# Automating Fact Extraction from Medical Records

Cheryl Clark

(781) 271-7975 • cclark@mitre.org

MSR



# Problem

- Critical medical information is buried in the free text of electronic medical records
- Medical researchers want to use automatic systems to analyze notes in large batches to uncover new insights about diseases, treatments, patient background and behaviors, etc.
- Current medical extraction systems identify certain medical concepts well, but serious gaps exist in these technologies that limit their ability to extract the relevant information
- Current technology often misinterprets expressions of negation and uncertainty, and disregards temporal information, leading to incomplete or misleading representations of concept status

# Background

He did not have chest pain until approximately 1-1/2 years ago.

Processing concludes *pain* is **negated** despite presence of *until*

Processing extracts *Dilaudid* as **current** med despite information conveyed by *discontinued*

Her pain was controlled with Dilaudid , but she developed some symptoms of respiratory depression , so this was discontinued.

**Disregarding temporal information results in incorrect status assignment for medical concepts**

# Objective

- **Build a system for automated fact extraction from medical free text that:**
  - **leverages the functionality of an existing open source extraction system (cTAKES)**
  - **supplies crucial information needed to understand meaning and determine the status of medical concepts**
- **Develop algorithms that interpret temporal information by**
  - **identifying temporal cues and their span of influence**
  - **evaluating interaction between temporal information and negation and uncertainty**
  - **computing status of medical problems and medications mentioned in free text of electronic patient records**

# Activities

- Using Mayo Clinic's Clinical Text Analysis and Knowledge Extraction System (cTAKES) to extract medical concepts (e.g., medications, diseases, procedures)
- Improving the negation and uncertainty detection module developed in previous year
- Collaborating with Brandeis University researchers to extend functionality of temporal detection toolkit built for news reports so that it can be applied to detect and interpret temporal information in clinical reports
- Collaborating with Boston and Bedford VA researchers to apply negation, uncertainty, and temporal detection technology in study of patients with ischemic stroke

# Highlight

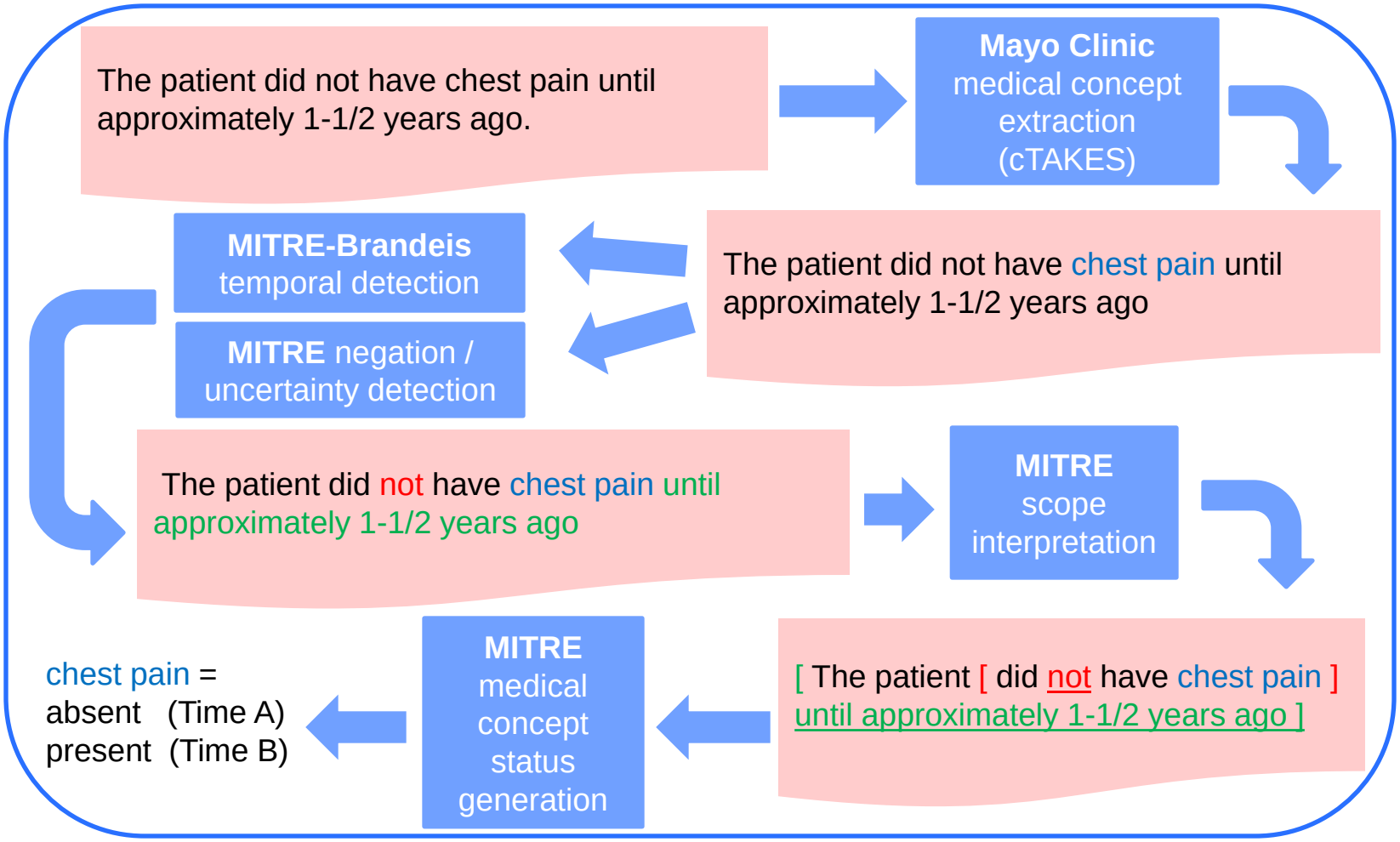
MITRE Innovation Program

## Classification of Assertions Made on Medical Problems

| | | |
|---|---|---|
| Patient had a **stroke** | → | *Present* |
| History inconsistent with **stroke** | → | *Absent* |
| We are unable to determine whether she has **leukemia** | → | *Possible* |
| Patient reports **shortness of breath** upon climbing stairs | → | *Conditional* |
| If you experience **wheezing** or **shortness of breath** | → | *Hypothetical* |
| Family history of **prostate cancer** | → | *Not Associated with Patient* |

**MITRE system for clinical assertion status classification ties for first place in Fourth i2b2/VA Challenge**

# Highlight



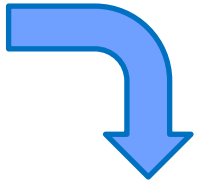

**Interpreting interaction of *not* and *until* to determine status of medical problem chest pain**

# Impacts

- We are developing a single core technology that can enable many applications that address needs expressed by our healthcare sponsors such as
  - the analysis of physicians' free-text notes for medical discovery
  - patient cohort selection
- We are working with Boston and Bedford VA researchers to develop and validate tools that can be applied to their clinical research needs
- Our research is positioning MITRE as a key pathfinder in the development of advanced language processing technologies for healthcare

# Future Plans

HOSPITAL COURSE BY SYSTEM:
**…**
Her urine was positive for greater than 100 , 000 group B hemolytic Streptococci and she was initially **started with Levaquin**. It was **switched to penicillin on January , 2005** , after the gram-positive organism was discovered and she was **switched back to Levaquin after the sensitivities came back** , and it was sensitive to Levaquin. The Levaquin is ideal because her coughing could also represent bronchitis. She has complained of a **cough since weeks before admission** , which is **now resolving**.



**Problems**

| | |
|---|---|
| **Cough** | |
| Present | Pre-encounter |
| Present (improving) | Encounter |

**Encounter Medications**

| | | |
|---|---|---|
| **Levaquin** | Taken | Time span A |
| | Not taken | Time span B |
| | Taken | Time span C |
| **Penicillin** | Not taken | Time span A |
| | Taken | Time span B |
| | Not taken | Time span C |

**Interpreting the language of change in free text**