

# 코딩을 통한 문예사조 분석

디지털 인문역량경진대회 우리42조

2022130875 고유진, 2022131104 김지연

2022131024 노명서, 2022130802 장세린, 2022130849 정지민

## 코딩을 통한 문예사조 분석

연구의 목적

선행 연구

코드

결과

## 연구의 목적

코딩을 통해 문예사조의 특징을 분석하는 것을 목적으로 한다.



고딕



낭만주의

## 고딕

- 1 인간의 어둡고 비이성적인 면
- 2 공포스러운 분위기, 어두운 것을 상징하는 요소
- 3 빈도 분석 시 고딕을 대표하는 단어가 등장할 것이다
- 4 부정적인 의미를 가진 단어의 빈도수가 높을 것이다
- 5 주인공을 중심으로 **concordance**를 출력하여 감성 분석을 진행하면 'Negative'의 비율이 높을 것이다

## 낭만주의

- 1 이상향적이고 상상력 중시
- 2 정신적, 지적 동향에서 자유로운 경향
- 3 빈도 분석 시 낭만주의를 대표하는 단어가 등장할 것이다
- 4 긍정적인 의미를 가진 단어의 빈도수가 높을 것이다
- 5 주인공을 중심으로 **concordance**를 출력하여 감성 분석을 진행하면 'Positive'의 비율이 높을 것이다

### 3. 코드

#### #PIPE 분석

```
from nltk.tokenize import word_tokenize
from nltk import pos_tag
from nltk.stem import WordNetLemmatizer
import nltk

from nltk.tokenize import word_tokenize
from nltk import pos_tag
from nltk.stem import WordNetLemmatizer
import nltk

nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('omw-1.4')
nltk.download('wordnet')

lemmatizer = WordNetLemmatizer()

with open('The Fall of the House of Usher .txt') as f:
    txt = f.read()

txt = txt.lower()
txt = word_tokenize(txt)
txt = pos_tag(txt)
    lemma = lemmatizer.lemmatize(token[0], pos=token[1])
    elif token[1].startswith('R'):
        lemma = lemmatizer.lemmatize(token[0], pos='r')
```

```
lemmas = list()
for token in txt:
    if token[1].startswith('J'):
    else:
        lemma = lemmatizer.lemmatize(token[0], pos='a')
    elif token[1].startswith('V'):
        lemma = lemmatizer.lemmatize(token[0], pos='v')
    elif token[1].startswith('N'):
        lemma = lemmatizer.lemmatize(token[0], pos='n')
    elif token[1].startswith('R'):
        lemma = lemmatizer.lemmatize(token[0], pos='r')
    else:
        lemma = lemmatizer.lemmatize(token[0])
    lemmas.append(lemma)

from nltk.text import Text
content_text = Text(lemmas)
A=content_text.concordance_list('usher')

with open('concor_mohican.txt', 'w') as f:
    for kwic in A:
        f.write(f"{kwic.line}\n")

with open('concor_mohican.txt', 'r') as f:
    last=f.read()

last=last.split("\n")

pipe = pipeline("text-classification")
pipe(last)
```

## 4. 결과

### #1차 결과

```
[('pattern', 4),  
 ('paper', 3),  
 ('great', 3),  
 ('shake', 3),  
 ('around', 2),  
 ('place', 2),  
 ('commit', 2),  
 ('dull', 2),  
 ('enough', 2),  
 ('eye', 2),  
 ('follow', 2),  
 ('woman', 2),  
 ('behind', 2),  
 ('sometimes', 2),  
 ('crawl', 2),  
 ('get', 2),  
 ('paint', 1),  
 ('look', 1),  
 ('boys', 1),  
 ('school', 1)]
```

#### The Yellow Wallpaper

##### 빈도 분석

##### 결과

- 고딕을 대표한다고 볼 수 있는 단어를 찾아볼 수 없음
- 긍정적이거나 부정적 의미와 상관없는 단어들의 비율이 높음

##### 감성 분석

##### 결과

- 'Negative'와 'Positive'의 비율이 유사하게 나타남
- 규모가 작아서 유의미한 결과라고 보기 어려움

```
[{'label': 'NEGATIVE', 'score': 0.9716004729270935},  
 {'label': 'NEGATIVE', 'score': 0.9993558526039124},  
 {'label': 'POSITIVE', 'score': 0.7481208443641663}]
```

### 실패 원인 분석과 1차 개선 방안



### #2차 결과

```
[('get', 42),  
 ('one', 34),  
 ('would', 27),  
 ('see', 27),  
 ('go', 24),  
 ('room', 23),  
 ('good', 22),  
 ('pattern', 22),  
 ('like', 21),  
 ('paper', 21),  
 ('think', 21),  
 ('little', 20),  
 ('know', 19),  
 ('creep', 19),  
 ('well', 18),  
 ('take', 18),  
 ('look', 18),  
 ('much', 16),  
 ('window', 16),  
 ('make', 15)]
```

{ 'label': 'NEGATIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'POSITIVE',
{ 'label': 'POSITIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'POSITIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'POSITIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'POSITIVE',
{ 'label': 'POSITIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'NEGATIVE',
{ 'label': 'POSITIVE',
{ 'label': 'POSITIVE',

### The Yellow Wallpaper(Whole)

#### 빈도 분석

##### 결과

- 1차 결과에서의 작은 결과값이라는 단점 보완
- 그러나 여전히 고딕을 대표한다고 볼 수 있는 단어를 찾기 어려움

#### 감성 분석

##### 결과

- 이 또한 1차 결과에서의 작은 결과값이라는 단점 보완
- 'Negative'와 'Positive'의 비율이 유사하다는 점은 개선되지 않았음

### 2차 개선 방안



### #최종 결과

```
[('felt', 2),  
 ('funeral', 2),  
 ('brain', 2),  
 ('kept', 2),  
 ('tread', 2),  
 ('till', 2),  
 ('break', 2),  
 ('beat', 2),  
 ('mourner', 1),  
 ('fro', 1),  
 ('seem', 1),  
 ('sense', 1),  
 ('seat', 1),  
 ('service', 1),  
 ('like', 1),  
 ('drum', 1),  
 ('think', 1),  
 ('mind', 1),  
 ('go', 1),  
 ('numb', 1)]
```

#### I felt a Funeral, in my Brain

##### 빈도 분석 결과

- 'funeral'이라는 고딕적 특징이 두드러지는 단어가 빈도 분석 결과에 나타남

##### 감성 분석 결과

- 'Negative'의 비율이 'Positive'의 비율보다 높게 나타남

##### 연구의 한계

- 시라는 작품의 특성상 데이터의 규모가 작다는 한계는 여전히 존재

```
[{'label': 'NEGATIVE', 'score': 0.974245548248291},  
 {'label': 'NEGATIVE', 'score': 0.9984663724899292},  
 {'label': 'NEGATIVE', 'score': 0.994211733341217},  
 {'label': 'NEGATIVE', 'score': 0.924787700176239},  
 {'label': 'POSITIVE', 'score': 0.9954480528831482},  
 {'label': 'POSITIVE', 'score': 0.7481208443641663}]
```

## 4. 결과

### #예외 분석

```
[{'label': 'NEGATIVE', 'score': 0.9916533827781677},  
{ 'label': 'NEGATIVE', 'score': 0.9941120743751526},  
{ 'label': 'NEGATIVE', 'score': 0.95807284116745},  
{ 'label': 'POSITIVE', 'score': 0.9987047910690308},  
{ 'label': 'NEGATIVE', 'score': 0.9945055842399597},  
{ 'label': 'POSITIVE', 'score': 0.9734799265861511},  
{ 'label': 'NEGATIVE', 'score': 0.9840548634529114},  
{ 'label': 'POSITIVE', 'score': 0.7471909523010254},  
{ 'label': 'NEGATIVE', 'score': 0.9253853559494019},  
{ 'label': 'NEGATIVE', 'score': 0.8564864993095398},  
{ 'label': 'POSITIVE', 'score': 0.9832224249839783},  
{ 'label': 'POSITIVE', 'score': 0.7481208443641663}]
```

#### <The Raven>의 감성 분석

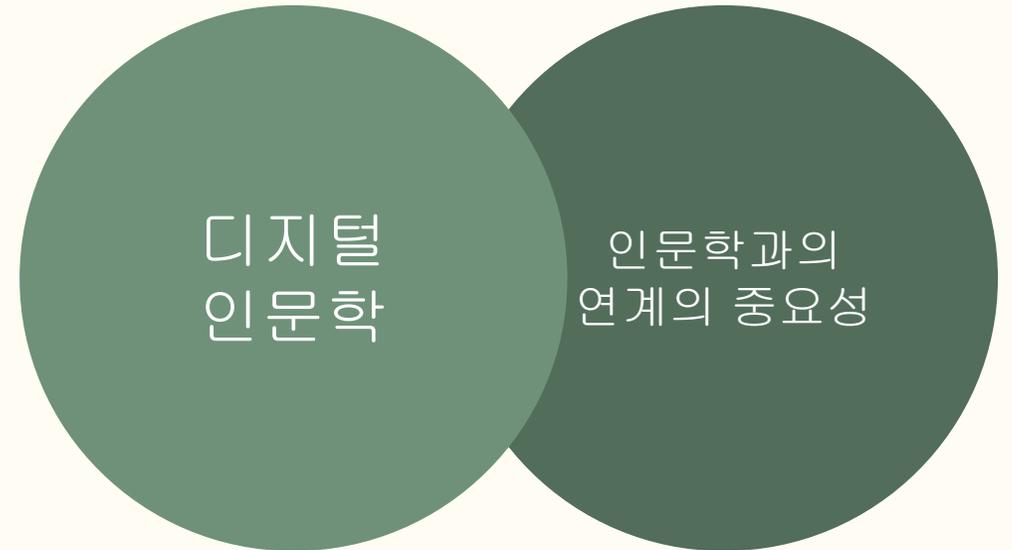
결과

```
[{'label': 'NEGATIVE', 'score': 0.9963341951370239},  
{ 'label': 'NEGATIVE', 'score': 0.9979506134986877},  
{ 'label': 'NEGATIVE', 'score': 0.714141845703125},  
{ 'label': 'POSITIVE', 'score': 0.7384368181228638},  
{ 'label': 'POSITIVE', 'score': 0.7481208443641663}]
```

#### <Proverbs of Hell>의 감성 분석

결과

결론



발표를 들어 주셔서 감사합니다.

디지털 인문역량경진대회 우리42조

2022130875 고유진, 2022131104 김지연

2022131024 노명서, 2022130802 장세린, 2022130849 정지민