Investigation of Mechanisms to Improve Question Answering

Ph.D. Thesis

Stanley J. Mlynarczyk

smlynarc@students.depaul.edu

DePaul University

School of Computing, College of Computing and Digital Media

Chicago, IL 60604

Winter, 2009

Committee In Charge:

Professor Steven Lytinen – Chair
Professor Hani Abu-Salem
Professor Peter Hastings
Dr. Boris Zibitsker

Dissertation of Stanley J. Mlynarczyk is approved:

Chair Date

Abstract

Investigation of Mechanisms to Improve Question Answering

By

Stanley J. Mlynarczyk

Dr. of Philosophy in Computer Science

DePaul University

School of Computing, College of Computing and Digital Media

Professor Steven Lytinen, Chair

This document is describes research in Internet Frequently Asked Question (FAQ) answering and Natural Language Processing (NLP). The research seeks to enhance the accuracy of FAQ file answer searches of an existing FAQ answering system. Secondly, this research seeks to increase breadth through implementation of a WEB search and classification mechanism. Accuracy improvements are intended to provide an increased ability to correctly match a user question with a corpus of previously asked questions residing in Internet FAQ files. Breadth increases will be accomplished through an automated approach to scanning and categorization of Internet FAQ files.

To:

My family: Yvonne, Greg and Chris who went through this experience with me.

Table of Contents

1 Introduction

The storage and retrieval of information has gone through a very large change in the recent past. This is partly due to the increased amount of information available and also to the fact that we now have a medium that facilitates information sharing (the Internet). Prior to computer based information storage and retrieval, documents were categorized through manual cross-referencing systems (e.g., title, author, subject, etc.). Computer systems and the Internet have greatly increased both the amount of data and access mechanisms for information retrieval.

While categorization of documents provides a coarse information retrieval mechanism, it is often insufficient for answering queries seeking specific answers to questions. In other situations, there may be too many potential documents retrieved with few indications as to which documents most closely match a user's needs.

Internet information poses a challenge because there is generally no categorization upon which to draw. A user who types in a few keywords into a search engine may occasionally be rewarded with a link to exactly the information that fits their needs, however, it often takes careful consideration about the choice of keywords to use.

The Internet contains various types of documents that might benefit from different approaches to document search. Examples of document types might include: Hyper Text Markup Language (HTML), text documents, binary and FAQ files. Not only is the content and format of such files different, the "amount" of information and its potential to be relevant to a user varies significantly. Only the content bearing text is most likely useful to a user; thus a search engine would typically filter out and discard HTM portions of the document. FAQ files pose a unique challenge in search, i.e., should the search focus on the "question", "answer" components or both? The question then becomes "what mechanisms should be used for searching a given content type?". The focus of this document is to describe mechanisms for search and retrieval of information contained within Internet FAQ files.

1.1 The Importance of Document Retrieval

With the widespread use of computers by the general public, especially due to the Internet, the volume of existing and newly added available information is so large that it is becoming increasingly difficult to locate information even with the introduction of the latest search engines. It is of little use that information is stored and available if retrieval methods cannot locate it.

The issue is not limited to the Internet. Even libraries that categorize books using computerized catalogs are not immune from a weakness in this area. While it may be possible for a user to search for a book by title, author, topic and keywords, it is still not likely that a user wanting to

answer a question such as "What kind of yeast should I use for a nut brown ale?" will be successful.

While keyword searches might bring up various books on brewing beer, there is no guarantee that a given book will contain an answer to the question. Even if a user locates a potentially helpful book, there is still the need to find the needed material within the book; it might be a single sentence in a table of ingredients for various recipes. In some cases one may enter reasonable keywords but a potentially useful book may be indexed using synonyms of the keywords the user has issued, resulting in a failed match. Lastly, the returned answer set may contain a large number of irrelevant matches.

These are only a few of the hurdles users face when attempting to navigate through the immense amount of information available. As information is constantly being added, the problems of efficient and effective retrieval also continue to grow. The need for tools to help in the management of information is critical. Without adequate tools, the benefits of having the information are greatly diminished.

1.2 Traditional Document Retrieval Approaches

Traditional methods in document retrieval generally attempt to search a document corpus using a word frequency based mechanism. This approach has its foundation in research (Gerard Salton, 1971) (more recently Korfhage, 1997) that focused on word frequency as a primary mechanism. Given a keyword(s) that a user wishes to search for, a large document corpus is scanned for the words appearing in each document. The thought is that with a "good" keyword (key words such as "the, it, are, he, she, can, etc..." would not be good keywords but ones such as "whale, vitamin, surgery, plastic" might be), a document retrieval system can then consider the number of occurrences of the keyword(s) within each document to determine its relevance.

Word frequency approaches to document retrieval also consider the relevance of words contained within a document. If a word occurs frequently in a document, this is generally a good indicator of document relevance. If a majority of documents within a document corpus also contain this word then its value as an indicator of relevance is questionable. Word frequency based retrieval seeks to identify those indicator words within a document that are unique to that document or rarely used within other documents. When a user wishes to search for a document using a set of keywords, an attempt is made to give preference to those keywords that match the indicator words.

Search engines that make use of word frequency can work quite well in identifying potentially relevant documents given a set of keywords. The problem often is that too many potential match candidates are returned in a search and the user is then faced with the problem of sifting through the returned set of links or restating the search criteria. While a user can enter a "question", generally most search engines discard the non-content bearing words, (e.g. prepositions, pronouns) and focus on the remaining words. Still, Internet search engines can do a credible job

depending on the skill of the person entering the keyword search criteria.

1.3 Limitations of Current Document Retrieval Approaches

There are situations where word frequency based Information Retrieval (IR) methods are not sufficient. One example is where a user wishes to ask a question and have only the "answer" returned instead of the document containing the answer. Such a system is faced with a very difficult challenge of extracting the question context, which may be part of a complex sentence structure, while also considering synonyms for the verbs, nouns and modifiers.

A second situation is where a user asks a question and the document to be matched against is a single sentence. This is the case where a user is asking a question that requires a system to find a match within a set of previously asked question files.

A third problem is that a user's selected keywords may be synonyms of words used in a desired document and thus will be missed by word frequency based IR methods that do not consider synonyms. Many of the approaches presented at the recent Text REtrieval Conference (TREC) (TREC-10, 2001) have attempted various mechanisms that make use of synonyms as well as word clustering (TREC-10) and focus, and there appears to be a general interest in pursing deeper level semantics.

These situations expose two issues. The first is that to find an "answer" to a question, a system must be able to determine the "intent" of the user's question and to determine the intent of the target answer. This is because simply returning a promising document in its entirety does not fulfill the user's requirements.

A second issue is that frequency based IR techniques work well when the amount of text in each document and in the document corpus is sufficiently large to support word frequency analysis; single sentence documents may not have enough words to even allow a word match to occur. If the words' intrinsic meanings cannot provide the means to yield a match then perhaps sentence semantics may fare better. Examples of a semantic approach might be synonym consideration and inter-word relationships.

1.4 FAQ File Question Answering

Internet FAQ files contain information about previously asked questions and resulting answers to these questions. The sources of these files vary but include USENET FAQ files as well as FAQ files that businesses have generated that are specific to products they sell and support. Typically, such files are organized as text files that follow a general pattern of {question, answer}; there is usually a non-standardized numbering scheme. While many FAQ files do follow a format convention, there is no standard for how text is organized within a FAQ file.

FAQFinder (Burke, et al., 1997) is a tool that was developed to provide answers to user questions through the recall of previously asked questions residing in Internet FAQ files, primarily USENET FAQ files. Instead of "generating" responses to a user question or to user supplied keywords, the FAQFinder approach is to analyze the user natural language query and attempt a match with questions that might have been asked and answered previously. The goal is not only to look for exact word matches, such as might be done with word frequency based document retrieval, but to return a question match based on semantic analysis of a user question and previously asked questions.

While the functionality of FAQFinder is in essence a form of document retrieval, traditional approaches successfully used in document retrieval, have some limitations when an attempt is made to use these methods in FAQFinder. This is not to say that such methods are useful in FAQFinder, rather, it has been empirically determined that a combination of approaches appears to yield the best results.

FAQFinder is an ongoing effort to test various Artificial Intelligence (AI) concepts in Natural Language Processing . The research described in this thesis proposal is an effort to increase accuracy and extend FAQFinder's functionality while analyzing effects of the implemented changes. Past efforts have successfully demonstrated the effects of grammar/parsing, traditional text retrieval approaches (document word frequency), word sense tagging (Brill, 1994), and use of semantic approaches such as WordNet (Miller, 1994).

Later work in FAQFinder (Lytinen et al, 2000) yielded conclusions that document retrieval accuracy can be improved through disambiguation of word senses and agrees with other research in this area (Sussna, 1993), (Kurohashi, 1998), (Resnik, 1993). All of these efforts on FAQFinder have produced promising results. The research described in this proposal seeks to build upon previous IR research and to find mechanisms to enhance the accuracy of Internet question answering.

1.5 Contributions Of This Dissertation

My research has pursued various options for increasing FAQFinder accuracy; these attempts are described within this dissertation. The research has yielded positive results in terms of increased FAQFinder accuracy and has demonstrated applicability to other text search tasks.

My efforts began with simple attempts to hone existing methods in an attempt to generate positive results. One of these was to increase the number of grammar rules used by the FAQFinder parser. Another was to increase the effectiveness of an early attempt at subject/verb/object (SVO) identification. These early efforts were important in that they revealed the extent of the problems that must be overcome to yield accurate results in question answering. While small positive increases in FAQFinder accuracy could in some cases be realized, these increases incurred a great deal of analysis and research. Furthermore, a given change might benefit a specific data and question set and then be of little value to another. These early attempts led me to the following observations:

- **Existing FAQFinder methods are effective to a point.** *Word frequency approaches are quite good at narrowing result sets.*

- **Existing FAQFinder methods produced quick positive returns.** *Word frequency is relatively easy to implement.*

- **Existing FAQFinder semantic approaches can sometimes be helpful.** *Word distance does help when the questions are non-ambiguous.*

- **Further increases in accuracy are limited due to ambiguity of questions.** *Real-world questions are often terse and ambiguous, which poses a challenge to methods that lack human level discernment.*

For there to be any hope of substantive increases in FAQFinder accuracy, it became apparent that ambiguity needed to be addressed. There were two areas of FAQFinder that seemed to hold promise, although early results were not significantly positive. The first was targeting of the subject, verb and object of a question. This seemed logical for human level understanding of a question, unfortunately, the early attempt with SVO did not have a sufficiently accurate way to properly identify the subject, verb and object.

The second area that seemed to hold promise was the answer component of a FAQ file. The answer components usually have a substantial amount of text compared to a user question. Early attempts at using the answer component failed to provide any tangible increase in accuracy. The approach taken by these early attempts were based on word frequency. Given that FAQ files are generally grouped such that questions/answers within the same domain comprise a FAQ, it is understandable that targeting of the answer component might not be able to discriminate between answers.

For there to be any hope of increasing FAQFinder accuracy, the difficult task of syntactic (and possibly semantic) analysis of the question and answer component needed to be undertaken. This type of approach was not the preferred method when this research began. There was still hope within the text analytics research community that there was a heuristic approach that could be found to produce good match results. My own conclusion after spending large amounts of time on this problem within a FAQFinder domain, was that deeper syntactic and semantic methods were necessary to achieve positive results. Thus began the effort to perform syntactic analysis on the user question and on the FAQ answer component.

Preliminary results with syntactic analysis were somewhat encouraging (refer to chapter 4). For the first time since I began working with FAQFinder, I was able to demonstrate some tangible accuracy benefit. These first steps were for one type of question ("How" questions), yet, there was a measurable positive contribution to FAQFinder accuracy. Encouraged by this success, I decided to pursue additional work in this area to include the "What" and "Who" question types.

Inclusion of the "What" and "Who" question types produced excellent results (refer to chapter 5). The research goal of increasing FAQFinder accuracy was met and exceeded. The new functionality was named "QUANDA" or (Question AND Answer). Additionally, I had the opportunity to

test QUANDA within a general text domain (TREC – Text Retrieval Conference) and against three web search engines (GOOGLE, YAHOO, ASK). QUANDA has generated very positive results in all cases.

QUANDA has proven to provide excellent accuracy improvements for FAQFinder and general text search. Research into QUANDA has shown that the hard task of syntactic analysis has yielded very positive benefits to accuracy. As a research direction, recent TREC (2007) question answering track publications confirm interest in a syntactic approach. My research into QUANDA has added to the body of knowledge in an early adoption of a syntactic approach to question answering. QUANDA as a tool, can provide benefit to future research in question answering and within practical applications of text search technology.

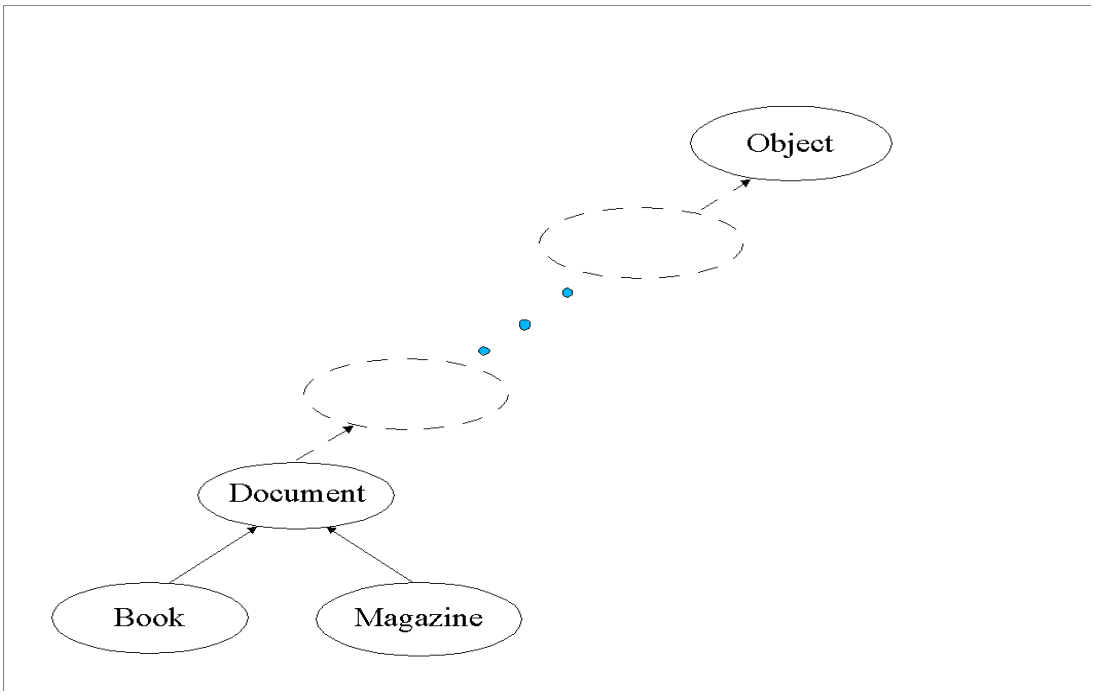
QUANDA accepts a user question, classifies it according to the question type, analyzes the question and determines parts-of-speech for question components. In addition to parts-of-speech, QUANDA identifies the subject, verb and object if the question contains these. Once a question is classified and analyzed, QUANDA attempts to match the question form within a question ontology. The ontology provides clues about potential answer forms that might satisfy the question. These clues (there may be many) are then used to match against a body of text that might contain a potential answer for the user question. As with the user question, QUANDA performs a similar analysis of the potential answer text in its attempt to match a question. QUANDA is comprised of a question classifier, question matcher and interfaces to a tagger, Wordnet and an integrated dictionary.

1.6 WordNet

WordNet (Miller, 1993) is an important part of FAQFinder's functionality. FAQFinder depends on semantic information to make judgments about sense of words it encounters in questions asked by users, and those residing in FAQ files. For example, the word "*bat*" and "*fly"* have more than one meaning (polysemy); both can serve as a verb or a noun in addition to there being multiple meanings for each. The noun form for "*bat*" can refer to a flying mammal or a baseball bat in addition to its verb form denoting an action, as but a few examples of this word's versatility. Because FAQFinder must match single sentences, it is vital that the meaning of every word in a sentence be correctly identified.

WordNet on its own is not capable of determining a given word's sense within a sentence as it is basically a repository of words and relationships between words. The actual sense determination belongs to FAQFinder. FAQFinder depends very heavily on the information that WordNet provides.

Since WordNet was first made available, there have been various additions to its functionality. Some of the key functionality that FAQFinder is dependent on is WordNet's hierarchical view of data (lexical inheritance). This hierarchy takes the form of a tree where each level within the tree points to higher and more general levels/concepts. For example:



In the above graphic, book and magazine are both a type of document. Document then is a generalization of the two words and it itself is a subcategory of a higher level term until the top of the tree is encountered.

Not all words within WordNet belong to a single tree, rather, words are organized by category. WordNet supports 25 categories of nouns, some of which are: act/action, animal, artifact, attribute/property, body, cognition/knowledge, etc..., 15 classes of verbs and 3 classes of adjectives. Lexical inheritance is used within FAQFinder when it attempts to determine how closely two words are related. This is accomplished by determining the lexical distance between two words in the inheritance tree (see the graphic above).

WordNet contains much more functionality than lexical inheritance. Functionality supported by WordNet includes:

- Synonyms
- Antonyms
- Hyponyms
- Hypernyms
- Meronyms
- Holonym
- Adjectives
- Verbs
- Verb Phrases
- Causation

Hyponymy implies a child class relationship to a word, e.g., tree might be hyponym of plant. Hypernym on the other hand implies a parent relationship to a word, e.g., a plant would be a hypernym of tree. A meronym defines a whole/part relationship between words. For example an automobile may have many parts (wheel, engine, tire, brake, windshield, etc...). Such words would be meronyms of the word automobile and automobile would be a holonym of the parts that comprise an automobile.

FAQFinder takes advantage of only a small subset of WordNet functionality. This leaves open the possibility of using additional WordNet features to improve FAQFinder. One issue with the implementation of any new functionality is the computational cost. Parsing of FAQ file answer text is one example of a computationally intensive task. This is not to say that performance issues would prevent implementation of additional functionality, only that performance would need to be taken into consideration (FAQFinder currently preprocesses much of the FAQ file question parsing already).

1.7 Brill Tagger

The Brill Tagger serves a very useful purpose in FAQFinder. The tagger is responsible for analyzing the words in a sentence and then tagging according to part-of-speech. This is quite useful within FAQFinder as it eliminates the need for computationally intensive analysis potentially involving WordNet.

The Brill Tagger accomplishes the tagging task efficiently using a set of heuristics. These heuristics look at words within a sentence and determine whether there might be a positional relationship between two words where this may provide a clue as to their part-of-speech. For example, there might be a rule stating that the word following an "a" is either an adjective or noun. Similarly it might be the case that where two words having noun senses are adjacent, the first word will be an adjective (the tagger first assumes all words are nouns and then uses heuristics to determine the final part-of-speech). Other examples of rules might be that words ending in "ed" are verbs or that words that begin with an upper case letter are proper nouns, or that a word is an adverb if adding the suffix "ly" results in a valid word.

The tagger does offer the ability to be retrained. It comes with a set of tools that allow it to accept a new vocabulary then "learn" a new set of rules. By default the tagger is linked to use the Brown and WSJ lexicon. The lexicon is a file of words that have been extracted and then tagged according to the parts of speech that an individual word might assume. For example, the word "preliminary" is tagged as:

preliminary JJ NN

which is an indicator that it may either be tagged as an adjective or noun.

2 FAQFinder

FAQFinder's focus is much narrower than a general WEB search engine. Unfortunately this does not mean that the issues FAQFinder must contend with to realize its functionality are any easier. In many ways the implementation issues of a question answering system are much more difficult than a word frequency based search mechanism. The reason for this is that FAQFinder must deal with a much smaller amount of text (a single sentence). While on the surface this may appear to be a plus, in reality it forces the need to consider a much deeper and difficult level of analysis based on the meaning of a single sentence as opposed to simple textual matches.

This is not to say that traditional word frequency based approaches are not used or are not appropriate; they are and FAQFinder has been quite successful using them. There are many sentences that cannot be handled through a straightforward word frequency matching approach and this will be addressed in detail further in this proposal.

An interesting thing to consider in any attempt to create a retrieval system to handle FAQ files, where the amount of text to be matched is so small, is the following: should the system focus on matching the question component or should it instead focus on the answer part? An answer to this question is not always clear. Ideally, one would want to find the "exact" question that someone else had previously asked and this might (generally) provide a good indication that the answer text is relevant. This situation, while it may occasionally occur, is rare.

In considering an approach that targets the answer component of a FAQ file, this appears to be more closely aligned with traditional word frequency based document retrieval. While this may seem like a worthwhile approach, FAQ files are arranged by topic and a single file on "brewing beer" may contain many questions related to this topic.

The difficulty with using a word frequency approach on the answer component is that because all of the questions in the FAQ file are about the same topic, in general, the answers in the files will contain the same words. So where the word "wort" might be an effective indicator word when attempting to find a single document on beer brewing within a large collection of general documents, it loses its effectiveness in a FAQ file on brewing.

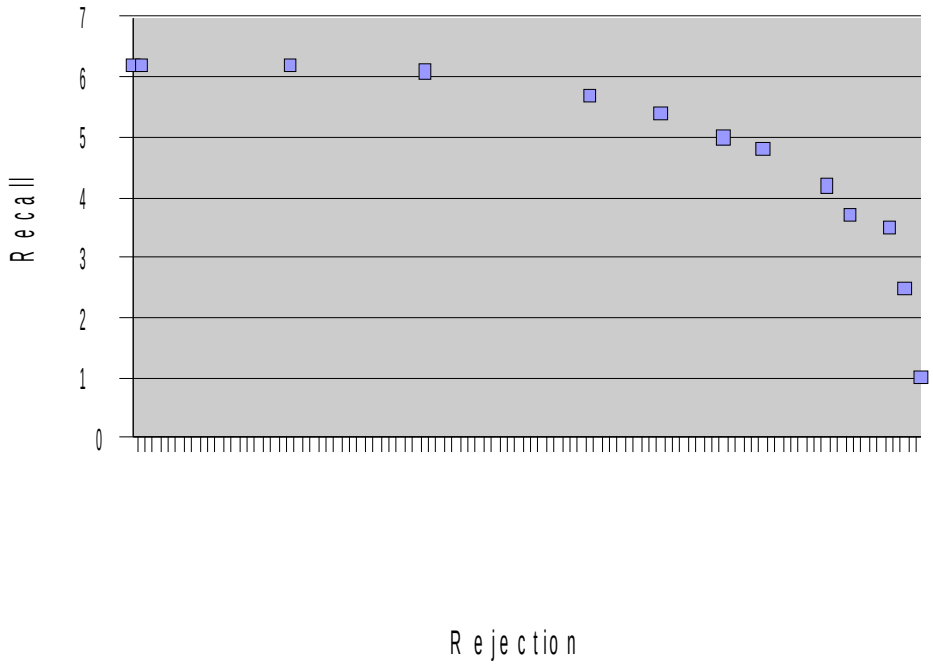
Past efforts in attempting to use the answer component have yielded mixed results. This does not mean that considering the answer component is worthless or that only the question component should be considered. Perhaps the key lies in making better use of one or both of these through different mechanisms. Perhaps taking a closer look at the information encoded in the text may provide a more effective means to know the user's true informational needs and provide a more accurate mechanism to match against a FAQ file's Question/Answer pairs.

The current level of FAQFinder accuracy is based on a combination of approaches: WordNet, Brill Tagger, and document word frequency analysis. Given the current level of accuracy, the research proposed in this document is an attempt to consider lessons learned by other researchers, both those directly and indirectly involved with FAQFinder, and extend that research in a way

that increases FAQFinder's accuracy. This effort will include an attempt to enhance existing techniques (grammar enhancements for example) and investigation of semantic analysis techniques.

FAQFinder accuracy has historically been measured by running a set of preselected questions against a pre-tagged set of 36 FAQ files. There are 123 questions that are considered "answerable" and 62 that are considered "unanswerable".

The following graph depicts the results prior to any modifications made as part of this research:



The graph depicts a relationship between recall and rejection where the Y-axis is recall and X-axis is rejection. "Recall" is a measure of the percentage of questions for which a correct answer was identified. A recall measure of 100% indicates that a correct result was returned for all of the questions asked by users. This differs from the standard measure of recall, which for document retrieval usually measures the percentage of relevant documents that a system retrieves. Previous FAQFinder researchers felt that this modified definition was a more relevant measure.

"Rejection" is an indicator of the selectivity of the system. A Rejection rate of 100% would be achieved if all of the questions users asked returned only valid results, and the system never made the mistake of returning an incorrect result.

In the above results, maximum recall is achieved when the rejection is "0". Ideally, both recall and rejection would be 100% which would mean that all answerable user questions were correctly associated with an appropriate FAQ file, and that none of the unanswerable questions were "incorrectly" associated with any of the test set.

While FAQFinder accuracy has been measured using the above mechanism, the current research will attempt to expand the accuracy measurement to include new and additional questions and data. One issue with using a static set of test data for too long is that there is the possibility that modifications to functionality might become specific to a given data set and not accurately reflect system performance in general. This is a labor intensive problem to resolve because the creation of a test data set requires manual effort to determine what an appropriate system response should be to any given user question.

As can be seen from the above graph, recall falls off significantly as rejection increases. Reasons for this are many. Findings by other FAQFinder researchers along with preliminary investigation into FAQFinder's accuracy has revealed the following:

1. Some questions in FAQ files are vague to the extent that even a human would find it difficult to understand what is being asked without prior explanation of what the subject area is. For example, the questions: "How do I start? What equipment do I need?", "What do oil names mean?", "Is CFS fatal?".
2. The Brill Tagger serves a very important function in FAQFinder because of its ability to "pre-tag" words according to part-of-speech. Unfortunately, mistakes in tagging will cause problems for down-stream semantic functionality within FAQFinder.
3. Acronyms pose a problem because these are not always tagged correctly as has been discovered by analysis of FAQFinder performance (such words may also not be defined in WordNet). (Burke et al., 1997) also stated:

   "Another category of failures comprises questions that contain domain specific knowledge. For example, the question "What is the best counting system for blackjack?" is answered under the question "What BJ counting system is the most effective?" which is not found by the system because of its inability to tell that "BJ" is a FAQ specific abbreviation of "blackjack" ".

   Acronym support in FAQFinder appears to be an important element.
4. There is no support for names of (persons, places, things, organizations), dates, currencies, numbers, products, slang.
5. FAQFinder performs matching of user questions to the question part of Internet FAQ files, ignoring the answer part. In some cases, the question component of a FAQ file contains little information that might identify the contents of a FAQ answer and it seems logical to consider the answer component. Attempts at considering the answer component have not produced tangible improvements. These previous efforts added answer text to the question component for consideration during computation of term weights for a user question. More recent research, (Chen, 2001) for example, approaches the problem by attempting to find text within a document that might be a specific "answer" to a user's question. Such an approach, while

more difficult than computing a term weight using Inverse Document Frequency Weight (IDF) (Korfhage, 1997), may hold more promise in the consideration of answer text.

6. Punctuation is a problem because users sometimes quote words and this is currently not handled properly in FAQFinder (a leading quote for example becomes part of the word). A similar problem occurs with parens "()". Hyphens are also problematic in that they are currently not stripped and two words separated by a hyphen are treated as a single word.
7. Multi-words are not recognized as a concept/term and thus terms such as "credit card" and "spark plug" are not given the special meaning they might deserve to maximize the possibility of a correct match. An example of why this might be a problem is in the following sentences: "What is the best spark plug for my Olds Cutlass?" and "Why do I get spark every time I plug in my lighter in my Olds Cutlass?". In the case of "spark plug", FAQFinder does not know that the terms "spark" and "plug" are different concepts as used here (WordNet does support phrases).
8. As with multi-words, phrases are often not recognized for their true meaning within a sentence. This can occur if the collective meaning of words within a phrase implies a meaning apart from the constituent words, e.g., "in the bag", "in hot water".
9. There is no support for causal relationships. In some cases there are strong causal relationships between sentences but FAQFinder has no way of knowing of these. For example: "Is HIV contagious?" and "Is AIDS contagious?", "How do I make beer?" and "How long should wort be allowed to cool?". In both of the above examples, we would want FAQFinder to indicate a match. The problem is that there are no causal relationships between HIV/AIDS and wort/beer.
10. There is no support for multiple sentence questions. A multiple sentence question will fail to be parsed. Even if support for multiple sentences were implemented in the parser's grammar, there is no linkage between referenced concepts between sentences.
11. Pronouns in sentences pose a special problem for FAQFinder especially in attempts at semantic based disambiguation. This is because the matching process has to deal with very short amounts of texts; basically one sentence against another. Should either the subject or object in a sentence resolve to a pronoun, the potential for a semantic match with another sentence diminishes significantly. There has been some success by other researchers to resolve pronouns (Cardie, 2001), (Azzam, 1998).

The combination of the above shortcomings causes difficulties because, while attempts to increase the FAQFinder's semantic capabilities may increase recall across parts of the recall/rejection graph, this however also potentially increases undesirable matches.

2.1 Hypothesis

Given the issues discovered during initial research into FAQFinder, there are a number of ways that accuracy might be improved. All of these focus on a core theme, which is to enhance FAQFinder's ability to perform a deeper level of question analysis. The Hypothesis that the current research is based on is:

Hypothesis: A combination of semantic analysis techniques and other enhancements can improve FAQFinder's accuracy.

It needs to be clarified that FAQFinder currently makes use of semantic analysis techniques. The approach taken by this research is:

1. Analysis of the limiting factors to existing semantic techniques.
2. Attempt FAQFinder modifications that will improve FAQFinder accuracy both generally and specifically in the semantic analysis functions.
3. Implement additional features that might improve FAQFinder accuracy. An existing semantic approach that focuses on the identification of sentence subject/verb/object will receive special attention in this area; the goal being to increase recall for all points on the recall/rejection curve.
4. To extend FAQFinder functionality to work on a broader corpus of Internet data and determine whether the implemented techniques perform similarly to the current small set of data.

Justification:

Current research into document retrieval falls into one to two categories; one focused on the data itself and the other on semantic based methods. Methods that focus on data, make use of term/document frequency. Methods that attempt to disambiguate text, do so by analyzing the meanings of words. FAQFinder is one example of such an approach, as are efforts by (Mihalcea, 1998 and 1999), (Harabagiu, 2001), (Chen, 2001) , (Moldovan, 1999 and 2000), (Azzam, 1998), (Cardie, 1999), etc...

Frequency methods work well when the amount of data is large enough such that reasonably accurate word probabilities can be determined. Such probabilities may yield an indication of a document's relevance in a user query. They may also yield statistics indicating the likelihood of words being co-located within a sentence or document or a given word's part-of-speech. Furthermore, given a user query, if that query is unambiguous, in that the terms used are ones that conform to the term probabilities that have been calculated, then frequency-based methods can work well.

If the assumptions made for frequency based methods are not valid, either because of data limitations or because of query ambiguity, accuracy may suffer. For this reason (evidenced by the large number of semantic based methods used in recent TREC proceedings), recent approaches for document retrieval do attempt at least some measure of semantic analysis, if for no other reason than to categorize a user question type, e.g.: (Roth, 2001); (Buchholz, 2001); (Chen et al., 2001); (Alpha et al., 2001); (Brill et al., 2001).

A semantic approach is not without problems. First is the need to determine the POS (Part Of

Speech) of individual words. Tagging functions such as the Brill Tagger (Brill, 1994) have demonstrated very good accuracy (96%).

Another approach used by some semantic methods is the parsing of data to determine the role that words assume within a sentence; such as whether a noun might be a modifier or whether a group of words might be part of a phrase or have a collective meaning. Accuracy of parsers can vary with sentence complexity. An assumption that a sentence is grammatically correct or that it conforms to an expected grammar may not be valid, resulting in an incorrect parse or the lack of a parse altogether.

The attractiveness of a semantic approach is also demonstrated through the use of word similarity such as in FAQFinder, (Burke et al., 1997) using WordNet and more recently (Mihalcea, 1999). Similarity as shown by such research can improve user query disambiguation.

The accuracy of word/sentence disambiguation depends to a large extent on the accuracy of the above mentioned approaches. While individually (the Brill Tagger for example) the accuracy may be high (96%), when used in combination with methods such as parsing and word similarity, the overall accuracy of disambiguation (using simple probability) is based on the combined accuracy of the individual factors (inaccuracy is multiplicative - assuming that the factors are mutually independent which may not be strictly true depending on the relationships between factors):

$A = M1 * M2*...Mn$

where A=Accuracy, M(1-n) = contributing factors.

If for example, we use the Brill Tagger with accuracy of 96%, a parser that is 85 % accurate and our WordNet similarity look-ups help in 90% of the cases, the overall accuracy would be:

$A=96\% * 85\% * 90\% = 73\%$

If the desired goal is an overall disambiguation accuracy of 90% and there are three contributing factors, the average accuracy of all of the factors must be at least 96.66%.

The need for high accuracy for tagging and parsing was stated by Mark Sanderson (Sanderson, 1994), and more recently by Eric Brill (Brill, 2001). He asserted that data focused approaches (large amounts of data are assumed) will generally produce a result while semantic approaches often fail to even produce a result.

Given the current trend toward semantic based methods, and realizing the importance of maintaining a very high accuracy in the individual components used in a semantic approach, it becomes obvious that the effort to improve already "well" performing components is necessary if the overall disambiguation result in a semantic approach is to be improved.

2.2 Previous Work In FAQFinder

The following sections will present the existing functionality of FAQFinder and past efforts to increase accuracy. Where appropriate, references to research upon which existing FAQFinder functionality is either based on, parallels or extends will be provided.

2.2.1 Vector Based Term Weight Computation

One of the methods used by FAQFinder computes weights for each term in a user question and is based on word document frequency (Korfhage, 1997). Each sentence's words (terms) are first stemmed and then tagged according to Part Of Speech (POS) using the Brill Tagger (Brill, 1992) . Each FAQ "question" is then considered to be a "document" and a user's question will be matched against the question in a FAQ file document (the question component of a FAQ file).

Question terms are converted to a term vector and a POS tagged "term set". The vector components' individual weights are computed using the formula:

$$W_i = (1 + \boldsymbol{log}(tf_i))\boldsymbol{log} N/df_i$$

Where,

i = the term index
tf_i=the number of times this term occurs in the question
df_i=the number of questions in which term i appears in the FAQ file (question components of the FAQ file)

The above is computed for each term in both the user question's vector and for each FAQ question's vector in a given FAQ file. Comparisons between the user question vector and each FAQ file's question vector is accomplished through the use of the cosine measure:

$$\cos\left(V_{users}, V_{faq}\right) = \sum W_{useri} \frac{W_{faqi}}{\left(\sqrt{\sum W_{useri}^2} x \sqrt{W_{faqi}^2}\right)}$$

Where,

W_i=weight of term i (where there is a matching term)
v_{user}=weight vector for the user question
v_{faq}=weight vector for the FAQ question

The vector approach is well suited to large documents but is also helpful in FAQFinder's smaller document dimensions (basically a FAQ question). Its limitations become apparent when a user question is semantically similar to a question in a FAQ file but where there are no common

terms. For this reason, FAQFinder uses other mechanisms that do address this limitation.

2.2.2 Word Sense Categorization

An integral part of question matching is the need to correctly determine the sense of words comprising a sentence. One example of this might be: "The bird likes to fly high in the sky". Considering each word in the sentence individually, the word fly could either be a verb or a noun. FAQFinder considers each sentence's words as a group to see whether there exists a relationship between them. Facilitating the search for relationships between words is functionality within WordNet that classifies words according to their hierarchy within a tree structure. Higher levels of the tree correspond to a general or super-class relationship with a given word. WordNet is implemented with a first level of 25 noun classifications, under which there are various subclasses.

FAQFinder makes use of this classification capability to ascertain the "sense" of each word in the sentence while considering its relationship to the other sentence words. In the example "The bird likes to fly high in the sky", the desired goal would be to have a classification of "verb" for fly, and in this case FAQFinder will correctly tag the word "fly" as a verb. This is possible because of FAQFinder's use of the Brill Tagger (Brill , 1992).

Given that two sentences' word components can be correctly tagged, the matching process is at the level where words such as "fly" allow FAQFinder to distinguish between two sentences that might use the same word but where the they differ in part of speech. For example, this would allow differentiation of the above-mentioned sentence with one such as "The fly has wings", where FAQFinder will tag the word fly in this case as a noun.

Words tagged according to part-of-speech are then further processed according to semantic similarity (distance). Tagged terms may have multiple WordNet senses . A comparison is made between two terms' vector representations of those senses, where the minimum distance between the senses in one vector vs. the senses in the other is taken as the measure of similarity between two words.

FAQFinder currently uses the Brill Tagger. There are recent successful attempts at using other taggers such as the SNoW tagger used by (Roth, 2001) and Collins Tagger (Collins, 1997). Roth's results indicate that SNoW is significantly faster (Roth claims 3000 times speed improvement over the Brill tagger) and may in fact provide better accuracy (96.86% versus 96.49%) when run on the WSJ and Brown Corpora. While accuracy would not be a reason to switch taggers, the performance claim (if accurate) might be.

2.2.3 Word Semantic Consideration

While word sense categorization is one example of FAQFinder's capability, another aspect of question matching is the ability to match on synonyms or near-synonyms. The English language is rich with many words that share a synonym relationship. FAQFinder's current implementation considers this fact during sentence matching.

The design of WordNet is based on the "meanings" of words. If one were to consider a set of words and their meanings as a two-dimensional lexical matrix (Miller, 1993), not only are synonyms more easily apparent but also the polysemy, or multiple meanings of a given word. Representation of this many-to-many relationship between words is accomplished in WordNet through the organization of words into "synonym sets", where such synonym sets can have semantic relationships between them.

In comparing two sentences, FAQFinder considers the closeness of meaning of words. In the sentence "The bird likes to fly high in the sky", the word fly for instance might have a synonym relationship with the word "soar". Although not a synonym, the word "travel" might also be a consideration when attempting a sentence match. Such relationships (semantic distances) vary in strength and this variance can be important in determining the best match between sentences.

Distance or closeness between words can be represented by organizing words into a list where it ascends in generality, such as:

sedan->car->motor_vehicle->vehicle->conveyance->artifact->object->entity

A large number of words can be grouped into a tree structure with higher levels implying more general meanings. WordNet nouns are grouped into 25 categories or "beginners". FAQFinder makes use of the ordering of word meanings by considering the "semantic distance" between two words. For example, the semantic distance between sedan and car is a closer relationship than between car and conveyance.

Sense tagging in FAQFinder has taken the additional step to consider the meaning of a given sentence prior to comparing two sentences in hopes of a match. Each sentence is first analyzed to determine which senses of a word to use in comparing two sentences. For example, the sentence "The black couple danced at the ball" has the potential to consider multiple meanings for the word "ball" and it would be incorrect to match on the sentence "The bowler rolled the black ball". FAQFinder will attempt to disambiguate each sentence and assign the correct sense to "ball" in both.

Semantic similarity is addressed in FAQFinder as a separate computation apart from a term vector measure. As was mentioned in "Vector Based Term Weight Computation" above, FAQFinder computes a "tagged term set" along with a term vector. The "tagged term set" is used to determine a semantic similarity between a user question and questions in a FAQ file.

Each word in a question (both user and FAQ file) can have synonyms (retrieved from WordNet) that are used in determining how closely a word in a user question is aligned with a word in a FAQ question. As was mentioned above, WordNet provides a distance measure between words

and for each term in the synset of the user and FAQ question, a distance measure is taken to see how closely a given term matches. The minimum distance between synset terms for a given term is then combined for each question to generate a similarity measure at a question level. This is accomplished by locating the closest matching synset term in a FAQ question to a corresponding synset term in the user's question. This is done for all terms in the user question, generating a similarity weighting between two compared sentences.

Because a word can have different senses (e.g., bat, ball, band, etc...), FAQFinder computes word senses based on the combination of terms within a given sentence, thus "The pitcher threw the ball" and "Cinderella danced at the ball" would have different senses for the word "ball". Determination of which sense of ball to be used in the above examples would be based on the sentence's co-words. This provides an additional level of disambiguation prior to distance computation.

Computation of word sense involves a tagging function that looks at all of the words within a sentence and its synonyms. Given that a word can have many different meanings, e.g., "bat", the tagging function looks at all of the senses that WordNet returns for all words within a sentence and determines whether there is a common meaning (sense) between words. For example, "bat" and "ball" when used in the same sentence have a sense related to baseball and not to a nocturnal flying mammal. Computation of this meaning involves use of WordNet returned senses, where the objective is to select a meaning for two words where there is a minimum distance between them. For a given sentence, this computation involves not just two words but the minimum distance in combination for all words in the sentence. The following formula is used to find the minimal set of sense combinations:

$$\Delta(S) = \sum_{s_i \in S} min_{s_{j \in S}, i \neq i} D(s_i, s_j)$$

Where S = Set of combinations of all synsets
si,sj=synsets
D=distance

3 Related Work

FAQFinder is essentially a document retrieval system, although, the approach taken differs from standard word frequency based document retrieval methods. Interest in AI techniques for Natural Language Parsing has understandably found a place in document retrieval application. As the amount of data to be cataloged and maintained has grown (especially as it is made available on the Internet), the need to provide accurate retrieval mechanisms has become an important area of business need and research. While the success of early search engines such as Yahoo and AltaVista has provided users with unprecedented access to data, along with this access come a few issues.

Oddly, the first issue is born out of the success of the solution, in that, as the amount of data has grown, search engines tend to return "too much" data in response to keyword based searches. To combat this problem, search engines provide Boolean expression capability in an attempt to give the user a way to distill the incoming data. While this can be helpful, results vary according to the skill of the user in their ability to compose search expressions.

A second issue is that keyword searches require exact matches between the keywords specified by a user and the target data. There is no semantic capability in a keyword match nor would the results in general be welcome due to the expanded result sets that might accompany such an approach. Yet, most users would no doubt appreciate a system that might correctly anticipate the relatedness of keyword terms and corresponding target data, and then return a valid result utilizing an educated guess.

Yet a third issue is that keywords, like any other words in the English language, are blessed (or cursed) with a degree of synonymy. Not only do nouns have multiple meanings , but also words may play the role of a different part-of-speech. For example, the words "fly", "run", "block", "trash", all have noun, verb and adjective contexts. This presents a problem in that returned matches will contain references for all three parts of speech; clearly not what a user might want.

The issues listed above are not an exhaustive look at the issues facing document retrieval. What can be said however is that there is a need for intelligence in document matching and an AI approach is one potential avenue to achieve this. Progress has been made in this area by many researchers, however, the road ahead does not afford a clue as to when a user will be able to converse with a system having a human level of understanding. Yet, progress has been made. SMART (Buckley, 1994), WordNet (Miller, 1993), Brill Tagger (Brill, 1994), Ask Jeeves (http://www.ASK.COM), and FAQFinder (Burke et al., 1997) have contributed to this progress. Such progress differs from keyword search and frequency based methods of document retrieval in that it seeks to support the analysis of embedded meanings of the words contained in a user's statement.

I will begin with a statement that was made by Mark Sanderson (Sanderson, 1994): "There is anecdotal evidence to suggest that in general, tools built for computational linguistics tasks need to operate with at least 90% accuracy before they are of practical use and in general, we can conclude that the performance of such systems is insensitive to ambiguity yet very sensitive to erroneous disambiguation". Sanderson made these statements after trying to improve upon results achieved through word/document frequency based analysis through the use of a technique proposed by (Yarowsky, 1992). Later research by others does show that semantic techniques can improve disambiguation and retrieval results, however, achieving positive results is a generally difficult process. Sanderson's statement about the negative consequences of erroneous disambiguation are somewhat relevant to FAQFinder in that current accuracy is impacted not only by how changes positively impact recall but also by how such changes generate false matches.

(Resnik, 1999) researched noun associations as a way to disambiguate natural language. His approach, which was based on earlier work by Kurohashi & Nagao, (Kurohashi et al., 1992), was to look at noun combinations and determine whether there are relationships that might yield semantic information. For example, "peanut" and "butter", collocated in a sentence, have a specific meaning that is unique from their individual meanings as do "securities fraud", "corn oil", "crude oil", etc. Resnik attempted various approaches that include number agreement, statistical grouping and similarity. He devotes special focus to number agreement as he states that "similarity of form is to a great extent captured by agreement in number". The example he uses is:

a. several *business* and *university* groups
b. several *businesses* and *university* groups

where in a, "business and university" may be conjoined and because of non-agreement in number in sentence b, "businesses and university" cannot be. This method is quite accurate at 90% however, coverage suffers at 53% due to situations where agreement could not be definitively decided. In an attempt to remedy this, Resnik elected to implement defaults, which increase coverage to 100% at the expense of a decrease in accuracy 82%.

An interesting aspect of Resnik's research was that he also attempted a semantic approach to disambiguating nouns. His method was to use WordNet, comparing combinations of words in an attempt to determine noun senses. He used an edge counting approach to ascertain similarity:

$$wsim_{edge}(w_1, w_2) = (2\,xMAX) - \left[min_{c1,c2}\, len(c1,c2)\right]$$

where w1,w2 are two words being compared, wsim is the edge similarity and c1, c2 are path lengths to a common WordNet entry that subsumes w1 and w2.

Resnik found that his semantic approach was not as effective as the number agreement method; accuracy 71.2%, coverage 66%. When using default, the results were 100% coverage and 72% accuracy.

What is surprising (or perhaps not so surprising) here is that the semantic similarity correlates very well with the values FAQFinder has been able to achieve ~70% recall with 0% rejection. This is interesting because FAQFinder does use WordNet similarity (sense tagging) as one of the approaches to question matching.

One issue with Resnik's work in using the semantic similarity is that for word groupings such as "peanut butter", use of WordNet on the individual words "peanut" and "butter" will yield a poor result. This is because individually, the two words' senses *are not* closely related. What is missing here is the need to represent multi-words and phrases as entities. This is not as much of a problem for concepts such as "wool clothing" as the individual words may be subsumed at some level of the WordNet hierarchy. As with "peanut butter", it is clearly a mistake to assume that concepts expressed by collocation of two nouns can be inferred from matching their senses

in the WordNet hierarchy; this does occur but as Resnik has indicated, the accuracy is as given above.

An important element here is that the number agreement approach used was actually quite successful at 90%. This does indicate an avenue that might be tried in FAQFinder to improve its accuracy along with consideration of phrasal tagging.

Along with noun-noun collocation, a similar case can be made for verb/noun collocation. Considering the verb/noun combinations: "turn corner", "raise hell", "eat crow", "change mind", "break bread", "fix flat", etc. All of these collocated verb/noun combinations pose a problem for taxonomies based on single words because the phrasal combinations' meanings cannot be determined. (Montemagni et al, 1998) alluded to the need to consider such relationships and suggested the need for a solution taking the form of a taxonomy patterned after WordNet. Their approach (SENSE) was to use distribution information (distributions trained from a corpus of verb/noun pairs) in combination with a taxonomy based on WordNet. Their findings indicate that the best accuracy is achieved by first applying a distribution approach and using taxonomy only in those situations where ambiguities remained.

Their conclusion "Semantic similarity is not simply a relation between two words in isolation, but rather a relation between two words and their context" is in-line with results seen by Resnik, that phrasal disambiguation can be improved through mechanisms that extend the functionality provided by a taxonomy such as WordNet.

The desire to consider additional words and context was formalized as a concept of "Semantic Density" which was proposed by (Mihalcea et al., 1998). Semantic density as defined by Mihalcea analyzed two words to determine the number of other words that the two considered words have in common. The thought is that two words that are similar in meaning will often be associated with other words in common. This approach is attractive from the standpoint that it can be applied to words with a different part of speech, e.g. verb/noun.

Mihalcea uses two methods for obtaining the related words that two compared words might have in common. The first is to make use of WordNet's usage examples for words (the gloss). Thus, the words in the gloss for one word can be used in comparisons with the gloss returned for another word. The similarity of the glosses is then used to determine how similar the two compared words are. A second method used by Mihalcea is to make use of the vast amount of data residing on the Internet; an approach also taken by (Buchholz, 2001). AltaVista is used to search for target words which then returns sentences from which an analog to WordNet's gloss is extracted and scored.

(Burke et al., 1997) stated that:

"When people talk about a topic, they tend to use the same sets of words over and over again. For instance, when one talks about "baseball", there is a high probability that one will use the words like “ball” and “bat”. Thus if one knows the words that tend to co-occur with the words of an input question one can use them in matching operations. The text of the audio_FAQ file test

data set shows that the words "speaker" and "listen" have a high degree of co-occurrence, which can be used in finding the correct match. We are looking for effective heuristics to find co-occurrences and use them in matching operations. One such heuristic would be to augment WordNet's semantic relations with a weaker domain-specific relation like co-occur. For example, if it is known that a particular FAQ is about baseball and words " ball" and "bat" show a high degree of co-occurrence, one can add the relation co-occur ("ball", "bat", "baseball") to WordNet's database of semantic relations"

Mihalcea's approach using semantic density appears to be the kind of functionality sought by the Burke et al. Mihalcea's approach is especially attractive because of its ability to address noun/verb word pairs and not simply noun/noun.

Mihalcea & Moldovan put this concept to use in 1999 (AAAI). They proposed to use semantic density to produce a large tagged corpus for NLP use. Dagan (Dagan et al., 1997) also took the same approach - that similarity of training set data to similar words in text could be used to disambiguate the text using statistics computed on a training corpus. Mihalcea's assertion is that that SemCor/WordNet is perhaps not large enough for statistical approaches at disambiguation and that one way to create such a corpus is to use the large amount of information on the Internet. Their approach was to collect data from the WEB and then use semantic density approaches (WordNet glosses in combination with semantic distance) to extract and categorize text. Categorized text would not be 100% accurately categorized, thus their approach does have a manual operation to verify the system's findings.

The semantic density approach used by Mihalcea/Moldovan is interesting also because one of the deficiencies mentioned in (Burke et al., 1997) for FAQFinder was that rejection might be improved through "condensation clustering". Condensation Clustering as defined by Burke, et al is in some ways similar to semantic density. The basic idea behind this technique is to look at the distribution of a term over a set of receptacles, i.e., sentences, paragraphs, Q&A pairs, chapters, etc. and see if the distribution has a statistical significant deviation from the random distribution. If that is the case, the term is likely to be content-bearing". One difference is that Mihalcea/Moldovan go a step further and make use of the information available in WordNet glosses as a way to qualify the encountered text.

Mihalcea continued this effort with Semantic density and WordNet glosses in (Moldovan, 1999) by applying the concept to Internet search. FAQFinder does attempt to determine the senses of words that comprise a question and Mihalcea takes an interesting approach to the same goal. Mihalcea first takes a word pair (verb and noun) and then obtains similarity lists for each word from WordNet. Using all possible combinations of words in the two similarity lists, these pairs are used in an Internet search and ranked according to frequency; the top two rated senses for the noun and verb are kept. Next, glosses are extracted for the two verb senses, and the nouns from those glosses are extracted to form noun contexts for each verb by assigning a weight according to WordNet Hierarchy. Finally the conceptual dependency is computed for a verb/noun pair by comparing the noun hierarchy of the verb's noun context and with the hierarchy of the original noun's WordNet sub-hierarchy. The goal being to determine which are the most likely senses for a verb/noun pair.

Mihalcea's effort is interesting and important because it attempts to establish senses using WordNet's hierarchies along with written text from the web (actual usage). It is an alternative to efforts that wish to establish sense using word frequency (Stetina, 1998). The attractiveness of Mihalcea's approach of over Stetina is that there is not a need for a large training corpus as the sense computation can occur in real time. This is probably also a weak point given the current technology's inability to handle the large computational tasks required. Mihalcea does not discuss performance although the fact that verb senses are limited to the top two is an indication of this being an issue and Mihalcea does state "Only the first t possible senses of this ranking will be considered. The rest are dropped to reduce computational complexity".

Mihalcea's results were quite encouraging (based on 50 TREC-6 topics and 50 actual user questions) using natural language queries in comparison to logical (AND/NEAR). The important aspect of the results were that not only did the system consistently return answers in the top ten results but it also consistently returned significantly fewer unrelated documents (precision) than a traditional search engine (AltaVista). Not measured, although also significant is the fact that Mihalcea's system returned not only the relevant document but also the paragraph that contained the answer to the user's question. While it is not appropriate to compare FAQFinder to an Internet search engine, Mihalcea's research using WordNet in this way is very encouraging.

(Stetina et al., 1998) also attempted disambiguation using WordNet senses with an interesting approach. They operated under the premise that words in a sentence have both a local sense component and one external to the sentence. Locally, a word's sense is usually constrained by specific relations within the sentence. An example might be in the case of a verb, where its sense is constrained via a subject/verb or verb/object relationship. They propose an approach to identify such relationships and then to use these relationships both in training an algorithm that assigns weights to words participating in a training corpus and then later to disambiguate text. External sense then, is the consideration of similar constraints, where the words in a given constraint are compared against constrained words that were encountered in a training corpus.

The training was done on the Brown Corpus where one of the 103 files was used as the test data and the remaining 102 files were used to train the system. As words are encountered in the training files, a set of matrices is created to hold weights accumulated for words participating in the identified relationships. Weights were computed using WordNet edge distance and frequency of occurrence in the training files. For example, the word combination "new airport" might have a measure of similarity to "modern building" because of the WordNet relationships between new/modern and airport/building plus the relationship that both phrases subscribe to ADJ/NOUN. In this way, their approach works even though there might not be an exact match between words in the training examples and in the text data being analyzed.

One difficulty with this approach is the need to properly tag words in a sentence according to part-of-speech. This is not always possible. While it is true that an unidentified word often may turn out to be a noun (such as in the case of an acronym) this still is a potential issue. Furthermore, not only is polysemy a concern, but also situations where words have both noun/verb contexts as an example. This poses a problem in parsing and part-of-speech tagging.

Still, their approach is interesting and the results they achieved 80% are encouraging. Use of this approach in FAQFinder might be applicable, especially if FAQ files' answer components were used as the training sets. The difficulty might be with properly parsing training data and subsequent loss of accuracy.

FAQFinder in some cases might actually do better here because of its consideration of "all" of the words within a sentence given that other words might help in "sentence" disambiguation. Still, there appears to be a need for phrasal concepts to be uniquely identified and Stetina's approach may be of value within FAQFinder. This is especially true for those cases where phrases and multi-words have a specific meaning (Resnik, 1999), phrase focus (Ittycheriah, 2000), phrase focus in LASSO (Moldovan, 2000).

Research (1999-2001) has focused on analysis of user questions that parallels the effort in FAQFinder, (Roth, 2001), (Buchholz, 2001), (Chen et al., 2001), (Alpha et al., 2001), (Brill et al., 2001). These submissions to TREC approach the problem of sentence disambiguation by attempting to ascertain the question type (who, what, when,...). Once the question type is known, an attempt is made to determine the target of the question (such as person, place, date, etc...) in an attempt to find an "answer" in a body of text. In the case of (Roth, 2001), and (Buchholz, 2001), questions are given a shallow parse to identify the key elements of a sentence while (Chen, 2001) used the Collins Parser (Collins 1996).

In (Chen et al., 2001), a sentence is scanned for clues about its type (who, what, when, ...) using predefined regular expressions. When a sentence's type is identified, an attempt is made to determine the focus or target of the question type. For example, in the sentence "What is the deepest lake in the US?", the sentence would be classified as a "what question" where the required answer will be for "deepest lake". An answer might then be searched for that makes a statement about the "deepest lake" in the "US".

An interesting point here is that there exists functionality within FAQFinder where the subject/verb/object of questions is identified. Given the sentence above, FAQFinder would pull out NIL/is/lake as the subject/verb/object triple. Intuitively, attempting a match with this triple against other questions in FAQ files might be difficult for the following reasons:

1. We really need the adjective "deepest" but in the context of a single term "deepest lake"
2. We also need the prepositional phrase "in the US" to qualify the question
3. We don't really have a subject

Without actually running this question through FAQFinder, a "guess" would be that the FAQFinder component that considers the subject/verb/object would not fare well (although the vector functions would probability do well). An additional point to be made and stressed here is the potential need for grouping the adjective and noun "deepest"/" lake" into a single term.

FAQFinder's approach to identification of the subject/verb/object is similar in concept to methods used by Chen/Roth/Buchholz, although, there are differences. While the recent efforts by Chen/Roth/Buchholz attempt a deeper level of semantic analysis by identifying the question

type and answer target, they do not attempt to support similarity such as the semantic distance used in FAQFinder or semantic density approaches (Mihalcea, 1998).

This presents an interesting question about whether FAQFinder would perform better if its functionality were extended to support phrases (this includes named entities and multi-words) and question type identification (who, what, when, ...). Questions would then be either matched against FAQ questions that are of similar type or perhaps with FAQ answers that might provide answers to a user's question.

It is not always evident what the subject and object are. While it is clear that these must be "nouns", "which" noun is not always easily resolved. If for example the head noun is a pronoun, identification of the subject or object using word is a fruitless effort. For example, the question about Napoleon: "When Napoleon attacked Russia, which battle did he lose the most soldiers?". This shows the difficulty of identifying the subject, verb and object. We would like them to be (Napoleon, lose, soldiers) but more than likely, the result will be (he, lose, soldiers). A potential solution to the above was described by Claire Cardie (Cardie, 1999) through a method called "co-referencing" where noun phrases contained within a sentence would be analyzed for semantic agreement.

Cardie's observation is that noun phrases that are within a short distance within text are often semantically related. In the above sentence about Napoleon for example, the pronoun "he" can be resolved by referencing "Napoleon". Cardie attempted a simple heuristic that identified all noun phrases in a sentence and then selected the "head" noun in each noun phrase (the last noun in a noun phrase).

Once head nouns were identified, WordNet was used to determine whether there was a relationship between head nouns. In the above case, the pronoun "he" and "Napoleon" both resolve to "human" and thus "he" "refers" to "Napoleon". The word "he" would not be identified as co-referring with either Russia or battle because they are not semantically related and "he" would not pertain to "soldiers" because "he" is singular and "soldiers" is plural. (Azzam, 1998) also focused on pronoun resolution, basing his work on the assumption that "... utterances in discourse are usually about something" and then attempting to resolve pronouns via the discourse focus.

Cardie's approach is something to consider for FAQFinder for one other situation. This is to disambiguate multiple sentence questions where the second question might use a pronoun that refers to a noun in a preceding question. (Harabagiu, 2001) also addressed this point (discussed later in this paper) where an attempt is made to connect concepts that span multiple sentences.

The concept of "focus", was a consideration not only for (Azzam, 1998) in pronoun resolution; (Ittycheriah, 2000) and (Moldovan, 1999) also made use of this. Ittycheriah uses the concept of focus as a mechanism to weigh and differentiate various options for word disambiguation. If for example, a sentence discusses a flowing river and a focus can be established that deals with "river", then preference can be given to text that may contain terms with hyponym/hypernym relationships to "river"; "Mississippi" for example.

Focus in one sense is an analog to what FAQFinder already does when it attempts to disambiguate terms within a sentence, e.g. "The fly is a small bug" and "The fly ball was caught by the pitcher". While FAQFinder computes a minimum semantic distance between terms in a sentence to establish the senses for terms, a "focus" approach attempts to make a similar determination for a sentence; in the above examples this might be "insect" and "baseball" respectively. Based on such research examples, it seems to be the case that determining the "focus" or "sentence topic" is relevant to both disambiguation of term meanings and in attempts at sentence matching.

This should not be a surprise. Human discourse involves determining the meaning or focus of sentences. At times a human may need to read more than one sentence (paragraph for example) to determine the meaning of a given sentence; the end goal being
to determine "what a given sentence is referring to" or in short, "focus". It thus is encouraging to know that FAQFinder and other research has deemed this to be an important factor, and it brings to surface the question, whether the greater use of sentence focus within FAQFinder might not yield positive results.

Recent efforts in FAQFinder (Lytinen, et al. 2002) have focused on the application of question type matching. Lytinen's approach was to classify a user question into one of twelve categories (definition, reference, time, location, entity, reason, procedure, manner, degree, atrans, interval, yes-no). A distance-weighted K-Nearest-Neighbor (KNN) algorithm (Dudani, 1976) was used as the classification mechanism. Use of KNN required the manual selection of a set of words (90) from example questions. These words were considered to be the most likely indicators of question type.
Results achieved by Lytinen were an error rate (after training) of 23%. The same training set yielded an error rate of 30% when used to classify questions from other FAQ files that were not included in creation of the training set. Question type classification functionality was then integrated with existing question matching logic in FAQFinder to determine how this might impact FAQFinder's accuracy (FAQFinder uses a vector, coverage and semantic metric combination). By including question type classification as the fourth metric, there was a significant improvement in FAQFinder accuracy. This is encouraging and is an indication that question type classification holds promise as a method to pursue.

Although not directly applicable to FAQFinder, it is interesting to note that (Chen, 2001) - mentioned above - (in his attempts to answer a question by scanning text that might provide a direct answer) searched answer sets via paragraph frequency tagging (using a sliding paragraph window) instead of whole document frequency tagging and found that it did improve the results (although an explanation for this was not provided). The attempt was to find key paragraphs that defined a document and then to find sentences within the paragraph that might be possible answers to a given question. As in (Cardie, 1999), the thought is that human speech (and written text) tends to address a specific topic within a "short" span of text. In the case of (Cardie, 1999), this is usually one sentence and for (Chen, 2001) this is a sliding paragraph. Chen's approach is also interesting because WordNet is used to compute a semantic distance between terms in the question and answer text, thus being similar to FAQFinder's use of WordNet.

Another example of an approach that used question typing is (Lee, 2001) and SiteQ. Question processing involves the determination of question type and conclusion type. The conclusion type refers to the type of answer that is expected based on the question type. Text is tagged using the POSTAGE/E English Tagger. Useless words are stemmed using a stop list which may remove words based on where a given word resides in a sentence; e.g., "kind", "sort", "one" are removed if they occur on the left side of a preposition.

One unique aspect of SiteQ is its use of the Lexico-Semantic Patterns and grammar such that a question such as "Who was President Cleveland's wife?" would be parsed via the rule: (%who) (%be)(@person) ->PERSON. In this case, %who matches the word "Who", %be matches the word "was" and @person matches "President Cleveland's wife". In this case, "wife" is determined to be the head of the noun phrase and thus categorizes the answer type to be PERSON.
Based on the amount of information provided in the documented results, it is not immediately clear how well LSP might function, but the attempt at formalization is at least is a positive step given that some formalism would be desirable.

Alpha - TREC-2001 (Oracle Corporation), is yet another instance of where current research appears to be headed in that it too attempts disambiguation through classification of a sentence into question type (who, what, where, etc...). Unlike other current research, Alpha makes use of an in-house developed knowledge base as opposed to relying on WordNet, although there is mention that WordNet was used for question disambiguation. The Oracle based knowledge base supports 425, 000 concepts grouped into 2000 major categories organized in a hierarchy with 6 top terms (business and economics, science and technology, geography, government and military, social environment and abstract ideas and concepts). The knowledge base was used to filter and pre-classify the document corpus into a set of document theme vectors. This approach is interesting because of its attempt to use a knowledge base other than WordNet, and it's approach of using the Oracle Database. Interestingly, it performed better than average on TREC-10.

(Brill et al., 2001) further supports the concept of sentence classification. The system (AskMSR) is a Question/Answer system funded by Microsoft Corporation). Brill classifies questions into seven categories (who, what, where, ...). While the question classification is an attempt at semantic analysis, AskMSR was implemented to minimize reliance on semantic techniques. Instead, Brill has elected to target results by relying on data: "In contrast to many question answering systems that begin with rich linguistic resources (e.g., parsers, dictionaries, WordNet), we begin with data and use that to drive the design of our system".

Brill's contention is that while there will be many instances where a direct data approach may fail, such as where a question could only be matched with semantic approaches, he feels that in most cases there are usually multiple answer matches using a data approach. Furthermore, he contends that the probability of getting a match somewhere in the document is good and that usually there will be multiple matches. Brill views that a semantic approach might fail altogether to find a match and return nothing because the syntactic or semantic rule weaknesses; a view not shared by (Moldovan, 2000) based on his research into the "LASSO" question answering system

which was not dependent on a single semantic method.

While Brill's approach with AskMSR is not currently applicable to FAQFinder because FAQFinder performs (question<->question) matching, this poses an interesting thought: "Could a simplistic approach based on question typing and subsequent question->answer_text matching be used effectively in FAQFinder?". Brill's approach relies on large amounts of data (answer text) but FAQ files usually contain shorter answer sections for a given user question. Still, if a user question (either in the FAQ file or at run-time) is weak, for example:

runtime question: "Who is the president of China?"
Relevant FAQ question: "Information about Asia"
Relevant FAQ Answer:

.....
..... The president of China is Zhu.
.....

In the above, we may not have a good match between user and FAQ question but the answer section would provide a good match. Still, given the relatively small amount of text in FAQ answers, it might turn out that a distance based semantic approach might be needed for FAQ files, e.g., the answer text might contain "The premier of China is Zhu" and unless we correlated president/premier, Brill's approach would fail.

Not all attempts at semantic disambiguation bear fruit as evidenced by (Ferret, 2001). The approach by Ferret began as an attempt at question "type" classification. Next, an anticipated answer type was calculated along with a "focus" (or object) which they assumed was usually either the noun or noun phrase following the main verb. An attempt was made to identify the noun head within the noun phrase to retrieve a focus "head" and any modifiers of the focus head that might be present.

Results were a disappointment to the authors, however, in analyzing the results the authors were able to identify reasons for the weak showing. Some of the lack of accuracy was traced to the type of questions their system (QALC) expected (and didn't expect). In the TREC research previously mentioned, many of the described systems were run on older TREC data and there seems to be consensus among the researchers that TREC question difficulty and ambiguity have increased. It is therefore plausible to assume that weaknesses in a system's components (such as a parser or question classifier) might result in a disappointing result (as was pointed out in the justification section of the thesis hypothesis). There is however something of value in this researcher's accomplishments in that the research supports the need to maintain high accuracy in components that comprise a semantic approach to disambiguation.

A more ambitious approach to the use of semantics was by (Harabagiu, 2001) where an attempt was made to preserve context between multiple questions as one might encounter in a conversation. Harabagiu also points out the need to consider the following in the context of a

conversation: demonstrative pronouns, third person pronouns, possessive pronouns, definite nominals , nominalizations of verbs, elliptical references, causal-effect reference and meronymic reference.

An example of a Definite Nominal might be "*What company* did the work?" where the company might have been introduced in a previous question.

An example of a nominalization might be where one question is "When was *construction* begun?" , followed by a second question "Where was it *constructed*?".

An elliptical reference according to Harabagiu, is one in which the expected answer type is inherited from the previous question; e.g., "What ***kind of sports car*** is the best handling?" followed by "What kind is the safest".

The FAQFinder answerable question set does include questions of the type addressed by Harabagiu. This poses a problem for FAQFinder as this makes it difficult to determine the subject/verb/object of a user question and of multiple sentence questions in FAQ files. Currently, FAQFinder does not support multiple sentence questions and any questions of this type fail to produce an answer.

Harabagiu's approach is important because in multiple sentence situations, the second question, while related in topic, can be searching for a different answer. This is of limited value to the current implementation of FAQFinder because FAQFinder makes no attempt to determine question type (who, what, where, etc..). If FAQFinder were modified to perform a deeper level of semantic analysis of a sentence where question classification were deemed important, then Harabagiu's approach would be of potential value.

Another interesting aspect of Harabagiu's research is the attempt to make use of "causal relationships". Not only does Harabagiu attempt to make use of causality, but also causal relationships are identified between references in multiple sentences; this theme was also mentioned by (Harabagiu, 1995). For example, a sentence such as "Which explosive was used?" might be related to "Which museum was damaged by a bomb explosion?" because of the causal relationship between explosive and "bomb explosion". Harabagiu's implementation focuses on establishment of a relationship between two successive questions instead of in the attempted match of two questions as is the case with FAQFinder, the concept is still intriguing because of its potential value in situations where WordNet similarity may not suffice.

Conclusions by Harabagiu on lessons learned are interesting; two statements in particular. The first is that "Open domain resources such as WordNet can be fully exploited to process more and more complex definition questions or for processing questions in context". The author is expressing satisfaction with an approach that attempts semantic disambiguation. The second lesson "such resources are not exhaustive, thus Q/A systems need to robustly process questions even when lexico-semantic information is not available". Here the author is recognizing that while WordNet provides very valuable functionality, a still more powerful knowledge representation would be desirable to aid in the processing of unrestricted domain text.

Information retrieval generally focuses on returning full documents that might pertain to a user query. As has already been mentioned, FAQFinder does not fit this paradigm in that the text that is to be matched is much smaller (a single sentence) and thus the need to consider semantic analysis techniques. (Moldovan et al., 2000) (LASSO System) stated that "Finding the answer to a question by returning a small fragment of a text, where the answer lies, is profoundly different from the task of information retrieval or information extraction". He was referring to the task of seeking an "answer" that might match a question posed by a user, however, FAQFinder has to deal with the same fundamental issue - that of the need to match very small amounts of text.

Moldovan's approach with LASSO is interesting from a number of aspects; first in its attempts to use semantics to parse a question to determine the user's intent. Secondly, it attempts to locate text that might potentially yield an answer. While this approach is not unique as other researchers in TREC-10 have also used this approach, there are two aspects of Moldovan's work that merit additional mention. The first is his use of a modified version of the NIST Zprise system that attempts to analyze and rank text (he uses paragraphs) according to whether the text contains a given set of keywords. This is done as a preliminary step to limit the amount of data the system has to process.

Secondly, LASSO determines the question type (who, what, where, ...) and this in turn drives the expected answer type ("who" for instance requires a person as the expected answer type). Lastly, LASSO attempts to determine the question focus, which will then be used to generate the list of keywords and to further quantify the type of answer being sought.

Moldovan did recognize that there were potential limitations with focus determination. His example "In 1990, what day of the week did Christmas fall on?" where the focus would be "day of the week", mentioned that such a focus would yield a set of terms not likely to be found in the target text. To resolve this issue, Moldovan uses a set of eight heuristics to increase the quality of the focus keywords.

Another comparison with FAQFinder can be made in the keywords that LASSO's heuristic functions target. For example, while it appears to be sophisticated in its identification and inclusion of proper nouns, question focus and quoted expressions, it then includes all nouns, verbs and adjectives. Interestingly, LASSO does not make use of synonyms or word semantic distance measures, instead relying on exact word matches. Also, unlike FAQFinder, it does not attempt the disambiguation of word senses, thus, FAQFinder appears stronger in this area.

It does use an interesting approach to parsing through the use of a modified version of the Brill Tagger where additional tagging rules were implemented and semantic dictionaries from WordNet and Gazetteers were unified. Additional heuristics were also added to recognize names of persons, organizations, locations, dates, currencies and products. This approach, modifying the tagger, is an attractive and potential option for modifications to FAQFinder to support the same word tagging enhancements.

Question type classification by LASSO and others brings to the surface the question of whether searching for word patterns (such as who, what, where, ...) could be extended to aid in searching for answers to user questions. (Soubbotin, 2001) elected to take this approach. He chose to "look" for patterns in text that might yield a clue in the answer text. While his approach is not unique (Moldovan, Harabagiu, Chen and others had to do the same to recognize potential answers), Soubbotin took this a step farther. Soubbotin decided to forgo NLP and knowledge-based analysis of the question text altogether and focus on text to recognize patterns residing within it.

In an example given by Soubbotin "What year was Mozart born in?", his approach is to predefine a text pattern such as "What year" and equate it to the category "When". Next, the pattern search continues and finds the word "born" which matches a predefined pattern that then recognizes the need to find a date. Given this information, the system will then search a document corpus matching any number of patterns that have been defined for date-of-birth specification, e.g.,

x was born in yyyy
x was born in jj/kk/llll
Mozart (1756-1791)
etc...

While Soubbotin's approach on the surface may appear simplistic, his results are respectable and satisfy some of Brill's concerns (Brill, 2001) about semantic methods being weak because of their propensity to fail to find an answer. Soubbotin's approach was extended to include looking for things such as "Mr.", "Jr.", punctuation marks, parens, $, date fields, various phrases and used a formalized grammar to express the patterns used.

Of 492 questions, Soubbotin's system supplied an answer for 372 (75%) of the questions and supplied a correct response for 289 questions. Soubbotin states that the system expressed "confidence" in these 372 questions being correct answers, thus 289/372 is 77% and a measure of the system's accuracy.

The semantic components of FAQFinder are dependent on the accuracy of word tagging (part-of-speech and sense). As a result, any modifications that might improve this accuracy have the potential to improve the semantic capabilities within FAQFinder.

The key components of the tagging are currently the Bill Tagger and a parser that provides sentence structure clues. An interesting parser approach by (Collins, 1997), is to improve the efficiency and accuracy of Context Free Grammar Parser (CFGP) through the use of statistical analysis of a training corpus. The approach used by Collins, appears to hold promise in those situations where a Context Free Grammar (CFG) yields many potential parses. This is perhaps not a problem for short simple sentences, as sentences increase in length and complexity, the potential for multiple parse interpretations can increase significantly. Collins' approach is to use prior knowledge of sentence structure and the probabilities associated with various structures.

Given a situation where a parser might need to choose between multiple parses, Collins proposes

that a choice based on probability (where the probabilities are based on term co-locations within a training corpus) is more accurate than a pure guess, or the use of the first parse found.

In the Collins parser, POS tagging of sentence terms is accomplished not as a separate step but along with the steps that determine sentence structure (the search for a parse tree that has the highest frequency of occurrence in the training corpus). The parser then will only return a POS tag if it has encountered the given word in the training corpus in a specific parse tree. This then points out a potential limitation (the need for a large training corpus) but also a positive aspect of the parser in that tagging is "automatic" and guaranteed to have been used at some point in the current context.

One downside with the use the Brill Tagger (Brill, 1994) in FAQFinder is that POS tagging occurs "first" before any analysis of appropriateness with it's collocation with other sentence terms. Once the tagger returns a POS tag for a given sentence, FAQFinder will then use this information in a attempt to find the subject/verb/object and takes no further steps to ensure word sense correctness.

A parser in combination with the tagger may fare better in ensuring correct word sense assignment for two reasons. The first is because the tagger is not 100% accurate, it might be that the sentence structure may yield clues about part-of-speech assignment. Secondly, sentence structure may provide clues about which words to select for the subject/verb/object.

It should be pointed out that the Brill Tagger is very accurate (96% as reported by Brill) and so it is not necessarily the case that a probabilistic parser is better. The attractiveness of a tagger and parser approach is the use of some other method to verify the Brill Tagger results. That a parser may provide additional clues about subject/verb/object selection is an added bonus that may improve FAQFinder accuracy.

Machine learning is an active area of artificial intelligence research for natural language processing. Machine learning techniques may be used in place of or as an enhancement to heuristic and frequency based text matching. For example, given a set of training examples that adequately define the answer to a set of input questions, it may be possible to learn classification rules which could then be used to classify additional sentences that are variations of the training examples. This is possible because a learning system attempts to identify a set of feature sets that are found to be important discriminators and the presence of such feature sets is not necessarily dependent on sentence structure.

Machine learning has performed very well in named entity recognition. Litman (1994) recorded good results using C4.5 to classify cue phrases in both text and speech. (Siegel, 1997) tried various learning methods including decision trees, genetic programming and log-linear regression to classify verbs into state and event categories, and found that all produced good results; especially decision trees with an accuracy of 93%.

(Marquez, 1999) experienced very good results increasing the accuracy of a POS tagger using a decision tree based machine learning approach. Marquez used a combination of functions (e.g.,

Gain Ratio, Chi-square statistic) to identify a set of features. These features then became decision tree features used to classify words according to part-of-speech.

Bikel et al, (1999) applied a machine learning approach for named entities based on Hidden Markov Models (HMMs). Bikel's approach (IdentiFinder) is to use a small feature table (hints) about whether a given word might be a named entity. A training set is used to calculate a set of probabilities of a given event (word combination) occurring. The calculated frequencies are then used to process the MUC-6 WSJ (Wall Street Journal) documents (MUC-6, 1995).

Bikel's results are very encouraging, achieving an F measure (where F=Precision*Recall/ (1/2(Precision + Recall)) of 94.92 (Recall=96%, Precision=93%). The attractiveness of Bikel's approach is that there is no need to create and maintain potentially complex rules. Furthermore, IdentiFinder was used to identify named entities for Spanish language text with excellent results (very little modification to the hint table). A machine learning approach to various aspects of FAQ question answering may merit consideration, especially given that named entities are an area of research targeted by this proposal.

An avenue of current research that also may deserve consideration is boosting which seeks to make use of simple and less accurate hypotheses (weak classifiers). These are combined to yield a single more accurate classifier. (Escudero 2000) is an example of the application of this method where an attempt is made to disambiguate word sense. Escudero compares the Schapire and Singer AdaBoost.MH algorithm with other approaches (Naive Bayes, K-Nearest Neighbor, Most Frequent Sense) and finds that of the three approaches, boosting performs the best. Furthermore, as the number of weak classifiers/rules is increased boosting continues to improve. This is encouraging as a possible approach to consider for various aspects of FAQFinder functionality, but perhaps not so much for word sense disambiguation due to the lack of a large enough training corpus.

Yet another interesting approach is one taken by (Hovy, 2001) for semantic based answer match. Hovy's approach to question answering is based on the idea that question forms fall into a set of general patterns for which potential answer forms can be predicted. A set of potential answer patterns is thus predicted and searched for within a corpus of knowledge. This approach was used in the TREC-9 competition where it tied for second place.

4 Phase I Research – Proof of Concept

The research perspective that this thesis is based upon is a quantitative approach, as much of the research is focused on measurement of various techniques' impact on question answering accuracy (Glatthorn, 1998). The general research approach is *Causal-Comparative* as the impact of selected techniques were analyzed to determine the most promising research direction.

The following table lists the main tasks that were undertaken as a part of this research effort:

Task Number	Task Description
1	Investigation of question/question matching improvements
2	Investigation of Question/Answer matching approaches
3	WEB harvest of Additional FAQ Files

The goal of the above tasks was to determine how various methods impact the performance of question answering systems (FAQFinder in this case). As a first step, an analysis was undertaken to fully understand the reasons for FAQFinder's current level of performance:

1. Where does each method do well and why?
2. Is a given method's performance lacking and why?
3. How did the various modifications impact accuracy?

FAQFinder improvements focused on enhancements to existing approaches and on new approaches that target the answer component of a FAQ. The goal was to enhance the effectiveness of existing and new approaches through mechanisms shown to have an influence (by current research in this area) and as a result of research described within this proposal. The following table lists the modifications that were made to FAQFinder for the Phase I prototype::

Item	Description
Integration of additional parsing rules into FAQFinder	The current CFG parser was enhanced to include additional grammar rules. There are now 450 rules compared to the 250 rules at the start of my research.
Ontology Layer*	A network-based protocol has been developed to provide a generalized access to Wordnet and a dictionary. This access facilitates Part-Of-Speech operations (POS). An active agent listens for requests on a TCPIP port and responds with POS information for a given word.
Question Classifier*	Function providing the ability to parse a user question and identify key parts-of-speech (e.g.,

	subject, verb, object, nouns, adejectives, etc…)
Question Matcher*	Function that accepts the output of the Question Classifier and attempts a match with an expected answer form. The question and answer forms reside in a question/answer rules respository.
Question/Answer Rules Repository*	The respository was populated with rules for the HOW question type.
Phrase & Named Entity Support*	A Phrase and Named Entity capability was created and populated with approximately 54,000 forms.
Integration of Question/Answer with FAQFinder	The Question/Answer Component (QUANDA) was integrated with FAQFinder. The integration approach was to trust QUANDA if QUANDA produces a result. If QUANDA does not produce a result, standard FAQFinder prodessing is used.
FAQ File processing utilities	A set of utilities were developed to assist in FAQ file manipulation. This includes a FAQ file verifier and question/answer display utility.
QUANDA rule utilities*	A set of rules checks were implemented as a set of utilities. This allows automated checking of QUANDA rules for a variety of syntax issues.

* These components comprise the core QUANDA functionality.

4.1 Investigation of question/question matching improvements

Analysis of the existing mechanisms within FAQFinder was undertaken in an attempt to match a user question and a FAQ question. The major focus was to determine whether new approaches may positively impact question matching accuracy. For example, perhaps modifications made to support question type classification may positively impact the SVO functionality.

Preliminary research yielded encouraging results that were the basis for further research with FAQFinder. The initial research has found a number of limitations in question/question matching, some of which are addressed by changes to FAQ finder (phrase support, named entity support and improvements to SVO identification for example) and others for which the limitation is not easily solved requiring deeper semantics or new approaches (e.g., question context requiring world knowledge to properly resolve).

.

4.1.1 Subject/Verb/Object Consideration

Sentence semantic analysis focuses on extraction of sentence meaning. One aspect of this is the process of determining which words to consider. In some cases this is obvious. Prepositions for example often serve to preserve proper grammar but in most cases impart little influence on sentence semantics. (Sussna, 1993) stated that "prepositions, articles and conjunctions are considered extraneous" when one is attempting to identify the words in a sentence that most contribute to a sentence's meaning.

Given the following sentence: "Do frequent oil changes prolong a car's life span?". In this case the word [a] only serves to preserve proper grammar but imparts no obvious semantic information. In looking further at the sentence, one might distill the text to [Do] [oil changes] [preserve] [life span]. Certainly, the word "car" is also important as it clarifies that the context being referred to is automotive. It becomes obvious that the relationship that is important here is [oil changes] [preserve] [life span]. In short what matters is the subject, verb and object (in this case it would be desirable to consider "oil changes" and "life span" as concepts instead of discrete words).

This is not always all that is necessary as adverbs and adjectives as well as secondary phrases enhance the meaning of sentences. Sentence semantics however, key off of the subject, verb and object. It is with this in mind that FAQFinder has been augmented to consider the Subject/Verb/ Object.

FAQFinder's use of SVO is a potentially important feature for disambiguation. Unlike traditional document retrieval, Internet FAQ questions are typically a single sentence and the matching process focuses on matching a single sentence within a FAQ file. This very small amount of text often presents a problem for a word frequency based match approach. With so little text, each word is valuable and the intended collective meaning of all of the words, or in short, the semantic content becomes important.

The SVO approach attempts to identify the SVO components in the user question, and in questions that have previously been asked residing in FAQ files. Once identified, a matching process attempts to match the SVO components in a user question with their counterparts in a FAQ question. Each SVO component's contribution is weighted equally, and the collective score for a potential match has a range of between 0 and 1.

Progress made in the initial research found limitations in existing techniques that identify the subject/verb/object followed by implementation of the Question/Answer parsing grammar. While the grammar is primarily to support research Question/Answer component matching, an important byproduct is the potential reuse of the results to increase the accuracy of the question/question matching. One of the limitations found in SVO question/question matching was that the rudimentary technique used in identification of the subject/verb and object was

limiting its accuracy.

4.1.2 Integration of Parsing into FAQFinder

The parser effort focused on three areas. First, the grammar used by an existing/available parser was enhanced with additional rules. A second effort that was completed, was to implement a preprocessing step to purge various special characters that affect the accuracy of parsing. Lastly, because parsing accuracy is key to the existing SVO semantic functionality, potential replacements for the existing parser were investigated (FaqFinder depends on CFG a parser that was developed to support basic semantic analysis). The existing parsing accuracy was found to be be sufficient to support the QUANDA matching approach as part of this research.

Parsing of natural language is a difficult task and the use of a Context Free Grammar is the current approach taken within FAQFinder. An accurate parse of question text may increase the potential to identify a sentence's subject, main verb and object. Identification of the SVO for simple sentences is easily achieved if the words can be classified by the Part Of Speech to which they belong. For example, the sentence "The bird flew into the nest" is relatively easily handled because each word can be categorized into its corresponding part of speech, without ambiguity, and because the sentence conforms neatly to a familiar subject/verb/object pattern.

If the English language were generally this well structured, Natural Language Processing would be a simple task. Unfortunately, parsing of language is made much more difficult when sentences include additional words and when the structure of the sentences changes from the simple subject/verb/object paradigm. As sentence complexity grows, words coalesce into logical groupings, modifiers appear, words' part-of-speech may become ambiguous if a given word can serve as both a verb and a noun depending on context, and the orderly pattern of subject/verb/object may be transposed. In essence, the task of properly identifying even the part-of-speech of a word in a sentence can become a difficult task - even for a human being - when the sentence is ambiguous.

In the case of FAQFinder, there is the need to determine whether two sentences are similar in meaning. If there were no synonymy in the English language then simply matching of each sentence's words might suffice in the matching process. This is not the case however, and not only must the matching process consider synonyms but also the possibility that two words might or might not be of the same part of speech even though they are spelled the same. Each word in a sentence must be categorized into the part of speech it belongs in order to proceed in considering synonyms. Two approaches (Brill, 1992) and WordNet (Miller, 1993) are used in FAQFinder to accomplish the POS tagging.

In matching two sentences, the initial inclination is to attempt a match between words in the sentence, first directly if there is an exact match between words and then via consideration of synonyms. Programmatically accomplishing this task requires proper classification according to part-of-speech, however, this often requires that consideration be given to a given word's

function within the structure of a sentence. There is then a logical need to consider the structure of a sentence and this leads to consideration of grammar in a sentence.

The existing Context Free parser was enhanced. As two sentences are considered for a match, each is sent through a parser in an attempt to ascertain the structure of the sentence. Not only must all words in the sentence be categorized according to part-of-speech, but also the parser must be able to match the sentence structure with one or more existing grammar rules. This becomes important when FAQFinder attempts to identify the subject/verb and object of the sentence. If the parsing is not accomplished correctly, a proper selection of Subject/Verb/Object either cannot occur because there are no candidates for the three categories or the wrong words might have been selected. The parser and its grammar, are critical components of the matching process if the SVO components are to depend on them.

SVO semantic functionality depends on the ability to identify the subject/verb/object of a sentence. Should the parser fail to parse a sentence, this will yield a NULL result for the SVO functions to deal with and thus poor performance by the SVO component. With a sufficient number of grammar rules, the hypothesis is that this will give the SVO component an enhanced opportunity to correctly identify the subject, verb and object of a sentence.

The existing parser had about 250 rules. An additional 200 grammar rules were added. These additional rules have increased the effectiveness of the parsing step. While the parsing ability has been increased, this has not yielded significant benefit to FAQFinder accuracy. This is in part due to the inability to accurately identify the subject/verb/object within a user question but it is also highly dependent on the questions themselves. Results presented in Phase I and Phase II of this research do make use of the additional grammar rules.

4.1.3Named Entity Support

Named Entity enhancements center around the following types of words: acronyms, proper names and locations. Such words are of concern in FAQFinder for two reasons. First, any words that are not recognized during tagging have the potential of being mis tagged which leads to parsing failures. Secondly, a word such as an acronym will not have a valid WordNet sense, which will cause it to be incorrectly considered when FAQFinder is attempting to determine senses of words within a sentence being matched.

Named Entities are in some cases, phrases (collocated Ngrams). For example: "John Quincy Adams" and "January 3, 1942" are a proper name and a date. The concept of a phrase as used here, refers to any multi-word group of words that collectively have meaning. Regardless of named entity type, it will be desirable to support syntax variations (e.g., 1/3/1942, 1/3/42, 01-03-1942, etc...).

Named entities are resolved similarly to phrases. As with phrases, early research into FAQFinder accuracy revealed that named entities were not properly recognized and contributed to

FAQFinder inaccuracy. The inaccuracy was partly due to some named entities being misidentified as to their POS. An additional source of error was that named entities often had important semantic relevance that was ignored or misinterpreted.

An example of some of these issues is the named entity “Stock Exchange”. The words have meaning individually and collectively. Individually, “stock” has multiple meanings as a noun and also can be used as a verb. “Exchange” also has both noun an verb meanings. Collectively, “Stock Exchange” has a meaning apart from the meanings of the individual words.

Named entity resolution seeks to preprocess text, recognize named entities and provide the ability to properly handle the terms. Named entities are defined within a named entity form file with the following syntax:

Named Entity Type:POS:Translation(s)

For example:

ENTITY:NNP:NYSE:New_York_Stock_Exchange
PERSON:NNP:Abraham Lincoln:Abraham_Lincoln
PLACE:NNP:United States Of America:United_States_Of_America, USA
THING:NNP:Dawn of The Dead:Dawn_Of_The_Dead, movie

.
Named Entiry support has been added in Phase I of this research. This was accomplished with a common mechanism for named entities and phrases. There are approximately 54,000 named entities and phrases included for both Phase I and Phrase II results..

4.1.4Phrase Support

Phrase support, as defined here, broadly refers to collocated Ngrams as opposed to grammatical phrases. The implementation of phrase resolution within this research effort collectively supports named entities, multi-words and grammatical phrases; although in many cases a multi-word can function as a grammatical noun phrase and a named entity may be a multi-word. The underlying knowledge representation does not need to make such distinctions, though there can be grammatical differences between these constructs.

Phrases often have a collective meaning apart from the meanings of the constituent words. Various researchers have approached word sense disambiguation through the consideration of combinations of words (e.g. Mihalcea, Resnik). The rationale for considering phrases in FAQ question answering (and in FAQFinder) differs from that of Mihalcea's (semantic density) and Resnik (co-located noun agreement). In both Mihalcea and Resnik, the body of text they worked with was much larger and the goals of semantic density and noun agreement was primarily single

word, word sense disambiguation. In the phrase support implemented here, the goal was to provide the correct meaning to a group of words. For example, the adverb phrase "under the table" might equate to "clandestine" if used in the sentence "The money transaction happened under the table". Not only would the words' true meaning be resolved, but the phrase might also be replaced by its resolved single word synonym (clandestine in this case).

If semantic approaches are to be successful in FAQ question answering, it is essential to extract as much meaning as possible from the small amount of text that is involved. This is especially important in the parsing step that identifies sentence components (SVO) and secondarily, some phrases and multi-words have meanings apart from their constituent words, e.g., "top shelf brew", "figure out", "on the other hand", hard and fast", "by default", "component sets", "when in doubt". Of these actual FAQ file phrases/multi-words listed, WordNet only contains a definition for "by default".

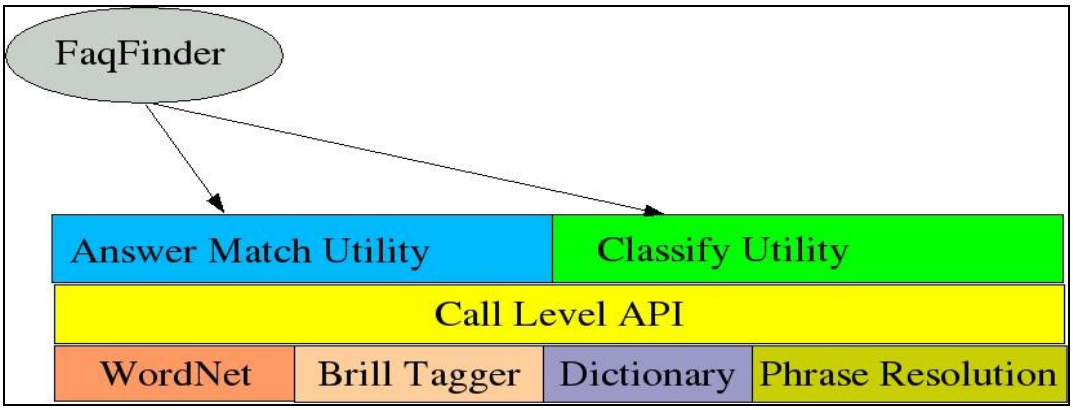
Prior to this research, FAQFinder assigned senses to phrase words individually without considering the meaning imparted by the word combination. This may lead to incorrect sense tagging and later incorrect sentence matching. Such phrases and multi-words also complicate the grammar of the parser, as these word combinations sometimes require unique parser rules to be correctly parsed.

While WordNet does support phrases, this support is not extensive. Furthermore, WordNet does not easily support insertion of new phrases. As a result, another mechanism was needed and has been implemented and used for both phase I and phase II results. Both named entities and phrases are supported for both Phase I and Phase II results and use a common mechanism. (54,000 named entities/phrase rules)

4.1.5POS and SVO Tagging

Part of speech tagging and SVO tagging is performed on both the question and answer text. Both the Brill Tagger and WordNet were found to be needed, as the tagger focuses on the POS a given word plays within a sentence (it is usually accurate but not always) and WordNet provides clearer identification for the potential POS classifications for a given word. It was also found to be the case that WordNet is not comprehensive enough and there was a need to include a full dictionary. The Webster Dictionary was integrated to provide word POS lookup.

All of the word lookup capability and POS tagging is integrated into a tagging and preprocessing layer. This preprocessing occurs as the first step as user supplies a question and as answer text is read from FAQ files. The preprocessing layer is transparent to the answer component matching. The following diagram illustrates this association.



To support the generation of phase I results, none of the extended capability of WordNet was used; both WordNet and the dictionary are used used for word lookup only. During investigation of the matching process, it was found that use of WordNet distance would have potentially resulted in additional question matches. Another observation is that WordNet's synonym capability would have resulted in additional matches. While these are potential future considerations, both phase I and phase II results only make use of basic Wordnet POS capability.

4.1.6Integration of Answer Component with FaqFinder

Two options were considered for FaqFinder integration. One of these was to create a fourth component to complement the Vector, Semantic and SVO components. This approach was discarded because the answer component match was designed with the intent to only return a result when there is is a high degree of certainty that an answer match is correct.

Generation of a match only during situations of high certainty was deemed desirable in light of the experience gained in trying to get answer component word frequency to generate positive results. A lack of discrimination here had a negative impact on accuracy. A similar result occurred when attempting to improve SVO matching because the function used to identify the Subject/Verb/Object could not be made accurate enough. The lack of accuracy in SVO tagging had a negative impact on SVO matching.

To prevent a re-occurrence of the accuracy issues, a decision was made to have the answer component match only return a result when there is a high degree of certainty. In this way, the presence of a result from the answer match component is assumed to have a high probability of being correct. This is intuitively attractive as FaqFinder is a “question” matcher and much success has been achieved using this approach. It is therefore desirable to not to diminish the weight of the question component match functions.

Such an approach allows the vector, semantic and SVO components to equally contribute and the SVO contribution to be overridden if an answer match is present. In the examples that were analyzed, this works quite well. This is because FaqFinder always had difficulty with questions that were terse and colloquial and there are many FAQ questions that fit this pattern. In such cases, a question<-->question match is often not possible and the ability to rely on the answer component affords another avenue.

The following question statistics illustrate the issue:

Question Type	***Number***
SVO Candidates (good potential for SVO match)	209
Terse (difficult matches)	326
Strong Match Potential	407

The total number of questions within the FAQ file test suite was 733. The SVO match therefore has at best the potential to match 28% of the questions within the FAQ files. SVO candidacy is based on the ability to extract a subject, verb and object.

Terse questions were those with 3 or less words within the question or where there were more than 3 words but described a single concept (e.g., Unit Investment Trusts and SPDRS). Terse questions do well with vector based matching.

The Strong Match Potential category is one with questions that are complete or nearly complete sentences. SVO candidates are a subset of this category.

4.2 QUANDA Question/Answer Matching

Although previous attempts at use of FAQ answer text had not yielded any significant increase in question matching accuracy, further research was conducted into mechanisms that have the potential to produce a positive result. In those situations where a successful question<=>question match is unlikely, a match between a question and a FAQ answer component (with QUANDA) has enhanced results.

A question<=>answer match differs from that of a question<=>question match. Question<=>question matching can in many cases focus on the similarity of words (and their meanings) between questions. Word similarity between question and potential answer is generally not helpful because most answers in a given FAQ file use similar terms. Question<=>answer matching can be a more difficult task because it may involve looking for potential answers to a user question within the answer text or the use of other semantic based methods. Searching for specific answers to a question is a difficult process, especially given that

FAQ answers are small amounts of text compared to what is the case with general document search.

This research effort has investigated various techniques for matching a user question with the answer component of a FAQ. This includes question type classification and answer search; both of which are components of QUANDA. It also includes the other components of QUANDA (named entity resolution, phrase resolution, ontology layer). Results achieved in Phase I are provided later in this chapter.

4.2.1 Question Type Identification

User questions often provide clues about what a user is asking for. It is sometimes possible to perform a basic level of disambiguation by identifying whether a sentence begins with an interrogative pronoun (what, who, when, where, which) or adjective (which, what, how) or adverb (what, when, where, why). Such a clue can be valuable in determining for example that a question is asking for a location (where). A "what" question on the other hand provides very little insight due to its generality.

While such simplistic questions occur often, in many cases user question syntax varies significantly due to the English language's support for rich expressionism. Although helpful, it may not be sufficient to rely simply on the presence of an interrogative pronoun or adverb/adjective. One option would be to make use of recent research (Tomuro 2001,2002) that approaches question typing using "paraphrasing". Paraphrasing may improve the results as this addresses the issue of varying sentence structure in an interrogative sentence.

Question Type Identification (QTI) performs a basic level of question analysis and where appropriate, provides this information to a function that scans the answer component of FAQ files. While prior attempts at analyzing the answer component have met with little success, these methods were based on word frequency. My approach with QUANDA is to provide an anticipated answer form when a question form is identified.

Depending on the question type, this requires pattern-matching rules to:

(1)Determine target of the question.
(2)Scan the answer text looking for an appropriate response sentence

For example, if the question is "Where is Belize?" then the pattern (where XXXX resolves to Belize) to be searched for in the answer might be "XXXX is in", "XXXX is located near", "XXXX is close to", "XXXX lies", etc., such that there might be a set of rules defining these possibilities. Such sets of rules are associated with each question type (who, what, where, etc.) and provide for meta fields for nouns or phrases/multi-words that might be unified with actual text.

Other question types such as how questions, (e.g., "How do I get started making beer?"), are much more difficult to answer. This is because the answer to a "how" question may not always provide a clue sufficient to correlate this back to the question. Ideally, the answer to the above question might contain text such as "you can get started making beer by...". Unfortunately, this may occur in only situations where there is a very close match between the user question and one previously asked, where it is more likely that the expected word sequence "you can get started making beer by" might be found. In many cases, a question answering system such as FAQFinder will need to provide an answer that is in the general subject area but perhaps not an exact match in the hopes that the information may be useful to the user.

4.2.2QUANDA Question/Answer Forms

Question Type Identification and Question match share a common capability; the QUANDA question/answer rules. In the section that follows, detailed information is provided on the QUANDA rules repository and the grammar specification (meta language).

Each user question must undergo a step that identifies the question type and resolves the SVO components. The identification of question types is accomplished using a question recognition module that takes direction from a set of rules files encoded using a meta language. This meta language supports not only question classification but also an expected potential answer form to the question.

For example, the following simple entry:

HOW:How can $NPP/$SUBJ $VERB [my] $NN/$OBJ:One may $VERB..$VERB [a] $NN/$OBJ..$OBJ

The above rule contains three colon delimited sections (Question Type, Question Form and Answer Form).

4.2.2.1Meta Language

The language for defining question forms described above has the following syntax:

<variable | word>
[<variable | word>]
<variable | word>..<referrer>
<variable | word> / <qualifier>
<variable | word> / <qualifier>..<referrer>

4.2.2.1.1 Meta Language Syntax

Term	*Definition*
/	Indicates that a qualifier immediately follows
..	Indicates that a referrer immediately follows
Qualifier	Identifier tag (e.g. $VERB)
Referrer	Tag used in answer component to refer to a tag in the question component (e.g. $VERB)
Word	Mandatory word - must be present.
[ ]	Optional word – may/may-not be present. If a word is present, this strengthens the match.
$	Indicates a tag (e.g., $VERB)
!	Indicates exact term match required for this match form. When present at the start of an answer component.
%	Word match overrides POS in the event that the POS of two matched words does not agree.

4.2.2.1.2 Meta Language Tags

Term	*Definition*
$SUBJ	Sentence subject
$VERB	Sentence verb but can also act as a qualifier
$OBJ	Sentence object
$VERBx where x = 1,2,3,...	Secondary verbs
$ADVx where x (if present) =1,2,3,...	adverb
$ADJ where x (if present) =1,2,3,...	adjective
$NNx where x (if present) =1,2,3,...	noun
$NNPx where x (if present) =1,2,3,...	proper noun
$CDx where x (if present) =1,2,3,...	numeric value
$PREPx where x (if present) =1,2,3,...	preposition
$PRONx = where x (if present) =1,2,3,...	pronoun

4.2.2.1.3 Meta Language Examples

$NPP/$SUBJ – A proper noun that is the subject of a sentence.
has/$VERB – specific word that is qualified/identified as a variable ($VERB)
$VERB – The sentence verb.
[my] – Brackets indicate an optional word.
$NN/$OBJ – A noun that is the object of a sentence.
$VERB..$VERB – This answer sentence verb matches the verb of the user question.
$NN/$OBJ..$OBJ – This answer sentence object is a noun and references the object specified in the user question.
How do I/$SUBJ VERB $NN/$OBJ:!You $VERB..$VERB $NN/$SUBJ..$OBJ using a %bobby

Given the question form rule that was previously specified, *"HOW:How can $NPP/$SUBJ $VERB [my] $NN/$OBJ:One may $VERB..$VERB [a] $NN/$OBJ..$OBJ"* one may input the following user question:

How can I sharpen my pencil?

The ClassTag function will produce the following output:

((HOW) (SUBJ.I) (VERB.sharpen) (OBJ.pencil) (How.WRB) (can.$VERB) (I.$SUBJ) (sharpen.$VERB) (pencil.$OBJ))

There are three parts to the ClassTag output; a question TYPE, the SVO identification and the POS and SVO tagging of the sentence words.

The SVO identification field plays an important role in that this component can be used multiple ways. First, the fields are utilized within the new Question/Answer matching function. Secondly, the SVO can be utilized to more accurately identify Subject/Verb/Object for use within the pre-existing SVO question/question matching functions.

Once a sentence is classified (there can be multiple question forms that match a given sentence), an SVO resolution is performed on the answer form. In the question, the answer form would resolve to:

One may sharpen [a] pencil ...

In the above question, the matching answer form may/may-not specifically request the presence of a sentence subject. If it does, then in this case "One" and "I" are not a match. It is also possible that it does not or that the referrer in the answer form is not present, in which case a match is possible. The meta language, ClassTag and matching functions support the concept of

missing SVO components as this naturally occurs in speech and in FAQ files. There is also no requirement that an answer form have referrers. For example, the following is an acceptable answer form:

One/$SUBJ may $VERB [a] $NN/$OBJ..$OBJ

In this example, there is a referrer for the sentence object but not one for the subject.

No attempt is made at paraphrasing of user questions although if this were done it would speed up the ClassTag and matching phases by decreasing the number of rules that must be searched and then applied.

4.2.2.2Meta Language Linkage

Question/Answer rules provide linkage between a question and expected answer form. Form however, is not the only thing needed to successfully match a user's question with a previously asked question's answer. The user's question will in general contain context that will be needed to determine whether a potential answer form contains the same context. For example, the question "How can I change my oil?" may match the following question form:

How can I/$SUBJ $VERB my $NN/$OBJ

The above sentence may match the following sentence within a FAQ file:

"You can change your oil at a gas station"

To facilitate a proper match, we must first find a rule within the question component. There may be many potential matches; many of these identical. This is due to the possibility that the same question form may have many potential answer forms. Once a set of potential forms have been identified, the next step is to determine whether a specific answer is a match for a given answer form. For example, the answer form might be:

You/$SUBJ can $VERB..$VERB your $NN/$OBJ..$OBJ at

Note the $VERB..$VERB and $OBJ..$OBJ elements. The ".." notation indicates linkage between the question and answer components. For the match between a question and potential answer, the following must be true:

4. The user question must match the question form.
5. The potential answer must match the answer form.
6. The references in the answer form must correctly resolve to the question form.

Only if all three of the above requirements are met, is there a match. It should also become clear that the linkage resolution implies both a POS match as well as a word match. There is thus an

opportunity to consider synonyms instead of just a simple pattern match. While the preliminary research has only considered pattern word match, there is an opportunity to increase accuracy by considering synonyms.

The following "HOW" forms have been defined and used to test the answer component matching functions. There were 51 rules used for the preliminary tests. Currently there are 1468 rules to provide coverage for the HOW Question/Answer forms. The following are the initial set of 51 forms:

FILE: ~data/how.forms
#
The matching must try to match the word "first" before trying to
match the POS type. For example, "how soon" is also "how %RB%", however
we must give over-riding preference to "how soon" as this is a specific
instance of %RB% Equivalent Forms
#
#
#
NOTE TO SELF: NEED TO IMPLEMENT ADVERB COMBINATION, e.g.,
#
is made, is purchased, is brewed.
#
This includes the case where the verb/adverb are separated, e.g.,
"How is beer brewed?"
#
#
HOW:What is/$VERB $NN/$SUBJ:$NN/$SUBJ..$SUBJ is/$VERB
HOW:How do $SUBJ $VERB:$SUBJ..SUBJ $VERB..$VERB HOW:How
do I/$SUBJ $VERB [a] $NN/$OBJ:To $VERB..$VERB [a] $NNP/$OBJ..$OBJ HOW:How
do I/$SUBJ $VERB [a] $NN/$OBJ:$NN/$SUBJ..$OBJ $VERB..$VERB HOW:How
do I/$SUBJ $VERB on $NN/$OBJ:$NN/$SUBJ..$OBJ $VERB..$VERB HOW:How
to $VERB $OBJ:To $VERB..$VERB $SUBJ..$OBJ HOW:How much will
[a] $NN/$SUBJ $VERB:[A] $NN/$SUBJ..$SUBJ will $VERB..$VERB HOW:How is/
$VERB [the] $NNP/$OBJ different from [the] $NNP/$OBJ:[The] $NNP/$SUBJ..$SUBJ differs/$VERB from [the] $NNP/$OBJ..$OBJ
HOW:How is/$VERB [the] $NNP/$OBJ different from [the] $NNP/$OBJ:[The] $NNP/$SUBJ..$OBJ differs/$VERB from [the] $NNP/$OBJ..
$SUBJ
HOW:How do I/$SUBJ $VERB [my] $NN/$OBJ:$SUBJ..$OBJ is/$VERB defined as
HOW:How do I/$SUBJ $VERB [my] $NN/$OBJ:This $NN/$SUBJ presents the formula for $VERB..$VERB $OBJ..$OBJ
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:How do you/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How do you/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How do you/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:How do you/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:They [can] $VERB/$VERB..$VERB [the] $NNP/$OBJ..$OBJ
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:If you [want] [to] $VERB..$VERB [the] $NNP/$OBJ..$OBJ
HOW:How do I/$SUBJ $VERB [a] $NNP/$OBJ:You [can] $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How can I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How can I/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How can I/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:How can I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:How can one/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:How can one/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ

HOW:How can one/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:How can one/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:In what way do I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way do I/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way do I/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:In what way do I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:In what way do you/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way do you/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way do you/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:In what way do you/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:In what way can I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way can I/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way can I/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:In what way can I/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:In what way can one/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way can one/$SUBJ $VERB [a] $NNP/$OBJ:You can $VERB..$VERB [a] $NNP/$OBJ..$OBJ
HOW:In what way can one/$SUBJ $VERB [a] $NNP/$OBJ:One way to $VERB..$VERB [a] $NNP/$OBJ..$OBJ is to
HOW:In what way can one/$SUBJ $VERB [a] $NNP/$OBJ:One can $VERB..$VERB [a] $NNP/$OBJ..$OBJ by
HOW:How is/$VERB the $NNP/$SUBJ different from the $NNP/$OBJ:The $NNP/$SUBJ..$SUBJ differs/$VERB from [the] $NNP/$OBJ..$OBJ
HOW:How is/$VERB the $NNP/$SUBJ different than the $NNP/$OBJ:The $NNP/$SUBJ..$SUBJ is very different/$VERB than [the] $NNP/$OBJ..$OBJ
HOW:How is/$VERB the $NNP/$SUBJ different from the $NNP/$OBJ:The $NNP/$SUBJ..$SUBJ is very different/$VERB from [the] $NNP/$OBJ..$OBJ
HOW:How are/$VERB $NNP/$SUBJ $NNP/$OBJ:$NNP/$SUBJ..$SUBJ are/$VERB..$VERB $NNP/$OBJ..$OBJ
HOW:How are/$VERB $NNP/$SUBJ $NNP/$OBJ:$NNP/$SUBJ..$SUBJ is/$VERB..$VERB $NNP/$OBJ..$OBJ

As can be seen from the above rules, it is possible for there to be many matching rules. There is a trade-off (specificity & generality). For the purpose of this test, the rules were left more verbose/specific to more closely align with a specific answer form. The alternative is to implement paraphrasing or modify the rules to be more general. For example, if disjunction were introduced into the language, then "How may I" and "How can I" could be handled by modifying the rule to be "How may|can I". This "logical or" functionality is planned for but has not implemented for the proposal. Additionally, one may more generally state " How $VERB $NNP". This generalization opens the possibility to significantly decrease the number of rules, There is also value in providing a more specific relationship between the user question and the target answer form. For this reason, the logical or "|" will most likely be used during the later stages of this research.

4.2.2.3 Meta Language Grammar Specification

The following describes the Context Free Grammar that fully describes the Meta Language.

Let G be the grammar use to describe the FaqFinder linkage language. G is defined by:

A->δ, where A ε (N ∪ T)*, the set of all strings that are a combination of N and T

where T defines a set of terminals
T={:, ' ', $SUBJ, $OBJ, $VERB, $NP, $ADJ, $ADV, $NNP. $CD, $PRON, $ADV, $PREP, $NN, 1,2,3,4,5,6,7,8,9,0}

and N defines a set of non terminals N={ σ, <WORD>, <META>, <META_R>, <REFERENCE>,<QUALIFIER>, <SPACE>, $SUBJ_TYPE, $OBJ_TYPE,

$VERB_TYPE, $NP_TYPE, $ADJ_TYPE, $NNP_TYPE, $PRON_TYPE,
$ADV_TYPE, $PREP_TYPE, $NN_TYPE, $CD_TYPE}
λ = Ø
<SPACE> -> ' ' | λ
<WORD> ε {any string separated by a space}

with productions:

σ=<QTYPE> :A :B
σ=<QTYPE> :A :<EXACT_MATCH> B

A -> <WORD> <META> <WORD>
A -> A <WORD>
B -> <WORD> <META_REFERENCE> <WORD>
B-> B <WORD>

<QTYPE> -> “HOW” | “WHEN” | “WHERE” | “YESNO” | “WHO”
<WORD> -> <WORD> <SPACE> <WORD>

<META> -> <WORD> <META>
<META> -> <META> <QUALIFIER> <META>

<EXACT_MATCH> -> '!'

<META> -> $OBJ_TYPE
<META> -> $CD_TYPE
<META> -> $VERB_TYPE
<META> -> $SUBJ_TYPE
<META> -> $ADJ_TYPE
<META> -> $ADV_TYPE
<META> -> $NP_TYPE
<META> -> $NNP_TYPE
<META> -> $NN_TYPE
<META> -> $PRON_TYPE
<META> -> $PREP_TYPE

There are three forms to specify a reference on the right side of <COLON2>. In the first case, there is no linkage. In the second, there is a qualifier and linkage reference and the third has only a linkage with no qualifier. All are acceptable syntax.

<META_REFERENCE> -> <META> <QUALIFIER> <META>
<META_REFERENCE> -> <META> <QUALIFIER> <META> <REFERENCE> <META_R>
<META_REFERENCE> -> <META> <REFERENCE> <META_R>

<QUALIFIER> -> '/'
<REFERENCE> -> '..'
<META_R> -> <META>
<$CD_TYPE>-> $CD<NUMBER>
<$VERB_TYPE>-> $VERB<NUMBER>
<$SUBJ_TYPE>-> $SUBJ<NUMBER>
<$ADJ_TYPE> ->$ADJ<NUMBER>
<$ADV_TYPE> ->$ADV<NUMBER>
<$NP_TYPE>-> $NP<NUMBER>
<$NNP_TYPE>->$NNP<NUMBER>
<$NN_TYPE>->$NN<NUMBER>
<$PRON_TYPE>->$PRON<NUMBER>
<$PREP_TYPE>->$PREP<NUMBER>
<NUMBER> -> 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | λ
<NUMBER> -> <NUMBER> <NUMBER>

The above grammar specification is lacking one feature that must be explicitly stated as follows:

Linking Rule:

If there exists a <META_REFERENCE> in the form:

<META_REFERENCE><QUALIFIER><REFERENCE><META_R>

e.g.

σ -> *<QTYPE> <COLON1> <WORD> <META> <COLON2> <META_REFERENCE> <QUALIFIER> <REFERENCE> <META_R>*

then <META_R> specified in the above form must have a corresponding matching string to the left of <COLON2>; i.e., <META_R> := <META> .

It must be stated that there are no positional limitations upon where <META> and <META_R> appear with respect to other <META> elements on their respective sides of <COLON2>.

Example:

The following example will use the standard C syntax /* text */ to represent informational comments. This information is not be considered part of the grammar.

YESNO: $VERB the $NN/$SUBJ $NN/$OBJ ?:The $NN/$SUBJ..$SUBJ $VERB..$VERB $OBJ..$OBJ

If the above production would be used to match the sentence:

Is the sky blue?

The ClassTag function would transform the sentence into:

Is/$VERB the/$PREP sky/$NN blue/$NN

Note that the above is not yet tagged with SVO tags; this step follows next. SVO tags are realized from the question match within the Question/Answer forms.

The matching function would then attempt to match both the question and answer components simultaneously. For the purposes of this example, we only consider the question component for brevity's sake. The match would find a question form of the following:

$VERB the $NN/$SUBJ $NN/$OBJ

This is clearly straight forward given the output of the ClassTag function.

4.2.3QUANDA Answer Match

Various approaches to question matching were attempted and discarded. The complexity of the matching process is significant and can quickly become unmanageable. Matching of user a user question to an answer component sentence occurs once a question is classified. This matching makes use of the resolved potential answer forms, of which there may be many (refer to the prior section on question classification). The key to the matching functionality is that it focuses on the SVO and other text within the user question and FAQ file sentences.

In Phase I, the “HOW” question type was chosen because it can pose a challenge in the matching process. This is because the universe of potential answer forms is generally much larger than for “WHO” or “WHEN” for example. Given the reliance on SVO and the meta language, this allows for simpler matching that focuses on the key elements within the matching text.

In addition to "HOW", a preliminary implementation of " WHAT" was attempted to test the core matching capability for other question types; this resulted in the match of one "WHAT" question type.

The approach described within this thesis differs from existing approaches such as Webclopedia (Hovy, 2001). Where Hovy identifies a corpus of approximately 140 Question/Answer types and seeks to match user questions into these 140 question types, the approach described within this research uses a much larger corpus that is very specific in its target of Question/Answer types. This allows very precise specification of Question/Answer forms.

A further distinction is that the FAQFinder research is foundationally based on consideration of

Subject/Verb/Object as the focus of semantic content. The sentence subject/verb/object are given special consideration in this respect. Added to this is a much more granular consideration of individual words within the sentence allowing the development of linkage/relationship between a user question and potential answer candidate. Partly because of this, this research requires a much larger corpus of Question/Answer matching rules. The goal is to produce accurate results which is an important consideration due to the need to integrate Question/Answer match into FAQfinder's existing suite of match capability; inaccuracy in Question/Answer match directly impacts the overall FAQfinder accuracy which is dependent on multiple matching techniques.

The answer component match functionality is implemented as a two step process. The first step is the tagging process defined by the ClassTag function. Every user question must first pass through the ClassTag function. Upon exit, a question is classified as to question type (e.g., HOW, WHEN, ...), each word is POS tagged, and a potential answer form is returned. This result then becomes the basis for all matching functionality against the FAQ answer components.

The answer component itself plays a role the second step of the matching process. It is at this step that the potential answer form(s) that were identified in the first step are matched against each sentence in the answer component. We are thus looking for a sentence in the answer component that matches a potential answer form. It is however more than just a match of a potential answer form. If this were the case then there would be a very large number of answer candidates.

The potential forms define in essence a template, but this template must then conform to the linkage rules (the references) that point back to the original user question. The linkage mandates that specific sentence items (such as subj/verb/object) are a match between the original user question and the potential answer form. It is this linkage that makes the match highly selective.

The answer component sentences are not put through the ClassTag() function. Every word in an answer component is sent through a tagging process to determine potential POS roles; e.g., the word fly has both a noun and verb usage.

The following is the pseudo code for the matching function used for the (basic) answer component matcher:

```
Let     Q=User Question
        QP=Parsed User Question
        PA_Array=Potential Answer Forms

QP=ClassTag(Q);       /* Tagged user question */
PA_Array=PotentialAnswerForms(QP);       /* Get potential answer forms for question*/

for each PA in PA_Array
do
        if (MatchAnswer(QP, PA, MatchResult) == yes)
        then
```

```
                /* Add to result */
                TotalMatchResult=AddToMatch(MatchResult)
        end_if
done

/*
 * The MatchAnswer function
 */

MatchAnswer(QP, PA, MatchResult)
{
        let AnswerText = the answer component text
        result=no

        for each sentence in the answer text
        do
                TaggedAnswerSentence=ClassTag(sentence)

                if(MatchSentence(QP, PA, TaggedAnswerSentence, MatchResult) = yes)
                then
                        result=yes
                end_if
        done
}

/*
 * MatchSentence attempts to match the expected answer component (there are most
 * likely multiple) with answer text that has been sent through the ClassTag() function
 * to be tagged for POS and SVO.  Note that having a match for Subject,Verb,Object
 * is considered a requirement if the anticipated answer form specifies one.  Additionally,
 * All expected words & Meta Tags in an expected answer form are also required,
 * however, extra words within the answer text are acceptable as long as the expected
 * answer form did not specify an exact/strict match.
 *
 * The matching proceeds with the understanding that an answer text sentence may
 * have multiple false starts; i.e., we may have a sentence that had a leading phrase
 * that must be ignored.  In this situation, the algorithm will restart at the
 * beginning of the potential answer form.
 *
 * Not shown in the logic below due to the desire to keep the pseudo code simple,
 * is the "exact match" feature – indicated with "!".  The exact match would require
 * the seek functions to result in a failure if there were words in the answer text that
 * were not expected in the answer form.
 */
MatchSentence(QP, PA, TaggedAnswerSentence, MatchResult)
```

```
{
        StartPosition=TaggedAnswerSentence

restart:
        /*
         * Check to see if we are at the end of the TaggedAnswerSentence. If so
         * return “no” indicating that the match failed.
         */
        if (StartPosition == NULL)
                return no

        /*
         * If we come to the end of the Answer Component it means
         * that we've successfully matched/resolved all of the tags and
         * words thus far and that there is a match.
         */
        for each word in PA Answer Component
        do
                switch(word) {
                case PlainWord:
                        if(SeekToWord(word, TaggedAnswerSentence, StartPosition) == NULL)
                                StartPosition=SeekToNextWord(StartPosition)
                                goto restart
                        break
                case POS MetaTag:
                        /* here we will examine each word in the answer text to
                         * see if the POS is a match with what the answer form
                         * requires. A MetaTag may for example require a verb.
                         * We must therefor verify that the current answer text
                         * word is a verb.
                         */

                        if(POS(word) != SeekToTag(word, TaggedAnswerSentence,
StartPosition) == NULL)
                                StartPosition=SeekToNextWord(StartPosition)
                                goto restart
                        break
                case QualifierTag:
                        /*
                         * The qualifier is a way to specify the POS that is required
                         * for a successful match and to identify the role a word plays
                         * in a sentence.  For example, $NN/$SUBJ indicates that a
                         * noun is expected and that it will fulfill the role of the sentence
                         * subject.  In this case $SUBJ is the qualifier.  Another example
                         * is move/$VERB as is $VERB/$VERB1.  In the latter case,
```

```
                        * $VERB1 is specifies an auxiliary verb (not the primary
                        * verb) in the sentence.
                        */
                       if(POS(word) != SeekToTag(word, TaggedAnswerSentence,
StartPosition) == NULL)
                              StartPosition=SeekToNextWord(StartPosition)
                              goto restart
                       break
               case ReferralTag:
                       /*
                        * Special situation - we need to resolve with the user question.
                        * The issue to be resolved is – does the referral tag in the
                        * current answer text sentence match the target in the user
                        * question? For example, in this question form snippet:
                        * HOW:...$NN/$SUBJ...:$NN/$SUBJ..$SUBJ
                        * The ..$SUBJ in the answer form must be matched with
                        * $NN/$SUBJ in the question form. Note that not only is
                        * there a need to analyze the referral but also to resolve
                        * the tags to the actual words.
                        */
                       if ((tag = SeekToTag(word, TaggedAnswerSentence, StartPosition)) ==
NULL)
                              StartPosition=SeekToNextWord(StartPosition)
                              goto restart
                       if(POS(word) != POS(tag))
                              StartPosition=SeekToNextWord(StartPosition)
                              goto restart
                       /*
                        * OK – the POS is a match.  Now look at the referral to
                        * make sure that there is a match with the word in the user
                        * question.
                        *
                        */
                       if(Referral(tag, PA) != SeekToTag(tag, PA, StartPosition))
                              StartPosition=SeekToNextWord(StartPosition)
                              goto restart
               End_Switch
       End_Do
       return yes      /* A Match!  */
}
```

The approach described above is a simplistic representation of the matching process. The actual code is considerably more complex. Additionally, the computational requirements are substantial as each sentence within the answer component of a FAQ file must be individually parsed/tagged and compared against each potential answer form; of which there usually are

multiple. Large FAQ files and those with large answer components require sizable computational resources. The computation is easily partitioned into parallel streams and may lend itself to distribution across multiple computing elements.

4.3WEB Harvest of FAQ Files

The corpus of FAQ files in use with FAQFinder contains both "answerable" and "unanswerable" test question sets. While many of these FAQ files are still relevant, it was desirable to increase the number of files as this may produce a more accurate measure of FAQFinder performance. One of the potential problems with continuous use of the same data is that enhancements may become too heavily "tuned" to a given set of data, and not truly reflect results that might be achieved for a more generalized data set. . By introducing multiple FAQ files and user questions, this provides a way to yield a more accurate measure of improvement..

Another need for additional FAQ files is to satisfy the need for a large corpus of information that can be used once the FAQFinder WEB page is implemented. If FAQFinder is to be useful (and used), it will need to have a large body of knowledge from which to draw relevant answers on a broad topic set.

The following is a list of tasks that were performed:

1) A set of tools were created that provide assistance with cleansing and verification.
2) A set of tools were created to help automate the process of formatting of FAQ files so that they conform to a standard format (Question/Answer tags).
3) FAQ files were collected from the WEB and categorized
4) A new list of answerable and unanswerable questions was extracted for a new set of test/control questions.

The newly gathered and formatted FAQ files were added to the existing corpus of FAQ files.

4.4Phase I Results

An effort was undertaken in Phase I to verify the results for the research outlined in this proposal. The first major assumption was that FaqFinder question answering could be improved by focusing on the answer component of a FAQ file (QUANDA). Prior research into consideration of the answer component did not yield significant improvements in FaqFinder accuracy. The QUANDA approach differs from previous attempts that focused on word frequency consideration. Instead of word frequency, QUANDA focuses on deeper analysis of the FAQ question and text within the answer component. This analysis attempts to parse the question, identify the Subject, Verb and Object and then use the Subject, Verb and Object to identify a likely match within the answer component. The positive contribution of the SVO is the second major assumption that was tested by this Phase I effort.

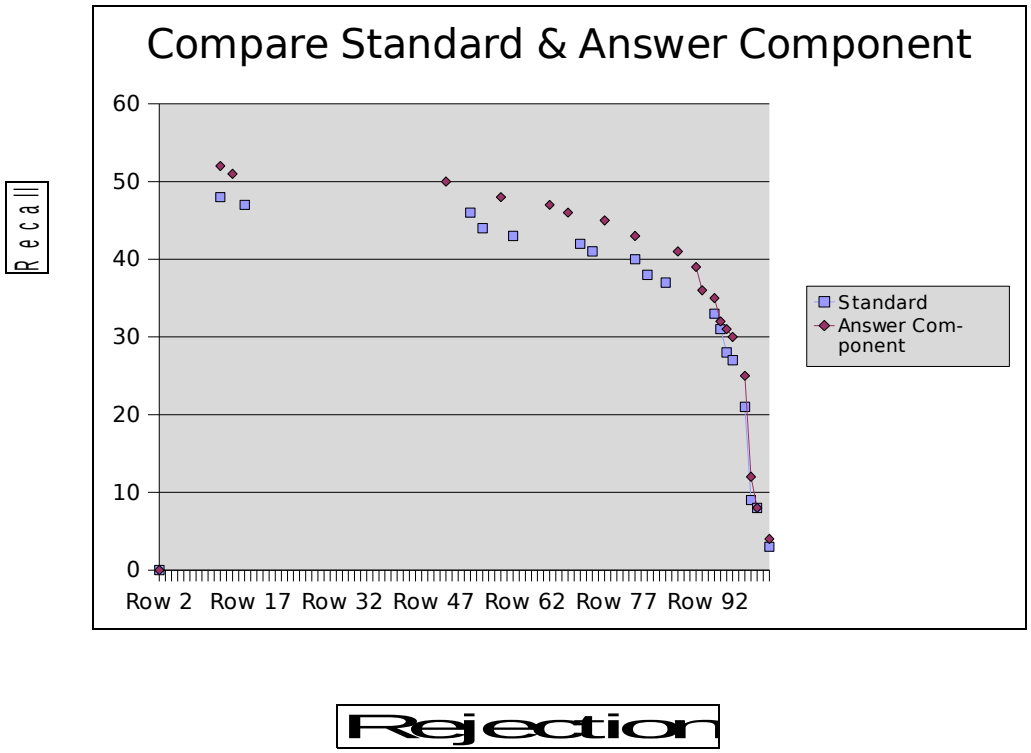
Below is a table that indicates the questions that were successfully matched by QUANDA and their related FAQ file.

Question	***FAQ File***	***Question Matched***
How Are P/E Ratios Computed?	investment03.tag	3
What is CCD?	digitalcam01.tag	1
How is the NYSE different than the NASDAQ?	investment06.tag	6
How do I compute my capital?	investment03.tag	6
How to make coffee?	coffeeMisc.tag	**, 1, 2
How do I compute future value?	investment02.tag	*, 16
How do viruses spread?	virusPart4_4.tag	**, 13, 17, 25
How do I buy a CD?	investment04.tag	*, 10
How much will a tax adviser cost?	investment02.tag	7

*Multiple matches of the same correct target answerable FAQ question.
**Additional valid matches found by matching function that were not part of the target answerable set.

Integration of answer component matching functionality has produced a positive result in overall FaqFinder accuracy. It must be pointed out that overall, the results achieved using the new FAQ question suite were lower than early results that used the pre-existing question/FAQ file suite. (The pre-existing FAQ file suite has been the standard suite used by other FAQFinder researchers). This can be explained by the lower quality of the FAQ files that comprise the new test suite. The new suite was needed to ensure that various enhancements made to FAQfinder had not become dependent on the specific FAQ file data for their accuracy. It is necessary from time to time to run tests using new sets of data to ensure that achieved results are data independent.

Questions listed within FAQ files are seldom complete and grammatically correct sentences. This is the case with the FAQ question that was used to produce these results. There was no attempt to clean up FAQ files, or modify the questions within them to provide more insight in the answer component. The only modifications were to embed tags to denote where a question and answer began/ended. This is in the spirit of FAQ files that currently exist on the Internet.



The above results show that answer component match has the potential to increase FaqFinder accuracy. This is intuitively obvious given that question matching often is destined to fail as the questions themselves (either in the FAQ file or the user's question) provide little or no information about the user's intent (without context/understanding of the subject). Secondly, the answer component of the FAQ file does provide the potential to enhance the question match in one case; where the user question is grammatically reasonable. Previous attempts at targeting the answer component using word frequency analysis produced little benefit because questions within a FAQ file are too closely related to each other (subject grouping).

While word frequency analysis of the answer component has not yielded positive results, the answer component holds the potential for enhanced accuracy. These positive results provide hope that question answering accuracy can be increased further within FAQFinder. The primary requirement for additional benefit is that the user question must provide enough information to deduce the user's intent. If this requirement is met, additional techniques for enhancing accuracy

can be employed (semantic analysis of the text within an answer component for example). User "intent" in this case implies that a question can be classified into a question category and matched with a potential answer format.

It must be stated that answer component matching is a complementary technique to question<-->question match. In situations where there is a high degree of similarity between a FAQ question and a user's question, it would be short sighted to ignore a direct question<--->question match.

In the current implementation, the answer component match is given a 33% contribution weight by the fact that it can override the SVO result, if there is a match returned by the answer match component. In the case where there is a question <---> question match (especially an exact or near exact match), the SVO match and hence the answer component match should have less impact on the result. In practice, this is what happens. If a match is returned by the answer component matching functions, it will at most wield a 33% contribution with the vector and semantic components contributing the remainder.

4.4.1 Raw Results

Please refer to Appendix B for a listing of the RAW results of the matching process.

4.4.2 Near Matches

The following are questions that that have potential matches with additional modifications (that have not been attempted because these are outside the scope of this research)..

Potentially Matching Question	***What is needed***
How do I slow down on blades?	Support for "ing" or stemming
How are viruses created?	WordNet distance or synonym support
What is a trojan horse?	DEFINITION type support
What is Ginsing?	Stemming or plural/singular handling
What are grind plates?	DEFINITION type support

4.4.3 Expansion of Question/Answer Forms

To support a broad category of WHO question types, an effort was undertaken to collect a large list of question forms and then to generate potential answer forms. A list of HOW question forms were extracted from the Webclopedia corpus (Hovy, 2001). For each question form, a potential set of answer forms were then generated, using the Question/Answer meta language. This effort yielded a total of 1468 meta language rules.

To verify that the large number of language rules had not negatively impacted the previously achieved results, the results were retested. It was was found that the results did not decrease and that the new meta language forms positively impacted the previously achieved results by matching an additional question. This is a welcome result as it demonstrates that the Question/Answer forms offer sufficient discrimination. There is thus hope that as the language rule corpus grows, its accuracy should not decrease and that the results achievable by FaqFinder will continue to improve.

4.4.4Abandoned Functionality

A number of things were tried and discarded to improve Brill Tagger accuracy. This is especially important in Question/Answer match because the Question/Answer forms are looking for specific POS tags. Both question and answer forms for example expect that the subject will generally be some noun/pronoun form. If the tagger misidentifies a noun as a verb for example, there is little hope that a proper match can be made. If the mis-tag is within the question classifier, the question will not be considered.

It is not that the tagger fails altogether. It is most often that the tagger provides the most likely POS tag, followed by other potential taggings. The research did attempt to make use of these additional taggings. In some instances this was helpful. In other situations, consideration of alternate POS tags had undesirable consequences in that inappropriate question classification matches resulted.

After much consideration, it was decided to only use the Brill Tagger's “best” (first) guess. This implies that the matchers will have a lower yield, but the net effect is that inappropriate matches are minimized (but not eliminated). If the tagger's best guess is wrong, as the analysis of the questions/answers indicates, this will lead to a loss of question classification accuracy and will impact answer match as well.

If another tagger is found to be more accurate, this would clearly have a positive impact. Unfortunately, the existing tagger cannot provide this level of assurance.

4.4.5Phase I Summary

The focus of the Phase I effort was to determine the efficacy of QUANDA as a component that would enhance FAQFinder accuracy. Based on what I have been able to demonstrate with Phase

I, the effort will continue with a Phase II. Phase II builds upon Phase I by expanding the capability of QUANDA. QUANDA rules in Phase II include additional WHO and WHAT question forms.

5 Phase II Research Final Results

The Phase I research was my initial attempt at producing favorable results using a variety of techniques including QUANDA. In this chapter, the Phase II results are presented. Phase II extended the capability of QUANDA to include additional question type support with the hope of generating additional improvements in FAQFinder accuracy. This hope has yielded a positive outcome as described further in this chapter.

In addition to FAQFinder, Phase II results are also provided for QUANDA's application to general text search.. This includes applying QUANDA to TREC (Text Retrieval Conference) question track data. Additonally, QUANDA was compared with the three major search engines (YAHOO, GOOGLE and ASK) to see how it compares in accuracy. All of these results are provided in sections below.

The following table describes new functionality added for Phase II.

Item	Description
WHAT and WHO question type support	3539 WHAT rules and 782 WHO forms were added.
Redesign of the interface to the Brill Tagger	This was necessary to overcome reliability issues with the Brill Tagger.
WEB Page	A WEB page was developed to test QUANDA question classification capability.
QUANDA SVO tagging migrated back into FAQFinder.	QUANDA's more accurate SVO tagging was migrated to FAQFinder. Unfortunately, this did not yield a positive result (refer to the relevant results later in this chapter)

5.1 FAQFinder Phase II Final Results

There are two sets of results produced. The first set (Question Set One) is from rerunning the same question set that was used in chapter four. The chapter four questions were collected by having users enter questions related to a narrowly defined topic set. Answer component matching included support for HOW question types.

A second question set was collected from the WEB by searching for FAQs within a given domain. Half of the FAQs were used as a question source and the other half as the FAQs against which matching is accomplished. The most recent run utilizing the WEB questions added support for WHO and WHAT question types.

The two sets of results are quite interesting in that they appear to differ significantly. After analysis of the results, there are good reasons for the differences as will be revealed later in this section.

5.1.1 FaqFinder+

As improvements are made to FaqFinder, it would be appropriate to distinguish these improvements from the original effort. For this reason, the FaqFinder with additional enhancements provided by this research, will be referred to as "FaqFinder+"

5.1.2 QUANDA

FaqFinder+ enhancements that are focused on Question/Answer are integrated with FaqFinder+ but are in reality a separate stand-alone capability. From this point forward, the Question/Answer functionality will be called "QUANDA".

5.1.3 Question Set One – Results With HOW/WHO/WHAT Support

Below are the results for the question set described in chapter four. The goal with retesting the older question set was to see any potential improvement that the QUANDA component might yield with the addition of WHO/WHAT.

5.1.3.1 Question Set One - FaqFinder+ Data

The following are the FAQ files that were used in testing the question list described in chapter 4 and in retest using additional support for WHO and WHAT question types. The file names are descriptive of the FAQ topics.

File Number	FAQ File
1	Cancer01.tag
2	CancerAnimals.tag
3	coffeeAndCaffeine.tag
4	coffeeMisc.tag
5	CullinaryHerb.tag
6	DietAndCancer.tag
7	digitalcam01.tag
8	DVD01.tag
9	DVD02.tag
10	DVD_Law.tag
11	InternetExplorer01.tag
12	investment02.tag
13	investment03.tag
14	investment04.tag
15	investment05.tag
16	investment06.tag
17	investment07.tag
18	investment08.tag
19	investment09.tag
20	investment11.tag
21	investment12.tag
22	investment13.tag
23	investment14.tag
24	investment15.tag
25	investment16.tag
26	investment17.tag
27	investment18.tag
28	investment19.tag
29	investment20.tag
30	KidneyCancer.tag
31	linux01.tag
32	linux02.tag
33	linux03.tag
34	linux04.tag
35	linux05.tag
36	linux06.tag
37	LinuxMeta.tag

File Number	FAQ File
38	LungCancer.tag
39	MedicinalHerb.tag
40	MSOffice01.tag
41	OpenOfficeAsian.tag
42	OpenOfficeBuild.tag
43	OpenOfficeCommunityCouncil.tag
44	OpenOfficeCommunity.tag
45	OpenOfficeDevelopment.tag
46	OpenOfficeIssue.tag
47	OpenOfficeLicensing.tag
48	OpenOfficeMain.tag
49	OpenOfficeMirrors.tag
50	OpenOfficeOverview.tag
51	OpenOfficeTechnology.tag
52	OpenOfficeXML.tag
53	ProstateCancer.tag
54	PubMed.tag
55	rollerblade01.tag
56	Scientology.tag
57	SpiceIsle.tag
58	TesticularCancer.tag
59	USB01.tag
60	virusMac.tag
61	virusMini.tag
62	virusPart1_4.tag
63	virusPart2_4.tag
64	virusPart3_4.tag
65	virusPart4_4.tag
66	Windows01.tag
67	Windows02.tag
68	Windows03.tag
69	Windows04.tag
70	Windows05.tag
71	Windows06.tag
72	WindowsForms.tag
73	WindowsMediaPlayer01.tag
74	WindowsSecurity01.tag

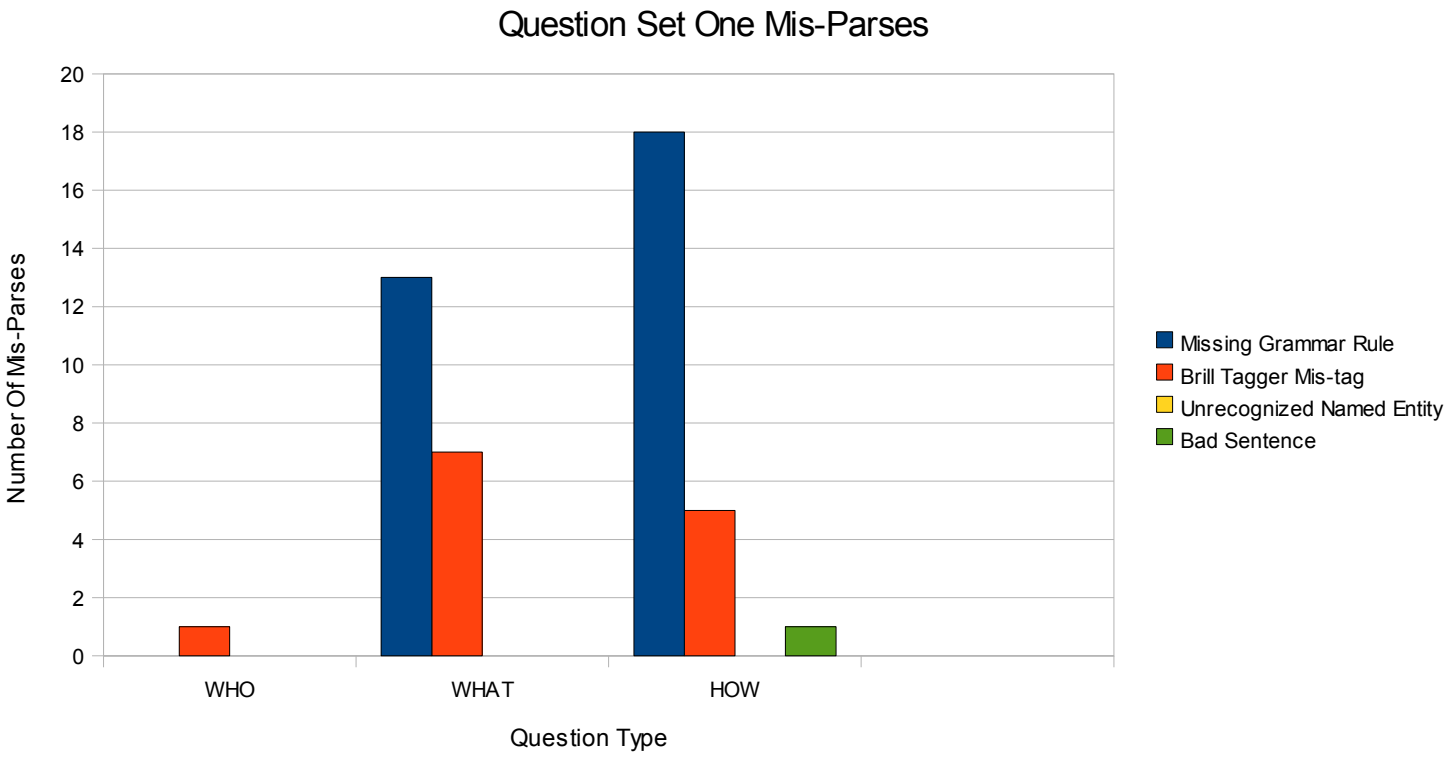
5.1.3.2 Questions Set One and FaqFinder+

The list of questions was the same as was used in chapter 4 and may be found in Appendix F. There a total of 120 answerable and 62 unanswerable questions.

5.1.3.3 Question Set One - FaqFinder+ Parse Failures

There were quite a few parse failures but it must be stated that a majority of the questions were not of WHAT/WHO/HOW type. The prevalent question type was the WHAT category. There was no effort to select only specific question types. This is true for both FaqFinder+ questions and for the TREC suite. In both cases, WHAT question types were the most common.

The following graph provides a break-down of the reasons for parse failures.



Detailed information on each question and reason for failure is provided in Appendix F.

5.1.3.4 Question Set One - FaqFinder+ Match Successes

The following questions were matched correctly by the QUANDA match logic:

Question Number	Question	FAQ Question Match	FAQ File
1	How do viruses spread?	25	virusPart4_4.tag
2	How do I buy a CD?	10	investment04.tag
3	How do I compute my capital?	6	investment03.tag
4	How do I compute future value?	16	investment02.tag
5	How do I buy stock options?	4	investment05.tag
6	How is the NYSE different than the NASDAQ	1	investment06.tag
7	How do I buy bonds?	6	investment04.tag
8	How are P/E ratios calculated?	3	investment03.tag
9	What is IRR?	1	investment03.tag
10	What is panax	29	MedicinalHerb.tag
11	What are limit orders?	4	investment19.tag
12	What is the NYSE?	1	investment06.tag
13	What are stock options?	1	investment05.tag
14	What is CCD?	33	digitalcam01.tag
15	What is the NASDAQ?	5	investment06.tag
16	What is the AMEX?	7	investment06.tag
17	What are I Bonds?	6	investment04.tag
18	How do computer viruses spread?	25	virusPart4_4.tag
19	What is an option?	1	investment05.tag
20	What is an IPO?	5	investment13.tag
21	What is a virus?	4	virusMac.tag
22	How much will a tax adviser cost?	2	investment02.tag
23	How to make coffee	1	coffeeMisc.tag

The new matches were all of question type WHAT and HOW. There were no questions of type WHO that matched but there were only two questions of this type in the collected question set: "Who owns LINUX?" and "Who can help me with LINUX questions?"; both of which were not answerable.

WHO and WHAT question types were anticipated to be easier to handle as the questioner is often asking for a very specific/precise answer. As an example, answer forms for "What is IIR?" might be:

IIR is a type of

IIR is a form of

IIR is a manner of

IIR is an indication of

IIR is a sign of

IIR is a kind of

IIR is a style of

and so on... These are all similar forms and QUANDA rules can be parameterized and generalized to handle a variety of potential word matches.

There were four additional matches of type "HOW" than what was reported in chapter 4 results. This is due to fixes to match answer logic that were discovered during testing of TREC results. When the original results were reported, the match logic did not correctly interpret linkages between potential answers and the question for some POS tags (ADJ/ADVERB/PREP).

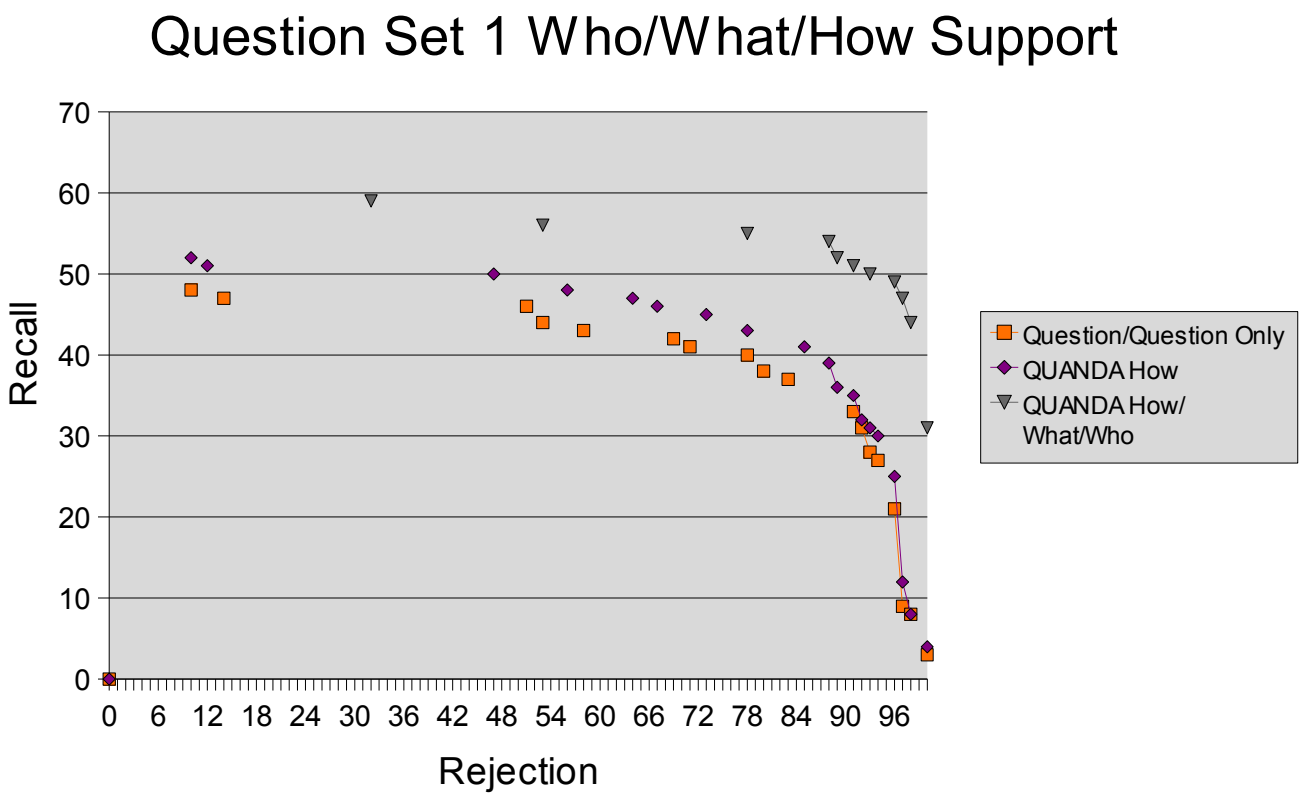
5.1.4 Question Set One - FaqFinder+ Results Analysis

Below are the results that were achieved when QUANDA component functionality was applied to the question set used in section 4.5.10. This is the same question set that was collected from a web interface of users asking questions from a predefined set of topics. There was a significant increase in FaqFinder+ recall when this question set was tested with the addition of WHAT and WHO question types. Analysis of the results indicates the following reasons for the increased accuracy:

a) The nature of the questions was such that question/question match had difficulty making matches on the question only.

b) The addition of "WHAT" question type support in QUANDA component provided additional matches.

c)There were no matches on "WHO" question type but there were only two WHO questions in the question mix.



The above graph shows that QUANDA logic has made a significant contribution to the overall results in terms of recall in general and in recall at higher levels of rejection. The latter is more significant in that the goal of this research was to increase accuracy of FaqFinder+. The accuracy, as show above, has improved substantially for this question set.

There are three sets of results depicted. The first is question/question only and is the baseline. The question/question results were improved by the inclusion of QUANDA support for the HOW question type; this is as was reported in chapter four. Addition of WHAT QUANDA support has produced results that are significantly better.

There is similarity between these results and what was achieved with TREC results reported later in this chapter. HOW and WHAT question types appear to be the most often encountered in question answering and in the TREC question suite.

It must also be mentioned that QUANDA component's contribution varies depending on:

1) Whether question/question produced a good match.
2) Whether the question was parsed correctly by QUANDA.
3) Whether the match pattern corresponded to a potential match.

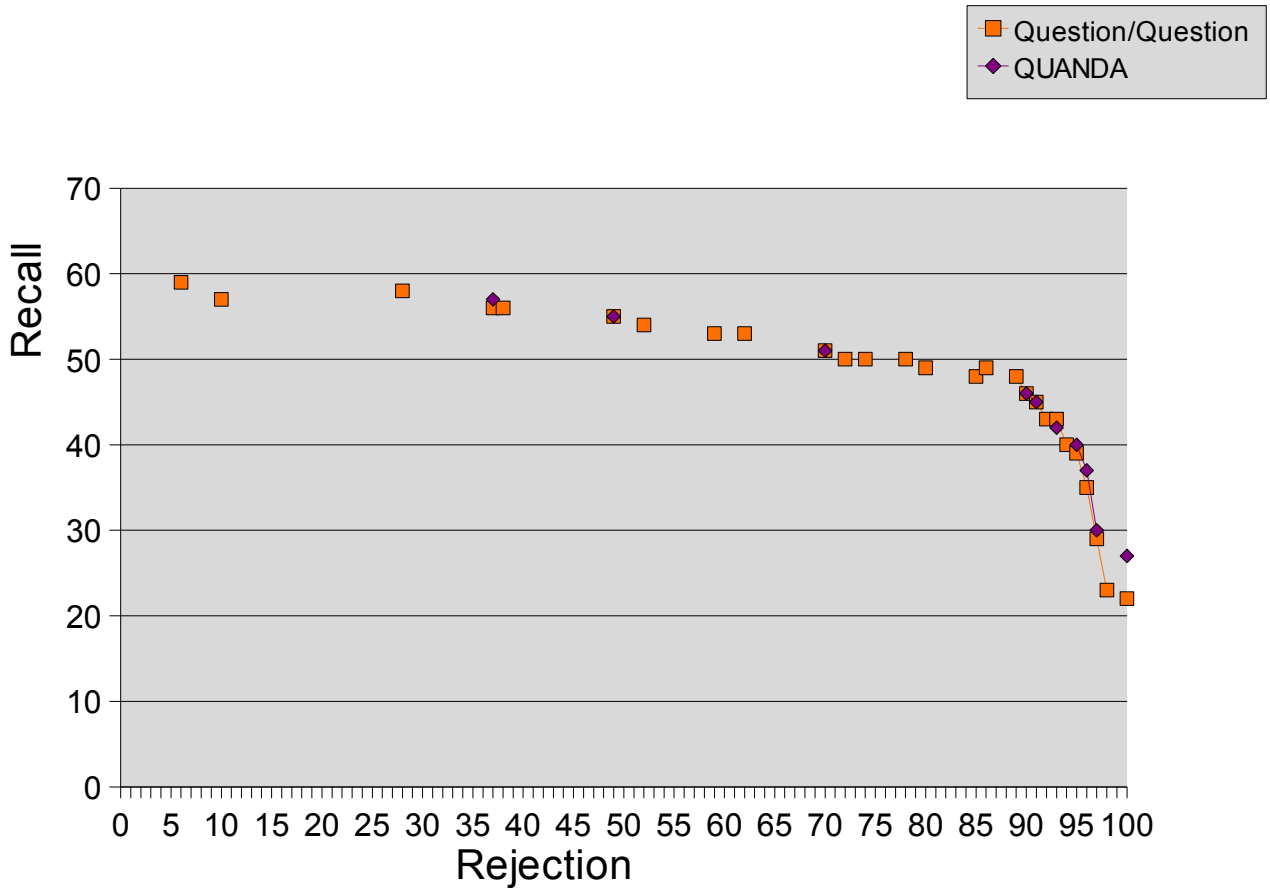
Later in this chapter, results for a new question set will be reported and these will show that QUANDA component was less of a factor. This must be qualified. It is not that

questions/answer did not contribute to the results; it did. QUANDA's contribution in the later results were masked because the question/question component was also successful on the same questions.

There is synergy between the question/question and QUANDA components within these results. Both provide benefit. There are times when a user question matches very closely a question that was previously asked, and taking this into account provides positive results. There are also situations where an answer to a question resides within a body of text that does not closely correlate with a previously asked question. It is here that QUANDA provides the most benefit. The two mechanisms working together have produced the best results as demonstrated by this research.

5.1.5 Web Question Set With HOW/WHO/WHAT Support

The following graphic details a comparison of the results achieved for two test runs; one with and one without QUANDA. The results indicate that the recall is very similar. The significant difference occurs at 100% rejection where clearly the addition of QUANDA has provided benefit. The difference is 27% for QUANDA vs 22% for question/question. This is positive and indicates that QUANDA has achieved its goal of increasing the accuracy of the FaqFinder+. Later in this chapter, a deeper look into why the results for this new question set do not show the same significant improvement that was seen for the previously reported question set. In essence, question/question performed very well on this question set and thus was not as dependent on QUANDA for an improvement.



5.1.5.1 Web Question Set - FaqFinder+ FAQ Data Files

The following table lists the FAQ files that were used for testing the new question set. In total there were 37 FAQ files containing a total of 1446 questions.

File	# of questions
Cancer01.tag	43
CancerAnimals.tag	9
DietAndCancer.tag	46
DVD01.tag	182
DVD02.tag	81
DVD_Law.tag	155
InternetExplorer01.tag	44
KidneyCancer.tag	8
LungCancer.tag	7
MSOffice01.tag	22
OpenOfficeAsian.tag	7
OpenOfficeBuild.tag	13
OpenOfficeCommunityCouncil.tag	17
OpenOfficeCommunity.tag	6
OpenOfficeDevelopment.tag	38
OpenOfficeIssue.tag	2
OpenOfficeLicensing.tag	27
OpenOfficeMain.tag	8
OpenOfficeMirrors.tag	8
OpenOfficeOverview.tag	11
OpenOfficeTechnology.tag	13
OpenOfficeXML.tag	13
ProstateCancer.tag	10
PubMed.tag	21
Scientology.tag	81
SpiceIsle.tag	8
TesticularCancer.tag	10
USB01.tag	16
Windows01.tag	50
Windows02.tag	12
Windows03.tag	31
Windows04.tag	41
Windows05.tag	42
Windows06.tag	37
WindowsForms.tag	80
WindowsMediaPlayer01.tag	118
WindowsSecurity01.tag	129
Total	1446

5.1.5.2 Web Question Set - FaqFinder+

The list of questions were gathered using Google and Yahoo WEB search. Selection of questions to use had to be coordinated with searches for relevant FAQ files. This is because there were instances where websites had borrowed questions and FAQ files and incorporated them.

The collection approach was to find FAQ files within a particular topic area that were created independently but within the same topic area. Given two independent FAQ files, the first was used as a FAQ and the second was used as a collection source for answerable and unanswerable questions.

No attempt was made to determine how successful FaqFinder+ might be in actually matching a given question or to select questions/FAQs that were better suited. This approach does potentially penalize the results that could be achieved if the FAQ files' question components are terse or not descriptive of the actual FAQ contents. For example, the question

"What is anamorphic DVD?"

has a potential match in the "DVD01.tag" with the question:

"Why is the picture squished, making things look too skinny?".

In the above, while the question and FAQ file question are potentially related, this relationship is not obvious without deeper topic knowledge. In some cases, a relationship is not obvious even with deep topic knowledge and human logic. FAQ files often do contain appropriate answers; it is just not obvious by looking at the FAQ's question. It is also often the case that a given FAQ may answer *other* questions because the author chose to expound beyond a specific answer to a question.

The approach taken by this research was to assume a real-world situation where there are no guarantees about question/topic agreement, and then to see how well the mechanism***s*** work. Given that FaqFinder+ has mechanisms to address both situations, this seemed an appropriate approach.

A list of questions was collected that were deemed "unanswerable" even though they were relevant to the topic area. Here, questions might have words in common with FAQ text, however, a human reader would not be able say that their question was truly answered. Proper disposition of unanswerable questions is as important as correctly identifying an answer – if not more so. For large bodies of text, a large number of false matches render results unusable. FaqFinder+ mechanisms include word frequency based approaches and it is possible for these to be misled by words that are in common between a user question and FAQ questions. For example, the question:

Which studios support DVD?

Has little to with the following question:

Since the Sony PS3 will support BDs, will the Xbox 360 support HD DVD?

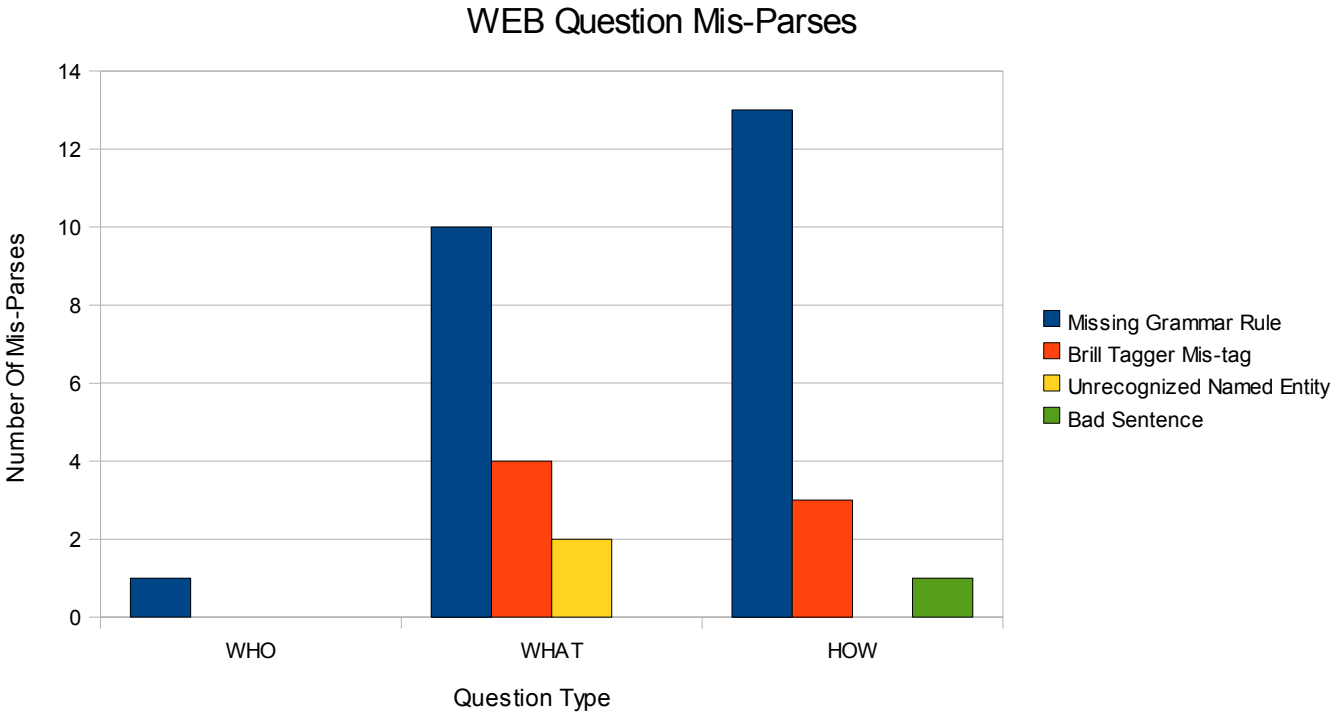
These two questions have two words in common; *DVD* and *support*. There is a possibility that a match between these two words may happen and thus yield a false answer. The goal of this FaqFinder+ research is to minimize the possibility that such matches will occur. Testing of mechanisms through introduction of appropriate "unanswerable" questions is thus an important component of FaqFinder+.

A detailed list of the WEB question is provided in Appendix G.

5.1.5.3Web Question Set - FaqFinder+ Parse Failures

As with the previous question set, not all questions were parsed correctly by QUANDA. The reasons for this include complexity of sentence structure for which there is no parse rule, and tagger mis-tags where a word is not correctly identified as to its part-of-speech. If a question cannot be parsed correctly, then the contribution of QUANDA is negated. In this case, QUANDA does not negatively impact FaqFinder+ results.

Below are the the statistics for parse failures of the "answerable" question set.



The WHAT question parse failures were primarily due to missing rules and Tagger mis-tags.

There was an issue with terms AC-3 and MPEG-2 not being identified properly due to pre-parse stemming that is not handling the '-' as part of the name. Given that all other testing thus far ran with the same issue, this was not remedied for this final test. It is possible that proper identification might have improved the results. For the most part, the HOW question mis-parses were due to missing grammar rules. There were three Tagger mis-tags and one sentence had a double question that was not handled by QUANDA. Refer to Appendix G for details of all WEB Question parse failures.

The FAQFinder "unanswerable" question set had parse failures as well, and while less of an issue for QUANDA precision, it deserves mention. Of the 91 unanswerable questions, 64 were not parsed correctly. While this seems low, 26 out of the 64 were of question types that are not currently supported by QUANDA. This then leaves 27 out of a possible 65 or approximately 41% which parallels parse results for the answerable set.

5.1.5.4 Web Question Set – Performance

Inclusion of QUANDA in FaqFinder+ has resulted in a significant increase in run-time to answer a question. This was expected as there is a significant computational effort in parsing and analyzing the much larger body of text within the answer component. The final test ran for almost 7 full days on a single AMD CPU equipped PC with a clock speed of 1.8Ghz. This equates to:

7 x 24 x 3600 = 604800 seconds

Given that there were 100 answerable and 91 unanswerable questions, this roughly equates to 3166 seconds per question or on average just under one hour. In practice, the run time is directly impacted by the amount of text within the answer component. Also, not all questions are parsed correctly by QUANDA and a (NIL) result is returned much more quickly. As a result, run times for correctly parsed questions exceeded an hour of CPU elapsed time on average.

The performance of the mechanism is not factor that would limit its use. There are a variety of optimizations that can be made to make it function and return valid results in seconds instead of the protracted run-times reported here. A variety of techniques have been used in the past that provide benefit. Pre-parsing of text is one mechanism that can help significantly. Another option is to distribute the computational effort across multiple CPUs. This could be achieved by having individual CPUs be responsible for subsets of the data analysis. For the purposes of this research, performance is not a limiting factor.

5.1.6 Web Question Set - FaqFinder+ Results Analysis

The result achieved is a little different than what was found in prior testing with the question set used in chapter four. One of the interesting things to note in this result is that FaqFinder+ question/question match did identify the correct answer in a number of questions where QUANDA also provided the correct answer. As a result, QUANDA did not provide the differentiation that would have been achieved had the question/question match guessed incorrectly. The two positives here are that (1) QUANDA did not detract from the results that

are achievable with question/question match and (2) QUANDA did improve the results.

Both sets of results taken together (older and newer question sets), have demonstrated improved accuracy and recall of FaqFinder+. This is encouraging as it shows that deeper analysis of text does provide benefit. Progress is not "easy" as the mechanisms that generate this benefit are more complex than statistical methods.

Complexity is one aspect but so is the computational effort in producing an answer. Computational requirements for the data sets and questions used in this research have been substantial. Large scale production quality versions of this functionality would require both optimization of the approach and massively parallel computing platforms. Optimization might be achieved by preprocessing of data sets and question forms. This would provide performance more appropriate for a user accustomed to search engine response times.

Optimization might also take the form of parallelizing the computation of individual data files and questions. Instead of having a single logic stream to examine a data set, perhaps splitting up the search such that sections (or lines) of a data set are handled in parallel would be appropriate. QUANDA often produces multiple potential answer forms. These too would be candidates for parallel processing.

In summary, the benefit for FaqFinder+ is evident in the results that were achieved. For the approach to be practically usable this will require additional effort.

5.2 TREC Data Results

This section provides the results that were achieved with QUANDA on TREC data from TREC-9. Testing was limited to HOW, WHAT and WHO question types. This limitation is primarily due to the large effort required to create QUANDA forms for each question type. The following table lists statistics about QUANDA forms:

Que Type	Number of Que Forms	Complexity	
		Avg Variables Per Que/Ans	Avg Terms Per Que/Ans
WHO	790	18.9	12.7
HOW	1838	20.8	17.1
WHAT	3525	23.2	15.9

In looking at the above table, WHO question types appear to have simpler QUANDA forms and in general, this was demonstrated to be true in the types of questions encountered both within TREC and FaqFinder+ question suites. As for the HOW question type, this required more variables (and thus descriptive information) than WHO types but less than WHAT types.

An interesting observation is that in comparing HOW and WHAT question answer forms, HOW QUANDA forms utilized fewer variables and WHAT types but averaged more terms. This can be explained by the fact that WHAT type complexity is high even though the average number of terms is lower than the HOW type. WHAT type forms utilized more variables and thus accepted generalized terms where HOW types were written to be more specific in the actual match terms. An increase of variable use enhances match potential but also can lead to undesirable matches. For example, specifying a $VERB variable as opposed to the explicit verb "ran" enhances the re-usability potential of the QUANDA form.

It was found that the decision to use variables must be weighed against the attractiveness of using explicit terms where ambiguity and undesirable matches will not occur. This is of special importance where the Question/Answer pair is very weak, as weak questions and/or answers can introduce inappropriate matches. For example, the TREC question "Who was Buffalo Bill Cody?". This can make use of the following QUANDA forms:

WHO:Who/$SUBJ $VERB $NNP/$OBJ:$ANSWER $VERB..$VERB $OBJ..$OBJ
WHO:Who/$SUBJ $VERB $NNP/$OBJ:$OBJ..$OBJ $VERB..$VERB $ANSWER

The above rules are generalized, allowing matches that are in close proximity to a potentially desirable answer but are not exactly what we are looking for. For example, "Buffalo Bill Cody was born in... " would not be an appropriate match but would potentially be reported by the above rules.

In the table below, the HOW question type shows a low precision of 50%. This is due in part to the presence of a weak rule similar to the above WHO rule, and also is due to the smaller set of correctly parsed HOW question types.

Ambiguity of the question and its equally non-specific answer form can be problematic as this result demonstrates. Additional specificity may be called for in this situation. Simply expecting an answer ($ANSWER) may not always be specific enough since in essence, the only references to the question are the object ($OBJ) and the verb which in this case is a weak verb. For this reason, attention must be given to creating appropriate and specific rules.

5.2.1 TREC Data Results For Top 5 Documents

Two different sets of results are presented for the TREC data where these correspond to runs considering the top five documents and top ten documents. This was done to determine the effect to increasingly larger bodies of text.

Below are the results that were achieved when the system considered the top five documents. The runtime utilizing a dedicated 2 CPU AMD 64bit system was approximately 60 hours.

The below table shows matches that correspond to the TREC top five answers from the *topdocs*

files.

Que Type	Total	Num Parsed	Correct Answer	Wrong Answer	Prec	Recall	#1 Top TREC Answer*	#2 Top TREC Answer*	#3 Top TREC Answer*	#4 Top TREC Answer*	#5 Top TREC Answer*
WHO	113	97	18	1	95%	18%	9(8)	7(6)	1(1)	8(1)	4(3)
HOW	52	41	2	3	40%	10%	1(1)	1(1)	0	0	0
WHAT	319	264	18	2	90%	6%	12(9)	4(3)	1(1)	7(4)	4(1)
Other	209	0	0	0	NA		NA	NA			
Total	693	402	38	6	86%	10%	22(18)	12(10)	2(2)	15(5)	8(4)

NOTE: A given question may have matched more than one answer and thus the total number of answers may be larger than the number of correctly matched questions. The table above lists both total answers and the (best) answers where weight is given to the better answer. If for example, a question is answered by both a TREC top#1 document and a TREC top#3 document, then the column "***#1 top TREC Answer***" will contain ***1(1)*** and the "***#3 Top TREC Answer"*** column will contain ***1(1)***.

It also should be pointed out that analysis of the top five TREC answers has revealed that in some cases, the *top answer* should not have been the top answer and QUANDA was in these instances able to perform a more accurate match. The following is a list of questions for which QUANDA determined a better candidate for the top answer:

Question #	Question	Better Answer
395	What is saltpeter	TREC topdoc #4
402	What nationality was Jackson Pollock	TREC topdoc #2
439	What is Nine Inch Nails	TREC topdoc #4

5.2.2Computing The Mean Reciprocal Rank – Top 5 Documents

The Mean Reciprocal Rank is given by the formula:

Mean Reciprocal Rank

$$MRR = \frac{1}{n}\sum_{n}^{i=1}(1/rank_i)$$

Where n=number of attempted questions

MMR for parsed questions

MRR_WHO=1/97 (8/1 + 6/2 + 0/3 + 1/4 + 3/5) = 11.85/97 =.122
MRR_HOW=1/39 (1/1 + ½) = 1.5/39 .038
MRR_WHAT=1/225 (8/1 + 3/2 + 1/3 + 5/4 + 1/5) = 10.95/225= .048

MRR_TOTAL=1/361(8/1 + 10/2 + 1/3 + 6/4 + 4/5) = 23.45/361 = .054

MMR using strict TREC measurement

MRR_WHO=1/113 (8/1 + 6/2 + 0/3 + 1/4 + 3/5) = 11.85/113 = .104
MRR_HOW=1/52 (1/1 + ½) = 1.5/52 = .028
MRR_WHAT=1/319 (9/1 + 2/2 + 1/3 + 4/4 + 1/5) = 12.2/319= . 038

MMR_TOTAL=1/693(18/1 + 9/2 + 1/3 + 5/4 + 4/5) = 24.883/693 = .036

Type	MRR Parsed Questions Only	MMR TREC Compliant
WHO	.122	.104
HOW	.038	.028
WHAT	.054	.038
Total	.070	.036

The above Mean Reciprocal Rank results are given to provide an indication of how QUANDA fared using the TREC metric. Two measurements are provided; one that only considers parsed questions and the other a strict TREC measurement. Comparing results against the published TREC results shows that QUANDA scores surprisingly well. QUANDA was competitive with other researchers' published results. While 90% of the published results exceeded QUANDA's

results, QUANDA was run "as is" on the TREC data. Given an opportunity to optimize for TREC, there is a reasonable expectation that these results could be improved.

There is not a way to determine and compare recall and precision for TREC results. It can however be stated that QUANDA yields a high precision.

5.2.3 TREC Data Results For Top 10 Documents

Below are the results that were achieved when the system considered the top 10 documents. The runtime utilizing a dedicated 2 CPU AMD 64bit system was approximately 130 hours.

The results were moderately higher than when the top 5 documents were considered. Examination of the documents, shows that TREC data files are not as ideally suited to question answering as FAQ files where a FAQ author's goal was to provide an answer to a question. While an article within the TREC corpus may be on-topic, an answer to a question may not reside within the document or it may require consideration of the document as a whole to ascertain an answer.

This is not to say that the approach is not appropriate. While the match yield may be low, the results show that the precision is excellent. This leads to the potential for QUANDA to be used in conjunction with other methods to produce benefit.

Que Type	Total	Num Parsed	Correct Answer	Incorrect Answer	Prec	Rec	TREC Top Answer					
							#1	#2	#3	#4	#5	>#5
WHO	113	97	19	3	86%	23%	11(8)	7(6)	1(1)	3(1)	10(3)	10(1)
HOW	52	41	2	3	40%	12%	1(1)	1(1)	0	0	0	
WHAT	319	264	22	3	88%	9%	14(8)	4(3)	1(1)	8(6)	4(0)	11(4)
Other	209	0	0	0	NA		NA	NA				
Total	693	402	43	9	83%	13%	27(17)	12(10)	2(2)	11(7)	14(3)	20(5)

5.2.4 Computing The Mean Reciprocal Rank – Top 10 Documents

Two measurements are given below; one for only those questions that are parsed and one that is in strict compliance with TREC measurement. In either case, QUANDA fared at the bottom if we are to compare the approach against other TREC results. This is not the only consideration as

precision is the number one goal for QUANDA. QUANDA did outperform some of the reported TREC results as can be seen in Appendix I.

MMR for parsed questions

MRR_WHO=1/97 (8/1 + 6/2 + 0/3 + 1/4 + 3/5 + 1/6) = 11.77/97 =.121
MRR_HOW=1/39 (1/1 + ½) = 1.5/39 = .038
MRR_WHAT=1/225 (8/1 + 3/2 + 1/3 + 6/4 + 0/5 + 4/6) = 12/225 = .05

MRR_TOTAL=1/361(17/1 + 10/2 + 1/3 + 7/4 + 3/5 + 5/6) = 25/361 = .07

MMR using strict TREC measurement

MRR_WHO=1/113 (8/1 + 6/2 + 0/3 + 1/4 + 3/5 + 1/6) = 11.77/113 = .104
MRR_HOW=1/52 (1/1 + ½) = 1.5/52 = .029
MRR_WHAT=1/319 (8/1 + 3/2 + 1/3 + 6/4 + 0/5 + 4/6) = 12/319 = .04

MMR_TOTAL=1/693(17/1 + 10/2 + 1/3 + 7/4 + 3/5 + 5/6) = 25/693 = .04

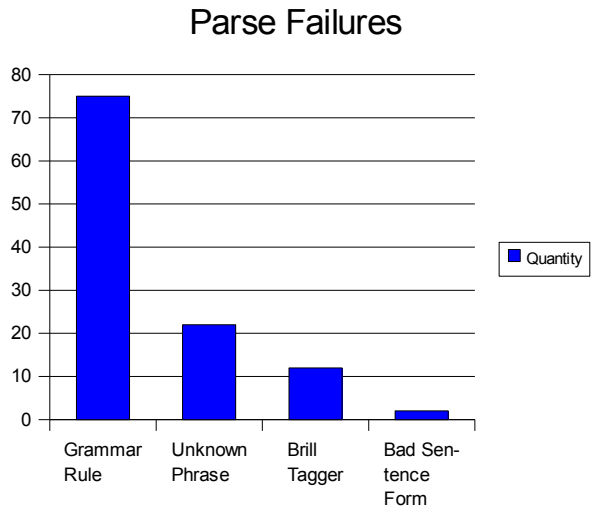
Type	MRR Parsed Questions Only	MMR TREC Compliant
WHO	.121	.104
HOW	.038	.029
WHAT	.05	.04
Total	.07	.04

When the top ten documents were considered, the MMR values did go up slightly. This is because there were additional matches in the additionally considered documents. Precision did go down slightly as there were incorrect matches as well. These incorrect matches were due to overly simplistic match rules that it might have in retrospect been better to eliminate. Still, the results remain quite good.

Although not allowed to complete due to the long run time, a run of top 15 documents was attempted and did not appear to be finding any additional matches. It appears from the three runs that the top five documents are most fruitful and match potential declines steeply beyond five documents.

5.2.5TREC Parse Failures

The tables below list the TREC questions that failed to parse correctly along with an explanation. Reasons for failure fall into a narrow set of categories, the largest of which are: missing grammar rules, unrecognized phrases/entities and Brill Tagger mis-tags. The following graph categorizes the parse failures:



As the above graph indicates, the majority of parse failures were due to missing rules. In some cases, there were multiple reasons for a parse failure. Poor sentence form confused both the Brill Tagger and caused POS mis-tags. Given the poor sentence form, this also proved problematic for the sentence matching logic. The sentence matcher is reasonably able to handle differences gracefully as long as the POS tags are correct.

Detailed information on each question is provided in Appendix I.

5.2.6 TREC Results Analysis

Running QUANDA against the TREC data was a very challenging test of the research approach and of the implementation itself. TREC data is much different from FaqFinder+ data. FaqFinder+ data is a set of targeted responses to questions and thus it is expected to contain text that answers the target question. This data's format/syntax is to a large degree easier to anticipate. For example, if a FAQFile deals with changing the oil in an automobile, it is anticipated that it might answer a question such as “How do I change my oil?”. Additionally, we might expect that somewhere within this FAQFile, there may be potential answer forms such as:

To change your oil...

You can change your oil...

Changing your oil is ...

Changing your car's oil can ...

Changing your engine oil...

etc...

It is thus easier to create a set of QUANDA forms that might not only address changing your oil but also generically an action that addresses a "How to" question.

TREC data on the other hand was not created to answer a specific question. TREC data is comprised of a set of periodicals that were written to report news and provide for interesting reading. As such, the expectation that TREC data will contain "answers" to questions is not frequently rewarded.

Before beginning the effort to analyze TREC data, the expectation was that it might not perform as well here as on FAQ files. While this has been the case, the analysis has shown the approach to have benefit. It is true that the percentage of questions that were correctly answered is small (48 out of a total pool of 693). There is another aspect of the results that warrants consideration; accuracy.

If a question was properly parsed, the system did a reasonably good job (approximately 10%) of finding an appropriate answer. This is significant and in sync with the original goal that was set for FaqFinder+ improvement. The problem and challenge in question answering is precision. It is not helpful to return a large number of possible answers where only a small subset are appropriate. The system only returns an answer when it actually finds what its rules indicate as an appropriate response. If an appropriate response is not found, the system does not return anything. This is a desirable feature within FaqFinder+ where the goal is to improve precision. It is also valuable within a search of TREC documents as well.

Text analysis benefits from a variety of techniques as other researchers have demonstrated (Moldovan, 2000), (Cardie, 1999), (Chen, 2001), (Harabagiu, 2001)). For example, frequency-based word analysis has multiple benefits including speed and recall. While many techniques have shown benefit, any single approach has both strengths and limitations. Word frequency analysis without synonym consideration may not provide adequate coverage for example. A limitation of word frequency approaches is lack of precision.

Combining multiple approaches has yielded benefits demonstrated by FaqFinder+ research. While a word frequency approach may be used for an initial scan of documents, the returned answer set might then be fed to another mechanism such as the semantic approaches within FaqFinder+ The problem that is encountered in bringing different approaches together is to determine which mechanism to trust when each method proposes different results.

It is here that this research can provide benefit. QUANDA can be trusted to provide an accurate result if it provides any result at all. This is helpful because it can be a component of much

broader solution and can be trusted to improve precision (when it contributes to a result).

The challenge with making the system function on TREC data is the large Question/Answer form repository that must be created to address the much richer variety of text that is encountered within periodicals.

Another challenge with TREC style data is the large amount of data that the system must deal with. While FAQ files are neatly categorized and a potential search space for any given question is fairly small, TREC data is extensive and the search space is orders of magnitude larger. Preprocessing and indexing is helpful but the search space is still potentially large and will continue to grow as the body of knowledge (periodical repository) grows (some of the test runs took weeks to complete). These issues are all solvable with existing well known approaches.

5.2.6.1 Results Comparison With TREC 9 Results

Expectations for QUANDA having a significant positive result with TREC style questions and data sets were not high when first considered. The nature of TREC datasets is that they were not written to answer a specific question as FAQ files are. Periodicals and other publications were authored to entertain and to report on topics from a news perspective and with the intent to topically educate the reader. As such, TREC data is semantically (the richer vocabulary used within periodicals for example imparts a variety of meanings) and syntactically different (sentence structure within FAQ files is much more pointed toward answering a question) from FAQ files.

It was encouraging to see that QUANDA performed as it did. While the number of question matches was not high, the precision was very good. This was the research goal: to increase the accuracy of FaqFinder+ The result achieved on the TREC data set supports this goal. The guiding principle/requirement was that it is better to (not) return an answer than to return a poor or incorrect answer. This approach has generated favorable results even when applied as a general purpose search tool within the TREC corpus.

5.3 TREC Questions using ASK, YAHOO and GOOGLE

One of the committee recommendations was to compare QUANDA with other search tools. ASK.COM was selected as the primary tool because it most closely resembles FaqFinder+ functionality in that users often just enter questions instead of focusing on keywords. This appeared to be a more appropriate comparison than Yahoo or Google as these have traditionally been associated with keyword type searches. As the results below show, this selection was appropriate and has shown not only that QUANDA is a viable approach, but also that ASK.COM is a very good search tool that is implemented with more than just a frequency-based approach.

Results are also presented for GOOGLE and YAHOO. While GOOGLE and YAHOO are keyword-based general search engines, they are the largest and most popular. These additional results provide depth to the results via validation of the word frequency based approach used by (ASK.COM) and by showing that ASK.COM's approach is more precise. This was the expectation.

The challenge with attempting this comparison is that ASK.COM does not attempt matches against the same data that TREC results are based upon. To some extent, this is not an apples-to-apples comparison. This limitation is not of critical importance to the results as we are not trying to compare the which approach returns more answers. Instead, the goal is to see how the approaches fare in terms of "precision". We thus are primarily interested in whether ASK.COM returns a correct result as the top result.

5.3.1 Parse Failure Results for ASK, GOOGLE and YAHOO

The below table lists the results achieved by sending the TREC test questions through the ASK.COM, GOOGLE and YAHOO websites. Only the questions that failed to be parsed by QUANDA are listed below. Results for the questions that were correctly parsed are provided in the section following. This was done to provide a more consistent comparison between the two tools.

ASK.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	ASK TOP 1 Success	ASK TOP 1 Fail	ASK TOP 1 Precision	QUANDA Precision*
WHO	29	8	78.00%	12	25	32.00%	95.00%
HOW	5	12	30.00%	2	15	12.00%	40.00%
WHAT	71	35	67.00%	33	73	31.00%	90.00%
Total	104	55	65.00%	47	113	29.00%	87.00%

*NOTE: QUANDA precision in this column is only provided for comparison only. These numbers are applicable to (correctly parsed questions) described later in section 5.4 .

GOOGLE.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	GOOGLE TOP 1 Success	GOOGLE TOP 1 Fail	GOOGLE TOP 1 Precision	QUANDA Precision*
WHO	35	2	94.00%	11	26	30.00%	95.00%
HOW	13	4	76.00%	4	13	23.00%	40.00%
WHAT	92	14	87.00%	34	72	32.00%	90.00%
Total	140	20	87.00%	49	111	31.00%	87.00%

*NOTE: QUANDA precision in this column is only provided for comparison only. These

numbers are applicable to (correctly parsed questions) described later in section 5.4 .

YAHOO.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	YAHOO TOP 1 Success	YAHOO TOP 1 Fail	YAHOO TOP 1 Precision	QUANDA Precision*
WHO	29	8	78.00%	17	20	46.00%	95.00%
HOW	12	5	70.00%	4	13	23.00%	40.00%
WHAT	75	31	67.00%	38	68	36.00%	90.00%
Total	116	44	72.00%	59	101	37.00%	87.00%

*NOTE: QUANDA precision in this column is only provided for comparison only. These numbers are applicable to (correctly parsed questions) described later in section 5.4 .

The results above show a number of interesting things. The first is that ASK.COM is quite good at finding an appropriate answer as shown in the "Precision" column. This is assuming that we use the first page of returned results by ASK.COM. For the purposes of this test, we are most interested in how ASK.COM is at returning a valid answer as the "top" answer. This is a fair comparison because QUANDA is judged this way and must be if we wish to preserve accuracy.

When analyzed for return of the "top" answer, ASK.COM's precision drops significantly (on average about 50%). When compared against QUANDA, ASK.COM is only about 30% as accurate. This must be qualified in that ASK.COM provided better coverage of answering questions. These results will be discussed in more detail below. QUANDA is providing value from a precision standpoint.

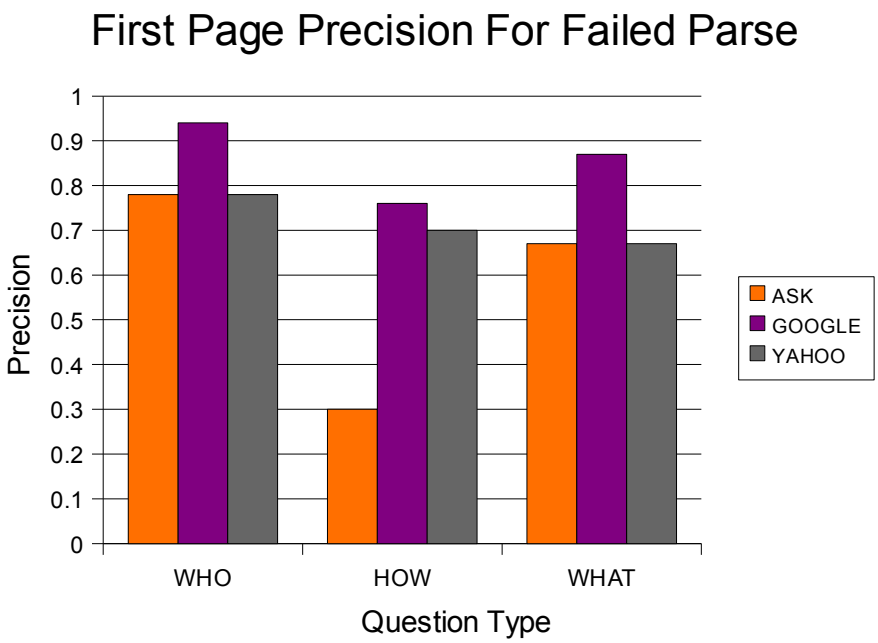
In comparing ASK.COM to GOOGLE and YAHOO, it is interesting to see that these two search engines provided more precision than ASK.COM. The results show the following:

1) GOOGLE and YAHOO outperformed ASK.COM on precision in both the first page results and in finding the "TOP 1" answer.

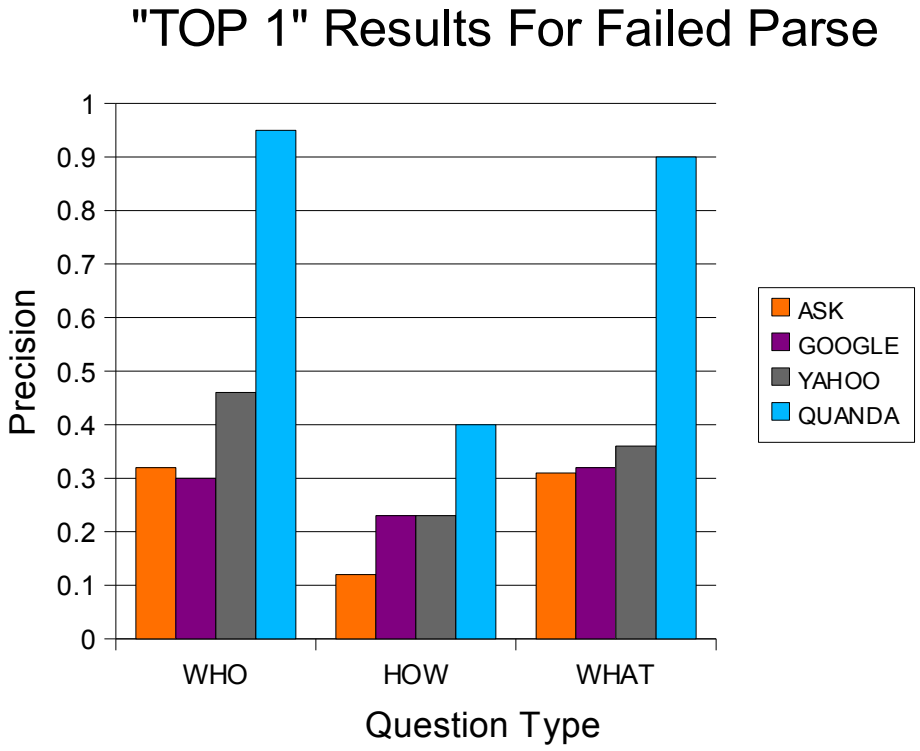
2) GOOGLE more consistently returned an appropriate answer in the first page (top 10) than both ASK.COM and YAHOO.

3) YAHOO, although not as consistent in returning an answer in the first page of results, returned more "TOP 1" answers than both ASK.COM and GOOGLE.

The following graph shows each of the methods' relative precision:



While the above graph shows the precision for the first page, below is the graph comparing the 3 search engines and QUANDA where only the "TOP 1" answer (best answer) is considered:



As was mentioned above, YAHOO seems to do very well in comparison to GOOGLE and ASK.COM – at least for the failed questions.. ***Even though QUANDA did not provide results for "failed questions", its results for "successfully parsed" questions are being included to provide a frame of reference for the type of precision QUANDA can provide.*** The above graph shows that when we consider only the top answer, QUANDA precision fares very well against the search engines.

5.3.2 Parse Success Results for ASK.COM, GOOGLE and YAHOO

The below table lists the results that were achieved by sending the TREC test questions through the ASK.COM, GOOGLE and YAHOO websites. Below are the results for the questions that were successfully parsed by the QUANDA parser for ASK.COM, GOOGLE and YAHOO search engines in separate tables.

ASK.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	ASK TOP 1 Success	ASK TOP 1 Fail	ASK TOP 1 Precision	QUANDA Precision
WHO	86	11	90.00%	43	54	44.00%	95.00%
HOW	18	23	44.00%	4	37	10.00%	40.00%
WHAT	211	53	80.00%	115	149	43.00%	90.00%
Total	315	87	78.00%	162	240	40.00%	87.00%

GOOGLE.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	GOOGLE TOP 1 Success	GOOGLE TOP 1 Fail	GOOGLE TOP 1 Precision	QUANDA Precision
WHO	87	10	90.00%	42	55	43.00%	95.00%
HOW	29	12	71.00%	8	33	19.00%	40.00%
WHAT	231	33	87.00%	115	149	43.00%	90.00%
Total	347	55	86.00%	165	237	41.00%	87.00%

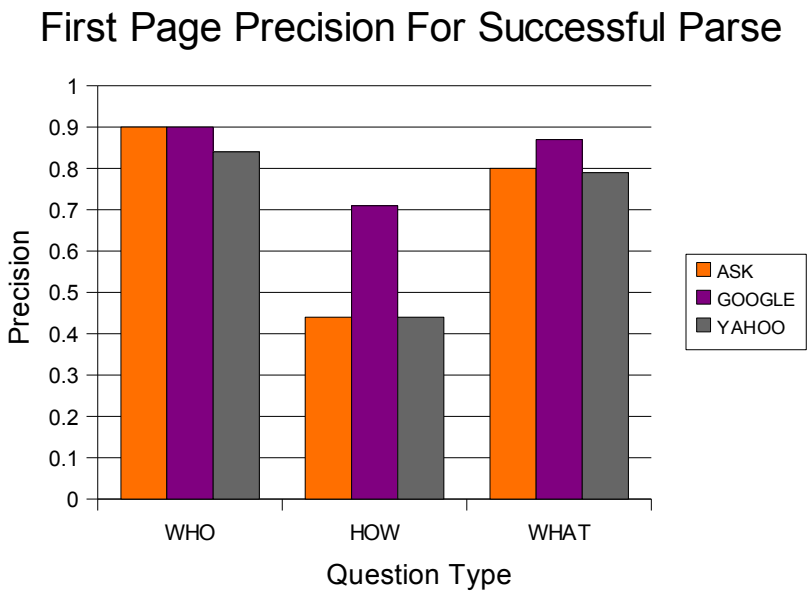
YAHOO.COM

Question Type	First Page Total Success	First Page Total Fail	First page Precision	YAHOO TOP 1 Success	YAHOO TOP 1 Fail	YAHOO TOP 1 Precision	QUANDA Precision
WHO	82	15	84.00%	39	58	40.00%	95.00%
HOW	19	24	44.00%	2	41	5.00%	40.00%
WHAT	237	62	79.00%	103	196	34.00%	90.00%
Total	339	101	77.00%	144	296	33.00%	87.00%

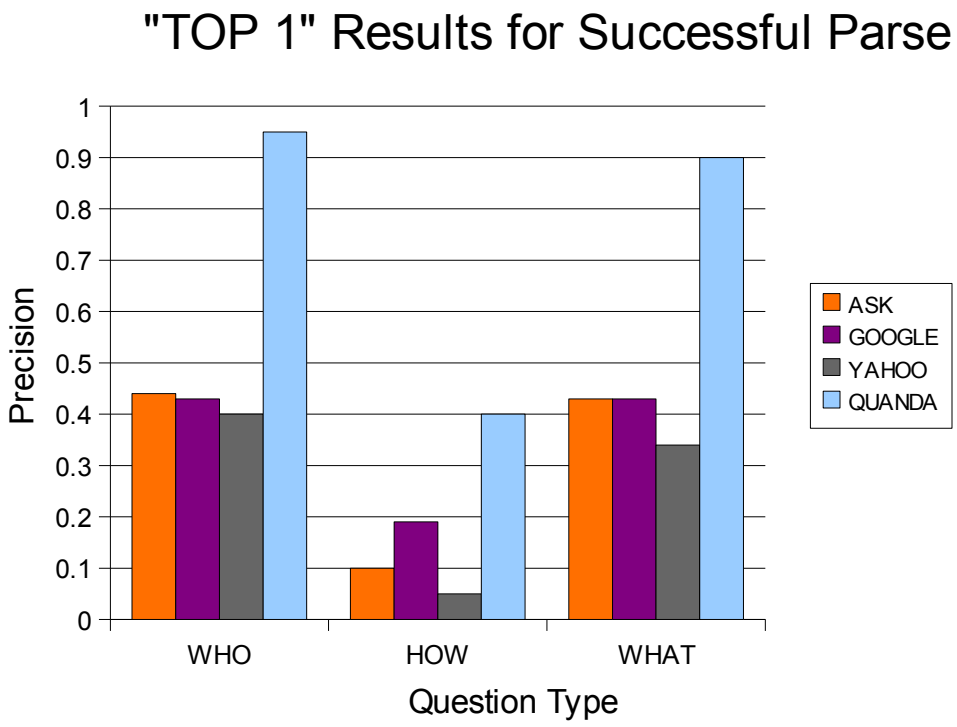
The results above show a number of interesting things. The first is that ASK.COM is quite good at finding an appropriate answer as shown in the "Precision" column. This is assuming that we use the first page of returned results by ASK.COM. For the purposes of this test, we are most interested in how ASK.COM is at returning a valid answer as the "top" answer. This is a fair comparison because QUANDA is judged this way and must be if we wish to preserve accuracy.

When analyzed for return of the "top" answer, ASK.COM's precision drops significantly (on average about 50%). When compared against QUANDA, ASK.COM is only about 39% accurate. This must be qualified in that ASK.COM provided better coverage of answering questions. These results will be discussed in more detail below. QUANDA is providing value from a precision standpoint.

In comparing ASK.COM to GOOGLE and YAHOO, there were some differences in how GOOGLE and YAHOO fared with the "successfully parsed" questions compared to the questions that failed to be parsed by QUANDA. In this case, ASK.COM seemed to do better than YAHOO in being able to provide a viable answer in the first page of results (10 results in a page). GOOGLE still seemed to provide better ability than ASK.COM. When considering the "TOP 1" answer, here ASK.COM was about equal to GOOGLE and significantly better than YAHOO.



While the above graph shows the precision for the first page, below is the graph comparing the 3 search engines and QUANDA where only the "TOP 1" answer (best answer) is considered:



The above graph shows that when we consider only the top answer, QUANDA precision fares very well against the search engines. Here ASK.COM and GOOGLE seem very closely matched in their precision with ASK.COM producing better results for "WHO" question types and GOOGLE doing besting ASK.COM on "HOW" question types.

5.3.3 Summarized Search Engine Results For All Questions

Below all of the questions (those that failed to be parsed by QUANDA and those that were parsed by QUANDA) are considered together. This is less relevant to a comparison with QUANDA but it is interesting to see how well the search engines performed relative to each other. The following information provides insight into how well ASK, GOOGLE and YAHOO performed on the TREC question set relative to each other.

ASK.COM

Question Type	First Page Total Success	First Page Total Fail	First Page Precision	ASK TOP 1 Success	ASK TOP 1 Fail	ASK TOP 1 Precision	QUANDA Precision*
WHO	115	19	86.00%	55	79	41.00%	95.00%
HOW	23	35	40.00%	6	52	10.00%	40.00%
WHAT	282	88	76.00%	148	222	40.00%	90.00%
Total	419	142	75.00%	209	353	37.00%	87.00%

**NOTE: QUANDA Precision is provided here for comparison but is only statistically relevant to questions that were successfully parsed.*

GOOGLE.COM

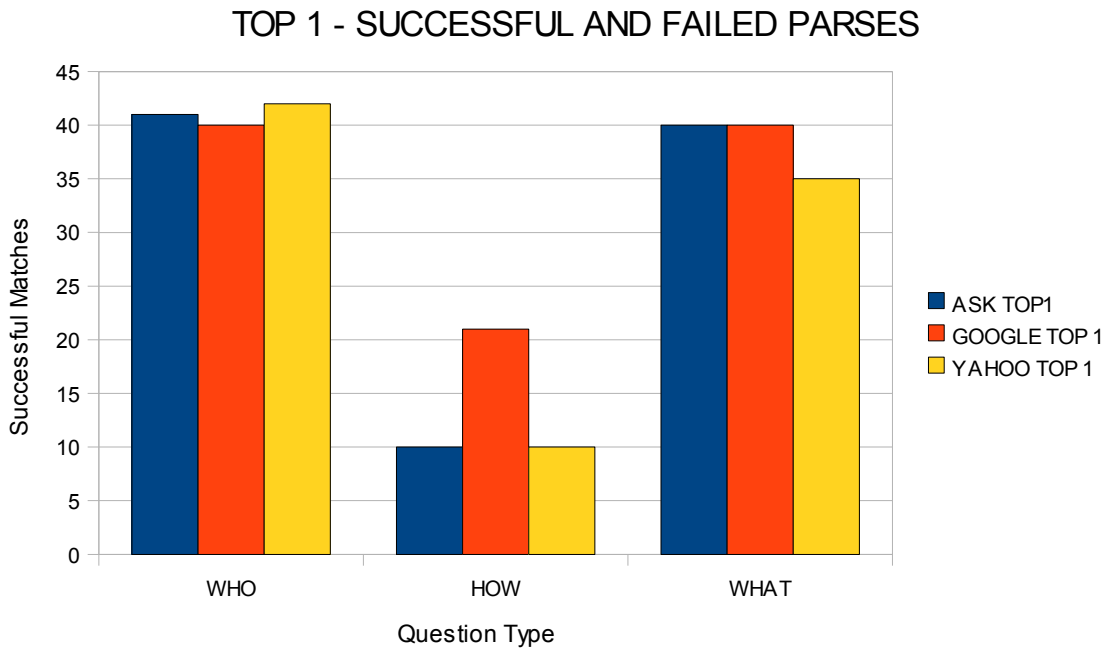
Question Type	First Page Total Success	First Page Total Fail	First Page Precision	GOOGLE TOP 1 Success	GOOGLE TOP 1 Fail	GOOGLE TOP 1 Precision	QUANDA Precision*
WHO	116	12	91.00%	53	81	40.00%	95.00%
HOW	42	16	72.00%	12	46	21.00%	40.00%
WHAT	323	47	87.00%	149	221	40.00%	90.00%
Total	487	75	87.00%	214	348	38.00%	87.00%

**NOTE: QUANDA Precision is provided here for comparison but is only statistically relevant to questions that were successfully parsed.*

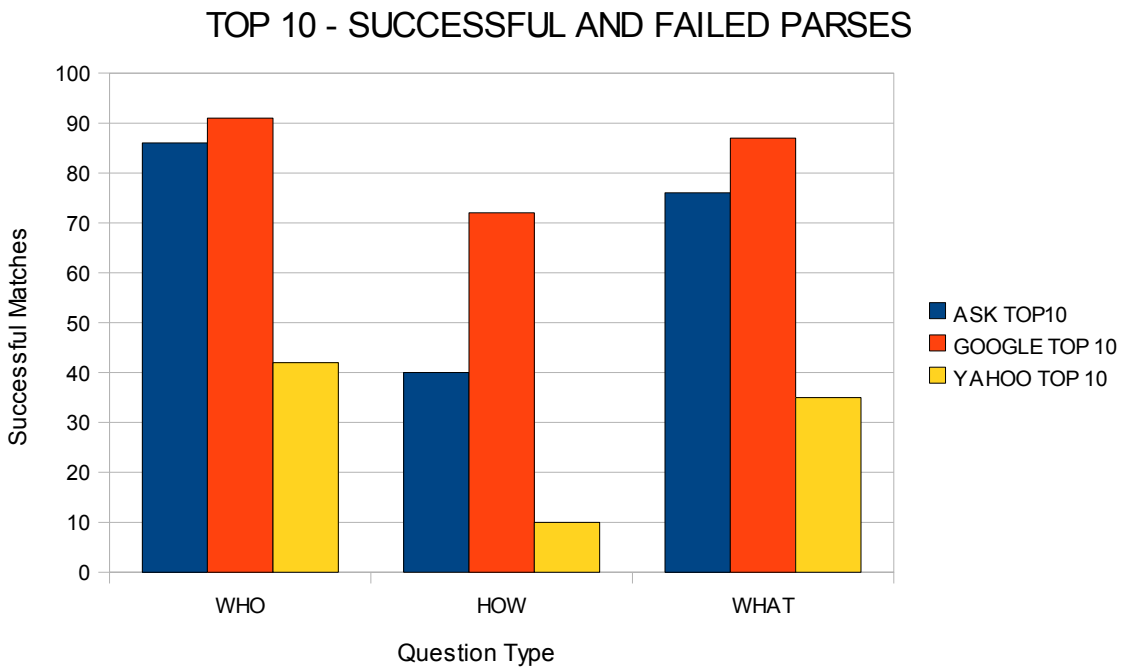
YAHOO.COM

Question Type	First Page Total Success	First Page Total Fail	First page Precision	YAHOO TOP 1 Success	YAHOO TOP 1 Fail	YAHOO TOP 1 Precision	QUANDA Precision*
WHO	111	23	83.00%	56	78	42.00%	95.00%
HOW	31	29	52.00%	6	54	10.00%	40.00%
WHAT	312	93	77.00%	141	264	35.00%	90.00%
Total	454	145	76.00%	203	397	34.00%	87.00%

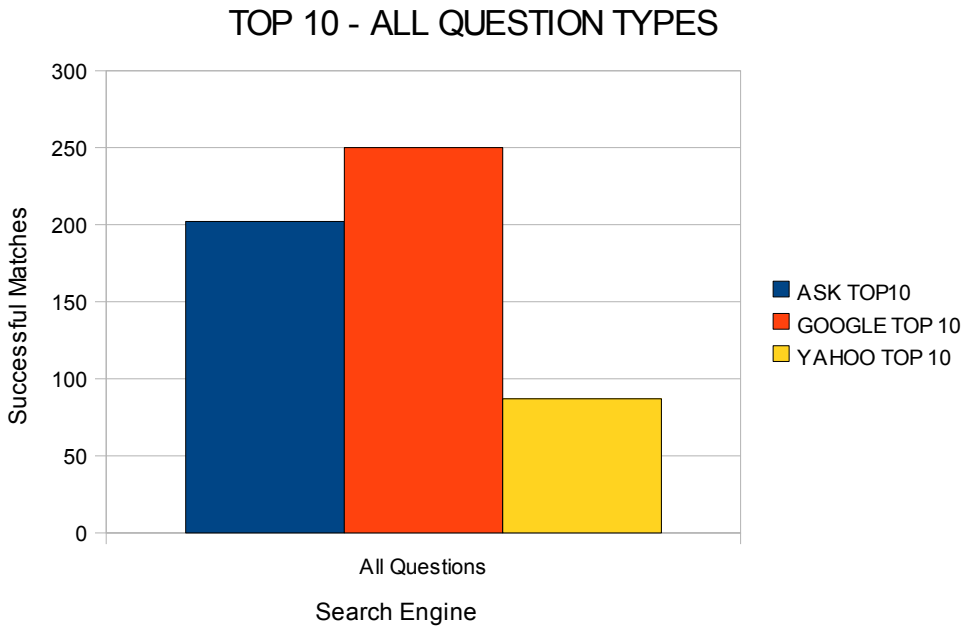
**NOTE: QUANDA Precision is provided here for comparison but is only statistically relevant to questions that were successfully parsed.*



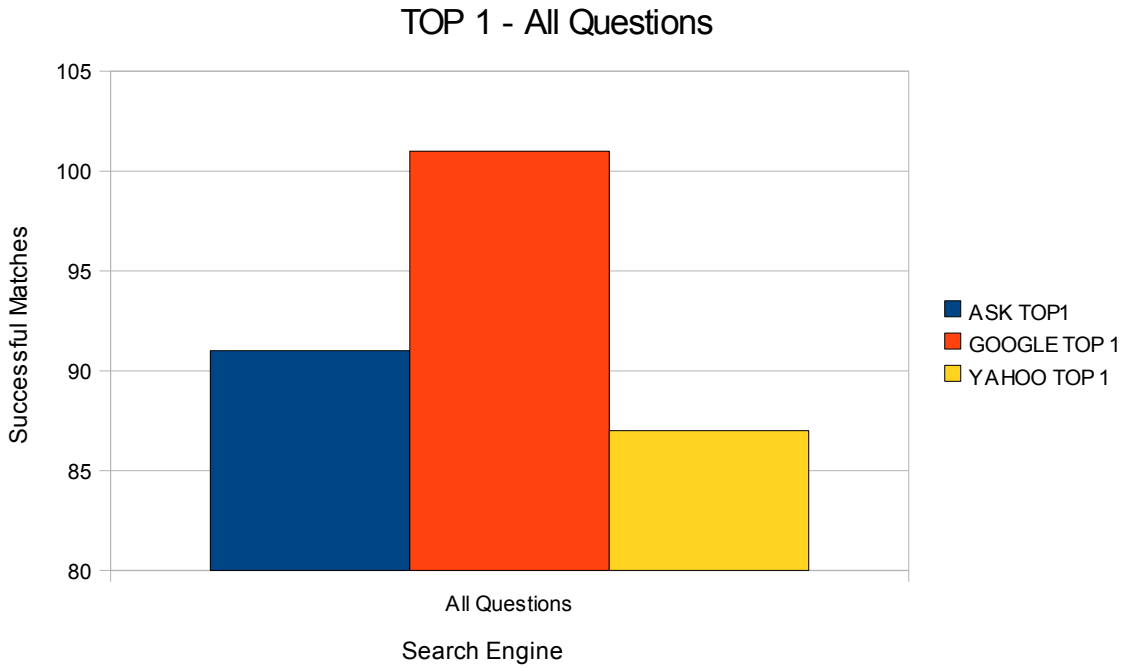
The above graph shows that all three engines are close in performance for WHO question types. There was a clear advantage noted for GOOGLE being able to answer the HOW question types in this test suite. GOOGLE and ASK performed similarly for the WHAT question types.



It was apparent that when a full page of results (TOP 10) was considered, GOOGLE and ASK performed better on the TREC question suite than YAHOO.



The above graph considers all question types a single group and again, GOOGLE appears to have an advantage overall.



As shown above, when precision counts, GOOGLE again performed the best of the three engines. The interesting thing to note is that the difference between ASK and YAHOO was not as pronounced as the prior graph that showed the TOP 10 results. The conclusions thus are:

1) GOOGLE was the overall best search engine for precision.

2) GOOGLE was the best search engine for returning an appropriate answer in the first page (top 10).

3) YAHOO seemed to have difficulty in returning the top answer and the top 10 answers

In summary, GOOGLE performed the best on the TREC question suite.

5.3.4 ASK.COM Results Analysis

Search engine results were achieved by typing TREC questions into the ASK.COM, GOOGLE

and YAHOO question window and manually retrieving results. A number of assumptions had to made about the results. These assumptions are listed below:

1) ***Only non-sponsored results were considered***. ASK.COM, GOOGLE and YAHOO, being commercial websites, sometimes present results from sponsors as the first set of returned answers. In some cases these were relevant but more often, the answers would skew the results negatively. As a result, answers are only reported for the non-sponsored portions of the results.
2) ***Match on any relevant term.*** To be fair to ASK.COM, GOOGLE and YAHOO results, a match attempt would be deemed a success if the search engine result simply contained the answer terms regardless of whether the context was an appropriate answer.
3) ***Main Page text match only.*** All three search engines return text that may be part of a broader document. Only the text that was returned on the main page was considered as an answer. This is a fair comparison as it is what QUANDA returns and text outside of this match window is not used in grading QUANDA either.
4) ***First Results Page only.*** The search engines may return many thousands of results. Only the first page of results was scanned for a potential answer. If the first page of results does not contain a valid answer, then it is assumed that the search engine has failed to find an answer. This is appropriate because for the purpose of comparison with QUANDA we are interested in precision and not coverage.

Applying TREC questions to the search engines at first glance appears to be an invalid comparison with past TREC results because the datasets are not the same (e.g., Tipster Corpus and web search space). While it would be unfair to directly compare these, our goal here is to determine how well ASK.COM, GOOGLE and YAHOO perform against QUANDA from a precision perspective. While the datasets are different, the results do provide valuable information about the precision of the two approaches.

Specifically, the answers we are seeking are:

1) How precise are the answers of ASK.COM, GOOGLE and YAHOO compared to QUANDA?
2) Are there situations where QUANDA may complement the results that a search engine may benefit?

5.3.5 Detailed ASK.COM Results Analysis

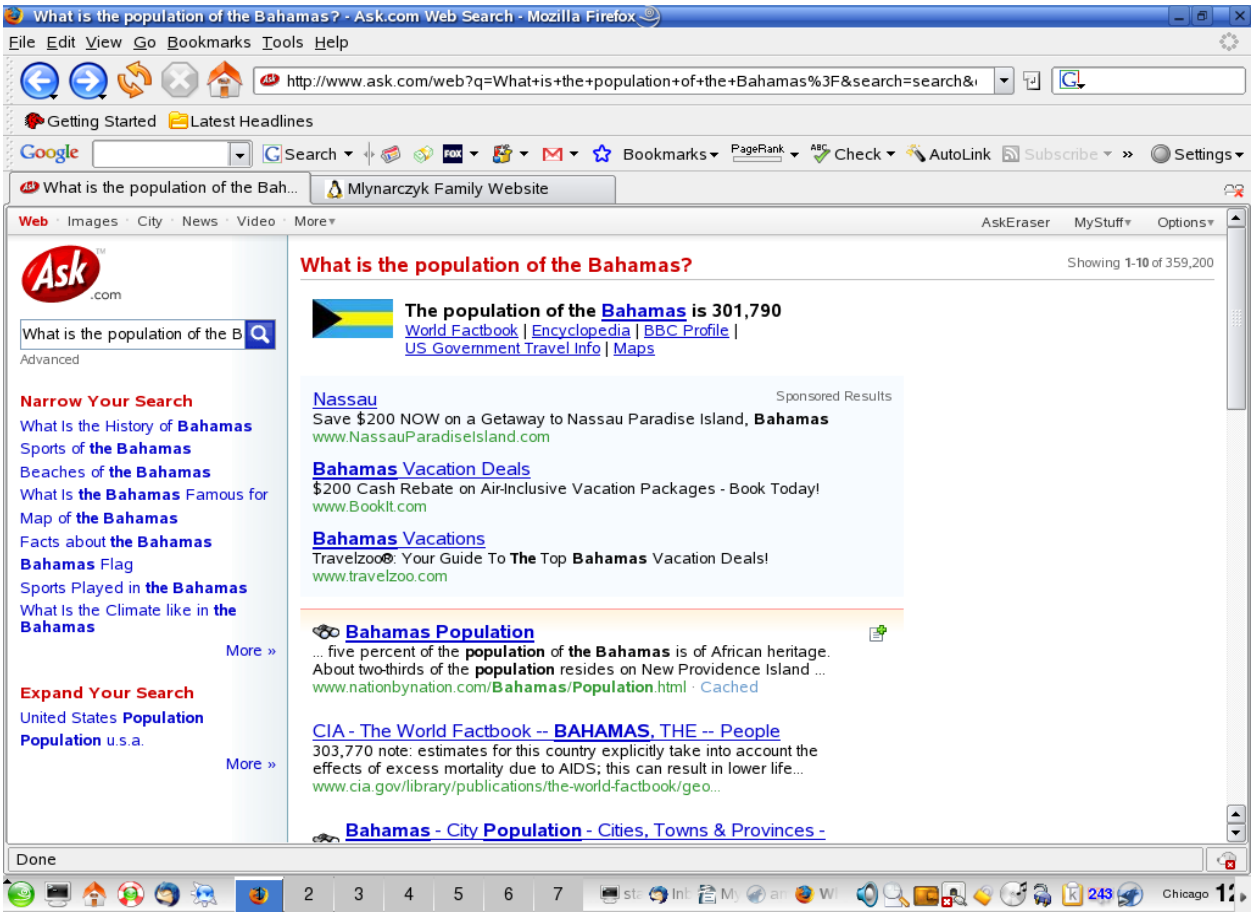
A detailed analysis of ASK.COM is being done because ASK.COM is closer to QUANDA's. Even though all of the search engines produced results that were close in terms of precision, ASK.COM seemed to possess built-in ability to distinguish between question types and this is discussed below.

5.3.6 ASK.COM Predefined Questions

An interesting observation of ASK.COM results is that it appears that a core set of questions have been anticipated by ASK.COM and a "comprehensive" and rich answer is presented. Good examples of this are the questions: "What is the population of the Bahamas?" and "What is the nickname of the state of Pennsylvania?". These "canned" Question/Answer pairs are displayed before the standard formatted results. The implication here is that the standard statistical methods employed by the search engine are abandoned in favor of "quick hits".

The ASK.COM search engine in most cases (based on the TREC questions that were tried) always provides an answer, however, the answer may not be the first/top answer. Additionally, the user may need to access the underlying document one layer below or lower to find the desired answer. With respect to the "quick hits", ASK.COM appears to be parsing the user question and attempting a quick question/question match and then presenting a "canned"/"polished" answer.

Please refer to the below answer for "What is the population of the Bahamas?". The top answer includes a direct answer to the question - "The population of the Bahamas is 301,790". Additionally, "Bahamas" is linked to additional information about Bahamas statistics while the Bahamas flag is an icon that provides a link to a CIA government website. This is a very obviously an anticipated question and carefully constructed answer. In one sense, this is validation that there is merit in anticipating user questions as is done in the QUANDA approach used by this research. Unlike ASK.COM's approach, QUANDA provides a much more general approach to question match providing an answer. It does this by anticipating answer forms but is not bound by topic-specific/canned results.



5.3.7 ASK.COM HOW Question Limitation

ASK.COM seemed not to do as well with questions such as “How hot does it get in a volcano?” and “How many miles in a 10K run?”. These questions did yield documents about the topics but were not appropriately answered. There were many other such questions where ASK.COM seemed to get to the topic but did not address the specific question. This might be explained by the approach used by ASK.COM which appears heavily focused on keyword frequency. This behavior is similar to that of FaqFinder+'s word frequency based analytics in that it is useful for filtering a set of documents to ascertain the probable set. As with FaqFinder+, word frequency does not always correlate well with a good answer to a specific question. Still, the approach is useful for filtering a large document repository.

5.3.8 ASK.COM Provides Better Answer Than TREC

In one case, ASK.COM provided the same correct answer that QUANDA did where the TREC

answer was clearly inappropriate. The question was “What is “Nine Inch Nails”? The TREC answer was incorrect in that it was related to common-nail hardware instead of the musical group. In general, ASK.COM did a very credible job of returning appropriate answers. It should be pointed out that QUANDA also produced a better answer than what TREC called for.

5.3.9 ASK.COM Question Word Count

An interesting observation with ASK.COM is that as the size of the question increases, so does the likelihood that it will not return an appropriate answer as the top answer. For example, the question: Who portrayed "Rosanne Rosanna-Dana" on the television show "Saturday Night Live". In this case, ASK.COM did not find an answer, however, when a new search was initiated using just “ Rosanne Rosanna-Dana Saturday Night Live”, a viable result was returned within the TOP 10 answers.

The reason for this may be that other words within the sentence were significant enough to preempt correct matches in this case. In short, more words within a sentence do not necessarily increase the likelihood of an appropriate match. This is not surprising as construction of appropriate keywords is often viewed as the key to successful web searches. Further evidence that this might be the case is that when “Rosanne” was removed from the search string (Rosanna-Dana Saturday Night Live), the results improved again so that an appropriate answer was found in the TOP 2 results. This further illustrates that while ASK.COM may have some ability to respond to “questions” that have been determined to be common/frequent, it does fall back on traditional word frequency approaches.

A question such as “Who reports the weather on the "Good Morning America" television show?” illustrates the challenge that a search engine based on word frequency has. In this question, ASK.COM does return a variety of documents about the television show but the tool does not specifically search for the person who “reports” the weather. In some cases, a match of this type might by chance contain the reporter's name but in this case it does not. QUANDA on the other hand might more specifically search for an occurrence of:

????? reports the weather on “Good Morning America”
the weather is reported by ????? on “Good Morning America”
“Good Morning America's” weather reporter is ??????
weather reporter for “Good Morning America's” is ??????

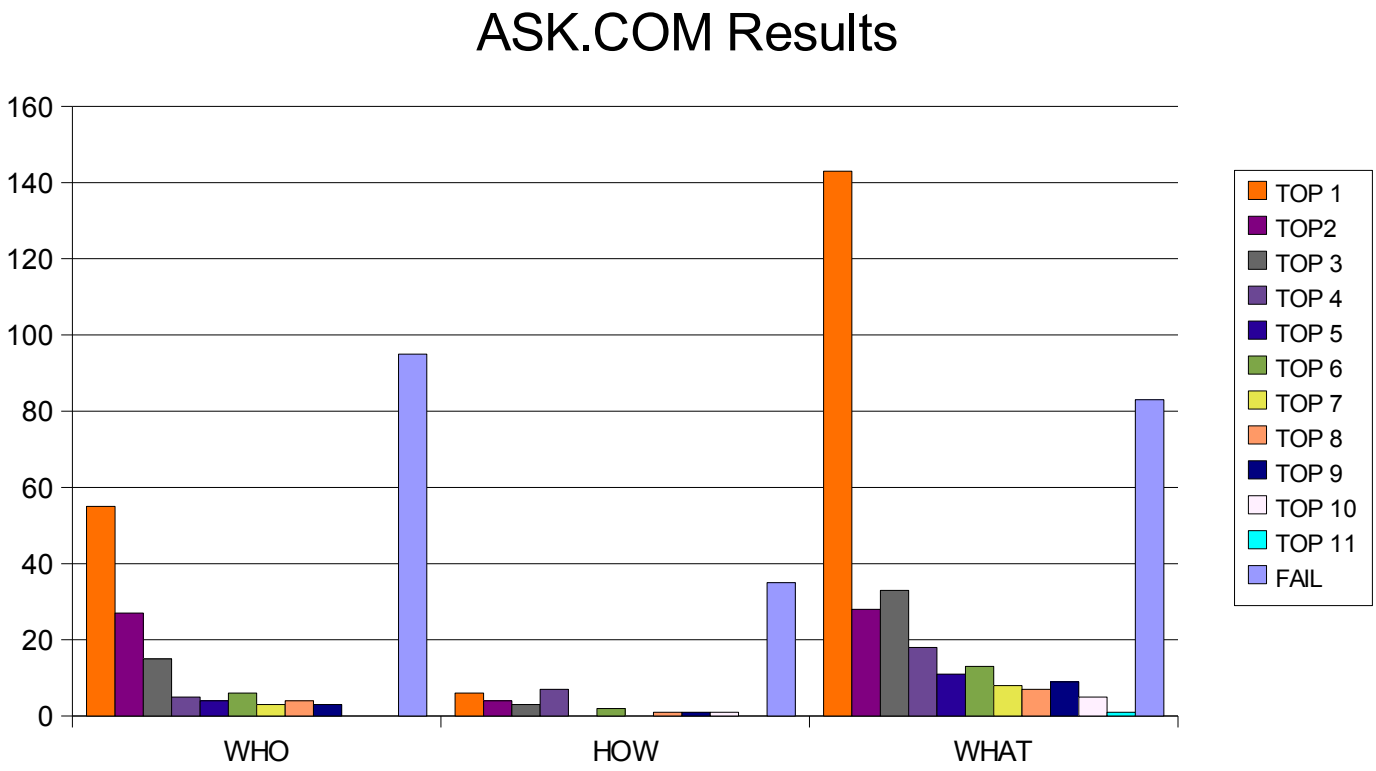
There are trade-offs between accuracy and coverage. Perhaps, if a frequency based approach returned the set of documents ASK.COM did, then a much simpler QUANDA rule might be successful in matching “????? reports the weather”.

5.3.10 ASK.COM Phrase Support

ASK.COM does possess the ability to recognize phrases. For example, “Witch Hazel” and “limited partnership” are instances where the terms could stand on their own but ASK.COM did associate them as phrases. This does show that ASK.COM is more than just a word frequency based approach.

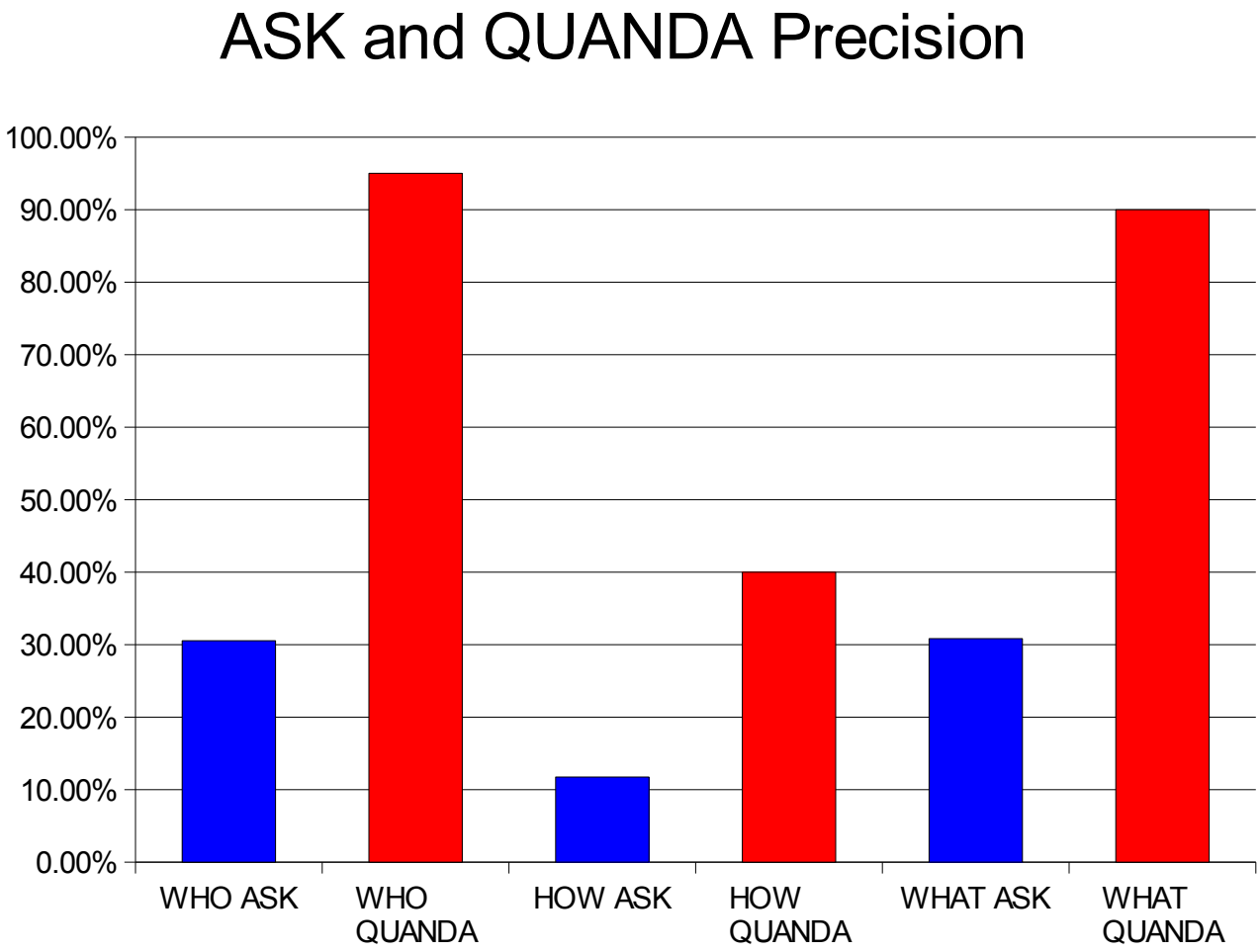
5.3.11 Comparing ASK.COM and QUANDA

So how are we to compare ASK.COM and QUANDA? It might be informative to first look at the distribution of results within the ASK.COM. The graphs below show the answer distribution for the 3 different question types (WHO/HOW/WHAT).



For all three question types, iet seems that ASK.COM's results will produce the “correct” answer more often if the correct answer will be found. As mentioned in the observations above, this seems logical in that if the proper “keywords” are present within a frequency-based text search approach, then the most likely candidate answer will be the presented first. Additionally, ASK.COM's canned responses to predefined question patterns also aids in this respect.

Also a point to note is that the different question types have different match success/fail characteristics. "WHAT" question types appear to have more successes than both HOW and WHO types. This can be seen in the above graph in that the FAIL (no answer on the first page) is much higher in WHO/HOW relative to the "TOP 1" results for the "WHAT" type. This is more easily seen in the below graph.



In looking at the precision that was achieved both by QUANDA on the TREC questions and by ASK.COM on it's WEB based knowledge store, it is clear that there are significant and consistent differences. The blue shaded areas signify ASK.COM results and the red correspond to QUANDA using TREC data. There are a few observations that are worth pointing out.

First, both approaches seem to find "HOW" questions as being more difficult to resolve and "WHAT" and "WHO" questions as being the easier. Intuitively this does make sense as the answer to a HOW question type may have less in common (term-wise) with the question itself

than a WHO type. This explanation is less satisfactory in comparing WHO and WHAT question types. WHAT question types, like HOW types, have broad application within the English language.

The reason WHAT question types appear to have fared as well as they did is that the types of WHAT questions asked were generally simple “What is a...” and “What newspaper servers Salt Lake City” and the subject/verb/object of the sentence had significance in the potential answer. This was certainly helpful to QUANDA in that potential answers refer back to the Subject/Verb/ Object in the question. It was also helpful to the word frequency-based match logic in ASK.COM.

While both approaches correlated in terms of question type results, the precision is where they differed. The goal with QUANDA was to improve the precision of FaqFinder+ and thus the mechanism was intended to be highly precise. A search engine such as ASK.COM has multiple requirements. While precision is desirable, so is the ability to return answers even if the answers aren't exactly what the user is looking for. The answer sets returned by a search engine are also much larger and it is expected that a user may wish to page through many results to decide which might be appropriate.

It is not fair nor productive to compare the two approaches as competing mechanisms. Both have a place and are useful as approaches to question answering. Text analytics research has sometimes diverged on which approach is better; whether to use word frequency, (Brill, 2001) for example, and achieve good results with minimal effort or whether to seek more precise results using more challenging approaches such as (Mihalcea, 2001).

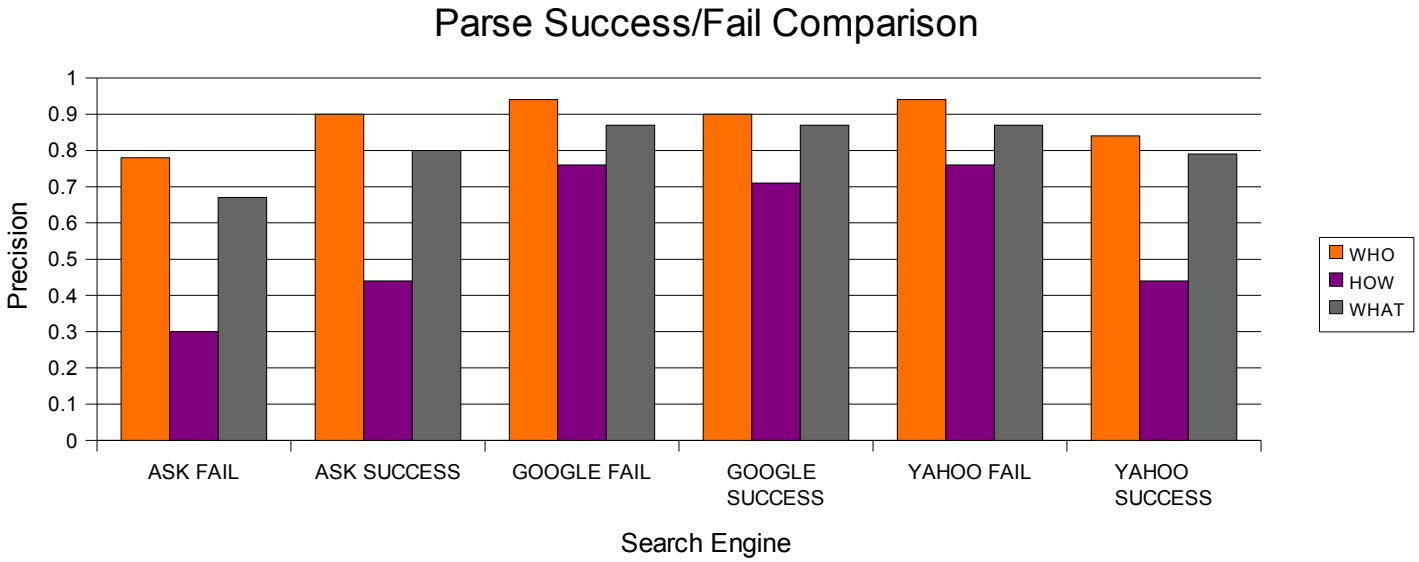
This argument is perhaps best resolved by the implementation challenges that appear in trying to build a system that must support a very large knowledge repository and to also return precise results. While FaqFinder+ is an example of this competing requirement, so is ASK.COM. While ASK.COM utilizes a word frequency approach, it also demonstrated some higher level analytics in the type of predefined answer sets it returned for often asked questions (such as questions on population of cities/states/countries).

In summary, QUANDA has demonstrated itself to be of value in increasing the precision of question answering. This research effort has also demonstrated that a word frequency based system such as ASK.COM is quite adept at returning relevant answers. In practical application, use of both approaches can provide benefit.

5.3.12 GOOGLE and YAHOO Results

Google results were surprisingly strong compared to ASK.COM. As can be seen from the graphs above, Both GOOGLE and YAHOO produced good results when answers provided within the first page were considered. There were differences between GOOGLE's and YAHOO's precision when the two classes of questions were applied; those that failed to be parsed by QUANDA and those that were parsed successfully. It is not completely clear why these differences were

present.

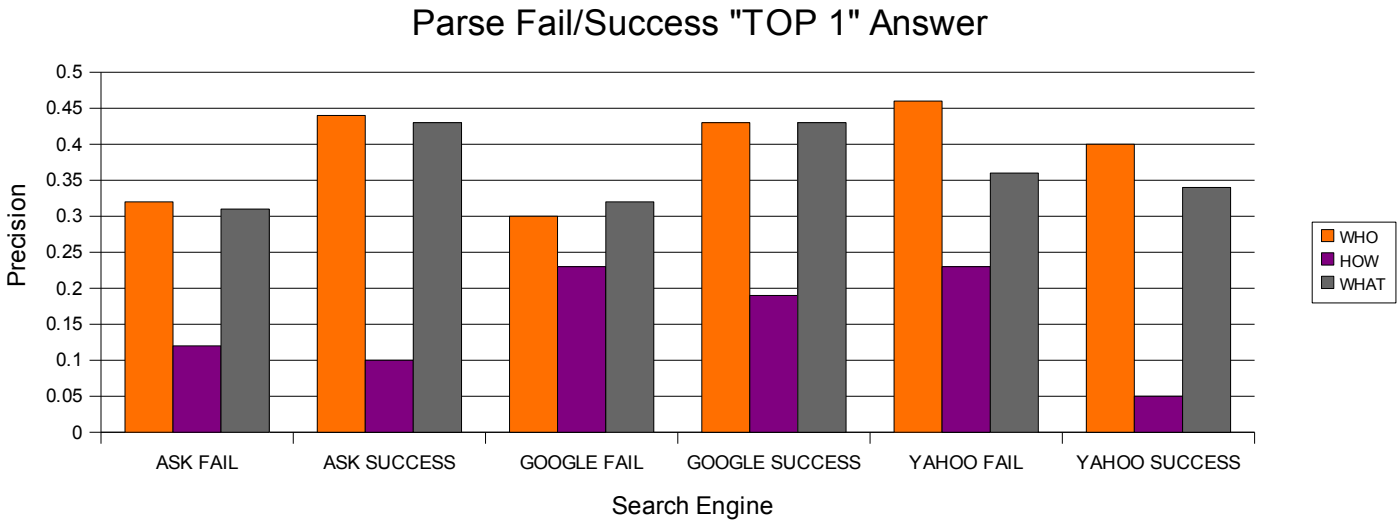


In the above graph, YAHOO performs significantly better on the questions that failed to be parsed correctly by QUANDA. GOOGLE also exhibits this behavior but to a lesser extent. ASK.COM on the other did better with the questions that were successfully parsed by Question/ Answer. A definitive statement about why this is the case cannot be made without a deeper understanding of each search engine's implementation.

The one conclusion that would be attractive to make is that ASK.COM's parsing of some question may be influencing its performance. Unfortunately, investigation of answer results does not show a consistent pattern that would support this.

An interesting observation with respect to YAHOO's results is that YAHOO is being heavily influenced by an unexpected source; the TREC government website itself. Questions are producing matches (often as the top answer) with the identical question that is stored on the TREC website. Instead of YAHOO producing an potential answer match, it is matching on the question. Both GOOGLE and ASK.COM appear to suppress the relevance of the TREC website. By not filtering the TREC question site, YAHOO's results are being negatively skewed. This is unfortunate since YAHOO did very well with producing the "TOP 1" answer for the questions that failed to be parsed correctly by QUANDA as is demonstrated below:

Some measure of consistency was evident in all the search engines in that all three produced better results with "WHO" question types for all questions (failed and succeeded parse). "HOW" questions showed the least success and this too was uniform across both questions that were successfully parsed and those that were not.



5.3.13 Search Engine and QUANDA Performance

There was no noticeable difference in performance of any of the three search engines. All three returned results quickly and any latencies noticed were more likely due to network congestion.

The performance of QUANDA was good when consideration is given to the hardware platform on which it was run and the complexity of the tasks being performed. To complete 693 questions and limiting the search to 5 documents took 60 hours. Given that only 402 questions were successfully parsed, the per question performance of the system is roughly:

(60 hours x 60)/402 or 8.9 minutes/question

Given that there are 5 documents, this then yields a figure of 1.8 minutes per question/document.

A figure of 1.8 minutes/document is not very competitive with the three mainstream search engines and a direct comparison is neither appropriate nor needed. QUANDA would not likely ever replace word frequency based approaches to text search. Instead, QUANDA might augment word frequency based approaches to provide increased accuracy.

There is still the need to consider performance as 1.8 minutes per document would not be usable/ useful. As with YAHOO, GOOGLE and ASK.COM, QUANDA performance can be improved through the implementation of parallelism. GOOGLE for example has tens of thousands of servers that both scan the web and provide answers. QUANDA testing was performed on a

single server but can be partitioned to function within a multi-processor environment. The functionality is already partitioned into separate parse and search steps.

Parsing of questions and text can be an operation that is distributable a single sentence level. With a farm of servers, the worst case response time would be the time to process an individual sentence (on the order of a second or two of elapsed CPU time). There are also opportunities to pre-parse text that would further enhance performance.

In summary, although performance of the system was exceedingly slow in comparison to the three search engines, this effort is a research exercise and was not intended to provide a production quality implementation. Should there be a need to enhance the performance of QUANDA, this can be accomplished and performance/response results, on par with GOOGLE/YAHOO/ASK.COM, are achievable. Given the increased accuracy that QUANDA can provide, integration of this type of functionality with existing search capability may be of value.

5.4 Recommendations for Improvement of TREC Results

Analysis of the results has revealed a number of modifications that would benefit the approach. Some of changes are obvious in their ability to positively impact the results and others may help. For others, additional testing would be needed to ensure that newly proposed functionality would provide improvement. The following is a list of recommendations:

5.4.1 Additional Rules

While there are already a large number of rules in the rules repository, the results indicate that some basic rules were missed. Existing rules were created from a data set of questions that was collected from the work of (Hovy, 2001). While already a large set comprised of thousands of questions, many of these have very specific and unique syntax. Focus on much more basic questions would provide benefit. For Example:

What year was the first automobile manufactured?

What was the name of the first Russian astronaut to do a spacewalk?

What time of year is air travel the heaviest

What time of year is air travel at a peak

What time of year has the most air travel

All of the above rules are basic rules but were missing from the repository. Had they been supported, these could have contributed positively to the results.

5.4.2Additional answer forms

While some of the questions were parsed correctly and had associated answer forms, the answer forms did not result in a match. In many of the cases analyzed, it is not feasible to resolve this with additional answer forms. This is because the text being searched does not specifically yield an answer or only peripherally yields clues about the topic. In these situations, reading an entire paragraph or article may be required along with human level reasoning and world knowledge. Still, there were situations where a richer set of answer forms would have helped. For this reason, additional effort in adding new answer forms to existing questions would help. It must also be mentioned that the answer forms should be "specific" to ensure that precision is maintained (to guard against to-general answer forms producing false matches).

5.4.3 Additional Phrase Recognition Support

QUANDA capability relies on a phrase recognition/resolution feature that identifies and transforms phrases. This is a step that is applied to both the questions and potential answers using the same mechanism, which then maximizes the chances for a match. Additional phrase forms need to be populated. For the purposes of this research, a very large phrase ontology was used, however, only a small subset of this is actually populated. As evidenced by the results analysis, additional phrase support would help and should be added.

Phrase recognition is not a replacement for paraphrase support but does provide some of the benefits. For example, in the following sentence:

Name a major city in Illinois?

the phrase matching rules would convert this to be:

What is a major city in Illinois?

While not paraphrasing in the strict sense, it does provide a way to minimize the number of QUANDA rules. In the current results, the match potential would have been increased had the phrase ontology supported:

Alyssa Milano

Nobel Peace Prize

Vietnam War

Darth Vader

Martin Luther King Jr.

etc...

While the argument can be made that named entities might justify the need for a more advanced algorithm than the simple string replacement that is used, this also requires careful consideration.

While terms such as Alyssa Milano and Martin Luther King Jr. might be easily accommodated and capitalized words as well, there are a variety of phrases that are lower case and are not easily recognized algorithmically. (e.g., “a good deal of”, “in the presence of”, “a great deal”, “about turn”, “in place of”, “operating manual”, etc...).

Given that the goal is high precision, guessing incorrectly is less desirable than not providing an answer and providing a good match/replacement is welcome regardless of the approach used.

The repository of phrases needs to be more fully populated and this would enhance the results.

5.4.4 Synonym Support

Synonym support is a capability that would need to be tested for benefit. On the surface, it is reasonable to assume that it would help if synonyms were tried by the pattern matchers. For example, the question:

“How do I realize a benefit from using a tax program?”

We might suggest that alternate forms of this question might be:

“How do I get a benefit from from using a tax program?”

While this is a simplistic change “realize=get”, there are implications in the matching functions when the target text is considered. The answer form will also have to support this similar equality and this is where problems may arise. This is due to sometimes unforeseen match anomalies, especially when using semantic distance; “get” is semantically close to “receive”, “buy”, “purchase”. Clearly, it would lead to an inappropriate match if “realize” were equated to “buy” or “purchase”.

Given some of these concerns, synonym support would be an interesting thing to test.

An interesting observation in the above example is that the word “realize” has additional meanings, one of which is more closely aligned with “understand”. Ideally, the above example would have been translated by the phrase replacement functions to yield something like:

“How do I benefit from using a tax program?”

thereby altogether eliminating the ambiguity of “realize a benefit from” via phrase replacement with “benefit from”.

5.4.5 Stemming Support

In some cases there is a need to match a question form to a potential answer form where the part of speech changes. For example, the main verb in a question might become the subject within an answer form. While this functionality is already supported within QUANDA matching logic, there is a missing component that would be required for a successful match.

If for example, we have the following TREC question:

Who manufactures synthroid?

The main verb is "manufactures". We might however wish to match the following potential answer form:

Merck and Ely Lily are manufacturers of synthroid.

The issue here is that we would want to match manufactures (VERB) with manufacturers (NOUN). While the association can be represented by "$NN/$NN1..$VERB" and the match logic will permit this association, the match will fail because the words are not currently stemmed to their base for the word comparison. This is a simple example where stemming would help, however, some words change more dramatically between POS changes. For example:

create/creation

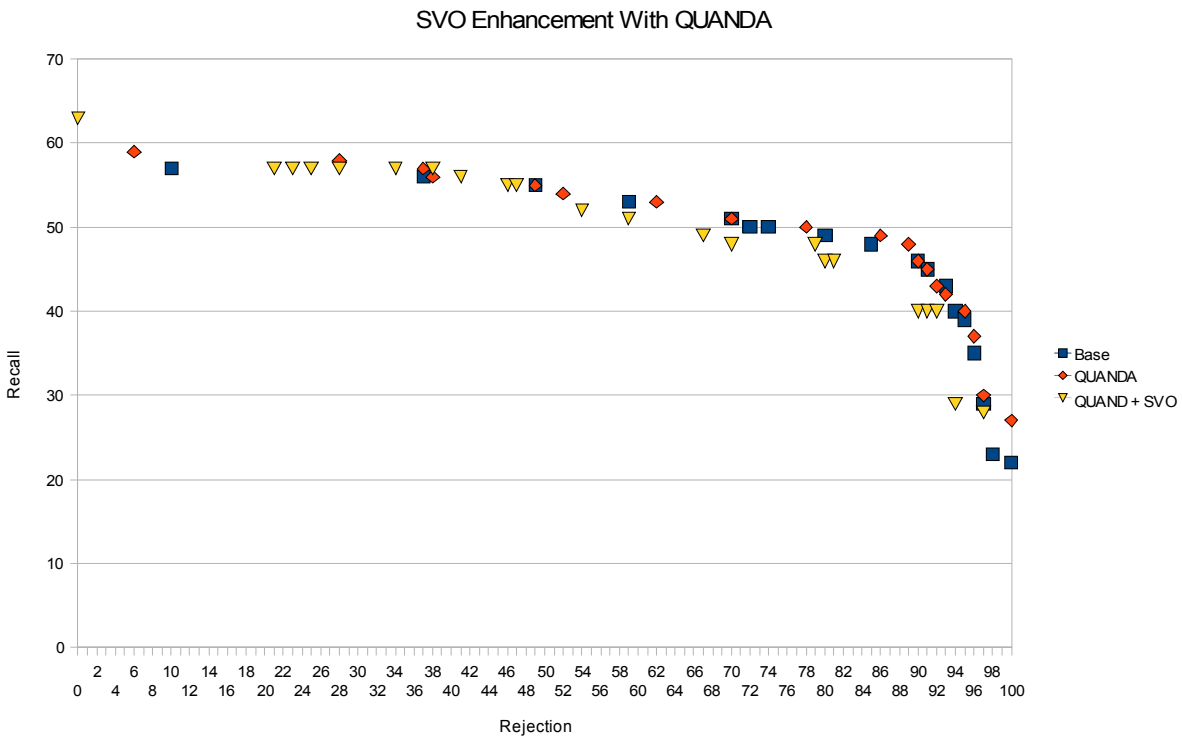
accept/acceptance

tax/taxation

so that to be truly effective, a much more complete support of stemming would be required.

5.5 SVO FaqFinder Integration

The QUANDA ability to more accurately identify subject/verb/object was integrated back into FAQFinder's original question/question SVO capability. The hope was that this would improve question/question match. The effort did not produce a positive result as can be seen in the below graph (QUANDA + SVO is the newly enabled functionality): The two other results are without QUANDA and with QUANDA but no enhanced SVO.



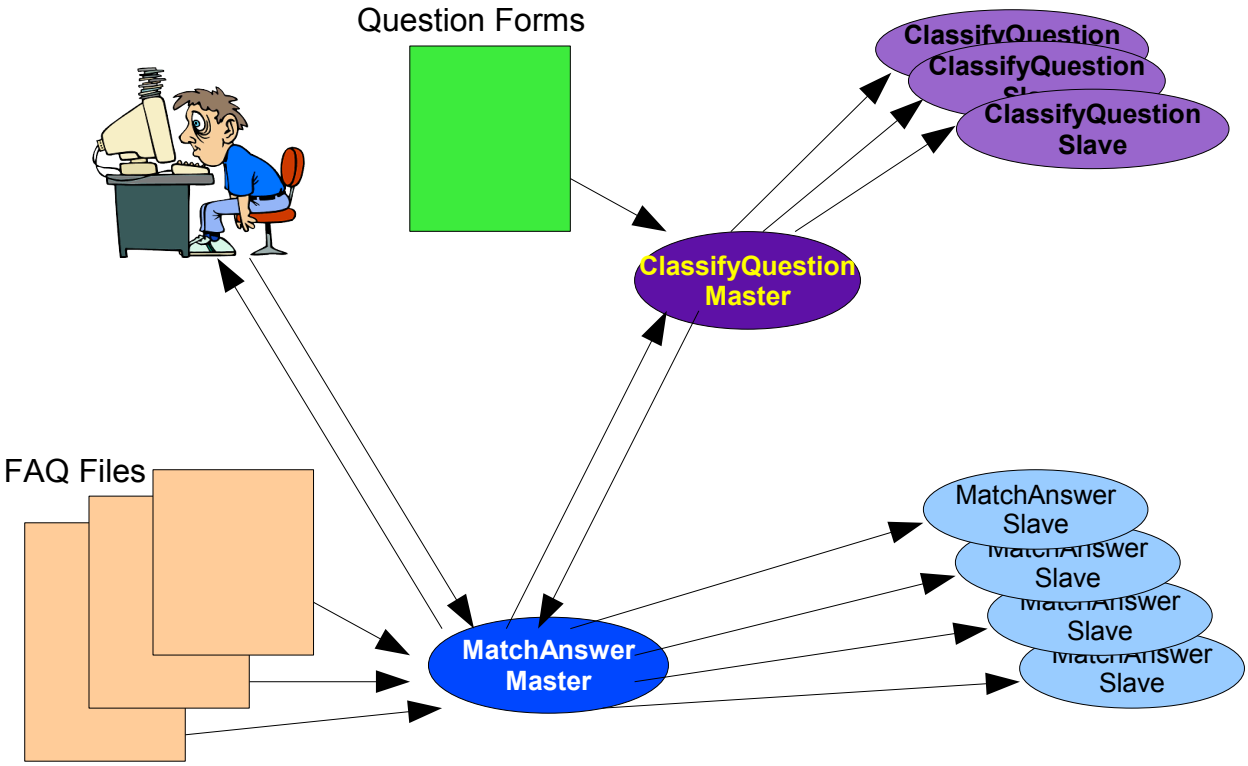
Analysis of the results has revealed that the more accurate SVO tagging **detracts** from a positive result. To understand why this is the case, one must look at how and what question types. The sentence subject that is often returned is "how" or "what". This contrasts with the currently less accurate mechanism that returns usually other sentence nouns. There is a higher probability for the less accurate SVO to provide a word match between two questions even though the correct subject has not been identified. In essence, the current SVO functionality is more in line with word frequency matching. In short, a small chance of matching a relevant word is better than no chance of matching QUANDA's return of HOW/WHAT, as the sentence subject is not helpful.

6 Recommendations For Future Improvement

During the research effort, there were a number of things learned and discovered for which time was not available to pursue. Among these are:

1) Taking advantage of WordNet synonyms and word distance within QUANDA. To support synonyms and word distance would require support for parallel processing within a multi-cpu environment in addition to the additional logic.

2) Parallel processing. This is a necessary step for the approach to be a viable future research direction. The computational effort is very large (7 days run time for the final test set) and support for multiple CPUs would greatly aid in reducing run times.



The above graphic shows how QUANDA can be extended for parallel processing. To accomplish the above would be not be a large effort as many of the existing components are already set up as client/server (such as the underlying dictionary and POS resolution functions)

3) Support for paraphrasing. This would potentially reduce the number of question/answer forms needed.

4) A tool to assist in the creation of question/answer forms. The current manual method is labor intensive and even a shell with command completion would greatly aid the rules development effort.

5) There are currently a number of utilities that scan the rule sets (similar to the lint command for C) for syntax errors. There is a need to increase the syntax checking suite's capabilities as the debug effort for new rule sets is not insignificant.

6) An attempt to integrate QUANDA into a WEB search engine. This might demonstrate the benefits of combining QUANDA with traditional WEB search.

7) Additional research into applying the existing methods to future TREC competitions. It

would be interesting to see how the approach would fare with some additional enhancements specifically targeted at TREC.

8) There is an opportunity to use the approach within business environment. Creation of natural language interfaces would benefit from the functionality demonstrated within question/answering. For example, a natural language interface to an SQL back-end database might be built where users might ask questions of the database using non-technical terms.

7 Conclusions

The research within this thesis has been at times complex, arduous, disappointing and finally, rewarding. A variety of techniques were tried with the hope that improvement could be made in FaqFinder+'s accuracy. In a majority of these attempts, little progress was made. With these attempts, knowledge was gained about the underlying issues with text analysis.

Statistical methods have been a significant focus of text analytics and for good reason; they produce good results. This has also been true for FaqFinder+ which makes use of statistical approaches. Attempts to improve FaqFinder+ accuracy beyond what has been achieved with statistical methods have met with some success but this success has been hard-won. For example, the efforts to utilize WordNet for word distance have helped. This help however has yielded modest results.

With every potential improvement comes the potential for degraded results. While it is beneficial for example to consider "word distance" in text analytics, it can also lead to problems. The problems stem from the larger pool of possibilities afforded by the techniques. While additional possibilities can be helpful (synonym consideration for example), it is also the case that we might be mistaken in the sense of a word being used.

The English language is rich with multiple meanings for words and selection of the wrong meaning increases the pool of potentially incorrect answers. Some of this word sense ambiguity can be mitigated by utilities such as the "Tagger". Unfortunately, a tagger is not 100% accurate. While it is true that the tagger can provide multiple potential meanings, one of which might be correct, incorrect tagging leads to invalid results.

The experience gathered during this research suggests that adding intelligence to the FaqFinder+ question matching process is not universally and automatically beneficial. If this is the case, then why focus on adding intelligence? Why not just utilize statistical methods? Following are conclusions that this research has realized:

1) Statistical methods provide good results quickly.
2) Statistical methods break down when the context of text indicates a meaning that is not evident from examination of individual terms.
3) Semantic analysis of text can help with disambiguation but is "hard".
4) Combining statistical and semantic methods has improved FaqFinder+'s accuracy.

There is clearly a benefit to the use of semantic methods and application of additional intelligence in text analytics. Attempts at applying these methods, as already mentioned, has yielded mixed results. Some of the reasons for the difficulty include:

1) Lack of ontologies to support the deeper analytics – especially world knowledge (WordNet helps but more is needed).
2) Logic that supports "reasoning" is difficult/complex to implement and is dependent on (1).
3) Pattern match based methods can be accurate but require a large investment in implementation and may not always provide adequate coverage. Narrowly defined rules can be accurate but may not cover variations while parameterization of terms may increase coverage but at the expense of accuracy.

As an example of one of the issues that was encountered, there was the initial attempt to identify the subject/verb/object (SVO) within a sentence and then to use this information to match against previously asked questions (refer to section 4.1.2). The approach was promising and often worked quite well but it did suffer from a dependence on logic that did not always perform the identification of subject/verb/object correctly. In situations where the logic failed, it was because the foundation to support proper SVO identification was absent. To a human that is able to consider the context of a question, this would pose little problem, however, the programmatic approach has no knowledge base upon which to make adequate judgments.

For example: two questions taken from the current question set:

What time of year is air travel the heaviest

What time of year is air travel at a peak

In the above sentences, identification of SVO is challenging. Ideally, we would know that "time of year" is a noun phrase, that "air travel" is a phrase identifying airline travel, that the sense of "heaviest" used here implies "busy" and not related to weight.

Similarly, "at a peak" would need to be recognized as referring to a "maximum" and not a mountain.

Much is taken for granted by human logic as many things humans reason about are resolved by base world knowledge. Programmatic approaches to simply identifying SVO as evidenced above, can be very difficult.

The approach taken by this research was to create a set of rules that tied together a user's question with a set of potential answers. Parameterized pattern match is the foundation. Accuracy is maintained by having the rules enforce sentence syntax along with referential compliance between question and potential answer. This approach seeks to strike a balance between ease of implementation (via parameterization of rule terms) and the specificity enforced by the question/ answer linkage.

The research approach did not seek to solve the world base knowledge problem. Instead, a set of

rules were created restricting the text forms that were acceptable. Tying a user question to a set of potential answer forms further restricted and clarified the acceptable "senses" of terms. In this case, we are not specifically "understanding" senses of words but we are correlating QUANDA relationships and thus restricting what is acceptable for a word based on its position within a sentence and its relationship within the question/answer pair.

The chosen approach has proved to be beneficial to FaqFinder+ accuracy. It is strict enough in its rules such that accuracy is maintained. It is parameterized to an extent that allows it to be flexible enough to handle a variety of question domains. This was demonstrated when QUANDA was applied to TREC data. While FaqFinder+ has benefited from inclusion of QUANDA, the mechanism is applicable to other environments.

One area that may specifically benefit from this approach is WEB search. A rudimentary form of this is already used by GOOGLE as was pointed out earlier in this research (GOOGLE maintains data for a select set of anticipated questions). Within a search engine, the benefit would be on the delivery of results. Search engines return an ordered list of results that are largely arranged according to word frequency weightings. QUANDA would help here because it might be able to further refine the returned list of results. Perhaps one of the results might contain the anticipated answer form and thus would be ranked at the top of the list.

Since this research effort began approximately eight years ago, the question answering research area has progressed. It is encouraging to see that the direction chosen by this research is being actively pursued by other researchers. The following are examples in recent TREC conferences where semantic analysis concepts are being studied that are related to this research.

(Rousinnov, et al., 2004) attempted a simplistic approach to semantic disambiguation in a question answering approach for TREC 2004. It was simplistic in the sense that no attempt was made to do part of speech tagging, phrase recognition or any other analysis of question or answer text. Only text pattern matching with potentially a simple verb such as "is" or "was" was considered. While the results were marginally positive, it did show that that there is merit in the approach.

(Schone, et al. 2004) and (Schone, et. al. 2005) used a system called QACTIS with some encouraging results. System had 3 major components. The first two were on the question interpretation and the third on answer matching. There was a template matching approach used. Like this research's use of the Brill Tagger, the system used a tagger called "Charniak Parser". Target answers would also be parsed by the parser and then an attempt was made in building relationships between the question and potential answers. The results were somewhat encouraging but the system had better luck with certain question types (no specification of which types). Results improved in 2005 on the factoid answering track from 20.4% to 25.4%.

(Gaizauskas, et. al. 2005) attempted an approach that attempted semantic disambiguation as well. As a first step, it utilized question rewrite by resolving sentence references. If for example the question is "How do I do it?", there were attempts to look in prior text to determine what "it" refers to. As for question matching, there was an attempt to determine what the target answer was topic (qvar) was and then to attempt a match (EAT). There was no deep semantic analysis done but the system did show promise by achieving .21 on the factoid answer track.

(Ahn, et. al. 2005) used the QED system that performs a low level NLP tagging of both question and potential answers. The approach did show promise, however, there were issues with accuracy. This may be partly due to the logic that is used to match question/answer. The logic did not attempt to use deeper semantics. Instead, it attempted to word match between question/answer and utilized a scoring method that reward "perfect matches" and lowered scores for less that perfect matches. Still, the approach did show promise.

The approach used by Molla in AnswerFinder (Molla 2006) is to parse both the question and potential answers. Once parsed, named entities (concepts) and relationships (possibly verbs) are used to determine potential matches. Potential matches are ranked and returned as possible answers. The results were disappointing to the authors, however, the approach has merit. It may be that the matching approach is too simplistic. The approach did utilize filtering of potential documents and considering only sentences that were filtered for relevance.

MITRE Corporation's approach shows some interesting features in its approach. In the question analysis component, various features are extracted. An attempt is made to identify the object or agent of the sentence, salient entity (what the question is about), answer restriction (e.g., first woman in space), geographical and temporal restrictions (e.g., in the nineteenth century). The results achieved by MITRE scored in the median of all results (.208).

(Saxena, et. al. 2007) elected to extract information from the user question. An attempt is made to ascertain the sentence object. The question is classified according to eight categories (person, location, organization, number title, jobtitle, date and money). The mechanism used for matching is not explained, but there is mention of pattern matching with potential answers. Results achieved were in the median of all results for TREC 16.

All of the above recent activity indicates good interest in exploration of deeper analytics. This is both encouraging and an affirmation that this research direction has merit and value.

In conclusion, the research presented within this thesis has achieved its goal of improving frequently asked question answering. The effort has shown that there is applicability of the research methods to other problem domains such as TREC or WEB search. This has also been a learning experience for this researcher, in that it has provided an understanding of the difficulties and opportunities in the area of text search.

While this research effort has concluded, the approach demonstrated within these pages has additional possibilities that will be explored.

I wish to thank Dr. Steven Lytinen, Dr. Hani Abu-Salem, Dr. Peter Hastings and Dr. Boris Zibitsker for their willingness to serve on my committee. Special thanks to Dr. Lytinen for his patience during this prolonged effort.

Appendix

8 Appendix A - Proposed Time Line

June, 2006: Distribute proposal to committee

Sept, 2006: Propose

Oct - Dec, 2006: (Task I) Question/Question enhancements

Scheduled Work:

1. Establish an environment to rerun all FaqFinder+ functionality and obtain a baseline to document existing accuracy. *(Completed)*
2. Add grammar and parsing enhancements and measure performance. *(Completed)*
3. An analysis will be done to determine the best ways to enhance Subject/Verb/Object identification *(Completed).*
4. Measure results and write a paper reviewing the findings. *(Completed: Paper presented at the 2nd Language Technology Conference)*

Jan-Dec, 2007: (Task I) Question/Question enhancements

Scheduled Work:

1. Enhancements to Subject/Verb/Object will be implemented. *(Completed)*
2. Investigate options for named entity support and implement. *(Completed)*
3. Investigate options for question type identification and implement. *(Completed)*
4. Measure results of Subject/Verb/Object improvements *(Completed)*
5. WHAT/WHO support *(Completed)*

Jan - May, 2008: (Task III) - FAQ Harvest – New data set creation

Scheduled Work:

1. Gather new test data set *(Completed)*
2. Gather new question set *(Completed)*
3. Retest old data set with WHO/WHAT support *(Completed)*
4. Retest new data set *(Completed)*

Jun-Aug, 2008: Complete Thesis

Scheduled Work:

1. I will complete the Thesis and get it ready for distribution.*(Completed)*

Dec, 2008: Distribute Thesis *(Completed)*

March, 2009: Defend Thesis

9 Appendix B - Preliminary FaqFinder+ RAW Results

Raw Results are available online at:
http://shrike.depaul.edu/~smlynarc/phd/phd.html

10Appendix C – TREC Top 5 Raw Results

Raw Results are available online at:
http://shrike.depaul.edu/~smlynarc/phd/phd.html

11Appendix D – TREC Top 10 Raw Results

Raw Results are available online at:
http://shrike.depaul.edu/~smlynarc/phd/phd.html

12Appendix E – TREC 9 General Results

Participant	50byte -MMR	250byte – MMR	Questions Answered	Questions Missed
MTR00S1 – Miter Corporation	0.038		52	630
MTR00L1 – Miter Corporation		0.102	124	558
INQ9WSUM (University of Massachussetts		0.325	341	341
INQ9AND (University of Massachussetts		0.339	338	344
MSQ9L50 (Microsoft Corporation)	0.196		207	475
MSQ9L250 (Microsoft Corporation)		0.264	265	417
IBMHLT0050 (IBM Corporation)	0.290		290	392
IBMHLT00250 (IBM Corporation)		0.457	419	263
SHEF50EA (University of Sheffield)	0.159		161	521
SHEF50 (University of Sheffield)	0.206		212	470
SHEF250 (University of Sheffield)		0.330	331	351
SHEF250P (University of Sheffield)		0.345	347	335
PIR0QAS1 (Queens College)	0.263		248	434
PIR0QAS2 (Queens College)	0.282		281	401
PIR0QAL1 (Queens College)		0.433	389	293
PIR0QAL2 (Queens College)		0.464	418	264
ICRJC99A (Imperial College)	0.231		228	454
ICRJC99B (Imperial College)		0.385	334	348
clr00b1-2 (CL Research)	0.119		113	569
clr00b1-2 (CL Research)	0.135		132	550
CLR00S1 (CL Research)		0.287	289	393
CLR00S2 (CL Research)		0.296	296	386
KUQA250A (Korea University)		0.371	344	338
LCCSMU1 (Southern Methodist University)		0.760	587	95
LCCSMU2 (Southern Methodist University)	0.580		453	229
ALI9A250 (Alicante University)		0.349	352	330
ALI9A50 (Alicante University)	0.227		231	451
ALI9C250 (Alicante University)		0.356	361	321
ALI9C50 (Alicante University)	0.230 .		231	451
NTTD9QAa1S (NTT DATA Corp)	0.216		221	461
NTTD9QAa2S (NTT DATA Corp)	0.231		243	439

Participant	50byte -MMR	250byte – MMR	Questions Answered	Questions Missed
KAIST9qa1 (Korea Advanced Institute of Science and Technology)	0.212		214	468
KAIST9qa2 (Korea Advanced Institute of Science and Technology)		0.327	320	362
NTTD9QAa1L (NTT DATA Corp)		0.391	383	299
NTTD9QAb1L (NTT DATA Corp)		0.381	369	313
Icat050 (LIMSI)	0.178		183	499
Icat250 (LIMSI)		0.407	375	307
LCIX250 (LIMSI)		0.375	372	310
FDUT9QS1 (Fudan University)	0.192		200	482
FDUT9QL1 (Fudan University)		0.339	313	369
UDEMING1 (University of Montreal)		0.352	340	342
UDEMING2 (University of Montreal)		0.366	357	325
UDEMSHRT (University of Montreal)	0.179		196	486
UDEMEXCT (University of Montreal)	0.149		152	530
PISA0 (Dipartimento di Informatica – Pisa)		0.227	226	456
IBMKR250 (IBM T.J. Watson Center – Hawthorne)		0.416	375	307
IBMKA250 (IBM T.J. Watson Center – Hawthorne)		0.425	388	294
IBMKA50 (IBM T.J. Watson Center – Hawthorne)	0.309		276	406
IBMKR50 (IBM T.J. Watson Center – Hawthorne)	0.315		280	402
FDUT9QL2 (Fudan University)		0.319	288	394
PISAB (Dipartimento di Informatica – Pisa)		0.238	240	442
XEROXQA9A (Xerox Research Center Europe)	0.227		229	453
XEROXQA9L (Xerox Research Center Europe)		0.353	333	349
PISAS (Dipartimento di Informatica – Pisa)	0.084		105	577
FDUT9QS2 (Fudan Unversity)	0.195		187	495
SUT9P2C3C050 (Syracuse U. Center for Natural Language Processing)	0.249		243	439
SUT9P2C3C250 (Syracuse U. Center for Natural Language Processing)		0.385	363	319
SUT9BN3C050 (Syracuse U. Center for Natural Language Processing)	0.247		246	436
SUT9BN3C250 (Syracuse U. Center for Natural Language Processing)		0.365	348	334
UWMT9QAS0 (University of Waterloo MultiText Project)	0.321		287	395
UWMT9QAL0 (University of Waterloo MultiText Project)		0.456	410	272
QNTUA02 (National Taiwan University)		0.315	306	376
QNTUA01 (National Taiwan University)		0.315	305	377
QNTUA03 (National Taiwan University)		0.278	288	394
CNXROLE2 (Conexor OY)	0.064		48	634
SCAI9QNA2 (Seoul National University)	0.101		105	577
SCAI9QNA3 (Seoul National University)		0.217	216	466
UALBERTA (University of Alberta)	0.180		203	479
ISI0A50 (USC/Information Sciences Institute)	0.318		297	385
SUNONE (Sun Microsystems)		0.340	315	367
SUNTWO (Sun Microsystems)		0.345	320	362
UWMT9QAS1 (Unversity of Waterloo MultiText Project)	0.257		252	430
UWMT9QAL1 (Unversity of Waterloo MultiText Project)		0.460	417	265
UIQA002 (University of Iowa)		0.212	263	419
UIQA001 (University of Iowa)		0.227	262	420

13Appendix F – Question Set One

13.1 Question Set One - FaqFinder+ Answerable Questions

The following is the list of questions determined to be "answerable". This determination required reading of each FAQ file to decide whether a particular question was addressed within the FAQ. No prior judgment was made about whether FaqFinder+ mechanisms would have difficulty with a given question, only that the FAQ (either directly or directly) provided insight into the question.

Number	Question
1	What are the differences, relative advantages, and disadvantages, of the various digital audio data reduction schemes, including delta-sigma 1-bit encoding, Philips' DCC (Digital Compact Cassette) compression, and Sony's MD (MiniDisc) compression?
2	What is the best single-CD player currently available?
3	Why did DAT (Digital Audio Tape) not catch on as a pre-recorded, or consumer, format?
4	How do I get started making beer?
5	Is making beer hard?
6	What are the advantages and disadvantages of corn sugar vs. powdered malt extract for the final in-bottle fermentation?
7	What sort of equipment do I need to make some homemade beer?
8	Where can I find out how to make beer?
9	Can a non-drug using lesbian contract AIDS?
10	Can an infected partner give someone HIV by performing oral sex on him/her?
11	Can you get AIDS from kissing someone?
12	Can you get HIV from a mosquito bite?
13	Does the diaphragm provide any protection from AIDS?
14	Do I have to pay for satellite service, or can I receive it for free after I buy a dish?
15	Are those new miny satellite dishes for real?
16	Is the picture quality of satellite TV identical to cable or broadcast tv?
17	For cutting full sized sheets of plywood, what tool is better, a circular saw or a table saw? (better as in safer & more accurate, not necessarily faster)
18	What's the best table saw to buy for an amateur who doesn't want to spend big bucks?
19	Is diabetes always inherited?
20	Is there any hope of a treatment for juvenile-onset diabetes which does not require frequent injections? I have heard of some kind of abdominal implant which will emit regulated doses of insulin?
21	How can I erase bad credit from my credit reports?
22	How safe is my credit record?
23	What are the phone numbers or addresses of the major credit reporting agencies, and what are their policies for giving out credit reports?
24	Will a mortgage payment that is a week late show up on a credit report?
25	How can I separate my credit history from my ex-husband's?
26	How do you pick a lock?
27	How much caffeine is there in tea relative to coffee?
28	Is caffeine linked to high blood pressure or asthma?
29	Is there really caffeine in chocolate?

Number	Question
30	What are the long-term effects of high caffeine consumption?
31	What is the chemistry behind a caffeine high?
32	I want to build a futon frame. Where can I get plans?
33	Are all the satellite TV channels scrambled?
34	how do I make ale?
35	what makes a speaker good?
36	Is it legal to own a lockpick?
37	Is it possible to contract AIDS from saliva?
38	What is the difference between ale and lager?
39	What do oil names mean?
40	What is insulin?
41	Are Sears table saws a good buy?
42	Which credit reporting agencies can I write to to get free credit reports?
43	Is there any caffeine at all in decaf?
44	What kind of saw should I buy to build bookcases?
45	How can I crack a krypto lock?
46	What kind of coffee has the most caffeine?
47	How do I brew beer?
48	How do i use perl?
49	what is perl language?
50	how do I pick a lock?
51	how does one get diabetes?
52	where can I get Perl?
53	how much caffeine does coke have?
54	what does a caffeine molecule look like?
55	are bose speakers as good as people say they are?
56	can you tell me where to find perl?
57	Is Satellite TV better than cable?
58	How can I pick a combination lock?
59	What beer kits are good?

Number	Question
60	can you get AIDS from sweat
61	can insects transmit AIDS
62	can AIDS be transmitted by saliva
63	can I get AIDS from kissing someone?
64	so was deckard a replicant
65	Who is Noam Chomsky?
66	how can i learn more about diabetes
67	does garnet work on solaris?
68	What is contact juggling?
69	I would like to learn more about ballet. What should I do?
70	How much caffeine can I safely consume on a daily basis?
71	i need to buy some stereo equipment. how can i tell a good speaker from a bad one?
72	i have developed a software package. how can i copyright it?
73	is there any difference between esperanto and a natural language
74	How fast is a PowerPC?
75	Is CFS fatal?
76	do you have anything on AIDS
77	what happens if i don't get my ferret descented?
78	What are hops?
79	how do i write a shell script
80	do the same viruses affect all computers or are they specific to particular machines or operating systems?
81	I've heard about cohousing, but it seems like it might be difficult to find other people that are sufficiently compatible to make it work. How is this typically done?
82	what is a good crossword puzzle dictionary?
83	What does the caffeine molecule look like?
84	Who makes the best mid-priced CD player
85	How do I enable the Perl debugger?
86	are there any good introductory books on lisp?
87	can you suggest a good reading list on diabetes?
88	when did Virginia secede?
89	Can I unscramble video?

Number	Question
90	How can I print out a string in Perl?
91	How do I compile my Perl script?
92	How do I setup anonymous ftp on Solaris?
93	How do I use Perl to interface to Oracle?
94	How do I use ftp?
95	How do I set up a network
96	How frequent should I change my car engine oil?
97	How many calories are in a hot fudge sundae?
98	How many calories are in a pound?
99	How often should I change my car's oil?
100	How to cook tofu?
101	How to install a dish?
102	Is CFS a real disease?
103	Is Chronic Fatigue Syndrome a real disease?
104	Is downshifting a good way to slow down my car?
105	Is downshifting a good way to stop my car?
106	Is there more caffeine in coffee or in tea?
107	Should I downshift my car?
108	What about DOT-5 brake fluids?
109	What is the easiest guide to PERL
110	When should I change my oil filter?
111	Are there any good lisp to c translators?
112	Do ferrets make good pets
113	How do I change hostname on a Sun machine
114	How do I create a perl module?
115	How do I kill fleas?
116	How do I open a file for read and write in perl?
117	How do I search for an item in an array in perl?
118	How often should I change my oil?
119	Satellite TV
120	should I downshift when braking?

13.2 Question Set One - FaqFinder+ Unanswerable Questions

The following is a list of questions were determined to be “unanswerable” even though they were relevant to the topic area.

Number	Question
1	Are motorcycles more dangerous than cars? How much more dangerous?
2	How do I get rust off of the chrome exhaust pipes on my motorcycle?
3	I just got my motocycle out of the garage after the winter, and it won't start. What should I do?
4	In which states is wearing a helmet on a motorcycle compulsory?
5	Where can I get motorcycle insurance at a reasonable cost?
6	What are the ten top jazz albums?
7	What is more common, \hot 2 or \hot 3\, in going from balanced to unbalanced audio signal cables? That is, when the ground wire has to be strapped onto one of the signal wires in order to plug into an ungrounded inlet, which pin is more commonly used?
8	Why does one of my speakers come in and out at random?
9	Why is my portable cd player skipping?
10	In perl, when doing a regular expression search and replace how do you include a portion of the serch expression in the replace string?
11	At what point in the process would I add the fruit to my wort if I wanted to make a fruit beer?
12	Can one really make good beer from a kit?
13	Doesn't homemade beer taste awful?
14	Doesn't homemade beer smell bad when its brewing?
15	How much does homebrewed beer cost in comparison to store bought beer?
16	When making beer, what ingredients are specific to a \dark\ beer?
17	Why does commercial American beer taste so awful, when imported beers are so much better, and even American \microbrews\ are so much better too?
18	At what point in the process would I add the fruit to my wort if I wanted to make a fruit beer?
19	Can one really make good beer from a kit?
20	Doesn't homemade beer taste awful?
21	Doesn't homemade beer smell bad when its brewing?
22	How much does homebrewed beer cost in comparison to store bought beer?
23	When making beer, what ingredients are specific to a \dark\ beer?
24	Why does commercial American beer taste so awful, when imported beers are so much better, and even American \microbrews\ are so much better too?
25	Will bleach clean infected needles?
26	How come unfinished mahogany is used for external applications, like on boats or on decks?
27	How do I prevent my work from splintering when routing edges into my work?

Number	Question
30	What sorts of stains and finishes are appropriate for various kinds of wood used in home furniture?
31	What is the best brand of fat free cheddar cheese?
32	What if a vegetarian has food allergies?
33	How do medeco locks differ from normal pin tumbler locks?
34	Is there any way to reverse the effects of caffeine so that one can get to sleep?
35	Is it a better deal to buy from Saturn, with their no-haggle fixed-price policy and possibly slightly higher profit margin, or to haggle from a regular dealer with the possibility of getting a slightly better deal?
36	What do I need to know to get the best deal on a used car?
37	What is different about an \overhead cam\ engine, compared to more traditional designs?
38	Why do automatic cars have fewer gears than standards?
39	How much does it cost to get a new transmission?
40	How do I paint Warhammer miniatures?
41	How can I prevent my beer bottles from exploding?
42	What's the greatest selling beer in Germany?
43	Is a butterfly a good boat?
44	What is a pilsner?
45	What do I do about a yeasty taste in my beer?
46	How slow are perl gdbm associative arrays?
47	Is yogurt a substitute for sour cream?
48	Can I pick a dead bolt lock?
49	Will used car dealers try to cheat me?
50	How can I improve the performance of my sunos httpd server?
51	Why should I choose a computer with a PowerPC processor?
52	What sort of wood should I use to build a shelf?
53	How do I build a shelf?
54	How do I build a bookshelf?
55	How do I build a desk?
56	Do all conservatives love Rush Limbaugh?
57	How do I make a white sauce?
58	Where can I get a white sauce?
59	What kind of cheese has the most mold?
60	How do I disinfect needles to prevent aids? does bleach work?
61	How do I set up a serial printer under Solaris?
62	How do I rotate a 3d point?

13.3 Question Set One - FaqFinder+ WHO Parse Failures

Number	Question	Reason
1	Who makes the best mid-priced CD player	“priced” Tagger mistag as VERB

13.4 Question Set One - FaqFinder+ WHAT Parse Failures

Number	Question	Reason
1	What are the differences, relative advantages, and disadvantages, of the various digital audio data reduction schemes, including delta-sigma 1-bit encoding, Philips' DCC (Digital Compact Cassette) compression, and Sony's MD (MiniDisc) compression?	Missing Rule
2	What is the best single-CD player currently available?	Missing Rule
3	What are the advantages and disadvantages of corn sugar vs. powdered malt extract for the final in-bottle fermentation?	Missing Rule
4	What sort of equipment do I need to make some homemade beer?	Missing Rule
5	What's the best table saw to buy for an amateur who doesn't want to spend big bucks?	Missing Rule
6	What are the phone numbers or addresses of the major credit reporting agencies, and What are their policies for giving out credit reports?	Missing Rule
7	What are the long-term effects of high caffeine consumption?	
8	What is the chemistry behind a caffeine high?	“high” Tagger mistag as ADJ
9	What makes a speaker good?	Missing Rule
10	What kind of saw should I buy to build bookcases?	“saw” Tagger mistag as VERB
11	What kind of coffee has the most caffeine?	Missing Rule
12	What does a caffeine molecule look like?	“look” Tagger mistag as NOUN
13	What beer kits are good?	“good” Tagger mistag as ADJ
14	What is contact juggling?	“juggling” Tagger mistag as VERB
15	I would like to learn more about ballet. What should I do?	Missing Rule
16	What happens if i don't get my ferret descented?	Missing Rule
17	What are hops?	“hops” Tagger mistag as VERB
18	What is a good crossword puzzle dictionary?	Missing Rule
19	What does the caffeine molecule look like?	“look” Tagger mistag as NOUN
20	What about DOT-5 brake fluids?	Missing Rule

13.5 Question Set One - FaqFinder+ HOW Parse Failures

Number	Question	Reason
1	How can I erase bad credit from my credit reports?	Missing Rule
2	How safe is my credit record?	Missing Rule
3	How can I separate my credit history from my ex-husband's?	“separate” Tagger mistag as NOUN
4	How much caffeine is there in tea relative to coffee?	Missing Rule
5	How do i use perl?	“use” Tagger mistag as NOUN
6	how much caffeine does coke have?	Missing Rule
7	how can i learn more about diabetes	Missing Rule
8	How much caffeine can I safely consume on a daily basis?	Missing Rule
9	i need to buy some stereo equipment. how can i tell a good speaker from a bad one?	Missing Rule
10	i have developed a software package. how can i copyright it?	Missing Rule
11	How fast is a PowerPC?	Missing Rule
12	How do I enable the Perl debugger?	Missing Rule
13	How can I print out a string in Perl?	Missing Rule
14	How do I setup anonymous ftp on Solaris?	“setup” Tagger mistag as NOUN
15	How do I use Perl to interface to Oracle?	Missing Rule
16	How do I set up a network	Missing Rule
17	How frequent should I change my car engine oil?	Missing Rule
18	How many calories are in a pound?	Missing Rule
19	How often should I change my car's oil?	C
20	How to install a dish?	Missing Rule; bad sentence
21	How do I change hostname on a Sun machine	Missing Rule
22	How do I open a file for read and write in perl?	“open” Tagger mistag as ADJ and “write” as a VERB
23	How do I search for an item in an array in perl?	Missing Rule
24	How often should I change my oil?	Missing Rule

14Appendix G – WEB Questions

14.1 Web Question Set - FaqFinder+ Answerable Questions

The following is the list of questions that were deemed "answerable". This determination required reading of each FAQ file to determine whether a particular question was addressed within the FAQ. No judgment was made about whether FaqFinder+ mechanisms would have difficulty with a given question; only that the FAQ (either directly or directly) provided insight into the question.

Question #	Question
1	Are clinical trials free?
2	Can I install Java after I install OpenOffice ?
3	Can I play regular DVD discs on my Divx player?
4	Can prostate cancer be found before a man has symptoms?
5	Chemotherapy and Pain: Does Chemotherapy Hurt
6	Did Hubbard say that the way to make a million dollars is to start a religion?
7	Does beta-carotene prevent cancer?
8	Does NT support Plug and Play?
9	Hair Regrowth After Chemotherapy
10	How can I access LinkOut?
11	How can I contact the OpenOffice.org project ?
12	How can I copy a DVD?
13	How can I get OpenOffice.org in some other language other than English?
14	How can I help the OpenOffice.org? project?
15	How can I help with OpenOffice.org?
16	How does LinkOut work?
17	How do I access CAM on PubMed?
18	How do I access the parental controls on my DVD?
19	How do I copy a DVD?
20	How do I get OpenOffice.org for Mac OS X?
21	How do I install Internet Information Server?
22	How do I install the Recovery Console?
23	How do I make my DVD-ROM regionfree?
24	How do I play a DVD on my computer?
25	How do I run an application as a service?

Question #	Question
26	How is prostate cancer diagnosed?
27	How is testicular cancer detected?
28	How is testicular cancer treated?
29	How many citations are included within CAM on PubMed?
30	How much does Scientology cost?
31	How often should women get a mammogram?
32	Is HPV a Sexually Transmitted Disease?
33	Is OpenOffice really free?
34	Is there database support in OpenOffice.org?
35	Wasn't L. Ron Hubbard a science fiction writer?
36	What about my existing DVD collection?
37	What are symptoms of testicular cancer?
38	What are the risk factors for testicular cancer?
39	What are the symptoms of prostate cancer?
40	What causes bladder cancer?
41	What does DVD mean?
42	What does DVD stand for?
43	What does OpenOffice.org have to offer?
44	What does Scientology teach about drugs? Are tobacco and alcohol drugs?
45	What do LinkOut buttons look like?
46	What features does OpenOffice.org have ?
47	What is AC-3?
48	What is a dual-layer disc and does it have any advantages?
49	What is a jewel case?
50	What is a keep case?

Question #	Question
51	What is anamorphic DVD?
52	What is a prostate and why do men need it?
53	What is a Thetan and an Operating Thetan?
54	What is auditing?
55	What is CSS?
56	What is DTS?
57	What is Duplication?
58	What is DVD?
59	What is DVD video?
60	What is HD DVD?
61	What is IIS?
62	What is LinkOut?
63	What is melanoma and is it life threatening?
64	What is MPEG-2?
65	What is MPEG?
66	What is OpenOffice?
67	What is OpenOffice.org?
68	What is prostate cancer?
69	What is regional coding?
70	What is Replication?
71	What is the CCHR?
72	What is the difference between OpenOffice and StarOffice ?
73	What is the OSA?
74	What is the prostate?
75	What is the RPF?

Question #	Question
76	What is the Sea Org?
77	What is the Windows XP task switcher?
78	What is Video CD?
79	What operating system does OpenOffice run on ?
80	What resolution will the video on a movie HD DVD be?
81	What's an E-meter?
82	What's Longhorn?
83	What's the difference between Dianetics and Scientology?
84	What's the storage capacity of DVD?
85	What's WebDVD?
86	What's Windows Server 2003?
87	Where can a person find more information about prostate cancer and its treatment?
88	Where can I find help using CAM on PubMed?
89	Where can I find the resource kit?
90	Where should I install OpenOffice ?
91	Which studios support DVD?
92	Who is at risk for prostate cancer?
93	Who is Derek Fawcus?
94	Who is Jon Johansen?
95	Why do Scientologists believe in past lives?
96	Why do some DVD players cost more than others?
97	Will DVD replace CD-ROM?
98	Will DVD replace laserdisc
99	Will DVD replace VCRs?
100	What is alternative medicine?

14.2 Web Question Set - FaqFinder+ Unanswerable Questions

Question #	Question
1	What are the symptoms of prostate cancer?
2	How reliable are the screening tests for prostate cancer?
3	How is localized prostate cancer treated?
4	What is a colposcopy and what happens during the procedure?
5	What the difference between colon cancer and colorectal cancer?
6	What are the symptoms of thyroid cancer?
7	How Much Does the HPV Vaccine Cost?
8	Why do radiation patients get dry mouth and what can be done to remedy it?
9	What exactly does it mean when a person has a high PSA level?
10	Are HPV and Genital Warts the Same Thing?
11	HPV Vaccine Side Effects
12	Pap Smears After a Hysterectomy
13	Can wearing deodorants and antiperspirants cause breast cancer?
14	I have stage IIA breast cancer. What are my chances of survival?
15	I was fired from my job after I told my boss about my diagnosis. Is this lawful?
16	What are the stages of cervical cancer and what do they mean?
17	Can a Virgin Get HPV?
18	What is a penectomy, how is it performed and are there any side effects?
19	What about those VCR-style DVD recorders such as the Panasonic DMR-E30S?
20	Which will win, Blu-ray or HD DVD?
21	What is Segment Re-Encoding?
22	How do I reset the DVD region count?
23	How do I set the DVD region in Windows 2000?
24	What about my existing DVD collection?
25	Who is behind HD DVD?

Question #	Question
26	What is FACT`s role in the DVD market?
27	Since the Sony PS3 will support BDs, will the Xbox 360 support HD DVD?
28	How can I obtain a fix for my Proline DVD-1000/Grundig GDV-200/210 Player?
29	Where can I rent DVD`s in the UK?
30	Why should I want high definition DVD?
31	Will EVD players be able to playback the other high definition DVD formats?
32	When will HD DVD be officially launched?
33	How do I uninstall OpenOffice.org?
34	How does OpenOffice.org work over NFS?
35	What schools have already committed to using OpenOffice?
36	What is the current version of OpenOffice.org?
37	How do I buy an OpenOffice.org CD?
38	Is there a groupware component for OpenOffice.org?
39	Is the Lycoris ProductivityPak compatible with OpenOffice.org?
40	How do I add more fonts to OpenOffice.org?
41	Does OpenOffice support WordPerfect file formats ?
42	Does OpenOffice come with an equivalent of Outlook ?
43	How do I get rid of the OpenOffice.org splash screen?
44	What is PubMed?
45	I see a journal listed on the Library's Electronic Journals page, but I don't see the Library's button for it in PubMed. Why?
46	What electronic journals are available through LinkOut?
47	I'm trying to view LinkOut full text from home. Why does the electronic journal keep asking me for a password?
48	Who's in charge of the church, now that Hubbard is dead?
49	What is the purpose of alt.religion.scientology? Is it for or against Scientology?
50	What scientific evidence is there for Dianetics and Scientology?

Question #	Question
51	Does Scientology use front groups? What are they?
52	What do the confidential levels of Scientology consist of and why are they secret?
53	What is the personality test the Scientologists offer for free? What does it measure?
54	Is Scientology a religion or a cult?
55	What is the Free Zone?
56	What is a Squirrel?
57	Who are the skeptics? Why do they spend so much time here if they don't even believe in Scientology?
58	How Does Scientology Deal with critics of the church?
59	What lawsuits has Scientology been involved in?
60	What is an SP?
61	What is PTS?
62	What is a Wog?
63	I have a family member involved in Scientology. How do I help them get out?
64	How is testicular cancer diagnosed?
65	Is followup treatment necessary? What does it involve?
66	What are the side effects of treatment?
67	How can I print to a USB printer from the command prompt?
68	Why do I hear hissing through my USB speakers when I play sounds at high volume on my laptop computer?
69	What are the Pocket PC serial cradle and USB cradle file transfer speeds?
70	Why can't I install Windows 2000 from certain USB CD-ROM drives?
71	What's Microsoft USB Flash Drive Manager for Windows XP?
72	How can I mark my USB storage devices as read-only?
73	How can I extend the USB polling period on my laptop?
74	Why did my USB 1.1 devices stop working after I updated a driver for a USB 2.0 device?
75	Why do I receive Stop Error 0x0000007E in Windows XP when I add a new USB device?

Question #	Question
76	Why doesn't my HP 8200 Series USB CD-ROM Writer install on my Dell Latitude CPt with Windows 2000?
77	When I connect a USB device to a USB hub, why do I receive an error message about available power?
78	Have you ever used an OnStream USB or IDE tape backup? If so, what?s your impression?
79	What are the IRQ's used for?
80	What is Java needed for?
81	Can multiple users edit a file concurrently?
82	Why can't I open Microsoft Office password protected files?
83	Can CAD drawing files be imported?
84	How do I turn AutoSave on?
85	Is there a grammar checker?
86	How can I record a macro ?
87	What does DVD stand for?
88	How does WINS work?
89	What is Bootcfg?
90	How do I change my picture for the Welcome screen?
91	How do I enable NTLM 2.0 support in Windows 98?

14.3 Web Question Set - FaqFinder+ WHO Parse Failures

Question #	Question	Reason
1	Who is at risk for prostate cancer?	Missing rule

14.4 Web Question Set - FaqFinder+ WHAT Parse Failures

Question #	Question	Reason
1	What about my existing DVD collection?	Missing rule
2	What does DVD mean?	Missing rule
3	What does OpenOffice.org have to offer?	Missing rule
4	What does Scientology teach about drugs? Are tobacco and alcohol drugs?	Missing rule
5	What features does OpenOffice.org have ?	“features” Tagger mis-tag as VERB
6	What is AC-3?	Misidentification of AC-3
7	What is a dual-layer disc and does it have any advantages?	Missing rule
8	What is a prostate and why do men need it?	Missing rule
9	What is a Thetan and an Operating Thetan?	Missing rule
10	What is melanoma and is it life threatening?	“melanoma” Tagger mis-tag as ADJ
11	What is MPEG-2?	Misidentification of MPEG-2
12	What is regional coding?	“Regional coding” mis-tag as VERB
13	What license is OpenOffice.org released under?	Missing rule
14	What operating system does OpenOffice run on ?	“operating” Tagger mis-tag as VERB
15	What resolution will the video on a movie HD DVD be?	Missing rule
16	What's Windows Server 2003?	Missing rule

14.5 Web Question Set - FaqFinder+ HOW Parse Failures

Question #	Question	Reason
1	How can I contact the OpenOffice.org project ?	Missing rule
2	How can I get OpenOffice.org in some other language other than English?	Missing rule
3	How can I help the OpenOffice.org? project?	Invalid sentence syntax
4	How can I help with OpenOffice.org?	Missing rule
5	How does LinkOut work?	“work” Tagger mis-tag as NOUN
6	How do I access CAM on PubMed?	“access” Tagger mis-tag as NOUN
7	How do I access the parental controls on my DVD?	“access” Tagger mis-tag as NOUN
8	How do I get OpenOffice.org for Mac OS X?	Missing rule
9	How do I install Internet Information Server?	Missing rule
10	How do I install the Recovery Console?	Missing rule
11	How do I make my DVD-ROM regionfree?	Missing rule
12	How do I play a DVD on my computer?	Missing rule
13	How do I run an application as a service?	Missing rule
14	How is testicular cancer detected?	Missing rule
15	How is testicular cancer treated?	Missing rule
16	How many citations are included within CAM on PubMed?	Missing rule
17	How often should women get a mammogram?	Missing rule

15Appendix I – TREC Questions

15.1 TREC WHO Parse Failures

Question	Reason For Parse Failure
Who created potlatch	Brill Tagger mistag of “potlatch” as a VERB
Who wrote The Pit and the Pendulum	“The Pit and the Pendulum “ is not recognized as a title and mis-parsed.
Who won the Nobel prize in literature in 1988	Missing grammar rule
Who was the king who was forced to agree to the Magna Carta	Missing grammar rule
Who was the king who signed the Magna Carta	Missing grammar rule
Who was the founding member of the Pink Floyd band	Missing grammar rule
Who was the first Russian to do a spacewalk	Missing grammar rule
Who was the author of the book about computer hackers called The Cuckoo Egg: Tracking a Spy Through the Maze of Computer Espionage	Missing grammar rule;complex sentence.
Who was the 33rd president of United States	Missing grammar rule
Who thought of teaching people to tie their shoe laces	Brill Tagger mistag of “teaching” as an adjective.
Who the lead singer of the Led Zeppelin band	Missing verb caused mis-parse
Who portrayed the man without a face in the movie of the same name	Missing grammar rule
Who played the teacher in Dead Poet Society	Unrecognized title “Dead Poet Society”
Who manufactures synthroid	Brill Tagger mistag of synthroid as an adjective.
Who manufacturers Magic Chef appliances	Brill Tagger mistag of “manufacturers” as noun. Can't fully place the blame on the tagger though since the word should have been “manufactures” and appliance is also misspelled.

15.2 TREC HOW Parse Failures

Question	Reason For Parse Failure
How did Bob Marley die	"Bob Marley" not recognized and mis-parse
How is OEM abbreviated	Brill Tagger mistag of "abbreviated" as an adjective.
How long do hermit crabs live	Missing grammar rule
How long is human gestation	Missing grammar rule
How many highway miles to the gallon can you get with the Ford Fiesta	Missing grammar rule
How many miles is it from London England to Plymouth England	"London England" and "Plymouth England" not recognized as entities
How many months does a normal human pregnancy last	Brill Tagger mistag of "last" as an adjective.
How many people lived in Nebraska in the mid 1980s	Missing grammar rule
How much does one ton of cement cost	Brill Tagger mistag of "cost" as a noun.
How much folic acid should an expectant mother take daily	Brill Tagger mistag of "take" as a noun.
How much folic acid should a pregnant woman get each day	Missing grammar rule
How was Teddy Roosevelt related to FDR	Missing grammar rule

15.3 TREC WHAT Parse Failures

Question	Reason For Parse Failure
What actor first portrayed James Bond	Missing grammar rule
What actress starred in The Lion in Winter	Unrecognized phrase "The Lion in Winter"
What are the names of Jacques Cousteau two sons	Missing grammar rule
What attracts tourists to Reims	Missing grammar rule
What card company sells Christmas ornaments	Missing grammar rule
What city airport is named Logan International	Missing grammar rule
What city houses the US headquarters of Procter and Gamble	Missing grammar rule

What city is 945 KDGE Radio located in	Missing grammar rule
What city is served by Logan International Airport	Missing grammar rule
What city is served by McCarren Airport	Missing grammar rule
What city is the Kentucky Horse Park near	Unrecognized Phrase “Kentucky Horse Park”, Missing grammar rule
What city the Kentucky Horse Park is near	Unrecognized Phrase “Kentucky Horse Park”, Missing grammar rule
What company produced rotary engine vehicles	Missing grammar rule
What country are Godiva chocolates from	Missing grammar rule
What Cruise Line does Kathie Lee Gifford advertise for	Unrecognized phrase “Kathie Lee Gifford”
What day is known as the national day of prayer	Missing grammar rule
What did Delilah do to Samson hair	Missing grammar rule
What does caliente translate to in English	Missing grammar rule
What does Final Four refer to in the sports world	Unrecognized phrase “Final Four”. Missing grammar rule
What does Knight Ridder publish	Unrecognized phrase “Knight Ridder”. Missing grammar rule
What does the acronym CPR mean	Missing grammar rule
What English word contains the most letters	Missing grammar rule
What English word has the most letters	Missing grammar rule
What ethnic group introduced the idea of potlatch	Missing grammar rule
What famous model was married to Billy Joel	Missing grammar rule
What flower did Vincent Van Gogh paint	Unrecognized phrase “Vincent Van Gogh”
What format was the major competition of VHS	Missing grammar rule
What instrument is Ray Charles best known for playing	Unrecognized phrase “Ray Charles”. Missing grammar rule
What is a ballet company Mikhail Baryshnikov has danced for	Unrecognized phrase “Mikhail Baryshnikov”. Missing grammar rule.
What is a film in which Jude_Law acted	Unrecognized phrase “Jude_Law”. Missing grammar rule.

What is a tiger that is extinct	Missing grammar rule
What is Martin Luther King Jr real birthday	Unrecognized phrase “Martin Luther King Jr”
What is of the major gods of Hinduism	Missing grammar rule.
What is one of the Seven Wonders of the Ancient World	Unrecognized phrase “Seven Wonders of the Ancient World”. Missing grammar rule. Could have used a phrase conversion of “is one of” to “is a”.
What is Pittsburg baseball team called	Missing grammar rule
What is the cultural origin of the ceremony of potlatch	Missing grammar rule
What is the daily requirement of folic acid for an expectant mother	Missing grammar rule
What is the equivalent of the Red_Cross in the Middle East	Missing grammar rule
What is the hair style called that new military recruits receive	Missing grammar rule
What is the most heavily caffeinated soft drink	Missing grammar rule
What is the name given to a group of geese	Missing grammar rule
What is the name given to the Islamic counterpart of the Red_Cross	Missing grammar rule
What is the name of the actress who starred in the movie Silence of the Lambs	Missing grammar rule
What is the name of the art of growing miniature trees	Missing grammar rule
What is the name of the Islamic counterpart to the Red_Cross	Missing grammar rule
What is the name of the Lion King son in the movie The Lion King	Missing grammar rule
What is the name of the Michelangelo painting that shows two hands with fingers touching	Missing grammar rule
What is the name of the song Will Smith sings about parents	Missing grammar rule
What is the name of the star of the cooking show Galloping Gourmet	Missing grammar rule
What is the name of the tallest mountain in the world	Missing grammar rule
What is the normal resting heart rate of a healthy adult	Missing grammar rule

What is the recommended daily requirement for folic acid for pregnant women	Missing grammar rule
What is the university of which Woodrow_Wilson was president	Missing grammar rule
What is the US location of Procter & Gamble corporate offices	Missing grammar rule
What is worth seeing in Reims	Missing grammar rule
What made Jane Goodall famous	Missing grammar rule
What part of your body contains the corpus callosum	Missing grammar rule
What President became Chief Justice after his presidency	Missing grammar rule
What radio station did Paul Harvey work for	Brill Tagger mis-tag of work as a NOUN.
What soft drink is most heavily caffeinated	Missing grammar rule
What soft drink would provide me with the biggest intake of caffeine	Missing grammar rule
What sport do the Cleaveland Cavaliers play	Unrecognized phrase “Cleaveland Cavaliers”.
What state in the United States covers the largest area	Missing grammar rule
What task does the Bouvier breed of dog perform	Missing grammar rule
What was the alternate to VHS	Brill Tagger mis-tag of “alternate” as an adjective. The question however should really be “What was the alternative to VHS”.
What was Darth Vader son named	Unrecognized phrase “Darth Vader”
What video format was an alternative to VHS	Missing grammar rule
What US Government agency registers trademarks	Brill Tagger mis-tag of “registers” as a NOUN.
What type of hunting are retrievers used for	Missing grammar rule
What type of horses appear on the Budweiser commercials	Missing grammar rule
What time of year is air travel the heaviest	Missing grammar rule
What time of year is air travel at a peak	Missing grammar rule
What time of year has the most air travel	Missing grammar rule
What time of year do most people fly	Missing grammar rule

What year was the first automobile manufactured	Missing grammar rule
What year was the Avery Dennison company founded	Unrecognized phrase “Avery Dennison company”
What year was Janet Jackson first album released	Missing grammar rule
What year was Desmond Mpilo Tutu awarded the Nobel Peace Prize	Unrecognized phrases: “Desmond Mpilo Tutu” and Nobel Peace Prize”
What year did the Vietnam War end	Brill Tagger mis-tagged “end” as a NOUN. Unrecognized phrase “Vietnam War”
What wrestling star became The Incredible Hulk	Unrecognized phrase “The Incredible Hulk”. Missing grammar rule.
What was the name of the television show staring Karl Malden that had San Francisco in the title	Missing grammar rule
What was the name of the sitcom that Alyssa Milano starred in with Tony Danza	Missing grammar rule. Unrecognized phrases: “Alyssa Milano”, “Tony Danza”
What was the name of the first Russian astronaut to do a spacewalk	Missing grammar rule

15.4Parse Failure Results

15.4.1 WHO Results For Parse Failures

Question	Comment	GOOGLE	YAHOO
Who created potlatch	FAIL	FAIL	FAIL
Who wrote The Pit and the Pendulum	TOP 1	TOP 1	TOP 1
Who won the nobel prize in literature in 1988	FAIL	TOP 2	TOP 1
Who was the king who was forced to agree to the Magna Carta	TOP 1	TOP 5	TOP 1
Who was the king who signed the Magna Carta	TOP 1	TOP 1	TOP 1
Who was the founding member of the Pink Floyd band	TOP 1	TOP 1	TOP 1
Who was the first Russian to do a spacewalk	TOP 1	TOP 3	TOP 2
Who was the author of the book about computer hackers called The Cuckoo	TOP 3	TOP 2	TOP 6
Who was the 33rd president of United States	TOP 3	TOP 1	TOP 3
Who thought of teaching people to tie their shoe laces	FAIL	TOP 8	TOP 2
Who the lead singer of the Led Zeppelin band	FAIL	TOP 3	TOP 3
Who portrayed the man without a face in the movie of the same name	FAIL	TOP 2	FAIL
Who played the teacher in Dead Poet Society	FAIL	TOP 1	TOP 1
Who manufactures synthroid	TOP 2	TOP 3	TOP 3
Who manufacturers Magic Chef applicances	FAIL	TOP 7	TOP 3
Who built the first pyramid	TOP 1	TOP 2	TOP 2
Who created baseball	TOP 4	TOP 8	FAIL
Who created television	TOP 2	FAIL	FAIL
Who created the electric guitar	FAIL	TOP 3	FAIL
Who created the paper clip	TOP 2	TOP 5	FAIL
Who found Hawaii	TOP 2	TOP 2	FAIL
Who is Anubis	TOP 1	TOP 2	TOP 2
Who is Barbara Jordan	TOP 5	TOP 5	TOP 1
Who is Charles Lindbergh	TOP 5	TOP 6	TOP 1
Who is Colin Powell	TOP 2	TOP 1	TOP 1
Who is Desmond_Tutu	TOP 1	TOP 1	TOP 1
Who is William Wordsworth	TOP 1	TOP 5	TOP 1
Who killed Billy_the_Kid	TOP 2	TOP 2	TOP 2
Who made the first airplane	TOP 3	TOP 1	TOP 1
Who owns CNN	TOP 7	TOP 3	TOP 6
Who was Buffalo_Bill	TOP 3	TOP 1	TOP 1
Who was Charles Lindbergh wife	TOP 1	TOP 1	TOP 1
Who was Jane Goodall	TOP 1	TOP 2	TOP 1
Who was Picasso	TOP 2	TOP 2	TOP 1
Who was the first king of England	TOP 1	TOP 3	TOP 2
Who won the Battle of Gettysburg	TOP 2	TOP 1	FAIL
Who wrote An Ideal Husband	TOP 3	TOP 3	TOP 1

15.4.2 WHO Results For Parse Successes

Question	ASK.COM	GOOGLE	YAHOO
CNN is owned by whom?	FAIL	TOP 4	TOP 7
Who assassinated President McKinley?	TOP 2	TOP 1	TOP 1
Who built the first pyramid?	TOP 1	TOP 1	TOP 6
Who coined the term "cyberspace" in his novel "Neuromancer"?	TOP 2	TOP 1	TOP 2
Who created "The Muppets"?	TOP 3	TOP 1	TOP 1
Who created the character James Bond?	TOP 1	TOP 1	TOP 5
Who created the character of Scrooge?	FAIL	TOP 3	TOP 2
Who created the comic strip, "Garfield"?	TOP 1	TOP 4	TOP 2
Who first broke the sound barrier?	TOP 1	TOP 1	TOP 2
Who found Hawaii?	TOP 2	TOP 2	FAIL
Who invented "The Muppets"?	TOP 1	TOP 1	TOP 1
Who invented baseball?	TOP 3	TOP 2	TOP 3
Who invented basketball?	TOP 1	TOP 4	TOP 2
Who invented paper?	TOP 1	TOP 1	TOP 2
Who invented silly putty?	TOP 1	TOP 2	TOP 1
Who invented television?	TOP 3	TOP 2	TOP 7
Who invented the electric guitar?	TOP 6	TOP 2	TOP 10
Who invented the game bowling?	FAIL	FAIL	FAIL

Question	ASK.COM	GOOGLE	YAHOO
Who invented the game Scrabble?	TOP 8	TOP 1	TOP 6
Who invented the paper clip?	TOP 2	TOP 2	TOP 6
Who invented the radio?	TOP 1	TOP 2	TOP 4
Who is Anubis?	TOP 2	TOP 1	TOP 1
Who is Barbara Jordan?	TOP 3	TOP 4	TOP 1
Who is buried in the great pyramid of Giza?	FAIL	FAIL	TOP 2
Who is Charles Lindbergh?	TOP 1	TOP 6	TOP 1
Who is Colin Powell?	TOP 1	TOP 4	TOP 1
Who is Coronado?	TOP 3	TOP 2	TOP 1
Who is Desmond Tutu?	TOP 1	TOP 2	TOP 1
Who is Henry Butler?	TOP 1	TOP 1	TOP 1
Who is Langston Hughes?	TOP 2	TOP 3	TOP 1
Who is Peter Weir?	TOP 4	TOP 7	TOP 1
Who is Secretary-General of the United Nations?	TOP 1	TOP 1	TOP 1
Who is Terrence Malick?	TOP 1	TOP 2	TOP 1
Who is the creator of "The Muppets"?	TOP 1	TOP 1	TOP 1
Who is the emperor of Japan?	TOP 3	TOP 1	TOP 1
Who is the fastest swimmer in the world?	TOP 3	TOP 5	FAIL
Who is the founder of the Wal-Mart stores?	TOP 7	TOP 1	TOP 2
Who is the Greek God of the Sea?	TOP 1	TOP 1	TOP 2
Who is the leader of India?	TOP 1	TOP 1	TOP 2
Who is the monarch of the United Kingdom?	FAIL	FAIL	FAIL
Who is the owner of CNN?	TOP 2	TOP 1	TOP 1
Who is the president of Bolivia?	TOP 1	TOP 1	TOP 1
Who is the prime minister of Australia?	TOP 1	TOP 1	TOP 1
Who is the prophet of the religion of Islam?	TOP 2	TOP 2	TOP 2
Who is the richest person in the world?	TOP 2	TOP 2	FAIL
Who is the richest woman in the world?	TOP 1	TOP 5	TOP 9
Who is William Wordsworth?	TOP 1	TOP 6	TOP 2
Who is Zebulon Pike?	TOP 4	TOP 8	FAIL
Who killed Caesar?	TOP 2	TOP 1	TOP 5
Who killed Martin Luther King?	TOP 1	TOP 2	TOP 9
Who made the first airplane that could fly?	TOP 1	TOP 1	TOP 1
Who made the first airplane?	TOP 3	TOP 1	TOP 1
Who made the rotary engine automobile?	TOP 1	FAIL	TOP 1
Who manufactures the software, "PhotoShop"?	TOP 2	TOP 4	TOP 6
Who owns CNN?	TOP 8	TOP 3	TOP 5
Who owns the St. Louis Rams?	TOP 6	FAIL	TOP 3

Question	ASK.COM	GOOGLE	YAHOO
Who portrayed Fatman in the television show, "Jake and the Fatman	FAIL	FAIL	FAIL
Who portrayed Jake in the television show, "Jake and the Fatman"?	FAIL	TOP 6	FAIL
Who provides telephone service in Orange County, California?	FAIL	FAIL	FAIL
Who reports the weather on the "Good Morning America" television s	FAIL	FAIL	TOP 1
Who shot Billy the Kid?	TOP 1	TOP 1	TOP 2
Who started the Dominos Pizza chain?	TOP 6	TOP 2	FAIL
Who used to make cars with rotary engines?	TOP 1	TOP 2	TOP 1
Who was Buffalo Bill?	TOP 3	TOP 3	TOP 3
Who was Charles Lindbergh's wife?	TOP 5	TOP 1	TOP 2
Who was considered to be the father of psychology?	TOP 2	TOP 1	TOP 1
Who was Darth Vader's son?	TOP 2	TOP 2	TOP 1
Who was Jane Goodall?	TOP 1	TOP 3	FAIL
Who was Lacan?	TOP 5	TOP 7	TOP 2
Who was Maria Theresa?	TOP 2	TOP 1	TOP 1
Who was Monet?	TOP 1	TOP 1	TOP 1
Who was Picasso?	TOP 1	TOP 1	TOP 1
Who was Quetzalcoatl?	TOP 9	TOP 2	TOP 1
Who was Samuel Johnsons's friend and biographer?	FAIL	TOP 3	TOP 3
Who was the 21st U.S. President?	TOP 1	TOP 2	TOP 1
Who was the architect of Central Park?	TOP 1	TOP 2	TOP 4
Who was the first coach of the Cleveland Browns?	TOP 1	TOP 1	TOP 2
Who was the first king of England?	TOP 1	FAIL	FAIL
Who was the first U.S. president ever to resign?	TOP 6	TOP 1	TOP 4
Who was the first woman in space?	TOP 2	TOP 1	TOP 1
Who was the inventor of silly putty?	TOP 1	TOP 1	TOP 2
Who was the oldest U.S. president?	TOP 6	TOP 3	TOP 4
Who was the president of Vichy France?	TOP 1	TOP 2	TOP 1
Who was the tallest U.S. president?	TOP 1	TOP 2	TOP 3
Who was Whitcomb Judson?	TOP 1	TOP 1	TOP 1
Who won the Battle of Gettysburg?	TOP 2	TOP 1	FAIL
Who won the rugby world cup in 1987?	TOP 3	TOP 1	TOP 2
Who won the Superbowl in 1982?	TOP 1	TOP 3	TOP 2
Who wrote "An Ideal Husband"?	TOP 1	TOP 2	TOP 3
Who wrote "The Scarlet Letter"?	TOP 1	TOP 1	TOP 1
Who wrote the book, "Huckleberry Finn"?	TOP 1	TOP 1	TOP 2
Who wrote the book, "Song of Solomon"?	TOP 4	TOP 2	TOP 3
Who wrote the book, "The Grinch Who Stole Christmas"?	TOP 6	TOP 1	TOP 4

15.4.3 HOW Results For Parse Failures

Question	ASK.COM	GOOGLE	YAHOO
How did Bob Marley die	TOP 4	TOP 2	TOP 1
How is OEM abbreviated	FAIL	TOP 4	TOP 3
How long do hermit crabs live	FAIL	TOP 1	FAIL
How long is human gestation	TOP 4	TOP 1	TOP 3
How many highway miles to the gallon can you get with the Ford Fiesta	FAIL	FAIL	FAIL
How many miles is it from London England to Plymouth England	FAIL	TOP 9	TOP 10
How many months does a normal human pregnancy last	FAIL	TOP 3	TOP 7
How many people lived in Nebraska in the mid 1980s	FAIL	FAIL	FAIL
How much does one ton of cement cost	FAIL	TOP 8	FAIL
How much folic acid should an expectant mother take daily	TOP 1	TOP 10	TOP 5
How much folic acid should a pregnant woman get each day	FAIL	TOP 3	TOP 1
How was Teddy Roosevelt related to FDR	TOP 1	FAIL	TOP 1
How many home_run did Babe_Ruth hit in his lifetime	TOP 3	TOP 1	TOP 1
How many soldiers died in World_War_II	FAIL	TOP 2	TOP 2
How many states have a lottery	FAIL	TOP 9	FAIL
How old is the sun	FAIL	TOP 1	TOP 3
How many islands does Fiji have	FAIL	FAIL	FAIL

15.4.4 HOW Results For Parse Successes

Question	ASK.COM	GOOGLE	YAHOO
About how many soldiers died in World War II?	TOP 5	FAIL	FAIL
How big is Australia?	TOP 1	FAIL	FAIL
How big is our galaxy in diameter?	TOP 4	TOP 1	TOP 1
How big is the Electoral College?	FAIL	FAIL	FAIL
How did Bob Marley die?	TOP 4	TOP 2	TOP 3
How do you abbreviate "Original Equipment Manufacturer"?	TOP 2	TOP 7	TOP 8
How far away is the moon?	FAIL	TOP 2	TOP 3
How fast can a Corvette go?	TOP 2	TOP 6	FAIL
How hot does the inside of an active volcano get?	FAIL	TOP 4	FAIL
How hot is the core of the earth?	FAIL	FAIL	TOP 8
How is thalassemia defined?	TOP 2	TOP 2	TOP 7
How large is Missouri's population?	TOP 2	TOP 4	TOP 1
How long would it take for a $50 savings bond to mature?	FAIL	TOP 3	TOP 3
How long would it take to get from Earth to Mars?	TOP 4	TOP 4	FAIL
How many astronauts have been on the moon?	TOP 3	TOP 1	TOP 2
How many casinos are in Atlantic City, NJ?	FAIL	FAIL	FAIL
How many continents are there?	TOP 1	TOP 1	TOP 7
How many copies of an album must be sold for it to be a gold album?	FAIL	TOP 2	FAIL
How many counties are in Indiana?	TOP 8	TOP 3	FAIL
How many dogs pull a sled in the Iditarod?	FAIL	FAIL	FAIL
How many films did Ingmar Bergman make?	FAIL	FAIL	FAIL
How many hexagons are on a soccer ball?	TOP 6	TOP 2	FAIL
How many home runs did Babe Ruth hit in his lifetime?	TOP 9	TOP 1	TOP 4
How many home runs did Lou Gehrig have during his career?	FAIL	TOP 9	FAIL
How many islands does Fiji have?	FAIL	FAIL	FAIL
How many people die from snakebite poisoning in the U.S. per year?	TOP 10	TOP 2	FAIL
How many people live in Chile?	TOP 1	FAIL	FAIL
How many people watch network television?	FAIL	TOP 6	FAIL
How many states have a "lemon law" for new automobiles?	FAIL	FAIL	TOP 7
How many states have a lottery?	FAIL	TOP 9	FAIL
How many Stradivarius violins were ever made?	FAIL	TOP 9	FAIL
How many Superbowls have the 49ers won?	FAIL	TOP 1	FAIL
How many types of lemurs are there?	TOP 4	TOP 3	FAIL
How many years ago did the ship Titanic sink?	TOP 1	FAIL	TOP 10
How many zip codes are there in the U.S.?	FAIL	TOP 2	FAIL
How much calcium should an adult female have daily?	FAIL	TOP 4	TOP 5
How much folic acid should an expectant mother get daily?	TOP 6	TOP 10	TOP 9
How much in miles is a ten K run?	FAIL	TOP 5	FAIL
How much money does the Sultan of Brunei have?	FAIL	FAIL	FAIL
How old is the sun?	FAIL	TOP 1	TOP 3
How tall is kilamanjaro?	FAIL	TOP 1	TOP 7
How tall is the giraffe?	TOP 3	TOP 1	TOP 2
How wide is the Atlantic Ocean?	FAIL	TOP 2	TOP 2

15.4.5 WHAT Results For Parse Failures

Question	ASK.COM	GOOGLE	YAHOO
What actor first portrayed James Bond	TOP 1	TOP 2	TOP 1
What actress starred in The Lion in Winter	TOP 1	TOP 5	TOP 3
What are the names of Jacques Cousteau two sons	TOP 2	TOP 6	TOP 5
What attracts tourists to Reims	FAIL	TOP 8	TOP 7
What card company sells Christmas ornaments	FAIL	TOP 1	TOP 1
What city airport is named Logan International	TOP 2	TOP 1	TOP 1
What city houses the US headquarters of Procter and Gamble	TOP 1	TOP 3	TOP 3
What city is 945 KDGE Radio located in	TOP 7	TOP 3	TOP 1
What city is served by Logan International Airport	TOP 1	TOP 1	TOP 2
What city is served by McCarren Airport	TOP 9	TOP 2	TOP 5
What city is the Kentucky Horse Park near	TOP 1	TOP 1	TOP 2
What city the Kentucky Horse Park is near	TOP 1	TOP 1	TOP 1
What company produced rotary engine vehicles	TOP 9	TOP 2	TOP 1
What country are Godiva chocolates from	TOP 4	FAIL	FAIL
What Cruise Line does Kathie Lee Gifford advertise for	TOP 7	TOP 4	TOP 1
What day is known as the national day of prayer	TOP 1	TOP 4	TOP 1
What did Delilah do to Samson hair	TOP 1	TOP 1	TOP 1
What does caliente translate to in English	TOP 1	TOP 4	FAIL
What does Final Four refer to in the sports world	FAIL	TOP 1	TOP 2
What does Knight Ridder publish	FAIL	TOP 1	TOP 10
What does the acronym CPR mean	FAIL	TOP 5	TOP 5
What English word contains the most letters	TOP 1	FAIL	FAIL
What English word has the most letters	TOP 5	FAIL	FAIL
What ethnic group introduced the idea of potlatch	FAIL	FAIL	FAIL
What famous model was married to Billy Joel	FAIL	TOP 1	TOP 1
What flower did Vincent Van Gogh paint	TOP 6	TOP 1	FAIL
What format was the major competition of VHS	TOP 10	TOP 7	TOP 6
What instrument is Ray Charles best known for playing	FAIL	FAIL	TOP 6
What is a ballet company Mikhail Baryshnikov has danced for	TOP 1	TOP 1	FAIL
What is a film in which Jude_Law acted	TOP 9	TOP 3	TOP 9
What is a tiger that is extinct	TOP 3	TOP 1	TOP 3
What is Martin Luther King Jr real birthday	TOP 9	TOP 2	TOP 10
What is of the major gods of Hinduism	TOP 3	TOP 2	TOP 1
What is one of the Seven Wonders of the Ancient World	FAIL	TOP 2	TOP 1
What is Pittsburg baseball team called	FAIL	TOP 2	TOP 1
What is the cultural origin of the ceremony of potlatch	FAIL	TOP 4	FAIL

Question	ASK.COM	GOOGLE	YAHOO
What is the daily requirement of folic acid for an expectant mother	FAIL	TOP 1	FAIL
What is the equivalent of the Red_Cross in the Middle East	TOP 1	FAIL	FAIL
What is the hair style called that new military recruits receive	FAIL	TOP 1	FAIL
What is the most heavily caffeinated soft drink	FAIL	TOP 5	FAIL
What is the name given to a group of geese	FAIL	TOP 1	FAIL
What is the name given to the Islamic counterpart of the Red_Cross	TOP 1	TOP 2	FAIL
What is the name of the actress who starred in the movie Silence o	TOP 1	TOP 2	FAIL
What is the name of the art of growing miniature trees	TOP 1	TOP 2	TOP 6
What is the name of the Islamic counterpart to the Red_Cross	TOP 1	TOP 2	FAIL
What is the name of the Lion King son in the movie The Lion King	FAIL	TOP 3	FAIL
What is the name of the Michelangelo painting that shows two hand	FAIL	FAIL	FAIL
What is the name of the song Will Smith sings about parents	FAIL	TOP 2	FAIL
What is the name of the star of the cooking show Galloping Gourm	TOP 1	TOP 4	TOP 9
What is the name of the tallest mountain in the world	TOP 1	FAIL	FAIL
What is the normal resting heart rate of a healthy adult	TOP 2	TOP 1	TOP 1
What is the recommended daily requirement for folic acid for pregn	TOP 5	TOP 3	TOP 3
What is the university of which Woodrow_Wilson was president	TOP 5	TOP 2	TOP 1
What is the US location of Procter & Gamble corporate offices	TOP 2	TOP 2	TOP 5
What is worth seeing in Reims	FAIL	TOP 3	FAIL
What made Jane Goodall famous	TOP 5	TOP 5	TOP 3
What part of your body contains the corpus callosum	FAIL	TOP 3	TOP 3
What President became Chief Justice after his presidency	TOP 8	TOP 1	TOP 1
What radio station did Paul Harvey work for	FAIL	TOP 7	TOP 3
What soft drink is most heavily caffeinated	FAIL	TOP 10	FAIL
What soft drink would provide me with the biggest intake of caffein	FAIL	TOP 10	FAIL
What sport do the Cleaveland Cavaliers play	FAIL	TOP 2	TOP 5
What state in the United States covers the largest area	TOP 4	TOP 2	TOP 2
What task does the Bouvier breed of dog perform	FAIL	TOP 8	FAIL
What was the alternate to VHS	FAIL	FAIL	FAIL
What was Darth Vader son named	TOP 4	TOP 10	FAIL
What video format was an alternative to VHS	TOP 2	TOP 1	FAIL
What US Government agency registers trademarks	TOP 1	TOP 4	TOP 1
What type of hunting are retrievers used for	FAIL	TOP 2	TOP 1
What type of horses appear on the Budweiser commercials	FAIL	TOP 1	TOP 2
What time of year is air travel the heaviest	FAIL	TOP 3	TOP 5
What time of year is air travel at a peak	FAIL	FAIL	FAIL
What time of year has the most air travel	FAIL	FAIL	TOP 1
What time of year do most people fly	FAIL	FAIL	FAIL
What year was the first automobile manufactured	TOP 7	TOP 3	FAIL
What year was the Avery Dennison company founded	FAIL	TOP 2	TOP 5
What year was Janet Jackson first album released	FAIL	TOP 4	TOP 1

Question	ASK.COM	GOOGLE	YAHOO
What year was Desmond Mpilo Tutu awarded the Nobel Peace Prize	TOP 1	TOP 1	TOP 3
What year did the Vietnam War end	TOP 2	TOP 3	TOP 4
What wrestling star became The Incredible Hulk	FAIL	FAIL	TOP 4
What was the name of the television show staring Karl Malden that had S	TOP 10	TOP 2	FAIL
What was the name of the sitcom that Alyssa Milano starred in with Tony	TOP 8	TOP 2	TOP 1
What was the name of the first Russian astronaut to do a spacewalk	TOP 2	FAIL	TOP 4
What are chloroplasts	TOP 3	TOP 1	TOP 1
What are geckos	TOP 1	TOP 1	TOP 1
What are Pomegranate	TOP 1	TOP 1	TOP 1
What are the Poconos	TOP 1	TOP 1	TOP 1
What city is Logan Airport in	TOP 1	TOP 3	TOP 2
What does laser stand for	TOP 6	TOP 3	TOP 5
What does NASA stand for	TOP 3	TOP 2	TOP 2
What is Anorexia_Nervosa	TOP 1	TOP 1	TOP 1
What is Black_Hills_South_Dakota famous for	TOP 9	TOP 1	TOP 1
What is California state bird	TOP 1	TOP 1	TOP 1
What is California state tree	TOP 3	TOP 4	TOP 1
What is Java	TOP 1	TOP 1	TOP 1
What is leukemia	TOP 3	TOP 2	TOP 1
What is Molybdenum	TOP 1	TOP 1	TOP 1
What is Nine_Inch_Nails	TOP 3	TOP 1	TOP 1
What is sake	FAIL	TOP 2	TOP 2
What is ouzo	TOP 1	TOP 1	TOP 1
What is saltpeter	TOP 5	TOP 1	TOP 1
What is TCI	TOP 1	TOP 1	TOP 2
What is the population of Mississippi	TOP 1	TOP 3	TOP 1
What is typhoid fever	TOP 1	TOP 1	TOP 1
What king signed the Magna Carta	TOP 1	TOP 2	TOP 3
What monarch signed the Magna Carta	TOP 2	TOP 3	TOP 2
What nationality was Jackson Pollock	TOP 2	TOP 9	TOP 1

15.4.6 WHAT Results For Parse Successes

Question	ASK.COM	GOOGLE	YAHOO
What actor first portrayed James Bond	TOP 1	TOP 2	TOP 1
What actress starred in The Lion in Winter	TOP 1	TOP 5	TOP 3
What are the names of Jacques Cousteau two sons	TOP 2	TOP 6	TOP 5
What attracts tourists to Reims	FAIL	TOP 8	TOP 7
What card company sells Christmas ornaments	FAIL	TOP 1	TOP 1
What city airport is named Logan International	TOP 2	TOP 1	TOP 1
What city houses the US headquarters of Procter and Gamble	TOP 1	TOP 3	TOP 3
What city is 945 KDGE Radio located in	TOP 7	TOP 3	TOP 1
What city is served by Logan International Airport	TOP 1	TOP 1	TOP 2
What city is served by McCarren Airport	TOP 9	TOP 2	TOP 5
What city is the Kentucky Horse Park near	TOP 1	TOP 1	TOP 2
What city the Kentucky Horse Park is near	TOP 1	TOP 1	TOP 1
What company produced rotary engine vehicles	TOP 9	TOP 2	TOP 1
What country are Godiva chocolates from	TOP 4	FAIL	FAIL
What Cruise Line does Kathie Lee Gifford advertise for	TOP 7	TOP 4	TOP 1
What day is known as the national day of prayer	TOP 1	TOP 4	TOP 1
What did Delilah do to Samson hair	TOP 1	TOP 1	TOP 1
What does caliente translate to in English	TOP 1	TOP 4	FAIL
What does Final Four refer to in the sports world	FAIL	TOP 1	TOP 2
What does Knight Ridder publish	FAIL	TOP 1	TOP 10
What does the acronym CPR mean	FAIL	TOP 5	TOP 5
What English word contains the most letters	TOP 1	FAIL	FAIL
What English word has the most letters	TOP 5	FAIL	FAIL
What ethnic group introduced the idea of potlatch	FAIL	FAIL	FAIL
What famous model was married to Billy Joel	FAIL	TOP 1	TOP 1
What flower did Vincent Van Gogh paint	TOP 6	TOP 1	FAIL
What format was the major competition of VHS	TOP 10	TOP 7	TOP 6
What instrument is Ray Charles best known for playing	FAIL	FAIL	TOP 6
What is a ballet company Mikhail Baryshnikov has danced for	TOP 1	TOP 1	FAIL
What is a film in which Jude_Law acted	TOP 9	TOP 3	TOP 9
What is a tiger that is extinct	TOP 3	TOP 1	TOP 3
What is Martin Luther King Jr real birthday	TOP 9	TOP 2	TOP 10
What is of the major gods of Hinduism	TOP 3	TOP 2	TOP 1
What is one of the Seven Wonders of the Ancient World	FAIL	TOP 2	TOP 1
What is Pittsburg baseball team called	FAIL	TOP 2	TOP 1
What is the cultural origin of the ceremony of potlatch	FAIL	TOP 4	FAIL

Question	ASK.COM	GOOGLE	YAHOO
Winnie the Pooh is an imitation of which animal?	FAIL	FAIL	FAIL
Winnie the Pooh is what kind of animal?	FAIL	TOP 4	FAIL
A corgi is a kind of what?	TOP 5	TOP 5	TOP 2
Aspartame is also called what?	TOP 1	FAIL	TOP 2
Aspartame is also known as what?	TOP 1	TOP 1	TOP 1
Aspartame is known by what other name?	FAIL	TOP 2	TOP 5
At what speed does the Earth revolve around the sun?	TOP 4	TOP 7	TOP 10
Jude Law acted in which film?	FAIL	TOP 4	TOP 6
Jude Law was in what movie?	TOP 1	TOP 9	TOP 5
Name a civil war battlefield.	FAIL	TOP 1	TOP 7
Name a female figure skater.	TOP 2	TOP 8	TOP 1
Name a flying mammal.	TOP 2	TOP 1	TOP 3
Name a Gaelic language.	TOP 1	TOP 1	TOP 1
Name a golf course in Myrtle Beach.	TOP 1	TOP 2	TOP 1
Name a movie that the actress, Sandra Bullock, had a role in.	TOP 11	TOP 1	TOP 2
Name a symptom of mononucleosis.	TOP 1	TOP 5	TOP 1
Name the creator of "The Muppets".	TOP 1	TOP 1	TOP 2
Name the first Russian astronaut to do a spacewalk.	FAIL	FAIL	FAIL
Name the highest mountain.	FAIL	FAIL	FAIL
Name the Islamic counterpart to the Red Cross.	TOP 3	TOP 2	FAIL
The Jewish alphabet is called what?	TOP 1	TOP 2	TOP 2
CNN is the abbreviation for what?	TOP 1	TOP 7	FAIL
CPR is the abbreviation for what?	TOP 2	TOP 1	TOP 5
For what disease is the drug Sinemet used as a treatment?	TOP 1	TOP 1	TOP 1
Hazmat stands for what?	TOP 2	TOP 1	FAIL
CNN is an acronym for what?	TOP 3	FAIL	TOP 9
What animal do buffalo wings come from?	TOP 9	TOP 5	TOP 2
What are birds descendents of?	TOP 1	TOP 1	TOP 1
What are chloroplasts?	TOP 1	TOP 1	TOP 1

Question	ASK.COM	GOOGLE	YAHOO
What are Cushman and Wakefield known for?	TOP 1	TOP 1	TOP 9
What are geckos?	TOP 1	TOP 2	TOP 1
What are pomegranates?	TOP 3	TOP 1	TOP 1
What are the birth dates for scorpios?	FAIL	TOP 1	FAIL
What are the Black Hills known for?	TOP 1	TOP 4	TOP 3
What are the names of the tourist attractions in Reims?	FAIL	FAIL	TOP 6
What are the Poconos?	TOP 1	TOP 1	TOP 1
What are tourist attractions in Reims?	TOP 4	TOP 5	FAIL
What breed of dog was the "Little Rascals" dog?	TOP 1	TOP 1	TOP 2
What can one see in Reims?	FAIL	TOP 7	TOP 5
What caused the death of Bob Marley?	TOP 4	TOP 3	TOP 1
What caused the Lynmouth floods?	TOP 2	TOP 10	TOP 2
What cereal goes "snap, crackle, pop"?	TOP 1	TOP 4	FAIL
What city did the Flintstones live in?	FAIL	TOP 3	TOP 1
What city does McCarren Airport serve?	TOP 9	TOP 2	FAIL
What city in Florida is Sea World in?	TOP 3	TOP 2	TOP 2
What city is Logan Airport in?	TOP 1	TOP 3	TOP 2
What city is Massachusetts General Hospital located in?	TOP 1	TOP 1	TOP 1
What city is the Orange Bowl in?	TOP 2	TOP 1	TOP 2
What company sells the most greeting cards?	FAIL	FAIL	FAIL

Question	ASK.COM	GOOGLE	YAHOO
What continent is Bolivia on?	TOP 1	TOP 1	TOP 1
What could I see in Reims?	FAIL	TOP 3	TOP 10
What culture developed the idea of potlatch?	FAIL	TOP 7	FAIL
What date is Boxing Day?	TOP 1	FAIL	TOP 2
What did brontosauruses eat?	TOP 1	TOP 1	TOP 2
What did Vasco da Gama discover?	TOP 1	TOP 6	TOP 3
What do ladybugs eat?	TOP 2	TOP 1	TOP 1
What do manatees eat?	TOP 3	TOP 3	TOP 3
What do most tourists visit in Reims?	FAIL	FAIL	TOP 2
What do penguins eat?	TOP 3	TOP 1	TOP 1
What do river otters eat?	TOP 3	TOP 1	TOP 1
What do the initials CPR stand for?	FAIL	FAIL	TOP 1
What do you call a group of geese?	TOP 1	TOP 1	TOP 2
What does caliente mean (in English)?	TOP 6	TOP 4	TOP 6
What does CNN stand for?	TOP 6	TOP 4	TOP 1
What does CPR stand for?	TOP 1	TOP 4	TOP 5
What does EKG stand for?	TOP 7	TOP 2	TOP 6
What does hazmat stand for?	TOP 4	TOP 4	FAIL
What does laser stand for?	TOP 6	TOP 4	TOP 6
What does NAFTA stand for?	TOP 3	TOP 2	TOP 1
What does NASA stand for?	TOP 3	TOP 2	TOP 2
What does Nicholas Cage do for a living?	TOP 1	TOP 4	TOP 10
What does SIDS stand for?	TOP 3	TOP 7	TOP 2
What does the abbreviation OAS stand for?	FAIL	TOP 5	FAIL
What effect does a prism have on light?	TOP 2	TOP 4	TOP 1
What film or films has Jude Law appeared in?	TOP 3	TOP 1	TOP 3
What film was Jude Law in?	TOP 3	TOP 6	TOP 3
What format was VHS's main competition?	TOP 1	TOP 6	FAIL
What hair color can I use to just cover a little gray?	FAIL	TOP 6	TOP 5
What hockey team did Wayne Gretzky play for?	TOP 10	FAIL	TOP 1
What instrument does Ray Charles play?	TOP 3	TOP 3	TOP 8
What is "Nine Inch Nails"?	TOP 1	TOP 1	TOP 1
What is a caldera?	TOP 1	TOP 1	TOP 1
What is a film starring Jude Law?	TOP 2	TOP 2	TOP 1
What is a meerkat?	TOP 1	TOP 1	TOP 2
What is a nanometer?	TOP 4	TOP 1	TOP 1
What is a nematode?	TOP 6	TOP 1	TOP 1
What is a nickname for Mississippi?	TOP 1	TOP 1	TOP 3
What is a stratocaster?	TOP 3	TOP 1	TOP 3
What is a synonym for aspartame?	TOP 1	TOP 3	TOP 3
What is Alice Cooper's real name?	TOP 3	TOP 3	TOP 5
What is anise?	TOP 1	TOP 1	TOP 1
What is anorexia nervosa?	TOP 1	TOP 1	TOP 1
What is another name for nearsightedness?	TOP 3	TOP 1	TOP 9
What is Archimedes famous for?	TOP 1	FAIL	FAIL
What is Betsy Ross famous for?	TOP 1	TOP 2	TOP 1
What is Black Hills, South Dakota most famous for?	TOP 7	TOP 2	TOP 2
What is California's capital?	TOP 1	TOP 2	TOP 6
What is California's state bird?	TOP 1	TOP 1	TOP 1
What is California's state tree?	TOP 1	TOP 7	TOP 8
What is Chiricahua the name of?	TOP 1	TOP 3	TOP 3
What is Colin Powell best known for?	TOP 1	TOP 1	TOP 10
What is cribbage?	TOP 5	TOP 1	TOP 1
What is D.B. Cooper known for?	TOP 1	TOP 2	TOP 1
What is Dick Clark's birthday?	TOP 1	TOP 1	TOP 2

Question	ASK.COM	GOOGLE	YAHOO
What is Dick Clark's date of birth?	TOP 1	TOP 1	TOP 1
What is Dr. Ruth's last name?	TOP 2	FAIL	TOP 2
What is Francis Scott Key best known for?	TOP 6	TOP 1	TOP 3
What is Giorgio Vasari famous for?	TOP 6	TOP 1	TOP 1
What is Jane Goodall famous for?	TOP 1	TOP 1	TOP 1
What is Jane Goodall known for?	TOP 3	TOP 1	TOP 1
What is Java?	TOP 6	TOP 1	TOP 2
What is leukemia?	TOP 3	TOP 1	TOP 2
What is measured in curies?	TOP 1	TOP 5	TOP 4
What is molybdenum?	TOP 1	TOP 1	TOP 1
What is Nicholas Cage's occupation?	FAIL	TOP 2	FAIL
What is Nicholas Cage's profession?	TOP 8	TOP 3	TOP 10
What is one of the cities that the University of Minnesota is located in?	TOP 1	FAIL	TOP 5
What is ouzo?	TOP 1	TOP 1	TOP 1
What is pandoro?	TOP 5	TOP 1	TOP 1
What is platinum?	TOP 1	TOP 1	TOP 1
What is porphyria?	TOP 1	TOP 1	TOP 1
What is sake?	FAIL	TOP 2	FAIL
What is saltpeter?	TOP 7	TOP 1	TOP 1
What is TCI?	TOP 1	FAIL	FAIL
What is thalassemia?	TOP 6	TOP 1	TOP 1
What is the abbreviation for Original Equipment Manufacturer?	TOP 1	TOP 1	TOP 4
What is the airport code for Los Angeles International?	TOP 1	TOP 1	TOP 1
What is the busiest air travel season?	TOP 4	TOP 1	TOP 2
What is the capital of Burkina Faso?	TOP 1	TOP 1	TOP 1
What is the capital of Haiti?	TOP 1	TOP 1	TOP 1
What is the chemical formula/name for napalm?	FAIL	FAIL	FAIL
What is the chemical symbol for nitrogen?	TOP 1	TOP 1	TOP 2
What is the collective noun for geese?	TOP 8	TOP 6	TOP 6
What is the collective term for geese?	FAIL	TOP 4	FAIL
What is the date of Bastille Day?	TOP 1	TOP 1	TOP 1
What is the date of Boxing Day?	TOP 1	TOP 1	TOP 1
What is the definition of hazmat?	TOP 1	TOP 1	TOP 3
What is the English meaning of caliente?	TOP 4	TOP 2	FAIL
What is the English translation for the word "caliente"?	FAIL	TOP 6	FAIL
What is the exchange rate between England and the U.S.?	FAIL	FAIL	TOP 5
What is the federal minimum wage?	FAIL	TOP 1	TOP 6
What is the gestation period for human pregnancies?	TOP 7	TOP 4	TOP 2
What is the gestation period for humans?	TOP 1	TOP 1	TOP 2
What is the habitat of the chickadee?	TOP 1	TOP 1	TOP 1
What is the highest mountain in the world?	FAIL	TOP 6	TOP 3
What is the Islamic counterpart to the Red Cross?	TOP 4	TOP 2	TOP 9
What is the Islamic equivalent of the Red Cross?	TOP 1	TOP 5	TOP 2
What is the Jewish alphabet called?	TOP 1	TOP 4	FAIL
What is the largest snake in the world?	TOP 1	TOP 1	TOP 2
What is the life expectancy of an elephant?	TOP 1	TOP 1	TOP 4
What is the location of McCarren Airport?	TOP 1	TOP 2	TOP 1
What is the location of Rider College?	TOP 7	TOP 9	TOP 1
What is the longest English word?	FAIL	FAIL	FAIL

Question	ASK.COM	GOOGLE	YAHOO
What is the meaning of caliente (in English)?	TOP 1	TOP 4	FAIL
What is the meaning of thalassemia?	TOP 1	TOP 1	FAIL
What is the medical condition of hypertension?	TOP 3	TOP 1	TOP 1
What is the most common cancer?	FAIL	TOP 1	TOP 2
What is the most common kind of skin cancer in the U.S.?	TOP 1	TOP 1	TOP 1
What is the most expensive car in the world?	TOP 5	TOP 9	TOP 5
What is the name for clouds that produce rain?	TOP 4	TOP 1	TOP 6
What is the name of a Greek god?	TOP 5	TOP 4	FAIL
What is the name of a Salt Lake City newspaper?	TOP 1	TOP 1	TOP 1
What is the name of Joan Jett's band?	TOP 1	TOP 1	TOP 3
What is the name of the company that manufactures the "American Girl"	TOP 1	TOP 1	TOP 1
What is the name of the inventor of silly putty?	TOP 1	TOP 1	TOP 2
What is the name of the Jewish alphabet?	TOP 9	FAIL	FAIL
What is the name of the longest ruling dynasty of Japan?	FAIL	FAIL	FAIL
What is the name of the second space shuttle?	TOP 1	TOP 4	TOP 9
What is the name of the vaccine for chicken pox?	TOP 5	TOP 1	TOP 7
What is the nickname for the state of Mississippi?	TOP 1	TOP 1	TOP 3
What is the nickname of Pennsylvania?	TOP 1	TOP 4	TOP 4
What is the nickname of the state of Mississippi?	TOP 1	TOP 3	TOP 3
What is the occupation of Nicholas Cage?	TOP 1	TOP 5	TOP 8
What is the Pennsylvania state income tax rate?	FAIL	TOP 1	TOP 1
What is the population of Japan?	TOP 1	TOP 3	TOP 1
What is the population of Kansas?	TOP 1	TOP 2	TOP 1
What is the population of Mexico?	TOP 1	TOP 1	TOP 1
What is the population of Mozambique?	TOP 1	TOP 2	TOP 2
What is the population of Ohio?	TOP 1	TOP 1	TOP 1
What is the population of the Bahamas?	TOP 5	TOP 1	TOP 2
What is the population of the United States?	TOP 1	FAIL	TOP 1
What is the primary language of the Philippines?	TOP 3	TOP 1	TOP 1
What is the purpose of a car bra?	TOP 2	TOP 1	TOP 3
What is the real name of the singer, Madonna?	FAIL	TOP 3	TOP 6
What is the salary of a U.S. Representative?	FAIL	TOP 1	TOP 1
What is the size of Argentina?	TOP 2	TOP 2	FAIL
What is the state nickname of Mississippi?	TOP 1	TOP 9	TOP 1
What is the tallest mountain?	TOP 2	TOP 2	TOP 7
What is the telephone number for the University of Kentucky?	TOP 1	TOP 1	FAIL
What is the wingspan of a condor?	TOP 3	TOP 1	TOP 2
What is the world's highest peak?	FAIL	TOP 2	TOP 6
What is the zip code for Fremont, CA?	TOP 3	TOP 1	TOP 1
What is the zip code for Parsippany, NJ?	TOP 1	TOP 1	TOP 1
What is titanium?	TOP 1	TOP 1	TOP 1
What is typhoid fever?	TOP 1	TOP 1	TOP 1
What is tyvek?	TOP 1	TOP 1	TOP 1
What is witch hazel?	TOP 1	TOP 1	TOP 1

Question	ASK.COM	GOOGLE	YAHOO
What king signed the Magna Carta?	TOP 1	TOP 1	TOP 1
What king was forced to agree to the Magna Carta?	TOP 1	TOP 1	TOP 1
What language is mostly spoken in Brazil?	TOP 3	TOP 1	TOP 2
What makes Black Hills, South Dakota a tourist attraction?	TOP 1	FAIL	FAIL
What monarch signed the Magna Carta?	FAIL	TOP 7	TOP 1
What movie did Madilyn Kahn star in with Gene Wilder?	FAIL	TOP 5	FAIL
What nationality was Jackson Pollock?	TOP 1	TOP 10	TOP 1
What newspaper serves Salt Lake City?	TOP 1	TOP 4	TOP 1
What ocean did the Titanic sink in?	TOP 8	TOP 1	TOP 2
What other name were the "Little Rascals" known as?	TOP 1	TOP 1	TOP 1
What party was Winston Churchill a member of?	TOP 1	TOP 3	TOP 2
What province is Edmonton located in?	TOP 1	TOP 3	TOP 4
What soft drink contains the largest amount of caffeine?	FAIL	FAIL	FAIL
What species was Winnie the Pooh?	FAIL	FAIL	FAIL
What state does Martha Stewart live in?	FAIL	TOP 7	FAIL
What state does MO stand for?	TOP 4	TOP 2	TOP 3
What state has the most Indians?	FAIL	FAIL	FAIL
What state is Niagra Falls located in?	TOP 2	TOP 1	TOP 2
What state is the Filenes store located in?	FAIL	FAIL	FAIL
What state produces the best lobster to eat?	FAIL	TOP 6	TOP 2
What store does Martha Stewart advertise for?	TOP 2	TOP 4	FAIL
What tourist attractions are there in Reims?	FAIL	FAIL	FAIL
What type of bridge is the Golden Gate Bridge?	TOP 10	TOP 3	TOP 4
What university was Woodrow Wilson President of?	TOP 4	TOP 1	TOP 6
What was Poe's birthplace?	TOP 6	TOP 1	TOP 6
What was the ball game of ancient Mayans called?	TOP 1	TOP 1	FAIL
What was the birthplace of Edgar Allen Poe?	TOP 1	TOP 3	TOP 1
What was the cause of Bob Marley's death?	FAIL	FAIL	TOP 6
What was the date of CNN's first broadcast?	FAIL	FAIL	FAIL
What was the death toll at the eruption of Mount Pinatubo?	TOP 4	TOP 10	FAIL
What was the distinguishing mark on the "Little Rascals" dog?	FAIL	TOP 5	FAIL
What was the man's name who was killed in a duel with Aaron Burr?	TOP 8	TOP 1	TOP 2
What was the name of "The Muppets" creator?	TOP 1	TOP 1	TOP 5
What was the name of Darth Vader's son?	TOP 2	TOP 4	TOP 5
What was the name of Jacques Cousteau's ship?	TOP 2	TOP 2	TOP 1
What was the name of the "Little Rascals" dog?	TOP 1	TOP 1	TOP 5
What was the name of the famous battle in 1836 between Texas and Mex	TOP 1	TOP 1	TOP 2
What was the name of the movie that starred Sharon Stone and Arnold S	TOP 2	TOP 1	TOP 1
What was the name of the Titanic's captain?	TOP 1	TOP 1	TOP 3
What was the nationality of Jackson Pollock?	TOP 1	TOP 3	TOP 1
What was the purpose of the Manhattan project?	TOP 1	TOP 2	TOP 3
What was the species of Winnie the Pooh?	FAIL	FAIL	FAIL
What were the names of the three ships used by Columbus?	TOP 4	TOP 8	FAIL
What year did Hitler die?	TOP 1	TOP 3	FAIL
What year did Montana become a state?	TOP 1	TOP 3	FAIL

Question	ASK.COM	GOOGLE	YAHOO
What's the average salary of a professional baseball player?	TOP 3	TOP 3	FAIL
What's the farthest planet from the sun?	TOP 2	TOP 1	TOP 1
What's the formal name for Lou Gehrig's disease?	TOP 2	TOP 5	TOP 4
What's the longest river in the world?	TOP 1	TOP 1	TOP 2
What's the most famous tourist attraction in Rome?	TOP 6	TOP 1	FAIL
What's the name of a golf course in Myrtle Beach?	TOP 3	TOP 1	TOP 1
What's the name of Pittsburgh's baseball team?	TOP 4	TOP 1	TOP 3
What's the name of the Tampa newspaper?	TOP 1	TOP 3	TOP 4
What's the name of the tiger that advertises for Frosted Flakes cereal?	TOP 10	TOP 1	TOP 1
What's the name of the Tokyo Stock Exchange?	FAIL	TOP 5	TOP 1
What's the population of Biloxi, Mississippi?	TOP 1	TOP 2	TOP 1
What's the population of Mississippi?	FAIL	FAIL	TOP 10
What's the tallest building in New York City?	TOP 4	TOP 2	TOP 1

16 Bibliography

Allen, James (1995). *Natural Language Understanding.* The Benjamin/Cummings Publishing Company, Inc.

Alpha, Shamim; Dixon, Paul; Liao, Ciya; Yang, Changwen (2001). Oracle at TREC 10: Filtering and Question Answering. *Proceedings of The 10th Conference on Text Retrieval (2001)*.

Argamon-Engelson, Shlomo; Dagan, Ido; Krymolowski, Yuval (1998). A Memory Based Approach to Learning Shallow Natural Language Patterns. *COLING ACL*, 67-73, Montreal Canada 1998.

Azzam, Saliha; Humphreys, Kevin; Gaizauskas, Robert (1998). Evaluating a Focus-Based Approach to Anaphora Resolution. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98*, 74-78, Association or Computational Linguistics.

Beckwith, Richard (1993). *Design and Implementation of the WordNet Lexical Database and Searching Software.* (From the WordNet home page).

Bikel, Daniel M.; Richard Schwartz; Weischedel, Ralph (1999). *An Algorithm that Learns What's in a Name.* Machine Learning (1999) *34,* 211-231.

Brill, Eric (1994). Some Advances in Information-Based Part Of Speech Tagging. *AAAI* (1994).

Brill, Eric; Lin, Jimmy; Banko, Michael; Dumais, Susan; Ng, Andrew (2001). Data-Intensive Question Answering. *Proceedings of the 10th Conference on Text REtrieval (2001)*.

Brundige, Ellen N (not dated). *The Legend of the Library*. Retrieved June 18, 2001, from http://www.perseus.tufts.edu/GreekScience/Students/Ellen/Museum.html.

Buchholz, Sabine (2001). Using Grammatical Relations, Answer Frequencies and the World Wide Web for TREC Question Answering. *Proceedings of the 10th Conference on Text REtrieval (2001).*

Buckley, Chris; Allan, James; Salton, Gerard (1994) . Automatic Routing and Ad-hoc Retrieval Using SMART: TREC 2. *Proceedings of the Second Text REtrieval Conference (TREC-2 1994)*.

Buckley, Chris; Mandar Mitra; Walz, Janet; Cardie, Claire (1998). *SMART High Precision: TREC 7 . Proceedings of the Seventh Text REtrieval Conference (TREC-7, 1998)*.

Burger, John D (2006). MITRE's Qanda at TREC-15. *Proceedings of the Fifteenth Text*

REtrieval Conference (TREC-15, 2006)

Burke, Robin D.; Hammond, Kristian J.; Kulyukin, Vladimir, A.; Lytinen, Steven L.; Tomuro, Noriko; Schoenberg, Scott (1997). Question Answering from Frequently-Asked Question Files: Experiences with the FAQ Finder System. *AI* (1997).

Cardie, Claire (1996). Automating Feature Selection for Case-Based Learning of Linguistic Knowledge. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 113-126 University of Pennsylvania, (1996).

Cardie, Claire; Wagstaff, Kiri (1999). Noun Phrase Coreference as Clustering. *Proceedings of the Joint SIG DAT Conference on Empirical Methods in Natural- Language Processing and Very Large Corpora,* 82-89, Association for Computational Linguistics, (1999).

Chen, Jiangping; Diekema, Anne R.; Taffet, Mary D.; McCracken, Nancy; Ozgencil, Necati Ercan; Yilmazel, Ozgur; Liddy, Elizabeth D. (2001). Question Answering: CNLP at the TREC-10 Question Answering Track. *Proceedings of the Tenth Text Retrieval Conference (TREC-10, 2001*).

Cohen, Paul; Schrag, Robert; Jones, Eric; Peese, Adam; Lin, Albert; Starr, Barbara; Gunning, David; Burke, Murray (1998). The Darpa High-Performance Knowledge Bases Project. *AI Magazine*, *19(*4).

Cohen, Paul; Chaudhri, Vinay; Pease, Adam; Schrag, Robert (1999). Does Prior Knowledge Facilitate the Development of Knowledge-based Systems? *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-1999).*

Colins, Michael (1997). Three Generative, Lexicalized Models for Statistical Parsing . *Proceedings of the 35th Annual Meeting of the ACL (jointly with the 8th Conference of the ACL) (1997).*

Dagan, Ido; Lee, Lillian; Pereira, Fernando (1997). Similarity-Based Methods For Word-Sense Disambiguation. *Proceedings of The 35th Annual Meeting of The Association of Computational Linguistics and 8th Conference of The European Chapter of The Association For Computational Linguistics (1997).*

Dai, H., Luo, T., Sung Y., Zhu, J. (2000). *Integrating Web Usage and Content Mining for More Effective Personalization. Proceedings of the International Conference on E-Commerce and Web Technologies (ECWeb2000)*, September 2000, Greenwich, UK.

Dudani, S. (1976). The Distance-Weighted K-Nearest-Neighbor Rule. *IEEE Transactions on Systems - Man and Cybernetics SMC, 6*(4), 325-327.

Escudero, G.; Marquez, L. ; Rigau, G. (2000). Boosting Applied to Word Sense Disambiguation . *Proceedings of the 11th European Conference on Machine Learning (ECML-00).*

Fellbaum, Christiane (1993). English Verbs as a Semantic Net (From the WordNet home page) Update of original paper (1990). *International Journal of Lexicography, 3*(4), 278-301.

Ferret, O.; Grau, B; Hurault-Planter, M.; Illouz, G; Monceaux, L; Robba, I; Vilnat, A. (2001). Finding An Answer Based On The Recognition Of The Question Focus. *Proceedings of The 10th Conference on Text Retrieval (2001).*

Gaizauskas, Mark A. Greenwood, Henk Harkema, Mark Hepple, Horacio Saggi on, Atheesh Sanka (2005). *The University of Sheffield's TREC 2005 Q&A Experiments (2005)*

Glatthorn, Allan (1998). *Writing the Winning Dissertation*. Corwin Press Inc.

Harabagiu, Sanda; Moldovan, Dan (1995). A Marker-Propagation Algorithm for Text Coherence . *Proceedings of the Workshop on Parallel Processing in Artificial Intelligence*, *IJCAI-91 (1995).*

Harabagiu, Sanda; Moldovan, Dan; Pasca, Marius; Surdeanu, Mihai; Mihalcea, Rada; Girju, Roxana; Rus, Vasile; Lacatusu, Finley; Mararescu, Paul; Bunescu, Razvan (2001). Answering Complex, List and Context Questions With LCC's Question-Answering Server. *Proceedings of the Tenth Conference on Text Retrieval (2001).*

Hovy, E.H., L. Gerber, U. Hermjakob, M. Junk, and C.-Y. Lin. 2001. Question Answering in Webclopedia. *Proceedings of the TREC-9 Conference. NIST, Gaithersburg, MD.*

Hovy, E.H., L. Gerber, U. Hermjakob, C.-Y. Lin, and D. Ravichandran. (2001). Toward Semantics-Based Answer Pinpointing. *Proceedings of the DARPA Human Language Technology Conference (HLT)*.

Ravichandran, D. and E.H. Hovy. (2002). Learning Surface Text Patterns for a Question Answering System. *Proceedings of the 40th ACL conference. Philadelphia, PA*.

Ahn, Kisuh; Bos, Johan; Curran R., James; Kor, Dave; Nissim, Malvina; Webber, Bonnie. (2005). Question Answering with QED at TREC-2005. *Proceedings of the Fourteenth Conference on Text Retrieval (TREC 14).*

Huyck, Christian; Lytinen, Steven (1993). Efficient Heuristic Natural Language Parsing. *AAAI (1993).*

Ittycheriah, Abraham; Franz, Martin; Roukos, Salim (2001). IBM's Statistical Question Answering System. *Proceedings of the 10th Conference on Text Retrieval (2001).*

Jurafsky, Daniel; Martin, James H. (2000). *Speech and Language Processing.* Prentice Hall Incorporated.

Khan, L., D. McLeod, and E.H. Hovy. (2004). Retrieval Effectiveness of an Ontology-Based Model for Information Selection. *Journal for Very Large Data Bases (VLDB). 13*(1), 71–85.

Katz, Boris (1997). From Sentence Processing to Information Access on the World Wide Web. *AAAI Spring Symposium.*

Korfhage, Robert R. (1997). *Information Storage and Retrieval*. John Wiley & Sons, Inc. (1997).

Kulyukin, Vladimir (1999). Application-Embedded Retrieval from Distributed Free-Text Collections. *AAAI (1999).*

Lee, Gary Geunbae; Seo, Jungyun; Lee, Seungwoo; Jung, Hanmin; Cho, Bong-Hyun; Lee, Changki; Kwak, Byung-Kwan; Char, Jeongwon; Kim, Dongseok; An, JooHui; Kim, Harksoo; Kim, Kyungsun (2001). SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP. *Proceedings of the 10th Conference on Text Retrieval (2001).*

Leidner , Jochen; Bos L., Johan; Dalmas, Tiphaine; Curran, James R.; Clark, Stephen; Bannard, Colin J.; Steedman, Mark; Weber, Bonnie (2003). The QED open-domain answer retrieval system for TREC 2003. *In Proceedings of the Twelfth Text Retrieval Conference (TREC 2003) , NIST Special Publication 500-255, pages 595–599, Gaithersburg, MD*.

Leong, H., Kapur and de Vel, O (1997). Text Summarization for Knowledge Filtering Agents in Distributed Heterogeneous Environments. *AAAI Spring Symposium.*

Lewis, David D.; Ringuette, Marc (1994). A Comparison of Two Learning Algorithms for Text Categorization. *Proceedings of the 3rd Annual Symposium on Document Analysis and Information Retrieval (1994).*

Litman, Diane (1994). Classifying Cue Phrases in Text and Speech Using Machine Learning. *Proceedings of the 12th National Conference on Artificial Intelligence, AAAI (1994).*

Lytinen, Steven L; Tomuro, Noriko (1996). Left Corner Unification-based Natural Language Processing. *AAAI/IAAI Vol. 2 1996.*

Lytinen, Steven L.; Tomuro, Noriko; Rapede, Tom (2000). The Use of WordNet Sense Tagging in FAQFinder. *AAAI-2000 Workshop on AI and Web Search (2000).*

Lytinen, Steven L.; Tomuro, Noriko (2002). The Use of Question Types to Match Questions in FAQFinder. *AAAI-2002 Spring Symposium on Mining Answers From Text (2002).*

Marquez, L.; Rodriguez, H.; Carmona, J.; Montolio, J. (1999). *Improving POS Tagging Using Machine Learning Techniques. Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora EMNLP/VLC (1999).*

Mihalcea, Rada; Moldovan, Dan I., (1998). Word Sense Disambiguation based on Semantic Density. *COLING/ACL (1998).*

Mihalcea, Rada; Moldovan, Dan I., (1999). An Automatic Method for Generating Tagged Corpora. *AAAI (1999).*

Miller, George A. (1993). Nouns in WordNet: A Lexical Inheritance System. From the WordNet home page. Update of original paper (1990). *International Journal of Lexicography, 3(4), 245-264.*

Miller, George A.; Beckwith, Richard; Fellbaum, Christiane; Gross, Derek; and Miller, Katherine A. (1993). *Introduction to WordNet: An On-line Lexical Database.* From the WordNet home page. Update or original paper (1990) *International Journal of Lexicography, 3*(4), 235-244.

Mitchell, Tom M. (1997). *Machine Learning.* WCB/McGraw-Hill.

Mlynarczyk, Stanley J.; Lytinen Steven L. (2005). FaqFinder Question Answering Improvements Using Question/Answer Matching. *Proceedings of the 2nd Language and Technology Conference (2005).*

Moldovan, Dan I; Mihalcea, Rada (1999). *Improving the search on the Internet by using WordNet and lexical operators*. Unpublished (1999).

Moldovan, Dan; Harabagiu, Sanda; Pasca, Marius; Mihalcea, Rada; Goodrum, Richard; Girju, Roxana; Rus, Vasile. (2000). LASSO: A tool for surfing the answer set. *Proceedings of the 9th Conference on Text Retrieval (2000).*

Montemagni, Simonetta and Vitto Pirelli (1998). Augmenting WordNet-like lexical resources with distributional evidence. An application-oriented perspective. *Proceedings of the COLING/ ACL Workshop on Usage of WordNet in Natural Language Processing Systems.*

MUC-6 (1995). The Sixth Message Understanding Conference. http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html.

Niemann, Michael (1998). Determining PP Attachment through Semantic Associations and Preferences. *Proceedings of CoNLL-98, Conference on Computational Language Learning Jan 22-24, 1998 Sydney, Australia,* http://www.uid.ac.be/conll98/proceedings.html.

Neyhart, David & Karper Erin (2001). Using American Psychology Association (APA) Format (Updated to 5th Edition). Retrieved March 10th, 2006, from http://owl.english.purdue.edu/handouts/research/r_apa.html.

Quinlan, J. R. (1992). MiniBoosting Decision Trees. *Journal of Artificial Intelligence Research (1998).*

Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research 4 (1996).*

Quinlan, J. R. (1996). Bagging, Boosting and C4.5. *AAAI (1996).*

Resnik, Philip (1993). *Using Information Content to Evaluate Semantic Similarity in a Taxonomy. Proceedings of LJCAI (1995).*

Roth, Dan; Kao, Gio Kao; Li, Xin; Nagarajan, Ramya; Punyakanok, Vasin; Rizzolo, Nick; Yih, Wen-Tau; Alm, Cecilia Ovesdotter, Moran, Liam Gerard (2001). Learning Components for A Question-Answering System. *Proceedings of The 10th Conference on Text Retrieval (TREC-10, 2001).*

Resnik, Philip (1999). Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence - 1999*, *11*, 95-130

Russell, Stuart; Norvig, Peter (1995). *Artificial Intelligence A Modern Approach.* Prentice Hall – (1995).
Roussinov, Robles-Flores, Ding (2004). *Experiments with WEB QA System and TREC 2004 Questions (TREC 2004)*.

Sanderson, Mark (1994). Word Sense Disambiguation and Information Retrieval. *17th ACM International Conference on Research and Development in Information Retrieval (SIGIR-94).*

Saxena, Ashish Kumar; Sambhu, Ganesh Viswanath; Subramaniam, L. Venkata; Kaushik, Saroj. (2007). IITD-IBMIRL System for Question Answering using Pattern Matching, Semantic Type and Semantic Category Recognition (2007). *Proceedings of The 16th Conference on Text Retrieval (TREC-16).*

Schone, Ciany, McNamee, Mayfield, Bassi, Kulman (2004). *Question Answering with QACTIS at TREC-2004 (TREC 2004).*

Schone, Ciany, Cutts, McNamee, Mayfield, Smith (2005). *Question Answering with QACTIS at TREC-2004 (TREC 2005).*

Siegel, Eric V. (1997). Learning Methods for Combining Linguistic Indicators to Classify Verbs . *Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing, EMNLP (1997).*

Stetina, Jiri; Kurohashi, Sadao; Nagao, Makoto (1998). General Word Sense Disambiguation Method Based on a Full Sentential Context. *Workshop COLING/ACL (1998).*

Sussna, (1993). Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic

Network. *Proceedings of the Second International Conference of Information and Knowledge Management (CIKM-93).*

Tomita, Masaru (1991). Generalized LR Parsing. *Kluwer Academic Publishers.*

Tomuro, Noriko (1989). Semi-automatic Induction of Systematic Polysemy from WordNet. *17th International Conference on Computational Linguistics (COLING '98).*

Tomuro, Noriko; Lytinen, Steven (2001). Selecting Features for Paraphrasing Question Sentences. *NLPRS 2001.*

Tomuro, Noriko (2002). Question Terminology and Representation for Question Type Classification. *19th International Conference on Computational Linguistics (COLING '02).*

Toutanova, Kristina; Manning, Christopher D. (2001). Feature Selection for a Rich HPSG Grammar Using Decision Trees. *CoNLL-2001.*

TREC-10 (2001). NIST Special Publication 500-250: The Tenth Text REtrieval Conference (TREC 2001) – http://trec.nist.gov/pubs.html.