Dataset Creation Framework for Personalized Type-based Facet Ranking Tasks Evaluation

Esraa Ali[1] ✉[0000-0003-1600-3161], Annalina Caputo[2] ✉[0000-0002-7144-8545], Séamus Lawless[1][0000-0001-6302-258X], and Owen Conlan[1][0000-0002-9054-9747]

[1] ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin
esraa.ali,seamus.lawless,owen.conlan@adaptcentre.ie
[2] ADAPT Centre, School of Computing, Dublin City University
annalina.caputo@adaptcentre.ie

Abstract. Faceted Search Systems (FSS) have gained prominence in many existing vertical search systems. They provide facets to assist users in allocating their desired search target quickly. In this paper, we present a framework to generate datasets appropriate for simulation-based evaluation of these systems. We focus on the task of personalized type-based facet ranking. Type-based facets (t-facets) represent the categories of the resources being searched in the FSS. They are usually organized in a large multilevel taxonomy. Personalized t-facet ranking methods aim at identifying and ranking the parts of the taxonomy which reflects query relevance as well as user interests. While evaluation protocols have been developed for facet ranking, the problem of personalising the facet rank based on user profiles has lagged behind due to the lack of appropriate datasets. To fill this gap, this paper introduces a framework to reuse and customise existing real-life data collections. The framework outlines the eligibility criteria and the data structure requirements needed for this task. It also details the process to transform the data into a ground-truth dataset. We apply this framework to two existing data collections in the domain of Point-of-Interest (POI) suggestion. The generated datasets are analysed with respect to the taxonomy richness (variety of types) and user profile diversity and length. In order to experiment with the generated datasets, we combine this framework with a widely adopted simulated user-facet interaction model to evaluate a number of existing personalized t-facet ranking baselines.

Keywords: Type-based Facets · Faceted Search · Personalization · Dataset Collection · Evaluation Framework · Simulated Users.

1 Introduction

In Faceted Search Systems (FSS), facets associated with the information objects being searched are used to decompose the information space into compounds of subjects [9]. They allow users to filter and narrow down the search space quickly. However, as the size of the collection increases, so does the number of facets, making it impractical to display them all at once. To tackle this problem, FSS usually employ ranking methods to find and promote relevant facets. Personalized facet ranking approaches exploit current user interactions as well as historical feedback to identify and rank facets of interest to the user. This paper looks at this problem by focusing on the specific approaches to type-base facet (t-facet) ranking, in which facets are further organised in a hierarchy for better readability. T-facets are derived from structured data organised in hierarchies, such as ontologies or taxonomies, and are usually extracted from `isA` or `type` attributes associated with the information objects.

Although the current literature presents a wealth of research in FSS, this area lacks standard datasets with relevance judgments for the specific problem of *personalised* facet ranking. This problem is even more relevant for personalized t-facet ranking tasks. Personalized FSS vary on the experimental setup they use to evaluate their ranking methods. Furthermore, none of the existing setups involves a rich hierarchical type-based taxonomy nor deals with the hierarchical nature of t-facets. A unified systematic methodology to build and evaluate collections suitable for this task is needed.

This research solves this problem by introducing a framework that customizes existing data collections to make them suitable for the assessment of such methods. It is dedicated to evaluate personalized t-facet ranking approaches where the past user's selections are used in the ranking process. The ranking methods focuses on type-based facets to leverage both their categorical and hierarchical nature. We study how datasets for personalised t-facet ranking should be selected and customized to fit the purpose of this task, and which simulation methods and IR metrics should be adopted to evaluate such FSS.

The proposed framework is concerned with search tasks that aim at minimizing user effort in precision-oriented FSS. The assumption is that the search task is fulfilled as soon as the user finds their intended search target. T-facet ranking approaches are evaluated by using a simulation-based methodology proposed by Koren et al. [8], which is well established and widely used in faceted search literature. The evaluation assumes that the searcher can identify the intended target and their associated facets as soon as they see it.

We contribute to this research area by proposing a dataset creation framework to evaluate personalized t-facet ranking methods using simulated interaction models. The framework outlines the eligibility criteria for existing collections, as well as the required data structure of the underlying documents (or information objects) and associated taxonomy of types. The framework also details the pre-processing and transformation steps required to implement this customization. Using this framework, we introduce two datasets created for this evaluation task. Finally, we show the feasibility of the proposed framework by analysing the generated datasets and using them to evaluate several personalized t-facet ranking baselines.

2 Related Research

Personalized t-facet ranking is an unexplored area in literature, hence we give a brief overview of existing personalized facet ranking evaluations. Chantamunee et al. [4] suggested a personalized facet ranking based on Collaborative Filtering (CF). They used user ratings and Matrix Factorization via SVM to learn facet ranks. The MoviesLens dataset was used in their evaluation. The average rating given by the user to the facet is used as ground truth, they reported RMSE values to measure the effectiveness of the ranking method. This experimental setup might be useful in prediction tasks, but it does not assess how the final facet list will assist the user in reaching their target.

Koren et al. [8] argues that task-based studies, while undoubtedly useful, are very limited, because they are expensive to conduct, hard to repeat, and the number of users is usually limited, which makes their results inconclusive and not reproducible, especially in personalized search systems. They instead suggest an approach that simulates the clicking behavior of users in the FSS. They attempt to measure the amount of effort required by users to satisfy their search needs. A User information need is considered fulfilled when the target resource is located by using the ranked facets. Based on this idea, the proposed evaluation counts users actions taken towards finding this intended target. The goal of the evaluation is to minimize the effort needed by the users to fulfill their search needs.

This is the most adopted simulation model for precision-oriented FSS present in literature [11, 1, 12, 10] and others have followed. Adaptive Twitter search system [1] adopted this approach for finding tweets. User profiles were built from users' previous tweets, and the evaluation assessed whether or not the personalized ranking approach could predict the latest retweets. Also non-personalized FSS adopted the same simulated user evaluation method. Vandic et al. [10] suggested an approach to rank facets based on query relevance and information structure features in the e-commerce domain. Different models for clicking behavior were used and metrics measuring user effort to scan facets and their values were computed.

In general, existing literature seems to follow two different paths to obtain evaluation collections in faceted search. The first is to utilize existing ad-hoc IR datasets with relevant judgments provided on the resource level. In this case, it is assumed that relevance travels from the resource (document) to the facets to which they belong. This is the path followed by the INEX 2011 Data Centric Track [12]. The task consisted of two sub-tasks: an ad-hoc search task and a faceted search one. In the faceted search track, the evaluation metrics measured the effort needed to reach the first relevant result. The evaluation was based on the user simulation interaction model proposed by Koren et al. [8]. We follow this path in transforming the TREC-CS 2016 dataset in section 4.1. Our framework customizes the dataset to fit to the type-based facet ranking task, existing personalized relevance judgments were useful to evaluate the facet ranking approach based on the same INEX 2011 Data Centric track assumptions.

The second path is to transform existing real-life datasets to fit facet ranking evaluation. This path was followed by Koren et al. [8] on the MovieLens evaluation. In order to generate query requests, they used the most recent user's ratings as search targets. The simulation approach was used to measure the user effort to reach those targets. The MovieLens dataset is not suitable for our task as movies genre (types) are limited and do not have a multilevel hierarchical taxonomy.

Following the steps in this last approach, our framework customizes real-life collections into a TREC-like format and then applies the INEX evaluation method. The customization framework formalizes the dataset generation process and extends it to suit the case of type-based facets. As an example of the second path, we adopt The Yelp dataset described in section 4.2. In both paths, the same aspects need to exist in the dataset in order to be a good fit for this research task, these are discussed in section (Sec. 3.2).

For both use cases, we adopt the metrics proposed in the INEX 2011 task: 1) The number of actions (#Actions), which counts how many clicks the user has to perform on the ranked facets in order to reach the first relevant document in the top results; 2) The faceted scan (F-Scan), which measures the user's effort to scan facets and documents until the user reach the same first relevant document in the top results.

3 Dataset Customization Framework

Before formalising the desiderata of the data and the processing procedure to generate a ground-truth dataset, we define the personalised t-facet ranking problem in the context of a FSS. We assume that when the user submits a search query, the underlying search engine starts by retrieving and ranking the top relevant documents. [3,4] Then, the FSS collects the t-facets associated with all retrieved documents. These collected t-facets are reckoned to be relevant and represent the input for the t-facet ranking approach. This research assumes that the ranking of t-facets occurs during the initial population of the result page, and that the t-facets are not reshuffled during the navigation process unless the user submits a new query, which will re-initiate the facet ranking step.

3.1 Eligibility Criteria

The applicability of the proposed framework is subjected to a number of criteria that pertain the domain, search task and type of data, listed as follows:

1. The underlying data collection is structured. It contains objects and each object has properties and types, which can be used as type-based facets.

[3] In the scope of this work, the term 'documents' is used to refer to the information objects being searched. According to the FSS domain, documents can be places, web pages, products, books or images, etc.

[4] How the document ranking is performed is outside scope of this research

2. The searchable objects belong to a rich taxonomy of categories, from which stems the need for ranking. This is a crucial requirement, as the categories act as type-based facets.
3. Data contain plenty of user feedback, ratings and reviews, which are also useful for personalization.
4. The data is accessible and available online, as some datasets needs further data collection or have no pre-defined taxonomy of types, which makes them unsuitable for this task.
5. The dataset's domain should be suitable for faceted search, e.g. product shopping, digital libraries, venue suggestion, social event search, etc.
6. There is room for personlization in both facet ranking and search result ranking.

3.2 Designated Data Structure

The intended dataset contains a collection of documents, or resources, to be searched. Each document has:

1. Textual description, for example a web page content or a description written by the document's owner.
2. A set of reviews and ratings given to this document by the users. The reviews reflect users' experiences or opinions about this document.
3. A set of categories assigned to the document by the document owners or system admins. The categories belong to a large hierarchical taxonomy of categories and they are treated as type-based facets.

Every document in the collection must be associated to at least one category. Documents may belong to more than one category. Each category must match with only one node in the hierarchical taxonomy. The FSS operates on a single unified taxonomy of categories for all the documents in the collection. When the facet types belong to a large, multilevel taxonomy, the FSS need to select the appropriate levels in the t-facet taxonomy to present to the user. In that case, we refer to them as level-n t-facets, where n is the level of the t-facet in the original taxonomy.

This hierarchical taxonomy can be seen as an directed acyclic graph, or tree, of categories, meaning that each node must have exactly one parent and can have zero or multiple children. Figure 1 demonstrates the tree structure with an emphasis on the levels. The taxonomy tree has a single root at level zero, this is the top of the tree. Level-n is the lowest level and contains the end leaf nodes. Categories (types) at the same level have the same distance from the root node.

This categorical taxonomy serves as the type hierarchy from which all the type-based facets are derived. Defining this taxonomy, its levels and the relationship between its nodes is crucial as this governs the t-facet ranking process.

Since this research is concerned with evaluating t-facet ranking rather than facet generation, the t-facets are directly collected and aggregated from the data. How the t-facets taxonomy is created or assigned to documents is outside of the

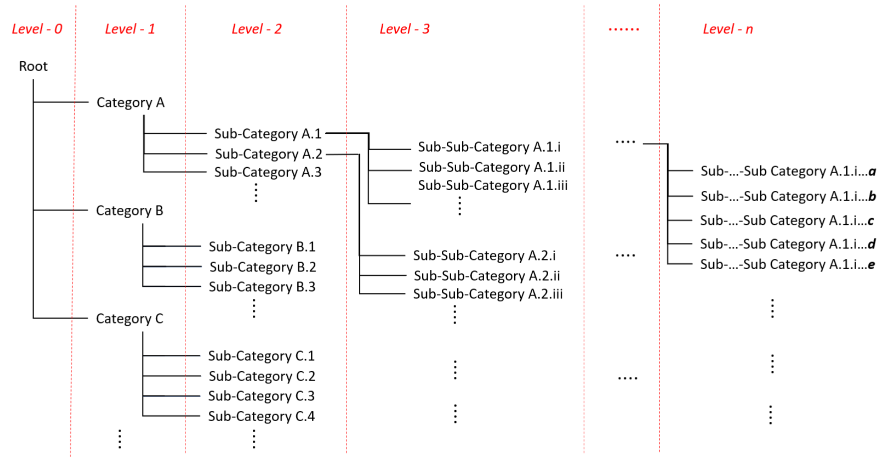

Fig. 1: Example of a multilevel hierarchical taxonomy of categories.

scope of this research. However, the ordering of t-facets is decided by the ranking algorithm.

The desired dataset should also includes user profiles. Each profile might contain basic information about the user, like name, age and gender, if available. It also has historical ratings the user gave to a number of documents within the collection. User ratings reflect whether they favored this document or not. Rating values belong to a numerical scale where the minimum value means dissatisfaction, and the maximum value reflects complete satisfaction. This scale values can also be mapped or classified to `positive`, `negative`, or `neutral` labels. We assume the middle point of the scale to be neutral, while values above it are positive, and values below it are considered as negative.

3.3 Required Preprocessing

The pre-processing, performed on the document categories, ensures that all the categories and their ancestors are linked to the document. Missing ancestors are added to the list of linked document categories. Ancestors common to multiple categories associated with the document are added only once. Pseudo-code explaining this preprocessing is shown in algorithm 1.

This preprocessing step is mandatory to ensure that the ranking method uses consistent category levels during the ranking process, as the ranking approaches consider a pre-configured number of levels. Let us consider the case of a document originally assigned to a fifth level category with a system operating at only the first two levels of the taxonomy. The pre-processing will ensure that the second level parent of the level-5 category is also included in the list. Without this step, the ranking approach might disregard this document and its t-facets from the ranking process.

Algorithm 1: Pre-processing Document Categories

```
Input: document, taxonomy_tree
Result: Expanded categories list
categories_list = retrieve_categories(document);
;                          // Returns list of categories for document
complete_category_list= {};
for category ∈ categories_list do
    complete_category_list.append(category);
    ancestors_list = find_ancestors(category);
    ;          // Returns all ancestors of category but the root node
    for ancestor ∈ ancestors_list do
        if ancestor ∉ categories_list then
            complete_category_list.append(ancestor);
        end
    end
end
Output: complete_category_list
```

3.4 Generating Evaluation Requests

Typically, existing IR datasets already contain requests and their relevance judgments at document level. In addition, datasets like the TREC-CS, provide the current search context and the user's historical ratings. However, datasets adapted from real-life require an additional step to create requests that imitate this type of information.

To achieve this, user information including user historical picks[5] can be utilized. Lets assume the users has m historical picks recorded in the original collection. We consider the most recent n picks as the intended search target. The n picks are then grouped according to their context (for example venues in the same city, or season of visit). Each context group that has a minimum threshold of t candidates will form a separate request. In order to produce a relevance judgment for each candidate in a request, the candidate's user rating is mapped into a relevance score; i.e. if the user rated this pick positively then it is considered relevant, otherwise it is considered irrelevant.

The personalized t-facet ranking task consists then in predicting the type-based facet sub-tree to which these relevant picks belong. The remaining of the ratings are part of the user's history and added to the user profile in the request. To avoid creating poor user profiles, only users with a minimum of r ratings in their profile are considered for this setup.

When the dataset under consideration does not provide explicit information needs, the framework generates artificial queries for each user. The queries are collected from the text associated with documents that the user has positively favored in the past (excluding the documents considered as candidates for eval-

[5]User picks are the user's interaction with the system that expresses a preference, like a rating, review, or feedback

uation). For this purpose, NLP methods for extracting keywords or tags can be employed to generate the top phrases which reflects the user's interests.

The contexts and the textual query will be used as input to the search engine. Both the quality of the generated queries and the retrieval model affect the evaluation of the facet ranking method. The search engine must be able to retrieve the intended search target in the relevant document set, otherwise, the appropriate facet needed to reach that document could be omitted in the ranked sub-tree. On the other side, assuming that such a document is in the initial pool retrieved by the search engine, it is the objective of the t-facet ranking approach to promote it at the top of the result list.

4 Experiments

In this section we demonstrate how the proposed framework can be applied in two different scenarios to obtain appropriate datasets for personalised t-facet ranking evaluation. The tourism domain, specifically the point of interest (POI) suggestion task, is chosen for the search task. In addition to the the availability of several online datasets, POI suggestion is a well-known personalization task where categories have already proven to play an important role [3, 2]. Moreover, we were able to identify two datasets in this domain that satisfy all the criteria listed in section 3.2: TREC-CS [7] and the Yelp datasets [6]. The following two sub-sections describe the two datasets, how they meet the criteria, and how they were customized to fit the t-facet ranked process.

4.1 Use Case 1: TREC-CS Dataset

The first dataset to which apply the proposed framework is the TREC Contextual Suggestion (TREC-CS) track dataset [7]. TREC-CS is a personalized Point-Of-Interest (POI) recommendation task in which participants develop systems to provide a ranked list of suggestions related to a given user profile and a context. We tackle the POI suggestion problem by ranking the types of venues as t-facets. The t-facet taxonomy is derived from the Foursquare venue category hierarchy [7]. To link as much Foursquare venues to TREC-CS POIs as possible, we complement the original data with three Foursquare supplementary datasets from [2, 3] and our own crawled POIs.

The contexts and requests are given by the dataset. In order to implement the document ranking, the input queries are formed by combining the user's tags weighed by their most common ratings provided by the same user. For the document ranking step, POIs web pages and reviews are indexed with Solr using BM25 with a NDCG value of 0.4023. The existence of relevance judgments makes it possible to evaluate our approach against a well established ground-truth. We follow the evaluation strategy used in the Faceted Search task of INEX 2011 Data-Centric Track [12].

[6] https://www.yelp.com/dataset , accessed June 2021

[7] https://developer.foursquare.com/docs/resources/categories, version: 20180323

4.2 Use Case 2: Yelp Dataset

In this use case, we apply the framework to the Yelp Open Dataset. In order to be comparable to TREC-CS dataset, we use it as a POI suggestion dataset. The user reviews, ratings, and POIs information are provided with the original dataset. POIs are assigned to categories derived from Yelp categories tree [8], which we use as t-facet taxonomy.

To ensure rich user profiles, only users with more than 170 reviews are included ($r = 170$). This threshold is suitable to have lengthy user profiles and a reasonable number of users in the dataset. We cap the user review at 1000 most recent reviews. For each user, we take the most recent 50 reviews ($n = 50$). To create visit context, we group the reviews by their city and state. Any context with more than 20 candidates is considered as a separate request ($c = 20$), this ensures a high number of relevant search targets for each request.

Unfortunately the Yelp dataset does not provide textual description for the POIs. Instead, we index all reviews, tips and attributes collected for each POI with Solr. The location is used as the initial filter to the search engine. In order to build a query, for each user we generate the top keywords from the latest 20 reviews in the user history (excluding all candidate target POIs). The query keywords are created using the Rapid Automatic Keyword Extraction algorithm (Rake)[9].

We created the relevance judgment for each request by mapping the user rating in the candidate target POIs into relevance score ($score = rating - 2$), thus POIs rated 2 and 1 will be considered irrelevant. This is useful to evaluate the document ranking separately. Also in this case the search engine is implemented in Solr using BM25, resulting in a NDCG value of 0.1608.

4.3 Personalized T-Facet Ranking Baselines

This section introduces the personalization baselines used to experiment with the two generated datasets. For methods that do not handle the hierarchical nature of the t-facets, we followed the two-step approach suggested by Ali et al. [6]. The first step scores each individual type-based facet. The second step uses the generated score to build the final t-facet tree to be displayed to the user.

T-Facet Scoring Methods

Probabilistic Scoring (Prob. Scoring) [6]. This is a probabilistic model to personalize t-facet ranking. Topic-based user profiles are collected from users' historical interactions with the system. The method assigns a score to the t-facet according to its relevancy to the user and query. We experiment using the no-background model with cosine similarity.

[8]https://www.yelp.com/developers/documentation/v3/all_category_list/categories.json
[9]https://github.com/csurfer/rake-nltk

Rocchio-BERT [5]. A lightweight method which utilizes Rocchio formula to build a vector representing the user interests. In this model, the user's profile is expressed in a category space through vectors that capture the users' past preferences. The BERT embeddings are used as t-facet representation in vector space. The t-facet score is the cosine similarity between its BERT vector and the user profile vector.

Most Prob. (Person) [8]. Most probable scoring method utilizes the user historical ratings. It is defined as the probability that the user will rate this facet positively. It is the number of time a t-facet was associated with a positive review by the user divided by the total number of POIs rated by the user.

Most Prob. (Collab) Also suggested by Koren et al. [8]. It is similar to the previous method, but computes the probabilities considering all the ratings from all the users in the system. It counts how many times this t-facet was rated positively by all the users divided by the number of POIs rated in the system.

MF-SVM . Matrix Factorization (MF) using SVM [4]. The matrix is built by adding the users and their t-facet ratings. T-facet ratings are collected from the POIs' ratings to which they are associated. Usually, the same facet may be associated with several POIs, thus has multiple ratings from the same user. In this case, this method takes the mean of the t-facet rating values.

T-Facet Tree Building Method .
The tree construction algorithm re-orders the original taxonomy tree by using the generated scores from the previous step [6]. It follows a bottom-up approach where the t-facets at the lower level in the taxonomy are sorted first, followed by all the ancestors of those t-facets, and so on up to the root of the hierarchy.

To build a final t-facet tree with v levels, we adopted a *fixed level* strategy [6]. The strategy respects the original taxonomy hierarchy and uses a predefined fixed page size for each t-facet level. It starts by grouping t-facets at level-v by their parent. Then, it sorts the parent nodes at level-$(v-1)$ by aggregating the scores of their top k children, the children are ordered by their relevance score generated in the previous step, and so on, up to level-1. We use *Max.* aggregation function to keep the top ranked t-facet at the top of the final tree.

5 Results and Discussion

Table 1 shows the statistics for the two generated datasets, Yelp and TREC-CS 2016. Both datasets operate on large multilevel taxonomies. Yelp taxonomy provides more categories, which make the ranking task more challenging. The statistics also show that the user profiles generated from Yelp dataset contain larger number of rated POIs per user; as a result we have more diverse t-facets rated by users. This provides richer data for the ranking algorithms to use in building the personalization model.

Table 1: **Comparison of Yelp and TREC-CS 2016 statistics after customization with the proposed framework.**

Item	TREC-CS 2016	Yelp
Total number of POIs	778K	160K
Total number of Taxonomy Types	942	1,566
Total number of Taxonomy Types in first two levels	459	994
Number Taxonomy Levels	5	4
Total number of Users	27 (209)	1,456
Average number of POIs rated per user	35.5 (54.1)	247.69
Total number of unique POIs rated by users	60 (4072)	81,163
Average number of Rated T-Facets in user history	38.18	135.8
Total Number of Requests	61	1495
Average number of t-Facets to be ranked per request	208	168.14

Yelp dataset also overcomes the limited availability of user profiles in TREC-CS 2016, in which users always rated the same 60 POIs. This affected the category distribution of the rated t-facets. In order to minimize the limited profile issues, we included users and ratings from TREC-CS 2015 dataset (statistics of this dataset are shown in the table between brackets). The results reported below use this improved user profiles.

Table 2 reports the evaluation results obtained for both datasets using the baselines mentioned in section 4.3. All results are reported by adopting the Fixed Level-Max tree building strategy, with level-1 and level-2 page size set to three. Several methods behave consistently across both collection. Rocchio-BERT outperformed the other personalization baselines on both datasets. Second in performance is the Prob. Scoring method; it is the second best method with Most Prob. (Collab). On the Yelp dataset, it provides the second best #Actions while its F-Scan results are worse than the Most Probable (Person.).

Table 2: **Results for baselines using Fixed-level (Max) strategy.**

Scoring Method	TREC-CS		Yelp	
	F-Scan	#Actions	F-Scan	#Actions
Rochio-BERT	**3.28**	**1.28**	**9.33**	**2.66**
Prob. Scoring	3.45	1.33	10.32	2.98
MF-SVM	3.91	1.49	19.29	4.07
Most Prob. (Person)	3.73	1.61	9.65	3.03
Most Prob. (Collab)	3.35	1.33	10.49	3.00

The affect of the quality of personal profiles is more evident in the Most Probable (Person.) performance, since this ranking method mainly depends on the individual user historical ratings. The approach improved with respect to both metrics in the Yelp dataset. On the other hand, MF-SVM performances dropped in the Yelp dataset; this indicates that the adoption of the average of documents rating as a t-facet rating is a poor heuristic to rate t-facets when many diverging ratings need to be aggregated.

6 Conclusions

This work presented a framework to generate benchmarks that can be used in evaluating personalized type-based facet ranking methods. The framework employs a fixed predefined t-facet taxonomy, which avoids propagating errors from the t-facet generation step to the facet ranking step. This enables the assessment and evaluation of the t-facet ranking process in isolation from other FSS components. We demonstrated the feasibility of this customization method by applying it in two different datasets. The first is TREC-CS dataset with existing relevance judgment at the document level, and the other is a larger dataset released by Yelp. In this last dataset, users' historical interactions are employed to compensate the lack of relevance judgments. As future plan, we intend to experiment with additional datasets in other domains, like product shopping or digital libraries.

Acknowledgements This work was supported by the ADAPT Centre, funded by Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106; 13/RC/2106_P2) and co-funded by the European Regional Development Fund.

References

1. Abel, F., Celik, I., Houben, G.J., Siehndel, P.: Leveraging the semantics of tweets for adaptive faceted search on twitter. The Semantic Web (2011)
2. Aliannejadi, M., Mele, I., Crestani, F.: A cross-platform collection for contextual suggestion. In: SIGIR. ACM (2017)
3. Bayomi, M., Lawless, S.: Adapt_tcd: An ontology-based context aware approach for contextual suggestion. In: TREC (2016)
4. Chantamunee, S., Wong, K.W., Fung, C.C.: Collaborative filtering for personalised facet selection. In: IAIT. ACM (2018)
5. Esraa Ali, Annalina Caputo, S.L., Conlan, O.: Personalizing type-based facet ranking using bert embeddings. In: SEMANTiCS 2021
6. Esraa Ali, Annalina Caputo, S.L., Conlan, O.: A probabilistic approach to personalize type-based facet ranking for poi suggestion. In: ICWE 2021
7. Hashemi, S.H., Clarke, C.L., Kamps, J., Kiseleva, J., Voorhees, E.M.: Overview of the trec 2016 contextual suggestion track. In: TREC (2016)
8. Koren, J., Zhang, Y., Liu, X.: Personalized interactive faceted search. In: WWW. ACM (2008)
9. Tunkelang, D.: Faceted search. Synthesis Lectures on Information Concepts, Retrieval, and Services pp. 1–80 (2009)
10. Vandic, D., Aanen, S., Frasincar, F., Kaymak, U.: Dynamic facet ordering for faceted product search engines. IEEE Transactions on Knowledge and Data Engineering **PP**(99), 1–1 (2017). https://doi.org/10.1109/TKDE.2017.2652461
11. Vandic, D., Frasincar, F., Kaymak, U.: Facet selection algorithms for web product search. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. pp. 2327–2332. ACM (2013)
12. Wang, Q., Ramírez, G., Marx, M., Theobald, M., Kamps, J.: Overview of the INEX 2011 Data-Centric Track. In: Focused Retrieval of Content and Structure. Springer (2012)