Chapter 7

A Data Driven Approach to Interactive Question Answering

Sharon Small, Tomek Strzalkowski, Ting Liu, Nobuyuki Shimizu
ILS Institute, State University of New York at Albany
1400 Washington Avenue, Albany, NY 12222
{small,tl7612,ns3203,tomek}@albany.edu

Boris Yamrom
Lehman College of the City University of New York
Bronx, NY 10468
byamrom@lehman.cuny.edu

Abstract

HITIQA is an interactive question answering technology designed to allow intelligence analysts and other users of information systems to pose questions in natural language and obtain relevant answers, or the assistance they require in order to perform their tasks. Our objective in HITIQA is to allow the user to submit exploratory, analytical questions, such as "What has been Russia's reaction to U.S. bombing of Kosovo?" The distinguishing property of such questions is that one cannot generally anticipate what might constitute the answer. While certain types of things may be expected (e.g., diplomatic statements), the answer is heavily conditioned by what information is in fact available on the topic. From a practical viewpoint, analytical questions are often underspecified, thus casting a broad net on a space of possible answers. Therefore, clarification dialogue is often needed to negotiate with the user the exact scope and intent of the question.

1. Introduction

HITIQA project is part of the ARDA AQUAINT program that aims to make significant advances in the state of the art of automated question answering. In this paper we focus on two aspects of our work:

1. Question Semantics: how the system "understands" user requests.
2. Human-Computer Dialogue: how the user and the system negotiate this understanding.

We will also discuss very preliminary evaluation results from a series of pilot tests of the system conducted by intelligence analysts via a remote internet link.

2. Factoid vs. Analytical

The objective in HITIQA is to allow the user to submit and obtain answers to exploratory, analytical questions. There are very significant differences between factoid or fact-finding, and analytical question answering. A factoid question seeks pieces of information that would make a corresponding statement true (i.e., they become facts): "How many states are in the U.S.?" / "There are X states in the U.S." In this sense, a factoid question usually has just one correct answer that can generally, be judged for its truthfulness. By contrast, an analytical question is when the "truth" of the answer is more a matter of opinion and may depend upon the context in which the question is asked. Answers to analytical questions are rarely unilateral, indeed, a mere "correct" answer may have limited value, and in some cases may not even be determinate ("Which college is the best?", "How do I stop my baby's crying?"). Instead, answers to analytical questions are often judged as helpful, or useful, or satisfactory, etc. "Technically correct" answers (e.g., "feed the baby milk") may be considered as irrelevant or at best unresponsive.

Factoid questions display a fairly distinctive "answer type", which is the type of the information piece needed to fulfill the statement. Recent automated systems for answering factoid questions deduct this expected answer type from the form of the question and a finite list of possible answer types. For example, "Who was the first man in space" expects a "person" as the answer, while "How long was the Titanic?" expects some length measure as an answer, probably in yards and feet, or meters. This is generally a very good strategy, which has been exploited successfully in a number of automated QA systems that appeared in recent years, especially in the context of TREC QA[1] evaluations (Harabagiu et al., 2000; Hovy et al., 2000; Prager at al., 2001).

The above process is not easily applied to analytical questions. This is because the type of an answer for analytical questions cannot always be anticipated due to their inherently exploratory character. In contrast to a factoid question, an analytical question has an unlimited variety of syntactic forms with only a loose connection between their syntax and the expected answer. Given the unlimited

---

[1] TREC QA is the annual Question Answering evaluation sponsored by the U.S. National Institute of Standards and Technology www.trec.nist.gov.

potential for the formation of analytical questions, it would be counter-productive to restrict them to a fixed number of question/answer types. Even finding a non-strictly factual answer to an otherwise simple question about Titanic length (e.g., "two football fields") would push the limits of the answer-typing approach. Therefore, the formation of an answer should instead be guided by the topics the user is interested in, as recognized in the question and/or through the interactive dialogue, rather than by a single type as inferred from the question in a factoid system.

3. Document Retrieval

When the user poses a question to a system sitting atop a huge database of unstructured data (text files), the first order of business is to reduce that pile to perhaps a handful of documents where the answer is likely to be found. This means, most often, some form of document retrieval, using fast but non-exact selection methods. In order to accomplish this, users' natural language questions need to be tokenized, converted into search queries and sent to a document retrieval engine, such as Google (www.google.com), SMART (Buckley, 1985), InQuery (Callan et al., 1992), OKAPI (Robertson et al., 2000), etc. In the experiments with the HITIQA prototype, document retrieval is used to return the top fifty documents from a large text archive comprised of the AQUAINT corpus[2] and a smaller collection of web-harvested documents. Currently, this archive can be searched using one of the two available IR systems (SMART, InQuery); in the future, multiple search engines will return documents and other information objects from different parts of distributed, multimodal data sources.

4. Data Driven Semantics of Questions

The set of documents and text passages returned from the initial search is not just a random subset of the database. Depending upon the quality of the text retrieval system available, this set can be considered as a first stab at understanding the user's question by the machine. Again, given the available resources, this is the best the system can do under the circumstances. Therefore, we may as well consider this collection of retrieved texts (the Retrieved Set) as

---

[2] The AQUAINT corpus of news articles from a variety of sources covering 1998-2000 is available to research community through the Linguistic Data Consortium.

the meaning of the question as understood by the system. This is a fair assessment: the better our search capabilities, the closer this set would be to what the user may accept as an answer to the question.

It should be clear that better search does not mean just better precision or better recall; it means both and it also means utility, novelty and context. In other words, while new information has the highest value to the user, an independent confirmation makes it more reliable. On the other hand, the importance of relevant information is often relative to other, perhaps only marginally related information; such additional information provides a context or backdrop for the current topic and gives an analysts a fuller picture of relative importance of the key facts. Therefore, the search needs to cast a net wide enough to capture more than just the most directly relevant pieces.

5. Clustering and Framing

5.1. Clustering

Information contained in the documents in the Retrieved Set is first clustered to uncover any salient themes running through it. This classification is partially orthogonal to the relevance ranking used in document search because it is reinforced by mutual similarities of text fragments within the set (i.e., by repetition of the same theme or pattern) rather than by similarity to the question. While occasionally these two classifications do converge, more often they form a two-dimensional space of things that are important because they are highly similar to the question and things that are important because they show up in numbers. Other dimensions can be added as well, for example time, geographical location, etc., as well as information quality including source, reliability and so forth. Other thrusts in HITIQA projects address these issues (Tang et al., 2003).

We use n-gram-based clustering of text passages and concept extraction to uncover the main topics, themes and entities in the retrieved set. Retrieved documents are first broken into naturally occurring paragraphs. Duplicate paragraphs are filtered out and the remaining passages are clustered using a two-step process, which is a combination of hierarchical clustering and n-bin classification (details of the clustering algorithm can be found in Hardy et al., 2002). The hierarchical clustering phase forms a small number of highly connected clusters based on strong similarity criteria derived from variable length word n-gram overlap. Typically three to six clusters are generated out of the top 50 documents, which may yield as many as 1000 passages, although only a small subset of these will enter the initial clusters (cluster seeds). The goal here is to have each cluster represent a salient theme within the retrieved set: usually an alternative or complimentary interpretation of the user's question. This is of course not always possible, and some clusters may end up merely accidental aggregations of text snippets.

5.2. Framing

In HITIQA we use a text framing technique to delineate the gap between the meaning of the user's question and the system "understanding" of this question. The framing is an attempt to impose a partial structure on the text that would allow for systematic comparison among different text pieces and also against the question. In particular, the framing process may reveal that some topics and themes within the retrieved set are not what the user had explicitly asked for, and thus may be unaware of their existence. Nonetheless these may be highly correlated with the desired answer, and may indeed carry important information.

In the current version of the system, most frames are fairly generic templates, consisting of a small number of attributes, such as location, person, country, organization, etc. The intent is of course to capture an event or a relation (for which we reserve a special attribute topic), but the generic frame makes no specific assumptions about the roles of its attributes with respect to topic event or relation. We have also developed a small number of relation templates, specifically for the weapons non-proliferation domain, which is one of the sources of our data. Within this domain, topically specialized frames can be defined, such as transfer, treaty, development, etc. For example the transfer frame (capturing all transfer events of weapons, technologies, components, etc.) assigns semantic roles to its attributes (e.g., source, destination, weapon type). While this may not capture the entire meaning of the text fragment, it nonetheless allows the system to make a crude comparison: this document says Egypt obtained missile components from North Korea (transfer missile-parts from N. Korea to Egypt), while another document says Egypt transferred weapon technology to Korea.

Most of the frame attributes are defined in advance, however dynamic frame expansion is also possible. Each of the attributes in a frame is equipped with an extractor function which specializes in locating and extracting instances of this attribute in the running text. The extractors are implemented using information extraction utilities which form the kernel of Sheffield's GATE[3] system. We have modified GATE to separate organizations into companies and other organizations, and we have also expanded by adding new concepts such as industries. More recently we have been using BBN's Identifinder software, which extracts 24 types of entities.

The framing process resembles strongly the template filling task in information extraction (cf. MUC[4] evaluations), with one significant exception. The MUC template task was to fill in a single template using information from an entire document (Humphreys et al., 1998). In the framing process, on the other hand, templates are filled by entities extracted from clusters of small size, very similar passages. The added redundancy of information makes this a less error-prone task – our experience has thus far been more positive than the MUC evaluation results might indicate. Furthermore, clusters containing adjacent text passages would be framed independently and then merged by combining values of frame attributes.[5]

A very similar process is applied to the user's question, resulting in a Goal Frame which can be subsequently compared to the data frames obtained from the retrieved set. For example, the Goal Frame generated from the question, "How has pollution in the Black Sea affected the fishing industry, and what are the sources of this pollution?" is shown in Figure 1. We should note that this is a generic frame and the attributes have no special roles beyond their usual interpretation.

The frames obtained from the clusters in the retrieved set (or Data Frames) are then compared to the Goal Frame. We pay particular attention to matching the topic attribute, before any other attributes are considered. As explained before, the topic attribute is expected to contain the main event or relation expressed in text. If any of the Goal Frame topics is present anywhere in the data cluster, then

---

[3] GATE is Generalized Architecture for Text Engineering, an information extraction system developed at the University of Sheffield (Cunningham, 2000).

[4] MUC, the Message Understanding Conference, funded by DARPA, involved the evaluation of information extraction systems applied to a common task.

[5] We should note that selecting the right frame type for a passage is an important pre-condition to "understanding".

it becomes the data frame's topic as well. If more than one match is found, the subsequent matches become sub-topics of the data frame. On the other hand, if no explicit match is found against the Goal Frame topic, we attempt to select the topic from the list of the Wordnet generated hypernyms.

Topic:*[pollution, industry, sources]*
Location*: [Black Sea]*
Industry*:[fishing]*
Text*: How has pollution in the Black Sea affected the fishing industry, and what are the sources of this pollution?*

Figure 1: HITIQA generated Goal Frame

Topic*: pollution*
Sub-topic*: [sources]*
Location*: [Black Sea]*
Industry *:[fisheries, tourism]*
Text: *[In a period of only three decades (1960's-1980's), the* ***Black Sea*** *has suffered the catastrophic degradation of a major part of its natural resources. Particularly acute problems have arisen as a result of* ***pollution*** *(notably from nutrients, fecal material, solid waste and oil), a catastrophic decline in commercial fish stocks, a severe decrease in* ***tourism*** *and an uncoordinated approach towards coastal zone management. Increased loads of nutrients from rivers and coastal* ***sources*** *caused an overproduction of phytoplankton leading to extensive eutrophication and often extremely low dissolved oxygen concentrations. The entire ecosystem began to collapse. This problem, coupled with* ***pollution*** *and irrational exploitation of fish stocks, started a sharp decline in* ***fisheries*** *resources.]*
Relevance*: Matches on all elements found in goalframe*

Figure 2: A HITIQA generated data frame. Words in bold were used to fill the frame.

A sample data frame generated from a passage retrieved in response to the query about the Black Sea is shown in Figure 2.

After the initial framing is done, an additional frame merging step is attempted for the frames assigned to the same topic. Frames judged to be related to the same event (same topic, overlap on key attribute values), are merged and values of their attributes are combined.

5.3. Judging Frame Relevance

We judge a particular data frame as relevant, and subsequently the corresponding segment of text as relevant, by comparison to the Goal Frame. A data frame is given a negative point for each value mismatches (called conflicts) between its attributes and the Goal Frame. If a data frame has no conflicts, it is given the highest relevance rank, and a conflict score of zero. All other data frames are scored with an incrementing conflict value, one for frames with one conflict with the Goal Frame, two for two conflicts etc. Frames that conflict with all information found in the query are given a score of 99 indicating the lowest relevancy rank. We should note here that such frames may sit over passages with some keyword overlap with the query, but presumably such overlap is semantically meaningless. Currently, frames with a conflict score of 99 are excluded from further processing.

6. Enabling Dialogue with the User

Framed information represents the system's partial understanding of the user question; specifically, zero-conflicts frames constitute elements of the answer. An answer can now be composed from the underlying text passages, or by generation, using attribute values from the frame. However, a large and diverse population of near-miss frames may indicate that only part of the answer has been captured; conversely a large number of zero frames may suggest that a narrower interpretation of the question is warranted. Whether any of these conditions require revisions to the answer space can be determined through a dialogue with the user.

HITIQA automatically evaluates the semantic correlation between frames and initiates a clarification dialogue with the user. The purpose of the dialogue is to help bridge the gap between the user's expectations and the system heretofore "understanding" of the question. In clarification dialogue the system solicits from the user more details about what information he or she is seeking, based on the composition of the entire answer space, including the near-miss frames. The main principle here is that the dialogue is at the information semantic level, not at the information organization level. Moreover, the dialogue relates directly to the subject matter addressed by the question: it is of course a challenge to make sure that what happens next is indeed a logical follow up to the question. Consider the following exchange:

> USER: *What is being done to combat AIDS epidemic in Africa?*
> HITIQA: *Are you interested in information about International AIDS Conference?*
> USER: *yes, but only as related to access to AIDS medication and prevention.*

Thus, the dialogue relates directly to the topic of the user inquiry, and of course it may subsequently proceed in any direction. Here is another example recorded in a recent session with HITIQA:

USER: *How did al-Qaida conducted its efforts to acquire weapons of mass destruction?*
HITIQA: *We have this information about bin Laden, not about al-Qaida; would you like to see this?*
USER: *Yes, sure*.

On the other hand, the user should never be asked if a particular keyword is useful or not, or if a document is relevant or not. The only exception to this rule is when a misspelling or omission of a key concept is detected in the user input – we could occasionally try "Did you mean …".[6] However, our experience shows that these can be quite irritating to most users.

In HITIQA we adopted a three-pronged dialogue strategy, based on the organization of the answer space:

1. *Narrowing dialogue*: ask questions that would allow the system to reduce the size of the answer set. This is to narrow the interpretation of the question and to reduce the number of zero frames.
2. *Broadening dialogue*: ask questions that would allow the system to decide if the answer set needs to be expanded by information just outside of it (near-misses). This is to move some of the 1-conflict and maybe 2-conflict, etc. frames into the zero-conflict zone.
3. *Fact seeking dialogue*: ask the user to supply missing values for certain attributes such as date or location, etc. Furthermore, allow the user to ask questions seeking additional facts and specific examples, or similar situations.[7]

Of the above, we have thus far implemented the first two options as part of the preliminary clarification dialogue. The clarification dialogue is when the user and the system negotiate the task that needs to be performed. We call this a "triaging stage" of the interaction, as opposed to the actual problem solving stage (item 3 above). In practice, these two stages are not necessarily separated and may be overlapping throughout the entire interaction.

[6] This option is not currently implemented in HITIQA.
[7] Please refer to examples in Dialogue One given at the end of this chapter.

6.1. Clarification Dialogue

Data frames with a conflict score of zero form the initial kernel answer space. Depending upon the size of this set and the presence of other frames outside of it, the system either proceeds to generate the answer or initiates a dialogue with the user. For example, if the answer space appears too large or varied, e.g. consists of many different topics, the system may ask the user how to narrow it. Alternatively, the presence of large groups of texts frames with near-miss frames assigned to them (i.e., frames with 1 or 2 conflicts with the Goal Frame) may indicate that the answer space is actually larger, and the user will be consulted about a possible broadening of the question. Currently, we only initiate a clarification dialogue for 1-conflict frames.

A 1-conflict frame has only a single attribute mismatch with the Goal Frame. This could be a mismatch on any attribute, for example, location, or organization, or time, etc. A special case arises when the conflict occurs on the topic attribute. Since all other attributes match, we may be looking at potentially different events or situations involving the same entities, or occurring at the same location or time. The purpose of the clarification dialogue in this case is to probe which of these topics may be of interest to the user. This is illustrated in the exchange below recorded during an evaluation session with an intelligence analyst:

user: Who is Elizardo Sanchez?
HITIQA: Are you interested in seeing information about <u>civil rights</u> as it is related to Elizardo Sanchez?

Topic: *civil-rights*
Sub-topic*: []*
Location*: [Cuba, Cuba, Cuba]*
Companies*: []*
People*: [Sanchez, Sanchez]*
Document date*: [2000, 1, 11]*
Source*: HAVANA (AP) –*
Text*: [``I consider that the situation for civil and political rights in* ***Cuba*** *has worsened over the* ***past year****... owing to that* ***Cuba*** *continues to be the only closed society in this hemisphere,''* ***Sanchez*** *said. ``There have been no significant release of prisoners, the number of people sanctioned or processed for political motives increased.* ***Sanchez****, who himself spent many years in* ***Cuban*** *prisons, is among the communist island's best known opposition activists. The commission he heads issues a report on civil rights every six months, along with a list of people it considers to be imprisoned for political motives.]*

Figure 3: A 1-conflict frame utilized in generating the Sanchez dialogue.

In order to understand what happened here, we need to note first that the Goal Frame for the user question does not have any specific value assigned to its topic attribute. This of course is as we would expect it: the question does not give us a hint as to what information we need to look for or may be hoping to find about Sanchez. This also means that all the data frames obtained from the retrieved set for this question will have at least one conflict with the Goal Frame, and thus will be classified as either near-misses or outliers. One such 1-conflict near-miss frame is shown in Figure 3: its topic is "civil rights", from the list of Wordnet generated hypernyms, and it is about Sanchez. HITIQA thus asks if "civil rights" is a topic of interest to the user. If the user responds positively, this topic will be added to the Goal Frame, and thus to the answer space. Subsequently, all 1-conflict frames with a civil-rights topic will be re-scored to zero conflicts. Similarly, 2-conflict frames with civil-rights topic will be re-scored to 1-conflict, etc. The above dialogue strategy is applicable to other attribute mismatch cases, and produces intelligent-sounding responses from the system.

The clarification dialogue will continue on the topic level until all the significant sets of NEAR-MISS frames are either included in the answer space (through user broadening the scope of the question that removes the initial conflicts) or dismissed as not relevant. When HITIQA reaches this point it will re-evaluate the data frames in its answer space. If there are too many answer frames now (more than a pre-determined upper threshold), the dialogue manager will offer to the user to narrow the question using another frame attribute. On the other hand, if the size of the new answer space is still small and there are many unresolved near-miss frames, the dialogue manager will suggest to the user ways of further broadening the answer space, thus making more data frames relevant, or possibly retrieving new documents by adding terms acquired through the clarification dialogue. When the number of frames is within the acceptable range, HITIQA will generate the answer using the text from the frames in the current answer space. The user may end the dialogue at any point and have an answer generated given the current state of the frames.

6.2. Answer Generation

In the first version of HITIQA, the answer is simply composed of text passages underlying the zero conflict frames. The text of these frames is ordered by date and output to the user. Typically, an answer to analytical type questions will require a few pages of information. The example below shows the first portion of the answer generated by HITIQA for the Black Sea query.

USER: *How has pollution in the Black Sea affected the fishing industry, and what are the sources of this pollution?*

HITIQA:

The Black Sea is widely recognized as one of the regional seas most damaged by human activity. Almost one third of the entire land area of continental Europe drains into this sea (...) The management of the Black Sea itself is the shared responsibility of the six coastal countries: Bulgaria, Georgia, Romania, Russia, Turkey, and Ukraine…

Our current work on answer generation is focused on fusion of more succinct and informative answer summaries. The first step in this process is to define a process by which the frame information can be utilized to automatically generate a "headline" for the underlying text passage. For example:

USER: *Do we have any data about Syria having nerve gas manufacturing capability?*

HITIQA:

SYRIA REPORTED TO DEVELOP SARIN NERVE GAS, WARHEADS AND MISSILES:
Syria is said to possess a large number of warheads filled with VX and sarin nerve gas, which can be mounted on Scud-B and Scud-C missiles. It is reported to have factories near Damascus and Homs that are producing hundreds of tons of mustard gas and sarin nerve agent.

This "headline" is generated by exploiting the semantics of thematic frames where attributes are assigned specific roles. In general, the more precise the semantics of a frame, the more satisfactory headline is generated. For generic frames (like those used in Sanchez examples in section 9) it would much harder to automatically come up with the right wording for a headline.

The work on answer fusion is in the initial stages. Eventually, we envision that headlines of multiple text passages will form a fused summary of the answer, something that the user can view first.

7. Visual Navigation through Answer Space

We use visualization techniques to supplement the dialogue and give the user more control over the data if they so choose. Visual display allows the user to see all the frames in a compact visual form and also to directly manipulate frame relevance. Figure 6 shows typical top level visualization panels. Each panel is split into two halves. The top half displays frames or group of frames, the lower half displays the legend. Frames are represented by colored circles, and are arranged by their scores and by their topic attribute values. Scores are represented using a color palette (bottom, left), from the most relevant (blue) to least relevant (red), using the accepted convention for denoting elevations on geographical maps. Annotated spokes on the circles represent key attributes of the frame. Each attribute has a fixed location on circle perimeter, so they can be visually located even when annotation is removed. This multi-parametric visualization technique has been shown to be effective (Grinstein et al., 1993). The text window in the lower right corner is used to display frame text.

The user can change relevancy of selected frame groups or individual frames. By positioning the cursor over a frame group or a frame and clicking the right mouse button the user changes the conflict-score of the selection according to the following logic. If the score is not equal to zero, it will become zero, and if the score is zero it becomes 99. Users also can reset the scores of all frames to their original (at the visual initiation time) values. With the left mouse button click when the cursor is positioned over the group of frames the user can switch to the display of the individual frames of this group. In Figure 6, the left screen shows three groups of frames (two 0-conflict and one 99-conflict), while the right panel displays four individual frames in one of the groups (represented by the lower blue circle on the left panel). By clicking the left mouse button again the user returns to the group level display, thus the left mouse button is used as a toggle between the individual frames and the groups of frames display.

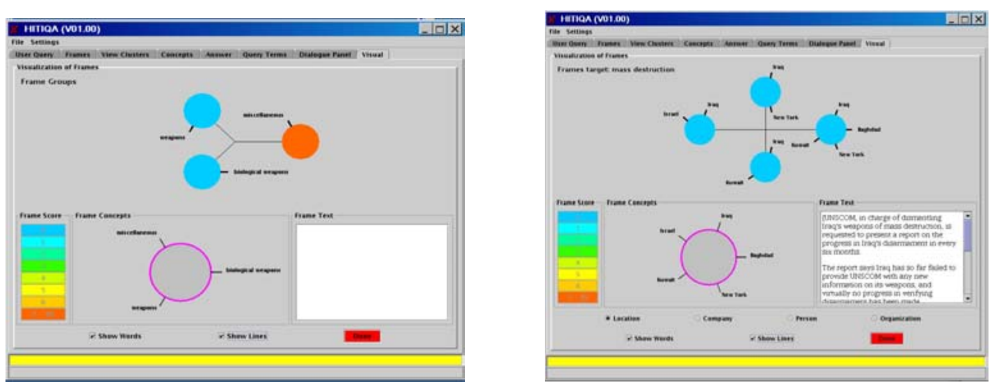

Figure 6. Group Level Display and Frame Level Display

When individual frames are displayed, the user may choose which attributes should be visualized by selecting from the menu in the legend window. This is controlled by a set of radio buttons with the names of attributes. When the user rolls the mouse over a frame circle, the actual document text is displayed in the small window on the right. The combination of attribute, frame, group and score manipulations allows for fast traversal of the answer space. Moving from the visual dialog to the textual dialog and back is seamless in a sense that any changes to the frame scores in one modality are immediately accessible to the user in another modality. When the “Done” button is clicked, HITIQA will either generate a new dialogue query if needed or generate the answer. We noted that users frequently return to visualization panel after the final answer has been generated, in order to verify that no important item has been overlooked. They can add and remove frames from the answer space and HITIQA will always seamlessly pickup a new dialogue or generate a new answer.

13. Future Work

This chapter describes a work in progress. We expect that the initial specification of content frame will evolve as we subject the initial system to more demanding evaluations. Currently, the frames are not topically specialized, and this appears the most logical next refinement, i.e., develop several (10-30) types of frames covering different classes of events, from politics to medicine to science to international economics, etc. This is expected to increase the accuracy of the dialogue as is the interactive visualization which is also under development. Answer generation will involve fusion of information on the frame level, and is currently in an initial phase of implementation.

14. Acknowledgements

This paper is based on work supported by the Advanced Research and Development Activity (ARDA)’s Advanced Question Answering for Intelligence (AQUAINT) Program under contract number 2002-H790400-000. We would like to thank Ralph Weischedel of Verizon/BBN for making Identifinder available for this project.

13. References

Chris Buckley. May 1985. Implementation of the SMART information retrieval system. Technical Report TR85-686, Department of Computer Science, Cornell University, Ithaca, NY.

James P. Callan, W. Bruce Croft, Stephen M. Harding 1992. The INQUERY Retrieval System. Proc. of DEXA-92, 3rd International Conference on Database and Expert Systems Applications. 78-83.

Cunningham, H., D. Maynard, K. Bontcheva, V. Tablan and Y. Wilks. 2000 Experience of using GATE for NLP R&D. In Coling 2000 Workshop on Using Toolsets and Architectures To Build NLP Systems.

Grinstein, G.G., Levkowitz, H., Pickett, R.M., Smith, S. 1993. "Visualization alternatives: non-pixel based images," Proc. of IS&T 46th Conf. 132-133.

H. Hardy, N. Shimizu, T. Strzalkowski, L. Ting, B. Wise and X. Zhang 2002. Cross-Document Summarization by Concept Classification. Proceedings of SIGIR-2002, Tampere, Finland.

Harabagiu, S., M. Pasca and S. Maiorano. 2000. Experiments with Open-Domain Textual Question Answering. In Proc. of COLING-2000. 292-298.

Humphreys, R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, Y. Wilks. 1998. Description of the LaSIE-II System as Used for MUC-7. In Proceedings of the Seventh Message Understanding Conference (MUC-7.)

Hovy, E., L. Gerber, U. Hermjakob, M. Junk, C-Y. Lin. 2000. Question Answering in Webclopedia. Proc. of Text Retrieval Conference (TREC-9).

John Prager, Dragomir R. Radev, and Krzysztof Czuba. Answering what-is questions by virtual annotation. In Human Language Technology Conference, Demonstrations Section, San Diego, CA, 2001.

Stephen E. Robertson, Steve Walker, Micheline Hancock-Beaulieu. (2000). Experimentation as a way of life: Okapi at TREC. Information Processing and Management 36(1): 95-108.

Rong Tang, K.B. Ng, Tomek Strzalkowski and Paul Kantor (2003). "Automatic Prediction of Information Quality in News Documents." Proceedings of HLT-NAACL 2003, Edmonton, Canada, May 27-June 1.