FEDERATED SEARCH AND Z39.50 AS INFORMATION RETRIEVAL STANDARD

Pravin N. Doibale
Research scholar S.R.T.M.U Nanded
Enail- pdoiale@yahoo.in Mob no-8698180671

ABSTRACT:-

It is the simultaneous search of multiple online databases and is and emerging feature of automated, web based library and information retrieval systems. It generally called as portal or web based search engine. Some time called as the broadcast search, parallel search, cross reference search etc. Various terminology are used for these tools in the literature including, cross searchers, cross database searchers, portals, broadcast searchers or parallel searchers, and meta searchers.

KEYWORDS:- federated search, Z39.50, standard

INTRODUCTION:-

Today's world is the digital world. So that generation of information is more than the need. In various ways the creation of information occurs that information is useful for the research purpose and general purpose, this information can be retrieves' by the researchers and by users, for their use. The searching of information is the big task because specific information searching requires specific keywords or specific query to search, then we can search particular information on specific topics. There for we need different techniques to search particular and specific information, one of the techniques is the federated search in which we can search parallel and cross search this technology useful for the researchers for their research because it gives the multiple information resources simultaneously through one search query. So the user can view search results in a single integrated list. In different words users not need to consult information on resource individually. Instead they can search various library catalogs web site, subscription and citation databases all at once. In this assignment I have describes what is federated search, and Z39.50 is the information retrieval standard.

General meaning of federated search:-

Federated means - Content is combined from different sources saving the effort of searching sources one at a time. Federated search is an information retrieval technology in which the simultaneous search of multiple searchable resources. Single query of users which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for view to the user. Federated search technology is an integral component of an information portal, which gives the multiple information resources. When the user put a search query in the search box of the search engines the system uses federated search technology to send the search string to each resource that is incorporated into the search engines. The individual information resources then send the search engines a list of results from the search query. User can view the lost of documents retrieved in each resource and link directly to each search result. Then we can say shortly the federated search is the search for information in multiple information resources through a single query. view of single list, link to directly to resources.

What is federated search:- it is the simultaneous search of multiple online databases and is and emerging feature of automated, web based library and information retrieval systems. It generally called as portal or web based search engine. Some time called as the broadcast search, parallel search, cross reference search etc. Various terminology are used for these tools in the literature including, cross searchers, cross database searchers, portals, broadcast searchers or parallel searchers, and meta searchers. The term Meta searcher is particularly prominent and is the term adopted by the United States national information standards organization. Federated searching is the function of search tools which search a number of databases, particularly subscription databases, simultaneously with one interface. The content search by federated

search tools is content that could not normally be searched via a web search engine.
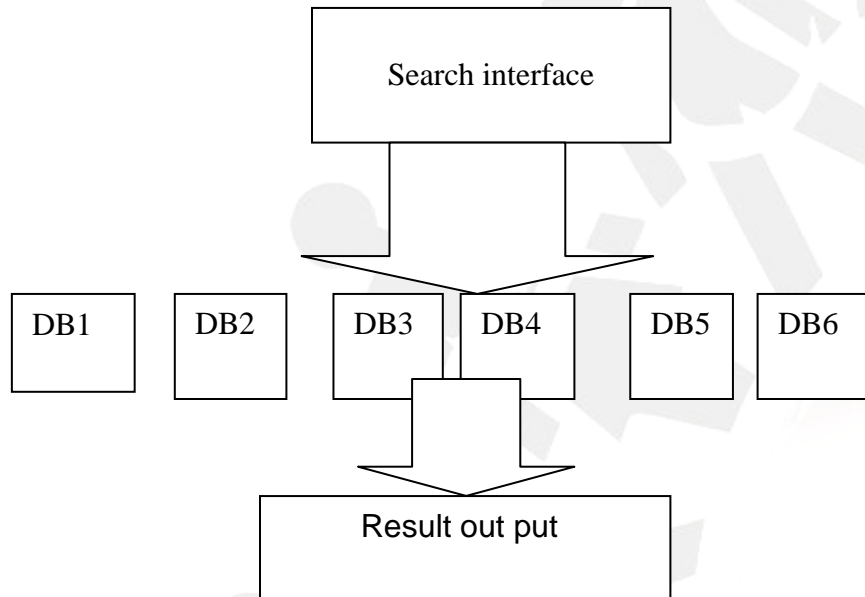

Fig. 1 federated search model

Definition of federated search:-

A. **Peter Jasco defines federated search as**, "Transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax, merging the results collected from the databases, presenting them in a succinct and unified format with minimal duplication, and allowing the library patron to sort the merged result set by various criteria".

B. Federated search is the process of performing a simultaneous real-time search of multiple diverse and distributed sources from a single search page, with the federated search engine acting as intermediary

Let's look at the key words in the definition and their influence on the value of federated search:

- **Federated** - Content is combined from different sources saving the effort of searching sources one at a time.

- **Simultaneous** - Federated search queries all user-selected sources at once. It would be unacceptably slow if it waited for all of the results from one source before querying the next.

- **Real-time** - Federated search occurs live and results are current. There's no stale content.

- **Multiple** - The value of federated search to the researcher increases as the number of sources increases.

- **Diverse sources** - Federated search engines typically can search sources containing documents of different types, e.g. PDF, Word, PowerPoint. The process of extracting text from documents of different types is hidden from the user.

- **Distributed sources** - Federated search engines expect to search content that lives in different locations.

- **Single search page** - Federated search engines provide a single point of searching.

Need and Purpose of Federated Search Tools:-

The need and purpose of the federated searching are as follows:-

- Tremendous growth of databases and different suppliers, with numerous interfaces and logins means that library users can find it confusing when attempting to access information.

- Library users confusing when trying to access information due to the library OPACs and web-pages have been alienating users with their use of library terminology and due to long list of databases.

- The demand of students using academic libraries. The growth of different types of databases, produced by different suppliers, with numerous interfaces and logins means that library users can find it confusing with attempting to access information.

There are certain purposes which can be served by the federated search are as follows:-

- Transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax.
- Merging the results collected from the databases,
- Presenting them in a succinct and unified format with minimal duplication,
- Providing a means, performed either automatically or by the portal user, to sort the merged result set.

Advantages of Federated Search:-

There are certain advantages of using Federated Searches. Some of them are as follows:

- It's required less time to do a basic search is benefit enough.
- Various ways searching across all sources.
- Advanced and simple search cam be occurs.
- It shows integrated results which are easy to view and use.
- Its gives direct links to the native source for advance searching.
- It's have the ability to filter, save, print sort, and e-mail search results.

Applications of federated search include:-

- Mednar.com - Searches medical information sources.
- Biznar.com - Searches business-related sources.
- WorldWideScience.org - Searches science content from all over the world, from government agencies, as well as other quality research and academic organizations.
- http://search.smartlib-bibliogen.ca/zengine?VDXaction=ZSearchSimple - Searches Capital Smart Library Consortium of Libraries.

- http://osulibrary.oregonstate.edu/metafind/about.html - Searches Oregon State University's Library.
- Scitopia.org - Searches digital libraries of leading science and technology societies.

Discussion about Z39.50 as information retrieval standard:-

Basic information about Z39.50:-

The Z39.50 standard, which is defined simply as a standard that enables two computer systems on a network to communicate for the purpose of information retrieval. The Z39.50 standard for information retrieval is important from a number of perspectives. While still not widely known within the computer networking community, it is a mature standard that represents the culmination of two decades of thinking and debate about how information retrieval functions can be modeled, standardized, and implemented in a distributed systems environment. And importantly it has been tested through substantial deployment experience. The governing body of the Z39.50 is the LC, the Z39.50 having various standards.1,2, and 3.

What is a Standard?

A standard represents an agreement on how to do something or carry out some activity to arrive at predictable results. All standards published by the National Information Standards Organization (NISO) are developed by a consensus process that draws on the expertise of implementers and vendors, product developers, and users of those products; they are approved by the American National Standards Institute (ANSI). Z39.50 is one of many NISO standards that address the application of both traditional and new technologies to information management, retrieval, and storage.

Definition of Z39.50:-

- LYNCH,CLIFFORD A. :Z39.50 properly "Information Retrieval (Z39.50); Application Service Definition and Protocol Specification, ANSI/NISO Z39.50" – is a protocol which specifies data structures and interchange rules that allow a client machine (as called an "origin" in the standard) to search databases on a server machine (called a "target" in the standard) and retrieve records that are identified as a result of such a search.

- FINNIGAN, SONYA AND WARD, NIGEL :ANSI/NISO Z39.50 – 1995 (ISO 23950) is one of a set of standards produced to facilitate the interconnection of computer systems. The standard specifies formats and procedures governing the exchange of messages between a client and server, enabling the user to search remote databases, identify records which meet specified criteria, and to retrieve some or all of the identified records and is concerned, in particular with the search and retrieval of information in databases. One of the major advantages of using Z39.50 is that it enables uniform access to a large number of diverse and heterogeneous information sources.

What is Z39.50:-

Z39.50 is an international standard protocol used by networked computer systems for information retrieval. It enables information seekers to search different systems on a network or the Internet through the use of a single user interface. Software and system vendors offer access to information from a diversity of unique systems with different hardware, software, interfaces, and database search commands. Compounding matters for the information seeker, the Internet provides access to a mind-boggling array of databases that grows daily. The challenge for users becomes how to find the right information painlessly amidst this vast array. The goal of Z39.50 is to reduce the complexity

and difficulties of searching and retrieving information. Z39.50 makes it easier to use the wealth of information resources on the Internet. When using Z39.50 enabled systems, a user in one system can search for electronic information in another system without having to know how that system works.

A Short History of Z39.50:-

Z3.9.50 come out of the linked systems project in 1980s for standardize searching of the major bibliographic databases of OCLC, many libraries in USA working in parallel to the LSP initiative was the standardization effort around an information retrieval protocol for library application under the auspices of NISO. Protocol develops by LSP moved to NISO and developed into the Z39.50 information retrieval standard, approved as a NISO standard in 1988. The Library of congress develops the official maintenance agency and registration authority for the Z39.50 standard. So the LC provides information pertaining to the development and maintenance of the Z39.50 existing and future versions as the implementation and use of the z39.50 protocol. A group called the Z39.50 implementers Group (ZIG) assumed a primary role in ongoing development. In conjugation with the Z39.50 maintenance Agency, the ZIG developed version2 and 3 of the Z39.50 protocol in 1992 and 1995. Version 1 defined the core services of Z39.50. Version 2 formalized the structure of information to be exchanged based on the ISO standard data description language and encoding rules known as ASN.1 and BER. With consensus on Z39.50 version 2, vendors began to create applications rapidly and momentum grew for compliance with Z39.50 functionality. Version 3, the current version of the standard, builds on and includes version 2. Version 3 is extremely powerful in its support for simple to highly complex applications. The current version of the Z39.50 standard is formally known as ANSI/ NISO Z39.50-1995, Information Retrieval— Application service definition and protocol specification. The

International Organization for Standardization (ISO) approved two information retrieval standards in 1993 (ISO 10162 and ISO 10163-1). In 1998, ISO adopted the Z39.50 protocols and issued ISO 23950 Information and documentation - Information retrieval (Z39.50), withdrawing the two previous standards.

How Z39.50 Works:-

In this process the server is known as standard as the "**target**". A client is known as an "**origin**" in the standard. User selects the target library from the menu.

- Enters search terms, the query is then sent to server by client.
- The client's user interface will help to build the query and will translate it in the proper syntax.
- There is a primary negotiation between the client and server.
- This particular query will ask the server for records where the terms are found.
- The server translates the search request for the targets library's database and receives a response about the numbers of machines. The client's user interface will tell how many records are in the set but not transferred at this point.
- If the client decides to ask for the records in the result set a transfer is initiated. Client receives the records.
- Records get presented in the interface for the user. The protocol provides generalized facilities for the communication of queries and results. One important feature is the attribute set. This allows terms and several of their characteristics to be defined. They capture semantics of a particular area. Z39.50 Protocol also uses generalized search syntax. The user application will convert queries accepted on the interface into the form required by Z39.50; these will be communicated from client to the server. t the

Page 153

server end , another application will take this and convert them into an appropriate form for running against databases. Results are communicated as number of records, a number of record formats are acknowledged.

Features Provided By Z39.50:-

Initialization:- it provides the establishment between client and server, means for initiating option that are to be used through the reminder this includes the degfault character set default language. And protocol.

Searching: - searching a more databases using a structural query using a well known search format. The query may contain Boolean operators, fielded search terms, proximity searching.etc.

Presenting Records:-

An extensive means of accessing information from a set of search results is provided through the protocol. This includes requesting specific ranges of search results, specific elements in records, specific variants in records, search term highlighting etc.

Maintaining Multiple Search Results:-

Z39.50 provides the capability of creating, naming storing and retrieving from one or more search result sets. This also provides facility to client to apply a search criterion to previously created results set.

Browsing:-

Z39.50 provides the ability to browse a window of index term or fields within the database.

Sorting Of Results:-

Z39.50 offers a means to sort a set of search results based on any given sort criterion.

Controlling Access:-

Not only does Z39.50 enable authentication on per-session basis but it also allows Authentication on a per-operation basis for the cases where the access to specific databases or records is Controlled.

Controlling Resources:-

Z39.50 provides a means for clients to cancel a search or a presentation request in the middle of an operation while continuing to maintain an open session with the server. It also permits clients to request resource reports that include accounting information on the number of searches, retrievals etc performed by the user.

Extended Services:-

Z39.50 provides the ability to Performs database maintenance operations, such as database updates, record insertion deletion etc. It also includes persistent result sets, queries and period queries.

Implications for the Libraries;-

Z39.50 protocol is main communication standard for library systems.

OPACS:-

Z39.50 Protocol to web gateways has been around for a few years. They allow OPACS to be available through web. Z39.50 Protocol provides access any and all the of the world's major library catalogues or just locate sources with a single search.

Cataloguing:-

Z39.50 searching for and downloading bibliographic records is becoming simple and efficient since multiple sources can be searched simultaneously and records easily compared. Currently libraries are often\ "Looked in" via service agreement and proprietary software to a b bolographic utility's Z39.50 Protocol will allow user to establish relationships with a variety of sources without penalties.

Union Catalogues:-

It helps for the union catalogue. The Z39.50 allows libraries within dissimilar catalogues to be grouped together without having to physically replicate their databases. A user may sit at OPAC screen and search several catalogues simultaneously Useful material and its location can be displayed with no additional work.

Limitations of Z39.50:-

1) Lack of agreed structure for representation of local holdings and availability data which is crustal to the completion of most end user searches i.e. user wants to know not that a remote library has a copy of given item in its catalogue but whether the item itself is available and at which location.

2) Loss of branding so that the originator cannot guarantee acknowledgement on the end-users display. This could make Z39.50 approach unacceptable to some commercial companies.

3) Complexity and thus high overheads to use.

CONCLUSION:-

It is concluded that the federated search is the Meta search or parallel search technique useful for the information retrieval. Its give the facility to users in single query user can search directly the information. Also the Z39.50 is a standard for client architecture in which a search engine and interface are divided into independent parts. When the client and server conform to the standard then the Z39.50 client can search any brand of Z39.50 server. Same cline can search widely dispersed databases on different native systems. The linkage of library systems to the internet and the maturation of the Z39.50 protocol offered the prospect of accessing an ever increasing array of bibliographic databases and full text databases through the local automated systems. This ability to directly link users with resources that represented different computing platforms reinforced the attractiveness of the Z39.50 protocol for libraries engaged in linking inter institutional or multivendor systems.

REFERENCES:-

i.   Arguello, J. Federated search in heterogeneous environments. *ACM SIGIR Forum*, 78.

ii.   George, C. A. Lessons learned: usability testing a federated search product. *The Electronic Library*, 5-20.

iii. Joint, N. Federated search engines and the development of library systems: ANTAEUS. *Library Review*, 653-659.

iv. Kraner, M. The Z39.50 information retrieval standard. *Computing & Control Engineering Journal*, 143.

v. Lynch, C. A. The Z39.50 Information Retrieval Standard. *D-Lib Magazine*.

vi. Shokouhi, M. Federated Search. *Foundations and Trends® in Information Retrieval*, 1-102.

vii. www.niso.org

viii. http://www.dlib.org/dlib/april97/04lynch.html

ix. http://old.cni.org/pub/NISO/docs/Z39.50-brochure/50.brochure.part02.html

x. http://archive.ifla.org/VI/5/op/udtop3/udtop3.htm