# 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.),
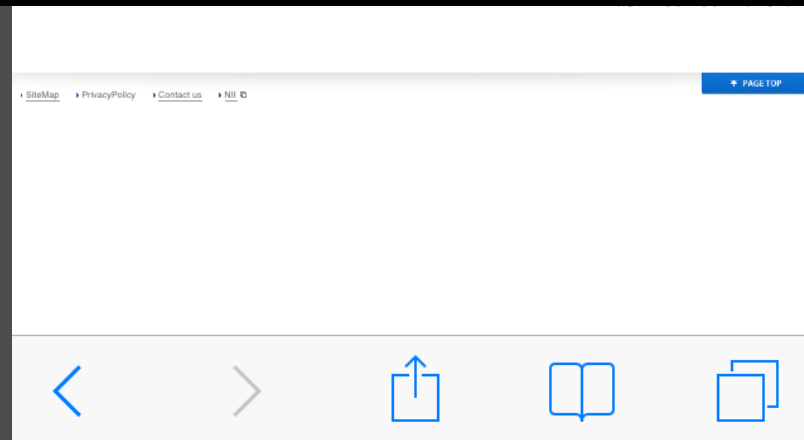Hajime Morita (Kyoto U.), and Sumio Fujita (Yahoo Japan Corporation)

# Let's see the current mobile search
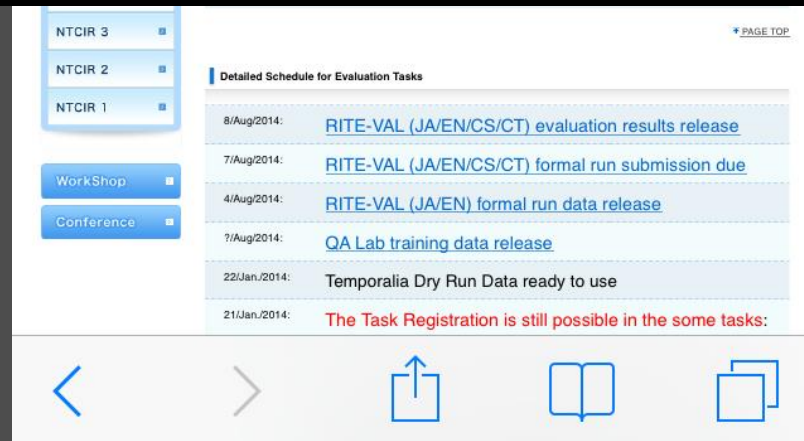
# 30sec are too long for mobile users

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**

| iUnit |
| --- |
| A series of evaluation workshops |
| Designed to enhance IA research |
| ... |

**Input: Intents**

| 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 ➡

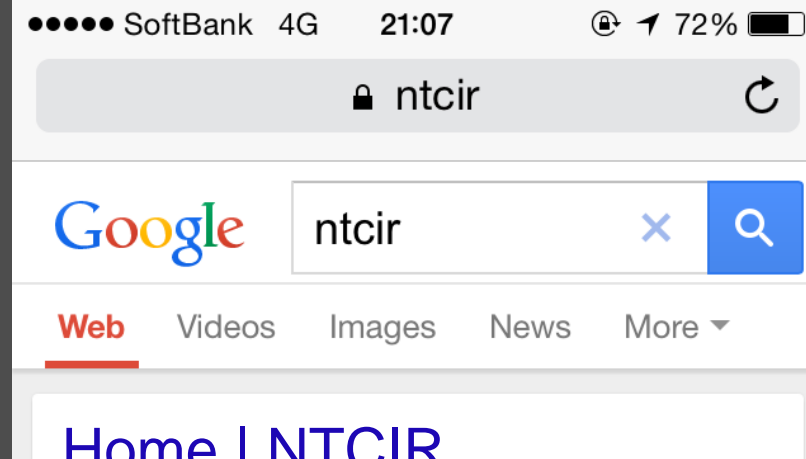| | |
| --- | --- |
| 20/Jan./2016: | Task Registration Due |
| 06/Jan./2016: | Document Set Release |
| Jan.-May/2016: | Dry Run |
| Mar.-July/2016: | Formal Run |
| 01/Aug./2016: | Evaluation Results Due |
| 01/Aug./2016: | Task overview release |
| 15/Sep./2016: | Paper submission Due |

**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**

NTCIRのホーム
research.nii.ac.jp/**ntcir**/index-ja.html ▾
**NTCIR**-12 カンファレンス & EVIA 2016 会議参加登録受付を開始しました（Online登録締切： 5月27
日、 早期登録締切：5月10日）... 2015.07.14: **NTCIR**-12 QA Lab日本語タスクでは、Phase 1の論述問題
のシステム解答について、予備校講師の方々による ...

NTCIR-12
2016.04.12 NTCIR-12カンファレンス
の基調講演ページを公開しました ...

NTCIR-11
NTCIR-11追加パイロットタスク募集
は、2014年1月31日に締切ました ...

NTCIR-10
NTCIR-10 カンファレンス:2013年6月
18-21日 東京 学術総合 ...

NTCIR-11 カンファレンス
第11回 NTCIR カンファレンス.
NTCIR-11 カンファレンス

データ
NTCIRでこれまでに構築したテスト
コレクション（実験用データ ...

会議参加者募集
会議参加者募集. NTCIR-11 カンファ
レンス 情報アクセス技術の評価 ...

nii.ac.jp からの検索結果 »

Home | 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
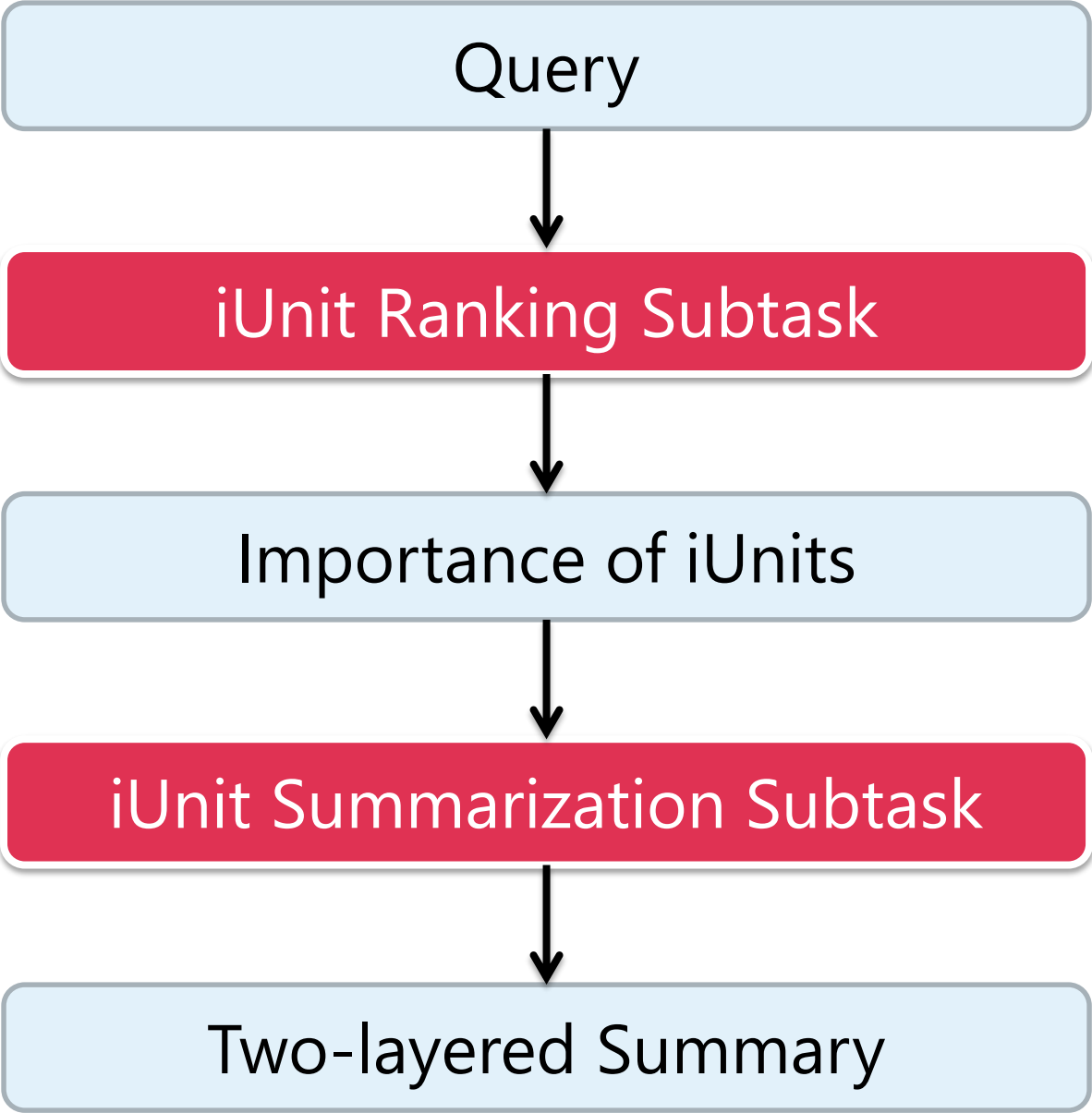
NTCIR-12 News

NTCIR-12 Schedule

NTCIR-12 Tasks

# No. Thus, using this interface is not very unrealistic.

# Goal of MobileClick

## Provide Direct and Immediate Mobile Information Access

# SUBTASKS

# Two Subtasks

Query

↓

**iUnit Ranking Subtask**

↓

Importance of iUnits

↓

**iUnit Summarization Subtask**

↓

Two-layered Summary

NTCIR

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

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
Tasks

2nd layer
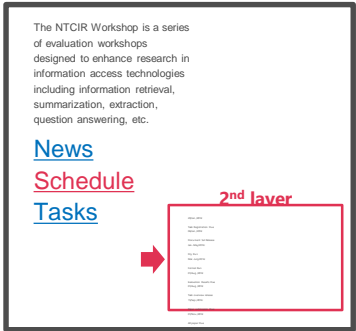
# iUnit Ranking Subtask

- ## 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**

| iUnit |
|---|
| A series of evaluation workshops |
| Designed to enhance IA research |
| Task Registration Due 20/Jun./2016 |

Output: **iUnit list**

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

nDCG
**0.5**

**Challenge**
# Predict the importance of strings rather than documents

# 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**

| iUnit |
| --- |
| A series of evaluation workshops |
| Designed to enhance IA research |
| ... |

Input: **Intents**

| 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.

News
Schedule ➡

**2ⁿᵈ layer**

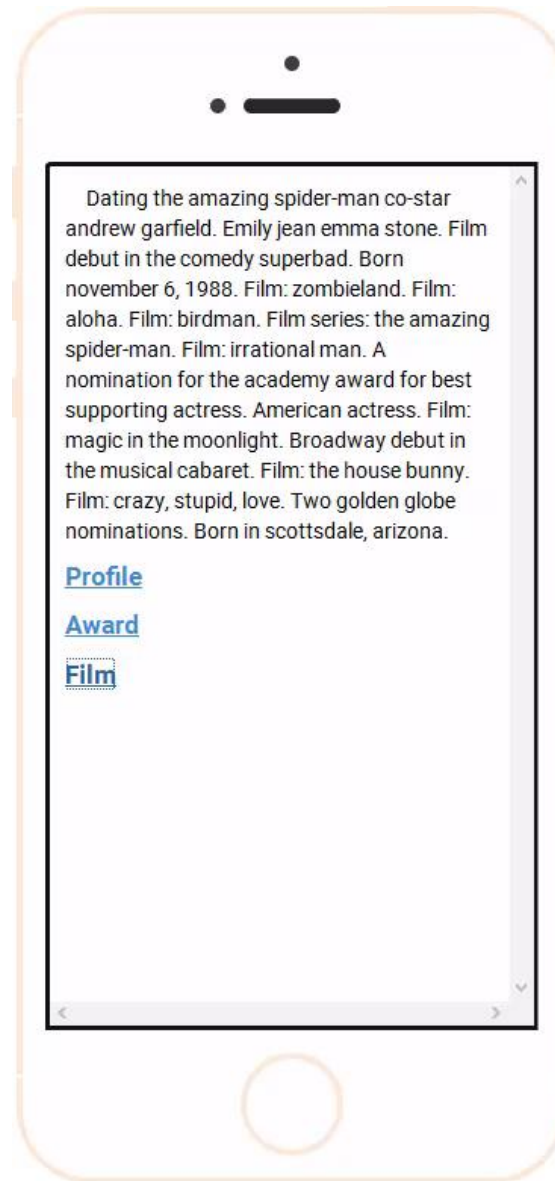| 20/Jan./2016: | Task Registration Due |
| --- | --- |
| 06/Jan./2016: | Document Set Release |
| Jan.-May/2016: | Dry Run |
| Mar.-July/2016: | Formal Run |
| 01/Aug./2016: | Evaluation Results Due |
| 01/Aug./2016: | Task overview release |
| 15/Sep./2016: | Paper submission Due |

**Evaluation metric designed for mobile information access**

## M-measure
### 0.5

**Challenge**
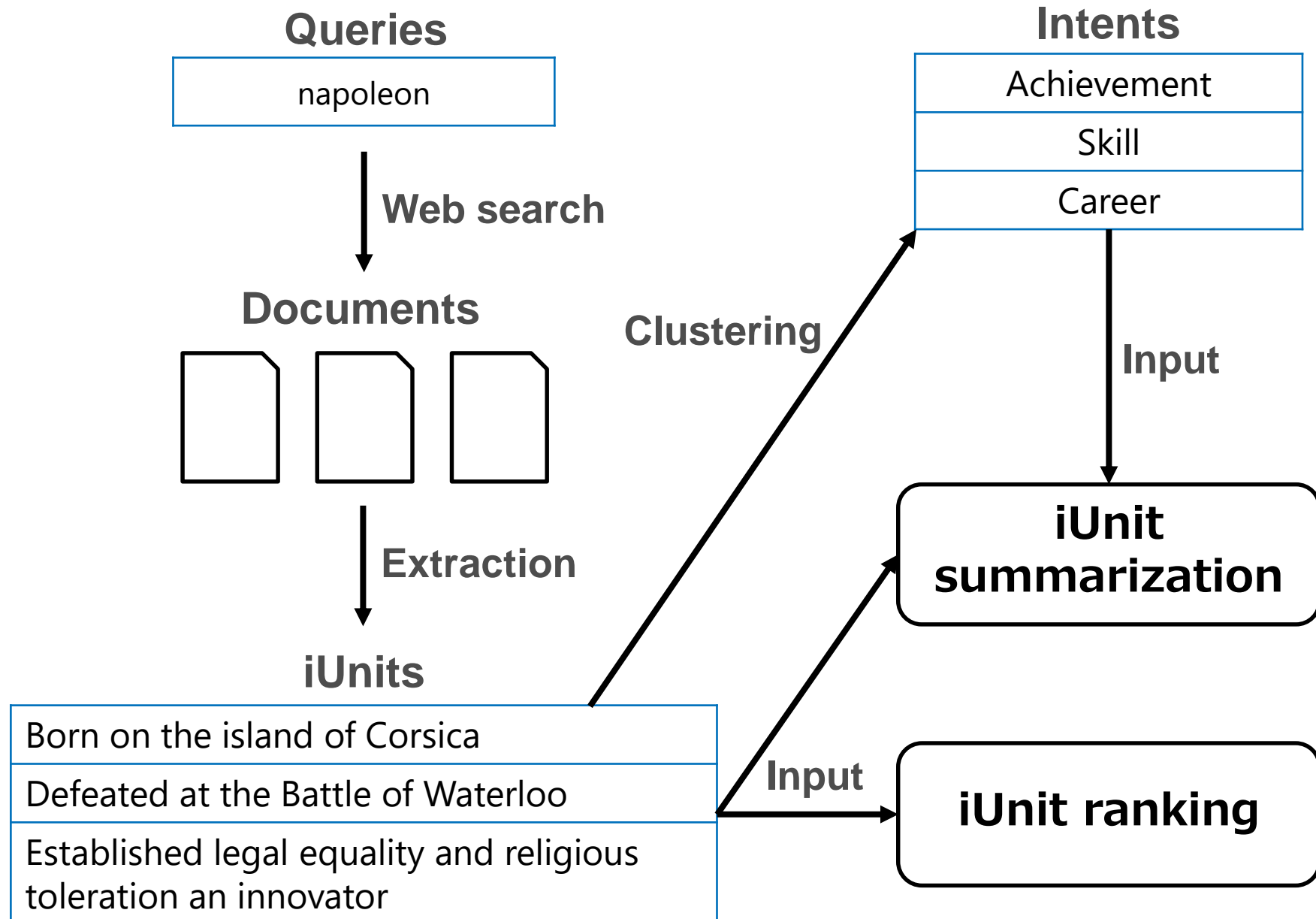## Lay out iUnits so that any types of users can be immediately satisfied

# Two-layered Summary in Action

Dating the amazing spider-man co-star andrew garfield. Emily jean emma stone. Film debut in the comedy superbad. Born november 6, 1988. Film: zombieland. Film: aloha. Film: birdman. Film series: the amazing spider-man. Film: irrational man. A nomination for the academy award for best supporting actress. American actress. Film: magic in the moonlight. Broadway debut in the musical cabaret. Film: the house bunny. Film: crazy, stupid, love. Two golden globe nominations. Born in scottsdale, arizona.

Profile

Award

Film

# DATA

# Overview of Data

**Queries**

| napoleon |
|---|

**Intents**

| Achievement |
|---|
| Skill |
| Career |

**Web search**

**Documents**

**Clustering**

**Input**

**Extraction**

**iUnit summarization**

**iUnits**

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

**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

| CELEBRITY | LOCATION | DEFINITION | QA |
|---|---|---|---|
| hulk hogan | bank adelanto | bitcoin | what is mirror made of |
| bruno mars | cafe killeen | divers disease | how to cook coleslaw |
| sharon stone | cincinnati art museum | windows 7 | role of animal tail |

- **Documents**
  - 500 commercial search engine results for each query
  - From which iUnits were extracted

# iUnits

- **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"**

| | |
|---|---|
| Born on the island of Corsica | General of the Army of Italy |
| Defeated at the Battle of Waterloo | One of the most controversial political figures won at the Battle of Wagram |
| Established legal equality and religious toleration an innovator | Baptised as a Catholic |
| Absent during Peninsular War | Cut off European trade with Britain |

# iUnit Extractor

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

- **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**

**iUnits**

**Intents**

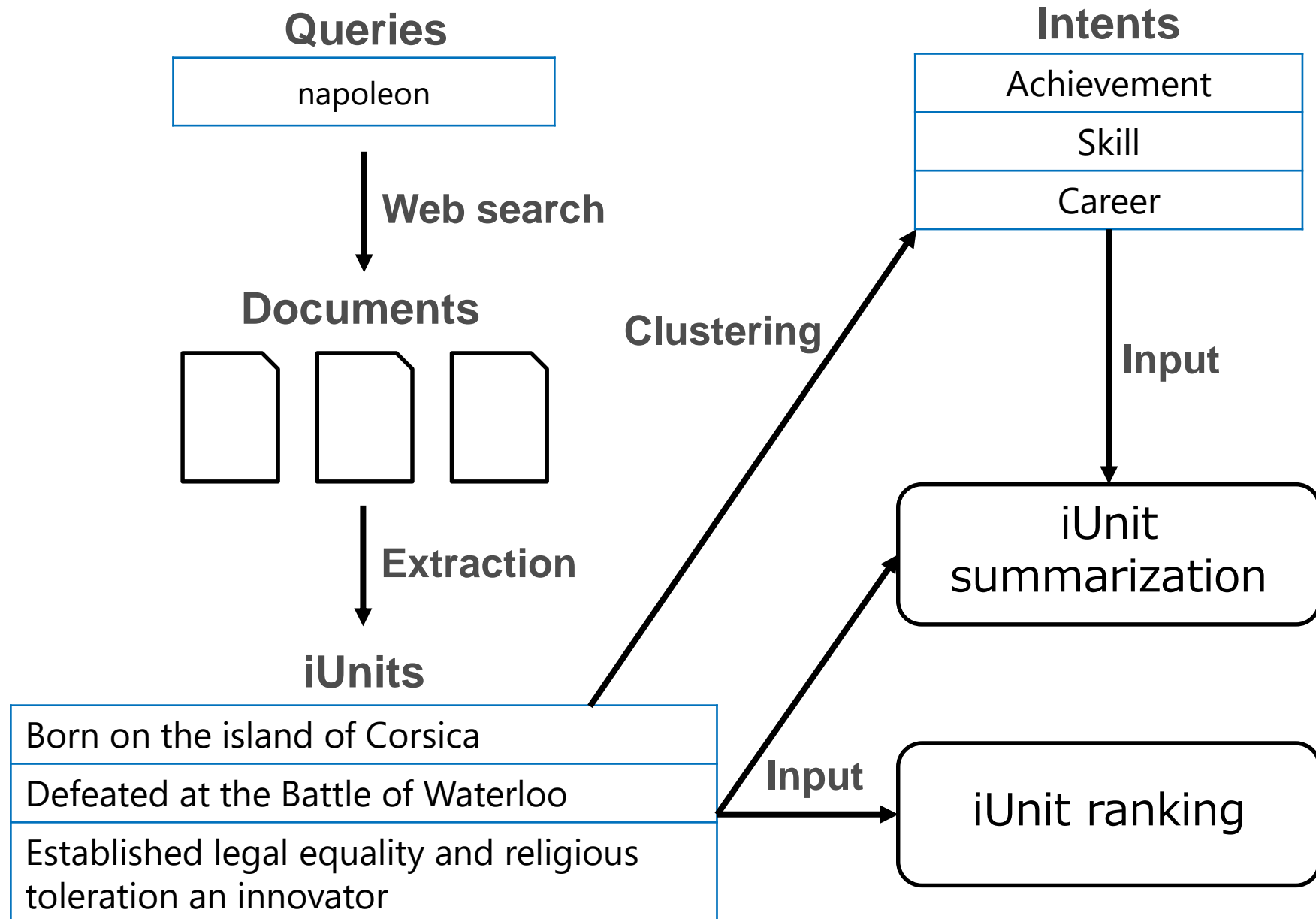| Born on the island of Corsica |
|---|
| Defeated at the Battle of Waterloo |
| Established legal equality and religious toleration an innovator |
| Absent during Peninsular War |

**Clustering** →

| Achievement |
|---|
| Skill |
| Career |

- 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 |
| --- |

↓ **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 |

**Clustering**

**Intents**

| Achievement |
| --- |
| Skill |
| Career |

↓ **Input**

**iUnit summarization**

**Input** → **iUnit ranking**

# EVALUATION

# Per-intent iUnit Importance

- **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"*

| iUnit | Importance |
|---|---|
| A series of evaluation workshops | 5 |
| Task Registration Due 20/Jun./2016 | 3 |

### *In terms of intent "Schedule"*

| iUnit | Importance |
|---|---|
| A series of evaluation workshops | 2 |
| Task Registration Due 20/Jun./2016 | 5 |

- **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**

| Intent | # of votes |
|---|---|
| Definition | 4 |
| Schedule | 3 |
| Tasks | 3 |

**Intent Probability**

| Intent | Prob. |
|---|---|
| Definition | 0.4 |
| Schedule | 0.3 |
| Tasks | 0.3 |

- ## **Evaluated in the same way as ad-hoc retrieval**

Output: **iUnit list**

| | iUnit |
|---|---|
| 1 | A series of evaluation … |
| 2 | Task Registration Due … |

➡

| | iUnit | GI |
|---|---|---|
| 1 | A series of evaluation … | **3.8** |
| 2 | Task Registration Due … | **2.5** |

**nDCG@10**
**Q-measure**

➡ **0.87**

### Per-intent iUnit Importance

*In terms of intent "Schedule"*

| iUnit | Importance |
|---|---|
| A series of evaluation workshops | **2** |
| Task Registration Due 20/Jun./2016 | **5** |

### Intent Probability

| Intent | Prob. |
|---|---|
| Definition | **0.4** |
| Schedule | **0.3** |

### **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 $q$

| iUnit | Importance |
|---|---|
| A series of evaluation workshops | **3.8** |
| Task Registration Due 20/Jun./2016 | **2.5** |

28

# 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

**Summary**

$u_1$   $u_2$
$u_3$

➡

**Trailtext
(reading path)**

$u_1$   $u_2$   $u_3$

$10_{chars}$  $5_{chars}$  $10_{chars}$

➡

**U-measure
$G(u_1)(1-10/L)$
$+ G(u_2)(1-15/L)$
$+ G(u_3)(1-25/L)$**

**Create a list of iUnits
by assuming that users
read text from left to right,
from top to bottom**

$$U = \sum_{r=1} G(u_r)(1 - \mathrm{pos}(u_r)/L)$$

$u_r$: r-th iUnit
$G(u)$: importance of u
$\mathrm{pos}(u)$: offset of u from the beginning
$L$: patience parameter

- **M-measure**
  - **Expectation of U-measure over multiple *trailtexts***

$$M = \sum_{\mathbf{t}} P(\mathbf{t})U(\mathbf{t})$$

$P(\mathbf{t})$: probability of trailtext $\mathbf{t}$
$U(\mathbf{t})$: U-measure of trailtext $\mathbf{t}$
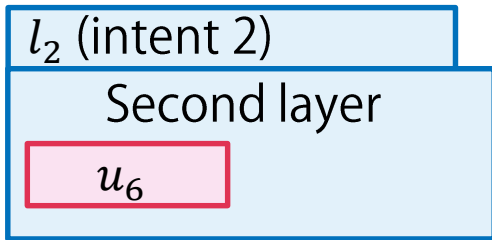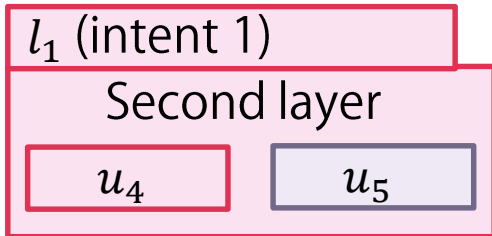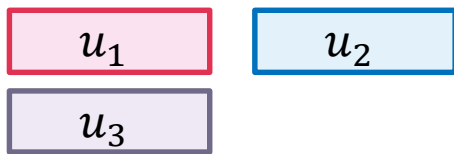
- **Generate trailtexts by assuming that**
  - Users read a summary from the top of the first layer
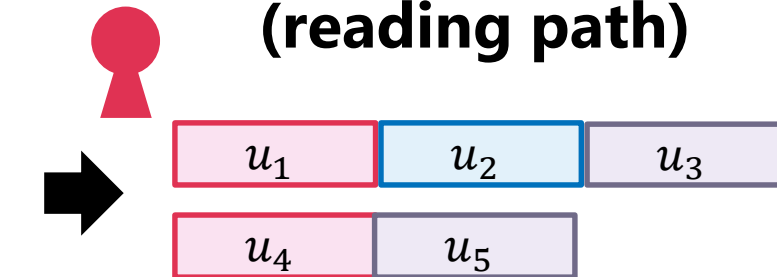  - Users click on an intent if they are interested in it

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 $(P(i_2|q))$
→
$u_1$ | $u_2$ | $u_3$

30

- **Compute the expectation of U-measure**

First layer

Trailtext (t)
(reading path)

U    M-measure

$u_1$    $u_2$

$u_3$

$l_1$ (intent 1)
  Second layer
  $u_4$    $u_5$

$l_2$ (intent 2)
  Second layer
  $u_6$

$u_1$    $u_2$    $u_3$

$u_4$    $u_5$

$$P(\mathbf{t}_1) = P(i_1|q) = 0.75$$

0.44

0.36

$u_1$    $u_2$    $u_3$

$u_6$

$$P(\mathbf{t}_2) = P(i_2|q) = 0.25$$

0.12
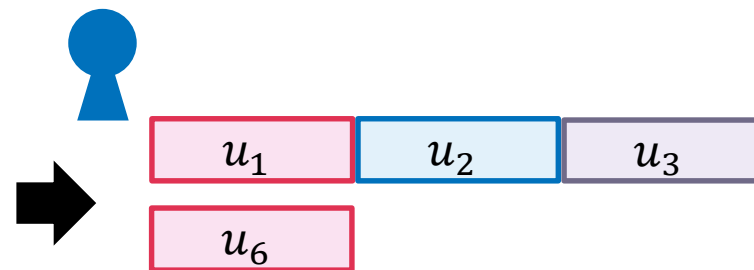
$$M = \sum_{\mathbf{t}} P(\mathbf{t})U(\mathbf{t})$$

Because trailtext $\mathbf{t}_2$ is read
by users interested in $i_2$

# RESULTS

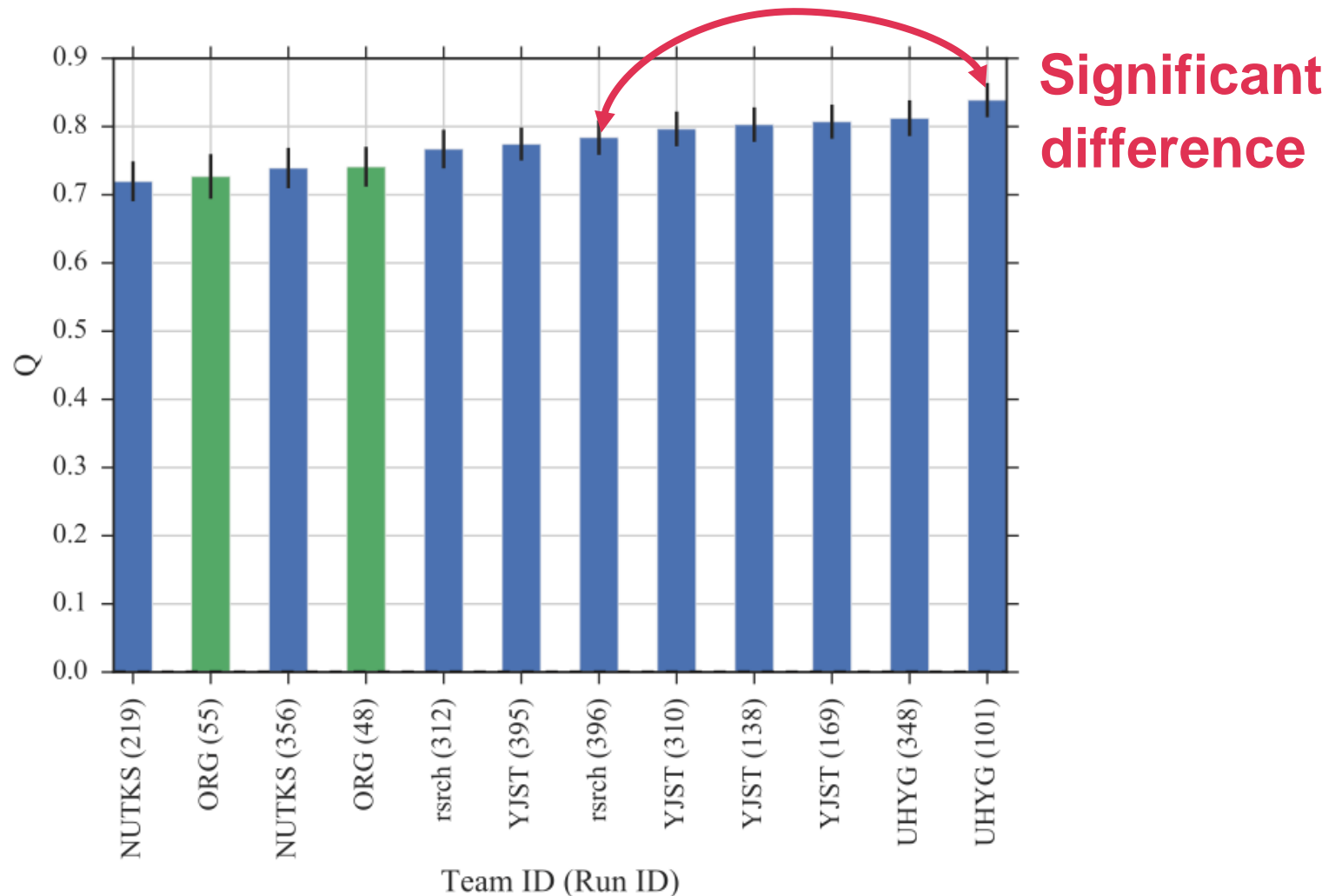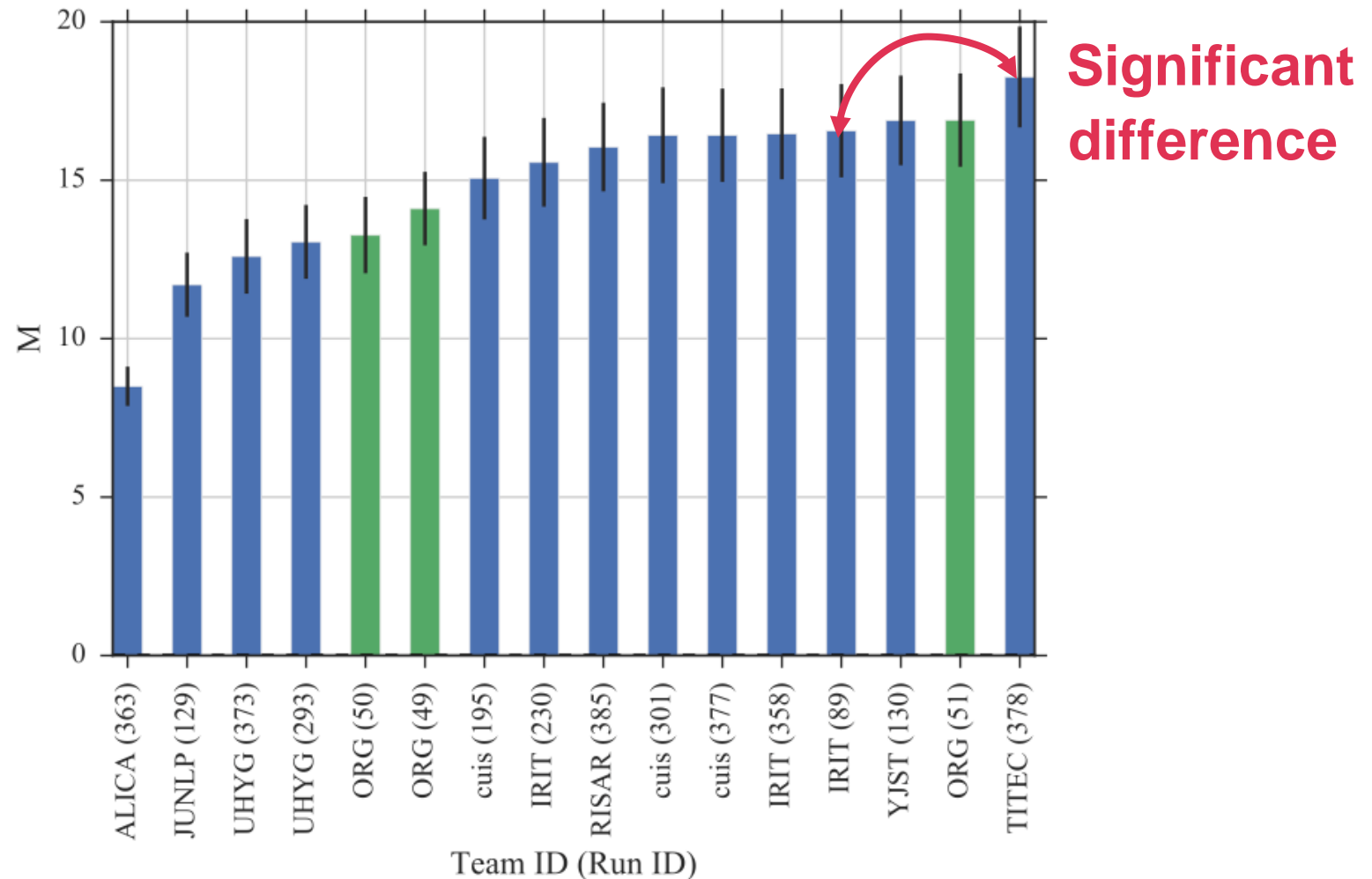# iUnit Ranking (English)

Q

Team ID (Run ID)
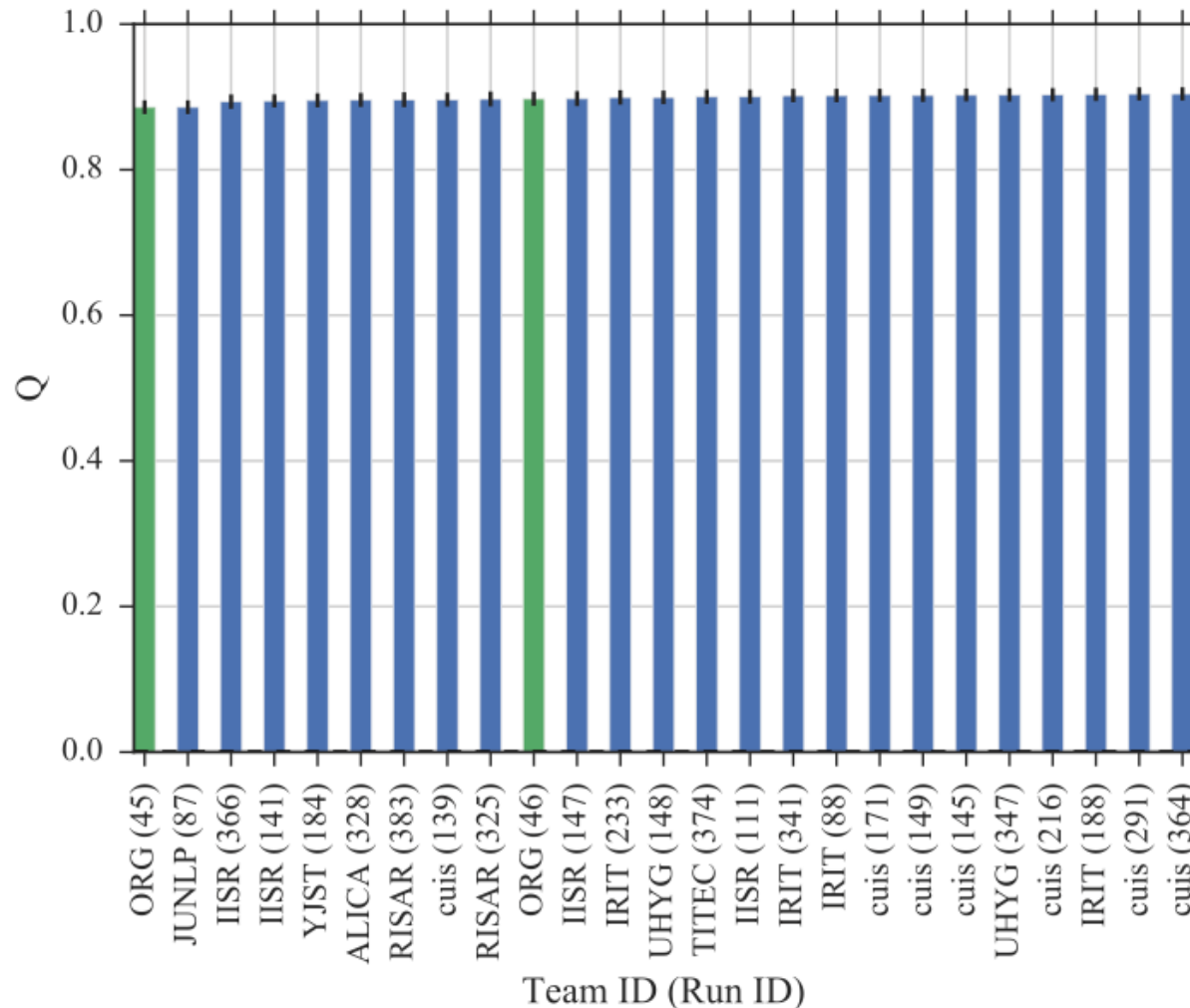
**Submitted runs showed similar performance
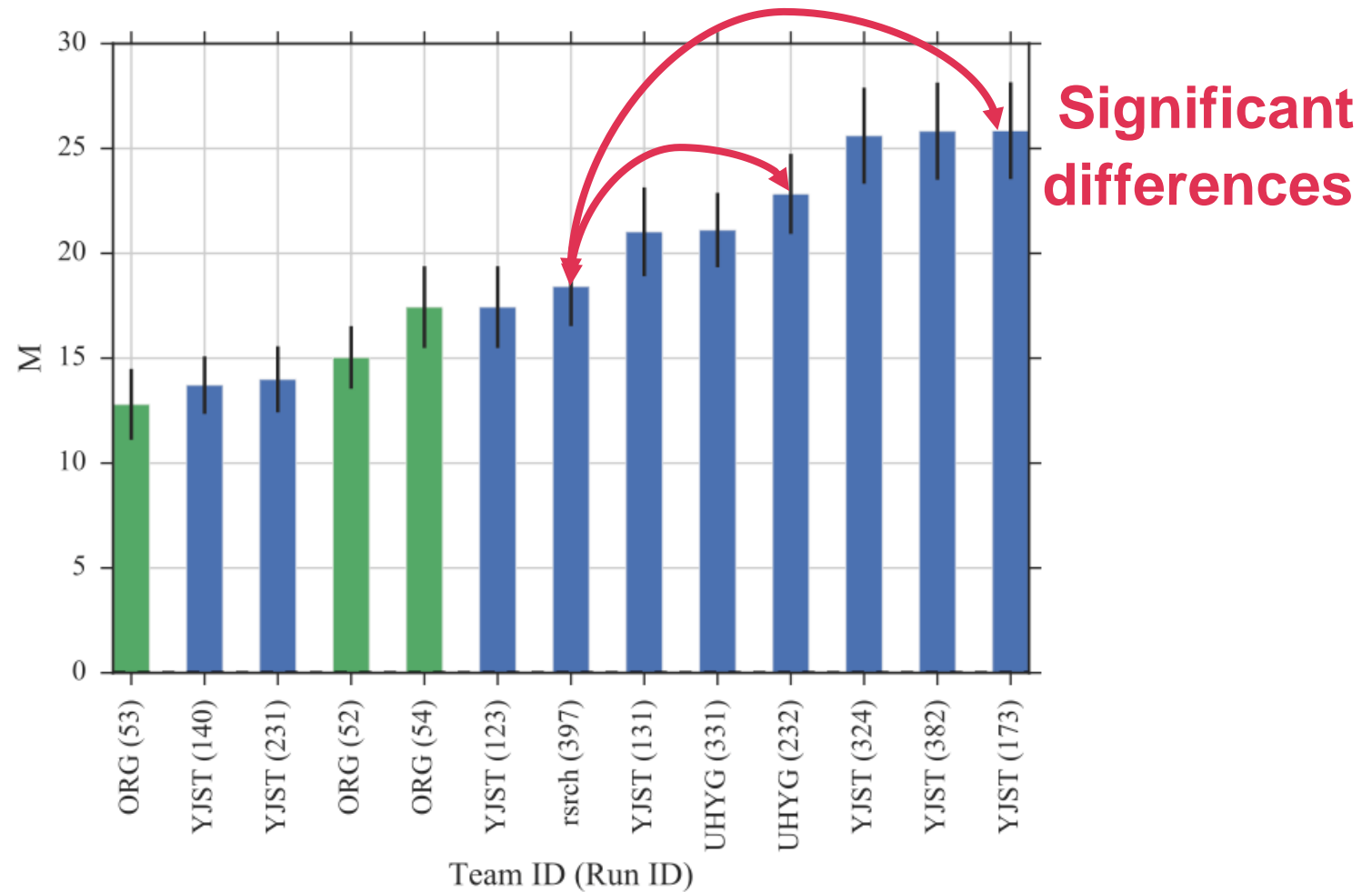(a few statistically significant differences)**

# iUnit Ranking (Japanese)



UHYG, YJST, and rsrch significantly outperformed the baseline method

# iUnit Summarization (English)



**TITEC and YJST are the top and are not statistically distinguishable, but did not significantly outperform the best baseline**

# iUnit Summarization (Japanese)



**YJST and UHYG significantly outperformed the baseline, and are not statistically distinguishable**

# NEW TRIALS

## README.md

# mobileclick

mobileclick provides baseline me

http://www.mobileclick.org/

`pypi package` `0.1.4` `circleci` `passing`

# 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 \
--iunit data/MC2-training/en/1C2-E-iunits.tsv \
--indexdir data/MC2-training-documents/1C2-E.INDX \
--pagedir data/MC2-training-documents/1C2-E.HTML
```

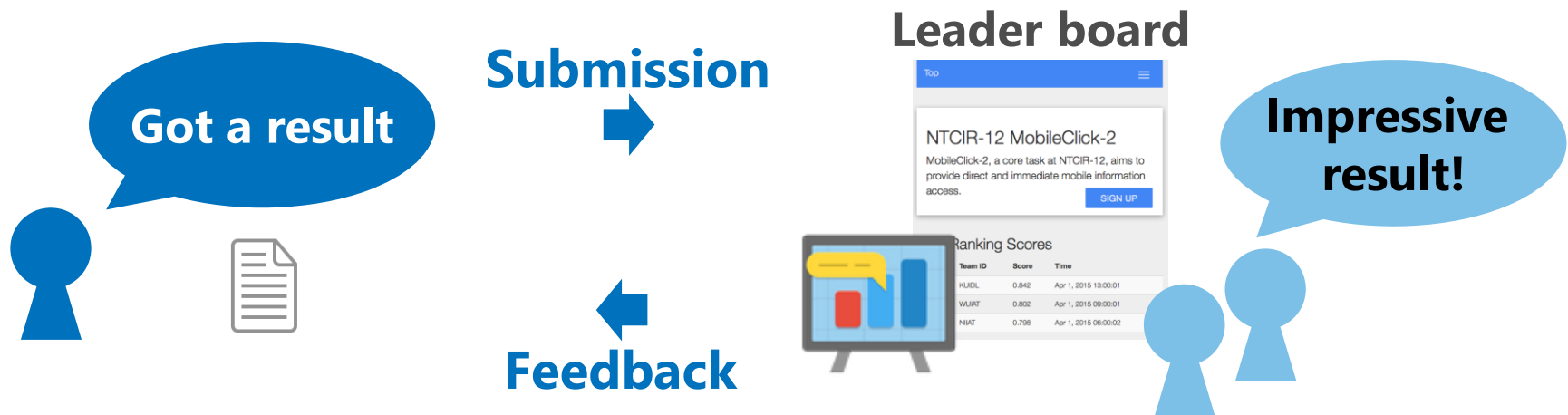Replicate the LM-based iUnit ranking baseline:

```
$ mobileclick_lan
--query data/MC2-
--iunit data/MC2-
--indexdir data/M
--pagedir data/MC
--language englis
```

**1** line for downloading the data

**5** lines to generate baseline results
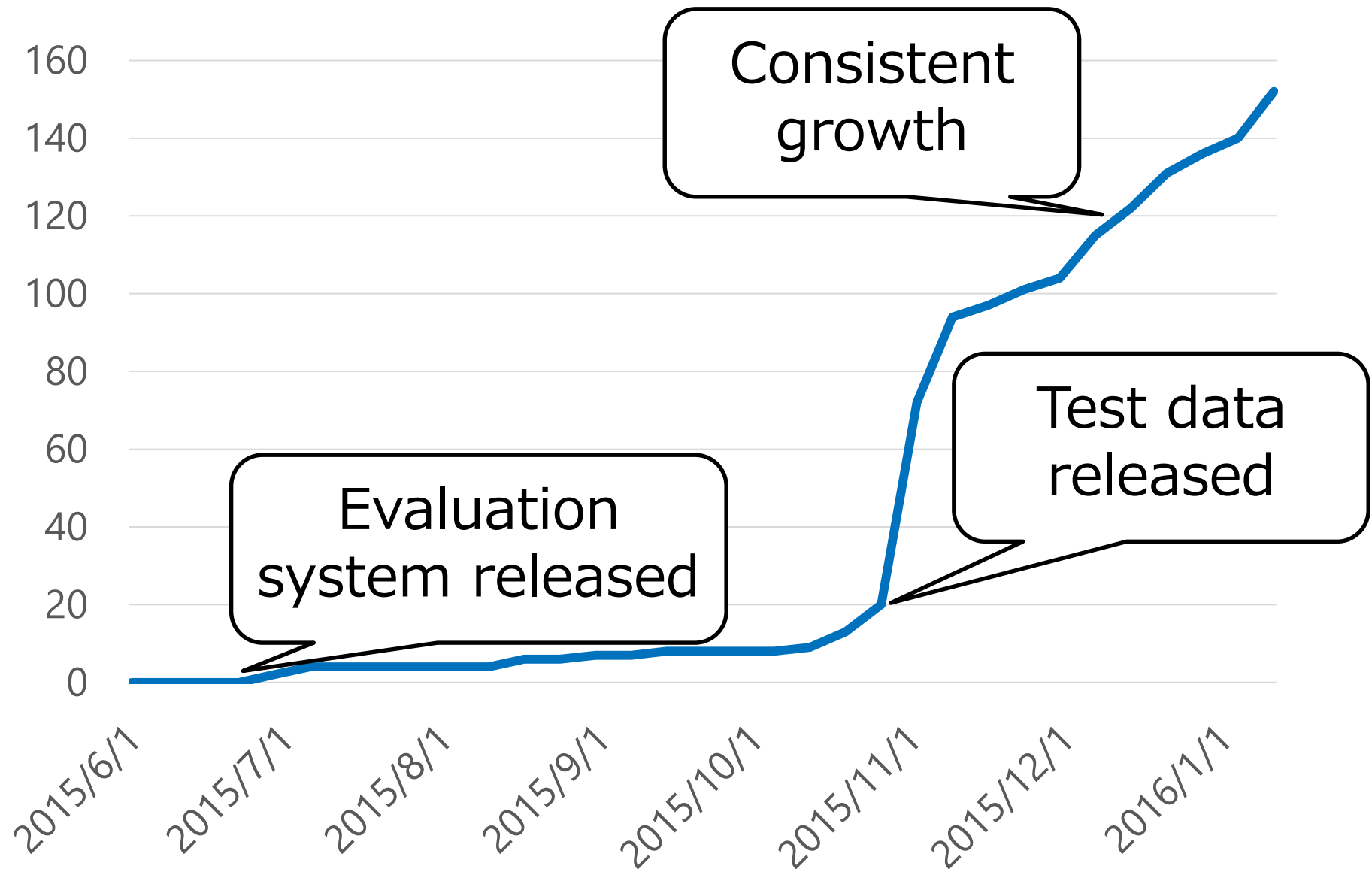
## Q. When can we get our evaluation result?
## A. Right after you submit your run!



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

# NTCIR-12 MobileClick-2 `Click`

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? | How to participate? | How to get started? |
|---|---|---|
| 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. | You can register at the NTCIR online registration system. Please also make an account of this website for getting datasets and more information. | 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 | Number of users | Number of participating groups |
|---|---|---|
| 455 | 52 | 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)

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

## English iUnit Summarization (Test)

| Rank | Group ID | Score | Time |
|---|---|---|---|
| 1 | TITEC | 18.2596 | 2016-02-04 22:33:51 UTC |
| 2 | ORG | 16.8975 | 2015-10-31 15:08:56 UTC |
| 3 | RISAR | 16.047 | 2016-02-05 06:27:34 UTC |
| 4 | ORG | 14.1051 | 2015-10-31 15:08:30 UTC |
| 5 | ORG | 13.2689 | 2015-10-31 15:08:41 UTC |

## Japanese iUnit Ranking (Test)

| Rank | Group ID | Score | Time |
|---|---|---|---|
| 1 | UHYG | 0.8388 | 2015-11-09 09:30:18 UTC |
| 2 | UHYG | 0.8123 | 2016-02-02 03:24:36 UTC |
| 3 | ORG | 0.7411 | 2015-10-31 15:03:44 UTC |
| 4 | ORG | 0.7269 | 2015-10-31 15:10:56 UTC |

## Japanese iUnit Summarization (Test)

| Rank | Group ID | Score | Time |
|---|---|---|---|
| 1 | UHYG | 22.8342 | 2016-01-22 09:54:39 UTC |
| 2 | UHYG | 21.1107 | 2016-02-01 03:26:58 UTC |
| 3 | ORG | 17.4376 | 2015-10-31 15:10:05 UTC |
| 4 | ORG | 15.0373 | 2015-10-31 15:09:39 UTC |
| 5 | ORG | 12.799 | 2015-10-31 15:09:52 UTC |

## MobileClick-1

- No team outperformed the baseline

- 4 teams participated

- 14 runs were submitted
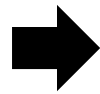
## MobileClick-2

- **Statistically significant differences**

- **11 teams participated**

- **66 runs were submitted**

# Summary

- **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