Published in final edited form as:
On Move Meaningful Internet Syst. 2016 October ; 10033: 699–708. doi:10.1007/978-3-319-48472-3_43.

An Ontology-Enabled Natural Language Processing Pipeline for Provenance Metadata Extraction from Biomedical Text (Short Paper)

Joshua Valdez[1], **Michael Rueschman**[2], **Matthew Kim**[2], **Susan Redline**[2], and **Satya S. Sahoo**[1]

[1]Division of Medical Informatics and Electrical Engineering and Computer Science Department, Case Western Reserve University, Cleveland, OH, USA

[2]Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard University, Boston, MA, USA

Abstract

Extraction of structured information from biomedical literature is a complex and challenging problem due to the complexity of biomedical domain and lack of appropriate natural language processing (NLP) techniques. High quality domain ontologies model both data and metadata information at a fine level of granularity, which can be effectively used to accurately extract structured information from biomedical text. Extraction of provenance metadata, which describes the history or source of information, from published articles is an important task to support scientific reproducibility. Reproducibility of results reported by previous research studies is a foundational component of scientific advancement. This is highlighted by the recent initiative by the US National Institutes of Health called "Principles of Rigor and Reproducibility". In this paper, we describe an effective approach to extract provenance metadata from published biomedical research literature using an ontology-enabled NLP platform as part of the Provenance for Clinical and Healthcare Research (ProvCaRe). The ProvCaRe-NLP tool extends the clinical Text Analysis and Knowledge Extraction System (cTAKES) platform using both provenance and biomedical domain ontologies. We demonstrate the effectiveness of ProvCaRe-NLP tool using a corpus of 20 peer-reviewed publications. The results of our evaluation demonstrate that the ProvCaRe-NLP tool has significantly higher recall in extracting provenance metadata as compared to existing NLP pipelines such as MetaMap.

Keywords

Ontology-based Natural Language Processing; Provenance Metadata; Scientific Reproducibility; Named Entity Recognition

1. Introduction

Provenance metadata describing the contextual information of scientific experiments can facilitate the evaluation of the soundness and rigor of research studies [1]. In addition, provenance metadata can support scientific reproducibility, which is one of the core foundations of scientific advancement [2]. The importance of scientific rigor and reproducibility is highlighted by the recent concerted efforts led the US National Institutes of Health (NIH) called “Rigor and Reproducibility” to ensure that results from research studies can be reproduced by other researchers in the community [2, 3]. As part of a NIH funded Big Data to Knowledge (BD2K) project on data provenance, we have defined the Provenance for Clinical and Healthcare Research (ProvCaRe) framework that extracts, analyzes, and stores provenance of research studies to support scientific reproducibility.

The ProvCaRe framework has identified three core elements of provenance that are necessary for scientific reproducibility, namely: (1) **Study Method** that captures the design of the study, the method used to collect, and method used for data analysis; (2) **Study Data** that represents the different data elements and their respective categories collected and analyzed in a research study; and (3) **Study Tools** that captures provenance of the different instruments and their parameters used in a research study. The ProvCaRe framework is being developed in close collaboration with biomedical domain experts involved with the National Sleep Research Resource (NSRR), which is the largest data repository of publicly available research sleep medicine studies in the world, with a goal of aggregating more than 50,000 studies collected from 36,000 participants [4]. Using this dataset will ensure that the ProvCaRe framework reflects the practical requirements of biomedical domain and it is also scalable with increasing volume and variety of biomedical data. The NSRR project has publicly released data from more than six sleep research studies and we use peer-reviewed publications associated with these studies to extract, model, and analyze provenance information associated with these research studies.

Peer-reviewed literature is one of the largest sources of domain knowledge across all disciplines in biomedicine and healthcare. Results from research studies are often published in peer-reviewed articles and are used by researchers to replicate the results together with the relevant experiment data. Therefore, provenance metadata can be extracted from biomedical literature and analyzed for supporting scientific reproducibility. However, extracting structured information from biomedical literature is difficult and it is the focus of extensive research in computer science as well as in biomedical informatics [5–7]. The key challenges in extracting structured data from biomedical text are the use of domain-specific terminology, complex sentence structure including use of negation and modifiers that significantly modify the meaning of terms or phrases in a sentence, and identification of complex relations that link terms across one or more sentences [5].

2. Background and Related Work

The Medical Language Extraction and Encoding (MedLEE) system is one of the earliest biomedical text processing tools that was developed to extract structured information from clinical unstructured text in the context of tuberculosis [8]. The US National Library of Medicine (NLM) developed the SPECIALIST system to detect anatomical information in coronary catheterization reports using the well-known Unified Medical Language System (UMLS) as a reference knowledge model [9]. The NLM also developed the widely used MetaMap tool that mapped terms in biomedical text with concepts modeled in UMLS Metathesaurus [10]. The automated identification of terms in biomedical text or NER is key task in biomedical NLP and different approaches have been applied for effective NER [11]. A high quality knowledge model such as an ontology plays a central role in NER, for example the MetaMap tool uses the UMLS Metathesaurus as a reference knowledge model. The UMLS integrates more than 60 biomedical vocabularies and the Metathesaurus is an important component of UMLS that stores biomedical concepts and relations linking the concepts together [12].

Similar to the MetaMap tool, the National Center for Biomedical Ontologies (NCBO) Open Biomedical Annotator leverages a large set of open source biomedical ontologies listed in NCBO to annotate text with ontology concepts [13]. In addition to MetaMap and NCBO Annotator, the clinical Text Analysis and Knowledge Extraction System (cTAKES) is a well known multi-component pipeline developed at the Mayo Clinic, which uses both rule-based and machine learning approaches for biomedical text processing [14]. The cTAKES platform builds on the Apache Unstructured Information Management Architecture (UIMA) framework [15] and openNLP toolkit [16]. The primary advantage of the cTAKES pipeline is an extensible architecture that allows developers to implement customized versions that focus on specific application domains. Therefore, cTAKES is a suitable platform to extend and develop the ProvCaRe-NLP tool that focuses on extraction of provenance information from biomedical text. It is important to note that none of these NLP tools focus on extraction of provenance information. We describe the details of this ProvCaRe-NLP pipeline in the following section.

3. Extracting Provenance Metadata from Biomedical Text Using Provenance Ontology

The ProvCaRe-NLP pipeline is a flexible system that aims to combine multiple ontologies, and text processing tools for extracting provenance information. In particular, we leverage the extensive coverage of existing text processing tools such as MetaMap and NCBO Annotator together with provenance-specific focus of ProvCaRe-NLP. The two primary objectives of the ProvCaRe NLP pipeline are: (1) **Provenance-specific NER** from biomedical text that correspond to the three components of the ProvCaRe framework; and (2) **Map extracted provenance terms to ontology classes**. We plan to use terms and relations in the next phase of the ProvCaRe project to generate provenance RDF graphs corresponding to provenance of research studies described in published articles. Figure 1 illustrates an overview of the ProvCaRe-NLP pipeline. The ProvCaRe-NLP tool extends the cTAKES pipeline and uses the following modules to process biomedical text:

1. **Sentence Detection**: Using the cTAKES sentence boundary detector module, which is based on the OpenNLP maximum entropy sentence detector tool, the ProvCaRe-NLP uses the occurrence of punctuation marks to detect the end of a sentence. The punctuation marks include period, question mark, and exclamation mark [14].

2. **Tokenization**: The ProvCaRe-NLP uses the cTAKES tokenizer to parse tokens from a sentence using space and punctuation marks for words and acronyms as well as dates as special case [14].

3. **Part-of-speech Tagging**: The ProvCaRe-NLP part of speech (POS) tagging module uses normalization rules to map variations in lexical properties of tokens as part of the preprocessing step, which builds on the cTAKES POS tagging module.
4. **Shallow Parsing**: The noun phrases detected from the previous steps are parsed by the cTAKES shallow parser module.
5. **Provenance Named Entity Recognition (NER)**: The output from the shallow parsing step is annotated using ontology classes that correspond to the terms defined in the ProvCaRe framework. As we discussed earlier, the provenance-specific NER is a key task of the ProvCaRe-NLP pipeline and requires use of multiple ontologies as reference terminology.

In the following section, we describe the details of the ProvCaRe framework and its use to characterize provenance metadata described in biomedical article.

3.1 The ProvCaRe Framework for Representing Provenance Metadata Associated with Research Studies

In collaboration with sleep medicine researchers working in the NSRR project, we identified the requirement set of provenance metadata elements, which can be classified as subcategories of the three core ProvCaRe components:

1. **Study method and its subcategories**: The various aspects of a scientific experiment, including how the data are collected as well as processed; how the data are analyzed; the method used to perform interventions in a biomedical research study (e.g., use of a new clinical drug); and the protocol used in the study (e.g., observational study or interventional study) are modeled as part of the Study method. The provenance information modeled in the Study method category also consists of data processing techniques used before and during data analysis.
2. **Study data and its subcategories**: A biomedical research study involves collection and analysis of a variety of data, for example the participants in a research study are often selected based on a set of *inclusion* and *exclusion criteria*. Similarly, the time values associated with the study data are also important provenance information.
3. **Study tool and its subcategories**: The various categories of instruments, both hardware and software tools, used in a research study represents essential provenance information, which is required for scientific reproducibility. For example, multi-contact intracranial electrodes used to record EEG data in epilepsy patients. The Study tool category also includes the software data analysis tools used in a research study, including statistical analysis package such as R or SAS.

Table 1 shows the provenance information extracted from a research study that evaluated the effects of providing oxygen at night and continuous positive airway pressure (CPAP) in patients with obstructive sleep apnea on measures of cardiovascular risk [17].

3.2 The ProvCaRe Ontology: Extending the W3C PROV Ontology

The W3C PROV specifications define a set of common classes and relations to model provenance information in various application domains [18]. As part of the ProvCaRe project, we have extended the PROV-O to model provenance information required to support scientific reproducibility in biomedical domain. The ProvCaRe ontology models the three core provenance metadata elements together with subcategories of provenance information described in the previous section as subclasses and subproperties of PROV-O. The `provcare:StudyMethod`[1] is modeled as a subclass of PROV-O `Activity` class. The `provcare:StudyData` is modeled as subclass of the PROV-O `Entity` class. The `provcare:StudyTool` class is modeled as subclass of the PROV-O `Agent` class and models different types of instruments based on their function (e.g., `provcare:DataAnalysisInstrument`) and their modality of recording data (e.g., `provcare:ImagingInstrument`). A key design feature of the ProvCaRe ontology is that it can be used together with existing biomedical ontologies, for example SNOMED CT and GO, which model the biomedical domain are a far greater level of granularity and coverage. We have developed a post coordination-based ontology modeling approach that allows the ProvCaRe ontology classes to be combined with existing biomedical ontology classes in a postcoordinated expression.

4. Results

A simple search for the keywords “sleep research” in the National Center for Biotechnology Information (NCBI) PubMed tool returns more than 69,000 articles, which demonstrate the large amount of research studies conducted in the sleep medicine domain. Therefore, the evaluation of the ProvCaRe-NLP pipeline using peer-reviewed articles in the sleep medicine domain is a representative approach for the wider biomedical research domain.

3.2 Comparison of MetaMap, NCBO Annotator and ProvCaRe-NLP tool

As described earlier, the MetaMap tools uses the UMLS as a reference knowledge model for the NER task, while the NCBO Annotator tool uses the biomedical ontologies listed at NCBO for the NER task. During our preliminary evaluation, we used the research study reported by Gottlieb et al. [17] as example and processed the full article using MetaMap and NCBO Annotator. A review of the results showed that although both MetaMap and NCBO Annotator successfully identify many domain-specific terms in the two articles, they did not detect provenance-specific information required to support scientific reproducibility. The primary reason for this limitation of these two tools is the lack of a provenance-focused ontology in both UMLS and the NCBO. Following our preliminary evaluation, we use the full corpus of 20 articles to evaluate the recall of MetaMap, NCBO Annotator, MetaMap and NCBO Annotator together, and finally the ProvCaRe-NLP together with MetaMap as well as NCBO Annotator.

[1] `Courier New` font is used to represent ontology classes. The provcare namespace refers to http://www.case.edu/ProvCaRe/provcare. The ProvCaRe ontology is available at: https://sites.google.com/a/case.edu/bmhinformaticsgroup/research/provcare/

Table 2 shows the comparative evaluation of three tools using "recall" as a measure of comparison. The recall measure M is computed using the total number of provenance-related terms extracted by each NLP pipeline, $P_{MetaMap}$, P_{NCBO}, and $P_{ProvCaRe}$, and the provenance terms in the gold standard $P_{goldstandard}$, which can be represented as

$$M=(P_{MetaMap}, P_{NCBO}, P_{ProvCaRe})/P_{goldstandard}$$

The results in Table 2 shows the effectiveness of the ProvCaRe-NLP pipeline, which aggregates entities that are identified using UMLS, the ontologies listed at NCBO, and the ProvCaRe ontology. The ProvCaRe-NLP achieves nearly 100% recall in terms of extracting provenance information from biomedical text (using the manually extracted provenance metadata by domain expert as "gold standard"). During our initial evaluations, we noticed that the primary source of provenance information in published articles are three sections: the *Abstract* of the article, the *Method* section that describes the details of the research study, and the *Results* section that describes the data analysis methods used in a research study (e.g., statistical model). Table 2 lists the recall measure for processing the complete article and the three specific sections. The results show that there is minor improvement in the recall measure if the NLP pipeline processes the full article as compared to only three sections of the article.

5. Conclusions and Future Work

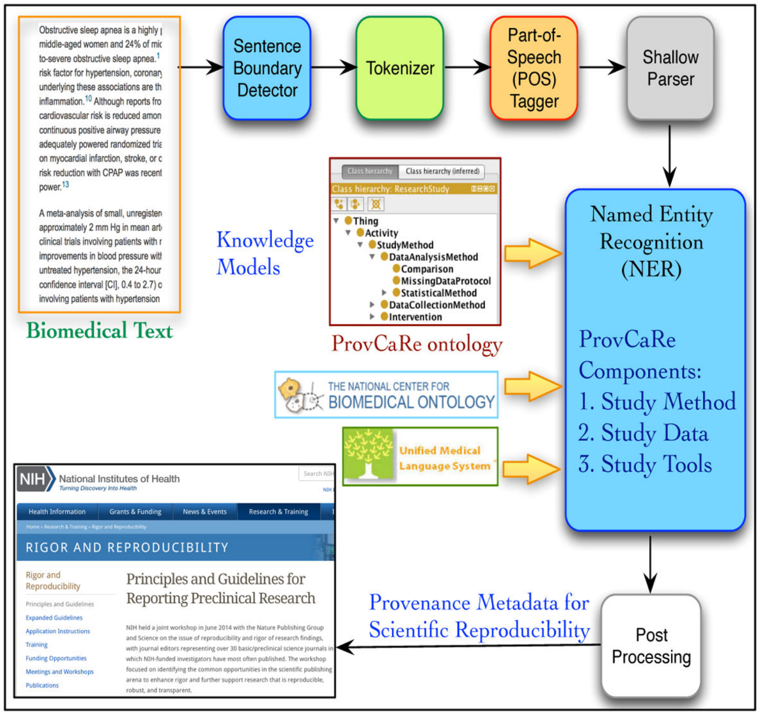
The increasing significance of provenance metadata in supporting transparency and scientific reproducibility can be assisted by the development of automated techniques to extract provenance from biomedical literature. Therefore, we have developed a NLP pipeline that leverages both provenance and biomedical domain ontologies to accurately identify and extract provenance metadata as part of the ProvCaRe project. The ProvCaRe-NLP pipeline combines the extensive coverage of the biomedical domain by existing NLP tools such as MetaMap and the NCBO Annotator together with focused provenance NER task to accurately identify provenance terms corresponding to the three components of the ProvCaRe framework. We demonstrate the effectiveness of the ProvCaRe-NLP pipeline using 20 peer-reviewed articles as text corpus. As part of our ongoing and future work, we are extending the functionalities of the ProvCaRe-NLP pipeline to identify and extract relationships linking the provenance terms by leveraging properties modeled in the ProvCaRe ontology. We aim to use the ProvCaRe-NLP pipeline to populate a provenance knowledgebase for use by biomedical researchers for querying and exploration.

Acknowledgments

This work is supported in part by the NIH-NIBIB Big Data to Knowledge (BD2K) 1U01EB020955 and NIH-NHLBI R24 HL114473 grants.

References

1. Sahoo, SS., Valdez, J., Rueschman, M. Scientific Reproducibility in Biomedical Research: Provenance Metadata Ontology for Semantic Annotation of Study Description. American Medical Informatics Association (AMIA) Annual Symposium; Chicago. 2016;
2. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature. 2014; 505:612–613. [PubMed: 24482835]
3. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitz AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley RT 3rd, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012; 490:187–91. [PubMed: 23060188]
4. Dean DA, Goldberger AL, Mueller R, Kim M, Rueschman M, Mobley D, Sahoo SS, Jayapandian CP, Cui L, Morrical MG, Surovec S, Zhang GQ, Redline S. Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource. SLEEP. 2016; 39:1151–1164. [PubMed: 27070134]
5. Meystre S, Savova G, Kipper-Schuler K, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. IMIA Year Book of Medical Informatics. 2008; 47:128–44.
6. Crowley RS, Castine M, Mitchell KJ, Chavan G, McSherry T, Feldman M. caTIES—A Grid Based System for Coding and Retrieval of Surgical Pathology Reports and Tissue Specimens In Support Of Translational Research. J Am Med Inform Assoc. 2010; 17:253–64. [PubMed: 20442142]
7. Friedman, C. A broad coverage natural language processing system. AMIA Fall Symp; 2000; p. 270-274.
8. Jain, NL., Knirsch, CA., Friedman, C., Hripcsak, G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. AMIA Fall Symp; Philadelphia. 1996; p. 542-546.
9. Sneiderman, CA., Rindflesch, TC., Bean, CA. Identification of anatomical terminology in medical text. AMIA Fall Symp; 1998; p. 428-432.
10. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. Journal of American Medical Informatics Association. 2010; 17:229–36.
11. Aronson, AR. MetaMap: Mapping Text to the UMLS Metathesaurus. US NLM2006;
12. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004; 32:267–70.
13. Jonquet, C., Shah, NM., Musen, MA. The Open Biomedical Annotator. presented at the AMIA Summit on Translat Bioinformatics; San Francisco. 2009;
14. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010; 17:507–13. [PubMed: 20819853]
15. Ferrucci, d, Lally, A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering. 2004; 10:327–348.
16. OpenNLP. Available: http://opennlp.sourceforge.net/index.html
17. Gottlieb DJ, Punjabi NM, Mehra R, Patel SR, Quan SF, Babineau DC, Tracy RP, Rueschman M, Blumenthal RS, Lewis EF, Bhatt DL, Redline S. CPAP versus oxygen in obstructive sleep apnea. New England Journal of Medicine. 2014; 370:2276–85. [PubMed: 24918372]
18. Moreau, L., Missier, P. PROV Data Model (PROV-DM). World Wide Web Consortium W3C2013;



Obstructive sleep apnea is a highly p
middle-aged women and 24% of mic
to-severe obstructive sleep apnea.[1]
risk factor for hypertension, coronary
underlying these associations are th
inflammation.[10] Although reports fro
cardiovascular risk is reduced amon
continuous positive airway pressure
adequately powered randomized tria
on myocardial infarction, stroke, or c
risk reduction with CPAP was recent
power.[13]

A meta-analysis of small, unregister
approximately 2 mm Hg in mean art
clinical trials involving patients with r
improvements in blood pressure with
untreated hypertension, the 24-hour
confidence interval [CI], 0.4 to 2.7) c
involving patients with hypertension

NIH held a joint workshop in June 2014 with the Nature Publishing Group and Science on the issue of reproducibility and rigor of research findings, with journal editors representing over 30 basic/preclinical science journals in which NIH-funded investigators have most often published. The workshop focused on identifying the common opportunities in the scientific publishing arena to enhance rigor and further support research that is reproducible, robust, and transparent.

Biomedical Text
Sentence Boundary Detector
Tokenizer
Part-of-Speech (POS) Tagger
Shallow Parser
Knowledge Models
Class hierarchy
Class hierarchy (inferred)
Class hierarchy: ResearchStudy
Thing
Activity
StudyMethod
DataAnalysisMethod
Comparison
MissingDataProtocol
StatisticalMethod
DataCollectionMethod
Intervention
ProvCaRe ontology
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY
Unified Medical Language System
Named Entity Recognition (NER)
ProvCaRe Components:
1. Study Method
2. Study Data
3. Study Tools
Post Processing
Provenance Metadata for Scientific Reproducibility
NIH National Institutes of Health
Turning Discovery Into Health
Search NIH
Health Information
Grants & Funding
News & Events
Research & Training
Home » Research & Training » Rigor and Reproducibility
RIGOR AND REPRODUCIBILITY
Rigor and Reproducibility
Principles and Guidelines
Expanded Guidelines
Application Instructions
Training
Funding Opportunities
Meetings and Workshops
Publications
Principles and Guidelines for Reporting Preclinical Research

Figure 1.

The ProvCaRe-NLP pipeline consists of five distinct phases that extracts provenance information from biomedical text. The pipeline extends the cTAKES tool to use a provenance-specific ontology together with UMLS (MetaMap) and biomedical ontologies listed at NCBO.

Table 1

Provenance Metadata Extracted by Domain Expert from Biomedical Article on Sleep Medicine Research

Study Method	**Data Analysis**	**Interventions**	**Comparison**	**Data Collection Method**
CPAP versus Oxygen in Obstructive Sleep Apnea, Gottlieb et al. 2014	Analysis, effect, CPAP, nocturnal, oxygen, supplementation, 24-hour mean arterial pressure, markers of cardiovascular risk, patients, established, coronary heart disease, multiple, cardiovascular risk factors	CPAP, Nocturnal, supplemental, oxygen, Healthy Lifestyle education	CPAP, nocturnal, supplemental, oxygen, 24-hour mean arterial pressure	Ambulatory, blood pressure monitoring, Blood, samples, fasting, glucose, insulin, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, total cholesterol, triglycerides, protein, C-reactive protein
Study Data	**Study Population**	**Study Variables**		
CPAP versus Oxygen in Obstructive Sleep Apnea, Gottlieb et al. 2014	Patients, years, age, established, coronary heart disease, multiple, cardiovascular risk factors, Berlin, questionnaire, scores, home, sleep testing, apnea-hypopnea index, events per hour	24-hour mean arterial blood pressure, Fasting, glucose, insulin, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, total cholesterol, triglycerides, protein, C-reactive protein, levels		
Study Tools	**Instrument**	**Software Tools**		
CPAP versus Oxygen in Obstructive Sleep Apnea, Gottlieb et al. 2014	Berlin Questionnaire. Epworth Sleepiness Scale. Embletta Gold portable sleep monitor with nasal cannula pressure transducer, oronasal thermal sensor, inductance plethysmography, and finger-tip pulse oximetry sensor. Three-lead electrocardiography	None specified		

Table 2

Comparative Evaluation of the Recall of the MetaMap, NCBO Annotator, and ProvCaRe-NLP Pipelines for Extracting Provenance Terms Using 20 Published Articles

Paper (Author last name and year of publication)	MetaMap		NCBO Annotator		MetaMap + NCBO Annotator		ProvCaRe-NLP + MetaMap+NCBO Annotator	
	Full Article	Abstract +Method+ Result	Full Article	Abstract+ Methods+ Results	Full Article	Abstract+ Methods+ Results	Full Article	Abstract+ Methods+ Results
O’Connor et al., 2009	0.44	0.52	0.89	0.93	0.9	0.95	0.92	0.97
Gottlieb et al., 2014	0.82	0.85	0.86	0.91	0.86	0.91	0.92	0.97
Redline et al., 1999	0.82	0.89	0.88	0.89	0.88	0.89	0.94	0.95
Patel et al., 2009	0.76	0.8	0.94	0.97	0.94	0.97	0.95	0.98
Wang et al., 2012	0.68	0.82	0.9	0.97	0.9	0.97	0.97	0.99
Patel et al., 2014	0.71	0.75	0.86	0.92	0.86	0.93	0.86	0.93
Mehra et al., 2007	0.76	0.93	0.91	0.98	0.92	0.98	1	1
Kwon et al., 2016	0.68	0.8	0.88	0.88	0.88	0.88	0.89	0.95
Archbold et al., 2010	0.72	0.77	0.79	0.85	0.83	0.85	0.85	0.87
Ramos et al., 2015	0.78	0.84	0.92	0.95	0.92	0.95	0.94	0.97
Brown et al., 2002	0.76	0.8	0.74	0.86	0.74	0.87	0.78	0.88
Quante et al., 2015	0.73	0.82	0.93	0.99	0.95	0.99	1	1
Yeboah et al., 2011	0.55	0.69	0.84	0.91	0.84	0.91	0.85	0.92
Kwon et al., 2014	0.68	0.74	0.87	0.91	0.87	0.91	0.87	0.91
Dean et al., 2015	0.91	0.93	0.87	0.93	0.87	0.93	0.99	1

Paper (Author last name and year of publication)	MetaMap		NCBO Annotator		MetaMap + NCBO Annotator		ProvCaRe-NLP + MetaMap+NCBO Annotator	
	Full Article	Abstract +Method+ Result	Full Article	Abstract+ Methods+ Results	Full Article	Abstract+ Methods+ Results	Full Article	Abstract+ Methods+ Results
Chen et al., 2015	0.66	0.74	0.86	0.94	0.86	0.94	0.94	0.95
Bertisch et al., 2015	0.77	0.83	0.75	0.78	0.78	0.84	0.98	0.99
Walia et al., 2014	0.71	0.8	0.88	0.89	0.88	0.89	0.98	0.99
Chervin et al., 2015	0.65	0.69	0.89	0.96	0.89	0.96	0.89	0.96
Rosen et al., 2015	0.74	0.8	0.88	0.89	0.88	0.89	0.99	1