

**Connect ▶ Find ▶ Act**

## **Search Server Relevancy Ranking What You Need to Know**

**Technical Article**

October 2009

**Guy Mounier**  
CEO

## Abstract

Search engine relevancy ranking optimization should be a mission-critical activity to any organization that implements SharePoint or Search Server with the goal of promoting content sharing and reuse. Yet, it is often largely overlooked. The first part of this article goes through the reasons why such relevancy optimization is not done. The second part will then drill down specifically into how to tune relevancy ranking in SharePoint and Search Server, with tips and tricks, free tools available, and best practices. Finally, the third part explains the importance of measuring search effectiveness and how to define the related KPI's. Are users finding what they are looking for? Which content sources are most useful? Who are my top contributors? What are my most and least popular documents?

The information contained in this document represents the current view of BA-Insight on the issues discussed as of the date of publication. Because BA-Insight must respond to changing market conditions, it should not be interpreted to be a commitment on the part of BA-Insight, and BA-Insight cannot guarantee the accuracy of any information presented after the date of publication. This document is for informational purposes only. BA-INSIGHT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, AS TO THE INFORMATION IN THIS DOCUMENT. BA-Insight may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from BA-Insight, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

©2009 BA-Insight Inc.. All Rights Reserved.  
<http://www.ba-insight.com/> Longitude v3.0 for MOSS

## Contents

<b>Overview</b>	<b>4</b>
<b>Part 1 – Why Search Relevancy Ranking Optimization is Overlooked</b>	<b>4</b>
<b>Part 2 – How to Tune Search Relevancy in SharePoint or Search Server</b>	<b>4</b>
<b>Part 3 – How to Measure Search Effectiveness and Result Relevancy</b>	<b>7</b>
Manual Method of Measuring Search Effectiveness	7
Automatic Method of Measuring Search Effectiveness	7

## Overview

Search engine relevancy ranking optimization should be a mission-critical activity to any organization that implements SharePoint or Search Server with the goal of promoting content sharing and reuse. Yet, it is often largely overlooked. The first part of this article goes through the reasons why such relevancy optimization is not done. The second part will then drill down specifically into how to tune relevancy ranking in SharePoint and Search Server, with tips and tricks, free tools available, and best practices. Finally, the third part explains the importance of measuring search effectiveness and how to define the related KPI's. Are users finding what they are looking for? Which content sources are most useful? Who are my top contributors? What are my most and least popular documents?

## Part 1 – Why Search Relevancy Ranking Optimization is Overlooked

Many SharePoint Administrators and System Engineers, who invest a significant amount of time in deploying the infrastructure, will largely ignore the tuning of the search engine. There are several reasons for this, yet many of them are not justified. The table below separates out the myths from reality based on a BA-Insight online survey of 511 System Administrators and IT Managers.

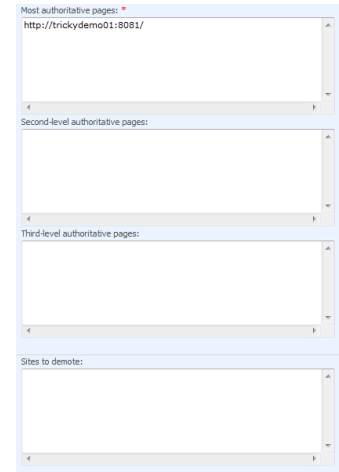
Reasons given for not investing in enterprise search relevancy ranking optimization techniques	The Reality
<b>#1 Reason.</b> If the search algorithm is good, then it really doesn't need any tuning, it just works out-of-the-box.	<b>Myth.</b> One size does not fit all. Your information architecture, the richness of the metadata describing those documents, the weights you give each property greatly impact the search result.
<b>#2 Reason.</b> I don't have the bandwidth to monitor user queries and affect relevancy ranking accordingly.	<b>Fact.</b> It is indeed very time consuming to manually monitor user queries and derive enough intelligence out of the usage data to improve search relevancy.
<b>#3 Reason.</b> I don't know how to do it. There is no Admin UI for adjusting the weights of the ranking algorithm. I wouldn't know what to adjust them to anyway.	<b>Fact.</b> The Admin UI is indeed non-existent. BA-Insight provides a free tool to manage such weights. We also provide guidance on what to set the values to. See the end of the article for free tool download location.

## Part 2 – How to Tune Search Relevancy in SharePoint or Search Server

There are two types of complementary strategies to tune and improve Search Relevancy in a search engine such as SharePoint or Search Server. The first strategy consists of improving the content quality and contextual facets. Better description of content with tags and categories make for more precise information retrieval. The creation of Best Bets for a given set of keywords also falls into this overall strategy. The topic of taxonomy and information architecture is out-of-scope in this article.

The second strategy, and the one we are going to expand on, is the ability of the search engine Administrator to affect the various levers of the relevancy ranking algorithms.

**Method 1:** Boosting Authoritative Sites. A content source that is deemed “authoritative” can be boosted. Assuming everything else is equal, a document stored in an Authoritative content source will rank higher. In the SharePoint Search Administration page, you can navigate to Authoritative Pages link to set a content source URL to most authoritative, second-level authoritative, third-level authoritative, or completely demote the content source. The screenshot on the right shows the user interface for listing the content source URL’s.



**Method 2:** Adjustments to Rank Parameters. As described in the screenshot below, a series of constant values governing the ranking algorithm may be reset. These constants are usually adjusted one at the time. Then the impact is measured over several queries and over time before adjusting other variables. Here are some of the most important constants you can adjust:

- The importance of term frequency in a document (k1). If your corpus is generic and single topic, where concepts are spread across documents evenly, you will likely want to decrease the saturation value for term frequency, as it is not a determining factor for relevancy. Instead, properties, currency of information, type of documents will play a stronger role.
- Weight of URL depth for calculating relevance (wud). If you are primarily indexing web sites or SharePoint sites for which the site structure breaks down a topic area in greater details, the URL depth weight is critical at distinguishing high value content residing near the root of the site from lower value content. Such URL depth weight will need to be doubled.
- File type (filetypepriordoc/html/XML/XLS/PPT). There may be certain type of documents, for example PowerPoint, Word, or PDF that you know contains more current or valuable content as it is the publishing format of choice in your organization. You may adjust the various weights of each file type accordingly.
- Language (languageprior). In multi-cultural environments, you may want to boost documents with a locale ID (language) that matches the end-user browser locale ID. This value is set to zero by default.

**Method 3:** Adjustment to individual Property Rank Weighting/Length Normalization and Rank Parameter.

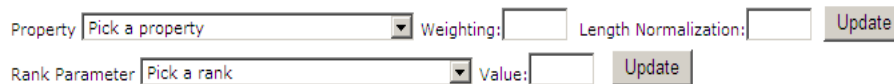
In most cases, modifying the property weight has a greater impact than modifying the length normalization setting. The range of possible values for this setting is 0 to infinity. Typically, you would want to configure this setting to a value between 1 (the weight setting for the body of a document) and 70 (the weight setting for the Title managed property). When the value is set to 0, the property is essentially removed from being used in the ranking algorithm.

Modifying the length normalization setting applies only to properties that contain text. For example, consider a scenario where relevance is calculated for two content items containing the query term within

the body of the content item, within a book, and within a document containing only a short paragraph. The book is likely to contain more instances of the query term, so it might receive a higher rank value, even if the shorter document is just as relevant to the user's search. The length normalization setting addresses this issue, providing consistency in ranking calculations for text properties, regardless of the amount of text within that property.

The range of possible values for this setting is 0 to 1. If this setting is 0 for a managed property, length normalization is turned off for that property; length normalization has the greatest influence for properties that have a setting of 1. For long text-managed properties, you usually want to set this to a value near to 0.7, which is the approximate setting for the body property. For managed properties that contain a small amount of text, but which are important to relevance, use the Title managed property's value for this setting, which is approximately 0.5.

In conclusion, it is worth noting that Microsoft only provides a programming interface to the adjustment of Weights and Rank parameters. As a result, BA-Insight developed a simple UI tool deployed in your Search Administration site to accomplish the tuning we discussed in this section during its own search deployments. The screenshot below depicts the UI. You may find more information on how to download this free tool at the bottom of the article.



The Enterprise Search ranking parameters are described in the following table.

Parameter	Defaults	Description
<b>k1</b>	16.404	Saturation constant for term frequency.
<b>Kqir</b>	2.12766	Saturation constant for click distance.
<b>wqir</b>	36.032	Weight of click distance for calculating relevance.
<b>Kud</b>	9.174312	Saturation constant for URL depth.
<b>wud</b>	31.468	Weight of URL depth for calculating relevance.
<b>languageprior</b>	0	Weight for ranking applied to content in a language that does not match the language of the user.
<b>filetypepriorhtml</b>	166.983	Weight of HTML content type for calculating relevance.
<b>filetypeprior doc</b>	163.109	Weight of Microsoft Office Word content type for calculating relevance.
<b>filetypepriorppt</b>	163.367	Weight of Microsoft Office PowerPoint content type for calculating relevance.
<b>filetypepriorxls</b>	153.097	Weight of Microsoft Office Excel content type for calculating relevance.
<b>filetypepriorxml</b>	158.943	Weight of XML content type for calculating relevance.
<b>filetypeprior txt</b>	153.051	Weight of plain text content type for calculating relevance.
<b>filetypepriorlistitems</b>	0	Weight of list item content type for calculating relevance.
<b>Filetypeprior message</b>	160.76	Weight of Microsoft Outlook e-mail message content type for calculating relevance.

## Part 3 – How to Measure Search Effectiveness and Result Relevancy

Measuring how well a search engine is doing comes down to understanding how well a particular result set matches the user query. Such matching process can be monitored automatically or manually. The two methods are discussed in further detail in this section.

### Manual Method of Measuring Search Effectiveness

There are a number of metrics out there that attempt at measuring how relevant a search engine is at answering user queries. The most commonly used is called Precision. Precision is the ratio of the number of relevant results in the first 10 search results divided by 10. The number of search results may be increased from 10 to 20, 40, or 100 to increase the accuracy of the measure. This being said, do realize that 90%+ of users will not go beyond the first page of 10 results. Therefore precision at 10 is an excellent way of measuring how relevant your users perceive the search engine to be, regardless of how it is actually doing on the top 100 results. Relevant results on page 2 to 10 will simply be ignored. Here is a simple example of a calculation:

- Two relevant results in the first page of 10 results: Precision = 20%
- Three relevant results in the first page of 10 results: Precision = 30%

For the precision calculation to be statistically meaningful, you will need to take a sample of real users—not subject matter experts, taxonomists, or search experts. Let them review a sample of real users' queries, culled from real query logs. Then give them a sample of real documents and ask them to grade and tag the documents in a SharePoint list as “Relevant” or “Not Relevant” against each specific query. Finally, run the queries against the search engine and sum the number of documents tagged “Relevant” for that query. Aggregate the results to derive an average Precision. The value of this method is that users are asked to grade and tag documents once. Running of the search engine precision test can be done by an Administrator, repeated as many times as desired, and possibly even automated. It can be valuable to automate the test script if the Administrator intends for example to adjust the weights of the ranking algorithm and measure the impact on Precision.

This method works well. It is obviously time-consuming and getting this type of feedback from real users may be a challenge. The method below describes how to automatically capture such feedback.

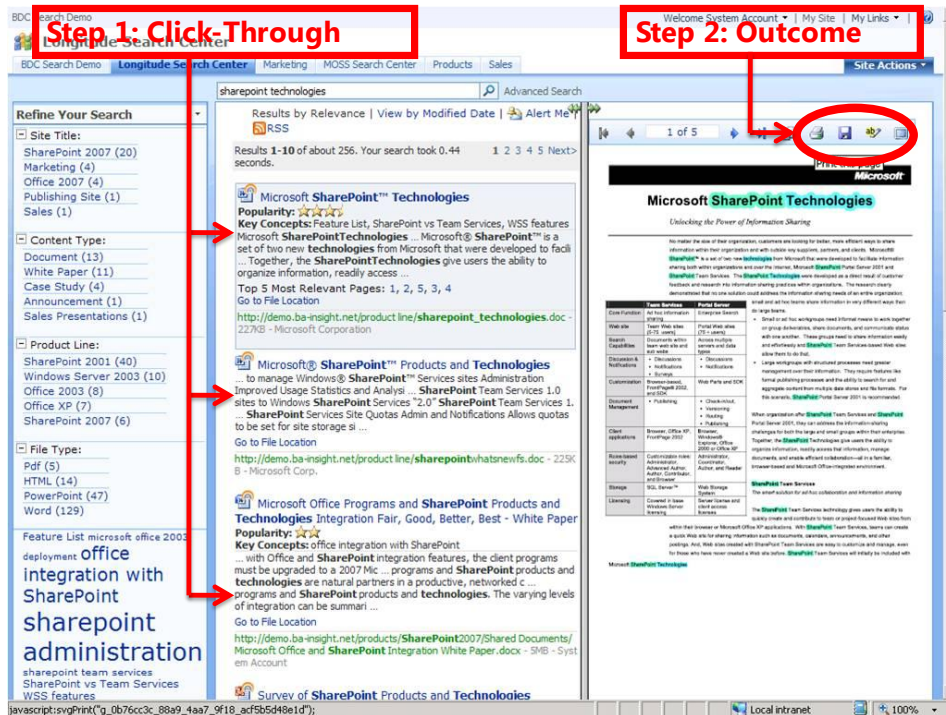
### Automatic Method of Measuring Search Effectiveness

The key to this method is the ability to get end-user feedback about a particular search result in the context of the search session. From experience, we find that it is very hard to ask for manual feedback. Less than 2% of users will actually provide such feedback, making the feedback data not statistically relevant because most documents will have no feedback linked to them, unless you are dealing with an extremely large number of users and quite small document corpus. It is typically the other way around.

The only true option is to systematically measure such feedback. E-Commerce site Amazon.com for example captures how many times a product is viewed in a page and then added to the Shopping Cart. Such conversion ratio enables Amazon.com to accurately promote popular products to the general public, but also to you specifically based on your buying pattern. By comparison, the Google Search Appliance

(GSA) will automatically measure how many time a user clicks on a link after being listed in a search result. Such conversion from Listed to Downloaded is an essential metric in calculating search ranking in GSA. Unfortunately, it hasn't proven to work as well as Amazon.com personalization engine. The main reason is that the feedback is noisy. Most users download a document to assess relevancy, not to work with it. Very often, the search result context with document tags and snippet is simply not enough to precisely assess relevancy.

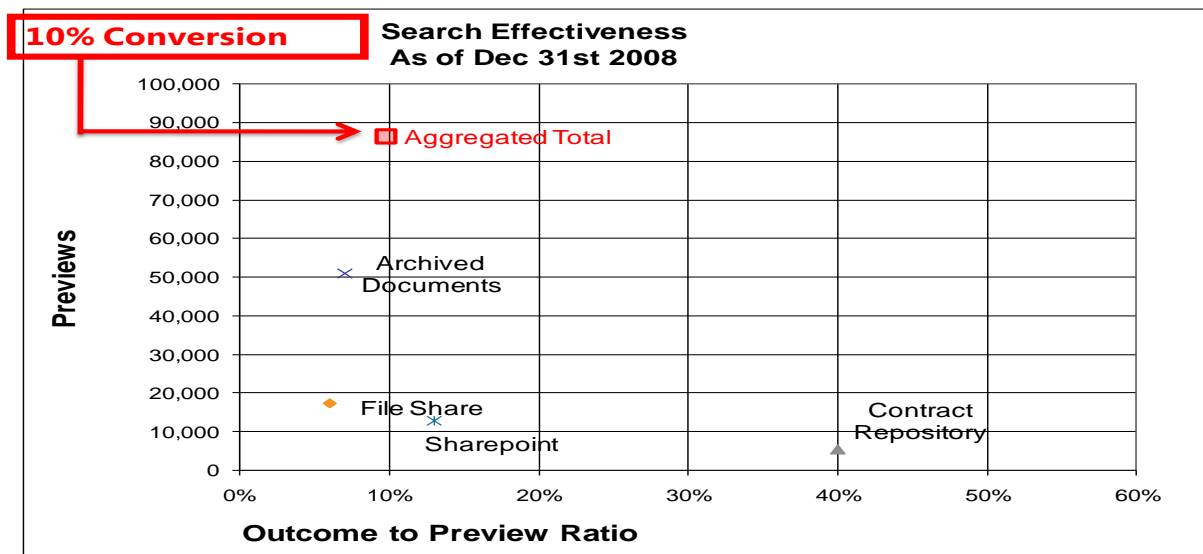
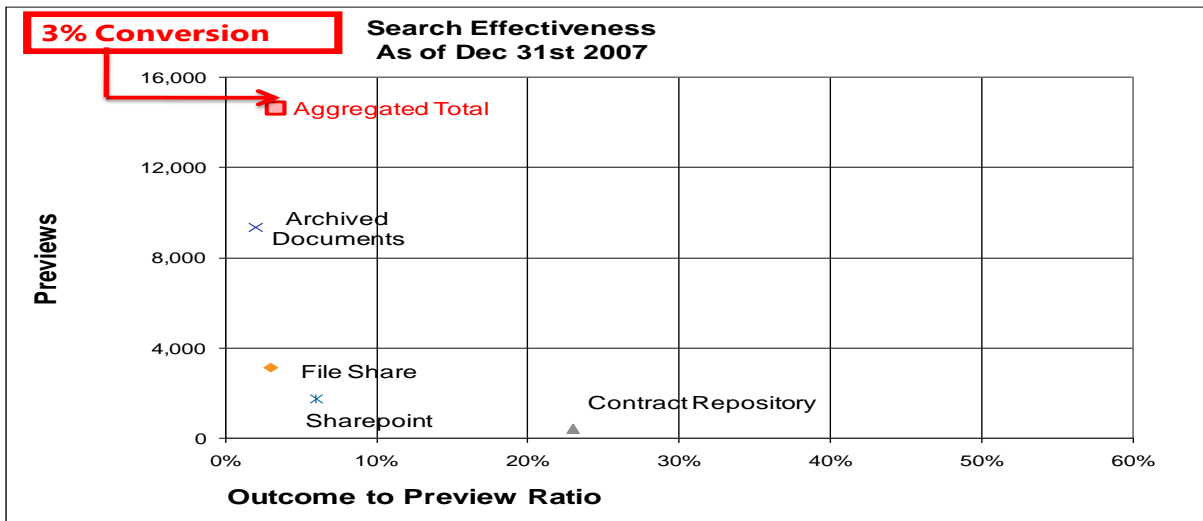
BA-Insight has invented a unique patent-pending method to accurately measure end-user relevancy feedback as depicted in the picture below.



With the ability to preview a document, relevancy assessment is now guaranteed before any other action is taken. In fact, when clicking on a SharePoint or Search Server result link, the user is brought instantly to the most relevant pages of the document. As a result, any subsequent actions such as Download, Print, E-mail, or Copy & Paste of text will be recorded as the document is actually relevant to my query, and the connection between the user query and the document is recorded in our usage log.

A detailed list of reports and KPI's enabled by our Search Analytics solution is available at the bottom of the article with a link pointing to the upcoming v3.6 release with Search Analytics. These reports include usage trend analysis, most relevant content sources, top contributors, most and least popular content, etc. In conclusion, we simply want to point out how such relevancy feedback can be automatically fed back to the search engine by adjusting the various weights and parameters in the search result. The net result is a two- to three-fold relevancy improvement over a 6 to 12 month period. It underscores the importance of search relevancy tuning, as from an ROI perspective, a 100% to 200% search effectiveness boost means

2 to 3 times faster ability to find information and 2 to 3 times more content reuse. With users spending 2.5 hours per day on average searching for information, according to IDC, that is a very big deal. The graphs below show you the net improvement of search effectiveness at a major consulting organization. The X-axis represents the search effectiveness (Conversion) as the ratio of number of relevant documents downloaded divided by total number of clicks or previews. The Y-axis shows the number of clicks or previews (in thousands). The Conversion ratio jumps across all content sources from 3% to 10% between Dec. 31st 2007 and Dec. 31st 2008, meaning that for every one hundred clicks to preview the document, the number of downloads jumped from an average of 3 to 10. Remarkable result which is repeated across all organizations who are serious about the ROI of Enterprise Search.





**United States**

445 Hamilton Avenue, SUITE 1102  
White Plains, New York 10601  
United States of America  
Tel | 914.220.8395

**Europe**

Vester Voldgade 106,  
DK-1552 København V  
Denmark  
Tel | +45 2161 6500

Web | [www.BA-Insight.net](http://www.BA-Insight.net)  
Email | [info@ba-insight.net](mailto:info@ba-insight.net)