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)

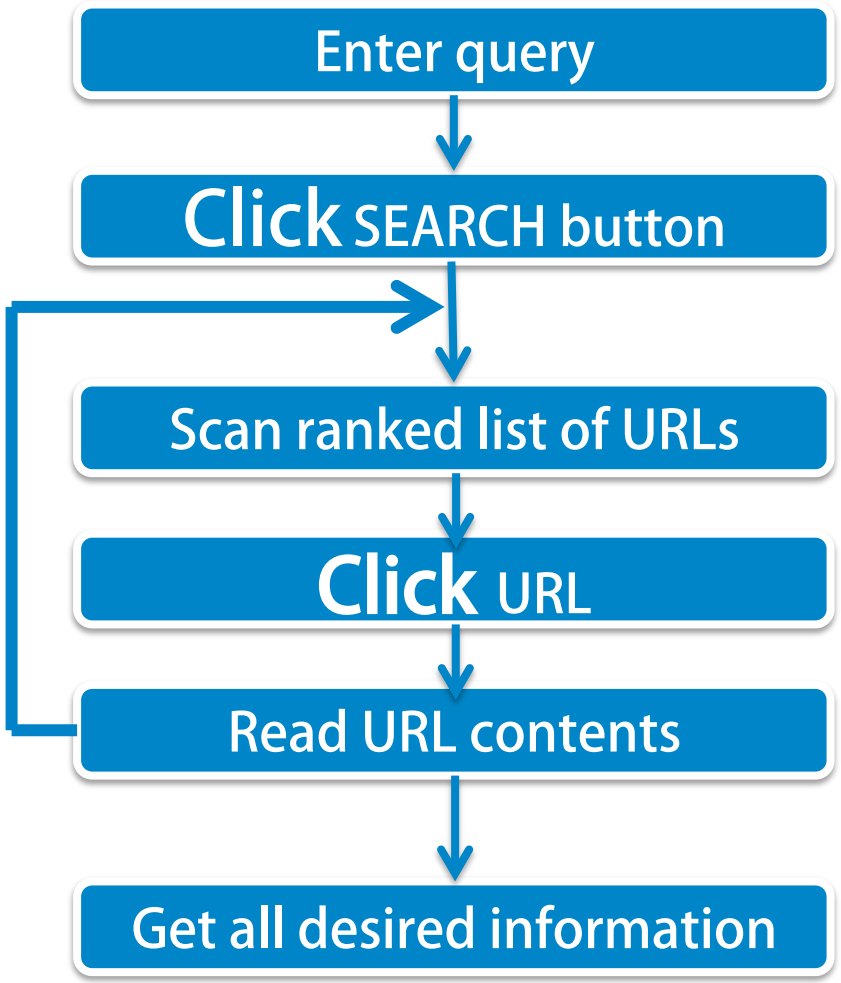





Suppose that …

- 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

Enter query

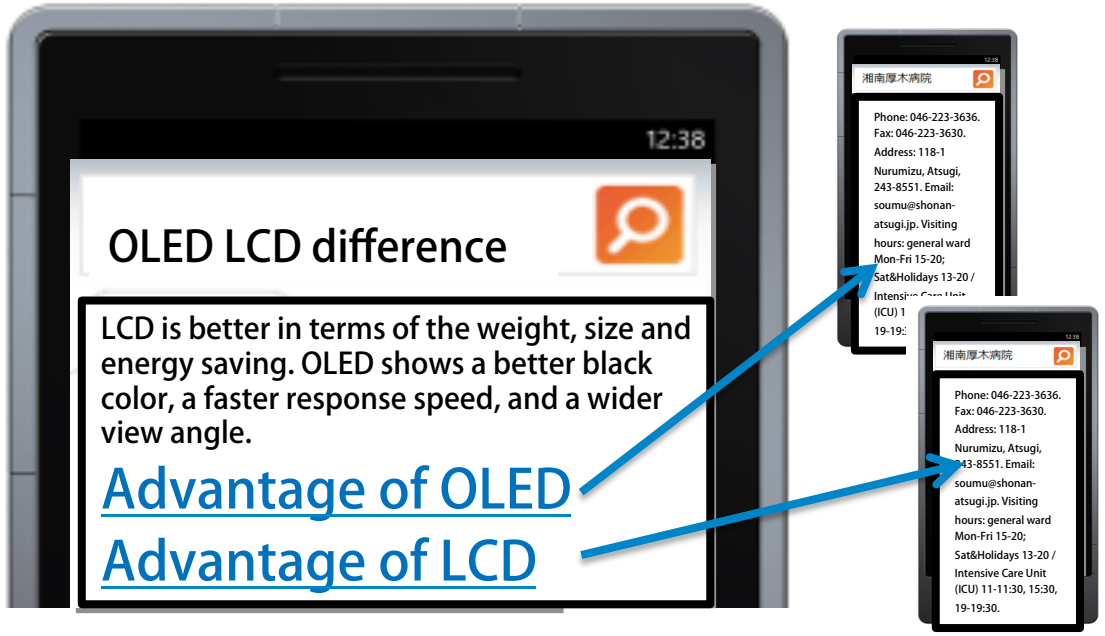
↓

Click SEARCH button

↓

Get all desired information

System output

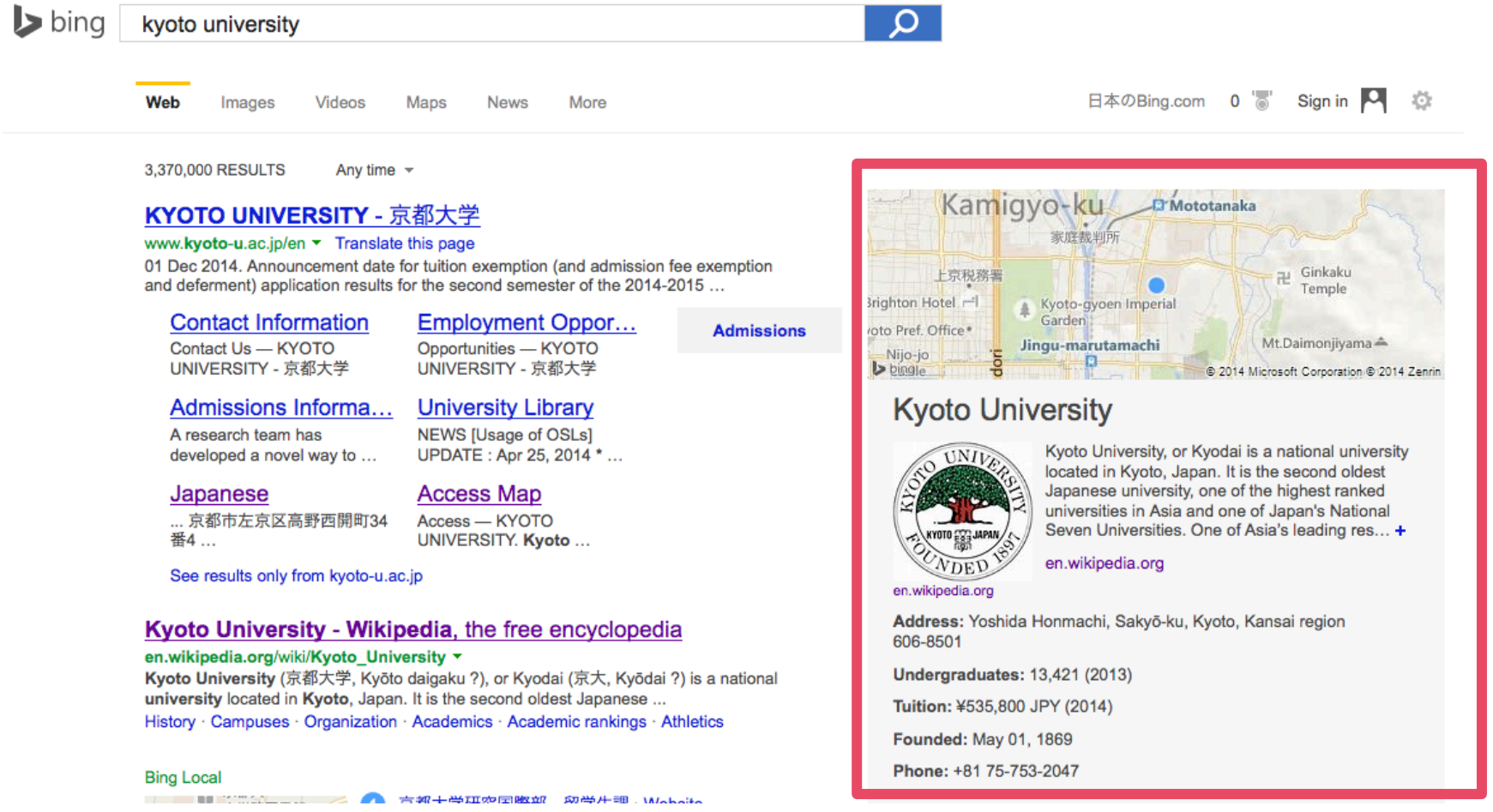


Task: Given a search query, return a two-layered textual output

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

Vision

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



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

Query

↓

iUnit Retrieval Subtask

↓

iUnits

↓

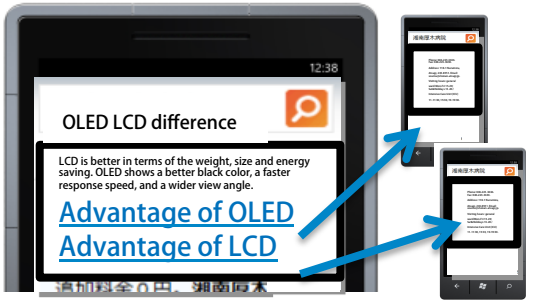
iUnit Summarization Subtask

↓

Two-layered Summary

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**

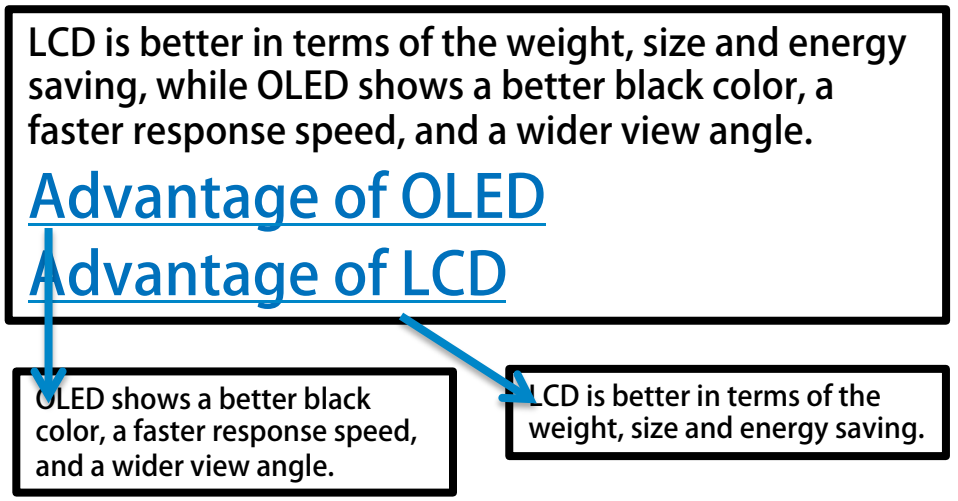
OLED LCD difference

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
…	…

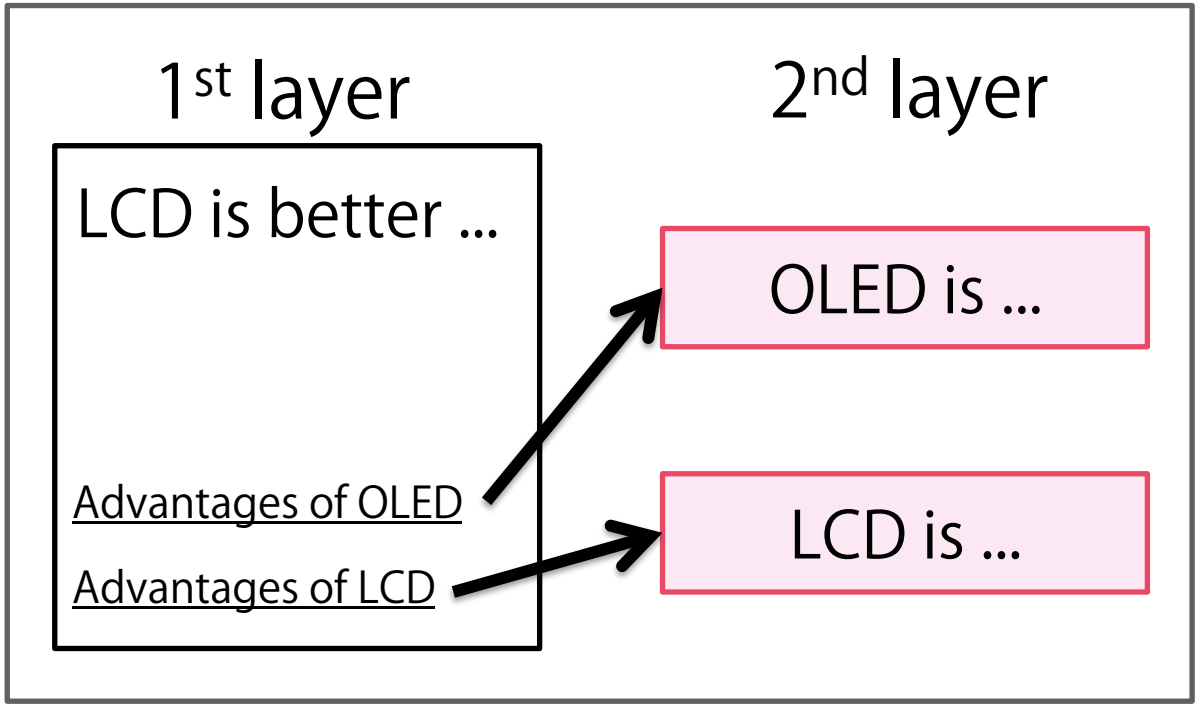
Output:

Two-layered textual output



Example of iUnit Summarization Run

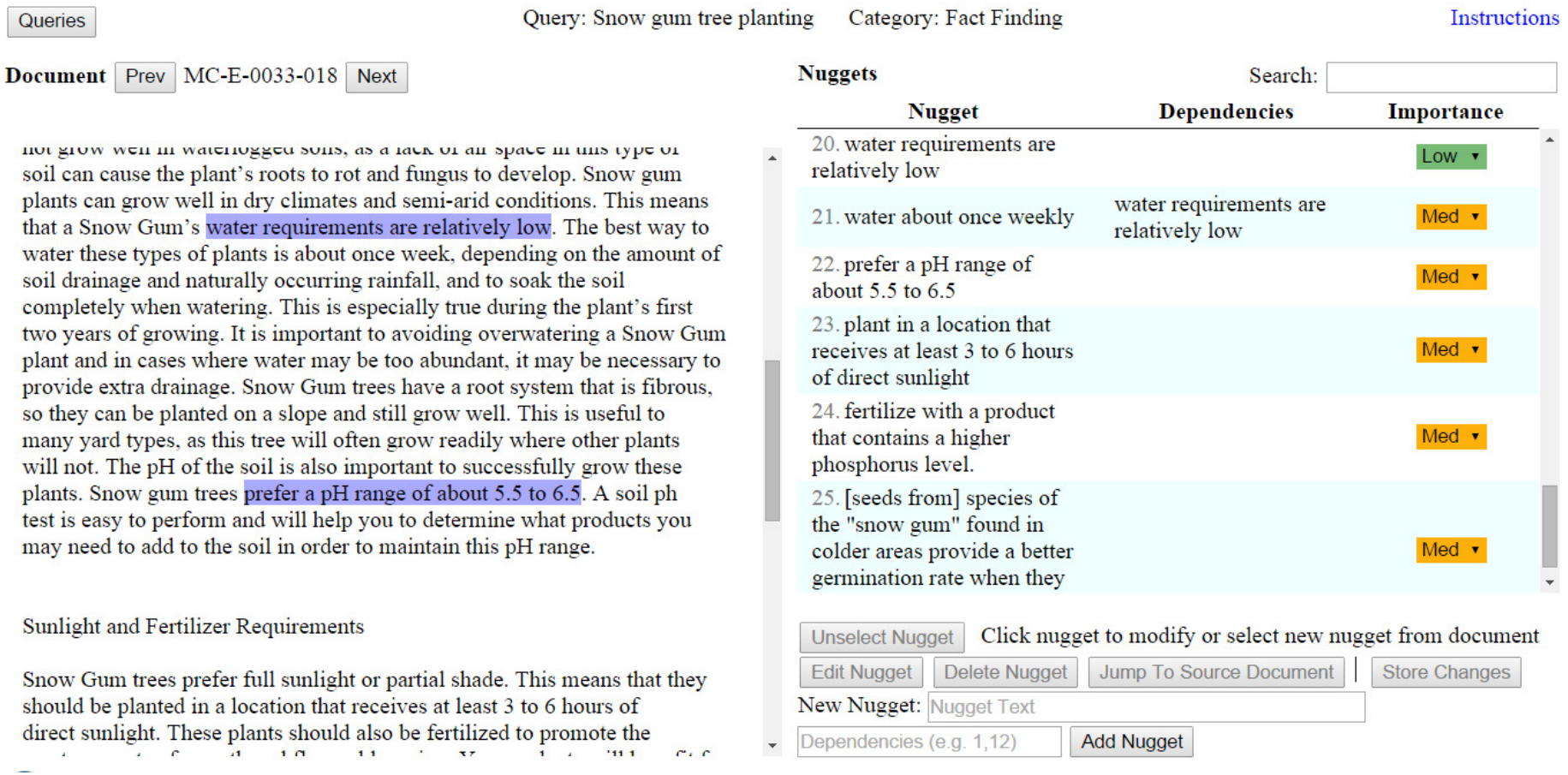
```
<result qid="MC-SAMPLE-E-0001">
  <firstlayer>
    LCD is better in terms of the weight, size and energy saving.
    OLED shows a better black color, a faster response speed, and a wider view
angle.
    <link id="1">Advantages of OLED</link>
    <link id="2">Advantages of LCD</link>
  </firstlayer>
  <secondlayer id="1">
    OLED is...
  </secondlayer>
  <secondlayer id="2">
    LCD is ....
  </secondlayer>
</result>
```



Evaluation Methodology

- **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)



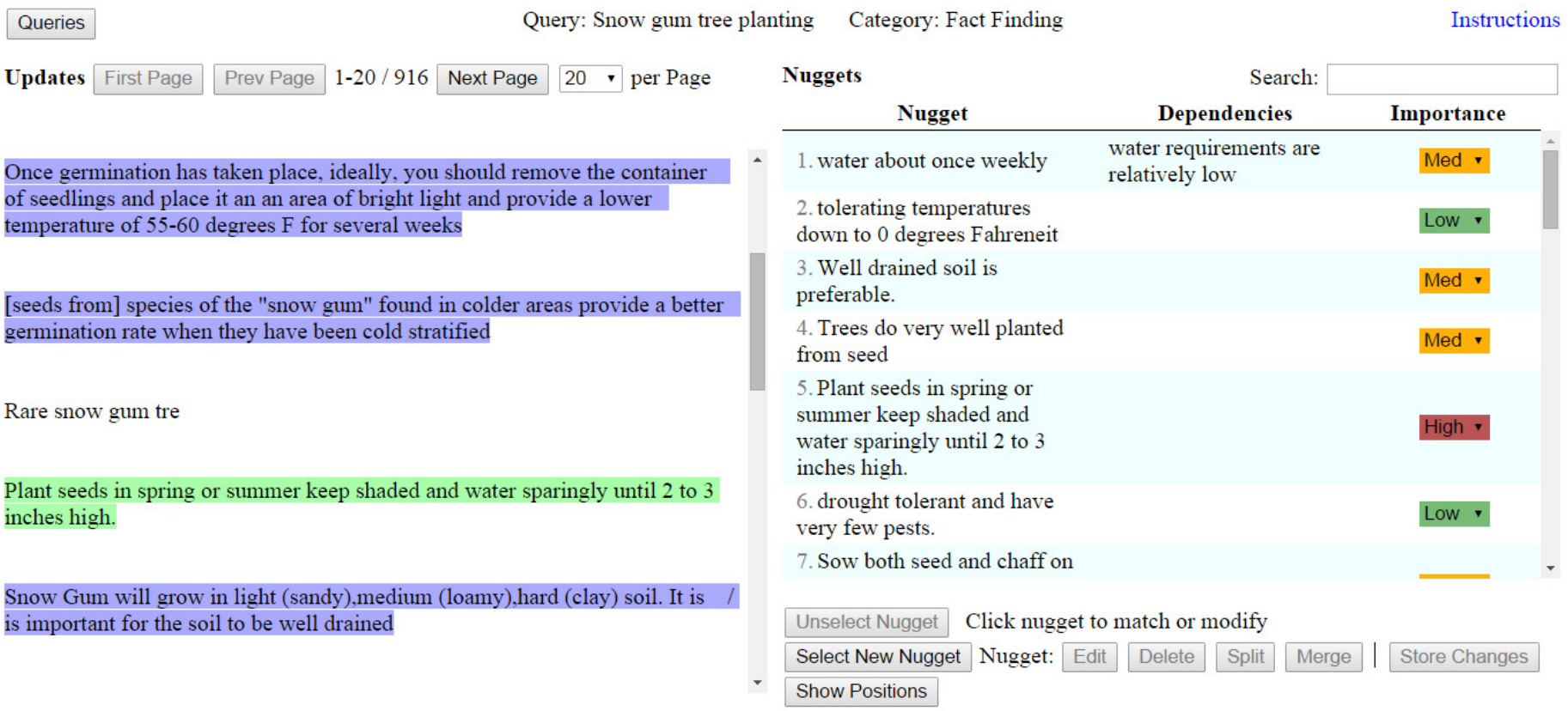
- Atomic and relevant information pieces were recorded as GiUnits with their importance

GiUnit Examples

- 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

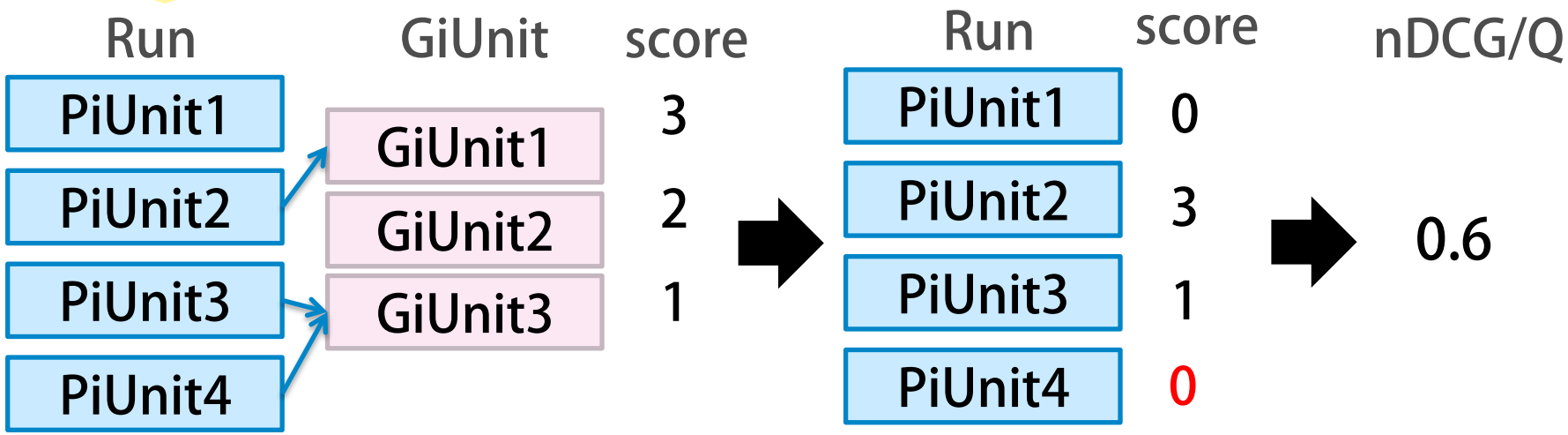
iUnit Matching



- 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

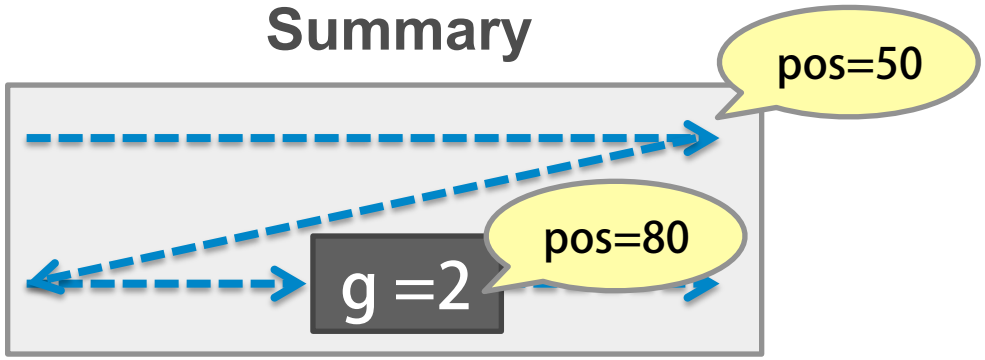
Evaluation Principles for iUnit Summarization (1/2)

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

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

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

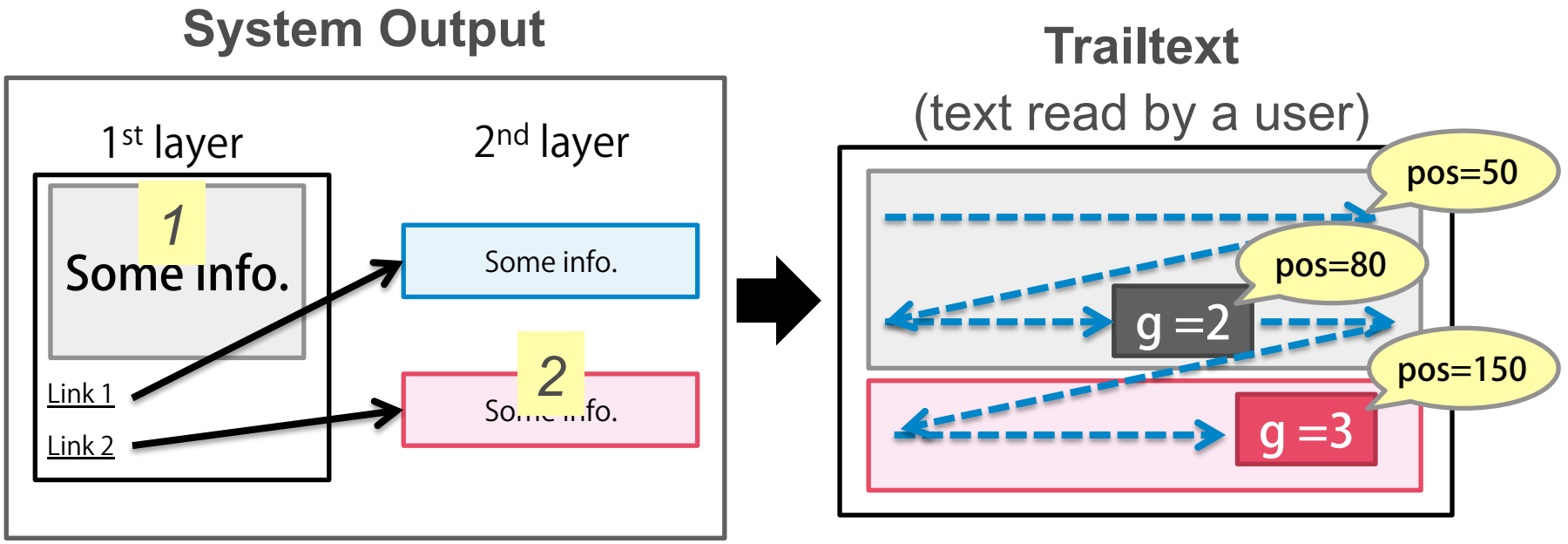
- *pos:* position in text
- *g*(*pos*): importance of GiUnit at *pos*
- *D*(*pos*): decay function
  = how much effort is required to reach *pos*
- *L:* patience parameter (e.g. 280)



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

Illustration of U-measure Computation

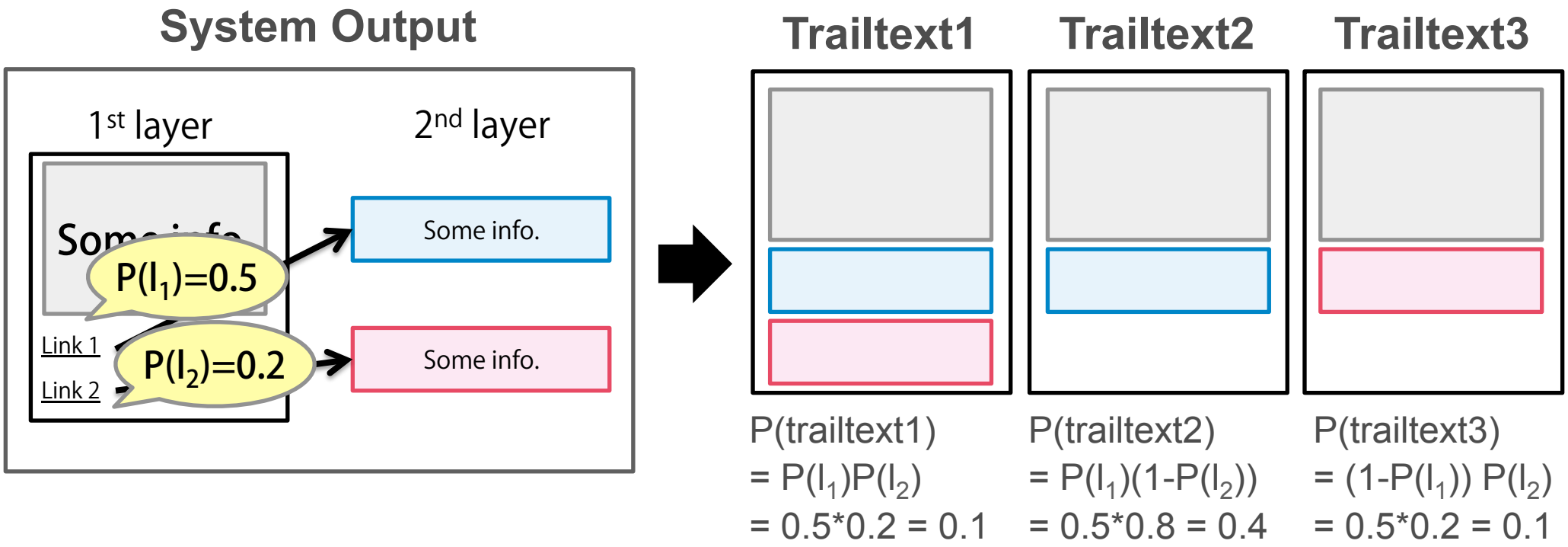
- Suppose that a user reads the first layer and the 2nd second layer of system output



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

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 $P(l_i)$

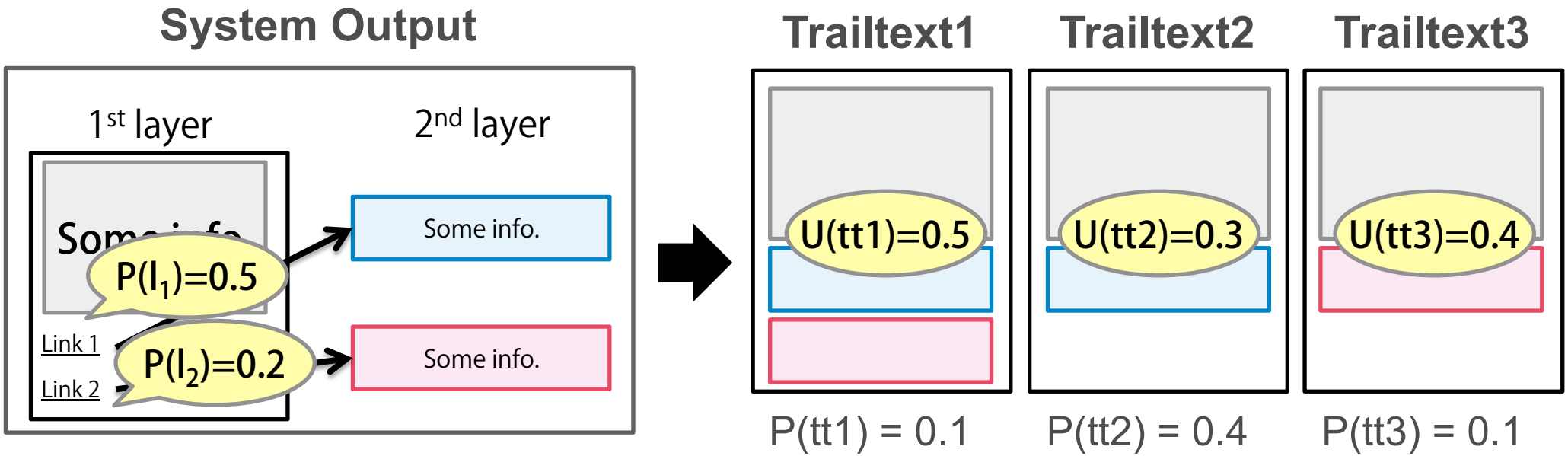


Evaluation Metric for iUnit Summarization

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

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

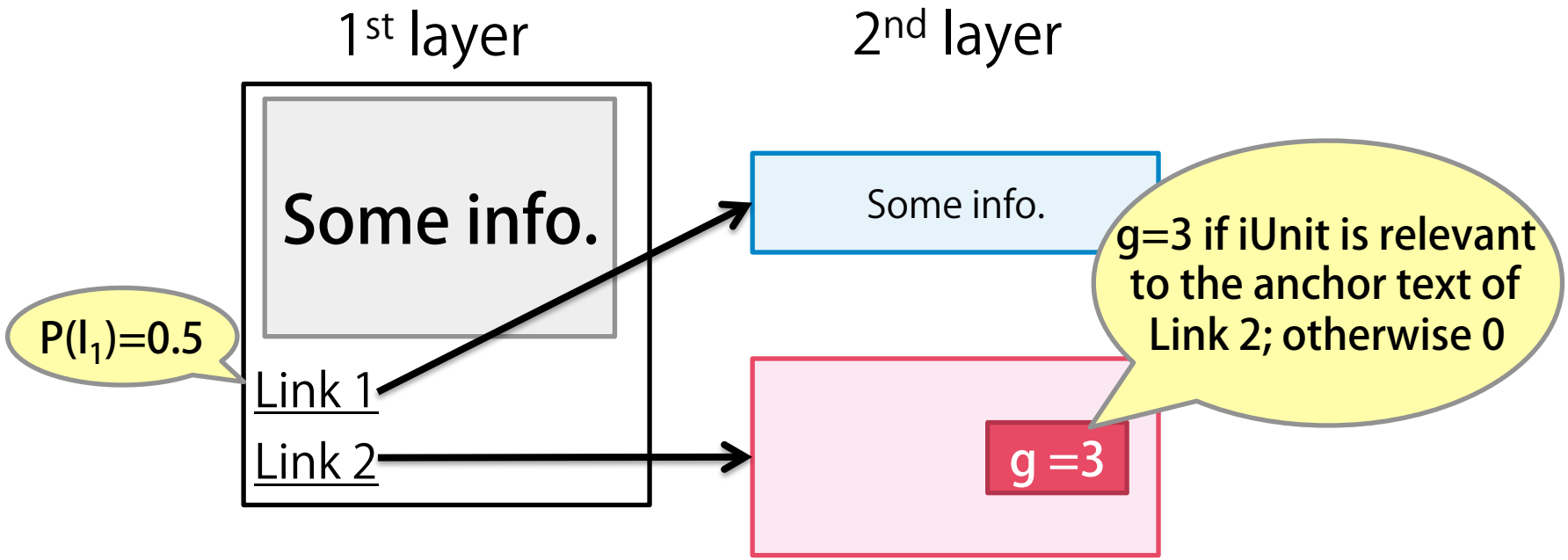
T: all possible trailtexts



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

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



Challenges in MobileClick Task

- **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.

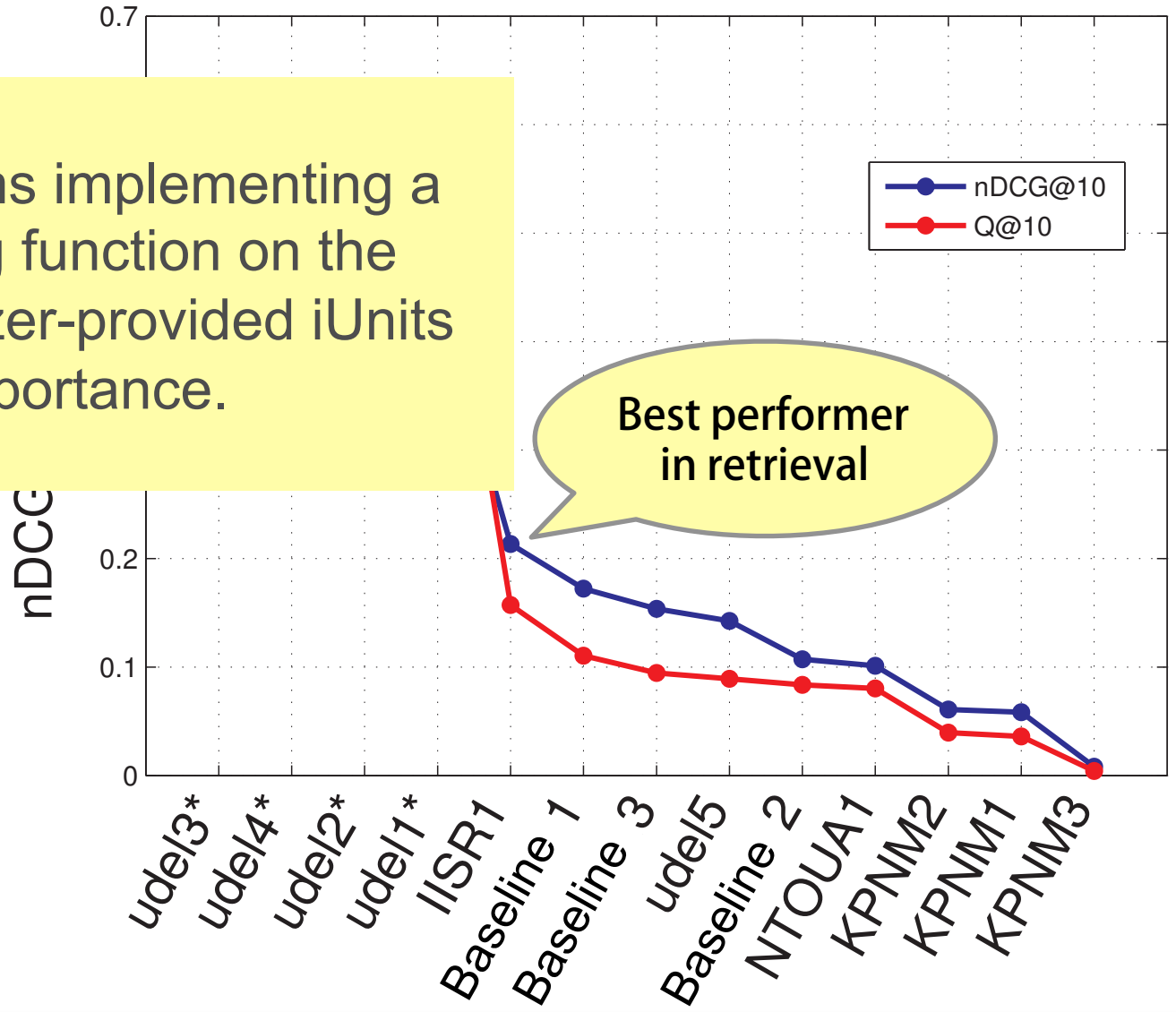
Baselines

- **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

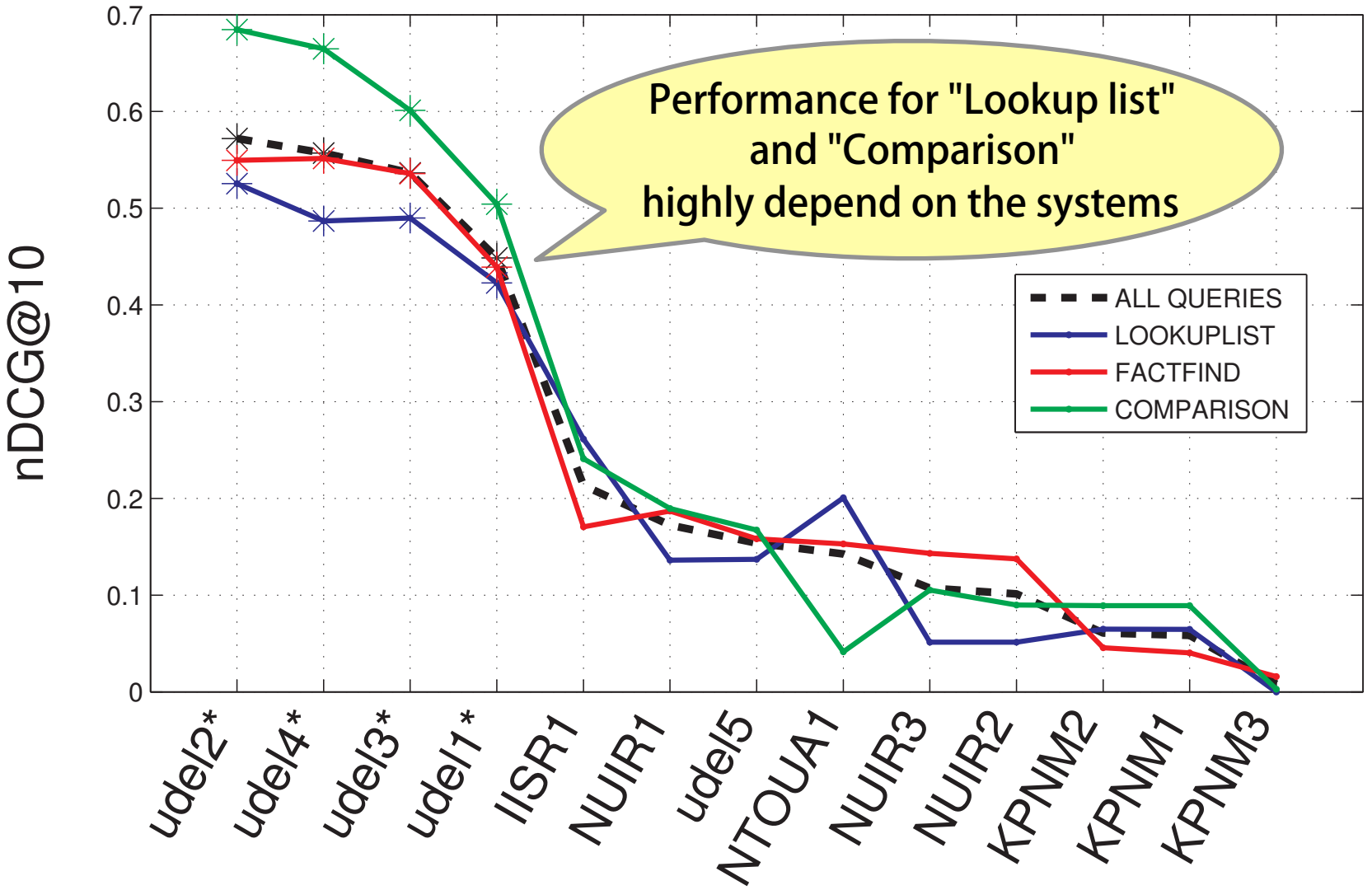
iUnit Retrieval Results

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



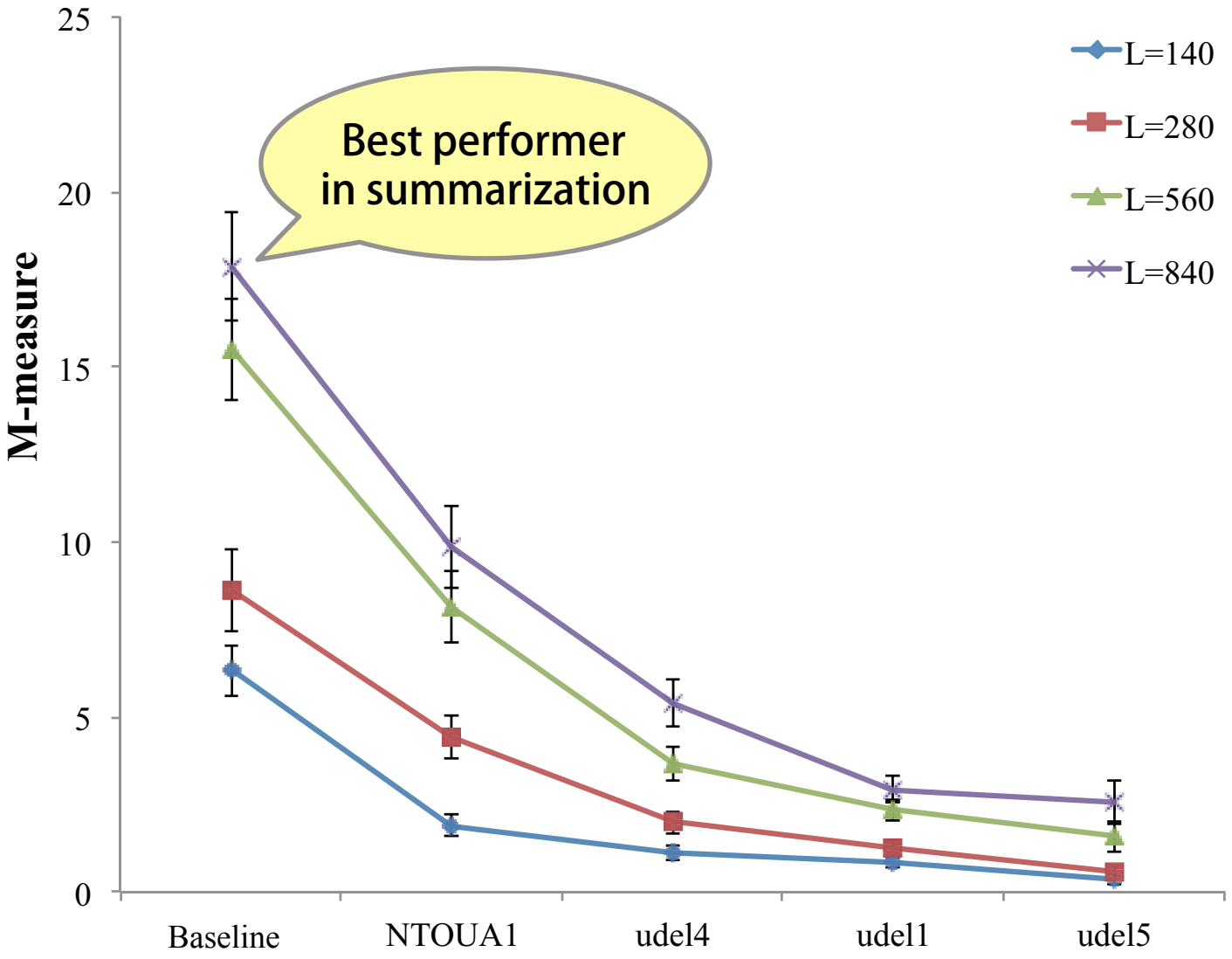
IISR achieved the highest performance

iUnit Retrieval Results per Query Category



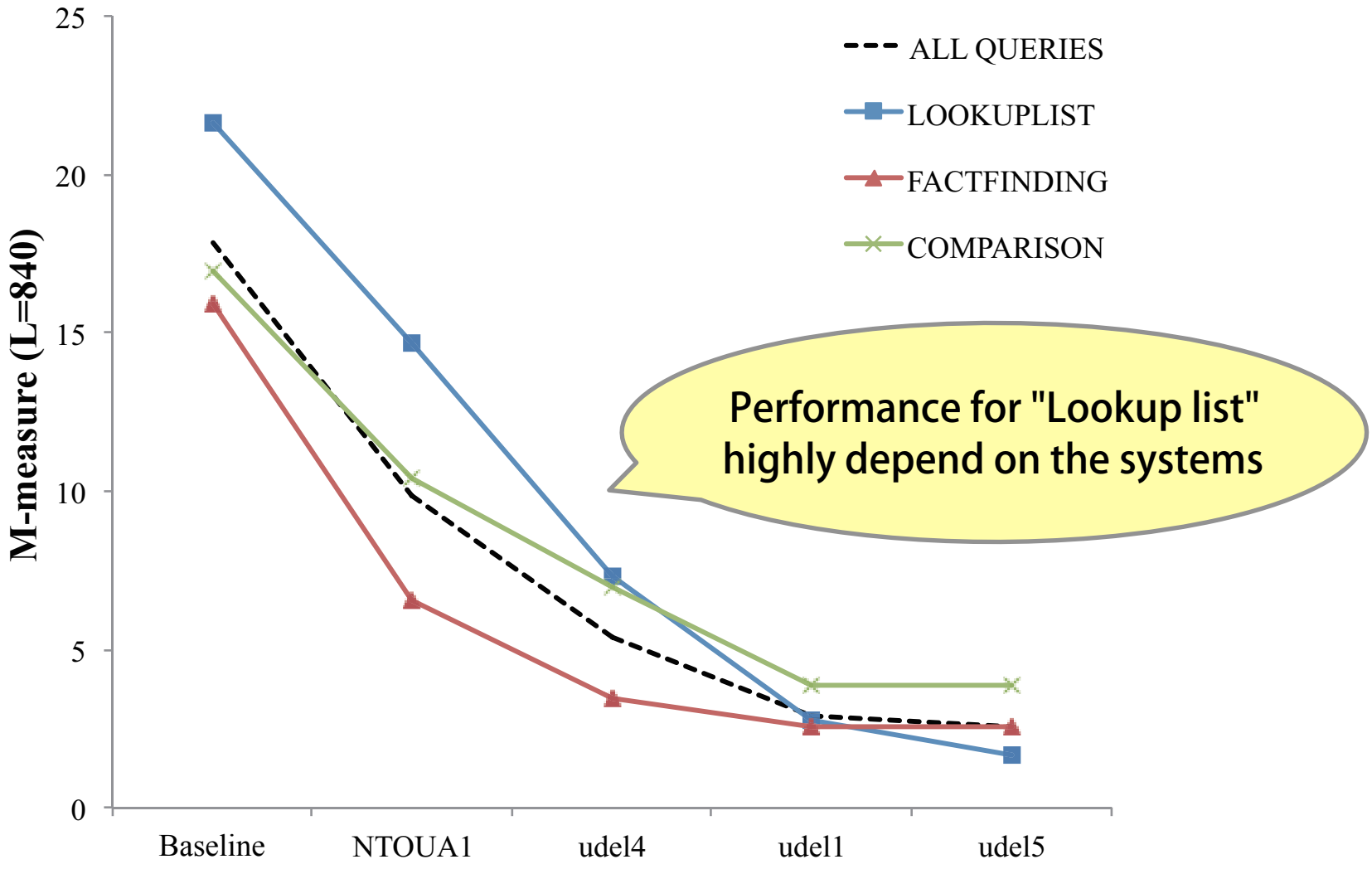
LOOKUPLIST	FACTFIND	COMPARISON
data mining course online	obamacare penalty	ivy bridge vs haswell

iUnit Summarization Results



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

iUnit Summarization Results per Query Category



Summary

- 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

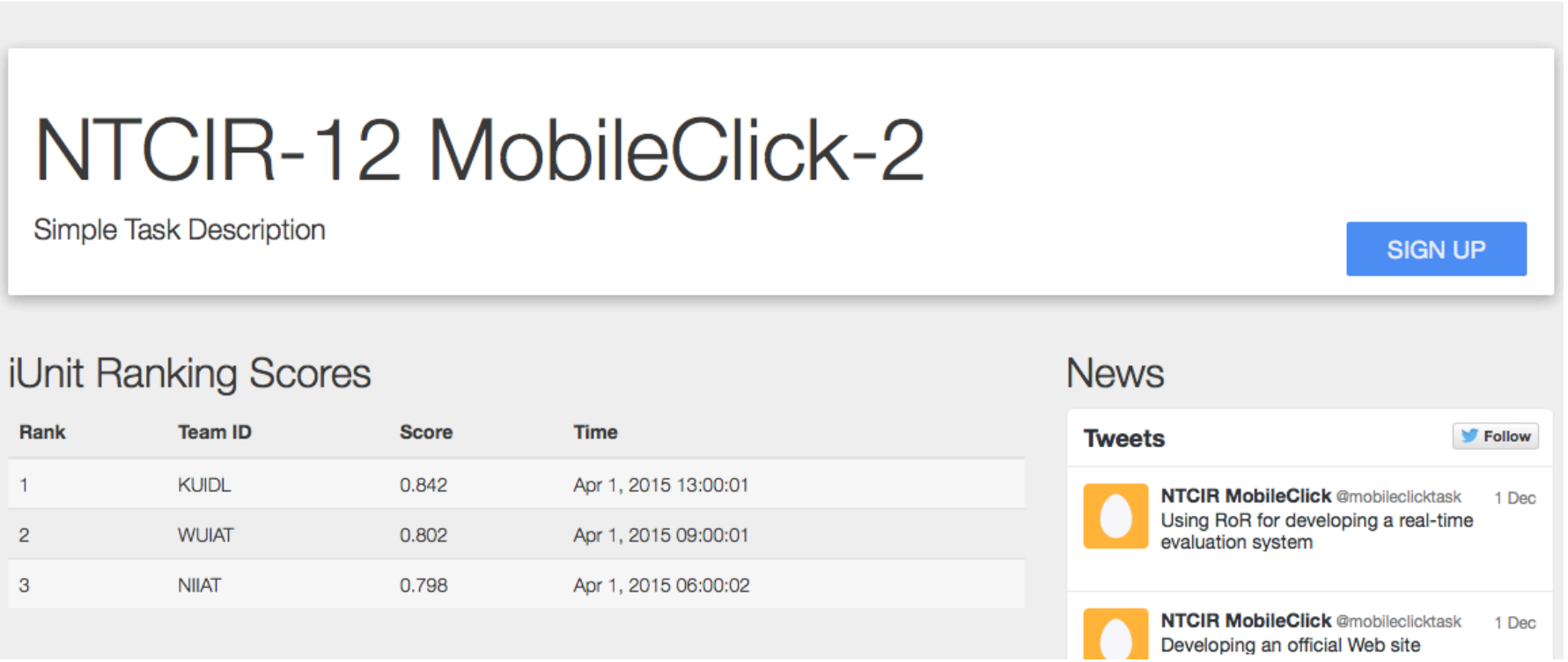
Round Table Session

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

"Invisible Baseline" Problem

MobileClick-2

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



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