Intelligent Clinical Notes System

Min LI, mili9528@uni.sydney.edu.au
Supervisor: Prof. Jon Patrick

School of IT
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

1. Objective

• There is great demand for highly accurate and timely Information Retrieval (IR) and Information Extraction (IE) in medicine and health care. To meet this need, we have developed a novel system, Intelligent Clinical Notes System (ICNS) to assist doctors in retrieving clinical notes based on concept searching.

• This system is a start of clinical Question and Answering (Q&A) system. Currently, this system is also used to collect clinical questions for different specialists.

2. Introduction

• In this poster, we present an Intelligent Clinical Notes System (ICNS), built to Intensivists requirements that can retrieve patient notes and extract useful information from them. The main objective of this system is to use natural language processing (NLP) to serve clinicians and improve their productivity and efficiency thus contributing to patient quality and safety. For example, the concept based search engine can automatically identify synonyms for use in a search request. In addition, automatic spelling correction, as well as most of the abbreviations and acronyms are identified and expanded, which makes the clinical notes more readable for the non-author.

• This system has been installed and integrated into the existing clinical information system in the Intensive Care Unit, Royal Prince Alfred Hospital, Sydney.

3. External Resources

• During the development of this system, several external resources were used. They are a medical ontology SNOMED CT, a process for converting text into SNOMED CT terminology, gazetteers and dictionaries.

• SNOMED CT: a comprehensive medical ontology constituting a reference terminology in a relationship hierarchy with approximately 350,000 concepts and 1.4 million relationships.

• Text to SNOMED CT Converter (TTSCT): is utilised in the ICNS by scanning through the whole patient note, identifying all the medical concepts within the free text and mapping them into concept IDs of SCT.

• Gazetteers: word lists of one class of content, such as a list of staff names, new medical terminology, etc which is used to improve the quality of information extraction. The gazetteers have been trained, from a 60 million token ICU corpus, with the content that is unknown to SCT or expressed in forms that are not readily identifiable in SCT

• Dictionaries:

1. Unified Medical Language System (UMLS)

2. Common words (Moby) dictionary,

3. Abbreviation dictionary. It contains nearly 1000 abbreviations with their matching full expressions.

4. Acronym dictionary. It contains more than 500 acronyms with their matching expressions.

5. Misspellings dictionary (a list of misspelt words and their corrected spellings trained from an ICU corpus), with more than 80,000 misspelt words.

3. System Architecture

• The system architecture can be divided into three main components (see figure1). They are the data warehousing component, server system component, and user interface component.

• User interface is used to submit client requests and receive the reorganized results.

• Server system is mainly responsible for processing the client requests, managing the user accounts and organizing the results for presentation.

• Data warehouse processing is responsible for all the substantive tasks, such as indexing, retrieval, proof reading which provides corrections of lexical formation, spelling corrections and canonicalization of non-words such as measurements. It also adds attributes to each token such as its semantic category (see figure 2).
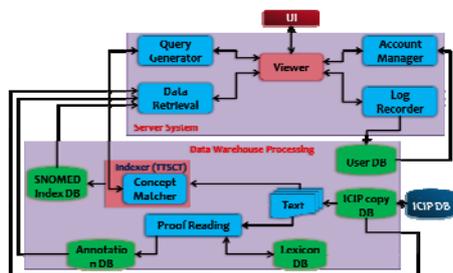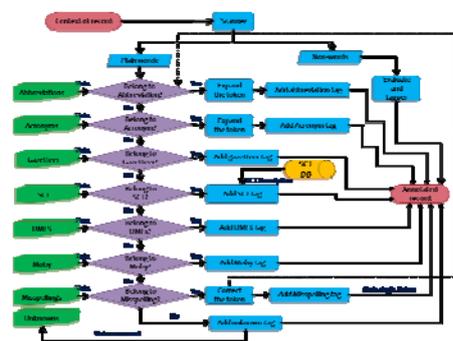

Fig. 1. System Architecture for ICNS

Fig. 2. Proof Reading in ICNS

4. Presentations

• Figure 3 is a screen shot of the retrieval page. Figure 4 is the screen shot for Abbreviation annotation, Acronym annotation and Unknown-words annotation. The question collection page is demonstrated in figure 5.

5. Conclusion

• Currently, comprehensive evaluation for the ICNS has not been completed. However, a small user interview was held for an initial
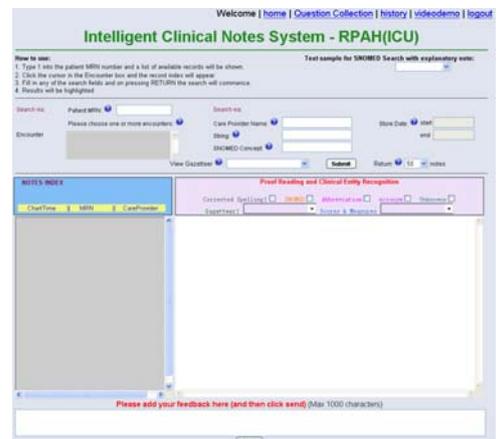

Fig 3. Retrieval page

Fig 4. User Interface selections for Abbreviations (pink), acronym (purple) and unknown-word (blue) annotations, auto-corrected misspellings (in grey), and Scores and Measurements recognition (in turquoise) dropdown list of "Scores & Measurements".

Fig. 5. clinical question collection page

evaluation. 4 doctors were interviewed while they were operating the ICNS, all of them believe the ICNS brings worthwhile convenience to their daily work.

• In the next stage, a fully evaluation for the ICNS will be done. Furthermore, this system will be extended into a clinical Q&A system to answer clinical question based on the patient medical records.

Demonstration page available at:

http://hitrl.it.usyd.edu.au/ICNS/