In this paper, we describe the IMTKU (Information Management at TamKang University) textual entailment system for recognizing inference in text at NTCIR-10 RITE-2 (Recognizing Inference in Text). We proposed a textual entailment system using a hybrid approach that integrates semantic features and machine learning techniques for recognizing inference in text at NTCIR-10 RITE-2 task. We submitted 3 official runs for BC, MC and RITE4QA subtask. In NTCIR-10 RITE-2 task, IMTKU team achieved 0.509 in the CT-MC subtask, 0.663 in the CT-BC subtask; 0.402 in the CS-MC subtask, 0.627 in the CS-BC subtask; In MRR index, 0.257 in the CT-RITE4QA subtask, 0.338 in the CS-RITE4QA subtask. IMTKU is ranked #1 in the CS-RITE4QA subtask of NTCIR-10 RITE-2 task.

**System Architecture**

**Methods for Official Runs**

**RITE-2-IMTKU-CT-BC-01**
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity, Text Based Global Distance) in SVM.

**RITE-2-IMTKU-CT-BC-02**
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

**RITE-2-IMTKU-CT-BC-03**
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

**RITE-2-IMTKU-CS-RITE4QA Subtask**
* Resource: Stanford Parser
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Longest Common Substring, Word Length Ratio, Text Length, Similarity between t1 and t2, Text Edit Distance) in SVM.

**RITE-2-IMTKU-CS-RITE4QA-01**
* Resource: Stanford Parser
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity) in SVM.

**RITE-2-IMTKU-CS-RITE4QA-02**
* Resource: Stanford Parser
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity, Text Based Global Distance) in SVM.

**RITE-2-IMTKU-CS-RITE4QA-03**
* Resource: Stanford Parser
* Feature Extraction from normalized t1 and t2. Measure similarity match between t1 and t2.
* Multiple Features used (Antonym, Negation, Word Based Similarity, Token Based Similarity, Lexical overlap, Text Pair Length, Token Length, WorkNet Similarity, Text Based Global Distance) in SVM.

**Performance**

**RITE2-IMTKU-CT-BC-01**
* Top1 MRR Top5
* 0.4067 0.7221% 0.3377
* 0.509 0.7229%

**RITE2-IMTKU-CT-BC-02**
* Top1 MRR Top5
* 0.4267 0.6885% 0.3377
* 0.338 0.6885%

**RITE2-IMTKU-CT-BC-03**
* Top1 MRR Top5
* 0.2800 0.3267% 0.2003
* 0.308 0.3267%

**RITE2-IMTKU-CS-RITE4QA-01**
* Top1 MRR Top5
* 0.1067 0.1991% 0.0500
* 0.3377 0.3267%

**RITE2-IMTKU-CS-RITE4QA-02**
* Top1 MRR Top5
* 0.1467 0.2144% 0.0500
* 0.3377 0.3267%

**RITE2-IMTKU-CS-RITE4QA-03**
* Top1 MRR Top5
* 0.1733 0.2603% 0.0500
* 0.2003 0.2603%

**Discussion**

**Issues of Definition in RITE MC between NTCIR-9 and NTCIR-10:**
* Definition of NTCIR-9 MC subtask: “A 5-way labeling subtask to detect (forward / reverse / bidirection) entailment or no entailment (contradiction / independence) in a text pair.”
* Definition of NTCIR-10 MC subtask: “A 4-way labeling subtask to detect (forward / bidirection) entailment or no entailment (contradiction / independence) in a text pair.”
* Misused NTCIR-9 MC labels on NTCIR-10 MC test datasets where “Reverse” label should be included.

**Cross Validation of Development and Test datasets of NTCIR-10 RITE-2 Task**

**Datasets**

<table>
<thead>
<tr>
<th>RITE2-CT_dev_test_bc_g.txt</th>
<th>68.85%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RITE1_CT_r1000_dev_test_bc_g.txt</td>
<td>73.83%</td>
</tr>
<tr>
<td>(Random select 1000 pairs from RITE1 BC Dev+Test Dataset)</td>
<td></td>
</tr>
<tr>
<td>RITE1_CT_dev_test_bc_g.txt</td>
<td>72.29%</td>
</tr>
<tr>
<td>(RITE1 BC Dev+Test Dataset: 421+950 = 1321 pairs)</td>
<td></td>
</tr>
<tr>
<td>RITE1_CT_dev_bc_g.txt (gold standard)</td>
<td>72.21%</td>
</tr>
<tr>
<td>(RITE1 BC Development Dataset: 421 pairs)</td>
<td></td>
</tr>
</tbody>
</table>

**Acknowledgements:**
This research was supported in part by the National Science Council of Taiwan under Grants NSC101-3113-P-032-001 and TKU research grant.