**Unstructured to structured information conversion for extracting meaningful clinical information from medical notes**

A Ranjan1\*, R Bista2\*\*

1 Kathmandu University, Dhulikhel, Nepal 2 Kathmandu University, Dhulikhel, Nepal

\*awa.ran@gmail.com \*\*rbista@ku.edu.np

**ABSTRACT:** In medical domain, one of the most important document is the notes that doctors, nurses or other medical practitioners take during patient interview. These notes contain vital information about the patient current condition, symptoms, family history, disease, procedures (like x-ray, lab test etc.), medication on and so on. These notes are in plain text English language which is in an unstructured way. An unstructured text can be defined as text which contains information not in a common structured format.

Example - Pt. presents with hyperlipidemia and strong family hx of CAD. Keeps active with job, kids, and softball, but no routine cardio exercise.

Since these note contain very useful clinical information, everyone are willing to extract such vital clinical information. But lack of defined structure, it makes these texts to be interpreted only by humans and not by any computer program

In structured information, we have only the useful information extracted from the text and other unnecessary information are thrown away. This makes the information to be presented in quite definite structure which can be stored in some files or database for further processing. Structured information from above example notes can be extracted as -

• Diagnosis - Hyperlipidemia

• Family History - CAD

• Actions - Job, playing softball, no cardio exercise.

Some advantages of structured information -

• Extracting the useful information with accuracy and speed removing error prone manual intervention

• Presenting various report automatically like diagnosis report , Drug report etc

• Suggesting course of action with some automated suggestion module.

**References**

1. Savova G.K,; Masanz,J.J.; Ogren, V.P.; Zheng, J.; Sohn, S.; Kipper-Schuler, C.K.;Chute, G.C. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications; ; Journal of the American Medical Informatics Association (JAMIA), 2010; pp. 507-513.
2. Liddy, E.D. Natural Language Processing; Encyclopedia of Library and Information Science 2nd Edition,2001
3. Ronan Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu,K.; Kuksa, P. Natural Language Processing (Almost) from Scratch; Journal of Machine Learning Research 12, 2011
4. Wu, Y.; Denny, C.J; Rosenbloom, S.T.; Miller, R.A.; Giuse, D.A.;Dr.Ing; Xu, H. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries; Department of Biomedical Informatics, Department of Medicine, School of Medicine, Vanderbilt University, Nashville, TN

**Key-words:** unstructured texts, structured texts, nature language processing, e-health