

Searching words that may be hyphenated in OPAC

- **Article Type:** General
 - **Product:** Aleph
 - **Product Version:** 18.01
-

Description:

When we search by title words for "19th-century us newspapers" we expect to retrieve our record 001149107. However, the record itself does not contain the string 19th-century; it instead contains "19th century" with no hyphen. It is obviously not indexed with a hyphen (nor should it be), and thus does not show up in the search results.

Is there a way to configure the OPAC to strip a hyphen from a search string (e.g. if we submitted the search 19th-century it would pass it to the server as 19th century, substituting a space for the hyphen)?

Resolution:

The "tab_word_breaking" table, section 90, is consulted for parsing the incoming OPAC data string. You can add a hyphen (and maybe an apostrophe) to the "to_blank" procedure's parameters for the 90 stanza.

The table header includes dire warnings about adding a hyphen with to_blank, as processing of hyphens and apostrophes is automatic in indexing. It explains the idea of "triple-posting", where hyphenated words are automatically indexed with the hyphen (e.g. 19th-century), with a blank instead (e.g. 19th century) and with the words run together (e.g. 19thcentury). If you add a hyphen to the to_blank parameters for indexing, it could mess up this processing.

But "triple-posting" does not happen with the incoming search string, so it would not cause the same problem if you add it to the 90 stanza. Words that are actually hyphenated in the database should still be retrieved, since search terms that get the hyphen stripped away should match one of the "triple-posted" index strings whether the incoming search string includes a hyphen or not. But those that do not include the hyphen in the indexed term should also be retrieved this way, although they will not be retrieved if you do not include a hyphen in the to_blank routine.

- **Article last edited:** 10/8/2013