Automatically Generating a Large, Culture-Specific Blocklist for China

Austin Hounsel, Prateek Mittal, Nick Feamster
Princeton University
Internet censorship is regularly measured.
Test lists are used as input

- Manually curated
  - Citizen Lab Block List
- Automatically generated
  - Darer 2017, “FilteredWeb”
  - Sfakianakis 2011, “CensMon”
- Crowdsourced
  - Herdict
  - GreatFire.org
However, these lists are…
Not frequently updated

- Manually curated:
  - CLBL has 8 commits in past 3 years

- Automatically generated:
  - CensMon published in 2011

- Crowdsourced lists limited to what users submit
Not all publicly available

- FilteredWeb
  - We were shown the results for validation
- CensMon
  - Complete block list from Chinese government
Mainly focused on English websites

- CLBL has limited # of China-specific websites
  - Mostly human rights orgs
- FilteredWeb doesn’t parse Chinese websites
Our contributions

• Large, up-to-date, publicly available list of censored domains in China
  • None of these websites are in the largest blocklist
• Qualitative analysis of China-specific websites
• Pull-request into CLBL (in progress)
“Censorship findings are only as interesting as the sites and services that you test.”

– Open Observatory of Network Interference
So.. what are we missing?
God's Management Plan for Saving Mankind—The Three Stages of Work

- Jehovah Decrees Laws to Lead Man's Life
- Jesus Redeems Man Through Crucifixion
- Almighty God Expresses the Truth to Judge and Purify Man

Christ of the Last Days Has Appeared in China

“For as the lightning comes out of the east, and shines even to the west; so shall also the coming of the Son of man be” (Matthew 24:27). “And at midnight there was a cry made, Behold, the bridegroom comes; go you out to meet him” (Matthew 25:6). “Behold, I stand at the door, and knock: if any man hear my voice, and open the door, I will come in to him, and will sup with him, and he with me” (Revelation 3:20).
Hong Kong: Banning of political party marks serious escalation of political repression

3 weeks ago

Tiny Hong Kong National Party is being used as a test case for a wider political crackdown
Feng Zhenghu went to

Who is destroying the election

- Feng Zhenghu went to Shangjiuyuan Zhou...
- Citizen Supervisory Group's 10th blocked ...
- The 8th notice of the citizen supervision gr...
- Notice of the 5th time that the civil rights l...
- Feng Zhenghu went to Suzhou to sweep th...
- Feng Zhenghu sent a letter to Li Qiang for ...
- Feng Zhenghu sent a letter to Li Qiang for ...

- The 10th blocked notice of the essay | citizenshipi...
- Litigation case | guard Feng Zhenghu's financial ...
- Announcement of the 8th blocked notice of the es...
- Miscellaneous essays to | protect the civil rights l...
- Miscellaneous essays | Feng Zhenghu went to Su...
- The painstaking work of Feng Zhenghu's hairc...
- Li Zhengzheng and others sued Cui Yadong, Presi...
- The painstaking work of the | illegal police escort...
- The painstaking work of Feng Zhenghu was ill...
- The painstaking work of Feng Zhenghu was det...

- Feng Zhenghu's election for the Shanghai People's ...
- The opening of personal information of the National ...
- Who can find the NPC deputies in China? (March 4, ...
- Establish legal authority and abolish "black jails" ...
- What should China's NPC deputies do?
- Who can find the NPC deputies in China? (March 6, ...
- Constitutional rights

Good life - Feng Zhenghu's

The page you have

Feng Zhenghu de Twitter

Tweets by @lzhenghu
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

Bootstrap w/ block list
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

0. Bootstrap w/ block list
What do we mean by “phrases”?

- We use “n-grams” of Chinese words/phrases as search terms.
- Example: “Chinese human rights violation”
  - Bigrams: Chinese human, human rights, rights violation
  - Trigrams: Chinese human rights, human rights violation
What do we mean by “phrases”?

• However, because Chinese is written without spaces, word boundaries are probabilistic

• Example: “特 首”
  • Bigram translation: Chief executive
  • Unigram translations: Special, first

• We use whatever phrases Stanford CoreNLP outputs
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

Bootstrap w/ block list

Bing

China
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

Bootstrap w/ block list

China
Bing
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

Bootstrap w/ block list

Bing

China
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

0. Bootstrap w/ block list
Our approach

1. Extract phrases
2. Perform search queries
3. Test for censorship
4. Store censored pages

Bootstrap w/ block list

Bing

China
Finding #1: Blocklists are incomplete.
Wu Tong commits Taiwan... Emperor's fear of eating evil fruit? US experts: China can't

In the middle school, "Roric 6th Bureau has no responsibility to lose the Lions 10 to 6 defeat."
Maoxun 2018-08-10 23:49

Trump has approved the "chain immigration". The parents of the Yue are naturalized as US citizens.
2018-08-10 23:00

Even the US government has fallen? The Chinese map of the State Council website is included in the interpretation of Taiwan's Washington.
2018-08-10 22:45

The ruling of Oualcomm’s
The mirror is pleasing! You may find your stage here. Mirror fire recruiting: marketing, advertising, program host, TV reporter.
The Weizheng Dasi Temple in Tongxin County, Ningxia Hui Autonomous Region, was triggered by the local government's order to dismantle the temple and the police confrontation with the temple and the suppression. Thousands of people have vowed to defend the mosque, and the police who have to execute the order today must have a bloody crisis.

According to a report by the Beijing-based reporter Haile, the Huizhou Great Mosque in Tongxin County, Ningxia Hui Autonomous Region was threatened with demolition and caused thousands of Hui people to defend. The mosque was recently renovated and expanded. A total of nine mosques have a new round of domes, and the official allegations of illegal construction of the mosque were ordered to be dismantled.

Chinese authorities are increasingly restricting religious supervision. The just-concluded National Religious Association meeting not only requires religious flags to fly the national flag, but also in the hall where the believers pray, and also hangs the axe and the Communist Party flag. But so far, the Hui people are slightly different from the Xinjiang Uyghur residents, and they can still pray for relatively loose prayers. However, the Ningxia authorities ordered the rectification of the Weizhou Great Mosque, which was considered by the local Hui people to be a sign of the knife.

The local party and government departments all pointed out that the illegal buildings of Weizhou Dazhen Temple in Tongxin County of Ningxia Hui Autonomous Region were not allowed to persuade the local Hui people. Because the mosque was rebuilt for a long time, it was not accused by the government, and no government department allowed it to rebuild. It is impossible to start work on the project.

Residents of the Huizheng Temple in Weizhou, Tongxin County, Ningxia Hui Autonomous Region, said that they still support the government's policies. The slogans not only support the Chinese Communist Party, but also support the unity
China News

Latest News

- Does Mount Everest’s record wet summer point to a ‘profound change’ in the Earth’s climate? - scmp
- 2018 Shanghai Isaac Stem International Violin Competition Launches - Asia Today
- Protest by Chinese Hui Muslims at Weizhou Grand Mosque goes on as sides seek compromise - scmp
- Pachinko king Kazuo Okada denies charges in Hong Kong arrest - scmp
- ‘Asian August’ comes to Hollywood but will it last? - scmp
- Bright industry outlook echoed by positive exhibitor recruitment at SPS – Industrial Automation Fair Guangzhou - Asia Today
- Ecodecly: The choice of the right cleaning technology decides Solvents are often a superior alternative in terms of cost efficiency and sustainability - Asia Today
- ‘Just tell me why my father died.’ Daughter of dead security guard lashes out at Hong Kong government for ignoring his family in their time of grief - scmp
- Chinese corruption inquiry finds US$39 million in cash at homes of ex-boss of Huarong Asset Management - scmp
- How Can a Small Business Sell Overseas - Asia Today
- In vaccination push, 184 Hong Kong primary schools to give free flu jabs - scmp
- Cambodian PM Hun Sen’s Facebook page was hacked to give away seats in
Finding #2: Many unpopular websites are blocked
Home

Pseudo-objective, pseudo-fair, pseudo-flat, pseudo-academic, special purpose sarcasm

Pseudo-base

Everyone can edit the spoof encyclopedia

There are 6,567 bug entries

Featured entry

- Tao Yuanming
  Tao Yuanming, the world's most famous [redacted] writer, the world's poorest writer, once served as a small official, resigned because of dissatisfaction with official culture [redacted].

- Xi Jinping
  Or abide by the buns, the master of the common horror empire
  Qingfeng emperor, referred to as Xi Di, surname Xi Wei Wei Ni, the full name of the curfew evaluation. It was inaugurated in the Western calendar year, and was the chairman of the Central Committee of the Communist Party of China and the Chairman of the Central Military

News

- On July 31, the Public Security Bureau caught the [redacted] which is famous for its rumors, recently promulgated the burial and cremation. It is not only consiered to be a, but even an incident in which the squadron forcibly protest against the government's push for cremation was burned down by the villagers [redacted].

- On July 26, popcorn exploded? At one o'clock time, the US Embassy in Beijing, China, was attacked with white powder, and the ground was full of broken glass, the ground was full of heart, showing the last piece of popcorn
Islam Question and Answer
General Supervisor: Shaykh Muhammad Saalih al-Munajjid

Fri 29 DhQ 1439 - 10 August 2018

Check out the new version of Islam Q&A

New Fatwas

He left Makkah before doing the tawaf and sa'i of Hajj.

If someone reaches Makkah at Zuhr on the day of ‘Arafah and goes out to ‘Arafat and does not do tawaf al-qudoom (tawaf of arrival).

Is it prescribed to kiss the Black Stone without doing tawaf?

He formed the intention to do sa'i during the first lap.

Should a bald man who has no hair pass a razor over his head?

When he reaches Muzdalifah he should pray Maghrib and ‘Ishaa’, putting them together at the time of the earlier or later prayer.

Do they have to stay in Mina afterstoning the Jamaraat on the 13th of Dhu’l-Hijjah?

He is not sure whether he threw six or seven pebbles when Stoning the Jamrah on behalf of someone else.

Is it permissible to take stones or earth from Muzdalifah or from the Haram?

She was not sure how many circuits of tawaf she had done, and she did another circuit after one.
Southern Mongolia Congress
Official website

ABOUT SOUTHERN MONGOLIA CONGRESS
REPORT ON HUMAN RIGHTS SITUATION IN SOUTHERN MONGOLIA
THE COURSE OF ACTION

THE POLITICAL PLATFORM OF SOUTHERN MONGOLIA CONGRESS

OTHERS
SITUATION OF VIOLATION OF UNITED NATIONS CONVENTION ON THE ELIMINATION OF ALL FORMS OF DISCRIMINATION IN PEOPLE'S REPUBLIC OF CHINA

Chinese characteristics and Chinese model
Political Report of the Third National Congress of the Chinese Labor Party
square round

Guests, consultants, delegates: You are good!

The Third National Congress of the Chinese Labor Party is an important conference that will lead the way. According to our two major assumptions, this conference should be held from Hong Kong, the second largest venue, to the mainland. Because as early as eleven years ago, we have already raised the issue of revolution and improvement. The situation at the time was a reform and revolutionary race. We pointed out that the choice of improvement, or the choice of revolution, the initiative is not in the hands of the people, but in the hands of the rulers. We declare that if the rulers choose to improve, the Chinese Labor Party is willing to give up the revolution and embrace reform. If the rulers give up the reform, the Chinese Labor Party will have to rise to the revolution.

Since Hu Wen’s pro-government launched the scientific development concept of “people-oriented, harmonious society”, on the second day, we established the reconciliation route of casting swords into plows. At that time, we thought that after the improvement of Hu Wen, China should make great progress, and we can return to the mainland to hold the three major. It seems that our expectations for Hu Wen are too high.

Since the founding of the party on November 17, 2000, the Chinese Labor Party has been nearly fourteen years old. Two years before the founding of the Party, that is, in 1998, Dr. Wang Bingzhang had a sword in ten years, and he returned to China to establish the modern China Democracy Party. On the eve of the founding of the Democratic Party, Bing Zhang brother personally flew to Australia to negotiate with me, hoping that the Labour Party in preparation for the party will become part of the upcoming Democratic Party in order to fight a unified banner. We agreed with the advice.
Finding #3: Political phrases are very effective as search terms

<table>
<thead>
<tr>
<th>Chinese</th>
<th>English</th>
<th>Censored domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>王岐山</td>
<td>Wang Qishan</td>
<td>74%</td>
</tr>
<tr>
<td>李洪志</td>
<td>Li Hongzhi</td>
<td>64%</td>
</tr>
<tr>
<td>郭伯雄</td>
<td>Guo Boxiong</td>
<td>62%</td>
</tr>
<tr>
<td>胡锦涛</td>
<td>Hu Jintao</td>
<td>56%</td>
</tr>
<tr>
<td>胡平</td>
<td>Hu Ping</td>
<td>54%</td>
</tr>
<tr>
<td>Morty</td>
<td>Morty</td>
<td>52%</td>
</tr>
<tr>
<td>命案</td>
<td>Homicide</td>
<td>52%</td>
</tr>
<tr>
<td>特首</td>
<td>Chief executive</td>
<td>52%</td>
</tr>
<tr>
<td>Vimeo</td>
<td>Vimeo</td>
<td>50%</td>
</tr>
<tr>
<td>中情局</td>
<td>CIA</td>
<td>50%</td>
</tr>
</tbody>
</table>

Table 2: Sample of unigrams with significant blockrates
Finding #3: Political phrases are very effective as search terms

<table>
<thead>
<tr>
<th>Chinese</th>
<th>English</th>
<th>Censored domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>中共 威胁</td>
<td>Chinese Communists threaten</td>
<td>50%</td>
</tr>
<tr>
<td>声明的 反共产主义</td>
<td>Declared anti-communist</td>
<td>44%</td>
</tr>
<tr>
<td>中国共产党的公共安全</td>
<td>Public security of the CPC</td>
<td>42%</td>
</tr>
<tr>
<td>北京 清洁</td>
<td>Beijing clean-up</td>
<td>40%</td>
</tr>
<tr>
<td>江泽民 胡锦涛</td>
<td>Jiang Zemin Hu Jintao</td>
<td>40%</td>
</tr>
<tr>
<td>迫害 活动</td>
<td>Persecution activities</td>
<td>40%</td>
</tr>
<tr>
<td>官员 呼吁</td>
<td>Officials called on</td>
<td>36%</td>
</tr>
<tr>
<td>重的 公民</td>
<td>Heavy Citizen</td>
<td>36%</td>
</tr>
<tr>
<td>非法 拘留</td>
<td>Illegal detention</td>
<td>36%</td>
</tr>
<tr>
<td>不同 的民主</td>
<td>Different Democratic</td>
<td>34%</td>
</tr>
</tbody>
</table>

Table 3: Sample of bigrams with significant block rates
Finding #3: Political phrases are very effective as search terms

<table>
<thead>
<tr>
<th>Chinese</th>
<th>English</th>
<th>Censored domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>北戴河会议</td>
<td>BEIDAIHE meeting</td>
<td>54%</td>
</tr>
<tr>
<td>中国共产党的宗教政策</td>
<td>The Chinese Communist Party’s religious policy</td>
<td>42%</td>
</tr>
<tr>
<td>采取暴力 镇压</td>
<td>To take a violent crackdown</td>
<td>38%</td>
</tr>
<tr>
<td>香港政治</td>
<td>Hong Kong Politics</td>
<td>34%</td>
</tr>
<tr>
<td>欧洲议会决议</td>
<td>European Parliament Resolution</td>
<td>32%</td>
</tr>
<tr>
<td>新唐王朝</td>
<td>New Tang Dynasty</td>
<td>32%</td>
</tr>
<tr>
<td>恐怖事件。</td>
<td>A terrorist event.</td>
<td>32%</td>
</tr>
<tr>
<td>天安门广场示威</td>
<td>Tienanmen Square Demonstrations</td>
<td>32%</td>
</tr>
<tr>
<td>敦促美国政府</td>
<td>Urging the US government</td>
<td>32%</td>
</tr>
<tr>
<td>1989 民主运动</td>
<td>1989 democracy movement</td>
<td>30%</td>
</tr>
</tbody>
</table>

**Table 4:** Sample of trigrams with significant blockrates
Future work

• Add support for other languages to find censored websites pertaining to other countries

• Measure churn; do additional websites become (un)censored over time?

• Compare results across search engines
Conclusion

• We built a large, publicly-available, culture-specific list of censored domains in China

• It is important to keep testing lists up to date to understand how censorship in a given country changes over time

• https://github.com/ahounsel/auto-blocklist