Natural Language Processing
Supporting Clinical Decision Support

Applications for Enhancing Clinical Decision Making
NIH Workshop; Bethesda, MD,
April 24, 2012

Stephane M. Meystre, MD, PhD
Department of Biomedical Informatics, University of Utah
stephane.meystre@hsc.utah.edu
Why bother about NLP?

**Fast growing electronic clinical data**
Rapid adoption of Electronic Health Records (EHR).
Increasing requirements for electronic documentation of healthcare.
Exponentially growing quantities of electronic investigation results information (imaging, genetic testing, etc.)

**So what's the problem?**
Mostly unstructured information, like *narrative text*, dictated and transcribed or typed in.
Most structured information coded with administrative and reimbursement-oriented terminologies.
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information

Unstructured
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries
- Imaging / Radiology
- Prescriptions (pharmacy; CPOE)
- Laboratory results
- Administrative information
Why bother about NLP?

What does the EHR contain?

- Documents
  - History and Physicals
  - Clinical notes, Consult notes
  - Operative reports
  - Surgical pathology reports
  - Progress notes, Letters
  - Orders
  - Discharge summaries

- Imaging / Radiology

- Prescriptions (pharmacy; CPOE)

- Laboratory results

- Administrative information
Why bother about NLP?

CDS information needs

**Structured** (i.e., using some data model) and **coded** (i.e., labeled with some standard terminology) clinical information

...but most EHR content is narrative text, unstructured, and is therefore inaccessible for CDS.

**Detailed** clinical information, at various levels of granularity

...but most terminologies used were created for mortality and morbidity public health statistics, and for reimbursement (i.e., ICD-9-CM, CPT-4), and don't allow for detailed, clinical care-oriented coding.
Why bother about NLP?

Possible solution to these problems

Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content

⇒ Clinical Information Extraction (IE)
Why bother about NLP?

Possible solution to these problems

Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content

Clinical Information Extraction (IE)
Possible solution to these problems

Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content

⇒ Clinical Information Extraction (IE)
Why bother about NLP?

Possible solution to these problems
Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content
⇒ Clinical Information Extraction (IE)
Why bother about NLP?

Possible solution to these problems

Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content

⇒ Clinical Information Extraction (IE)
Why bother about NLP?

Possible solution to these problems

Natural Language Processing (NLP) can be used to automatically extract various types of clinical information from the EHR narrative text content

⇒ Clinical Information Extraction (IE)
Why bother about NLP?

Clinical information extraction
Information Extraction (IE) involves extracting predefined types of information from text (e.g., people, diseases, symptoms, medication doses).

The development of IE applications in the biomedical domain has been far more advanced than in the clinical domain, mostly for two reasons:

– Extremely limited availability of clinical text because of patient confidentiality and privacy reasons.

– Characteristics of clinical text and related difficulties to analyze it.
The Automated Problem List

What was the problem?

An electronic problem list was already available at IHC, but
• it was often incomplete, inaccurate, not timely…not used!
• the problem list was becoming a central component of the
  EHR and was to be used by many applications (CPOE,
  documentation, knowledge access, etc.).
Clinical Information Extraction

The Automated Problem List

What was the problem?

An electronic problem list was already available at IHC, but
• it was often incomplete, inaccurate, not timely...not used!
• the problem list was becoming a central component of the
  EHR and was to be used by many applications (CPOE,
documentation, knowledge access, etc.).

▷ Need for a problem list of good quality
   (complete, accurate, timely, and coded)
## The Automated Problem List

![Screen capture of the Automated Problem List](image.png)

<table>
<thead>
<tr>
<th>Status</th>
<th>Rw</th>
<th>Problem</th>
<th>Type</th>
<th>Onset Dt</th>
<th>Noted Dt</th>
<th>Rw/Dt</th>
<th>Clinician</th>
<th>Body System</th>
<th>POC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>Dyspnea</td>
<td>Drag</td>
<td>10/29/04</td>
<td></td>
<td></td>
<td>MEYSTRE, STE</td>
<td>LDS Hospital</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>Back pain</td>
<td>Drag</td>
<td>10/29/04</td>
<td></td>
<td></td>
<td>MEYSTRE, STE</td>
<td>LDS Hospital</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>Arrhythmia</td>
<td>Drag</td>
<td>10/29/04</td>
<td></td>
<td></td>
<td>MEYSTRE, STE</td>
<td>LDS Hospital</td>
<td></td>
</tr>
</tbody>
</table>
The Automated Problem List

Information Extraction application components

- Pre-processing
  - Section detection
  - Sentence detection
  - Disambiguation

- NLP module
  - Problems detection
    - MMTx
    - The patient has <C0013404> but no <C0002962>.
  - Negation detection
    - NegEx
    - The patient has <C0013404> but no <C0002962>.

- Post-processing
  - Disambiguation
  - Negation reconciliation
  - Mapping to local codes
  - CDA document creation

Medical problems:
- Dyspnea C0013404 82573 ...
- Pneumonia C0032285 83253 ...
- Anemia C0002871 82531 ...
- Asthma C0004096 82565 ...
The Automated Problem List

[Image of the Automated Problem List interface]

- **Dyspnea**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Back pain**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Anxiety**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Pulmonary embolus**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Deep vein thrombosis**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Pulmonary Edema**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Heart failure**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Heart Block**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Aortic valve insufficiency**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Mitral insufficiency**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
- **Mitral stenosis**: Diag: 10/29/04, Comments: Problem was originally automatically extracted from text, Enc: 0-LDS Hospital
The Automated Problem List

HOSPITAL COURSE
On arrival in my office, he was cyanotic and had a saturation of 70%. He was transported by ambulance to the hospital where he was admitted in respiratory distress. His HR was 170 at that time. During his hospital stay, he was thought to be in congestive heart failure due to his ascites in his extremities and his enlarged heart with evidence of pulmonary edema on chest X-ray. He was intubated for several days down to 2 to 3 liters. He was given diuretics. His baseline creatinine was 1.6 and the goal was to diurese him until that increased. Despite loss of significant weight, he was still hypoxic. He was on adequate anticoagulation but concern was raised that he might still have a pulmonary embolus, so he was sent to PE lab, and was found to have chronic DVTs in both legs. He was assumed to be having pulmonary embolus because of this and ultimately I got a spiral CT scan which demonstrated this pulmonary embolus. Coumadin was reversed with vitamin K and his heparin was continued with a goal of heparinizing it and placing an IVC filter. On 10/03/2000, he had increased oxygen requirements so it was thought it was not going to be possible to discharge him to home.

DISPOSITION
Social work discussed his illness with his son and daughter and arrangements were made to transfer him to St. John’s Hospital for rehabilitation.

DISCHARGE MEDICATIONS
One baby aspirin a day, and oxygen at 8 liters by nasal cannula. He was to follow up with me one week after his discharge from the nursing home.

DISCHARGE DIAGNOSES
1. Pulmonary embolus
2. Chronic deep vein thrombosis
3. Inferior vena cava filter placement
4. Chronic hypoxia
5. Sleep apnea
The Automated Problem List

Information Extraction application versions comparison

- MPLUS
- MMTx + NegEx
- Keyword search
- Reviewers

Precision vs. Recall

Fallout vs. Recall
The Automated Problem List

Information Extraction application versions comparison

- MPLUS
- MMTx + NegEx
- Keyword search
- Reviewers

Precision vs Recall

Fallout vs Recall
The Automated Problem List

Information Extraction application versions comparison

- MPLUS
- MMTx + NegEx
- Keyword search
- Reviewers

![Graph showing comparison of recall and precision for different methods.](image-url)
The Automated Problem List

Information Extraction application versions comparison

- **MPLUS**
- **MMTx + NegEx**
- **Keyword search**
- **Reviewers**

Precision vs. Recall and Fallout vs. Recall graphs.
Study population

- Adult inpatients in the ICU and the Cardiovascular Surgery unit (LDS Hospital, Salt Lake City, UT)
- Control group: physicians had access to the standard electronic problem list
- Intervention group: physicians had access to the APL (i.e. with proposed problems)

<table>
<thead>
<tr>
<th></th>
<th>All patients</th>
<th>ICU patients</th>
<th>CVS patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial controls</td>
<td>76</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>RCT: Tests</td>
<td>88</td>
<td>54</td>
<td>34</td>
</tr>
<tr>
<td>RCT: Controls</td>
<td>83</td>
<td>51</td>
<td>32</td>
</tr>
<tr>
<td>TOTAL</td>
<td>247</td>
<td>149</td>
<td>98</td>
</tr>
</tbody>
</table>
# The Automated Problem List

## APL system RCT results

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.102 (0.069-0.135)</td>
<td>0.998 (0.995-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.266 (0.192-0.34)</td>
<td>0.993 (0.988-0.999)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.815 (0.771-0.859)</td>
<td>0.957 (0.947-0.966)</td>
</tr>
<tr>
<td><strong>ICU patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.089 (0.049-0.129)</td>
<td>0.999 (0.998-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.41 (0.308-0.512)</td>
<td>0.989 (0.98-0.998)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.774 (0.714-0.835)</td>
<td>0.963 (0.95-0.976)</td>
</tr>
<tr>
<td><strong>CVS patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.123 (0.063-0.182)</td>
<td>0.995 (0.989-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.037 (0.013-0.063)</td>
<td>1</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.88 (0.823-0.938)</td>
<td>0.947 (0.933-0.96)</td>
</tr>
</tbody>
</table>
## The Automated Problem List

### APL system RCT results

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.102</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>(0.069-0.135)</td>
<td>(0.995-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.266</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>(0.192-0.34)</td>
<td>(0.988-0.999)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.815</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>(0.771-0.859)</td>
<td>(0.947-0.966)</td>
</tr>
<tr>
<td><strong>ICU patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.089</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>(0.049-0.129)</td>
<td>(0.998-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.41</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>(0.308-0.512)</td>
<td>(0.98-0.998)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.774</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>(0.714-0.835)</td>
<td>(0.95-0.976)</td>
</tr>
<tr>
<td><strong>CVS patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.123</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>(0.063-0.182)</td>
<td>(0.989-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.037</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(0.013-0.063)</td>
<td></td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.88</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>(0.823-0.938)</td>
<td>(0.933-0.96)</td>
</tr>
</tbody>
</table>
## The Automated Problem List

### APL system RCT results

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.102</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>(0.069-0.135)</td>
<td>(0.995-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.266</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>(0.192-0.34)</td>
<td>(0.988-0.999)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.815 (0.771-0.859)</td>
<td>0.957 (0.947-0.966)</td>
</tr>
<tr>
<td><strong>ICU patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.089</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>(0.049-0.129)</td>
<td>(0.998-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.41</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>(0.308-0.512)</td>
<td>(0.98-0.998)</td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.774 (0.714-0.835)</td>
<td>0.963 (0.95-0.976)</td>
</tr>
<tr>
<td><strong>CVS patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>0.123</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>(0.063-0.182)</td>
<td>(0.989-1)</td>
</tr>
<tr>
<td>Tests</td>
<td>0.037</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.013-0.063)</td>
<td></td>
</tr>
<tr>
<td>Tests with proposed probs.</td>
<td>0.88 (0.823-0.938)</td>
<td>0.947 (0.933-0.96)</td>
</tr>
</tbody>
</table>
Clinical Information Extraction

**T extractor (for medications extraction)**

Why extract medications?

– CPOE and e-Prescribing systems becoming widely available in the healthcare system.

– Large proportion of medications actually taken by the patient only mentioned in narrative clinical text documents in the EHR:
  • Prescribed in another institution or private practice,
  • Bought over-the-counter,
  • Prescribed before the introduction of CPOE.
Textractor (for medications extraction)

Why extract medications?
- CPOE and e-Prescribing systems becoming widely available in the healthcare system.
- Large proportion of medications actually taken by the patient only mentioned in narrative clinical text documents in the EHR:
  - Prescribed in another institution or private practice,
  - Bought over-the-counter,
  - Prescribed before the introduction of CPOE.

⇒ This is where automated medication information extraction could (and will) help!
2009 i2b2 Medication Extraction Challenge Task:
Identification of medications and related details in discharge summaries:

- Medication name (brand name, generic name, drug class)
- Dosage
- Route
- Frequency of the administration (incl. “as needed”)
- Duration of the treatment
- Reason(s) for the prescription

Corpus
1249 discharge summaries, de-identified and re-identified with realistic surrogates, split in a training corpus (696 documents) and a testing corpus (553 documents).
**Textractor**

### Application details

Built on UIMA, with functionalities developed as modules (TAEs), organized in a pipeline.

Several components based on OpenNLP tools, wrapped and retrained.

Uses MMTx (version 2.4.C), wrapped for UIMA in its first version. Dictionary lookup now based on Lucene.

Context analyzer based on the ConText algorithm, with improvements for this task.

Regular expressions used to extract dosage, route, etc.
# Textractor (Medications Extraction)

## 2009 i2b2 Medication Extraction Challenge Evaluation:

<table>
<thead>
<tr>
<th>Fields</th>
<th>Exact match</th>
<th></th>
<th></th>
<th></th>
<th>Inexact match</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>F$_1$-measure</td>
<td>Recall</td>
<td>Precision</td>
<td>F$_1$-measure</td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>Medication</td>
<td>0.746</td>
<td>0.772</td>
<td>0.759</td>
<td>0.763</td>
<td>0.784</td>
<td>0.773</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td>0.757</td>
<td>0.916</td>
<td>0.829</td>
<td>0.786</td>
<td>0.925</td>
<td>0.850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route</td>
<td>0.817</td>
<td>0.920</td>
<td>0.865</td>
<td>0.803</td>
<td>0.926</td>
<td>0.860</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>0.789</td>
<td>0.892</td>
<td>0.837</td>
<td>0.742</td>
<td>0.924</td>
<td>0.823</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>0.326</td>
<td>0.397</td>
<td>0.358</td>
<td>0.326</td>
<td>0.501</td>
<td>0.395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reason</td>
<td>0.169</td>
<td>0.669</td>
<td>0.270</td>
<td>0.148</td>
<td>0.703</td>
<td>0.245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.715</td>
<td>0.832</td>
<td>0.769</td>
<td>0.693</td>
<td>0.839</td>
<td>0.759</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Textractor (Medications Extraction)

2009 i2b2 Medication Extraction Challenge Evaluation:

<table>
<thead>
<tr>
<th>Fields</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>0.746</td>
<td>0.772</td>
</tr>
<tr>
<td>Dose</td>
<td>0.757</td>
<td>0.916</td>
</tr>
<tr>
<td>Route</td>
<td>0.817</td>
<td>0.920</td>
</tr>
<tr>
<td>Frequency</td>
<td>0.789</td>
<td>0.892</td>
</tr>
<tr>
<td>Duration</td>
<td>0.326</td>
<td>0.397</td>
</tr>
<tr>
<td>Reason</td>
<td>0.169</td>
<td>0.669</td>
</tr>
<tr>
<td>Overall</td>
<td>0.715</td>
<td>0.832</td>
</tr>
</tbody>
</table>

Duration and reason for a prescription annotations were difficult to extract, and difficult to annotate by humans!

<table>
<thead>
<tr>
<th></th>
<th>Our team’s exact match IAA</th>
<th>Experts’ Exact F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>0.16</td>
<td>0.47</td>
</tr>
<tr>
<td>Reason</td>
<td>0.31</td>
<td>0.40</td>
</tr>
</tbody>
</table>
**Textractor (Medications Extraction)**

2009 i2b2 Medication Extraction Challenge Evaluation:

<table>
<thead>
<tr>
<th>Fields</th>
<th>Exact match</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>0.746</td>
<td>0.772</td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td>0.757</td>
<td>0.916</td>
<td></td>
</tr>
<tr>
<td>Route</td>
<td>0.817</td>
<td>0.920</td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>0.789</td>
<td>0.892</td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td><strong>0.326</strong></td>
<td><strong>0.397</strong></td>
<td></td>
</tr>
<tr>
<td>Reason</td>
<td>0.169</td>
<td>0.669</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.715</td>
<td>0.832</td>
<td></td>
</tr>
</tbody>
</table>

Duration and reason for a prescription annotations were difficult to extract, and difficult to annotate by humans!

<table>
<thead>
<tr>
<th></th>
<th>Our team’s exact match IAA</th>
<th>Experts’ Exact F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td><strong>0.16</strong></td>
<td><strong>0.47</strong></td>
</tr>
<tr>
<td>Reason</td>
<td><strong>0.31</strong></td>
<td><strong>0.40</strong></td>
</tr>
</tbody>
</table>
Conclusions

For customers of automatically extracted clinical information, or CDS based on this information, trust can be based on several factors:

- Ensuring that there is a **human expert in the loop** for all clinical information generated by NLP and becoming part of the EHR.
- Allowing users to know the **origin of the extracted information** (and the methods used; and therefore use NLP methods that make it possible).

With trust in the NLP-generated information, even moderate performance is acceptable (moderate sensitivity, but sufficient PPV).