Vision Statement to Guide Research in Question & Answering (Q&A) and Text Summarization

by
Jaime Carbonell[1], Donna Harman[2], Eduard Hovy[3], and Steve Maiorano[4], John Prange[5], and Karen Sparck-Jones[6]

1. INTRODUCTION

Recent developments in natural language processing R&D have made it clear that formerly independent technologies can be harnessed together to an increasing degree in order to form sophisticated and powerful information delivery vehicles. Information retrieval engines, text summarizers, question answering systems, and language translators provide complementary functionalities which can be combined to serve a variety of users, ranging from the casual user asking questions of the web (such as a schoolchild doing an assignment) to a sophisticated knowledge worker earning a living (such as an intelligence analyst investigating terrorism acts).

A particularly useful complementarity exists between text summarization and question answering systems. From the viewpoint of summarization, question answering is one way to provide the focus for query-oriented summarization. From the viewpoint of question answering, summarization is a way of extracting and fusing just the relevant information from a heap of text in answer to a specific non-factoid question. However, both question answering and summarization include aspects that are unrelated to the other. Sometimes, the answer to a question simply cannot be summarized: either it is a brief factoid (*the capital of Switzerland is Berne*) or the answer is complete in itself (*give me the text of the Pledge of Allegiance*). Likewise, generic (author's point of view summaries) do not involve a question; they reflect the text as it stands, without input from the system user.

This document describes a vision of ways in which Question Answering and Summarization technology can be combined to form truly useful information delivery tools. It outlines tools at several increasingly sophisticated stages. This vision, and this staging, can be used to inform R&D in question answering and text summarization. The purpose of this document is to provide a background against which NLP research

---

[1] Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA 15213-3891. jgc@cs.cmu.edu
[2] National Institute of Standards and Technology, 100 Bureau Dr., Stop 8940, Gaithersburg, MD 20899-8940. donna.harman@nist.gov
[3] University of Southern California-Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292-6695. Hovy@isi.edu
[4] Advanced Analytic Tools, LF-7, Washington, DC 20505. Stevejm@ucia.gov
[5] Advanced Research and Development Activity (ARDA), R&E Building STE 6530, 9800 Savage Road, Fort Meade, MD 20755-6530. JPrange@ncsc.mil
[6] University of Cambridge, New Museums Site, Pembroke Street, Cambridge, CB2 3QG, ENGLAND. Karen.Sparck-Jones@cl.cam.ac.uk

sponsored by DRAPA, ARDA, and other agencies can be conceived and guided. An important aspect of this purpose is the development of appropriate evaluation tests and measures for text summarization and question answering, so as to most usefully focus research without over-constraining it.

2. **BACKGROUND**

Four multifaceted research and development programs share a common interest in a newly emerging area of research interest, Question and Answering, or simply Q&A and in the older, more established text summarization.

These four programs and their Q&A and text summarization intersection are the [7]:

- Information Exploitation R&D program being sponsored by the Advanced Research and Development Activity (ARDA). The "Pulling Information" problem area directly addresses Q&A. This same problem area and a second ARDA problem area "Pushing Information" includes research objectives that intersect with those of text summarization. (John Prange, Program Manager)
- Q&A and text summarization goals within the larger TIDES (Translingual Information Detection, Extraction, and Summarization) Program being sponsored by the Information Technology Office (ITO) of the Defense Advanced Research Project Agency (DARPA) (Gary Strong, Program Manager)
- Q&A Track within the TREC (Text Retrieval Conference) series of information retrieval evaluation workshops that are organized and managed by the National Institute of Standards and Technology (NIST). Both the ARDA and DARPA programs are providing funding in FY2000 to NIST for the sponsorship of both TREC in general and the Q&A Track in particular. (Donna Harman, Program Manager)
- Document Understanding Conference (DUC). As part of the larger TIDES program NIST is establishing a new series of evaluation workshops for the text understanding community. The focus of the initial workshop to be held in November 2000 will be text summarization. In future workshops, it is anticipated that DUC will also sponsor evaluations in research areas associated with information extraction. (Donna Harman, Program Manager)

Recent discussions by among the program managers of these programs at and after the recent TIDES Workshop (March 2000) indicated the need to develop a more focused and coordinated approach against Q&A and a second area: summarization by these three programs. To this end the NIST Program Manager has formed a review committee and separate roadmap committees for both Q&A and Summarization. The goal of the three committees is to come up with two roadmaps stretching out 5 years.

The Review Committee would develop a "Vision Paper" for the future direction of R&D in both Q&A and text summarization. Each Roadmap Committee will then prepare a response to this vision paper in which it will outline a potential research and development path(s) that has (have) as their goal achieving a significant part (or maybe all) of the ideas laid out in the Vision Statement. The final versions of the Roadmaps, after evaluation by

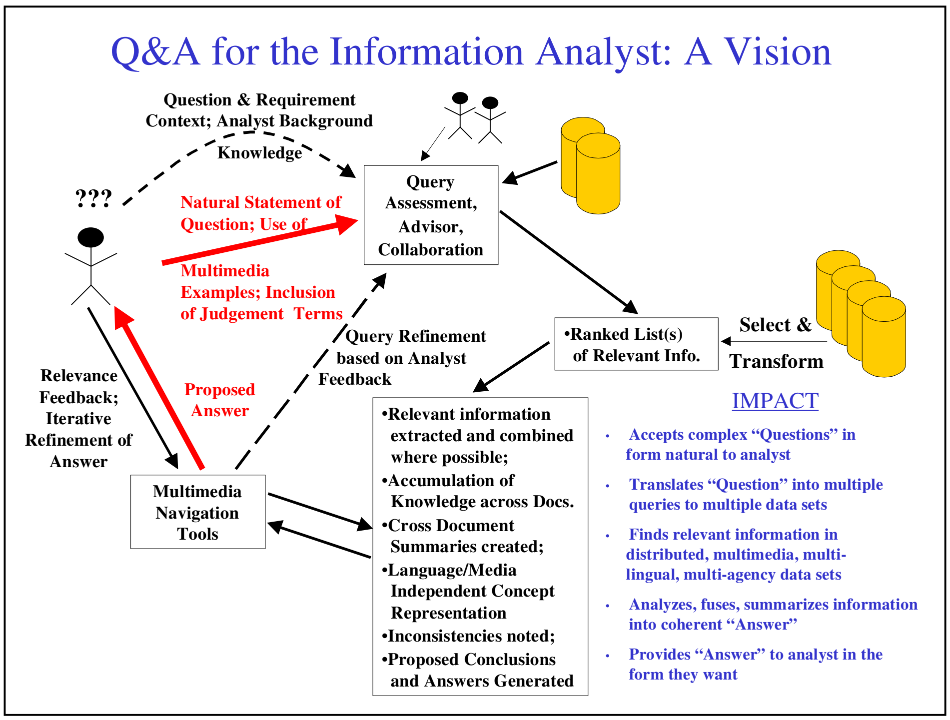
[7] Additional background information on the ARDA Information Exploitation R&D program, on DARPA TIDES program, on the TREC Program and its Q&A Track are attached as Appendix 1 to this document.

the Review Committee, and the Vision Paper would then be made available to all three programs, and most likely also to the larger research community in Q&A and Summarization areas, for their use in plotting and planning future programs and potential cooperative relationships.

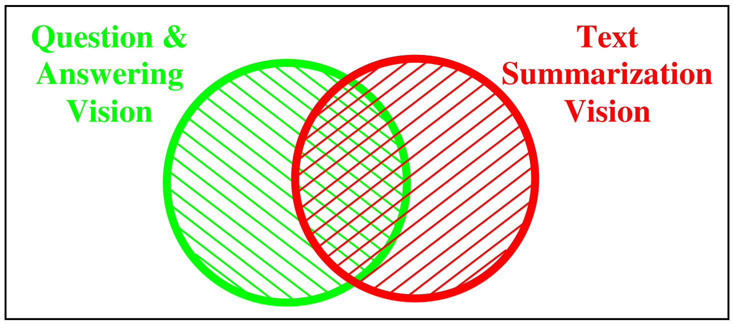
Vision Paper for Q&A and Text Summarization

This document constitutes the Vision Paper that will serve to guide both the Q&A and Text Summarization Roadmap Committees.

In the case of Q&A, the vision statement focuses on the capabilities needed by a high-end questioner. This high-end questioner is identified later in this vision statement as a "Professional Information Analyst". In particular this Information Analyst is a knowledgeable, dedicated, intense, professional consumer and producer of information. For this information analyst, the committee's vision for Q&A is captured in the following chart that is explained in detailed later in this document.



As mentioned earlier the vision for text summarization does intersect with the vision for Q&A. In particular, this intersection is reflected in the above Q&A Vision chart as part of the process of generating an Answer to the questioner's original question in a form and style that the questioner wants. In this case summarization is guided and directed by the scope and context of the original question, and may involve the summarization of information across multiple information sources whose content may be presented in more than one language media and in more than one language. But as indicated by the following Venn diagram, there is more to text summarization than just its intersection with Q&A. For example, as previously mentioned generic summaries (author's point of view summaries) do not involve a question; they reflect the text as it stands, without input from the system user. Such summaries might be useful to produce generic "abstracts" for text documents or to assist end-users to quickly browse through large quantities of text in a survey or general search mode. Also if large quantities of unknown text documents are clustered in an unsupervised manner, then summarization may be applied to each document cluster in an effort to identify and describe that content which caused the clustered documents to be grouped together and which distinguishes the given cluster from the other clusters that have been formed.



the process of generating an Answer to the questioner's original question in a form and style that the questioner wants. In this case summarization is guided and directed by the scope and context of the original question, and may involve the summarization of information across multiple information sources whose content may be presented in more than one language media and in more than one language. But as indicated by the above Venn diagram, there is more to text summarization than just its intersection with Q&A. For example, as previously mentioned generic summaries (author's point of view summaries) do not involve a question; they reflect the text as it stands, without input from the system user. Such summaries might be useful to produce generic "abstracts" for text documents or to assist end-users to quickly browse through large quantities of text in a survey or general search mode. Also if large quantities of unknown text documents are clustered in an unsupervised manner, then summarization may be applied to each document cluster in an effort to identify and describe that content which caused the clustered documents to be grouped together and which distinguishes the given cluster from the other clusters that have been formed.

Summarization is not separately discussed again until the final section of the paper (Section 7: Multidimensionality of Summarization.) In the intervening sections (Sections 3-6) the principal focus is on Q&A. Summarization is addressed in these sections only to the extent that Summarization intersects Q&A.

This Vision Paper is Deliberately Ambitious

This vision paper has purposely established as its challenging long-term goal, the building of powerful, multipurpose, information management systems for both Q&A and Summarization. But the Review Committee firmly believes that its global, long-term vision can be decomposed into many elements, and simpler subtasks, that can be attacked in parallel, at varying levels of sophistication, over shorter time frames, with benefits to many potential sub-classes of information user. In laying out a deliberately ambitious vision, the Review Committee is in fact challenging the Roadmap Committees to define program structures for addressing these subtasks and combining them in increasingly sophisticated ways.

3. FULL SPECTRUM OF QUESTIONERS

Clearly there is not a single, archetypical user of a Q&A system. In fact there is a full spectrum of questioners ranging from the TREC-8 Q&A type questioner to the knowledgeable, dedicated, intense, high-end professional information analyst who is most likely both an avid consumer and producer of information. These are in a sense then the two ends of the spectrum and it is the high end user against which the vision statement for Q&A was written. Not only is there a full spectrum of questioners but there is also a continuous spectrum of both questions and answers that correspond to these two ends of the questioner spectrum (labeled as the "Casual Questioner" and the "Professional Information Analyst" respectively). These two correlated spectrums are depicted in the following chart.

SOPHISTICATION LEVELS OF QUESTIONERS

Level 1	Level 2	Level 3	Level 4
"Casual Questioner"	*"Template Questioner"*	*"Cub Reporter"*	*"Professional Information Analyst"*

COMPLEXITY OF QUESTIONS & ANSWERS RANGES:

FROM:		**TO:**
Questions: *Simple Facts*	↔	***Questions:*** *Complex; Uses Judgement Terms Knowledge of User Context Needed; Broad Scope*
Answers: *Simple Answers found in Single Document*	↔	***Answers:*** *Search Multiple Sources (in multiple Media/languages); Fusion of information; Resolution of conflicting data; Multiple Alternatives; Adding Interpretation; Drawing Conclusions*

But what about the other levels of questioners between these two extremes? The preceding chart identifies two intermediate levels: the "Template Questioner" and the "Cub Reporter". These may not be the best labels, but how they are labeled is not so important for the Q&A Roadmap Committee. Rather what is important is that if the ultimate goal of Q&A is to provide meaningful and useful capabilities for the high-end questioner, then it would be very useful when plotting out a research roadmap to have at least of couple of intermediate check points or intermediate goals. Hopefully sufficient detail about each of the intermediate levels is given in the following paragraphs to make them useful mid-term targets along the path to the final goal.

So here are some thoughts on these four levels of questioners:

Level 1. "Casual Questioner". The Casual Questioner is the TREC-8[8] Q&A type questioner who asks simple, factual questions, which (if you could find the right textual document) could be answer in a single short phrase. For Example: Where is the Taj Mahal? What is the current population of Tucson, AZ? Who was the President Nixon's 1st Secretary of State? etc.

Level 2. "Template Questioner". The Template Questioner is the type of user for which the developer of a Q&A system/capability might be able to create "standard templates" with certain types of information to be found and filled in. In this case it is likely that the answer will not be found in a single document but will require retrieving multiple documents, locating portions of answers in them and combining them into a single response. If you could find just the right document, the desired answer might all be there, but that would not always be the case. And even if all of the answer components were in a single document then, it would likely be scattered across the document. The questions at this level of complexity are still basically seeking factual information, but just more information than is likely to be found in a single contiguous phrase. The use of a set of templates (with optional slots) might be one way to restrict the scope and extent of the factual searching. In fact a template question might be addressed by decomposing it into a series of single focus questions, each aimed at a particular slot in the desired template. The template type questions might include questions like the following:

- "What is the resume/biography of junior political figure X" The true test would not be to ask this question about people like President Bill Clinton or Microsoft's Chairman Bill Gates. But rather, ask this question about someone like the Under Secretary of Agriculture in African County Y or Colonel W in County Z's Air Force. The "Resume Template" would include things like full name, aliases, home & business addresses, birth, education, job history, etc.
- "What do we know about Company ABC?" A "resume" type template but aimed at company information. This might include the company's organizational structure - both divisions, subsidiaries, parent company; its product lines; its key officials, revenue figures, location of major facilities, etc.
- "What is the performance history of Mutual Fund XYZ?"

You can probably quickly and easily think of other templates ranging from very simple to very involved and complex.

Not everything at this level fits nicely into a template. At this level there are also questions that would result in producing lists of similar items. For instance, "What are all of the countries that border Brazil?" or "Who are all of the Major League Baseball Players who have had 3000 or more hits during their major league careers?" One slight complication here might be some lists may be more open ended; that is, you might not know for sure when you have found all the "answers". For example, "What are all of the consumer products currently being marketed by Company ABC." The Q&A System might also need to resolve finding in different documents overlapping lists of products that may include variations in the ways in which the products are identified. Are the similarly

---

[8] For more information on the Q&A Track in TREC-8 check out the following web site: http://www.research.att.com/~singhal/qa-track.html. More information on both TREC and the Q&A Track is available at the NIST website: http://trec.nist.gov/.

named products really the same product or different products? Also each item in the list may in fact include multiple entries, kind of like a list of mini-templates. "Name all states in the USA, their capitals, and their state bird."

Level 3. "Questioner as a 'Cub Reporter'". We don't have a particularly good title for this type of questioner. Any ideas? But regardless of the name this next level up in the sophistication of the Q&A questioner would be someone who is still focused factually, but now needs to pull together information from a variety of sources. Some of the information would be needed to satisfy elements of the current question while other information would be needed to provide necessary background information. To illustrate this type and level of questioner, consider that a major, multi-faceted event has occurred (say an earthquake in City XYZ some place in the world). A major news organization from the United States sends a team of reporters to cover this event. A junior, cub reporter is assigned the task of writing a news article on one aspect of this much larger story. Since he or she is only a cub reporter, they are given an easier, more straightforward story. Maybe a story about a disaster relief team from the United States that specializes in rescuing people trapped within collapsed buildings. Given that this is unfamiliar territory for the cub reporter, there would a series of highly related questions that the cub reporter would most likely wish to pose of a general informational system. So there is some context to the series of questions being posed by the cub reporter. This context would be important to the Q&A system as it must judge the breadth of its search and the depth of digging within those sources. Some factors are central to the cub reporter's story and some are peripheral at best. It will be up the Q&A system to either decide or to appropriately interact with the cub reporter to know which is the case. At this level of questioner, the Q&A system will need to move beyond text sources and involve multiple media. These sources may also be in multiple foreign languages (e.g. the earthquake might be in a foreign country and news reports/broadcasts from around the world may be important.) There may be some conflicting facts, but would be ones that are either expected or can be easily handled (e.g. the estimated dollar damage; the number of citizens killed and injured, etc.) The goal is not to write the cub reporter's news story, but to help this 'cub reporter' pull together the information that he or she will need in authoring a focused story on this emerging event.

Level 4. Professional Information Analyst. This would be the high-end questioner that has been referred to several times earlier. Since this level of questioner will be the focus of the Q&A vision that is described in a later section of this paper, our description of this level of questioner will be limited. The Professional Information Analyst is really a whole class of questioners that might include:

- Investigative reporters for national newspapers (like Woodward and Bernstein of the Washington Post and Watergate fame) and broadcast news programs (like "60 Minutes" or "20-20");
- Police detectives/FBI agents (e.g. the detectives/agents who investigated major cases like the Unibomber or the Atlanta Olympics bombing);
- DEA (Drug Enforcement Agency) or ATF (Bureau of Alcohol, Tobacco and Firearms) officials who are seeking to uncover secretive groups involved in illegal activities and to predict future activities or events involving these groups;

- To the extent that material is available in electronic form more current event historians/writers (e.g. supporting a writer wishing to author a perspective on the air war in Bosnia, or to do deep political analysis of the Presidential race in the year xxxx);
- Stock Brokers/Analysts affiliated with major brokerage houses or large mutual funds that cover on-going activities, events, trends etc. in selected sectors of the world's economy (e.g. banking industry, micro-electronic chip design and fabrication);
- Scientists and researchers working on the cutting edge of new technologies that need to stay up with current directions, trends, approaches being pursued within their area of expertise by other scientists and researchers around the world (e.g. wireless communication, high performance computing, fiber optics, intelligent agents); or
- The national-level intelligence analysts affiliated with one of the Intelligence Community agencies (e.g. the Central Intelligence Agency, National Security Agency, or Defense Intelligence Agency) or the military intelligence analyst/specialist assigned to a military unit that is forward deployed.

Two of the government members of the Review Committee are affiliated with agencies within the Intelligence Community. Because of their level of expertise and experience with intelligence analysts within their respective agencies, the intelligence analyst has been selected as the exemplar for this class of high-end questioners or Professional Information Analysts. The following section provide a more in-depth description of the intelligence analyst and of the capabilities that a Q&A system would need to provide to fully satisfy the Q&A needs of a archetypical intelligence analyst. While the review committee believes that almost all of the intelligence analyst's needs and characteristics, as described, directly translate to each of the other Professional Information Analysts types identified above, the committee has chosen to write this next section from its base of expertise and to encourage individual readers to interpret these intelligence analyst within the context of another type of high-end questioner types with whom the reader may be more familiar.

4. THE PERSPECTIVE OF THE INTELLIGENCE ANALYST

The vision statement that will be provided in the next section is written from the perspective of intelligence analysts whose primary work responsibilities is the analysis and production of intelligence from human language or linguistically-based information. As mentioned in the preceding section the intelligence analysts was selected as the exemplar for the larger class of Professional Information Analysts because of the significant knowledge and experience of two of the Review Committee's members with intelligence analysts. The Review Committee believes that by understanding the perspective of the Intelligence Analysts will permit the members of the Roadmap Committee and other readers of this vision statement to appropriately extrapolate the intelligence analyst's perspective to the reader's favorite exemplar from the class of Professional Information Analyst. (Several other potential exemplars from this class are described at the very end of the preceding section.)

The stereotypical intelligence analyst that we are considering in this section, performs his or her analytic tasks at one of the national level Intelligence Community Agencies in order to produce strategic level intelligence that is principally directed towards the intelligence needs and requirements of the National Command Authority (NCA) (e.g. the President, his aids, National Security Council, Cabinet Secretaries, etc.).

Generalization about Strategic Level Intelligence Analysts

Before providing with what we believe to be important generalizations and observations about Strategic Level Intelligence Analysts, we need to identify two caveats:

- First, there are clearly other Intelligence Community organizations (see next section) and other levels of intelligence besides strategic (e.g. operational and tactical that is the focus of the TIDES hypothetical scenario provided earlier). And while believe that much of what follows applies to these latter analysts as well, we are in no way claiming that the following vision statement adequately addresses the capabilities that such analysts would need in a Q&A environment of the future.
- And second, there is clearly not a single, stereotypical analyst who is performing strategic level intelligence production within the national level Intelligence Community Agencies. But we believe that it is fair to make the following generalizations since have wide applicability even if they don't have universality. Also we believe that these generalization are important to describe since they individually and collectively have significant impact on the vision statement that follows in the next section.

So here are my generalizations. (Note: In the bullets that follow, all references to Intelligence Analysts are really references to Intelligence Analysts working at a national level Intelligence Community agency to produce strategic level intelligence for the National Command Authority or NCA.)

- Intelligence analysts are not casual consumers of information. Raw data and information is their lifeblood, the central focus of their professional efforts. They are often totally immersed in information and in their interpretation of this information against specific requirements that have been generated by the ultimate consumers of the intelligence that they produce. The analysis and production of intelligence from information is their full time job.

- Intelligence analysts are almost always subject matter experts within their assigned task area. They have typically worked in this task area for a significant number of years. It is not uncommon for the senior analysts within a given area or organization to have more than 20 years of experience. In some agencies more than others, these analysts may also be skilled linguists in multiple foreign languages or they have close access to such linguists. The point is that they have both broad and deep knowledge of the subject area they have been working for a significant time period and they are highly skilled analysts and linguists. They are consummate professionals who are highly dedicated to their assigned intelligence production tasks.
- Many Intelligence Analysts perform all source analysis and production. That is, their efforts require that they analyze and exploit information from multiple media (text, voice, image, etc.), from multiple languages, and different styles and types and then fuse their interpretation of these multiple information items into a single intelligence report.. Even when “single item reporting” is done, the analyst undoubtedly uses his or her past experience and knowledge that has been previously accumulated in an all source environment. Also while some information is automatically routed to analysts’ workstations, it is still the case that these analysts must know how to retrieve important information from a number of different databases and on-line archives, some of which might not be resident within their organization or even agency.
- Many Intelligence Analysts track and follow a given event, scenario, problem, situation within their assigned task area for an extended period of time. In this regards they frequently develop extensive “notes” and “working papers” that help them keep track of their evolving investigation. So when they develop a query for retrieving additional new information they are doing so within an extensive context, that is known to the analyst but which may not be specifically expressed within the current query. (Typically, the problem is that the retrieval system is not capability of accepting and using such contextual information even if the analyst provided it.)
- Many Intelligence Analysts need to coordinate their analysis and production tasks with other analysts who are working within the same subject domain or in a highly related subject domain area. These other analysts may be working in different organizations and even in different agencies. Unfortunately analysts do not always know who these analysts are that they would benefit from coordinating with and hence, in some situations, this may be an under utilized resource.
- Intelligence Analysts typically work with overwhelming volumes of information. Frequently the quality of the raw data that produces this voluminous information is far less than ideal. These analysts must often work with “dirty” data (e.g. data whose signal to noise ratio makes its intelligibility difficult), errorful data (e.g. the raw data may contain errors itself, or new errors may be introduced when the data

is collected or during subsequent processing steps), missing or incomplete data, conflicting data, data that is intentionally deceptive or whose validity is questionable, and data whose value degrades over time.

- Given all of the difficult conditions facing our Intelligence Analysts, their production of intelligence is judged against the following standards (called the "Tenets of Intelligence"): [9] (And you thought the CNN reporter had it tough!)
  - Timeliness. Intelligence must be made available in time for the NCA to act appropriately on it. Late intelligence is as useless as no intelligence.
  - Accuracy. To be accurate, intelligence must be objective. It must be free from any political or other constraint and must not be distorted by pressure to conform to the positions held by the NCA. Intelligence products must not be shaped to conform to any perceptions of the NCA's preferences. While intelligence is a factor in determining policy, policy must not determine intelligence.
  - Usability. Intelligence must be tailored to the specific needs of the NCA and provided in forms suitable for immediate comprehension. The NCA must be able to quickly apply intelligence to the situation at hand. Providing useful intelligence requires the intelligence producers to understand the circumstances under which their intelligence products are used.
  - Completeness. Complete intelligence answers the NCA's questions about the adversary and current situation to the fullest degree possible. It also tells the NCA what remains unknown. To be complete, intelligence must identify all the adversary's perceived capabilities. It must inform the NCA of possible future courses of action and it must forecast future adversary actions and intentions. Uncertainties and degrees of belief in each of these elements of the intelligence report must be clearly and understandably identified.
  - Relevance. Intelligence must be relevant to the planning and execution of responses to an adversary or to a situation. Intelligence must contribute to the NCA's understanding of both the adversary and the current situation. It must help the NCA to decide how to accomplished its policy goals and objectives without being unduly hindered by the adversary and within the constraints of the current situation.

[9] This description of the "Tenets of Intelligence" was extracted from "Intelligence Support to Operations", J-7 (Operational Plans and Interoperability Directorate), Joint Chiefs of Staff, March 2000.

5. A VISION FOR A FUTURE, ADVANCED QUESTION AND ANSWERING SYSTEM

In the most recent Text Retrieval Conference (TREC-8; Nov 1999) the Question and Answer (Q&A) Track included the following question:

- "Question 73: Where is the Taj Mahal?"

This is a simple factual question whose most obvious answer ("Agra, India") could be found in a single, short character string within at least one text document within the approximately 1.9 gigabyte text collection of primarily newswire reports and newspaper stories. That is the answer was a "simple answer, in a single document, from a single source consisting of a single language media".

Within the context of discussion provided in the previous section, questions that an Intelligence Analyst might wish to pose might be more on the order of:

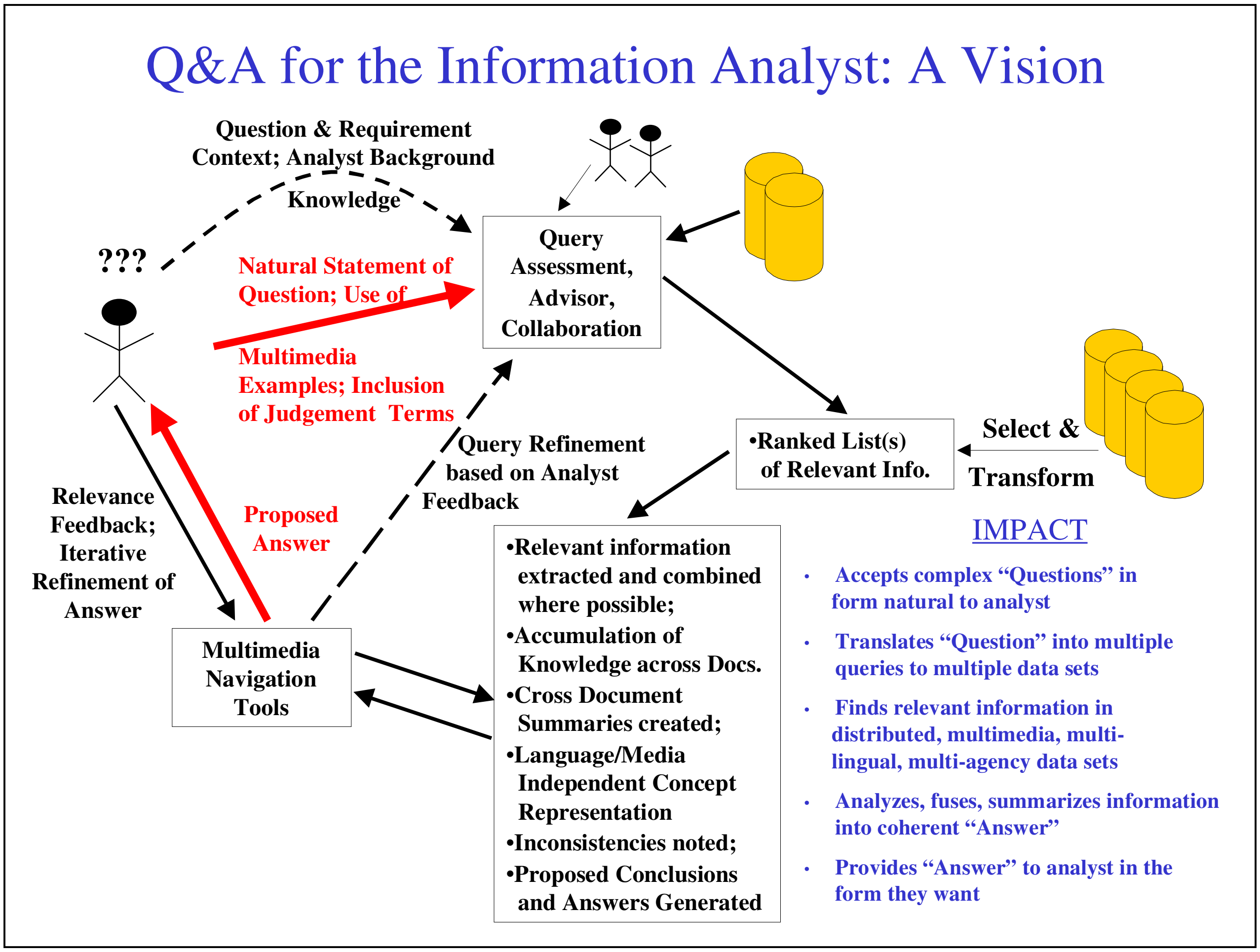
- While watching a video clip collected from the state television network of foreign power, the analyst becomes interested in a senior military officer who appears to be acting as an advisor to the country's Prime Minister. The analyst, who is responsible for reporting any significant changes in the political power base of the Prime Minister and his ruling party in this foreign country, is unfamiliar with this military officer. The analyst wishes to pose the questions, "Who is this individual? What is his background? What do we know about the political relationship of this unknown officer and the Prime Minister and/or his ruling party? Does this signal a significant shift in the influence of the country's military over the Prime Minister and his ruling party?"
- After reading a newswire report announcing the purchase of small foreign based chemical manufacturing firm (the processes used by this firm are dual use, capable of producing both agricultural chemicals as well as chemicals used in chemical weapon systems) by a different foreign based company (Company A). The analyst wishes to pose the following questions, "What other recent purchases and significant foreign investments has Company A, its subsidiaries, or its parent firm made in other foreign companies that are capable of manufacturing other components and equipment needed to produce chemical weapons? Has it openly, or through other third parties, purchased other suspicious equipment, supplies, and materials? What is the intent, purpose behind these purchases or investments?"
- While reading intelligence reports written by two different analysts from two different agencies, a third analyst has an "ah hah" experience and uncovers what she believes might be evidence of a strong connection between two previously unconnected terrorist groups that she had been tracking separately. The analyst wishes to pose the following questions, "Is there any other evidence of connections, communication or contact between these two suspected terrorist groups and its known members? Is there any other evidence that suggests that the two terrorist groups may be planning a joint operation? If so, where and when?"

When one compares these Intelligence Analyst questions provided above, along with the hypothesized answers that the reader might contemplate being produced for each, one quickly sees that the content, nature, scope and intent of these hypothesized Intelligence Analyst Q&A's are significantly more complex that those posed in the Q&A Track of TREC-8. The nature of these differences is discussed in more detail in the paper's final section. It is sufficient at this point to indicate that the factual components of these questions are unlikely to be found in a single text document. Rather the response to these factual components will require the fusing, combining, summarizing of smaller, individual facts from multiple data items (e.g. a newspaper article, a single news broadcast story) of potentially multiple different language media (e.g. relational databases, unstructured text, document images, still photographs, video data, technical or abstract data) presented in possibly different foreign languages. Then since the Q&A system will be required to fuse together multiple, partial factual information from different sources, there is a strong likelihood that there will be conflicting facts, duplicate facts and even incorrect facts uncovered. This may result in the need to develop multiple possible alternatives, each with their own level of confidence. In addition, the reliability of some factual information may degrade over time and that factor would need to be captured in the final answer. These are all difficult complications that were purposely (and correctly) avoided in the formulation of the TREC-8 Q&A task but which can not be avoided if a meaningful Q&A system is to be developed for Intelligence Analysts. And if these complications are not enough, each set of questions also contains some level of judgement or intent and some prediction of possible future courses of action to be rendered and included in the final answer.

So from the perspective of the Intelligence Analyst the ultimate goal of the Question and Answer paradigm would the creation of an integrated suite of analytic tools capable of providing answers to complex, multi-faceted questions involving judgement terms that analysts might wish to pose to multiple, very large, very heterogeneous data sources that may physically reside in multiple agencies and may include:

- Structured and unstructured language data of all media types, multiple languages, multiple styles, formats, etc.,
- Image data to include document images, still photographic images, and video; and
- Abstract/technical data.

A pictorial description of the data flow associated with the integrated suite of analytic tools capable of providing answers to complex questions is depicted below and then is described and explained verbally in the paragraphs following the diagram.



The Intelligence Analyst would pose his or her questions during the analysis and production phase of the Intelligence Cycle. In this case the analyst is typically pursuing known intelligence requirements and is seeking to "pull" the answer out of multiple data sources. More specifically, the components of this problem include the following:

- Accept complex "Questions" in a form natural to the analyst. Questions may include judgement terms and an acceptable answer may need to be based upon conclusions and decisions reached by the system and may require the summarization, fusion, and synthesis of information drawn from multiple sources. Analyst may supplement the "Question" with multimedia examples of information that is relevant to some aspect of the question. For example, in the first example of the Intelligence Analyst question, the analyst would need to supply an annotated example of a portion of the television broadcast that captures the image and possibly voice of the unknown senior military officer and may choose to also include that portion of the broadcast that caused the analyst to suspect that the officer was acting as an advisor.
- Translate "Complex Question" into multiple queries appropriate to the various data sets to be searched. This translation process will require the Q&A system to take into account the following information when it translates the analyst's original question into these multiple queries:

    - Information about the context of the current work environment of the analyst, the scope and nature of the intelligence requirement that prompted the current question, and the analyst's current level of understanding and knowledge of information related to this requirement.
    - Information about the nature, location, query language, and other technical parameters related to the data sources that will need to be search or queried to find relevant data. It could easily be that the analyst is totally unaware of the existence of multiple, relevance data sources, but the Q&A system still needs to generate appropriate queries against these sources. The Q&A system also needs to be capable of understanding and dealing with multiple-levels of security and need-to-know access consideration that could potentially be associated with some data sources.
    - Information obtained during the analysis and question formulation process from collaboration with other analysts (these would include analysts in the same organization, but could also include previously unknown analysts in other external organizations or even agencies) who are working closely related intelligence requirements. The information obtained could come from the analyst's personally held knowledge or from his or her working notes, papers and records. In particular, the Q&A system could propose appropriate collaboration with selected analysts based upon its knowledge and understanding of which analysts have previously posed similar and related questions.
- <u>Find relevant information in distributed, multimedia, multilingual, multi-agency data sources.</u> These multiple data sources can each be large data repositories to which a stream of new data is continuously being added or these data sources could be the original data sources against which new or modified data collection could be initiated. It is a very dynamic environment in the both the data sources and the data contained within these sources is constantly changing.

  When multiple, highly heterogeneous data sources are searched for relevant information, It is likely that significantly different retrieval and selection algorithms and weighting and ranking approaches will be used to locate relevant information in these heterogeneous data sources. In some cases it may still be possible and practical to create a single merged list of relevant information. This would seem to be the preferred outcome at this point in the process. But because of these differences, it may not be practical or even useful to merge the various ranked lists of retrieved and selected information. This might occur for example when one data source consists of text documents while another data source may consist of video only segments (e.g. surveillance or reconnaissance video).

  For ease of reference, the individual retrieved information items are referred to as "documents" regardless of their language media or type.
- <u>Analyze, fuse and summarize information into a coherent "Answer.</u> At this point the primary focus shifts to the creation of a coherent, understandable "answer" that is responsive to the originally posed question. This probably means that all information that is potentially relevant to the given question needs to extracted from each

"document". The accuracy, usability, time sensitivities and relevance of each extracted piece of information must be assessed. To the extent possible these individual information objects need to be combine and the assessments of their accuracy, usability, time sensitivities and relevance needs to be accumulated across all relevant retrieved "documents". Cross "document" summaries may be appropriate and useful at this point.

At the point where the data objects potentially containing the relevant information have been identified, an alternate or complementary approach might be to create a language and media independent representation of the relevant information and concepts that can be meaningfully extracted. That is, create a conceptual or information-based common denominator.

Inconsistencies, conflicting information, and missing data need to be noted. Proposed conclusions and answers, including multiple alternative interpretations, would need to be generated. (Note: Issues related to these topics were discussed earlier in this section.)

In all cases direct links would need to be maintained back to specific data item from which all relevant information or concept was extracted

- <u>Provide (Proposed) "Answer" to analyst in the form they want</u>. Generation of an appropriate answer may take various forms.
    - Under certain conditions and in response to particular types of questions, predetermined "Answer Templates" may be developed that are satisfactory to the analyst. In the best of worlds these Answer templates would be user defined. Maybe something along the lines of the Report Wizard in Microsoft's Access database program. This would probably be possible for the simpler factual Q&A's, even when the answer's must be achieved by combining, fusing, and summarizing factual information extracted from multiple information items.
    - For the more complex questions, the form of an appropriate answer may be too unique to be generated by a single answer template. In this case it may be possible for the Q&A system to subdivide the original question into a collection of simpler, more factually oriented subquestions. Specific subquestions may be chosen because an existing answer template can be associated with these subquestions. Then relationships that exist between the subquestions might help guide the manner in which the filled in answer templates are presented to the analyst.
- <u>Provide Multimedia Visualization and Navigation tools.</u> These tools would allow the Analyst to review the answer being proposed by the Q&A system, all of supporting information, extracted data, interpretations, conclusions, decisions, etc, and all of originally retrieved relevant "documents". The proposed answer may in fact generate additional questions or may suggest ways in which the original query (queries) need to be refined. In this case the Q&A cycle could be repeated. The analyst needs to have the ability to alter or override decisions, interpretations, and conclusions made automatically by the Q&A system and to modify the format, structure, content of the proposed answer. At some point, the analyst either rejects or discards the Q&A

system generated answer or accepts the jointly produced system/analyst answer . The analyst then moves on to other analytic and production tasks that could entail posing a new question.

The manner in which the analyst interacts and uses the Q&A system as well as the choices and changes that he or she makes needs to be automatically captured, analyzed by the Q&A system, and then used by the Q&A system to modify its future behavior. This use of this relevance feedback should permit the Q&A system to learn to be more effective in responding to future questions from this same analyst or by other analysts when asking similar or related questions.

6. QUESTION AND ANSWERING -- A MULTI-DIMENSIONAL PROBLEM

a. Observations on Classification and Taxonomy by the SMU Lasso System.

One of the best performing systems at the TREC-8 Q&A Track was the Lasso system developed by Dan Moldovan, Sanda Harabagiu, et al, Department of Computer Science and Engineering, Southern Methodist University, Dallas, Texas. Two very interesting and informative tables were included in the paper that SMU presented at TREC-8.[10] (Copies of these two tables are included immediately following this paragraph.)

- **Table 1**: As part of its Question Processing component, the Lasso system attempts to first classify each question by their type or "Q-class" (what, why, who, how, where, etc.) and then further classify the question within each type into its "Q-subclass. For example, for the Q-class "what", the Q-subclasses were "basic what", "what-who", "what-when", and "what-where". Table 1 (Types of questions and statistics) breaks down each of the 200 Q&A test question used in TREC-8 into their "Q-class" and "Q-subclass".

  I believe that this is a very useful question classification scheme, but one which will need to be significantly expanded as the Q&A task is opened up to broader classes of questions. This effort might be greatly enhanced through an effort to first methodically collect a large number of operational questions developed by real intelligence analysts working on significantly different intelligence requirements across a number of different agencies and then to systematically work towards the development of a workable question classification scheme. The perceived or actual difficulty in generating an appropriate answer should be evaluated as well for each question.

- **Table 5**: The following section is quoted from the SMU TREC-8 paper.

  "In order to better understand the nature of the QA task and put this into perspective, we offer in Table 5 a taxonomy of question answering systems. It is not sufficient to classify only the types of questions alone, since for the same question, the answer may be easier or more difficult to extract depending on how the answer is phrased in the text. Thus we classify the QA systems, not the questions. We provide a taxonomy based on three criteria that we consider important for building question answering systems: (1) knowledge base, (2) reasoning, and (3) natural language processing indexing techniques. Knowledge bases and reasoning provide the medium for building question contexts and matching them against text documents. Indexing identifies the text passages where answers may lie, and natural language processing provides a framework for answer extraction."

  In its Table 5, SMU identifies a taxonomy of 5 Classes for Q&A Systems. The degree of complexity increases from Class 1 to Class 5. Of the 153 test questions that the Lasso System correctly answered in TREC-8, 136 were assigned to Class

---

[10] Dan Moldovan, Sanda Harabagiu, et al. Lasso: A Tool for Surfing the Answer Net. TREC-8 Draft Proceedings, NIST, November 1999, pages 65-73. Table 1 : Types of questions and statistics is found on page 67 and Table 5: A taxonomy of Question Answering Systems is found on page 73 of the TREC-8 Draft Proceedings.

1 (the easiest class) and 17 were assigned to Class 2. None were assigned to the higher Classes in the SMU taxonomy. Again we believe that the Q&A research area would greatly benefit from an extensive, methodical study into the creation of separate taxonomies for both questions and answers (which SMU did not propose) as well as a similar study into a further refinement and extension of the Q&A system taxonomy that SMU has begun.

Q-class	Q-subclass	Nr. Q	Nr. Q answered	Answer type	Example of question	Focus
what		64	54			
	basic what	40	34	MONEY/NUMBER/ DEFINITION/TITLE/ NNP/UNDEFINED	*What was the monetary value of the Nobel Peace Prize in 1989?*	monetary value
	what-who	7	7	PERSON/ ORGANIZATION	*What costume designer decided that Michael Jackson should only wear one glove?*	costume designer
	what-when	3	2	DATE	*In what year did Ireland elect its first woman president?*	year
	what-where	14	12	LOCATION	*What is the capital of Uruguay?*	capital
who		47	37	PERSON/ ORGANIZATION	*Who is the author of the book "The Iron Lady: A Biography of Margaret Thatcher"?*	author
how		31	21			
	basic how	1	0	MANER	*How did Socrates die?*	Socrates
	how-many	18	13	NUMBER	*How many people died when the Estonia sank in 1994?*	people
	how-long	2	2	TIME/DISTANCE	*How long does it take to travel from Tokyo to Niigata?*	–
	how-much	3	2	MONEY/PRICE	*How much did Mercury spend on advertising in 1993?*	Mercury
	how-much- <modifier>	1	0	UNDEFINED	*How much stronger is the new vitreous carbon material invented by the Tokyo Institute of Technology compared with the material made from cellulose?*	new vitreous carbon material
	how-far	1	1	DISTANCE	*How far is Yaroslavl from Moscow?*	Yaroslavl
	how-tall	3	3	NUMBER	*How tall is Mt. Everest?*	Mt. Everest
	how-rich	1	0	UNDEFINED	*How rich is Bill Gates?*	Bill Gates
	how-large	1	0	NUMBER	*How large is the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast?*	Arctic refuge
where		22	16	LOCATION	*Where is Taj Mahal?*	Taj Mahal
when		19	13	DATE	*When did the Jurassic Period end?*	Jurassic Period
which		10	8			
	which-who	1	1	PERSON	*Which former Klu Klux Klan member won an elected office in the U.S.?*	former Klu Klux Klan member
	which-where	4	3	LOCATION	*Which city has the oldest relationship as sister-city with Los Angeles?*	city
	which-when	1	1	DATE	*In which year was New Zealand excluded from the ANZUS alliance?*	year
	which-what	4	3	NNP/ ORGANIZATION	*Which Japanese car maker had its biggest percentage of sale in the domestic market?*	Japanese car maker
name		4	4			
	name-who	2	2	PERSON/ ORGANIZATION	*Name the designer of the show that spawned millions of plastic imitations, known as "jellies"?*	designer
	name-where	1	1	LOCATION	*Name a country that is developing a magnetic levitation railway system?*	country
	name-what	1	1	TITLE/NNP	*Name a film that has won the Golden Bear in the Berlin Film Festival?*	film
why		2	0	REASON	*Why did David Koresh ask for a word processor?*	David Koresh
whom		1	0	PERSON/ ORGANIZATION	*Whom did the Chicago Bulls beat in the 1993 championship?*	Chicago Bulls
Total		200	153 77%			

Table 1: Types of questions and statistics. In this table we considered that a question was answered correctly if its answer was among top five ranked long answers.

Class	KB	Reasoning	NLP/Indexing	Examples and Comments
1	dictionaries	simple heuristics, pattern matching	complex noun, apposition, simple semantics, keyword indexing	Q33: *What is the largest city in Germany?* A: .. *Berlin, the largest city in Germany..* Answer is: simple datum or list of items found verbatim in a sentence or paragraph.
2	ontologies	low level	verb nominalization, semantics, coherence, discourse	Q198: *How did Socrates die?* A: .. *Socrates poisoned himself..* Answer is contained in multiple sentences, scattered throughout a document.
3	very large KB	medium level	advanced nlp, semantic indexing	Q: *What are the arguments for and against prayer in school?* Answer across several texts.
4	Domain KA and Classification, HPKB	high level		Q: *Should Fed raise interest rates at their next meeting?* Answer across large number of documents, domain specific knowledge acquired automatically.
5	World Knowledge	very high level, special purpose		Q: *What should be the US foreign policy in the Balkans now?* Answer is a solution to a complex, possible developing scenario.

Table 5: A taxonomy of Question Answering Systems. The degree of complexity increases from Class 1 to Class 5, and it is assumed that the features of a lower class are also available at a higher class.

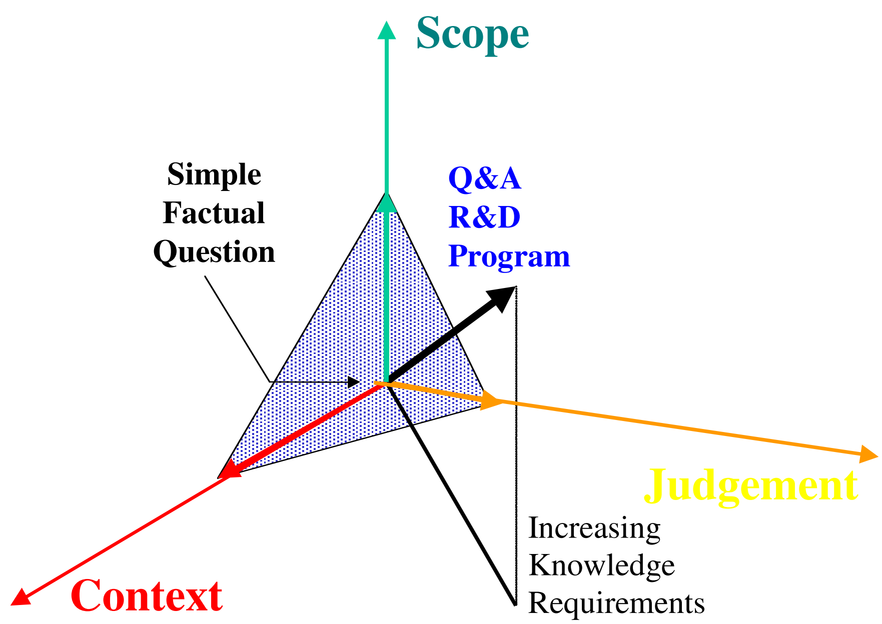
b. **Multi-Dimensionality of Both Questions and Answers.**

We have highlighted the SMU observations on both the classification of questions and on the taxonomy of the Q&A systems, because we believe that both of these efforts are significant, very insightful and useful. Unfortunately they are only preliminary. There is much more work in these areas yet to be done. We believe that these classifications and taxonomy are preliminary because the simplified nature of the Q&A task in TREC-8 masks the true multi-dimensionality of the advanced Q&A task that we have attempted to outline in the previous section.

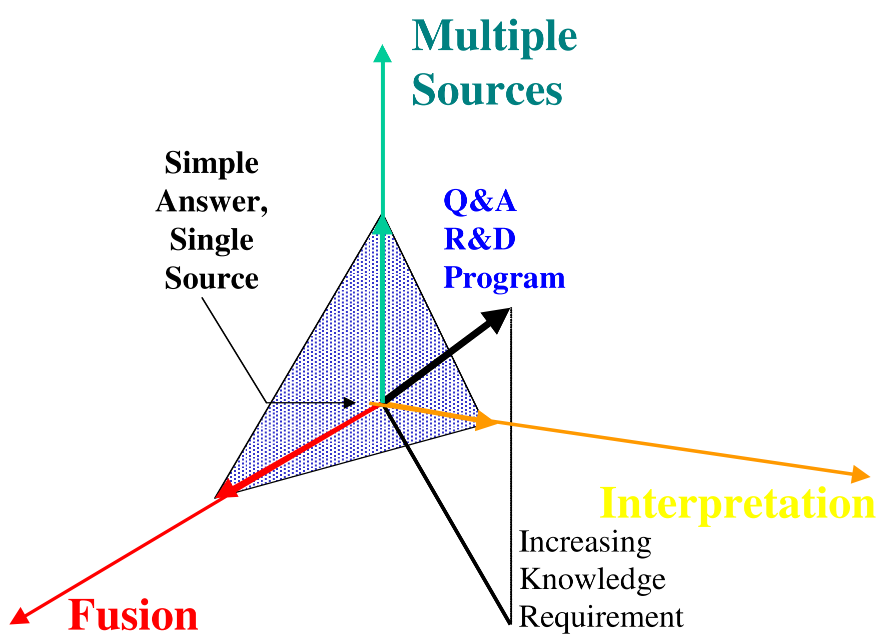
From an intelligence analyst perspective TREC-8 Q&A task placed severe restrictions on the both the dimensionality of the Questions and on the Answers. It is as if both the Q&A track questions and answers are operating very close to the origin of higher dimensional spaces.

The questions in the Q&A track were restricted to be simple factual questions. These questions required no or little context beyond the simple question statement provided. It was not necessary to know why the question was being generated. Nor was it necessary to consider expanding the any of the questions with knowledge available to the question asker but not explicitly included in the question. Also the TREC-8 Q&A questions were factual and contained no judgement terms. The scope of the questions was limited to a single 50 byte or shorter character string within the single text document. The net result was that in final analysis each document in the collection being searched for a possible answer could be analyzed independent of all other documents and that even within documents, the span of detailed searching for an answer was limited a relatively small number of connected paragraphs. In a diagram on a following page we suggest that the question part of the Q&A task has at least three important dimensions in its more general setting. One possible set of dimensions for a higher dimension question space is:

DIMENSIONS OF THE QUESTION PART OF THE Q&A PROBLEM



DIMENSIONS OF THE ANSWER PART OF THE Q&A PROBLEM



- Context (that is, in what manner and to what degree does the given question need to be expanded to adequately reflect the context in which it was asked?). For example, recall the TREC-8 question cited earlier, "Where is the Taj Mahal?" The obvious answer under most circumstances would be "Agra, India", the location of the famous architectural wonder built as a memorial to Mumtaz Mahal, a Muslim Persian princess, by her Mughal emperor husband, Shah Jahan, in the seventeenth century. But given a different context, background and interest, the questioner may consider the correct answer to be "Atlantic City, NJ" (the location of the Trump Casino, Taj Mahal) or "Bombay, India" (the location of Taj Mahal Hotel) or some other "Taj Mahal" that may exist elsewhere in the world.
- Judgement (that is, how should the question be translated into potentially multiple queries into the search space in order that the appropriate relevant information will to retrieved so that the judgement, intent, motive, etc. terms found in the original question can be adequately resolved during the answering part of this task?)
- Scope (that is, how broadly or narrowly should the question be interpreted; how wide of a search net must be cast in order to locate information relevant to the given question? Which of the available data sources are likely to contain information relevant to the given questions?)

Similarly, the answers in the Q&A Track were restricted to simple answers that could be found in a single source from a single data collection. In more advanced Q&A environments the desired answers would require the system to:

- Retrieve information from multiple, heterogeneous, multi-media, multi-lingual, distributed sources,
- Fuse, combine, and summarize smaller, individual facts into larger informational units where these facts have been extracted from multiple data items (e.g. a newspaper article, a single news broadcast story), involve multiple different language media (e.g. relational databases, unstructured text, document images, still photographs, video data, technical or abstract data), and have their language component expressed in English and multiple foreign languages. Since the Q&A system will be required to fuse together multiple, partial factual information from different sources, there is a strong likelihood that there will be conflicting facts, duplicate facts and even incorrect facts uncovered. This may result in the need to develop multiple possible alternatives, each with their own level of confidence. In addition, the reliability of some factual information may degrade over time and that factor would need to be captured in the final answer.
- Interpret the retrieved information appropriately so that the answering component can deal appropriately with the judgement, intent, motive, etc. terms found in the original question. This is clearly the dimension of the answering component that requires the greatest level of cognitive skill. It is the dimension along which progress and advances will be the hardest and slowest, but it is not one to be completely ignored when planning future R&D efforts.

Based upon this discussion, we suggest in the diagram on a preceding page that the answering part of the Q&A task has at least three important dimensions in its more general setting. One possible set of dimensions for this higher dimension question space that is based upon the preceding discussion would be:

- Multiple Sources
- Fusion
- Interpretation.

c. Increasing Requirement for Knowledge in Q&A Systems [11]

TREC-8 sponsored its inaugural Q/A Track in November 1999, which was significant for several reasons. It was a first step in the post-TIPSTER era in which retrieval and extraction technologies emerged from their individual sandboxes of MUC and TREC evaluations to play together. The stated goal in the Government's funding agreement with NIST in support of TREC was to provide a problem and forum in which traditional information retrieval techniques would be coerced into partnership with different natural language processing technologies. Having systems answer specific questions of fact ("Who was the 16th president of the U.S.?" "Where is the Taj Mahal?" etc.), was thought to be an appropriate way to accomplish this coercion, and, at the same time, set the stage for a much more ambitious evaluation.

The Q/A Track was an unqualified success in terms of achieving the above-stated goal. The top-performing systems incorporated MUC-like named-entity recognition technology in order to categorize properly the questions being asked and anticipate correctly the type of responses that would be appropriate as an answer. Traditional IR techniques alone were insufficient. In addition, participants were required to be more creative with their indexing methods. Paragraph indexing combined with Boolean operators proved to be surprisingly efficient. In short, the Q/A Track not only moved TREC out of the shadow of TIPSTER, but also helped demonstrate that returning a collection of relevant documents is a limited view of what "search strategy" means. In many instances, a specific answer to a specific question is the epitome of a user's "search."

[11] Observation: Clearly as Q&A Systems develop beyond the capabilities that they have exhibited in the TREC-8 environment, there will be increasing requirements placed on the knowledge on which these more advanced Q&A systems. In the two diagrams associated with subsection b, a single line was originally depicted as emanating from the origin out into the first octant at a 45-degree angle away from each of the three coordinate axes. This line was simply labeled as the direction of "increasing difficulty". The description in this subsection describes very well a particular aspect of this difficulty --that being the level and sophistication of the knowledge that would be required by more ambitious Q&A systems. After considering this increasing knowledge requirement section more carefully this line was relabeled as "Increasing Knowledge Requirements".(This is its current depiction.) The feeling was that the level, sophistication, and type of knowledge required by Q&A systems are jointly dependent on how far out on each of the Question axes (Content, Judgement and Scope) that the Q&A system is coupled with how far out on the Answer axes (Fusion, Interpretation, and Multiple Sources) you are as well. So because of these dependency relationships "Knowledge" has been depicted in both diagrams as a first octant vector in these two three dimensional spaces. An unanswered question is whether or not it is more meaningful to depict all knowledge as a single vector or as a set of three separate vectors -- one each for the three types that are described in this section: Explanatory, Modal and Serendipitous. In either case the depiction raises some interesting questions related to the implications of projecting any or all of these knowledge types onto any of the six Q &A dimensions that were identified in subsection b.

Mention was made above to a more ambitious evaluation. TIPSTER had many success stories, but the implicit goal of natural language understanding was not one of them. It was thought that having systems tackle tasks involving the extraction of events and relationships and fill complex template structures would a priori necessitate solving the natural language understanding problem, or at least major portions of it. Extraction developers, however, discovered that partial parsing and syntax alone (pattern recognition) would score well in the MUCs. The syntactic approach – what Jerry Hobbs' characterized as getting what the text "wears on its sleeve" – extracted entity names with human-level efficiency. However, the 60% performance ceiling on the more difficult event-centered Scenario Task is indicative of what was not successfully extracted or even attempted: Those items that crossed sentence boundaries and required robust co-reference resolution, evidentiary merging, inferencing, and world knowledge; that is, language understanding.

Semantics, in short, was put on the back burner. The contention here is that Q/A properly conceived and implemented as a mid- to long-term program (5-8 years) puts natural language understanding in the forefront of researchers and systems' developers. As Wendy Lehnert stated over 20 years ago: "Because questions can be devised to query any aspect of text comprehension, the ability to answer questions is the strongest possible demonstration of understanding." It is time to "go back to the future" (Bill Woods' LUNAR): The ultimate Q/A capability facilitates a dialogue between the user and machine, the ultimate communication problem and all that that entails. There would be many technological issues to be addressed along the way.

The TREC Q/A Track dealt with what could be called **"factual knowledge,"** answering specific questions of fact. TREC-9 proposes to reduce answer strings from 50 and 200 bytes to 20. TREC-8 avoided difficult questions of motivation and cause –Why and How – limiting most of the questions to What, Where, When, and Who. A more ambitious evaluation would require moving beyond questions of fact to **"explanatory knowledge."**

Anyone who has considered a typology[12] of questions readily understands that questions are not always in the form of a question. But answers sometimes are. Consider a telephone conversation that begins with the pseudo-question "Well, I sort of need your name to handle this matter." Since the questioner obviously is eliciting a name, a rhetorical response might be "What, you expect me to provide personal information?" Also, sometimes the person with the answers asks the questions. The previous examples

[12] The examples provided here are discussed in terms of lexical categories which only scratches the surface of any proposed typology. Wendy Lehnert discusses 13 conceptual question categories in her taxonomy. ("The Process of Question Answering," LEA 1978) In the context of MUC where slot fills in general answer the question Who did What to Whom When Where and sometimes How, the How question would most likely be considered in terms of **instrumentality**, e.g., How did the dignitary arrive? (By what means did he come?), or **causal antecedent**, e.g., How did the F-15 crash? (What caused the plane to crash?). These categories may cover the predominant number of cases in a bounded MUC task, but not in an open domain. After Lehnert, one would need to deal with such categories as: **Quantity** – How often does it rain in Seattle? (Everyday. Twice a day. Most evenings between November and May.); **Attitude** – How do you like Seattle? (It's fine. I like rain. I haven't seen anything yet.); **Emotional/Physical State** – How is John? (A bit queasy after the 6-hour flight in coach.); **Relative Description** – How smart is John? (Not smart enough to fly business.); **Instructions –** How does one get to Seattle? (Take a left onto Constitution and go straight for 3,000 miles.)

point to other challenges in turn. The problem is not simply one of dealing with more difficult questions; the data source is not limited to narrative texts or factual news articles. Speech is also a part of the ultimate Q/A terrain. Furthermore, the “search strategy” cannot be thought of only as one user trying to acquire answers from a system. Communication in the “real world” does not work that way. Communication is interactive and necessitates clarification upon clarification. The same should be true for our Q/A framework.

Further analysis of the question typology reveals the sort of knowledge characterized by DARPA’s High-Performance Knowledge Base (HPKB) Program: **“Modal knowledge.”** The program’s knowledge base technology employed ontologies and associated axioms, domain theories, and generic problem-solving strategies to handle proactively crisis management problems. In order to evaluate possible political, economic, and military courses of action before a situation becomes a crisis, one needs to ask such hypothetical questions as “What might happen if Faction Leader X dies?” Or, “What if NATO jets strafe Serb positions along Y River?” More difficult still are counterfactual questions such as “What if NATO jets do not strafe Serb positions along Y River?” In addition to the problems surrounding knowledge base technology like knowledge representation and ontological engineering, questions of this type require sophisticated logical and probabilistic inference techniques with which the text-processing community has not traditionally been involved. Since knowledge bases form a repository of possible answers – factual, explanatory, or modal -- more would have to be done in this area. Needless to say, there would be a need for greater cooperation between the text-processing and knowledge base communities.

Other requirements: Text and speech are not the only sources of information. Databases abound on the Web and in any organizational milieu. A Q/A system must not require the user to be educated in SQL, but enable the user to query any data source in the same natural and intuitive manner. The system should also handle misspellings, ambiguity (“What is the capital of Switzerland?”), world knowledge (“What is the balance on my 401K plan?”). Other data -- transcribed speech and OCR output – is messier. And, most of the world’s data is not in English. TIPSTER demonstrated that retrieval and extraction techniques could be ported to foreign languages. This being the case, a TREC-like Q/A track for answering questions of fact in foreign languages could be initiated soon, especially in those languages for which extraction systems possess a “named-entity” capability.

In the new world of e-Commerce and knowledge management, the term “infomediary” has come to replace “intermediary” in the old world of bricks and mortar and the physical distribution of goods. A Q/A system viewed as a search tool would require an infomediary to handle the profiling function of traditional IR systems. One can envision perhaps intelligent agents that would run continuously in the background gathering relevant answers to the original query as they become available. In addition, these agents would provide ready-made Q/A templates, generated automatically, and stored in a FAQ archive. This archive would actually constitute a knowledge base, a repository of information to retain and augment cooperate/analytical expertise.

One does not want list all, but rather suggest some of the techniques that may be useful in accomplishing natural language understanding. We've already mentioned linguistic pragmatics. But there are a host of other areas in pragmatics: discourse structure, modeling speakers' intents and believes, situations and context – in short, everything that theoretical linguists have not idealized in terms of what it is to communicate and what Carnap would have considered to have been idiosyncratic. In the area of IR, certainly conceptually based indexing and value filtering would be an important component to any Q/A system. Also, applying statistical or probabilistic methods in order to get beyond pattern recognition by grasping concepts beneath the surface (syntax) of text is a promising area of research; and, not only for the text processing community. The same techniques might help tackle the thorny issue identified in HPKB that language-use linguists following Wittgenstein have understood for years: Many words or concepts are not easily defined because they change canonical meaning in communicative situations and according to context and their relationship with other terms. Statistical and probabilistic methods can be trained to look specifically at these associative meanings.

Finally, there is a third type of knowledge we will label "**serendipitous knowledge.**" One could equate this search strategy to a user employing a search engine to browse document collections. However, a sophisticated Q/A system coupled with sophisticated data mining, text data mining, link analysis, and visualization/navigation tools would transform our Q/A system into the ultimate man-machine communication device. These tools would provide users with greater control over their environment A query on, e.g. Chinese history, might lead one to ask questions about a Japanese school of historiography because of unexpected links that the "knowledge discovery" engine discovered, and proffered to the user as valuable path of inquiry. That path of inquiry, moreover, could be recorded for future use, and traversed by others for similar, related, or even different reasons.

The same function – suggesting questions – is the final piece of the Q/A system in its interactive, communicative mode. So many times the key to the answer is asking the right question. One thinks of the genius of a philosopher such as Kant who asked questions of a more fundamental and powerful sort than his predecessors, challenging assumptions that had hitherto gone unchallenged. One wonders about the conventional question "How do we use technology?" Asking instead "How does technology use us?" is more than semantic chicanery. Perhaps what we needed to think about in the early 20th Century was not how to drive cars, but what they would do to our air, landscape, social relations, cities, etc. More and more frequently one hears the call for analysts to "think out of the box." Perhaps the ultimate Q/A system is one way to compensate for an individual's limitations in terms of experience and expertise; another tool for thinking, not just searching.

d. **Final Observation on Q&A.**

A final observation about these two diagrams. On each diagram we have depicted a Q&A R&D program as a plane that cuts across all three dimensions of each diagram. Clearly we are moving in the direction of increasing difficulty as this plane is moved further away from the origin. But it is our belief and the contention of this vision paper that this is exactly the direction in which the roadmap for Q&A should envision the R&D community moving. How far away from the origin along all three axes we should move the R&D plane and how rapidly can we discover technological solutions along such a path? These are exactly the questions that the Q&A Roadmap Committee should consider and deliberate.

7. MULTIDIMENSIONALITY OF SUMMARIES [13]

Summarizing is a complex task involving three classes of factor: the nature of the input source, that of the intended summary purpose, and that of the output summary. These factors are many and varied, and are related in complicated ways. Thus summary system design requires a proper analysis of such major factors as the following:

Input: Characteristics of the source text(s) include:

- *Source size*: Single-document *vs.* Multi-document. A single-document summary derives from a single input text (though the summarization process itself may employ information compiled earlier from other texts). A multi-document summary is one text that covers the content of more than one input text, and is usually used only when the input texts are thematically related.
- *Specificity*: Domain-specific *vs.* General. When the input texts all pertain to a single domain, it may be appropriate to apply domain-specific summarization techniques, focus on specific content, and output specific formats, compared to the general case. A domain-specific summary derives from input text(s) whose theme(s) pertain to a single restricted domain. As such, it can assume less term ambiguity, idiosyncratic word and grammar usage, specialized formatting, etc., and can reflect them in the summary. A general-domain summary derives from input text(s) in any domain, and can make no such assumptions.
- *Genre and scale*. Typical input genres include newspaper articles, newspaper editorials or opinion pieces, novels, short stories, non-fiction books, progress reports, business reports, and so on. The scale, often correlated with the genre, may vary from paragraph-length to book-length. Different summarization techniques may apply to some genres and scales and not others.
- *Language*: Monolingual *vs.* Multilingual. Most summarization techniques are language-sensitive; even word counting is not as effective for agglutinative languages as for languages whose words are not compounded together. The amount of word separation, demorphing, and other processing required for full use of all summarization techniques can vary quite dramatically.

Purpose: Characteristics of the summary's function include: (Note that these are the most critical constraints on summarizing, and the most important for evaluation of summarization system output.)

- *Situation*: Tied vs. Floating. Tied summaries are for a very specific environment where the who by, what for, and when of use is known in advance so that summarizing can be tailored to this context; for example, product description summaries for a particular sales drive. Floating situations lack this precise context specification, e.g. summaries in technical abstract journals are not usually tied to predictable contexts of use.

[13] For additional background on summarization, the reader is directed to "Advances in Automatic Text Summarization"; Inderjeet Mani and Mark Maybury, editors; MIT Press; 1999.

- *Audience*: Targeted vs. Untargeted: A targeted readership has known/assumed domain knowledge, language skill, etc, e.g. the audience for legal case summaries. A (comparatively) untargeted readership has too varied interests and experience for fine tuning e.g. popular fiction readers and novel summaries. (A summary's audience need not be the same as the source's audience.)
- *Use*: What is the summary for? This is a whole range including uses as aids for retrieving source texts, as means of previewing texts about to be read, as information-covering substitutes for source texts, as devices for refreshing the memory of an already-read sources, as action prompts to read their sources. For example a lecture course synopsis designed for previewing the course may emphasize some information e.g. course objectives, over others.

Output: Characteristics of the summary as a text include:

- *Derivation*: Extract *vs.* Abstract. An extract is a collection of passages (ranging from single words to whole paragraphs) extracted from the input text(s) and produced verbatim as the summary. An abstract is a newly generated text, produced from some computer-internal representation that results after analysis of the input.
- *Coherence*: Fluent *vs.* Disfluent. A fluent summary is written in full, grammatical sentences, and the sentences are related and follow one another according to the rules of coherent discourse structure. A disfluent summary is fragmented, consisting of individual words or text portions that are either not composed into grammatical sentences or not composed into coherent paragraphs.
- *Partiality*: Neutral *vs.* Evaluative. This characteristic applies principally when the input material is subject to opinion or bias. A neutral summary reflects the content of the input text(s), partial or impartial as it may be. An evaluative summary includes some of the system’s own bias, whether explicitly (using statements of opinion) or implicitly (through inclusion of material with one bias and omission of material with another).

The explicit definition of the purpose for which system summaries are required, along with the explicit characterization of the nature of the input and consequent explicit specification of the nature of the output are all prerequisites for evaluation. Developing proper and realistic evaluations for summarizing systems is then a further material challenge.

Appendix 1

PROGRAM DESCRIPTIONS

Four multifaceted research and development programs share a common interest in a newly emerging area of research interest, Question and Answering, or simply Q&A and in the older, more established text summarization.

These four programs and their Q&A and text summarization intersection are the:

- Information Exploitation R&D program being sponsored by the Advanced Research and Development Activity (ARDA). The "Pulling Information" problem area directly addresses Q&A. This same problem area and a second ARDA problem area "Pushing Information" includes research objectives that intersect with those of text summarization. (John Prange, Program Manager)
- Q&A and text summarization goals within the larger TIDES (Translingual Information Detection, Extraction, and Summarization) Program being sponsored by the Information Technology Office (ITO) of the Defense Advanced Research Project Agency (DARPA) (Gary Strong, Program Manager)
- Q&A Track within the TREC (Text Retrieval Conference) series of information retrieval evaluation workshops that are organized and managed by the National Institute of Standards and Technology (NIST). Both the ARDA and DARPA programs are providing funding in FY2000 to NIST for the sponsorship of both TREC in general and the Q&A Track in particular. (Donna Harman, Program Manager)
- Document Understanding Conference (DUC). As part of the larger TIDES program NIST is establishing a new series of evaluation workshops for the text understanding community. The focus of the initial workshop to be held in November 2000 will be text summarization. In future workshops, it is anticipated that DUC will also sponsor evaluations in research areas associated with information extraction. (Donna Harman, Program Manager)

As further background information a short description is provided of each of the first three R&D Program.

a. **ARDA's Information Exploitation R&D Thrust**

The Advanced Research and Development Activity (ARDA) in Information Technology was created as a joint activity of the Intelligence Community and the Department of Defense (DoD) in late November 1998. At this time the Director of the National Security Agency (NSA) agreed to establish, as a component of the NSA, an organizational unit to carry out the functions of ARDA. The primary mission of ARDA is to plan, develop and execute an Advanced R&D program in Information Technology, which serves both the Intelligence Community and the DoD. ARDA's purpose is to incubate revolutionary R&D

for the shared benefit of the Intelligence Community and DoD. ARDA originates and manages R&D programs which:

- Have a fundamental impact on future operational needs and strategies;
- Demand substantial, long-term, venture investment to spur risk-taking;
- Progress measurably toward mid-term and final goals; and
- Take many forms and employ many delivery vehicles.

Beginning in FY2000, ARDA established a multi-year, high risk, high payoff R&D program in Information Exploitation that when fully implemented will focus against three operationally-based information technology problems of high interest to its Intelligence Community partners. Within the ARDA R&D community these problems are referred to as "Pulling Information", "Pushing Information" and "Navigating and Visualizing Information". It is the "Pulling Information" problem that is aimed squarely at developing an advance Q&A system. The ultimate goals for each of these three problem focuses are:

- "Pulling Information": To provide supported analysts with an advanced question and answer capability. That is, starting with a known requirement, the analyst would submit his or her questions to the Q&A system which in turn would "pull" the relevant information out of multiple data sources and repositories. The Q&A system would interpret this "pulled" information and would then provide an appropriate, comprehensive response back to the analyst in the form of an answer.
- "Pushing Information": To develop a suite of analytic tools that would "push information" to an analyst that he or she had not asked for. This might involve the blind or unsupervised clustering or deeper profiling of massive heterogeneous data collections about which little is known. Or it might involving moving present day data mining techniques into the realm of incomplete, garbled data or in novelty detection where we might uncover previously undetected patterns of activity of significant interest to the Intelligence Community. Or it might involve providing alerts to an analyst when significant changes have occurred within newly arrived, but unanalyzed massive data collections when compared against previously analyzed and interpreted baseline data. And tying it all together, it might involve creating meaningful ways of portraying linked, clustered, profiled, mined data from massive data sources.
- "Navigating and Visualizing Information": To develop a suite of analytic tools that would assist an analyst in taking all of the small pieces of information that he or she has collected as being potential relevant to a given intelligence requirement, and then creating an appropriate information space (potentially tailored to the needs of either the analyst or current situation) through which the analyst can easily "navigate" while exploring the assemble information as a whole. Using visualization tools and techniques the analyst might seek out previously unknown links and connections between the individual pieces, might test out various hypotheses and potential explanations or might look for gaps and inconsistency. But in all cases the analyst is using these "navigating and visualizing" tools to help put the relevant pieces of the requirements puzzle together into a larger, more comprehensive mosaic in preparation for producing some type of intelligence report.

More information on ARDA and its current R&D Programs will be available very shortly (hopefully by the end of April 2000) on the following Internet website:

www.ic-arda.org

b. **DARPA's TIDES PROGRAM** [14]

Within DARPA's Information Technology Office (ITO), the TIDES Program is a major, multiyear R&D Program directed at **T**ranslingual **I**nformation **D**etection, **E**xtraction, and **S**ummarization. The TIDES program's goal is to develop the technology to enable an English-speaking U.S. military user to access, correlate, and interpret multilingual sources of information relevant to real-time tactical requirements, and to share the essence of this information with coalition partners. The TIDES program will increase users' abilities to locate and extract relevant and reliable information by correlating across languages and will increase detection and tracking of nascent or unfolding events by analyzing the original (foreign) language(s) reports at the point of origin, as "all news is local." The accomplishment of the TIDES goals will require advances in component technologies of information retrieval, machine translation, document understanding, information extraction, and summarization, as well as the integration technologies which fuse these components into an end-to-end capability yielding substantially more value than the serial staging of the component functions. The mission extends to the rapid development of retrieval and translation capability for a new language of interest. Achievement of this goal will enable rapid correlation of multilingual sources of information to achieve comprehensive understanding of evolving situations for situation analysis, crisis management, and battlespace applications.

The TIDES Program has included the following among its challenges:

- Exhaustive search is typically expected, in order to avoid missing a key fact, event, or relationship.
  - Exhaustive search is however a very inefficient process
- Most information is in text:
  - www, newswire, cables, printed documents, OCR'd paper documents, transcribed speech.
  - It is impossible exhaustively read and comprehend all of the available text from critical information sources
- Critical information sources occur in unfamiliar languages
  - There are always many simmering pots… it is unpredictable which will heat up
  - For example, there are over 70 languages of critical interest in PACOM's area of responsibility
  - And there are over 6,000 languages in the world
- Commercial machine translation is inadequate
  - Essentially non-existent (not commercially viable) for all but the major world languages (e.g., Arabic, Chinese, French, German, Italian, Japanese, Korean,

[14] The information in this section was extracted from the DARPA TIDES Program website located on the Internet at: http://www.darpa.mil/ito/research/tides/. More information on TIDES is available at this same website.

Portuguese, Russian, Spanish) Very low quality for languages unlike English (e.g., Arabic, Chinese, Japanese, Korean)

In order to better understand the perspective and focus of the TIDES program the following hypothetical scenario was extracted from the TIDES website.

> "A crisis has erupted in Northern Islandia that is disrupting the economic and political stability of the region. While Northern Islandia has never attracted much attention within the Department of Defense, its proximity to areas of vital interest make its current unrest a cause for concern. Unfortunately, there is no machine translation capability into English for Islic, the native language of the indigenous population, and there are very few individuals available to the Defense community who have any proficiency in the native language. While information available from neighboring sources, in languages for which machine translation is available, is providing a degree of insight into the unfolding situation, the primary sources of information are the impenetrable Islic web pages. Using TIDES technologies, information retrieval systems are adapted within a week to be able to retrieve Islic materials using English queries. Within a month, these systems have also progressed to the point where topics can be tracked, named entities (people, places, organizations, …) can be identified and correlated, and coherent summaries generated. Integrated into the Army's FALCON-2 (Forward Area Language Conversion - 2) mobile OCR units enables analysts additional access to local printed materials. This rapid adaptation into Islic now enables analysts to track events directly based both on information retrieved from the Web and in situ documents. The situation in Northern Islandia, while still critical, is no longer as vexing. Analysts now can identify the issues at stake and the stakeholders, leading to informed decision options for the Commanders in Chief."

More information on TIDES can be found at its' Internet website:
http://www.darpa.mil/ito/research/tides/

c. **NIST's TREC Program**[15]

The Text REtrieval Conference (TREC), co-sponsored by the National Institute of Standards and Technology (NIST), by DARPA, and beginning in 2000 by ARDA, was started in 1992 as part of the TIPSTER Text program. Its purpose is to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. In particular, the TREC workshop series has the following on-going goals:

- to encourage research in information retrieval based on large test collections;
- to increase communication among industry, academia, and government by creating an open forum for the exchange of research ideas;
- to speed the transfer of technology from research labs into commercial products by demonstrating substantial improvements in retrieval methodologies on real-world problems; and

[15] The information in this section was extracted from the NIST TREC Program website located on the Internet at: http://trec.nist.gov/. More information on TREC is available at this same website.

- to increase the availability of appropriate evaluation techniques for use by industry and academia, including development of new evaluation techniques more applicable to current systems.

TREC is overseen by a program committee consisting of representatives from government, industry, and academia. For each TREC, NIST provides a test set of documents and questions. Participants run their own retrieval systems on the data, and return to NIST a list of the retrieved top-ranked documents. NIST pools the individual results, judges the retrieved documents for correctness, and evaluates the results. The TREC cycle ends with a workshop that is a forum for participants to share their experiences.

This evaluation effort has grown in both the number of participating systems and the number of tasks each year. Sixty-six groups representing 16 countries participated in TREC-8 (November 1999). The TREC test collections and evaluation software are available to the retrieval research community at large, so organizations can evaluate their own retrieval systems at any time. TREC has successfully met its dual goals of improving the state-of-the-art in information retrieval and of facilitating technology transfer. Retrieval system effectiveness has approximately doubled in the seven years since TREC-1. The TREC test collections are large enough so that they realistically model operational settings, and most of today's commercial search engines include technology first developed in TREC.

In recent years each TREC has sponsored a set of more focused evaluations related to information retrieval. Each track focuses on a particular subproblem or variant of a retrieval task. Under its Track banner, TREC sponsored the first large-scale evaluations of the retrieval of non-English (Spanish and Chinese) documents, retrieval of recordings of speech, and retrieval across multiple languages. In the TREC-8 (1999) a new Question Answering Track was established. The Q&A Track will continue as one of the Tracks in TREC-9 (2000).

Question and Answering Track [16]

The Question and Answering Track of TREC is designed to take a step closer to *information* retrieval rather than *document* retrieval. Current information retrieval systems allow us to locate documents that might contain the pertinent information, but most of them leave it to the user to extract the useful information from a ranked list. This leaves the (often unwilling) user with a relatively large amount of text to consume. There is an urgent need for tools that would reduce the amount of text one might have to read in order to obtain the desired information. This track aims at doing exactly that for a special (and popular) class of information seeking behavior: QUESTION ANSWERING. People have questions and they need answers, not documents.

---

[16] Information in this section was extracted from the Q&A Track website located on the Internet at the following address: http://www.research.att.com/~singhal/qa-track.html. More information on the Q&A Track is available at this same website and at the TREC website at: http://trec.nist.gov/.

As a first attack on the Q&A problem, the Q&A Track in TREC-8 devised a simple task: Given 200 questions, find their answers in a large text collection. No manual processing of questions/answers or any other part of a system is allowed in this track. All processing must be fully automatic. The only restrictions on the questions are:

- The exact answer text *does* occur in some document in the underlying text collection.
- The answer text is less than 50 bytes.

Some example questions and answers are:

Q: Who was Lincoln's Secretary of State?
A: William Seward

or

Q: How long does it take to fly from Paris to New York in a Concorde?
A: 3 1/2 hours

The data for this task in TREC-8 consisted of approximately 525K documents from the Federal Register (94), London Financial Times (92-94), Foreign Broadcast Information Service (96), and Los Angeles Time (89-90). The collection was approximately 1.9 GB in size.

In TREC-8, participants were required to provide their top five ranked responses for each of 200 questions. Each response was a fixed length string (either a 50 or 250 byte string) extracted from a single document found in the collection. Multiple human judges scored all responses submitted to NIST. If the correct answer was found among any the top five responses the system's score for that question was the reciprocal of the highest ranked correct response. The system received a score of zero for any question for which the correct answer was not found among the top five ranked responses. The system's score for an entire run was the average of its question scores across all 198-test questions. (Two of the 200 original questions were belatedly deleted from the test set.) Twenty different organizations from around the world participated in thisTREC-8 Q&A Track and a total of 45 different runs were submitted by these organizations and evaluated by NIST. The highest scoring Q&A system on runs using a 50 byte window was developed by Cymfony of Williamsville, New York (Run Score: 66.0%; Correct answer was not found in 54 of the 198 questions.) The highest scoring system using a 250 byte window was developed by Southern Methodist University, of Dallas, Texas (Run Score: 64.5% ; Correct answer was not found in 45 of the 198 questions.)

More information on TREC along with detailed results on all evaluations conducted during TREC-8 can be found at the following Internet web site:

http://trec.nist.gov/

More information on the Q&A Track can be found at the following Internet web site:

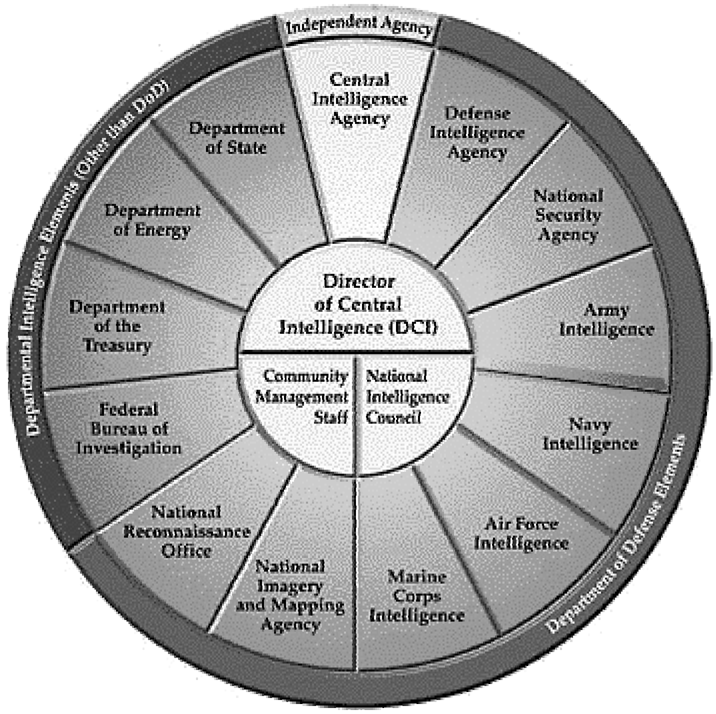
http://www.research.att.com/~singhal/qa-track.html

Appendix 2

<u>Intelligence Community</u>

The Intelligence Community is a group of 13 government agencies and organizations that carry out the intelligence activities of the United States Government. The figure below graphically depicts the Intelligence Community. In the reader is interested he or she is directed to the following Internet web site for the Intelligence Community:

http://www.odci.gov/ic/



INTELLIGENCE COMMUNITY
OF THE US GOVERNMENT

The Intelligence Community is headed by the Director of Central Intelligence (DCI), who also leads the Central Intelligence Agency (CIA), one of 13 members of the Community.

Appendix 3

Intelligence Cycle

The analysis and production activity that our Intelligence Analysts must perform is just one element with a larger, more general process called the Intelligence Cycle. In order to understand the perspective of intelligence analysts with respect to the Q&A task, we believe the reader needs to have an appreciation of the larger process and environment in which this Q&A task is to be performed. [17]

INTELLIGENCE CYCLE

The Intelligence Cycle is the process of developing raw information into finished strategic intelligence for the National Command Authority (e.g. President, his aids, National Security Council, Cabinet Secretaries, etc.) for national level policy and decision making, into operational intelligence for major military commanders and forces to use in the planning and execution of military operations of all types and sizes, and into tactical intelligence for use by tactical level military commanders who must plan and conduct battles and engagements. There are five steps that constitute the Intelligence Cycle. These same five steps are followed at all three intelligence levels (strategic, operational

[17] The description of the Intelligence Cycle was produced by blending together description of the Intelligence Cycle found in "Intelligence Support to Operations", J-7 (Operational Plans and Interoperability Directorate), Joint Chiefs of Staff, March 2000 with the description found on the following internet web site: http://www.odci.gov/cia/publications/factell/intcycle.html.

and tactical), by organizations ranging from large, national level intelligence agencies to the intelligence sections of the smallest military unit.

Two additional observations before turning to the Intelligence Cycle itself. First, the Intelligence Cycle is a highly simplified model of intelligence operations in terms of five broad, general steps. As a model, it is important to note that intelligence actions do not always follow sequentially through this cycle. However the intelligence cycle does present intelligence activities in a structured manner that captures the environment and ethos of the overarching intelligence process. Second, it is vitally important to recognize the clear and critical distinction between information and intelligence. Information is data that have been collected but not further developed through analysis, interpretation, or correlation with other data and intelligence. It is the application of analysis that transforms information into intelligence. They are not the same thing. In fact they have very different connotations, applicability and credibility.

Step 1: Planning and Direction

This step covers the management of the entire effort, from identifying the need for data to delivering an intelligence product to a consumer. It is the beginning and the end of the cycle--the beginning because it involves drawing up specific collection requirements and the end because finished intelligence, which supports policy decisions and hopefully satisfies an existing requirement, may also generate new requirements. The whole process depends on guidance from public officials and military commanders. Policymakers--the President, his aides, the National Security Council, and other major departments and agencies of government—and Military Commanders—Secretary of Defense, Chairman of Joint Chief of Staff, combatant commanders (CINCs) and other commanders and forces--initiate requests for intelligence. These requests for intelligence can be on going, standing requirements or very specific, time-sensitive requests.

Once generated, this phase of the Intelligence Cycle also matches requests for intelligence with the appropriate collection capability. It synchronizes the priorities and timing of collection with the required-by-times associated with the requirement. Collection planning registers, validates, and prioritizes all collection, exploitation, and dissemination requirements. It results in requirements being tasked or submitted to the appropriate organic, attached, and supporting external organizations and agencies.

Step 2: Collection

Intelligence sources are the means or systems used to observe, sense, and record or convey raw data and information on conditions, situations, and events. There are six primary intelligence disciplines: imagery intelligence (IMINT), human intelligence (HUMINT), signals intelligence (SIGINT), measurement and signature intelligence (MASINT), technical intelligence (TECHINT), and open-source intelligence (OSINT).

During the collection phase, those intelligence sources identified during collection planning (described above) collect the raw data and information needed to produce finished intelligence.

Collection may be both classified and unclassified. In almost all cases both the specific means/methods/locations of collection and the collected information itself are classified. But collection does also includes the overt gathering of information from open sources such as foreign broadcasts, newspapers, periodicals, and books.

Step 3: Processing

During this step, the raw data obtained during the collection phase is converted into forms that can be readily used by intelligence analysts in the analysis and production phase. Processing actions include initial interpretation, signal processing and enhancement, data conversion and correlation, transcription, document translation and decryption. Processing includes the filtering out of unwanted or unusable data, decisions on the routing and distribution of the processed data from the point of collection to analytic organizations and to individual analysts or to data repositories for possible retrieval by an analyst at a later date. Processing may be performed by the same element that collected the information or by multiple elements in multiple, separate steps. By the end of processing the final product may have been significantly altered from its original raw data state at the time and point of collection, but it is still basic information and not intelligence.

Step 4: Analysis and Production

Analysis and Production is the conversion of basic information into finished intelligence. It includes integrating, evaluating, and analyzing all available data--which is often fragmentary and even contradictory--and preparing intelligence products. Analysts, who are subject-matter specialists, consider the information's reliability, validity, and relevance. They integrate data into a coherent whole, put the evaluated information in context, and produce finished intelligence that includes assessments of events and judgments about the implications of the information for the United States.

The national level agencies within the Intelligence Community devotes the bulk of their resources to providing strategic intelligence to policymakers. It performs this important function by monitoring events, warning decision-makers about threats to the United States, and forecasting developments. The subjects involved may concern different regions, problems, or personalities in various contexts--political, geographic, economic, military, scientific, or biographic. Current events, capabilities, and future trends are examined.

These national level intelligence agencies produce numerous written reports, which may be brief--one page or less--or lengthy studies. They may involve current intelligence, which is of immediate importance, or long-range assessments. Some finished intelligence reports are presented in oral briefings. The CIA also participates in the drafting and production of National Intelligence Estimates, which reflect the collective judgments of the Intelligence Community.

Step 5: Dissemination

The last step, which logically feeds into the first, is the distribution of the finished intelligence to the consumers, the same policymakers whose needs initiated the intelligence requirements. Finished intelligence is hand-carried daily to the President and key national security advisers. The policymakers, the recipients of finished intelligence, then make decisions based on the information, and these decisions may lead to the levying of more requirements, thus triggering the Intelligence Cycle.