

# SSEM X UNIST health care center Project Progress presentation

Improving UNIST Healthcare Center  
counseling Service using text mining



20176008 Junmo Nam  
20176019 Jingyu Jang

# INTRODUCTION

## **Problems due to lack of counseling staff compared to the demand of students**

- Most of students give up counseling because their waiting time is too long
- The counseling center has 5,000 students per counselor
- Most of students get better by simple counseling

→ Improve the efficiency of initial counseling using the counseling Q & A data

# Overview

UNIST freshmen survey data  
(undergraduate /graduate )

Survey 1						
No.	Question	Disagree				agree
1	Q1	1	2	3	4	5
2	Q2	1	2	3	4	5
...	...	...	...	...	...	...

Survey 1						
No.	Question	Disagree				agree
1	Q1	1	2	3	4	5
2	Q2	1	2	3	4	5
...	...	...	...	...	...	...

...

Survey 6						
No.	Question	Disagree				agree
1	Q1	1	2	3	4	5
2	Q2	1	2	3	4	5
...	...	...	...	...	...	...

Survey 1 Results example

Score	Document
~69	Doc 1
70~92	Doc 2
93~115	Doc 3
116~	Doc 4

The purpose of the client's visit dataset  
(UNIST Healthcare center)

No	Topic1	Topic2	Topic3
Student 1	W1	w2	w3
...	...	...	...
...	...	...	...

Naver Q&A dataset

Question	Answer
Q1	A1
...	...
...	...
...	...

3종류의 데이터 셋

Assumption

1. 설문 질문 조합으로 만든 문서가 실제 결과 설명 문서와 연관성이 있을 것
2. 지식인 Q&A Pair의 신뢰성이 있을 것
3. 지식인 답변과 설문 결과 문서가 같은 Topic에 대해 유사할 것

# Overview

## Student 1

Question	Disagree				Agree
Survey1 Q1	1	2	3	4	<b>5</b>
Survey1 Q2	<b>1</b>	2	3	4	5
Survey1 Q3	1	2	<b>3</b>	4	5
Survey1 Q4	1	<b>2</b>	3	4	5
...	...	...	...	...	...

- 학생들의 답변을 기준으로 강한 동의 일수록 높은 확률 값 설정
- 이 확률 값을 기준으로 각 설문 문항에 대한 학생들의 text 생성
- 설문 결과에 해당하는
- 학생별로 모든 설문 결과 text merge

0%

~

100%



Generate text based on answer (0~100%)

Question	Survey 1 text	Survey 1 label	...
Student 1	Q1,Q3...	Text 1	...
Student 2	Q1,Q5...	Text 2	...
Student 3	Q2,Q3...	Text 3	...
...	...	...	...



Question	Survey text	Survey label
Student 1	Q1,Q3...	Results document1
Student 2	Q1,Q5...	Results document2
Student 3	Q2,Q3...	Results document3
...	...	...

# Overview

Naver Q&A dataset

Question	Answer
Q1	A1
...	...
...	...
...	...

The purpose of the client's visit dataset  
(UNIST Healthcare center)

No	Topic1	Topic2	Topic3
Student 1	W1	w2	w3
...	...	...	...
...	...	...	...

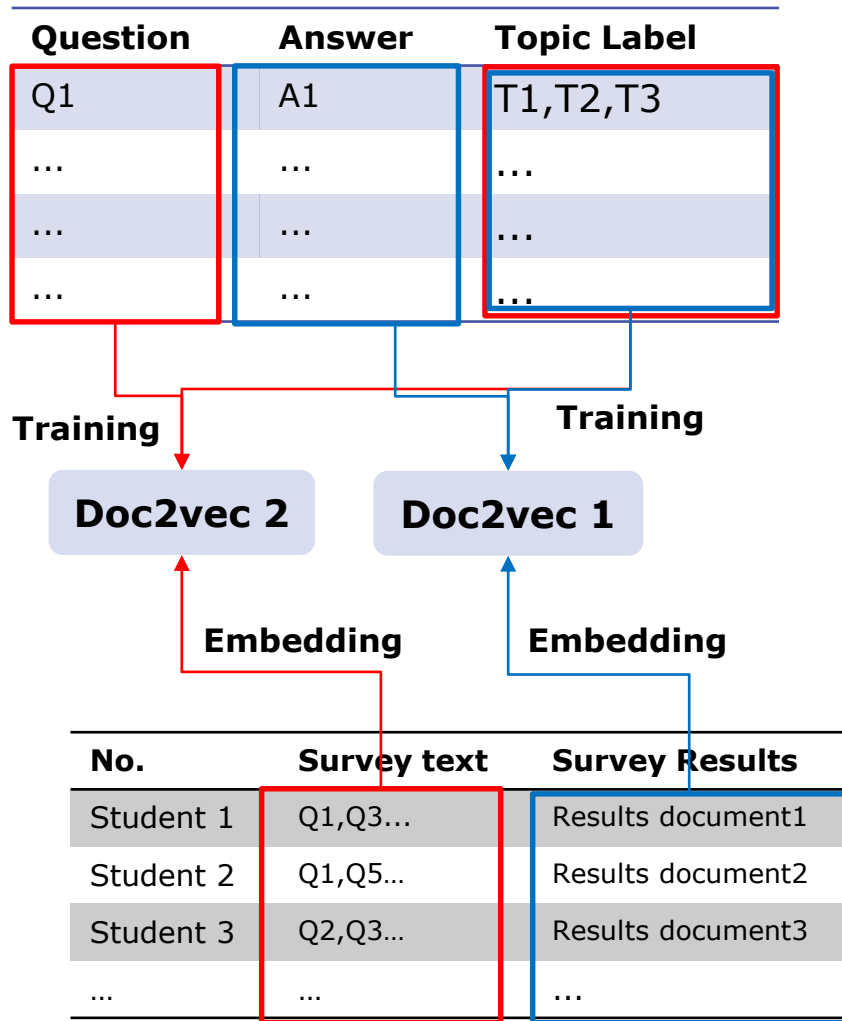
Prior probability =  $\{w1:0.1, w2:0.05, \dots\}$

Topic modeling  
(LDA)

Naver Q&A dataset

Question	Answer	Topic Label
Q1	A1	Topic1
Q2	A2	Topic2
Q3	A3	Topic1
...	...	...

# Overview



## Model results

No.	Survey text	Model1 Topic label	Model 2 Topic Label	Top n accuracy
Student1	Text1	T1,T2,T3 ...	T1,T3,T5...	0.66
Student2	text2	...	...	0.55
...	...	...	...	...
...	...	...	...	...

모델 정확도가 높다는 것은 ?

