Research Note 02-14

Search Strategy A-Z List of Topics

Martin White
Intranet Focus Ltd.

September 2014

Introduction

Only a minority of organisations have a search strategy. However as search becomes more widely recognised as a business-critical application for decision and task support it will be increasingly important to document a strategy that will have an impact on the work of most, if not all, employees.

Every organisation takes a different approach to strategy development and it is not possible to set out a recommendation for the structure of a search strategy. The aim of this Research Note is to set out a check-list of the 40 topics that should be included in a search strategy. Information on most of the topics can be found in my book Enterprise Search. The second edition of this book will be published in April 2015 and will include more advice about search strategy development than the first edition.

For convenience this Research Note refers to an individual search application. In reality organisations will usually have multiple search applications, many of which are embedded in applications such as Customer Management Relations, Document Management or Records Management. Even these embedded applications should be included in a corporate search strategy.

At the end of the Research Note is a checklist of all the headings, with a column that can be used to decide on what topics are relevant, a column for the person responsible for preparing the section and column to show that the work has been completed. In reality it is unlikely that more than two or three sections are not relevant.

Of course a search strategy should not stand on its own. Ideally it should reflect the information management strategy of the organisation, especially with regard to content quality, metadata management and user requirements. It should also be seen as a companion to an intranet strategy, where search is usually of significant value but is often not addressed in any level of detail.

Intranet Focus Ltd
12 Allcard Close
Horsham
RH12 5AJ, UK
+44 1403 26707

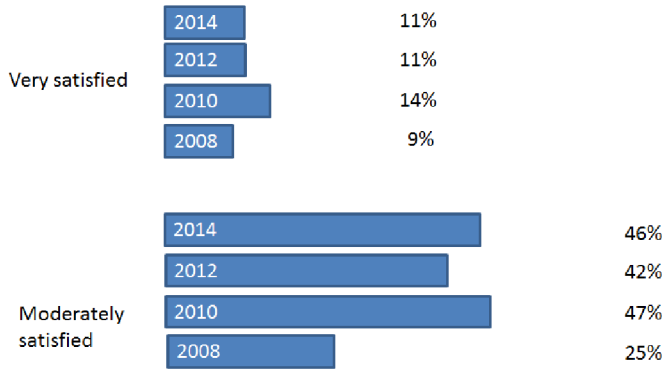
Contacts
www.intranetfocus.com
contactus@intranetfocus.com

Search Strategy A-Z

Analytics

Search logs are a good example of big data, given the volume of data, the speed of change and the variety of the queries that will be recorded. Analysing the data sets, presenting the information and then if required taking action to improve the search experience require a considerable amount of patient and detailed work.

Over the last six years the results from the Global Intranet Trends report and more recently the Digital Workplace Trends report indicate that organisations with a high level of search satisfaction are in the minority, to the importance of establishing the extent of dissatisfaction and the reasons for it are likely to be a major element of the work of the search team.



From The Digital Workplace in the Connected Organisation 2014 © NetStrategy/JMC and used with permission http://www.digital-workplace-trends.com

A typical analytics plan might be

Weekly

- Crawl and indexing performance, and up-time

Monthly

- Assessment of top 100 queries
- Assessment of zero hits
- Assessment of new search terms

Quarterly

- Survey of specific user groups
- Analysis of user feedback and help desk calls
- Usability testing

Annually

- Track search satisfaction as an element of an employee engagement survey

Architecture

This section has to be the responsibility of the IT department, as it should set out the server architecture for search applications. This is not just about ensuring that where needed there will be development and production servers but also about the extent to which indexes will be managed on a distributed basis and the location and availability of document and other

repositories. Distributed index management is a common issue with SharePoint 2013 implementations. The objectives are to ensure that performance service levels are achieved and that there is an appropriate disaster recovery plan in place.

Best Bets
In certain circumstances best bets can be of value to a search user. Any search for a hazardous chemical might usefully bring up a company policy on chemical risk management. Best bets often seem to be a reaction either to a problem with the intranet information architecture or poor metadata which results in important documents being pushed so far down the list of results that most users will not find them.

Best bets not only need to be chosen with care but need to retain their credibility for the long term. Each best bet should have a specified owner with a responsibility for ensuring that the best bet is indeed the best bet for a particular query or set of queries.

Big Data
The hype around Big Data has now subsided but the topic has caused many organisations to look at how data and information can be accessed, analysed and used to make good business decisions. Enterprise search and Big Data should be seen as complementary business applications. In the 2014 AIIM survey on search implementation there seems to be an indication that search is now a higher priority than content analytics. Many of the underlying core elements of search, Big Data and eDiscovery are common and so let themselves to being managed by a search centre of excellence.

Budget
The rapid adoption of open source search applications has inevitably put pressure on commercial search software vendors to reduce their pricing or to change their business models. As a result even high-end search applications are significantly less expensive than a few years ago. It is however very difficult to get an initial indication of the cost of a search application from a vendor or a systems integrator. To some extent that is understandable because of the range of variables involved. However the costs of search implementation are increasingly related to people costs, either in terms of open source software development and the requirement for an appropriate level of support for the search application post-implementation.

Business Cases
In an ideal world there would be a single business case for search based on the importance of information as an asset to the organisation and the need to provide effective access. Rarely is that the case. A good alternative is to set out a set of business cases, each of which make a contribution to meeting the business objectives of the organisation and make a case for a direct or indirect contribution to meeting the costs of the technology investment and the on-going staff costs for maintenance and support. These business cases might typically developed around

- Decision support
- Risk reduction
- Knowledge management
- Customer service
- On-boarding
- Acquisition

Each business case should be owned by a stakeholder and the stakeholder should be either a member of, or represented on, the Search Steering Group. This creates a robust governance structure for search and ensures that changes to business requirements and search capabilities are kept in step.

Business Impact
Although it is important to monitor the technical performance of the search application and the use that is being made of it through analytics, in the final analysis the issue is whether the application is having an impact on the business. This requires a search team that has the time and resources to survey, and more important talk to, users about the business impacts they are achieving through high-quality search.

Cloud Search
There are often very good business reasons for adopting cloud-based applications but it is important to understand the potential limitations of cloud-based search applications. At present they may not have the full functionality of the on-premise versions and may not index the entire content of documents. There can also be some latency, support and disaster recovery implications. These will undoubtedly be addressed in the future but should be clearly set out in the strategy.

Communications
Search applications will be among the most heavily used enterprise applications and yet in most organisations the search team and the actions that are being taken to improve search performance are all but invisible. These actions may well be invisible to many, if not all, search users as they may be subtle changes in the ranking of particular document sets or queries.

At the core of the communications strategy should be a dot-release approach to search improvement across all applications. The current version should be designated 2.0 and then a programme of enhancements can be constructed around 2.1, 2.2, 2.3 etc. The release level should be shown on the search home page. In this way users can be assured that enhancements are taking place and the communications programme can be used to highlight the changes and the potential impact on the user experience and user satisfaction.

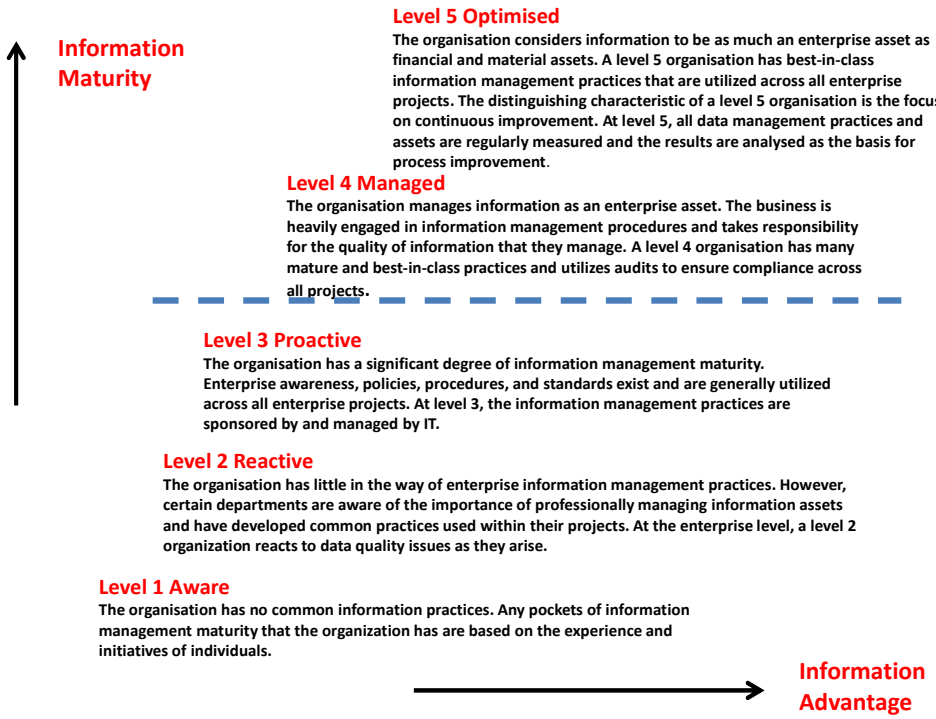
It may be useful to consider a search team blog in which enhancements to the applications can be outlined and advice given on how to get the best out of search.

Connectors
Connectors are a very technical element of the search architecture but are a very common point of failure for search and that is why they have a specific heading in this list. Connectors enable a search application to crawl and index content on a specific file server or application. They can be very susceptible to small configuration changes to either the search application or to the server being crawled and indexed.

For example the intranet may just be on a single web server but the people search application is on an Oracle HR application. To index the Oracle HR application may require the use of a connector. Connectors also tend to be version specific, so that a change in the version release of the Oracle HR application could require a new connector to be installed and tested. Connectors can be expensive to purchase and need constant attention to make sure that they are working correctly. They also connect applications developed by two different vendors and tracking down where the problems lie if the connector seems not to be working correctly can be a difficult discussion which has to be managed by getting the appropriate experts from both vendors around the same table. It can be very time-consuming.

The purpose of this section of the strategy is to define the requirements for connectors and the source of the connectors in use.

Content Quality
Ideally a search strategy should be framed within the context of an information management strategy that sets out standards and guidelines for content quality. In 2000 the Meta Group developed a five level information maturity model.



The dotted line in this diagram marks the point between Levels 3 and 4 where the business takes responsibility for managing information. Arguably organisations operating at Levels 1 – 3 are unlikely to be able to achieve high levels of search performance and satisfaction because the quality of content will be less than optimal and very variable.

There are a number of areas where setting out content quality guidelines can have a significant impact on search performance in areas such relevance ranking, the optimisation of filters and facets and the design of user interfaces.

These guidelines should include as a minimum

- Defined authorship to confirm the authority of the content
- Dates that are unambiguous in format and definition
- Consistent metadata, ideally based on taxonomies and controlled terms
- Protective marking to manage secure access
- A review period so that users can be certain that older content is still relevant
- Informative titles enhance ranking effectiveness and are easy to scan

Content Scope

This section should documented in reasonable detail the content that is being crawled and indexed so that there is no dispute about whether this is the case. Note that this list should include both the content (e.g. project reports for 2008-2012) and the server on which this content resides. This is because it is not uncommon for file sets to be moved between servers and this list should enable these changes to be managed.

Crawl Management

The crawl schedule and the content scope need to be matched together within the context of user expectations and the rate of change of business-critical content being added to a particular server. The complexity of crawl management is well illustrated by the section in Microsoft TechNet on crawl management for SharePoint 2013. Most other commercial and open source search applications will have a similar crawl management capability.

Development Plan
The development of search applications never comes to an end. There will always be changes in the requirements of users as a result of changes in business objectives and the search application will itself offer additional functionality over time. This is where the dot-release approach mentioned in the section on Communications can be of assistance in not only setting out a programme of development but also indicating dependences.

Disaster Recovery
Disaster recovery plans are related to business requirements and are usually expressed by using two measures: Recovery Time Objective (RTO) and Recovery Point Objective (RPO). RTO and RPO requirements are derived by determining the downtime cost to the organisation if a disaster happens. In the case of search working out what the costs might be and therefore what the best option for disaster recovery might be is not an easy calculation to make.

There are three related issues

- Loss of the search application service.
- Loss of index integrity
- Loss of crawled content

The concept of crawling is unique to search applications and the implications of either the search application server or a content server crashing on the extent to which content has been crawled and indexed may take quite a bit of unravelling. Establishing a disaster recovery strategy should involve a substantial number of "What if" scenarios. One of the particular challenges with search is that it may not be at all obvious to the user that a 'disaster event' has taken place.

eDiscovery
eDiscovery relates to a formal search to identify content and documents that relate to a particular incident, customer, contract or risk. This is somewhat broader than the use of eDiscovery in the support of legal actions, especially in the USA. It is not just about finding the content but about being able to place a hold on this content and documents so that no changes are made. The underlying technologies of eDiscovery are similar to those for search though very few vendors operate in both the eDiscovery and enterprise search markets.

Emails
Email content features in both enterprise search and eDiscovery and can be a very contentious topic as employees may feel (incorrectly!) that their email exchanges with colleagues, let alone external contacts, is confidential to the individual. Searching emails is not easy because of the number of 'content items' that are involved. In an email discussion each email will be repeated many times and that can have a major impact on relevance determination.

Feedback
Providing effect feedback channels for search users is important as it could be that just a very small number of users (perhaps in strategic planning roles) are experiencing problems finding content that they know is available. There could be many reasons for this, including inappropriate security permissions in this particular example. Users have to be assured that the search team welcome feedback and will respond with an appropriate degree of urgency. In effect there needs to be a service level agreement on the period within which the user can expect a response. Clearly there needs to be the resources to respond, investigate the nature of the problem and come up with a solution.

Governance
In the majority of organisations the search budget is held by IT. However if the organisation has recognised the importance of a search support team the question that has to be answered is who is going to pay the personnel costs. Search needs to be directed from both a strategic and operational perspective by a cross-departmental Search Steering Group with representation

from major stakeholders and from IT. The Search Manager should into this Group for direction on priorities and budget management even if the line management responsibility for the Search Manager is held by someone outside of the Steering Group. Inevitably this can end up with some complicated reporting lines.

One way to understand the governance requirements is to list out the decisions that are likely to have to be made over the next year and then work through who is going to take these decisions and what information they need to ensure that the optimum decision is being taken. Then the people with the responsibility for the decision and for providing the information should be involved in the Steering Group.

The composition, scope and protocols for a Search Steering Group will be very dependent on the way in which an organisation conducts business.

Help Desk Governance
There will need to be two help desks. One of these will be the IT Help Desk and the other will be the Search Help Desk. The IT Help Desk will be able to respond to queries about technical issues with the search applications and the Search Help Desk will respond to queries that are related to content and query management. The reality is that most search problems involve an element of both IT and content issues and so the relationship between the two help desks needs to be as clearly set out as possible.

IT Help Desks usually have a 'ticket' system which enables staff to track the progress and eventual resolution of an issue. It could be that a particular ticket needs to be passed on to the Search Help Desk and this may require some fine tuning of the IT Help Desk application. The volume of queries being managed by the Search Help Desk will be substantially smaller but so will the resources available to address the issue.

Information management
Reference has already been made to the importance of having a corporate information management policy and strategy. The term 'information management' covers many different aspects of the business. A survey in a recent report from the Information Governance Initiative indicated that the term Information Governance was seen by respondents as covering around twenty enterprise application areas. Intranet Focus Ltd has prepared a Research Note on information management.

IT Liaison
Search applications need the closest possible liaison between IT and the business and it is advantageous to have a senior IT manager with responsibility for the management of IT resources for search. It is most certainly not just a case of having access to the on-call technician even though their expertise may be important in issue resolution.

Another aspect of IT liaison is to be able to anticipate future requirements for search applications and also to be ready to respond to changes being made in the IT enterprise architectures which could have an impact on search performance.

Legal conformance
Data privacy legislation in the European Union is going to be substantially revised over the next few years, and the impact of these changes on cross-border access to information, and especially to personal information, should be addressed in the strategic plan. In the case of public sector organisations in many countries of the world search applications could be very important in meeting Freedom of Information requests.

Another important consideration is access to copyrighted information, and in the defence industry there could be implications for export control.

Licences
Licenses for search software come in many different formats. Some are server based, some user based and some are document volume based, with many hybrids and varying levels of support associated with them. The search strategy needs to set out the main elements of all current licences and consider what actions may be required in the longer term to ensure that the licence conditions align with each other and with business requirements. Cloud licences in particular need to be carefully reviewed.

Metadata
The management of metadata is of fundamental importance to effective search. Ideally standards and guidelines on metadata management should be set out in an information management strategy. Many organisations have either enterprise or departmental taxonomies and controlled term lists, for example of product designations. The search strategy should document these taxonomies and term lists, identifying the owners and the processes for updating them.

Mobile
The importance of search on mobile devices is going to increase substantially in the years ahead. A search strategy for mobile needs to cover not only how search is going to be implemented on Android, iOS and Windows smartphones and tablets but how users will be able to undertake cross device searches in which they start a search in a laptop or desktop PC and then refine it on-site using their smartphone or tablet.

Open Source
Even if the organisation is not currently using open source search applications the chances are that at some point in the future it will be, even if only for specialised applications. The search strategy should at least consider the implications of implementing open source applications, primarily in terms of the extent to which the organisation will be able, or needs to be able, to develop open source search applications internally.

People Search
Search strategies tend to focus on finding information, whereas an important user requirement is to be able to track down people by name, role or experience. One of the immediate issues to be considered is whether there should be a separate query box for people search. Almost certainly people search will involve some degree of federated search across a number of databases and there will be a number of issues where the HR department may have views on the way in which people search is implemented. Expertise search is a separate but related topic under this heading.

Performance
In this section the performance criteria covering the time taken to respond to query with a list of search results and the time taken to call up documents or applications associated with any of the results should be set out. Search users expect a sub-one second response to a query and it is important to monitor whether any delays in meeting an agreed response level are back-end or front-end issues.

One common reason for latency in search is because of the security management protocols. Inevitable late-binding security, where each result is matched against the security permissions of the user, will result in a greater degree of latency. Performance will also be dependent on the search system architecture and the network bandwidth availability. The bandwidth availability could well change during the day.

Project Management
The search strategy should never be a project plan. Projects have a finite duration whereas search needs constant attention. Treating it as a project often means that at the end of the project, often the implementation of a new search application, the project team is dispersed,

along with the knowledge of the implementation, leaving in effect a 'black box' to be maintained by IT.

Query Management
The adoption of type-ahead query management on Google and Bing has highlighted the benefit at an enterprise level of managing the query and not just managing the set of results generated by what is often a poor query. Different categories of content may benefit from different types of query management, and that requires an understanding of how the indexes are being managed. In some cases the type-ahead feature is not security trimmed so even the query may alert users to content that they are not able to access.

Return on Investment
This is one topic that should not be included. There is no way that the investment can be measured. It is not just technology licences or search team head count but also the time that has been spent creating the content that is now held in repositories that cannot be searched effectively. The returns include better decision making, faster on-boarding and fewer interruptions by staff hunting down information through their networks.

Risk
The risk strategy should have two components. The first of this is operational risks for the search application. These will not only be technical risks (potentially leading to a disaster recovery requirement) but the availability of staff resources to manage the search applications.

The second component is the risks to the business of a search application which does not meet both the requirements of users and the business itself. These requirements should link back to the business cases used in the search strategy.

Risks are traditionally assessed using a risk score that is a product of the potential impact of the risk occurring and the potential probability of the risk occurring. It can be very difficult to score probability and therefore the emphasis should always be towards impact rather than risk.

Roadmap Management
Commercial vendors are often reluctant to divulge their product roadmap to avoid competitive threats. In the case of open source applications there is a very rapid deployment of new versions of core Lucene, Solr and Elastic Search software but often these are bug fixes and performance upgrades. A view needs to be taken on what is known, or has been disclosed, regarding the roadmap of current applications and the opportunities and challenges involved in upgrading to a new version.

Scope
Of course the scope of the strategy should be stated at the outset. In most organisations there will be multiple search applications, some of them stand-alone and others embedded in other enterprise applications or as search-based applications. It is important to define which come within the scope of the strategy and therefore of the Search Steering Group.

Search Support Team
To achieve high levels of search performance requires

- Ensuring that search supports business objectives
- Regular assessment of analytics
- A good understanding of why and how people search
- Dependable IT performance and support
- Training and help desk support
- Anticipating future requirements

Even in smaller organisations if information is a business-critical asset there should be at least a full-time search manager. However even a full-time manager will have difficulty in fulfilling all of the tasks in this list, each of which has related sub-tasks.

The only way to determine search team roles, responsibilities and staffing levels is to develop a list of tasks and then match these against the potential risks if they are not undertaken. Using this approach also helps define the roles of individual team members, whether they need to be full-time or part-time positions and the skills and training that each member of the team requires.

Security
In most organisations there will be documents and other information that can only be accessed by specific people or people in specific roles. All too often security is based around managed email circulation lists with no indication on the document as to the circulation of the document. Document level security should ideally be compliant with ISO 27001, which was revised in 2013.

The key issues are

- Confidentiality: Protecting information from unauthorized parties.
- Integrity: Protecting information from modification by unauthorized users.
- Availability: Making the information available to authorized users.

In the case of a search application access is managed through Access Control Lists which add permissions to access based on either an individual's Active Directory address or a group AD list. Implementing a document security (more correctly data and information security) into ACLs can be a complex task. Equally complex is updating the ACLs to reflect changes in the security schemes. It is an area of search management where IT and business departments (including HR) need to work very closely together.

SharePoint
SharePoint is now so ubiquitous that there will probably be some level of use in almost any organisation, even those nominally using IBM or Oracle enterprise applications. The management requirements of search in SharePoint have changed substantially between SharePoint 2010 and SharePoint 2013. Business units with departmental SharePoint implementations may not appreciate the power and complexity of SharePoint 2013 search and may need convincing that at least the search element of SharePoint should be included in the corporate search strategy.

Stakeholders
The strategy should list the stakeholders with a direct interest in either the management of search applications or the need for high quality search to be delivered to teams and departments that they have responsibility for.

Training
In theory Google has set the standard for 'intuitive search' but the search engine has access to a very substantial amount of information about each user that can be used to enhance the query performance. Even with this contextual information there are many books on who to get the best out of Google for specific types of enquiry. The same approach should be taken with enterprise search applications. Many users will be able search very effectively with little or no training, but with support from Help pages and access to the Search Help Desk.

Other groups of users, perhaps using a search application for very specific purposes, may benefit from virtual or on-site training. The training programme should set out a range of training solutions and take account of the fact that perhaps 10% of the employees of an organisation will leave/be replaced in the course of a year.

Usability Tests
Search user interfaces, especially when customised for a specific user group, are often quite complex. It is important to be undertaking usability testing on a regular basis. This can be quite time consuming and the staff requirements need to be taken into account when considering the staffing levels for the Search Support Team.

User Requirements
Search has to be user driven. The clarity with which user requirements are stated in a search strategy will have a direct impact on the quality of the search performance which is delivered. In this A-Z list user requirements come towards the close of the document but in reality they are the starting place for the development of a search strategy. User requirements need to be set within broad business objective and information management contexts so that search is seen as one element of the information discovery process. This ensures that if there is a change in the architecture of an intranet, as an example, the search user requirements can be restated to take account of this change. Search should never be seen as a 'standalone' task.

User Interface
Unlike many enterprise applications the search user interface is usually able to be customised to meet the requirements of specific user groups. SharePoint 2013 is a good example and with an open source application there is very considerable scope to design user interfaces. As the search application continues to be developed the only visible sign of this development may well be changes to the user interface, hopefully taking into account feedback from users in the development process.

A major change in the user interface (such as the addition of facets and filters) needs to be clearly communicated and appropriate training put in place.

An aspect of user interface design that is often overlooked is to ensure that the interfaces are compliant with the Web Accessibility Initiative. The density of text (often a result of wishing to display 10 results on a page) can result in users who have difficult reading small print sizes being presented with a major readability challenge. Users with some degree of dyslexia may also encounter problems with search user interfaces.

Website Search
This is one search application that often gets overlooked in a strategy because it is often not the responsibility of the IT department. Marketing may well have outsourced site search to Google or another hosted service.

Checklist

Topic	Relevant (Y/N)	Owned by	Completed
Analytics			
Architecture			
Best Bets			
Big Data			
Business Cases			
Business Impact			
Cloud Search			
Communications			
Connectors			
Content Quality			
Content Scope			
Crawl Management			
Development Plan			
Disaster Recovery			
eDiscovery			
Emails			
Feedback			
Governance			
Help Desks			
Information Management			
IT Liaison			
Legal conformance			
Metadata			
Mobile			
Open Source			
People Search			
Performance			
Query Management			
Risk			
Roadmap Management			
Scope			
Search Support Team			
Security			
SharePoint			
Stakeholders			
Training			
Usability Tests			
User Requirements			
User Interface			
Website Search			