Indian Journal of Science and Technology, Vol 9(36), DOI: 10.17485/ijst/2016/v9i36/102055, September 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Search Ranking for Heterogeneous Data over Dataspace

Niranjan Lal[1*], Samimul Qamar[2] and Savita Shiwani[1]

[1]Computer Science Department, Suresh Gyan Vihar University, Jaipur - 302017, Rajasthan, India; niranjan_verma51@yahoo.com, savitashiwani@gmail.com
[2]Computer Network Engineering Department, King Khalid University, Kingdom of Saudi Arabia; drsqamar@rediffmail.com

Abstract

Traditional relational database systems queries works over structured data, whereas information retrieval systems are designed for additional versatile and flexible ranked keyword queries, works over unstructured data, Semi-structured, Streamed data, Social networking data and data without any format, known as heterogeneous data. However, several new and emerging applications need data management capabilities that mix the advantages both approaches. In this paper, we have proposed and initiate steps to combine heterogeneous statistics and information retrieval systems over Dataspace, which are the collection heterogeneous data, data from various sources and in different format. In several enterprise, the heterogeneity among information at different levels has becomes a difficult job. In an organization, data exist in structured, semi-structured or unstructured format or combination of all these. The existing heterogeneous data management systems are unsuccessful to deal with such information in efficient manner. Dataspace approach gives the solution of the problem of presence of heterogeneity in information and a variety of drawbacks of the existing systems. The main motive of this paper is to explain searching ranking mechanism in Dataspace. We also investigate how structured, semi structured or unstructured data can be take advantages for ranking of search on Web and Dataspace with their research challenges.

Keywords: Algorithms, Dataspace, Information Retrieval, Internet of Thing, Ranking, Ranking of Database, Search Engines Algorithms

1. Introduction

Searching structured information uses simple database queries for retrieval of information. In present scenario data is available in heterogeneous form or collection of unstructured, semi-structured or structured format that is in several format and embedded in each other, or underline the qualities of each other, the questions are arises here, How we can search heterogeneous data, How ranking mechanism will take decision about information which is appropriate for a given query and How combined search queries for heterogeneous data consisting of structured keyword.

Dataspace are collection of unstructured, semi-structured structured, and heterogeneous data and set of correlation among them[3], information from multiple sources in different formats or unidentified contents and contents without any proper format, For example rtf , text, ms word contents, electronic mails, rich site summary feeds, Audio files, Video files, portable document format, hyper text markup language files), semi-structured (For example extensive markup language, LaTEX files, web data, Electronic Data Interchange, scientific research data), or structured i.e., data in proper format and system readable, like relational 'data model' real-world entity (e.g., Relational database, Customer Relationship Management, Enterprise Resource Planning data, a software package, and an email repository. "A Dataspace should be totally helpful for data portability"[2]. The Dataspace Participants is shown in the Figure 1[21]. Ranking mechanisms is one of the most necessary mechanisms of every search engine and also require high attention during the development

*Author for correspondence

of search engine. Different search engines may apply different ranking algorithm and an information retrieval system give higher rank to relevant documents.

In today's era billions of web pages available over the internet and Dataspace or Big data systems, and volumes of multi-structured data also generated every day. several organization are looking for better way to capture appropriate data, while when a end user entered a query, a user's query return thousands of web page so to retrieve a relevant document from the Web or Dataspace, it is important to rank these results to displayed first to find the optimum solution. Currently, heterogeneity is increasing day by day. The availability of data in various multiple formats and extraction of information from these heterogeneous resources is challenging task. For example "What is the population of India?" Here, we get the response in number and another example for textual documents, "***Why Amitabh Bachan is Famous?"*** here answers will be because of his Good Acting.

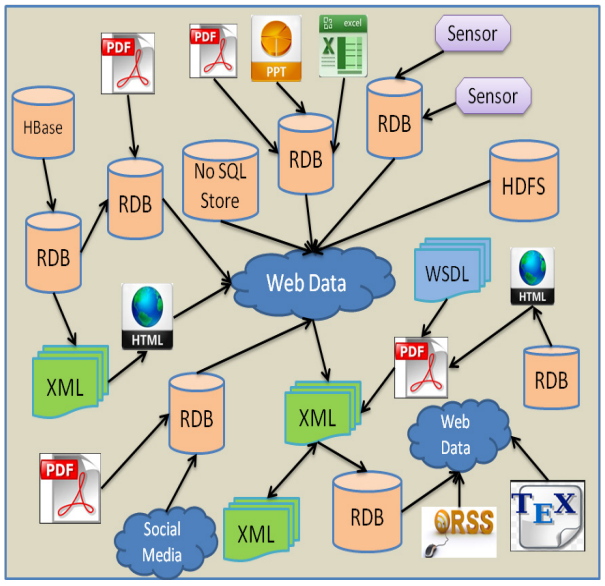

Figure 1. Dataspace participants.

In this paper we evaluate search and ranking of heterogeneous data from Dataspace by hybrid query i.e., heterogeneous data for combined queries dealing with heterogeneous data over Dataspace, and it will give the redundant results from the Dataspace to end user. In this paper we also investigate ranking framework for heterogeneous search with combining keywords with structure elements and Dataspace research challenges for search ranking.

We have organized this paper as follows. Section 2, explains previous work done in searching and ranking, Section 3 defines the Dataspace, with some examples of Dataspace system and Dataspace Management System Architecture, Section 4 explains the Ranking problem in Dataspace, problem in the Text, structured and in heterogeneous data over Dataspace. In Section 5 we have given our proposed solution for search ranking in Heterogeneous Data. Section 6 discussed about release problems and possible guidelines for future study in the field of Dataspace. Final Section 7 concludes the paper with future direction.

2. State of the Art

This section briefly discussed about related work of Information retrieval, ranking of structured data on combined query.

Author in[8], Provides a ranking algorithms for finding statistics about document in the form of term frequency, using probability approach and foundation for information retrieval.

Author in[9] mentions efficient approach for structured data retrieval by combining structured query and phrase keywords.

Author in[10,11] uses link analysis for connected data over the web for searching and ranking.

Author in[12,13] describes how to rank data, how to use ontology approach for web search, and how rank web data using hierarchical link analysis.

Author in[14], explained the concepts for the retrieval of semi-structured data like, XML (Extensive Markup Language) as the extraction of semi-structured data is discussed.

Author in[15] describes the idea of correlations for ranking of db query outcome using probability model between text and structured data.

Author in[16] describe the hybrid approaches for searching data on semantic web, using ontology for retrieval of documents.

Author in[17] explained how to search data by combines keywords and semantic search.

Author in[18] mention approach ranking data using RDF graph.

Ranking developed in the IR community, where as extracted information is combination of heterogeneous data, ranking for Dataspace dataset does not deal with

pre-defined format data only, but also deals with the different datasets and format data on Dataspace and web.

3. Dataspace

Dataspace are a way to manage information over the web and make it possible to run certain types of queries those are inappropriate or unrealistic in traditional indexes and databases. Dataspace construct and separate indexes, their database information i.e., metadata and heterogeneous data

Dataspace is to activate collaboration between database researcher through the sharing of any organization data, text, images, audio, video, sound, movies, models, social media data, technical reports, new research, white papers, models, multimedia data, sensor data, administrative data, conferences, medical data, and simulations etc.

Do you want to discover organization, enterprises or university data? 1. Share your existing large dataset, 2. Make your all data visible globally or public, which is not confidential, 3. Make your data visible to only those your collaborators, and 4. Make it uniform, no matter where the heterogeneous data is located.

If we create a Dataspace for any university, it can store, 1. Students Projects and Reports, 2. Administrative Works, 3. Examination Data, 4. Technical Reports, 5. Conferences and Workshop Proceedings, 6. Research Datasets that are Independent, 7. Digital Collection of Images, Audios, Videos, Emails, PPT (Power Point Presentations), PDF (Portable Document Format), excel, or other assets created by members of the university.

Dataspace will produce innovation in several areas for the commercial and society for the future challenges, like management of Personal Information (e.g., PIM), data management of Scientific data, Environmental observation and forecasting, Personal health information, Ecological data analysis, E-science, E-commerce, M-commerce, The web, Military applications, Medical applications, Social media, Multimedia, Structured queries and content on the WWW.

3.1 Dataspace Management System Architecture

This section describes architecture of a Dataspace Management System (DSMS). The Dataspace Management System Architecture is shown in Figure 2. A DSMS work as a mediator between Dataspace and users to provide services to them and also deal, manage, supports heterogeneous data, make relationship and manipulate these relation among the users. Dataspace architecture provide the service to the organizations, enterprises or companies, so they can manage heterogeneous data, which is scalable over time to meet the different needs of the organization and remove the diversity of the real world problems related to heterogeneity, interoperability and data integration. DSMS architecture support any type of heterogeneous data with associated data modeling and mapping of Dataspace.

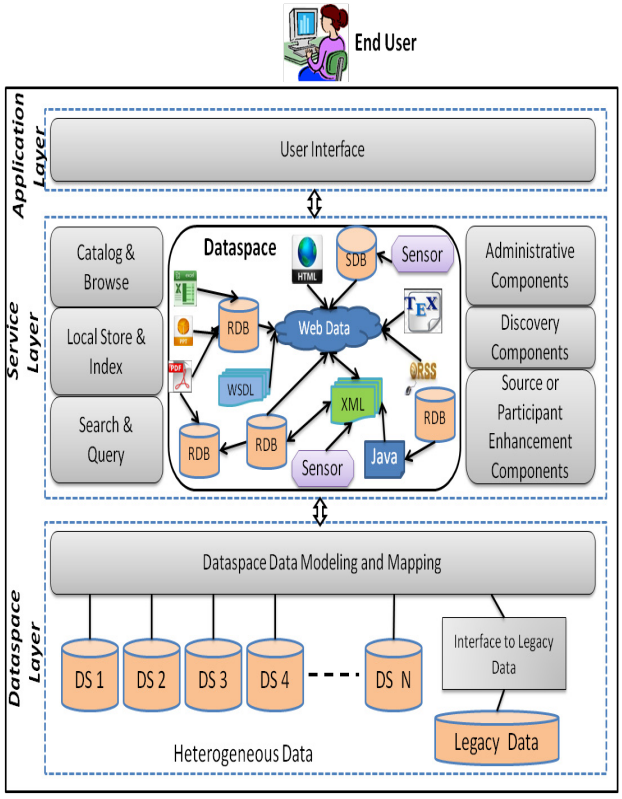

Figure 2. Dataspace management system architecture.

Dataspace architecture is consisting mainly three layers which contains different components to manage the data, named as: 1. Dataspace Layer, 2. Service layer, and 3. Application layer.

3.1.1 Dataspace Layer

This is the first bottom layer which maintains and stores the heterogeneous data and represent schemas which contains the following components: Dataspace Data Modeling and Mapping, Data sources (DS), metadata or

repository for the Dataspace, legacy data and interface to legacy data.

- Dataspace Data modeling and Mapping: Data integration is significant field of novel research, which is undergo from important data modeling, Dataspace system predict a unified system that offers utile service on its heterogeneous data. Dataspace modeling is done by the Dataspace data model which is able to bring forward or present heterogeneous data and able to represent heterogeneous data in a single data model, which also allow for the integration of all information like RDB, XML, Social Data, PPT, RSS etc. inside a single Dataspace data model, it is the foundation of Dataspace system. Dataspace data model map the schemas
- Data sources (DS): Data sources can be any format like; HTML , PPT, PDF, DOC, EXCEL, XML, RDB, Web Data., Social Data, Logs, RSS, NoSQL, File Systems , Sensor data HBase, HDFS, Latex Files, WSDL (Web Service Description Language), Conference Papers, Technical reports, any type of data which are managed by the organizations.
- Legacy Data: A legacy database is usually something that you will have to receive or obtain design schema or design decision and which can be replaced by a suitable technique. In essence it's is legacy when newer technologies have been developed, and we are still using the old system, old database or old techniques, where Information stored in an old or no longer format or computer system, where difficult to access process and retrieve these data.
- Interface to Legacy Data: Most organization or enterprise have an environment of different kind of legacy systems, legacy processes, legacy application and legacy data sources. Maintaining legacy data is one of the unmanageable and difficult challenges that advanced organization or companies are facing today. So by using Dataspace technology we can solve this problem of legacy data modernization, to handle the legacy data, Dataspace layer provides an interface to legacy data. Legacy data interface is used to identify the interaction perimeter of the legacy system, the legacy asset(s) and analyze the interaction to make the system agile, legacy data interface helps to read and updates the external data with their structure, semantic and modeling of legacy data.

3.1.2 Service Layer

This layer provides the conceptual way to store the information in the Dataspace layer and way to access this information. This is the mediation layer consists of several mediators and provides services for Dataspace. The required components of this layer are, 1. Catalog and Browse: stores metadata of Dataspace participants, 2. Search and Query: Querying and searching from Dataspace, 3. Local store and Index: Store data locally in cache and index heterogeneous data for fast accessing of data, 4. Discovery component: discovers new Participants and relationship of them with old or new participants, 5. participant enhancement component: Provides Backup, replication and recovery of heterogeneous data in Dataspace, and 6. Administration component : Interaction among all components of Dataspace management system[1,2].

3.1.3 Application Layer

This layer interacts with end users. This layer responsible for providing user interface to access the services provided by DSMS.

4. Problems in Dataspace

This section shows the problem in Dataspace like Ranking problem, text and structured data retrieval problem with heterogeneity of data in Dataspace.

4.1 Ranking Problem in Dataspace

Ranking gives an appropriate and efficient a result of documents searched on the web, for given information is obtained by their relativity is an intrinsic part of each search engine. The ranking function evaluates the power of searching by giving good chance of relevant results are found on Dataspace.

In information retrieval, rank documents according to relevance to some query or phrase keywords. Ranking problems arise in Dataspace due to heterogeneity of data; in this case end user can retrieve any type of information's from the Dataspace according to their preference like, Reports, Research papers, Titles, E-books, PDF, PPTS, Email etc., from University Dataspace. Songs, Lyrics,

Movies etc. from Multimedia Dataspace and medical data or biological data from the health Dataspace. There are many applications where it is desirable to rank rather than to classify instances, for instance: Ranking the importance of web pages, evaluating the financial credit rating of a person, and ranking the risk of investments. When the end user entered a query, the index of information is used to get the documents most relevant to the end user query and resulted documents are then ranked according to importance and their degree of relevance.

In case of Dataspace, how does search take place over a heterogeneous data known as Ranking, how schema mapping works on heterogeneity data that support search to query with given keywords and phrase and what data may be relevant even we know there is no single data model and mediated schema available .

4.2 Text and Structured Data rertieval problem

Conventional IR models are using keyword query to retrieve relevant or useful documents. One of the leading problems in retrieval of structured data is that it returns results are not in predefined or proper format as same as in document retrieval. There are varieties of ways to produce all possible existing results. Then the results are ranked using any one of the given technique as described here 1. Traditional content-based models, For e.g., TF-IDF, 2. Vector space model, or 3. Content structure-based ranking models. However, there is still absence of common estimation methods for contrast all the above retrieval models for keyword search on structured data. So it is very difficult to illustrate any conclusions on the above diverse approaches.

4.3 Data Heterogeneity on the Dataspace

On the web multiple datasets are store and available in multiple formats and each dataset is represented in the form of data graph. Real-world entities and datasets show heterogeneity at two levels one is the schema level due to same entity store in different datasets and the data level due to same data in different format with different descriptions. Figure 3 shows an example of this heterogeneity by real-world datasets like Amazon a, and Imdb 1. Mappings between different data sources can be compute using[20].

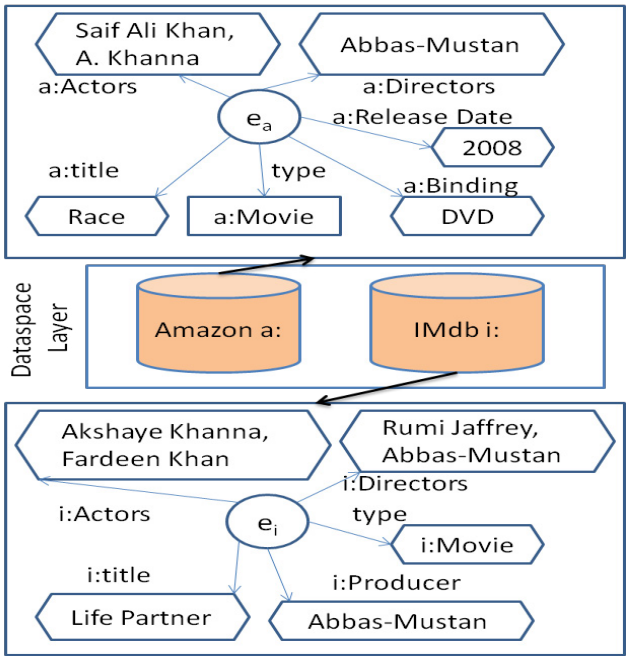

Figure 3. Data heterogeneity on the dataspace.

5. Proposed Solution

In this section first we have proposed data model for heterogeneous data, and then we have given the solutions of text data, structure data and heterogeneous data over Dataspace.

5.1 Retrieval Data models for Text and Structured Data

The main problem in the Dataspace how combined the queries for ranking for heterogeneous data. We have divided this in two models 1. Heterogeneous Data Model and 2. Heterogeneous Query Model.

5.1.1 Heterogeneous Data Model

Representation:

Name Graph- G (R, L, E_R, E_L, E_G, G')

Where

R-Resource nodes,

E_R-Represents edges used for connecting resource nodes,

E_L-Edges used for connecting resource nodes to literal nodes L,

E_G-Edges used for connecting R to G'

Named Graphs- G' (R', L',)- Contains elements of Graph G (R'⊂R.)

Labeling of Edges:

One edge (Labeled Contents)-Point a literal node to hold textual Information

Other edges-Points *headline*, *paragraph*, etc.

One textual entity of textual information form a named Graph g'∈G' is described in Figure 4 using dashed circles.

5.1.2 Heterogeneous Query Model

Query to heterogeneous data model is differs due to heterogeneity of data from textual keyword queries to combined queries to retrieve structure, semi-structure or unstructured data.

Example: A Combined query q can includes textual data q_t, structured Data q_s and a text Data i.e., $q = q_s \wedge q_t$. If one component is unfilled, the query is converted into purely textual or purely structured or semi-structured.

The structured part q_s follows:

Where,

$q_s = \{q_{s1}, q_{s2},\}$.

. $q_t = \{ \mid = (x_i, kw), x_i \in variable(q)\}$.

For example, Find the query: *"Cricketer who shifted to North India"*, is described in Figure 5. The outcome of such query is uses different variables. By using heterogeneous query model we can represent fully structured, heterogeneous, and textual query. Textual query is a keyword query, shown in Figure 5. This query model is related to the model given by[18,22].

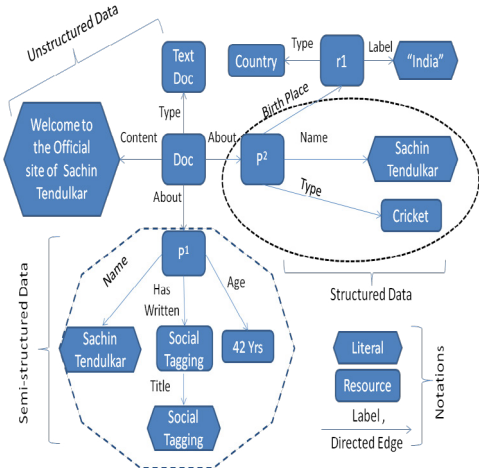

Figure 4. Illustration of the data model, with structured, semi-structured and unstructured data.

```
Select
    ?x where {
?x RDF: type ns: Cricketer          # (1)
?x {moved to North India}           # (2)
?x {has written Social Tagging}}    # (3)
```

Figure 5. Illustration of the heterogeneous query model, consists of three patterns. 1. Structure part. 2. Unstructured part. 3. Semi-structured part.

5.2 Solution for Text and Structured Data

Author in[9] explains the hybrid query for information retrieval, they have proven that hybrid query approach is good in IR.

In our heterogeneous model, we rank outcome of graphs according to the possibility of specified query q via graph g, i.e., P(g|q). By using q_s we also evaluate the shape of the outcome graphs, and these likely graphs contribute to the similar formation, then outcome will depends only on the aspect, which distinguish the outcome, using variables and their relations of textual data query q_t. Consequently, we can lower the ranking to

$$P(g \mid q) \propto \prod_{i=1}^{n} P(q_i \mid x_i), with, q_i = q_{t_j} \wedge q_{s_k}, x_i \in q_{t_j} q_{s_k}$$

$$P(g \mid q) \propto P(g).P(q \mid g) \propto P(g) . \prod_{i=1}^{n} P(q_i \mid x_i) \quad (1)$$

Computation of P (q_i\|x_i):

We can evaluate in two ways based on q is either purely textual or purely structure or not.

(a) Purely structured: If the query is purely structured, we can capture full information and then ranking will be depend on the demand of the resulting data. If a data is available in text format and the data is not understandable then rank outcome by evaluating the relation of each keyword k of the text data element to the related variable, represent in Equation (2).

$$P(q_i|x_i) = \begin{cases} \prod_{k \in q_{t_i}} \alpha P_t(k|x_i) + (1-\alpha) P_t(k) \; if \; q_t \neq \varnothing \\ \prod_{x_i} P_s(x_i) \qquad\qquad\qquad\qquad if \; q_t = \varnothing \end{cases} \quad (2)$$

We have given simple ranking heterogeneous model for combined query scenario. There may be more possible solution for the text and structured data in future.

5.3 Solution for Heterogeneous Data

Clearly, if all datasets in Dataspace are in same format and uses same schema level for storing, representing and managing heterogeneous data, then by using structured query q_s we can easily retrieve information from G_t. When datasets in are different and multiple formats, we can use, the following solutions to search data from Dataspace.

5.3.1 Keyword Search (KW)

The first and widely use solution for heterogeneous data searching make bag of words and make each query term for the representation of each object with their metadata and description for use of keyword search. Keyword search query is simple but also flexible due to same keyword query is used for Gs and.

5.3.2 Rewriting the Structured Query

Another solution for Information Retrieval crisis from Dataspace is use of database. Here, we are using the relational database to store data in Dataspace. In this case query also will be in structured format and we rephrase the structured query q_s in a query q_t that grasp to the terminology of the objective dataset G_t. In heterogeneity compute the entity detail and schema mapping using any tool. Then, predicates and constants in q_s entities in Gs and replaced these predicates and constants with.

A Dataspace management system we can find the result using 1. Keyword queries, 2. Structured queries, 3. Meta data queries, 4. Monitoring queries etc.

6. Research Challenges of Dataspace

In this section we have discuss some research challenges for future work for information retrieval and search ranking over Dataspace.

- **Top-k result calculation:** For searching data on semi structured, we require top-K results for rising effectiveness on key search from the data space. The ranking function for weighted tree size and ranking using aggregation of particular node is previously explained. Finding the top result for other outcome and ranking of these functions in Data space is an open challenge.
- **Estimation of Hybrid Keyword searching:** In this problem we represent how effectively XML or SQL keyword search is performed over a huge heterogeneous datasets and queries. We require a framework for finding Hybrid keyword search for heterogeneous data from the Dataspace.
- **Implementing of emerged views**: The efficiency of critical search engine can be improved using emerged views. The existing challenge is to use the emerged views to upgrade the processing time of query and also apply the emerged view for calculation of overall result from Dataspace or Big data.
- **By using significant feedbacks in information retrieval:** Evaluation for analyzing users interest and also user relevance response is commonly require for achieving success in search feature. Exploring the structural relationships between keywords using explicit feedbacks for XML or SQL search is completed. By using relevance feedback, implicit feedback in dissimilar aspects of hybrid keyword searching in Dataspace is the biggest asset.

There are existences of lot of opportunity for exploiting user feedback to get better search feature with reference to the heterogeneous data over Dataspace

- **Diverse Data Models:** There are multiple types of data modes exist for fetching data from structured, unstructured, semi-structured individually. Existing models target on exploring relational databases and data-centric XML information. There may be other solutions for information retrieval from heterogeneous data using keyword or phrase queries. Furthermore, develop a mechanism that permit user to access extensive collections of available heterogeneous data sources.
- **Query format for the Heterogeneous Data:** Keyword queries are commonly used and also easy for using it but it deficit the communicative power. The SQL and XQuery, are complex, demonstrative and hard to learn; There is no format of writing query for the keywords, labeled keyword search, analytical keyword queries, and for natural language query.
- We need to make an appropriate format for the heterogeneous data from the Dataspace or Big data.
- **Search Quality Improvement over the Dataspace:** Using some available work the supe-

riority of exploration results has high opinion to intended user requirement for this purpose we can use Information Retrieval field. But for the improvement of the search quality over the dataspace, the involvement of the user can improve the search engine design, such as study of query log and user click-through streams. In contrast, keyword search on structured data, unstructured, semi-structured or combination of all these i.e., heterogeneous information poses a exclusive asset on analyzing user aspect from the Dataspace or Big data.

- **Evaluation for the Dataspace Datasets:** With the rising popularity of keyword search on structured, unstructured, semi-structured data or combination of all these i.e., heterogeneous data, become an increasing need to develop an estimation framework and conduct the system blueprint to access the data from Dataspace.

7. Conclusion

In our paper we have discuss the Dataspace, Dataspace system architecture, Information retrieval from unstructured, semi-structured and structured data or combination of all three i.e. heterogeneous data over the Dataspace, which includes approaches for keyword search, and search ranking. We have given the problem definition of the Textual Data, structured Data, Unstructured and Heterogeneous data with their solution. In this paper we have also given two information retrieval data models for Textual data and Structured Data search ranking, one is heterogeneous data model and another one is heterogeneous query model with their mathematical formulations using probability.

In final section we have find out some research challenges in the area of information retrieval, search ranking of heterogeneous datasets over Dataspace. Using our approached we can find the better information using search ranking of heterogeneous data over Dataspace.

We will do result analysis in our next paper, that will based on our proposed approach.

8. References

1. Franklin M, Halevy A, Maier D. From databases to dataspaces: A new abstraction for information management. ACM Sigmod Record. 2005; 34(4):27–33.
2. Podolecheva M, Prof T, Scholl M, Holupirek E. Principles of Dataspaces. Seminar From Databases to Dataspaces Summer Term; 2007.
3. Daniel MH. Ranking for web data search using on-the-fly data integration. KIT scientific publishing; 2014.
4. Chen Y, Wang W, Liu Z, Lin X. Keyword search on structured and semi-structured data. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data; 2009. p. 1005–10.
5. Cleverdon C. The cranfield tests on index language devices. In Aslib Proceedings. 1967; 19(6):173–94.
6. Pound J, Mika, Zaragoza H. Ad-hoc object retrieval in the web of data. In Proceedings of the 19th International Conference on World Wide Web. 2010. p. 771–80.
7. Weikum G. DB&IR: Both sides now. In Proceedings of ACM SIGMOD International Conference on Management of Data; 2007. p. 25–30.
8. Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. 2010; 3(4):333–89.
9. Zhao L, Callan J. Effective and efficient structured retrieval. Proceedings of the 18th ACM Conference on Information and Knowledge Management. 2009. p. 1573–6.
10. Kleinberg JM. Authoritative sources in a hyperlinked environment. JACM. 1999; 46(5):604–32.
11. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Computer Networks. 2012; 56(18):3825–33.
12. Harth A, Kinsella S, Decker S. Using naming authority to rank data and ontologies for web search. International Semantic Web Conference; Berlin Heidelberg: Springer. 2009 Oct. p. 277–92.
13. Delbru R, Toupikov N, Catasta M, Tummarello G, Decker S. Hierarchical link analysis for ranking web data. Extended Semantic Web Conference; Berlin Heidelberg: Springer. 2010. p. 225–39.
14. Lalmas M. XML retrieval (synthesis lectures on information concepts, retrieval, and services). San Francisco: Morgan and Claypool Publishers; 2009.
15. Chaudhuri S, Das G, Hristidis V, Weikum G. Probabilistic ranking of database query results. Proceedings of the 13th International Conference on Very Large Data Bases. 2004; 30:888–99.
16. Rocha C, Schwabe D, Aragao MP. A hybrid approach for searching in the semantic web. In Proceedings of the 13th International Conference on World Wide Web; 2004. p. 374–83.
17. Bhagdev, R, Chapman S, Ciravegna F, Lanfranchi V, Petrelli D. Hybrid search: Effectively combining keywords and semantic searches. European Semantic Web Conference; 2008. p. 554–68.
18. Elbassuoni S, Ramanath M, Schenkel R, Sydow M, Weikum G. Language-model-based ranking for queries on RDF-

graphs. Proceedings of the 18th ACM conference on Information and Knowledge Management; ACM. 2009. p. 977–86.
19. Singh M, Jain SK. Transformation rules for decomposing heterogeneous data into triples. Journal of King Saud University-Computer and Information Sciences. 2015; 27(2):181–92.
20. Doan A, Halevy AY. Semantic integration research in the database community: A brief survey. AI Magazine. 2005 Mar; 26(1):83–94.
21. Lal N, Qamar S. Comparison of ranking algorithm with dataspace. Proceeding of International Conference on Advances in Computer Engineering and Application (ICACEA); 2015 Mar; p. 565–72.
22. Dwivedi S, Shiwani S. Evaluation of bitmap index compression using data pump in oracle data base. IOSR-JCE. 2014 May/Jun: 16(3):43–8. e-ISSN: 2278-0661, p- ISSN: 2278-8727.