Universität des Saarlandes
Max-Planck-Institut für Informatik
AG5

Learning Rules With Numerical Constants in Large Uncertain Knowledge Bases

Masterarbeit im Fach Informatik
Master's Thesis in Computer Science
von / by

André de Oliveira Melo

angefertigt unter der Leitung von / supervised by

Prof. Dr. Gerhard Weikum

betreut von / advised by

Dr. Martin Theobald

begutachtet von / reviewers

Dr. Max Mustermann

Prof. Dr. Gerhard Weikum

November / November 2012

Hilfsmittelerklärung

Hiermit versichere ich, die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Non-plagiarism Statement

Hereby I confirm that this thesis is my own work and that I have documented all sources used.

Saarbrücken, den 14. November 2012,

(André de Oliveira Melo)

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlichtwird.

Declaration of Consent

Herewith I agree that my thesis will be made available through the library of the Computer Science Department, Saarland University.

Saarbrücken, den 14. November 2012,

(André de Oliveira Melo)

To my father Cicero, my mother Marlene and my sister Carolina

- Andre

Abstract

With millions of articles in multiple languages, Wikipedia has become the de-facto source of reference on the Internet today. Each article on Wikipedia contains encyclopedic information about various topics (people, events, inventions, etc.) and implicitly represents an entity. Extracting the most important facts about such entity will help users to find desired information more quickly and effectively. However, this task is challenging due to the incomplete and noisy nature of Wikipedia articles. This calls for a mechanism to detect and summarize the most important information about an entity on Wikipedia.

This thesis proposes and implements CATE (**C**ontext-**A**ware **T**imeline for **E**ntity Exploration), a framework that utilizes Wikipedia to summarize and visualize the important aspects of entities in a timeline fashion. Such a system will help users to draw quickly an informative picture of an entity (e.g. life of a person, or evolution of a research topic, etc.). The novelty of CATE lies in seeing the entity in different contexts, synchronous with contemporaneous events. In addition, CATE puts the entity in a relationship with other entities, and thus offers a broader portrait about it. In order to efficiently query and visualize the events related to the entity, a number of techniques have been developed, combining information extraction and information retrieval with a novel ranking model. The thesis also discusses several experiments and evaluation results to show the effectiveness of the methods proposed.

Acknowledgements

First and foremost, I would like to thank my advisors Dr. Nicoleta Preda and Shady Elbassuoni, for their invaluable guidance. I feel deeply indebted to their strong advises, encouragements and instructions. They have not only assisted me in completing this thesis, they have also helped me to broaden my attitude towards research, and to develop my personality.

A special note of thanks to Prof. Gerhard Weikum for giving me the opportunity to pursue this thesis under his supervision and for his valuable inputs. I am very grateful to work with his group in the Max-Planck Institute for Informatics. It is a priceless experience to me.

I am grateful to Dr. Martin Theobald to review my master thesis.

I am also very grateful to my friends and classmates for their company and moral support. I learnt a lot from them in last 2 years which is a very important factor for successful completion of my thesis. Special thanks are sent to the colleague students in the lab room 4.11 for the memorable time working with them.

Finally, I would like to thank IMPRS-CS for the assistance and financial support that I received from them. It gives me total freedom to work on my master thesis.

Inhaltsverzeichnis

Abstract vi

Acknowledgements viii

List of Figures 1

List of Tables 3

Literaturverzeichnis 5

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis