Automated Narrative Extraction from Administrative Records*,**

Karine Megerdoomian
The MITRE Corporation
McLean, VA, USA
karine@mitre.org

Karl Branting
The MITRE Corporation
McLean, VA, USA
lbranting@mitre.org

Charles E. Horowitz
The MITRE Corporation
McLean, VA, USA
chorowitz@mitre.org

Amy B. Marsh
The MITRE Corporation
McLean, VA, USA
amarsh@mitre.org

Nick Modly
The MITRE Corporation
McLean, VA, USA
nmodly@mitre.org

Stacy J. Petersen
The MITRE Corporation
McLean, VA, USA
spetersen@mitre.org

Eric O. Scott
The MITRE Corporation
McLean, VA, USA
escott@mitre.org

Sujit B. Wariyar
The MITRE Corporation
McLean, VA, USA
swariyar@mitre.org

ABSTRACT

The U.S. Probation and Pretrial Services Office staff produce billions of pages of information on defendants' and offenders' profile and conduct. While it is critical for probation officers and district chiefs to have up-to-date knowledge on their clients to better assist and reduce risk of recidivism, the data are often stored in narrative texts in multiple large documents. As a result, these records remain mostly out of reach without the use of painstaking manual review. This paper describes an analytic prototype developed to automatically acquire structured information from natural language text in probation office documents through the application of PDF content extraction, text mining, and language analytics. Since serious mental illness is very prevalent in the U.S. corrections system, the first phase of the project focused on extracting information and constructing timelines from narrative text regarding the defendants' mental health conditions, substance use and treatment history.

Automated narrative extraction and the construction of an event timeline for defendants' mental and emotional health history have allowed the probation office to have a better understanding of their client population and to perform analyses that were previously unavailable to the organization. This technical approach can be applied across organizations, legal institutions, clinical administrations, and government agencies that maintain large amounts of information in the form of free text narratives.

*In Proceedings of the Workshop on Artificial Intelligence and the Administrative State (AIAS 2019), June 17, 2019, Montreal, QC, Canada.

Published at http://ceur-ws.org
** Approved for Public Release; Distribution Unlimited 19-1482. Throughout this document, all names of people, places, facilities and dates are replaced with fictitious ones to anonymize the information.

1 Introduction

The U.S. Probation and Pretrial Services Office (PPSO) staff supervise more than 300,000 people a year and collect and produce billions of pages of information on defendants' and offenders' profile and conduct, as well as on the strategies and actions of officers and their outcomes. While it is critical for probation officers to have up-to-date knowledge on their clients to reduce the risk of recidivism, the data are often stored in narrative texts in multiple large documents, making it very challenging and time-consuming to collect all relevant case information manually. This renders 70 terabytes of mostly unstructured data on more than a million defendants, and strategies used by thousands of officers over decades, mostly unusable by PPSO [1]. As a result, policy makers, program evaluators, and probation and pretrial services staff have been denied valuable data with which to do their jobs.

A significant number of offenders supervised by the U.S. probation services have a current mental health condition, most of them with co-occurring substance use disorders.

Defendants who suffer from mental disorders often require more intensive monitoring and specialized treatment [2]. We therefore focus on addressing important PPSO business questions to better understand the nature of the mental conditions in the officers' caseload and gain knowledge of the defendants' diagnosis and treatment history. The information was automatically obtained from the free text sections of Presentence Investigation Reports (PSIR), which represent investigations into the history of the person convicted of a crime before sentencing to determine if there are extenuating circumstances. To automatically extract and analyze the free text information in the PSIRs, we applied language analytics technology to detect the events of interest (substance use, diagnosis, treatment sessions, prescriptions) in the defendant's life and visualized them as a timeline of activities that could be reviewed by the probation and parole officers.

The system leverages Apache cTAKES (clinical Text Analysis and Knowledge Extraction System), an open-source Natural Language Processing (NLP) system developed specifically to extract and analyze clinical information from unstructured text [3]. cTAKES identifies clinical terms such as drugs, diseases and disorders, symptoms, and medical and treatment procedures. It also performs deep textual analysis and can identify, for instance, if a sentence is negated or not, or if the person being discussed is the patient or a family member. The prototype system combines the results of cTAKES with rich linguistic analysis from other open source systems such as concept ontologies and the Stanford CoreNLP parser and entity recognizer [4]. These syntactic and semantic analyses are then enhanced to adapt to the use case, by identifying significant terms for the events of interest for the mental health domain, applying linguistic analysis to improve argument and negation detection, and implementing recent advances in NLP to improve precision (e.g., vector space semantics, algorithms for building a narrative timeline).

All extracted information on a defendant's narrative is stored in a graph database and displayed on a dynamic map, allowing filtering of results based on judicial district, defendants' demographic information (age, education, citizenship), criminal category, mental conditions or medications prescribed.

As large amounts of information in business, government and administration are maintained in the form of narratives (clinical records, legal and financial summaries, progress reports, human resources assessments, etc.), the approach described in this paper for acquiring structured information from narrative text can be reapplied across organizations and government agencies.

2 Background

Past clinical information extraction systems have tended to rely on shallow NLP techniques (pattern-matching, simple parses, linear pattern interpretation rules). More recently, however, several projects have adopted knowledge-based approaches adapted for the clinical domain.

While the advantages of machine learning methods for information extraction cannot be denied, they also present a number of limitations in applications for narrative extraction from clinical data. To begin with, machine learning algorithms require large amounts of training data which are pre-tagged for the relevant features and parameters. Preparing the pre-annotated data sets can be time-consuming and expensive. In addition, such probabilistic approaches might miss rare phenomena that need to be identified since they do not occur often enough in the training data to be picked up by the learning algorithms. Another challenge for using machine learning methods in the clinical domain is that users often expect high level of consistency in the results and precise information on how the computational decisions were made. In such instances, a rule-based approach might be more transparent and easier to understand and modify.

The approach described in this paper leverages in-depth linguistic and semantic analysis to detect the domain information in narrative text, more in line with recent knowledge-based approaches [5] [6]. Machine learning approaches often require a large amount of pre-annotated data on which to train the algorithms. Since the PSIR data had not previously been tagged for the events of interest and mental conditions, a purely machine learning approach was not readily available. Hence, the prototype applies a hybrid method. It leverages rich linguistic and semantic information through the application of open-source Natural Language Processing systems, adapted for the existing use case by applying a combination of rule-based linguistic analysis, vector space semantics, and machine learning techniques to enhance the results. These were used to improve negation detection and argument identification (i.e., entities the events refer to), and to develop temporal reasoning algorithms. Ontologies (lexicons) of mental health and medication terms, vetted by a subject matter expert, were used for concept identification. The rest of this section provides a detailed description of the technical steps in building the analytic prototype.

3 Technical Approach

The technical approach is a hybrid one, leveraging open source NLP applications often developed by training machine learning algorithms, and refining the syntactic and semantic analyses with a combination of knowledge-based and probabilistic approaches.

3.1 Analytic Pipeline

The presentence reports undergo several steps in order to extract the defendant's mental health and substance use narratives. These are shown in **Error! Reference source not found.** and are described in detail in the rest of this section. The specific steps involved are:

1. **Content Extraction:** parsing the different sections of the PDF documents and extracting the structured profile and criminal information as well as all free text content. This component also "cleans" the data by normalizing the textual content to maximize processing.
2. **Language Analytics:** The extracted text for each PSIR is run through the Natural Language Processing components, providing a full linguistic parse, a list of entities and events of interest, and semantic relationships.
3. **Knowledge Discovery:** This step is the heart of the textual analytics where the system identifies all concepts, events, and their relationships for the domain of interest.
   - Identifies the events of interest associated with the defendant (arrests, diagnoses, treatments, prescriptions, drug use, suffering from a mental condition);
   - Determines whether the information is obtained from medical records or if it is reported by the defendant, by a medical professional, or by a third party;
   - Provides full event description including date, location, persons involved, treatment provider, nature of treatment and medication prescribed;
   - Computes the temporal relationships between the various events to build a narrative timeline for a defendant.

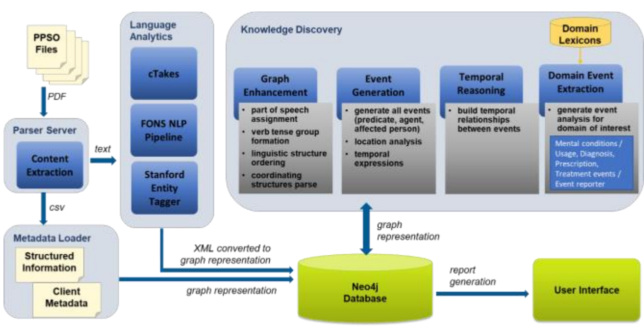

Figure 1: Analytic pipeline for narrative extraction and timeline development

4. **Neo4j Database:** Neo4j is a graph database management system and is available as open source software. All extracted information from the Knowledge Discovery component, as well as the client demographic metadata, and structured information on arrest history and federal offenses extracted from the presentence reports are loaded into the database.
5. **User Interface (UI):** This component interacts with the Neo4j database and displays results on a Google Earth map. The UI allows the user to run queries, to review the details on particular defendants, and to see aggregate results on the data set.

3.2 Content Extraction

The Content Extraction component parses the PDF presentence reports, identifies all subsections and extracts the textual content. To analyze the mental health and substance use information of defendants, the text content of the Mental and Emotional Health (MEH) and Substance Abuse (SA) sections in presentence reports are automatically extracted. In addition, this step identifies and extracts all federal charges from the cover sheet of the PSIR, criminal history information from the Juvenile Adjudications and Adult Criminal Convictions sections of the report, Arrest Dates and associated charges from the Criminal History information, and Criminal History Score and Category from the Criminal History Computation section.

The prototype's Content Extraction component successfully extracted information from 92% of the original PDF documents, providing us with a data set of 11,243 extracted narrative text documents to analyze. Given that some defendants have more than one presentence report associated with them, the successfully extracted content corresponds to 10,973 defendants. The free text content extracted from the MEH and SA sections amount to 22,486 text items. These can range from a few sentences to several paragraphs depending on the report.

3.3 Language Analytics

The Language Analytics component leverages existing Natural Language Processing software to perform various linguistic analyses on a piece of text. NLP is a subset of Artificial Intelligence (AI) and is fast becoming an essential technology in modern-day organizations to gain significant insights from unstructured content, such as email communications, social media, videos, customer reviews, customer support request, and administrative records in business and government. Natural Language Processing tools and techniques help to automatically process, analyze, and understand large amounts of data, providing structure and meaning to information that originally was in unstructured form.

In this step of the analysis, the texts extracted from the Mental and Emotional Health and Substance Abuse sections of the PSIRs are run through several NLP software tools. The software packages currently in use are Apache cTAKES (clinical Text Analysis and Knowledge Extraction System), Stanford Named Entity Recognizer, and FONS (Framework for Operation NLP Services) – a software package pipeline leveraging open source tools and was built by a research team at MITRE to detect events of interest to national security.

cTAKES output forms the primary basis for further analytics. It was chosen primarily because of its entity recognition capabilities in the clinical domain, which aligned with the desire to obtain data about PPSO clients' mental and emotional health and substance use. Entities identified by cTAKES include medical conditions, drugs/medications, medical procedures, and medical symptoms. The entities identified by cTAKES out-of-the-box were supplemented with additional entities frequently encountered by analysts in PSIRs. We worked closely with a PPSO subject matter expert to review the list of conditions and medications that cTAKES recognized, and identify the ones that were of interest in the mental and emotional health and substance use domain. The subject matter expert also identified a more general superclass for each of these specific mental and emotional conditions so that further analysis could be conducted at the appropriate level of granularity. For example, conditions such as *depression, chronic depression,* and *major depressive disorder* were all mapped to the more general term *depressive disorder.*

cTAKES also provides domain-independent NLP capabilities of syntactic parsing, dependency parsing, and semantic role labelling – it can give the base forms of words, their parts of speech, mark up the structure of sentences in terms of phrases and syntactic relations, detect negation in the sentence and identify the role of the entities in a sentence (e.g., agent of event). The results of all these capabilities were used to identify events of interest in a client's mental and emotional health and substance use history. However, we found it useful to supplement the cTAKES output with other natural language processing systems to achieve the most accurate analysis. The Stanford Named Entity Recognizer was applied to identify people, places, organizations, dates, times, and locations, none of which are identified by cTAKES. Additionally, the FONS system, which also generates entities, syntactic parsing and dependency parsing output, was used to supplement cTAKES' output to obtain a higher level of accuracy. In particular, FONS was applied to the PSIR text data to tag entities (people, facilities, locations, dates and times), and to categorize all events into conceptual classes by detecting event types (e.g., state, transfer, communication) and different verb meanings (e.g., *prescribe* can either be the verb denoting the prescription of medication by a medical professional or a communication event meaning 'to advise', 'to recommend').

3.4 Domain-Specific Entity and Event Identification

The Knowledge Discovery phase of the analytics involves processing the output from the Natural Language Processing systems to perform several steps in knowledge discovery in natural language text:

1. Identify concepts (entities and events) of interest associated with the client, including mentions of a client suffering from a mental condition, diagnoses, treatments, prescriptions and drug use.
2. Detect the event description such as the date and location when it occurred, the persons involved, the treatment provider, the nature of treatment (e.g., inpatient or outpatient, anger management, drug rehabilitation) and the medication prescribed.
3. Detect the source of the information – was the information reported by the client, was it obtained from medical records or a medical professional, or reported by a third party?

As described, cTAKES detects these entities of interest in the mental and emotional health domain. However, to identify whether a client is suffering from a mental condition, it does not suffice to simply retrieve sentences with a mental condition mention. It is also important to detect the subject of the sentence to distinguish cases where a family member is mentioned to suffer from a mental condition (e.g., "*the defendant's mother suffered from Schizophrenia*"), and to exclude any negated events (e.g., "*the defendant does not suffer from a severe mental disease or defect*"). Fortunately, when cTAKES identifies a concept, it also identifies that sentence's polarity (whether the entity appears in a negated context or not), and the event's subject (whether that event or concept should be ascribed to the client described in the text, a family member of the client, or someone else). Some modifications to the cTAKES source code were made to improve the accuracy of these attribute identifications.

While the cTAKES entities can be counted to obtain statistics on the prevalence of various mental conditions among the defendant population, further processing is necessary to identify more complicated events, such as receiving a diagnosis, attending treatment, being prescribed medication, or using drugs. To identify the events of interest, a small sample of PSIRs was reviewed to identify the verbs commonly associated with these events. An iterative process was used in reviewing the event detection results and updating the predicates for the domain. The verbal predicates associated with each type of event are listed in Table 1.

Event Type	**Predicate**
Diagnosis	diagnose
Prescription	prescribe, treat (with)
Treatment	admit, attend, complete, discharge, enroll, enter, hospitalize, meet, participate, place, receive, see, seek, speak, treat, undergo
Usage	abuse, addict, consume, drink, experiment, ingest, inhale, relapse, smoke, snort, take, try, use

Table 1: Verbs used to identify events related to mental and emotional health and substance use

Once the predicates are identified, the semantic roles associated with each occurrence of the predicate are automatically extracted to enable the identification of the predicate's agent, affected entity, and whether the predicate was negated. The sentence in which the predicate appeared was also examined to identify medications, drugs, mental conditions, medical procedures, and treatments associated with that event.

To detect the source of the information, all sentences with Communication events identified by the FONS software package were analyzed and the subject of the verbs extracted. For example, in "*Dr. Gray stated that the defendant has never been hospitalized for emotional disorders of any kind",* the communication verb *stated* is detected and its subject, *Dr. Gray* (a medical professional), is identified as the source of the information. Similarly, in the example "*the defendant's mother also reported he was diagnosed with Bi-Polar Disorder several years ago",* the source of information is identified as *the defendant's mother* (a third party).

If the subject of the communication verb is mentioned as *the defendant*, the system treats it as a self-reported event. In the writing style of the presentence reports, mentions of *he* or *she* tend to refer overwhelmingly to the defendant. Since the current version of the analytic system does not include a "coreference resolution" component that can accurately identify who the pronouns refer to, the assumption is made to treat these cases as self-reported events. This can be seen in the following example where the events in both sentences are automatically labeled as self-reported: "T*he defendant expressed feelings of depression, helplessness, and hopelessness. He also admitted to occasional auditory hallucinations."*

If the name of the defendant is mentioned as the subject of the communication verb (e.g., "*McKenna could not recall being prescribed medication to treat his Depression"*), an additional step is performed to verify the name *McKenna* against the defendant metadata information – if the system finds a match, then the information is labeled as self-reported.

Certain automated enhancements had to be made to the Communication event detection, however, since the automatic classification by FONS included verbs such as *stuttered* and *snorted*. In order to improve the results, we computed semantic vector measures that capture the similarity in usage of verbs against canonical Communication events such as *reported* and *stated*. The verbs that are closest in the context of use within the text and thus have closer meaning to the *report/state* verbs produce higher values and are thus more likely to be indicative of the source of information. The top verbs identified as Communication events are listed in Table 2.

Event Type	**Predicate**
Communication	state, indicate, note, explain, report, say, acknowledge, discuss, identify, confirm, deny, address, agree, communicate, question, suggest, tell, describe, claim, mention, inform, disclose
Other formulation	according to

Table 2: Terms used to identify the source of information.

This linguistically rich event-based narrative analysis methodology allows the Language Analytics component to extract information of interest including the people involved in the event, the time it occurred, and the places mentioned. A sample analyzed sentence is shown in the following example:

> The defendant<source-of-info> reported she<affected-entity/diagnose-event> was diagnosed<diagnose-event> at the age of 14<time> with depression<mental-condition>, schizophrenia<mental-condition> and bi-polar disorder<mental-condition> and was not prescribed<prescribe-event|NEGATIVE> any medication<medication-mention>.

3.5 Generalized Event Analysis

While cTAKES proved very useful for identifying events in the clinical domain, it is not specifically tuned for identifying more general events. Events that are not directly related to diagnoses, prescriptions, substance abuse, or treatment may still be of interest when analyzing a client's mental and emotional health history. For example, in *"He became depressed when his infant brother died"*, the event of the infant brother's death does not fall into one of the domain-specific event categories, but it is still relevant to indicate a trigger or risk factor. To try to capture these types of events, a more general approach to parsing free text was used, producing an event-based analysis for every verb encountered in the Mental and Emotional Health and Substance Abuse sections.

As part of the Knowledge Discovery phase, the linguistic output from the NLP systems loaded into the Neo4j graph database is used as the basis for generating events that do not rely on a domain-specific vocabulary. In this framework, events are generally identified by the presence of a verb and an event-based analysis is performed on the sentence. In simple sentences, this means that one event corresponds to the entire sentence. However, if a sentence contains multiple clauses, each clause could potentially represent one event. In the sentence *"he became depressed* ***when*** *his infant brother died"*, *becoming depressed* is one event, and *his infant brother died* is a separate event. The two clauses are linked by the conjunction *when,* which indicates the temporal relationship between them. To handle sentences such as these, a list of terms that signify a subordinate clause was created and sentences were divided into clauses when one of these terms was found. The list of terms used is in Table 3 below. These terms are used in further analytics to identify temporal or causal relations between events.

Relationship Type	Clause Marker Terms
Temporal	after, before, during, following, prior to, throughout, until, upon, when, while
Causal	although, as a result of, because, due to, in order to, since
Other	according to, along with, in addition to, relating to

Table 3: Terms signifying the presence of a subordinate clause in a sentence.

After all clauses have been identified, an event is generated for each clause. If the clause contains a verb, the verb phrase forms the basis of the event. If there is no verb phrase in the clause, (e.g., in the sentence "w*hile in prison, the defendant used heroin",* "while in prison" is a clause without an explicit verb), the phrase after the clause marker forms an event description which is the basis of the event. Then, information from the syntactic parses, dependency parses, semantic roles, and named entities are used to identify agents, affected entities, indirect objects, locations, and temporal mentions related to the basis of the event for a complete narrative analysis.

3.6 Temporal Reasoning

Once all relevant events have been extracted from the text of the PSIRs, it is possible to make a timeline of the relevant events with temporal mentions in a client's history. To accomplish this, we adapted TimeML (Markup Language for Temporal and Event Expressions) standards to the narratives generated [6]. TimeML is designed to provide a standard way to annotate events with a time stamp and place events in chronological order; it is thus optimal for the problem of timeline generation. In TimeML, events are typically described by verbs, which aligns with our approach to narrative generation. In the actual TimeML specification, temporal expressions are marked as separate entities, falling into the categories of ***date*** (for events that take place at a specific time, which might be a date, month, or year), ***time*** (for events that take place at a specific time of day), ***duration*** (for events that have clear start and end points), and ***set*** (for periodic events). In our adaptation, we recorded the type of temporal expression as an attribute of the event it was associated with, and did not use the category of ***time*** since the specific time of day of various events is not typically specified in PSIRs. The category of ***set*** was recorded but is not currently used for timeline generation. The start date and end date of each event are also recorded as additional attributes.

To place events in order, TimeML uses the TLINK tag, which records the id's of two related events and the temporal relationship between the two. In this project, temporal relationships are marked as an attribute of the event rather than a separate entity, and an abbreviated set of seven temporal relationships are used, rather than the fourteen defined in TimeML. The temporal relationships utilized are listed in Table 4.

Relationship	Description
AFTER	Event 1 occurs some time after Event 2
BEFORE	Event 1 occurs some time before Event 2
BEGINS_AT	Event 1 occurs immediately after Event 2
ENDS_AT	Event 1 occurs immediately before Event 2
INCLUDES	Event 1 starts before and ends after Event 2
IS_INCLUDED	Event 1 starts after and ends before Event 2
SIMULTANEOUS	Event 1 and Event 2 start and end at the same time

Table 4: Temporal relationships used for timeline generation.

Determining the values of the temporal type, start time, end time, and temporal relationships for the generated events is a three-step process. In the first pass through the events, any temporal mentions associated with each event were parsed with regular expressions and used to set the event's type, start date, and end date. Next, temporal relationships were identified by examining events to see if they contained any of the subordinate clause markers listed in Table 5. Rules were then applied to relate two events connected by a subordinate clause marker. One final pass through the events was used to set any additional start and end dates that could be inferred after the temporal relationship was determined.

We can follow this entire process on the sentence "*He began smoking marijuana at the age of 16 until his arrest in 2014",* which contains the events *he began smoking marijuana* and *his arrest in 2014.* The first step after the identification of the two clauses is to identify the presence of the temporal expressions in each clause – *at the age of 16* in the *smoking* event (EV1) and *in 2014* in the *arrest* event (EV2). In EV1, the start time can be obtained from the defendant's date of birth in the profile information available in the database. In EV2, the Knowledge Discovery component establishes that the temporal expression is of type ***date***, with a start and end time set to span the whole year as shown in Table 5, since the time is not more clearly specified than that. The second step will identify the subordinate clause marker *until*, and follow a rule that establishes that the *smoking marijuana* event ended at *his arrest in 2014.* The final step will use the presence of the

ENDS_AT relationship to set the end time of *he began smoking marijuana* to the start time of *his arrest in 2014*. The final event analysis associated with a temporal range is then used to build a timeline and visualize on the web-based interface.

Clause/Event detection	[He began smoking marijuana]$_{\textbf{clause1/EV1}}$ **until** <clause-marker> [his arrest in 2014] $_{\textbf{clause2/EV2}}$
Step 1: Detect temporal expressions when available	EV1: <type: "date", startAt: {year: '1992', month: '6', day: '12'}, endAt: None> EV2: <type: "date", startAt: {year: '2014', month: '1', day: '1'}, endAt: {year: '2015', month: '1', day: '1'}>
Step 2: Establish temporal relation between events	id: "EV1", relType: "ENDS_AT EV2"
Step 3: Temporal reasoning to set temporal expression	EV1: <type: "date", startAt: {year: '1992', month: '6', day: '12'}, endAt: {year: '2014', month: '1', day: '1'}> EV2: <type: "date", startAt: {year: '2014', month: '1', day: '1'}, endAt: {year: '2015', month: '1', day: '1'}>

Table 5: Temporal reasoning process in Knowledge Discovery phase.

4 Graph-Based Representation

The main motivation for using a graph database to store the parse output is that syntactic parse outputs are often modeled in linguistic theory in the form of trees (a graph in which each node has a single parent) and dependency parses capture the semantic relationship associated with two nodes, so storing the parse outputs as a graph allows to use Neo4j API (Application Programming Interface) and CQL (Cypher Query Language) to directly access these grammatical relationships and handle the recursion inherent in language. Additionally, once the natural language parsing outputs are stored in graph format, it is easy to align and merge the outputs from the different NLP systems being used. Finally, Neo4j provides a visualization of the graph for linguists and developers that assists in understanding the structure of the language.

Once the output from the NLP systems is stored in the database, we apply several enhancements to the raw system output to improve the parses' accuracy and generalizability. These enhancements include labelling all nodes with a more coarse-grained part-of-speech tag, grouping together multi-word verb phrases into a single entity (e.g. merging the nodes for the terms in the phrase *has been attending* into a single node *attend* with appropriate tense and aspect information), and combining coordinated phrases with conjunctions into a single entity (e.g. merging the nodes for the terms in the phrase *mental and emotional* into a single node to facilitate further analysis).

The User Interface interacts with the Neo4j database to access all content and narrative analytics output and displays the results on a Cesium Server. The web-based interface allows the user to run queries of interest, filter based on the defendant's profile information, and view the retrieved information on a spatial map of judicial districts or States.

The UI displays an aggregate report of the data for the provided query as shown in Figure 2. This display can be further filtered based on the mental conditions, medications and substances of interest, as well as the defendant's demographic information and criminal category.

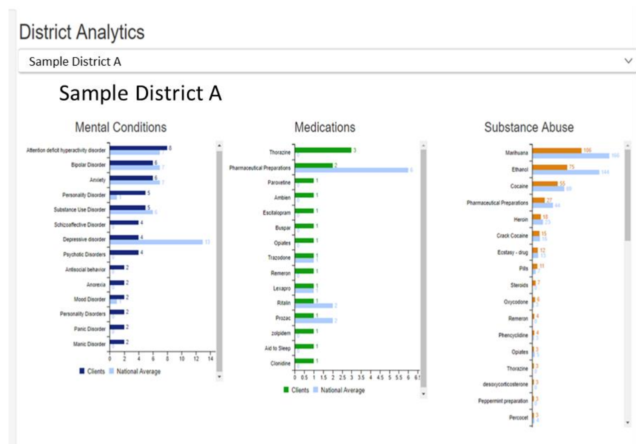

Figure 2: Narrative analytics results viewed by judicial district.

The user can then select to view the identified defendants on a map to the level of street detail. The user may also select to view a particular client's information in more detail, such as mental conditions reported, and see associated text from the Presentence Investigation Report with relevant sections highlighted. In addition, the data are used to visualize a timeline of the defendant's life events including arrests, diagnoses, substance use, and treatments.

5 Results of Analytics

The programmatically important questions of interest to PPSO that are addressed in the current prototype are (i) determining how many defendants sentenced had a mental health condition; (ii) the types of conditions present; (iii) the source of the diagnosis; (iv) prior treatment exposure; and reporting of that information by demographic, offense and prior criminal history information.

To identify the number of defendants with a mental illness, the system extracts all the client cases where a mental illness was mentioned as attributed to the defendant (whether officially diagnosed or not). It was found that 3,959 defendants in the data set (about 36% of the studied population) had a history of one or more mental conditions. If Substance Use Disorder is included as a mental health condition, that number increases to 58%. Figure 3 provides the heuristics for the mental health conditions mentioned in the Mental and Emotional Health sections of presentence reports studied. However, the total number of defendants who have officially been diagnosed with a mental condition is 2,238 (20% of the studied population). In addition, 82% of the defendants had a history of substance use (mainly Marijuana and alcohol), and 53% of cases had a prior criminal record cited. Most common prescriptions are Prozac, Ritalin, Seroquel and Xanax, and top substances include Marijuana, Alcohol, Cocaine, Methamphetamine and Heroin.

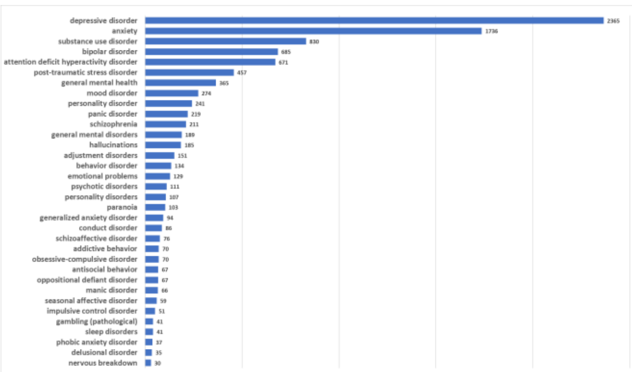

Figure 3: Mental health conditions associated with about 11,000 defendants.

As described earlier, the analytic prototype identifies the source of information for each detected event of interest. There are five distinct categories for the source: (i) self (client self-report), (ii) medical professional, (iii) medical records, (iv) report (official non-medical records, including evaluations and assessments), and (v) third party (third party corroboration such as a family member, defense counsel, probation agent, or pretrial services agency). In the presentence reports studied, the majority of the events (about 89% of all events found) are self-reported.

The full set of results in response to the PPSO business questions is shown in Table 6.

Category	Count	Percent	Accuracy*
PSRs with Mental Health and Substance Abuse sections [set P]	11,243		
Total number of clients corresponding to set P	10,973	100%	
Total number of clients for which an event of interest was found (diagnosis, prescription, treatment, substance use, suffering from mental condition)	10,743	98%	
Total number of clients with a mental condition	3,959	36%	98%
Total number of clients who had a diagnosis made, with an explicit mental condition	2,238	20%	99%
Total number of clients who attended some sort of treatment or assessment/evaluation (mental health)	3,469	32%	87%
Total number of clients who attended some sort of treatment or assessment/evaluation (substance abuse)	3,817	35%	90%
Total number of clients with medication prescribed	2,057	19%	92%
Total number of clients using/abusing substances	8,998	82%	*Not Evaluated*
Total number of clients with historical arrest information (30,566 arrests total, extracted from metadata)	5,764	53%	N/A

Table 6: Automatically obtained responses to the PPSO business questions on defendants' mental health and substance use history.

System performance was evaluated by creating a small reference sample of about 500 sentences to measure the accuracy of the information extracted for each event type. The 500 sentences were manually annotated by team members indicating the expected mental conditions, event types (diagnosis, treatment, prescription, usage), and medications. The annotations also included important event-related information such as the agent (prescriber, diagnoser), polarity (whether the event is negated or not), and the temporal expression associated with the event. The language analytics results were then compared to the pre-annotated reference set to measure how many of the detected elements were accurate and to also calculate how many of the expected elements were not picked up by the system.

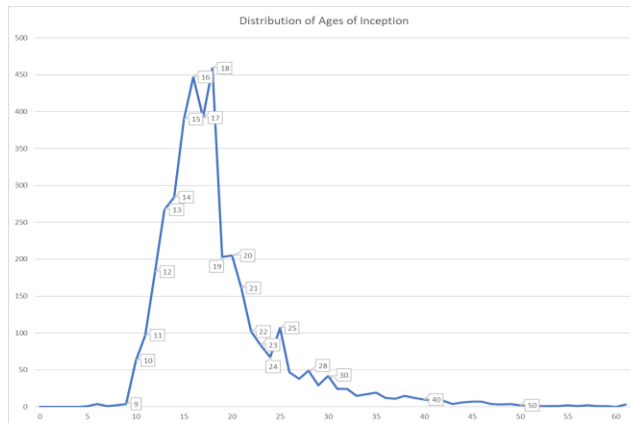

Figure 4: Correlation analysis shows defendants' onset of substance use. The x-axis represents the defendant's age and the y-axis is the number of times the onset of substance consumption is found in the text.

We also explored the aggregated national and district data for potential correlations and analyses across defendants. Figure 4 illustrates one such analysis, which shows the onset of substance use among the defendants studied. This examination automatically detects any mentions of the age of the defendants in the Substance Abuse texts and identifies any sentences that refer to the onset of using a drug or alcohol by the defendant. For instance, a sentence such as "*he began using cocaine at age 17*", is labeled as an "inception predicate" and associated with the age of the defendant (i.e., 17). The results show that the onset of substance use among defendants starts at age 10, with a steady increase to age 16 and peaks at age 18.

This analysis is only one example of the types of aggregated correlations and computations that are available after the full language analytics have been performed on the data. Other correlations explored include automatically detecting instances of co-morbidity to understand which mental conditions tend to co-occur most often among the population, automatic detection of defendants with previous suicide attempts or history of suicidal ideation, and identification of events that may trigger mental health issues (e.g., death of a family member, history of sexual or domestic abuse, fatal medical diagnosis, divorce).

6 Application to Risk Assessment

Analysts working in the Probation and Pretrial Services domain leverage a variety of data-driven instruments to measure trends, train officers, and assess the recidivism risk in individual clients. At a high level, these efforts are typically described in terms of the popular Risk, Needs, and Responsivity model, which dictates that effective offender supervision ought to allocate more treatment resources to high-risk clients, that treatment should target specific criminogenic needs in the client's case, and that officers should apply cognitive-behavioral techniques to respond to the details of a client's particular situation [7, 8]. In recent years there has been a trend toward using data-driven approaches for the first step, and actuarial risk assessment instruments such as "Levels of Service" surveys [9] and the federally developed Post-Conviction Risk Assessment (PCRA) [10] have played an increasingly important role in the allocation of treatment resources. These tools are typically based on survey questions that must be administered and recorded by the officer, which then serve as inputs to traditional statistical modeling techniques (e.g., logistic linear regression). Such tools are time-consuming to use, and they offer only a limited, static snapshot of the specific criminogenic needs that are present in a client's case. Risk/needs assessment is an active area of research, and efforts are ongoing to identify next generation tools that can offer improved data-driven methods that can help support probation officer responses during their regular interactions with clients. Leveraging the wealth of unstructured information that is present in the existing documentation that is available in probation case tracking systems is one promising approach to solving this problem.

Any application of AI or data analysis to officer decision-making can end up having a significant impact on the population under supervision, and so it is important to be aware of the various ethical concerns that surround the application of data analysis software to social issues [11]. Such concerns include the need for general algorithmic accountability [12], the need for assurance that algorithms that are used for such important tasks as recidivism prediction do not exhibit unacceptable biases [13], the need for judicial review of algorithm-assisted decision-making (where such review may be called for), and more practically, the need to inspire trust in users, who tend to be unwilling to rely on algorithms whose inner workings are poorly understood. Some of these issues are of greater concern than others in a probation domain. Judicial review, for example, is a legal necessity when algorithms directly impact a judge's decisions, but risk/needs assessments for offenders who are on supervised release are not normally referred to in judicial decision-making.

In this work, we focus on the foundational question of extracting information from unstructured text that can inform the decisions of officers and analysts working within the federal probation system. We defer questions about automated risk assessment, and the fairness thereof, to future research. The current work focuses instead on extracting and arranging raw facts from various sources in a visualization that a human can use to support their professional judgement in a particular case and that an officer can potentially leverage in detecting patterns that had previously been unavailable.

7 Future Directions

The paper describes a successful approach to the automatic extraction and analysis of narrative text in the mental health and substance use domain. The approach has since been applied to other domains such as employment history and financial history. The results provide evidence that the use of technology in identifying important information in free narrative text in administrative records is feasible and cost-effective, and any adaptations to new domains can be accelerated through probabilistic methods. These analytics can be further developed in various directions, depending on the mission needs of the organization. This section provides some directions to pursue.

The current results of the analytics can further be improved upon by annotating more data and performing a larger-scale evaluation and refinement cycle. Although event extraction accuracy ranged in the 90-percentile, an evaluation conducted on a larger data set will provide better accuracy measures and can identify low frequency events that may have been missed in the current version of the analytics. Further work can also be performed on negation and argument detection to achieve higher precision. In addition, the analytics results have not yet been fully validated by a subject matter expert to ensure that the data identified and the way the results are presented are valuable for the PPSO officer or mental health analyst.

Building a timeline of a defendant's life events from narrative text is a very complex task and the topic of much current research in the field of NLP. We successfully identified the temporal expressions associated with events and introduced a temporal reasoning component which is tightly integrated into the system's syntactic parse and semantic relations output. Yet, identifying the temporal relations between events is not an easy task and oftentimes, the system needs to infer a relationship that is not overtly mentioned in the sentence. Building a hybrid method combining knowledge-based linguistic analysis with a statistical machine learning approach will provide more robust temporal relationship analyses.

One of the issues that were left unaddressed in the current version of the analytics was the distinction between events (e.g., diagnoses, treatments) that occurred in the past and those that are currently valid. This can be accomplished by leveraging the tense and aspect information that the system computes and adding a filter on the UI to allow the user to view only events that are current.

Building a complete timeline of a defendant's life events will provide the important information at the individual level for PPSO officers to view and analyze, helping them identify precursor events and triggering factors. For instance, in addition to the mental health and substance use information, the personal history of the defendant (e.g., whether he or she graduated high school, history of domestic violence or neglect), existence of dependents (e.g., number of dependents and their age, learning issues, custodian), family relations (e.g., siblings and whether they have a criminal or substance abuse history), employment status, gang or terrorism activity, etc. are all important information elements that could shed light on the defendant's situation and allow probation officers to provide more efficient supervision and intervention measures to reduce recidivism. This requires fusing all events and information extracted from presentence documents onto a single timeline to view and analyze.

An important goal for analytics research is to leverage the large amount of data from diverse sources available to the probation office—including treatment reports, Chrono notes, social media, structured metadata, risk assessments, and court documents—to obtain a more complete picture of the defendant's history, conduct and status. The data analytics methods will be applied to these data sources and all results combined into a unified database available for query and analysis. Building on a multi-source analysis, the system can begin identifying precursor events to criminal activity or noncompliance, or detecting triggers for mental health issues or substance use relapses, and leverage that information to build a predictive model to forecast potential risk and generate automatic alerts. Such an alerting system can help direct an officer's attention to elements of a client's case history that indicate a special cause for concern.

ACKNOWLEDGMENTS

We would like to acknowledge the guidance and support of the U.S. Probation and Pretrial Services Office throughout this project. In particular, we would like to thank Steve Levinsohn for providing essential subject matter expertise to the team. The project was led and funded by the Technology Solutions Office at the Administrative Office of the U.S. Courts as part of the Applied Technology Research and Development program managed by the Judiciary Engineering and Modernization Center operated by the MITRE Corporation.

This (software/technical data) was produced for the U. S. Government under Contract Number USCA16D2019, and is subject to Judiciary Policy Clause 6-60, Rights in Data—General (JUN 2012).

For further information, please contact The MITRE Corporation, Contracts Management Office, 7515 Colshire Drive, McLean, VA 22102-7539, (703) 983-6000.

REFERENCES

[1] Matthew G. Rowland (2018). Federal Probation and Pretrial Services: What's Going On and Where Are We Going? (Presentation by the Chief of PPSO). Probation and Pretrial Services Office, Administrative Office of the U.S. Courts.
[2] Probation and Pretrial Services Office (2016). Overview of Probation and Supervised Release Conditions. Administrative Office of the United States Courts.
[3] Guergana K. Savova, James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *Journal of American Medical Informatics Association* 17: 507–513.
[4] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60
[5] Shervin Malmasi, Nicolae L. Sandor, Naoshi Hosomura, Matt Goldberg, Stephen Skentzos, and Alexander Turchin (2017). Canary: An NLP Platform for Clinicians and Researchers. *Applied Clinical Informatics* 08(02): 447-453.
[6] Hyuckchul Jung, James Allen, Nate Blaylock, Will de Beaumont, Lucian Galescu, and Mary Swift (2011). Building timelines from narrative clinical records: Initial results based on deep natural language understanding. *Proceedings of BioNLP 2011 Workshop,* pp. 146-54. Association for Computational Linguistics.
[7] Don A. Andrews, James Bonta, and Robert D. Hoge (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal justice and Behavior* 17.1: 19-52.
[8] Bonta, James, and Donald A. Andrews (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation* 6.1: 1-22.
[9] J. Stephen Wormith and James Bonta (2018). The Level of Service (LS) Instruments. In *Handbook of Recidivism Risk/Needs Assessment Tools*, pp. 117-145. Wiley & Sons.
[10] Christoper T. Lowenkamp, James L. Johnson, Alexander M. Holsinger, Scott W. VanBenschoten, and Charles R. Robinson (2013). The federal Post Conviction Risk Assessment (PCRA): A construction and validation study. *Psychological Services* 10(1): 87-96.
[11] Scott W. VanBenschoten (2008). Risk/needs assessment: Is this the best we can do. *Federal Probation* 72: 38.
[12] Nicholas Diakopoulos (2014). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*. Columbia University, Tow Center for Digital Journalism. New York: Columbia University Academic Commons.
[13] Alexandra Chouldechova (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. Big Data(5): 153-163.