

Semantic Web Technologien für Fachinformation

Prof. Dr. York Sure-Vetter
28. Juni 2017

Institute of Applied Informatics and Formal Description Methods (AIFB)



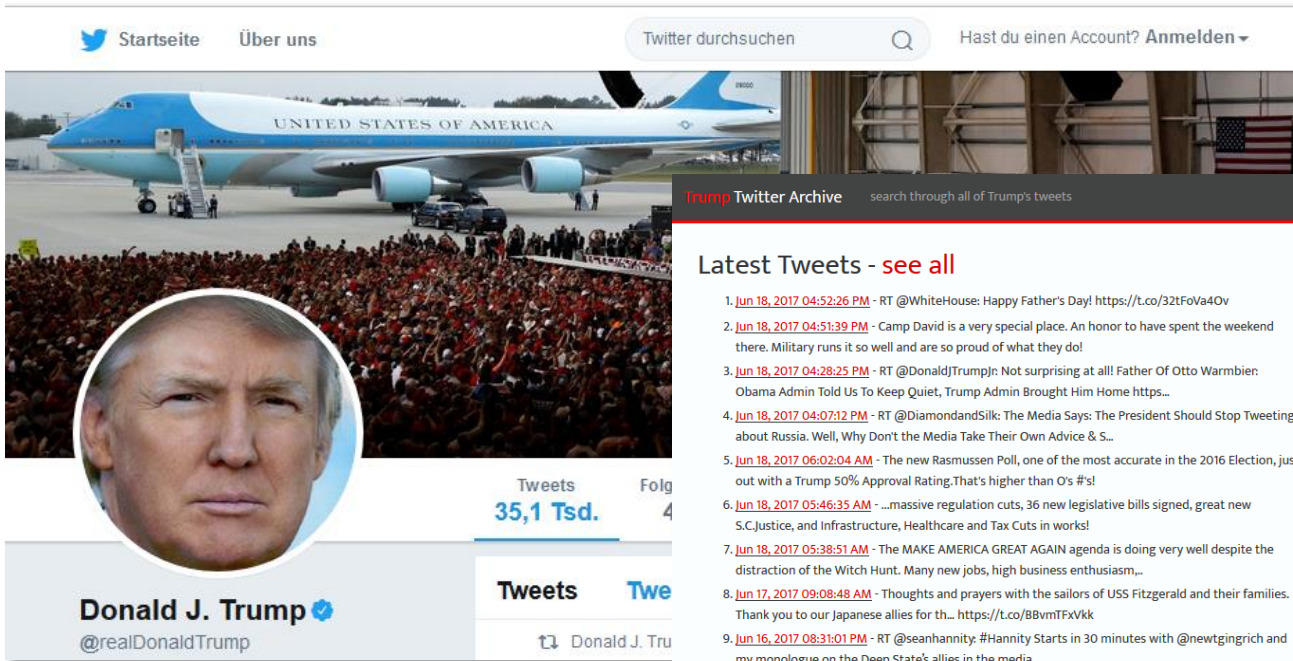
Agenda

- Beginners guide to the Web, the Semantic Web and the Social Web
- Semantics in practice: the Semantic CorA approach


Trivia

Obama has 60 Millions more  than Trump.

Trump twitters



Startseite Über uns Twitter durchsuchen Hast du einen Account? Anmelden



Trump Twitter Archive search through all of Trump's tweets

Latest Tweets - see all

1. Jun 18, 2017 04:52:26 PM - RT @WhiteHouse: Happy Father's Day! <https://t.co/32tFoVa40v>
2. Jun 18, 2017 04:51:39 PM - Camp David is a very special place. An honor to have spent the weekend there. Military runs it so well and are so proud of what they do!
3. Jun 18, 2017 04:28:25 PM - RT @DonaldJTrumpJr: Not surprising at all! Father Of Otto Warmbier: Obama Admin Told Us To Keep Quiet, Trump Admin Brought Him Home <https://t.co/32tFoVa40v>
4. Jun 18, 2017 04:07:12 PM - RT @DiamondandSilk: The Media Says: The President Should Stop Tweeting about Russia. Well, Why Don't the Media Take Their Own Advice & S...
5. Jun 18, 2017 06:02:04 AM - The new Rasmussen Poll, one of the most accurate in the 2016 Election, just out with a Trump 50% Approval Rating. That's higher than O's #s!
6. Jun 18, 2017 05:46:35 AM - ...massive regulation cuts, 36 new legislative bills signed, great new S.C. Justice, and Infrastructure, Healthcare and Tax Cuts in works!
7. Jun 18, 2017 05:38:51 AM - The MAKE AMERICA GREAT AGAIN agenda is doing very well despite the distraction of the Witch Hunt. Many new jobs, high business enthusiasm...
8. Jun 17, 2017 09:08:48 AM - Thoughts and prayers with the sailors of USS Fitzgerald and their families. Thank you to our Japanese allies for th... <https://t.co/BBvmTFxVkk>
9. Jun 16, 2017 08:31:01 PM - RT @seanhannity: #Hannity Starts in 30 minutes with @newtingrich and my monologue on the Deep State's allies in the media
10. Jun 16, 2017 08:30:08 PM - Back from Miami where my Cuban/American friends are very happy with what I signed today. Another campaign promise that I did not forget!

Fake News!

1. Jun 16, 2017 07:23:08 AM - The Fake News Media hates when I use what has turned out to be my very powerful Social Media - over 100 million people! I can go around them
2. Jun 13, 2017 07:48:44 AM - Fake News is at an all time high. Where is their apology to me for all of the incorrect stories???
3. Jun 13, 2017 05:35:55 AM - The Fake News Media has never been so wrong or so dirty. Purposely incorrect stories and phony sources to meet their agenda of hate. Sad!
4. Jun 6, 2017 07:15:36 AM - Sorry folks, but if I would have relied on the Fake News of CNN, NBC, ABC, CBS, washpost or nytimes, I would have had ZERO chance winning WH
5. May 30, 2017 06:04:48 AM - Russian officials must be laughing at the U.S. & how a lame excuse for why the Dems lost the election has taken over the Fake News.

<https://twitter.com/realdonaldtrump>

<http://www.trumptwitterarchive.com/>

Get Active - www.contactingcongress.org

234 tweets with "loser"

I feel sorry for Rosie 's new partner in love whose parents are devastated at the thought of their daughter being with @Rosie—a true loser.

— Donald J. Trump (@realDonaldTrump) December 14, 2011

222 tweets with "dumb" or "dummy"

You must admit that Bryant Gumbel is one of the dumbest racists around - an arrogant dope with no talent. Failed at CBS etc-why still on TV?

— Donald J. Trump (@realDonaldTrump) August 21, 2013

204 tweets with "terrible"

I loved beating these two terrible human beings. I would never recommend that anyone use her lawyer, he is a total loser!

— Donald J. Trump (@realDonaldTrump) May 23, 2013

183 tweets with "stupid"

@michellemalkin You were born stupid!

— Donald J. Trump (@realDonaldTrump) March 22, 2013

156 tweets with "weak"

Obama twitters



The screenshot shows the Twitter profile of Barack Obama. At the top, there are navigation links for 'Startseite' and 'Über uns', a search bar, and a link to 'Anmelden'. The profile picture is a circular image of Barack Obama smiling. The header image shows him and a woman standing on a rocky outcrop overlooking a vast mountain range. Below the header, statistics are listed: 15,5 Tsd. Tweets, 629 Tsd. Follower, 91 Mio. Gefällt mir, and 10 Listen. A 'Folgen' button is visible. The bio reads: 'Dad, husband, President, citizen.' A recent tweet is displayed: 'Of all that I've done in my life, I'm most proud to be Sasha and Malia's dad. To all those lucky enough to be a dad, Happy Father's Day!' with a dropdown arrow. A 'Neu bei Twitter?' section is also present.

<https://twitter.com/barackobama>

7.	<i>Four more years.</i> pic.twitter.com/bAJE6Vom	@BarackObama ^[13]	0.937	November 6, 2012	[G]
8.	<i>Thank you for everything. My last ask is the same as my first. I'm asking you to believe—not in my ability to create change, but in yours.</i>	@POTUS44 ^[15]	0.882	January 10, 2017	[H]

https://en.wikipedia.org/wiki/List_of_most_retweeted_tweets_on_Twitter

2017: Frank-Walter Steinmeier on Facebook



Frank-Walter Steinmeier als Bundespräsident

York Startseite 12

Frank-Walter Steinmeier als Bundespräsident

@steinmeierbundespraesident

Startseite

Gefällt mir Abonnieren Teilen ...

Nachricht senden

<https://www.facebook.com/steinmeierbundespraesident/>

The Web

- 1989 Web Idea
- 1994 Amazon founded
- 1998 Google founded
- 2001 .com Bubble
- 2004 Facebook founded
- 2008 Obama becomes president due to Web strategy
- 2009 SAP defines Web strategy
- 2010 Instagram
- 2011 Google+
- 2012 Coursera
- 2017 Bundespräsident on FB

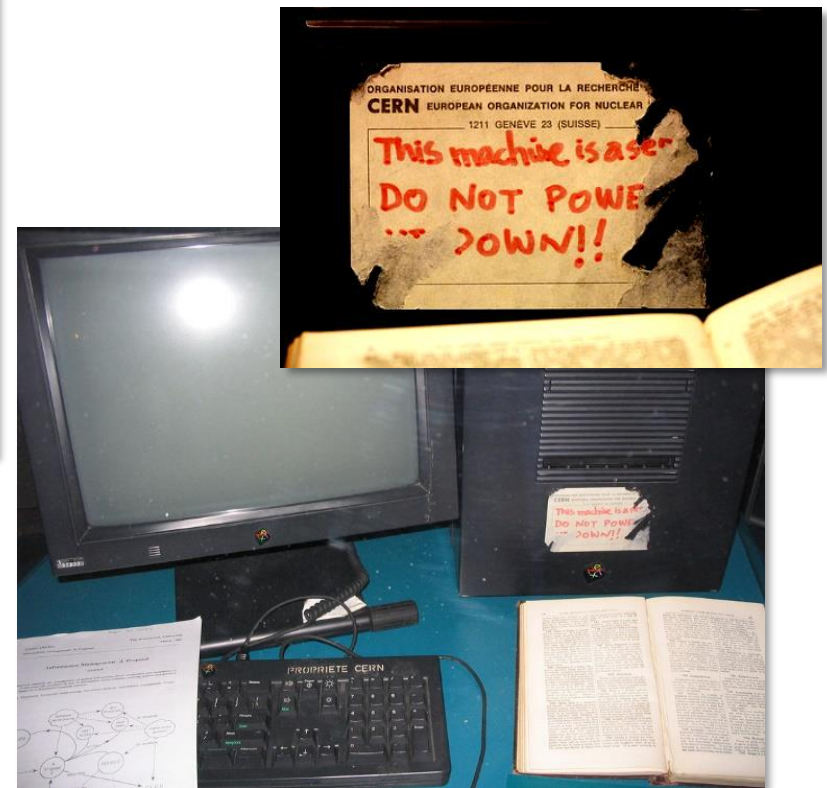
Advertising & Buy/sell

Live and Work

The inventor of the World Wide Web



Sir Tim Berners-Lee
(Turing Award winner 2017)



The first Website ... ever!

- The „first Website“, such as shown by Tim Berners-Lee at the Hypertext 91 surfaces conference:

<http://info.cern.ch/>

- First Website:

<http://info.cern.ch/hypertext/WWW/TheProject.html>

The first Website ... ever!

```
The World Wide Web project

                                WORLD WIDE WEB

The WorldWideWeb (W3) is a wide-area hypermedia[1] information retrieval
initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this
document, including an executive summary[2] of the project, Mailing lists[3] ,
Policy[4] , November's W3 news[5] , Frequently Asked Questions[6] .

    What's out there?[7]Pointers to the world's online information,
                                subjects[8] , W3 servers[9], etc.

    Help[10]                on the browser you are using

    Software                A list of W3 project components and their current
    Products[11]            state. (e.g. Line Mode[12] ,X11 Viola[13] ,
                                NeXTStep[14] , Servers[15] , Tools[16] , Mail
                                robot[17] , Library[18] )

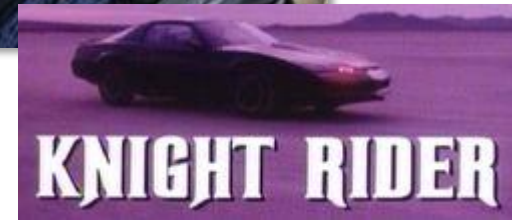
    Technical[19]          Details of protocols, formats, program internals
                                etc

<ref.number>, Back, <RETURN> for more, or Help: | █
```

Computers – some 30 years back (1987)



```
**** COMMODORE 64 BASIC V2 ****  
64K RAM SYSTEM 38911 BASIC BYTES FREE  
READY.
```



Search (like it's 1997!)

- <http://web.archive.org/web/19981111183552/google.stanford.edu/>



Web 1.0: Vision of the Web

- Is a space in which
 - **Resources** are identified by Uniform Resource Identifiers (URIs)
 - **Protocols** support interaction between agents (HTTP)
 - **Formats** represent information resources (HTML)

Web 1.0: Uniform Resource Identifiers (URI)

- **Uniform Resource Identifier**
- Resources may be anything that can be linked to or spoken of
 - Resources can be abstract or physical objects
 - Resources can contain a reference to another resource
- *Identifiable*, but not necessarily *retrievable*
 - (e.g. protected access)
- A single global system of identifiers
 - Each URI ideally identifies a single resource in a context-independent manner
- URIs act as names and addresses
- URIs require institutions
 - E.g. the registry that handles domain names

Web 1.0: Uniform Resource Identifiers (URI)

- Structure:

- URI = scheme ":" hier-part ["?" query] ["#" fragment]

- Example:

- foo://example.com:8042/over/there?name=ferret#nose



Web 1.0: HTTP

- **HTTP:** A protocol that is basically stateless, a transaction consisting of:
 - **Connection**
 - The establishment of a connection by the client to the server - when using TCP/IP port 80 is the well-known port, but other non-reserved ports may be specified in the URL;
 - **Request**
 - The sending, by the client, of a request message to the server;
 - **Response**
 - The sending, by the server, of a response to the client;
 - **Close**
 - The closing of the connection by either both parties.

Web 1.0: HTML

- HTML: A representation format
 - Idea: Decoupling of content and representation
 - Cues for graphical presentation of content

	Prof. Dr. York Sure-Vetter
	Professor Tel.: +49 (721) 608 44283 Email: york.sure-vetter@kit.edu
	Wissensmanagement Raum: 219 (Geb. 11.40)
	Sekretärin: Beate Kühner , Gisela Schillinger

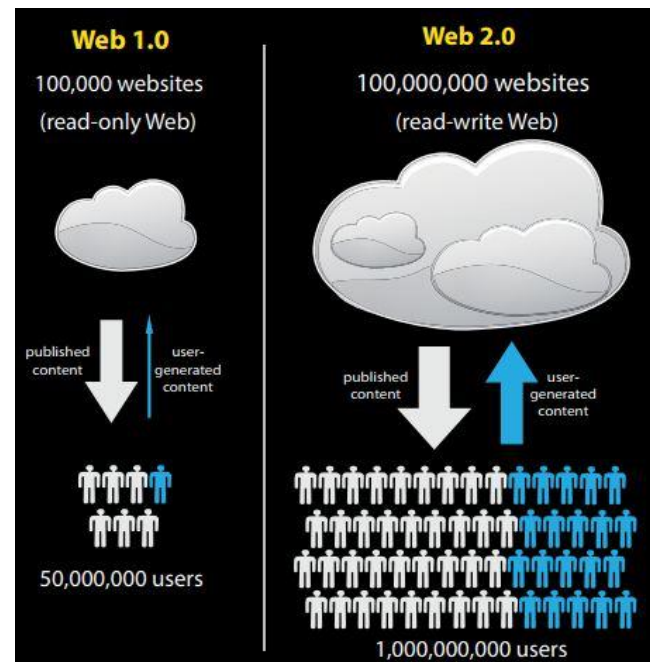
```

<p>Prof. Dr. York Sure-Vetter</p>
</tr>
<tr>
<td class="row-left">
<p><span class="ein-em">
Professor
</span>Tel.:+49 (721) 608 44283
<br />Email: york.sure-vetter@kit.edu
<br />
<a href="/web/Wissensmanagement" title="Wissensmanagement">Wissensmanagement</a>
  
```

Difference Web 1.0 and Web 2.0

- Web 1.0 focused on one-to-many information publishing
 - Very static information

- Web 2.0 focuses on many-to-many information publishing
 - Collaborate and share Information like e.g. in wikis, facebook, blogs, etc.



Web 2.0: e.g. Wikis

- Everyone can edit content on the Web
- ... without access to the server



The image shows a screenshot of a web page for the Karlsruhe Institute of Technology (KIT). The page header includes the KIT logo and the name of the Institute for Applied Informatics and Formal Description Methods (AIFB). A search bar is visible at the top left.

The main content area displays a profile for Prof. Dr. York Sure, including a photograph and contact information:

- Prof. Dr. York Sure
- Professor
- Tel.: +49 (721) 608 44283
- Email: york.sure@kit.edu
- Wissensmanagement
- Raum: 219 (Geometrie)
- Sekretärin: Barbara

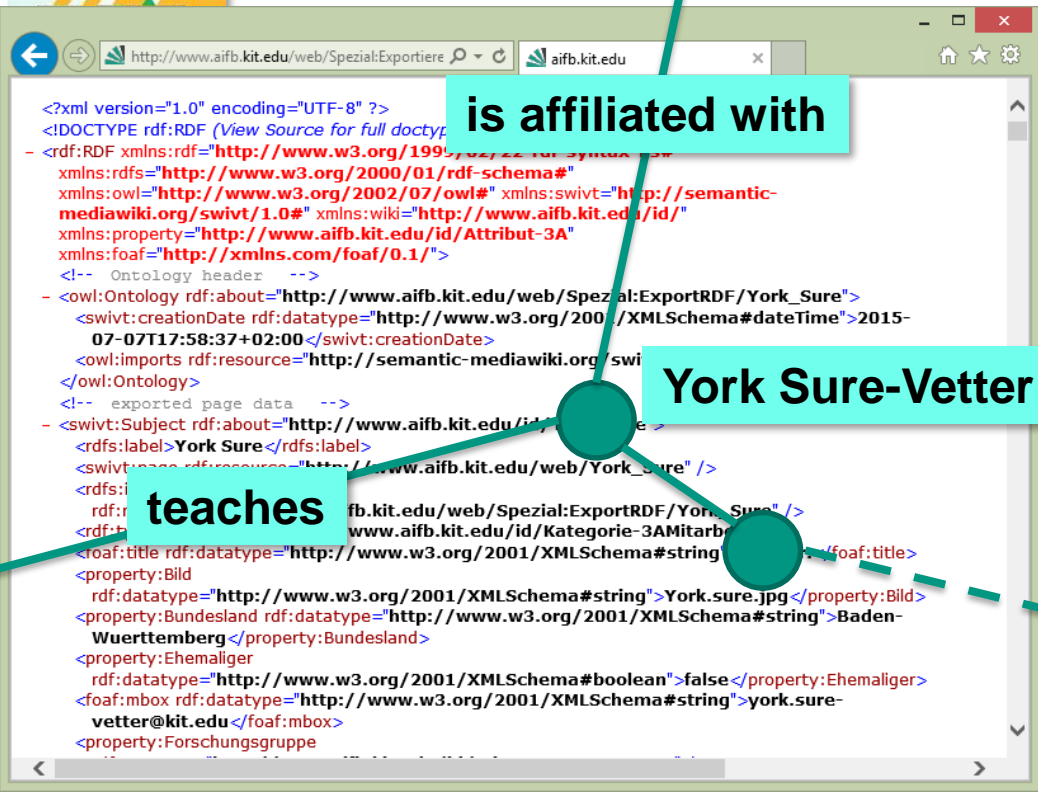
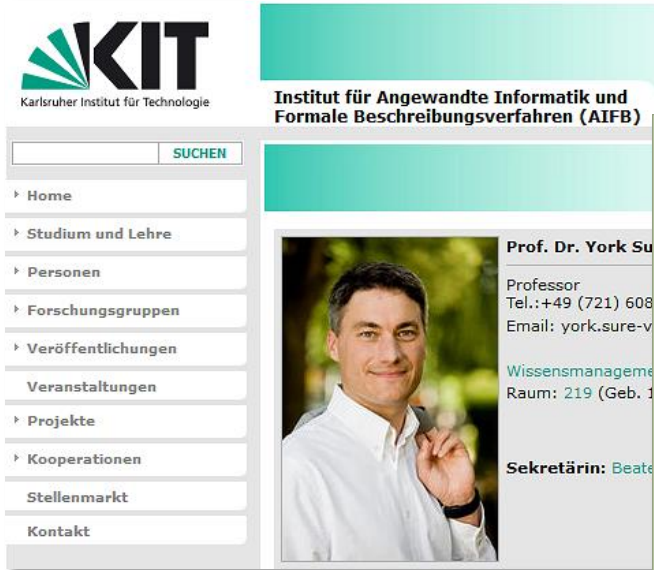
Below the profile, there is a 'Bearbeiten von York Sure' (Edit) interface. It features a rich text editor toolbar with icons for bold, italic, underline, link, unlink, list, and other formatting options. The editor contains a code block with the following text:


```

    {{Mitarbeiter
    |Vorname=York
    |Nachname=Sure-Vetter
    |Akademischer Titel=Prof. Dr.
    |Kürzel=ysu
    |Forschungsgruppe=Wissensmanagement
    |Stellung=Professor
    |Ehemaliger=False
    |Telefon=+49 (721) 608 44283
    |Email=york.sure-vetter@kit.edu
    |Raum=219
    |Bild=York.sure.jpg
    }}
    
```

Semantic Web: RDF and OWL

- Semantic descriptions of the content
- Machine readable and executable



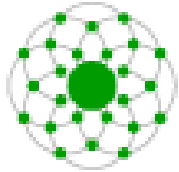
Web Science (WS2017-18)

Semantic MediaWiki (SMW)



Max Völkel, Markus Krötzsch, Denny Vrandečić, Heiko Haller, Rudi Studer:
Semantic Wikipedia. WWW 2006: 585-594

VRE Semantic CorA



Semantic CorA

Semantic Collaborative Corpora Analysis for Humanities and Social Sciences

- www.semantic-cora.org
- Partners: DIPF, KIT, Uni Göttingen

- Main Objectives:
 - Development of a Virtual Research Environment (VRE)
 - Exemplary realization in the research field History of Education
 - Enhancing quantitative + qualitative analysis
 - Infrastructural distribution into other disciplinary contexts

- Technological Platform:
 - MediaWiki (MW) + extension Semantic MediaWiki (SMW)

Credits



Thanks to **Dr. Basil Eil**
for the slides about
Semantic CorA

Step 1: Importing Research Data

Scripta Paedagogica Online



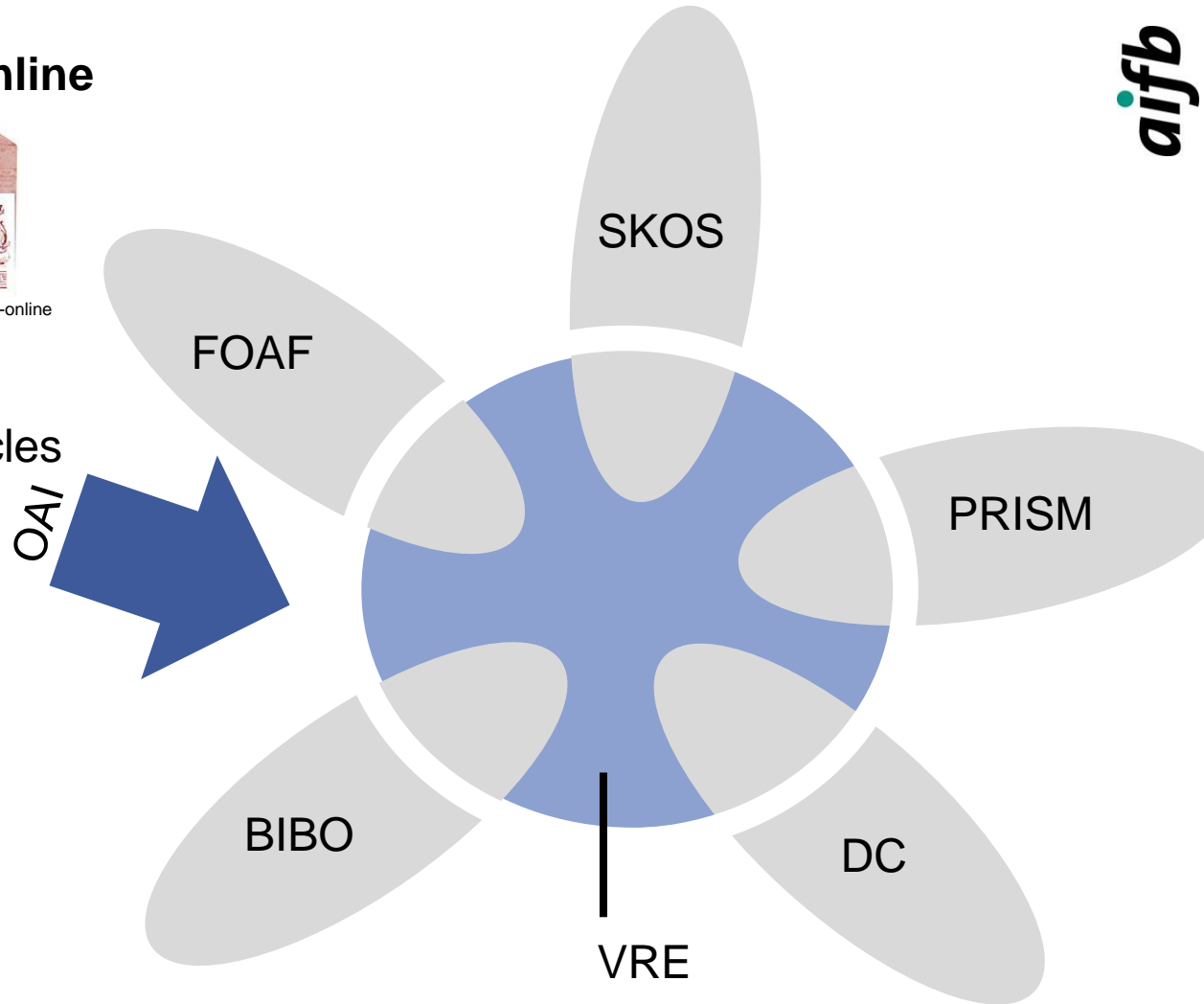
<http://bbf.dipf.de/retro-digit0.htm/digitale-bbf/scripta-paedagogica-online>



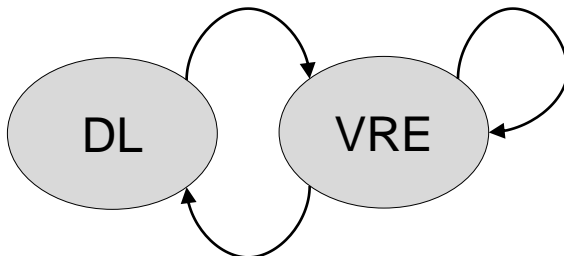
25 lexica
22.000 articles

+ manual import of
62 undigitized lexica

OAI



Life-Cycles & Interactions



Step 2: Enriching Research Data



U b s i c h t.

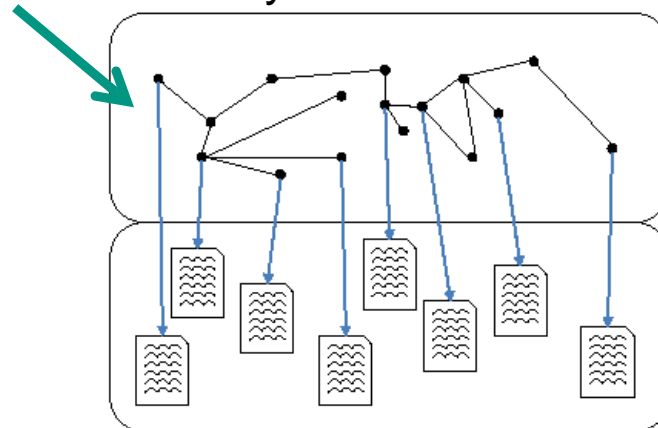
Wohl manche Leser die Worte eines Au-
ches überschlagen, so will ich in die-
sem ersten Artikel das Erheblichste über
die Absicht dieser kleinen Schrift sagen:
Wer die edle Absicht hat, gesunde, geschickte
und wohlgestirtere Kinder zu erziehen, der vertheile
seine weisen und übertreuen Bemühungen in folgen-
de drei **Grundsätze**:

1) Er suchet seiner holden Jugend eine **edelpet-
liche gute Erziehung** zu verschaffen.
Diese Sorge vor die Gesundheit und vor die
Ausbildung des Leibes fängt mit der Geburt
eines Kindes an, und höret erst in den Zah-
nen auf, da der Verstand und das Herz des
Sündlings Stärke genug besiget, selbst der
fluge Reiterer eines Körpers zu seyn!

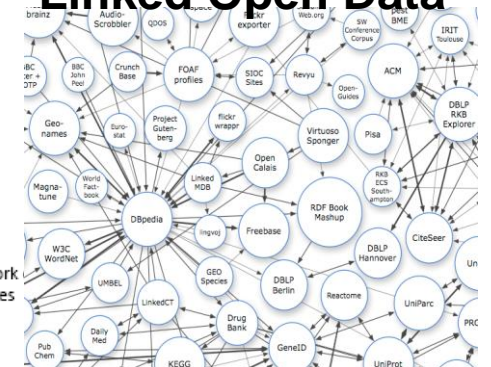
2) Die **wissenschaftliche gute Erziehung**
ist der zweite wichtige Theil der Pädagogie.
Sie begreiffet alle Bemühungen unter sich,
welche Eltern anwenden: den Verstand ihrer
Kinder

- * Semantically annotate parts of images
- * Create data about entities
- * Link entities
- * Open coding

Queryable network



Linked Open Data

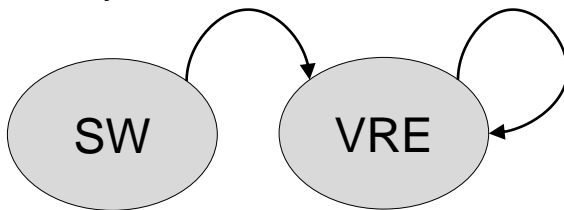


http://linkeddata.org/static/images/lo-datasets_2009-07-14_cropped.png

Semantic network
metadata entities

Imported research objects:
scanned pages

Life-Cycles & Interactions



Step 3: Data Cleansing

- Imported data may contain errors
- Inform the provider & other consumers and provide patches

1

Property	Value
Title:	Abbitte
Language:	ger
First Page:	6
Last Page:	800

2

Title:	Abbitte
Category:	1863-84_Roitus
Part of Volume:	1863-84_Roitus/Volume_01
Language:	ger
First Page:	6
Last Page:	800
Free text:	{{DataCorrection}}

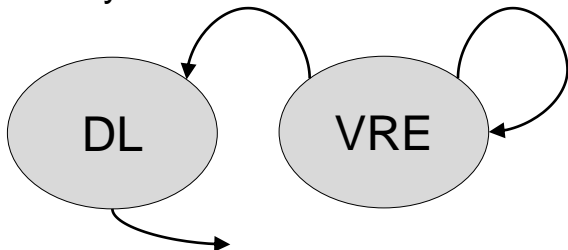
4

ErrorType:	wrong property value
ErrorOrigin:	BBF
CorrectionType:	edited in template call
Reporter:	Basil
CorrectionStatus:	done
ConcernedProperty:	Last Page
ConcernedPValueOld:	800
ConcernedPValueNew:	7
Explanation:	lemma is not that long, see last page of article

3

[DataCorrection \(Edit\)](#)

Life-Cycles & Interactions

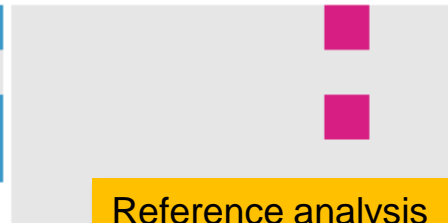


Step 4: Exploring and Analyzing

Internal Reference



Reference to Author

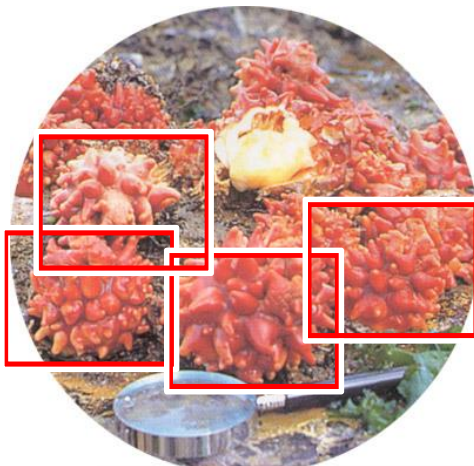


Reference to Literature



Reference analysis

top-15 der Frauenberufe:

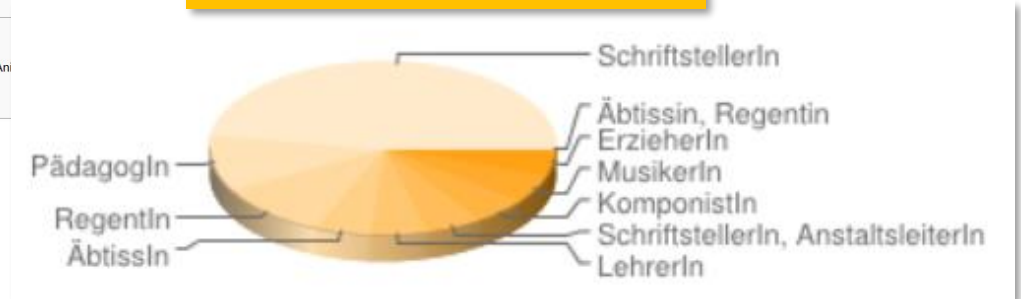


Fine-grained image analysis



Snip	ImageAnnotationText
	Animal 1
	Animal 2
	Animal 3
	Animal 4

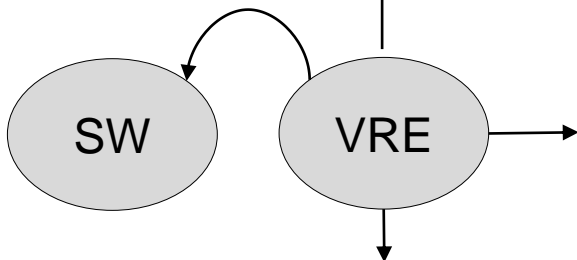
Visualization and exploration



Step 5: Export and Sharing

- Selectively export content using imported vocabularies:
 - Corrected data
 - Annotations / Semantic network
 - Aggregated data
 - Evidence data related to research hypothesis

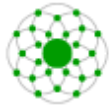
Life-Cycles & Interactions



Next Steps

- Support **more resource types** besides lexica and images:
 - Video, audio, interview transcripts
- Vision:
 - future **ecologies** of small to large research projects and data providers
 - **data flows** between the participants
 - leads to **enriched and improved data**
 - providing benefits in a **multitude of collaborations**

Summary



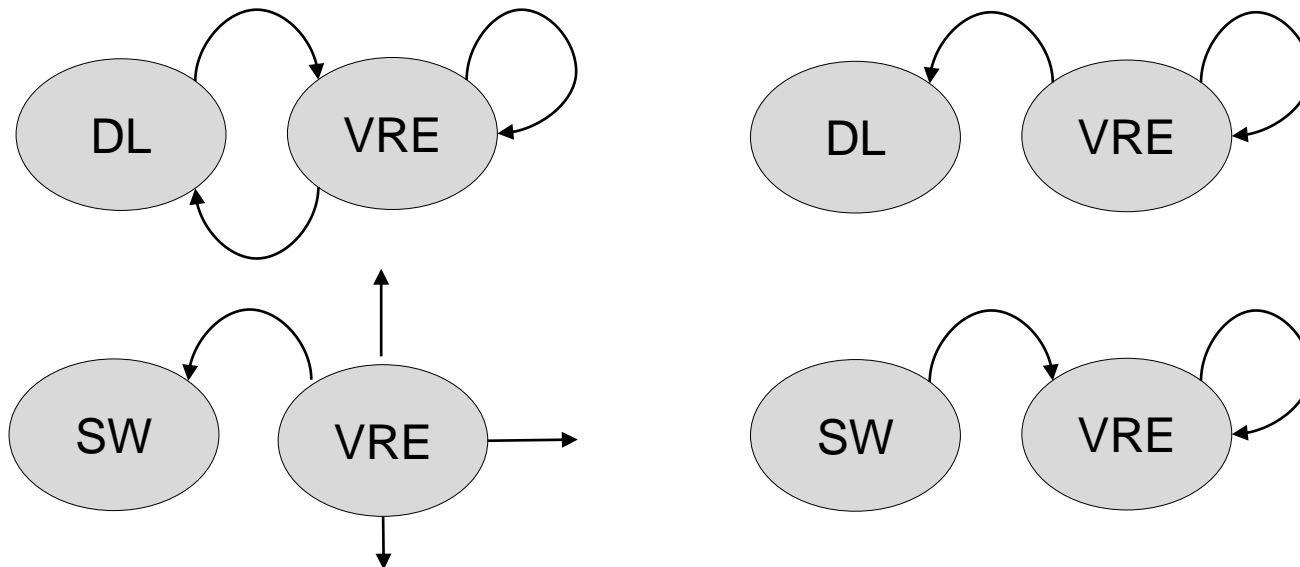
Semantic CorA

Semantic Collaborative Corpora Analysis for Humanities and Social Sciences

Activities:

1. Importing Research Data
2. Enriching Research Data
3. Data Cleansing
4. Exploring and Analyzing
5. Export and Sharing

Life-Cycles & Interactions:



Trivia – revisited

Obama has 60 Millions more **followers** than Trump.

The World Record in Retweeting is currently being held by 

2017: New World Record in Retweets

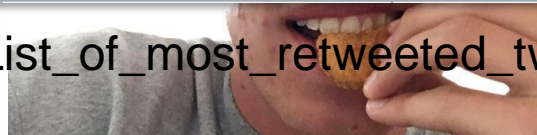
afin

Top tweets [\[edit\]](#)

The following table lists the top ten most retweeted tweets on Twitter, the account that tweeted it, the total rounded to the nearest thousand, and the date it was originally tweeted. The notes include the details surrounding the tweet.

Rank ↕	Tweet ↕	Account name ↕	Retweets (millions) ↕	Date tweeted ↕	Notes ↕
1.	<p>HELP ME PLEASE. A MAN NEEDS HIS NUGGS</p> <p>pic.twitter.com/4SrfHmEMo3</p>	@carterjwm ^[2]	3.641	April 5, 2017	[A]

https://en.wikipedia.org/wiki/List_of_most_retweeted_tweets_on_Twitter



“Carter Wilkerson tweeted [Wendy's](#) asking how many retweets he needed for a year of free chicken nuggets. When Wendy's replied with "18 Million", he accepted the challenge. The tweet received unlikely support from several major companies including [Microsoft](#), [Amazon](#) and [Google](#), prompting the hashtag #NuggsForCarter and propelling the tweet to become the most retweeted of all time within 34 days, at which point Wendy's gave him a year of free nuggets.”

See also: <http://www.npr.org/sections/thetwo-way/2017/05/09/527597422/quest-for-free-chicken-nuggets-inspires-twiters-most-retweeted-tweet>

Vielen Dank für Ihre Aufmerksamkeit!