

曖昧なクエリと(不) 明快なクエリ: NTCIR-10 INTENT-2と1CLICK-2 タスクへの誘い

酒井 哲也

(Microsoft Research Asia)

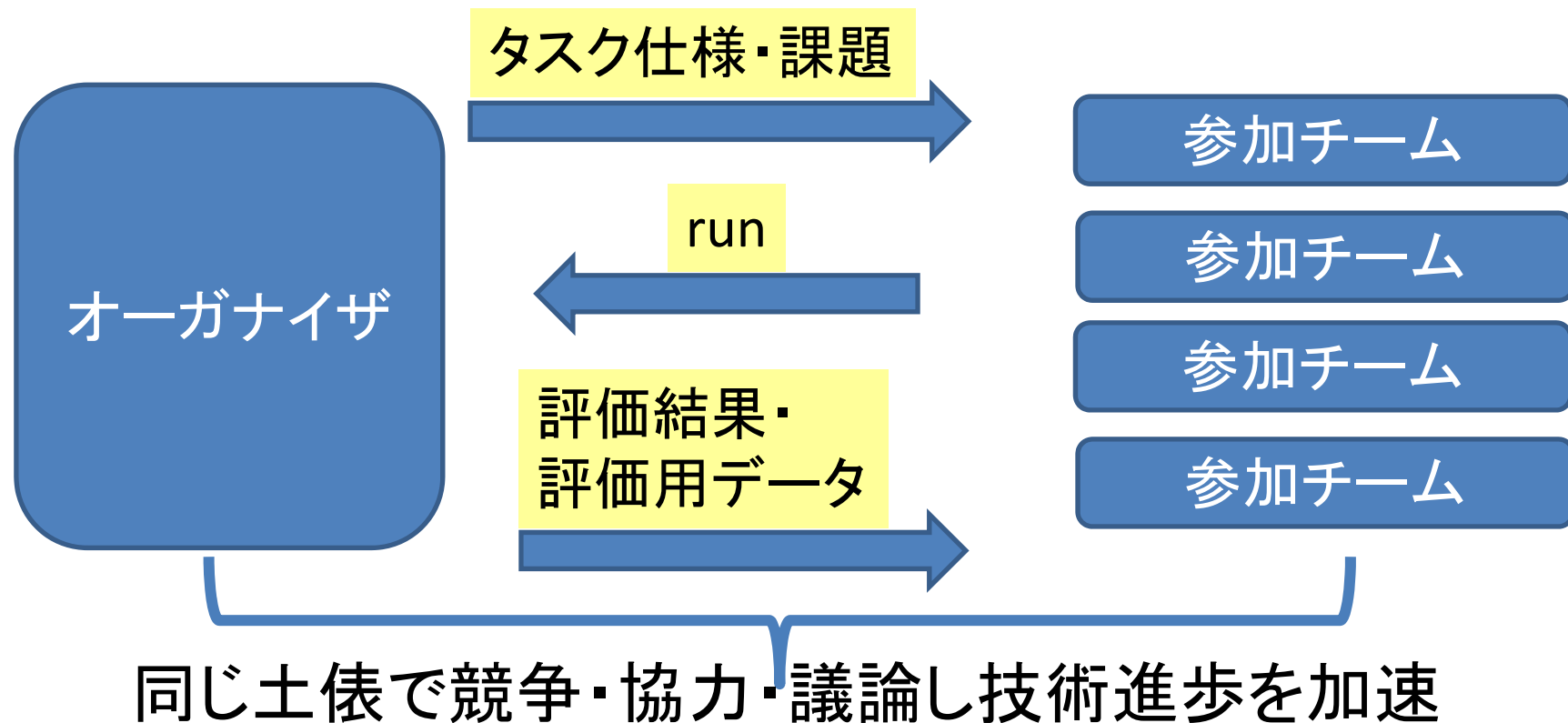
March 26, 2012@IPSJ IFAT/DD

発表概要

1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには

評価型ワークショップとは

- 複数の研究チームが参加、共通の課題を解く。
システム出力(run)を期限までに提出
- オーガナイザが全チームの出力を評価、比較



情報検索の評価型ワークショップ

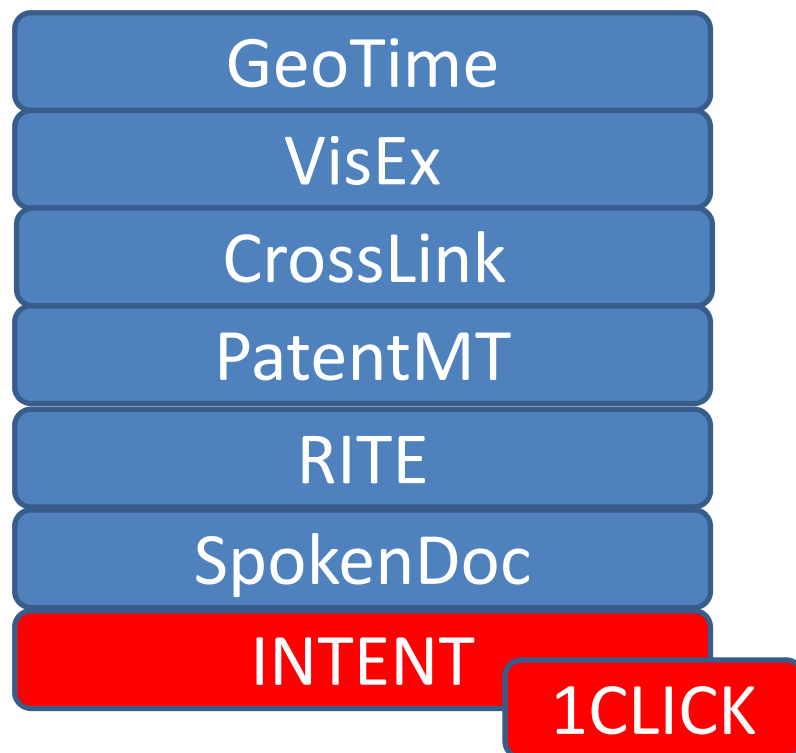
- TREC (Text Retrieval Conference) 1992-
- NTCIR 1999- [一年半毎に開催]
- NII Testbeds and Community for Information access Research
- CLEF (Cross-Language Evaluation Forum) 2000-
- INEX (Initiative for the evaluation of XML retrieval) 2002-
- TRECVID 2003-
- FIRE (Forum for IR Evaluation) 2008-
- MediaEval (Benchmarking Initiative for Multimedia Evaluation) 2010-

詳細は
神門先生の
ご発表で！

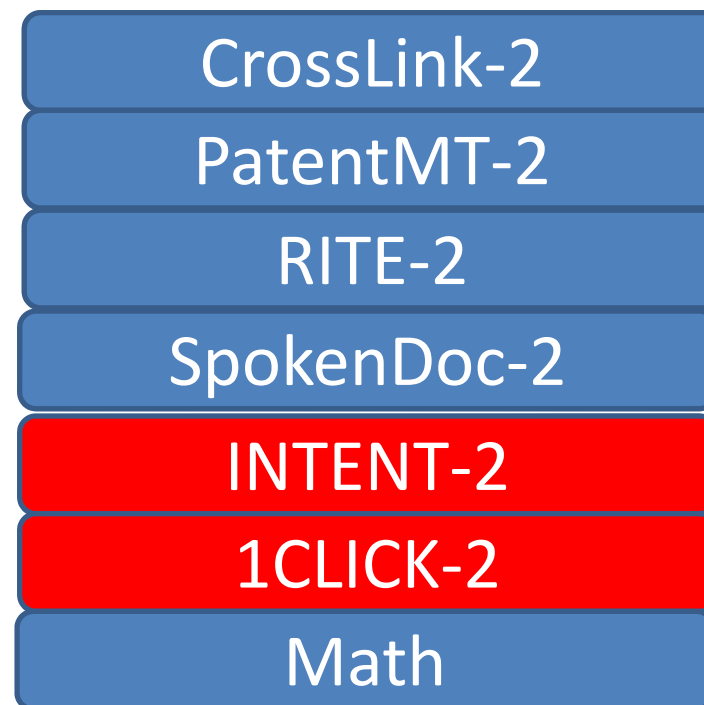
NTCIR-9, 10のタスク(競技種目)

NTCIR-9 (～'11年12月)

NTCIR-10 (～'13年6月)



INTENT, 1CLICK以外の
タスクについてはこの後の
NTCIRセッションで!



発表概要

1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには

このクエリのユーザの意図は？

オフィス



曖昧な (ambiguous) クエリ

オフィス



マイクロソフト製品！



仕事場！



このクエリのユーザの意図は？

ハリーポッター

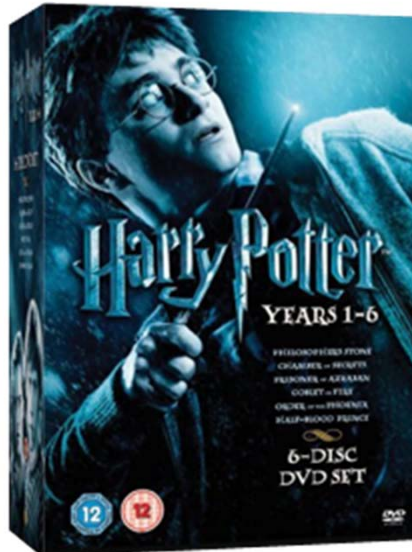


不明快な (underspecified) クエリ

ハリーポッター



ハリーポッターの
映画！



ハリーポッターの
本！



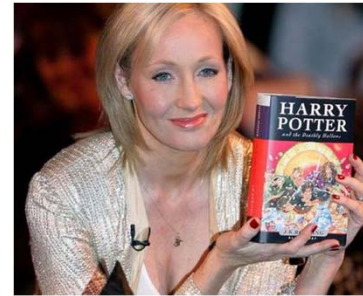
ハリーポッターの
キャラクター！



検索結果多様化

曖昧/不明快なクエリに対して、
異なるユーザの検索意図を満足させるような
検索結果リストを生成

検索結果ページ(SERP)



INTENT-1 サブタスク

- Subtopic Mining (日、中)

入力: クエリ (例: “ハリーポッター”)

出力: ランク付けされたサブピック文字列のリスト
(例: “ハリーポッター 本, ハリーポッター映画,
ハリーポッター キャラクター”)

- Document Ranking (日、中)

入力: クエリ (例: “ハリーポッター”)

出力: 多様化された、ウェブページのランキング

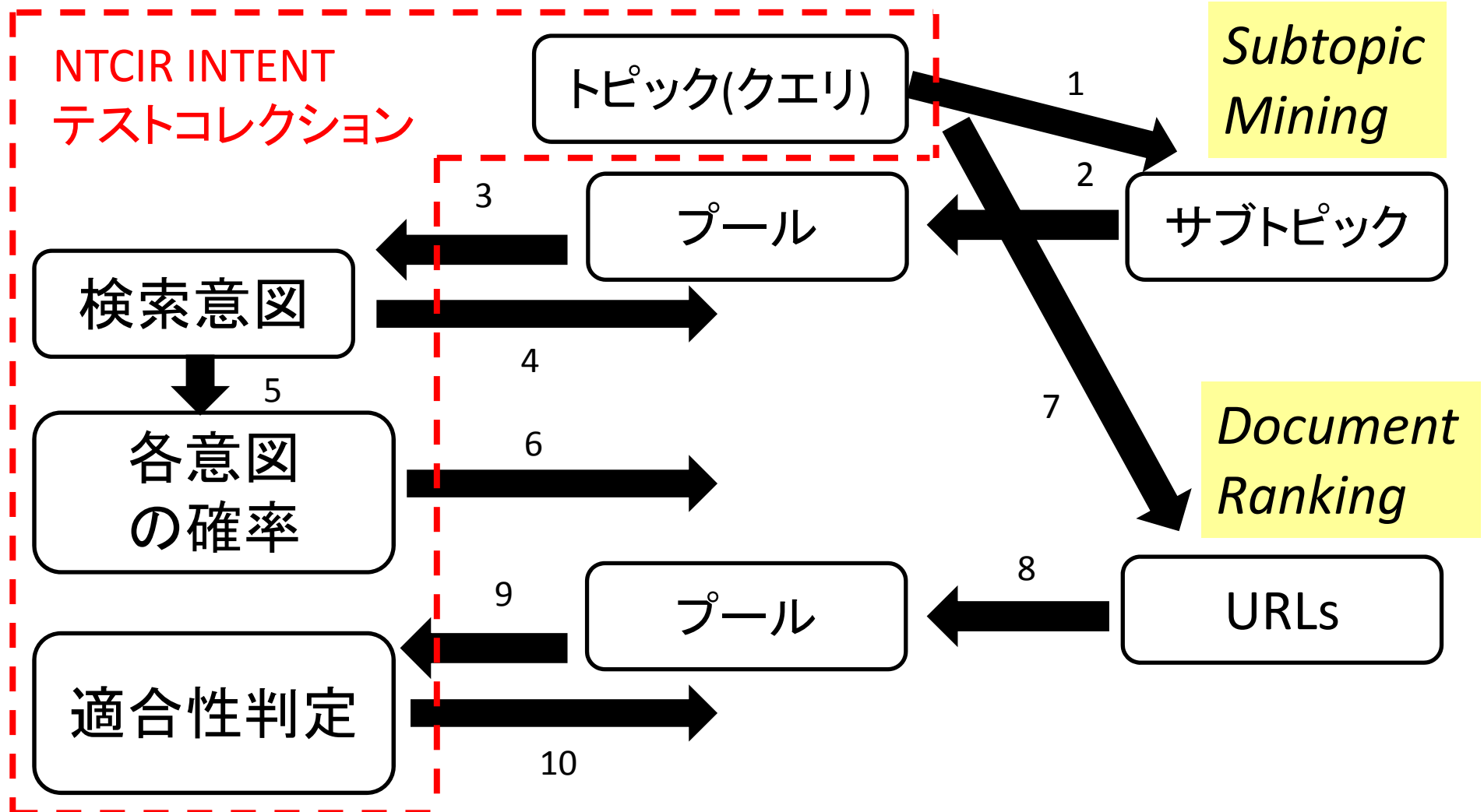
研究課題: 適合性と多様性のバランスの最適化

INTENT タスクの流れ

判定作業者

オーガナイザ

参加チーム



Q: “モーツァルト”

意図の数=3

意図:

コンサート

映画

音楽学校

確率:

0.3

0.5

0.2

多様化
された
検索結果

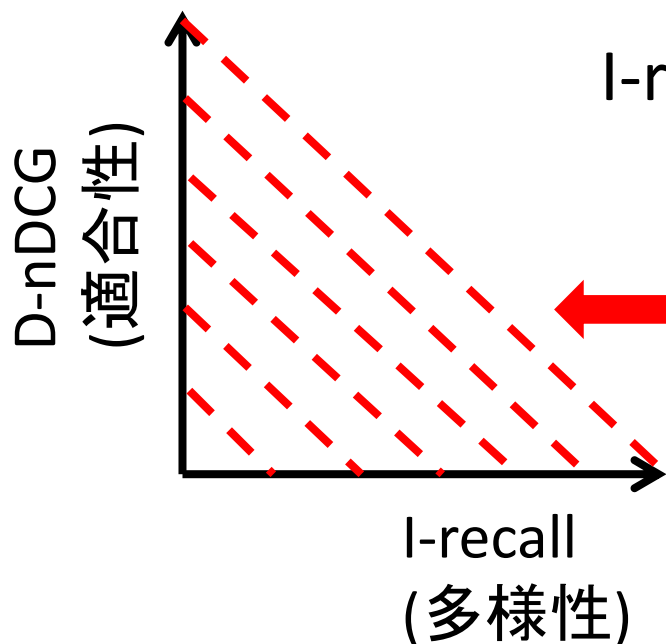
URL1	L1-Relevant	L1-Relevant	
URL2	L2-Relevant		
URL3			L2-relevant

Global Gain (URL1)

$$\begin{aligned} &= 0.3 * \text{Gain}(\text{L1}) + \\ &\quad 0.5 * \text{Gain}(\text{L1}) \\ &= 0.3 * 1 + 0.5 * 1 \\ &= 0.8 \end{aligned}$$

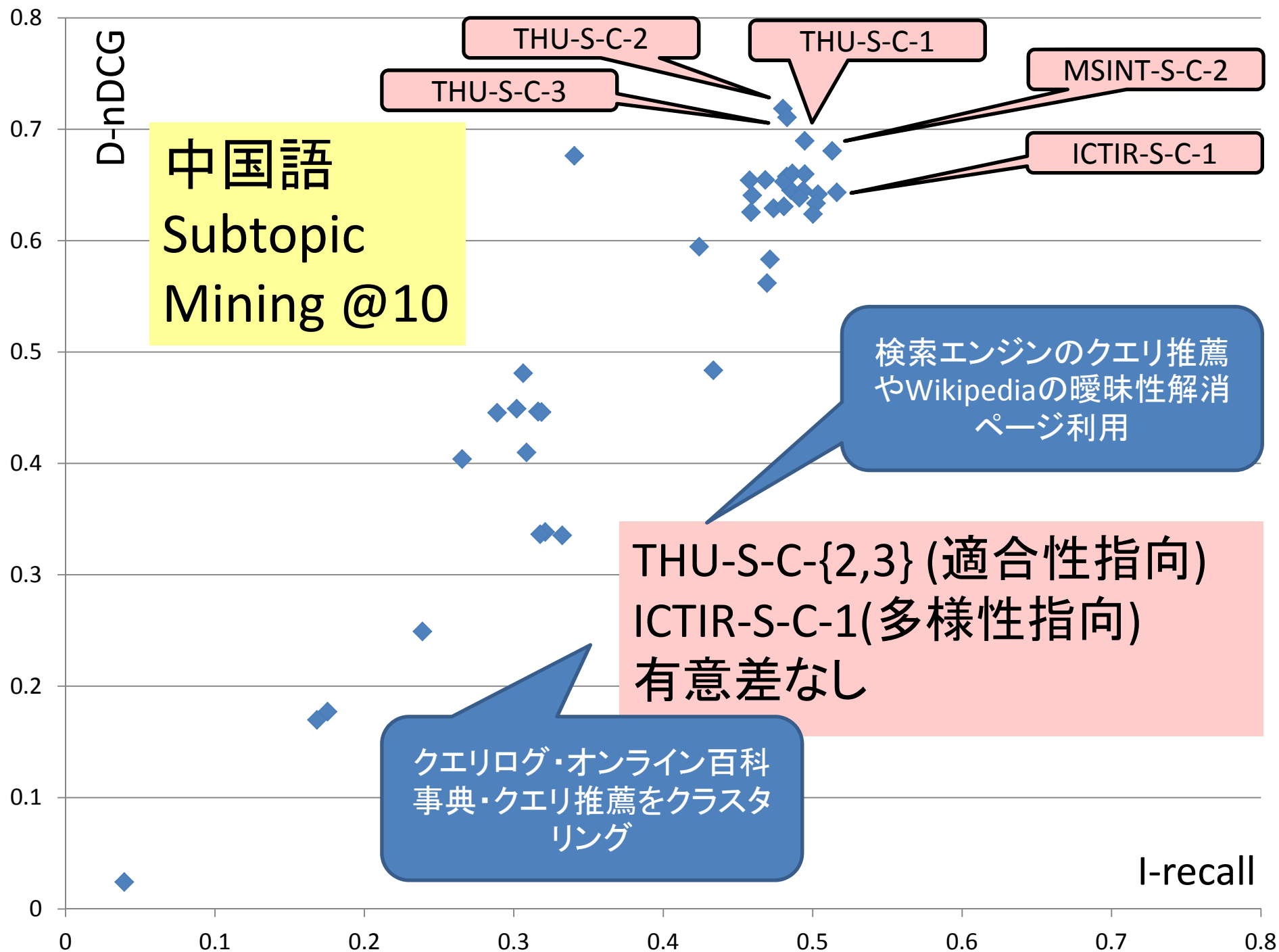
$$\text{l-recall} = 3/3 = 1$$

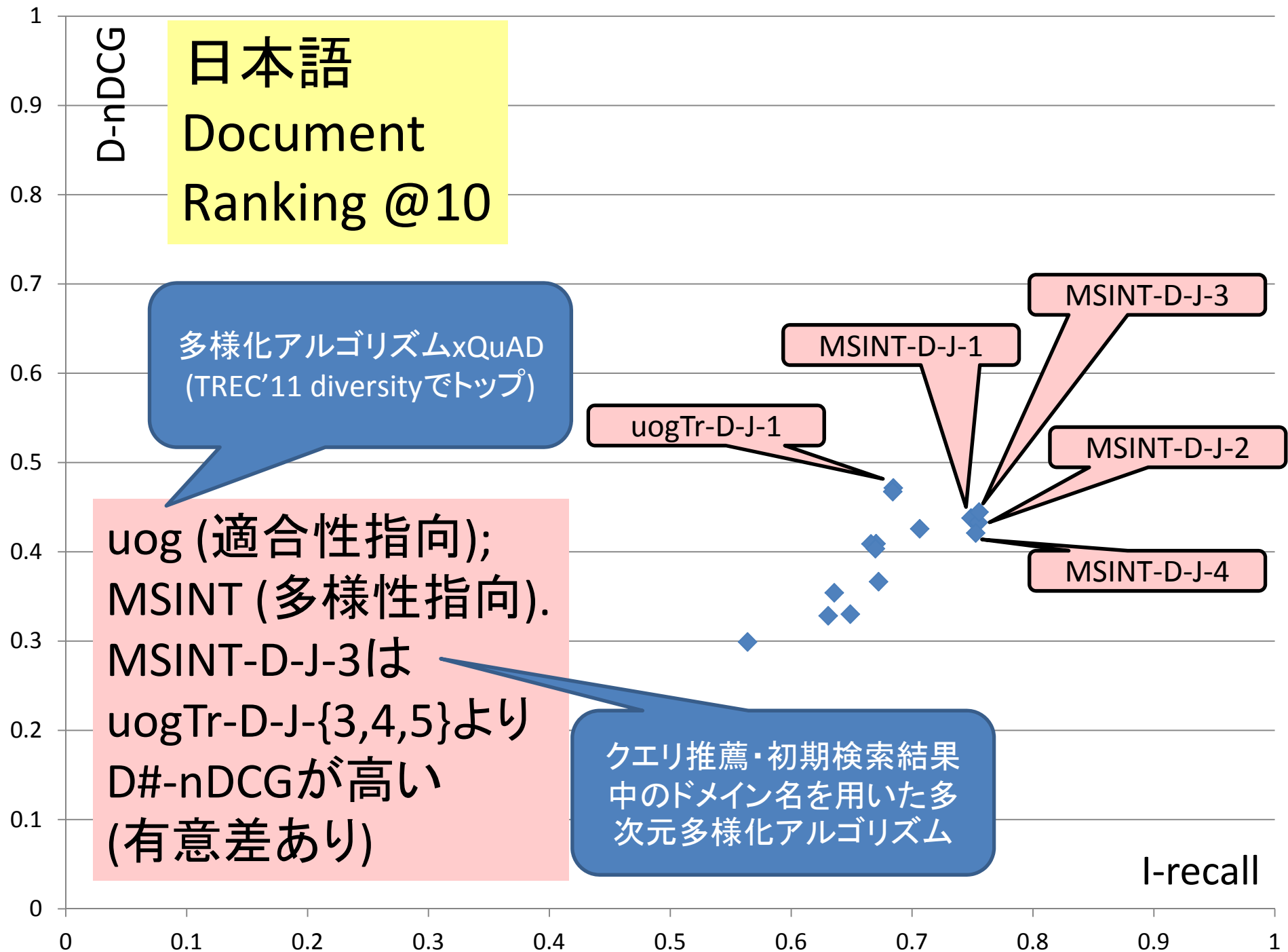
Global Gainにより
D-nDCGを計算



$$D\#-nDCG = (\text{l-rec} + \text{D-nDCG}) / 2$$

[Sakai/Song SIGIR'11]





発表概要

1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには



INTENT-2@NTCIR-10 キックオフ

オーガナイザ:

酒井 哲也, Ruihua Song, Zhicheng Dou

(Microsoft Research Asia)

Min Zhang, Yiqun Liu (精華大)

山本 岳洋, 加藤 誠 (京大)

岩田 麻佑 (阪大)

<http://research.microsoft.com/en-us/people/tesakai/intent2.aspx>

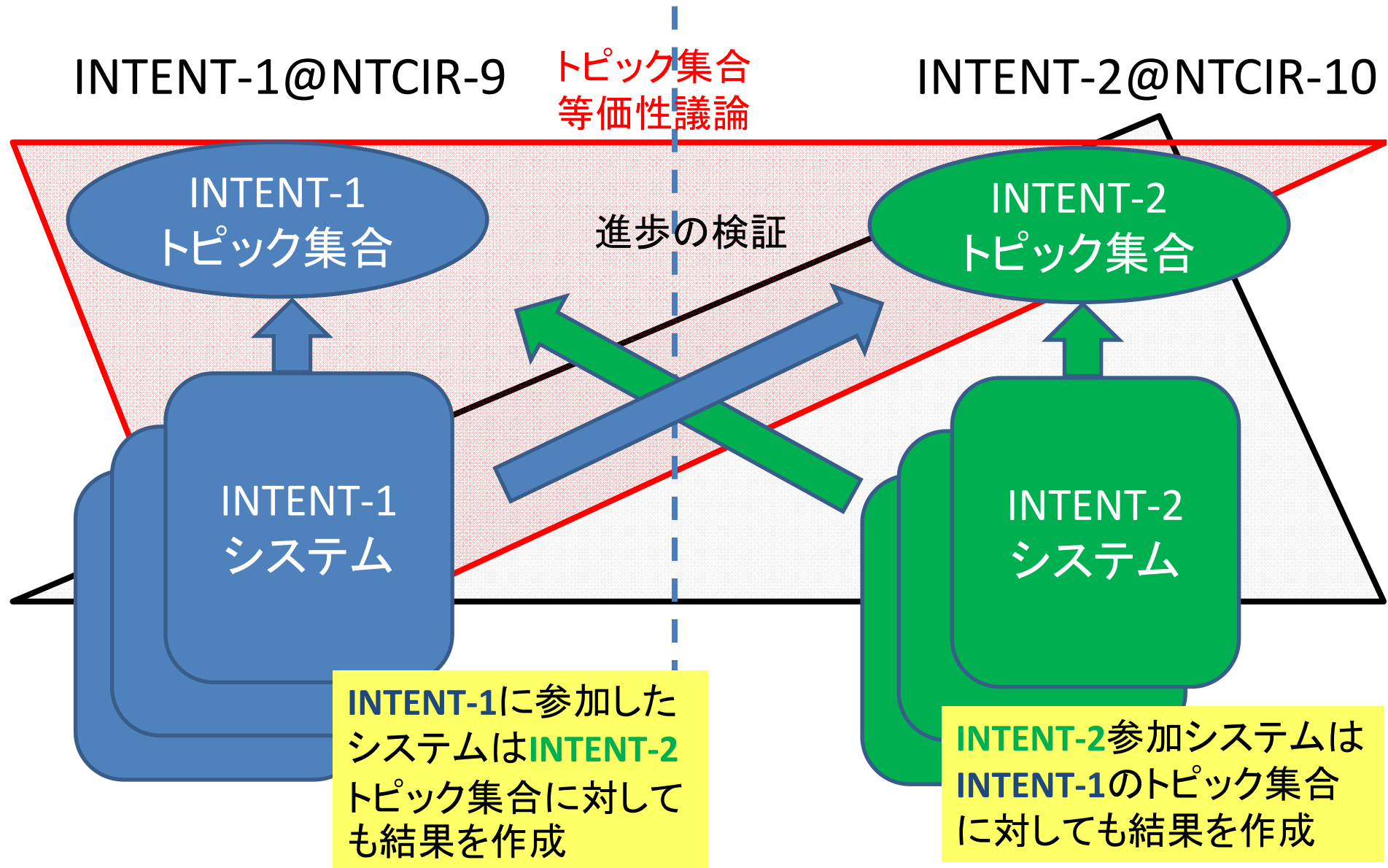


2012年3月8日@NII, 東京

INTENT-1 との相違点 (1/2)

- 英語サブトピックマイニングを導入、
TREC 2012 diversity task と同じトピックを利用
- 検索結果多様化を必要としないトピック(one item search クエリ)あり
 - 選択的多様化を奨励
- オーガナイザ側がクエリ推薦および検索結果を
ベースラインとして参加者に配布
 - 自前で検索エンジンを持たない研究チームでも
気軽に参加可

INTENT-1 との相違点 (2/2)



スケジュール

参加者

オーガナイザ

2012年 5月	トピックおよびベースラインの配布
2012年 7月	Run提出締切
2012年8～12月	サブトピックをもとに意図を同定 意図毎の適合性判定
2013年 1月	評価結果を参加者に配布
2013年 3月	論文初稿締切
2013年 5月	論文最終稿締切
6月18～21日	NTCIR-10

INTENT-1では**7つの国・地域**から**17の研究チーム**が参加！
INTENT-2ではより多くの参加者を求めています！！

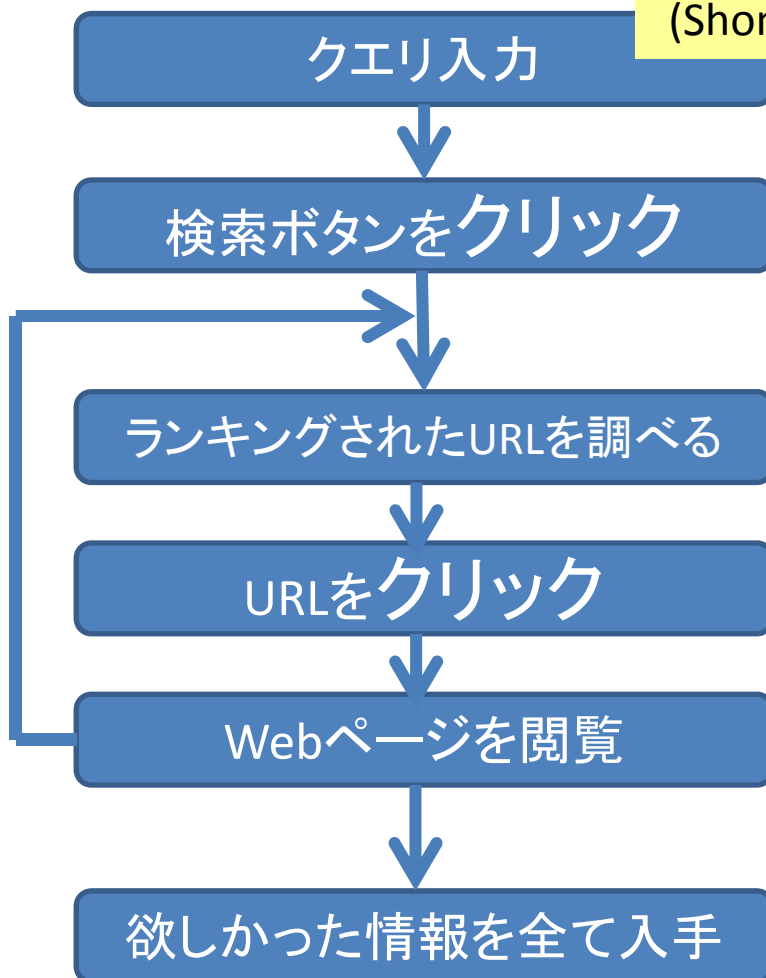
発表概要

1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには

従来の検索

= More-than-One Click Access

湘南厚木病院
(Shonan Atsugi Hospital)



The screenshot shows a search engine interface with the query '湘南厚木病院' (Shonan Atsugi Hospital). The results list several links related to the hospital, including general information, medical services, and contact details. The first result is titled '湘南厚木病院' and provides information about the hospital's location, contact numbers, and a link to the website. The second result is titled '診療科紹介 | 湘南厚木病院' and lists various medical departments. The third result is titled '徳洲会グループ - 特定医療法人 沖縄徳洲会 湘南厚木病院' and provides information about the hospital's affiliation and services. The fourth result is titled '湘南厚木病院 看護部案内' and provides information about the nursing department. The fifth result is titled 'トップページ of 湘南鎌倉総合内科+湘南厚木内科' and provides information about the hospital's general internal medicine department.

One Click Access

比較的
明快なクエリ

湘南厚木病院
(Shonan Atsugi Hospital)

クエリ入力

検索ボタンをクリック

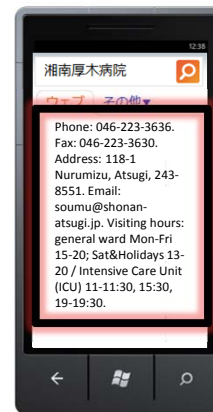
欲しかった情報を全て入手

システムはX文字の文章を出力

湘南厚木病院

ウェブ その他▼

Phone: 046-223-3636. Fax: 046-223-3630. Address: 118-1 Nurumizu, Atsugi, 243-8551. Email: soumu@shonan-atsugi.jp. Visiting hours: general ward Mon-Fri 15-20; Sat&Holidays 13-20 / Intensive Care Unit (ICU) 11-11:30, 15:30, 19-19:30.



特にモバイル検索
に有効

「ランキングされたURLリスト」ではなく「情報」そのものを検索

1クリックアクセスの利点

欲しい情報へ素早く直接アクセス



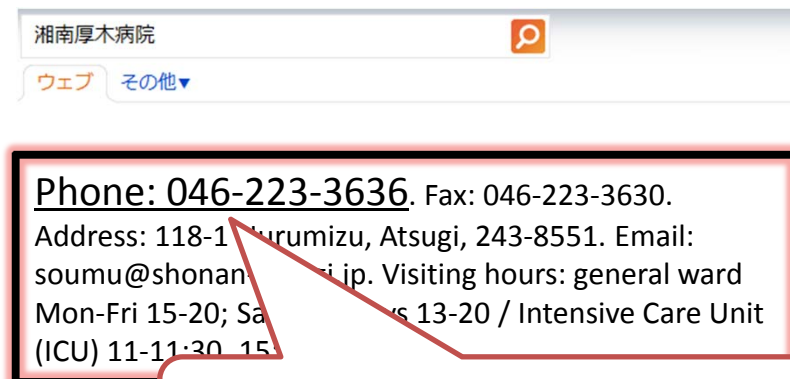
すぐに湘南厚木病院に電話しなくては！

従来の検索エンジン



やっと見つけた

1クリックアクセスシステム



すぐ見つかった！

文書検索≠情報検索 iUnitによる評価

Submission Evaluation

Job ID: 1501

Query: 北川景子

Description: 1C0-0001 CE 1 北川景子

[Back to job list](#)

Content: 407 chars

2011年3月1日...公式サイト。、出...報、ブログ。/
の公式サイト。ニュース、フォト...の売れ筋商
品、出演DVD、写真集、CD、関連書...
ど。<http://www.amazon.co.jp/pf/heRosebud>。ファ
ンサイト。笑顔の写真集、事典、など.../2008年7月
14日...名前：(きたがわけいこ) 出身地：兵庫県神戸
市愛称：けいこちゃん、きーたん生年月日：1986年8月
22日血液型：o型身長：160cm体重：非公開所属事務所：
スターダストプロモーション.../2011年3月1日...
Yahoo!人物名鑑.2011年3月3日...の、画像、動画、
ニュース、作品情報、ブログ、Feペディア（出典：
Wikipedia）等、Yahoo!人物名鑑でチェック！
<http://talent.yahoo.co.jp/pf/detail/.../2010年1月8日>...きたがわ・けいこ・ヒートアイランド・
(c) 2007「ヒートアイランド」製作委員会・木村了/城
田優/。■生年月日：1986/08/22。■出身地：兵庫
県。IMDBで検索[IMDBSearch]。をYahoo

システム出力(X-string)

判定者が
iUnitにマッチ
する部分を選
択

Readability

☒ -2

☐ -1

☐ 1

☐ 2

Trustworthiness

☒ -2

☐ -1

☐ 1

☐ 2

Done

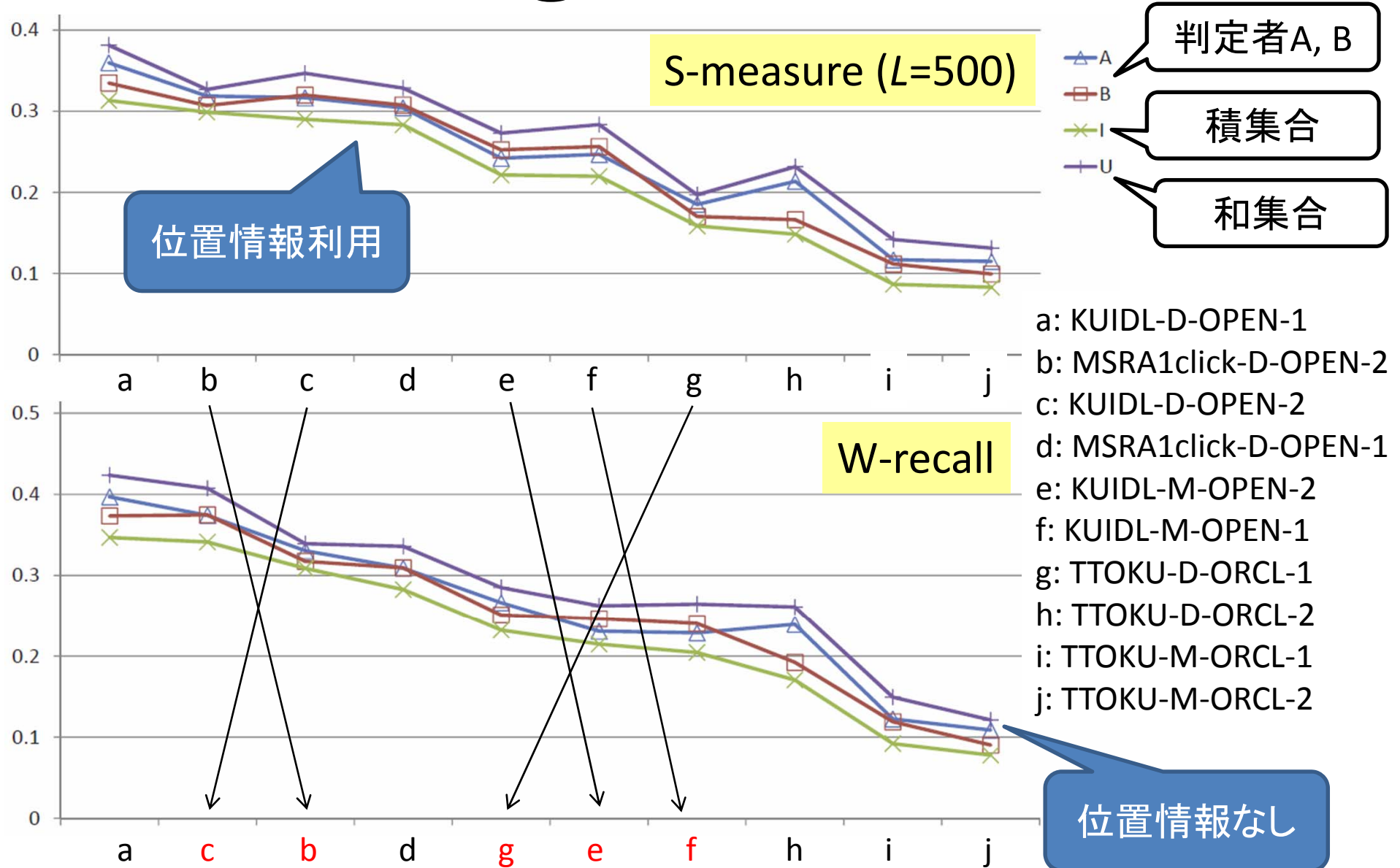
Memo

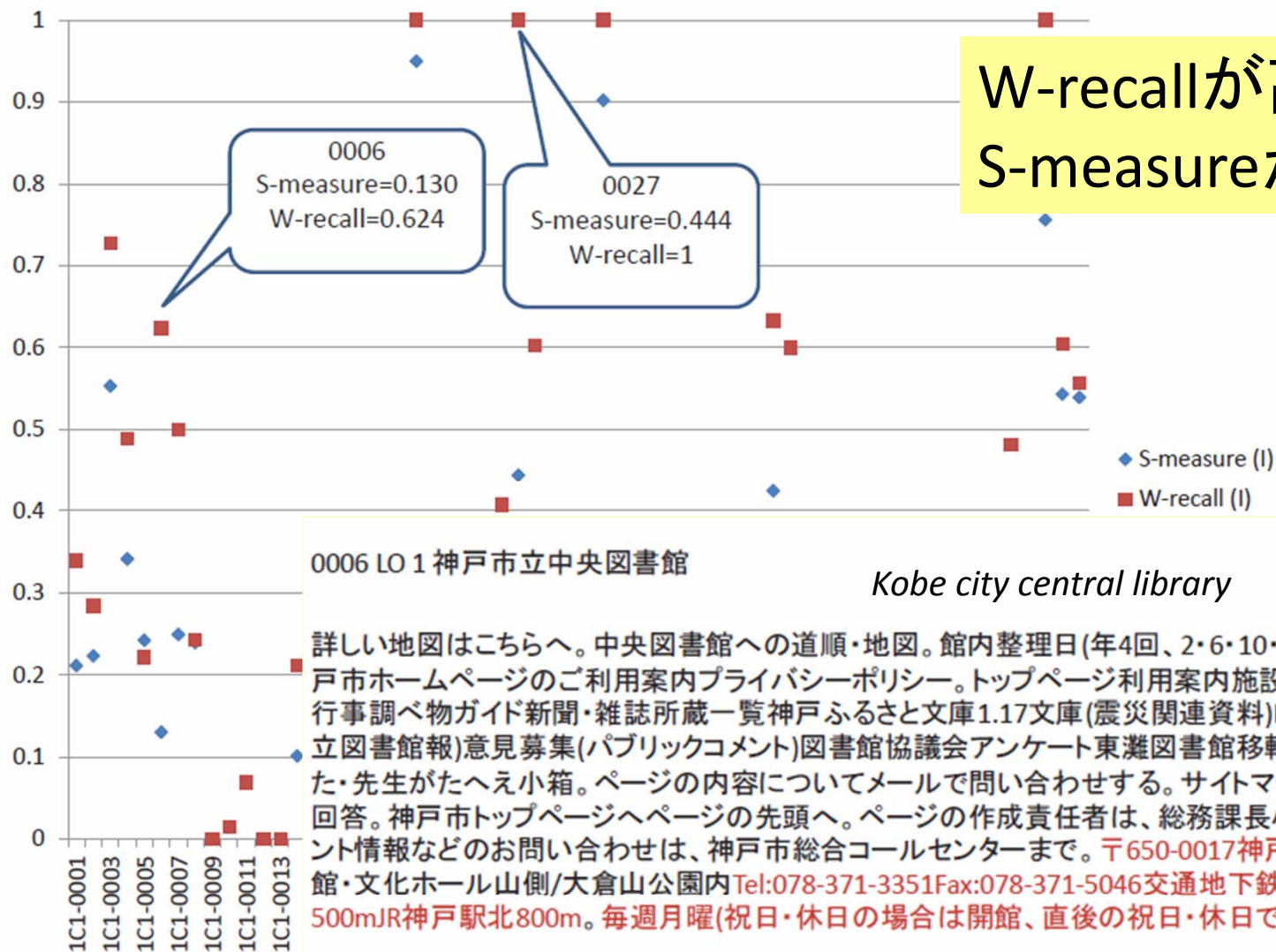
1. ☒ 1986年8月22日生
・生年月日:1986年8月22日 [184,199]
2. ☒ 兵庫県出身
3. ☒ 血液型O型
4. ☒ twitterアカウント kitagawa_keiko
5. ☒ 公式ブログ <http://star-studio.jp/kitagawa-keiko/>
6. ☒ 女優
7. ☒ 公式モバイルサイト <http://stardust-web.net/>
8. ☒ 身長160cm
9. ☒ 2003年 ミスSEVENTEENに選ばれデビュー
10. ☒ 北川景子は本名
11. ☒ スターダストプロモーション所属
12. ☒ 出演映画 Dear Friends ディア フレンズ
13. ☒ 出演TV モップガール
14. ☒ 2006年 SEVENTEENモデル卒業
15. ☒ 2009年 明治大学卒業
16. ☒ 趣味・特技 映画鑑賞
17. ☒ 趣味・特技 音楽鑑賞
18. ☒ 趣味・特技 絵画鑑賞
19. ☒ 趣味・特技 読書
20. ☒ 出演映画 チェリーパイ
21. ☒ 出演映画 花のあと

位置情報を記録

正解iUnits

1CLICK-1@NTCIR-9公式結果





W-recallが高いのに
S-measureが低い事例

適合情報が
出力の末尾
にしかない!

0006 LO 1 神戸市立中央図書館

Kobe city central library

詳しい地図はこちらへ。中央図書館への道順・地図。館内整理日(年4回、2・6・10・12月、3木曜)。サイトマップ神戸市ホームページのご利用案内プライバシーポリシー。トップページ利用案内施設案内アクセス開館時間と休館日行事調べ物ガイド新聞・雑誌所蔵一覧神戸ふるさと文庫1.17文庫(震災関連資料)KOBE本棚事業概要書燈(神戸市立図書館報)意見募集(パブリックコメント)図書館協議会アンケート東灘図書館移転整備子どものページうちのかた・先生がたへえ小箱。ページの内容についてメールで問い合わせする。サイトマップお問い合わせよくある質問と回答。神戸市トップページへページの先頭へ。ページの作成責任者は、総務課長小林史明です。市政やくらし、イベント情報などのお問い合わせは、神戸市総合コールセンターまで。〒650-0017神戸市中央区楠町7丁目2-1中央体育館・文化ホール山側/大倉山公園内Tel:078-371-3351Fax:078-371-5046交通地下鉄大倉山駅北200m高速神戸駅北500mJR神戸駅北800m。毎週月曜(祝日・休日の場合は開館、直後の祝日・休日でない日を休館)。

N003 467 休館日 毎週月曜(祝日・休日の場合は開館、直後の祝日・休日でない日を休館)。毎週月曜だけではNG

N007 353 住所 〒650-0017神戸市中央区楠町7丁目2-1

N008 388 phone 078-371-3351

N009 403 fax 078-371-5046

N010 417 アクセス 地下鉄大倉山駅 北200m

N011 427 アクセス 高速神戸駅 北500m

N012 437 アクセス JR神戸駅 北800m

正解iUnits

発表概要

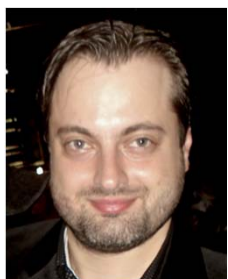
1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには



1CLICK-2@NTCIR-10 キックオフ

TASK ORGANISERS: 加藤 誠, 山本 岳洋 (京大),
酒井 哲也, Young-In Song, Zhicheng Dou (MSRA),
Virgil Pavlu, Matthew Ekstrand-Abueg, Shahzad Rajput (Northeastern U)
岩田 麻佑 (阪大)

<http://research.microsoft.com/en-us/people/tesakai/1click2.aspx>



March 8, 2012@NII, Tokyo

1CLICK-1との相違点

- 日本語に加え**英語**のタスクを導入
- クエリタイプを4種類から**8種類**に
- 情報の基本単位**iUnit**の明確な定義
- 手動**および**自動によるシステム出力の評価
- より豊富なサブタスク
 - より公正な評価を目指しつつ気軽な参加も可
 - 検索結果・ウェブページデータの提供
 - クエリ分類サブタスクのみの参加も可

1CLICK-2クエリ

- 8種類のクエリタイプ

(これらの種類は特に検索結果ページを見ただけでユーザが満足する割合が高い [Li/Huffman/Tokuda SIGIR'09])

- 芸術家, スポーツ選手, 政治家, 俳優
- 地物, 地域情報
- 定義 (語の)
- QA (自然言語で書かれた質問)

- 実際の大規模モバイルクエリログから選出したクエリ 日・英それぞれ100件

1CLICK-2 サブタスク(日・英)

参加チームは下記のいずれかに参加

- メインタスク

- **Mandatory Runs**: オーガナイザが配布する
Web検索結果からX-stringを生成
- **Oracle Runs (optional)**:
iUnitsの抽出元URLリストからX-stringを生成
- **Open Runs (optional)**:
任意の検索エンジンを利用してX-stringを生成

- クエリ分類サブタスク

- 与えられたクエリのクエリタイプを推定
(芸術家, QA, etc.)

スケジュール

参加者

オーガナイザ

2012年 4月 サンプルクエリとiUnitの配布

2012年 8月 本番用クエリ配布

2012年 10月 Run提出締切

2012年 11月 iUnitsによる評価期間
-2013年 1月

2013年 2月 評価結果配布

2013年 3月 ドラフト論文締切

2013年 6月 NTCIR-10

情報検索, 情報抽出, モバイル検索, 要約,
次世代検索に興味がある方はぜひご参加ください

発表概要

1. NTCIR (エンティサイル)について
2. INTENT-1@NTCIR-9
3. INTENT-2@NTCIR-10
4. 1CLICK-1@NTCIR-9
5. 1CLICK-2@NTCIR-10
6. タスクに参加するには

タスクに参加するには

NTCIRホームページから参加登録(準備中)!
各タスクホームページをチェック!

<http://research.microsoft.com/en-us/people/tesakai/intent2.aspx>

ntcadm-intent@nii.ac.jp

<http://research.microsoft.com/en-us/people/tesakai/1click2.aspx>

ntcadm-1click@nii.ac.jp

Follow @tetsuyasakai #INTENT2 #1CLICK2

検索エンジンをもっていない
チームでも
簡単に
参加可能!

よくばりなあなたへ

'12年4月	1CLICKサンプルクエリ配布
'12年5月	INTENTトピック・ベースライン配布
'12年7月	INTENT run締切
'12年8月	1CLICK本番クエリ配布
'12年10月	1CLICK run締切
'12年11-'13年1月	1CLICK iUnitsによる共同評価
'13年1月	INTENT結果配布
'13年2月	1CLICK結果配布
'13年3月	論文初稿締切
'13年5月	論文最終稿締切
'13年6月18-21日	NTCIR-10

ご清聴ありがとうございました。
ご参加お待ちしております！