RESEARCH PLAN PROPOSAL

Enhancing Information Access and Retrieval Techniques to improve the Effectiveness of Information Architecture

For registration to
Doctor of Philosophy

IN THE FACULTY OF COMPUTER SCIENCE

to

THE IIS UNIVERSITY, JAIPUR

Submitted By:

Anubha Jain

ICG/2010/11449

Under the Supervision of:

Supervisor:
Dr. Swati V. Chande
Professor and Principal-MCA
International School of Informatics
& Management,
Jaipur

Co-supervisor:
Dr. Ashwani K. Ramani
Professor and Chairman
School Of Computer Science
& Information Technology,
Devi Ahilya University, Indore

Department of CS & IT

April, 2011

RESEARCH PROPOSAL PLAN

1. Topic:

Enhancing Information Access and Retrieval Techniques to improve the Effectiveness of Information Architecture

2. Introduction:

Information Architecture (IA), in simple terms, can be described as "a process for describing and classifying information". IA involves the processes of planning, overseeing implementation, and managing the growth of organizational structures that provide meaningful context in support of user understanding and mediate interaction between system content and user structures such as digital libraries, web sites, databases, information retrieval (IR) systems, educational software.

Information architecture provides a proactive basis for Information Systems development. In the age of Internet, it is most commonly associated with websites and intranets. The Information Architecture Institute (IAI, 2007) has defined Information architecture as:

1. The structural design of shared information environments.
2. The art and science of organizing and labeling web sites, intranets, online communities, and software to support findability and usability. (Rosenfeld and Morville, 2006)
3. An emerging community of practice focused on bringing principles of design and architecture to the digital landscape.

An information architecture (IA) is a personnel, organizational and technology independent profile of the major information categories used within an enterprise (Dickson et al., 1985). The goal of information architecture is to enable the provisioning of right information in appropriate context to the stakeholders who need it.

Information Architecture is one of the four domains of Enterprise architecture (EA) and the most important one. Enterprise architecture (EA) is a thorough description of the various aspects related to the structure of an enterprise. Enterprise Information Architecture (EIA) is the application of information in enterprise environment.

IA Components

The information architecture components can be categorized into four broad categories (Morville and Rosenfeld, 2006):

- *Organization Systems*: How we categorize information (e.g., by subject, chronology).
- *Labeling Systems*: What we call information (e.g., scientific terminology ("Acer") or lay terminology ("maple")).
- *Navigation Systems*: How we browse through information (e.g., clicking through a hierarchy).
- *Search Systems*: How we search information (e.g., executing a search query against an index).

An alternative method of categorizing information architecture components (Morville and Rosenfeld) is comprised of:

- *Browsing Aids*: They present users with a predetermined set of paths to help them navigate the site, through menus and links. These include organization systems, site-wide and local navigation system, Sitemaps, etc.
- *Search Aids*: These components allow the entry of a user-defined query and display a customized set of results that match their queries. These are dynamic and mostly

automated counterparts to browsing aids. These include Search interface, Query language, Query builders, Retrieval algorithms, search zones, and search results.

- *Content and tasks*: These are the users' ultimate destinations, as opposed to separate components that get users to their destinations. The information architecture components are usually embedded in content and tasks. E.g., Headings, Embedded links, and embedded metadata.
- "*Invisible" Components*: Certain key architectural components are manifest completely in the background; users rarely interact with them. These components often help other components, such as a thesaurus that is used to enhance a search query.

Data and information being two valuable resources for any organization is a key area to be focused upon. The organizations are struggling with how to handle boundless legacy data and how to create taxonomies within a new product. Web has become a repository of functionality & content, physical boundaries between different kinds of information blurring rapidly. The augmentation of data has led to an amalgam of information and documents over the web. This information overload has also resulted in greater complexities and loss of control over information. The major challenge being faced by the organizations is how to guide people through the vast amount of information on offer, so they can successfully find the information they want and thus find value in the system. Information Architecture aims at the development of interfaces that facilitate the flow of useful and relevant information to the user.

Thus, access to relevant information is a key issue in IA. The Association for Information and Image Management (AIIM) does not use the term information architecture but rather the phrase information organization and access (IOA). The major concepts of IOA (AIIM, 2008) are:

- Preparing and organizing information
  - Architecture - structure and composition of a repository, information collection or individual document
  - Intelligence - enriched content (i.e., metadata, categorization)
- Accessing information
  - Search and Retrieval - querying information and obtaining matching results
  - Findability - quality of being locatable or navigable

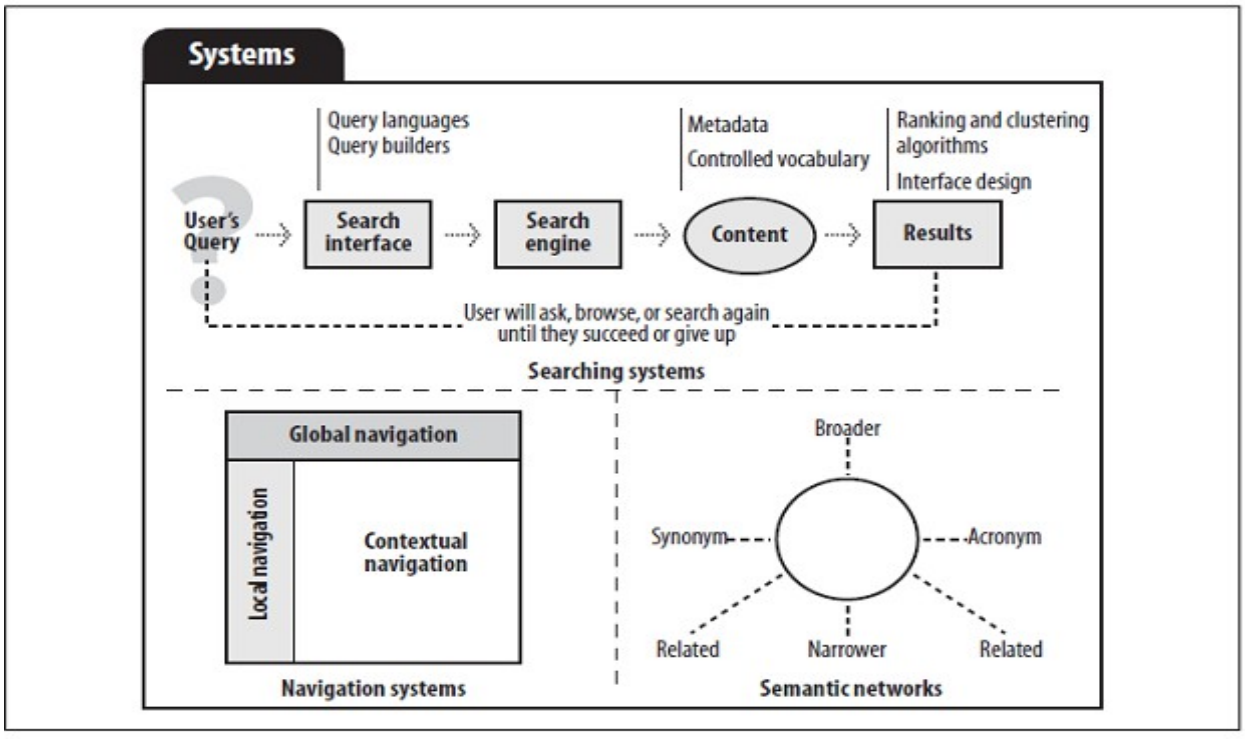

Fig 1. Information architecture Systems (Morville and Resenfeld, 2006)

As more and more organizations are putting their data over Web, they are facing the problem of information overload. Marshall (2009) observed that most organizations lag in accessing useful information from the billions of gigabytes accumulated everyday. Such organizations are 'data rich but information poor'. This is due to lack of an effective Information Architecture. The problem of information overload could be minimised by developing an effective information architecture, which provides a suitable framework for the information strategy of an organization. It will allow effective flow of information within and outside the organization; to the stakeholders as and when demanded.

An Effective Information Architecture is indispensable to an enterprise or a web based system, leading to high user experience and the goals of findability and accessibility are achieved. It is the blueprint of the site upon which all other aspects are built namely, form, function, metaphor, navigation and interface, interaction, and visual design. It is of utmost importance to create an effective Information Architecture, so that the users can step logically through a system, confident that they are getting closer to the information they require. Effective information architectures enable people to quickly, easily and intuitively find content. (Barker, 2005).

Effective information architecture is evolved from the understanding of the business objectives and constraints, the content of the system, and the requirements of the people that will use the system. An effective information architecture design depends upon: Context, Content and User.

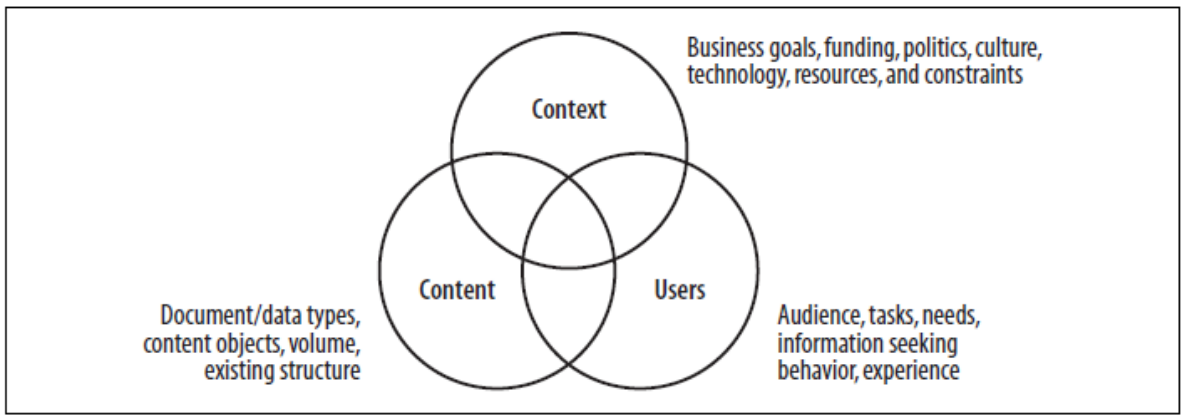

Fig 2. The infamous circles of information architecture (Morville and Rosenfeld)

Information Architecture plays a significant role in aligning the business goals with IT. It also plays a part in different managerial activities like development of DSS, business process reengineering, e-governance.

The digitally stored information is increasing in volume and variety, as well as the ways of combining information from different sources to derive insights. The most critical technological and business problem is finding the right information in this sea of information. IA has its main focus on the structure and design of web sites, such that the key aspects of usability and findability are appropriately addressed. This is because for information to be fit for purpose, it is fundamental that it is able to be found by the person who needs it.

A good information architecture design should increase the usability of a website, so that the needed information can be searched. Searching (and more broadly, information retrieval) is an expansive, challenging, and well established field (Morville and Rosenfeld), because of Web being the biggest repository of knowledge. And search is all about IR and its technologies. Thus, IR has gained a place with other technologies at the center of the stage. Information retrieval (IR) deals with the representation, storage, organization of, and access to information items such as documents, Web pages, online catalogs, structured and semi-structured records, multimedia objects. The representation and organization of the information items should be such as to provide the users with easy access to information of their interest. The research areas in IR

include modeling, Web search, text classification, systems architecture, user interfaces, data visualization, filtering, languages (Baeza-Yates and Ribeiro-Neto, 2006).

In this study, we aim at evaluating and refining information access and retrieval techniques, and thus enhance the information architecture, so as to provide accurate and timely information to the users of an information system. Information is a critical asset for any organization. An IA provides a framework which can guide decision making at all levels of management. Many researchers have pointed out that 'developing and implementing an information architecture' may be ranked as the most important IS management issue. For successful decision making, it is crucial that the right amount of information is generated and consumed, and that should be with respect to the context. This necessitates further development of technologies aimed at making sense of information spaces - so that the right information can be delivered at the right time to the right people. The refinement in the field of information retrieval and information access is an ever challenging task, which would lead to improvement of the effectiveness of information architecture. The internet, e-commerce, organizational interdependencies, knowledge management and systems thinking have helped drive Information architecture viewpoint further.

3. Review of Related Literature

Brancheau et al.(1989) presented an approach to develop global information architecture, which lays a foundation for building IA for a large-scale organizations. They explained architectural view of information, by taking example of construction architecture. They stressed the need for a target architecture that is based on business functions, organization structure and existing applications. They identified the importance of information architecture as a guiding tool for information requirements determination at all levels of management. They recognized that the data model is always incomplete because of the large size of organization. Earl (1993) also supported their argument to focus on a manageable subset of the organization that is of current relevance and interest, and then gradually expand it for the entire organization. They emphasized the need for appropriate level of accuracy, timing for business initiatives and ever-present need for commitment from senior management as key success factors for IA development. With growth in Internet technologies, another factor that contributes to success of IA development is the access and discovery of information.

Because IA has wide organizational scope, the development of IA typically calls for a comprehensive planning and design exercise generally known as strategic data planning (SDP), to guide future systems development efforts " by building " a model of the enterprise, its functions, its processes, and its underlying data (Goodhue et al., 1992). Numerous conventional IAs were developed using such data-driven methodologies such as Business Systems Planning (BSP) or Information Engineering (IE) and Strategic Systems Planning (SSP) before the network era. However, these methods were function oriented and ignored the cross-department, cross-organization and customer reach functions that are necessary in an e-commerce context. These exhibited only internal data processing needs of the firms and overlooked the information requirements from external sources. In today's global conditions, the enterprise cannot rely only on intra-organisation information but also on inter-organisation information. It is of utmost importance that the corporate workers are provided access to collections of information resources via a powerful information access framework that will result in effective information architecture.

The internet, e-commerce, organizational interdependencies, knowledge management and systems thinking have helped drive the view of information as a critical organizational asset (Evernden & Evernden, 2003). Most important is the focus on information, followed by technology, emphasizing the use and value of information content as a competitive resource distinct from the systems supporting its use.

Martin et al. (2010) stressed that IA has been rather neglected amongst the architecture domains in practice. The organizations usually implement e-commerce systems in a haste to have Internet presence, and creation of a soundly managed information base is rarely found. The authors have highlighted six domains of research issues in IA, which could help researcher as well as practitioner. These are technical issues like information quality and security, storage and retrieval, searchability and accessibility, metadata, relationship between information and process; and organizational issues like ownership, governance and organizational structure. Earl (1993) included within the domain of information architecture: determination of data storage locations, use and access, design and administration of databases, the definition and coding of data and communication protocols for interchange between organizations.

Downey and Banerjee (2011) suggest that IA should be included as a major component of architectural planning and assessment. They have prepared an IA checklist which can be used to incorporate IA into the architectural review process. The IA checklist provides a comprehensive framework for assessing the activities of information architecture of a proposed or existing system and should be embedded in the organizational processes. The authors have emphasized upon Information Organization and Information access as being most important areas of checklist. The emphasis is on three key aspects namely classification, semantics and search.

Knabe and Tunkelng (2007) identified information retrieval as a dominant paradigm for addressing the business problem of finding of relevant information. The authors discussed a few fundamental weaknesses of information retrieval, particularly in the enterprise search context. They stressed that Information access systems should seek more information from the user with reference to the information task to be performed. They discussed query refinement, aggregation and visualization as a few techniques that guide the user toward information goals. However, the authors have been confined to the techniques related to visual interaction.

Bhatia & Khalid (2008a) discuss the Web information retrieval paradigm, as a variant of classical information retrieval. Because of the heterogeneous and dynamic nature of Web, the user search behaviour & expectations change. They illustrate the basics of web information retrieval by putting light on the components, model categories, tools, tasks and the performance measures that quantify the quality of retrieval results. This paper is an introduction to the fundamental concepts related to web retrieval and the measures of evaluation.

In another paper, Bhatia & Khalid (2008b) have reviewed how Machine Learning (ML) and Information Retrieval (IR) can be applied to web mining, and how it will help develop applications that can drive the next generation of Web search with the key to support relevant search results by effectively and efficiently digging out user- centric information. Future research areas include multimedia data mining, multilingual knowledge extraction and semantic web.

PageRank is a hyperlink-based retrieval algorithm that calculates document scores by considering the entire hyperlink connected graph represented by all the links in the entire document collection (Brin & Page, 1998). This paper provides a high-level overview of an early implementation of the Google web search engine. It also highlights the importance of content matches like anchor text, plain text in large font, plain text, etc and proximity to arrive at an overall ranking. However, there is no systematic evaluation of the ranking algorithm.

Kleinberg (1999) illustrated how hyperlink information in web pages can be used for web search when using a set of retrieved documents. He showed how information inherent in the underlying network structure of the web could be utilized. Kleinberg based his model on the authorities for a topic, and on hubs for pages that link to a large number of thematically related authorities. This work was significant in providing an algorithmic approach to quantifying the quality of web pages, a key issue in the web environment where the massive size of the database, information redundancy and the uncertain quality and source of information make retrieval difficult. Further research area was of other algorithmic possibilities to facilitate the discovery of information.

Borodin et al. (2005) proposed new families of algorithms to improve the ranking of Web search results, within the hubs and authorities framework. These algorithms used a Bayesian approach, as opposed to the usual algebraic and graph theoretic approaches. They also introduced a theoretical framework for the study of Link Analysis Ranking algorithms. The framework allowed for the definition of specific properties of Link Analysis Ranking algorithms, as well as for comparing different algorithms. Future research can be to characterize the degree of similarity of the two algorithms.

Richardson et al. (2006) made use of another ranking machine learning algorithm to outperform PageRank algorithm. They used features that were independent of the link structure of the Web and other static features based on anchor text and domain characteristics to propose a resulting model. However, the experiments being conducted in the paper were biased toward pages that have higher than average quality. Also, this algorithm could only be applied to pages that have already been crawled, and were primarily useful for index ordering and improving relevance, not for directing the crawl. Further investigations can be done for a machine learning approach for crawl prioritization.

Moffat and Zobel (1996) described how to build a document index for fast text retrieval. They explained how queries are resolved using simple Boolean and ranked document retrieval methods. They also discussed fast retrieval methods using skipping for Boolean queries and reduced-memory ranking for ranked queries. They presented an alternative structure for compressed inverted indexes that allows fast query processing. The methods provided are useful in building a large scale document retrieval system and form the basis of large scale information retrieval.

Zobel and Moffat (2006) have provided a tutorial which introduces the key techniques and their enhancements underlying text search engines over the last decade. The development of a family of new index representations has led to a wide range of innovations in index storage, index construction, and query evaluation.

Callan (2000) discussed about distributed information retrieval, where the documents of interest appear scattered over multiple text databases, which can be heterogeneous along a number of dimensions, such as topical focus, intended audience, supported query evaluation models. He has presented an overview of key research addressing the main challenges in distributed information retrieval.

Moffat et al. (2007) described a term-oriented system in which partial answers are transferred between servers rather than inverted lists, in case of distributed information retrieval. Their results show that the revised method is competitive with document distribution in terms of query throughput, but has problems with load balancing and is thus not very scalable.

Broder (2007) has discussed a shift from the classical information retrieval model which supports only informational queries to navigational queries and then towards semantic queries. He has further talked about upcoming trends in Web IR. The first trend is towards context driven Information Supply to include the supply of relevant information from multiple sources without requiring the user to make an explicit query. The second trend is "social search" driven by the fact that the Web as a huge repository of knowledge and a vast social environment can be used to ask the members of a given web milieu rather than construct elaborate queries. These are future areas of research which reduce the task of the query formulation.

Hawking (2004) has focused upon the challenges of how to bring highly effective search to the complex information spaces within enterprises. He state some of the unsolved open research questions in the domain of enterprise search like defining of an appropriate enterprise search test collection, effective ranking over heterogeneous documents in an enterprise, effective search over collection of e-mails, exploiting of search context within enterprise searches. However, the author has proposed an enterprise search test collection only for solving a small sub-problem

related to navigational search, but a thorough testing was not done. The other research question paves a new direction for research.

Zhu et al. (2007) proposed methods for improving the effectiveness of enterprise search on navigational queries using dictionaries and geographically tagging intranet pages. Their approach was based on offline identification of navigational pages, intelligent generation of term-variants to associate with each page, and the construction of separate indices exclusively devoted to answering navigational queries. They have used a simple rank-merge algorithm, which has a few inadequacies. Other rank-merge algorithms should have been investigated to improve the mean reciprocal rank. Another area of research is ranking based on organizational hierarchies and job roles. Also, determining variants based on statistical analysis could be further investigated.

Demartini (2007) proposed a research plan that aims at using Information Retrieval, Semantic Web, and User Modelling techniques to improve the current state of enterprise search. He aimed to make the search experience independent from the structure and type of the searched data, and dependent on the user. Further research area can be related to expert search, and the optimum number of experts to be retrieved.

Balog et al. (2009) proposed language modeling approaches that integrate the representation, association and search of experts using various textual data sources into a generative probabilistic framework. They proposed two search strategies to find experts, which are further evaluated in the context of enterprise search systems within an intranet environment. These models provide the basic framework which can be extended to incorporate other variables and sources of evidence for better performance. However, the models used assumed uniform priors on candidate. Also, the parameter used did not give optimal result on all topic sets. The priors should be estimated and integrated in the modeling.

For expert finding, Smirnova and Balog (2011) presented a user-oriented approach that balances two factors that influence the user's choice: time to contact an expert, and the knowledge value gained after. They used the distance between the user and an expert in a social network to estimate contact time, and consider various social graphs, based on organizational hierarchy, geographical location, and collaboration, as well as the combination of these. They demonstrated the improvements over other measures, using a realistic test set.

Macdonald and Ounis (2007) made use of pseudo-relevance feedback to improve retrieval performance in ad hoc search tasks. They defined two approaches for query expansion, one based on the initial of ranking of documents for the query topic. The second approach is based on the final ranking of candidates. Further analysis can be done on relevance of feedback mechanisms.

An effective Information Architecture can guide in organizing a large amount of information to provide an optimal user experience. Spagnolo et al. (2010) extend the concept of classical Information architecture with features provided by search patterns and advanced interfaces in the case very large content intensive websites. This enhances user experience of searching and navigation to sense making, at-a-glance understanding, playful exploration and serendipitous browsing. They proposed new design strategies to develop enhanced IA which can transform (parts of) websites into Rich Internet Applications (RIAs). However, the authors did not take into consideration the content items which were hidden or shown according to several criteria. It does not order a content item as being high or low in ranking or assigning a scaled classification to. Two open research issues are classification and relevance of content items.

Yuan et al. (2009) has emphasized upon the prominence of IA methodology in the organization of information portal content. IA methodology will guide in building "user-centered and service-oriented" design. The purpose was to construct an information path which enables information visible and understandable to the users. IA integrates the information services and resources and various information systems of the campus in a manner that personalizes web interface and

provides understandability and clarity along with usability. The stages are Concept Design, Organize information content, Generate access path to information, Information display and release, Information content maintenance. This methodology is used with an aim to provide updated information and to continuously reduce upon the information silos created due to integration of different systems. However, the given methodology is not sufficient to improve upon several specific aspects related to search and personalization.

4. Motivation/Justification and Relevance:

Information architecture is a discipline, which has always been a prevailing topic of interest for practitioners and researchers. Information architecture is considered as a discipline to develop high quality software applications and web based applications. The growth in Internet also resulted in the growth of e-commerce and e-business. E-commerce and E-learning are overarching concepts that have attracted attention of stakeholders (researchers, practitioners, educational institutions and corporate institutions) for many years. Enterprise Information Architecture is playing a dominant role in shaping up of such applications.

The problems of information architecture are social issues as much as technical. They include issues of ownership, governance and organisational structure, maturity, culture, responsibility and accountability for the qualities of information in organizations, interoperability, flexibility and searchability, distributed nature of data over a network, integration problems due to differences in structure, information layout, finding of relevant information due to volume of data and variety of formats available. The technological changes have resulted in restructuring of business architecture due to rapid business changes such as mergers & acquisitions in banking industries, globalization and outsourcing, resulting in new e-business models (Kamogawa and Okada, 2004); which consider the issues of system integration and collaboration with an enterprise architectural viewpoint.

As information proliferates exponentially, finding of relevant information due to volume of data and variety of formats available is one of the prevalent issues. Good IA lays the necessary groundwork for an information system that makes sense to users. Thus, it is indispensible to have enterprise search as a component of every business application in an organization or any web site. The primary goal of an IR system is to retrieve all the documents that are relevant to a user query while retrieving as few non relevant documents as possible.

Information Retrieval is a major component in every business as well as Web. The research that is being done in IR today focuses on the Web and on the needs and challenges of businesses. Searching within an enterprise, and finding what you need, is different from that on the Web. Enterprise Information retrieval differs from Web retrieval in terms of scale, hyperlink, ranking, user roles and needs, tasks and security.

The exponential growth of digital information available in Enterprises and on the Web creates the need for search tools that can respond to the most sophisticated informational needs. Retrieving relevant documents is not enough anymore and finding entities rather than just textual resources provides great support to the final user both on the Web and in Enterprises.

However, the expectations of corporate workers are increasing due to the rising modern business demands and the ever-changing Web. In case of enterprises, as the data being more specific to given search tasks, new techniques are needed. There are still open issues and characteristics that do not allow reusing the same techniques used in Web search.

As information retrieval and access being a central issue in enterprise information architecture, an effort is made to refine the techniques of search systems being used for enterprise Applications and web sites. The literature review has shown that there is a lot of scope in the field of information retrieval and information architecture, which make a room for further research to be taken up in this area.

5. Objectives:

The objectives of the proposed research are:

- To examine the performance and shortcomings of existing information architecture, if any, within the organisation, in terms of information organization, access, and retrieval.
- To systematically review the techniques and algorithms for information access and retrieval and to refine the existing techniques and algorithms, or apply a new algorithmic approach, to improve the effectiveness of enterprise information architecture.
- To evaluate the effect of an improved search system on the effectiveness of an information architecture.
- To make use of the evolved technique(s) of information retrieval or knowledge management in enterprise Applications and web sites.
- To evaluate the information architecture on the basis of parameters such as findability, storage and retrieval, usability and customer satisfaction after the application of IR technique.
- To compare the efficiency and effectiveness of the proposed approach and the existing approaches.

6. Plan of Work and Methodology

In order to achieve the above mentioned objectives the following methodology will be followed:

- Systematic Review of Research Work: A systematic review of IR models and ongoing research work in information retrieval in enterprise would be performed, by keeping track of latest research. This would be done by reading the related articles being published in the journal.
- Analysis of Role of information retrieval techniques: The significance of Information retrieval techniques would be analysed in business applications.
- Identification of the areas of improvement: The issues related to information access which needs to be tackled, and which need further research would be explored.
- Use of information retrieval methods: The algorithms may be refined and would be applied to improve upon the effectiveness of enterprise search.
- Assessment of the proposed methods: The proposed methods would be evaluated upon various criteria.
- Experiments would be performed to check the performance of the solution, using test cases.
- Conclusion: On the basis of the result, the conclusions shall be drawn.
- Report Generation: A thesis report shall be prepared on the basis of the research work undergone. The work being proposed & the results of the research shall be published.

7. Tools and Techniques:

To carry out purposed research, a few techniques and tools shall be required for performing different tasks. A brief summary of these tools and techniques is given below. This is not an exhaustive list. During research, if a new technique is found, it may be integrated into the work.

A. Statistical Tools:

For performing comparative data analysis any one of the statistical analysis tools may be used:

- MS-Excel
- MATLAB
- SPSS

B. Graphical Tools:

For diagrams and flowcharts, any one of the tools may be used:

- MS Visio
- Gliffy
- LucidChart
- DrawAnywhere

C. Design/Modeling Techniques:

For modeling purpose, any of the following approaches may be used:

- Probabilistic or Statistical Model
- Filtering algorithms
- Ranking algorithms
- Indexing algorithms
- Language models
- Semantic Web
- Machine Learning
- Soft Computing Techniques

D. Evaluation Techniques:

For evaluating, any one or combination of the following techniques shall be used:

- Think-aloud methodology
- Test Collections
- Heuristic evaluation

8. Time Plan

The work is proposed to be carried out in following phases in an approximate time span of two years:

Activity\ Months	2	4	6	8	10	12	14	16	18	20	22	24
Literature survey												
IR techniques Review												
Refinement												
Application of IR techniques												
Evaluation & Comparison												
Thesis writing												

9. References, Bibliography, Webliography:

- Association for Information and Image Management (AIIM). (2008). IOA Workbook 1: Practitioner Level, Washington, 12.

- Balog, K., L. Azzopardi and M. Rijke (2009) A language modeling framework for expert finding. *Information Processing and Management* **(45)**, 1–19.
- Broder A. (2007). The next generation web search and the demise of the classic IR model. In *Proceedings of the 29th European conference on IR research* (ECIR'07). Springer-Verlag, Berlin, Heidelberg, 1-1.
- Baeza-Yates R. and B. Ribeiro-Neto. (2006). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd Edition), Addison-Wesley, Harlow, England.
- Bhatia, MPS, and A.K. Khalid. (2008a). A primer on the Web information retrieval paradigm. *Journal of Theoretical and Applied Information Technology*, **4(7)**, 657-662.
- Bhatia, MPS and A.K. Khalid (2008b). Information retrieval and machine learning: Supporting technologies for web mining research and practice. *Webology*, **5(2)**, Article 55. Available at: http://www.webology.org/2008/v5n2/a55.html.
- Borodin, A., G. O. Roberts, J. S. Rosenthal, and P. Tsaparas (2005) Link analysis ranking: algorithms, theory, and experiments. *ACM Trans. Internet Technoogy.* **5(1)**, 231-297.
- Brancheau, J.C. and J.C. Wetherbe (1986) Information Architecture: Methods and Practice, *Information Processing and Management,* **22(6),** 453–463.
- Brancheau, J. C., L. Schuster, and S.T. March. (1989) Building and Implementing an Information Architecture. *ACM SIGMIS Database*, **20(2)**, 9–17.
- Brin, S., and L. Page (1998) The anatomy of a large-scale hyper-textual web search engine. *In Proceedings of the 7th WWW Conference*, Brisbane, Australia, 107-117.
- Callan J. (2000) Distributed information retrieval, In *Advances in Information Retrieval* (W. Bruce Croft, editor), 127– 150. Kluwer Academic Publishers, New York.
- Demartini G. (2007), Leveraging semantic technologies for enterprise search, In *Proceedings of the ACM first Ph.D. workshop in CIKM*, Lisbon, Portugal.
- Dickson, G. W., and J. C. Wetherbe (1985), *The Management of Information Systems,* McGraw-Hill, New York.
- Downey, L., and S. Banerjee. (2011). Building an Information Architecture Checklist. Journal of Information Architecture. **2(2)**, 25-42. Available at: http://journalofia.org/volume2/issue2/03-downey/.
- Earl M.J. (1993) Experiences in Information Systems Strategic Planning, *MIS Quarterly* **17(1),** 1–25.
- Evernden, R. and E. Evernden (2003). Third-generation information architecture. *Communications of the ACM*, **46(3)**, 95-98.
- Goodhue, D.L., L.J. Kirsch, J. A. Quillard and M.D. Wybo. (1992) Strategic Data Planning: Lessons from the Field. *MIS Quarterly,* **16(1)**, 11-34.
- Iain Barker, (2005) *What is Information Architecture*. Available at: http://www.steptwo.com.au/papers/kmc_whatisinfoarch/index.html
- Information Architecture Institute (IAI). *(2007) What is IA*. Available at: http://iainstitute.org/documents/learn/What_is_IA.pdf.
- Kamogawa, T. and H. Okada (2004). Issues of e-business Implementation from Enterprise Architecture Viewpoint. *In International Symposium on Applications and the Internet Workshops* (SAINT 2004 Workshops). 47 - 53.

- Knabe F. and D. Tunkelang. (2007). Enterprise information access and user experience, *IT Professional*, IEEE Computer Society, **9(1)**, 21 - 28.
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment, *Journal of the ACM*, **46(5)**, 604-632.
- Hawking D. (2004). Challenges in Enterprise Search. In *Proceedings of the 15th Australasian database conference* **(27)**. Australian Computer Society, Inc., Darlinghurst, Australia, 15-24.
- Macdonald, C.and I.Ounis (2007) Using Relevance Feedback in Expert Search, In Advances in Information Retrieval, Lecture Notes in Computer Science, Springer, Berlin **(4425)**, 431-443.
- Marshall, R. (2009) Global digital data growth beats expectations. Available at: http://www.v3.co.uk/v3-uk/news/1995604/global-digital-growth-beats-expectations.
- Martin A., D. Dmitriev and J. Akeroyd. (2010) A Resurgence of Interest in Information Architecture. *International Journal of Information Management*, **30(1)**, 6-12.
- Moffat A. and J. Zobel (1996) Self-indexing inverted files for fast text retrieval. *ACM Trans. Inf. Syst.* **14(4)**, 349-379.
- Moffat, A., W. Webber, J. Zobel and R. Baeza-Yates (2007) A pipelined architecture for distributed text query evaluation. *Journal of Information Retrieval.* **10(3)**, 205-231.
- Morville P. and L. Rosenfeld. (2006). Information architecture for the World Wide Web (3rd ed.). O'Reilly Media, CA.
- Matthew, R., A. Prakash and E. Brill (2006) Beyond PageRank: machine learning for static ranking. In *Proceedings of the 15th international conference on World Wide Web* (WWW '06). ACM, New York, USA, 707-715.
- Smirnova, E. and K. Balog (2011) A User-Oriented Model for Expert Finding. In *Advances in Information Retrieval*, Lecture Notes in Computer Science, Springer, Berlin, **(6611)**, 580 – 592.
- Spagnolo, L., D. Bolchini, P. Paolini and N. Di Blas (2010) Beyond Findability. *Journal of Information Architecture.* **2(1)**, 19-36. Available at: http://journalofia.org/volume2/issue1/03-spagnolo/.
- Yuan F., Ping X, Liu Q., and Fu X (2009). Digital Campus Information Portal Content Organization Based On "Information Architecture". In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human* (ICIS '09). ACM, New York, USA, 1330-1334**.**
- Zhu H., S. Raghavan, S. Vaithyanathan, and A. Löser. (2007). Navigating the Intranet with High Precision. In *Proceedings of the 16th international conference on World Wide Web* (WWW '07). ACM, New York, 491-500.
- Zobel J. and A. Moffat (2006) Inverted files for text search engines. *ACM Computing Survey.* **38(2)** Article 6.