The Unified Medical Language System: A Scoping Review of its Use in Research

Brody, Stacy (NIH/NLM) [C]

Associate Fellow, 2018-2019

Project Sponsors

Liz Amos (NIH/NLM/OD)

Anna Ripple (NIH/NLM/LHC)

August 2019

Contents

Abstract

Background: The Unified Medical Language System (UMLS) has greatly impacted biomedical and bioinformatics research.

Objective: The goal of this project is to provide an analysis and synthesis of the research conducted using the UMLS within the past 15 years to provide a foundation for strategic planning for the next version of the UMLS.

Methods: The Associate Fellow and project sponsors searched for scientific literature which used the UMLS as a tool in research. The Fellow then led the development of a protocol and data extraction form.

Results: The Fellow and project sponsors identified 3,510 citations. A sample of 348 articles were reviewed during the development and refinement of a protocol and data extraction form.

Conclusions: The work presented in this report demonstrates a proof of concept, including tested screening criteria, protocol, and data extraction form. The preliminary results provide insights into the research being conducted using the UMLS and derived products and demonstrate the need to continue support for those parts of the UMLS used in linking or mapping terms and processing texts. An extension of this work, using the methodology presented here and applied to the remaining citations, could uncover additional trends or bolster those identified in the random sample to inform the visioning of the UMLS.

Background

Purpose/Objective/Research Question

The goal of this project is to provide an environmental scan of research conducted using the Unified Medical Language System (UMLS). Specifically, the Fellow aims to describe how the UMLS and derived products are being employed in research and in what broad research categories. The results of this scoping review will influence the strategic visioning for the UMLS.

An additional aim of this work is to describe the development of a scoping review protocol, including the development of screening criteria and a data extraction form.

The Unified Medical Language System

The UMLS was started in 1986 and launched in 1990 with a vision to help users contend with the problem of multiple terminologies. It consists of three knowledge sources, the Metathesaurus, Semantic Network, and the SPECIALIST Lexicon and Lexical Tools. The Metathesaurus brings together the terms and codes of over 200 vocabularies, grouping them into concepts and enabling crosswalking between terminologies. The Semantic Network provides a hierarchical organizational structure by which to organize terms and concepts. The SPECIALIST Lexicon and Lexical Tools support natural language processing applications. For the purposes of this review, UMLS broadly includes the Unified Medical Language System, its components, and related tools and applications, such as MetaMap and SemRep.

Upon its 30-year anniversary, the UMLS is to be re-envisioned to enable the future of biomedical and computational science research. The results of this scoping review, which intends to describe how researchers use the UMLS and derived products in research, along with other information sources such as the annual survey and a spring 2020 planning workshop, will inform the visioning process.

Scoping Reviews

Scoping reviews are conducted to provide environmental scans of a research area and to provide context for decision making and strategic planning.

A type of review methodology, the scoping review provides a landscape view of research. Scoping reviews can be used to determine “the way the research has been conducted” and “are designed to provide an overview of the existing evidence base regardless of quality” (p. 142, Peters et al., 2015). Scoping reviews consist of five steps, with an optional sixth:

1. Identify the research question
2. Identify relevant studies
3. Select studies
4. Extract data from included studies
5. Collate, summarize and report results
6. (optional) Consult with stakeholders

(Tricco et al., 2016; Arksey and O’Malley, 2003).

Scoping reviews are typically conducted by teams and involve an iterative protocol development process, contrasted with the linear systematic review methodology (Arksey and O’Malley, 2003). In this report, the Fellow describes the iterative nature of the review and the results of screening and extracting data from a random sample.

Methods and Materials

The aim of this project is to describe how researchers use the UMLS and in what research areas the UMLS is being used. In this section, the search methodology, sampling protocol, and screening criteria and data extraction form development are described.

Search: Identify Relevant Studies

To identify published papers using the UMLS in research, the Fellow and project sponsors used a two-prong search methodology. Searches were executed in February and March 2019.

First, Fellow and project sponsors searched PubMed (ncbi.nlm.nih.gov/pubmed), Web of Science Core Collection, and Scopus

'UMLS OR Unified Medical Language System OR Unified Medical Language Systems'.

The PubMed search was conducted for all fields. In Web of Science, fields were restricted to topics. In Scopus, fields were restricted to Title, Abstract, Keyword. The search was necessarily broad in order to capture the fullest scope of research.

Additionally, one project sponsor conducted a search

"MetaMap" OR "Metathesaurus" OR "specialist lexicon"

This search added 148 unique citations beyond those captured in the UMLS search above.

An alternative method not employed in this study is to use the PubMed Central functionality for searching within the Methods and Materials sections of papers. This method was not used because this review seeks to capture research beyond that funded by the National Institutes of Health and in the biomedical domain.

Second, one project sponsor identified articles citing seminal papers. A combination of address field searching (for National Library of Medicine variants) AND author name (from NLM) searching yielded 3,000 citations. This was manually reviewed and reduced to 957 citations. From the 957, the top 10 highest-cited UMLS publication were identified:

Author	Title	Source	Total Citations	Publication Year
Bodenreider, O.	The Unified Medical Language System (UMLS): integrating biomedical terminology	Nucleic Acids Research	771	2004
Lindberg, D.A.B.; Humphreys, B.J.; McCray, A. T.	The Unified Medical Language System	Methods of Information in Medicine	621	1993
Aronson, A.R.; Lang, F.M.	An overview of MetaMap: historical perspective and recent advances	Journal of the American Medical Informatics Association	366	2010

Aronson, A.R.	Effective mapping for biomedical text to the UMLS Metathesaurus: the MetaMap program	Proceedings. AMIA Symposium.	358	2001
Humphreys, B.L.; Lindberg, D.A.B.; Schoolman, H.M.; Barnett, G.O.	The Unified Medical Language System: an informatics research collaboration	Journal of the American Medical Informatics Association	247	1993
Bodenreider, O.; Stevens, R.	Bio-ontologies: current trends and future directions	Briefings in Bioinformatics	168	2006
Rindflesch, T.C.; Fiszman, M.	The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text.	Journal of Biomedical Informatics	160	2003
Demner-Fushman, D.; Chapman, W.W.; McDonald, C.J.	What can natural language processing do for clinical decision support?	Journal of Biomedical Informatics	142	2009
McCray, A.T.; Nelson, S.J.	The representation of meaning in the UMLS	Methods of Information in Medicine	115	1995
Humphreys, B.L.; Lindberg, D.A.B.	The UMLS project: making the conceptual connection between users and the information they need.	Bulletin of the Medical Library Association	98	1993

Table 1: Web of Science Top 10 Most Frequently Cited UMLS publications

The 10 publications (Table 1) had a total of 3,046 citing publications. The Web of Science functionality 'Citation Reports' was used to remove self-citations, reducing the 3,046 publications to approximately 2,550 .

This second methodology was included to capture articles not returned in the UMLS string search. Articles may, for instance, describe the use of MetaMap without describing the UMLS. Though the UMLS website instructs users to cite the Bodenreider, 2004, article, this varies in practice. Authors cite papers in the list above, provide URLs, or provide neither indirect nor direct citation.

We merged the UMLS string and citation search result sets in EndNote X9. Using built-in EndNote functionalities, duplicates were removed, and citations were limited to publication dates of 2005 or later.

Sampling

After discussions of the timeline and the feasibility of developing and executing a protocol on the full sample, the Fellow and project sponsors opted to take a sample. This would facilitate the development of a clear methodology and provide preliminary results for stakeholders.

Using a confidence interval of 95% with a 5% margin of error, and employing the calculator from Survey Monkey (surveymonkey.com/mp/sample-size-calculator/), a sample size of 347 was calculated. Citations were exported from EndNote X9 to Excel. The RAND function was used and citations ordered by newly-generated random numbers. The top 348 citations constituted the sample used in the remaining steps to develop and refine screening and charting processes and provide proof of concept and preliminary results.

Screening: Select Studies

The 348 citations were loaded into Colandr, a web-based tool designed for conducting collaborative reviews (colandrapp.com/, Cheng et al., 2018). The Fellow learned of the tool through the National Network of Libraries of Medicine webinar series on systematic reviews (nnlm.gov/scr/training/systematic-review-series).

Three reviewers, the Fellow and two project sponsors, manually screened the title, abstract, and full text. Each citation was screened by two independent reviewers. Colandr includes a functionality for two screeners per article. The screening criteria was initially unclear and developed as the team read and discussed articles. The Fellow and project sponsors gained familiarity with the diverse corpus and refined the screening criteria. As Arksey and O'Malley (2003) describe in their seminal work on scoping review methodology, the search is necessarily broad, and "decisions… can be made once some sense of the volume and general scope of the field has been gained" (p. 23). In their study, too, "criteria were devised *post hoc*, based on increasing familiarity with the literature" (Arksey and O'Malley, 2003, p. 26).

The Fellow and project sponsors met weekly to discuss issues that arose during screening. Because this scoping review aims to provide a landscape report of the research being conducted using the UMLS, articles which referred to the UMLS and did not use the UMLS as a research tool were excluded. Ultimately, articles were excluded based on four criteria:

1. Language: Non-English language articles were excluded.
2. Article type: Articles which did not describe original research, i.e. review articles and commentaries, were excluded.
3. Not related to UMLS: Articles which did not refer to or describe the UMLS or related tools. Many were irrelevant and referenced, for instance, Unified Modeling Language (UML).
4. UMLS not used: Articles which referred to the UMLS and did not employ the UMLS as a research tool were excluded.

The final criterion, UMLS not used, is essential to fulfilling the purpose of the project. However, making a judgment according to this criterion required careful reading. Relevant information was often found in the materials and methods sections of the articles. Full-text review was necessary, even at the initial screening stage. Screening disagreements were recorded on over 50 articles because one screener used only the abstract and one used full text. Reliance on abstracts alone could have resulted in the wrongful exclusion of these and perhaps more articles. For the purposes of this methods-dependent review, title and abstract screening is insufficient.

As noted, each article was screened by two independent reviewers. Reviewers were permitted to select more than one reason for exclusion. Reviewers were not required to agree on the reason for exclusion, only on the decision to include or exclude the article. Disagreements over inclusion and exclusion were decided by discussion among reviewers at weekly meetings. The functionality of Colandr enables reviewers to change screening decisions.

Data Charting: Extract Data from Included Studies

A data extraction form evolved from weekly discussions and reading articles in the sample set. The form included use cases identified on the UMLS website (nlm.nih.gov/research/umls/). The form also included research category, with scope notes drawn from MeSH.

The team tested the form by orally reviewing several articles during a weekly meeting. Additionally, the form was tested by two groups of two reviewers on unique sets of 25 articles. The 25-article sets were randomly selected from the included articles of the screening sample. The random selection was completed similarly to our sample selection, using the RAND() function in Excel.

Inputting the data extraction form into Colandr required patience and some trial and error. Certain characters are prohibited. Documentation for the tool is unclear. Once the form was entered successfully and the 25-article sets uploaded into independent Colandr projects, reviewers began data extraction. Each reviewer logged onto their own Colandr project to ensure the extraction was independent. A Colandr tutorial was developed to assist new users (see Appendix 1).

Interrater reliability within reviewer pairs was calculated to identify areas in which instructions, scope notes, or field options needed modification.

Disagreements were discussed to refine scope notes and instructions before proceeding to data extraction on the remainder of the sample. MeSH topics were referenced for indexed articles to settle some disagreements on research type.

After completing data extraction for the initial sets of 25, the data extraction form was revised. Categories were merged and scope notes added for clarification.

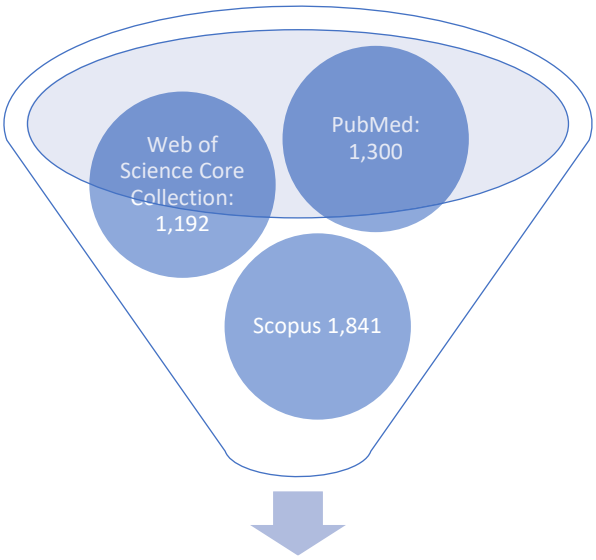
Data extraction continued with one reviewer per article and in a unified Colandr project to allow reviewers to log on and off as time allowed.

Results

Search Results

The PubMed search yielded approximately 1,300 results, Web of Science Core Collection topic search 1,192 results, and Scopus TITLE-ABS-Key 1,841 results.

Deduplicating and limiting by publication date resulted in 3,510 unique citations with dates 2005 or later. Excluded for pre-2005 publication date were 1,238 articles.



3,510 unique post-2005 citations

(plus 1,238 pre-2005 citation)

Figure 1: Search Results

Sampling

The reviewers took a random sample of 348 citations from the search results set. This sample set had similar percentages of reference types as the full citation set.

Reference Type	Percentage of Sample Set	Percentage in Full Set
Conference Proceedings	39%	37%
Periodical	58%	60%
Book	2%	3%
Book Series	1%	1%

Table 2: Reference types, as percentage of the sample and full results sets

The sample set also had similar trends for article publication year frequencies.

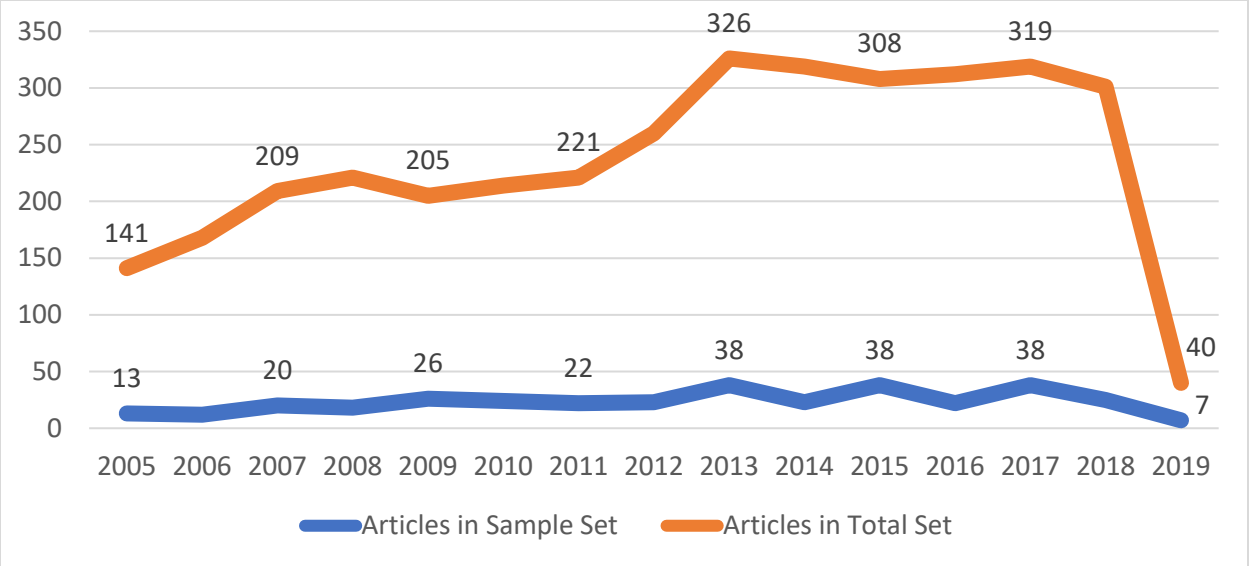


Figure 2: Counts of articles by publication year in the sample and full results sets

Publication Year	Articles in Sample Set	Articles in in Full Set
2005	13	141
2006	12	168
2007	20	209
2008	18	221
2009	26	205
2010	24	214
2011	22	221
2012	23	260
2013	38	326
2014	23	319
2015	38	308
2016	22	312
2017	38	319
2018	25	301
2019	7	40

Table 3: Counts of articles by publications year in the sample and full results sets

Screening

Included Articles

After screening the sample set of 348 citations, 198 citations were included. Most citations, 114 of the 198, were periodicals. The most common periodical titles were the *Journal of Biomedical Informatics* and the *Journal of the American Medical Informatics Association*. Conference proceedings accounted for 83 included citations. The most common proceedings title was the American Medical Informatics Association. There were long tails of over 50 journal titles and over 30 conference titles with fewer than 10 citations each. There was one book citation.

Excluded Articles

At the screening stage, 150 articles were excluded. Reviewers agreed on at least one reason for exclusion for 102 of the 150 articles. For 48 articles, reviewers agreed the articles needed to be excluded but not on the reason for exclusion.

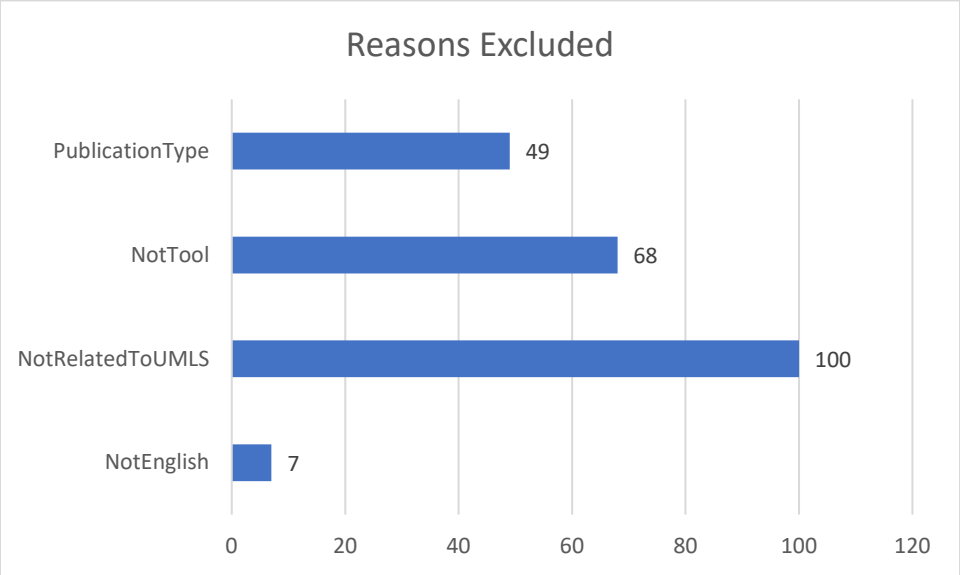


Figure 3: Counts of articles excluded according to screening criteria. The most common reason for exclusion was not being related to the UMLS.

An example of a paper excluded for Publication Type, that is, a paper not describing original research, is "Biomedical Informatics for Cancer Research: An Introduction" (Ochs, Casagrande, and Davuluri, R.V., 2010). The search strategy retrieved, in addition to research articles, commentary, perspectives, and materials such as this, which do not describe novel research.

The article "Consistency analysis of UMLS terminological and conceptual relations" (Barriere, 2012). was excluded for the reason Not Tool. Both reviewers agreed that, in this paper, the UMLS is treated as a topic of research instead of being used as a tool to conduct research.

As seen in the figure above, many articles were excluded because reviewers identified that they were not related to UMLS. For instance, in the article "Model-driven development based transformation of stereotyped class diagrams to XML schemas in a healthcare context" (Domínguez, Eladio, et al., 2007), the authors "propose the use of class diagrams of the Unified Modeling Language (UML) with stereotypes and eXtensible Markup Language (XML) schemas". We saw many irrelevant results which included the Unified Modeling Language rather than the Unified Medical Language System.

Data Charting

A data extraction form with 13 fields was developed (Appendix 2)

As noted in the Methods section, the form was revised after the initial sets of 25 articles were completed (Appendix 3). Due to inconsistencies and confusion in coding, the UMLS uses "facilitate mapping" and "link terms and codes" were consolidated. Scope notes for UMLS uses were also added.

Subcategories of research types were included in the revision. These were initially offered as scope notes. In coding the initial 25, the group found these granular categories helpful and noted that any grouping could be done in the analysis stage after extraction.

Interrater Reliability Measures

Interrater reliability measures were calculated for the initial sets of 25 citations completed as independent reviewer pairs.

	Percent Agreement	Kappa
Set 1 (BH & AR)	88.8%	0.76
Set 2 (SB & LA)	86.5%	0.5984

Table 4: Interrater reliability metrics for the 25-article sets

In set one, one article showed no agreement (Kappa 0 to 0.20), six articles showed weak agreement (Kappa 0.40 to 0.59), 12 showed moderate agreement (Kappa 0.60 to 0.79), and six showed strong agreement (Kappa 0.80 to 0.90).

In set two, the reviewers identified one article which should have been excluded at screening for publication type. Of the remaining 24, four articles reflected minimal agreement (Kappa 0.21 to 0.39), four articles showed weak agreement (Kappa 0.40 to 0.59), 12 showed moderate agreement (Kappa 0.60 to 0.79), three showed strong agreement (Kappa 0.80 to 0.90), and one showed almost perfect agreement (Kappa above 0.90).

Questions for which “N/A” and “No” were common answers had lower levels of agreement and lower Kappa values because reviewers often left these fields blank.

It was also noted that the field Research Category had lower percent agreements (87.2%, 78%) and Kappa values (0.72, .3612).

Discussions led to refining scope notes and instructions and revising data for consistency, as described in the section above.

Extraction Results

Four reviewers singly extracted data from 110 articles.

Research Types

Research type broadly describes the category or area of research described in the paper.

Artificial Intelligence emerged as the leading research category, followed by Information Storage and Retrieval, and Information Services. An example of an article classified as Artificial Intelligence is “Classifying free-text triage chief complaints into syndromic categories with natural language processing” (Chapman, et al., 2005).

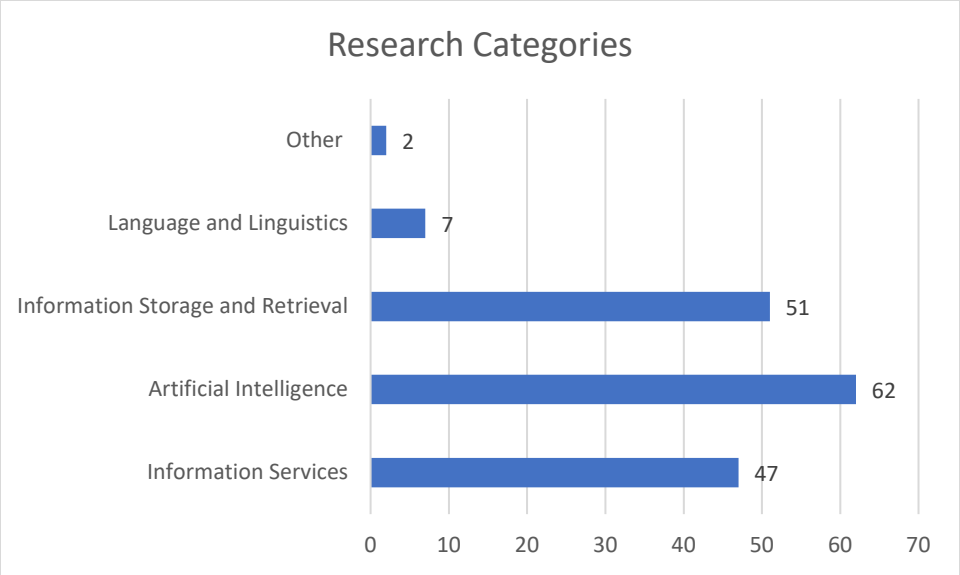


Figure 4: Counts of articles by research category

UMLS Products

Reviewers extracted the UMLS products and derived tools used, as named by the authors or inferred from the methods.

Eighty-three articles used the Metathesaurus. The second most commonly used UMLS product was MetaMap, employed in 48 of 110 articles. In two articles, the reviewers were unable to infer which UMLS product was used.

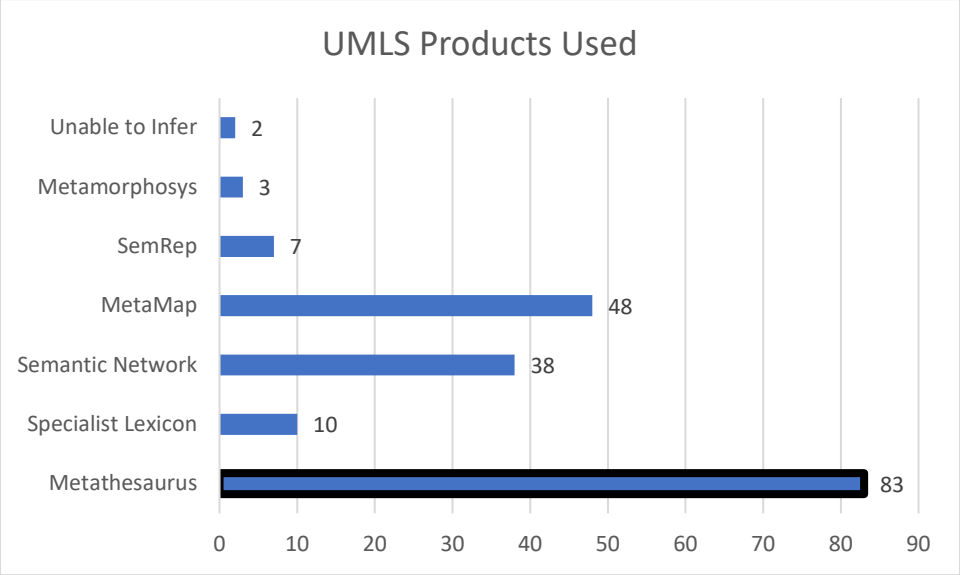


Figure 5: Counts of articles using various UMLS products

Corpus

The corpus is the textual dataset upon which the authors conducted their research, as described in the article.

Forty-six articles used scientific literature as the research corpus. Electronic health records were used in 35 of the 110 articles. Other articles used data from genomic or protein databases or user-generated content, such as that found in online health fora.

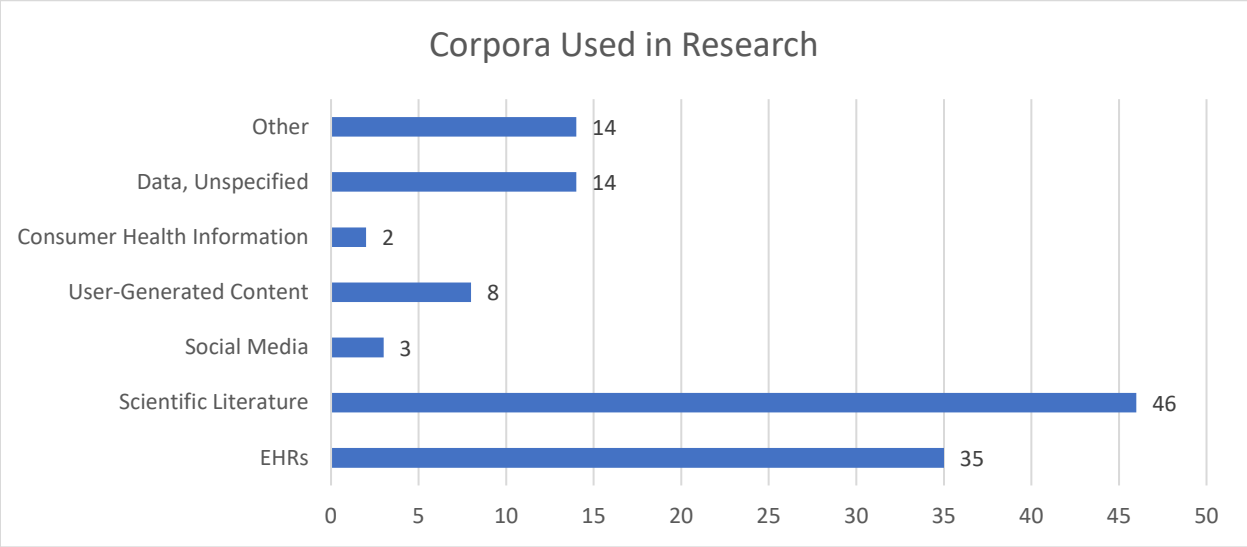


Figure 6: Counts of articles using different types of corpora

Other NLM Products

Other NLM products was a free-text field in the data extraction form, allowing reviewers to include a variety of products, as named by the article authors.

Sixty-nine articles used other NLM products. These include PubMed, MEDLINE, and PMC, as would be captured in the corpus Scientific Literature. Researchers also used a variety of other NLM products and services:

> PubMed, MEDLINE, PMC, Clinical Query Filters, ClinicalTrials.gov
>
> Vocabularies, i.e. SNOMED CT, RxNorm, MeSH
>
> NCBI & LHNCBC Databases, i.e. RefSeq, GenBank, Gene, dbSNP, OMIM, PubChem, OpenI
>
> Tools and Products Derived from UMLS, i.e. MetaMap, SemRep, API, Medical Text Indexer, and relationship files

UMLS Uses

Text processing was the most common use of the UMLS, reported for 67 of the 110 articles. The second most common use of the UMLS was to facilitate mapping or linking, reported for 43 articles.

For some articles, it was clear the UMLS was being used to facilitate mapping

> *we will also map ATC drugs to NDF-RT via UMLS as the intermediate identifier"*
> (Zhu and Tao, 2014)

Presentation of Results to Stakeholders

Arksey and O'Malley (2003) emphasize the importance of the optional sixth step in their methodological framework for scoping reviews. This is the consultation exercise. The Fellow presented the results to one of the key stakeholders at NLM and hopes the work completed during the Associate Fellowship Program can be combined with other datasets, such as bibliographic data and annual user survey results, to inform discussion among stakeholders at a spring 2020 workshop for visioning the next UMLS.

Discussion, Limitations, and Recommendations

Compared to Other Findings

The current work relied on the scholarly literature to describe uses of the UMLS. Prior studies have used surveys of UMLS users.

In 2006, a report delivered at the American Medical Informatics Association Annual Symposium described the users and uses of the UMLS from the annual user reports data. Their findings, which use data collected from 1,427 users, indicate that most users employ the UMLS to process clinical text (81%), though users also process nonclinical text, including bibliographic information (14%) and consumer health information (10%) (Fung, Hole, and Srinivasan, 2006). In the current work, electronic health records were used as corpora for 32% of articles, bibliographic information or scientific literature for 42%, and consumer health information in only 2%. The surveys may include more data from users beyond traditional researchers, for instance, those in the healthcare industry.

In 2007, independent researchers analyzed 70 responses to a 26-question survey of users and uses of the UMLS. Respondents overall reported using the UMLS most frequently in research (73% of respondents). The next most frequently reported operation was prototype design (31%). Terminology research was the most frequently reported purpose of use, followed by information retrieval and terminology translation. These results differ from those of the current work, in which terminology research is reported for only 6%. In the 2007 survey, about 44% of respondents reported using the UMLS as a mapping tool (Chen et al. 2007). Similarly, the current work reports that 39% of articles employed the UMLS to facilitate mapping.

The table below contrasts the results from the 2006 and 2007 surveys with the results of data extraction from the 110 articles as described in the previous section. The categories are not directly comparable. Notes are made below to provide clarification for any indirect comparisons, and it is noted when the category is not applicable.

	Data Extraction Field	Current Work	Fung, Hole, and Srinivasan 2006	Chen et al. 2007, Table 9, overall
Data Source		110 research articles	Survey results from 1,427 UMLS users	Survey results from 70 users
EHR	Corpus	32%	81%[1]	14%
Information Retrieval	Research Category	46%	31%	27%

Artificial Intelligence/ Natural Language Processing	Research Category[2]	56%	21%	17%
UMLS Research	Screening Criteria[3]	20%[4]	*N/A*	19%
Terminology research	UMLS Use	6%	53%	53%
Terminology translation	Research Category[5]	*N/A*	*N/A*	20%
System development	*N/A*	*N/A*	*N/A*	4%
Decision support	*N/A*	*N/A*	*N/A*	4%
Mapping	UMLS Use	39%	35%	44%
Create Terminology/ Terminology Building	UMLS Use	8%	33%	*N/A*
Knowledge Acquisition	*N/A*	*N/A*	20%	*N/A*
Concept Discovery	*N/A*	*N/A*	19%	*N/A*
Terminology Service	UMLS Use	7%	14%	*N/A*
Access to Terminologies/ Extract Terminologies	UMLS Use	6%	13%	80%
Terminology Publishing	*N/A*	*N/A*	6%	*N/A*
Other	*N/A*	*N/A*	5%	*N/A*

Table 5: Comparing reported and extracted use of the UMLS from the current work and prior surveys

[1] "Far more users used UMLS to process clinical information (81%) than non-clinical information (45%)". (Fung, Hole, and Srinivasan, 2006).

[2]Natural Language Processing was included in the broader research category Artificial Intelligence

[3]Articles which researched the UMLS were excluded from the current work.

[4] 68 articles were excluded because the UMLS was not used as a tool.

[5] Translation was included in the broader research category of Language and Linguistics and might also be captured though the Yes/No question of multiple languages.

Similarities across datasets can be identified. For instance, the current work found that 39% of papers used the UMLS to facilitate mapping. Fung, Hole, and Srinivasan (2006) noted that 35% of users reported using the UMLS for mapping, and Chen, et al. (2007), found that 44% of users surveyed employed the UMLS for mapping. However, the data also reflects major differences. For instance, the

prior studies found slightly more than half (53%) of licensees use the UMLS for terminology research. The current work reflects a smaller percentage, 6 of the 110 articles reviewed.

Fung, Hole, and Srinivasan (2006) also reviewed which parts of the UMLS users reported using.

	Current work	Fung, Hole, and Srinivasan 2006
Metathesaurus	75%	94%
Semantic Network	35%	41%
Specialist Lexicon	9%	28%

Table 6: Comparing reported and extracted use of UMLS products from the current work and a prior survey

As in the 2006 study, the current work shows a high frequency of Metathesaurus use.

Limitations

The scholarly literature captures only a portion of the work using the UMLS and related tools. This review includes a significant number of conference proceedings. Additional uses might be captured in white papers or internal business reports, which might reflect difference uses or research types.

The broad search returned many articles which mentioned the UMLS or cited key UMLS papers in the introduction, background, related works, and future directions sections. Many, however, did not use the UMLS in the research described. Broad searches are common among scoping reviews, though they pose obstacles to screening.

A limitation of the data extraction portion of this review is that the research categories are neither granular nor exhaustive. The research categories were drawn from a small sample of articles and from MeSH terms and hierarchy. These choices created limits in assigning categories and capturing the breadth of research. The project team attempted to find a balance between feasibility and granularity.

Additional limitations include the time constraints and the lack of experience and knowledge of the Fellow. The iterative nature of the development of scoping review protocols required substantial time. Additionally, the Fellow did not have prior experience conducting scoping reviews. Overall, consulting with stakeholders and experts took about one month. Searching and cleaning the search results data took another month, as did trialing tools and methods. Developing criteria and screening the sample set spanned six weeks. Developing the data extraction form and extracting data from 110 articles took another two months. Data analysis was completed within two weeks. Considering the overlap of some of these stages, the process took about six months.

The project sponsors plan to use the protocol to continue the scoping review. Results will be combined with other knowledge sources to inform redevelopment and strategic planning for the UMLS.

Recommendations

The Fellow and project sponsors worked collaboratively using NIH Box accounts. This allowed for document-sharing, collaboration, and note-taking.

The Fellows recommends discussing proposed measures and protocols with others experienced in the research methodology as methods are developed. For instance, the team learned after completing the 25-article sets using the data extraction form that a smaller sample would have been reasonable.

Fortunately, the team learned in ample time that it was not necessary to have two independent reviewers per paper going forward.

Conclusion and Next Steps

The work presented in this report demonstrates a proof of concept. The protocol, screening criteria, data extraction form, and tool, have been revised and tested on a sample of the scholarly literature retrieved with our broad search strategy. The preliminary results provide some insights into the research being conducted using the UMLS and derived products. An extension of this work, using the methodology presented here and applied to the remaining citations could uncover additional trends and bolster those identified in the random sample. The results from such a work would inform the visioning of the UMLS.

This work, and the results from any extensions thereof, will be presented to stakeholders and experts at a planning workshop in spring 2020, in the year of UMLS's thirtieth anniversary.

References

Arksey, Hilary, and Lisa O'Malley. "Scoping studies: towards a methodological framework." International journal of social research methodology 8.1 (2005): 19-32.

Barriere, Caroline. "Consistency analysis of UMLS terminological and conceptual relations.". Proceedings of the 10th Terminology and Knowledge Engineering Conference: New Frontiers in the Constructive Symbiosis of Terminology and Knowledge Engineering, TKE 2012.

Chapman, Wendy W., et al. "Classifying free-text triage chief complaints into syndromic categories with natural language processing." Artificial intelligence in medicine 33.1 (2005): 31-40.

Chen, Yan et al. "Analysis of a study of the users, uses, and future agenda of the UMLS." Journal of the American Medical Informatics Association : JAMIA vol. 14,2 (2007): 221-31. doi:10.1197/jamia.M2202

Cheng, S. H., et al. "Using machine learning to advance synthesis and use of conservation and environmental evidence." Conservation biology: the journal of the Society for Conservation Biology 32.4 (2018): 762.

[software] Colandr. https://www.colandrapp.com/

Domínguez, Eladio, et al. "Model-driven development based transformation of stereotyped class diagrams to XML schemas in a healthcare context." International Conference on Conceptual Modeling. Springer, Berlin, Heidelberg, 2007.

[software] EndNote X9.

[software] Excel. Microsoft Office 16.

Fung KW, Hole WT, Srinivasan S. "Who is using the UMLS and how - insights from the UMLS user annual reports." AMIA Annu Symp Proc. 2006;2006:274–278.

Ochs, M., J. Casagrande, and R. Davuluri. *Biomedical informatics for cancer research*. Springer Science+ Business Media, 2010.

Peters, Micah DJ, et al. "Guidance for conducting systematic scoping reviews." International journal of evidence-based healthcare 13.3 (2015): 141-146.

[software] SurveyMonkey. Sample Size Calculator. surveymonkey.com/mp/sample-size-calculator/

Tricco, Andrea C et al. “A scoping review on the conduct and reporting of scoping reviews.” BMC medical research methodology vol. 16 15. 9 Feb. 2016, doi:10.1186/s12874-016-0116-4

Zhu, Qian, and Cui Tao. "Pharmacological class data representation in the Web Ontology Language (OWL)." 2014 IEEE International Conference on Big Data (Big Data). IEEE, 2014.

Acknowledgements

Betsy Humphreys for sharing her knowledge and insights. She was one of two preparers of the 1986-1996 Current Bibliographies in Medicine on the UMLS and was a foundational figure in the development of the UMLS. Humphreys attended meetings with the Fellow and project sponsors to discuss the project purpose and protocol and conducted data extraction on articles in the sample of 348.

Olivier Bodenreider for his insights into the UMLS.

Dina Demner-Fushman for her insights into potential search and sorting strategies.

Liz Amos and Anna Ripple for their support as project sponsors, weekly discussions, reviewing prowess, and Excel skills.

Kathel Dunn for her continuous support and guidance throughout the Associate Fellowship Program.

This research was supported in part by an appointment to the NLM Associate Fellowship Program sponsored by the National Library of Medicine and administered by the Oak Ridge Institute for Science and Education.

Appendices

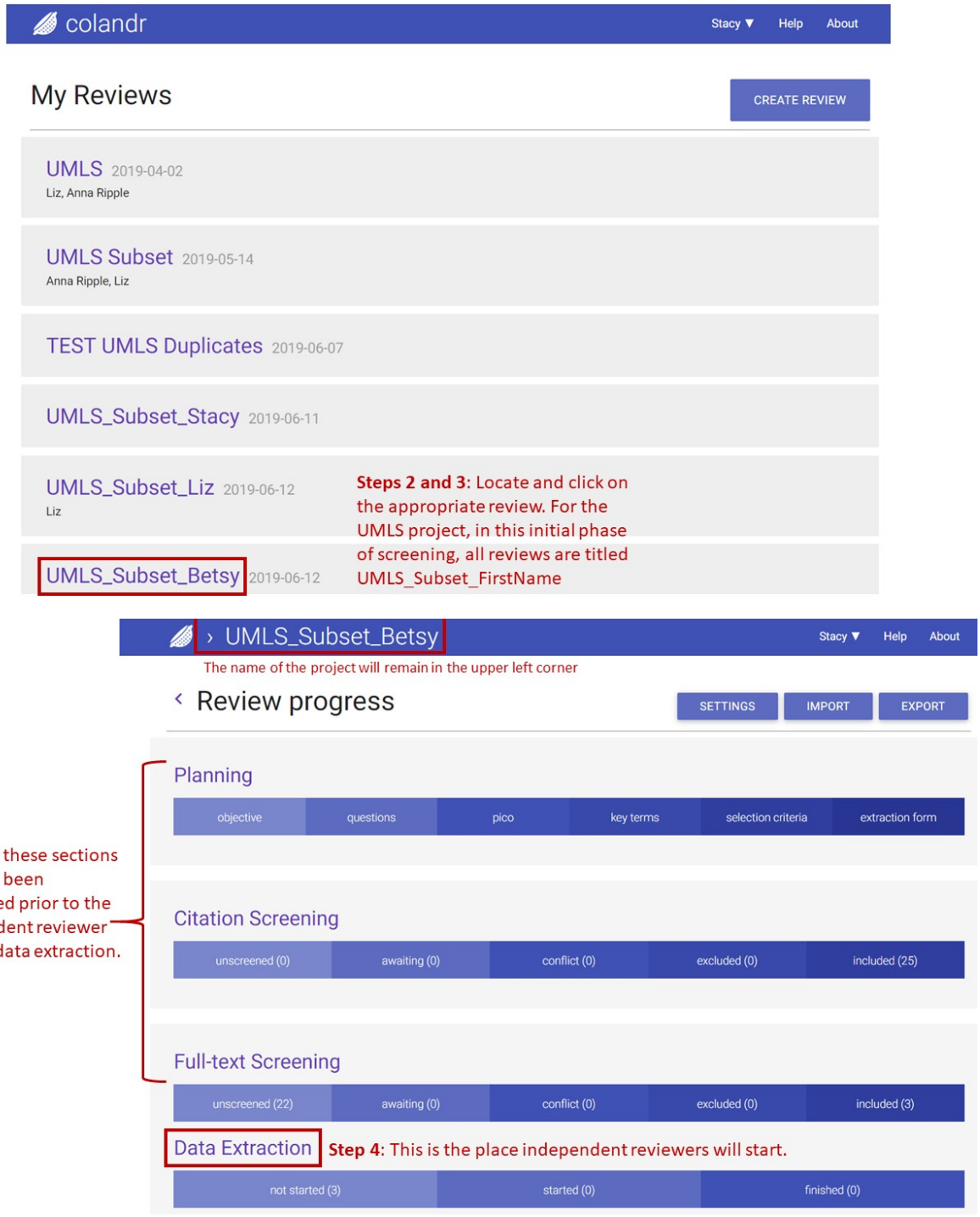
Appendix 1: Colandr Data Extraction Tutorial

Colandr Instructions for Independent Reviewers in 10 Easy Steps

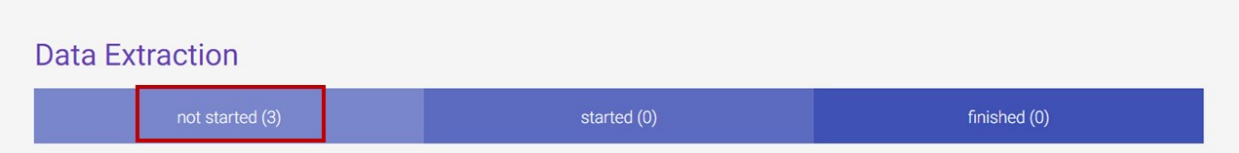
Following this summary, you will find a series of screenshots to guide you through the steps of data extraction using Colandr.

1. Sign up for a Colandr account. Let the project team know when this step is completed.
2. Log in to Colandr. On your dashboard, look for a review titled UMLS_Subset_FirstName.
3. Click the review title.
4. On the review page, scroll down to the last row, Data Extraction. This is where you will work.
5. Underneath Data Extraction, click Not Started.
6. On the data extraction page, you can click the article title to view the abstract. Beneath the article title, click the gray box Review Labels to access the data extraction form.
7. The screen will automatically open with the PDF window. Click Toggle PDF to close and open the PDF viewer.
8. Enter your answers to the data extraction questions:
    a) For free text, type directly or copy and paste. (Use quotation marks for direct quotes.). Click the save icon when completed.
    b) For controlled fields, select an item from the dropdown menu and click the plus sign. Repeat as many times as necessary for multiple selection fields.
    c) Be sure to refer to the data extraction form Word document or contact the Project Leads with any questions.
9. When you have completed the form, click Finalize.
    a) If extraction is not finalized, you may return to the article in the Started, or In Progress, section.

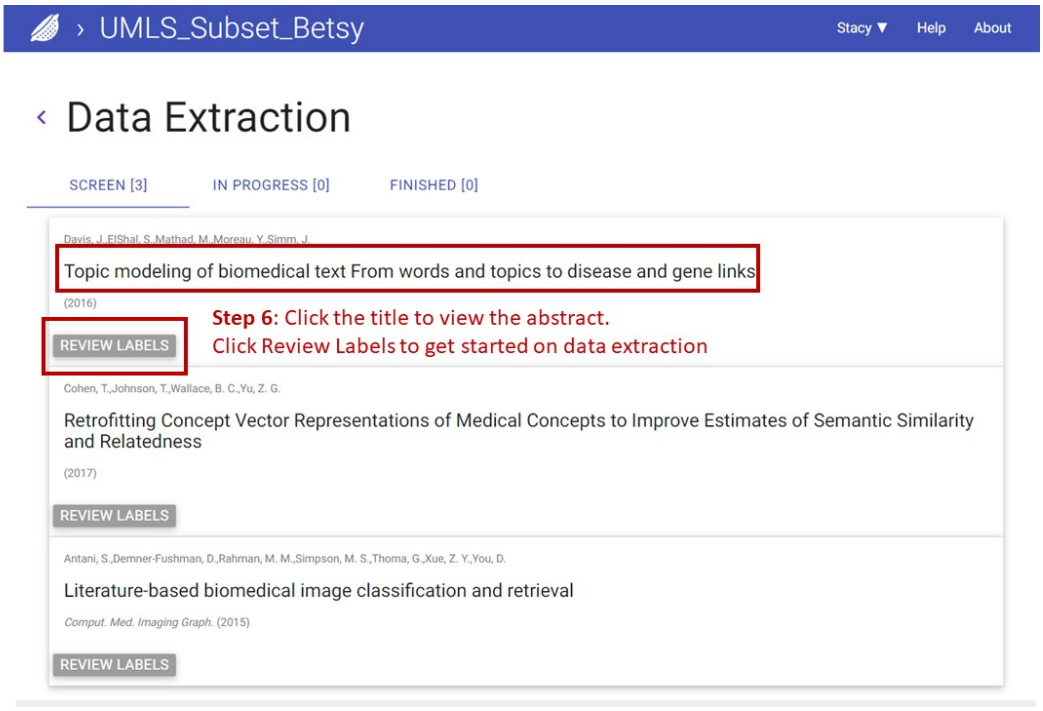
Citations, once finalized, are moved to Finished. If answers are to be modified, the form can be reopened.

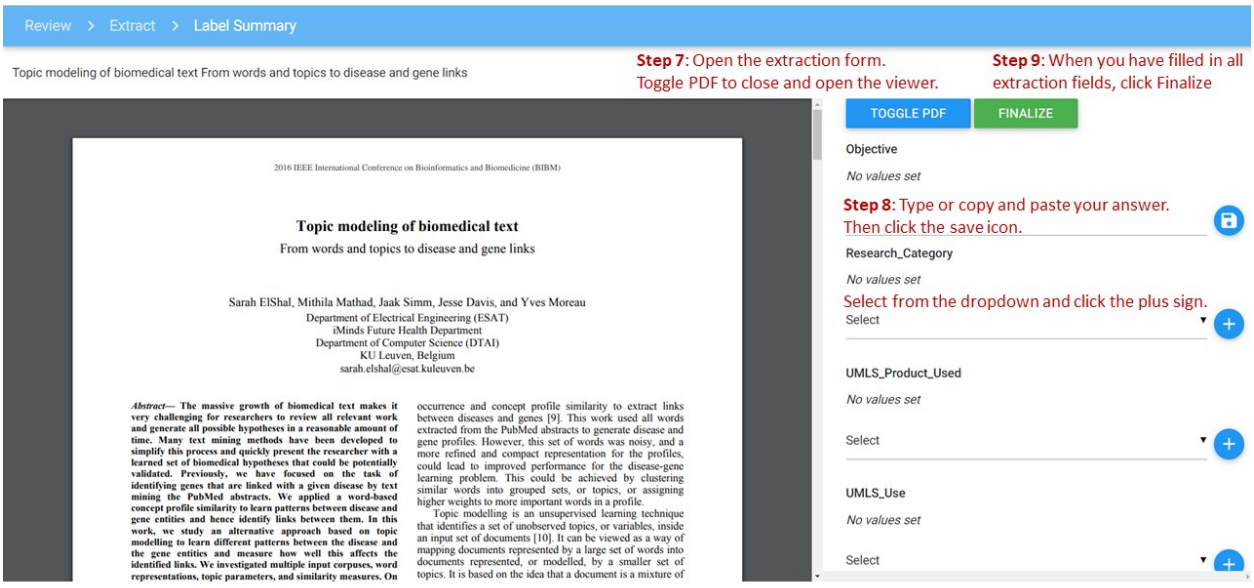

colandr
Stacy ▼
Help
About
My Reviews
CREATE REVIEW
UMLS 2019-04-02
Liz, Anna Ripple
UMLS Subset 2019-05-14
Anna Ripple, Liz
TEST UMLS Duplicates 2019-06-07
UMLS_Subset_Stacy 2019-06-11
UMLS_Subset_Liz 2019-06-12
Liz
UMLS_Subset_Betsy 2019-06-12
Steps 2 and 3: Locate and click on the appropriate review. For the UMLS project, in this initial phase of screening, all reviews are titled UMLS_Subset_FirstName
› UMLS_Subset_Betsy
Stacy ▼
Help
About
The name of the project will remain in the upper left corner
‹ Review progress
SETTINGS
IMPORT
EXPORT
Planning
objective
questions
pico
key terms
selection criteria
extraction form
Work for these sections will have been completed prior to the independent reviewer starting data extraction.
Citation Screening
unscreened (0)
awaiting (0)
conflict (0)
excluded (0)
included (25)
Full-text Screening
unscreened (22)
awaiting (0)
conflict (0)
excluded (0)
included (3)
Data Extraction
Step 4: This is the place independent reviewers will start.
not started (3)
started (0)
finished (0)

The UMLS Project Leads may continue to upload PDFs as independent reviewers begin extracting data. Citations for which PDFs have not been uploaded appear in “unscreened”.



Step 5: Independent reviewers start here.


› UMLS_Subset_Betsy
Stacy ▼
Help
About
‹ Data Extraction
SCREEN [3]
IN PROGRESS [0]
FINISHED [0]
Davis, J.,ElShal, S.,Mathad, M.,Moreau, Y.,Simm, J.
Topic modeling of biomedical text From words and topics to disease and gene links
(2016)
REVIEW LABELS
Step 6: Click the title to view the abstract.
Click Review Labels to get started on data extraction
Cohen, T.,Johnson, T.,Wallace, B. C.,Yu, Z. G.
Retrofitting Concept Vector Representations of Medical Concepts to Improve Estimates of Semantic Similarity and Relatedness
(2017)
REVIEW LABELS
Antani, S.,Demner-Fushman, D.,Rahman, M. M.,Simpson, M. S.,Thoma, G.,Xue, Z. Y.,You, D.
Literature-based biomedical image classification and retrieval
Comput. Med. Imaging Graph. (2015)
REVIEW LABELS


Review › Extract › Label Summary
Topic modeling of biomedical text From words and topics to disease and gene links
Step 7: Open the extraction form.
Toggle PDF to close and open the viewer.
Step 9: When you have filled in all extraction fields, click Finalize
TOGGLE PDF
FINALIZE
2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
Topic modeling of biomedical text
From words and topics to disease and gene links
Sarah ElShal, Mithila Mathad, Jaak Simm, Jesse Davis, and Yves Moreau
Department of Electrical Engineering (ESAT)
iMinds Future Health Department
Department of Computer Science (DTAI)
KU Leuven, Belgium
sarah.elshal@esat.kuleuven.be
Abstract— The massive growth of biomedical text makes it very challenging for researchers to review all relevant work and generate all possible hypotheses in a reasonable amount of time. Many text mining methods have been developed to simplify this process and quickly present the researcher with a learned set of biomedical hypotheses that could be potentially validated. Previously, we have focused on the task of identifying genes that are linked with a given disease by text mining the PubMed abstracts. We applied a word-based concept profile similarity to learn patterns between disease and gene entities and hence identify links between them. In this work, we study an alternative approach based on topic modelling to learn different patterns between the disease and the gene entities and measure how well this affects the identified links. We investigated multiple input corpuses, word representations, topic parameters, and similarity measures. On
occurrence and concept profile similarity to extract links between diseases and genes [9]. This work used all words extracted from the PubMed abstracts to generate disease and gene profiles. However, this set of words was noisy, and a more refined and compact representation for the profiles, could lead to improved performance for the disease-gene learning problem. This could be achieved by clustering similar words into grouped sets, or topics, or assigning higher weights to more important words in a profile.
Topic modelling is an unsupervised learning technique that identifies a set of unobserved topics, or variables, inside an input set of documents [10]. It can be viewed as a way of mapping documents represented by a large set of words into documents represented, or modelled, by a smaller set of topics. It is based on the idea that a document is a mixture of
Objective
No values set
Step 8: Type or copy and paste your answer.
Then click the save icon.
Research_Category
No values set
Select from the dropdown and click the plus sign.
Select
UMLS_Product_Used
No values set
Select
UMLS_Use
No values set
Select

Step 10: Once the reviewer completes the form and clicks Finalize, the citation moves to Finished.

Appendix 2: Data Extraction Form

Data Extraction Form for the UMLS Research Review

This is the protocol and form to be used to extract information from the full text of the screened corpus of articles. The questions in this form will be incorporated into the web-based tool Colandr for ease-of-use. This document is to be used for reference and clarification. If you have any questions, please contact stacy.brody@nih.gov.

For select sections, as indicated, text should be copied and pasted directly from the paper. Reviewers must use quotation marks. When the text is long, ellipses are permitted.

For items that cannot be located or identified, the reviewer should note in that field, or in the comments field at the end, the item that could not be found. Blanks can be misinterpreted. Please indicate unsure or not applicable in fields, as indicated.

Notes about Using Colandr

- The timeout period is relatively short. You may frequently be prompted to sign back in. Be sure to save your work.
- Once you finish entering answers to the form, click Finalize, the green button towards the top of the screen.

Extraction Form

Field	Instructions	Example(s)
Objective	Scan the paper for a sentence or phrase, i.e. "The purpose of this research is to…" "The objective of this paper is…" "In this study, we…" If the paper has a structured abstract, it may include an objective. The objective should be copied directly from the article.	"Objective: The aim is to investigate the use of information retrieval techniques in recommending patient education materials for diabetic questions of patients" (Zeng et al. 2017) "The objective of this study is to understand the types of health information (health topics) that users search online for Cardiovascular Diseases, by performing categorization of health search queries (from Mayoclinic.com) using UMLS MetaMap based on UMLS concepts and semantic types" (Jadhav et al. 2014)
Research Category	Select the most appropriate category from the dropdown menu. Multiple selections are allowed.	Select from: 1. Information Storage and Retrieval

	See descriptions and scope notes following this table.	2. Information Services 3. Language and Linguistics 4. Artificial Intelligence 5. Other
UMLS Product Used	Select from the dropdown list the UMLS product(s) used in the research, as indicated in the methods section. Multiple selections are allowed.	Select from: • MetaMap • Metathesaurus • Semantic Network (semantic types) • MetamorphoSys • Specialist Lexicon • SemRep • Not indicated, unable to infer
UMLS Use/ Purpose/ Application	Select the purpose/use/application from the checkbox. Multiple selections are permitted.	Select from: • Link terms and codes • Process texts to extract concepts, relationships, or knowledge • Facilitate mapping between terminologies • Extract specific terminologies from the Metathesaurus • Create and maintain a local terminology • Develop a terminology service • Research terminologies or ontologies
Corpus	In this field, indicate the corpus of text or data used in the research, as described in the methods or materials. Multiple selections are permitted.	Select from: • Electronic health records • Scientific literature (make note in comments section if only specific parts are used) • Consumer health information/patient education materials • Social media • User-generated content, including questions and online health forums • Data, unspecified, i.e. drug data, protein data, genetic data, etc. • Other
Name of Tool Produced	Enter the name of any new tool produced from the research. If no tool is produced, enter N/A	HMPAS: http://fcode.kaist.ac.kr/hmpas Chinese-language knowledgebase

Other NLM Products Used	Enter the name(s) of the other NLM product(s) used in the research. If multiple, separate names by a comma. If none, enter N/A. For complete list, see eresources.nlm.nih.gov/nlm_eresources/	ClinicalTrials.gov, PubMed/MEDLINE
How the UMLS is Cited	Check all those that apply.	Select from: • URL • Citation • Both • Neither
Multiple languages or translations	Select yes if a non-English language is used and/or if the goal is to translate into or build a resource in a non-English language.	
System/tool operational	Is the tool or system in production/operational (as opposed to a research prototype)? Check yes or no or unsure	Select from: • Yes • No • Unsure • N/A
Reviewer	Enter your first and last initials	SB
Comments/Other	Enter any comments on the paper or clarifications to entries in the data extraction form.	
Corpus Comments	Enter comments on the corpus used in the research. You may copy and paste directly from the article, if necessary. If so, please use quotation marks.	For the corpus, the authors extracted and used data tables from the full text of scientific articles.

Research Type Categories (drawn from MeSH)

Information Storage and Retrieval

Including "organized activities related to the storage, location, search, and retrieval of information". Included in this category are the following:

Information Retrieval: methods for extracting concepts from biomedical texts and retrieving information, including data extraction; systems or applications that utilize information retrieval methods to retrieve texts and concepts, inclusive of search engines

Data Mining: "use of tools to sort, organize, examine, and combine large sets of information"

Information Services

Organized services to provide information on any questions an individual might have using databases and other sources. (From Random House Unabridged Dictionary, 2d ed)

Classification: "the systematic arrangement of entities in any field into classes based on common characteristics such as properties, morphology, subject matter, etc."

Knowledge Bases: the building or development of knowledge bases, defined as "collections of facts, assumptions, beliefs, and heuristics that are used in combination with databases to achieve desired results, such as a diagnosis, an interpretation, or a solution to a problem (From McGraw Hill Dictionary of Scientific and Technical Terms, 6th ed)"

Vocabulary, Controlled: "A specified list of terms with a fixed and unalterable meaning, and from which a selection is made when CATALOGING; ABSTRACTING AND INDEXING; or searching BOOKS; JOURNALS AS TOPIC; and other documents. The control is intended to avoid the scattering of related subjects under different headings (SUBJECT HEADINGS). The list may be altered or extended only by the publisher or issuing agency. (From Harrod's Librarians' Glossary, 7th ed, p163)"

Language and Linguistics

Translation the study focuses on translating UMLS terminologies into other languages.

Semantics the study focused on meanings and the relationships of meanings

Artificial Intelligence

"Theory and development of COMPUTER SYSTEMS which perform tasks that normally require human intelligence. Such tasks may include speech recognition, LEARNING; VISUAL PERCEPTION; MATHEMATICAL COMPUTING; reasoning, PROBLEM SOLVING, DECISION-MAKING, and translation of language."

Natural Language Processing: Studies that employ natural language processing techniques, methods, and algorithms. "Computer processing of a language with rules that reflect and describe current usage rather than prescribed usage."

> Note: NLP includes word sense disambiguation. Many studies conducting this type of work will note this in the title, abstract, or keywords.

Machine Learning: machine learning, including supervised and unsupervised methods. "A type of ARTIFICIAL INTELLIGENCE that enable COMPUTERS to independently initiate and execute LEARNING when exposed to new data."

Appendix 3: Revised Data Extraction Form

Data Extraction Form for the UMLS Research Review

This is the protocol and form to be used to extract information from the full text of the screened corpus of articles. The questions in this form will be incorporated into the web-based tool Colandr for ease-of-use. This document is to be used for reference and clarification. If you have any questions, please contact stacy.brody@nih.gov.

For select sections, as indicated, text should be copied and pasted directly from the paper. Reviewers must use quotation marks. When the text is long, ellipses are permitted.

For items that cannot be located or identified, the reviewer should note in that field, or in the comments field at the end, the item that could not be found. Blanks can be misinterpreted. Please indicate unsure or not applicable in fields, as indicated.

Notes about Using Colandr

- The timeout period is relatively short. You may be prompted to sign back in. Save your work.
- For each answer, click the save icon to enter that answer.
- Answer all fields. If the question does not apply, type or select N/A. Blanks are ambiguous.
- Once you finish entering answers, click Finalize, the green button towards the top of the screen.

Extraction Form

Field	Instructions	Example(s)
Objective	Scan the paper for a sentence or phrase, i.e. "The purpose of this research is to…" "The objective of this paper is…" "In this study, we…" If the paper has a structured abstract, it may include an objective. The objective should be copied directly from the article.	"Objective: The aim is to investigate the use of information retrieval techniques in recommending patient education materials for diabetic questions of patients" (Zeng et al. 2017) "The objective of this study is to understand the types of health information (health topics) that users search online for Cardiovascular Diseases, by performing categorization of health search queries (from Mayoclinic.com) using UMLS MetaMap based on UMLS concepts and semantic types" (Jadhav et al. 2014)
Research Category	Select the most appropriate category from the dropdown menu. Multiple selections are allowed.	Select from: 6. Information Storage and Retrieval

	Select the narrowest level that comprehensively describes the research. See descriptions and scope notes following this table.	a. Information retrieval b. Data mining 7. Information Services a. Classification b. Knowledge Bases c. Vocabulary, Controlled 8. Language and Linguistics 9. Artificial Intelligence a. Natural Language Processing b. Machine Learning 10. Other
UMLS Product Used	Select from the dropdown list the UMLS product(s) used in the research, as indicated in the methods section. Multiple selections are allowed.	Select from: • MetaMap • Metathesaurus • Semantic Network (semantic types) • MetamorphoSys • Specialist Lexicon • SemRep • Not indicated, unable to infer
UMLS Use/ Purpose/ Application	Select the purpose/use/application from the checkbox. Multiple selections are permitted.	Select from: • Link terms and codes, facilitate mapping between terminologies • Process texts to extract concepts, relationships, or knowledge • Extract specific terminologies from the Metathesaurus • Create and maintain a local terminology • Develop a terminology service • Research terminologies or ontologies
Corpus	In this field, indicate the corpus of text or data used in the research, as described in the methods or materials. Multiple selections are permitted.	Select from: • Electronic health records (any piece) • Scientific literature (make note in comments section if only specific parts are used) • Consumer health information/patient education materials • Social media • User-generated content, including questions and online health forums

		• Data, unspecified, i.e. drug data, protein data, genetic data, etc. • Other
Name of Tool Produced	Enter the name of any new tool produced from the research. If no tool is produced, enter N/A	HMPAS: http://fcode.kaist.ac.kr/hmpas Chinese-language knowledgebase
Other NLM Products Used	Enter the name(s) of the other NLM product(s) used in the research. If multiple, separate names by a comma. If none, enter N/A. For complete list, see eresources.nlm.nih.gov/nlm_eresources/	ClinicalTrials.gov, PubMed/MEDLINE
How the UMLS is Cited	Check all those that apply.	Select from: • URL • Citation • Both • Neither
Multiple languages or translations	Select yes if a non-English language is used and/or if the goal is to translate into or build a resource in a non-English language.	
System/tool operational (at time of writing)	Is the tool or system in production/operational (as opposed to a research prototype)? Check yes or no or unsure	Select from: • Yes • No • Unsure • N/A
Reviewer	Enter your first and last initials	SB
Comments/Other	Enter any comments on the paper or clarifications to entries in the data extraction form.	
Corpus Comments	Enter comments on the corpus used in the research. You may copy and paste directly from the article, if necessary. If so, please use quotation marks.	For the corpus, the authors extracted and used data tables from the full text of scientific articles.

Research Type Categories (drawn from MeSH)

Information Storage and Retrieval

Including "organized activities related to the storage, location, search, and retrieval of information". Included in this category are the following:

Information Retrieval: methods for extracting concepts from biomedical texts and retrieving information, including data extraction; systems or applications that utilize information retrieval methods to retrieve texts and concepts, inclusive of search engines

Data Mining: "use of tools to sort, organize, examine, and combine large sets of information"

Information Services

Organized services to provide information on any questions an individual might have using databases and other sources. (From Random House Unabridged Dictionary, 2d ed)

Classification: "the systematic arrangement of entities in any field into classes based on common characteristics such as properties, morphology, subject matter, etc."

Knowledge Bases: the building or development of knowledge bases, defined as "collections of facts, assumptions, beliefs, and heuristics that are used in combination with databases to achieve desired results, such as a diagnosis, an interpretation, or a solution to a problem (From McGraw Hill Dictionary of Scientific and Technical Terms, 6th ed)"

Vocabulary, Controlled: "A specified list of terms with a fixed and unalterable meaning, and from which a selection is made when CATALOGING; ABSTRACTING AND INDEXING; or searching BOOKS; JOURNALS AS TOPIC; and other documents. The control is intended to avoid the scattering of related subjects under different headings (SUBJECT HEADINGS). The list may be altered or extended only by the publisher or issuing agency. (From Harrod's Librarians' Glossary, 7th ed, p163)"

Language and Linguistics

Translation the study focuses on translating UMLS terminologies into other languages.

Semantics the study focused on meanings and the relationships of meanings

Artificial Intelligence

"Theory and development of COMPUTER SYSTEMS which perform tasks that normally require human intelligence. Such tasks may include speech recognition, LEARNING; VISUAL PERCEPTION; MATHEMATICAL COMPUTING; reasoning, PROBLEM SOLVING, DECISION-MAKING, and translation of language."

Natural Language Processing: Studies that employ natural language processing techniques, methods, and algorithms. "Computer processing of a language with rules that reflect and describe current usage rather than prescribed usage."

Note: NLP includes word sense disambiguation. Many studies conducting this type of work will note this in the title, abstract, or keywords.

Machine Learning: machine learning, including supervised and unsupervised methods. "A type of ARTIFICIAL INTELLIGENCE that enable COMPUTERS to independently initiate and execute LEARNING when exposed to new data."

UMLS Use/Application

UMLS uses are drawn from the list on the UMLS homepage: nlm.nih.gov/research/umls/

- Link terms and codes between your doctor, your pharmacy, and your insurance company AND Facilitate mapping between terminologies
    - The UMLS product is used to process strings, phrases, or individual words and link these to codes; or
    - to map entire terminologies to one another
- Process texts to extract concepts, relationships, or knowledge
    - The UMLS product is used to process entire "chunks" of text, larger than phrases or words.
- Extract specific terminologies from the Metathesaurus
    - The research describes pulling an entire terminology or terminologies from the UMLS Metathesaurus.
- Create and maintain a local terminology
    - The research describes using a UMLS product in the development or maintenance of a novel terminology for local use.
    - For instance, developing a consumer health vocabulary
- Develop a terminology service
    - The article describes "software methods… that allow other systems to determine the locally acceptable term to use for a given purpose" (Shortliffe 2006)
    - A terminology service is "a service that lets healthcare applications make use of codes and value sets without having to become experts in the fine details of code system, value set and concept map resources, and the underlying code systems and terminological principles" (hl7.org/fhir/terminology-service.html)
- Research terminologies or ontologies
    - The paper describes research about a terminology or ontology or compares multiple terminologies or ontologies.

References

Shortliffe, Edward H. *Biomedical informatics*. Ed. James J. Cimino. Springer Science+ Business Media, LLC, 2006.