

# Exposé

outlining a Master Thesis on:

Semantic approaches to scientific citation recommendation (tentative title)

Tarek Saier

Reviewer: Prof. Dr. Georg Lausen

Advisor: Dr.-Ing. Michael Färber

## 1. INTRODUCTION

This exposé will outline a prospective Master Thesis in the area of scientific citation recommendation and argue for its value. The approach will encompass the creation of a dataset and development of supervised learning methods with a focus on semantic analysis of citation contexts. Evaluation of the resulting implementation will follow the most prevalent methods in the field.

The remainder of this document is structured as follows. Section 2 will provide some theoretical background on relevant areas and give a quick overview of related work. A detailed description of the planned methodology and approach will be given in section 3; followed in section 4 by an outline of the planned evaluation. Section 5 and 6 conclude the exposé by listing the foreseen contributions of the Thesis and a proposed schedule.

## 2. BACKGROUND

foo

### 2.1 Related work

- survey stuff (take more narrow look) - wiki page news article papers (for entitiy stuff)

## 3. METHODOLOGY AND APPROACH

foo bar MAG[7][3][4][8] entity[6] claim[5] argument[2]

## 4. EVALUATION

foo bar

## 5. CONTRIBUTIONS

- apparently semantic stuff not very explored (cite survey if possible, look at tables) - creation of another nice (exact citation markers, large citation context, etc.) dataset like gold standard paper[1] - a nice dataset like gold standard paper[1] but not restricted to CS domain

## 6. SCHEDULE

foo bar

## 7. REFERENCES

- [1] M. Färber, A. Thiemann, and A. Jatowt. A High-Quality Gold Standard for Citation-based Tasks. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, LREC 2018, 2018. r.
- [2] T. Goudas, C. Louizos, G. Petasis, and V. Karkaletsis. Argument extraction from news, blogs, and social media. In A. Likas, K. Blekas, and D. Kalles, editors, *Artificial Intelligence: Methods and Applications*, pages 287–299, Cham, 2014. Springer International Publishing. r.
- [3] D. Herrmannova and P. Knoth. An analysis of the microsoft academic graph. *D-Lib Magazine*, 22(9/10), 2016. r.
- [4] S. E. Hug, M. Ochsner, and M. P. Brändle. Citation analysis with microsoft academic. *Scientometrics*, 111(1):371–378, Apr 2017. r.
- [5] R. Levy, Y. Bilu, D. Hershcovich, E. Aharoni, and N. Slonim. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics. r.
- [6] A. Mishra and K. Berberich. Leveraging semantic annotations to link wikipedia and news archives. In N. Ferro, F. Crestani, M.-F. Moens, J. Mothe, F. Silvestri, G. M. Di Nunzio, C. Hauff, and G. Silvello, editors, *Advances in Information Retrieval*, pages 30–42, Cham, 2016. Springer International Publishing. r (ch 1-3).
- [7] B. Paszcza. Comparison of microsoft academic graph with other scholarly citation databases, 11 2016. r (ch 1,""3"").
- [8] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. P. Hsu, and K. Wang. An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 243–246, New York, NY, USA, 2015. ACM. r.

Winter semester 2018/19

Albert-Ludwigs Universität Freiburg  
Technische Fakultät, Institut für Informatik  
Lehrstuhl für Datenbanken & Informationssysteme