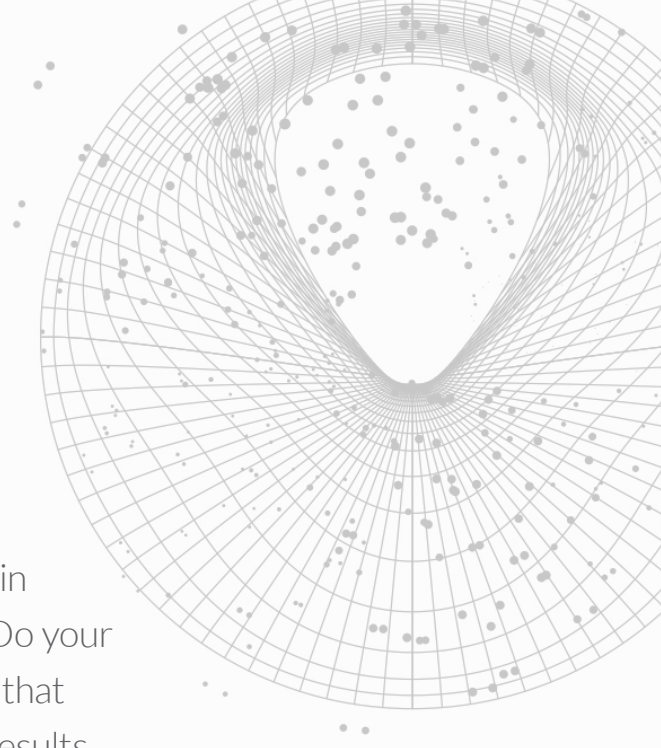


# SharePoint Search Tuning Guide

By John Lenker, Senior Sales Engineer, Lucidworks



Do you experience difficulty with SharePoint Search in getting the results you want, in the order you want? Do your users complain about seemingly nonsensical reasons that the documents they are looking for appear lower in results than they expect? Does the search engine seem to exert too much control and influence over your results? Do you wish your team had more control?

Here is a search tuning guide for SharePoint that hopes to resolve, or at least lessen the impact of, the plights listed above. First, we'll do a quick rundown of 7 ways to tune your SharePoint deployment—and then go into deeper detail of the how-to for each one.

## What to Tune

When SharePoint Search was a standalone product, especially when it was still independent and sold as FAST ESP (Enterprise Search Platform), it performed more like one would expect of a modern enterprise search engine. However, through Microsoft's "blackboxing" and insulation of features and hiding general configurability from the search administrator, many of the features of FAST have been relegated to mysteries and require modification to optimize.

### Tune 1: Boost on Keywords

If you use taxonomies or managed metadata in SharePoint, by default any tags applied to documents do not influence ranking in search

results. In many cases, it's desirable for this influence to be applied, which is a main reason for using taxonomies or managed metadata in the first place.

## **Tune 2: Boost on Proximity**

SharePoint shows one of its weaknesses in the relatively small impact that proximity of search terms has on the rank of a document in a search result set. Documents with frequent occurrences of a three-term search query in the body often outrank a document with those three terms present in the title. Even worse, if the title matches that three-term query exactly, it is still pushed below other results!

## **Tune 3: Boost on Freshness**

The age of a document may have immense importance to your site's users (for instance, a news site), or it may have none (an online encyclopedia). Freshness should have at least some influence on search results, and SharePoint does not account for it by default.

## **Tune 4: Customize Spelling Dictionary**

Unfortunately, you cannot actually change the spelling dictionary, but you can change the Query Spelling Inclusions and Exclusions lists, which may offer some level of control over your corrections.

## **Tune 5: Boost on FileType**

Often, you will find that PowerPoint documents have more weight in SharePoint search results vs. PDF or Word, and with all other factors being equal, there doesn't seem to be a clear reason why. However, in most cases, especially when it comes to enterprise search, PDF and Word doc content is more relevant to a user for most of their queries than PowerPoint.

## Tune 6: What about XML, CSV, and other structured content?

This is not exactly a tuning option, but if you have worked with SharePoint for even a short time, you probably already know that it's not very good at ingesting non-native content, such as non-Office documents and other structured data like XML, CSV, or JSON. To ingest such content properly, you'll need to use custom parsers, for example, to index the first row of a CSV as a header or parse individual nodes in your XML as metadata (crawled) properties.

## Tune 7: Equality for Synonyms

SharePoint does not give equal weight to synonyms, even if you use its synonym dictionary (also known as the thesaurus). The weight given to a synonym is only 20% the weight of the original term. This can be corrected with some manual intervention so the weight of both the original term and its synonym are equal.

## How to Tune

Keep in mind this is not a comprehensive guide. Guidelines will be given, but step-by-step instructions to achieving these tunes are beyond the scope of this guide. However, when a web reference with more implementation detail is available, it will be provided.

## Tune 1: Boost on Keywords

There are multiple ways to accomplish this, you have several options:

1. Use a custom rank profile
2. Use dynamic ranking via XRANK or
3. Change the weight of managed properties.

The easiest is 3) the last one, changing the weight of managed properties.

In SharePoint, weights in the default rank model are set in the following way:

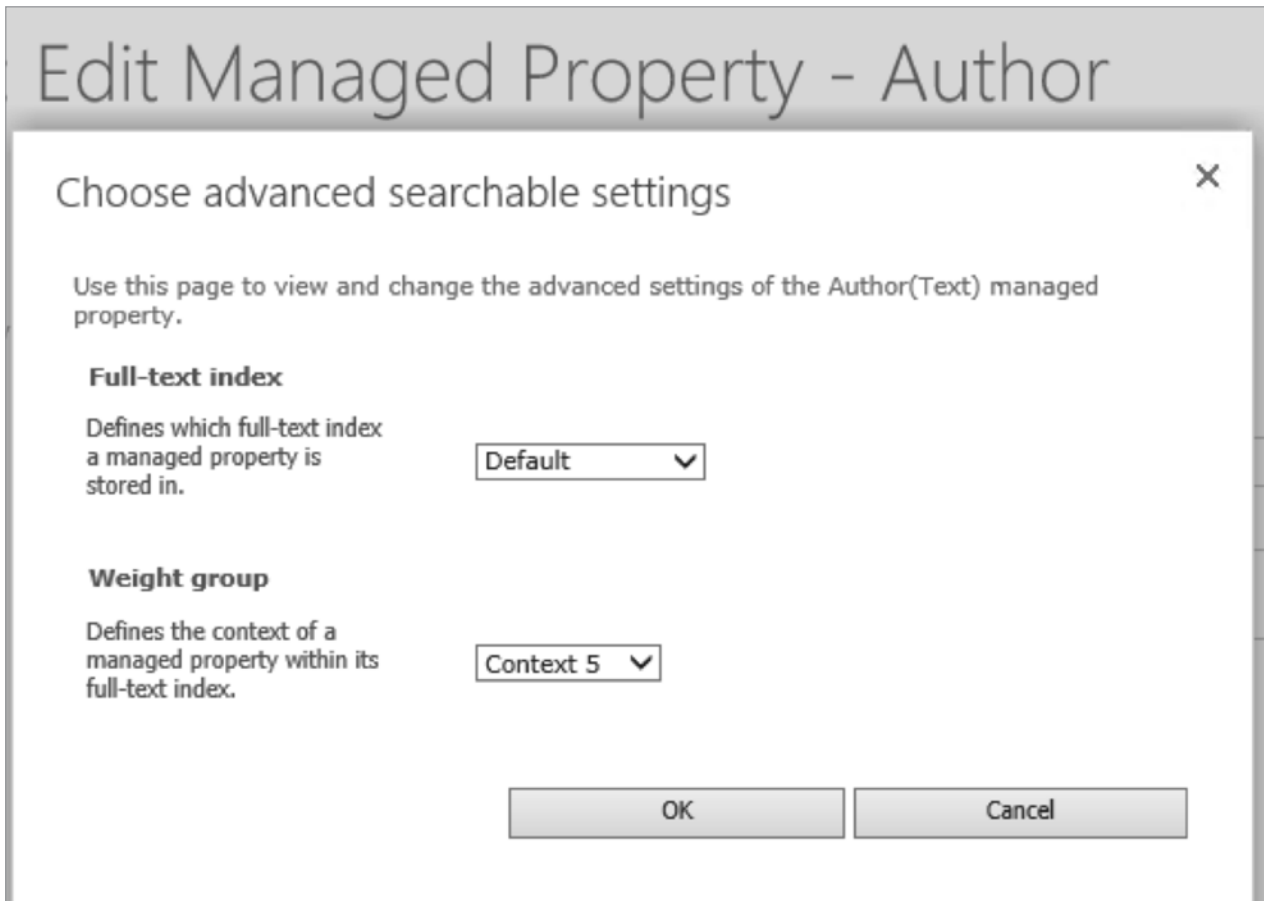
- Title 100%
- QLogClickedText 72%
- SocialTag 59%
- Filename 52%
- Author 41%
- AnchorText 18%
- Body 7%

Note that those numbers are n relative importance to each other, not anything else.

To see where a particular managed property falls in weight, you need to inspect its **Weight Group**. This is found in the Advanced Searchable Settings for that managed property. For example, Author is mapped to “Context 5”. This is merely a lookup value and does not correspond to anything else. Here are the Contexts:

- Title 1
- QLogClickedText 3
- SocialTag 14
- Filename 2
- Author 5
- AnchorText 6
- Body 7

To change the weight of *Author* so it's of the same relative importance as *Title*, for example, you just need to change its Weight Group to “Context 1”.



Giving more weight to tags and managed metadata is accomplished in the same way, either with the use of a new managed property (set to searchable) that maps to `ows_taxId_MetadataAllTagsInfo`, or, if you prefer, editing one of the core managed properties (`owstaxidmetadataalltagsinfo`) to be searchable.

## Tune 2: Boost on Proximity

This is a two-part solution. The first is to use XRANK to boost any document that contains the query terms in the Title property. This does not do anything for proximity, but is a necessary first step to achieve proximity boosting. Here is the syntax that should be used in your Query Text section under “Change Query” after editing the Search Results web part:

```
{searchTerms} XRANK(cb=100) Title:"{searchTerms}"  
XRANK(cb=5000) Title:"{searchTerms}"
```

This will use the submitted terms as the query, but dynamically boost any document where those terms appear in the Title property. You may change the value of cb (“constant boost”) as you see fit.

The second part is modifying the rank profile entitled “Search Ranking Model with Two Linear Stages.” Make a copy and use it as a basis for your own. Modify the boosts for proximity using the following strategy:

1. If all of the query terms are found in the item’s content (body), in the same order, and with a maximum of one word in between (maximum distance is 1), the item is given extra rank.
2. If all of the query terms are found in the item’s title, in the same order, and with a maximum of one word in between (maximum distance is 1), the item is given extra rank.
3. If all of the query terms are found in the item’s title, in the same order, and with no words in between (maximum distance is 0), the item is given extra rank.
4. If the query terms match the item’s complete title exactly, the item is given extra rank.

Custom rank profiles are a more intensive process than what fits within the scope of this document. For more detail on modifying a rank profile, please see:

<https://docs.microsoft.com/en-us/sharepoint/create-custom-ranking-model>  
<https://docs.microsoft.com/en-us/sharepoint/dev/general-development/customizing-ranking-models-to-improve-relevance-in-sharepoint#proximity>



## Example prior to boosting:



[Everything](#) [People](#) [Conversations](#) [Videos](#)

Preference for results in English ▼

 [Spirit 2013.pdf](#)  
SPIRIT 2013 Statement: Defining Standard Protocol Items for ... Annals of **Internal Medicine** Research and Reporting Methods 200 5 February 2013 Annals of **Internal Medicine** Volume 158 ...  
[localhost/C\\$/Users/.../Downloads/LifeSciences/Spirit 2013.pdf](#)

 [medi-97-e0459.PMC5916650.pdf](#)  
Seoul Paik Hospital, d Department of **Internal Medicine**, Yonsei University College of **Medicine**, e Department of **Internal Medicine**, University of Ulsan College of **Medicine**, Asan ...  
[localhost/C\\$/Users/.../PubMed/4e/d6/medi-97-e0459.PMC5916650.pdf](#)


 [Equine idiopathic hemorrhagic cystitis: Clinical features...](#)  
This syndrome may be difficult to ... Journal of Veterinary **Internal Medicine** published by Wiley Periodicals, Inc. on behalf of the American College of Veterinary **Internal Medicine** ...  
[localhost/C\\$/Users/.../Downloads/PubMed/JVIM-32-1202.PMC5980356.pdf](#)

 [kjim-2015-267.PMC5339463.pdf](#)  
Jung-Hwan Yoon, M.D. Department of **Internal Medicine** and Liver Research Institute, Seoul National University College of **Medicine**, 101 Daehak-ro, Jongno-gu, Seoul 03080 ...  
[localhost/C\\$/Users/.../PubMed/3f/kjim-2015-267.PMC5339463.pdf](#)

 [kjme 28-3.hwp](#)  
Chan Soo Shin 1 1 Department of **Internal Medicine**, 2 Office of Medical Education, and 3 ... Seoul National University College of **Medicine**, 103 Daehak-ro, Jongno-gu, Seoul 03080 ...  
[localhost/C\\$/Users/.../Downloads/PubMed/kjme-2016-34.PMC5016264.pdf](#)





After boosting:





[Everything](#) [People](#) [Conversations](#) [Videos](#)


Preference for results in English ▼

 [Internal Medicine Vol.56](#)  
Two hours after iliac crest bone harvesting, the patient experienced syncope and severe ... in the right common femoral artery, an **internal** mammary (IM) 5 Fr diagnostic catheter was ...  
[localhost/C\\$/Users/.../PubMed/4f/03/1349-7235-56-2639.PMC5658532.pdf](#)

 [Internal Medicine Vol.57](#)  
Division of Nephrology, Department of **Internal Medicine**, Shinshu University School of **Medicine**, Japan and 3 Department of Biomedical ...  
[localhost/C\\$/Users/.../PubMed/4d/ad/1349-7235-57-0127.PMC5799070.pdf](#)

 [Internal Medicine Vol.56](#)  
The **Internal Medicine** is an Open Access article distributed under the Creative Commons ... 2017 The Japanese Society of **Internal Medicine**  
<http://www.naika.or.jp/imonline/index.html>  
[localhost/C\\$/Users/.../PubMed/4f/2a/1349-7235-56-0231.PMC5337474.pdf](#)

 [Internal Medicine Vol.57](#)  
which is asso 1 Department of **Internal Medicine**, Kitakyushu Municipal Medical Center, Japan, 2 Department of **Internal Medicine**, Japan ...  
[localhost/C\\$/Users/.../PubMed/4d/9e/1349-7235-57-1533.PMC6028685.pdf](#)

 [Internal Medicine Vol.56](#)  
Japan and 3 Department of **Internal Medicine**, Yokote Municipal Hospital, Japan Received for ... 1982 estimated that the incidence of **internal** organ perforation due to toothpick ingestion ...  
[localhost/C\\$/Users/.../PubMed/4f/51/1349-7235-56-0657.PMC5410475.pdf](#)

## Tune 3: Boost on Freshness

This can be done in your new rank profile, but can also done via XRANK. It's best not to get carried away with XRANK as it can impact query performance, but it is quite useful. You will need to add XRANK parameters based on the write managed property similar to this:

```
XRANK(cb=1) write>2018-01-01
```

So, anything with a modified date in the year 2018 will get a small boost. This can get much more detailed, giving different weights to different time spans. For more detail, see the following:

<https://www.techmikael.com/2013/10/adding-freshness-boost-to-sharepoint.html>

<https://docs.microsoft.com/en-us/sharepoint/dev/general-development/customizing-ranking-models-to-improve-relevance-in-sharepoint#freshness>

Using the same search as the above example (“internal medicine”) without boosts to the Title, here are the results with a freshness boost. Note that the document with a Title beginning “Spirit 2013” has fallen out of the top 5 results (was previously the top result):

internal medicine XRANK(cb=1) write>2018-01-01

Everything People Conversations Videos

Preference for results in English ▾

[medi-97-e0459.PMC5916650.pdf](#)  
Seoul Paik Hospital, d Department of **Internal Medicine**, Yonsei University College of **Medicine**, e Department of **Internal Medicine**, University of Ulsan College of **Medicine**, Asan ...  
[localhost/C\\$/Users/.../PubMed/4e/d6/medi-97-e0459.PMC5916650.pdf](#)

[Equine idiopathic hemorrhagic cystitis: Clinical features...](#)  
This syndrome may be difficult to ... Journal of Veterinary **Internal Medicine** published by Wiley Periodicals, Inc. on behalf of the American College of Veterinary **Internal Medicine** ...  
[localhost/C\\$/Users/.../Downloads/PubMed/JVIM-32-1202.PMC5980356.pdf](#)

[jkms-33-e112.PMC5890082.pdf](#)  
Guri Hospital, Guri, Korea 2 Department of **Internal Medicine**, Chonbuk National University School of **Medicine**, Jeonju, Korea 3 Department of **Internal Medicine** ...  
[localhost/C\\$/Users/.../PubMed/4f/18/jkms-33-e112.PMC5890082.pdf](#)

[medi-97-e9782.PMC5882450.pdf](#)  
The impact of IL-28B in Asian ... University Hospital, c Faculty of **Internal Medicine**, College of **Medicine**, Kaohsiung Medical University, d Department of **Internal Medicine**, Kaohsiung ...  
[localhost/C\\$/Users/.../PubMed/4e/54/medi-97-e9782.PMC5882450.pdf](#)

[jkms-33-e196.PMC6000598.pdf](#)  
Sang Hong Baek Cardiovascular **Medicine** Division, Department of **Internal Medicine**, The Catholic University of Korea ...

## Tune 4: Customize Spelling Dictionary

Please see the following guide for managing the Query Suggestions in SharePoint:

<https://docs.microsoft.com/en-us/sharepoint/search/manage-query-suggestions>

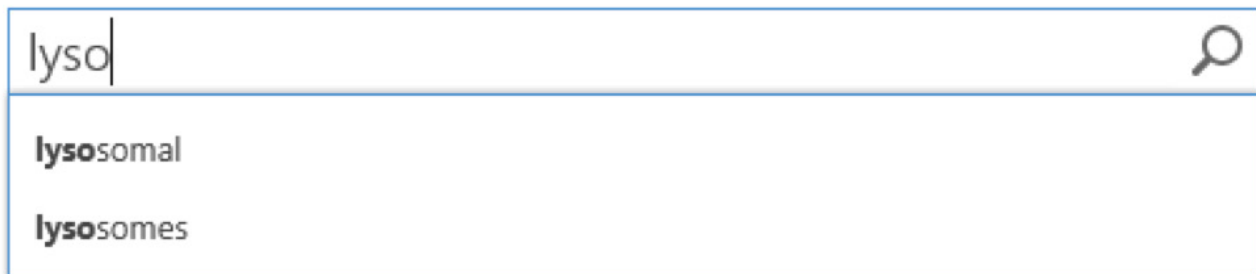
This guide covers the inclusion and exclusion lists for spell correction:

<https://docs.microsoft.com/en-us/sharepoint/search/manage-query-spelling-correction>

This is the general linguistic features for SharePoint, including the languages that are supported for correction. Note that dynamic correction occurs when an index contains several thousand items.

### Example suggestions in text file:

- lysosomes
- internal medicine
- lysosomal
- microbical
- electric potential difference



## Tune 5: Boost on FileType

A rank profile customization is appropriate for this boost, like the others, but also like the others, it can be achieved via XRANK. Note, using XRANK for this is a bit of a hack since it will boost the same way for every query. In that sense, using a custom rank profile is better. Here is an example of XRANK nonetheless:

```
XRANK(cb=5) FileType=pdf XRANK(cb=3) FileType:doc XRANK(cb=-1)
FileType:ppt
```

## Tune 6: What about XML, CSV, and other structured content?

You can certainly ingest XML and other structured content with SharePoint's default crawling mechanisms. However this is not entirely useful as XML tags are not assigned to managed properties without using Business Connectivity Services and SharePoint Designer—and that's only for display, never mind full-featured search. As an example, consider the following medical clinical trial data, available from <https://clinicaltrials.gov/ct2/resources/download> :

```
<required_header>
  <download_date>ClinicalTrials.gov processed this data on July 11,
2018</download_date>
  <link_text>Link to the current ClinicalTrials.gov record.</link_text>
  <url>https://clinicaltrials.gov/show/NCT00002535</url>
</required_header>
<start_date>July 1993</start_date>
<completion_date type="Actual">July 2004</completion_date>
<phase>Phase 2</phase>
<study_type>Interventional</study_type>
<has_expanded_access>No</has_expanded_access>
```

Without a lot of help, SharePoint cannot put these XML tags into searchable properties. Most admins are not in the business of making something like this possible. When this XML document is crawled using the fileshare crawl, the entirety of the XML gets put into the body of the searchable document, and the filename is assigned to Title, but that's all. No property is created for phase, has\_expanded\_access, etc.:

[NCT00002535.xml](#)

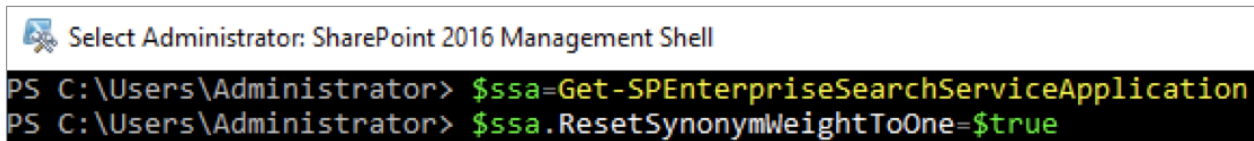
TIL; IL-2. Regimen B: **Prostaglandin** Inhibition Therapy plus Biological Response ...  
expectancy: At least 3 months **Hematopoietic**: (unless tumor involvement of bone ...  
[localhost/C\\$/Users/.../Downloads/ClinicalTrials0/NCT00002535.xml](#)

The content listed in bold appears in a tag hierarchy called `<detailed_description><textblock>`, but there is no way to scope our search via these tags as properties. To achieve this, you will need an XML parser. So, in essence, this “tune” is really overcoming a problem in SharePoint so that you might exert more control over how results sets are formed or how users might interact with search. In other words, you have a much more capable search engine if you can offer scoped searches on content like what we’ve seen here.

## Tune 7: Equality for Synonyms

For all of SharePoint 2016 and the September and October 2016 CUs of SharePoint 2013, Microsoft provided a way to set the weight of any synonym to 100%, just like its corresponding “parent” term. This is done with simple Powershell:

```
$ssa = Get-SPEnterpriseSearchServiceApplication <your SSA name>  
$ssa.ResetSynonymWeightToOne = $true
```



```
Select Administrator: SharePoint 2016 Management Shell  
PS C:\Users\Administrator> $ssa=Get-SPEnterpriseSearchServiceApplication  
PS C:\Users\Administrator> $ssa.ResetSynonymWeightToOne=$true
```

## Alternatives?

Fortunately, there are alternatives in the marketplace that provide easier and often more capable solutions to these same tuning problems. One of those is Fusion Server and Fusion AI from Lucidworks, which provides a scalable platform for building powerful search and big data applications. In the next installment, we will explore how Fusion can accomplish these tasks in better ways.

### Get Started or Learn More

For more information or to start using Fusion, contact us today at [lucidworks.com/contact](http://lucidworks.com/contact) or give us a call at **415-329-6515**.