Classifying *Mental-functioning* using a Named Entity Recognition Tool: Based on the Ecological Mental-functioning Model Ontology

Guy Divita[1], Bart Desmet[1], Maryanne Sacco[1], Kathleen Coale[1], Rafael Jimenez Silva[1], Rebecca Parks[1], Elizabeth Rasch[1]

[1] Rehabilitation Medicine Department, National Institutes of Health Clinical Center, Bethesda, MD

Abstract

Objective

Mental-functioning information extracted and characterized from clinical documentation is sought and valued by clinicians, however, it is sparsely found in clinical records. The Ecological Mental-Functioning Ontology (EMFO) was developed as a mechanism to explore and define mental-functioning from a clinical rehabilitation perspective. By leveraging the terminology and knowledge derived from the EMFO, a named entity recognition (NER) was developed and applied to extract mental- functioning activity information in health records.

Methods and Materials

Optical character recognition (OCR) processed pages were sampled from a large set of disability claims records that had been provided by the Social Security Administration (SSA) for extraction of functioning information. A sub-sample was manually annotated for Communication & Cognition (ComCog), and another sub-sample manually annotated for Interpersonal Interactions and Relationships (IPIR), both of which are components of mental-functioning. These samples provided training and benchmarking for the NER.

Results

Weighted average ComCog Subclassification F-Score on this set was 0.62. Weighted average IPIR subclassification F-Score was 0.63.

Discussion

Challenges included noisy OCR'd data, issues with segmentation, and a cognitively challenging task for manual annotation.

Conclusions

The benchmarked NER is one part in a series of tasks to provide the informatics community with resources related to functioning broadly and mental-functioning specifically.

1 Introduction

Mental-functioning information extracted and characterized from clinical documentation is a valuable source of data for purposes such as disability determination, health research, and health outcomes evaluation, and is information that clinicians express interest in having[1]. However, *mental-functioning* activity information can be challenging to identify in health records, due to over emphasis of documentation at the mental body function level. Accurately characterizing a patient's current and past *mental-functioning* has been elusive to date, in part, because it has been inadequately described within standard terminologies and ontologies and often confused with the related topic of mental-function.[1,2] The Ecological Mental-functioning Ontology (EMFO)[3-5] was developed as a mechanism to explore and define *mental-functioning* from a clinical rehabilitation perspective. The aim of this paper is to present the development and benchmarking of a NER that identifies *mental-functioning*, leveraging terminology and knowledge derived from the EMFO, benchmarked on SSA claimant records. The NER is an evolution of v3NLP Framework, tuned for this application.

2 Prior Work

The rule-based NLP platform employed for this work, referred to as java-nlp-Framework[6], was adapted from the V3NLP Framework[7] and Sophia[8] which were used for symptom extraction[9] and finding mentions of sexual trauma in clinical notes of veterans[10]. The framework employed is built upon Apache's Unstructured Information Management Architecture[11] (UIMA) NLP platform, therefore resembles the cTAKES[12] system closely, but has a pedigree from Unified Medical Language System (UMLS) concept extraction in biomedical literature, MetaMap[13] as the java code base was derived from MMTx[14]. More recently, Java-nlp-Framework has been crafted to find body-function information within clinical texts in support of efforts by the Social Security Administration (SSA) to identify such information[6].

2.1 Ecological Mental-functioning Ontology (EMFO) Background

The Ecological Mental-functioning Ontology (EMFO) describes classes and relationships having to do with the domain of *mental-functioning*. *Mental-functioning* is an individual's behaviors, activities and participation in daily life. These behaviors can be observed by others, such as clinicians, and documented in health records, and could also be reported by the patient themselves. This ontology was created initially to identify what is and is not *mental-functioning* in a larger effort to find *mental-functioning* mentions within clinical records for evidence to potentially aid social security disability adjudicators in making informed decisions from medical evidence.

It is important to point out that *mental-functioning* is different but related to mental-functions. Mental health domain experts have stressed the importance to distinguish and highlight mentions within clinical text that have to do with *mental-functioning* at the activity level and not include mentions solely at the body/mental-function level. Mental-functions are classified as body functions, a person's intrinsic physiological capability, and are not evidence of activities nor participation. Documenting that a patient has the capability to do calculations (evidence of mental-function), is different from documentation based on observation of calculating the tip for a restaurant bill (evidence of *mental-functioning*).

Standard medical terminologies, including Systematized Nomenclature of Medicine - Clinical Terms[15] (SNOMED-CT) and Medical Subject Headings[16] (MeSH), are distillations of language used within clinical notes. Leveraging existing medical terminologies to identify language of *mental-functioning* in clinical records was limited, as terminologies include much coverage around mental-functions and less coverage of activity functioning, further exacerbated by definitions muddling the distinctions between the two.

Initially, the ontology was based on a basic structure defined in the International Classification of Functioning, Disability and Health (ICF)[17] that explicitly separates functions of the body from functions at the activities and participation level, or what we call functioning. Overall, the ontology does recognize the ICF structure broadly with a few deviations and additions reflecting an evolved understanding of the field since the ICF's introduction in 2001.

Mental-functioning is related to mental-functions, which is related to body function, which is related to body structures, and so on. All of this exists within contextual and environmental factors that affect functioning in daily life activities. As such, the EMFO includes classes and relationships that have to do with body function, body structures, context and environment along with feedback mechanisms that round out the description of the theoretical model. The Ecological part of this ontology's name is an acknowledgement of the contextual, environmental, and feedback components incorporated within the ontology.

2.2 NER Background

The java-nlp-Framework contains a traditional NLP pipeline based on the UIMA platform.
This platform is structured as a pipeline composed of individual modules or software-annotators that perform work on input documents, where each input document gets turned into an encapsulation of the document and its meta-information. Each of these document containers, referred to as common analysis structures (CAS), are passed from one software-annotator to the next in sequence within the pipeline. The document container consists of the text that is immutable, and a list of stand-off annotations or mark-ups that reference offsets within the text. The software-annotators add, edit or delete annotations. that decomposes the text of input files into constituent parts including tokens and sentences. The employed pipeline contains annotators that do an admirable job of identifying semi-templated components, including clinically significant sections, section headings, slots and their values, questions and answers, and to a lesser extent, checkboxes.

The pipeline described in this paper includes a dictionary-lookup software-annotator that provides the bulk of the computational effort that the rest of the pipeline relies upon (see Figure 1).

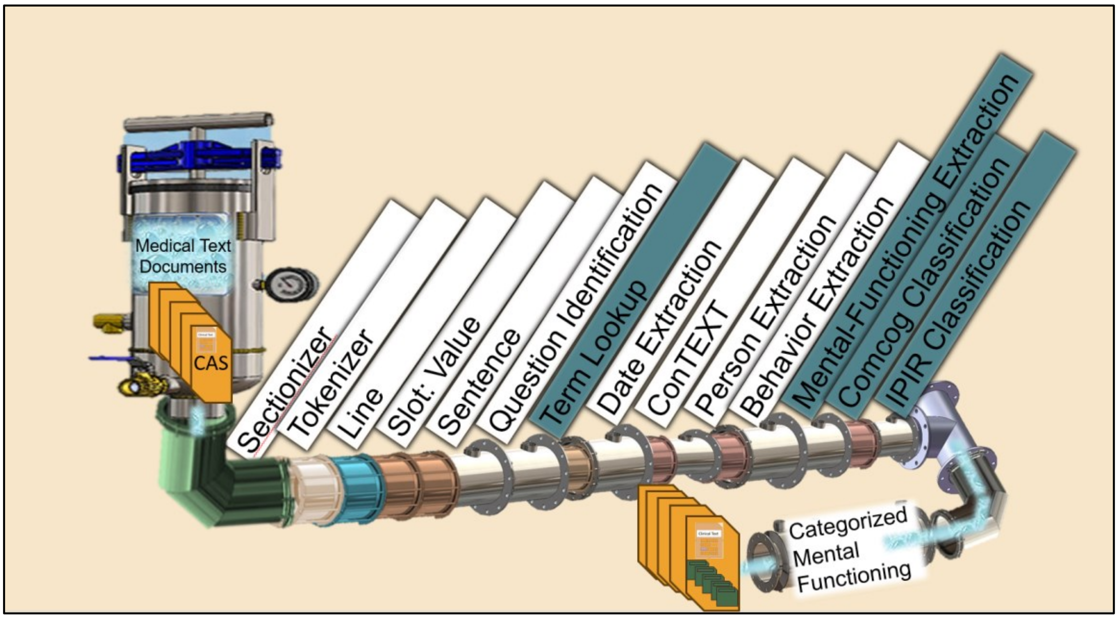

FIGURE 1: *Mental-functioning* NER Pipeline

This dictionary-lookup has four significant qualities worth mentioning. Dictionary entries can be multi-word keys, called terms, allowing terms like *diabetes mellitus* to be chunked together into a single unit. The lookup algorithm chunks from right to left in a greedy style within sentential boundaries. While (human) English readers read left to right, invoking the parsing from right to left within a sentence favors chunking multi-word terms more accurately due to the nature of English multi-word terms where the head or most important word of the term is the last word. The dictionaries, elsewhere called lexicons, are crafted and curated for specific purposes, including finding general English terms, to dictionaries to support finding disease names, symptoms, proper names and the like. Each dictionary entry includes meta information that is carried along with matches, most importantly, potential semantic categories by which the entry can be classified. For example, the entry for *listening* has the category "Communication" associated with it. It is the categories associated with matched terms that downstream software-annotators utilize.

Dictionaries have been crafted and deployed in the general pipeline to chunk terms that are known to not be relevant to downstream tasks, for example, a dictionary of people's names, a dictionary of place names, a role dictionary, dictionaries for behavior and emotion to name a few. Non-dictionary terms are also found using software-annotators that are regular-expression-based to find date, time and number expressions. These pattern-based chukers result in terms that get classified via the pattern that picked them up.

Task-specific software annotators work off document decomposition annotators that identify sentences, section slot:values, and terms. These downstream software-annotators follow the general pattern of looping through each section, and within each section, looping through each sentence or slot: value, or question/answer, and within each of those, looping through each term looking for relevant term categories. The offset boundaries of terms that match the sought after categories then are used to create new machine-created annotation instances with specific labels.

The NER framework can take formatted input documents from text, to csv, and General Architecture of Text Engineering[18] (GATE) types and output not only UIMA flavored xml, but also GATE formats. For this paper, GATE formatted manual annotations were the input format and both GATE and UIMA formats were output.

2.3 Manual Annotations on a Clinical Corpus

OCR processed pages were sampled from a large set of disability claims records that had been provided by the Social Security Administration (SSA) for tasks related to this work[2]. The pool of 65,514 PDF documents included almost 2 million pages. The original source was 16,000 adult disability claims with the inclusion criteria of: allegations of musculoskeletal, neurological, or mental impairments; a disability determination decision issued between 2013 to 2018; and spanning five geographically diverse regions of the United States. Many documents in this collection included notes from sequential clinical encounters over time with a particular healthcare provider. The quality of the OCR varied requiring us to develop a mechanism to assess and filter out poorly OCR'd documents[19]. While SSA did provide categories for the documents, they were not classifications related to traditional clinical document types, such as History and Physical, or Discharge Summary. Domain experts provided a word-list of relevant ComCog and IPIR terms from which to pull pages. Samples were created from word frequency lists from the OCR'd pages based on frequency distributions of relevant words found in the pages of text. Records were selected following a uniform distribution whose range included some records having many relevant words and some records having only a few relevant words.

Guidelines were created in conjunction with the EMFO to codify how the samples get annotated. Domain experts, including physical and occupational therapists and a medical doctor, manually annotated using the developed guidelines. The guidelines used for the ComCog and IPIR annotation tasks are published on Zenodo[20 21]. Each page was annotated by one of two domain experts working on the task. Because there were only resources sufficient to have the samples annotated by one annotator, an inter-annotator agreement task was performed across the annotators working within each sample to ensure consistency before the full sets were annotated. See Table 3c for the inter-rater reliability (IRR) results for the ComCog sample, and Table 4c for the IRR results for the IPIR set. For the ComCog task, 324 SSA pages were manually annotated, including 24 IRR pages, 240 training pages and 60 pages used for testing. For the IPIR task, 1011 SSA pages were manually annotated, including 51 pages for IRR, 780 pages for training, and 180 pages for testing.

3 Methods

The work that is described in this paper represents the task of further classifying sentences that previously had been identified with *mental-functioning* content from ComCog and IPIR categories via leveraging terminology derived from the EMFO. This work relies on prior annotations done in an upstream process that used a modified spaCy tokenizer/sentence splitter[22] and relied on manual *mental-functioning* annotations on those sentences. The sections that follow are descriptions and specifics of the downstream annotators that identified and created the *mental-functioning* annotations.

3.1.1 Mental-functioning NER Pipeline and Annotator

The *mental-functioning* NER pipeline (see Figure 1) includes software-annotators that create evidence of behavior, emotion and support (financial, social, institutional) from the terms in these semantic categories. This is followed in the pipeline by a software-annotator that creates *mental-functioning* evidence and *mental-functioning* mention mark-ups. The Mental-Functioning Ontology Annotator identifies and creates *mental-functioning* evidence mark-ups from terms that have semantic categories that generalize up to *mental-functioning* within the EMFO. Specifically, a *mental-functioning* mention encapsulates the outer spans of any EMFO evidence found within the scope of a sentence, or slot:value, or checkbox.

The basic algorithm within the annotator is over-generative, and filters are applied to weed out spurious mentions. There are ad-hoc filter annotators in the pipeline to filter out documents that include SSA forms, where a templated text contains what looks like *mental-functioning* mentions, but are not, since the mentions are within templated text instructions to providers on how to fill out the form. Similarly, mentions that can be attributed to the author of the document, not the client or patient, are filtered out. These mentions have evidence that includes the use of pronouns such as *we* and *our* along with use of pronouns such as *you*, *your*, *his*, *her*, *she*, and *him*. These mentions are also tagged with *provider* attribution as a feature at the term and sentence level.

The Mental-Functioning Ontology software-annotator creates *ComCog_yes* and *IPIR_yes* mark-ups for sentences that contain such content for sentences that have not otherwise been filtered out. This NER can take input from previously machine-annotated documents with *ComCog_yes*, and *IPIR_yes* that were either manually annotated from machine segmented sentences or from the Bidirectional Encoder Representations from Transformer[23] (BERT) based models trained on the manual annotations. The NER can utilize these upstream broad classifications in addition to the evidence gathered to this point in the pipeline.

3.1.2 Communication & Cognition (ComCog) Software-annotator

The ComCog software-annotator looks for terms that the lexicon has categorized with a ComCog subcategory within the span of the *ComCog_yes* sentence. When found, a subcategory evidence markup is created, and a subcategory markup is created for the span of the *ComCog_yes*. The ComCog Sub-Categorization annotator further sets 20 attributes to each *ComCog_yes* annotation. For each *ComCog_yes* sentence span, all the EMFO generated evidence that generalizes to ComCog activities from the Ontology, along with behavior and support evidence that cover that span, are gathered and if there is any of these found, further processed. There are 16 ComCog attributes taken from 3 chapters in the ICF activities and participation component having to do with activity adaption, cognition, and communication, and 4 attributes specifically requested by the SSA use case (see Table 1)[17].

Table 1: ComCog categories

ComCog Categories **ICF Codes and Code Name**	**ICF Code Description**
d110-d129 Purposeful Sensory Experiences	Includes watching and listening and other basic senses intentionally to experience stimuli
d130-d159 Basic Learning	Acquiring language, rehearsing, acquiring information, learning to read, write or calculate
d160 Focusing Attention	Intentionally focusing on specific stimuli, such as by filtering out distracting noises
d163 Thinking	Formulating and manipulating ideas, concepts, and images
d166 Reading	Performing activities involved in the comprehension and interpretation of written language for the purpose of obtaining general knowledge or specific information
d170 Writing	Using or producing symbols or language to convey information
d172 Calculating	Performing computations by applying mathematical principles to solve problems that are described in words and producing or displaying the results
d175 Solving Problems	Finding solutions to questions or situations by identifying and analyzing issues, developing options and solutions, evaluating potential effects of solutions, and executing a chosen solution
d177 Making Decisions	Making a choice among options, implementing the choice, and evaluating the effects of the choice
d179 Applying Knowledge Other	Represent information about managing money
d210-d220 Undertaking Tasks	General aspects of carrying out single or multiple tasks
d230 Carrying Out Daily Routine	Carrying out simple or complex and coordinated actions in order to plan, manage and complete the requirements of day-to-day procedures or duties
d240 Handling Stress	Carrying out actions to manage and control the psychological demands required to carry out tasks demanding significant responsibilities and involving stress, distraction, or crises
d310-d329 Receiving Comm	Receiving spoken, nonverbal, written or sign language messages
d330-d349 Producing Communication	Speaking, non-speech vocal expression, singing, or producing non-verbal, written, or sign language
d350-d369 Conversation	Conversations, discussion, or using communication devices
SSA Variables	**Variable description**
Applied Memory	Memory observed in everyday communication and cognition
Adaptation	Responding appropriately to novel and unexpected situations, demands and changes while performing communication and cognitive activities
Pacing	The speed of which information is received, processed and a response through motor actions are generated
Persistence	The ability to stick with a task over time (including work tasks) and sustaining an activity over a period of time for a cognitive reason

The ComCog annotator creates evidence mark-ups for each term it deems ComCog evidence, and the sentence span with the evidence in it (that has not been filtered out by non-relevant counter evidence) is used to create one or more of the ComCog sub-category mark ups.

3.1.3 Non-Relevant Evidence

There is a check to throw out mentions that are not IPIR, ComCog, or *mental-functioning* related due to some known context. For example, if the sentence has the word *invoice*, or *billing* in it, this is not an EMFO mention. There were 1389 terms marked as *NotEMFO* which include *case number*, *department of social services*, *social security act,* along with terms like *swelling* and *please*.
These non-relevant evidence terms were garnered and added from frequency distributions of terms appearing in false-positive mentions and never appearing in true-positive or false-negative mentions.

3.1.4 ComCog and IPIR Dictionary/Lexicon Creation and Curation

Vocabularies from the UMLS[24] with potential of having relevant *mental-functioning* information were reviewed. Beyond the ICF[17], these included the MeSH[16], SNOMED-CT[15], Medical Dictionary for Regulatory Activities[25] (MedDRA), and Thesaurus of Psychological Index Terms[26] (PSY). Seed concepts from each of the EMFO's 4 quadrants (i.e., input, throughput, output, and feedback) were identified to traverse through the UMLS as a whole. The UMLS was systematically traversed through descending hierarchical and non-relationships as they existed via the content within MRREL and MRHIER[27]. Paths were followed based on UMLS concept-level relationships, rather than the more exact or accurate method of only following relationships asserted within and among the same source at the atom level. This approach was an effort to broaden our ability to extract terms at the expense of retrieving some non-relevant terms during development. Manual culling steps were taken to remove the more egregious fallacious terms afterwards. Some seed term areas were fruitful, picking up much content. Other seed terms were less fruitful, picking up terminology that is still covered only by the ICF.

3.1.4.1 VerbNet and Additional Coverage

In the initial pilot annotation tasks, the domain expert annotators were highlighting observable behaviors within the sample clinical corpus, but these functioning mentions had no coverage within any UMLS sources. They were, not surprisingly, mostly verb and adverbial phrases. VerbNet[28] is a catalog of English Verb classes, by thematic role which covers almost all known English verbs. Within VerbNet, the annotators identified 162 of the 322 classes related to *mental-functioning* and then subsequently classified each class back as either ComCog or IPIR. The annotators subsequently sub-classified the VerbNet classes with ComCog and IPIR categories, resulting in an additional lexicon used within the NER. This addition greatly increased coverage.

3.1.4.2 Terminology from the Manual Annotations

The lexicons were augmented and tuned by utilizing terms within the manual annotations which had no coverage in the initial dictionaries, (i.e., no UMLS coverage) using only terms discovered in the training sets, in particular, discovered from a review of the NER's false-negatives. These annotations gave us the opportunity to further curate the lexica by culling terms which spuriously identified *mental-functioning* from a review of the high-frequency false-positives in the training set.

3.1.5 Interpersonal Interactions and Relationships (IPIR)

The IPIR Sub-Categorization annotator further sets 8 attributes to each *IPIR_yes*. For each *IPIR_yes* sentence span, all the EMFO generated evidence that generalizes to IPIR Activities from the Ontology, along with IPIR Participation evidence, behavior and Support evidence that cover that span are gathered, and if there is any of these found, further processing is done (see table 2)[29].

Table 2: IPIR Categories

IPIR Categories **ICF Code and Code Names**	**ICF Code Description**
d710-d729 General Interpersonal Interactions	Interacting, maintaining and managing interactions with people in a contextually and socially appropriate manner, such as by showing consideration and esteem when appropriate, or responding to the feelings of others as well as, as by regulating emotions and impulses, controlling verbal and physical aggression, acting independently in social interactions, and acting in accordance with social rules and conventions
d730 Relating with Strangers	Engaging in temporary contacts and links with strangers for specific purposes, when asking for directions or other information, or making a purchase
d740 Formal Relationships	Creating and maintaining specific relationships in formal settings
d7400 Relating with Persons in Authority	Creating and maintaining formal relations with people in positions of power of a higher rank or prestige relative to one's own position
d750 Informal Social Relationships	Entering into relationships with others, such as casual relationships with people living in the same community or residence, or with co-workers, students, playmates, people with similar backgrounds or professions.
d760 Family Relationships	Creating and maintaining kinship relationships, such as those with members of the nuclear family, extended family, foster and adopted family and step-relationships, more distant relationships such as second cousins, or legal guardians.
d770 Intimate Relationships	Creating and maintaining close or romantic relationships between individuals, such as husband and wife, lovers or sexual partners
d779 IPIR, other	General relationships where no particular relationship can be specified such as “others”, “other people”, “crowds”, “someone”, etc.

The IPIR annotator creates evidence mark-ups for each term it deems IPIR evidence, and the sentence span with the evidence in it (that has not been filtered out by non-relevant counter evidence) is used to create one or more of the IPIR sub-category mark-ups.

3.1.6 IPIR Heuristics

The pipeline includes software-annotators for each of the IPIR categories. Each software-annotator works roughly the same, where the terms of each sentence of each section are iterated thru accumulating evidence to be acted upon for each sentence and each section. Below are specifics about what evidence each software-annotator is looking for and triggered on.

ICF d710-d729 General Interpersonal Interaction mark-ups are created when there is emotion evidence present. Interaction mentions are also created when the sentence span included person evidence and behavior evidence.

ICF d730 Relating with Stranger mark-ups are created when there is evidence of stranger relationship evidence. These kinds of terms come from a lexicon of 100 terms like *shopper*, *public places*, and *visitor*. The bulk of these terms were garnered from examples seen in the data.

ICF d740 and d7400 Formal Relationship mark-ups are created when terms categorized as non-authority or sometimes-authority are present. These kinds of terms come from a person-role and title lexicon that was augmented with_non-authority, sometimes-authority and authority-position semantic categories, garnered from the Bureau of Labor Statistics and other public sources[30-32]. Those terms categorized as authority-position had d7400 mentions made.

ICF d750 Informal Social Relationship mark-ups are created when there is informal relationship evidence found in the span of each *IPIR_yes* sentence. If there is no such evidence, but there is the presence of pronouns, an informal social relationship markup was made. This works only because we already know this is an *IPIR_yes* sentence from an upstream software-annotator, so cases where the pronouns relate to possessions rather than to other people would have already been filtered out. The UMLS terms categorized with family history and person roles were the impetus for the informal relationships lexicon employed here.

ICF d760 Family Relationship mark-ups are created when there is family history evidence found in the span of the *IPIR_yes* sentence. The underlying terminology used here are the terms in the UMLS categorized with their Family-History semantic type.

ICF d770 Intimate Relationship mark-ups are created when there is intimate relationship evidence found in the span of the *IPIR_yes* sentence. A small lexicon of 130 terms was created mostly from the guideline examples, observations, introspection and use of thesauri to garner adjectives and nouns that indicate an intimacy. Such terms include *significant other*, *old lady*, and *inseparable*.

ICF d779 Particular interpersonal relationships, other specified and unspecified mark-ups are also created when there is “other” relationship evidence found in the span of *IPIR_yes* sentence. There were only 30 or so terms identified with this tag, which included *others*, *nobody*, and *withdrawn*. The guidelines and training data were source for this terminology.

4 Results

Note that the annotations and efficacy were done at the sentence level: See table 3a for the ComCog test efficacy. Table 3b shows the inter-rater reliability for each category. See table 4a for the IPIR test efficacy, 4b for the IRR. The test, training and IRR sets were distinct draws from the larger SSA set.

Table 3a: SSA **Test** Sample ComCog Subcategory Efficacy

Table 3b: SSA **IRR** Sample ComCog

	SSA Test Sample Results				**IRR Results**	
ComCog Subcategories	**F-Score**	**Recall**	**Precision**	**Support**	**F-Score**	**Support**
d110-d129	0	0	0	0	0	0
d130-d159	0.77	0.71	0.83	7	0	1.5
d160	0	0	0	5	0	0.5
d163	0.4	0.41	0.38	41	0.43	13
d166	0.57	1	0.4	2	0.67	1.5
d170	0.75	0.75	0.75	4	0	0
d172	0	0	0	0	0	0
d175	0.22	0.33	0.17	3	0	0
d177	0.49	0.59	0.42	27	0.81	16
d179	1	1	1	1	0	0
d210 d220	0.36	0.52	0.27	29	0.93	33.5
d230	0.29	0.2	0.5	5	0	0
d240	0	0	0	3	0	0
d310-d329	0.59	0.56	0.62	36	0.63	30
d330-d349	0.86	0.93	0.79	216	0.93	55
d350-d369	0.6	0.65	0.56	23	0.96	11.5
Applied Memory	0.33	0.43	0.27	37	0.81	30
Adaptation	0	0	0	4	0	1
Pacing	0	0	0	3	0	0
Persistence	0.29	0.25	0.33	4	0	0.5
Micro Avg	0.58	0.64	0.56		0.77	
Macro Avg	0.38	0.42	0.36	22.5	0.31	9.7
Weighted Avg	0.64	0.7	0.59		0.8	
Blue font	Indicate there were no IIR results for this category, implying that it is not known if this is a cognitively difficult task or not					
Grey cell backgrounds	Indicate either that the task was cognitively difficult or there was little exposure to examples to train on					
Peach cell backgrounds	Highlight categories that were seen in the IRR and adequate training examples					

Table 4a: SSA Test Sample IPIR Subcategorization Efficacy

Table 4b: SSA IIR Sample Subcategory Results

IPIR Subcategories	SSA Test Sample Results F1 score	Precision	Recall	Support	IRR Results F1 score	Support
d710-d729 General Interpersonal Interactions	0.72	0.77	0.68	247	0.88	61.5
d730 Relating with Strangers	0.54	0.47	0.64	15	0	1
d740 Formal Relationships	0.75	0.77	0.72	203	0.91	47.5
d7400 Relating with Persons in Authority	0.52	0.47	0.58	15	1	2
d750 Informal Social Relationships	0.59	0.61	0.57	101	0.91	17.5
d760 Family Relationships	0.74	0.72	0.75	93	0.86	18.5
d770 Intimate Relationships	0.67	0.63	0.7	41	0	1
d779 IPIR, other	0.49	0.72	0.37	79	0.86	18.5
Micro Avg	0.67	0.63	0.72		0.88	
Macro Avg	0.63	0.65	0.63	99.25	0.68	27.92
Weighted Avg	0.68	0.65	0.72		0.87	

5 Error Analysis

This was a challenging set of tasks. The OCR was very noisy, and many documents included multi column formats where the OCR would read across the columns, scrambling sentence segments. We tried two sentence segmenters, neither of which was attuned to the OCR challenges. One segmenter did better on prose sentence boundaries but poorer on correctly chunking semi-structured text that included slot: values, questions and answers, and checkboxes. Consequently, one source of error was the slot or the fact that the question was in one sentence and the answer was in another. As the task was bounded by looking for sentences that had mentions in them, single word sentences with the answer *yes* in them routinely got missed by the machine. From forms that had check box answers or hand-writing we did not pick up what was checked or handwritten answers. Several failures also were the result of out-of-dictionary spellings due to missing or transposed character OCR transformations. One source of false positives were the instructions and information sections of forms, where activity mentions were found, but they were not about the patient. There were multiple categories that had little or no representation either in the initial inter-annotator agreement set, training set, and/or the test set. This raised issues when categorized mentions showed up in the test set, but the category did not show up in the training set. No priority had been taken to augment terminology around those kinds of mentions. Lastly, there were no resources to alter the annotations found by the NER that the annotators upon review noted to be true positives. SSA provided guidance to favor recall, so the NER was thus tuned to be over-generative.

6 Discussion

Beyond the OCR and segmentation issues, this task was not just difficult for the NER, the IRR task also indicated that except for five majority classes within ComCog, the rest of the 20 categories had too little exposure or were very difficult for the annotators to agree upon. The two ComCog majority class exposure on their own raised the average across all the classes giving a misleading confidence that this task has an acceptable level of difficulty. The IPIR task was both cognitively more achievable and had enough context coming through the noise to achieve reasonable machine performance.

Going forward, when the next corpus is analyzed, data selection should be augmented to include at least a minimum number of mentions within each category to train from, and a feedback task to review output from the NER so as to alter mentions that were suggested but got missed. It is not too surprising that there were low cell counts for multiple ComCog categories, as *mental-functioning* is rarely mentioned in clinical records that are not specifically mental health notes. As we could not pull specific note types, our ability to create a data set by term extraction was hampered. However, the task did point out that *mental-functioning* shows up at some level of capture, even in very noisy data.

7 Future Work

The quest to refine *mental-functioning* extraction and classification within clinical text continues with efforts underway to apply this NER to a more traditional clinical corpus without the OCR and document type issues. Further, efforts are proceeding to make the text, manual annotations for this next corpus, as well as the NER tuned to this additional corpus available.

8 Conclusions

Mental-functioning extracted and characterized from clinical documentation is a valuable source of functional data for purposes such as disability determination, health research, health outcomes evaluation, and it is information that clinicians express interest in having. However, mental-functioning activity information can be challenging to identify in health records due to the over emphasis of documentation at the mental body function level.

The EMF ontology was developed to aid in accurately finding these mentions through semantic information that distinguishes mental-function from evidence of *mental-functioning*. A rule-based NER was developed from an existing codebase, leveraging the EMFO to extract and classify *mental-functioning* mentions. This NER is benchmarked on sampled pages pulled from OCR'd SSA claimant records. The weighted average ComCog Subclassification F-Score on this set was 0.63. The weighted average IPIR subclassification F-Score was 0.63. The benchmarked NER is one part in a series of tasks to provide the informatics community with resources related to functioning.

9 Acknowledgements

9.1.1 Author contributions

GD oversaw the development of the EMFO and NER. BD oversees the deliverables and provided analytics. MS developed the EMMF and the EMFO, is one of the domain experts, and an annotator

of this corpus. KC developed the EMFO, is one of the domain experts and an annotator. RJS provided statistics, is a domain expert and an annotator of this corpus. RP is a domain expert and annotator of this corpus. ER is the head of this branch, is a domain expert, provided the framing for this task.

9.1.2 Supplementary material

9.1.3 Funding

This work was funded by the Intramural Research Program of the National Institutes of Health and the US Social Security Administration.

9.1.4 Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9.1.5 Software Availability

The NER and EMFO ontology is on github, available at:
https://github.com/CC-RMD-EpiBio/EcologicalMentalFunctioningOntology

9.1.6 Data Availability

The data used for this study came from the Social Security Administration. Our organization does not have permission or rights to share it. Our organization has no intellectual property rights or privileges to the data we were given access to. In addition, this data contained sensitive information including PHI and PII which required additional security and access safeguards be put into place prior our use of it.

10 References

1. Goldman HH, Porcino J, Divita G, et al. Informatics research on mental health functioning: decision support for the social security administration disability program. Psychiatric Services 2023;**74**(1):56-62.
2. Zirikly A, Desmet B, Newman-Griffis D, et al. Information extraction framework for disability determination using a mental-functioning use-case. JMIR Medical Informatics 2022;**10**(3):e32245.
3. Divita G, et al. Ecological Mental-functioning Ontology. Secondary Ecological Mental-functioning Ontology 2024. https://github.com/CC-RMD-EpiBio/EcologicalMentalFunctioningOntology/.
4. Sacco MJ, Divita G, Rasch E. Development of an ontology to characterize mental-functioning. Disability and rehabilitation 2024;**46**(16):3739-48.
5. Building an ecological mental-functioning ontology: An Informatics Perspective. Proceedings of the International Conference on Biomedical Ontologies 2024; Enschede, The Netherlands.
6. Divita G, Coale K, Maldonado JC, Silva RJ, Rasch E. Extracting body function information using rule-based methods: Highlighting structure and formatting challenges in clinical text.

Frontiers in Digital Health 2022;**4**:914171.
7. Divita G, Carter ME, Tran L-T, et al. v3NLP framework: tools to build applications for extracting concepts from clinical text. eGEMs 2016;**4**(3):1228.
8. Sophia: a expedient UMLS concept extraction annotator. AMIA Annual Symposium Proceedings; 2014.
9. Divita G, Luo G, Tran L-TT, Workman TE, Gundlapalli AV, Samore MH. General symptom extraction from VA electronic medical notes. MEDINFO 2017: Precision Healthcare through Informatics: IOS Press, 2017:356-60.
10. Divita G, Brignone E, Carter ME, et al. Extracting sexual trauma mentions from electronic medical notes using natural language processing. MEDINFO 2017: Precision Healthcare through Informatics: IOS Press, 2017:351-55.
11. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering 2004;**10**(3-4):327-48.
12. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association 2010;**17**(5):507-13.
13. Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association 2010;**17**(3):229-36.
14. Failure analysis of MetaMap transfer (MMTx). MEDINFO 2004; 2004. IOS Press.
15. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics 2006;**121**:279.
16. Lipscomb CE. Medical subject headings (MeSH). Bulletin of the Medical Library Association 2000;**88**(3):265.
17. WHO. *International Classification of Functioning, Disability and Health*. Geneva: World Health Organization.
18. Cunningham H. GATE, a general architecture for text engineering. Computers and the Humanities 2002;**36**:223-54.
19. Angelfish: Building a Structurally Diverse Clinical Document Corpus. AMIA; 2020.
20. Section NCREB. Annotation Guideline for Free-Text Activity Functioning Information: Communication & Cognition. Secondary Annotation Guideline for Free-Text Activity Functioning Information: Communication & Cognition 2025a. https://zenodo.org/records/13910167.
21. Section NCREB. Annotation Guideline for Free-Text Activity Functioning Information: Interpersonal Interactions & Relationships. Secondary Annotation Guideline for Free-Text Activity Functioning Information: Interpersonal Interactions & Relationships 2025b. https://zenodo.org/records/13774684.
22. Vasiliev Y. *Natural language processing with Python and spaCy: A practical introduction*: No Starch Press, 2020.
23. Koroteev MV. BERT: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943 2021.
24. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research 2004;**32**(suppl_1):D267-D70.
25. Jaasu NM, Kamaraj R, Seetharaman R. MedDRA (medical dictionary for regulatory activities). Research journal of pharmacy and technology 2018;**11**(10):4751-54.

26. Gallagher LA. Thesaurus of psychological index terms. (No Title) 2005.
27. NLM. UMLS Reference Manual [Internet]. Secondary UMLS Reference Manual [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK9685/.
28. Palmer M, Bonial C, Hwang JD. 17 VerbNet: Capturing English Verb Behavior, Meaning, and Usage. The Oxford handbook of cognitive science 2016:315.
29. Organization WH. *International Classification of Functioning, Disability, and Health: Children & Youth Version: ICF-CY*: World Health Organization, 2007.
30. Job titles for social workers. Secondary Job titles for social workers. https://www.wcupa.edu/education-socialwork/gradSocialWork/documents/socialworkjobtitles.pdf.
31. U.S. Bureau of Labor Statistics 2010 Census Occupational Titles and Code List. Secondary U.S. Bureau of Labor Statistics 2010 Census Occupational Titles and Code List 2010. https://www.bls.gov/cps/cenocc2010.htm.
32. ONGIG Job Titles: The Difinitive Guide. Secondary ONGIG Job Titles: The Difinitive Guide 2025. https://www.ongig.com/job-titles#/#hierarchy.