# Hitachi India Developed a Technology to Extract Precisely Designated Information from Electronic Medical Records

*Hitachi team won 1st prize in a technical competition held by a European research forum for extraction of appropriate medical information from electronic medical records by achieving the precision of 86.8%*

**New Delhi, India, September 17, 2014** --- Hitachi India Pvt. Ltd. ("Hitachi India") and Hitachi, Ltd. (TSE: 6501, "Hitachi") today announced that they have developed a technology to extract precisely designated information from electronic medical records with International Institute of Information Technology, Hyderabad ("IIIT-H"). To apply the technology makes it possible to analyze electronic medical records including various mixed-forms such as free-form texts written by doctors or laboratory data written in given formats. It is expected that more advanced medical services will be realized by that, such as prediction of the incidence rates of disease in the future.

Hitachi India, Hitachi and IIIT-H won 1st prize with an extraction precision of 86.8% in CLEF[1] e-Health 2014 Task2[2], a technical competition on medical information extraction, which was held by a European research forum CLEF from Jan. to May 2014. The technology will be presented at CLEF Conference 2014 in UK from 15th to 18th Sep. 2014.

To predict the incidence rates of disease, it is necessary to analyze medical record texts such as doctors' remarks as well as the input information in the given format such as blood test results. Many research institutes have been developing the technology to extract necessary information because the medical record texts are written in different ways by doctors.

This competition has been held since 2013 aiming at enabling patients to easily read medical records containing many medical terms, and it is divided into 3 tasks based on the technical objectives. Ten research organizations including a team of Hitachi India, Hitachi and IIIT-H, participated in Task2, which competes in the precision to extract information such as body location, severity, etc. of disease, from electronic medical records. The team won 1st prize with the precision of 86.8% in the competition. Features of the technology developed are as follows.

1. Automatic learning technologies to construct classification rules automatically
In order to improve extraction precision, machine learning technologies were installed, which can construct rules to determine answers from preset learning data. Moreover,

the extraction precision is improved by utilizing the information on document structures such as the text information of chapter, section, etc. as well as main texts.

2. Post-processing modules employing context information
In order to increase the precision more, 2 post-processing modules are added. The first module converts the extracted information into more precise one by considering contexts around it. The next module determines the meanings of ambiguous phrases by using appropriate dictionaries to the type of input medical records. These modules improved the precision to extract body location information from 60% to 74%.

In the development of the technology, the team adopted Apache cTAKES[3], an open source program for medical text analysis, which improved efficiency of development, and enabled further expansion of features.

Hitachi India and Hitachi will continue to improve this extraction technology, and develop new technologies leading to the creation of advanced medical services.

**Notes**

(1) CLEF (Conference and Labs of the Evaluation Forum) Initiative is a European research body to promote research and development of multilingual and multimodal information access systems, and it supports technical competitions, such as e-Health, and annual conferences. In 2014, CLEF held 8 technical competitions such as news recommendation, image annotation, etc. including e-Health.    http://www.clef-initiative.eu/home

(2) CLEF e-Health 2014 is a technical competition aimed at the development of information access technology in electronic medical records, and is divided into 3 tasks. Task1 is on the technology to display medical records for patients to understand easily, and Task2 is to extract disease-related information from medical records, and Task3 is to retrieve relevant documents to queries on medical records.   http://clefehealth2014.dcu.ie/

(3) Apache cTAKES (clinical Text Analysis and Knowledge Extraction System) is an open source program to analyze English texts in medical domains.   http://ctakes.apache.org/

**About Hitachi in India**

Hitachi started its business in India in the 1930's. Currently, Hitachi has approximately 30 business bases and approximately 9,200 employees in India. In addition to being a leader in construction machinery and air-conditioning systems, the Hitachi Group in India is expanding on its 'Social Innovation Business', such as information & telecommunication systems, power systems, industrial, transportation and urban development systems. Together with further localization, Hitachi aims to contribute to a sustainable society in India as well as the country's economic growth.
For more information about the Hitachi Group in India, please visit the website at http://www.hitachi.co.in/.

**About Hitachi, Ltd.**

Hitachi, Ltd. (TSE: 6501), headquartered in Tokyo, Japan, delivers innovations that answer society's challenges with our talented team and proven experience in global markets. The company's consolidated revenues for fiscal 2013 (ended March 31, 2014) totaled 9,616 billion yen ($93.4 billion). Hitachi is focusing more than ever on the Social Innovation Business, which includes infrastructure systems, information & telecommunication systems, power systems, construction machinery, high functional materials & components, automotive systems, healthcare and others. For more information on Hitachi, please visit the company's website at http://www.hitachi.com.

# # #