Welcome!

NTCIR-10 Conference & EVIA 2013

Noriko Kando, Tsuneaki Kato, Douglas W. Oard, Mark Sanderson
NTCIR-10 General Chairs

19th June 2013
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

2,347 Users!

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

NTCIR-10 (2012-2013)
135 Teams Registered to Task(s)
973 Teams Registered so far
## Tasks in the NTCIR

<table>
<thead>
<tr>
<th>Domain</th>
<th>Year</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media・UGC</td>
<td>'99 '01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Community QA, Opinion Analysis</td>
</tr>
<tr>
<td>Module-Based</td>
<td>'02 '04 '05 '07 '08 '10 '11 '13</td>
<td>IR + Inference, QA + IR</td>
</tr>
<tr>
<td>Special Domain</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>MedNLP, Math Retrieval, Entrance Exam, Spoken Doc Retrieval, GeoTime Retrieval, Patent Retrieval</td>
</tr>
<tr>
<td>Questions Answering</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Any types of Qs, Dialogue, Cross-Lingual IR, Factoid, List</td>
</tr>
<tr>
<td>IE &amp; IR</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Link Discovery</td>
</tr>
<tr>
<td>Semantic</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Inference</td>
</tr>
<tr>
<td>Summarization</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Text Mining, Trend Info Visualization, Text Summarization</td>
</tr>
<tr>
<td>Interactive</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Interactive IR and Visualization</td>
</tr>
<tr>
<td>Web</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Intent mining, Diversified Search, SERPT Quality, Web Search</td>
</tr>
<tr>
<td>Cross-Lingual</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Machine Translation, Cross-Lingual IR, Non-EnglishIR</td>
</tr>
<tr>
<td>Text Retrieval</td>
<td>'01 '02 '04 '05 '07 '08 '10 '11 '13</td>
<td>Ad Hoc, IR for QA</td>
</tr>
</tbody>
</table>

Year that the conference was held, The Tasks started 18 Months before
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
About NTCIR-10

What was new?
NTCIR-9 & 10: Objectives

- Solid foundation
  - New structure
- Task diversity
  - Covers a wide context in Information Access
  - Studies rich media types
- Community-led task organisation
  - Sustainability of research
- Promotion of research resources
  - Show case in the NTCIR-9 Meeting
NTCIR-10: Structure

• General Co-Chairs
  – Noriko Kando (NII)
  – Tsuneaki Kato (The Univ of Tokyo)
  – Douglas W. Oard (Univ of Maryland)
  – Mark Sanderson (RMIT Univ.)

• Program Co-Chairs
  – Hideo Joho (Tsukuba University)
  – Tetsuya Sakai (MSR)

• Task Organisers
  – 58 researchers worldwide
  – 135 Groups Registered, 101 Active Participating Groups

• EVIA Co-Chairs
  – Ruihua Song (MSR)
  – William Webber (Univ of Maryland)
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
NTCIR-10 Program Committee

• Charles Clarke
• Kalervo Järvelin
• Hideo Joho, chair
• Gareth Jones
• Noriko Kando
• Tsuneaki Kato
• Tetsuya Sakai, chair
• Mark Sanderson
• Ian Soboroff
NTCIR-10 Data Provider/Cooperation

CMU (Clue Web 09)
Hong Kong Institute of Education
Japan Patent Office
Linguistic Data Consortium
Mainichi Newspaper
Microsoft Research
National Center for University Admissions
National Institute of Information and Communication Technology (NICT)

National Institute for Japanese Language and Linguistics (NIJL)
Sogou.com
Spoken Document Processing WG, SIG Spoken Language Processing, Information Processing Society Japan
Tokyo Shoseki
UDN.COM
Zhuhai Chilinstar

And all the organizations of the task organizers!
NTCIR-10 Conference

http://research.nii.ac.jp/ntcir/
NTCIR-10 Conference

• Conference!
• 218 Pre-Registered Participants!
• Keynote: Shashi Thakker (Google)
  Charles Clarke (U of Waterloo)
• Invited talks:
  – Ian Soboroff (TREC/TAC)
  – Gareth Jones (MedEval)
NTCIR-10 Conference Organization Committee

General Co-Chairs:
Noriko Kando, Tsuneaki Kato, Douglas W. Oard, Mark Sanderson

Accounting Chair : Hiroaki Ohshima
Exhibitions and Poster Chair : Masaharu Yoshioka
Local Arrangements Chair : Emi Ishita
Publications Chair : Kazuaki Kishida, Kazuko Kuriyama
Sponsor Chair : Tsuneaki Kato
Travel Support Chair : Takehito Utsuro
NTCIR-10 Conference Sponsors

Silver
Hitachi, Ltd.

NEC Information Systems, Ltd.

Bronze
Japan Patent Information Organization (Japio)
IR-Advanced Language Technologies, Inc.
Yahoo Japan Corporation
A9.Com / Amazon Japan
NTT Data Corporation

Special Thanks to
10 Travel Supports!
NTCIR-10

Assessors

Thanks indeed!
NTCIR-10
NTCIR Project Office

Mariko Okada
Kazuko Matsuo
Noriko Kubo
Miho Sugimoto (by June 2012)

Thanks!
NII Testbeds and Community for Information access Research

Why?

- 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)
Mid-term Vision

Datasets
- Large-Scale Reusable TCs
- Cross-lingual
- Users’ Information Seeking
  - Tasks + Genres
  - NTCIR-1, NTCIR-3

Integrating
NLP • IR
- Documents -> Answer, Information

Opinion/Attitudes
- NTCIR-6, NTCIR-9

Weaving out Knowledge

Users Context

Intension

Interaction & Visualization
- Module-based
- NTCIR-12

Interaction & Visualization
- Module-based Platform (Open Advancement)
Thank you for your participation! Enjoy!

For further enquiries, contact the NTCIR office
ntc-secretariat@nii.ac.jp

http://research.nii.ac.jp/ntcir/ntcir-10/