Semantic Web Technologien für Fachinformation

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

Institute of Applied Informatics and Formal Description Methods (AIFB)
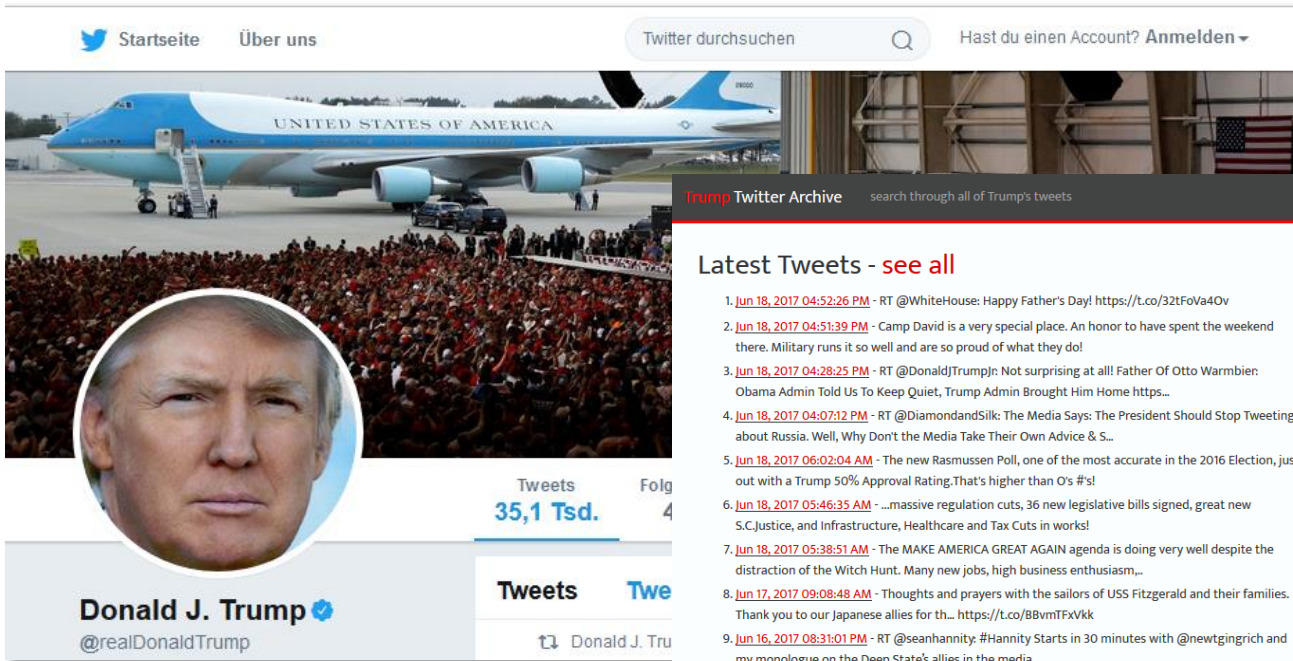
Agenda

- Beginners guide to the Web,
the Semantic Web and the Social Web

- Semantics in practice: the Semantic CorA approach

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Trivia

Obama has 60 Millions more ⬜⬜⬜ than Trump.

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Trump twitters

https://twitter.com/realdonaldtrump

http://www.trumptwitterarchive.com/

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Obama twitters

https://twitter.com/barackobama

7.	*Four more years.* *pic.twitter.com/bAJE6Vom* ⤴	@BarackObama[13]	0.937	November 6, 2012	[G]
8.	*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.*	@POTUS44[15]	0.882	January 10, 2017	[H]

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

Institute of Applied Informatics and Formal
Description Methods (AIFB)

2017: Frank-Walter Steinmeier on Facebook

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

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

The Web

1989   Web Idea

1994   Amazon founded

1998   Google founded   **Advertising & Buy/sell**

2001   .com Bubble

2004   Facebook founded

2008   Obama becomes president
        due to Web strategy

2009   SAP defines Web strategy

2010   Instagram   **Live and Work**

2011   Google+

2012   Coursera

2017   Bundespräsident on FB

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

The inventor of the World Wide Web

Sir Tim Berners-Lee
(Turing Award winner 2017)

Prof. Dr. York Sure-Vetter
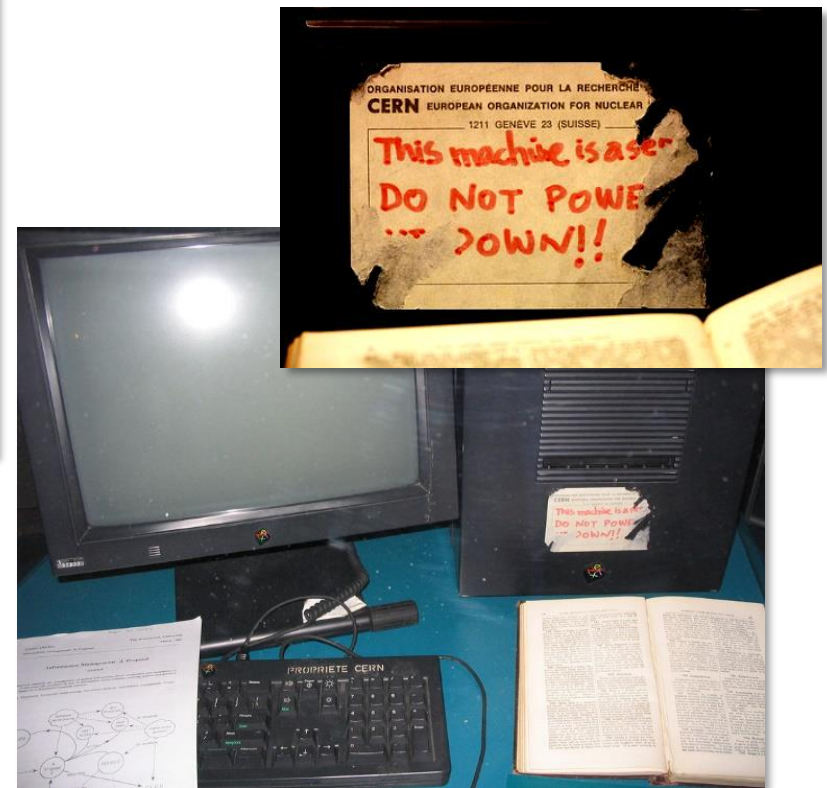
The first Website … ever!

- 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!

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Computers – some 30 years back (1987)

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

KNIGHT RIDER

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Search (like it's 1997!)

- http://web.archive.org/web/19981111183552/google.stanford.edu/
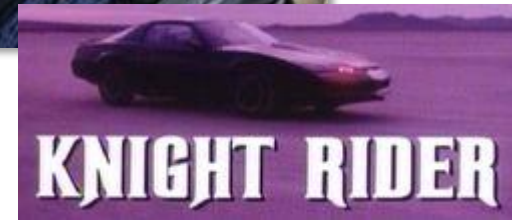
Web 1.0: Vision of the Web

- Is a space in which

  - *Resources* are identified by Uniform Resource Identifiers (URIs)

  - *Protocols* support interaction between agents (HTTP)

  - *Formats* represent information resources (HTML)

Prof. Dr. York Sure-Vetter   Institute of Applied Informatics and Formal
Description Methods (AIFB)

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

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Web 1.0: Uniform Resource Identifiers (URI)

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

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

      scheme   authority   path   query   fragment

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

```html
<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>
```
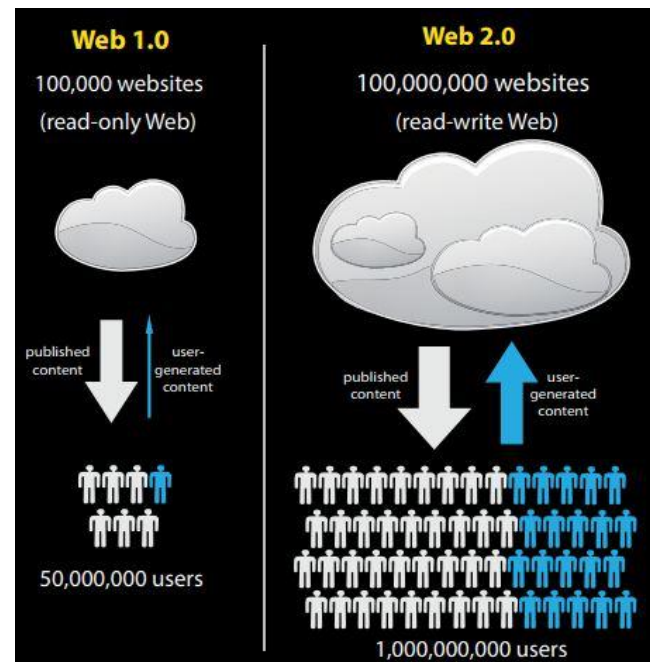
Prof. Dr. York Sure-Vetter

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.

Prof. Dr. York Sure-Vetter   Institute of Applied Informatics and Formal
Description Methods (AIFB)

Web 2.0: e.g. Wikis

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

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Semantic Web: RDF and OWL

- Semantic descriptions of the content
- Machine readable and executable

KIT

is affiliated with

York Sure-Vetter

teiches

Web Science (WS2017-18)

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)
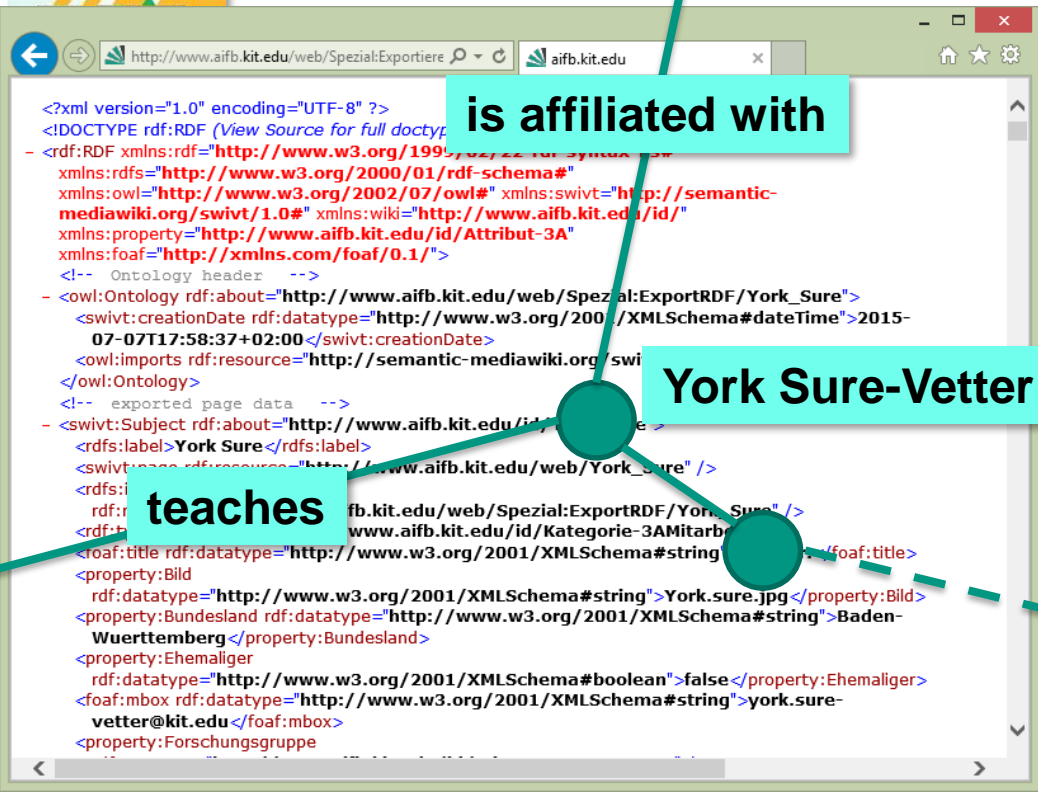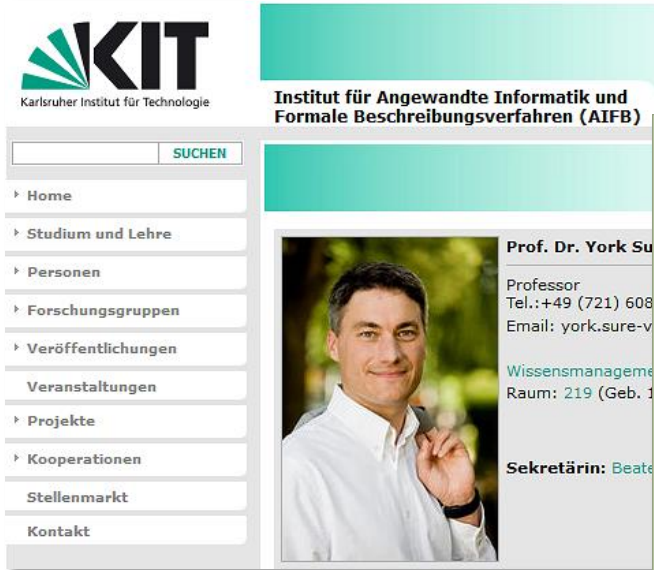
Semantic MediaWiki (SMW)

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

Prof. Dr. York Sure-Vetter
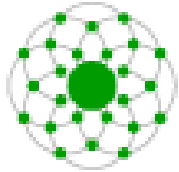
Institute of Applied Informatics and Formal Description Methods (AIFB)

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)

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Credits

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

Prof. Dr. York Sure-Vetter

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

Life-Cycles & Interactions

DL ⟷ VRE

OAI

SKOS

FOAF

PRISM

BIBO

DC

VRE

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)

Step 2: Enriching Research Data

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

Linked Open Data

Queryable network

Semantic network metadata entities
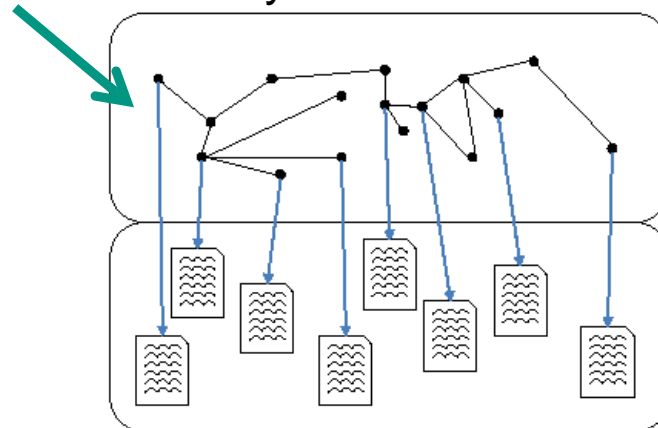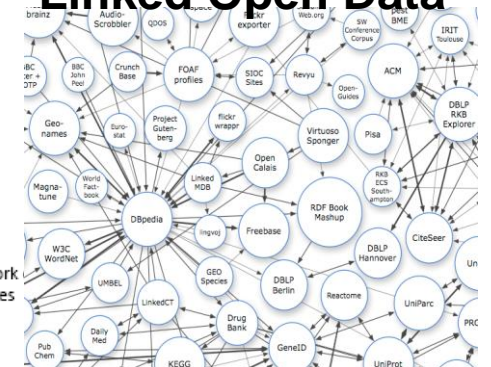
http://linkeddata.org/static/images/lod-datasets_2009-07-14_cropped.png

Imported research objects:
scanned pages

Life-Cycles & Interactions

SW   VRE

Step 3: Data Cleansing

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

Life-Cycles & Interactions

Step 4: Exploring and Analyzing

Internal Reference   Reference to Author   Reference to Literature
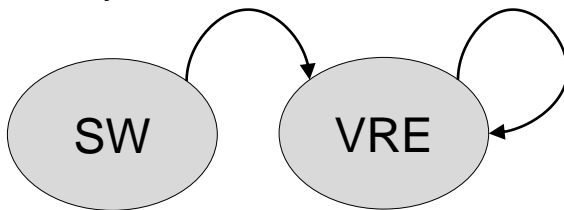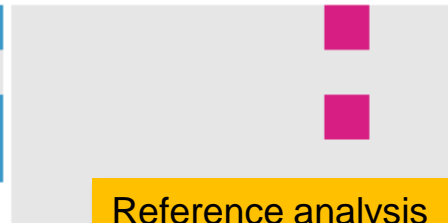
Reference analysis
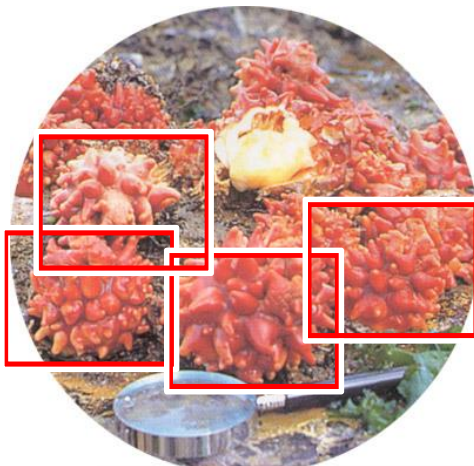
Fine-grained image analysis

Visualization and exploration

top-15 der Frauenberufe:
- Verleger
- ÄrztIn
- Professur Gräzistik
- Psychologie und Ethik
- Äbtissin
- Äbtissin, Regentin
- ErzieherIn
- MusikerIn
- KomponistIn
- SchriftstellerIn, AnstaltsleiterIn
- LehrerIn
- ÄbtissIn
- RegentIn
- PädagogIn
- SchriftstellerIn

Snip	ImageAnnotationText
	Animal 1
	Animal 2
	Animal 3
	Ani...

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal Description Methods (AIFB)

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

SW   VRE
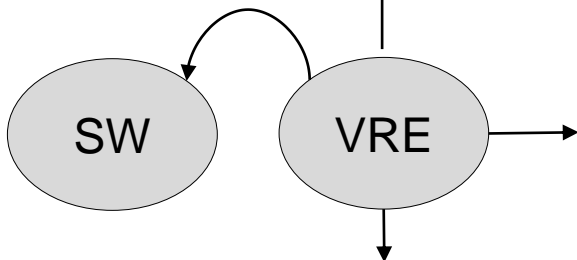
Next Steps

- 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
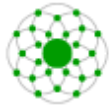
Semantic CorA
Semantic Collaborative Corpora Analysis for Humanities and Social Sciences

Activities:

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

Life-Cycles & Interactions:

Prof. Dr. York Sure-Vetter
Institute of Applied Informatics and Formal Description Methods (AIFB)
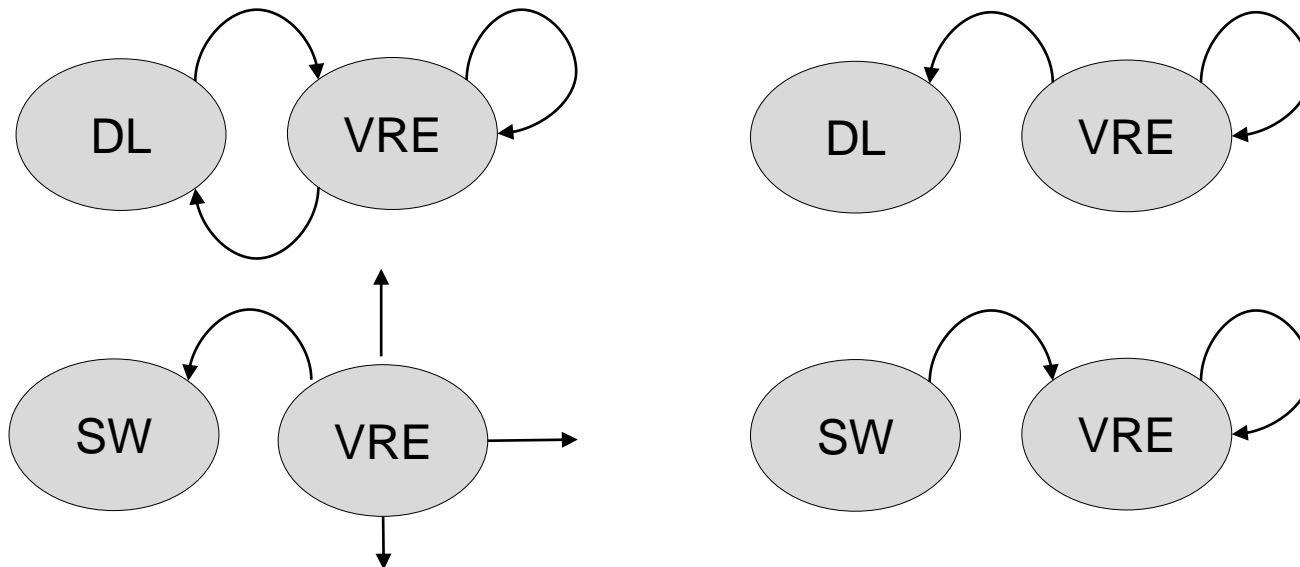
Obama has 60 Millions more **followers** than Trump.

The World Record in Retweeting is currently being held by

2017: New World Record in Retweets

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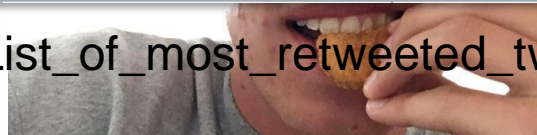
Rank ⬍	Tweet ⬍	Account name ⬍	Retweets (millions) ⬍	Date tweeted ⬍	Notes ⬍
1.	*HELP ME PLEASE. A MAN NEEDS HIS NUGGS* pic.twitter.com/4SrfHmEMo3	@carterjwm[2]	3.641	April 5, 2017	[A]

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-twitters-most-retweeted-tweet

Vielen Dank für Ihre Aufmerksamkeit!

Prof. Dr. York Sure-Vetter

Institute of Applied Informatics and Formal
Description Methods (AIFB)