Frequently asked questions (FAQ)

1. The instruction of hw4 suggests that we should install ubuntu on windows so as to install solr more conveniently. However. if I do not want to install ubuntu, may I install solr and other software needed directly to windows system?

   Answer –

   You can. However, if you run into some issues when installing directly on Windows, we may not be able to guide you. You will be on your own.

2. Are we allowed using any language and platform instead of PHP for the Web pages?

   Answer –

   You can use node.js, however no help will be provided for the code implementation for this homework.

3. There is no step-wise instruction given for installing solr. Can you guide me in installation process? I downloaded Solr-7.1.0 zip.
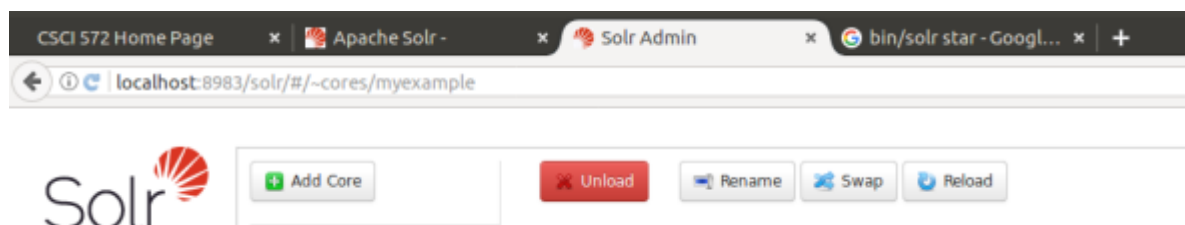
   Answer –

   You need to unzip it and then follow the instructions in "Using Solr to Index a Web Site" document.

   Once you are done with the Ubuntu installation steps, the Solr setup instructions are to be executed inside Ubuntu. So, setting up Java 8, downloading, unzipping Solr, Apache, PHP etc. all should be done inside Ubuntu.

4. How to reload our core and query with new configuration.

   Answer –

   There will be a reload button in the Solr instance's core admin:

   I suggest exploring the Solr instance, it's various configurations and settings.

5. Do I need to install an Apache server on our computer in addition to running the Solr server?

    Answer –

    Apache server will be installed by default in Linux machines. in the assignment we are using PHP script to communicate to the Solr. As we know PHP is a server-side scripting language and it requires Apache or any other web servers.

    However, if you run into any problem then use the below commands on Ubuntu terminal.

    To install Apache on Ubuntu

    sudo apt-get install apache2

    To install PHP on Ubuntu

    sudo apt-get install libapache2-mod-php

6. Can I run solr in cloud mode?

    Answer –

    You don't need cloud mode for this assignment. Only use "bin/solr start" to start running the Solr instance in stand-alone

    To Restart Solr-

    First do, "bin/solr stop -all". Then "bin/solr start"

7. I am not mentioning my top 10 results in the report. I am maintaining separate text files for each query. Is that allowed or shall I mention the results only in the report?

    Answer-

    It would make grading much easier if you included all in one file.

8. I am getting zero overlaps between Default and pagerank algorithms. Could you please let us know how many points would be deducted for zero overlaps?
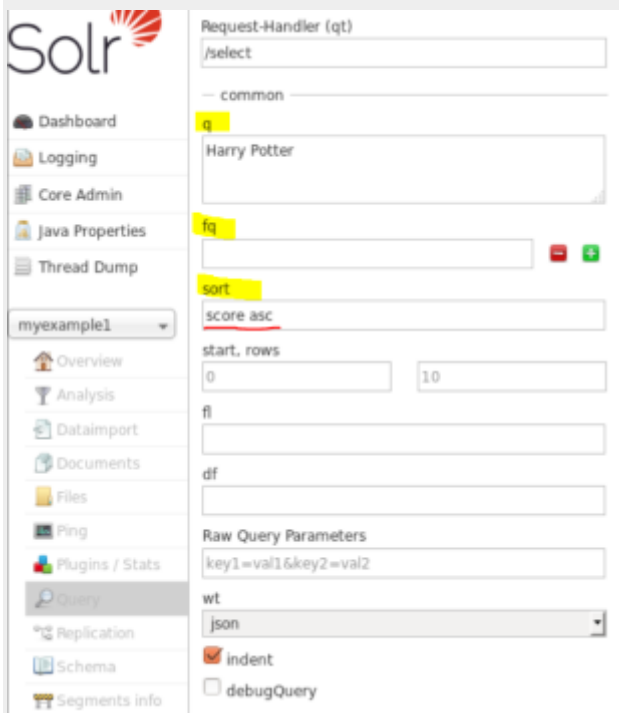
    Answer –

    As long as you follow the assignment steps and implement it correctly, you won't be penalized. Just make sure your steps are clear in your report and you should be fine.

9. How to use additional parameters from code?

   Answer –

Here is an example of using Additional Parameters:

Please notice the below parameters when querying Solr from the dashboard: q, fq, sort, etc.,

When querying Solr from the dashboard, you enter the query in the q parameter input box-> "Harry Potter".
You may use the other parameters as well to further refine the query.
Ex: sort parameter's default value when nothing is specified is "score desc", i.e. the results are sorted and displayed from highest score to lowest score. You can change to "score asc".(For demo purposes)

How do you specify this in the PHP code? By sending it as an additional parameter in the call for Solr:
$additionalParameters = array(
 'sort' => 'score asc'
);
$results = $solr->search($query, 0, $limit, **$additionalParameters**);

URL   ->

http://localhost:8983/solr/myexample1/select?indent=on&q=Harry Potter&sort=score asc&wt=json

Hope this clears your doubts on how to use additional parameters from code, and why you were getting the error.
Further reading -> https://wiki.apache.org/solr/CommonQueryParameters

10. Do we need sort results by any attribute in the first algorithm?

    Answer - No. It is already sorted by default value -> "score desc".

11. What is the use of Java code?

    Answer - The Java code creates the edgelist.txt file.
            You must use this as input for Python code, which creates a graph using NetworkX library.

12. Do we have to write page rank algorithm?

    Answer –

    The default search result is returned based on Lucene's ranking algorithm.

    The next step is to get the results based on the page ranking score. You do not need to write
    the algorithm yourself. It is implemented in NetworkX library. You use it to obtain an external file
    with page ranking scorings, which you add to Solr. Please see the document.

13. If that page doesn't contain description information, we don't need to show description?

    Answer –
        That's ok. You can show Null or N/A.

14. Getting much better results with 'pageRankFile asc' as opposed to 'pageRankFile desc'. Which
    one shall I use?

    Answer –

        Solr uses Boolean model to narrow down the range of the documents that needs to be
        scored. This will be done based on the query term. Page rank "desc" will next sort and
        return these from pages from highest score to lowest score.
        The pages having highest page rank scores (maybe a homepage, or some other popular
        page with many links to it) may not be most relevant to query term.