Dependency Parsing

Synthesis Lectures on Human Language Technologies

Editor
Graeme Hirst, *University of Toronto*

Synthesis Lectures on Human Language Technologies publishes monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is placed on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Dependency Parsing
Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2009

© Springer Nature Switzerland AG 2022
Reprint of original edition © Morgan & Claypool 2009

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Dependency Parsing

Sandra Kübler, Ryan McDonald, and Joakim Nivre

ISBN: 978-3-031-01003-3 paperback
ISBN: 978-3-031-02131-2 ebook

DOI 10.1007/978-3-031-02131-2

A Publication in the Springer Nature series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #2
Series Editor: Graeme Hirst, *University of Toronto*

Series ISSN
Synthesis Lectures on Human Language Technologies
ISSN pending.

Dependency Parsing

Sandra Kübler
Department of Linguistics, Indiana University

Ryan McDonald
Google Research

Joakim Nivre
Department of Linguistics and Philology, Uppsala University
School of Mathematics and System Engineering, Växjö University

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #2

ABSTRACT

Dependency-based methods for syntactic parsing have become increasingly popular in natural language processing in recent years. This book gives a thorough introduction to the methods that are most widely used today. After an introduction to dependency grammar and dependency parsing, followed by a formal characterization of the dependency parsing problem, the book surveys the three major classes of parsing models that are in current use: transition-based, graph-based, and grammar-based models. It continues with a chapter on evaluation and one on the comparison of different methods, and it closes with a few words on current trends and future prospects of dependency parsing. The book presupposes a knowledge of basic concepts in linguistics and computer science, as well as some knowledge of parsing methods for constituency-based representations.

KEYWORDS

parsing, syntax, dependency parsing, dependency grammar

Contents

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES iii

Contents vii

Preface xi

1 Introduction 1

1.1 Dependency Grammar 1

1.2 Dependency Parsing 6

1.3 Summary and Further Reading 9

2 Dependency Parsing 11

2.1 Dependency Graphs and Trees 11

2.1.1 Properties of Dependency Trees 13

2.1.2 Projective Dependency Trees 16

2.2 Formal Definition of Dependency Parsing 18

2.3 Summary and Further Reading 19

3 Transition-Based Parsing 21

3.1 Transition Systems 21

3.2 Parsing Algorithm 25

3.3 Classifier-Based Parsing 27

3.3.1 Feature Representations 28

3.3.2 Training Data 31

3.3.3 Classifiers 33

3.4 Varieties of Transition-Based Parsing 34

3.4.1 Changing the Transition System 34

3.4.2 Changing the Parsing Algorithm 36

3.5 Pseudo-Projective Parsing 37
3.6 Summary and Further Reading 38

4 Graph-Based Parsing 41

4.1 Arc-Factored Models 42
4.2 Arc-Factored Parsing Algorithms 43
4.2.1 Reducing Labeled to Unlabeled Parsing 45
4.2.2 Non-Projective Parsing Algorithms 46
4.2.3 Projective Parsing Algorithms 49
4.3 Learning Arc-Factored Models 54
4.3.1 Parameter and Feature Representations 54
4.3.2 Training Data 55
4.3.3 Learning the Parameters 56
4.4 Beyond Arc-Factored Models 56
4.5 Summary and Further Reading 61

5 Grammar-Based Parsing 63

5.1 Context-Free Dependency Grammar 64
5.1.1 Parsing with Bilexical Grammars 65
5.2 Constraint Dependency Grammar 69
5.2.1 Weighted Constraint Dependency Grammar 71
5.2.2 Transformation-Based Constraint Dependency Parsing 72
5.3 Summary and Further Reading 75

6 Evaluation 79

6.1 Evaluation Metrics 79
6.2 Dependencies for Cross-Framework Parser Evaluation 80
Converting Constituent Treebanks into Dependency Formats 81
The CoNLL Shared Tasks on Dependency Parsing 82
6.5 Summary and Further Reading 86

7 Comparison 87

7.1 Comparing Transition-Based and Graph-Based Models 87

7.2 Comparing Grammar-Based and Data-Driven Models 90

7.3 Summary and Further Reading 93

8 Final Thoughts 95

A Resources 97

A.1 Parsers 97

A.2 Treebanks 98

A.3 Dependency Parsing Wiki 99

Bibliography 101

Author Biographies 115

Preface

Dependency-based methods for syntactic parsing have become increasingly popular in natural language processing in recent years. One of the reasons for their success is that they have been shown to work reliably for a wide range of typologically different languages. The increased interest in dependency-based parsing has led to investigations into a range of different parsing algorithms.

The aim of this book is to give readers new to this field an introduction to the parsing algorithms used in dependency parsing. The aim is not to propose new methods or new findings nor to promote a single algorithm but rather to give an overview of existing algorithms and a comparison of their major similarities and differences. Additionally, we will touch upon matters of evaluation and data representation.

This book is aimed at graduate students and researchers in computer science, linguistics, and computational linguistics. It expects familiarity with basic concepts in linguistics and computer science, as well as some knowledge of parsing methods for constituency-based representations. Thus, we expect the reader to be familiar with basic chart parsing algorithms such as the Cocke-Kasami-Younger algorithm or Earley's algorithm. Additionally, we expect the reader to be familiar with the basic concepts of probability theory.

It is helpful, but not necessary, for the reader to be familiar with concepts from machine learning. The book concentrates on supervised approaches to dependency parsing, which rely on a range of different learning approaches: memory-based learning, support vector machines, and perceptron learning, to name just a few. However, these approaches are not central to the understanding of the dependency parsing algorithms.

The book is partly based on material from two courses: The ACL/COLING 2006 tutorial on *Dependency Parsing*, presented by Sandra and Joakim, and the ESSLLI 2007 introductory course *Introduction to Data-Driven Dependency Parsing*, given by Ryan and Joakim. Other material is derived from earlier publications, and we are grateful to the Association for Computational Linguistics for giving us permission to reuse material previously published in various conference proceedings. We also want to thank our co-authors in those publications: Atanas Chanev, Koby Crammer, Gülşen Eryiğit, Jan Hajič, Johan Hall, Kevin Lerman, Svetoslav Marinov, Erwin Marsi, Jens Nilsson, Fernando Pereira, Kiril Ribarov, Sebastian Riedel, Giorgio Satta, Mario Scholz, and Deniz Yuret. In addition, we have drawn on our experience from teaching dependency parsing in courses on computational linguistics and parsing, and we are grateful to all the students who attended these classes, and whose questions and comments helped shape the material presented here.

We owe a special debt to the organizers of the 2006 CoNLL Shared Task, Sabine Buchholz, Amit Dubey, Yuwal Krymolowski, and Erwin Marsi, who set the stage for many of the recent developments in dependency parsing by creating a common platform for research and evaluation.

Finally, we want to thank Gerald Penn, Marco Kuhlmann, and Liu Haitao, who read the first complete draft of the book and suggested numerous improvements. All remaining errors, however, are our own responsibility. We hope you enjoy the book.

Sandra Kübler, Ryan McDonald, and Joakim Nivre
Bloomington, New York City, and Uppsala
December 2008