

# Learning to Rank Definitions to Generate Quizzes for Interactive Information Presentation



Ryuichiro Higashinaka, Kohji Dohsaka, Hideki Isozaki  
NTT Communication Science Laboratories, NTT Corporation, Japan

## Overview

- We propose ranking biographical facts (definitions) to automatically generate “Who is this?” quizzes.
- Such ranking would realize an information presentation dialogue (Fig.1) in which users can be highly motivated to receive information, which could lead to better understanding by users.
- Supervised learning approach achieves better ranking than those using co-occurrence statistics. (Tab.1)
- Content of a definition (e.g., birth, occupation) must be taken into account for better ranking. (Tab.2)
- Quiz-style information presentation is actually effective for learning. (to be presented at Interspeech 2007)

## Figure1 A procedure for realizing a quiz-style information presentation dialogue

### Entry of Natsume Soseki in an encyclopedia

**Natsume Soseki:** Novelist and scholar of British literature. Real name: Kinnosuke. Born in Ushigome, Edo. Graduated from the University of Tokyo. Master of early-modern literature along with Mori Ogai. After the success of “I Am a Cat”, quit all teaching jobs and joined Asahi Shimbun. Published masterpieces in Asahi Shimbun. Familiar with Haiku, Chinese poetry, and calligraphy. Works include “Botchan”, “Sanshiro”, etc.

### Quiz-style ordering of biographical facts

1. Graduated from the University of Tokyo.
2. Born in Ushigome, Edo.
3. Novelist and scholar of British literature.
4. Familiar with Haiku, Chinese poetry, and calligraphy.
5. Published masterpieces in Asahi Shimbun.
6. Real name: Kinnosuke.
7. Master of early-modern literature along with Mori Ogai.
8. After the success of “I Am a Cat”, quit all teaching jobs and joined Asahi Shimbun.
9. Works include “Botchan”, “Sanshiro”, etc.

### Quiz-style information presentation dialogue

S1 Who is this? First hint: Graduated from the University of Tokyo.  
U1 Yoshida Shigeru?  
S2 No, not even close! Second hint: Born in Ushigome, Edo.  
U2 I don't know.  
S3 OK. Third hint: Novelist and scholar of British literature.  
U3 Murakami Haruki?  
S4 Close! Fourth hint: Familiar with Haiku, Chinese poetry, and calligraphy.  
U4 Mori Ogai?  
S5 Very close! Fifth hint: Published masterpieces in Asahi Shimbun.  
U5 Natsume Soseki?  
S6 That's right!

(S stands for system utterance and U for user utterance in the dialogue.)

## Approach: Training a ranking model by supervised learning

### Reference Ranking Data

- 11 annotators ranked definitions of 150 people
- Mean Kendall's coefficients of concordance: 0.76

### Features (4,991 features in total)

- **R-related features** (co-occurrence statistics of content words)
- **Positional features** (positions of content words in *Wikipedia*)
- **Bag-of-Words features**
- **Semantic category features**

### Learning Algorithm

- Ranking SVM (Joachims, 2002)

Rank	Description	Ranking Error
1	<b>Proposed ranking model</b>	<b>0.185</b>
2	Wikipedia-Sentence-PMI-max	0.299
3	Wikipedia-Section-PMI-max	0.309
4	Wikipedia-Document-PMI-max	0.312
5	Mainichi-Sentence-PMI-max	0.318
6	Mainichi-Document-PMI-max	0.325

Table1. Performance of the proposed ranking model and 5 best performing baselines (cf. Ranking Error: rate of incorrectly ordered pairs)

Rank	Feature Name	Weight
1	Wikipedia-Sentence-PMI-max	0.723
2	SemCat:33 (others/someone)	-0.559
3	SemCat:186 (creator)	0.485
4	Bag-of-Words:bakufu (feudal government)	0.451
5	SemCat:163 (sovereign/ruler/monarch)	0.422
6	Wikipedia-Document-PMI-max	0.409
7	SemCat:2391 (birth)	-0.404
8	Wikipedia-Section-PMI-max	0.402
9	SemCat:2595 (unit; e.g., numeral classifier)	0.374
10	SemCat:2606 (plural; e.g., plural form)	-0.368

Table2. 10 most salient features in the trained ranking model