




Welcome!



Twitter: #ntcir10
Ust: ntcir10-kickoff

NTCIR-10 Kick-Off Event

2012.03.08

日本語セッション: 13:30-

English Session: 15:30-

プログラム

- NTCIRの概要
- NTCIR-10の特徴
- 採択タスクの紹介
- 参加のメリット
- 参加方法
- 主な日程
- Q & A

NTCIRの概要

情報検索・アクセス技術の評価と性能比較のための研究基盤 NTCIR(エンティサイクル)

目的: 大規模な評価実験用の研究基盤を提供することによって情報
アクセス技術研究の促進

■ 基盤: データ、評価手法、コミュニティ

■ システム間の性能比較, 手法の特性の相互比較 技術移転 互いに学びあう場

1997年末にプロジェクト開始

■ 一年半毎に成果報告会を開催

再利用可能な大規模データセット(テストコレクション)

■ 学術, 新聞記事, 特許, Web, ヤフー知恵袋, Wikipedia, 大学入試など

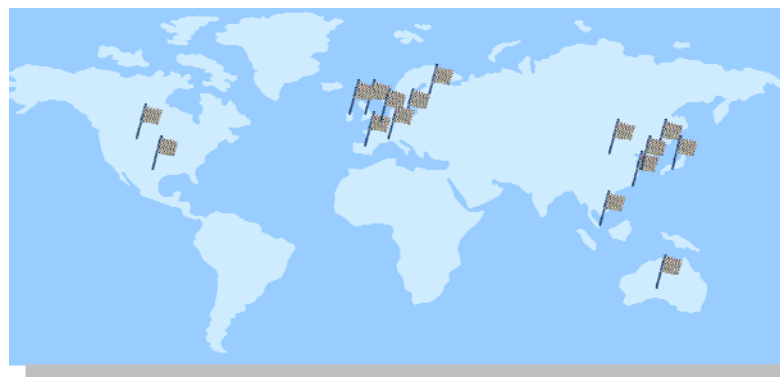
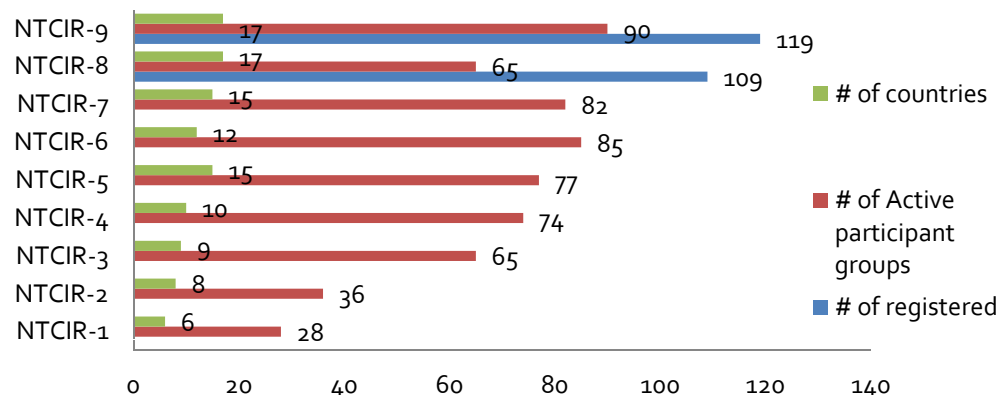
■ 研究目的利用で公開

研究部門(タスク)

■ 情報検索: 言語横断, 特許, Web

■ 質問応答: 単言語, 言語横断

■ 要約, 動向情報, 特許マップ自動生成



情報検索 Information Retrieval (IR)

- 蓄積された大量の文書から利用者の情報要求に**レレバント**な(適合する) 情報を選びだす
 - 構造化されていない、自然言語テキスト
 - 伝統的には文書検索→文書中の情報活用支援へ
- コンピュータの使用 1950年代から
- 人間の主観的判定を成否の基準にした最初の計算機科学

情報アクセス Information Access (IA)

- 蓄積された大量の文書中の情報を利用者が利用できるようにするまでの全過程
- 検索、要約、質問応答、テキストマイニング、クラスタリングなど

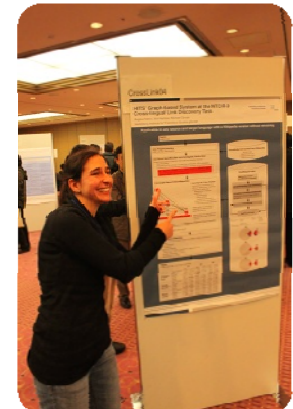
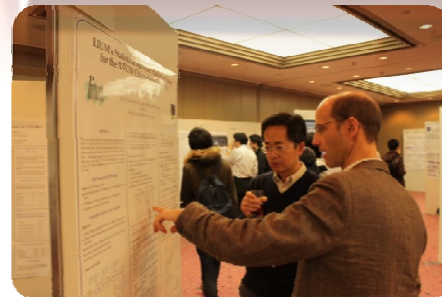
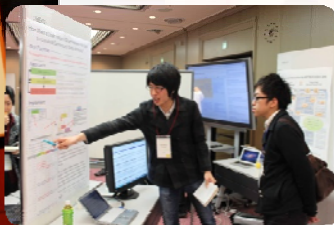
NTCIR 同次 技術分野	年	1 '99	2 '01	3 '02	4 '04	5 '05	6 '07	7 '08	8 '10	9 '11	10 '13	研究部門(タスク)
ソーシャルメディア・UGC									■			コミュニティQA
							■	■	■			意見分析
モジュール評価								■	■		■	検索+推論
												質問応答+検索
特定ドメイン										■	■	入試問題
										■	■	音声文書検索
									■	■		地理・時間情報検索
				■	■	■	■	□	□			特許検索
質問応答							■	■	■	□		あらゆるタイプの質問
					■	■						対話
						■	■	■	■			言語横断
				■	■	■	■					事実を尋ねる、リスト型
情報抽出と検索										■	■	リンク発見
意味処理										■	■	推論
要約・統合					■	■	■	■	■			テキストマイニング・分類
						■	■	■				動向情報の抽出・可視化
			■	■	■							テキスト要約
インタラクティブ										■		インタラクティブ検索と可視化
Web										■	■	検索意図・多様な意図の検索
										■	■	検索結果一覧ページの質
				■	■	■						Web検索
言語横断		■	■	■	■	■	■	■	■	■	■	機械翻訳
		■	■	■	■	■	■	■	■			言語横断検索
		■	■	■	■	■	■	■	■			英語以外の検索
テキスト検索		■	■	■	■	■	■	■				日本語検索

年は成果報告会開催年。研究部門は18カ月前に開始

NTCIRワークショップの手順

- 研究部門（タスク）のプロポーザル募集
- 研究部門・対象技術の選定
 - プログラム委員会
 - データ提供機関との協議・契約
- 研究部門ごとに、実験・評価手法の提案・議論
- 参加申込
 - 訓練用データ（文書、課題、正解）配布
 - 各参加者が実験、チューニング
 - 評価用データ（文書、課題）配布
- 検索結果の提出
- 提出結果から正解候補収集、人手判定
- 正解公開、評価結果発表
- 成果報告会
 - 次回について協議

NTCIR: 検索システムのかくらべ



<http://research.nii.ac.jp/ntcir/>

NTCIR-10 Organization

<NTCIR-10 Program Co-Chairs>

- 上保 秀夫 (筑波大学)
- 酒井 哲也 (Microsoft Research Asia, China)

<EVIA2013 Co-Chairs>

- Ruihua Song (Microsoft Research Asia, China)
- William Webber (University of Maryland, USA)

<NTCIR-10 General Co-Chairs>

- 神門 典子 (国立情報学研究所)
- 加藤 恒昭 (東京大学)
- Douglas W. Oard (University of Maryland, USA)
- Mark Sanderson (RMIT University, Australia)
- ex-general chair: Eiichiro Sumita (NICT)

-48 Task Organizers from all over the world

NTCIR テストコレクション(評価実験用データセット)

	Ad Hoc/ CLIR [Scientific Abstracts] (Japanese/ English IR)	Chinese IR	CLIR [News] (Cross- Lingual QA)	CLQA (Cross- Lingual IR)	MuST (Multimodal Summarization for Trend Information)	OPINION (Opinion Analysis)	PATENT	QAC (Question Answering)	TMREC (Term Recognition)	TSC (Summa- rization)	WEB
NTCIR-1	NTCIR-1 Ad Hoc/ CLIR	-	-	-	-	-	-	-	NTCIR-1 TMREC	-	-
NTCIR-2	NTCIR-2 Ad Hoc/ CLIR	CIRB 010	-	-	-	-	-	-	-	NTCIR-2 SUMM	-
NTCIR-3	-	-	NTCIR-3 CLIR	-	-	-	NTCIR-3 PATENT	NTCIR-3 QA	-	NTCIR-3 SUMM	NTCIR-3 WEB
NTCIR-4	-	-	NTCIR-4 CLIR	-	-	-	NTCIR-4 PATENT	NTCIR-4 QA	-	NTCIR-4 SUMM	NTCIR-4 WEB
NTCIR-5	-	-	NTCIR-5 CLIR	NTCIR-5 CLQA	-	-	NTCIR-5 PATENT	NTCIR-5 QA	-	-	NTCIR-5 WEB
NTCIR-6	-	-	NTCIR-6 CLIR	NTCIR-6 CLQA	NTCIR-6 MuST	NTCIR-6 OPINION	NTCIR-6 PATENT	NTCIR-6 QA	-	-	-

テストコレクションの多くはワークショップ終了後、研究目的で公開

NTCIRの焦点

再利用可能な実験用データ セット

日本語、東アジア諸語
言語横断／対訳コーパスの利用
多様な文書ジャンル

新しい情報アクセス技 術へのチャレンジ


検索と言語処理の融合
情報活用支援
より現実的な／利用者志向

研究者のフォーラム

研究アイデアの交換
評価手法・指標の研究・議論

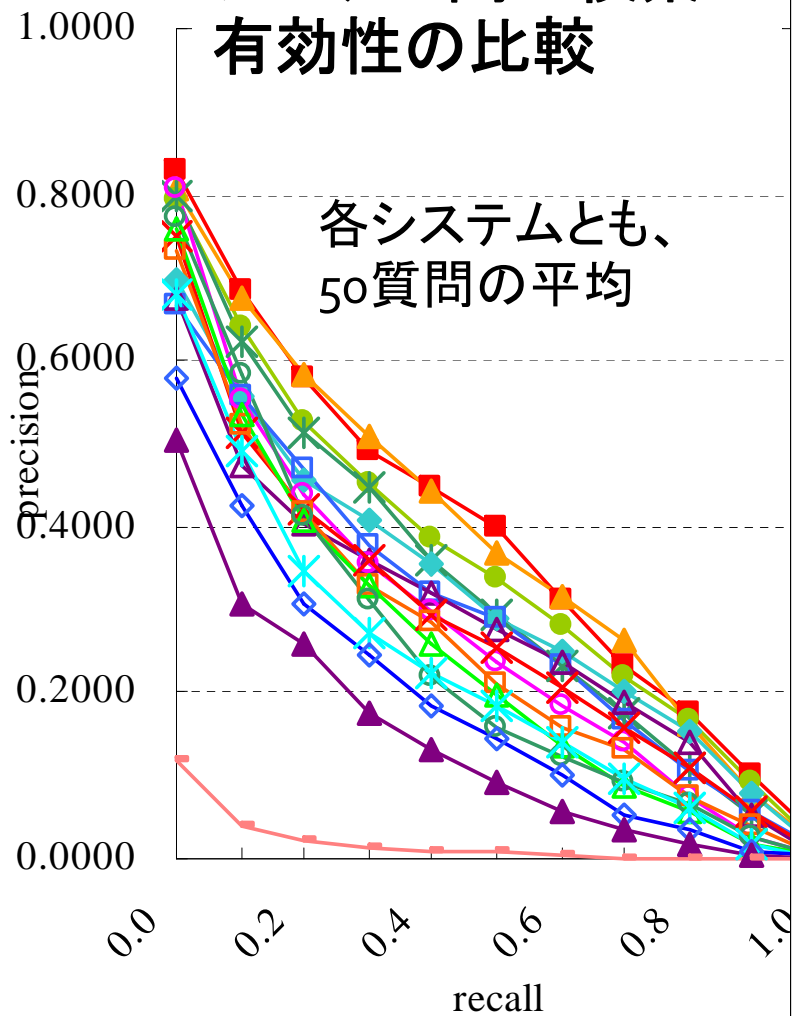
- どの要素により重点をおくかは次第に変化している
- 開始当初：データセット →次第にチャレンジ→ 今後はフォーラムと評価手法の研究がますます重要

情報アクセス技術の評価の6レベル

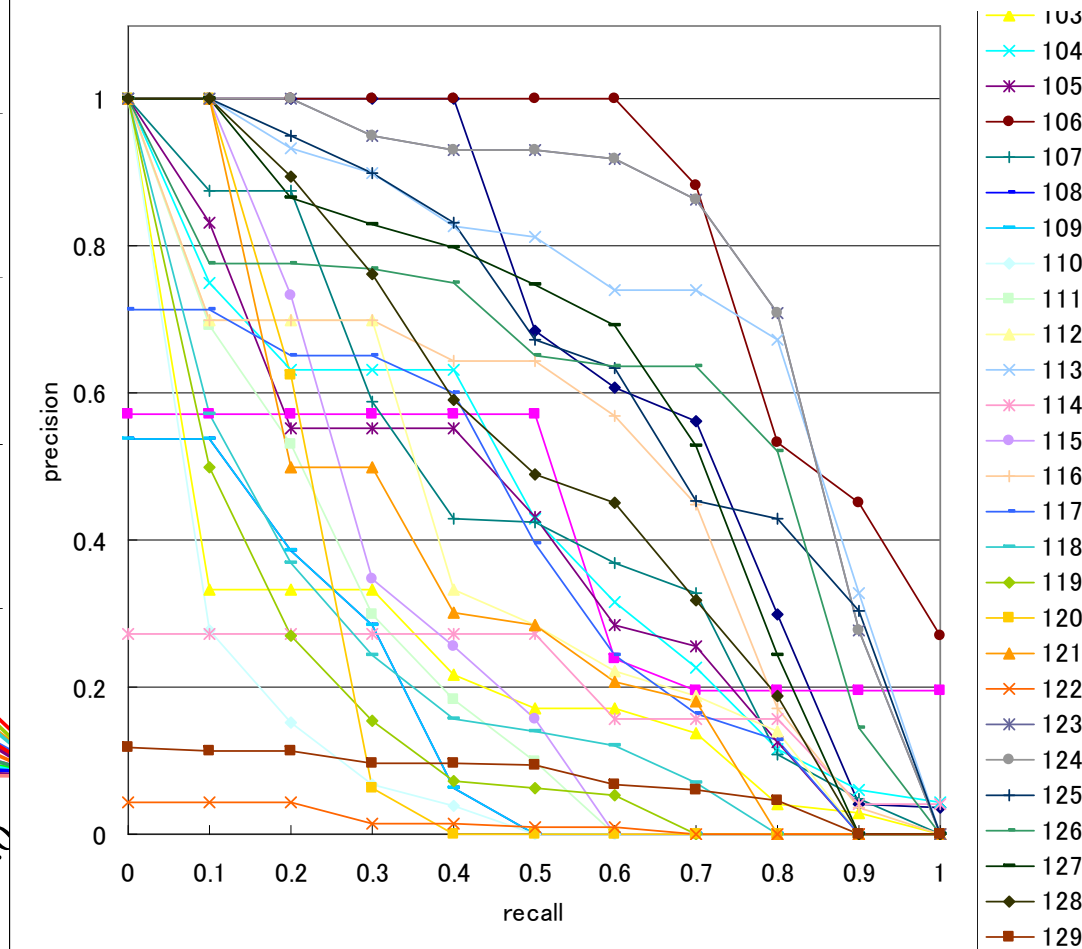
1. 工学レベル：効率*efficiency* *ex.*速さ
2. 入力レベル： *ex.*データベースの網羅性、質、新しさ…
-  3. 処理レベル:検索有効性*effectiveness* *ex.*再現率、精度
4. 出力レベル：結果の提示
5. 利用者レベル： *ex.* 利用者が必要とする労力
6. 社会レベル：重要性

(Cleverdon & Keen 1966)

システム間の検索有効性の比較



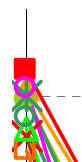
1つの検索システム上で、50個の検索質問の検索有効性の比較



検索の難しさは、検索質問ごとに異なる

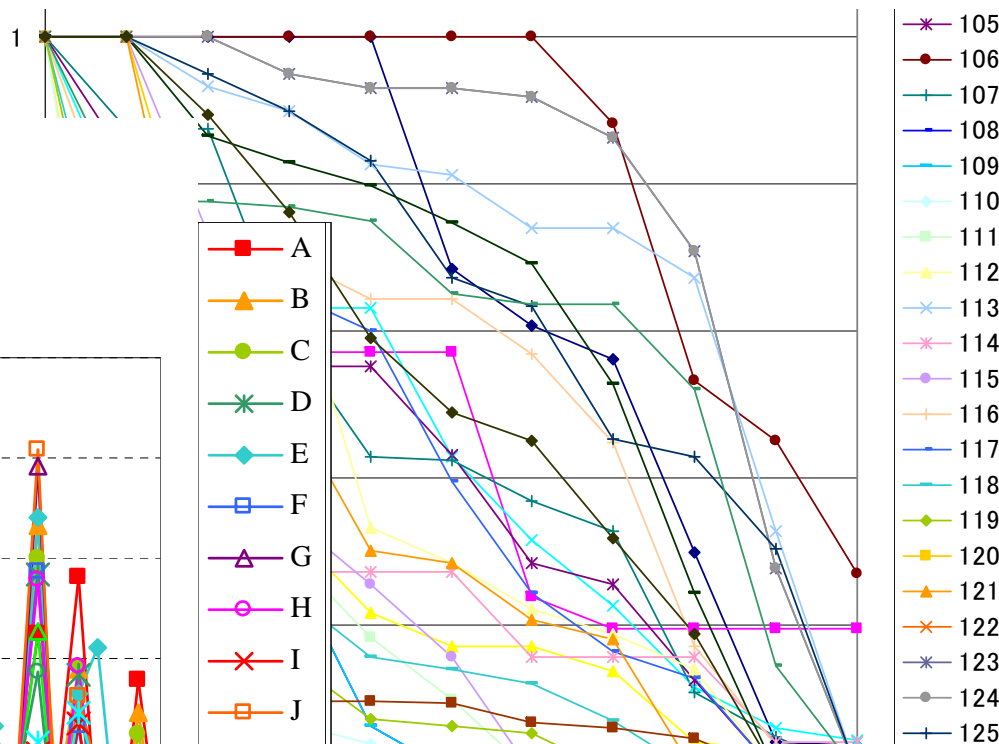
システム間の検索有効性の比較

1.0000
0.8000

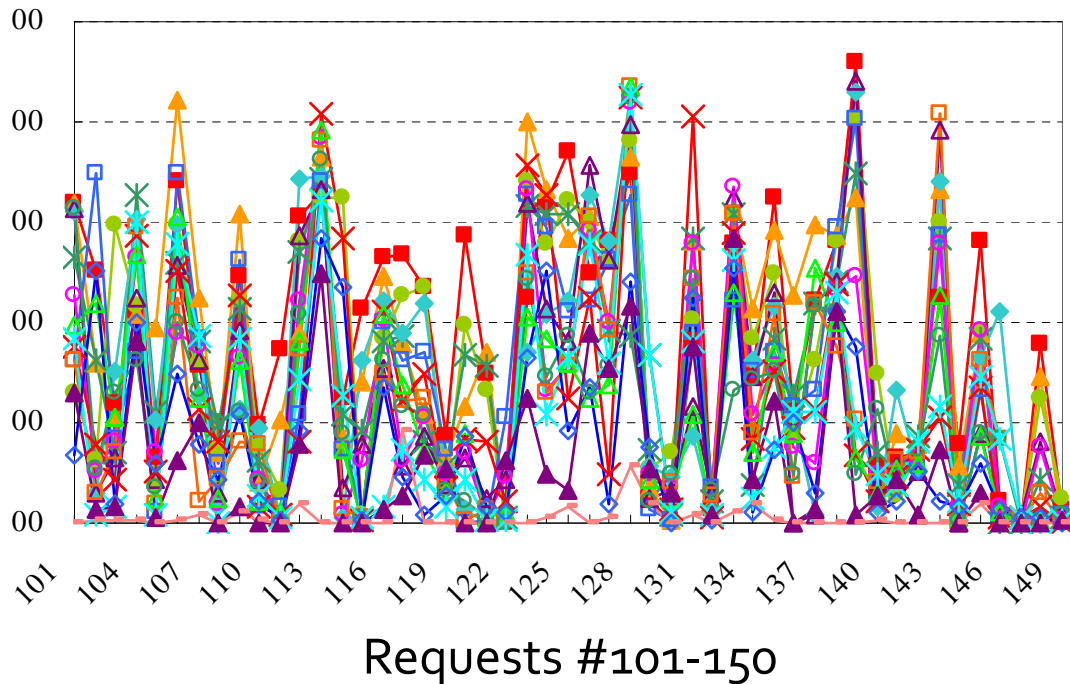


「難しい質問」は、システムによって異なる

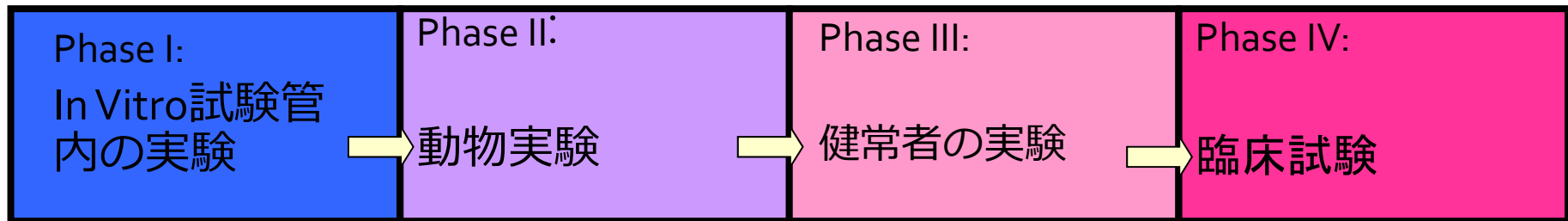
1つの検索システム上で、50個の検索質問の検索有効性の比較



信頼性の高い評価をするには、多数の質問を用いる必要がある。評価指標によって必要な質問数は異なる



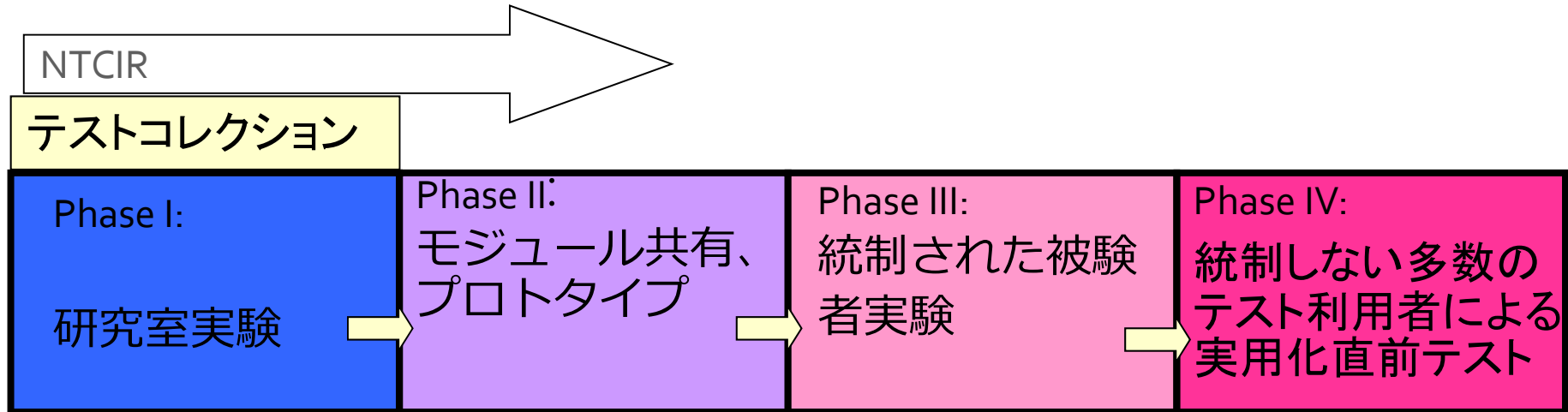
製薬企業の研究開発



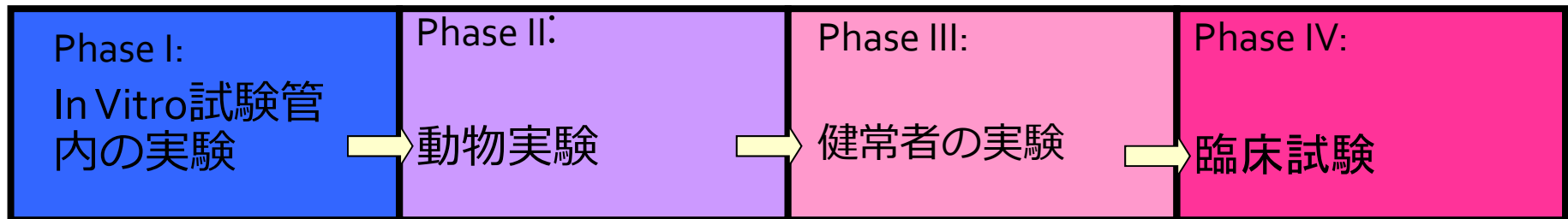
統制しやすい

実験設計：
評価する側面を選定

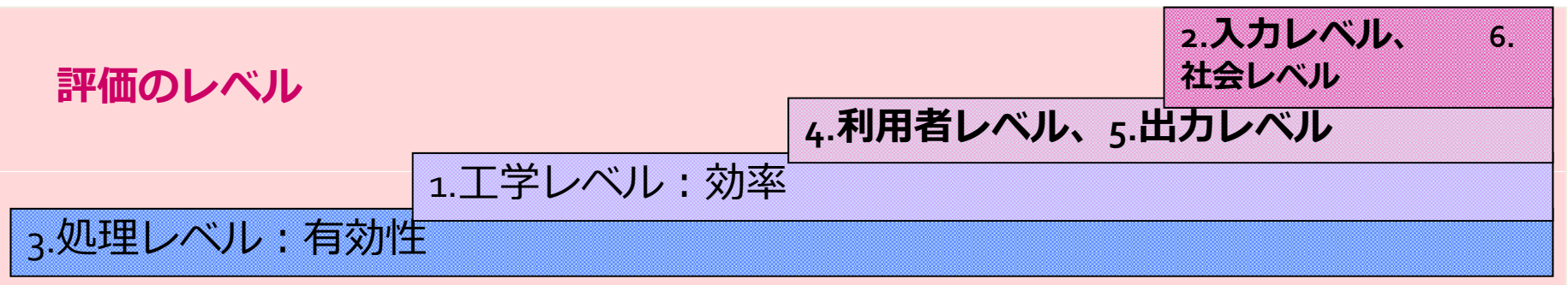
統制できない
要因 + + +



製薬企業の研究開発



評価のレベル



実験設計

○利用者の情報探索タスク

- 文書データのジャンル(種類)+典型的な利用者の利用
- 利用者の状況、アクセスの目的

実験は現実世界での情報アクセスを抽象化したもの

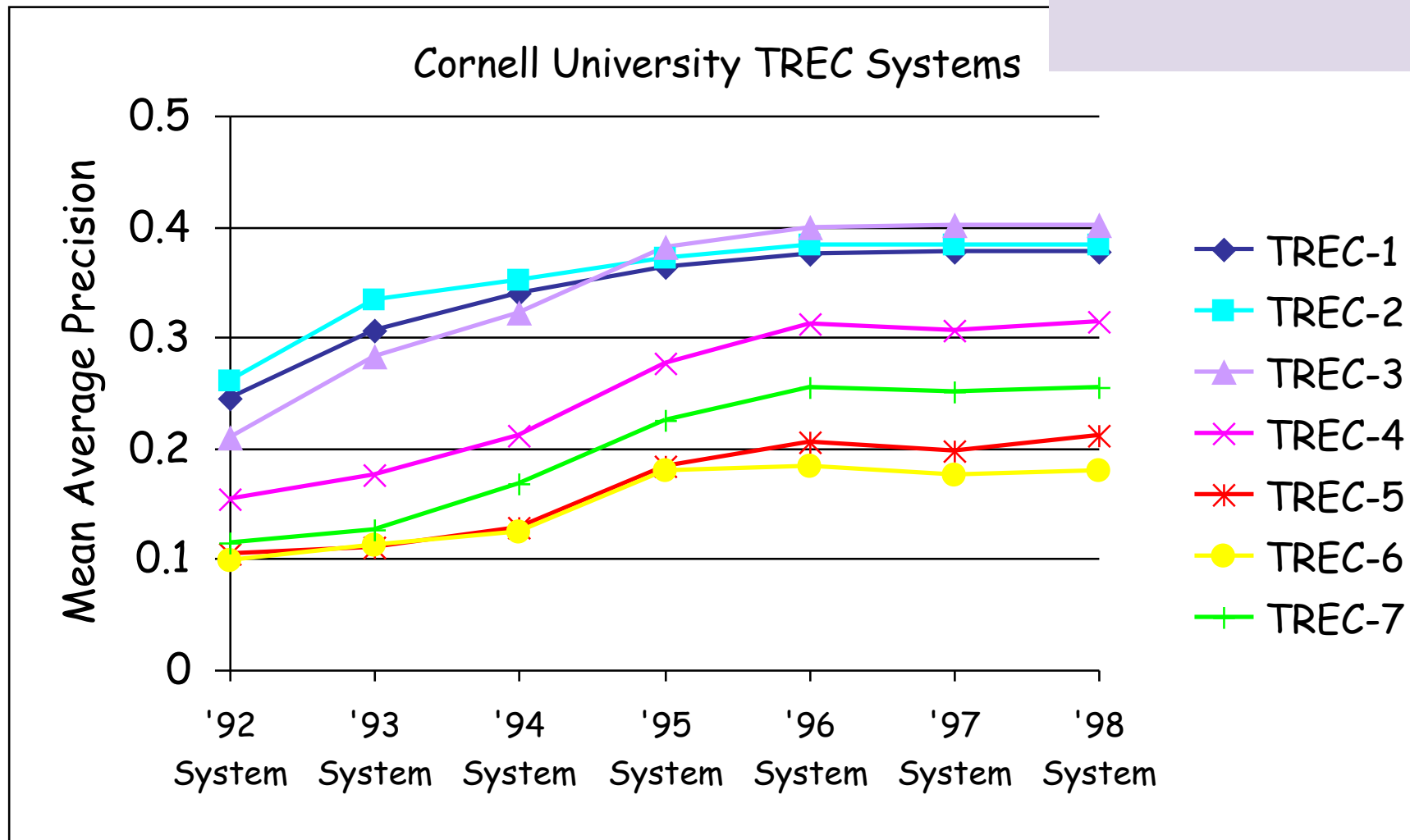
評価する側面を選定。現実世界に近づくほど、統制できない要因が増加する

○検証とベンチマーク

- なぜ、どの程度、このシステムは他よりよく（悪く）働くのか
- どのようにシステムを改善すればよいのかを学ぶ
- システムの有効性について、科学的な理解を提供する

評価ワークショップによる技術の有効性の進展

およそ3年で1.5~2倍



今後の課題

- 評価への個別的なニーズへの対応
- 探索的・対話的な情報アクセス
 - 利用者の意図
 - 協調探索
 - 経験・知識・リテラシー
- SWIRL 2012で議論された重要トピック
 - 対話検索
 - モバイル検索
 - ゼロクエリ検索
 - スマート検索
 - Query by walking around
 - 構造, エンティティ, 関係

NTCIR-10の特徴

What's new?

NTCIR-10: ねらい

- 強固な運営基盤
 - 運営組織の強化
- タスクの多様性
 - 情報アクセスにおける幅広いコンテンツを対象
 - 多様なメディアを研究対象に
- コミュニティ主導のタスク運営
 - 研究の持続性
- 研究資源利用の促進
 - 成果報告会で特別ブースの設置

NTCIR-10: 運営組織



TO

- General Co-Chairs
 - 加藤 恒昭 (東京大学)
 - 神門 典子 (NII)
 - Douglas W. Oard
(University of Maryland)
 - Mark Sanderson (RMIT)
- Program Co-Chairs
 - 酒井 哲也 (MSRA)
 - 上保 秀夫 (筑波大学)
- タスクオーガナイザ
 - 世界中の48人の研究者
 - 参加者 (あなたです!)
- EVIA 2013 Co-Chairs
 - Ruihua Song (MSRA)
 - William Webber
(University of Maryland)



<http://research.nii.ac.jp/ntcir/ntcir-10/organizers.html>

NTCIR-10: これまでの展開

2011.10	NTCIR-10組織の発足
2011.11	タスクプロポーザルの募集開始 10のプロポーザルが応募
2011.12	NTCIR-9 成果報告会
2012.01	プログラム委員会とProgram Co-Chairsの審査の結果、7つのタスクが採択
2012.02	各タスクにおける参加募集要項の準備
2012.03	NTCIR-9 Kick-Off Event

NTCIR-10 プログラム委員

- Charles Clarke (University of Waterloo, Canada)
- Kalervo Järvelin (University of Tampere, Finland)
- Hideo Joho (Co-chair, University of Tsukuba, Japan)
- Gareth Jones (Dublin City University, Ireland)
- Noriko Kando (NII, Japan)
- Tsuneaki Kato (The University of Tokyo, Japan)
- Douglas W. Oard (University of Maryland)
- Tetsuya Sakai (Co-chair, Microsoft Research Asia, PRC)
- Mark Sanderson (RMIT, Australia)
- Ian Soboroff (NIST, US)

NTCIR-10 採択タスク

NTCIR-10 採択タスク一覧

コアタスク

- [Intent-2] Search intent and diversification
- [1Click-2] One-Click Access
- [RITE-2] Recognizing Inference in Text
- [SpokenDoc-2] IR for Spoken Documents
- [PatentMT-2] Cross-lingual access to Patent Docs
- [CrossLink-2] Cross-lingual Link Discovery

パイロットタスク

- [Math] Access to mathematical contents

CrossLink-2

Crosslink Introduction

- Cross-lingual link discovery (Crosslink or CLLD) is concerned with automatically finding potential links between documents in different languages.
- CLLD algorithms actively recommend a set of meaningful anchors in the context of a source document and establish links to documents in an alternative language.
- CLLD is helpful for complimentary knowledge discovery in different language domains.

Cross-lingual Link Discovery

Article: Australia

...
Ranked third in the [Index of Economic Freedom](#) (2010),[178] Australia is the [world's thirteenth largest economy](#) and has [the ninth highest per capita GDP](#); higher than that of the United Kingdom, Germany, France, Canada, Japan, and the United States. The country was ranked second in the United Nations 2010 [Human Development Index](#) and first in [Legatum's 2008 Prosperity Index](#). [179] All of Australia's major cities fare well in global comparative livability surveys: [180] Melbourne reached first place on [The Economist's 2011 World's Most Livable Cities](#) list, followed by Sydney, Perth, and Adelaide in sixth, eighth, and ninth place respectively. [181] Total government debt in Australia is about \$190 billion. [182] Australia has among the highest house prices and some of the highest household debt levels in the world.
...

No link was created for this term, for finding articles in languages we prefer traditionally we do:

Search

Translate

经济学家
...

The
Economist
...

エコノミスト
...

이코노미
스트
...

How to automatically create cross-lingual links for a document if no links existing yet?

- Links in other languages?
- New articles?
- Missing links?
- Not what we are looking for?
- What about other relevant links?

Cross-lingual Link Discovery

- ❖ Cross-lingual Links
- ❖ New Links
- ❖ Better Links
- ❖ More options



- All about **multi-lingual knowledge discovery** in knowledge bases (e.g. Wikipedia)
- All about **easy** and **efficient** information access

The Goal of Crosslink Task

- It is aimed to create a reusable resource for evaluating automated cross language link discovery approaches. The results of this research will be used in building and refining systems for automated link discovery.

Crosslink task at NTCIR-9

- Cross-lingual link discovery (Crosslink) as a pilot task of NTCIR-9 has been successfully held in 2011
- Below is a list of participating teams with submissions:

GROUP	ORGANISATION
DUIIS	Daegu University
HITS	Heidelberg Institute for Theoretical Studies
IISR	Yuan Ze University
ISTIC	Institute of Scientific and Technical Information of China
KMI	The Open University
kslab_nut	Nagaoka University of Technology
KSLP	Kyungsung University
nthuisa	Academia Sinica
QUT	Queensland University of Technology
UKP	TU Darmstadt
WUST	Wuhan University of Science and Technology

Crosslink task at NTCIR-9 (Cont.)

- There were in total 57 runs from 11 teams were received.

Group	En-2-Zh	En-2-Ja	En-2-Ko
DUIIS	0	0	2
HITS	3	3	3
IISR	0	0	5
ISTIC	1	0	0
KMI	4	0	0
kslab_nut	0	1	0
KSLP	0	0	5
nthuisa	3	0	0
QUT	5	2	1
UKP	5	5	5
WUST	4	0	0
Sub-total	25	11	21
Total	57		

Crosslink Task at NTCIR-10

- CLLD is a core task of NTCIR-10
- New Subtasks
 - Chinese to English CLLD (C2E)
 - Japanese to English CLLD (J2E)
 - Korean to English CLLD (K2E)
- New Document Collections
 - ECJK Wikipedia collections
- New Topics
 - 25 topics for each language

Evaluation Framework

- Assessment Types: Automatic (Wikipedia Ground Truth), Manual (human in the loop)
- Test data, training data, gold standard, ECJK Wikipedia collections, validation tool, assessment tool and evaluation tool (with system evaluation metrics).
- Snapshots of the validation, assessment and evaluation tools:

The image displays three screenshots of software tools used in the evaluation framework:

- Validation Tool:** A web browser window showing the Wikipedia page for "Australia". The page content includes information about the country, its history, and its status as a member of the Commonwealth of Nations. A blue box labeled "Validation Tool" is overlaid on the page.
- Assessment Tool:** A web browser window showing a manual assessment interface. It displays a list of topics, including "海藻" (seaweed) and "網鰐刺魚" (net shark). A blue box labeled "Assessment Tool" is overlaid on the interface.
- Evaluation Tool:** A web browser window showing a table of evaluation metrics. The table has columns for various metrics and rows for different topics. A blue box labeled "Evaluation Tool" is overlaid on the table.

System Evaluation Metrics

- $LMAP = \left(\sum_{t=1}^n \frac{\sum_{k=1}^m p_{kt}}{m} \right) / n$

- $R - Prec = \sum_{t=1}^n P_t @ R / n$

- *Precision-at-N* is computed using the average precision for all topics (source articles) at a pre-defined position N in the results list. Values of N were chosen as: 5, 10, 20, 30, 50, and 250.

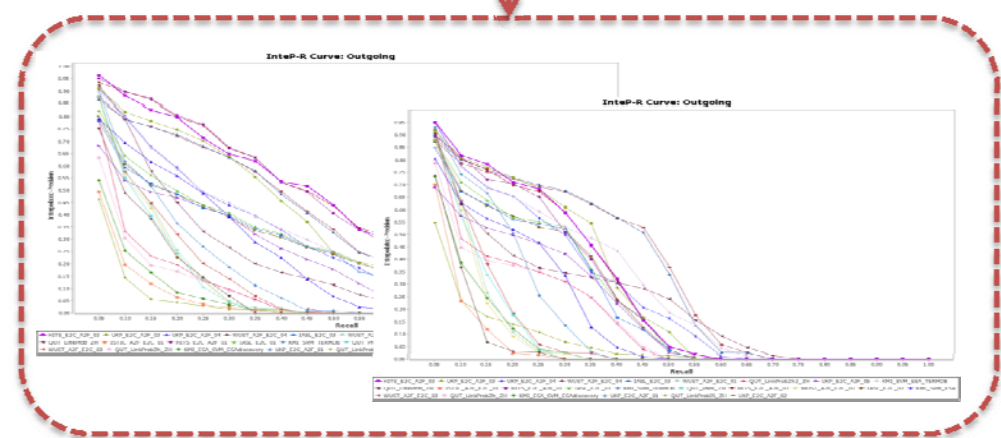
where n is the number of topics; m is the number of identified items (links or anchors); P_{kt} is the precision at top k items (links or anchors) for topic t ; $P_t @ R$ (= number of correct items (links or anchors) / number of items (links or anchors) in *qrel*) is the precision calculated using number of links / anchors in *qrel* as denominator for topic t .

Expected Outcomes

The screenshot shows a web interface with multiple language entries for 'Custard'. On the left, there's an image of a croissant labeled 'La Parisienne Almond Croissant'. In the center, there's a large image of a custard dish. To the right, there's another image of a custard dish. The text is in multiple languages: English, Chinese, Korean, and Japanese. A red box highlights the word 'crema pasticceria' in the Italian text. Green arrows point from the English text to the Chinese text, and from the Chinese text to the Japanese text, indicating cross-lingual links.

- More good submissions
- More original and innovative approaches can be seen in identifying meaningful anchors and suggesting high quality cross-lingual links
- The research results can really help the cross-lingual knowledge discovery in knowledge bases

Evaluation Framework



- The evaluation framework will be further refined
- The evaluation methods will be further perfected to distinguish the good and the bad CLLD algorithms for the new subtasks

Expected Participants

- Previous active participants of NTCIR-9 Crosslink task
- The registered participants of NTCIR-9 Crosslink task.
- Participants of CLIR or IR4QA task in previous NTCIR workshops.
- Participants of previous INEX Link-the-Wiki track
- Other researchers in the CLIR field

Contacts

- Organisers:

Shlomo Geva	Queensland University of Technology, Australia
Andrew Trotman	University of Otago, New Zealand
Yue Xu	Queensland University of Technology, Australia
Eric Tang	Queensland University of Technology, Australia
In-Su Kang	Kyungsoong University, South Korea
Fuminori Kimura	Ritsumeikan University, Japan
Haitao Mi	Chinese Academy of Sciences, China

- Mailing list: crosslink@lists.otago.ac.nz

INTENT-2



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

このクエリのユーザーの意図は？

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ーの意図は？

ハリーポッター

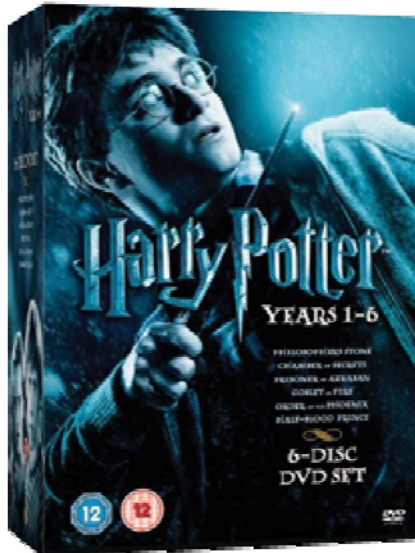


不明快な (underspecified) クエリ

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すること

検索結果ページ(SERP)



INTENT-2 サブタスク

- サブトピックマイニング (日、中、英)

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ト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”)

- ドキュメントランキング (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

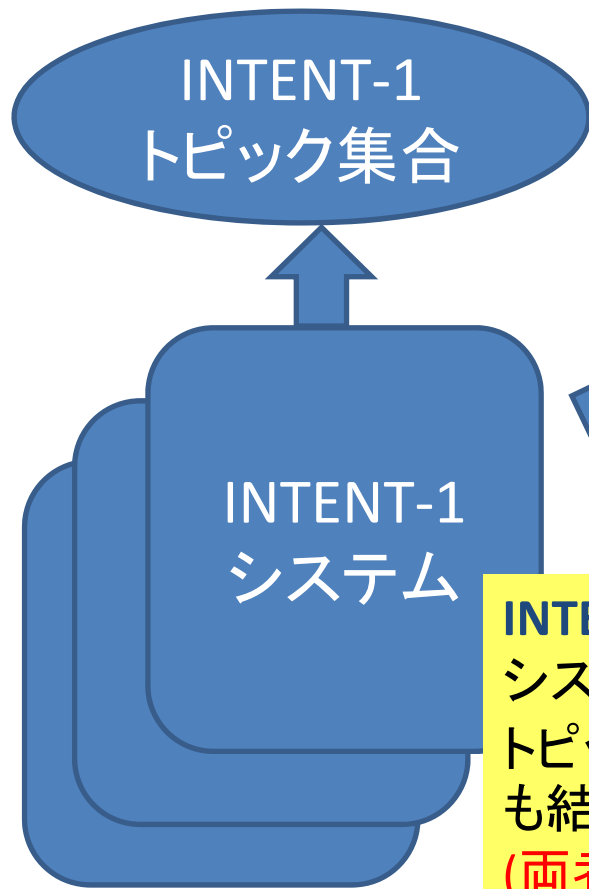
INTENT-1 との相違点 (1/2)

- 英語 (new!) のサブトピックマイニングタスクでは **TREC 2012 diversity task** と共通のトピックを一部利用
- トピック集合に、検索結果の多様化を必要としないようなクエリ (one item search クエリ) を含める
 - システムは各トピックについて多様化の必要性を判定する必要
- 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能に

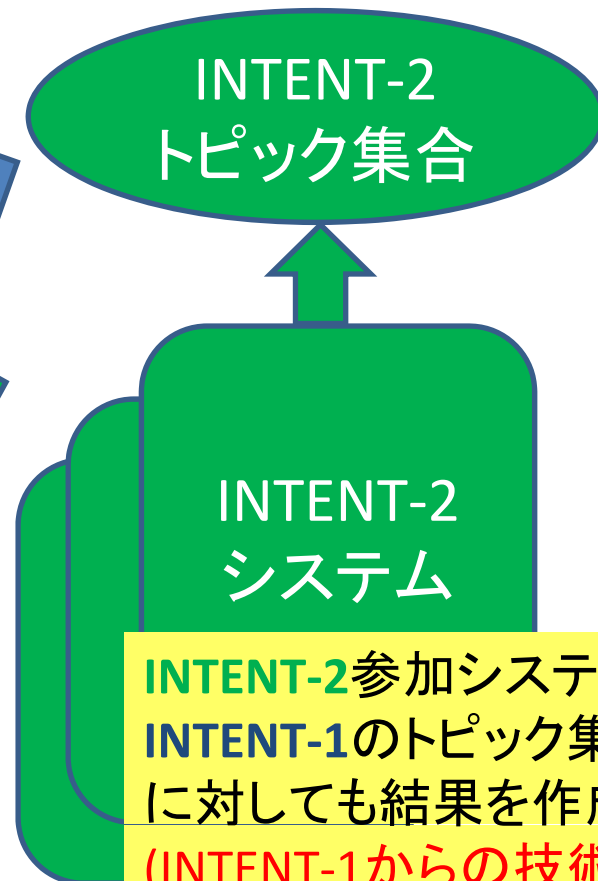
INTENT-1 との相違点 (2/2)

INTENT-1@NTCIR-9

INTENT-2@NTCIR-10



INTENT-1に参加したシステムはINTENT-2トピック集合に対しても結果を作成
(両者のトピック集合の等価性を議論)



INTENT-2参加システムはINTENT-1のトピック集合に対しても結果を作成
(INTENT-1からの技術の進歩について議論)

スケジュール

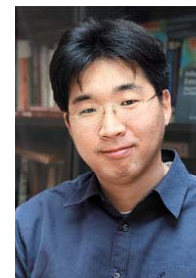
参加者

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	全てのランの提出締切
2012年8~12月	サブトピックからのintent作成 intent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?	初稿論文締切
2013年 5月?	最終原稿締切
6月18~21日	NTCIR-10

INTENT-1では**7つの国と地域**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！！

1Click-2



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)

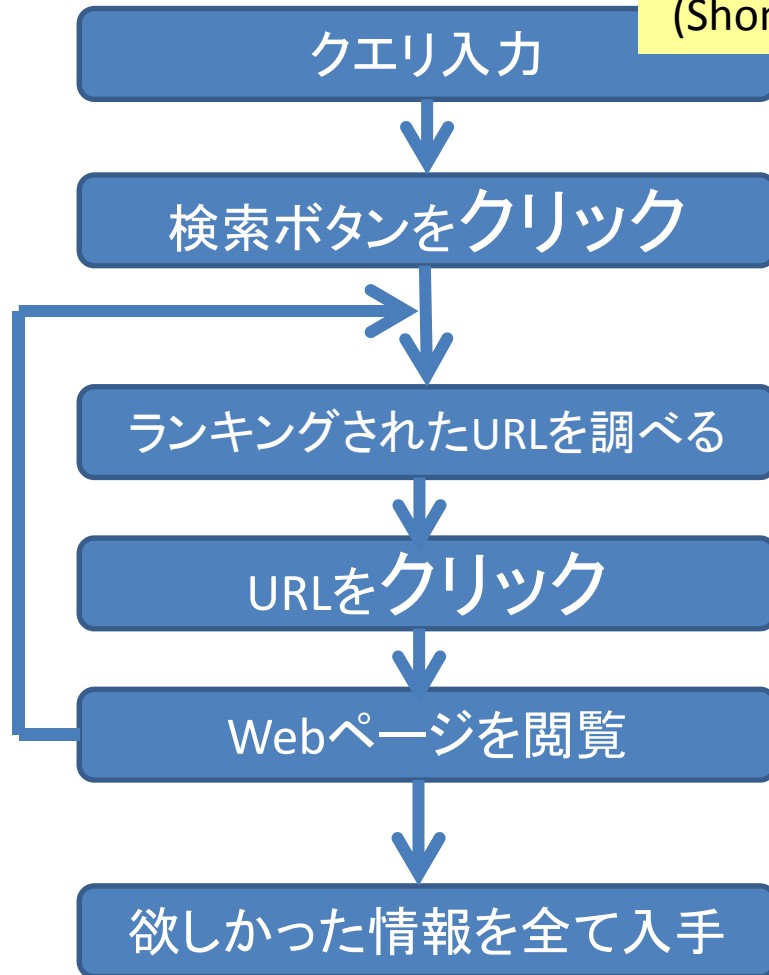
<http://research.microsoft.com/en-us/people/tesakai/1click2.aspx>



March 8, 2012@NII, Tokyo

従来の検索 = More-than-One Click Access

湘南厚木病院
(Shonan Atsugi Hospital)



One Click Access

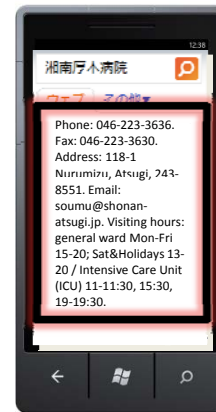
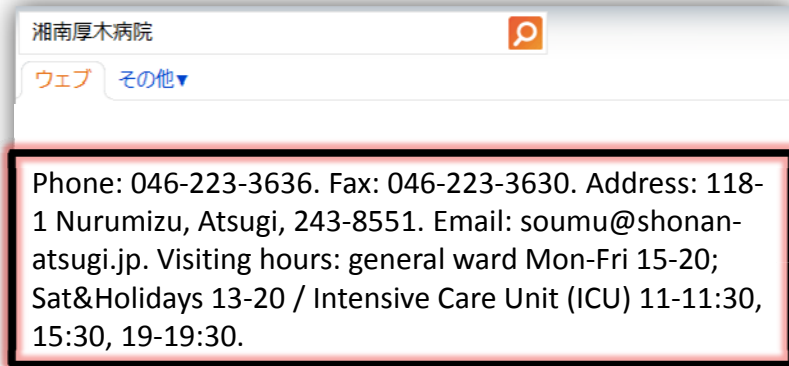
湘南厚木病院
(Shonan Atsugi Hospital)

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



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



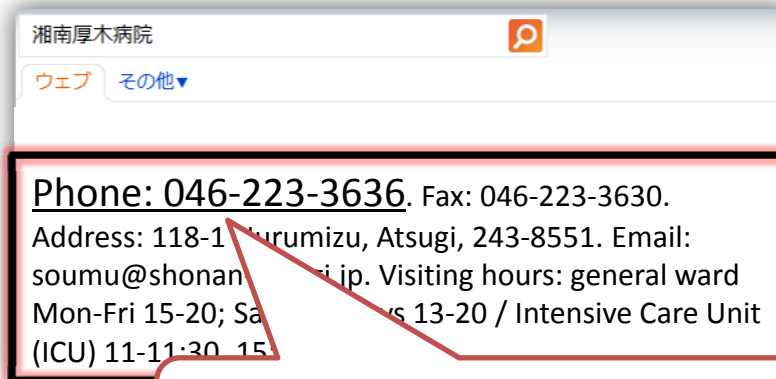
すぐに湘南厚木病院に電話しなくては！

従来の検索エンジン

1クリックアクセスシステム



やっと見つけた



すぐ見つかった!

クエリ

- **8種類のクエリタイプを用意 [Li et al., SIGIR09]**
(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)
 - 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)
- **クエリは商用検索エンジンの大規模ログデータから選出**

iUnitsによる評価

- 手動/自動でシステム出力と*iUnits* (我々が定義する情報の最小単位)を比較して評価

電話: 046-223-3636. Fax: 046-223-3630.
住所: 〒243-8551神奈川県厚木市温水
118-1. Email: soumu@shonai-ct.ac.jp

システム出力

- 電話番号: 046-223-3636
- Fax番号: 046-223-3630
- 住所: 〒243-8551神奈川県厚木市温水

iUnits

- システムは文書検索ではなく情報検索の観点から評価される

1CLICK-2 サブタスク

参加者は下記のどちらかへの参加していただきます

- メインタスク

- **Mandatory Runs**: 共通のWeb検索結果から結果を生成
- **Oracle Runs (optional)**:
iUnitsが記載されているURLリストから結果を生成
- **Open Runs (optional)**:
任意の検索エンジンを利用して結果を生成

OR

- クエリ分類サブタスク

- クエリが与えられた時にそのクエリタイプを推定
(芸術家, QA, etc.)

1CLICK-1との相違点

- 日本語に加え**英語**のタスク (留学生の方にぜひ)
- クエリタイプを4種類から**8種類**に
- **iUnit**という新しい情報単位での評価
- **自動**/手動によるシステム出力の評価
- タスクのバリエーションが増えました
 - より公正な評価を目指しつつ気軽な参加も可能に
 - クエリ分類タスク (Web検索には非常に重要なタスク)
だけでも参加できます

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 システム出力締切

2012年 11月 iUnit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

PatentMT-2

The Patent Machine Translation Task (PatentMT)

Isao Goto (NICT)

Bin Lu (City Univ. of Hong Kong / Hong Kong Institute of Education)

Ka Po Chow (Hong Kong Institute of Education)

Eiichiro Sumita (NICT)

Benjamin K. Tsou (Hong Kong Institute of Education / City Univ. of Hong Kong)

Motivation

- There is a significant **practical need** for patent translation.
 - to understand patent information written in foreign languages.
 - to apply for patents in foreign countries.
- Patents constitute one of the **challenging domains**.
 - Patent sentences can be quite **long** and contain **complex structures**.
 - Translation between **languages with largely different word order** is difficult for **long** sentences.

Goals of PatentMT

- To develop **challenging** and **significant practical** research into patent machine translation.
- To **investigate** the **performance** of state-of-the-art machine translation systems in terms of patent translations involving Japanese, English, and Chinese.
- To **compare** the effects of **different methods** of patent translation by applying them to the same test data.
- To **create** publicly-available **parallel corpora of patent documents** and human evaluations of MT results for patent information processing research.
- To **drive machine translation research**, which is an important technology for cross-lingual access of information written in unknown languages.
- The ultimate goal is **fostering scientific cooperation**.

Findings of PatentMT at NTCIR-9

- **SMT** was the **best** system for **Chinese to English** and **English to Japanese** patent translation.
 - This is the **first time** for **SMT** to be **demonstrated equal or better** quality than that of the top-level RBMT for **English to Japanese** patent translation.
 - The **pre-ordering** method of NTT-UT for SMT is very effective for English to Japanese patent translation.
- **80%** of patent sentences could be understood in the best system for **Chinese to English** patent translation.
- **RBMT** was the best system for **Japanese to English** patent translation.

Remaining Issues of NTCIR-9

- Practical evaluation
 - The quality of translated sentences was evaluated at NTCIR-9.
 - More practical evaluations are also expected.

Outline of the Plans for NTCIR-10

- Three subtasks:

Subtasks	Training data
Chinese to English	1 million sentence pairs
Japanese to English	Approximately 3.2 million sentence pairs
English to Japanese	

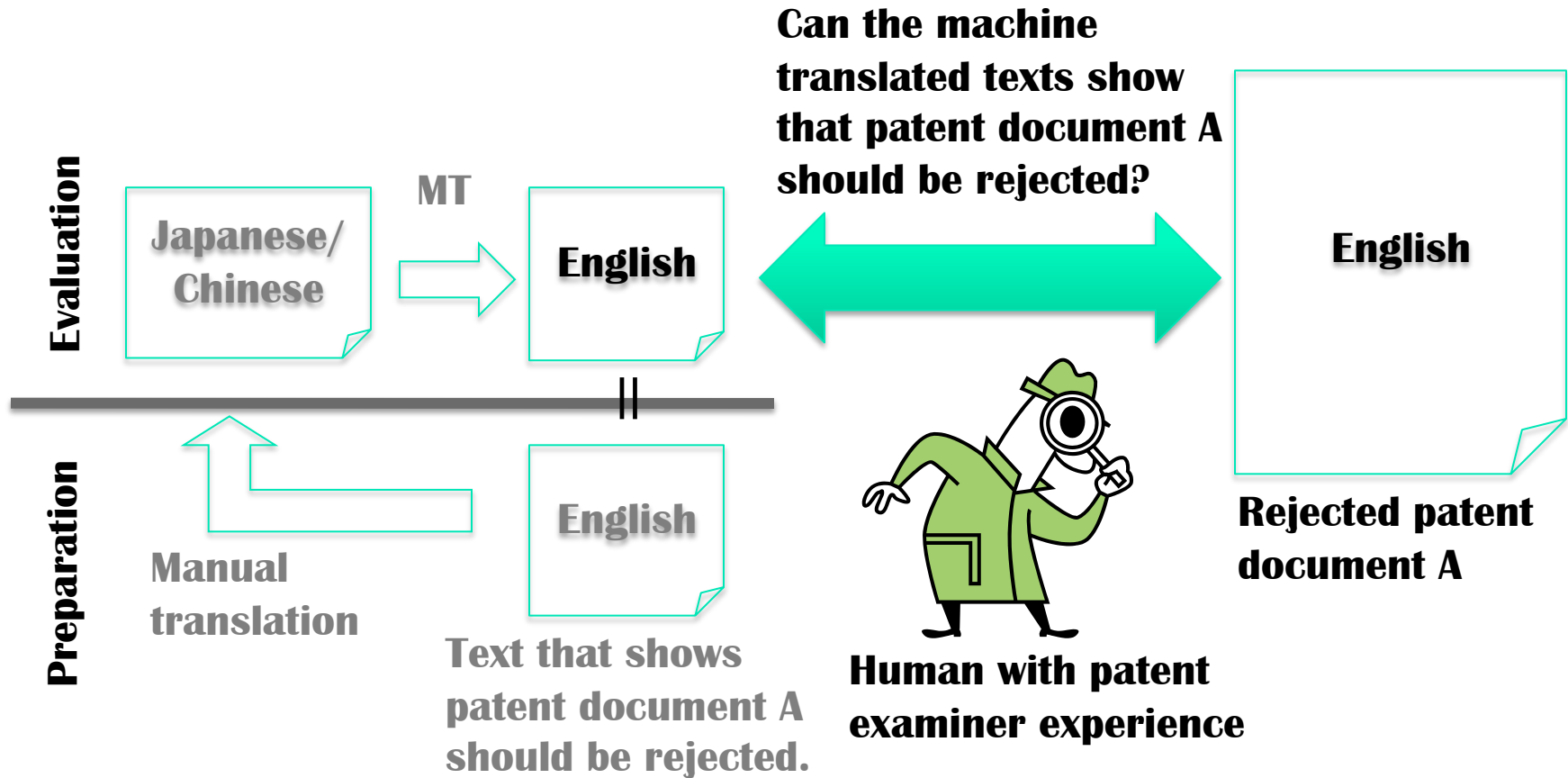
(Subtasks and training data are the same as at NTCIR-9)

- Participants select subtasks in which they wish to participate.
- Large scale **parallel corpora** and **new test sets** will be **provided**.
- **Practical evaluation** will be added (under consideration).
- **Human evaluation** will be carried out.

Differences from NTCIR-9

<p>Practical Evaluation (under consideration)</p>	<p>New: To explore practical MT performance in appropriate fields for patent machine translation.</p>	
<p>Intrinsic Evaluation</p>	<p>Similar to the NTCIR-9 evaluation. Quality of translated sentences will be evaluated. Additions:</p>	
	<p><i>Chronological evaluation</i></p>	<p>Comparison between NTCIR-10 and NTCIR-9 to measure progress.</p>
<p><i>Multilingual evaluation</i></p>	<p>Comparison of CE and JE translations using the same English reference will be added.</p>	

Possible Approach to Practical Evaluation



(The feasibility of this is under investigation.

We are working hard to make necessary arrangements.)

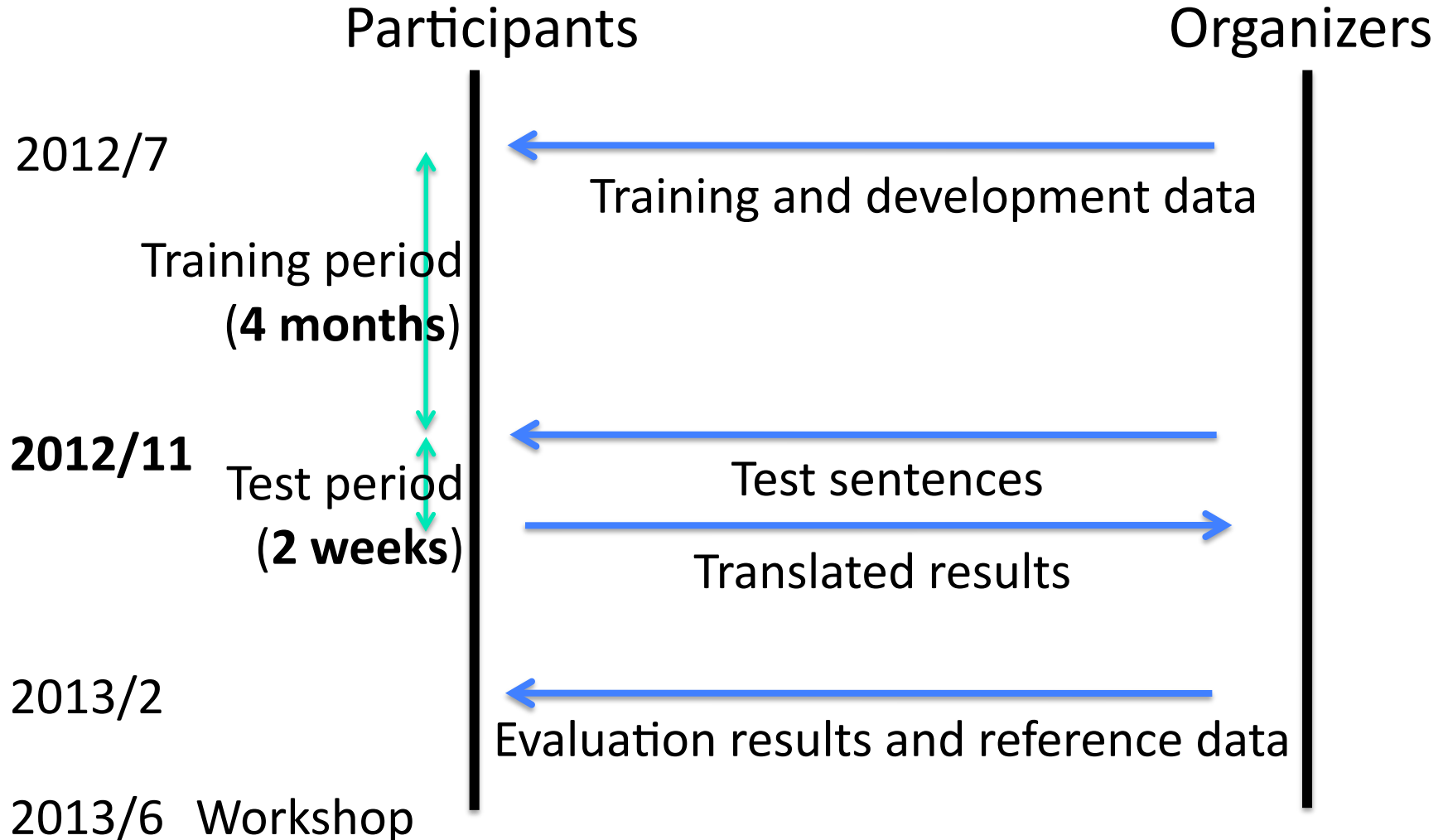
Chronological Evaluation

- In addition to the new NTCIR-10 test sets, the NTCIR-9 test sets will be also translated.
- Translations of the NTCIR-9 test sets at NTCIR-10 will be compared to the NTCIR-9 submissions.
- This allows **measurement of the progress from NTCIR-9.**

Multilingual Evaluation

- We will produce a **C-J-E multilingual** test set.
- A C-J-E multilingual test set enables **comparison** of the **CE** translations and the **JE** translations using the **same English reference data**.

The Flow and Tentative Schedule of the Task



Why is it so exciting to participate in?

- **Patents** are one of the **challenging domains** for MT.
 - Patent sentences could be quite **long** and contain **complex structures**.
 - Translation between **languages with largely different word order** is difficult for **long** sentences.
- Participants will receive **evaluation results** for their MT quality.
- Participants can use **large-scale patent parallel** and **monolingual corpora**.
- Participants can choose subtasks from three language directions, including **the language direction of Chinese to English**.
- We look forward to many groups participating in PatentMT at NTCIR-10!

SpokenDoc-2



NTCIR-10

SpokenDoc-2

IR (Information Retrieval) for Spoken Documents

Kiyoaki Aikawa¹, Tomoyosi Akiba², Seiichi Nakagawa²,
Hiroaki Nanjo³, Hiromitsu Nishizaki⁴, Yoichi Yamashita⁵

¹Tokyo University of Technology, Japan

²Toyohashi University of Technology, Japan

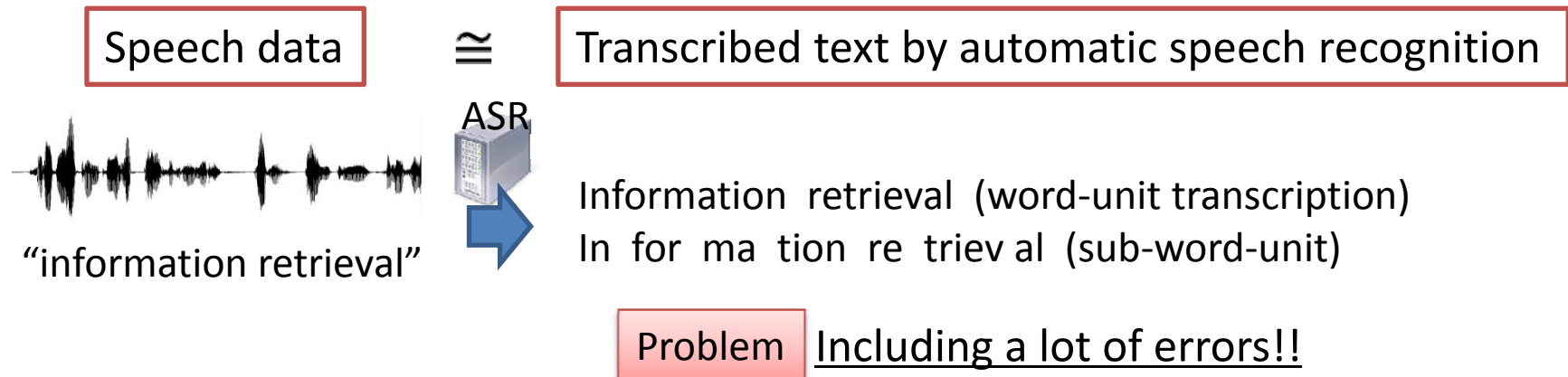
³Ryukoku University, Japan

⁴University of Yamanashi, Japan

⁵Ritsumeikan University, Japan

What is “SpokenDoc-2”?

- Second round of the IR for spoken documents
- Finding the information related to given a query from too much speech data



Participants of SpokenDoc-2 will challenge

Information retrieval from very noisy text data

Techniques for SpokenDoc may be used for OCR or Machine Translated text retrieval

Target Speech Data

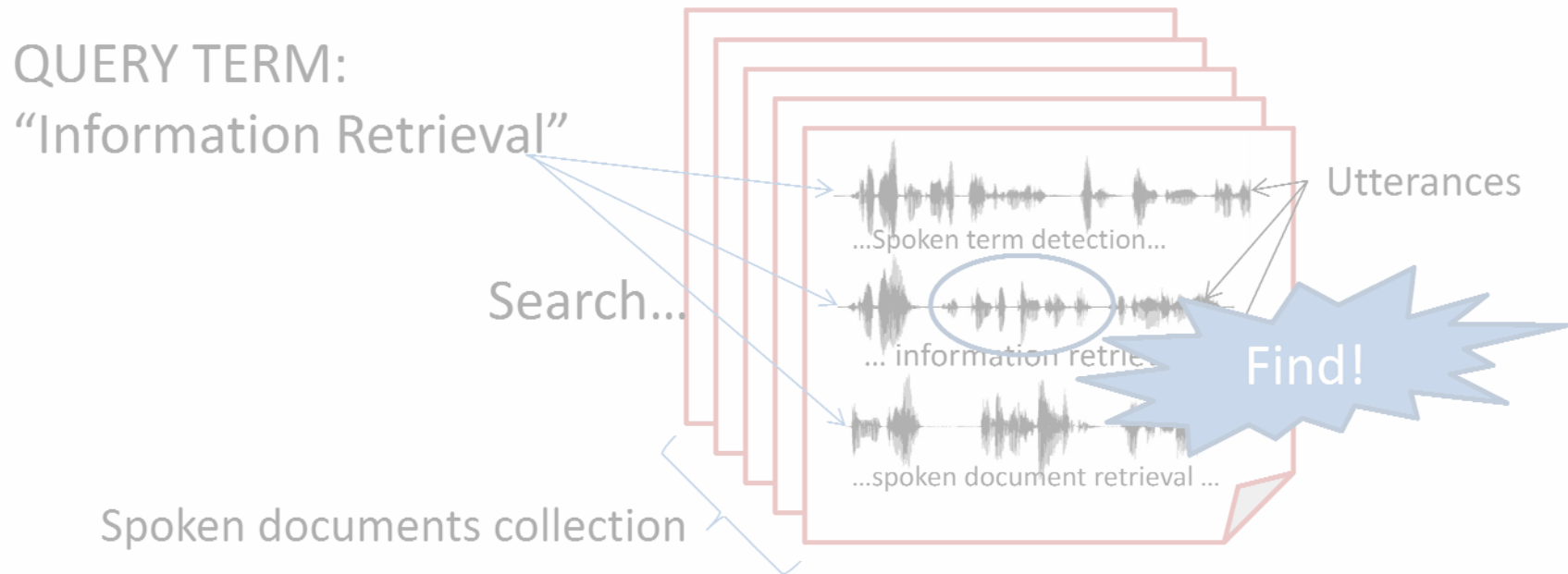
- Type of speech data
 - Broadcast news speech, podcast, lecture speech...
Having noisier words
Our target
- Databases
 - CSJ (Corpus of Spontaneous Japanese)
 - 2,702 lecture speeches, 628 hours
 - **New target!** Real academic meeting lectures collection
 - Over 70 speeches from the spoken document processing workshops

SpokenDoc-2 Task Overview

- Sub-task1: Spoken Term Detection
- Sub-task2: Spoken Document Retrieval

Spoken Term Detection

1. Finding utterances including a query term



2. A query term is included or not included in a spoken document?

Spoken Document Retrieval

- Ad-hoc Information Retrieval from lecture speeches
- Finding the **passages** including the relevant information related to a given query topic
- Query
 - Text query
 - Spoken Query (optional)



Merits for Task Participants

- The organizers will **provide the rich transcriptions** by an automatic speech recognition system
 - Participants can focus on developing IR technique for noisy data, although you have not any techniques for speech recognition
- Participants will **get the real lecture speeches**:
 - Academic meeting lectures
 - All lectures have reference transcriptions
 - ~~– Some lectures have slide data and digitalized papers~~

All data is Japanese

We are welcome to join SpokenDoc-2



- Schedule
 - Mar. 2012: release of the task description
 - June 2012: release of the reference automatic transcription
 - Sept. 2012: dry-run
 - Nov. 2012: formal-run evaluation
 - Nov. 2012-Feb.2013: relevance judgment
 - Feb. 2013: release of the evaluation results
 - May 2013: camera ready submission due
 - June 2013: NTCIR-10 workshop meeting
- Contact
 - SpokenDoc-2 organizers:
 - E-mail: ntcadm-spokendoc2@cl.ics.tut.ac.jp
 - Web (coming soon!)
 - <http://www.cl.ics.tut.ac.jp/~sdpwg/index.php?ntcir10>
 - Twitter: @spokendoc2



RITE-2



NTCIR-10 RITE-2 タスク

(Recognizing Inference in TExt)



Yotaro Watanabe¹
Junta Mizuno¹
¹Tohoku University



Yusuke Miyao²
²National Institute of Informatics



Tomohide Shibata³
³Kyoto University



Cheng-Wei Lee⁴
⁴Academia Sinica



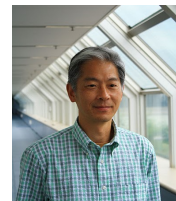
Chuan-Jie Lin⁵
⁵National Taiwan Ocean University



Shuming Shi⁶
⁶Microsoft Research Asia



Hiroshi Kanayama⁷
⁷IBM Research



Koichi Takeda⁷



Hideki Shima⁸
⁸Carnegie Mellon University



Teruko Mitamura⁸

RITE-2

Recognizing
Inference in
Text@NTCIR10

RITE-2 タスクの概要

- RITEはテキスト間の含意(=推論), 換言(=同じ意味), 矛盾の判断を対象とした, テキスト理解システムのためのベンチマークタスク
- 含意: テキスト t_1 が与えられたとき,
 仮説 t_2 も真であると推論可能
 - t_1 : 川端康成は「雪国」などの作品でノーベル文学賞を受賞した
 - t_2 : 川端康成は「雪国」の著者である
- 言語: 日本語, 中国語 (Simplified, Traditional)

RITE-2 のサブタスク

- **Binary-Class (BC) サブタスク**

- 1組の文対 $\langle t_1, t_2 \rangle$ を与えられたとき、 t_1 から仮説 t_2 が真であるかどうかを出力

- **Multi-Class (MC) サブタスク**

- 1組の文対 $\langle t_1, t_2 \rangle$ を与えられたとき、4種類の関係に分類

含意関係あり	t_1 が t_2 を含意	t_1 と t_2 は同義
含意関係なし	t_1 と t_2 は矛盾	t_1 と t_2 は意味的に独立

- **大学入試サブタスク**

- **BC サブタスク**

- 大学入試の問題を対象に、BCと同じ課題を解く

- **検索サブタスク**

- 仮説 t_2 と、Wikipedia・教科書が与えられたとき、 t_2 が真であるかどうかを出力

RITE-2

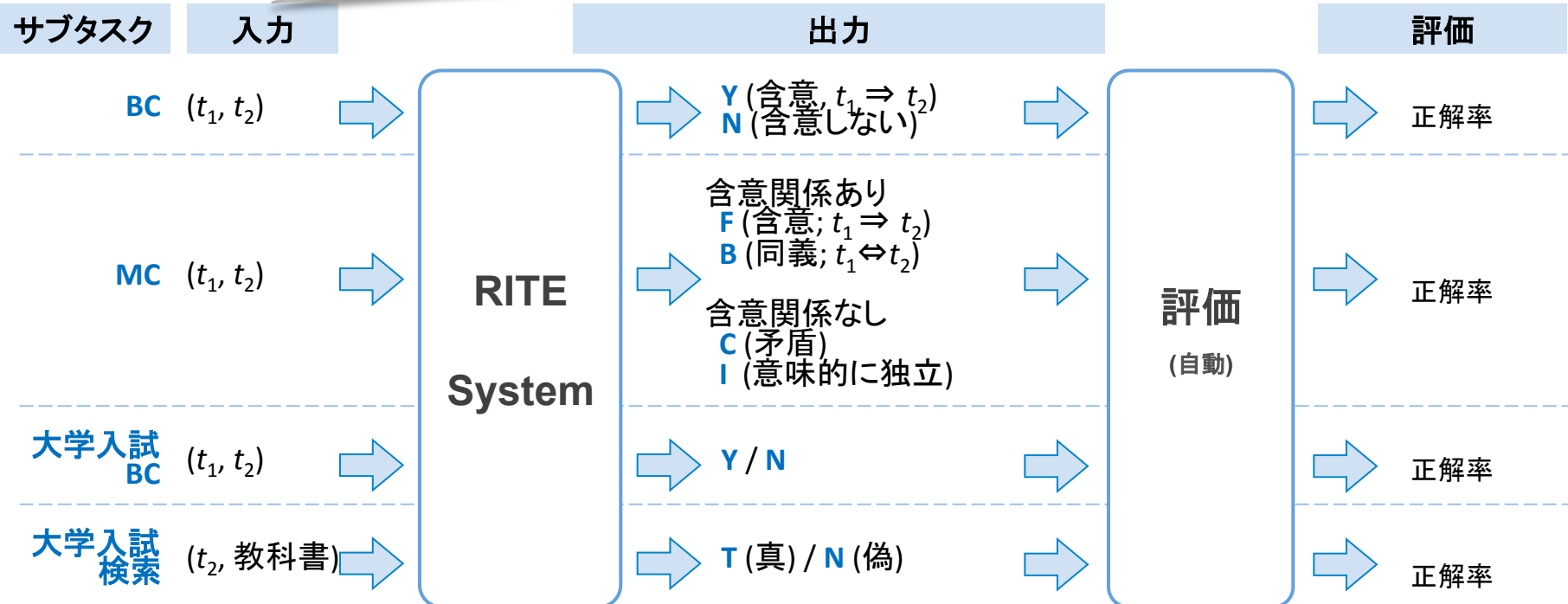
Recognizing
Inference in
TExt@NTCIR10

RITE-2 評価の流れ

t_1 が t_2 を含意？

t_1 : 川端康成は「雪国」などの作品でノーベル文学賞を受賞した

t_2 : 川端康成は「雪国」の著者である



RITE-2

Recognizing
Inference in
Text@NTCIR10

ご参加をお待ちしております

RITEは含意関係認識, 言い換え認識の研究だけでなく, 幅広い研究分野と関連があります

- **基盤技術**: 言語・意味解析, 知識獲得, 機械学習
- **応用技術**: 情報検索, 質問応答, 自動要約 など

大学生から企業の研究者に至るまで, 幅広く参加していただけるよう工夫をします

- 既存のリソースやツールの情報を共有

詳しくはウェブサイトにて

<http://www.cl.ecei.tohoku.ac.jp/rite2>

Math

NTCIR-10 kickoff

March 8, 2012

NTCIR-10 MATH PILOT TASK

The Goal of NTCIR-10 Math Task

- NTCIR Math Task aims at exploring methods for mathematical content access through its task design and the construction of the evaluation dataset.

[Formula]

a mathematical relationship or rule expressed in symbols
(Oxford Dictionary)

In science, a formula is a concise way of expressing information, or a general relationship between quantities.
(Wikipedia)



**INFORMATION
ACCESS TO
MATHEMATICAL
CONTENT**

?

Math Information Access

Representations

Embedded image (png, gif, ...)

$$\log(z_1) + \log(z_2) = \log(z_1 z_2) \ ; \ z_1 + z_2 \geq 0$$

Character sequence (latex source)

$$\log(z_1) + \log(z_2) == \log(z_1 z_2) \ ; \ z_1 + z_2 \geq 0$$

Web-browsable XML

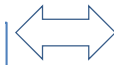
```
<math xmlns='http://www.w3.org/1998/Math/MathML' mathematica:form='TraditionalForm' xmlns:mathematica='http://www.wolfram.com/XML/'>
<semantics>
<mrow>
<mrow>
<mrow>
<mrow>
<mi>log</mi>
<mo>+</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mo>=</mo>
<mi>log</mi>
<mo>(</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mo>)</mo>
<mo>(</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mo>)</mo>
<mo>(</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mo>)</mo>
<mo>(</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mo>)</mo>
</semantics>
</math>
```

XML for math semantics

```
<annotation-xml encoding='MathML-Content'>
<apply>
<ci>Condition</ci>
<apply>
<eq />
<apply>
<plus />
<apply>
<ln />
<apply>
<ci>Subscript</ci>
<ci>z</ci>
<cn type='integer'>1</cn>
</apply>
</apply>
<apply>
<ln />
<apply>
<ci>Subscript</ci>
```

Resources

mathematical knowledge-base and math ontology



Strict Content MathML (W3C recommendation)

Wolfram MathWorld:
13,081 entries (Sep. 13, 2011)
Wolfram Functions site:
307,409 formulas (Sep. 15, 2011)



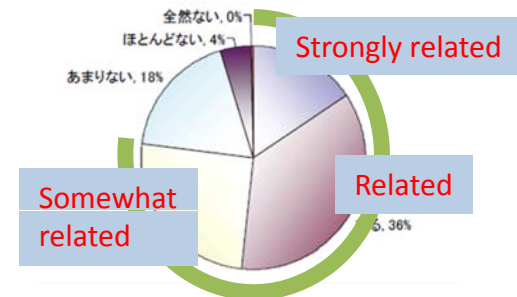
Wikipedia:
26,566 mathematics articles



Requirement

NISTEP Policy Study
Mathematics as deserted science in Japanese S&T policy
- Current situation on mathematical sciences research in major countries and need for mathematical sciences from the science in Japan (2006.5)

Q. Is mathematics related to your research?



77% researchers across diversity of disciplines answered 'YES'.

Task Overview

- **[Math Retrieval Subtask]**
 - Given a document collection, retrieve relevant mathematical formulae or documents for a given query.
- **[Math Understanding Subtask]**
 - Extract natural language definitions of mathematical expressions in a document for their semantic interpretation.

Math Retrieval Subtask

- **Dataset (scheduled)**

- Scientific Articles from ArXiv e-print server

<http://arxiv.org/>

- Converted into XML+MathML by arXMLiv project

<http://kwarc.info/projects/arXMLiv/>

- 10,000 docs for a dry run, additional 100,000 docs for a formal run

- **Search Types**

- The Math retrieval task uses the above 110K docs and can be envisaged in three different search scenarios

- Formula Search

- Search for formula queries within the formulae database of the used dataset

- Full-Text Search

- Search the document collection using formula queries. Combinations of keywords and formulae.

- Open Information Retrieval

- Search the document collection using free textual queries.

Math Retrieval Subtask : Dataset Example

```

<m:mi id="id57134">S</m:mi>
<m:mi id="id57136">|</m:mi>
<m:mo id="id57138">&prime;</m:mo>
</m:msubsup>
</m:mrow>
<m:mo id="id57141">&ne;</m:mo>
<m:msup id="id57143">
<m:mi id="id57144">G</m:mi>
<m:mo id="id57146">&prime;</m:mo>
</m:msup>
</m:mrow>
<m:mo id="id57149">,</m:mo>
</m:mrow>
<m:annotation-xml id="id57151"
<m:apply id="id57154">
<m:neq id="id57155">
<m:apply id="id57156">
<m:apply id="id57157">
<m:csymbol id="id57158" cd="ambiguous">superscript</m:
<m:apply id="id57163">

```

math representation (MathML)

Let $\{G_\gamma \mid \gamma \in \Gamma\}$ be a family of abelian groups. If G_γ is not a proper union of then $G = \bigoplus_{\gamma \in \Gamma} G_\gamma$ is also not a proper union of finitely many cosets.

Demonstration Proof

To prove it by transfinite induction we have two cases to distinguish. If Γ is with some Γ' and for Γ' the statement is true. Then we get $G = G_\gamma \oplus G'$, which is a proper union of finitely many cosets of G with respect to a subgroup H such that $b + G_\gamma \subseteq S$ with some $b \in G$ and $S = G_\gamma + S'$, where S' is a proper coset of G' .

Suppose that G is a proper union of the cosets S_1, \dots, S_n . If S_i contains a coset of G_γ written as $G_\gamma + S'_i$; otherwise, S'_i is the empty set. By induction

$$\bigcup_{i=1}^n S'_i \neq G',$$

therefore, there is a $d \in G'$, such that $d + G_\gamma$ is not contained in any S'_i . Moreover, then it contains an $r_1 + d$ and $S_i = r_1 + d + G_i$, where $r_1 \in G_\gamma$ and G_i is a subgroup of G_γ .

$$S_i \cap (d + G_\gamma) = (r_1 + d + G_i) \cap (r_1 + d + G_\gamma) = (r_1 + d) + G_i \cap G_\gamma$$


and this implies that G_γ is a proper union of some of the cosets $r_1 + (G_i \cap G_\gamma)$, which contradicts the induction hypothesis.

In the second case Γ is a limit ordinal. For a $\Gamma' < \Gamma$ set

$$G_{\Gamma'} = \bigoplus_{\alpha \in \Gamma'} G_\alpha.$$

Assuming G is a proper union of the cosets T_1, \dots, T_k we obtain

$$G_{\Gamma'} = \bigcup_{i=1}^k (G_{\Gamma'} \cap T_i).$$


Cornell University Library
We gratefully acknowledge supporting institutions

arXiv.org > math > arXiv:0801.0652
Search or Article-id (Help | Advanced search)

Mathematics > Rings and Algebras

Covering theorems for Artinian rings

A. Borbely, V. Bovdi, B. Brindza, T. Krausz

(Submitted on 4 Jan 2008)

The covering properties of Artinian rings which depend on their additive structure only, are investigated.

Comments: 5 pages

Subjects: Rings and Algebras (math.RA)

MSC classes: 16P20

Journal reference: Publ. Math. Debrecen, 51/3-4, 1997, p.323-329

Cite as: arXiv:0801.0652v1 [math.RA]

Submission history

From: Victor Bovdi [view email]

[v1] Fri, 4 Jan 2008 10:27:23 GMT (5kb)

[Which authors of this paper are endorsers?](#)

Link back to: arXiv, form interface, contact.

Download:

- Source
- PDF
- PostScript
- Other formats

Current browse context:

math.RA

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [0801](#)








Change to browse by:

math

References & Citations

- NASA ADS

Bookmark (what is this?)

papers from arXiv.org

xhtml/xml



Math Retrieval Subtask :

Query example

Formula search

1.

$$\int_{[l]}^{[h]} [f](x)^2 dx$$

2.

$$[a] + [b] = [b] + [a]$$

3.

$$\lim_{[x] \rightarrow \infty} [a] + [b] = [c]$$

Full-text search

1.

Pythagorean Theorem

2.

Bell curve in the form of

$$\frac{1}{[\sigma]\sqrt{2\pi}} \exp \left(-\frac{([x]-[\mu])^2}{2[\sigma]^2} \right)$$

3.

NOT Bayes Theorem,
but written as

$$P([X] | [Y])$$

Open IR

1.

What is the fifth
summand in the
Taylor expansion of
sinus hyperbolicus?

2.

For which n and
 k is $PSL(n,k)$ not
commutative.?

Math Retrieval Subtask : Formula Search Example

Query formula

The screenshot shows a search engine interface with a navigation bar (Questions, Activity, Sign in, Books, Articles, MWS Engine BETA). The search input field contains the LaTeX query $\lim_{x \rightarrow 0} y$. Below the input, there are two columns of search results. The left column shows the raw LaTeX code: `<m:apply><m:apply><m:csymbol cd="ambiguous">subscript</m:csymbol><m:limit/><m:apply><m:ci>→</m:ci>`. The right column lists search categories: Examples - LaTeX queries, Generic subscript search, Specific subscript search, Specific integral search, Physical constant search, All limits approaching zero, and Text in math search. A 'Search' button is located at the bottom left of the results area.

Retrieved formula

The document page displays several mathematical formulas and text. The first formula, $\chi(t, t_w) = \lim_{h_0 \rightarrow 0} \frac{m[h](t)}{h_0}$, is circled in red. Below it, the text says "Then we have that" followed by formula (9): $\chi(t, t_w) = \beta \int_{t_w}^t dt' X[C(t, t')] \frac{\partial C(t, t')}{\partial t'}$. The text continues: "and by performing the change of variables $u = C(t, t')$ we finally obtain the key equation" followed by formula (10): $\chi(t, t_w) = \beta \int_{C(t, t_w)}^1 du X(u)$. The text then says "where we have used the fact that $C(t, t) \equiv 1$ in Ising models." Below this, there is more text: "can be easily extracted simply measuring the integrated response to a small external field" followed by formula (11): $w) = S[C(t, t_w)]$. Then formula (12): $= \int_C^1 du X(u)$. The text continues: "eed is encoded in the shape of the function $S(C)$. ion factor is equal to one and the relation becomes $C(t, t_w)$ or $S(C) = 1 - C$." followed by formula (13): $S(C) = 1 - C$. At the bottom, there is a section header "2.2 LINK between the statics and the dynamics" and the start of a paragraph: "To get information on the thermodynamical properties of the model we should match the violation factor $X(C)$ to some static observable. This can be done using the following

Math Understanding Subtask

- **Task definition**
 - Extract natural language definitions of mathematical expressions in a document (Basic Task) with their semantic interpretation (Challenge Task, TBA).
- **Dataset (scheduled)**
 - Development Data
 - 10 papers selected from ACL-Anthology Reference Corpus
 - 30 papers selected from ArXiv.org dataset which will be also used in Math Retrieval Task.
 - Data for Formal Run (submission period: five days)
 - 10 papers selected from ACL-Anthology Reference Corpus
 - 10 papers selected from ArXiv.org dataset which will be also used in Math Retrieval Task.

Math Understanding Subtask : Dataset Example

Original text (pdf)

The similarity between G_1 and G_2 is then defined as follows:

$$groupSim(G_1, G_2) = RI(G_1, G_2) \times RC(G_1, G_2)$$

where

$$RI(G_1, G_2) = \frac{2AI(G_1, G_2)}{II(G_1) + II(G_2)}$$

is the relative interconnectivity and

$$RC(G_1, G_2) = \frac{AC(G_1, G_2)}{\frac{|G_1|}{|G_1| + |G_2|} IC(G_1) + \frac{|G_2|}{|G_1| + |G_2|} IC(G_2)}$$

is the relative closeness.

Example of Presentation MathML

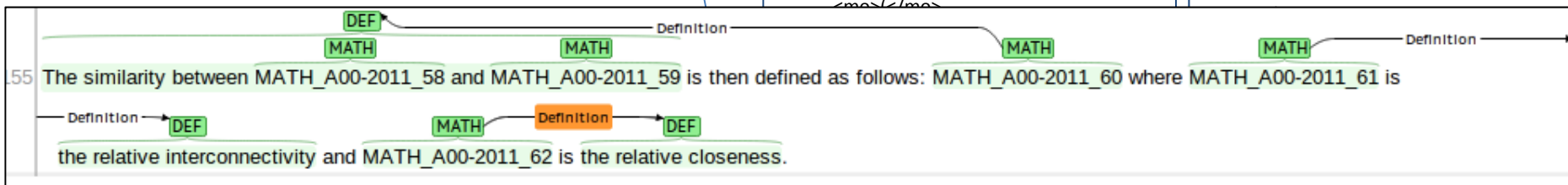
```
<mrow>
  <mi id="math_b6a1eb64.1">
    <mi mathvariant="italic">groupSim</mi>
  </mi>
  <mo>=</mo>
  <mrow>
    <mi>RI</mi>
    <mn>1</mn>
    </mrow>
    <mo>,</mo>
    <mrow>
      <mi>RC</mi>
      <mn>2</mn>
    </mrow>
  </mrow>
  <mo></mo>
</mrow>
</mrow>
<mo>=</mo>
<mrow id="math_b6a1eb64.2">
  <mrow>
    <mi mathvariant="italic">RI</mi>
  </mrow>
  <mo>*</mo>
  <mrow>
    <mi>RC</mi>
  </mrow>
  <mo></mo>
</mrow>
</mrow>
```

MathML parallel markup

Example of Content MathML

```
<annotation-xml encoding="MathML-Content">
  <apply>
    <eq />
    <apply xref="math_b6a1eb64.1">
      <ci>groupSim</ci>
    </apply>
    <selector />
    <ci>G</ci>
    <cn>1</cn>
  </apply>
  <apply>
    <selector />
    <ci>G</ci>
    <cn>2</cn>
  </apply>
</apply>
<apply xref="math_b6a1eb64.2">
  <times />
  <apply>
    <ci>RI</ci>
  </apply>
  <apply>
    <selector />
    <ci>G</ci>
    <cn>1</cn>
  </apply>
  <apply>
    <selector />
    <ci>G</ci>
    <cn>2</cn>
  </apply>
</apply>
```

Definitions of mathematical objects in text



* Annotation tool (<http://www.nactem.ac.uk/brat-annotation/>)

```
<mi>G</mi>
<mn>2</mn>
</msub>
</mrow>
<mo></mo>
```

Schedule

February-April, 2012	Task framework development
April, 2012	Call for participation
May, 2012	Dataset and example topics release
Early-mid October, 2012	Topic release for Open IR search type
Late October, 2012	Topic release for Formula, Full-text search types, and Math understanding subtask
Early November, 2012	Results submissions due for Formula, Full-text search types, and Math Understanding subtask
Mid November, 2012	Results submissions due for Open IR search type
February, 2013	Evaluation Results Released
March, 2013	Draft papers for NTCIR-10 Proceedings Due
May, 2013	Camera ready for NTCIR-10 Proceedings Due
June, 2013	NTCIR-10 Meeting

Task information

- **Contact**

- ntcir10adm-math@nii.ac.jp

- **Task Web page**

- <http://ntcir-math.nii.ac.jp/>

- **Task Organizers**

- Akiko Aizawa (National Institute of Informatics, Japan)

- Michael Kohlhase (Jacobs University Bremen)

- Iadh Ounis (University of Glasgow)

- **Task Advisors**

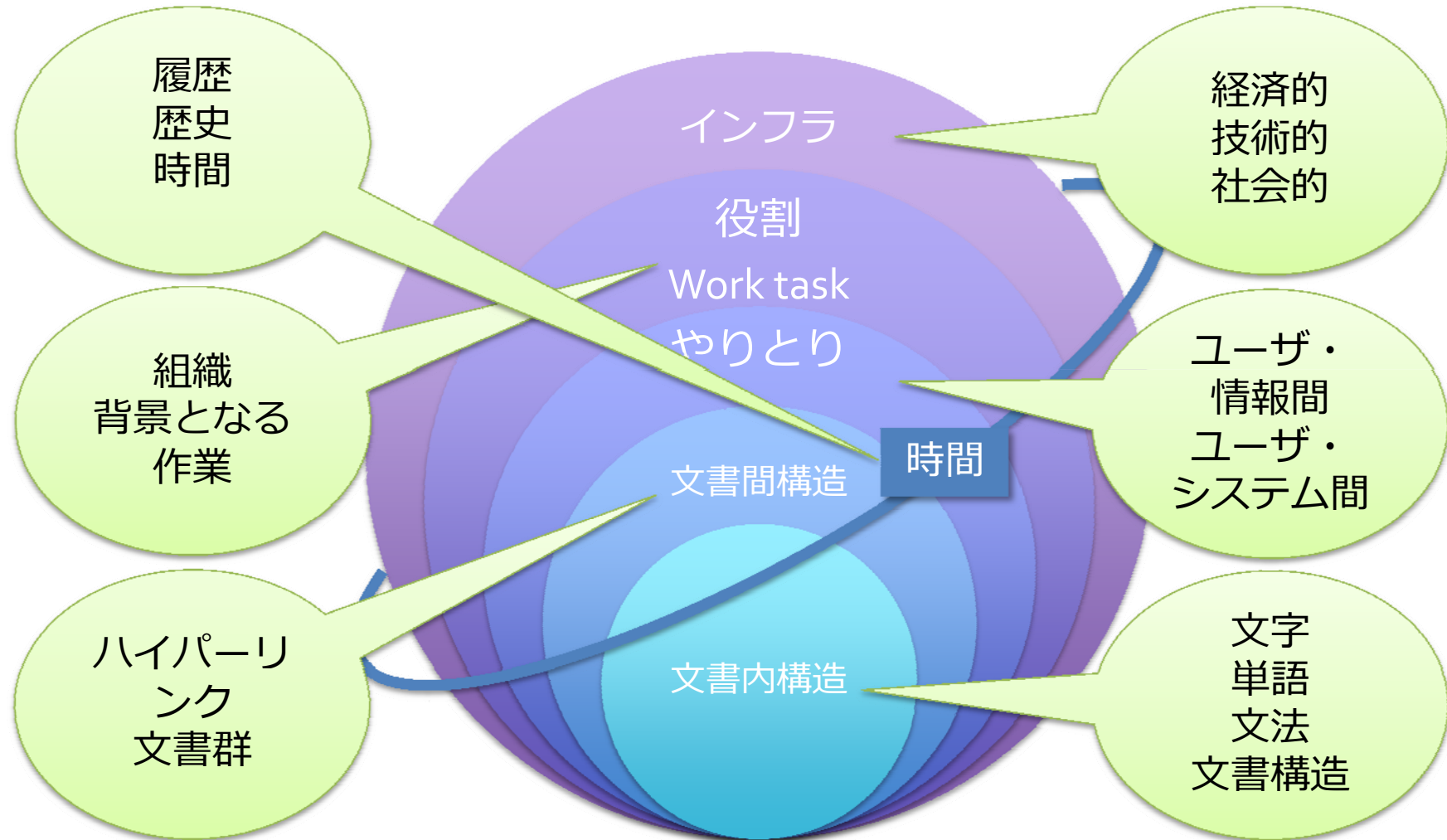
- Noriko Kando (National Institute of Informatics, Japan)

- Fredric C. Gey (University of California, Berkeley)

NTCIR-10 タスク・マップ

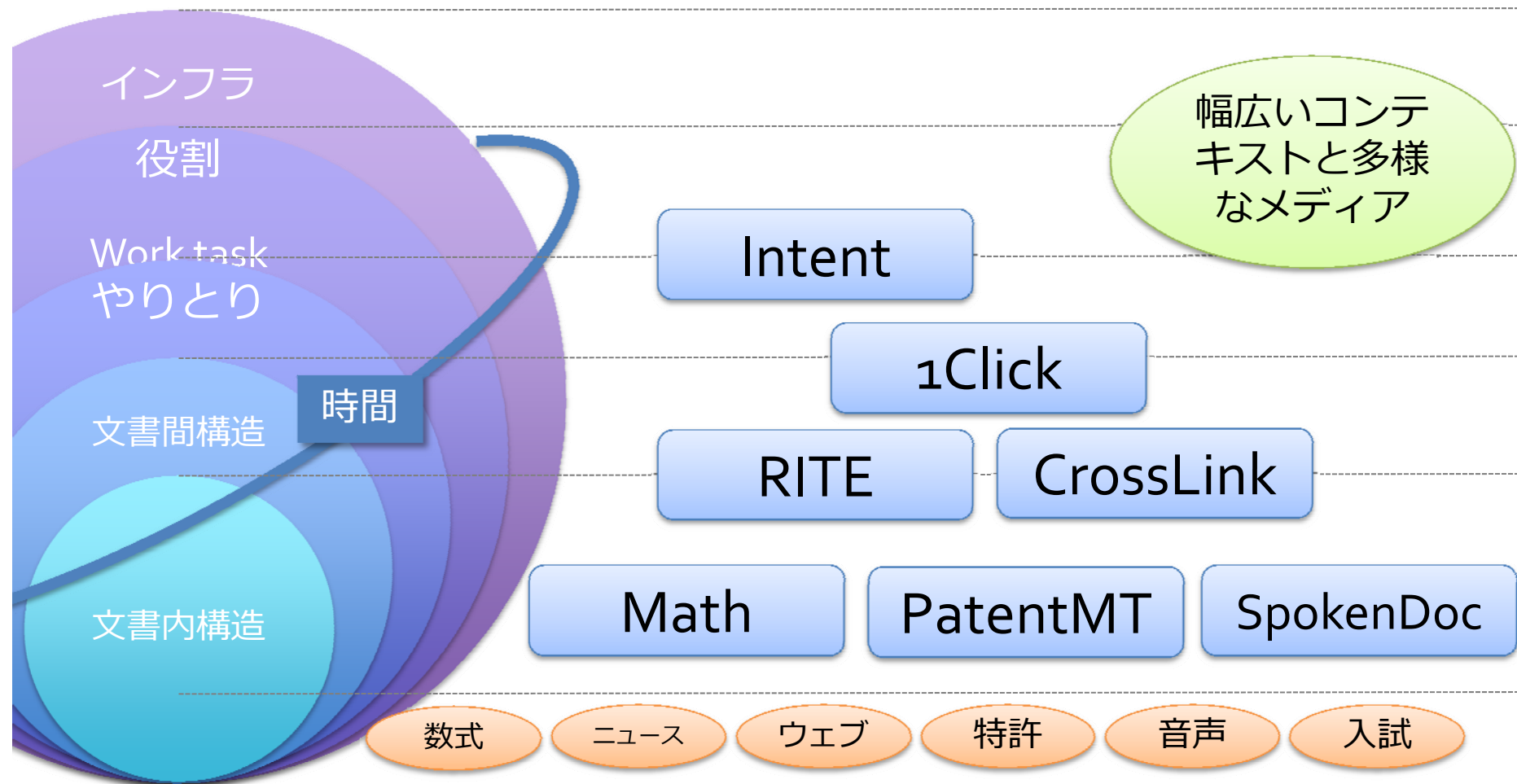
まとめ

情報アクセスのコンテキスト

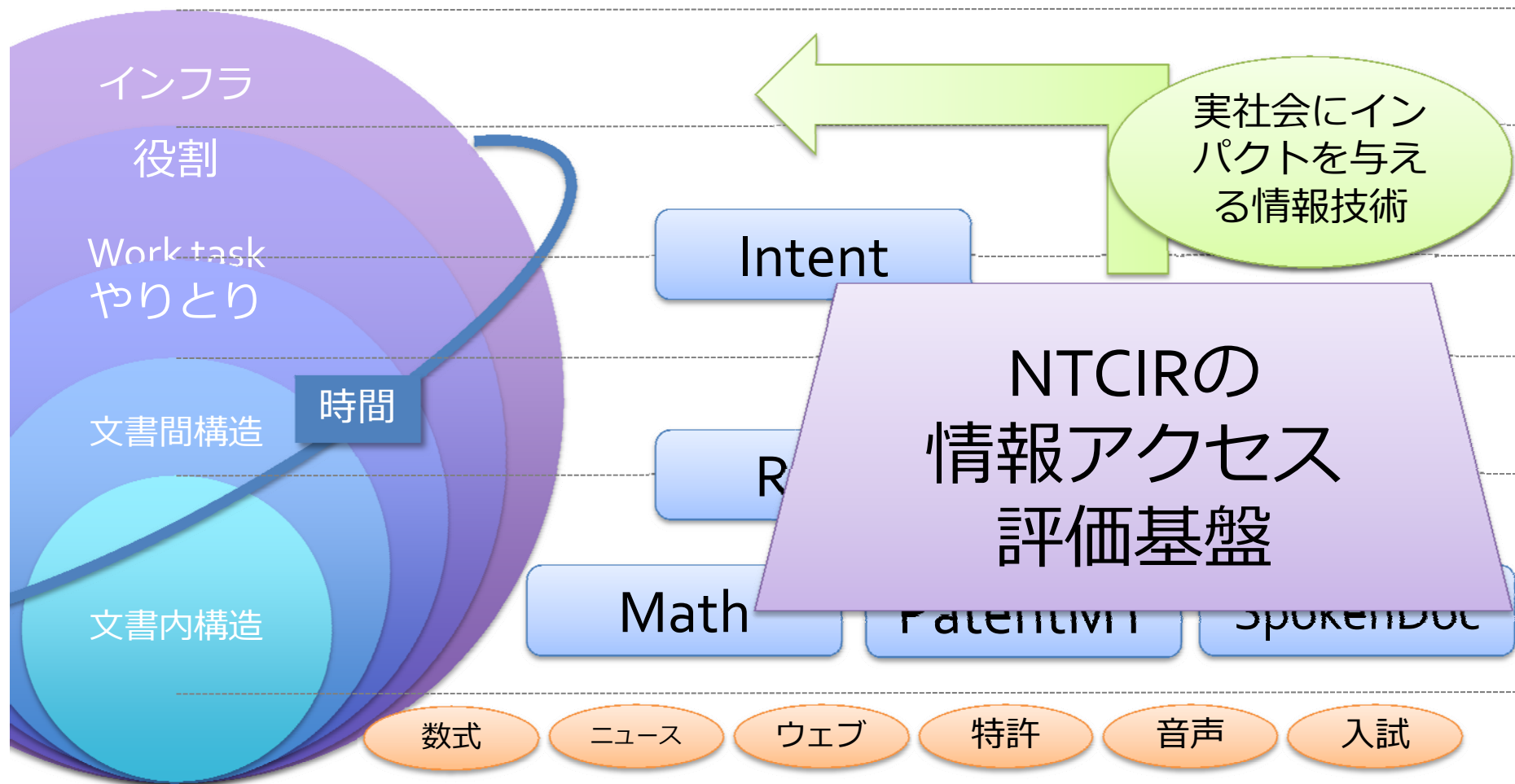


Adapted from Ingwersen & Järvelin (2005)

NTCIR-10 タスク



NTCIR-10の長期的展望



タスク参加のメリット

学生編・企業編

参加のメリット（学生編）

- すぐに始められる
 - 実験データの提供
 - （ある程度）定義された評価指標
- 業績
 - 他の参加者との比較による公正な議論
 - 次のステップについて国際コミュニティと意見交換
- 多様なタスク
 - 新しい課題への挑戦もあり、これまでの応用もあり

参加のメリット（企業編）

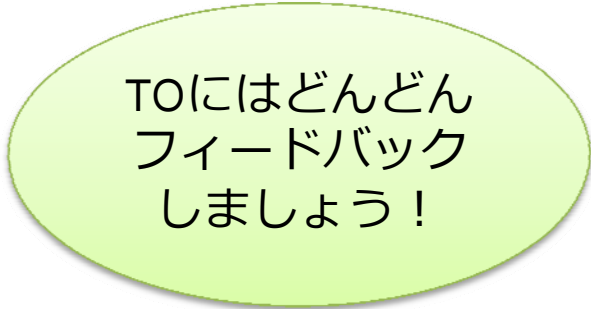
- プレゼンスの向上
 - 顧客に対して、同業他社に対して
- 公正なベンチマーク
 - 社内ベンチマークは自社製品に有利になりやすい
 - 入札時のアピール
- 開発の促進
 - 自社製品のブラッシュアップやバグ取りが短期間で集中的にできる
- 研究資源への早期アクセス
 - タスクで構築された様々なリソースを一般公開より先に利用可能

NTCIR-10への参加方法

簡単な6ステップ

参加方法

1. Task descriptionとCFPをよく読む
2. 質問はTOに問い合わせる
3. 参加するタスクを決める
4. NTCIRのウェブサイトから参加者登録をする
5. データ利用やその他の覚書を記入する
6. タスクのMLやウェブサイトフォローする



TOにはどんどん
フィードバック
しましょう！

NTCIR-10参加に関するお願い

参加者にしていただくこと

- 成果報告用論文の提出
- 成果報告会への出席 + ポスター発表

重要な日程

スケジュール帳にメモを

重要な日程

2012.03.08	キックオフイベント
2012.06.30	タスク参加登録締切
2012.07.01	文書データ配布開始
2012.07-2012.12	予備テスト
2012.09-2013.02	本テスト
2013.02.01	評価結果返送
2013.02.01	タスク概要報告一部公開
2013.03.01	成果報告会用論文 提出締切
2013.05.01	成果報告会用論文 最終原稿締切
2013.06.18-21	第10回NTCIR成果報告会 (東京)

詳細はTOから

まとめ

記念
すべき

- NTCIRの10回目のサイクル
 - 運営組織も更に強化！
- 7つの魅力的なタスク
 - 世界中の48人の研究者によって運営
- 斬新な研究をする土台を準備
 - 多様なコンテキストとメディア

**皆さんの
参加を
お待ちしております！**

NTCIR



<http://research.nii.ac.jp/ntcir/ntcir-10/>

ご清聴ありがとうございました！

お問い合わせはNTCIR事務局まで
ntc-secretariat@nii.ac.jp

Q & A