

Mulgara - Bug #191

Search is slow when Lucene writes to disk

08/13/2009 10:22 PM - Paula Gearon

Status:	New	Start date:	
Priority:	High	Due date:	
Assignee:	Paula Gearon	% Done:	0%
Category:	Mulgara	Estimated time:	0.00 hour
Target version:			
Resolution:			
Description			
<p>Often, if a search is exceptionally slow, much of that time is wasted when Lucene decides that its results need to be sorted. If the quantity of data to be sorted does not fit into memory, then Lucene swaps it out to disk, radically slowing what would otherwise be a fast set of comparisons.</p> <p>Most of this sorting is useless and can occur anywhere in the search process (e.g., when combining two datasets in the midst of a larger query, or sorting the entire result set before returning it).</p> <p>Tuning Lucene (telling it to sort only when necessary) and/or changing the OQL used in searches may vastly improve response times in some cases.</p> <p>Example: on a full corpus, enable date searching (currently disabled in SearchAction.java lines 155 to 164 due to excessive query slowness) and search PLoS One for:</p> <p>gene AND date:[2008-05-16 TO 2009-01-26]</p>			