SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

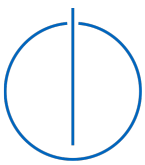
TECHNISCHE UNIVERSITÄT MÜNCHEN

Expose for Bachelor's Thesis in Informatics

User-centric design of an exploratory search system for scholarly entities in Natural Language Processing

Nutzerzentierter Design eines explorativen Suchsystems für wissenschaftliche Entitäten in Natural Language Processing

Author:	Ferdy Dermawan Hadiwijaya
Supervisor:	Prof. Dr. Florian Matthes
Advisor:	Tim Schopf
Submission Date:	17.07.2023



Expose

1 Introduction

Natural Language Processing (NLP) is a rapidly growing research field. It has seen exponential growth over the years. The rate of publication release poses a problem, making it difficult for new and experienced researchers alike. Professional researchers have difficulty keeping up with state-of-the-art papers. Other researchers face challenges getting into a new research subfield of NLP due to the sheer amount of existing publications.

There are multiple tools researchers use to explore and search papers. Researchers follow journal pages or use their physical and virtual social networks to explore and discover the latest advancements in NLP. Then, they would use the publications search engine, such as google scholar, to find these papers. This approach requires the researchers to use multiple tools simultaneously, making it challenging to understand connections between publications. The existing tools are generic and don't provide features specific to NLP researchers' needs.

This thesis delves into a web application solution to improve researchers' current exploration and discovery process. The web application contains scholarly entities such as publications, research fields, and researchers. This web application aims to highlight connections between scholarly entities and inform researchers of trends in NLP research. We plan to streamline NLP researchers' exploration and search process, simplifying the existing approaches.

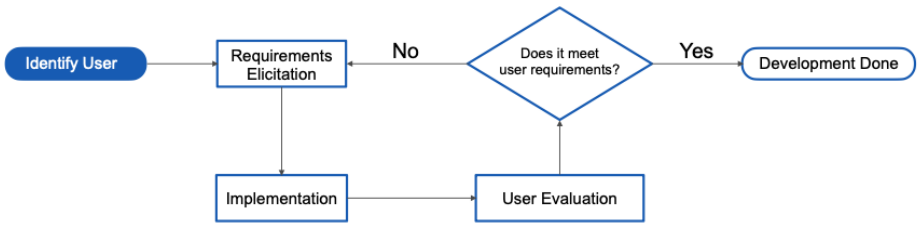
2 Research Questions and Methodology

This thesis will explore these four research questions:

1. **What are the existing approaches of researchers to search, explore, and keep up with the NLP research?**

   We conduct interviews with NLP researchers to understand their pain points, analyze their needs, and validate if an exploratory search solution can help tackle their problems.

2. **How can we curate and present information in a web application to support user-friendly search and exploration of scholarly entities in NLP?**



We employ an iterative user-centric design process. We start with identifying target users through persona. Then, we start the iterative process by analyzing the user's need through requirements elicitation. We then implement according to this, which will be evaluated by another set of users to remove bias. If fitting, we start a new iteration process. By the end of these iterations, we will have developed a usable system that adds value to the user.

3. **What approaches can we use to achieve a performant semantic search and exploration of relevant scholarly entities in Natural Language Processing?**

   Within the iterative user-centric design, we also apply an iterative development process through technical literature review, rapid implementation, and systematic evaluation. Due to the time limit of this thesis, we maintain the fragile balance of rapid development and user experience.

4. **How can we systematically evaluate the usefulness of our proposed approach to our target users?**

   This research question aims to analyze the user's experience with the product and evaluate its usability. This evaluation is achieved by an extended version of System Usability Testing that employs Likert Scale and open-ended questions.

3 Conclusion

In conclusion, this thesis proposes a web application solution to address the challenges faced by researchers in the exponentially growing field of Natural Language Processing (NLP). By addressing these research questions and developing a user-centric designed web application, this thesis aims to contribute to the improvement of the exploration and discovery process for NLP researchers.