

# The KLE's Subtopic Mining System for the NTCIR-11 IMine Task

Se-Jong Kim, Jaehun Shin, Jong-Hyeok Lee  
 {sejong, rave0206, jhlee}@postech.ac.kr

## INTRODUCTION

### Key Ideas

- Reflect the characteristics of resources in terms of relevance, popularity & diversity
- Generate first-level subtopics using keywords in second-level subtopics
- Keep the quality of second-level subtopics

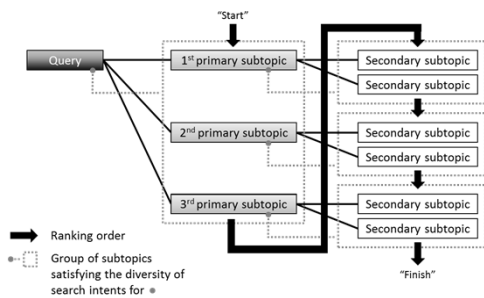
## SECOND-LEVEL SUBTOPIC MINING

### Second-level Subtopic Extraction

- Use simple patterns based on noun phrases & partial queries
- P1: ((ADJ)?(NN)+ (non-NN))?(query)((non-NN)\*(ADJ)?(NN)+)?  
 P2: ((ADJ)?(NN)+ (non-NN))?(q<sub>left</sub>)(word)(q<sub>right</sub>)((non-NN)\*(ADJ)?(NN)+)?  
 P3: (q<sub>right</sub>)(non-NN)\*(ADJ)?(NN)+  
 P4: (ADJ)?(NN)+ (non-NN)\*(q<sub>left</sub>)

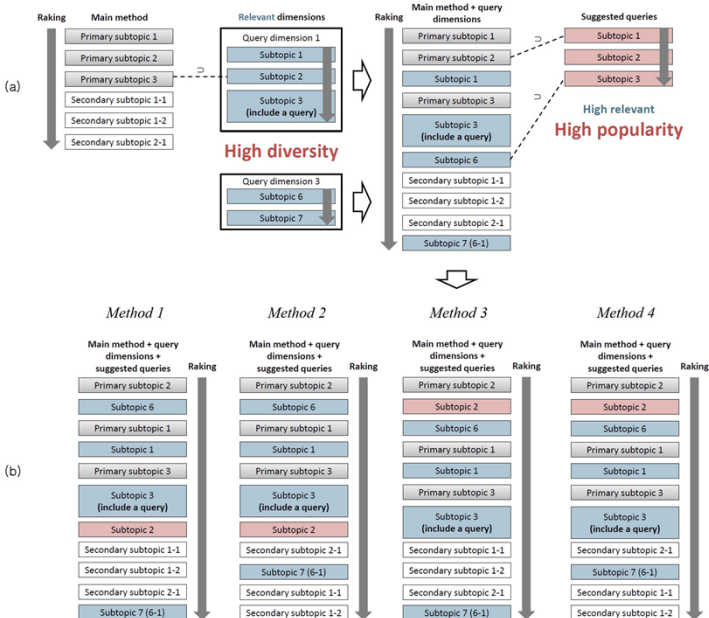
### Second-level Subtopic Ranking

- Construct a three-level hierarchical structure of subtopics
- Maintain a balance between popularity & diversity



### Result Combining and Re-ranking

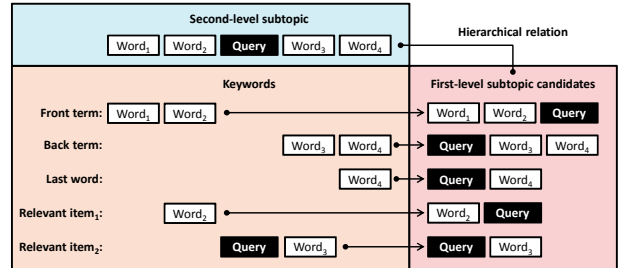
- Combine & re-rank the ranked lists of subtopics using the provided resources based on following assumptions
- suggested query = subtopic with **high popularity**  
 relevant query dimension = subtopic set with **high diversity**



## FIRST-LEVEL SUBTOPIC MINING

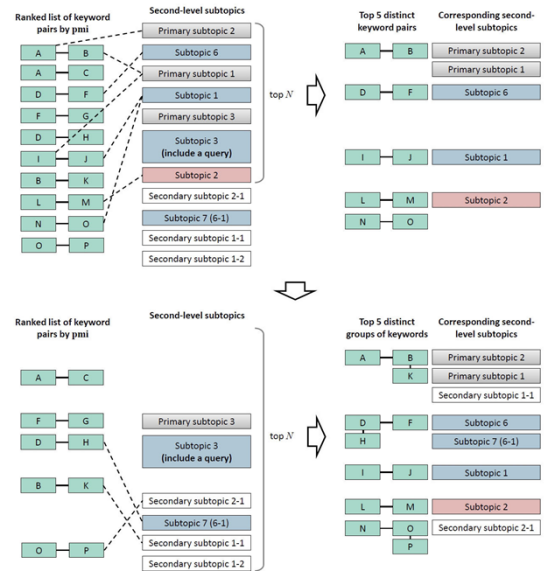
### First-level Subtopic Generation

- Select keywords & attach them to appropriate positions



### Result Grouping and Ranking

- Use the Pointwise Mutual Information of keywords



## EVALUATION RESULTS

### Results on unclear topics for English

Run	Hscore	Fscore	Sscore	H-measure
KLE-S-E-3A	0.1291	0.6539	0.7317	0.0980
KLE-S-E-4A	0.1260	0.6511	0.7294	0.0938
KLE-S-E-2A	0.1200	0.5698	0.7342	0.0893
KLE-S-E-1A	0.1185	0.5591	0.7298	0.0873

### Results on unclear topics for Japanese

Run	Hscore	Fscore	Sscore	H-measure
KLE-S-J-3A	0.2030	0.4416	0.5086	0.1038
KLE-S-J-4A	0.2025	0.3920	0.4997	0.1008
KLE-S-J-2A	0.1867	0.4502	0.4697	0.0908
KLE-S-J-1A	0.1759	0.4372	0.4509	0.0853

### Results on unclear topics for Chinese

Run	Hscore	Fscore	Sscore	H-measure
KLE-S-C-2A	0.5413	0.5736	0.6339	0.3360
KLE-S-C-1A	0.5306	0.5666	0.6360	0.3303
KLE-S-C-4A	0.5148	0.4986	0.6640	0.3279
KLE-S-C-3A	0.5072	0.4817	0.6718	0.3255

- Achieve high performances on second-level subtopic
- Show low Hscores due to considering word correlation only