Book Reviews

Interactive Multi-Modal Question-Answering

Antal van den Bosch* **and Gosse Bouma‡ (editors)**
(*Tilburg University and ‡University of Groningen)

Berlin: Springer (Theory and Applications of Natural Language Processing series, edited by Eduard Hovy), 2011, xii+279 pp; hardbound, ISBN 978-3-642-17524-4, $124.00; e-book, ISBN 978-3-642-17525-1; paperbound, $24.95 or €24.95 to members of subscribing institutions

Reviewed by
Constantin Orăsan
University of Wolverhampton

Processing and presentation of multimodal information was one of the important directions pursued by researchers in the areas related to information processing and management in the first decade of this century (Stock and Zancanaro 2005; Maragos, Potamianos, and Gros 2008; Lalanne et al. 2009). The Interactive Multimodal Information eXtraction (IMIX) Programme, a research program that ran between 2004 and 2009 and was funded by the Netherlands Organisation for Scientific Research (NWO), adhered to this direction of research. This book contains a collection of articles describing research carried out in the IMIX Programme. Given the large scale of the program, the book covers only parts of it, arguably the most important ones: question answering, (spoken) dialogue systems, and human–machine interaction.

The book is organized into four parts and an epilogue. The first part introduces the IMIX Programme and the demonstrator developed by it. The main purpose of the program was to bring together research groups from the Netherlands to build an interactive multimodal question answering (QA) system that is able to answer general encyclopedic medical questions. The IMIX Programme funded seven individual projects that worked in a common field and contributed to a common demonstrator. The fact that these projects ran largely independently is also apparent from the book because there are few links between its chapters.

The architecture of the demonstrator is presented in the second article of the book, "The IMIX Demonstrator" (Dennis Hofs, Boris van Schooten, and Rieks op den Akker). The demonstrator showed users a fully functional system and allowed them to ask questions using text, speech, and gestures. The answers produced by the system were presented in the form of text, speech, or images, and could be used in follow-up questions. The article features a detailed description of the architecture, as well as several screenshots and diagrams; these can be useful to researchers who want to find out more about the demonstrator. In addition to the technical details, there is an interesting discussion about the role of demonstrators in large projects and problems that need to be addressed when building them. I think this brief discussion could be very useful for anyone involved in a medium or large project that includes several research groups and needs to build demonstrators.

The second part of the book focuses on dialog managers and covers most of the interaction discussed in this book. First, the Vidiam (DIAlogue Management and the VIsual channel) project is described in the article entitled "Corpus-Based Development of a Dialogue Manager for Multimodal Question Answering" (Van Schooten and

Op den Akker). In addition to the corpora built in the project and the dialog manager developed on the basis of these corpora, the article also contains a very good discussion about how it is possible to integrate a dialog manager with a QA engine as a way of developing an interactive QA system. I am not aware of any other articles that contain all the information presented here in one place and in such detail. The second article in this part, "Multidimensional Dialogue Management" (Simon Keizer, Harry Bunt, and Volha Petukhova), is more theoretical and presents a dialog manager built using the framework of Dynamic Interpretation Theory (Bunt 2000) which is able to both interpret and generate utterances using dialog acts. The article also presents briefly the way in which this dialog manager was integrated in the IMIX demonstrator.

In my opinion, the editors of the book could have chosen a better title for the third part of the book: "Fusing Text, Speech, and Images." Both articles in this part present work done in the IMOGEN (Interactive Multimodal Output GENeration) project,[1] one of the subprojects embedded in the IMIX Programme that focused on producing multimodal presentations that combine text, speech, and graphics. Only the first article focuses on the multimodal aspect of the project, however. The other one discusses only text processing. The article "Experiments in Multimodal Information Presentation" (Charlotte van Hooijdonk, Wauter Bosma, Emiel Krahmer, Alfons Maes, and Mariët Theune) presents three experiments for finding the appropriate way of combining text and images when answering questions from the medical domain. In one of these experiments, the multimodal answers are produced automatically. The other article, "Text-to-Text Generation for Question Answering" (Bosma, Erwin Marsi, Krahmer, and Theune), discusses sentence fusion and could fit very well in a book dedicated to text summarization, as the method presented there is tested not only on data specific to IMIX, but also on the DUC 2005 data.[2]

The fourth and the largest part of the book is "Text Analysis for Question Answering." It contains five articles, none of which describe a full QA system. Instead, as the title suggests, they focus on various ways of processing texts that can help with answering questions. One common feature of these articles is that they describe methods to extract entities or relations between entities from texts. Most of the articles also briefly discuss how this information is used in QA systems.

Most of the methods described in the fourth part of the book are now widely used in computational linguistics, but when they were proposed a few years ago many of them were rather innovative. For brevity, I give only a succinct indication of the methods presented in the articles. "Automatic Extraction of Medical Term Variants from Multilingual Parallel Translations" (Lonneke van der Plas, Jörg Tiedemann, and Ismail Fahmi) describes how to acquire medical terms and their variants from parallel corpora. "Relation Extraction for Open and Closed Domain Question Answering" (Bouma, Fahmi, and Jori Mur) shows how it is possible to extract relations between entities using dependency paths in a large collection of newspaper articles and in a much smaller and closed domain corpus of medical documents. A sequence labeling method for entity recognition is presented in the article "Constraint-Satisfaction Inference for Entity Recognition" (Sander Canisius, Antal van den Bosch, and Walter Daelemans). Large newspaper corpora and the Web are used in "Extraction of Hypernymy Information from Texts" (Erik Tjong Kim Sang, Katja Hofmann, and Maarten de Rijke) to determine hypernymy relations between entities. The last article in the fourth

1 http://wwwhome.cs.utwente.nl/∼theune/IMOGEN/.
2 http://www-nlpir.nist.gov/projects/duc/intro.html.

part, "Towards a Discourse-Driven Taxonomic Inference Model" (Piroska Lendvai) looks at how the structure of discourse can be used for knowledge discovery from encyclopedic texts. All the articles are well written and could be very interesting for researchers working on information extraction.

The book finishes with an epilogue written by three members of the international review panel of IMIX (Eduard Hovy, Jon Oberlander, and Norbert Reithinger) who give a very good overview of the project, providing information that is not covered in any other article of the book. For example, it expands on the multimodal research carried out in the program and presents some details from the point of view of project management. An objective evaluation of the overall program is also included.

The book is interesting and I enjoyed reading it. I have to point out, however, that the research presented here is rather old. The IMIX Programme effectively ended in 2008, so it can be argued that most of the articles refer to work that is more than 5 years old. The authors of the epilogue praise the researchers involved in IMIX for the large number of publications they produced. This means that most of the information presented in the book was already published in one form or another somewhere else. Despite this, the book compiles in one place information about the IMIX Program which otherwise could take a while to collect.

When I started reading the book, I expected to find more about *interactive multimodal question answering*. Each of these topics is presented individually, but with the exception of the article about the IMIX demonstrator, they are not discussed as a whole. I was particularly disappointed by how little space was dedicated to multimodal processing.

The articles in the book are written by different groups of authors and they are more or less stand-alone. To achieve this, they all present brief background information about the IMIX Programme. Despite the extra space required for it and the overlap between the information presented in the articles, this is not necessarily bad because it means that researchers who do not have the time to read the whole book can focus on only the articles that are most relevant for them.

The potential readers of this book are likely to be researchers interested in the processing of Dutch texts. Researchers in question answering, dialog processing, and information extraction would also benefit from the book.

References

Bunt, Harry. 2000. Dialogue pragmatics and context specification. In *Abduction, Belief and Context in Dialogue*. John Benjamins, Amsterdam.

Lalanne, Denis, Laurence Nigay, Philippe Palanque, Peter Robinson, Jean Vanderdonckt, and Jean-François Ladry. 2009. Fusion engines for multimodal input: a survey. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, ICMI-MLMI '09, pages 153–160, New York, NY.

Maragos, Petros, Alexandros Potamianos, and Patrick Gros, editors. 2008. *Multimodal Processing and Interaction: Audio, Video, Text*. Springer, Berlin.

Stock, Oliviero and Massimo Zancanaro, editors. 2005. *Multimodal Intelligent Information Presentation*. Springer, Berlin.

Constantin Orăsan is a Senior Lecturer in Computational Linguistics at the University of Wolverhampton, UK. His current research interests include anaphora and coreference resolution, automatic summarization, and question answering. Orăsan's address is RIILP, Wulfruna St., University of Wolverhampton, WV1 1LY, UK; e-mail: C.Orasan@wlv.ac.uk.