

Doctoral Dissertation Academic Year 2016

Entity Linking with a Knowledge Base

Graduate School of Media and Governance
Keio University

Ikuya Yamada

Copyright © by Ikuya Yamada, 2016
All Rights Reserved.

Abstract

Entity linking (EL) is the task of detecting entity names from a document and resolving the detected names into corresponding entries in a knowledge base (e.g., Wikipedia, DBpedia, and Freebase). EL resolves the ambiguity in entity names by linking them to corresponding unambiguous identifiers of a knowledge base. Further, through EL, a document can be enriched with quality information available in a knowledge base. Therefore, EL has been widely used as a base component in various natural language processing tasks such as information extraction, information retrieval, and question answering.

In this dissertation, we propose two novel methods to significantly enhance the performance of EL. First, we describe our EL method, which is based on the joint embedding of words and entities. We also attempt to improve the performance of EL in Twitter messages. In our experiments, our proposed methods successfully outperform other state-of-the-art methods.

Furthermore, in order to validate the applicability of EL for existing applications, we provide two application examples of EL. We first propose a novel approach for enhancing the performance of named entity recognition (NER) in Twitter messages using EL. Further, we describe a new EL application that enhances a user's text reading experience by automatically converting entity names in a document into links. As a result, we confirm that EL can be effectively used for these applications.

Keywords: Entity linking, wikification, Wikipedia, named entity recognition, knowledge base

Acknowledgements

I want to express my sincere appreciation towards my supervisor Professor Yoshiyasu Takefuji for his valuable guidance throughout my doctoral studies. I would also like to thank Professor Hideaki Takeda, Professor Yasushi Kiyoki, and Professor Hideyuki Tokuda for their kind assistance. Without their support, this work would have not been possible. I am also grateful to Keio University for providing me with a great working environment.

Contents

List of Figures	8
List of Tables	9
1 Introduction	11
1.1 Motivation	11
1.2 Contributions	12
1.3 Organization	16
2 Entity Linking Using Embedding of Words and Entities	17
2.1 Introduction	17
2.2 Joint Embedding of Words and Entities	18
2.3 Entity Linking Using Proposed Embedding	22
2.4 Experiments	26
2.5 Summary	33
3 End-to-End Entity Linking for Tweets	35
3.1 Introduction	35
3.2 The Proposed System	36
3.3 Experiments	39
3.4 Summary	40
4 Enhancing Named Entity Recognition Using Entity Linking	41

4.1	Introduction	41
4.2	The Proposed System	42
4.3	Experiments	45
4.4	Summary	46
5	Enhancing Text Reading Experience by Linking Entities	49
5.1	Introduction	49
5.2	Application Architecture	51
5.3	Evaluating the Helpfulness of Entities	53
5.4	Summary	67
6	Related Work	69
6.1	Learning Representation of Entities for Entity Linking	69
6.2	Entity Recognition and Linking for Tweets	70
6.3	Enhancing Text Reading Experience by Linking Entities	71
7	Conclusions and Future Works	73
7.1	Conclusions	73
7.2	Future Works	74
A	Source Code of Joint Embedding of Words and Entities	77
	Bibliography	103