

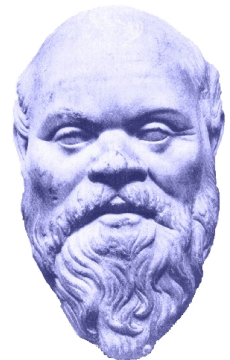


Workshop on Wikipedia Research



Jakob Voss

WikiSym 2006, Odense, Denmark
23 August 2006





Workshop on **Wikipedia Research**

[[**/Intro**]]

{{ **Definition** }}

- * Wikipedia
- * Wikimedia
- * MediaWiki

{{ About [[user:nichtich|me]] }}

- * Active in German Wikipedia since 2002
- * Founding board member of Wikimedia Germany
- * M.A. Computer Science and Library & Information Science
- * System librarian in Göttingen

I bring order out of chaos, I shine light into the dark
because power comes from knowledge just like fire from a spark
and like Gutenberg and Luther with press and pen in hand
I take the message to the masses in a form they understand

Jonathan Rundman: *I'm a librarian*

{{ **Agenda** }}

- * **[[/Intro]]**

- * You are here

- * **[[/Topics]]** & Results

- * What?

- * **[[/Methods]]**

- * How? Where? Who?

{ { **Wikipedia Research?** } }

- * **New discipline**

- * Wikipedia is a singular phenomenon
- * Very heterogenous

- * **General meta-questions**

- * Research vs. development?
- * Inside or outside Wikipedia?
- * What science?

{ { **Wikipedia Research?** } }

“Understanding and fostering the growth of the World Wide Web, both in engineering and societal terms, will require the development of a new interdisciplinary field.”

Berners-Lee, Hall, Hendler, Shadbolt, Weitzner:
Creating a Science of the Web.

In: *Science* 11 Aug 2006. v 313, n 5788, p 769-771

<http://dx.doi.org/10.1126/science.1126902>

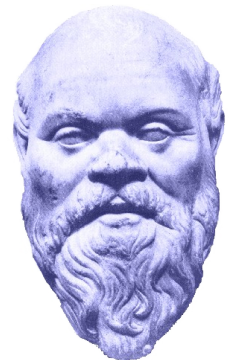


Workshop on Wikipedia Research



Jakob Voss

WikiSym 2006, Odense, Denmark
23 August 2006





Workshop on **Wikipedia Research**

[[**/Topics**]]
(&Results)

{{ The real object of science }}

***Please apologize if I haven't
mentioned your publication!***

Tell me if your paper is missing in the bibliography!

{{ Infrastructure : Bibliography }}

Wiki Research Bibliography

[Wikindx](#) [File](#) [Resources](#) [Metadata](#) [Help](#)

Browse Categories

Bibliography: WIKINDEX Master Bibliography

[Authorship](#) [5] [Content](#) [19] [Ethics](#) [1] [General](#) [36] [History](#) [4] [Impact](#) [14] [Knowledge](#)
[Management](#) [19] [Quality](#) [14] [Relatives](#) [5] [Roles](#) [4] [Semantics & KO](#) [34] [Software](#) [31]
[Teaching & Learning](#) [29] [Usability](#) [3] [Usage as Corpus](#) [6] [User Interaction](#) [6] [Users](#) [27]
[Wikipedia](#) [12]

* <http://bibliography.wikimedia.de/>

* 230 resources (also about wikis in general)

{{ **Topics** }}

* Content*

- * Structure, Semantics, Quality...

* Users

- * Motivation, Roles, Interaction...

* Impact

- * Society, Recipitation...

{{ **Content** }}

- * Structure

- * Network analysis, Visualization...

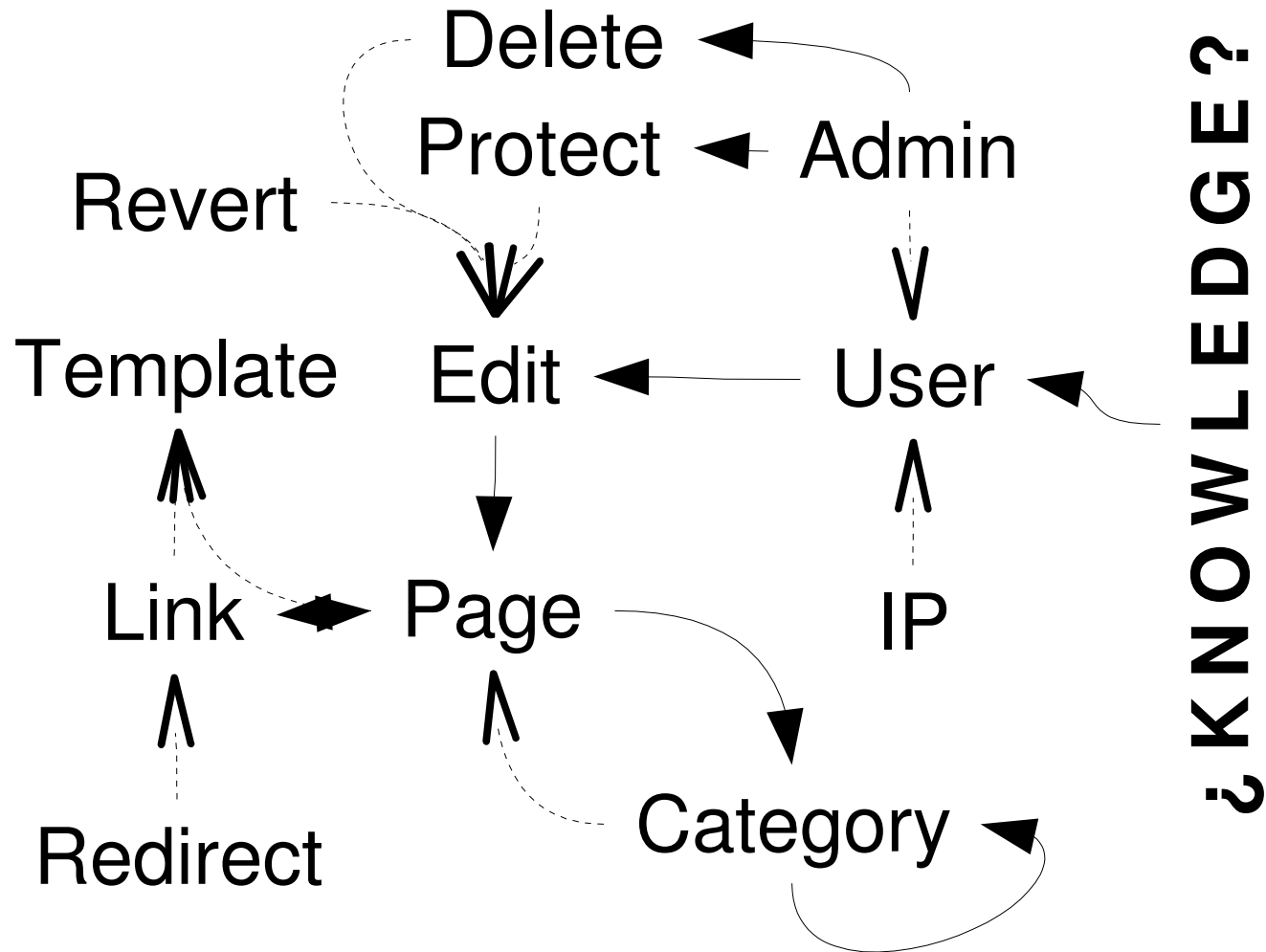
- * Semantics and Knowledge Organization

- * Categories, Semantic Wikis...

- * Quality

- * Other usage (as corpus)

{{ Content : Structure }}



{{ Content : Structure }}

* Network analysis

- * Zlatic et al. 2006: *Wikipedia as complex networks.*
- * Copacci et al 2006: *Preferential attachment in the growth of social networks: the case of Wikipedia.*

* Visualization

- * Hollaway et al 2005: *Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors.*
- * Viégas, Wattenberg, Dave 2004: *Studying Cooperation and Conflict between Authors with History Flow Visualizations*

{{ **Semantics & KO** }}

- * **Semantics:**
computers understand what people meant
- * **Knowledge Organization:**
structure the creation of meaning

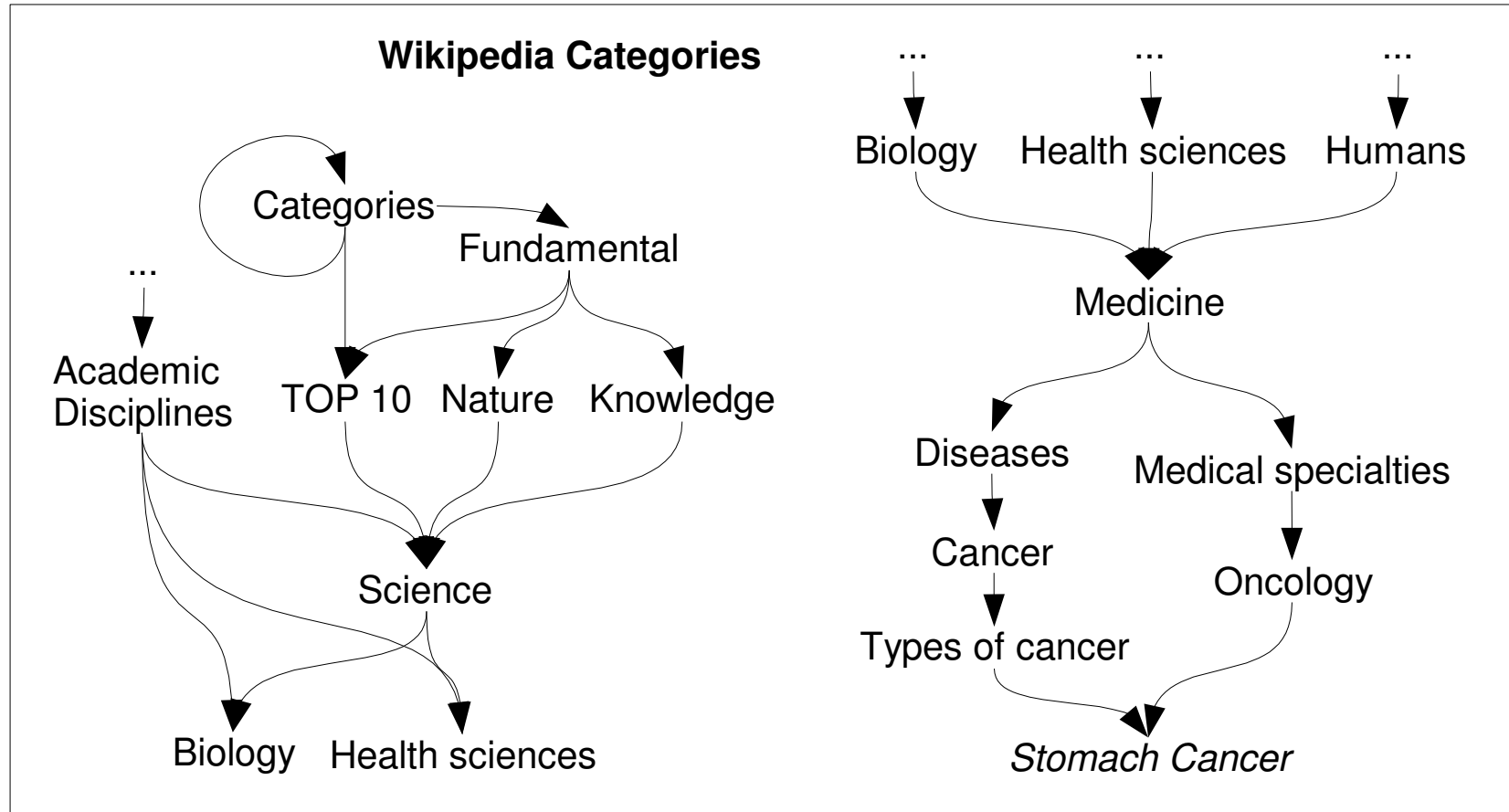
{{ **Semantics & KO** }}

- * **Semantics:**
computers understand what people meant
- * **Knowledge Organization:**
structure the creation of meaning
- * **Knowledge Management:**
people understand what people meant
- * **Teaching & Learning:**
people understand the creation of meaning

{{ **Semantics & KO** }}

- * **Semantics:**
computers understand what people meant
- * **Knowledge Organization:**
structure the creation of meaning
- * **Knowledge Management:**
people understand what people meant
- * **Teaching & Learning:**
people understand the creation of meaning
- * **The fable of Strong AI:**
computers understand the creation of meaning

{{ Semantics : Categories }}



[<http://arxiv.org/cs/0604036> Collaborative thesaurus tagging the Wikipedia way]

{{ Semantics : Categories }}

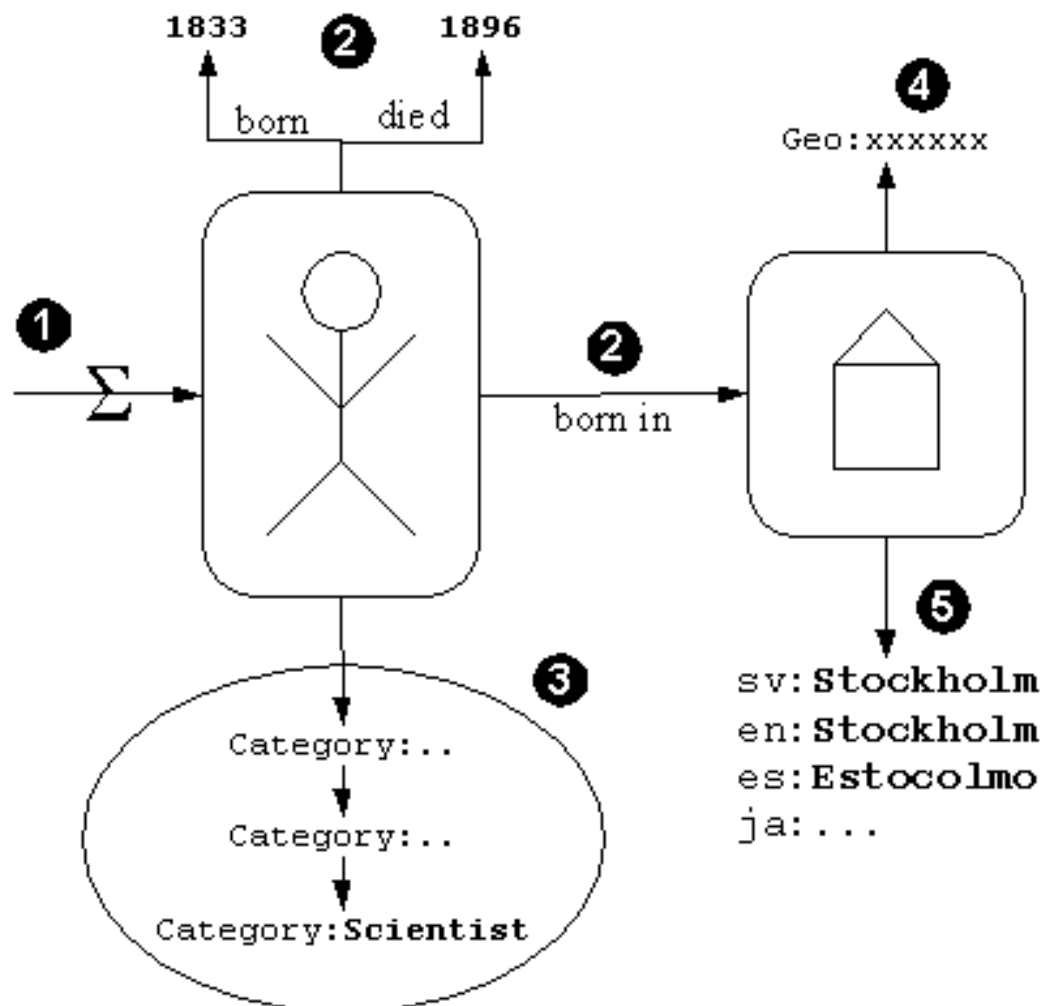
- * Chernov et al. 2006:
Extracting Semantics Relationships between Wikipedia Categories.
- * Holloway et al. 2006:
Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors.
- * Voss 2006:
Collaborative thesaurus tagging the Wikipedia way.

{{ **Semantics** }}

- * Krötzsch, Vrandečić & Völkel 2005: *Wikipedia and the Semantic Web – The missing Links.*
- * Voss 2005: *Metadata with Personendaten and beyond.*
- * Ruiz-Casado et al. 2005: *Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets.*
- * Hepp et al. 2006: *Harvesting Wiki Consensus: Using Wikipedia Entries as Ontology Elements.*
- * Strube & Ponzetto 2006: *WikiRelate! Computing Semantic Relatedness Using Wikipedia.*
- * ...

{{ Semantics : Extract }}

„Famous 19th century scientists of this town?“

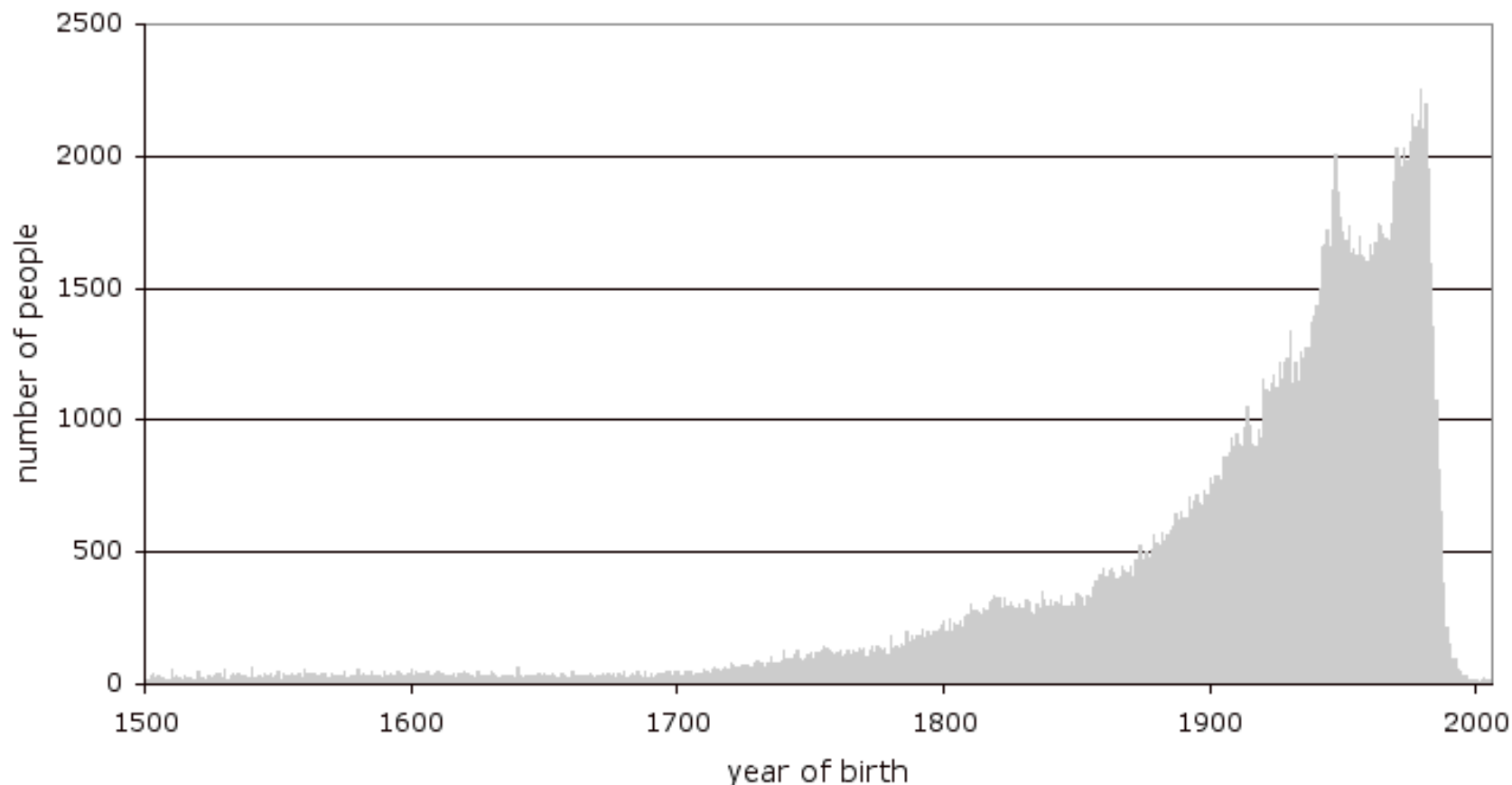


Semantic Information in Wikipedia

- 1** *famous*: sum of **inlinks**
- 2** *19th century*: **personendaten**
~ 1780 < born < 1880
- 3** *scientist*: **categories**
- 4** *this town*: **GEO...**
- 4** *town*: **Interwiki-links**

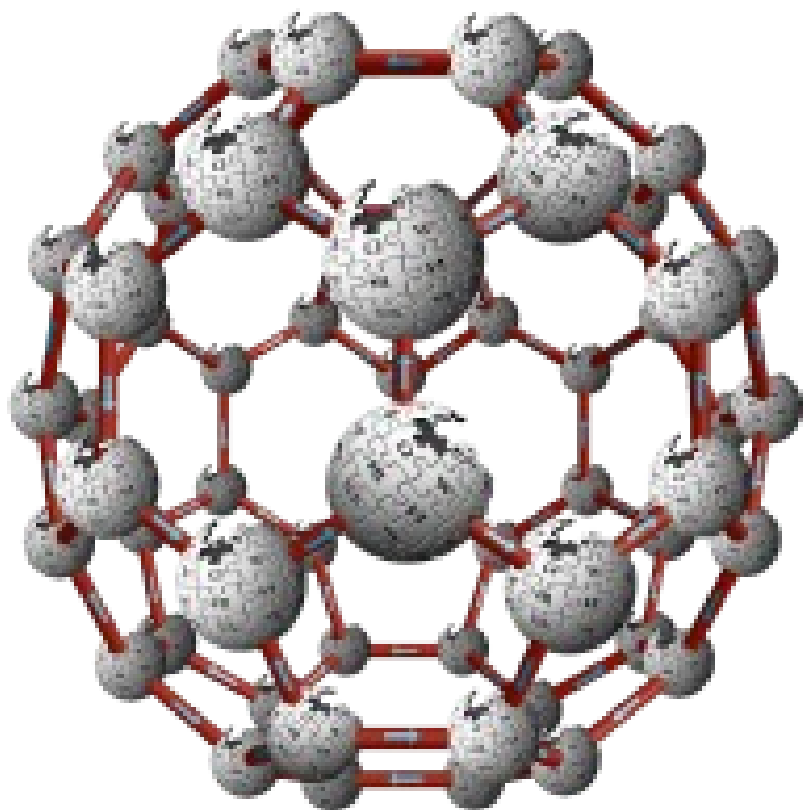
{{ Semantics : Extract }}

biographic articles with year of birth



[<http://en.wikipedia.org/wiki/User:Dsp13> **Biographical coverage in Wikipedia**]

{{ Semantics : Build }}



Relations

[[located in::Fyn]]

Attributes

[[has population:=186 595]]

it's [[Odense]]

[\[http://wiki.ontoworld.org/wiki/Semantic_MediaWiki\]](http://wiki.ontoworld.org/wiki/Semantic_MediaWiki)

{{ Quality }}

* „*Good Wikipedia, Bad Wikipedia*“



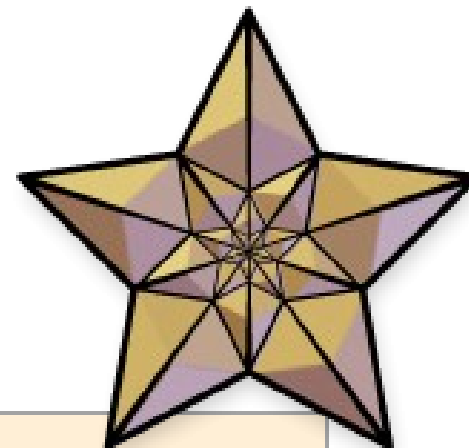
The **factual accuracy** of this article or section is **disputed**.

Please see the relevant discussion on the [talk page](#).



The **neutrality** of this article is **disputed**.

Please see the discussion on the [talk page](#).



* What **is** Quality, Neutrality, Usefulness...?

{{ Quality }}

- * Lots of articles in the press & blogosphere
 - * Seigenthaler '05, House of Representatives '06, etc.
 - * http://en.wikipedia.org/wiki/Criticism_of_Wikipedia
- * Study of Science Journal Nature
 - * Experts reviewed 42 selected articles of Encyclopædia Britannica and Wikipedia
 - * EB had 123 errors (average 2.9)
 - * Wikipedia had 162 errors (averages 3.9)
- * Better *studies* on information quality???

{{ Quality }}

- * Sanger 2004:
Why Wikipedia Must Jettison Its Anti-Elitism.
- * Brändle 2005:
Zu wenige Köche verderben den Brei.
- * Anthony et al. 2005:
Explaining Quality in Internet Collective Goods.
- * Stvilia et al. 2005:
Information Quality: Discussions in Wikipedia.
- * Reagle 2005: *Is the Wikipedia Neutral?*
- * Lanier 2006: *Digital Maoism:
The Hazards of the New Online Collectivism.*

{ { Content : Usage as corpus } }

- * INEX Test corpus for XML Retrieval
<http://inex.is.informatik.uni-duisburg.de/2006/>
- * TREC-13, 2004
<http://staff.science.uva.nl/~gilad/pubs/uams-trec-2004-final-qa.pdf>
- * WiQA at CLEF 2006
<http://ilps.science.uva.nl/WiQA/>
- * Human Knowledge Compression Contest
<http://prize.hutter1.net/>

linguistic (in contrast to logic) approach to AI

{{ **Users** }}

- * Authorship
- * Roles
 - * Powerusers, Fluctuation
- * Interaction, Roles, Motivation...

This is social science – Please help me!

{{ Users : Authorship }}

- * Default diff function
 - * Not very intelligent (newlines, moves etc.)
- * Main authors:
 - * <http://de.wikipedia.org/wiki/Wikipedia:Hauptautoren/S>
- * Better diff feature:
 - * <http://217.147.83.36:9000/history::171=169>
- * ***What is*** authorship in Wikipedia?

{{ Authorship : WhodunitQuery }}

Query History for Text (I'm really new here, but near as i can tell if it says on a webpage it can be reprinted what's the problem?)

User talk:AmiDaniel Load Stop

we still have today because we didn't address them back then. Overall it "seems to me" talking about population is one of the most note worthy things a person can do.

So what you are saying is that it has to be from a published library reference work before it can be included here?

I'm really new here, but near as i can tell if it says on a webpage it can be reprinted what's the problem?

== There's Been Some Confusion ==

{{!unblock}}My apologies. I was trying to fix a bug in VandalProof, whereby [[User:That Guy, From That Show!]] has been unable to log-in. I needed a username with a similar character structure to test that I had fixed the problem; however, I failed to declare it on my userpage as I have with my other alternate accounts. I do, however, confirm that this account ([[User:Der Ami Daniel, With an Exclamation Point!]]) is one of my legitimate alternate accounts. Sorry for all of the inconvenience and confusion.

By the way, my IP address (71.210.212.81.) has been autoblocked as a result of [[User:Der Ami Daniel, With an Exclamation Point!]]'s block, in case I didn't make that clear above. Thanks! [[User:AmiDaniel|AmiDaniel]] ([[User talk:AmiDaniel|talk]]) 03:26, 15 May 2006 (UTC)

:You should be good to go now. -[[User:PS2pcGAMER|PS2pcGAMER]] ([[User talk:PS2pcGAMER|talk]]) 03:50, 15 May 2006 (UTC)

::No problem. I thought you were on a wikibreak for another few days? Maybe I should have left the block in place to enforce it. :p -[[User:PS2pcGAMER|PS2pcGAMER]] ([[User talk:PS2pcGAMER|talk]]) 03:53, 15 May 2006 (UTC)

:::Now... you are all good to go :) --[[User:Lightdarkness|light]][[User:Lightdarkness|darkness]]^{[[User_talk:Lightdarkness|talk]]} 04:10, 15 May 2006 (UTC)

Oldid	User	Date	
<input checked="" type="checkbox"/>	cur	Lightdarkness	2:10, 14 May 2006
<input checked="" type="checkbox"/>	53258919	PS2pcGAMER	1:53, 14 May 2006
<input checked="" type="checkbox"/>	53258551	PS2pcGAMER	1:50, 14 May 2006
<input checked="" type="checkbox"/>	53255787	AmiDaniel	1:26, 14 May 2006
<input checked="" type="checkbox"/>	53254671	AmiDaniel	1:17, 14 May 2006
<input checked="" type="checkbox"/>	53213638	Lee Wells	6:11, 14 May 2006
<input checked="" type="checkbox"/>	53211518	AmiDaniel	5:55, 14 May 2006
<input checked="" type="checkbox"/>	53211005	AmiDaniel	5:52, 14 May 2006
<input checked="" type="checkbox"/>	53210875	Lee Wells	5:51, 14 May 2006
<input checked="" type="checkbox"/>	53210622	Lee Wells	5:49, 14 May 2006
<input checked="" type="checkbox"/>	53175270	AmiDaniel	1:37, 14 May 2006
<input checked="" type="checkbox"/>	53174413	AmiDaniel	1:31, 14 May 2006
<input checked="" type="checkbox"/>	53141661	220.240.131.192	6:23, 14 May 2006
<input checked="" type="checkbox"/>	53097187	AmiDaniel	1:24, 13 May 2006
<input checked="" type="checkbox"/>	53096848	24.226.54.173	1:21, 13 May 2006
<input checked="" type="checkbox"/>	53096325	24.226.54.173	1:17, 13 May 2006
<input checked="" type="checkbox"/>	53060774	AmiDaniel	6:15, 13 May 2006
<input checked="" type="checkbox"/>	53045608	AmiDaniel	4:23, 13 May 2006
<input checked="" type="checkbox"/>	53045277	AmiDaniel	4:21, 13 May 2006
<input checked="" type="checkbox"/>	53044932	AmiDaniel	4:19, 13 May 2006
<input checked="" type="checkbox"/>	53043113	AmiDaniel	4:07, 13 May 2006
<input checked="" type="checkbox"/>	53042964	AmiDaniel	4:06, 13 May 2006
<input checked="" type="checkbox"/>	53042138	AmiDaniel	4:00, 13 May 2006

[<http://en.wikipedia.org/wiki/User:AmiDaniel/WhodunitQuery> **WhodunitQuery**]

{{ Powerusers }}

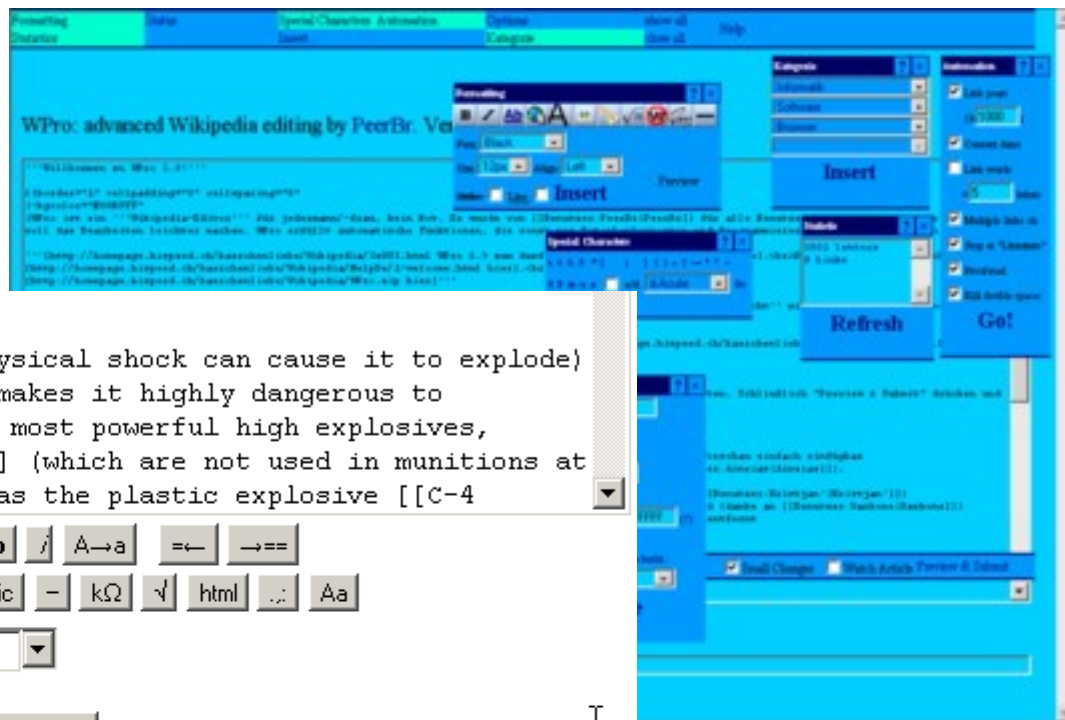
- * Hardcore Wikipedians / Core community
 - * Definition ?
 - * Elite?
 - * Influence?
 - * Addiction?
- * may have other working environments, for instance **user scripts**
 - * [<http://en.wikipedia.org/wiki/Wikipedia:Scripts>]
- * may have other focus for instance arbitration, VfD, Wikimedia...

{{ Powerusers : Definition }}

- * Above average activity
 - * What average?
 - * What activity?
- * Straight Power-law (Lotka's law)
 - * $<10\%$ accounts do $>90\%$ edits
 - * $>50\%$ accounts do < 10 edits

{{ Powerusers : Editors }}

[<http://de.wikipedia.org/wiki/Benutzer:PeerBr/WPro> **WPro**]



```
== Instability and desensitization ==  
In its pure form, it is a [[contact explosive]] (i.e., physical shock can cause it to explode) and degrades over time to even more unstable forms. This makes it highly dangerous to transport or use. In this undiluted form it is one of the most powerful high explosives, comparable to the military explosives [[RDX]] and [[PETN]] (which are not used in munitions at full concentration because of their sensitivity) as well as the plastic explosive [[C-4
```

Get ←Find Find→ ↑↓ ← → Undo all **b** / A→a == ← → ==
All ←Repl. Repl.→ Case Regexp Fix: Basic - kΩ √ html ⋮ Aa

Edit summary:

This is a minor edit (?) Watch this page

Save page Instant: Server: [Editing help](#) (opens in new window)

"Nitroglycerin", also known as "nitroglycerine", "trinitroglycerin", and "glyceryl trinitrate", is a [[chemical compound]]. It is an extremely dangerous heavy, colorless, poisonous, oily, explosive liquid obtained by [[nitration | nitrating]] [[glycerol]]. >
Nitroglycerin is used in the manufacture of [[explosive]]s, specifically [[dynamite]], and as such is employed in the [[construction]] and [[demolition]] industries. It is also used medically as a [[vasodilator]] to treat [[heart]] conditions. It is colored yellow when it is decomposing due to acidic [[pH]].

[<http://en.wikipedia.org/wiki/User:Cacycle/editor> **WikEd**]

{{ Powerusers : AWB }}

The screenshot displays the AutoWikiBrowser (AWB) application interface. The main window shows a Wikipedia article titled "Demenika" in Greek, with the current revision and "Your text" side-by-side. The article text is partially highlighted in green. Below the article, there are sections for "Line 1:", "Line 28:", and "See also".

On the right side, a "Wiki Data Dump Searcher" window is open. It features a search interface with options for "Text matches" (Simple text search, Regular expression search) and "Other" (Don't check length, Don't count links, Limit no. of results, Article starts with). A list of search results is displayed, including "Aquaculture", "Cincinnati Bengals", "Mahjong", "NBA (disambiguation)", "Rudolf Diesel", "Solar power", "Seventh-day Adventist Church", "The Matrix", "Howard Walter Florey", "Religion and sexuality", "Boston Celtics", "70 mm film", "Madison, Wisconsin", "Isaac Brock", "Missoula, Montana", "Leeds United F.C.", "Culture of Iraq", and "Nuclear fuel cycle".

At the bottom left, there is a "Make list" window showing a list of items to be processed, including "Edward Franklin Abee", "Corporation tax", "Labrador tea", "Galen Peirson", "Milton Brown", "Machelen (Zulte)", "Pahon", "Kazza", "Leonardo Leonardo", "Thiele", "HDF", "DigitalGlobe", "Richard Rodney Bennett", "Victoria Park (Warrington)", "Coachella Festival", "Discipline (BDSM)", "Eternity puzzle", "Carebear", and "Spinal anaesthesia".

The bottom status bar shows "Ready to save Bot timer: 0" and "en.wikipedia User:Bluebot".

[<http://en.wikipedia.org/wiki/Wikipedia:AutoWikiBrowser> **AutoWikiBrowser**]

Jakob Voss : **Workshop on Wikipedia Research. WikiSym 2006-08-23**, Odense, Denmark

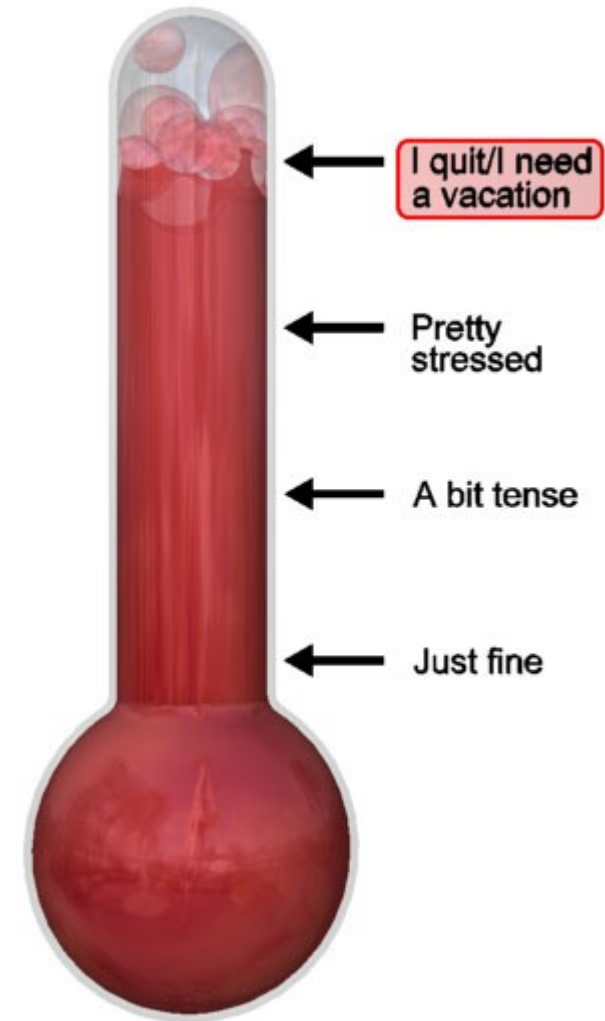
{{ Powerusers : VandalProof }}

The screenshot displays the VandalProof 1.2 (AmiDaniel) application window. The interface is divided into several sections:

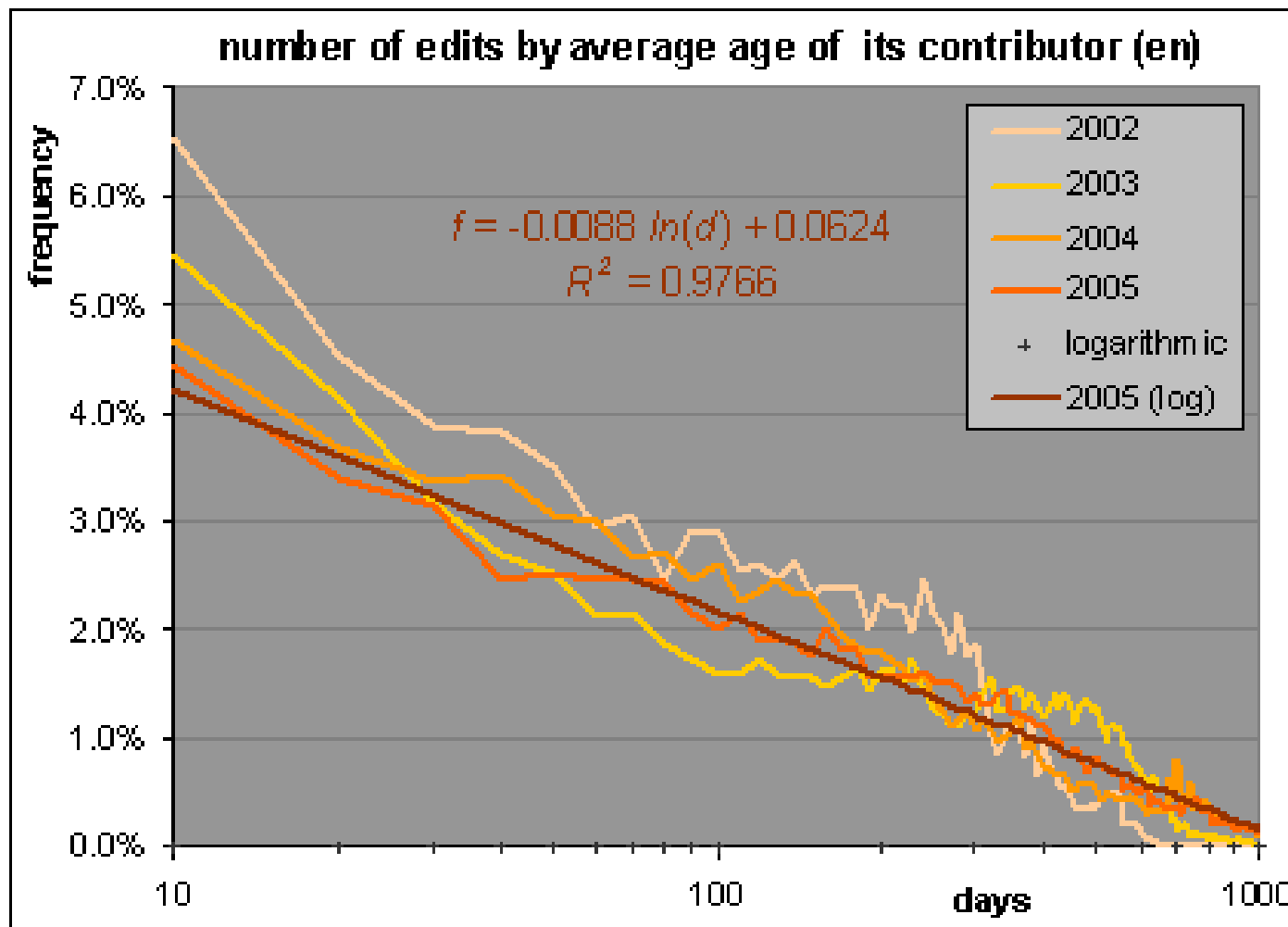
- Top Bar:** Contains navigation links (File, Navigate, Vandalism Log, Article Tools, User Tools, Help) and a series of "Rollback" buttons for different revision ranges (e.g., Rollback ((test1-n)), Rollback ((test1a-n)), etc.).
- Left Panel:** Features a "Recent Changes" tab with a table of article changes. The table has columns for "Article", "User", and "Summary". A yellow highlight is present on the row for user 222.154.55.105.
- Right Panel:** Shows the Wikipedia article for "MMORPG". It includes the article title, a description, and a list of revisions. A yellow highlight is visible on the article text.
- Bottom Panel:** Shows the Windows taskbar with various open applications, including "VandalPro...", "10,000 Days...", "Windows Me...", "Binary Funct...", "VP newlook2...", "My Playlist", "Media", "VBasic", and a system clock showing 17:01.

[<http://en.wikipedia.org/wiki/User:AmiDaniel/VandalProof> **VandalProof**]

{{ Users : Fluctuation }}



{{ Users : Newbies and edits }}



~25% of all edits are by “newbies” that have done their first edit less than 100 days ago

[<http://wm.sieheauch.de/?p=44> **Days since first edit**]

[http://meta.wikimedia.org/wiki/Days_since_first_edit **Days since first edit**]

{{ **Users** }}

* Motivation

- * Frost 2005: *Zivilgesellschaftliches Engagement in virtuellen Gemeinschaften*
- * Forte et al 2005: *Why do people write for Wikipedia?*

* Roles

- * Ciffolilli 2003: *Phantom authority, self-selective recruitment and retention of members in virtual communities*
- * Reagle 2006: *Do as I do: leadership in the Wikipedia*

* Collaboration, Discussion

- * Matei et al 2006: *Ambiguity and conflict in the Wikipedian knowledge production system*
- * Pentzold et al 2006: *Focault at Wiki.*

{{ **Users : More** }}

- * More roles
- * Usage patterns
- * ...
- * **BUT: Privacy!**

{{ **Semantics & KO** }}

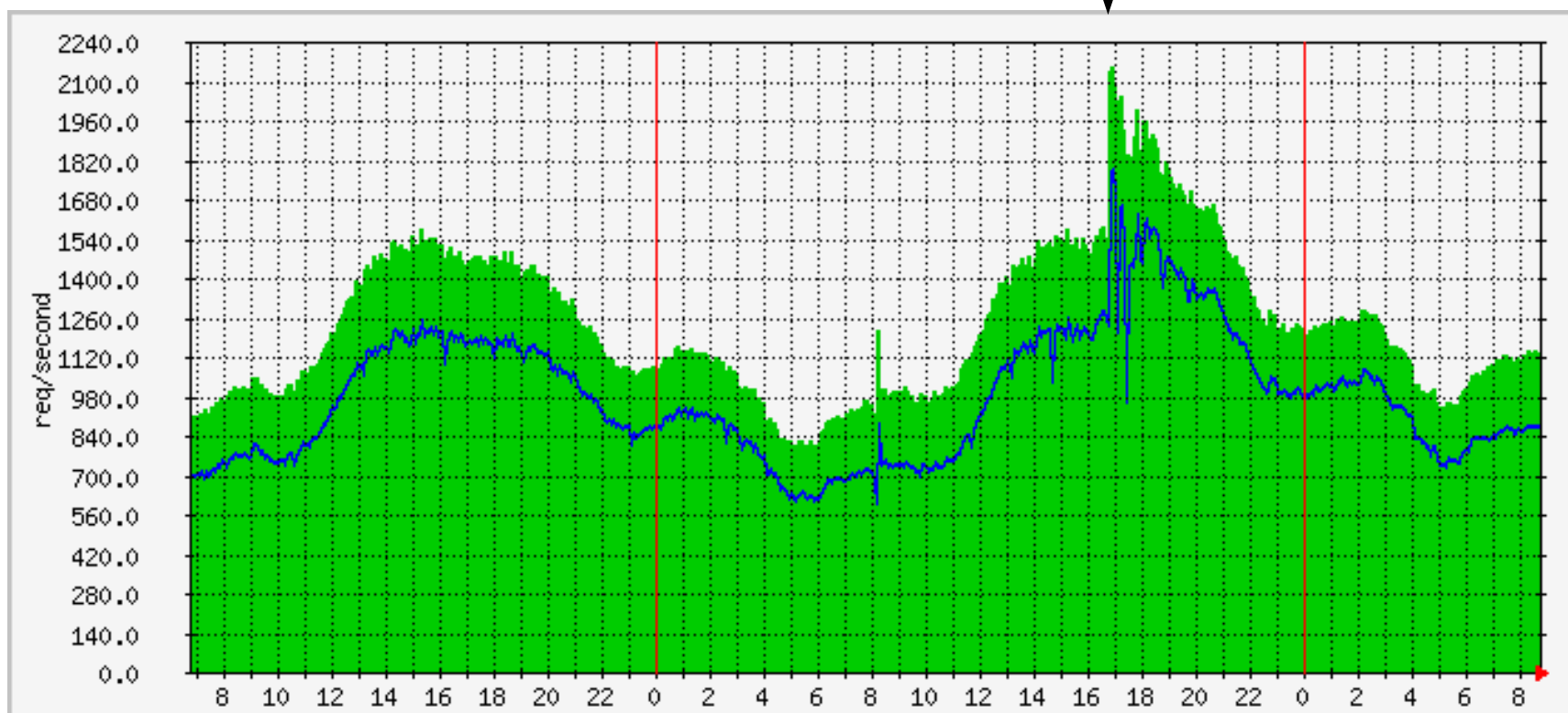
- * **Semantics:**
computers understand what people meant
- * **Knowledge Organization:**
structure the creation of meaning
- * **Knowledge Management:**
people understand what people meant
- * **Teaching & Learning:**
people understand the creation of meaning
- * **The fable of Strong AI:**
computers understand the creation of meaning

{{ **Impact** }}

- * Attention
- * Open Content, open society?
- * History of Wikipedia and Encyclopaedias
- * Usability
- * Ethics
- * ...

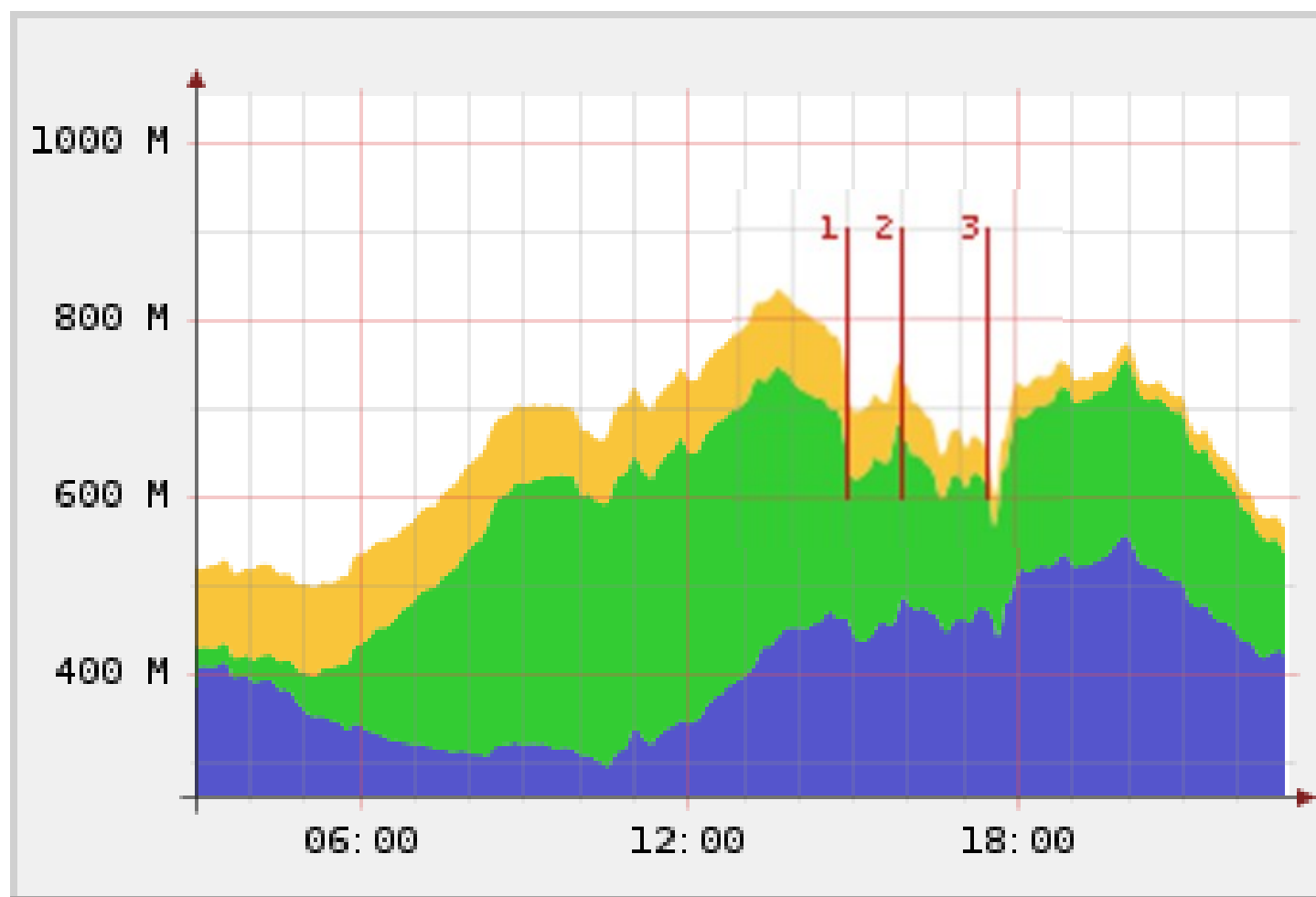
{{ Impact : Attention }}

2005-05-19: *habemus papam*



{{ Impact : Attention }}

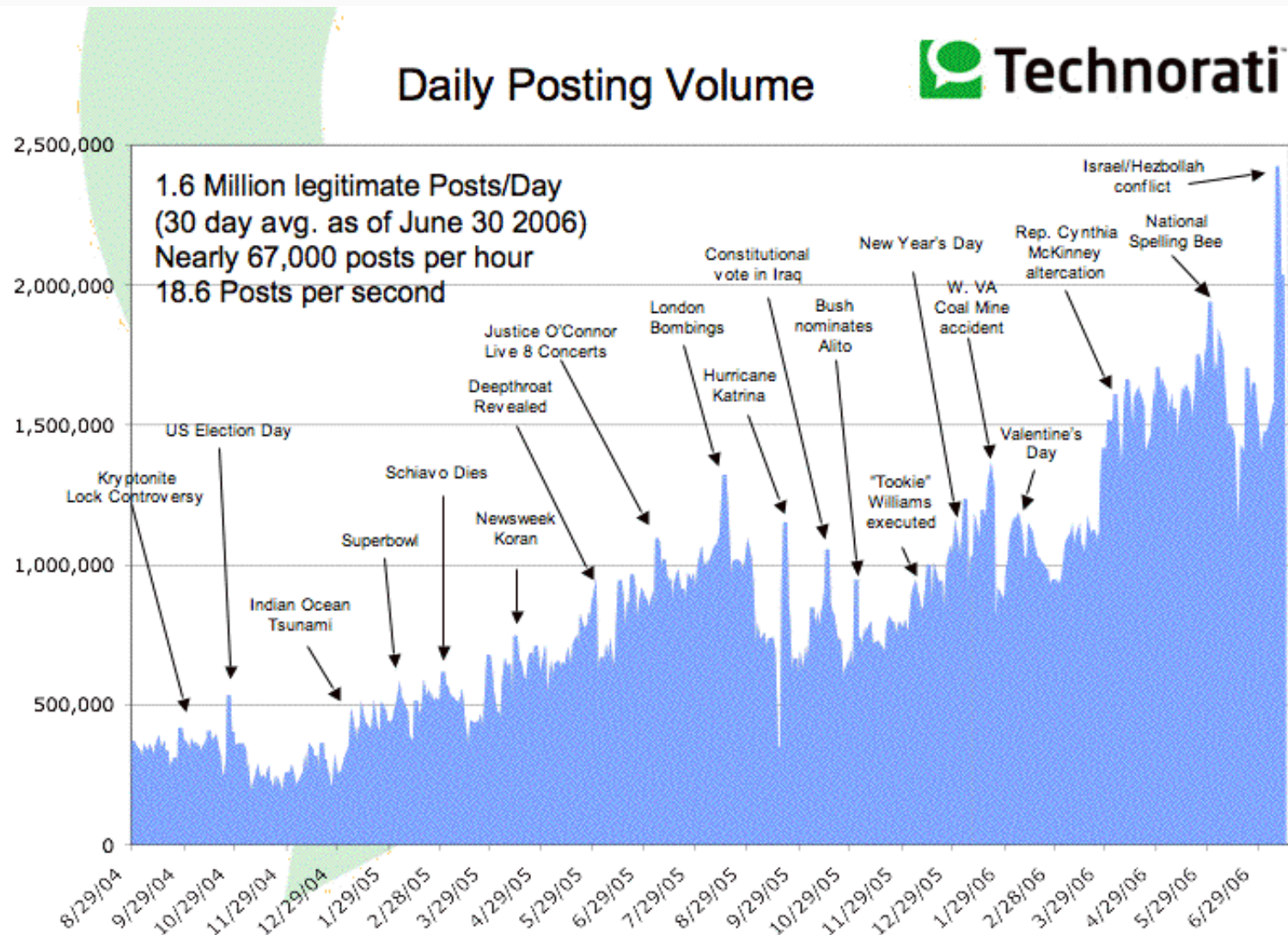
2006-06-30: Germany-Argentina (start, half, penalty)



[<http://wikipedistik.de/2006/07/01/enzyklopaedie-fussball-01/> **Tim Bartel**]

Jakob Voss : **Workshop on Wikipedia Research**. *WikiSym 2006-08-23*, Odense, Denmark

{{ Impact : Attention }}



[<http://www.sifry.com/alerts/archives/000436.html> **State of the Blogosphere**]

{ { **Research Needed** } }

- * Comparision between different languages (intercultural studies)
- * Impact of gender to Wikipedia and knowledge (gender studies)
- * Ethical aspects (responsability, privacy...)
- * *...what are you interested in?...*

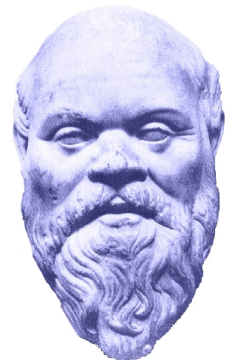


Workshop on Wikipedia Research



Jakob Voss

WikiSym 2006, Odense, Denmark
23 August 2006





**Workshop on
Wikipedia Research**

[[/Methods]]

{{ **Agenda** }}

- * Scientists

- * Methods

 - * statistics, surveys, interviews

 - * bots, API, database

- * Infrastructure

 - * Mailing lists, blogs, conferences, bibliography

{{ **Scientists** }}

- * Community
- * Students (Diploma/Master theses)
- * Journalists
- * PhD theses
- * University projects
- * Papers at journals and conferences

{{ **Scientists : Disciplines** }}

- * Media Studies, Journalism
- * Psychology, Sociology
- * Library & Information Science
- * Computer Science
- * Pedagogy, Knowledge Management

- * Law, Economics, Gender studies

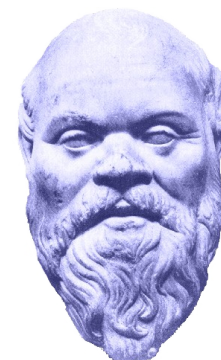
{{ **Methods** }}

- * Comparisons
- * Statistics
- * Surveys
- * Interviews
- * Literature review

- * ...depending on the discipline...
- * *Don't use tools if you don't know how they work!*

{{ **Philosophy** }}

- * Foucault @ wiki
- * Habermas's Theory of Communicative Action
- * Heidegger's Hammer



{{ **Statistics** }}

- * Statistics of Wikipedia's size and growth
<http://stats.wikimedia.org/>
- * Statistics of communication on mailing lists:
<http://www.infodisiac.com/Wikipedia/ScanMail/>

Erik's the best! :-)

- * Many simple / smaller statistics in the projects
- * Don't even trust your own statistics!

{{ Statistics : Tools }}

- * User edit counter

<http://tools.wikimedia.de/~interiot/cgi-bin/>

PRIVACY!

- * CoCat

<http://tools.wikimedia.de/~voj/cgi-bin/coca>

slow!

- * Build in MediaWiki methods

{{ Statistics : Usage }}

- * Too much
- * No stats, no manipulation
- * WikiCharts
<http://tools.wikimedia.de/~leon/stats/wikicharts/>
- * Main page, Current news (Hisbollah, Harry Potter...), Countries, Sexuality

{{ Surveys }}

- * Difficult

 - * Validity

 - * Acceptance

- * Motivation of Contributors (2005)

 - <http://wy2x05.psychologie.uni-wuerzburg.de/ao/research/wikipedia.php?lang=en>
=> 88% men, 50% singles

- * General (Official) User Survey

 - http://meta.wikimedia.org/wiki/General_User_Survey

{{ **Bots** }}

- * (semi)automatic processes interacting with Wikipedia – navigation and editing
- * get a special account flag
- * Correct spelling, add categories, change links, detect vandalism etc.

{ { **Bot Frameworks** } }

- * [<http://pywikipediabot.sourceforge.net/>
PyWikipediaBot] (Python)
- * [<http://sourceforge.net/projects/dotnetwikibot>
DotNetWikiBot Framework] (.NET, C#)
- * **Query API**

{{ Query API }}

- * Query data directly from MediaWiki
- * [<http://en.wikipedia.org/w/query.php> en]
- * Returns many, many properties
- * Multiple output formats (XML, JSON, TXT...)

{{ API : Example }}

#Question: Main authors of an article?

#Idea: Count number of edits

#Implementation: evaluate revision

{ { API : Implementation } }

* <http://en.wikipedia.org/w/query.php>
?**what**=revisions &**format**=json (for testing: jsonfm)
&**rqlimit**=200 &**titles**=*article's title*

```
* { "pages":  
    "3173243": {  
        "title": "article's title"  
        "revisions": [  
            {  
                "timestamp": "2006-08-13T07:53:48Z",  
                "user": "user's name"  
            }, {  
                "timestamp": "2006-08-13T07:41:02Z",  
                "anon": ""  
                "user": "ip number"  
            }  
        ]  
    }  
}
```

{ { API : Implementation } }

```
...  
eval("var queryResult=" + data);  
var page = anyChild(queryResult.pages);  
var revisions = page.revisions;  
...  
for(var i=0; i<revisions.length; i++) {  
    var rev = revisions[i];  
  
    if (rev.minor != null) minorCount++;  
    if (rev.anon != null) anonCount++;  
    else  
        if (!editorCounts[rev.user]) {  
            editorCounts[rev.user] = 1;  
        } else {  
            editorCounts[rev.user]++;  
        }  
}  
...  
...
```

{{ API : Example }}

#Question: Main authors of an article?

#Idea: Count number of edits

#Implementation: evaluate revision

#Conclusion: edit-counting is inaccurate

{{ Database }}



- * Toolserver

 - <http://meta.wikipedia.org/wiki/Toolserver>

- * (almost) direct access (read-only)

- * MediaWiki database layout

 - http://meta.wikimedia.org/wiki/Database_layout

{{ Database : Dump }}

* Database Dumps (XML)

<http://download.wikimedia.org>

{{ Database : Example }}

* **Question:** Sizes of categories?

* **Idea:** table categorylinks

cl_from

cl_to

cl_sortkey

cl_timestamp

* **Method 1:** MySQL statement

```
SELECT cl_to, COUNT(cl_from) FROM  
categorylinks GROUP BY cl_to
```

* ...and wait

{{ Database : Example }}

* **Method 2:** Short script

#Extract categorylinks

```
echo "SELECT cl_from, cl_to FROM  
categorylinks" | mysql enwiki_p > catlinks
```

#Skip 1st line & 2nd column, count, sort

```
awk "NR!=1 {print \$2;}" catlinks |  
perl count.pl |  
sort -r -n -k 2 > catsizes
```

#Show result

```
head catsizes
```

{{ Database : Example }}

- 12 NPOV_disputes
- 12 Articles_which_may_contain_original_research
- 12 Forms_of_government
- 12 Anarchism
- 12 Political_ideologies
- 12 Political_philosophies
- 12 Social_philosophy

[Categories: NPOV disputes](#) | [Articles which may contain original research](#) | [Forms of government](#) | [Anarchism](#) | [Political ideologies](#) | [Political philosophies](#) | [Social philosophy](#)

{{ Database : Example }}

```
#!/usr/bin/perl
# count.pl - create histogram

my %counter = ();

while (<STDIN>) {
    $_ =~ s/\n$//; # remove linebreak
    $counter{$_}++;
}
foreach $key (keys %counter) {
    print "$key\t" . $counter{$key} . "\n";
}
```

{{ Database : Example : Results }}

GFDL_images.....	102445
Living_people.....	97298
Disambiguation.....	55204
Redirects_from_US_postal_abbreviation.....	40890
User-created_public_domain_images.....	36300
Album_covers.....	34007
Public_domain_images.....	29220
Logos.....	27666
Free_use_images.....	20310
Promotional_images.....	20013
User_en.....	17015
Screenshots_of_television.....	15628
Screenshots_of_films.....	13862
Open_proxies_blocked_on_Wikipedia.....	13152
Unprintworthy_redirects.....	12755
1911_Britannica.....	11991
Screenshots_of_computer_and_video_games....	11648
Articles_with_unsourced_statements.....	11553
Articles_lacking_sources.....	10857

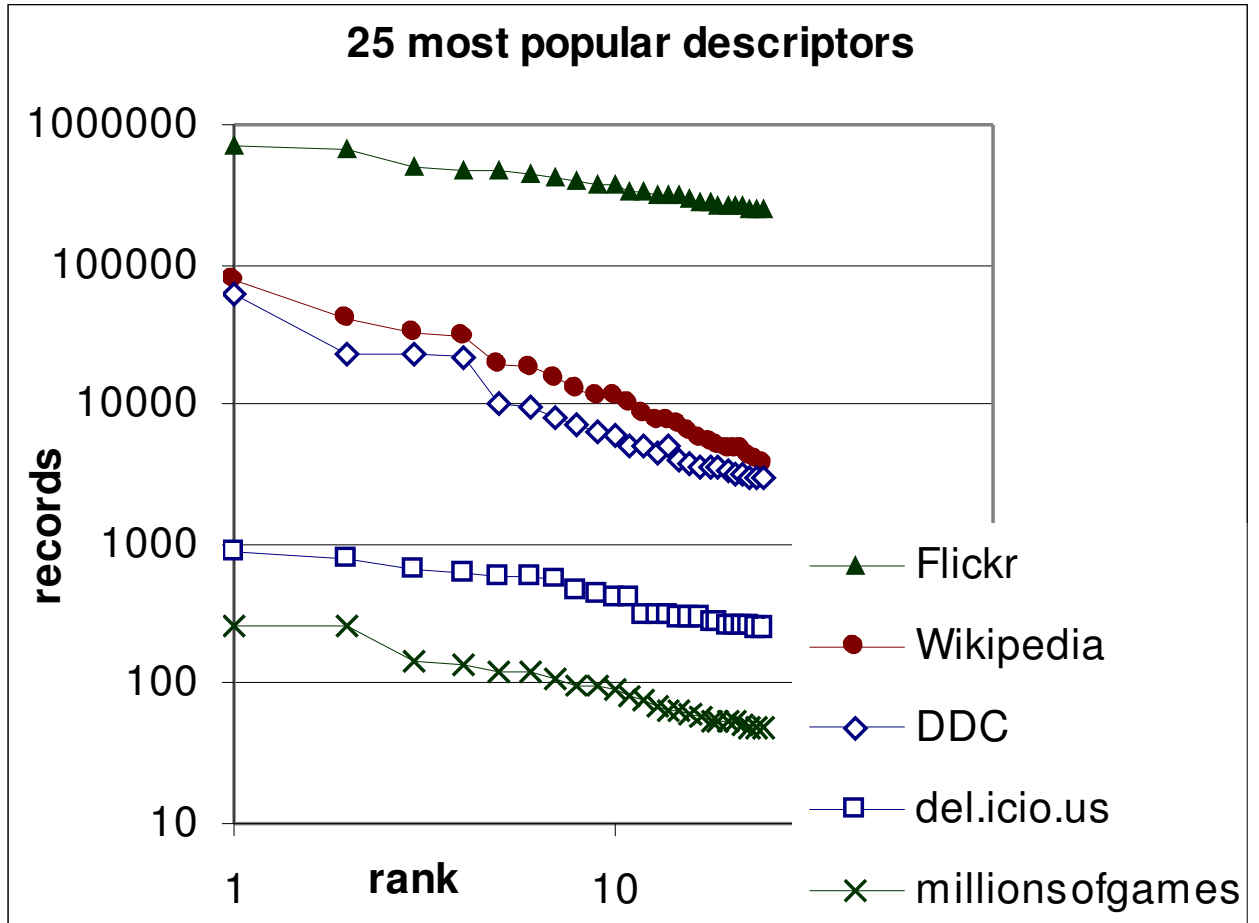
{{ Database : Example : Results }}

* Pages per
Category

power law

Wikipedia $\lambda = 0.96$

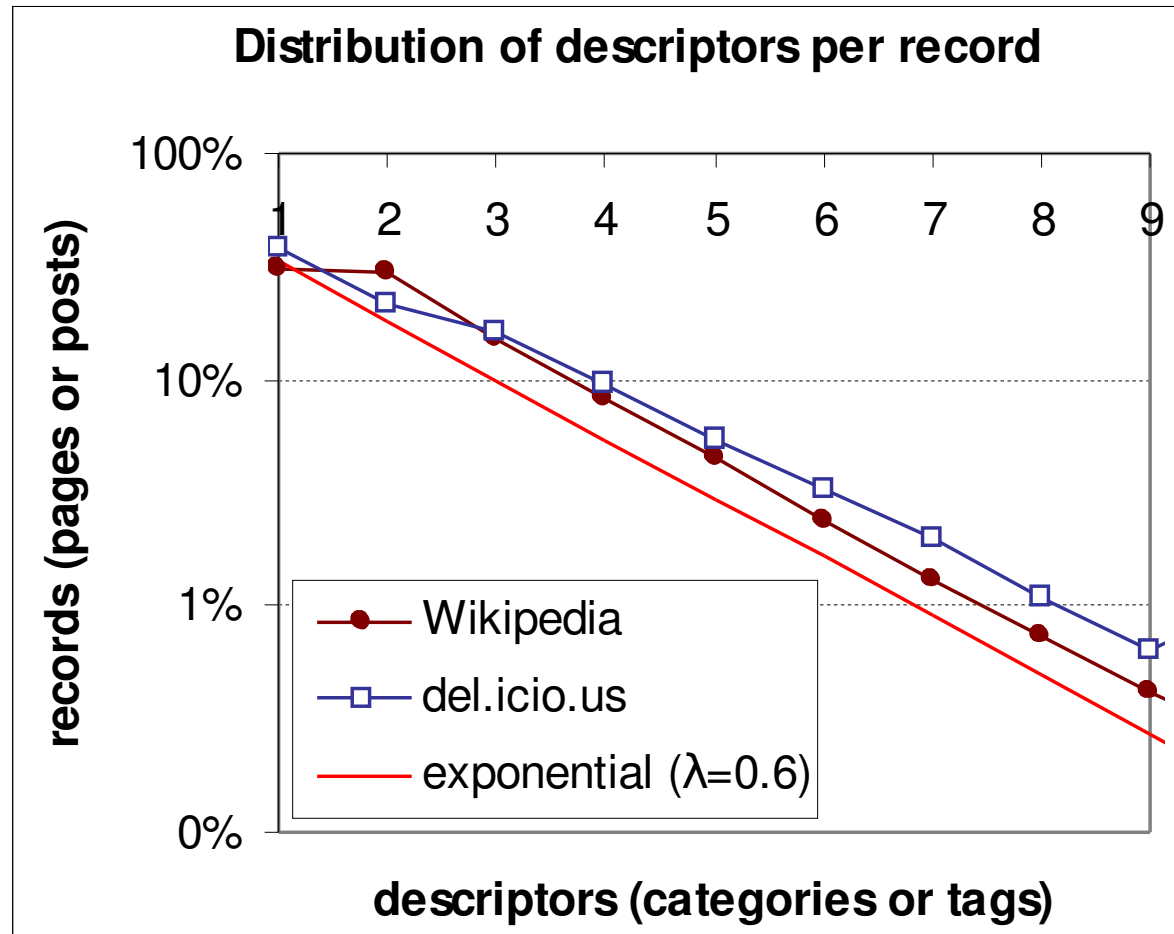
DDC $\lambda = 0.94$



{{ Database : Example : Results }}

* Categories
per page:

**exponential
distribution**



{{ Infrastructure }}

- * Mailinglists
- * Weblogs
- * Wikis
- * Wiki Research Bibliography
- * Conferences
- * Journals
- * Institutes ?

{ { Infrastructure : Mailing lists } }

- * **wiki-research-l**

<http://mail.wikipedia.org/pipermail/wiki-research-l/>

- * **wiki-research**

<http://www.wikisym.org/cgi-bin/mailman/listinfo/wiki-research>

- * **wikipedia-l, foundation-l...**

{{ Infrastructure : Wikis }}

- * [\[\[meta:Research\]\]](#)
- * [\[\[meta:Research/Social_Research_Collaborations\]\]](#)
- * [\[\[en:Wikipedia:WikiProject Wikidemia\]\]](#)
- * [\[\[de:Wikipedia:Wikipedistik\]\]](#)

- * Many distributed pages

{{ Infrastructure : Blogs }}

- * Joseph Reagle
<http://reagle.org/joseph/blog/>
- * Wikimetrics (me)
<http://wm.sieheauch.de/>
- * Wikipedistik (Tim Bartel)
<http://www.wikipedistik.de/>
- * Andrew Lih
<http://www.andrewlih.com/blog/>
- * ...

{{ Infrastructure : Conferences }} }



WIKIMANIA 2006

The International Wikimedia Conference

August 4 - 6

Cambridge, Massachusetts, USA

The International Symposium on Wikis

- * **WikiSym 2005**
- * **WikiSym 2006**
- * **[[WikiSym 2007]]**

{{ Infrastructure : Bibliography }}

Wiki Research Bibliography

[Wikindx](#) [File](#) [Resources](#) [Metadata](#) [Help](#)

Browse Categories

Bibliography: WIKINDEX Master Bibliography

[Authorship](#) [5] [Content](#) [19] [Ethics](#) [1] [General](#) [36] [History](#) [4] [Impact](#) [14] [Knowledge](#)
[Management](#) [19] [Quality](#) [14] [Relatives](#) [5] [Roles](#) [4] [Semantics & KO](#) [34] [Software](#) [31]
[Teaching & Learning](#) [29] [Usability](#) [3] [Usage as Corpus](#) [6] [User Interaction](#) [6] [Users](#) [27]
[Wikipedia](#) [12]

* <http://bibliography.wikimedia.de/>

* 230 resources (also about wikis in general)