

# Wiki(pedia) and neuroinformatics

Finn Årup Nielsen

Lundbeck Foundation Center for Integrated Molecular Brain Imaging;  
Informatics and Mathematical Modelling,  
Technical University of Denmark;  
Neurobiology Research Unit,  
Copenhagen University Hospital Rigshospitalet  
August 29, 2006

# Myself — Finn Årup Nielsen — fnielsen

Engineer with Ph.D. thesis “Neuroinformatics in Functional Neuroimaging” (Nielsen, 2001)

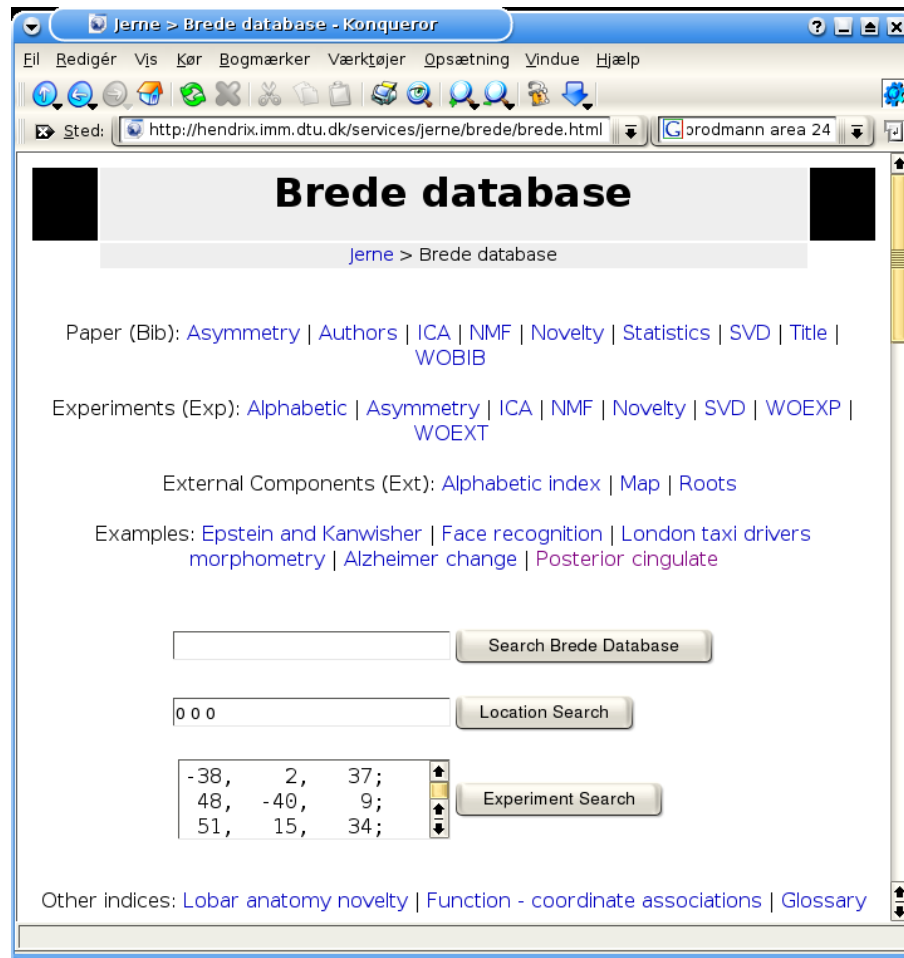
Building mathematical models and computer programs to analyze brain scans.

Building a database and data mining tools for meta-analysis: the “Brede Database” (Nielsen, 2003) and “Brede Toolbox” in the Matlab programming environment (Nielsen and Hansen, 2000). Both distributed on the Internet.

Wikipedia authoring as “fnielsen” of English and Danish versions since 2002. Small edits in private and well as professional interests.

Almost 1000 Danish edits which makes for a rank about 75 disregarding robots.

# Brede Database



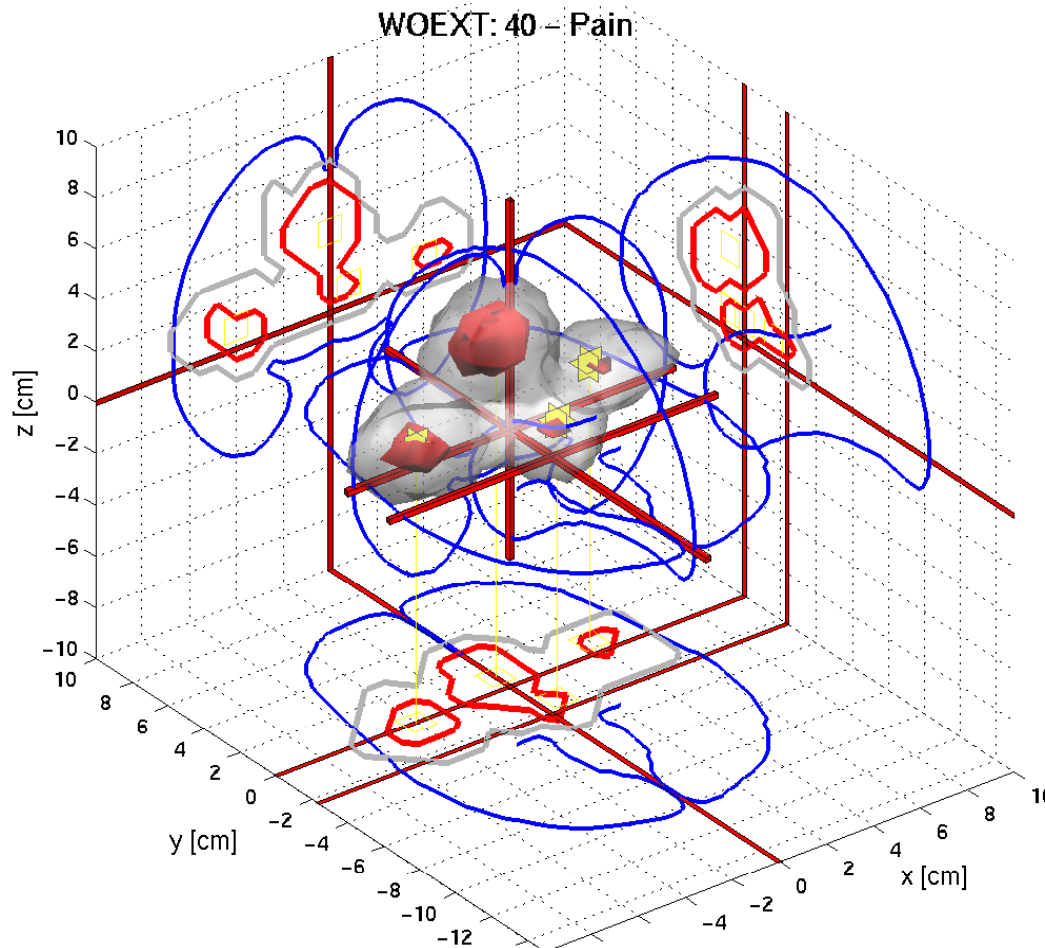
Neuroinformatics database with information from published scientific articles.

Information stored in a simple-format XML

Construction of static web-pages with 3-D renderings with Matlab available on the Internet.

Accompanying Toolbox in Matlab

# Example analysis



Automatic analysis of information from the Brede Database requiring numerical/statistical processing with computer clusters (Nielsen, 2005).

Text mining: multivariate analysis of bag-of-words matrices (Nielsen et al., 2005).

The burden of data entry is large.

# Brede Database and Wikipedia

## WOROI: 5 - Posterior cingulate gyrus

Bib -> [Asymmetry](#) | [Author](#) | [ICA](#) | [NMF](#) | [Novelty](#) | [Statistics](#) | [SVD](#) | [Title](#) | [WOBIB](#) ]

Roi -> [Alphabetic](#) | [Hammers](#) | [Tzourio-Mazoyer](#) ]

[ [Brede](#) ]

### WOROI: 5 - Posterior cingulate gyrus

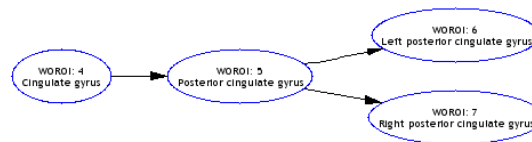
Abbreviation: PCgG

Variation: *posterior cingulate*

Variation: *posterior cingulate area*

Variation: *posterior gyus cinguli*

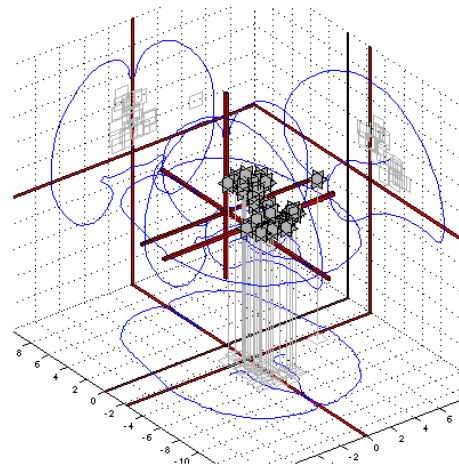
BrainInfo: [144](#)



Parents	Siblings	Children
<a href="#">Cingulate gyrus</a>		<a href="#">Left posterior cingulate gyrus</a> <a href="#">Right posterior cingulate gyrus</a>

#### Talairach coordinates

x	y	z	Lobar anatomy	WOBIB	WOEXP
6	-29	38	Right posterior cingulate gyrus and precuneus	<a href="#">21</a>	<a href="#">66</a>
9	-53	14	Right posterior cingulate gyrus	<a href="#">32</a>	<a href="#">109</a>
4	-53	14	Right posterior cingulate gyrus	<a href="#">32</a>	<a href="#">110</a>
2	-40	40	Posterior cingulate gyrus	<a href="#">35</a>	<a href="#">117</a>
52	-30	20	Right postcentral gyrus/posterior cingulate gyrus	<a href="#">35</a>	<a href="#">119</a>
-4	-36	24	Left posterior cingulate gyrus	<a href="#">41</a>	<a href="#">135</a>
-4	-35	29	Left posterior cingulate gyrus	<a href="#">41</a>	<a href="#">137</a>
-4	-35	40	Left posterior cingulate gyrus	<a href="#">41</a>	<a href="#">138</a>
0	-26	29	Posterior cingulate gyrus	<a href="#">41</a>	<a href="#">140</a>
-2	-48	20	Left posterior cingulate gyrus	<a href="#">49</a>	<a href="#">164</a>
-9	-33	46	Posterior cingulate gyrus	<a href="#">57</a>	<a href="#">183</a>
0	-17	28	Right posterior cingulate gyrus	<a href="#">60</a>	<a href="#">186</a>
3	-53	15	Right posterior cingulate gyrus	<a href="#">71</a>	<a href="#">223</a>



Hard coded deep links in brain region taxonomy of the Brede Database to Wikipedia entries, NeuroNames (another taxonomy) (Bowden and Martin, 1995), CoCoMac (another database) (Kötter, 2004), NIH Mesh terms and labeled volumes (Hammers et al., 2002; Tzourio-Mazoyer et al., 2002; Svarer et al., 2005).

# Wikipedia and neuroinformatics

Collaborative and incremental web-based entering would be useful in a neuroinformatics database.

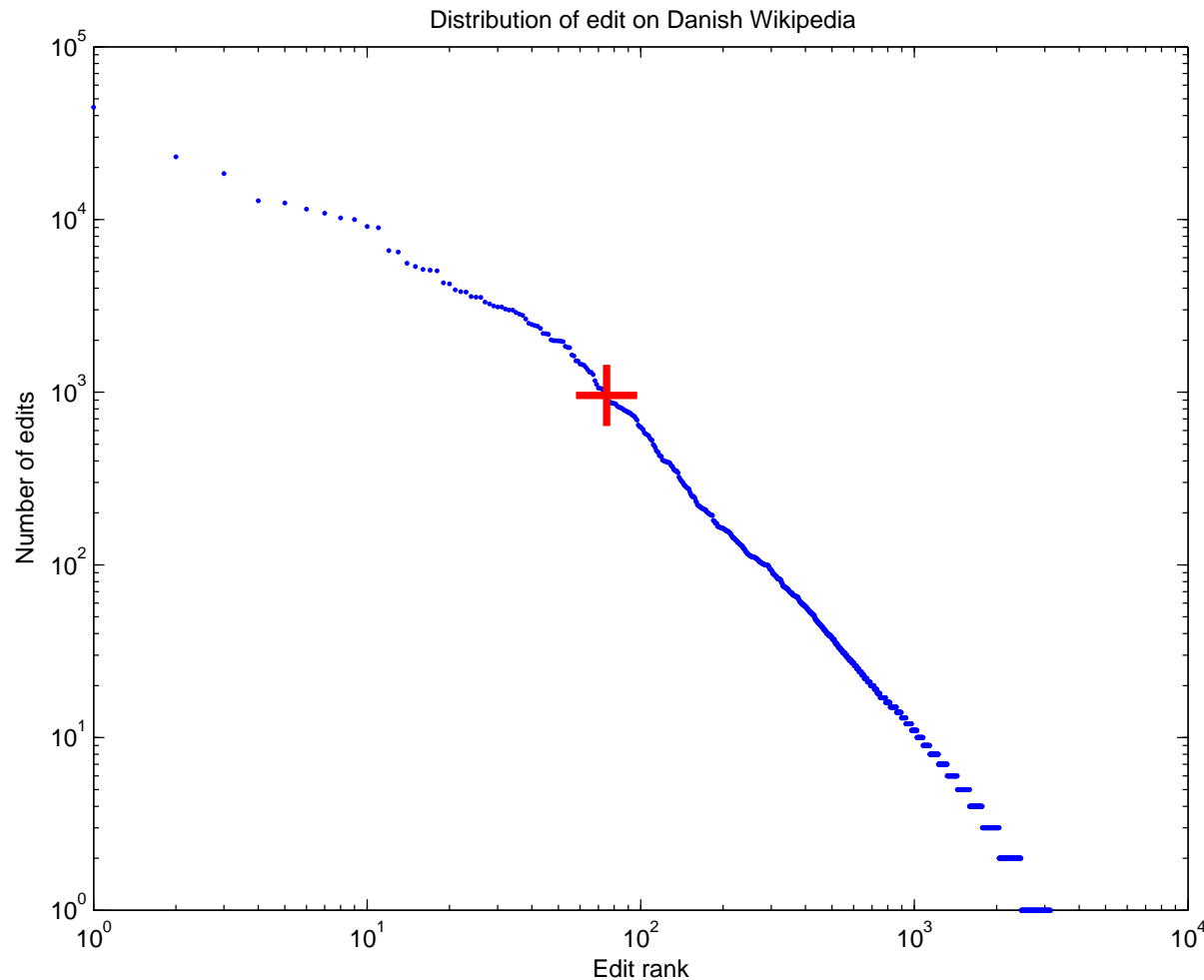
Structured fields are important. Templates, infoboxes? Semantic Wikipedia or Wikidata may be interesting.

Extensible database: Flexible fields to accomodate new ideas that are generated in research

Specialized interface for entering data.

Online numerical processing? And generation of visual elements? Specialized searches.

# Wikipedia research?



Distribution of edits by users on the Danish Wikipedia. Rank on x-axis. (Myself indicated with the red cross.)

Similar to (Voss, 2005, Fig. 6)

# Wikipedia clustering? Preliminaries

Construct binary matrix  $\mathbf{X}$ (articles  $\times$  authors) with 1 indicated an edit.

Excluding usernames matching “bot” and documents beginning with “Wikipedia”.

Exclude articles with less than three different authors.

Danish Wikipedia:  $\mathbf{X}$ (12774  $\times$  3149) with density 0.0025

Some kind of normalization? The results may depend on the exact kind.

Non-negative matrix factorization (Lee and Seung, 2001) — one of the algorithms “off the shelf” in the Brede Toolbox (Nielsen and Hansen, 2000).



## Wikipedia clustering. Some cluster example

Danish Kings: Christian 3., Christoffer 1., Erik Klipping, Frederik 1.

Countries: Portugal, Slovenien, Polen, Tyskland, Belgien, Estland

2006: Skabelon:Aktuelle begivenheder 2006, FC København, Fodbold, Tour de France, Lordi, VM i fodbold 2006, Muhammed-tegningerne, Michael Rasmussen

Danish municipalities and counties: Roskilde Amt, Birkerød Kommune, Frederikssund Kommune

Years: 2003, 2001, 2004, 2005

Discussion: Bruger diskussion:User#1, Bruger diskussion:User#2, Jesus fra Nazaret, Kristendom, Anders Fogh Rasmussen, Diskussion:Dansk Folkeparti, Diskussion:Muhammed-tegningerne, Kreationisme

## Wikipedia clustering

The cluster results will depend critical on the weighting of authors and titles.

With no weighting very active authors will dominate the cluster results.

Changing the weighting will show different aspects of the corpus.

Some of the clusters are related to the Category pages of Wikipedia.

Applications?

# References

- Bowden, D. M. and Martin, R. F. (1995). NeuroNames brain hierarchy. *NeuroImage*, 2(1):63–84. PMID: 9410576. ISSN 1053-8119.
- Hammers, A., Koepp, M. J., Free, S. L., Brett, M., Richardson, M. P., Labbé, C., Cunningham, V. J., Brooks, D. J., and Duncan, J. (2002). Implementation and application of a brain template for multiple volumes of interest. *Human Brain Mapping*, 15(3):165–174. DOI: 10.1002/hbm.10016. <http://www3.interscience.wiley.com/cgi-bin/abstract/89013541/>. ISSN 1065-9471. Describes a segmentation of the MNI single subject brain. Assessment of the method by using manual labeling of landmarks and exemplified on a FMZ PET study.
- Kötter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics*, 2(2):127–144. PMID: 15319511. <http://www.cocomac.org/cocomac2004.pdf>.
- Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Leen, T. K., Dietterich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, pages 556–562, Cambridge, Massachusetts. MIT Press. <http://hebb.mit.edu/people/seung/papers/nmfconverge.pdf>. CiteSeer: <http://citeseer.ist.psu.edu/-lee00algorithms.html>.
- Nielsen, F. Å. (2001). *Neuroinformatics in Functional Neuroimaging*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Nielsen, F. Å. (2003). The Brede database: a small database for functional neuroimaging. *NeuroImage*, 19(2). <http://208.164.121.55/hbm2003/abstract/abstract906.htm>. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.
- Nielsen, F. Å. (2005). Mass meta-analysis in Talairach space. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 985–992, Cambridge, MA. MIT Press. [http://books.nips.cc/papers/files/nips17/NIPS2004\\_0511.pdf](http://books.nips.cc/papers/files/nips17/NIPS2004_0511.pdf).

- Nielsen, F. Å., Balslev, D., and Hansen, L. K. (2005). Mining the posterior cingulate: Segregation between memory and pain component. *NeuroImage*, 27(3):520–532. DOI: 10.1016/j.neuroimage.2005.04.034.
- Nielsen, F. Å. and Hansen, L. K. (2000). Experiences with Matlab and VRML in functional neuroimaging visualizations. In Klasky, S. and Thorpe, S., editors, *VDE2000 - Visualization Development Environments, Workshop Proceedings, Princeton, New Jersey, USA, April 27–28, 2000*, pages 76–81, Princeton, New Jersey. Princeton Plasma Physics Laboratory. [http://www.imm.dtu.dk/pubdb/views/edoc\\_download.php/1231/pdf/imm1231.pdf](http://www.imm.dtu.dk/pubdb/views/edoc_download.php/1231/pdf/imm1231.pdf). CiteSeer: <http://citeseer.ist.psu.edu/309470.html>.
- Svarer, C., Madsen, K., Hasselbalch, S. G., Pinborg, L. H., Haugbøl, S., Frøkjær, V. G., Holm, S., Paulson, O. B., and Knudsen, G. M. (2005). MR-based automatic delineation of volume of interest in human brain PET imaging using probability maps. *NeuroImage*, 24(4):969–979. PMID: 15670674. DOI: 10.1016/j.neuroimage.2004.10.017.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289. DOI: 10.1006/nimg.2001.0978.
- Voss, J. (2005). Measuring wikipedia. In *Proceedings International Conference of the International Society for Scientometrics and Informetrics : 10th*. <http://eprints.rclis.org/archive/00003610/>.