

## **LawHelp Search Engine Upgrade Release Notes January 2010**

On January 22, 2010, a new search engine was implemented on LawHelp. The new search facility uses the SolR open source search platform. This is a natural language, full-text search that supports both simple word searches and phrase searches.

### **How does natural language searching work?**

The new search engine has natural language parsing built in. This means that users are able to enter their searches using common spoken or written language. There are no syntax rules or search operators to remember. For example, a search for "divorcing" will match resources containing the word "divorce." Similarly, a search for "what is elder abuse" will match resources containing those search terms as well as related words such as "abusive" and "elderly." Currently the search engine supports natural language searching in English and Spanish.

The search engine also contains an index of common synonyms to support natural language processing. For example, a search for "dissolution" will match resources containing the word "dissolution" as well as resources containing "divorce." The synonym list can be expanded over time.

The new search facility also has a suggested spelling feature for commonly misspelled words. For example, a search for "afadavit" will result in the message "Did you mean to search for *affidavit*?"

### **Which content items does the search engine include?**

The search engine looks at LawHelp resources, organizations and news articles. Expired news articles, inactive items or items in inactive channels are not included.

### **Which parts of each content item does the search engine examine?**

The search extends to all of the major content areas of the site.

When examining resources, the search engine includes the following fields:

- Resource Title
- Resource Author (organization and Other)
- Resource Description
- All text in resource attachments (including PDF documents, as long as they've been scanned "OCR," which converts image to text)
- If the resource is an FAQ, all text and link fields are also included.
- If the resource is a Short Profile, all description and name fields are included.

When examining news items, the search engine includes the following fields:

- Title
- Teaser

- Body Text
- Author/Source

When examining organizations, the search engine includes the following fields:

- Name
- Description
- City

### **How are search results displayed?**

Resources located by the search are displayed first, ranked in order of relevance. Relevance is determined by a number of factors, including in which field(s) and how often the keyword(s) appear. For example, if a keyword matches a word in the resource title, that item will rank higher than an item with the word matching in the resource body text only. Similarly, the more frequently a search term occurs in a document, the higher it will rank.

Resources have been manually “boosted” in the search results and will generally rank higher than News and Organizations. However, an Organization may be ordered above Resources and News items if it is a strong match.

### **What search filters are available?**

Several new search filters are available for users to tailor their search results:

- Ordering by Date orders the results by date created with the most recent at the top of the results

In the right navigation, users can filter search results according to:

- Content type (resource or organization)
- Channel /Category
- Topic / Subtopic
- Language

The number of matching items displays in parentheses next to the filter.

### **How often does the search index update?**

The index is updated shortly after any changes are made to Resources, News, or Organizations. This indexing process happens in the background and may take up to 5 minutes to be reflected in the search results.

### **How does the search engine take into account geographic limitations on resources and organizations?**

The search engine does not look at geographic coding at all. A search will return all results matching the keywords submitted, without regard to county, city, or zip code restrictions.

## **How are non-English resources searched and returned?**

Non-English resources are searched the same way English language resources are. However, if the search is conducted from the English site, the search engine returns the English resource as the primary result, with the alternate language version(s) linked below.

Each search on LawHelp is analyzed against all language versions of resources. This means that a search in English on a foreign language mirror site will match against all language versions of the resource. The results will display the resource in the site language if it is available. If the resource is not available in the site language, the English version will be displayed. Similarly, if Spanish text is searched on the English version of the site, the input will be analyzed and searched against the Spanish versions of resources, but the results will be displayed in English since that remains the site language.

**Note:** The search engine currently only supports searches in languages that use the Latin (aka Roman) alphabet. The search cannot match search terms in languages that use different alphabets. For example, if you enter a search using Russian characters, the search cannot match Russian resources containing that text. For more information on what languages are supported, please email [support@probono.net](mailto:support@probono.net).

## **Where are the admin tools associated with the search engine?**

The following admin pages contain fields associated with the search engine. (Note: to edit these fields on bilingual mirror sites, use the View Language drop-down menu.)

Add Content -> Search Tips

- You can make changes to the Search tips page on this admin page. PBN will circulate sample text for this page soon.

Page Setup -> Search Labels

- On this page you can control all of the various labels associated with the search engine.

Page Setup -> Navigation

- At the bottom of this page you can control turning the search engine on or off as well as its name on the site.

## **How can I tell what search terms users are entering?**

Summary usage reports are currently being modified to include lists of the top search terms and phrases. When this development work is done, admins will be able to view a list of the top 50 search terms, and a separate list of the top 50 search phrases, in summary usage reports.