Introduction

A common strategy for text-based information retrieval is to use a ranking function to rank all texts according to search terms and select the top n. This report describes and discusses our results using different textual similarities for topics in the IR subtask to calculate how well topics match documents and returning a sorted list.

Methods

LM Jelinek Mercer Similarity algorithm:
Under the query-likelihood approach, language models for IR try to estimate for each document the probability that the query Q was generated by the underlying language model. If it is assumed that terms occur independently, then the probability becomes the product of the individual query terms given the document mode.

Abstract

- We choose a simple base model for the IR subtask and using a document-based storage method.
- Formulating a retrieval strategy with Elastic Search.
- This strategy uses embedded retrieval algorithms to retrieve topics and calculate text similarity.

Experiments

- Statistical analysing

<table>
<thead>
<tr>
<th>L2 label</th>
<th>Number of topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum</td>
<td>105.36</td>
</tr>
</tbody>
</table>

Table 1: The number of L2 labels in the training set.

<table>
<thead>
<tr>
<th>L2 label</th>
<th>training set</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum</td>
<td>141</td>
</tr>
</tbody>
</table>

Table 2: The number of topics with L2 labels.

Experiments

- Search process

Conclusion

In the final performance results, the effect presented by our team is moderate in the overall performance.

Reference

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.ISBN:978-1-4503-0000-0/18/06 [J]