



THE ENTERPRISE SEARCH PRIMER



WHITE PAPER

Chirag Gandhi

July 2008

Enterprise Search is gaining ground within the organization; thanks to the popularity and productivity gains observed while using web search engines like Google. This paper introduces the concept of Enterprise Search and some interface patterns.

1. Introduction

Ever since the first data-centric applications were created, users have been searching data for information; be it a customer contact or a part number. In the early days, when data was stored only in files, search capability was primitive but critical nonetheless. Relational databases provided a quantum leap in the ability of users to structure and manipulate data. Relational databases increased overall system usability by making it easier and faster to find specific data.

With the advent of word processing systems, information that had always been stored and accessed on paper became digital documents. At the same time, as e-mail flourished, communication became digital artifacts of business processes. When the number of documents were small, users could remember where they were located and fetch them as required. However, over time, the number of electronic documents grew steadily. As work became more automated and less manual, companies reduced staff and centralized business functions. This led to an exponential increase in non-structured content handled per user.

In the late 90's, as companies were putting up generation one websites, they faced a similar problem on the internet. There was a rapid proliferation of content - most of it HTML pages. Lycos, Altavista, and Yahoo provided gateways to the internet by creating directories which classified web sites. These were the first portals.

In 1998 a small startup named Google introduced a new concept on the internet - a site dedicated to search that was just two pages.

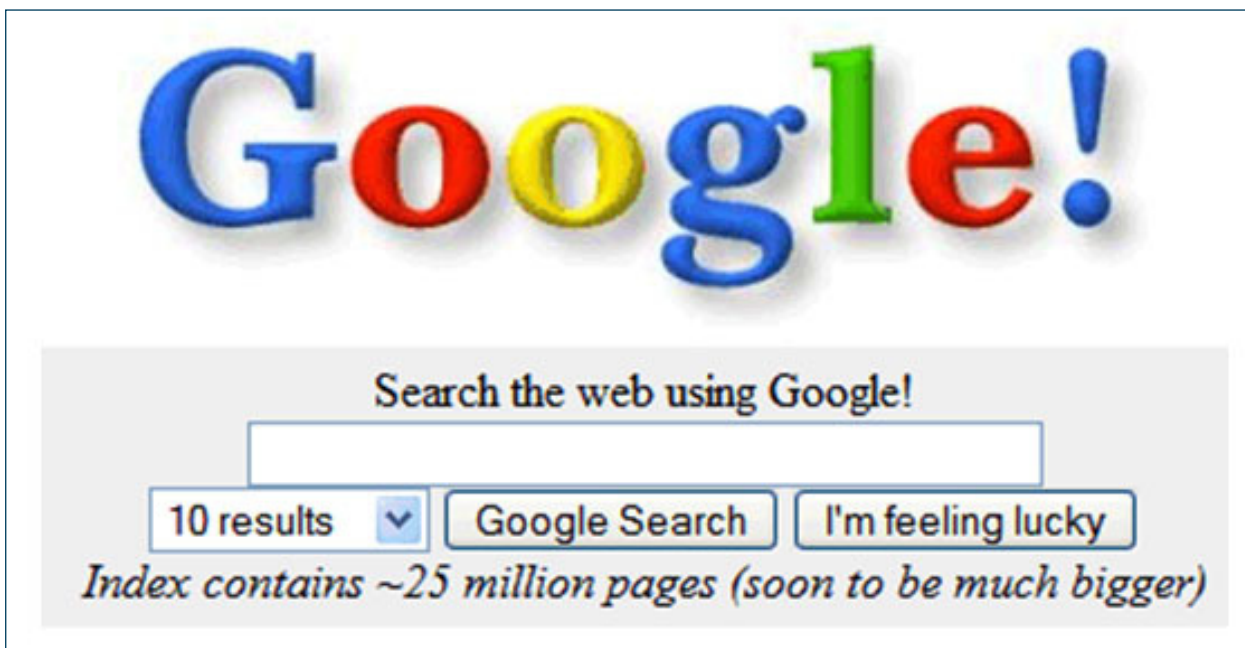


Figure 1. Google Home Page 1998

Google's home page was a search box; the second page was a search result page. Google's page rank algorithm made millions of pages on the web easily searchable. The most relevant results could be generally found on the first result page. Usage of Google skyrocketed and it quickly became the destination of choice when anything needed finding on the internet. Search became the gateway to the internet.

As the size of content repositories grew within organizations, they started looking for ways to utilize this information more effectively. A study by IDC in 2001¹ estimated that for a 1000 person company, the implementation of an effective enterprise search solution would provide direct cost savings in excess of 15 million USD, while the intangible benefits of being more competitive in the market could not be measured. The recent increased awareness among companies on the importance of search can be attributed to personal experiences with Google. Enterprise Search is steadily moving from a nice-to-have technology to a mission-critical one.

¹ Susan Feldman, Chris Sherman. (2001). The High Cost of Not Finding Information. Framingham, MA: IDC

2. Enterprise Search is NOT Web Search

Early adopters of the technology soon learned that enterprise search is not the same as web search - what worked on the internet does not work within the company. So what is different between these two seemingly similar needs?

1. Security - Company information assets are strictly controlled by access control rules. Search results must adhere to these rules.
2. Diversity of content - While most of the content on the Internet is HTML, the company has many different content formats - HTML, text documents, spreadsheets, databases, e-mails, etc. A search engine needs to handle all these different formats effectively.
3. Integration with LOB applications - While search on the internet is largely stand-alone, within the organization, search needs to be integrated into LOB applications.
4. Nature of content - The internet has millions of linked documents. Google created the page rank algorithm that used the nature of page content to refine its results. In the organization, documents are diverse and generally are not linked to each other. Because of this, Google's page rank algorithm does not work well in the enterprise.
5. Custom vocabulary and taxonomy - Each company has its customized vocabulary and taxonomy. The ability to incorporate such custom linguistic needs into processing and querying logic is essential.
6. Precision vs. Recall - Text-based searching necessarily fails to yield accurate and comprehensive results due to the inherent elasticity of language. For machines to be able to achieve comparable precision and recall, they need semantics - controlled vocabularies and rich syndetic networks.
7. Information life cycle - Enterprise information follows controlled lifecycles from content creation to archival, ultimately to being deleted. Additionally, live content in a company can undergo versioning. These various states must be effectively handled by an enterprise search solution, which is not the case for search.
8. Compliance - Recent legislation mandates that companies maintain records on information access and usage. Enterprise search engines need to handle these requirements at a user level of granularity.

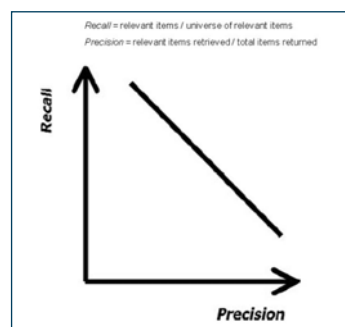


Figure 2. Precision v/s Recall

3. Enterprise Search Interface Patterns

Companies want the search to be simple - "like Google". This user expectation means that at the query entry end it is really just the search box - it has to be as simple as that. There are three main patterns for the search interface. Each pattern has its pros and cons. Companies need to evaluate their circumstances and key business drivers and select the appropriate pattern for user-facing applications.

The Locator Pattern

The locator pattern is used when users simply wish to locate relevant information. They are aware of the terms they need to perform the search. This pattern is appropriate for sites which have a very limited quantity of data.

Google Enterprise
Google enterprise products make your users more productive by combining the innovation ... Check out Google's latest Enterprise Search ideas on our Google ...
www.google.com/enterprise/ - 10k - [Cached](#) - [Similar pages](#) - [Note this](#)

Google Enterprise: Google Search Appliance
Google Search Appliance 5.0 delivers secure search across more content sources, with enhanced search quality and new user interface enhancements. ...
www.google.com/enterprise/gsa/ - 13k - [Cached](#) - [Similar pages](#) - [Note this](#)
[More results from www.google.com »](#)

Enterprise Search Summit 2008
ENTERPRISE SEARCH SUMMIT is a highly intense, in-depth, 2-day conference that covers how to develop, implement and enhance cutting-edge internal search ...
www.enterprisearchsummit.com/ - 36k - [Cached](#) - [Similar pages](#) - [Note this](#)

Enterprise Search from Microsoft
Enterprise Search from Microsoft. Empower people to find information. FREE Search Server 2008 Express download.
www.microsoft.com/enterprisearch/ - 41k - [Cached](#) - [Similar pages](#) - [Note this](#)

FAST - Enterprise Search
The Norwegian company that developed the AllTheWeb search engine. Corporate background, products, technology, news and events.
www.fastsearch.com/ - 25k - [Cached](#) - [Similar pages](#) - [Note this](#)

Enterprise search - Wikipedia, the free encyclopedia
Enterprise search is the practice of identifying and enabling specific content across the enterprise to be indexed, searched, and displayed to authorized ...
en.wikipedia.org/wiki/Enterprise_search - 27k - [Cached](#) - [Similar pages](#) - [Note this](#)

Figure 3. Search Result for the Locator pattern

This interface is very similar to the Google search results page, and from a search behavior perspective it provides value similar to Google search on the Google's simplicity and at times its frustration. The interface is minimal and allows the user to engage in very basic search behavior. Users follow this behavior every time they perform a

search: first they predict and use keywords that are most likely to return relevant results. Then, they revise these search terms based on the earlier result set. This cycle is repeated until relevant results are returned. Search engine algorithms provide targeted suggestions by ranking results.



Figure 4. User search behavior

Implementations of this kind generally do not require special considerations and hence can be deployed quickly with minimal cost.

The Finder Pattern

The finder pattern is used when users need to find or discover their search terms. This pattern is required when there are large quantities of content and/or the content is fairly complex, preventing the user from guessing the appropriate search terms. The advantage of this pattern is that it greatly enhances the efficiency of the natural search behavior of the users.

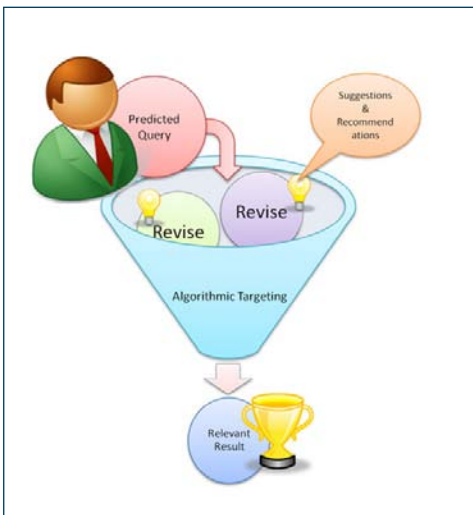


Figure 5. User search behavior enhanced for discovery

Search results are enhanced by navigational aids that are computed by analyzing the metadata of the content in the search result. Such suggestions are called **facets**, **navigators**, and **clustering**. Additionally, site administrators or search managers can provide recommendations for specific search terms. Such aids improve the usability of the search solution and are extremely relevant in an enterprise context.

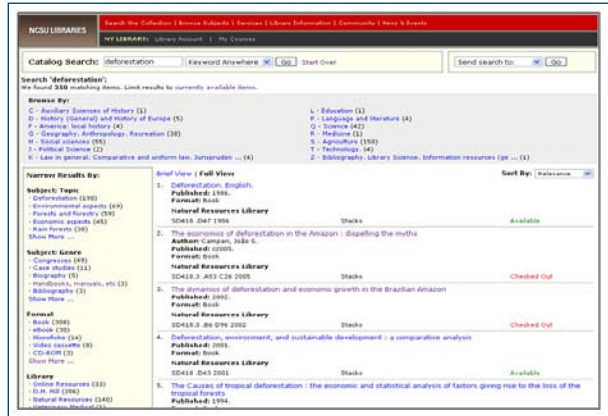


Figure 6. Search Result for the Finder pattern

A recent enhancement to the finder pattern is the ability for some enterprise search engines to act as information integration hubs. These provide automated suggestions that lead to *serendipitous discovery*. A good use case within the enterprise for this would be “When the users searches with a product name, the main search result should display all the documents related to the product. However, if the user belongs to the sales team, it should also display (as a sidebar) a list of clients who have bought this product”.

The implementation of this pattern requires careful planning and analysis. This pattern is very effective when used in conjunction with Content Management and Metadata Management solutions, and can provide high user satisfaction scores when implemented correctly. To reduce the cost of these implementations, some search engines are configured to handle a few pre-defined metadata fields and can perform automated *entity extractions* to populate these fields.

The Social Pattern

With the adoption of web 2.0 principles into the enterprise - called Enterprise 2.0, a new pattern for search interface has emerged, the social pattern. This pattern utilizes the concepts of *collective intelligence* and *social networks* effectively to improve search interface usability.

Collective intelligence is harnessed by creating reverse-intelligence filters for usage patterns of the content. These filters can provide a list of relevant documents based on what other users have downloaded for the same search terms.

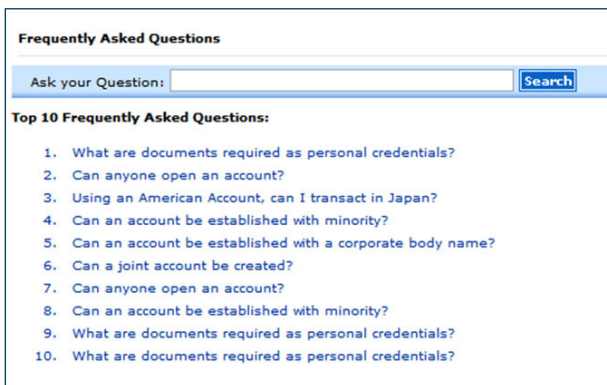


Figure 7. TOP 10 utilizing the Social pattern



Figure 8. Enhancing the search experience by using the social pattern

Collective intelligence can be explicitly elicited by the use of social tagging and Folksonomy solutions.

Social networking data also provides meaningful inputs to control relevance algorithms, providing a more context-sensitive set of results.



Figure 9. User search behavior enhanced by Enterprise 2.0

The social pattern is very recent and is currently supported by a very limited number of Enterprise Search vendors. It requires a higher degree of investment and careful planning. However, this pattern provides maximum user satisfaction and will prove to be extremely effective in organizations having highly collaborative work environments. This pattern also provides extremely high user satisfaction in the B2C context where the organization has large amounts of customer facing content in scenarios such as e-commerce and e-service.

4. New Trends in Enterprise Search

As the acceptance of Enterprise Search as a mission-critical solution is growing, companies are finding new ways to deploy this versatile tool.

Enterprise Search as a Service

Companies with multiple portals will need to deploy search functionality for each of these stakeholder touchpoints. It would be prudent to consider deploying a search solution as a shared service to be used by multiple sites. Most Enterprise Search engines today provide Webservice interfaces for search functionality, and so are capable of deployment as a service. In such scenarios, careful consideration must be given to the scalability and security features provided by the product selected.

Enterprise Search as an Information Integration Layer

A new trend, which has found some early adopters, is the use of Enterprise Search as an intelligent, agile, and non-invasive information integration and access layer, aka "Search as a Platform". There are certain merits to this approach:

- Enterprise Search can handle dirty and/or incomplete data because it was originally designed for non-structured content.
- It is highly optimized for a "write-once, read many" scenario, providing data retrieval with sub second response times.
- Linguistic capabilities are highly evolved.
- Schemas are fairly flexible and hence can accommodate most content types.

Utilizing these strengths of Enterprise Search, some vendors have built Business Intelligence front-ends to provide a "BI Lite" application. Some other applications of Enterprise Search as an Information Integration Layer are shown in the diagram below.

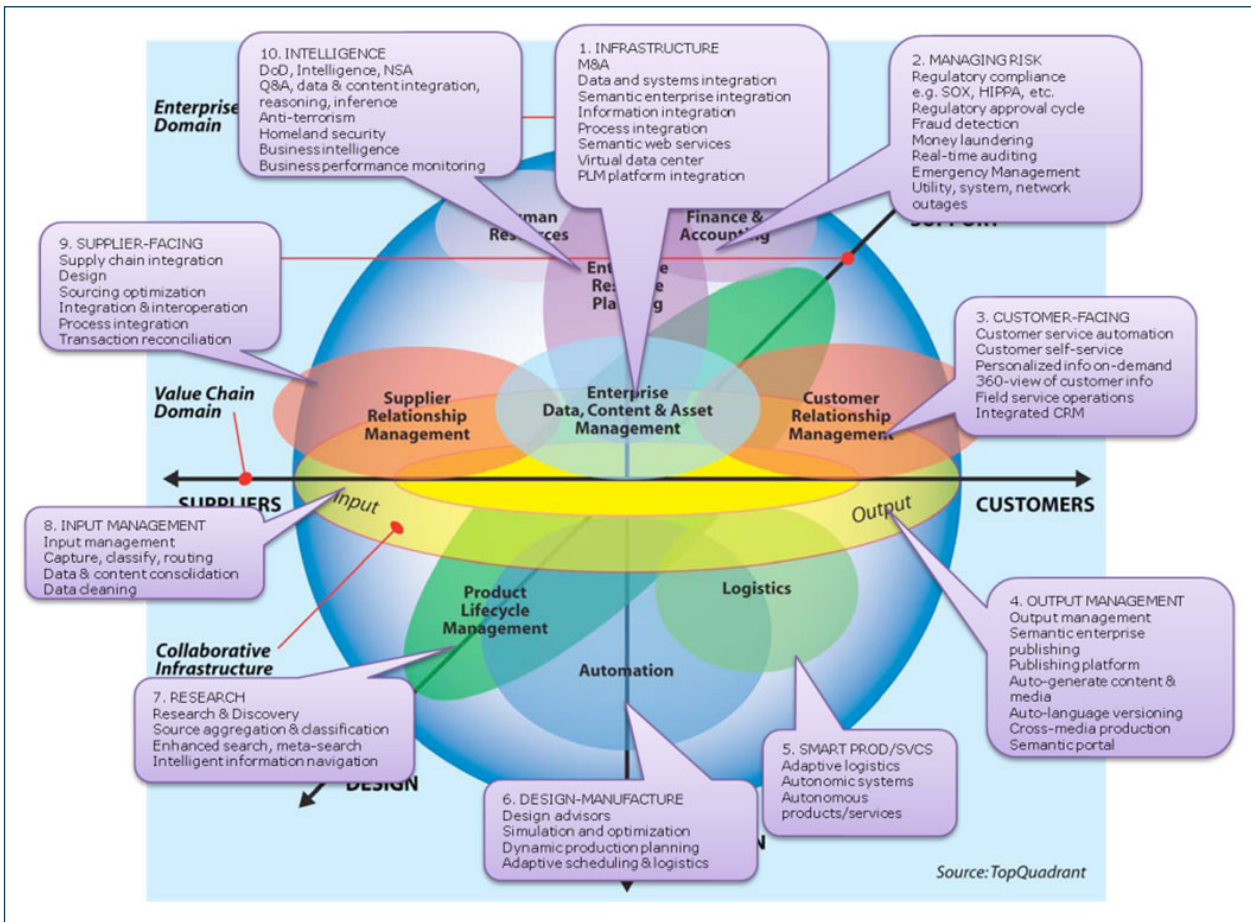


Figure 10. Application universe for Enterprise Search as an information access layer

Federated Search

Many companies need to allow their users to search across multiple systems, both internal and external. Federated search is a use of the Enterprise Search engine to pass on the user query to downstream search engines and provide an aggregated search result.

5. Search Center of Excellence

One of the most common mistakes made by an organization implementing a search solution is to treat it as an “appliance” – once deployed it is forgotten. Unfortunately, unlike database applications, search is a living organism and needs to react to changes in the business environment. New content and types of content are constantly added to the search engine, and except in its simplest implementation – that of a keyword search, the algorithms that drive the engine need to be tuned periodically to produce relevant results. Subtle changes in the business environment also drive the need to tune to relevance models.

If such constant monitoring and maintenance is not performed, search implementation loses its efficacy and leads to user dissatisfaction. In order to maintain an

implementation, it is prudent to invest in the creation of a Search Center of Excellence. The Search CoE should address critical areas such as a user’s search experience, collection and analysis of metrics, tuning, and system response. If the content can be indexed by Internet search engines such as Google, an additional responsibility for the CoE would be Search Engine Optimization (SEO) – the optimization of site design and interaction with vendors to promote organizational content on the internet.

6. Conclusion

Enterprise Search is evolving from a “nice to have” technology to a mission-critical business enabler. There are many capable vendors providing products to meet the needs of the enterprise. Selecting the right product and then deploying it in the most appropriate manner to maximize return on investment is a challenge as there is no “one size fits all” solution. It is also important for companies to recognize that Enterprise Search needs to be maintained as a strategic tool. Establishing a Search Center of Excellence within the organization will provide the right environment to harvest the true potential of this technology.

7. Further Reading

Battelle, J. (2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Portfolio Hardcover.

Ledford, J. L. (2007). *SEO: Search Engine Optimization Bible*. Wiley.

Morville, P. (2005). *Ambient Findability: What We Find Changes Who We Become*. O'Reilly Media, Inc.

8. Acknowledgements

I would like to thank Paula Thornton, Bert Hooyman and Dennis Pierson for the review and excellent suggestions on this topic - it is always a pleasure engaging her in a dialog around this topic. I would also like to thank Philippe Pineau for giving me the push which made me actually get down to penning my thoughts.

Contact us

USA

460 Park Avenue South
Suite #1101, New York
NY 10016, USA
Tel.: +1 212 686 6655
Fax: +1 212 686 2422

UK

88 Wood Street
London EC2V 7RS,, UK
Tel.: +44 20 85281000
Fax: +44 20 85281001

Australia

410 Concord Road
Rhodes, NSW 2138, AUS
Tel.: +61 290 221 146,
Fax: +61 290 221 134

INDIA

Bagmane Technology Park
Byrasandra Village,
C.V. Raman Nagar
Bangalore 560 093, India
Tel.: +91 80 4004 0404,
Fax: +91 80 4004 9999

About Mphasis

Mphasis is a leading Applications, Infrastructure Technology, and BPO services provider.

The company delivers real improvements in business performance for clients through a combination of technology know-how, domain and process expertise. With currently over 36,000 people, Mphasis services clients in Financial Services, Healthcare, Communications, Media & Entertainment, Transportation & Logistics, Energy & Utilities, Consumer & Retail, and Governments around the world. To know more, visit www.mphasis.com.

Mphasis and the Mphasis logo are registered trademarks of Mphasis Corporation. All other brand or product names are trademarks or registered marks of their respective owners. Mphasis is an equal opportunity employer and values the diversity of its people. Copyright © Mphasis Corporation. All rights reserved.

