iJRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: IV Month of publication: April 2018

DOI: <http://doi.org/10.22214/ijraset.2018.4570>

www.ijraset.com

Call: 08813907089

E-mail ID: ijraset@gmail.com

Research Paper on Implementation of Mining Facets Automation for the Searched Queries

Shraddha Londhe¹, Dr. Hemant Deshmukh², Prof. Pallavi Dhande³

¹M.E 2ND Year Computer Science and Engineering, ²Head of the Department, ³Assistant Professor, Dr. Rajendra Gode Institute of Technology & Research

Abstract: Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. The process of finding query facets which are in the form of multiple groups of words or phrases will be address as a problem to explain and summarize the content covered by a query. Facet is a set of items which describe and summarize one important aspect of a query, here a facet item is typically a word or phrase. A query may have multiple facets that summarize the information about the query from different perspectives. It is assumed that the important aspects of a query are usually presented and repeated in the query's top retrieved documents in the style of lists and query facets can be mined out by aggregating these significant lists. Query facets provide interesting and useful knowledge about a query and thus can be used to improve search experiences in many ways. To assist information for finding faceted queries a technique is explore that represent interesting facets of a query using groups of semantically related terms extracted from search results. Web search queries are often multi-faceted, which makes a simple ranked list of results inadequate. So, a method is used, refer to as QDMiner to automatically mine query facets by extracting and grouping frequent lists from free text, HTML tags, and repeat regions within top search results. Search results based on used method will simply improve the efficiency of users ability to find information easily.

I. INTRODUCTION

A query facet is a set of items which describe and summarize one important aspect of a query. Here a facet item is typically a word or a phrase. A query may have multiple facets that summarize the information about the query from different perspectives. Query facets provide interesting and useful knowledge about a query and thus can be used to improve search experiences in many ways. users can understand some important aspects of a query without browsing tens of pages. For example, a user could learn different brands and categories of watches. We can also implement a faceted search [1], [2], [3], [4]based on the mined query facets. User can clarify their specific intent by selecting facet items. Then search results could be restricted to the documents that are relevant to the items. Important pieces of information about a query are usually presented in list styles and repeated many times among top retrieved documents. Thus we propose aggregating frequent lists within the top search results to mine query facets and implement a system called QDminer. More specifically QDMiner extracts lists from free text, HTML tags and repeat regions contained in the top search results groups them into clusters based on the items they contain then rank the clusters and items based on how the lists and items appear in the top results. Query facets are mined by the following four steps:

- 1) *List and Context Extraction* Lists and their context are extracted from each document.
- 2) *List Weighting* All extracted lists are weighted and thus some unimportant or noisy lists that occasionally occurs in a page can be assigned by low weights
- 3) *List Clustering* Similar lists are grouped together to compose a facet. We do not use individual weighted lists as query
- 4) *Facet and Item Ranking* Facets and their items are evaluated and ranked. Based on our motivation that a good facet should frequently appear in the top results, a facet is more important.

Figure 1 : Working Of Proposed System

In project, there are two ecommerce website that are ecommerce 1 and ecommerce 2. Both website have its own admin and various users. Admin have authority to upload product that uploaded product are going through 4 steps that are list extraction, list weighting, list clustering and data ranking. These 4 steps are very important steps throughout the project because these 4 steps helps to convert unstructured data to structured data then this structured data submit to the database of websites and with the help of admin structured data in database of websites are finally submitted to database of search engine. And if user search any product then data comes in structured format in proper way from database of search engine

In this paper we also implement the impression and click technique.

Impression is a measurement of responses from a Web server to a page request from the user browser, which is filtered from robotic activity and error codes, and is recorded at a point as close as possible to opportunity to see the page by the user. An impression is when an ad is fetched from its source, and is countable. In this topic by using impression we show most visited items by user in our search engine. Click-through rate (CTR) is the ratio of users who click on a specific link to the number of total users who view a page, email. It is commonly used to measure the success of an particular website as well as the effectiveness of email campaigns.

II. IMPLEMENTATION

In project, there are two ecommerce website that are ecommerce 1 and ecommerce 2. Both website have its own admin and various users. Admin have authority to upload product that uploaded product are going through 4 steps that are list extraction, list weighting, list clustering and data ranking. These 4 steps are very important steps throughout the project because these 4 steps helps to convert unstructured data to structured data then this structured data submit to the database of websites and with the help of admin structured data in

database of websites are finally submitted to database of search engine. And if user search any product then data comes in structured format from database of search engine This shows various screenshots how the implementation of the system is done. There are six modules in the project.

A. Ecommerce 1(Website 1)

1) Ecommerce 1 Home Page

Screenshot 1: Ecommerce 1 Home Page:

Screenshot 6.1 is of home page of ecommerce 1 that is website 1.

2) Ecommerce 1 Registration Form

Screenshot 2: Ecommerce 1

3) *Registration Form:* In Registration Form Of Ecommerce 1, new user have to enter full name, email address, password, mobile number, city, pin code, state, country, address. In this form all fields are necessary.

4) Ecommerce 1 Login Page

Screenshot 3:Ecommerce 1 Login Page

This is the login page of the ecommerce 1 website from where user can login with the help of email address and password

5) Ecommerce 1 Products

Screenshot 4 : Ecommerce 1 Products

In product page there is list of all product as well as there is also link of categories so that user can directly go to that page of particular category if user wants.

B. Admin of Ecommerce 1

1) Login Page

Screenshot 6: Login Page Of Admin 1

This is login page of admin of ecommerce 1 in which username and password of admin is require.

2) Home Page

Screenshot 7: Home Page Of Admin 1

This is home page of admin of ecommerce Admin have authority to handle all the work of ecommerce 1 such as upload product, view product.

3) Upload Product

The screenshot shows the 'Upload Product' form in the Ecommerce 1 Admin interface. The form includes fields for Product Name, Product Price, Product Category (a dropdown menu), Product Brand (a dropdown menu), Product Description, Product Color, and two image upload slots (Image 1 and Image 2). A 'Submit' button is located at the bottom right of the form.

Screenshot 8: Upload Product Admin1

In this upload form we have to fill product name, product price, product category, product brand, product description, product colour, product image and after submitting the product will show in ecommerce 1.

4) Product Details

The screenshot shows the 'Product Details' page in the Ecommerce 1 Admin interface. It displays a table with product information, including Product Name, Price, and Description. The table has columns for Product Name, Price, and Description. The first row shows a product named 'Smart Watch' with a price of '10,000' and a description.

Screenshot 9: Product Details Admin 1

This is product details of admin of ecommerce 1. In this product details admin can view uploaded product.

C. Ecommerce 2 (Website 2)

1) Login Page

The screenshot shows the login page of the Ecommerce 2 website. It features a 'Login' section with input fields for 'Email and Username' and 'Enter your Password'. There is a 'Forgot Password?' link and a 'Login' button. The page has a purple header and a blue footer.

Screenshot 10: Ecommerce 2 Login

This is the login page of the ecommerce 2 website from where user can login with the help of email address and password

2) Registration Form

The screenshot shows the registration form of the Ecommerce 2 website. It includes input fields for 'Full Name', 'Email Address', 'Password', 'Mobile Number', 'City', 'Pin Code', 'State', and 'Country'. There is a 'Registration for service' button at the bottom.

Screenshot 11: Registration Form

In Registration Form of Ecommerce 2, new user have to enter full name, emailaddress, password, mobile number, city,pin code, state, country, address. In this form all fields are necessary.

3) Product

Screenshot 12 : Product Of Ecommerce

In product page there is list of all product as well as there is also link of categories so that user can directly go to that page of particular category if user wants.

D. Admin2

1) Login Page

Screenshot 14: Login Page Of Admin Of Ecommerce 2

This is login page of admin of ecommerce 2 in which username and password of admin is require. Admin have authority to handle all the work of ecommerce 2

2) Home Page

Screenshot 15: Home Page Of Admin 2

This is home page of admin of ecommerce2. Admin have authority to handle all the work of ecommerce website 2 such as upload product, view product.

3) Upload Product

Screenshot 16: Upload Page Of Admin Of Ecommerce 2

This is upload page of admin of ecommerce 2. In this upload form we have to fill product name, product price, product category, product brand, product description, product colour, product image and after submitting the product will show in ecommerce 2.

4) Product Details

Screenshot 17: Product Details Of Admin Of Ecommerce 2

This is product details of admin of ecommerce 2. In this product details admin can view uploaded product.

5) Payment Details

Screenshot 18: Payment Details

If user wants to do online payment then user have to enter name of card, card number, date, cvv number, and pin number. And after payment order place successfully.

E. Search Engine

1) Search Engine Home

Screenshot 19: Search Engine Home

This is main part of project, search engine which gives the proper result to the user which store in the database of search engine. There are two ecommerce websites that are ecommerce 1 and ecommerce 2. Both websites have their own admin and various users. Admin upload product that uploaded product are going through 4 steps that are list extraction, list weighting, list clustering and data ranking. These 4 steps help to convert unstructured data to structured data then this structured data submit to the database of websites and with the help of admin structured data in database of websites are finally submitted to database of search engine. And if user search any product then data comes in structured format in proper way from database of search engine.

2) Product Clicks Portal Wise

Screenshot 20: Product Clicks Portal Wise

The click-through rate is the number of times a click is made on the product divided by the total impressions. Product clicks are given separated for both the ecommerce website. Each item name is given with its product click portal wise

3) Product Impression Portal Wise

Product ID	Product Name	Portal	Impression
118	VIVO V7 (340, 32 GB) (4 GB RAM)	Web 1	0
115	Mi Mix 2 (Black, 128 GB) (6 GB RAM)	Web 1	0
119	Moto C Plus (Fine Gold, 16 GB) (2 GB RAM)	Web 1	3
124	HP Envy Core (3.7m Gen - (4 GB/128 GB SSD/Windows 10 Home) 13-4679TU Thin and Light Laptop (13.3 inch, Gold, 1.32 kg)	Web 1	0
123	Acer Swift 3 Ryzen 5 Quad Core - (8 GB/1 TB HDD/Linux) SF315-41 Laptop (15.6 inch, Black, 2.1 kg)	Web 1	0
125	Moto M (Silver, 64 GB) (4 GB RAM)	Web 1	2
126	Moto M (Silver, 64 GB) (4 GB RAM)	Web 1	0
127	Moto Z2 Force (Super Black, 64 GB) (6 GB RAM)	Web 1	2
128	Moto Z2 Force (Super Black, 64 GB) (6 GB RAM)	Web 1	0
129	Apple iPhone 6 (Space Grey, 32 GB)	Web 1	0
130	VIVO V7+ (Matte Black, 64 GB) (4 GB RAM)	Web 1	0

Screenshot 21: Product Impression Portal Wise

An impression is when an product is fetched from its source, and is countable. Whether the product is clicked is not taken into account. Each time an product is fetched, it is counted as one impression

F. Admin Section of Search Engine

1) Login page

Admin Section

Member Login

Please provide your details

☐ Remember me

Screenshot 22: Login page

This is login page of admin of search engine in which username and password of admin is require. Admin have authority to submit the structured data to the database of the search engine.

2) Home Page Of Admin Of Search Engine

SEARCH ADMIN SECTION

Extract Data Product Clicks Product Impressions

Search Your Query

Screenshot 23: Home page

In home page of admin of search engine, there are three links one is of extract data which is direct to home page which contain web1 and web2 two buttons, second is product clicks and third is product impression. There is also log out button to log out from the session.

3) Product Clicks Portal Wise In Admin Of Search Engine

Product Clicks Portal Wise			
Product	Impressions	Clicks	CTR
Portal Web 1			
115	100	10	10%
116	100	10	10%
117	100	10	10%
118	100	10	10%
119	100	10	10%
120	100	10	10%
121	100	10	10%
122	100	10	10%
123	100	10	10%
124	100	10	10%
125	100	10	10%
126	100	10	10%
127	100	10	10%
128	100	10	10%
129	100	10	10%
130	100	10	10%
131	100	10	10%
132	100	10	10%
133	100	10	10%
134	100	10	10%
135	100	10	10%
136	100	10	10%
137	100	10	10%
138	100	10	10%
139	100	10	10%
140	100	10	10%
141	100	10	10%
142	100	10	10%
143	100	10	10%
144	100	10	10%
145	100	10	10%
146	100	10	10%
147	100	10	10%
148	100	10	10%
149	100	10	10%
150	100	10	10%
151	100	10	10%
152	100	10	10%
153	100	10	10%
154	100	10	10%
155	100	10	10%
156	100	10	10%
157	100	10	10%
158	100	10	10%
159	100	10	10%
160	100	10	10%
161	100	10	10%
162	100	10	10%
163	100	10	10%
164	100	10	10%
165	100	10	10%
166	100	10	10%
167	100	10	10%
168	100	10	10%
169	100	10	10%
170	100	10	10%
171	100	10	10%
172	100	10	10%
173	100	10	10%
174	100	10	10%
175	100	10	10%
176	100	10	10%
177	100	10	10%
178	100	10	10%
179	100	10	10%
180	100	10	10%
181	100	10	10%
182	100	10	10%
183	100	10	10%
184	100	10	10%
185	100	10	10%
186	100	10	10%
187	100	10	10%
188	100	10	10%
189	100	10	10%
190	100	10	10%
191	100	10	10%
192	100	10	10%
193	100	10	10%
194	100	10	10%
195	100	10	10%
196	100	10	10%
197	100	10	10%
198	100	10	10%
199	100	10	10%
200	100	10	10%
201	100	10	10%
202	100	10	10%
203	100	10	10%
204	100	10	10%
205	100	10	10%
206	100	10	10%
207	100	10	10%
208	100	10	10%
209	100	10	10%
210	100	10	10%
211	100	10	10%
212	100	10	10%
213	100	10	10%
214	100	10	10%
215	100	10	10%
216	100	10	10%
217	100	10	10%
218	100	10	10%
219	100	10	10%
220	100	10	10%
221	100	10	10%
222	100	10	10%
223	100	10	10%
224	100	10	10%
225	100	10	10%
226	100	10	10%
227	100	10	10%
228	100	10	10%
229	100	10	10%
230	100	10	10%
231	100	10	10%
232	100	10	10%
233	100	10	10%
234	100	10	10%
235	100	10	10%
236	100	10	10%
237	100	10	10%
238	100	10	10%
239	100	10	10%
240	100	10	10%
241	100	10	10%
242	100	10	10%
243	100	10	10%
244	100	10	10%
245	100	10	10%
246	100	10	10%
247	100	10	10%
248	100	10	10%
249	100	10	10%
250	100	10	10%
251	100	10	10%
252	100	10	10%
253	100	10	10%
254	100	10	10%
255	100	10	10%
256	100	10	10%
257	100	10	10%
258	100	10	10%
259	100	10	10%
260	100	10	10%
261	100	10	10%
262	100	10	10%
263	100	10	10%
264	100	10	10%
265	100	10	10%
266	100	10	10%
267	100	10	10%
268	100	10	10%
269	100	10	10%
270	100	10	10%
271	100	10	10%
272	100	10	10%
273	100	10	10%
274	100	10	10%
275	100	10	10%
276	100	10	10%
277	100	10	10%
278	100	10	10%
279	100	10	10%
280	100	10	10%
281	100	10	10%
282	100	10	10%
283	100	10	10%
284	100	10	10%
285	100	10	10%
286	100	10	10%
287	100	10	10%
288	100	10	10%
289	100	10	10%
290	100	10	10%
291	100	10	10%
292	100	10	10%
293	100	10	10%
294	100	10	10%
295	100	10	10%
296	100	10	10%
297	100	10	10%
298	100	10	10%
299	100	10	10%
300	100	10	10%
301	100	10	10%
302	100	10	10%
303	100	10	10%
304	100	10	10%
305	100	10	10%
306	100	10	10%
307	100	10	10%
308	100	10	10%
309	100	10	10%
310	100	10	10%
311	100	10	10%
312	100	10	10%
313	100	10	10%
314	100	10	10%
315	100	10	10%
316	100	10	10%
317	100	10	10%
318	100	10	10%
319	100	10	10%
320	100	10	10%
321	100	10	10%
322	100	10	10%
323	100	10	10%
324	100	10	10%
325	100	10	10%
326	100	10	10%
327	100	10	10%
328	100	10	10%
329	100	10	10%
330	100	10	10%
331	100	10	10%
332	100	10	10%
333	100	10	10%
334	100	10	10%
335	100	10	10%
336	100	10	10%
337	100	10	10%
338	100	10	10%
339	100	10	10%
340	100	10	10%
341	100	10	10%
342	100	10	10%
343	100	10	10%
344	100	10	10%
345	100	10	10%
346	100	10	10%
347	100	10	10%
348	100	10	10%
349	100	10	10%
350	100	10	10%
351	100	10	10%
352	100	10	10%
353	100	10	10%
354	100	10	10%
355	100	10	10%
356	100	10	10%
357	100	10	10%
358	100	10	10%
359	100	10	10%
360	100	10	10%
361	100	10	10%
362	100	10	10%
363	100	10	10%
364	100	10	10%
365	100	10	10%
366	100	10	10%
367	100	10	10%
368	100	10	10%
369	100	10	10%
370	100	10	10%
371	100	10	10%
372	100	10	10%
373	100	10	10%
374	100	10	10%
375	100	10	10%
376	100	10	10%
377	100	10	10%
378	100	10	10%
379	100	10	10%
380	100	10	10%
381	100	10	10%
382	100	10	10%
383	100	10	10%
384	100	10	10%
385	100	10	10%
386	100	10	10%
387	100	10	10%
388	100	10	10%
389	100	10	10%
390	100	10	10%
391	100	10	10%
392	100	10	10%
393	100	10	10%
394	100	10	10%
395	100	10	10%
396	100	10	10%
397	100	10	10%
398	100	10	10%
399	100	10	10%
400	100	10	10%
401	100	10	10%
402	100	10	10%
403	100	10	10%
404	100	10	10%
405	100	10	10

In list weighting, all extracted lists are weighted and thus some unimportant or noisy lists, Such as the price list that occasionally occurs in a page can be assigned by low weights. Some of the extracted lists are not informative or even useless. There are two weighted list ecommerce1 and ecommerce2.

7) List Clustering

Screenshot 28: List Clustering

In list clustering, similar lists are grouped together to compose a facet. For example, different lists about watch gender types are grouped because they share the same items “men’s” and “women’s”. There are two clustered list one is of ecommerce website1 and second is of ecommerce website2.

8) Data Ranking

Screenshot 29: Data Ranking

In data ranking, facets and their items are evaluated and ranked . For example, the facet on brands is ranked higher than the facet on colours based on how frequent the facets occur and how relevant the supporting documents are.

III. ACKNOWLEDGEMENT

It is matter of great pleasure by getting the opportunity of highlighting a fraction of knowledge I acquired during my technical education through this dissertation. The making of the project needed co-operation and guidance of a number of people. I therefore consider it my prime duty to thank my guid, co-guid and principal.

IV. CONCLUSION

We study the problem of finding query facets. We propose a systematic solution which we refer to as QDMiner to automatically mine query facets by aggregating frequent lists from free text, HTML tags and repeat regions within top search results. the first approach of finding query facets, QDMiner can be improved in many aspects. We further analyze the problem of duplicated lists, and find that facets can be improved by modeling fine-grained similarities between lists within a facet by comparing their similarities.

REFERENCES

- [1] O.Ben-Yitzhak, N.Golbandi, N.Har’El, R. Lempel, A.Neumann, S.Ofek-Koifman , D.Sheinwald, E.Shekita, B.Sznajder and S.Yogev, “Beyond basic faceted search,” in Proceedings of WSDM ’08, 2008.
- [2] M. Diao, S. Mukherjea, N.Rajput, and K.Srivastava, “Faceted search and Browsing of audio content on spoken web,” in Proceedings of CIKM ’10, 2010.
- [3] D.Dash, J.Rao, N.Megiddo, A. Ailamaki, and G.Lohman “Dynamic faceted search for discovery-driven analysis,” in CIKM ’08, 2008.
- [4] W. Kong and J. Allan, “Extending faceted search to the general web,” in Proceedings of CIKM ’14, ser. CIKM ’14. New York,NY, USA: ACM, 2014, pp. 839–848.
- [5] T. Cheng, X. Yan, and K. C.-C. Chang, “Supporting entity search: a large-scale prototype search engine,” in Proceedings of SIGMOD ’07, 2007, pp. 1144–1146.
- [6] K. Balog, E. Meij, and M. de Rijke, “Entity search: building bridges between two worlds,” in Proceedings of SEMSEARCH’10, 2010, pp. 9:1–9:5.

10.22214/IJRASET

45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)