



10 years ESJ
Special edition

High Throughput Neurological Phenotyping with MetaMap

Daniel B. Hier, MD

Raghu Yelugam, MS

Missouri University of Science and Technology, USA

Michael D. Carrithers, MD PhD

University of Illinois at Chicago, USA

Donald C. Wunsch II, PhD

Missouri University of Science and Technology, USA

[Doi:10.19044/esj.2022.v18n4p37](https://doi.org/10.19044/esj.2022.v18n4p37)

Submitted: 10 November 2021

Accepted: 20 January 2022

Published: 08 February 2022

Copyright 2022 Author(s)

Under Creative Commons BY-NC-ND

4.0 OPEN ACCESS

Cite As:

Hier D.B., Yelugam R., Carrithers M.D. & Wunsch D.C. (2022). *High Throughput Neurological Phenotyping with MetaMap*. European Scientific Journal, ESJ, 18 (4), 37.

<https://doi.org/10.19044/esj.2022.v18n4p37>

Abstract

The phenotyping of neurological patients involves the conversion of signs and symptoms into machine readable codes selected from an appropriate ontology. The phenotyping of neurological patients is manual and laborious. MetaMap is used for high throughput mapping of the medical literature to concepts in the Unified Medical Language System Metathesaurus (UMLS). MetaMap was evaluated as a tool for the high throughput phenotyping of neurological patients. Based on 15 patient histories from electronic health records, 30 patient histories from neurology textbooks, and 20 clinical summaries from the Online Mendelian Inheritance in Man repository, MetaMap showed a recall of 61-89%, a precision of 84-93%, and an accuracy of 56-84% for the identification of phenotype concepts. The most common cause of false negatives (failure to recognize a phenotype concept) was an inability of MetaMap to find concepts that were represented as a description or a definition of the concept. The most common cause of false positives (incorrect identification of a concept in the text) was a failure to recognize that a concept was negated. MetaMap shows potential for high throughput phenotyping of neurological patients if the problems of false negatives and false positives can be solved.

Keywords: MetaMap, phenotyping, high throughput, ontologies, natural language processing

Introduction

Valuable clinical data is held in electronic health records in the form of unstructured text (Esteva et al., 2019). To unlock the value of this data, it must be converted to a computable form (Xiao, Choi, & Sun, 2018; Miotto, Wang, Wang, Jiang, & Dudley, 2018). One approach to making unstructured clinical data computable has been called deep phenotyping (Kohler et al., 2019). Deep phenotyping is the conversion of the signs and symptoms of a patient to concepts from a target ontology (Kohler et al. 2014; Robinson 2012). Phenotyping is important to the precision medicine initiative that individualizes treatment and prognosis to specific patients (Collins & Varmus, 2015). Although phenotyping can be done by manual methods, high throughput methods are needed to phenotype large numbers of patients in precision medicine (Robinson, 2012). High throughput phenotyping depends on the rapid extraction of signs and symptoms from large text sources. The text mining of electronic health records relies on methods of named entity recognition (NER) derived from natural language processing (Marrero, Urbano, Sanchez-Cuadrado, Morato, & Gomez- Berbis, 2013; Kimia, Savova, Landschaft, & Harper, 2015). Fu et al. (2020) defined concept extraction as a two-stage process in which medical concepts are first identified in text and then mapped to a concept in a disease ontology.

MetaMap is a natural language processing (NLP) tool developed by the National Library of Medicine that utilizes linguistic and statistical methods to recognize medical terms in text and map them to concepts in the UMLS Metathesaurus (Aronson & Lang 2010; Aronson, 2000, 2001a, 2001b; Mork & Aronson, 2006). MetaMap works by detecting sentence boundaries, tokenizing words, tagging parts of speech, and parsing sentences into smaller phrases. Candidate phrases are compared to strings in the UMLS Metathesaurus and similarity scores are calculated. In a final processing stage, word disambiguation occurs, negation is evaluated, and the best concept matches are selected. Although the primary purpose of MetaMap is to index the scientific literature, MetaMap has been re-purposed for other uses, including the identification and extraction of medical concepts in electronic medical records, text mining of clinical concepts in various documents, identification of medical concepts in clinical guidelines, and the extraction of medical concepts from email (Fu et al., 2020, Bashyam et al., 2009, Osborne et al., 2007, Brennan & Aronson, 2003, Gooch & Roudsari, 2011).

This study explores the utility of MetaMap for the high throughput phenotyping of neurological patients. We have previously phenotyped neurological patients by manual methods (Hier et al., 2020, Hier & Brint,

2020). The phenotyping process involves converting a neurological history and examination in narrative form to concepts and machine-readable codes selected from an appropriate ontology (NCBO BioPortal, 2021). The process of deep phenotyping is illustrated with this patient with multiple sclerosis.

A 55-year-old woman with multiple sclerosis complains of fatigue and blurred vision. Examination shows increased reflexes in all limbs, decreased visual acuity in the right eye, appendicular ataxia in the left arm, an unsteady gait, an afferent pupil defect in the right eye, and bilateral extensor plantar responses.

The patient can be phenotyped using concepts and CUI codes selected from the UMLS Metathesaurus (Table 1).

Table 1: Example Phenotype of a multiple sclerosis patient

Sign or Symptom	CUI code
appendicular ataxia	C0750937
hyperreflexia	C0151889
afferent pupil defect	C0339662
blurred vision	C0344232
bilateral extensor plantar	C0422917
unsteady gait	C0231686
decreased visual acuity	C0234632
fatigue	C0015672

Methods

MetaMap as a tool to extract concepts from for phenotyping was evaluated on three different types of text files. The first file consisted of the brief histories and examinations of 30 neurology cases from two electronic textbooks (Gondolo, 2005; Ubogu, 2005). The second file consisted of de-identified progress notes from 15 neurology outpatients seen in the Multiple Sclerosis Clinic of the University of Illinois at Chicago. The third file consisted of 20 clinical descriptions of inherited neurological diseases selected from the Online Mendelian Inheritance of Man (OMIM) (Hamosh, Scott, Amberger, Valle, & McKusick, 2000; Amberger & Hamosh, 2017).

Text files were spell-checked. Run-on words and separated words were eliminated. Each file was batch entered into MetaMap (Table 2). The negation option and the word disambiguation options were used to reduce false negatives and false positives. Semantic types were restricted to those relevant to phenotyping. Each concept identified in the files by MetaMap was compared to concepts in a target neuro-ontology (Hier & Brint, 2020; NCBO BioPortal, 2021). The neuro-ontology is a subset of 1531 concepts from the

UMLS Metathesaurus. Concepts were marked *relevant* to phenotyping if they were in the target ontology and were marked *extraneous* if they were not (Table 3). Extraneous concepts were excluded from further analysis. An expert coder reviewed all assignments using these criteria:

- 1) Ground truth positives were concepts identified by the expert coder in the textfiles.
- 2) Ground truth negatives were all negated concepts found by MetaMap and confirmed by the expert coder plus false positives.
- 3) False positives were concepts found by MetaMap that the expert coder did not confirm.
- 4) True negatives were concepts that were correctly negated by MetaMap.
- 5) False negatives were concepts missed by MetaMap and added by the expert coder.
- 6) True positives were concepts found by MetaMap in the text and that were confirmed by the expert coder.

Errors (false positives or false negatives) were further classified as follows (Table 4).

Failure to recognize a concept as a definition or description. Some phenotypic concepts were described or defined without explicit mention of the concept name. Consider the UMLS concept *anomia* which is a disorder of naming. MetaMap recognizes *anomia* as a valid concept but may fail to recognize its description in the text "she can name none of the everyday items that are shown to her."

Failure to recognize colloquialisms or figures of speech. MetaMap recognizes the concept *ankle reflex* but may not recognize the colloquialism *ankle jerk*. Similarly, MetaMap recognizes the concept *extensor plantar response* and its synonym *Babinski sign* but may not recognize the colloquial expression *upgoing toe*.

Failure to join coordinated concepts. When encountering two concepts from the UMLS Metathesaurus, MetaMap will sometimes fail to join two coordinated concepts into a more specific concept. For example, when MetaMap processes the phrase *ataxic gait*, it may decompose it into *gait* and *ataxia* but fail to join these two concepts into the more specific concept (*ataxia + gait = ataxic gait*).

Errors in assessing negation. For purposes of this paper, we use negation in a restricted sense. When phenotyping, the focus is on collecting abnormal findings (weakness, incoordination, imbalance, etc.), and normal findings are ignored (normal balance, normal sensation, normal strength, normal coordination, etc.) However, negation reverses the meaning of a phrase. For example, *balance is normal* is a normal finding. However, *imbalance, poor balance, impaired balance, and no balance* are abnormal findings. Some concepts are intrinsically abnormal such as *alexia* which means *unable to read*. Thus, the phrases *no alexia* and *able to read* are normal findings. MetaMap has a built-in option called NeGex, which assists in evaluating negated concepts. Negex marks negated concepts in text with the symbol N.

Failure to disambiguate between homographs. Homographs are words that are spelled the same but have different meanings. For example, MetaMap might fail to disambiguate *looking blue* (hypoxic) from *feeling blue* (depressed). We used the word disambiguation option in MetaMap.

Failure to recognize acronyms or abbreviations. MetaMap may fail to recognize a concept expressed as an acronym. For example, MetaMap may not recognize that ON is an acronym for *optic neuritis*.

Attribution error. An attribution error occurs when a finding that belongs to someone other than the patient (e.g., a family member) is attributed to the patient. For example, hypertension in the father of the patient can be erroneously allocated to the patient.

Standard formulas were used to calculate recall, precision, accuracy, and specificity (Table 5). Confusion matrices (Table 6) and performance rates (Table 7) were calculated separately for the textbook neurology cases, the EHR neurology notes, and the OMIM clinical diseases synopses.

Table 2. MetaMap Batch Mode Settings

Option	Setting
relaxed model	-C
enable negex	-negex
allow concept gaps	-g
use word disambiguation	-y
ignore word order	-i
single line delimited input	-l
show CUIs	
restrict semantic types	-J
conjunction processing	-conj
composite phrases	-Q
restricted semantic types	acquired abnormality, disease or syndrome, finding, mental or behavioral dysfunction, sign or symptom, mental process
restricted vocabulary sources	MTH, MSH, HPO, OMIM, MEDCIN ICD10CM, SNOMEDCT US †

† For vocabulary sources, see
<https://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html>

Table 3. Formulas for recall, precision, and accuracy

recall	$TP/(TP + FN)$
precision	$TP/(TP+FP)$
accuracy	$TP+TN/(TP+FP+TN+FN)$

Results

MetaMap identified UMLS concepts in the three different types of text files: neurological case histories from textbooks, physician notes from an EHR, and neurological clinical summaries from the OMIM. On average, each textbook case had 38 UMLS concepts, each EHR physician note had 138 UMLS concepts, and each OMIM clinical summary had 126 UMLS concepts. Most of the concepts identified were not relevant to the phenotyping process (identified concepts were valid UMLS concepts but were not signs or symptoms) and were extraneous to the phenotyping (Table 4). Examples of extraneous concepts include *d cognition, eye, ocular, death, short-term memory, diagnosed, patient outcome, problem*, and many others. These extraneous concepts appear in the UMLS Metathesaurus but are not in the limited ontology of 1531 neurological concepts used for phenotyping (Hier & Brint 2020). Extraneous concepts were filtered out and excluded from the calculations of accuracy, precision, and recall.

True positives were defined as concepts that were valid signs and symptoms as identified by the expert coder. MetaMap had a hit rate for these true positives of 69% on the textbook cases, 76% on the EHR notes, and 87% on the OMIM clinical summaries (Table 4).

Most of the false negatives (4.8-6.2 per case) occurred due to a failure of MetaMap to recognize concepts expressed as a definition of a concept (e.g., *unable to read* for *alexia*) or as a description of a concept (e.g., *the patient could not swallow food* for *dysphagia*) (Table 4).

After filtering out extraneous concepts, the specificity of MetaMap was high (98-99%) due to a high true negative rate and a low false-positive rate (Table 5). False positives had several causes, including failure to recognize that a concept was negated correctly, failure to disambiguate homographs, failure to join concepts into a more specific concept (some have called these errors in post-coordination of concepts (Brown 2001, Hedeler et al. 2014, Nasvas et al. 2010) and errors in attribution (attributing a concept to the wrong person).

Table 4: Textfile sources used for MetaMap evaluation

Measure	Textbook Source	EHR Source	OMIM Source
Number of cases	30	15	20
Word per cases	120.2	616.6	963.1
Lines per case	23.7	137.9	126.1
MetaMap concepts found per case	38.1	302.6	206.2
Extraneous concepts per case	30.7	290.7	165.3

Table 5. Analysis of Errors made by MetaMap

FP# due to error in interpreting negation	0.5	1.4	0.7
FP due to attribution to the wrong person	0.3	0.0	0.1
FP due to failure to join concepts	0.3	0.3	0.0
FP due to failure to disambiguate homographs	0.3	0.0	1.6
FN† due to failure to recognize colloquialism	0.3	0.0	0.0
FN due to failure to recognize the description of the concept	3.2	3.5	3.3
FN due to failure to join concepts	0.9	0.8	0.3
FN due to error in interpreting negation	0.1	0.1	0.0
FN due to failure to disambiguate homographs	0.4	0.1	0.0

False positives (FP) were concepts found by MetaMap in the target neuro-ontology that the expert coder did not confirm.

† False negatives (FN) were concepts missed by MetaMap and added by the expert coder.

Table 6. Confusion Matrix for MetaMap

Source	Ground Truth	Coder Assignment	MetaMap Positive	MetaMap Negative
Textbook	Positive	12.5	TP (7.6)	FN (4.9)
	Negative	1.7	FP (1.4)	TN (0.3)
EHR	Positive	16.4	TP (11.9)	FN (4.5)
	Negative	2.6	FP (1.7)	TN (0.9)
OMIM	Positive	34.7	TP (31.1)	FN (3.6)
	Negative	2.9	FP (2.4)	TN (0.5)

Where TP = true positives, TN = true negatives, FP = false positives, and FN = false negatives

Table 7. Performance metrics for MetaMap Neurology Phenotyping Task

Rates	Textbook Source	EHR Source	OMIM Source
Recall	60.8%	72.6%	89.6%
Accuracy	55.6%	67.4%	84.0%
Precision	83.9%	87.5%	92.8%

Discussion and Conclusions

The results suggest that MetaMap can extract clinical concepts for neurological phenotyping from electronic health records, published neurological case studies, and online clinical summaries (Table 7). Its most important limitation is that recall of the ground truth concepts is only 60.8-89.6% depending on the file type. The most important reason for misses was the failure of MetaMap to recognize concepts when they appeared as a description of the concept or as a definition of the concept. MetaMap had the best recall on the OMIM clinical summaries and the worst recall on the textbook neurology cases. The better performance of MetaMap on the OMIM clinical summaries compared to the textbook cases reflects different writing styles. The OMIM clinical summaries are written concisely with less use of descriptive text. For example, one of the neurology textbooks writes that "his wife stated that he had progressive weakness in the left face and upper extremity with milder involvement of the lower extremity for about three weeks." MetaMap finds the concept of facial weakness but misses arm weakness and leg weakness. The more economical writing style of the OMIM summaries, such as "additional features included central hypoventilation, postural hypotension, bladder incontinence, and myoclonus," allowed MetaMap to correctly recognize all four phenotypic concepts: central hypoventilation, urinary incontinence, myoclonus, and orthostatic hypotension. False positives (spuriously added concepts) were unusual with MetaMap. Although many UMLS concepts in the text files are irrelevant to phenotyping, two strategies to reduce false positives were used. First, semantic types were restricted to concepts relevant to phenotyping (Table 2). Second,

the use of a small target neuro-ontology permitted a filtering of extraneous concepts (Table 4).

The main causes of false positives were errors in evaluating negation, failure to join multiple concepts into a more specific concept (e.g., *absent + ankle + reflex = absent ankle reflex*), attribution errors in which the concept was assigned to the wrong person (e.g., hypertension in the father of the patient is attributed to the patient), and errors due to a failure to disambiguate homographs (Table 5). Disambiguation of homographs (i.e., failure to distinguish “fall down” from “Fall season”) has proven to be a difficult problem in NLP (Schuemie, Kors, & Mons, 2005; Agirre & Edmonds, 2007). Although MetaMap uses rule-based methods to resolve word mean ambiguity, a few false positives and false negatives occurred due to word meaning ambiguity (Table 5).

The primary reason for false negatives (Table 5) was the inability of the MetaMap to reliably recognize concepts encoded as either a description of the concept or a definition of the concept. For example, MetaMap can recognize concepts such as *bradykinesia* or *hypomimia* when stated explicitly but may fail to recognize concepts expressed as a description such as “movements were slow and halting” (bradykinesia) or “the face was minimally expressive” (hypomimia). Friedlin & Overhage (2011) noted that some false negatives occurred due to an unexplained inability to recognize a concept that is in the UMLS Metathesaurus. They also found that valid concepts in the text were sometimes absent from the UMLS. Baud et al. (2004) have commented that variability in how concepts are represented in medical text contributes to false negatives. They tested MetaMap on 25 variants of the expressions *scapular fracture and* found that MetaMap recognized only 68-88% of the variants as valid.

Divita et al. (2004) evaluated the ability of MetaMap to find UMLS concepts in genetic condition documents of the National Library of Medicine. They found a recall of 53% and that most false negatives were due to failure to identify definitional phrases and failure to recognize concepts that had been expressed implicitly. These findings mirror our findings. Similarly, Pratt & Yetisgen-Yildiz (2003) examined the ability of MetaMap to find UMLS concepts in medical article titles. They found a 53-93% recall that depended upon the stringency of their evaluation criteria. Okumura & Tateisi (Okumura & Tateisi, 2012) and Cohen et al. (2011) have evaluated MetaMap for high throughput phenotyping.

Other causes of error by MetaMap included failure to recognize colloquialisms (upgoing toes for Babinski sign) and failure to recognize figures of speech (feeling blue for depression). Paterson et al. (2010) have emphasized that just as medical English is a sub-language of English, each medical subspecialty has its own sub-language. For example, in orthopedics,

a *pin* is a device for fixing a fracture, whereas, in neurology, a *pin* is a device for testing sensation. Knowledge of which medical sub-language is in use (neurology, orthopedics, etc.) might help recognize concepts represented as colloquialisms, figures of speech, or ambiguous terms. Using a target ontology limited to a specific specialty could improve precision and accuracy without reducing recall.

We are exploring whether additional pre-processing of input to MetaMap or post-processing of output from MetaMap can improve phenotyping accuracy. Some failures of MetaMap may be addressable with additional NLP processing. We are currently investigating whether hybrid methods that combine rule-based algorithms with neural networks can improve the recall of phenotypic concepts by MetaMap. The recognition of phenotypic concepts as definitions or descriptions poses a special challenge for rule-based NLP methods like MetaMap. The potential advantages of specialty ontologies (for example, the neuro-ontology) for phenotyping patients from a specific disease class such as neurology needs further exploration. If accuracy could be improved to 95% or better high throughput phenotyping of electronic health record clinical notes using MetaMap could become workable.

Human Studies

The Institutional Review Board of the University of Illinois approved this study (Neuroimmunology Biobank Study protocol 2017-0520).

References:

1. Agirre, E., & Edmonds, P. (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33). Springer Science & Business Media.
2. Amberger, J. S., & Hamosh, A. (2017). Searching online mendelian inheritance in man (OMIM): a knowledge base of human genes and genetic phenotypes. *Current Protocols in Bioinformatics*, 58(1), 1–2.
3. Aronson, A. R. (2000). The MetaMap mapping algorithm. *National Institutes of Health (NIH), Technical Report* (<https://lhncbc.nlm.nih.gov/ii/information/Papers/mm.mapping.pdf>)
4. Aronson, A. R. (2001a). Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp.* 2001:17-21.
5. Aronson, A. R. (2001b). MetaMap variant generation. *2014-08-15. nlm.nih.gov/papers/references/mm.variants.pdf*.

6. Aronson, A. R., & Lang, F.-M. (2010, 05). An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17 (3), 229-236. Retrieved from <https://doi.org/10.1136/jamia.2009.002733>
7. Bashyam, V., Divita, G., Bennett, D. B., Taira, R. K., & Browne, A. C. (2007). A normalized lexical lookup approach to identifying UMLS concepts in free text. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems* (p. 545). IOS Press
8. Brown P. (2001). Principles of Post-coordination Concept Construction and Retrieval. *Proceedings of the AMIA Symposium*, 869.
9. Brennan, P. F., & Aronson, A. R. (2003). Towards linking patients and clinical information: detecting UMLS concepts in email. *Journal of biomedical informatics*, 36(4-5), 334-341.
10. Cohen, R., Gefen, A., Elhadad, M., & Birk, O. S. (2011). CSI-OMIM-Clinical synopsis search in OMIM. *BMC Bioinformatics*, 12 (1), 1-10.
11. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372 (9), 793–795.
12. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, 25 (1), 24–29.
13. Friedlin, J., & Overhage, M. (2011). An evaluation of the UMLS in representing corpus-derived clinical concepts. In *AMIA annual symposium proceedings* (Vol. 2011, p. 435).
14. Fu, S., Chen, D., He, H., Liu, S., Moon, S., Peterson, K. J., and others (2020). Clinical concept extraction: a methodology review. *Journal of Biomedical Informatics*, 103526.
15. Gondolo, T. (2005). *Neurology study guide: Oral board examination review*. Springer Nature.
16. Gooch, P. and Roudsari, A. (2011). *A tool for enhancing MetaMap performance when annotating clinical guideline documents with UMLS concepts*. Paper presented at the IDAMAP Workshop at 13th Conference on Artificial Intelligence in Medicine (AIME'11), 02-07-2011 - 06-07-2011, Bled, Slovenia.
17. Hamosh, A., Scott, A. F., Amberger, J., Valle, D., & McKusick, V. A. (2000). Online mendelian inheritance in man (OMIM). *Human mutation*, 15 (1), 57–61.
18. Hedeler, C., Parsia, B., & Brandt, S. (2014, May). Estimating and analyzing coordination in medical terminologies. *IEEE*. In *2014 IEEE*

27th International Symposium on Computer-Based Medical Systems (pp. 357-362).

19. Hier, D. B., & Brint, S. U. (2020). A neuro-ontology for the neurological examination. *BMC Medical Informatics and Decision Making*, 20(1), 1–9.
20. Hier, D. B., Kopel, J., Brint, S. U., Wunsch, D. C., Olbricht, G. R., Azizi, S., & Allen, B. (2020). Evaluation of standard and semantically-augmented distance metrics for neurology patients. *BMC Medical Informatics and Decision Making*, 20(1), 1–15.
21. Kimia, A. A., Savova, G., Landschaft, A., & Harper, M. B. (2015). An introduction to natural language processing: how you can get more from those electronic notes you are generating. *Pediatric emergency care*, 31(7), 536–541.
22. Kohler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., and others (2014). The human phenotype ontology project: linking molecular biology and disease through phenotype data—*Nucleic Acids Research*, 42(D1), D966–D974.
23. Kohler, S., Øien, N. C., Buske, O. J., Groza, T., Jacobsen, J. O., McNamara, C., and others (2019). Encoding clinical data with the human phenotype ontology for computational differential diagnostics. *Current protocols in human genetics*, 103(1), e92.
24. Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., & Gómez-Berbís, J. M. (2013). Named entity recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces*, 35(5), 482–489.
25. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities, and challenges. *Briefings in bioinformatics*, 19(6), 1236–1246.
26. Mork, J. G., & Aronson, A. R. (2006). Filtering the UMLS Metathesaurus for MetaMap. *National Library of Medicine Technical Report*. (<https://lhncbc.nlm.nih.gov/ii/information/Papers/filtering09.pdf>)
27. Navas, H., Lopez Osornio, A., Gambarte, L., Elías Leguizamón, G., Wasserman, S., Orrego, N., Luna, D. and de Quirós, F.G.B. (2010). Implementing rules to improve the quality of concept post-coordination with SNOMED CT. In *MEDINFO 2010* (pp. 1045-1049). IOS Press.
28. NCBO BioPortal. (2021). Neurologic examination ontology. <https://bioportal.bioontology.org/ontologies/NEO>.
29. Okumura, T., & Tateisi, Y. (2012). A lightweight approach for extracting disease-symptom relation with MetaMap toward

- automated generation of disease knowledge base. *International Conference on Health Information Science*. pp.164–172.
30. Osborne, J. D., Lin, S., Zhu, L. J., & Kibbe, W. A. (2007). Mining biomedical data using MetaMap transfer (MMtx) and the unified medical language system (UMLS). In *Gene Function Analysis* (pp. 153-169). Humana Press
 31. Patterson, O., Igo, S., & Hurdle, J. F. (2010). Automatic acquisition of sublanguage semantic schema: towards the word sense disambiguation of clinical narratives. In *AMIA Annual Symposium Proceedings* (Vol. 2010, p. 612). American Medical Informatics Association.
 32. Pratt, W., & Yetisgen-Yildiz, M. (2003). A study of biomedical concept identification: MetaMap vs. people. 2003, 529. (AMIA Annual Symposium Proceedings)
 33. Robinson, P. N. (2012). Deep phenotyping for precision medicine. *Human Mutation*, 33 (5),777–780.
 34. Schuemie, M. J., Kors, J. A., & Mons, B. (2005). Word sense disambiguation in the biomedical domain: an overview. *Journal of Computational Biology*, 12 (5), 554–565.
 35. Ubogu, E. E. (2005). *Neurology oral boards review*. Humana Press.
 36. Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, 25(10), 1419–1428.