fast SBP™

Reporting and benchmarking search

A search engine may be viewed as a set of servers that will run with little or no intervention. However, to guarantee user satisfaction, search performance and quality must continually be improved. The prerequisite is to monitor the performance and behavior of the system.

A technical white paper

by FAST Search Best Practices

5 things you should know about benchmarking search

1. It's hard to improve something that's not measured
2. Continuous benchmarking is vital to the ongoing success of a search system
3. Too many metrics can confuse administrators, so you should measure only what is relevant and manageable
4. Different stakeholders (users, systems administrators, business-unit managers, executives) will have difference benchmarking needs
5. The best search engines supply a wide variety of reporting tools and APIs that can be leveraged.

A search engine may be viewed as a set of servers that will run with little or no intervention. However, to guarantee user satisfaction, search performance and quality must continually be improved. The prerequisite is to monitor the performance and behavior of the system. For example, by tracking the average response times for queries over time, the service provider can see whether the system has enough capacity to meet peak loads.

Different stakeholders have quite different objectives. For example, a Yellow Pages user will be looking for local information while someone adding a search function to an IYP (Internet Yellow Pages) site wishes to increase advertising sales.

Reporting and benchmarking creates a structured method for measuring and validating the success of search with respect to these varied stakeholders. Without adequate measurements, enhancements may not be possible. The most popular queries, broken links, and user surveys are often used to evaluate the quality of search.

Within each vertical industry sector there are four major categories of people interested in benchmark data: the consumer, the executive team, the business unit manager, and IT. This paper will address how different areas of reporting and benchmarking are useful to these groups, and how they can be used to ensure that the needs of all players are satisfied.

What information is most useful to…

Reporting can be divided into two categories: searchable data and index information, and search engine usage metrics. The first category includes the number of documents in target repositories, an audit of those documents, and information on hardware usage. Usage measurements include hard numbers such as click-counting, and subjective measurements about the quality of the interface and the results ranking.

These categories can be split up further according to the stakeholder groups and the reports relevant to each. Identifying which group has an interest in which metrics helps to assign the ownership of search components to different groups; who is responsible for identifying sources of content; whose role it is to determine what good and bad results are; who is responsible for building and maintaining the platform?

Below, the different benchmarks and which category they are important to will be examined in turn.

… the consumer?

Search engines must keep track of how many documents the index contains, what the growth rate is, and how fresh the information is. These are typical of the factors that influence how users choose a Web search engine. So it's clear that good reporting of index volume can be used to influence a site's overall traffic.

Q: My e-commerce generates a lot of traffic, but not as much revenue as expected. How do I rectify this?

A: It's likely that visitors are coming to your site intending to purchase, but they're not finding compelling products to buy. Do many of their queries result in zero hits? If so, synonyms, lemmatization, and spell-checking can be used to increase recall. If lots of queries result in pagination, the ranking model may need to be changed.

By monitoring user behavior, it's possible to understand why the sales are low, and then act on that information to improve your site's profitability.

Comprehensive user-facing index statistics reports give information discovery users the confidence to use search for their professional investigations. For instance, if an attorney is using a litigation support tool to find all WorldCom e-mails that mention former CEO

Bernie Ebbers during 2003, he or she will need to know that the search is for 100% of all of the e-mails in that period. More realistically, it will be acceptable that some documents are not searchable, but there will have to be a detailed report of why those documents aren't searchable.

From a query point of view, it is sometimes worthwhile to show end users the most popular searches or items. For e-commerce, for example, the most popular sales reflect what the user is most likely to be looking for or may be tempted buy. Reporting that information directly to end consumers can drive sales by placing popular items prominently – the Internet equivalent of candy bars at the checkout line.

… to IT?

The IT team is responsible for maintaining the health of the search service. So the team must be able to monitor the vital statistics of the system, see what values are causes for concern, and identify the remedies for these anomalies.

The size of the index (in gigabytes of data on disk) and other information about the hardware will be monitored. The percentage of downtime for vital components or the repair time will help determine whether the high-availability strategy is appropriate for the SLAs (Service Level Agreements) that the IT team has committed to.

Search-specific reports have to be monitored including: Index throughput: A sudden drop in the number of documents indexed per day often indicates a problem with a connector. For example, the password to the repository may have expired or the quality of the network connection to the source has degraded. A surge in the index rate may change the hardware requirements of the search installation.

Index freshness: This means the delay between adding documents to a repository and making them searchable. This is critical for certain applications such as news search.

Index size: This refers to the number of documents indexed and the size of the search nodes on disk. It's important to monitor this data because the number of servers and their configuration will have been planned based on a sizing estimate. If the index size greatly exceeds the prediction, performance will suffer.

IT is also concerned with total query volume, query complexity, average query response time and its distribution over time. By monitoring query speed over peak periods, the user experience will meet the targets set by the project's business leaders.

Q: I get lots of complaints from my intranet users that they can't find what they're looking for.

A: What are your users looking for? The first step is to mine your query logs to learn what are your most popular queries. Then you should approach test users to find out what documents they would expect from their queries.

Either the correct corporate resources have not been indexed, or the popular and authoritative sources need to be boosted. You should use a specific document for query boosting for "expenses form", "lunch menu" and other queries where the text may not appear in the document.

Reports on the size of the index distributed by department (or other numbers such as registered users or query volume per division) can be used by IT systems administrators to charge back IT resources to other business units. Correct reporting from the search engine is essential to streamline this accounting process.

… the business-unit manager?

The business manager's targets are usually the financial and qualitative success of the search implementation, which can be measured with relevant query-related reports including:

Top queries: - the most popular keyword searches. These should be tested most frequently for relevancy. The topics searched may help when deciding what types of content should be added to the index.

Query volume, number of unique users, and average number of searches per user: These figures are used to measure the pick-up of search.

Average click distance: This describes the number of clicks it takes to find the correct information after initiating a search. For instance, repeatedly hitting "next" counts towards the number of clicks. A low click distance is a good indicator of an effective results presentation and appropriate rank tuning.

Futile queries: The most common queries that return zero results and the proportion of queries that are futile. These metrics are critical to the success of any system. An empty hit list indicates an unsuccessful search, and must be rectified by adding data sources or changing searchable fields. For example, if users often search by department ID because it was the easiest way to find information in a legacy database, then the department ID field should be indexed. Synonym lists can be used if the search user's vocabulary does not correspond to that of the content.

Abandonment rate: the percentage of times an end user leaves a search page without clicking on a result. This indicates that users are dissatisfied with the results.

Click-through rates (CTR) – the number of times a user clicks on at least one result. For sales-driven sites, CTR may mean the proportion of searches that result in a purchase. For Web sites with pay-per-click revenue models, the number of times that someone follows an advertising link is vital. A poor advertising CTR may indicate that the banners are not aligned with the users – products do not correspond to the user demographic or the algorithm used to link a search to a banner is not appropriate.

Q: I have embedded a search engine in my DMS. Should I expose any reports to end users?

A: Showing too many statistics to end users will confuse them. Search should simply be a seamless part of your DMS.

The uniqueness of a DMS is that users often search documents that they created, or that they heard of from a document's author. So it's important to provide user-facing reports on the document-processing stage – for example, an audit of documents where indexing is pending, or of documents that weren't indexed for some reason.

In both cases it is important to monitor page impressions, i.e. how often documents or products are viewed. This may be fed back to relevancy models for example.

Focus groups and test users can determine how "good" results are. Typically, the top N searches will be reviewed to determine whether the results were appropriate. It can help to perform the same tests on competitors' sites.

The business manager must define what constitutes satisfactory quality. To score relevancy models, a more structured approach involves devising rules or tests.

Example tests include:

At least two of the top ten results should be no more than 48 hours old

All test queries should return at least 20 hits.

Two types of rules exist: global rules and query-specific rules. Global rules are evaluated against result sets; query-specific rules are evaluated against result sets for specific queries. Automated benchmarking tests can be used to monitor the quality of search without daily manual intervention.

Once tangible numbers are measured, they become critical information for the search provider, and a feedback loop can be constructed to improve user satisfaction. In addition to these query satisfaction measures, the business manager will likely be given objectives such as revenue targets.

For an OEM launching a search-derivative application, the target will be linked to the volume and average price of sales of the final software product. For an e-commerce site that tunes ranking models, a change in the value and number of sales per search will demonstrate its effectiveness – or lack thereof.

… to executive management?

The executives are focused on the strategic goals of the organization, not on the detail of IT implementations. Nevertheless, search is often a strategic ally. Small companies rely on word of mouth and personal relationships to propagate expertise. If a business experiences rapid growth, an intranet project with search can be used to spread knowledge and best practices among the employees. A drop in the time spent looking for information and the amount of duplicate work done would both be measures of the success of such a search project. The uptake of search, such as the volume of traffic, will also be of interest.

When the core business function of certain employees is linked to search, their efficiency will be representative of the efficacy of the search tool. Information discovery workers and litigation support staff are two such examples.

For both OEMs and e-commerce sites, sales revenues and profit margins will be monitored. Executives will be interested in measuring incremental sales following the addition of a search function or an upgrade to a more advanced search engine.

Search implementations all share one point in common – the decision to invest in the software, hardware, and manpower required for the project will have been made based on an ROI (Return on Investment) calculation. This shows that the cost of implementation was outweighed by, for example, the cost of the increased productivity, increased traffic, advertising income, improved litigation success rate, or higher sales, which are all measurable quantities.

How to generate reports

Each stakeholder will have an interest in different reports to quantify the value and success of search. IT managers will monitor QPS rates, business managers will have an eye on the quality of results, and the chief technology officer will want to calculate the total ROI of the implementation. It is normally the responsibility of the technical teams to provide adequate reporting tools for each group. There are varying methods and applications to generate reports, and some general types of reporting to consider:

Content reporting

Basic content statistics include the total number of documents indexed, the number of documents per collection, and the last time the index was refreshed. Information on the content aggregation and document processing will be viewed using each connector's administrative UI. More advanced reports will need to use callbacks from the content API of logs to provide the most complete vision of indexing.

Query reporting

Query reporting will often be a mixture of directly accessible numbers and those culled from log files. This includes QPS, average and individual response times, most popular searches, and futile searches. More flexible search systems will provide an API to access log information and a mechanism to store past audits of statistical values. This must be aligned with the organization's log file retention policies.

Click-through reporting

Recording user behavior is UI-dependent since the search engine only has a record of searches, query refinement, and pagination requests, not clicks-through. A link proxy page or similar mechanism must be built into the interface to allow CTR logging. To fully leverage this information, it may be useful to store the search that led to the click-through and information about the user (IP address, time of query. etc). This will allow more complete usage profiling.

Mini case study
E-commerce site uses product popularity to boost sales

Who
Large multi-brand e-commerce site aimed at teenagers

Challenge
To tap trends in teenage fashion to boost sales.

Solution
Monitoring of the most popular sales, and segmenting that information by different criteria, such as sales of brand X or sales of items in category Y. This information is then used to prioritize the brands and categories which are most popular at any point.

Technology
Sales reporting and monitoring using industry standard e-commerce analysis tools. This information is then used by the content managers to tweak the ranking in the search engine for particular areas using static boosting.

Performance monitoring

Performance numbers are required to scrutinize the behavior of the search engine's hardware and software elements. Operating-system monitoring and enterprise management tools will provide information on disk, memory, and CPU usage, network bandwidth capacity, and other information with which to benchmark the system's performance. To some extent, this is true regardless of the nature of the software. For search engines, local disk I/O activity or the performance of the storage area network are critical to indexing performance while the document processing function is the most CPU-dependent.

To create reports, it is important to leverage the search engine's out-of-the-box reporting tools as much as possible. These tools include a simple graphical interface using data driven by internal log parsing and generated statistics. Should the system use an open architecture, the development team may decide to modify or augment the off-the-shelf tools to meet the enterprise's needs.

Alternatively the data available in a standard format such as XML or tabulated format can then be input into the reporting and trend analysis tool of choice.

Guidelines and recommendations for different segments

An understanding of the different benchmarking tools and the business goals is necessary in order to optimally monitor and improve search. This applies to the design of the system as well as to ongoing surveillance and improvements.

E-commerce and other business-to-consumer applications will be most interested in analyzing user behavior. Understanding the user base will lead to targeted improvements. Commercial considerations are also critical. Use of reporting and a feedback loop will allow for improvement.

In knowledge management scenarios such as archiving and compliancy, it's essential to monitor completeness of the data set. So the most important statistics to track relate to the connectors' behavior – crawl and index rate, document processing or indexing failures, and the size of the index.

The success of an internal search project is gauged by user satisfaction. If, after installation of search, pick-up is lower than expected, this may mean that employees are not satisfied with the results, preferring to rely on other methods of finding information. Satisfaction surveys and looking at top queries can also improve the search. There may be some common queries ("lunch menu", "expenses form", etc) where boosting can be used to make sure a certain document always comes to the top of the list.

Application integration (OEM) projects are an area where it is difficult to generalize reporting. Sales growth is a strong indicator, although others will be key depending on the final application. With any OEM project, the important decision is to determine how much of the search engine should be exposed to the end users and administrators of the product. This applies to reporting information – too little will increase maintenance and support issues but too much will confuse.

The fundamental steps to improve reporting

Precise and clear gathering and analysis of index and user statistics is essential for benchmarking and reporting. Without these measurements, there is no concrete method for defining the success of search. Without metrics, it is difficult to determine which parts of the application require tweaking and whether system modifications have had any impact on the user experience – or on the profitability of the implementation.

Other problems can arise when benchmarking becomes excessive. Too much information will likely confuse the system's administrators; it usually indicates a fundamental lack of understanding of the challenges and goals of the implementation. So with any search initiative, it's crucial to identify exactly what to measure, and how and when to measure it.

First, the key goals and challenges for the project can be identified, which leads to identifying the critical metrics. Then it's necessary to map the responsibility of the feedback loop for each area to the various project stakeholders. This "responsibility mapping" must be assigned to a trusted team member - someone who has the leverage and knowledge to get the necessary changes made.Unfortunately, search reporting and benchmarking is often forgotten. By choosing a search engine that provides the appropriate benchmarking tools, stakeholders can more easily measure clearly identified sub-systems and act on the results to continually assess and improve search.

Frequently asked questions

Q: Who's in charge and what do they do?

A: Different stakeholders are responsible for different areas of search. Make a plan for how your team will use reports and summary data to improve and refine the user's search experience.

Q: What do I use to test the system?

A: The utilities and administrative APIs of the search engine are the first points to consider. Then we recommend that you leverage production system testing and analysis tools to gain valuable insights into the system's operational profile. Also, deploying an off-the-shelf Web server monitoring tool can be a quick way to monitor traffic information.

Q: What do I change?

A: Based on the data in the reports, you can update the document processing, index schema and rank profile, and update the query and results processing to improve the user experience and profitability of the system.

Q: How do I fix empty result sets?

A: Futile queries are those that return zero results. You should investigate why these might occur – for example, because of spelling errors, semantics, or UI difficulties – and then work to correct the underlying causes .

Q: Do I have to deploy enough hardware to maintain peak loads to ensure a timely and high-quality search response?

A: As you look to improve the search system, you need to consider how to minimize TCO while still providing world-class service. Consider renting servers for peak loads – for example during holiday seasons for an e-commerce site.

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

SWP.001.T.01.011206