Toward a Question Answering Roadmap

Mark T. Maybury[1]
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730
maybury@mitre.org

Abstract

Growth in government investment, academic research, and commercial question answering (QA) systems is motivating a need for increased planning and coordination. The internationalization of QA research, and the need to move toward a common understanding of resources, tasks and evaluation methods provided motivate a need to facilitate more rapid and efficient progress. This paper characterizes a range of question answering systems and provides an initial roadmap for future research, including a list of existing resources and ones under development. This roadmap was initiated at the LREC 2002 Workshop on QA and we propose to update the roadmap during the workshop with moderated group brainstorming sessions.

Characterizing Question Answering Systems

Figure 1 characterizes a range of characteristics of question answering systems. The set of dimensions the distinguish various question answering systems which might range from systems for on-line help to access encyclopedic or technical manual information, to open web-based question answering, to very sophisticated QA in support of business or military intelligence analyses. Characteristics that distinguish QA environments include but are not limited to:

- the nature of the query, including the question form (e.g., keyword(s), phrase(s), full question(s)) the question type (e.g., who, what, when, where, how, why, what-if), and the intention of the question (e.g., request, command, inform).
- the level of complexity of the question and answer,
- characteristics of the source(s) and/or supporting corpora (e.g., size, dynamicity, quality),
- properties of the domain and/or task (e.g., degree of structure, complexity),
- the potential for answer reuse,
- the degree of performance required (e.g., precision and recall),
- the nature of the users (e.g., age, expertise, language proficiency, degree of motivation) and the important of usability,
- the purposes of the users (e.g., help with homework or cooking, strategic analysis),

---

- nature of supporting knowledge sources (e.g., degree of necessary linguistic, world knowledge)
- reasoning requirements (e.g., inference required for question analysis, answer retrieval, presentation generation)
- the degree of multilinguality and cross linguality (e.g., questions might be asked in one language),
- the user model (e.g., stereotypical vs. individualized user models)
- the task model (e.g., structured vs. unstructured tasks)
- the type of answers provided (e.g., named entities, phrases, factoid, link to document summary)
- the nature of interaction (e.g., user reactivity, mixed initiative, question and answer refinement, answer justification)

Figure 1 distinguishes question answering systems by various characteristics. For example, we can have QA from a selected document collection as in the Text Retrieval (TREC) QA track, retrieval of answers from semi-structured sources such as dictionaries, encyclopaedia or fact books, QA from massive, unstructured sources such as the web, and multimedia QA. As Figure 1 shows, there is a range of question/answer complexity, corpus volume, and degree of answer integration. Systems may address a variety of question forms (e.g., keyword, phrase, question) and types (e.g., who, what, why). Questions might encode a range of intentions such as a request for information, a command to perform some action such as a calculation, or also even information within the question (e.g., "What type of Titleist balls does Tiger Woods use?"). The answers might come in the form of a named entity, a phrase, a factoid, a link to a document or documents, or a generated summary. Additional characteristics include the degree of world knowledge in the system, its use of context and support for QA dialogue, if it has a user model and its nature (e.g., stereotypical, individualized, overlay), its task model, the structure of the domain, the degree of answer reuse in the system, and the degree of expected performance.

Kinds of QA	Question/ Answer Complexity	Source Volume, Quality	Corpus, Resource Model	Answer Integration & Generation	Type of User Query	Type of System Answers
TREC QA **On-line dictionaries, encyclopedias** **On line manuals (e.g,. UC)** **Web QA (e.g., Kiwilogic linguabot, e-Gain mail, Firepond, IONaut and NSIR)** **Multimedia, Multilingual QA**	Moderate Q., Easy A.	small (100s MB), static, high quality source	Encyclo-pedic technical manuals	easy	FORM: - Keyword(s) - Phrase(s) - Question(s)	- Named entities - Phrase - Factoid - Link to document(s) - Summary
	Easy to Moderate Q., Moderate A.	small to high (10 GB), dynamic, variable quality sources	Web	moderate	TYPE: - Who - What - When - Where - How - Why - What-if	
	Hard Q., Hard A. [Interacts w/ Multilinguality]	very high, real-time, streaming, dynamic variable quality sources	Varied Multi-linguality (in Response and in Query)	hard	INTENT: - request - command - inform	

Figure 1. Question Answering Characteristics

Question Answering Roadmap

Figure 2 is a roadmap jointly created by participants of the LREC Q&A workshop. The roadmap is divided into three lanes dealing with resources necessary to develop or evaluate QA systems, methods and algorithms, and systems (including their performance and evaluation). The roadmap starts now and runs until 2006. Each lane leads to outcomes (indicated by sign posts) such as measurable progress from having shared resources, a composable QA toolkit, and personalized QA. An overall, long term outcome of QA systems that become high quality and enhance productivity.

Sign posts along the road indicate intermediate outcomes, such as a typology of users, a topology of answers, a model of QA tasks (from both a system and user perspective), QA reuse across sessions, and interactive dialogue. Roadblocks along the way include the need to manage and possibly retrain user expectations, the need for reusable test collections and the need for evaluation methods. Overall workshop participants felt that general natural language processing and inference were limiters to progress, and so these were represented as speed limits signs on the left hand side of the road map. Here also we can see an arrow that indicates that feasibility testing and requirements determination are continuous processes along the road to productive, quality QA.

On the right hand side of the road map we can see the progression of question and answer types. Questions progress from simple factoid questions to how to why then to what-if questions, whereas answers start out as simple facts but move to scripted or templated answers and then progress further to include multimodal answers.

Related fields such as high performance knowledge bases (HPKB), topic detection and tracking (TDT), databases, virtual reference desks, and user modeling were noted as having particular importance for solving the general QA problem which will require cross community fertilization. Individual activities within the lanes are either currently planned or future desired events progressing toward longer term objectives.
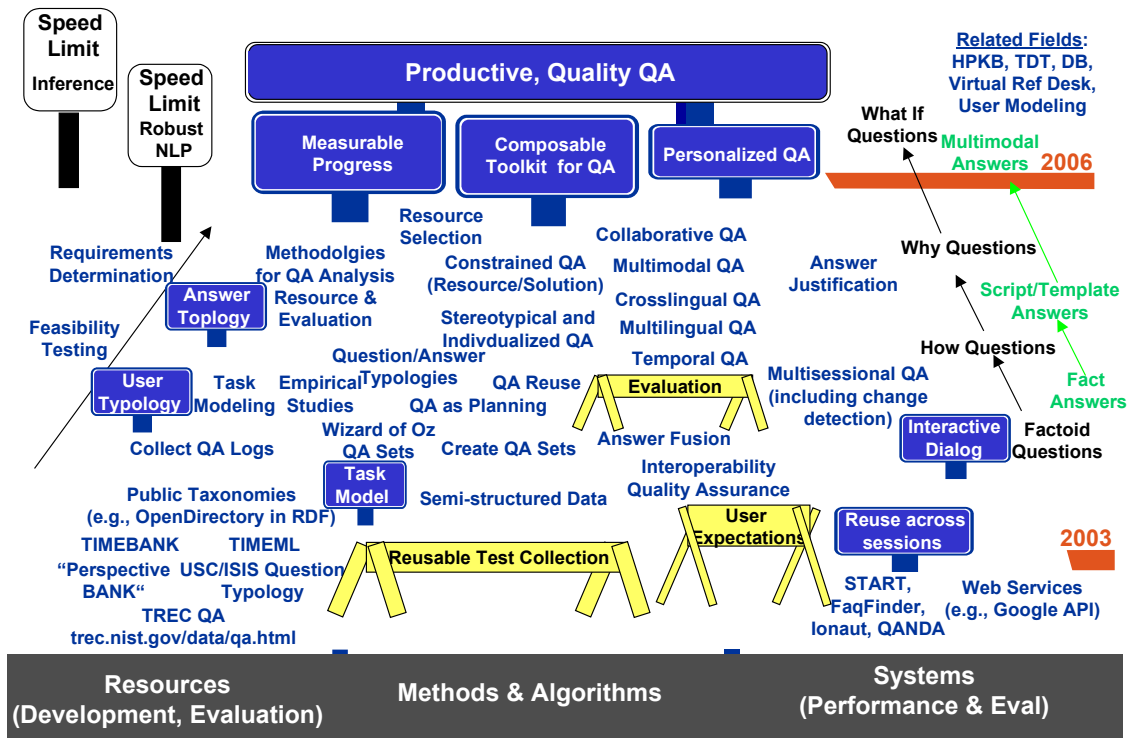

Figure 3. Question Answering Roadmap

Future

A workshop on multilingual summarization and question answering was planned at COLING in Taipei in August and a Japanese NTCIR Q&A workshop is being planned together with a future release of a Japanese QA corpora (see research.nii.ac.jp/ntcir/workshop/qac/cfp-en.html). We intend to publish this roadmap and regularly update it as new resources and tools emerge and as new QA challenges emerge.

References

ARDA QA Roadmap (www.nlpir.nist.gov/projects/duc/papers/ qa.Roadmap-paper_v2.doc).

ARDA AQUAINT Program - http://www.ic-arda.org/InfoExploit/aquaint/index.html

ARDA Q&A Roadmap - www-nlpir.nist.gov/projects/duc/papers/qa.Roadmap-paper_v2.doc

LREC Q&A Roadmap Workshop - http://www.lrec-conf.org/lrec2002/lrec/wksh/QuestionAnswering.html

ARDA NRRC Summer 2002 workshops on temporal and multiple perspective question answering - nrrc.mitre.org

TREC QA track - http://trec.nist.gov/presentations/TREC10/qa/

Bos, J. and Gabsdil, M. 2000. "First-Order Inference and the Interpretation of Questions and Answers", Proceedings of Gotelog 2000, Goteborg, Sweden, pages 43-50.

Dragomir Radev, Weiguo Fan, Hong Qi, and Amardeep Grewal, "Probabilistic Question Answering on the Web", Proceedings, 11th International WWW Conference (Honolulu, Hawaii, May 2002).

Maybury, M. T., Sparck Jones, K. Voorhees, E., Harabagiu, S., Liddy, L., and Prange, J. Tuesday May 28, 2002. Workshop on Strategy and Resources for Question Answering in conjunction with the Third International Conference on Language Resources and Evaluation (LREC). Palacio de Congreso de Canarias, Canary Islands, Spain. http://www.lrec-conf.org/lrec2002/lrec/wksh/QuestionAnswering.html