# NLP Challenges for Detecting Medication and Adverse Drug Events from Electronic Health Records

Adverse Drug Events (ADEs) are common and occur in approximately 2-5% of hospitalized adult patients. Each ADE is estimated to increase healthcare cost by more than $3,200. Severe ADEs rank among the top 5 or 6 leading causes of death in the United States. Prevention, early detection and mitigation of ADEs could save both lives and dollars. Electronic Health Records (EHRs) provide real-time pharmacovigilance and drug safety surveillance by employing Natural language processing (NLP)techniques in EHR narratives.

We’ve annotated 1089 EHR notes with medications, as well as relations to their corresponding attributes, indications and adverse events. It provides valuable resources to develop NLP systems to automatically detect those clinically important entities. Therefore we are announcing a public NLP challenge aiming to promote deep innovations in related research tasks, and bring researchers and professionals together exchanging research ideas and sharing expertise. The ultimate goal is to further advance ADE detection techniques to improve patient safety and health care quality.

Tentative timelines

* Registration: begins August 1st, 2017
* Training data release: November 15th, 2017
* System submission
* Workshop

# Training Data

Training split of the annotated corpus contains around 900 de-identified, natural text Electronic Health Record documents. The documents are annotated with named entities and relations. The types of annotations are detailed below:

## NER Annotations

The named entities annotations can be divided into two broad categories, namely, medical entities and attributes. Medical entity annotations are either medical signs/symptoms or medication names. Medical attributes are annotations that provide details about the medical entities. The different type of NER annotations are:

## Adverse Drug Event [Entity]

Adverse drug events (ADEs) are annotations of adverse side-effects of medication. They are medical signs or symptoms resulting from the normal use of a drug at a normal dose. This includes drugs as a single dose, with prolonged use, withdrawal/discontinuation of a drug or from drug combinations.

## Indication [Entity]

An indication annotation is a medical sign or symptom that is the basis or direct cause of treatment. Alternatively, it can be described as medical condition for which a drug has been prescribed in the past or present.

## Other SSLIF [Entity]

## Other SSLIF annotations are medical sign, symptom or disease names that are neither being actively treated (Indication) nor are they an adverse side effect (ADE) of a drug. These annotations can be a medical conditions recorded in the previous history of the patient by the hospital staff or any recorded condition that is not caused by or a cause of any mentioned drug prescription.

## Severity [Attribute]

## Severity is an attribute of ADEs, Indications and SSLIFs. It denotes the severity of the affliction. It is often indicated by modifying words such as mild, minimal, markedly, severe, end-stage, small, extremely, substantial.

## Drugname [Attribute]

## Drugname annotations is name of a prescribed substance or over-the-counter drug, used in the diagnosis, cure, mitigation, treatment, or prevention of disease.

## Dosage [Attribute]

## Dosage is the amount of drug in a unit dose. It is a numerical value and is an attribute of Drugname.

## Route [Attribute]

Route is the path through which a drug is taken into the body. It is an attribute of Drugname.

## Frequency [Attribute]

Frequency is the rate of administration of the drug and is an attribute of Drugname. It is an attribute of Drugname.

## Duration [Attribute]

It is the duration or time range for the administration of the Drugname. It is an attribute of Drugname.

## Relation Annotations

## The annotation of this section are comprised of relations between different Medical Entities and Medical Attributes. The relation types are:

### Drugname – Attribute Relations

### Dosage (relation code “do”)

### Route (relation code “manner/route”)

### Frequency (relation code “fr”)

### Duration (relation code “du”)

In these type of relations, attributes (Dosage, Route, Frequency, Duration) describe an aspect of the Drugname Entity. These are relations between a drug and its corresponding prescription details.

### ADE/Indication/Other SSLIF – Severity: (relation code “severity\_type”)

This type of relation captures a modifying detail about the sign or symptom. It relates the modifier (for example “mild”) with the annotated entity (for example “fever”).

### ADE – Drugname: (relation code “adverse”)

This type of relation denotes that the ADE entity is an adverse side effect of the prescription of the Drugname Entity.

### Indication – Drugname (relation code “reason”)

This type of relation denotes that the Drugname entity has been prescribed as a direct treatment for the Indication Entity.

# Challenges

The learning algorithms for this task have to be modeled keeping in mind the challenges that the clinical domain entails. Unlike the text in the open domain, EHR notes are frequently noisy, containing incomplete sentences, phrases and irregular use of language. These notes incorporate abundant abbreviations, rich medical jargons, and their variations, which make recognizing semantically similar patterns in EHR notes difficult. Additionally, different events exhibit different patterns and possess different prevalences. For example, while a medication comprises of at most a few words of a noun, an ADE annotation (e.g., “has not felt back to his normal self”) may cover a significant part of a sentence. Moreover, while medication information is frequently described in EHRs, some annotations types such as ADE or Indication are typically rare events.

Another Major challenge inherent in this domain, is that different medical annotations like medication, diagnosis or adverse drug reaction require different context window sizes for identification. For example, while a medication can be determined by a context of two or three words containing the drug name, an adverse drug reaction would require the context of the entire sentence. This requires modeling long term dependencies between labels and relations. For example, in the sentence “the patient exhibited A secondary to B”, the label for A is strongly related to the label prediction of B. A can either be labeled as an adverse drug reaction or a symptom if B is a Medication or a Diagnosis respectively.

# Tasks

The challenge is divided into three different tasks:

## Task 1: Named Entity Recognition

## Task 2: Relation Extraction without NER

## Task 3: Relation Extraction with NER