International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 01-05

Personalized Searching Technique Using Automatic Facet Selection Algorithm

R.Saranya[1], Dr.R.Suguna[2]
[1]*(Department of computer science, Theivanai Ammal College for Women, India)*
[2]*(Department of computer science, Theivanai Ammal College for Women, India)*

Abstract: *The price play a role when a consumer decides to choose where to buy a product online. Therefore, online retailers pay special attention to the usability and efficiency of their Web shop user interfaces. Now a day, many Web shops make use of the so-called faceted navigation user interface, which is in literature also sometimes referred to as 'faceted search'. Facets are usually grouped by their property in user interfaces, in order to prevent them from being scattered around, and, thereby, confusing the user. Multifaceted search is a commonly used interaction paradigm in e-commerce applications, such as Web shops. Because of the large amount of possible product attributes, Web shops usually make use of static information to determine which facets should be displayed. Unfortunately, this approach does not take into account the user query, leading to a non-optimal facet drill down process. In this paper, propose on automatic facet selection, with the goal of minimizing the number of steps needed to find the desired product. These articles propose several algorithms for facet selection, which a paper evaluates against the state-of-the-art algorithms from the literature.*
Keywords: *Faceted Search, collaborativerecommendation, Facet Selection, personalization.*

I. Introduction

Online product search has now a day become more important than ever, as consumers purchase more often on the Web. One explanation for this is that the Web facilitates the user finding products that better match their.ing channels. Not only do the users have access to more information (e.g., user reviews, exact product information), they also found it easier to shop from their homes. On the other hand, because of the many options, users are often overwhelmed and found it difficult to browse through the available products. Multifaceted search, also sometimes referred to as `guided navigation', is a popular interaction paradigm that allows users to navigate through multidimensional data. One of the main uses of multifaceted search is in the domain of e-commerce, i.e., Web shops. It is being employed to solve the parametric product search problem for Web shops that have collected local openings and product information. For example, in a Web shop the user might enter a query like `Samsung, gps' in order to search for a Samsung phone that has built-in GPS capabilities. After showing the initial result set, most Web shopping interfaces display the facets of the products in the result set, which can be used to further drill down into the results set. The facets in this case are product attribute and value combinations. An important problem of multifaceted search is the, Selection of facets that should be displayed for each query. Because products have so many attributes that could be displayed as facets, Web shops usually have some static business logic to display certain facets for each result set. Although this works for local Web shops that do not have many product categories, the creation of this business logic is a time consuming process and is not appropriate for Web wide product search. One solution to this problem is to employ an optimized facet selection process. The goal of such an optimization process is to show facets that effectively partition the product search space so that the user can easily drill down and found its desired product. In literature, this is referred to as the facet selection problem, which can be expressed as the optimization of a hyperactive media link generation process.

II. Facet Selection

The goal of this paper is to reduce the search effort of a user that is searching for a product that meets his needs. In this section, we first give the formal problem formulation and then we present the considered facet selection algorithms. We assume that the number of results scanned by a user (before finding his desired product)represents the search effort. The main use case that we consider is that a user submits a query to the product search engine Next, the search engine computes a ranked list of products and a set of facets of size that are to be displayed. Furthermore, the set represents all facets that belong to all products Similar to; we incorporate four assumptions about the user in the considered simulation strategies for the evaluation. First, we assume that for each user, who has submitted query q, there exists a single target product that full fills the user's needs. This target product is assumed to always be present in the initial result set, but can be ranked very low. Second, the user ends the search session if it founds in the top results. Third, a user exactly knows which facets (if any) from are associated to. Last, if there are any facets that are associated with, we assume that the user then also selects these. We incorporate four assumptions about the user in the considered simulation strategies for the evaluation. First, we assume that for each user u, who has submitted query, there exists a single target product that full fills the user's needs. This target product is assumed to always be present in the initial result set, but can be ranked very low. Second, the user ends the search session if it founds in the top results. Third, a user exactly knows which facets (if any) from are associated to. Last, if there are any facets that are associated with what, we assume that the user then also selects these. In this paper, we also assume that multiple clicks (drill downs) can occur. More specifically, we assume that the above described process can repeat itself a maximum of times (iterations). If the user founds the desired product in the top results in less than iterations, then the search session ends prematurely, otherwise it ends after iterations. If we let it to remain unchanged, then the result set at any iteration can be denoted as where represents the previously selected facets. Similarly, the proposed facets by the search engine at any iteration are denoted

2. DATASET

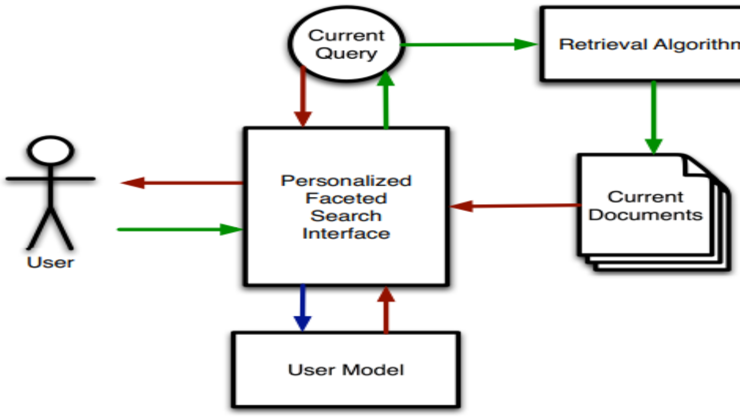
2.1 Data Collection

We use a data set that is gathered from Kieskeurig.nl, The largest price comparison site in the Netherlands. This service does not only provide price comparisons, but also has very detailed information on products. For this evaluation, we focused on consumer electronics and chose mobile phones to be the category of products that we use in the experiments. The data set contains 980 products for which we have key/value pairs, i.e., product attributes. All product information is in Dutch, but should be understandable also for non-Dutch speaking people because of the frequently used English terminology in the product attributes. Using the product attributes, we created the facets using the following rules. A facet is a combination of a product attribute and a value (or range of values. For product properties that represent multivalve qualitative values, such as `Supported Video formats', we created a binary facet for each value. Similarly, for single-valued qualitative product attributes, we created a single facet for each value. For all the quantitative properties, we manually defined the ranges that would represent the different facets. As a result of this facet creating process, we obtained 487 facets for the 980 products. The size and variety of this data set allows for a thorough evaluation of the facet selection algorithms

2.2 Personalized facet search mechanism

An efficient personalized faceted search mechanism can be used to: 1) solve millions of e-commerce users' immediate information needs. 2) Help users better understand the data, especially the data space relevant to the user, and 3) help users better understand how the engine works through the simple interactive interface, and there by train users in how to make more effective use of the interface over time.In literature, can find approaches in personalized automatic faceted search. [1] Faceted search applications require faceted data, namely the existence of facet hierarchies and the mapping of documents on to those hierarchies.

Figure 1: Personalized Interface Block Diagram



III. Related Works

Based on our existing work [2], this study focuses on implementing a dynamic faceted navigation system, as a decision support system, which groups results based on search context using the Semantic Web technology. Similarity matching to the searched product [3], Existing studies have shown that users find faceted search interfaces Intuitive and easy to use [4], However, the unstructured webpage make it is difficult to solve the issue very good. WisamDakka presents an unsupervised technique for automatic extraction of Facets useful for browsing text databases [5]. In [6] the approach was extended and improved with a focus on product search.

Using additional user assumptions and the same theoretic approach as [7], two new methods for facet sorting were developed in previous article. Faceted search has become a popular technique in commercial search applications, especially for online retailers and libraries [8, 9]. One of primary reason is its ability to let a user iteratively define the intended query by adding or removing constraints (facet bindings) while browsing the data. This considerably reduces the burden on the user, as she needs not create the perfect query upfront - often a problem most users face when querying over third-party data sources whose underlying data distribution is hidden from the user [10]. The facet suggestion model is driven by our intent to provide a minimum effort database exploration solution for enterprise users. We focus on a simple but intuitive metric for measuring effort: the expected number of queries that the user has to answer in order to reach the tuples of interest [11]. Thus the first challenge is to judiciously select the facets to be suggested, so that the user reaches the desired tuples with minimum effort. The second challenge we have to deal with is the problem of uncertainty in user knowledge in the facet selection process. In particular, the uncertainty over an attribute refers to the probability of the user being able to provide a desired value for that attribute. We solve this challenge by extending the decision tree model to account for such attribute uncertainties [12].

IV. Proposed Evaluation Methodology

This Article proposes new algorithms for the facet selection problem in product search. We evaluate several approaches and compare our proposed algorithms against several state-of-the-art facet selection algorithms from the literature. Our proposed algorithms aim to partition the space in the most effective manner and thus allow the user to drill down in the least amount of time. We perform the evaluation on a large data set and analyze the results, differently from previous works, across three different measures.

4.1User Interface Utility

While the specifics vary between individual faceted search interfaces, every interface shares certain characteristics. In general, a faceted search interface is divided into three parts. The first part is a list of the facets in the document collection. Each facet has a list of available values associated with it. When there are a large number of values for a facet, interfaces tend to display only a fraction of the available values, but allow the user to view the complete list upon request. The user can restrict the current query by selecting facet-value pairs from this list subset of the search space the user is examining, but also allows the user to broaden the current query by removing some previously selected facet-value pairs.

4.1.1 Personalized Facet-value pair Recommendation - As stated in the introduction, one of the keys to building an effective faceted search interface is presenting the user with facet-values that that a relevant to the user's current search task. If the presented facet-values are not relevant to the task, the user could be forced to spend extra effort to find his/her document(s), or in the worst case not findhis/her document(s) at all. This section describes several possible algorithms to select facet-value pairs.
4.1.2 Facet-Value Pair Suggestions -After a user performs an action in the middle of a session, the system needs present a list of facet value pairs. We can view this as a feature selection ask, and the following is an incomplete list of algorithms that can be used
4.1.3Most Frequent - This is the simplest suggestion method. In this method, the facet-value pairs that are found in the currently selected documents are counted, and the most frequent values for each facet are presented to the user for query refinement. This method is popular among many commercially available faceted search interfaces, and thus provided an appropriate baseline for comparison.
4.1.4 Most Probable - In this method, the facet-value pairs in the currently selected documents are ranked according to their probability of being included in a document relevant to the user. These probabilities can either be determined by the relevance judgments by the community of users (*Collaborative Prob.*), or personalized for each individual user (*Personal Prob)*.This method was examined as it can be easily integrated into adaptive and personalized retrieval algorithms.
4.1.5 Mutual Information -The point wise mutual information between the presence of a facet-value pair appearing in a document and a document's relevance is calculated. The most informative values are then presented to the user for query refinement. Mutual information was considered as a facet-value suggestion method and since it is a common method used for feature selection.

4.2 Starting/Landing Page for Faceted Search

A faceted search system needs to present a good starting/landing page for a user. Since a user's profile describes the qualities that define a relevant document for that user, this information can be used by the systemto initially place each user nearer to his/her relevant documents even before the user begins to formulate a query. This is accomplished by examining the user's profile and automatically constructing an initial query based on facet-value pairs that are likely to be contained in the user's relevant documents. This automatically constructed query, along with the documents returned by this query, creates the *start state*. The following three methods to determine user's start state are proposed.

4.2.1 Null Start State -This is the simplest start state creation method. In this method, each user begins in a state with no face-value pairs selected by default and no pre-fetched documents. This method served as a baseline for comparing the other start state creation methods as this is the most prevalent method for beginning user initiated searches in information retrieval systems.

4.2.2 Collaborative Start State -In this method the system automatically issues a query containing the facet-value pairs that are the mostly likely to be contained in relevant document as determined by the common Bayesian prior. This query is then issued to the underlying retrieval algorithm and the matching documents are initially suggested to the user. Since the start state is created by the common prior, every user is presented thesame art state.

4.2.3 Personalized Start State -This method is similarly to the method above, except that the default query is determined by each user's profile.

4.3 Empirical Evaluation

In order to demonstrate these ideas, a set of experiments were carried out using documents from the Internet Movie Database (IMDB) corpus along with real user relevance judgments for each document from the Movie Lens and Netflix Prize corpora. These data sets were chosen since they provided documents containing facets, such as director and *actor*, along with real user relevance judgment son these documents. The IMDB corpus was trimmed to contain only documents found in either the Movie Lens or Netflix corpora. This led to approximately 8,000 documents containing 367,417 facet-value pairs spread among 19facets.Both the Movie Lens and Netflix corpora were reduced to approximately 5,000 unique users through uniform sampling, giving 742, 036 and 633,257 user judgments respectively. In both data sets users expressed a preference for retrieved documents based on a 1 to 5 rating scale. Each rating was converted into Boolean relevance judgments by assuming all movies that were rated 4 or greater were relevant, and all movies rated 3 or less were non-relevant. These user judgments were randomly divided 90% for training and 10% for testing. For simplicity, all facets were assumed to be nominal. Each user model was a multivariate Bernoulli distribution, with the shared prior being a multi variant Gamma distribution. Without loss of generality, a simple reward mechanism shown in Table 4 is used for evaluation. The interface was configured to return a maximum of 10 matching documents per page, and a maximum of 5 values for each facet.

V. Experimental Results

For the evaluation, we simulate a user that is in a faceted search session. There are two aspects that are important in this type of simulation. First, we need a way to generate queries that are sufficiently realistic for theexperiments. Second, we need one or more simulation strategies of users in order to simulate the clicking on a facet. Before we go into the details of these two aspects, let us first explain on a high level how we have designed the simulation. Given a query, we submit it to the product search engine, after which, for every product that we consider as a possible target product, we simulate a faceted search session. The set of possible target products consists of the 100 products after the top products. We set which results in performing the simulation with each product ranked in the range as a target product. The reason for this is that we want to measure how the algorithms perform for many different target products. Next, the ranked search results are obtained and a faceted search session is simulated, where a user is aware of the target product, but is only able to recognize it when it appears in the top-10 results. The user keeps clicking on a facet (described shortly) until either the target product.Action corresponds to a bigger user utility. Four notes worthyconclusions can be made from these results. First, point wise mutual information (PMI) significantlyunder Performed when compared to the other facet value selection methods. Mutual information measures the correlation between two random variables, in this case the presence or absence of a facet-value pair and a document's relevance. Point wise mutual information rather than complete mutual information was used because only facet-value pairsthat have a positive correlation should be suggested to the user. PMI breaks down when there are facet-value pairs thatare strongly correlated with relevant documents, but occur in only tiny fraction of all relevant documents. For example,if all films containing the facet-value pair *genre=film no ire* ranked highly, then *genre=film will* be suggested early for query refinement, even if *genre=filmnoir*is contain edibles than 1 percent of all relevant documents. These resultsshow that correlation measure such as mutual information are not a good choice for this type of selection problem, since the probability of utilizing the suggested features is more important than how tightly correlated they are with relevance. Finally, simply suggesting the most frequent values foreach facet performed well when compared to the personalized suggestion methods. There are two possible reasons for this. First, the frequency of facet value pairs in the documents is correlated with users' idea of what makes a document relevant. In general this may not be the case, and thus frequency may not be a good facet-value selectionmechanism for when the users' expectations do not closely match what is contained in the document repository. Inthis case frequency would fail to provide good suggestions, and the personalized probability models would be shown tobe superior. Second, the *Personal Prob*algorithm used to select facet-value pairs is not good enough and far from optimal. This is not surprising since the probabilistic model proposed in this paper is based on strong assumptions thatmay not be true on the evaluation data sets.

VI. Conclusions

This article focused on automatic facet selection in the domain of e-commerce, for the purpose of minimizing the number of steps required by the user in order to found its desired product. This paper proposed several facet selection algorithms, which we evaluated against the state-of-the-art algorithms from literature. Furthermore, we implemented all considered facet selection algorithms in a freely available Web application called faccy.net.The evaluation was performed with simulations employing 1000 queries, 980 products, 487 facets, and three drill down strategies. We used three different evaluation metrics. The experimentalenvironment is repeatable and controllable, which makes it abench mark able evaluation environment. Although the simulated users differ from real users, the evaluation methodology does provide insight into understanding how various faceted interface design algorithms perform. This paperdoes not intend to claim whether this evaluation methodis better or worse than user studies. Instead, the outlinedapproach serves to complement user studies by being cheap,repeatable, and controllable.How to select a set of facet-value pairs at each step of the interaction process to optimize a user utility is a more fundamental that requires future research. This paper serves a first step towards personalized faceted search. The facet value pair selection algorithms examined in this paper is far from optimal.

References

[1]. S. Liberman and R. Lempel. Approximately optimal facet selection.In Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC 2012), pages 702{708.ACM, 2012.

[2]. H.-J. Kim, Y. Zhu, W. Kim, and T. Sun, "Dynamic Faceted Navigation in Decision Making using Semantic Web Technology," Decision Support Systems, vol. 61, pp. 59–68, 2014..

[3]. S. Johnson Lim, Y. Liu, and W. Lee. Multi-facet product information search and retrieval using semantically annotated product family ontology. Information Processing & Management, 46(4):479{493,2010.

[4]. J. Koren, Y. Zhang, and X. Liu. Personalized Interactive faceted search.In Proceedings of the 17th International conference on World Wide Web (WWW 2008), pages 477{486.ACM, 2008.

[5]. WisamDakka, Panagiotis G. Ipeirotis, "Automatic Extraction of Useful Facet Hierarchies from Text Databases," icde, pp.466-475, 2008 IEEE 24th International Conference on Data Engineering, 2008.

[6]. S. Liberman and R. Lempel, "Approximately optimal facet value selection," Sci. Comput. Program., vol. 94, pp. 18–31, 2014.

[7]. D. Vandic, F. Frasincar, and U. Kaymak, "Facet selection algorithms for Web product search," in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., 2013, pp. 2327–2332.

[8]. G. Marchionini and B. Brunk. Toward a general relation browser: A GUI for information architects. Journal of Digital Information, 4, 2003.

[9]. D. Tunkelang. Dynamic Category Sets: An Approach for Faceted Search. In Faceted Search Workshop (SIGIR '06), 2006.

[10]. E.Stoica, M.Hearst, and M.Richardson, "Automating creation of hierarchical faceted metadata structures," in NAACL-HLT, 2007.

[11]. S. Roy, H. Wang, G.Das, U.Nambiar, and M.K.Mohania, "Minimumeffort driven dynamic faceted search in structured databases," in CIKM, 2008.

[12]. V.T.Chakravarthy, V.Pandit, S.Roy, P.Awasthi, and M.Mohania, "Decision trees for entity identification: Approximation algorithms and hardness results," in PODS, 2007.