

~~Wiki~~leaks Media Research Project

Christine Corbett Moran



@corbett_inc

Christine Corbett

I have not yet developed my position on
[@wikileaks](#). I have the cables, the internet,
now where to start? where to finish?

4 Dec via [Tweetie for Mac](#) ☆ [Favorite](#) ↩ [Reply](#) 🗑 [Delete](#)

4 Dec via [Tweetie for Mac](#) ☆ [Favorite](#) ↩ [Reply](#) 🗑 [Delete](#)

now where to start? where to finish?

Motivation



@ario

Ario Jafarzadeh

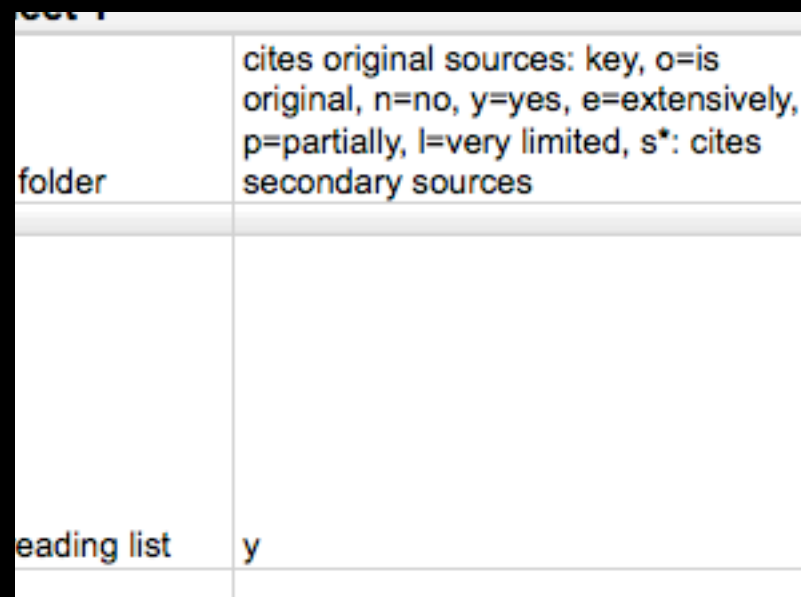
@corbett_inc Start here: <http://bit.ly/fnOtWK> Finish with balaclavas and molotovs.

4 Dec via TweetDeck ☆ Favorite ↻ Retweet ↩ Reply

4 Dec via TweetDeck ☆ Favorite ↻ Retweet ↩ Reply

molotovs.

Motivation



folder	cites original sources: key, o=is original, n=no, y=yes, e=extensively, p=partially, l=very limited, s*: cites secondary sources
reading list	y

```

- 133x80

#09diarywar.html.to instapaper
has_thousands_of_secret_documents;_427international_embarrass
#fountlees47_lives,_Obama_administration_warns_Wikileaks_to
_attached_millions_of_files_to_be_released_tonight to instap
/01/wikileaks-knocked-off-net-dns-everyday to instapaper
818622208 to instapaper
pi/www.wikileaks.org/wiki/Wikileaks>About to instapaper
to instapaper
o instapaper
e74dswmch.html_y=14hp to instapaper
pi/www.wikileaks.org/wiki/Wikileaks>AboutWikileaks_has_1.2
Wikileaks_has_1.2_million_documents.3F to instapaper
wikileaks-us-army-iraq-attack to instapaper
are-of-index-on-censorship-freedom-of-expression-award-announ
lood_-_Report_on_Etra-Judicial_Killings_and_Disappearances,_
=18227 to instapaper
010-05-20_5_phoneering_web_sites_that_would_totally_change_th
paper
an-security/ to instapaper
202168/the-roger-tighe-as-round-wikileaks.html to instapaper
for/Wikileaks-hit-by-new-online-leakage-2111570.html to inst
.2034276-3,00.html to instapaper
/100607fa_fact_khatzadourian/fourcurrentPage=8 to instapaper
325865,500-how-to-leak-a-secret-and-not-get-caught.html to in
use-cyberdisidents-launch-Wikileaks-a-site-forwhistleblowers
urrent-blacklist-hacker-monitors-police/story-eftfrgpye-12257
ange.html to instapaper
board to instapaper
make-julien-assange-iraq-video?page=2 to instapaper
pi/www.scriamedia.edu.edu/article-eastasia.asp?parentid=6085
A.xag?parentid=60857 to instapaper
pi/www.nowpublic.com/wikileaks_and_untraceable_document_disc
nable_document_disclosure to instapaper
/100607fa_fact_khatzadourian/printable=true to instapaper
instapaper
akea_for_snooping_denial/ to instapaper
per
pi/www.scents.co.uk/scents/news.cfm?rit_id=143229af8Areal=w
o-1432293afAAreal-widgets.content_view_1 to instapaper
432314fpj1fyd1 to instapaper
ide-21642 to instapaper
3 to instapaper

tycentral/2011/jan/29/wikileaks-shut-down to instapaper
l-o-sony-the-economy-of-wikileaks/ to instapaper

8 to instapaper
7 to instapaper
366 to instapaper
/100607fa_fact_khatzadourian/currentPage=all to instapaper
/17/knight-foundation-hands-out-grants-to-32-groups-but-not-w
we/pnews.td477.3 to instapaper
Leaks_repairs/ to instapaper
html to instapaper
lywQZ8CQqpsD8GDTs_713qaeT24JfEJ_YFeQ to instapaper
hp to instapaper
exange.ted/ to instapaper
3.html to instapaper
oundation-sheds-light-on-wikileaks-donationsFixrsPtd0dxhds to
s-weee-grow-as-spokesman-quits-site/19651296?test=latestnews

#!/usr/bin/env python
# encoding: utf-8
'''
Add links to Instapaper.py

Simple script to add all urls referenced in root url to inst
BeautifulSoup. Install with
$ easy_install beautifulsoup

Author: Christine Corbett Moran
License: BSD
'''

import sys
import getopt
from BeautifulSoup import *
from sets import Set
import urllib
import urllib2

help_message = '''
./AddLinksToInstapaper.py -u username -p password page_url

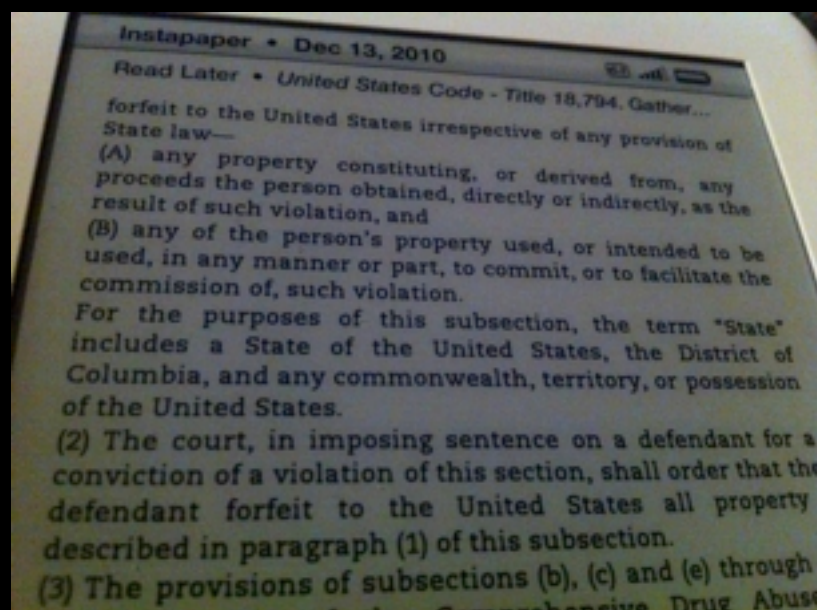
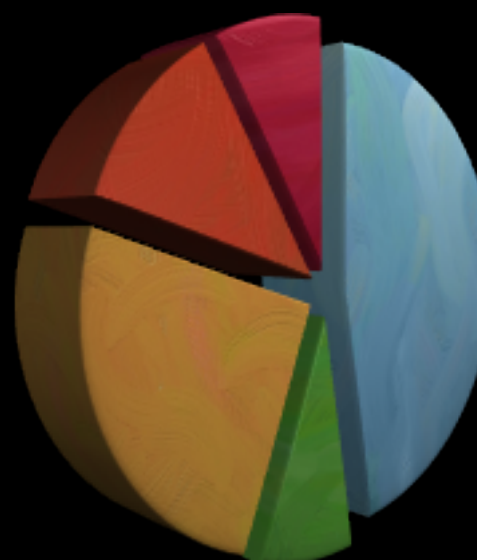
Adds all urls referenced on page_url to instapaper. If -w op
wikipedia urls are retained, if not they are ignored. The u
password option refer to instapaper username and optional pa
'''

class Usage(Exception):
    def __init__(self, msg):
        self.msg = msg

def add_links_to_instapaper(url,username,password=None,
                           wikilinks=False,verbose=False):
    def valid_link(link):
        return ('http' in link) and (wikilinks or not 'wiki'
INSTAPAPER_ADD=https://www.instapaper.com/api/add'
instapaper_params={'username':username}
if password:
    instapaper_params['password']=password
# spoofing user agent headers, some sites like Wikipedia
 opener.addheaders=[('User-agent', 'Mozilla/5.0')]
 wikipedia_page=BeautifulSoup(opener.open(url).read())
 all_links=Set()
 for link in wikipedia_page.findAll('a'):
     try:
         link=link['href']
         if not link in all_links and valid_link(link):
             instapaper_params['url']=link
             result=urllib.urlopen(INSTAPAPER_ADD+'%s'%
urllib.urlencode(instapaper_params))
             if verbose:
                 print "status code %s for adding %s to i
(result.read(),link)
                 all_links.add(link)
             except:
                 continue

def main(argv=None):
    if argv is None:
        argv = sys.argv
    try:
        try:
            opts,args=getopt.getopt(argv[1:], "hwu:p:")

```

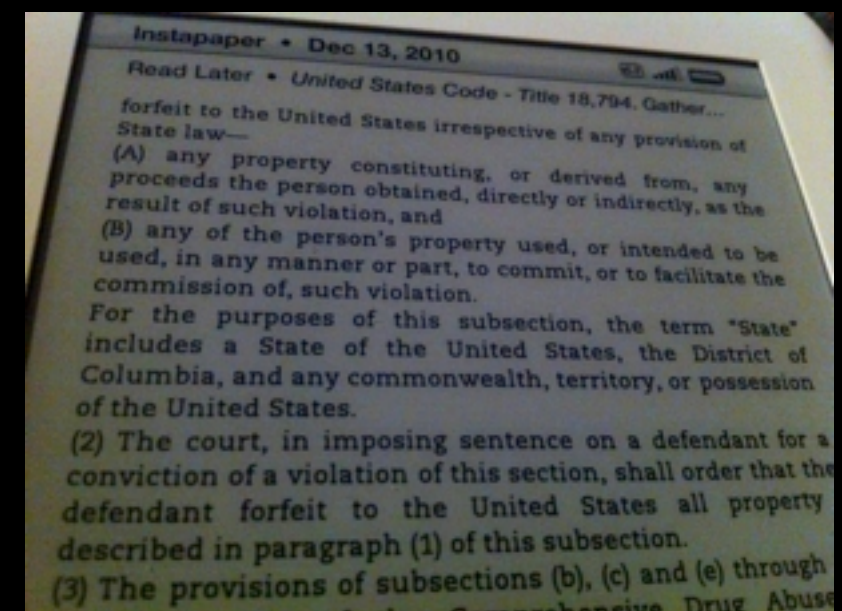
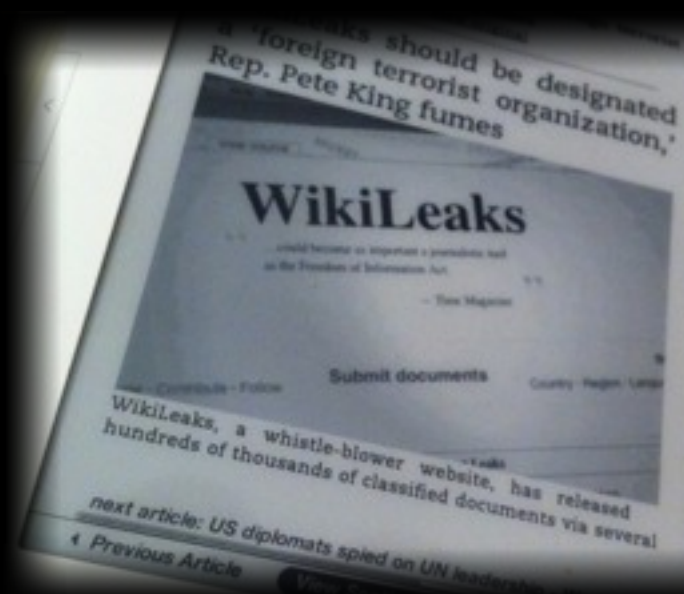
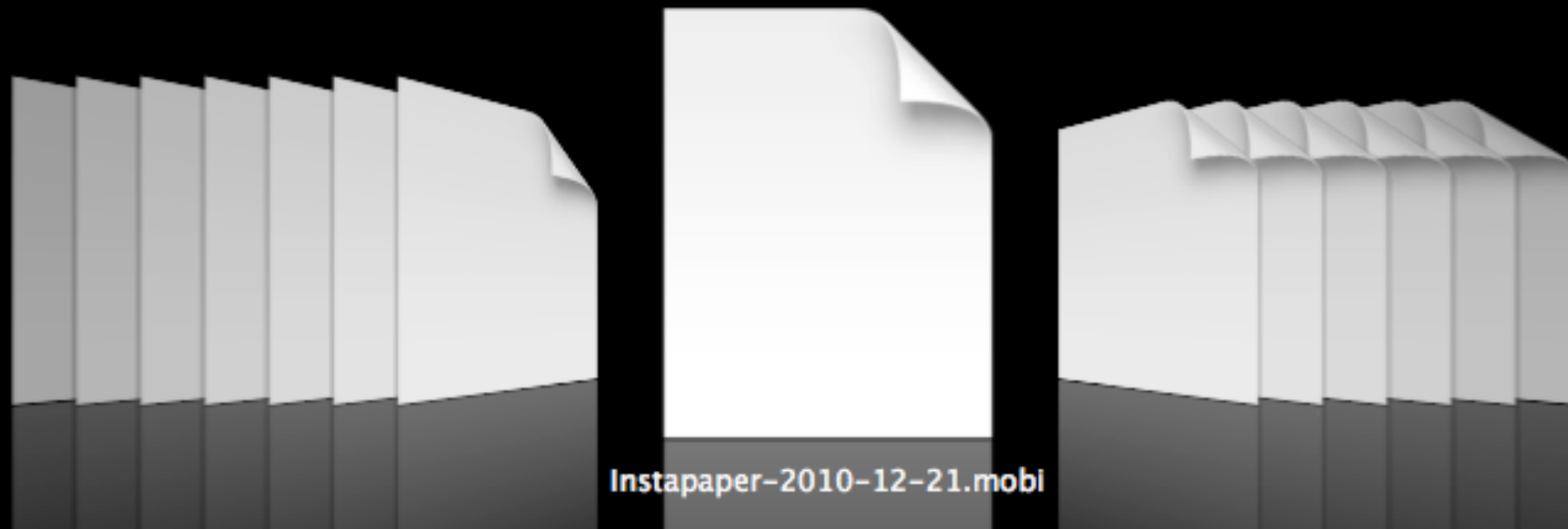



```

def add_links_to_instapaper(url, username, password=None, wikilinks=False):
    INSTAPAPER_ADD='https://www.instapaper.com/api/add'
    instapaper_params={'username':username}
    if password:
        instapaper_params['password']=password
    # spoofing user agent headers, some sites like Wikipedia require
    opener=urllib2.build_opener()
    opener.addheaders=[('User-agent', 'Mozilla/5.0')]
    original_page=BeautifulSoup(opener.open(url).read())
    all_links={}
    for link in original_page.findAll('a'):
        try:
            link=link['href']
            if not all_links.has_key(link) and valid_link(link, wikilinks):
                instapaper_params['url']=link
                result=urllib.urlopen('%s?%s' % \
                    (INSTAPAPER_ADD, urllib.urlencode(instapaper_params)))
                all_links[link]=True
        except Exception, e:
            continue

```

I. Obtain articles

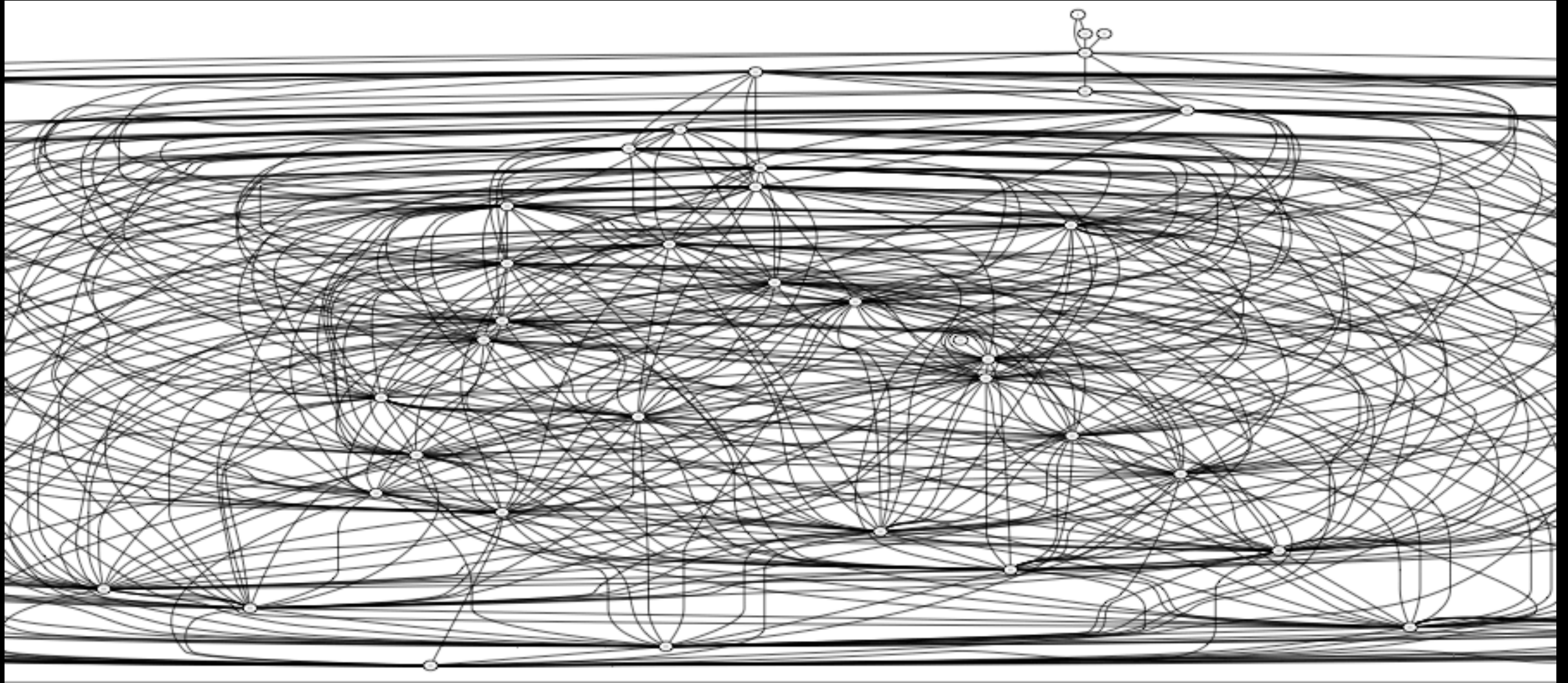


2. Read articles

Wikileaks Research Project - Citation classification and assorted notes, original URLs : Sheet 1






link	title	my notes/highlights: forgive typos they were typed quickly on the tiny Kindle keyboard	instapaper folder	cites original sources: key, o=is original, n=no, y=yes, e=extensively, p=partially, l=very limited, s*: cites secondary sources
http://zunguzungu.wordpress.com/2010/11/29/julian-assange-and-the-computer-conspiracy-%E2%80%9Cto-destroy-this-invisible-government%E2%80%9D/	Julian Assange and the Computer Conspiracy; "To destroy this invisible government" « zunguzungu	Read original essay first, think about it, then read the above post. http://cryptome.org/0002/ja-conspiracies.pdf Ja-conspiracies.pdf definitely on the top reading list.	Wikileaks reading list	y
http://news.cnet.com/8301-13578_3-20000469-38.html	8301-13578_3-20000469	CNET article--more valuable for its links to original sources and sections of the espionage act	Wikileaks reading list	y
http://www.firstamendmentcenter.org/analysis.aspx?id=17318	firstamendmentcenter.org analysis		Archive	y
http://www.fas.org/blog/secretcy/2010/06/wikileaks_review.html	Wikileaks Fails "Due Diligence" Review Secrecy News		Archive	y

3. Classify articles



4. Analyze relationships

Recreate

WikileaksResearchProject /			
name	age	message	history
 AddUrlsToInstapaper/	about 15 hours ago	adding urls to instapaper code in this reposito... [Christine Corbett Moran]	
 GraphImages/	about 15 hours ago	Python script to submit all urls referenced on ... [Christine Corbett Moran]	
 WikileaksKindle/	about 2 hours ago	latest version of analysis [Christine Corbett Moran]	
 WikileaksOriginalHtml/	about 15 hours ago	Python script to submit all urls referenced on ... [Christine Corbett Moran]	
 README	about 16 hours ago	first commit [Christine Corbett Moran]	

- <https://github.com/corbett/WikileaksResearchProject>
- contact:
 - **twitter** @corbett_inc
 - **email** corbett@physik.uzh.ch

Extend

- Named entities analysis and other tasks building on NLTK
- Graph-incorporate citation information, strength of connection
- Automatic citation type analysis
- Integrated script
- Level up: point to topic, tells you what to read to educate yourself