Syracuse University
SURFACE

The School of Information Studies Faculty Scholarship

School of Information Studies (iSchool)

2003

What do You Mean? Finding Answers to Complex Questions

Anne R. Diekema
Utah State University

Ozgur Yilmazel
Syracuse University

Jiangping Chen
Syracuse University

Sarah Harwell
Syracuse University, scharwel@syr.edu

Elizabeth D. Liddy
Syracuse University, liddy@syr.edu

See next page for additional authors

Follow this and additional works at: http://surface.syr.edu/istpub

Part of the Library and Information Science Commons, and the Linguistics Commons

Recommended Citation
Diekema, A. R., Yilmazel, O., Chen, J., Harwell, S., He, L., Liddy, E. D. 2003. What Do You Mean? Finding Answers to Complex Questions.

This Conference Document is brought to you for free and open access by the School of Information Studies (iSchool) at SURFACE. It has been accepted for inclusion in The School of Information Studies Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Authors/Contributors
Anne R. Diekema, Ozgur Yilmazel, Jiangping Chen, Sarah Harwell, Elizabeth D. Liddy, and Lan He

This conference document is available at SURFACE: http://surface.syr.edu/istpub/60

What do You Mean? Finding Answers to Complex Questions

Anne R. Diekema, Ozgur Yilmazel, Jiangping Chen, Sarah Harwell, Elizabeth D. Liddy, Lan He

Center for Natural Language Processing
Syracuse University, School of Information Studies
4-206 Center for Science and Technology
Syracuse, NY 13214-4100
www.cnlp.org
{diekemar, oyilmaz, jchen06, scharwel, liddy, lhe}@syr.edu

Abstract

This paper illustrates ongoing research and issues faced when dealing with real-time questions in the domain of Reusable Launch Vehicles (aerospace engineering). The question-answering system described in this paper is used in a collaborative learning environment with real users and live questions. The paper describes an analysis of these more complex questions as well as research to include the user in the question-answering process by implementing a question negotiation module based on the traditional reference interview.

Introduction

While there was significant early research in question-answering in the fields of logic and linguistics (Belnap 1963; Belnap and Steel 1976); automatic question-answering research has been largely driven by the Text REtrieval Conference (TREC), co-sponsored by the National Institute of Standards and Technology (NIST) and the Defense Advanced Research Projects Agency (DARPA). The purpose of TREC is to support research in the area of information retrieval by organizing a yearly large-scale system evaluation on a variety of retrieval related tasks (tracks). Although progress has been made since question-answering was first added as a track at TREC in 1999 (Voorhees 2000; Voorhees and Tice 2000), the research has largely converged on shorter fact based general domain questions. This means that having a relatively successful question-answering system tailored to the TREC question-answering task (Diekema et al. 2001) does not necessarily ensure success in question-answering outside TREC.

This paper will illustrate ongoing research and the issues we face when dealing with real-time questions in the domain of Reusable Launch Vehicles (aerospace engineering). Reusable Launch Vehicles (RLVs) are advanced launch systems, developed in a joint effort by NASA, the Air Force and the aerospace industry, to make space travel substantially cheaper. This particular domain, the actual users of the system, and the questions asked all demanded a change in question-answering strategy that is quite unlike the direction on which TREC has focused.

Copyright © 2002, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

Background

We have developed a QA system (Liddy 2001) with funding from NASA and AT&T for use within a collaborative learning environment for undergraduate students majoring in aeronautical engineering from two universities. The students are taking courses that are taught within the AIDE (Advanced Interactive Discovery Environment for Engineering Education). The students are able to ask questions and quickly get answers in the midst of their hands-on collaborations within the AIDE. The collection against which the questions are searched consists of textbooks, technical papers, and websites that have been pre-selected for their relevance and pedagogical value.

The students' questions are not typically simple factoid questions, but tend to be more complex and require more than bare answers, such as:

What are the changes made to the design of the Shuttle SRM since the Challenger Accident?

The system provides five short answers on the answer page. The student can then click on a link that provides access to the full document. In case the link is a dead link, or the student is otherwise having trouble accessing the page, a cached version is also provided. The system is currently undergoing user testing.

System Overview

The prototype of the CNLP AIDE question-answering system (see Figure 1) consists of four different modules: document processing, Language-to-Logic (L2L), Search Engine, and Answer Providing Passages. Document processing is done offline. When a user submits a question to the system, the question is first sent to the Language-to-Logic module, which generates the L2L query representation and identifies the question focus. The Search Engine

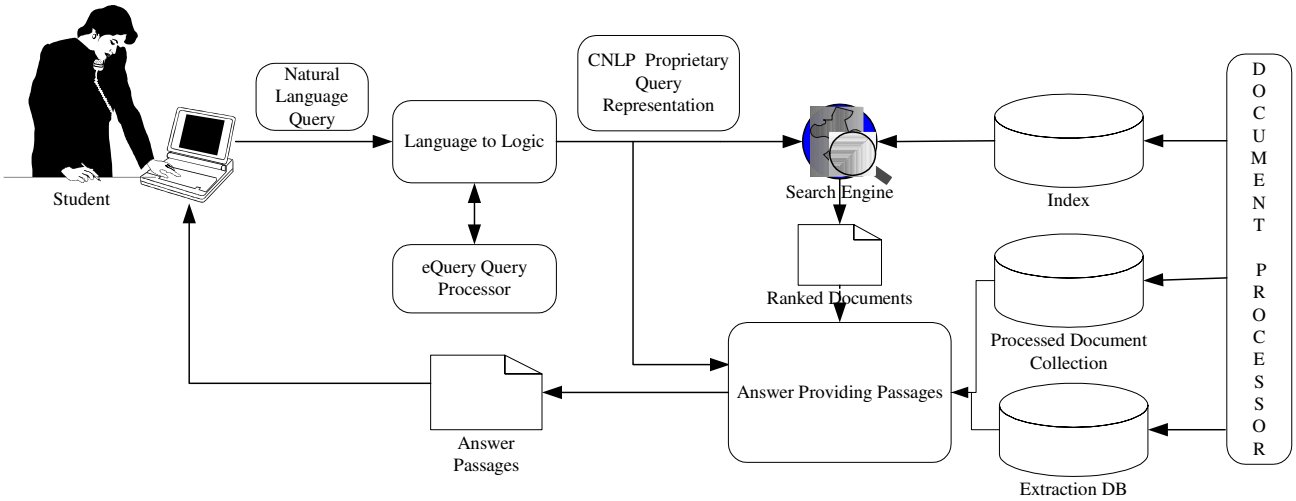

Figure 1: System Overview.

module then searches the index and returns the top 200 relevant passages. The L2L query representation, question focus, and the retrieved passages are passed to the Answer Providing Passages module, which returns the top five most relevant answer passages. Each of the modules is described in more detail below.

Document Processing Module

The documents in the system consist of chapters from textbooks, technical papers, and web sites. The document processing includes several stages such as preprocessing, splitting, and tagging. At the preprocessing stage, documents are converted into XML format. They are then divided into passages of approximately 100 words in size. These passages are sent to the information extraction server, which performs such NLP processes as stemming, part-of-speech tagging, and Named Entity detection. The passages are also indexed by our Search Engine Module for passage retrieval.

Language-to-Logic Module

During question processing, the system converts a natural language question into a logical query representation used for passage retrieval. Before creating the logical query representation, the L2L module carries out stemming, stopword removal, phrase and Named Entity recognition, and abbreviation expansion. It also determines the focus of each question used for answer finding. For example, the question *"What types of materials are used for TPS tiles in the space shuttle?"* results in the following output:

L2L: material* +TPS* ("Thermal Protection System") tile* "space shuttle"
focus: object, substance, materials

where the first line represents the logical form of the question and third line represents the question focus. Here the focus asks about substances, materials, or objects.

Search Engine Module

The system uses passages rather than documents for answer finding. The search engine indexes the passages after which they are ready for retrieval in response to a user's question. The 200 top ranked passages are returned by the system and sent to the Answer Providing Passages module.

Answer Providing Passages Module

The Answer Providing Passages Module takes the tagged passages that are retrieved by the search engine and identifies answer candidates based on the question focus. A weighting formula is used to assign weight to each answer candidate. The weighting formula takes into account the following factors: number of keywords occurring in the passage, the confidence level of focus assignment, and the distance between the candidate and the keywords. The process is repeated for all 200 passages and up to 5 passages with the highest weighted answer candidates are returned to the user.

The Questions

NASA questions differ from TREC questions in several respects. First, a NASA question is live, written by a student whose question can be ambiguous, or dependent upon implicit knowledge that isn't explicitly stated in the question. Real-time questions are often hurried and rife with malformed syntax and spelling errors. Due to the nature of the subject area, the NASA questions are complex, needing complex answers or sometimes returning information from

which the answer needs to be inferred once the answer-providing passages are read.

For example the simple question: *"How dsose the shuttle fly?"* (leaving aside its obvious typo) is so broad as to thoroughly confound a reference librarian, let alone an automatic question-answering system. Does the student wish to know that the shuttle flies upside down, i.e. the physical orientation of the space shuttle as it flies? Or is the student looking for specifications related to the way the space shuttle flies during its launch, the way it orbits when it arrives in space or its re-entry into our atmosphere? Or does the student need information about the way the shuttle navigates?

The question *"Do welding sites yield any structural weaknesses that could be a threat for failure?"* has no subject, i.e. it doesn't specify where or on what the welding sites are located. We can assume (as humans who know what course the students are taking) that the welding sites are probably located on the space shuttle, but the system is unable to make this assumption.

Another type of question appears simple, e.g. *"At what temperatures do liquid metals typically exist?"* The question answering system would typically look for "liquid metals", a particular type of verb and a temperature (determined by the L2L focus analyzer). However, the actual answer is much more complex. Melting points depend on the type of liquid metal, with binary liquids having a sharp melting point (e.g. mercury –39 C), liquid metals made of heavier elements having a lower melting point (unspecified) and alkali metals having melting points below 200 C. This answer can be found in one document, but over several paragraphs, and it is still not the complete answer—as it fails to specify the exact temperatures of liquid metals made of heavier melting points.

A fourth type of complex question found comparison of two different elements from two different documents where the answer has to be synthesized by the actual questioner. For example:

What advantages/disadvantages does an Aluminum alloy have over Ti alloy as the core for a Honeycomb design?

It is unlikely that the system will find a particular sentence or paragraph that will answer this question thoroughly. This type of question requires higher order thought processes that require synthesis and analysis of existing information within the document collection. To help the questioner, the system must be able to parse the question into different parts, e.g. return a page on the strengths and weaknesses of "Aluminum alloy" for Honeycomb design as well as return a document that talks about the advantages and disadvantages of "Ti alloy" for Honeycomb design. It will then be necessary for the questioner to deduce an answer from the pieces of returned information.

Analysis of NASA Questions

For our analysis we closely examined 406 questions that were asked of the system by students in the aeronautical engineering program. This analysis found these questions are similar in language usage to scientific writing generally. They are:

- Objective - personal pronouns seldom appear in the questions, and even if they do, are not very useful in representing the semantics of the questions
- Plain - the adjectives and adverbs used are necessary, not superfluous, modifiers and are used either to convey a certain feature or to specify a level.
- Accurate – the questions require certain prepositional phrases to convey the temporal, spatial or conceptual domain of an occurrence.

These questions present the following linguistic features:

1. A large number of domain-specific noun phrases, including Proper Noun phrases, and verb phrases.
2. There are clear syntactic patterns that can be used to categorize questions into classes.
3. These questions are comparatively longer, with complex syntax containing several prepositional phrases as modifiers.
4. Question focuses are complicated, but are identifiable based on lexico-semantic information.

Question Type Classification

To identify the focus of a question, the L2L Module first determines the question type (Chen et al. 2002). Whereas the TREC questions all fall neatly into one of our 14 categories, the NASA questions we have logged defy this organization entirely and thus require a different approach. Particularly, we have noted that there are a wider range of question types; fewer Proper Name entities; the noun phrases and verb phrases are domain specific, and; the prepositional phrases are important in specifying exactly what the questioner is asking for.

As shown in Table 1, eight different question types emerged from the questions analyzed. In addition, the question focuses of each question type were identified based on the lexical and / or syntactic information. Noun phrase, verb phrases, and prepositional phrases in the questions were categorized into classes with attached semantic relations. For this classification, a domain expert was consulted for the definition of phrases with which our team was unfamiliar. In this paper we report only on the question classification, details of our approach to focus identification and phrase analysis can be found in Diekema et al, 2001.

TYPE	DEFINITION	EXAMPLE
Wh-	Simpler factoid questions of the *what, when, where* type	*When was the concept of glass transition temperatures first discovered?*
Yes / No	Require a yes or no response, but may mask a complex inquiry	*Doesn't the simplification of the complex honeycomb design for the thermal protection system of a Reusable Launch Vehicle jeopardize the accuracy of results?*
How	Require an explanation.	*How are layers in TABI bonded together?*
Quantification	Looking for a specific amount, such as cost, weight, number, maximum, volume, etc	*What is the highest temperature the space shuttle undersurface experiences during its mission?*
Conditional	Inquiry indicates a condition that the answer needs to take into account. Indicated by phrases such as: *in addition to, aside from, other than, etc*	*Aside from contact of two tiles that can be damaging, are there any other reasons why insulating tiles on Reusable Launch Vehicles must be isolated from one another?*
Alternative	User provides several alternatives, one of which needs to be proven true, e.g. *A or B or C*	*Are Thermal Protection systems of spacecrafts commonly composed of one panel or a collection of smaller tiles?*
Why	Require an explanation	*Why all shear loads and twisting moments set to zero for the preliminary design phase of TPS?*
Definition	Looking for a formal or semi-formal definition of an element, process, material, etc.	*What is a liquid metal?*

Table 1. Classification of NASA Question Types.

Query Negotiation

Reference librarians have successfully fielded ambiguous or open-ended questions for years using the reference interview to narrow a broad question. One possible solution for a question-answering system faced with broad or ambiguous questions is query clarification, where the system asks the questioner for more information in order to return better results. We are currently exploring the possibility of utilizing reference interview theory to provide a framework for automatic query negotiation between the system and the questioner.

A reference interview has three parts: questioning the questioner, locating the answer, and returning the answer to the patron (Bopp and Smith 2001). The reference interview begins by restating the question in order to allow the patron to refine his or her thoughts and allow the librarian to understand the query better. The librarian might respond with an open-ended question, that is, a what, where or how question (Bopp and Smith 2001). For example, when the system is faced with a question as ambiguous as *"How does a shuttle fly?"* it could respond, *"What part of flying a shuttle would you like to explore?"* thus allowing the patron to rephrase his or her question and make it more specific.

We are currently investigating the potential of adding a user-system interaction step to the question-answering process. In this step the user will have the opportunity to refine his or her question during an exchange with the system in order to obtain better results for ambiguous and open-ended questions. We intend to base the automatic query negotiation on the model of the reference interview. Since the common reference interview is an actual conversation between librarian and patron, certain modifications of the question negotiation process are in order.

The main goal of the reference interview is to determine what information the user needs. As is true in a library setting, it is important to verify first that the question has been understood correctly. This active listening process requires paraphrasing the question back to the user to ensure question comprehension. After the information need has been established, follow-up interview questions might be in order to further clarify what the user is looking for. Finally, once the answer has been provided, the user is asked whether the answer is what he or she was looking for.

Receiving the Question

It is important to first make sure the question is spelled and formulated in an understandable way. The system will check question spelling and ask the user whether certain strange spellings were intended and suggest alternative spellings. At this stage the system will also ask the user to pick the full form of any acronym in their question.

Some questions are in fact multiple questions separated by a comma: "In reference to the plate geometries, is the "flat stiffened plate analysis approach" used as an assumption for simplification or will all thermally protected surfaces actually be flat?" In this case, the user will be advised that the system can only handle one question at a time and will be requested to ask the questions separately or otherwise rephrase the question.

Paraphrasing Questions

For some questions, the information need is easy to determine by the system. The question "*What is the weight of the space shuttle?"* clearly asks for a weight of a certain item. This type of question is currently recognized by our system. It is familiar with weight measures (i.e. tons, kilos, pounds) and can provide a short factual answer. It is therefore fairly straightforward to add an extra step where the system paraphrases a weight question: "Do you want to know how much the <OBJECT> weighs?" This narrow question requires a yes or no answer. If the user answers yes, the system can display the answer to the user and have the user specify whether the answer is satisfactory. If not, the system will then treat the question as more complex and proceed to the more complicated reference interview.

Paraphrasing the question becomes more complicated when questions are open-ended (*"Why must there be a buffer between tiles on the Thermal Protection System surface?"*), or ambiguous (*"How does an X-Ray spectrometer locate stress fields?"*).

For those questions that cannot be paraphrased as easily as the one about the weight, the question needs to be processed in more detail. During this process, information about the entities, events, and relations is extracted and presented in human-readable form.

The question: "Are Thermal Protection systems of spacecrafts commonly composed of one panel or a collection of smaller tiles?" will result in system feedback in the following format:

1. The subject of your question is a Thermal Protection System which is a system associated with a spacecraft.
2. You would like an answer back about a part.
3. The answer should have to do with one panel or smaller tiles.

The user can quickly see whether the system has understood the question. The system knows the answer is related to the Thermal Protection System and that the answer has to do with panels or tiles of a spacecraft. The user will be asked to read the information extractions, and apply corrections where necessary. Once it is clear the question has been understood correctly, the system proceeds to the next step in the reference interview.

Follow-up Questions

Based on the information that the system has extracted, the system can either decide to start answer-finding or, in case the question has been flagged as problematic, ask follow-up questions for further clarification.

In case the information extractions seem to be missing key information, the user will be asked to supply additional information pertaining to the question. For example, it might help to know where the *"stress fields"* might be located when trying to answer the question about the spectrometer. Once enough information has been provided, the system begins the answer finding phase.

Answer Satisfaction

If the answer is satisfactory to the user, the system interaction has come to an end. Otherwise, the user will be asked to rephrase their own question, and hints will be provided as to how to go about doing this. Also, the system could provide information about the collection and the possibility that the answer can simply not be found within it.

Architecture of an Interactive Question Answering System

Based on our question analysis and the theory and practice of reference interviewing described in the previous section, we plan to modify the Q&A strategy by integrating query negotiation into the current system. In our new design, the user is an active component of the whole Q&A process. The execution of any procedure is determined by the user's response and/or judgment to the alternative questions or the answers suggested by the system. Furthermore, in order to facilitate the interaction between the users and the system, information about questioning techniques, questions that the system can better answer, and the terminology from the collection will be provided to the user. Figure 2 shows the system architecture based on our new Q&A strategy.

The modified Q&A process as shown in Figure 2 can be described as follows: The domain specific documents are collected, preprocessed, split into sections, indexed, and tagged offline as before. A user who comes to the system can ask a natural language question immediately, or go to the pages containing help information first and then submit a question. The system will conduct spell checking and focus identification after receiving the question. Next, the query negotiation process will be activated to refine the question. The user will have total control of the negotiation process and can stop at any point. The finalized question will then go through passage retrieval and answer finding. At last, the system will return 5 possible answers and display them to the user. The user will be given the opportunity to evaluate these answers. The system will provide

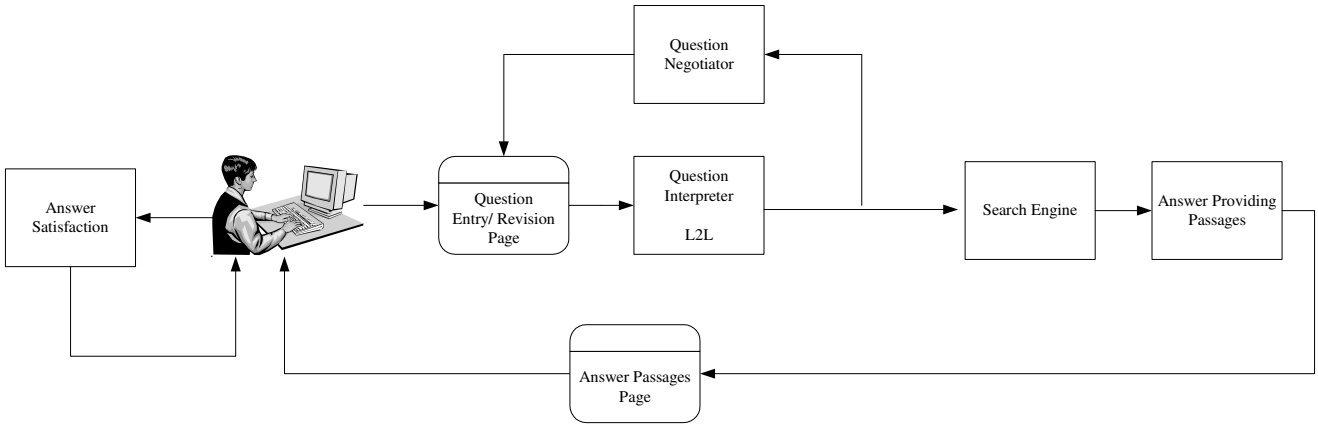

Figure 2: Modified System Architecture.

several options if the user is not satisfied. These options will include rephrasing the question for another search attempt, asking an expert for the answer, or returning answers from other resources.

The new system will have the following modules: 1) Document Processing; 2) Question Interpretation & Negotiation; 3) Search Engine; 4) Answer Providing Passage; and 5) Answer Satisfaction. The Document Processing, Search Engine, and Answer Providing Passage modules stay the same as the current system, which have been described earlier in this paper. Question Interpretation & Negotiation and Answer Satisfaction modules are new and are described in detail below:

Question Interpretation & Negotiation module

This module has three components: Question Entry/Revision Page, Question Interpreter and Question Negotiator.

Question Entry/Revision Page

On this page the user may type in the question, review the rephrased question returned by the system, or go to the help page for more information. This page will be brought up at each interaction between the system and the user. The help information includes a description of question techniques, an explanation of what kinds of questions can be better answered by the system and the sample questions, and a collection of domain specific terms. The user can ask a natural language question immediately, or go to the pages containing help information.

Question Interpreter

This procedure coordinates spell checking and Language-to-Logic conversion. A lexicon based on WordNet 1.7 will be used for spell checking. If strange spellings are detected, the system will ask the user whether they are intended and suggest alternative spellings. At this stage the system will also ask the user to pick the full form of any acronym used in the question.

The Language-to-Logic module is the same as described previously. This procedure will determine whether the question has a focus as determined by previously written rules. If it cannot decide the focus for the question, it will pass the process to the Question Negotiator.

Question Negotiator

Question Negotiator begins the conversation between the system and the user. Each time the system cannot understand the focus for a question, it turns to the Question Negotiator for help in refining the question. The main goal of the Question Negotiator is to determine the needs of the user and translate those needs in order for the system to understand them. In each interaction, the system will carry out one of the following actions; it will rephrase the question, ask for more information from the user, or end the interaction to start the answer-finding process.

If the language to logic conversion process returns a question missing key information (e.g. there were no entities extracted or no focus), the system will then ask the user to supply additional information pertaining to the question. The user will be able to alter the question as understood by the system. For example if the system returned the wrong subject, the user can type over the incorrect subject and resubmit the question with the corrections.

Answer Satisfaction

The user is then shown a page that allows feedback about the answers provided. If the answer is unsatisfactory, the user will be provided with three options: 1) the user can choose to return to the Question Entry/Revision page to begin the process with an alternative question 2) the user will be able to ask the question using the resources from the Web or 3) the user can choose to get help from a subject specialist.

Conclusion

We have found that the question-answering paradigm driven by TREC does not necessarily prepare question-answering systems for real-world applications. The nature of queries generated by real users, as well as the breadth vs. narrowness of what constitutes a useful answer, diverges substantially from the TREC setup. Answering complex questions requires a different approach. Our current NASA research attempts to incorporate the reference interview from library science into the automatic question-answering process. By helping the user to reformulate complex questions we hope to improve the question-answering process.

Acknowledgments

This research was supported by funding for National Aeronautics and Space Administration under NCC1-01-004, the State of New York, and AT&T.

References

Belnap, N. D. 1963. An analysis of questions: Preliminary report. Scientific Report TM-1287. Santa Monica, CA.

Belnap, N. D.; Steel, T. B. 1976. *The logic of questions and answers*. New Haven, CT., Yale University Press.

Voorhees, E.M. 2000. The TREC-8 Question Answering Track Report. In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*, 77-82. Gaithersburg, MD.: National Institute of Standards and Technology.

Voorhees, E.M.; Tice, D. 2000. The TREC-8 Question Answering Track Evaluation. In *Proceedings of the Eighth Text REtrieval Conference (TREC-8)*, 83-105. Gaithersburg, MD.: National Institute of Standards and Technology.

Diekema, A.; Liu, X.; Chen, J.; Wang, H.; McCracken, N.; Yilmazel, O.; Liddy, E.D. 2001 Question Answering: CNLP at the TREC-9 Question Answering Track. In: *Proceedings of the Ninth Text REtrieval Conference (TREC-9)*, 501-510. Gaithersburg, MD.: National Institute of Standards and Technology.

Liddy, E.D. 2001. *Breaking the Metadata Generation Bottleneck*. Joint Conference on Digital Libraries. Roanoke, VA., June 25, 2001. http://www.cnlp.org/presentations/

Chen, J.; Diekema, A.R.; Taffet, M.D.; McCracken, N.; Ozgencil, N. Ercan; Yilmazel, O.; Liddy, E.D. 2002 Question Answering : CNLP at the TREC-10 Question Answering Track. In *Proceedings of the Tenth Text REtrieval Conference (TREC 2001)*, 485-494. Gaithersburg, MD.: National Institute of Standards and Technology.

Bopp, R.E., Smith, L.C. 2001. *Reference and information services: An introduction*. 3rd Edition. Englewood, CO. Libraries Unlimited.