

Project Outline

Magdalena Fritz - Matr. Nr. 01630011

Supervisor: Dr. Jana Lasser, Co-supervisor: Dr. Allan Hanbury

Topic

The topic of my interdisciplinary project is “**Challenges of motherhood - Young MUM's urgent questions**”. In collaboration with the Institute for Interactive Systems and Data Science from the University of Technology of Graz and the Complexity Science Hub Vienna, where Dr. Jana Lasser is my supervisor, I'll try to provide answers to questions such as: What are the challenges of motherhood? Which questions concern young mothers?

The topic and these questions arose on request of the organization “Die Möwe - Kinderschutz hat einen Namen”, that wants to find out which questions are in fact really important to a majority of young mothers. As a result of my project they want to answer these questions to support young, inexperienced mums and to offer them useful information.

General Approach

As a first step of my project I'll explore the platform CrowdTangle and use it to collect data. CrowdTangle is owned by Facebook and enables a user to aggregate all meaningful Facebook, Twitter, Instagram, and Reddit posts for a given query. In my project I want to work with Facebook data, whereby I'll extract text data either from specific forums where young mothers are active or by searching for keywords related to motherhood, which I'll prepare together with domain experts from the organization “Die Möwe”.

An important part of my work will be “Vocabulary Expansion”. The purpose is to extend the vocabulary on young mother-related domains with user-generated slang and colloquial forms. Thus we probably notice the importance of enriching domain expertise with user-generated content and therefore fully incorporate the language complexity of online discussions. The vocabulary expansion will hopefully increase the richness of the terminology and, consequently, the volume of conversations on Facebook that contain these terms.

To extract the main topics (represented as a set of words) that occur in the corpus of posts collected from Crowd Tangle, I plan to use “Topic Modelling”. More specifically, the Latent Dirichlet Allocation (LDA) method can be used to cluster text into distinct topics and thus to extract the most important topics in the data set. The topic clusters can then be interpreted based on the most frequent words they contain. For each topic, questions can be identified by searching for sentences that contain a question mark. In this way questions that concern young mothers can be found out.

Domain-specific lecture

To gain theoretical knowledge in this area I'm attending the lecture “Entwicklungspsychologie I” at the university of Vienna this semester (2021W) that is held by Dr. Tina Eckstein-Madry.

Literature that could be used

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. “Latent dirichlet allocation.” the Journal of machine Learning research 3 (2003): 993-1022.

Balsamo, Duilio, et al. “Patterns of routes of administration and drug tampering for nonmedical opioid consumption: data mining and content analysis of reddit discussions.” Journal of Medical Internet Research 23.1 (2021): e21212.