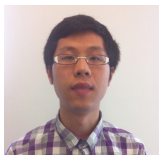
2016 / 06 / 08

Overview of the NTCIR-12 IMine-2 task

Takehiro Yamamoto	Kyoto University
Yiqun Liu	Tsinghua University
Min Zhang	Tsinghua University
Zhicheng Dou	Renmin University of China
Ke Zhou	Yahoo! Labs London
Ilya Markov	University of Amsterdam
Makoto P. Kato	Kyoto University
Hiroaki Ohshima	Kyoto University
Sumio Fujita	Yahoo! JAPAN Corporation



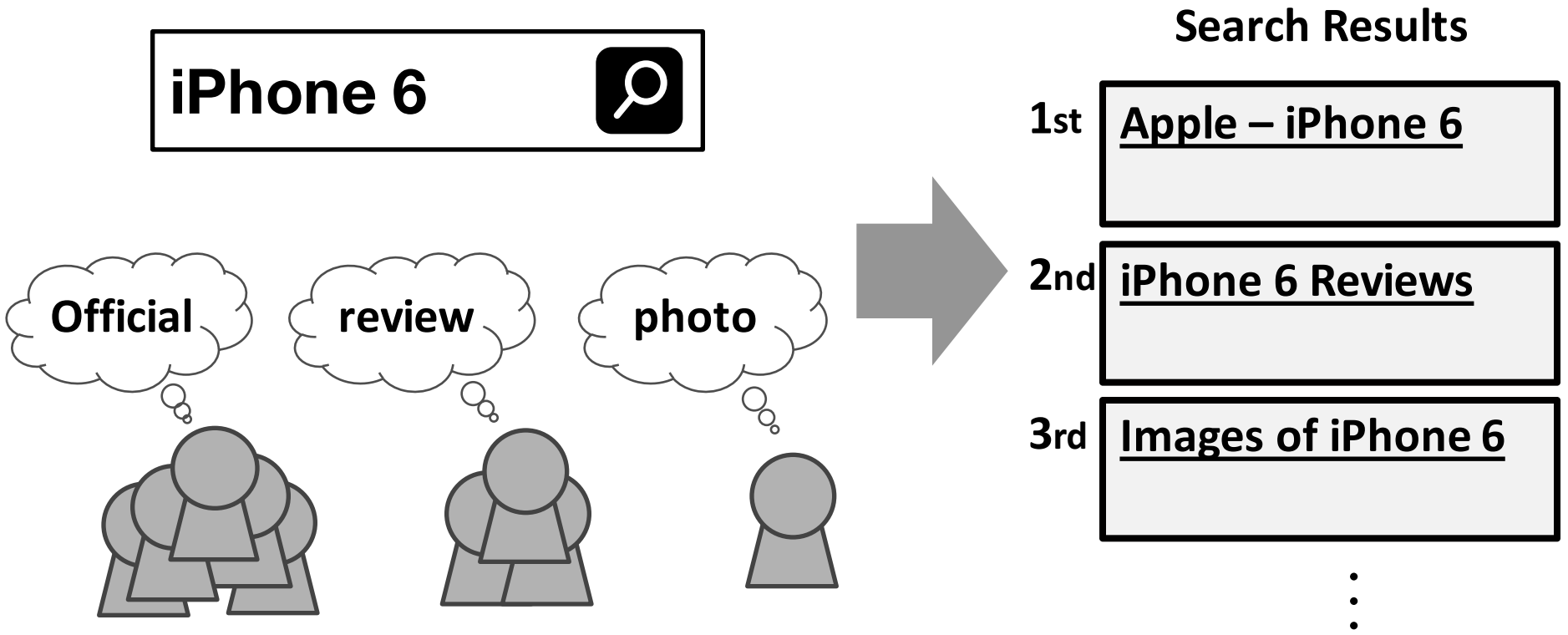
iPhone 6

Goal of IMine

Understanding
search intent
behind the query

Search Result Diversification (INTENT, INTENT-2, IMine-1)

Given an ambiguous/underspecified query, produce a single result page that **satisfies different user intents**



Key Challenges

- Identify possible **search intents** behind the query
  - review, official, photo, …
- Generate a **diversified ranked list** that satisfies different search intents

What's new in IMine-2?

IMine-2

=

Search Result Diversification **+** **Vertical Intents**

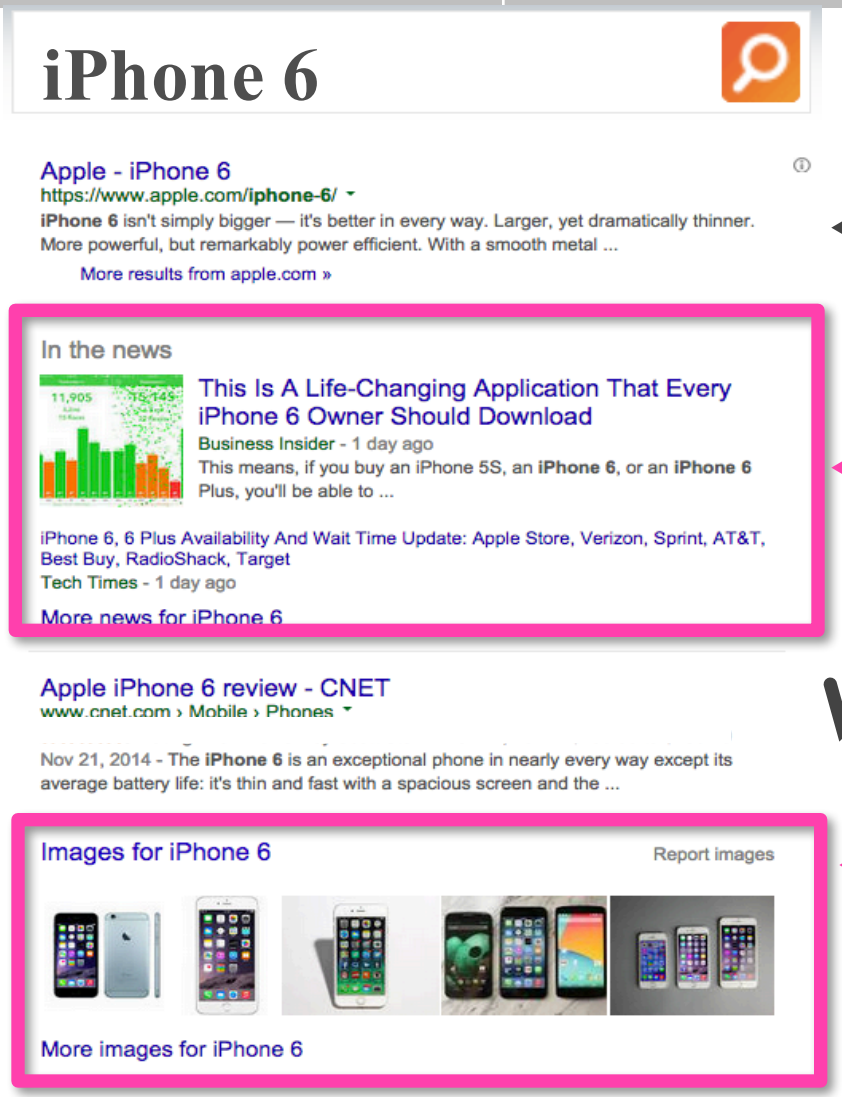
Vertical Intents

official

sales

review

photo



Web

News

Image

11 Verticals

Vertical Intents

We have to identify relevant verticals

as well as search intents of a query

SUBTASKS

Two Subtasks

- Query Understanding subtask
  - ≒ Subtopic Mining subtask
- Vertical Incorporating subtask
  - ≒ Document ranking subtask

Query Understanding Subtask (EN, CN, JA)

Given a query, the system is required to generate a diversified ranked list of subtopics with their relevant vertical intents.

iPhone 6

Subtopic	Importance	Vertical Intent
iPhone 6 Apple	*0.95*	*Web*
iPhone 6 sales	*0.80*	*News*
iPhone 6 image	*0.73*	*Image*
iPhone 6 review	*0.65*	*Web*
⋮	⋮	⋮

Query Understanding Subtask (EN, CN, JA)

Given a query, the system is required to generate a diversified ranked list of subtopics with their relevant vertical intents.

iPhone 6

Subtopic	Importance	Vertical Intent
iPhone 6 Apple	*0.95*	*Web*
iPhone 6 sales	*0.80*	*News*
[illegible] 6 image	*0.73*	*Image*

Challenges

- Subtopic Relevance
- Subtopic Diversity
- Vertical Relevance

Verticals

(a) Image

(b) News

(c) Encyclopedia

(d) Download

(e) QA

(f) Shopping

Vertical Incorporating subtask (EN, CN)

Given a query and the document collection, the system is required to return a diversified ranked list of documents

iPhone 6

Documents	Importance
doc-011	*0.95*
Vertical-news	*0.80*
doc-021	*0.73*
Vertical-image	*0.65*
⋮	⋮

Virtual vertical document

= virtual document that represents *ideal* vertical result of the query

= *Participants can focus on which vertical and where it should be incorporated into the ranked list*

Vertical Incorporating subtask (EN, CN)

Given a query and the document collection, the system is required to return a diversified ranked list of documents

iPhone 6

Documents	Importance
doc-011	*0.95*
Vertical-news	*0.80*

Virtual vertical document

= virtual document that represents *ideal* vertical

Challenges

- Document Relevance
- Document Diversity
- Vertical Relevance

DATA

Topics

- **100 queries** for each language
  - 50 queries in IMine-1
- Five types
  - Ambiguous (e.g. *jaguar*)
  - Faceted (e.g. *iPhone6*)
  - Very clear (e.g. *NII address*)
  - Task-oriented (e.g. *lose weight*)
  - Vertical-oriented (e.g., *iPhone photo*)

Data

- Document Corpus
  - for Vertical Incorporating subtask
  - 500 search results of a commercial search engine for each query
- Query suggestions
  - collected by several commercial search engines
- Sogou search user behavior data
- Yahoo! JAPAN related search query data

NEWS RELEASE

大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

平成 27 年（2015 年）7 月 17 日

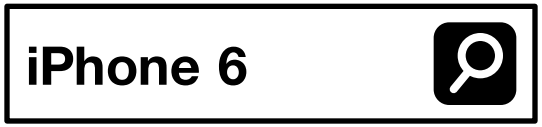
NII と Yahoo! JAPAN が検索技術研究のために新たな提携
「Yahoo!検索」の検索クエリデータを NII のワークショップに無償提供

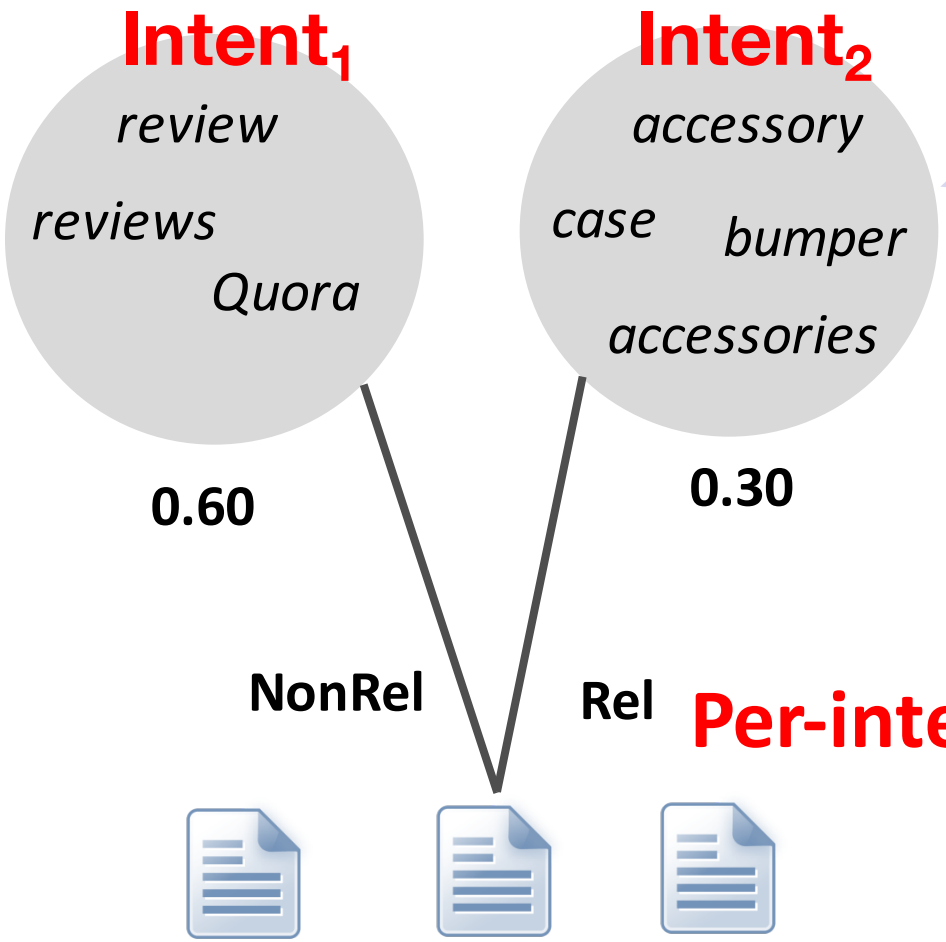
ヤフー株式会社（以下 Yahoo! JAPAN、代表取締役社長：宮坂 学、東京都港区）と大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（以下 NII、所長：喜連川 優、東京都千代田区）は、情報学研究の一層の推進に寄与するために、Yahoo! JAPAN より NII に「Yahoo!検索」の検索クエリデータを無償提供することになりました。本データは、情報アクセス技術の評価ワークショップ「NTCIR（エンティサイル）」（*1）の参加者向けに提供され、同ワークショップに参加する研究グループは無償で活用できます。

Yahoo! JAPAN が提供するデータは、平成 21 年（2009 年）7 月から平成 25 年（2013 年）6 月の期間に「Yahoo!検索」で検索された全クエリ（ユーザーが検索時に入力した単語やフレーズ）の中から、現在進行している NTCIR の第 12 サイクル（NTCIR-12）で設定された研究課題に対する関連度の高い

EVALUATION

Ground Truth Construction

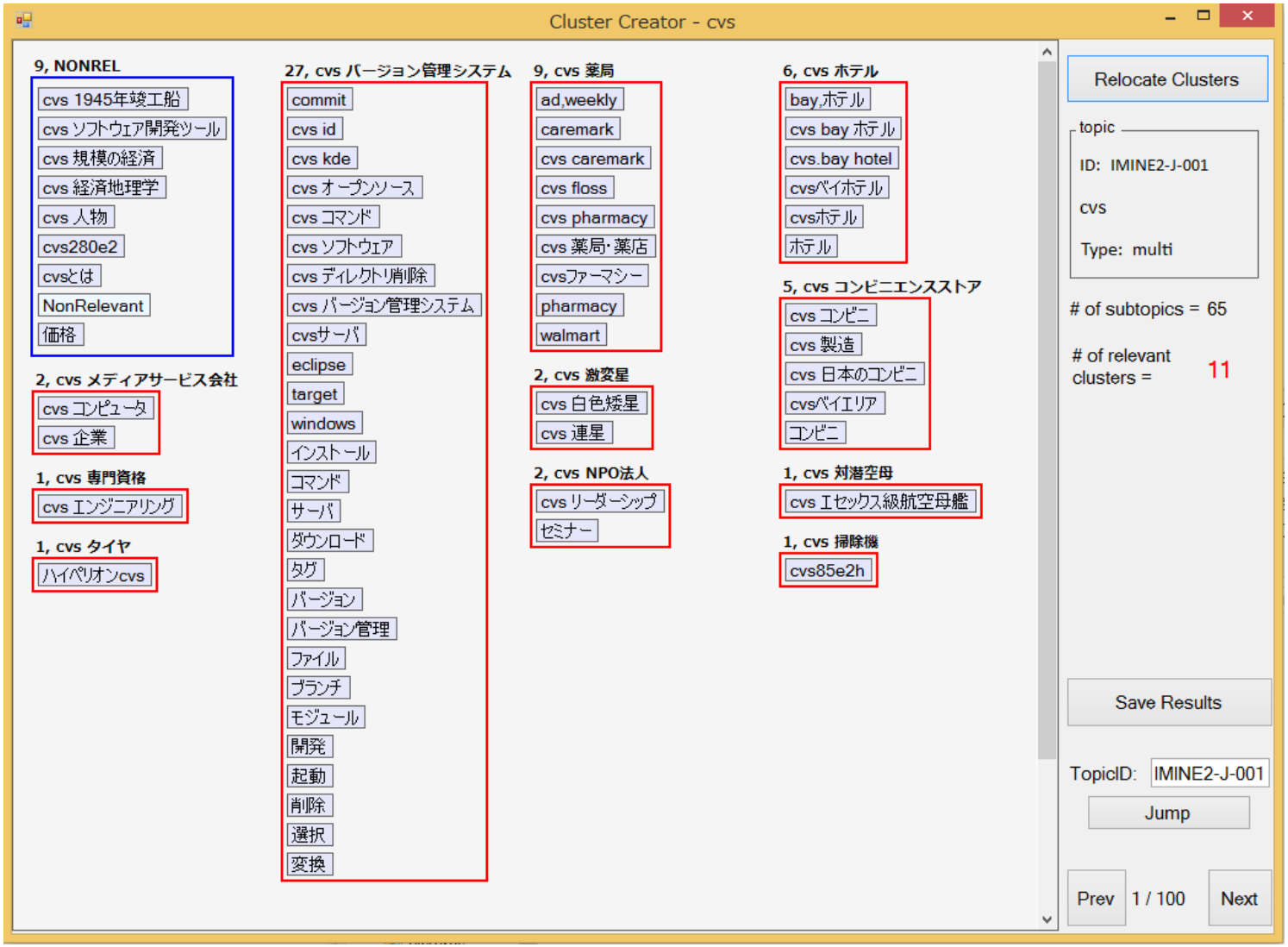




Vertical Importance

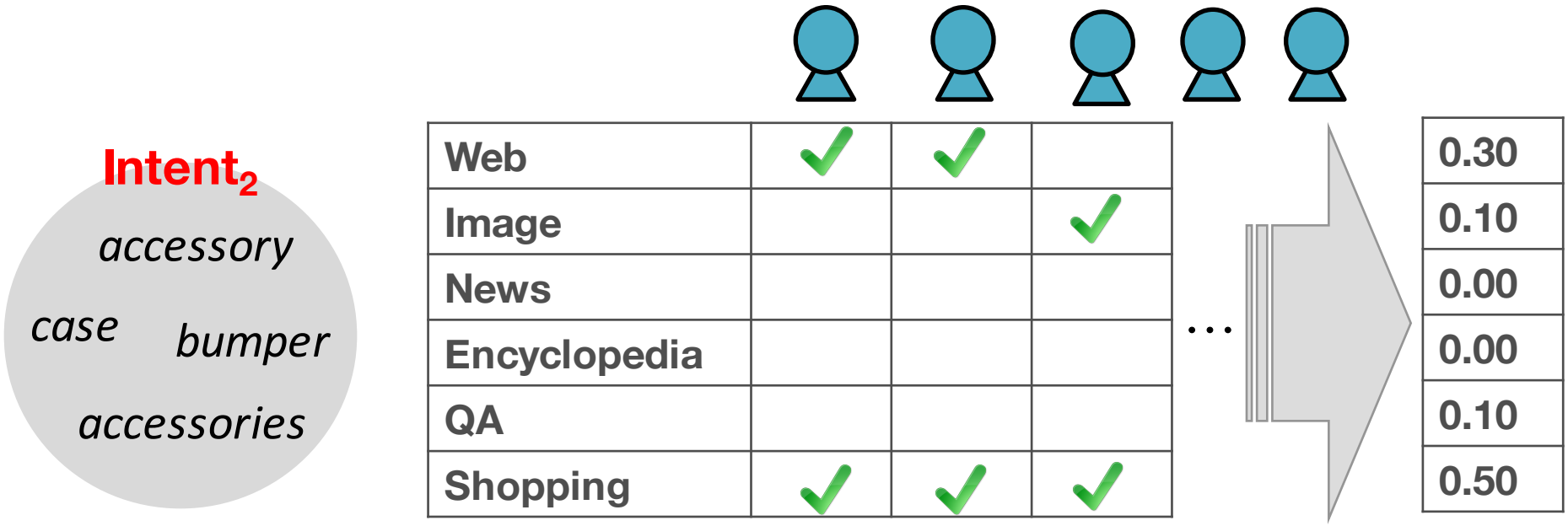
Web	0.30
Image	0.10
News	0.00
Encyclopedia	0.00
QA	0.10
Shopping	0.50

Clustering Pooled Subtopics into Intents

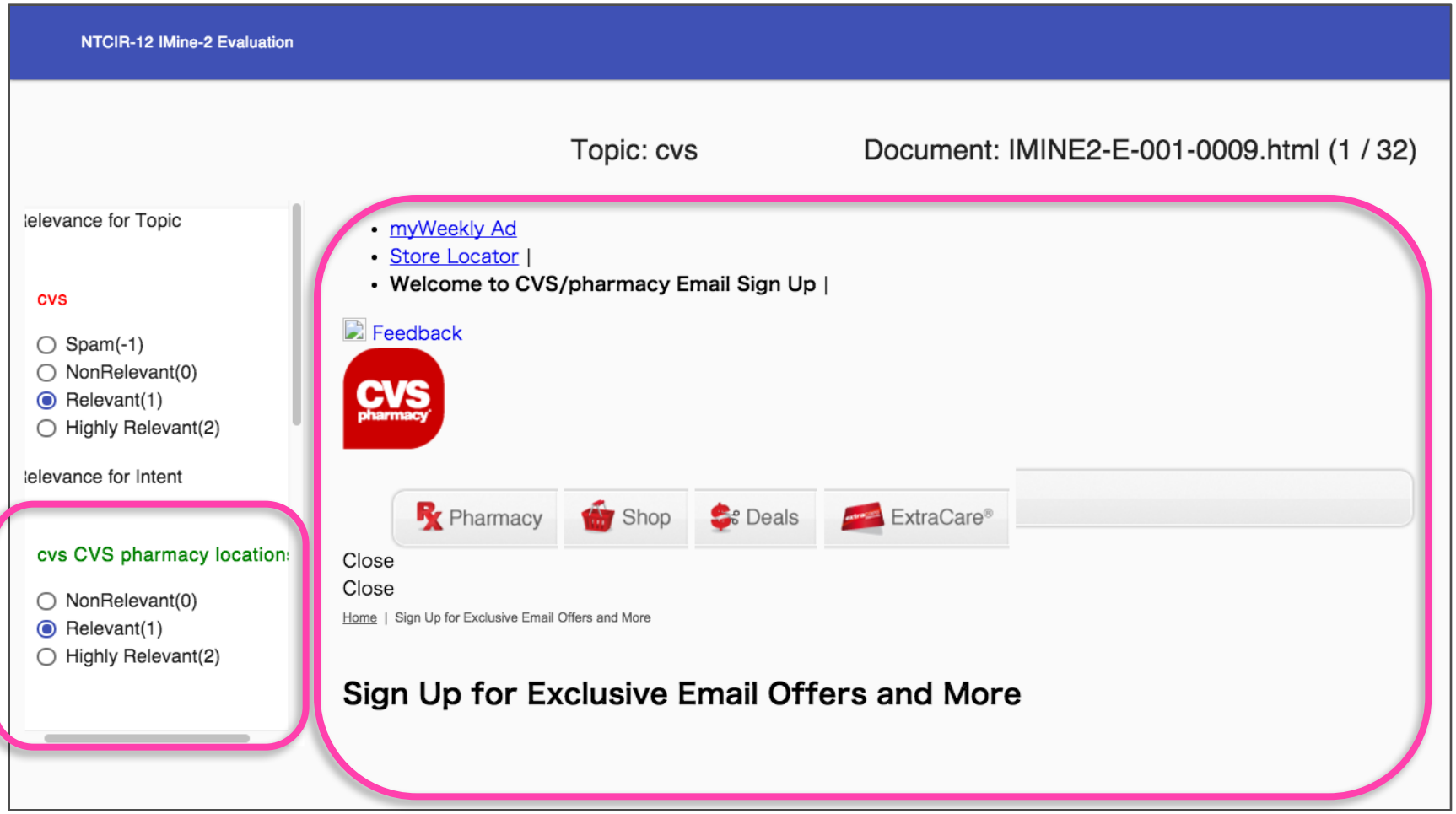


E.g. 6,119 subtopics for Chinese subtask

Vertical Importance



Per-Intent Document Relevance



Per-Intent Relevance

Document

E.g. 6,778 documents for Chinese subtask

EVALUATION METRIC

Query Understanding Subtask (again)

INPUT query

OUTPUT subtopics and their vertical intents

iPhone 6

Subtopic	Importance	Vertical Intent
iPhone 6 Apple	*0.95*	*Web*
iPhone 6 sales	*0.80*	*News*
iPhone 6 image	*0.73*	*Image*
iPhone 6 review	*0.65*	*Web*
⋮	⋮	⋮

Evaluation Metric for Query Understanding

- **Subtopic Diversity**
  - D# proposed by [Sakai+2011]

$$\text{D\#-nDCG} = \gamma I\text{-}rec + (1-\gamma) D\text{-}nDCG$$

How many intents are covered by the subtopics?

How much relevant are the subtopics?

- **Vertical Prediction Accuracy**
  - How accurate are the predicted verticals?

$$\text{V-score@}l = \frac{1}{l}\sum_{r=1}^{l} Accuracy(r)$$

Evaluation Metric for Query Understanding

- Subtopic Diversity
  - D#-nDCG proposed by [Sakai+2011]

$$\text{QU-score} = \lambda D\#\text{-}nDCG@l + (1-\lambda)\, V\text{-}score@l$$

Subtopic Diversity

Vertical Prediction Accuracy

$$\text{V-score}@l = \frac{1}{l}\sum_{r=1} Accuracy(r)$$

Vertical Incorporating subtask (EN, CN)

Given a query and the document collection, the system is required to return a diversified ranked list documents including verticals

iPhone 6

Documents	Importance
doc-011	*0.95*
Vertical-news	*0.80*
doc-021	*0.73*
Vertical-image	*0.65*
⋮	⋮

Evaluation Metric for Vertical Incorporating

- D# proposed by [Sakai+2011]

$$\text{D\#–nDCG} = \gamma I\text{–}rec + (1-\gamma) D\text{–}nDCG$$

How many intents are covered by the documents?

How much relevant are the documents?

$$g_i(d) = \sum_{v \in V} \delta_v(d) P(v|i) rel_i(d) \ ,$$

Gain for intent *i*

Vertical type

Vertical importance

Document Relevance

Participants

We received 42 runs from 9 teams

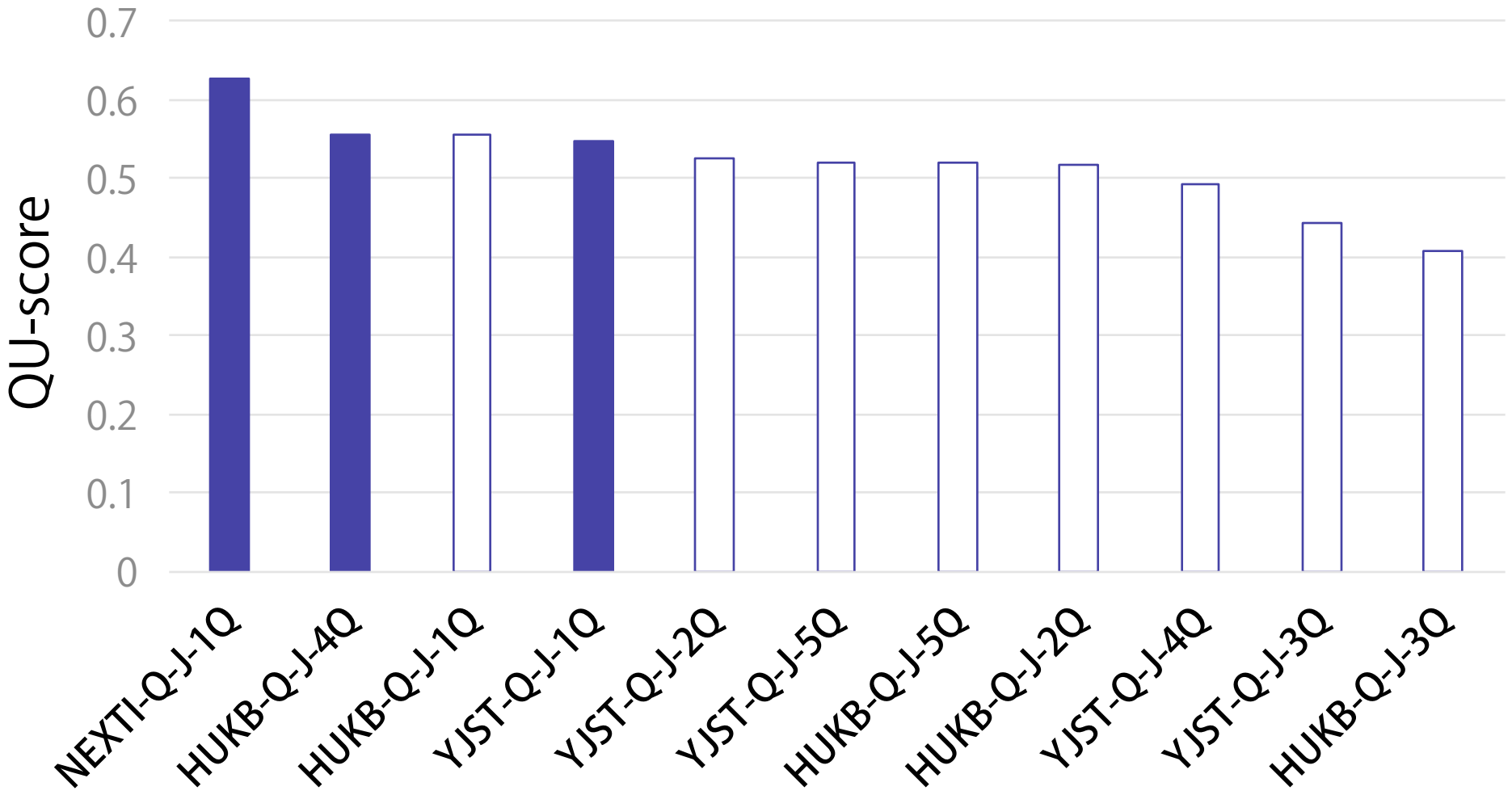
- Techniques
  - Topic Modeling
  - Syntactic Pattern
  - Word Embedding (*e.g.* word2vec)
  - Learning to Rank
- Resources
  - Web pages
  - Queries
  - Encyclopedia (e.g. Wikipedia, Baidu Baike, Hudong Baike …)

Please join IMine-2 session at

June 9th 11:00～12:30 !

RESULTS

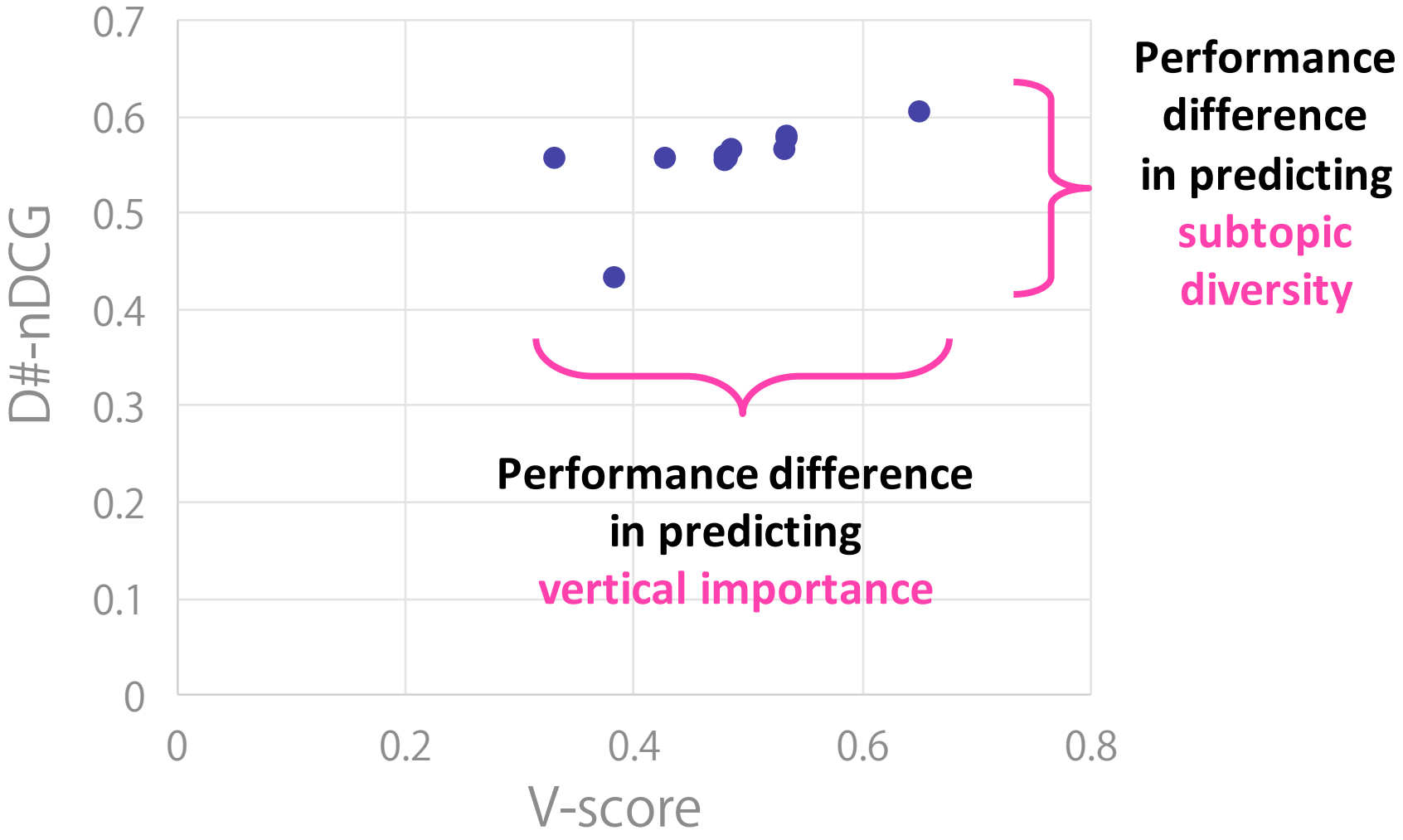
Japanese Query Understanding subtask



NEXTI achieved the best performance

– significantly different from the others

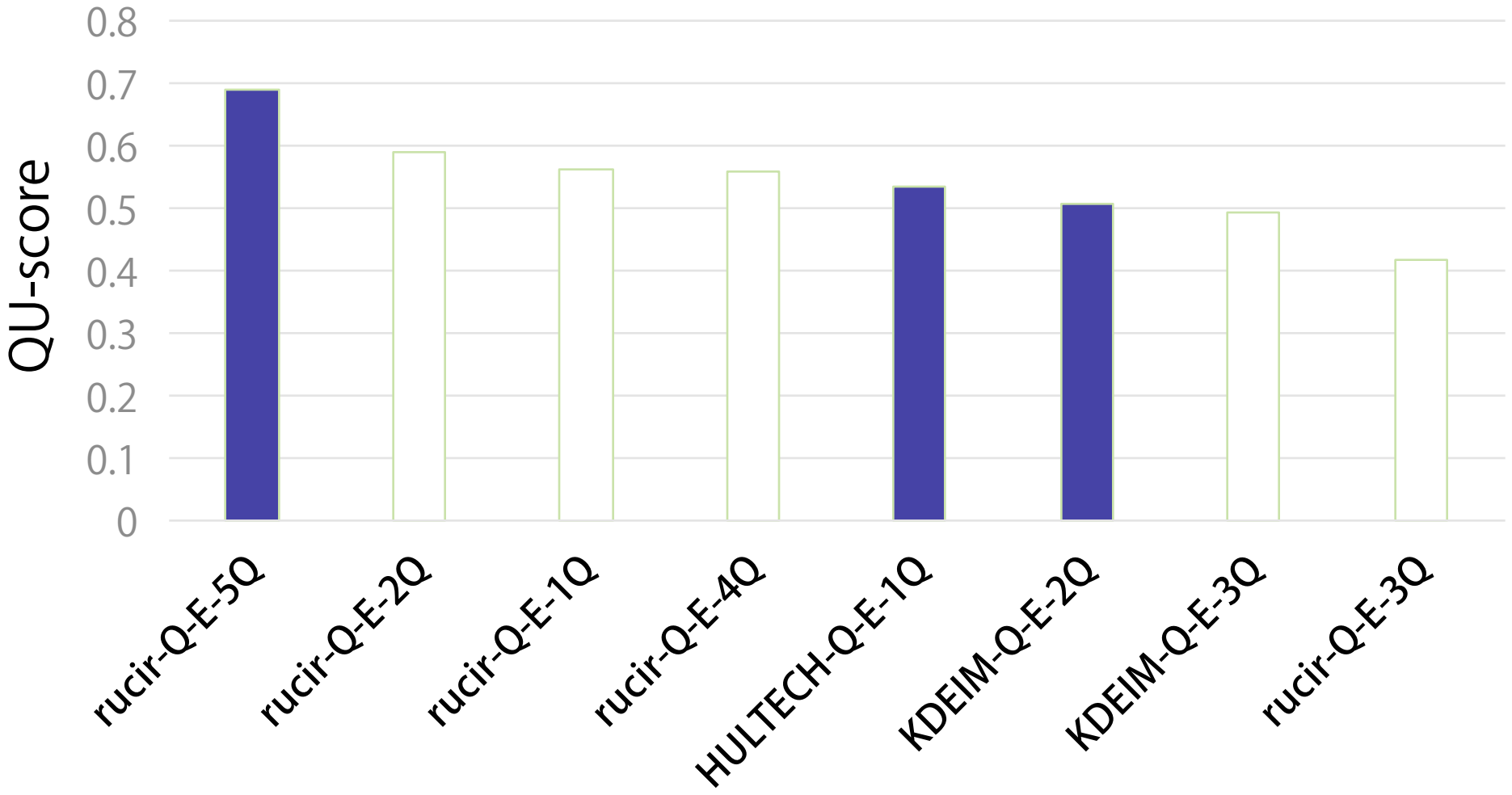
Japanese Query Understanding subtask



V-score diff. > D#-nDCG diff.

Participant runs have similar performances in subtopic diversity

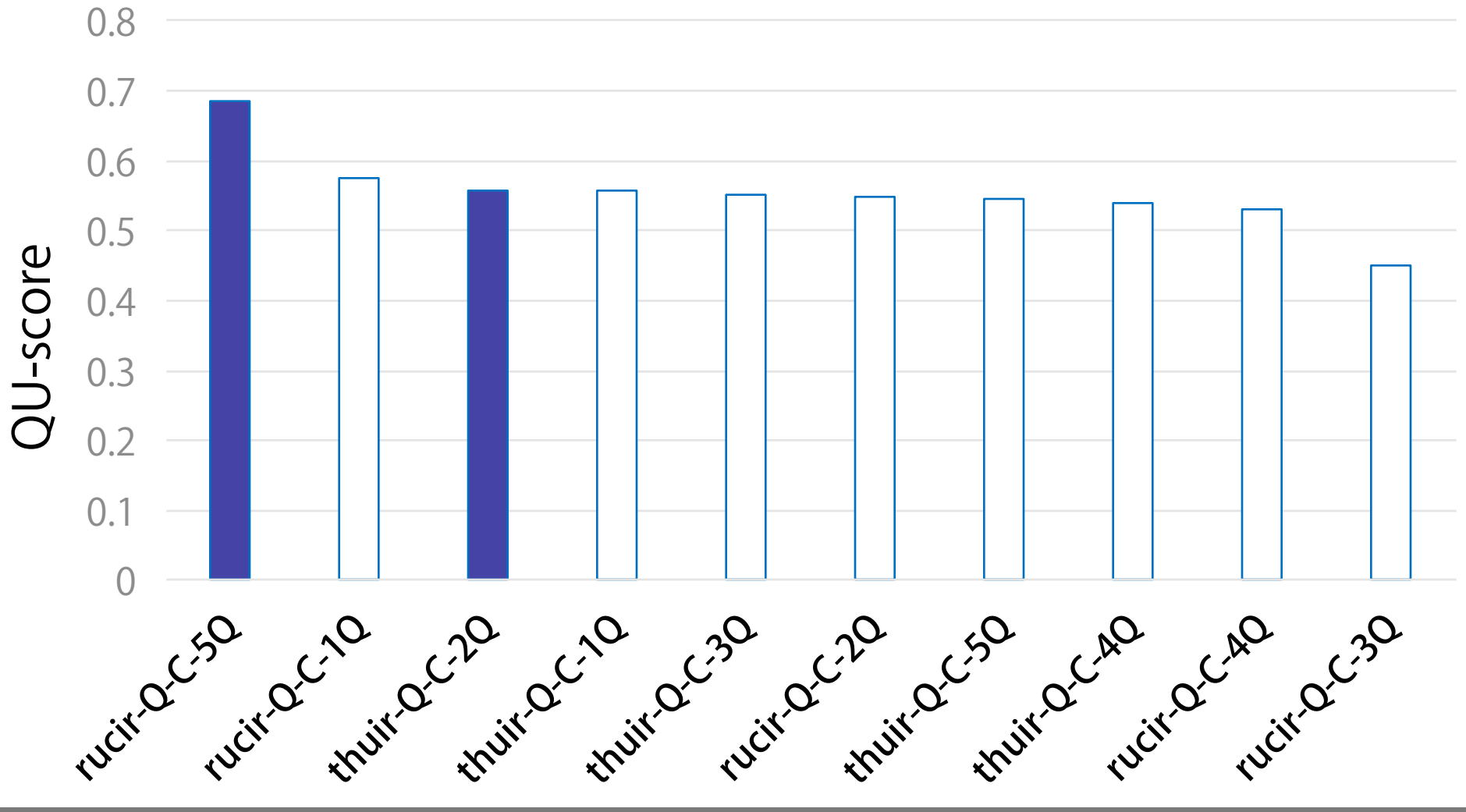
English Query Understanding Subtask



ruicir achieved the best performance

– significantly different from the others

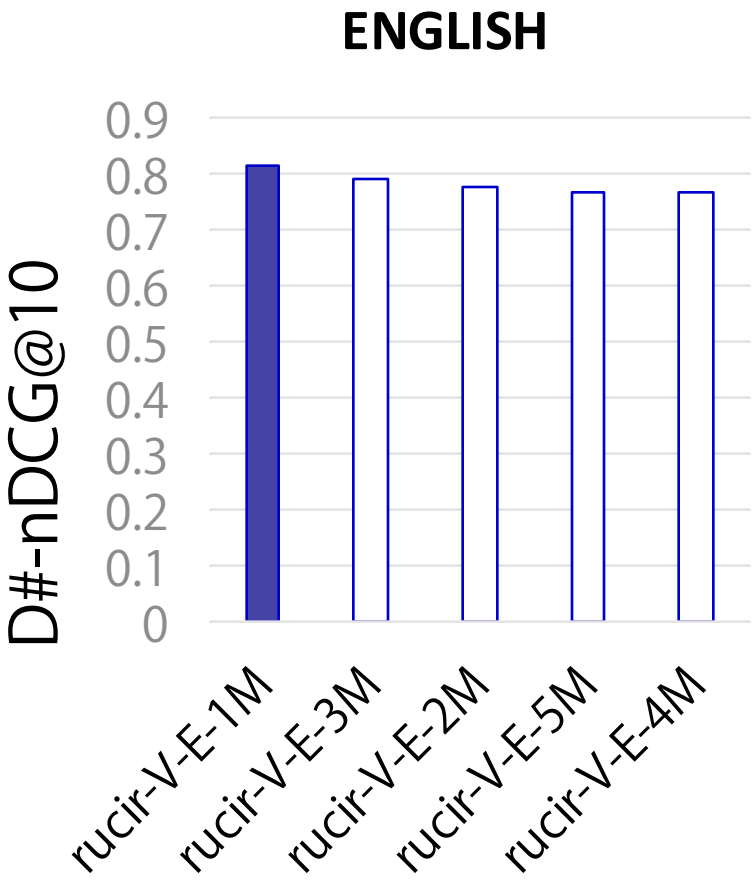
Chinese Query Understanding Subtask

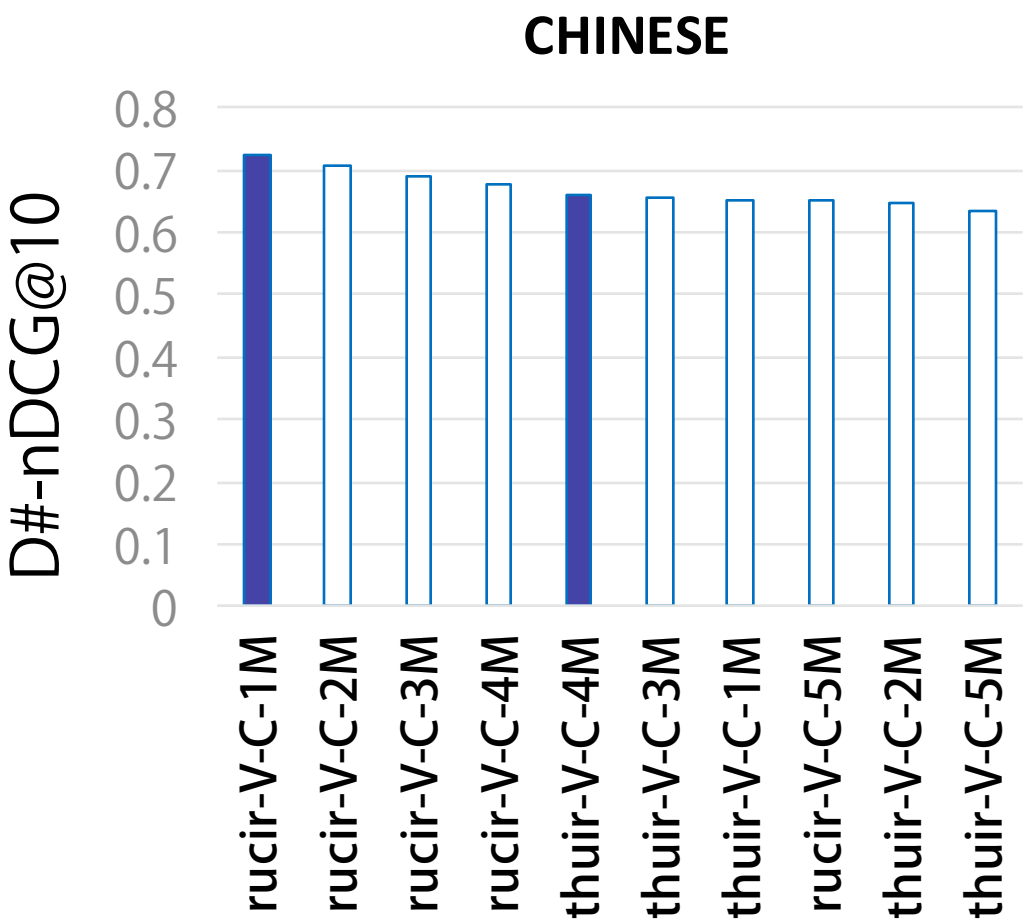


ruicir again achieved the best performance

– significantly different from the others

Vertical Incorporating Subtask





ruicir achieves the best performance

What we have done through INTENT/IMine tasks?

- Researches related to INTENT/IMINE
  - Many Diversification algorithms…
    - Selective Diversification [Sakai+WWW2012] [Tsukuda+AIRS2013]…
    - Hierarchical Intents[Hu+CIKM2015][Wang+SIGIR2016]…
  - Evaluation Metrics [Sakai+SIGIR2011] …
  - Task-oriented Search [Yan+SIGIR2015]…
  - Intrinsically Diverse Search[Raman+SIGIR2013][Umemoto+SIGIR2016]…

What would be the remaining research topics?

- Techniques
  - How can we generate vertical results which can satisfy different user intents?
- Evaluation
  - How can we evaluate the utility of a ranked list as a whole?
    - Appearance of a vertical result affects the utility of other Web documents

Summary

- IMine-2
  - Understand and satisfies diverse users' search intents
- Subtasks
  - Query Understanding
  - Vertical Incorporating
- Results
  - 42 runs from 9 teams
  - Some teams achieved significantly different performance than other teams