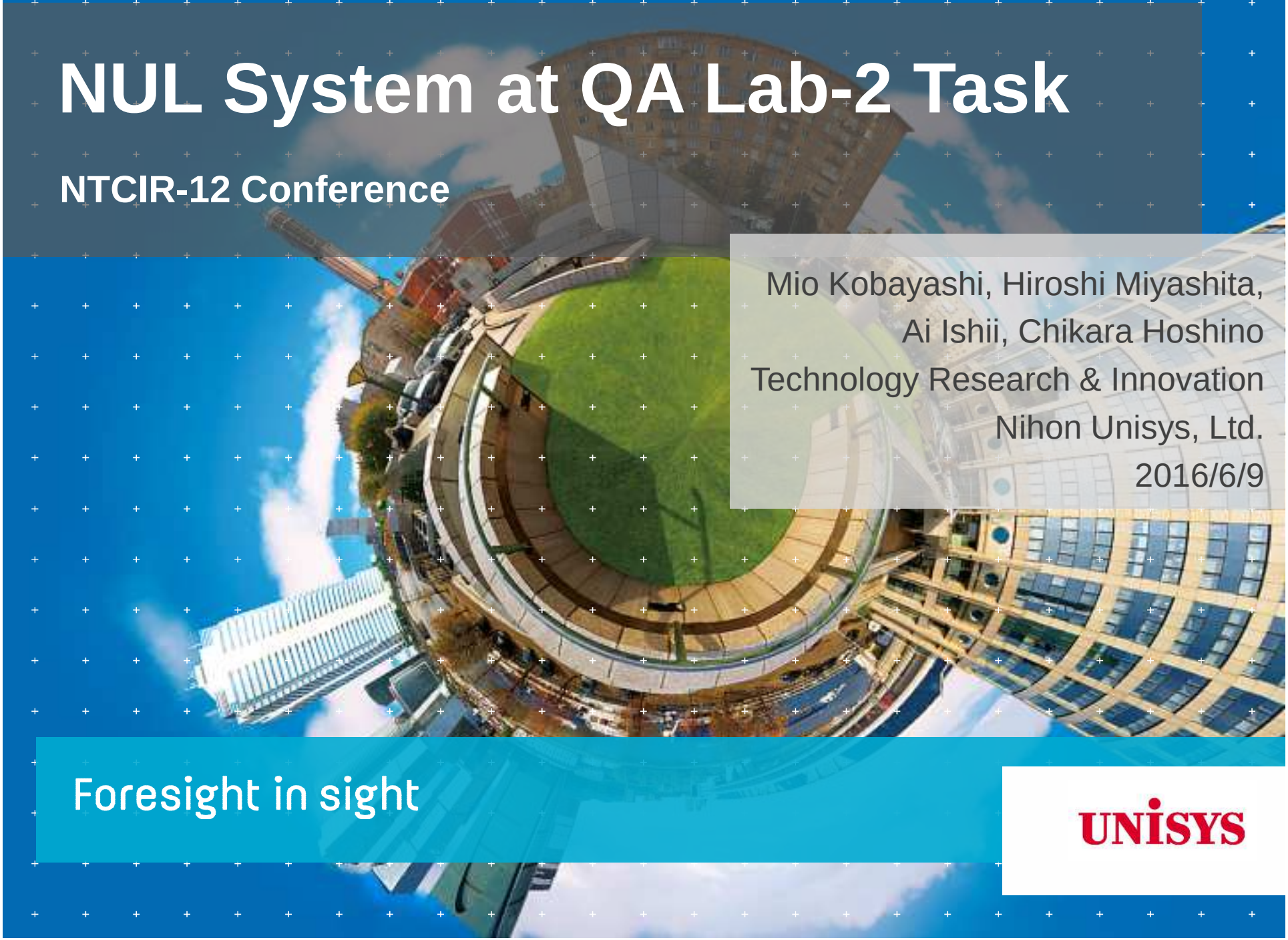
NUL System at QA Lab-2 Task

NTCIR-12 Conference

Mio Kobayashi, Hiroshi Miyashita,
Ai Ishii, Chikara Hoshino
Technology Research & Innovation
Nihon Unisys, Ltd.

2016/6/9



Foresight in sight

UNISYS

Agenda

Foresight in sight

1. Introduction
2. System Architecture
3. Experimental Results & Discussion
4. Future Efforts & Conclusion

1 Introduction

Our past challenge

- NTCIR-11 QA-Lab
  - Convert to Textual entailment recognition (RTE)
  - Shallow approach
  - Apache Solr

    →33 points (out of 100)

Previous studies

- Solve questions by keyword distributions
  - Kano (2014)
- Convert questions into factoid questions
  - Kanayama (2013), Okita (2014)
- Convert questions into textual inference
  - Tian (2014)

Observation of the task

- True-or-False : 70% of questions
  - Not necessary to read knowledge resources widely
    - 99.3% of questions have a paragraph which contain all of named entities in correct choice
  - Time descriptions are more abstract
  - Location descriptions are more abstract

2 System Architecture

Knowledge Resources

- 4 sets of high school textbooks
- Wikipedia
- World history ontology (Kawazoe 2014)
- Web sites of world history

Dictionaries

- Named entity dictionary
  - Class of words
  - Approximately 45,000 entries
- Synonym dictionary
  - Wikipedia Redirect
  - Wikipedia Hyponymy extraction tool
- Hypernym-hyponym dictionary
  - Wikipedia Hyponymy extraction tool
- Antonym dictionary
  - World history ontology (Kawazoe 2014)
- Suffix dictionary
- Year conversion dictionary
  - Years of nations and historical events

Common Modules

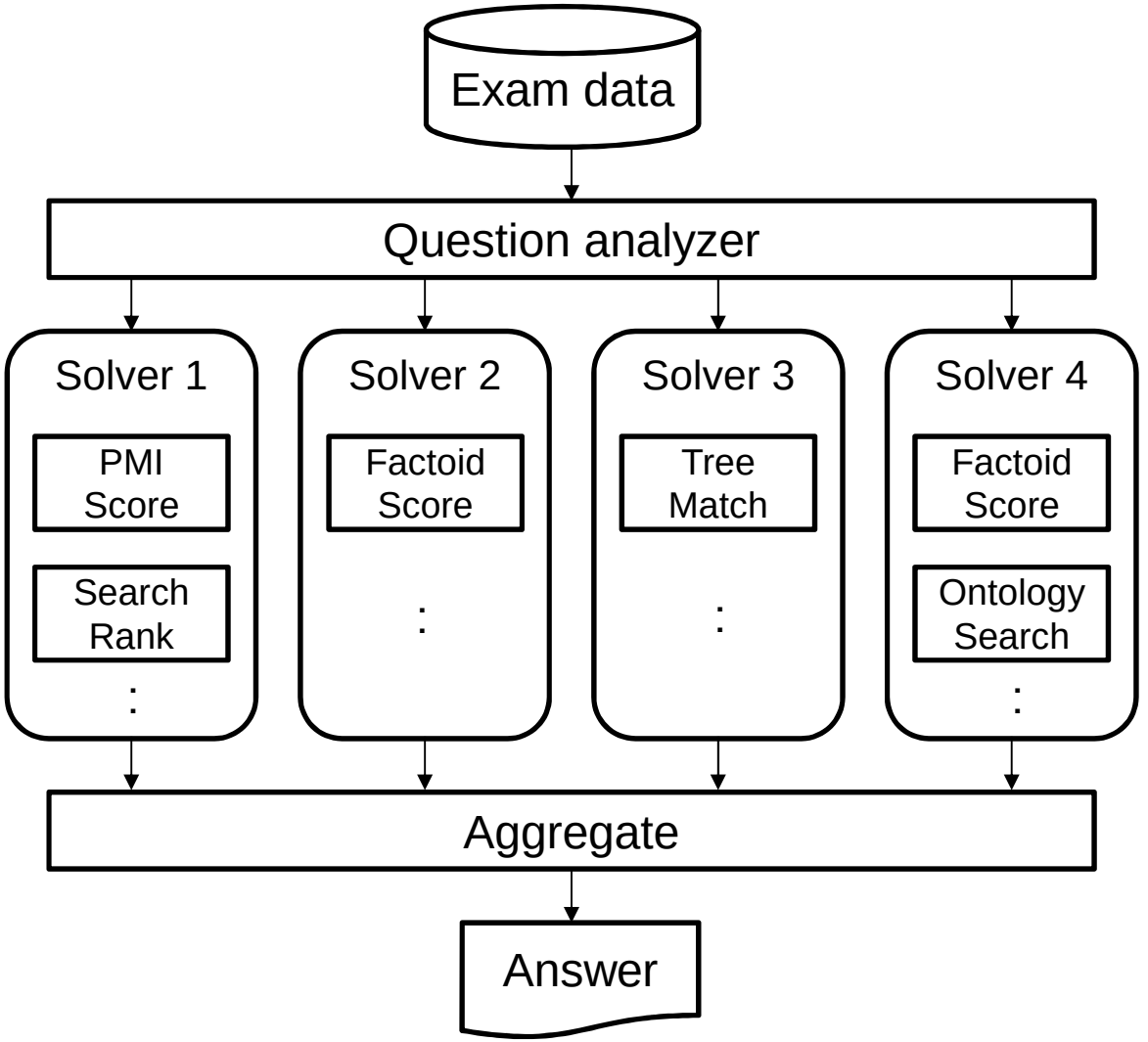
- Time expressions
  - “Charlemagne defeated the Magyar at the 8th century.”
  - 794 ⊂ 8th century
- Matching of words
  - Synonym, hypernym
  - Ignore suffixes
  - Exclusive relation (have same class & not matched)
- Extracting named entities from questions
  - “Charlemagne defeated the Magyar at the 8th century.”

Search engine settings

- Apache Solr
  - Index: paragraph / sentence
  - Weighting: Okapi-BM25
  - Morphological dictionary: UniDic (Den, 2008)

System Architecture

Foresight in sight



Relation of Type of Questions and Solvers

Relation of Type of Questions and Solvers

Type of questions	So.1	So.2	So.3	So.4
Relative True-or-False Questions	✓	✓	✓	✓
Relative True-or-False in Focus Word Questions	✓	✓	✓	✓
Absolute True-or-False Questions	✓	✓	✓	
Factoid Questions	✓	✓		✓
Slot-Filling Questions	✓	✓		✓
Time Reordering Questions				✓
What-Time Questions				✓
Unique Image, Mixed and Other Questions	✓			

Aggregate – weighted Borda count

Foresight in sight

Question: What is the name of banknote of the Yuan dynasty?
(元が発行した紙幣の名称として最も適当なものを選べ)
[1] Spade money (布銭) [2] Huizi (会子) [3] Jiaozi (公子) [4] Chao (交鈔)

Solver 1 (weight:0.6)	1st: [4], 2nd: [1], 3rd: [2], 4th: [3]	[4]2.4pt [1]1.8pt [2]1.2pt [3]0.6pt
Solver 2 (weight:0.5)	1st: [4], 2nd: [2], 3rd: [1], 4th: [3]	[4]2.0pt [2]1.5pt [1]1.0pt [3]0.5pt
Solver 3 (weight:0.4)	1st: [3], 2nd: [2], 3rd: [1], 4th: [4]	[3]1.6pt [2]1.2pt [1]0.8pt [4]0.4pt
Solver 4 (weight:0.3)	1st: [1], 2nd: [3], 3rd: [4], 4th: [2]	[1]1.2pt [3]0.9pt [4]0.6pt [2]0.3pt

[4] 5.4pt
[1] 4.8pt
[2] 4.2pt
[3] 3.6pt

“Answer is [4]”

Solver 1 – Score of PMI

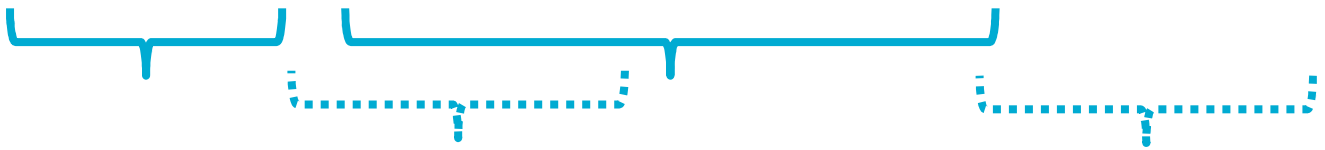
- Pointwise Mutual Information (PMI) values
  - S = pairs; NE and next NE or CW(which has antonym)
  - $PMIScore = \frac{1}{|S|}\sum_{(w_1,w_2)\in S} log \frac{P(w_1,w_2)}{P(w_1)P(w_2)}$

- In correct choice, each word has relationship each other

17**世紀後半**、**ロシア**では **ピョートル1世**が **即位**し、**西欧化政策**を **開始**した。

Late 17 century Russia Peter the Great enthronement Westernized policy start

time$_1$ *ne*$_1$ *ne*$_2$ *cw*$_1$ *ne*$_3$ *cw*$_2$



Solver 1 – Score of Search Rank

Foresight in sight

Search Rank

- Calculate score about combinations of 3 or more words
- $RankScore = \frac{1}{|Q|}\sum_{(ne_i,q_i)\in Q} -Rank(ne_i, q_i)$
- $Rank(ne_i, q_i) = \begin{cases} rank_{ne_i,q_i} & (rank_{ne_i,q_i} < k) \\ 2 * k & (otherwise) \end{cases}$

17世紀後半、[　]ではピョートル1世が即位し、西欧化政策を開始した。
Late 17 century [　] Peter the Great enthronement Westernized policy start

- Omit 1 NE and search the sentence
- Check the rank which contain omitted NE

Solver 1 – Score of Time expression

Foresight in sight

- Time expression
  - $TimeScore = 10.0 - 20.0 * \frac{rank(W_{time})}{length(list_{time})}$

[　　]、ロシアでは ピョートル1世が 即位し、西欧化政策を 開始した。
[　　] Russia Peter the Great enthronement Westernized policy start

- Omit a time expression and search the sentence
- Check the search result and add some rules such as:
  - Time Expression is not found: -10.0
  - Time Expression and all NE are found in result: +10.0
  - All NE are found but it contains other Time Expression: -100.0

Solver 1

Foresight in sight

- Calculate total score of a choice

17世紀後半、 イギリスの エリザベス1世が 統一法を 発布した。

Late 17 century United Kingdom Elizabeth I Act of Uniformity issue

$time_1$ ne_1 ne_2 ne_3 cw_1

					Average
PMIScore		4.01	1.56	--	2.79
RankScore	1.0	1.0	1.0		-1.00
TimeScore	-10.0				-10.00
				total score:	**-8.21**

Solver 2

Foresight in sight

- Convert to virtual factoid question

Charlemagne defeated the Magyar at the 8th century.
(8**世紀にカール大帝はマジャール人を撃退した。**)

Charlemagne defeated the Magyar at the (Time).
((**時代**)**にカール大帝はマジャール人を撃退した。**)

(Person) defeated the Magyar at the 8th century.
(8**世紀に**(**人物**)**はマジャール人を撃退した。**)

Charlemagne defeated the (Nationality) at the 8th century.
(8**世紀にカール大帝は**(**民族**)**を撃退した。**)

Solver 2 - Answering factoid questions

Foresight in sight

- $QAScore(Q, d_k) = \sum_i \alpha_{q_i,d_k} exp\{-\gamma l(q_i, d_k)^\beta\}$
- Q : virtual factoid question
  - q_i : ith word of Q
  - d_k : kth word of searched resource
- $l(q_i, d_k)$: the nearest neighbor distance from d_j to d_k
  - d_j : in synonym of q_i, nearest word from d_k

Q: Charlemagne defeated [personType].

- Score about named entity “Avars”:

$$QAScore(Q, Avars)$$

$$= \alpha_{Charlemagne,Avars} exp\{-\gamma l(Charlemagne, Avars)^{\beta}\}$$

- There are 22 steps between “Avars” and “Charlemagne” which is nearest from “Avars”

$$l(Charlemagne, Avars) = 22$$

Textbook

…fighting the Avars, …(21 words)… Charles the Great was crowned by Pope Leo III…

Solver 2 – cost of factoid questions

- Calculate costs of each choice

Rank	Word	Score
1	Avar	3.2
2	Mongolian	2.6
…		
5	Magyar	1.1

Cost: 3.2-1.1=2.1

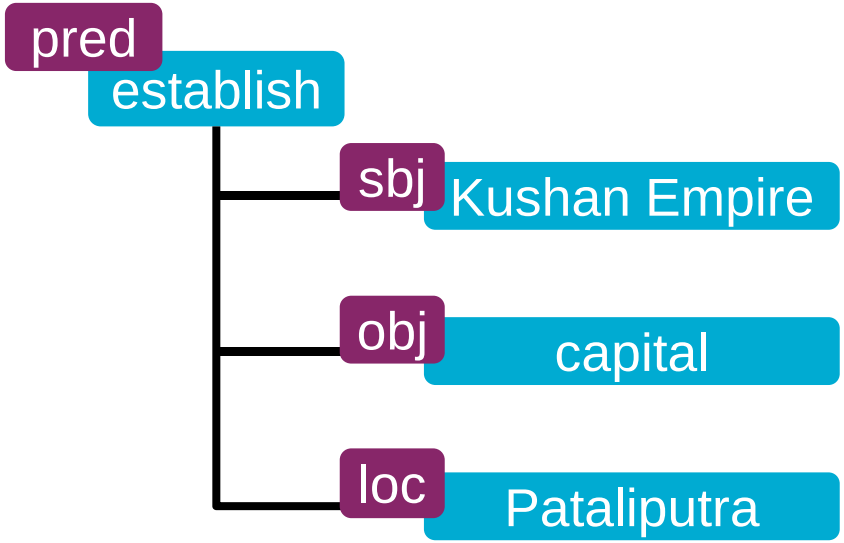
- Cost of a choice = A mean of the each cost of virtual question
- Answer = Choice which has lowest cost

Solver 3 – Syntax tree matching

Foresight in sight

- Convert to syntax tree

[1] Kushan Empire established capital in Pataliputra.
(**クシャーナ朝はパータリプトラに都をおいた。**)



hypothesis

An abstract expression for a choice is similar to an abstract expression in knowledge resources
→the choice is correct.

Solver 3 – Syntax tree matching

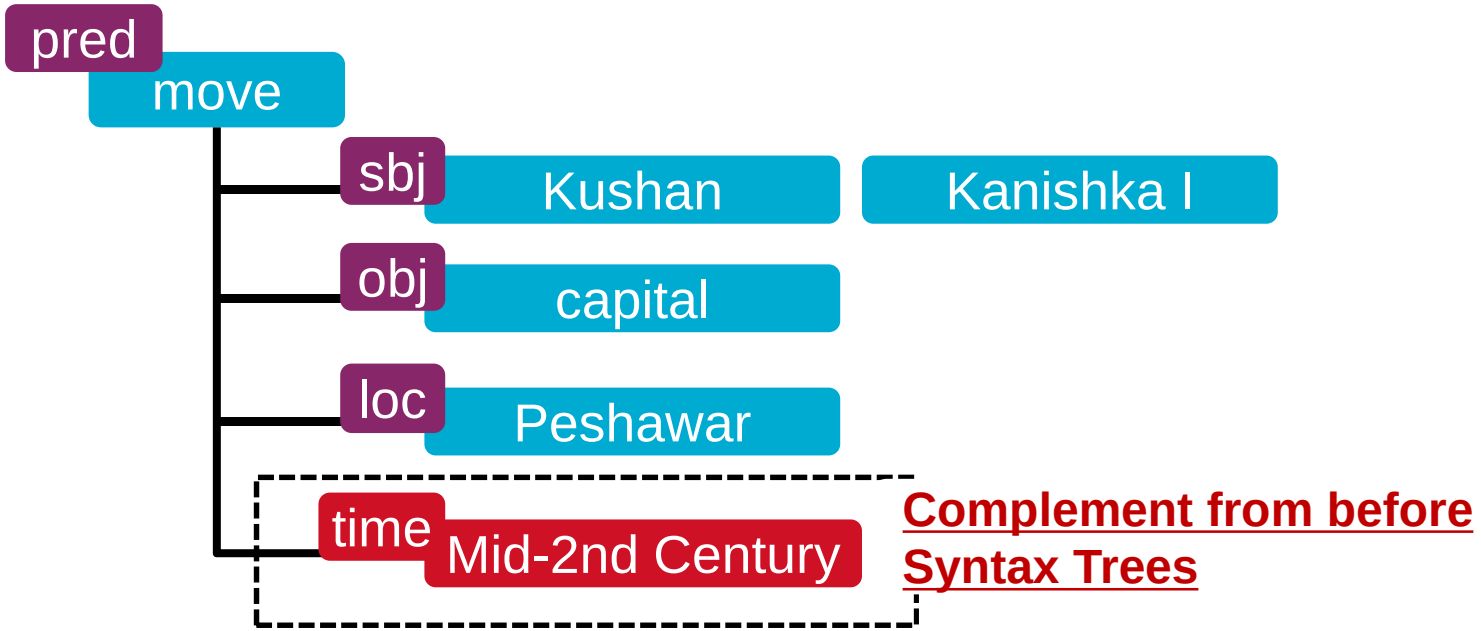
Foresight in sight

- Search for keywords in knowledge resources

Kushan Empire | Pataliputra | Ancient India | capital

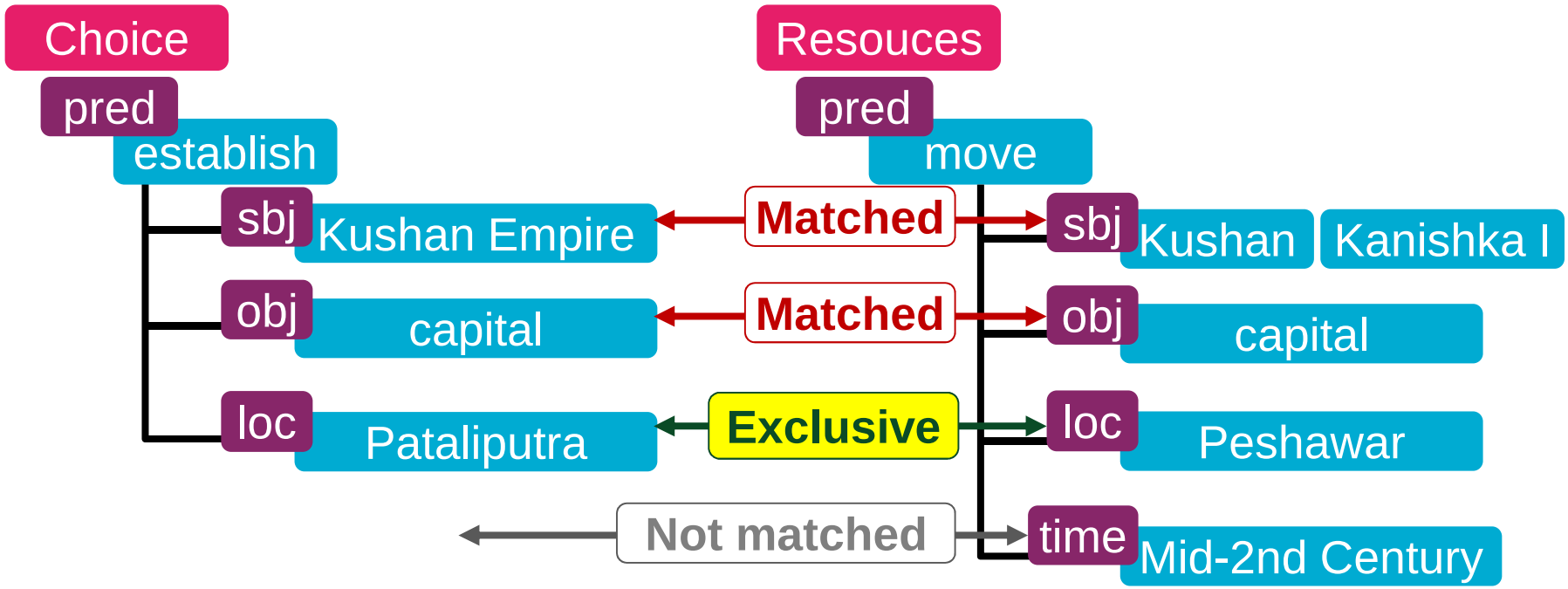
Searched sentence

The Kushan King Kanishka I moved capital to Peshawar.
(クシャーナ朝の王、カニシカ1世は都をペシャワールに移した。)



Solver 3 – Syntax tree matching

- Matching of Syntax Tree



Same class & not synonym/hypernym
→ Pataliputra ∩ Peshawar = Φ

Solver 4

Foresight in sight

- Convert to virtual factoid question (similar to solver 2)

Charlemagne defeated the Magyar at the 8th century.
(8**世紀にカール大帝はマジャール人を撃退した。**)

Charlemagne defeated the Magyar at the (Time).
((**時代**)**にカール大帝はマジャール人を撃退した。**)

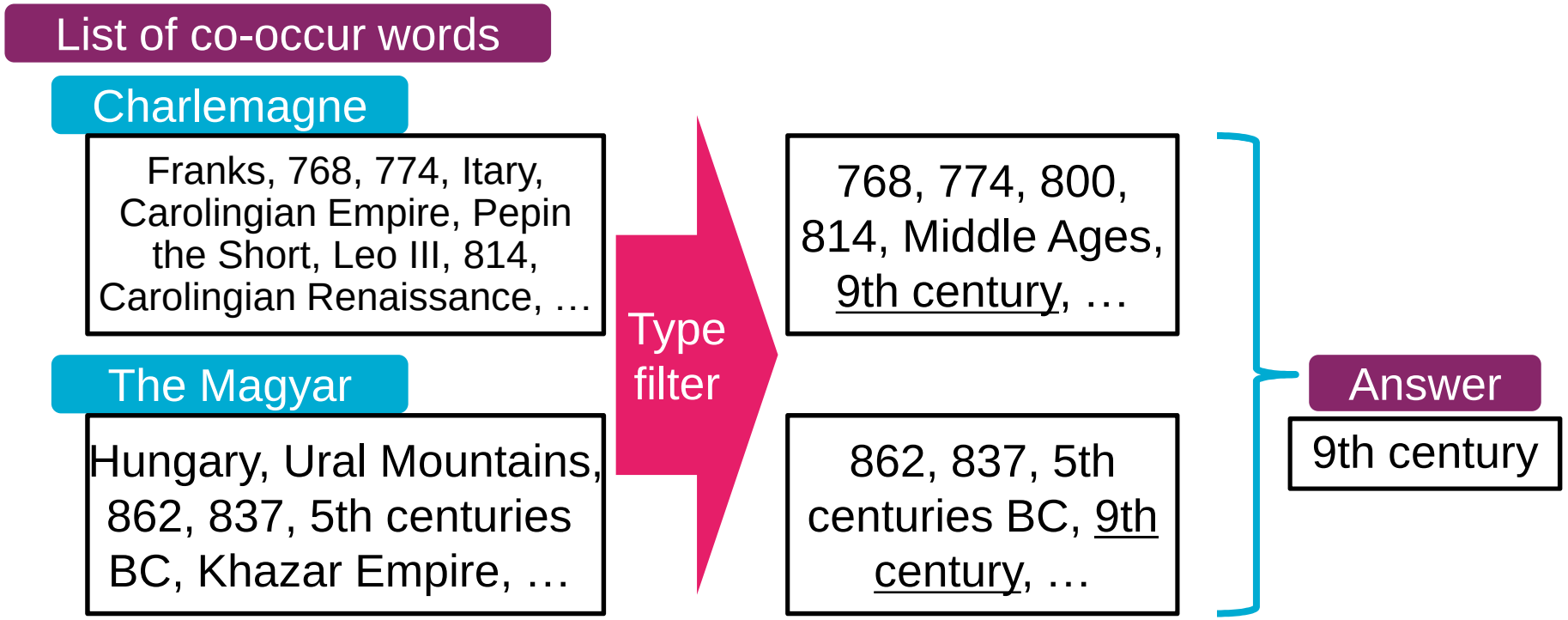
(Person) defeated the Magyar at the 8th century.
(8**世紀に**(**人物**)**はマジャール人を撃退した。**)

Charlemagne defeated the (Nationality) at the 8th century.
(8**世紀にカール大帝は**(**民族**)**を撃退した。**)

Solver 4 – co-occurrence score

- Solve factoid from co-occurrence

Charlemagne defeated the Magyar at the (Time).
((時代)にカール大帝はマジャール人を撃退した。)



3 Experimental Results & Discussion

Experimental Results

Foresight in sight

- NUL scores of formal run

Ph.	Exams	Run #1	Run #2	Run #3
1	National Center Test (1999)	43	**49**	36
2	Benesse mock exam (2015 Jun/All/out of 175)	**121**	121	118
(2)	Benesse mock exam (2015 Jun/Pattern 1)	76	76	76
(2)	Benesse mock exam (2015 Jun/Pattern 2)	64	64	61
3	National Center Test (2011)	65	65	**68**
3	Benesse mock exam (2014 Sep/All/out of 125)	**77**	76	76
(3)	Benesse mock exam (2014 Sep/Pattern 1)	60	57	60
(3)	Benesse mock exam (2014 Sep/Pattern 2)	58	60	54

Torobo Result

Experimental Results

- Scores by each solvers of phase 3

Ph.	Exams	So.1	So.2	So.3	So.4	Combined
1	National Center Test (1999)	46	56	30	40	52 / 100
2	Benesse mock exam (2015 Jun)	121	104	82	56	125 / 175
3	National Center Test (2011)	62	62	43	39	65 / 100
3	Benesse mock exam (2014 Sep)	58	65	36	33	76 / 125
-	**Total**	**287**	**287**	**191**	**168**	**318 / 500**

Experimental Results

- Numbers of correct (each solvers / type of questions)

Type of questions	So.1	So.2	So.3	So.4	Combined
Relative True-or-False Questions	53	54	48	32	56 / 75
Relative True-or-False in Focus Word Questions	3	4	2	0	2 / 4
Absolute True-or-False Questions	9	6	7	-	9 / 17
Factoid Questions	8	7	-	4	6 / 10
Slot-Filling Questions	7	10	-	4	9 / 13
Time Reordering Questions	-	-	-	6	6 / 9
What-Time Questions	-	-	-	0	0 / 3
Unique Image, Mixed and Other Questions	7	-	-	-	7 / 13

By National Center Test(2011) and Benesse Mock Exam(2014 Sep, 2015 Jun)

Error Analysis

- Patterns of incorrect
  - Misjudgment of words dependency (solver 1)
  - Confused by many-to-many relationship (solver 2 and 4)
  - Error because of homonyms (solver 3)
  - Failed to anaphora resolution (solver 3)
  - Cannot match Syntax Trees by rephrasing (solver 3)
  - Too frequent named entities (All solvers)
  - Failed to extract necessary named entity from instruction (All solvers)

4 Conclusion & Future Works

Conclusion & Future Works

Foresight in sight

- Conclusion
  - Observe the task
    - Make 4 strategies from observation
  - Combination strategy
    - To reduce variance in generalization error

  →Get 1st place(JA)
- Future Works
  - Strict analysis of sentence
  - Determination of important words
  - Optimal strategy of solver combination
  - Question analysis