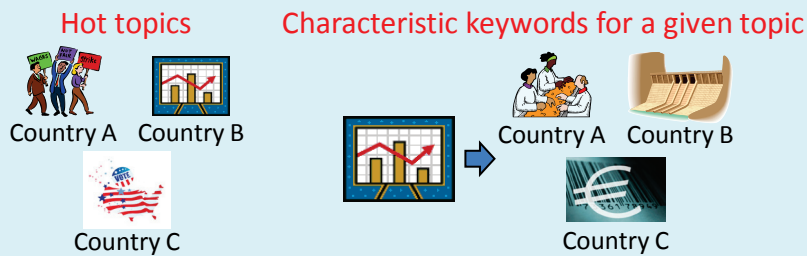


Multifaceted Analysis of News Articles by Using Semantic Annotated Information

Masaharu Yoshioka(Hokkaido University), Noriko Kando(National Institute of Informatics)

Background: NSContrast: World News Surveillances System

- News surveillance system that focuses on the difference of each news resource.
- Comparison of characteristic features (keywords) among different countries



Objectives

- Proposal of comparative multifaceted analysis interface for news articles from different countries
- Usage of semantic annotated text**
 - Person, Place, Organization, Sentiment, News source, ...
- Support interactive construction of multifaceted structured query for comparative analysis
- For example
 - Key **persons** related to Barack Obama about "Global Warming"
 - Comparison of **negative issues** related to Barack Obama in each **country**
 - Comparative **trend** analysis on **positive and negative articles** for "Economics" that contains different keywords (e.g., Euro, Japan, China, ...)

NSContrast

- News surveillance system that focusing on the difference of each news resource.
- Proposal of a framework to analyze multiple news resources for understanding their characteristics

Hot topics

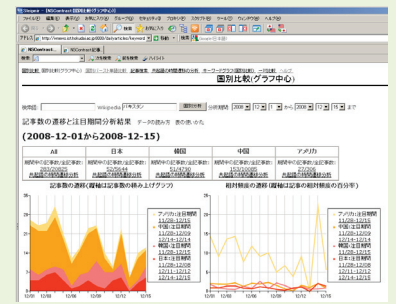
- Japan
 - Influenza
 - Demonstration of Egypt
 - Snow

Hot topics	Characteristic keywords for a given topic
Japan Influenza Demonstration of Egypt Snow	China Chinese New Year Inflation Evacuation from Egypt Korea Demonstration of Egypt Foot and Mouth Disease Derailment of KTX

- China
 - Chinese New Year
 - Inflation
 - Evacuation from Egypt

- Korea
 - Demonstration of Egypt
 - Foot and Mouth Disease
 - Derailment of KTX

Trend analysis



Semantic Annotation with Wikipedia Ontology

Wikipedia has a large number of entries about named entities
Usage of Category Information for estimating type of the entry

Classify named entities into several semantic categories by using Wikipedia category

Semantic category

Person: 日本の人物 (Japanese Person)、...

Place: 北海道の地理 (Geography of Hokkaido)、...

Organization: 国連機関 (Organization of United Nation)、...

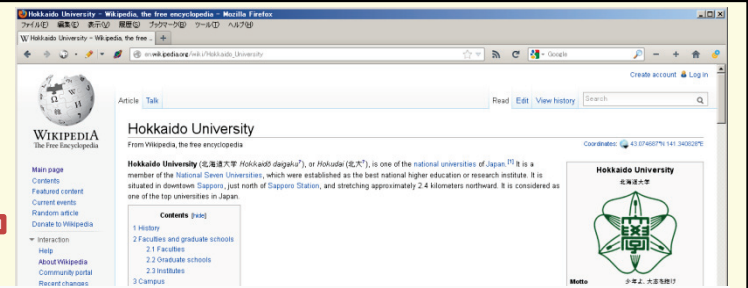
Varieties of description

e.g., Barack Obama and President Obama

Usage of Wikipedia redirect for identify synonym

Wikipedia category

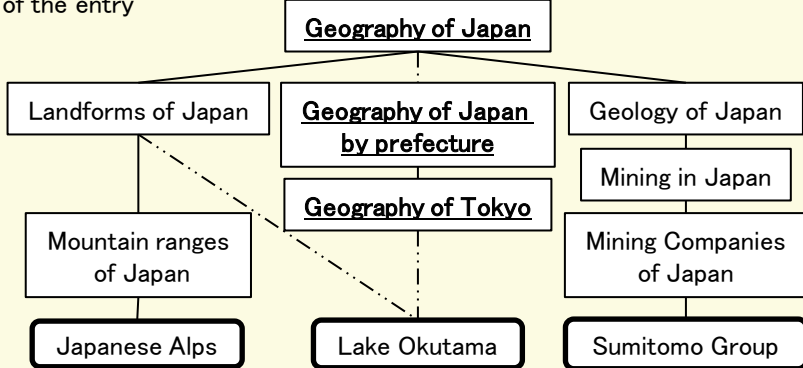
Categories: Educational institutions established in 1918 | Japanese national universities | National Seven Universities | Hokkaido University | Universities and colleges in Sapporo



Semantic Annotation with Geographic Information by using GeoNames and Wikipedia

Automatic Link Discovery between GeoNames and Wikipedia

Usage of Category Information for estimating location and type of the entry



Example of Link Discovery

'Narosura' populated place in Kenya
<http://en.wikipedia.org/wiki/Narosura>

Categories: Populated places in Rift Valley Province | Rift Valley Province, Kenya geography stubs

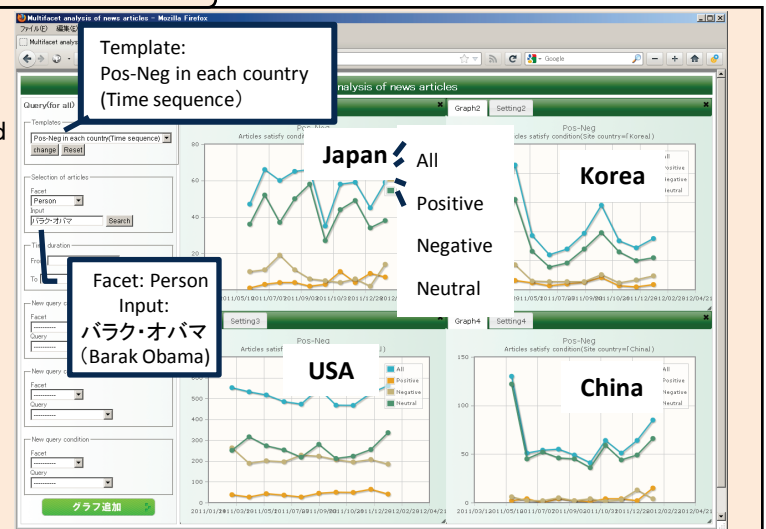
Name	Country	Feature class	Latitude	Longitude
1 Narosura	Kenya, Rift Valley	populated place	S 1° 32' 0"	E 35° 52' 0"
2 Narosura	Kenya, Rift Valley	hill	S 1° 33' 19"	E 35° 48' 0"
3 Koitiko	Kenya, Rift Valley	hill	S 1° 30' 0"	E 35° 49' 0"
4 Isanda	Kenya, Rift Valley	stream	N 0° 3' 0"	E 35° 44' 0"
5 Narosura	Kenya, Rift Valley	stream	S 1° 33' 41"	E 36° 1' 40"

Extract new link with higher precision and detect errors



Multifaceted Interface for Comparison

- Multiple windows** that represents different information based on different structured queries
- Structured query
 - Topic keywords** for selecting articles
 - Multiple windows for representing different characteristic information related to the topic keyword
 - Type of information
 - Representation style
 - Additional query keywords for selecting articles to compare
 - All windows are updated when topic keywords are modified
- Template for analysis
 - Support novice users to construct structured query for common analysis cases
 - Comparison among positive and negative articles trends in different country
 - Type of information: Sentiment (positive, negative and neutral)
 - Representation style: Time sequence graph for representing number of articles
 - Additional query keywords: News resource (Japan, Korea, USA and China)



Conclusion and future works

- Proposed a novel framework that enables multifaceted analysis of news articles.
- Evaluation is important, but it takes some time to use this interface in effective way