

# Wikipedia — nørdernes sejr over vandalerne?

Finn Årup Nielsen

9. april 2008

‘Wikipedia kan ikke virke i teorien, men gør det i praksis’ lyder det om verdenens— i skrivende stund—ottende mest besøgte websted.\* Wikipedia’s artikler står ofte øverst i søgemaskinernes resultatlistor, artiklerne bliver citeret i aviser,<sup>2</sup> i videnskabelige artikler, Indiens højesteret<sup>3</sup> og amerikanske domstole<sup>4</sup> har brugt Wikipedia i deres kendelser, Wikipedia’s grundlægger Jimmy Wales har vidnet for det amerikanske senat<sup>5</sup> og i Tyskland anvendes offentlige midler til at forbedre indholdet af Wikipedia’s tyske version.<sup>6</sup> Men hvorfor virker Wikipedia når hvilkensomhelst person kan redigere i artiklerne, ekspert såvel som vandal? Den åbne arkitektur går direkte mod normal computersikkerheds credo, og Jimmy Wales har dog også modtaget e-mails fra bekymrede brugere der ville påpege den store ‘sikkerhedsbrist’.<sup>7</sup>

Flere forskere er begyndt at kigge nærmere på Wikipedia, dels for at udnytte den rige information, dels for at undersøge hvordan Wikipedia bliver opbygget og brugt, og om muligt svare på hvorfor Wikipedia virker. Disse bestræbelser involverer fysikere, dataloger, ingeniører, sociologer og folk fra biblioteksvidenskaben.

---

\*Alexa rankede som nummer 8 i slutningen af 2007. I begyndelse af 2007 var den allerede brudt igennem til top ti over verdenens mest besøgte hjemmesider.<sup>1</sup>

†Blandt undersøgte karakteristikkere er mange af de moderne ‘fancy’ tal: PageRank og Kleinberg’s HITS,<sup>12</sup> den såkaldte ‘lilleverdens-koefficient’, med cluster-koefficient og kortestevejmiddelværdi,<sup>14,16</sup> ‘butterfly-komponenternes’ størrelser,<sup>13,14</sup> potenslov-koefficienter,<sup>8,13-15,17</sup> reciprocitet,<sup>14</sup> assortativitets-koefficient<sup>14</sup> og triade-signifikansprofil.<sup>14</sup>

## Et kæmpe netværk

Forfatterene af Wikipedia udgør et kæmpe socialt netværk, der kan studeres ud fra deres spor i Wikipedia’s revisionshistorik. Flere forskere har studeret det varierende antal redigeringer som wikipedianere foretager.<sup>8-10</sup> Andre har også undersøgt linkstrukturen i Wikipedia.<sup>11-16</sup> Wikipedia’s rige netværksstrukturer udviser nogle slående træk som også er set blandt andre ‘naturlige’ netværk, for eksempel elnettet, netværk i forbindelse med neuroner i orme eller netværket blandt filmskuespiller.<sup>†</sup> På trods af at Wikipedia fungerer uden central styring opstår der generelle mønstre i brugerens adfærd og Wikipedia’s vækst. For eksempel: Antallet af en artikels redigeringer forudsiger simpelt hvor mange den vil modtage fremover.<sup>18,19</sup>

## Wikipedia som vidensbase

De fleste Wikipedia artikler er rubriceret i en eller flere kategorier, og kategorierne kan være ordnet i et hierarki. Der er ingen central styring på hierarkiet, og taksonomiet opstår kun ved at Wikipedia’s brugere *tagger* hver enkelt side med emneord.<sup>22</sup> Orde-

## Historien

Allerede i idéerne fra webbens fader Tim Berners-Lee ligger mekanismer der tillader information at flyde både frem og tilbage mellem web-serveren og brugeren. Først i 1995 opstår dog den egentlige wiki, der tillader nem oprettelse og redigering af web-sider.<sup>20</sup> Det er Ward Cunningham der sætter den først wiki op, og til formatering og linkning benytter han et enklere opmærkningsprog end det sædvanlige internet's HTML. Egentligt var det blot til at lette samarbejde med kollegaer omkring den specielle form for strategi til computerprogrammering han

undersøgte. Wiki-idéen bredte sig relativt langsomt, indtil Larry Sanger hørte om wiki-idéen. Sanger var ansat af Jimmy Wales til at udvikle en Internet encyklopædi: *Nupedia*. Det gik dog langsomt med oprettelsen af artiklerne, og derfor beslutter de to at prøve at supplere encyklopædien med en wiki. I løbet af kort tid bliver denne wiki-udgave af encyklopædien—Wikipedia—større end originale, og senere nedlægger de da også Nupedia.

Oprindeligt var Wikipedia finansieret af Jimmy Wales der havde skabt sig en formue ved

børshandel og—øjensynligt—gennem hans firma Bomis hvis aktiviteter blandt andet inkluderede 'Bomis Babe Report'.<sup>21</sup> Wales har dog nu overdraget Wikipedia til en fund og funden modtager donationer fra brugerne og firmaer.

Med Wikipedia som model har wiki-idéen bredt sig, så mange større virksomheder nu har en wiki som en del af deres intranet. Softwaren MediaWiki, som Wikipedia anvender, kan downloades frit, modificeres og sættes op som en del af ens egen web-server. Der findes også mange andre wiki-softwarepakker.

ne 'folksonomi'<sup>23</sup> eller 'folketaksonomi' er brugt som denne type taksonomi. Det er muligt at få fat i dette taksonomi, da hele Wikipedia ligger i en kæmpe fil på Internettet, lige til at downloade og fodre sin algoritme med.

Forskere har således foreslået at bruge Wikipedia's folketaksonomi til at beregne ords *semantiske slægtskab*, for eksempel om ordene 'FBI' og 'investigation' er beslægtede.<sup>24</sup> Det kan være et vigtigt element i computeres 'forståelse' af naturligt sprog. På engelsk findes der allerede et stort velstruktureret ordtaksonomi, *WordNet*, men det mangler mange egenavne. Ved at kombinere WordNet og Wikipedia kunne de tyske forskere opnå de bedste resultater.

Det tyske eksempel er blot en af flere måder hvorpå Wikipedia bliver udnyttet til computer-baseret sprogprocessering og

avanceret informationssøgning.<sup>16,25–30</sup> På DTU Informatik brugte to studerende, Magnús Sigurdsson and Søren Christian Halling, delen af Wikipedia om musikgrupper til deres Internet-baserede søgemaskine for at gruppere søgeresultater i forhold til relevante emner.<sup>31</sup>

Men hvis Wikipedia bruges som vidensbase skal man kunne stole på den, — eller ihvertfald skal en stor del være kvalitetsfuld. Er den det?

## Kvaliteten?

Flere formelle studier af Wikipedia's kvaliteter har været udført.<sup>‡</sup> I den måske mest omtalte undersøgelse havde *Nature's* videnskabsjournalister sakset 42 af Wikipedia's og Encyclopaedia Britannica's naturviden-

<sup>‡</sup>Se den engelske Wikipedia's egen oversigte på 'Wikipedia:External peer review' (oldid=169010073) og 'Reliability of Wikipedia' (oldid=179243736).

## Carl Hewitt vs. Wikipedia

Carl Hewitt, en pensioneret professor fra MIT, kom i veltent i slutningen af 2007 efter at have redigeret Wikipedia.<sup>32</sup> Han har en lang karriere indenfor programmering, men på trods af hans store bidrag i dette område, har flere andre Wikipedianere anset nogle af hans redigeringer som selv-promovering. Han har blandt andet indsat beskrivelser af hans forskning (den såkaldte *actor model*) i fysikartikler, hvor andre mente at de ikke

hørte hjemme. Uoverensstemmelsen førte til lange diskussioner og en sag for *Arbitration Committee*, Wikipedia's højesteret, der besluttede at at bandlyse ham.

Sagen viser et eksempel på at det kan være svært at skrive fra et neutralt synspunkt hvis man har en interessekonflikt og er dybt optaget af sit forskningsfelt. Den officielle politik på Wikipedia tillader dog at specialister indsætter deres egen forskning, hvis det

ellers har baggrund i en troværdig udgivelse og overholder Wikipedia's andre regler. Jeg selv har indtil videre sluppet afsted med at liste referencer ind til nogle artikler jeg har 'været med på'. Universiteterne forventer at deres forskere gør deres arbejde kendt for offentligheden, og Wikipedia er en måde hvor man både kan nå andre forskere og almindelige informationssøgende.

skabelige artikler, og derefter sat blindede eksperter til at vurdere dem.<sup>33</sup> Sammenligningen mellem de to encyklopædier viste at Wikipedia's artikler indeholdt flest mangler og unøjagtigheder, — men overraskende nok ikke særlig mange flere end Encyclopaedia Britannica. *Stern* foretog i december 2007 en tilsvarende undersøgelse med den tyske Wikipedia overfor online-udgaven af den tyske encyklopædi *Brockhaus*.<sup>34,35</sup> Ugemagasinet havde bedt et uafhængigt forskningsinstitut, Wissenschaftlicher Informationsdienst Köln, om at vurdere 50 artikler ud fra kriterier om korrekthed, fuldstændighed, aktualitet og forståelighed. I 43 tilfælde blev Wikipedia's artikler vurderet som bedre end Brockhaus', og Wikipedia fik det bedste karaktergennemsnit.<sup>36,37</sup> Wikipedia blev anset for bedre i aktualitet og — måske overraskende — i korrekthed, mens Brockhaus lå bedst i fuldstændighed. Desuden blev nogle af Wikipedia's artikler betragtet som for komplicerede for læg-

mand.<sup>§</sup>

Blog-skribenter har 'eksperimentelt' redigeret faktuelle fejl ind i Wikipedia for at iagttage hvor hurtigt de bliver rettet, — eller om de overhovedet bliver rettet.<sup>39,40</sup> Om denne sabotage-journalistik lykkes kan være afhængig af hvor 'snedigt' og 'snigende' man gør det. For åbenbar vandalisme, hvor store dele af en artikels tekst blive slettet, har et studie foretaget af forskere fra MIT og IBM fundet at det typisk tager blot et par minutter før artiklen bliver gendannet af årvågne wikipedianere.<sup>41</sup>

På trods af at den typiske vandalisme hurtigt bliver rettet betyder det dog ikke at al form for vandalisme og fejlagtig information hurtigt bliver korrigeret: Da journalisten John Seigenthaler fandt sig omtalt i Wikipedia havde der i ikke mindre end 132 dage fejlagtigt stået at han var involveret i Kennedy-mordene.<sup>42</sup> Man kan heller ikke være sikker på at kvalitetsartikler vedbliver med at være det. En relativ stor del, der har

<sup>§</sup>Hverken Nature's eller Stern's undersøgelser var i sig selv peer-reviewet. Der er udført et peer-reviewet studie af Wikipedia's kvalitet.<sup>38</sup> Det var dog ikke et sammenlignende blindet studie.

## Wikipedia i undervisning

Mange brugere vil nok mene at Wikipedia fungerer udmærket til baggrundslæsning. Men kan man bruge Wikipedia som kilde? I begyndelsen af 2007 forbød et institut på Middlebury College de studerende at citere Wikipedia som kilde efter at en hel række studerende havde brugt fejlagtig information på Wikipedia om en bestemt episode i Japan's historie.<sup>43</sup> Men i slutningen af 2007 udtalte Jimmy Wales at han ikke så problemer i at yngre studerende brugte den som reference, og at encyklopædien

burde ses som en trædesten til andre kilder.<sup>44</sup>

Det er forholdsvis nemt for en lærer at angive korrekt information på Wikipedia. Det er muligt at lave et permanent link til en bestemt version af en artikel. Læreren kan linke til en version han selv har bedømt, så de studerende kan føle sig nogenlunde sikre på at Wikipedia's information er til at stole på og derved bruge det som supplement til en lærebog.

I både England og USA har studerende fået 'lektier for' i

Wikipedia. På University of East Anglia skal de studerende i International Relations således redigere og skrive artikler om Mellemøsten.<sup>45</sup> De bliver bedømt på forbedringen af otte artikler og skal selv skrive en. De ofte meget kontroversielle emner i dette område kræver at de studerende balancere de forskellige holdning og understøtter dem med kilder. Lignende opgave bliver stillet i miljøhistorie på University of Washington-Bothell.<sup>46</sup>

været erklæret 'featured article' på den engelske Wikipedia, er på et senere tidspunkt degraderet til 'former featured article'.

## Teknisk tillid?

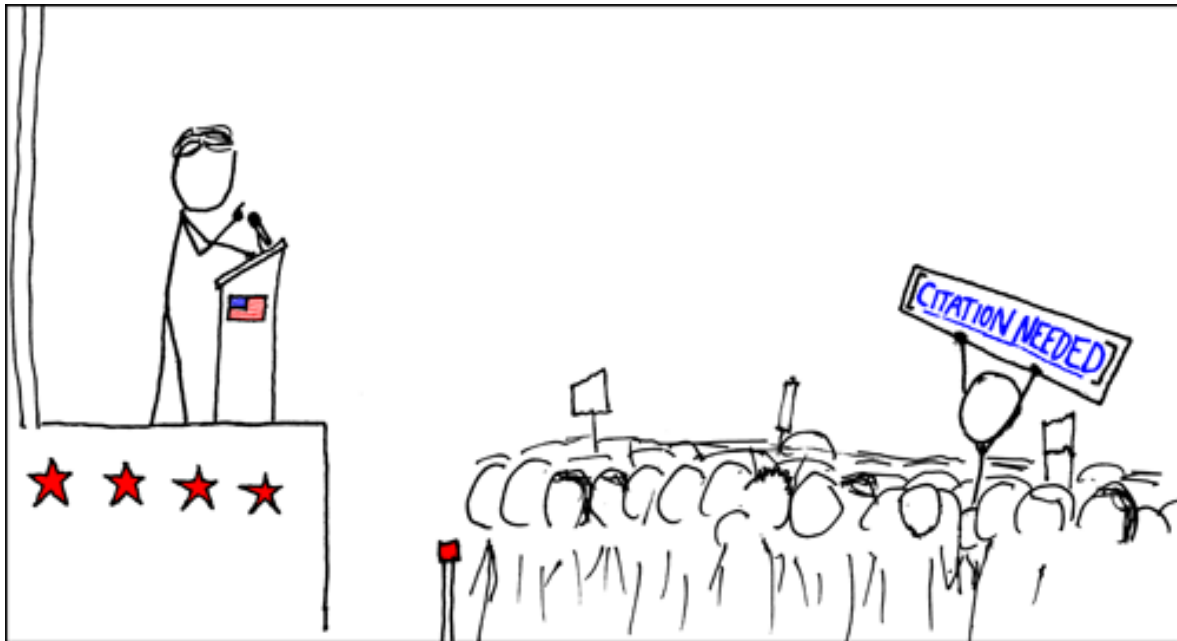
Flere forskellige tekniske måder til at sikre eller i det mindste vurdere kvaliteten af artikler i wikier på er foreslået.

Umiddelbart skulle man tro at hvis man ikke tillod anonyme brugere at redigere i Wikipedia ville man få bedre kvalitet, men besynderlig nok tyder resultater fra en forskergruppes analyse af de franske og nederlandske Wikipediaer på det modsatte: At anonyme redigering faktisk er *bedre* end redigeringer fra registrerede brugere.<sup>47</sup>¶ Der er dog flere andre wikier der har givet køb på en af de grundlæggende wiki-elementer og begrænset redigeringsadgan-

gen: Eugene Izhikevich—en forsker i teoretisk neurobiologi—administrerer en anden wiki, *Scholarpedia*, og tillader kun udvalgte eksperter at skrive artiklernes endelige version. Han har haft held med at tilstrække adskillige Nobelpristagere som skribenter. *Ganfyd* begrænser redigeringen til registrerede læger,<sup>48</sup> mens den generelle encyklopædi *Citizenium* for eksempel ikke tolererer anonymitet. Andre tiltag fokuserer på godkendelse af artiklerne gennem peer-review. Der er allerede mekanismer til peer-review i selve Wikipedia, men derudover er der foreslået kombinationer med eksterne systemer.<sup>49</sup> Et eksempel er *Veropedia*, der har lidt over 5000 checkede Wikipedia-baserede artikler: De kalder dem 'verificerede' artikler.

Problemer som åbenbar vandalisme kan håndteres af 'recent changes' patruljen, der er en løs gruppe af frivillige der konstant

¶ Det er dog ikke alle der finder at anonyme redigeringer er af det gode: En stor del af vandalismen kommer fra anonyme brugere, se for eksempel User:SallyScot/Anonymous\_edit\_analysis (oldid 178592175) på den engelske Wikipedia.



Figur 1: Kilde mangler! I Wikipedia er det nemt at angive at man ikke finder et udsagn understøttet ved at indsætte et ‘kilde mangler’. Det kan være mere besværlig i ‘virkeligheden’. Fra Randall Munroe’s tegneserie XKCD. © Randall Munroe.

overvåger de seneste ændringer på Wikipedia og korrigerer dårlige ændringer. Men mere tekniske løsninger er også i værk: Wikipedianerne har konstrueret programmer der automatisk checker nyeste redigeringer for vandalisme, og disse programmer kan gendanne en vandaliseret side indenfor et minut.<sup>50</sup> De simple tommelfingerregler som bruges i programmet *ClueBot* kan fange mange vandalisme-forsøg, og programmet overvåger konstant Wikipedia med stor success. Andre har udviklet mere avancerede metoder blandt andet ved at opbygge statistiske modeller på baggrund af artiklernes revisionshistorik.<sup>51–53</sup> Sådanne systemer kan farvelægge tekstfragmenter læseren ikke kan stole på og således angive at det er usikkert at ‘Anders Fjogh Rasmussen’ er Danmarks Statsminister.<sup>54–56</sup> Systemerne kan udnytte at der er en vis grad af sammenhæng mellem en artikels kvalitet, antallet af dens redigeringer og antallet af dens

skribenter,<sup>18,19,57</sup> og at et tekstfragment der har overlevet gennem mange redigeringer ‘nok’ er troværdig. Også fra DTU Informatik kommer et foreslag: Thomas Korsgaard beskrev i sit afgangsprøveprojekt et anbefalings-system baseret på tilbagemeldinger fra betroede brugere.<sup>58</sup>

Kildehenvisninger er en indirekte måde at sikre kvaliteten af en Wikipedia artikel på. Disse henvisninger kan referere til blandt andet bøger, nyhedsartikler, eller troværdigt materiale på Internettet. Er der ikke angivet en kilde til et udsagn i en artikel og tror man ikke på det, kan man som skribent hurtigt indsætte `{{fact}}` i den engelske eller `{{kilde mangler}}` i den danske Wikipedia og dermed angive at ‘nogen’ bør sørge for en kildeangivelse, — eller slette informationen. Efterhånden er ganske mange artikler understøttet med henvisninger. I slutningen af 2007 talte jeg at der var over 70.000 strukturerede henvisninger i

<sup>50</sup>Wikipedia trust coloring demo, <http://trust.cse.ucsc.edu/>

## Kan det blive ved?

Det bliver interessant at om Wikipedia i fremtiden kan modstå 'angreb' fra hvad man kunne kalde 'kompetente vandaler', — skribenter der ikke laver den vanlige grafitti-vandalisme, men skriver kompetente bidrag der blot er vægtet bort fra Wikipedia's 'neutrale synspunkt'. En virksomhed eller en organisation, der ønsker et godt omdømme, vil være interesseret i at de fremstår positivt på Wikipedia,<sup>59</sup> — og de ville kunne slette negativ omtale og fremhæve positiv. Fortsætter Wikipedia's popularitet med at stige

vil det blive mere og mere relevant at redigere i ens 'Wikipedia image'. Kan de mange frivillige Wikipedianere modstå det pres der nok opstår fra fuldtidsansatte PR-folk?<sup>60</sup>

En af de første historier der synliggjorde problemet var Microsoft's forsøg på at betale en skribent til at rette i en artikel om åbne dokumentstandarder.<sup>61</sup> Senere gjorde den amerikanske ph.d.-student Virgil Griffith det nemt at opspore virksomheder's Wikipedia-redigeringer med sin Internet-service *Wiki-Scanner*. Det medførte en del

omtale i pressen. Blandt andet kunne *BBC* berette at 'nogen' fra Diebold, fabrikanten af problematiske afstemningsmaskiner, havde slettet information.<sup>62</sup> Nogle måneder efter kunne avisen *The Independent* skrive at 'nogen' fra BBC havde rettet i artiklen 'Criticism of the BBC'...<sup>63</sup> Muligheden for at opspore redigeringerne fik det nederlandske justitsministerium til midlertidigt at forbyde dets 30.000 ansatte i at redigere i Wikipedia.<sup>64</sup> En Wikiscanner for den danske Wikipedia findes også.<sup>65</sup>

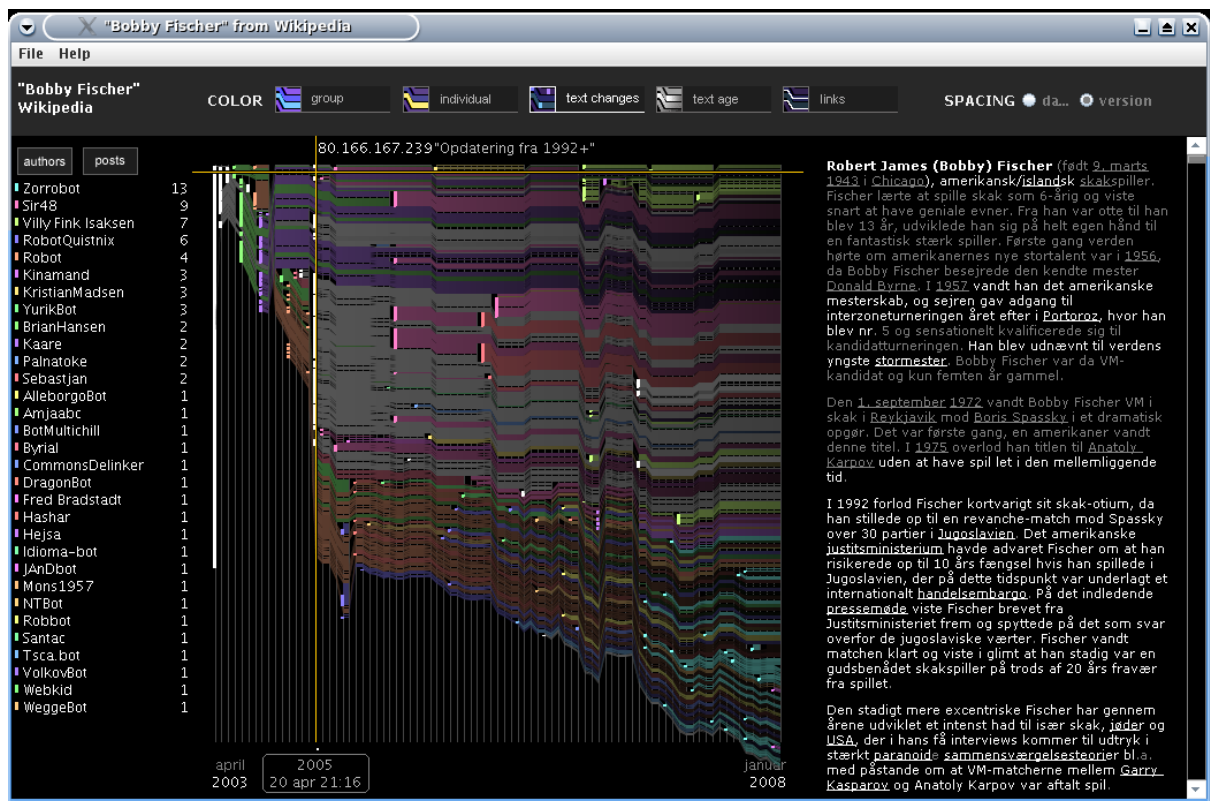
Wikipedia til videnskabelige tidsskriftsartikler. Tallet var mere end en fordobling på et år. Jeg undersøgte videre i hvor høj grad disse henvisninger stemte overens med citeringerne man ser indenfor den videnskabelige litteratur,<sup>66</sup> Det viser sig at det gør de ganske godt. Når antallet af henvisninger til de forskellige tidsskrifter bliver sammenlignet med tal fra *Journal Citation Reports* er der en ganske god overensstemmelse således at et tidsskrift der modtager mange Wikipedia-henvisninger også får mange 'videnskabelige' citeringer. For eksempel var de to mest citerede tidsskrifter fra Wikipedia *Nature* og *Science*. Dog er specielt astronomi og astrofysiske tidsskrifter mere citeret fra Wikipedia end tallene fra Journal Citation Reports ville forudsige. Og overraskende er også australske botaniske tidsskrifter mere citeret end forventet. Det skyldes at en gruppe wikipedianere har sat sig for at skrive vel-understøttede artikler om den farvestrålende planteslægt *Bank-*

*sia*, og ikke mindre end fire *Banksia*-artikler har opnået Wikipedia's ærefulde 'featured' status.

## Hvorfor virker Wikipedia?

Trods de mange tekniske forslag til vurdering af tiltro er grunden til at Wikipedia virker nok den at Wikipedia-systemet har været i stand til at tilstrække og fastholde de entusiastiske samarbejdende encyklopædionørder og været i stand til at gøre det besværlig for ondsindede eller skødesløse vandaler: At der er flere nørder end vandaler.<sup>54</sup>

At Wikipedia kan tiltrække skribenter skyldes blandt andet den lave barriere for at gå i gang.<sup>67,68</sup> For eksempel tager det blot to museklik og et tastetryk at rette en kommafejl. En del wikipedianer nævner også at de begyndte at redigere i Wikipedia fordi de opdagede fejl eller mangler ved en artikel som de vidste noget om.<sup>68</sup> Betyder det så



Figur 2: Programmet *history flow* fra IBM kan visualisere en artikels udvikling og nemt give et overblik over hvornår store ændringer forekommer: Den danske Wikipedia's artikel om Bobby Fischer bliver markant udvidet den 20. april 2005 af en anonym bruger. 'Zorrobot' står dog for de fleste redigeringer. Zorrobot er et norsk program der automatisk opdaterer de små henvisninger til Wikipedia-artikler om Bobby Fischer på andre sprog.

at jo mere korrekt og fuldstændig Wikipedia bliver jo svære vil den have i at tiltrække nye brugere?

Et vigtigt element i en åben wiki som Wikipedia er revisionssystemet, der bevarer gamle versioner af artiklerne. Således kan vandaler godt nok slette skrevet information, men det er dog let at genetablere ud fra gamle gemte versioner. Det er samtidig også nemt at holde øje med andre brugeres bidrag. En andet teknisk element der har betydning for kvaliteten er diskussionssiderne der er knyttet til hver enkelt artikels egentlige side. På denne type sider kan wikipedianerne disku-

tere og klage over problemer med for eksempel unøjagtigheder, læsbarhed og relevans.<sup>69</sup> Revisionssystemet og diskussionssiderne muliggør også at en wikipedianer kan opbygge et godt omdømme. Gratisarbejdende Open Source programmører kan være drevet af at få anerkendelse.<sup>70</sup> Lignende gør sig nok til en vis grad gældende for wikipedianere.<sup>67,71</sup> Der findes flere web-servicer der kan optælle en wikipedianers antal redigeringer, og statistik med 'List of Wikipedians by number of edits' opgøres, hvor en vis Rich Farmbrough har haft 'bragging rights' som den ivrigste med sine 176.393 redigeringer i december 2007.\*\*

\*\*Redigeringsstatistikken er på Wikipedia's artikel 'Wikipedia:List of Wikipedians by number of edits' (oldid = 180423533). Web-servicen *WikiDashboard* ved Palo Alto Research Center giver også en ganske god oversigt.

Desuden kan wikipedianere give hinanden de såkaldte *barnstars*, der er udemærkelser for godt arbejde.

Skabelsen af Wikipedia er en decentral proces uden økonomisk incitament, hvor individuelle bidrag bliver sammenføjet til et kollektivt resultat. Visse web-spil har samme karakter: Hollywood Stock Exchange er et web-baseret markedsspil hvor det blandt andet gælder om at forudsige hvor meget en film vil indtjene. Deltagere kan gøre dette ved frit at købe og sælge aktier i den kunstige valuta 'Hollywood dollars'. Fire uger efter filmpremieren indløses Hollywood aktierne i forhold til den virkelige indtjening. Prisen på aktien kommer således til at virke som en forudsigelse på en films virkelige indtjening. Tilbage i 2000 var jeg i en gruppe der undersøgte kunstige markedsspil's evne til at spå.<sup>72,73</sup> Det gør de overraskende godt. For eksempel kunne Hollywood Stock Exchange forudsige Oscar-vindere bedre end udpegede eksperter. Tilsyneladende tiltrækker spåmarkederne vel-informerede og vel-motiverede spillere hvis indbyrdes handel skaber information, også selv om de frivillige deltagere ikke vinder virkelige penge. Der er ingen egentlig handel i Wikipedia, men wikipedianerne får skabt et miljø hvor en gensidig udveksling af information kollektivt skaber en af verdens mest besøgte hjemmesider.<sup>††</sup>

## Litteratur

- [1] Juan Carlos Perez. Wikipedia breaks into U.S. top 10 sites. *PCWorld*, February 2007.
- [2] Donna Shaw. Citing Wikipedia. *American Journalism Review*, February/March 2008.

---

<sup>††</sup>Der er en hel del flere videnskabelige artikler om og med Wikipedia end dem jeg har beskrevet her.<sup>74-85</sup> På dansk har universitetsstuderende også skrevet om Wikipedia.<sup>86,87</sup> I øjeblikket er der også en lang liste på den engelske Wikipedia-side 'Wikipedia:Wikipedia in academic studies'.

- [3] S. B. Sinha and Harjit Singh Bedi. Commr. Of Customs, Bangalore vs. m/s ACER India Pvt. Ltd. The Judgement Information System, December 2007. Appeal (civil) 2321 of 2007.

ANNOTATION: A case in the Supreme Court of India where Wikipedia is used for the definition of 'laptop'.

- [4] Noam Cohen. Courts turn to Wikipedia, but selectively. *The New York Times*, January 2007.
- [5] Jimmy Wales. Testimony of Jimmy Wales founder of the Wikipedia and of the Wikimedia Foundation regarding "e-government 2.0: Improving innovation, collaboration, and access" before the U.S Senate committee on homeland security and governmental affairs, December 2007.
- [6] Torsten Kleinz and Craig Morris. German Wikipedia receives state funding. *heise online*, July 2007.
- [7] Chris Taylor. It's a wiki, wiki world. *TIME*, May 2005.
- [8] Jakob Voß. Measuring Wikipedia. In *Proceedings International Conference of the International Society for Scientometrics and Informetrics : 10th*, 2005.
- [9] Aaron Swartz. Who writes Wikipedia. Raw Thought, September 2006. Blog.
- [10] Felipe Ortega, Jesús M. González-Barahona, and Gregorio Robles. On the inequality of contributions to Wikipedia. In *Proceedings of the 41st Hawaiian International Conference on System Sciences (HICSS-2008)*, 2008.



ANNOTATION: Computation of the Gini coefficient for author contributions for the top ten largest Wikipedias. The find the Gini coefficients to be from 0.92 to 0.97. Have they excluded bots? Which namespaces are they looking at?

- [11] Wray Buntine. Static ranking of web pages, and related ideas. In Michel Beigbeder and Wai Gen Yee, editors, *Open Source Web Information Retrieval*, October 2005.
- [12] Francesco Bellomi and Roberto Bonato. Network analysis of Wikipedia. In *Proceedings of Wikimania 2005 — The First International Wikimedia Conference*, 2005.

ANNOTATION: Describes results of application of PageRank and Kleinberg’s HITS algorithm on the English Wikipedia corpus. “United States” scored highest in both. Entries related to religion scored high PageRank.

- [13] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74:036116, September 2006.
- [14] V. Zlatić, M. Božičević, H. Štefančić, and M. Domazet. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74(1):016115, July 2006.
- [15] Todd Holloway, Miran Božičević, and Katy Börner. Analyzing and visualizing the semantic coverage of Wikipedia

and its authors. *Complexity*, 12(3):30–40, January 2007.

- [16] Torsten Zesch and Iryna Gurevych. Analysis of Wikipedia category graph for NLP applications. In *TextGraphs-2: Graph-Based Algorithms for Natural Language Processing. Proceedings of the Workshop*, pages 1–8, New Brunswick, New Jersey, USA, April 2007. The Association for Computational Linguistics.

ANNOTATION: Description of the use of the Wikipedia category graph for determining semantic relatedness. Several different similarity and distance measures on the graph are examined on several human labeled datasets from the German Wikipedia. Graph characteristics (average shortest path, cluster coefficient and power law exponent) are also shown.

- [17] A. Capocci, F. Rao, and G. Caldarelli. Taxonomy and clustering in collaborative systems: The case of the on-line encyclopedia Wikipedia. *EPL*, 81(2):28006, January 2008.
- [18] Dennis M. Wilkinson and Bernardo A. Huberman. Cooperation and quality in Wikipedia. In *Proceedings of the 2007 international symposium on Wikis*, pages 157–164, New York, NY, USA, 2007. ACM.

ANNOTATION: An investigation of what signals quality articles on the English Wikipedia. A growth model is suggested for the number of edits on an article, and this lognormal model shows

good agreement with the observed number of edits. It is also shown that older articles typically have more edits than newer articles and that quality articles have comparably more edits, many distinct editors, many edits in its talk page, a quick ‘turnaround’ and a high number of edits per editor when correcting for article age and Google PageRank. They write that featured articles with high Google PageRank have a lower number of talk page edits than articles with low PageRank. It is not clear if that effect is due to archiving.

- [19] Dennis M. Wilkinson and Bernardo A. Huberman. Assessing the value of cooperation in *Wikipedia*. *First Monday*, 12(4), April 2007.
- [20] Mike Rogoway. Inventor of the wiki has a new job in Portland. *OregonLive.com*, May 2007.
- [21] Curtis Krueger. There’s no end to it. *Floridan*, November 2004.
- [22] Jakob Voß. Collaborative thesaurus tagging the Wikipedia way. arXiv, April 2006.
- [23] Daniel H. Pink. Folksonomy. *The New York Times*, December 2005.
- [24] Michael Strube and Simone Paolo Ponzetto. WikiRelate! computing semantic relatedness using Wikipedia. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, pages 1419–1424, Menlo Park, California, 2006. AAAI Press.
- [25] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatic extraction of semantic relationships for WordNet by means of pattern learning from Wikipedia. In *Natural Language Processing and Information Systems*, volume 3513 of *Lecture Notes in Computer Science*, pages 67–79, Berlin/Heidelberg, May 2005. Springer.
- [26] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets. In *Advances in Web Intelligence*, volume 3528 of *Lecture Notes in Computer Science*, pages 380–386, Berlin/Heidelberg, 2005. Springer.
- [27] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. *Data & Knowledge Engineering*, 61(3):484–499, June 2007.
- [28] Andrew Krizhanovsky. Synonym search in Wikipedia: Synarcher. In *11th International Conference “Speech and Computer” SPECOM’2006. Russia, St. Petersburg, June 25–29*, pages 474–477, 2006.
- ANNOTATION: Description of the Synarcher program that analyze Wikipedia with the Kleinberg HITS algorithm and present related terms to a query term in a graph visualization.
- [29] Evgeniy Gabrilovich and Shaul Markovitch. Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge. In *Proceedings of the Twenty-First National Conference*

- on *Artificial Intelligence*, pages 1301–1306, Menlo Park, California, 2006. AAAI Press.
- [30] Evgeniy Gabrilovich and Shaul Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proceedings of The Twentieth International Joint Conference for Artificial Intelligence*, pages 1606–1611, 2007.
- [31] Magnús Sigurdsson and Søren Christian Halling. Zeeker: A topic-based search engine. Master’s thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Kongens Lyngby, Denmark, 2007.
- [32] Jenny Kleeman. Wikipedia ban for disruptive professor. *The Observer*, December 2007.
- [33] Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, December 2005.
- ANNOTATION: Report of a comparison of the accuracy in Wikipedia and Encyclopedia Britannica
- [34] Wikipedia schlägt Brockhaus. *Stern*, December 2007.
- [35] K. C. Jones. German Wikipedia outranks traditional encyclopedia’s online version. *InformationWeek*, December 2007.
- [36] Wikipedia schlägt die Profis. *Spiegel Online*, December 2007.
- [37] Torsten Kleinz. Reputation und Wikipedia: von Einfluss, Anerkennung und Vergleichen. *heise online*, December 2007.
- [38] Thomas Chesney. An empirical examination of Wikipedia’s credibility. *First Monday*, 11(11), 2006.
- [39] Alex Halavais. The Isuzu experiment. *a thaumaturgical compendium*, August 2004. Blog.
- [40] How authoritative is Wikipedia. Dispatches from the Frozen North, September 2004. Blog.
- [41] Fernanda B. Viégas, Marting Wattenberg, and Kushal Dave. Studying cooperation and conflict between authors with *history flow* visualizations. In Elizabeth Dykstra-Erickson and Manfred Tscheligi, editors, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 575–582, New York, NY, USA, 2004. ACM.
- [42] John Seigenthaler. A false Wikipedia ‘biography’. *USATODAY.com*, 2005.
- [43] Noam Cohen. A history department bans citing Wikipedia as a research source. *The New York Times*, February 2007.
- ANNOTATION: Middlebury College history department has banned students from using Wikipedia as a source
- [44] Alistair Coleman. Students ‘should use Wikipedia’. *BBC NEWS*, December 2007.
- [45] Donald MacLeod. Students marked on writing in Wikipedia. *Guardian Unlimited*, March 2007.
- [46] Associated Press. University class has students updating Wikipedia. *FOX-News.com*, November 2007.
- [47] Denise Anthony, Sean W. Smith, and Tim Williamson. The quality of open

source production: Zealots and good samaritans in the case of *Wikipedia*. Technical Report TR2007-606, Department of Computer Science, Dartmouth College, Hanover, New Hampshire, April 2007.

ANNOTATION: An investigation of 7058 contributors on the French and Dutch Wikipedias, to answer whether contributions from registered users have higher quality than anonymous users. Each user is measured wrt. retention of their added text, while not considering, e.g., edit wars. They find that the retention is higher for anonymous users.

- [48] Brandon Keim. WikiMedia. *Nature Medicine*, 13(3):231–233, February 2007.

ANNOTATION: An editorial describing wikis used in medicine and biology. Wikipedia, Digital Universe, Citizendium, Ganfyd, Ask-DrWiki and OpenWetWare are described.

- [49] Kermit K. Murray. Mass spectrometry on Wikipedia: Open source and peer review. Presented at the 55th ASMS Conference on Mass Spectrometry, June 3–7, 2007, Indianapolis, Indiana, June 2007.
- [50] Jacobi Carter. ClueBot and vandalism on Wikipedia, 2007.
- [51] Pierpaolo Dondio, Stephen Barrett, Stefan Weber, and Jean Marc Seigneur. Extracting trust from domain analysis: A case study on the Wikipedia project. In *Autonomic and Tru-*

*sted Computing*, volume 4158 of *Lecture Notes in Computer Science*, pages 362–373, Berlin/Heidelberg, October 2006. Springer.

- [52] Honglei Zeng, Maher Alhossaini, Li Ding, Richard Fikes, and Deborah L. McGuinness. Computing trust from revision history. In *Proceedings of the 2006 International Conference on Privacy, Security and Trust*, October 2006.

ANNOTATION: Sets up a statistical model (a dynamic Bayesian network) for modeling of Wikipedia article trust by looking at the revision history.

- [53] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. Investigations into trust for collaborative information repositories: A Wikipedia case study. In *Proceedings of the Workshop on Models of Trust for the Web*, Edinburgh, United Kingdom, May 2006.
- [54] B. Thomas Adler and Luca de Alfaro. A content-driven reputation system for the Wikipedia. In *Proceedings of the 16th international conference on World Wide Web*, pages 261–270, New York, NY, USA, 2007. ACM.
- [55] B. Thomas Adler, J. Benterou, K. Chatterjee, Luca de Alfaro, I. Pye, and V. Raman. Assigning trust to Wikipedia content. Technical Report UCSC-CRL-07-09, School of Engineering, University of California, Santa Cruz, CA, USA, November 2007.
- [56] Brian Bergstein. New tool measures Wikipedia entries. *USATODAY.com*, September 2007.

- [57] Andrew Lih. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In *5th International Symposium on Online Journalism*. The University of Texas at Austin, April 2004.
- ANNOTATION: A study examining the inbound citations to Wikipedia from press articles published in 2003 and parts of 2004. It displays “diversity” (number of individual editors of an article) against “rigor” (number of edits for an article) in some Wikipedia articles and the study regard these as indicators of quality. World War II, Islam and Astronomy scored high on these dimensions. The indicators of article quality increased after press citation for some Wikipedia articles.
- [58] Thomas Rune Korsgaard. Improving trust in the Wikipedia. Master’s thesis, Technical University of Denmark, Kongens Lyngby, Denmark, 2007.
- [59] Jens Schnabel. Reklamering på Wikipedia er en gråzone. *Comon.dk*, July 2007.
- [60] Thomas Claburn. Law professor predicts Wikipedia’s demise. *InformationWeek*, December 2006.
- [61] Brian Bergstein. Microsoft offers cash for Wikipedia edit. *Washingtonpost.com*, January 2007.
- [62] Jonathan Fildes. Wikipedia ‘shows CIA page edits’. *BBC NEWS*, August 2007.
- [63] Brian Brady. BBC staff rewrote Wikipedia pages to water down criticism. *The Independent*, December 2007.
- [64] Associated Press. Ministry bans Wikipedia editing. *Guardian Unlimited*, November 2007.
- [65] Asger Westh. Wikipedia: Den elektroniske slagmark. *eptn.dk*, January 2008.
- ANNOTATION: A news article about a version of the WikiScanner for the Danish Wikipedia.
- [66] Finn Årup Nielsen. Scientific citations in Wikipedia. *First Monday*, 12(8), August 2007.
- ANNOTATION: Statistics on the outbound scientific citation from Wikipedia with good correlation to the Journal Citation Reports from Thomson Scientific.
- [67] Andrea Cifollilli. Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12), December 2003.
- [68] Susan L. Bryant, Andrea Forte, and Amy Bruckman. 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, 2005. ACM.
- ANNOTATION: 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.

- [69] Besiki Stvilia, Michael B. Twidale, Les Gasser, and Linda C. Smith. Information quality discussions in Wikipedia. In Suliman Hawamdeh, editor, *Knowledge Management. Nurturing Culture, Innovation, and Technology. Proceedings of the 2005 International Conference on Knowledge Management*, pages 101–113, Singapore, October 2005. World Scientific.

ANNOTATION: An analysis with respect to information quality of a sample of the discussion pages on Wikipedia. The mostly qualitative analysis is based on the their own information quality assessment model, and they provide example quotations from the pages. They conclude that ‘the Wikipedia community takes issues of quality very seriously’.

- [70] Tor Nørretranders. *Det generøse menneske: En naturhistorie om at umage giver mage*. People’sPress, 2006.

ANNOTATION: A book dubbing ‘Homo generosus’ and discussing issues such as altruism, gift economy, generosity, reciprocity, cooperation, moral, costly signals,

the handicap principle, sexual selection and economic games.

- [71] Bernhard Hoisl, Wolfgang Aigner, and Silvia Miksch. Social rewarding in wiki systems — motivating the community. In *Online Communities and Social Computing*, volume 4564 of *Lecture Notes in Computer Science*, pages 362–371, Berlin/Heidelberg, August 2007. Springer.

ANNOTATION: A suggestion for a computation of reputation of the individual authors in a wiki-system. Formulas are shown but it is not clear if the suggestion is tested on a real wiki.

- [72] David M. Pennock, Steve Lawrence, C. Lee Giles, and Finn Årup Nielsen. The real power of artificial markets. *Science*, 291(5506):987–988, February 2001.

ANNOTATION: A brief note reporting that artificial markets, such as Hollywood Stock Exchange and Foresight Exchange, are good predictors of future outcomes.

- [73] David M. Pennock, Steve Lawrence, Finn Årup Nielsen, and C. Lee Giles. Extracting collective probabilistic forecasts from web games. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 174–183, New York, NY, USA, August 2001. SIGKDD, Association for Computing Machinery, ACM.

- [74] Andreas Neus. Managing information quality in virtual communities of practice. In E. Pierce and R. Katz-Haas,

- editors, *Proceedings of the 6th International Conference on Information Quality at MIT*, Boston, MA, 2001. Sloan School of Management.
- [75] Sisay Fissaha Adafre and Maarten de Rijke. Discovering missing links in Wikipedia. In Jafar Adibi, Marko Grobelnik, Dunja Mladenic, and Patrick Pantel, editors, *Proceedings of the 3rd international workshop on Link discovery*, pages 90–97, New York, NY, USA, 2005. ACM.
- [76] William Emigh and Susan C. Herring. Collaborative authoring on the web: A genre analysis of online encyclopedias. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences ([HICSS'05]*, page 99.1, Washington, DC, USA, 2005. IEEE Computer Society.
- [77] Jenneke Fokker, Johan Pouwelse, and Wray Buntine. Tag-based navigation for peer-to-peer Wikipedia. Collaborative Web Tagging Workshop, WWW2006.
- [78] Nikolaos Th. Korfiatis, Marios Poulos, and George Bokos. Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review*, 30(3):252–262, 2006.
- [79] Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113, October 2006.
- [80] Klaus Stein and Claudia Hess. Does it matter who contributes: a study on featured articles in the german Wikipedia. In *Proceedings of the 18th conference on Hypertext and hypermedia*, pages 171–174, New York, NY, USA, 2007. ACM.
- [81] Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He says, she says: Conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 453–462, New York, NY, USA, April 2007. ACM.
- [82] Aniket Kittur, Ed Chi, and Todd Mytkowicz Bryan A. Pendleton, Bongwon Suh. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Computer/Human Interaction 2007*. ACM SIGCHI, 2007. This is not found in the CHI 2007 proceedings.
- [83] Fernanda B. Viégas. The visual side of Wikipedia. In *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, page 85a, Washington, DC, USA, 2007. IEEE Computer Society.
- ANNOTATION: Description of how images are constructed and used on Wikipedia, as well as a survey among 29 ‘featured picture’ contributors.
- [84] Fernanda B. Viégas, Martin Wattenberg, and Matthew M. McKeon. The hidden order of Wikipedia. In *Lecture Notes in Computer Science*, volume 4564, pages 445–454, Berlin/Heidelberg, 2007. Springer.
- [85] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A nucleus for a web of open data. In *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer, 2008.

ANNOTATION: Description of a system that extracts information from the templates in Wikipedia, process them and presents them in various ways. Some of the methods and services they use are MySQL, Virtuoso, OpenCyc, GeoNames, Freebase, SPARQL and SNORQL. The system is available from <http://DBpedia.org>

- [86] Niels Mølbjerg Lund Pedersen and Anders Due. Wikipedia — viden som social handle. Master's thesis, Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark, August 2006.
- [87] Anders J. Jørgensen. Den epistemiske hensigtsmæssighed af åbne kontra lukkede videnskabelige samarbejdsformer — med speciel hensynstagen til peer-reviews og internetfænomenet Wikipedia, January 2007.

\$Id: Nielsen2008Wikipedia.tex,v 1.39 2008/04/09 10:00:29 fn Exp \$