Domain-Sensitive Temporal Tagging

Synthesis Lectures on Human Language Technologies

Semantic Similarity from Natural Language and Ontology Analysis
Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain
2015

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition
Hang Li
2014

Ontology-Based Interpretation of Natural Language
Philipp Cimiano, Christina Unger, and John McCrae
2014

Automated Grammatical Error Detection for Language Learners, Second Edition
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2014

Web Corpus Construction
Roland Schäfer and Felix Bildhauer
2013

Recognizing Textual Entailment: Models and Applications
Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto
2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax
Emily M. Bender
2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing
Anders Søgaard
2013

Semantic Relations Between Nominals
Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
2013

Computational Modeling of Narrative
Inderjeet Mani
2012

Natural Language Processing for Historical Texts
Michael Piotrowski
2012

Sentiment Analysis and Opinion Mining
Bing Liu
2012

Discourse Processing
Manfred Stede
2011

Bitext Alignment
Jörg Tiedemann
2011

Linguistic Structure Prediction
Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing
Hang Li
2011

Computational Modeling of Human Language Acquisition
Afra Alishahi
2010

Introduction to Arabic Natural Language Processing
Nizar Y. Habash
2010

Cross-Language Information Retrieval
Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce
Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling
Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems
Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics
Graham Wilcock
2009

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

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2008

Domain-Sensitive Temporal Tagging

Jannik Strötgen and Michael Gertz

ISBN: 978-3-031-01035-4 paperback
ISBN: 978-3-031-02163-3 ebook

DOI 10.2200/S00721ED1V01Y201606HLT036

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

Lecture #36
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

Domain-Sensitive Temporal Tagging

Jannik Strötgen
Max Planck Institute for Informatics, Saarbrücken, Germany

Michael Gertz
Heidelberg University, Germany

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #36

ABSTRACT

This book covers the topic of temporal tagging, the detection of temporal expressions and the normalization of their semantics to some standard format. It places a special focus on the challenges and opportunities of domain-sensitive temporal tagging. After providing background knowledge on the concept of time, the book continues with a comprehensive survey of current research on temporal tagging. The authors provide an overview of existing techniques and tools, and highlight key issues that need to be addressed. This book is a valuable resource for researchers and application developers who need to become familiar with the topic and want to know the recent trends, current tools and techniques, as well as different application domains in which temporal information is of utmost importance.

Due to the prevalence of temporal expressions in diverse types of documents and the importance of temporal information in any information space, temporal tagging is an important task in natural language processing (NLP), and applications of several domains can benefit from the output of temporal taggers to provide more meaningful and useful results.

In recent years, temporal tagging has been an active field in NLP and computational linguistics. Several approaches to temporal tagging have been proposed, annotation standards have been developed, gold standard data sets have been created, and research competitions have been organized. Furthermore, some temporal taggers have also been made publicly available so that temporal tagging output is not just exploited in research, but is finding its way into real world applications. In addition, this book particularly focuses on domain-specific temporal tagging of documents. This is a crucial aspect as different types of documents (e.g., news articles, narratives, and colloquial texts) result in diverse challenges for temporal taggers and should be processed in a domain-sensitive manner.

KEYWORDS

temporal tagging, temporal expressions, temporal annotation, time, time extraction, time recognition, time normalization, temporal information, domain sensitivity, cross-domain temporal tagging, domain-sensitive temporal tagging, information extraction, temporal information extraction, TIMEX3, TimeML, HeidelTime, timelines

Contents

List of Figures

List of Tables

Preface

Time matters! Whatever document we read, be it a news article, biography, some microblog, or a patient's record, to name but a few examples, temporal information embedded in the documents typically helps us determine the course of events and actions, to correlate events, and eventually to get an overview of the documents' content. Driven by the continuously increasing amount of textual data that is available on the Web, in electronic archives, and Intranet document repositories the computer-supported analysis and exploration of textual data has become a necessity and also a challenge in numerous application domains. Named Entity Recognition (NER), that is, the task of information extraction that aims at detecting and classifying elements in some text into predefined classes, such as locations, persons, organizations, and temporal expressions, has become a cornerstone of tools and techniques that help to address this challenge.

Only in the past two decades has the topic of temporal tagging as a specific type of NER task become a major focus in research and development. Temporal tagging addresses the extraction, classification, and normalization of temporal expressions that occur in text documents, and it is the prerequisite for temporal information extraction. By now, the important role of temporal tagging has been well recognized in application domains such as text summarization, question answering, information retrieval, and topic detection and tracking. In these applications of temporal tagging, results can be as simple as the fully automated construction of a timeline of events detected in a document's content and can be as complex as revealing the temporal discourse structure in documents.

To date, there is no book that provides a comprehensive overview of the various methods, tools, evaluation competitions, and challenges the tasks of temporal tagging are faced with in the presence of diverse types of textual data and application domains. This book aims at closing this gap. Starting from the very fundamental role and concepts of time in documents, it provides an up-to-date overview of annotation standards, techniques, and competitions for evaluating the quality of temporal taggers, annotated corpora (including non-English texts) used for evaluations and developments, as well as a detailed overview of temporal taggers.

As the title indicates, this book focuses particularly on temporal tagging of documents from different domains, including text data different from the well-studied domain of news articles. For this, we discuss the challenges and approaches temporal taggers have to consider when processing news-style, narrative-style, colloquial-style, and so-called autonomic-style documents, the latter covering documents that contain many temporal expressions that cannot be normalized to real points in time, but only according to some local or autonomic time frame. Examples of autonomic-style documents are specific types of scientific texts and literary works.

We believe that this book provides researchers, practitioners, and developers a valuable resource for designing and improving temporal tagging techniques and tools, or just for applying them in a useful manner as part of more complex text analysis and exploration pipelines. While publicly available temporal taggers already provide sophisticated output for several application scenarios, there is still a lot of work in this area ahead of us. This book aims at providing a solid foundation on which such work can be built.

Jannik Strötgen and Michael Gertz
Saarbrücken, Germany and Heidelberg, Germany
July 2016

Acknowledgments

This book gives an in-depth overview of methods, tools, and techniques of temporal tagging in different domains. Based on the number of publications and evaluation competitions, the past few years clearly show that this field is taking on an enormous interest in the research community and industry. We thus would like to thank all researchers who actively contribute new ideas to this field, organize evaluation competitions, and provide temporal tagging tools and resources for other researchers and the public.

Although this book is about temporal tagging in general and not just about our temporal tagger HeidelTime, we want to take the opportunity to thank all contributors of HeidelTime for their great work and many users for helpful feedback to further improve the tool. We also would like to thank the many students at Heidelberg University who contributed in the form of student projects, and bachelor and master theses.

In particular, we thank Anne-Lyse Minard and Steven Bethard for their valuable reviews of the draft of this book. They put a lot of effort into the reviews and provided numerous valuable comments as well as suggestions to significantly improve the book. Finally, we want to thank the series editor Graeme Hirst for his great support and his instant replies to all our questions. It is time for a big thank you!