

# Federated Search

---

## IMERGE Consulting, Inc

---

*Bernard Chester  
March/April 2008  
bchester@imergeconsult.com  
As Appeared in e-Doc Magazine  
www.imergeconsult.com*

As organizations grow and/or combine, they collect legacy systems, often built using different technologies with their contents organized in different ways. This complicates locating and integrating information, and unifying an organization. Getting a global hit list for organizational information isn't as easy as Google Web search leads many users to expect.

Data warehousing tools link databases and permit answering complicated and multi-source questions. While challenging, unstructured information is an order of magnitude more difficult. The variety of content formats and storage systems, and the absence of a universal unifying standard, like SQL, hinder a similar approach. How can we tie this information together, so we can answer simple questions like "show me all of the documents that relate to a particular account" or "all the places where a product referenced?" The answer is found in federated search systems (FSS). Like Web search products, these tools will take an inquiry and bring back a list of matches from a large list of possible sources. But federated search systems take a different approach than most Web searches.

Web search sites, like Google and Yahoo, are continually crawling the Web, searching for new or changed information, and updating a master index of their contents. Pages are rated based upon techniques such as textual analysis and examination of references to or from it. The master index is used to quickly resolve inquiries, but it must be understood that it is always out of date, since it uses the contents as of the last examination. To resolve an inquiry in real-time would be impossible, due to the number of items that would need to be referenced. The same challenge occurs when you use the MS Windows® search feature—if you haven't enabled indexing on the drive, then it must look at every file. By contrast, a federated search engine will take your request and send it out to as many different repository and search systems as it is configured to interface with. The search will get transformed (as best as is possible) to match the indices and searching mechanisms of each target. When the results are returned from each system, they are merged into a consolidated result set, and organized according to some criteria. By exploiting the capabilities of servicing repositories, an FSS can utilize non-content information like metadata and relevancy ranking. However, by working with a non-homogeneous set of applications, federated search systems face a number of challenges that a Web search does not:

- **Application integration features.** Federated search systems need to be able to integrate with a range of sources, often built using out-of-date technology. This could range from screen scraping to database interfaces to using messaging and transaction systems. Content access mechanisms need to be supported to prevent the FSS from becoming a backdoor to restricted information. A FSS needs to address error conditions such as when system does not reply in a timely manner to a request, or has a connection problem that needs to be resolved.
- **Data standardization.** In order to transcribe queries and results between the FSS and the different servicing systems, an FSS needs to support metadata mapping and transformations. In addition, it will need to have rules for handling situations where a query term is not available—should it exclude the query, exclude the repository, or just proceed without the term?
- **Evaluating and merging result sets.** While an FSS could concatenate the result sets it obtains, that isn't very useful. It needs to be able to merge the results based upon metadata and relevance to the request. While document repositories may return relevancy rankings, they only refer to relevancy within, and may be not be useful when comparing the results with those from another system. Elimination of duplicate hits will clarify the choices. However it is generally impractical for the FSS to examine all of the documents that have been located.

- **Providing access to information stored in a variety of places.** An FSS results screens needs to provide a mechanism for the user to obtain identified documents. This may involve a built-in viewer or invoking the owning product's interface.

In summary, a FSS consolidates multiple document storage systems with a "hyper-engine" that maintains no indices of its own, but instead relies upon the capabilities of all the linked systems.

*--Bernard Chester, CDIA+, ICP, edp, (bchester@imergeconsult.com) is a consultant on ECM who focuses on implementation and integration issues. He is a principal with IMERGE Consulting (www.imergeconsult.com), an independent ECM consulting firm. imergeconsult.com), an independent ECM consulting firm.*