POLIRURAL

Evaluation of Hämeenlinna region's local development strategy for 2014-2020: Text mining results

Häme pilot's focus in the PoliRural project is on supporting new and growing business opportunities in farms and rural areas, especially in entrepreneurship of circular economy, services and welfare.

In the policy evaluation study, we looked at policy measure of one of the Häme Leader local action group´s strategy for the region. The name of the strategy from which the measure was chosen is LOCAL TIME - Hämeenlinna region's local development strategy for 2014-2020, implemented by Linnaseutu Leader action group. This Leader strategy was chosen because it contains policy measures for supporting new and growing business opportunities in rural areas.

Sources

The reading list for text mining was created in Semex. The list contained 31 sources from HAMK's library of 122 sources on Semex. The sources were national and regional policies focusing on entrepreneurship and rural development; regional strategies; research studies and discussion groups on the Internet. Some documents were not processed due to errors.

Results

The polarity scores by Topics were quite green, so meaning they were positive. The negatives were topics like "city center", "promotion of trade and industry", "rural habitat", "association" and "economic sector" and "city service" (from left to right). The results were not expected and may be not credible. Many topics in the picture are dealing with different kinds of areas such of Antarctic/ Mediterranean/ contaminated/ Vulcanic/ rural/dangerous. The most frequent topic is "project".

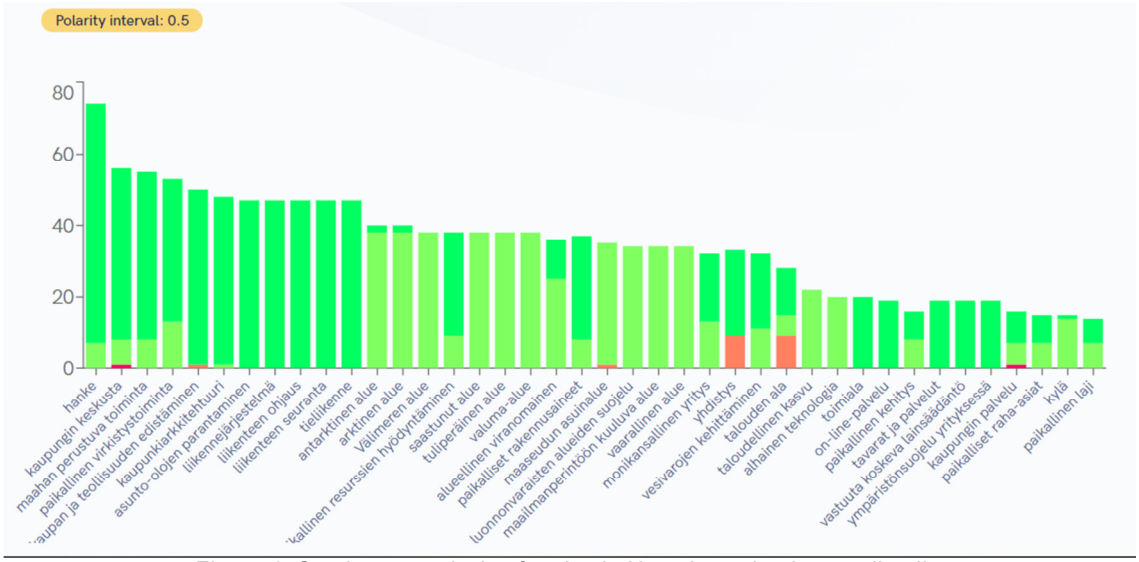

Figure 1. Sentiment analysis of topics in Hame's evaluation reading list

When source type is selected only to Policies, the picture is a little different, and the negative words are “city center”, “promotion of trade and industry” and “city service”.

When narrowing the region to Häme (FI), there were no results by polarity scores.

When using Keywords instead of Topics, the picture seems odd, words are like loose words put together as “promotion walk” or words with no meaning e.g. “lutuesna”.

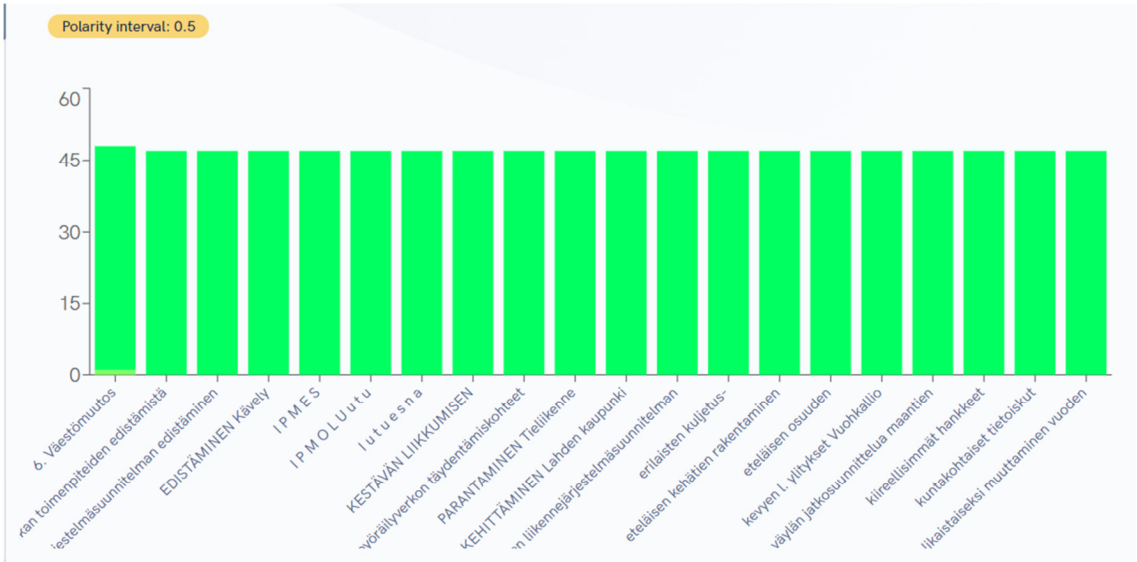

Figure 2. Semantic analysis of keywords in Hame evaluation reading list

Semantic diagram

It was possible to produce a custom semantic diagram for Reading List when using Analytics. To explore semantic relationships between keywords in the library the term ‘rural area, sparsely populated area’ appeared in the center. The main associations were linked to fallow area, temporary settlement, densely populated area, settlement, urban settlement, rural habitat, built residential area, sparsely populated area.

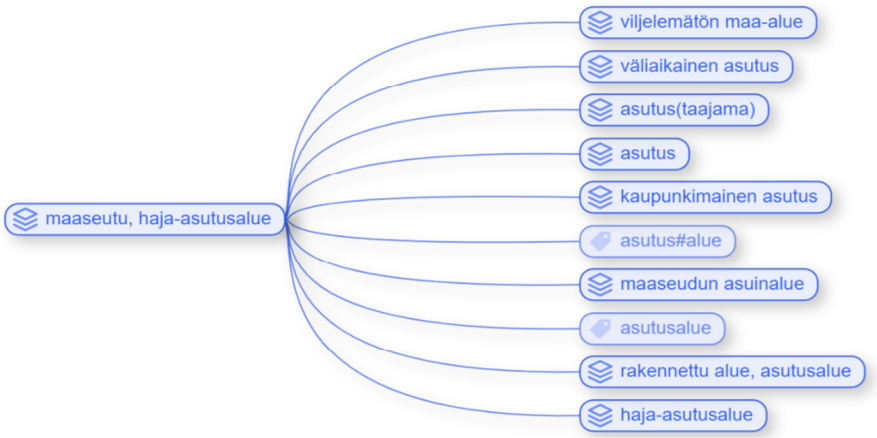

Figure 3. Semantic diagram

Named entity

Testing the Named Entities failed with Reading List, it didn´t open. Instead some tests were made with individual sources. In Linnaseutu Leader strategy there were three names from 13 words in the results. There were some wrong results in “topics” but “keywords” looked right.

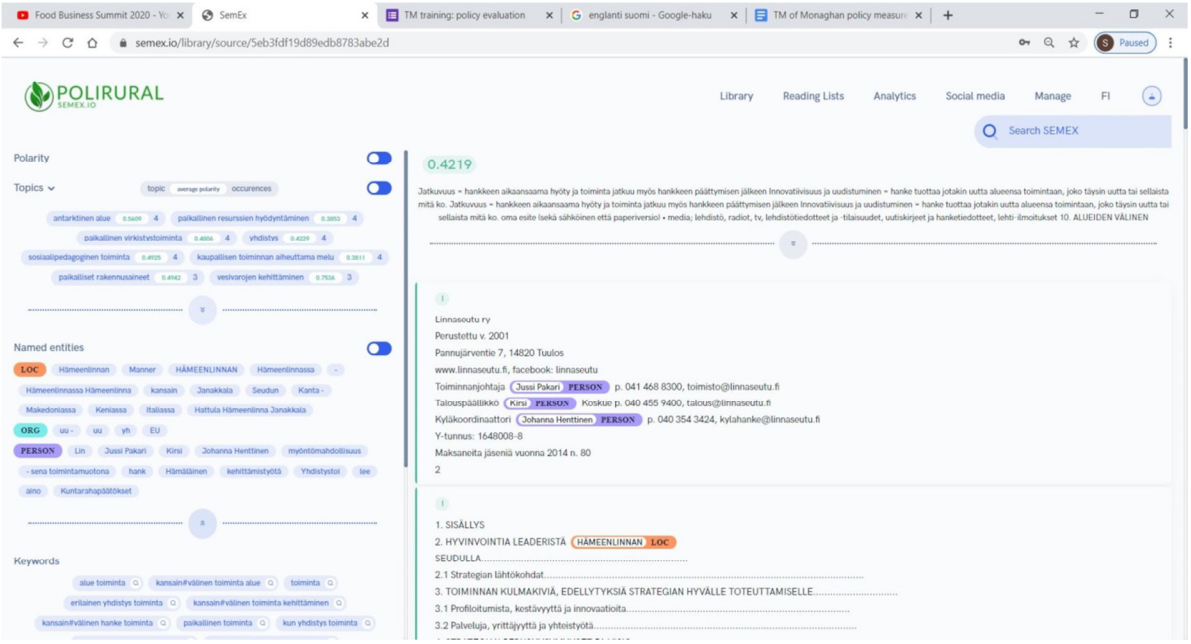

Figure 4. Named entities

In another trial in a different short article, the persons were right and their opinion was negative when disagreeing with the Airbnb actions as rural tourism entrepreneur.

Conclusion

In summary, these test trials didn´t provide further information on the policy evaluation. The tool is not reliable yet and has many mistakes at this stage of the development process, partly because of the language issue. Nevertheless, in the future, Semex can be a very useful tool for ranking and extracting details from large data sources.