

# Book Review

## Domain-Sensitive Temporal Tagging

Jannik Strötgen, Michael Gertz

(Max Planck Institute for Informatics, Heidelberg University)

Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 36), 2016, xvii+133 pp; paperback, ISBN 978-1-62705-495-1; ebook, ISBN 978-1-62705-499-7; doi:10.2200/S00721ED1V01Y201606HLT036, \$55.00

*Reviewed by*

*Ruihong Huang*

*Texas A&M University*

Understanding time as expressed in text is an important goal of natural language understanding and extremely important for many applications, including information extraction, information retrieval, and question answering. This book provides a comprehensive overview, the challenges, available data resources, and existing systems for the task, with a special emphasis on sensitivity of temporal tagging to domains. This is a well-written book with contents well structured and organized. The discussions on temporal tagging for different domains are valuable and inspirational not only to readers interested in the specific subject of temporal tagging, but also to a broader range of readers interested in Natural Language Processing (NLP) in general. Although it is well known that domain changes often cause significant performance reductions of various NLP systems, little work has been conducted to understand and further explain specific ways such influences have been applied. To a great extent, this book fills the gap in the context of temporal tagging.

This book consists of six chapters, which I will group into three parts. The first three chapters give a clear introduction to temporal tagging, including subtasks, characteristics of time, realizations of temporal expressions, data annotation standards, data sets, and evaluation metrics. This initial part provides sufficient background knowledge for further discussion of domain influences on temporal tagging. The fourth chapter is the core of the book, and it identifies four major domains—news-style, narrative-style, colloquial-style, and autonomic-style documents—and elaborates on unique characteristics of each domain and their implications on temporal tagging. The last two chapters describe a list of temporal taggers (full-fledged or focusing on one stage of temporal tagging), compare their designs (rule-based vs. learning-based) and their capabilities of addressing multiple domains and even multiple languages, and conclude with future directions of temporal tagging.

Chapter 1 specifies the two subtasks of temporal tagging, temporal expression extraction, and normalization, and explains that temporal tagging can be viewed as a specific type of named entity recognition and normalization. Chapter 1 also briefly describes several temporal tagging applications, including information extraction, information retrieval, and question answering.

Chapter 2 clearly describes key characteristics of time. Specifically, time can be normalized and temporal information can be organized in a hierarchical structure based

---

doi:10.1162/COLL\_r.00319

© 2018 Association for Computational Linguistics

Published under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license

on their granularities. The chapter goes on to describe four categories of temporal expressions in real text (i.e., date, time, duration, and set). I found the discussion on differences between a point in time that may have a duration and a duration of time very interesting. The subsection on realizations of temporal expressions is at the core of this chapter, and clearly defines four types of temporal expressions—explicit, implicit, relative, and underspecified expressions. It is important to distinguish between these types before we examine differences of temporal tagging across domains. Note that recognizing and normalizing each type of temporal expression requires different strategies and is at a different difficulty level. Meanwhile, the authors discuss uncertainty or fuzziness of some temporal expressions. For instance, in “He visited Germany in 2010,” it is rather unlikely that the visit took place the whole year. The exact point or period in 2010 is not known.

Chapter 3 surveys annotation standards, describes several evaluation metrics, and provides a comprehensive list of research competitions and annotated news-style corpora. Although I found the description of annotation standards to be generally well thought out and organized, I occasionally felt it was difficult to understand some of the description. For instance, it is difficult to immediately understand the major differences between TIMEX2 and TIMEX3, the description of TIMEX3, and its various tags and abstract tags with no extent in TIMEX3. More examples would be helpful. Instead of sequentially reading each section in this chapter, it may help to first read Section 3.4 on annotated corpora. This chapter also includes an extensive description of different metrics used for measuring temporal tagging performance. The list of research competitions covers all the recent major efforts that were indicated by their adopted corpora, including news-style corpora (MUC, ACE, and TempEval), biomedical texts, QA TempEval (news, wiki, and blogs), and multi-language annotated corpora.

Chapter 4 is the core of the book, and defines four major types of domains, examines their unique characteristics, and discusses strategies of temporal tagging for each domain. The chapter starts by describing specific characteristics of news-style documents, which is the dominant type in most studies on temporal tagging. This is followed by a general discussion of genres or domains, covering news, Wikipedia, dialogs, short messages, and clinical reports. Specifically, this book defines a domain as a group of documents that have the same characteristics relevant for the task of temporal tagging. After providing a list of annotated corpora that includes non-news texts, the chapter identifies four broad types of domains—news-style, narrative-style, colloquial-style, and autonomic-style documents—which I found fascinating, especially the fourth domain that features unresolvable time expressions due to local time frames. The discussions of unique features for each domain with respect to temporal tagging are clear and well organized, and directly lead to strategies as suggested by the authors for addressing the task in each domain. Note that each domain is defined in a broad sense and covers texts created in several scenarios. For instance, the news-style documents include “not only news articles but also many other types of documents (e.g., letters and formal blog posts), which are written similarly and thus belong to the same domain from a temporal tagging point of view.”

Chapter 5 describes several widely used temporal taggers, both rule-based and learning-based taggers, and compares their temporal tagging performance on documents in different domains. Clearly, taggers prepared or trained for a particular domain do not perform well when applied to a different domain. Consistent with the authors’ vision, this chapter emphasizes that temporal taggers developed with all the domains in mind are preferable. This chapter also argues for developing highly multilingual temporal taggers.

Chapter 6 concludes the book by pointing out more directions for the future work of temporal tagging, in order to achieve accurate and complete temporal understanding.

To summarize, this book provides timely references to most recent advances of temporal tagging, including both annotated corpora from different domains and systems, as well as insightful discussions and vision in terms of categorizing effects of domains and designing generalized temporal taggers. This book is recommended not only to students, researchers, and developers who work in the field of temporal tagging, but also to a wide range of readers interested in NLP in general.

*Ruihong Huang* is an assistant professor in the Department of Computer Science and Engineering, Texas A&M University, College Station, Texas, USA. Her research interests are mainly in natural language processing, including information extraction, event recognition and extraction, deceptive and abusive language detection, semantics, and discourse understanding. Huang's e-mail address is [huangrh@cse.tamu.edu](mailto:huangrh@cse.tamu.edu).