

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329492737>

Product Rank Based Search Engine for E-Commerce

Conference Paper · April 2018

DOI: 10.1109/I2CT.2018.8529503

CITATIONS

2

READS

920

3 authors, including:



Urvesh Rathod

National Institute of Technology Karnataka

4 PUBLICATIONS 2 CITATIONS

[SEE PROFILE](#)



Aruna Pavate

St Franics Institute of Technology

45 PUBLICATIONS 34 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Connecting Researchers on the Globe [View project](#)



Online and Distance Learning [View project](#)

Product Rank Based Search Engine for E-Commerce

Unification of E-Commerce

Urvesh Rathod

Computer Engineering
Atharva College of Engineering
Mumbai, Maharashtra
urveshrathod25@gmail.com

Prof. Aruna Pavate

Computer Engineering
Atharva College of Engineering
Mumbai, Maharashtra
arunapavate@gmail.com

Vaibhav Patil

Computer Engineering
Atharva College of Engineering
Mumbai, Maharashtra
Vappatil88@gmail.com

Abstract— In today's world, e-commerce is the strongest and the fastest growing platform. People inclined to buy products from websites now and then. Up to the present time price comparisons of the same product on different websites are available but now, moving a step ahead, our proposed system tends to serve as a smart web spider for all e-commerce websites, which crawls the whole website for you and find out the cheapest yet best quality products of your choice in real time. Moreover, this assures the union of all e-commerce website in form of one rigorously filtered system. It is also capable of making good recommendations as per customer's requirement and serves as an active agent in the system. This product can also prove beneficial for e-commerce companies because it will be able to track down dead pages, Seller's wrongly entered price or any kind of glitches/errors/fraudulent activities on their websites in real time.

Keywords: Web Search Engine, Web Crawlers, PageRank Algorithm, Product Rank Algorithm.

I. INTRODUCTION

E-commerce in India is the fastest growing platform where nearly more than thousands of products are purchased in every second. Indian e-commerce is growing at an annual rate of 51%, the highest in the world, and is expected to jump from \$30 billion in 2016 to \$120 billion in 2020. [1] Companies like Amazon, Flipkart, Paytm and Snapdeal makes billions of profits every year and most of them are during festive seasons. At each website, there are millions of products are available and customers are not able to find products which are really offered at cheap prices and with more discount. Even filters fail to sort the products from low-high rate due to large and dynamic set of database. Therefore, it is necessary to filter out unnecessary and overpriced products and provide customers with few quality products at discounted rates available on websites during sales and festive seasons.

The proposed system merges all the products available at different websites and rigorously filters out the best product for the customer every hour. Using highly focused web crawlers, only those products are available to a customer who comes under predefined parameters like price, brand, product discount and popularity of the product. This narrows down the product listing from million too few hundreds at each website.

Since our system unifies e-commerce websites, the overall result will always be under thousand products giving customer lots of option to choose the product amongst them.

II. REVIEW OF LITERATURE

In this section, we discuss some areas where such research needs to be extended to work better on the web.

Customers prefer online shopping rather than offline shopping because of better price and product availability. Due to an increase in healthy competition in e-commerce market, it is also important to know which e-commerce website is most ideal, in terms of quality and price, amongst all. [1][2]

Products on E-commerce websites are growing rapidly day by day. Many comparison websites are available on the internet which compares the price of a product with different websites. Most of them are operated manually thus it becomes very tedious to update each product manually. Thus it is very much important to have a robust automated search engine available in which products can be compared with an automated engine with minimal human intervention. [3]

Google uses PageRank Algorithm to rank web pages based on the graph of the web. It helps discriminate among several web pages available and provides the best and highly relevant result in return. [4]

Web data mining is a very fast developing area. Using data mining, system will be able to determine number of visits and purchases made for the product listed on e-commerce websites via analyzing the statistics report of our proposed system. It will be able to keep track of trends and popularity of products which will help in providing multi-level filtering. [5]

In order to get maximum efficiency, it is necessary to trim millions of products on websites carefully. In order to achieve this, our system needs to train web-crawlers continuously and make them highly efficient. The parameters chosen for our web crawlers will decide the overall competence of our system. Few key parameters on which a product will depend are price of the product, brand and popularity of the product. This will make our web-crawlers more focused and highly efficient. These parameters are dynamic in nature so our automated

engine will make sure that web-crawlers remain highly focused. [6] [12]

Price prediction is important now-a-days to improve consumers experience on any e-commerce website. It will help customer to decide the best time to buy that product. Using SVM Algorithm, it is possible to plot graph and predict the price of the product with minimal error for few months. [7]

Dr. Daya Gupta, Devika Singh, proposed a new ranking algorithm named as User Preference Based, Page Rank Algorithm (UPBR) which employs Web Usage, Web Structure and Web Content Mining techniques to order the web pages with the help of agents and specialized crawlers. The UPBR is compared with other ranking algorithm with advantages, shortcomings, methodologies used as well as input parameters used. [13]

The new method an Intelligent Search Method (ISM) was proposed by Anish Jain. This research is based on, to index the web pages using an intelligent search strategy. It is essential to infer the meaning of the search query and then index the web pages based on the analysis. This method can be integrated with any of the Page Ranking Algorithms to produce improved and relevant search results.[14]

In order to reduce the problem of author had proposed a hybrid approach of Enhanced –Ratio-Rank and page level keyword algorithm. This model considers the in links and out links of pages and keywords of each page based on that calculates the page rank [15]

Anish Gupta, described different approaches of Focused Crawling classified according to their methodology or dependency on determining appropriate pages to: Priority based focused crawler, learning based crawler, Structure based focused crawler, context based crawler, and other focused crawler approaches.[17]

As per the above study, future work we need to design the new page ranking algorithms that meets out global standards of ranking challenges efficiently and is able to provide both user specific as well as relevant information easily and quickly.

III. INSIGHTS OF SYSTEM

This study focuses on saving time and money of the customers and searching for the best product on behalf of them. Highly trained focused web crawlers will scrape out products every hour from all the listed websites. Minimum priced product will be selected as a threshold value and ranking will be given in descending order. Higher the ranking, lower the price, better the deal and will be available first to the user. The system filters each product every time at multiple levels to get the best amongst millions of products.

The projected system will give best outcomes in no time as well as keep track of the price for each product listed and help in predicting the future price of the product enabling customer

to decide when to purchase the product. Customers will be able to track previous price of the product and future price of the product which will help them decide the nature of the product.

The main motive behind this work is, there will be a unified and centralized system where customers can find products with ideal price which can help improve shopping experience of the consumers so that they don't need to search for a product back and forth at different websites.

IV. PROPOSED SYSTEM

The existing system is unable to display products available at huge discounts in real time unless ecommerce websites promotes about the discounts to customers. Filtering out products on the existing system does not guarantee availability of product at best price. An unusual behavior in topic drift that often gets influenced by the unreasonable relationships between searched product and recommended product. Nonexistence of prediction of product price is also a disadvantage. Moreover, sometimes due to technical glitches/wrong entered price by sellers, token of Appreciation/Compensation is provided to customers for the misconception/glitches occurred, resulting in intangible loss to the company.

The proposed system is a Real-Time Search Engine where users will find best products among multiple e-commerce websites available in India. It will reduce time and money wasted in manual filtering and will provide customers a better shopping experience. Moreover, it will also help e-commerce companies to identify price errors on their website and help provide better service to customer. It will be automated, user-friendly and easy to use.

In order to achieve maximum efficiency system uses highly efficient and focused dynamic web crawlers which will filter out products at first level based on the predefined set of parameters integrated into the system. The scraped products will pass through Product Rank Algorithm where each product will be ranked accordingly. Filtration process during these stages will be highly rigorous and this will result in availability of best products to the customers. [7][9]

Customers being unaware of product discounts available everyday on the e-commerce websites and have to wait till festive season to buy the desired product. With the help of this system customers can perceive highly discounted or determined discounted products of any category. This system proves to save a lot of time and money of the users and narrow down the possibilities of product occurrences/glitches in just a few clicks. Henceforth, customers won't wait for any flash sale or discounted sale at actual season instead they can buy them whenever they like by spotting at one place.

This system is similar to any of the exiting e-commerce system where user will search for the product and will get desired product within few seconds on the screen. As

soon as user enters the search parameters highly focused web crawlers will perform web scraping and start filtration process.

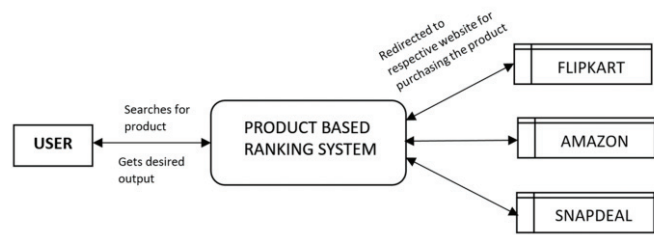


Fig. 1.1: Product based ranking system

Fig. 1.1 shows the basic overview of system. Users will search for products. The products will be scraped and processed in system every hour to get updated results. When users decide to purchase the product control will be redirected to the respective website for purchasing the product.

Scraping out millions of products from different websites will occur only once every hour for an individual search entry. The following cases system will encounter:

Case 1: Product is searched for the first time.

Suppose n users are trying to search n different products at the same time then lots of computational power is needed for operation which is practically impossible so before user searching for any product system will go under maintenance every day for few hours and store the entries of each product in database.

Case 2: If the product is already in the database.

If the product is already available in the database, web crawlers will only update price of the product after every hour which takes no time and will display it to the user. Since the scraped products are in few thousand the amount of work done by web crawlers is comparatively less and achievable.

After scraping out the necessary products is to be filtered using Product Rank Algorithm as shown in Figure 1.2. Product Rank Algorithm starts ranking products according to the popularity of the product and lowest price of the product. It works in the same way as Page Rank Algorithm works in Google search engine to rank pages [4]. Here the only difference, this system ranks each product based on how popular the product is on that day and what is the price of that product instead of web pages. This is the second stage filtration processes in which products having higher ranking displayed first.

As shown in Fig. 1.2 once user has given the desired input for search, web crawlers start crawling websites to collect details of the product. Thereafter the search results are stored into database in order to process it into product rank algorithm. Finally, after computation, filtered products will be displayed to the user.

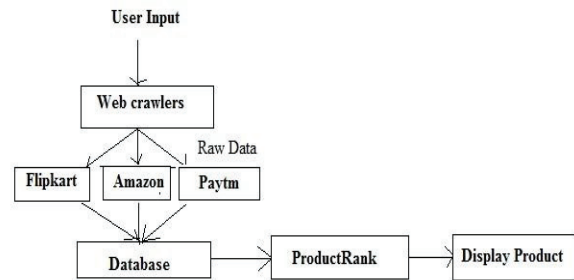


Fig. 1.2: Input processing inside system

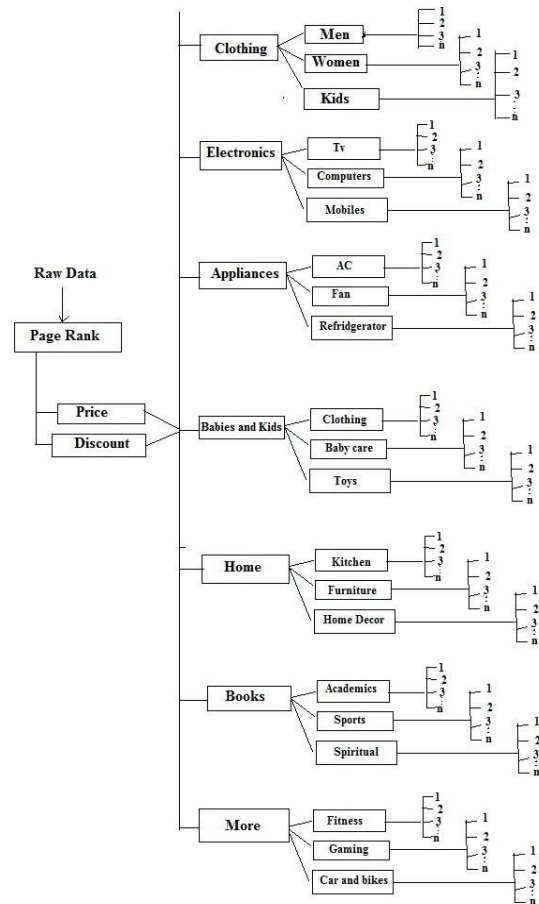


Fig. 1.3: Product Rank algorithm on different products

Fig. 1.3 shows working of product Rank algorithm. Here price and discount are considered as key parameters in filtration process. 1-n in the graph are leaf nodes (products) and internal nodes are their associated categories. For a particular search there are “n” different products available and algorithm places those products individually into each subcategory. Let us assume user is searching for Men’s t-shirt then t-shirt will be stored into database under category clothing ->Men’s -> T-shirts. Products will be sorted in ascending order w.r.t price

and ranks will be given to them. Products with highest rank will be displayed first.

Products are listed as leaf nodes and corresponding categories are internal nodes. When the user enters any product to search, Product Rank Algorithm starts ranking the products. This system checks whether the scraped product is relevant to searched category or not. Consider n products after web scraping, each product has some parameters associated with it like price, discount, etc. each product will be passed concurrently through the graph seen in Figure 1.3 to check which product falls under which category. Product with the least price, best brand, and highest popularity will be ranked from top to bottom. Figure 1.4 shows the working flow of the proposed system.

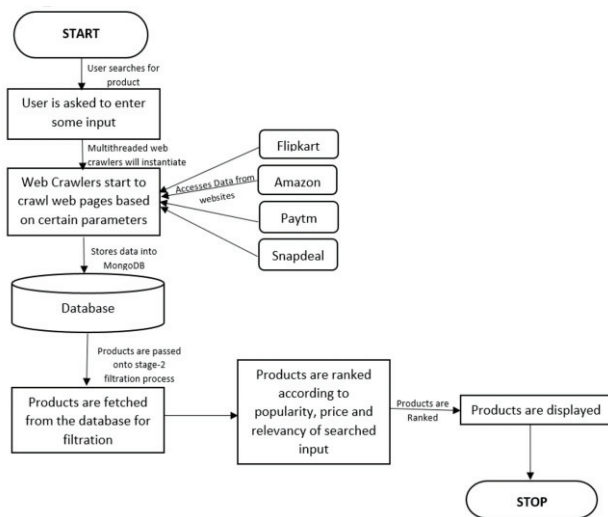


Fig. 1.4 working flow of the system

- Step 1: User enters input into the system that needs to be searched.
- Step 2: Multithreaded web crawler activated and in parallel product searched on e-commerce websites taking user's input as key parameter. This is Level-1 filtration process where the entire unnecessary and irrelevant product won't be considered.
- Step 3: Web Crawlers stores all the details of the selected products into the database.
- Step 4: All the scraped products passed through Product Rank algorithm (Level-2) where ranking to the product starts.
- Step 5: If the product comes under parameters pre-defined as then the rank of the product increases else decreases.
- Step 6: According to the product rank items listed in decreasing order.
- Step 7: After collecting all the products with its ranking displayed to the user.

V. CONCLUSION

Due to the increase of numerous products in e-commerce, customers are unable to find proper product or in many of the cases customers buy overpriced product which is non-beneficial. In order to increase customers experience at e-commerce, filtration of unnecessary products is needed and listings of quality products are needed. This system will provide customers the crystal clear view of the searched product throughout the year which help customers to decide whether products are really at discounted rate/on sale or not. This system will be a revolt in e-commerce and will definitely led an increase in sales report.

VI. FUTURE SCOPE

With the influx of big data era, the facts on web have become much more assorted and dynamic, leading to better-quality challenges in Ranking. Therefore, as a future work, we need to design the new page ranking algorithms that meet our global standards of ranking challenges efficiently and are able to provide both user-specific as well as relevant information easily and quickly

In future, we would like to implement this approach and then compare our proposed model with different algorithms in order to analyze and evaluate it with respect to time and accuracy.

REFERENCES

- [1]. <https://www.export.gov/article?id=India-e-Commerce> 08th August 2017
- [2]. <https://www.ibef.org/industry/ecommerce.aspx> 07th December 2017
- [3]. Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine". Computer Science Department, Stanford University, Stanford, CA 94305, USA, pp.1-20 1999.
- [4]. Page Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry, "The PageRank Citation Ranking: Bringing Order to the Web", Tech. Rep., pp.1-17, 1998.
- [5]. Yonghua Zhao, "WEB data mining applications in e-commerce".The 9th International Conference on Computer Science & Education (ICCSE 2014)), Department of Computer Science and Technology, Wuhan University of Technology, China, pp.557-559,2014
- [6]. Anish Gupta, "Survey on topic-focused Web Crawler" 2015 1st International Conference on The futuristic trend in Computational Analysis.Asst.professor, Noida, pp.1-4,2015
- [7]. Dai Wei , PengGeng, Liu Ying (Univ. of Chinese Acad. of Sci., Beijing, China) "A prediction study on e-commerce sales based on structure time series model and web search data". 2014 26th Chinese Control and Decision Conference (CCDC), Pp.5346- 5351, 2014.
- [8]. Buyhatke- <https://compare.buyhatke.com/> 29 January 2017
- [9]. A. Gelbukh , "Natural language processing", 2013 International Conference on Advances in Computing, Communications, and

- Informatics (ICACCI), National Polytech. Inst., Mexico, IEEE pp.2112-2121,2013
- [10]. Sanjay Kumar Malik, SAM Rizvi, “ Information Extraction Using Web Usage Mining, Web Scraping and Semantic Annotation” 2011 International Conference on Computational Intelligence and Communication Systems. pp.465-469, 2011.
- [11]. David Hawking, ”Web search engines part 1 and part 2”, 2015 International Conference on Advances in Computing, Communications, and Informatics (ICACCI), a scientist at CSIRO ICT Centre, Canberra, Australia, ICACCI, pp.86-90, 2015.
- [12].S. Chakrabarti, M. van den Berg, and B. Dom, ” Focused crawling: a new approach to topic-specific web resource discovery”, 8th World Wide Web Conference, Toronto, Canada, Computer Networks (Amsterdam, Netherlands: 1999), pp.1623-1640, 1999.
- [13].Dr. Daya Gupta, Devika Singh, ”User Preference Based Page Ranking Algorithm”. International Conference on Computing, Communication, and Automation (ICCCA2016), Department of Computer Science and Engineering, Delhi, ICCCA, pp.166-171, 2016
- [14].Ashish Jain, Rajeev Sharma, Gireesh Dixit, Varsha Tomar, ” Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages”, International Conference on Communication Systems and Network Technologies, 978-0-7695-4958-3/13 IEEE DOI 10.1109/CSNT.2013.137 pp.640-645, 2013.
- [15].Lissa Rodrigues, Shree Jaswal, “Hybrid Model for Improvised Page Ranking”. 2015, International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp.466-469. 2015.
- [16].Upul Senanayake, Peter Szot, Mahendra Piraveenan, Dharshana Kasthurirathna. “The Performance of Page Rank Algorithm under Degree Preserving Perturbations”. Centre for Complex Systems Research, Faculty of Engineering and IT, Australia, pp.1-4. 2006.
- [17]. Anish Gupta, Priya ,” FOCUSED WEB CRAWLERS AND ITS APPROACHES “, 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management IEEE (ABLAZE-2015) pp 619-622. 2015