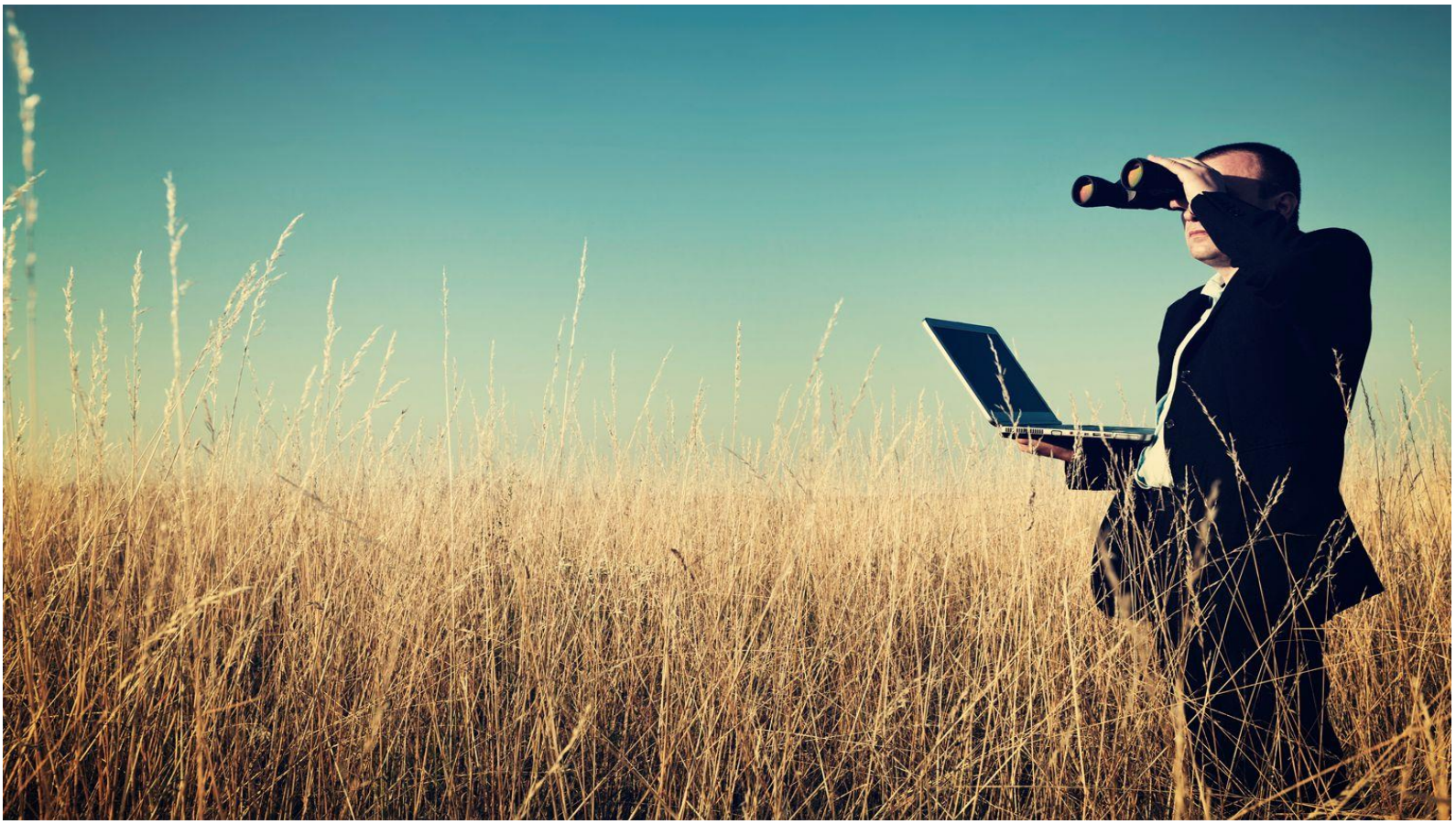


Non-Traditional
Enterprise Search
Implementation
The Approach

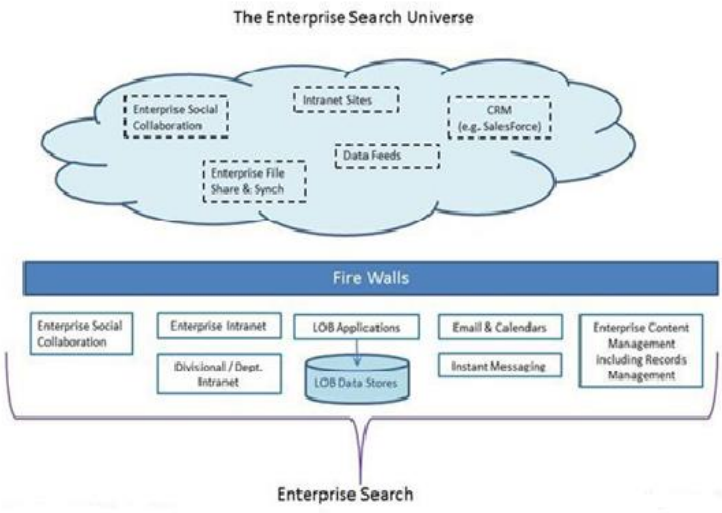
Contents

1 Introduction

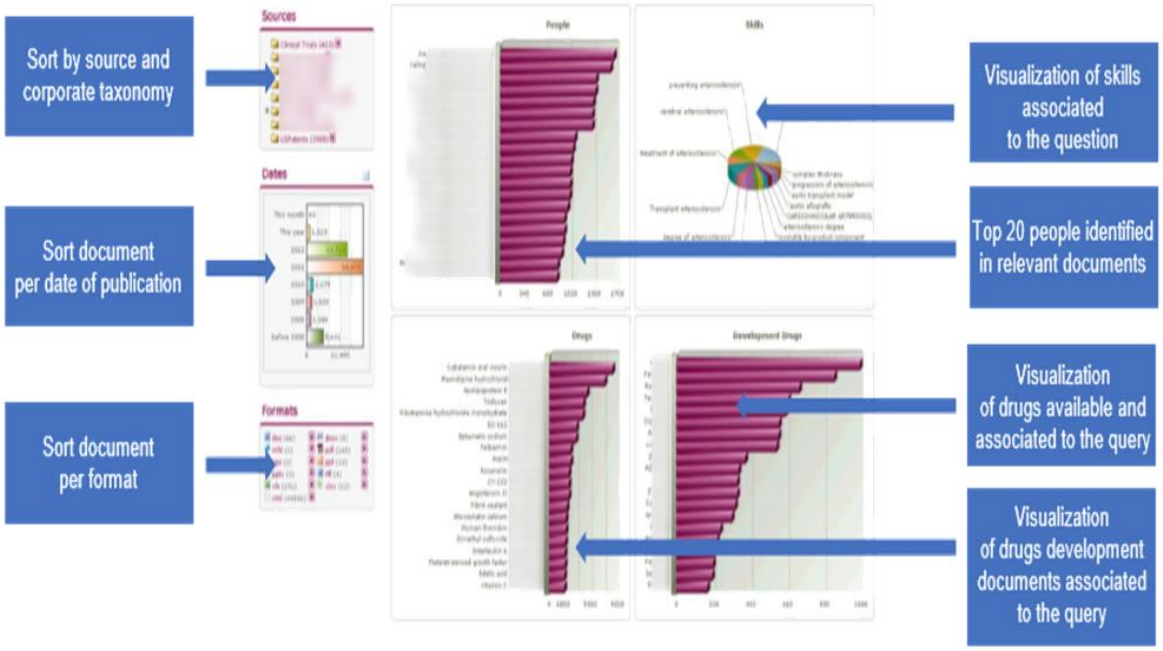
Search engines are everywhere. Every business application or website that contains a significant amount of content has one. From embedded search engines to sophisticated, enterprise-scale platforms, there are a number of established principles that can guide organizations to use search effectively.

This is what happens when an enterprise doesn't have right Information Discovery strategy rolled out.

Research says on an average 8-10 hours/week of an Employee goes into searching information within or outside the network; out of which 20% results are fruitful



Below is what we wish to have



Providing big data functionality without overhauling the data center is achievable with a two-pronged approach that combines augmented enterprise search and distributed computing. Search is very good at finding interesting things that are buried under mounds of information. Enterprise search can provide near-real-time access across a wide variety of content types that are managed in many otherwise siloes systems. It can also provide a flexible and intuitive interface for exploring that information.

2 Today's Enterprise Search

Nowadays, the market is flooded with solutions that try to efficiently tackle Enterprise Search (sometimes Enterprise CMS and Search at the same time), but more often than not companies end up facing costly integrations which only provide partially effective results.

Surveys (year 2015-16) give insight into the difference between companies with happy search users versus those whose employees prefer to avoid using internal search. One particularly interesting finding last year was that there are three levels of 'search maturity', identifiable by how search is implemented across content.

- The least mature search organizations, roughly 25% of respondents, have search for specific repositories (siloes), but they generally treat search as 'fire and forget', and once installed, there is no ongoing oversight.
- More mature search organizations that represent about 60% of respondents, have one search for all silos; but maintaining and improving search technology has very little staff attention.
- The remaining 15% of organizations answering the survey invest in search technology and staff, and continuously attempt to improve search and findability. These organizations often have multiple search instances tailored for specific users and repositories.

Ok, but what is an issue?? Well let's have a look –

1. Until few years back, Enterprise content meant, typical email, file-share, and content from standard ECM products. And some standard textual search engine with few Boolean operators, long list of facets, filters, suggestions (basically correcting spellings and grammar that too not all search engine does) etc. drives the information (search results). Now the widening scope of Enterprise content getting changed from "structured content" to semi-structured and un-structured; conventional and unconventional sources, Existing Enterprise Search engine turns out to be too expensive with sophisticated integration and sometimes incapable
2. Result relevancy is still a problem, degree of pain varies based on the investment and effort put in Search area. The result of it is waste of employees' effort search for results "Actually what they want" with different keyword combinations…trying their Luck ☺
3. Search has been mostly considered as an artefact searching tool for reactive actions and not proactive actions, which means decision made are reactive ones and not based on the prediction model. Hence, not being fully able to curb the problem, or see the trends and work accordingly
4. Not all content is meant for all users; also at the same time it is equally essential to get a 360-degree view of few areas by few users. Both type of Enterprise (search engine confined to its own specific application, or one-search engine configuration for all) suffers from poor information and analytics
5. With the culture of BYOD (Bring your own device) for work and companies not investing in RWD (Responsive Web Design) of search has a major hit on finding information, and act real time. Not every time one carries a laptop or iPad, but yes they do carry cell phones/smartphones

In short, an efficient Enterprise Search system should be a versatile search tool, capable of understanding complex search queries, compartmentalizing information to target multiple audiences and being able to prioritize through different types of sources.

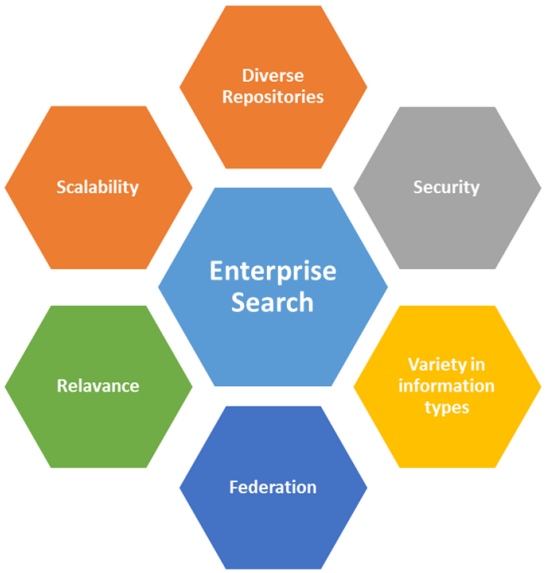
3 Things to be considered for Enterprise Search

According to Gartner Analyst Darin Stewart, enterprise search can bring "big data" within reach http://blogs.gartner.com/darin-stewart/2014/04/01/enterprise-search-can-bring-big-data-within-reach/.

So understanding the foundations for enterprise search success can help with big data initiatives too, especially if unstructured content is involved.

Things that should be considered and revisited periodically for effect Enterprise search are:

- **Diverse repositories:** The repositories for processing the information vary from a simple web server to a complex content management system. The enterprise search engine must be capable of dealing with diverse repositories.
- **Security:** Security in the enterprise has been one of the primary concerns along with fine-grained access control while dealing with enterprise search. Corporates expect data privacy from enterprise search solutions. This means two users running the same search on enterprise search may get two different sets of results based on the document-level access
- **Variety of information:** The information in any enterprise is diverse and has different dimensions, such as different types (including PDF, doc, proprietary formats, and so on) of document or different locale (such as English, French, and Hindi). An enterprise search would be required to index this information and provide a search on top of it. This is one of the challenging areas of enterprise searches



- **Scalability:** The information in any enterprise is always growing. Continual data growth is a challenge to search system accuracy and speed. And enterprise search has to support its growth without impacting on accuracy and speed. Best practices exist and should be applied consistently. In-short Scalability should an inherent nature
- **Relevance:** Public web searches can identify relevance from various mechanisms such as links across web pages, whereas enterprise search solutions differ completely in the relevance of entities. The relevance in case of enterprise search involves understanding of current business functions and their contributions in the relevance ranking calculations
- **Federation:** Any large organization would have a plethora of applications. Some of them carry technical limitations, such as proprietary formats and inability to share the data for indexing. Many times, enterprise applications such as content management systems provide inbuilt search capabilities on their own data. Enterprise search has to consume these services and it should provide a unified search mechanism for all applications in an enterprise. A federated search plays an important role while searching through various resources

Equally important is the **data quality injected**. The phrase "Garbage In is Garbage Out" still holds true. So expectation to have data to be classified at the time on indexing (but not right away at the source) is too demanding with no guarantee for 100% accuracy.

It is better to classify and tag the content at source level even before indexing.

The **user experience** governs how users interact with the system and find the content they are looking for. Without a clean and easy to understand user experience, users will not use or adopt the enterprise search platform. Additionally, without applications that provide value and increase usability to users, adoption will also suffer.

One has to also think on the factors on how to approach for **Post-Acquisition Enterprise Search.**

4 What we are heading to

The scope of enterprise search is expanding. For any enterprise search solution to succeed in today's enterprise, it must go well beyond the traditional boundaries of enterprise search. Since facets, auto-completion, and 'did-you-mean' features were introduced to search engines, only small changes have taken place over the past few years from a user interface point of view. And although future changes will take place 'under the radar,' they will all lead to improved search quality for a better user experience and improved satisfaction.

Search Engine	Artificial Intelligent Knowledge Engine
Current methods of evaluating search accuracy are usually based on golden query sets, key documents, zero result searches, or top results	A more sophisticated and automated approach is using log file information from the search engine (search queries) and web servers (clicks) to create a holistic view on search accuracy instead of focusing on limited amounts of information
Search Result Personalization- Right now the personalization is based on the user profile from within the enterprise or from external networking sites like linked-in, FB etc.	Search Result Personalization- Search result quality may also be improved by tailoring the search results to individual users or user groups. With log analytics, users may be clustered based on their search behaviour, documents downloaded, user data (profiles), or content contributions
Most of the legacy search engines rely on matching words, one to one, with the exception of stop-words and some linguistic base form reduction	With increasing computing power, it is now possible to effectively extract entities (statistical phrases or acronyms), compute semantic co-occurrence (semantically related terms occurring together), and perform semantic weighting (based on the document corpus, the user's history, or both)
We search for something and make decisions to perform actions based on the results we get back	About to happen with search- the search technology will be performing the next step. In the same way, as information systems have improved the quality of knowledge work, the next generation of information systems will start to do the low-level knowledge work, allowing us to focus on higher-level tasks

5 Few points around defining a checklist for Enterprise Search

A-Z checklist sequence are the topics that should be considered in the development of a search strategy. Not all may be relevant to a specific organization, but if they are to be excluded then the reason for doing so should be discussed and agreed. All the topics in the list are considered at length in Enterprise Search.

Points	Details
Accessibility	Sets out the extent to which the search applications meet the Web Accessibility Initiatives
Acquisition	The extent to which the search applications could be extended in an acquisition or merger situation
Architecture	Server and network architecture requirements and server availability

Points	Details
Best bets/suggested	The user requirements for best bets and how they will be reviewed and revised
Big Data	Integration of the search and Big Data strategies, especially around common metadata schemas
Budget	License costs, vendor maintenance and support, and staff costs
Business cases	Summary of the evidence from user requirements research to support and prioritize specific business cases, including the potential business impact
Cloud search	The potential benefits and challenges from implementing cloud (or hybrid) search
Communications	The communications strategy and forward communications program for stakeholders and users
Connectors	Requirements for connectors and associated support from suppliers
Content analytics	The extent to which the organization will benefit from implementing content analytics solutions and the relationship of these solutions to search
Content quality	Requirements for content quality and content curation to enhance search performance, ideally placed within an information life cycle framework
Content scope	A list of the content being crawled and indexed
Crawl management	Optimal crawl schedules to balance user requirements with any architecture/performance constraints
Dependencies	Business or technical dependencies that could impact search performance and search satisfaction
Development plan	The opportunities for enhancing the search environment over the following two years, based on user requirements research and business objectives matched against resources
Disaster recovery	Disaster recovery plans with Recovery Time Objective (RTO) and Recovery Point Objective (RPO) requirements
eDiscovery	If appropriate, the touch points between the eDiscovery strategy and the search strategy, especially regarding the sharing of skills
Expertise search	Linking the requirements for expertise search from a knowledge management strategy with the search strategy
External search	The requirements for search access to external information resources on e.g. Research, competitors, and market opportunities
Federated search	Current and potential opportunities and challenges for implementing federated search
Feedback	How users will be able to feedback comments and suggestions to the search team
Governance	The ownership of the search budget and search strategy, together with roles, responsibilities, and lines of reporting for members of the search team

Points	Details
Help desk management	The relationship between the IT Help Desk team and ticket system and the search help desk
Information management	A summary of the organization's information management strategy with particular reference to the requirements and objectives for the search strategy
IT liaison	Service-level agreements with IT departments for support and development, including the requirement for staff with specific skills to be available
Key performance indicators	Definition of a set of periodic key performance indicators that relate to the business cases and business impact requirement
Language	Setting out any requirements for indexing and searching in languages other than the nominal corporate language
Legal conformance	Requirements to conform to data privacy, Freedom of Information, and export license controls
Licenses	List of licenses by vendor and license renewal date so that the implications of a merger or acquisition of the vendor can be quickly assessed
Metadata	A summary of metadata schema, controlled term lists, taxonomies, thesauri, and relevant master data schema
Metrics	A summary of the suite of performance, discovery, satisfaction, and impact metrics, together with required benchmark levels
Migration	The implications for search as an element in a content migration strategy
Mobile	How the search applications will be implemented on mobile devices, together with an assessment of the need for cross-device support
Open source	Sets out the organization's approach to using open source applications
People search	The requirements for people search
Performance	The technical (network/server) performance benchmarks for crawl, index, query, and result display
Risk register	A risk analysis relating both to operational and strategic risks for the search application, and the consequential risks to the organization
Roadmap	Release dates for upgrades to search applications, the basis on which they would be implemented, and development roadmaps for other enterprise applications
Scope	Confirmation of the repositories to be crawled and indexed in order to meet user and business requirements, the search applications to be included in the strategy, and the search applications that are being excluded
Search support team	Operational responsibilities and reporting lines for the search support team, including requirements for training
Security	Summary of security requirements covering confidentiality, integrity, and availability in line with ISO standards

Points	Details
Stakeholders	Confirmation of the stakeholders and other members of the search community, using the RACI model
User training	Provision of training courses for search users, especially new joiners and staff in search-intensive roles
Usability tests	The scope and schedule for on-site and remote usability testing
User requirements	Summarizes the core user requirements
User interface	Sets out the rationale for the user interface to meet user requirements
Website search	Sets out the management and operational links between internal and external search