From Passages Into Elements in XML Retrieval

Kelly Itakura
Charles L. A. Clarke

Reality

- Trotman and Geva (2006):
  - In INEX adhoc track, returned results are XML elements.
  - We confuse XML retrieval with XML-element retrieval.

What Users Want

- Human assessors perform relevance judgments based on passages
- What we really need to return is passages or elements based on passages.

Turning Reality into Ideal

- To transition into passage-based XML retrieval, we need ways to convert passages or elements.

  - TG+: passages \(\rightarrow\) elements
  - TG-: elements \(\rightarrow\) passages
Our Implementation of TG+

- Compute the scores of all possible passages.
- Score:
  \[ \text{score}(P) = \sum_{i \in Q} W_i \frac{f_{P,i}(k_i + 1)}{f_{P,i} + k_i(1 - b + b \frac{|P|}{\text{avgdl}})} \]
  \[ W_i = \log \frac{\text{total documents containing } i}{\text{documents containing } i} \]
- Assign the scores of elements as follows.

Other Algorithms for Comparison

- Element Retrieval
  - score each elements, get rid of nesting, and get the top 1500 results.
- Passage Retrieval
  - get the positions of all query term occurrences, score all possible passages defined by these positions, get rid of nesting, return the top 1500 results.

Experimental Setup

- Corpus: INEX2005 IEEE collection
  - 16,819 files from IEEE journals from 95-04.
- Queries: 39 topics from INEX2005 adhoc focused task.
- Evaluation: nxCG metric in generalized quantization.
- Others: Wumpus search engine to retrieve positions of query terms.
Experimental Results #1

- Measures: Percent Excess (PE) and Percent Lack (PL) of results returned by Passage Retrieval.

PE: 

PL: 

Experimental Results #2
Experimental Results #2

Conclusions

- TG+ performs worse than element retrieval because it returns a lot of extra terms.
- TG- may miss a lot of terms but because Focused task prefers specificity over exhaustivity, the performance will be better.
- Ultimately, the best elements to return are those that best fit the best passages,
- Or, return multiple elements that cover the passages just enough.

Happily Ever After