A knowledge model for the interpretation and visualization of NLP-parsed discharge summaries

Michael Krauthammer, MD[1] and George Hripcsak, MD MS[1]

[1]Department of Medical Informatics, Columbia University, New York

At our institution, a Natural Language Processing (NLP) tool called MedLEE is used on a daily basis to parse medical texts including complete discharge summaries. MedLEE transforms written text into a generic structured format, which preserves the richness of the underlying natural language expressions by the use of concept modifiers (like change, certainty, degree and status). As a tradeoff, extraction of application-specific medical information is difficult without a clear understanding of how these modifiers combine. We report on a knowledge model for MedLEE modifiers that is helpful for a high level interpretation of NLP data and is used for the generation of two distinct views on NLP-parsed discharge summaries: A 'physician view' offering a condensed overview of the severity of patient problems and a 'data mining view' featuring binary problem states useful for machine learning.

INTRODUCTION

Natural Language Processing (NLP) in the medical domain offers unique opportunities for extending the use of the electronic medical record (EMR) by converting narrative text into coded data [1]. The codes can be used for such tasks as automated decision support and classification of medical information using machine-learning algorithms, which automatically generate rules for classifying medical conditions [2].

It has been shown that machine-learning techniques critically depend on the underlying medical text representation [3]. Normally, the coded output of NLP parsers is very complex, and may therefore not be useful for machine learning without some kind of interpretation and simplification.

This paper describes a knowledge model for interpretation of NLP-parsed discharge summaries. The model enables the extraction of information at different levels of complexity. The rationale behind this approach is that queries on the parsed discharge summaries are completely application dependent and require different level of representational granularities. For example, a machine-learning algorithm is best served with a simple binary representation of diseases (disease present or absent, respectively) while a 'physicians view' on the same discharge summary may involve different levels of disease severities. The task of the knowledge model therefore is to condense the rich representation of the original NLP output to the desired level of granularity for interfacing with different kinds of applications. We present two such interfaces in the form of dedicated views on the underlying discharge summaries.

BACKGROUND

At Columbia Presbytarian Hospital, a NLP processor, MedLEE[4], originally designed for parsing radiology reports, has been recently expanded to cover many relevant medical domains and patient discharge summaries [5]. MedLEE's architecture is fairly flexible and has been integrated in such systems like decision support[6], coding [7] and vocabulary design[8].

MedLEE generates a conceptual representation of discharge summaries. In the MedLEE XML output (Fig. 1), concepts are represented as higher lever nodes in the XML tree (problems, findings, etc.) while concept-modifiers are represented as child-elements (degree, certainty, etc.). A MedLEE XML DTD defines how modifiers can be combined: For example, the disease concept *Myocardial infarction* can be combined with status-modifier *past history*, indicating that the disease is part of the patient's medical history. On the other hand, the same concept with the certainty-modifier *not likely* indicates the probable absence of the disease. It is also possible to nest modifiers, which gives rise to a multitude of possible concepts-modifiers pairs.

It has been shown that applications critically depend on the exact structure of MedLEE's output[9], and mechanisms have been integrated to 'simplify' the output by mapping modifier values to a limited set of terms[10].

This paper extends the latter idea by proposing a taxonomic hierarchy for MedLEE modifiers (Fig.2) that can be used to extract high level data representations such as the absence or presence of problems or findings mentioned in the discharge summary. For example, the model evaluates whether

1067—5027/01/$5.00 © 2001 AMIA, Inc.

a problem is part of the personal history of a patient by examining status modifiers with values *previous* or *post*, which indicate the (current) absence of a problem. Degree and certainty modifiers may also indicate the absence of a disease, as is the case with modifier values *negligible* or *very low certainty*, respectively. These state conditions can be modeled in a knowledge base where modifier values with similar meanings are grouped together. Such a taxonomy differs from previous attempts to model clinical (time) data [11] as it models structured statements from discharge summaries instead of time-stamped clinical events. It also differs form a general description of the content of discharge summaries, as is provided by MENELAS [12] and MedLEE's[4] own semantic grammar, as its main purpose is to simplify the interfacing with other applications.

Recently, many knowledge representation tools such as Protégé-2000[13] have become available that facilitate the construction of taxonomic knowledge bases. Protégé-2000 is a frame-based knowledge base editor which features multiple inheritance for classes (concepts), distinct views for slots, forms, classes and instances (see Fig. 2) as well as a JAVA interface for integration into larger projects. In this paper, Protégé-2000 interfaces with a XML parser and a XSL processor [14] which facilitate the access and manipulation of XML documents.

Fig.1. MedLEE XML output

```
<problem v = "disease" idref = "p64"> <bodyloc v =
"coronary artery" idref = "p60"></bodyloc>
<certainty v = "high certainty" idref =
"p52"></certainty><status v="family history">
</status> </problem>
```

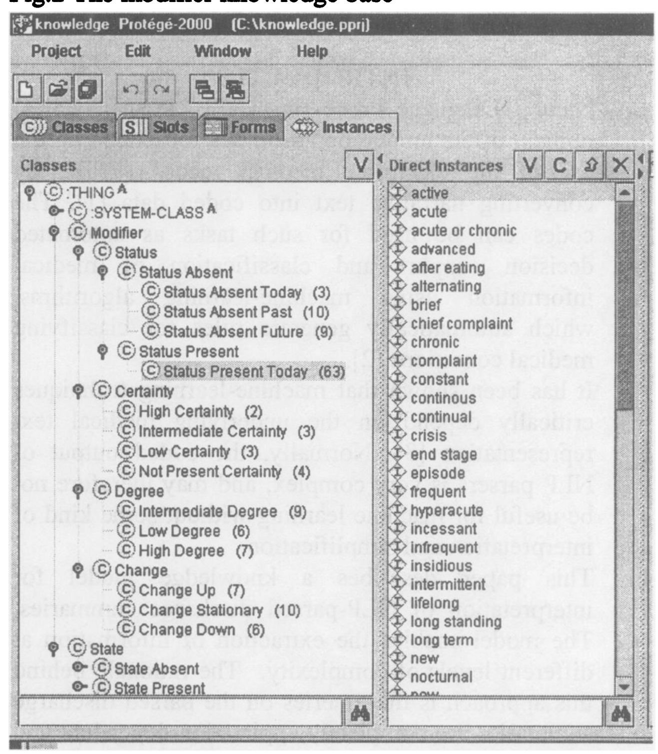
METHODS

We used Protégé-2000 to model the knowledge base of MedLEE modifiers (Fig. 2). The model integrates 4 different MedLEE modifiers: *Status*, *Change*, *Degree* and *Certainty* which are modeled as unique concepts in the knowledge base. They are subsumed by a top-level concept, called *Modifier*. Individual modifier values are added as individuals or instances of these concepts: For example, *Status Present Today* – a grandchild of *Status* - groups modifier values that indicate the current presence of a problem, such as *constant*, *lifelong* and *acute*. Fig.2 shows the modifier value *active* listed as an instance of *Status Present Today*.

In addition to the modifier concepts and instances, the knowledge base includes a top-level concept *State*, which defines the state of a finding or problem according to its modifier combinations. Two subclasses of *State*, *State Absent* and *State Present*, subsume concepts with relationships to different modifier concepts. For example, *State Absent* subsumes a concept with a relationship to the modifier concept *Status Absent*. This indicates that problems with modifier values that are instances of *Status Absent* – or its subclasses *Status Absent Today*, *Status Absent Past* and *Status Absent Future* - should be assigned a state *absent*. The model also handles more complex situation, such as a problems with two modifiers influencing the presence or absence of a problem. For example, degree modifier *low degree* with change modifier *regression* can represent an absence of a problem.

The following steps are used to generate the different views on the discharge summaries (Fig. 4.): First, a XML parser, which has access to the Protégé knowledge base, parses and *transforms* the MedLEE XML output as follows: The state of a problem is determined according to the modifier combinations. For example, a problem *fever* with modifier value *acute* would be assigned a state *present* based on the fact that *acute* is listed under *Status Present Today* which has a relationship to the state concept *State Present*. After determining the problem state, the XML output is annotated with a new state attribute, which codes the absence or presence of a problem (see Fig. 3).

Fig.2 The modifier knowledge base



After determining the state of a problem, the knowledge model annotates each problem modifier with its type. For example, *family history* is annotated with type *Status Absent Past* (see Fig 1. and 3.)

As a final step, two kinds of XSL style sheets are used to generate the resulting views (Fig. 5 and 6) on the discharge summaries. The style sheets use the annotations from the previous steps and transform the complex XML format into a much simpler HTML document that can be displayed in any HTML client. Style sheets encode filter rules for the XML documents, and can be easily adapted for new purposes.

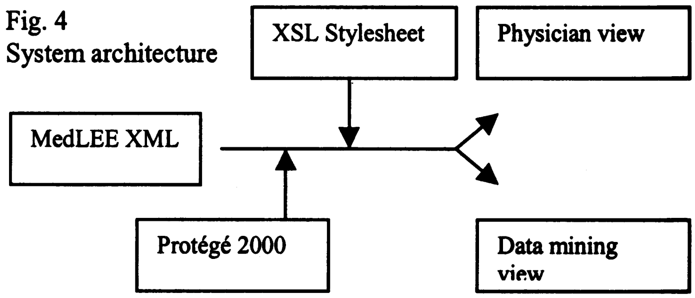
Fig. 3. Transformed and annoted XML output

```
<problem v = "disease" idref ="p64" state="absent">
<bodyloc v = "coronary artery" idref = "p60">
</bodyloc> <certainty v = "high certainty" idref =
"p52" type="High Certainty'> </certainty> <status
v="family history" type="Status Absent
Past"></status></problem>
```

RESULTS

Fig. 5. and 6. show two views of the same discharge summary. The 'physician view' (Fig. 5.) depicts the written text of the discharge summary in the right-hand column. The column to its left shows problems and findings as detected by the NLP processor. The categories *Problems (absent/present), Finding (absent/present)*, *History* and *Risk* reflect the changes applied to the original XML output by the knowledge model. Problems and findings are separated according to their state (absent/present) and ordered as follows: Severe problems (underlined) are listed on top, followed by intermediate-degree problems with increasing severity (arrow up), and so on. Problems that occurred previous to the current hospitalization are listed under *History*, while problems that may occur in the future are listed under *Risk*. The 'data mining' view is simpler (Fig. 6.): It lists problems and findings with their respective state (absent or present).

Fig. 4
System architecture



EVALUATION

A preliminary evaluation of the system involved an assessment of how well the system discriminates between different state categories. We compiled a list of MedLEE-identified problems and findings from 4 randomly chosen discharge summaries.
The list was stripped from any additional information (i.e., MedLEE modifiers). 2 physicians read the same discharge summaries with the task of placing each item on the list in the category ***state absent*** or ***present***.
The system agreed with the physicians in 58 out of 63 instances about the presence or absence of a problem or findings. This results in a system accuracy of 93.5%. The physicians were also asked to assess any changes in problem severity. Unfortunately, a quantitative comparison to the system's assessment of change was not possible because MedLEE did annotate only selected problems with change modifiers. An example of a situation where the physician report change information on a problem where no MedLEE change information is present will be discussed in the next section.

DISCUSSION

We present a knowledge model for a high level interpretation of NLP data. The model is useful for assessing the state (presence or absence), degree and

Fig. 5.'Physican view'

report history of present illness item	**Problems present:** diarrhea discomfort (abdomen) ⇑ vomiting lightheaded unconscious **Problems absent:** pain (chest) seizure **Findings present:** demo (67 year) **History:** syncope	HISTORY OF PRESENT ILLNESS: This 67 year old with a history of syncope in 1987 and 1989. She reported that she was evaluated both times and the work up was negative for any specific etiology. On the day of admission she reports having one episode of severe diarrhea and she was having increasing abdominal discomfort with flatulence and one episode of vomiting. When she returned to the bath room to move her bowels again she felt light headed and called for a family member. The family member reported that the patient was unconscious at that time and was placed in bed and recovered within 1-2 minutes. there was no history of any precipitating shortness of breath, chest pain or any seizure activity. At the time the patient was seen in the hospital she already felt fine.

severity change of medical problems and findings.

Fig. 6.'Data mining view'

report history of present illness item	Demo (67 years)	present
	Syncope	absent
	diarrhea	present
	discomfort (abdomen)	present
	vomiting	present
	lightheaded	present
	unconscious	present
	pain (chest)	absent
	seizure	absent

The complexity of the underlying NLP output is reduced to condense summary views (Fig. 5 and 6.) on the same data.

The usefulness of generating a high level representation of medical text has been recently discussed in studies about inductive learning techniques [2, 3]. Wilcox and Hripcsak [3] show that differences in the performance of learning algorithms chiefly depend on the medical text representation format. Using a simplified NLP format with just one disease state, Chapman and Haug [2] demonstrate that machine-learning algorithm were able to produce rules performing similar as expert written rules for pneumonia identification. In light of the latter study, which is limited to the domain of radiology, an important finding of this paper is the feasibility of 'simplifying' NLP data in an *extended* domain of medicine, as represented by a discharge summary.

In light of problems with other sources of medical information, such as provided by claims data, which proved to be unreliable for such tasks as outcome prediction [16, 17], this simplification or abstraction of data is important for successfully mining of complete discharge summaries. This seems somewhat paradoxical, given that until recently, the *lack* of sufficient granularity of EMR data [15] was thought to be a major obstacle for a successful introduction of large scale medical data mining. This situation has remarkably changed: With the introduction of NLP to the medical domain, many new information sources became accessible. But the new wealth of medical data has its drawbacks: How to best select parameters of importance remains an open question. In this light, this study provides a framework for accessing and converting rich medical information to a simpler, higher-level representation. The open architecture of our system is helpful: The knowledge base of concept modifiers can be easily adapted for new questions that require a different abstraction of the underlying information. For example, exploring data about medical errors may necessitate the modeling of findings that were not adequately pursued during a hospital stay. These findings may be tagged in the text with modifiers *very low certainty* or *borderline certainty*. In the current model, these findings would be considered absent.

The two views on the discharge summaries, a detailed 'physician view' and a simpler 'data mining view' demonstrate the capability of the model to extract different levels of complexity. While the 'data mining' view represents the high level simplification and abstraction of the discharge summary discussed above, the 'physician view' offers a quick overview of patient problems and findings. The ordering of problems, as well as high lightening of parameters such as degree and change, is helpful to conveniently browse through a large amount of discharge summaries. This may be useful for collecting a training set of a few hundred discharge summaries for machine learning: Test cases can be easily selected by just looking at inclusion criteria among the problem and finding list.

The evaluation showed a high degree of system accuracy for determining the presence or absence of a problem or finding. Physician often disagreed with the system in regard of changes in problem severity. For example, a sentence "At the time the patient was seen in the hospital she already felt fine" (see Fig.5) prompted the physicians to consider problem severities in the 'history of present illness' section as *decreasing*. The NLP processor does not provide this kind of reasoning: none of the problems or findings in Fig.5 has a downward arrow, indication decreasing severity. Language ambiguity was another reason for disagreement. In the sentence "The patient was admitted for increasing shortness of breath, cough and fever" it is unclear which symptoms are increasing in severity. The language processors considered only shortness of breath as increasing, while the physicians considered all three symptoms as worsening.

LIMITATIONS

The study does not address the problem of redundant statements in medical texts. Often, the same problem may occur more than once in a parsed discharge summary. In some instances, these statements may contradict each other, like pneumonia with certainty modifiers *rule out* and *minimal criteria*. Considering whether the contradicting statements occurred in related or in unrelated sections of the discharge summary may be helpful. Otherwise, contradictions may be resolved by summing up statements sharing the same opinion and weighting them against statements with different opinions. Some sort of a majority rule would then decide the absence or presence of a disease.

CONCLUSION

We present a knowledge model for a high level interpretation of NLP data. The model is useful for assessing the state (presence or absence), degree and severity change of medical problems and findings. We evaluated the usefulness of the model in a system for generating different views on NLP-parsed discharge summaries: The system accurately categorized a problem as being absent or present in over 90% of cases. The ability to accurately extract high-level abstractions from NLP-parsed discharge summaries is important, as efficient machine learning is dependent on adequate medical text representation.

This work was supported by National Library of Medicine grants R01-LM06910 and R01-LM06274

1. Spyns, P., *Natural language processing in medicine: an overview.* Methods Inf Med, 1996. **35**(4-5): p. 285-301.
2. Chapman, W.W. and P.J. Haug, *Comparing expert systems for identifying chest x-ray reports that support pneumonia.* Proc AMIA Symp, 1999: p. 216-20.
3. Wilcox, A. and G. Hripcsak, *Medical text representations for inductive learning.* Proc AMIA Symp, 2000: p. 923-7.
4. Friedman, C., et al., *A general natural-language text processor for clinical radiology.* J Am Med Inform Assoc, 1994. 1(2): p. 161-74.
5. Friedman, C., *Towards a comprehensive medical language processing system: methods and issues.* Proc AMIA Annu Fall Symp, 1997: p. 595-9.
6. Hripcsak, G., et al., *Automated tuberculosis detection.* J Am Med Inform Assoc, 1997. 4(5): p. 376-81.
7. Lussier, Y., et al. *Automated ICD-9-CM Encoding Using Medical Language Processing: A feasability study.* in *AMIA.* 2000.
8. Liu, H. and C. Friedman, *A method for vocabulary development and visualization based on medical language processing and XML.* Proc AMIA Symp, 2000: p. 502-6.
9. Hripcsak, G., G.J. Kuperman, and C. Friedman, *Extracting findings from narrative reports: software transferability and sources of physician disagreement.* Methods Inf Med, 1998. **37**(1): p. 1-7.
10. Friedman, C., *A broad-coverage natural language processing system.* Proc AMIA Symp, 2000: p. 270-4.
11. Shahar, Y. and C. Cheng, *Knowledge-based visualization of time-oriented clinical data.* Proc AMIA Symp, 1998: p. 155-9.
12. Zweigenbaum, P., *MENELAS: an access system for medical records using natural language.* Comput Methods Programs Biomed, 1994. **45**(1-2): p. 117-20.
13. Noy, N.F., R.W. Fergereson, and M.A. Musen. *The knowledge model of Protege-2000: Combining interoperability and flexibility.* in *2th International Conference on Knowledge Engeneering and Knowledge Management.* 2000. Juan-les-Pins, France.
14. Seol, Y.H., S.B. Johnson, and J. Starren, *Use of the Extensible Stylesheet Language (XSL) for medical data transformation.* Proc AMIA Symp, 1999: p. 142-6.
15. Maklan, C.W., R. Greene, and M.A. Cummings, *Methodological challenges and innovations in patient outcomes research.* Med Care, 1994. **32**(7 Suppl): p. JS13-21.
16. Jollis, J.G., et al., *Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research.* Ann Intern Med, 1993. **119**(8): p. 844-50.
17. Iezzoni, L.I., *Assessing quality using administrative data.* Ann Intern Med, 1997. **127**(8 Pt 2): p. 666-74.