

Technical Disclosure Commons

Defensive Publications Series

December 2020

Automated Extraction of Pivot Topics for Sideways Expansion of Search Scope

Alankar Jain

William McMeans

William M Leszczuk

Sitaram Iyer

Seher Aylin Altiok

See next page for additional authors

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Jain, Alankar; McMeans, William; Leszczuk, William M; Iyer, Sitaram; Altiok, Seher Aylin; Sharifi, Mehrbod; Miglino, Jerome; and Barlow, James K., "Automated Extraction of Pivot Topics for Sideways Expansion of Search Scope", Technical Disclosure Commons, (December 04, 2020)

https://www.tdcommons.org/dpubs_series/3851



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Inventor(s)

Alankar Jain, William McMeans, William M Leszczuk, Sitaram Iyer, Seher Aylin Altiok, Mehrbod Sharifi, Jerome Miglino, and James K. Barlow

Automated Extraction of Pivot Topics for Sideways Expansion of Search Scope

ABSTRACT

Users benefit from mechanisms that can help them refine their queries to facilitate searching for information connected to their underlying intent. Apart from refinements to narrow the scope of a query, users can benefit from suggestions that can help them pivot their information seeking by expanding their search sideways to related topics. This disclosure describes computational techniques for automated determination of suitable topics and/or queries for helping users expand the scope of their information search by pivoting to topics related to their query. The techniques involve selecting a meta-query, performing query expansion, identifying, aggregating, and deduplicating related entities. The identified entities are clustered and ranked to enable selection of particular entities that can be shown to users as pivot topics.

KEYWORDS

- Search engine
- Query refinement
- Query expansion
- Query suggestion
- Pivot topic
- Search scope
- Entity extraction
- Entity clustering
- Entity filtering
- User journey

BACKGROUND

Search engine users express their information needs by entering a query. However, queries are, at best, a noisy expression of user intent. Therefore, users benefit from mechanisms that can help them refine their queries to facilitate searching for information connected to the underlying intent. For instance, a user who enters a higher level query such as “college education” can benefit from query refinements that add relevant specificity, such as “college

education in California.” Current search engines support query refinement via a number of mechanisms, such as query suggestions, related searches, etc., that help users narrow the search space by increasing specificity. Such features are primarily based on the searches of other users who issued the same or similar queries.

Apart from refinements to narrow the scope of a query, users can benefit from suggestions that can help them pivot their information seeking by expanding their search sideways to related topics. For instance, a user who searches for “college education” may also be interested in queries on related topics such as student loans, college entrance exam preparation, etc. While such topics may not reflect an immediate need, they might still be relevant to the user’s underlying intent in issuing the initial query. Search engines support such sideways expansion of the search scope by providing suggested related topics and/or queries that other users search for when making the same or similar queries as the user’s original query. Such suggestions are derived based on available entity similarities. As a result, the suggestions are not always available, and, when available, their relevance and utility can be limited.

DESCRIPTION

This disclosure describes computational techniques for automated determination of suitable topics and/or queries for helping users expand the scope of their information search by pivoting to topics related to their search query.

With user permission, the determination of pivot topics and/or queries is carried out via the following steps:

1. **Select meta-query:** A given user query can be associated with one or more meta-queries.

For example, if the query “college education” reflects an underlying intent to seek college education, it can be associated with meta-queries such as “financing college education,”

“selecting a field of study,” etc. Such meta-queries can be manually curated. One or more of these meta-queries is selected for further processing.

2. **Expand meta-query:** The selected meta-query or meta-queries can be expanded to similar queries using suitable approaches, such as nearest-neighbor queries in a query embedding space. Since the determination of pivots is dependent on the respective search results for such queries, the set of expanded queries is large enough to ensure the robustness of the pivots.
3. **Retrieve search result pages:** A search engine is used to retrieve the top N pages corresponding to each of queries in the set of expanded queries constructed in the previous step..
4. **Extract relevant entities:** The pages retrieved in the previous step are parsed using suitable entity extraction techniques to determine salient entities within the page contents. For example, entities within pages related to financing college education can include terms such as “student loans,” “financial aid,” “scholarships,” etc.
5. **Aggregate and filter extracted entities:** The extracted entities are aggregated across all documents and subsequently filtered by frequency counts. For instance, a joint query-document embedding space can be used to compute similarities between entities and documents and the similarity scores can be summed across different documents. In addition, specific entities, such as persons, locations, etc., can be blocklisted to avoid irrelevant or unwanted pivots.
6. **Cluster aggregated entities:** Similar entities within the set of aggregated and filtered entities are clustered using suitable techniques such as Hierarchical Agglomerative Clustering applied to entities represented as queries/phrases. The clustering process can help identify and merge semantic duplicates within the set of aggregated and filtered entities generated in

the previous step. For instance, clustering can group together similar queries such as “financial aid” and “financial assistance.”

7. **Rank entity clusters:** The clusters produced in the previous step can be ranked based on relevance and/or diversity and/or frequency. The ranking is then used to select the one or more entities that are shown to the user as pivot topics.

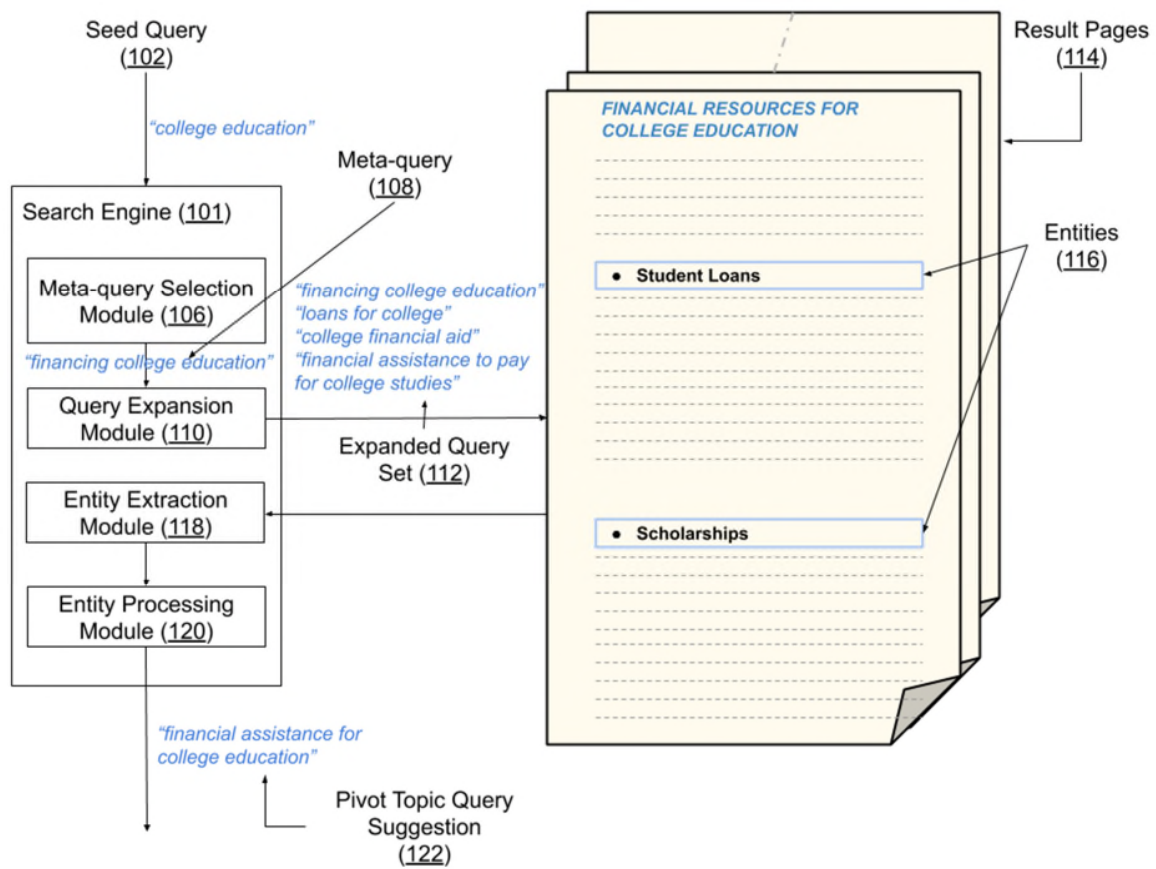


Fig. 1: Generating pivot topics for sideways expansion of search scope for a query

Fig. 1 shows an operational implementation of the techniques described in this disclosure. A user invokes a search engine (101) for a seed query (102, “college education”) regarding the topic of college education. With the user’s permission, a meta-query selection module (106) is used to select a meta-query (108, “financing college education”) associated with the seed query.

The selected meta-query is then used to create an expanded set of queries (112) via a query expansion module (110). The pages of the results (114) for the expanded set of queries are parsed to extract relevant entities (116) within the content of the pages via an entity extraction module (118).

The extracted entities are further processed by an entity processing module (120). As described above, the processing can include aggregation, filtering, clustering, and ranking. Based on the results of the ranking, one or more pivot topics are selected as a suggestion. For instance, Fig. 1 shows the pivot topic (122) of financial assistance connected to the original query regarding college education.

With user permission, the described techniques can be implemented within any application or service that provides search functionality for content such as documents, files, webpages, posts, etc. Various operational parameters, such as the number of search result pages (N), the number of meta-queries, the number of queries in the expanded set query set, the number of pivots, etc., can be set by the developers and/or specified by the users and/or determined dynamically at runtime based on the topic of the seed query and/or the content of the initial set of pages retrieved based on the seed query. The outcomes of the process described above can be presented using any suitable user interface (UI) techniques, such as query refinement suggestions, information snippets, search links, etc.

Implementation of the described techniques can make it faster and easier for users to expand their search scope to include additional topics connected to the topic of their original query, thus enhancing the search engine user experience (UX).

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may

enable collection of user information (e.g., information about a user's queries, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

Users benefit from mechanisms that can help them refine their queries to facilitate searching for information connected to their underlying intent. Apart from refinements to narrow the scope of a query, users can benefit from suggestions that can help them pivot their information seeking by expanding their search sideways to related topics. This disclosure describes computational techniques for automated determination of suitable topics and/or queries for helping users expand the scope of their information search by pivoting to topics related to their query. The techniques involve selecting a meta-query, performing query expansion, identifying, aggregating, and deduplicating related entities. The identified entities are clustered and ranked to enable selection of particular entities that can be shown to users as pivot topics.