In this assignment, we have three main parts need to implement, that is spelling correction, autocompletion for search queries and snippet for search results.

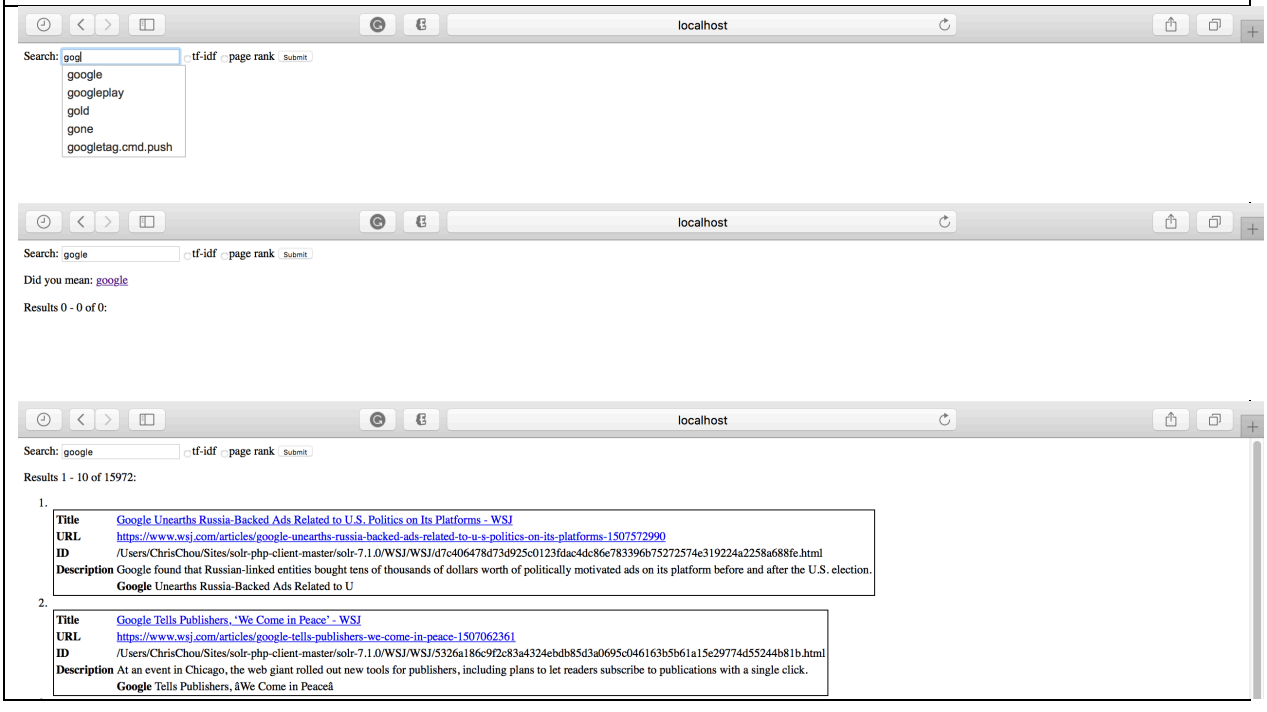
In the first part, I use Apache Tika to parse the Wall Street Journal webpages, Apache Tika is a library used for content type detection and data extraction from various file format. With Tika, we can easily extract all the contents and metadata from the news for which webpages we are responsible into a text file, called big.txt. After generating this file, I use Peter Norvig's spelling corrector with PHP version to implement the correction task. The previous big.txt file will be fed to the spelling corrector as the lexicon for this job. It uses insertion, deletion, transposition and substitution to compute all word within two edit distance and return the word with the most probability. Therefore, when we input a word that is not in our lexicon, the spelling corrector will return a word that has the most probable word from the entered word to us.

In the second part, I use Solr's built-in SuggestCompnent to implement the autocompetion. First, configure our solrconfig.xml with the search component and the corresponding the request handler for using SuggestComonent follow by the instructions on the document provided. Then, we can make use of the suggest request handler to get the suggested terms from Solr. I use AJAX to request the suggested term and dynamically update the suggested terms from the search box in our PHP file.

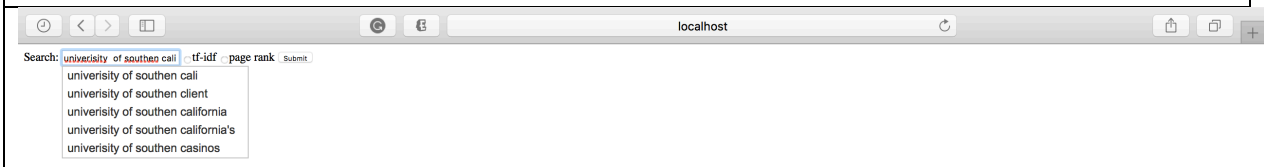
In the third part, I use the metadata description of the result webpage and check with query to see if there exist a match. If there is one, then print out the metadata as the snippet.

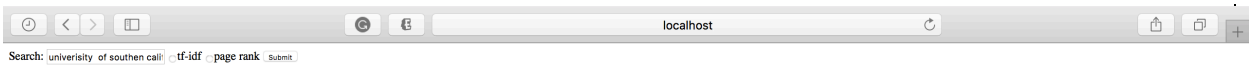
Five examples for autocompletion, correction and the results after clicking on the suggestion:

1



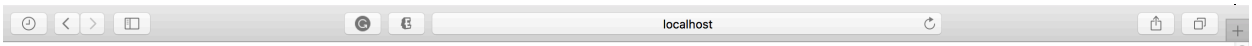
2





Did you mean: university of southern california

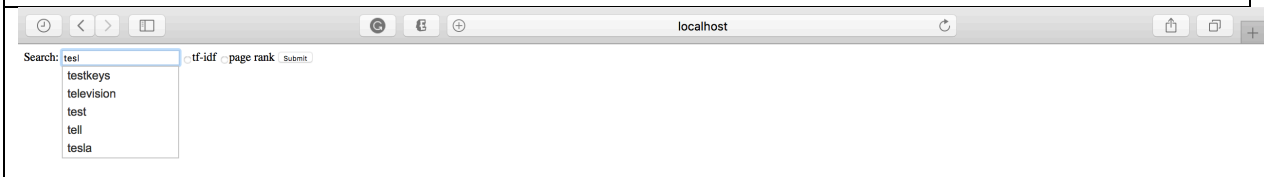
Results 0 - 0 of 0:

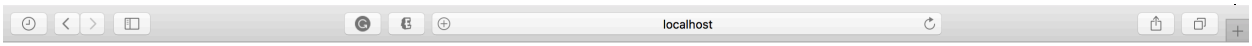


Results 1 - 10 of 263:

1. **Title** NFL Hall of Famer Lynn Swann Buys in Los Angeles - WSJ
   URL https://www.wsj.com/articles/nfl-hall-of-famer-lynn-swann-buys-in-los-angeles-1493304865
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/9aee394a3f19b1592bda0076d7cf34e534cc662380f6194edacffb0faed4e175.html
   Description The home is about 6 miles from the University of Southern California, where Mr. Swann, 65, was named athletic director last year.
2. **Title** News Article Archive from June 05, 2017 - Wsj.com
   URL http://www.wsj.com/public/page/archive-2017-6-05.html
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/a29e56c9e17ad055c309492c7f17ed14e7e5705fbfd44412c0be386e9977b033.html
   Description Archive of top headlines and current news articles.

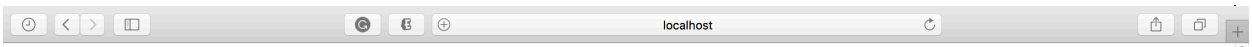
3





Did you mean: tesla

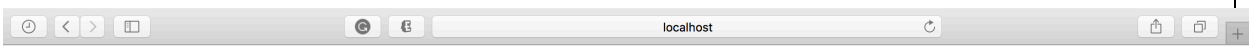
Results 0 - 0 of 0:



Results 1 - 10 of 1699:

1. **Title** The Truth Is Catching Up With Tesla - WSJ
   URL https://www.wsj.com/articles/the-truth-is-catching-up-with-tesla-1507399374
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/b4acf636bf5da99e5bc15d912e1ad4b3fa562f56290d745de6260ab785da5cb5.html
   Description Tesla CEO Elon Musk is a visionary, which has endeared him to Wall Street analysts and investors alike, but there is a fine line between setting aggressive goals and misleading shareholders. The Truth Is Catching Up With **Tesla**
2. **Title** Tesla Readies for Model 3 Launch by Adding Hundreds of Repair Vans - WSJ
   URL https://www.wsj.com/articles/tesla-readies-for-model-3-launch-by-adding-hundreds-of-repair-vans-1499778004
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/e67f47d2fd6d40be7a5e95429ab37218bcf8eae5cd1810dcd991c69e92a19841.html
   Description As Tesla begins a launch of its first mass-market car, the company said it plans to triple its capacity to repair vehicles, adding 1,400 technicians, dozens of new service centers and hundreds of maintenance vans. com","nofollow":"false"}}}],"isLoggedIn":false,"region":"na,lat","section":"Tech","twitterText":"**Tesla** readies for Model 3 with plan to add hundreds of repair vans, 1,400 technicians","shareUrl":"https://www

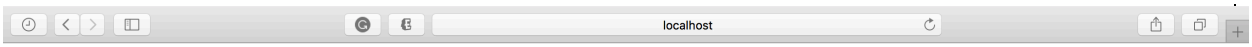
4



Search: tawan tf-idf page rank Submit

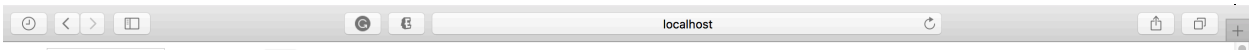
Did you mean: taiwan

Results 0 - 0 of 0:



Search: tawa tf-idf page rank Submit

tamaño
toward
taxation
taiwan
tata

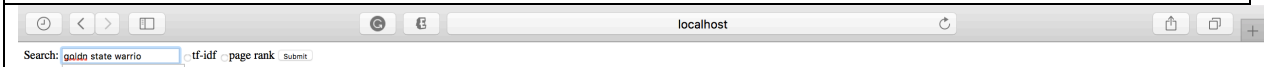


Search: taiwan tf-idf page rank Submit

Results 1 - 10 of 281:

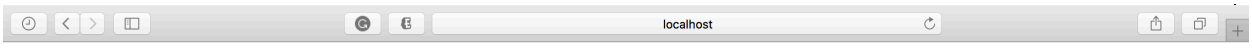
1. **Title** Taiwan Needs Submarines - WSJ
   URL https://www.wsj.com/articles/taiwan-needs-submarines-1491954190
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/5243c4883e9e692cf81400868955bfcdc8510cbcd3fe5148e36ee1784a82e792.html
   Description As China increases its threats, the U.S. can help the island's self-defense.
   com","nofollow":"false"}]},"isLoggedIn":false,"region":"na,lat","section":"Opinion","twitterText":"**Taiwan** needs submarines","shareUrl":"https://www
2. **Title** President Tsai: Taiwan Will Not Succumb to China
   URL http://www.wsj.com/video/president-tsai-taiwan-will-not-succumb-to-china/C3868C8E-EE14-4E73-B2CE-2FC74F36CD7F.html
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/9f0adf1996d3fd5bc4837c21f6cc57be18c3ce9600e00e1a41357ad29e8446e3.html
   Description Taiwan's first woman president, Tsai Ing-wen, in an exclusive interview with The Wall Street Journal, discusses the island's fraught relations with China, tensions over territorial disputes, the President Tsai: **Taiwan** Will Not Succumb to China

5



Search: goldn state warrio tf-idf page rank Submit

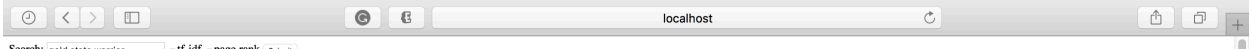
goldn state warriors
goldn state warrior
goldn state warring
goldn state warrior's
goldn state warrick



Search: goldn state warrior tf-idf page rank Submit

Did you mean: gold state warrior

Results 0 - 0 of 0:



Search: gold state warrior tf-idf page rank Submit

Results 1 - 10 of 65:

1. **Title** Despite Setbacks, Trump's Trade Warrior Peter Navarro Is Fighting On - WSJ
   URL https://www.wsj.com/articles/despite-setbacks-trumps-trade-warrior-peter-navarro-is-fighting-on-1494275645
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/9427937043b763ba768e88b6ab282d1e967821368 62f563218c695a9e100b4a5.html
   Description The White House's most hawkish trade adviser, Peter Navarro, says the administration is still pushing to win concessions from trading partners even though the president has softened some positions.
2. **Title** News Article Archive from January 11, 2014 - Wsj.com
   URL http://www.wsj.com/public/page/archive-2014-1-11.html
   ID /Users/ChrisChou/Sites/solr-php-client-master/solr-7.1.0/WSJ/WSJ/1ad67ee0486f97ed0ca7abbf0eadfd1ac9b524f411efd915c08c86a97ed89af1.html
   Description Archive of top headlines and current news articles.