Journal of King Saud University – Computer and Information Sciences xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics

Dheeraj Malhotra *, O.P. Rishi

Department of Computer Science and Informatics, University of Kota, Kota, Rajasthan 324 005, India

ARTICLE INFO

Article history:
Received 18 October 2017
Revised 22 February 2018
Accepted 28 February 2018
Available online xxxx

Keywords:
E-Commerce website ranking
IMSS- AE tool
RV page ranking algorithm
Second generation big data analytics
Hadoop-MapReduce
Personalized page ranking

ABSTRACT

The purpose of this research work is to explore various limitations of conventional search and page ranking systems in an E-Commerce environment. The key objective is to assist customers in making an online purchase decision by providing personalized page ranking order of E-Commerce web links in response to E-Commerce query by analyzing the customer preferences and browsing behavior. This research work first employs an orderly and category wise literature review. The findings reveal that conventional search systems have not evolved to support big data analysis as required by modern E-Commerce environment. This work aims to develop and implement second-generation HDFS- MapReduce based innovative page ranking algorithm, i.e. Relevancy Vector (RV) algorithm. This research equips the customer with a robust metasearch tool, i.e. IMSS-AE to easily understand personalized search requirements and purchase preferences of customer. The proposed approach can well satisfy all critical parameters such as scalability, partial failure support, extensibility as expected from next-generation big data processing systems. An extensive and comprehensive experimental evaluation shows the efficiency and effectiveness of proposed RV page ranking algorithm and IMSS-AE tool over and above other popular search engines.

© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In this modern era of big data, the shopping activity is modified a lot because of enormous growth in online shopping websites, also known as E-Tailers. The new age customers prefer to shop through these online portals because of various attractions in countries like India such as easy and cheap availability of the Internet. The primary reason is intense competition between telecoms, for instance, *Reliance Jio* prime membership offers free unlimited internet data usage for three months for all of its users in nominal charges. Some of the other reasons include lucrative cash back and easy returns without deduction of shipping charges from portals like *PayTm*, Cash on delivery type regular features from E-Commerce sites like *Flipkart*, *Amazon*, and other E-Tailers. Moreover, searching a suitable E-Commerce website to best suit the customer purchase requirements is not so easy as customers are primarily dependent on conventional search engines like *Google*, *Bing* to find a suitable E-Commerce web site. However, when different users search the same E-Commerce query, even a most advanced and popular search engine retrieves the same result as discussed by Gomez-Nieto et al. (2014). Thus, irrespective of the background and personalized tastes of customer submitting the query as most of the modern search engines tend to return the results by interpreting the E-Commerce query in various possible ways. Moreover, if the query is ambiguous or incomplete, then the situation will get even worse as discussed by Malhotra and Verma (2013). For instance, for the incomplete E-Commerce search query *"Galaxy"*, some customers may be interested in links to buy a new *Samsung Galaxy* series mobile phone, while another customer may be interested in searching links for online booking of tickets for a movie *Guardians of the Galaxy Vol. 2*. Hence, there is an urgent need for personalized E-Commerce search system. The personalized system may modify the E-Commerce search query by keeping track of customer's preferences by maintaining his/her profile, search preferences through browsing history, etc. over a period and return results in correct order of ranking with customer's relevant output links on top to best suit the customer requirements (See Fig. 1).

E-Commerce data is explosively increasing on the scale of TeraBytes (TB) to PetaBytes (PB) on a daily basis due to the continuous

* Corresponding author.
E-mail address: dheerajmalhotra4@gmail.com (D. Malhotra).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2018.02.015
1319-1578/© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

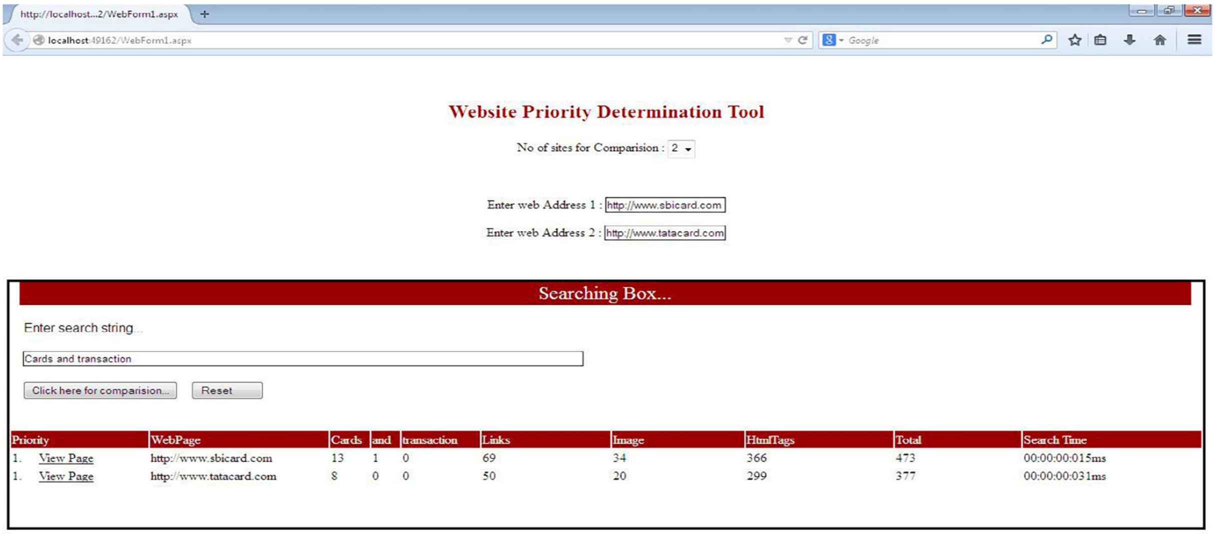

Fig. 1. Website Priority Determination Tool by Verma et al., (2015).

increase in WWW traffic. For instance, to purchase an item on the web, a customer may explore many websites to have satisfactory E-Commerce transaction which not only provides the high quality branded product but also at best possible discounted price or maximum wallet cash back. Hence, as a result, many of the online shopping portals getting big data on daily basis like *Amazon* or *PayTm Mall* android based E-Commerce portal, which handles around a million customer transaction logs on a regular basis, resulting into many TB of data generated on a daily basis. This excessive online generated data is commonly known as '*Big Data*' with emphasis on high values of various popular V's, i.e., Value, Velocity, Variety, Veracity, and Volume. Big data may be defined as a collection of a huge number of data sets, the speed of incoming data before processing, outgoing data after processing and range of data sources are beyond the capabilities of conventional relational databases systems for processing and management. Verma and Singh (2017a,b) proposed that '*Big Data*' consist of many useful patterns in the form of association rules which are never extracted and hence advanced big data analytics is required to explore these hidden patterns. These patterns are helpful for E-Commerce websites. E-Tailer may utilize such patterns for market basket analysis and hence to increase sales by extracting customer's favorite purchase patterns, efficient and easy inventory management to avoid situations like overstock or out of stock by identifying significant purchase trends for a specific product from various sources such as social media trend analysis. An online merchandise seller may use a big data analytics tool to analyze multiple posts on social media like *Instagram, Facebook*. The images of popular celebrities that are most shared / most liked recently to determine the latest fashion of dress material and hence can order more stock for similar dresses to quickly satisfy the increased demand of the market.

Market basket analysis using big data analytics for ranking of E-Commerce websites can be easily accomplished by using –RV-Map-Reduce framework, which is robust and scalable and is an open source platform for efficient processing of E-Commerce based big data. Hadoop cluster is characterized by some parallel machines that can easily store and process big data sets, a significant number of clients may easily submit their processes to distributed Hadoop cluster from different locations. Map -Reduce is a simplified programming model that can be used to process big data in Hadoop cluster with the help of primary functions known as Map and Reduce to process big data in (Key, Value) pair format. Hadoop and Map - Reduce based cloud computing framework may be used for efficient deployment of big data based advanced E-Commerce website ranking system.

The overall objective of this research work is to assist the customer in easy searching and correctly ranking E-Commerce websites to buy genuinely priced authentic products as well as to help E-Tailers in optimizing the structure of their websites to take advantage over competitors.

2. Literature review

Advance adaptive E-Commerce search is a personalized search for retrieval and ranking of relevant E-commerce websites by using intelligent technologies like semantic web, neural networks. The personalized search mechanism requires big data analytics to retrieve useful association rules from data in text, images or videos format as available on social media and purchase history of various customers to retrieve customer specific E-Commerce website ranking patterns efficiently. There are different types of traditional personalized search systems as discussed in the literature.

2.1. Review of hyperlink based adaptive search methods

In general, E-Commerce applications employ hyperlink personalization to assist the client by recommending E-Commerce websites that are more relevant as determined by feedback obtained through his/her buying history and explicit or implicit ratings. It is assumed that consumers who gave similar ratings to related products have similar preferences and accordingly algorithm recommend various website links to the users that are most popular in the similar category as determined by previous customers. E-Commerce portals/websites like *Paytm Mall, Myntra* uses hyperlink personalization to aid their clients in searching, ranking and purchasing appropriate products. Aoki et al., (2015) discussed the architecture of Web index (WIX) system for hyperlink generation that can be used to insert links to web pages by replacement of keywords as per customer's choice. This, in turn, will reduce his/her load to go through all the web links produced in search engine result. However, if multiple web links can be associated with a keyword, then relevance computation takes time which is the major limitation of the proposed system. Alam and Sadaf (2015) discussed that fetching the significant information from WWW is moderately difficult. The modern search engines may return quite a huge number of web pages in response to user's query, and the result becomes unmanageable and irrelevant if the query is erroneous as general purpose search engines retrieve documents consequent to all probable meanings of a query. They discussed heuristic search mechanism to extract a group of the pages to assist a user in locating her / his required information effortlessly. They worked out significant cluster label from the title of various documents sharing similar hyperlinks by using Apriori algorithm. However, the usefulness of proposed method using only title information is not guaranteed on heterogeneous data sets. Verma et al., (2015) developed SNEC page ranking algorithm based on various intelligent technologies like artificial neural networks and semantic web. In this paper, we had discussed website priority tool for easy evaluation of E-Commerce search queries and to obtain the relevant ranking of E-Commerce websites. The proposed tool may be used to get E-Commerce website correct ranking concerning its competitor websites efficiently. However, as discussed in future work of this paper, we are going to incorporate various capabilities in our presently proposed IMSS- AE tool such as page loading speed, image-based search, security comparison to rank the E-Commerce websites as required by modern day customer. Hence, proposed algorithm and tool in present research work is an improvement of previously published SNEC algorithm and website priority determination tool.

2.2. Review of search methods based on content personalization

Content personalization on WWW refers to the process of showing different content to different customers on same portal/website. Sugiyama et al., (2004) explained that sites like *Yahoo* present the relevant information to users in which they are probably more interested. Users/customers may specify the tabs of his/her choice on such websites that may include Bollywood/Hollywood movies, news, fashion updates, forecasting- sun sign/weather. Users may build their favorite page outline as per their requirement on content personalized portals. However, such systems usually suffer from various limitations like constant effort from the user is requisite as such systems are directly reliant on user inputs. Moreover, these portals cannot automatically adapt to changing needs of the user unless he/she explicitly modify his previously registered preferences. Kuppusamy and Aghila (2014) proposed all-purpose CaSePer, an adaptive website change detection architecture to help the users who frequently browse a specific website and are concerned in knowing the most recent changes rather than considering the complete content of the websites on repeated visits. This model requires being adapted as a custom-made personal search system. Moreover, the experimental efficiency of such a search system is required to be evaluated.

2.3. Review of search methods based on recommender system

In this current age of big data, there is an emerged need of recommender system to deal with information explosion on the web. Wasid and Kant (2015) discussed that recommender systems might help users by suggesting entertainment material like games, shopping deals to make efficient use of their usual search time on the web. They suggested a technique known as particle swarm optimization to determine priorities of various users and accordingly to present recommendations personalized for a specific user. They also suggested different filtering techniques usable by Recommender System, i.e., demographic filtering, collaborative filtering, content-based filtering and hybrid filtering techniques for web-based personalization. Adamopoulos (2014) proposed Probabilistic *Neighborhood* approach to conquer the regular problems of development of K nearest neighbor's method. They discussed the concept of *unexpectedness* in popular recommender systems for easily satisfying the requirements of the user. Cacheda et al., (2011) suggested an efficient method for collaborative filtering based on differences between customers and products rather than based on their similarities. They suggested latest metrics, GPIM and GIM to calculate the accuracy of prediction for web personalization and unwanted biased prediction of recommendation system. They carried out a detailed comparison between various collaborative filtering algorithms to differentiate their strengths and weakness in diverse conditions. Guy et al., (2010), suggested that recommender system may be merged into search engines for implementation of personalized search. They also discussed that user experience is more important than the performance of recommender systems. Jung et al., (2004), discussed a prototype SERF developed for a university library. This system learns from the user regarding document relevancy corresponding to a search query. It motivates the customer to enter meaningful and non-ambiguous queries and then asks for explicit ratings of search results to measure the level up to which could system satisfy requirements of the user. However, the success of the proposed system depends on the fact that how easily it can compel the user to provide the ratings. Hence, extensive research is required for recommender systems utilization as a personalized search system.

2.4. Review of search methods based on contextual relevance feedback

The contextual systems use user's implicit and explicit data to develop a contextual knowledge base through gathering different customer contextual profiles. Limbu et al., (2006) suggested modification/expansion of queries to appropriately reveal the user's interest and hence to obtain contextually personalized search results. The proposed approach efficiently improves various search criteria like recall and precision by expanding the incomplete/ambiguous query appropriately using thesaurus approach and by adding meta keywords to search query respectively. Tanapaisankit et al., (2012) suggested a personalized search model, the Query in Context (QIC) which improves search query by including user preferences and hence ranking search results with context enrichment to cut down the number of contextually inaccurate search results. The proposed model can be implemented by allowing search query terms with multiple meanings to get weighted towards correct contexts. Vinay et al., (2005) compared three different types of contextual relevance based feedback algorithms by employing target testing procedure and experimentally established that the Bayesian algorithm is more efficient than RSJ and Rocchio algorithms. They also explored that modern search engines do not provide the option for Relevance Feedback and hence users are often dissatisfied with the returned results and are required to modify their query to obtain relevant results manually.

2.5. Review of intelligent technologies based search methods

Singh and Vélez (2014) discussed the model of a search engine Simha to competently search over different cloud platforms for unstructured and structured data using backend elastic search engine. They also reviewed the significance of cautiously designed processes such as Extraction, Transform and Load while indexing big data. Malhotra (2014) explored that huge size of web and SEO interference leads to difficulty in retrieving valuable information from the internet through search engines. However, an artificial neural network can be efficiently trained to provide better search results by implementing supervised learning. Zhang et al., (2012) discussed cloud-based semantic++ search framework to provide results from social networks. They explored the failure of general purpose search engines to establish the relationships between objects, people and web pages by various social networking portals such as *Facebook, Instagram, Twitter*. Wang et al., (2011) proposed a methodology for search engine optimization based on customer' feedback which may be implicit or explicit and artificial neural networks and hence their use in the implementation of unbiased website ranking model.

3. Motivation

The vast repository of data on the web may be termed as big data. In the present situation, it sometimes becomes quite difficult for a customer to search relevant E-Commerce website on the Internet easily. One of the commonly followed temporary measures is to use a popular search engine like *Google*. However, as discussed in the literature, none of the search engines can completely solve retrieval problem as no search engine can index entire information available on the web. Bo and Yang-Mei (2014) discussed that most of the conventional search engines suffer from various limitations such as incomplete indexing, low precision, SEO manipulated page rank, low recall. Moreover, a conventional search engine presents the same output consequent to the same query, despite current requirements or personalized preferences of customer submitting the query as discussed by Rasekh (2015). This approach is not suitable for customers with a different set of requirements. Let us take an example, a female or male customer searching for "Online purchase of Belt" on a conventional search engine. The customer will get the same rank of various listed web links in output without any consideration to the fact that one of the customers usually make queries for products meant for ladies and another one for males. Hence, ideally, the search query should be intermediately expanded to "Online purchase of men belts" or " Online purchase of women belts" to make the output more personalized and relevant to E-Commerce customer. A few of the modern search engines provide an option for personalized search. However, they usually fail to adapt to continuously changing needs of the customer as discussed by Wang and Wong (2014). Moreover, users are frequently required to modify their E-Commerce search query number of times to retrieve relevant web links in correct order of ranking as discussed by Verma et al., (2015)

Metasearch engines can address partial indexing problem of conventional search engines to a modest extent. They are built on the top of some search engines, and they search for a query on all of supporting search engines followed by integration and ranking of output links retrieved from each of the search engines to display the result and hence improving recall and precision. However, metasearch engine approach has its own set of limitations. The usual number of web links returned in output for E-Commerce query by each of the supporting search engines is outsized. Youssif et al., (2011) discussed as if the search query is ambiguous, output links in result becomes even more massive as traditional search engines try to retrieve web links corresponding to all probable meanings of a query hence, integrating and correctly ranking vast number of E-Commerce websites require enormous efforts. Moreover, E-Commerce website ranking using conventional data mining techniques is not efficient as discussed by Verma and Singh (2017a,b) and require to deal with many problems like:

- The credibility of high-ranking E-Commerce websites in search engines output appeared to have declined as a customer is not usually able to find the suitable and genuine product at reasonable price. For example, some of the E-Commerce websites are selling goods without acquiring preauthorization from the manufacturer of the product at unreasonable prices leading to various difficulties for the customer while applying for guarantee/warranty services from the manufacturer. Moreover, E-Tailer also finds it complicated to structure their E-Commerce websites appropriately to survive in this modern age of intense competition.
- Conventional website ranking systems do not focus on essential features as required by big data management systems. These features include partial failure support, infrastructure and application scalability, component recovery, data recoverability and ability to respond in real time as needed by modern metasearch systems or search engines to search in today's age of big data as discussed by Tsai et al.,(2015).
- Traditional search engines usually perform semantic less page ranking process regarding frequency count of keywords, the proximity between candidate website and E-Commerce query. The queries which can be interpreted in various contexts are likely to produce unexpected results, and user ends up either with lot many website links and sometimes not even a single link in the output.

The proposed research work focuses on addressing above mentioned problems as faced by various stakeholders viz. E-Tailers, End Users, and Search engine developers. The research problem can be summarized to develop a personalized metasearch engine for the benefit of all stakeholders. Moreover, the proposed approach will overcome the restrictions of traditional data mining approaches to extract useful E-Commerce web links from big databases of various search engines by providing essential features of the second generation big data systems like partial failure support, scalability, real-time response.

4. Comparison of platforms for big data analytics

To appropriately choose deployment framework for a web search and ranking application, we need to compare various aspects such as capabilities for partial failure support, fault tolerance, scaling, real-time processing and efficiency in iterative execution. Here, we compared various existing deployment paradigms in Sections 4.1,4.2 and 4.3 to explain some of the characteristics of different cloud-based platforms useful for deployment of *E-Commerce website search and ranking system*.

4.1. Types of deployment platforms

Various existing cloud-based deployment platforms are explained as follows (Khurana 2014; Malhotra et al., 2017a,b)

- In one of a kind, cluster utilizes blob storage space as a primary storage space such as Azure blob store, S3. Here temporary clusters are implemented, and they exist only till the period of workflow execution. Blob store act as a source and destination of the workflow. Here, virtual machines may be considered as task execution containers.
- In another type, first generation HDFS (Hadoop Distributed File System) is used as a primary storage space. In contrast, here, persistent clusters are used for long-term storage. Moreover, virtual machines are persistent, and they can perform execution as well as data storage. This type may even use blob storage for cyclic backups and to give data to HDFS. This kind of cloud deployment platform is useful for workloads of type SLA batch workloads, Ad Hoc Interactive and Ad Hoc Batch. For instance, interactive SLA workloads are usually deployed on HDFS due to virtual machines requirement as servers and blob storage requirement as a backup.

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015

ARTICLE IN PRESS

4.2. Second generation HDFS

With the recent technological shifts, second generation big data processing systems need to support multiple analytic methods on varied data types, and the ability to respond in real time. Malhotra and Rishi (2017) discussed the essential characteristics of first-generation HDFS like partial failure support, scalability through data streaming and global memory scheduling is also required to be continued by second-generation HDFS as shown in Fig. 2.

There are two significant trends of Second Generation HDFS based Big data search and ranking systems (Gebara et al., 2015; Malhotra and Rishi, 2016)

- There is rapid growth in network bandwidth as compared to hard drive bandwidth.
- Development of In-Memory computation models such as Spark allow intermediate results to be kept in memory and hence reduces overhead of iterative analytics

Second Generation HDFS is adapted as a long-term store from where web applications read their initial data and write back their final results. The data layer is subdivided into various segments for steady storage and provides storage for intermediate objects separately. However, one of the limitations of HDFS lies in running iterative algorithms efficiently. Map function requires to read data at the start of iteration and to write back the results to the disk at the end of the iteration. This frequent access to disk in writing and reading data is responsible for performance and efficiency degradation as discussed by Singh and Reddy (2015).

4.3. Ranking comparison of existing and proposed deployment platforms

Table 1 shows a ranking comparison of various possible big data deployment frameworks on different characteristics such as scaling, fault tolerance. Here Rank -1 shows the best option and Rank - 5 for worst option among all of the listed platforms. It may be noted that this ranking table provides a general idea regarding strengths and weakness of various platforms and it mainly depends on the specific application/purpose. In general, big data applications, there is a tradeoff between Scaling and Real-Time Processing capabilities.

For example, in web search applications, indexing process requires a highly scalable platform to handle billion of web pages returned by some supporting search engines. This indexing accomplished via HDFS and Spark are the optimal choices for web search applications as discussed by Shou et al., (2014), and hence these are preferred and proposed deployment frameworks for E-Commerce website search and ranking applications. The detailed ranking comparison between various deployment platforms is shown in Table 1. In the implementation of our proposed IMSS-AE tool, we have chosen HDFS platform due to its high scaling and fault tolerance rank which are two most important requisites in an E-Commerce environment. We have given preference to HDFS over SPARK platform due to easy availability and adaptability of hardware and software related infrastructural requirements for HDFS- Map reduce environment and hence to improve the probability of increased usage and popularity among retailers.

5. System Design

This proposed research work addresses above mentioned E-Commerce website search and ranking problem as discussed in Section 3 using *Intelligent Technologies based Personalized Big data Analytics*. The simplified modular block diagram of the system is shown in Fig. 3.

5.1. Phase 1: Query preprocessing using semantic analysis

The proposed E-Commerce website ranking system can easily keep track of customer preferences, i.e., short term and long term preferences by building customer's profile. This system can closely monitor customer's browsing history, and the system will automatically update customer's profile with a change in his/her browsing patterns of websites without requiring any additional effort from the customer. Here long-term preferences can be retrieved using customer's past browsing history and registered preferences while short-term preferences will be retrieved using browsing history of last two days only. This phase can extract search queries and visited web links from browsing history by fetching meta keywords and hence by developing customer's profile which can be further used to establish customer's contextual database. These Meta keywords can be utilized for selecting concepts with the ontology-based database. These Meta keywords through selected ideas will be used to disambiguate the search query and hence to expand a simple keyword query into more meaningful customer personalized query to improve the search results through backend search engines as discussed by Malhotra and Rishi (2017). The Semantic Relevancy Vector (SRV) is determined by using Longest Common Subsequence (LCS) to determine

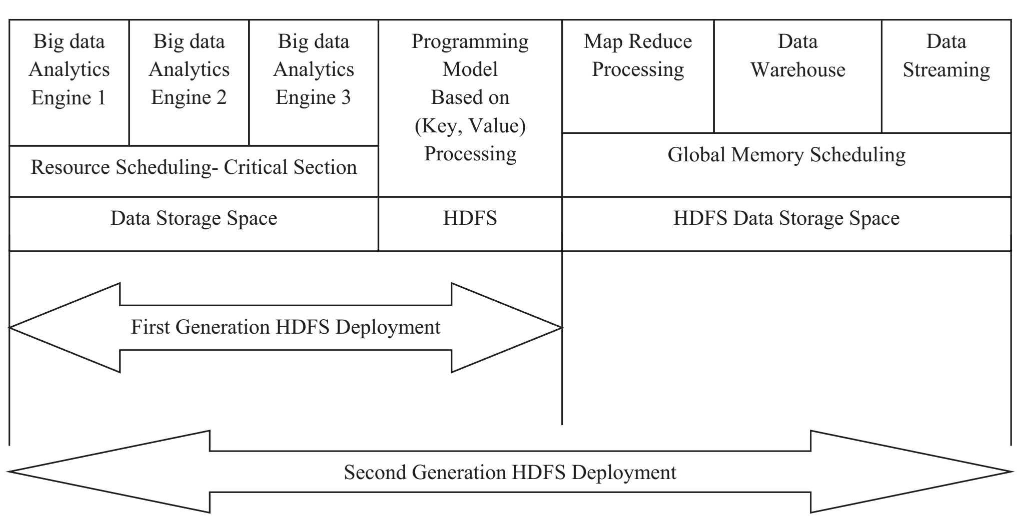

Fig. 2. Second Generation HDFS V/S First Generation HDFS (Malhotra and Rishi, 2017).

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015

Table 1
Ranking Comparison of Existing and Proposed Deployment Platforms.

Platform	Scaling Rank (Type)	Fault Tolerance Rank	Real-Time Processing Rank	Iterative Tasks Rank
HDFS	**1** (Horizontal)	**1**	4	4
SPARK	**1** (Horizontal)	**1**	4	3
PEER TO PEER	**1** (Horizontal)	5	5	4
HPC CLUSTERS	3 (Vertical)	2	3	2
MULTICORE	4 (Vertical)	2	3	2
GPU	4 (Vertical)	2	**1**	2
FPGA	5 (Vertical)	2	**1**	2

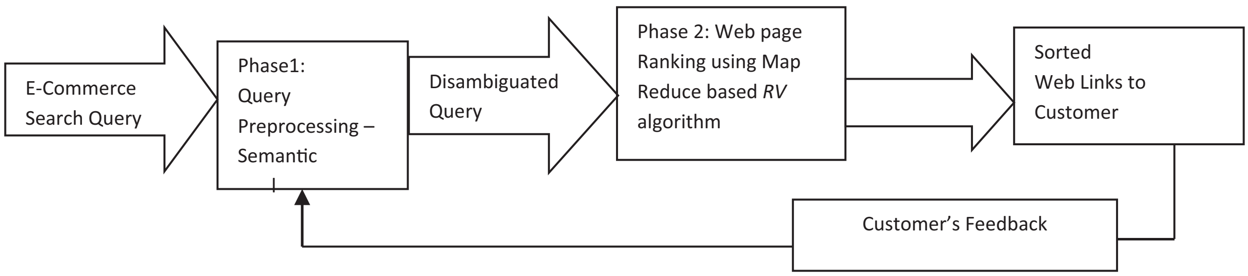

Fig. 3. Simplified system Design.

the proximity of web page and contextual similarity concerning customer requirement. The detailed conceptual flow of this step is shown in Fig. 4.

5.2. Phase 2: Website ranking using map reduce based RV page ranking algorithm

This research paper uses Hadoop-RV- Map Reduce based big data mining and analytics framework to simplify the E-Commerce website personalized search and ranking process through the implementation of *Intelligent Meta Search System for Advanced E-Commerce. IMSS-AE* tool is built on the top of some other popular search directories like Yahoo, Meta search engines like *Dogpile* and search engines like *Google*. This proposed research work is implemented at middle layer of public cloud for service level agreement. This phase accepts preprocessed disambiguated query as generated in the last step. In this step, we will first search for user-specified query on each of the back end search engines and will assign a unique id to each of the retrieved clusters of web pages from 1 to n. These clusters are then compared with user specifications such as privacy/security, response time and ease of accessibility to find relevant cluster list L which should be further processed for ranking purpose. Short listing of clusters can be performed by performing a parametric match. The very first criterion is to determine accessibility which may be public, private or community type cloud. The second criterion is related to security which can be determined by https: transmission capability or SSL availability followed by the third criterion of response time which should be less than that of customer specified value. The first stage of ranking will be

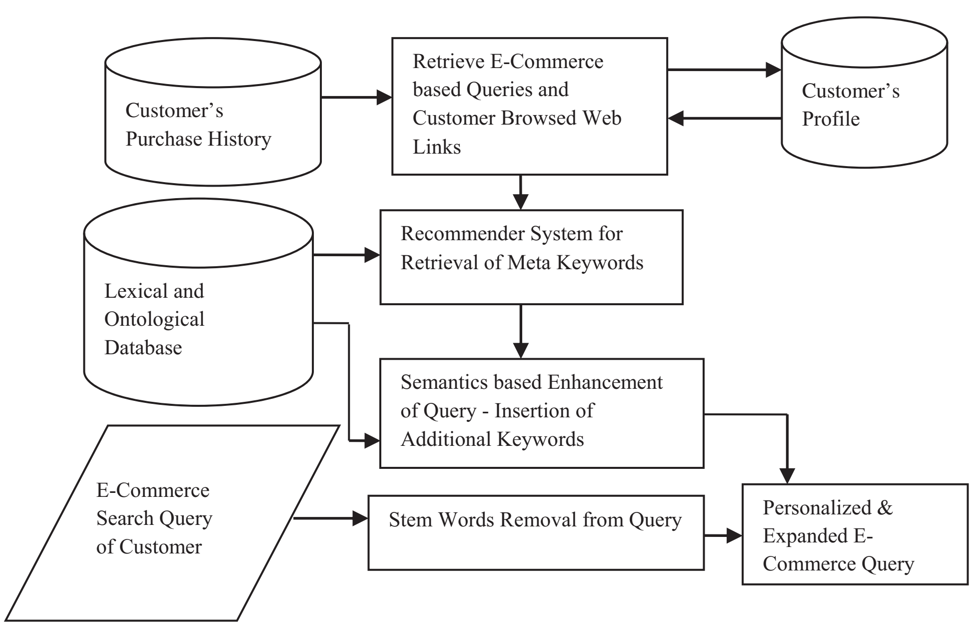

Fig. 4. System Design- Query Preprocessing.

implemented by determination of content relevancy in two-phase programming model called Map and Reduce to support HDFS based cloud framework. Map and Reduce code used in the proposed algorithm is as follows:

```
Map (SEngine_ID: Integer, Web_Log: String)//Web Log Cluster
   processing
{
List < String> TL: = Tokenize (Web_Log) // TL- Token List
   While (Web_Token in TL)
   {
   Insert ((String) KL, (Integer) 1)// KL- Keyword List
   }
}
Reduce (KL: String, count: List < Integer>)
{
Integer Freq = 0
While(KL)
{
Freq = Freq + 1
}
Insert ((String) Web_Token, (Integer) Freq)
}
```

Here, Map method will accept a key as search engine ID for each retrieved web links cluster from various background search engines and the second argument is weblog to tokenize each of the entry of link entry in the weblog for counting frequency of each of the keyword in E-Commerce search query. Insert () method is used to generate elements in the list by inserting numeric one corresponding to each occurrence of a keyword as we token. However, Reduce method is implemented to cumulate over all the occurrence of each keyword. This is accomplished by the insertion of numeric 1 (one) to determine the frequency of the keyword in each of the web document and hence to conclude the content relevancy vector of retrieved web documents from various search engines. The second stage of ranking concludes the time relevancy vector (TRV) for each web page using its last time of update on the web as well as by considering previous customer time spend statistic with similar E-Commerce search query. The third stage of ranking includes feedback relevancy vector (FRV) which may include explicit and implicit feedback of past customer. Some of the previous research results show that explicit feedback of a product / E-Commerce website in the form of online reviews can significantly impact the purchase decision of a customer. Liu et al., (2017) discussed that it is quite difficult for a customer to review a large number of online reviews easily. Hence, there is an urgent need to develop a method to rank E-Commerce websites based on sentiment analysis. Online reviews usually expressed in sentences and hence dictionary based semantic analysis is used in this research work to determine neutral, negative or positive reviews. The Semantic Relevancy Vector (SRV) is already determined in step1. At last, all of these vectors with their weighted contribution as mentioned by customer helps in determination of rank for each of the cluster of web pages as discussed in proposed RV page ranking algorithm. The various evaluation metrics involved in the determination of rank of an E-Commerce website are Semantic Relevancy Vector (SRV), Feedback Relevancy Vector (FRV), Content Relevancy Vector (CRV), Privacy Vector (PV) and Accessibility Vector (AV). The step by step calculation of various evaluation metrics and their role in the determination of overall rank and personalized search precision of an E-Commerce website along with system design of phase 2 is shown in Fig. 5.

5.3. Relevancy Vector (RV) page ranking algorithm

Relevancy Vector, the page ranking algorithm is an extended algorithm of earlier published CPR algorithm by Malhotra et al., (2017a,b). RV algorithm is an improvement over CPR algorithm due to two main reasons (i) RV algorithm is designed to take benefit of cloud technology (ii) RV algorithm unlike CPR algorithm is specially tailored for E-Commerce website ranking. RV algorithm is discussed in detail as follows:

- Start
- Accept E-Commerce query from a customer.
- Personalize search query using customer profile database and semantic enhancement.
- Split the query into various keywords W_1, W_2,……..,W_n and remove stem words from the query.
- Determine minimum and maximum length of each of the keyword as follows

```
    Set min = strlen(W1), max = strlen(W1)
    Set c = 2
      While (c < n) do
        If MIN > Wc then
          MIN = strlen(Wc)
  EndIf
        If MAX < Wc then
          MAX = strlen(Wc)
  EndIf
  EndWhile
```

- Execute E-Commerce query on various backend search engines and assign ID to retrieved websites can easily
- Determine customer navigation session. This process can be accomplished by comparing customer's query with each of the past E-Commerce, and other search queries present in customer profile database using LCS. The LCS, i.e., Longest Common Subsequence is used to determine proximity between website and customer preferences and store the same in SRV[ID] to represent the semantic rank of particular E-Commerce website identified by ID.
- Calculate timestamp T_S of creation and average time spent by past customer T_P to calculate Time relevance vector TRV[ID] = (Ts + Tp)/2

  // Calculation of CRV [ID]

- For x = 1 to n do // n refer to total number of websites
  - o Calculate frequency of each keyword using web dictionary
  - o Eliminate all those websites with frequency of found keywords less than not found keywords
  - o Call **Map**(WebPage_ID, WebPage_Content)
  - o Call **Reduce**(Web_Link, Count)
  - o Calculate average frequency of the frequency of individual keywords
  - o Store average frequency in CRV[ID]

  EndFor

- For x = 1 to r do // r refer to remaining websites after elimination in last step
- Calculate Privacy vector, PV[ID] = 0; If(linkprivacy = privacy(website(ID)) then set PV[ID] = 1

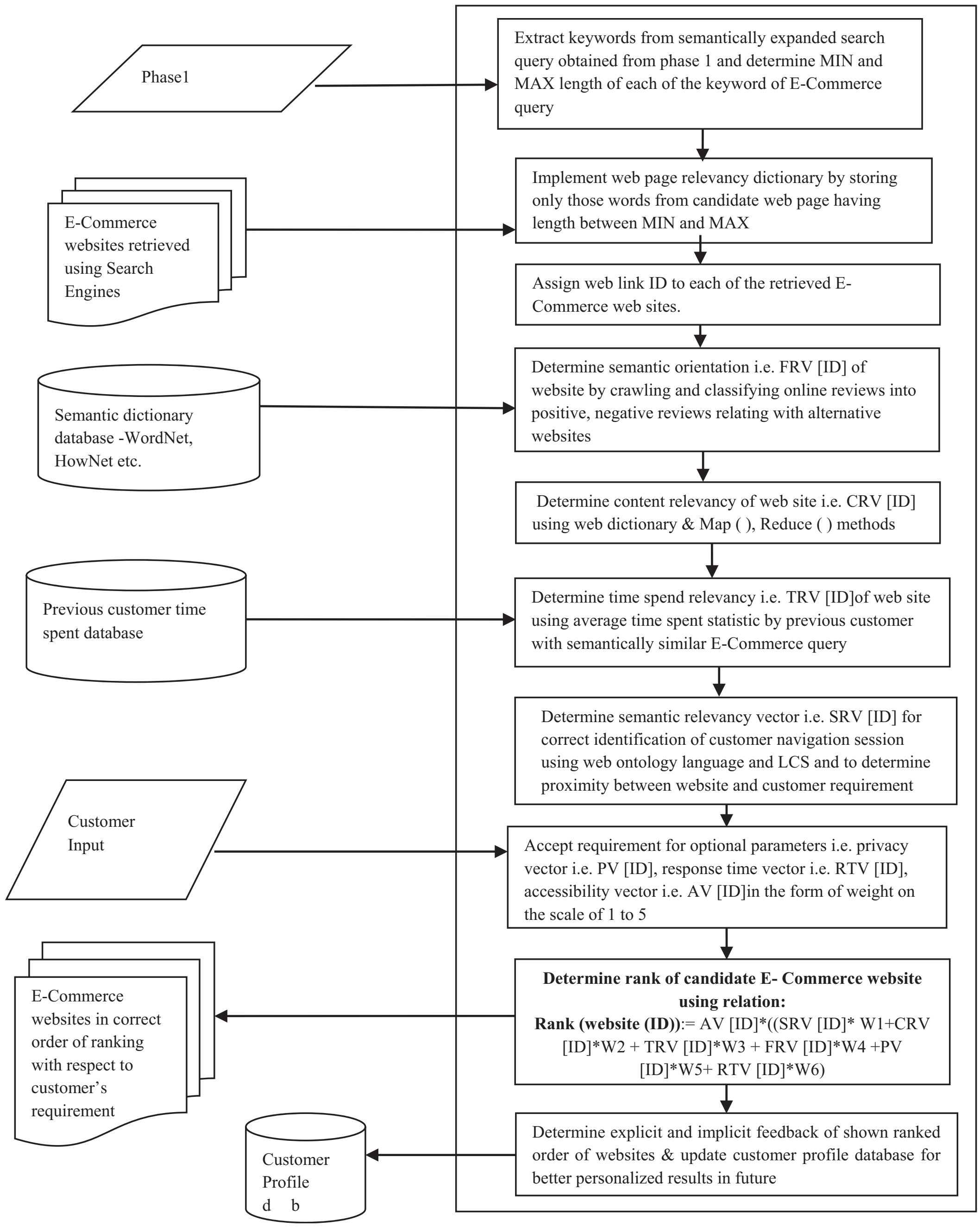

Fig. 5. Web Site Ranking using Map Reduce based RV page ranking Algorithm.

- Calculate Accessibility Vector, AV[ID] = 0; If (Cloud = Public) then set AV[ID] = 1
- Calculate Reply Time Vector, Set RTV[ID] = 0
- If (linkresponse > ReplyTime(website(ID)) then

  RTV [ID] = strresponse – ReplyTime(website(ID))
  EndFor

- Eliminate all E-Commerce websites with either RTV[ID] = 0, PV [ID] = 0 or AV[ID] = 0
- Determine Feedback Relevancy Vector i.e. FRV[ID] using semantic dictionaries to analyze online reviews and categorize them into negative, positive and neutral reviews and calculate FRV as follows:

Set Count = 0
If (Review is Positive) then // *Well Satisfied Experience of Past Customer*
Count = Count + 2
Else If (Review is Negative) then //*Unsatisfied Experience of Past Customer*
Count = Count -2
Else If (Review is Neutral) // *Hesitant or Confused Past Customer*
Count = Count – 1
EndIF
Set FRV [ID] = Count

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015

- Calculate Rank (website (ID)):= AV [ID]∗((SRV [ID]∗W1 + CRV [ID]∗W2 + TRV [ID]∗W3 +
- FRV [ID]∗W4 + PV [ID]∗W5 + RTV [ID]∗W6)
- Accept feedback from the customer about the shown ranking order and update customer profile database.

The RV page ranking algorithm determines the relevancy of an E-Commerce website for a specific customer using the calculation of various relevancy vectors such as Content Relevancy Vector, Semantic Relevancy Vector, Reply Time Vector, Feedback relevancy Vector, Privacy Vector. The algorithm starts with personalized expansion of search query as discussed in Section 5.1. After removal of stem words viz. *a, the, an* from the query. The RV algorithm will calculate the minimum and maximum length of each of the keyword of the search string. The SRV is determined using Longest Common Subsequence. The CRV is determined using Map and Reduce functions. Moreover, the algorithm will remove all those E-Commerce websites from final output with Reply Time Vector = 0, Accessibility Vector = 0 or Privacy Vector = 0. The previous step is further followed by calculation of Feedback Relevancy Vector depending on the experience of past customer. In the last, rank of a website is calculated by weighted summation of various relevancy vectors.

5.4. Intelligent meta search system for advanced E-Commerce - IMSS-AE tool

IMSS-AE tool using second generation HDFS, Map-Reduce framework for big data analytics is implemented in the ASP.NET framework to assist the customer while performing E-Commerce transaction. This tool is also used to determine the performance of RV page ranking algorithm. The interface of IMSS-AE tool is shown below in the Fig. 6. After registration/Sign In/Sign Up, the interface of the tool on authentication will allow the customer to select few or all of the mentioned metasearch engine/search engine/search directory, i.e., *Dogpile, Yahoo*, and *Google* respectively for background retrieval of E-Commerce websites. Here, IMSS-AE tool will act like metasearch engine; the customer can specify search string in the search box on the interface of IMSS-AE tool. The tool will first expand the search query to more meaningful personalized search query. This tool will further assign the rank to some of the top web links retrieved from back-end search engines based on the calculation of various ranking vectors such as AV, FRV, SRV, CRV, RTV, TRV with appropriate weight age as determined from customer specified parameters. The detailed discussion about the calculation of ranking vectors and weighted contribution is discussed in Section 5.2. The tool will output

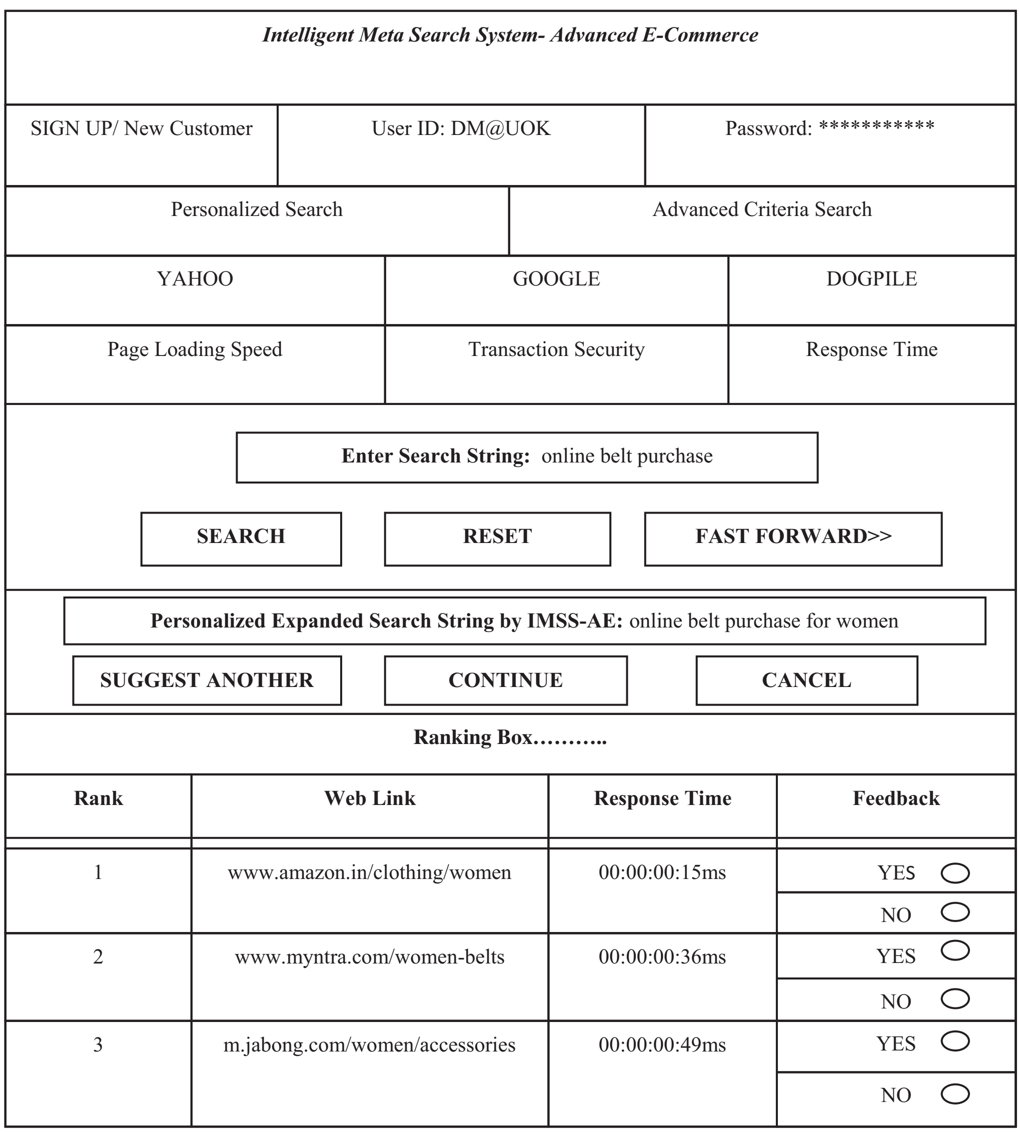

Fig. 6. Interface of IMSS-AE Tool.

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015

E-Commerce web links in the sorted order of their ranking along with various statistics as selected by the customer in advanced search criterion such as page loading speed, response time, transaction security as well as background search engine. However, *Personalized Search* tab will not allow choosing any of the search criteria and will give result directly by referring customer registered past preferences. This tool will suggest personalized expanded search string by using browsing history. Moreover, in output, the tool will rank the various links using customer's preferred search criteria along with details of selected statistics, the tool will also allow the customer to provide feedback about the ranked order of web links and hence to improve its personalized ranking capabilities to better match with changing customer preferences.

6. Experimental and graphical analysis

The personalized relevancy of an E-Commerce website to a specific customer for a given product query depends upon its position in the output of search results. To compare the IMSS- SE Tool with other popular search tools, Precision of search at X metric is considered, which is here shown by P (X). Various search tools used for comparison in this study are metasearch engine, search engine and search directory, i.e., *Dogpile, Google, Yahoo* and IMSS-SE Tool by Malhotra et al.,(2017a,b). For a given E-Commerce query, P (X) reports how many fractions of output links in the result, labeled as significant are presented in the top X results. Here, it is assumed that a web link ranked upper is more relevant for the customer. The tool rank is then compared with human volunteer's judgment to verify the relevancy reported by the tool as well as professional search engine/tool, and at last the difference in precision of IMSS-AE tool and professional search tool is plotted.

To evaluate the efficiency and effectiveness of proposed RV algorithm and IMSS- AE tool. Here, we employed 20 human volunteers in various age groups from 15 years to 45 years, and with a minimum of 3 years' experience of carrying out numerous E-Commerce transactions, nine of them are males, and eleven of them are females. They were asked to use personal laptops with installed IMSS-AE tool followed by sign up/registration process on the tool, we asked volunteers to repeat the following steps for at least five trial runs on each of *Dogpile, Yahoo, Google* and proposed IMSS- AE Tool:

1. Initially, we asked volunteers to search for intentional incomplete E-Commerce query, for instance, a query like *Samsung online purchase* rather than *Samsung mobile online purchase*
2. In next step, volunteers were asked to rank output links from 1 (worst) to 5(best) individually to all of the considered engines and proposed IMSS- AE tool. The basis of ranking is various precision parameters such as personalized relevancy, page update time and response time to the top 10 web links in the output.
3. After gathering ranked data from each of the volunteers, normalization of various precision parameters is carried out by using the following expression:

$NP_{ab} = (MAX\ (PP_{ab}) - PP_{abr})/(MAX\ (PP_{ab}) - MIN\ (PP_{ab}))$

Where, PP_{ab} = Value of b_{th} precision parameter of $_{metasearch}$ webpage; NP_{ab} = Normalized value of b_{th} precision parameter of a_{th} webpage; MIN, MAX = Minimum and Maximum value of each of the precision parameter.

1. After that, we calculated the overall weighted precision of each E-Commerce web link retrieved by the volunteer as $\mathbf{N_a = \sum W_b. NP_{ab}}$, Where, N_a = weighted precision of a_{th} webpage; W_b = Weight assigned to b_{th} parameter by customer, where $0 <= W_b <= 1$
2. At last, the overall precision of search engine/tool is determined by calculating the average of all the weighted precisions as gathered by volunteers for a given parameter among Response time; Page updated content, Personalized Relevancy at a time, **Precision(ID) = AVERAGE ($\mathbf{N_a}$).**

The graph is shown in Figs. 7, 8 and 9 demonstrate the average precision metric comparison between proposed IMSS-AE Tool with a popular search directory, i.e., Yahoo as calculated by volunteers by following above steps for various precision parameters, i.e., Response Time, Page Freshness and Personalized Relevancy respectively. The graphs shown indicate that initial average precision of IMSS-AE is lesser than *Yahoo* for page freshness, response time and personalized relevancy. However, soon after few trial runs, the precision of tool improves in comparison to Yahoo. This improved precision demonstrates the semantic-based learning capabilities of IMSS-AE. This tool can build customer profile database by monitoring his/her personalized browsing preferences with some trial runs and hence tool will be able to calculate various important relevancy vectors as discussed in Section 5.2 such as SRV, FRV, TRV more accurately. However, precision improvement in Fig. 7 for response time is not as significant as in Figs. 8 and 9, i.e. for page freshness and personalized relevancy. This difference in precision statistics is because of background implementation of Map-Reduce based second generation HDFS used in the tool. This lagging is due to time delay occurring on account of iterative analytics. This suspension can be improved further by use of Spark based HDFS system as in-memory computation models implemented through Spark allow intermediate results to be kept in memory and hence reduces the overhead of iterative analytics as

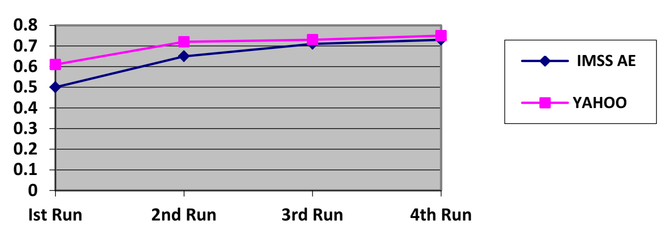

Fig. 7. Precision Comparison between IMSS-AE and YAHOO- Response Time.

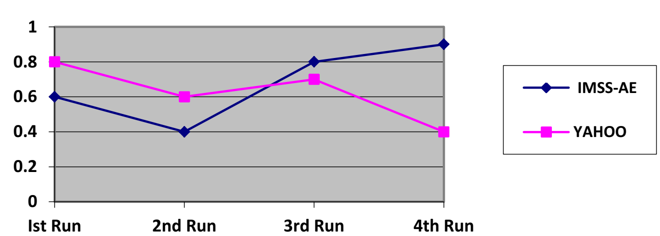

Fig. 8. Precision Comparison between IMSS-AE and YAHOO - Page Freshness.

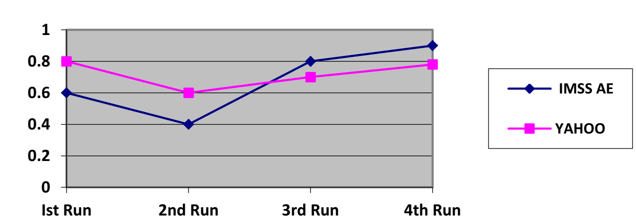

Fig. 9. Precision Comparison between IMSS-AE and YAHOO- Personalized Relevancy.

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015

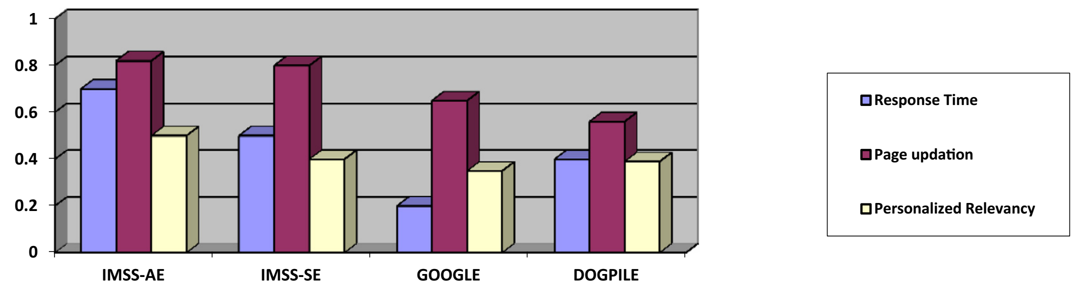

Fig. 10. Search Precision Comparison between IMSS-AE with *Dogpile, IMSS- SE,* and *Google.*

discussed by Malhotra and Rishi (2017). Similarly, we carried out an extensive experimental comparison between proposed IMSS-AE tool with Google, Dogpile and IMSS-SE regarding various precision parameters, i.e., Response Time, Page Updation and Personalized Relevancy. The extensive experimental evaluation discussed and its graphical demonstration in Figs. 7, 8, 9, and 10 indicates the improvement in various precision parameters at much faster pace when a personalized search is accomplished with proposed IMSS-AE over other professional and popular search engines, i.e., Google, Yahoo and Meta search engines, i.e., Dogpile, IMSS-SE.

7. Conclusion and future work

This research paper presents a Hadoop- Map Reduce based personalized E-Commerce search framework for the second generation big data analytics. The research gap is shown in this study by submitting various conventional search systems in the form of detailed category wise literature review. This research work proposes a novel RV page ranking algorithm and implements the same as an E-Commerce website ranking tool, i.e., Intelligent Meta Search System for Advanced E-Commerce. The IMSS- AE tool can assist modern day customer in choosing appropriate E-Commerce website for online purchase of a product. The efficiency of proposed ranking approach is justified by experimental analysis. The graphical evaluation for comparison of personalized precision of IMSS-AE tool over *Yahoo, Dogpile, Google*, and IMSS- SE tool demonstrates the effectiveness of proposed approach over conventional & professional page ranking methods. The practical implications for three different audiences of this research work are as follows:

Practical Implication for End User- The end user of this research work is an online customer willing to make an online transaction. The result of this research work in the form of IMSS-AE tool can assist the customers in the suitable ranking of E-Commerce websites for the purchase of a specific product. The end user will be benefitted by personalized website ranking output and hence can easily select a website that is most appropriate for satisfying the online purchase needs of a user.

Practical Implications for E-Tailers/Retailers: The E-Tailers, i.e., E-Commerce websites or Retailers will be benefitted from this research work as they can use IMSS- AE tool to improve the structure of their websites to satisfy their customers easily and hence to take the lead over the competition.

Practical Implications for Search Engine Developers: This research work can assist the developers in search engine domain to bring out their best in the form of meta search tool. They can take advantage of vast databases of various search engines and can employ Big Data Analytics to fetch out personalized page ranking patterns using an innovative algorithm like proposed RV page ranking algorithm.

In future, RV page ranking algorithm and IMSS-AE tool can further be enhanced to implement market basket analysis through extraction of association rules from big data stored in online transactional databases. These association rules will assist various stakeholders, i.e., E-Tailers/Retailers for launching various promotional schemes such as Buy One Get One; combo discounted offers, appropriate product suggestions to online customers, target marketing. The retail transactional databases are often quite huge. The traditional data mining approaches to mine useful patterns from such voluminous databases to launch promotional schemes are quite time-consuming and inefficient when compared with Hadoop/Map reduce like big data analytics framework. These promotional schemes from E-Tailers/Retailers will be quite beneficial for another stakeholder, i.e., end user and hence will result in not just increase in sales but also in better satisfaction of end user. Moreover, for the benefit of another stakeholder, i.e., search engine developers, next-generation big data analytics may be incorporated in future editions. These future versions may include (i) SPARK model may be used for reduction in processing overhead. The overhead is on account of iterative analytics and can be reduced by keeping intermediate results in memory To overcome various limitations of conventional HDFS- Map Reduce such as lack of real-time response and dynamic initialization of multiple analytic engines (ii) proportional growth in network bandwidth requirements along with secondary storage needs. The precision of proposed IMSS-AE tool can further be improved by incorporating artificial neural networks for implementation of supervised learning of customer preferences for the better-personalized experience.

References

Adamopoulos, P., 2014. On discovering non-obvious recommendations: Using unexpectedness and neighborhood selection methods in collaborative filtering systems. In: Proceedings of the 7th ACM international conference on Web search and data mining, ACM, pp. 655–660.

Alam, M., Sadaf, K., 2015. Labeling of Web Search Result Clusters using Heuristic Search and Frequent Itemset. Procedia Computer Science, Elsevier, pp. 216–222.

Aoki, Y., Koshijima, R., Toyama, M., 2015. Automatic Determination of Hyperlink Destination in Web Index. In: Proceedings of the 19th International Database Engineering & Applications Symposium, ACM, pp. 206–207.

Bo, C., Yang-Mei, L., 2014. Design and Development of Semantic-Based Search Engine Model. Intelligent Computation Technology and Automation (ICICTA), 2014 7th International Conference, IEEE, pp. 145–148.

Cacheda, F., Carneiro, V., Fernández, D., Formoso, V., 2011. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. ACM Trans. Web (TWEB) 5 (1). https://doi.org/10.1145/1921591.1921593.

Gebara, F., Hofstee, H., Nowka, K., 2015. Second Generation Big data Systems: Cover Feature Outlook. IEEE Computer Society, IEEE, pp. 36–41.

Gomez-Nieto, E., San Roman, F., Pagliosa, P., Casaca, W., Helou, E.S., de Oliveira, M.C.F., Nonato, L.G., 2014. Similarity preserving snippet-based visualization of web search results. IEEE Trans. Visualization Comput. Graph. 20 (3), 457–470.

Guy, I., Jaimes, A., Agulló, P., Moore, P., Nandy, P., Nastar, C., Schinzel, H., 2010. Will recommenders kill search?: Recommender systems-an industry perspective. In: Proceedings of the fourth ACM conference on Recommender systems, ACM, pp. 7–12.

Jung, S., Harris, K., Webster, J., Herlocker, J.L., 2004. SERF: integrating human recommendations with search. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management, ACM, pp. 571–580.

Khurana, A., 2014. Bringing big data systems to the cloud. IEEE Cloud Comput. 1(3), IEEE, 72–75.

Kuppusamy, K.S., Aghila, G., 2014. CaSePer: an efficient model for personalized web page change detection based on segmentation. J. King Saud Univ. Comput. Information Sci. 26(1), Elsevier, 19–27.

Limbu, D.K., Connor, A., Pears, R., MacDonell, S., 2006. Contextual relevance feedback in web information retrieval. In: Proceedings of the 1st International Conference on Information Interaction in Context, ACM, pp. 138–143.

Liu, Y., Bi, J.W., Fan, Z.P., 2017. Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. Information Fusion 36, 149–161.

Malhotra, D., 2014. Intelligent web mining to ameliorate Web Page Rank using Back-Propagation neural network. Confluence the Next Generation Information Technology Summit (Confluence), 2014 5th International Conference, IEEE, pp. 77–81.

Malhotra, D., Verma, N., 2013. An ingenious pattern matching approach to ameliorate web page rank. Int. J. Comput. Appl. 65 (24), 33–39.

Malhotra, D., Malhotra, M., Rishi, O.P., 2017. An Innovative Approach of Web Page Ranking Using Hadoop- and Map Reduce-Based Cloud Framework. In: Proceedings of Advances in Intelligent Systems and Computing, Vol. 654, CSI, Springer, pp. 421–427.

Malhotra, D., Rishi, O.P., 2016. IMSS-E: An Intelligent Approach to Design of Adaptive Meta Search System for E-Commerce Website Ranking. Proceedings of the International Conference on Advances in Information Communication Technology & Computing, ACM. https://doi.org/10.1145/2979779.2979782.

Malhotra, D., Rishi, O.P., 2017. IMSS: A Novel Approach to Design of Adaptive Search System Using Second Generation Big data Analytics. In: Proceedings of International Conference on Communication and Networks, Springer, pp. 189–196.

Malhotra, D., Verma, N., Rishi, O.P., Singh, J., 2017b. Intelligent Big data Analytics: Adaptive E-Commerce Website Ranking Using Apriori Hadoop–BDAS-Based Cloud Framework. Maximizing Business Performance and Efficiency Through Intelligent Systems, IGI Global, pp. 50–72.

Rasekh, I., 2015. A new competitive intelligence-based strategy for web page search. Elsevier, Procedia Computer Science, pp. 450–456.

Sugiyama, K., Hatano, K., Yoshikawa, M., 2004. Adaptive web search based on user profile constructed without any effort from users. In: Proceedings of the 13th International Conference on World Wide Web, ACM, pp. 675–684.

Tanapaisankit, P., Watrous-deVersterre, L., Song, M., 2012. Personalized query expansion in the QIC system. In: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries, ACM, pp. 259–262.

Shou, L., Bai, H., Chen, K., Chen, G., 2014. Supporting privacy protection in personalized web search. IEEE Trans. Knowledge Data Eng. 26(2), IEEE, 453–467.

Singh, A., Vélez, H.G., 2014. Hierarchical multi-log cloud-based search engine. In: Complex, Intelligent and Software Intensive Systems (CISIS), Eighth International Conference, IEEE, pp. 211–219.

Singh, D., Reddy, C.K., 2015. A survey on platforms for big data analytics. J. Big Data 2 (1), 8. https://doi.org/10.1186/s40537-014-0008-6.

Tsai, C.W., Lai, C.F., Chao, H.C., Vasilakos, A.V., 2015. Big data analytics: a survey. J. Big Data 2 (1), 21. https://doi.org/10.1186/s40537-015-0030-3.

Verma, N., Malhotra, D., Malhotra, M., Singh, J., 2015. E-commerce website ranking using semantic web mining and neural computing. Procedia Computer Science, Science Direct, Elsevier, pp. 42–51.

Verma, N., Singh, J., 2017. An intelligent approach to Big Data analytics for sustainable retail environment using Apriori-MapReduce framework. Ind. Manage. Data Syst. 117(7), Emerald, 1503–1520.

Verma, N., Singh, J., 2017. A comprehensive review from sequential association computing to Hadoop MapReduce parallel computing in a retail scenario. J. Manage. Analytics, Taylor and Francis. doi:10.1080/23270012.2017.1373261.

Vinay, V., Wood, K., Milic-Frayling, N., Cox, I.J., 2005. Comparing relevance feedback algorithms for web search. In Special interest tracks and posters of the 14th international conference on World Wide Web, ACM, pp. 1042–1053.

Wang, S., Xu, K., Zhang, Y., Li, F., 2011. Search engine optimization based on algorithm of BP neural networks. In Computational Intelligence and Security (CIS), 2011 Seventh International Conference, IEEE, pp. 390–394.

Wang, H., Wong, K., 2014. Personalized search: An interactive and iterative approach. In Services (SERVICES), 2014 IEEE World Congres, IEEE, pp. 3–10.

Wasid, M. and Kant, V., 2015. A particle swarm approach to collaborative filtering based recommender systems through fuzzy features. Procedia Comput. Sci. 54, Elsevier, 440–448.

Youssif, A.A., Ghalwash, A.Z., Amer, E.A., 2011. HSWS: Enhancing efficiency of web search engine via semantic web. In: Proceedings of the International Conference on Management of Emergent Digital EcoSystems, ACM, pp. 212–219.

Zhang, G., Li, C., Xing, C., 2012. A Semantic++ Social Search Engine Framework in the Cloud. In Semantics, Knowledge and Grids (SKG), 2012 Eighth International Conference, IEEE, pp. 270–278.

Please cite this article in press as: Malhotra, D., Rishi, O.P. An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.02.015