

## SEEK, AND YE SHALL FIND: **ENTERPRISE SEARCH AND WEBSITE SEARCH**



Be it on the internet or just within companies: The amount of digital information keeps growing constantly and rapidly. And for both a company's website and within a company it is true that the more information there is, the more important becomes a proper search solution. This document shall provide a closer look and a discussion basis on Enterprise Search and Website search.

**Enterprise search** has become a vastly common term, even a buzzword, and most of the time it is used when it comes to drawing on diverse information from different sources. This angle puts Enterprise Search and its specialists on the field of Big Data: A huge amount of information growing in different directions by different rules or speed can hardly be tackled with a taxonomy-based approach. The information access that search appliances provide does hit that nail perfectly - because it is flat and very performant for almost every kind of information.

A bigger brand website will give a good example for a system drawing on different sources such as

- Press texts and other content, coming from a CMS or static html pages
- Social media application and user generated content such as communities, ratings, user reviews
- Product catalogues or web shops (assortments, prices, marketing data, availability checks)
- Third-party applications such as dealer locator or service descriptions
- Download files
- Websites that e.g. belong to the same company but are technically independent

This information variety has always been a good basis for search engine optimization and to bring organic traffic on the page, and that's how most companies already use it. But it also offers very interesting options for search solutions within a company. In order to understand these it is well worth the time to take a closer look on the mechanism of a search application:

On the internet (with regards to classical, crawler-based search engines), a document is treated like a website.

- Search terms will be looked up in the URL, in meta fields or page content.
- Scoring and relevancy will be calculated based upon hit location, hit distances and hit matching. In addition, the hidden algorithms that search engine providers keep to themselves come into play – generally trying to esteem content quality with the eyes of human users.
- SEO measurements aim to making the content particularly valuable for certain search terms - and thus boost the score and relevancy of a page, depending on the search algorithm.

Usually, the ranking will also consider the frequency of occurrence for a search term and put it in proportion to the absolute text length. That is, in short, the reason for optimizing the keyword density of a text.

The search solutions that companies might want to provide are usually subject to more complex requirements than crawler-based ones. Why? For a start, when performing a search on a company website, the three major criteria (URL, META-fields, content) are just not sufficient for a proper ranking:

- Content from a certain source might be more important than content from other sources. Should a hit from the shop part of the website not be more important than from the press archive?
- A hit matching with the product name should be probably more important than other ones (e.g. from scope of supply)
- The length of press or CMS texts (and the density of certain keywords) depends rather on the editor than on the actual scoring that would be intended.

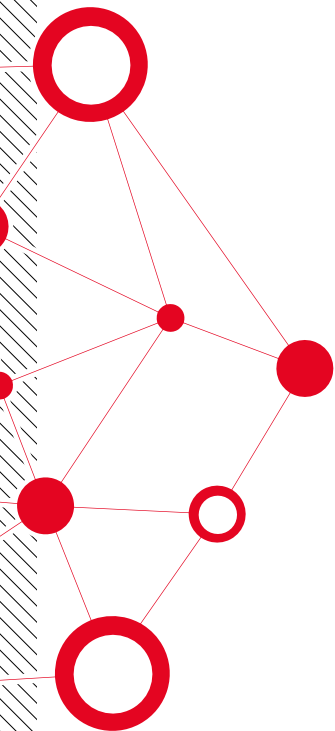
Taking even only those aspects in consideration, a configurable solution that accounts for those aspects is of the essence in order to be efficient.

As the data mostly come from systems that hold much more than the content that is being published, the potential is huge. One example for this would be to integrate hidden ranking criteria into a shop search so it will display the products with the lowest return rates (or the best ROI) more relevant than others.

Also, not all data-delivering systems have the same update cycles, particularly when it comes to multilingual systems: Social Media systems should feed the index constantly while e.g. Online catalogues would be updated only once a week and the CMS would be published daily.

For the search frontend, however, expectations are already set by our daily user experience and the big ecommerce portals and search engines: Quick and easy access, fast indexing, facet search options (just to name a few).

A proper tracking (Webtracking or custom tracking within the search) will deliver click paths and hits/nohits statistics so the search function can be improved constantly – because content, SEO measurements and product portfolio will keep changing over the time as well. You see: A good and convenient search is not a one-time technology decision but a system that needs care and attention.



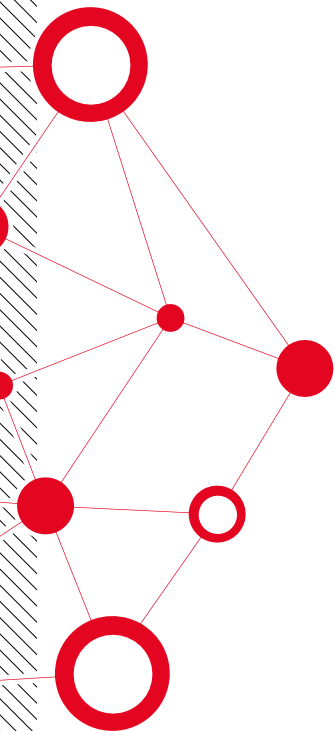
It helps to keep in mind that the search box on a website lets the user communicate directly with the company. And often enough it is the only option for the user to sidestep the existing navigation paths.

To illustrate this thought, just try to remember when you last visited amazon or another big ecom-merce portal without using the search function. A search function is the simplest way to explore a universe of different taxonomies and unknown content structures.

When it comes to realization, one will quickly come across the two big players SOLR and elastic search, both of which base on the Lucene Index. At this point, we do not want to argue the pros and cons of both frameworks, because they depend on technical requirements and also because they are, at least up to a certain point, subject to taste, operation and system landscape.

Both frameworks, however, can be set up and scaled in a way that is state of the art and robust:

- High performance and speed, both in indexing (near real-time) and response
- Good handling of huge amounts of data
- Stable and simple operation options
- Flexible setup based on feeds and web services
- Scoring manipulation and boosting
- Additional functions such as fuzzy search, facets, synonyms, truncated search, stop words etc.



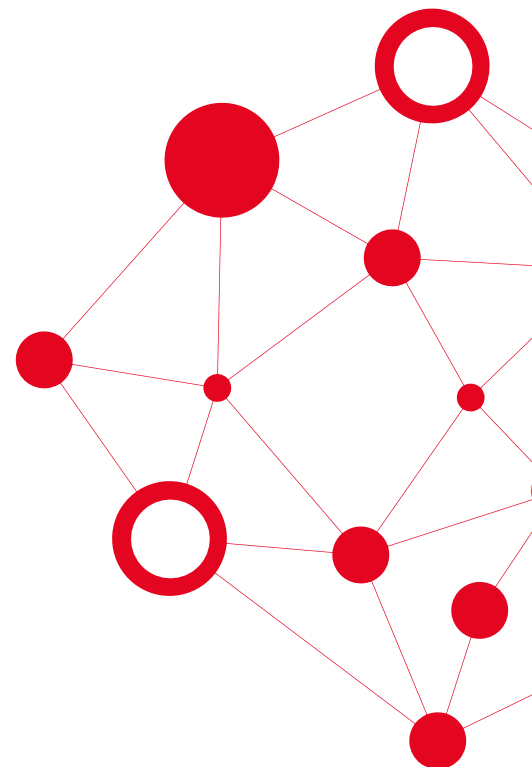
The simple and robust setup of search solutions is also an attractive option for the realization of huge product catalogues. This is why Kittelberger media solutions and their advastamedia® suite offer a native SOLR XML Export that can be directly imported into a solr server. The whole data persistency of a catalogue, including all media objects, can thus be done in a simple, performant and scalable way.

**Summary:** Realizing a professional search application on company websites or within companies offers huge potentials for easy and intelligent access to diverse information.

Current frameworks offer the options to cover most of the requirements a modern Enterprise search application would have.

Managing and knowing the to-be-searched content is, however, just as important – since any search bases on content. Website search in particular can thus be used to improve UX and to boost conversion.

Copyright belongs to Kittelberger media solutions GmbH. If you would like to use parts of this article, please inform us.



Bayernstraße 8  
D-72768 Reutlingen  
Tel: +49 7121 6289-0  
[info@kittelberger.de](mailto:info@kittelberger.de)