Abstract

"Text Knowledge Extraction" maps natural language texts onto a formal representation of the facts contained in the texts. Common text knowledge extraction methods show a severe lack of methods for understanding natural language "degree expressions", like "expensive hard disk drive" and "good monitor", which describe gradable properties like price and quality, respectively. However, without an adequate understanding of such degree expressions it is often impossible to grasp the central meaning of a text.

This book shows concise and comprehensive concepts for extracting degree information from natural language texts. It researches this task with regard to the three levels of *(i)* analysing natural language degree expressions, *(ii)* representing them in a terminologic framework, and *(iii)* inferencing on them by constraint propagation. On each of these three levels, the author shows that former approaches to the degree understanding problem were too simplistic, since they ignored by and large the role of the background knowledge involved. Thus, he gives a constructive verification of his central hypothesis, viz. that the proper extraction of grading knowledge relies heavily on background grading knowledge.

This construction proceeds as follows. First, the author gives an overview of the ParseTalk information extraction system. Then, from the review of relevant linguistic literature, the author derives two distinct categories of natural language degree expressions and proposes knowledge-intensive algorithms to handle their analyses in the ParseTalk system. These methods are applied to two text domains, viz. a medical diagnosis domain and a repository of texts from information technology magazines. Moreover, for inferencing the author generalizes from well-known constraint propagation mechanisms. This generalization is especially apt for representing and reasoning with natural language degree expressions, but it is also interesting from the point of view where it originated, viz. the field of temporal reasoning. The conclusion of the book gives an integration of all three levels of understanding showing that their coupling leads to an even more advanced — and more efficient — performance of the proposed mechanisms.

Foreword

If you are sitting in a basement room without a view — not to mention the bars in front of the windows — and writing a book, then you better have good company. I had the best company you could imagine. Waltraud Hiltl, Katja Markert, Martin Romacker, Klemens Schnattinger, Andreas Klee and I shared very little office space, but plenty of chocolate, coffee, champagne, and enthusiasm for our research. North German coolness and creativity sprang mostly from my colleagues in the second floor. I learned a lot from and laughed a lot with Nobi Bröker, Susanne (Sue) Schacht, Manfred Klenner, Peter Neuhaus, Stefan Schulz, and Michael Strube.

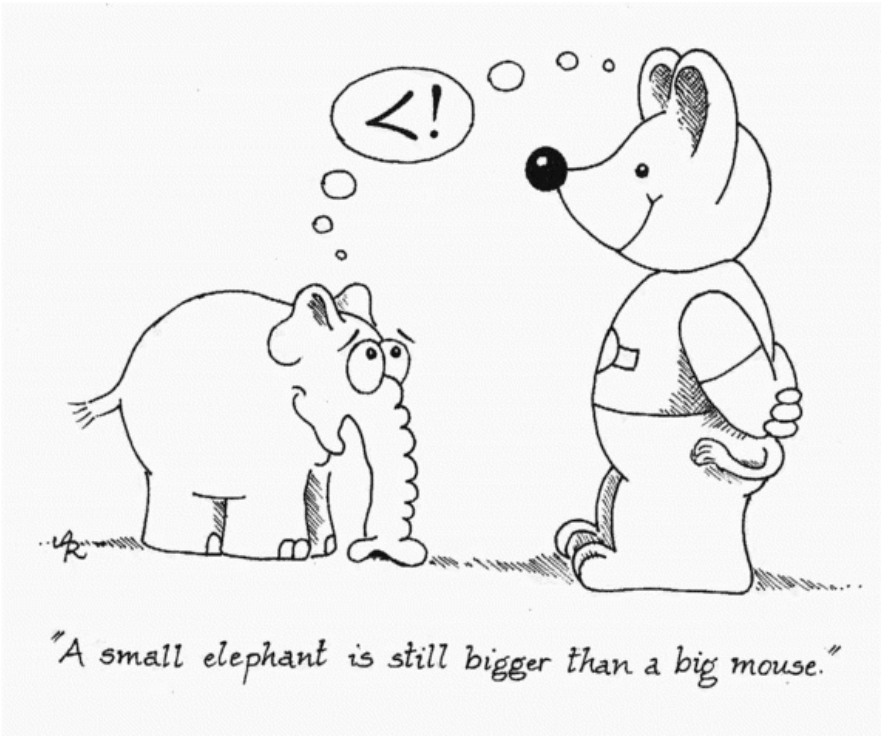
I thank my friend and partner Angela Rösch for motivational and technical support and for living together with someone who cares about strange things, works too much and does not improve in any way over the years. Special thanks go to my family who sometimes wondered what was going on when I started talking enthusiastically about "semantics", but they never let wane their encouragement for me.

Kornel Marco provided great service by implementing parts of the system presented in this book. Joe Bush helped me polish up the text with his capabilities as an American native speaker. Remaining errors are entirely my fault and due to my lack of diligence.

This book would not have seen the light of day without the dissertation grant through the Graduiertenkolleg "Menschliche & Maschinelle Intelligenz" funded by the Deutsche Forschungsgemeinschaft (DFG).

Mostly, however, I must thank my advisor Udo Hahn. My perspective on research in Artificial Intelligence, Computational Linguistics, and Cognitive Science grew under his auspices. He fostered the work reported in this book in so many ways that I cannot name them all.

Steffen Staab
Karlsruhe, Germany, October 1999



Misleading Norms of Expectation
By a courtesy of Angela Rösch