Overview of the NTCIR-12 MobileClick-2 Task

Makoto P. Kato (Kyoto U.), Tetsuya Sakai (Waseda U.), Takehiro Yamamoto (Kyoto U.), Virgil Pavlu (Northeastern U.), Hajime Morita (Kyoto U.), and Sumio Fujita (Yahoo Japan Corporation)
Let's see the current mobile search
What's NTCIR?

Your Search Stats

Clicks: 2
Time: 00:31
When is the deadline of NTCIR?
30 sec are too long for mobile users.
Let's do better!
**iUnit Summarization Subtask**

- Given a query, a set of iUnits, and a set of intents, generate a **two-layered summary**

  **Input:** Query
  - NTCIR

  **Input:** iUnit set
  - A series of evaluation workshops
  - Designed to enhance IA research

  **Input:** Intents
  - News
  - Schedule

  **Output:** Two-layered summary
  - The NTCIR Workshop is a series of evaluation workshops designed to enhance research in information access technologies including information retrieval, summarization, extraction, question answering, etc.

  **Evaluation metric designed for mobile information access**
  - M-measure
  - 0.5

**Challenge**

Lay out iUnits so that any types of users can be immediately satisfied
What's NTCIR?

The NTCIR Workshop is a series of evaluation workshops designed to enhance research in information access technologies including information retrieval, summarization, extraction, question answering, etc.

NTCIR-12

Held on June 9(Tue) - 12(Fri), 2016 at National Center of Sciences, Tokyo, Japan

NTCIR-12 News

NTCIR-12 Schedule

NTCIR-12 Tasks

Your Search Stats

Clicks: 0
Time: 00:03
When is the deadline of NTCIR?

NTCIR-12 Schedule

- 20/Jan./2016: Task Registration Due
- 06/Jan./2016: Document Set Release
- Jan.-May/2016: Dry Run
- Mar.-July/2016: Formal Run
- 01/Aug./2016: Evaluation Results Due
- 01/Aug./2016: Task overview release
- 15/Sep./2016: Paper submission Due
- 01/Nov./2016: All paper Due
- 09-12/Dec./2016: NTCIR-11 Conference

Your Search Stats
Clicks: 1
Time: 00:15
Is This Interface So Different from That of the Current Search Engine?

No. Thus, using this interface is not very unrealistic.
Goal of MobileClick

Provide Direct and Immediate Mobile Information Access
The NTCIR Workshop is a series of evaluation workshops designed to enhance research in information access technologies including information retrieval, summarization, extraction, question answering, etc.

**News**

**Schedule**

- 20/Jan./2016: Task Registration Due
- 06/Jan./2016: Document Set Release
- Jan.-May/2016: Dry Run
- Mar.-July/2016: Formal Run
- 01/Aug./2016: Evaluation Results Due
- 01/Aug./2016: Task overview release
- 15/Sep./2016: Paper submission Due
- 01/Nov./2016: All paper Due
- 09-12/Dec./2016: NTCIR-11 Conference

**Tasks**

1. A series of evaluation workshops
2. Task Registration Due 20/Jun./2016
3. Designed to enhance IA research

---

Two Subtasks

- Query
- iUnit Ranking Subtask
- Importance of iUnits
- iUnit Summarization Subtask
- Two-layered Summary

---

NTCIR
Given a query and a set of iUnits, rank them based on their estimated importance.

Note: iUnits are information pieces relevant to a given query.

**Input:** Query

NTCIR

**Input:** iUnit set

<table>
<thead>
<tr>
<th>iUnit</th>
<th>1</th>
<th>A series of evaluation workshops</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>Task Registration Due 20/Jun./2016</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Designed to enhance IA research</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Output:** iUnit list

**Challenge**

Predict the importance of strings rather than documents.
Given a query, a set of iUnits, and a set of intents, generate a two-layered summary.

Input: Query
Input: iUnit set
Input: Intents

Output: Two-layered summary

The NTCIR Workshop is a series of evaluation workshops designed to enhance research in information access technologies including information retrieval, summarization, extraction, question answering, etc.

2nd layer

News
Schedule

Evaluation metric designed for mobile information access

M-measure 0.5

Challenge
Lay out iUnits so that any types of users can be immediately satisfied
Two-layered Summary in Action
DATA
Overview of Data

Queries
- napoleon

Web search

Documents

Extraction

iUnits

- Born on the island of Corsica
- Defeated at the Battle of Waterloo
- Established legal equality and religious toleration an innovator

Intents
- Achievement
- Skill
- Career

Clustering

iUnit summarization

Input

iUnit ranking
Queries and Documents

• Queries
  – 100 English/Japanese queries
  – Most of which were ambiguous/underspecified
  – **Selected from five categories:**
    celebrity, location, definition, and QA (similar to NTCIR 1CLICK-2)

Examples

<table>
<thead>
<tr>
<th>CELEBRITY</th>
<th>LOCATION</th>
<th>DEFINITION</th>
<th>QA</th>
</tr>
</thead>
<tbody>
<tr>
<td>hulk hogan</td>
<td>bank adelanto</td>
<td>bitcoin</td>
<td>what is mirror made of</td>
</tr>
<tr>
<td>brunon mars</td>
<td>cafe killeen</td>
<td>divers disease</td>
<td>how to cook coleslaw</td>
</tr>
<tr>
<td>sharon stone</td>
<td>cincinnati art museum</td>
<td>windows 7</td>
<td>role of animal tail</td>
</tr>
</tbody>
</table>

• Documents
  – 500 commercial search engine results for each query
  – From which iUnits were extracted
• **Definition**
  – Atomic information pieces relevant to a given query

• **The number of iUnits**
  – **2,317** (23.8 iUnits per query) for English
  – **4,169** (41.7 iUnits per query) for Japanese

**Examples of iUnits for query “Napoleon”**

<table>
<thead>
<tr>
<th>Born on the island of Corsica</th>
<th>General of the Army of Italy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defeated at the Battle of Waterloo</td>
<td>One of the most controversial political figures won at the Battle of Wagram</td>
</tr>
<tr>
<td>Established legal equality and religious tolerance an innovator</td>
<td>Baptised as a Catholic</td>
</tr>
<tr>
<td>Absent during Peninsular War</td>
<td>Cut off European trade with Britain</td>
</tr>
</tbody>
</table>
iUnit Extractor

- https://addons.mozilla.org/ja/firefox/addon/iunit-extractor/
  – Useful for nugget extraction, etc.

Napoleon was born in Corsica to a relatively modest family of noble Tuscan ancestry. He supported the French Revolution from the outset in 1789 while serving in the French army, and he tried to spread its ideals to his native Corsica, only to find himself banished from the island by the authorities in 1793. Two years later, he saved the governing French Directory by firing on royalist insurgents with cannons. The Directory rewarded Napoleon by giving him command of the Army of Italy at age 26. In April 1796, he began his first military campaign against the Austrians and their
An intent can be defined as
- A specific interpretation of an ambiguous query ("Mac OS" and "car brand" for "jaguar"),
- An aspect of a faceted query ("windows 8" and "windows 10" for "windows")

Obtained by clustering iUnits

<table>
<thead>
<tr>
<th>iUnits</th>
<th>Intents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Born on the island of Corsica</td>
<td></td>
</tr>
<tr>
<td>Defeated at the Battle of Waterloo</td>
<td></td>
</tr>
<tr>
<td>Established legal equality and religious</td>
<td></td>
</tr>
<tr>
<td>toleration an innovator</td>
<td></td>
</tr>
<tr>
<td>Absent during Peninsular War</td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td></td>
</tr>
<tr>
<td>Skill</td>
<td></td>
</tr>
<tr>
<td>Career</td>
<td></td>
</tr>
</tbody>
</table>
Yahoo Search Query Data

• Queries and their statistics related to our training and test query sets were provided by Yahoo Japan Corporation
  – Co-Click Queries
    Queries that share clicks with the query sets
  – Co-topic Queries
    Queries that include a query string in the query sets
  – Co-Session Queries
    Queries that appeared in the same session as the query sets
• Used by participants for ranking iUnits and generating two-layered summaries
Overview of Data (Repeated)

Queries
- napoleon

Documents
- Web search

Extraction

iUnits
- Born on the island of Corsica
- Defeated at the Battle of Waterloo
- Established legal equality and religious toleration an innovator

Intents
- Achievement
- Skill
- Career

Clustering

iUnit summarization

Input

iUnit ranking
EVALUATION
Importance of iUnits in terms of an intent was given by two assessors at a 5-point scale:

- An iUnit is more important if it is more necessary for more users who are interested in the intent.
- The inter-rater agreement: 0.556 (weighted kappa).

### Per-intent iUnit Importance

**In terms of intent “Definition”**

<table>
<thead>
<tr>
<th>iUnit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A series of evaluation workshops</td>
<td>5</td>
</tr>
<tr>
<td>Task Registration Due 20/Jun./2016</td>
<td>3</td>
</tr>
</tbody>
</table>

**In terms of intent “Schedule”**

<table>
<thead>
<tr>
<th>iUnit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A series of evaluation workshops</td>
<td>2</td>
</tr>
<tr>
<td>Task Registration Due 20/Jun./2016</td>
<td>5</td>
</tr>
</tbody>
</table>
• Intent probability was estimated by voting
  – $P(i|q)$: probability of having intent $i$ given $q$
  – 10 assessors voted for one or more intents for a given query

### Intent Voting

<table>
<thead>
<tr>
<th>Intent</th>
<th># of votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>4</td>
</tr>
<tr>
<td>Schedule</td>
<td>3</td>
</tr>
<tr>
<td>Tasks</td>
<td>3</td>
</tr>
</tbody>
</table>

### Intent Probability

<table>
<thead>
<tr>
<th>Intent</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>0.4</td>
</tr>
<tr>
<td>Schedule</td>
<td>0.3</td>
</tr>
<tr>
<td>Tasks</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Evaluation of iUnit Ranking

- Evaluated in the same way as ad-hoc retrieval

Output: iUnit list

<table>
<thead>
<tr>
<th>iUnit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A series of evaluation workshops</td>
</tr>
<tr>
<td>2</td>
<td>Task Registration Due ...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iUnit</th>
<th>GI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A series of evaluation workshops</td>
</tr>
<tr>
<td>2</td>
<td>Task Registration Due ...</td>
</tr>
</tbody>
</table>

nDCG@10 Q-measure = 0.87

Per-intent iUnit Importance

In terms of intent “Schedule”

<table>
<thead>
<tr>
<th>iUnit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A series of evaluation workshops</td>
<td>2</td>
</tr>
<tr>
<td>Task Registration Due 20/Jun./2016</td>
<td>5</td>
</tr>
</tbody>
</table>

Intent Probability

<table>
<thead>
<tr>
<th>Intent</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>0.4</td>
</tr>
<tr>
<td>Schedule</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Global Importance

\[ G(u) = \sum_{i \in I_q} P(i|q) g_i(u) \]

- \( P(i|q) \): intent probability
- \( g_i(u) \): per-intent importance
- \( I_q \): intents for query \( q \)

Output:

<table>
<thead>
<tr>
<th>iUnit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A series of evaluation workshops</td>
<td>3.8</td>
</tr>
<tr>
<td>Task Registration Due 20/Jun./2016</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Evaluation of iUnit Summarization (Single-layer Case)

- Consider single-layered summary evaluation
- **U-measure** [Sakai and Dou. SIGIR2013]
  - Higher if more important iUnits appear earlier

\[
U = \sum_{r=1}^{r} G(u_r) \left(1 - \frac{\text{pos}(u_r)}{L}\right)
\]

\( u_r \): r-th iUnit
\( G(u) \): importance of u
\( \text{pos}(u) \): offset of u from the beginning

\( L \): patience parameter

Create a list of iUnits by assuming that users read text from left to right, from top to bottom.
M-measure

- **M-measure**
  - Expectation of U-measure over multiple *trailtexts*
  
  \[ M = \sum_t P(t)U(t) \]
  
  - *\( P(t) \): probability of trailtext \( t \)*
  - *\( U(t) \): U-measure of trailtext \( t \)*

- **Generate trailtexts by assuming that**
  - Users read a summary from the top of the first layer
  - Users click on an intent if they are interested in it

**First-layer**

- \( u_1 \)
- \( u_2 \)
- \( u_3 \)
- \( l_1 \) (intent 1)

**Second layer**

- \( u_4 \)

**Trailtext**

- User interested in Intent 1 *\( P(i_1|q) \)*
  
  - \( u_1 \)
  - \( u_2 \)
  - \( u_3 \)
  - \( u_4 \)

- User interested in Intent 2 *\( P(i_2|q) \)*
  
  - \( u_1 \)
  - \( u_2 \)
  - \( u_3 \)
Compute the expectation of U-measure

Evaluation of iUnit Summarization (Two-layer Case)

First layer

\[ u_1 \quad u_2 \]
\[ u_3 \]

Second layer

\[ l_1 \] (intent 1)
\[ u_4 \quad u_5 \]

\[ l_2 \] (intent 2)
\[ u_6 \]

Trailtext \( t \) (reading path)

\[ u_1 \quad u_2 \quad u_3 \]
\[ u_4 \quad u_5 \]

\[ u_1 \quad u_2 \quad u_3 \]
\[ u_6 \]

Because trailtext \( t_2 \) is read by users interested in \( i_2 \)

\[ P(t_1) = P(i_1 | q) = 0.75 \]

\[ P(t_2) = P(i_2 | q) = 0.25 \]

\[ M = \sum_t P(t)U(t) \]

\[ M = 0.36 \]

\[ U = 0.44 \]

\[ U = 0.12 \]
RESULTS
Submitted runs showed similar performance (a few statistically significant differences)
UHYG, YJST, and rsrch significantly outperformed the baseline method.
TITEC and YJST are the top and are not statistically distinguishable, but did not significantly outperform the best baseline.
YJST and UHYG significantly outperformed the baseline, and are not statistically distinguishable.
Approaches of Participants

Please come to our session! (DAY-3 (Thu) 9:00 – 10:30)
MobileClick tool available at https://github.com/mpkato/mobileclick

```
mobileclick

mobileclick provides baseline metrics for http://www.mobileclick.org/

Requirements

Minimum requirements:
- Python 2.7
- NumPy
- nltk
- BeautifulSoup

Requirements for Japanese runs:
- mecab-python

Download MobileClick data (Please sign up at http://www.mobileclick.org/):

```
$ mobileclick_download_data
Please input the email and password for http://www.mobileclick.org
Email: <Your email address>
Password: <Your password>
```

Replicate the random iUnit ranking baseline:

```
$ mobileclick_random_ranking_method --runname random_ranking_method \
--query data/MC2-training/en/1C2-E-queries.tsv \n--iunit data/MC2-training/en/1C2-E-iunits.tsv \n--indexdir data/MC2-training-documents/1C2-E.INDX \n--pagedir data/MC2-training-documents/1C2-E.HTML
```

Replicate the LM-based iUnit ranking baseline:

```
$ mobileclick_language_model \n--query data/MC2-training/en/1C2-E-queries.tsv \n--iunit data/MC2-training/en/1C2-E-iunits.tsv \n--indexdir data/MC2-training-documents/1C2-E.INDX \n--pagedir data/MC2-training-documents/1C2-E.HTML \n--language english
```

1 line for downloading the data
5 lines to generate baseline results
Q. When can we get our evaluation result?
A. Right after you submit your run!

- Evaluation for test queries started from **Nov 2015**
  - Participants were allowed to submit a run per week
Leader Board Timeline

- Evaluation system released
- Test data released
- Consistent growth
MobileClick-2, a core task at NTCIR-12, aims to provide direct and immediate mobile information access, by automatically generating a concise two-layered summary for a given query.

**What is the task?**
There are two subtasks: iUnit ranking and iUnit summarization. You may participate in either one of the two, or both. Please take a look at [task page](#).

**How to participate?**
You can register at the NTCIR online registration system. Please also make an account of this website for getting datasets and more information.

**How to get started?**
After logging in, you can download data and upload results for getting evaluation scores. Please read guideline before submission.

---

**Evaluation system for the test data is now running!** (Submission deadline: Feb. 4, 2016)

**Number of submissions:** 455
**Number of users:** 52
**Number of participating groups:** 22

---

**Announcements**

**Evaluation System for Test Data Launched**
Please submit your runs for the test data and get evaluation results! Please check guideline before submission.

**MobileClick-2 Tools Ver0.2 Released**
Python package “mobileclick” is now ready for test runs.
## Latest Leader Board

### English iUnit Ranking (Test)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group ID</th>
<th>Score</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TITEC</td>
<td>0.9003</td>
<td>2016-02-04 06:40:47 UTC</td>
</tr>
<tr>
<td>2</td>
<td>UHYG</td>
<td>0.8994</td>
<td>2016-01-06 05:34:43 UTC</td>
</tr>
<tr>
<td>3</td>
<td>ORG</td>
<td>0.8975</td>
<td>2015-10-31 15:03:26 UTC</td>
</tr>
<tr>
<td>4</td>
<td>RISAR</td>
<td>0.8972</td>
<td>2016-01-29 04:13:47 UTC</td>
</tr>
<tr>
<td>5</td>
<td>RISAR</td>
<td>0.8962</td>
<td>2016-02-05 04:19:19 UTC</td>
</tr>
</tbody>
</table>

### English iUnit Summarization (Test)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group ID</th>
<th>Score</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TITEC</td>
<td>18.2596</td>
<td>2016-02-04 22:33:51 UTC</td>
</tr>
<tr>
<td>2</td>
<td>ORG</td>
<td>16.8975</td>
<td>2015-10-31 15:03:56 UTC</td>
</tr>
<tr>
<td>3</td>
<td>RISAR</td>
<td>16.047</td>
<td>2015-02-05 06:27:34 UTC</td>
</tr>
<tr>
<td>4</td>
<td>ORG</td>
<td>14.1051</td>
<td>2015-10-31 15:03:30 UTC</td>
</tr>
<tr>
<td>5</td>
<td>ORG</td>
<td>13.2669</td>
<td>2015-10-31 15:03:41 UTC</td>
</tr>
</tbody>
</table>

### Japanese iUnit Ranking (Test)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group ID</th>
<th>Score</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UHYG</td>
<td>0.8388</td>
<td>2015-11-09 09:30:18 UTC</td>
</tr>
<tr>
<td>2</td>
<td>UHYG</td>
<td>0.8123</td>
<td>2016-02-02 03:24:36 UTC</td>
</tr>
<tr>
<td>3</td>
<td>ORG</td>
<td>0.7411</td>
<td>2015-10-31 15:03:44 UTC</td>
</tr>
<tr>
<td>4</td>
<td>ORG</td>
<td>0.7269</td>
<td>2015-10-31 15:10:56 UTC</td>
</tr>
</tbody>
</table>

### Japanese iUnit Summarization (Test)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group ID</th>
<th>Score</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UHYG</td>
<td>22.8342</td>
<td>2016-01-22 09:54:39 UTC</td>
</tr>
<tr>
<td>2</td>
<td>UHYG</td>
<td>21.1107</td>
<td>2016-02-01 03:25:58 UTC</td>
</tr>
<tr>
<td>3</td>
<td>ORG</td>
<td>17.4376</td>
<td>2015-10-31 15:10:06 UTC</td>
</tr>
<tr>
<td>4</td>
<td>ORG</td>
<td>15.0373</td>
<td>2015-10-31 15:09:39 UTC</td>
</tr>
<tr>
<td>5</td>
<td>ORG</td>
<td>12.799</td>
<td>2015-10-31 15:09:52 UTC</td>
</tr>
</tbody>
</table>
MobileClick-1
• No team outperformed the baseline
• 4 teams participated
• 14 runs were submitted

MobileClick-2
• Statistically significant differences
• 11 teams participated
• 66 runs were submitted
• **Goal of MobileClick:**
  Provide direct and immediate mobile information access

• **Subtasks:**
  – iUnit ranking
  – iUnit summarization

• **Results:**
  – 11 teams submitted 66 runs
  – Participants outperformed the baseline in all the subtasks
  – Some teams showed significant improvement

• **Acknowledgements**
  – Yahoo Japan Corporation
  – Wider Planet