Importance of User Studies for Information Retrieval: A Case Study of Assam Agricultural University, Jorhat

Nitumika Gogoi[1] and Uma Kanjilal[2]

Abstract

User studies over the years have attempted to explain the information use phenomena, understand and predict information use behavior and improve the information search results by manipulating essential conditions. Librarians are exhorted to develop traits in identifying the distinct groups of users along with their information-seeking habits, sources, and needs. These, in turn, facilitate the successful evaluation of search results in Information Retrieval. The present user study on a sample of 184 users is highlighted as one of the essential tasks that lead to search-oriented query collection from a heterogeneous group of users in the Assam Agricultural University and consider the specific databases for the Information Retrieval process. The users' satisfaction level can judge the attitudinal approach to information search after retrieving information relevant to the queries submitted by the users.

Keywords: User Study, Queries, Information-Seeking Behavior, Information Retrieval, Agricultural Library

Introduction

The continuous monitoring of user behavior determines the degree of library-user interaction during the information retrieval process. The changes in the information search pattern are dependent on factors like information needs and the implementation of new services in the library. In terms of the usage of library resources, user behavior presents many overlapping issues that reflect on the incapability of the users lacking in the specificity of the search queries, consequently ending up with vague

[1] Research Scholar, Asstt.Librarian, Rev.B.M.Pugh Library, Assam Agricultural University, Jorhat
[2] HOD, Faculty of Library & Information Science School of Social Science, IGNOU, New Delhi

Received Date: 29.01.2022 Accepted Date: 05.02.2022

and general search strategies. Besides implementing services in the library, promotion and publicity of these services to the users must also be emphasized to avoid dispensing non-essential information to the users during information retrieval. Many documents on information retrieval validate that the standard approach to IRS evaluation revolves around the degree of relevancy of documents. As a rule of thumb, a minimum of 50 (fifty) information needs is realized to be the minimum requirement for deciding upon the relevance factor. A document is said to be relevant if it addresses the stated information need not just because it contains almost all the words in the query but due to certain stated statistical measurements applied in the various information retrieval models.

Hence "relevance" is assessed relative to an information need, not a query. Hence when the information need is framed improperly, it is difficult for the IRS to process and produce the relevant result as expected by the user.

Objective of study

The objective is to study the user community of agriculture and allied fields to draw inferences on the taxonomy of different user groups and their information-seeking approaches. To highlight the interdependence between the user profile and query collection and justify the selection of databases i.e., CeRA and DELCON for the Information Retrieval Process.

Scope and limitation of the present study

The study throws light on the attitudinal readiness to information use on which further research can be attempted. Also, the stratification of the user groups of agricultural information in the study can be considered for any research involving the users.

As for the concerned limitations, since the queries collected for the Information Retrieval Process are considered research-oriented queries, so out of the undergraduate and postgraduate students, only the postgraduate students are considered for the study as they are assigned with preparing

dissertation a compulsory credit course. Secondly, the Assam Agricultural University constitutes the colleges of Agriculture, Community Science, Sericulture, and Horticulture within its campus. The Colleges of Veterinary Science and Fisheries are located in different districts. Therefore maximum data received comprises users of the in-campus colleges and sparsely from users from Veterinary and Fishery.

Literature Review

Wilson (1981), in his model of factors influencing the needs and information-seeking behavior, opines that the full range of personal human needs is at the root of motivation towards information-seeking behavior. He suggests that as part of the search for the satisfaction of these needs, an individual may engage in information-seeking behavior. However, many factors other than these needs are also recognized to play a part in the information retrieval process. Gwizdka & Lopatovska (2009) defined "subjective aspects of information searching that include all aspects related to users' feelings and perceptions (in contrast to their actions directed towards the system in achieving their goal)." They included studies of searchers' feelings associated with the search stages, searchers' satisfaction, and relevance judgments (Kuhlthau, 1991; Tessier, Crouch, & Atherton, 1977; Su, 2003; Saracevic, 2007 as cited by Gwizdka & Lopatovska, 2009). Their primary goal was to examine searchers' behavior while monitoring changes in their 'subjective' states, specifically the effects of changing search circumstances on their feelings and experiences. Therefore users' information-seeking behavior depends mainly on the types of users. In the present context, the users of agricultural information are discussed in detail to make the base for performing IR nearly appropriate.

The perception of library users happened to be constricted in types, but the approach in seeking information is realized to be unchanged, as witnessed by Ramachandra and Somasekhar (1979) in their work "Bibliographical organization of literature in agricultural sciences" where they stated that the agricultural users form a single user group. Although

the subject interests of such users overlap, their information needs have much in common. A significant point was highlighted by Ford (1980), who defined the information needs in terms of conceptual incognitas, the parameters used to correlate the user interests with cognitive, social psychologies. They stated that the satisfaction of such needs is attributed to the access to varying ranges of information sources, from individual learning sources to large-scale databases. Guha (1983) quoted Herbert Menzel, who grouped the user need studies into three main categories.

- Communication behavior studies in which surveys are conducted to investigate the pattern of the overall intention of the user community with the communication system, without reference to any specific information receiving event;
- Use studies in which surveys are conducted to find out the use of any communication medium, such as primary or secondary sources and
- Information pattern studies in which surveys are conducted to find out the pattern of flow of information in the communication system.

In the book "Agricultural Librarianship in India," Subbaiah, R. (1988) mentioned that the users of agricultural information can be classified as end-users and intermediaries. According to him, the end-users are the various types of persons connected with agriculture in one way or the other such as teachers, students, scholars, extension workers, farmers, entrepreneurs, etc. while the librarians, documentalists, even publishers and distributors act as the intermediaries who make the information available to the end-users. French, B.A. (1990), speaking on the taxonomy of users in "User needs and library services in agricultural sciences," stated that a relative dearth of real knowledge about the users of agricultural information exists, which is further complicated by the expanding scope of agriculture. He mentions Metcalfe's views published in 1981 who proposed eleven categories of users arranged into two groups viz. primary literature users and secondary/tertiary literature users. Based on various studies on the types of users, French classified and explained the users under five

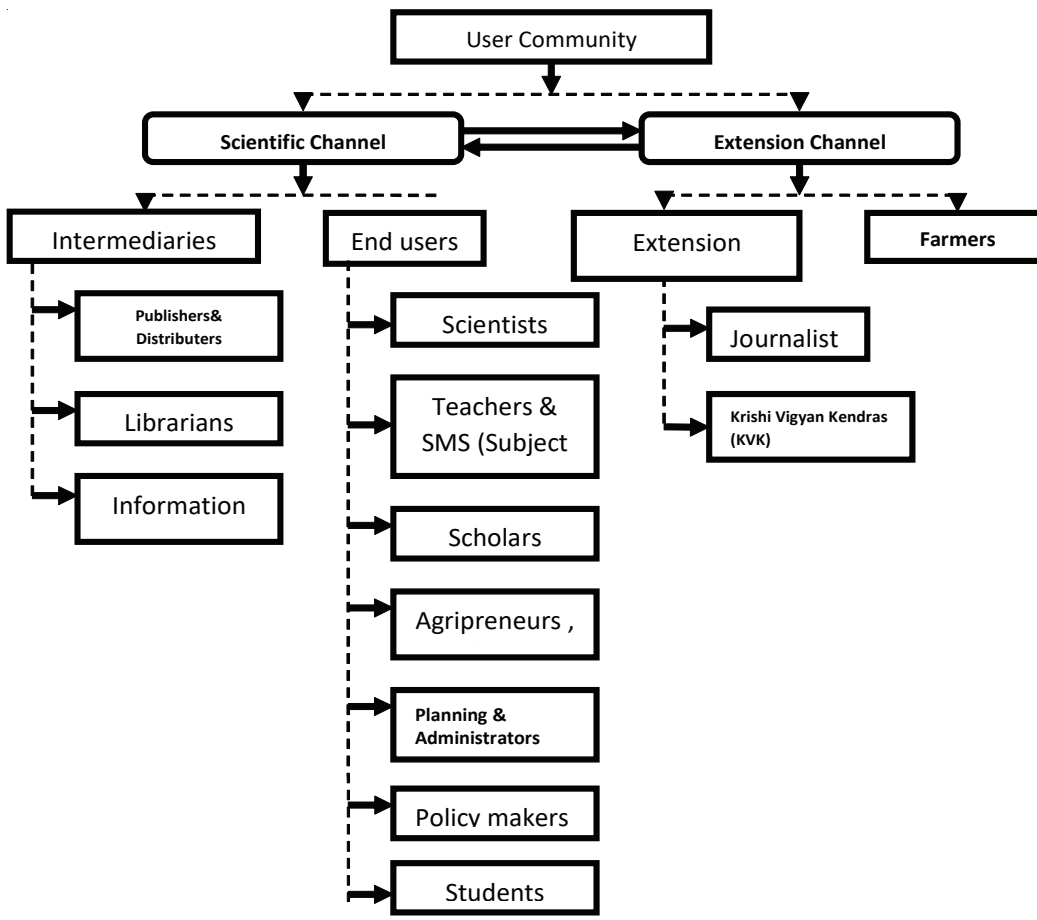
categories with some variations and overlap viz. scientists and researchers, extension scientists or researchers, extension advisers, farmers/producers, and others like policymakers, educators, journalists, agribusiness workers, and consumers. According to Gadge, Sane, and Kekre (2015), end-users present additional troubles in the information retrieval process by submitting an unstructured query. The authors stated that the "user's main concern is the methods to procure the pertinent and accurate information from the web. The meticulous conveyance of content is subject to user preferences and interpretation." Some of the numerous reasons for failure to receive relevant documents by the users can be cited from the works of Sahami et al. (2004), Baeze-Yate (2003), and Ravikumar and Singh (2010) as follows:

1. Information needs are frequently imprecisely defined, generating a semantic gap between user needs and their specifications (Sahami et al.; Baeze-Yate).
2. The user queries are limited to a couple of words and the users often do not have the foggiest idea about the best query to retrieve the information they require (Ravikumar and Singh).
3. There are various types of users and they have their perspectives and interpretation. Besides, the user's needs for information change with time. (Baeze-Yate; Ravikumar and Singh).

Al-Maskari and Anderson (2010) considered user satisfaction as a subjective variable that can be influenced by several factors such as system effectiveness, user effectiveness, user effort, user characteristics, and expectations. They suggested that information retrieval evaluators should consider all these factors in obtaining user satisfaction. They validated that most studies are done before theirs have conflicting conclusions on the relationship between user satisfaction and system effectiveness. This study has substantiated this relationship and supports using user satisfaction as a criterion of system effectiveness.

Stratification of the User Community of Agricultural Information

The agricultural universities that span across many disciplines viz agriculture, veterinary, sericulture, fisheries, horticulture, and community sciences apart from imparting agri-based education play important roles in extension services to the common man specifically to farmers and are also involved in government schemes and projects. The Krishi Vigyan Kendras (KVKs) under the administration of the universities are involved in conducting training programs for the farmers and producers, extension activities like field days, field visits, diagnostic field visits, and rural exhibitions as well as work in collaboration with the departments and NGOs. The user community classified as end-users and intermediaries forms an integral part of disseminating agricultural information.



The methodology adopted for the user study in the present work

A survey-based study with a questionnaire as the tool for user data simply does not fulfill the purpose of determining the user information needs. A questionnaire can only be considered one of the many methods of data collection involved in the user study. Rocio Herrera C. et al. (1987) suggested that developing appropriate techniques and methodologies for user studies is necessary because most libraries hitherto have not been fully conversant with the problems and techniques associated with such studies. Moreover, such studies often result in a considerable collection of disparagingly interpreted data, more likely from inadequate and poorly chosen samples. The questionnaire is supplemented with an interview because a small sub-sample should be interviewed to check the validity of the answers given to the questionnaire.

(a) Hypothesis

H_0 =There exist no significant difference in the approach of the different categories of users w.r.t.

1. Usage, fulfilment of the purpose of their use, and their satisfaction level from the different types of library services and facilities. Their primary interest in using the services is acquiring the appropriate information based upon their need during information search.
2. Usage of the different e-resources subscribed and purchased by the library

H_1 =There exist a significant difference in the approach of the different categories of users w.r.t.

1. Usage, fulfilment of the purpose of their use, and their satisfaction level.
2. Usage of the different e-resources subscribed and purchased by the library.

(b) User Categories

From the researcher's perspective, the possible user groups in the agricultural university can be categorized as

- Scientists and researchers
- Extension scientists and subject matter specialists (SMS)
- Faculties and scholars
- Students (PG)

These four categories of users are considered for query collection mainly because they are the constant users of the library and fulfill the first criteria of the scope i.e., taking into account the digital collection in the library that serves the purpose of the title of this study. The study was conducted by adopting the purposive sampling method and data collected through a survey-based questionnaire containing several closed-ended, open-ended, and multiple-choice questions designed to collect the relevant research data from the registered users of the Central Library of Assam Agricultural University. "Google forms" are used to prepare the questionnaire, and the interaction and conversation are jotted down later to derive the conclusion. Likert's five-point scale was used in designing the questionnaire.

Data Collection

The sampling done was purely purposive as the primary goal was to evaluate the respondents' awareness about the e-content the library has subscribed, purchased, and available through CeRA; their knowledge about the databases viz Krishikosh DelCON, India Agristat. The variables in this study are measured using a 5 point Likert scale with the following variations: agreement, quality, and likelihood to allow for correlations. In this regard, five different factors are proposed with the following parameters and their respective items:

Factors	Parameters	No. of items
1	Library visit	8
2	Library infrastructure, ambiance & ICT facility	12
3	Library services (usage)	8
4	Users & information seeking pattern (purpose & satisfaction)	8
5	e-resources & database	6

(c) Response Trend

A sample of 184 users depending on the class attendance was considered. The following findings are drawn from 161 respondents, out of which 13 responses for missing data and 9 responses for being casual and identical are removed. The final datasets decided to be used for analysis are 139 (161-13-9=139) categorized as

1. Post Graduate Students = 73
2. PhD Scholars = 35
3. Faculties (teachers/ scientists/SMS etc) = 31

Table-1: Demographic data

Variables	Number	Percentage
Questionnaire distributed	184	100
Questionnaire returned	161	87.5
Usable questionnaire	139	75.54
Categories of respondents		
PG Students	73	52.51
Research Scholars	35	25.17
Teachers/Scientists/SMS	31	22.3
Total respondents	139	100

The collected data from the users was organized, tabulated, analyzed, and interpreted using the following tools and techniques: frequency distribution, percentages, summated mean scores through the weighted index and ranking techniques, MS-Excel, and the chi-square test with a significance level of 0.05 i.e., ∞=0.05 for which we accept the null hypothesis (H_0) framed for this study when (x^2)tabulated > (x^2)calculated

Result and Analysis

The collected data was analyzed and interpreted to draw the results and test hypotheses framed for this study.

Visit the Library

a. No. of regular visitors: 63 (students and scholars); Reason for visit expressed as $\mathbf{R_1>R_2>R_3>R_4}$ where

   R1= Preparation of notes for class, assignment, research activities, articles for subject-related magazines, daily agriculture-based newspaper articles, self-help groups, etc

   R2= Spending time in the e-resource section

   R3= Preparation for competitive exams in the banking sector, agricultural officer, NET-JRF, Ph.D. entrance, etc

   R4= Regular newspaper reading

b. No. of frequent visitors: 27 (faculties, extension scientists, and scientists); Reason for visit expressed as $\mathbf{F_1>F_2>F_3>F_4}$ where

   F1=Newspaper reading (interest and liking)

   F2= Gaining awareness with the recent trends, changes, and implementation in the agricultural and allied sectors) through the newsletter, reports, and other updated records

   F3= Class preparation in digital mode (prefer the e-resource and computer laboratory section)

   F4= Group discussion (faculty and a few of his/her students) and other queries

Usage, the purpose of use, and satisfaction level w.r.t the library services

From table 2, it is evaluated that under the "usage" criterion, it is seen that 80% and above users of all the three categories use the circulation service, DDS, photocopy service and the online service as all the users are always in need of prompt information; students are focused in preparing class notes, assignments, literature search for their dissertation topics as well as photocopying of the reference collection and fractions of dissertations that are not for loan; research scholars are more occupied and concentrated with their research and investigation as well as cracking of competitive exams due to which it can be ascertained from the above table that online service, DDS, CAS, reference service, CD ROM service and photocopying service are in high demand; Faculties for whom teaching, supervising and research are part and parcel of their academic pursuit have additional responsibilities of projects, agricultural camps, extension service to the other categories of users involved in agriculture viz. farmers, entrepreneurs etc. and therefore their usage statistic of online service, reference service and CAS are in high demand. Regarding "purpose," when compared among the three user categories based on the allotted ranks, it is seen that the circulation service and online service have best served the purpose of use. Furthermore, it can again be analyzed from the ranking that reference service, CAS, DDS, etc. are towards the decreasing mode in terms of the purpose served. When the "satisfaction level" of all the three user categories is compared based on the ranks, online service hits the top of the ranking list. In contrast, other services show the minimum difference in meeting the satisfaction of all the users. Therefore, it can be interpreted that all the three categories of users do have very much common in the services when usage, purpose served, and satisfaction level is taken into account, differing only in the degree of importance in information seeking level. Hence it can be assumed there is no significant difference in the approach for seeking information, and therefore, the null hypothesis is accepted.

Table - 2: Usage, the purpose of use, and satisfaction level w.r.t the library services

Services	Usage			Purpose			Satisfaction		
	P.G. students	Research Scholars	Faculties	PG	Research Scholars	Faculties	PG	Research Scholars	Faculties
	F (%)	F (%)	F(%)	mean(rank)	mean(rank)	mean(rank)	mean(rank)	mean(rank)	mean(rank)
Circulation service	69(94.52)	35(100)	25(80.64)	4.203(1)	4.17(1)	4.08(3)	3.94(4)	3.91(3)	3.4(7)
Reference service	46(63.01)	31(88.57)	27(87.09)	3.07(7)	3.22(8)	3.11(8)	3.52(7)	3.33(8)	3.26(8)
CAS	53(72.60)	33(94.28)	30(96.77)	3.6(5)	3.75(5)	3.8(6)	3.96(3)	3.9(4)	3.63(5)
DDS	61(83.56)	33(94.28)	25(80.64)	3.44(6)	3.69(6)	3.84(5)	3.81(5)	3.63(6)	4.08(2)
Photocopy service	73(100)	35(100)	23(74.19)	4.03(3)	3.91(4)	4.34(1)	4.01(2)	3.88(5)	3.56(6)
Catalogueue service	53(72.60)	29(82.85)	17(54.83)	2.9(8)	3.55(7)	3.58(7)	3.71(6)	3.48(7)	3.76(4)
CD-ROM database service	35(47.94)	31(88.57)	13(41.93)	3.8(4)	4.06(2)	3.92(4)	3.42(8)	4.06(2)	3.92(3)
Online service	64(87.67)	35(100)	31(100)	4.06(2)	4(3)	4.19(2)	4.96(1)	4.91(1)	4.77(1)
Chi-square calculated value	13.98			0.12			0.2		
Degree of freedom	14			14			14		
Significance level	0.05			0.05			0.05		
Chi-square table value	23.68			23.68			23.68		
H_0 accepted/ rejected	Accepted			Accepted			Accepted		

Use of e-resource collection

The interpretation that can be drawn from the analysis in Table 3 below is that the mean frequency of usage of e-resources and the ranks suggest that CeRA is most effective in catering to the information needs of the three user categories. The PG and scholars use Krishikosh, but they almost wind up their search in the Theses and Dissertation section because the library has held a strong collection since its inception. On the other hand, the faculties suggest the above two user categories to browse Krishikosh. Also, from the data collected, it was found that scholars and faculties have introduced themselves to all the e-resources mentioned above and databases as compared to PG students. As CeRA and DelCON have nearly fulfilled their search, IndianJournals.com is hardly used by the users. In this case, the null hypothesis is accepted as has been derived from the "x^2-test" that there exists no significant difference in the use of e-resources by all the three categories of users.

Use of e-resources & databases

e-resources/database	**PG mean(rank)**	**Research Scholar mean(rank)**	**Faculty mean(rank)**
e-books	3.93(2)	4.06(2)	4.65(2)
CeRA	4.58(1)	4.86(1)	4.74(1)
DelCON	2.88(5)	3.89(3)	4.06(4)
Krishikosh	3.18(3)	3.74(4)	3.32(5)
IndiaAgriStats	2.95(4)	3.57(5)	4.26(3)
IndianJournals.com	1.47(6)	2.26(6)	2.97(6)
Chi-square calculated value	0.5		
Degree of freedom	10		
Significance level	0.05		
Chi-square table value	18.307		
H_0 accepted/rejected	Accepted		

I. Information gathered from oral communication

Although it may be realized that it is not judgmental to include the oral communication between the researcher and the user group "extension scientists of KVKs" in a study purely based on query-document effectiveness in Information Retrieval, it has, however been felt necessary to state that focus on research on the ground level issues will prepare a strong base for an efficient Information System. Following issues that show scope for necessary initiatives to be undertaken are

- The present agricultural trend demands high competition in the agricultural market, and hence agriculture has become an information-intensive occupation. It is, however, understood from the conversation with the extension scientists that the farmers they work with,
- Lack of confidence in the accuracy and reliability of the information because they like being stuck to the age-old traditional methods of agriculture and financial and climatic issues sometimes create serious barriers. The KVKs thus provide their central source of information.
- A significant setback for the extension scientists and the farmers is linguistic communication, both verbal and nonverbal, which is otherwise the benefit of lingua franca enjoyed by the scientists and researchers
- Translating information to local languages and successfully disseminating and applying information is another time-consuming impediment faced.
- The collection of information in the local language is way far less than the ratio of the information available in English. Adding up is the wave of e-resources, both national and international, posing a more significant technological challenge.
- This focuses on the need for selective translation and dissemination services. To improve users' quality of information-seeking nature as

one of the highly considered visibility indicators, various innovative steps need to be adopted. In this regard a few obstacles identified are:

- The imbalance between document delivery capabilities and bibliographic capabilities. Bibliographic capabilities being more developed proves that the accessibility to agriculture outreach the other branches of sciences. However, from interactions with the users, they report difficulties obtaining the information when needed.
- Examining bibliographic information retrieval as an approach to measure user satisfaction. In this regard, two principles need to be adopted, viz., the selection of the data sources and the precision of the automatically extracted resources. Precision in this context means the relevance of displayed resources must be of the same subject coverage as the article selected by the user.
- The users' online information search tends to focus on precision and recall and the purpose of the request rather than on the structure and content of a particular database.

II. Interpretation

Following were the interpretations that are drawn from the data gathered with regards to users' approach to accessing information

- Most users start their search with a search engine and not preferably with the library portal
- Prefers to browse and read online
- Purchased and subscription-based e-resources of the library are equally accessed

From table 3 "Use of e-resource collection", it is decided that for e-resource retrieval, CeRA and DelCON databases will be taken into account for Information Retrieval Process for query-document pairs.

The outcome of the study

1. The search factor "Usage of the online databases" from the questionnaire acts as the criteria in selecting the databases from which documents are to be retrieved based on the frequency of use by the users.
2. The criteria viz. usage, purpose, and satisfaction from library services and the use of different e-resources are intentionally taken into account to make the selection of databases for retrieval of documents based on the queries collected from the users much easier time-saving. This is hoped to make the central theme of the present research study on "The application of document similarity measures for e-resource collection in Agriculture and its allied fields" comprehensive.

Conclusion

An exhaustive user study based on the pattern of search strategy, search platform, and search techniques makes most of the retrieval successful, decreasing the probability of search failures. At any point of information search, the user community should realize the demand for the librarian's guidance. Tonta (1992) has concluded in his study that "user input should be sought when analyzing search failures with retrieval effectiveness measures such as precision and recall. The same can be said for failure analysis studies that are based on user satisfaction measures. We should strive for full-scale user involvement as much as possible in every stage of analysis of search failures. Despite user participation in the evaluation process, search failures in document retrieval systems are unlikely to be eliminated altogether

The present study is a section followed by the query classification and preparing the search process to be accurately implemented with the application of the Vector space model of Information Retrieval, the main topic of doctoral research undertaken by the researcher.

References

Al-Maskari, A. and Sanderson, M. (2010). A review of factors influencing user satisfaction in information retrieval. Journal of the Association for Information Science & Technology, Association for Information Science & Technology, vol. 61(5), pages 859-868, May.

Baeze-Yate, Ricardo. (2003). Information Retrieval in the Web: Beyond Current Search Engines. International Journal of Approximate Ranking. Vol 34 pp 97-104.

Ford, N. (1980). Relating information needs to learner characteristics in higher education. Journal of Documentation, 36(2). Pp. 99-114.

French, B.A. (1990). User needs and library services in agricultural sciences. Library Trends, vol.38 no. 3 pp. 415-41

Gadge, Jayant, Sane, Suneeta, Kekre, H.B. (2015). "Performance Analysis of Layered Vector Space Model in Web Information Retrieval". International Journal of Applied Information Systems. Foundation of Computer Science. New York, USA, Vol 8 no. 5 feb.2015. ISSN 2249-0868

Guha, B. (1983).Documentation and information. Calcutta, World Press, pp.61

Gwizdka, J., & Lopatovska, I. (2009). The role of subjective factors in the information search process. Journal of the American Society for Information Science and Technology. 60(12). (doi:10.1002/asi.21183)

Hensley, C.B., Sowarby, A.J. and Resnick, A. (1962). Selective Dissemination of Information. IRE Transactions of the Professional Group on Engineering Management, EM-9:2.

Herrera C.,Rocio, Lotero M. , Libia and Rua R., Ivan (1987). User studies in university libraries. International Reader in the Management of Library, Information and Archive Services.

Ramchandra, H.R & Somasekhar, K.T. (1979). Bibliographical organization of literature in agricultural science. AALDI Bulletin, 7: 24

Ravikumar, P. & Singh, Ashutosh K. (2010). Web Structure Mining: Exploring Hyperlinks and Algorithms for Information Retrieval. American Journal of Applied Science. Vol 7(6), pp 840-845.

Sahami, Mehran et al. (2004). The Happy Searcher: Challenges in Web Information Retrieval. PRICAI 2004: Trends in Artificial Intelligence pp 3-12.

Subbaiah, R. (1988).Agricultural Librarianship in India: An overview. New Delhi

Tonta, Yasar (1992). Analysis of Search Failures in Document Retrieval Systems: A Review. The Public-Access Computer Systems Review 3, no. 1 pp 4-53.

Wilson, T.D. (1981). On user studies and information needs. Journal of Documentation. Vol. 37, pp.3-15.