Bridging the gap between coordinate- and keyword- based search of neuroscientific databases by UMLS-assisted semantic keyword extraction

Bartłomiej Wilkowski, Marcin Marek Szewczyk, Lars Kai Hansen
Center for Integrated Molecular Brain Imaging
Technical University of Denmark, DTU Informatics
bw@imm.dtu.dk, msz@imm.dtu.dk, lkh@imm.dtu.dk

Introduction

The rapid growth of the neuroimaging literature brings the demand for integration, organization and dissemination among a highly interdisciplinary community of researchers, see e.g. (Wager, Lindquist & Kaplan 2007). Since functional localization in brain is normally represented in form of stereotaxic coordinates, it can be used directly in the process of retrieving related literature in a given functional context by the measure of coordinate distance. Current neuroimaging databases which provide coordinate-based search capabilities (Brede Database, BrainMap) contain relatively small number of publications (Szewczyk 2008), therefore an interconnection with more comprehensive bibliographical databases can extend the results pool. Recently, the BredeQuery plug-in for SPM pipeline was presented as a tool which enables coordinate-based querying of the Brede Database directly from SPM (Wilkowski, Szewczyk, Rasmussen, Hansen & Nielsen 2009). As an extension of the current plugin's functionality, we propose methods for integration of the Brede Database with the almost complete medical publication database - PubMed (`http://pubmed.org`).

Methods

The first step towards the integration of PubMed with the BredeQuery plug-in is efficient keyword extraction from abstracts returned by the Brede Database (Nielsen 2003) after coordinate-based searching. The extracted keywords can be later used for modelling PubMed's query. Keywords are concatenated using OR, AND logical operators. We are developing Semantic KEyword Extraction Pipeline for Medical Documents (SKEEPMED) web service for mapping terms from abstracts to the UMLS Metathesaurus concepts using the MetaMap08 program (Aronson 2001). As we focus on the neuroscientific literature, we extract two types of keywords: brain parts (**brain_part**) and other significant domain terms (**term**). The final query is constructed with the following structure: *(**brain_part_1** OR **brain_part_2** OR ...) AND (**term_1** AND **term_2** AND ...).*

Results

We queried the Brede Database with a test coordinate in Talairach (Talairach & Tournoux 1988) space (-8,1,9), which relates to the thalamus brain region. The highest ranked experiment returned by the database belongs to article "Neuroanatomical Correlates of Happiness, Sadness, and Disgust" by Richard D. Lane et al. (1997), in the following referred to as the "source". "Source" contains description of 10 experiments with a total of 90 reported coordinates. The following keywords were extracted by SKEEPMED: **brain_part** keywords: *cerebral cortex, thalamus, insula, frontal lobe* ; **term** keywords: *disgust, sadness, happiness, emotion.* The PubMed query based on these keywords returned the "source" and 20 additional articles closely related to the topic of "source". We inspected only articles published later than "source" (16), 8 of which do not contain any experiment coordinates. To investigate the relevance of the remaining 8 articles, we compared spatial closeness of experiment coordinates from these 8 articles with the "source" experiments by querying Brede Database. Experiment coordinates from 6 of them are located in a close neighbourhood of "source's" experiments. Results are presented in Table 1.

#	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.

Conclusions

Current neuroimaging databases are limited and we have discussed a new way of enhancing their usability. We use the highly refined information in the Brede database to form an informed query into the literature at large. The case story showed the viability of the approach and gives us confidence that coordinate based search can be combined with language processing for a productivity enhancing tool for all neuroimagers. Current work concerns quantitative testing of the method.

References

Aronson, A. (2001), 'Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program', *JOURNAL OF BIOMEDICAL INFORMATICS* **35**, 17–21.

Nielsen, F. Å. (2003), The Brede database: a small database for functional neuroimaging, *in* 'NeuroImage', Vol. 19, Elsevier. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19-22, 2003, New York, NY.

Szewczyk, M. M. (2008), Databases for neuroscience, Master's thesis, Technical University of Denmark.

Talairach, J. & Tournoux, P. (1988), *Co-planar Stereotaxic Atlas of the Human Brain*, Thieme Medical Publisher Inc, New York.

Wager, T. D., Lindquist, M. & Kaplan, L. (2007), 'Meta-analysis of functional neuroimaging data: current and future directions', *Social Cognitive and Affective Neuroscience* **2**(2), 150–158.

Wilkowski, B., Szewczyk, M. M., Rasmussen, P. M., Hansen, L. K. & Nielsen, F. Å. (2009), Coordinate-based meta-analytic search for the SPM neuroimaging pipeline - The BredeQuery plugin for SPM5, *in* 'HEALTHINF 2009'.