

Evaluation of Development strategy of the territorial district of the Central Bohemian Region, for the period 2019-2024, up to 2030: Text mining results

The Development strategy of the territorial district of the Central Bohemian Region, for the period 2019-2024, up to 2030 (hereinafter the Strategy) was evaluated in Task 4.5 as one of the Central Bohemian pilot activities. We separately examined the Strategy as a core policy document of the Region in the field of regional development and then looked at all other policy documents.

Sources

Although in English, Italian and Spanish language the text mining using the SEMEX software brought very good results, in Czech language the results were confusing. However, this is not surprising, because Czech is a very complicated language in general, especially with regard to conjugation and declension. Also the semantic combination of expressions seems to be very problematic and currently makes the results of text mining difficult to use.



Figure 1. The average polarity scope of “Topics” found via TM process in the Development strategy of the territorial district of the Central Bohemian Region, for the period 2019-2024, up to 2030

Results

The results of the text mining showed invalid semantic combination of expressions which are hard to use for any practical purposes. For example, the word “oblast” (which means “area”)

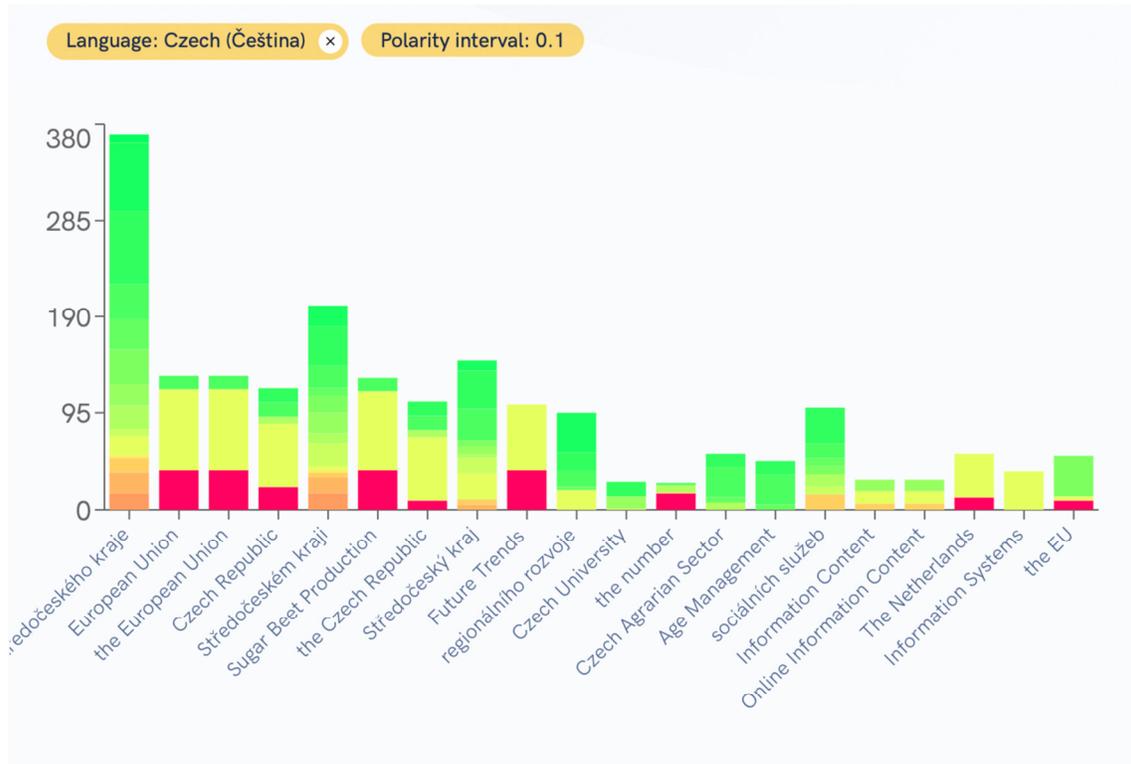


Figure 3. The 0.1 polarity scope of “Keywords“ found via TM process in all policy documents in Czech language presented on the SEMEX library

There was no option to create a chart after selecting the region, particularly the Central Bohemian Region.

Conclusion

Despite the above experience, the potential of the text mining via SEMEX software is promising. The Czech language is very complicated and causes many issues, as demonstrated above. This fact is described in many scientific publications and can be confirmed by experts.