

# One search to search them all?

Martin White, Managing Director, Intranet Focus Ltd

*[martin.white@intranetfocus.com](mailto:martin.white@intranetfocus.com)*

This article is perhaps more technical than usual, but as well as a chunk of search technology, I am covering some important information management issues which may end up on your desk in the near future. Probably much sooner than that, if you are using O365.

In the initial stages of a project it is quite common for a client to say (more in hope than anything else) that it would be 'really good' to have just one search box to search across all applications. The vision is for 'Silo-Breaker Search' that provides a single point of access to all the information resources in the organisation. In principle this is an admirable vision, but in practice it is very difficult to deliver. The technology is often referred to as 'federated search' and many vendors claim they can deliver on this requirement.

## **Aggregated or federated**

On the way to federated search [aggregated search](#) (a PDF download - 'Aggregated Search: A New Information Retrieval Paradigm' - from the OATAO open access repository) is worth a mention. A search query presents a list of links to content items (usually referred to as 'information nuggets') that the software assembles into a page for display. An information nugget can be in multiple formats (text, image, video, etc.) and granularity (document, passage, word, etc.). Each of these formats is defined as a 'vertical'. This is the approach behind the 'search cards' you see in Google and Bing. The selection of nuggets, and establishing an order of presentation that takes into account the outcomes of the pages that users clicked after their initial search, is not as easy as it might seem.

The objective of [federated search](#) is to query multiple sources of information and present the user with the most relevant to their requirements. This approach is widely used in e-commerce comparison sites but translating it to enterprise search is a significant challenge.

## **A brief introduction to federated search technology**

There are two approaches to providing a federated search. The first is 'One Big Index'. In principle it is possible to crawl and index any number of individual search applications, or

business applications with a search component, and create a single index. That is not difficult. What is difficult is creating a ranking list of results that make any sort of sense to the user, including presenting them in a consistent way. The results lists are also likely to be quite long, given the size of the index, and delivering results with high precision is very difficult.

The second is 'Lots of Simultaneous Searches'. The query is managed by one search application, which then sends out the query to other search applications within an application, such as a library management database or an intranet. Results from all the applications are then either integrated, or more usually presented in several different sections of the search results page. It is not sensible to produce a 'ranked' list of results as ranking cannot be calculated as an algorithm of the ranking in each of the individual applications.

Both options require the use of 'connectors', which are pieces of software which convert the query into a format applicable to each application and then return the result to the master search application.

### **Microsoft goes enterprise-wide**

This article was sparked into life by an announcement from Microsoft that it was now offering one hundred or more [connectors for O365](#) so that (in principle at least) an organisation using O365 can search almost any other repository or application in the organisation. Wow! That's a real advance and it will not be long before your IT Team come down to tell you that you can throw away all your other search applications and just use O365.

Slow down and contain your excitement!

Connectors are challenging to write and maintain. A small change in configuration in one of the queried applications may end up disconnecting the connector. When a connector between two search applications fails (though often they just fail to perform as expected) then there is always an interesting discussion between the vendors concerned about which end of the connector has failed. In many cases what will happen is that a third-party vendor gets involved. For example, you may use [Box](#) as a document file management application. Microsoft offers connectors from [BAInsight](#), [Raytion](#) and [Accenture](#). How will the choice be made and by whom? When it does not work as expected, do you blame Box, the connector vendor, or Microsoft. Happy days!

Here are some questions you need to ask IT:

1. Are they able to show that the connectors they have selected work in other organisations using the same combination of applications? If not, then you need to consider if you want to be the first to do so.
2. What will the user interface look like? It is difficult to present more than ten results on a screen and now IT are suggesting that it will display results from many more information collections. Ask for some believable screen mock-ups.
3. Will the facets that you are using on O365 (date, author, department etc.) work the same way across all the other applications? If this is not the case users are going to find it challenging to handle the large results list that will inevitably arise with searching multiple applications.
4. How easy will it be to analyse the search logs, especially as you will need to optimise the search against each individual application and the total collection?
5. Are there defined owners of the other applications that you can work with to ensure that information and data quality are good enough and that security permissions are maintained?
6. If one of the applications goes down will that affect the overall search experience? Indeed, will the user even realise that the application has gone AWOL?

There are many more but probably the most important of all is what is the business case for federated search. In my experience the number of times that an employee needs to search across all the applications in your organisation is actually very small. So, what user research has IT undertaken to justify the complex implementation and management of federated search?

This is where an information management strategy (and of course a digital workplace strategy) becomes so important. A federated search implementation must add value to the user experience and to the achievement of business objectives. With such a strategy you will have a framework either to push back against the connector onslaught or to identify where federated (and for that matter aggregated) search applications will transform the performance of employees and the organisation.