Mechanisms to Discover the Real News on the Internet

Yuta Nemoto and Vitaly Klyuev
Software Engineering Lab
University of Aizu, Japan
Introduction

- Growing impact of the online news media
- Generation Z accelerates this trend with Internet media on their smartphones
- Available quick access to information sources
- Risk of false information spreading on the Internet
Hot Topics

- **Shared Information without evidence** on Facebook…
  
  “Hold your breath for more than 10 seconds without coughing, discomfort, stiffness or tightness, etc., it basically indicates no infection.”

- **Shared misinformation** on Twitter…
  
  “Soon we ran out of toilet paper because novel coronavirus affects its supply.”

“Share” or “Like” accelerate the spread and make the effect of wrong information bigger.
Motivation

- Online media users tend to use one application such as Google or Facebook
- Centralized platforms may hide key documents and affect user’s decision making

Tools to support user’s careful thinking are strongly demanded
Metasearch Engine

Internet

DuckDuckGo

Yahoo!

Google

Bing

Metasearch Engine

Query

Response
Goal of the Research

Develop a tool to help users who want to find real news related to their information needs

Provide different pieces of information to create the real vision of users

Metasearch System

- Increases plurality in the search results
- Reduces biased searches
Approach

Filtering schema for the search results

- Classify the result items into
  - Encyclopedias
  - Famous news agencies
  - Online newspapers
  - Portals
  - Blogs

- Max. 9 documents: 1 Encyclopedia, 2 for each of 4 other categories, must be presented
- Latest document must be prioritized for presentation
- ...

Implementation

Application process consists of 3 layers

- Metasearch
- Result Item Selection
- Presentation
Implementation

Metasearch

Data Collection

Parse the obtained web page and specify the result items

BeautifulSoup Library

Send query to Yahoo, Bing, Yandex...

Classification of items by the domain of the URL $+\alpha$

Pass the set of result items to the Result Item Selection process
Implementation

for all categories except encyclopedia:
  if the category has multiple items which have the same domain:
    for duplicated source of items:
      latest ← pick the latest item
      remove all others
      put latest back to category
  result list append 1 item randomly picked from encyclopedia category

for all categories except encyclopedia:
  if the number of items in the category is 1:
    result list append the item in the category
  else:
    result list append 2 items randomly picked from the category

return result list
Implementation

- Web application with responsive web design
- Smartphone-native application
Current Results

Experiment

➢ Keyword: “iran nuclear deal”
➢ Total 41 results collected
  • 6 from Bing
  • 15 from Yandex
  • 10 from Yahoo!
  • 10 from DuckDuckGo
➢ Classified as...
  • 4 Encyclopedia
  • 18 Famous News Agencies
  • 7 Online News Papers
  • 8 Portal Websites
  • 2 Blogs

Result for Mobile Device

Query: iran nuclear deal
<Result for Mobile Devices>
  - https://en.wikipedia.org/wiki/Iran_nuclear_deal_fr...
  - retrieved from Bing
[FamousNews] The Iran nuclear deal explained – RT World ...
  - https://www.rt.com/news/425589-iran-nuclear-deal-e...
  - retrieved from Yandex
[NewsPaper] Trump Abandons Iran Nuclear Deal He Long Sco...
  - https://www.nytimes.com/2018/05/08/world/middleeast...
  - retrieved from Yahoo
[Portal] Iran nuclear deal – Conservapedia
  - https://www.conservapedia.com/Iran_nuclear_deal
  - retrieved from Yandex
[Blog] The Historic Deal that Will Prevent Iran from Acq...
  - https://obamawhitehouse.archives.gov/issues/foreign...
  - retrieved from Yahoo
...
Discussion

Difficulties and the Next Work:

- Implementation for search engines: Google and Baidu
- How to deal with search platform having the same search system? Ex. Yahoo and Bing
- Evaluation of the efficiency of the tool
Discussion

Two Platforms Having the Same Engine

AND

Yahoo! and Bing

Two quick solutions:

- Remove one of Yahoo! or Bing from our search engine list
- Reduce the weighting of results given by Yahoo! or Bing by half
Discussion

How to evaluate the quality of the search outcomes?

General search engine:
TREC-Style Average Precision (TSAP) – score by analyzing the relevance of top N result items

How about metasearch engine?
A way to measure the neutrality or fairness is needed
Still it depends on experts’ judgment in many case
Conclusion

Done:

• The information retrieval from multiple search engines
• The organized search results with predefined filtering schema

Challenges:

• Obtaining search results from Google, Baidu, etc.
• Classification of Portal/Blog based on the basic page contents
• Evaluation of the performance of this tool
Thank you for your attention