# README for the NTCIR-10 RITE-2 Test Collection

April 12, 2013

by Yotaro Watanabe yotaro-w [at] ecei.tohoku.ac.jp


[ Updated: May 30, 2014]

- (A) The asterisks (*) refer to open access data (use and/or redistribution permitted under the conditions of Creative Commons Attribution-Share-Alike License 3.0(Unported)).
- (B) The data with double asterisks (**) refer to the data included in NTCIR-9 RITE.


## [Test Collection]

The NTCIR-10 RITE-2 Test Collection is intended to evaluate systems that automatically recognize semantic relations (i.e entailment, paraphrase, and contradiction) between sentences.

The test collection includes:

(1) RITE2_JA_bc-mc-unittest: The Development data and the formal run data for Japanese BC, MC and UnitTest subtasks

(2) RITE2_JA_exam: The Development data and the formal run data for Japanese Entrance Exam subtasks (ExamBC and ExamSearch)

(3) RITE2_CS: The Development data and the formal run data for Simplified Chinese BC, MC and RITE4QA subtasks

(4) RITE2_CT: The Development data and the formal run data for Traditional Chinese BC, MC and RITE4QA subtasks


➤ **RITE2-JA-bc-mc-unittest includes the following files:**

- RITE2_JA_{dev,test}_{bc,mc}.xml ({development or test} data for the {BC or MC} subtask) *

- RITE2_JA_{dev,test}_{bc,mc}.parsed.KNP.xml (the {BC or MC} subtask data with linguistic analysis results by KNP) *

- RITE2_JA_{dev,test}_{bc,mc}.parsed.KNP_bp.xml (the {BC or MC} subtask data with linguistic analysis results by KNP) *

- RITE2_JA_{dev,test}_{bc,mc}.parsed.unidic.xml (the {BC or MC} subtask data with linguistic analysis results by MeCab & CaboCha) *

- RITE2_JA_testlabel_{bc,mc}.xml (test data with gold standard labels for the {BC or MC} subtask) *

- RITE2_JA_{dev,test}_unittest.xml (development data for the UnitTest subtask) *

- RITE2_JA_{dev,test}_unittest.parsed.KNP.xml (the UnitTest data with linguistic analysis results by KNP)*

- RITE2_JA_{dev,test}_unittest.parsed.KNP_bp.xml (the UnitTest data with linguistic analysis results by KNP) *

- RITE2_JA_{dev,test}_unittest.parsed.unidic.xml (the UnitTest data with linguistic analysis results by MeCab & CaboCha) *

- RITE2_JA_testlabel_unittest.xml (test data with gold standard labels for the UnitTest subtask) *

- RITE1_JA_{dev,test}_{bc,mc}.txt (the {BC, MC} data used in RITE1) **

- RITE1_JA_test_rite4qa.txt (the RITE4QA data used in RITE1) **

➢ **RITE2-JA-exam includes the following files:**
- RITE2_JA_dev_examsearch_wiki_docid.tsv  (text ids which are t1 candidates) *
- RITE2_JA_wiki.xml.bz2 (Wikipedia corpus for the Japanese Entrance Exam subtask) *

➢ **RITE2-CS includes the following files:**
- RITE2_CS_dev_bc.xml (development data for the BC subtask)
- RITE2_CS_dev_mc.xml (development data for the MC subtask)
- RITE2-TestData-CS-BC.xml (formal run data for the BC subtask)
- RITE2-TestData-CS-MC.xml (formal run data for the MC subtask)
- RITE2-TestData-CS-ExtraBC.xml (additional data for the BC subtask)
- RITE2-TestData-CS-ExtraMC.xml (additional data for the MC subtask)
- RITE2-TestData-CS-RITE4QA.xml (formal run data for the RITE4QA subtask)
- RITE2-TestData-CS-OptionalRITE4QA.xml (additional data for the RITE4QA subtask)

➢ **RITE2-CT includes the following files:**
- RITE2_CT_dev_bc.xml (development data for the BC subtask)
- RITE2_CT_dev_mc.xml (development data for the MC subtask)
- RITE2-TestData-CT-BC.xml (formal run data for the BC subtask)
- RITE2-TestData-CT-MC.xml (formal run data for the MC subtask)
- RITE2-TestData-CT-ExtraBC.xml (additional data for the BC subtask)
- RITE2-TestData-CT-ExtraMC.xml (additional data for the MC subtask)
- RITE2-TestData-CT-RITE4QA.xml (formal run data for the RITE4QA subtask)
- RITE2-TestData-CT-OptionalRITE4QA.xml (additional data for the RITE4QA subtask)

For more details, please refer to the NTCIR-10 RITE-2 overview paper available at the NTCIR online proceedings:

http://research.nii.ac.jp/ntcir/publication1-en.html