Clustering cliques for graph-based summarization of the biomedical research literature

Background: Graph-based notions are increasingly used in biomedical data mining and knowledge discovery tasks. In this paper, we present a clique-clustering method to automatically summarize graphs of semantic predications produced from PubMed citations (titles and abstracts). Results: SemRep is used to extract semantic predications from the citations returned by a PubMed search. Cliques were identified from frequently occurring predications with highly connected arguments filtered by degree centrality. Themes contained in the summary were identified with a hierarchical clustering algorithm based on common arguments shared among cliques. The validity of the clusters in the summaries produced was compared to the Silhouette-generated baseline for cohesion, separation and overall validity. The theme labels were also compared to a reference standard produced with major MeSH headings. Conclusions: For 11 topics in the testing data set, the overall validity of clusters from the system summary was 10% better than the baseline (43% versus 33%). While compared to the reference standard from MeSH headings, the results for recall, precision and F-score were 0.64, 0.65, and 0.65 respectively.

Publication information
Journal: BMC Bioinformatics
Volume: 14
Article number: 182
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.54 SJR 1.467 SNIP 0.946
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.656 SNIP 1.077 CiteScore 2.77
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.836 SNIP 1.202 CiteScore 2.91
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.932 SNIP 1.335 CiteScore 3.38
ISI indexed (2013): ISI indexed yes
MEDLINE MeSH Indexing: Lessons Learned from Machine Learning and Future Directions

Due to the large yearly growth of MEDLINE, MeSH indexing is becoming a more difficult task for a relatively small group of highly qualified indexing staff at the US National Library of Medicine (NLM). The Medical Text Indexer (MTI) is a support tool for assisting indexers; this tool relies on MetaMap and a k-NN approach called PubMed Related Citations (PRC). Our motivation is to improve the quality of MTI based on machine learning. Typical machine learning approaches fit this indexing task into text categorization. In this work, we have studied some Medical Subject Headings (MeSH) recommended by MTI and analyzed the issues when using standard machine learning algorithms. We show that in some cases machine learning can improve the annotations already recommended by MTI, that machine learning based on low variance methods achieves better performance and that each MeSH heading presents a different behavior. In addition, there are several factors which make this task difficult (e.g. limited access to the full-text of the citations) which provide direction for future work.
A bottom-up approach to MEDLINE indexing recommendations

MEDLINE indexing performed by the US National Library of Medicine staff describes the essence of a biomedical publication in about 14 Medical Subject Headings (MeSH). Since 2002, this task is assisted by the Medical Text Indexer (MTI) program. We present a bottom-up approach to MEDLINE indexing in which the abstract is searched for indicators for a specific MeSH recommendation in a two-step process. Supervised machine learning combined with triage rules improves sensitivity of recommendations while keeping the number of recommended terms relatively small. Improvement in recommendations observed in this work warrants further exploration of this approach to MTI recommendations on a larger set of MeSH headings.

General information

State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling, US National Institute of Health
Authors: Jimeno-Yepes, A. (Ekstern), Wilkowski, B. (Intern), Mork, J. (Ekstern), Van Lenten, E. (Ekstern), Demner-Fushman, D. (Ekstern), Aronson, A. (Ekstern)
Publication date: 2011

Host publication information

Title of host publication: AMIA 2011 Proceedings
Main Research Area: Technical/natural sciences
Links:
http://www.amia.org/amia2011
Source: orbit
Source-ID: 275905
Publication: Research - peer-review › Article in proceedings – Annual report year: 2011
Graph-Based Methods for Discovery Browsing with Semantic Predications

We present an extension to literature-based discovery that goes beyond making discoveries to a principled way of navigating through selected aspects of some biomedical domain. The method is a type of "discovery browsing" that guides the user through the research literature on a specified phenomenon. Poorly understood relationships may be explored through novel points of view, and potentially interesting relationships need not be known ahead of time. In a process of "cooperative reciprocity" the user iteratively focuses system output, thus controlling the large number of relationships often generated in literature-based discovery systems. The underlying technology exploits SemRep semantic predications represented as a graph of interconnected nodes (predication arguments) and edges (predicates). The system suggests paths in this graph, which represent chains of relationships. The methodology is illustrated with depressive disorder and focuses on the interaction of inflammation, circadian phenomena, and the neurotransmitter norepinephrine. Insight provided may contribute to enhanced understanding of the pathophysiology, treatment, and prevention of this disorder.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling, National Library of Medicine, Institute for Biostatistics and Medical Informatics
Authors: Wilkowski, B. (Intern), Fiszman, M. (Ekstern), Miller, C. M. (Ekstern), Hristovski, D. (Ekstern), Arabandi, S. (Ekstern), Rosemblat, G. (Ekstern), Rindflesch, T. C. (Ekstern)
Pages: 1514-1523
Publication date: 2011
Conference: AMIA Annual Symposium, Washington, USA, 01/01/2011
Main Research Area: Technical/natural sciences

Publication information
Journal: AMIA Annual Meeting. Proceedings
Volume: 2011
ISSN (Print): 0146-146X
Original language: English
Electronic versions:
D26AAAd01.pdf
Links:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243228/
http://www.amia.org/amia2011

Bibliographical note
This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose.
Source: orbit
Source-ID: 316649
Publication: Research - peer-review › Conference article – Annual report year: 2011

Semantic Approaches for Knowledge Discovery and Retrieval In Biomedicine

This thesis discusses potential applications of semantics to the recent literature-based informatics systems to facilitate knowledge discovery, hypothesis generation, and literature retrieval in the domain of biomedicine. The approaches presented herein make use of semantic information extracted from biomedical texts by natural language processing systems supported by biomedical ontologies. The thesis is divided into two main parts: First, a field of literature-based discovery is introduced, with a review of recent approaches of the field; second, literature retrieval in the domain of neuroimaging (neuroscience) is discussed with the emphasis put on the coordinate-based searching of related publications. My own contribution to the first part is a novel literature-based ‘discovery browsing’ methodology incorporating semantic predications, graph theory and path analysis for guiding researchers through the relevant literature on a user-specied biomedical phenomenon. Moreover, the additional analyses of the methodology show its potential application as a support for the recent probabilistic retrieval methods. In the second part of the thesis, I present the BredeQuery plugin which integrates a coordinate-based literature retrieval system with the common in neuroimaging statistical analysis environment. It is followed by the detailed description of a prototype of context-dependent neuroscientific literature retrieval methodology, which thanks to the employment of ontologies, allows the user to define context of interest for a search. The peer reviewed research articles, included in the appendices, discuss further the details of the presented methods, case studies, and provide other related information.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Wilkowski, B. (Intern), Hansen, L. K. (Intern)
Number of pages: 133  
Publication date: 2011

**Publication information**  
Place of publication: Kgs. Lyngby, Denmark  
Publisher: Technical University of Denmark (DTU)  
Original language: English

Series: IMM-PHD  
Number: 261  
ISSN: 0909-3192  
Main Research Area: Technical/natural sciences  
Electronic versions:  
phd261_bw.pdf  
Source: orbit  
Source-ID: 277868  
Publication: Research › Ph.D. thesis – Annual report year: 2011

---

**BredeQuery: Coordinate-Based Meta-analytic Search of Neuroscientific Literature from the SPM Environment**  
Large amounts of neuroimaging studies are collected and have changed our view on human brain function. By integrating multiple studies in meta-analysis a more complete picture is emerging. Brain locations are usually reported as coordinates with reference to a specific brain atlas, thus some of the databases offer so-called coordinate-based searching to the users (e.g. Brede, BrainMap). For such search, the publications, which relate to the brain locations represented by the user coordinates, are retrieved. We present BredeQuery – a plugin for the widely used SPM data analytic pipeline. BredeQuery offers a direct link from SPM to the Brede Database coordinate-based search engine. BredeQuery is able to ‘grab’ brain location coordinates from the SPM windows and enter them as a query for the Brede Database. Moreover, results of the query can be displayed in a MATLAB window and/or exported directly to some popular bibliographic file formats (BibTeX, Reference Manager, etc).

**General information**  
State: Published  
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling  
Authors: Wilkowski, B. (Intern), Szewczyk, M. M. (Intern), Rasmussen, P. M. (Intern), Hansen, L. K. (Intern), Nielsen, F. Å. (Intern)  
Pages: 314-324  
Publication date: 2010

**Host publication information**  
Title of host publication: Biomedical Engineering Systems and Technologies : Communications in Computer and Information Science  
Volume: 52  
Place of publication: Berlin  
Publisher: Springer  
ISBN (Print): 3-642-11720-1  
Main Research Area: Technical/natural sciences  
Conference: International Joint Conference, BIOSTEC : HEALTHINF 2009, Porto, Portugal, 01/01/2009  
coordinate based database, literature retrieval  
Links:  
http://www.springer.com/computer/database+management+%26+information+retrieval/book/978-3-642-11720-6  
Source: orbit  
Source-ID: 275874  
Publication: Research - peer-review › Article in proceedings – Annual report year: 2010

**Bridging the gap between coordinate- and keyword- based search of neuroscientific databases by UMLS-assisted semantic keyword extraction**  
General information  
State: Published  
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling  
Authors: Wilkowski, B. (Intern), Szewczyk, M. M. (Intern), Hansen, L. K. (Intern)  
Publication date: 2009

**Host publication information**  
Title of host publication: NeuroImage : Supplement 1  
Volume: 47
COORDINATE-BASED META-ANALYTIC SEARCH FOR THE SPM NEUROIMAGING PIPELINE: The BredeQuery plugin for SPM5

Large amounts of neuroimaging studies are collected and have changed our view on human brain function. By integrating multiple studies in meta-analysis a more complete picture is emerging. Brain locations are usually reported as coordinates with reference to a specific brain atlas, thus some of the databases offer so-called coordinate-based searching to the users (e.g., Brede, BrainMap). For such search, the publications, which relate to the brain locations represented by the user coordinates, are retrieved. In this paper we present BredeQuery – a plugin for the widely used SPM5 data analytic pipeline. BredeQuery offers a direct link from SPM5 to the Brede Database coordinate-based search engine. BredeQuery is able to ‘grab’ brain location coordinates from the SPM windows and enter them as a query for the Brede Database. Moreover, results of the query can be displayed in an SPM window and/or exported directly to some popular bibliographic file formats (BibTeX, Reference Manager, etc.).

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Wilkowski, B. (Intern), Szewczyk, M. (Intern), Rasmussen, P. M. (Intern), Hansen, L. K. (Intern), Nielsen, F. Å. (Intern)
Pages: 11-17
Publication date: 2009

Host publication information
Title of host publication: HEALTHINF 2009: Proceedings of the Second International Conference on Health Informatics
Place of publication: Portugal
Publisher: Institute for Systems and Technologies of Information, Control and Communication
ISBN (Print): 98-98-11163-1
Main Research Area: Technical/natural sciences
SPM, fMRI, PET, neuroinformatics, Talairach, MNI, brain region, brain function, brain activity, article retrieval, coordinate-based search, bibtex, Brede Database, ontology
Electronic versions:
BredeQueryPaper.pdf
Links:
http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=5706
Source: orbit
Source-ID: 227998
Publication: Research - peer-review › Article in proceedings – Annual report year: 2009

Knowledge Discovery in Neuroinformatics

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling
Authors: Wilkowski, B. (Intern)
Number of pages: 589
Publication date: 2009

Host publication information
Title of host publication: MIE
Main Research Area: Technical/natural sciences
Conference: Medical Informatics in a United and Healthy Europe 2009, Sarajevo, Bosnia and Herzegovina, 01/01/2009
Neuroinformatics, Data sharing, Medical ontology
Source: orbit
Source-ID: 240054
Publication: Research - peer-review › Article in proceedings – Annual report year: 2009
Neuroscientific literature search based on the location coordinates in brain - BredeQuery plugin for SPM environment

General information
State: Published
Organisations: Cognitive Systems
Authors: Wilkowski, B. (Intern)
Publication date: 2009

Host publication information
Title of host publication: Frontiers in Neuroinformatics
Main Research Area: Technical/natural sciences
Conference: 2nd INCF Congress of Neuroinformatics, Pilsen, Czech Republic, 06/09/2009 - 06/09/2009
DOIs:
10.3389/conf.neuro.11.2009.08.071
Links:
http://frontiersin.org/conferences/individual_abstract_listing.php?conferid=155&pap=2104&ind_abs=1
Source: orbit
Source-ID: 255632
Publication: Research - peer-review › Conference abstract in proceedings – Annual report year: 2009

MIAWARE Software: 3D Medical Image Analysis with Automated Reporting Engine and Ontology-based Search
This article presents MIAWARE, a software for Medical Image Analysis With Automated Reporting Engine, which was designed and developed for doctor/radiologist assistance. It allows to analyze an image stack from computed axial tomography scan of lungs (thorax) and, at the same time, to mark all pathologies on images and report their characteristics. The reporting process is normalized - radiologists cannot describe pathological changes with their own words, but can only use some terms from a specific vocabulary set provided by the software. Consequently, a normalized radiological report is automatically generated. Furthermore, MIAWARE software is accompanied with an intelligent search engine for medical reports, based on the relations between parts of the lungs. A logical structure of the lungs is introduced to the search algorithm through the specially developed ontology. As a result, a deductive report search was obtained, which may be helpful for doctors while diagnosing patients’ cases. Finally, the MIAWARE software can be considered also as a teaching tool for future radiologists and physicians.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling, Aveiro University, Centro Hospitalar do Baixo Alentejo - Hospital Jose Joaquim Fernandes de Beja, Technical University of Lodz
Authors: Wilkowski, B. (Intern), Pereira, O. N. M. (Ekstern), Dias, P. (Ekstern), Castro, M. (Ekstern), Janicki, M. (Ekstern)
Pages: 201-206
Publication date: 2008

Host publication information
Title of host publication: HEALTHINF 2008 : Proceedings of the First International Conference of Health Informatics
Volume: Volume 2
Place of publication: Portugal
Publisher: Institute for Systems and Technologies of Information, Control and Communication
ISBN (Print): 978-989-8111-16-6
Main Research Area: Technical/natural sciences
Conference: International Joint Conference on Biomedical Engineering Systems and Technologies : International Conference on Health Informatics, Funchal - Madeira - Portugal, 01/01/2008
Image visualization, Computed axial tomography, Radiological report, Ontology
Electronic versions:
151_MIAWARE_SOFTWARE.pdf
Links:
http://www.healthinf.org
Source: orbit
Source-ID: 211154
Publication: Research - peer-review › Article in proceedings – Annual report year: 2008

Projects:
Knowledge Discovery in Neuroinformatics

Department of Informatics and Mathematical Modeling
Period: 01/05/2008 → 30/11/2011
Number of participants: 5
Phd Student:
Wilkowski, Bartlomiej (Intern)
Main Supervisor:
Hansen, Lars Kai (Intern)
Examiner:
Larsen, Jan (Ekstern)
Andreasen, Troels (Ekstern)
Honkela, Timo (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: 1/3 DTU-stip, 2/3 FUR/andet
Project: PhD