USES OF MESH AND MESH INDEXING IN BIOMEDICAL INFORMATICS RESEARCH

Kate Masterton, Associate Fellow 2013-14

Project Sponsors

Caitlin Sticco, Bibliographic Services Division

John Kilbourne, MeSH Section

CONTENTS

ABSTRACT

Background: The MeSH thesaurus is used within NLM primarily to provide subject indexing for articles in MEDLINE. This project seeks to explore how MeSH indexing and the MeSH vocabulary are being used beyond PubMed in informatics and computational research. These types of uses are likely growing, and we are interested in improving services to this user group, as well as understanding the scope and impact of these kinds of works.

Objective: Identify and characterize uses of MeSH and MeSH indexing beyond literature retrieval in PubMed and collect information about the experiences of researchers using these tools for novel applications.

Methods: Research projects published within the past five years were identified through literature searches in IEEE, PubMed, and Scopus. In addition to literature review, a formal interview process was conducted in order to collect experiences from researchers who use MeSH and/or MeSH indexing for bioinformatics applications. Over 50 candidates for interview were identified through published literature and discussions with experts in the field.

Results: The literature retrieval set (583 citations) was screened and citations that met relevancy criteria were further analyzed. Each relevant citation (225) was assigned one or more categories of research based on how MeSH and/or MeSH indexing was utilized. In addition, broader areas of application were identified. Interview data uncovered high-level themes regarding the value of MeSH and MeSH indexing to bioinformatics researchers, as well as varied challenges experienced by this user group.

Conclusions: Uses of MeSH and MeSH indexing in novel applications outside of PubMed are significant, diverse, and widespread. Specific needs of this user group are often dependent on individual research applications. This project has identified many ideas for further consideration.

INTRODUCTION

The Medical Subject Headings (MeSH) were created by the National Library of Medicine (NLM) in 1960 to serve as the controlled vocabulary for MEDLINE, and also for subject cataloging of the NLM collection. Indexers at NLM use MeSH to provide article level indexing in MEDLINE (Libscomb, 2000). Therefore, the primary use of MeSH at NLM has been to aid literature retrieval. While searchers of PubMed are certainly the largest user group of MeSH and MeSH indexing, this project and report focus on a different user group - biomedical informatics researchers.

The field of biomedical informatics can be described as the development and application of computational tools and methods for expanding the use of biological data (Makalowski 2003). Biomedical informatics researchers use controlled vocabularies to organize and leverage biomedical data. This project was developed to discover how biomedical informatics researchers use MeSH and MeSH indexing in their work and to what extent.

The goal of this project was twofold: learn about uses of MeSH and MeSH indexing in biomedical informatics research that go beyond the originally intended purpose, and develop an understanding of the needs of the user group that uses MeSH and MeSH indexing in biomedical informatics research.

Therefore, this project tries to address questions such as:

- What are the types of uses for MeSH and MeSH indexing beyond retrieving the literature in PubMed?
- What is the scope and impact of these types of works?
- What is the size of this user group?
- Does this user group have any unmet needs for advancing their research?

The project is exploratory and represents the first attempt NLM has made to identify and characterize uses of MeSH and MeSH indexing, flagship NLM products, in biomedical informatics research. The findings of this project will potentially inform indexing policy and MeSH development, and have uncovered further areas of investigation for future attention.

METHODS

The project had two primary components – a review of published literature and interviews with biomedical researchers using MeSH or MeSH indexing for purposes beyond literature retrieval.

LITERATURE REVIEW

The literature review included three sources – PubMed, Scopus, and IEEE. For each, a search string (Appendix) was created to find research that used MeSH or MeSH indexing beyond literature retrieval; for example, in some type of computation application or statistical analysis. The search included literature published within the past five years.

Citations from each resource were exported to EndNote Web for automatic de-duplication, and then exported to Microsoft Excel for further analysis. The original retrieval set included 583 citations after de-duplication. Two reviewers analyzed the 583 citations. Despite the use of search strings to find only research that used MeSH or MeSH indexing beyond retrieval, the original retrieval set included publications that were not relevant for the project; for example, systematic reviews using MeSH indexing to retrieve literature on a specific topic. All publications that did not use MeSH or MeSH indexing beyond retrieval purposes were discarded. After the initial screening for relevancy, 225 citations were included for further analysis. Considering that this only represents publications from three sources published within the past five years, the number of relevant projects exceeded initial expectations.

The two reviewers then examined the 225 relevant citations. Using research categories developed during the review process, reviewers assigned one or more research categories to each citation. Research categories described how MeSH or MeSH indexing was used in the research. Reviewers also took notes on broader areas of research application (for example, clinical support, translational science, etc.), which pieces of MeSH or MeSH indexing was used in the research, and any other resources (ie: OMIM, UMLS, Gene, etc.) used for the research. This detailed information is recorded in the citation analysis spreadsheet (Supplement 1).

INTERVIEWS

The citation analysis spreadsheet provides useful information about the uses of MeSH and MeSH indexing for purposes beyond literature retrieval. While this information is valuable, the publications provide variable levels of detail about why researchers chose MeSH and MeSH indexing for their research, how exactly they were used, and if researchers are satisfied with MeSH and MeSH indexing for research purposes. To obtain this type of information, NLM received approval from the Office of Management and Budget to interview biomedical informatics researchers about their uses of MeSH and MeSH indexing beyond literature retrieval (entire application for approval is available in supplemental files, see Masterton_MeSH and MeSH indexing files.zip). Six open ended interview questions (Appendix) were developed to gain further insight about user experience with MeSH and MeSH indexing for non-retrieval purposes. For the interviews, 62 researchers were contacted via email and ultimately 25 researchers were interviewed for approximately 30 minutes each. The researchers were identified via published literature, and a contact directory with notes was created to manage potential interviewees. The researchers identified and contacted came from all over the world; slightly over half were outside of the US (Figure 1). The resulting transcripts and notes are available as supplemental files (see Masterton_MeSH and MeSH indexing files.zip).

Country	No. of researchers contacted
USA	28
UK	4
Canada	4
Netherlands	3
Japan	3
Spain	3
France	3
Germany	2
India	2
China	2
Portugal	2
Romania	1
Israel	1
Switzerland	1
Brazil	1
Turkey	1
Italy	1

Figure 1: Number of researchers contacted by country. It was interesting to discover that MeSH and MeSH indexing are being used in biomedical informatics research across the globe.

RESULTS

LITERATURE REVIEW

The breakdown of the categories of research is illustrated in Figure 2. This represents the total number of citations assigned to each category. Each citation could have one or more categories. For many of the categories, more granular subcategories were also created.

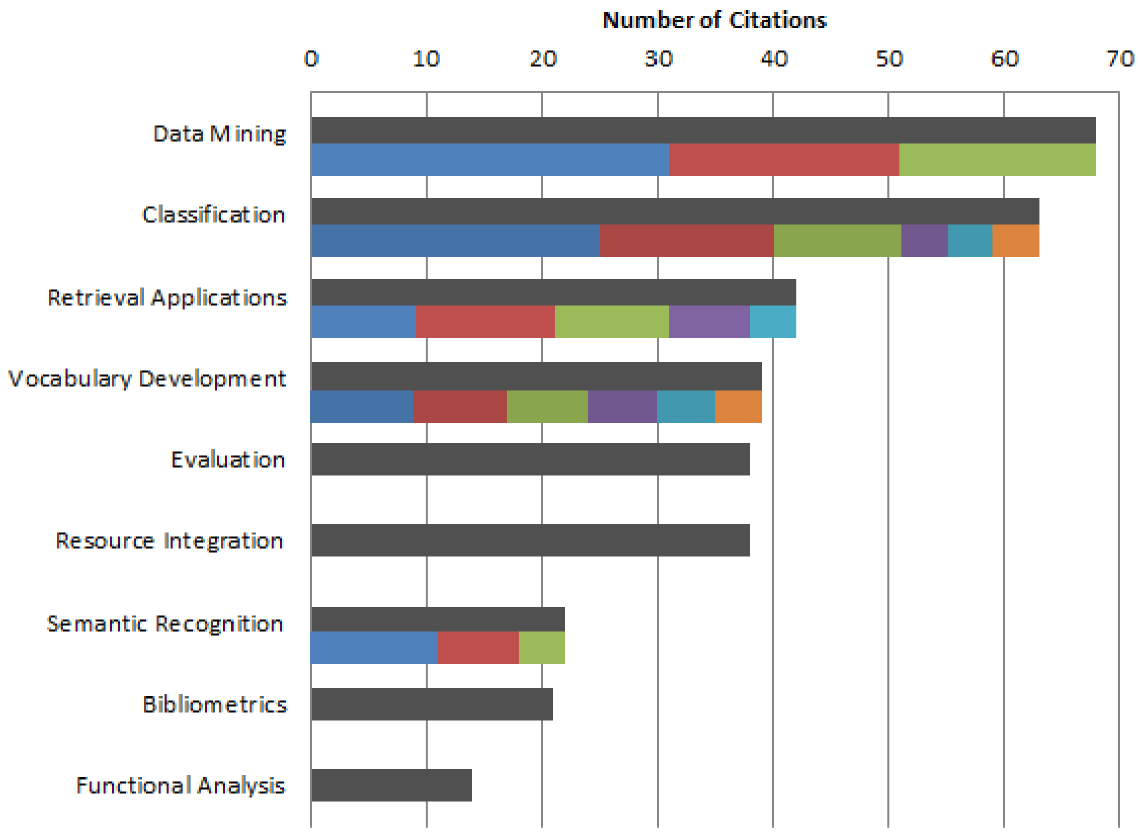


Figure 2: total number of citations assigned to each research category. Subcategories represented where applicable (shown in colored bars – for a breakdown of all subcategories see the category descriptions beginning on page 8). Each of the 225 citations reviewed could have one or more research categories. The total number of assigned categories was 347, so each of the 225 citations had an average of 1.5 research categories.

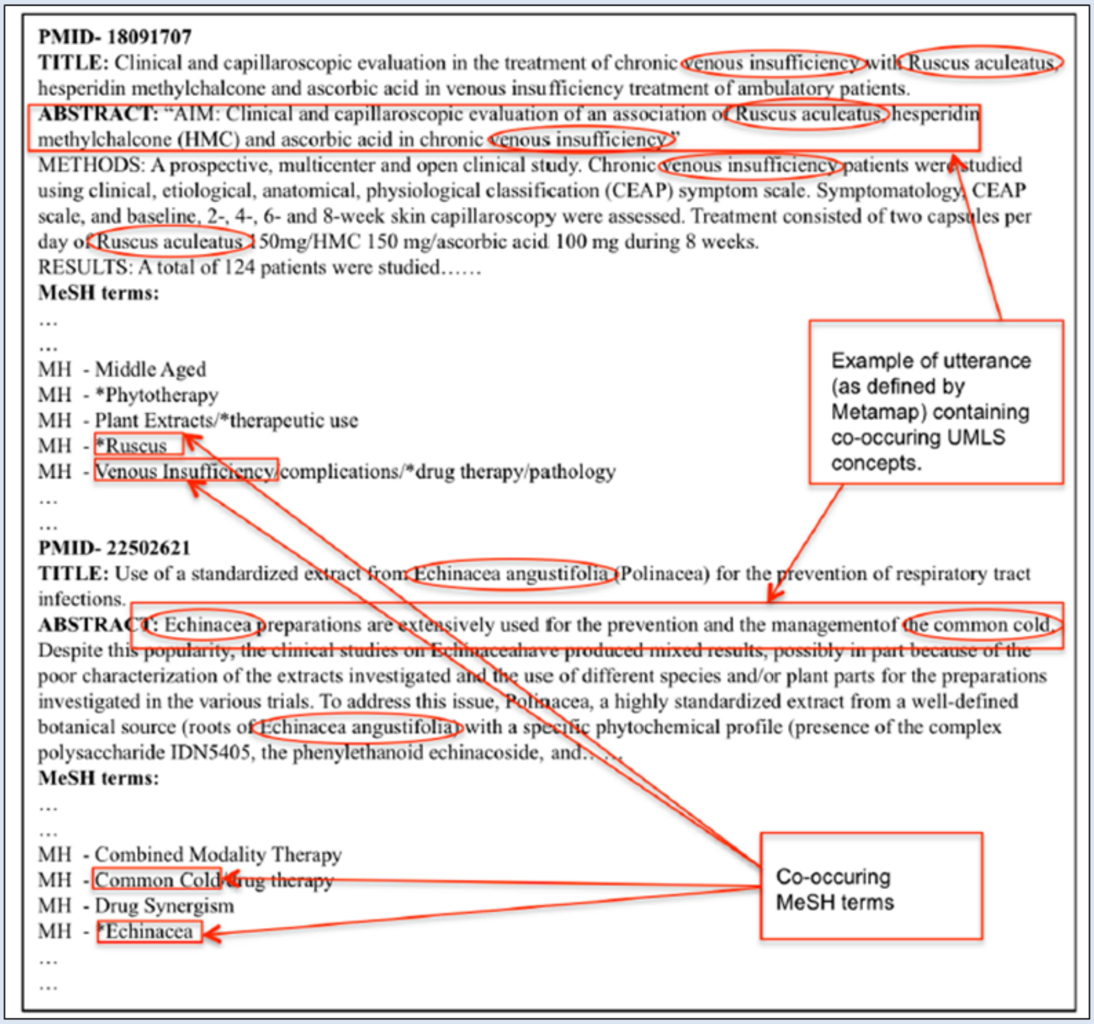
DEFINITIONS OF RESEARCH CATEGORIES

Data Mining (68 Citations): Research using sophisticated analysis tools to sort through, organize, examine, and combine large sets of information. Contains the subcategories:

> **Relationship Extraction (31 Citations)**: Extracts semantic triples and other meaningful standard representations of information and relationships described explicitly in free text or sets of information.
>
> **Literature Based Discovery (20 Citations)**: Searches for hidden, implicit connections among information embedded specifically in published literature.
>
> **Graph Analysis (17 Citations)**: Generates and studies graphs that represent relationship networks between entities to identify new knowledge.

Sharma V, Sarkar IN. Leveraging concept-based approaches to identify potential phyto-therapies. J Biomed Inform 2013;46(4):602-14.

This project uses MeSH indexing in combination with UMLS concepts extracted from free text for retrieval and ranking of both direct (concept relationships found from direct correlations) and inferred (concept relationships found from shared signs and symptoms) associations between plants and human diseases.



PMID- 18091707
TITLE: Clinical and capillaroscopic evaluation in the treatment of chronic venous insufficiency with Ruscus aculeatus, hesperidin methylchalcone and ascorbic acid in venous insufficiency treatment of ambulatory patients.
ABSTRACT: "AIM: Clinical and capillaroscopic evaluation of an association of Ruscus aculeatus, hesperidin methylchalcone (HMC) and ascorbic acid in chronic venous insufficiency."
METHODS: A prospective, multicenter and open clinical study. Chronic venous insufficiency patients were studied using clinical, etiological, anatomical, physiological classification (CEAP) symptom scale. Symptomatology, CEAP scale, and baseline, 2-, 4-, 6- and 8-week skin capillaroscopy were assessed. Treatment consisted of two capsules per day of Ruscus aculeatus 150mg/HMC 150 mg/ascorbic acid 100 mg during 8 weeks.
RESULTS: A total of 124 patients were studied……
MeSH terms:
…
…
MH - Middle Aged
MH - *Phytotherapy
MH - Plant Extracts/*therapeutic use
MH - *Ruscus
MH - Venous Insufficiency/complications/*drug therapy/pathology
…
…
PMID- 22502621
TITLE: Use of a standardized extract from Echinacea angustifolia (Polinacea) for the prevention of respiratory tract infections.
ABSTRACT: Echinacea preparations are extensively used for the prevention and the managementof the common cold. Despite this popularity, the clinical studies on Echinaceahave produced mixed results, possibly in part because of the poor characterization of the extracts investigated and the use of different species and/or plant parts for the preparations investigated in the various trials. To address this issue, Polinacea, a highly standardized extract from a well-defined botanical source (roots of Echinacea angustifolia) with a specific phytochemical profile (presence of the complex polysaccharide IDN5405, the phenylethanoid echinacoside, and……
MeSH terms:
…
…
MH - Combined Modality Therapy
MH - Common Cold/drug therapy
MH - Drug Synergism
MH - *Echinacea
…
…

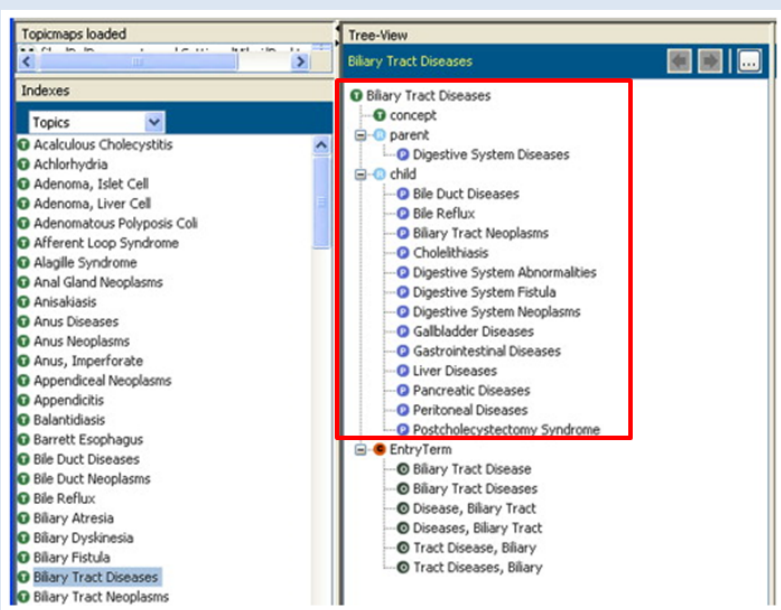
Figure 3: Example of a Data Mining project using MeSH and MeSH indexing (http://www.ncbi.nlm.nih.gov/pubmed/23665360)

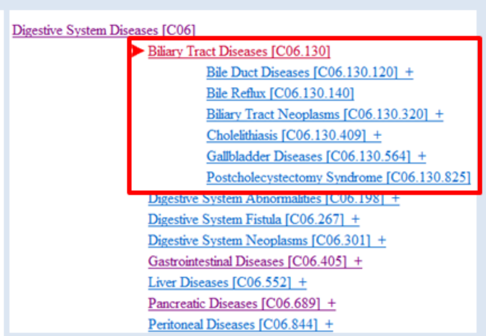
<u>Classification (63 Citations)</u>: Research to systematically arrange or retrieve entities in categories based on common characteristics such as properties, morphology, subject matter, etc. Contains the subcategories:

> **<u>Other (25 Citations)</u>**: Covers a wide array of other types of classified entities not covered in the other subcategories. Some examples include classification of documents based on security level, classification of clinical trials by specialty, etc.
>
> **<u>Automated Literature Indexing (15 Citations)</u>**: Attempts to characterize (or index) published literature (MEDLINE) as NLM does, but with automated methods.
>
> **<u>Datasets (11 Citations)</u>**: Attempts to characterize datasets
>
> **<u>Images (4 Citations)</u>**: Attempts to characterize image data
>
> **<u>Multi-Terminology Indexing (4 Citations)</u>**: Researches the application of indexing using more than one terminology within a given resource
>
> **<u>Websites (4 Citations)</u>**: Attempts to characterize websites

Burdescu D, Mihai C, Stanescu L, Brezovan M. Automatic image annotation and semantic based image retrieval for medical domain. Neurocomputing 2013;109(3): 33-48.

This research attempts to automatically annotated gastrointestinal images. The classification task here is not text based, instead they use image shapes and other visual features to generate automatic annotations by training with similar, pre-labeled examples. The researchers used MeSH as the vocabulary for annotation, but instead of using the hierarchical tree structure they manually created an ontology out of the portion of MeSH they needed.





Left figure shows MeSH as an ontology that was used by Burdescu et al. (2013), in comparison to the MeSH hierarchical tree structure (left).

Figure 4: Example of a Classification project using MeSH and MeSH indexing (http://www.sciencedirect.com/science/article/pii/S0925231212006698)

Retrieval Applications (42 Citations): Research that produces applications that enhance retrieval in various systems. Contains the subcategories:

Interface (12 Citations): Presents and evaluates search interfaces incorporating MeSH and MeSH indexing.

Query Expansion (10 Citations): Presents and evaluates applications to expand original user queries using MeSH or MeSH indexing.

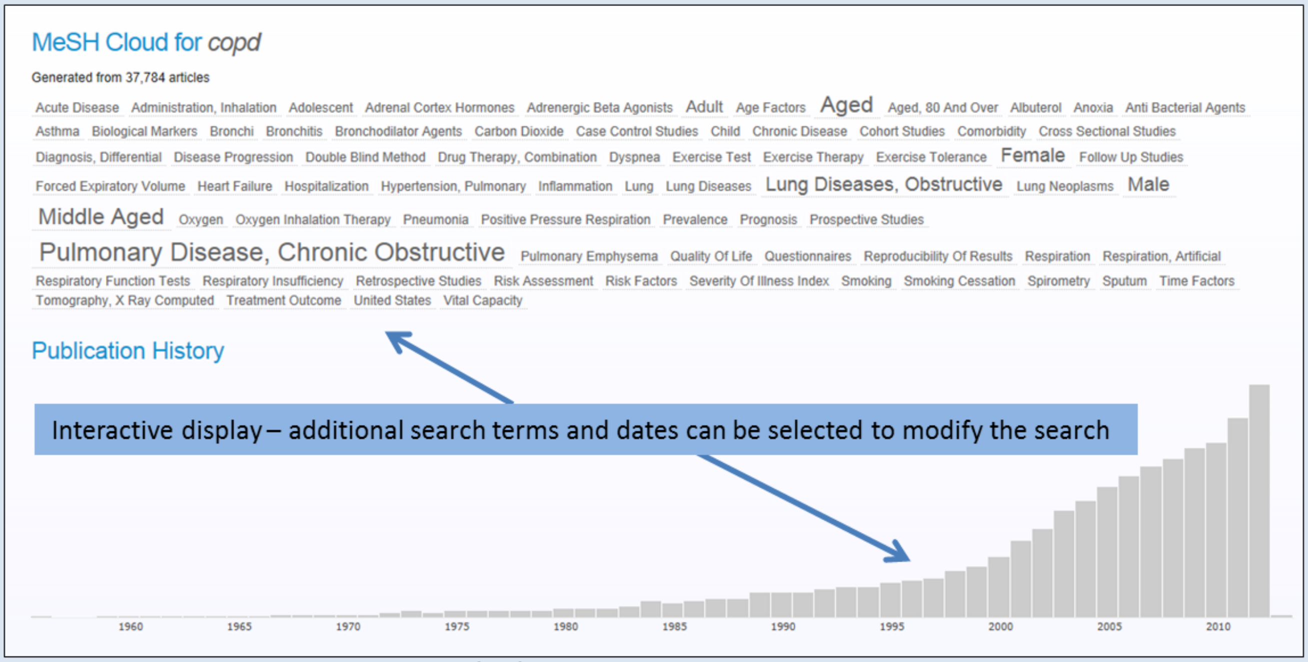
Filter (9 Citations): Presents and evaluates applications to filter retrieval results to a given criteria.

Rank (7 Citations): Presents and evaluates applications to rank retrieval results, usually by relevancy to a given query.

Semantic Search (4 Citations): Applies the meaning of language to improve retrieval accuracy.

Sarkar IN, Schenk R, Miller H, Norton CN. LigerCat: using "MeSH Clouds" from journal, article, or gene citations to facilitate the identification of relevant biomedical literature. AMIA Annu Symp Proc 2009: 563-567.

LigerCat creates tag clouds for queries based on significant MeSH terms extracted from the articles returned by a Medline search. The user can manipulate the search by selecting terms in the tag cloud. LigerCat is still maintained and open access (http://ligercat.ubio.org/articles).



Result of LigerCat search for COPD on 8/26/2014

Figure 5: Example of a Retrieval Application project using MeSH and MeSH indexing (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815376/)

Vocabulary Development (39 Citations): Research to develop, integrate, or evaluate vocabularies or ontologies. Contains the subcategories:

Vocabulary Mapping (9 Citations): Maps or aligns one vocabulary to another

Indexing Analysis (8 Citations): Evaluates the application of the MeSH vocabulary in indexed published literature for a particular topic

Automated Vocabulary Expansion (7 Citations): Automatically extends vocabularies

Comparative Analysis (6 Citations): Compares two or more existing vocabularies

Vocabulary Evaluation (5 Citations): Evaluates MeSH as a vocabulary (usually with a specific purpose for evaluation criteria)

Other (4 Citations): Research focused on other types of vocabulary development

Kahn CE. Annotation of figures from the biomedical imaging literature: a comparative analysis of RadLex and other standardized vocabularies. Acad Radiol 2014;21(3): 384-92.

This example shows a comparative analysis of multiple vocabularies for use in annotating radiology images. The analysis used the National Center for Biomedical Ontology Annotator (NCBO Annotator) to assign terms to images . The terms were derived automatically from text in the article title and image caption. Six biomedical ontologies were compared, including MeSH. Evaluation included the number of annotations per image and the number of annotations per term to show overall coverage for the field of radiology. This project asserts that RadLex fills a gap in medical terminologies.

TABLE 1. Comparison of Six Ontologies for Annotation of the ARRS GoldMiner Corpus of Figure Captions Using the Most Recent Version of Each Ontology at the NCBO BioPortal Site

				Annotated Terms		Annotated Figures		Annotations		
Ontology	Version	Release Date	No. of Terms	Number	Percent	Number	Percent	Number	Per Term	Per Figure
FMA	3.1	March 3, 2010	83,281	5398	6.5	324,376	84.2	1,288,568	15.5	3.3
ICD-10-CM	2011_01	January 1, 2011	91,590	1635	1.8	84,987	22.1	104,095	1.1	0.3
LOINC	236	June 1, 2011	171,399	7683	4.5	380,834	98.9	5,008,536	29.2	13.0
MeSH	2012	September 9, 2011	229,698	15,792	6.9	381,978	99.2	3,097,452	13.5	8.0
RadLex	3.8	February 19, 2013	39,218	8504	21.7	380,338	98.8	3,871,573	98.7	10.1
SNOMED CT	2011_07_31	July 31, 2011	395,036	41,371	10.5	384,492	99.9	11,588,578	29.3	30.1
Total			1,010,222	80,383	8.0	385,018		24,958,802	24.7	64.8

FMA, Foundational Model of Anatomy; ICD-10-CM, International Classification of Diseases, Version 10, Clinical Modification; LOINC, Logical Observation Identifier Names and Codes; MeSH, Medical Subject Headings; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

The Annotated Terms column shows the number of terms from each ontology that appeared in the annotations. The Annotated Figures column shows the number of figures captions from the collection that were annotated.

Figure 6: Example of a Vocabulary Development project using MeSH and MeSH indexing (http://www.ncbi.nlm.nih.gov/pubmed/24507425)

Evaluation (38 Citations): Applies to research that uses MeSH and/or MeSH indexing as a baseline for evaluation.

Ruau D, Mbaqwu M, Dudley JT, Krishnan V, Butte AJ. Comparison of automated and human assignment of MeSH terms on publicly-available molecular datasets. J Biomed Inform 2011;44(Suppl 1)

The increasing availability of publically available molecular datasets can be used to advance scientific research, but discoverability relies on descriptive annotations of such datasets, which the majority of publicly available datasets lack. This research attempts to present partial solutions to this problem. The National Center for Biomedical Ontology Annotator (NCBO Annotator) and the NLM MetaMap programs were applied to free text associated with datasets (from the PRIDE data repository) in order to generate automated MeSH annotations. These automated annotations are compared to manually-assigned MeSH terms already existing in PRIDE. To provide a gold standard for comparison, the MEDLINE indexing was extracted from papers associated with PRIDE datasets.

Overall, automated methods achieved higher recall (number of relevant annotations retrieved) of MeSH annotations than human dataset submitters, but the human dataset submitters achieved higher precision (fraction of retrieved annotations that are relevant). The authors conclude that human dataset submitters are using too few terms to characterize their datasets. An indication is that perhaps automated concept identification programs could assist human dataset submitters to select appropriate annotations for molecular datasets.

Figure 7: Example of a project using MeSH and MeSH indexing for evaluation purposes (http://www.ncbi.nlm.nih.gov/pubmed/21420508)

<u>Resource Integration (38 Citations)</u>: These projects link together disparate data resources, in order to find and describe relationships between many different types of information.

Ruau D, Dudley JT, Chen R, Phillips NG, Swan GE, et al. Integrative approach to pain genetics identifies pain sensitivity loci across diseases. PLoS Comput Biol 2012;8(6)

The process of identifying human genes relevant to pain requires difficult-to-conduct and expensive clinical trials. This research presents a method for discovering pain gene candidates for further investigation that takes advantage of freely available data (disease-related clinical literature and gene expression microarray data). They use MeSH annotation co-occurrences (using the MeSH term "Pain" and disease headings) to build a disease-specific pain index representing the relative painfulness of a disease compared to others. Then they mined data from GEO and ArrayExpress and extracted differentially expressed genes associated with the specific diseases in the pain index. Finally, they organize these genes according to the disease-specific pain index in order to determine significant associations between gene expression patterns and pain.

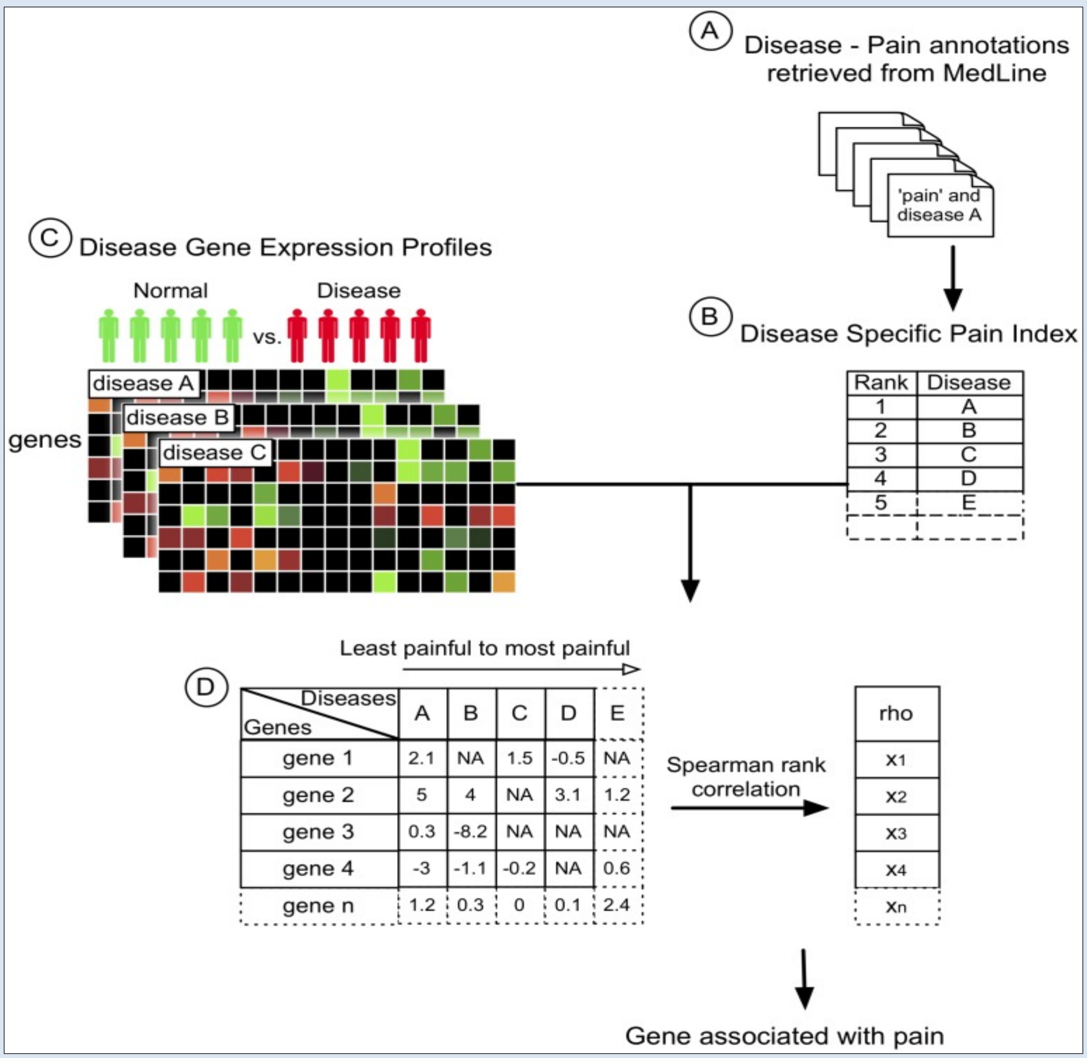


Figure 8: Example of a Resource Integration project using MeSH and MeSH indexing (http://www.ncbi.nlm.nih.gov/pubmed/22685391)

<u>Semantic Recognition Tasks (22 Citations)</u>: Tasks that try to understand some type of meaning in unstructured text. Contains the subcategories:

> **<u>Semantic Similarity (11 Citations)</u>**: Research that attempts to identify the closeness of meaning between sets of documents or terms
>
> **<u>Word Sense Disambiguation (7 Citations)</u>**: These projects try to determine which meaning of a word applies in a specific instance.
>
> **<u>Named Entity Recognition (4 Citations)</u>**: Attempts to identify and classify mentions of named things (such as specific diseases, genes, treatments, etc.) within unstructured text.

Sánchez, D., Solé-Ribalta, A., Batet, M., Serratosa, F. Enabling semantic similarity estimation across multiple ontologies: an evaluation in the biomedical domain. J Biomed Inform 2012; 45(1)

Semantic similarity estimation between a pair of terms contributes to the better understanding of textual resources, and ontology-based methods have been shown to be effective. However, comparing terms from different ontologies is challenging because of varying structures, granularity, etc. This research presents two approaches for semantic similarity estimation of terms from heterogeneous ontologies. The two approaches are evaluated using MeSH and WordNet (which is a general purpose vocabulary).

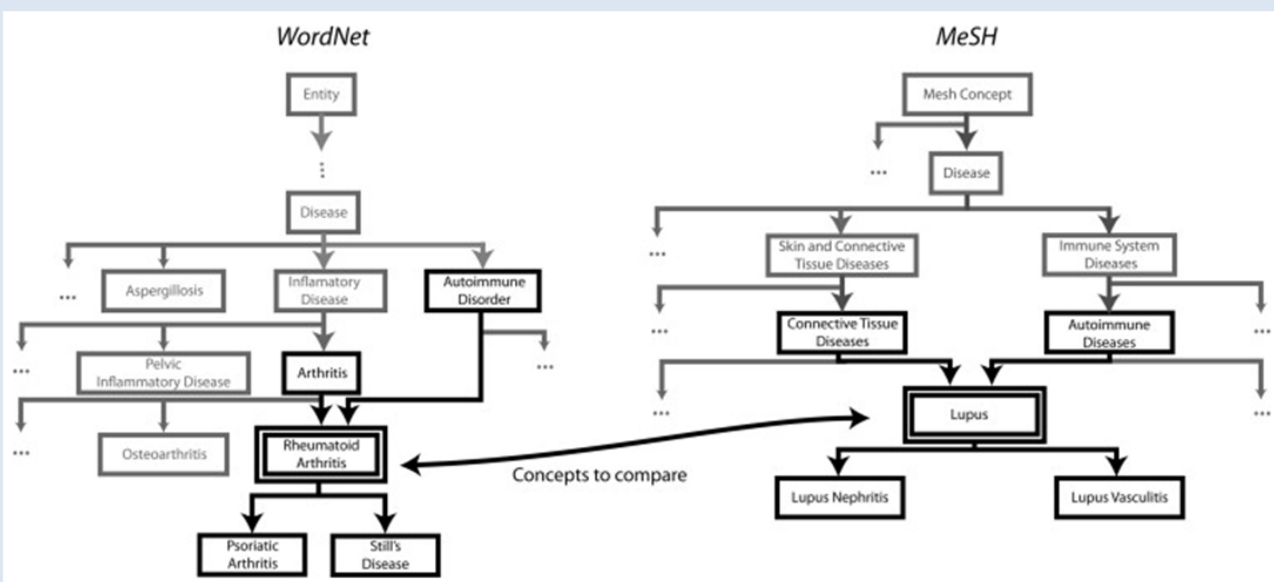
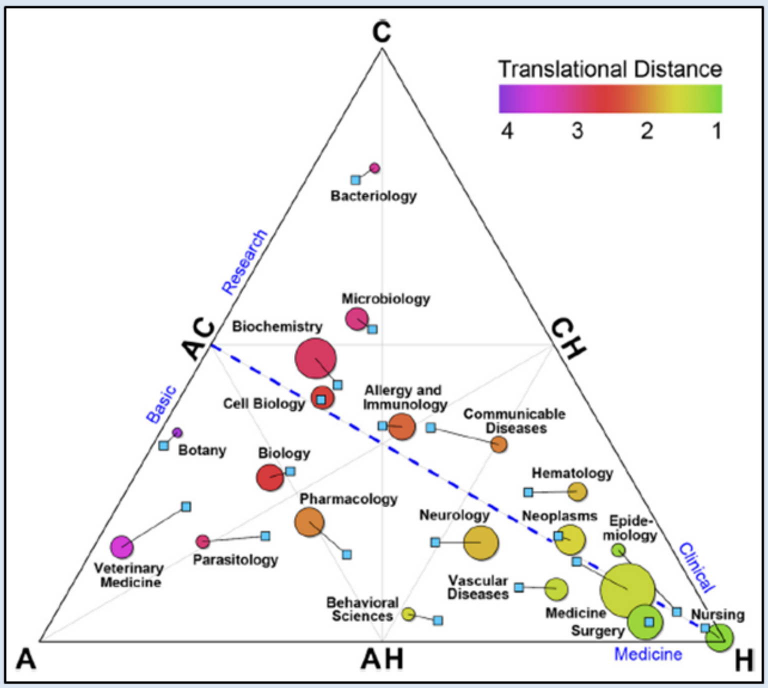


Figure 9: Example of a Semantic Recognition project using MeSH and MeSH indexing (<u>http://www.ncbi.nlm.nih.gov/pubmed/22056693</u>)

Bibliometrics (21 Citations): Research using statistical methods to analyze a body of literature to reveal the historical development of subject fields and patterns of authorship, publication, and use. Note that for the current project bibliometrics research using MeSH and/or MeSH indexing for retrieval only is discarded. The research must include some type of computational or statistical application of the MeSH and/or MeSH indexing for inclusion.

Weber, G. Identifying translational science within the triangle of biomedicine. J Transl Med 2013; 11(126)

This project used MEDLINE indexing and the MeSH hierarchy to determine whether articles in PubMed contained cell, animal, or human research. The project developed a simple categorization schema based on these types of research (cell, animal, human) in order to visualize research in a "triangle of biomedicine." In addition to visualization, the project also attempted to show research paths towards translation. This type of work could be useful to policy makers in evaluating the impact of funding provided to advance the pace of translation.



Disciplines mapped onto the Triangle of Biomedicine. The corners of the triangle correspond to animal (A), cellular or molecular (C), and human (H) research. The dashed blue line indicates the Translational Axis from basic research to clinical medicine.

Figure 10: Example of a Bibliometrics project using MeSH and MeSH indexing (http://www.translational-medicine.com/content/11/1/126)

Functional Analysis (14 Citations): These projects apply MeSH annotations to genomics or proteomics datasets to interpret the themes of the dataset.

Umezawa M, Tainaka H, Kawashima N, Shimizu M, Takeda K. Effect of fetal exposure to titanium dioxide nanoparticle on brain development – brain region information. J Toxicol Sci 2012;37(6):1247-52.

Researchers wanted to identify which regions of the brain had downregulated genes when prenatally exposed to titanium dioxide (TiO_2). To perform the analysis, they mapped 87 MeSH terms associated with brain regions to 2,037 genes on a microarray test by using the gene reference database PubGene (now CoreMine). The analysis provides data to support the hypothesis that maternal TiO_2 exposure results in alternation to the cerebral cortex, olfactory bulb and the regions intimately related to dopamine systems of offspring mice.

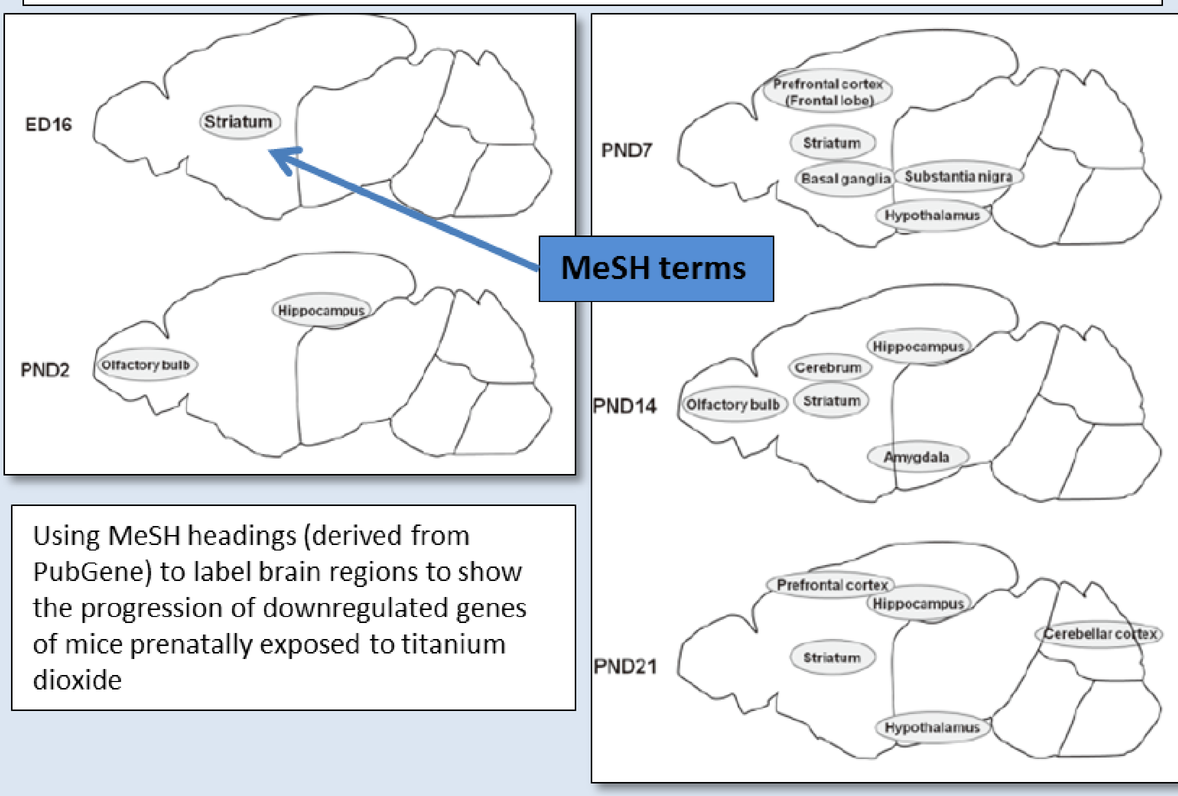


Figure 11: Example of a Functional Analysis project using MeSH and MeSH indexing (http://www.ncbi.nlm.nih.gov/pubmed/23208439)

ADDITIONAL FINDINGS

The research categories developed attempted to characterize how MeSH and MeSH indexing was used in the research. Not only was the number of projects discovered more numerous than originally anticipated, but the research was also extremely varied and diverse. Reviewers also attempted to identify broader areas of application to discover what areas these projects are useful for. The areas of application turned out to be so diverse that the reviewers could not create clean, neat categories. These areas ranged from various fields in biomedicine (such as clinical support, genetic association studies, and translational research) all the way to fields outside of biomedicine (such as patents, knowledge management, and funding policy).

In addition, the reviewers found research at largely varying stages of development. For example in the 225 relevant citations was research similar to "basic science" – for instance, an article about a specific algorithm (Wang 2014). But within the same 225 relevant citations were projects presenting fully-fledged consumer-ready tools like integrated databases or clinical information support tools. For example, the tool presented in Cheung 2012 is being used to prioritize genes in the area of personalized medicine. One interviewee using the tool explained:

> "The motivation is that if we exclude the tools that use MeSH terms right now, we would have say over 100 genes per patient to show to the clinician that these 100 genes may or may not be of interest to the patient's phenotype. But in order for a clinician to go through every one of those 100 is too time consuming. So from the bioinformatics perspective we try to make this easier. So [*this tool*] automatically extracts genes that are more likely to be relevant given the phenotype. So the clinician will provide the MeSH terms that best describe the patient's phenotype and then the tool would then take those as input and then the output would be a set of genes that best match those MeSH terms based on the PubMed literature and we return those genes as the higher priority."

The reviewers did not make any attempt to quantify the degree of research development, but it was an interesting discovery. Reviewers did attempt to identify web based tools and store them in a document for further development and analysis (Supplement 2).

In addition to the research categories, reviewers attempted to identify and note which pieces of MeSH and MeSH indexing were being used. This task was more difficult than originally imagined, because details provided by published literature vary widely. Some articles contained explicit and detailed information about which pieces of MeSH and MeSH indexing was used while other articles contained scanty details.

Despite this irregularity, the reviewers were able to identify an important finding. The fact that through MEDLINE MeSH hooks into the literature is very valuable for biomedical informatics researchers. The reviewers found that over half the projects of the 225 publications reviewed relied on the indexing in MEDLINE. MeSH as it exists in MEDLINE underpins many of these studies (Figure 12).

Pieces of MeSH / Indexing Used	No. of Citations
MeSH Thesaurus	152
Indexing (MEDLINE)	126
MeSH Hierarchy	109
SCR	23
Subheadings	16
Major Topics (IM)	13

Figure 12: The reviewers noted which pieces of MeSH and/or MeSH indexing was used for the research identified via published literature. These numbers represent what reviewers were able to extract, but may not be truly representative due to the fact that many publications lacked sufficient detail for this task.

INTERVIEWS

The goal of the interviews was to identify common themes and ultimately recommendations for NLM directly from the user group of interest – biomedical informatics researchers. However, in many areas biomedical informatics researchers did not share universal ideas about how to improve MeSH and MeSH indexing, and in fact in many cases these researchers actually had completely opposite opinions or needs.

For example, the interviewees collectively made many comments about the MeSH structure. Some of the researchers asserted that the MeSH vocabulary should be made into a true ontology, while other researchers defended the current MeSH structure. A true ontology formally represents knowledge and follows strict rules in order to make knowledge computationally useful. MeSH currently does not follow the rules of a true ontology, because it was built as a retrieval vocabulary. An example that illustrates this difference well is found in the MeSH heading "Accidents" (Figure 13).

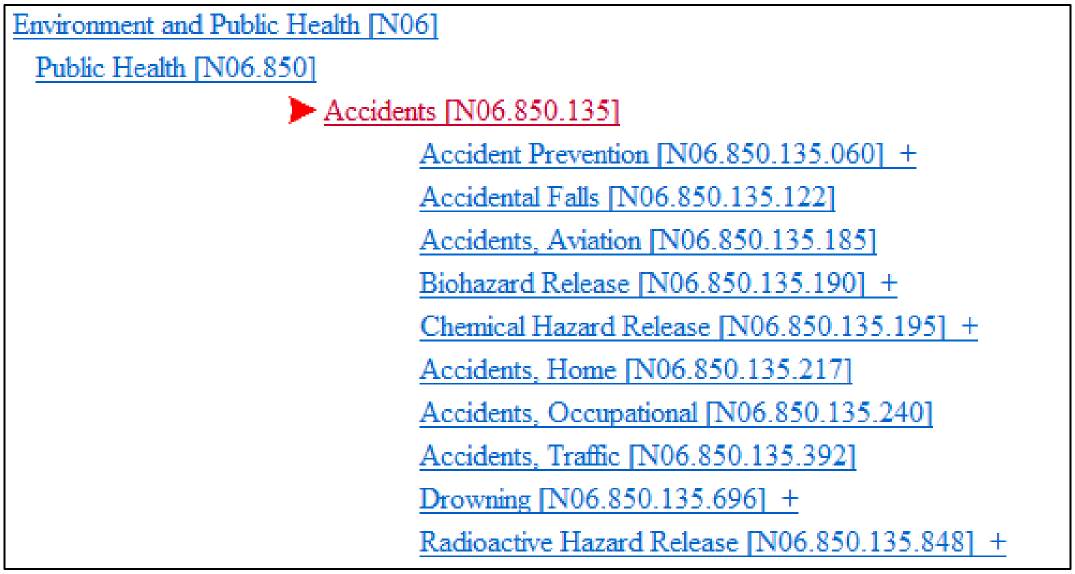

Environment and Public Health [N06]
Public Health [N06.850]
Accidents [N06.850.135]
Accident Prevention [N06.850.135.060] +
Accidental Falls [N06.850.135.122]
Accidents, Aviation [N06.850.135.185]
Biohazard Release [N06.850.135.190] +
Chemical Hazard Release [N06.850.135.195] +
Accidents, Home [N06.850.135.217]
Accidents, Occupational [N06.850.135.240]
Accidents, Traffic [N06.850.135.392]
Drowning [N06.850.135.696] +
Radioactive Hazard Release [N06.850.135.848] +

Figure 13: MeSH heading for "Accidents" illustrates the retrieval oriented structure of the MeSH tree hierarchy. "Accident Prevention" is not a type of Accident.

The MeSH heading "Accidents" has multiple child headings. Most of these children are types of accidents (for example, "Drowning" and "Accidental Falls"). However, the heading "Accident Prevention" is not a type of accident. This structure works for retrieval because one would assume that a user searching for information about accidents would also be interested in information about accident prevention. However, this is not consistent, formal knowledge representation and therefore breaks the rules of a true ontology.

One researcher who has used MeSH and MeSH indexing extensively for literature based discovery and retrieval applications defends the current retrieval-based hierarchical structure, saying:

> "If you think of any indexing terminology or subject heading vocabulary as an ontology you're going to get yourself in a lot of trouble very quickly. Doesn't matter if it's MeSH or the LoC subject headings, any subject headings is designed for organizing content matter, not knowledge. It's for retrieval purposes, which is very different. I think the ontology people don't often appreciate that difference and if you want to get into the ontology space you use a proper ontology…You shouldn't be using MeSH for representing knowledge. And that's of course my own personal view but I would worry if MeSH started to worry about its semantic relationships and taxonomic incorrectness. I mean I think that's not the job of an indexing terminology."

In contrast, other researchers (both in the literature and the interviews) try to leverage the MeSH vocabulary as an ontology. For example, Burdescu (2013) creates an ontological structure for MeSH from scratch in order to use the MeSH vocabulary for automatically annotating gastrointestinal images based on visual features (not text) to enable semantic image retrieval. Another group of researchers interviewed use MeSH for a variety of tasks, but do not use the MeSH hierarchy for these tasks. Instead, depending on the use case, they hang the MeSH vocabulary onto other ontological structures (for example, the SNOMED graph structure). These use cases involve processing text from Electronic Health Records (EHRs) for purposes such as pharmacovigilance (drug safety), phenotypic profiling, and answering unstudied clinical questions. In most cases (but not 100%), the research projects found that used MeSH as an ontology (either by creating a new structure from scratch or using a pre-existing ontological structure) were trying to use it in some capacity outside of the literature indexed in MEDLINE (for instance, images, EHRs, etc.).

In addition to the MeSH structure, researchers also had differing opinions on expansions to the MeSH vocabulary. These opinions tended to be very research specific. For example, one interviewee who researches methods to filter huge retrieval sets for systematic reviews expressed a need for MeSH to expand its coverage of study designs. Another interviewee felt that MeSH should not expand greatly, but that it could do more to align with other U.S. government initiatives. For example, this individual mentioned that Meaningful Use is not well covered by MeSH. Another interviewee has developed a pathogen database to describe the infectious pathogens of humans and animals. The database contains information about the pathogens, host animals, and geographical location of pathogens. For these researchers, the inclusion or linking of the NCBI Taxonomy database to MeSH, in addition to expanded geographical information, would be very helpful.

Furthermore, researchers do not share universal opinions about data formats. Most of the researchers agreed that the MeSH and MEDLINE data are easy to access, but there was not complete agreement on what data format is the best for biomedical informatics researchers. An example of this is illustrated by the varying responses received when interviewers asked the researchers about the usefulness of MeSH in Resource Description Framework (RDF). Some researchers were very interested in MeSH in RDF, some had not thought about it, and some said they would not use MeSH in RDF and are happy with the current data formats available.

While a lack of consensus prevents this project from making direct recommendations about large changes to MeSH and MeSH indexing, the fact that differing opinions exist has implications for future work. It is useful to understand the general themes that are important to biomedical informatics researchers. The lack of consensus also implies that many of the needs of this user group are research specific. And perhaps the largest takeaway is that because of the differing opinions found, NLM would benefit from continuing the conversation with biomedical informatics researchers.

Finally, while researchers expressed differing opinions on major topics of discussion, there was a nearly universal opinion that MeSH and MeSH indexing are valuable to biomedical informatics research. Using data from the literature reviews and the interviews, features that contribute to the value of MeSH and MeSH indexing for biomedical informatics are:

- MeSH connects to the published literature via MEDLINE indexing (see Figure 12)
- Because MeSH has existed since 1960 and has been regularly maintained, it has become familiar and trusted
- MeSH and MeSH indexing are used by many researchers, and this is valuable because it allows researchers to build off existing work and make direct comparisons
- MeSH and MeSH indexing are free of charge and license
- MeSH as a vocabulary is relatively small and general compared with other vocabularies such as SNOMED, which is larger and clinically oriented, and this is valuable in some types of biomedical informatics research
- The MeSH hierarchical tree structure is valuable, as illustrated by the number of projects found that try to use OMIM and MeSH together. OMIM is popular but unstructured, so researchers have tried many maneuvers to utilize the best of both (the structure of MeSH and the terms of OMIM - see Buizer-Voskamp 2010 for an example). However, a comparative analysis of vocabulary structures was not conducted, so it is unknown whether or not researchers prefer the MeSH structure to other available structures
- MeSH contains disease vocabulary, and there does not appear to be another better simple list of diseases that researchers are using

Identifying why MeSH and MeSH indexing are appreciated by biomedical informatics researchers is helpful because NLM can better understand what strengths these products have already in order to leverage and build upon these valuable features in the future.

LIMITATIONS

An interesting finding of this project is that it would be nearly impossible to conduct a comprehensive review of all the biomedical informatics research projects utilizing MeSH and MeSH indexing. This discovery is supported by two findings:

1. Full text searching yields additional relevant results - One of the resources used for the literature review, IEEE, allowed for full text searching. When IEEE was searched in the Title, Abstract, and Metadata, only 12 relevant citations were found. However, when using full text searching, 60 relevant citations were retrieved. It is not possible to know whether or not this 5-fold increase could be extrapolated to the other resources, but it does illustrate that many publications may mention the use of MeSH and MeSH indexing outside of the title and abstract.

2. Some research does not mention uses of MeSH or MeSH indexing - Through direct conversations with researchers, NLM discovered projects that incorporated MeSH and MeSH indexing that would not have been found in a literature search. For example, the article titled *Clinical Assessment Incorporating a Personal Genome* (Ashley 2010) uses MeSH to visualize a patient's genetic risk factors in the context of known environmental and behavioral factors, but the use of MeSH is not mentioned in the article (the use was mentioned in an interview). In addition, other project may use MeSH indirectly by using products that have already incorporated MeSH and MeSH indexing - such as PubGene, which is now COREMINE (Jenssen 2001). It is unknown how many articles may cite the use of a product such as PubGene but not mention MeSH or MeSH indexing directly.

Furthermore, because there were only two reviewers available for this project, individual citations were only reviewed by one person. If this project were to be expanded, a future step would be to recruit additional reviewers, preferably with informatics backgrounds, to review the research categories and "double annotate" the publications. The result of this could be published as a review article, and the citation analysis spreadsheet would be a supplementary document.

In addition, it is noted that 8 of the 25 interviewees were affiliated with Stanford University. This may introduce some bias into our interview data. However, considering that even with this potential bias no consensus was found in the interview data, the results are still valuable.

The results of this project are admittedly heavily qualitative. As the project unfolded, it became obvious that it would be very difficult to attempt to quantify the answers to the questions addressed in this research. For example, finding an exact number of biomedical researchers who have utilized MeSH and MeSH indexing would be near impossible. Establishing a number of users for derived products was similarly elusive, because many of the researchers interviewed did not have a clear idea of how many users their own derived products had. This project, therefore, serves as a starting point. It leads to many future areas of analysis and investigation. It portrays a "picture" of uses of MeSH and MeSH indexing in biomedical informatics research, not a statistical report.

RECOMMENDATIONS

The synthesized findings of the literature review and the interviews with biomedical informatics researchers illuminate both immediate recommendations for NLM as well as additional questions and areas for further exploration. For some of the recommendations that require further analysis, separate proposals could be drafted for future work (see Masterton_MeSH and MeSH indexing files.zip for some potential additional projects).

IMMEDIATE RECOMMENDATIONS

Write a paper to serve as a citation for MeSH and MeSH indexing

This recommendation was generated from a comment from one of the interviews:

> "It's very hard to cite MeSH in a paper. When I write a paper and I want to cite MeSH I'm like, you know, they don't have an official, published paper about MeSH. And so I have to go back and use these old papers or historical documents. If MeSH would write a paper in say the NAR (Nucleic Acid Research) Database Issue, or something like that, so there was an official paper there that other databases could consistently cite, it would probably help MeSH track who is citing MeSH and how it's being used. Right now it's very hard to properly cite MeSH. I think that would be a GREAT paper in the NAR Database Issue."

In addition, an informal review of how researchers cite MeSH and MeSH indexing further exposed the need for a citation that biomedical informatics researchers feel comfortable using to cite the use of MeSH and MeSH indexing in their work.

Eliminate the MeSH Memorandum of Understanding (MOU)

Currently, MeSH has a MOU that a user must fill out every time MeSH is downloaded (even if the same person has downloaded it in the past). While the inconvenience is minor, it represents a slight barrier to access. Furthermore, at the outset of the project the team tried to use the data from the MOU to contact researchers, but the data was not useful and would have been difficult to prepare for our purposes (de-duplication, etc.). Therefore, the recommendation is to either eliminate the MOU altogether, or alter it to provide useful information for NLM. Also, if the MOU is kept, it would be helpful if it could remember returning users.

Release MeSH in Resource Description Framework (RDF)

The fact that a fair number of biomedical informatics researchers verbally expressed interest in MeSH in RDF during interviews, compounded by the fact that several projects were found that were already using MeSH in RDF, leads to the recommendation that NLM release an authority version of MeSH in RDF (see Nolin et al. 2012 for an example illustrating a use case for MeSH in RDF). The NLM Linked Data Infrastructure Working Group is already moving in this direction. In addition, if MeSH were also release in Web Ontology Format (OWL), it could be added to Ontobee (http://www.ontobee.org/) - which is produced and managed by one of the researchers interviewed - and potentially to other linked data repositories. This may increase access and use.

Provide better documentation, including the indexing manual

The interview data exposed that biomedical informatics researchers would prefer if MeSH and MeSH indexing were more "transparent." A recommendation to address this is to provide better documentation for users trying to apply MeSH and MeSH indexing in downstream uses. One example of documentation that would be very useful for this group is the indexing manual. Another interviewee mentioned that there are certain assumptions that computer scientists make about the rules of a vocabulary, and since MeSH does not follow these rules consistently it would be helpful to know about this upfront (this also ties into the discussion about the MeSH structure).

<u>Improvements to the MeSH browser and thesaurus</u>

This recommendation covers improvements to the MeSH browser and thesaurus not related to specific vocabulary expansions. For example, several interviewees mentioned that having a machine-readable versioning history of MeSH would be very useful for certain data mining tasks. The need for machine-readable formatting could apply to all elements of the thesaurus as well. Other interviewees mentioned that the MeSH browser itself is outdated and difficult to navigate. And finally, several researchers suggested that it would be useful if the MeSH browser and thesaurus linked to additional resources.

<u>Promote use of MeSH and MeSH indexing in biomedical informatics research</u>

During the interviews, several researchers suggested that MeSH and MeSH indexing are underutilized resources in biomedical informatics research. Additional comments in published literature outside of the retrieval set for the review highlight this fact as well:

> "A conspicuously underutilized resource in systems biology is MeSH, whose coverage overlaps with GO and Online Mendelian Inheritance in Man (OMIM). It is straightforward computationally to link annotation from databases via unique citation identifiers…The absence of MeSH in systems biology studies may reflect the extra steps required to reach MeSH terms from more familiar annotation. Alternatively, it may reflect magnification of errors generated by mapping from one terminology to another" (Roberts 2006)
>
> "It should be noted that although using MeSH terms for querying PubMed records results in more specific retrieval than using words found in abstracts [8], querying with MeSH terms is an option seldom chosen by PubMed users [9]. So ultimately, it is possible that the huge value of the work that the NLM curators do in creating and using MeSH terms lies in enabling computational mining methods" (Andrade-Navarro 2012)

Continuing the investigation begun by this project will help to eventually make improvements to MeSH and MeSH indexing for these types of research, but an additional effort can be made to promote the use of these flagship NLM tools in biomedical informatics research.

<u>Continue the conversation</u>

Because the interviews did not achieve consensus, but provided a great deal of unpublished information about the needs of this user group, it is very important to continue this conversation with biomedical informatics researchers using MeSH and MeSH indexing for research purposes beyond retrieval of the literature. While this continued conversation could take the form of additional interviews and follow up surveys, it would be useful to explore alternative strategies to stay in long-term, ongoing contact with this user group. Some ideas for future communication could include a community website, a liaison from NLM to serve this specific user group, or some type of network, like the NLM National Network for Medical Libraries.

ADDITIONAL QUESTIONS AND AREAS OF EXPLORATION

MeSH expansions and integration

There was no consensus identified about what areas of the MeSH vocabulary would be most valuable to expand for biomedical informatics researchers, yet there is an obvious need for biomedical informatics researchers to tailor the MeSH vocabulary for specific research purposes. NLM should investigate how to make this integration and tailoring as easy as possible for these researchers, even when the vocabulary is not expanded for NLM's purposes. In addition, it may be valuable to explore and analyze the concept of unofficial MeSH extensions that researchers can create and share, but that would not be used for indexing the literature in MEDLINE.

Comparative vocabulary analysis

The entry point for the current research was limited in that it sought out projects already using MeSH and MeSH indexing in biomedical informatics research. This means that the information and opinions gathered are likely incomplete. A comparative analysis with other vocabularies that are heavily used in biomedical informatics research could help inform NLM about how additional vocabularies are used my biomedical informatics researchers, what features of these vocabularies are most valued by this user group, etc.

Further analysis of potential changes to MeSH structure

This project was unable to provide specific recommendations about alterations in the MeSH structure. However, this area seems to be one of high value for continuing research. For example, several biomedical informatics researchers expressed the opinion that MeSH should be a true ontology. Further analysis could be conducted to discover why this would be valuable to biomedical informatics researchers (ie: what can they do with true ontologies that they cannot with a retrieval-based structure?), what the process of making MeSH into an ontology would be (and what it would cost), and how such a change would affect literature retrieval.

Additionally, analysis could be conducted to discover smaller changes to the current MeSH hierarchy that may make rules more consistent and therefore easier for biomedical informatics researchers to use.

Establishing the future value of uses of MeSH and MeSH indexing in biomedical informatics research

An analysis in this topic would try to establish whether or not this is a growth area for MeSH and MeSH indexing. Methods for this could include a bibliometric analysis over time, as well as comparative analyses with other resources used in biomedical informatics research.

Expansion of literature annotations and other curation efforts

Through the interviews, the project attempted to establish what types of additional literature annotation would be most valuable for biomedical informatics research. There was, however, no clear consensus. The closest this research can come to a recommendation is that if possible, NLM's current gene indexing should be expanded. This is an area where further investigation is needed. Recommendations for future analysis include identification and characterization of internal and external biocuration efforts, and a continued evaluation of what literature annotations would provide the greatest value to the research community.

Prioritization of users

The research conducted for this report focused solely on biomedical informatics researchers as users of MeSH and MeSH indexing. However, this may not be the largest user group of MeSH and MeSH indexing. Further analysis is needed to establish the utility of MeSH and MeSH indexing to other user groups in order to understand where and how NLM can provide the greatest value.

CONCLUSIONS

The primary findings of this project are:

- The user group of biomedical informatics researchers using MeSH and MeSH indexing is larger than originally anticipated
- The types of research using MeSH and MeSH indexing in biomedical informatics research beyond literature retrieval is very diverse
- Because the research is so varied, the audience for the biomedical informatics research using MeSH and MeSH indexing is broad
- Biomedical informatics researchers do not have universal opinions about how MeSH and MeSH indexing can be improved
- MeSH and MeSH indexing have great utility beyond literature retrieval

This project represents a starting point for a potentially much larger investigation. The key takeaways as outlined above point to a need to continue the exploration.

REFERENCES

Andrade-Navarro MA. Mining the literature: new methods to exploit keyword profiles. Genome Med 2012;4(10).

Ashley E, Butte A, Wheeler M, Chen R, Klein T. Clinical assessment incorporating a personal genome. Lancet 2010; 375(9725):1525-35

Buizer-Voskamp JE, Franke L, Staal WG, van Daalen E, Kemner C, Ophoff RA, Vorstman JA, van Engeland H, Wijmenga C. Systematic genotype-phenotype analysis of autism susceptibility loci implicates additional symptoms to co-occur with autism. Eur J Human Genet 2010;18(5): 588-95.

Burdescu D, Mihai C, Stanescu L, Brezovan M. Automatic image annotation and semantic based image retrieval for medical domain. Neurocomputing 2013;109(3): 33-48.

Cheung WA, Ouellette BF, Wasserman WW. Quantitative biomedical annotation using medical subject heading over-representation profiles (MeSHOPs). BMC Bioinformatics 2012;13(249).

Jenssen TK, Laegreid A, Komorowski J, Hovig E. A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 2001;28(1): 21-8.

Kahn CE. Annotation of figures from the biomedical imaging literature: a comparative analysis of RadLex and other standardized vocabularies. Acad Radiol 2014;21(3): 384-92.

Libscomb, CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000; 88(3):265-66.

Makalowski, W. (2003). Bioinformatics. In Encyclopedia of the human genome.

Nolin M, Dumontier M, Belleau F, Corbeil J. Building an HIV data mashup using Bio2RDF. Brief. Bioinform. 2012;13(1):98-106.

Roberts PM. Mining literature for systems biology. Brief Bioinform 2006;7(4):399-406.

Ruau D, Dudley JT, Chen R, Phillips NG, Swan GE, et al. Integrative approach to pain genetics identifies pain sensitivity loci across diseases. PLoS Comput Biol 2012;8(6)

Ruau D, Mbaqwu M, Dudley JT, Krishnan V, Butte AJ. Comparison of automated and human assignment of MeSH terms on publicly-available molecular datasets. J Biomed Inform 2011;44(Suppl 1)

Sánchez, D., Solé-Ribalta, A., Batet, M., Serratosa, F. Enabling semantic similarity estimation across multiple ontologies: an evaluation in the biomedical domain. J Biomed Inform 2012; 45(1)

Sarkar IN, Schenk R, Miller H, Norton CN. LigerCat: using “MeSH Clouds” from journal, article, or gene citations to facilitate the identification of relevant biomedical literature. AMIA Annu Symp Proc 2009: 563-567.

Sharma V, Sarkar IN. Leveraging concept-based approaches to identify potential phyto-therapies. J Biomed Inform 2013;46(4): 602-14.

Umezawa M, Tainaka H, Kawashima N, Shimizu M, Takeda K. Effect of fetal exposure to titanium dioxide nanoparticle on brain development – brain region information. J Toxicol Sci 2012;37(6):1247-52.

Wang Y, Fang H, Yang T, Wu D, Zhao J. Degree-adjusted algorithm for prioritization of candidate disease genes from gene expression and protein interactome. IET Syst Biol 2014;8(2):41-6.

Weber G. Identifying translational science within the triangle of biomedicine. J Transl Med 2013; 11(126)

APPENDIX

SEARCH STRINGS

PubMed

("medical subject headings" OR "medical subject heading") AND (information science[mesh] OR indexing OR "hypothesis generation" OR "literature based discovery" OR "data mining" OR "natural language processing" OR "data visualization" OR "network analysis" OR "graph analysis" OR "artificial intelligence" OR rdf OR semantic* OR integration OR "text mining" OR annotation OR "knowledge base" OR bibliometric* OR microarray) AND "last 10 years"[dp] NOT (editorial[pt] OR biography[pt] OR comment[pt] OR guideline[pt])

Yield: 577 on 5/13/2014

("medical subject headings" OR "medical subject heading") AND (information science[mesh] OR indexing OR "hypothesis generation" OR "literature based discovery" OR "data mining" OR "natural language processing" OR "data visualization" OR "network analysis" OR "graph analysis" OR "artificial intelligence" OR rdf OR semantic* OR integration OR "text mining" OR annotation OR "knowledge base" OR bibliometric* OR microarray) AND "last 5 years"[dp] NOT (editorial[pt] OR biography[pt] OR comment[pt] OR guideline[pt])

Yield: 299 on 5/30/2014

Scopus (Advanced Search)

TITLE-ABS-KEY(**"medical subject headings"** OR **"medical subject heading"**) AND (**"information science"** OR **indexing** OR **"hypothesis generation"** OR **"literature based discovery"** OR **"data mining"** OR **"natural language processing"** OR **"data visualization"** OR **"network analysis"** OR **"graph analysis"** OR **"artificial intelligence"** OR **rdf** OR **semantic*** OR **integration** OR **"text mining"** OR **annotation** OR **"knowledge base"** OR **bibliometric*** OR **microarray**) AND (PUBYEAR > **2004**) AND (LIMIT-TO(LANGUAGE, **"English"**))

Yield: 476 on 5/13/2014

TITLE-ABS-KEY(**"medical subject headings"** OR **"medical subject heading"**) AND (**"information science"** OR **indexing** OR **"hypothesis generation"** OR **"literature based discovery"** OR **"data mining"** OR **"natural language processing"** OR **"data visualization"** OR **"network analysis"** OR **"graph analysis"** OR **"artificial intelligence"** OR **rdf** OR **semantic*** OR **integration** OR **"text mining"** OR **annotation** OR **"knowledge base"** OR **bibliometric*** OR **microarray**) AND (PUBYEAR > **2009**) AND (LIMIT-TO(LANGUAGE, **"English"**))

Yield: 259 on 5/30/2014

IEEE (Institute of Electrical and Electronics Engineers)

“medical subject headings” in Metadata Only

Yield: 27 on 5/13/2014

“medical subject headings” in Full Text & Metadata

Yield: 277 on 5/13/2014

Yield: 232 (filtered to 2004 and after)

There are definitely things in the full text we miss in just the metadata, but there will be more junk too.

“medical subject headings” in Full Text & Metadata

Yield: 277 on 5/30/2014

Yield: 145 (filtered to 2009 and after)

INTERVIEW QUESTIONS (WITH CHECKLIST)

1. Could you tell me about your research, including the problems this research is trying to solve and the estimated impact of the research?
    - ☐ What is your ultimate goal for this project?
    - ☐ How many users do you think you have?
    - ☐ Is your product open source?
    - ☐ Can you summarize the impact the product or research has had so far?

2. How did you decide to use MeSH and/or MeSH indexing for this research, and how specifically do you utilize MeSH and/or MeSH indexing in your research?
    - ☐ Did you consider other options?
    - ☐ What is most useful about MeSH and MeSH indexing for your research?
    - ☐ Which elements of MeSH and/or MeSH indexing do you use in your research and what is the significance or each element?
        - ☐ Subject Headings (Descriptors)
        - ☐ Subheadings(Qualifiers)
        - ☐ Major Topics (IM concepts, “starred” concepts)
        - ☐ Substances
        - ☐ Entry terms (thesaurus)
        - ☐ Tree structure (parents and children relationships)
        - ☐ Other parts of the MeSH record? (See also, previously indexed as…)

3. What do you wish MeSH and/or MeSH indexing had or could do that it currently does not? For your research purposes, how do you work around any shortcomings or insufficiencies?
    - ☐ What other vocabularies and resources do you use in this research?
    - ☐ Could your research be enhanced by linking or mapping MeSH in some way to other resources?
    - ☐ Have you found any additional resources you would like to map to but cannot, and why?
    - ☐ Are there changes you would like to see made to MeSH and/or MeSH indexing?

4. Besides the current subject indexing, are there other types of curation or annotation of the literature that you would find helpful?

5. NLM is very interested in making our data accessible and usable. For your purposes, are you satisfied with the way the data you need is delivered? Please discuss any issues with access to MeSH and/or MEDLINE data that you have.
    - ☐ Would it have been helpful to have MeSH and/or MEDLINE available in Resource Description Framework (RDF)?
    - ☐ Would you like to be able to pick and choose specific parts of MeSH and/or MEDLINE for download, or is it easier to download everything for your purposes?
    - ☐ Did our licensing or memorandums of understanding create any obstacles for you?
    - ☐ How are updates to MeSH reflected in derived products or research?

6. Do you know any other researchers using MeSH or MeSH indexing?