NTCIR Evaluation Activities: Recent Advances on RITE (Recognizing Inference in Text)

Min-Yuh Day, Ph.D.
Assistant Professor

Department of Information Management
Tamkang University

http://mail.tku.edu.tw/myday
Outline

• Overview of NTCIR Evaluation Activities
• Recent Advances on RITE (Recognizing Inference in Text)
• Research Issues and Challenges of Empirical Methods for Recognizing Inference in Text (EM-RITE)
Overview of NTCIR Evaluation Activities
NTCIR

NII Testbeds and Community for Information access Research

NII: National Institute of Informatics

http://www.nii.ac.jp/en/
A series of evaluation workshops designed to enhance research in information-access technologies by providing an infrastructure for large-scale evaluations.

Data sets, evaluation methodologies, forum
• Project started in late 1997
  – 18 months Cycle

Source: Kando et al., 2013
• **Data sets (Test collections or TCs)**
  - Scientific, news, patents, web, CQA, Wiki, Exams
  - Chinese, Korean, Japanese, and English
Tasks (Research Areas)

- IR: Cross-lingual tasks, patents, web, Geo, Spoken
- QA: Monolingual tasks, cross-lingual tasks
- Summarization, trend info., patent maps,
- Inference,
- Opinion analysis, text mining, Intent, Link Discovery, Visual

Source: Kando et al., 2013
NTCIR-10 (2012-2013)

135 Teams Registered to Task(s)

973 Teams Registered so far

Source: Kando et al., 2013
Procedures in NTCIR Workshops

- Call for Task Proposals
- Selection of Task Proposals by Program Committee
- Discussion about Task Design in Each Task
- **Registration to Task(s)**
  - Deliver Training Data (Documents, Topics, Answers)
    - Experiments and Tuning by Each Participants
  - Deliver Test Data (Documents and Topics)
    - Experiments by Each Participants
- **Submission of Experimental Results**
- Pooling the Answer Candidates from the Submissions, and Conduct Manual Judgments
- **Return Answers (Relevance Judgments) and Evaluation Results**
- **Conference** Discussion for the Next Round
- Test Collection Release for non-participants

Source: Kando et al., 2013
## Tasks in NTCIR (1999-2013)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Year</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media·UGC</td>
<td>'99</td>
<td>Community QA</td>
</tr>
<tr>
<td></td>
<td>'01</td>
<td>Opinion Analysis</td>
</tr>
<tr>
<td>Module-Based</td>
<td>'02</td>
<td>IR + Inference</td>
</tr>
<tr>
<td></td>
<td>'04</td>
<td>QA + IR</td>
</tr>
<tr>
<td>Special Domain</td>
<td>'05</td>
<td>MedNLP</td>
</tr>
<tr>
<td></td>
<td>'07</td>
<td>Math Retrieval</td>
</tr>
<tr>
<td></td>
<td>'10</td>
<td>Entrance Exam</td>
</tr>
<tr>
<td></td>
<td>'11</td>
<td>Spoken Doc Retrieval</td>
</tr>
<tr>
<td></td>
<td>'13</td>
<td>Geotime Retrieval</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patent Retrieval</td>
</tr>
<tr>
<td>Question Answering</td>
<td></td>
<td>Any types of Qs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dialogue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cross-Lingual IR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Factoid, List</td>
</tr>
<tr>
<td>IE &amp; IR</td>
<td></td>
<td>Link Discovery</td>
</tr>
<tr>
<td>Semantic</td>
<td></td>
<td>Inference</td>
</tr>
<tr>
<td>Summarization</td>
<td></td>
<td>Text Mining/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trend Info Visualization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text Summarization</td>
</tr>
<tr>
<td>Interactive</td>
<td></td>
<td>Interactive IR and Visualization</td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td>Intent mining, Diversified Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SERP Quality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Web Search</td>
</tr>
<tr>
<td>Cross-Lingual</td>
<td></td>
<td>Machine Translation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cross-Lingual IR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-EnglishIR</td>
</tr>
<tr>
<td>Tetxt Retrieval</td>
<td></td>
<td>Ad Hoc. IR for QA</td>
</tr>
</tbody>
</table>

Year that the conference was held, The Tasks started 18 Months before

Source: Kando et al., 2013
### Evaluation Tasks from NTCIR-1 to NTCIR-10

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Term Recognition and Role Analysis (TMREC)</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ad hoc/Crosslingual IR(1) - Chinese/English/Japanese IR(2) - CLIR(3-6)</td>
<td>28</td>
<td>30</td>
<td>20</td>
<td>26</td>
<td>25</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text Summarization Challenge (TSC)</td>
<td></td>
<td>9</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Retrieval (WEB)</td>
<td></td>
<td></td>
<td>7</td>
<td>11</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question Answering Challenge (QAC)</td>
<td></td>
<td>16</td>
<td>18</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patent Retrieval [and Classification] (PATENT)</td>
<td></td>
<td></td>
<td>10</td>
<td>18</td>
<td>7</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multimodal Summarization for Trend Information (MUST)</td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>15</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crosslingual Question Answering (CLQA)(5,6) -&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>12</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>Advanced Crosslingual Information Access (ACILIA)(7,8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>Opinion(6) - Multilingual Opinion Analysis (MOAT)(7,8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patent Mining (PAT-MN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Community Question Answering (CQA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Geotemporal IR (GeoTime)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Interactive Visual Exploration (Vis-Ex)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Patent Translation (PAT-MT)(7,8) - Patent Machine Translation (PatentMT)(9,10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>INTENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>One Click Access (1CLICK)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Recognizing Inference in Text (RITE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Crosslingual Link Discovery (Crosslink)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>IR for Spoken Documents (SpokenDoc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Medical Natural Language Processing (MedNLP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

Source: Joho et al., 2013
Source: Kando et al., 2013
The 10th NTCIR Conference
Evaluation of Information Access Technologies
June 18-21, 2013
National Center of Sciences, Tokyo, Japan
Organized by:
NTCIR Organizing Committee
National Institute of Informatics (NII)
NII Testbeds and Community for Information access Research

- Data sets / Users’ Information Seeking Tasks
- Evaluation Methodology
- Reusable vs Reproducibility
- User-Centered Evaluation
- Experimental Platforms
- Open Advancement
- Advanced NLP (Knowledge- or Semantic-based)
- Diversified IA Applications in the Real World
- Best Practice for a technology
  - Best Practice for Evaluation Methodology
- Big Data (Documents + Behaviour data)

Source: Kando et al., 2013
NTCIR-11
Evaluation of Information Access Technologies
July 2013 - December 2014

NTCIR-11
Evaluation of Information Access Technologies
July 2013 - December 2014

Task Participation

NTCIR-11 Tasks are as follows:
- CORE TASKS
  IMine  Math-2  MedNLP-2  MobileClick  RITE-VAL  SpokenQuery&Doc
- PILOT TASKS
  QALab  Temporaila

Task participation will start from October 2013. Please see slides for task introduction at the NTCIR-11 Kick-Off Event.

NTCIT-11 Evaluation Tasks
(July 2013 - December 2014)

• Six Core Tasks
  – Search Intent and Task Mining ("IMine")
  – Mathematical Information Access ("Math-2")
  – Medical Natural Language Processing ("MedNLP-2")
  – Mobile Information Access ("MobileClick")
  – Recognizing Inference in TExt and Validation ("RITE-VAL")
  – Spoken Query and Spoken Document Retrieval ("SpokenQuery&Doc")

• Two Pilot Tasks
  – QA Lab for Entrance Exam ("QALab")
  – Temporal Information Access ("Temporalia“)

NTCIR-11 Important Dates

(Event with * may vary across tasks)

- 2/Sep/2013  Kick-Off Event in NII, Tokyo
- 20/Dec/2013  Task participants registration due *
- 5/Jan/2014  Document set release *
- Jan-May/2014  Dry Run *
- Mar-Jul/2014  Formal Run *
- 01/Aug/2014  Evaluation results due *
- 01/Aug/2014  Early draft Task overview release
- 01/Sep/2014  Draft participant paper submission due *
- 01/Nov/2014  All camera-ready copy for proceedings due
- 9-12/Dec/2014  NTCIR-11 Conference in NII, Tokyo

http://research.nii.ac.jp/ntcir/ntcir-11/dates.html
NTCIR-11 Organization

• **NTCIR-11 General Co-Chairs:**
  - Noriko Kando (National Institute of Informatics, Japan)
  - Tsuneaki Kato (The University of Tokyo, Japan)
  - Douglas W. Oard (University of Maryland, USA)
  - Tetsuya Sakai (Waseda University, Japan)
  - Mark Sanderson (RMIT University, Australia)

• **NTCIR-11 Program Co-Chairs:**
  - Hideo Joho (University of Tsukuba, Japan)
  - Kazuaki Kishida (Keio University, Japan)

Recent Advances on RITE  
(Recognizing Inference in Text) 

NTCIR-9 RITE (2010-2011)  
NTCIR-10 RITE-2 (2012-2013)  
NTCIR-11 RITE-VAL (2013-2014)
Overview of the Recognizing Inference in TExt (RITE-2) at NTCIR-10

Yotaro Watanabe
Tohoku University

Yusuke Miyao
NII

Junta Mizuno
Tohoku University

Tomohide Shibata
Kyoto University

Hiroshi Kanayama
IBM Research

Cheng-Wei Lee
Academia Sinica

Chuan-Jie Lin
National Taiwan Ocean University

Shuming Shi
MSRA

Teruko Mitamura
CMU

Noriko Kando
NII

Hideki Shima
CMU

Kohichi Takeda
IBM Research

Overview of RITE-2

• RITE-2 is a generic benchmark task that addresses a common semantic inference required in various NLP/IA applications.

\[ t_1: \text{Yasunari Kawabata} \text{ won the Nobel Prize in Literature for his novel “Snow Country.”} \]

Can \( t_2 \) be inferred from \( t_1 \) ?

\( t_2: \text{Yasunari Kawabata} \text{ is the writer of “Snow Country.”} \)

Source: Watanabe et al., 2013
Yasunari Kawabata
Writer

Yasunari Kawabata was a Japanese short story writer and novelist whose spare, lyrical, subtly-shaded prose works won him the Nobel Prize for Literature in 1968, the first Japanese author to receive the award.

http://en.wikipedia.org/wiki/Yasunari_Kawabata
RITE vs. RITE-2

Source: Watanabe et al., 2013
Motivation of RITE-2

• Natural Language Processing (NLP) / Information Access (IA) applications
  – Question Answering, Information Retrieval, Information Extraction, Text Summarization, Automatic evaluation for Machine Translation, Complex Question Answering

• The current entailment recognition systems have not been mature enough
  – The highest accuracy on Japanese BC subtask in NTCIR-9 RITE was only 58%
  – There is still enough room to address the task to advance entailment recognition technologies

Source: Watanabe et al., 2013
BC and MC subtasks in RITE-2

$t_1$: Yasunari Kawabata won the Nobel Prize in Literature for his novel “Snow Country.”
$t_2$: Yasunari Kawabata is the writer of “Snow Country.”

• **BC subtask**
  - Entailment ($t_1$ entails $t_2$) or Non-Entailment (otherwise)
  
• **MC subtask**
  - Bi-directional Entailment ($t_1$ entails $t_2$ & $t_2$ entails $t_1$)
  - Forward Entailment ($t_1$ entails $t_2$ & $t_2$ does not entail $t_1$)
  - Contradiction ($t_1$ contradicts $t_2$ or cannot be true at the same time)
  - Independence (otherwise)

Source: Watanabe et al., 2013
Development of BC and MC data

1. Retrieve pairs of sentences
2. Edit pairs if needed
3. For each example, 5 annotators assigned its semantic label
4. Accept an example if 4 or more annotators assigned the same label to the example

Source: Watanabe et al., 2013
Entrance Exam subtasks
(Japanese only)

Entrance exam problem
National Center Test for University Admission
(Daigaku Nyushi Center Shiken)

第1問 モニュメントや歴史的建造物について述べた次の文章A～Cを読み、下の問い（問1～11）に答えよ。（配点 33）

A 現在、アテネの中心部の丘にその地名を誇る①パルテノン神殿は、古代ギリシアを象徴する歴史的建造物である。この神殿は、②オスマン帝国の支配下でモスクとして利用されたこともあったが、18世紀には廃墟となっていた。1799年にイギリスの大使としてイスタンブールに赴任したエルギン侯は、③ギリシアを訪れ、パルテノン神殿の遺跡から壁画を収集し、本国に送った。今日、大英博物館で「エルギン・マーブル」として展示されているものがそれである。1987年、パルテノン神殿は、世界遺産として登録された。

問3 下線部②の国について述べた文として最も適当なものを、次の①～⑤のうちから一つ選べ。[3]

① スレイマン1世の時代が最盛期であった。
② 国教はシーゾー派のイスラム教であった。
③ バルカン半島に誕生した後、小アジアへ進出した。
④ ベルリン会議により、ボスニア＝ヘルツェゴビナの統治権を得た。
⑤ スルタン・スレイマン1世(Kanuni Sultan Süleyman, オスマン語 Sultan Süleyman, トルコ語 Süleyman, 1494年11月6日 - 1566年9月6日)は、オスマン帝国の第10代皇帝(在位:1520年 - 1566年)。

46年の長期にわたる在位の中で19国もの対外遠征を行い、数多くの軍事的成績を収めてオスマン帝国を最盛期に導いた。英語では、「壮麗な Magnificent」の名で呼ばれ、日本ではしばしば「スレイマン大帝」と呼ばれる。トルコでは法典を編纂し帝国の制度を整備したことから「立法帝(カー ヌーニー al-Qanuni)/ Kanuni」のあだ名で知られている。

$t_1$: スレイマン1世は数多くの軍事的成績を収めてオスマン帝国を最盛期に導いた。（Suleiman I contributed in a lot of military successes and led the Ottoman Empire to its peak.

$t_2$: オスマン帝国ではスレイマン1世の時代が最盛期であった。（The Ottoman Empire’s peak was during the reign of Suleiman I).
Entrance Exam subtask: BC and Search

- **Entrance Exam BC**
  - Binary-classification problem (Entailment or Nonentailment)
  - t1 and t2 are given

- **Entrance Exam Search**
  - Binary-classification problem (Entailment or Nonentailment)
  - t2 and a set of documents are given
    - Systems are required to search sentences in Wikipedia and textbooks to decide semantic labels

Source: Watanabe et al., 2013
UnitTest (Japanese only)

- **Motivation**
  - Evaluate how systems can handle linguistic phenomena that affects entailment relations

- **Task definition**
  - Binary classification problem (same as BC subtask)

Source: Watanabe et al., 2013
RITE4QA (Chinese only)

• Motivation
  – Can an entailment recognition system rank a set of unordered answer candidates in QA?

• Dataset
  – Developed from NTCIR-7 and NTCIR-8 CLQA data
    • t1: answer-candidate-bearing sentence
    • t2: a question in an affirmative form

• Requirements
  – Generate confidence scores for ranking process

Source: Watanabe et al., 2013
Evaluation Metrics

• Macro F1 and Accuracy (BC, MC, ExamBC, ExamSearch and UnitTest)

\[
\text{MacroF1} = \frac{1}{|C|} \sum_{c \in C} F1_c \\
\text{Accuracy} = 100 \times \frac{N_{correct}}{N_{examples}}
\]

• Correct Answer Ratio (Entrance Exam)
  – Y/N labels are mapped into selections of answers and calculate accuracy of the answers

• Top1 and MRR (RITE4QA)

\[
\text{Top1} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \left[ \text{top answer is correct} \right] \\
\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}
\]

Source: Watanabe et al., 2013
Countries/Regions of Participants

- **China**: 3 groups
- **Japan**: 15 groups
- **Ireland**: 1 group
- **India**: 1 group
- **Taiwan**: 8 groups

Source: Watanabe et al., 2013
Formal Run Results: BC (Japanese)

- The best system achieved over 80% of accuracy (The highest score in BC subtask at RITE was 58%)
- The difference is caused by
  - Advancement of entailment recognition technologies
  - Strict data filtering in the data development

Source: Watanabe et al., 2013
The top scores are almost the same as those in NTCIR-9 RITE

Source: Watanabe et al., 2013
RITE4QA
(Traditional/Simplified Chinese)

Source: Watanabe et al., 2013
Participant’s approaches in RITE-2

• **Category**
  – Statistical (50%)
  – Hybrid (27%)
  – Rule-based (23%)

• **Fundamental approach**
  – Overlap-based (77%)
  – Alignment-based (63%)
  – Transformation-based (23%)

Source: Watanabe et al., 2013
Summary of types of information explored in RITE-2

- Character/word overlap (85%)
- Syntactic information (67%)
- Temporal/numerical information (63%)
- Named entity information (56%)
- Predicate-argument structure (44%)
- Entailment relations (30%)
- Polarity information (7%)
- Modality information (4%)

Source: Watanabe et al., 2013
Summary of Resources Explored in RITE-2

- **Japanese**
  - Wikipedia (10)
  - Japanese WordNet (9)
  - ALAGIN Entailment DB (5)
  - Nihongo Goi-Taikei (2)
  - Bunruigoihyo (2)
  - Iwanami Dictionary (2)

- **Chinese**
  - Chinese WordNet (3)
  - TongYiCi CiLin (3)
  - HowNet (2)

Source: Watanabe et al., 2013
Advanced approaches in RITE-2

- **Logical approaches**
  - Dependency-based Compositional Semantics (DCS) [BnO], Markov Logic [EHIME], Natural Logic [THK]

- **Alignment**
  - GIZA [CYUT], ILP [FLL], Labeled Alignment [bcNLP, THK]

- **Search Engine**
  - Google and Yahoo [DCUMT]

- **Deep Learning**
  - RNN language models [DCUMT]

- **Probabilistic Models**
  - N-gram HMM [DCUMT], LDA [FLL]

- **Machine Translation**
  - [JUNLP, JAIST, KC99]

Source: Watanabe et al., 2013
NTCIR-11
RITE-VAL
(Recognizing Inference in Text and Validation)

https://sites.google.com/site/ntcir11riteval/
NTCIR-11 RITE-VAL Task
(Recognizing Inference in Text and Validation)

Suguru Matsuyoshi¹
¹University of Yamanashi

Yotaro Watanabe²
²Tohoku University

Yusuke Miyao³
³National Institute of Informatics

Tomohide Shibata⁴
⁴Kyoto University

Teruko Mitamura⁵
⁵Carnegie Mellon University

Chuan-Jie Lin⁶
⁶National Taiwan Ocean University

Cheng-Wei Shih⁷
⁷Academia Sinica

RITE is a benchmark task for automatically detecting the following **semantic relations between two sentences:**

- entailment, paraphrase and contradiction.

Given a text $t_1$, can a computer infer that a hypothesis $t_2$ is most likely true (i.e., $t_1$ entails $t_2$)?

- $t_1$: **Yasunari Kawabata** won the Nobel Prize in Literature for his novel **“Snow Country.”**
- $t_2$: **Yasunari Kawabata** is the writer of **“Snow Country.”**

**Target languages:**

- Japanese, Simplified Chinese, Traditional Chinese, and English.

Source: Matsuyoshi et al., 2013
Main two tasks of RITE-VAL

Source: Matsuyoshi et al., 2013
IEEE IRI 2013 Workshop Program

Chair: Min-Yuh Day

- Rank Correlation Analysis of NTCIR-10 RITE-2 Chinese Datasets and Evaluation Metrics
  Chuan-Jie Lin (1), Cheng-Wei Lee (2), Cheng-Wei Shih (2) and Wen-Lian Hsu (2)
  (1) National Taiwan Ocean University, Taiwan
  (2) Academia Sinica, Taiwan

- Chinese Textual Entailment with Wordnet Semantic and Dependency Syntactic Analysis
  Chun Tu and Min-Yuh Day
  Tamkang University, Taiwan

- Entailment Analysis for Improving Chinese Textual Entailment System
  Shih-Hung Wu (1), Shan-Shun Yang (1), Liang-Pu Chen (2), Hung-Sheng Chiu (2) and
  Ren-Dar Yang (2)
  (1) Chaoyang University of Technology, Taiwan
  (2) Institute for Information Industry, Taiwan

- Interest Analysis using Social Interaction Content with Sentiments
  Lun-Wei Ku and Chung-Chi Huang
  Academia Sinica, Taiwan

- Clustering and Summarization Topics of Subject Knowledge Through Analyzing Internal Links
  of Wikipedia
  I-Chin Wu, Chi-Hong Tsai and Yu-Hsuan Lin
  Fu-Jen Catholic University, Taiwan
IMTKU System Architecture for NTCIR-9 RITE

Preprocessing

RITE Corpus (T1, T2 Pairs)

CKIP AutoTag (POS Tagger)

HIT Dependency Parser

SINICA BOW

HIT TongYiCiLing

Chinese Antonym

Feature Extraction

Machine Learning Module

Knowledge-Based Module

Voting Strategy Module

Predict Result (BC)/(MC)
IMTKU System Architecture for NTCIR-10 RITE-2

XML Train Dataset of RITE Corpus (T1, T2 Pairs)

Train

- Preprocessing
- Feature Generation
- Feature Selection
- Training Model (SVM Model)
- Evaluation of Model (k-fold CV)

XML Test Dataset of RITE Corpus (T1, T2 Pairs)

Predict

- Preprocessing
- Feature Generation
- Feature Selection
- Use model for Prediction
- Predict Result (Open Test)

- CKIP AutoTag (POS Tagger)
- HIT TongYiCiLing
- WordNet
- Dependency Parser
- Negation Antonym
Discussions

• 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.”
## IMTKU Experiments for NTCIR-10 RITE-2 Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>10 Fold CV Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RITE2_CT_dev_test_bc_g.txt</td>
<td>68.85%</td>
</tr>
<tr>
<td>(RITE2 BC Dev + Test Dataset: 1321 + 881 = 2202 pairs)</td>
<td></td>
</tr>
<tr>
<td>RITE1_CT_r1000_dev_test_bc_g.txt</td>
<td>73.83%</td>
</tr>
<tr>
<td>(Random select <strong>1000 pairs</strong> 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 + 900 = 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>
## IMTKU Experiments for NTCIR-9 RITE Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>10 Fold CV Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RITE1_CT_dev_bc_g.txt (gold standard)</td>
<td>76.48%</td>
</tr>
<tr>
<td>(BC Development Dataset: 421 pairs)</td>
<td></td>
</tr>
<tr>
<td>RITE1_CT_test_bc_g.txt</td>
<td>66.33%</td>
</tr>
<tr>
<td>(BC Test Dataset: 900 pairs)</td>
<td></td>
</tr>
<tr>
<td>RITE1_CT_dev_test_bc_g.txt</td>
<td>67.67%</td>
</tr>
<tr>
<td>(BC Dev+Test Dataset: 421+900 = 1321 pairs)</td>
<td></td>
</tr>
</tbody>
</table>
Tamkang University

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-10 RITE-2

Demo

http://rite.im.tku.edu.tw

Min-Yuh Day *, Chun Tu, Shih-Jhen Huang, Hou-Cheng Vong, Shih-Wei Wu

myday@mail.tku.edu.tw

2013/06/19

NTCIR-10 Conference, June 18-21, 2013, Tokyo, Japan
Demo

Textual (T1): 一九九七年香港回歸中國

Hypothesis (T2): 香港的主權和領土是在一九九七由英國歸還給中國的。

Result: No 0.653509

Detail:
- Longest Common subsequence: 7
- T1 Length: 11
- T2 Length: 24
- Length Difference: -11
- Length Ratio: 0.4583
- T1 Token Length: 5
- T2 Token Length: 16
- Token Length Ratio: 0.3125
- Token Length Difference: -11
- March-Based Edit Distance: 10

Referenced Paper: "IMTKU Textual Entailment System for Recognizing/Inference in Text at NTCIR-10 RITE2" [Source]
IMTKU Textual Entailment System
Department of Information Management, Tamkang University

Demo

Example:  ● Sample 1  ● Sample 2  ● Sample 3

Textual (T1):
一九九七年香港回歸中國

Hypothesis (T2):
香港的主權和領土是在一九九七年由英國歸還給中國的。

Result: No 0.653509

Detail:
Word Similarity   : 0.8747
Word Net Similarity: 18.55
Word Net Similarity Ratio: 23.083333333333333
Word Net Similarity Short: 30.333333333333333
Negation Number Difference: 0
Antonym Number Difference: 0
T1: 一九九七年香港回歸中國
T2: 香港的主權和領土是在一九九七年由英國歸還給中國的。
T1 CKIP: ？(QUESTIONCATEGORY) 一九九七年 (N) 香港 (N)

http://rite.im.tku.edu.tw

In conjunction with IEEE IRI 2013

San Francisco, USA
August 14, 2013

https://sites.google.com/site/emrite2013/
IEEE International Workshop on Empirical Methods for Recognizing Inference in TExT

In conjunction with IEEE IRI 2013

Welcome to IEEE EM-RITE 2013

IEEE International Workshop on Empirical Methods for Recognizing Inference in TExT
(IIEEE EM-RITE 2013)

In conjunction with IEEE IRI 2013

San Francisco, USA
August 14, 2013

Textual Entailment and Paraphrase are inference tasks of natural language processing (NLP) for automatically detecting entailment, paraphrase, and contradiction in texts. The aim of this workshop is to provide a forum for original high-quality research contributions on empirical methods for recognizing inference in text as well as multidisciplinary research opportunities.

Topics of interest include but are not limited to practical areas that span a variety of aspects of empirical methods for recognizing inference in text including:

- Guidelines, standards, best practices and models for the construction and annotation of Textual Entailment datasets
- Evaluation of Knowledge Resources for Textual Entailment
- Recognizing Inference in Text
- Recognizing Textual Entailment
Conclusions

• Welcome to join **NTCIR-11 RITE-VAL**

• Online demo system **RITE.IM.TKU**
  – [http://rite.im.tku.edu.tw](http://rite.im.tku.edu.tw)

• Welcome to join **IEEE EM-RITE** 2014, 2015, ...
References


• Hideo Joho and Tetsuya Sakai, Overview of NTCIR-10, Proceedings of NTCIR-10, 2013


Q & A

NTCIR Evaluation Activities: Recent Advances on RITE (Recognizing Inference in Text)

Min-Yuh Day, Ph.D.
Assistant Professor

Department of Information Management
Tamkang University

http://mail.tku.edu.tw/myday