

Information Extraction based Approach for the NTCIR-9 1CLICK Task

Makoto P. Kato Meng Zhao Kosetsu Tsukuda Yoshiyuki Shoji
Takehiro Yamamoto Hiroaki Ohshima Katsumi Tanaka
Graduate School of Informatics, Kyoto University

Methods

Query Classifier

multi class SVM

query →

Feature	# of features
Has Wikipedia article	1
Frequency of parts-of-speech	44
Query unigram	85
Sentence pattern	2
Number of documents containing expanded query	15
Has travel services	1
Number of search results	1
Terms in search results	39
Total	185

CELEBLITY

Attribute-based



Attribute name	Value
Country	USA
Residence	Las Vegas, Nevada, U.S.

Sentence-based



RelativePosition
The position of a sentence in a section

LOCATION

query: Kyoto Royal Hotel

www.royalhotel.com
+
address OR access OR ...

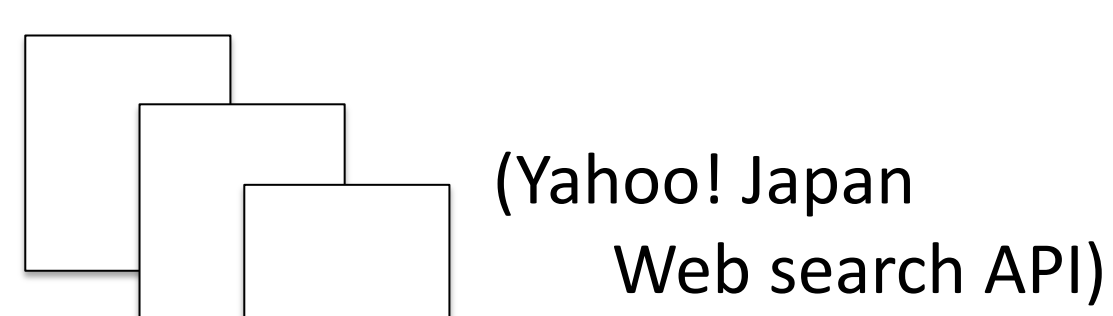


Table based extraction etc.

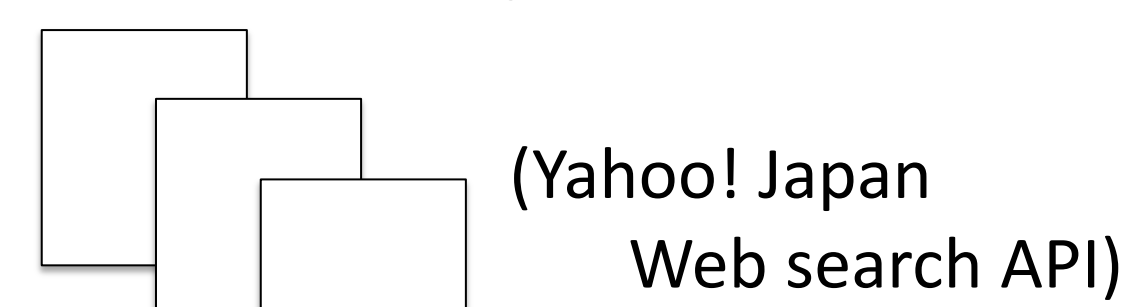
check in	15:00
check out	10:00

Check in : 15:00
Check out : 10:00

DEFINITION

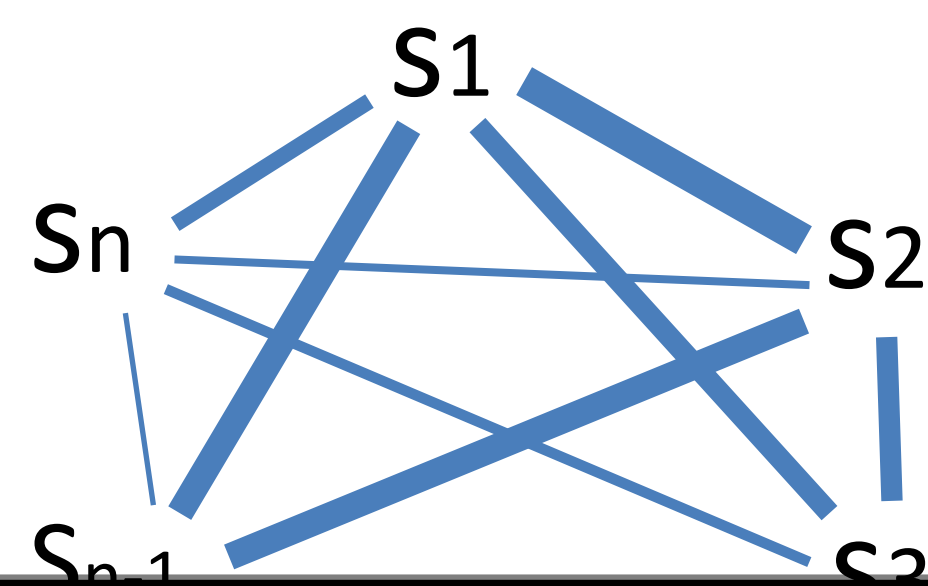
query: hospitality

"hospitality towa"



$S = \{s_1, s_2, s_3, \dots, s_n\}$
 s_i : hospitality towa ...

LexRank

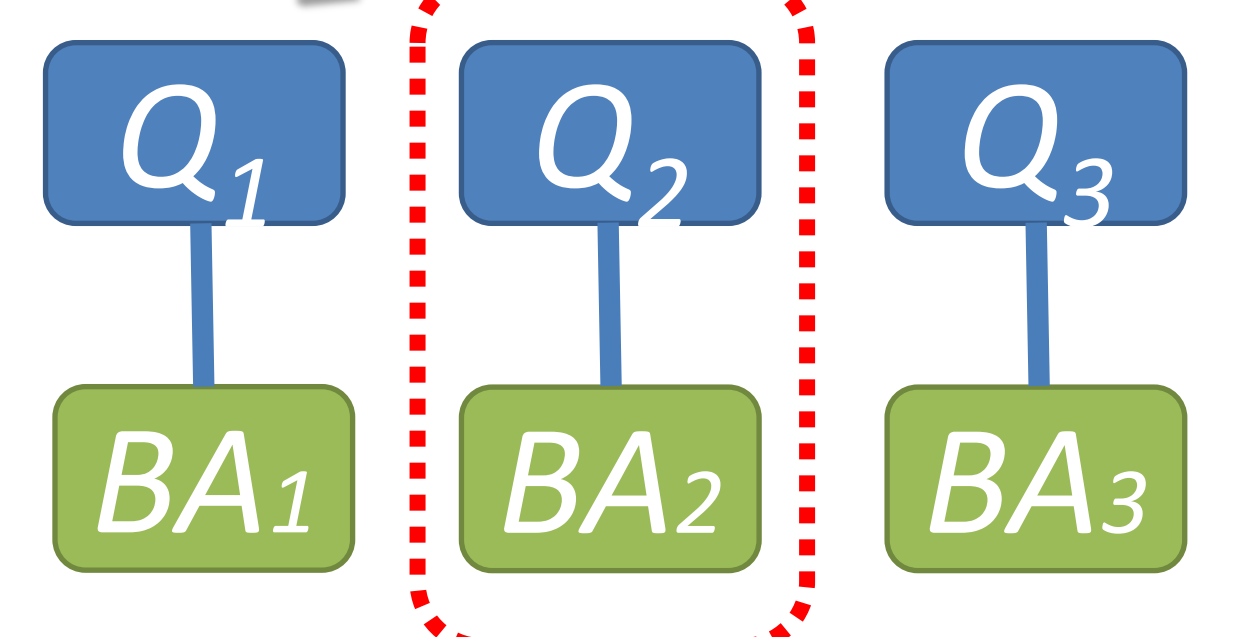


QA

query: Which is taller, Tsutenkaku or Utsunomiya tower?

Yahoo!
Chiebukuro

most similar Q



Information Summarizer

Because it is **too similar** to the already selected sentences

Sentence 1
Sentence 2
Sentence 3
...
Sentence N

MMR-based Information Summarization Method

Sentence 1
Sentence 3
...
Sentence K

Shorter than the limitation

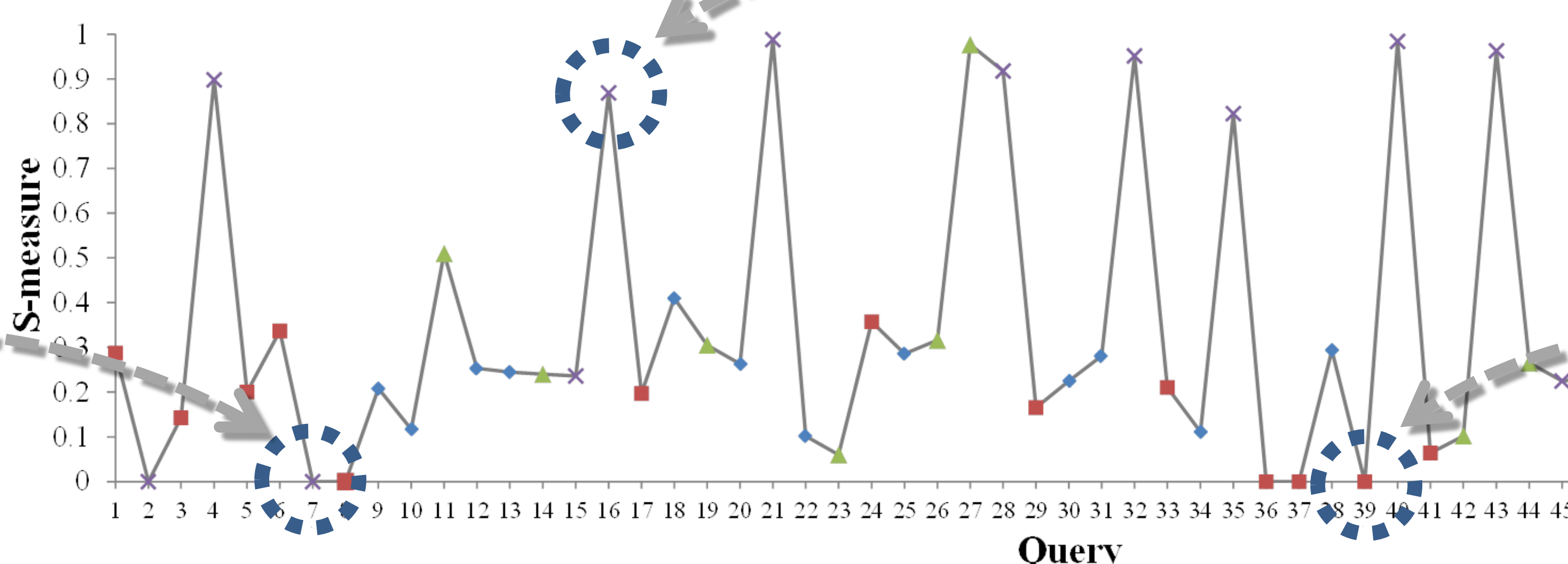
Summarized Text

Because of the output length **limitation**

Results

Which is taller, Tsutenkaku or Utsunomiya tower?

Where is Yumura-onsen?



Kanazawa University

- The **object identification** problem should be tackled (e.g. Kanazawa University)
- QA queries achieved high score **only when exactly the same questions** are available