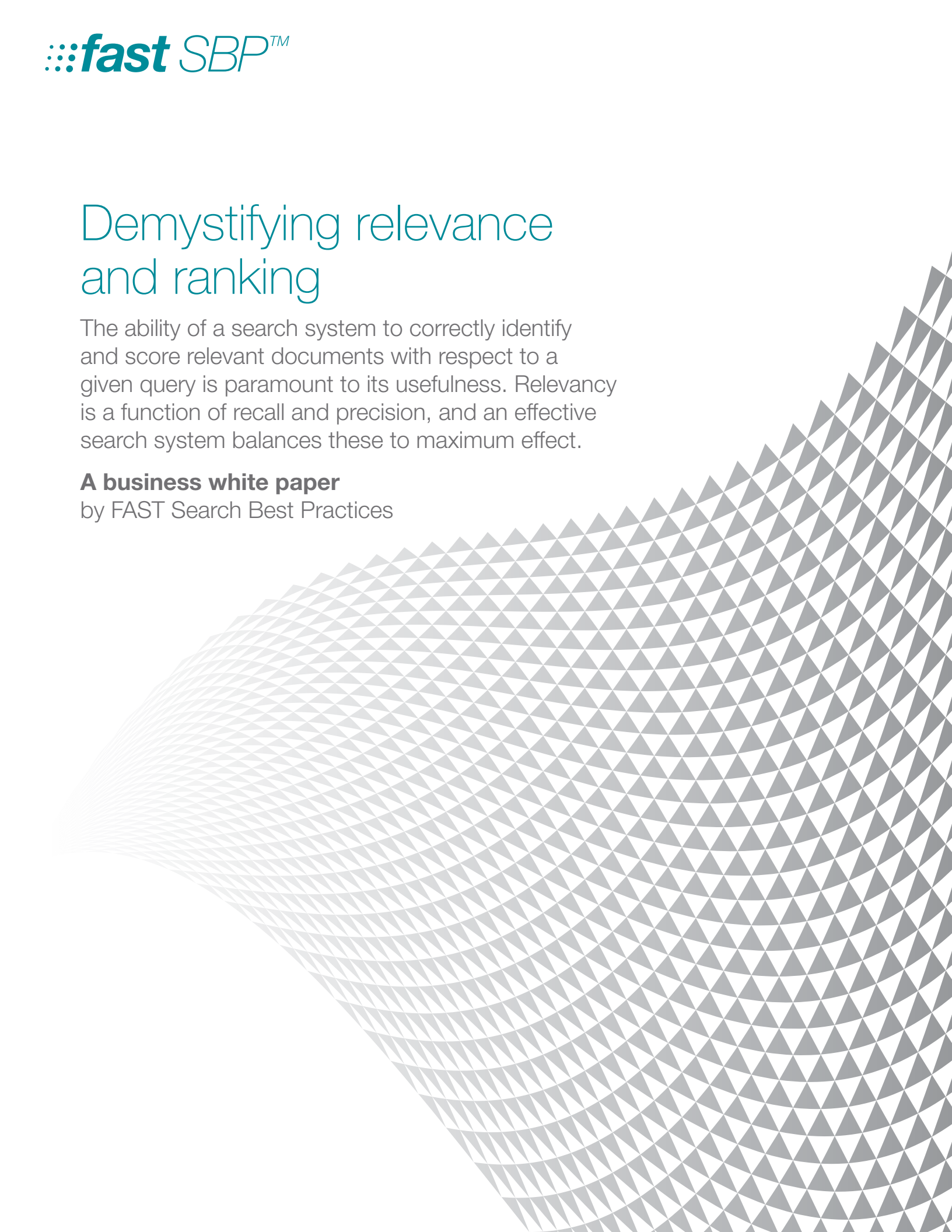
fast SBP™

Demystifying relevance and ranking

The ability of a search system to correctly identify and score relevant documents with respect to a given query is paramount to its usefulness. Relevancy is a function of recall and precision, and an effective search system balances these to maximum effect.

A business white paper

by FAST Search Best Practices



Information Overload

We are drowning in a sea of data. The world produces between 1 and 2 exabytes[1] of unique information each year, which is roughly 250 megabytes for every man, woman, and child on earth. Printed documents of all kinds comprise only .003% of this total[2], with magnetic storage being the most popular medium for storing all manner of information.

To create a vast majority of this information, we use modern, digital tools. These tools make it simple and cheap to produce information quickly and effectively. The ability to create information so rapidly has created a unique imbalance; we can create information much more swiftly than we can consume it. To help balance the information scales, we must develop and use technologies that enable us to discover, find, correlate and consume information at an ever-increasing rate.

One technology that comes to the rescue, helping to balance the scales, is enterprise search. Enterprise search is a broad term that encompasses a set of modern software technologies that enable us to navigate the sea of data in our professional and personal lives. These search systems provide mechanisms for accessing, classifying, and retrieving information, regardless of source or format, in a timely, efficient and cost-effective manner.

What is Relevancy?

One key measure of a search system is the degree to which a set of documents answers the information need expressed by a group of query terms. We colloquially refer to this as the system's "relevancy" and as you evaluate enterprise search products, you will hear this term again and again. So, what is relevancy, and why is it important?

People use the term "relevancy" as a way of describing how well the system addresses their information needs. They say the system provided "relevant results" or a "good set of documents" that helped them to answer their questions, find specific information, or further their research. People tend to judge the overall usefulness of the system based almost entirely on this experience. So, if the experience is positive, the system is viewed favorably, and is used often.

Determining the appropriate set of documents to retrieve, based on a very small set of search query terms, is the most significant job a search engine performs. It is the primary mechanism that determines your having a great vs. a terrible search experience. A high-quality relevancy algorithm seems almost magic - you type in 2 or 3 words describing what you're looking for and voilà, it delivers answers to your questions – with the system seeming to "know" what you wanted. On the other hand, a suboptimal relevancy calculation results in frustration. We've all experienced the search that gives us everything but what we are looking for! So, why do some systems answer our questions quickly and effectively, while others leave us wondering? The answer lies how the system calculates which documents to return, and the degree to which we can influence this calculation to align with our business objectives.

How a search system calculates relevancy can be based on a variety of algorithms and techniques. The most popular approaches are based on a variety of mathematical models: Boolean or exact match algebra, term-frequency and inverse document frequency statistics, vector spaces, and numerical probabilities.

Each one offers its own set of pros and cons, and manifestations in particular vendor implementations. In fact, by in large, search companies compete on the basis of their relevancy algorithms, the specifics of most are closely guarded trade secrets. Some of these relevancy systems are configurable while other are not, and this will to a large degree affect how closely these search systems can be adjusted to the particular content, users and the business needs of the information provider.

In order to understand a search system, you need to have some idea of how it calculates relevancy – the basics of the implementation. You must also understand whether you have any influence over the algorithm. Vendors vary widely here – some offer no control whatsoever. Their algorithms are complex black boxes – it "just works". Other vendors offer relevancy model adjustments via their professional services, and others offer open models that enable you to fluidly adapt the system to changing business needs and user communities. When evaluating search systems, make sure you understand these critical factors.

[1] An exabyte is a billion gigabytes.

[2] Source: UC Berkeley, School of Information Management & Systems, How Much Information?

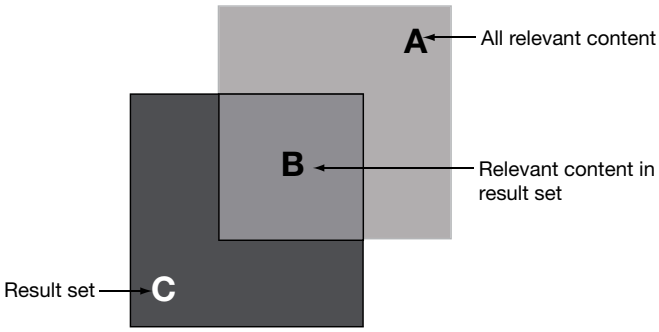
Although a complete explanation of search relevancy models and vendor approaches is well beyond the scope of this document, it is important to understand that regardless of technique applied, there are several indicators that help us to measure the effectiveness of a given relevancy system.

How does Relevancy work?

Search performance (a slightly more formal way of thinking about relevancy) is most frequently described in terms of precision and recall. Precision is an indication that a retrieved document is germane (relevant), and recall is an indication that a pertinent (relevant) document has been retrieved. Recall and precision are calculated as follows:

Recall = (number of relevant items retrieved)/(number of items in collection)
Precision = (number of relevant items retrieved)/(number of items retrieved)

Figure 1 helps to illustrate this idea:



Traditionally, recall and precision are inversely related, which means that for an increase in precision, there is a decrease in recall, and vice versa. For example, web search engines tend to have great recall, but poor precision – which of the 3.2 million results are right for you? Whereas, a good medical search application tends to give fewer results that are specific to the patient in question.

In addition to recall and precision, there is one more key metric that we use to determine search performance. Historically, this has been called the F-measure and is calculated as follows:

$F = (2)(\text{precision} \times \text{recall})/(\text{precision} + \text{recall}) = 2PR/(P + R) = 2/(1/R) + (1/P)$

As demonstrated, the F-measure is the harmonic mean of recall and precision. Using a harmonic mean helps minimize the influence of any one variable, while at the same time penalizing uneven contribution. In terms of measuring search performance, this means that a highly performing search system returns a high percentage of relevant documents, with minimized noise. A "good" F-measure is "large" with respect to the values of P and R.

And … how does Ranking work?

After a search system determines a result set for a given query, it assigns a relevancy score or rank value to each document in the set. It then orders the results according to the values of the relevancy score, highest (most relevant) to lowest (least relevant). Here, good ranking models increase the (perceived) precision of the results.

It is important to note that these scores are meaningful only to the system that computed them, are correlated to the query, and their values are generally relative to the set of documents returned. What all this means is that a given rank value is meaningless outside the context of a given query and result set, making it rather difficult to display them in user interfaces, or to map them to a percentage scale. In addition, since different systems calculate rank in their own way, it is complex to normalize the rank values across different search systems.

Furthermore, ranking means different things to different people. Internet patrolling applications where hate groups, terrorist activities or the illegal trading of child pornography are tracked, ranking becomes important to help determine threat assessment, or in the conviction of criminals. On the other hand, in litigation support, minimal ranking is preferred – here a document is either 'in', or 'out', and precision is of utmost importance.

So What? What Should I Be Looking For?

So, what does all this mean to you? What should you be looking for in a search system with respect to relevancy? In general, look for systems that offer a tunable relevancy model, and mechanisms by which to favor precision over recall.

People cannot effectively deal with millions (or even hundreds) of documents simultaneously – we need small sets of precise documents to help us answer our questions. We need systems that enable us to define queries in explicit terms, and that return clear-cut answers. We need systems that increase the signal-to-noise informational ratio. Look for systems that offer the ability to use statistical modeling, document structure, proximity, freshness, and lexical analysis to help formulate sound queries, surface key documents, and eliminate redundant and superfluous data.

Conclusion

Relevancy and ranking are central topics in search. The ability of a search system to correctly identify and score relevant documents with respect to a given query is paramount to its usefulness. As we have seen, relevancy is a function of recall and precision, and an effective search system balances these to maximum effect. The degree of balance between these relevancy components is captured in the F-measure.

After all of this explanation, it is interesting and important to note that for all practical purposes, relevancy is quite subjective. When we search, we have an information need – we are looking for information or an answer to a question. If the search system we use helps to answer our question quickly, we feel that we received "relevant" answers. The extent to which we feel this is highly correlated to what we expect or want the system to return.

This means that two people, using the same search system, against the same corpus, and issuing identical queries could perceive the results from opposite perspectives; one could find the results perfectly relevant and useful, while the other one could conclude that the exact same results are wholly irrelevant. Keep this in mind when developing your search system, as it places implicit demands on other aspects of your system, including information architecture, content processing and user interface models.

Happy searching!

About FAST SBP™ (Search Best Practices)

SBP consulting is a highly focused transfer of search knowledge and experience from FAST to its prospects and customers. SBP workshops aim to help enterprises realize the full potential of search, by creating optimal strategic, functional and technical roadmaps, delivered in the form of business model, solution and architecture designs.

For any feedback or questions related to this paper, please contact us at sbp@fastsearch.com.

Fast Search & Transfer
www.fastsearch.com
info@fastsearch.com

Regional Headquarters

The Americas
+1 781 304 2400

Europe, Middle East & Africa (EMEA)
+47 23 01 12 00

Japan
+81 3 5511 4343

Asia Pacific
+612 9929 7725

SWP.015.B.01.020806