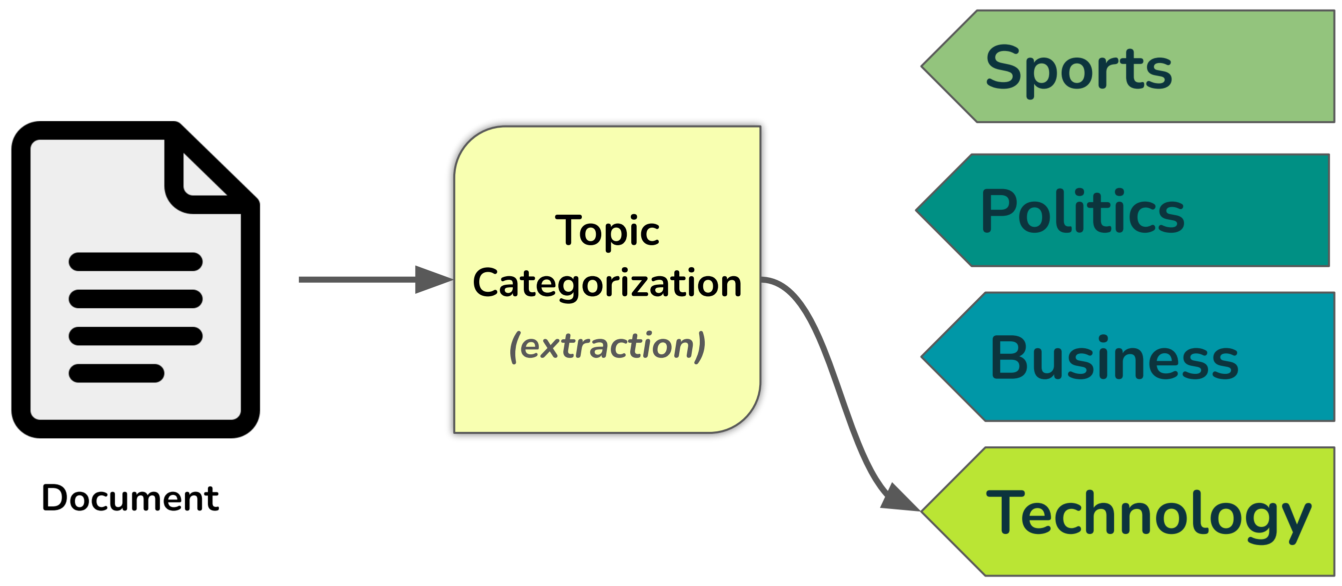
Explainable Zero-Shot Topic Extraction (ZeSTE) Using a Common-Sense Knowledge Graph

Ismail Harrando and **Raphaël Troncy**
EURECOM - Sophia Antipolis, France

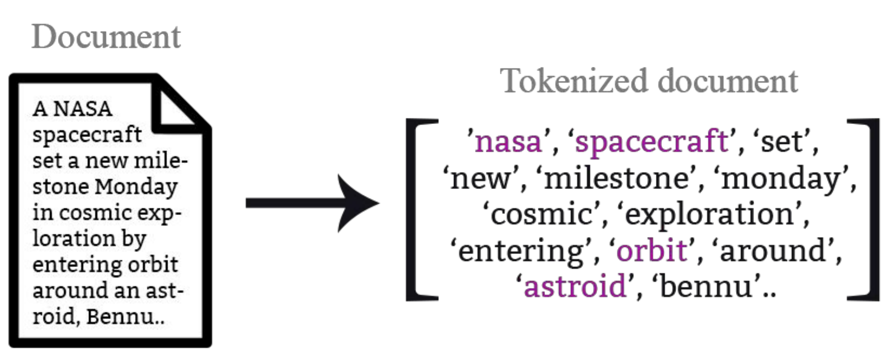
Topic Categorization

A text classification task where the goal is to label a textual document (e.g. news articles, video transcripts, etc.) into one of multiple predefined *topics*, i.e. labels that are related to the topical content of the document.

Common examples for news topics are "Politics", "Sports" and "Business".



ZeSTE Approach

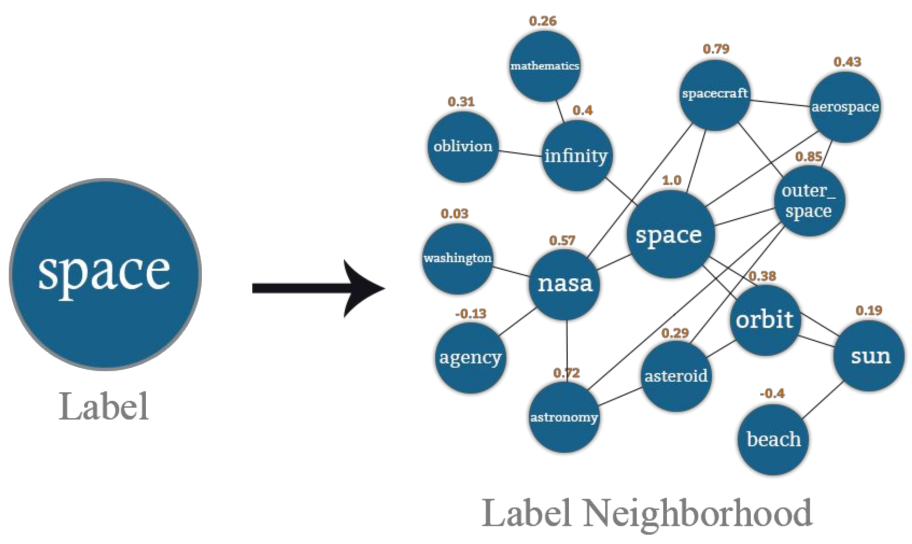


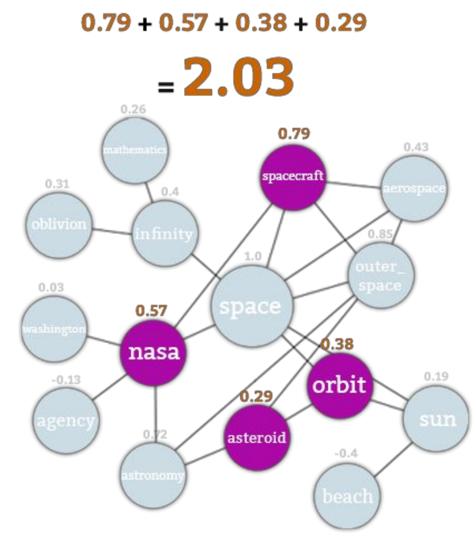
1. Preprocessing and Tokenization :

- Lowercase the text
- Remove all non-alphabetical symbols and stopwords.
- Tokenize the strings using the space as separator
- Lemmatize the word using WordNetLemmatizer

2. Generating the Topic Neighborhood :

by querying ConceptNet [1] to get every node that is *N* hops away from the label node. Every node is then given a score that is based on the cosine similarity between the ConceptNet Numberbatch [2] embeddings of the label and the node. This score represents the relevance of any term in the neighborhood to the main label, and would also allow us to refine the neighborhood and produce a score.





3. Generate a topic score :

Once the neighborhood is generated, we can predict the document label by adding up the scores of each node in the overlap between the document content and the label neighborhood. Each node score is computed based on ConceptNet Numberbatch embeddings similarity.

[1] ConceptNet: https://conceptnet.io/
[2] ConceptNet Numberbatch: https://github.com/commonsense/conceptnet-numberbatch

Experiments & Results

	Method	BBC News	AG News	20NG	AFP News	YQA-v0	YQA-v1
Zero-shot	BL-GWA	26.1	26.7	53.5	60.0	51.8	36.2
	BL-EN	40.2	63.9	36.7	32.8	49.9	43.4
	Entail	71.0	64.0	45.8	61.8	52.0	49.3
	ZeSTE	**94.0**	**72.0**	**63.0**	**80.9***	**60.3**	**58.4**
	Supervised	96.4	95.5	88.5	-	72.6	58.4

Performance on 5 topic modeling datasets (accuracy score)

* *On the french version of AFP, the Accuracy is 78.2*

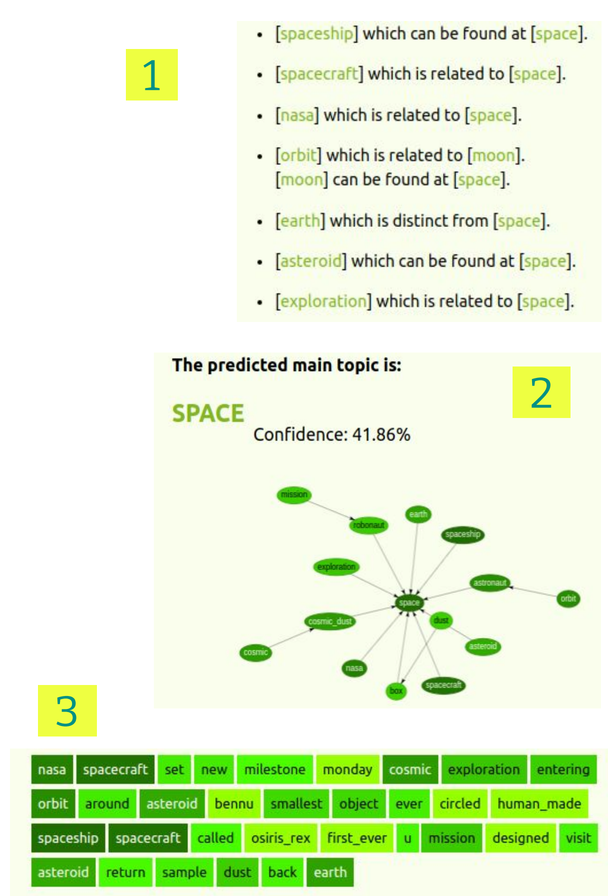
	BBC	AG	20NG	AFP
ZeSTE	80.6	71.0	61.6	73.8
ZeSTE + BERT	**94.3**	**84.2**	**70.1**	**83.0**

Accuracy of ZeSTE and a bootstrapped BERT classifier on the test set of four datasets

Explaining Predictions

Given the label neighborhood, we can generate an explanation as to why a document has been given a specific label. It can be generated in natural language (1) or shown as the subgraph of ConceptNet (2) that connects the label node and every word in the document that appears within its neighborhood, and hence counted towards its score.
Since this graph is usually quite big, we can generate a more manageable summary by picking up the closest *N* terms to the label in the graph embedding space, as they constitute the nodes contributing most to the score of the document.
Another method of explaining the predictions of the model is to highlight the words (or n-grams) that contributed to the classification score in the document (3).

Go to zeste.tools.eurecom.fr for a live demo!



Contact & Info

✉ ismail.harrando@eurecom.fr
github.com/D2KLab/ZeSTE
zeste.tools.eurecom.fr/

EURECOM Sophia Antipolis
MeMAD Methods for Managing Audiovisual Data
SILKNOW
ASRAEL

Check out our demo!