

## MedLEE

### A Medical Language Extraction and Encoding System

#### What is MedLEE?

MedLEE is a text processor that extracts and structures clinical information from textual radiology reports and translates the information to terms in a controlled vocabulary, such as the UMLS or SNOMED. Clinical information can then be accessed by further automated procedures. Although the processor has been applied to the domains of radiology, discharge summaries, sign out notes, pathology reports, electrocardiogram reports, and echocardiogram reports, the design is extensible so that it can readily be ported to other clinical domains. MedLEE generates XML output and has the ability to incorporate local terminology.

#### MedLEE's Unique Advantage

MedLEE is the only comprehensive and operational natural language processor (NLP) system formally shown to be as accurate as physicians in interpreting narrative medical records. MedLEE is recognized as the most effective NLP for medical narratives in many medical subspecialties. In numerous evaluations carried out by independent users, MedLEE performed well. Evaluators concluded

that the system was safe for use in real-world clinical applications.

#### Applications

MedLEE is a crucial component of an innovative knowledge management tool. It allows for the automation and simplification of decisions when integrated into a clinical information system.

MedLEE\* can be configured for applications other than medical reports.

In recent clinical use, MedLEE has been used in a patient safety program to detect a broad range of medical events: the 45 patient-specific events defined in the New York Patient Occurrence Reporting and Tracking System. The system achieved very high performance, enabling broad screening for medical events.

MedLEE has also been applied to improve safety through intervention. For example, MedLEE detects patients at high risk for having active tuberculosis and recommends respiratory isolation. Its use has reduced the rate of missed tuberculosis by almost one half at New York-Presbyterian Hospital.

#### Summary

##### *MedLEE*

- Extracts information allowing data to be used in automated processes
- Has accuracy comparable to that of physicians
- Is safe for real-world clinical use
- Is able to encode data in a variety of formats, including XML
- Has been proven effective in use at New York Presbyterian Hospital
- Has a simple method to incorporate local terminology
- Is the product of over 20 years of cutting edge clinical and natural language research
- Is available for evaluation and licensing

\*A new service is in the offing that will provide customized MedLEE applications.

## Example of MedLEE Natural Language Processing

The frames below illustrate the use of MedLEE to extract structured information from a plain text file. Figure 1 is input of a patient history file, and Figure 2 is the MedLEE output for the same text file. Note that the granularity of the structured output augments the value of the information and enables automated applications.

Figure 1. Input to MedLEE

HISTORY OF PRESENT ILLNESS: The 52 year old women presented to the emergency room with three episodes of severe right orbital headache lasting 2 hours and occurring every day with rapid onset, accompanied with nausea and vomiting, **without paresia**. Patient **denies myalgia**. Positive **family history for SLE**.  
 PHYSICAL EXAMINATION: Alert and oriented times three. Normal neurological examination.  
 LABORATORY DATA: ESR 37 mm/hr.  
 ASSESMENT and PLAN: The patient was hospitalized with a **preliminary diagnosis of temporal arteritis** treated with Prednisone 60 mg twice a day. Head CAT scan **ruled out subarachnoid hemorrhage**. Temporal artery biopsy planned tomorrow.

Figure 2. Output of MedLEE

Class		Qualifier	
Name	Value	Name	Value , Unit
finding:	women	Age =	52 , year
problem :	headache	bodyloc =	orbital
		region =	right
		degree =	high degree
		duration =	2 , hours
		quantity =	3 , episodes
		change_rate =	fast
problem:	nausea	certainty =	high certainty
problem:	vomiting	certainty =	high certainty
problem:	<b>paresia</b>	certainty =	<b>absence of</b>
problem:	<b>myalgia</b>	certainty =	<b>absence of</b>
problem:	<b>systemic lupus erythematosus</b>	status =	<b>family history</b>
normalfinding:	alert and oriented times 3		
normalfinding:	examination	bodyfunc =	neurologic
labtest:	erythrocyte sedimentation rate	measure =	37 , mm/hr
problem:	<b>temporal arteritis</b>	certainty =	<b>moderate certainty</b>
		status =	<b>preliminary</b>
		time period =	admission
medication:	prednisone	certainty =	high certainty
		dose =	60 , mg
		frequency =	bid
problem:	<b>subarachnoid hemorrhage</b>	certainty =	<b>absence of</b>
procedure:	computerized axial tomography	bodyloc =	head
procedure:	biopsy	bodyloc =	temporal artery
		timeper =	tomorrow
		status =	plan

### Intellectual Property / Technology Position

US patent No. **6,055,494** for *System And Method For Medical Language Extraction And Encoding*, issued on Apr. 25, 2000 and No. **6,182,029** for *System and Method For Language Extraction And Encoding Utilizing The Parsing Of Text Data In Accordance With Domain Parameters*, issued on Jan. 30, 2001. Inventor: Carol Friedman

MedLEE is available for licensing.

### Technical Contact

Carol Friedman, Ph.D., Professor, Department of Biomedical Informatics, Columbia University  
 e-mail: carol.friedman@dbmi.columbia.edu

### For more information

Vincent Tomaselli, Deputy Director; Center for Advanced Information Management, voice: 212.305.2944;  
 e-mail: tomaselli@cat.columbia.edu