

# InteractiveMediaMINE at the NTCIR-11 IMine Search Task

Shohei MINE, Takuma MATSUMOTO, Tomofumi YOSHIDA, Takuya SHINOHARA, Daisuke KITAYAMA

Graduate School of Engineering, Kogakuin University

## Background

TaskMine aims to explore the methods of automatically finding subtasks of a given task

※ We use Yahoo! Chiebukuro(web-based Q&A service) as our system resource

- In general, Q&A services aim to collect answers that solve the user's problems
- We expect Yahoo! Chiebukuro to be useful as our system resource for mining tasks

## Step 1. Extending Query

1. Executing a morphological analysis for the inputted query by user, and extracts nouns and verbs from the query
2. Joining the extracted nouns, a single-byte blank, and verbs to make a sentence
3. Adding “方法(hoho)” at the end of the query

※方法 is a word that means method or way in Japanese, and it is effective in retrieving pages that include questions about methods used for solving certain problems

## Step 2. Retrieving

1. Using Yahoo! Chiebukuro with the extended search query
2. Retrieving the top 10 pages of the search results

## Step 3. Extracting Tasks

1. Performing dependency parsing for all the extracted sentences in order to extract a pattern  
Extracted verb end of the sentence are converted into its dictionary form
  2. Extracting chunks that contain the following pattern:  
Chunks that depend on noun(a) + noun(b)+ “を(wo)”(c)  
+ the chunk that includes certain verb(d) + chunks that depend on the verb(e)
  3. Extracting sentences as tasks
- ※ Considering “を(wo)” is a Japanese particle pointing to a direct object, it is useful to extract tasks that can solve the user's problem

## Step 4. Ranking

- Based on the supposition that the words that appeared frequently are important in solving problems
- We define the evaluation scores as follows:

$$Score(t_i) = \sum_{noun \in nouns_{t_i}} \sum_{ans \in A} NounFreq(noun, ans) + \sum_{task \in T} VerbFreq(verb_{t_i}, task)$$

nouns	Frequency	verbs	frequency
米	4	加える	2
水	4	止める	1
鍋	3	入れる	2
火	3		
方	1		
ザル	1		
分量	1		
中火	1		
沸騰	1		
キープ	1		
...	...		
火力	1		

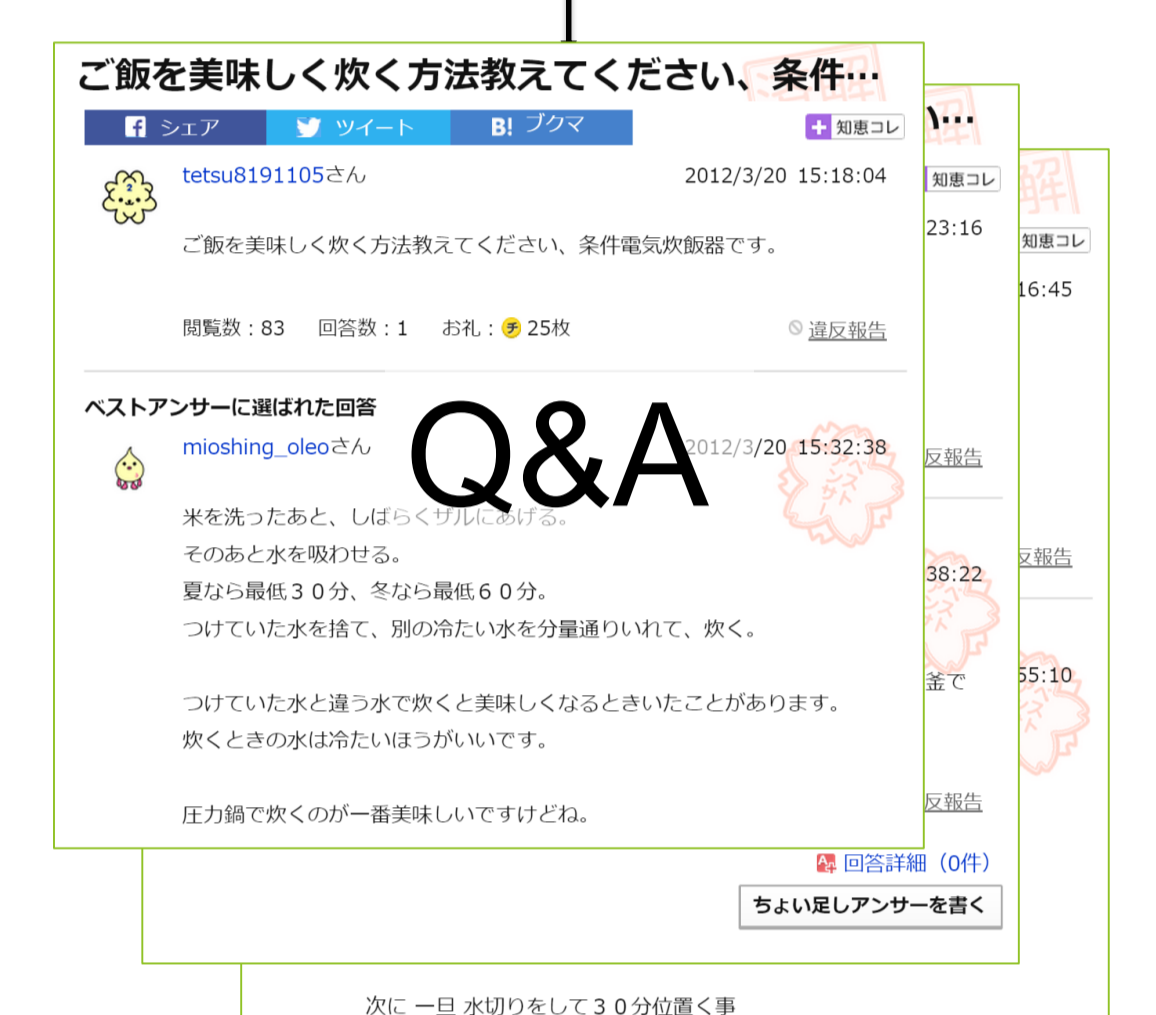
Query: “ご飯を炊く”※

※I want to cook the rice

“ご飯”, “炊く”

ご飯 炊く

ご飯 炊く 方法



炊き方は、洗った米を  
(a) (b) (c)  
ザルにあげて水を切って30分したら、分量の水を加える  
(d) (e)

お米2合を普通にといで鍋に入れて水を  
2カッププラス大さじ2杯入れる  
Tasks

Rank	Candidate tasks	NounFreq	VerbFreq	Score
1	鍋を中火にかけて沸騰すれば1~2分間キープして直ぐに火を最低限まで絞って15分間、最後に一瞬強火にして火を止める	17	1	18
2	炊き方は、洗った米を“ザルにあげて水を切って”30分したら、分量の水を加える	14	2	16
3	言われるように、鍋に米を入れて指の第一関節や手首までお水を入れる	14	2	16
4	お米2合を普通にといで鍋に入れて水を2カッププラス大さじ2杯入れる	12	2	14
5	米を炊く時は、基本的に吸水させたあとに、最大火力になるんですが、火を止める	8	2	10

## Experiment

- We mined tasks for the query set of the TaskMine subtask
- Dataset is 50 tasks

	nDCG@1	nDCG@5	nDCG@10	nDCG@50
InteractiveMediaMINE	0.323	0.330	0.320	0.289
Organizer	0.013	0.040	0.053	0.096

## Results

nDCG is computed for a variety of cutoff thresholds k(=1,5,10,50)

→InteractiveMediaMINE perform better than the organizer's baseline method among all the runs

Our system makes it easy to collect answers accurately for ordinary questions

TM-019, 歯周病を治療する, has one of the highest scores for all metrics, but extracted task is long sentence as table below

※ “I want to treat periodontal disease”, in English

The accuracy of the extracted tasks decreases for queries that are not the type of questions that ordinary users ask

TM-023, レーザーカッターを使う, has one of the lowest scores for all metrics

※ “I want to use the laser cutter”, in English

→Yahoo! Chiebukuro is a general Q & A service that is not specialized in any specific field