Interactive Cohort Identification of Sleep Disorder Patients Using Natural Language Processing and i2b2

W. Chen1; R. Kowatch2; S. Lin1; M. Splaingard3; Y. Huang1

1Research Information Solutions and Innovations; 2Center for Innovation in Pediatric Practice; 3Sleep Disorder Center, Nationwide Children’s Hospital, Columbus, OH

Keywords
Sleep disorder, cohort identification, natural language processing (NLP), i2b2, clinical ontology

Summary
Nationwide Children’s Hospital established an i2b2 (Informatics for Integrating Biology & the Bedside) application for sleep disorder cohort identification. Discrete data were gleaned from semi-structured sleep study reports. The system showed to work more efficiently than the traditional manual chart review method, and it also enabled searching capabilities that were previously not possible.

Objective: We report on the development and implementation of the sleep disorder i2b2 cohort identification system using natural language processing of semi-structured documents.

Methods: We developed a natural language processing approach to automatically parse concepts and their values from semi-structured sleep study documents. Two parsers were developed: a regular expression parser for extracting numeric concepts and a NLP based tree parser for extracting textual concepts. Concepts were further organized into i2b2 ontologies based on document structures and in-domain knowledge.

Results: 26,550 concepts were extracted with 99% being textual concepts. 1.01 million facts were extracted from sleep study documents such as demographic information, sleep study lab results, medications, procedures, diagnoses, among others. The average accuracy of terminology parsing was over 83% when comparing against those by experts. The system is capable of capturing both standard and non-standard terminologies. The time for cohort identification has been reduced significantly from a few weeks to a few seconds.

Conclusion: Natural language processing was shown to be powerful for quickly converting large amount of semi-structured or unstructured clinical data into discrete concepts, which in combination of intuitive domain specific ontologies, allows fast and effective interactive cohort identification through the i2b2 platform for research and clinical use.

Correspondence to:
Yungui Huang PhD MBA
Research Information Solutions and Innovations
Nationwide Children’s Hospital
575 Children’s Crossroad
Columbus, OH 43215 United States
E-mail: Yungui.Huang@nationwidechildrens.org
Introduction

Cohort identification is commonly used to find patients with shared characteristics. It is important for early identification of disease risks [1–4] and patient recruitment for clinical trials [5–7]. Cohort identification usually requires searching a large clinical database for a small subset of subjects; therefore, it is often time consuming. The process could be more expensive when manual chart reviews are needed to confirm diagnosis or other clinical features using unstructured data such as natural language-based clinical notes [4, 8–10]. The labor-intensive nature of cohort identification creates major barriers for time-critical decision making in clinical practice [11].

Natural language processing (NLP) techniques have been widely used to quickly analyze a large amount of texts in support of humans [12]. Techniques such as part-of-speech tagging, parsing, and named entity recognition (NER) may speed up the process of identifying diagnoses, procedures and medications in clinical notes with satisfactory accuracy [13–15]. Syntactic structures obtained from parse trees provide useful indicators for extracting concepts and values from natural language sentences as well as for building queries [16]. Knowledge frameworks such as Unstructured Information Management Architecture (UIMA), clinical Text Analysis and Knowledge Extraction System (cTAKES) and MetaMap Transfer (MMTx) have been widely used to provide terminology services with a natural language processing system [17–21]. Unified Medical Language System (UMLS) often serves as the backend knowledge base for clinical text processing purposes [22]. Machine learning-based approaches have also been integrated into NLP systems for clinical NER tasks [23–25]. However, their overall efficacy may be compromised sometimes by the lack of in-domain knowledge possessed by human experts [26, 27].

Sleep study summary reports (polysomnograms) are semi-structured documents that require either manual or automated natural language analysis. These documents contain information extracted from different sources such as clinical charts, GRASS® sleep study programs, and human edits [4]. Previous methods of finding cohorts for sleep study included searching sleep study documents using keyword-based file search functions on a Windows system. Human experts later manually verified these documents and pulled useful data from validated corresponding documents. This manual error-prone process may take up to 60–70 hours to finish. It is highly necessary to automate the process so that cohorts can be identified more effectively for various sleep research purposes.

I2b2 is an interactive medical informatics system that is widely used for patient cohort identification [28, 29]. It features a user-friendly web-based interface, an ontology browser and a search tool [30]. Although i2b2 has been used in various clinical applications, it is still new to sleep study, especially with unstructured sleep report data [31, 32]. I2b2 comes with its own natural language processing component as optional but this component is limited to specific purposes and is not flexible enough to produce discrete data [33].

This study aimed to leverage i2b2’s rich self-service query builder functionalities. It attempted to overcome i2b2’s NLP limitations by building a customized NLP workflow to populate i2b2 tables. It also constructed in-domain ontologies to speed up cohort identification by defining queries through dragging and dropping concepts into a web-based query interface. Natural language processing techniques were employed to extract both standard and non-standard clinical terms from sleep study documents collected at Nationwide Children’s Hospital for the past 11 years. The accuracy of natural language processing results was evaluated using standard precision, recall and F1 measures against results provided by human experts. Screenshots were included to illustrate the use cases of the system. The contribution of this study included a working demo of the i2b2 self-service cohort identification system for sleep study and easily adaptable source code for parsing text data for new applications.

Data

15,683 sleep study reports were collected from January 2004 to September 2014. Figure 1 represents the distribution of sleep study reports over the years. Considering the fast-growing number of documents, there is urgent need from our physicians to automate the cohort identification process by using informatics system to overcome the poor performance of traditional manual means.
Sleep reports were generated through a pipelined process (▶ Figure 2). Basic patient information was collected from smart text in Epic™ (Epic Systems, Verona, WI), a commercial medical informatics system. These smart texts are annotations that wrap a collection of information into one short phrase. For example, typing the smart phrase #vitals# will pull all vitals information of a patient. Sleep study lab results were obtained from the GRASS™ system (Grass Instruments, Quincy, MA) that was used to assess the quality of sleep using sensors and monitors. These results mainly contained data such as sleep architecture, respiratory and EEG in both tabulated and free text formats. In the sleep domain, sleep architecture, for example, represents the cyclical pattern of sleep between different sleep stages, including non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. Finally, both Epic text and sleep data were transferred into a Word template file for reporting.

All documents had sections listed in ▶ Table 1. Each section contained concepts that could be either text or numeric. Textual concepts were concepts of diagnosis, procedure and medication while numeric concepts are those with numeric values. NLP parser and regular expression parsers were developed respectively to extract textual concepts and numeric concepts.

**Methods**

**Document processing flow**

There were several steps in processing documents (▶ Figure 3). First, all Word documents were converted into text format. Second, documents were split into sections as listed in ▶ Table 1. Third, for the indication section where previous diagnosis, procedure and medication information would be extracted, we developed a heuristic classifier to classify sentences into four predefined categories: diagnosis, procedure, medication and others. Sentence classification was necessary to make our terminology parser work correctly.

Two parsers were implemented to extract information from different sections: a regular expression parser and a NLP parser. For sections that contained only numeric concepts a regular expression (RE) parser was implemented. In this case, concept names and values were adjacent in the text. For sections that contained only textual concepts, a NLP parser was implemented. Medical terminologies were extracted using a tree parser. In this case, the concept name was the terminology name and the concept value was the same as the concept name. For sections that contained both types of concepts, both parsers were employed.

Once terminologies were extracted, they were matched against Epic terminologies, an ad hoc terminology database we used for classifying terminologies into standard and non-standard. All those with exact matches in EPIC terminology database were tagged as standard and others as non-standard.

**Regular expression parser and numeric concept extraction**

A regular expression parser was implemented by extracting concepts and values that followed certain patterns in a sentence. For example, in the sentence REM apnea index is 0 the concept name is REM apnea index and the concept value is 0. The following regular expression pattern was used to capture such a concept value pair:

```
(?i)rem\s+apnea\s+index\s+(is\s+)?[0–9.]+  
```

This regular expression would capture any caseless text that begins with rem followed by apnea followed by index followed by an optional is and a number. For each concept, we constructed a regular expression parser using the similar pattern-based approach. All numeric concept names were predefined, as they were the same as the numeric concepts from the GRASS sleep study system. In our sleep reports, numeric concepts mainly came from the Lab Results section. As lab results were semi-structured text, we achieved over 98% accuracy of extracting numeric concepts based on a manual evaluation of 100 documents.
Sentence classification

The indication section is the only section that contains textual concepts and requires NLP parsing analysis. Sentences in the indication section were classified into one of the four categories: diagnosis, procedure, medication and others. Below is an example of the sentences in the indication section of a sleep study report.

“This is a 09 yrs 10 mos old white male with a history of sleep onset insomnia, nocturnal awakenings and painful legs at rest for the past two years. (Diagnosis) Previous surgeries include tracheostomy and UPPP. (Surgeries) Current medications include Protonix, Zyrtec, Bactrim and Flovent. (Medications). Wake up time: 09:30am. Bed time: 08:00 pm. BMI: 35. (Others)”

Sentences in the indication sections were classified using heuristic rules by detecting lexical features such as with a medical history of, surgeries include and medication include, among others. Given the limited language patterns of constructing sentences of each category, the overall classification accuracy was above 97% based on 100 manually verified samples.

Using the method above, we observed 99.9% of documents containing diagnosis sentences, 80.2% containing procedure sentences, 92.6% containing medication sentences. We have not tested against machine learning-based classifiers, as our heuristic based classifier was much easier to implement and worked well enough for our purposes.

NLP parser and textual concept extraction

To extract medical terminologies from classified sentences, we parsed the sentence using NLP parsing techniques. We utilized the Stanford Parser Java library to generate the parse tree of a sentence [34]. On the top level of the tree was the original sentence and on the leaf level were the words in the sentence. Nodes in between were phrases such as noun phrases, verb phrases or prepositional phrases. The recursive neural network (RNN) dependency parser was chosen to parse the sentences [35, 36]. This RNN parser is a new parser recently developed by researchers from Stanford University. It is shown to be scalable to variable-sized inputs and aware of sentence context for parsing. Moreover, it leverages information about the semantic structures of a sentence to deal with unseen phrases during parsing [36]. Given that medical reports include a significant amount of phrases that are not common in traditional training corpus such as Wall Street Journals, we found this parser to be a good candidate to suit our needs. The RNN parser turned out to perform better at least in our case than the other two parsers available in the Stanford NLP libraries, the PCFG parser [38] and the factored parser [39]. Parsing a sample of 100 sentences showed that the RNN parser achieved over 90% accuracy, and was about 14.3% more accurate than the PCFG parser and 12.5% more accurate than the factored parser.

Compared with the traditional named entity recognizer, the NLP parser does not have to rely on knowledge base to extract terminologies. It works by examining the syntactic structures of a parse tree; therefore it works independently from the knowledge base. For example, we assume sleep apnea is the exact entry in a knowledge dictionary, the phrase diagnosed with possible sleep apnea is different from diagnosed with sleep apnea. The traditional named entity recognizer relies on terminology dictionary or lexical patterns and therefore is not designed to capture the modifiers such as possible in this case [40]. Such modifiers, however, provide important information for clinical decision-making.
We used the following sentence from our dataset to demonstrate the NLP parsing process: *Patient is a white male with a history of asthma, mild obesity and restless sleep. Although obesity matched exactly to an entry in our knowledge base, restless sleep did not have any exact matches. The fact that they are both noun phrases, however, helps us to capture these terminologies regardless of their existence in the knowledge database. Therefore, by extracting all noun phrases in the sentence, we were able to obtain a set of candidate terminologies. The parser tree of this sentence is shown below:

```
(ROOT
  (S
    (NP (NNP Patient))
    (VP (VBZ is)
      (NP (DT a) (JJ white) (NN male))
      (PP (IN with)
        (NP (DT a) (NN history))
        (PP (IN of)
          (NP (NN asthma))
          (, ,)
          (NP (JJ mild) (NN obesity))
          (CC and)
          (NP (JJ restless) (NN sleep)))))

(, . . )
```

Table 2 shows the noun phrases extracted from the sentence. As we can see, the candidate list includes some overlapping phrases. Overlapped phrases may convey different meanings. By including them, we gave the parser certain flexibility for interpretations. To filter out irrelevant noun phrases, we defined a list of skipped words. This skipped list included 98% of the most frequent noun phrases that were surely not medical terminologies such as *patient*, *history*, *male*, and so on. Even if certain phrases that were not be completely skipped, we may still be able to filter them when we search and select phrases in i2b2.

Given our Epic knowledge base, *Asthma* had an exact match but the other two did not (Table 3). Therefore, they were classified into the two categories for building the ontologies: standard terms and non-standard terms. Both terms will be made available for searching through i2b2.

**Plurality, case, abbreviation and negation handling**

Case variation and plurality may cause unnecessary differentiation of ontologies. For example, *leg pain* and *Legs pain* should be under the same ontology. Case issue was handled by converting all words to their lowercases and plurality was resolved by converting all words into their lemmas. The lemma of a word is the base form a word. For example, the lemma of a plural noun is its singular form. The lemma of a past tense verb is its present tense. Final terminologies were represented as lowercase lemmas. This way, we ended up with a much smaller list of terminologies for easier searching and browsing.

Some abbreviations were actual medical terminologies such as T&A (tonsillectomy with adenoidectomy) while others were not such as *s/p* (status post). Previous approaches compared different classic NLP systems (i.e. MetaMap, MedLEE and cTAKES) for medical abbreviation recognition and concluded with suboptimal overall results [41]. In our case, abbreviations were handled automatically by the NLP parser through the tree parsing process. In most cases, these abbreviation will be tagged as nouns or noun phrases and therefore result in candidate terminologies.

Whether a phrase should be considered as candidate terminology depends on whether it is negated in the sentence. For example, in the sentence *patient does not have leg pain*, *leg pain* should not be considered as a candidate terminology. Negation is handled by a negation library called NegEx [42]. This library can detect negation of a phrase by analyzing the relationship between a negation phrase and its surrounding phrases. In Table 2, we show how NegEx handles negation.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Negation Handling</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Patient is not a white male</em></td>
<td>Negated</td>
</tr>
<tr>
<td><em>Patient is a white male</em></td>
<td>Not Negated</td>
</tr>
<tr>
<td><em>Obesity is mild</em></td>
<td>Negated</td>
</tr>
<tr>
<td><em>Obesity is not mild</em></td>
<td>Not Negated</td>
</tr>
<tr>
<td><em>Sleep is restless</em></td>
<td>Negated</td>
</tr>
<tr>
<td><em>Sleep is not restless</em></td>
<td>Not Negated</td>
</tr>
</tbody>
</table>

Although *leg pain* and *Legs pain* should be under the same ontology, the parser handled them differently because of case variation. This is a limitation of the current NLP system. To overcome this, we propose to use a more robust case-insensitive NLP system that can handle case variation and negation automatically.
trigger (e.g. does not have) and a phrase (e.g. leg pain). Previous experiments showed that this library could reach above 94% of accuracy [42].

**Populating i2b2 data repository**

i2b2 (Informatics for Integrating Biology & the Bedside) is widely used to access clinical data for knowledge discovery [29]. i2b2 backend data repository is based on a special relational database schema called the star schema. A star schema makes concepts and observations more extensible [29]. To use i2b2 for our sleep disorder project, we mapped all extracted concepts and their values to i2b2's star schema tables.

A central fact table and a few dimension tables represent the schema (Figure 4). The central table is the observation fact table that contains all concepts and their values. Five surrounding tables, called dimension tables, provide supplemental information to the concepts in the fact table. The concept table contains all concepts extracted from all documents. The patient table contains all patient information the same as the provider table contains all provider information. The encounter table contains the encounter number field which uniquely identifies each encounter. A visit may have multiple encounters. i2b2 tables relate to each other through foreign keys such as concept code, medical record number, and so on.

All extracted concepts are mapped to the concept table while values of the concepts are mapped to the fact table. In the fact table, there is a type field to indicate whether the concept is a textual or numeric concept. The value of a numeric concept will be number. The value of a textual concept will be the concept itself.

Typically, i2b2 is popular using discrete data collected from a clinical system, clinical encounters or other relational databases. Compared with relational databases, the star schema of i2b2 is famous for its simplified query logic and improved performance against aggregation operations. For further information, one may take a look at the detailed documentations (www.i2b2.org).

**Ontology development**

The i2b2 cohort identification process relies heavily on the design and development of ontologies. The i2b2 ontology is hierarchically organized collection of concepts. It is represented by a concept path and a concept code [28]. A concept path is a \"\" delimited string that separates the concept by different levels of ontologies. From left to right, the ontology becomes more specific. The concept code could be any string that can uniquely identify the concept for example, the ICD9 or ICD10 code.

Concepts in i2b2 also have values that could be either numeric or textual. Consequently, we defined two types of concepts that were extracted from the text: textual concept and numeric concept. Both types of concepts were extracted from sleep documents to build the ontologies for sleep disorder study.

Sleep disorder concepts are grouped into categories such as patient information, indications, sleep architecture, respiratory, ventilation, cardiovascular, EEG and sleep disorder diagnosis. Each category corresponded to the fixed headings of the document and constituted the top-level hierarchy of the ontology path. Under each top ontology were nested ontologies that corresponded to more granulated concepts. The general form for all ontology paths was defined as

\[
\text{\{Section Header\}\{Concept Category\}\{Concept Class\}\{Concept Name\}}
\]

In the indication section, the section header was Patient Indication and Identification. But since the indication section mainly includes medical history information of a patient, we updated the section header name to Medical History in i2b2 to make it more intuitive to browse.

**Ontology building**

The knowledge base we used was a terminology database from Epic systems. The knowledge base was used during ontology mapping but not during ontology extraction. The reason we used such an
ad hoc knowledge base rather than ones in the open domain (such as SNOMED or UMLS) was because physicians in our hospital are all familiar with Epic systems. In addition, we found the terms in our sleep documents are better matches with those in the Epic database. Therefore, using the Epic knowledge base we may develop the most useful ontologies for our patients.

The Epic knowledge base also includes generic information such as classes of diagnosis, procedures and medications. Using this information, we further developed the concept path required by i2b2 as follows:

- Medical History\Diagnosis\standard\diagnosis type\diagnosis name
- Medical History\Diagnosis\non-standard\diagnosis name
- Medical History\Procedure\standard\procedure type\procedure name
- Medical History\Procedure\non-standard\procedure name
- Medical History\Medication\standard\medication type\medication name
- Medical History\Medication\non-standard\medication name

Given the above concept path structure, users are able to query any concept on the path using both concept browser and search functions provided by i2b2. The type information is not part of the parser but part of the ontology. The typing information comes directly from Epic systems. Once the program detects a diagnosis term, all the ontological information related to that term would automatically come from the Epic knowledge base.

Results

Top terminologies

From 15,683 sleep documents we extracted 13,095 patients, 26,550 concepts, and 1.03 million facts. Both textual and numeric concepts were extracted from different sections of sleep documents. From the indication section, previous textual concepts of diagnosis, procedures and medications were extracted. Figure 5 and Figure 6 show the top standard and non-standard medications found among all documents. Albuterol and Singulari are the most common standard medications that can be found in our Epic knowledge base while Periactin and Elavil are the most common non-standard medications.

In the sleep disorder diagnoses section, only the sleep disorder related diagnoses were extracted. Results in Figure 7 showed that snoring and periodic limb movement (plm) were two leading sleep disorder diagnoses. Numeric results were mainly extracted from the lab results sections. For example, highest plm index found is 146 and waso (wake after sleep onset) time ranges from 1 to 596 minutes.

Terminology extraction evaluation

Our evaluation process was based on the comparison of NLP parsed results against human expert annotated results. Standard precision, recall and F1 measures were calculated for a random sample of 100 documents. Our evaluation contained two types of accuracy metrics: exact match accuracy and partial match accuracy. This applies to both standard and non-standard terminologies.

Exact and partial matches were defined against human experts rather than entries in the knowledge base. For example, if obstructive sleep apnea should be extracted according to human experts, the extracted term sleep apnea will only result in a partial match. In most cases, the boundary was clear and easy to identify consecutive words that consisted of a noun phrase.

- **Exact match:** If the term is the same as the human gold.
- **Partial match:** If only a few words from the term match the human gold.

String similarity could be measured using metrics such as edit distance [43] but we found the distance-based similarity measure was not as useful here as in other cases. For example, a larger overlap between a candidate term and human gold does not guarantee a higher rate of correctness. Accord-
ing to our human experts, as long as the partial term retained the same major meaning as the gold, it is a valid partial match. For example, *sleep apnea* and *apnea* are both valid partial matches to *obstructive sleep apnea* but *obstructive* is not a partial match. A partial match was decided by the head noun of the term according to our human experts. Given the availability of our human expert resources, we asked two domain experts to manually annotate concepts in the sentences together at the same time. If they have any disagreement, they will resolve it to their best-agreed result possible. From 100 human annotated sample documents, we obtained 576 diagnosis concepts, 182 procedure concepts and 416 medication concept in total.

Precision, recall and F1 measures were calculated to evaluate the accuracy of exact and partial matches (Table 4). Precision was calculated as the percentage of correct terms over all the terms that were actually found. Recall was calculated as the percentage of correct terms over all terms that should be found. F1 was calculated as $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$. The results were evaluated using 100 new samples.

We also calculated the accuracy of extracted numeric concepts. As they were mainly machine-generated snippets from the GRASS sleep study system, a high accuracy of 99% was achieved based on 100 samples.

I2B2 sleep disorder portal

The ontology browser in i2b2 could be used to select concepts extracted from documents (Figure 8). Each top node was named after the corresponding section heading. Nested nodes corresponded to different levels of ontologies in the concept path. For illustration purpose, a numeric concept was marked with (N) while textual concepts were marked with (T).

The concept search function of i2b2 allowed searching concepts by either names or ICD9 codes. Figure 9 shows the function of searching concepts containing “sleep apnea”. Keyword-based concept search is much faster than using the ontology browser.

Figure 10 shows the interactive query builder for finding any patients with any type of *sleep apnea* and whose *BMI* is greater than 25 and *PLM index* is greater than 0. This resulted in 21 patients in our dataset. These results were the same as those done by manual methods but it was simply much faster. Figure 11 and Figure 12 represent the views of results summary and previous queries. To protect the privacy of our patients, we did not expand their medical record numbers in the view.

Discussion

Cohort identification is an important task for time-critical decision-making. This paper developed an i2b2 application to speed up the process of cohort identification by converting textual information from sleep study documents into discrete concepts. By using our system, physicians may browse different sleep study ontologies and find patient cohorts using an interactive query builder. The system greatly reduced the labor cost of cohort identification and made data more accessible than using traditional keyword-based methods of document search. A lot of “what if” questions can be answered directly in real time instead of waiting for weeks for manual chart review.

Two types of parsers were developed to extract data from documents and were demonstrated to be effective. Regular expression parser was developed to parse semi-structured numeric concepts. It achieved very high accuracy given the fact that they were mainly machine generated. By comparison, the NLP parser parsed sentences into tree structures and each node in the tree corresponded to a meaningful text segment. Based on the part of speech tag of the node, one may filter out all noun phrases as candidate terminologies. A major benefit of using the NLP parser was that it allowed capturing of both standard and non-standard terms, which by default are not supported by traditional dictionary-based named entity recognition methods.

The overall accuracy of the terminology parsing was above 80%. For cohort identification purposes, this was within an acceptable range if physicians did not aim to accurately find all eligible patients. False positives can be further reduced by manual review of a much smaller cohort; although, false negatives are harder to improve. The accuracy was subject to several factors. First, sentences
may be misclassified; however, given the semi-structure nature of our document the misclassification cases were rare. Second, terminologies of different classes may be in the same sentence. This may be the major issue that decreased the accuracy. If a medication name was mentioned in a diagnostic sentence, it was not detected given our current implementation. Finally, sentences that were classified in the others categories may also include diagnoses, medications or procedures. In this case, we entirely missed the terminologies in those sentences. Despite these limitations, we still found the NLP parser to be useful and it could be further improved by implementing more sophisticated classifiers.

**Conflicts Of Interest**
The authors declare that they have no conflicts of interest.

**Human Subjects Protections**
The project is not qualified as Human Subjects research, as defined by the United States Department of Health and Human Services and Food and Drug Administration. Therefore, an application was not submitted to the IRB per our institution’s policy on Quality Improvement projects.

**Acknowledgement**
The authors would like to acknowledge Megan Reynolds, our domain expert, for providing the gold standard, Richard Hoyt for the discussion on i2b2, and Florine Shivers for the editing. We also appreciate the constructive comments from four anonymous reviewers. We also thank the continuous support from Research Information Solutions and Innovations department at Nationwide Children's Hospital.
**Research Article**

**Fig. 1** Number of Reports From 2004 to 2014

**Fig. 2** Sleep Study Reports Generation Pipeline
Fig. 3  NLP Parsing Pipeline

Fig. 4  Star Schema for Sleep Project
Fig. 5  Top standard medications

Fig. 6  Top non-standard medications
Fig. 7  Top sleep disorder diagnoses

Fig. 8  Ontology browser of sleep i2b2 portal
Fig. 9  Search Sleep Apnea Ontologies

Fig. 10  i2b2 Query Interface
Fig. 11  Result view

Fig. 12  Previous query view
### Table 1  Sleep report sections

<table>
<thead>
<tr>
<th>Section name</th>
<th>Structure</th>
<th>Concept types</th>
<th>Example concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Info</td>
<td>Semi-structured</td>
<td>Text and numeric</td>
<td>patient name, date of birth, medical record number, referring md, date of study.</td>
</tr>
<tr>
<td>Indication</td>
<td>Unstructured</td>
<td>Text and numeric</td>
<td>previous diagnoses, medication, procedures, bmi, bedtime, wakeup time.</td>
</tr>
<tr>
<td>Architecture</td>
<td>Semi-structured</td>
<td>Numeric</td>
<td>lights out time, lights on time, total record time, number of REM (rapid eye movement) episodes.</td>
</tr>
<tr>
<td>Respiratory</td>
<td>Unstructured</td>
<td>Numeric</td>
<td>respiratory rate, number of central, obstructive and mixed apneas.</td>
</tr>
<tr>
<td>Gas Exchange</td>
<td>Unstructured</td>
<td>Numeric</td>
<td>O2 saturation.</td>
</tr>
<tr>
<td>Ventilation</td>
<td>Unstructured</td>
<td>Numeric</td>
<td>end tidal, ph</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Unstructured</td>
<td>Numeric</td>
<td>heart rate.</td>
</tr>
<tr>
<td>EEG</td>
<td>Unstructured</td>
<td>Numeric</td>
<td>plm (periodic limb movement) index.</td>
</tr>
<tr>
<td>Sleep Diagnosis</td>
<td>Unstructured</td>
<td>Text</td>
<td>current sleep disorder diagnoses.</td>
</tr>
</tbody>
</table>

### Table 2  Candidate terminology list

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Skipped or not</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient</td>
<td>yes</td>
</tr>
<tr>
<td>a white male with a history of asthma, … sleep</td>
<td>yes</td>
</tr>
<tr>
<td>a white male</td>
<td>yes</td>
</tr>
<tr>
<td>a history</td>
<td>yes</td>
</tr>
<tr>
<td>asthma</td>
<td>no</td>
</tr>
<tr>
<td>mild obesity</td>
<td>no</td>
</tr>
<tr>
<td>restless sleep</td>
<td>no</td>
</tr>
</tbody>
</table>

### Table 3  Final Candidate Terminology List

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Standard or non-standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>asthma</td>
<td>standard</td>
</tr>
<tr>
<td>mild obesity</td>
<td>non-standard</td>
</tr>
<tr>
<td>restless sleep</td>
<td>non-standard</td>
</tr>
</tbody>
</table>
### Table 4  Terminology extraction accuracy

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Precision (Exact)</th>
<th>Recall (Exact)</th>
<th>F1 (Exact)</th>
<th>Precision (Exact and Partial)</th>
<th>Recall (Exact and Partial)</th>
<th>F1 (Exact and Partial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>75.10%</td>
<td>71.18%</td>
<td>73.09%</td>
<td>82.10%</td>
<td>77.33%</td>
<td>79.64%</td>
</tr>
<tr>
<td>Procedure</td>
<td>77.72%</td>
<td>73.21%</td>
<td>75.40%</td>
<td>85.72%</td>
<td>78.10%</td>
<td>81.73%</td>
</tr>
<tr>
<td>Medication</td>
<td>86.12%</td>
<td>79.91%</td>
<td>82.90%</td>
<td>93.95%</td>
<td>85.34%</td>
<td>89.44%</td>
</tr>
<tr>
<td>Overall</td>
<td>79.65%</td>
<td>74.77%</td>
<td>77.13%</td>
<td>87.26%</td>
<td>80.26%</td>
<td>83.60%</td>
</tr>
</tbody>
</table>
28. Murphy SN, Wilcox A. Mission and Sustainability of Informatics for Integrating Biology and the Bedside (i2b2), eGEMs (Generating Evidence & Methods to improve patient outcomes) 2014; 2(2): 7.