Academic Concept Extractor (ACE): Unsupervised concept based entity extraction from scientific titles

Adit Krishnan, Aravind Sankar, Shi Zhi, Jiawei Han
Department of Computer Science
University of Illinois at Urbana-Champaign
{aditk2,asankar3,shizhi2,hanj}@illinois.edu

ABSTRACT

This paper studies the extraction and typing of entities from titles of academic literature, in order to gain a deeper understanding of their specific contributions and automate the construction of a problem-solution database to analyze the relations between them. To achieve this goal, we propose an unsupervised, domain independent, two phase algorithm to extract entity mentions and type them into appropriate concepts. In the first phase of our algorithm we propose a generative model which exploits textual and syntactic features to broadly segment titles and type them into concepts. In the second phase, we propose an unsupervised approach based on adaptor grammars to extract fine grained entities of interest without the need for any external resources or human effort, in a purely data driven manner. We analyze literature from diverse scientific domains and show significant gains over state-of-the-art techniques. We also present an analysis and summarization of the knowledge base constructed as part of our algorithm.

CCS CONCEPTS

•**Computer systems organization** → **Embedded systems;** *Redundancy;* Robotics; •**Networks** → Network reliability;

KEYWORDS

ACM proceedings, LaTeX, text tagging

ACM Reference format:
Adit Krishnan, Aravind Sankar, Shi Zhi, Jiawei Han. 2017. Academic Concept Extractor (ACE): Unsupervised concept based entity extraction from scientific titles. In *Proceedings of ACM CIKM, Singapore , Nov 2017 (CIKM'17),* 1 pages.
DOI: 10.475/123_4

1 INTRODUCTION

2 CONCLUSIONS

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

ACKNOWLEDGMENTS

The authors would like to thank ...

CIKM'17, Singapore

DOI: 10.475/123_4