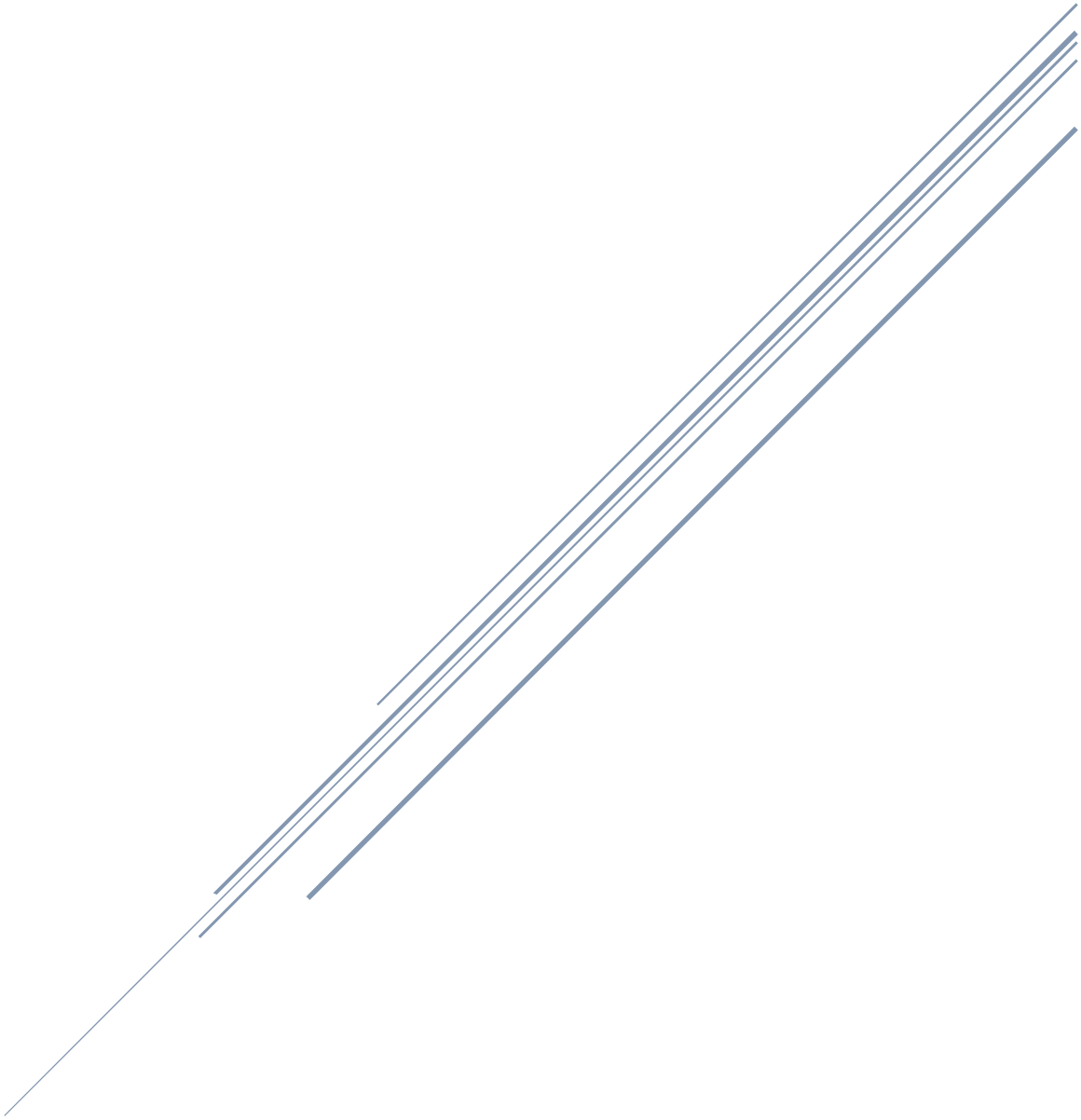
# WARC PORTAL V 1.0



# Table of Contents

# Features

## Login

1. Click the “Person” icon in the top right corner of the screen (beside visualizations)
2. Click **Login** in the drop down menu
3. On the login page, enter your username and password, and click **Log In**
    a. If your password is incorrect, the system will not allow you to log in
    b. If you are successful, you will be returned to the homepage
    c. If you do not have a username set up, please contact your system administrator
4. To log out when complete, click the “Person”

## Creating Collections

*Before you create a collection, ensure you are logged in.*

1. Click the “Person” icon in the top right corner of the screen (beside visualizations)
2. Click **Collections** in the drop down menu
3. On the create collections screen, there will be two columns
    a. The left hand column will list all collections currently existing in the system
    b. The right hand column will be a form to create a new collection
    c. To create a new collection:
        i. Give your collection a descriptive name: enter it into the “Name” box of the form
        ii. Click the “Files” box, and select a series of WARC files from the list to be added to this collection. You may select multiple files to be added. If you change your mind about a file, click the small blue **x** beside the file name in the “Files” box
            1. If your files are not present, the system is still processing them. Give the system time to process the files before continuing.
        iii. Click the **Add Collection** button. The page will refresh, and the collection will be added.
        iv. Click the WARC Portal logo to return to the homepage.

## Searching Archives

On the homepage, you will be presented with a paged list of all documents in the system.

1. To search for a specific term, enter that term in the search bar at the very top of the screen, and click the blue “Magnifying Glass”
2. The screen will reset and show documents matching your query

1. To filter results by criteria:
    a. On the left hand site of the screen, there will be a list of “Categories” and “Domains”
        i. Select one to many categories by clicking to restrict the types of documents that will appear in the results

ii. Select one to many domains by clicking to restrict the websites from which the results will appear

b. On the bar below the Documents/Images tabs, there will be an "Advanced Search" bar that you can use to further restrict results. In any case, stipulating a filter will refresh the results to reflect your choice.
    i. Click the "collection" box, and select a specific collection you would like to search through.
        1. Selecting a collection will restrict all of the results to pages only in that collection.
        2. Click the **x** beside the collection name to clear the collection selection.
    ii. Click the "Publish Date" box, and select a specific "From" and "To" date that you would like to search within
        1. Selecting a range (or endpoint) for these dates will restrict all of the results to documents only published in that date range.
        2. Click the **x** near the top of the box to clear the date.
    iii. Click the "Crawl Date" box, and select a specific "From" and "To" date that you would like to search within
        1. Selecting a range (or endpoint) for these dates will restrict all of the results to documents only crawled in that date range.
        2. Click the **x** near the top of the box to clear the date.
c. For pages:
    i. Scroll to the bottom to see the page selection toolbar:
        1. You may select a specific page by clicking on it
        2. You may advance one page by clicking the **>** button
        3. You may advance to the last page by clicking the **>>** button
        4. You may go back one page by clicking the **<** button
        5. You may go back to the first page by clicking the **<<** button

## Searching Images

On the homepage, click the "images" tab on the top toolbar. Clicking this will take you to a page showing all images available in the system.

1. To search for a specific term, enter that term in the search bar at the very top of the screen, and click the blue "Magnifying Glass"
2. The screen will reset and show images matching your query

1. To filter results by criteria:
    a. On the bar below the Documents/Images tabs, there will be an "Advanced Search" bar that you can use to further restrict results. In any case, stipulating a filter will refresh the results to reflect your choice.
        i. Click the "collection" box, and select a specific collection you would like to search through.

1. Selecting a collection will restrict all of the results to images only in that collection.
2. Click the **x** beside the collection name to clear the collection selection.

ii. Click the "Publish Date" box, and select a specific "From" and "To" date that you would like to search within
1. Selecting a range (or endpoint) for these dates will restrict all of the results to images only published in that date range.
2. Click the **x** near the top of the box to clear the date.

iii. Click the "Crawl Date" box, and select a specific "From" and "To" date that you would like to search within
1. Selecting a range (or endpoint) for these dates will restrict all of the results to images only crawled in that date range.
2. Click the **x** near the top of the box to clear the date.

b. For pages:

i. Scroll to the bottom to see the page selection toolbar:
1. You may select a specific page by clicking on it
2. You may advance one page by clicking the **>** button
3. You may advance to the last page by clicking the **>>** button
4. You may go back one page by clicking the **<** button
5. You may go back to the first page by clicking the **<<** button

## TF-IDF Functionality

TF-IDF: In information retrieval, tf–idf, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval and text mining. The tf-idf value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. (https://en.wikipedia.org/wiki/Tf%E2%80%93idf)

1. On the homepage, click the "Visualizations" button, which will give you a menu, on which you should click "TF-IDF"
2. You will be taken to the TF-IDF page, which will present you with two drop-down selections boxes, one for the "Collection", and one for the "Document"
    a. Click the Collection box, and select a collection from the list that you would like to view an analysis on
    b. Click the Document box, and select a document from the list (of documents in that collection) that you would like to view an analysis on
    c. The graph will appear in the box below. You can make changes to collection or document to view a different graph.
    d. If you would like to restart, you may click the **x** beside the document or collection name to remove your selection.
    e. Hover over a bar of the word ranking to see that word's ranking.

## Word Cloud Functionality

1. On the homepage, click the “Visualizations” button, which will give you a menu, on which you should click “Word Cloud”
2. You will be taken to the Word Cloud page, which will present you with two drop-down selections boxes, one for the “Collection”, and one for the “Document”
    a. Click the Collection box, and select a collection from the list that you would like to view an analysis on
    b. Click the Document box, and select a document from the list (of documents in that collection) that you would like to view an analysis on.
    c. The graphic will appear in the box below. You can make changes to collection or document to view a different graph.
    d. If you would like to restart, you may click the **x** beside the document or collection name to remove your selection.

# License

## WARC Portal



## Other Packages

### WARCBASE



### Pywb

This program comes with ABSOLUTELY NO WARRANTY
This is free software, and you are welcome to redistribute it
under certain conditions

## D3

Copyright (c) 2014-2015 Eric. S Bullington, Lim Yang Wei, and project contributors

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

## REACT

BSD License

For React software

Copyright (c) 2013-present, Facebook, Inc.
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

* Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

* Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

* Neither the name Facebook nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

## Django

Copyright (c) Django Software Foundation and individual contributors.
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of Django nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS

SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

## POSTGRESQL

PostgreSQL Database Management System
(formerly known as Postgres, then as Postgres95)