

Overview of NTCIR-10

Hideo Joho (University of Tsukuba)

Tetsuya Sakai (Microsoft Research Asia)

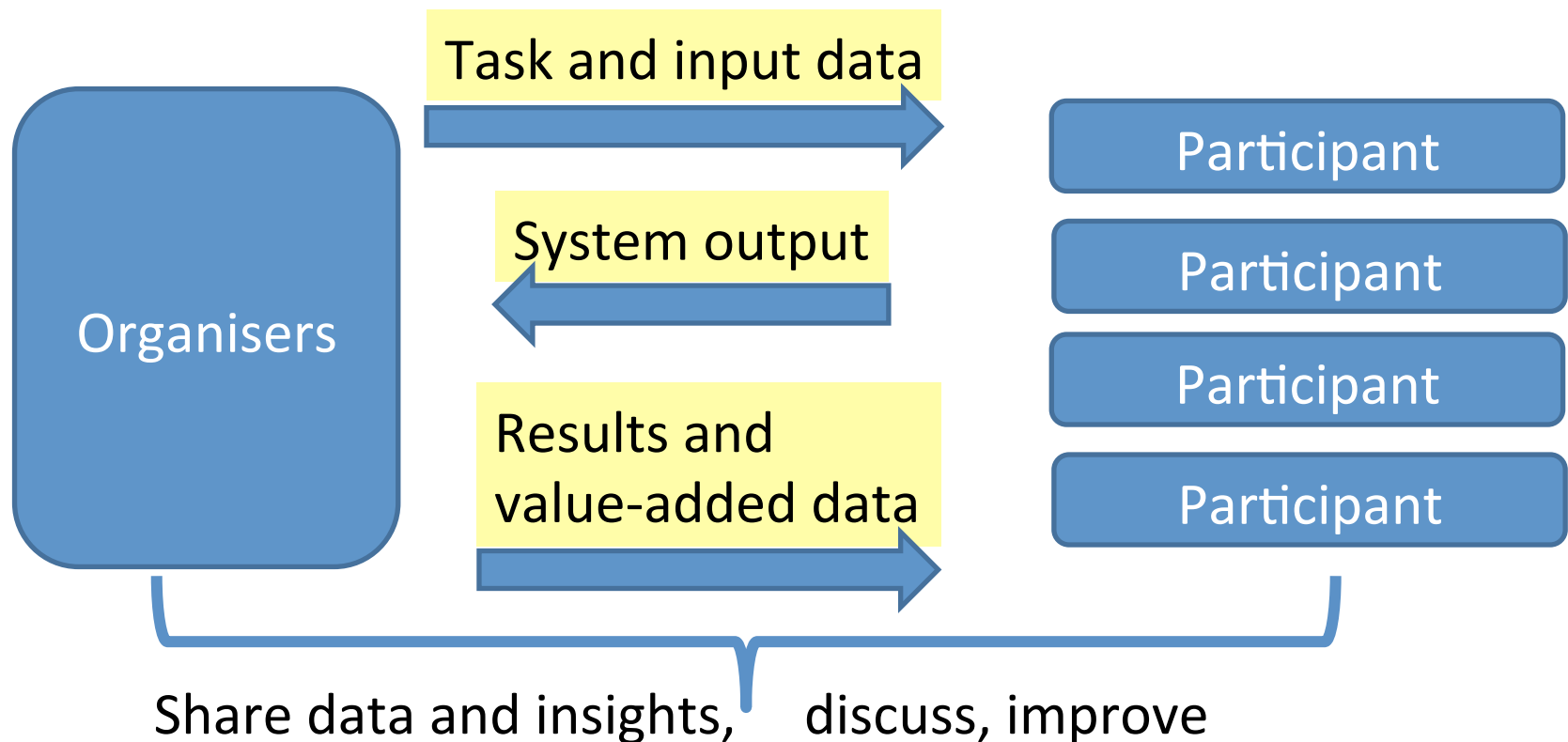
Outline

- Basic concepts
- Overview of NTCIR-10
- NTCIR-10 Evaluation Tasks
- What's Next? NTCIR-11
- Announcements
- Wrap-up

BASIC CONCEPTS

Evaluation Forums

- Research teams gather up to solve shared problems; submit system output before deadline
- Systems evaluated and compared across teams



Why Evaluation Forums?

- Compete and collaborate, accelerate research
- Build large-scale test collections through collective efforts e.g. *pooling*
- Foster interdisciplinary research towards grand challenges, build new research communities

NII Testbeds and Community for Information access
Research

Information Retrieval Evaluation Forums

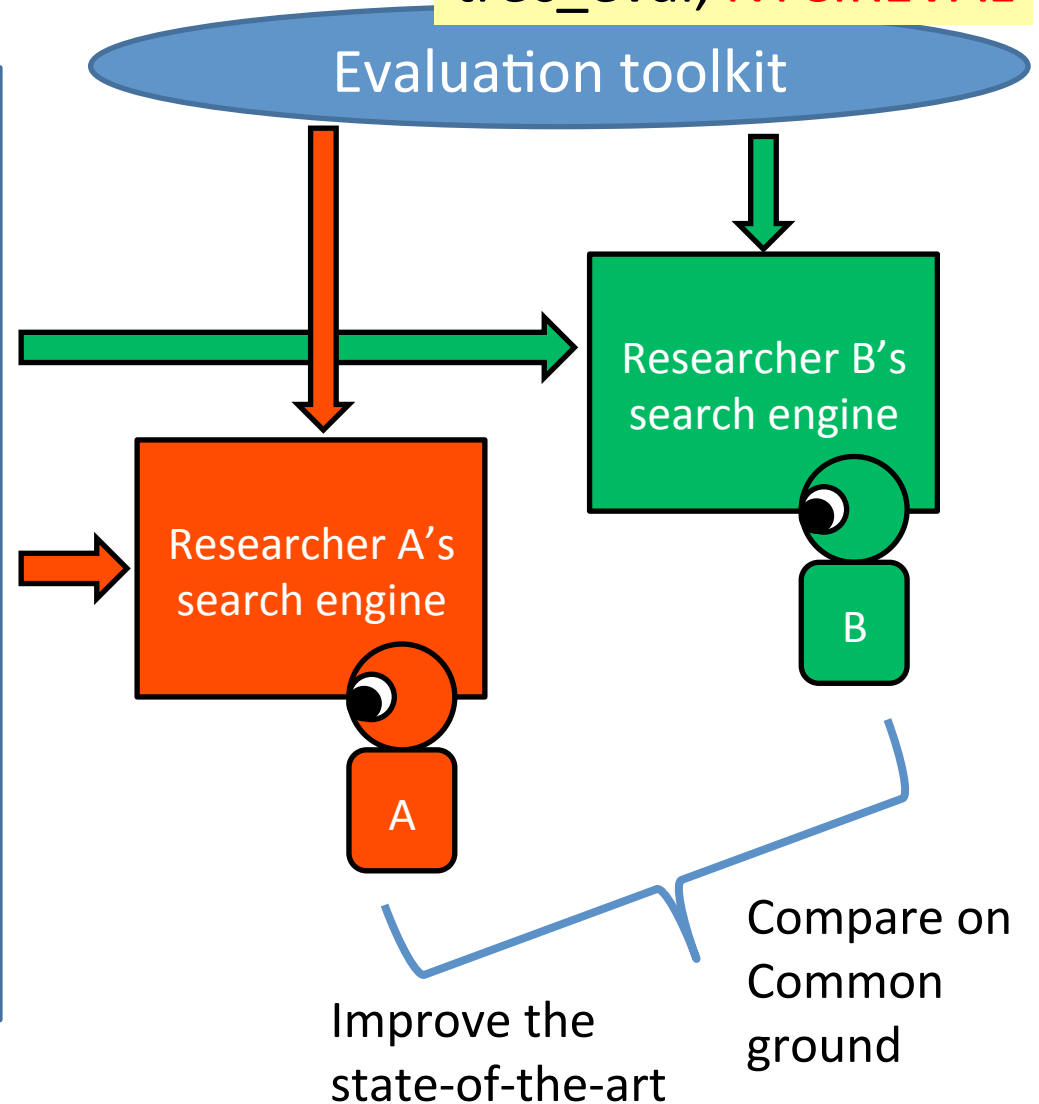
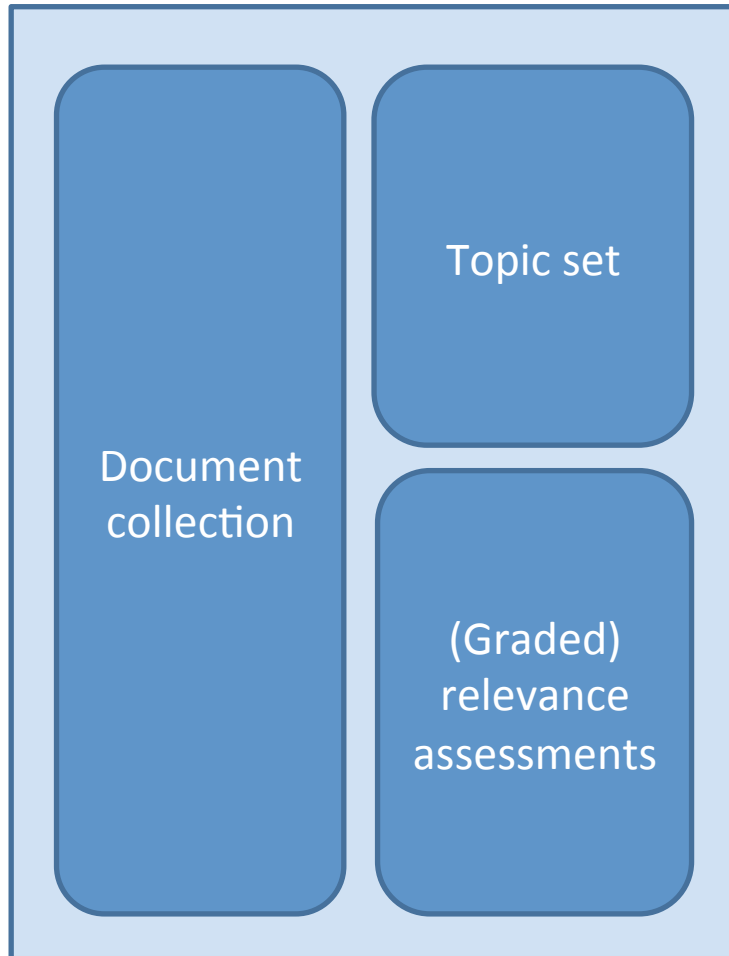
- TREC (Text Retrieval Conference) 1992-
- NTCIR 1998/9- [sesquiannual]
- CLEF (Cross-Language Evaluation Forum) 2000-
- INEX (Initiative for the evaluation of XML retrieval) 2002-
- TRECVID 2003-
- FIRE (Forum for IR Evaluation) 2008-
- MediaEval (Benchmarking Initiative for Multimedia Evaluation) 2010-

[not exhaustive]

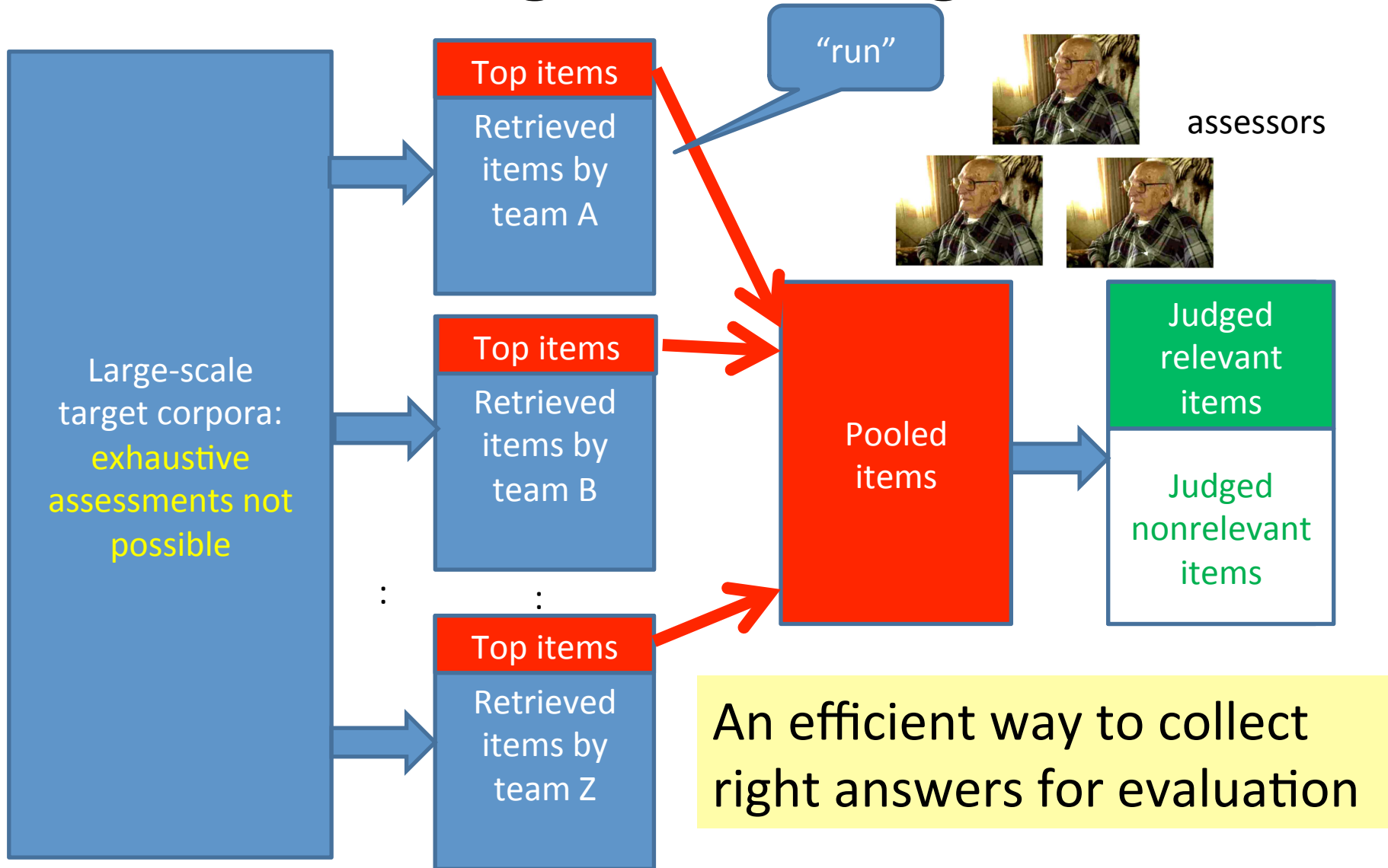
Jargon: Test Collection

trec_eval, NTCIREVAL

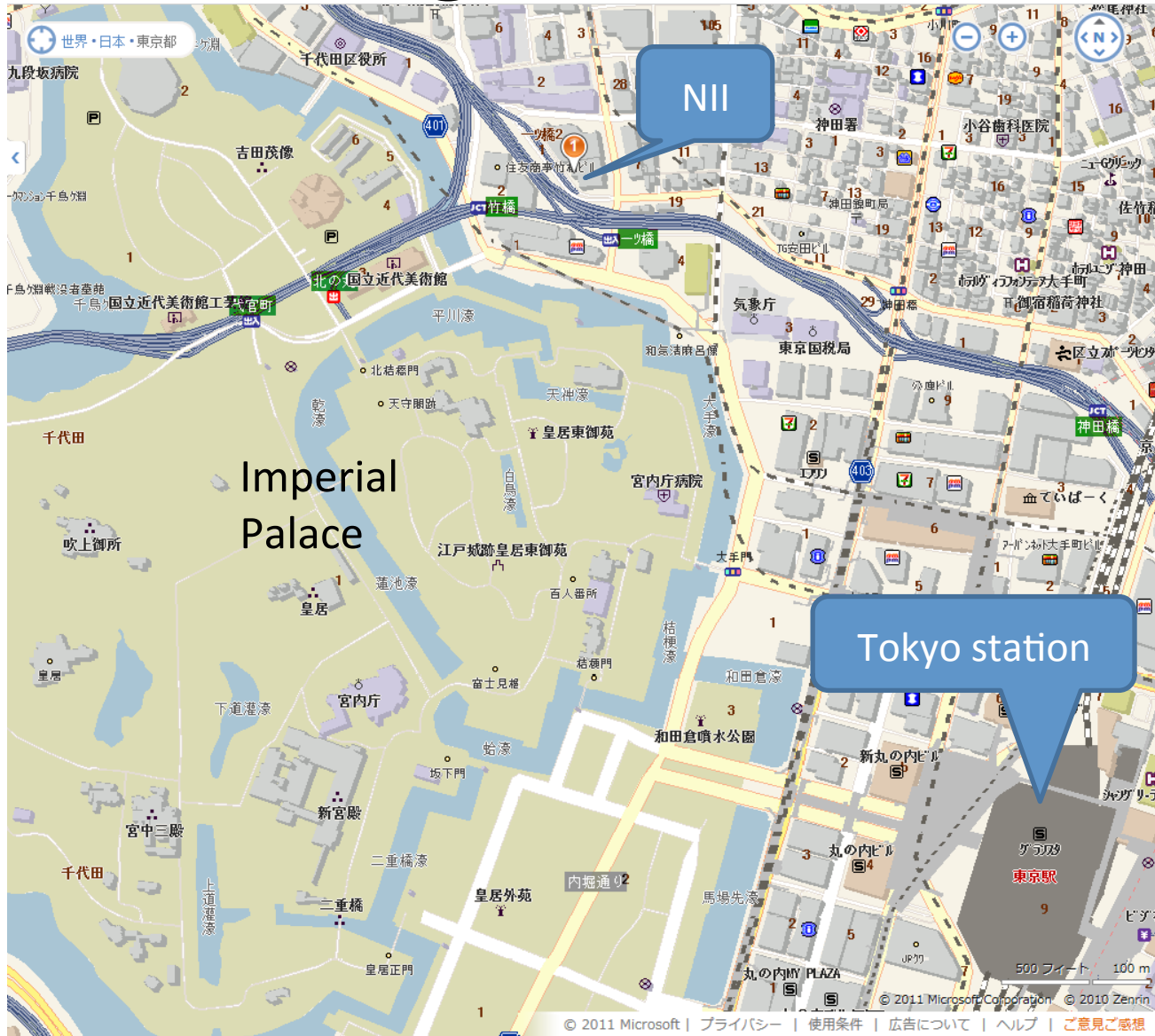
Test collection



Jargon: Pooling



Jargon: MAP



Query:
(I want to eat) sushi

Jargon: MAP

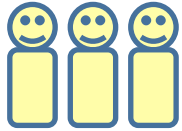
Mean
Average
Precision

[寿司 - Wikipedia](#)

語源説・種類・地方の寿司・歴史

寿司（すし、鮓、鮓、寿斗、寿し、壽司）と呼ばれる食品は、酢飯と主に魚介類を組み合わせた日本料理である。大別すると、生鮮魚介を用いた「早鮓（早ずし）」と、魚介類に米を加えて乳酸発酵させた「なれ鮓（なれずし）」に区分される。そのなかでも代表的な寿司は前者の握り寿司（江戸前寿司）であり、すでに“sushi”で通じるほど世界中に...

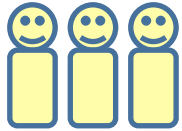
ja.wikipedia.org/wiki/寿司・スパムとして報告



[元気寿司グループ | 回転寿司のパイオニア](#)

11/09/14 千両・廻鮮日本海【秋期間限定メニュー】販売中!! 11/09/01 すしおんど【秋期間限定メニュー】販売中!! 9月1日(木)より 11/09/01 元気寿司【秋期間限定メニュー】販売中!! 9月1日(木)...

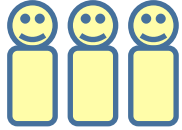
www.genkisushi.co.jp/index.php・スパムとして報告



[かっぱ寿司へようこそ!!](#)

かっぱ寿司で使用している国産の魚介類産地について かっぱ寿司で使用している魚介類の産地につきましては、「原産地情報」ページで日々最新情報をご提供しております。厳選素材の美味しさをお楽しみください「秋のデカ旨」開催中!

www.kappa-create.co.jp・スパムとして報告



[寿司\(鮓\)ランキング \[食ベログ\]](#)

全国にある寿司(鮓)のお店32,640件を一般ユーザーの口コミをもとに集計した様々なランキングから探すことができます。美味しい寿司(鮓)のお店や、人気の寿司(鮓)のお店が簡単に見つかります!

r.tabelog.com/sushi・スパムとして報告



[東京の寿司 マグロが自慢【すしざんまい】 tsukiji sushi-zanmai](#)

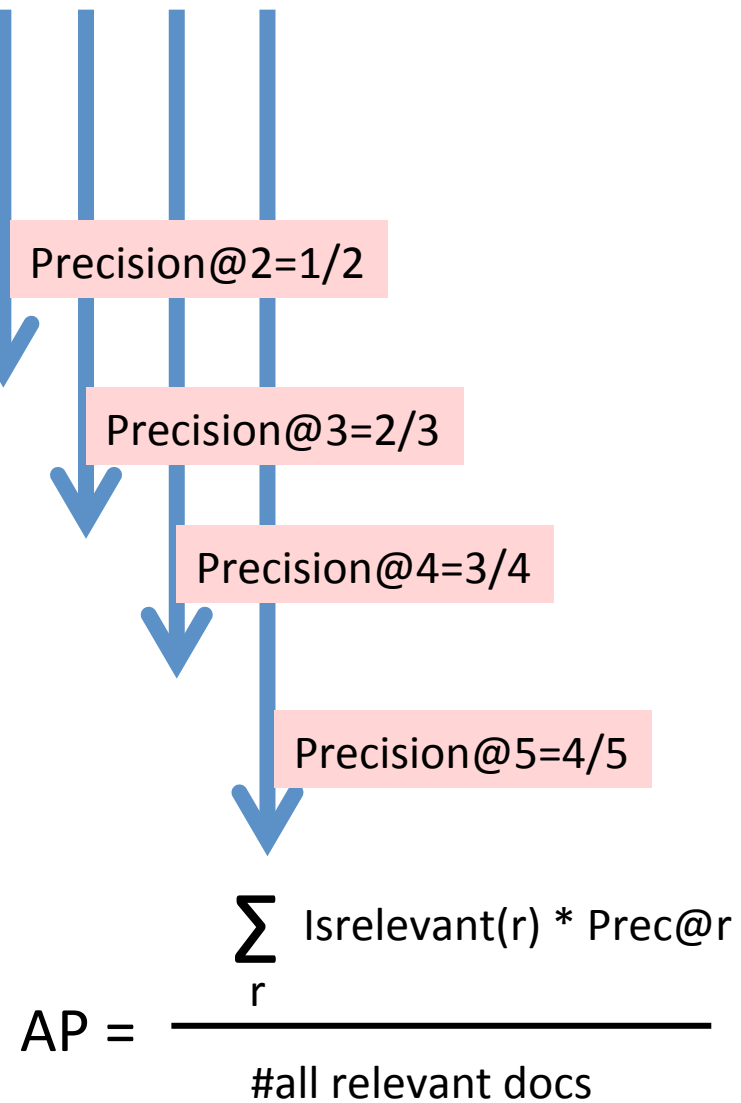
お寿司といえば『すしざんまい』!! 江戸前寿司の本場、東京で美味しい寿司が食べたいそんなあなた、『すしざんまい』へお越しください!! tsukiji sushi-zanmai ... 《What's NEW》特選おせち【三段重】予約申込受付中! 11/13(日)...

www.kiyomura.co.jp・スパムとして報告

[寿司 レシピ 689品 \[クックパッド\] 簡単おいしいみんなのレシピが109万品](#)

人気の寿司レシピ(作り方)が689品集まった寿司コミュニティ。毎日更新の寿司レシピランキング、みんなが料理した写真付きレポートなどが充実している。... 寿司 推薦レシピ: 689品 つくれぽ1000 人気手作り寿司のレシピ。お寿司のアイデア...

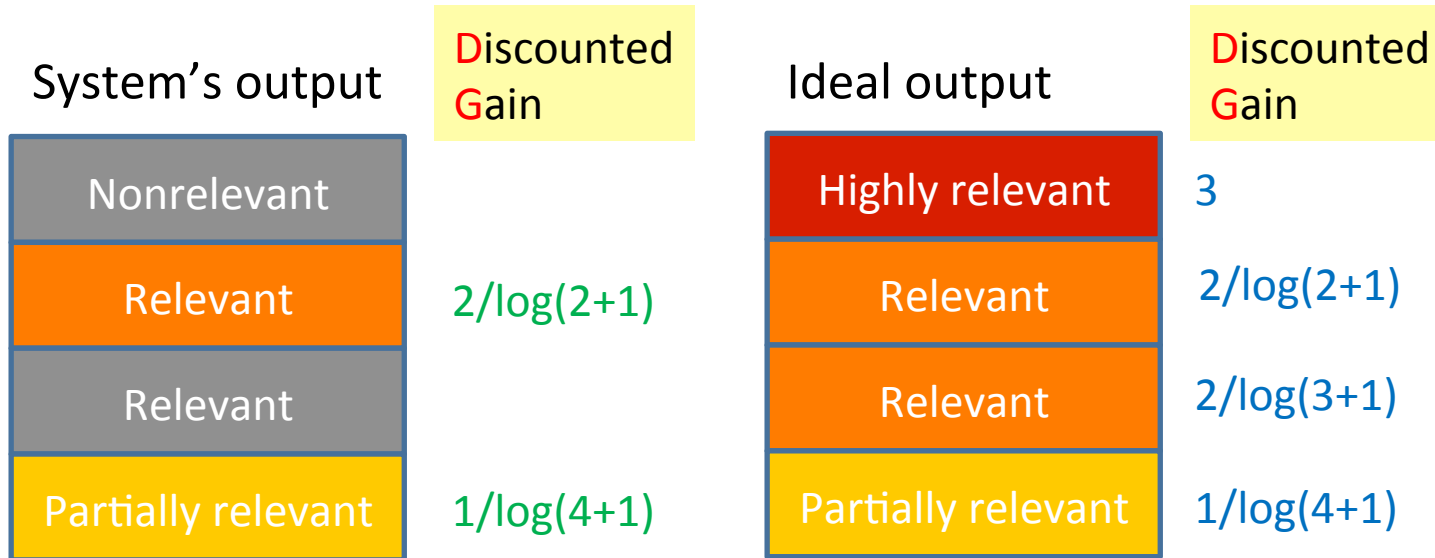
cookpad.com/category/165・スパムとして報告



User's stopping probability: uniform

Jargon: nDCG

n
D
C
G
normalised
Discounted
Cumulative
Gain



$$\text{nDCG} = \frac{\sum \text{System's discounted gains}}{\sum \text{Ideal discounted gains}}$$

Unlike AP, nDCG can utilise graded relevance assessments (widely used in Web search evaluation)

OVERVIEW OF NTCIR-10

NTCIR-10 Tasks

Core

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

Pilot **** NEW!! ****

- [Math] Access to mathematical contents
- [MedNLP] MedNLP

Tasks: NTCIR-1 to NTCIR-10

Year	1999	2001	2002	2004	2005	2007	2008	2010	2011	2013
Task/NTCIR round	1	2	3	4	5	6	7	8	9	10
Automatic Term Recognition and Role Analysis (TMREC)	9									
Ad hoc/Crosslingual IR(1) -> Chinese/English/Japanese IR(2) -> CLIR(3-6)	28	30	20	26	25	22				
Text Summarization Challenge (TSC)		9	8	9						
Web Retrieval (WEB)			7	11	7					
Question Answering Challenge (QAC)			16	18	7	8				
Patent Retrieval [and Classification] (PATENT)			10	10	13	12				
Multimodal Summarization for Trend Information (MUST)					13	15	13			
Crosslingual Question Answering (CLQA)(5,6) -> Advanced Crosslingual Information Access (ACLIA)(7,8)					14	12	19	14		
Opinion(6) -> Multilingual Opinion Analysis (MOAT)(7,8)						12	21	16		
Patent Mining (PAT-MN)							12	11		
Community Question Answering (CQA)								4		
Geotemporal IR (GeoTime)								13	12	
Interactive Visual Exploration (Vis-Ex)									4	
Patent Translation (PAT-MT)(7,8) -> Patent Machine Translation (PatentMT)(9,10)							15	8	21	21
INTENT									16	11
One Click Access (1CLICK) (subtask of INTENT at NTCIR-9)									4	8
Recognizing Inference in Text (RITE)									24	28
Crosslingual Link Discovery (Crosslink)									11	10
IR for Spoken Documents (SpokenDoc)									10	12
Math										6
Medical Natural Language Processing (MedNLP)										12

Number of active participants

Languages in NTCIR-10

	RITE	Patent MT	INTENT	1CLICK	CrossLink	Spoken Doc	Math	MedNLP
Chinese (simplified)		CE						
Chinese (traditional)								
English		EJ			CE, JE, KE			
Japanese		JE						
Korean								

Number of participating teams

Table 1: Number of participating teams by country/region

Country/Region	# of Teams
Japan	54
China	16
Taiwan	11
Germany	6
USA	4
Canada	3
France	3
Ireland	2
Korea	2
Australia	1
India	1
Czech Republic	1
United Kingdom	1
TOTAL	105

NTCIR-10 EVALUATION TASKS

Cross-lingual Link Discovery (CrossLink-2)

<http://ntcir.nii.ac.jp/CrossLink-2/>

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
...

이코노미스트
...

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

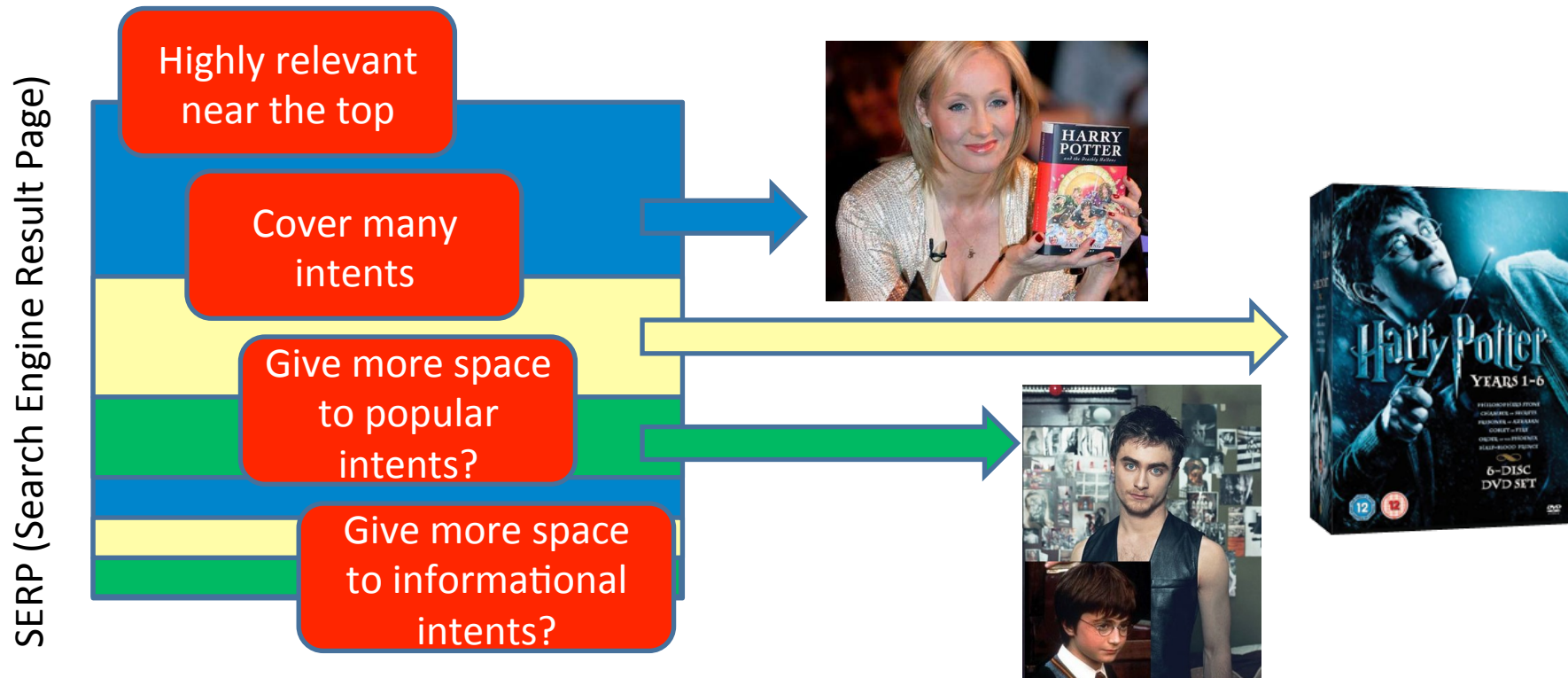
- ❖ Cross-lingual Links
- ❖ New Links
- ❖ Better Links
- ❖ More options



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

INTENT-2

- Given an ambiguous/underspecified query, produce a single Search Engine Result Page that satisfies different **user intents!**
- Challenge: balancing relevance and diversity



One Click Access (1Click-2)

English , Japanese

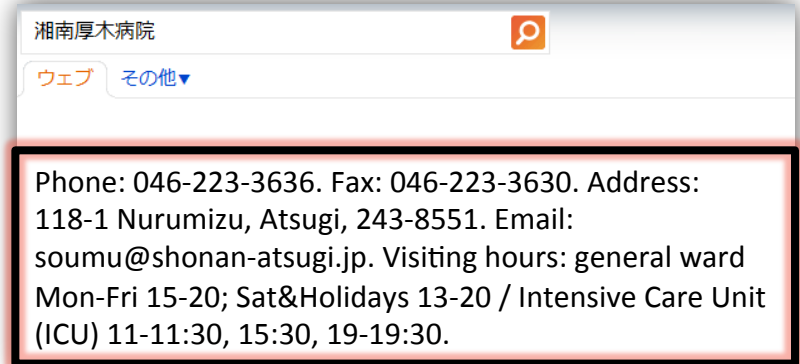
(Shonan Atsugi Hospital)

Enter query

Click SEARCH button

Get all desired information

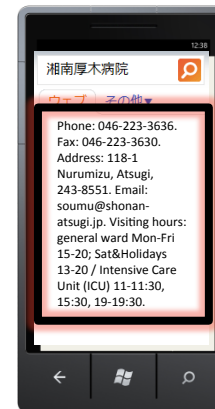
The system outputs *X-string*



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.



Particularly important for
mobile search

Go beyond the "ten-blue-link" paradigm in Web search

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**.

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.

- **English Scientific Papers**

[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
?**

MedNLP

■ De-identification task

Add Tags <a>: age, <x>: sex, <t>: time

■ Complaint and diagnosis task

Add Tags <c>: complaint, <d>: diagnosis

■ Free task

The dataset : **Japanese** Dis-Charge Summaries

■ Example input data: plain text

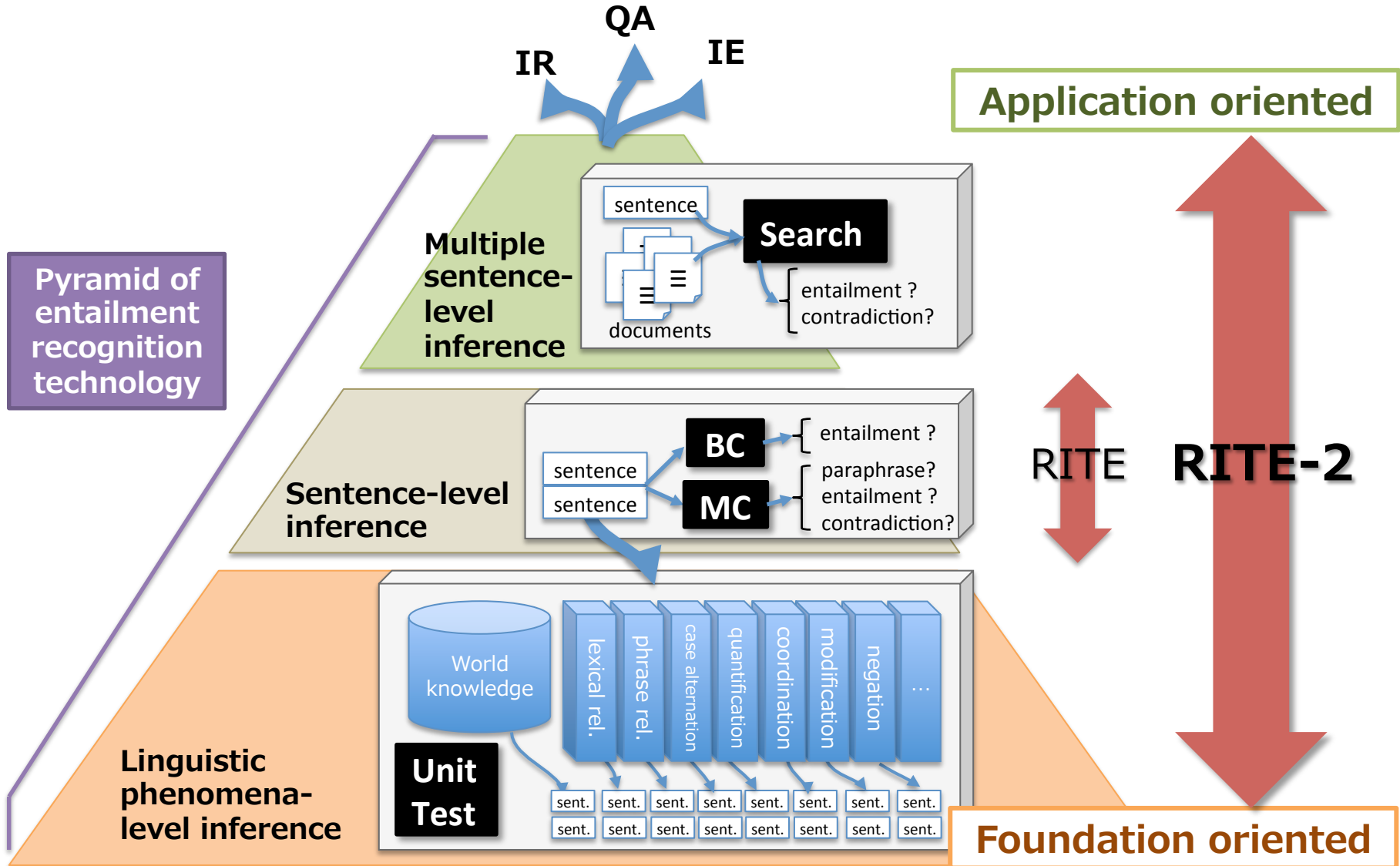
工場に勤めている64歳の男性。
2025年月8月2日(来院5日前)頃から腹痛が生じるとともに、食欲不振、嘔気・嘔吐出現した。
体幹は温かいが、抹消は湿潤冷汗でショック状態。明らかな運動麻痺はみられず。翌日、意識障害出現し、腎機能障害の増悪を認めて徐々に尿量低下し、8月9日18時10分に心肺停止。8月9日21時44分死亡確認。

■ Example output data: tagged text

工場に勤めている<a>64歳の<x>男性</x>。
<t>2025年月8月2日(来院5日前)頃から</t><c>腹痛</c>が生じるとともに、<c>食欲不振</c>、<c>嘔気</c>・<c>嘔吐出現</c>した。体幹は温かいが、抹消は<c>湿潤冷汗</c>で<c>ショック状態</c>。明らかな<c modality="negation">運動麻痺</c>はみられず。<t>翌日</t>、<c>意識障害出現</c>し、<c>腎機能障害</c>の増悪を認めて徐々に<c>尿量低下</c>し、<t>8月9日18時10分</t>に<c>心肺停止</c>。<t>8月9日21時44分</t><c>死亡確認</c>。

Best Systems will be incorporated in a UIMA workflow for NTCIR-11
Thank for Elsevier

NTCIR-10 RITE-2



SpokenDoc-2

- 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)
- Databases (**Japanese Only**)
 - 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



Highlights of NTCIR-10 Tasks

- Diverse Challenges
 - Smart Web Access, Cross-Languages, Multimedia, Knowledge Extraction, and Domain Expertise
- Chronological Evaluation
 - Repeat the experiments on NTCIR-9 dataset with new systems at NTCIR-10
- Collaboration with other Evaluation forums
 - Sharing relevant data, insights, etc.

WHAT'S NEXT? NTCIR-11

NTCIR-11

- 11th Cycle of NTCIR
 - July 2013 – December 2014
- NTCIR-11 Conference: **9-12th Dec 2014**
- 11 Task Proposals were submitted
 - Scheduled to announce the result on 26th June
- Task Proposal Presentation
 - **14:10 on 21st June (Day-4)**

ANNOUNCEMENTS

Tetsuya to join General Co-Chairs



- Program Co-Chairs of NTCIR-9 and NTCIR-10
- Participated all NTCIRs!
- Task organiser of 1Click and INTENT
- Q-measure, D-measure, I-measure

Kazuaki to join Program Co-Chairs

- Prof. Kazuaki Kishida at Keio University
- IR, Clustering, Bibliometrics, Evaluation
- CLIR Task organiser at NTCIR-4, 5, 6
- Program Co-Chairs of EVIA 2011



Please join us to celebrate **10**

NTCIR (NII Testbeds and Community for Information Access Research) Project

NTCIR Japanese

About NTCIR FAQ Search

Publications/Online Proceedings Data/Tools NTCIR CMS Site Related URLs Contact us

NTCIR Home > NTCIR-10 > NTCIR-10 Conference

NTCIR-10 Conference

The 10th NTCIR Conference

Evaluation of Information Access Technologies

EVA2013

The 5th International Workshop on Evaluating Information Access (EVA 2013)
A satellite workshop of the NTCIR-10 Conference

**June 18-21, 2013 at National Center of Sciences, Tokyo, Japan
National Institute of Informatics (NII)**

[CO-LOCATED: ASSIA 2013: Asian Summer School in Information Access](#)
June 22-24, 2013 at Tsukuba, Japan

Welcome

The memorable 10th NTCIR Conference and The 5th International Workshop on Evaluating Information Access (EVA 2013) will be held in June 18-21, 2013, at NII in Tokyo, Japan. Anybody is welcome to attend and join the conference. We had valuable contributions and warm discussions during the previous conference with 225 attendees from 16 different countries. The results of participating teams, as well as comparative evaluation results will be presented during the 10th NTCIR Conference. We are planning to have exciting invited talks, including talks by speakers from the Web search engine companies.

[Leave a message to celebrate the 10th NTCIR](#)

Over 15 Years of Solid Work NTCIR-10 Conference website

Celebrating the 10th Cycle of NTCIR!

Please leave your message to join us celebrating its 15 years of activity and 10th cycle of NTCIR! Just a quick message is great. Telling us how NTCIR helped your research and career would be perfect.

[Leave a message](#)

Please note all comments will be screened for spam removal (Texts won't be edited).

第10回の節目を迎えたNTCIRへのお祝いメッセージを募集しています。日本語で結構です。簡単なメッセージだけでも構いません。もしNTCIRがどのように皆様のお役に立ったかを書いていただけるとなうれしいです。

[メッセージを残す](#)

コメントはスパム防止のために一度確認致しますので、あらかじめご了承ください (文章は編集致しません)。

NTCIR-10 General Co-Chairs

- Noriko Kando (National Institute of Informatics)
- Tsuneaki Kato (The University of Tokyo)
- Douglas W. Oard (University of Maryland)
- Mark Sanderson (RMIT University)

NTCIR-10 Program Co-Chairs

- Hideo Joho (University of Tsukuba)
- Tetsuya Sakai (Microsoft Research Asia)

Wrap up

- It's been 15 years, 10th running of NTCIR
- Largest number of participants ever
- Infrastructural change made in NTCIR-9 was a huge success
- Strive to develop the communities further
- Your feedback is crucial!

