question refer to words in previous questions or answers and hence use these other words when formulating the IR query. For example, the question "Is he married?" cannot be used *as is* in order to select documents, as the only word passed to the IR engine would be "married" (possibly the root version "marry") which would return too many documents to be of any use. Noting that the "he" refers to a previously mentioned person (e.g. "Sean Connery") would enable the answerer to seek an answer in a smaller number of documents. Moreover, given that the current question is asked in the context of a previous question, the documents retrieved for the previous related question could provide a context in which to initially seek an answer.

In order to verify the usefulness of constraining the set of documents from in which to seek an answer, a subset made of 15 clarification dialogues (about 100 questions) from the given question data was analyzed by taking the initial question for a series, submitting it to the Google Internet Search Engine and then manually checking to see how many of the questions in the series could be answered simply by using the first 20 documents retrieved for the first question in a series. The results are summarized in the following diagram (Fig. 1):

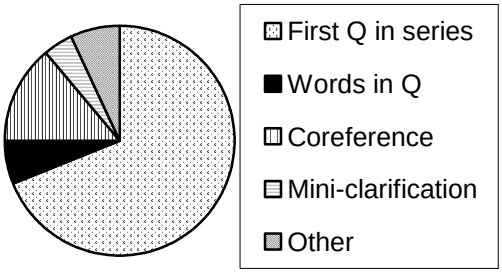

Fig. 1: Search technique used for Question

- 69% of clarification questions could be answered by looking within the documents used for the previous question in the series, thus indicating the usefulness of noting the occurrence of clarification dialogue.
- The remaining 31% could not be answered by making reference to the previously retrieved documents, and to find an answer a different approach had to be taken. In particular:
- 6% could be answered after retrieving documents simply by using the words in the question as search terms (e.g. "What caused the boxer uprising?");
- 14% required some form of coreference resolution and could be answered only by combining the words in the question with the words to which the relative pronouns in the question referred (e.g. "What film is he working on at the moment", with the reference to "he" resolved, which gets passed to the search engine as "What film is Sean Connery working on at the moment?");
- 7% required more than 20 documents to be retrieved by the search engine or other, more complex techniques. An example is a question such as "Where exactly?" which requires both an understanding of the context in which the question is asked ("Where?" makes no sense on its own) and the previously given answer (which was probably a place, but not restrictive enough for the questioner).
- 4% constituted mini-clarification dialogues within a larger clarification dialogue (a slight deviation from the main topic which was being investigated by the questioner) and could be answered by looking at the documents retrieved for the first question in the mini-series.

Recognizing that a clarification dialogue is occurring therefore can simplify the task of retrieving an answer by specifying that an answer must be in the set of documents used the previous questions. This is

consistent with the results found in the TREC context task (Voorhees 2002), which indicated that systems were capable of finding most answers to questions in a context dialogue simply by looking at the documents retrieved for the initial question in a series. As in the case of clarification dialogue recognition, therefore, simple techniques can resolve the majority of cases; nevertheless, a full solution to the problem requires more complex methods. The last case indicates that it is not enough simply to look at the documents provided by the first question in a series in order to seek an answer: it is necessary to use the documents found for a previously asked question which is related to the current question (i.e. the questioner could "jump" between topics). For example, given the following series of questions starting with Q_1:

Q_1: When was the Hellenistic Age?
[…]
Q_5: How did Alexander the great become ruler?
Q_6: Did he conquer anywhere else?
Q_7: What was the Greek religion in the Hellenistic Age?

where Q_6 should be related to Q_5 but Q_7 should be related to Q_1, and not Q_6. In this case, given that the subject matter of Q_1 is more immediately related to the subject matter of Q_7 than Q_6 (although the subject matter of Q_6 is still broadly related, it is more of a specialized subtopic), the documents retrieved for Q_1 will probably be more relevant to Q_7 than the documents retrieved for Q_6 (which would probably be the same documents retrieved for Q_5)

9 Conclusion

It has been shown that recognizing that a clarification dialogue is occurring can simplify the task of retrieving an answer by constraining the subset of documents in which an answer is to be found. An algorithm was presented to recognize the occurrence of clarification dialogue and is shown to have a good performance. The major limitation of our algorithm is the fact that it only considers series of questions, not series of answers. As noted above, it is often necessary to look at an answer to a question to determine whether the current question is a clarification question or not. Our sentence similarity algorithm was limited by the number of semantic relationships in WordNet: for example, a big improvement would come from the use of noun-verb relationships. Future work will be directed on extending WordNet in this direction and in providing other useful semantic relationships. Work also needs to be done on using information given by answers, not just questions in recognizing clarification dialogue and on coping with the cases in which clarification dialogue recognition is not enough to retrieve an answer and where other, more complex, techniques need to be used. It would also be beneficial to examine the use of a similarity function in which similarity decayed in function of the distance in time between the current question and the past questions.

References

Ardissono, L. and Sestero, D. 1996. "Using dynamic user models in the recognition of the plans of the user". *User Modeling and User-Adapted Interaction*, 5(2):157-190.

BNCFreq. 2003. *English Word Frequency List*. http://www.eecs.umich.edu/~qstout/586/bncfreq.html (last accessed March 2003).

Budanitsky, A., and Hirst, G. 2001. "Semantic distance in WordNet: and experimental, application-oriented evaluation of five measures", in *Proceedings of the NAACL 2001 Workshop on WordNet and other lexical resources*, Pittsburgh.

De Boni, M. and Manandhar, S. 2003. "The Use of Sentence Similarity as a Semantic Relevance Metric for Question Answering". *Proceedings of the AAAI Symposium on New Directions in Question Answering*, Stanford.

De Boni, M. and Manandhar, S. 2002. "Automated Discovery of Telic Relations for WordNet". *Proceedings of the First International WordNet Conference*, India.

Fellbaum, C. 1998. *WordNet, An electronic Lexical Databas*e, MIT Press.

Ginzburg , J. 1998. "Clarifying Utterances" In: J. Hulstijn and A. Nijholt (eds.) *Proceedings of the 2nd Workshop on the Formal Semantics and Pragmatics of Dialogue*, Twente.

Ginzburg and Sag, 2000. *Interrogative Investigations*, CSLI.

Green, S. J. 1997. *Automatically generating hypertext by computing semantic similarity*, Technical Report n. 366, University of Toronto.

Harabagiu, S., Miller, A. G., Moldovan, D. 1999. "WordNet2 - a morphologically and semantically enhanced resource", In *Proceedings of SIGLEX-99*, University of Maryland.

Harabagiu, S., et al. 2002. "Answering Complex, List and Context Questions with LCC's Question-Answering Server", *Proceedings of TREC-10*, NIST.

Hirst, G., and St-Onge, D. 1998. "Lexical chains as representations of context for the detection and

correction of malapropisms", in Fellbaum (ed.), *WordNet: and electronic lexical database*, MIT Press.

Jiang, J. J., and Conrath, D. W. 1997. "Semantic similarity based on corpus statistics and lexical taxonomy", in *Proceedings of ICRCL*, Taiwan.

Lee, G. G., et al. 2002. "SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP", *Proceedings of TREC-10*, NIST.

Lin, D. 1998. "An information-theoretic definition of similarity", in *Proceedings of the 15th International Conference on Machine Learning*, Madison.

Mihalcea, R. and Moldovan, D. 1999. "A Method for Word Sense Disambiguation of Unrestricted Text", in *Proceedings of ACL '99*, Maryland, NY.

Miller, G. A. 1999. "WordNet: A Lexical Database", *Communications of the ACM*, 38 (11).

Moldovan, D. and Rus, V. 2001. "Logic Form Transformation of WordNet and its Applicability to Question Answering", *Proceedings of the 39th conference of ACL*, Toulouse.

Resnik, P. 1995. "Using information content to evaluate semantic similarity", in *Proceedings of the 14th IJCAI*, Montreal.

Soubbotin, M. M. 2002. :"Patterns of Potential Answer Expressions as Clues to the Right Answers", *Proceedings of TREC-10*, NIST.

van Beek, P., Cohen, R. and Schmidt, K., 1993. "From plan critiquing to clarification dialogue for cooperative response generation", *Computational Intelligence* 9:132-154.

Voorhees, E. 2002. "Overview of the TREC 2001 Question Answering Track", *Proceedings of TREC-10*, NIST.