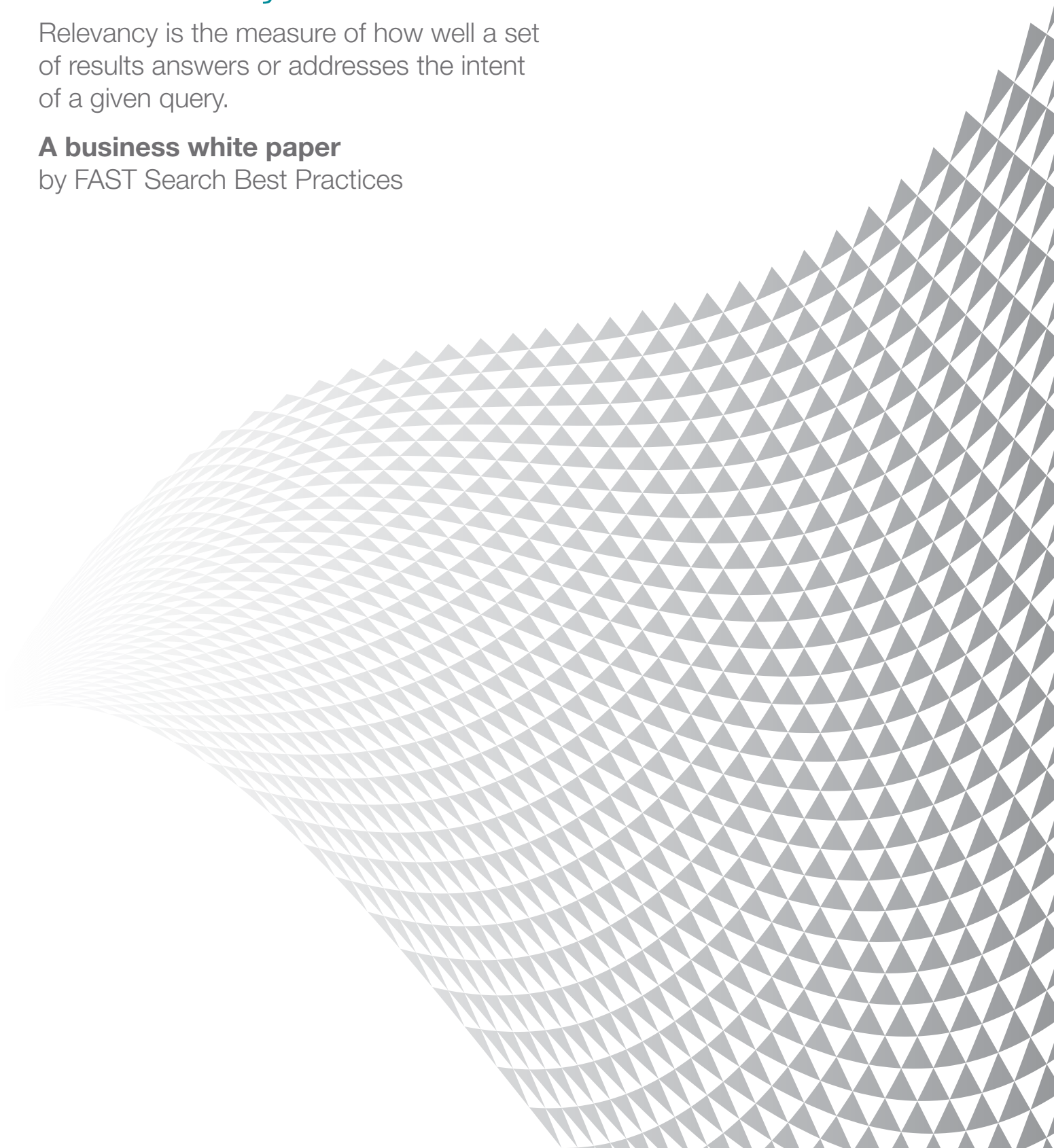


Relevancy and search

Relevancy is the measure of how well a set of results answers or addresses the intent of a given query.

A business white paper

by FAST Search Best Practices



5 things you should know about relevancy

1. Search users search in order to use information – not for the sake of searching
2. Relevancy success depends on understanding the context and characteristics of the search user
3. Adjusting relevancy models is like adjusting the graphic equalizer on an audio system – key components can be adjusted independently
4. Linguistic tools can improve precision and recall
5. Relevancy can be tested and tuned by using a “golden set” of well-known documents and queries

Relevancy is what really matters

In most companies, the role of search has evolved from basic keyword search in external Web pages to include comprehensive information retrieval solutions –static and dynamic – that cover a wide range of internal and external systems, applications, and networks. As a result, there is much more demand for sophisticated information retrieval functionality and performance.

Essentially, search engines must produce accurate and relevant results to meet the expectations of increasingly demanding users and applications – and to meet the needs of the business. That means it must now be possible to tune the relevance of the search engine.

Relevancy is the measure of how well a set of results answers or addresses the intent of a given query.

Search technology relies on assessment tools called ranking models to determine how closely content matches a particular query and whether it should be included in the search results. However, the ranking models of most search engines are inaccessible, so it's difficult and sometimes impossible to alter or tune them to meet different needs. In every case, they use the same yardstick.

However, search is not limited to one uniform environment. Business users search in multiple contexts: e-commerce sites, corporate Web sites, intranets, extranets, portals, etc. Each has distinctive objectives and each user community values content differently. So it makes sense to be able to adjust the yardstick used to evaluate content in order to get results that are aligned with the objectives of the search context and needs of the users.

What determines relevance? The relevance of a document is represented by a number or ranking value that helps determine how closely the document matches the characteristics implied by the query. The value is constructed from a number of factors based on the detailed analysis of all parts of the document. These include, for example, title, author, date, body, meta-tags, key concepts, classification, etc. These factors can be tuned independently, or in groups, to create ranking models for particular content – for instance, where the date of a highly relevant news article may be ranked higher than for news on the same topic.

A document's relevancy should be decided by the extent to which its content satisfies the user's need for information – not simply on the basis of the words common to both query and document.

Most search technologies employ a fixed ranking model that is designed for generalized use in a common context. A fixed model works very well within its frame of reference because it's optimized for that specific content. However, as you move away from its design base, a fixed model rapidly loses its effectiveness. An example: a public Web search solution would use a relevancy model that is geared to rank Web pages. However, it cannot effectively deliver the goods in an e-commerce context because the evaluation mechanism is incorrectly calibrated for e-commerce searches and usually can't be adjusted.

Similarly, in a knowledge management setting, many documents come from sources other than the Web: document management systems, customer relationship management systems, e-mail, file systems, etc. These documents need a different “quality comparison” mechanism.

Multiple dimensions enable full control

Enterprise search solutions are evolving far beyond standard Web search capabilities. They are now able to base ranking models on multi-faceted and tunable measurements of the quality or the match between the query and a possible result document. In this environment, relevancy is determined partly on the parameters discussed above, but also on concepts, sentiment, or additional levels of abstraction, such as freshness, completeness, authority, statistics, quality, and geography. (We will describe these levels in more detail later).

Q: I want one search platform to support my Internet, extranet, and intranet presence. How do I ensure that users receive relevant results?

A: If you use multiple rank profiles (with different relevancy parameters) based on the user's information needs, you'll be able to fully control the weight of each rank component for a given query. Information can be locked down so that only the users with the correct security profiles can gain access.

A multi-parameter model enables full control of the relative effect of each ranking component for a given query. Ideally, the search solution should provide a set of pre-defined relevancy model profiles that align with specific uses or audiences – site search, news, shopping, self-service, market intelligence, surveillance, etc. Organizations need to be able to apply these types of capabilities “out-of-the-box” or be able to easily tune relevancy models so they can be optimized for their target audiences. For example, for a Web site, page popularity is key and should be a priority, whereas for a news-monitoring application, freshness should be the primary factor, and sentiment context is important for market intelligence applications.

A convenient way to understand the importance of relevancy models is to visualize a graphic equalizer on an audio system, which has pre-sets for audio environments such as concert hall, car, home, classical, and rock, for example, and which also allows for individual adjustment to meet the needs of the listener. Similarly, search solutions must be capable of providing pre-set relevancy models where each of the parameters can be independently adjusted, and a change in one does not affect the others.

Many factors affect the relevancy of a document with respect to a given query. In general, it is impacted by the content of the document, the language in which the content is written, the degree to which lemmatization is employed, the extent to which stop words are removed, and the degree to which synonym expansion, thesauri, link cardinality, proximity boosting, and dynamic ranking are utilized. It should be noted that a document's relevancy with respect to a query is not necessarily decided on the basis of words common to both query and document, but rather by the extent to which its content satisfies the user's need for information.

The tuning parameters that define relevancy ranking models can have a profound impact on the search results themselves, and also on the associated business environment. Six key parameters stand out:

Freshness – how fresh is the document compared to the time of query? A number of new search applications require sub-second or non-stop updates to the index, so traditional batch processing is not an option. These new-generation solutions need to be able to search for news, receive stock alerts, or update the index with new products or pricing. In such cases, freshness is of paramount importance to users.

Completeness – how well does the query match superior document contexts such as a document title or URL? What matches the query? Is it the document title, the author, a mention in the body text, metadata linked to the document, root and expanded forms of words, etc.? For example, if the query is “Boston College”, then results citing “Boston University” would come up less readily if completeness is a key factor to the rank profile.

Authority – is the document considered an authority for this query? Many items can be part of the analysis of documents to determine this parameter – items such as Web link cardinality, article references, page impressions, and product revenue, to name a few. For example, with link cardinality, the search application will monitor the number of links in or out of a document, with high frequencies of links indicating the authority of a document.

Statistics – how well does the content of the overall document match the query? One simple example is the number of times the query term appears in the document. Another is the proximity of the words in the document – how close they are to one another.

Quality – what’s the quality of the document? How important is it when viewed from the perspective of the content owner or search application? For a corporate Web site, product landing pages and press releases are typical of the types of documents that have higher “quality” ratings. For a market intelligence application, documents from news services and industry analysts may be considered more important.

Location – how important is location in relation to the query term? For an Internet Yellow Pages company, there’s significant value in being able to maintain location entities at a contextual level to enable extreme precision in search and contextual navigation. The feature also offers and user stickiness – for example, if a user is searching for “BMW dealership within 15km of my workplace”.

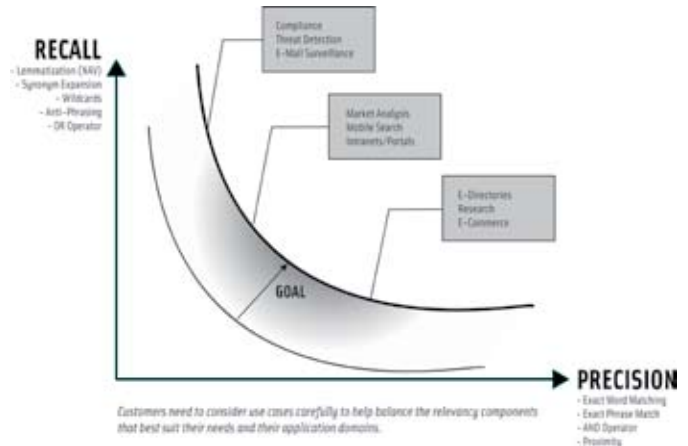
Tuning relevancy to meet the business need

Tuning the relevancy of search can lead to significant improvements not only for the search user (improved results to query terms), but also for the underlying business application (enhanced revenues, providing transparency for compliance initiatives, etc.).

One caveat, however: simply maximizing each of these controls will generally not lead to improved relevancy. If they are to achieve the goal of relevancy - balancing recall and precision - organizations need to understand the search application, the underlying information, and the search user. In general, customers need to strike a balance between recall, (finding everything related to a query) and precision (finding only those documents or entities that relates to a given query). More scientifically, recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the index. Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved.

When using search systems, the search application must strike a balance between recall and precision. The aim is to dramatically improve precision without sacrificing recall.

Optimizing recall or precision - or both?



For example, knowledge discovery or compliance applications (upper left of the curve in the diagram above) will rate recall as being more important than precision – in other words, customers do not want to “miss” any important documents when performing compliance-related searches. Here, customers may want to enhance the importance of the completeness and statistical parameters.

But in an e-commerce or e-directory environment, users prefer much more precise search results so that customers are not swamped with too many non-specific results. For instance, searching for “iPod” should return only results that are Apple Computer’s iPod players or accessories, not the similar players made by Creative or Sony. The business objective is to convert the browsing/researching user into a paying customer in as few clicks as possible.

Furthermore, by tuning the relevancy model or rank profile, e-commerce sites can promote particular products towards the top of the result set. These may be excess inventory items that the retailer wants to sell as quickly as possible, or products that generate the highest margin.

Linguistic capabilities (lemmatization, synonym expansion, phrase detection, etc.) become important here in order to correct spelling mistakes, and to effectively turn a “bad” query into a “good” query, avoiding zero

hit results. Detection of implicit phrases and proper names in the query protects them from further query transformation. By creating a list of product names, for example, the search provider can ensure that queries are directed to the desired pages that match the implicit product name phrase.

Q: I run an e-commerce site and I have excess inventory of an LCD monitor that will soon be superseded. How can I move this inventory out quickly?

A: Relative query boosting allows you to promote the ranking score to ensure that a particular document is always displayed in the first N hits on the result list, provided the user searches with a specific query. For all other queries, the ranking position will not be impacted by any boost, thereby maintaining high levels of user satisfaction.

Relevancy tuning becomes an important tool for OEM providers that leverage search to offer a value-added information access layer across their own application frameworks, along with integration with third-party structured and unstructured data sources.

Given the wide range of implementation scenarios and user profiles, OEM providers need to be au fait with rank profile concepts so they can control the ranking and sorting of query results, linguistic tools, entity extraction, and boosting techniques to achieve the best results.

The impact of relevancy

The primary aim of applying relevancy models is to bring forward the most useful information for a particular query. In other words, relevancy enables the system to best address the user's need for information. These models provide what are sometimes referred to as a "good-and-plenty" result profile, which refers to the precision, accuracy, and completeness of the result set relative to a given query set. Over time, customers can tune the relevancy models to meet the needs of diverse user environments and to shift the relevancy curve to the right, improving both recall and precision as a result.

From the standpoint of search providers serving organizations' needs, relevancy models are often underestimated.

Unfortunately, many organizations apply a one-size-fits-all methodology, which does not serve users' varied needs at all well.

In an e-commerce or e-directory setting, for instance, failure to find the right information at the right time can lead to lost sales when customers cannot locate the products and services they want. The costs are also significant for any large enterprise that depends on the skills and capabilities of its knowledge workers. In an IDC study titled "The high cost of not finding information", the research firm noted that poor search functionality can cost a business dearly if employees make poor business decisions based on faulty or poor information. And productivity plummets when different divisions and project teams reinvent the wheel. These tangible costs can be summarized as: 1) the time wasted on searching; 2) the cost of reworking information; and 3) the opportunity cost to the organization.

It is vital that content owners understand search users' profiles and their search expertise. The average user will type only one or two words into the query (and will often mis-spell at least one of them.) So content owners must combine linguistic mechanisms, advanced search, and relevancy tuning so the appropriate content can be brought forward to meet the needs of, say, a knowledge worker or an account executive.

Good enterprise search engines will place the power of relevancy in the hands of the business owner, not the IT department. The business manager or content owner defines and adapts relevancy models with respect to the user base, building in the appropriate business rules.

Customers may have to make relevancy trade-offs, such as the ingestion speed versus recall and precision, and the index size versus recall and precision.

In general, it's useful to process the content before indexing – using tools such as lemmatization – to maximize relevancy during query time. It's then helpful to store the enhanced metadata in the index that is searched. This helps reduce ingestion rates and expand the index, but it results in an unparalleled search experience – the measure against which search platforms are measured by users. One point worth noting: disk space is far less expensive than servers when considering the tradeoff between index size and ingestion rates.

Mini case study

IT company boosts sales with highly relevant search platform

Who

Global provider of IT hardware and services that uses its online channel for direct sales and service support.

Challenge

To provide customers with efficient access to the information needed to make informed purchasing decisions on over 14,000 products and solutions.

Solution

An intelligent and flexible search solution that provides highly relevant results by understanding the user's information needs and query patterns. The solution avoids futile queries by applying multiple rank models via multiple rank profiles and query mappings.

Technology

Advanced index profiles and multiple rank profiles to support product or support queries. Also, advanced linguistics influence rank based on the user's spoken language and locale.

In order to measure the efficacy of relevancy models, and to know how to tweak the models appropriately, search customers should understand the relationships among the body of information, the particular queries, the processing of documents, and the index and processing on the query side.

Four tips for improving relevancy

There is no escaping the need to understand the user base and the importance of providing different relevancy frameworks to suit different information needs and business models. Search providers cannot forget that users don't search for the fun of searching – they're looking for specific information.

The following tips are helpful for improving relevancy and search effectiveness:

Understand rank profiles – this clearly ties to understanding the user base, but it goes further. By leveraging rank profiles, organizations will be able to influence the relevancy calculations that are applied to documents. It is possible to have individual profiles configured for each of several user groups.

Influence the rank models – search providers can alter the rank calculation assigned to documents for given queries. This is particularly useful in e-commerce and paid-for-placement scenarios because they will boost the ranking of higher-paying advertisers.

Use linguistic tools – Apply lemmatization to improve precision and recall; synonym expansion to improve recall; and spell checking to prevent futile (0 hit) queries. Activate anti-phrasing to remove the “noise” from the query, such as the text of the phrase “how do I.”

Test, measure and refine – use a “golden set” of well-known documents and queries to test and tune relevancy. Providers should use at least 2,000 documents and more than 50 queries.

Guidelines and recommendations

Since relevancy is a complex subject, it's not surprising that many mistakes are made in addressing it properly.

For example, some organizations discount relevancy as a tool to bring out the right information at the right time. Users can be quite skeptical that machine-derived relevancy can produce the best results for their queries. These users tend to have had experience with systems that use taxonomy and classification as a primary means of information retrieval, so they tend to fall back on what's comfortable.

At the same time, some organizations tend not to understand the tools available to them. Consequently, they either do not use them or they configure them improperly (many customers wrongly configure the profile of the indexed content.) And others fail to use what we'll call a “golden set” of documents and queries – a mechanism used as a control in relevancy tuning exercises.

By understanding relevancy, organizations gain the room to be creative with their underlying business models and to experiment in the presentation of results. They can apply entity extraction to unstructured data and dynamic drill-down to structured content, for example. Tools like

these allow search users to determine what's relevant to them (price, rating, availability, etc.). They also provide a navigation experience, guiding users to the right answers in as few clicks as possible.

Of course, the final determination of relevancy is subjective, and measurements of relevancy must always include user satisfaction surveys. In the end, superlative search is possible when there are open, high performance relevancy models.

Frequently asked questions

Q: Where is the best place to use synonym expansion?

A: It depends on the nature of the search application. Synonym expansion can be applied at query time or when the document is processed. In order to provide flexibility and ease of management, this expansion is normally performed at query time. This also avoids the need to re-index when new synonyms are introduced.

Q: How can I ensure that mis-spelled product queries are directed to the relevant page?

A: Custom dictionaries can be developed to align with your specific content. Query terms can be detected and corrected with alternative tokenization. For example, if the dictionary contains the term "thinkpad", a query for "think pad" will be corrected to search for "thinkpad".

Q: When should I activate a wildcard search?

A: Activate wildcards only for fields that are not data-rich (that is, they lack appropriate metadata). Activating wildcards for data-rich fields can impair relevancy.

Q: Will it be time-consuming to manage many different rank profiles for my user base?

A: Not necessarily. The critical step is to understand the nature of the content and the user base. Assigning different rank parameters to specific types of content will simplify this process.

Q: How do I deal with duplicates?

A: A good search system will remove duplicates from the result set. Duplicates can be detected on several levels, differentiating between real duplicates and perceived duplicates. Field collapsing allows for the de-duping of results with identical value in any given result field.

Q: What is a static rank?

A: It's a query-independent, fixed rank that is given to a document during document processing. For instance, the CEO's quarterly newsletter may be static-ranked as #1 in the result set for any query within an organization.

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.

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

© 2006 Fast Search & Transfer ASA. All rights reserved.

Fast Search & Transfer, FAST, FAST ESP, and all other related logos and product names are either registered trademarks or trademarks of Fast Search & Transfer ASA in Norway, the United States and/or other countries. All other company, product, and service names are the property of their respective holders and may be registered trademarks or trademarks in the United States and/or other countries.

SWP.009.B.01.011206