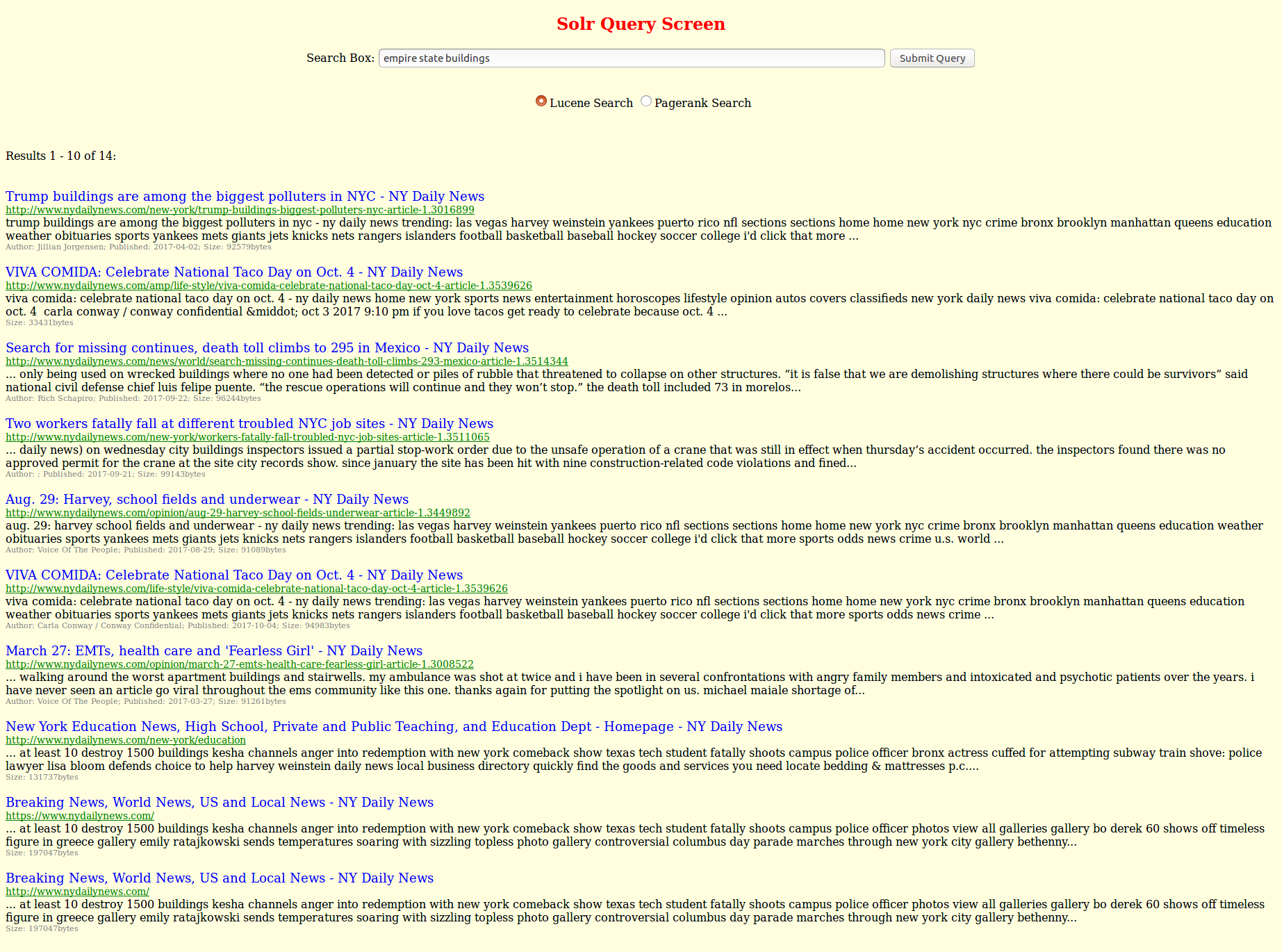
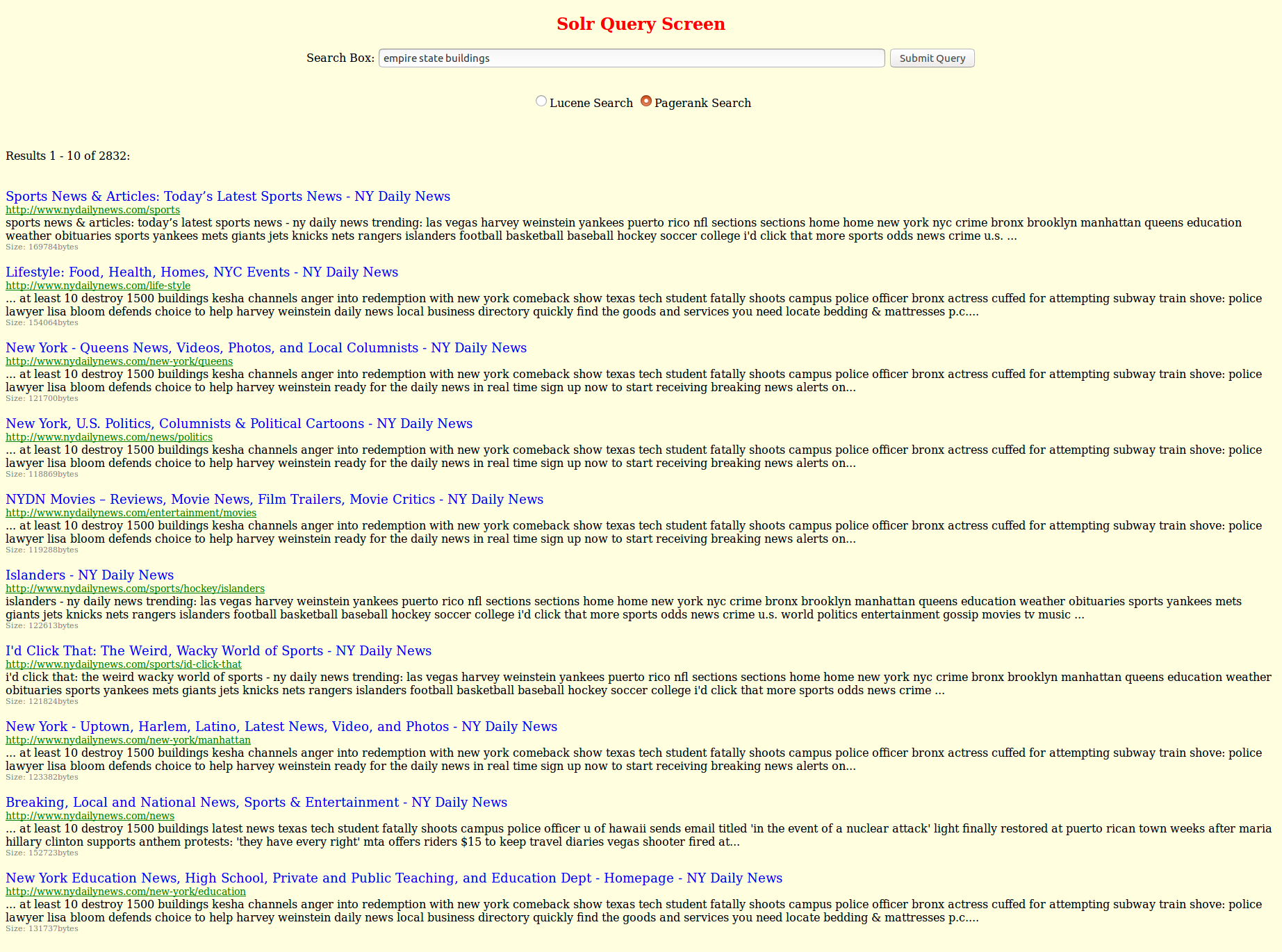
1. **Steps Taken to Complete the assignment:**  Enumerating the steps for the added functionality after assignment-4
2. Created a PHP page called solrquery.php which run on top of Apache Server. This php page uses Apache-solr-php client to access solr and based on query type i.e either Lucene based or Page Rank based, returns results of the query showing 10 at a time on the page.
3. For **Spellcheck** Download Norvigs spellchecker from his website. Include his 'SpellCorrector.php' which lets you use the correct() function.
4. Use Apache TIKA to parse through all the HTML file to obtain content and clean this to obtain only valid strings ( for example alphanumeric strings are valid but URLs and strings with special characters are not ) and populate them in big.txt.
5. Added common words like 'Politics' and 'sports' and all the queries from assignment 4 such as l to 'big.txt' to increase their probability in the spell corrector program.
6. In PHP, pass each term of the query to correct() to obtain the most likely correct spelling. Store both this and the old query with incorrect query terms. Query solr using the new correct query and display results for the new correct query but ask the user whether she/he wants the results for the incorrect query. If so, then pass this query to solr instead and display the results if any.
7. Created a PHP program called generatesnippet.php which use ‘simple\_html\_dom.php’ module to get the content of a webpage. Saved the content as array of words and searched the array for the query terms. If query terms are found in order or only 1 query term is found it selects the index from the array and create a snippet of length 60 words.
8. For **Autocomplete** made the requisite changes in 'solrconfig.xml' by adding a search component which uses the Fuzzy Lookup algorithm on the text being entered and obtain suggestions using '\_text\_ which is defined in managed-schema.xml.
9. Added a request Handler on the suggest component which displays as many suggestions as defined in 'suggest.count' while making use of 'suggest.dictionary'.
10. Condition an AJAX call to the solr server on the change in value of search box. The call is made to 'localhost:8983/solr/csci572/suggest' followed by the value in search box as well as a parameter requesting the result in JSON format.
11. On successful return of the AJAX call, parse the JSON and populate the suggestions in a drop-down list box
12. **Results of the page:**
13. Lucene Query Result

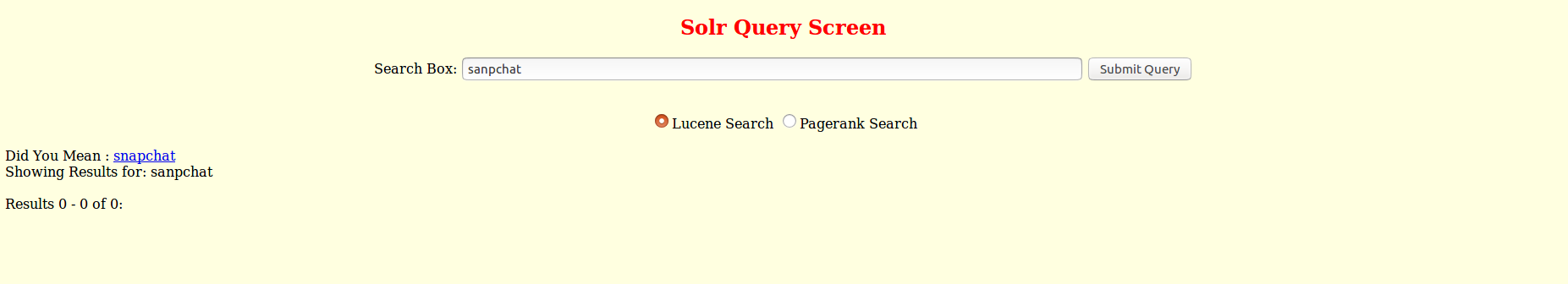


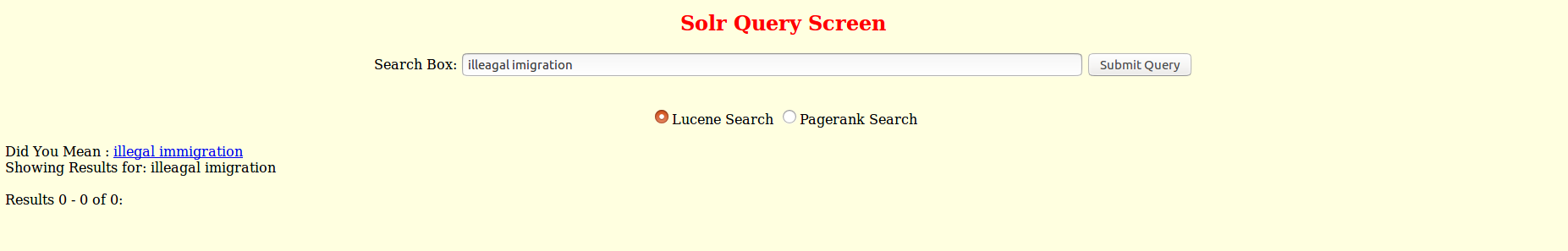
1. PageRank result of the query

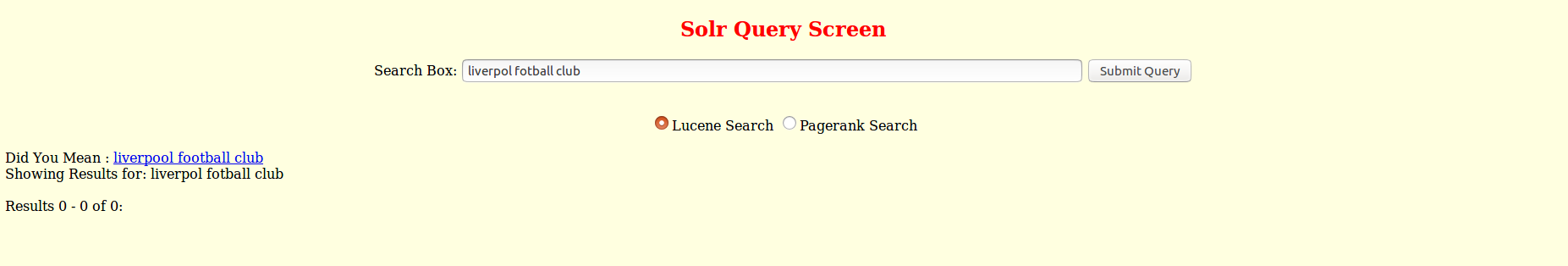


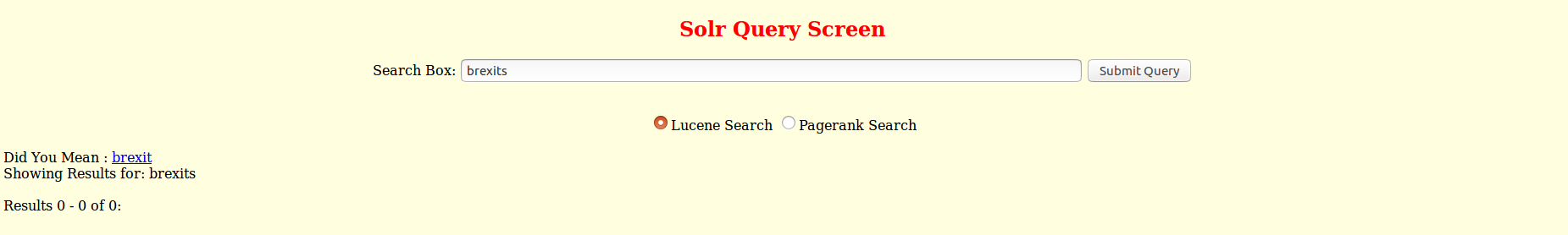
1. **Results of the 5 Spelling Corrections**

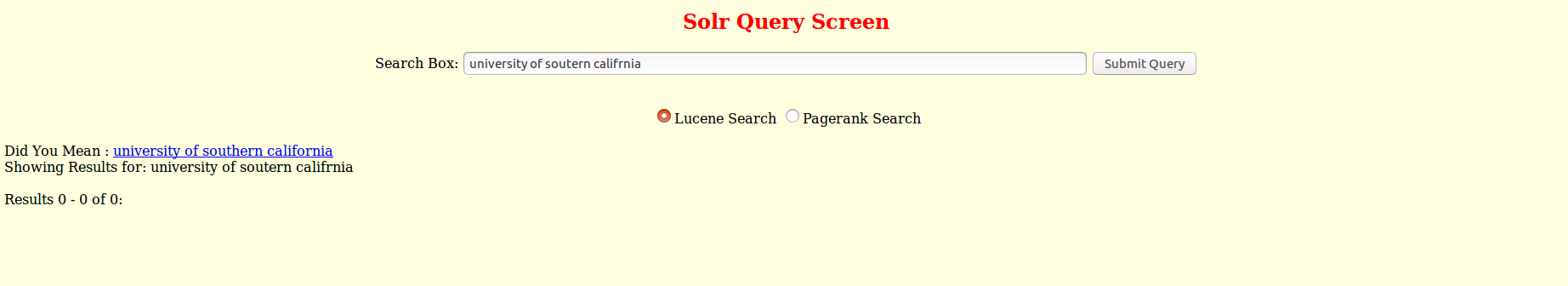
|  |  |
| --- | --- |
| **Query Entered** | **Correct Query** |
| Snpachat | snapchat |
| Illeagal imigration | Illegal immigration |
| Liverpol football club | Liverpool football club |
| brexits | brexit |
| University of soutern califrnia | University of southern california |











1. **Results of the 5 AutoComplete:**

|  |  |
| --- | --- |
| **Input** | **Expected Suggestion** |
| univ | university |
| Po | policy |
| ente | entertainment |
| fac | Facebook |
| Gymn | gymnastics |

