

# Preface to the 2nd edition

 <https://doi.org/10.1075/nlp.5.preface>

Pages ix–x of

**Natural Language Processing for Online Applications.**  
**Second revised edition : Text retrieval, extraction and categorization**

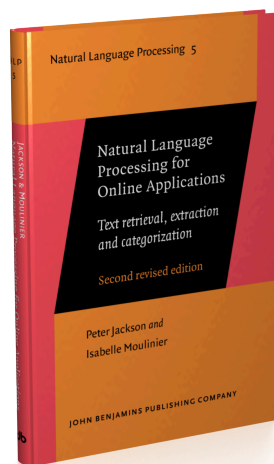
**Peter Jackson and Isabelle Moulinier**

[Natural Language Processing, 5] 2007. x, 231 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

For further information, please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website at [benjamins.com/rights](http://benjamins.com/rights)



## Preface to the 2nd edition

In the new edition, we have made a number of changes to both the style and content of the volume. The original intent was to make the book accessible to a wider audience, and so each chapter began with some fairly light fare. However, it appears that most of the uptake of the First Edition was by academics, and so we have revised most of the chapter openings to make them less informal and more focused upon the applications context. We continue to provide a historical perspective of each area, but we have also updated the references with work done since 2002. New topics that we have introduced include sentiment analysis, language modeling for information retrieval, the GATE toolkit for information extraction, and support vector machines for classification. We maintain our focus upon results and systems that are being used, or about to be used, in a commercial setting. This means that we continue to make relatively scant reference to new research that has yet to prove its worth from a practical standpoint.

Since 2002, we have seen the Internet economy return from the bust of 2000, the continuing rise of Web search, and the advent of online advertising as big business. What we have not seen are fundamental changes in the contribution of natural language processing to the way in which information is accessed and organized, despite the continuing claims of a new wave of start-ups that their search engines ‘understand’ human language, and the like. Nor has the concept of the Semantic Web made significant inroads into the way transactions are conducted or the way information is structured, although that remains a promise for the future. What has changed since 2002 is the scale and complexity of both the open Web and the so-called ‘dark’ Web of proprietary information, and the nature of the business models that power Web 2.0 initiatives.

I have spent the intervening years continuing to run an R&D group of 35 people in a \$7 billion online information business. During that time, I have seen the technologies described in this book begin to make a real difference to the kinds of products and services that can be offered to professionals on the Web. But, to get the full benefit, the prosecution of research needs to be aligned with business goals and its output needs to well-integrated with both front-end applications development and back-end editorial processes. These aspects of successful deployment do not form the subject matter of this book, but they do influence the selection of topics and the way in which we think about total systems

evaluation. As a result, we typically go beyond academic measures of precision and recall to consider broader aspects of utility and usability. We hope that readers will find this perspective both useful and informative.

Peter Jackson  
Saint Paul, MN