



## Q&A on the Upcoming Text Search Changes to the Knowledge Mosaic SEC Filings page

### What's changing?

In October 2012, we're adding advanced capabilities to text searching on the SEC Filings page. The main addition is a new **relevance** component, which will supplement our existing Boolean search syntax. We'll continue to offer our full range of Boolean search capability – that is, the use of proximity operators; other relational operators like AND, OR, and, NOT; and the ability to organize or “nest” phrases using parentheses.

Our engineers have been working on the development of this new text search technology, called **Qmax™**, for some time. Although it will initially be limited to the SEC Filings page, we'll soon roll it out across the Knowledge Mosaic website.

### Why is Qmax good?

Here are three important things Qmax will bring to the table:

- 1) It will allow us to **unify our site with a single search syntax** across all search pages. Before the end of the year, users will only need to know one set of text search rules on Knowledge Mosaic.
- 2) It will allow us to introduce advanced Boolean search capability to relevance-based pages such as Laws, Rules, & Agencies Materials and SEC Comment Letters.
- 3) It will allow us to introduce to our SEC Filings page (and other pages) the power and efficiency of relevance-based search, in which documents that are a stronger match are boosted over those that are a weaker match. More on that below.

### How is it special?

Qmax is special because it is a **best-of-both-worlds approach to text searching**. Those two worlds are relevance-based searching and traditional Boolean searching. Since each of these approaches has distinct advantages, we wanted to find a way to incorporate both, thus offering the best search experience possible.

Essentially, Qmax works by recognizing whether or not a text search query is Boolean. If the user does not input any Boolean operators, the search acts much as it does on Google (or on our current Laws, Rules, & Agency Materials page). If Boolean operators are used, the search acts essentially as it does now on the SEC Filings page.

## **This sounds complicated. Will I have to learn even more operators? Will Qmax be harder to use?**

No. In fact, quite the opposite: the Knowledge Mosaic SEC Filings page will become simpler to use, because Boolean operators – though they will still be supported – will no longer be necessary to get good results. On the other hand, users who like the precision and control offered by Boolean searching will still have that option. There will be no new operators to learn; just the ones you already use. (There are, however, two minor changes to search syntax. See the “search syntax” question below.)

## **What if I don't want relevance boosting of my results? What if I'm old school and just want to live in the Boolean world?**

You can easily exclude relevance boosting from your list of results before (or after) running your search. You can still live entirely in the Boolean world if you want to.

## **How will relevance boosting work? What constitutes a “stronger match”?**

Relevance boosting on the SEC Filings page will be based primarily on these three criteria:

- 1) **Frequency or density of word/phrase matches.** The more times a word, phrase, or text string comes up in the document, the higher its relevance ranking.
- 2) **Phrase boosting.** If you enter only a string of words (no Boolean operators), the search engine will boost results based on adjacency (words right next to each other). For example, suppose you type in:

```
merger agreement
```

Documents that contain the exact phrase “merger agreement” will tend to rank higher than others. Documents that contain the phrase “agreement and plan of merger” will also be returned, but will tend to rank lower.

- 3) **Document date.** Relevance is also impacted moderately by the date of the document. All else being equal, more recent documents will rank higher than older documents. (This is the current behavior on our Laws, Rules, & Agency Materials search page).

## **How will search syntax on the SEC Filings page change?**

The most important point to be made is that *users will be able to do everything on the SEC Filings page they can do now.* In many if not most queries, no change in search grammar or syntax will be necessary to get exactly the same results. There are only two exceptions:

- 1) Quotation marks will now be necessary if you require a string of words to be treated as a unified phrase. For instance, in the “merger agreement” example above, you can specify an exact match of the phrase by using quotation marks. Otherwise, phrase boosting will kick in as described above.

2.) Quotation marks will now be necessary if you require a word to be an exact match as opposed to a “stemmed” variant of the word. “Stemming” means that grammatical variants of a word will be regarded as a match. For example, if you enter the word

acquire

without quotation marks, then documents including such variants as *acquired*, *acquiring*, and *acquisition* will also be returned. Quotation marks allow you to opt out of stemming.

### **Why is Knowledge Mosaic making this change?**

We’re motivated by the opportunity to bring two major benefits to the Knowledge Mosaic user. The first is to be able to offer a single search syntax across the entire Knowledge Mosaic site.

The second is to improve text searching on every page of the Knowledge Mosaic website. We believe in the value of relevance searching because it combines simplicity of use with extreme power. It is the future of search, and is quickly becoming industry standard.

On the other hand, we understand that Boolean searching gives users pinpoint precision and complete control over their queries. So Qmax allows us to offer both relevance and precision together – the best of both worlds.