



Bachelor's Thesis / Master's Thesis

Evaluating Different Approaches to Infer a Person's Gender and Ethnicity from the Name

In many situations of daily life, we read a person's name and infer the gender and ethnicity. For example, if we receive an e-mail from Heike Müller, it is immediately clear to us that this is a German woman, and we will automatically reply to her with Ms. Müller without explicitly thinking about it. Reading the name Dimitrij Acelya on a list of participants, we expect a man of Russian ethnicity. There are many more situations in which we get a person's name as the only information about a person and unconsciously indent on their gender and ethnicity.

It has already been shown in some situations, such as in the recruitment process, that there is discrimination against groups of people. Since it is interesting to investigate whether there is discrimination against people based on their name, it is important to automatically assign the gender and ethnicity on the basis of their name. As there are ambiguous names such as Kim which is not clearly attributable to a gender or Peter which is not clearly attributable to an ethnicity, a heuristik is necessary. Thus, the question arises:

How can a computer automatically determine the name and the ethnicity based on the name?

In order to answer this research question, a literature review should be carried out to identify different methods. Afterwards the identified methods should then be compared and evaluated on a test dataset.

Recommended literature:

- Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M., & Strohmaier, M. (2016). Inferring gender from names on the web: A comparative evaluation of gender detection methods. In Proceedings of the 25th International conference companion on WWW. pp. 53-54.
- Santamaría, L., & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. PeerJ Computer Science.
- Sood, G., & Laohaprapanon, S. (2018). Predicting race and ethnicity from the sequence of characters in a name. arXiv preprint arXiv:1805.02109.
- Harris, J. A. (2015). What's in a name? A method for extracting information about ethnicity from names. Political Analysis. pp. 212-224.
- Carsenat, E. (2019). Inferring gender from names in any region, language, or alphabet. doi:10.13140/RG.2.2.11516.90247.

Supervisor: Stefanie Krause, M.Eng. & Sebastian Duda, M.Sc.