

## CONTEXT

We wanted to find a method to support annotation experience when the text is unknown and of difficult comprehension, unless the reader is an expert. We were able to extract topics using Correlation Explanation (CorEx) and to classify documents by their topic similarity using Word Mover's Distance (WMD).

## USED METHODS

- From unstructured text, be able to retrieve topics, emojis, and important contextual words in more than one language. For this we used Correlation Explanation (CorEx)
- Using Word Mover's Distance (WMD) to annotate similar sentences and assist the annotators.

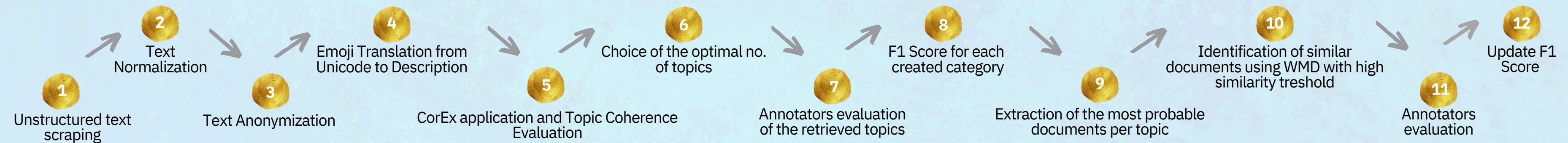
## DATASET

### Dataset #1:

- Language: English
- Content: 87,957 Tweets, plus associated emojis and hashtags
- Topic: Eating Disorders
- Statistics: Low Type-Token ratio, on AVG 11 words/tweet, 11.75% emojis usage

### Dataset #2:

- 3,064 documents
- Annotated in 10 categories recognized by CorEx, each representing an aspect of Eating Disorder (ED) : General ED, Weight, Body Representation, Body Description, Self Representation, Harm, Community, Diet, Sport, Consequences.



## GOAL

Creating a tool to assist annotators in their task, by providing a way to discover topics, important contextual multilingual words and to increment semi-automatically the number of annotated sentences, even in a complex context such as the Eating Disorders (ED) communities.

## LIMITATIONS

- people are self declaring that they suffer from an ED
- we handled only online content within 1 social media, Twitter
- some annotations fell under the General ED category: how to deal with topics minority?
- Clear annotations on this domain, by human or by an algorithm, are complex due to the co-presence of more topics in less than 20 words.

## COREX FINDINGS

- words related to Eating Disorders but mostly not understandable if read outside their context and without knowing the community (e.g. borboletana, OMAD, rexy, ABC, casuloana)
- foreign language words used in English communities, e.g. #waniliowemleko
- celebrities linked to ED fantasies (e.g. Lena Snow, Chloe Ting)
- emojis related to ED actions: [mouth], [nauseated face], [face with open mouth vomiting] for binge eating, [face upside down] for weight gain, [dizzy symbol] for excessive sport.

## WMD RESULTS

- we used WMD on the 30% of dataset no. 2 including all classes
- using a high similarity threshold (98%) we were able to automatically retrieve documents of the same topic, improving the overall F1-Score

	Weight	Body Representation	Self Representation	Harm	Consequences	Community	Sport	Diet	Body Description	General
F1-Score CorEx	0.0178	0.1291	0.0159	0.0275	0.0315	0.0676	0.0334	0.0564	0.1326	0.6412
F1-Score WMD	0.7686	0.7909	0.7544	0.51	0.9721	0.8319	0.7404	0.7756	0.6114	0.5955