NTCIR Pilot Task: Math Task

Math Understanding Subtask

1. Format of the Training Dataset

The training dataset is available in three data formats: two XML-based formats and one plain text-based format. All files shall be in UTF-8 encoding.
The detailed explanation of the two XML based format can be found at:
http://ntcir-math.nii.ac.jp/wp-content/uploads/2012/08/ntcir1.html

The plain text-based format consists of two parts separated by a blank line. The first part contains the tokenized content of the data (split by whitespace, as defined by Python's interpretation of the regular expression u"\s+", after replacing any tags in unannotated files with spaces) preceded by a zero-based index. For instance, the text "*Let MATH_0801.0652_19 be subgroups of the group MATH_0801.0652_20.*" will be converted into the list shown in Figure 1. The index and the text are separated by a tab (\t) character.

```
0	Let
1	MATH_0801.0652_19
2	be
3	subgroups
4	of
5	the
6	group
7	MATH_0801.0652_20.
```

Figure 1. Conversion Result

The second part, after the blank line, contains the mathematical expressions and their descriptions. Each mathematical expression is again separated from the next one by a blank line.
We separate information about full descriptions and short descriptions into different files; however, the format is similar, as described below.

a. Full Descriptions

For each mathematical expression, the first line contains its identifier. The description indices follow, one description per line. They are written in the 'list-style': they are enclosed in square brackets, and every number inside the list is the index of a word of a description. When a description consists of a sequence of words, the index of the first word and the last word will be shown separated by a hyphen symbol (-) instead of the index of every word.

For example, let us consider the two data snippets in Figure 2. From the left column, we know that MATH_0802.1661_219 is described as "the element" and "the commitment". The right column gives information that the description of MATH_0801.0652_94 is "the non-divisible part of MATH_0801.0652_95". These descriptions would be encoded into the plain text format as shown in Figure 3.

The discontinuous part of description of MATH_0801.0652_94 (i.e. "of MATH_0801.0652_94") is concatenated with the previous part by using comma (,).

206	sends	493	if
207	the	494	the
208	element	495	non-divisible
209	MATH_0802.1661_219	496	part
210	(the	497	MATH_0801.0652_94
211	commitment)	498	of
		499	MATH_0801.0652_95
		500	contains

Figure 2. Two snippets of data

MATH_0802.1661_219 [207-208] [210-211]	MATH_0801.0652_94 [494-496,498-499]

Figure 3. Full description of MATH_0802.1661_219 and MATH_0801.0652_94

b. Short Descriptions

A full description may have more than one short description. Besides, short descriptions never have a discontinuous part.

Short descriptions are written in a very similar way to the full descriptions. One line in the file will represent short descriptions of a corresponding full description (one line per full description). The short descriptions belonging to the same full description will be separated by using comma (,), just like parts of discontinuous full descriptions were. For instance, let us consider the text in Figure 4.

There are two full descriptions of MATH_0801.0652_175. The first one is “the lattice” (1011-1012). The second one is “a proper union of the lattices MATH_0801.0652_176 and MATH_0801.0652_177 spanned by MATH_0801.0652_178 respectively” (1015-1027).

The first full description has one short description, which is itself (1011-1012). The second full description has two full descriptions, i.e., (1017) and (1015-1023). This information is written in the plain text format as shown in Figure 5.

949	The
950	lattice
951	MATH_0801.0652_175
952	is
953	a
954	proper
955	union
956	of
957	the
958	lattices
959	MATH_0801.0652_176
960	and
961	MATH_0801.0652_177
962	spanned
963	by
964	MATH_0801.0652_178
965	respectively,

Figure 4. Text surrounding MATH_0801.0652_175

```
MATH_0801.0652_175
[949-950]
[955,953-961]
```

Figure 5. Short descriptions of MATH_0801.0652_175

2. Submission

For the submission, we will distribute files in plain text-based format that is already pre-filled with the content data. The predicted descriptions should be added beneath, according to the specification above. The general format for the submission is shown in the Figure 6.

Any text following the expression identifier (separated by a tab) will be ignored by the evaluation script. Similarly, any text in a description line after the closing bracket is also ignored. The participants can use this comment syntax freely for human readability or debugging purposes.

0 aa
1 bb
2 cc
.
.
.
999 zz
blank line
MATH_{paperNumber}_1 *put any comment after a tab*
[0-9] *put any comment here*
[10,15-18]
blank line
MATH_{paperNumber}_2
//Description 1
…
//Description N
blank line
next mathematical expression

Figure 6. Submission Format

3. Evaluation

For the evaluation, we will match the positions of the extracted descriptions against the positions of gold-standard descriptions. There are two matching scenarios, namely strict matching and soft matching. Briefly, these scenarios are explained as follow:

- The extracted description will pass the strict matching evaluation if its position, consisting of start index and length, same as the position of the gold-standard description.
- The extracted description will pass the soft matching evaluation if its position contains, is contained in, or overlaps with the position of the gold-standard description.

We provide the participants a python script, so that they can measure the effectiveness of their method at any time for themselves. To run it, one should specify the file containing gold-standard data and the file(s) containing our prediction. There are three arguments for this script:

- -t *file containing prediction*
- -f *file containing gold-standard data of full descriptions*
- -s *file containing gold-standard data of short descriptions*

If we specify all of the arguments, our predictions will first be evaluated by using full descriptions. Those predictions that fail in the full description evaluation will be evaluated by considering the short descriptions. Since one mathematical expression may have more than one description, we will search for the best match between the extracted descriptions and the gold-standard descriptions.

Each evaluation will be executed by using strict matching and soft matching scenarios automatically. The metrics are precision and recall.