26

Answering Causal Questions and Developing Tool Support

Sodel Vazquez-Reyes and Perla Velasco-Elizondo
Autonomous University of Zacatecas, Centre for Mathematical Research
México

1. Introduction

People explore the world by asking questions about what is seen and felt. Thus, Question Answering is an attractive research area as a distinctive combination from a variety of disciplines, including artificial intelligence, information retrieval, information extraction, natural language processing and psychology. Psychological approaches focus more on theoretical aspects, whereas artificial intelligence, information retrieval, information extraction and natural language processing approaches investigate how practical Question Answering systems can be engineered.

A simple query to a search engine will return hundreds of thousands of documents. This raises the need for a new approach to allow more direct access to information. Ideally, a user could ask any question (open domain) and instead of presenting a list of documents, Question Answering technology could present a simple answer to the user.

In contrast to Information Retrieval systems, Question Answering systems involve the extraction of answers to a question rather than the retrieval of relevant documents.

At present, there are textual Question Answering systems working with factual questions. These systems can answer questions of the type: *what, who, where*, *how many* and *when*. However, systems working with complex questions, such as *how* and *why,* are still under research, tackling the lack of representations and algorithms for their modelling (Burger et al., 2001; Maybury, 2003; Moldovan et al., 2003).

One kind of complex question is "*why*" (causal). One reason why causal questions have not been successfully treated is due to their answers requiring more elaboration (explanations) instead of short answers.

The idea for our research "*detecting answers to causal questions through automatic text processing*", was born, and the hypothesis is that, whilst "*why*" questions appear to require more advanced methods of natural language processing and information extraction because their answers involve opinions, judgments, interpretations or justifications, they can be approached by methods that are intermediate between Information Retrieval and full Natural Language Understanding. We advocate the use of methods based on information retrieval (bag-of-words approaches) and on limited syntactic and/or lexical semantic

analysis as a first step towards tackling the problem of automatically detecting answers for causal questions.

For this research, a *"why"* questions, is defined simply as interrogative sentences in which the interrogative adverb *why* occurs in the initial position. For example: *Why is cyclosporine dangerous?*

The topic of a *"why"* question is the proposition that is questioned and it is presupposed to be true according to the document collection; otherwise, the *"why"* question cannot be answered. The proposition of the *"why"* question shown previously is: *Cyclosporine is dangerous*.

The topic of the *"why"* question that is questioned should be the event that needs an explanation according to the questioner. Identifying the question's topic and matching it to an item (event, state, or action) in the text is a prerequisite for finding the answer. The response to the *"why"* question stated previously is:

Why is cyclosporine dangerous?
It can harm internal organs and even cause death.

Inappropriate answers to *"why"* questions are mainly due to misunderstanding the questions themselves (Galambos & Black, 1985). For example, the answer to the following question varies depending on how the question is understood.

Why did Sodel dine at the Mexican Restaurant?

If we understand that this question concerns Sodel's motivation for eating, we could reply that he ate there "because he was hungry". If we understand that the question relates to why Sodel chose that particular restaurant, we could answer that "He had heard that it is a good restaurant and he wanted to try it". If we understand that the question is about Sodel going to a restaurant instead of eating at home, we could answer that "Sodel's wife is out of town and he can't cook". Neither having the same topic nor being expressed by the same sentence constitutes a criterion of identity for *"why"* questions. In other words, the same sentence can express different *"why"* questions. There are contextual factors and background knowledge for the description of its interpretation. These answers to *"why"* questions may be subjectively true and each answer have boundary conditions for itself. The accuracy of the answer is in the mind of the perceiver.

The ultimate goal in textual Question Answering systems should be to answer any type of question; consequently, the research direction should move towards that goal, and a focus on causal questions serves this agenda well.

2. Definition of problem

Many members of the information retrieval and natural language processing community believe that Question Answering is an application in which sophisticated linguistic techniques will truly shine, owing to it being directly related to the depth of the natural language processing resources (Moldovan, et al., 2003).

However, Katz and Lin (2003) have shown that the key to the effective application of natural language processing technology is to employ it selectively only when helpful, without

abandoning the simpler techniques. They proved that syntactic relations enable a Question Answering system successfully to handle two linguistic phenomena: semantic symmetry and ambiguous modification. That is, the incorporation of syntactic information in Question Answering has a positive impact.

Voorhees shows that the research in Question Answering systems has made substantial advances (Voorhees, 2001, 2002, 2003, 2004), answering factual questions with a high degree of accuracy. Several forms of definition questions are processed appropriately, and list questions retrieve sequences of answers with good recall from large text collections, although it is necessary to answer hard and complex questions too, such as procedural, comparative, evaluative and causal questions.

On the other hand, (Moldovan, et al., 2003) showed that the main problem with Question Answering systems is the lack of representations and algorithms for modelling complex questions in order to derive as much information as possible, and for performing a well-guided search through thousands of text documents.

Some of the complex questions types seem to need much semantic, world knowledge and reasoning to be handled properly, e.g. for automatically resolving ambiguities or finding out which measure or granularity a user would prefer. This is beyond the scope of the current research and even those in the near future. However, in this research, we advocate the use of methods based on information retrieval (bag-of-words approaches) and on limited syntactic and/or lexical semantic analysis as a first step towards tackling the problem of causal questions.

Our goal was not to implement a functional (fully-fledged) textual Question Answering system but to investigate how methods based on information retrieval (bag-of-words approaches) and on limited syntactic and/or lexical semantic analysis can contribute to the real-world application of causal text and causal questions. This enables us to focus on the key matter of how the answer is contained in the document collection.

3. Related work

In the literature on Question Answering, the system developed in the Southern Methodist University and Language Computer Corporation (Harabagiu et al., 2000) has been considered to have the most sophisticated linguistic techniques due to the depth of its natural language processing resources. This system classifies questions by expected answer type, but also includes successive feedback loops that attempt to make progressively larger modifications to the original questions until they find an answer that can be justified as abductive proof —semantic transformations of questions and answers are translated into a logical form for being analysed by a theorem prover.

The system of the Southern Methodist University and Language Computer Corporation first parses the question and recognises the entities contained in it to create a question semantic form. The semantic form of the question is used to determine the expected answer type by finding the phrase that is most closely connected to other concepts in the question. The system then retrieves paragraphs from the corpus, using boolean queries and terms drawn from the original question, related concepts from WordNet, and an indication of the expected answer type.

Paragraph retrieval is repeated using different term combinations until the query returns a number of paragraphs in a pre-determined range. The retrieved paragraphs are parsed into their semantic forms, and a unification procedure is run between the question semantic form and each paragraph semantic form. If the unification fails for all paragraphs, a new set of paragraphs is retrieved using synonyms and morphological derivations of the previous query.

When the unification procedure succeeds, the semantic forms are translated into logical form, and a logical proof in the form of an abductive backchaining from the answer to the question is attempted. If the proof succeeds, the answer from the proof is returned as the answer string. Otherwise, terms that are semantically related to important question concepts are drawn from WordNet and a new set of paragraphs is retrieved.

While research in Question Answering mainly focussed on responding to factual questions, definition questions and list questions using stochastic processes, a more recent trend in Question Answering aims at responding to other types of question that are of great importance in everyday life or in professional environments such as procedural, causal, comparative or evaluative questions. These have not yet been studied in depth; they require different types of methodologies and formalisms, particularly at the level of the linguistic models, knowledge representation and reasoning procedures.

However, the ideal system does not exist yet although approaches to support that goal have been created. We will demonstrate a small number of approaches to Question Answering working with complex questions or advanced methods. The explanations are based on their general ideas.

In order to answer *"why"* questions, the aim of an ideal system should be to address a form of Question Answering that does not focus on finding facts, but rather on finding the identification and organisation of opinions, to support information analysis of the following types: *(a)* given a particular topic, find a range of opinions being expressed about it; *(b)* once opinions have been found, cluster them and their sources in different ways, and *(c)* track opinions over time.

Verberne et al. (2007) demonstrate an approach to answering *"why"* questions, based on the idea that the topic of the *"why"* question and its answer are siblings in the rhetorical structure of the document, determined according to Rhetorical Structure Theory (RST) (Mann & Thompson, 1988), connected by a rhetorical relation that is relevant for *"why"* questions —"discourse-based answer extraction". They implemented an algorithm that: *(a)* indexes all text spans not from the source document but from a manually analysed representation of it into RST relations that participate in a potentially RST relation relevant; *(b)* matches the input question to each of the text spans in the index; and *(c)* retrieves the sibling for each of the found spans as the answer. The result is a list of potential answers, ranked using a probability model that is largely based on lexical overlap. For the purpose of testing their implementation, they created a test collection consisting of seven texts from the RST Treebank and 372 *"why"* questions elicited from native speakers who had read the source documents. From this collection, they obtained a recall of 53%, with a mean reciprocal rank of 0.662. On the basis of the manual analysis of the question-answer pairs, they argued that the maximum recall that can be obtained for this data set, from the use of RST relations as proposed, is 58.0%. They declare that, although there are no reference data

for the performance of automatic Question Answering working with *"why"* questions, they considered a recall of 53% (and a maximum recall of 58%) to be mediocre at best.

Waldinger et al. (2004), Benamara and Saint-Dizier (2004), and McGuinness and Pinheiro da Silva (2004) delve into knowledge-based Question Answering and support inferential processes for verifying candidate answers and providing justifications. That is, systems that target the problem of Question Answering over multiple resources have typically taken the approach of first translating an input question into an intermediate logical representation, and, in the realm of this intermediate representation, matching parts of the question to the content supplied by various resources.

Light et al. (2004), provide an empirical analysis of a corpus of questions that enables the authors to identify examples of reuse scenarios, in which future questions could be answered better by using information previously available to the system (e.g., in the form of previously submitted questions or answers already returned to the users). The authors acknowledge that some of the proposed categories of reuse are very difficult to implement in working system modules.

Schlaefer (2007) has used ontologies for extracting terms from questions and corpus sentences and for enriching the terms with semantically similar concepts. In order to improve the accuracy of Question Answering systems, semantic resources have been used. Semantic parsing techniques are applied to transform questions into semantic structures and to find phrases in the document collection that match these structures.

Vicedo and Ferrandez (2000) have demonstrated that their evaluation improvements when pronominal references are solved for IR and Question Answering tasks. That is, they are solving pronominal anaphora.

Mitkov (2004) describes that coreference resolution has proven to be helpful in Question Answering, by establishing coreferences links between entities or events in the query and those in the documents. The sentences in the searched documents are ranked according to the coreference relationships.

Castagnola (2002) shows that for the purpose of improving the performance of Question Answering, he resolves pronoun references via the use of syntactic analysis and high precision heuristic rules.

Galitsky (2003) introduces the reasoning mechanism as the background of the suggested approach to Question Answering, particularly, scenario-based reasoning about mental attitudes. Default logic is used for correction of the semantic representations. He describes the process of representing the meaning of an input query in the constructed formal language.

Setzer et al. (2005) address the role that temporal closure plays in deriving complete and consistent temporal annotations of a text. Firstly, they discuss the approaches to temporal annotation that have been adopted in the literature, and then further motivate the need for a closed temporal representation of a document. No deep inferencing, they argue, can be performed over the events or times associated with a text without creating the hidden relations that are inherent in it. They then address the problem of comparing the diverse

temporal annotations of the same text. This is far more difficult than comparing, for example, two annotations of part-of-speech tagging or named entity extent tagging, due to the derived annotations that are generated by closure, making any comparison of the temporal relations in a document a difficult task. They demonstrate that two articles cannot be compared without examining their full temporal content, which involves applying temporal closure over the entire document, relative to the events and temporal expressions in the text. Once this has been achieved, however, an inter-annotator scoring can be performed for the two annotations.

Nyberg et al. (2004) aim to capture the requirements of advanced Question Answering and its impact on system design and the requirements imposed on the system (e.g., time-sensitive searches and the detection of obscure relations). The challenges that face the push towards the development of Question Answering systems of increased complexity are especially the challenges of practicability and scalability. Indeed, such issues become important for any system that would actually attempt to perform planning in a broad domain. Similarly, it may be very challenging to find common linguistic representations to use across highly modular systems for encoding internal information, as the information sources themselves can vary widely, from unstructured text at one end of the spectrum to full-blown knowledge bases at the other end.

Planning structures explicate how a person does certain things, and how he or she normally tries to achieve some goal. Plans cannot be built from the story itself but have to be taken from some world knowledge module. Studying instructional texts seems to be very useful for answering procedural questions —"how". Aouladomar (Aouladomar, 2005a, 2005b; Aouladomar & Saint-Dizier, 2005a, 2005b) incorporated concepts from linguistics, education, and psychology to characterise procedural questions and content to produce an extensive grammar of the ways in which a procedural text may be organised, a framework that appears to show much promise. Although her work is on French texts, the procedural features she identified included general ones, e.g. the distinct morphology of verbs in procedures. However, her research did not directly address the task of classifying texts as either procedural or non-procedural.

Aouladomar mentioned that questions beginning with how should not be neglected, since, according to recent usage data from a highly-trafficked web search engine, queries starting with how alone is the most popular category of queries beginning with question words.

The approaches to Question Answering mentioned above are likely to become relatively more language-dependent, as they require larger and more complex resources of various kinds.

3.1 Challenges of Questions Answering

The ultimate goal in textual Question Answering systems is to answer any type of question. If the information needs are very simple ones (e.g. factoid, definition or list), then the answer can be simple word(s), phrase(s) or sentence(s). If the information needs are more complex, then the answers may come from a deep documentary analysis, or from multiple documents. Where candidates answer from different corpora, these could be merged or possibly summarised.

Moreover, if we recognise that users can obtain valuable information through inference and construction from new material, combined with what they know already, the scope for Question Answering is far wider. This can be far more than simply deriving the kind of exact answer that is required due to the rich knowledge or complex inference it requires.

Alternatively, candidate answers from different languages could be translated into the user's native language. For example, if we retrieve answers from Spanish, French and Italian and translate them into English, we could compare the nature of the answers drawn from different geographic and cultural contexts.

The research direction is moving gradually towards these goals and it is our hope that the Question Answering research groups can collaborate in order to achieve these goals.

4. Approach

Most causal questions are of the form "*Why Q*?", where *Q* is an observation or fact to answer (which we have identified as an effect). If a *"why"* question is an effect, then we are searching for its explanations (which we have identified as causes). So, we have called the cause and its effect a causal relation.

The *"why"* question (effect) has an infinite number of different answers. Each answer contains an explanation of a cause for the question. In particular, causes explain their effects. For this reason, a cause tells us why its effect occurs.

The natural complexity of a question depends on how the question is understood, and the accuracy of the answer is in the mind of the perceivers, depending mainly on their knowledge level (contextual factors and background knowledge) for the description of its interpretation. Some users prefer a more accurate explanation, while others look for explanations with a broader perspective and better explanatory resources. Any answer that appeals to a cause is taken to be highly relevant and, therefore, to provide an explanation of the effect —a *"why"* question.

In order to get answers to a *"why"* question, we should try to detect causal relations. Although textual Question Answering systems are evolving towards providing exact answers only, for *"why"* questions the answers should be surrounded by some context, with the purpose of supporting the answer.

4.1 Methodology

We have used the lexico-syntactic classification for *"why"* questions proposed by Verberne et al. (2007). The categories used are existential "there" questions, process questions, questions with a declarative layer, action questions and have questions. The result of question analysis task is not used in the answer candidate extraction task. However, it gives a category to each *"why"* question.

The answer candidate extraction task provides an approach to tackling a subset of causal questions. We used the following procedure for detecting possible answers to *"why"* questions:

Identify the topic of the question.
In the list of sentences of source document, identify the clause(s) that express(es) the same proposition as the question topic.

Select the best three clauses as answers.
Detect cause-effect information expressed in the answers selected.

Step 1. The topic of the question (which we have identified as an effect) is the observation or fact that is questioned. In other words, it is the premise of question.

Step 2 and 3. We suggest that decomposition of the complex task of recognizing which source text expresses the same proposition as the question topic would make a step towards better understanding the process for answering causal questions. This should involve making use of set of measures (see 4.1.1), and using each one as a weighting factor within the whole evaluation for ranking of possible answers. The sum of factors is the final value. To be precise, each measure is applied to the words belonging to the question-text pair. The best three answers are selected.

Step 4. We used a rule-based approach to identify and extract cause-effect information expressed in the answers selected.

4.1.1 Matching formulae

The answer extraction process relies on the computation of four measures:

1. *Simple matching*. The stop words are not removed; for this reason, non stop words are weighted with *1.9* and stop words with *0.1*. The final weight is calculated as the sum of all values and normalized dividing it by the length of the question and text (total number of words). In which, *Q* is a question and *T* is a text with possible answer, see (1).

$$dist_{jcm} = IC(N1) + IC(N2) - 2 * IC(LCS) \tag{1}$$

However, if simple matching is not possible, we are working with stems. All the occurrences in the question's stems set that also appear in the text's stems set will increase the accumulated weight in a factor of one unit. The stems are weighted with *1.9* for non stop words and *0.1* for stop words.

Longest consecutive subsequence. This process measures the surface structure overlap between the text with possible answer and the question (only consecutive words). In order to compute this overlap we extract the longest consecutive subsequence (LoCoSu) between the question and the text with possible answer, *LoCoSu(Q, T)*, see (2).

$$structure\_Overlap = \left| LoCoSu(Q,T) \right| \tag{2}$$

For example, if we have *Q*= {"AA", "BB", "CC", "FF" } and *T*= {"AA", "BB", "DD", "FF"} then *LoCoSu* = {"AA", "BB"} = 2.

In order to calculate *LoCoSu*, we have used a third party implementation, the longest common substring tool (Dao, 2005).

This feature indicates the presence of the same word with *1*, or otherwise *zero*. We are removing stop words. We are using stems if simple matching is not possible.

One should note that this measure assigns the same relevance to all consecutive subsequences with the same length. Furthermore, the longer the subsequence is, the more relevant it will be considered.

We have used a threshold of 2, that is, *LoCoSu (Q, T)* bigger or equal to 2.

Sorensen's similarity coefficient. This only considers non stop words. The value is increased by one per word in the intersection or union. Sorensen's similarity coefficient is a distance measure, see (3).

$$Sorensen = 1 - D \tag{3}$$

Where D is Dice's coefficient, which is always in [0,1] range, see (4).

$$D = \frac{2|Q \cap T|}{|Q| + |T|} \tag{4}$$

Where Q is a set of words of question and T is a set of words of sentence, possible answer. Note that if simple matching is not possible, we use their stems for evaluation.

WordNet-based Lexical Semantic Relatedness. The measure uses WordNet (Miller, 1995) as its central resource. Here, we are in fact considering similarities between concepts (or word senses) rather than words, since a word may have more than one sense. Measures of similarity are based on information in is-a hierarchy. WordNet only contains is-a hierarchies for verbs and nouns, so similarities can only be found where both words are in one of these categories. WordNet includes adjectives and adverbs but these are not organised into is-a hierarchies, so similarity measures cannot be applied.

Concepts can, however, be related in many ways apart from being similar to each other. These include part-of relationships, as well as opposites and so on. Measures of relatedness make use of this additional, non-hierarchal information in WordNet, including the gloss of the synset. As such, they can be applied to a wider range of concept pairs including words that are from different parts of speech.

If we want to compute lexical semantic relatedness between pairs of lexical items using WordNet, we can find that several measures have been reported in the literature. According to the evaluation of Budanitsky and Hirst (2006), the measure proposed by Jiang and Conrath (1997) is the most effective. The same measure is found the best in word sense disambiguation (Patwardhan, Banerjee, & Pedersen, 2003). We confirm that the Jiang and Conrath measure was the best for the task of patter_induction for information extraction (Stevenson & Greenwood, 2005).

The Jiang and Conrath metric (jcm) uses the information content (IC) of the least common subsumer (LCS) of the two concepts. The idea is that the amount of information two concepts share will indicate the degree of similarity of the concepts, and the amount of information the two concepts share is indicated by the *IC* of their *LCS*. Thus, they take the sum of the *IC* of the individual concepts and subtract from that the *IC* of their *LCS*, see (5).

$$dist_{jcm} = IC(N1) + IC(N2) - 2 * IC(LCS) \tag{5}$$

Where *N1* is the number of nodes on the path from the *LCS* to concept *1* and *N2* is the number of nodes on the path from the *LCS* to concept *2*.

Since this is a distance measure, concepts that are more similar have a lower score than the less similar ones. The result with the smallest distance is taken to disambiguate the senses between two words. In order to maintain consistency among the measures, they convert this measure to semantic similarity by taking its inverse, see (6).

$$sim_{jcm} = \frac{1}{IC(N1) + IC(N2) - 2 * IC(LCS)} \tag{6}$$

In order to calculate *jcm* similarity, we have used a third party implementation, the WordNet Relatedness tool (Pedersen, Patwardhan, & Michelizzi, 2004).

We use three types of words for calculating this feature in order to discover lexical semantic relatedness:

a. *jcm* similarity between question-word against text-word.
b. *jcm* similarity between question-word against synonyms of text-word. The idea is to show that if synonyms are different words with identical or at least similar meanings, then we can use them to calculate *jcm* similarity in order to provide extra resources for disambiguation process.
c. *jcm* similarity between question-word and the synonyms of text-word antonyms. We think that a word pair where the individual words are opposite in meaning could help in the disambiguation process, identifying cause-effect (text-question) described with opposite words. In other words, an additional exploration of potential associative relations for text-question pairs.

4.2 Implementation

The system architecture is depicted in Fig. 1. As can be seen, it has a base client-server architecture (Shaw & Garlan, 1996) hosting several components that support different duties. There are three main actors in the environment surrounding the system: (*i*) the User —which is the person who issues the questions to be answered by the system, (*ii*) the Administrator —which is the person whose main duty is to update the system's databases and (*iii*) CAFETIERE —which is an external tool used to support the query/document processing work. All the communication among the main architectural parts of the system, is carried out in a synchronous request-response mode, i.e. a "source part" submits a request and waits until the response is returned from the "target part". All the components of the architecture are written in Java.

As depicted in Fig. 1, the client-side of the architecture hosts two user interface (UI) components: the User UI and Administration UI. These UI components enable the communication of the User and Administrator with the system.

A User issues a question to the system via the User UI component. This question is passed up to the server as a user request. In the server side, all user requests are processed by the Query Processing component. Internally, this component has a Pipe-and-Filter like architecture (Shaw & Garlan, 1996), which allows splitting the query-processing job into a series of well-defined low-coupled sequential steps. Three filter components constitute the Query Processing component: Question Analysis, Answer Candidate Extraction and Cause-Effect Detection. The Question Analysis filter enables the classification of the issued

question with regard to a category that corresponds to a syntactic pattern. The Answer Candidate Extraction filter automatically maps the question onto the sentences of document, mainly by measuring lexical overlapping and lexical semantic relatedness between the question topic and the sentences of document to detect possible answers for the question evaluated. Finally the Cause-Effect Detection filter uses a rule-based approach in order to identify the cause and effect information expressed in the selected answers. Both the Question Analysis and Cause-Effect Detection filters interact with the CAFETIER tool (Black et al., 2003). Specifically, they use CAFETIER's lexico-syntactic analysis pipeline.

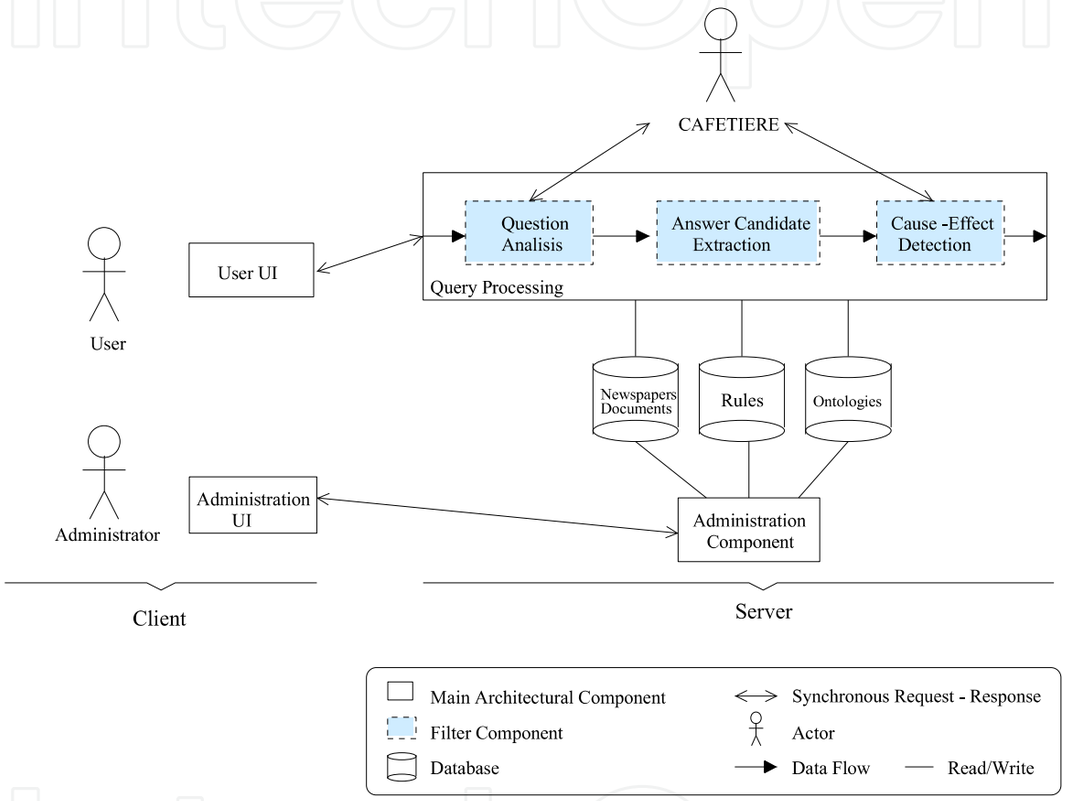

Fig. 1. The architecture of the system.

The Administration UI component is the means by which the Administrator maintains the system's databases. As shown in Fig. 1, there are three databases: (*i*) Newspaper Documents —which contains a list of candidate text passages (possible answers) that, in all likelihood, match the original question, (*ii*) Rules —which contains a set of lexico-syntactic and basic semantic rules for the English language that are used to produce phrasal and conceptual annotations as well as representations of elements of interest, events and relations and (*iii*) Ontologies —which are lists of known names of places, people, organizations, artifacts, etc.; that help to assign conceptual classes to single and/or multi word phrases as additional information for the information extraction analysis.

In the following sections we will focus on describing the elements of the architecture supporting Query Processing.

4.2.1 Question Analysis

As mentioned before, we use CAFETIERE's lexico-syntactic analysis pipeline for the classification of *"why"* questions. This pipeline is constituted by: sentence splitter, tokenizer, orthography tagger, stemmer, POS tagger, gazetteer lookup (single and multi-word names and terms), and rule-based analyzer (context sensitive rule-based analysis).

The Question Analysis filter that implements the required logic to interact with CAFETIERE in the following way:

1. *Modifying the resources for an already existing analysis engine.* For example, we have done the following:

- added words to the lexicon used by the part-of-speech tagger.
- added rules to the file used by the part-of-speech tagger.
- added patterns to the file used by the part-of-speech tagger.

The previous three points have been executed for improving the result of part-of-speech tagger to our research.

2. *Creating a new instance of an existing analysis engine type, with its own set of resources*. This does not involve changing any of the code of the analysis engine, only declaring what specific resource instance(s) it will use. We have for example:

- created one lookup analysis engine for using our six gazetteers: list of cue words, modal verbs, auxiliaries, process verbs, declarative verbs and agentive nouns. The objective was to support the phrase-level analysis of our research.
- created three different rule-based analysis engines, each with its own set of rules for performing higher level analysis up to the clause and sentence level. They have been divided into three levels: tags, phrase level and clause level.

3. *Creating a new aggregate analysis engines to run modules in different sequences, or to run different permutations of modules that are used in the question analysis*. Basically, the analysis engines integrating the aggregate are: sentence splitter, tokenizer, orthography tagger, stemmer, POS tagger, gazetteer lookup, rule-based analyzer and concept collector.

This question analysis process relies on CAFETIERE tool in order to assign a question category that corresponds to the types of entities, which constitute the category —each question category corresponds to a syntactic pattern. The question focus is its premise.

The process is a rule-based approach, which uses hand-crafted rules that look for lexical and syntactic clues in the question. We have sets of rules for tags, phrases, and clauses (which include the question types). The rules use six gazetteers as knowledge source: cue words, modal verbs and auxiliaries, process verbs, declarative verbs and agentive nouns. We have 143 rules, which are constituted by 32 rules for tags, 76 rules for phrases and 35 rules for clauses.

The output generated for this filter, which is the annotated and original question, are passed up to the answer candidate extraction filter.

4.2.2 Answer candidate extraction

We implemented an answer candidate extraction from text to identify the three best answers to the question. In order to do that, this component implements the logic to perform the

computation of four matching formulae: simple matching, longest consecutive subsequence, Sorensen's similarity coefficient and WordNet-based lexical semantic relatedness. This process matches the question against the sentences in the text to identify the clauses that express the same proposition as the question.

Within the entire set of measures, each one of them is considered as a feature with the same weight. For selecting the best three possible answers for each question, we have combined the four metrics, see (7).

$$matching\_formulae = \frac{f_{sm} + f_{lcs} + f_{ssc} + f_{lsr}}{4} \quad (7)$$

The extraction of possible answers is accomplished by selecting the top three possible answers for each question. The answers are saved into a file, which is passed up to the Cause-Effect Detection filter to do the corresponding processing.

4.2.3 Cause-Effect Detection

The detection of cause-effect information expressed in the identified best three possible answers is done by the Cause-Effect Detection filter. As depicted in Fig. 1, this filter contains the required logic to use the CAFETIERE tool to support this job.

We have investigated how cause-effect information could be extracted from newspaper text using rule-based approach without full parsing of sentences. A set of rules that usually indicate the presence of a causal relationship was constructed and used for the extraction of cause-effect information.

No inferencing from common sense knowledge or domain knowledge was used. Knowledge-based inferencing of causal relationships requires a detailed knowledge of the domain, and newspaper text covers a very wide range of topics. Only linguistic clues were used to identify causal relationships. For example, we are using explicit linguistic indications of cause and effect, such as *because, however, due to, so, therefore, but, as a result of this,* and so on, instead of inferencing from common sense knowledge or domain knowledge.

We identified the following two ways of explicitly expressing cause-effect:

1. Using causal links to link two phrases, clauses or sentences.
2. Using causative verbs.

The detection of cause-effect information (rule-based approach) expressed in the identified answers was implemented in the same way that our question analysis (see Section 4.2.1), *(a)* modifying the resources for an already existing analysis engine; *(b)* creating a new instance of an existing analysis engine type, with its own set of resources; and *(c)* creating a new aggregate analysis engine.

Rules were created to identify sentences containing *causal links and causative verbs*. The cause-effect information was then extracted. The implementation can identify causal relations in newspaper text when it focuses on the causal relations that are explicitly indicated in the text using linguistic means.

A complete example is the question, "Why did researches compare changes in oxygen concentration?" which belongs to the document wsj_0683. It is an action question. The best three possible answers detected by our answer candidate extraction task are:

1. Researchers at Ohio State University and Lanzhou Institute of Glaciology and Geocryology in China have analyzed samples of glacial ice in Tibet and say temperatures there have been significantly higher on average over the past half-century than in any similar period in the past 10,000 years.
2. According to greenhouse theories, increased carbon dioxide emissions, largely caused by burning of fossil fuels, will cause the Earth to warm up because carbon dioxide prevents heat from escaping into space.
3. To compare temperatures over the past 10,000 years, researchers analyzed the changes in concentrations of two forms of oxygen.

The correct answer is the third one. The verb phrase *to compare temperatures over the past 10,000 years* is the cause, and the clause *researchers analyzed the changes in concentrations of two forms of oxygen* is the effect.

5. Evaluation

We have shown previously that a "*why*" question has more than one answer since there does not exist only one correct way of explaining things; therefore, it is quite difficult to determine whether a string of text provides the correct answer. Human assessors have legitimate differences of opinions in determining whether a response actually answers a question. If human assessors have different opinions, then eventual end-users of the Question Answering technology have different opinions as well because some users prefer a more accurate explanation, while others look for explanations having a broader perspective and better explanatory resources.

We can highlight that the time and effort required to manually evaluate a Question Answering application is considerable, owing to the need for human judgment. This issue is compounded by the fact that there is no such thing as a canonical answer form. Assessors' decision on the correctness of an answer makes resulting scores comparative, not absolute.

For our evaluation, we used the collection of Verberne et al. (2007), It has a relative preponderance of questions (typically expressed in the past tense) about specific actions of their motivations or relations to them, because their source documents are Wall Street Journal articles (news) and they describe a series of events that are specific to the topic, place and time of the text. The collection consists of seven texts from the RST Treebank of 350-550 words each and 372 "*why*" questions.

Throughout the evaluation of the answers detected, we adopt a manual approach whereby an assessor determines if a response is suitable for a question, with two possible outcomes. The response is either correct (i.e. the answer string must contain exactly the information required by the question), or incorrect (i.e. it is a wrong answer or no answer at all). To evaluate a question, an assessor was required to judge each answer string in that question's answer pool (set of three answers).

The kind of evaluation executed for our research is a post-hoc evaluation on analyzed data —question-answer pairs. It is divided in three sections: question classification, which is a lexico-syntactic classification where each question category corresponds to a pattern, answer candidate extraction, which maps the question onto the correct source text in order to detect possible answers for the question evaluated, and the detection of cause-effect information expressed in the answers selected.

5.1 Question classification

We classified the 372 "*why*" questions of the collection using the question analysis filter (rule-based approach). The lexico-syntactic classification is constituted of 5 categories: *existential "there" questions, process questions, questions with a declarative layer, action questions and have questions*.

The following questions are examples that were classified with the existential "there" category:

- Why is there resistance to the Classroom Channel?
- Why is there a reference to the musical "The Music Man"?
- Why is there controversy in this Dallas suburb?

Examples of questions in the process question category are:

- Why did Cincinnati Public Schools reject the subscription offer?
- Why did the US Coast Guard close part of the Houston Ship Channel?
- Why did the petroleum plant explode?

The following questions are examples of declarative layer category:

- Why does Whittle think he can reach subscription goals within one year?
- Why did the report say that advertisers were showing interest?
- Why does Mr Hogan think the company will be successful?

Examples of questions in the action question category are:

- Why do researches conclude that the earth is warming?
- Why did Dr. Starzl advise against buying Fujisawa stock?
- Why were town officials embarrassed?

The following questions are examples that were classified with the have question category:

- Why did firefighters have difficulties getting the fire under control?
- Why does Whittle have reason for concern?
- Why did the research team have no financial stake in the drug?

For some categories, the question analysis filter only needs fairly simple cues for choosing a category. For example, the presence of the word *there* with the syntactic category *EX* leads to an the category existential "there" question.

For deciding on questions with a declarative layer, action questions and process questions, complementary lexical-syntactic information is needed. In order to decide whether the question contains a declarative layer, the filter checks whether the main verb is in

declarative verbs list, and whether it has a subordinate clause. The distinction between action and process questions is made by looking up the main verb in a list of process verbs. This list contains the 529 verbs from Levin verb index (1993). If the main verb is not determined to be process, declarative or have, it is assigned to the action verb category.

Questions with a declarative layer need further analysis because they are ambiguous. For example the question "Why did they say that migration occurs?" can be interpreted in two ways: "Why did they say it?" or "Why does migration occur?". Our answer candidate extraction filter should try to find out which of these two questions is supposed to be answered. In other words, the filter should decide which of the clauses contains the question focus. For this reason, questions with a declarative layer are most difficult to answer.

Table 1 shows the results of our question classification. We observe that the five categories of classification (existential "there" questions, process questions, questions with a declarative layer, action questions and have questions) had a performance of 54.56%. In other words, for the 372 questions of collection, 203 questions were classified correctly. We want to highlight that question classification uses only questions, without answers.

Category	Question per Category	Questions Classified Correctly	Performance
Existential "there" questions	11	5	45.45%
Process questions	145	102	70.34%
Questions with a declarative layer	93	17	18.27%
Action questions	102	65	63.72%
Have questions	21	14	66.66%
TOTAL	372	203	54.56%

Table 1. Lexico-syntactic classification for the 372 "*why*" questions of collection.

We have assigned correctly the existential "there" category to 45.45% of the questions; 70.34% were labelled as process questions; 18.27% of the questions had a declarative layer; the category of action questions was assigned to 63.72% of the questions because if the main verb of the questions is not a process, declarative or a have verb, then we are assumed that its type is action. And 66.66% were labelled, as have questions.

We observe that question with a declarative layer are most difficult to identify because of clausal object, that is, a subordinate clause must be detected after declarative verb.

The general rules (only a lexico-syntactic analysis) for categories of classification are:

- *Existential "there"* category:

WRB + VP + EX + NP + [ADVP|PP|NP] + ?

- *Process question* category:

WRB + VP + + NP + Process-VERB + [ADVP | PP | NP] + ?

- *Declarative layer* category:

WRB + VP + NP + Declarative-VERB + S + ?

- *Action question* category:

WRB + VP + NP + Action-VERB + [NP] + [ADVP | PP | NP] + ?

- *Have question* category:

WRB + VP + NP + Have-VERB + NP + [ADVP | PP | NP] + ?

If the lexico-syntactic pattern of the question did not correspond to any category, then it was assigned to the default category. However, we detected that questions can also be classified in the default category due to *(a)* errors of part-of-speech tagger; and *(b)* errors of rule-based analyzer. The default category contained 45.43% of the questions. A few examples of questions assigned to the default category are:

- Why is the sago expensive?
- Why is the Sago a pricey lawn decoration?
- Why is rowdy behavior unlikely at the Grand Kempinski?

5.2 Answer candidate extraction

In order to evaluate the answer candidate extraction filter, most previous Question Answering work has been evaluated using traditional metrics as recall, and mean reciprocal rank (Voorhees, 2003, 2004). We follow the standard definition of recall, see (8).

$$recall = \frac{c}{t} \tag{8}$$

Where c is the number of correct annotations produced, and t is the total number of annotations that should have been produced.

Using this formula, the recall obtained by our lexical overlapping and lexical semantic relatedness approach is 36.02%. In our test corpus, t=372, the number of *"why"* questions in the collection, and c=134, the number that were answered correctly by those techniques.

We hypothesized some of the *"why"* questions could have been unanswered because the collection's questions were created by native speakers who might have been tempted to formulate *"why"* questions that did not address the type of argumentation that one would expect of questions posed by persons who needed a practical answer to a natural *"why"* questions.

For this reason, working from the premise that "our lexical overlapping and lexical semantic relatedness approach can only answer *'why'* questions with explicit and ambiguous causation because it uses basic external knowledge for disambiguation", we recalculated recall for the 218 *"why"* questions with explicit and ambiguous causation. The rate of our recall increased to 61.46% of the former. The rate of recall thus increases considerably and

we reach similar results as the RST method (Verberne, et al., 2007), which relies on texts where all causal relations have been pre-analysed.

The second metric used was mean reciprocal rank (MRR), see (9). The original evaluation metric used in the Question Answering tracks TREC 8 and 9 (Voorhees, 2000) was mean reciprocal rank (MRR), which provides a method for scoring systems which return multiple competing answers per question. We used MRR because our implementation returns a list with the best 3 possible answers that have been found to each question.

$$MRR = \frac{\sum_{i=1}^{|Q|} \frac{1}{r_i}}{|Q|} \tag{9}$$

Where Q is the question collection and r_i the rank of the first correct answer to question i or *0* if no correct answer is returned.

We showed that our answer candidate extraction filter found answers for 134 *"why"* questions on undifferentiated texts. The distribution for 134 *"why"* questions is 48 correct answers located in the first position, 29 correct answers located in the second position and 57 correct answers located in the third position. Consequently, the MRR is 0.219

Working from the previously mentioned premise that "our lexical overlapping and lexical semantic relatedness approach can only answer *'why'* questions with explicit and ambiguous causation because it uses basic external knowledge for disambiguation", then for the 218 *"why"* questions with explicit and ambiguous causation, the rate of our MRR increases to 0.373 of the former.

5.3 Cause-effect information

The cause-effect information in the answers selected is mainly expressed by causal links. The reason for this could be the fact that the events discussed in newspaper texts use connectives between two adjacent clauses. We detected the following causal links: *before*, *after*, *where*, *due to*, *because of*, *but*, *about*, *because*, *for*, and *from*. Three examples are presented:

Why are the Mayor and two members of the Council worried?

Mayor Lynn Spruill and two members of the council said they were worried ***about*** setting a precedent that would permit pool halls along Addison's main street.

Why would the interior regions of Asia heat up first?

Some climate models project that interior regions of Asia would be among the first to heat up in a global warming ***because*** they are far from oceans, which moderate temperature changes.

Why could the number of people known injured increase?

Nearby Pasadena, Texas, police reported that 104 people had been taken to area hospitals**, *but*** a spokeswoman said that toll could rise.

The 89.01 % of question-answer pairs of the collection contain causal links.

In order to evaluate cause-effect information, we manually identified 50 question-answer pairs in the collection. After that, we used the answer candidate extraction filter for evaluating the cause-effect information detected, using the same 50 question-answer pairs manually identified. Table 2 shows we detect 34% of cause-effect information.

	Total questions-answer pairs	% of questions-answer pairs
Question-Answer pairs analysed manually	50	100%
Questions for which we identified a text (possible answer)	39	78%
Questions for which the identified text is a correct answer	17	34%

Table 2. Outcome of cause-effect evaluation.

Questions, which contain modals, constitute 7.52% of the total of collection. The function of modals is important in defining the semantic class of question. We cannot solve this issue because our answer candidate extraction filter works with lexico-syntactic level. For example:

1. Why did Nando's not use actors to represent chefs in funny situations?
2. Why can Nando's not use actors to represent chefs in funny situations?

Answer to Question *1* is a motivation, and answer to question *2* is a cause.

6. Conclusion

We introduced an approach which draws on linguistic structures, enabling the classification of "*why*" questions and the retrieval of answers for "*why*" questions from a newspaper collection. The steps to summarize our approach are:

1. Assign one category to each "*why*" question, using lexico-syntactic analysis. Each question corresponds to a syntactic pattern (rule-based approach).

   The lexico-syntactic classification is constituted of 5 categories: *existential "there" questions*, *process questions*, *questions with a declarative layer, action questions* and *have questions*.

2. Detecting three possible answers to each "*why*" questions:

   2.1. Identify the topic of the question (effect).
   2.2. In the list of sentences of source document, identify the clause(s) that express(es) the same proposition as the question topic (making use of a set of measures).
   2.3. Select the best three clauses as answers.
   2.4. Detect cause-effect information expressed in the answers selected (rule-based approach).

The output for each question is a question category and three possible answers.

6.1 Contributions

We have hypothesised that these methods will also work for *"why"* questions, and have attempted to discover to what extent methods based on information retrieval ('bag of words' approaches), and on limited syntactic and/or lexical semantic analysis can find answers to *"why"* questions. So, our research contributes new knowledge to the area of automatic text processing by the following:

- We have developed an analysis component for feature extraction and classification from questions (a rule-based approach as a first step towards tackling the problem of question analysis of *"why"* questions using an Information Extraction Analyzer.
- An original answer candidate extraction filter has been developed that uses an approach that combines lexical overlapping and lexical semantic relatedness (lexico-syntactic approach) to rank possible answers to causal questions. On undifferentiated texts, we obtained an overall recall of 36.02% with a mean reciprocal rank of 0.219, indicating that simple matching is adequate for answering over one-third of *"why"* questions. We analyzed those question-answer pairs where the answer was explicit, ambiguous and implicit, and found that if we can separate the latter category, the rate of recall increases considerably. When texts that contain explicit or ambiguous indications of causal relations are distinguished from those in which the causal relation is implicit, recall can be calculated as 61.46% of the former, with a mean reciprocal rank of 0.373, which is comparable to results reported for texts where all causal relations have been pre-analyzed. This plausible result shows the viability of our research for automatically answering causal questions with explicit and sometimes ambiguous causation.
- We have found that people have conflicting opinions as to what constitutes an acceptable response to a *"why"* question. Our analysis suggests that there should be a proportion of text in which the reasoning or explanation that constitutes an answer to the *"why"* question is present, or capable of being extracted from the source text. Consequently, the complexity of a *"why"* question depends on the knowledge level of users. While some users prefer a more accurate explanation, others look for explanations with a broader perspective and better explanatory resources. Any answer that appeals to a cause is taken to be highly relevant and, therefore, to provide an explanation of the effect − a *"why"* question. In order to provide a context with which to support the answer, the paragraph from which the answer was extracted should be returned as the answer.

We conclude that this research offers a greater understanding of *"why"*. It provides an approach to tackling a subset of *"why"* questions (with explicit and sometimes ambiguous causation) which combines lexical overlapping and lexical semantic relatedness. It further considers the detection of cause-effect information that is explicitly indicated in the text using causal links and causative verbs. For these reasons, this research contributes to a better understanding of automatic text processing for detecting answers to *"why"* questions and to the development of future applications for answering causal questions.

6.2 Further work

To improvements the answer candidate extraction, we could experiment with ambiguous and implicit causation since our lexico-syntactic approach has not been successful for these

types of causation. In order to generate correct answers, we would need to go beyond the co-occurrence of terms and lexical semantic relatedness due to the mismatch between the expressions used in the question and the expressions used in the source text.

We should contribute to the implementation and evaluation of fundamental techniques representing knowledge and reasoning.

When we use a causal relation to describe the interaction between two sentences, it would be more interesting and informative if we could present an answer that offers chain of explanations connecting the two events, that is, the entire answer to the *"why"* question should be a chain of explanation of its causal relation.

When considering the relevance of answers to causal questions, we should involve carrying out inferences (we view inference as the process of making implicit information explicit) to arrive at the required answer. A relevant answer requires the provision of an appropriate explanation, according to the questioner. The explanation should increase the questioner's existing knowledge rather than duplicate it. A filter for explanation generation that takes into account the descriptions already presupposed by the question could be implemented by using descriptions to generate explanations, which are themselves answers to the *"why"* question. The algorithm could begin to make use of the information explicitly encoded by the lexical and syntactic analysis.

Three questions need to be considered in order to advance our reasoning about the descriptions embedded in *"why"* question, (1) Where can we find the descriptions presupposed by the question?, (2) How can we recognise the descriptions presupposed by the question?, and (3) How can we represent and compute the descriptions presupposed by the question?

In order to understand this process in more detail, an analysis of epistemology (the theory of knowledge) would be necessary. This is the branch of philosophy concerned whit the nature, origin, and scope of knowledge. It addresses the question "how do you know what you know?"

7. References

Aouladomar, F. (2005a). A Preliminary Analysis of the Discursive and Rhetorical Structure of Procedural Texts, *Proceedings of the Symposium on the Exploration and Modelling of Meaning*, pp. 21-24, Biarritz, France, November 14-15, 2005.

Aouladomar, F. (2005b). Towards Answering Procedural Questions, *Proceedings of the Workshop on Knowledge and Reasoning for Answering Questions*, pp. 21-31, Edinburgh, Scotland, July 30, 2005.

Aouladomar, F., & Saint-Dizier, P. (2005a). An Exploration of the Diversity of Natural Argumentation in Instructional Texts, *Proceedings of the 5th International Workshop on Computational Models of Natural Argument*, pp. 12-15, Edinburgh, Scotland, July 30, 2005.

Aouladomar, F., & Saint-Dizier, P. (2005b). Towards Generating Procedural Texts: an exploration of their rhetorical and argumentative structure, *Proceedings of the 10th European Workshop on Natural Language Generation*, pp. 156-161, Aberdeen, Scotland, August 8-10, 2005.

Benamara, F., & Saint-Dizier, P. (2004). Advanced Relaxation for Cooperative Question Answering. In: *New Directions in Question Answering*, M. Maybury, (Ed.), 263-274, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

Black, W., McNaught, J., Vasilakopoulos, A., Zervanou, K., Theodoulidis, B., & Rinaldi, F. (2003). *CAFETIERE: Conceptual Annotations for Facts, Events, Terms, Individual Entities, and Relations,* Parmenides Technical Report No. TR-U4.3.1, In: The University of Manchester.

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics,* Vol. 32, No. 1, (March 2006), pp. 13-47, ISSN 0891-2017.

Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., et al. (2001). *Issues, Tasks and Program Structures to Roadmap Research in Question & Answering,* Technical Report, In: NIST, Available from http://www-nlpir.nist.gov/projects/duc/roadmapping.html.

Castagnola, L. (2002). *Anaphora Resolution for Question Answering,* Master Degree Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA, June, 2002.

Dao, T. (2005). The Longest Common Substring with Maximal Consecutive, In: *The Code Project*, Retrieved on 30/09/2011, from http://www.codeproject.com/KB/recipes/lcs.aspx?display=Print

Galambos, J. A., & Black, J. B. (1985). Using Knowledge of Activities to Understand and Answer Questions. In: *The psychology of questions,* A. C. Graesser & J. B. Black (Eds.), 148-169, ISBN 0898594448, Lawrence Erlbaum Associates.

Galitsky, B. (2003). *Natural Language Question Answering system*, Advanced Knowledge International Pty Ltd, ISBN 0868039799, London, UK.

Harabagiu, S., Moldovan, D., Pasca, M., Mihalcea, R., Surdeanu, M., Bunescu, R., Girju, R., Rus, V. and Morarescu, P. (2000). FALCON: Boosting Knowledge for Answer Engines. P*roceedigns of the Ninth Text REtrieval Conference (TREC-9),* pp. 479-488, online publication, 06/10/11, http://trec.nist.gov/pubs/trec9/t9_proceedings.html, Gaithersburg, Maryland, USA, November 13-16, 2000.

Jiang, J., & Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *Proceedings of the International Conference on Research in Computational Linguistics,* pp. 19-33, Taipei, Taiwan, August, 1997.

Katz, B., & Lin, J. (2003). Selectively Using Relations to Improve Precision in Question Answering, *Proceedings of EACL-2003 Workshop on Natural Language Processing for Question Answering*, pp. 43-50, Budapest, Hungary, April 12-17, 2003.

Levin, B. (1993). *English Verb Classes and Alternations: a preliminary investigation,* University of Chicago Press, ISBN 0226475336, USA.

Light, M., Ittycheriah, A., Latto, A., & McCracken, N. (2004). Reuse in Question Answering: A Preliminary Study. In: *New Directions in Question Answering*, M. Maybury, (Ed.), 162-182, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text,* Vol. 8, No. 3, (December 1998), pp. 243-281.

Maybury, M. T. (2003). Toward a Question Answering Roadmap. In: *New Directions in Question Answering*, M. Maybury, (Ed.), 3-14, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

McGuinness, D., & Pinheiro da Silva, P. (2004). Trusting Answers from Web applications. In: *New Directions in Question Answering*, M. Maybury, (Ed.), 275-286, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communication of the ACM,* Vol. 38, No. 11, (November 1995), pp. 39-41, ISSN 0001-0782.

Mitkov, R. (2004). Anaphora Resolution. In: *The Oxford Handbook of Computational Linguistics,* R. Mitkov (Ed.), 265-283, Oxford University Press, ISBN 0198238827, USA.

Moldovan, D., Pasca, M., Harabagiu, S., & Surdeanu, M. (2003). Performance Issues and Error Analysis in an Open-Domain Question Answering System. *ACM Transaction on Information Systems,* Vol.21, No.2, (April 2003), pp. 133-154, ISSN 1046-8188.

Nyberg, E., Burger, J., Mardis, S., & Ferrucci, D. (2004). Software Architectures for Advanced QA, In: *New Directions in Question Answering*, M. Maybury, (Ed.), 19-30, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

Patwardhan, S., Banerjee, S., & Pedersen, T. (2003). Using Measures of Semantic Relatedness for Word Sense Disambiguation*, Proceedings of the Fourth International Conference on Intelligence Text Processing and Computational Linguistics*, pp. 241-257, Mexico City, Mexico, February 16-22, 2003.

Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). WordNet::Similarity - Measuring the Relatedness of Concepts*, Proceedings of the 9th National Conference on Artificial Intelligence, (Intelligent Systems Demonstration)*, pp. 38-41, Stroudsburg, PA, USA, May 2-7, 2004.

Schlaefer, N. (2007). *A Semantic Approach to Question Answering*, VDM Verlag Dr. Mueller e.K., ISBN 3836450739.

Setzer, A., Gaizauskas, R., & Hepple, M. (2005). The Role of Inference in the Temporal Annotation and Analysis of Text, *Journal Language Resources and Evaluation,* Vol. 39, No. 2-3, (May 2005), pp. 243-265, ISSN 1574020X.

Shaw, M., & Garlan, D. (1996). *Software Architecture: Perspectives on an Emerging Discipline*, Prentice Hall, ISBN 0131829572, USA.

Stevenson, M., & Greenwood, M. (2005). *A semantic approach to IE pattern induction, Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 379-386, Stroudsburg, Pensilvania, USA, June 25-30, 2005.

Verberne, S., Boves, L., Oostdijk, N., & Coppen, P. (2007). Discourse-based answering of why-questions, *Journal Traitement Automatique des Langues. Special issue on Computational Approaches to Discourse and Document Processing,* Vol. 47, No. 2, (Octuber 2007), pp. 21-41.

Vicedo, J., & Fernandes, A. (2000). Applying Anaphora Resolution to Question Answering and Information Retrieval Systems*, Proceedigns of Web-Age Information Management*, pp. 344-355, ISBN 3-540-67627-9, Shanghai, China, June 21-23, 2000.

Voorhees, E. M. (2000). Overview of the TREC-9 Question Answering Track*, Proceedings of the Ninth Text REtrieval Conference*. pp. 1-13, online publication, 06/10/11, http://trec.nist.gov/pubs/trec9/t9_proceedings.html, Gaithersburg, Maryland, USA, November 13-16, 2000.

Voorhees, E. M. (2001). Overview of the TREC 2001 Question Answering Track, *Proceedings of the Tenth Text REtrieval Conference*. pp. 1-15, online publication, 05/10/11, http://trec.nist.gov/pubs/trec10/t10_proceedings.html, Gaithersburg, Maryland, USA, November 19-22, 2001.

Voorhees, E. M. (2002). Overview of the TREC 2002 Question Answering Track, *Proceedings of the Eleventh Text REtrieval Conference*. pp. 1-15, online publication, 04/10/11, http://trec.nist.gov/pubs/trec11/t11_proceedings.html, Gaithersburg, Maryland, USA, November 18-21, 2002.

Voorhees, E. M. (2003). Overview of the TREC 2003 Question Answering Track, *Proceedings of the Twelfth Text REtrieval Conference*. pp. 1-13, online publication, 03/10/11, http://trec.nist.gov/pubs/trec12/t12_proceedings.html, Gaithersburg, Maryland, USA, November 17-20, 2003.

Voorhees, E. M. (2004). Overview of the TREC 2004 Question Answering Track, *Proceedings of the Thirteenth Text REtrieval Conference*, pp. 1-12, Special Publication 500-261, Gaithersburg, Maryland, USA, November 16-19, 2004.

Waldinger, R., Appelt, D., Dungan, J., Fry, J., Hobbs, J., Israel, D., et al. (2004). Deductive Question Answering from Multiple Resources, In: *New Directions in Question Answering*, M. Maybury, (Ed.), 253-262, ISBN 0-262-63304-3, Cambridge, Massachusetts USA.

IntechOpen