Linguistic Information Processing Technology for Online Information Credibility Analysis

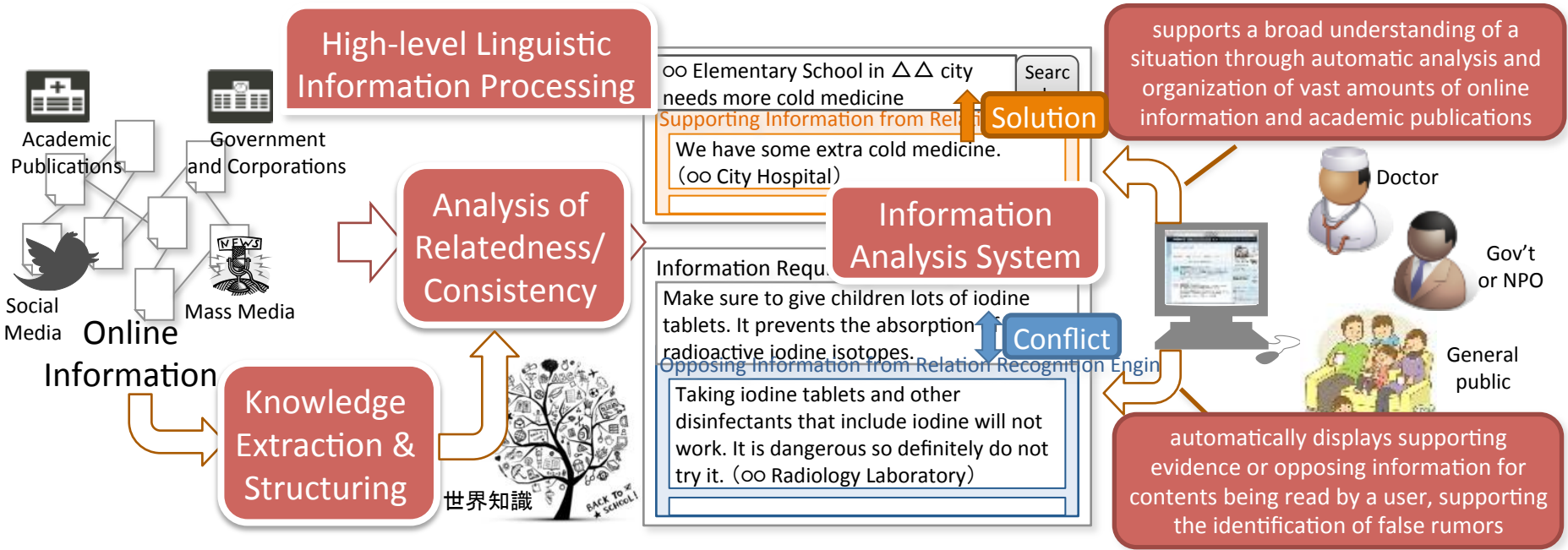
Kentaro INUI　Naoki OKAZAKI（Graduate School of Information Sciences）　TEL 022−795−7091
email inui@ecei.tohoku.ac.jp

Research Background

During the 2011 Tohoku Earthquake and its aftermath, gathering and distribution of information about the safety of disaster victims and the disaster status Online, including on Twitter and various support sites, greatly contributed to support efforts, such as reuniting people with their loved ones and supplying disaster areas. IT for facilitating online information sharing is extremely important to future disaster responses. However, in the Tohoku Earthquake, biased information and unfounded rumors were also spread on the Web, greatly contributing to people's confusion and unease. In order to effectively make use of online information and build a safe and comfortable society, technology for analysis of information credibility is essential.

Research Goals

- to construct an information environment that users can use safely and effectively, by automatically analyzing and collecting language content (unstructured information written in natural language) and then analyzing and classifying the underlying logical structure of circulating information
- Example: development of a health and medical information analysis system (see figure below)
  - to build an infrastructure that constantly gathers information from diverse sources including: DB of academic publications from the medical domain; the web sites of related institutions; social media including Facebook and Twitter
  - to automatically analyze the relatedness (e.g. *is a solution for a problem*) and consistency (e.g. *is evidence or conflicting information*) of the gathered information and present the results to the user in a bird's-eye-view visualization
  - by doing so, support the gathering and communication of health and medical information during a natural disaster, and the detection and verification of low-quality information such as false rumors



Kentaro INUI　Naoki OKAZAKI
（Graduate School of Information Sciences）

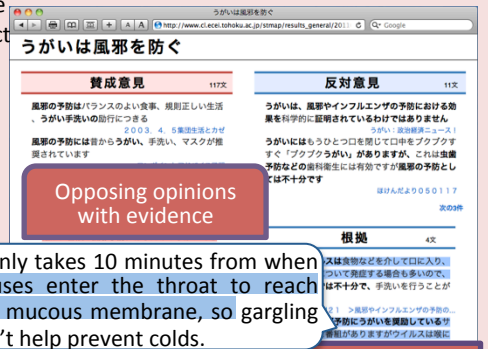
- **This research's core technology/features/patents (see below figure)**
  - Knowledge base construction: technology for automatically extracting target domain knowledge from a large-scale document collection, such as the Web, and constructing a large-scale knowledge base
  - Analysis of the logical structure in documents: technology for analyzing the logical structure of target documents by interpreting their contents with the knowledge base
  - Monitoring of consistency between documents: technology for automatically detecting contradictions and other inconsistent information between documents by comparing their logical structures
- **Potential influence and applications of this research**
  - For specialists (e.g. medical or governmental institutions, aid organizations, etc.): improved accessibility to disaster area information and academic publications, easier to obtain a broad understanding of health and medical-related situations
  - For the public (e.g. patients and their families, concerned citizens, etc.): improved accessibility to information distributed by specialists, easier detection of unfounded rumors and other poor quality information. For the health and medicine domain, citizens will be able to find treatments without worrying about unfounded rumors and folk medicine, addressing the lack of coordination between patients and physicians.

Automatic Analysis of Linguistic Information

High-level linguistic information processing automatically supports analysis of online information credibility

- semantic analysis technology detects the terminology and various expressions used on the Internet
- logical inference technology makes the most of vast knowledge bases to detect agreeing and conflicting opinions

Statement Map generation system



Does gargling really prevent colds?!

Opposing opinions with evidence

It only takes 10 minutes from when viruses enter the throat to reach the mucous membrane, so gargling can't help prevent colds.

Bird's-eye-view visualization including supporting and opposing evidence

Core Technology for Linguistic Analysis

Language processing that links the variety of linguistic expressions to the real world

- multiple expressions can refer to the same thing (diversity)
- a single expression can have multiple meanings (ambiguity)

渡辺陽太郎さん、二高へ避難との連絡あり。

person　place　action

identification of entity names and event information

standardization of different representations of person, place, and event names

生存（渡邉陽太郎, 仙台第二高校）

mapping to DB of real world entities

Accumulation and search of unified entity and event information

Nihon Keizai Shimbun "Tech Trend Survey" (Top 3) (6/2011)
The Sixteenth Annual Meeting of the Association for Natural Language Processing
Excellent Presentation Award, Excellent Young Researcher Award (3/2011)

The 72nd National Convention of the Information Processing Society of Japan
Excellent Research Award (3/2011)
CICLING 2010 Best Paper Award (First Place) (2/2010)