TextCritical.net - Feature #539

Find similar forms when expanding word into related forms

02/24/2013 01:02 AM - Anonymous

Status:	Closed		**Start date:**	02/24/2013
Priority:	Normal		**Due date:**	
Assignee:	Luke Murphey		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	0.6			

Description

The search ought to find variations ignoring diacritical marks if a form could not be found. This useful because some searches fail to match for some reason even though they seem like thy ought to match.

For example, Josephi Vita 22:110 includes the word πιστὸς yet a search for "work:josephi-vita pisto)s" fails. A search for "work:josephi-vita no_diacritics:pistos" succeeds. Note that a search for "work:josephi-vita pisto\s" also fails.

History

#1 - 02/24/2013 07:52 PM - Luke Murphey

It looks like the SimpleAnalyzer is not allowing the parentheses through.

#2 - 02/24/2013 08:24 PM - Luke Murphey

It looks like the following *are not* considered part of the word (per the Simple Analyzer): ( ) \ + &

The following *are* considered part of the word: / = |

#3 - 02/24/2013 08:28 PM - Luke Murphey

For some reason, a request for variations is repeated 48 times. Not sure why this is happening.

#4 - 02/24/2013 08:46 PM - Luke Murphey

It looks like multiple calls are being made to GreekVariations._words. The constructor of GreekVariations is called twice. I put in some code to cache the results. I would rather figure out why the variations class is getting called more than once but from what I can tell this is something in the Whoosh classes that I cannot change. I added a caching system for the results and this works well.

#5 - 02/24/2013 09:04 PM - Luke Murphey

I fixed the performance problem caused by multiple calls in [#540](#540).

#6 - 02/25/2013 07:47 AM - Luke Murphey

- % Done changed from 0 to 20

#7 - 02/25/2013 08:15 AM - Luke Murphey

I may be able reduce the number of database calls by not getting the verse information in the search results if the results are not for the page to be shown. From what I can tell, all verses are being loaded even if only a subset are being shown on the results page.

#8 - 02/25/2013 08:17 AM - Luke Murphey

- Assignee changed from Anonymous to Luke Murphey

#9 - 03/02/2013 06:16 AM - Luke Murphey

- Status changed from New to In Progress

#10 - 03/02/2013 05:21 PM - Luke Murphey

Rebuilding the indexes with SimpleAnalyzer allowed slashes to work provided they were escaped.

#11 - 03/02/2013 05:48 PM - Luke Murphey

All of the characters seem to come through now with the exception: parentheses, asterisks, amperstands. I believe these are being filtered because whoosh uses them as special characters for the the purpose of searching.

#12 - 03/02/2013 05:48 PM - Luke Murphey

- % Done changed from 20 to 50

#13 - 03/04/2013 05:27 AM - Luke Murphey

- Status changed from In Progress to Closed

- % Done changed from 50 to 100