IANCIS

Indexing of Anonymous Networks for Crime Information Search

Alessandro Celestini
Institute for Applied Computing (IAC-CNR)
a.celestini@iac.cnr.it

May 20, 2015

IANCIS

Objective: Developing a tool able to crawl onion websites and feed a semantic engine for indexing and clustering collected data.

Website: www.iancis.eu

Consortium:

Istituto per le Applicazioni del Calcolo "Mauro Picone”

Arma dei Carabinieri HQs-ICT Office

Expert System

IANCIS

The tool should:

- Automatically **explore** TOR hidden services
- **Extract** text from collected resources and **analyze** it
- **Visualize** analysis results and **identify** evidences of illegal activities

The tool is meant to be the first step towards the development of an investigation instrument for the TOR network

TOR

Tor is an **anonymity network**, it was originally designed, implemented, and deployed as a third-generation **onion routing** project of the U.S. Naval Research Laboratory.

Tor network is composed by volunteer-operated servers (relays) that allow people to improve their privacy and security on the Internet.

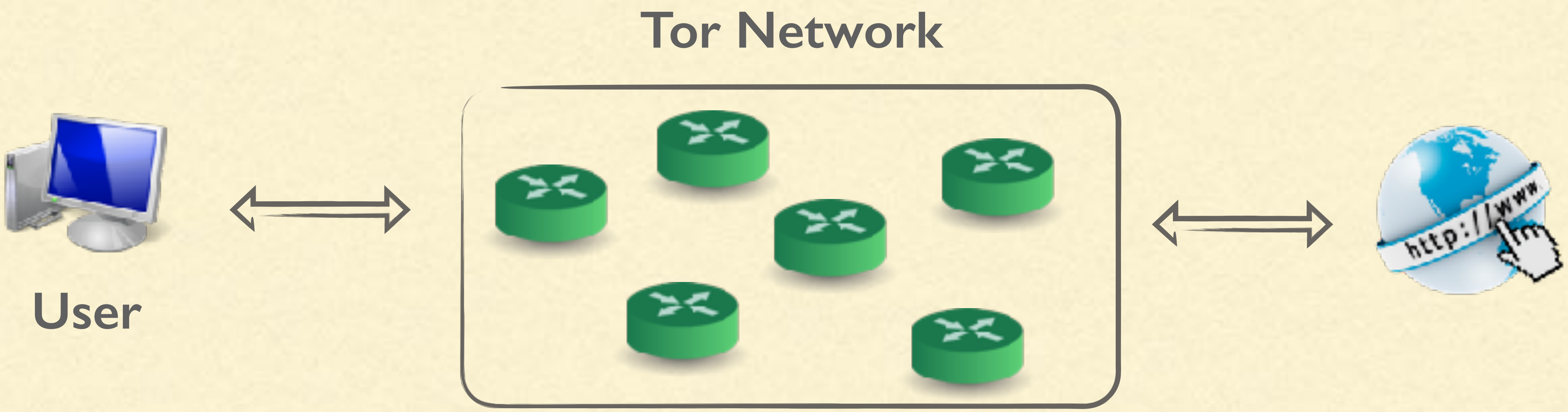
Tor protects users against a common form of Internet surveillance known as **"traffic analysis"**. Traffic analysis can be used to infer **who is talking to whom** over a public network.

TOR

Tor is an **anonymity network**, it was originally designed, implemented, and deployed as a third-generation **onion routing** project of the U.S. Naval Research Laboratory.

Tor network is composed by volunteer-operated servers (relays) that allow people to improve their privacy and security on the Internet.

Tor protects users against a common form of Internet surveillance known as **"traffic analysis”**. Traffic analysis can be used to infer **who is talking to whom** over a public network.



Tor Network

User

Hidden Services

Tor can provide **receiver privacy** for Internet services through a feature called “hidden services”.

Tor’s hidden services let users publish web sites and other services **without revealing the location of the site.**

Hidden Services

Tor can provide **receiver privacy** for Internet services through a feature called "hidden services".

Tor's hidden services let users publish web sites and other services **without revealing the location of the site.**

Hidden services are accessible only through the Tor network and each service is identified by its ***onion address.*** An onion address is a **16 characters** string followed by the domain .onion, e.g http://duskgytldkxiuqc6.onion.

The number of Hidden Services is estimated to be in the order of **tens of thousands**[1,2]

[1] Torproject Blog, "Some statistics about onions", https://blog.torproject.org/blog/some-statistics-about-onions
[2] A. Biryukov, I. Pustogarov, and R. Weinmann. 2013. Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization. In Proceedings of SP '13

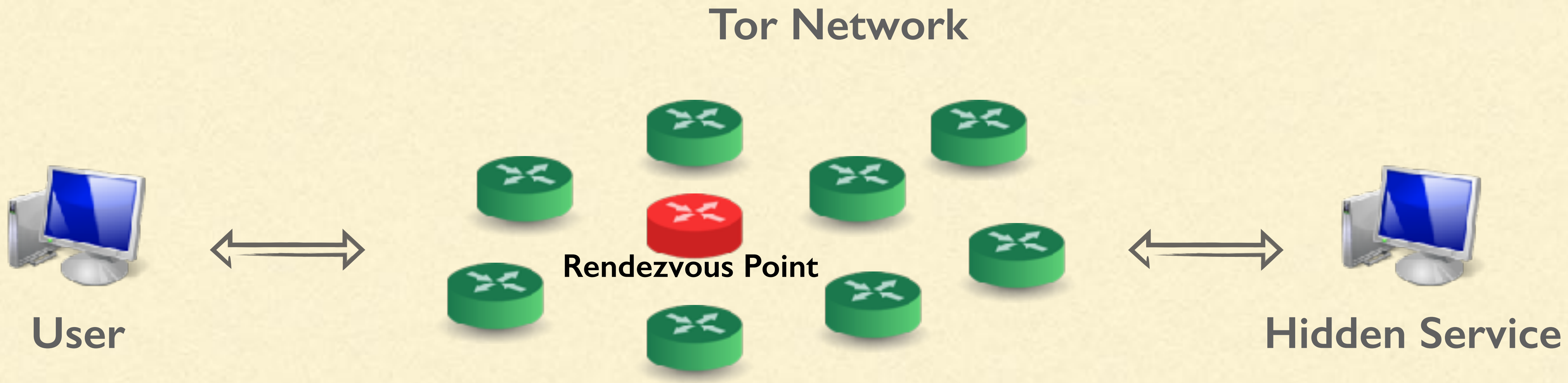
Hidden Services

Tor can provide **receiver privacy** for Internet services through a feature called "hidden services".

Tor's hidden services let users publish web sites and other services **without revealing the location of the site.**

Hidden services are accessible only through the Tor network and each service is identified by its ***onion address.*** An onion address is a **16 characters** string followed by the domain .onion, e.g http://duskgytldkxiuqc6.onion.

The number of Hidden Services is estimated to be in the order of **tens of thousands**[1,2]



[1] Torproject Blog, "Some statistics about onions", https://blog.torproject.org/blog/some-statistics-about-onions
[2] A. Biryukov, I. Pustogarov, and R. Weinmann. 2013. Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization. In Proceedings of SP '13

Hidden Services

Why LEA are interested in Hidden Services?

Several Hidden Services **promote illegal activities** such as:

- Illicit trafficking in weapons
- Illicit trafficking in narcotic drugs
- Child pornography
- Trafficking in human beings
- Terrorism

Hidden Services

Why LEA are interested in Hidden Services?

Several Hidden Services **promote illegal activities** such as:

- Illicit trafficking in weapons
- Illicit trafficking in narcotic drugs
- Child pornography
- Trafficking in human beings
- Terrorism

Silk Road was an online marketplace
mostly used for selling illegal goods.
In 2013 the FBI shut down the website.

Hidden Services

Why LEA are interested in Hidden Services?

Several Hidden Services **promote illegal activities** such as:

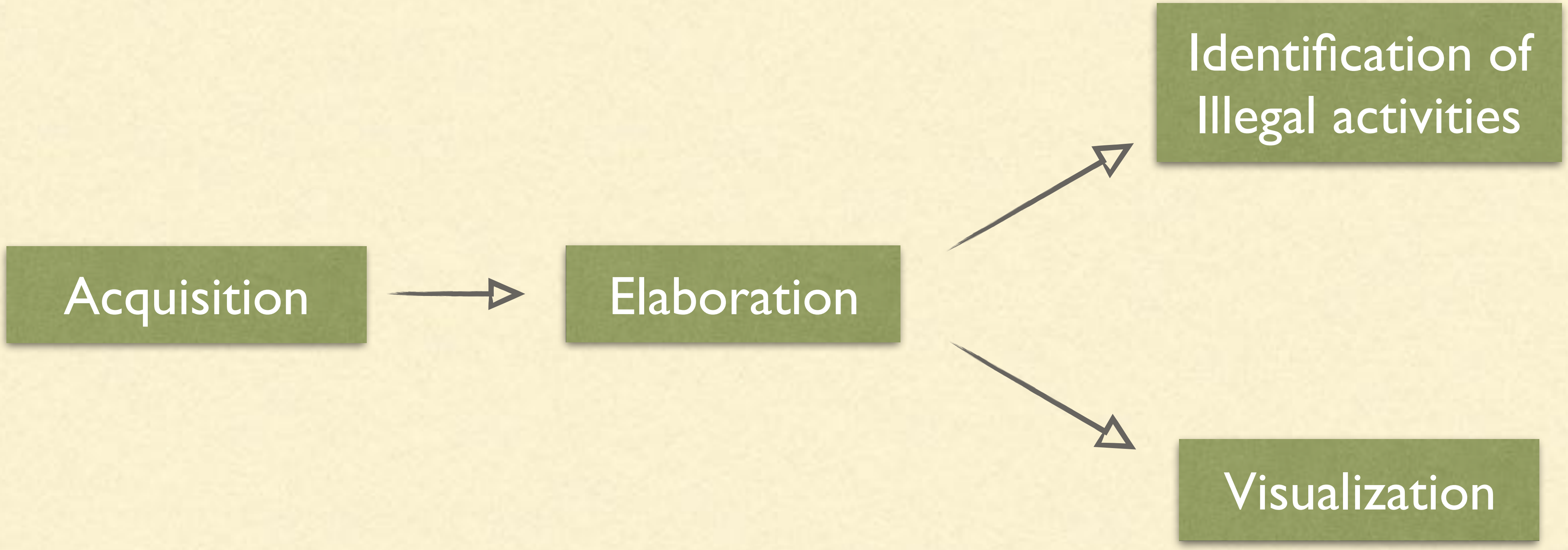
- Illicit trafficking in weapons
- Illicit trafficking in narcotic drugs
- Child pornography
- Trafficking in human beings
- Terrorism

Silk Road was an online marketplace mostly used for selling illegal goods.
In 2013 the FBI shut down the website.

Lolita City hosted child pornography images and videos.

Tool

Workflow



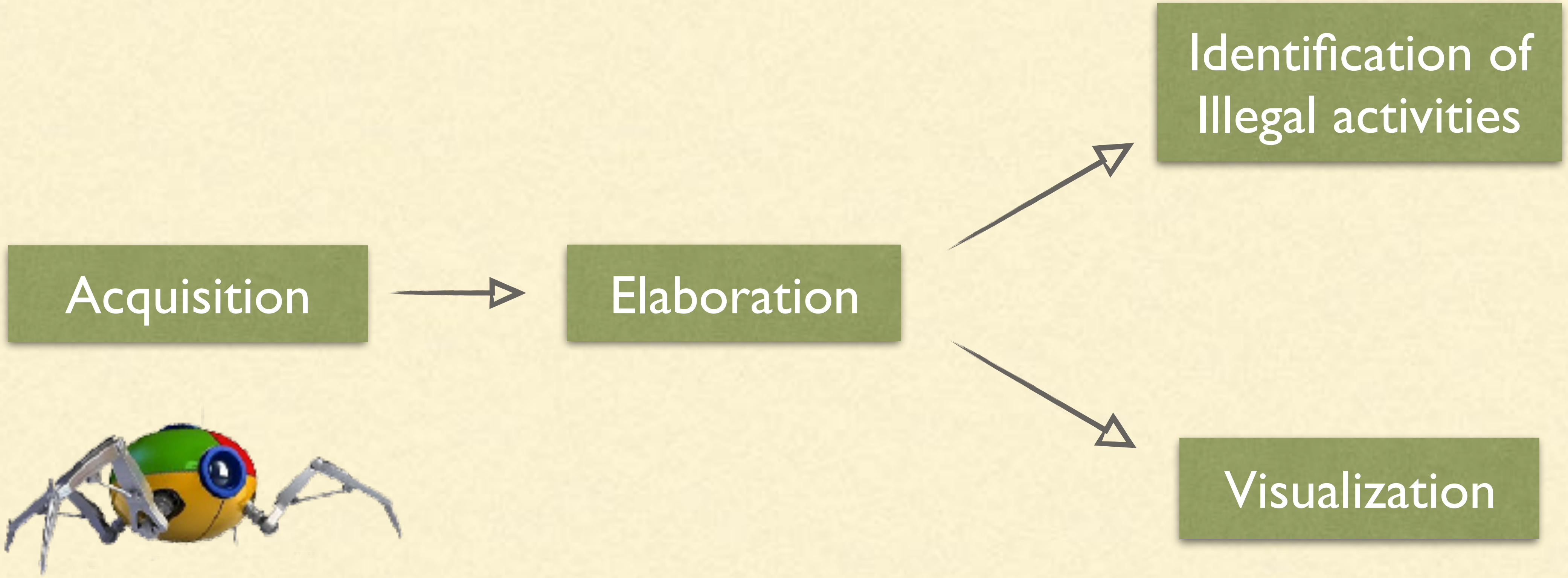
Acquisition

Elaboration

Identification of Illegal activities

Visualization

Workflow



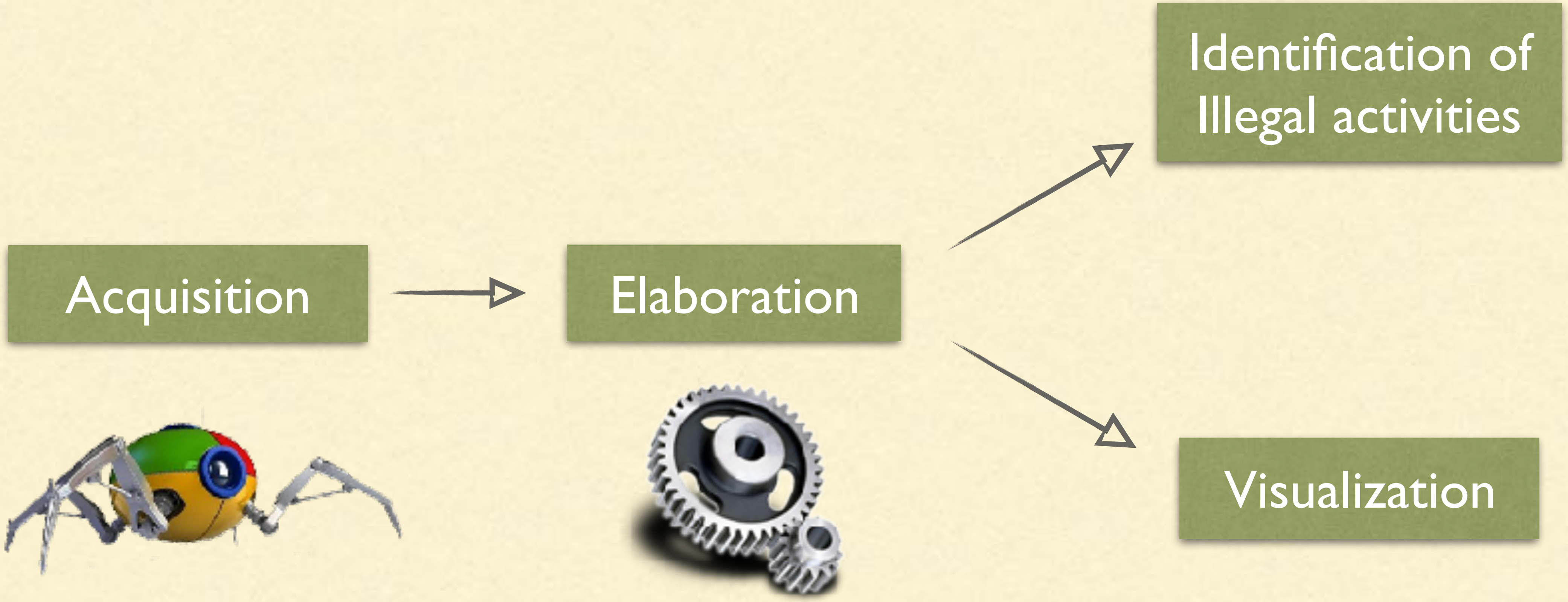
Acquisition

Elaboration

Identification of Illegal activities

Visualization

Workflow



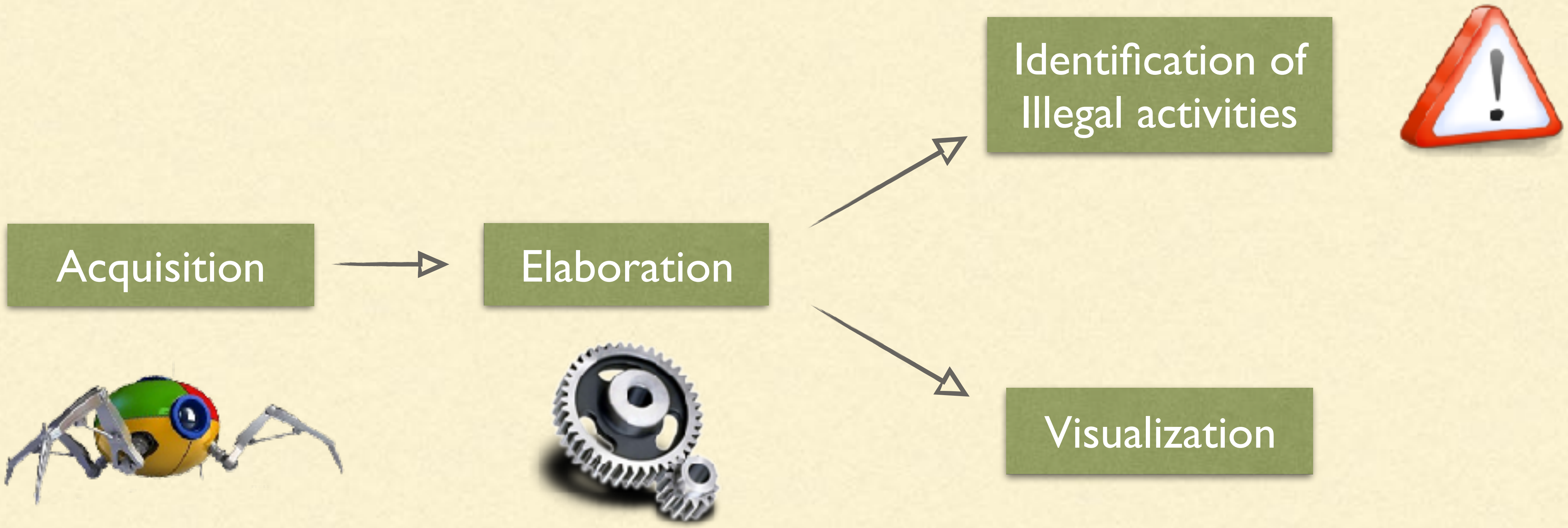
Acquisition

Elaboration

Identification of Illegal activities

Visualization

Workflow



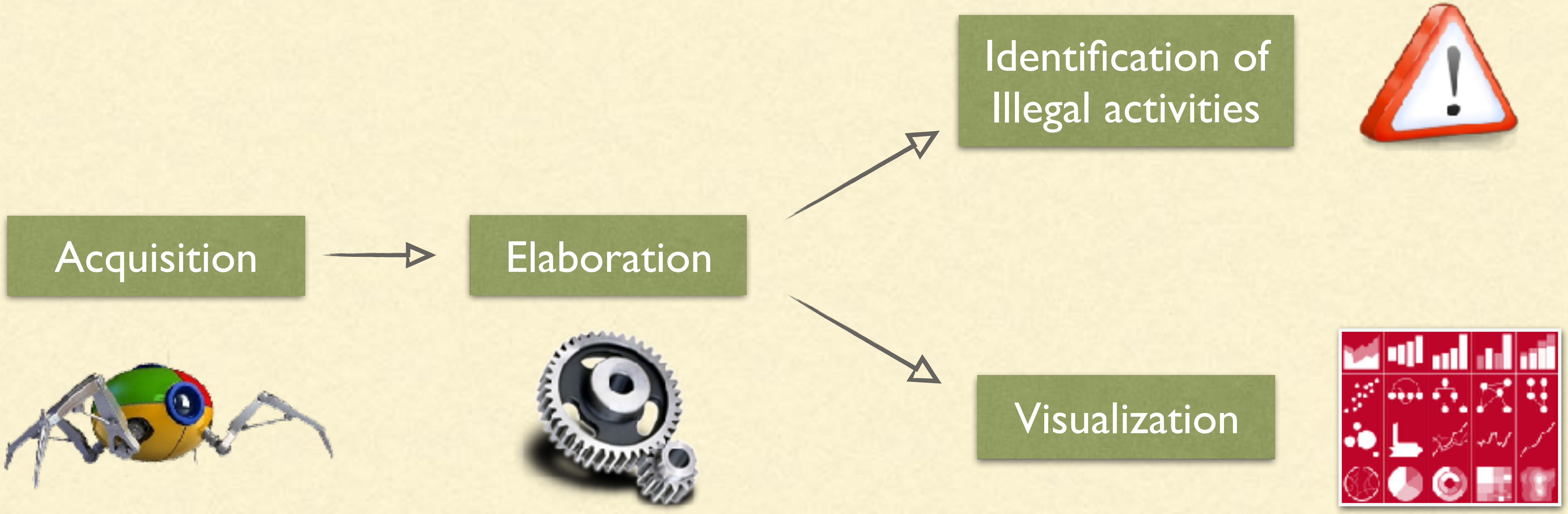
Acquisition

Elaboration

Identification of Illegal activities

Visualization

Workflow



Acquisition

Elaboration

Identification of Illegal activities

Visualization

Technologies

ACQUISITION

ELABORATION **Text**

ELABORATION **Text**

Technologies

ACQUISITION

BUBING: high performance crawler developed by the Laboratory for Web Algorithmics (LAW) of the University of Milan

ELABORATION **Text**

ELABORATION **Text**

Technologies

ACQUISITION

BUBING: high performance crawler developed by the Laboratory for Web Algorithmics (LAW) of the University of Milan

ELABORATION **Text**

TIKA: open-source software suite for the identification and extraction of text from more than 1000 different file types

ELABORATION **Text**

Technologies

ACQUISITION

BUBING: high performance crawler developed by the Laboratory for Web Algorithmics (LAW) of the University of Milan

ELABORATION **Text**

Apache Tika

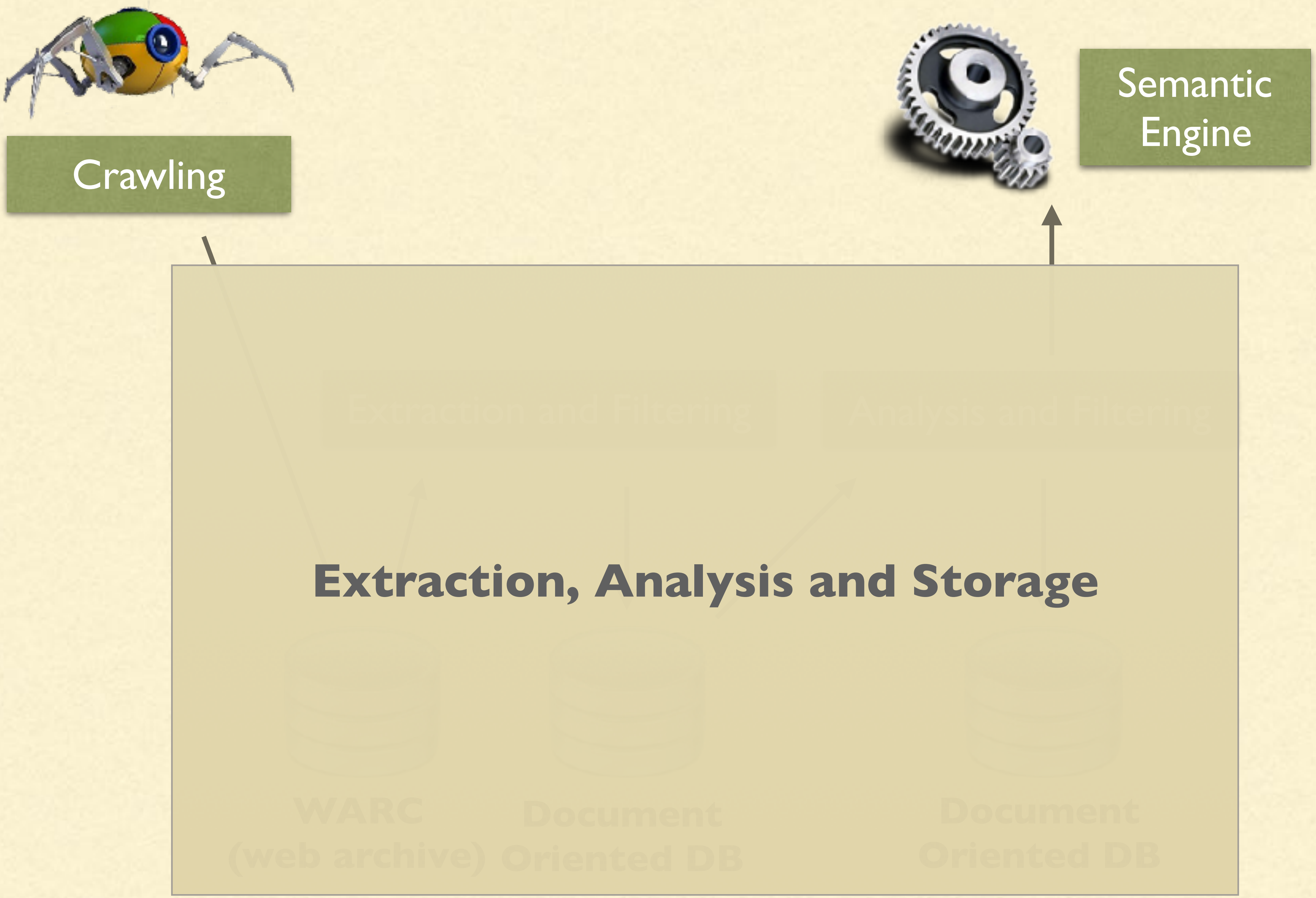
TIKA: open-source software suite for the identification and extraction of text from more than 1000 different file types

ELABORATION **Text**

COGITO

COGITO: semantical analysis engine by Expert System, that classifies a text according to a suitable taxonomy, providing both quantitative and qualitative information (*what topic* the text is about and *how much* the text discusses such topic)

Architecture



Crawling

Semantic Engine

Extraction, Analysis and Storage

Extraction and Filtering

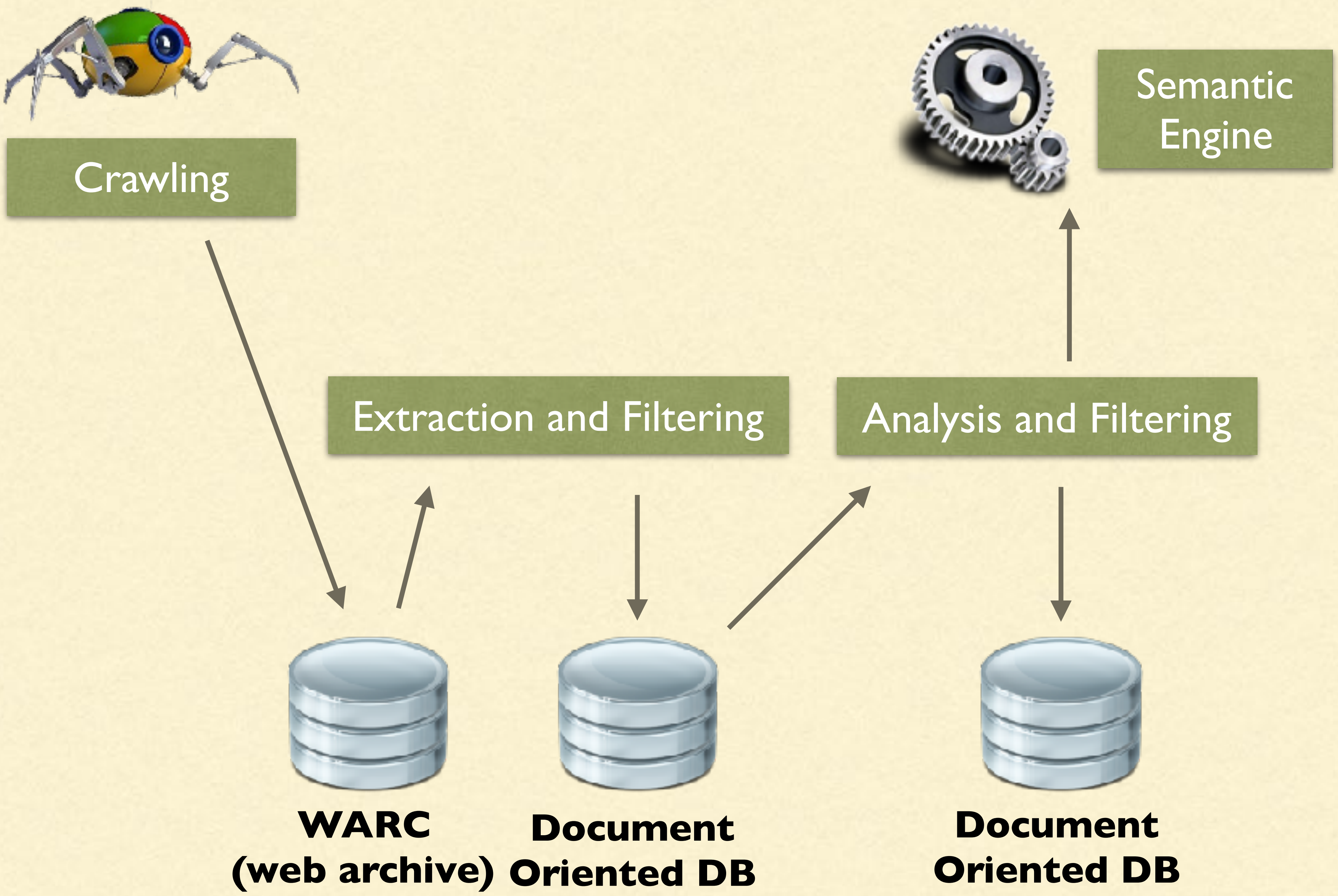
Analysis and Filtering

WARC (web archive)

Document Oriented DB

Document Oriented DB

Architecture



Crawling

Semantic Engine

Extraction and Filtering

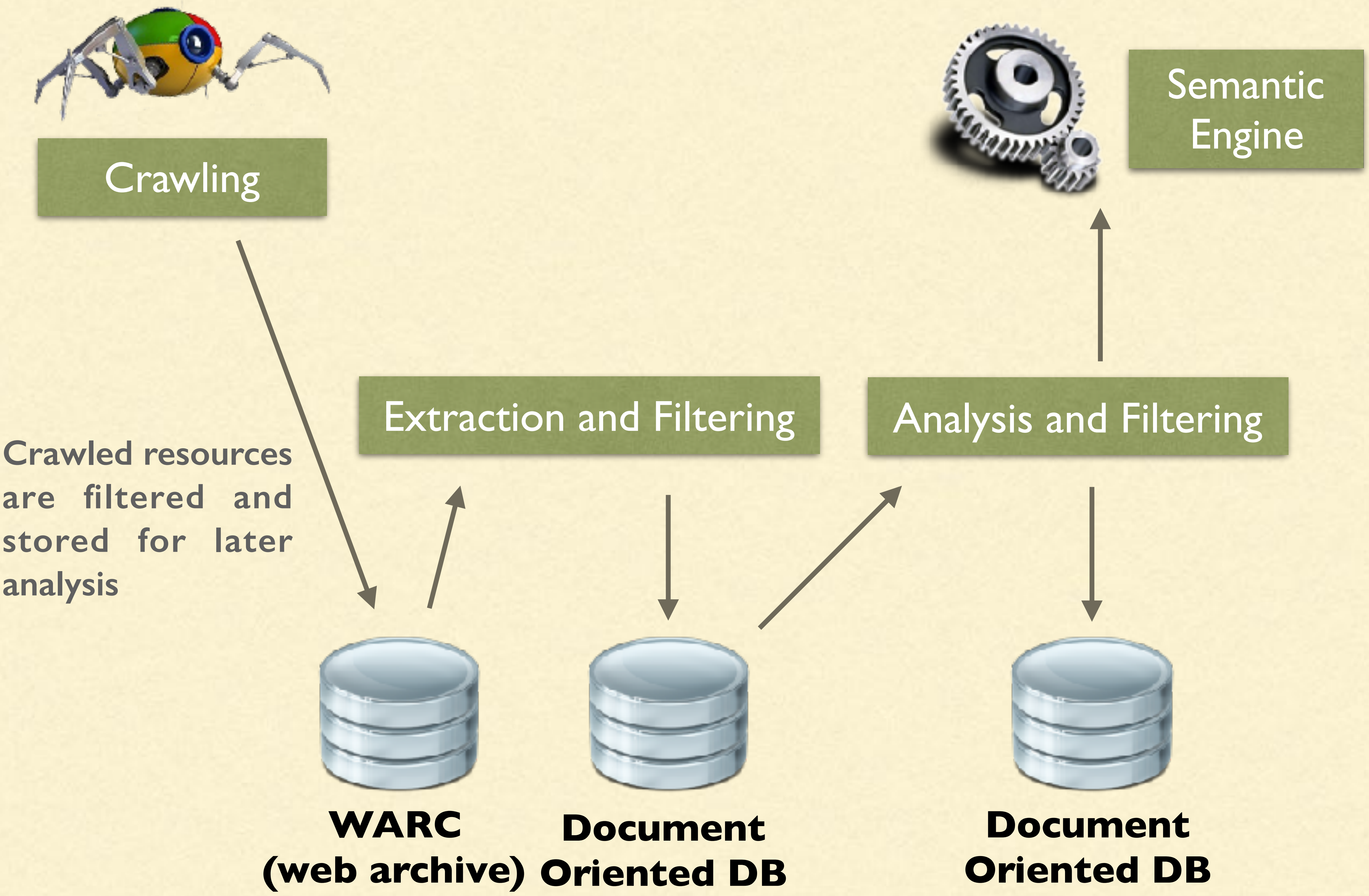
Analysis and Filtering

WARC (web archive)

Document Oriented DB

Document Oriented DB

Architecture



Crawling

Semantic Engine

Extraction and Filtering

Analysis and Filtering

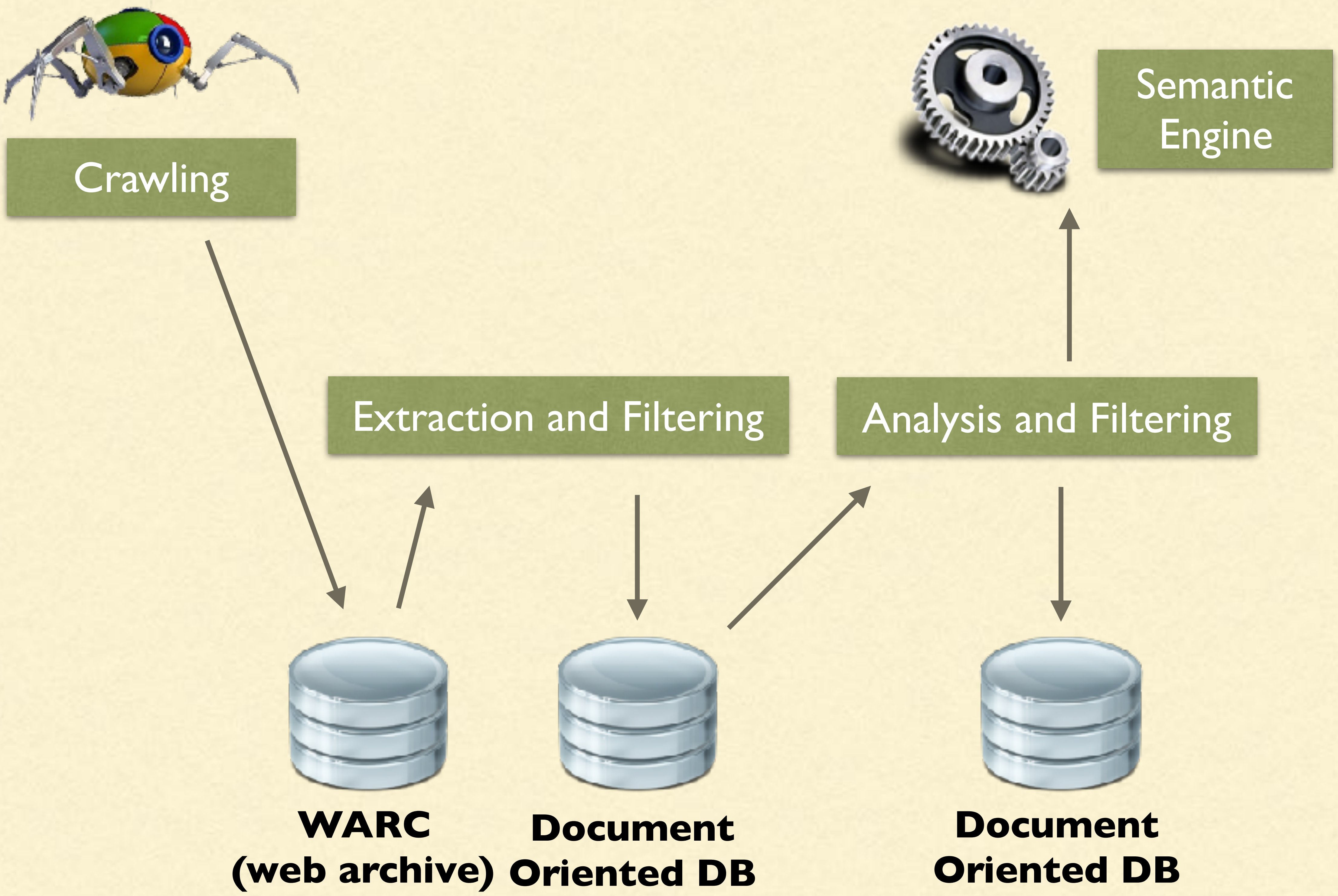
Crawled resources are filtered and stored for later analysis

WARC
(web archive)

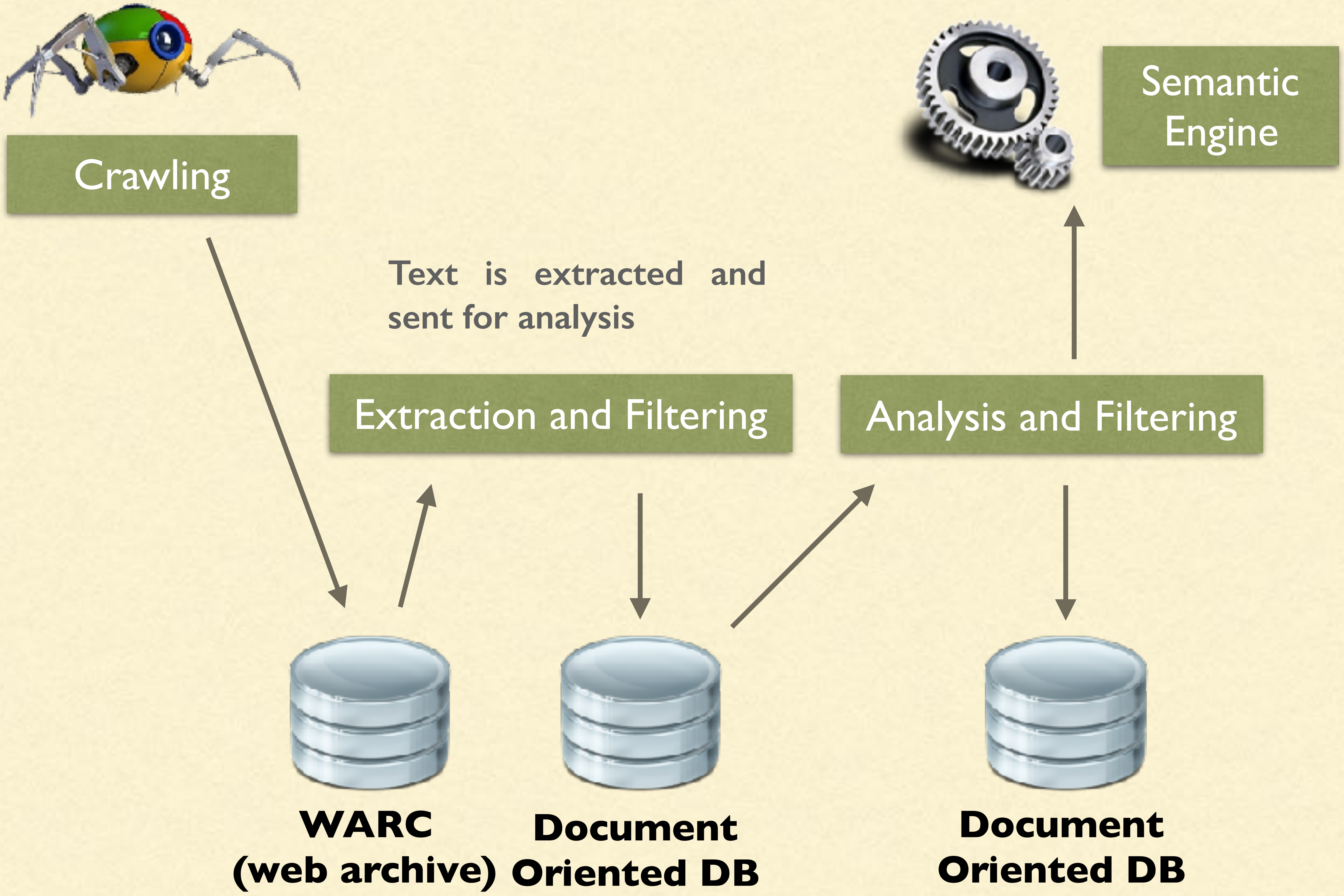
Document
Oriented DB

Document
Oriented DB

Architecture



Architecture



Crawling

Semantic Engine

Text is extracted and sent for analysis

Extraction and Filtering

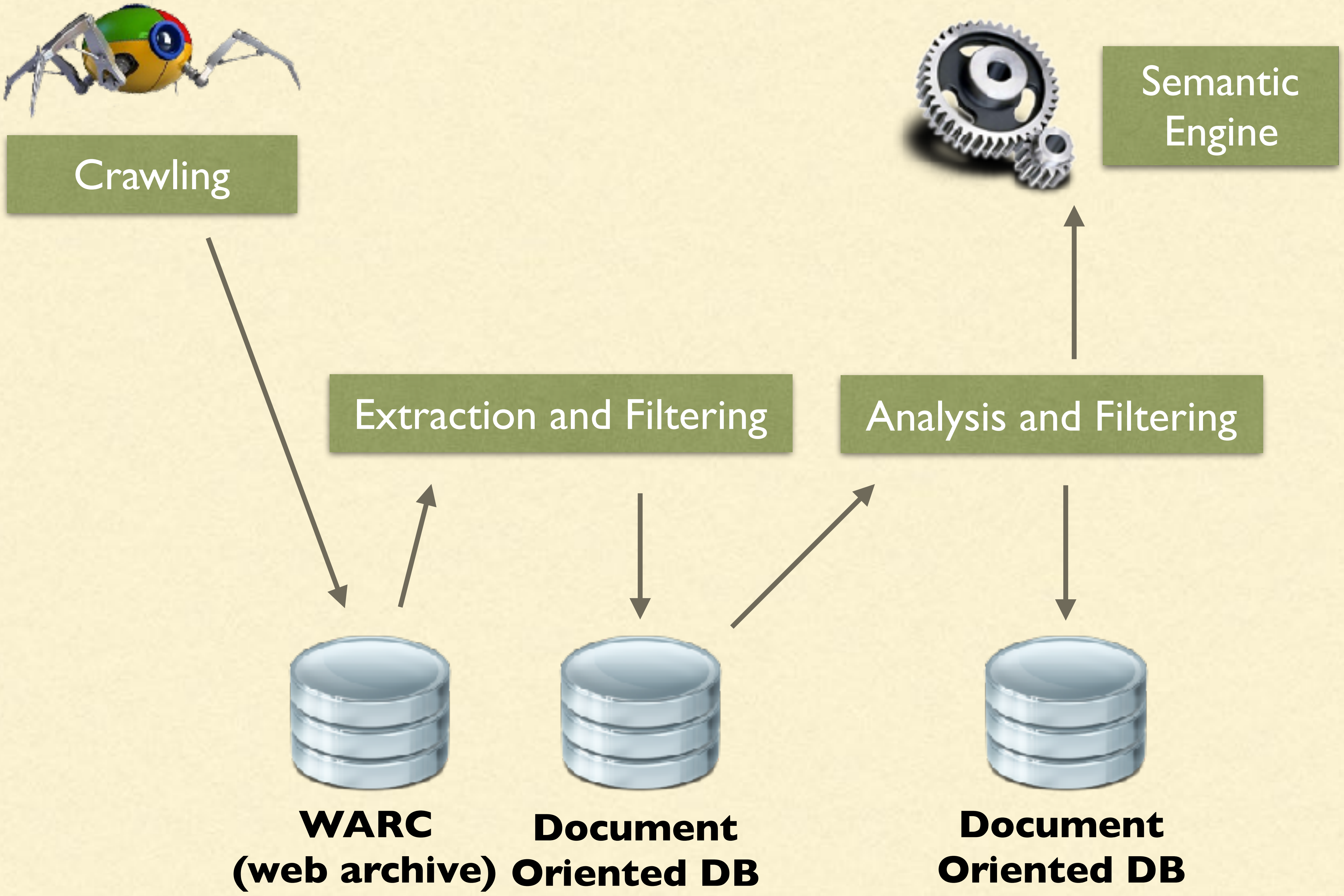
Analysis and Filtering

WARC (web archive)

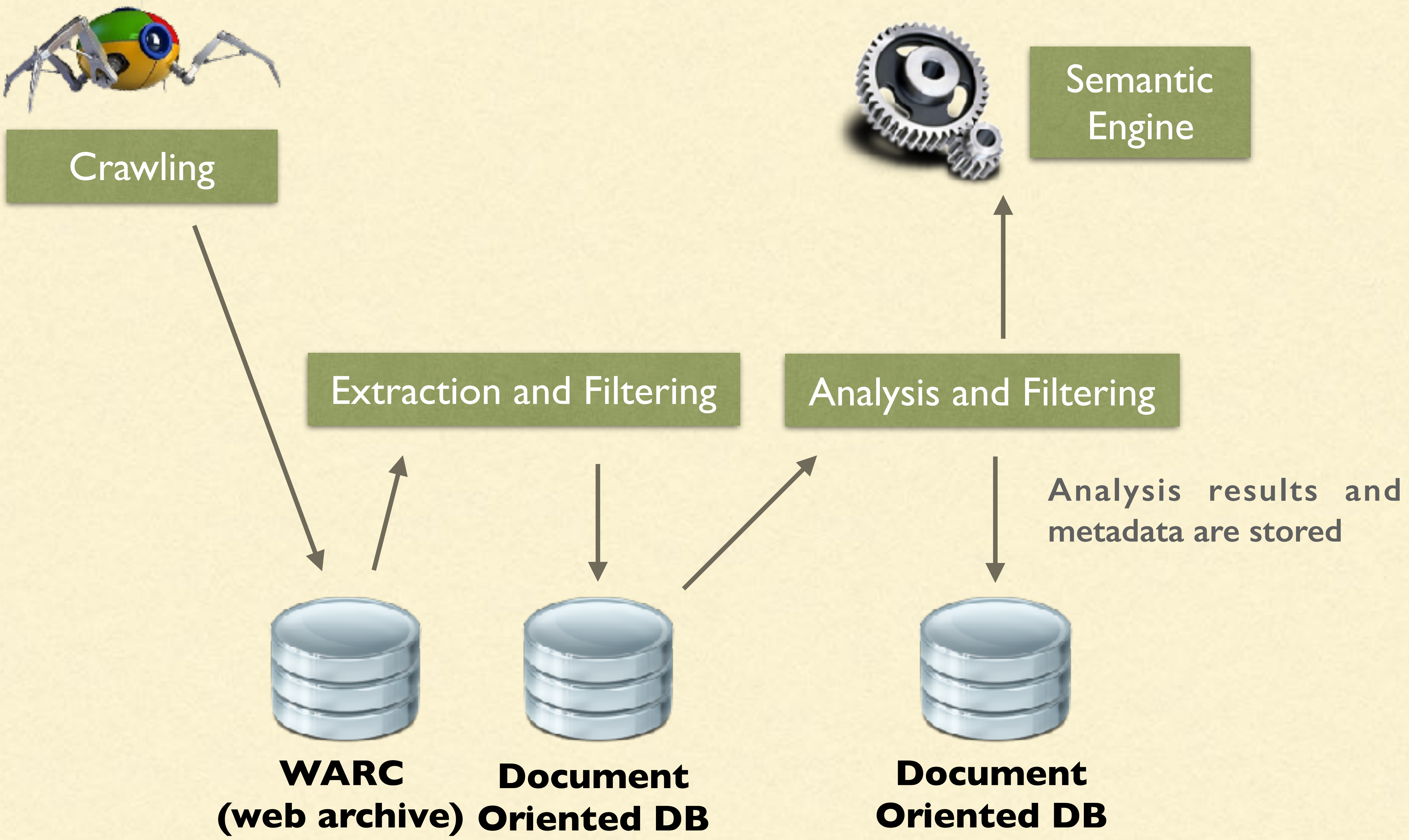
Document Oriented DB

Document Oriented DB

Architecture



Architecture



Crawling

Semantic Engine

Extraction and Filtering

Analysis and Filtering

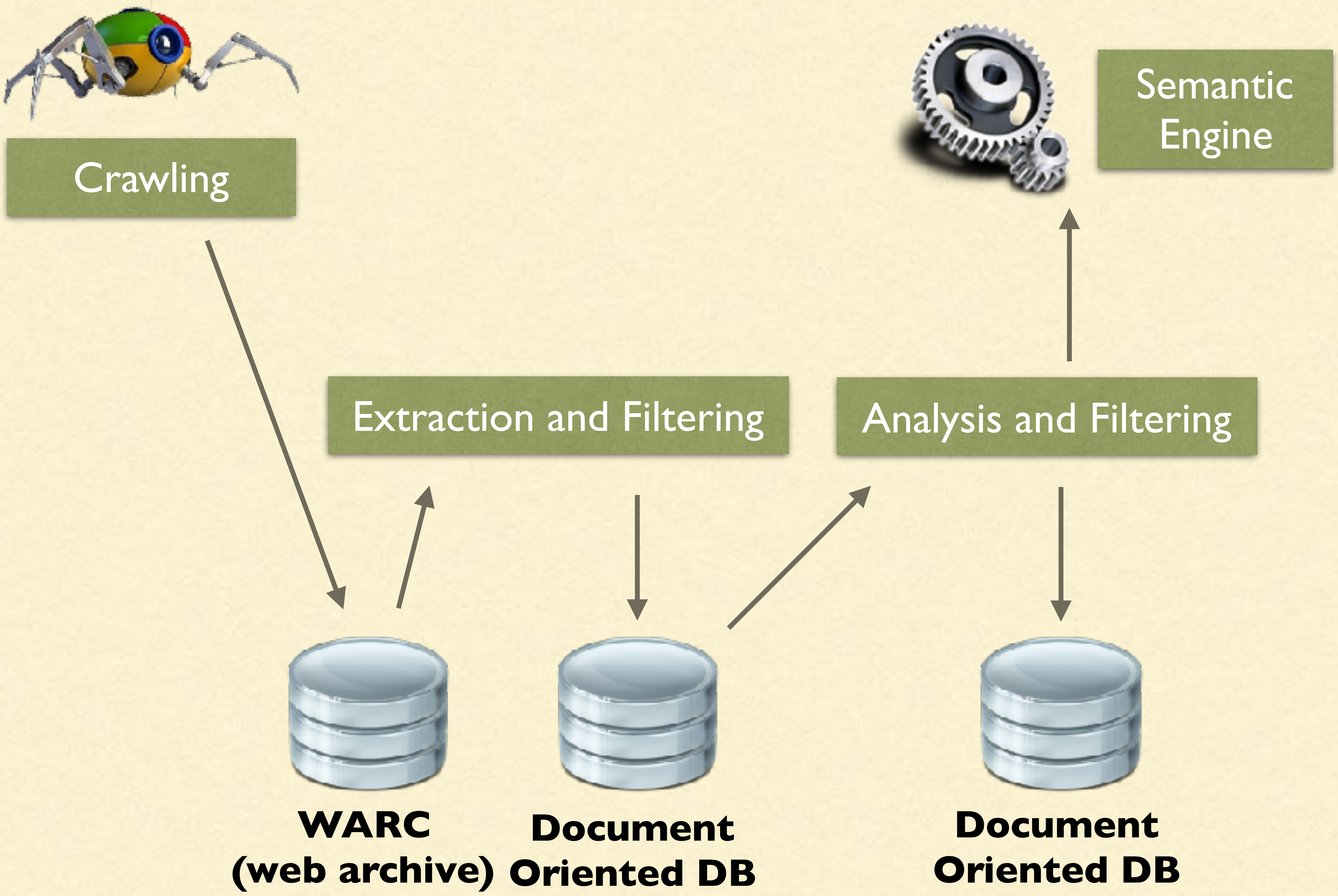
Analysis results and metadata are stored

WARC (web archive)

Document Oriented DB

Document Oriented DB

Architecture



Crawling

Semantic Engine

Extraction and Filtering

Analysis and Filtering

WARC (web archive)

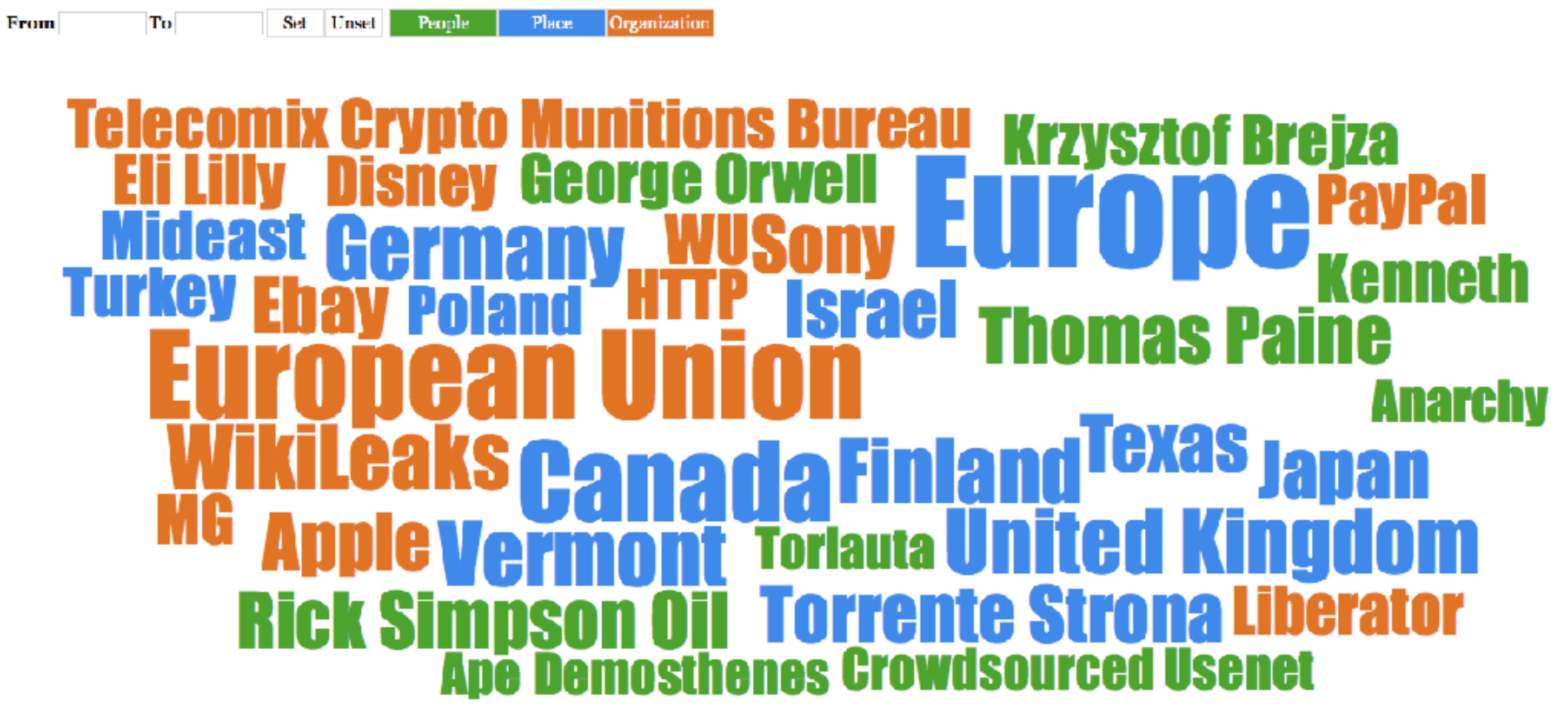
Document Oriented DB

Document Oriented DB

Visualization

Tag Cloud

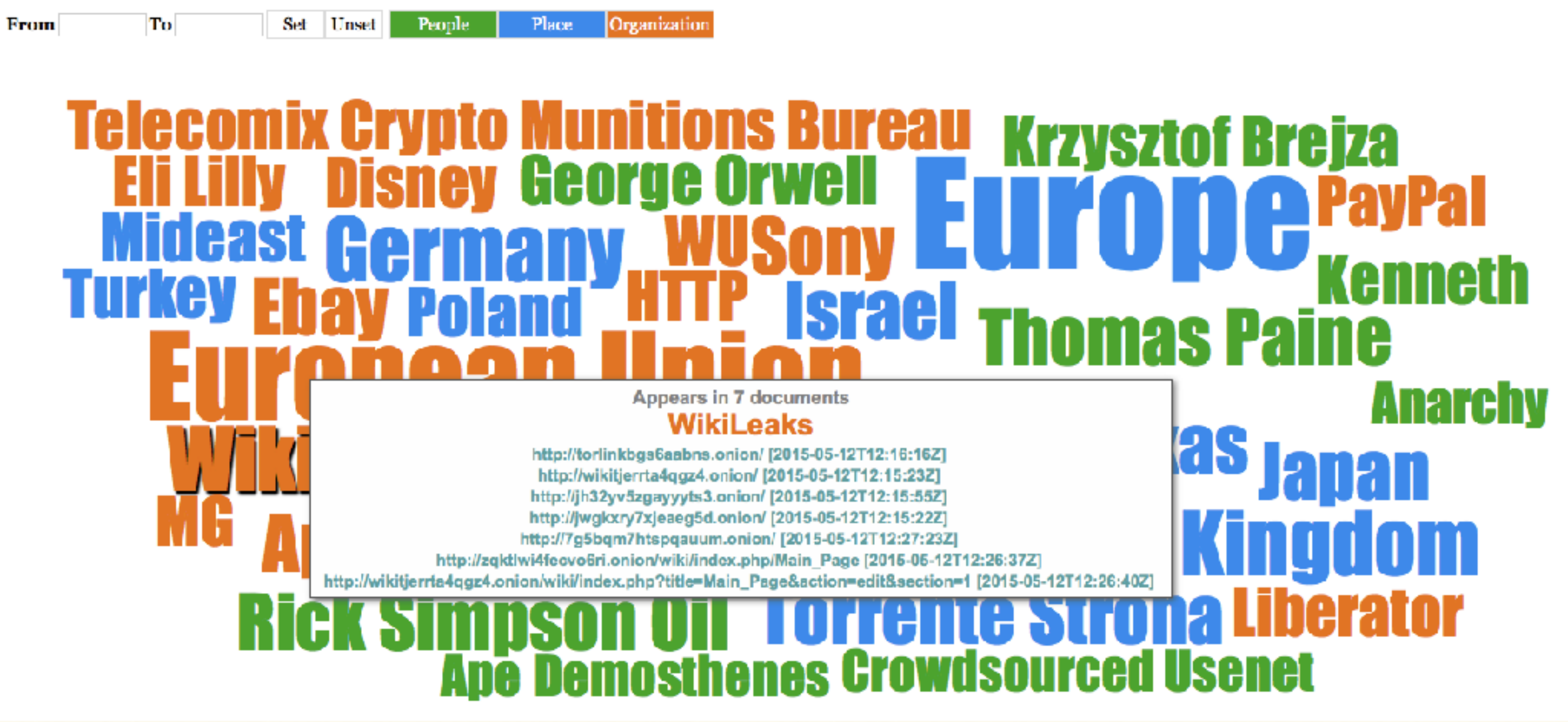
Shows the **entities extracted** by Cogito
(in the documents retrieved by the crawler)



Three types of entities are identified by the engine → **People**
Places
Organizations

Tag Cloud

Shows the **entities extracted** by Cogito
(in the documents retrieved by the crawler)



Appears in 7 documents

WikiLeaks

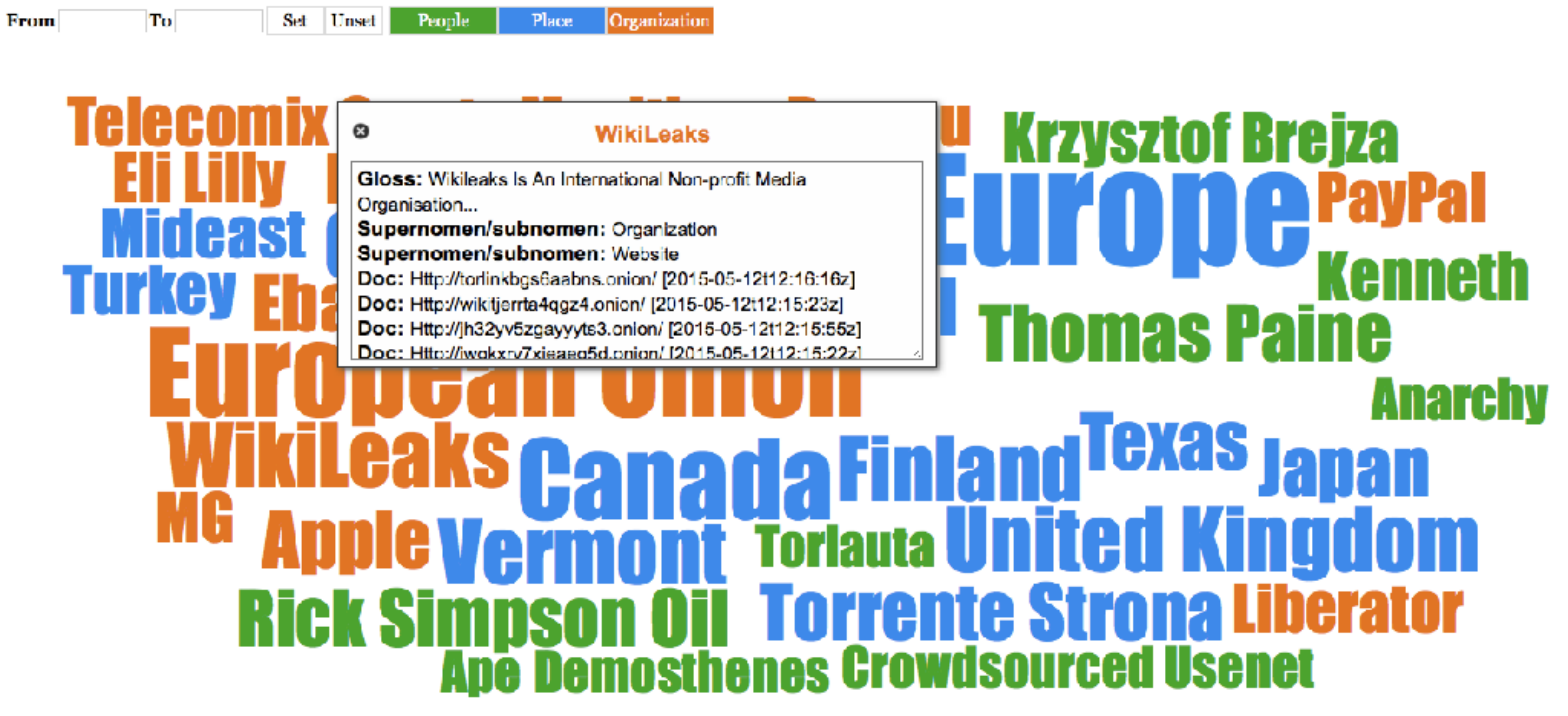
```
http://torlinkbgs6aabns.onion/ [2015-05-12T12:16:16Z]
http://wikitjerrta4qgz4.onion/ [2015-05-12T12:15:23Z]
http://jh32yv5zgayyyts3.onion/ [2015-05-12T12:15:55Z]
http://jwgkxry7xjeaeg5d.onion/ [2015-05-12T12:15:22Z]
http://7g5bqm7htspqauum.onion/ [2015-05-12T12:27:23Z]
http://zqktlwi4fecvo6ri.onion/wiki/index.php/Main_Page [2015-05-12T12:26:37Z]
http://wikitjerrta4qgz4.onion/wiki/index.php?title=Main_Page&action=edit&section=1 [2015-05-12T12:26:40Z]
```

Three types of entities are identified by the engine →

People
Places
Organizations

Tag Cloud

Shows the **entities extracted** by Cogito
(in the documents retrieved by the crawler)



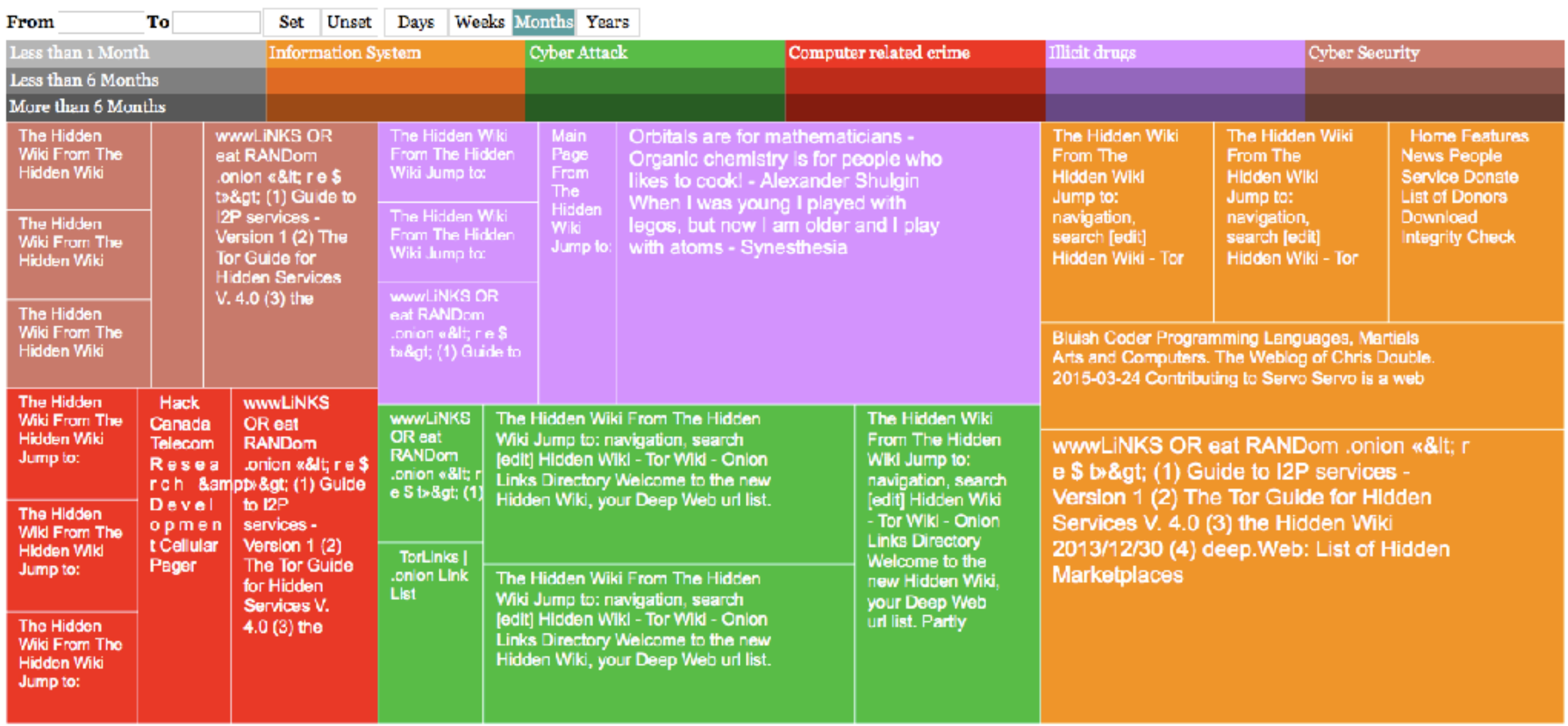
WikiLeaks

Gloss: Wikileaks Is An International Non-profit Media Organisation...
Supernomen/subnomen: Organization
Supernomen/subnomen: Website
Doc: Http://torlinkbgs6aabns.onion/ [2015-05-12t12:16:16z]
Doc: Http://wikitjerrta4qgz4.onion/ [2015-05-12t12:15:23z]
Doc: Http://jh32yv5zgayyyts3.onion/ [2015-05-12t12:15:55z]
Doc: Http://iwokxrv7xieaeg5d.onion/ [2015-05-12t12:15:22z]

Three types of entities are identified by the engine → **People**
Places
Organizations

Treemap

Shows **documents by category**
(categories are assigned by the engine)



The Hidden Wiki From The Hidden Wiki

The Hidden Wiki From The Hidden Wiki

The Hidden Wiki From The Hidden Wiki

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the

The Hidden Wiki From The Hidden Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b&gt; (1) Guide to

Main Page From The Hidden Wiki Jump to:

Orbitals are for mathematicians - Organic chemistry is for people who likes to cook! - Alexander Shulgin When I was young I played with legos, but now I am older and I play with atoms - Synesthesia

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor

Home Features News People Service Donate List of Donors Download Integrity Check

Bluish Coder Programming Languages, Martials Arts and Computers. The Weblog of Chris Double. 2015-03-24 Contributing to Servo Servo is a web

The Hidden Wiki From The Hidden Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

Hack Canada Telecom Research &amp; Development Cellular Pager

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b&gt; (1)

TorLinks | .onion Link List

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list.

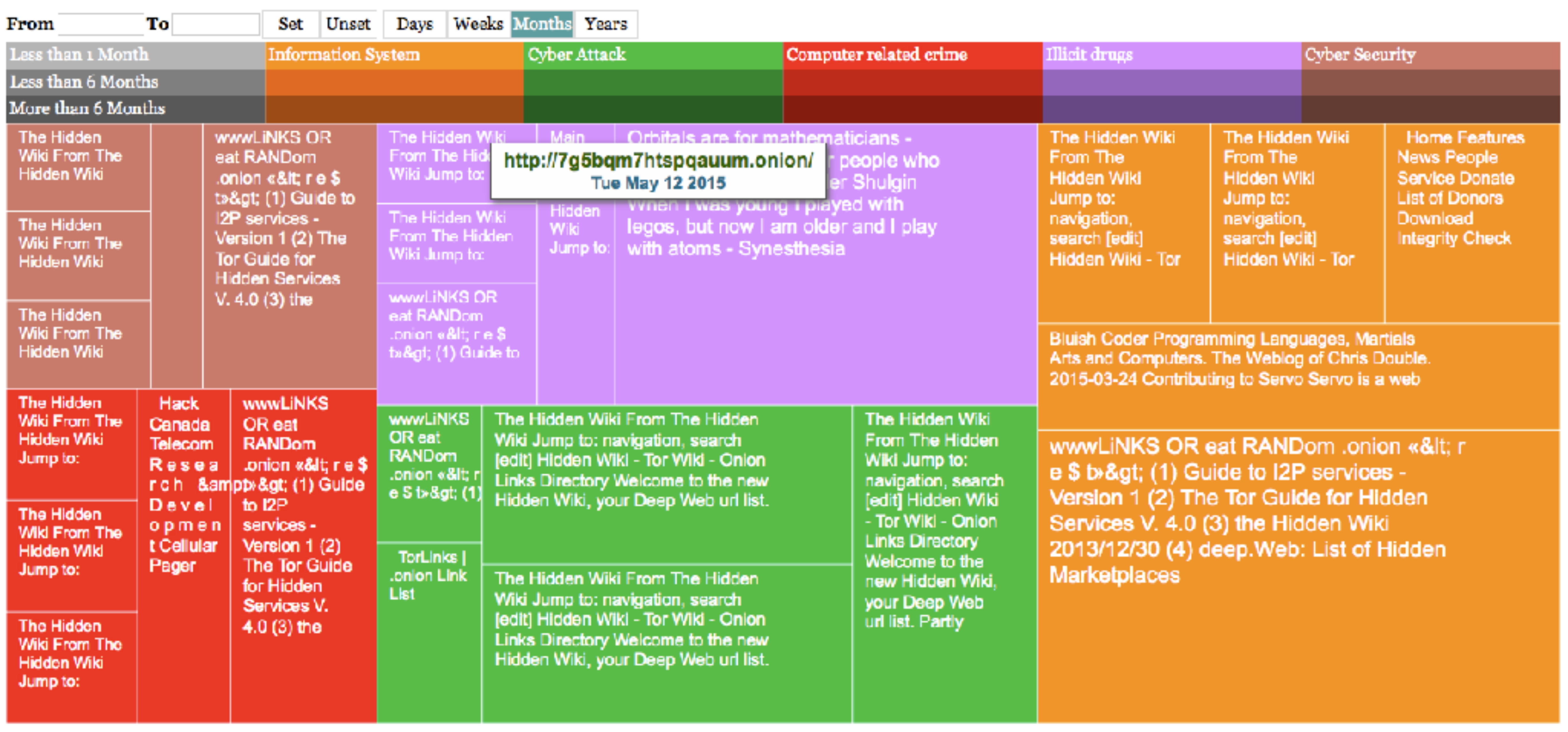
The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list.

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list. Partly

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the Hidden Wiki 2013/12/30 (4) deep.Web: List of Hidden Marketplaces

Treemap

Shows **documents by category**
(categories are assigned by the engine)



The Hidden Wiki From The Hidden Wiki

The Hidden Wiki From The Hidden Wiki

The Hidden Wiki From The Hidden Wiki

wwwLiNKS OR eat RANDom .onion «&lt; r e $ t›&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the

The Hidden Wiki From The Hidden Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

Hack Canada Telecom Research &ampt Development Cellular Pager

wwwLiNKS OR eat RANDom .onion «&lt; r e $ t›&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the

The Hidden Wiki From The Hid Wiki Jump to:

The Hidden Wiki From The Hidden Wiki Jump to:

wwwLiNKS OR eat RANDom .onion «&lt; r e $ t›&gt; (1) Guide to

Main Hidden Wiki Jump to:

Orbitals are for mathematicians - people who er Shulgin When I was young I played with legos, but now I am older and I play with atoms - Synesthesia

http://7g5bqm7htspqauum.onion/
Tue May 12 2015

wwwLiNKS OR eat RANDom .onion «&lt; r e S t›&gt; (1)

TorLinks | .onion Link List

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list.

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list.

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor Wiki - Onion Links Directory Welcome to the new Hidden Wiki, your Deep Web url list. Partly

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor

The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor

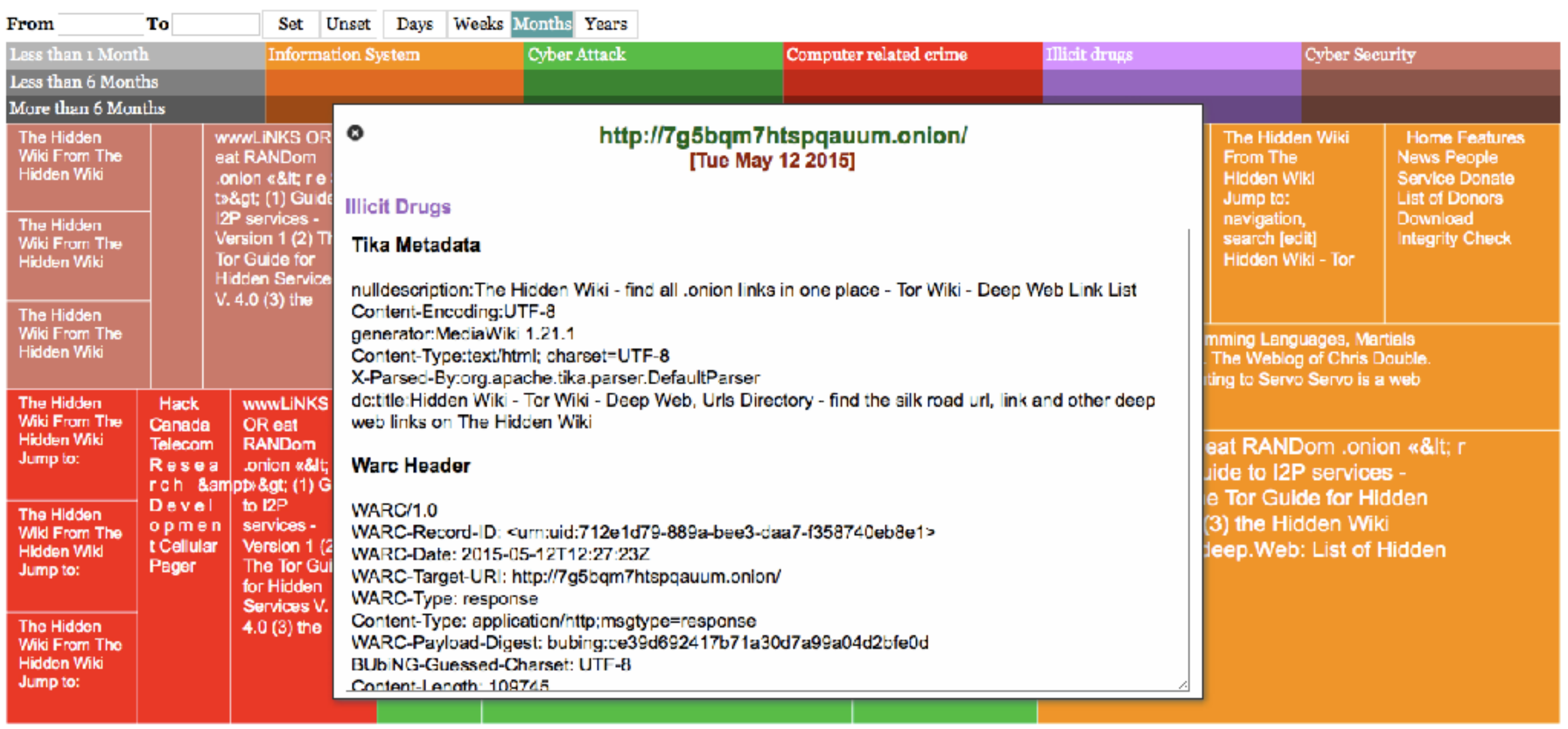
Home Features News People Service Donate List of Donors Download Integrity Check

Bluish Coder Programming Languages, Martials Arts and Computers. The Weblog of Chris Double. 2015-03-24 Contributing to Servo Servo is a web

wwwLiNKS OR eat RANDom .onion «&lt; r e $ b›&gt; (1) Guide to I2P services - Version 1 (2) The Tor Guide for Hidden Services V. 4.0 (3) the Hidden Wiki 2013/12/30 (4) deep.Web: List of Hidden Marketplaces

Treemap

Shows **documents by category**
(categories are assigned by the engine)


From
To
Set
Unset
Days
Weeks
Months
Years
Less than 1 Month
Less than 6 Months
More than 6 Months
Information System
Cyber Attack
Computer related crime
Illicit drugs
Cyber Security
The Hidden Wiki From The Hidden Wiki
The Hidden Wiki From The Hidden Wiki
The Hidden Wiki From The Hidden Wiki
wwwLiNKS OR eat RANDom .onion «&lt; re t»&gt; (1) Guide I2P services - Version 1 (2) Th Tor Guide for Hidden Service V. 4.0 (3) the
The Hidden Wiki From The Hidden Wiki Jump to:
The Hidden Wiki From The Hidden Wiki Jump to:
The Hidden Wiki From The Hidden Wiki Jump to:
Hack Canada Telecom Research &amp; Development Cellular Pager
wwwLiNKS OR eat RANDom .onion «&lt; »&gt; (1) G to I2P services - Version 1 (2 The Tor Gui for Hidden Services V. 4.0 (3) the
The Hidden Wiki From The Hidden Wiki Jump to: navigation, search [edit] Hidden Wiki - Tor
Home Features News People Service Donate List of Donors Download Integrity Check
mming Languages, Martials . The Weblog of Chris Double. ting to Servo Servo is a web
eat RANDom .onion «&lt; r uide to I2P services - e Tor Guide for Hidden (3) the Hidden Wiki deep.Web: List of Hidden

http://7g5bqm7htspqauum.onion/
[Tue May 12 2015]

Illicit Drugs

Tika Metadata

nulldescription:The Hidden Wiki - find all .onion links in one place - Tor Wiki - Deep Web Link List
Content-Encoding:UTF-8
generator:MediaWiki 1.21.1
Content-Type:text/html; charset=UTF-8
X-Parsed-By:org.apache.tika.parser.DefaultParser
dc:title:Hidden Wiki - Tor Wiki - Deep Web, Urls Directory - find the silk road url, link and other deep web links on The Hidden Wiki

Warc Header

```
WARC/1.0
WARC-Record-ID: <urn:uid:712e1d79-889a-bee3-daa7-f358740eb8e1>
WARC-Date: 2015-05-12T12:27:23Z
WARC-Target-URI: http://7g5bqm7htspqauum.onion/
WARC-Type: response
Content-Type: application/http;msgtype=response
WARC-Payload-Digest: bubing:ce39d692417b71a30d7a99a04d2bfe0d
BUbiNG-Guessed-Charset: UTF-8
Content-Length: 109745
```

Thank you for your time