Sreyas International Journal of Scientists and Technocrats, Vol.3 (3) 2019, pp.21 -24

Virtual Query Facets using Knowledge Bases

A Swathi[1], Akarapu Nikhil kumar[2] and S V S R Y Koundinya[2]

[1] Computer Science, Sreyas Institute of Engineering and Technology, Hyderabad, India
[2] Computer Science, Sreyas Institute of Engineering and Technology, Hyderabad, India
Received- 20 Mar 2019, Revised- 18 Apr 2019, Accepted- 15 May 2019, Published- 15 June 2019

ABSTRACT

A query facet is a significant list of information that explains an underlying aspect of a query. Existing algorithms on mining facets depends on extracting frequent lists contained in top search results. The coverage of facets and facet items mined by this kind of methods might be limited, because only a small number of search results are used. In order to solve this problem, we propose the algorithm on mining query facets by using knowledge bases which contain high-quality structured data. Second, we mine initial query facets from search results, then expanding them by finding similar entities from Freebase. Experimental results shown that our proposed method significantly improved the coverage of facet items over the state-of-the-art algorithms.

Keywords: Query, knowledge bases, facet

1. FACET GENERATION

Given a query Q issued to QDMKB, we try to find similar entities in Freebase and create properties of these entities. These properties are inherent summarizations of corresponding entity in different aspects thus can be used directly as facet candidates. We will use Freebase Search API to retrieve similar entities and use Topic API to retrieve properties. Kong et al. [1] proposed two supervised methods, namely QF-I and QF-J, to mine query facets from the results.

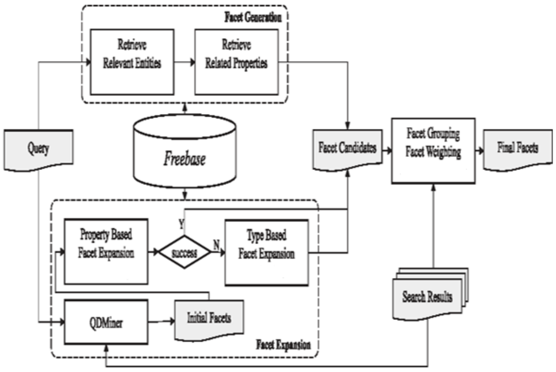

Fig.1 Architecture

1.1 Retrieve Relevant Entities

We use Freebase Search API 3 to retrieve entities for an input query string. Based on the introduction to the Search API, retrieved Freebase entities are ranked by a relevant score that is a function of its inbound and outbound link counts in Freebase. We can retrieve similar entities in top search results for most queries. In this research, we just want exactly matched entities in most cases.

*Corresponding author.
Email address: swathi@sreyas.ac.in (A Swathi)
Double blind peer review under responsibility of SERAYAS Publications
https://dx.doi.org/10.24951/sreyasijst.org/2019031005

1.2 Retrieving Similar Properties

Each Freebase entity belongs to different types, and each type contains different properties. For instance, entity Steven Spielberg belongs to types Film Director, TV Producer, and etc. Type Film Director includes properties such as Films Directed which has targets entities like Schindler's List.We call these properties and their target entities as direct properties which could be retrieved by Topic API.

2. FACET EXPANSION

Given a query r and its top search results D(r), assume F(r) are initial query facets extracted by QDMiner. Recall that QDMiner is a system that can automatically mine query facets by extracting and clustering frequent lists from top search results. QDMiner may suffer from the limitation of the search results. The well-known examples of web-based facet expansion systems are Google Sets [3] We propose further expanding the facets to find more peer items by leveraging a knowledge base. Assuming that F(r) is an initial facet generated by QDMiner.

2.1 Facet Expansion based on Properties

We obtain all properties P(r) for query r. Note that these properties include the direct properties, the merged second hop properties, and the separate second-hop properties. Initial facet f 2 F(r) , we try to find the most suitable property p 2 P(r) that can cover most items in f.

2.2 Facet Expansion based on Types

If P(r) is empty or no matched property is found, the method introduces fails. In this section, we propose directly finding similar items for a facet without the premise that query matches entities in Freebase. More specifically, we try to find a formalized query that can be used to retrieve similar entities. We use two existing datasets that are used by QDMiner [2], namely UserQ and RandQ, to evaluate the proposed method. Experimental results show that our proposed method QDMKB significantly outperforms all state-of-the art methods including QDMiner, QF-I, and QF-J in terms of rp-nDCG.finally, we combine the type and constraints to make up a single MQL query to retrieve more candidates from Freebase.

3. FACET GROUPING

After facet generation and expansion, we collect a set of facet candidates which consist of new facet candidates generated from Freebase facet candidates expanded by the property based and type based methods. There might be duplicate items within these facet candidates; hence we need to further group these facets. We use the QT algorithm (Quality Threshold) to cluster facet candidates [4]. QT is a clustering algorithm that groups data into high quality clusters. Compared to other clustering algorithms, diameters don't exceed a user-defined diameter threshold.

4. Facet Weighting

Finally, we sort all final facets by weights to form output of QDMKB. The reason why we use search results instead of knowledge bases to weight and rank each facet is that search results reflect users' attention and intent. While by search results, we could confirm which one appears more frequently to justify its importance.

5. Test Cases

Test Case	Description	Input	Expected Output	Actual Output	Status
Case 1.1	Admin login	1:Username 2:incorrect Password	Incorrect username or password.	Login failed.	Fail
Case 1.2	Admin login	1:Username 2:Password	Login successful with username and password.	Login successful.	Pass
Case 2	Admin Home Page	Select the option to proceed	Redirects to the selected page.	Displays desired page.	Pass
Case 3.1	Researcher registration	Password with spaces.	Spaces are not allowed.	Accepting password with spaces.	Fail

Fig.2 Test cases for Virtual Query Facets using Knowledge Bases

Case 4	Add Products Page	Product description.	Displays products to select and add.	Product added	Pass
Case 5	Query search	Query	Results using the keyword.	Multiple results using multiple queries.	Fail
Case 6	User profile picture upload	Word document.	Successfully uploaded the profile picture.	Successfully uploaded	Fail

Fig.3 Test cases for Display products and Uploading picture

6.Conclusion

The facets here are generated using four steps. Then from these generated pattern lists are extracted using the list extraction algorithm. This method can extract high quality lists from the top k query search results. Hence these high quality lists can be used to generate meaningful facets.

7.Future scope

Facets extraction has witnessed a booming interest recently and has been undergoing research for the past few years. Focusing Query Facets Mining, the future work possible in this area includes the part-of-speech information can be used to check the homogeneity of lists and improve the quality of query facets.

8. References

[1]. "Extracting query facets from search results," in Proceedings of SIGIR '13, 2013, pp. 93–102.

[2]. Z. Dou, S. Hu, Y. Luo, R. Song, and J.-R.Wen, ―Finding dimensions for queries,‖ in Proceedings of CIKM'11, 2011, pp. 1311–1320.0.

[3]. L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring Expression Data: Identification and Analysis of Coexpressed Genes," Genome Research, vol. 9, no. 11, pp. 1106–1115, November 1999.

[4]. Sumati Pathak, Rohit Raja, Vaibhav Sharma, and K. Ramya Laxmi, A Framework Of ICT Implementation On Higher Educational Institution With Data Mining Approach, European Journal of Engineering Research and Science, ISSN (Online) : 2506-8016.

[5]. Sumati Pathak, Rohit Raja, Vaibhav Sharma The Impact of ICT in Higher Education. Published in IJRECE Vol. 7 Issue 1 January-March, 2019. ISSN: 2393-9028 (PRINT) ISSN: 2348-2281 (ONLINE) ISSN: 2393-9028 (PRINT). Vol 7, Issue 1, pp 1650-1656. (UGC Approved)

[6]. Rakesh Kumar Lenka, Amiya Kumar Rath, Zhiyuan Tan, Suraj Sharma, Deepak Puthal, N V R Simha, Rohit Raja, Shankar Sharan Tripathi, and Mukesh Prasad Building Scalable Cyber-Physical-Social Networking Infrastructure Using IoT and Low Power Sensors, IEEE Access, Vol. 6, Iss. 1, pp. 30162-30173, Print ISSN: 2169-3536, Online ISSN: 2169-3536, Digital Object Identifier: 10.1109/ACCESS.2018.2842760. (SCI Index)

[7]. Rohit Raja, Tilendra Shishir Sinha, Raj Kumar Patra and Shrikant Tiwari(2018), Physiological Trait Based Biometrical Authentication of Human-Face Using LGXP and ANN Techniques, Int. J. of Information and Computer Security, Vol. 10, Nos. 2/3, pp. 303- 320. (Scopus Index)

[8]. Rohit Raja, Md Rashid Mahmood, Raj Kumar Patra (2018) Study and Analysis of Different Pose Invariant for face Recognition Under Lighting Condition, Sreyas International Journal of Scientists and

Technocrats, Vol. 2(2), pp 11-17, 2018.

[9]. Ayenala Prithvi Raj, Rohit Raja, Suresh Akella (2017) A New Framework for Trustworthiness of Cloud Services, International Journal of Research, Volume 04 Issue-1 December 2017. e-ISSN: 2348-6848, p-ISSN: 2348-795X.

[10]. Rohit Raja, Raj kumar Patra, T. S. Sinha (2017), Extraction of Features from Dummy face for improving Biometrical Authentication of Human, International Journal of Luminescence and Application, ISSN:1 2277-6362, Vol. 7, No. 3-4, Oct Dec 2017, Article 259, pp. 507-512.

[11]. Rohit Raja, Tilendra Shishir Sinha, Ravi Prakash Dubey (2016), Soft Computing and LGXP Techniques for Ear Authentication using Progressive Switching Pattern, Published in International Journal of Engineering and Future Technology, Vol. 2, Iss. 2, pp.66-86, ISSN: 2455-6432.

[12]. Rohit Raja, Tilendra Shishir Sinha, Ravi Prakash Dubey (2016), Orientation Calculation of human Face Using Symbolic techniques and ANFIS, Published in International Journal of Engineering and Future Technology, Vol. 7, Iss.7, pp. 37-50, ISSN: 2455-6432.

[13]. Rehmat Khan, Rohit Raja (2016) Introducing L1- Sparse Representation Classification for facial expression, Published in Imperial Journal of Interdisciplinary Research (IJIR), Vol. 2, Iss. 4, pp. 115-122, ISSN: 2454-1362.

[14]. Nikita Rawat, Rohit Raja (2016), A Survey on Vehicle Tracking with Various Techniques", International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), Vol. 5 Iss. 2, pp. 374-377, ISSN: 2278-1323.

[15]. Nikita Rawat, Rohit Raja (2016), Moving Vehicle Detection and Tracking using Modified Mean Shift Method and Kalman Filter and Research, International Journal of New Technology and Research (IJNTR), Vol. 2, Iss. 5, pp. 96-100, ISSN: 2454-4116.

[16]. Shraddha Shukla, and Rohit Raja (2016) A Survey on Fusion on Color Images, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 6, pp. 1707-1709, ISSN: 2278-1323.

[17]. Shraddha Shukla and Rohit Raja (2016) Digital Image Fusion using Adaptive Neuro-Fuzzy Inference System, International Journal of New Technology and Research (IJNTR), Vol. 2, Iss. 5, pp. 101-104, ISSN:-2454-4116.

[18]. Rohit Raja, Tilendra Shishir Sinha, Ravi Prakash Dubey (2015), Recognition of human-face from side-view using progressive switching pattern and soft-computing technique, Association for the Advancement of Modelling and Simulation Techniques in Enterprises, Advance B, Vol. 58, N 1, pp. 14-34, ISSN:-1240-4543. (Scopus Index)