



Topics in Internet Content Exploration - Searchub

Course: IERG4999

Student Name: LO Siu Kit

Supervisor: Prof. LEE, Yiu-bun Jack

Department of Information Engineering, The Chinese University of Hong Kong

Introduction

Searching on mobile devices is extremely common these days, but it is not efficient and intuitive enough.

- Come up with a search query
- Tamper with current search query
- Search on multiple platforms
- Search by non-textual input

Related Works: Results Post-processing

Similarities -> Clustering -> Pattern -> Re-rank

Objectives

- Create a way to search on mobile devices more efficiently & intuitively. (UI/UX)
- Assist in search query generation & refinement.
- Supports for non-textual input.
- Sort & organize search results from different platforms.

Methodology & Results

Implementation: Cross-platform Mobile Application - Searchub (Built with Flutter)

Efficient & Intuitive UI: Minimize actions required

- Load results directly
- ONE swipe to switch results / types
 - (General, Videos, SNS)

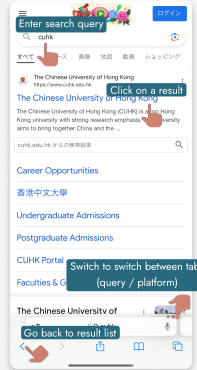


Figure 1: Conventional Search

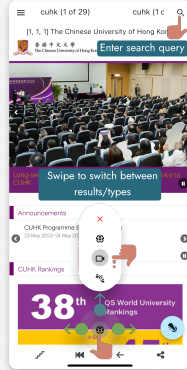


Figure 2: UI of Our App

Drill-down Search: New search based on current result

- Construct Search Query
 - Manually (A)
 - Directly appends to current search query
- Auto with draggable FAB (B)
 - Use position information of FAB from Flutter to extract contents of HTML element using JavaScript
 - Go through query refinements if necessary

Auto Query Refinements: Reduce hassle of tampering with query

- Targeted HTML element is not extractable, e.g., image, manual selection forbidden
 - Screenshot visible portion of current web page
 - Perform text detection (Google Vision API)
 - Concatenate all detected text as new search query
- If new query or content of targeted HTML element is too long
 - Keywords Extraction (Google Language API)

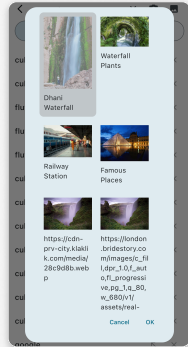


Image Search: Generate query by images

1. Take photo / Upload image from device
2. Bing Image API returns images with labels; Google Vision API returns images.
3. Image with label -> label = search query; Image -> image search -> loop until text search

Search Histories: Traverse between searches with ease

- Save user's search histories in current session
- Update last viewed search type & result index
- Can go back to any previous search with a tap

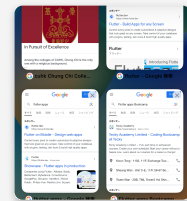


Figure 6: Multiple Tabs Opened

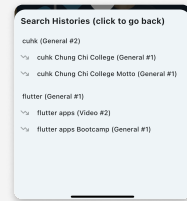


Figure 7: Search Histories

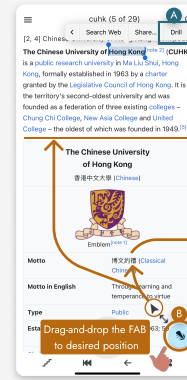


Figure 3: Drill-down Search

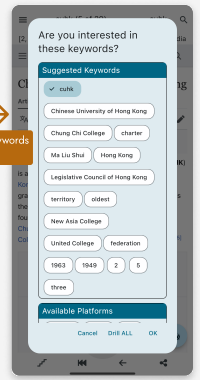


Figure 4: Keywords Extraction

Gathering Results: Minimize the chance of having 0 results

- Customizable: Google (API), Bing (API), DuckDuckGo (Web Scraping)

Merging & Re-ranking: Organize & sort results from different platforms

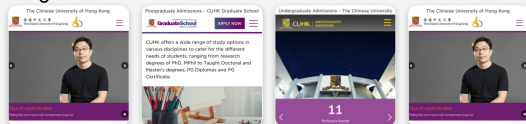
- **Frequency:** Reduce duplicated results
 - Merge results with identical URL
 - Sort descending by frequency
 - Ignored ranking (1, 1 =/= 1)
- **Further Merge:** Aggressively merge potentially identical results
 - Extension of **Original Rank**
 - Compare loosely (sub-string) by URL
 - www.cuhk.edu.hk VS www.cuhk.edu.hk/ VS www.cuhk.edu.hk/index.html
 - Merge after verify by snippet / title
 - Weighted = Base Score System, Not Weighted = Average Rank (2, 3 -> 2.5)
- **Original Rank:** Utilize ranking of results
 - Base Score = $\frac{\text{No. of total results}}{\text{No. of platforms}}$
 - Score of each result = $\frac{\text{Base Score}}{\text{Rank in original platform}}$
 - Sort descending by score of each result

Example: Searching for "cuhk"

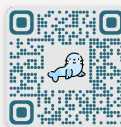
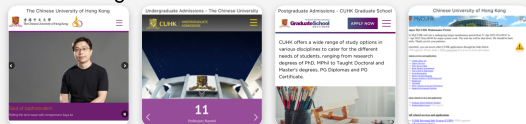
Frequency



Original Rank



Further Merge



Filter Results by Uniqueness: Avoid missing out results

- A search result appears on more one platforms = NOT unique
- Duplicated (not unique) results might imply that they are deemed related by more platforms
- Unique results might imply that they are not flavored by some platforms, doesn't imply not related
- Mark uniqueness of results during merging & re-ranking

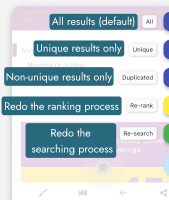


Figure 8: UI for filtering uniqueness



Figure 9: A unique result

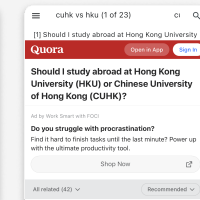


Figure 10: Another unique result

Conclusion

- Successfully built a cross-platform mobile application that is intuitive to use and requires a low learning curve.
- Combined existing APIs to create a more useful and powerful solution to the existing problems, such as keywords extraction.
- Achieved the goal of providing a more convenient, intuitive, and efficient searching experience that returns optimal results for everyone that searches on the Internet.

Future Directions

Enhanced Results Post-processing

- Fetch webpage content in background to accurately identify similar results
- Search for Images / Videos**
 - Dedicated UI for images / videos instead of website

References

- [1]. A. Sharma, N. Aggarwal, N. Duhan and R. Gupta, "Web search result optimization by mining the search engine query logs," 10 February 2011. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/5706716/authors#authors>. [Accessed 28 November 2022].