

Using Time Periods Comparison for Eliminating Chronological Discrepancies between Question and Answer Candidates at QALab NTCIR11 Task

Yasutomo Kimura
Otaru University of Commerce
3-5-21, Midori, Otaru,
Hokkaido, Japan
kimura@res.otaru-
uc.ac.jp

Keiichi Takamaru
Utsunomiya Kyowa University
1-3-18, Odori, Utsunomiya,
Tochigi, Japan
takamaru@kyowa-u.ac.jp

Hideyuki Shibuki
Yokohama National University
79-7 Tokiwadai, Hodogaya-ku,
Yokohama, Japan
shib@forest.eis.ynu.ac.jp

Fumito Masui
Kitami Institute of Technology
165, Koencho, Kitami, Japan
f-masui@mail.kitami-
it.ac.jp

Fumitoshi Ashihara
Hokkaido University
Kita-ku Kita 14 Nishi 9,
Sapporo, Japan
ashihara@media.eng.
hokudai.ac.jp

Yuzu Uchida
Hokkai-Gakuen University
Chuo-ku Minami 27 Nishi 11,
Sapporo, Japan
yuzu@eli.hokkai-s-
u.ac.jp

Michal Ptaszynski
Kitami Institute of Technology
165, Koencho, Kitami,
Japan
ptaszynski@cs.kitami-
it.ac.jp

Kenji Araki
Hokkaido University
Kita-ku Kita 14 Nishi 9,
Sapporo, Japan
araki@ist.hokudai.ac.jp

Arnaud Jordan
Hokkaido University
Kita-ku Kita 14 Nishi 9,
Sapporo, Japan
arnaud@media.eng.
hokudai.ac.jp

Hokuto Ototake
Fukuoka University
8-19-1 Nanakuma, Jonan-ku,
Fukuoka, Japan
ototake@fukuoka-u.ac.jp

Rafal Rzepka
Hokkaido University
Kita-ku Kita 14 Nishi 9,
Sapporo, Japan
rzepka@ist.
hokudai.ac.jp

ABSTRACT

This paper reports on our approach to the NTCIR-11 QALab task of answering questions from Japanese National Center Examinations for University Admissions. Our approach aims at identifying and comparing periods of world history in both the questions and the answer candidates. We created and applied a date identification method, which checks for temporal overlaps between time periods in questions and their answer candidates. In this paper we introduce details of this method and analyze the test results. When tested on the World History Dictionary that is used for preparing to the exams, our approach achieved 30% of correct answers in the 2007 Center Exam Task and 17% for the 2003 Center Exam Task.

Team Name

NNLP

Subtasks

Phase 1 Center Exam (Japanese)
Phase 2 Center Exam (Japanese)

Keywords

Temporal overlaps identification

1. OVERVIEW OF THE SYSTEM

To solve the Center Exams for world history in Japanese, our system firstly extracts time related phrases from questions and their answer candidates. Then the terms are compared from various viewpoints to the compared one. The candidate whose terms are the most consistent with each other is selected as most probable answer. There are a numerous viewpoints to find correctness of answer candidate, *e.g.* date, place, country, religion, etc. The NNLP team considered that dates of the terms are most important clue to solve an exam for history. Although there is some previous research on dates, these method are difficult to apply to characteristic expressions of dates in world history[1][2].

Therefore our system focuses on dates within the terms from world history. The system checks up temporal overlaps between periods of terms in questions and their answer candidates. Incorrect candidates are likely to contain terms that do not overlap with other terms.

The dates related to most of the terms usually span over a period of several years, although some terms are events within a year. Therefore, we needed to specify the Time

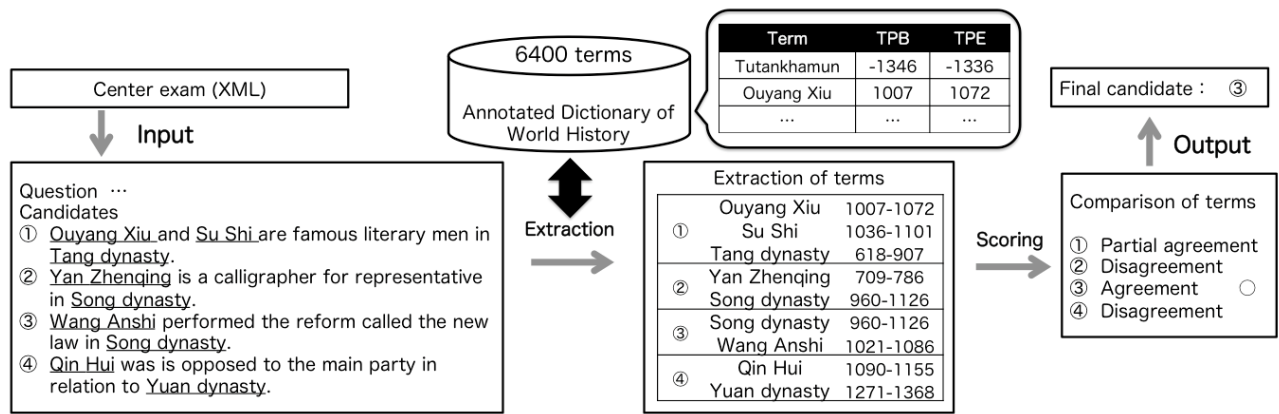


Figure 1: An example of processing using date identification method

Period Beginning point (TPB) and the Time Period Ending point (TPE) of all applicable terms. For example, the term “Ostrogothic Kingdom” represents a kingdom established by the Ostrogoths in the lands of present Italy and was established in 493, and collapsed in 553. Therefore the TPB for the term “Ostrogothic Kingdom” is 493 and the TPE is 553. In the case of person’s name, the date of the birth and the date of the death are used for TPB and TPE, respectively.

The NLP team created a database of World History terms which contains both TPB and TPE. Since a dictionary of world history clearly describes terms and their periods, our World History term database is created based on this dictionary. It was more efficient than using a school textbook, due to the fact that in a dictionary each definition is provided in a specified format, whereas working with a textbook would require applying additional knowledge acquisition techniques. A dictionary of world history does not only include items such as “term (entry word)” and their explanation. It also includes some additional information, such as “alias”, “reading of term (pronunciation)”, “area”, “date”, etc. Our method eliminates incorrect candidates by finding chronological discrepancies between terms in question and answer candidates.

Figure 1 shows an example of the way to solve a question using an identification of dates in the World History Exam B from year 2009.

The underlined sentence of the question is as follows:

“After Tang dynasty and Song dynasty, brilliant people which had cultural and artistic talents had passed the imperial examination.”

The four answer candidates are follows:

1. Ouyang Xiu and Su Shi are famous literary men in Tang dynasty.
2. Yan Zhenqing is a representative calligrapher in Song dynasty.
3. Wang Anshi performed the reform called the New Law in Song dynasty.
4. Qin Hui was opposed to the main party in relation to Yuan dynasty.

Each answer candidate contains the following terms (TPB and TPE are in parenthesis)

Candidate 1

- Ouyang Xiu (1007-1072)
- Su Shi (1036-1101)
- Tang dynasty (618-907)

Candidate 2

- Yan Zhenqing (709-786)
- Song dynasty (960-1126)

Candidate 3

- Song dynasty (960-1126)
- Wang Anshi (1021-1086)

Candidate 4

- Qin Hui (1090-1155)
- Yuan dynasty (1271-1368)

Firstly, our system confirms a consistency of dates by comparing the dates of terms in each answer candidate. In the case of first answer candidate, the system compares time periods between Ouyang Xiu (1007-1072), Su Shi (1036-1101) and Tang dynasty (618-907). Tang dynasty has no overlap period with other terms. The candidate is considered as incorrect candidate in this step. In the case of third answer candidate, all terms overlap each other’s time periods. Secondly, the system confirms a consistency of dates between terms in each answer candidate and the ones in the underlined texts. Then the system extracts terms from the underlined sentence. The following terms were extracted from the example sentence.

Underlined Sentence

- Tang dynasty (618-907)
- Song dynasty (960-1126)

The system confirms a consistency between terms in candidates and the ones in underlined sentence. Next, matching rate of overlapping terms is calculated. The answer candidate which has the highest matching rate is selected as the answer for the question.

2. DICTIONARY OF WORLD HISTORY

We considered Dictionary of World History as a sufficient database since almost each term contains such information as “time period”, “area” or “religion”. However, the Dictionary of World History was not applicable as a database in a straightforward manner, since the date expressions are usually not plain numerical values, but time-related expressions, such as “9th century” and “10th century BC”, etc. Therefore the NNLP team created a database which contains both TPB and TPE using the Dictionary of World History. Figure 2 shows the procedure that extracts and annotates both TPB and TPE.

2.1 Expression of “Century”

There are many expressions of dates in the Dictionary of World History. Specially, expressions using “century” have a variety of meanings of periods between TPB and TPE. Therefore, we developed a conversion method, which converts to numerical values time-related expressions, such as the following ones:

1. Nth century
2. begin (from | by | after) Nth century
3. finish (by | at | in) Nth century
4. in the middle of Nth century
5. in the beginning of Nth century
6. in the end of Nth century

2.2 TPB and TPE

Our system needs both TPB and TPE in order to compare time periods in questions and their answer candidates. Therefore we needed to create a database which contains both TPB and TPE before checking for temporal overlaps between time periods in questions and their answer candidates. If a term’s explanation includes date expressions such as “Nth century” or “Nth century BC”, it can be converted to numerical values. Table 1 shows examples of conversion procedure.

Table 1: Examples of conversion procedure.

Expression	TPB	TPE
Nth century	100N-99	100N
Nth century BC	-100N	-(100N-99)
Nth century ~ Mth century	100N-99	100M
Nth century BC ~ Mth century BC	-100N	-(100M-99)
Nth century BC ~ Mth century	-100N	100M
the first half of Nth century	100N-99	100N-50

2.3 Creation of database by date extraction

In this section, we illustrate the method used to extract both TPB and TPE from the Dictionary of World History¹. Although the Dictionary of World History does not always include date expressions such as “Nth century” and “N century BC”, we attempted to fill in the blanks of both TPB and TPE for all possible terms.

¹<http://www.jikkyo.co.jp/book/detail/85096>

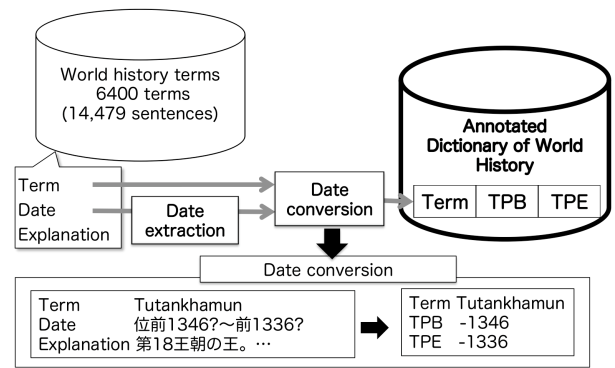


Figure 2: Date annotation for the Dictionary of World History.

To do this we used a list of time related terms. Firstly, we needed to choose a unique format to represent the term in the time span. Then, in aim to improve the number of dated terms in our database, we extracted dates from explanations available for each term. In particular, we split this task into seven steps according to their precision and the way of handling. In each of this level we used regular expressions to extract the dates and attached them to the corresponding terms. The seven steps are explain below.

Step 1 Formatting and converting already extracted date into a single start and possible end year unit.

Step 2 Extracting date from the explanation sentence, using regular expressions like $[0-9]\{1,4\}(\text{year})$ and all existing variations.

Step 3 Extracting already dated terms from the explanation sentence to estimate the target term date. For example, from the explanation content of the term “French revolution” we can assign it the starting date of 1789. To increase the precision of this step we only extract terms which were of the same class, section and chapter.

Step 4 Same as the step 3 using the term from the same section and chapter.

Step 5 Same as the step 3 using term from the same chapter.

Step 6 Same as the step 3 using all of the terms.

Step 7 Finally, if we could not extract any date from the explanation sentence, we assigned the category period to the term. This period is calculated by parsing all dated terms of the category to select the minimal and maximal date. For the step 7 we use the date of the caption index.

Step 8 Same as the step 7 using periods of the term section.

Step 9 Same as the step 7 using period of the term chapter.

Result checking To check the extraction algorithm we implemented a simple verification procedure, which verifies whether the starting date precedes the ending date.

The steps are executed in the order: step1 > step7 > step 2 > step7 > step3 > step4 > step5 > step6 > step8 > step9 The step 7 uses already dated terms, therefore why we execute it several time during the database improvement. The result of the extraction is represented in the Table 2.

Table 2: Number of annotated terms in each step.

Step	Annotated terms
step1	4,941
> step7	197
step2	2,253
> step7	2,856
step3	568
step4	521
step5	759
step6	1,163
step8	1,221
step9	0
Total	14,479

Table 3 shows examples of conversion. Time period expressions are not only "Century" but also other expressions such as "year", "month" and "day". Therefore we convert these expressions into numerical values. For example, we translate "March, 1802" into "1802.16". "1802.16" is calculated as follows: $1802 + \frac{31+28}{365}$

Table 3: Examples of conversion.

Raw data	TPB	TPE
March, 1802	1802.16	1802.16
1838/39, 97	1838/39	1897
1795, 1806	1795	1806
1861, 65	1861	1865
B.C 19 century , B.C. 612	-1900	-612

3. METHOD

Our method identifies time periods of historical terms in both questions and the answer candidates. The score for each answer candidate is calculated using world history term database. The terms from the database are extracted for each answer candidate.

3.1 Extraction of terms

At first, we use morphological analysis to match the terms from the database with the terms in questions. The extracted terms **t** are gathered by matching the database terms with the result of morphological analysis **m**, associated one to one. Moreover, we extract time-related terms such as 12th century, not present in the database.

3.2 Scoring

Each answer candidate is scored by judging whether the date of the term is intersected or not, so that the score is lower when the dates of terms are not close. Firstly, combinations of all pairs from the extracted terms are enumerated. The date relations between two terms is determined by whether the date of one term includes the date of another term (inclusion), or is intersected (intersection). For

example, for two terms T_1, T_2 with initial date being t_s and terminal date being t_e , intersection and inclusion are determined as follows.

Inclusion:

$$(t_{s(t_1)} \leq t_{s(t_2)} \wedge t_{e(t_1)} \geq t_{e(t_2)}) \vee (t_{s(t_1)} \geq t_{s(t_2)} \wedge t_{e(t_1)} \leq t_{e(t_2)})$$

Intersection : /

$(t_{e(t_1)} > t_{s(t_2)} \wedge t_{s(t_1)} < t_{e(t_2)})$ The score is calculated according to the following equation by using the number of intersection and inclusion.

$$s = \frac{\text{inclusion_number} + \text{intersection_number}}{\text{all_combination_of_termst}}$$

In addition, if the score is the same, the correct answer candidate becomes the one that has the lowest number of answer candidates. Moreover, the details of the scoring method is different according to question type. The question types we used is shown in Table 4. If the question is not of the type from Table 4, score of all answer candidates becomes 0.

Table 4: Question types.

Category	Question type
Sentence	sentence, term_location (no filling part)
True/False problem	(symbol-TF)*2, (symbol-TF)*2
Filling problem	term_other, term_person, term_location(filling part), (symbol-term_other)*2, (symbol-term_location) (symbol-term_other), (symbol-term_location)*2, (symbol-term_person)*2, (symbol-term_location) (symbol-sentence), (term_location) (term_location)
Time-ordering problem	o(sentence-sentence-sentence-sentence)

3.2.1 Sentence

The terms are extracted form answer candidates and underlined parts and the the score is calculated for each answer candidate.

3.2.2 Filling problem

We extract the terms only from the filled underlined parts with the candidate words.

3.2.3 Time-order

We judge whether the remainder between two events (A and B, B and C, C and D) is negative and divide the number of negative remainder by 3.0.

3.2.4 True/False Problem

We sum up the scores of item a and b, and divide the score by 2.0. If the score of the item becomes negative, the answer is considered as ‘False’.

4. RESULTS

Table 5 shows results for both Phase 1 and Phase 2. In Phase 1, we achieved 11 correct answers in 36 questions. The total score of the Center exam is 31, which was calculated by QALab. In Phase 2, we achieved 7 correct answers in 41 questions. The total score of the Center exam is 18, which was also calculated by QALab.

Moreover, Table 6 shows the result of the other center exam.

4.1 Example of Correct Answers

Below we explain an example of a correct answer using our method. It is the question number 35 in the exam from year 2009.

The underlined sentence of the question is as follows:

“Nomads settle down in the oasis city.”

The answer candidates are as follows:

1. In Sengoku period, Ethnic Groups called Wu Hu founded a country in Northern China.
2. Hephthalite was overthrown by both Gokturks and Sasanian Empire.
3. In Lushan Rebellion, Yuezhi sent for reinforcements in Tang dynasty.
4. To attack Xiongnu from both sides, Zhang Qian of Later Han Dynasty were dispatched in Western Regions.

Each answer candidate contains the following terms, which contains both TPB and TPE.

Candidate 1

- Sengoku period (-403.0, -221.0)
- Wu Hu (48.0, 308.0)
- Ethnic group (-4400000.0,1994.0)

Candidate 2

- Hephthalite (534.0, 567.0)
- Gokturks (501.0, 767.0)
- Sasanian Empire (226, 651)

Candidate 3

- Yuezhi (-100.0, -1.0)
- An Lushan Rebellion (755, 763)

- Tang dynasty (618, 907)

Candidate 4

- Xiongnu (-400.0, 308.0)
- Later Han Dynasty (25, 220)
- Zhang Qian (-214, -114)
- Western Regions (202.0, 907.0)

Score of the answer candidates are [1] 0.67, [2] 1.00, [3] 0.33, [4] 0.33 respectively. Therefore our system selected “candidate 2” as the answer for the question. It was the correct answer.

4.2 First and Second Run

In the first run, there were 11 questions for which our system could not compare the periods of terms correctly, though the solution is prepared for the question type. In the second run, there were 8 of such questions.

The reasons are as follows (the number of corresponding questions in parentheses):

- Lack of the sufficient number of extracted terms (1st:3, 2nd:3)
- Error in scoring process (1st:4, 2nd:3)
- Error in analyzing question sentences (1st:4, 2nd:2)

4.3 Question type

New question types are appeared in the formal run, which were not present in the training datasets. There are 3 of such questions in the first run and 3 questions in the second run. Our system is capable of solving only the questions of the type presented previously in Table 4. Therefore it could not solve the new types of questions. Unknown question types for our system are as follows.

- First run: #term_time, o(symbol-symbol-symbol), (symbol-symbol)(symbol-symbol)
- Second run: o(symbol-symbol-symbol), referenceSymbol, (symbol-sentence)(symbol-term_other)

5. CONCLUSIONS

We created and applied a date identification method, which checks for temporal overlaps between time periods in questions and their answer candidates. In this paper, we introduced details of this method and analyzed the test results. When tested on the World History Dictionary that is used for preparing to the exams, our approach achieved 30% of correct answers in the 2007 Center Exam Task and 17% for the 2003 Center Exam Task.

We found that the proposed system has some problems in the term extraction process, scoring process, and question analyzing process. We plan to improve them to obtain higher rate of correct answers. We also found that there are unknown question types for new exams. We also plan to take measures to deal with these unknown question types. A method which does not use the description of “question type” is needed to support our system.

Our system at the moment uses only time periods of a term. We plan to take other properties of a term, i.e. region, country and religion, into account as well.

Table 5: Results of the formal run.

Phase	Run ID	Question	Correct	Total Score	Correct answer rate
Phase1	Center-2007–Main-SekaishiB	36	11	31	0.3050
Phase2	Center-2003–Main-SekaishiB	41	7	18	0.1707

Table 6: Results of the dry run.

Exam ID	Question	Correct	Correct answer rate
Center-1997–Main-SekaishiB	40	10	0.2500
Center-2001–Main-SekaishiB	39	14	0.3590
Center-2005–Main-SekaishiB	36	7	0.1944
Center-2009–Main-SekaishiB	36	11	0.3056

6. ADDITIONAL AUTHORS

7. REFERENCES

- [1] S. Yasuda, H. Konishi, M. Asahara, M. Imada, and K. Maekawa. Temporal ordering annotation on the balanced corpus of contemporary written Japanese. *Natural Language Processing*, 20(2):201–221, 2013 (in Japanese).
- [2] K. Yasusi. Methods of extracting and representing year references from an encyclopedia for chronological-table-generating text searching. *The Special Interest Group Technical Reports of IPSJ*, 99(57):81–88, 1999 (in Japanese).