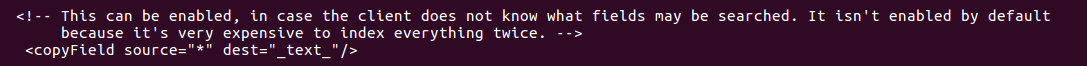
Name: Chunyu Peng

USC ID: 3745895256

CSCI 572 HW4 Report

I first downloaded Ubuntu and VirtualBox. After setting them up, I downloaded Solr 7.7.3 and unzipped it. I uncommented the following lines:



Text

Description automatically generated

Then I downloaded html files from foxnews and created my core named myexample.

Next, I went to the Solr dashboard to reload the settings.

After putting the html files into the shared folder, I started to index these files.





This is the dashboard of Solr with my core.

Graphical user interface, text, application, email

Description automatically generated

After the files are indexed, I tried to execute the default query.

Graphical user interface, text, application

Description automatically generated

Next, I executed another query with the word french.

Graphical user interface, text, application, email

Description automatically generated

Furthermore, I wrote the java code using Eclipse with JSoup library to extract links. Then the code generates list of edges in the EdgeList.txt file.Text

Description automatically generated

In addition, I wrote the python code using PyCharm with networkx library to calculate the page rank. The alpha value I used is 0.85, personalization is None, max\_iter is 30, tol is 1\*10^-6, nstart is None, weight is ‘weight’, dangling is None. As a result, I get the external\_pageRankFile.txt file.

I put external\_pageRankFile.txt into data folder in my core folder.



I then go to solrconfig.xml file to add listener.

Text

Description automatically generated

And go to managed-schema file to add the following to lines:





Now we go back to the dashboard and reload the settings.

Then add an additional parameter in sort: pageRankFile.txt desc. This will sort the result of pageRank in descending order.

Graphical user interface, text, application

Description automatically generated

I also tried it with the word french.

Graphical user interface, text, application

Description automatically generated

Next I installed apache and solr-php by git clone <https://github.com/PTCInc/solr-php-client.git>. Then I created a index.php file inside the solr-php-client folder. I used the template php provided in the instruction. And used a drop down list for user to choose between the two algorithms.

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

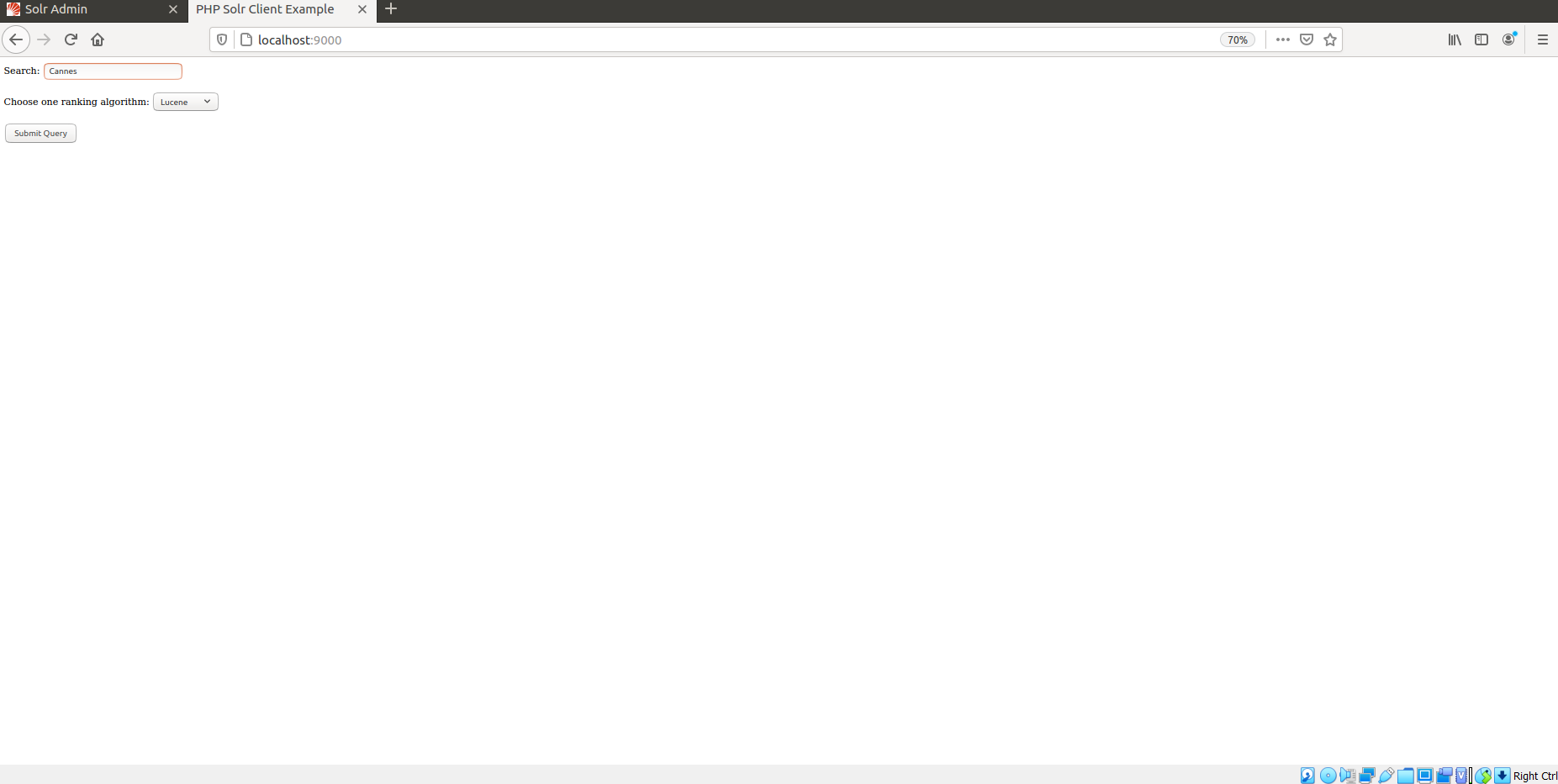
Why some pages have higher page rank?

Because these pages have larger number of pages linked to themselves. If there are larger amount of pages linked to one page. This page is considered higher ranking. Moreover, if a page is linked to a page with high rank, this page will also have higher rank.

Now I need to put different words into the query.

Cannes

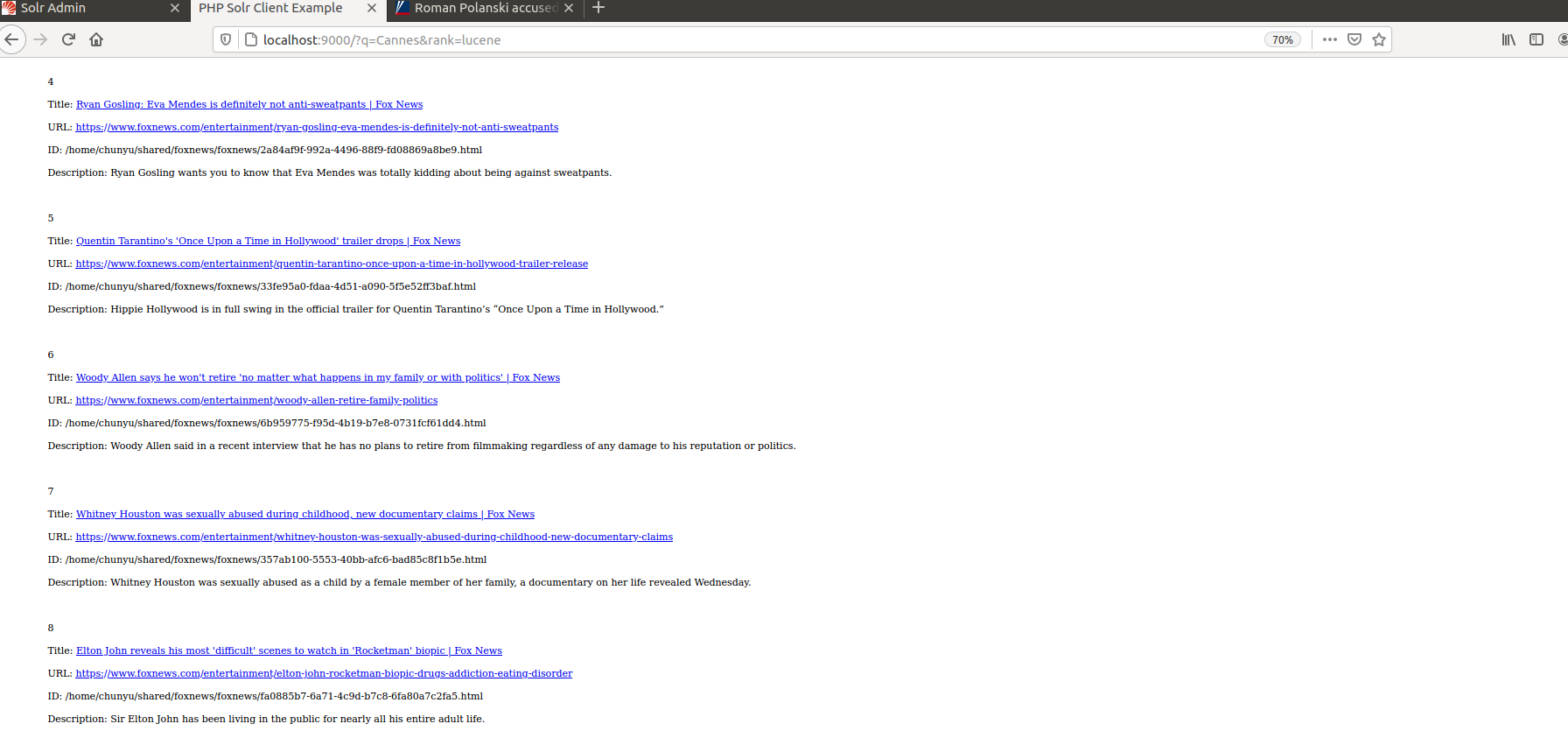
Initial page when enter query

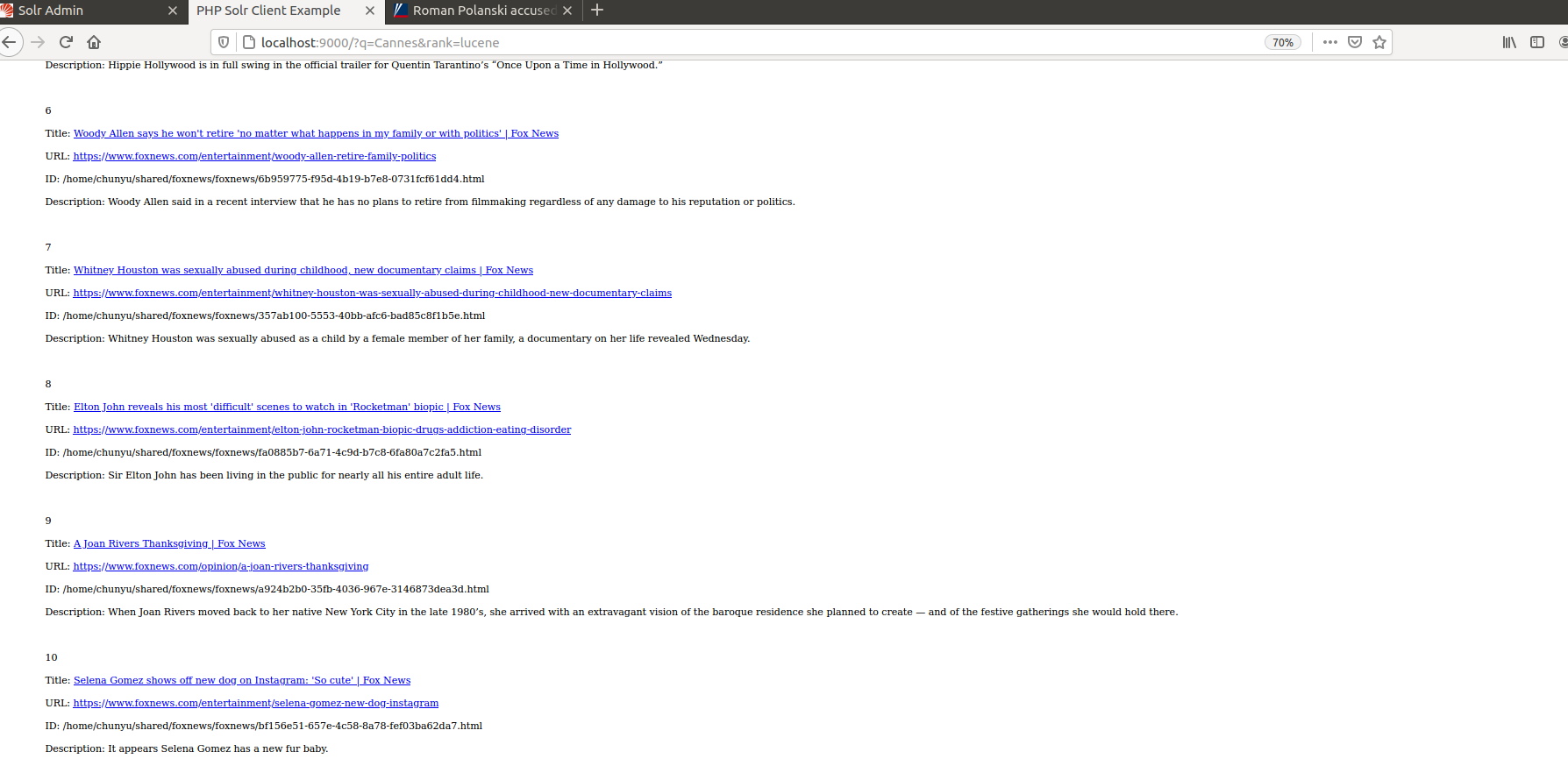


Lucene

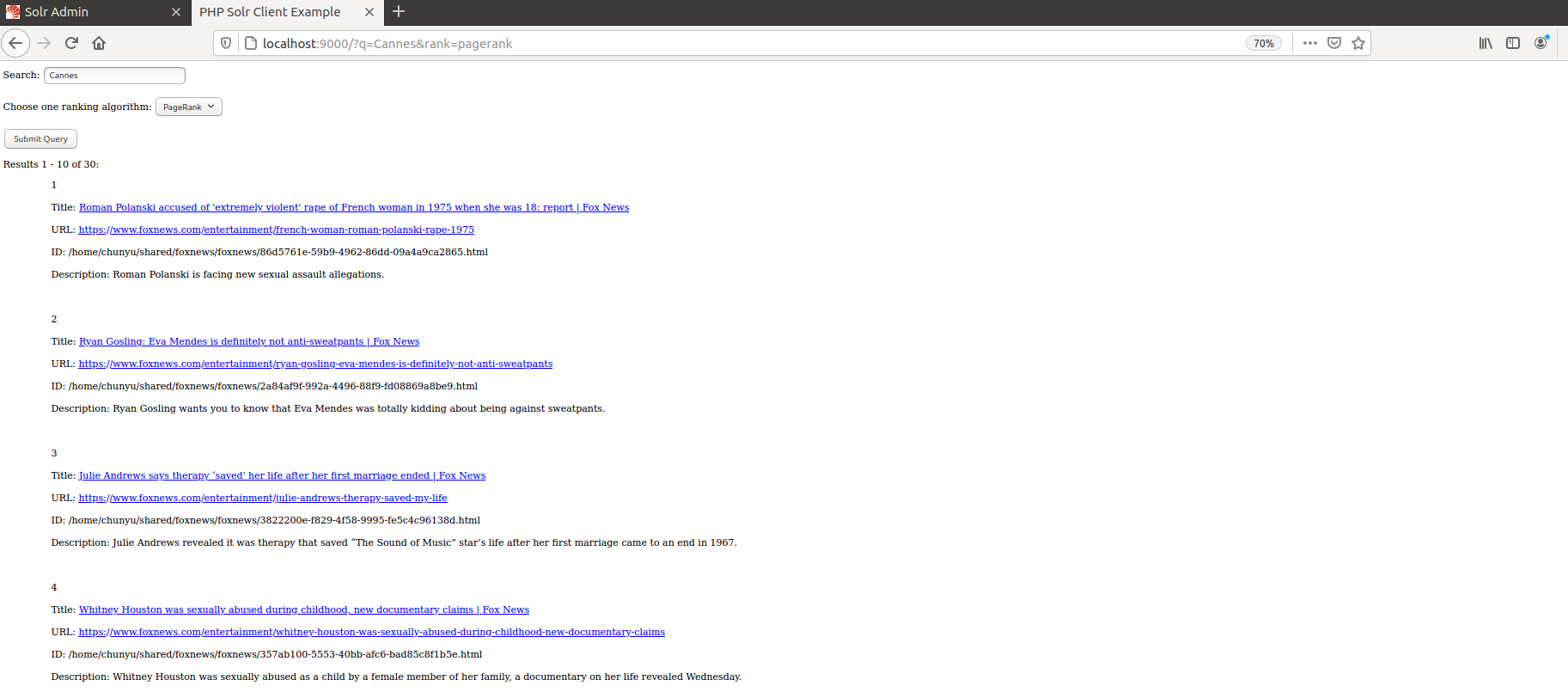
Graphical user interface, text, application, Teams

Description automatically generated





PageRank



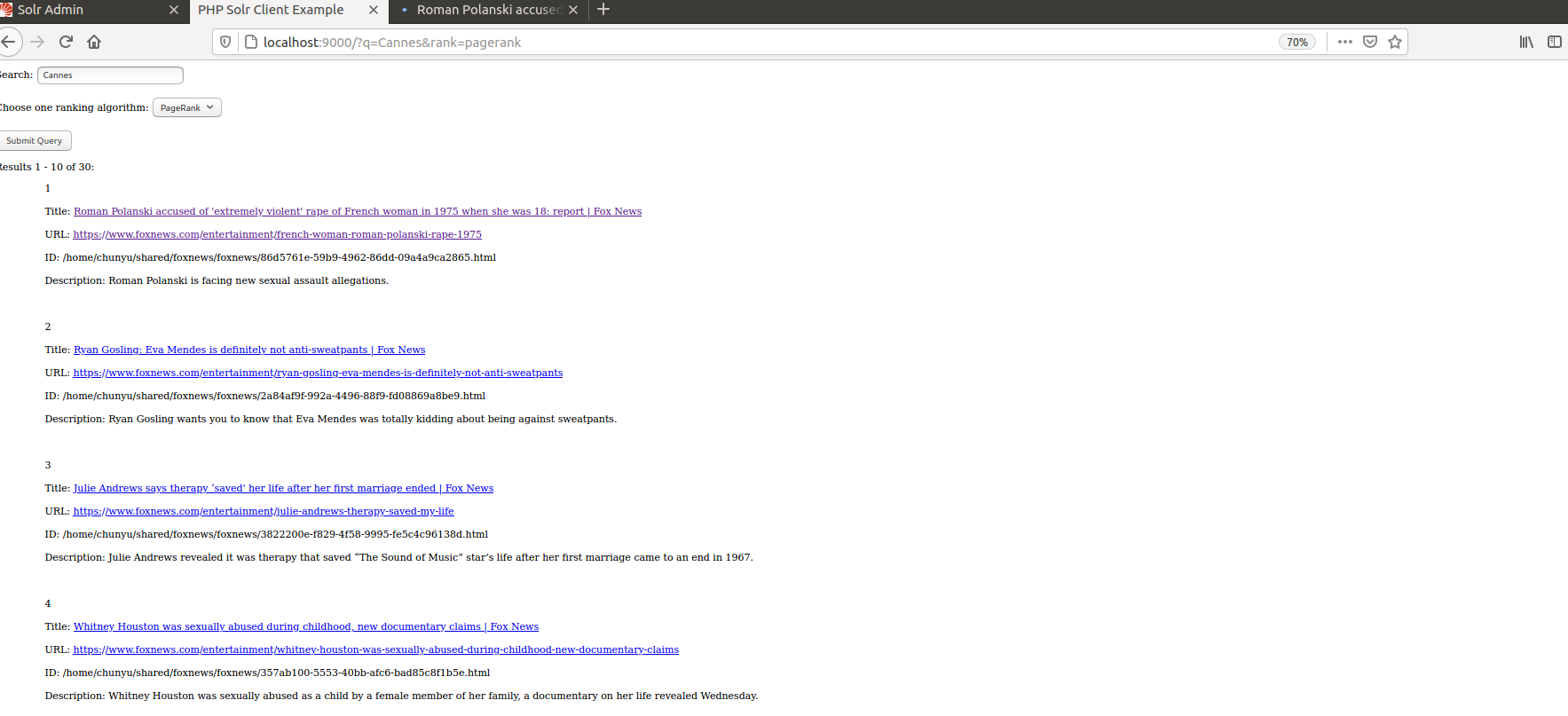
Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Click



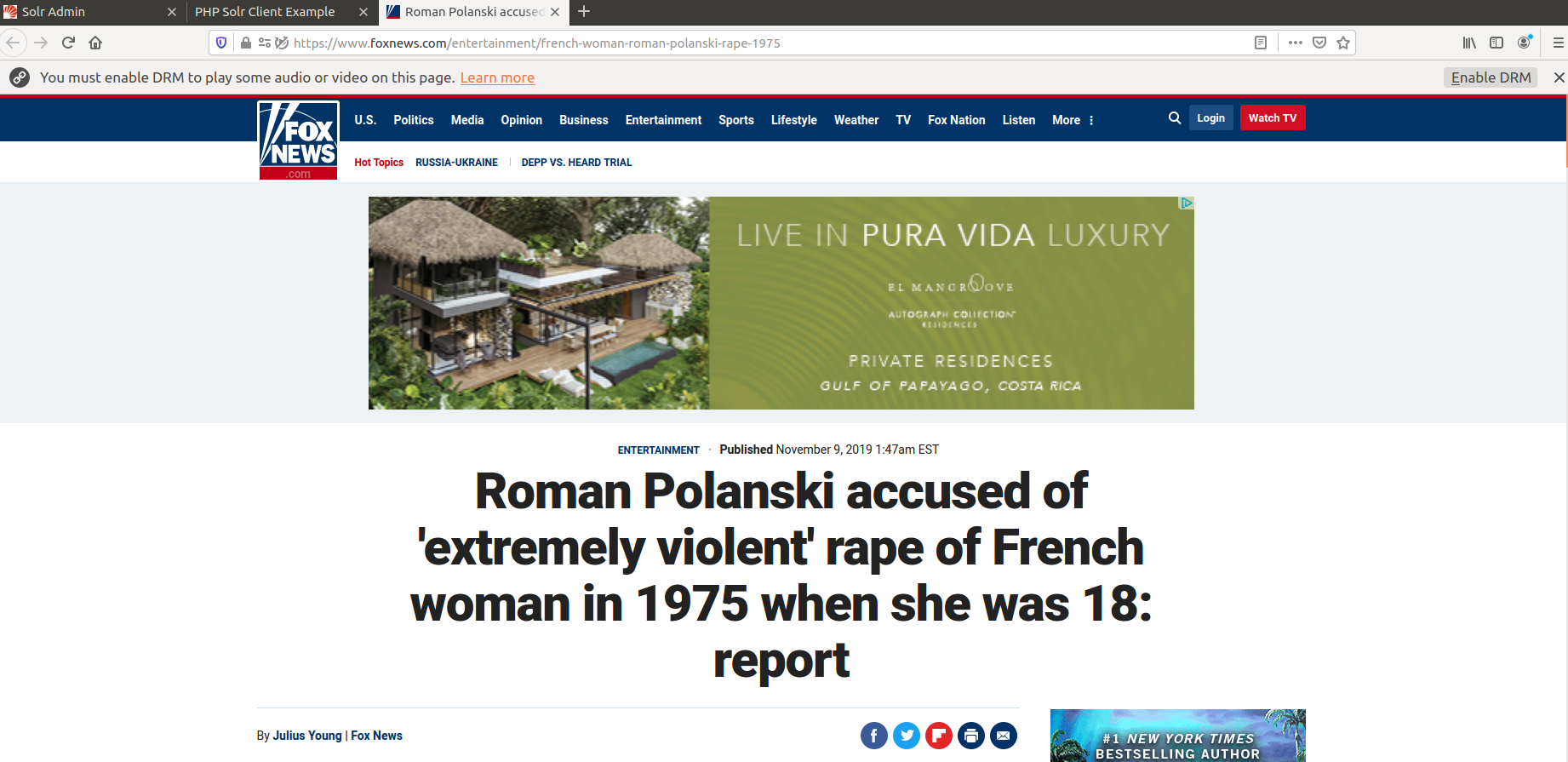


Table of query

Cannes

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/category/entertainment/events/cannes-film-festival> | <https://www.foxnews.com/entertainment/french-woman-roman-polanski-rape-1975> |
| 2 | <https://www.foxnews.com/entertainment/selena-gomez-bill-murray-whispered-cannes-film-festival-reveal> | <https://www.foxnews.com/entertainment/ryan-gosling-eva-mendes-is-definitely-not-anti-sweatpants> |
| 3 | <https://www.foxnews.com/entertainment/selena-gomez-cannes-film-festival-red-carpet> | <https://www.foxnews.com/entertainment/julie-andrews-therapy-saved-my-life> |
| 4 | <https://www.foxnews.com/entertainment/ryan-gosling-eva-mendes-is-definitely-not-anti-sweatpants> | <https://www.foxnews.com/entertainment/whitney-houston-was-sexually-abused-during-childhood-new-documentary-claims> |
| 5 | <https://www.foxnews.com/entertainment/quentin-tarantino-once-upon-a-time-in-hollywood-trailer-release> | <https://www.foxnews.com/entertainment/selena-gomez-mom-sister-concert> |
| 6 | <https://www.foxnews.com/entertainment/woody-allen-retire-family-politics> | <https://www.foxnews.com/entertainment/sharon-stone-rocks-bikini-see-her-many-ageless-looks> |
| 7 | <https://www.foxnews.com/entertainment/whitney-houston-was-sexually-abused-during-childhood-new-documentary-claims> | <https://www.foxnews.com/entertainment/dolph-lundgren-sylvester-stallone-rocky-iv-hospital> |
| 8 | <https://www.foxnews.com/entertainment/elton-john-rocketman-biopic-drugs-addiction-eating-disorder> | <https://www.foxnews.com/entertainment/michael-moore-endorses-bernie-sanders> |
| 9 | <https://www.foxnews.com/opinion/a-joan-rivers-thanksgiving> | <https://www.foxnews.com/entertainment/elton-john-rocketman-biopic-drugs-addiction-eating-disorder> |
| 10 | <https://www.foxnews.com/entertainment/selena-gomez-new-dog-instagram> | <https://www.foxnews.com/category/entertainment/events/cannes-film-festival> |

Congress

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/us/navy-vet-donates-gold-star-families-stories-to-library-of-congress> | <https://www.foxnews.com/politics/trump-tells-congress-to-immediately-approve-mexico-canada-trade-deal-says-us-moving-past-stupid-years-of-trade-policy> |
| 2 | <https://www.foxnews.com/opinion/judge-andrew-napolitano-congress-constitutional-power-declare-war> | <https://www.foxnews.com/tech> |
| 3 | <https://www.foxnews.com/opinion/rep-andy-biggs-congress-debt-ceiling-budget-deal> | <https://www.foxnews.com/category/world/conflicts> |
| 4 | <https://www.foxnews.com/opinion/gutfeld-congress-attacking-zuckerberg> | <https://www.foxnews.com/category/politics/foreign-policy> |
| 5 | <https://www.foxnews.com/politics/jon-stewart-blasts-congress-over-9-11-fund-hearing-drove-me-nuts-there-were-empty-seats> | <https://www.foxnews.com/category/tech/topics/security> |
| 6 | <https://www.foxnews.com/category/us/congress> | <https://www.foxnews.com/category/politics/executive> |
| 7 | <https://www.foxnews.com/faith-values/bible-school-drew-brees-day-congress> | <https://www.foxnews.com/category/politics/executive> |
| 8 | <https://www.foxnews.com/opinion/andy-biggs-congress-judges-immigration-laws> | <https://www.foxnews.com/opinion/mike-huckabee-lets-fight-for-fabricated-structural-steel-manufacturing-jobs> |
| 9 | <https://www.foxnews.com/opinion/bradley-blakeman-trade-pact-with-canada-and-mexico-offers-incredible-benefits-congress-should-approve-it> | <http://www.foxnews.com/trending> |
| 10 | <https://www.foxnews.com/politics/trump-tells-congress-to-immediately-approve-mexico-canada-trade-deal-says-us-moving-past-stupid-years-of-trade-policy> | <https://www.foxnews.com/media/elijah-cummings-seat-kimberly-klacik-baltimore-election-republican> |

Democrats

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/opinion/rnc-chair-mcdaniel-democrats-impeachment-vote-is-baseless-political-vendetta-trump-did-nothing-wrong> | <https://www.foxnews.com/us> |
| 2 | <https://www.foxnews.com/politics/federal-judge-house-democrats-mazars-subpoena-trump-financial-record> | <https://www.foxnews.com/opinion> |
| 3 | <https://www.foxnews.com/politics/democrats-arent-actually-impeaching-the-president-yet-but-they-are-digging-deep> | <https://www.foxnews.com> |
| 4 | <https://www.foxnews.com/media/democrats-disdain-working-class-trump-jr-fox-friends> | <https://www.foxnews.com/politics> |
| 5 | <https://www.foxnews.com/opinion/david-bossie-baseless-trump-impeachment-inquiry-by-desperate-democrats-makes-dramatic-flip-flop> | <https://www.foxnews.com/category/politics/elections> |
| 6 | <https://www.foxnews.com/opinion/rep-andy-biggs-lipstick-impeachment> | <https://www.foxnews.com/category/world/conflicts> |
| 7 | <https://www.foxnews.com/media/sen-lindsey-graham-every-american-should-be-bothered-by-what-democrats-are-doing-to-trump> | <https://www.foxnews.com/category/politics/foreign-policy> |
| 8 | <https://www.foxnews.com/opinion/david-bossie-trump-impeachment-vote-is-democratic-declaration-of-war-republicans-must-declare-war-on-dems> | <https://www.foxnews.com/category/world/scandals> |
| 9 | <https://www.foxnews.com/tech/google-facebook-and-twitter-staffs-splurge-on-democrats-ahead-of-midterms> | <https://www.foxnews.com/category/entertainment/music> |
| 10 | <https://www.foxnews.com/media/democrats-impeachment-abuse-ken-starr> | <https://www.foxnews.com/category/us/personal-freedoms> |

Patriot Movement

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/politics/trump-antifa-terrorist-portland-protests> | <https://www.foxnews.com/science> |
| 2 | <https://www.foxnews.com/world/iranian-cruise-missiles-drones-saudi-oil> | <https://www.foxnews.com> |
| 3 | <https://www.foxnews.com/transcript/gutfeld-whats-a-liberal-to-do-when-a-republican-turns-out-to-be-better-at-compassion> | <https://www.foxnews.com/tech> |
| 4 | <https://www.foxnews.com/media/fox-nation-patriot-awards-american-heroes-st-petersburg> | <https://www.foxnews.com/category/science/planet-earth> |
| 5 | <https://www.foxnews.com/media/marine-vet-most-valuable-patriot-awards-congress> | <https://www.foxnews.com/category/entertainment/movies> |
| 6 | <https://www.foxnews.com/media/fox-nation-patriot-awards> | <https://www.foxnews.com/entertainment/gwyneth-paltrow-relationship-harvey-weinstein> |
| 7 | <https://www.foxnews.com/media/tomi-lahren-patriot-awards-fox-nation> | <https://www.foxnews.com/politics/aoc-bloomberg-purchase-election> |
| 8 | <https://www.foxnews.com/faith-values/veterans-serve-disaster-relief> | <https://www.foxnews.com/world/ukraine-russia-separatists-pullback> |
| 9 | <https://www.foxnews.com/entertainment/netflix-pulls-patriot-act-with-hasan-minhaj-episode-after-getting-valid-legal-complaint-from-saudi-arabia> | <https://www.foxnews.com/science/earths-mantle-and-crust-are-in-a-fiery-battle-to-the-death-of-supercontinents> |
| 10 | <https://www.foxnews.com/tech/netflix-defends-pulling-patriot-act-episode-in-saudi-arabia> | <https://www.foxnews.com/auto/2020-ram-1500-built-to-serve-pickup-salutes-the-u-s-military> |

Republicans

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/politics/house-republicans-storm-impeachment-inquiry-deposition-what-went-down> | <https://www.foxnews.com/opinion> |
| 2 | <https://www.foxnews.com/opinion/tom-del-beccaro-republicans-must-win-the-impeachment-trial-and-they-can-by-following-these-five-steps> | <https://www.foxnews.com> |
| 3 | <https://www.foxnews.com/politics/republicans-hunter-biden-ukraine-whisteleblower-impeachment-witness> | <https://www.foxnews.com/politics> |
| 4 | <https://www.foxnews.com/politics/bill-weld-thrilled-other-republicans-are-challeneging-trump-in-primary> | <https://www.foxnews.com/category/politics/elections> |
| 5 | <https://www.foxnews.com/politics/house-republicans-plan-to-call-adam-schiff-to-testify-in-impeachment-inquiry-call-him-a-fact-witness> | <https://www.foxnews.com/category/world/conflicts> |
| 6 | <https://www.foxnews.com/politics/republicans-impeachment-hearings-lineup> | <https://www.foxnews.com/category/world/scandals> |
| 7 | <https://www.foxnews.com/politics/trump-settles-score-with-republicans-who-didnt-embrace-him-in-midterms> | <https://www.foxnews.com/category/politics/house-of-representatives> |
| 8 | <https://www.foxnews.com/opinion/newt-gingrich-house-republicans-2020-retirements> | <https://www.foxnews.com/category/politics/house-of-representatives> |
| 9 | <https://www.foxnews.com/opinion/matt-gorman-kavanaugh-republicans> | <https://www.foxnews.com/category/politics/senate> |
| 10 | <https://www.foxnews.com/media/joe-biden-republicans-epiphany-trump> | <https://www.foxnews.com/politics> |

Senate

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/politics/trump-senate-trial-impeachment-inquiry-reporters-notebook> | <https://www.foxnews.com/health> |
| 2 | <https://www.foxnews.com/category/politics/2020-senate-races> | <https://www.foxnews.com/entertainment> |
| 3 | <https://www.foxnews.com/media/jeff-sessions-former-embattled-trump-attorney-general-announces-bid-for-old-us-senate-seat> | <https://www.foxnews.com/shows> |
| 4 | <https://www.foxnews.com/opinion/newt-gingrich-left-viciously-attacks-mcconnell-because-hes-successful-senate-leader-and-trump-ally> | <https://www.foxnews.com/us> |
| 5 | <https://www.foxnews.com/faith-values/pastor-andrew-brunson-senate-trump-prayer> | <https://www.foxnews.com/world> |
| 6 | <https://www.foxnews.com/politics/jeff-sessions-alabama-senate-campaign-expected> | <https://www.foxnews.com/travel> |
| 7 | <https://www.foxnews.com/category/politics/elections/senate> | <https://www.foxnews.com/opinion> |
| 8 | <https://www.foxnews.com/politics/trump-sessions-against-senate-run-alabama> | <https://www.foxnews.com/science> |
| 9 | <https://www.foxnews.com/media/judge-pirro-jeff-sessions-senate-run-fox-friends> | <https://www.foxnews.com> |
| 10 | <https://www.foxnews.com/politics/virginia-democrat-accused-of-sex-with-teen-secretary-wins-state-senate-seat> | <https://www.foxnews.com/foxaroundtheworld> |

Olympics 2020

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/sports/alex-morgan-2020-olympics-child-birth> | <https://www.foxnews.com/entertainment> |
| 2 | <https://www.foxnews.com/sports/naomi-osaka-us-citizenship-japan-2020-olympics> | <https://www.foxnews.com/opinion> |
| 3 | <https://www.foxnews.com/category/sports/olympics> | <https://www.foxnews.com> |
| 4 | <https://www.foxnews.com/tech/russian-hackers-attack-possible-olympic-ban> | <https://www.foxnews.com/tech> |
| 5 | <https://www.foxnews.com/category/sports/soccer> | <https://www.foxnews.com/politics> |
| 6 | <https://www.foxnews.com/person/c/barnini-chakraborty> | <https://www.foxnews.com/lifestyle> |
| 7 | <https://www.foxnews.com/category/sports/tennis> | <https://www.foxnews.com/category/health/heart-health> |
| 8 | <https://www.foxnews.com/category/person/alex-morgan> | <https://www.foxnews.com/category/entertainment/style> |
| 9 | <https://www.foxnews.com/sports/kikkan-randall-new-york-city-marathon-skiing-cancer> | <https://www.foxnews.com/category/politics/elections> |
| 10 | <https://www.foxnews.com/sports> | <https://www.foxnews.com/category/world/global-economy> |

Stock

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/tech/jeff-bezos-to-be-no-2-richest-behind-bill-gates-after-amazon-stock-drop> | <https://www.foxnews.com> |
| 2 | <https://www.foxnews.com/transcript/cavuto-looking-back-at-the-stock-market-crash-30-years-ago> | <https://www.foxnews.com/category/us/economy> |
| 3 | <https://www.foxnews.com/transcript/cavuto-the-markets-are-easier-to-take-stock-of-than-life> | <https://www.foxnews.com/science/revolutionary-war-rifle-stolen-50-years-ago-on-display> |
| 4 | <https://www.foxnews.com/category/shows/your-world/transcript/common-sense> | <https://www.foxnews.com/entertainment/demi-lovato-first-interview-year> |
| 5 | <https://www.foxnews.com/person/c/gordon-chang> | <https://www.foxnews.com/media/chris-wallace-elizabeth-warren-medicare-for-all-plan> |
| 6 | <https://www.foxnews.com/category/tech/companies/microsoft> | <https://www.foxnews.com/tech/hp-recalls-more-laptop-lithium-ion-batteries-after-reports-of-overheating-melting-and-charring> |
| 7 | <https://www.foxnews.com/category/tech/companies/amazon> | <https://www.foxnews.com/tech/heres-how-the-trump-tariffs-hike-may-affect-prices-for-pc-parts> |
| 8 | <https://www.foxnews.com/category/person/sheryl-sandberg> | <https://www.foxnews.com/category/tech/companies/amazon> |
| 9 | <https://www.foxnews.com/sports/royals-hire-ex-cardinals-skipper-mike-matheny-as-manager> | <https://www.foxnews.com/opinion/max-lucado-fearless-parenting> |
| 10 | <https://www.foxnews.com/auto/fiat-chrysler-sales-boss-reid-bigland-sues-company-for-violating-whistleblower-protections> | <https://www.foxnews.com/auto/electric-ford-mustang-lithium> |

Virus

|  |  |  |
| --- | --- | --- |
| Rank | Lucene | PageRank |
| 1 | <https://www.foxnews.com/tech/computer-virus-cripples-tribune-publishing-los-angeles-times> | <https://www.foxnews.com/science> |
| 2 | <https://www.foxnews.com/science/deadly-seal-virus-spreading-melting-arctic-ice> | <https://www.foxnews.com/category/science/wild-nature> |
| 3 | <https://www.foxnews.com/health/connecticut-third-mosquito-eee-death> | <https://www.foxnews.com/category/science/planet-earth> |
| 4 | <https://www.foxnews.com/category/tech/topics/virus> | <https://www.foxnews.com/category/tech/topics/computers> |
| 5 | <https://www.foxnews.com/health/eee-eastern-equine-encephalitis-massachusetts-man-killed> | <https://www.foxnews.com/science/deadly-seal-virus-spreading-melting-arctic-ice> |
| 6 | <https://www.foxnews.com/health/what-is-norovirus-3-things-to-know-about-the-vomit-inducing-bug> | <https://www.foxnews.com/science/sandra-orangutan-legal-personhood-new-home-florida> |
| 7 | <https://www.foxnews.com/health/pregnant-women-who-get-flu-shot-are-less-likely-to-be-hospitalized-from-virus-cdc-says> | <https://www.foxnews.com/science/ancient-shipwrecks-stone-pyramid-anchors-discovered-greek-island> |
| 8 | <https://www.foxnews.com/health/flu-season-2019-vaccine> | <https://www.foxnews.com/science/earths-mantle-and-crust-are-in-a-fiery-battle-to-the-death-of-supercontinents> |
| 9 | <https://www.foxnews.com/category/health/infectious-disease/zika> | <https://www.foxnews.com/science/mammoth-traps-prehistoric-mexico> |
| 10 | <https://www.foxnews.com/health/british-mom-son-flu> | <https://www.foxnews.com/science/mercury-transit-2019-what-to-know> |

Overlap

|  |  |
| --- | --- |
| queries | overlaps |
| Cannes | 4 |
| Congress | 1 |
| Democrats | 0 |
| Patriot Movement | 0 |
| Republicans | 0 |
| Senate | 0 |
| Olympics 2020 | 0 |
| Stock | 1 |
| Virus | 1 |

Bar Graph