Advances in Open Domain Question Answering

Tomek Strzalkowski and Sanda Harabagiu (editors)
(SUNY Albany and University of Texas at Dallas)

Reviewed by
Paul Thompson
Dartmouth College

This edited volume provides an excellent, comprehensive introduction to recent research on open-domain question answering, meaning that the question-answering systems are not used merely in specific, narrow domains, but rather in domains such as news or the Internet. Most of the chapters describe systems that have been developed for the Advanced Question Answering for Intelligence, or AQUAINT, program sponsored by the U.S. Government Advanced Research and Development Activity (ARDA), or the related Text Retrieval Conference (TREC) Question Answering (QA) Track. Both of these activities began in 1999 and, as of this writing, still continue. Although much of the research described here has this common origin, a wide variety of approaches to question answering are represented in these chapters. The book is divided into six sections:

1. Approaches to question answering

2. Question processing

3. Question answering as information retrieval

4. Answer extraction

5. Evaluating question-answering systems

6. Perspectives on question answering

Each section consists of two or three chapters.

Automatic question answering has long been studied in artificial intelligence research. Early systems were constrained to answering questions about limited domains, for example, baseball statistics or lunar rocks. By the 1990s, question answering had become a less active field of research. During the 1990s the U.S. Government funded the Tipster program for research on information retrieval and information extraction. As discussed by Maiorano in his chapter "Question answering: Technology for intelligence analysis," it was hoped that the template-filling of Tipster's information-extraction tasks would require researchers to develop systems capable of deep natural language understanding. Instead, shallow techniques were able to perform well on those tasks. Similarly, the AQUAINT program is seen as building on information-retrieval and information-extraction technology to provide systems that can extract answers from open-domain free text for information seekers, rather than just ranked lists of documents that might answer the question when read. Again, there is the view that achieving a question-answering capability will require deep natural language

understanding. Although some of the chapters in this volume describe long-range goals of achieving levels of question answering requiring deep understanding, much of the research described here so far has focused on simpler question-answering tasks, such as the factoid question answering of the TREC QA track. As the book shows, however, factoid, or slightly more complex, question answering can benefit from a variety of more or less deep approaches.

Section 1 of the book contains three chapters presenting very different approaches to question answering. The first chapter in the book, by Moldovan et al., describes how definition-style questions can be answered using deep language processing and logic theorem proving. The next chapter, by Ittycheriah, describes the application of maximum entropy and other machine-learning approaches to answering factoid questions. The chapter by Vicedo and Ferrández shows how anaphora resolution can improve question-answering performance. Most contemporary question-answering systems use an information-retrieval component to find candidate documents, followed by an extraction component that produces an answer. Section 3 describes three approaches to question answering that rely primarily on information-retrieval techniques.

Contemporary question-answering systems generally include question-processing and answer-processing components, where a question type is first determined from an analysis of the question, and then answers are sought using algorithms specific to that question type. Sections 2 and 4 describe question processing and answer extraction. Harabagiu's chapter, "Questions and intentions," provides a particularly useful discussion of the importance of pragmatic knowledge in question answering. In Section 4, Prager et al.'s chapter, "Question answering by predictive annotation," discusses preprocessing of text, for example, marking up named entities, and Srihari et al. in "Question answering supported by multiple levels of information extraction" focus on applying information-extraction techniques at retrieval time. On the other hand, Echihabi et al., in their chapter "How to select an answer string," use statistical and information-theoretic techniques, which interpret an answer as a "translation" of the question.

The Tipster program, with its Message Understanding Conferences, and TREC both placed a strong emphasis on evaluation. The TREC QA Track and the Interactive Question Answering Track likewise have developed formal evaluation metrics, in particular for factoid question answering. Though evaluation for the more advanced forms of question answering in the AQUAINT program has not been as formal, some interesting new approaches to evaluation involving analysts performing realistic tasks have been developed, as discussed, for example, by Strzalkowski et al. in their chapter, "Question answering as dialogue with data." Voorhees's "Evaluating question answering system performance" and Hersh's "Evaluating interactive question answering" describe the TREC QA Track evaluations; Ogden et al.'s "Habitability in question-answering systems" takes a step back from current question-answering technology and considers user requirements with future question-answering systems.

The final section of the book considers question answering from the perspectives of intelligence analysis, statistical machine learning, and the long-range future. Maiorano's chapter has been mentioned earlier. Riloff et al. in "Reverse-engineering question / answer collections" describe a technique with which to bootstrap training data to support machine-learning approaches to question answering. Maybury's concluding chapter, "New directions in question answering," provides a long-range perspective on question-answering research, describing three future question-answering visions. The first is a user-centered one, where personal assistants model the user's interests and cognitive style. Another vision is of embedded intelligence, where intelligence and

question-answering capabilities reside in the environment. The third vision is of embodied agents, for example, virtual avatars or physical humanoid robots. Maybury notes that one of the crucial issues will be discovering the right mixed-initiative symbiosis among human and artificial agents.

This book is based largely on an ongoing, but soon to be completed, research program. The chapters necessarily reflect research from early in the program. It will be useful to compare the present volume to a volume that might appear once the program has been completed. So far, question answering has not had the commercial success of, say, information-retrieval technology. It will be interesting to see how research in this area proceeds once the stimulus from the AQUAINT program ends and whether the expected utility of question answering for analysts can be extended to the more general user.

Paul Thompson is a Research Associate Professor in the Department of Computer Science at Dartmouth College. His research interests include probabilistic information retrieval, natural language processing, and information assurance. Thompson's address is Department of Computer Science, Dartmouth College, Hinman Box 6211, Hanover, NH 03755-3510; e-mail: Paul.Thompson@dartmouth.edu.