WASP: Web Archiving and Search Personalized

Johannes Kiesel, Arjen P. de Vries, Matthias Hagen, Benno Stein and Martin Potthast
@KieselJohannes, @arjenpdeVries, @matthias_hagen, @bennostein, @martinpotthast

DESIRES, August 29th 2018
The Personal Search Engine: Motivation
The Personal Search Engine: Motivation
The Personal Search Engine: Motivation
The Personal Search Engine: Motivation
The Personal Search Engine: Motivation

Getting Started Training - Wasp Barcode
https://www.waspbarcode.com/training/getting-started-training
Available for MobileAsset, WaspTime and InventoryControl, training covers basics including installation, database setup, and overview of key features and icons.

Videos

1. WASP Basic Install and basic features Demo
   - waspAnalytics
   - YouTube - Jul 1, 2010

2. Unboxing and setup of the Wasp Delta 20X40
   - Simone | FNTSMN
   - YouTube - Feb 21, 2016

3. How to set up the WHY Trap for Wasps, Hornets & Yellowjackets
   - WhyRescue
   - YouTube - Jul 15, 2017

[PDF] Quick Start Guide - Wasp Barcode
https://www.waspbarcode.com/.../waspbarcode/.../mobile-asset-quick-start-guide.ashx?
without the written permission of Wasp Barcode Technologies, ... you through the installation process and help you set up your software. Most of the information ...
The Personal Search Engine: Motivation

GitHub - webis-de/wasp
https://github.com/webis-de/wasp

WASP: WASP is a fully functional prototype of a personal web archive and search system. Quickstart: You need to have Docker installed on your system.
The Personal Search Engine: Motivation

```
match setup visited: last week
```

GitHub - pakmor/wasp
https://github.com/pakmor/wasp

WASP. WASP is an application for the archiving and search system. Quickstart. You need to have Python installed on your system.
The Personal Search Engine: Motivation

Only you can see this data. Google protects your privacy and security. Learn more

Today

Some activity may not appear yet

ITEMS | SEARCH | YOUTUBE | ADS | HELP
--- | --- | --- | --- | ---
21 | | | | |

11:12 PM

support.google.com

Visited See & control your search activity - Computer - Google Search Help
The Personal Search Engine: Motivation

Search history:
- *wasp setup*
- Filter by date & product

Activity:
- *Visited GitHub - webis-de/wasp*
  - Details • Search
This is not the web page you are looking for.
The Personal Search Engine: Inspiration
The Personal Search Engine: Inspiration

Personal search engine!
WASP

WASP is a fully functional prototype of a personal web archive and search system.

Quickstart

You need to have Docker installed on your system. You can then immediately start WASP like this:

```
docker run -p 8001:8001 -p 8002:8002 -p 8003:8003 --name wasp -d webis/wasp:0.2.0
```

Depending on your Docker setup, you might need to run this command as an administrator or add `sudo` in front of it.

After a few seconds, you should already see the search interface in your browser at (http://localhost:8003/search)

Next, you will have to change the proxy settings of your browser to route the requests and responses through it. How this can be done depends on the browser (Chrome, Edge, Firefox, Opera, Safari). However, also consider using a tailored proxy switching plugin for your browser, which allows you to disable/enable WASP easily. In all cases, you need to specify (localhost:8001) as proxy for HTTP and HTTPS and have "localhost" as an exception (often so by default).

Now it is time to clear your browser's cache (or use incognito mode). Otherwise you will be disappointed when some elements are missing in the archive.
### Web Archive Search Personalized

<table>
<thead>
<tr>
<th>From:</th>
<th>beginning</th>
<th>month ago</th>
<th>week ago</th>
<th>day ago</th>
<th>beginning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Until:</td>
<td>month ago</td>
<td>week ago</td>
<td>day ago</td>
<td>now</td>
<td>now</td>
</tr>
</tbody>
</table>

**Query**

- **Go!**
WASP, is a fully functional prototype of a personal web archive and search system. Depending on your Docker setup, you might need to run this command. You are already set up to use WASP for HTTP connections. For HTTPS connections, you have to trust WASP to identify foreign web pages for you.

GitHub - webis-de/wasp [archive] [live]
https://github.com/webis-de/wasp

WASP, is a fully functional prototype of a personal web archive and search system. Depending on your Docker setup, you might need to run this command. You are already set up to use WASP for HTTP connections. For HTTPS connections, you have to trust WASP to identify foreign web pages for you.

wasp/README.md at master · webis-de/wasp · GitHub [archive] [live]
https://github.com/webis-de/wasp/...-name wasp -d webis/wasp:0.2.0 Depending on your Docker setup, you might need to run this command. You are already set up to use WASP for HTTP connections. For HTTPS connections, you have to trust WASP to identify foreign web pages for you.

GitHub - webis-de/wasp [archive] [live]
WASP

WASP is a fully functional prototype of a personal web archive and search system.

Quickstart

You need to have Docker installed on your system. You can then immediately start WASP like this:

```
docker run -p 8001:8001 -p 8002:8002 -p 8003:8003 --name wasp -d webis/wasp:0.2.0
```

Depending on your Docker setup, you might need to run this command as an administrator or add `sudo` in front of it.

After a few seconds, you should already see the search interface in your browser at (http://localhost:8003/search)

Next, you will have to change the proxy settings of your browser to route the requests and responses through it. How this can be done depends on the browser (Chrome, Edge, Firefox, Opera, Safari). However, also consider using a tailored proxy switching plugin for your browser, which allows you to disable/enable WASP easily. In all cases, you need to specify (localhost:8001) as proxy for HTTP and HTTPS and have "localhost" as an exception (often so by default).

Now it is time to clear your browser’s cache (or use incognito mode). Otherwise you will be disappointed when some elements are missing in the archive.
WASP

Search Interface

pywb

Index

WARCs

warcprox

World Wide Web

Browser
All requests (➡️) and responses (⬅️) while browsing are stored and indexed.
All requests (→) and responses (←) while browsing are stored and indexed.

Page on localhost allows to search. Result page links to archive where...
- All requests (→) and responses (←) while browsing are stored and indexed
- Page on localhost allows to search. Result page links to archive where...
- Visited pages are reproduced for the corresponding time
Personal search engine!
Insight 1: Not Indexing Near-duplicates

- What changes warrant a re-archiving?
Insight 2: Browsable and Deletable History

Web Archive Search Personalized

From: beginning  month ago  week ago  day ago  beginning
Until: month ago  week ago  day ago  now  now
Query

April 23rd, 2018
Insight 3: Easy (De-)activation of Archiving

Easy activation, deactivation, and status-check

Patterns

<table>
<thead>
<tr>
<th>Name</th>
<th>Pattern</th>
<th>Type</th>
<th>http(s)</th>
<th>On/Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>all URLs</td>
<td>*</td>
<td>wildcard</td>
<td>both</td>
<td>on</td>
</tr>
</tbody>
</table>

White Patterns

Black Patterns
Insight 4: Combined Indexing of Sub-pages

- Should visited sub-pages of a single article be indexed as one?
- If so, to which sub-page should be linked in the result list?
Insight 5: Personalized Search

Google search for "desires 2018"

About 145,000,000 results (0.41 seconds)

DESIRES 2018
desires.dei.unipd.it/

Vision. A systems-oriented biennial conference, complementary in its mission to the mainstream Information Access and Retrieval conferences, emphasizing the ...

You've visited this page many times. Last visit: 8/23/18
Insight 5: Personalized Search

You've visited this page many times. Last visit: 8/23/18
WASP: Web Archiving and Search Personalized

Insights overview

- Not Indexing Near-duplicates
- Browsable and Deletable History
- Easy (De-)activation of Archiving
- Combined Indexing of Sub-pages
- Personalized Search

Code and Instructions on Github

github.com/webis-de/wasp

Thank you for your attention!