

# RITE (Recognizing Inference in TExt)

## NTCIR-9総括と NTCIR-10へ向けて



**Yotaro  
Watanabe<sup>1</sup>**  
<sup>1</sup>Tohoku University



**Junta  
Mizuno<sup>1</sup>**  
<sup>1</sup>Tohoku University



**Yusuke  
Miyao<sup>2</sup>**  
<sup>2</sup>National Institute  
of Informatics



**Tomohide  
Shibata<sup>3</sup>**  
<sup>3</sup>Kyoto  
University



**Cheng-  
Wei Lee<sup>4</sup>**  
<sup>4</sup>Academia  
Sinica



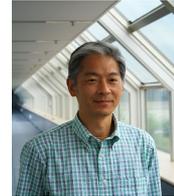
**Chuan-  
Jie Lin<sup>5</sup>**  
<sup>5</sup>National Taiwan  
Ocean University



**Shuming  
Shi<sup>6</sup>**  
<sup>6</sup>Microsoft  
Research Asia



**Hiroshi  
Kanayama<sup>7</sup>**  
<sup>7</sup>IBM Research



**Koichi  
Takeda<sup>7</sup>**



**Hideki  
Shima<sup>8</sup>**  
<sup>8</sup>Carnegie Mellon University



**Teruko  
Mitamura<sup>8</sup>**

# RITEの概要

- RITE (Recognizing Inference in TExt)
  - テキスト間の含意 (=推論), 換言 (=同じ意味), 矛盾の判断を対象とした, テキスト理解システムのためのベンチマークタスク
- 含意関係の定義
  - テキスト  $t_1$  が与えられたとき, 仮説  $t_2$  も真であると推論可能

$t_1$ : 川端康成は「雪国」などの作品でノーベル文学賞を受賞した

$t_2$ : 川端康成は「雪国」の著者である (含意)

- 対象言語
  - 日本語, 中国語 (Simplified, Traditional)

# Motivation

- 様々な情報アクセス(Information Access)技術への応用
  - 質問応答 (Question Answering)
    - 質問の回答が含まれる文を，含意関係認識技術を用いて発見し，回答を抽出
  - 情報検索 (Information Retrieval)
    - クエリを含意するテキスト，または矛盾するテキストを検索することで情報を集約，組織化
  - 文書要約 (Text Summarization)
    - 含意関係の認識により，同一内容が含まれる文を削除して冗長性を排除
- 言葉の意味を考慮した，高度な情報アクセス技術の実現

# RITE のサブタスク

- Binary-Class (BC) サブタスク
  - 1組の文対 $\langle t_1, t_2 \rangle$ が与えられたとき,  $t_1$ から仮説 $t_2$ が真であるかどうかを出力 (2値)
- Multi-Class (MC) サブタスク
  - 1組の文対 $\langle t_1, t_2 \rangle$ が与えられたとき, 5種類の関係に分類
    - 換言, 前向き含意, 後ろ向き含意, 矛盾, その他
- Entrance Exam (大学入試) サブタスク
  - 大学入試の問題を対象に, BCと同じ課題を解く
- RITE4QA サブタスク
  - QAシステムへ組み込むことを視野に入れた設定
  - 回答候補をランキング

# 大学入試サブタスク

## センター試験の問題

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偉容を誇る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ルに赴任したエルギン卿は、③ギリシアを訪れ、パルテノン神殿の遺跡から彫刻類を収集し、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、世界文化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①～④のうちから一つ選べ。

- ① スレイマン1世の時代が最盛期であった。
- ② 国教はシーア派のイスラーム教であった。
- 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＝ヘルツェゴヴィナの統治権を得た。

## Wikipedia

### スレイマン1世

スルタン・スレイマン1世(Kanuni Sultan Süleyman、オスマン語 سليمان Sulaymān, トルコ語 Süleyman, 1494年11月6日 - 1566年9月5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13回もの対外遠征を行い、数多くの軍事的成功を収めてオスマン帝国を最盛期に導いた。英語では、「**壮麗帝**(the Magnificent)」のあだ名で呼ばれ、日本ではしばしば「**スレイマン大帝**」と称される。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-Qānūnī / Kanuni)」のあだ名で知られている。

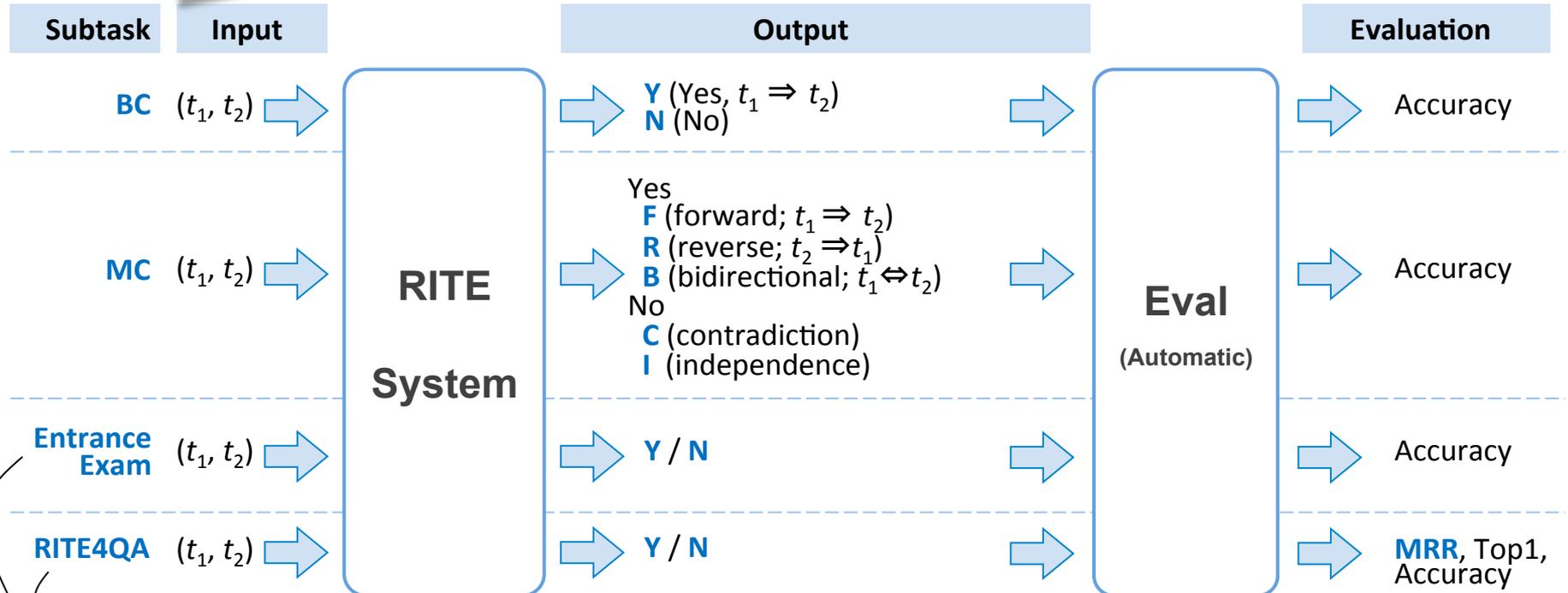
$t_1$ : スレイマン1世は数多くの軍事的成功を収めてオスマン帝国を最盛期に導いた。

$t_2$ : 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。

# RITE 評価の流れ

$t_1$ : 川端康成は「雪国」などの作品でノーベル文学賞を受賞した  
 $t_2$ : 川端康成は「雪国」の著者である

Does  $t_1$  entail (infer)  $t_2$ ?



application-oriented

# 関連研究との比較

	言語	含意	含意関係の方向	換言	矛盾	QAでの解の選択
TAC RTE (2-way)	英語	X				
TAC RTE (3-way)	英語	X			X	
MSR Paraphrase Corpus	英語			X		
CLEF AVE	英語					X
Kurohashi Lab's	日本語	X		(X)		
<b><u>NTCIR-9 RITE</u></b>	日本語 中国語 (S, T)	X	X	X	X	X
SemEval-2012 CLTE	言語 横断	X	X	X		

# RITEフォーマルラン

Number of submitted runs

Subtask	Language			Total
	JA	CS	CT	
BC	24	33	32	89
MC	10	27	22	59
Entrance Exam	18	-	-	18
RITE4QA	13	17	16	46
<b>Total</b>	<b>65</b>	<b>77</b>	<b>70</b>	<b>212</b>

- 全5カ国, 計24チームが参加

# 結果 (BC)

JA

Run	Accuracy
JAIST-01	0.5800
JAIST-02	0.5660
JAIST-03	0.5520
NTTCS-03	0.5480
LTI-03	0.5460
LTI-02	0.5420
LTI-01	0.5340
NTTCS-01	0.5320
IBM-02	0.5260
FX-02	0.5240
<i>Average</i>	<i>0.5233</i>
<i>Baseline (char overlap)</i>	<i>0.5160</i>

CS

Run	Accuracy
UIOWA-01	*0.9705
UIOWA-03	*0.9631
UIOWA-02	*0.9361
ICRC_HITSZ-03	0.7764
FudanNLP-02	0.7617
ICRC_HITSZ-02	0.7568
FudanNLP-01	0.7469
WHUTE-03	0.7371
NTU-01	0.7346
WHUTE-02	0.7322
WUST-01	0.7248
NTU-02	0.7224
NTU-03	0.7199
ZSWSL-01	0.7199
IASLD-01	0.7150
ICL-01	0.7150
<i>Average</i>	<i>0.7135</i>
<i>Baseline (char overlap)</i>	<i>0.7617</i>

CT

Run	Accuracy
UIOWA-01	*0.9078
UIOWA-02	*0.8844
IASLD-03	0.6611
IASLD-02	0.6533
III_CYUT_NTHU-02	0.6500
IASLD-01	0.6478
NTOUA-02	0.6422
<i>Average</i>	<i>0.6212</i>
<i>Baseline (char overlap)</i>	<i>0.6667</i>

\* UIOWA Systems contain manual intervention (not fully automatic).

Showing runs above the average.

# 結果 (MC)

JA

Run	Accuracy
IBM-02	0.5114
KYOTO-03	0.4841
KYOTO-02	0.4795
IBM-01	0.4545
NTTCS-03	0.4523
NTTCS-01	0.4477
IBM-03	0.4455
<i>Average</i>	<i>0.4124</i>
<i>Baseline</i> <i>(char overlap)</i>	<i>0.4682</i>

CS

Run	Accuracy
UIOWA-01	*0.8919
UIOWA-02	*0.8919
UIOWA-03	*0.8870
ICRC_HITSZ-03	0.6413
ICRC_HITSZ-02	0.6241
ZSWSL-02	0.6192
WHUTE-02	0.6093
<i>Average</i>	<i>0.5971</i>
<i>Baseline</i> <i>(char overlap)</i>	<i>0.5315</i>

CT

Run	Accuracy
UIOWA-01	*0.7867
UIOWA-02	*0.7744
UIOWA-03	*0.7244
MCU-01	0.5356
IMTKU-01	0.5222
IMTKU-02	0.5067
<i>Average</i>	<i>0.5019</i>
<i>Baseline</i> <i>(char overlap)</i>	<i>0.4885</i>

\* UIOWA Systems contain manual intervention (not fully automatic).

# 結果 (EXAM)

JA

Run	Accuracy
IBM-01	0.7217
TU-02	0.7183
TU-03	0.7042
IBM-02	0.6742
LTI-03	0.6674
KYOTO-02	0.6561
KYOTO-03	0.6561
LTI-02	0.6538
JAIST-02	0.6516
JAIST-03	0.6516
TU-01	0.6493
JAIST-01	0.6222
LTI-01	0.6018
KYOTO-01	0.5928
<i>Average</i>	<i>0.5863</i>
<i>Baseline (char overlap)</i>	<i>0.6516</i>

# 結果 (RITE4QA)

JA

Run	Acc	MRR
LTI-03	0.6753	0.2982
JAIST-01	0.5602	0.2765
JAIST-03	0.6940	0.2731
JAIST-02	0.6763	0.2604
LTI-02	0.6411	0.2563
JUCS-01	0.5954	0.2490
<i>Average</i>	<i>0.6148</i>	<i>0.2424</i>
<i>Baseline1 (char overlap)</i>	<i>0.4180</i>	<i>0.3192</i>
<i>Baseline2 (all yes)</i>	<i>0.1100</i>	<i>0.1657</i>
<i>Baseline3 (random)</i>	<i>0.5000</i>	<i>0.2320</i>
<i>Baseline4 (QA system)</i>	<i>0.1100</i>	<i>0.3917</i>
<i>Oracle</i>	<i>1.0000</i>	<i>0.5326</i>

CS

Run	Acc	MRR
UIOWA-01	*0.9010	0.4272
IMTKU-02	0.4090	0.3998
WHUTE-02	0.4876	0.3979
WHUTE-01	0.3886	0.3773
IMTKU-03	0.4716	0.3768
IMTKU-01	0.3319	0.3744
ICL-01	0.3231	0.3545
ICRC_HITSZ-01	0.6390	0.3520
WHUTE-03	0.3275	0.3494
ICRC_HITSZ-03	0.7293	0.3398
<i>Average</i>	<i>0.5192</i>	<i>0.3367</i>

CT

Run	Acc	MRR
UIOWA-01	*0.9010	0.4272
IMTKU-03	0.4003	0.3992
NTOUA-03	0.6346	0.3824
NTOUA-01	0.5459	0.3803
IMTKU-01	0.3246	0.3772
IMTKU-02	0.3392	0.3736
NTOUA-02	0.5124	0.3572
ICRC_HITSZ-01	0.6390	0.3520
ICRC_HITSZ-03	0.7293	0.3398
<i>Average</i>	<i>0.5514</i>	<i>0.3352</i>
<i>Baseline1 (char overlap)</i>	<i>0.2317</i>	<i>0.3844</i>
<i>Baseline2 (all yes)</i>	<i>0.1906</i>	<i>0.2378</i>
<i>Baseline3 (random)</i>	<i>0.5000</i>	<i>0.3454</i>
<i>Baseline4 (QA system)</i>	<i>0.1906</i>	<i>0.4852</i>
<i>Oracle</i>	<i>1.0000</i>	<i>0.5906</i>

\* UIOWA Systems contain manual intervention (not fully automatic).

# アプローチ・言語資源

- アプローチ
  - 機械学習
  - 述語項構造のマッチング (KYOTO, LTI, NTTCS, SITLP, WHUTE, ZSWSL)
  - Bilingual enrichment (JAIST, JUCS)
  - クラウドソースを用いたルールベース (UIOWA)
  - Lexical Functional Grammar (FX)
  - 文のアライメント (TU)
- 言語資源
  - Alexandria Digital Library, Baidupedia, CC-CEDICT, 日本語語彙体系, HowNet, NAIST jdic, REIKAI-SHOGAKU, Wikipedia, WordNet, etc...

# 様々な手がかり

- Overlap (character, word, bigram, trigram, head-word, POS, NE, numerical expression)
- String Similarity (Jaro distance, Jaro–Winkler distance, Jaccard Coefficient, Chebyshev Distance, Dice Coefficient, Manhattan Distance, Longest Common Subsequence, Cosine similarity, Levenshtein Edit Distance, BLEU score)
- Structural matching (predicate-argument matching, subtree matching, Tree Edit Distance)
- Verbs number mismatch
- Antonyms
- Negation / Polarity matching
- Temporal matching (5% improvement in EXAM [IBM])
- Quantification (*all, only, most, every...*)
- Quote (something just said might not be true...)

# RITE-1 まとめ

- 最も優れたシステム
  - ベースライン（文字列オーバーラップ）に勝る性能
- 様々な技術，資源の有効性が調査された
  - 教師あり機械学習，クラウドソースの活用，述語項構造のマッチング，文のアライメント，etc.
  - Alexandria Digital Library, Baidupedia, CC-CEDICT, 日本語語彙体系, HowNet, NAIST jdic, REIKAI-SHOGAKU, Wikipedia, WordNet, etc...
- 自動評価
  - Ablation studyなど，追加の評価実験が可能に
- 多くの参加者：合計24チーム

**RITE-2**

Recognizing  
Inference in  
Text@NTCIR10

# NTCIR-10 RITE-2 課題設計

## RITE-2

Recognizing  
Inference in  
Text@NTCIR10

# RITE-2のサブタスク

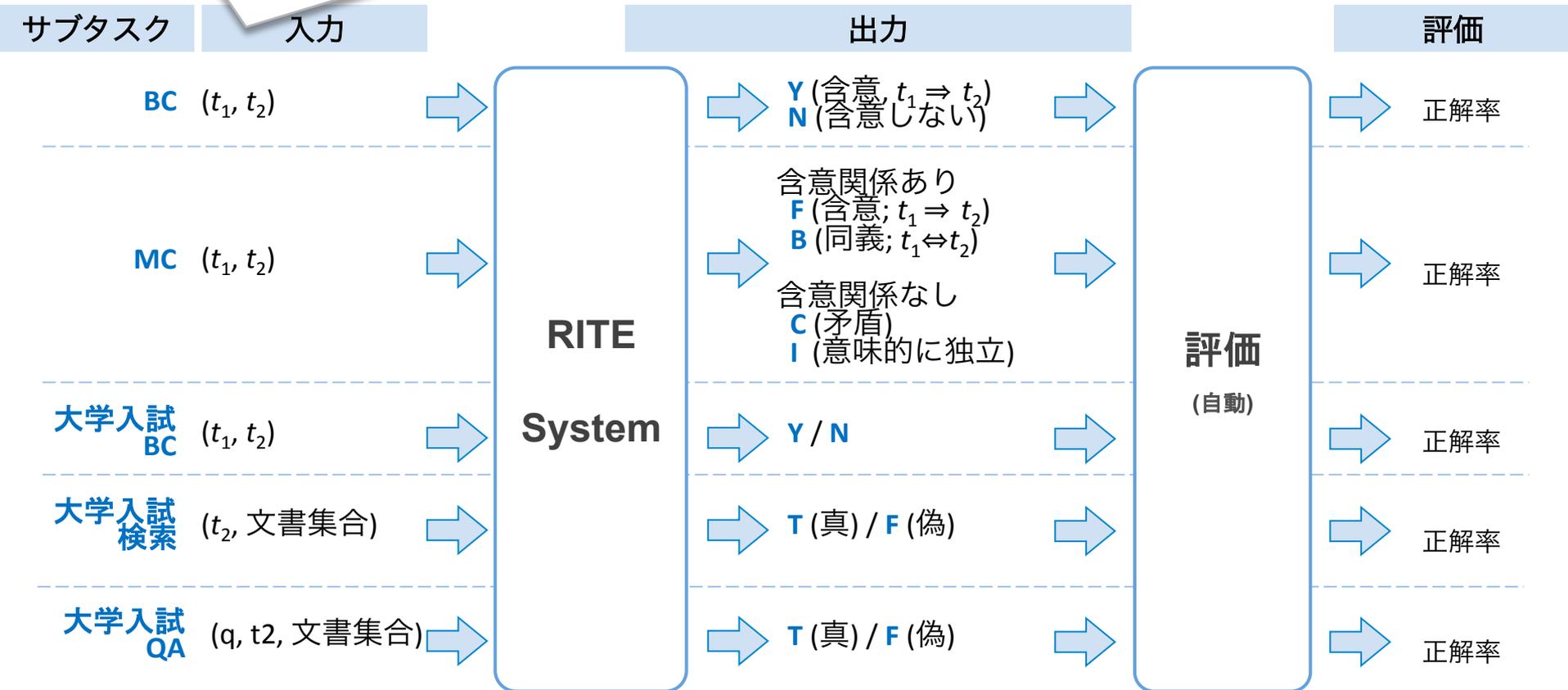
- BC, MC サブタスク: RITE-1とほぼ同様の問題設定
- 大学入試 (EXAM) サブタスク
  - BCタスク
  - Searchタスク\*
    - 仮説 $t_2$ と, Wikipediaや教科書などの文書集合が与えられたとき,  $t_2$ が真であるかどうかを出力
      - 例:  $t_2$ : コロッセウムは古代ローマを代表する建造物である。
  - QAタスク\*
    - 質問文 $q$ と仮説 $t_2$ , 文書集合が与えられた時,  $q$ の回答として $t_2$ が正しい (真である) かどうかを出力
      - $q$ : 古代ローマを代表する建造物として正しいものを, 次の $x \sim x$ のうちから一つ選べ。
      - $t_2$ : コロッセウム
- RITE4QA: 中国語(Simplified, Traditional)のみ実施予定

\*現在タスク設定を議論中.  
変更の可能性あり.

# RITE-2 評価の流れ

川端康成は「雪国」などの作品でノーベル文学賞を受賞した  
 川端康成は「雪国」の著者である

Does  $t_1$  entail (infer)  $t_2$ ?



# RITE-1 vs. RITE-2

	RITE-1	RITE-2
2値分類タスク (Binary-Class)	<ul style="list-style-type: none"><li>• 2値 (Yes or No)</li></ul>	<ul style="list-style-type: none"><li>• 2値 (Yes or No)</li></ul>
多値分類タスク (Multi-class)	<ul style="list-style-type: none"><li>• 5値分類 (換言, 前向き含意, 後向き含意, 矛盾, その他)</li></ul>	<ul style="list-style-type: none"><li>• <u>4値</u>分類 (換言, 前向き含意, 矛盾, その他)</li></ul>
大学入試タスク (Entrance Exam)	<ul style="list-style-type: none"><li>• BC (2値)</li></ul>	<ul style="list-style-type: none"><li>• BC (2値)</li><li>• <u>SEARCH</u> 仮説の真偽を, 与えられた文書集合から判断</li><li>• <u>QA</u> 質問文に対する回答 (仮説) の真偽を文書集合から判断</li></ul>
RITE4QA	<ul style="list-style-type: none"><li>• 日本語, 中国語</li></ul>	<ul style="list-style-type: none"><li>• <u>中国語のみ</u></li></ul>

**RITE-2**

Recognizing  
Inference in  
Text@NTCIR10

# RITE-2

## 今後のスケジュール

日程	イベント
2012/02 ~ 03	データ作成
2012/06/30	RITE-2 参加登録締め切り
2012/07/01	開発データリリース, 資源の公開
2012/11/14~21	フォーマルラン
2012/12/01	結果発表, テストデータリリース
2013/03/01	論文 (ドラフト) 提出締め切り
2013/05/01	論文 (カメラレディ) 提出締め切り
2013/06/18~21	NTCIR-10

**RITE-2**

Recognizing  
Inference in  
Text@NTCIR10

# ご参加をお待ちしております

RITEは含意関係認識，言い換え認識の研究だけでなく，幅広い研究分野と関連があります

- **基盤技術**：言語・意味解析，知識獲得，機械学習
- **応用技術**：情報検索，質問応答，自動要約 など

大学生から企業の研究者に至るまで，幅広く参加していただけるよう工夫をします

- 既存のリソースやツールの情報を共有

詳しくはウェブサイトにて

<http://www.cl.ecei.tohoku.ac.jp/rite2>

