



# tweets3

## "Concepts" Results

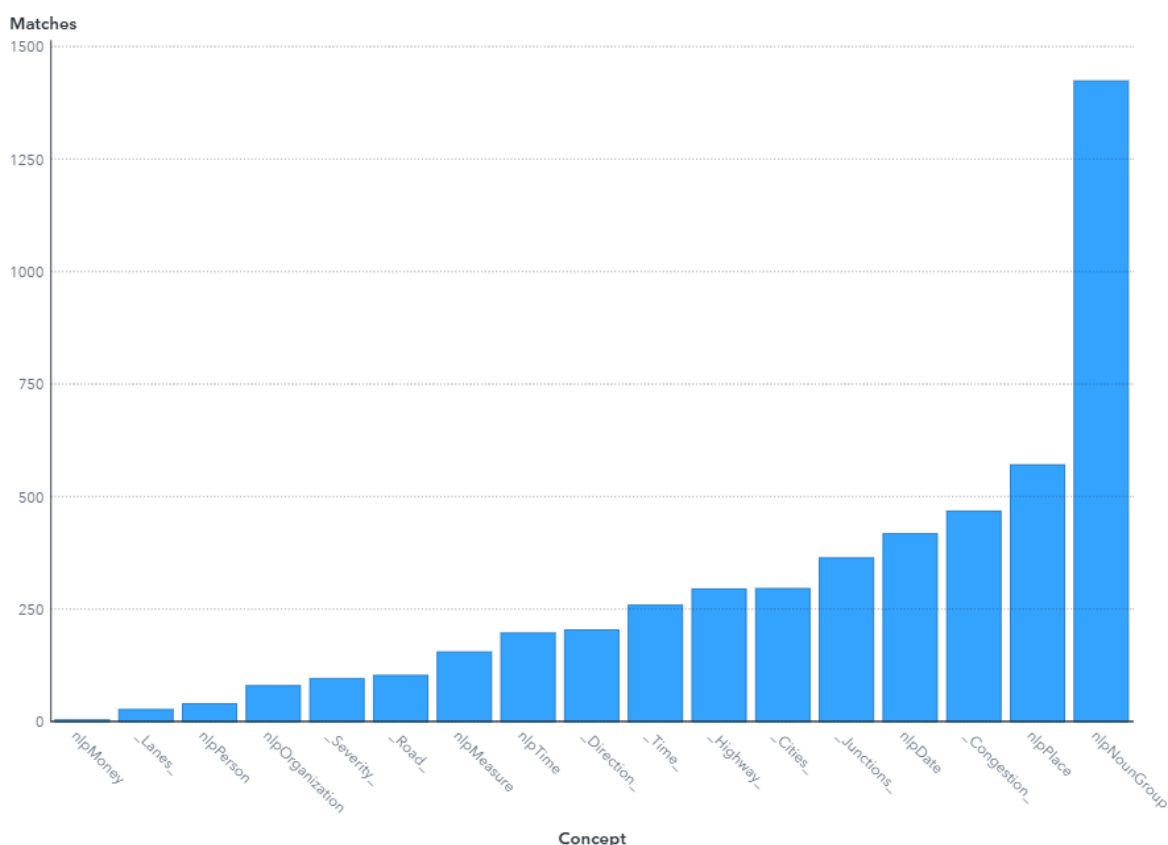
by: ta01468@surrey.ac.uk

---

# Contents

Number of Matches Per Concept	3
Number of Documents Per Concept	4
Average Number of Matches Per Document	5

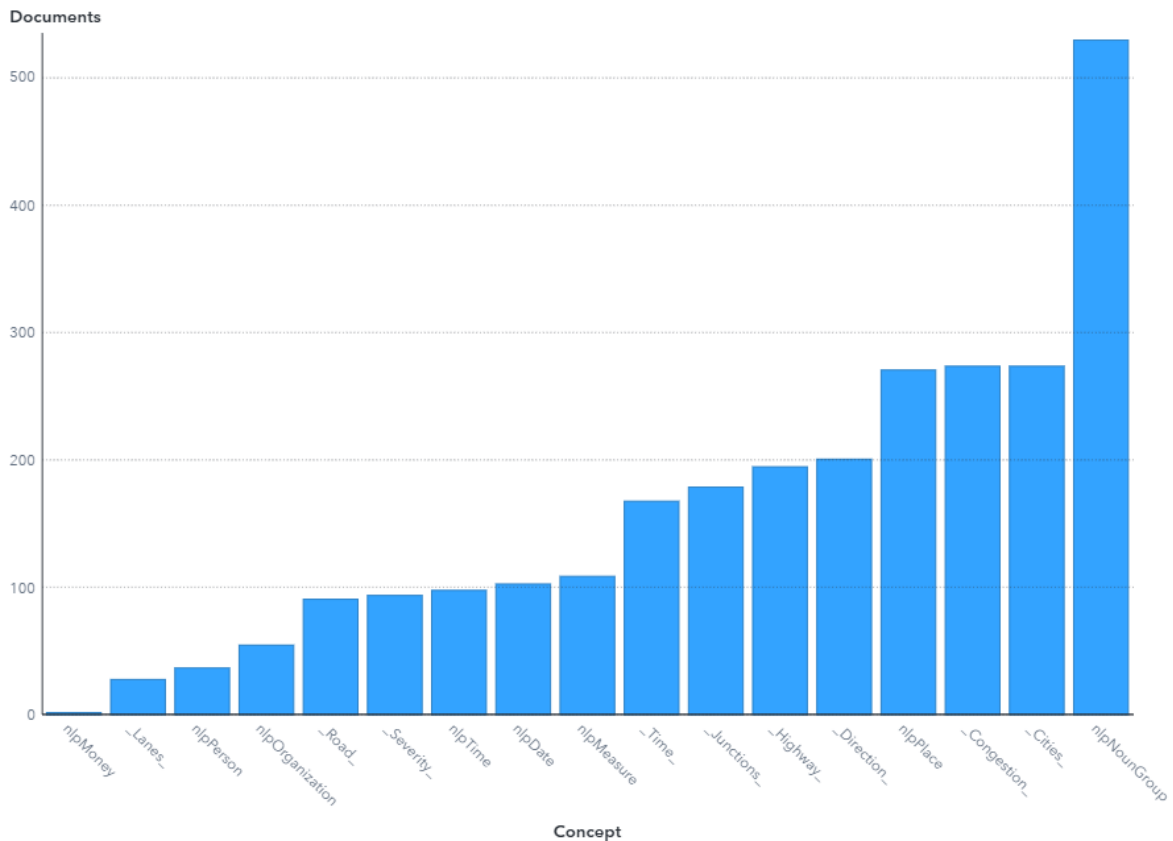
## Number of Matches Per Concept



The Number of Matches per Concept report depicts how useful each concept is for finding information in the data generally. The top matching concept in this data, nlpNounGroup, has 1,425 matches, while the least matched concept, nlpMoney, has 5 matches.

This information indicates how closely each concept, and the data in the documents are aligned. Many matches show that a concept is well-defined to extract information from the data set. Fewer matches for a concept indicate that either the data is not appropriate for the concept or that the concept needs further definition by adding or refining rules.

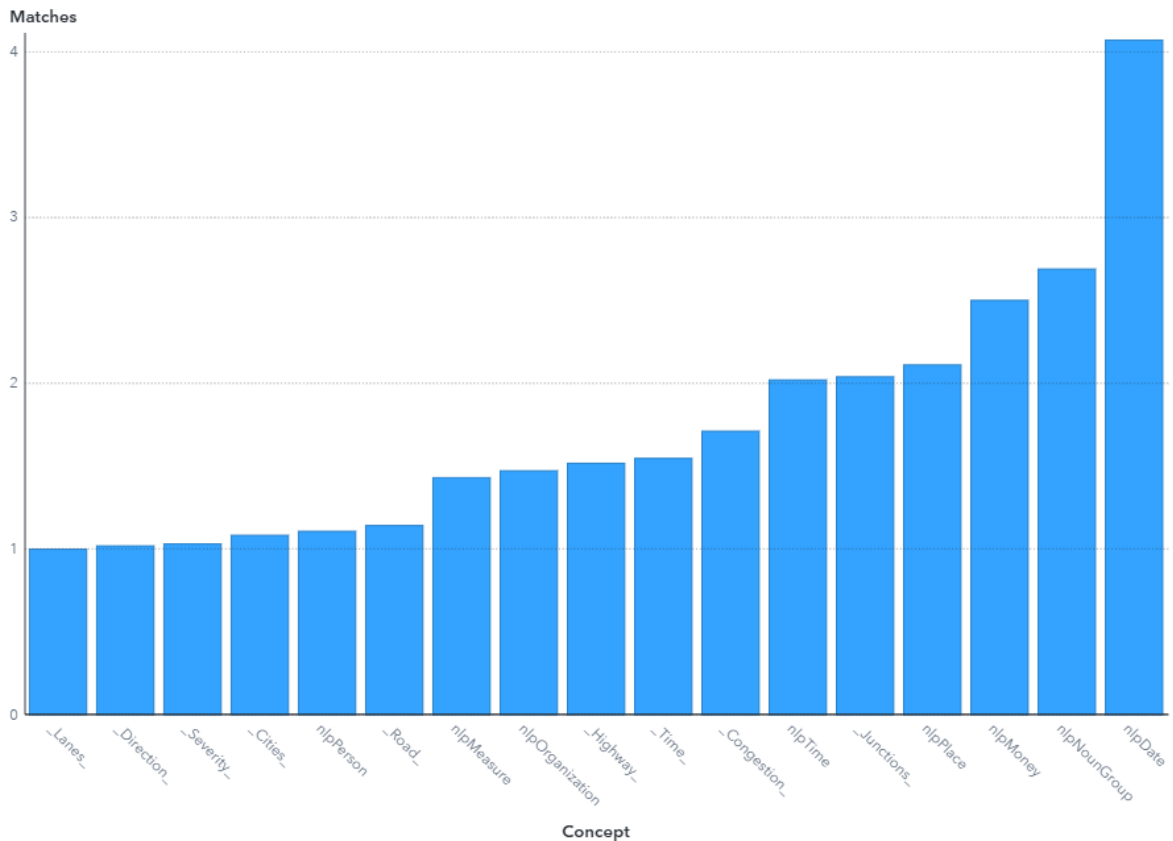
## Number of Documents Per Concept



The Number of Documents per Concept report depicts how well each concept covers the documents in the data set. In this data set, the two concepts with the greatest coverage are nlpNounGroup with matches spanning across 530 documents (88.63% of the total documents), and \_Congestion\_ with matches spanning across 274 documents (45.82% of total). The concepts with very light coverage (less than 5%) across documents in this data set are: \_Lanes\_, nlpMoney.

This information indicates how broad each concept is in terms of how many documents it matches. In a project that is expected to cover all the documents with each concept or a subset of concepts, this report can be used to gauge how close the model is to that goal and which concepts are more complete in their coverage.

## Average Number of Matches Per Document



The Average Number of Matches per Document report depicts the amount of information extraction performed by the concept inside each document where it matches. In this data set, for example, when nlpNounGroup matches in a document, it matches 2.69 times on average.

The calculation of the average does not include documents where the match count is zero for that concept. This report shows when a concept is getting the desired depth to help with prioritization during development of concepts within the model.