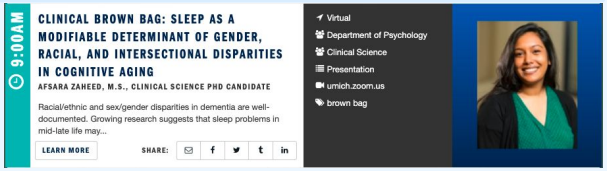
University of Michigan Events Search Engine

Tianying Zhang, Jiayang Ding, Cindy Yu, Haijian Wu, Hezhang Fan
{danniez, vicdjy, xinyuy, willwuhj, hzfan}@umich.edu

Introduction

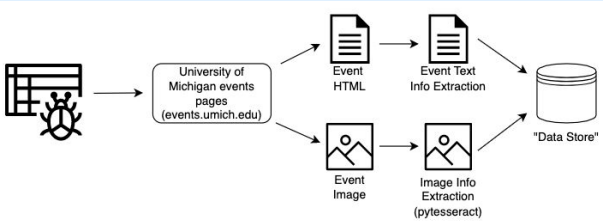


Event searching is a difficult task for students at the University of Michigan. Current events searching platforms offered by the University of Michigan often do not generate satisfactory query results e.g. only offer sorting by category but no searching by keyword.

Goal

Facilitate students' campus involvement through designing a UMich Ann Arbor campus-wide "activity searcher", which will capture all kinds of activities held by student organizations, major programs, or university departments.

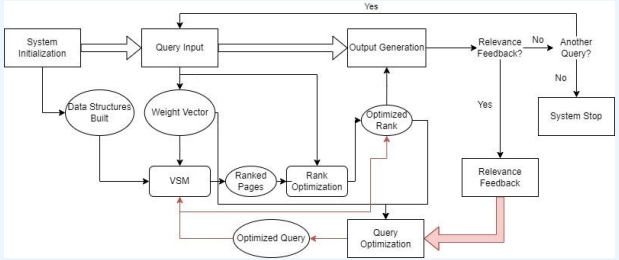
Data Collection and Data Samples



We used events.umich.edu and events.engin.umich.edu as seed URLS for our crawler in order to extract events posted by UM Student Life and the College of Engineering. For each event, we extracted event information[1], such as title, date and event description, from the HTML of the webpages. For each event, we also extracted image information. By using the python library, pytesseract[2], we extracted any event information that appears on the images associated with each event. Finally, all the data we extract is saved to a "data store" (in excel worksheet format) by the python library pandas.

Method

The search engine first semantically expands query based on word embeddings[3], then uses VSM with tf-idf to calculate cosine similarity scores for a query with all events stored in the database. The content of returned events are summarized using BART pretrained model[4], then the engine reranks the retrieved events based on the summarizations using sentence embedding[3]. Finally, the engine outputs events names, links and summaries based on the optimized ranking. Users can provide relevance feedback to optimize the query and gain an updated result improved by Standard Rocchio Method.



Evaluation

Main Metrics: Average Precision & NDCG

- Relevance depends on the user's subjectivity and personal preference for different pages. To avoid bias, the feedback from multiple users is used in the evaluation process. Users provide feedback In the format of (2,1,0), corresponding to (Very Relevant, Relevant, Irrelevant).

- The system considers average precision and NDCG. After users provide relevance feedback, the AP of the engine increases as expected; but NDCG would decrease because optimization of the query does not include rank optimization.

	AP	NDCG
Original Query	0.86995	0.96647
After Relevance Feedback	0.92398	0.89635

Results

Sample (Does not contain all outputs)
Input: Research Summer 2022
Output: *eventID*:464
Name: Become a Summer Research Mentor
Place: Off Campus Location Time:2022-04-19 5:00
Link: https://events.umich.edu/event/92672
Event Description:
UROP Research Mentors are faculty and post-doc researchers who provide undergraduate student researchers an opportunity to engage in research activities. This early exposure to research fosters a valuable academic experience for students. Through this collaboration, students gain research skills.
Relevance Feedback: 464

Conclusion

Achievements:
- Designed crawler to extract different event info from HTML pages and images included with each event.
- Designed a model that utilizes semantic expansion and employs VSM and Standard Rocchio Method. The content of events is summarized for users.
- Evaluated our model and provided results for various user queries

Future Work:
- Construct a front-end platform in order to facilitate user experience, most likely a website, that users can interact with in combination with the activity searcher.
- Include off-campus activities, such as Ann Arbor local events, for those who would like to explore the town.

References

[1]Hogenboom et al, "An overview of event extraction from text," CEUR Workshop Proceedings, 2011
[2]R.Smith. "An overview of the tesseract ocr engine", ICDAR 2007
[3]Mikolov et al, "Distributed Representations of Words and Phrases and their Compositionality", NeurIPS 2013
[4]Lewis et al, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension", ACL 2020