

The following are the steps I took to complete this project.

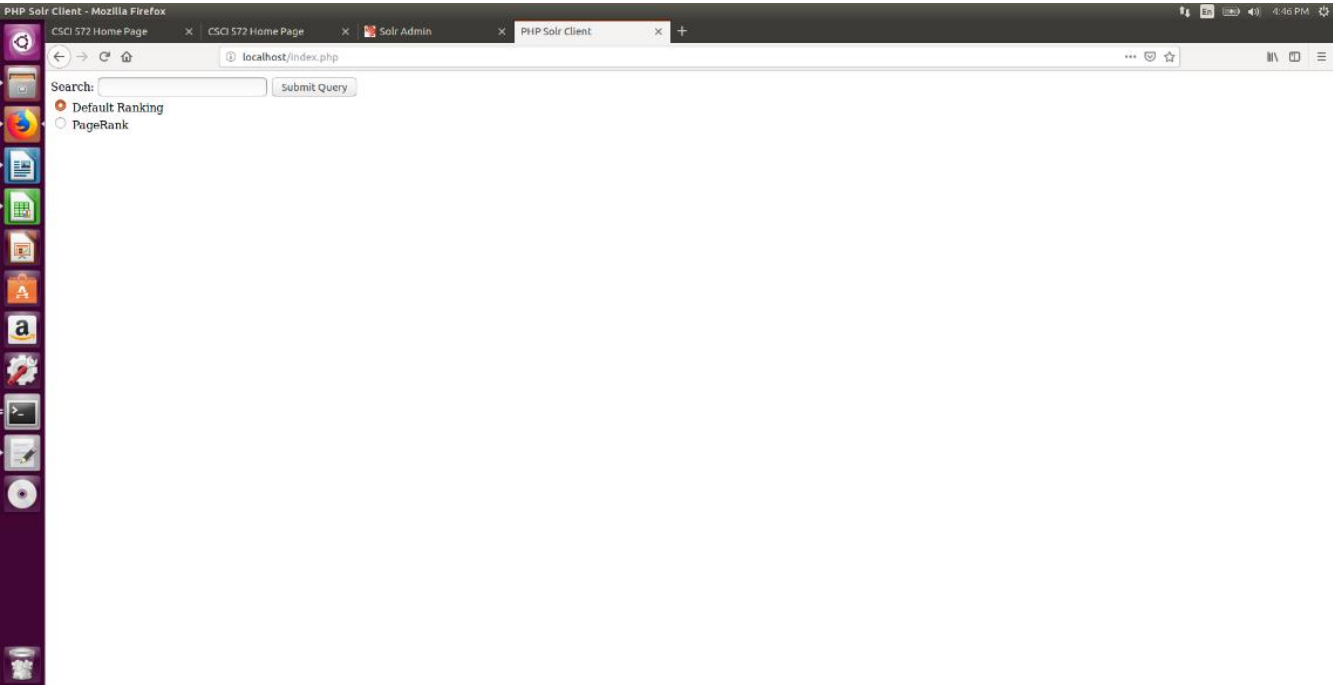
1. I first created an index for the Yahoo files with Solr per the instructions
2. I then modified solr-php-client to accept input from the user in a HTML form, and send the request to the Solr server. After the Solr server processes the query, it returns the results which are parsed by the PHP program and formatted for display. Adjustments made to the solr-php-client code were to add radio buttons and define specific filter parameters for the ranking algorithm that was chosen. I also modified the way results were displayed so that the title and og\_url fields linked to the actual webpages. This is all contained in one file: index.php
3. I then wrote a java program that extracted the links from the provided files and wrote the extracted links to a text file.
4. Next I used python and Networkx to create a graph with my extracted edges and calculate the pagerank for each node. The PageRank parameters I used were:  
alpha=0.85, personalization=None, max\_iter=30, tol=1e-06, nstart=None,  
weight='weight',dangling=None
5. The python code then wrote the full id path and page rank value to a text file (external\_pageRankFile.txt).
6. The file with PageRanks (external\_pageRankFile.txt) then had to be placed in the data folder of my Solr core and the managed-schema and solrconfig.xml files were modified to recognize the new file as per the instructions.

Overlap: I only had 1 overlap for the “Venezuela” query and 1 for the “Channel” query, the rest had 0. Many of the PageRank queries did not have og\_urls and I had to just list the high level web pages and correlate ids with urls in the provided URLtoHTML file.

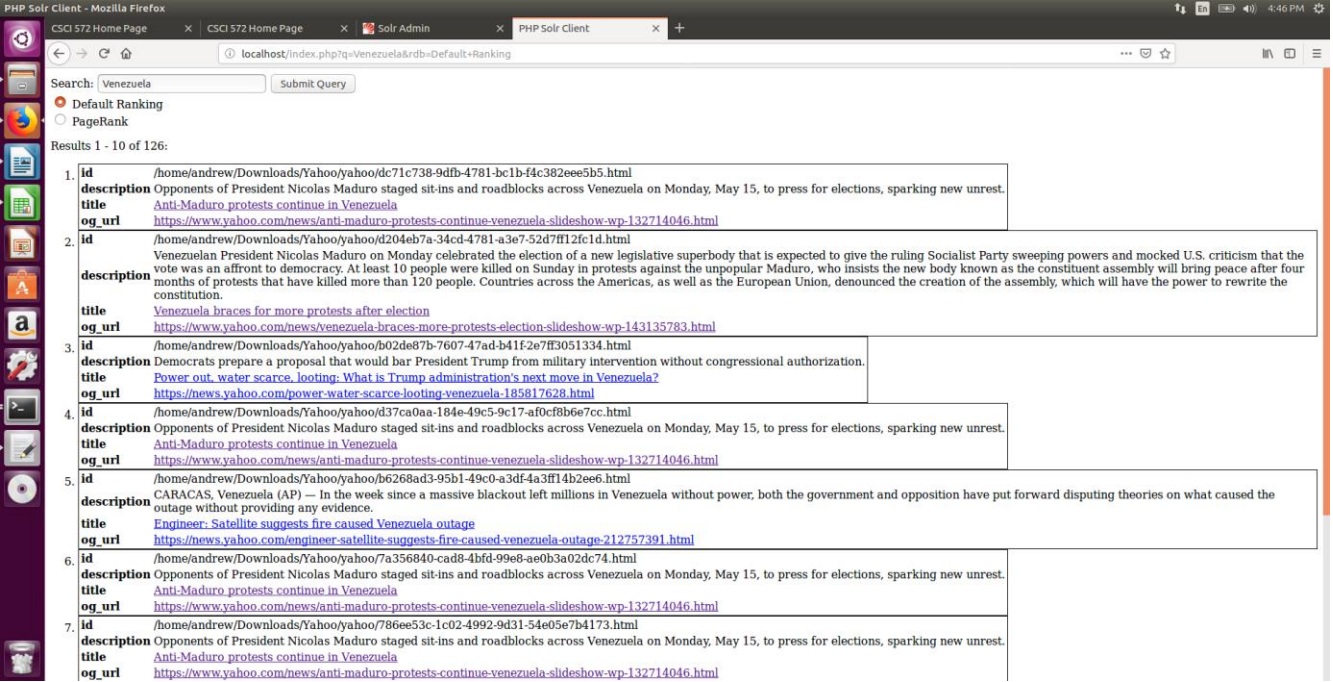
It seems that PageRank does not necessarily correlate with relevance. When using the PageRank algorithm, popular pages appeared in the results, but they weren't necessarily about the query, they just happened to mention the term. The default Lucene algorithm had more relevant results, but also had many duplicates. Some pages have higher page ranks than others because they are higher level domains and lots of other pages point to them.

Screenshots and additional data on following pages:

Initial Page:



Lucene Default



## PageRank (notice some files do not have og\_url)

PHP Solr Client - Mozilla Firefox

Search: Venezuela Submit Query

Default Ranking  
PageRank

Results 1 - 10 of 126:

- id** /home/andrew/Downloads/Yahoo/yahoo/73f85d91-7903-4110-a92f-90faa0380179.html

**description** The latest news and headlines from Yahoo! News. Get breaking news stories and in-depth coverage with videos and photos.

**title** Science | Yahoo News - Latest News & Headlines
- id** /home/andrew/Downloads/Yahoo/yahoo/f39d4318-1147-4165-bcc4-0b4aa3da4c8f.html

**description** Big stories dominated the news in 2016 like no other year: Donald Trump won. The U.K. backed the Brexit. Right-wing populism surged around the globe. And more.

**title** The top 10 undercover stories of 2016

**og\_url** <https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/fa310394-6c90-422f-ba1c-36fd81f4f5a6.html

**description** The canoe had been the idea of Tom Hewitt, the agents' case officer in the CIA's huge Miami station, where he waited for word of the mission. It would be Hewitt's job to guide their actions from afar, now that they were back in their homeland.

**title** How a CIA agent trained agents who found Soviet missiles in Cuba

**og\_url** <https://news.yahoo.com/operation-cobra-untold-story-cia-officer-trained-network-agents-found-soviet-missiles-cuba-100005794.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/951375d8-d56f4cf2-8d04-99471d76ff63.html

**description** The word "Trump" is conspicuously absent in Karen Handel's fierce campaign against Democrat Jon Ossoff in Georgia's Sixth Congressional District.

**title** In close House race, Republican Karen Handel gets help from Paul Ryan

**og\_url** <https://www.yahoo.com/news/close-house-race-republican-karen-handel-gets-help-paul-ryan-220451019.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/2d1ac33c-d16a-4d3a-bf98-2366cab760ca.html

**description** Democratic lawmakers on Tuesday tore into the Supreme Court's decision to uphold President Trump's travel ban targeting several Muslim-majority countries.

**title** Dems rip SCOTUS decision to uphold Trump's travel ban

**og\_url** <https://www.yahoo.com/news/marker-shame-democratic-lawmakers-rip-supreme-courts-travel-ban-decision-172209635.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/88b62b4e-aa86-49bb-bf1f-df1d65d04a66.html

**description** From around 2009 to 2013, the U.S. intelligence community experienced crippling intelligence failures related to its secret internet-based communications system, a key means for remote messaging between CIA officers and their sources.

**title** The CIA's communications suffered a catastrophic compromise

**og\_url** <https://news.yahoo.com/cias-communications-suffered-catastrophic-compromise-started-iran-090018710.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/b9dd22a1-2980-4916-a491-55d65f65a00f.html

**description** Condemning the "depraved" North Korean regime, the president mocked Kim Jong Un, saying: "Rocket Man" is on a suicide mission for himself and for his regime." Trump's stark message came in a speech in which he also sharply criticized China, mildly rebuked Russia, condemned Venezuela's government

**title** Trump warns U.S. may have to 'destroy' North Korea

**og\_url** <https://www.yahoo.com/news/trump-warns-u-s-may-totally-destroy-north-korea-161113772.html>
- id** /home/andrew/Downloads/Yahoo/yahoo/d6427a97-09ab-424f-bd3c-5f5f3bb2acd.html

**description** This is the first of two excerpts adapted from Russian Roulette: The Inside Story of Putin's War on America and the Election of Donald Trump (Twelve Books) by Michael Isikoff, Chief Investigative

## After clicking on link

The top 10 undercover stories of 2016 - Mozilla Firefox

https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html

Home Mail News Finance Sports Entertainment Search Mobile More

YAHOO! Search Search News Search web Sign in Mail

DigitalOcean Kubernetes is hard. We just made it easier for you to deploy faster. TRY FREE

### The top 10 undercover stories of 2016

YAHOO! News Yahoo News Staff, Yahoo News - December 30, 2016

Top, from left: Dakota Access pipeline (Stephanie Keith/Reuters), Arctic ice (NASA/AP), North Carolina flooding (Carlo Allegri/Reuters), Middle, from left: Drug prices (Dilley Mingasson/AP), Prison protest (Alex Milan Tracy/Sipa USA), Venezuela unrest (Carlos Garcia Rawlinz/Reuters), Voting (Elise Amendola/AP), Bottom, from

430

Kubernetes is hard. We just made it easier for you to deploy faster. DigitalOcean TRY FREE

What to Read Next

2 major cases add up to big doubts about

## **Table of queries:**

### **Lucene Default Query**

#### **Venezuela**

<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>  
<https://www.yahoo.com/news/venezuela-braces-more-protests-election-slideshow-wp-143135783.html>>

<https://news.yahoo.com/power-water-scarce-looting-venezuela-185817628.html>>

<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>

<https://news.yahoo.com/engineer-satellite-suggests-fire-caused-venezuela-outage-212757391.html>>

<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>  
<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>

<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>  
<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>  
<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>>

#### **Senate**

<https://www.yahoo.com/news/senate-skinny-repeal-obamacare-falls-apart-senate-floor-mccain-defects-072042708.html>>  
<https://www.yahoo.com/news/senate-judiciary-committee-vote-kavanaugh-slideshow-wp-170920887.html>>  
<https://www.yahoo.com/news/senate-judiciary-committee-vote-kavanaugh-slideshow-wp-170920887.html>>  
<https://www.yahoo.com/news/senate-judiciary-committee-vote-kavanaugh-slideshow-wp-170920887.html>>  
<https://www.yahoo.com/news/jeff-session-testifies-senate-hearing-slideshow-wp-191620787.html>>  
<https://www.yahoo.com/news/protesters-across-country-oppose-gops-slideshow-wp-133309283.html>>  
<https://www.yahoo.com/news/kavanaugh-ford-testify-senate-judiciary-slideshow-wp-145913238.html>>  
<https://www.yahoo.com/news/paul-ryan-premature-say-whether-house-pass-senate-healthcare-bill-174352395.html>>  
<https://www.yahoo.com/news/cbo-likely-say-millions-lose-coverage-senate-health-care-bill-215135603.html>>

### **PageRank Query**

#### **Venezuela**

<https://www.yahoo.com/news/science/>  
<https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html>  
<https://news.yahoo.com/operation-cobra-untold-story-cia-officer-trained-network-agents-found-soviet-missiles-cuba-100005794.html>  
<https://www.yahoo.com/news/close-house-race-republican-karen-handel-gets-help-paul-ryan-220451019.html>  
<https://www.yahoo.com/news/marker-shame-democratic-lawmakers-rip-supreme-courts-travel-ban-decision-172209635.html>  
<https://news.yahoo.com/cias-communications-suffered-catastrophic-compromise-started-iran-090018710.html>  
<https://www.yahoo.com/news/trump-warns-u-s-may-totally-destroy-north-korea-161113772.html>  
<https://www.yahoo.com/news/miss-universe-moscow-trumps-beauty-contest-spawned-business-deal-russians-bond-putin-100026386.html>  
<https://www.yahoo.com/news/anti-maduro-protests-continue-venezuela-slideshow-wp-132714046.html>  
<https://www.yahoo.com/news/washington-sets-new-sanctions-syria-regime-215119677.html>

#### **Senate**

<https://www.yahoo.com/news/>

<https://www.yahoo.com/news/originals/>

<https://www.yahoo.com/news>

<https://www.yahoo.com/news/politics/>

<https://www.yahoo.com/news/tagged/skullduggery>

<https://www.yahoo.com/news/topics>

<https://www.yahoo.com/news/politics>

<https://www.yahoo.com/news/topics/president-trump>

<https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html>

<https://www.yahoo.com/news/trump-senate-republicans-look-like-fools-dont-dump-60-vote-rule-135937103.html>>

## **Democrats**

<https://www.yahoo.com/news/midterms/house/seat/>  
<https://www.yahoo.com/news/midterms/house>  
<https://www.yahoo.com/news/democrats-rolling-2018-midterms-message-next-week-191517316.html>>  
<https://www.yahoo.com/news/will-democrats-filibuster-gorsuch-republicans-dont-think-so-111241581.html>>  
<https://www.yahoo.com/news/white-house-tries-tag-democrats-blame-possible-shutdown-203618659.html>>  
<https://www.yahoo.com/news/democrats-hope-better-deal-message-will-resonate-190849661.html>>  
<https://www.yahoo.com/news/democrats-2018-many-candidates-many-candidates-162905320.html>>  
<https://www.yahoo.com/news/democrats-vow-fight-lifetime-keep-brett-kavanaugh-off-supreme-court-051334537.html>>  
<https://www.yahoo.com/news/mcconnell-plan-fails-will-democrats-republicans-work-together-fix-health-care-220450895.html>>

<https://www.yahoo.com/news/democrats-wrest-power-away-republicans-state-level-183309710.html>>

## **Republicans**

<https://www.yahoo.com/news/midterms/house/seat/>  
<https://www.yahoo.com/news/midterms/house>  
<https://www.yahoo.com/news/mcconnell-plan-fails-will-democrats-republicans-work-together-fix-health-care-220450895.html>>  
<https://www.yahoo.com/news/grim-warnings-white-house-republicans-ahead-election-171354295-election.html>>  
<https://www.yahoo.com/news/democrats-wrest-power-away-republicans-state-level-183309710.html>>  
<https://www.yahoo.com/news/republicans-rush-obamacare-repeal-and-replace-through-house-committees-233538837.html>>  
<https://www.yahoo.com/news/gop-obamacare-replacement-narrowly-passes-out-of-committee-after-3-republicans-defect-145519718.html>>  
<https://www.yahoo.com/news/will-democrats-filibuster-gorsuch-republicans-dont-think-so-111241581.html>>  
<https://www.yahoo.com/news/house-republicans-reveal-obamacare-replacement-bill-013510100.html>>  
<https://news.yahoo.com/down-ticket-5-republicans-finally-000000015.html>>

## **Patriot Movement**

<https://www.yahoo.com/news/jeffersons-tree-liberty-blood-schoolchildren-171918999.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>

<https://news.yahoo.com/russians-targeted-reagan-presidential-campaign-heres-cia-stop-100048764.html>

## **Democrats**

<https://www.yahoo.com/news/>  
<https://www.yahoo.com/news/originals/>  
<https://www.yahoo.com/news>

<https://www.yahoo.com/news/politics/>

<https://www.yahoo.com/news/politics>  
<https://www.yahoo.com/news/topics/president-trump>  
<https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html>  
<https://news.yahoo.com/russians-targeted-reagan-presidential-campaign-heres-cia-stop-100048764.html>

<https://news.yahoo.com/ann-coulter-lunatic-trump-challenged-2020-right-165049018.html>  
<https://www.yahoo.com/news/64-hours-october-one-weekend-blew-rules-american-politics-2-162827162.html>

## **Republicans**

<https://www.yahoo.com/news/>  
<https://www.yahoo.com/news/originals/>

<https://www.yahoo.com/news>

<https://www.yahoo.com/news/politics/>  
<https://www.yahoo.com/news/tagged/when%20presidents%20lead/>

<https://www.yahoo.com/news/tagged/skullduggery>

<https://www.yahoo.com/news/tagged/through-her-eyes>

<https://www.yahoo.com/news/science/>

<https://www.yahoo.com/news/us/>

<https://www.yahoo.com/news/topics>

## **Patriot Movement**

<https://www.yahoo.com/news>  
<https://www.yahoo.com/news/tagged/through-her-eyes>

<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>  
<https://www.yahoo.com/news/wing-patriot-prayer-rally-counter-slideshow-wp-212529094.html>>

### **Oscar 2019**

<https://news.yahoo.com/spike-lee-pushes-back-trump-spat-continues-oscar-speech-162915413.html>>  
<https://news.yahoo.com/skullduggery-power-vice-000835154.html>>  
<https://www.yahoo.com/news/topics/star-wars>  
<https://news.yahoo.com/trump-bashes-spike-lee-racist-hit-oscars-123716258.html>>

<https://www.yahoo.com/news/topics/royal-family>  
<https://news.yahoo.com/kamala-harris-trump-racist-185542673.html>>  
<https://news.yahoo.com/the-names-behind-the-college-cheating-charges-141742295.html>>  
<https://www.yahoo.com/news/topics/technology>  
<https://news.yahoo.com/gore-us-getting-close-political-shift-climate-change-223933667--politics.html>>  
<https://www.yahoo.com/news/heres-know-orlando-victims-000000001.html>>

### **Channel**

<https://www.yahoo.com/news/topics/fox-news-channel>

<https://www.yahoo.com/news/topics/cnn>  
<https://www.yahoo.com/news/service-charlottesville-victim-heather-heyer-call-righteous-action-172114070.html>  
<https://www.yahoo.com/news/topics/megyn-kelly>  
<https://www.yahoo.com/news/topics/marijuana-legalization>  
<https://www.yahoo.com/news/topics/space-exploration>

<https://www.yahoo.com/news/topics/health>  
<https://www.yahoo.com/news/topics/sex-crimes>

<https://www.yahoo.com/news/2016-pictures-news-slideshow-wp-100040397.html>  
<https://www.yahoo.com/news/topics/president-trump>  
<https://www.yahoo.com/news/the-top-10-undercovered-stories-of-2016-100009854.html>  
<https://www.yahoo.com/news/hate-america-comes-back-100032161.html>

<https://www.yahoo.com/news/tagged/midterms>

<https://www.yahoo.com/news/tagged/midterms/>  
<https://www.yahoo.com/news/cory-booker-real-090003827.html>  
<https://news.yahoo.com/crackdown-opioids-victims-people-need-live-100058361.html>

### **Oscar 2019**

<https://www.yahoo.com/news/>

<https://www.yahoo.com/news/originals/>  
<https://www.yahoo.com/news>

<https://www.yahoo.com/news/politics/>  
<https://www.yahoo.com/news/tagged/when%20presidents%20lead/>

<https://www.yahoo.com/news/tagged/skullduggery>  
<https://www.yahoo.com/news/tagged/through-her-eyes>  
<https://www.yahoo.com/news/science/>

<https://www.yahoo.com/news/us/>

<https://www.yahoo.com/news/topics>

### **Channel**

<https://www.yahoo.com/news/topics>  
<https://www.yahoo.com/news/topics/president-trump>  
<https://www.yahoo.com/news/64-hours-october-one-weekend-blew-rules-american-politics-2-162827162.html>  
<https://www.yahoo.com/news/topics/supreme-court>  
<https://www.yahoo.com/news/topics/russia-investigation>

<https://www.yahoo.com/news/cory-booker-real-090003827.html>  
<https://www.yahoo.com/news/skullduggery-washington-think-tank-says-no-saudi-funds-124450269.html>

<https://news.yahoo.com/crackdown-opioids-victims-people-need-live-100058361.html>

<https://www.yahoo.com/news/topics/sanctuary-cities>

<https://www.yahoo.com/news/topics/adam-ripon>

## Wall

<https://www.yahoo.com/news/wall-wall-trump-defends-border-plan-kelly-suggests-changed-145144558.html>>

<https://www.yahoo.com/news/imagining-trumps-big-beautiful-wall-might-183350584.html>>

<https://www.yahoo.com/news/gop-leaders-back-at-least-12-billion-for-border-wall-153042941.html>>

<https://www.yahoo.com/news/trumps-big-beautiful-wall-collides-congress-102804089.html>>

<https://www.yahoo.com/news/border-wall-funding-threatens-government-shutdown-budget-battle-heats-180534239.html>>

<https://www.yahoo.com/news/on-the-border-waiting-for-the-big-beautiful-wall-225959295.html>>

<https://www.yahoo.com/news/sanders-calls-obamas-400k-wall-street-speaking-fee-unfortunate-150857262.html>>

<https://www.yahoo.com/news/steve-king-has-a-model-of-the-border-wall-he-wants-to-build-to-protect-our-superior-civilization-183248446.html>>

<https://www.yahoo.com/news/rebuilding-great-wall-china-slideshow-wp-163045771.html>>

<https://www.yahoo.com/news/rebuilding-great-wall-china-slideshow-wp-163045771.html>>

<https://www.yahoo.com/news/unconventional-17-trump-meets-man-000000272.html>

<https://www.yahoo.com/news/service-charlottesville-victim-heather-heyer-call-righteous-action-172114070.html>

## Wall

<https://www.yahoo.com/news/>

<https://www.yahoo.com/news/originals/>

<https://www.yahoo.com/news>

<https://www.yahoo.com/news/politics/>

<https://www.yahoo.com/news/tagged/when%20presidents%20lead/>

<https://www.yahoo.com/news/tagged/skullduggery>

<https://www.yahoo.com/news/tagged/through-her-eyes>

<https://www.yahoo.com/news/science/>

<https://www.yahoo.com/news/us/>

<https://www.yahoo.com/news/topics>

