DataSet generation

1. Each Systematic review has a unique ID in the format CD000254.pub2 and has a webpage associated with it

Example :

Review : Fisher-2014-Stem cell therapy fo.pdf

ID of review : CD007888.pub2

Web page link : http://onlinelibrary.wiley.com/wol1/doi/10.1002/14651858.CD007888.pub2/references

2. All such IDs of systematic reviews are obtained using grep tool by regular expression “CD[0-9]{6}.pub[0-9]” from text files of systematic review PDFs

3. Using all these IDs each systematic review web page is crawled and all reference blocks of each page are extracted.

Sample reference Block:

<div class="bibSection" id="CD011273-bbs2-0011">

<div class="headingCont"><h5>Balshem 2011</h5></div>

<ul class="plain"><li id="CD011273-bib-0011"> <cite id="cit11"> <span class="author">Balshem H</span>, <span class="author">Helfand M</span>, <span class="author">Schunemann HJ</span>, <span class="author">Oxman AD</span>, <span class="author">Kunz R</span>, <span class="author">Brozek J</span>, et al. <span class="articleTitle">GRADE guidelines: 3. Rating the quality of evidence</span>. <span class="journalTitle">Journal of Clinical Epidemiology</span> <span class="pubYear">2011</span>;<span class="vol">64</span>:<span class="pageFirst">401</span>-<span class="pageLast">6</span>. </cite><ul class="externalReferences">

<li>

<a href="/resolve/reference/XREF?id=10.1016/j.jclinepi.2010.07.015" shape="rect">CrossRef</a>,</li>

<li>

<a href="/resolve/reference/PMED?id=21208779" shape="rect">PubMed</a>,</li>

<li><a href="/resolve/reference/ISI?id=000288364300009" shape="rect">Web of Science&#174; Times Cited: 395</a></li>

</ul> </li></ul>

</div>

4. Titles of all the reference blocks are extracted

Example :

‘ <span class="articleTitle">GRADE guidelines: 3. Rating the quality of evidence</span>

5. List of these titles are obtained and are provided as input to getPM.java code which queries using eutils search api

6. Each query returns a list of PMIDs

7. All the results which have less than 5 PMIDs are taken

8. Using all these PMIDs a query string is generated of the form “"7474675[uid] OR 1739528[uid] OR 22999730[uid]”

9. This string is entered in the search field of this page manually <http://www.ncbi.nlm.nih.gov/pubmed/>

10. The results are exported using “send to”->”citation manager”. A citation.bib file is created by this