Knowledge discovery in neuroinformatics

DTU

Technical University of Denmark, DTU Informatics

COGNITIVE SYSTEMS SECTION

Neuroinformatics Research Group

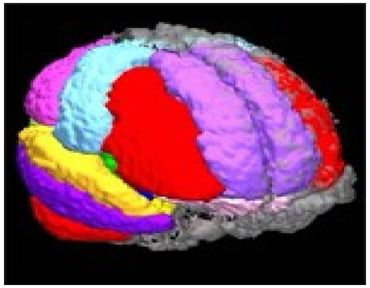
” Coordinate-based meta-analytic search of neuroscientific literature and its expansion using semantic keyword extraction”

Speakers: BARTŁOMIEJ WILKOWSKI
MARCIN SZEWCZYK

Cimbi

THE LUNDBECK FOUNDATION

Center for integrated molecular brain imaging



National Institutes of Health (NIH), 9000 Rockville Pike, Bethesda, Maryland 20892 – June 25, 2009

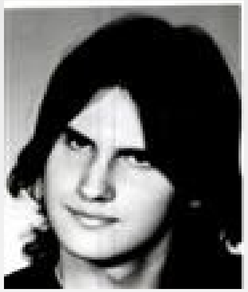
Neuroinformatics Research Group

Professor Lars Kai Hansen

Finn Årup Nielsen (Senior Researcher)

Bartlomiej Wilkowski (PhD Student)

Marcin Marek Szewczyk (Research Assistant)



Peter Mondrup Rasmussen (PhD Student)

Roadmap

- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

Roadmap

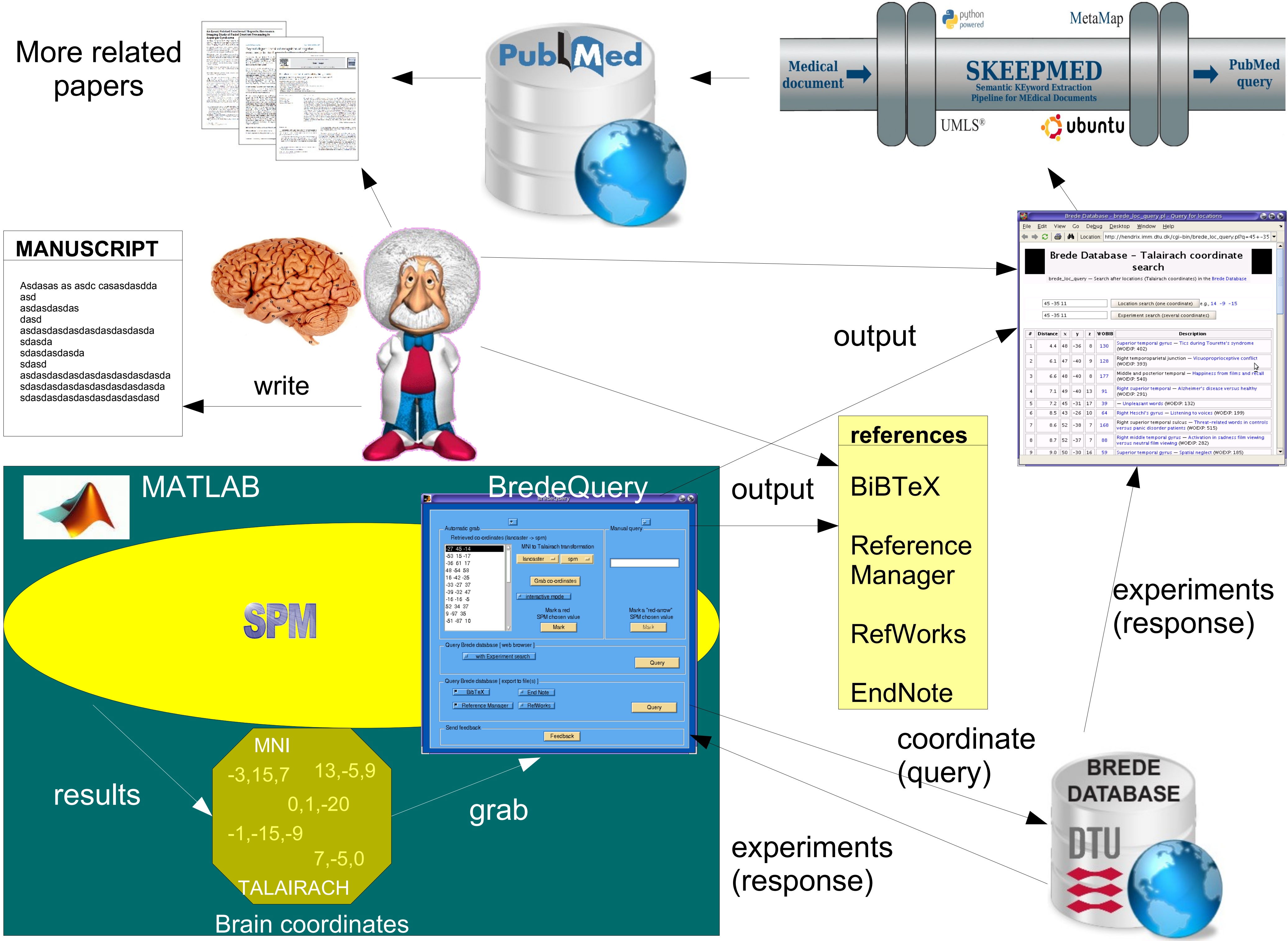
- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

Motivations

- Growing number of functional neuroimaging studies → demand for:
  - Data integration,
  - Data dissemination between research centers;
    (Ascoli, 2006) – „The Ups and Downs of Neuroscience Shares”
    (Teeters et al., 2008) - „Data Sharing for Computational Neuroscience”
- Functional localization hypothesizes that a given human behavior is established by a change in brain activity in a relatively limited number of spatially segregated processing units → → demand for:
  - Efficient (coordinate/localization-based) searching of references to any related literature;

Project overview

- Develop the tools for meta-analysis and efficient searching of related literature/experiments given coordinate(s) in brain (knowledge discovery):
  - Database offering coordinate-based querying service
  - Software to facilitate literature searching directly from neuroscientists' common environments (SPM, FSL, ...)
  - Extending coordinate-based search results by querying bigger, more comprehensive databases like PubMed
  - Creating a secure web-service for neuroscience for stimulation of data and experience dissemination among research groups



More related papers
PubMed
Medical document
SKEEPMED
Semantic KEyword Extraction Pipeline for MEdical Documents
PubMed query
python powered
MetaMap
UMLS®
ubuntu
MANUSCRIPT
Asdasas as asdc casasdasdda
asd
asdasdasdas
dasd
asdasdasdasdasdasdasdasda
sdasda
sdasdasdasda
sdasd
asdasdasdasdasdasdasdasdasda
sdasdasdasdasdasdasdasdasda
sdasdasdasdasdasdasdasdasd
write
Brede Database – Talairach coordinate search
output
references
BiBTeX
Reference Manager
RefWorks
EndNote
MATLAB
BredeQuery
SPM
output
experiments (response)
coordinate (query)
MNI
-3,15,7
13,-5,9
0,1,-20
-1,-15,-9
7,-5,0
TALAIRACH
results
grab
Brain coordinates
experiments (response)
BREDE DATABASE
DTU

Roadmap

- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

Brede Database

- Close to 4000 coordinates from 186 papers with a total of 586 experiments
- Firstly, data stored in XML files. Recently, moved to MySQL database.
- Web-based searching: http://hendrix.imm.dtu.dk/services/brededatabase/
- Recording published neuroimaging experiments that list stereotaxic coordinates in so-called MNI or Talairach space *(Talairach and Tournoux, 1988) - ”Co-planar Stereotaxic Atlas of the Human Brain”*

Coordinate-based searching in Brede DB

Brede Database - brede_loc_query.pl - Query for locations

File Edit View Go Debug Desktop Window Help

Location: http://hendrix.imm.dtu.dk/cgi-bin/brede_loc_query.pl?q=45+-35

Brede Database – Talairach coordinate search

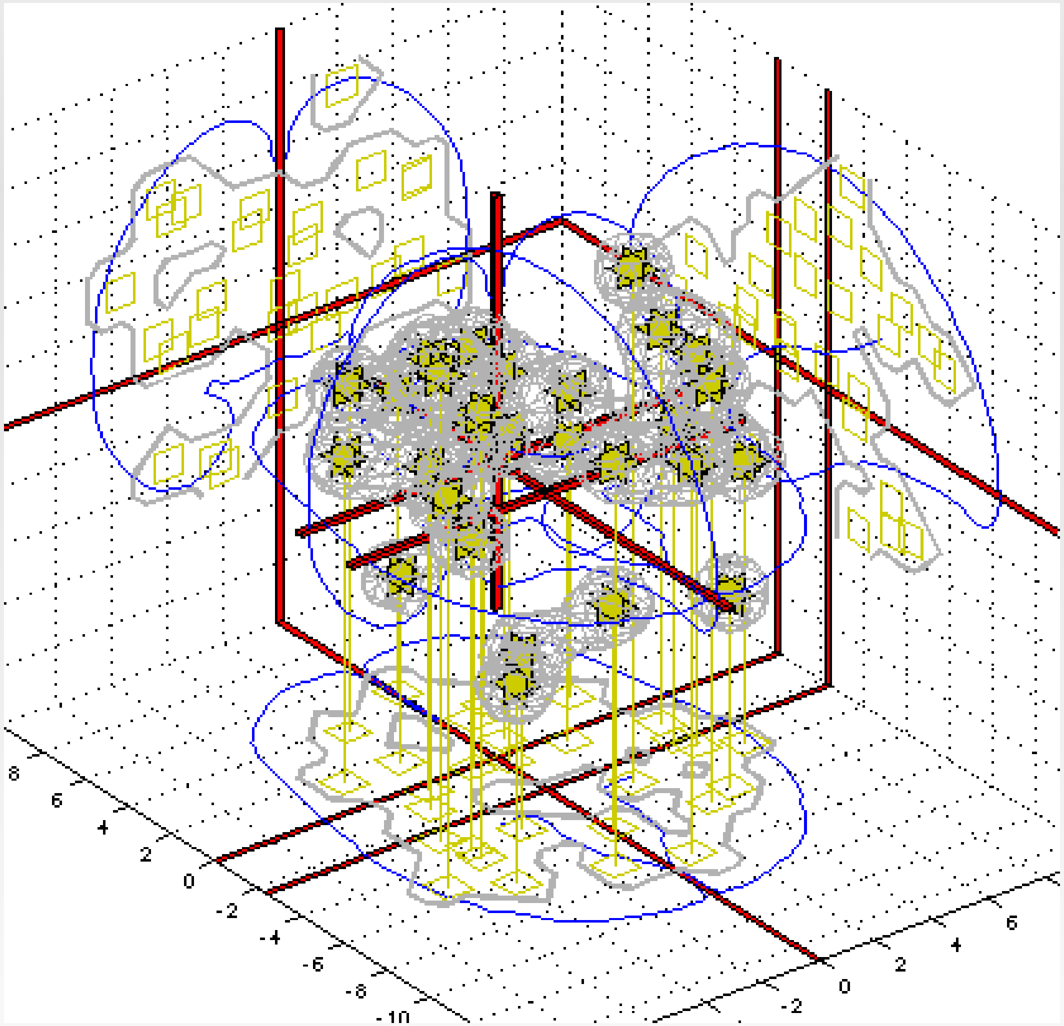
brede_loc_query — Search after locations (Talairach coordinates) in the Brede Database

45 -35 11 [Location search (one coordinate)] e.g., 14 -9 -15

45 -35 11 [Experiment search (several coordinates)]

#	Distance	x	y	z	WOBIB	Description
1	4.4	48	-36	8	130	Superior temporal gyrus — Tics during Tourette's syndrome (WOEXP: 402)
2	6.1	47	-40	9	128	Right temporoparietal junction — Visuoproprioceptive conflict (WOEXP: 393)
3	6.6	48	-40	8	177	Middle and posterior temporal — Happiness from films and recall (WOEXP: 540)
4	7.1	49	-40	13	91	Right superior temporal — Alzheimer's disease versus healthy (WOEXP: 291)
5	7.2	45	-31	17	39	— Unpleasant words (WOEXP: 132)
6	8.5	43	-26	10	64	Right Heschl's gyrus — Listening to voices (WOEXP: 199)
7	8.6	52	-38	7	168	Right superior temporal sulcus — Threat-related words in controls versus panic disorder patients (WOEXP: 515)
8	8.7	52	-37	7	88	Right middle temporal gyrus — Activation in sadness film viewing versus neutral film viewing (WOEXP: 282)
9	9.0	50	-30	16	59	Superior temporal gyrus — Spatial neglect (WOEXP: 185)

Database entry visualizations



An fMRI experiment resulting in 29 reported coordinates

Brede Database offers:

- location search (distance between coordinates)

- 'experimental' search (similarity between two sets of coordinates / volumes)

(Nielsen and Hansen, 2004) - "Finding related functional neuroimaging volumes"

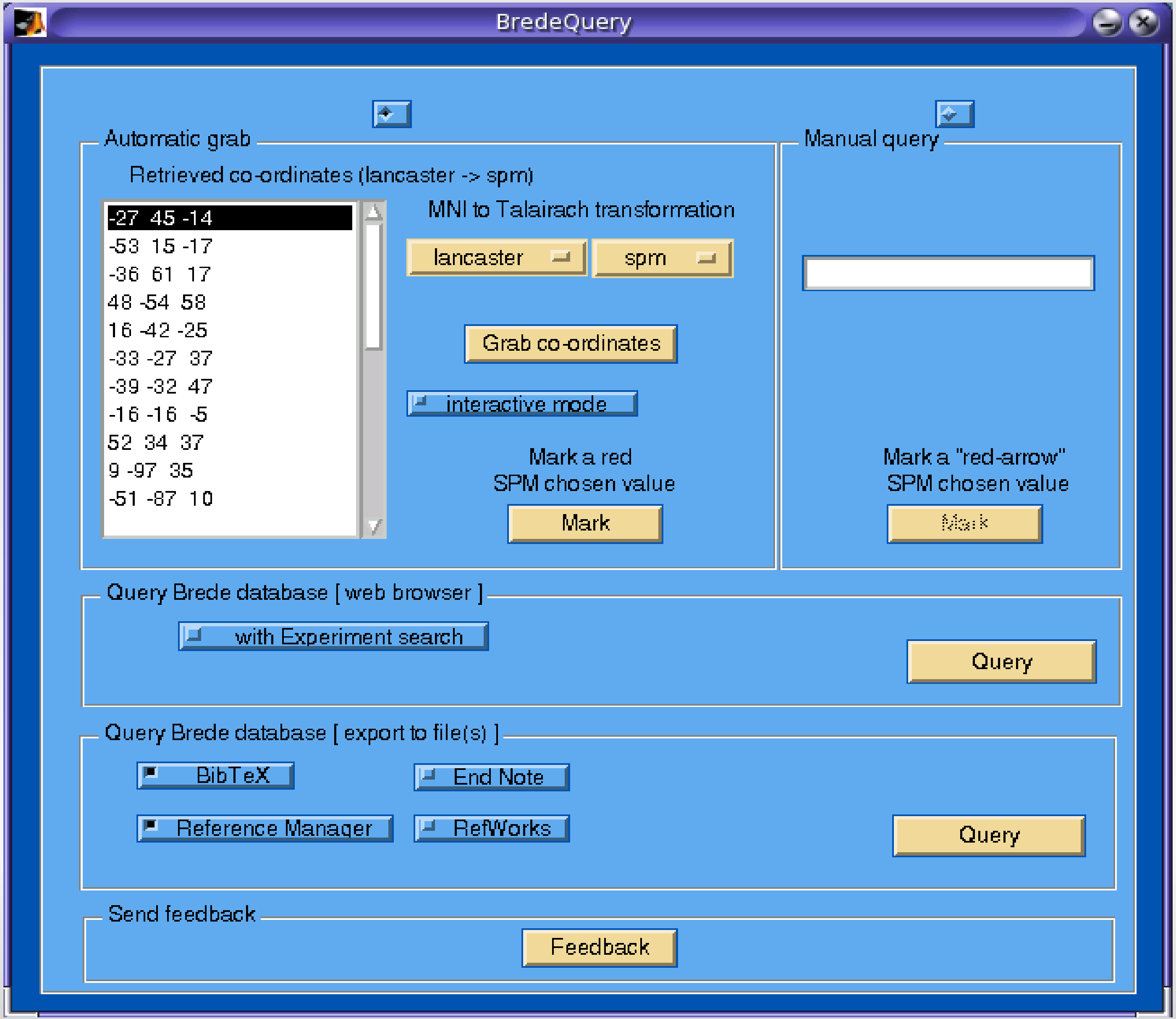
Statistical Parametric Mapping (SPM)

- ”Statistical Parametric Mapping refers to the **construction** and **assessment** of **spatially extended statistical processes** used to **test hypotheses** about **functional imaging data**. These ideas have been instantiated in software that is called SPM.”
- ”The SPM software package has been designed for the **analysis of brain imaging data sequences**. The sequences can be a series of images from different cohorts, or time-series from the same subject. The current release is designed for the analysis of fMRI, PET, SPECT, EEG and MEG.”

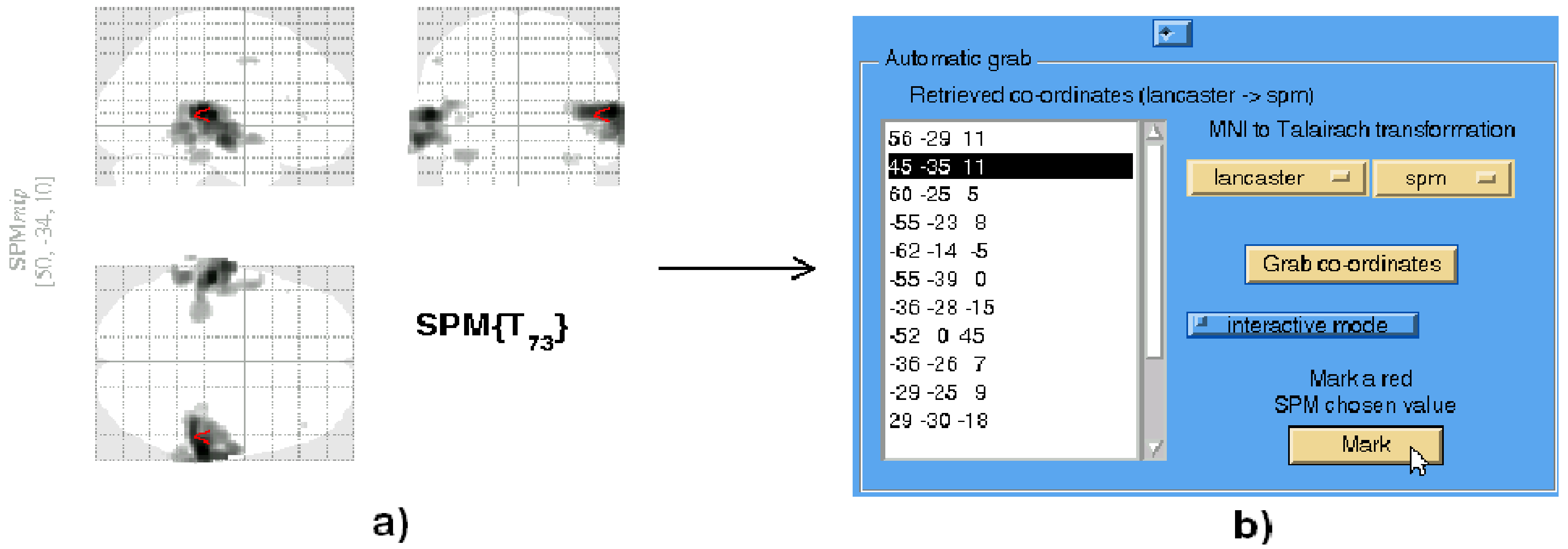
Taken from: http://www.fil.ion.ucl.ac.uk/spm/

BredeQuery plugin for SPM

http://neuroinf.imm.dtu.dk/BredeQuery/



Brain coordinates grabbing



The coordinates of the most significant activations in brain, found during an SPM analysis, are:

1. grabbed by the BredeQuery plugin,
2. transformed using any of MNI to Talairach transformations,
3. prepared for a coordinate-based searching with Brede Database;

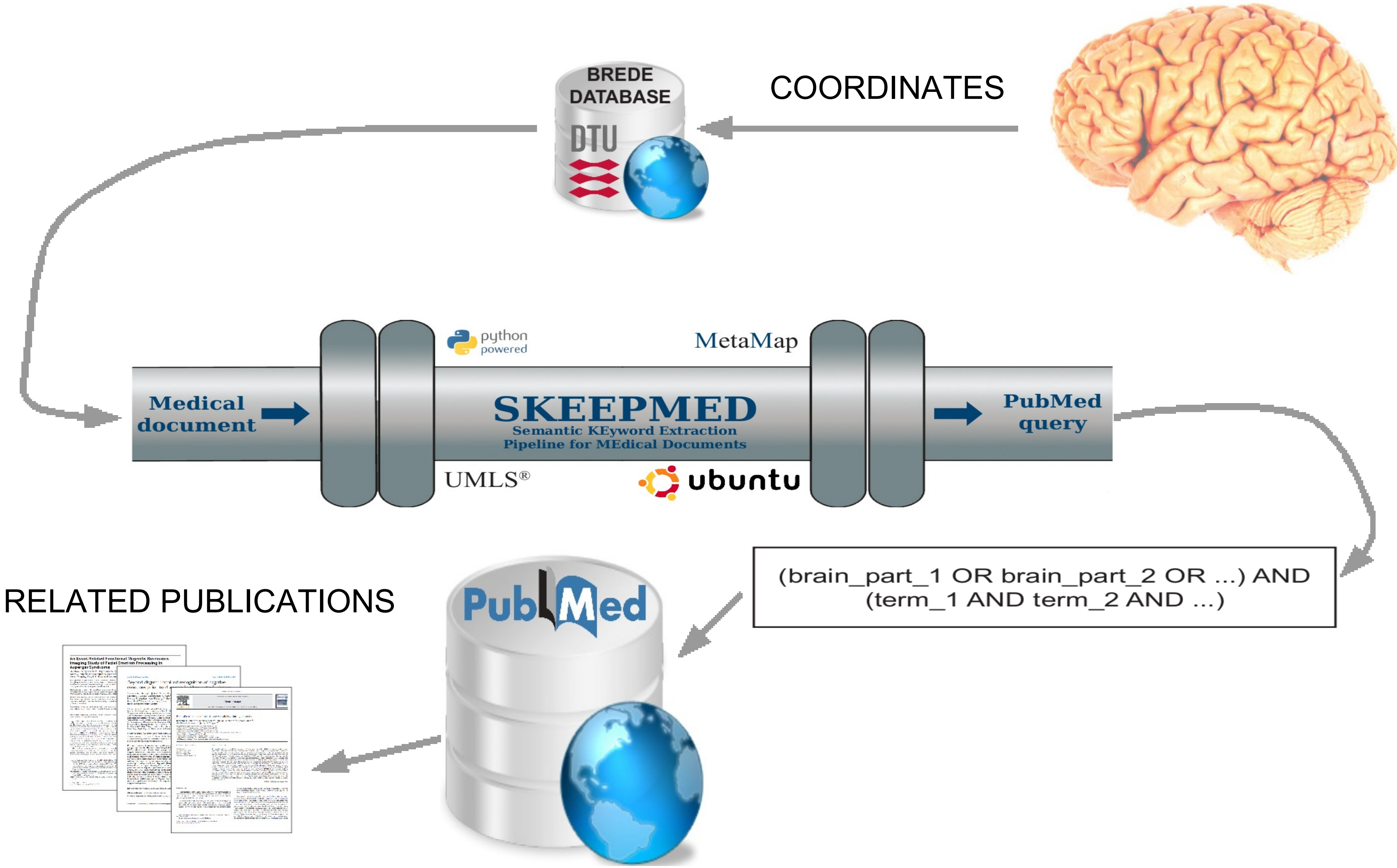
MNI-to-Talairach transformations

- **brett** - Piece-wise affine transformation by Matthew Brett *(Brett, 1999) - ”The MNI brain and the Talairach atlas.”*
- **lancaster** – affine transformation by Jack Lancaster et al. *(Lancaster et al., 2007) - ”Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template.”*
  - **SPM**
  - **FSL**
  - **POOLED** (combined)

Roadmap

- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

SKEEPMED


BREDE DATABASE
DTU
COORDINATES
python powered
MetaMap
Medical document
SKEEPMED
Semantic KEyword Extraction Pipeline for MEdical Documents
PubMed query
UMLS®
ubuntu
(brain_part_1 OR brain_part_2 OR ...) AND (term_1 AND term_2 AND ...)
RELATED PUBLICATIONS
PubMed

Architecture

- Load text (abstract, article):
  - skeepmed_input_xml = open(xml_file_path,'r')
- Run MetaMap:
  - metamap_file_exec_path = '/usr/local/bin/metamap08'
  - parameters = '-% format *abstract.txt metamap_out_file.xml*'
  - metamap_log = subprocess.Popen([metamap_file_exec_path, parameters],stdout=subprocess.PIPE).communicate()[0]
- Parse MetaMap XML and getListOfKeywords():
  - Check all Mappings and their Candidates, select those with sufficient NegScore, count frequency of each keyword occurence, store in a dictionary (keyword:freq)
- Create query, ask PubMed

Keywords

- Two types of keywords:
  - brain_parts
  - terms
- Brain_parts retrieval settings:
  - Only Neuronames Brain Hierarchy data source used
  - Threshold low
- Terms retrieval settings:
  - All data sources used
  - Threshold high = 1000 (max) (*only best matches*)
  - Minimum occurence frequency > 1

PubMed's query

(brain_part_1 OR brain_part_2 OR ...) AND
(term_1 AND term_2 AND ...)

Keyword extraction test

#	1. PubMed Article	2. Year	3. Position in Brede Database search
1	Neural correlates of heart rate variability during emotion (Lane RD et al.)	2009	#1 (70%)
2	Beyond disgust: impaired recognition of negative emotions prior to diagnosis in Huntington's disease (Johnson SA et al.)	2007	no coordinates reported
3	Disgust and happiness recognition correlate with anteroventral insula and amygdala volume respectively in preclinical Huntington's disease (Kipps CM et al.)	2007	#3 (20%)
4	An event related functional magnetic resonance imaging study of facial emotion processing in Asperger syndrome (Deeley Q et al.)	2007	–
5	Neurophysiological correlates of induced discrete emotions in humans: an individually oriented analysis (Aftanas LI et al.)	2006	no coordinates reported
6	Neurophysiological correlates of induced discrete emotions in humans: an individual analysis (Aftanas LI et al.)	2004	no coordinates reported
7	Functional neuroanatomy of emotions: a meta-analysis (Murphy FC et al.)	2003	no coordinates reported
8	Common and distinct neural responses during direct and incidental processing of multiple facial emotions (Winston JS et al.)	2003	#9 (20%)
9	A preferential increase in the extrastriate response to signals of danger (Surguladze SA et al.)	2003	#1 (10%)
10	Impaired facial emotion recognition in early-onset right mesial temporal lobe epilepsy (Meletti S et al.)	2003	no coordinates reported
11	Age-related differences in brain activation during emotional face processing (Gunning-Dixon FM et al.)	2003	–
12	An fMRI study of facial emotion processing in patients with schizophrenia (Gur RE et al.)	2002	#2 (60%)
13	Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI (Phan KL et al.)	2002	no coordinates reported
14	Deficits in recognition of emotional facial expression are still present in alcoholics after mid- to long-term abstinence (Kornreich C et al.)	2001	no coordinates reported
15	Activation of anterior paralimbic structures during guilt-related script-driven imagery (Shin LM et al.)	2000	#7 (50%)
16	Perception of emotion in frontotemporal dementia and Alzheimer disease (Lavenu I et al.)	1999	no coordinates reported

Table 1: Results of spatial closeness comparison between experiments from PubMed retrieved articles and "source". Column 3. shows the position of the best-matched "source's" experiment in the results list returned by Brede Database when querying a test article experiment coordinates. In the parentheses the percentage of matched "source's" experiments found in top 20 Brede Database results is shown.

Test coordinate: (-8,1,9) – thalamus brain region

Brede Database best match:
"Neuroanatomical Correlates of Happiness, Sadness, and Disgust"
by Richard D. Lane et al. (1997)

Keywords:

brain_part: *cerebral cortex, thalamus, insula, frontal lobe*

term: *disgust, sadness, happiness, emotion*

Roadmap

- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

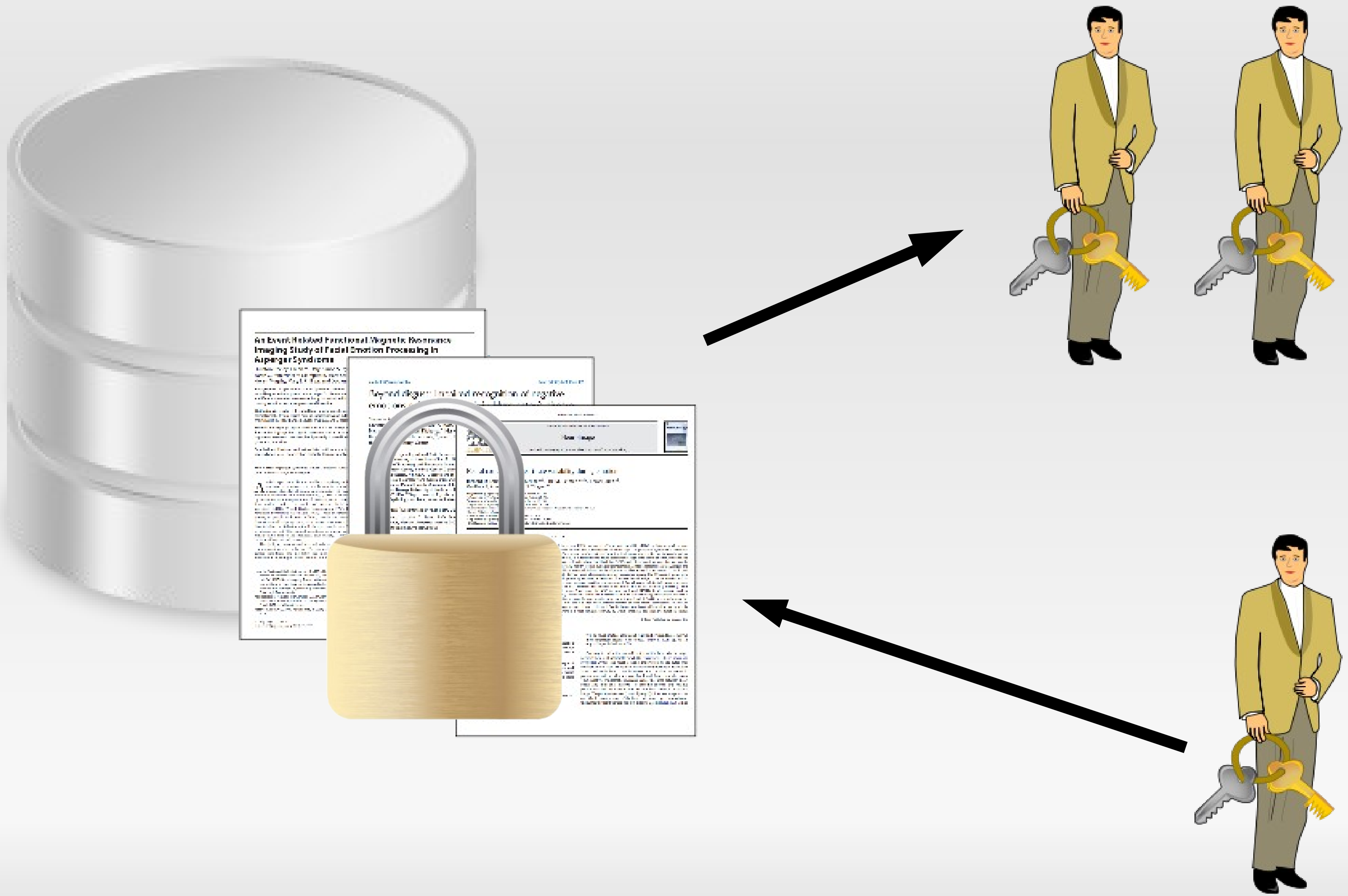
Functionality evaluation

- How well works our recent pipeline?
- Need for automatic evaluation of the results – how? (current consultations with professor Ingemar Cox)
- Find the best Metamap parameters settings (data sources, semantic types, thresholds) – employment of metaheuristics?
- Combine data mining, machine learning, statistical methods (LSA, NMF, etc.) with ontological mapping?

Metaheuristics

- Thousands of parameters: threshold value (0..1000), 135 Semantic Types, 148 UMLS Sources → $2^{10} \cdot 2^{135} \cdot 2^{148} = 2^{293}$
- Metaheuristics used for finding the best parameters' setting (very stable results)
- Algorithm type: tuned simulated annealing
- 3 random articles for tuning, 3 random articles for testing
- Evaluation (golden set – 20 papers from PubMed)

Secure portal for neuroscientists



Secure portal for neuroscientists

- Integrated toolkit for encrypted communication
- Mixture of symmetric and asymmetric cryptography protocols to securely exchange information within virtual groups and public
- Version control
- Ability to securely exchange documents, coordinates
- Peer review system
- Ability to easily publish given work

Hopes for the future of MetaMap

- Unicode support
- Native 64-bit platform
- Ability to query for semantic types
- Ability to query for UMLS sources

Hopes for the future of MetaMap

- Both stand alone application and service oriented
- Ability to extract UMLS mapping hierarchy
  - parent, child
  - siblings, synonyms
- Open Python API

Roadmap

- Motivations and project overview
- Coordinate-based searching (BredeDatabase & BredeQuery plugin for SPM)
- Semantic KEyword Extraction Pipeline for MEdical Documents (SKEEPMED)
- Future directions, bottlenecks, problems
  - Validation and evaluation
  - Machine learning & ontologies (hybrid approach)
  - Metaheuristics for finding the best MetaMap parameters setting
- Conclusions

Thank you for your attention!

Questions?

Bartłomiej Wilkowski - bw@imm.dtu.dk