

One-on-One Tutoring by Humans and Computers

Martha Evens and Joel Michael

(Illinois Institute of Technology and Rush Medical College)

Lawrence Erlbaum Associates, 2006, xix+475 pp; hardbound, ISBN 0-8058-4360-4, \$120.00; paperbound, ISBN 0-8058-4361-2, \$45.00

Reviewed by

Pamela Jordan

University of Pittsburgh

Tutorial dialogue is an unusual genre of natural language interaction in that one of the dialogue participants asks questions to which he or she already knows the answer. This book serves the dual purpose of introducing computational linguists to the unique character of tutorial dialogues and the challenges associated with building tutorial dialogue systems. Evens and Michael chronicle an ongoing 15-year effort that has succeeded in building a measurably effective tutorial dialogue system in the domain of cardiovascular physiology. Although the techniques employed in the many versions of the system produced over the years may seem somewhat dated, there are two exciting things about this book and the system it describes. One is that students who used the system learned more than those who engaged in other learning activities. The second is the companion CD with the latest, in-use version of the system along with the source code. Given the many remaining challenges, which are outlined at the end of the book, the CIRCSIM-tutor system can serve as a test bed and allow any researcher just starting out in the area of tutorial dialogue to test new techniques and approaches while avoiding the high costs of building an effective tutorial dialogue system. However, the authors never explicitly express the latter as one of their goals. Instead they state that they wish us to come away with the knowledge that a useful tutorial dialogue system can be built using existing technology and an appreciation of the power of one-on-one tutoring and the importance of natural language dialogue in one-on-one tutoring. I think the book succeeds with these first two goals but is slightly less convincing with respect to the third because they do not devote much time to it.

The book is divided into four parts. The first part, chapters 1 and 2, provide an introduction to the CIRCSIM-tutor project and the domain in which it tutors. Chapter 1 introduces one-on-one tutoring (since there are many forms of tutoring), the importance of natural language dialogue to tutoring, and background on how the CIRCSIM-tutor project arose. Chapter 2 is a brief, gentle introduction to the area of cardiovascular physiology that is the domain of instruction. Although I had previously read many of the corpus analysis papers from the CIRCSIM-tutor project, this introduction helped me better understand the many dialogue excerpts and the discussion, which are presented in part 2 of the book, to a greater degree than before.

Part 2 of the book, which is its strongest one and may be of the most interest to computational linguists, spans chapters 3–10. It covers all of the studies the project did of one-on-one tutoring by expert human tutors, where the main purpose of the studies was to inform the building of a tutorial dialogue system. The companion CD includes all the dialogues collected by the CIRCSIM-tutor project. As various studies of the collected dialogues are discussed, the parts of the system they influenced are introduced. At the same time, part 2 provides a good introduction to the learning sciences literature on tutorial dialogue. Chapters 3–6 provide the background for the more exciting chapters that

follow in 7–10. Chapter 3 provides details on the methodology followed in collecting human–human and human–computer tutorial dialogue, and will be of value to anyone who is just beginning to collect dialogues to analyze. Chapter 4 gives a high-level view of student and tutor language use and how they compare to one another. It compares number and length of turns, kinds of syntactic constructions used, and lexicons. The discussion on tutor usage of discourse markers went much deeper than any of the rest of the discussion in this chapter and for me was the highlight of a generally interesting chapter. Chapters 5 and 6 cover the domain knowledge needed to solve CIRCSIM-tutor problems and the knowledge of tutoring that is needed. These two chapters will be of interest only to those readers whose research touches on domain reasoning or cognitive science.

Chapters 7 and 8 were two of my favorites. Chapter 7 provides details on the intentions involved in tutorial dialogue at three levels: protocol, strategy, and tactic. Most of these are defined generally enough to be relevant to other tutorial domains. A protocol is the project's term for the overall tutoring plan, and impacts whether feedback during problem solving is immediate or delayed. Strategies and tactics address how to move the student from what he or she knows toward deriving a new bit of knowledge, for example, by exploring underlying concepts. All three levels of tutor planning revolve around keeping the student successfully working independently as much as possible. If a student fails too often, this signals the need to switch the tactic, strategy, or protocol. Toward the end of this chapter, the discussion of positive and negative acknowledgments found in human–human tutorial dialogue provides many useful details for system developers. Chapter 8 acts as a detailed extension of chapter 7 by thoroughly exploring hints in human tutorial dialogue. Hints are one of the main tactics available in tutorial dialogue. Chapter 9 then applies all that has been described in part 2 of the book in order to walk us through an analysis of a tutoring episode. Finally, chapter 10 is an excellent look at how effective human tutors are and the differences between expert and novice tutors. Much of the methodology used in evaluating the effectiveness of the human tutors is repeated in part 4 of the book when the CIRCSIM-tutor system is evaluated.

Chapters 11–17 constitute part 3 of the book and describe the overall architecture of the system and each of its major modules in more detail. Although this part of the book extends the introductions made in part 2, it does not provide quite enough detail. However, there is enough for a reader to understand what was tried and why, and there are plenty of pointers to relevant project papers that will be useful for anyone who wants to try using the CIRCSIM-tutor system as a test bed. Although the techniques used are not all measurably the most accurate and the corpus of human tutoring dialogues is never used to directly train the system, researchers will still benefit from at least scanning through the techniques used and the reasons why these techniques were selected.

The fourth part of the book spans chapters 18–21, the most important of which are 18 and 20. Chapter 18 provides detailed evaluations of the versions of the system tested over the years. Readers who are not familiar with the evaluation methodology adopted from the learning sciences community will find chapter 18 educational. Chapter 19 provides an interesting but perhaps too brief history of tutorial dialogue system development and briefly introduces a number of recent systems and projects that came into existence after the CIRCSIM-tutor project began, but which have influenced one another. Chapter 20 helpfully suggests possible research directions that would improve our understanding of tutorial dialogues (e.g., continued study of the differences between expert and novice tutors), explore the educational impact of tutorial actions

(e.g., the impacts of immediate vs. delayed feedback), and improve tutorial dialogue systems (e.g., better adaptation to individual students). Chapter 21 summarizes what the authors have learned about building a tutorial dialogue system and passes along words of wisdom, such as frequent testing by and feedback from student users, and encouragement to others who wish to pursue similar work.

Overall, Evens and Michael have written a delightful, interesting book. Although at times it reads like an internal project status report in its endeavor to document why certain design decisions were made, it captures some of what does not often get explicitly stated in research papers and articles: the limitations imposed by whatever our currently available hardware and software tools happen to be.

I recommend this book to all researchers who are interested in explaining problem-oriented human-human dialogues or who are interested in building and evaluating dialogue systems as one way of testing the generality of one's theories and systems. I also recommend this book as a resource for those of us who are already working in the area of dialogue for educational applications (e.g., peer problem solving as well as tutorial dialogue). It is one that I wish had come into existence at least five years earlier so that I would not have had to rediscover some of the many bits of wisdom sprinkled throughout the book. Finally, I recommend it as a guidebook for researchers who are interested in getting started in the area of dialogues for educational applications.

Pamela Jordan is a Research Associate at the Learning Research and Development Center at the University of Pittsburgh. Her most recent work in computational linguistics focuses on building and evaluating natural language dialogue systems for educational applications. She received an M.S. in Computational Linguistics from Carnegie Mellon University and a Ph.D. in Intelligent Systems from the University of Pittsburgh. Jordan's address is LRDC, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260; e-mail: pjordan+@pitt.edu.

