

# Wiki'er og Wikipedia

Finn Årup Nielsen

DTU Compute  
Technical University of Denmark

April 23, 2013

## Resumé

Wiki'er er efterhånden alle vegne. De nyeste udviklinger muliggør strukturering af data, så man kan opbygge sin egne specialiserede wiki-baserede database eller trække på den enorme mængde information i Wikipedia f.eks. via DBpedia, en af de centrale data resourcer i den såkaldte Linked Open Data cloud, som efterhånden linker alverdens åbne data sammen.

På gå-hjem-mødet vil den danske Wikipedia også blive inddraget: Hvordan man redigerer, hvilke problemer der opstår og hvorfor.

# Oversigt

Wikipedia forskning: f.eks. i brugermotivation, prædiktion, værktøjer, botter.

Brug af Wikipedia: DBpedia.

Wiki-teknologi: Semantic MediaWiki, Wikidata, analyse af data i en wiki

# Wikipedia vil uundværlig fejle!

“[...] Wikipedia inevitably will be overtaken by the gamers and the marketers to the point where it will lose all credibility. There are so many examples of community-driven communication tools that ultimately were taken over—USENET and the Open Directory Project are two that come top-of mind—that I didn't imagine that my statement would be controversial or debatable.”

# Wikipedia vil uundværlig fejle!

“[...] Wikipedia inevitably will be overtaken by the gamers and the marketers to the point where it will lose all credibility. There are so many examples of community-driven communication tools that ultimately were taken over—USENET and the Open Directory Project are two that come top-of mind—that I didn't imagine that my statement would be controversial or debatable.”

— (Goldman, 2005)

## Wikipedia's nulte lov:

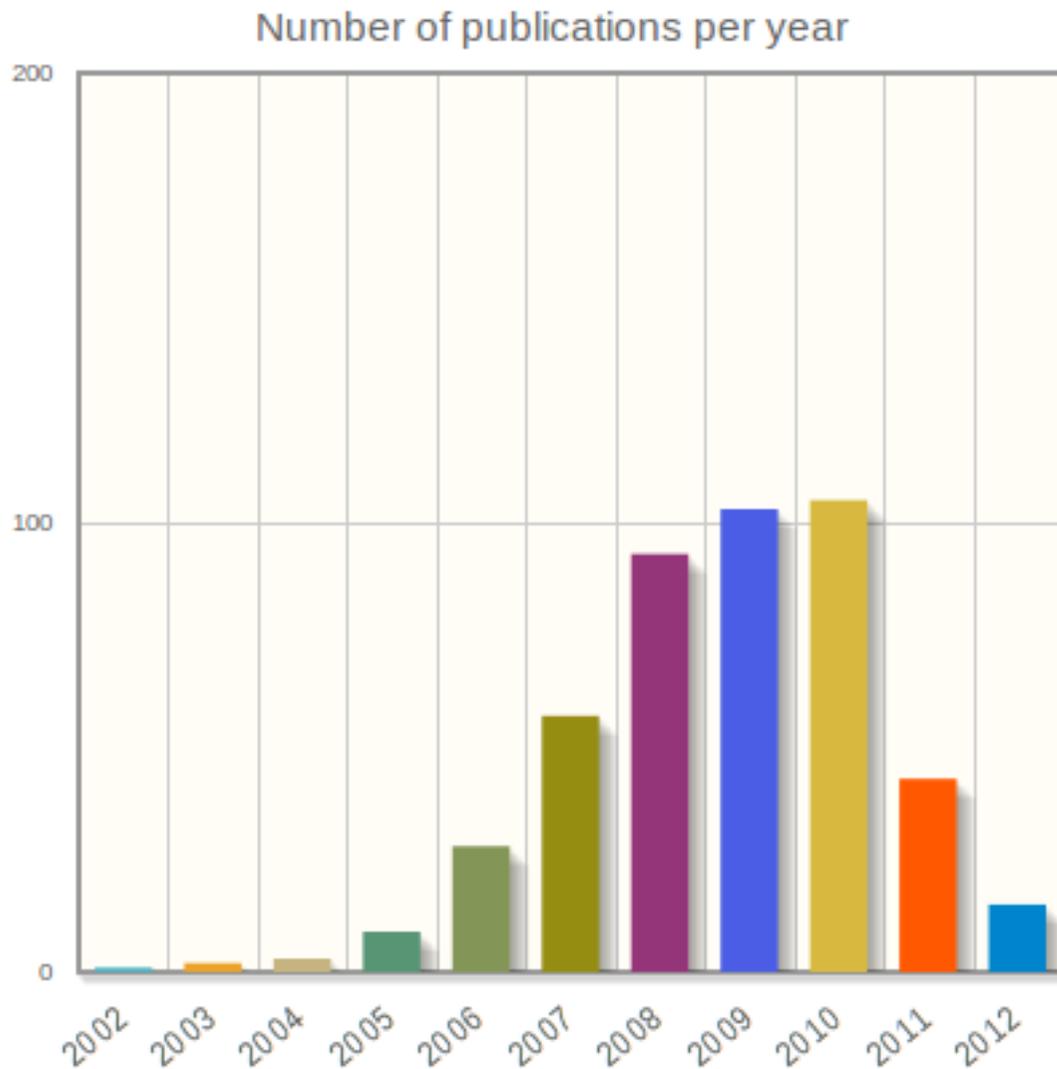
0. Wikipedia kan ikke virke i teorien men gør det i praksis.

## Wikipedia's nulte lov:

0. Wikipedia kan ikke virke i teorien men gør det i praksis.

Hvorfor?

# Wikipedia forskning



Forskningsartikler med Wikipedia som funktion af årstal.

Indenfor et væld af videnskabelige grene: Computer science, medicin, historie, lingvistik, jura, skovbrug, ...

Opgjort indtil juli 2011 og med langt fra alle konferenceartikler. Mange flere konferenceartikler, specielt indenfor computer science.

# Historien om wikier

Midten af 1990'erne: En wiki startes af Ward Cunningham

2001: Wikipedia startes, bog om wikier udgives

2002: Første forsknings artikel om Wikipedia.

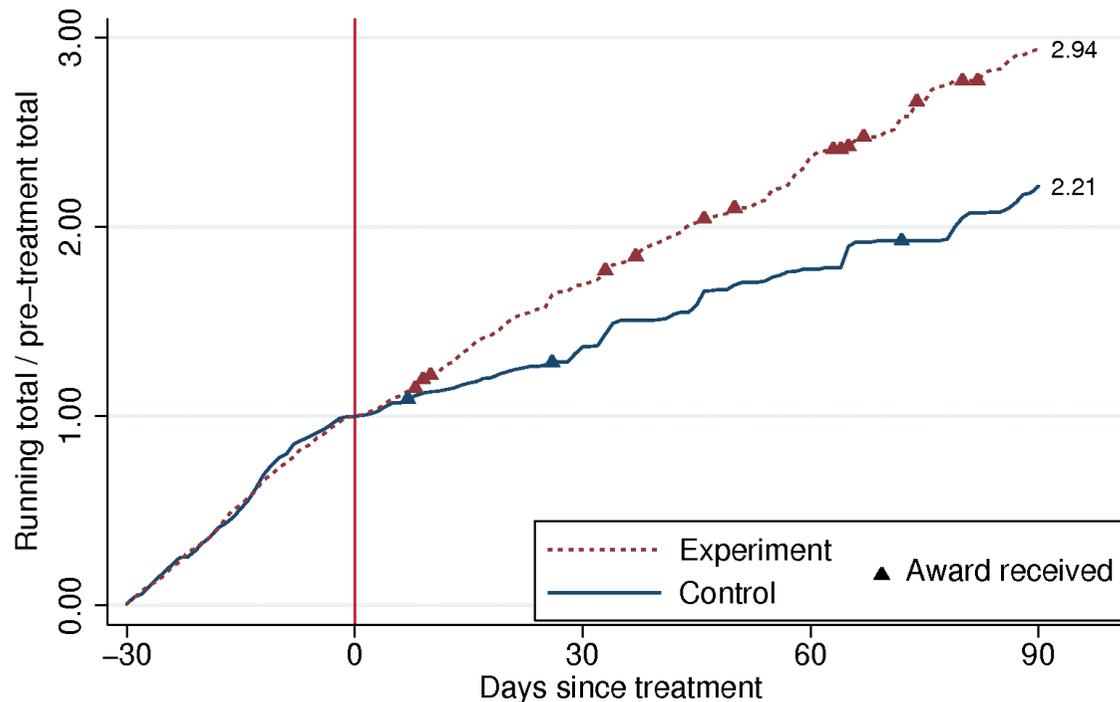
Midten af 2000'erne: Semantiske wikier foreslås, Semantiske service med Wikipedia data, Wikipedia bliver for alvor kendt

2013: Wikidata tages i brug, Wikipedia anslås at have 500 millioner besøgende om måneden

# Motivation

Hvorfor kan man for folk til at arbejde gratis?

# Motivation gennem sociale præmier



Sociale præmier, såkaldte *barnstars*, kan gives brugerne imellem.

Påvirker barnstars motivationen for at skrive?

Online eksperiment med to grupper Wikipedia-skribenter som var uvidende forsøgspersoner foretaget af ([Restivo and van de Rijt, 2012](#))

# Motivation

Gennem blandt andet interview og spørgeskemaer:

Lav barrier, ikke nødvendigt at lade registrere sig ([Ciffolilli, 2003](#); [Bryant et al., 2005](#); [Müller-Seitz and Reger, 2010](#))

Selv-uddannelse, bevise evner for potentielle arbejdsgivere ([Rosenzweig, 2006](#); [Müller-Seitz and Reger, 2010](#)).

Gruppe-identifikation ([George, 2007](#))

Social status, ry (barnstars) ([George, 2007](#); [Müller-Seitz and Reger, 2010](#); [Restivo and van de Rijt, 2012](#))

Altruisme

Kreativ nydelse, "flow" ([Rosenzweig, 2006](#); [George, 2007](#); [Müller-Seitz and Reger, 2010](#))

Penge, betaling(!)

Følelse af personlig præstation ([Yang and Lai, 2010](#))

Internal self-concept motivation: "I like to share knowledge which gives me a sense of personal achievement." ([Yang and Lai, 2010](#))

# Motivation: penge og betaling



## I Get Paid To Edit Wikipedia For Leading Companies

**ETHICALWIKI**  
 Wikipedia Marketing Blog | Why Wiki matters | Wiki Education | Wikipedia Consulting

### The five approaches to company policy on Wikipedia

Posted by [David King](#) on Fri, Jul 06, 2012

Marketing, public relations, legal and other risk management professionals are often in a tug-of-war, when comes to establishing corporate policy on Wikipedia. Without professional consulting, the team is left interpreting Wikipedia's conflict of interest guidelines or vague and sensationalized media headlines. The decision is like an old Shakespeare riddle, "to edit or not to edit." This presents companies with two equally poor decisions to either take the dangerous path of editing Wikipedia or ignoring the most in website on the planet. There are other alternatives.

Below is the company policy/strategy selection chart from EthicalWiki's [Overview of Wikipedia](#). The chart helps align the best approach to Wikipedia based on corporate attributes.

Corporate Attributes ↓	Strategy →	Hands-off	Monitoring & response	Public relations	Content marketing	Paid editing
------------------------	------------	-----------	-----------------------	------------------	-------------------	--------------

### Manipulating Wikipedia Content: Israeli Program to Train Editors to Ensure that "What is Written" is "Zionist in Nature"

By Prof. Michel Chossudovsky  
 Global Research, January 29, 2013

In August 2010, the **Yesha Council** together with **Israel Sheli (My Israel)**, organized a workshop in Jerusalem to teach people how to edit Wikipedia articles in a pro-Israeli way. (Wikipedia)

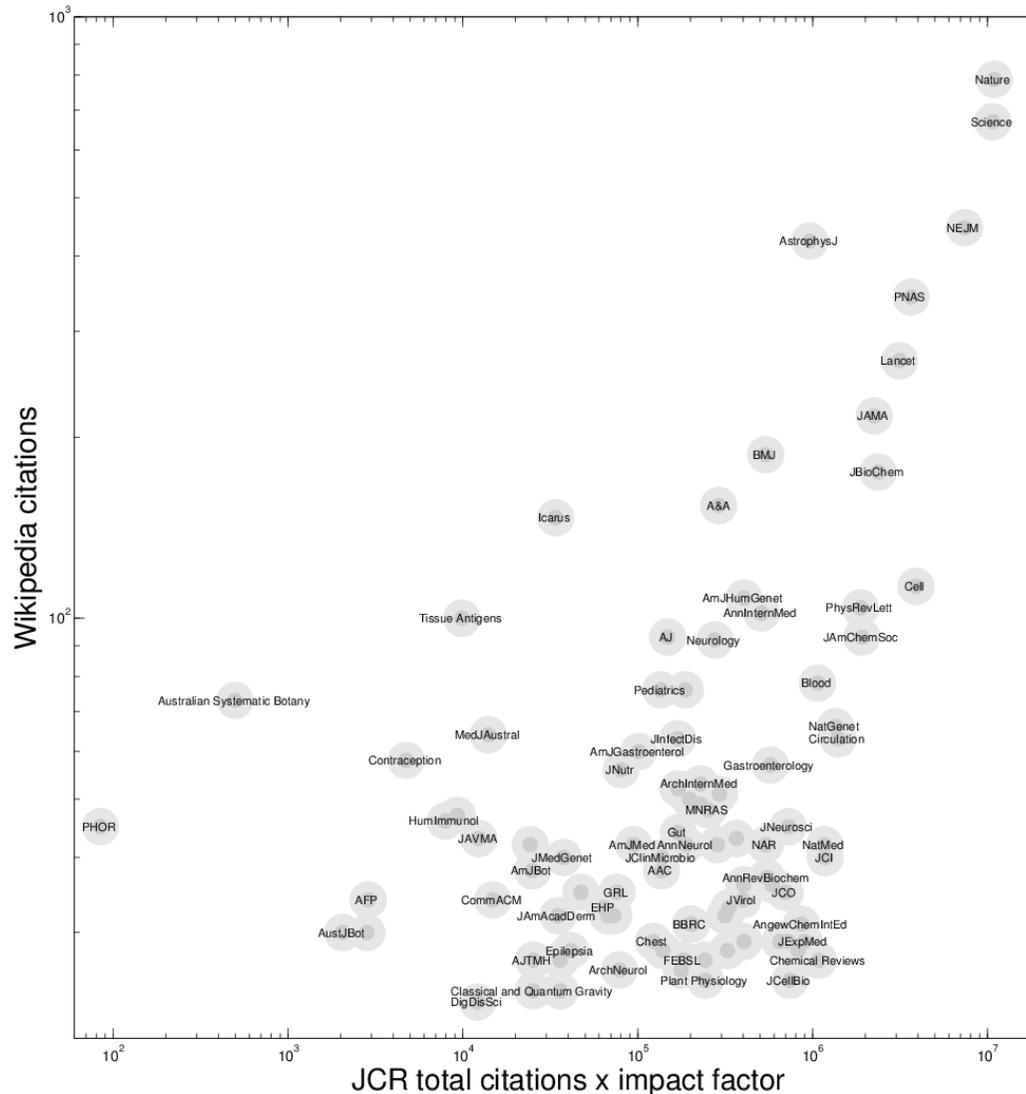
The **Yesha Council** which is the Hebrew acronym for **Yehuda Shomron, Aza**, lit. "Judea Samaria and Gaza Council" is an umbrella organization of municipal councils of Jewish settlements in the West Bank (and formerly in the Gaza Strip), known by the Hebrew acronym **Yesha**. (ibid)

**My Israel (Israel Sheli)** is "an Israeli right wing extra-parliamentary Zionist movement", which "deals with PR across the internet and especially social networks".

We bring to the consideration of Global Research readers this short video which features the Yesha Council-Israeli Sheli training program, pertaining to the art of manipulating online content on Wikipedia entries.

# Har Wikipedia relation til virkeligheden?

# Videnskabelige citeringer: IRL vs. Wikipedia



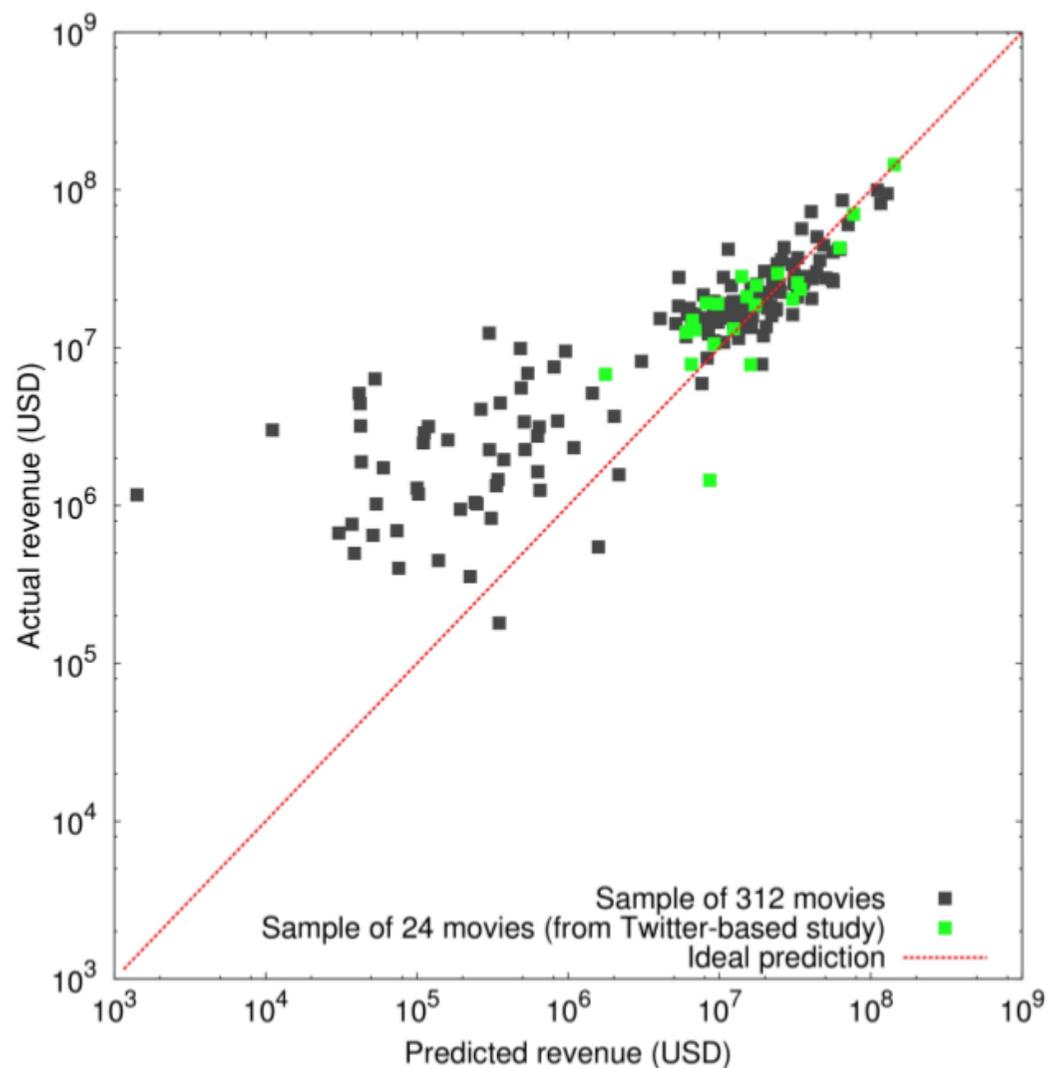
Strukturerede citering i Wikipedia til videnskabelige tidsskrifter (Nielsen, 2007).

Korrelation med IRL "journal impact"

2007: Rimelig korrelation, men med overvægt mod tidsskrifter indenfor astronomi og "Australian Systematic Botany".

Undervægt af biokemi.

# Prædiktion af billetsalg med Wikipedia



Studie med 312 film: Prædiktion af salgstal fra Wikipedia data

Salgstal for film i USA fra [Box Office Mojo](#)

Forskellige tal fra Wikipedia: Antallet af redigeringer, visninger og forskellige brugere.

Plot fra ([Mestyán et al., 2012](#)) af korrelation mellem prædikterede og faktiske værdi.

# Er Wikipedia påvirket væk fra "NPOV" ?

## Partifælle pyntede på Thornings profil: Fjernede et helt afsnit på wikipedia

Af Lars Fogt, lfo@bt.dk  
Mandag den 7. januar 2013, 09:04



SENESTE NYT  
Malala har ge kranieoperat  
Før St lillebr  
Agge skov

## Web-ups: Trane Nørbys ansatte rettede 31 gange i hendes leksikon

Af Lars Fogt  
Fredag den 4. januar 2013, 06:12



SIKJØNNMALING: Lars Løkke Rasmussens informationsmedarbejdere fjernede ufordelelige oplysninger om ham før valget i 2011.  
Foto: Keld Navntoft © NTB Scanpix

## Danmarks statsminister skjønnmalt på Wikipedia før valget

SENESTE NYT

Malala har genne kranieoperation  
Agger o skovtur  
Efter har  
Kontro kontar

Kort tid før folketingsvalget i 2011 ble siden om Lars Løkke Rasmussen endret 13 ganger.

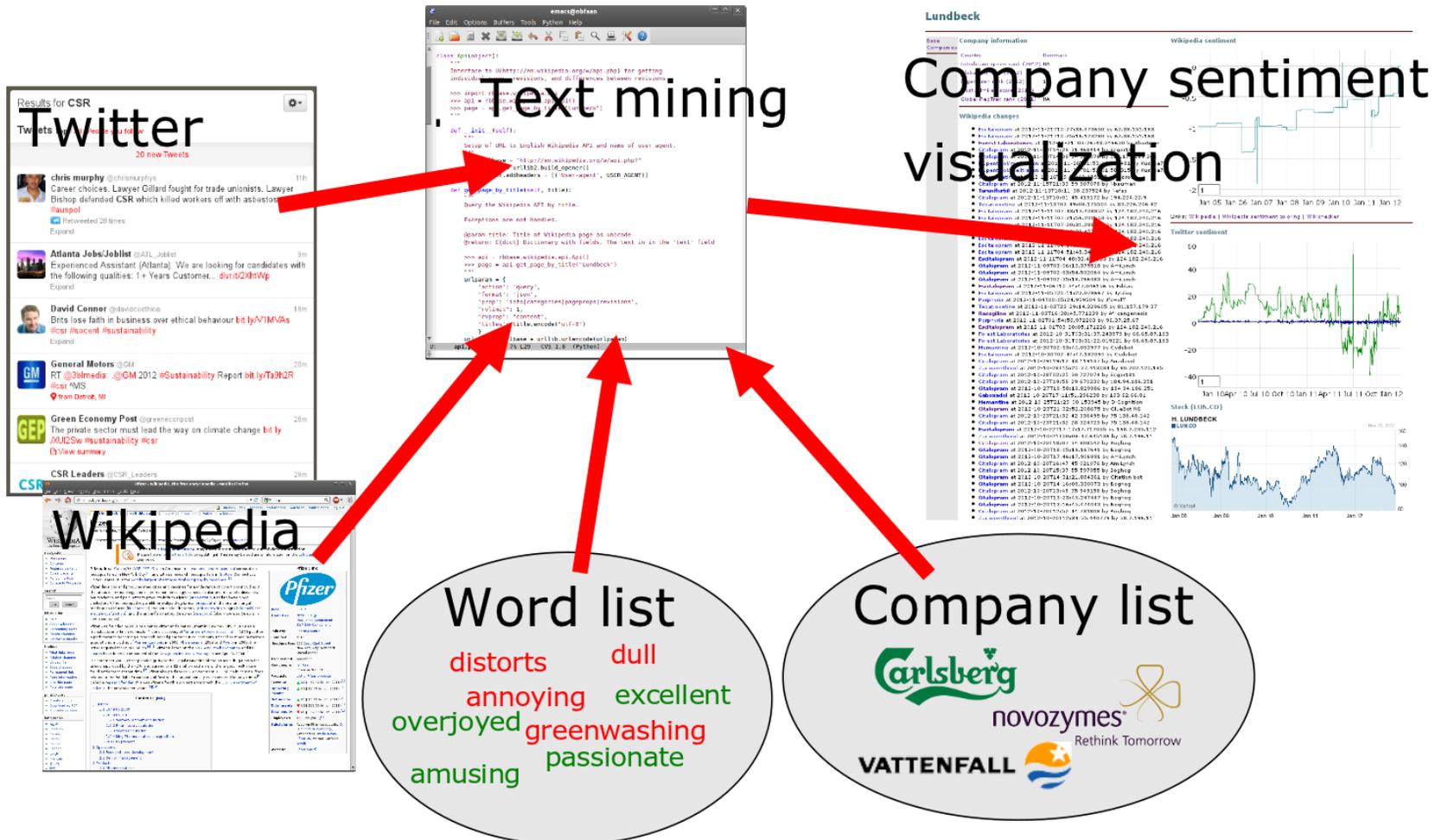
Del med andre:  
Anbefal 2  
Twitter  
Epost  
Del  
Facebook

Blant annet ble et avsnitt om problemsaker fjernet, skriver [den danske avisen B.T.](#)  
Opplysninger om liv og  
Opposisjonsleder trække sig 18:18  
Storfavoritterne snubbe Nations Cup 18:06  
Eastgate chipleader ved Jyske Mesterskaber 18:03  
toq pokalen - med et mål!



Lars Barth-Heyerdahl  
lbh@tv2.no  
Publisert 05.01.2013 17:22  
Oppdatert 05.01.2013 17:22

# System til overvågning af firmasider



# Simpel “stemningsanalyse”

Stemningsanalyse ved brug af en ordliste

Her eksempel med to afsnit fra den engelsk Wikipedia artikel om medicinalfirmaet Pfizer.

Negative ord: Lobbying, ban, lawsuits

Positive ord: honor, top

## [ [edit](#) ] Political lobbying

Pfizer is a leading member of the [U.S. Global Leadership Coalition](#), a Washington D.C.-based coalition of over 400 major companies and NGOs that advocates for a larger International Affairs Budget, which funds American diplomatic, humanitarian, and development efforts abroad. <sup>[ 85 ]</sup>

Pfizer is one of the single largest [lobbying interests](#) in [United States](#) politics. For example in the first 9 months of 2009 Pfizer spent over \$16.3 million on [lobbying](#) US congressional lawmakers, making them the sixth largest [lobbying interest](#) in the US (following Pharmaceutical Research and Manufacturers of America (PhRMA), which ranked fourth but also represents many of their [interests](#)). A spokeswoman for Pfizer said the company “wanted to make sure our voice is heard in this conversation” in regards to the company’s expenditure of \$25 million in 2010 to [lobby](#) health [care](#) reform. <sup>[ 86 ]</sup>

Pfizer’s primary [interests](#) are opposition of Congressional efforts to attach a prescription drug [benefit](#) to Medicare and opposition to generic drugs entering US markets. <sup>[ 87 ]</sup> Pfizer also purportedly proposed a [ban](#) on all [lawsuits](#) against [manufacturers](#) of [body implant](#) parts which was proposed in the [United States Congress](#) as part of [tort reform](#) legislation. <sup>[ 88 ]</sup>

According to [U.S. State Department](#) cables released by the [whistle-blowing](#) site [WikiLeaks](#), Pfizer “lobbied against New Zealand getting a [free trade agreement](#) with the [United States](#) because it objected to New Zealand’s restrictive drug buying rules and tried to get rid of New Zealand’s former health minister, [Helen Clark](#), in 1990. <sup>[ 89 ]</sup><sup>[ 90 ]</sup>

## [ [edit](#) ] Employment and diversity

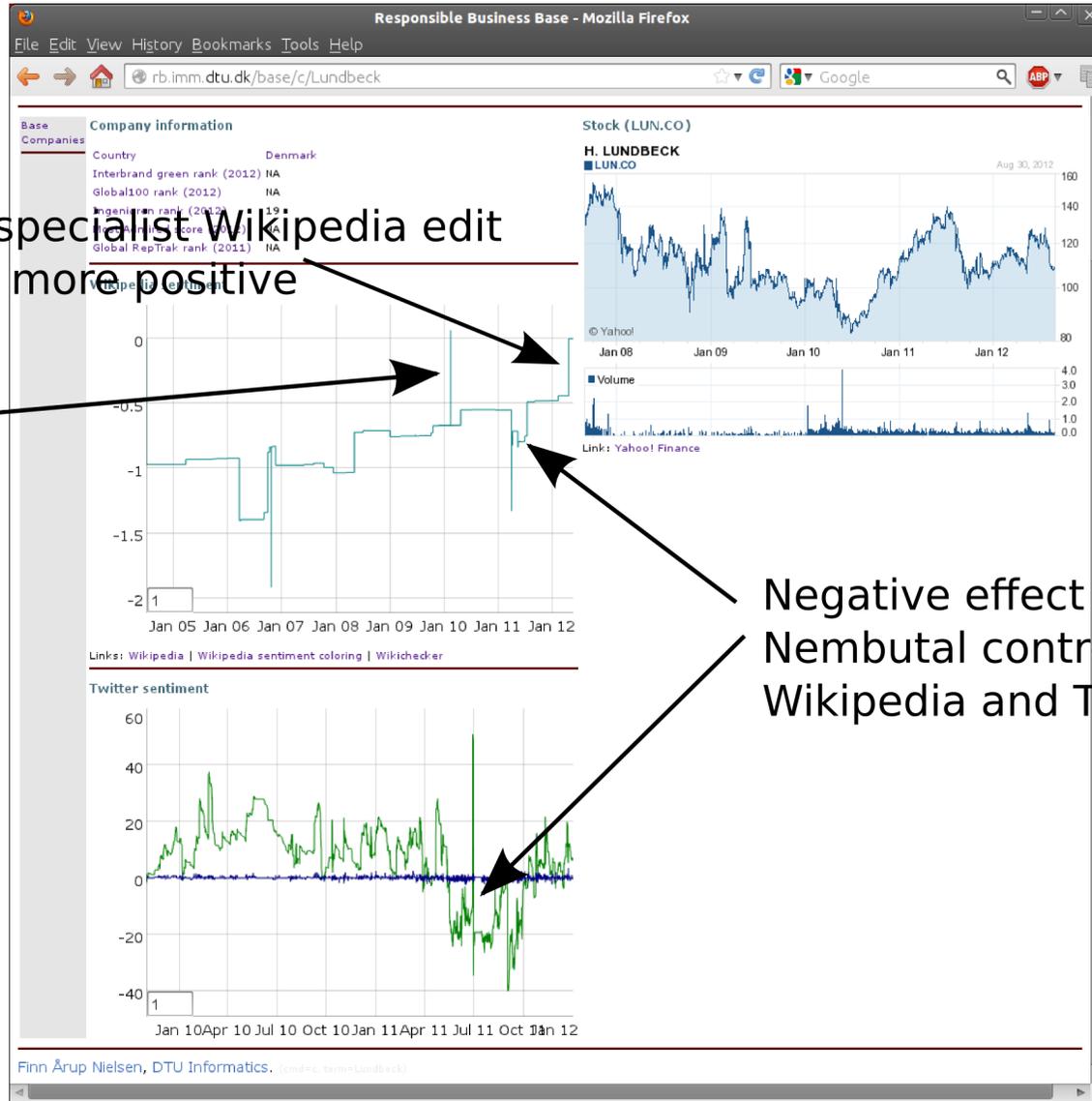
Pfizer received a 100% rating on the [Corporate Equality Index](#) released by the [Human Rights Campaign](#) starting in 2004, the third year of the report. In 2007, Pfizer’s Canadian division was named one of [Canada’s Top 100 Employers](#), as published in [Macleans](#) magazine, the only research-based pharmaceutical company to receive this [honor](#). <sup>[ 91 ]</sup> In 2008, there was controversy, including inquiries from members of Congress, around Pfizer’s practice of replacing US workers with H-1b guest workers <sup>[ 92 ]</sup>

# Hvad påvirker stemningen om et firma?

Lundbeck media specialist Wikipedia edit makes the article more positive

Vandalism

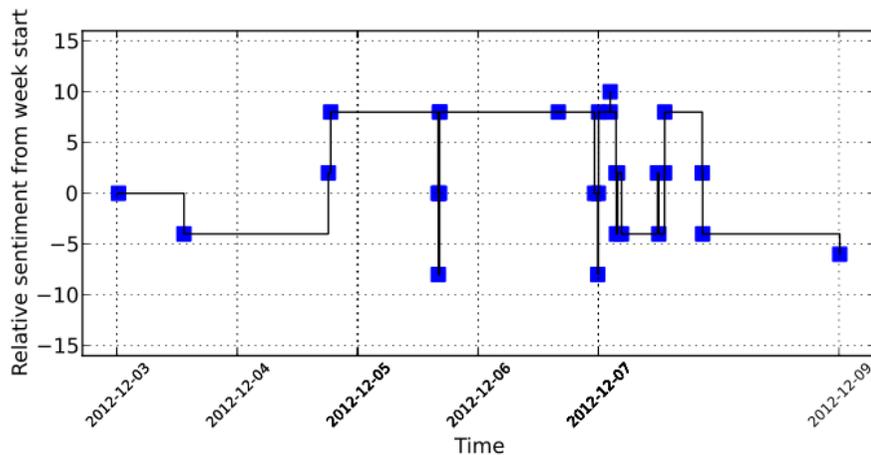
Negative effect of Lundbeck's Nembutal controversy in both Wikipedia and Twitter



# Periodisk stemningsrapport

## Lundbeck

27 change(s) with total sentiment -6.0 and total arousal 146.0 | [RBBBase](#) | [Wikipedia](#)



### Users

[AmiLynch](#) (-22.0), [71.224.176.33](#) (-12.0), [141.162.101.50](#) (-12.0), [166.137.88.30](#) (-8.0), [166.137.88.18](#) (-8.0), [Martinevans123](#) (-6.0), [Pauchpedia](#) (-4.0), [ClueBot NG](#) (-2.0), [66.214.64.91](#) (-2.0), [Dcirovic](#) (0.0), [C6541](#) (0.0), [98.185.157.123](#) (0.0), [63.156.194.193](#) (2.0), [Wingedsubmariner](#) (6.0), [Ssscience](#) (6.0), [82.13.137.193](#) (6.0), [Skinwalker](#) (16.0), [Ohnoitsjamie](#) (16.0), [Deli nk](#) (18.0).

### Individual edits

- [Rasagiline](#) by [66.214.64.91](#) (talk, contrib) at 2012-12-09 00:08 with sentiment -2. See [change on Wikipedia](#).
  - + These data suggest but do not prove a neuroprotective effect; the US Food and Drug administration has repeatedly denied Teva Pharmaceuticals' request for an on-label indication for neuroprotection in Parkinson disease.
- [Citalopram](#) by [141.162.101.50](#) (talk, contrib) at 2012-12-07 20:48 with sentiment -6. See [change on Wikipedia](#).
  - + Critics call this the industry's "Dirty Little Secret."
- [Escitalopram](#) by [141.162.101.50](#) (talk, contrib) at 2012-12-07 20:43 with sentiment -6. See [change on Wikipedia](#).
  - + Critics call this the industry's "Dirty Little Secret."
- [Escitalopram](#) by [Deli nk](#) (talk, contrib) at 2012-12-07 13:14 with sentiment 6. See [change on Wikipedia](#).
  - Critics call this the industry's "Dirty Little Secret."
- [Citalopram](#) by [Deli nk](#) (talk, contrib) at 2012-12-07 13:12 with sentiment 6. See [change on Wikipedia](#).
  - Critics call this the industry's "Dirty Little Secret."

Internet Relay Chat redigeringsstrøm monitoring.

Stemningsanalyse med "AFINN" (Nielsen, 2011)

Sprogprocessing med Python's NLTK (Bird et al., 2009)

Rapport med plot

## Andre værktøjer og botter

Den engelske Wikipedia's antivandalisme bot, ClueBot NG, benytter machine learning

Der er vandalism detektion konkurrencer ([Potthast et al., 2010](#); [Potthast and Holfeld, 2011](#))

WikiScanner (Virgil Griffith), ([Erenrich, 2008](#)) sammenholder anonyme redigering med whois information: Gav mange historier i medierne. Wiki-ganda ([Chandy, 2009](#)) var et lignende ordliste-baseret projekt.

Adskillige hundrede tools på *Toolserver* hvor der er adgang til dele af SQL-data

# Semantisk Web og Wikipedia (med DBpedia)

# Hvad er det Semantiske Web?

Semantisk Web =

Tripel datastruktur (der repræsenterer subjekt, verbum og objekt)

+ URI'er til at navngive elementer i tripel datastrukturen

+ standarder (RDF, N3, SPARQL, ...)

for maskinlæsbar halvstrukturerede data

Brug: IBM's Watson-program til Jeopardy

## Eksempel tripler

Subjekt	Verbum	Objekt
neuro:Finn	a	foaf:Person
neuro:Finn	foaf:homepage	<a href="http://www.imm.dtu.dk/~fn/">http://www.imm.dtu.dk/~fn/</a>
dbpedia:Charlie_Chaplin	foaf:surname	Chaplin
dbpedia:Charlie_Chaplin	owl:sameAs	fbase:Charlie Chaplin

hvor de såkaldte “prefix” er

```
PREFIX foaf:      <http://xmlns.com/foaf/0.1/>
```

```
PREFIX neuro:    <http://neuro.imm.dtu.dk/resource/>
```

```
PREFIX dbpedia:  <http://dbpedia.org/resource/>
```

```
PREFIX owl:    <http://www.w3.org/2002/07/owl#>
```

```
PREFIX fbase:    <http://rdf.freebase.com/ns/type.object.>
```

## DBpedia

DBpedia udtrækker halvstruktureret data fra Wikipedia (fra dens infobokse og kategorier) og lænker til et tripellager (*triple store*)

Data er gjort tilgængelig online i forskellige formatter: <http://dbpedia.org>

DBpedia navne (URI'er), f.eks., [http://dbpedia.org/resource/John\\_Wayne](http://dbpedia.org/resource/John_Wayne)

Menneskelig læsbar, f.eks., [http://dbpedia.org/page/John\\_Wayne](http://dbpedia.org/page/John_Wayne)

Maskinlæsbar, f.eks., [http://dbpedia.org/data/John\\_Wayne.json](http://dbpedia.org/data/John_Wayne.json)

# DBpedia forespørgsel

SPARQL endpoint for DBpedia:

<http://dbpedia.org/sparql>

Farmaceutiske firmaer med mere end 30'000 ansatte:

```
SELECT ?Company ?numEmployees ?industry ?page WHERE {  
    ?Company dbpprop:industry ?industry ;  
             dbpprop:numEmployees ?numEmployees ;  
             foaf:page ?page .  
    FILTER (?industry = dbpedia:Pharmaceutical_industry ||  
           ?industry = dbpedia:Pharmaceutical_drug) .  
    FILTER (?numEmployees > 30000) .  
}  
ORDER BY DESC(?numEmployees)
```



# Semantic MediaWiki

# Semantic MediaWiki

“Berlin” artikel på normal MediaWiki:

```
'''Berlin''' is the capital of [[Germany]] and has 3.443.570 inhabitants.
```

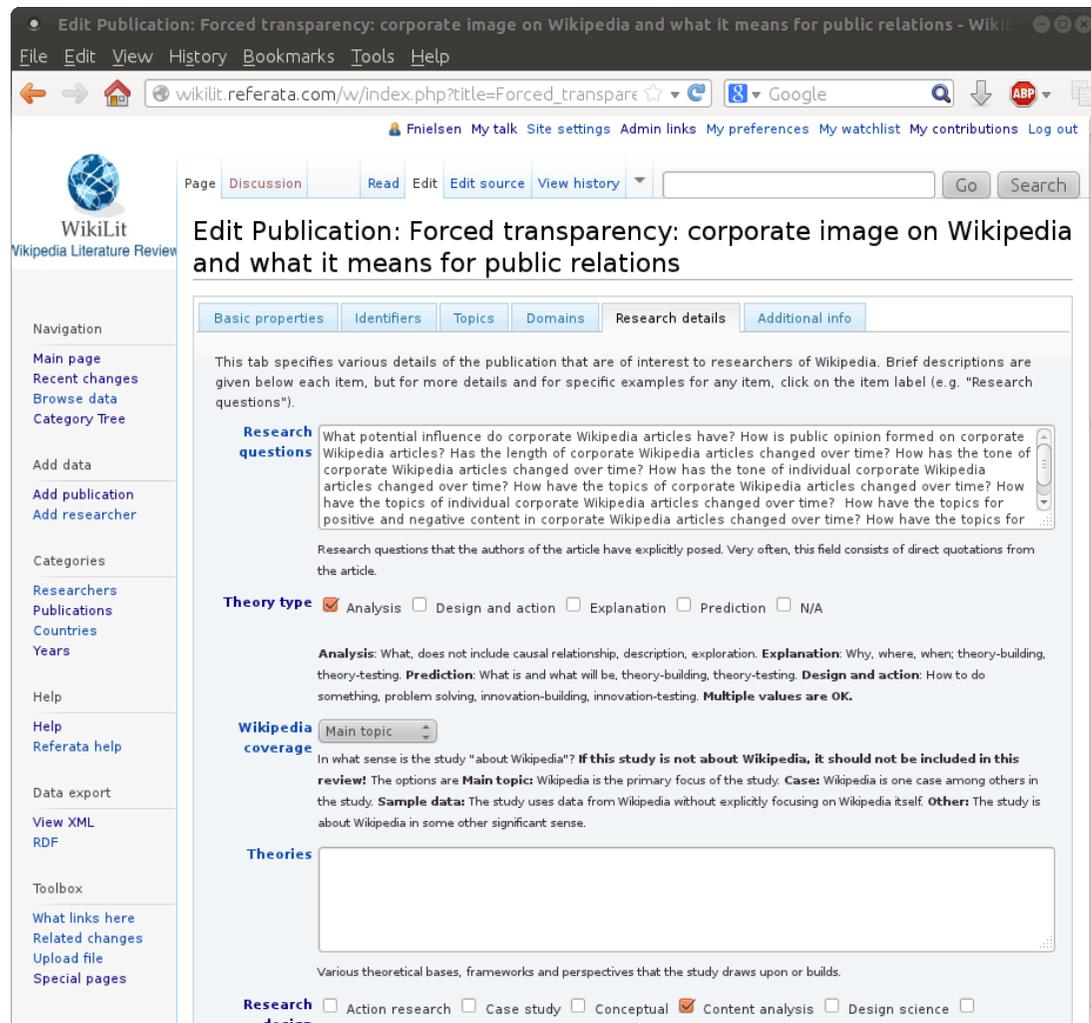
“Berlin” artikel på Semantic MediaWiki med dobbeltkolon:

```
'''Berlin''' is the capital of [[is capital of::Germany]] and has [[population::3443570]] inhabitants.
```

Resulterer i definitionen af følgende tripeldata:

Subjekt/side	Verbum/egenskab	Objekt/værdi
Berlin	is capital of	Germany
Berlin	population	3443570

# Semantic Forms



Forms kan defineres på specielle sider på en Semantic MediaWiki, — med Semantic Forms udvidelsen installeret

Muliggør definition af skabelonindhold med tekstfelter, popup-mener, radioknapper, ...

Her WikiLit som vi har brugt til at repræsentere videnskabelige artikler: Bibliografiske detaljer, forskningsspørgsmål, ...

# Resulterende definition

```
{{Publication
|type=Journal article
|title=Forced transparency: corporate image on Wikipedia and what ...
|authors=Marcia W. DiStaso, Marcus Messner
|published_in=Public Relations Journal
|year=2010
|month=Spring
|volume=4
|issue=2
|url=http://www.prsa.org/SearchResults/download/6D-040201/0/Force ...
|peer_reviewed=Yes
...}}
```

Skabelonen Publication definerer blandt andet:

```
{{#if: {{{title|}}} | [[title::{{{title}}}]] | Missing ''title'' }}
```

# Resulteret side med indhold

The screenshot shows a web browser window displaying a WikiLit article. The browser's address bar shows the URL: `wikilit.referata.com/wiki/Forced_transparency:_corporat`. The page title is "Forced transparency: corporate image on Wikipedia and what it means for public relations".

The article content includes:

- Publication (help)**: Forced transparency, corporate image on Wikipedia and what it means for public relations. Authors: Marcia W. DiStaso, Marcus Messner. Citation: *Public Relations Journal* 4 (2): . 2010 Spring.
- Contents [hide]**:
  - 1 [edit] Abstract
  - 2 [edit] Research questions
  - 3 Research details
  - 4 [edit] Conclusion
  - 5 [edit] Comments
  - 6 Further notes[edit]
  - 7 Summary
    - 7.1 Ranking and popularity
    - 7.2 Commercial aspects
    - 7.3 Other content topics
- Abstract**: Collaboratively edited information on social media that circumvents traditional media gatekeepers poses a challenge to public relations practitioners. The online encyclopedia Wikipedia gives corporate critics the opportunity to shape the public image of major corporations. This longitudinal panel study analyzed the framing of 10 Fortune 500 companies on Wikipedia between 2006 and 2010. It was found through content analyses of tonality and topics of more than 3,800 sentences in the articles for Wal-Mart, Exxon Mobil, General Motors, Ford, General Electric, Chevron, ConocoPhillips, Citigroup, AIG, and IBM that the negativity increased over time and that the focus shifted from historical information to legal concerns and scandals. The findings show that public relations practitioners need to pay close attention to the forced transparency about their companies on Wikipedia.
- Research questions**: "What potential influence do corporate Wikipedia articles have? How is public opinion formed on corporate Wikipedia articles? Has the length of corporate Wikipedia articles changed over time? How has the tone of corporate Wikipedia articles changed over time? How has the tone of individual corporate Wikipedia articles changed over time? How have the topics of corporate Wikipedia articles changed over time? How have the topics of individual corporate Wikipedia articles changed over time? How have the topics for positive and negative content in corporate Wikipedia articles changed over time? How have the topics for positive and negative

# Semantic MediaWiki forespørgsler

```
'''Unique values''': {{#arraydefine: my.array
| {{#ask: [[collected datatype::+]] | mainlabel =- | headers = hide
|? collected datatype | limit = 100000 }}
| , | print=list, sort=asc, unique
}}
```

**Unique values:** Archival records, Computer usage logs, Direct observation, Documents, Experiment, Interviews, Literature review, N/A, Survey, Websites, Wikipedia pages

```
{{#ask: [[Collected datatype::+]]
|? Collected datatype
| sort = Collected datatype
| limit = 5
}}
```

	↕ Collected datatype ↕
<a href="#">The work of sustaining order in Wikipedia: the banning of a vandal</a>	Archival records Computer usage logs
<a href="#">Extracting key terms from noisy and multitheme documents</a>	Archival records Wikipedia pages
<a href="#">Cross-language plagiarism detection</a>	Archival records Wikipedia pages
<a href="#">An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database</a>	Archival records Wikipedia pages
<a href="#">A semantic approach for question classification using WordNet and Wikipedia</a>	Archival records Wikipedia pages
<a href="#">... further results</a>	

Semantic MediaWiki udvidelsen er ikke sat til på Wikipedia.

# Wikidata

## Wikidata/Wikibase

Nylig forsøg på at strukturere Wikipedia's halvstrukturerede data

Udviklergruppen arbejder fra Berlin (Wikimedia Deutschland). Blandt andet Jeroen De Dauw, John Erling Blad, Daniel Kinzler

Flersproglig så etiket (*labels*) og beskrivelser kan være på flere sprog

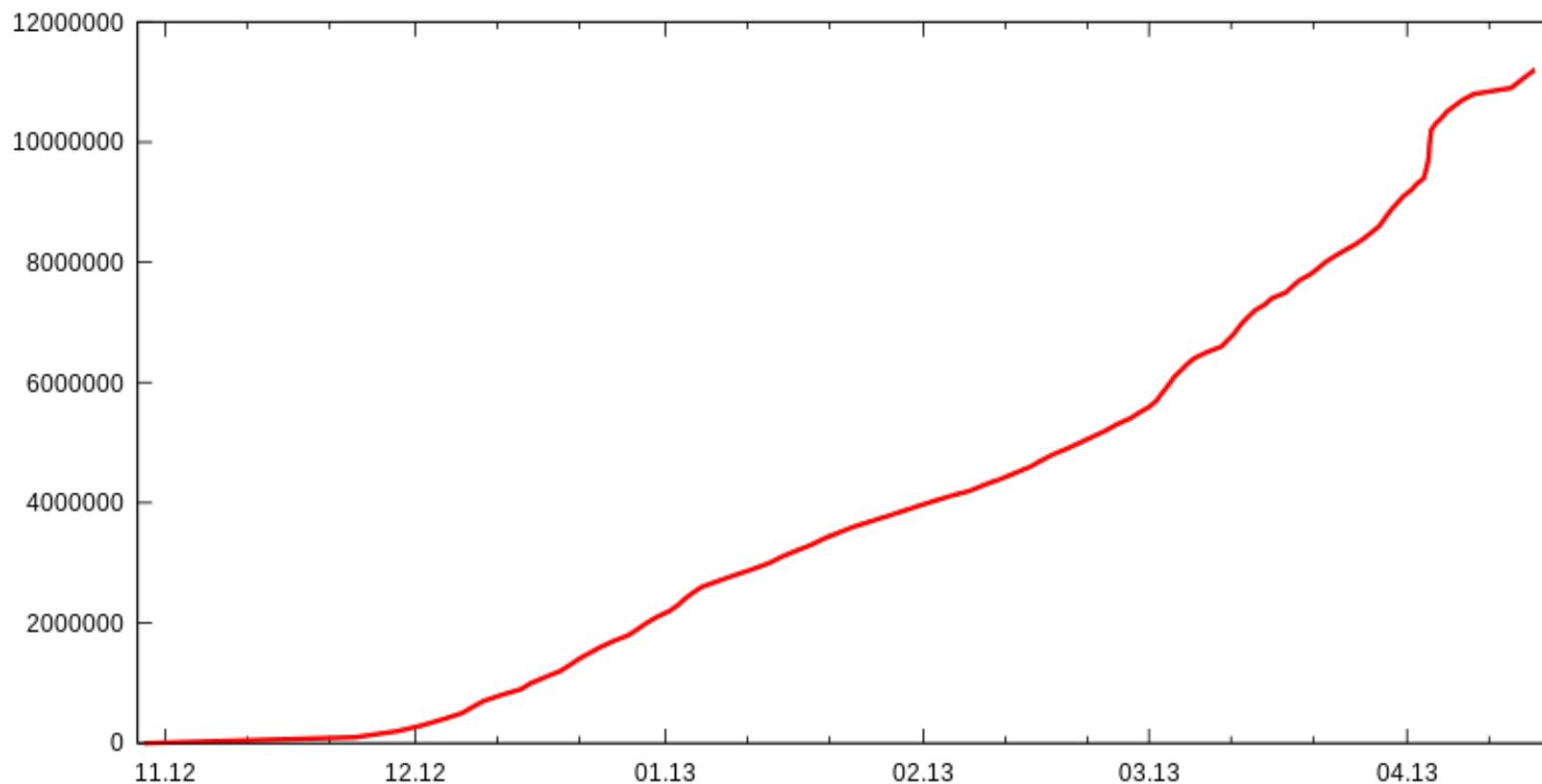
**Wikibase** hedder programmet/extensionen til MediaWiki

Instans på [wikidata.org](http://wikidata.org) under Wikimedia Foundation til Wikipedia

Wikidata har allerede over 10 millioner "items" (svare til sider/emner):  
Nr 10 million var "Pazardzhik"

Sproglinkning i Wikipedia er nu overtaget af Wikidata

# Vækst i Wikidata



Fra [Wikidata item creation progress no text](#) (Pyfisch, CC-BY-SA)

## Wikidata data model

Entity: Enten en “item” eller en egenskab

### 1. *Item* = Emne

- (a) Item identifier, f.eks. “Q1748” for København
- (b) Flersproglig etiket (*label*), f.eks “København”, “Copenhagen”
- (c) Flersproglig beskrivelse, “Danmarks hovedstad”
- (d) Flersproglig aliaser
- (e) Interwikilinks (links mellem de forskellige sprogversioner af Wikipedia)

## (f) Udsagn (*Claims*)

### i. Påstand (*Statement*)

A. Egenskab (*Property*), f.eks. "GND-type" (P107)

B. Værdi (*Property value*), f.eks. "geografisk objekt"

C. Kvalifikator (*Qualifiers*)

ii. Kilde (*Reference*), i øjeblikket under [Requests for comment](#)

## 2. Egenskaber (*Property*)

(a) Egenskabsidentifikator (*Property identifier*)

(b) Flersproglig label

(c) Flersproglig beskrivelse

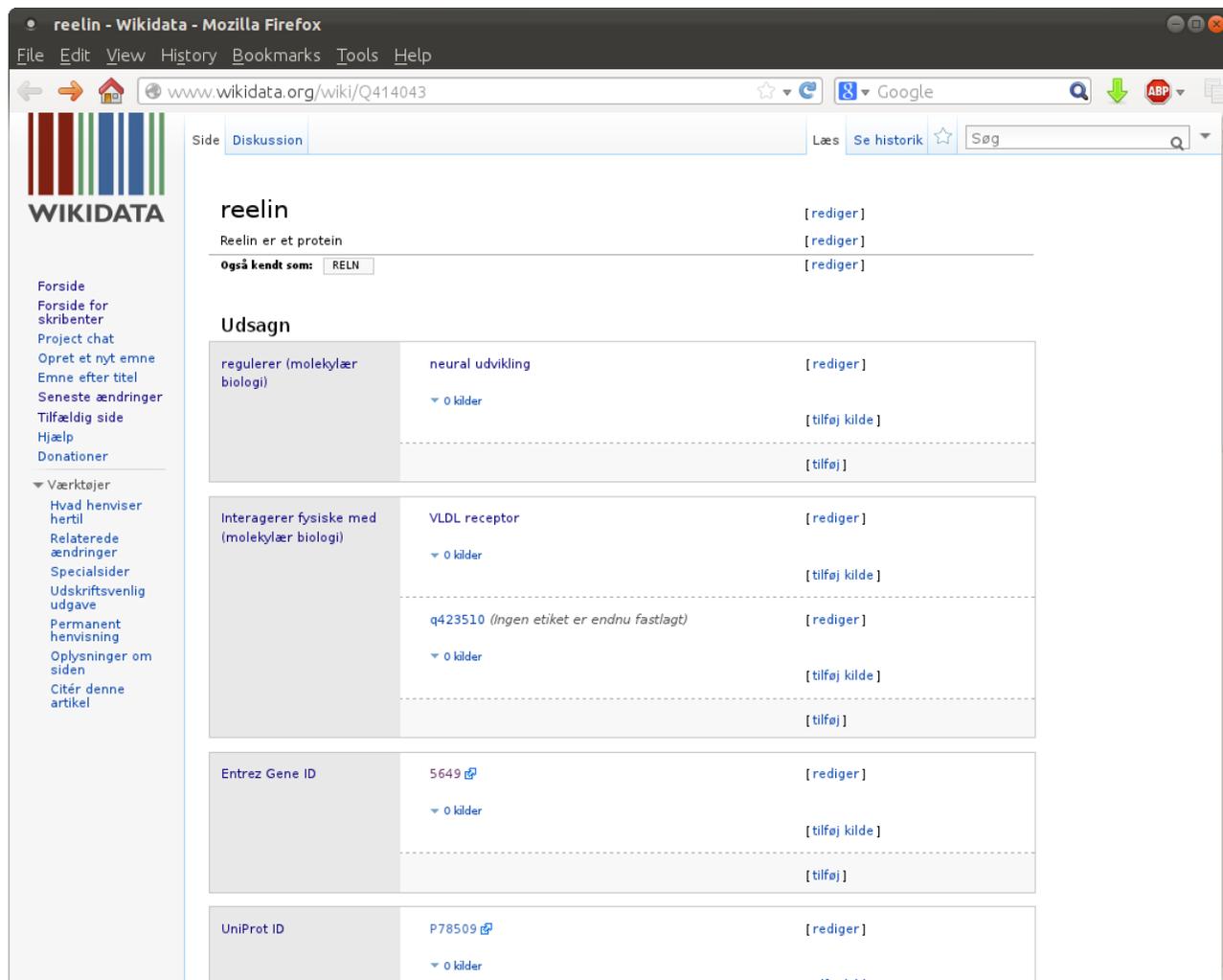
(d) Flersproglig aliaser

(e) Datatype

# Udsagn

Emne	Udsagn				Kilde	
	Påstand			Kvalifikator		
	Egenskab	Værdi	Kvalifikator			
			Egenskab	Værdi		
London	befolkningstal	8173900	metode	estimat	En angivet publikation fra Office for National Statistics	
			dato	juni 2012		
ulv	slægt	Canus				
Danmark	grænser op til	Tyskland				
<a href="#">Indien</a>	er en	selvstændig stat	fra	15. august 1947		
<a href="#">Thorvald Stauning</a>	er en	<a href="#">dansk statsminister</a>	fra	24. april 1924		
			til	14. december 1926		
		<a href="#">dansk statsminister</a>	fra	30. april 1929		
			til	3. maj 1942		

Eksempler på udsagn fra brugeren Byrial Jensen.



reelin - Wikidata - Mozilla Firefox

www.wikidata.org/wiki/Q414043

reelin [rediger]

Reelin er et protein [rediger]

Også kendt som: RELN [rediger]

Udsagn

regulerer (molekylær biologi)	neural udvikling [rediger]
	▼ 0 kilder [tilføj kilde]
	[tilføj]
Interagerer fysiske med (molekylær biologi)	VLDL receptor [rediger]
	▼ 0 kilder [tilføj kilde]
	q423510 (Ingen etiket er endnu fastlagt) [rediger]
	▼ 0 kilder [tilføj kilde]
	[tilføj]
Entrez Gene ID	5649 [rediger]
	▼ 0 kilder [tilføj kilde]
	[tilføj]
UniProt ID	P78509 [rediger]
	▼ 0 kilder [tilføj kilde]

Forskellige projekter søger kollaborativt at definere *egenskaber* i f.eks. bioinformatik og for bibliografiske data.

Her “reelin” med den danske beskrivelse “Reelin er et protein”: egenskaber: “regulerer”, “interagerer”, “Entrez Gene ID”, osv.

Endnu ikke oversat item “q423510 (Ingen etiket er endnu fastlagt)”

# Diskussion af egenskaber og items

**Participants** [redigér]

- Aubrey
- Micru
- Tpt
- EugeneZelenko
- User:Jarekt
- Maximilianklein (talk)
- Don-kun

**Book properties** [redigér]

This table groups properties from [en:Template:Infobox book](#) and [fr:Modèle:Infobox Livre](#). Please propose new properties here, and update this [mapping](#) created by Aubrey.

Wikidata property / Wikidata Eigenschaft / Propriété de Wikidata	Property / Eigenschaft / Propriété	Datatype / Datentyp / Types de données	qualifiers	FR	EN	DE	Creation level	Description Beschreibung Description
Original title		Monolingual string	-	X	X	?	created	
Original subtitle		Monolingual string	-	?	?	?	created	
Original language		Item	-	X	X	?	created	
Author		Item	-	X	X	?	created	
Translator		Item	-	X	X	?	Proposal	
Translated into	languages	Item	-	?	X?	?	Proposal	languages of translations of

Eksempler på ontologiske problemer:

Er en “bog” et værk, udtryk, manifestation eller en fysisk kopi? (jvf. *Functional Requirements for Bibliographic Records*)

Skal et protein indeholde gen-information, eller skal man lave forskellige *items* for (tilsvarende) gener og proteiner?

# Programmørens interface

Spørg efter København (Q1748), få flersproglige elementer på dansk og få data i JSON:

```
http://wikidata.org/w/api.php?  
    action=wbgetentities & ids=Q1748 & languages=da & format=json
```

Lille Python program til at hente København's land:

```
import json, requests  
url = "http://wikidata.org/w/api.php?" + \  
    "action=wbgetentities&ids=Q1748&languages=da&format=json"  
response = json.load(requests.get(url))  
property = response['entities']['q1748']['claims']['p17'][0]  
property['mainsnak']['datavalue']['value']['numeric-id']
```

Giver "35" (Q35=Danmark).

## Ekperimentelle service

MediaWiki-programmøren Magnus Manske's eksperimenterer med komplekse online forespørgsler til Wikidata data (<http://54.214.12.43:8085/>) eksempel:

“Places in the U.S. that are named after Francis of Assisi”

`http://54.214.12.43:8085/api?q=tree[30][150][17,131] and claim[138:676555]`

hvor Q30 er “USA”, P150 er “inddeles i”, P17 er “land”, P131 er “tilhører det administrative område”, P138 er “opkaldt efter”, Q676555 er “Frans af Assisi”

Andre eksempler:

“Bridges across the Danube”

“Bridges in Germany”

# Programmering og beregning med wikier?

## MediaWiki parser funktioner

Skabeloner i MediaWiki har et simpelt programmeringssprog med rudimentære funktioner såsom “if” og streng operationer.

Vældig obskur syntaks.

Eksempel med et af de mere forståelige til at formatere en citering:

```
{{#if: {{{journal|}}} | '''[[{{{journal}}} (journal)|{{{journal}}}]]'''  
| missing '''journal''' }} {{#if: {{{volume|}}} | '''{{{volume}}}'''  
| missing '''volume''' }} {{#if: {{{issue|}}} | ({{{issue}}}) }}:  
{{#if: {{{pages|}}} | {{{pages}}} | missing '''pages''' }}.  
{{#if: {{{year|}}} | {{{year}}} | missing '''year''' }}  
{{#if: {{{month|}}} | {{{month}}} }}
```

Måske på vej ud, pga Lua programmeringssprog.

# Lua programmeringssprog

[Lua eksempel](#) på siden “Module:Factorial” (fra brugeren “PiRSquared17”):

```
local p = {}
function p.factorial(frame)
    return p._factorial(tonumber(frame.args[1]))
end
function p._factorial(n)
    if n == 0 then
        return 1
    else
        return n * p._factorial(n - 1)
    end
end
```

Wikitext der giver resultatet “24”

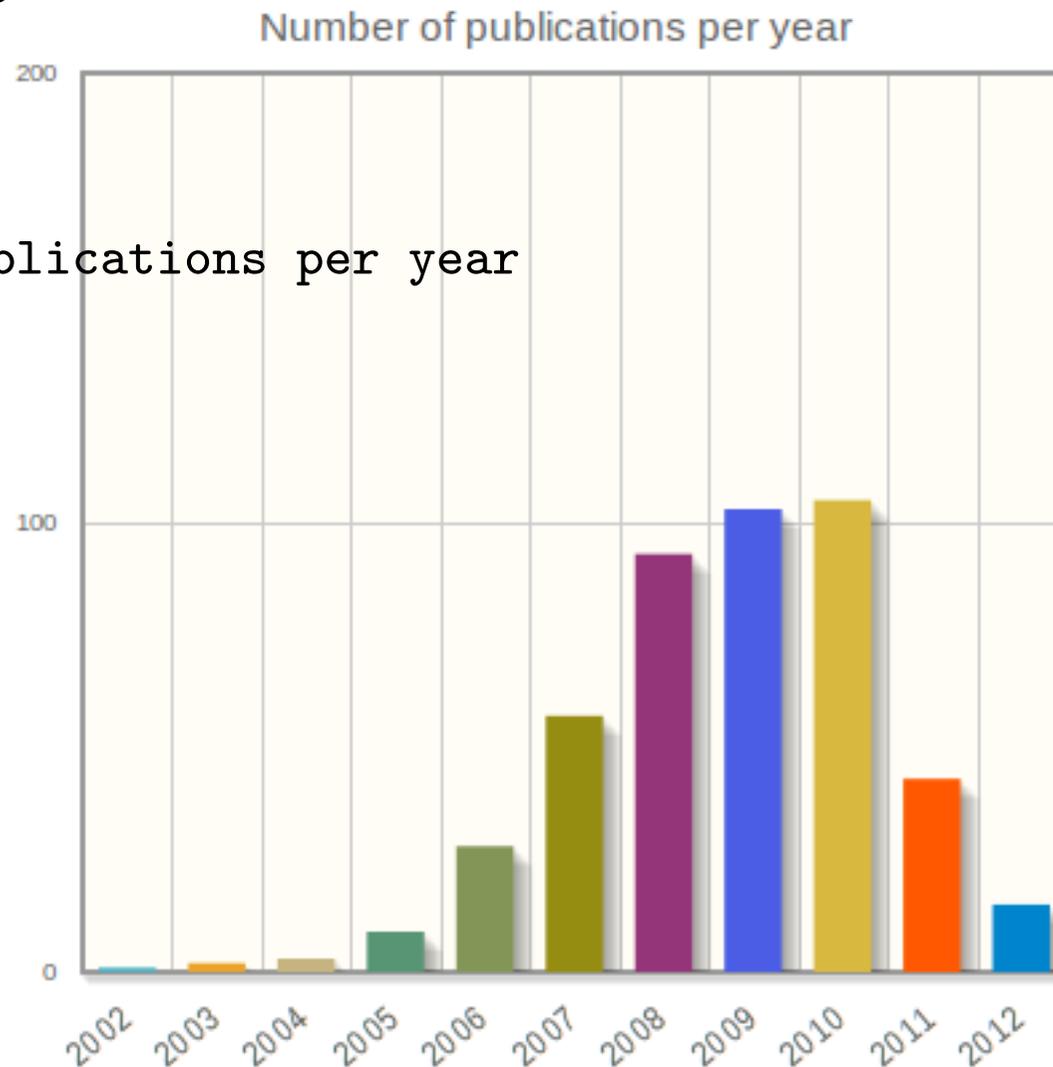
```
{{#invoke:Factorial|factorial|4}}
```

# Plot med Semantic MediaWiki

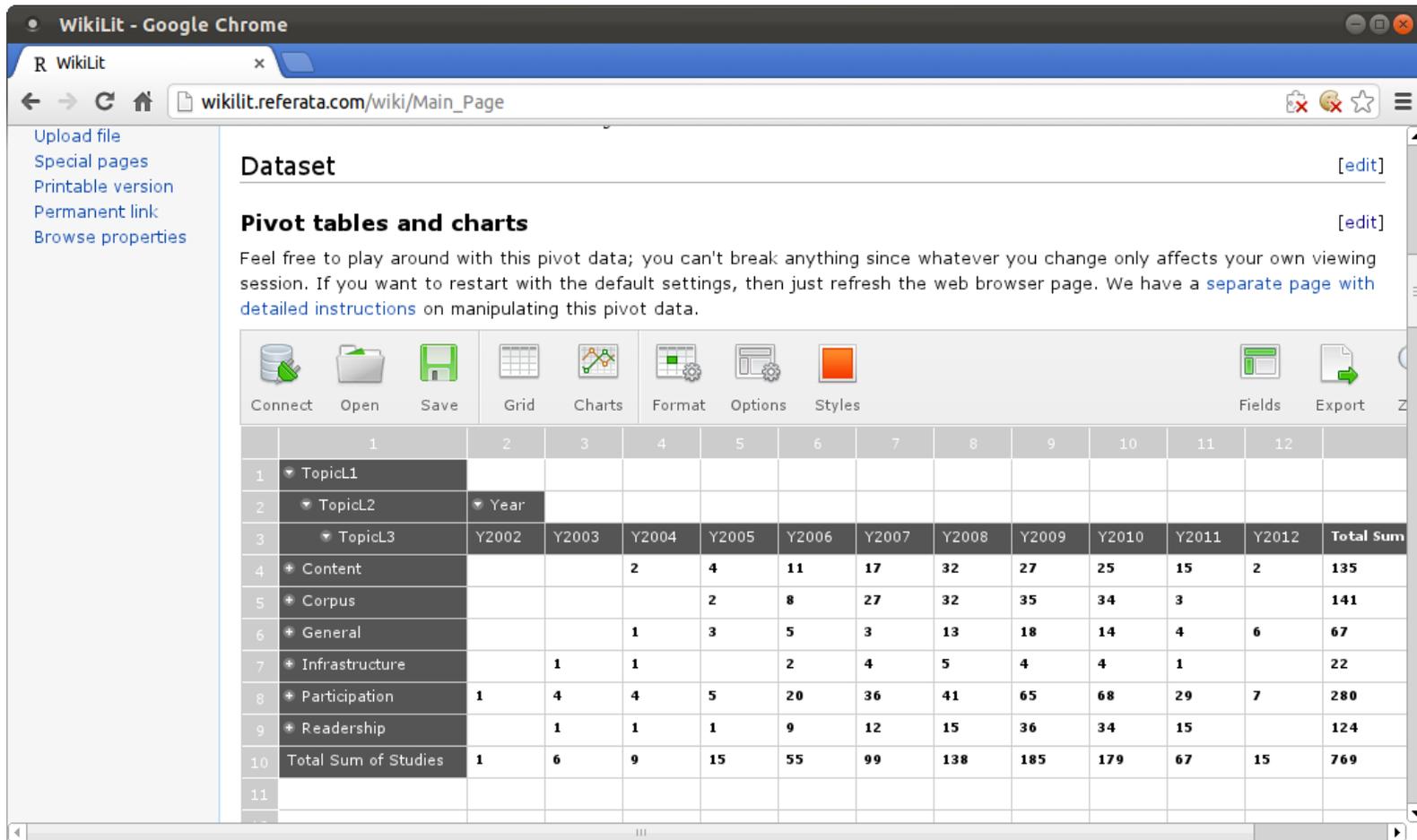
Semantic MediaWiki forespørgsel:

```
{{#ask: [[Category:Years]]
| ?number of publications
| charttitle = Number of publications per year
| format= jqplotbar
| pointlabels = yes
| min = 0
| width=50%
}}
```

Nødvendigt at optælle antallet af papers for hvert år på hvert "Category:Years" side.



# Pivotanalyse med Semantic MediaWiki



**Dataset** [edit]

**Pivot tables and charts** [edit]

Feel free to play around with this pivot data; you can't break anything since whatever you change only affects your own viewing session. If you want to restart with the default settings, then just refresh the web browser page. We have a [separate page with detailed instructions](#) on manipulating this pivot data.

	1	2	3	4	5	6	7	8	9	10	11	12	
1	▼ Topic1												
2	▼ Topic2	▼ Year											
3	▼ Topic3	Y2002	Y2003	Y2004	Y2005	Y2006	Y2007	Y2008	Y2009	Y2010	Y2011	Y2012	Total Sum
4	* Content			2	4	11	17	32	27	25	15	2	135
5	* Corpus				2	8	27	32	35	34	3		141
6	* General			1	3	5	3	13	18	14	4	6	67
7	* Infrastructure		1	1		2	4	5	4	4	1		22
8	* Participation	1	4	4	5	20	36	41	65	68	29	7	280
9	* Readership		1	1	1	9	12	15	36	34	15		124
10	Total Sum of Studies	1	6	9	15	55	99	138	185	179	67	15	769
11													

Pivottabel for papers over år og emne sat op af Chitu Okoli på vores cloud-baserede wiki med Flexmonster, semantisk forespørgsel og widget.

# Tabel-baseret wiki med data mining

The image shows a screenshot of a Wikipedia article titled "No association between 5-HTTLPR and harm avoidance in Korean college students" and a Brede Wiki table titled "Brede Wiki - Personality Genetics".

**Wikipedia Article:**

**No association between 5-HTTLPR and harm avoidance in Korean college students**

**Subjects**

Group of 586 healthy Korean students with 59 males and 85 females were included in the study. The average age was 23.8. The study was approved by the Institutional Review Board of Asan Medical Center.

**Results**

Category:Pages in need of cleanup

**Brede Wiki Table:**

Gene	Sample	Method	Year	Pub	Pub ID	Structure	Pub ID	Comment
5-HTTLPR	1000	SNP	2006	1	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals
5-HTTLPR	1000	SNP	2006	1	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals
5-HTTLPR	1000	SNP	2006	1	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals	16165	Association of 5-HTTLPR genotype and personality traits in a population of 1000 individuals

Red arrows indicate the flow of information: from the Brede Wiki table to the Wikipedia article, and from the Wikipedia article back to the Brede Wiki table.

“Brede Wiki for personality genetics”

Indtastning af data fra videnskabelige artikler om personligheds-genetik (Nielsen, 2010)

Indtastning i tabel-baseret wiki: Rimelig hurtigt indtastning, men slet ikke fleksibel.

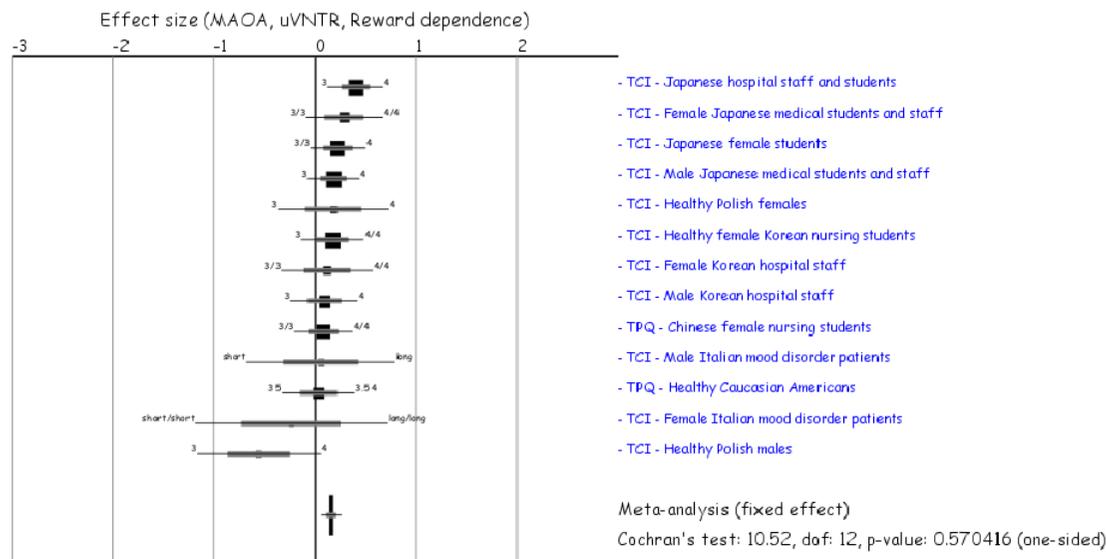
Mulighed for at regne på data.

# Beregning og plot i wikien

Tabel med beregning på tværs af data i wikien: Her over genetisk variationer og personlighedsdimensioner.

	Effect	Std	P	Studies	Subjects	Gene	Polymorphism	Trait
1	0.854	0.223	0.00013	2	107	ESR1	TA repeat	Harm avoidance
2	-1.102	0.289	0.00014	2	245	HTR3A	C178T	Harm avoidance
3	-0.779	0.220	0.00039	1	90	ESR1	TA repeat	Anxiety
4	-0.445	0.135	0.00098	1	247	TH	TCAT repeat	Extraversion
5	-0.401	0.123	0.00108	1	315	DRD4	Exon 3 VNTR	Positive emotions
6	0.165	0.051	0.00118	13	1747	MAOA	uVNTR	Reward dependence
7	-0.393	0.123	0.00135	1	315	DRD4	Exon 3 VNTR	Extraversion
8	-1.355	0.427	0.00152	1	125	HTR3A	C178T	Nonconformity
9	-0.758	0.240	0.00161	1	122	SLC6A4	5-HTTLPR	Activity
10	-0.174	0.055	0.00163	16	1791	SLC6A4	5-HTTLPR	Agreeableness

Plot af dele af data: Her et såkaldt forest plot i SVG med beregning af en meta-analytisk effektstørrelse



# Beregning med almindelig MediaWiki

Major Depressive Disorder Neuroimaging Database - Amygdala, total - Statistics

Major Depressive Disorder Neuroimaging Database - Amygdala, total - Statistics.csv

First Author	Year	Subgroup	Number of Patients	Number of Controls	Patient Mean	Patient SD	Control Mean	Control SD	Patient Age	% Female Patients	Age of Onset	Ham-D rating scale	Antidepressants %	Mood stabilizers %	Antipsychotics %	Drug Free %	Imaging	MRI Field Strength (T)	Slice Thickness (mm)	PfMD
Sheline YI	1998		20	20	3374	562.4546352	2534	560.6713183	54	100.0		5	70.0				MRI	1.5	1.25	9674587
Bremner JD	2000		16	16	1676	474	1341	449	43	97.5		100.0	0.0	0.0	0.0	12.3	MRI	1.5	3	10618023
Frod T	2003	first episode	30	30	3895	525.738338	3591	541.7137621	40.8	56.7	40	24.8				12.3	MRI	1.5	1.5	12586453
Frod T	2003	multiple episode	27	27	3542	458.2534233	3556	530.3728877	49.1	48.1	37.4	21.3				12.3	MRI	1.5	1.5	12586453
Caetano SC	2004		31	31	3.87	0.816259763	4.2	0.73054774	39.2	77.4	27.9	11.8	0.0	0.0	0.0	100.0	MRI	1.5	1.5	15598548
Hastings RS	2004		10	10	2865	531.5370166	2563	536.2704542	38.9	100.0		23	0.0	0.0	0.0	100.0	MRI	1.5	1.5	14997169
Lange C	2004		17	17	2.55	0.49	2.26	0.33	34	100.0	29	22	100.0			0.0	MRI	1.5	1.3	15564576
Xia J	2004		22	13	4477.23	247.7155727	4629.23	394.87261187	39.5	49.5		21.45				100.0	MRI	1.5	1.2	15641704
Russo IM	2005		20	24	4.42	0.76	5.26	0.785	15.35	85.0	12.8	16.55	0.0	0.0	0.0	100.0	MRI	1.5	1.5	15607296
Wolkowitz D	2006		12	87	3508	593.4489927	3010	443.1338863	22.6	41.7	22.5				0.0	MRI	1.5	1.5	16441856	
Wenger G	2006		21	23	2.6	0.565209979	2.3	0.379473313	34	100.0	29	28	100.0			0.0	MRI	1.5	1.3	16740316
Caetano SC	2007		19	24	3.02	0.440519196	3.16	0.502891638	13	31.6	10.3	47.4			52.6	MRI	1.5	1	17949501	
Nickles B (M)	2007		45	16	3.1	0.6	3.4	0.5	32	65.7	36.1	26.8	64.4				MRI	1.5	1.5	16930715
Murray MA	2007		26	18	28943.6	3425.689339	28377.87	3372.589347	20.54	100.0	15.58						MRI	1.5	1	17311571
Andrascu C	2008		71	32	0.22	0.04	0.25	0.04	72.2	69.0	52.3	18.3	16.9		1.4		MRI	1.5	1.5	18079490
Keller J	2008	psychosis	23	11	4.85	0.99627586	5.2	0.854025761	36.5	47.8	27.6	30.5	65.2	17.4	70.0	17.4	MRI	3	1.5	18450591
Keller J	2008	no psychosis	19	11	5.38	0.977188848	5.2	0.854025761	36.6	63.2	27	23.7	57.9	10.5	0.0	42.1	MRI	3	1.5	18450591
Nachreiner FP (B)	2008		32	35	3.01	0.588347725	2.72	0.550236313	14.08	62.5	11.77				100.0	0.0	MRI	1.5	1.5	17640621
Tamburo PJ	2008		14	11	3728	692.0411837	3100	590.1508166	69.8	35.7		13.8					MRI	1.5	1.5	19085964
Kronenberg G	2009		24	14	3.45	0.579120022	3.94	0.512328202	54.5	62.5		25.3	0.0	0.0	0.0		MRI	1.5	1.05	19384960
Lorenzetti V (B)	2009	depressed	29	15.5	3263.63	324.8962542	3206.12	289.1374099	35.52	75.9	21.07					16.1	MRI	1.5	1	19464062
Lorenzetti V (B)	2009	remitted	27	15.5	3309.7	368.1797727	3206.12	289.1374099	35.07	66.7	26.04					16.1	MRI	1.5	1	19464062
van Eindhoven P	2009	depressed	20	10	4747	515.6240879	4375	766.5528031	34.1	65.0	34.1	21.08	0.0				MRI	1.5	1	19028381
van Eindhoven P	2009	remitted	20	10	4086	561.6262102	4375	766.5528031	35.8	70.0	33.4	3.4	0.0				MRI	1.5	1	19028381
Weber K	2009		38	62	1.67	0.25	1.68	0.27	66.11	81.6	37.76	47.4					MRI	3	0.9	20018981

Download data as CSV | Edit data as CSV | Meta-analysis

Category: includes CSV

Repræsentation af data i et simpelt format i en MediaWiki-baseret wiki: komma-separerede værdier

Beskriv filen på en anden side.

I det her tilfælde: Data fra neuroimaging med måling af hjernestrukturers størrelse betinget på sygdom (Nielsen et al., 2012).

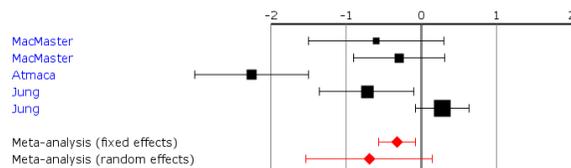
# Data analyse med en MediaWiki-baseret wiki

## Interpreted data and analysis

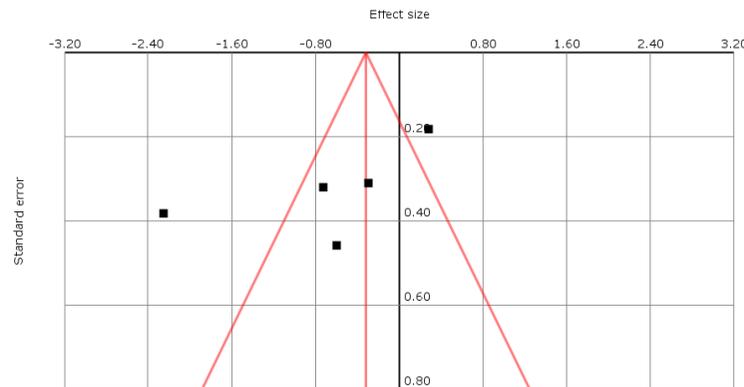
Study	Experimentals				Controls				Effects							
	Mean	SD	Events	N	Mean	SD	Events	N	SD <sub>pooled</sub>	Effect	SE	CI	Weight	Weight <sub>psl</sub>		
MacMaster 2006	0.434000	0.130230	nan	10	0.542000	0.209860	nan	10	0.174644	-0.592	0.459	-1.492 0.308	4.741110	7.674%	0.986339	18.096%
MacMaster 2006	0.623300	0.179700	nan	21	0.677600	0.186680	nan	21	0.183223	-0.291	0.310	-0.899 0.318	10.378958	16.798%	1.112003	20.401%
Atmaca 2009	691.000000	62.000000	nan	23	846.000000	73.000000	nan	23	67.723703	-2.249	0.384	-3.001 -1.498	6.797617	11.002%	1.052587	19.311%
Jung 2009	465.000000	55.800000	nan	12	543.000000	113.700000	nan	62	106.903327	-0.722	0.321	-1.352 -0.092	9.691549	15.686%	1.103616	20.248%
Jung 2009	577.800000	129.100000	nan	60	543.000000	113.700000	nan	62	121.515804	0.285	0.182	-0.072 0.641	30.176193	48.840%	1.196075	21.944%
Meta-analysis (fixed effect)				126				178		-0.316	0.127	-0.565 -0.067		100%		
Meta-analysis (random effects, DSL)										-0.685	0.428	-1.524 0.155				100%

$\tau^2$ : 0.896063 |  $Q$ : 38.262526 |  $df$ : 4.0 | **P-value**: 0.000

## Forest plot



## Funnel plot



Web service beskrivelse af data i MediaWiki-skabelon, henter data og beregner og generer plot.

Eksempel online på:

<http://neuro.compute.dtu.dk/wiki/Amygdala>

## Opsummering

Større og større muligheder for at strukturere data på wiki'er

Wikidata allerede godt i gang med at strukturere Wikipedia's halvstrukturerede data med mange millioner emner

Bedre programmeringsmuligheder efterhånden tilgængelig fra "almindelige" wikier: Lua på Wikipedia.

## Mere information

[Wikipedia research and tools: Review and comments](#) — en oversigtsartikel om forskning i Wikipedia og andre wiki'er.

Yaron Koren, *Working with MediaWiki*. Bog om MediaWiki og Semantic MediaWiki for administratore og brugere.

# References

- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly, Sebastopol, California. ISBN 9780596516499.
- Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 1–10, New York, NY, USA. ACM. [Link](#). *Describes interviews with 9 Wikipedia contributors and some of their characteristics: Most of the contributors tell that their initial edit was for correcting a problem or extending a weak article. As novices they were not aware of the Wikipedia community. As more experienced contributors they get a sense of community and decrease article writing and increase administration.*
- Chandy, R. (2008–2009). Wikiganda: Identifying propaganda through text analysis. *Caltech Undergraduate Research Journal*, 9(1):6–11. [Link](#). *Description of a opinion mining system for Wikipedia edits.*
- Ciffolilli, A. (2003). Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12). [Link](#).
- Erenrich, D. (2008). Wikiscanner: Automated conflict of interest detection of anonymous Wikipedia edits. In *Student-Faculty Programs: 2008 Abstract Book*, page 22. California Institute of Technology. *Student abstract about the Wikiscanner, that detects conflict of interest edits in Wikipedia also using Ip2location and USPTO trademark databases and computing link distance between pages and categories.*
- George, A. (2007). Avoiding tragedy in the wiki-commons. *Virginia Journal of Law and Technology*, 12(8):1–42. PMID: . [Link](#).
- Goldman, E. (2005). Wikipedia will fail within 5 years. *Technology & Marketing Law Blog*. [Link](#).
- Mestyán, M., Yasseri, T., and Kertész, J. (2012). Early prediction of movie box office success based on Wikipedia activity big data. ArXiv 1211.0970. [Link](#).
- Müller-Seitz, G. and Reger, G. (2010). 'Wikipedia, the free encyclopedia' as a role model? lessons for open innovation from an exploratory examination of the supposedly democratic-anarchic nature of

- Wikipedia. *International Journal of Technology Management*, 32(1):73–88. PMID: . DOI: [10.1504/IJTM.2010.035985](https://doi.org/10.1504/IJTM.2010.035985).
- Nielsen, F. Å. (2007). Scientific citations in *Wikipedia*. *First Monday*, 12(8). [Link](#). *Statistics on the outbound scientific citation from Wikipedia with good correlation to the Journal Citation Reports from Thomson Scientific*.
- Nielsen, F. Å. (2010). A fielded wiki for personality genetics. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, New York, NY, USA. ACM. [Link](#). ISBN 978-1-4503-0056-8.
- Nielsen, F. Å. (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In Rowe, M., Stankovic, M., Dadzie, A.-S., and Hardey, M., editors, *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, volume 718 of *CEUR Workshop Proceedings*, pages 93–98. [Link](#).
- Nielsen, F. Å., Kempton, M. J., and Williams, S. C. R. (2012). Online open neuroimaging mass meta-analysis. In Castro, A. G., Lange, C., van Harmelen, F., and Good, B., editors, *Proceedings of the 2nd Workshop on Semantic Publishing*, volume 903 of *CEUR Workshop Proceedings*, pages 35–39, Aachen, Germany. [Link](#).
- Potthast, M. and Holfeld, T. (2011). Overview of the 2nd international competition on Wikipedia vandalism detection. In Petras, V. and Clough, P., editors, *Notebook Papers of CLEF 2011 Labs and Workshops*. [Link](#). *Report from a prediction competition on Wikipedia vandalism detection. The corpus was based on both English, German and Spanish Wikipedias. Three systems participated*.
- Potthast, M., Stein, B., and Holfeld, T. (2010). Overview of the 1st international competition on Wikipedia vandalism detection. In *PAN 2010*. [Link](#).
- Restivo, M. and van de Rijt, A. (2012). Experimental study of informal rewards in peer production. *PLoS ONE*, 7(3):e34358. PMID: . DOI: [10.1371/journal.pone.0034358](https://doi.org/10.1371/journal.pone.0034358). [Link](#). *An experiment where Wikipedia editors were given informal awards to see how it affected their productivity*.
- Rosenzweig, R. (2006). Can history be open source? Wikipedia and the future of the past. *Journal of American History*, 93(1):117–146. PMID: . DOI: [10.2307/4486062](https://doi.org/10.2307/4486062). [Link](#). *Discuss several aspects of history on the English Wikipedia and how professional historians should regard that wiki. The author also make a quality assessment of a American history articles on Wikipedia and compare them against Encarta and American National Biography Online*.
- Yang, H.-L. and Lai, C.-Y. (2010). Motivations of Wikipedia content contributors. *Computers in Human Behavior*, 26(6):1377–1383. PMID: . DOI: [10.1016/j.chb.2010.04.011](https://doi.org/10.1016/j.chb.2010.04.011). *Reports on a survey among Wikipedia contributors about their motivation for sharing knowledge*.