



감	성	과		문	학		
---	---	---	--	---	---	--	--

.. 감성 어휘 중심의
출판 콘텐츠 마케팅 전략 제안

심영석 — 이혜린 — 조재연
빅데이터 청년인재 — 명지대 육조

목차



소개

분석 배경 — 분석 목적



분석 기획

데이터 소개 — 분석 방법



분석 결과

감성 분류 — 감성 마케팅 효과 — 감성 트렌드



제안점

Insight — 기대 효과 — 한계





소개
분석배경

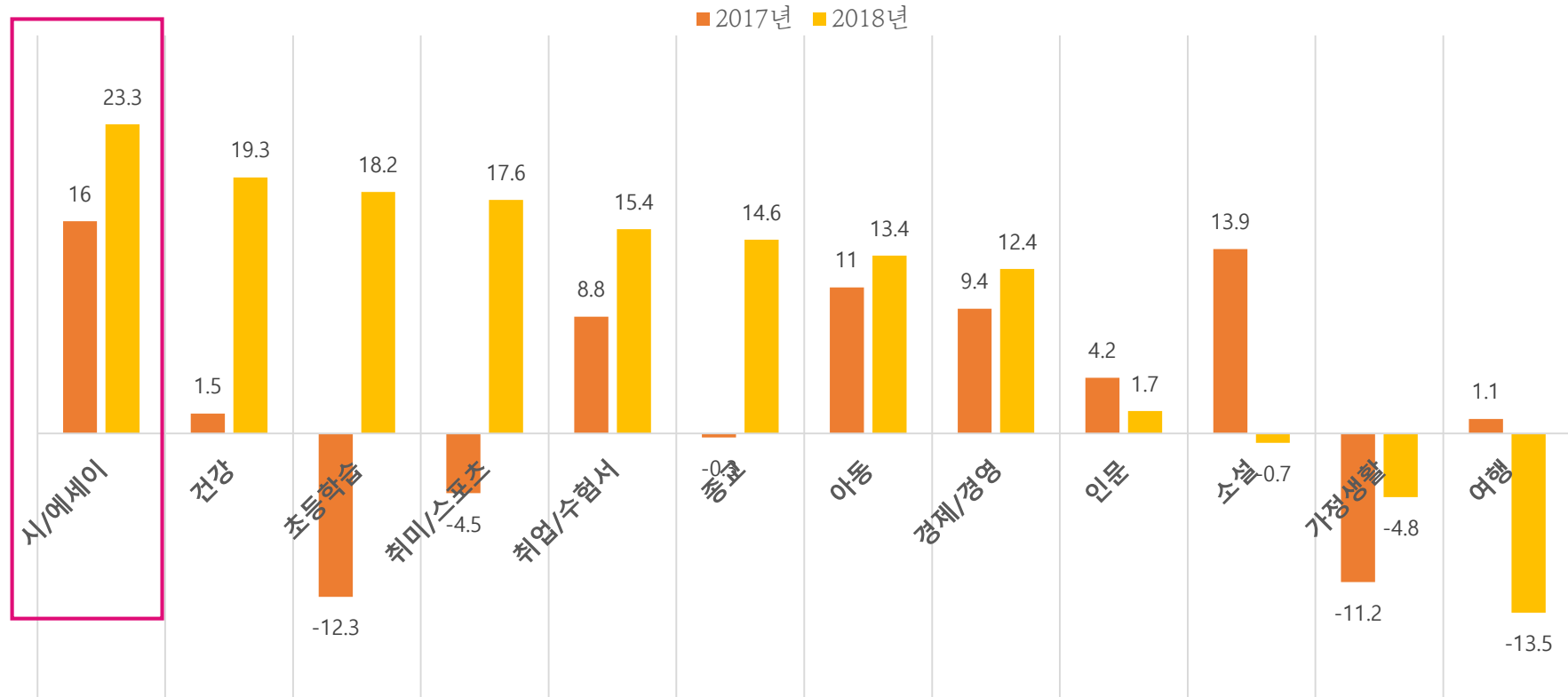
당신은 왜 ‘시(詩)’를 읽나요?

“사람들은 이제 더 이상 필요한 지식을 얻기 위해 책을 찾아 다니지 않는다”

소개

분석 배경

교보문고 분야별 연간 판매액 변화추이

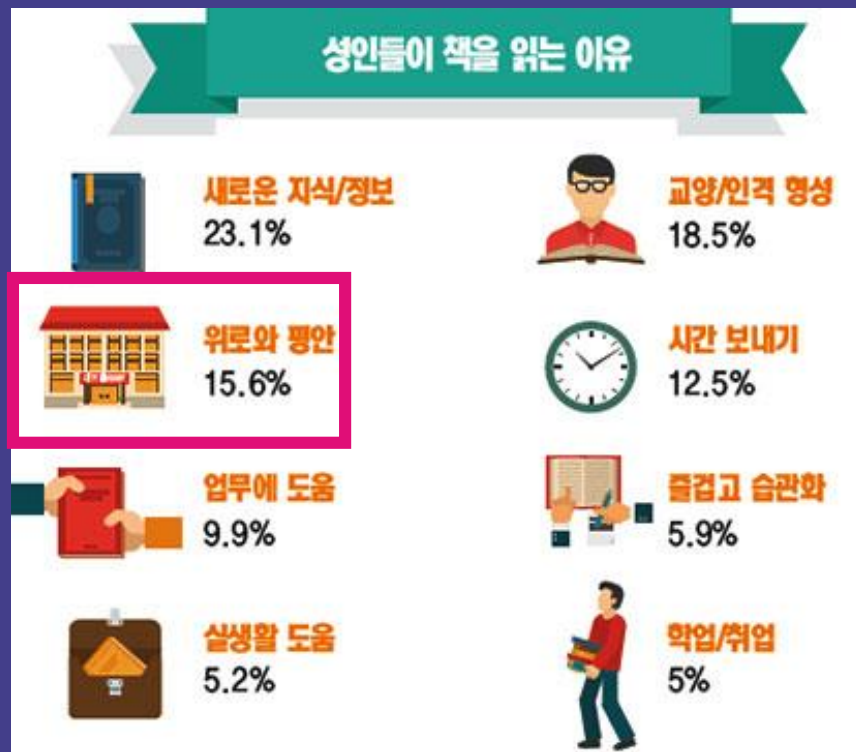


=> 2018년 가장 인기를 끈 분야는 시/에세이 분야로,
3년 연속 두 자릿수 신장률 기록하며 판매 권수와 판매액 모두 상승함.

소개

분석 배경

사람들은 왜 '시/에세이'를 읽을까?



Q 슬플때 기분이 씁쓸하고 안좋을때 읽으면 마음이 조금이라도... 2005,10,05,

슬플때 기분이 씁쓸하고 안좋을때 읽으면 마음이 조금이라도 평온해질만한 책 사람의 기분은 글로다스릴 수없나요?... 다수릴수있는것 같은데... 여러분들이 책좀 추천해주세요^^

Q 우울할때 슬플때 화날때 읽으면 좋은 동화 추천좀 해주세요.. 2013,10,02,

우울할때 화날때 슬플때 읽으면 아이들 동화 추천해주세요, 많으면 많을수록 좋아요.

Q 우울하고 슬플때, 무기력할때, 볼만한 영화와 애니와 책 추천해... 2008,06,29,

우울하고 슬프고 무기력하고 살기싫고 그런 축 처지는 기분 들 때 볼 만한 영화, 애니메이션, 책 추천해주세요.

Q 기분이 우울하거나 시도때도없이 슬플때 2018,10,15,

기분이 우울하거나 시도때도없이 슬플때 마음을 진정시킬만한 책이 있을까요? 보면 마음이 편해질까..

소개
분석배경

2017년 1년 동안 출판된 책만 8만 130종, 하루에 새로운 책이 200종 넘게 태어나고 있다.
이렇게 많은 책 중에서, 어떻게 하면 실패 없이 나에게 딱 맞는 책을 고를 수 있을까?

출처: 전안나(2019.06.25), "실패 없이 책 고르는, 고수의 비법", pub.chosun

추천도서

YES24 편집자 리뷰	<ul style="list-style-type: none"> · 가정 살림 · 만화/라이트노벨 · 역사 · 자기계발 · IT/모바일 	<ul style="list-style-type: none"> · 건강 취미 · 문학 · 예술 · 자연과학 	<ul style="list-style-type: none"> · 경제/경영 · 사회/정치 · 유아/어린이 · 종교 	<ul style="list-style-type: none"> · 국어 외국어 사전 · 여행 · 인문 · 청소년
YES24 오늘의 책	<ul style="list-style-type: none"> · 2019년 · 2015년 · 2011년 · 2007년 · 2003년 · 분야별 보기 	<ul style="list-style-type: none"> · 2018년 · 2014년 · 2010년 · 2006년 · 2002년 	<ul style="list-style-type: none"> · 2017년 · 2013년 · 2009년 · 2005년 · 2001년 	<ul style="list-style-type: none"> · 2016년 · 2012년 · 2008년 · 2004년 · 2000년
올해의 책	<ul style="list-style-type: none"> · 2018년 올해의 책 ... · 2016년 올해의 책 ... · 2014년 올해의 책 ... 	<ul style="list-style-type: none"> · 2018년 올해의 책 ... · 2016년 올해의 책 ... · 2014년 올해의 책 ... 	<ul style="list-style-type: none"> · 2017년 올해의 책 ... · 2015년 올해의 책 ... · 2013년 올해의 책 ... 	<ul style="list-style-type: none"> · 2017년 올해의 책 ... · 2015년 올해의 책 ... · 2013년 올해의 책 ...

편집자 추천, 판매량 기준, 문학상 수상 기준 등으로 도서 추천

소개
분석배경

다른 문화 콘텐츠들은 어떤 기준으로 콘텐츠를 추천해줄까?



넷플릭스, 알고리즘 아닌 사람 추천 기능 만든다

김민선 기자 입력 2019.08.25. 13:40 수정 2019.08.25. 14:19

10



| 넷플릭스 크리에이티브 팀 전문가가 추천

(지디넷코리아=김민선 기자)넷플릭스가 알고리즘이 아닌 사람 추천 콘텐츠 구역을 따로 만든다고 23일(현지시간) IT 매체 테크크런치가 보도했다.

넷플릭스는 사람이 추천한 콘텐츠를 표출하는 '컬렉션즈'를 iOS 기기 이용자들을 대상으로 실험 중이다.

출처: 김민선(2019.08.25), "넷플릭스, 알고리즘 아닌 사람 추천 기능 만든다", ZD Net Korea

노연정. (2018). 감성기반 영화 추천 서비스 및 시스템 연구 - 넷플릭스를 중심으로. 한국정보과학회 학술발표논문집, pp2,006-2,008.



소개
분석 배경

온라인 서점도 '감성' 기준으로 책을 추천해준다면 어떨까?

For what?

1. '위로와 평안'을 위해 읽는 시/에세이
2. 도서 추천 시스템의 부재
3. 정형데이터 분석으로는 한계가 있는 콘텐츠 분석

#행복



#우울



#좌절

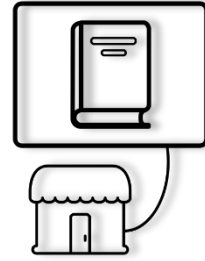




소개

분석 목적

‘감성과 문학’ 데이터 분석 목적



1. 출판 콘텐츠에 적합한 ‘문학 감성 사전’을 구축하여 도서 ‘감성’ 분류
 - > 정성적 분석(text분석 및 내용분석)을 통해 도서의 감성을 분석하고 분류
 - > 소비자에게 도서, 특히 문학작품을 추천하는 새로운 기준 제안
2. 도서 구매에 미치는 영향을 살펴봄으로써 실질적인 마케팅 전략 제안
 - > 도서 구매에 영향을 미치는 영향 가운데 ‘책표지’의 컬러와 감성이 도서 구매에 미치는 영향
 - > 도서 감성 트렌드가 도서 구매에 미치는 영향
 - > 책 표지색과 감성이 일치할 때, 판매에의 영향 확인
3. 문화콘텐츠 추천 시스템에 적합한 기준으로 ‘감성’을 제안
 - > 고객의 도서 구매 결정 시간을 줄이고, 구매 후 만족도를 높일 수 있는 감성추천 제안
 - > 감성 트렌드의 체계적인 분석을 통한 효과적인 감성 마케팅 제안



분석 기획

데이터 소개 | 분석 방법



분석 데이터 정의

1. 온라인 서점 매출액 기준 1위 사업자인 'yes24'의 국내도서 > 소설/시/희곡 > 시/희곡
2. 시/희곡 장르 도서 전체를 수집: 총 28,826권의 도서를 대상으로 아래의 데이터 수집

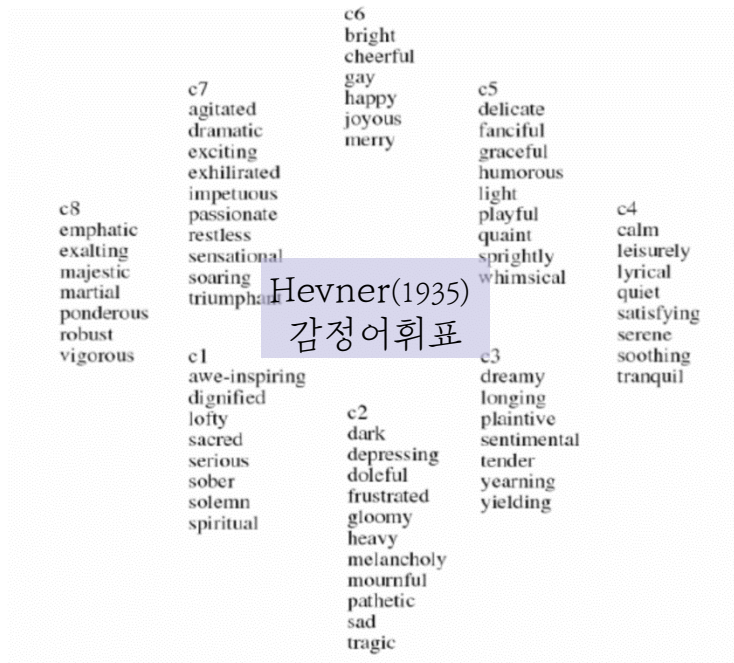
분석 도구

Python, R, SPSS

	데이터 정의	데이터 유형
Id	각 도서 페이지의 URL 고유값	명목척도, 주요 키
제목	책 제목	CHAR
출간일	책 출간일 (ex 2019.08.28)	DATE
누적판매순위	2019년 8월 23일 기준 누적 판매 순위(1위~28,502위)	서열척도
평점	도서를 구매한 고객의 책에 대한 평가 점수(5점 척도)	등간척도
소개글	판매자가 입력한 책 내용 소개 텍스트	CHAR
감성값	소개 글을 기반으로 추출한 도서의 감성 값	등간척도
표지 그림	책을 대표하는 표지 이미지	이미지(jpg)
RGB	표지 이미지에서 추출한 RGB값 (1~255)	등간척도

분석 기법 1) 감성값 분석 방법 (Python)

- 1) 크롤링(책 소개글)
- 2) 데이터 전처리
- 3) 감성어휘 추출
: 형태소분석_Konlpy.tag) (test data 1,000개의 도서 소개글에서 564개의 단어추출)
- 4) ‘문학 감성어휘 사전’ 구축
: 3명의 코더가 내용분석을 통해 Hevner의 감성어휘표를 기준으로 8개 그룹으로 나눔



감성그룹	단어 수	예시단어
G1 (담담, 진지, 경외)	31	깊은, 이해, 성찰, 사유
G2 (우울, 좌절, 어둠)	151	슬픔, 비극, 죽음, 고통
G3 (꿈, 갈망, 동경)	55	이상, 자유, 그녀, 그리워
G4 (고요, 만족, 잔잔)	83	잔잔하게, 서정, 위로, 공감
G5 (기묘, 환상, 우아)	63	새로운, 영원, 독특, 달
G6 (밝음, 행복)	67	희망, 용기, 행복한, 눈부신
G7 (들뜸, 놀람, 충동)	28	충동, 욕망, 절정, 폭발
G8 (단호, 강건, 호전)	54	외면, 불화, 긴장, 날카로운
기타(분류불가)	32	심리, 성의, 풍부한, 부끄럼

분석 기법 1) 감성값 분석 방법 (Python)

- 5) 전체 텍스트 데이터를 기준으로 ‘문학 감성어휘 사전’에 규정된 단어들을 매칭, 추출 후 그룹화 : 감성사전에 매칭되지 않는 결측 데이터 삭제 -> 17,549개의 데이터로 축소

분류 예시

시를 잊은 나와 그대에게 선물하는
감성 라이팅북

“꽃 피는 봄날,
비 내리는 여름날,
낙엽 뒹구는 가을날,
눈 쌓이는 겨울날,
시집을 안 사면 무엇을 살 것인가?”

작가이자 철학자인 김용규는 이야기한다. 시는 흑백영화처럼 진부해
진 일상과 낯익은 세계를 하나둘씩 형형색색 새롭게 태어나게 만든다
고. 고되고 권태롭고 무의미한 삶을 새롭게 구제한다고.

그러나 언제부터인가 삶에서 시는 잊혀져갔다. 정확히 말해 우리 주
변에는 여전히 시가 있건만, 가슴에 담을 여유가 없다. 지하철 스크린
도어에, 예쁜 음식점 벽에, 페이스북 페이지에 넘쳐나는 시를 읽어도
더 이상 가슴이 울리지 않는 우리. 시를 잊은 나와 그대에게 시를 되
돌려줄 방법은 없는 것일까?

G2에 카운트

G3에 카운트

감성그룹	감성빈도
G1 (담담, 진지, 경외)	0
G2 (우울, 좌절, 어둠)	2
G3 (꿈, 갈망, 동경)	0
G4 (고요, 만족, 잔잔)	1
G5 (기묘, 환상, 우아)	0
G6 (밝음, 행복)	1
G7 (들뜸, 놀람, 충동)	0
G8 (단호, 강건, 호전)	0

분석 기법 1) 감성값 분석 방법 (Python)

5) 전체 텍스트 데이터를 기준으로 ‘문학 감성어휘 사전’에 규정된 단어들을 매칭, 추출 후 그룹화

```
from openpyxl import load_workbook
from konlpy.tag import Kkma

kkma = Kkma()

compare = []

match_matrix = list()

for index in range(1,9):
    with open("sensi_dict_{}.txt".format(index),'r') as f:
        compare = [line.replace('\n', '') for line in f.readlines()]

    wb = load_workbook('yes24_all.xlsx')
    sheet = wb['Sheet1']

    row = 2 # 2번째 행부터 읽을

    match_count = list()

    while True:
        text = sheet.cell(row=row, column=2).value # row 행 2 열의 내용을 가져올
        print('***{}번째 진행중***'.format(row))
        if not text: # 내용이 비어있으면
            break # 더이상 읽지 않을

        #print(text) # 문장 출력

        poslist = kkma.pos(text) #행 별 형태소 분석한 파일 poslist에 할당

        match = list()

        #비교 대조를 위한 구분
        sum_count = 0
        for pos1 in compare:
            count = 0
            for pos2 in poslist:
                if pos1 == str(pos2):
                    count += 1

            if count > 0:
                match.append((pos1, count))

        sum_count += count

        match_count.append(sum_count)
```

	A	B	C	D	E	F	G	H	I	J
1	ID	g1	g2	g3	g4	g5	g6	g7	g8	max_group
2	24259730	0	0	0	0	1	0	0	0	5
3	41847244	0	0	2	7	8	0	3	0	5
4	69669326	3	3	17	35	7	2	1	0	4
5	18489338	0	2	6	7	0	3	2	1	4
6	18182099	0	2	6	7	3	5	1	0	4
7	67116558	0	0	0	2	1	4	0	0	6
8	23599602	0	0	0	4	9	0	0	0	5
9	75187745	3	1	1	8	1	1	1	0	4
10	77985305	1	0	6	1	1	1	0	1	3
11	8150462	3	4	0	3	0	0	0	0	2

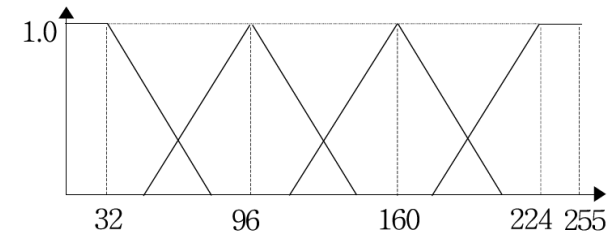
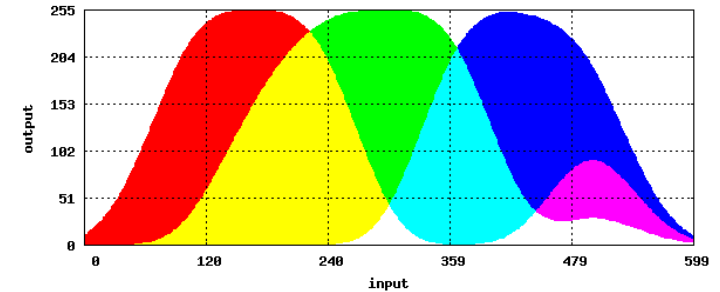
: 가장 많이 산출된 단어를 기준으로 해당 도서의 감성 규정
: 중복값의 경우

분석 기획

데이터 소개

분석 기법 2) 표지 색 추출 방법 (Python)

- 1) 크롤링(표지 이미지.jpg)
- 2) 데이터 전처리(NA제거) -> 총 17,425개 분석
- 3) 이미지 RGB값 추출(open CV사용) : 4구간으로 나눠 RGB 각 값의 median 추출



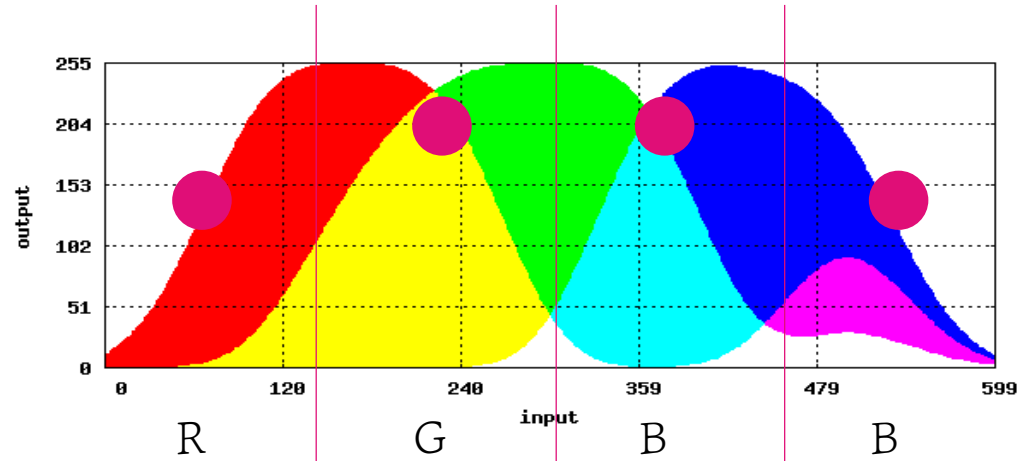
(그림 3) R,G,B 각각의 소속함수

<표 1> R,G,B 각각의 소속도

퍼지값 (R,G,B 각각의 색채 빈도수)	소속구간
적다 (L)	[0, 79]
조금 적다 (SL)	[49, 143]
조금 많다 (SH)	[113, 207]
많다 (H)	[177, 255]

분석 기법 2) 표지 색 추출 방법 (Python)

4) 추출된 각 4구간 별, 최대값을 갖는 값 만을 남겨둠



5) 각 4구간의 RGB 값은 각각 1~3의 명목변수로 전환하여 사용하였으며, 각 4구간 가운데에서도 최대로 등장한 RGB 가운데 하나를 해당 도서 표지의 대표색으로 정의하고 변수를 하나 더 생성

분석 결과

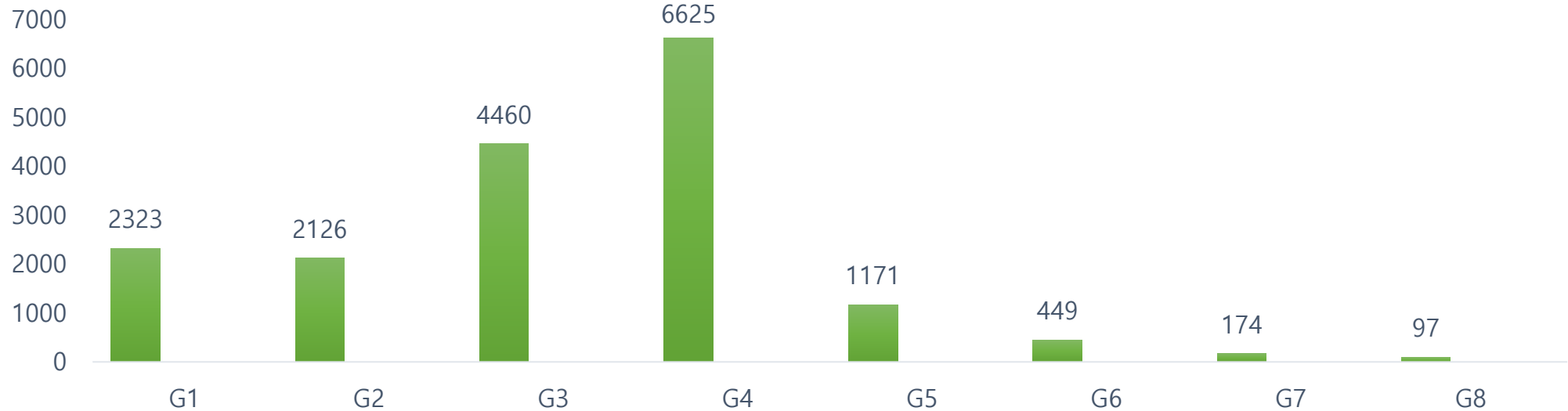
감성 분류 | 감성이 판매에 미치는 영향 | 감성 트렌드



분석 결과

도서 감성 분류

분석 0) 도서 감성 분류



감성그룹	분류 도서
G1 (담담, 진지, 경외)	침묵이 풍경이 되는 시간, 돈이라는 문제, 다윈의 거북이, 학식을 뿜내는 여인들 등
G2 (우울, 좌절, 어둠)	살아 있는 모든 것은 다 흔들린다, 그리움 외로움 그리고 고독, 마음의 행로, 목화밭의 고독 속에서 등
G3 (꿈, 갈망, 동경)	문득 사람이 그리운 날엔 시를 읽는다, 바람을 바라보다, 자유의 나무 한 그루, 그래서 사랑 등
G4 (고요, 만족, 잔잔)	파란을 넘어서, 이제야 내 뒤를 돌아본다, 오래된 맑음, 오늘 같이 있어, 풀꽃, 울고 싶어도 내 인생이다 등
G5 (기묘, 환상, 우아)	별을 만든 시인과 아이스크림, 빈센트 반 고흐, 빛의 색 고흐를 만나다, 꽃, 모여서 산다 등
G6 (밝음, 행복)	긍정의 빛, 나 이제 행복해도 되겠지, 행복을 주는 글, 봄의 왈츠, 모든 순간이 다 꽃으로 필 거예요 등
G7 (들뜸, 놀람, 충동)	녹색비단구렁이, 황홀한 숲, 치명적인 것들, 바람부는 날이면 압구정동에 가야 한다 등
G8 (단호, 강건, 호전)	분단과 통일의 성찰, 시와 리얼리즘, 방부제가 썩는 나라, 바알, are you sure? 등

도서 감성 분류
분석 결과

분석 1) 도서의 감성과 색

- > 색 표지 주요색 가운데에서는, B계열의 표지색이 과반수 이상을 차지
- > ANOVA 분석 결과,
감성 그룹별로 RGB에 차이가 있는 것으로 나타남 ($F=2.104$, $p=.040$)

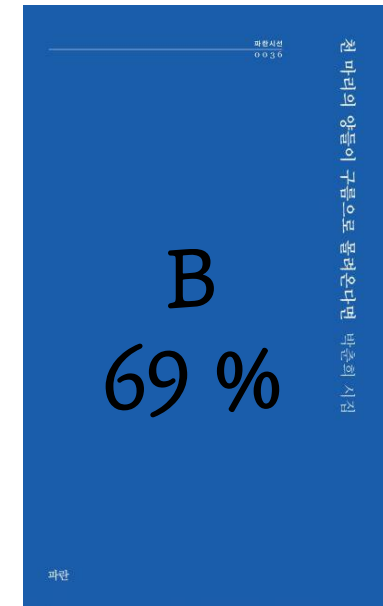
R
7 %



G
23 %



B
69 %



도서 감성 분류
분석 결과

분석 1) 도서의 감성과 색



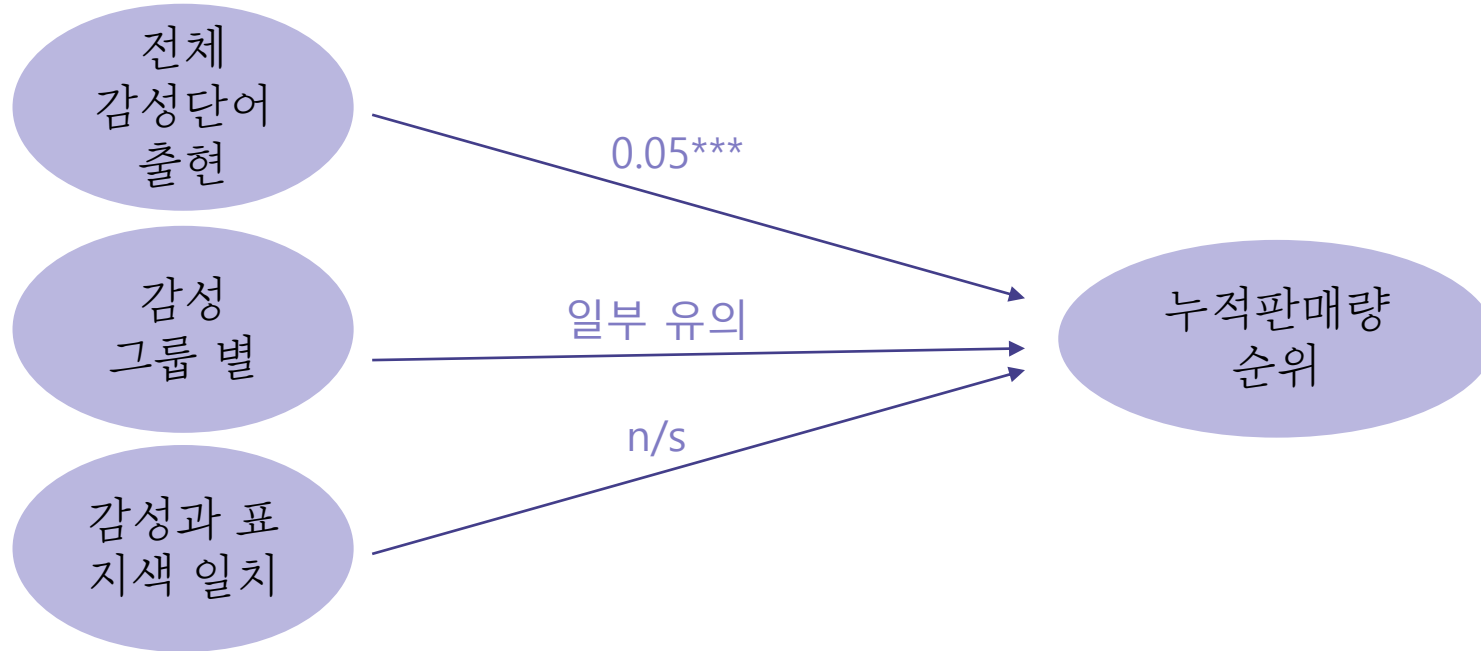
-> 상관관계 분석 결과,
RGB는 G1(-0.18*), G3(-0.20**), G4(-0.23**), G6(-0.18*), G7(-0.20**)으로 부(-)의 상관관계를 나타냄.

-> 즉, R=1, G=2, B=3 임을 고려했을 때, ‘진지’, ‘갈망’, ‘고요’, ‘행복’, ‘충동’과 관련한 도서들은 R 계열에 가까운 채표지색과 상관관계를 갖는 것으로 나타남.

-> 회귀분석에 있어서는 G7(-0.07*)(충동)의 감성의 도서 만이 R계열의 채표지색에 예측력을 가짐

도서 감성 분류
분석 결과

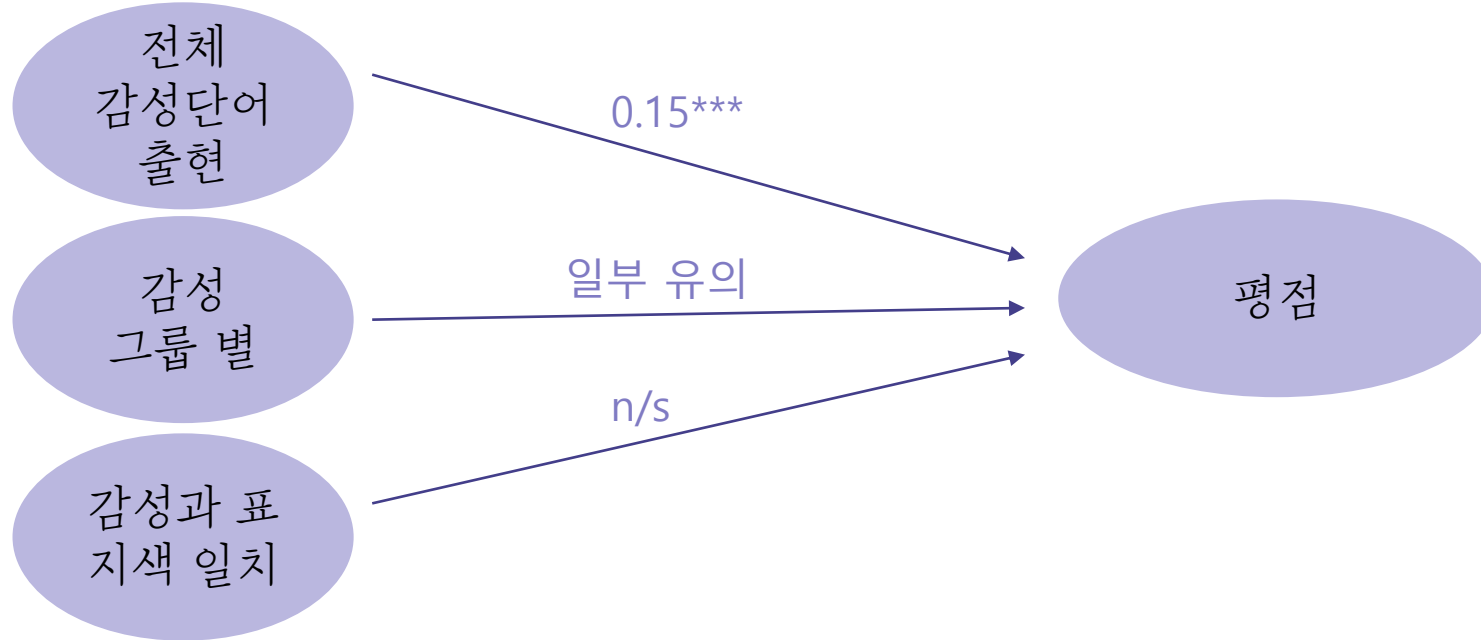
분석 2) 감성과 책표지가 판매에 미치는 영향



- 1) ‘감성단어 출현’: 감성단어가 많이 사용될수록 누적판매량 순위가 높아짐
- 2) ‘감성그룹 별’: 감성단어 별, 누적판매량 순위에 차이가 있는 것으로 나타났으며($F=10.762$, $p<.001$), 특히 G4(고요), G5(환상), G7(충동)의 감성단어가 사용될수록 누적판매량 순위가 높아짐
- 3) ‘감성과 표지색 일치’: 앞에서 분석한 상관관계분석에 근거하여 G1,3,4,6,7에 해당하는 감성의 도서 표지색이 B가 아닐 경우, 감성과 표지색이 일치한다는 더미변수를 생성하여 분석하였으나 감성과 표지색 일치 여부는 누적판매량 순위에 유의미한 영향을 주지 않는 것으로 나타남

도서 감성 분류
분석 결과

분석 3) 감성과 책표지가 평점에 미치는 영향

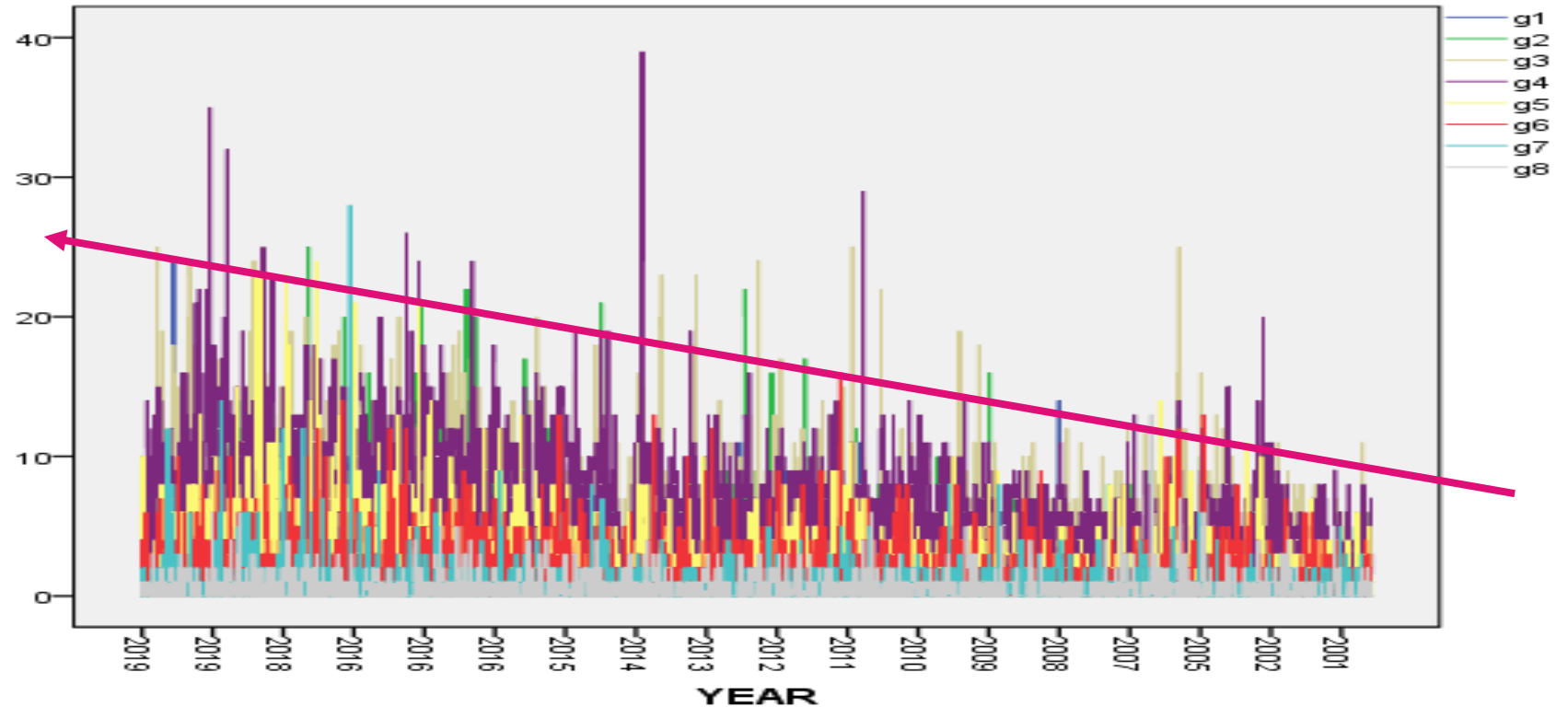


- 1) ‘감성단어 출현’: 감성단어가 많이 사용될수록 평점이 높아짐
- 2) ‘감성그룹 별’: 감성단어 별, 평점에 차이가 있는 것으로 나타났으며($F=13.416$, $p<.001$), 특히 **G1(진지), G3(갈망), G4(고요), G5(환상), G6(행복)**의 감성단어가 사용될수록 평점이 높아짐
- 3) ‘감성과 표지색 일치’: 앞에서 분석한 상관관계분석에 근거하여 G1,3,4,6,7에 해당하는 감성의 도서 표지색이 B가 아닐 경우, 감성과 표지색이 일치한다는 더미변수를 생성하여 분석하였으나 감성과 표지색 일치 여부는 누적판매량 순위에 유의미한 영향을 주지 않는 것으로 나타남

분석 결과

도서 감성 트렌드

분석 4) 도서 감성 트렌드



- 1) 시계열 분석 결과, 연도별 감성 트렌드의 주기성은 없는 것으로 나타남
- 2) 그러나 감성 단어 출현 빈도가 점차 증가하고 있음을 확인할 수 있음
- 3) 눈에 띄는 점은 2014년 유독 G4(고요, 위로, 평안)의 빈도가 늘었는데,
이는 2014년 발생한 세월호 참사에 대한 우울 등을 극복하기 위한 도서 출간이 많았을 것으로 예측

제안

Insight | 기대효과 | 한계

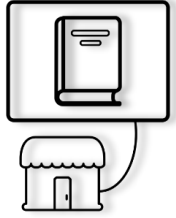




제안

Insight

Insight



문학 콘텐츠에 있어서의 '감성'의 중요성

감성단어의 출현이 판매량과 평점에 양(+)의 영향을 주는 것으로 나타남

-> 문학을 소비하는 목적이 '감성적' 측면에 있는 만큼, 감성에 초점을 둔 마케팅 필요

문학 감성 트렌드의 비주기성

문화 콘텐츠의 트렌드는 경제 등과 같이 일정한 주기성을 띄지 않는 경우가 많음

-> 트렌드를 예측하는데 초점을 두기보단 소비자들의 감성 흐름을 항상 주목하고

이에 영향을 미치는 거시 트렌드를 연구하여 빠르게 대처할 수 있는 '즉시성'을 높여야 함

표지색의 영향

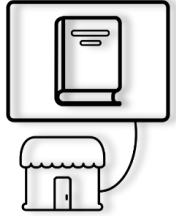
책표지의 감성과 도서의 감성의 일치여부는 판매량/평점에 큰 영향을 주지 않는 것으로 나타났으며, 표지색 그 자체 또한 판매량/평점에 큰 영향을 주지 않는 것으로 나타남

-> 표지 이미지 요소 가운데 '색'을 제외한 어떤 요소가 도서 구매에 영향을 주는지 연구할 필요가 있음



제안
기대효과

기대효과



온라인 서점 최초의 감성기반 추천 시스템

- 감성 기반 추천 시스템 고안을 통해 소비자의 만족도를 높일 수 있음
- 넷플렉스, 멜론 등의 문화 콘텐츠 시스템에서 보듯, 문화 콘텐츠는 소비자의 '취향'을 분석하여 제시해줌으로써 지속적인 구매를 이끌어 낼 수 있음
- 시/희곡과 같은 문학 장르에 있어 분류된 '감성'을 중심으로 효과적인 감성마케팅 가능

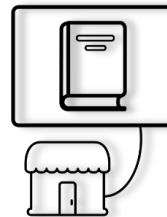
출판 기획에 있어서의 방향 제시

- '감성'이라는 정성적인 영역을 보다 데이터화하여 객관적인 지표 제시
- 문학감성사전을 중심으로 도서를 분석하여 작가들의 포지셔닝을 확인해보고 앞으로의 전략 수립을 제시할 수 있음

제안

한계

한계



문학 감성어휘 사전 분류의 한계점

- 감성사전 분류에 있어 3명의 코더들의 주관이 개입될 수 있으나, 해당 분류에 대한 정확도는 코더 간 신뢰도 이외에는 확인하지 못함
- 감성 어휘 분류에 있어 개별 도서의 맥락적인 의미를 알기 어려워 drop된 단어가 많음 : 추후에는 토픽 모델링 등으로 이를 보완할 수 있는 방법이 필요함
- 총 1,000권의 도서를 기준으로 기준이 되는 감성사전을 구축하였음에도 다양한 어휘를 구사하는 문학작품 특성상 보다 많은 테스트 데이터가 필요했을 것으로 생각

표지색 추출 기준의 한계

- 선행연구를 기반으로 4구간으로 나누어 RGB값을 추출하였으나, 이보다는 단 하나의 주조색을 추출하는 방식이 더 효과적이었을 것으로 생각됨
- 표지 이미지 요소에 있어 표지색 이외에 다양한 요소들을 함께 고려하지 못한 아쉬움

감사합니다.

감성과

문학

출판콘텐츠 마케팅 전략 제안