Lessons from Past INTENT Tasks

- About result diversification:
  - **INTENT@NTCIR9**: Top performers may be either relevance-oriented or diversity-oriented.
  - **INTENT2@NTCIR10**: Diversified ranking results significantly outperformed baseline non-diversified run.

- About the task settings:
  - SM task attracts much more participants than DR.
  - Subtopic annotation (clustering and importance voting) is rather difficult even for organizers and professional assessors.
  - Whether ordinary users are more satisfied with diversified search results remains unknown.
More User Behavior Data

- THUIR@INTENT/INTENT2: More user behavior data led to better performance.
- More raw click-through data: SogouQ has doubled its size to include click-through data collected from Sogou.com in 2011
  - 1.85GB => 3.85GB, over 40M user clicks
- More subtopic candidates generated from more recent user behavior data.
  - Search engine data provided by two major Chinese search engines will be adopted
- Find out whether more logs help improve SM/DR performance

Intent Annotation Using Logs

- SM results in pools will be clustered with click-through/pseudo RF data at first to generate preliminary candidate intent groups.
- Query frequency information will be taken into consideration during subtopic importance voting process.
- Data source: recently collected data from Sogou for Chinese SM, Bing for English/Japanese SM
  - More credible SM qrels with less annotation efforts
  - Perhaps the reusability of results could be increased
Crowd-Sourcing based Evaluation

- A search-engine-like annotation interface to collect feedback information from a relatively large number of unprofessional users (e.g. 50+ undergraduate students)
- Data collected from the interface: query, click, examination

\[ P(C_i = 1) = P(E_i = 1) P(R_i = 1) \]

- High correlation with examination
  - Preliminary results on 10 users.
  - KAPPA: 0.65, Accuracy: 0.83.
- Find out whether D# measures accords with user satisfaction

INTENT with/for Knowledge Graph

- Fuji in Wikipedia:
  - Fuji the actress
  - Fuji the Mountain
  - Fuji the apple
- Ambiguous queries: NTCIR INTENT #0205: 功夫 (kung fu), #0206: 生日快樂 (happy birthday)
- INTENT task with existing knowledge graphs such as wikipedia, freebase (e.g. THCIB and THUIS in INTENT2)
- INTENT task to help improve existing knowledge graphs
Summary of Proposal

<table>
<thead>
<tr>
<th>Number of Topics</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese: 100</td>
<td>•</td>
<td>Chinese: 100</td>
</tr>
<tr>
<td>Japanese: 100</td>
<td>•</td>
<td>Japanese: 100</td>
</tr>
<tr>
<td>English: 50</td>
<td>•</td>
<td>English: 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DR task setting</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese: ClueWeb JA</td>
<td>•</td>
<td>English: ClueWeb12-2B13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support from log analysis for annotation</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>Support from log analysis for SM annotation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crowd sourcing</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>Crowd sourcing for Chinese DR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtopic candidate</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query suggestions from Bing, Google, Sogou and Baidu</td>
<td>•</td>
<td>Query suggestions from Bing, Google, Sogou and Baidu</td>
</tr>
<tr>
<td>Candidates generated by MSRA from search engine result page</td>
<td>•</td>
<td>Candidates generated by MSRA from search engine result page</td>
</tr>
<tr>
<td>Candidates generated by THU from Sogou log data</td>
<td>•</td>
<td>Candidates generated by THU from Sogou log data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User behavior data</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SogouQ (data collected in 2008): 2GB approximately</td>
<td>•</td>
<td>SogouQ (data collected in 2008 and 2011): 4GB approximately</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DR Baseline</th>
<th>INTENT2</th>
<th>The NEW INTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese DR baseline</td>
<td>•</td>
<td>Chinese DR baseline</td>
</tr>
<tr>
<td>Japanese DR baseline</td>
<td>•</td>
<td>Japanese DR baseline</td>
</tr>
<tr>
<td>ClueWeb12-2B13 retrieval service is provided by CMU</td>
<td>•</td>
<td>ClueWeb12-2B13 retrieval service is provided by CMU</td>
</tr>
<tr>
<td>SogouT retrieval service is provided by THU/SOGOU</td>
<td>•</td>
<td>SogouT retrieval service is provided by THU/SOGOU</td>
</tr>
</tbody>
</table>

Task Mining Task (TASKMINE) Task Proposal
- As the NEW INTENT’s Subtask -

Takehiro Yamamoto
Makoto P. Kato
Hiroaki Ohshima
(Kyoto University)

June, 2013@NTCIR-10, Tokyo
Overview

Given a query, find the set of tasks to achieve it

Lose Weight

- Do physical exercise
- Control calories intake
- Take diet pills
- Have Surgery

Overview

Given a query, find the set of tasks to achieve it

Move to a new house

- Stop gas
- Stop electricity
- Submit notification of move out
- Submit notification of move in
Why Important?

**Demand for Supporting Complex Search Task**

- Domains such as Health, Education, Travel are tends to be complex tasks, and should be supported
  [Donate *et al.* WWW2010]

- Subgoal mining from Sponsored search data
  [Yamamoto *et al.* CIKM2012]

- Grouping relevant tasks by ODP Categories
  [Hassan *et al.* CIKM2012]

---

**Our Task Model**

*Task* is a set of *actions* that approach to the *goal state*

---

- Lose Weight

  - Initial State
    - Do physical exercise
    - Have surgery
    - Take diet pills

  - Goal State
    - slim
Our Task Model

Task is a set of actions that approach to the goal state

Possible Research Issues

- IR People
  - iUnit (information unit) based retrieval (1CLICK-2)
  - Subtopic mining techniques (INTENT-2)

- NLP People
  - Procedural information extraction
  - Entailment relation extraction (RITE), especially verb entailment

- QA People

- And more
Task Plans

• Task Settings
  • Manually prepare 100 queries (tasks)
  • Participants’ output format:
    Ranked list of tasks with their temporal orders

• Language
  • Japanese

• Evaluation Method
  • Manually Create gold-standards
    • Annotate task relationships
      (temporal, part-of, is-a, hierarchical, etc.)
    • Apply new evaluation metrics (hierarchical, and more)

Summary

TASKMINE as INTENT-3 Subtask
Given a query, find the set of tasks to achieve it

Research Challenges:

• How to Retrieve tasks from the Web?

• How to Model relationships among tasks?
Thank you