Overview of NTCIR-10

Hideo Joho (University of Tsukuba)
Tetsuya Sakai (Microsoft Research Asia)
Outline

• Basic concepts
• Overview of NTCIR-10
• NTCIR-10 Evaluation Tasks
• What’s Next? NTCIR-11
• Announcements
• Wrap-up
BASIC CONCEPTS
Evaluation Forums

- Research teams gather up to solve shared problems; submit system output before deadline
- Systems evaluated and compared across teams

Share data and insights, discuss, improve
Why Evaluation Forums?

- Compete and collaborate, accelerate research
- Build large-scale test collections through collective efforts e.g. *pooling*
- Foster interdisciplinary research towards grand challenges, build new research communities

NII Testbeds and Community for Information access Research
Information Retrieval Evaluation Forums

- TREC (Text Retrieval Conference) 1992-
- NTCIR 1998/9- [sesquiannual]
- CLEF (Cross-Language Evaluation Forum) 2000-
- INEX (Initiative for the evaluation of XML retrieval) 2002-
- TRECVID 2003-
- FIRE (Forum for IR Evaluation) 2008-
- MediaEval (Benchmarking Initiative for Multimedia Evaluation) 2010-

[not exhaustive]
Jargon: Test Collection

Test collection

- Document collection
- Topic set
- (Graded) relevance assessments

Evaluation toolkit

- `trec_eval`
- `NTCIREVAL`

Comparison on common ground

- Researcher A’s search engine
- Researcher B’s search engine

Improve the state-of-the-art
Jargon: Pooling

Large-scale target corpora: exhaustive assessments not possible

Top items
Retrieved items by team A

Top items
Retrieved items by team B

Top items
Retrieved items by team Z

Pooled items

“run”

Judged relevant items

Judged nonrelevant items

An efficient way to collect right answers for evaluation
Jargon: MAP
Jargon: MAP

Mean Average Precision

Precision@2 = 1/2
Precision@3 = 2/3
Precision@4 = 3/4
Precision@5 = 4/5

\[
\text{AP} = \frac{\sum_{r} I_{\text{isrelevant}(r)} \times \text{Prec@r}}{\#\text{all relevant docs}}
\]

User’s stopping probability: uniform

Query: (I want to eat) sushi
nDCG = \[
\frac{\sum \text{System’s discounted gains}}{\sum \text{Ideal discounted gains}}
\]

Unlike AP, nDCG can utilise graded relevance assessments (widely used in Web search evaluation)
OVERVIEW OF NTCIR-10
NTCIR-10 Tasks

Core
• [CrossLink-2] Cross-lingual Link Discovery
• [Intent-2] Search intent and diversification
• [1Click-2] One-Click Access
• [RITE-2] Recognizing Inference in Text
• [SpokenDoc-2] IR for Spoken Documents

Pilot ** NEW!! ***
• [Math] Access to mathematical contents
• [MedNLP] MedNLP
# Tasks: 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) -&gt; Chinese/English/Japanese IR(2) -&gt; 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>9</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Retrieval (WEB)</td>
<td></td>
<td>7</td>
<td>11</td>
<td>7</td>
<td></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>10</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crosslingual Question Answering (CLQA)(5,6) -&gt; Advanced Crosslingual Information Access (ACLIA)(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>Opinion(6) -&gt; 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></td>
<td></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></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></td>
<td></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></td>
</tr>
<tr>
<td>Patent Translation (PAT-MT)(7,8) -&gt; 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></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></td>
</tr>
<tr>
<td>One Click Access (1CLICK) (subtask of INTENT at NTCIR-9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></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></td>
<td></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></td>
<td></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></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></td>
</tr>
</tbody>
</table>

**Number of active participants**
## Languages in NTCIR-10

<table>
<thead>
<tr>
<th>RITE</th>
<th>Patent MT</th>
<th>INTENT</th>
<th>1CLICK</th>
<th>CrossLink</th>
<th>Spoken Doc</th>
<th>Math</th>
<th>MedNLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese (simplified)</td>
<td>CE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese (traditional)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>EJ</td>
<td></td>
<td></td>
<td>CE, JE, KE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>JE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of participating teams

<table>
<thead>
<tr>
<th>Country/Region</th>
<th># of Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>54</td>
</tr>
<tr>
<td>China</td>
<td>16</td>
</tr>
<tr>
<td>Taiwan</td>
<td>11</td>
</tr>
<tr>
<td>Germany</td>
<td>6</td>
</tr>
<tr>
<td>USA</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>3</td>
</tr>
<tr>
<td>Ireland</td>
<td>2</td>
</tr>
<tr>
<td>Korea</td>
<td>2</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>105</strong></td>
</tr>
</tbody>
</table>
NTCIR-10 EVALUATION TASKS
Ranked third in the *Index of Economic Freedom* (2010),[178] Australia is the *world’s thirteenth largest economy* and has the *ninth highest per capita GDP*; higher than that of the United Kingdom, Germany, France, Canada, Japan, and the United States. The country was ranked second in the United Nations 2010 *Human Development Index* and first in *Legatum’s 2008 Prosperity Index.*[179] All of Australia’s major cities fare well in global comparative livability surveys.[180] Melbourne reached first place on *The Economist’s 2011 World’s Most Livable Cities* list, followed by Sydney, Perth, and Adelaide in sixth, eighth, and ninth place respectively.[181] Total government debt in Australia is about $190 billion.[182] Australia has among the highest house prices and some of the highest household debt levels in the world.
Given an ambiguous/underspecified query, produce a single Search Engine Result Page that satisfies different user intents!

Challenge: balancing relevance and diversity.
One Click Access (1Click-2)

Enter query

Click SEARCH button

Get all desired information

The system outputs \textit{X-string}


Go beyond the "ten-blue-link" paradigm in Web search
PatentMT

• To develop **challenging** and **significant practical** research into patent machine translation.

• To **investigate** the **performance** of state-of-the-art machine translation systems in terms of patent translations involving Japanese, English, and Chinese.

• To **compare** the effects of **different methods** of patent translation by applying them to the same test data.

• To **create** publicly-available **parallel corpora of patent documents** and human evaluations of MT results for patent information processing research.

• To **drive machine translation research**, which is an important technology for cross-lingual access of information written in unknown languages.

• The ultimate goal is **fostering scientific cooperation**.
NTCIR-10 Math Task

• NTCIR Math Task aims at exploring methods for mathematical content access through its task design and the construction of the evaluation dataset.

• English Scientific Papers

[ Formula ]
a mathematical relationship or rule expressed in symbols
(Oxford Dictionary)

In science, a formula is a concise way of expressing information, or a general relationship between quantities.
(Wikipedia)
MedNLP

■ De-identification task
Add Tags  <a>: age, <x>: sex, <t>: time

■ Complaint and diagnosis task
Add Tags  <c>: complaint, <d>: diagnosis

■ Free task

The dataset:  Japanese Dis-Charge Summaries

■ Example input data: plain text
工場に勤めている64歳の男性。<br>
2025年8月2日（来院5日前）頃から腹痛が生じるとともに、食欲不振、嘔気・嘔吐出現した。<br>
体幹は温かいが、抹消は湿潤冷汗でショック状態。明らかな運動麻痺はみられず。翌日、意識障害出現し、腎機能障害の増悪を認めて徐々に尿量低下し、8月9日18時10分に心肺停止。8月9日21時44分死亡確認。

■ Example output data: tagged text
工場に勤めている<a>64歳</a>の<x>男性</x>。
<t>2025年8月2日（来院5日前）頃から</t><c>腹痛</c>が生じるとともに、<c>食欲不振</c>、<c>嘔気</c>・<c>嘔吐出現</c>した。体幹は温かいが、抹消は<c>湿潤冷汗</c>で<c>ショック状態</c>。明らかな<c>運動麻痺</c>はみられず。<t>翌日</t>、<c>意識障害出現</c>し、<c>腎機能障害</c>の増悪を認めて徐々に<c>尿量低下</c>し、<t>8月9日18時10分</t>に<c>心肺停止</c>。<t>8月9日21時44分</t><c>死亡確認</c>。
NTCIR-10 RITE-2

Linguistic phenomena-level inference

Sentence-level inference

Multiple sentence-level inference

Pyramid of entailment recognition technology

Search

entailment?

contradiction?

entailment?

paraphrase?

contradiction?

entailment?

World knowledge

Unit Test

BC

MC

IR

QA

IE

Application oriented

Foundation oriented

RITE

RITE-2
SpokenDoc-2

- Ad-hoc Information Retrieval from lecture speeches
- Finding the **passages** including the relevant information related to a given query topic
- Query
  - Text query
  - Spoken Query (optional)
- Databases (Japanese Only)
  - CSJ (Corpus of Spontaneous Japanese)
    - 2,702 lecture speeches, 628 hours
      - New target! **Real academic meeting lectures collection**
    - Over 70 speeches from the spoken document processing workshops
Highlights of NTCIR-10 Tasks

• Diverse Challenges
  – Smart Web Access, Cross-Languages, Multimedia, Knowledge Extraction, and Domain Expertise

• Chronological Evaluation
  – Repeat the experiments on NTCIR-9 dataset with new systems at NTCIR-10

• Collaboration with other Evaluation forums
  – Sharing relevant data, insights, etc.
WHAT’S NEXT? NTCIR-11
NTCIR-11

• 11th Cycle of NTCIR
  – July 2013 – December 2014

• NTCIR-11 Conference: 9-12th Dec 2014

• 11 Task Proposals were submitted
  – Scheduled to announce the result on 268th June

• Task Proposal Presentation
  – 14:10 on 21st June (Day-4)
ANNOUNCEMENTS
Tetsuya to join General Co-Chairs

- Program Co-Chairs of NTCIR-9 and NTCIR-10
- Participated all NTCIRs!
- Task organiser of 1Click and INTENT
- Q-measure, D-measure, I-measure
Kazuaki to join Program Co-Chairs

- Prof. Kazuaki Kishida at Keio University
- IR, Clustering, Bibliometrics, Evaluation
- CLIR Task organiser at NTCIR-4, 5, 6
- Program Co-Chairs of EVIA 2011
Please join us to celebrate 10
Wrap up

• It’s been 15 years, 10th running of NTCIR
• Largest number of participants ever
• Infrastructural change made in NTCIR-9 was a huge success
• Strive to develop the communities further
• Your feedback is crucial!
Thank you and enjoy the conference!