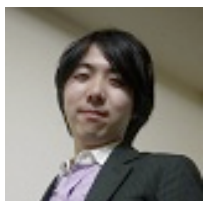


Overview of the NTCIR-11 MobileClick Task

Makoto P. Kato (Kyoto University),
Matthew Ekstrand-Abueg, Virgil Pavlu (Northeastern
University), Tetsuya Sakai (Waseda University),
Takehiro Yamamoto (Kyoto University) and Mayu Iwata
(KDDI Corporation)

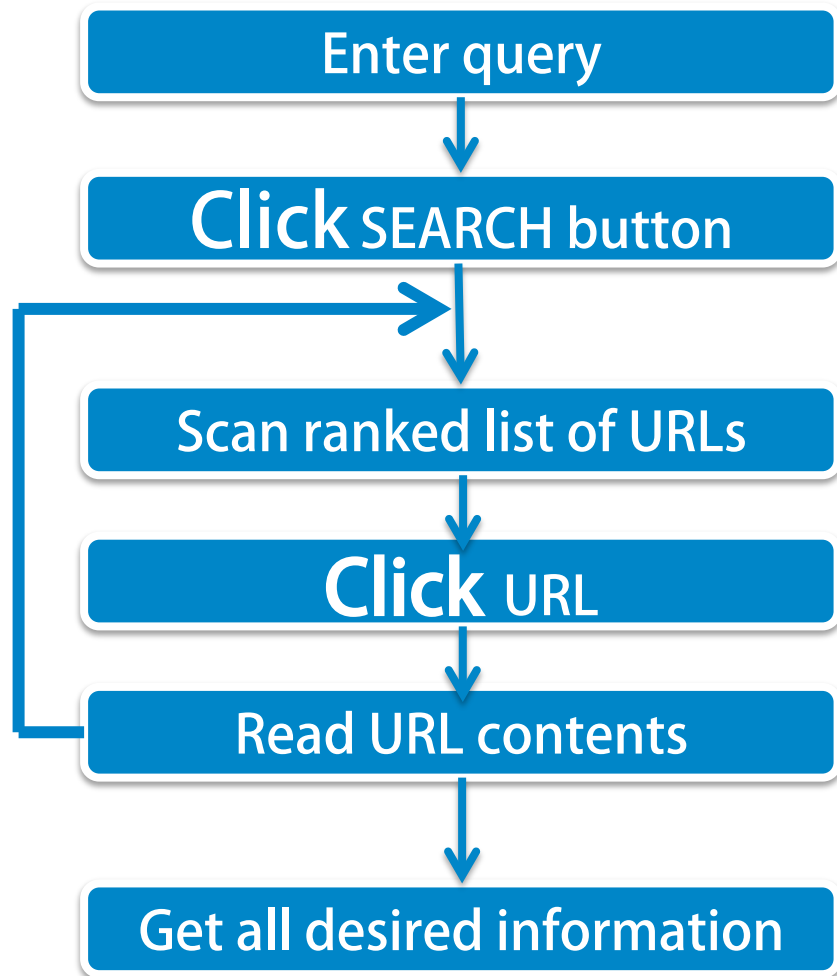


Suppose that ...

- You are finding answers for a question “what’s the difference between Organic EL and LCD?” in an electronics store

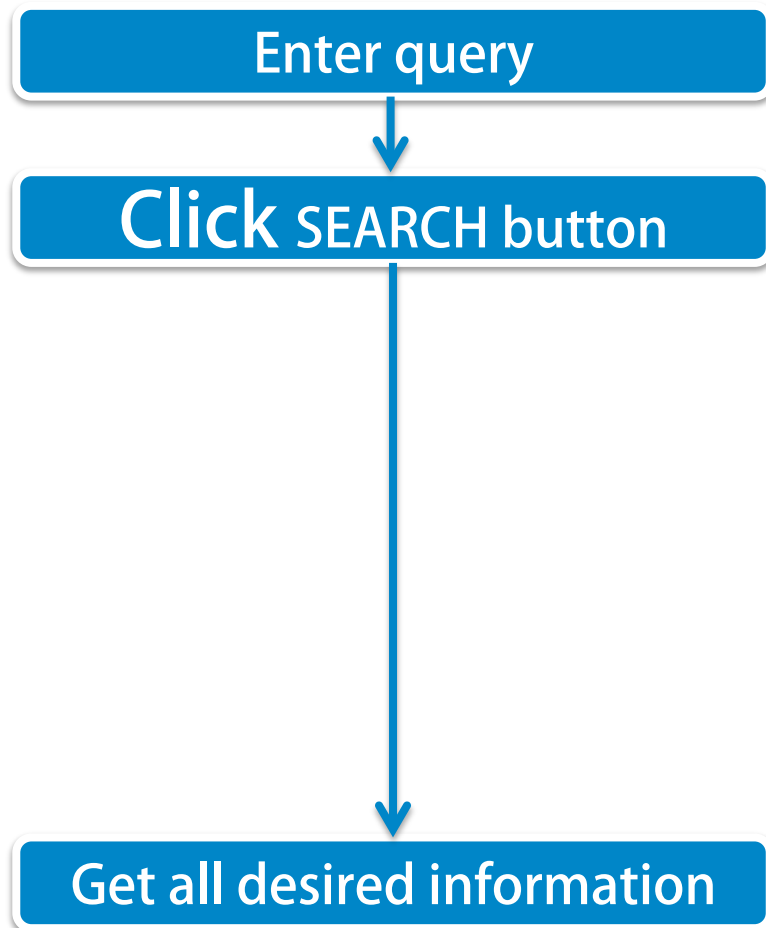


IR Systems in *Ten-Blue-Link* Paradigm

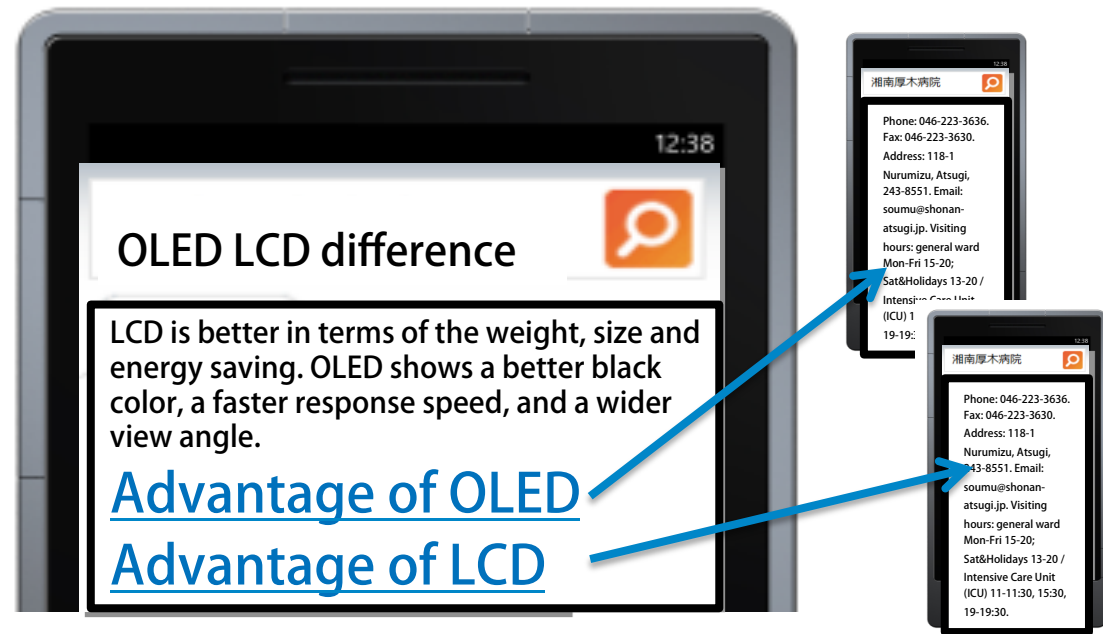


Long way to get all desired information

MobileClick System





System output





Task: Given a search query, return a **two-layered** textual output

Go beyond the "ten-blue-link" paradigm, and tackle **information** retrieval rather than document retrieval

- Present a concise summary (like "Knowledge Graph") to **any kind of queries** in a way that **any users** can easily reach their desired info.

Web Images Videos Maps News More

日本のBing.com 0 Sign in  

3,370,000 RESULTS Any time ▾


KYOTO UNIVERSITY - 京都大学
www.kyoto-u.ac.jp/en ▾ Translate this page
01 Dec 2014. Announcement date for tuition exemption (and admission fee exemption and deferment) application results for the second semester of the 2014-2015 ...

Contact Information Contact Us — KYOTO UNIVERSITY - 京都大学	Employment Oppor... Opportunities — KYOTO UNIVERSITY - 京都大学	Admissions
Admissions Informa... A research team has developed a novel way to ...	University Library NEWS [Usage of OSLs] UPDATE : Apr 25, 2014 * ...	
Japanese ... 京都市左京区高野西開町34番4 ...	Access Map Access — KYOTO UNIVERSITY. Kyoto ...	


[See results only from kyoto-u.ac.jp](#)

Kyoto University - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Kyoto_University ▾
Kyoto University (京都大学, Kyōto daigaku ?), or Kyodai (京大, Kyōdai ?) is a national university located in Kyoto, Japan. It is the second oldest Japanese ...
[History](#) · [Campuses](#) · [Organization](#) · [Academics](#) · [Academic rankings](#) · [Athletics](#)

Bing Local



Kyoto University



Kyoto University, or Kyodai is a national university located in Kyoto, Japan. It is the second oldest Japanese university, one of the highest ranked universities in Asia and one of Japan's National Seven Universities. One of Asia's leading res... +
en.wikipedia.org

en.wikipedia.org

Address: Yoshida Honmachi, Sakyo-ku, Kyoto, Kansai region 606-8501

Undergraduates: 13,421 (2013)

Tuition: ¥535,800 JPY (2014)

Founded: May 01, 1869

Phone: +81 75-753-2047

Queries and Documents

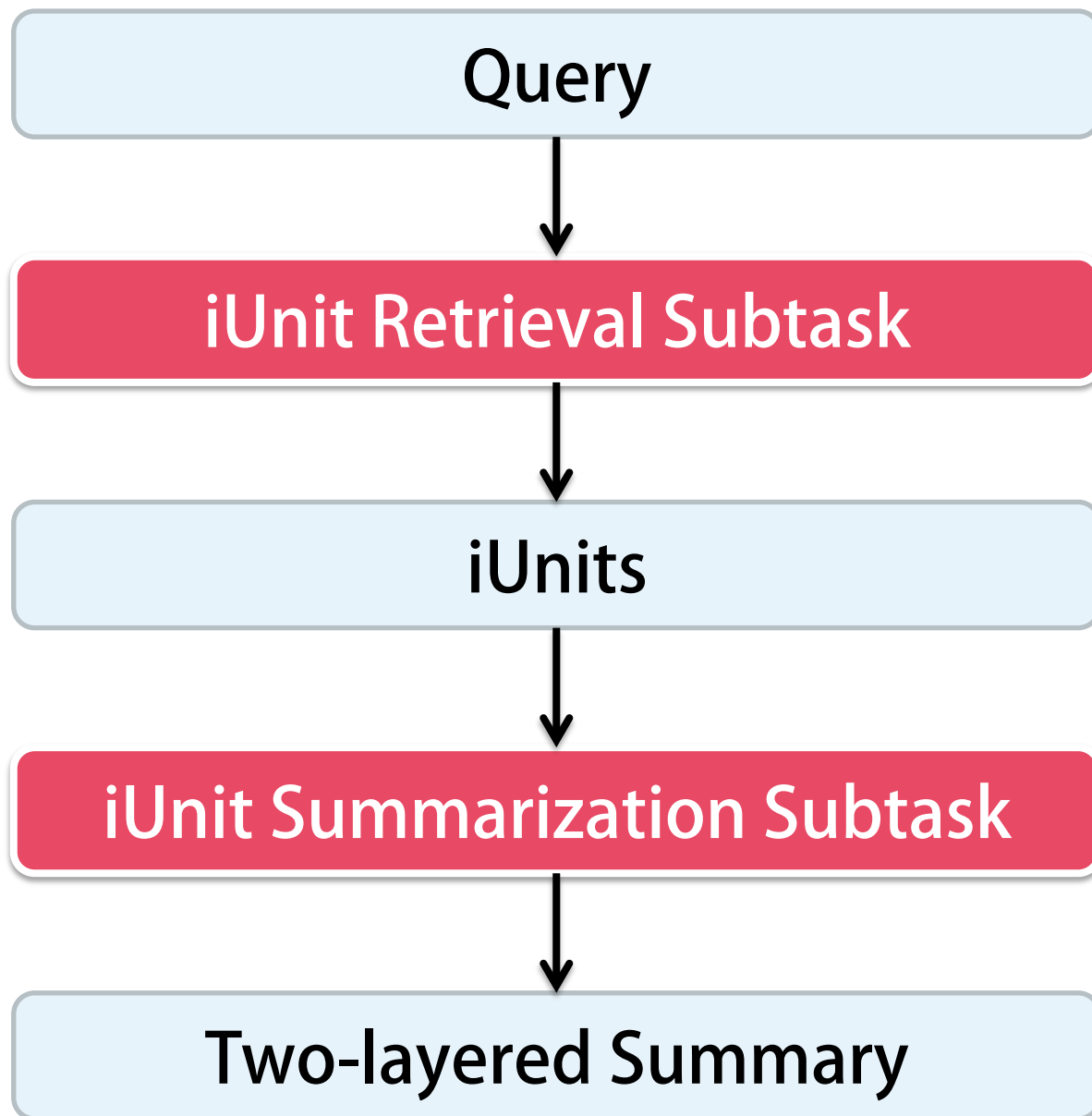
- Queries
 - 50 English/Japanese queries
 - Most of which were highly specific (e.g. java vs python text processing)
 - Based on 1CLICK-2 task at NTCIR-10, the first paragraph of Wikipedia articles were good enough for named entity queries

Examples

hiphop clubs barcelona	Michoacan crafts materials	sears illinois
why does turkey make you sleepy	ron paul tea party	aaron rodgers belt celebration
french landmarks	syrian civil war players	ukraine debt

- Documents
 - 300 - 400 commercial search engine results for each query
 - From which summaries are generated

Two Subtasks



OLED LCD difference



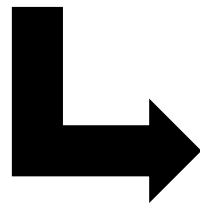
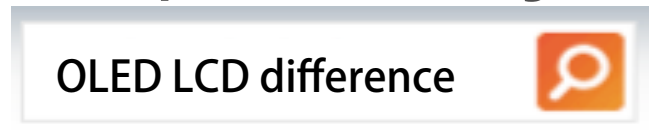
Score	iUnit
5	LCD is lighter than OLED
4	OLED shows a better black color
3	OLED has a wider view angle
...	...



iUnit Retrieval Subtask

- Generate a list of *iUnits* ranked according to their importance for a given query
 - iUnits:
atomic information pieces relevant to a given query

Input: **Query**



Output: **List of iUnits**

Rank	iUnit
1	LCD is lighter than OLED
2	OLED shows a better black color
3	OLED has a wider view angle
...	...

iUnit Summarization Subtask

- For a given query and a list of *iUnits*, generate a two-layered textual output

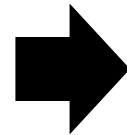
Input: **Query**

OLED LCD difference



Input: **List of iUnits**

Rank	iUnit
1	LCD is lighter than OLED
2	OLED shows a better black color
3	OLED has a wider view angle
...	...



Output:

Two-layered textual output

LCD is better in terms of the weight, size and energy saving, while OLED shows a better black color, a faster response speed, and a wider view angle.

Advantage of OLED

Advantage of LCD

OLED shows a better black color, a faster response speed, and a wider view angle.

LCD is better in terms of the weight, size and energy saving.

Example of iUnit Summarization Run

```
<result qid="MC-SAMPLE-E-0001">
```

```
<firstlayer>
```

LCD is better in terms of the weight, size and energy saving.

OLED shows a better black color, a faster response speed, and a wider view angle.

```
<link id="1">Advantages of OLED</link>
```

```
<link id="2">Advantages of LCD</link>
```

```
</firstlayer>
```

```
<secondlayer id="1">
```

OLED is...

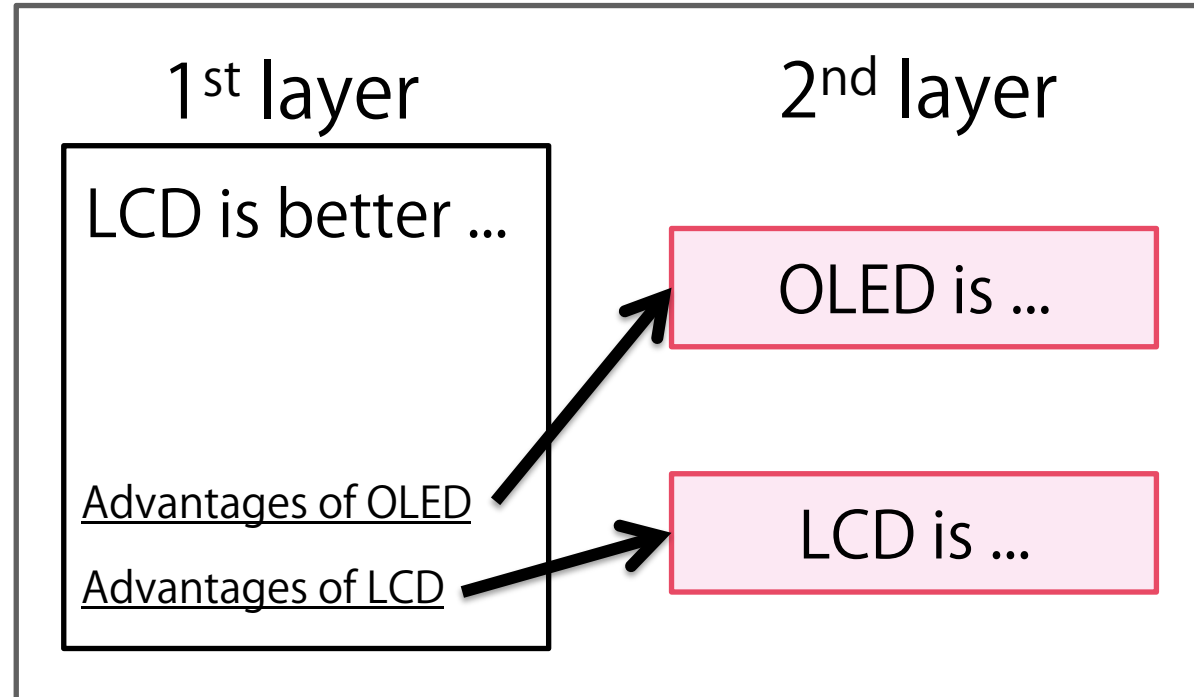
```
</secondlayer>
```

```
<secondlayer id="2">
```

LCD is

```
</secondlayer>
```

```
</result>
```



- **Gold Standard iUnits**
 - For each query, assessors created gold standard iUnits (GiUnits) based on the document collection
- **iUnit Matching**
 - Assessors recorded the position at which GiUnits appear in the system output
- **Evaluation Metrics**
 - iUnit Retrieval: nDCG and Q-measure
 - iUnit Summarization: M-measure

Creating Gold Standard iUnits (GiUnits)

Queries

Query: Snow gum tree planting Category: Fact Finding

[Instructions](#)

Document [Prev](#) MC-E-0033-018 [Next](#)

not grow well in waterlogged soils, as a lack of air space in this type of soil can cause the plant's roots to rot and fungus to develop. Snow gum plants can grow well in dry climates and semi-arid conditions. This means that a Snow Gum's water requirements are relatively low. The best way to water these types of plants is about once week, depending on the amount of soil drainage and naturally occurring rainfall, and to soak the soil completely when watering. This is especially true during the plant's first two years of growing. It is important to avoiding overwatering a Snow Gum plant and in cases where water may be too abundant, it may be necessary to provide extra drainage. Snow Gum trees have a root system that is fibrous, so they can be planted on a slope and still grow well. This is useful to many yard types, as this tree will often grow readily where other plants will not. The pH of the soil is also important to successfully grow these plants. Snow gum trees prefer a pH range of about 5.5 to 6.5. A soil ph test is easy to perform and will help you to determine what products you may need to add to the soil in order to maintain this pH range.

Sunlight and Fertilizer Requirements

Snow Gum trees prefer full sunlight or partial shade. This means that they should be planted in a location that receives at least 3 to 6 hours of direct sunlight. These plants should also be fertilized to promote the

Nuggets

Search:

Nugget	Dependencies	Importance
20. water requirements are relatively low		Low ▾
21. water about once weekly	water requirements are relatively low	Med ▾
22. prefer a pH range of about 5.5 to 6.5		Med ▾
23. plant in a location that receives at least 3 to 6 hours of direct sunlight		Med ▾
24. fertilize with a product that contains a higher phosphorus level.		Med ▾
25. [seeds from] species of the "snow gum" found in colder areas provide a better germination rate when they		Med ▾

[Unselect Nugget](#) Click nugget to modify or select new nugget from document

[Edit Nugget](#) [Delete Nugget](#) [Jump To Source Document](#) | [Store Changes](#)

New Nugget:

 [Add Nugget](#)

- Atomic and relevant information pieces were recorded as GiUnits with their importance

- For query "LCD OLED difference"

OLED is better in contrast

Less afterimage in OLED

Less power consumption for LCD

LCD is lighter

The display of OLED consists of gas

The display of LCD consists of liquid crystal

OLED is self-luminous

LCD uses backlight

LCD has a narrow view angle

OLED's contrast in dark places is better

OLED is excellent in expressive power of black

OLED is easily scalable

LCD shows smooth gradation

iUnit Matching

Queries

Query: Snow gum tree planting Category: Fact Finding

Instructions

Updates

First Page Prev Page 1-20 / 916 Next Page 20 ▾ per Page

Once germination has taken place, ideally, you should remove the container of seedlings and place it in an area of bright light and provide a lower temperature of 55-60 degrees F for several weeks

[seeds from] species of the "snow gum" found in colder areas provide a better germination rate when they have been cold stratified

Rare snow gum tree

Plant seeds in spring or summer keep shaded and water sparingly until 2 to 3 inches high.

Snow Gum will grow in light (sandy),medium (loamy),hard (clay) soil. It is important for the soil to be well drained

Nuggets

Search:

Nugget	Dependencies	Importance
1. water about once weekly	water requirements are relatively low	Med ▾
2. tolerating temperatures down to 0 degrees Fahrenheit		Low ▾
3. Well drained soil is preferable.		Med ▾
4. Trees do very well planted from seed		Med ▾
5. Plant seeds in spring or summer keep shaded and water sparingly until 2 to 3 inches high.		High ▾
6. drought tolerant and have very few pests.		Low ▾
7. Sow both seed and chaff on		

Unselect Nugget

Click nugget to match or modify

Select New Nugget

Nugget:

Edit

Delete

Split

Merge

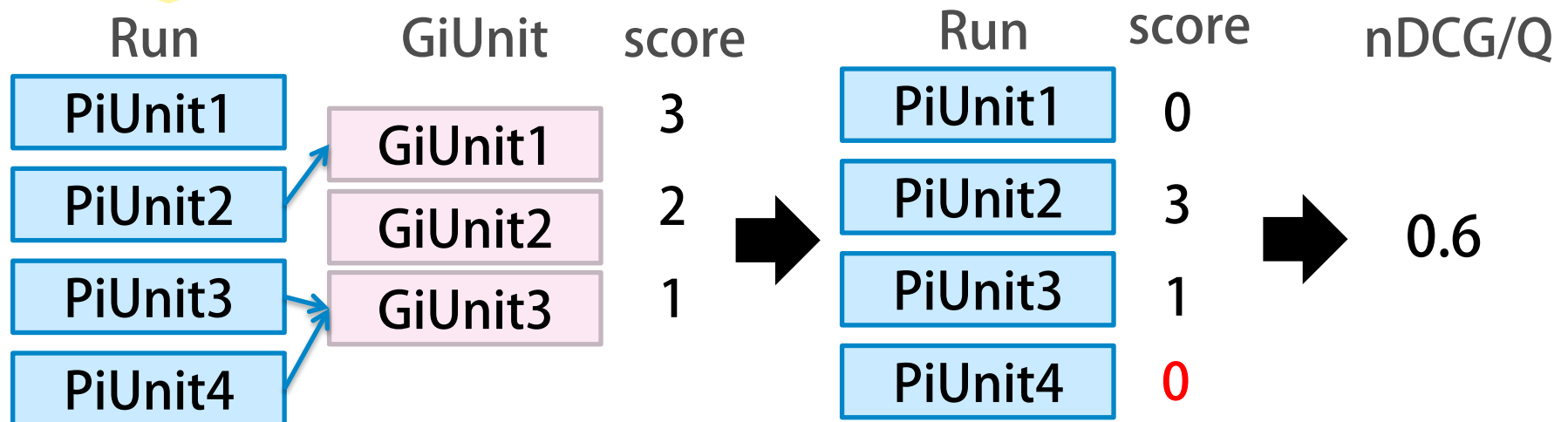
Store Changes

Show Positions

- GiUnits in the system output were manually recorded with their position

Evaluation of iUnit Retrieval Runs

Identify GiUnits from each participant iUnits (PiUnits)



Give scores to PiUnits based on the scores of identified GiUnits, where redundant PiUnits were ignored

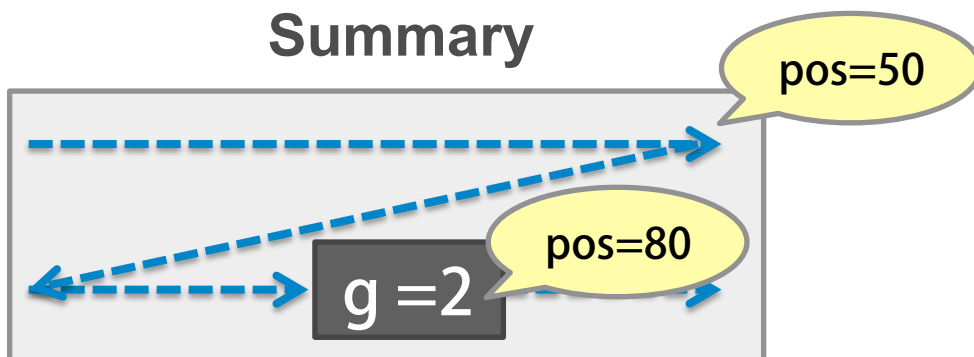
Evaluation Principles for iUnit Summarization (1/2)

- 1. The utility of a summary is measured by **U-measure**
 - Idea: "More important GiUnits earlier" is better

$$U = \sum_{pos=1}^{|tt|} g(pos)D(pos)$$

$$D(pos) = 1 - \frac{pos}{L}$$

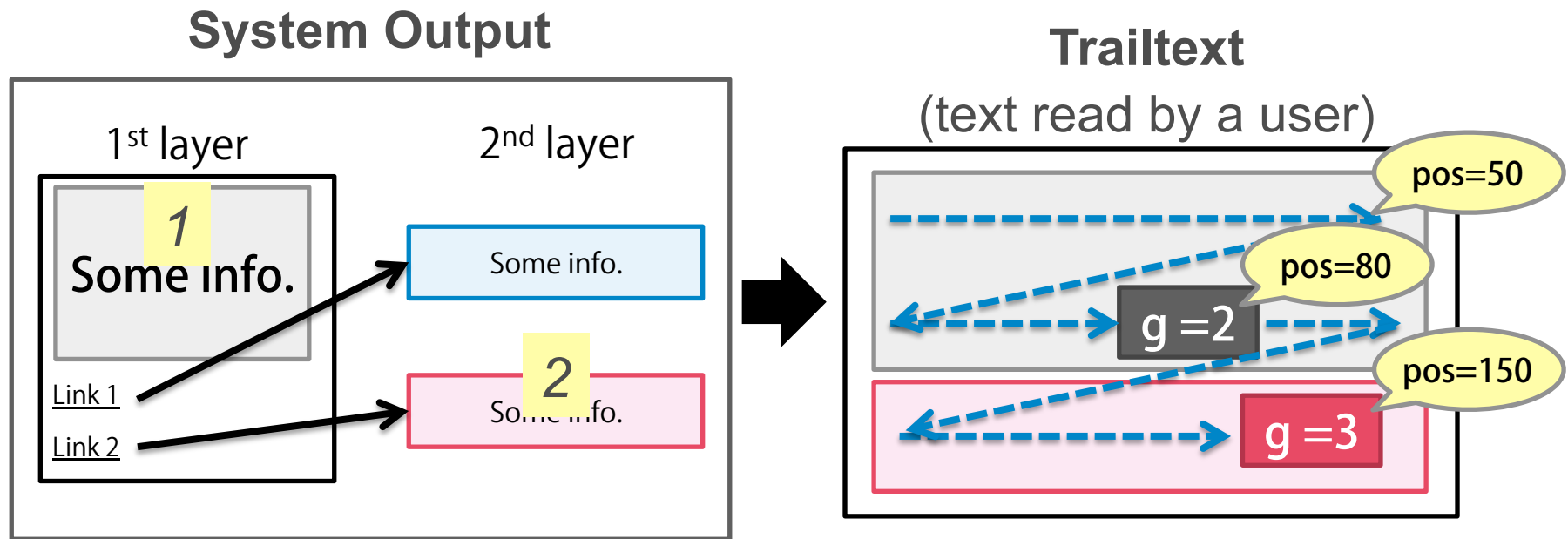
- pos : position in text
- $g(pos)$: importance of GiUnit at pos
- $D(pos)$: decay function
= how much effort is required to reach pos
- L : patience parameter (e.g. 280)



$$U = g(80)D(80) = 2\left(1 - \frac{80}{280}\right)$$

Illustration of U-measure Computation

- Suppose that a user reads the first layer and the 2nd second layer of system output

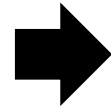
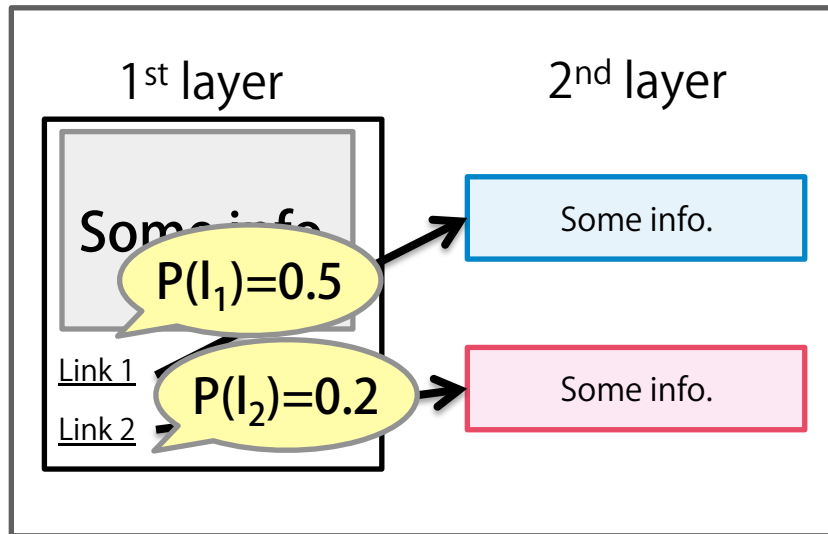


$$U = g(80)D(80) + g(150)D(150) = 2\left(1 - \frac{80}{280}\right) + 3\left(1 - \frac{150}{280}\right)$$

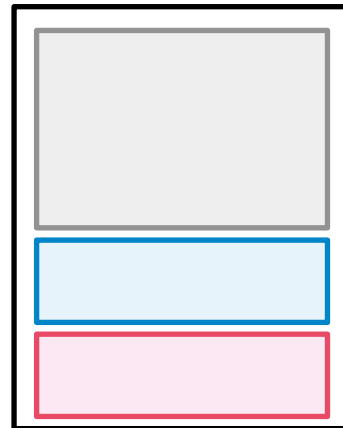
Evaluation Principles for iUnit Summarization (2/2)

- 2. Users probabilistically read a summary
 - When they reach link l_i , they click on the link with probability $P(l_i)$

System Output

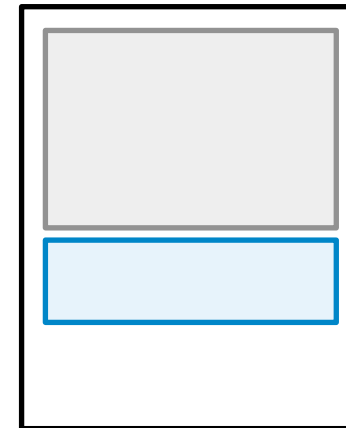


Trailtext1



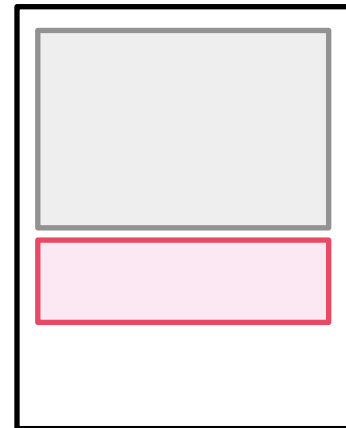
$$\begin{aligned} P(\text{trailtext1}) &= P(l_1)P(l_2) \\ &= 0.5 \cdot 0.2 = 0.1 \end{aligned}$$

Trailtext2



$$\begin{aligned} P(\text{trailtext2}) &= P(l_1)(1-P(l_2)) \\ &= 0.5 \cdot 0.8 = 0.4 \end{aligned}$$

Trailtext3



$$\begin{aligned} P(\text{trailtext3}) &= (1-P(l_1)) P(l_2) \\ &= 0.5 \cdot 0.2 = 0.1 \end{aligned}$$

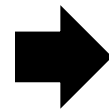
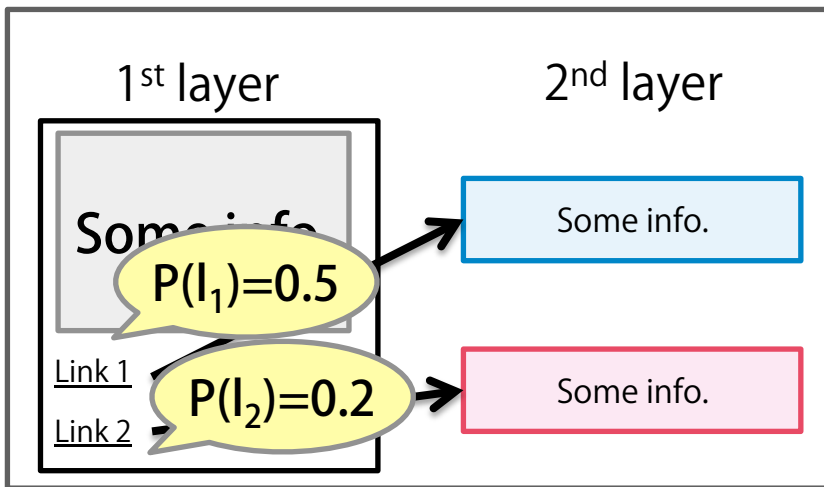
Evaluation Metric for iUnit Summarization

- **M-measure** = the expected utility of users who probabilistically click on links

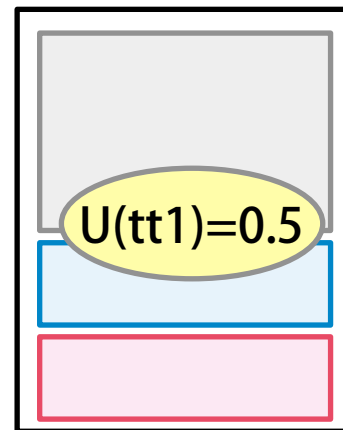
$$M = \sum_{tt \in T} P(tt)U(tt)$$

T : all possible trailtexts

System Output

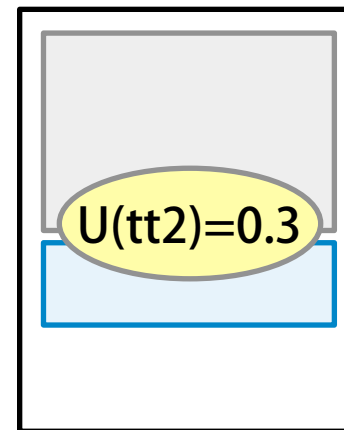


Trailtext1



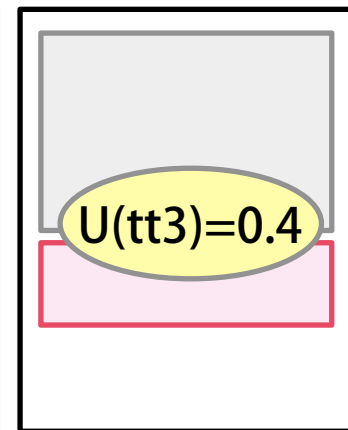
$$P(tt1) = 0.1$$

Trailtext2



$$P(tt2) = 0.4$$

Trailtext3

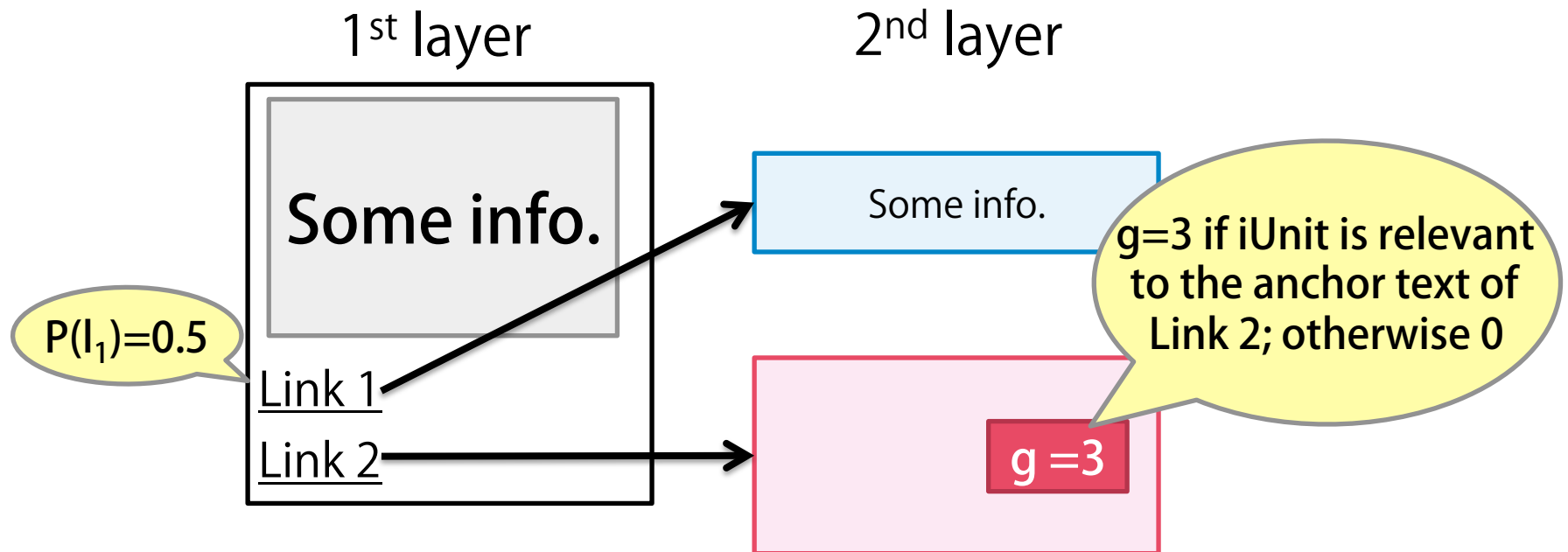


$$P(tt3) = 0.1$$

$$M = P(tt1)U(tt1) + P(tt2)U(tt2) + \dots = 0.1*0.5 + 0.4*0.3 + \dots$$

Click Probability and Importance of iUnits

- Click probability
 - Estimated by assessors' voting
- Importance of iUnits
 - $g = 0$ if the iUnit is irrelevant to the link



- **iUnit Retrieval**

- Estimating the importance of information pieces

- **iUnit Summarization**

- Help users navigate so that they can efficiently reach their desired information
 - Clustering iUnits
 - Creating meaningful links for clustered iUnits

Participating Teams

Team name	Organization/Method
KPNM	Hunan University of Science and Technology, China
	[Retrieval] Chain simple techniques based on statistical models and heuristic rules to extract significant text units
IISR	National Central University, Taiwan
	[Retrieval] Classify each query into eight query types and set the weights of the extraction methods accordingly
udel	University of Delaware, USA
	[Summarization] Simple re-ranking approach based on the cosine similarity between each iUnit and a dynamic "model" pseudo-document
NTOUA	National Taiwan Ocean University, Taiwan
	[Summarization] Grouping by longest leading substring.

- **iUnit Retrieval**

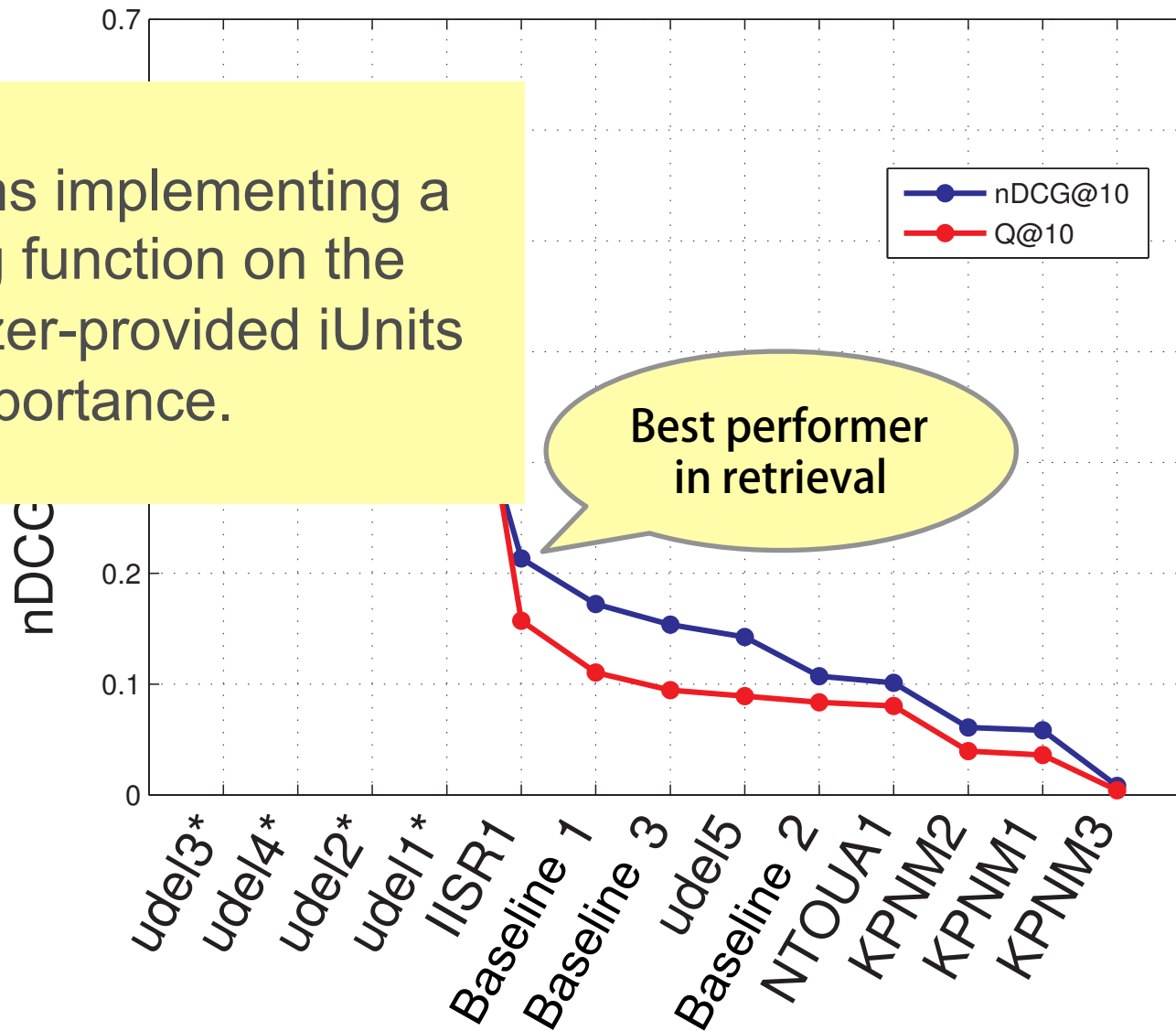
- Sentences in snippets of the search results
- In order of appearance in the results

- **iUnit Summarization**

- Output iUnits in descending order of the iUnit importance
- "Headers" (e.g. h1 and h2) used as links

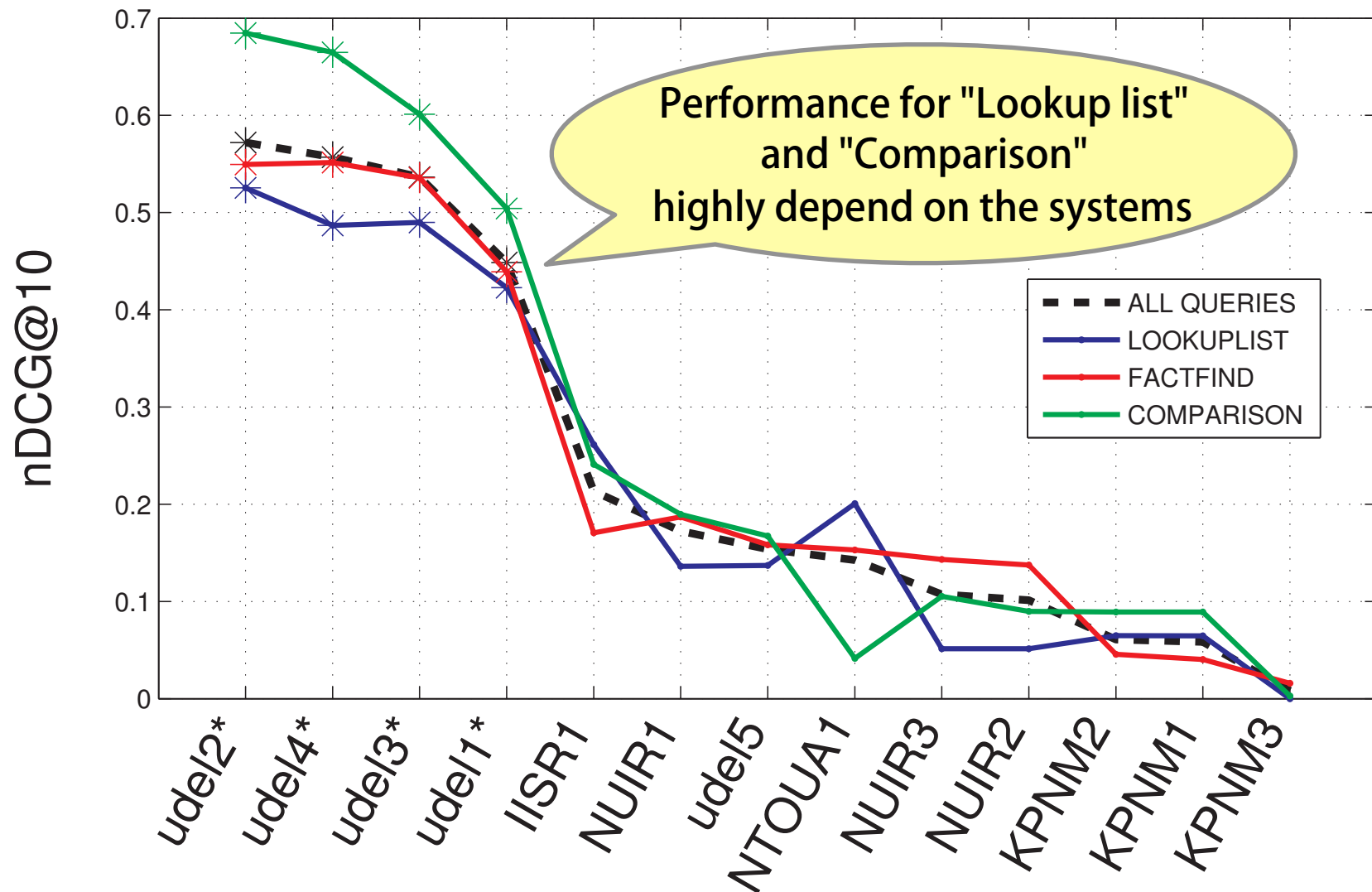
iUnit Retrieval Results

Systems implementing a ranking function on the organizer-provided iUnits and importance.



IISR achieved the highest performance

iUnit Retrieval Results per Query Category



LOOKUPLIST

FACTFIND

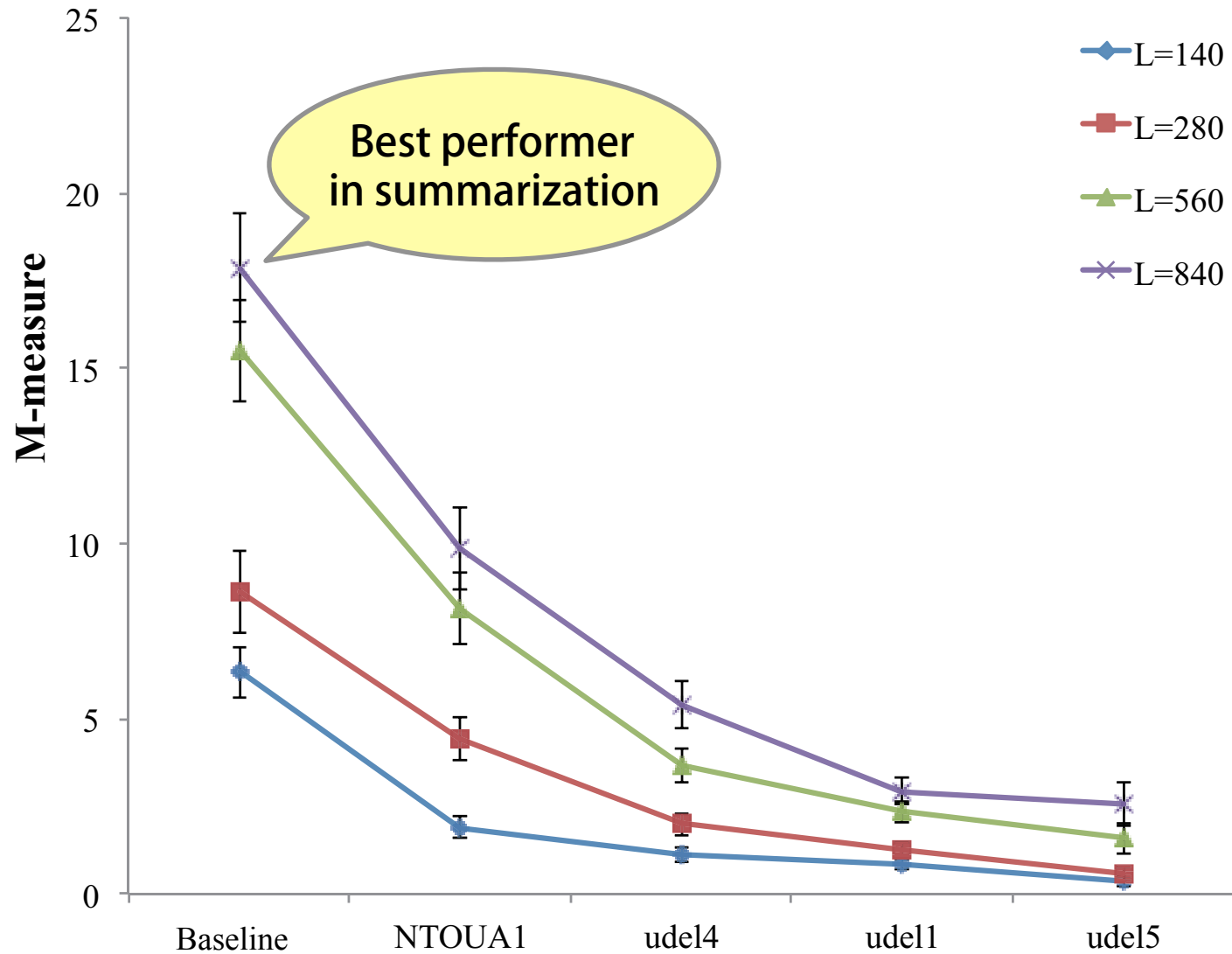
COMPARISON

data mining course online

obamacare penalty

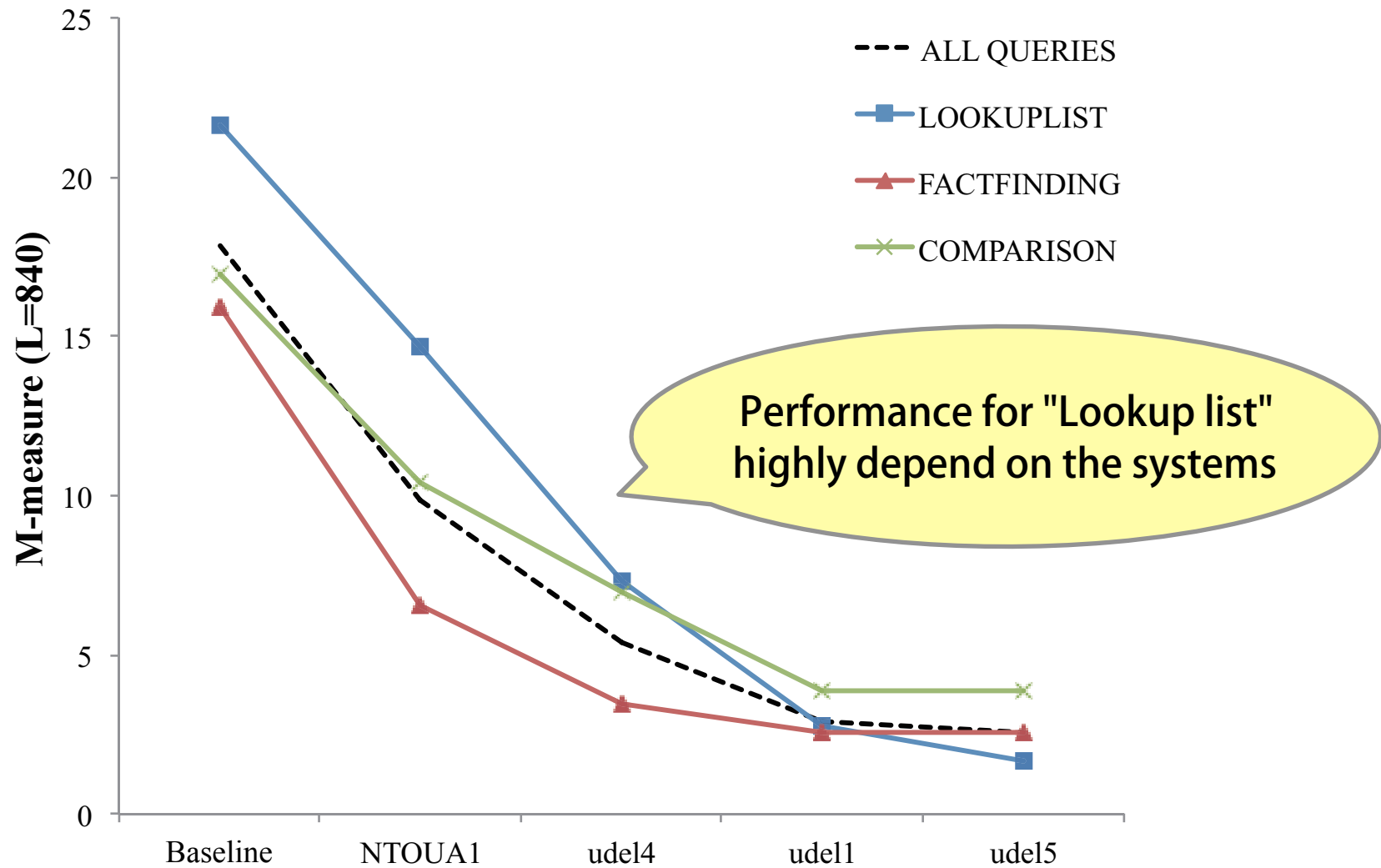
ivy bridge vs haswell

iUnit Summarization Results



Baseline achieved the highest performance possibly due to lack of details regarding the evaluation metric

iUnit Summarization Results per Query Category



- **MobileClick** task aims to achieve *information* retrieval systems that appropriately lay out information in two layers
- Evaluation was based on **iUnits** and **M-measure** that approximates utility of users who read the two-layered summary
- **IISR** achieved the highest performance in retrieval, while summarization was a difficult task to achieve high performance

More Details in MobileClick Session at Day-4 11:35 -

	Title	Organization/Method
1	Improving iUnit Retrieval with Query Classification and Multi-Aspect iUnit Scoring: The IISR System at NTCIR-11 MobileClick Task	National Central University, Taiwan
		[Retrieval] Classify each query into eight query types and set the weights of the extraction methods accordingly
2	Description of the NTOU MobileClick System at NTCIR-11	National Taiwan Ocean University, Taiwan
		[Summarization] Grouping by longest leading substring.
3	Udel @ NTCIR-11 MobileClick Track	University of Delaware, USA
		[Summarization] Simple re-ranking approach based on the cosine similarity between each iUnit and a dynamic "model" pseudo-document

- **DAY-3 (Dec. 11 (Thu))**
- **Location:**
Seminar Room 1904, 19F
- **Time:**
16:05 – 18:00

"Invisible Baseline" Problem

- In MobileClick-2, we will keep the basic task design, but **promise to help you continuously improve your systems**

The screenshot displays the NTCIR-12 MobileClick-2 website. At the top, the title "NTCIR-12 MobileClick-2" is prominently displayed, followed by the subtitle "Simple Task Description". A blue "SIGN UP" button is located in the top right corner. Below the title, the page is divided into two main sections: "iUnit Ranking Scores" on the left and "News" on the right. The "iUnit Ranking Scores" section features a table with four columns: Rank, Team ID, Score, and Time. The table lists the top three teams: KUIDL (Rank 1, Score 0.842), WUIAT (Rank 2, Score 0.802), and NIIAT (Rank 3, Score 0.798). The "News" section includes a "Tweets" subsection with a "Follow" button and two tweets from the NTCIR MobileClick account, dated December 1st.

NTCIR-12 MobileClick-2

Simple Task Description

[SIGN UP](#)


iUnit Ranking Scores


Rank	Team ID	Score	Time
1	KUIDL	0.842	Apr 1, 2015 13:00:01
2	WUIAT	0.802	Apr 1, 2015 09:00:01
3	NIIAT	0.798	Apr 1, 2015 06:00:02

News

Tweets

[Follow](#)

 **NTCIR MobileClick** @mobileclicktask 1 Dec
Using RoR for developing a real-time evaluation system

 **NTCIR MobileClick** @mobileclicktask 1 Dec
Developing an official Web site

Before the run submission deadline, our real time scoreboard will show the current performance of participant systems