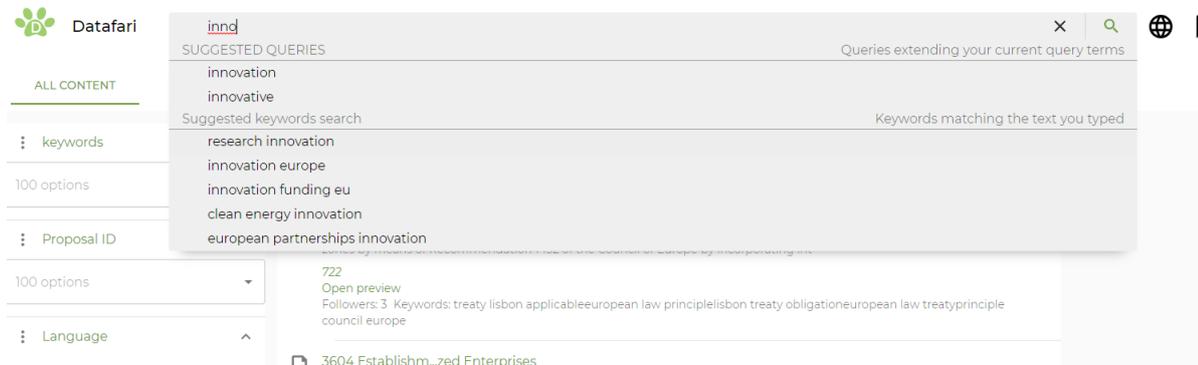## Challenge 5 – Team Datafari – Proposing an exploration tool for proposals, comments and extracted semantics
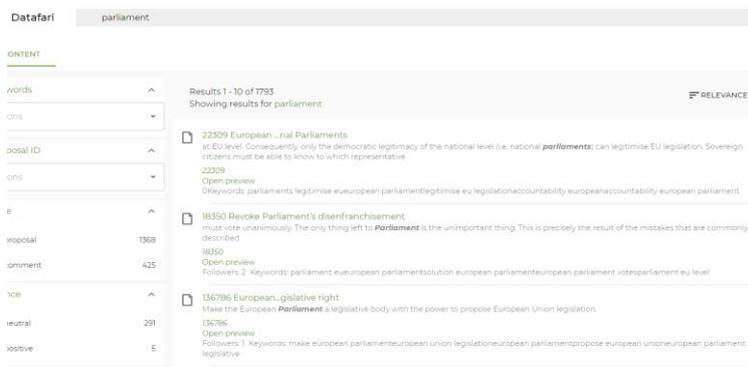
Extracting semantics such as stances or keywords is interesting, but not very valuable without an efficient way to use them. As part of challenge 5, we have setup an instance of Datafari (our open source enterprise search solution) to give an idea of what a modern search interface could look like. It is running live, and it could obviously be further enhanced had we had more time. We used existing NLP ML algorithms out of the box without particular optimisation, as our focus here is to show what a global solution could do. Note that due to lack of time, there is a little bug in the language used (it mixes French and English), and the language icon is not functional, sorry for the inconvenience.

What it does technically :
1. Indexing the proposals and comments csv files, keeping only the english content

2. Using spacy for automatic analysis of the proposals and comments, feeding the results to the search index
2.1 For keywords detections, we used KeyBERT [2]
2.2 For stances detections, we used SpacyTextBlob [3] : values below -0,4 were considered negative, between -0,4 and +0,4 neutral, higher than +0,4 positive

3. Allowing users to do full searches on
3.1 the text content of both proposals and comments to find any proposals/comments that contain the searched query (stemmed by default)
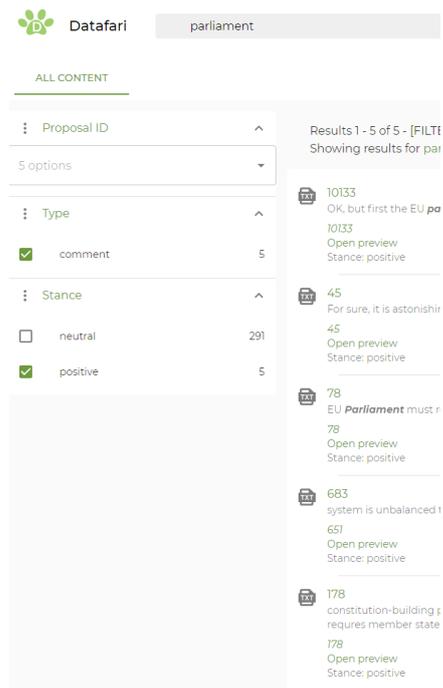3.2 the id of proposals to see a proposal and all of its comments
3.3 existing keywords

4. In terms of search results visualisation, users can
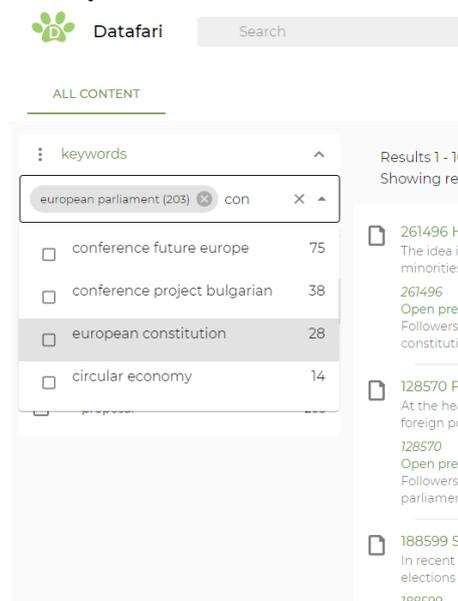4.1 see snippets of the results, with additional info like followers and id

4.2 quickly visualise stances of comments

4.3 filter to see only certain type of stances

4.4 filter through existing proposals keywords and do multiselection

Our demo is publicly available at https://datathon.datafari.com/datafariui
It is using the Community edition of Datafari which is fully open source, under Apache v2 licence.
The customisation of our UI is available in a dedicated gitlab public branche.

[1] Datafari Community Edition – https://www.datafari.com/en
[2] KeyBERT - https://github.com/MaartenGr/KeyBERT
[3] SpacyTextBlob - https://spacy.io/universe/project/spacy-textblob