Implementation of an educational tool: a multilingual search system and characterization of biased texts

Proposer(s) / Proposatzailea(k):

Itziar Aldabe

Nora Aranberri

Contact / Kontaktua:

itziar.aldabe@ehu.eus
nora.aranberri@ehu.eus

Description / Deskribapena

The main objective of the work is to create an educational tool, specifically, a search system and characterization of multilingual texts adapted for Secondary Education. The work will focus on the design, implementation and evaluation of a text search system. The system will be able to offer documents in multiple languages and characterized by bias so that teachers can select them according to the set learning objectives. In this project, the group of measures that can be used to identify and characterize bias will be explored first and the most promising measures implemented using Natural Language Processing (NLP) techniques. An approach based on neural networks will be explored.

How can we identify bias in a text? As Lucy et al. (2020) claim, not much research has been carried out on applying NLP techniques to answer sociological questions in education. As they describe "Some early efforts apply machine counting of words to scanned textbooks, such as Lachmann and Mitchell (2014)'s study on depictions of war. A number of recent studies outside education have used NLP methods to study the reflection of gender and other social variables in text: Fast et al. (2016) look at gender stereotypes in online fiction; Hoyle et al. (2019) measured the association of adjectives and verbs with different genders in a million digitized books; Garg et al. (2018) quantified a century of gender and ethnic stereotypes using word representations learned from books, newspapers, and other texts; and Ash et al. (2020) examine the role of gender slant in judicial behavior using text written by judges." (Lucy et al, 2020: 2). Building on the previous work, Lucy et al. (2020) examine depictions of social groups in history texts (see also Field et al., 2019; Joseph et al., 2017; Ornaghi et al., 2019), extending NLP methods to textbooks. These works present a good starting point to explore the domain.

Goals / Helburuak

The main objective of the project is to analyze and implement NLP approaches for bias detection and to integrate the defined characterization measures in a search engine. The key objectives are the following:

1. Analysis of the state of the art techniques for developing search engines and bias detection.
2. Design and implementation of a multilingual system to characterize bias in texts.
3. Design and implementation of a search engine.

Requirements / Betebeharrak

Machine learning. Good programming skills and basic math skills.

Although it is not a requirement, taking the course "**Seminar on language technologies. Deep Learning**" will allow the student to accomplish more ambitious goals. Contact us for further details.

The dissertation can be written in Basque, English or Spanish.

Framework / Esparrua

NLP applications for education

Python, pytorch/tensorflow

Tasks and plan / Atazak eta plana

- Analyze the state of the art on bias detection.
- Select the text data to work on.
- Design and implement a system to characterize bias in texts.
- Test and evaluate the implemented algorithm.
- Design and implement a search engine which integrates the measures to characterize bias in texts.
- Test and evaluate the complete system (search and characterization).
- Analyze the output of the system to 1) perform an error analysis and 2) propose possible improvements.
- Write up the report.

References

Ash, E., Chen, D. L., Ornaghi, A. (2020). Stereotypes in high-stakes decisions: Evidence from US Circuit Courts. National Bureau of Economic Research.

Fast, E., Vachovsky, T., Bernstein, M. S. (2016). Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In Tenth International AAAI Conference on Web and Social Media 2016.

Field, A., Bhat, G., Tsvetkov, Y. (2019). Contextual affective analysis: A case study of people portrayals in online #MeToo stories. In Proceedings of the international AAAI conference on web and social media (Vol. 13, No. 01, pp. 158–169).

Garg, N., Schiebinger, L., Jurafsky, D., Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences of the U S A, 115(16), E3635–E3644.

Hoyle, A., Wolf-Sonkin, L., Wallach, H., Augenstein, I., Cotterell, R. (2019). Unsupervised discovery of gendered language through latent-variable modeling. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 1706–1716). Association for Computational Linguistics.

Joseph, K., Wei, W., Carley, K. M. (2017). Girls rule, boys drool: Extracting semantic and affective stereotypes from Twitter. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing (pp. 1362–1374). Association for Computing Machinery.

Lachmann, R., Mitchell, L. (2014). The changing face of war in textbooks: Depictions of World War II and Vietnam, 1970–2009. Sociology of Education, 87(3), 188–203.

Lucy, L., Demszky, D., Bromley, P. and Jurafsky, D. (2020) Content analysis of textbooks via natural language processing: findings on gender, race, and ethnicity in Texas US history textbook, *AERA Open*, 6(3), 1-27.

Ornaghi, A., Ash, E., Chen, D. L. (2019). Stereotypes in high stake decisions: Evidence from US Circuit Courts (Working Paper 2). Center for Law & Economics.