Datathon COFE
Explore the citizens' propositions by Make.org

MAKE.
ORG

Overview

👉 An open, public interface

👉 In english only

👉 Powered by two semantic algorithms

Keyword Extraction

👉 Relevancy vs coverage balance

👉 Relevancy features:

- Consultation based Frequency features
- External corpus based frequency features
- POS tag based features
- NER based features
- Wikipedia based features

👉 Top 5 keywords by proposition
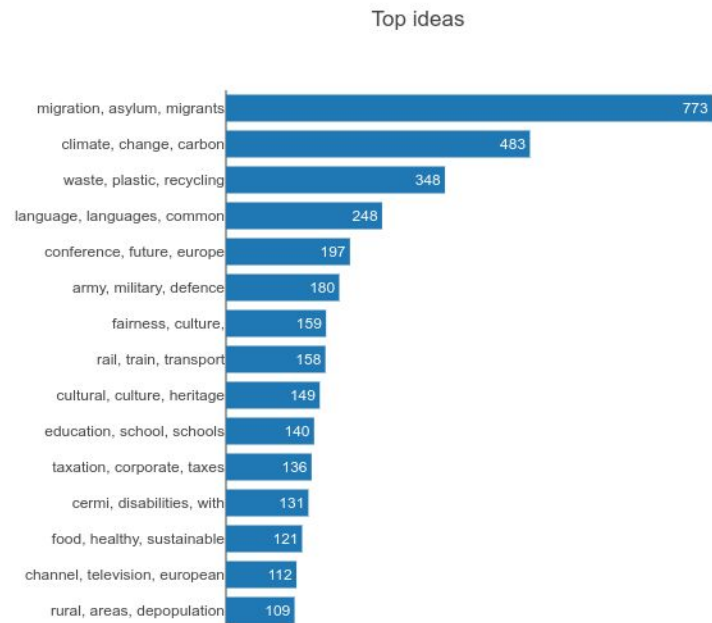
Clustering into top ideas

👉 Focus on top 100 ideas

👉 Main clustering steps:

- S-BERT embeddings
- UMAP projection => 150 dim.
- HDBSCAN clustering

👉 Cluster names:

- top 3 title words, tf-idf

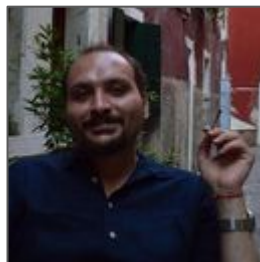Top ideas

Idea	Count
migration, asylum, migrants	773
climate, change, carbon	483
waste, plastic, recycling	348
language, languages, common	248
conference, future, europe	197
army, military, defence	180
fairness, culture,	159
rail, train, transport	158
cultural, culture, heritage	149
education, school, schools	140
taxation, corporate, taxes	136
cermi, disabilities, with	131
food, healthy, sustainable	121
channel, television, european	112
rural, areas, depopulation	109

Demo !

https://dial.make.org/future_eu

Thank you !

https://dial.make.org/future_eu

David MAS
dm@make.org

Hazem AL SAIED
hazem.al.saied@make.org

Valentine SANTARELLI
valentine.santarelli@make.org