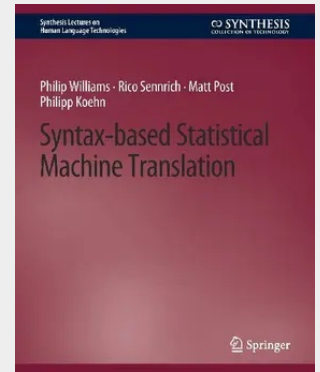


Syntax-based Statistical Machine Translation

This unique book provides a comprehensive introduction to the most popular syntax-based statistical machine translation models, filling a gap in the current literature for researchers and developers in human language technologies. While phrase-based models have previously dominated the field, syntax-based approaches have proved a popular alternative, as they elegantly solve many of the shortcomings of phrase-based models. The heart of this book is a detailed introduction to decoding for syntax-based models.

The book begins with an overview of synchronous-context free grammar (SCFG) and synchronous tree-substitution grammar (STSG) along with their associated statistical models. It also describes how three popular instantiations (Hiero, SAMT, and GHKM) are learned from parallel corpora. It introduces and details hypergraphs and associated general algorithms, as well as algorithms for decoding with both tree and string input. Special attention is given to efficiency, including search approximations such as beam search and cube pruning, data structures, and parsing algorithms. The book consistently highlights the strengths (and limitations) of syntax-based approaches, including their ability to generalize phrase-based translation units, their modeling of specific linguistic phenomena, and their function of structuring the search space.

This unique book provides a comprehensive introduction to the most popular syntax-based statistical machine translation models, filling a gap in the current literature for researchers and developers in human language technologies. While phrase-based models have previously dominated the field, syntax-based approaches have proved a popular alternative, as they elegantly solve many of the shortcomings of phrase-based models. The heart of this book is a detailed introduction to decoding for syntax-based models. The book begins with an overview of synchronous-context free grammar (SCFG) and synchronous tree-substitution grammar (STSG) along with their associated statistical models. It also describes how three popular instantiations (Hiero, SAMT, and GHKM) are learned from parallel corpora. It introduces and details hypergraphs and associated general algorithms, as well as algorithms for decoding with both tree and string input. Special attention is given to efficiency, including search approximations such as beam search and cube pruning, data structures, and parsing algorithms. The book consistently highlights the strengths (and limitations) of syntax-based approaches, including their ability to generalize phrase-based translation units, their modeling of specific linguistic phenomena, and their function of structuring the search space.



53,49 €

49,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783031010361

Medium: Buch

ISBN: 978-3-031-01036-1

Verlag: Springer

Erscheinungstermin: 11.08.2016

Sprache(n): Englisch

Auflage: Erscheinungsjahr 2016

Serie: Synthesis Lectures on Human Language Technologies

Produktform: Kartoniert

Gewicht: 406 g

Seiten: 190

Format (B x H): 191 x 235 mm

