Information Extraction based Approach for the NTCIR-10 1CLICK-2 Task

Tomohiro Manabe†, Kosetsu Tsukuda†, Kazutoshi Umemoto†, Yoshiyuki Shoji†, Makoto P. Kato,
Takehiro Yamamoto, Meng Zhao, Soungwoong Yoon, Hiroaki Ohshima, Katsumi Tanaka
Graduate School of Informatics, Kyoto University
†Research Fellow of Japan Society for the Promotion of Science

Framework

Query

Three IE Methods

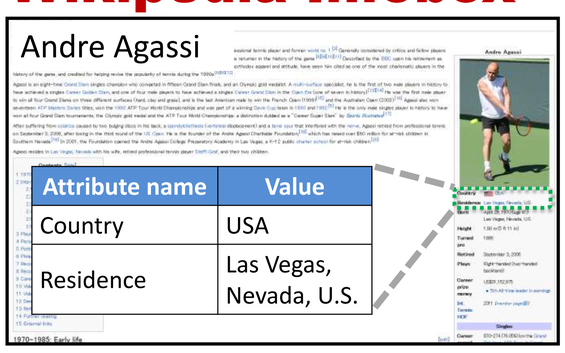
IE for Highly Structured Information

IE from Table

Check in	15:00
Check out	10:00

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

IE from Wikipedia Infobox



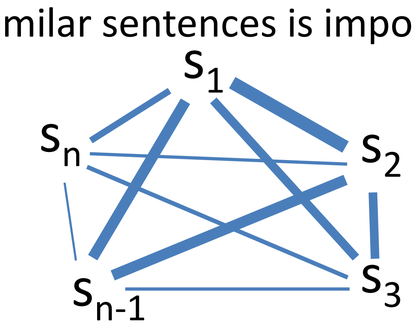
IE based on Regular Expression

Tel: *ddd-ddd-dddd*

Important Sentence Extraction

TextRank-based IE

- The sentence having many similar sentences is important



IE based on Access Information

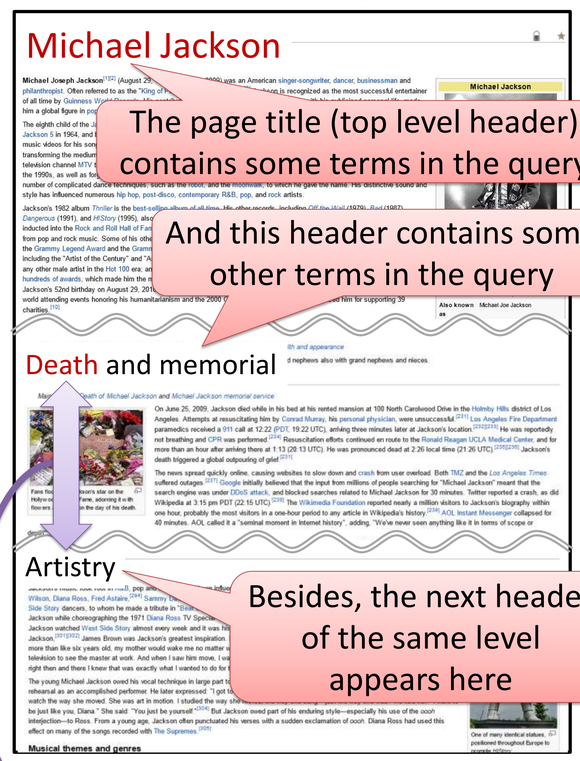
- Access information usually contains typical words

Station / By Car / 10 minutes

JR焼津駅よりお車で約10分
富士山静岡空港より約40分

Airport / 40 minutes

IE based on Hierarchical Headings



Query Classifier

Multi Class SVM

Features for Japanese Queries	# of Features	Only on Open Runs
Query length	2	
In dictionary	10	
Frequency of parts-of-speech	9	
Query Unigram	3	
Sentence pattern	2	
# of documents containing Expanded query	19	*
# of search results	1	*
Terms in search results	36	*
Sites in search results	44	*
Total	**126**	

Features for English Queries	# of Features	Only on Open Runs
In dictionary	11	
Frequency of parts-of-speech	12	
Sentence pattern	2	
Terms in search results	8	*
Sites in search results	9	*
Total	**42**	

Key-Value Pairs

Query Type

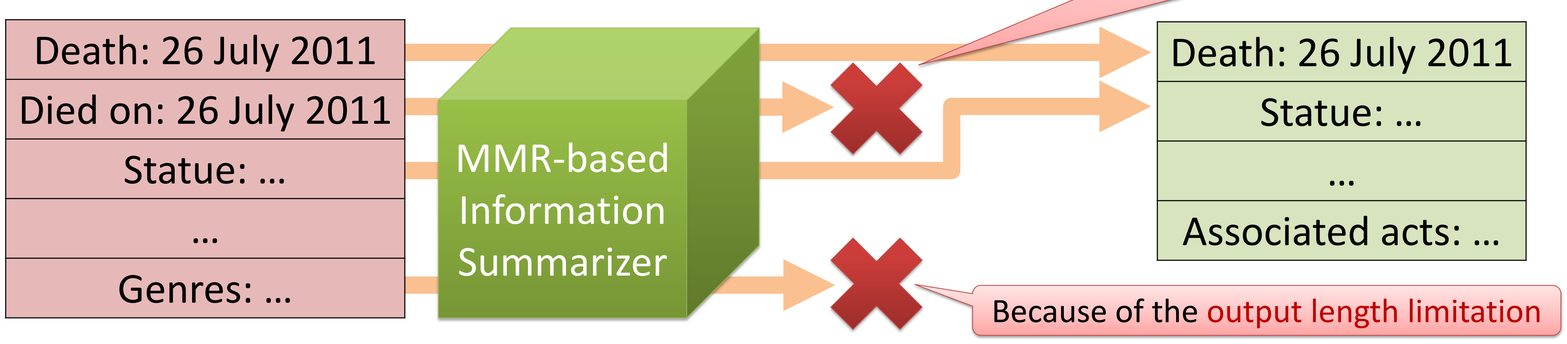
Weighting Function

$$\text{Score}'(u) = c_{t,e}\,\text{Score}(u),$$

where t is the query type and e is the extraction method

$c_{t,e}$	ARTIST	ACTOR	POLITICIAN	ATHLETE	FACILITY	GEO	DEFINITION	QA
IE for Highly Structured Information	2	2	1	2	**5**	**4**	2	1
Important Sentence Extraction	1	1	1	2	1	1	3	**4**
IE based on Hierarchical Headings	**4**	3	2	1	1	1	2	**4**

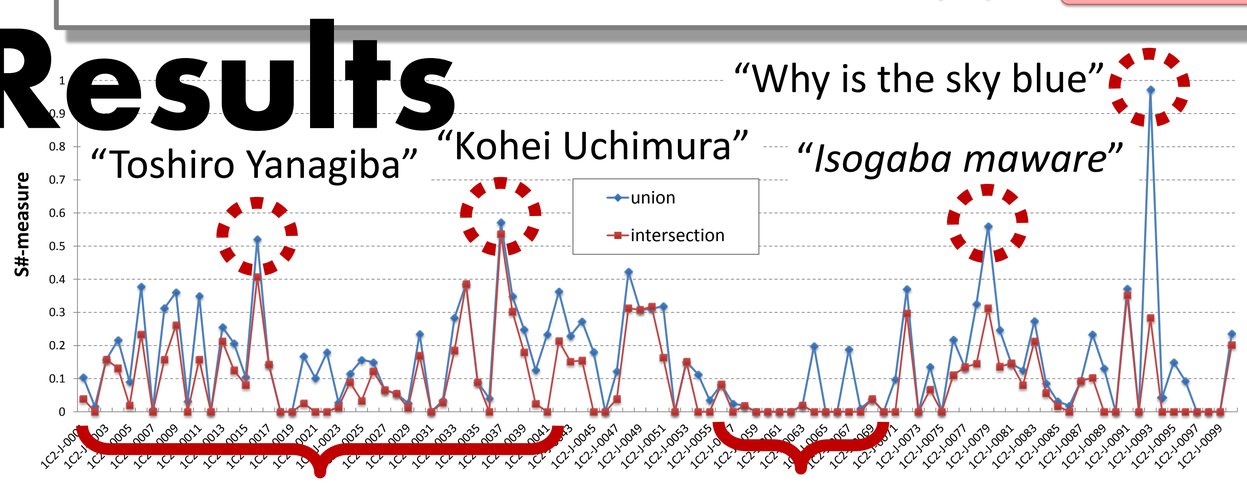
Information Summarizer



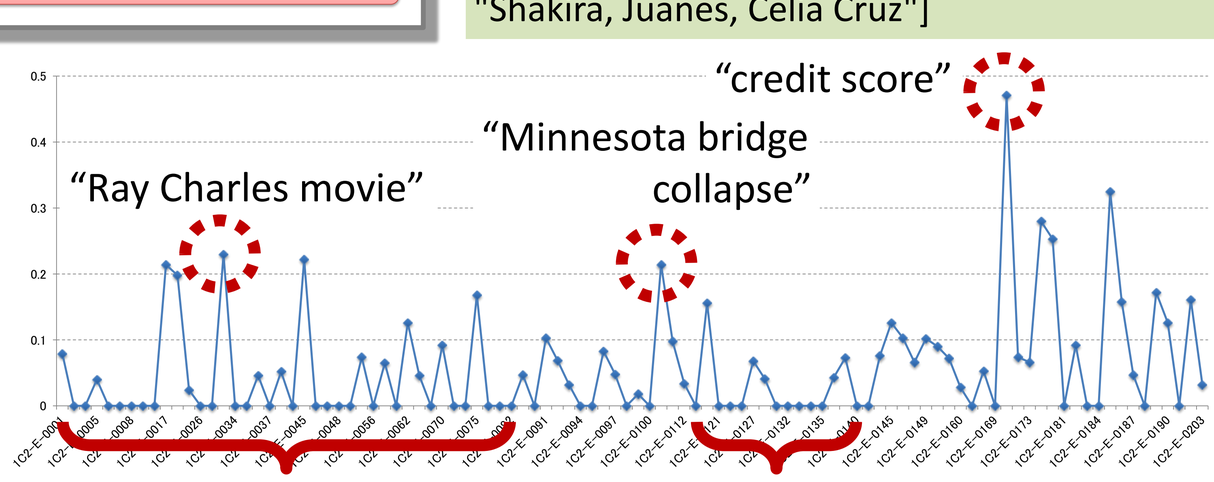
Summarized Text

["death": "26 July 2011"], ["statue": "Joe Arroyo's Statue on Barranquilla"], ["career": "Life and career"], ["music": "Joe Arroyo obituary"], ["Persondata": "Place of death"], ["Instruments": "Vocals, Wood block"], ["Born": "(1955-11-01)1 November 1955Cartagena de Indias, Bolivar, Colombia"], ["Associated acts": "Shakira, Juanes, Celia Cruz"]

Results



Per-query S#-measure of KUIDL-J-D-MAND-1



Per-query S#-measure of KUIDL-E-D-MAND-5

- The greater part of our methods seems effective for CELEBRITY queries.
- Our methods are ineffective for GEO queries which are for retrieving object sets.