



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

博士學位論文

# 빅데이터 분석을 통한 모바일 RPG의 개선안 연구

-구글 플레이 국내 상위 3개 RPG를 중심으로-

嘉泉大學校 大學院

IT融合工學科

GAME工學 專攻

廉 東 賢

博士學位論文

# 빅데이터 분석을 통한 모바일 RPG의 개선안 연구

-구글 플레이 국내 상위 3개 RPG를 중심으로-

Research on improvement of mobile RPG using big data  
analysis

-Focusing on high rank three RPG in Korean Google Play-

嘉泉大學校 大學院

IT融合工學科

GAME工學 專攻

廉 東 賢

博 士 學 位 論 文  
指 導 教 授 金 廷 潤

# 빅데이터 분석을 통한 모바일 RPG의 개선안 연구

-구글 플레이 국내 상위 3개 RPG를 중심으로-

Research on improvement of mobile RPG using big data  
analysis

-Focusing on high rank three RPG in Korean Google Play-

위 論文을 GAME工學 博士學位 論文으로 提出함.

2020 年    月    日

嘉泉大學校 大學院

IT融合工學科

GAME工學 專攻

廉        東        賢

이 論文을 廉 東 賢의  
工學博士 學位論文으로 認准함

2020 年    月    日

審査委員長      남 상 훈      ⑩

審 査 委 員      윤 태 복      ⑩

審 査 委 員      은 성 종      ⑩

審 査 委 員      오 석 희      ⑩

審 査 委 員      김 정 윤      ⑩

## 국문 초록

### 빅데이터 분석을 통한 모바일 RPG의 개선안 연구

-구글 플레이 국내 상위 3개 RPG를 중심으로-

가천대학교 일반대학원

IT융합공학과 게임공학전공

염 동 현

지도교수 김 정 윤

본 논문은 빅데이터 분석기법을 통하여 모바일 RPG(Role Playing Game)의 개선안을 연구하였다. 이를 위해 구글 플레이 국내 상위 3개 모바일 RPG 사용자 리뷰를 수집하고 이를 가공하여 유의미한 정보를 추출하였다.

빅데이터로부터 유의미한 정보를 추출하기 위해 구글 플레이 스토어에서 크롤링(Crawling)을 통해 사용자 리뷰를 수집하고, 토큰닝(Tokening) 과정을 거쳐 유의미한 텍스트 데이터를 추출하였다. LDA(Latent Dirichlet Allocation) 토픽 모델링(Topic Modeling)을 통하여 결과를 시각화 하였고, 시각화 결과를 바탕으로 이를 최종 해석하는 과정을 거쳐 주요 토픽을 찾아내었다.

DAU(Daily Active Users)와 매출에 따라 다른 성격의 게임 3개를 선정하여 직접 사례 분석을 진행하였으며 각 게임의 토픽을 분석하였고

게임별 방향성을 제시하였다. 또한 각 게임별 토픽을 분석 비교하여 모바일 RPG의 공통 토픽이 무엇인지 알아보았으며 현재 모바일 RPG의 개발 트렌드와 방향성은 뛰어난 그래픽, 과금 유도가 가능한 콘텐츠를 갖추는 것임을 알 수 있었다. 사용자들은 게임이 가진 고유의 게임성보다는 게임의 그래픽 요소와 보상 및 합리적인 과금에 관심이 많다는 것을 확인하였다.

결론에서는 빅데이터 분석 결과 모바일 RPG 제작에 필요한 주요 요소로써 첫 번째, 그래픽(비주얼)은 게임의 충성도에 큰 영향을 준다는 점. 두 번째, 사용자들은 캐릭터 관련 콘텐츠에 많은 관심을 갖고 있다는 점. 세 번째, 과금의 밸런스는 사용자들이 가장 민감하게 생각한다는 점. 마지막 네 번째, 전투 콘텐츠에 대한 욕구가 강하지 않다는 점을 꼽았으며 이를 바탕으로 모바일 RPG 개발을 개선할 것을 제안하였다.

추가적으로 개선할 사항으로는 빅데이터 수집시 크롤링의 정교함을 개선할 것과 LDA 토픽 모델링의 정확도를 높일 것, 다수의 분석자를 통한 비교를 제안하였다.

**핵심어 : 빅데이터, 모바일 RPG, 토픽 모델링, LDA**

# 목 차

국문 초록 .....	i
목 차 .....	iii
표 목차 .....	vii
그림 목차 .....	viii
 I. 서 론 .....	 1
1. 연구의 배경과 목적 .....	1
1) 연구의 배경 .....	1
2) 연구의 목적 .....	7
2. 연구의 문제 및 범위 .....	9
1) 연구의 문제 .....	9
2) 연구의 범위 .....	10
 II. 관련연구 .....	 14
1. 모바일 RPG 개선방안 연구 .....	14
2. 빅데이터 .....	14
1) 빅데이터의 의미 .....	14
2) 빅데이터의 활용 .....	16
3. 텍스트 마이닝(Text Mining) .....	20
4. 토픽 모델링 .....	21
1) 토픽 모델링 관련 연구 .....	21



2) LDA(Latent Dirichlet Allocation) .....	2
3) 언어 모델의 평가 방법 .....	24
5. 빅데이터 시각화 .....	25

### III. 모바일 RPG의 빅데이터 분석 .....72

1. 연구 방법 .....	27
2. Python을 이용한 빅데이터 수집 .....	28
1) 크롤링(Crawling) .....	28
2) 모듈 설치 .....	29
3. Python을 이용한 전처리 .....	33
1) 전처리(Preprocessing) .....	33
2) 인코딩(Encoding) .....	35
3) 자연어처리(Natural Language Processing) .....	36
4) 모듈 설치 .....	39
4. Python을 이용한 토픽 모델링 .....	40
1) 단어 사전 구축 .....	40
2) LDA 모델 훈련 .....	40
3) 최적의 passes 찾기 .....	41
4) 최적의 num_topics 찾기 .....	41
5. Python을 이용한 시각화(Visualization) .....	42
1) 모듈 설치 .....	42
2) LDA 모델 도식화 .....	43
3) 토픽 분석 .....	45

<b>IV. 실험 및 평가</b>	<b>46</b>
1. 사례분석 : A게임에 대해서	46
1) 게임소개	46
2) 빅데이터 분석 실험	46
3) 시각화 결과	52
4) 토픽 분석	61
2. 사례분석 : B게임에 대해서	63
1) 게임소개	63
2) 빅데이터 분석 실험	63
3) 시각화 결과	69
4) 토픽 분석	76
3. 사례분석 : C게임에 대해서	78
1) 게임소개	78
2) 빅데이터 분석 실험	78
3) 시각화 결과	84
4) 토픽 분석	91
4. 종합 분석	93
1) 분석의 객관성 확보	93
2) 게임별 토픽 분석 비교	93
3) 빅데이터 분석 결과	96
<b>V. 결론</b>	<b>98</b>

참고문헌 .....	101
부록 .....	114
ABSTRACT .....	<b>Ⅱ</b>

## 표 목차

<표 1> tag Package 로딩 시간 비교 .....	37
<표 2> tag Package 실행 시간 비교 .....	38
<표 3> tag Package 문장 해석 비교 .....	38
<표 4> 게임 A의 passes 선정 평가 .....	49
<표 5> 게임 A의 num_topics 선정 평가 .....	51
<표 6> 게임 B의 passes 선정 평가 .....	66
<표 7> 게임 B의 num_topics 선정 평가 .....	68
<표 8> 게임 C의 passes 선정 평가 .....	81
<표 9> 게임 C의 num_topics 선정 평가 .....	83
<표 10> 게임별 토픽 분석 비교 .....	94
<표 11> 각 게임에 대한 방향성 제시 .....	97

## 그림 목차

[그림 1] RPG의 평균 게임플레이 시간 .....	2
[그림 2] App Ape 매출 순위와 DAU 순위 그래프 .....	4
[그림 3] 주로 이용하는 모바일 게임 장르 .....	6
[그림 4] 구글 플레이 스토어 모바일 RPG .....	6
[그림 5] 연구 문제 .....	9
[그림 6] 연구의 범위 .....	10
[그림 7] 연구 흐름도 .....	13
[그림 8] Facebook 맞춤형 광고 예시 .....	17
[그림 9] Amazon 가격 변경 그래프 .....	18
[그림 10] 배송 서비스에서 고객이 원하는 것의 변화 .....	19
[그림 11] LDA 확률적 그래프 모델 .....	23
[그림 12] 구글플레이 스토어의 사용자 리뷰 화면과 HTML 소스 ..	31
[그림 13] 사용자 리뷰의 ‘더보기’ 버튼 .....	32
[그림 14] 사용자 리뷰의 ‘전체 리뷰’ 버튼 .....	32
[그림 15] Passes에 따른 Perplexity와 Coherence 그래프 예시 ..	41
[그림 16] Topic 수에 따른 Perplexity와 Coherence 그래프 예시	42
[그림 17] pyLDAvis를 이용한 Visualization 예시 .....	44
[그림 18] 검은사막 모바일 .....	46
[그림 19] 게임 A의 리뷰 수집 .....	47
[그림 20] 게임 A의 토크닝 처리 .....	48
[그림 21] 게임 A의 passes 선정 평가 .....	50

[그림 22] 게임 A의 num_topics 선정 평가 .....	51
[그림 23] 게임 A의 LDA Visualization 결과 .....	52
[그림 24] 게임 A의 Topic 1에 대한 결과 .....	53
[그림 25] 게임 A의 Topic 2에 대한 결과 .....	54
[그림 26] 게임 A의 Topic 3에 대한 결과 .....	55
[그림 27] 게임 A의 Topic 4에 대한 결과 .....	56
[그림 28] 게임 A의 Topic 5에 대한 결과 .....	57
[그림 29] 게임 A의 Topic 6에 대한 결과 .....	58
[그림 30] 게임 A의 Topic 7에 대한 결과 .....	59
[그림 31] 게임 A의 Topic 8에 대한 결과 .....	60
[그림 32] 리니지M .....	63
[그림 33] 게임 B의 리뷰 수집 .....	64
[그림 34] 게임 B의 토크닝 처리 .....	65
[그림 35] 게임 B의 passes 선정 평가 .....	67
[그림 36] 게임 B의 num_topics 선정 평가 .....	68
[그림 37] 게임 B의 LDA Visualization 결과 .....	69
[그림 38] 게임 B의 Topic 1에 대한 결과 .....	70
[그림 39] 게임 B의 Topic 2에 대한 결과 .....	71
[그림 40] 게임 B의 Topic 3에 대한 결과 .....	72
[그림 41] 게임 B의 Topic 4에 대한 결과 .....	73
[그림 42] 게임 B의 Topic 5에 대한 결과 .....	74
[그림 43] 게임 B의 Topic 6에 대한 결과 .....	75
[그림 44] 에픽세븐 .....	78

[그림 45] 게임 C의 리뷰 수집 .....	79
[그림 46] 게임 C의 토크닝 처리 .....	80
[그림 47] 게임 C의 passes 선정 평가 .....	82
[그림 48] 게임 C의 num_topics 선정 평가 .....	83
[그림 49] 게임 C의 LDA Visualization 결과 .....	84
[그림 50] 게임 C의 Topic 1에 대한 결과 .....	85
[그림 51] 게임 C의 Topic 2에 대한 결과 .....	86
[그림 52] 게임 C의 Topic 3에 대한 결과 .....	87
[그림 53] 게임 C의 Topic 4에 대한 결과 .....	88
[그림 54] 게임 C의 Topic 5에 대한 결과 .....	89
[그림 55] 게임 C의 Topic 6에 대한 결과 .....	90

# I. 서론

## 1. 연구의 배경과 목적

### 1) 연구의 배경

모바일 게임은 넓은 의미로는 모바일 기기(휴대폰, PDA, 휴대용 게임기 등)에서 하는 게임으로 정의된다. 좁은 의미로는 이동전화 단말기를 통하여 서버에 접속하여 이용하는 게임(무선 인터넷 서비스 환경에서 이용 가능한 게임)으로 정의된다[1].

#### (1) RPG(Role Playing Game)

본 연구에서는 모바일 RPG(Role Playing Game)를 대상으로 연구를 진행한다. RPG는 사용자들이 가장 많이 플레이하는 게임 장르중 하나로 2020년 6월 현재 구글 플레이스토어에 등록된 모바일 게임 중 최고 매출 상위 100개 중 56개가 RPG이다[2].

RPG가 다른 장르와 차별화 되는 지점으로는 첫 번째, 가상의 설정에서 사용자가 캐릭터의 역할을 맡는 게임으로 사용자에게 역할을 부여함으로써 그만큼 사용자의 관여와 몰입이 강하게 나타난다는 점[3]. 두 번째, 사용자는 문자적인 행동이나 캐릭터 개발에 관한 체계적인 의사결정 과정을 통해 이야기 내에서 이러한 역할을 수행할 책임이 있다는 점[4]. 세 번째, 긴 플레이시간, 뛰어난 연장성 그리고 강한 중독성이 온라인 플랫폼과 잘 맞아떨어져 지속적인 수익을 창출할 수 있다는 특성이 있다는 점이다[5].



모바일 또는 온라인 RPG는 플레이타임이 긴 대표적 장르이다. 사용자들의 게임 플레이 타임을 수집하는 HowLongToBeat.com에 의하면 온라인 MMORPG인 월드 오브 워크래프트(World Of Warcraft)의 경우 [그림 1]과 같이 평균 플레이타임이 3,000시간에 달하는 것으로 조사되며, 패키지 RPG의 경우에는 최근의 히트작인 젤다의 전설 : 야생의 숨결(The Legend of Zelda : Breath of the Wild)은 약 200시간이 채 안 되는 것으로 조사되었다[6].



[그림 1] RPG의 평균 게임플레이 시간

## (2) 높은 매출 구조를 가진 RPG

사용자와 게임 제작자간에 자동으로 이루어지는 피드백 루프를 성공적으로 만들려면 사용자가 게임과 상호 작용하는 방식을 측정해야 한다. 이 정보는 원격 계측 데이터를 다시 서버로 보내는 소프트웨어의 계측을 통해 수집된다. 이러한 특정 정보를 일반적으로 게임 메트릭(Game Metrics)이라고 한다[7].

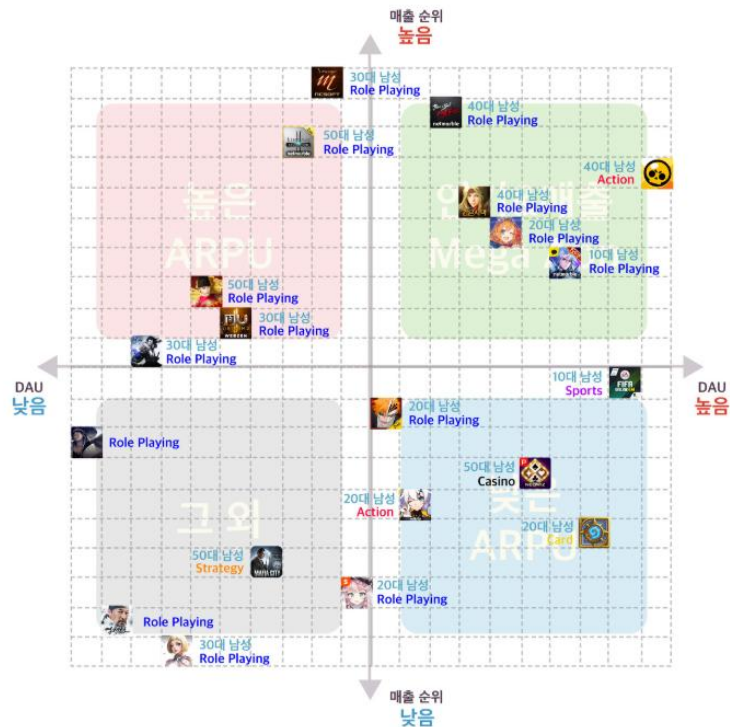
게임 메트릭이 보여주는 여러 지표 중에 DAU(Daily Active Users)와 ARPU(Average Revenue Per User)가 있다. DAU는 특정 날짜에 모바일 앱 또는 웹 제품을 열고 참여하는 총 사용자 수를 나타내는 용어이다. DAU를 통해 제품의 성장률을 측정할 수 있으며 트렌드를 드러내고 사용자 행동을 나타낼 수도 있다[8]. 게임에서는 하루에 게임을 즐기는 실제 사용자 수를 말하는 것으로, 게임을 설치만 하고 플레이하지 않는 사용자는 제외한 수치이다. 따라서 DAU가 높다는 것은 실제로 많은 사용자들이 게임을 자주 즐기고 있다는 의미이다.

무료게임의 경우에는 게임 내에서 추가로 구매하는 아이템이나 게임 이용권 등을 통해서 매출이 발생하기 때문에 꼭 DAU가 높다고 해서 매출까지 높은 것은 아니다. 일반적으로 DAU가 높다는 것은 사용자수가 많다는 것이고, 사용자수가 많다는 것은 유료아이템을 구매할 확률이 높아진다는 것을 의미한다. 이는 DAU가 높은 게임이 매출도 높을 가능성이 높다는 것인데, 현재 모바일 시장에서는 이러한 상식이 통하지 않는다. DAU는 상위권에 있으나 매출 순위는 낮은 경우 혹은 DAU는 낮지만 매출은 높은 경우처럼 DAU와 매출이 서로 다른 양상을 보이는 경우가 많다[9].

DAU가 높더라도 매출이 적다면 이는 게임 개발사 입장에선 효율적이지 못하다. 게임 라이브 서비스(live service)를 위해서는 유지인력 비용, 운영 인력 비용, 고객 응대 인력 비용, 마케팅 비용, 서버 및 네트워크 비용 등의 고정 비용이 발생한다. 어떤 게임이더라도 라이브 서비스를 하게 되면 이 고정 비용이 발생하는 것을 막을 수는 없으므로 고정 비용 대비 더 많은 매출을 내는 게임을 서비스하길 원하게 된다.

게임의 효율성을 나타내는 지표로는 ARPU가 사용되며 사용자당 평균 매출을 말한다. ARPU를 통해 회사의 수익 창출 능력 및 고객별 성장에 대한 분석을 세분화할 수 있다[10]. 게임의 전체 매출을 사용자수로 나누어서 사용자 1인당 얼마의 과금을 하는지 말해주는 것으로, 이 지표가 높을수록 효율적으로 과금요소가 설계된 게임이란 의미이다. 사용자가 적지만 사용자당 평균 매출이 높은 게임에 비해 사용자가 많지만 사용자당 평균 매출이 낮은 게임은 많은 사용자수로 인해 발생하는 고정 비용 대비 매출이 낮아 효율적이지 못하다.

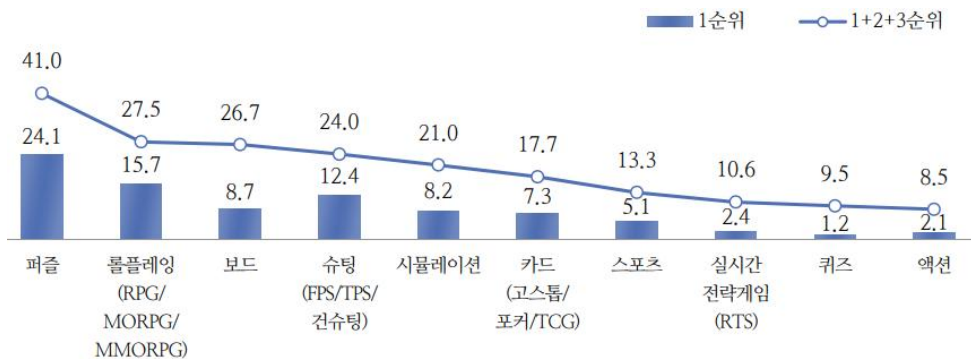
[그림 2]의 App Ape의 ‘2019년 한국 모바일 게임 시장 객단가 분석’에 따르면 2019년 4월 17일 구글플레이 한국 게임 매출 순위 20위권 중 14개가 RPG로 보고되었다[11]. 매출 순위 상위 20개의 게임에 대해서 매출과 DAU의 관계를 보면, 높은 ARPU를 얻은 게임들은 모두 RPG이며, 매출과 DAU가 동시에 높은 인기 게임들 분포에서도 1개를 제외한 게임이 모두 RPG로 분석되었다. 반면, 매출은 높지만 그보다 더 DAU가 높아 ARPU가 낮은 게임들에는 대부분 비RPG가 자리했다.



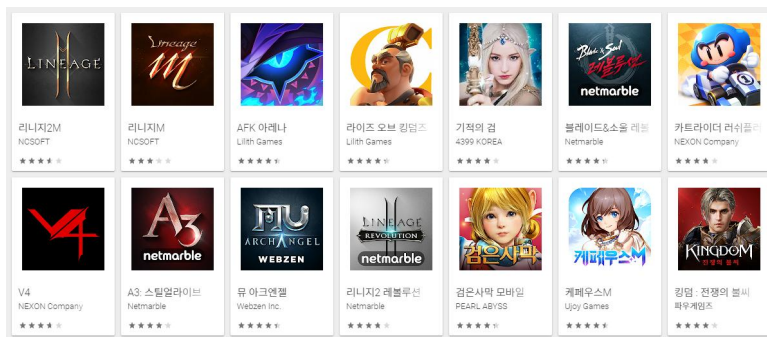
[그림 2] App Ape 매출 순위와 DAU 순위 그래프

### (3) RPG 공급과다로 인한 부작용

많은 게임 개발사들이 매출이 높은 장르의 게임 개발을 선호하면서 RPG의 공급이 많아지고 그에 따라 RPG를 즐기는 사용자수도 증가하는 순환구조가 만들어진다. [그림 3]의 그래프에서 볼 수 있듯이 한국 콘텐츠진흥원이 발간한 2019년 게임이용자 실태조사 보고서의 통계 자료를 통해서 확인이 가능하다[12]. [그림 4]는 2020년 6월 현재 구글 플레이에서 서비스되고 있는 RPG의 목록이다.



[그림 3] 주로 이용하는 모바일 게임 장르



[그림 4] 구글 플레이 스토어 모바일 RPG

하지만 RPG의 매출이 높고 사용자가 많아 비슷한 콘텐츠를 가진 게임들이 양산되는 형태를 보이며 선정적인 광고, 게임의 낮은 질, 과도한 과금 유도 등이 문제로 떠오르고 있다[13]. 양산된 게임이 게임 시장 및 게임 경험에 대해 부정적으로 작용하고 있음을 여러 연구를 통해 알 수 있다.

전영준은 집단 감성과 모바일 게임 사용경험 연구에서 저질 표절 게임 양산과 같이 콘텐츠의 질이 플랫폼에서 충분히 검토되지 않는다는 피드백은 사실상 서비스에 대한 신뢰도를 저하시킬 수 있는 심각한 사안으

로 판단한다고 보았다[14].

기대현은 소셜 네트워크 서비스와 소셜 네트워크 게임 사용자들 사이의 양방향 상호작용 연구에서 ‘저질 표절게임 양산’이라는 키워드가 주로 사용되고 있으며 이를 통해 카카오 게임에 포함된 게임들의 유사성 때문에 카카오 게임을 사용하는 사용자들이 카카오 게임에 대해 부정적인 인식을 가지고 있다고 하였다[15].

정성화의 연구에 의하면 국내 게임이 성장하지 못하는 근본 원인은 변화 없이 기존 게임 형태 그대로 양산되어 출시되는 현 시장의 잘못된 모순과 각종 플레이로 인해 생기는 문제점들이 서비스 실패의 주원인이라고 밝히고 있다[16].

## 2) 연구의 목적

게임 개발사는 과금을 한 사용자에게 비용을 지불하지 않은 사용자들에 비해 더 강한 무기, 더 좋은 장비, 더 빠른 성장과 같은 상대적으로 좋은 게임 환경을 제공하고 있다.

하지만 매출을 위해서 비슷한 게임이 계속 제작되는 현상은 다양성을 추구해야 하는 게임 업계 발전에도 나쁜 영향을 줄 것이며, 나아가서는 사용자들의 게임에 대한 불신으로 인해 게임 산업의 리더십이 미국에서 일본으로 전환되었던 것 같은 아타리 쇼크(Atari Shock)가 재현될 수 있음이 우려된다[17].

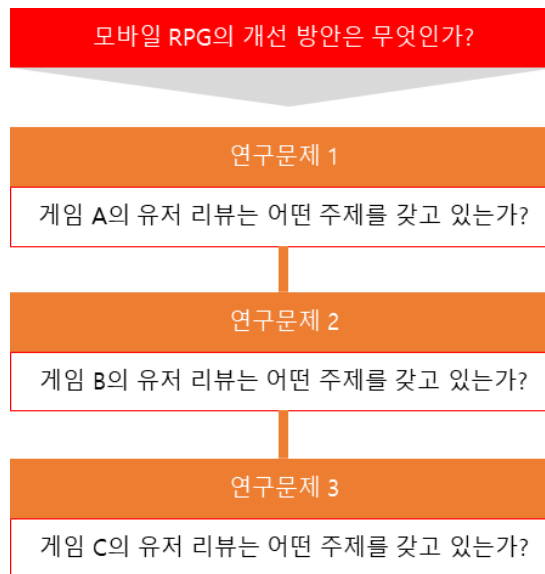
ARPU가 높은 게임을 만드는 것은 수익을 우선으로 하는 기업의 목표이므로 이를 강제할 수는 없다. 따라서 기업과 사용자 모두가 만족할 수 있는 게임 개발을 할 수 있도록 방향을 제시하는 것이 필요하다.

좋은 게임을 보급하기 위해 사용자들이 원하는 것이 무엇인지를 파악하는 것이 중요하다. 이를 위해 사용자들의 목소리를 가장 객관적이고 진솔하게 들을 수 있는 사용자 게임 커뮤니티를 중심으로 사용자들의 생각을 수집하고 분석하기로 하였다.

## 2. 연구의 문제 및 범위

### 1) 연구의 문제

앞서 명시한 연구의 목적은 [그림 5]와 같이 다음 3개의 세부적인 연구문제를 갖는다. 본 연구는 세부 연구문제들에 대해서 사례분석을 실시하고 이를 해석하는 과정으로 구성된다. 세부 연구문제들의 사례분석을 통한 해석이후 각 사례에서 드러난 내용을 종합적으로 결과를 제시한다. 연구의 결론은 각 연구문제별 도출된 결과의 요약으로 제안한다.



[그림 5] 연구문제



## 2) 연구의 범위

본 연구는 [그림 6]의 모바일 게임 장르 중에서 RPG 장르로 제한하여 모바일 RPG의 개선방안을 찾는데 초점을 둔다. 사용자 커뮤니티의 모바일 RPG에 대한 평가를 수집하고 이를 분석하여, 모바일 RPG의 개선안을 도출하는 것이다. 이를 위해 구글 플레이의 사용자 리뷰를 대상으로 하였다. 유효한 사용자 리뷰를 수집하기 위해서는 다수의 사용자를 확보할 수 있는 게임이 필요하며, 일반적으로 매출 순위 상위권에 있는 게임이 DAU가 높다고 볼 수 있으므로 매출 순위 상위권에 속한 모바일 RPG를 대상으로 하였다.



[그림 6] 연구의 범위

또한 비교대상이 될 수 있도록 DAU가 높고 매출도 높은 게임, DAU는 낮지만 매출은 높은 게임, DAU가 높지만 매출이 낮은 게임을 대상으로 하였다. DAU가 낮고 매출이 낮은 게임은 일반적으로 사용자들에

게 선택받지 못한 게임이므로 분석으로 유의미한 결과를 얻을 수 없기에 제외하였다. 본 논문에서는 편의상 각각 게임A, 게임B, 게임C로 통칭하기로 한다.

인터넷의 보급률이 높아짐에 따라 소비자가 상품을 구매하는 방식이 바뀌었다. 오프라인 상점에서 사용자가 구매 결정을 하기 전에 제품을 테스트하고 평가할 수 있지만, 온라인 상점에서는 제품을 직접 평가하는 기능이 훨씬 더 제한적이므로 사용자가 직접 생산한 리뷰에 더 의존하고 있다[18].

인터넷에 산재한 다양한 비정형적인 데이터들 중, 특정 상품들에 대해 개인이 평가한 온라인 리뷰들은 해당 기업에 의미 있는 정보를 제공할 뿐만 아니라 상품에 관심이 있는 잠재적 고객에게도 필요한 정보를 제공해준다[19].

본 연구의 사례 분석을 위해서 사용자들의 게임에 대한 생각이 진정성 있게 표현되어 있는 구글 플레이 스토어의 사용자 리뷰 데이터를 수집하였다.

본 연구의 구체적인 연구 및 분석방법은 다음과 같다.

제I장 서론에서는 연구 배경과 목적, 연구의 문제 및 범위에 대해 기술하였다.

제II장 관련연구에서는 먼저 모바일 RPG 개선방안을 위한 선행 연구를 알아보았고, 빅데이터에 대한 이론과, 빅데이터를 다루기 위한 텍스트 마이닝, 주제를 찾기 위한 토픽 모델링, 마지막으로 빅데이터 분석을 시각화하는 선행 연구에 대해 고찰하였다.

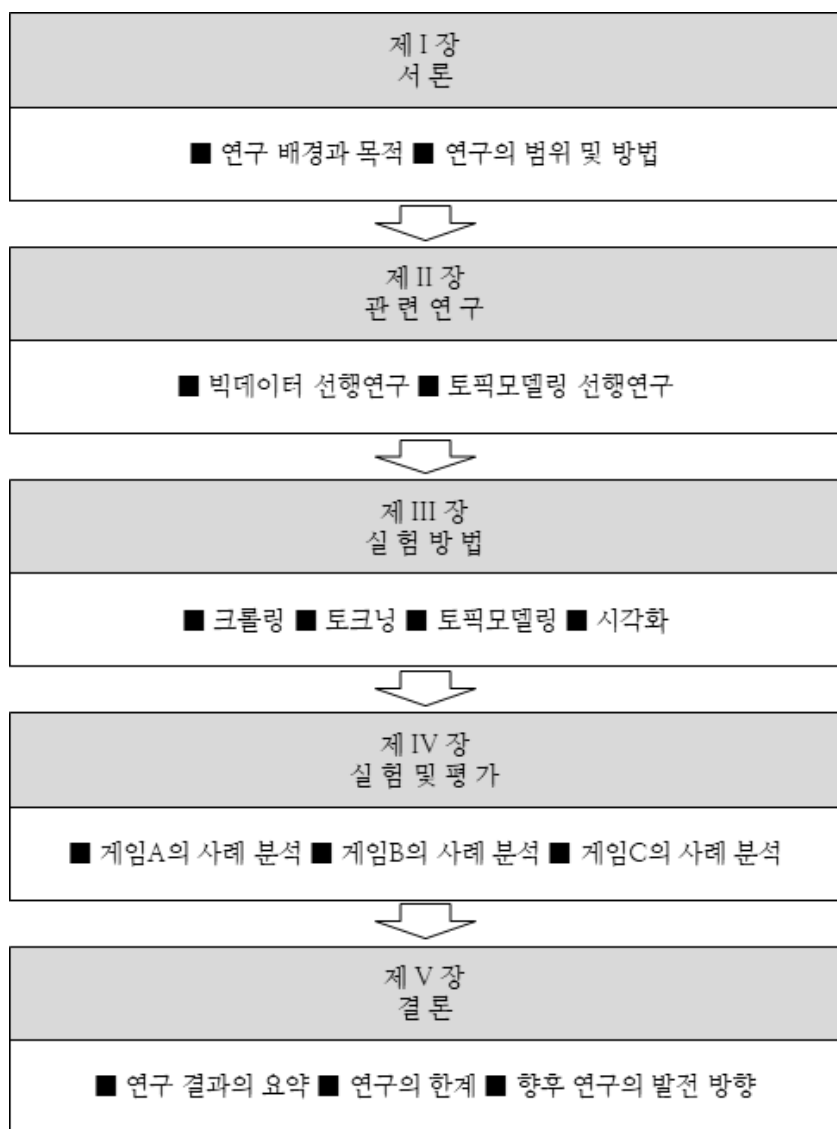
제III장에서는 모바일 RPG의 빅데이터 분석에 대해서 다루었다.

Python을 이용하여 빅데이터의 수집, 전처리, 토픽 모델링 그리고 시각화하는 방법을 기술하였다.

제IV장에서는 3개의 게임에 대해서 각각 실제 사례를 분석하였으며 그 내용을 정리하였다. 종합적으로 각 게임들에 대한 분석을 비교하고 게임별 개선방안을 제시하였다.

제V장에서는 본 연구에서 제안한 분석 결과의 내용, 개선사항과 향후 연구에 대해서 기술하였다.

본 연구의 흐름은 [그림 7]에서 보는 바와 같다.



[그림 7] 연구 흐름도

## II. 관련연구

### 1. 모바일 RPG 개선방안 연구

모바일 RPG의 개선방안에 대한 선행 연구를 알아보았다. 송두헌은 게임별 공식카페를 통해 설문조사를 실시하여 게임플레이의 동기, 주로 사용하는 캐릭터 선정 이유, 월 소비 금액을 조사하였다[20]. 이규복은 자동전투 시스템의 유형과 실제 운용 사례를 통해 문제점 해결방안을 모색하였다[21]. 김승언은 인앱 결제 서비스 이용자들에 대한 설문조사를 통해 사용성 평가를 하였으며[22], 전상현은 차별화된 게임 브랜드를 만들기 위해 인기 RPG 캐릭터를 분석하였다[23]. 전지훈은 게임 로그 데이터로부터 사용자의 행동과 심리정보를 나타내는 속성을 추출하여 분석하였다[24].

### 2. 빅데이터

#### 1) 빅데이터의 의미

빅데이터(Big Data)란 고용량, 속도 및 다양성을 특징으로 하는 정보 자산을 가치로 전환하기 위해 특정 기술 및 분석 방법을 요구하는 것이다.[25] 일반적으로 사람이 처리하기 어려운 정도로 방대한 정형 또는 비정형 데이터 집합에서 의미를 추출하고 결과를 분석하는 기술을 말한다. 이에 따라 세계 각국에서 창의적인 빅데이터 활용이 미래 경쟁력이 핵심 수단으로 인식하고, 국가 차원에서 빅데이터 활성화를 위해 다양한 노력을 기울이고 있다[26]. 세계 각국의 정부와 기업들은 빅데이터

가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하고 있으며, McKinsey, The Economist, Gartner 등은 빅데이터를 활용한 시장변동 예측과 신사업 발굴 등 경제적 가치창출 사례 및 효과를 제시하고 있다[27].

빅데이터 분석 기술을 사용하면 비정형 및 구조화 된 정보를 여러 소스에서 대량으로 처리할 수 있어 특히 사용자 프로필에 대한 정확하고 풍부한 정보를 수집할 수 있어 개인화(Personalization)에 중요하다. 예를 들어, 이러한 데이터는 사용자가 아직 고려하지 않은 요소를 예측하려는 추천 시스템에 적합하다[28].

빅데이터의 특징은 3V로 요약하는 것이 일반적으로 데이터의 양(Volume), 데이터 생성 속도(Velocity), 형태의 다양성(Variety)을 의미한다. 빅데이터는 말 그대로 대용량의 데이터(Volume)를 저장, 트랜잭션, 분류해야 한다. 또한 오늘날 디지털 데이터는 매우 빠른 속도로 생산되므로 자동화, 실시간, 빠른 처리, 스트리밍이 중요하기 때문에 대용량의 데이터를 빠르게(Velocity) 처리하고 분석해야 한다. 마지막으로 디지털 데이터는 텍스트, 음성, 사진, 영상 등 다양하게 존재(Variety)하므로 이것을 다 처리할 수 있어야 한다[29].

빅데이터는 이질적이고 복잡한 데이터 구조를 갖는다. 주로 관계형 데이터베이스로 구성된 기존의 데이터와는 달리 빅데이터 환경에서는 문자, 숫자, 그림 등 다양한 형태의 데이터가 서로 연관되어 한꺼번에 포함되어 있다. 이들 간의 숨겨진 패턴을 찾아내는 것은 빅데이터에서 매우 중요한 작업 중 하나이다. 이와 같은 빅데이터의 특성 때문에 통계학과 기계학습을 이용한 빅데이터의 분석에는 어려움이 따른다[30].

## 2) 빅데이터의 활용

### (1) Facebook

Facebook은 전 세계 월 14억 명이 사용하는 소셜네트워크서비스(SNS) 기업이며 서비스 이름이기도 하다[31]. 사용자로부터 얻어지는 빅데이터를 Facebook이 어떻게 사용하는지 예시를 알아본다.

첫 번째, 얼굴 인식이다. Facebook을 이용하는 사용자들은 자신이 찍은 사진들을 지인과 공유하는 것을 즐긴다. 지인 기반이므로 사진에는 자주 보이는 얼굴들이 찍히기 마련이고, Facebook은 이를 자동으로 인식해 친구의 얼굴을 태그 해준다. 2005년 10월 첫 사용 이후 2009년까지 300억장의 사진이 업로드 되었다[32].

두 번째, ‘좋아요’ 버튼을 통한 사람의 성향 파악이다. ‘좋아요’ 버튼을 통하여 직접적으로는 사용자가 선호하는 대상(인물, 장소, 물건 등)에 대해 알 수 있으며, 경제적인 측면에서 선호하는 브랜드와 취향을 알 수 있게 해주기 때문에 [그림 8]과 같이 개인 맞춤형 마케팅으로 활용할 수 있고, 정치적 의미에 대한 표현으로써 사용되어 사용자의 정치 성향을 알 수 있는 수단이 된다. 이런 맞춤형 광고가 사용자들의 구매 결정에 영향을 끼쳤으며 이에 대해서 사용자가 부정적인 반응을 보이지 않는 것으로 인식되었다[33].

세 번째, ‘좋아요’를 통해 파악한 사용자의 성향에 따라 사용자에게 보여주는 콘텐츠를 제어한다. ‘좋아요’를 많이 누른 지인의 글은 타임라인에 자주 보이게 되며, 한동안 ‘좋아요’를 누르지 않은 지인의 글은 타임라인에서 사라진다. 타임라인 중간에 보이는 광고도 사용자의 나이, 지

역, 성별 등에 따라 맞춤형으로 제공된다. 이렇게 응답에 대한 효과는 개인의 상호작용 파트너와 연결되어 있다고 느끼고 친숙함을 증가시켜 준다[34].



[그림 8] Facebook 맞춤형 광고 예시

## (2) Amazon

Amazon은 빅데이터를 전자상거래(e-Commerce)에 잘 활용하는 기업이다. 특히 미국 내 당일배송/5시간배송 이라는 불가능할 것 같은 일은 빅데이터가 아니면 해결할 수 없었을 것이다[35]. Amazon의 빅데이터 활용과 관련된 예시를 알아본다.

첫 번째, 유동형 가격 정책이다. 고객은 과거에는 웹사이트 방문 횟수와 관계없이 같은 가격의 상품을 접했다. 이제는 빅데이터 플랫폼이 소



비자의 구매 의지를 평가하기 때문에 가격이 자주 바뀐다. 온라인 소매 업체는 오프라인 소매 업체에 비하여 훨씬 자주 가격을 변경하는 것으로 나타났다[36]. [그림 9]에서 보는 것처럼 Amazon은 매일 약 250만 번 상품 가격을 변경한다. Amazon은 빅데이터로 경쟁사 가격부터 자사 재고까지 파악한 후 분석해 제품 가격을 변경할 수 있다[37].



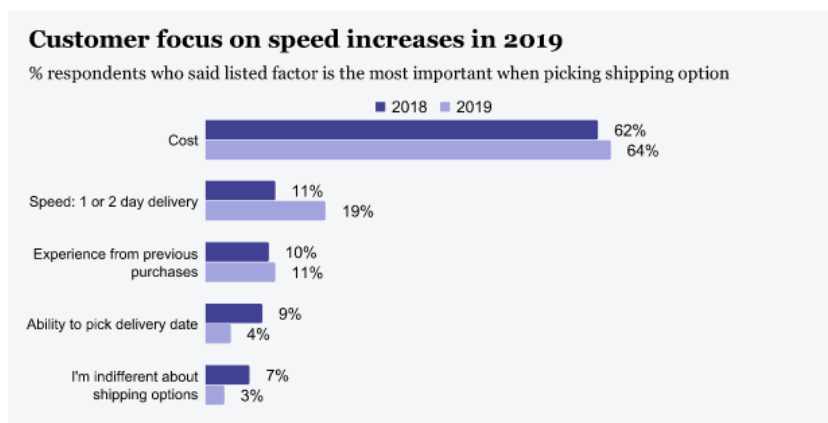
[그림 9] Amazon 가격 변경 그래프

두 번째, Amazon Fraud Detector 시스템을 통해 부정행위를 방지한다. 매년 전 세계에서 온라인 사기 행위로 수백억 달러의 피해가 발생한다. 가짜 계좌를 만들고 훔친 신용 카드로 결제하는 등 악의적인 공격을 받기가 특히 쉽다. 빅데이터, 머신러닝 및 20년간 축적된 Amazon의 부정행위 탐지 전문 지식을 활용하여 부정행위 가능성이 있는 온라인 활동을 자동으로 식별함으로써 이러한 문제를 해결하고 더 많은 부정행위를 신속하게 처리할 수 있다. 이전에 머신러닝과 관련한 경험이 없는 고객도 클릭 몇 번으로 부정행위 탐지 모델을 만들 수 있

다[38].

세 번째, Amazon Personalize는 20년 이상의 추천 경험을 바탕으로 실시간 제공되는 개인화된 제품 및 콘텐츠 추천과 맞춤형 마케팅 프로모션을 통해 고객 참여율을 높일 수 있도록 지원한다. 머신 러닝을 사용하여 웹 사이트 및 애플리케이션을 위한 고품질 추천을 생성하여 몇 번의 클릭만으로 정교한 개인화 기능을 손쉽게 구축할 수 있다[39].

마지막으로 프라임회원제를 대상으로 한 당일배송/5시간배송 정책이다. Supply Chain Dive가 2019년에 조사한 바에 따르면, [그림 10]에 것처럼 상품 구매 시 배송속도를 가장 중요하게 생각한 사람이 2018년 11%에서 2019년 19%로 상승했다[40]. 따라서 Amazon을 비롯한 유통업체들은 배송 속도에 투자를 해서 소비자들을 붙잡으려고 한다.



Matt Leonard / Supply Chain Dive, data from Convey

[그림 10] 배송 서비스에서 고객이 원하는 것의 변화

### 3. 텍스트 마이닝(Text Mining)

본 연구에서는 게임 리뷰를 수집하여 이를 분석한다. 대부분의 인터넷 상에서 수집할 수 있는 리뷰는 자연어 형태의 비정형 텍스트 형태로 존재한다. 정해진 형태가 없기 때문에 비정형 데이터에서 정보를 추출하는 텍스트 마이닝(Text Mining) 기법을 사용해야 한다[41].

Alison은 점점 증가하는 출판 연구에 대해 복잡하고 시간이 많이 걸리는 체계적인 검토를 위해 텍스트 마이닝을 이용하였다[42]. Wei Liu는 콘텐츠 단어를 기능 단어와 효과적으로 분리하고 클러스터 활성화를 위한 알고리즘을 도입하는데 텍스트 마이닝 기술을 사용하였다[43]. Salloum은 ML(Mobile Learning) 연구와 관련된 과학적 테스트의 주제를 식별하기 위해 텍스트 마이닝 기술을 사용하였으며[44], 아랍어 신문의 Facebook 페이지 분석 및 분류를 위해 텍스트 마이닝 기법을 사용하였다[45].

빅데이터 분석은 유의미한 정보를 얻기 위한 과정으로 분석의 목적에 따라 알고리즘과 수학적 처리과정을 거친다. 텍스트 마이닝은 데이터마이닝의 한 분야로 방대한 텍스트 데이터를 기반으로 의미 있는 패턴을 찾아내는 기술이다[46]. 인간의 말은 각 언어별로 어휘적, 문법적 독특성이 있을 뿐 아니라, 그 표현의 형태가 매우 다양하고 복잡하여 일괄적인 규칙으로 규정하기 힘든 경우가 많으며, 언어가 사용되는 환경에 따라 끊임없이 변화하는 특성을 지니고 있다. 이러한 인간 언어 중 문자로 표현된 언어를 컴퓨터로 분석 처리하고 그 구조와 의미를 이해하려는 기술이 자연언어처리(Natural Language Processing)이다[47].

비구조화된 형태의 문서들이 수집되면 전처리 과정을 거쳐서 텍스트

분석이 수월한 형태로 변환된다. 전처리 작업은 텍스트 분석을 위한 알고리즘의 성능을 결정짓는 중요한 요소로서 현재 자연어처리와 머신러닝 등을 응용하여 활발하게 연구되고 있는 분야이다. 전처리를 거친 텍스트 데이터는 컴퓨터가 처리할 수 있는 다양한 방식으로 표현되는데 일반적으로 문서의 단어들이 벡터 공간(Vector Space) 속에서 표현된다[48].

## 4. 토픽 모델링

### 1) 토픽 모델링 관련 연구

수집된 텍스트에서 단어의 빈도수를 추출할 수는 있지만 의미 없는 단어들도 많이 포함되어 있으며, 이 단어들의 집합에서 주요 주제(Topic)를 찾기는 쉽지 않다. 토픽 모델링 알고리즘은 방대한 양의 텍스트에서 사용된 단어들을 분석하여 주제들을 발견하고 그 주제들이 서로 어떻게 연결되어 있는지, 시간에 따라 어떻게 변화되어 가는지를 분석해 주는 통계적 방법이다[49]. 토픽 모델은 정성적, 정량적으로 정확한 것으로 증명되었다[50].

토픽 모델링은 그 출현과 함께 텍스트 마이닝 분야에서 큰 관심을 받아온 연구방법론이다. 특히 Blei가 제안한 LDA(Latent Dirichlet Allocation)는 토픽 모델링 연구에 있어 표준 도구로 자리 잡은 알고리즘이다. 소셜미디어는 공통된 관심사 혹은 사회적인 이슈를 공유하는 이용자들의 연결로 이루어져 있다는 점과 연일 생산되는 방대한 양의 텍스트 데이터 때문에, 다양한 함의를 가질 수 있는 잠재적인 정보를

발굴하기 위한 연구의 장으로 각광받아 왔다[51]. 생성 모델은 실제 문헌을 작성하는 과정으로 보고 문헌을 작성하기 위해 각 문헌에 어떤 주제들을 포함시킬 것인지, 또 그에 따라 어떤 단어들을 어떤 주제에서 선택하여 배치할 것인지를 각각의 파라미터로 모델링한다[52].

Chong Wang은 과학 기사 추천을 위해 협업 토픽 모델링을 사용하였으며[53], AC Calheiros는 호텔의 정서가 고객에게 어떤 감정을 유발시키는지 확인하기 위해 토픽 모델링을 사용하였다[54]. Lijun Sun은 토픽 모델링을 이용하여 운송 연구의 주제와 트렌드를 연구하였으며[55], N Pingclasai는 토픽 모델링을 사용하여 버그 및 기타 요청에 대한 버그 리포트 분류에 대해 연구하였다[56].

## 2) LDA(Latent Dirichlet Allocation)

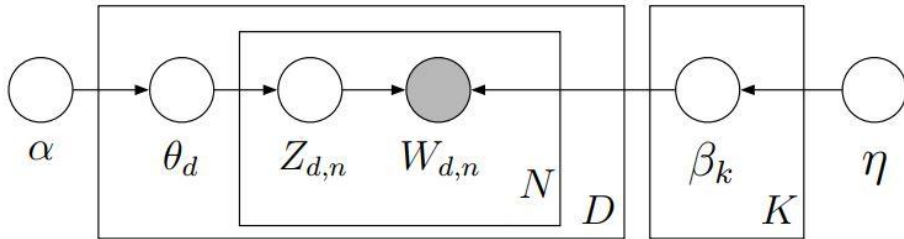
LDA는 Blei가 2003년 처음 발표하였다[57]. 그에 따르면 토픽의 분포와 토픽을 구성하는 단어들의 분포가 디리클레 분포(Dirichlet Distribution)를 따르는데 실수 벡터 중 벡터의 요소가 양수이며 모든 요소를 더한 값이 1인 경우에 대해 확률 값이 정의된다.

자연어처리에서 LDA 토픽 모델링은 가장 많이 사용되는 방법 중 한 가지이다. LDA는 단어 확률로 토픽을 나타내는데 각 토픽에서 확률이 가장 높은 단어는 일반적으로 어떤 토픽을 나타내는지 좋은 아이디어를 제공한다[58].

Kai Tian은 소프트웨어의 자동 분류를 위해 LDA를 이용하였으며[59], Tirunillai는 빅데이터의 전략적 브랜드 분석을 하여 온라인 채터의 마이닝 마케팅을 연구하기 위해 LDA를 이용하였다[60]. Bolelli는 텍스

트 컬렉션(collection)에서 토픽과 트렌드를 감지하는 연구를 위해 LDA를 이용하였으며[61], Somasundaram은 버그 리포트의 자동 분류를 위해 LDA를 이용하였다[62].

LDA 알고리즘은 생성모델로써 문서 내의 숨겨져 있는 토픽들을 찾아내는 알고리즘이다. 생성모델은 실제 문서를 작성하는 과정으로 보고 문서를 작성하기 위해 각 문서에 어떤 토픽들을 포함 시킬 것인지, 또 그에 따라 어떤 단어들을 어떤 토픽에서 선택하여 배치할 것인지를 각각의 파라미터로 모델링한다. 따라서 문서, 단어 등 관찰된 변수(observed variable)를 통해 문서의 구조와 같은 보이지 않는 변수(hidden variable)를 추론하는 것을 목적으로 하며 결과적으로 전체 문서 집합의 토픽들과 각 문서별 토픽 비율, 각 단어들이 각 토픽에 포함될 확률들을 알아낼 수 있다. 단어가 독립적이지 않다는 가정에서 단어를 생성하는 조건에 따라 사후확률을 추론한다[63]. 이러한 관계는 [그림 11]과 같은 확률적 그래프 모델로 표현할 수 있다[64].



[그림 11] LDA 확률적 그래프 모델

토픽은  $\beta_{1:K}$ 이며, 각각의  $\beta_k$ 는 단어의 분포이다.  $d$ 번째 문서의 토픽 비율은  $\theta_d$ 이며,  $\theta_{d,k}$ 는 문서  $d$ 의 토픽  $k$ 에 대한 토픽 비율이다.  $d$ 문서의

토픽 지정은  $z_d$ 이며,  $z_{d,n}$ 은 문서  $d$ 의  $n$ 번째 토픽 할당으로 고정 단어로 부터 얻을 수 있다[64].

### 3) 언어 모델의 평가 방법

토픽 모델에 의해 학습된 토픽은 종종 유용해 보이지만 특정 주제가 없기 때문에 최종 사용 시에는 덜 유용할 수 있다. 실제 문제를 해결하기 위해 토픽 모델링을 적용하려면 필요 없는(junk) 토픽을 쉽게 식별할 수 있는 방법이 필요하다[65].

LDA는 일반적으로 문서 분류 또는 정보 검색과 같은 보조 작업을 통해 성능을 측정하거나 훈련용 문서의 확률을 추정하여 평가된다. 평균적으로 더 나은 모델일수록 일반화된 문서의 확률 가능성이 높다[66].

학습된 토픽을 평가하기 위한 가장 초기의 작업은 혼란도(Perplexity) 결과를 기반으로 수행된다. 훈련용 문서의 모음에서 모델을 학습한 다음 보이지 않는 테스트 문서의 로그 확률을 해당 학습 모델을 사용하여 계산한다. 일반적으로 혼란도는 단어 당 기하 평균 가능성의 역수이다. 혼란도는 한 모델이 다른 모델보다 유리하다는 것을 보여주는 표준 방법이다[67]. 이 수치는 특정 확률모델이 실제로 관측되는 값을 얼마나 잘 예측하는지를 평가할 때 사용한다.

혼란도 값이 낮다는 것은 학습이 잘되었다는 의미이지, 그 결과가 사람이 해석하기에 좋다는 것은 아니다. Chang은 혼란도 측정은 예측 모델을 평가하기에 유용하지만 토픽 모델링의 더 나은 탐색 목적을 달성할 수는 없다고 밝히고 있다[68].

Newman은 WordNet, Wikipedia 및 Google 검색 엔진과 같은 리소

스 및 온톨로지 유사성에서 링크 중복 및 용어 동시 발생에 이르는 방법을 기반으로 토픽 일관성(Coherence)을 예측하는 모델을 제안하였다. 두 개의 서로 다른 데이터 세트에서 배운 토픽에 대해, 일관된 토픽이 무엇인지에 대한 주석 간 상호 합의가 있음을 보여주며, 또한 위키피디아 기반의 방법은 토픽 일관성의 평가에 대해 인간과 거의 완벽한 합의를 달성 할 수 있음을 보여준다고 하였다[69].

## 5. 빅데이터 시각화

빅데이터 분석의 주요 측면 중 하나는 거대한 데이터 세트에서 흥미로운 패턴을 찾을 수 있다는 것이지만, 실제로 분석 결과는 원시(raw) 숫자에 불과하며 이 숫자로 해석을 하는 것은 매우 어렵다. 하지만 이런 숫자가 시각적으로 표현되면 우리의 뇌는 의미 있는 패턴을 찾고 그에 따른 결정을 내리기 쉬워진다[69].

빅데이터 처리는 결코 쉬운 일이 아니며 특별한 방법과 접근 방식이 필요하다. 그래픽 사고는 인간에게 매우 단순하고 자연스러운 유형의 데이터 처리이므로 이미지 데이터 표현은 효과적인 방법이 될 수 있으며, 데이터 이해를 용이하게 하고 의사 결정에 충분한 지원을 제공할 수 있다[70].

빅데이터 시각화를 위해 많은 툴이 제공되고 있는데 용도에 맞는 툴을 이용하여 시각화에 도움을 줄 수 있다. FineReport는 이러한 시각화 기술에 대해서 2020년 오픈 소스 데이터 시각화 툴 10개를 선정하였다[71]. 빅데이터의 분석 결과를 시각화하는 것이 해석에 도움을 줄 수 있지만, 시각화 자체가 해석의 결과물이 되는 것은 아니다. 빅데이터



분석에는 시각화를 통한 결과물을 사람이 직접 해석하는 것이 필요하다.

Ekaterina는 5개의 시각화 기술을 설명하였다[72]. 첫 번째는 2D/3D 표준 이미지로 막대 바(bar)나 라인, 그래프, 차트 등을 표현한다. 두 번째는 기하학적 표현으로 다차원의 데이터 세트를 다룰 수 있어 수학적 통계를 표현하기에 적합하다. 세 번째는 아이콘으로의 표현으로 데이터의 특정 특성에 따라 텍스처 패턴을 바꿔 표현할 수 있다. 네 번째는 픽셀 표현으로 각 숫자 값을 컬러 픽셀로 표시하고 일부 값들을 병합하여 많은 양의 데이터를 시각화하기에 좋다. 마지막 다섯 번째는 계층적 이미지로 트리 맵(Tree map) 형태로 데이터를 계층 구조화하여 보여주기 좋다.

### III. 모바일 RPG의 빅데이터 분석

#### 1. 연구 방법

본 연구를 위해서는 모바일 RPG 사용자들의 게임 리뷰에서 유의미한 주제를 추출해야 한다. 이를 위해 몇 단계의 과정을 거치며 빅데이터를 가공해야 한다.

첫 번째 과정은 사용자들의 게임 리뷰를 수집하는 것이다. 사용자들이 작성한 게임 리뷰는 특정한 웹 페이지에 정형화된 방법으로 게시가 되어 있으며, 웹 크롤링을 통해 이를 수집하는 것이 가능하다.

두 번째 과정은 웹 크롤링을 통해 수집한 텍스트 데이터를 컴퓨터가 분석할 수 있는 형태의 수치 데이터로 변환하는 것이다. 텍스트 데이터를 수치 데이터로 변환하는 과정에서 불필요한 단어를 걸러내는 전처리(Preprocessing) 과정이 필요하다. 전처리 과정을 거친 텍스트 데이터는 수치로 변환하는 인코딩 과정을 거친다. 인코딩을 거친 텍스트 데이터는 자연어 처리 과정을 거쳐 의미 있는 글자만 남겨지게 된다.

세 번째 과정은 글자 집합에서 주제를 뽑아내기 위해 토픽 모델을 만드는 것이다. 이를 위해 LDA 알고리즘을 이용하여 토픽 모델을 구성하고 성능 평가를 실시하여 최적의 토픽 모델을 찾아낸다.

마지막으로 최적의 토픽 모델을 이용하여 위에서 추려낸 글자들의 집합에서 주제를 뽑아내는 것이다. 토픽 모델의 결과는 해석하기 용이하도록 시각화한다. 시각화 데이터를 통해 사용자들이 게임 리뷰에서 주로 이야기하는 주제가 무엇인지 추론해 낸다.

본 연구에서는 Python을 이용하여 연구를 수행하였다. Python은 가장 숙련된 프로그래머뿐만 아니라 초보자도 쉽게 액세스하고 사용할 수 있는 프로그래밍 언어이다. 크고 작은 프로젝트에 모두 사용할 수 있으며 적응력이 뛰어나고 효율적인 프로그래밍 기능으로 널리 사용되고 있다. 궁극적으로 생산성과 효율성을 높이는데 도움이 된다[73]. 또한 학습 곡선이 낮고 구문과 의미가 투명하고 문자열 처리 기능이 우수하다. 통역(interpreted) 언어로써 대화식 탐색을 용이하게 한다[74].

## 2. Python을 이용한 빅데이터 수집

### 1) 크롤링(Crawling)

크롤링은 웹(Web)상에 존재하는 콘텐츠를 수집하는 작업으로 인터넷상의 데이터를 가져와서 파싱하고 필요한 데이터만 추출하는 기법이다. 이런 작업을 위한 프로그램을 크롤러(Crawler)라고 부르며 웹 페이지를 대량으로 처리한다[75].

김성민은 웹 크롤링과 데이터 마이닝(Data Mining)을 사용하여 소규모 비즈니스 도메인 지식 자동 검색을 연구하였다[76]. Linares는 트위터(Twitter) 소셜 네트워크 서비스의 트윗을 웹 크롤링하여 관광 교통의 예측 도구를 구축하였다[77]. Musale은 프로필 매칭(Profile Matching)과 웹 크롤링을 이용하여 직업 추천 시스템을 연구하였다[78].

웹 크롤링에서의 어려운 점은 매일 많은 양의 페이지가 추가 및 변경, 삭제된다는 점이다[79]. 크롤러는 인터넷에 있는 웹페이지를 방문해서

자료를 수집하는 일을 하는데 한 페이지만 방문하는 것이 아니라 그 페이지에 링크된 또 다른 페이지를 차례대로 방문하여 엄청난 분량의 웹 문서를 검색하는데 사람이 이를 수행하는 것은 매우 비효율적이다.

웹은 기본적으로 HTML로 표현되며 HTML 내에 어떤 정형화된 형태로 관리된다. 따라서 정형화된 규칙을 분석해서 우리가 원하는 정보들만 뽑아오는 것이 가능해진다[80].

## 2) 모듈 설치

Python을 이용해 크롤링을 수행하기 위해 모듈을 설치한다.

### (1) requests 설치

HTML 페이지의 정보를 가져오기 위해서 requests 라이브러리를 설치한다. HTML 페이지를 불러와 텍스트 형태로 저장할 수 있다.

```
pip install requests
```

### (2) BeautifulSoup 설치

BeautifulSoup은 HTML 코드를 Python에서 사용할 수 있는 객체 구조로 변환할 수 있도록 Parsing 하는 역할을 한다. bs4 라는 라이브러리를 통해서 설치하는 방법이 추천된다[81].

```
pip install bs4
```

### (3) Selenium 설치

Selenium은 주로 웹앱(WebApp)을 다루는 프레임워크로 웹드라이버 API를 통해 크롬(Chrome) 브라우저를 제어한다. 구글 플레이 스토어는 동적 홈페이지로 플레이 리뷰의 양이 많아 한 번에 다 로드하지 못하고 사용자의 요청에 맞춰서 그때그때 동작하게끔 설계되어 있다. 따라서 일반적인 방법으로 크롤링을 시도하게 되면 화면에 보이는 리뷰의 일부만 수집할 수밖에 없다. Selenium은 직접 브라우저를 컨트롤 할 수 있으므로 숨겨져 있는 모든 리뷰를 꺼낼 수 있게 된다[82].

```
pip install selenium
```

### (4) 크롬 웹드라이버 설치

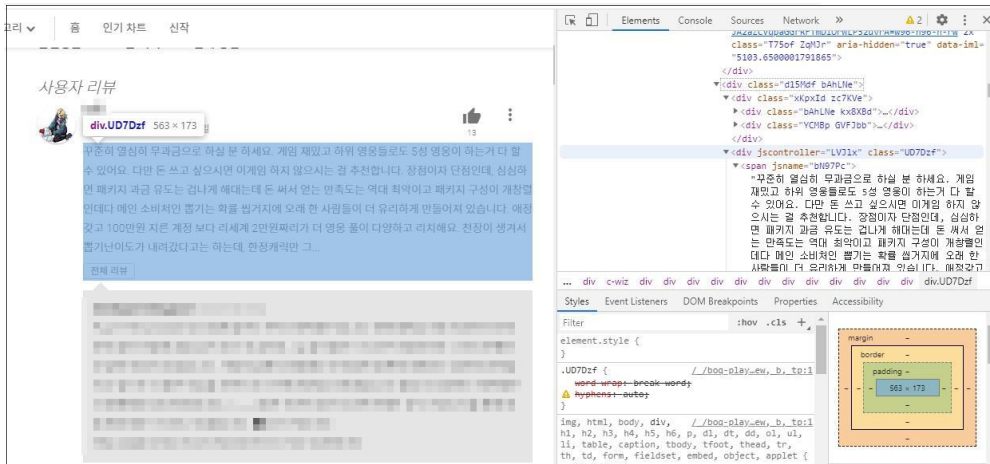
Selenium이 작동하기 위해서는 Chrome 브라우저를 위한 웹드라이버를 설치하여야 한다. 설치된 버전의 구글 크롬에 맞는 웹드라이버를 다운받아 설치한다.

### (5) HTML 소스 파싱

BeautifulSoup으로 가져온 HTML 소스를 파싱하기 위해 BeautifulSoup에서 제공하는 select를 이용한다. select는 CSS Selector의 조건과 일치하는 모든 객체들을 List 형태로 가져온다.

크롬 브라우저를 이용하여 구글 플레이를 접속하여 [그림 12]에서처럼 사용자 리뷰 화면에서 개발자모드(F12키를 누르면 개발자 페이지가 열린다.)를 들어간다. 크롤링 하고자 하는 데이터의 CSS를 찾아 class

값을 가져온 후 BeautifulSoup의 find\_all 함수를 이용하여 class 값이 일치하는 모든 데이터를 수집하여 List로 저장한다.

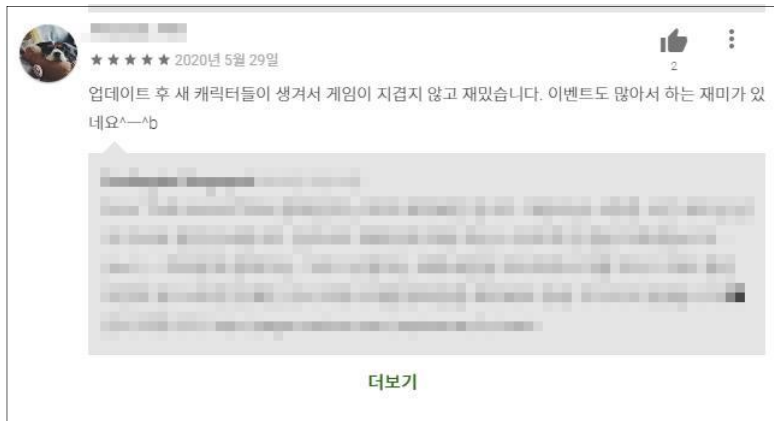


[그림 12] 구글플레이 스토어의 사용자 리뷰 화면과 HTML 소스

## (6) ‘더보기’ 제어

구글 플레이 스토어에는 리뷰를 더 보고 싶을 때 두 가지 방법이 존재한다. 첫 번째 방법은 화면을 아래로 스크롤 하는 것인데 계속 화면을 스크롤 하면 더 이상 리뷰를 볼 수 없게 된다. 이때 화면 하단에 [그림 13]과 같은 ‘더보기’ 버튼이 나오게 된다.

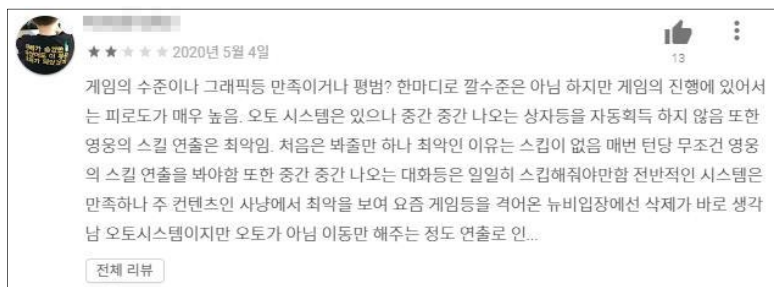
첫 번째 방법을 위해서는 Selenium으로 브라우저를 컨트롤해 윈도우를 스크롤시키면 된다. 두 번째 방법은 ‘더보기’ 버튼의 CSS(Cascading Style Sheet)를 이용해 Selenium으로 버튼을 클릭시키면 된다.



[그림 13] 사용자 리뷰의 ‘더보기’ 버튼

## (7) ‘전체 리뷰’ 제어

구글 플레이 스토어에서는 사용자의 리뷰 양이 많으면 자동으로 숨김 처리가 돼서 리뷰의 일부만 보이게 된다. 이 경우 파서(parser)가 숨겨진 리뷰와 전체 리뷰를 중복해서 가져오게 되므로 [그림 14]에서와 같이 ‘전체 리뷰’ 버튼이 있다면 이를 처리해 주어야 한다. 숨겨진 리뷰와 전체 리뷰가 모두 존재할 경우 전체 리뷰를 가져오도록 한다.



[그림 14] 사용자 리뷰의 ‘전체 리뷰’ 버튼

### (8) 제작사 답글 제거

구글 플레이 스토어의 사용자 리뷰에는 게임을 제작한 제작사가 답글을 달 수 있게 되어 있다. 본 실험을 위해서는 사용자의 리뷰만 필요하므로 제작사의 답글을 제거하도록 한다.

### (9) 특수문자 제거

추후에 단어 위주로 토큰을 잘라서 빈도 측정을 해야 하므로 사용자들의 리뷰에서 특수문자를 제거해 준다. 한글과 숫자, 영문자, 공백을 제외하고 모든 문자는 삭제한다.

### (10) 파싱 데이터 저장

파싱한 데이터는 텍스트 파일로 저장한다.

## 3. Python을 이용한 전처리

### 1) 전처리(Preprocessing)

크롤링으로 얻어낸 말뭉치 데이터를 필요에 맞게 토큰화(Tokenization), 정제(Cleaning), 정규화(Normalization) 등의 전처리를 한다.

토큰화는 일련의 입력 문자 섹션을 구분하고 분류하는 프로세스이다[83]. 문법적으로 더 이상 나눌 수 없는 언어요소를 뜻하며 주어진 말뭉치로부터 토큰(Token)을 분리하는 작업을 토큰화라고 부른다[84].

아래의 입력에서 특수문자를 제외시켜 토큰화를 하게 되면

입력: I am a student. You are a girl!



아래와 같은 출력을 얻게 된다.

출력 : "I", "am", "a", "student", "You", "are", "girl"

영어에서는 띄어쓰기를 기준으로 잘라내서 간단하지만, 한글은 영어와는 달리 조사, 어미 등을 붙여서 사용하는 교착어이기 때문에 띄어쓰기 만으로는 토큰화를 하게 되면 의미가 사라진다. 따라서 한글 토큰화에서는 뜻을 가진 가장 작은 말인 형태소(Morpheme)로 토큰화를 해야 한다. 형태소는 일정한 음성에 일정한 뜻이 결합되어 있는 말의 가장 작은 단위이다. 즉, 최소의 유의적 단위 (minimum meaningful unit)로 정의된다[85].

한글 철자 및 띄어쓰기 검사를 위해 고려해야 할 한국어의 특성은 다음과 같다[86].

- ✓ 음절의 특징으로서 한글은 모아쓰기 특징을 가진다. 즉 자음과 모음이 조합되어 단음절을 형성한다.
- ✓ 띄어쓰기가 일정하지 않고 복잡하다. 따라서 띄어쓰기 오류를 범하기 쉽다.
- ✓ 한국어는 첨가어로서 각 낱말의 의미 변화에 의해 문장의 성분을 결정하며, 첨용과 활용이 자유롭다.
- ✓ 한국어는 불규칙 현상 및 음운 현상이 발달하였다.
- ✓ 한글과 한문 및 영문을 혼용하여 사용하는 경우가 많다.

정제(Cleaning) 작업은 토큰화를 하는 과정 중에 불필요한 데이터를 제거하는 것을 말한다. 특수기호 같은 것들이 대표적으로 제거해야 할 데이터에 속한다.

정규화(Normalization)는 다른 단어이지만 의미가 같은 것들을 한데 묶는 작업을 말한다. 예를 들어, ‘너’, ‘니’, ‘당신’은 모두 상대를 지칭하는 명사이므로 ‘너’라는 단어로 통합할 수 있다.

불용어(Stopword) 처리는 유의미한 단어 토큰만을 선별하기 위해서 의미가 없는 단어 토큰을 제거하는 작업을 말한다. 예를 들어, 감탄사나 1인칭 대명사, 조사 등은 분석을 위해서는 크게 필요하지 않다.

## 2) 인코딩(Encoding)

컴퓨터는 문자를 처리하기 위해 이를 숫자로 바꾼다. 이를 위해 자연어 처리에서는 여러 가지 기법들이 있다. 많은 통계학습 알고리즘은 숫자를 입력으로 요구한다. 원-핫 인코딩(One-Hot Encoding)은 간단하고 널리 사용되는 인코딩 방법이다[87].

Granda는 유기 합성 로봇을 제어하기 위한 연구에서 신경망을 구축하고 머신 러닝을 위한 문헌 데이터를 인코딩하기 위해 원-핫 인코딩을 이용하였으며[88], 김광호는 원-핫 인코딩을 사용하여 딥 러닝 기반 단기 부하 예측 모델을 연구하였다[89].

원-핫 인코딩은 단어 집합(Vocabulary)을 만들 때 중복을 허용하지 않는 단어를 모아놓은 것이다. 이후 단어에 벡터 표현을 하는데 해당하는 단어에는 1의 인덱스를 부여하고, 해당하지 않으면 0의 인덱스를 부여한다. 이렇게 표현된 벡터를 원-핫 벡터(One-Hot vector)라고 한다.

원-핫 인코딩은 다음의 두 가지 과정을 거쳐서 완성된다.

- (1) 각 단어에 고유한 정수 인덱스를 부여한다.
- (2) 해당하는 단어에 1의 인덱스를 부여하고, 해당하지 않으면 0의 인덱스를 부여한다.

원-핫 인코딩의 한계는 단어의 개수가 늘어날수록, 벡터를 저장하기 위해 필요한 공간이 계속 늘어난다는 점으로 저장 공간 측면에서는 매우 비효율적인 표현 방법이다.

단어 주머니(Bag of Words, BoW) 모델은 커다란 단어 주머니가 있다는 개념이다. 단어들의 순서는 고려하지 않고 단어들의 출현 빈도에만 집중하는 텍스트 데이터의 수치화 표현 방법이다.

BoW는 다음 두 가지 과정을 거쳐서 완성된다.

- (1) 각 단어에 고유한 정수 인덱스를 부여한다.
- (2) 각 인덱스에 단어 토큰의 등장 횟수를 기록한 벡터를 만든다.

Bow는 각 단어가 등장하는 횟수를 벡터로 표현한 것이기 때문에, 단어의 빈도에 따라 문서의 내용을 파악하는데 쓰인다. 많이 사용된 단어들을 묶으면 해당 문서의 토픽을 파악하는데 도움이 된다.

### 3) 자연어처리(Natural Language Processing)

자연어처리(Natural Language Processing, NLP)는 텍스트에서 의미 있는 정보를 분석, 추출하고 이해하는 일련의 기술 집합이다. KoNLPy는

한국어 정보처리를 위한 Python 패키지이며 오픈소스로 제공되고 있다[90][91].

한글의 형태소를 분석하기 위해서 KoNLPy는 5가지의 tag Package를 제공한다. 5가지의 태그 패키지의 성능 비교 결과는 다음과 같다[92].

### (1) 시간 비교

tag Package별 클래스를 로딩하는 시간을 비교한 결과가 <표 1>에 정리되어 있다.

<표 1> tag Package 로딩 시간 비교

tag Package	로딩 시간(/seconds)
Kkma	5.6988
Komorán	5.4866
Hannaum	0.6591
Twitter(Okt)	1.4870
Mecab	0.0007

10만자의 문서를 대상으로 각 클래스의 pos 메소드를 실행하는데 소요되는 시간으로 결과가 <표 2>에 정리되어 있다.

<표 2> tag Package 실행 시간 비교

tag Package	실행 시간(/seconds)
Kkma	35.7163
Komoran	25.6008
Hannaum	8.8251
Twitter(Okt)	2.4714
Mecab	0.2838

<표 3> tag Package 문장 해석 비교

Hannanum	Kkma	Komoran	Mecab	Twitter
아버지가방에 들어가 / N	아버지 / NNG	아버지가방에 들어가신다 / NNP	아버지 / NNG	아버지 / Noun
이 / J	가방 / NNG		가 / JKS	가방 / Noun
시ㄴ다 / E	에 / JKM		방 / NNG	에 / Josa
	들어가 / VV		에 / JKB	들어가신 / Verb
	시 / EPH		들어가 / VV	다 / Eomi
	ㄴ다 / EFN		신다 / EP+EC	

## (2) 문장 해석 비교

샘플 문장 “아버지가방에들어가신다”를 비교하여 성능 검증을 하였다. 제대로 된 문장 해석은 “아버지가 방에 들어가신다”로 분석되어야 한다. <표 3>에 비교 결과를 정리하였다[93].

성능 비교상 가장 우수한 결과물을 낸 것은 Mecab이고, 이어서 Kkma와 Twitter가 우수한 성능을 보여주었다. Hannanum과 Komoran은 문장을 거의 해석하지 못하였다.

## (3) 최종 결과

시간 비교와 문장 해석 비교를 통해서 가장 우수한 성능을 보여준 tag Package는 Mecab이었다. 다만, KoNLPy에 따르면 Mecab은 윈도우 OS 환경에서 작동하지 않는다고 밝히고 있다. 본 실험 환경이 Windows 10 환경이었기 때문에 다음으로 성능이 좋은 Twitter를 사용하기로 한다. Kkma의 경우 문장 해석력은 Twitter와 비슷한 성능을 보여주었으나 실행 속도에서 큰 차이가 있다.

## 4) 모듈 설치

Java를 설치하고 설치된 폴더를 환경변수에 등록한다. 한국어의 형태소를 분석하기 위해서 KoNLPy의 tag Package중 Twitter(Okt)를 이용한다. 형태소 중에서 유의미한 단어만 골라내기 위해 명사와 동사만을 추리는 것으로 결정한다. Twitter(Okt) tag에서는 명사는 ‘Noun’, 동사는 ‘Verb’로 표현된다. 필요 없는 형태소는 모두 제거한다.

### (1) JPyPe1 설치하기

```
pip install JPyPe1
```

### (2) KoNLPy 설치하기

```
pip install konlpy
```

## 4. Python을 이용한 토픽 모델링

### 1) 단어 사전 구축

Gensim은 Python으로 구현된 토픽 모델링용 라이브러리로 LDA 알고리즘을 제공한다.

```
pip install gensim
```

### 2) LDA 모델 훈련

Gensim에서 LDA 모델을 만들기 위한 중요 파라미터로 두 가지를 선정하였다.

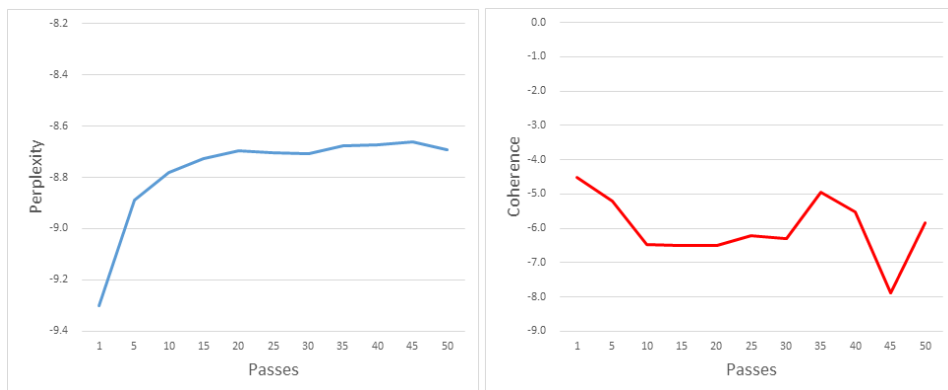
- (1) num\_topics - 가설로 세운 적절한 토픽의 개수
- (2) passes - LDA 모델 학습 횟수

두 개의 파라미터를 조절하여 혼란도(Perplexity) 점수와 일관성

(Coherence) 점수를 측정하여 좋은 LDA 모델이 만들어졌는지 확인할 수 있다. 가장 좋은 평가가 나온 파라미터를 정하여 최종 LDA 모델을 완성한다.

### 3) 최적의 passes 찾기

LDA 모델 파라미터 중 num\_topics를 10으로 가정하고, passes를 1부터 50까지 5의 배수로 단계를 구분하여 테스트 하여 혼란도 점수와 일관성 점수를 측정하면 [그림 15]의 그래프 결과를 얻을 수 있다. 그래프를 분석하여 가장 좋은 점수를 기록한 passes를 최종 LDA 모델 파라미터 값으로 지정한다. 테스트 결과 passes의 수가 30을 넘어가면 혼란도 점수가 거의 일정한 값에 수렴하는 것으로 확인되었다.



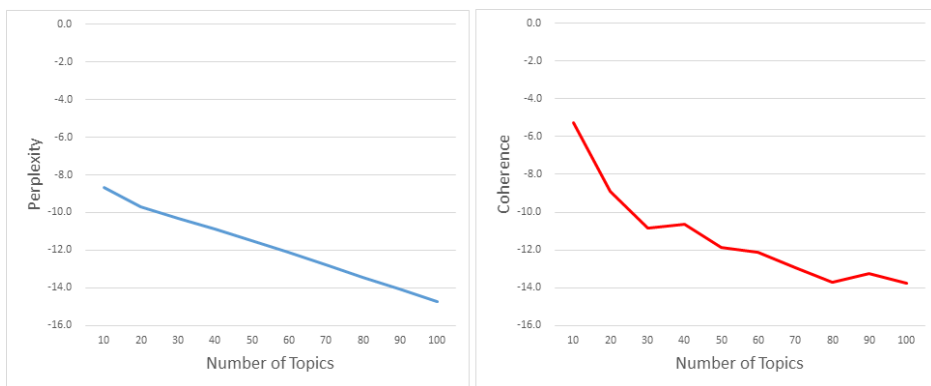
[그림 15] Passes에 따른 혼란도 점수와 일관성 점수 그래프 예시

### 4) 최적의 num\_topics 찾기

위 3)에서 구한 passes 값을 적용하고 LDA 모델 파라미터 중



num\_topics를 구한다. num\_topics를 2부터 20까지 2의 배수로 단계를 구분하여 테스트 하여 혼란도 점수와 일관성 점수를 측정하면 [그림 16]의 그래프를 얻을 수 있다. 그래프를 분석하여 가장 좋은 점수를 기록한 num\_topics를 최종 LDA 모델 파라미터 값으로 지정한다. 테스트결과 토픽의 개수가 20개 이하부터 혼란도 점수와 일관성 점수가 지속적으로 안 좋아지므로 20개 이상의 토픽을 검사하는 것은 의미가 없다.



[그림 16] Topic 수에 따른 혼란도 점수와 일관성 점수 그래프 예시

## 5. Python을 이용한 시각화(Visualization)

### 1) 모듈 설치

#### (1) IPython 설치

```
pip install IPython
```

## (2) pyLDAvis 설치

토픽 모델을 시각화하기 위한 Python 라이브러리이다. 사용자가 텍스트 데이터 모음에 적합한 토픽 모델의 토픽을 해석하도록 돕기 위해 설계되었다. 이 패키지는 대화식 웹 기반 시각화를 위해 적합한 LDA 토픽 모델에서 정보를 추출한다[94].

```
pip install pyLDAvis
```

## (3) 단어주머니(Bag of Words) 생성

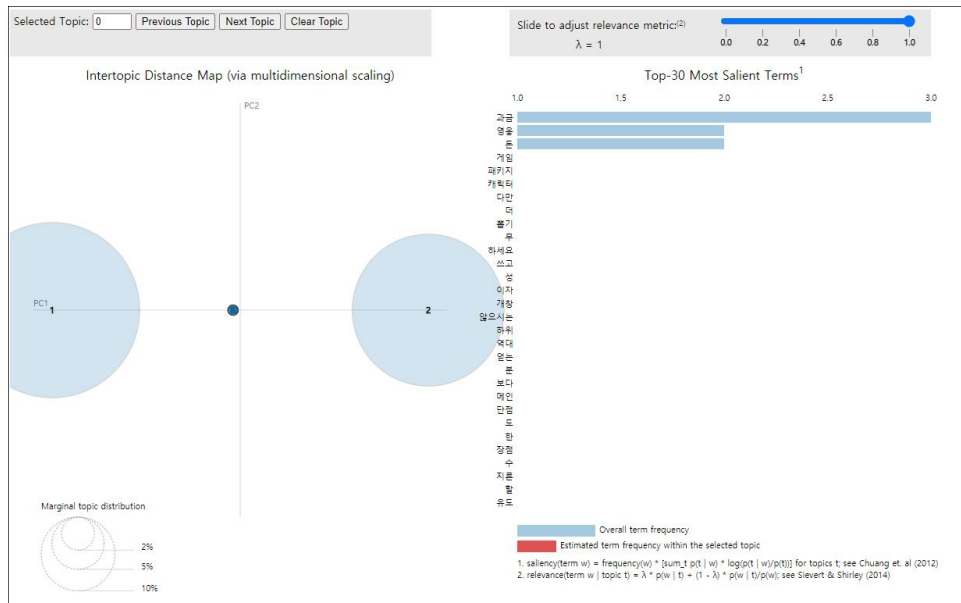
doc2bow 모듈을 이용하여 단어에 인덱스를 부여한다.

## (4) LDA모델 생성

앞선 평가에서 얻어진 토픽의 개수(num\_topics)와 패스의 숫자(passes)를 적용하여 LDA모델을 생성한다. LDA 모델 생성에는 gensim을 이용한다.

## 2) LDA 모델 도식화

pyLDAvis를 이용하면 LDA 모델을 도식화 할 수 있으며 이 정보는 html로 저장된다. 두 가지 형태의 정보를 제공하는데 [그림 17]의 왼쪽에 보이는 원형 그래프는 두 토픽의 연관성을 보여준다. 거리가 가까우면 두 토픽은 서로 비슷한 내용을 다루고 있는 것이고, 거리가 멀면 서로 다른 내용을 다루고 있는 것을 의미한다. [그림 17]의 오른쪽에 보이는 막대그래프는 토픽의 키워드를 보여준다.



[그림 17] pyLDAvis를 이용한 Visualization 예시

## (1) Salient

전체 문서에서 단어들이 얼마나 두드러지게 나타나는지를 보여준다. 일반적으로 많이 언급된 단어들이 salient 값이 높게 나타난다. 막대그래프에서는 하늘색으로 표현된다. 가장 두드러진 단어 30개의 리스트를 보여준다.

## (2) 람다( $\lambda$ )

람다( $\lambda$ )는 전체 문서에서 단어들의 관련성(Relevance)을 보여주는 것으로 오른쪽 막대그래프에서 0에서 1 사이의 값을 사용자가 조정하여

관련성을 표현할 수 있다. 람다 값은 기본적으로 1로 설정되어 있는데 이는 해당 토픽에서 가장 많이 발견된 단어들 위주로 보여준다는 의미이다. 반대로 0의 값은 토픽에서 가장 드물게 등장하는 단어들을 말한다.

람다 값 1은 가장 많이 발견되는 단어지만 의미 없는 단어가 포함되어 있을 가능성이 있다. 비록 불용어로 걸러냈다고 하지만 일부 의미 없는 단어가 포함되어 있기 때문이다. 오히려 람다값 0이 자주 등장하지 않지만 중요한 단어일 가능성도 있다.

본 논문에서는 사용자들의 성향을 분석하는 목적이므로 가장 많이 언급되는 단어들로 구성된 토픽에 중점을 두기로 한다. 따라서 람다값 1에서 추출되는 단어들로 토픽 분석을 실시한다.

### 3) 토픽 분석

LDA 모델 도식화를 기반으로 토픽을 찾아내야 한다. Salient 값이 높게 나온 단어들을 조합하여 유추하는 방법으로 토픽을 찾아낼 수 있다. 문서에 대해 어느 정도의 기반 지식이 없다면 토픽을 찾아내는 것이 어려울 수 있다. 또한 분석자의 주관적인 생각이 개입될 우려가 상당히 존재하여 분석자에 따라 다른 분석이 나올 가능성이 있다.

## IV. 실험 및 평가

### 1. 사례분석 : A게임에 대해서

#### 1) 게임소개

본 사례분석에 사용된 게임은 [그림 18]에서 보이는 ‘검은사막 모바일’로써 2019년 4월 구글플레이 한국 게임 매출 순위 20위 권 내의 RPG이다. 통계에 의하면 DAU와 매출이 모두 높은 게임으로 분석되고 있다[95].



[그림 18] 검은사막 모바일

#### 2) 빅데이터 분석 실험

##### (1) 크롤링

구글플레이 ‘검은사막 모바일’ 페이지에서 [그림 19]에서처럼 총 5,240개의 리뷰를 수집하여 1,050,234바이트의 텍스트 파일로 저장하였다[96]. 2020년 6월 28일 구글플레이의 robots.txt에 의하면 크롤링이 가능한 것으로 확인되었다. 크롤링 데이터는 2020년 6월 14일로부터

터 2018년 9월 16일까지의 과거 5,240개를 확보하였으며, 구글플레이 사용자 리뷰는 날짜 정렬이 되지 않아 해당 기간 내 게시물을 무작위 수집하였다.

수집한 데이터는 게임 A의 리뷰로써 모두 텍스트 데이터 형식을 취하고 있다. 해당 게임을 다운로드 한 사용자에게 한해서 작성이 가능하며 한 사용자가 몇 번이든 작성이 가능하다. 별도의 제목을 작성할 필요는 없으며 해당 게임에 대해 하고싶은 말이라면 무엇이든지 작성이 가능하다.

재미있긴한데 진입 보상이 좀 약한 듯 합니다  
뭔가 패치를 받았긴한데 아직도 제 캐릭터가 안보이네요 제 캐릭터가 안보인다고여여  
과금러인데 딱건다좋은데 진짜 pvp는 개똥망게임이네요 과금할 의지가 사라집니다  
현질하고 쪽해오면 계정이 실수로 삭제되고 복귀할려고 합니다  
설치를 했는데 캐릭터도안나오고 화면이 검정색으로 변하네요  
아니 제가 이 앱에 들어가면 검정바탕이고 응답이 없다고 합니다 이 것만 고쳐주면  
점나여 그래픽 예술임 쏜게임이라는 생각안들  
결재했는데 문제생겨서 문의날을라고하면 계속팅기네  
자꾸 게임 도중에 멈추는데 개선 해주세요  
사양이너무 높아서 접속률 많으면 렉발생  
새로 설치하는데 자꾸 업데이트하라고 드고 마켓에오면 업데이트는 안뜨는데요  
게임은 재미있는데 크야요 게임이 안되는디요 접속이 안되는디 지원다가 새로알아도 안되는디  
복귀할라니까 업데이트 하란소리만 반복하면서 실행부터 안됨 응 안해  
안녕하세요 제가 인게임에서 상품을 구매했는데 결제만 되고 상품이 들어오질 않았어요 여피해야하조  
조작도 간편하면서 방대한 게임현질유도도 읊따  
정말정말정말 재밌어요 오랜만에 싹껏 즐깁니다  
게임패드 작동이 되나요 아무래도 터치 조작은 불편해서  
깔리는데 로그인이 되질안아요 이유좀 알려주세요  
재미는 있네요 근데 넘 복잡하것 같은데요 모바일 겹치고는  
간만에 하니 점있네요 늦게 시작해서 쿠폰이 없어서 아쉽네요  
이벤트도 뽕뽕하고 점있음 초보자는 점에 많이 복잡하고 어려울수도 있는데 하다보면 암것도 아님  
정말재밋고 캐릭터가 너무 맘에 듭니다  
장개게임과 국산게임의 차이를보여준  
노트9인데 설치후 권한설정창에서 넘어가지질 않습니다 터치도안대고  
오류투성이게임이예요 허구헌날 낮에패치하고 엠티하고  
왜이렇게 자주 튕기나요 원 패치만 했다면 이런데 얼른 고치세요  
다른 3d모바일점 나오면서 느낀건데 검은사막이 넘사벽이네요 다시 검사나 합시다 여러분  
이벤트알림수신했는데 11월4주차 보상 문자로안왔는데 뭘중 받고 게임시작하려고했는데  
다른게임하다가 돌아오게되어  
재밋고 좋은데 한번씩 튕기고 끊기네요  
공부할게 많고 랩업은 느리지만단순하거나 지루하지 않아서 좋아요  
양산형 중엔 젤 나을걸요 님들 확률 극악이지만 불소 리니지보다 양호하고 소과금이라야 할만하다는데 무과금 뎅커들 꽤 있고 그래픽  
조금 허전한 것 같아요 컷테스트 만큼은 좋은데 불소 리니지도 꽤 좋은데 불소 안정한

[그림 19] 게임 A의 리뷰 수집

## (2) 토크닝

[그림 20]에서처럼 앞서 수집한 리뷰를 토크닝 처리하였다. 명사와 동사만 수집하고 불용어를 정리하여 597,896 바이트의 텍스트 파일로 저장하였다.

현재 두 약의 새도 유저 과금 줄거고 심연 등급 장전구 장 이상 장비 류 램 구성 이상 넘어가까엔 시간 모되 스토리 컨텐츠 소모 속도 조절 사 입장 이해 합니다만 월드 보스 컨텐츠 한정 횟수 비급 드롭을 확률 이유 성장 속도 늦어지고 개선 주시 현재 개편 예정 중인 돌 줄기는 항상 편의 고심 하시는 필 어비스 운영자 알아주셨으면 하여 남깁니다

설치 저장 공간 떠서 딴 어를 삭제 설치 했어요 서버 선택 직업 선택 위리어 검 아래 꽃고 고개 숙이는 장면 멈추어 접속 줄기기 시작 되네요

위장 채집 의상 캐릭 바꿀 사야는거 샀지만 저번 에도말 했지만 공룡 바꿔주 신안 달까 산 가지고있고 블랙필 주면서 반환 돈가 가요 태양 전장 예전 낮 점심시간 할수있게 시간 분배 저녁 시간 폭태 시간 평일 낮 하는거 고려 해주셨음

블레스 불스 유사 를 모두 해봤고 일주일 넘기 기전 이탈 되었다 전입 장벽 느껴졌고 그래픽 달지 않는데 최적화 기기 검사 수준 그래픽 부담 돌리며 컨텐츠 어려움 줄이고 목표 목표 달성 하며 캐릭터 재미 느낀다 가방 정보 정형 가 느꼈다 동선

시작 한지 일차 유저 그래픽 컨텐츠 보상 시작 루담 레벨 업 컨텐츠 줄기고싶어요 무게 차는거 이벤트 무게 증가 상점 인 임재 사거나 뿌리는 이벤트 했으면

운영 자동 사냥 질리는 감 알아 매번 물품 들어와서 무 과금 하고있네요

들어간 유저 저장 공간 확보 장공 부족 하다며 짜증 납니다 저장 공간 돌려주시길 바랍니다 여러분 장공 돈 뜯어 낼려고 만든 화나고 장공 줄였으면 되겠네요 고쳐 주기 화나서 올리겠습니다

무 과금 하다 보니 아이템 생기 들어요 업데이트 자주 하지만 이벤트 자주 레벨 낮 댄 노점 할수도 재미 붙여요

업 플레이어 하여 유저 월드 경영 이벤트를 잡기 핏습니 핏기 현상 개인 문제 월드 경영 최종 보상 나가더라도 받아질 주셨으면

자신 강점 를 모르거나 메인 가 월 별 병 걸려서 우린 승부 할수 착각 빠진 현실 리 따라가기에 시간 프로 이상 자동차 보는 담당자 조 전달 당시 월 를 운영 사실 입 사과 수준 유저 대적 운영 조작 모르겠고 유저 피드백 돈 무시 상위 프로 매출 프로 담당 하는건 알겠는데요 나머지 프로 유저 떠나가서 망점 분위기 생기 상위 프로 유저 남아 원가 착각 패치 수준 보 검사 모 매출 비중 큰데 신경 거의 쓰는거 리 따라갈 시간 당신 최고 강점 엔진 만든 리뉴얼 그래픽 최적화 시켜서 검사 모에 인식 시켜서 여름 넘겨 캐릭 노출 수명복 팔 깨도 매출 나오는데 과금 패키지 록 메달

퀘스트 완료 가요 완료 완료 터치 클리어 아우 스트레스 받습시다

모 계다 중반 가면 레업 하기가 나갑니다 단계 차 레벨 업 전략 쓰세요 스마트폰 장점

렐리 티나 모든 타 모바일 비교 해봐도 압도 만드신듯 최근 린 비교 해봐도 몇 만배 만드셨구요 운영 업데이트 하시느라 수고 그래픽 모두 최고점 굳이 단점 찾으면 할게 탓 갈피 못잡고 유튜브 교육방송 봐도 모르겠다는

격지 했던 뒤 월 속제 특유 컨텐츠 느껴지는 적정 루 컨텐츠 고갈 만큼 유저 벗어나는 업데이트 겹 쳐서 전형 꼴을 운영 보여줍니다

결과 애증

페이스북 접속 유저 로그 인시 곳 접속 본인 위해 신분증 보내 달라 여기 신분증 업로드 되지 않아 접속 하고있습니다

즐기면서 해온 유저 저가 시작 구글 연동 시작 했는데요 최근 안드로이드 아이폰 갈아탔는데 혹시 아이폰 계정 불러오는 시간 투자 했고 해를 알려주시면 감사하겠습니다

년 전하 다가 복귀 유저 그래픽 장 이벤트 사랑받는 모바일 되길 기원

진행 시 먹통 되면서 반응 없었는데 차 업데이트 되고 지우고 깔았는데 기기 갯 노트

구글 기프트카드 리 워드 이벤트 리 워드 받았는데 결제 되어있더라구요 환불 요청 경우

되었는데 이벤트 정착 지원 해주네요 감사합니다

개인 설치 접속 하니 기존 데이터 전부 초기 되 복구 시킬

다른 앱도 비해 운영 면 점차 개선 나가주실 믿습시다

무 과금 모바일 무 과금 줄길 회사 하겠지만

성장 특화 신규 유저 혜택 별것 성장 쉽지도 않고 전투력 올라요 광고 낡 시작

쟁 컨텐츠 줄기지 않는 유저 수도 상당수 존재 특정 재화 쟁 수급 문제 주술 사고싶어용

용량 떠서 재시 번은 해봐는데 지우고 깔고 용량 문제

나온지 되었는데도 한주 안거르 업데이트 및 이벤트 우리나라 대한민국 몇 무 과금 해볼수있는 안밖 사 마음 들니 재미나 감사합니다

유튜브 보다가 알았습니다만 전쟁 연가 란 검사 커스터 마이 쟁 시스템 캐릭터 선택 배 끼겨

[그림 20] 게임 A의 토크닝 처리

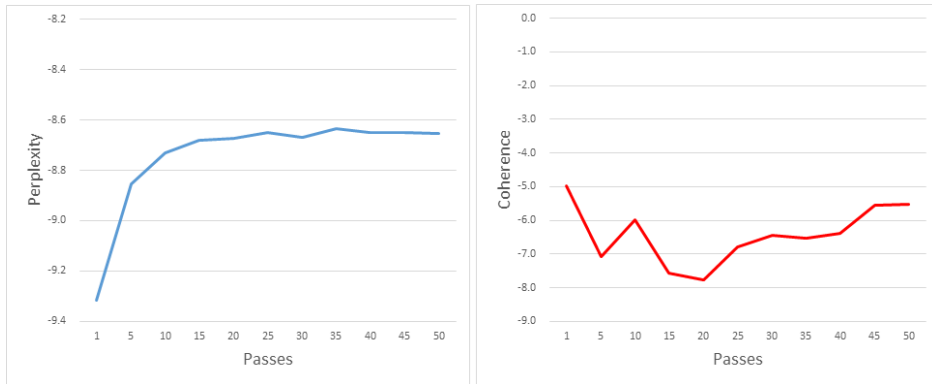
### (3) 좋은 LDA 모델을 위한 passes와 num\_topics 찾기

말뭉치 데이터에서 13,727개의 BoW가 수집되었다. <표 4>와 [그림 21]의 분석 결과 passes의 개수가 35개 일 때 perplexity가 - 8.63, coherence - 6.52로 적절한 passes의 개수는 35로 설정하였다.

<표 4> 게임 A의 passes 선정 평가

Number of Topics : 10		
Passes	Perplexity	Coherence
1	-9.31733058240904	-4.97822704307078
5	-8.85533264054937	-7.08837488674337
10	-8.72987158964666	-5.97760306040754
15	-8.67998583984564	-7.57764065286563
20	-8.67248388304137	-7.76321987343434
25	-8.64926669900657	-6.79417103818888
30	-8.67081857683750	-6.43441554243108
35	-8.63459062575451	-6.52871132251322
40	-8.65097062368819	-6.39405473319330
45	-8.64812449990869	-5.56232710901742
50	-8.65268198418890	-5.52872994170555



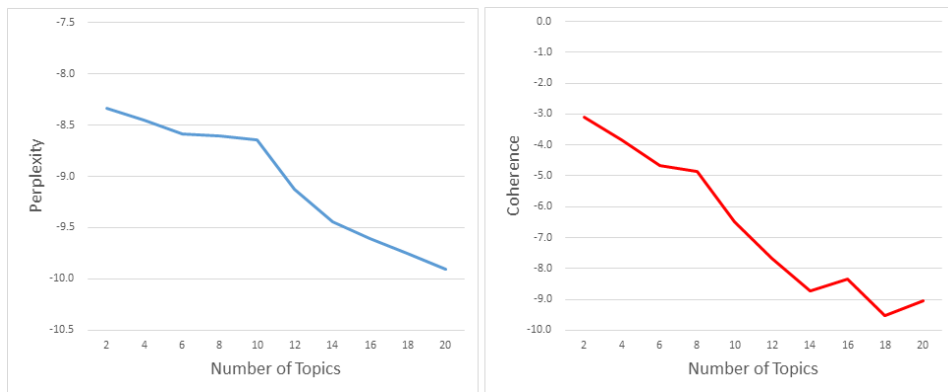


[그림 21] 게임 A의 passes 선정 평가

passes를 35로 설정하고 실행한 테스트에서는 <표 5>와 [그림 22]의 분석 결과 최적의 num\_topics는 2개로 확인되었다. 하지만 토픽의 수가 너무 적어 다양한 의견을 얻을 수가 없어서 범위를 넓히는 시도를 하였다. 토픽이 8개가 넘으면 급격하게 Coherence Score가 나빠지는 것으로 나타났으므로 8개의 토픽으로 LDA 모델을 테스트한다.

<표 5> 게임 A의 num\_topics 선정 평가

Number of Passes : 35		
num_topics	Perplexity	Coherence
2	-8.33850978111908	-3.08267221390819
4	-8.44744144116353	-3.82648995881847
6	-8.58400861235560	-4.65830563814543
8	-8.60793876850313	-4.87088801386873
10	-8.63928218273656	-6.50385551658380
12	-9.12589199656941	-7.67620109201471
14	-9.44575059747794	-8.74085583505371
16	-9.60364141133656	-8.32949969670364
18	-9.74802498382963	-9.52195720322489
20	-9.90354478338539	-9.03879106817835

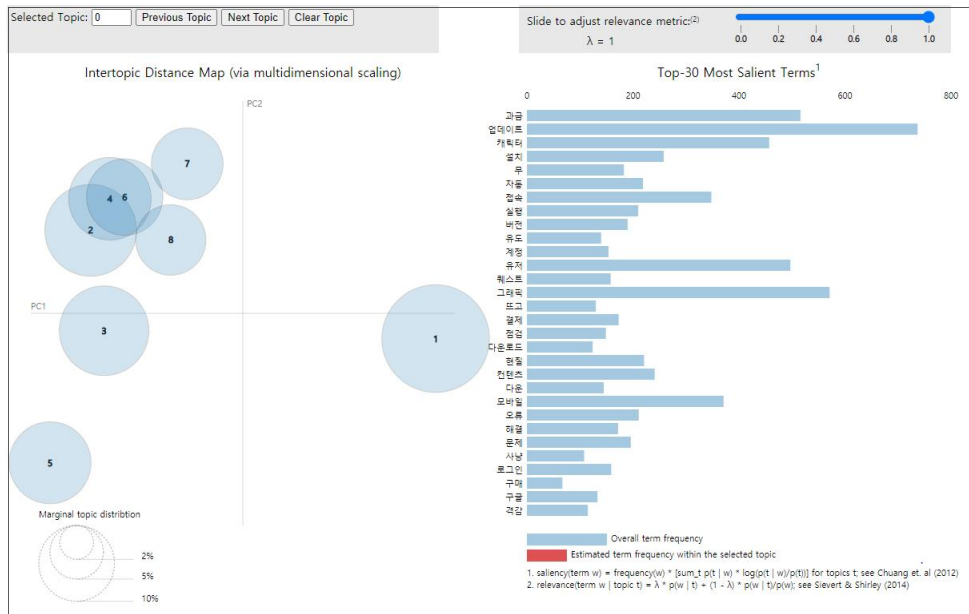


[그림 22] 게임 A의 num\_topics 선정 평가

### 3) 시각화 결과

passes는 35, num\_topics는 8의 파라미터로 LDA 평가를 실시하였으며 [그림 23]과 같은 결과를 얻었다.

#### (1) Top-30 Most Salient Terms

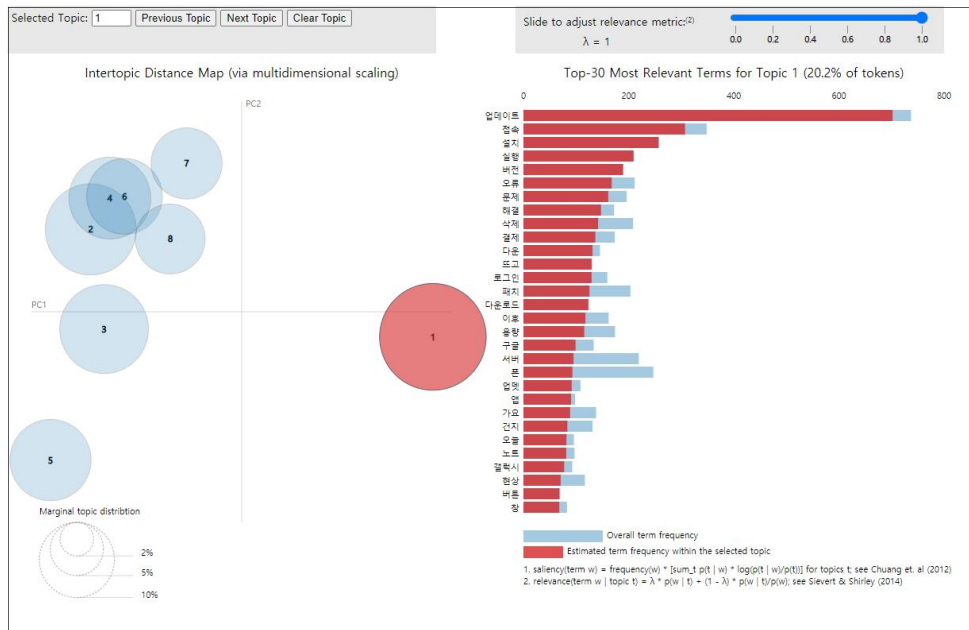


[그림 23] 게임 A의 LDA Visualization 결과

전체 문서에서 가장 두드러진 용어는 다음과 같다. 게임을 특정할 수 있는 용어는 삭제하였다.

과금, 업데이트, 캐릭터, 설치, 무, 자동, 접속, 실행, 버전, 유도, 계정, 퀘스트, 그래픽, 결제, 점검, 다운로드, 현질, 모바일, 오류, 해결, 문제, 사냥, 로그인, 구매

## (2) Top-30 Most Relevant Terms for Topic 1



[그림 24] 게임 A의 Topic 1에 대한 결과

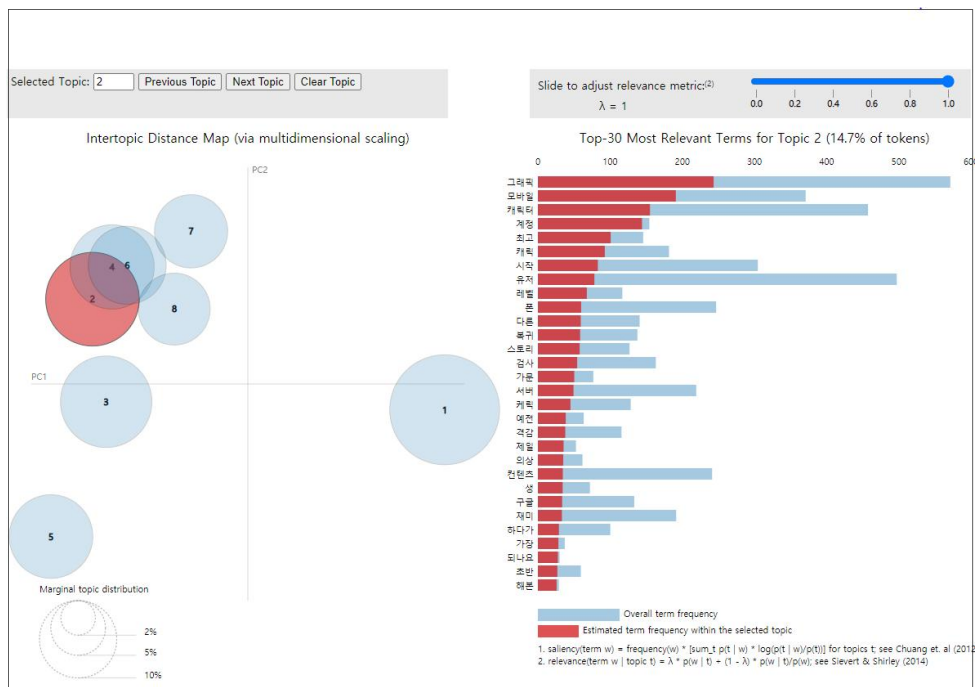
토픽 1의 결과는 [그림 24]에서 보는 바와 같다. 가장 관련성이 높은 단어들은 다음과 같다.

업데이트, 접속, 설치, 실행, 버전, 오류, 문제, 해결, 삭제, 결제, 다운, 로그인, 패치, 다운로드, 용량, 서버, 폰, 앱

분석 결과 토픽 1은 게임의 실행 문제와 관련이 있다. 게임을 업데이트하거나 패치를 할 때 잦은 오류 발생 등 문제가 발생하고 있는 것으로 유추할

수 있다. 일반적으로 게임의 실행에 문제가 있으면 사용자들에게 진입장벽으로 작용해 게임을 떠나게 된다. 하지만 사용자 리뷰에 이러한 내용이 많다는 것은 빠른 문제 해결을 바라는 사용자들의 바람이 적용된 것으로 볼 수 있어 게임에 대한 충성도가 매우 높음을 알 수 있다.

### (3) Top-30 Most Relevant Terms for Topic 2



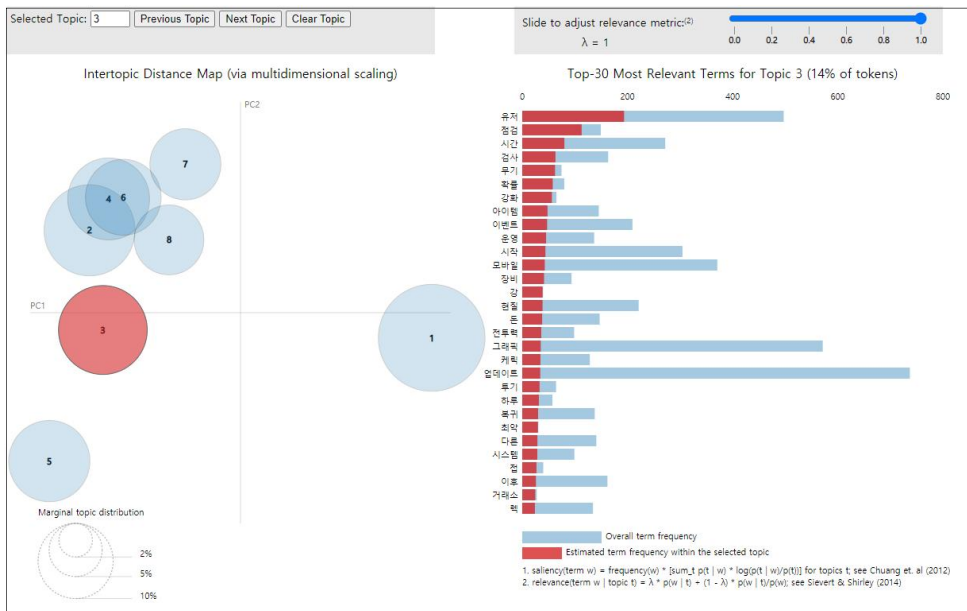
[그림 25] 게임 A의 Topic 2에 대한 결과

토픽 2의 결과는 [그림 25]에서 보는 바와 같다. 토픽 2에서 가장 관련성이 높은 단어들은 다음과 같다.

그래픽, 모바일, 캐릭터, 계정, 시작, 레벨, 폰, 복귀, 스토리, 가문, 서버, 의상, 재미

분석 결과 토픽 2는 게임의 그래픽과 관련되어 있다. 모바일이란 단어가 주로 언급된 것은 A 게임이 PC 온라인으로도 서비스되고 있기 때문이다. 사용자들은 PC 온라인 버전과 모바일 버전의 그래픽에 대해 비교하는 것으로 유추할 수 있다.

#### (4) Top-30 Most Relevant Terms for Topic 3



[그림 26] 게임 A의 Topic 3에 대한 결과

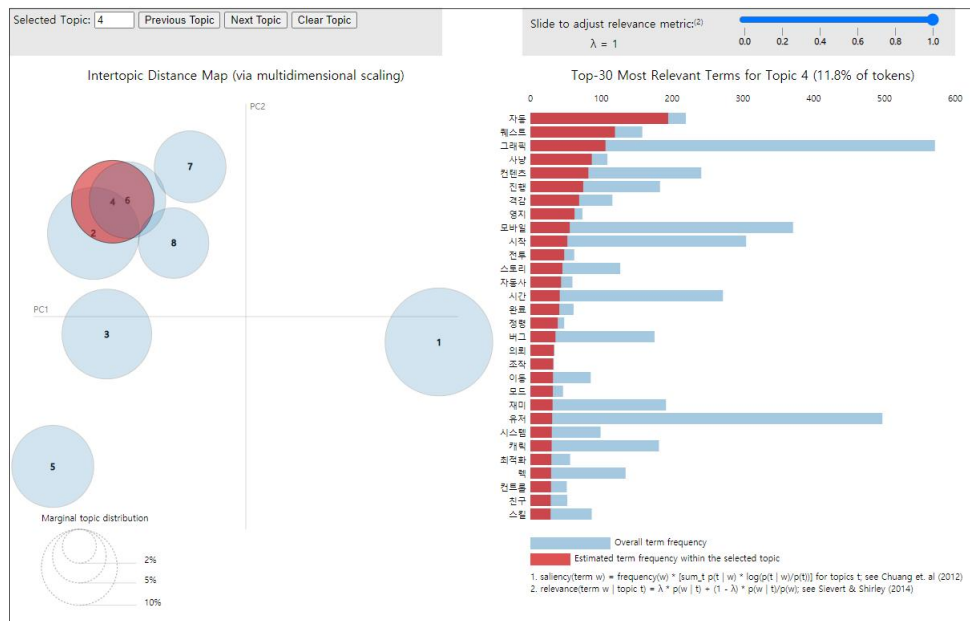
토픽 3의 결과는 [그림 26]에서 보는 바와 같다. 토픽 3에서 가장 관련성

이 높은 단어들은 다음과 같다.

점검, 검사, 무기, 확률, 강화, 아이템, 이벤트, 운영, 모바일, 장비, 강, 현실, 전투력, 돈, 그래픽, 거래소

분석 결과 토픽 3은 캐릭터의 강화와 관련된 BM과 관련된 것이다. RPG에서 캐릭터의 성장은 중요한 요소이고 회사는 이를 주요 수익요소로 삼고 있다. 얼마의 돈을 투자해서 얼마만큼 강해질 수 있는가에 관심을 보인다고 유추할 수 있다.

## (5) Top-30 Most Relevant Terms for Topic 4



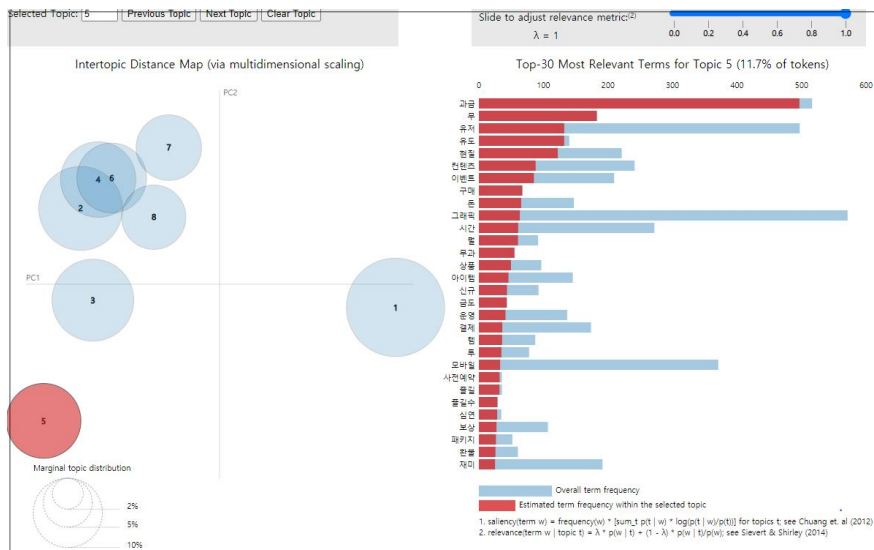
[그림 27] 게임 A의 Topic 4에 대한 결과

토픽 4의 결과는 [그림 27]에서 보는 바와 같다. 토픽 4에서 가장 관련성이 높은 단어들은 다음과 같다.

자동, 퀘스트, 그래픽, 사냥, 진행, 영지, 전투, 스토리, 시간, 완료, 정령, 버그, 의뢰, 조작, 이동, 재미, 최적화, 랙, 컨트롤, 친구, 스킬

분석 결과 토픽 4는 자동 사냥에 관련된 내용을 다루고 있다. 최근 RPG의 트렌드가 된 자동 사냥에 대해 이슈가 많은 것으로 보이며, 자동사냥으로 인한 조작의 편리성 대비 재미에 대한 논의가 진행되고 있음을 알 수 있다.

## (6) Top-30 Most Relevant Terms for Topic 5



[그림 28] 게임 A의 Topic 4에 대한 결과

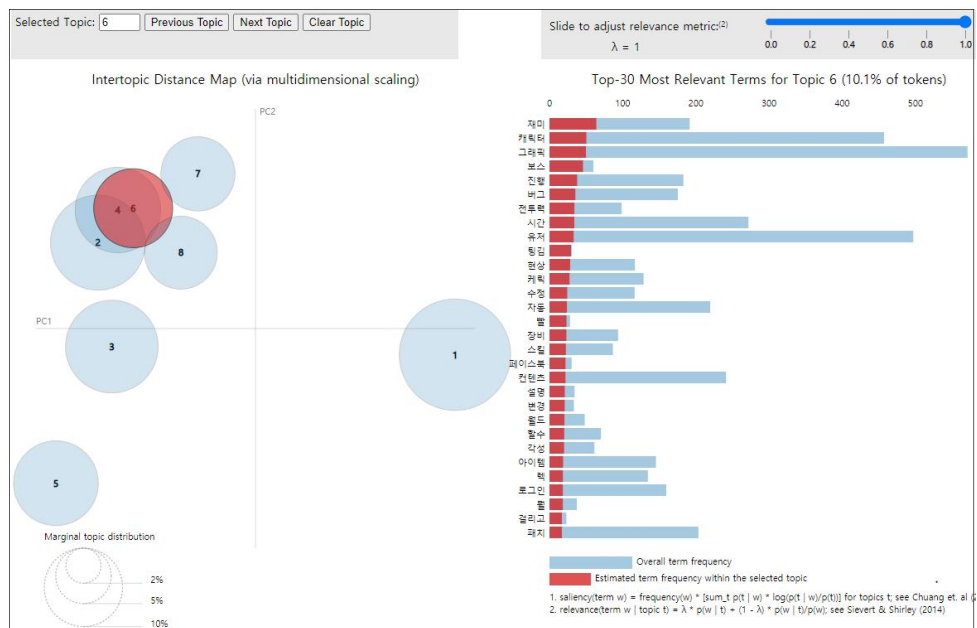


토픽 5의 결과는 [그림 28]에서 보는 바와 같다. 토픽 5에서 가장 관련성이 높은 단어들은 다음과 같다.

과금, 무, 유도, 현실, 콘텐츠, 이벤트, 구매, 돈, 그래픽, 시간, 상품, 아이템, 신규, 운영, 결제, 사전예약, 보상, 패키지, 환불

분석 결과 토픽 5는 게임의 과도한 구매유도와 관련된 것으로 보인다. 과금을 하지 않고(무과금) 게임을 즐길 수 있는지에 대해 이슈가 있음을 알 수 있으며, 매출이 높은 게임으로 분석된 만큼 과금을 유도하는 이벤트와 패키지 상품 등에 대해서 관심도가 높음을 알 수 있다.

## (7) Top-30 Most Relevant Terms for Topic 6



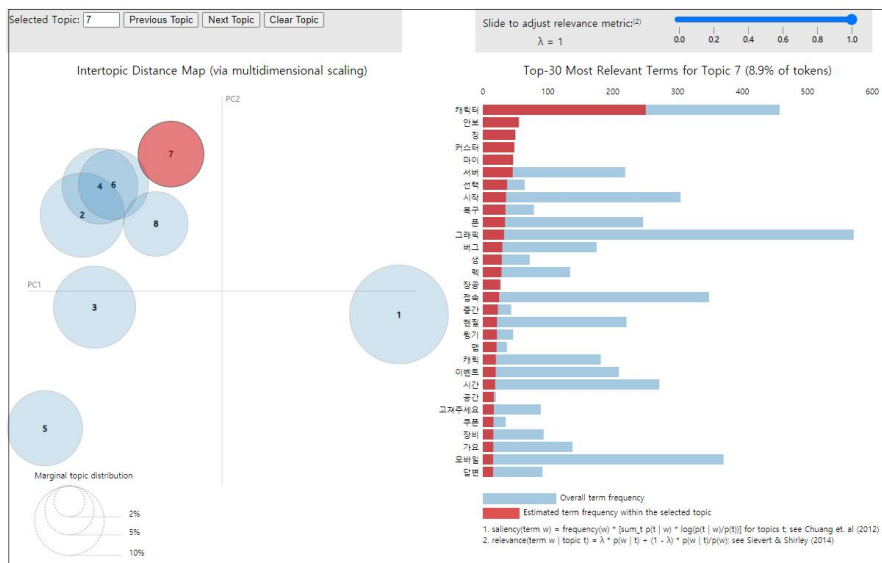
[그림 29] 게임 A의 Topic 6에 대한 결과

토픽 6의 결과는 [그림 29]에서 보는 바와 같다. 토픽 6에서 가장 관련성이 높은 단어들은 다음과 같다.

재미, 그래픽, 캐릭터, 보스, 진행, 버그, 전투력, 시간, 튕김, 자동, 장비, 스킬, 아이템, 렉, 패치

분석 결과 토픽 6은 게임의 보스 전투와 관련된 내용으로 본 게임의 보스 전투가 큰 재미요소임을 알려준다. 보스를 공략하기 위해 어떤 장비와 아이템이 필요한지 어떤 과정을 통해 공략할 수 있는지에 대해 이슈가 되고 있음을 알 수 있다.

## (8) Top-30 Most Relevant Terms for Topic 7



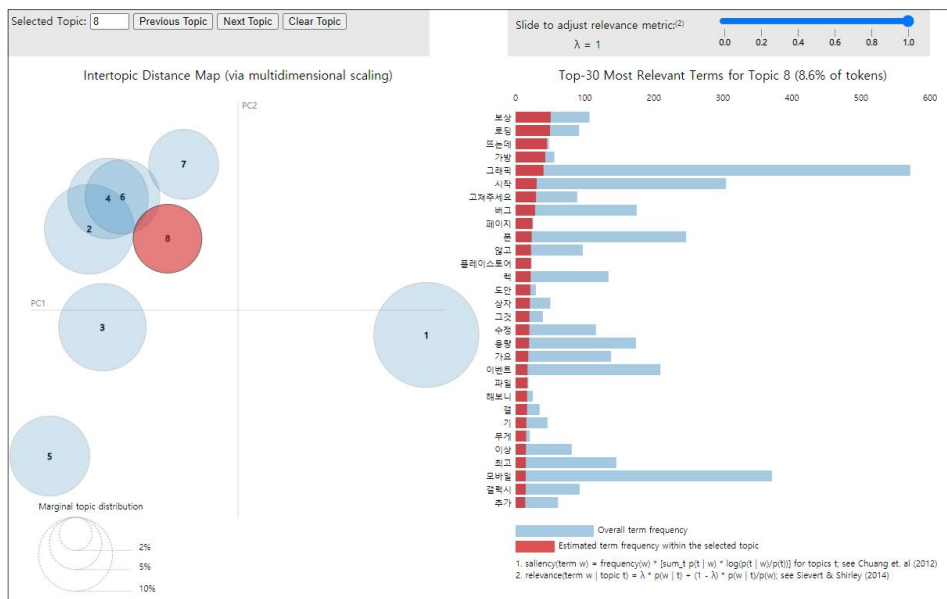
[그림 30] 게임 A의 Topic 7에 대한 결과

토픽 7의 결과는 [그림 30]에서 보는 바와 같다. 토픽 7에서 가장 관련성이 높은 단어들은 다음과 같다.

캐릭터, 커스터마이징, 서버, 선택, 복귀, 폰, 그래픽, 버그, 텍, 접속, 현실, 튕김, 답변

분석 결과 토픽 7은 캐릭터의 커스터마이징과 관련된 내용이다. 본 게임은 사용자 스스로가 자신의 캐릭터를 꾸밀 수 있는 커스터마이징 요소를 제공하고 있으며, 커스터마이징에 대한 사용자들의 관심도가 매우 높음을 알 수 있다.

## (9) Top-30 Most Relevant Terms for Topic 8



[그림 31] 게임 A의 Topic 8에 대한 결과

토픽 8의 결과는 [그림 31]에서 보는 바와 같다. 토픽 8에서 가장 관련성이 높은 단어들은 다음과 같다.

보상, 로딩, 버그, 폰, 렉, 가방, 도안, 상자, 수정, 용량, 파일, 모바일

분석 결과 토픽 8은 게임의 로딩 과정 중에 발생하는 렉에 대한 버그와 그에 따른 보상 문제를 다루고 있다. 토픽 1에서 게임의 실행 문제와 관련되어 잦은 오류가 있다고 분석되었는데 본 게임의 운영팀은 업데이트나 패치 등으로 인해 오류가 발생할 경우 적절한 보상을 통해 사용자들을 달래고 있다고 유추할 수 있다.

#### 4) 토픽 분석

토픽 1과 8에서는 게임의 오류가 큰 이슈가 되고 있으며 이를 무마하기 위한 보상이 적절하게 지급되고 있는 것으로 보인다. 또한 게임의 실행 오류에 대해 사용자들이 많은 리뷰를 남긴다는 것으로 보아 빨리 문제가 해결되기를 바라는 마음이 담겨 있다고 볼 수 있다. DAU가 높은 게임은 사용자들의 충성도가 매우 높다는 것을 보여주고 있다.

토픽 2와 7에서는 게임의 그래픽 품질과 관련된 내용으로 이슈가 되고 있음을 보여준다. PC 버전과 온라인 버전의 그래픽 품질, 그리고 본 게임의 주요 특징인 커스터마이징이 사용자들이 본 게임을 플레이하는 이유가 되고 있음을 알 수 있다.

토픽 3과 5는 캐릭터의 성장과 관련된 내용을 다루고 있는데 회사는 캐릭

터를 강화시키는데 과금을 유도함으로써 이에 대해 사용자들의 이슈가 많은 것으로 파악된다. 다만, 매출이 높은 게임으로 분석되었기 때문에 많은 사용자들이 과금을 하여 캐릭터를 강화시키는 데에 긍정적인 반응을 하고 있음을 알 수 있다.

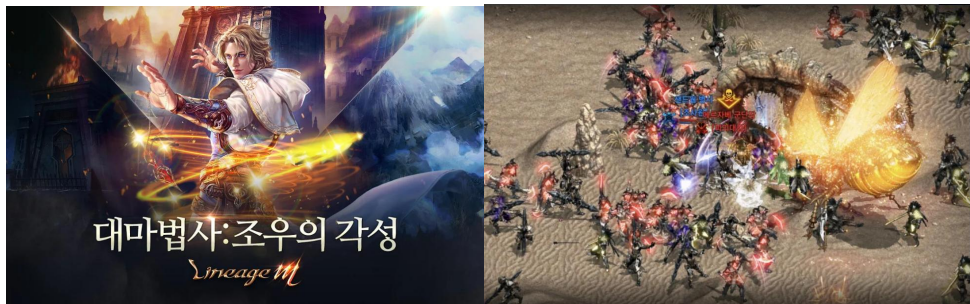
토픽 4와 6은 자동사냥 및 보스 전투와 관련된 것으로, 게임의 전투가 재미요소로 언급되고 있음을 알 수 있다.

전체 토픽을 분석한 결과 본 게임에 대해 사용자들은 안정적인 운영을 바탕으로 한 기술적 이슈 해소, 뛰어난 그래픽, 과금을 통한 캐릭터 강화, 재미있는 전투라고 인식하고 있음을 알 수 있다.

## 2. 사례분석 : B게임에 대해서

### 1) 게임소개

본 사례분석에 사용된 게임은 [그림 32]에서 보이는 ‘리니지M’으로써 2019년 4월 구글플레이 한국 게임 매출 순위 20위 권 내의 RPG이다. 통계에 의하면 DAU는 낮지만 매출이 높은 게임으로 분석되고 있다 [97].



[그림 32] 리니지M

### 2) 빅데이터 분석 실험

#### (1) 크롤링

구글플레이 ‘리니지M’ 페이지에서 [그림 33]에서처럼 총 5,200개의 리뷰를 수집하여 617,210바이트의 텍스트 파일로 저장하였다[98]. 2020년 6월 28일 구글플레이의 robots.txt에 의하면 크롤링이 가능한 것으로 확인되었다. 크롤링 데이터는 2020년 6월 14일부터 2018년 9월 20일까지의 과거 5,200개를 확보하였으며, 구글플레이 사용자 리뷰는 날짜 정렬이 되지 않아 해당 기간 내 게시물을 무작위 수집하였다.

수집한 데이터는 게임 B의 리뷰로써 모두 텍스트 데이터 형식을 취하고 있다. 해당 게임을 다운로드 한 사용자에게 한해서 작성이 가능하며 한 사용자가 몇 번이든 작성이 가능하다. 별도의 제목을 작성할 필요는 없으며 해당 게임에 대해 하고싶은 말이라면 무엇이든지 작성이 가능하다.

전서버 생성제한이면 신규유저는 게임하지말란건가 답변보니 언제 풀릴 예정이라는것도아니고 그냥 한번씩 확인해보면서 되면하라는답변은 뭐조 진짜 다운로드전에 미리알려주던지 5기거나 다운로드했는데 장난치나 게임 다시 시작하려고 접속 했더니서버마다 신규캐릭 생성제한이 걸려있어서 서버에 들어가지도 못하고 접속을 끈었네요 너무한거 아닌가요 최소한 몇개의섭은 캐릭 생성을 하게 해주셔야 하는거 아닌가요 전섭을 모두다 제한을 두시면 이젠 게임을 하지 말라는거 아니니까요 제한 해제 해 주세요 단 몇개의 섭만이라도요

돌아하는 그래픽에 열심히 키우고 있는 유저 한데 무과금 유저라 캐릭터 삭제가 너무 답이 없음 그나마 알함 베풀어 만든게 400 다이아 캐삭권 같은데 제발 60미만 말고 그냥 캐삭 가능하게 해주세요 못해도 50은 해줘야 하지 않나요

개 10 이틀전에 같았는데 점검끝나고 바로 접속해도 전서버 생성제한이네 작업자들 정리좀하고 신규유저 점은 하게해야하는거 아니냐 서버를 빠르게 늘리면가18 오픈한지가 언젠데 신규유저는 아예 점도 하지말란말이나 개10같은 nc 아 갤럭시S10E에서 갤럭시S20으로 바꿨는데 게임 텍이 왜이리 심한가요 S20이 리니지M만 텍 많다는 글도 많은데 왜 안고치는지

신형으로 바꿨는데 텍이 더 심한 정도가 아니라 없던 텍이 이리 생겼는데 구형으로 바뀌야 되는건지

문의들 드릴곳을 찾지 못해 리뷰에 남겨 드립니다 과금을 하여 최상급 마법인형 뽑기에서 에틴레드 뽑았습니다 근데 웬걸 영웅등급 마법인형 제외하고 나머지만 들어왔습니다 너무 신나 가지고 화면 캡처를 못하게 진짜 서러운데 누락인건가요 아니면 오류가 난건가요 문의 가능한 곳이나 정확한 답변 부탁드립니다 솔직히 진짜 어이가 없습니다

옛날 리니지 감성 생각하고 시작한 리니지m 하지만 이질감이 너무 컸고요 리니지하면 데스나이트인데 변신목록에서도 너무 저평가되어있는거 아닌가 싶네요 과금유도 좀 줄이고 통란하면서 즐길수있게 해주면 안되나요 리니지를 사랑하는 일반유저로서 너무 안타깝습니다

컴퓨터로만 하다가 모바일로 처음해보려고하는데 캐릭터생성제한이있네요 신규채널마저 그렇다니얼른 개선되길바라면서 별점드릴게요

진짜개뿔같네 오래만에 리니지 생각나서30분걸려서 깔았지만 서버제한은언데 신규는아예 안받겠다는거네 역시 고인물게임 고맙다 돈안쓰게해줘서 처음 입문하시는 분들은 다른 고인물에게 죽지 않도록 현질을 많이 해 장비를 맞추고 시작하실 권합니다 Nc에서 만든 자등사냥 기능을 사용할수가 없습니다 자등사냥 하는 유저들만 골라 죽이고 다니는 유저들막피때문에 게임을 원할하게 즐길수가 없습니다 저 또한 이유없이 막피 하는 고인물들에게 여러번 죽어 다시는 할 생각이 없어 지우기 전에 몇달동안 한 소감 남깁니다 처음 시작 하시는 분들에게 정보가 되길 바랍니다

처음깔고 처음접속하는데 서버원활한것 같나많은데 서버안정화를 위한 캐릭생성제한게다가 리뷰보면 결제사기겁나잖아서 나도 사기당할까봐 바로제거

개인거래는 진작에 거론도안하고있고 과금유도뿐 무과금못즐깁니다 하지마세요 현명한유저라면 극소수개돼지들과 매크로밖에없어서 매달몇백을써도 못즐깁니다

업데이트 지금 안되고 저장공간때문인가싶어서 다지우고 업햇해도 에러다지우고 새로깔아도 똑같네요 어케 해야되조

오랜만의 접속에 업데이트 아이디 찾기 후 로그인을 했는데 캐릭터가 없고 생성제한 서버 뭐 이런것만 따로 몇 년만에 접속도 아니고 한 두달 안했을 뿐인데 제 캐릭 어디 간거조 빠른 해결 요청합니다 여러 가입환영 메시지 이젠 뭐야 대체 그러고 새로 가입한적 없다고 답 메일 보내니 차단됐다고 오네 이거 무슨 상황이조 미치겠네 이런걸로 남비할 시간 없으니 빨리 해결해줘요 전화하니 근무시간 아니라고 나오고 도대체 뭘 어쩌라는건지 계정 헛갈렸네 어쩌고 저쩌고 하지마세요 피시로 게임한적 없고 로그인 한 적 당연히 없고 아이디 찾기에서 인증하고 하나 뜬 아이디 확인하고 비번 한다음 바로 로그인 후 엉망진창 대가리 섞여지게 한거니 진짜 기분 나쁘다 못해 짜증나고 화나네요 당장 해결 해주세요

작업장을 일부러 방지하는 게임입니다 패치나 이벤트는 모두 핵과금유저를 위한거구요 게임 내 아이템 드랍을 합성 확률등을 자기를 마음대로 대폭 하향시키고 그 아이템들을 패키지 같은 형태로 끼워서 팔아먹는 놈들입니다 중소과금 유저들은 절대 못하는 게임입니다

게임은 정말 ভাল물었는데요 게임내유저의70프로가작업장 운영진들이 돈에 환장해서 작업장들을 일부러 안잡고 게임 내 시세 유지하는데 이용하구요 이변에 신규 서버 신규 캐릭 또한 유입 되는 유저를 위한게 아니라 기존에 유저들 신규 캐릭으로 갈아 타게 만들어서

[그림 33] 게임 B의 리뷰 수집

## (2) 토크닝

[그림 34]에서처럼 앞서 수집한 리뷰를 토크닝 처리하였다. 명사와 동사만 수집하고 불용어를 정리하여 365,415 바이트의 텍스트 파일로 저장하였다.

저버 생 제한 신규 유저 하지말란 답변 토니 언제 풀릴 예정 해보면서 되면하라는 답변 다운 받기전에 미리 알려주면지 가다 다운 받았는데 장난 치 시작 하려고 접속 했더니 서버 신규 캐릭 생 제한 걸려있어서 서버 들어가지도 접속 곤었네요 최소한 몇개 생 캐릭 생 해주셔야 하는거 생 모두 제한 두시 제한 해제 단 몇개 생 그래픽 키우고 유저 무 과금 유저 캐릭터 삭제 답 알량 베풀어 만든게 다이아 캐삭 미만 캐삭 해줘야 않나요 이를 깔았는데 점점 끝나고 바로 접속 서버 생 제한 업자 정리 신규 유저 하계해 서버 놀리면가 오론 한지 젠편 신규 유저 아예 하지말란 걸럭시 걸럭시 바꿨는데 렉 렉 안고치는지 신형 바꿨는데 렉 렉 생겼는데 구형 바꿔야 건지 문의 드릴 곳 찾지 남겨 드립니다 과금 하여 최상급 마법 인형 뽑기 에틴 레드 뽑았습니다 영웅 등급 마법 인형 제외 나머지 들어왔습니다 가지 캡처 한게 누락 가요 오류 난건 가요 문의 곳 답변 어이 옛날 감성 시작 절감 컸고요 데스 나이트 번신 룩룩 평가 되어있는거 심네요 과금 유도 줄 통런 즐길수있게 해주면 사랑 일반 유저 로서 컴퓨터 로만 하다가 모바일 해보려고하는데 캐릭터 생 제한 신규 채널 얼른 개선 되길 바라면서 드릴게요 돛 생각나서 깔앗드 서버 제한 신규 는아예 받겠다는거네 역시 인물 돈 쓰게 해줘서 입문 하시는 다른 인물 죽지 않도록 현실 장비 맞추고 시작 하실 권합니다 만든 자동차 가능 할수가 자동차 유저 골라 죽 다니는 유저 막피 할하 줄길수가 또한 이유 막피 인물 죽어 다시는 지우기 탈동 소감 남깁니다 시작 하시는 정보 되길 바랍니다 깔고 접속 서버 곳 겁나 서버 안정화 위 캐릭 생 제한 게다가 결제 기갑 나도 사기 바로 제거 개인 거래 진작 거론 도안 과금 유도 무 과금 줄깁니다 유저 라면 개돼지 메 크로 매달 백 써도 줄깁니다 업데이트 장공 심어서 다지우고 업애헤 예러 지우고 깔아도 어케 되조 접속 업데이트 찾기 로그인 캐릭터 생 제한 서버 떠요 몇 년 접속 했을 뿐 캐릭 어디 간거조 해결 요청 여러 가입 환영 메세지 대체 가입 한적 답 메일 보내니 차단 됐다고 오네 상황 시간 해결 해줘요 근무시간 도대체 뭘 어찌 건지 계정 어찌 저찌 피시 한적 로그인 적 찾기 인증 돈 비번 다음 바로 로그인 엉망 진창 대가리 섞어지게 한거니 기본 화나네요 당장 해결 작업장 일부러 방치 패치 이벤트 모두 핵 과금 유저 위 거구 아이템 드랍을 합성 확률 자기 마음대로 대폭 하향 시키고 아이템 패키지 형태 끼워서 팔 놈 들입니다 중소 과금 유저 절대 만들었는데요 유저 프로 작업장 운영 돈 작업장 일부러 잡고 시세 유저 이용 하구요 신규 서버 신규 캐릭 또한 유입 유저 위 기존 유저 신규 캐릭 타게 만들어서 지르고 패키지 팔 먹게 하기 위 상술 캐릭 사기 캐릭 내놓아서 기존 클래스 유저 혹기사 타는 추세 한국 최초 결론 해보려고 하시는 절대 만원 현실 하시거들랑 하시는게 나야요 최소 백만원 써야 현실 티 매크로 몬스터 잡는데 자동차 클릭 배치 하다가 몬스터 잡네요 이유 특정 돌아가면서 포함 죽 다니네요 그래픽 만들어주세요 매크로 무작위 죽어시는 정 떨어져서 하기 관리자 좀더 관심 가지 없애주세요 성인 운영 되나 싶네요 그래픽 전혀 성인 않는데 만들어주세요 하시는 흥미 점점 잃어가시는데 뭔가 즐길 되었으면 확장시절 하면 유저 끼리 대화 뭔가 기억 성인 되어서 느낌 받아요 특정 인증 보니 현실 유도 수익 보신 만큼 투자 코로나 조심하시고 하루 되시길 바랍니다 주년 이벤트 꼭 장비 복구 나왔음 꼭 환불 신청 현금 금일 다이아 대리 결제 해주었는데 사기 결제 역 확인가능 하구요 결제 취소 결제 되고 아이템 들어왔는데 구공 환불 소비자보호 센터 환불 처리 일대일 교환 출시전부터 광고 해놓고 교환 되고 라지 모바일 유저 위 업데이트 패키지 팔 먹으려고만 업데이트 만해 되고있어요 진심 주역 시작 하려는 차라리 모바일 진심 후회 돈 써도 돈 쓴 티 폰 번호 기기 바 개서 구글 치면 과거 폰 인증 번호 전송 답 만들어 두셔서 폰 번호 변경 할수있게 해주시먼지 홈페이지 앱 전부 인증 하니 본인 폰 인증 번호 와야 인증 할텐데 해결 하셔 싶네요 운영자 코로나 고객 센터 콜도 문의 답변 주시 구글 개인정보 폰 번호 변경 다마 쳤는데 정답 안주시 하려는데 생 제한 기변 등록 채적 난국 삭제 하신건 가요 욕 잘못 된거 환불 해달라고 적은건데 삭제 답변 무시 하시는건 가요 시작 했지만 추억 돈고 무분별 결제 유도로 지치는건 사실 결제 외 시작 동시 지르고 번신 뽑는데 확률 현실 해야하나 고민 가차 확률 노리 사의 첫 유저 돈 하는것 이상하

영자 단신 답 매크로 답글 달면서 유저 호구 할수 쫓아 시작 했다고 들하고 운영 욕 봐도 제 무 섯 답글 안다셔도 되는데 회사 개인

[그림 34] 게임 B의 토크닝 처리

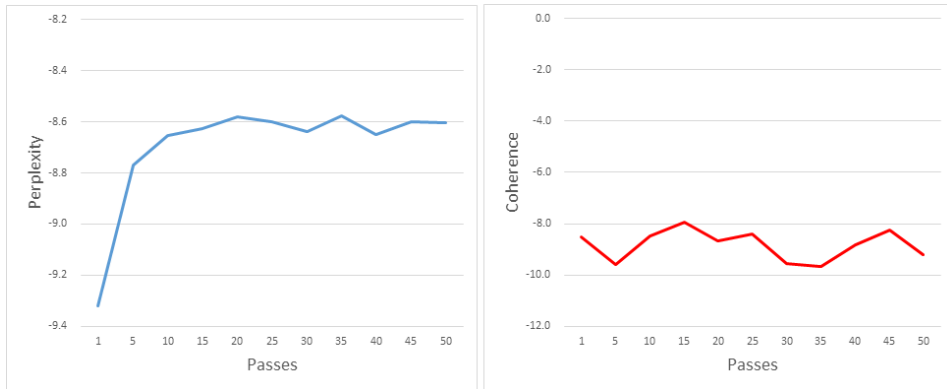


### (3) 좋은 LDA 모델을 위한 passes와 num\_topics 찾기

말뭉치 데이터에서 10,669개의 BoW가 수집되었다. <표 6>과 [그림 35]의 분석 결과 passes의 개수가 15개 일 때 perplexity는 - 8.62, coherence는 - 7.93으로 가장 낮았으므로 적절한 passes의 개수는 15로 설정하였다.

<표 6> 게임 B의 passes 선정 평가

Number of Topics : 10		
Passes	Perplexity	Coherence
1	-9.31848043974155	-8.49800841528890
5	-8.77001429455755	-9.59882721207194
10	-8.65442021577062	-8.46122019344012
15	-8.62673437305226	-7.93472841293811
20	-8.58045083050569	-8.67840348995151
25	-8.59859089661216	-8.38498516366664
30	-8.63747104711185	-9.56687639059628
35	-8.57865062825531	-9.68541707627071
40	-8.64928919940359	-8.80438441845706
45	-8.59846832780244	-8.24330998271807
50	-8.60244526143835	-9.18971886585067

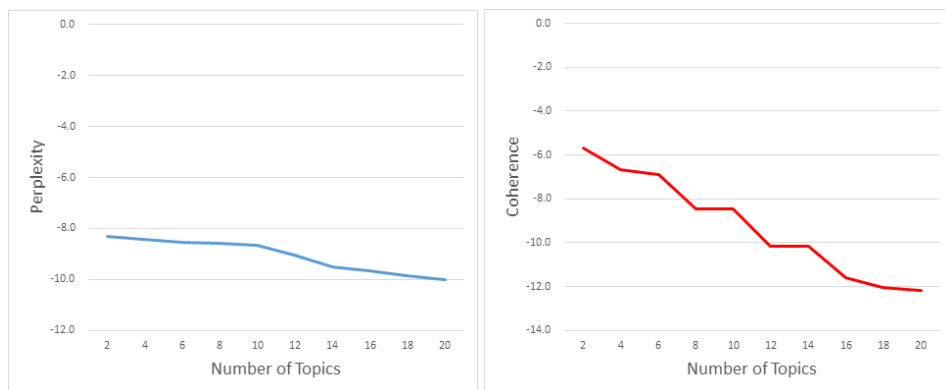


[그림 35] 게임 B의 passes 선정 평가

passes를 15로 설정하고 실행한 테스트에서는 <표 7>과 [그림 36]의 분석 결과 최적의 num\_topics는 2개로 확인되었다. 하지만 토픽의 수가 너무 적어 다양한 의견을 얻을 수가 없어서 범위를 넓히는 시도를 하였다. 토픽이 6개가 넘으면 급격하게 Coherence Score가 나빠지는 것으로 나타났으므로 6개의 토픽으로 LDA 모델을 테스트한다.

<표 7> 게임 B의 num\_topics 선정 평가

Number of Passes : 15		
num_topics	Perplexity	Coherence
2	-8.31320014025449	-5.69439681977374
4	-8.44368906022285	-6.68048217768729
6	-8.55156797601043	-6.87824728929037
8	-8.59752437207391	-8.46588643426722
10	-8.64844891496094	-8.45432001141816
12	-9.05539301884107	-10.14292383889440
14	-9.50822058526511	-10.14999271660070
16	-9.67522907694728	-11.61042831152570
18	-9.84623125749856	-12.04029213731690
20	-10.01449098372010	-12.17935993463980

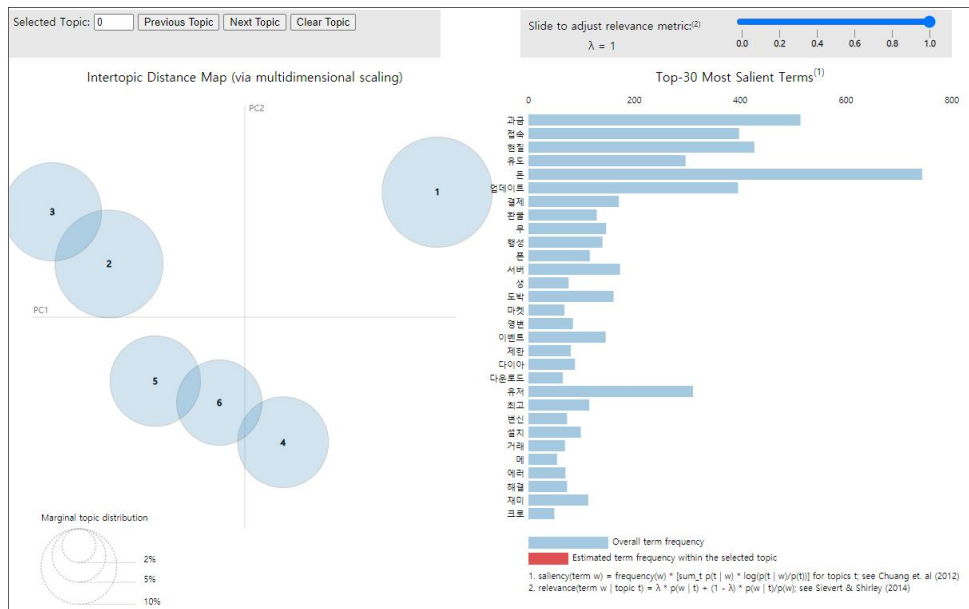


[그림 36] 게임 B의 num\_topics 선정 평가

### 3) 시각화 결과

passes는 15, num\_topics는 6의 파라미터로 LDA 평가를 실시하였으며 [그림 37]과 같은 결과를 얻었다.

#### (1) Top-30 Most Salient Terms

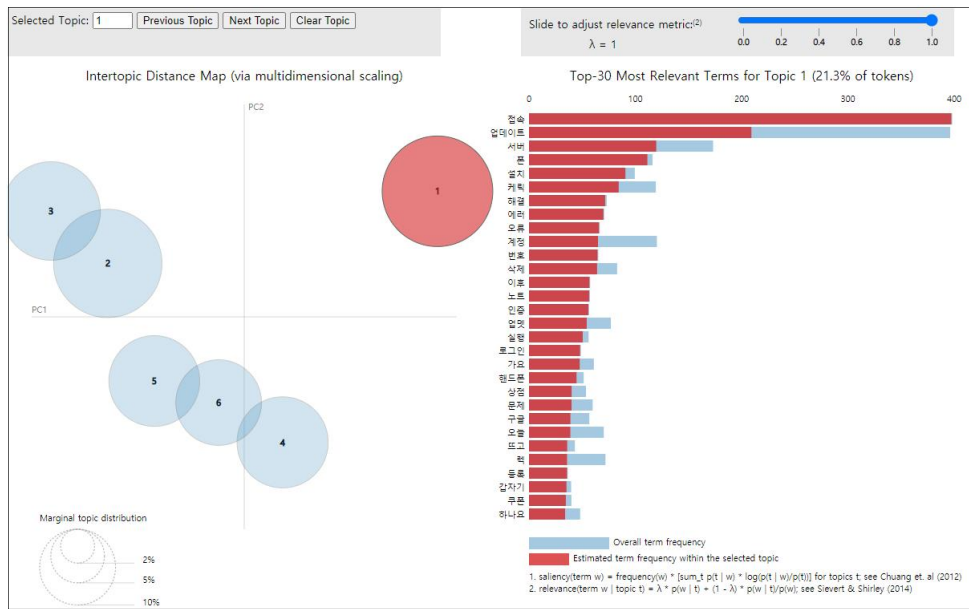


[그림 37] 게임 B의 LDA Visualization 결과

전체 문서에서 가장 두드러진 용어는 다음과 같다. 게임을 특정할 수 있는 용어는 삭제하였다.

과금, 접속, 현실, 유도, 돈, 업데이트, 결제, 환불, 무, 폰, 서버, 도박, 마켓, 이벤트, 제한, 다이어, 다운로드, 변신, 설치, 거래, 에러, 해결, 재미

## (2) Top-30 Most Relevant Terms for Topic 1



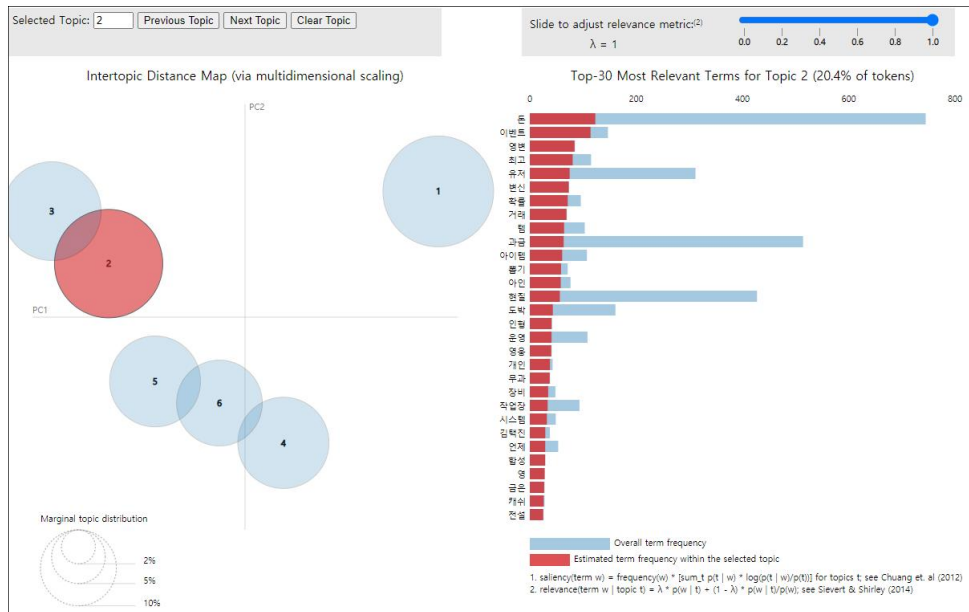
[그림 38] 게임 B의 Topic 1에 대한 결과

토픽 1의 결과는 [그림 38]에서 보는 바와 같다. 가장 관련성이 높은 단어들은 다음과 같다.

접속, 업데이트, 서버, 폰, 설치, 해결, 에러, 오류, 계정, 번호, 삭제, 인증, 실행, 로그인, 핸드폰, 문제, 렉

분석 결과 토픽 1은 게임의 실행과 관련된 이슈를 다루고 있다. 업데이트가 진행되어 사용자들이 몰려 접속이 원활하지 않음을 보여주는 것으로 인기 있는 게임임을 의미한다. 또한 사용자들이 몰려 플레이가 끊어지는 현상과 서버의 문제도 있는 것도 같은 맥락으로 볼 수 있다.

### (3) Top-30 Most Relevant Terms for Topic 2



[그림 39] 게임 B의 Topic 2에 대한 결과

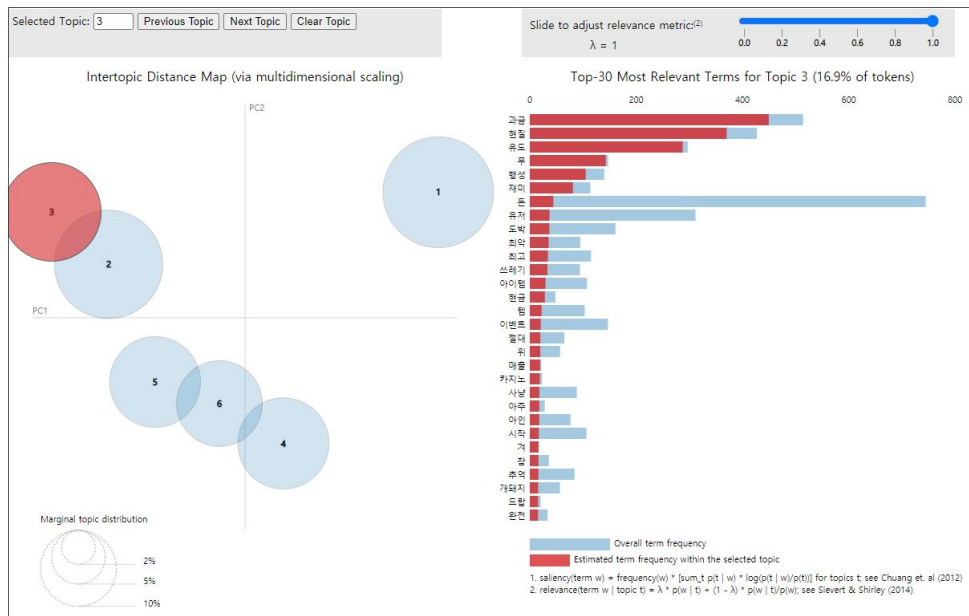
토픽 2의 결과는 [그림 39]에서 보는 바와 같다. 토픽 2에서 가장 관련성이 높은 단어들은 다음과 같다.

돈, 이벤트, 영변, 변신, 확률, 거래, 아이템, 과금, 뽑기, 현질, 도박, 운영, 무과금, 장비, 작업장, 시스템, 캐쉬

분석 결과 토픽 2는 캐릭터의 변신과 관련된 콘텐츠를 다루고 있다. 캐릭터의 변신을 위해서는 사용자들이 돈을 써서 즐겨야 하는 콘텐츠로 보이며 확률에 의해 결과가 달라질 수 있음을 알 수 있다. 이에 대해 현질, 도박, 작업

장 등 부정적인 이미지의 단어들 사용되고 있어 대체로 캐릭터 변신 콘텐츠의 과금 요소에 대해 부정적인 생각을 하고 있다고 유추할 수 있다.

#### (4) Top-30 Most Relevant Terms for Topic 3



[그림 40] 게임 B의 Topic 3에 대한 결과

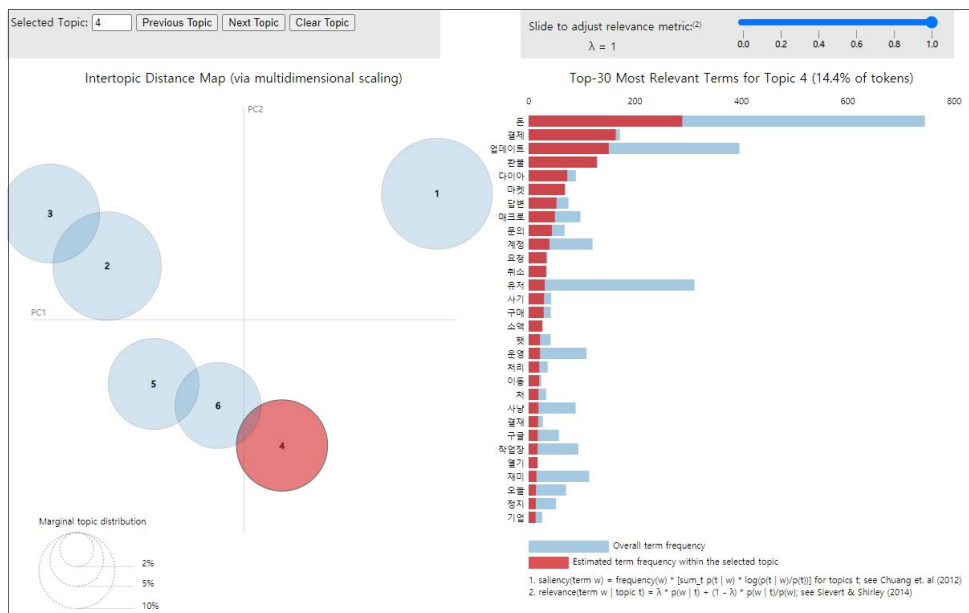
토픽 3의 결과는 [그림 40]에서 보는 바와 같다. 토픽 3에서 가장 관련성이 높은 단어들은 다음과 같다.

과금, 현질, 유도, 무, 재미, 돈, 도박, 최악, 최고, 쓰레기, 아이템, 현금, 이벤트, 매출, 카지노, 사냥, 개돼지

분석 결과 토픽 3은 게임의 전반적인 과금 유도에 대해 불만을 성토향는 내

용인 것으로 보인다. 게임의 현금 결제 유도가 도박적인 성격의 콘텐츠로 이루어져 있으며 이벤트를 통해 과금을 유도하고 있다고 보인다. 사용자들은 게임의 과금 유도에 대해 최악, 쓰레기, 개돼지라는 표현을 사용하며 매우 부정적인 인식을 가지고 있음을 알 수 있다.

## (5) Top-30 Most Relevant Terms for Topic 4



[그림 41] 게임 B의 Topic 4에 대한 결과

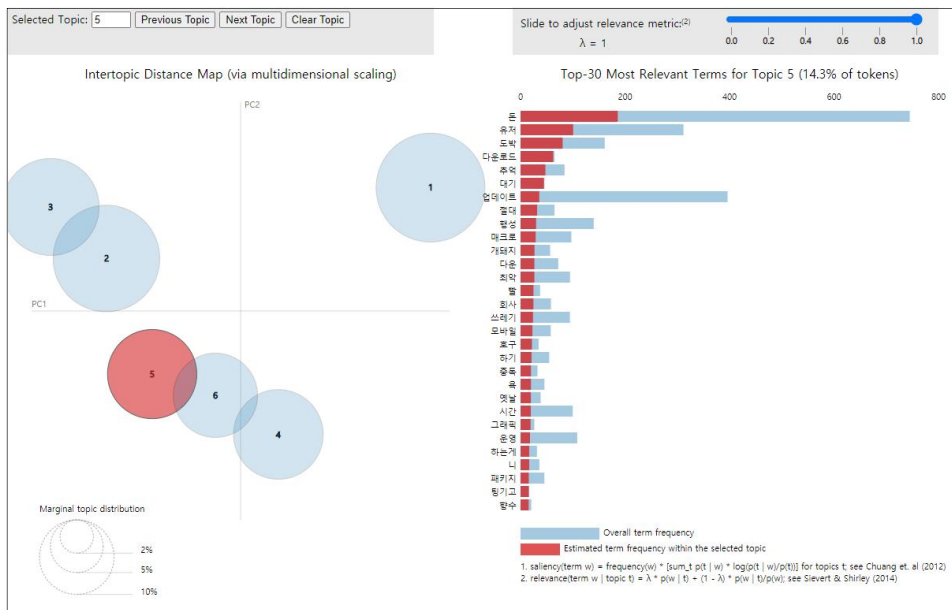
토픽 4의 결과는 [그림 41]에서 보는 바와 같다. 토픽 4에서 가장 관련성이 높은 단어들은 다음과 같다.

돈, 결제, 업데이트, 환불, 다이어, 마켓, 답변, 매크로, 문의, 계정, 요청, 취소, 사기, 구매, 소액, 운영, 처리



분석 결과 토픽 4는 결제와 환불과 관련된 내용으로 게임 콘텐츠에 대한 구매 취소에 대한 내용으로 보인다. 게임 운영에 대한 불만으로 집단으로 환불 및 구매 취소에 대한 항의가 이루어진 것으로 유추할 수 있으며 답변을 요구하거나 사기 등의 단어로 보아 사용자들이 매우 격앙되어 있는 것으로 유추할 수 있다.

## (6) Top-30 Most Relevant Terms for Topic 5



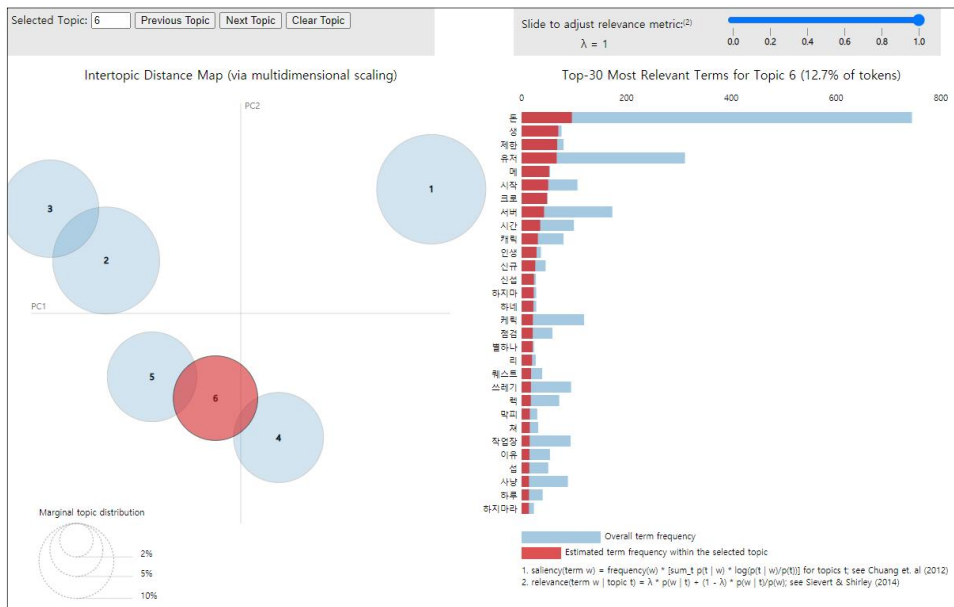
[그림 42] 게임 B의 Topic 5에 대한 결과

토픽 5의 결과는 [그림 42]에서 보는 바와 같다. 토픽 5에서 가장 관련성이 높은 단어들은 다음과 같다.

돈, 도박, 다운로드, 추억, 대기, 업데이트, 매크로, 개돼지, 다운, 최악, 회사, 쓰레기, 회사, 호구, 중독, 욕, 운영

분석 결과 토픽 5는 전반적인 게임성에 대한 평가로 보인다. 오래된 게임 IP를 이어오고 있는 게임이므로 예전에 플레이했던 사용자들을 타겟으로 게임이 만들어졌음을 의미하지만 게임의 도박성 및 과금유도에 대해 사용자들의 돈을 갈취하려는 게임 개발사에 대해 비난하는 내용으로 유추할 수 있다.

## (7) Top-30 Most Relevant Terms for Topic 6



[그림 43] 게임 B의 Topic 6에 대한 결과

토픽 6의 결과는 [그림 43]에서 보는 바와 같다. 토픽 6에서 가장 관련성이 높은 단어들은 다음과 같다.

돈, 제한, 매크로, 서버, 신규, 신섭, 점검, 캐릭터, 퀘스트, 작업장, 사냥

분석 결과 토픽 6은 게임 내에 새로운 서버가 추가된 내용으로 보인다. 일반적으로 서버가 추가되는 상황은 지속적인 사용자의 증가로 인해 쾌적한 플레이 환경을 보장하려는 개발사의 의지를 볼 수 있다. 또한 신규 서버에서의 매크로 사용에 대한 사용자들의 우려 또는 개발사의 공지가 있었던 것으로 보여 게임 전체적으로 매크로를 이용한 플레이가 매우 빈번한 것으로 유추할 수 있다.

#### 4) 토픽 분석

토픽 1은 게임의 업데이트를 사용자들이 기다리고 있으며 업데이트시 많은 사용자들이 몰려 게임 플레이가 불안해 지고 있음을 말해주는 것으로 게임이 매우 인기 있음을 보여주고 있다.

토픽 2, 3, 4, 5는 게임의 콘텐츠 및 게임의 전반적인 부분에서 과금에 대한 사용자들의 불만을 보여주고 있다. 특히 본 게임은 상대적으로 DAU는 낮지만 매출이 높은 게임으로써 게임 플레이를 위해 적극적인 과금 유도가 필수적으로 인식되고 있음을 알 수 있다. 게임의 일부 사용자는 적극적인 과금을 통해 플레이를 하고 있고, 또 일부 사용자들은 과도한 과금 유도에 대해 불만을 터뜨리고 있다는 것을 알 수 있다. 게임 내 과금 사용자와 비과금 사용자의 만족도가 매우 다를 수 있음을 의미한다.

토픽 6은 게임 내 매크로에 대한 이슈가 있음을 알 수 있다. 일반적으로

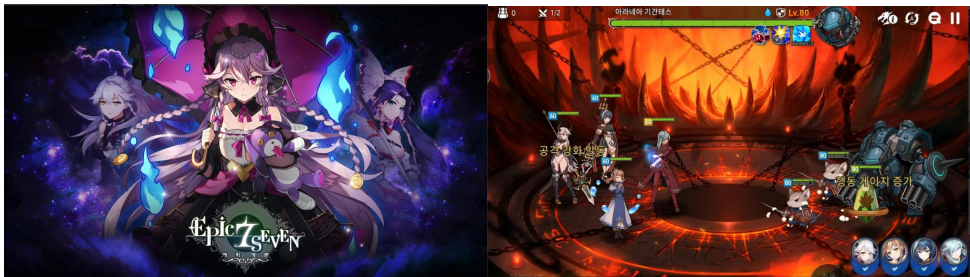
매크로를 통한 플레이는 게임 내 이득을 취할 수 있는 방법으로 알려져 있어 비용을 지불한 사용자들이 회사에 매크로에 대한 대응을 요구하는 것으로 유추할 수 있다.

전체 토픽을 분석한 결과 본 게임에 대해 사용자들은 지나친 게임의 과금 유도에 대해 매우 긍정적인 부류와 매우 부정적인 부류의 2가지 양상을 보이는 것으로 나타났다.

### 3. 사례분석 : C게임에 대해서

#### 1) 게임소개

본 사례분석에 사용된 게임은 [그림 44]에서 보이는 ‘에픽세븐’으로써 2019년 4월 구글플레이 한국 게임 매출 순위 20위 권 내의 RPG이다. 통계에 의하면 DAU는 높지만 매출이 낮은 게임으로 분석되고 있다 [99].



[그림 44] 에픽세븐

#### 2) 빅데이터 분석 실험

##### (1) 크롤링

구글플레이 ‘에픽세븐’ 페이지에서 [그림 45]에서처럼 총 5,200개의 리뷰를 수집하여 1,041,315바이트의 텍스트 파일로 저장하였다[100]. 2020년 6월 28일 구글플레이의 robots.txt에 의하면 크롤링이 가능한 것으로 확인되었다. 크롤링 데이터는 2020년 6월 14일부터 2018년 10월 9일까지의 과거 5,200개를 확보하였으며, 구글플레이 사용자 리뷰는 날짜 정렬이 되지 않아 해당 기간 내 게시물을 무작위 수집하였다.

수집한 데이터는 게임 C의 리뷰로써 모두 텍스트 데이터 형식을 취하고 있다. 해당 게임을 다운로드 한 사용자에게 한해서 작성이 가능하며 한 사용자가 몇 번이든 작성이 가능하다. 별도의 제목을 작성할 필요는 없으며 해당 게임에 대해 하고싶은 말이라면 무엇이든지 작성이 가능하다.

재미있음 처음했을때 에피소드 초반에서 막혀서 지웠다가 복귀이벤트로 장비를 주니까 스토리 연달까지 잘 봤습니다 그래픽이랑 캐릭터 블리타가 정말 좋은데 내 영웅들이 어느정도 장비 스펙이 있어야 메인서비스스토리로마톤pvp 등 컨텐츠를 제대로 즐길수있는게 단점 초반 미궁에서 얻는 장비들의 레벨을 2배 이상 올리고 속도치명공생명력 같은 범용성 높은 세트옵션을 달아서 미궁장비를 쓸만하게 만들었으면 좋겠습니다 진입장벽만 더 줄이면 초보자도 부담없이 재미있게 즐길 수 있다고 생각합니다 에픽세븐 화이팅

일단 게임성은 괜찮은데 뽑기시스템이 좀 나사가 빠진 것 같다 일단 아티랑 영웅이랑 붙어있는 건 넘기더라도 신비소환이나 월광소환이 상태가 많이 안 좋다 신비소환은 200회 뽑기를 하면 월광5성을 확정으로 주는데 돈으로 뽑는다면 신비 패키지 62만7천원으로 다 구매해도 138뽑기만하고 월광뽑기는 20번 뽑으면 하나주는 금빛 전승석 6개를 모으면 뽑기를 할 수 있는데 70퍼센트로 노율이 대부분인 월광3성이 뜨고 월광45성도 하자있어서 못쓰는게 있어서 잘 뽑아도 답없는게 뜨기도 한다 그리고 일반뽑기는 4성영웅도 뽑기확률이 5퍼센트가 안된다 말그대로 운빨존망록우점인 듯 그래서 웬만한면 계정사서 시작하는게 좋다고 생각합니다

원래 모바일게임에는 흥미가 아예없었는데 이 게임은 진짜 너무 재밌네요 유일하게 하는 모바일게임이 이게임인거 같아요 일단 스토리도 너무 마음에 들거 저같은 무과금 유저도 할수있게 서비스해주시는건 진짜 좋은거 같아요 그리고 자주자주 업데이트하는 것도 좋은거 같고 이번에 키즈나 아이 나온것도 뭐 중국 관련해서 나름 논란이 있는것 같지만 이미 키즈나 아이 중국 버전은 중국이 사갔고 그래서 중국이 멋대로 정치 성향을 내도 특별히 키즈나 아이쪽 잘못이 아니기 때문에 전 이번 업데이트 관찰하고 봅니다 요즘 심심치 않게 악플들이 보이는데 신경쓰지마시고 화이팅 하세요

게임스토리나 캐릭터디자인컨텐츠 다좋은데 5성이나 월광45성 확률이 너무극악이라 나올거라는기대도안된다 뽑기에서 아키텍트랑 영웅이랑 같이나오는것도 맘에안들고 쟈 답없는건 월광뽑기한번하려고 뽑기120번해서 모은거했는데 월광에서 45성나올확률이 그냥 돈벌아먹으려고 만든게임으로밖에안보입니다

나름 재밌게 플레이 하고있는 유저임 정점은 무과금 소과금으로도 충분히 플레이 할 수 있다고 생각함 본인은 무과금으로 에피소드1 클리어 6성승급 전직 등 복귀 하고서도 모험 외에 다양한 컨텐츠를 즐기면서 플레이 하고 있음시간은 조금 걸리지만 무과금으로 초고속 육성은 바라면 안됨 성역 전투 컨텐츠로 크리스탈 자금자축 중분하고 첩터 클리어 보상으로 나쁘지 않다고 생각함 단점은 캐릭터 및 아티팩트 소환물 극악의 축매체 드랍률 등 대표적으로 꼽아보자면 이정도임 재화도 종류가 많아서 나도 모르게 상점에서 재화별로 구매할 수 있는 것들이 많아서 놀랐음 토틀로 재료 얻어서 장비 제작 가능하고 전설등급도 나오다보니 장비는 좋은거 항상 꺼줄 다른 캐릭터 전용장비도 없네 나왔으면 좋겠네

가차게임 중에서도 손꼽히는 랜덤성을 가진 게임 이게임은 p2w가 아니라 lucky2w 게임이다 그만큼 만족스러운 상황보다 불만족스럽고 스트레스받는 일이 많다 특히 장비 캐릭도 마찬가지로 선택권 개념이 없고 언제 나올지도 모르는 픽업을 계속 기다려야하고 나온다해도 비용이 만만치 않은편

요새 개선 많이되었습니다 뉴비를 위해서 장비도 정말 좋은 템으로 지원해주고 편의성개편 스킬연줄 정말 맘에들입니다 하지만 아직 문제점인 밸런스패지 월광5성중복개선안등 유저들이 계속연급해오는것에 비해 너무 느리게 개선되어지고 있다는것 등등 개선이되고있지만 조금 부족한부분이 없을순없지만 차차 대응도 빨라지기 바라며 제발 여기서 더나빠지지않도록 운영해주세요 화이팅

꾸준히 열심히 무과금으로 하실 분 하세요 게임 재밌고 하위 영웅들로도 5성 영웅이 하는거 다 할 수 있어요 다만 돈 쓰고 싶으시면 이게임 하지 않으시는 걸 추천합니다 장점이자 단점인데 심심하면 패키지 과금 유도는 겁나게 해대는데 돈 써서 얻는 만족도는 역대 최악이고 패키지 구성이 개창렬인데다 메인 소비처인 뽑기는 확률 씹거지예 오래 한 사람들이 더 유리하게 만들어져 있습니다 애정값고 100만원 지른 계정 보다 리세계 2만원짜리가 더 영웅 풀이 다양하고 리치해요 천장이 생겨서 뽑기난이도가 내려갔다고는 하는데 한정캐릭만 그럽고 120회면 돌이 12000개 3만원에 돌 1500개입니다 120회마다 뽑는 월광 캐릭터는 쓸만한 캐릭터는 웬만하면 45성이라 땅이 수두룩함 다만 진득히 오래하면 그만큼 얻는다는 이야기이기도 하니 저처럼 돈쓰고 부들부들하지 말고 무과금 소과금 하세요 자살충동 읊니다

[그림 45] 게임 C의 리뷰 수집

## (2) 토크닝

[그림 46]에서처럼 앞서 수집한 리뷰를 토크닝 처리하였다. 명사와 동사만 수집하고 불용어를 정리하여 605,572 바이트의 텍스트 파일로 저장하였다.

했을 에피소드 초반 막혀서 지워다가 복귀 이벤트 장비 주 스토리 엔딩 봤습니다 그래픽 캐릭터 퀄리티 영웅 장비 스펙 메인 서브스토리  
오토마톤 컨텐츠 제대로 즐길수있는게 단점 초반 미궁 얻는 장비 레벨 배 이상 올리고 도치 생명력 범용 세트 옵션 달아서 미궁 장비 쓸  
만하 만들었으면 진입 장벽 줄이면 초보자 부담 줄릴 화이팅

일단 성은 뽑기 시스템 나사 빠진 일단 아티 영웅 불러있는 넘기더라도 신비 소환 월광 소환 상태 신비 소환 뽑기 월광 성 확정 주는데  
돈 뽑는다면 신비 패키지 구매 뽑기 월광 뽑기 뽑으면 주는 금빛 전승 석 모으면 뽑기 퍼센트 노 쓸이 대부분 월광 성 뜨고 월광 성도  
하차있어서 쓰는데 뽑아도 답 드 기도 뽑기 영웅 뽑기 확률 퍼센트 그대로 운빨 존망 폭우 계정 사서 시작 하는게

원래 모바일 흥미 아예 재밌네 모바일 일단 스토리 마음 들구 무 과금 유저 할수있게 서비스 해주신건 업데이트 키즈나 아이 나온것도  
중국 관련 나름 논란 키즈나 아이 중국 버전 중국 사갔고 중국 멋대로 정치 성향 내도 키즈나 아이 쪽 잘못 업데이트 뽑니다 심심 치  
않게 억울 보이는데 신경 쓰지 마시고 화이팅 하세요

스토리 캐릭 디자인 컨텐츠 성 월광 성 확률 극악 나올거라는 기대도 뽑기 아키텍트 영웅 나온것도 맘 절 답 월광 뽑기 뽑기 모은거  
셋 는데 월광 성나 확률 돈빨 먹으려고 만든 보입니다

나름 하고있는 유저 장점 무 과금 과금 본인 무 과금 에피소드 클리어 승급 전직 복귀 하고서도 모험 외 컨텐츠 즐기면서 시간 걸리지만  
무 과금 초 고속 육성 바라면 성역 전투 컨텐츠 크리스탈 자급자족 첵 클리어 보상 없다고 단점 캐릭터 및 아티 팩트 환율 극악 축  
매제 드랍률 대표 곱 보자면 재화 종류 나도 모르게 상점 재화 별로 구매 놀랐음 토벌 재료 얻어서 장비 제작 전설 등급 나오다 보니  
장비 항상 개 중 다른 캐릭터 전용 비드 얼른 나왔으면

가차 손꼽히는 런엄 가진 가 상황 불만족 스트레스 받는 장비 캐릭 마찬가지로 선택 개념 언제 나올지도 모르는 픽업 기다려야하고 나온  
다해 비용 만 얹은 편

요새 개선 되었습니다 뉴비 위해 장비 템 지원 해주고 편의 개선 스킬 연출 맘에드니 문제점 별론 패지 월광 중복 선안 유저 언급  
해오는것에 비해 개선 되어지고 등등 개선 되고있지만 손 차차 대응 지기 바라며 제발 여기 나빠지지 않도록 운영 화이팅

무 과금 하실 하세요 하위 영웅 성 영웅 하는거 돈 쓰고 싶으시면 얹으시는 추천 장점 이자 단점 패키지 과금 유도 겁나게 해대 돈  
써서 얻는 만족도 역대 최악 패키지 구성 개장 얼 메인 비처 뽑기 확률 씹거지예 만들어져 애정 갖고 지른 계정 보다 리세 영웅 풀이  
리치 해오 천장 생겨서 뽑기 난이도 내려갔다고는 한정 캐릭 돌이 돌 개입 니 뽑는 월광 캐릭터 쓸만 캐릭터 성 광 수두 진득 얻는다는  
이야기 하니 돈 쓰고 무 과금 과금 하세요 자살 중

서머 워 유저 서머 워 비해 확률 소환 재화 구 하기가 쉬움 무엇 재미 류 가장 중임 서머 는 망점 거르시길

뽑기 좋아짐 확정 시스템 느낌 이긴하지만 성취 확률 기대는것 보다 확정 영웅 얻는다는 의의 큼 월광 뽑에서나 오고 디자인 말다 했고  
방송 통해 소통 의견 반영 개발 진의 노고 느껴짐 핵 차단 영웅 밸런스 잡으면 될거 상향 오피 캐 오피 잡는 시스템 개선 되고  
캐릭터 활용 다양 활성화 해주고 활용 디 열순 맺 성 내놔서 자동 무제한 넣어주면 영웅 장비 칠때 사냥 넘사벽 포기 메인 서브 스토리  
재미 토벌 단 처럼 재밌는 컨텐츠 파이팅

수집 단계 이점 여러가지 밸런스 운영 보이긴 최근 할러는 성의 보임 아레나 쪽 갈려면 구 하기 캐릭터 올라가긴 캐릭터 잡고 애정 캐  
쓰면 캐릭터 매력 최근 걸질 걷는 중이 시작 후회 할듯 어서 들어와서 라비 맞춘 가라구

확률 몹시 구멍 운빨 될수도 뽑기 뽑는데 성 안돌 뽑기 확률 누적 몇번 뽑으면 성 영웅 아티 팩트 확정 추가 했음 뽑기 줬음  
달아요 모바일 개발자 했나 봐요 편의 캐릭 성 키우는데 시간 일주일 넘도록 해봤어요 대체 시간 동안 잡고 해야하냐요 반복 해야해  
조렘텐 심지어 반복 사냥 펫 얻는 다음 반복 사냥 되는데 그것 펫 올려야 판 문제 사냥 장비 잡 캐릭 무니 주머니 들어가 공간 돈  
써가지고 놀려야 맘 굴드 강화 및 장비 아티 교체 비용 모되 얻을려면 아우 뽑기 비용 성 얻어도 언제 키워 천장 찍고 아티 뽑고 캐릭  
얻었는데 나오게 캐릭 천장 토벌 해야하고 미궁 해야하고 사료 이구 자동 스킬 갓 하더니 리세 계정 산 해당 나도 그거 구입 그계 월광  
완전 도박 잭라 전직 키울수 해주지 는 뉴비는 뽀빠 스토리 난이도 성 라이거 뿌려라 스킬 찍을때 축메제 얻는것도 전직 장비 파는  
드레그 일일이 놀러 가야

오픈 초기 하다 최근 복귀 유저 예전 뽑았던 키워던 캐릭 거의 사라졌네요 중간 일이 모르겠지만 월광 나름 당시 만렙 키워 돈 캐릭  
사라져있으니

치즈 터지 기전 집고 복귀 한지 그때 운영 나아짐 방송 유저 소통 러고 이벤트 가끔 창렬 보상 예전 유저 대적 시절 느낌 들 컨텐츠  
는 할게 수준 배 개 서라도 냈으면

분명 여러가지 문제 흥여 않았었고 초기 문제 심각 인지 했지만 달라진 태도 유저 마음 읽고 소통 나아지려고 노력 과금 역시

[그림 46] 게임 C의 토크닝 처리

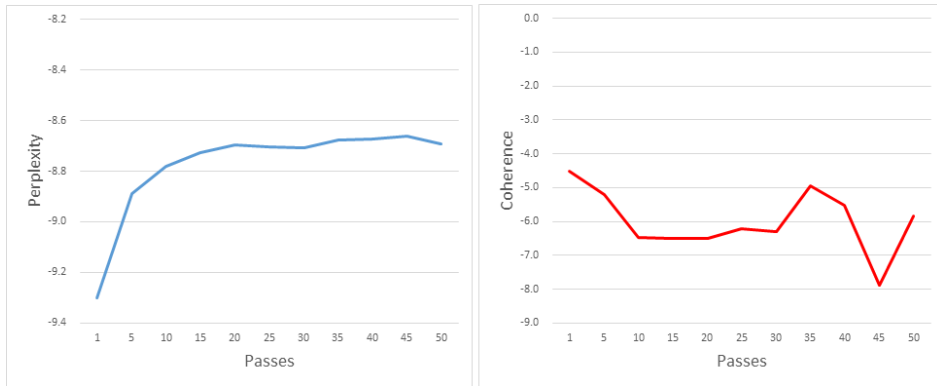
### (3) 좋은 LDA 모델을 위한 passes와 num\_topics 찾기

말뭉치 데이터에서 14,256개의 BoW가 수집되었다. <표 8>과 [그림 47]의 분석 결과 passes의 개수가 35개 일 때 perplexity가 - 8.67, coherence가 - 4.95로 가장 낮았으므로 적절한 passes의 개수는 35로 설정하였다.

<표 8> 게임 C의 passes 선정 평가

Number of Topics : 10		
Passes	Perplexity	Coherence
1	-9.30015333842946	-4.52193500518477
5	-8.88753465861192	-5.19648987675470
10	-8.78056584759916	-6.46376812356079
15	-8.72790287029703	-6.48888605020543
20	-8.69741502843649	-6.51072842414513
25	-8.70285736677712	-6.21316244593962
30	-8.70937470747556	-6.31258748688697
35	-8.67730072048907	-4.95217749759635
40	-8.67308347418137	-5.52898198562416
45	-8.66333221723036	-7.88315796945259
50	-8.69047091106161	-5.83359856649781



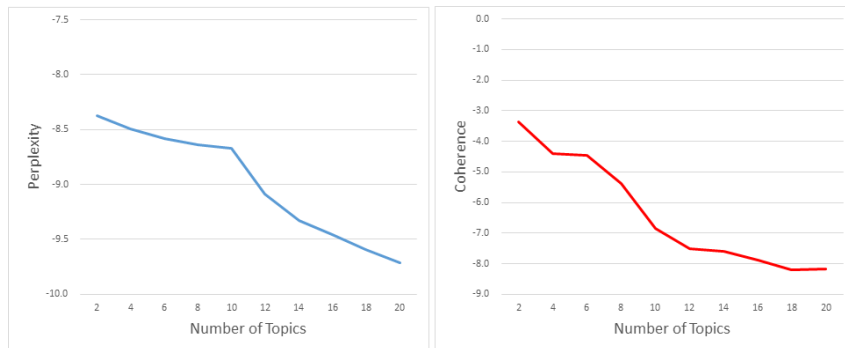


[그림 47] 게임 C의 passes 선정 평가

passes를 35로 설정하고 실행한 테스트에서는 <표 9>와 [그림 48]의 분석 결과 최적의 num\_topics은 2개로 확인되었다. 하지만 토픽의 수가 너무 적어 다양한 의견을 얻을 수가 없어서 범위를 넓히는 시도를 하였다. 토픽이 6개가 넘으면 급격하게 Coherence Score가 나빠지는 것으로 나타났으므로 6개의 토픽으로 LDA 모델을 테스트한다.

<표 9> 게임 C의 num\_topics 선정 평가

Number of Passes : 35		
num_topics	Perplexity	Coherence
2	-8.37551556012199	-3.37063384926481
4	-8.49069146948325	-4.40431232273009
6	-8.58538627749991	-4.45248126355892
8	-8.63821745567219	-5.36503480568969
10	-8.67221358079873	-6.85673520801466
12	-9.08385150527451	-7.50661593725249
14	-9.32650190973797	-7.60581039029182
16	-9.45376910737312	-7.89382152743748
18	-9.59204750202744	-8.19974913852788
20	-9.71483134497554	-8.17611892723626

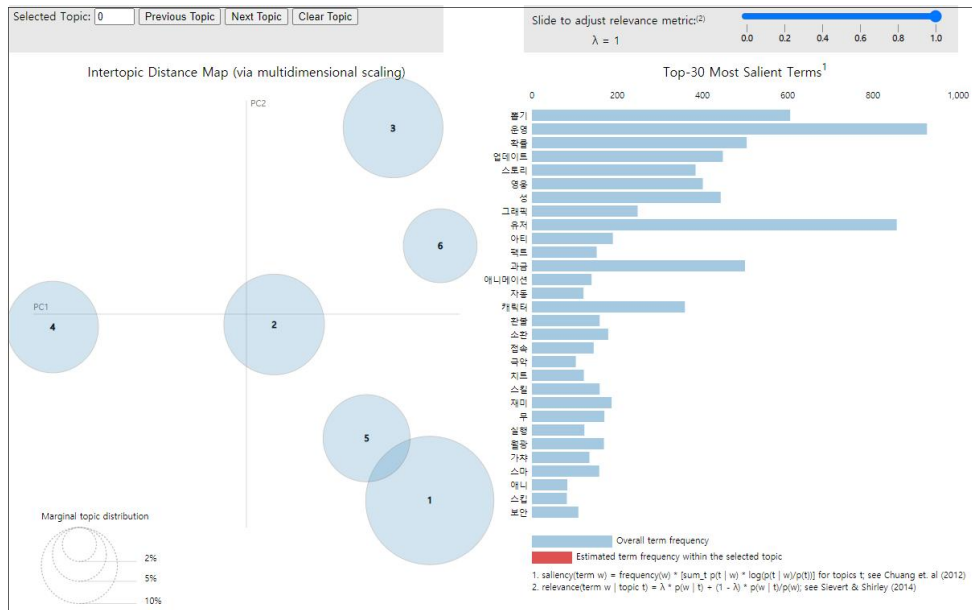


[그림 48] 게임 C의 num\_topics 선정 평가

### 3) 시각화 결과

passes는 35, num\_topics는 6의 파라미터로 LDA 평가를 실시하였으며 [그림 49]과 같은 결과를 얻었다.

#### (1) Top-30 Most Salient Terms

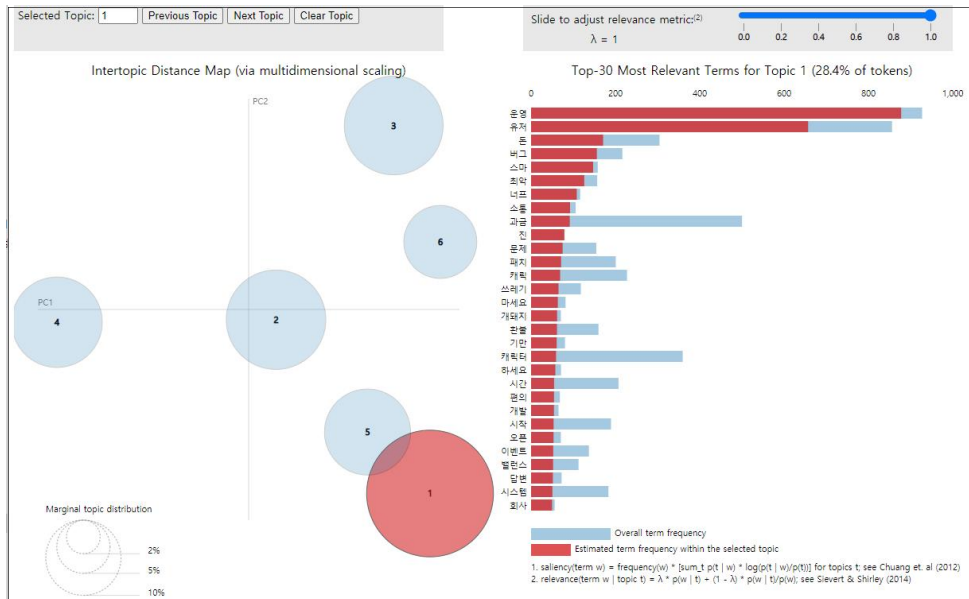


[그림 49] 게임 C의 LDA Visualization 결과

전체 문서에서 가장 두드러진 용어는 다음과 같다. 게임을 특징할 수 있는 용어는 삭제하였다.

뽑기, 운영, 확률, 업데이트, 스토리, 영웅, 성, 그래픽, 유저, 아티팩트, 과금, 애니메이션, 자동, 캐릭터, 환불, 소환, 접속, 극악, 치트, 스킬, 재미, 무, 실행, 가차, 보안

## (2) Top-30 Most Relevant Terms for Topic 1



[그림 50] 게임 C의 Topic 1에 대한 결과

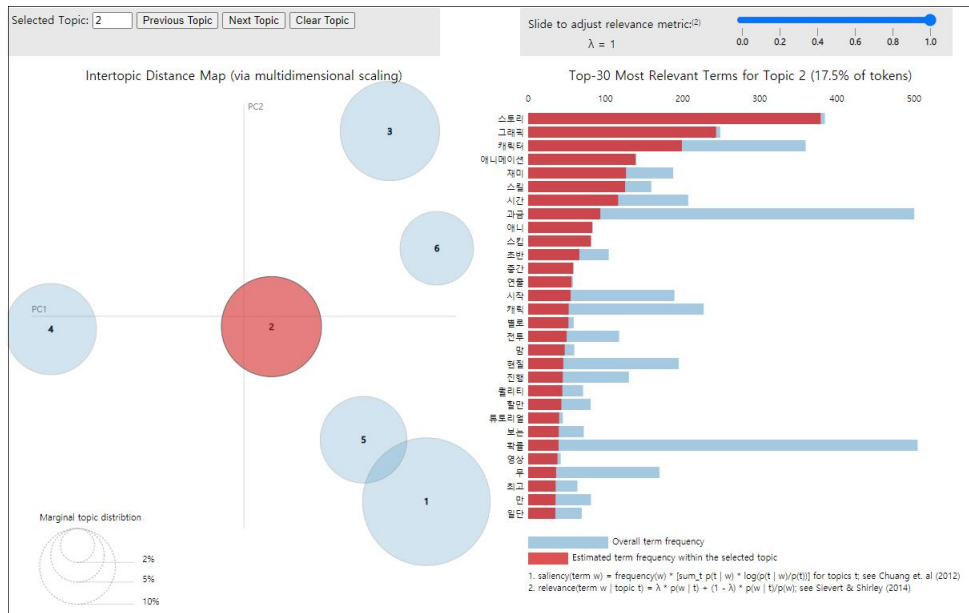
토픽 1의 결과는 [그림 50]에서 보는 바와 같다. 토픽 1에서 가장 관련성이 높은 단어들은 다음과 같다.

운영, 유저, 돈, 버그, 최악, 너프, 소통, 과금, 문제, 패치, 쓰레기, 개돼지, 환불, 기만, 캐릭터, 시간, 편의, 개발, 오픈, 이벤트, 밸런스, 답변, 시스템, 회사

분석 결과 토픽 1은 게임의 운영 이슈에 대한 내용을 다루고 있다. 게임 내 버그와 밸런스 문제로 잦은 패치와 밸런싱 작업이 사용자들에게 스트레스를 주고 있음을 알 수 있다. 또한 이 과정에 사용자들과의 소통이 잘 이루어지지 않고 있으며 이에 대해 사용자들은 환불 등의 이슈

를 제기하고 있음을 알 수 있다.

### (3) Top-30 Most Relevant Terms for Topic 2



[그림 51] 게임 C의 Topic 2에 대한 결과

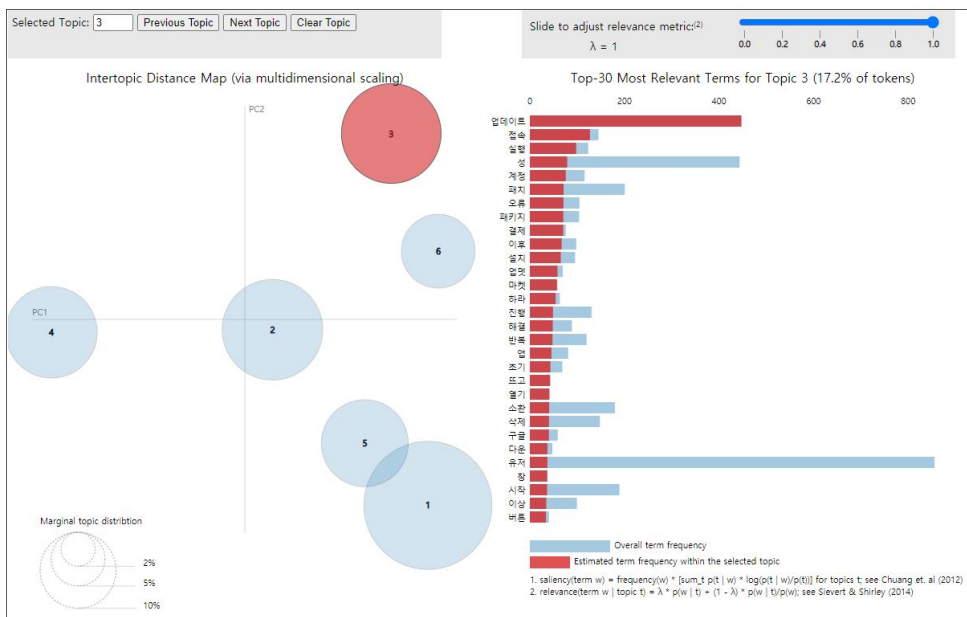
토픽 2의 결과는 [그림 51]에서 보는 바와 같다. 토픽 2에서 가장 관련성이 높은 단어들은 다음과 같다.

스토리, 그래픽, 캐릭터, 애니메이션, 재미, 스킬, 시간, 과금, 스킵, 초반, 중간, 연출, 별로, 전투, 현질, 진행, 퀄리티, 튜토리얼, 확률, 영상

분석 결과 토픽 2는 게임의 콘텐츠에 대한 평가가 주를 이루고 있으며, 게임의 스토리와 그래픽, 캐릭터, 애니메이션 등에 대해 많은 의견

이 개선되고 있음을 알 수 있다. 다만, 게임 내의 튜토리얼이나 영상 연출 등의 퀄리티가 매우 낮아 이를 건너뛰고(스킵)자 하는 욕구가 있음을 알 수 있다. 본 게임의 개발자들은 게임의 스토리 요소를 연출로 녹여 게임의 강점으로 삼고자 했으나 충분히 사용자들에게 어필하지 못하는 것으로 유추할 수 있다.

#### (4) Top-30 Most Relevant Terms for Topic 3



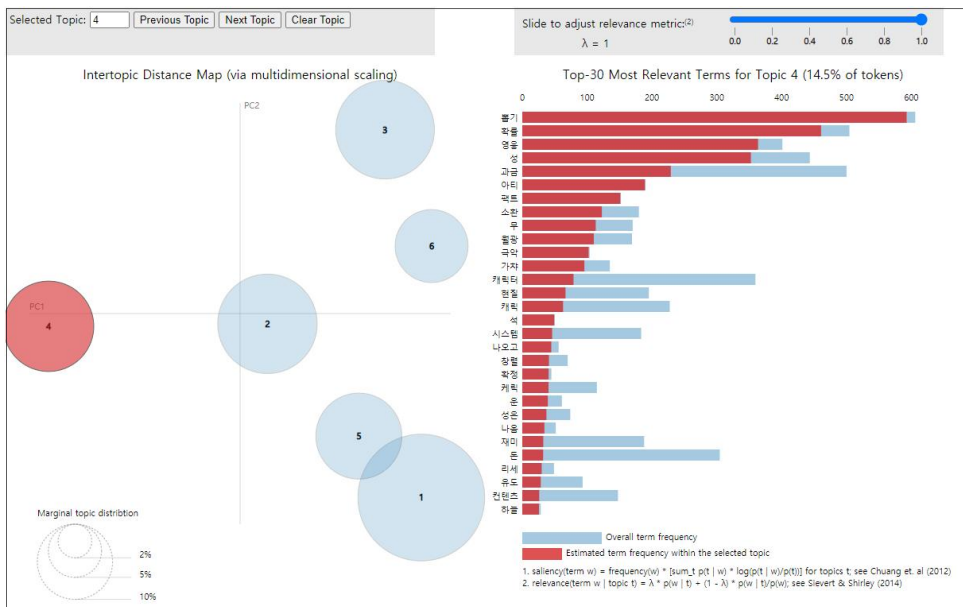
[그림 52] 게임 C의 Topic 3에 대한 결과

토픽 3의 결과는 [그림 52]에서 보는 바와 같다. 토픽 3에서 가장 관련성이 높은 단어들은 다음과 같다.

업데이트, 접속, 실행, 성, 계정, 패치, 오류, 패키지, 결제, 설치, 마켓, 진행, 해결, 반  
복, 앱, 소환, 삭제, 다운, 이상

분석 결과 토픽 3은 게임의 실행과 관련되어 문제가 있음을 보여준다. 업데이트나 패치 등이 진행되며 오류가 빈번하게 발생하고 있고 이 때문에 게임을 삭제하는 경우도 있음을 알 수 있다. 다만, 게임의 DAU가 높기 때문에 게임의 원활한 실행을 바라는 사용자들의 마음이 반영되어 게임에 대한 인기를 보여 주는 것으로 해석할 수 있다.

## (5) Top-30 Most Relevant Terms for Topic 4



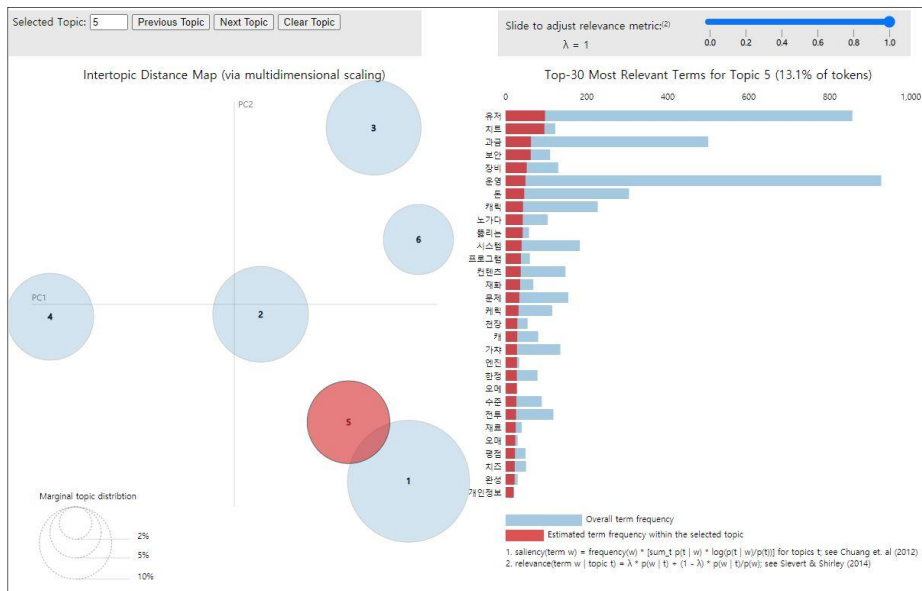
[그림 53] 게임 C의 Topic 4에 대한 결과

토픽 4의 결과는 [그림 53]에서 보는 바와 같다. 토픽 4에서 가장 관련성이 높은 단어들은 다음과 같다.

뽑기, 확률, 영웅, 성, 과금, 아티팩트, 소환, 무, 극악, 가차, 캐릭터, 현실, 시스템, 창  
 렬, 확정, 운, 재미, 돈, 유도

분석 결과 토픽 4는 게임의 과금과 관련된 불만이 나타난 것으로 보인다. 특히 게임 내 캐릭터의 뽑기에 확률이 매우 낮게 적용되어 있어 이에 대한 사용자들의 불만이 매우 높은 것으로 나타났다. 본 게임은 DAU가 높지만 상대적으로 매출이 낮은 것으로 분석되었는데 과금 요소가 사용자에게 불합리하게 적용되어 있는 것으로 유추할 수 있다.

## (6) Top-30 Most Relevant Terms for Topic 5



[그림 54] 게임 C의 Topic 5에 대한 결과

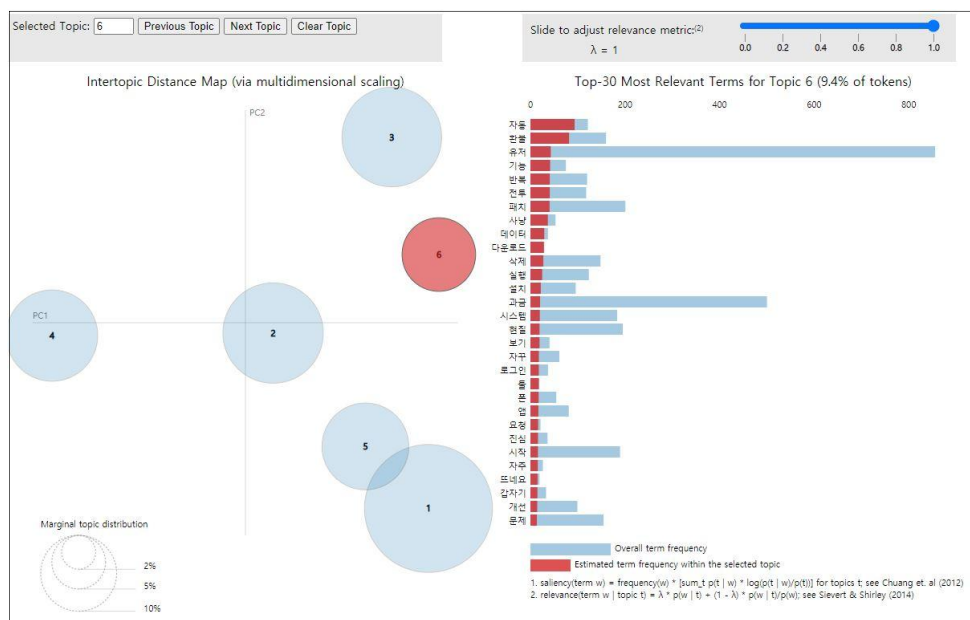


토픽 5의 결과는 [그림 54]에서 보는 바와 같다. 토픽 5에서 가장 관련성이 높은 단어들은 다음과 같다.

치트, 과금, 보안, 장비, 운영, 돈, 캐릭터, 노가다, 시스템, 재화, 콘텐츠, 문제, 가차, 한정, 수준, 평점, 개인정보

분석 결과 토픽 5는 게임의 보안에 관련된 이슈로 보인다. 과금을 하지 않아도 치팅(cheating)을 통한 재화의 비정상적인 획득 등에 문제가 있는 것으로 유추할 수 있다.

## (7) Top-30 Most Relevant Terms for Topic 6



[그림 55] 게임 C의 Topic 6에 대한 결과

토픽 6의 결과는 [그림 55]에서 보는 바와 같다. 토픽 6에서 가장 관련성이 높은 단어들은 다음과 같다.

자동, 환불, 기능, 반복, 전투, 패치, 사냥, 데이터, 삭제, 실행, 설치, 과금, 현실, 로그인, 요청, 개선
------------------------------------------------------------------

분석 결과 토픽 6은 게임의 반복되는 전투와 사냥 콘텐츠에 대한 불만이 나타난 것으로 보인다. 온라인으로 서비스되는 RPG에서 반복적인 전투와 사냥은 필수불가결한 요소라고 여겨지며, 이를 반복 실행해야 하는 사용자에게 적극적인 동기부여를 할 수 있는 요소가 필요하지만 본 게임에서는 사용자들이 전투와 사냥에 대해 큰 매력을 느끼지 못하고 있음을 알 수 있다.

#### 4) 토픽 분석

토픽 1, 3을 통해 전반적으로 게임의 운영적인 부분에서 게임 개발사와 사용자들이 소통을 제대로 하지 못하고 있음을 알 수 있다. 하지만 게임의 인기가 높다는 것을 반증하는 것으로 보인다.

토픽 2는 본 게임의 캐릭터, 그래픽 등의 비주얼적인 콘텐츠가 강점인 것으로 보여 많은 사용자들이 게임을 좋아하고 있는 것으로 나타났다. 하지만 콘텐츠를 강제로 사용자들에게 전달하고자 하는 욕구에 사용자들의 반발이 있는 것으로 보인다.

토픽 4는 사용자들이 게임의 과금요소에 대해 불합리함을 느끼고 있으며 이로 인해 게임의 매출에 부정적인 영향을 주고 있음을 알 수 있다. 전형적으로 게임성은 부족하지만 비주얼은 예쁜 게임으로 여겨진

다.

토픽 5는 게임에서 비정상적인 방법으로 재화 획득이 이루어지고 있음을 보여준다. 토픽1의 운영의 불만과 연계하면 게임 내에 벌어지는 이상 현상에 대해서 개발사가 적극적인 대응을 하지 못하고 있음을 나타낸다.

토픽 6은 전투 및 사냥 콘텐츠에 대해 사용자들이 만족감을 느끼지 못하고 있음을 보여준다. 게임정보다는 콘텐츠가 좋은 게임으로 볼 수 있다.

전체 토픽을 분석한 결과 본 게임은 뛰어난 비주얼을 바탕으로 높은 인기를 끌고 있지만 상대적으로 게임성은 이를 따라주지 못하는 것으로 나타났다. 또한 과금을 통해 매우 낮은 확률로 보상을 얻는 요소는 적극적인 과금을 방해하고 있는 것으로 나타났다.

## 4. 종합 분석

### 1) 분석의 객관성 확보

앞선 사례분석에서 객관적인 해석을 위해 분석자가 게임에 대한 정보를 최소화 할 수 있도록 분석자가 기존에 플레이하지 않았던 게임을 선정하였으며, 게임에 대한 기본적인 정보인 게임명, 매출 순위, DAU 순위만을 알고 있는 상황에서 분석을 진행하였다. 또한 사전에 게임의 리뷰를 읽게 되면 주관적인 정보가 전달될 것을 우려하여 리뷰를 읽지 않은 상태로 진행하였다.

### 2) 게임별 토픽 분석 비교

게임 A, 게임 B, 게임 C는 매출 순위 상위권에 있는 게임들로써 기본적으로 사용자들의 충성도가 높은 게임이다. 따라서 게임의 실행 불가 현상에 대한 불만이 공통적으로 나타나고 있음을 알 수 있었다. 일반적으로 충성도가 낮은 게임의 경우에는 실행 불가 현상이 진입 장벽으로 작용하여 리뷰를 남길 필요를 느끼지 못한다.

또한 세 게임 모두 공통적으로 과금에 대해 매우 많은 리뷰가 수집되었으며, 과금 요소의 적절성이 게임의 매출에 크게 영향을 주고 있음을 보여준다.

게임 A, 게임 B, 게임 C의 토픽 분석 비교 결과를 <표 10>에 정리하였다.

<표 10> 게임별 토픽 분석 비교

	토픽 분석
게임 A	<ul style="list-style-type: none"> <li>✓ DAU가 높으며 매출도 높음</li> <li>✓ 게임의 인기가 높아 실행 불가 현상에 예민하게 반응</li> <li>✓ 게임의 그래픽 품질과 캐릭터 커스터마이징이 특징</li> <li>✓ 긍정적인 매출 유도. 합리적인 과금으로 인식</li> <li>✓ 전투 콘텐츠에 대한 재미요소 확립</li> </ul>
게임 B	<ul style="list-style-type: none"> <li>✓ DAU는 낮지만 매출은 높음</li> <li>✓ 게임의 인기가 높아 실행 불가 현상에 예민하게 반응</li> <li>✓ 게임의 지나친 과금 유도가 있으나 이를 긍정적으로 생각하는 부류와 부정적으로 생각하는 부류가 공존</li> <li>✓ 과금이 영향이 큰 게임인 만큼 부정행위에 대해 민감하게 반응</li> </ul>
게임 C	<ul style="list-style-type: none"> <li>✓ DAU는 높지만 매출은 낮음</li> <li>✓ 비주얼 콘텐츠가 매우 뛰어나서 인기를 끌고 있음</li> <li>✓ 반면, 게임성은 매우 떨어짐</li> <li>✓ 과금 요소의 확률이 매우 낮아 사용자들을 적극적으로 과금하도록 유도하지 못하고 있음</li> </ul>

게임 A의 경우 리뷰의 텍스트 양은 1,050,234 바이트, 게임 B의 경우 617,210 바이트, 게임 C의 경우 1,041,315 바이트가 수집되었다. 세 게임 모두 약 5,200여개의 리뷰를 수집하였지만 리뷰에 적힌 글자의 양에서 A게임과 C게임은 유사한 반면 B게임은 두 게임에 비해 절반가량의 텍스트만을 기록했다. A와 C게임에서 사용자는 긴 텍스트를 통해

많은 정보를 전달한 것으로 보이며, B게임의 사용자는 정보량이 적어 텍스트가 짧은 단순 감상이나 불만을 털어놓았다고 유추할 수 있다. 게임 B의 토픽들에서 과금에 대한 불만이 강하게 나타나 있었던 것이 원인으로 지목될 수 있다.

게임 A와 게임 C의 경우 그래픽 요소가 토픽으로 등장하였으며 게임의 인기를 보여주는 주요 요소로써 사용자들에게 인식되고 있음을 알 수 있다. 반면, 게임 B의 경우 이전에 인기 있던 게임의 IP를 활용하였기 때문에 비주얼적인 부분에서의 유사성이 작용한 것이라 해석된다. 이는 RPG에 있어서 그래픽 또는 비주얼적인 강점이 게임 선택에 중요하게 반영되고 있음을 알 수 있다.

세 게임에서 공통적으로 발견된 토픽으로는 캐릭터와 관련된 것들이 있다. 게임 A의 경우에는 커스터마이징, 게임 B의 경우에는 변신 및 강화, 게임 C의 경우에는 캐릭터 수집(뽑기)이 주요 토픽에 등장했다. RPG에서 사용자들이 캐릭터에 많은 관심을 두고 있는 것으로 볼 수 있다.

반면, RPG의 주된 콘텐츠인 전투는 사용자 리뷰에서 중요하게 언급되지 않았다. 게임 A의 경우에만 보스 전투에 대한 언급이 등장하고 있을 뿐이며, 특히 최근 게임들의 트렌드인 자동전투가 세 게임에 공통적으로 적용되어 있지만 자동전투는 세 게임의 어느 토픽에도 등장하지 않아 상대적으로 자동전투에 대해 큰 관심을 보이지 않았다. RPG에서 전투는 필수 요소지만 자동전투가 일반화된 이후 전투 자체에 대해 사용자들이 크게 관심을 갖지 않은 것으로 해석된다.

### 3) 빅데이터 분석 결과

빅데이터 분석 결과의 대부분이 긍정적인 내용 보다는 부정적인 내용이 주를 이룬 것으로 보아, 구글 플레이 사용자 리뷰는 사용자들의 직접적인 게임의 불만 사항을 표출하는 장소로 이용되는 것을 알 수 있었다.

분석을 통하여 뛰어난 그래픽 요소, 과금유도가 가능한 콘텐츠가 현재 모바일 RPG의 개발 트렌드와 방향성이라고 확인 할 수 있었다. 또한 사용자는 게임이 가진 고유의 게임정보다는 게임의 그래픽 요소와 보상 및 합리적인 과금에 대해 관심이 많았다. 이는 본 연구에서 우려한 바와 같이 비슷한 게임성을 가진 모바일 RPG에서는 게임정보보다 보상과 과금요소가 더 중요하게 여겨지고 있다는 것을 확인하였다.

토픽 분석 결과를 바탕으로 각 게임에 대해 <표 11>과 같이 방향성을 제시할 수 있다.

<표 11> 각 게임에 대한 방향성 제시

	방향성 제시
게임 A	<ul style="list-style-type: none"> <li>✓ 게임 구동 불가에 대한 사용자들의 불만이 많아 게임의 업데이트와 패치 등의 완성도에 힘을 써야 함</li> <li>✓ 캐릭터의 강화 콘텐츠가 매출에 큰 도움을 주고 있으며 거부감이 덜한 상황으로 이를 활용한 콘텐츠를 장점으로 활용할 수 있음</li> <li>✓ 전투에 대한 긍정적 반응을 바탕으로 전투 콘텐츠를 특징으로 내세울 수 있음</li> </ul>
게임 B	<ul style="list-style-type: none"> <li>✓ 과도한 과금 유도로 인해 스트레스를 받는 사용자가 많은 상황임. 다만 이 부분은 게임이 컨셉으로 가져가는 요소로 보임</li> <li>✓ 과금 사용자들의 매크로 이용에 대한 불만이 크기 때문에 매크로 이용에 대한 적절한 대응이 필요함</li> </ul>
게임 C	<ul style="list-style-type: none"> <li>✓ 비주얼 콘텐츠가 매우 인기 있는 것으로 보이므로 이 부분을 지속적인 장점으로 끌고 갈 필요가 있음</li> <li>✓ 과금요소가 매출에 부정적인 영향을 주도록 설계되어 있는 것으로 보이므로 이를 개선할 필요가 있음</li> <li>✓ 전투 및 사냥 콘텐츠에 대해 만족하지 못하는 것으로 파악되므로 이에 대한 개선이 필요해 보임</li> </ul>



## V. 결론

RPG에서의 사용자들의 관심사에 대해 분석하기 위해 빅데이터 분석 방법을 이용하였다. 설문은 일반적으로 질문에 대해 답을 달아야 하는 책임감을 갖고 있어 무의식적으로 설문 의도에 따라 답을 할 수 있다. 따라서 보다 진정성 있는 데이터를 얻기 위해 구글 플레이의 사용자 리뷰를 활용하였다. 구글 플레이 사용자 리뷰는 사용자 자발적으로 작성하며 사용자와 개발사간의 직접적인 소통 창구로 활용되고 있어 사용자들의 관심사가 직접적으로 반영되는 곳이다.

빅데이터를 확보하기 위해서는 많은 사용자 리뷰가 필요하였으며 사용자들의 피드백이 활발한 게임을 선정하기 위해 매출순위 상위권에 속하는 게임을 선정하였다.

빅데이터를 확보하고 분석하기 위해 Python과 오픈소스 모듈을 활용하였다. 게임마다 5,000건 이상의 사용자 리뷰를 크롤링하여 확보하였으며 이를 토큰화하여 의미 있는 단어를 추출하였다.

단순 단어의 출현 빈도만으로는 정확한 토픽(주제)을 파악하기 어려웠으므로 유의미한 단어를 찾아내기 위해 LDA 토픽 모델링 기법을 활용하였다. LDA 모델의 성능 평가를 위해 혼란도 점수(Perplexity Score)와 일관성 점수(Coherence Score)를 비교하여 LDA 모델의 최적 성능을 찾아내었다.

LDA 모델로 분석된 빅데이터는 비주얼 도식화 하여 토픽간의 상관관계를 보여주고 토픽을 구성하는 주요 단어들을 확보하였다. LDA 모델

에 의해 분석된 토픽을 통해 해당 토픽이 어떤 내용인지 분석하였으며 그 내용을 정리하였다.

앞선 분석을 바탕으로 본 연구를 통해 모바일 RPG 제작에 필요한 주요 요소를 분석하면 다음과 같으며 이를 바탕으로 RPG 개발을 개선할 수 있을 것으로 기대한다.

- (1) 그래픽(비주얼)은 게임의 충성도에 큰 영향을 준다.
- (2) 사용자들은 캐릭터 관련 콘텐츠에 많은 관심을 갖고 있다.
- (3) 과금의 밸런스는 사용자가 가장 민감하게 생각한다.
- (4) 전투 콘텐츠에 대한 욕구가 강하지 않다.

본 연구에서 추가적으로 개선할 사항에 대해서 크게 세 가지로 정리할 수 있다.

첫 번째, 크롤링을 통해 수집한 빅데이터에서 유의미한 단어를 뽑아내는 토크닝 과정의 정교함을 개선하면 더 높은 확률의 정확성을 가지는 분석을 할 수 있다. 한글의 특성상 형태소를 추출했을 때 의미를 잃어버리는 경우가 많이 발생하여, 유사 의미의 단어를 표제어으로써 정제하는 과정 및 쓸데없는 단어를 처리하는 불용과정 등에 더 정교함을 가할 수 있다.

두 번째, LDA 토픽 모델링 기법은 정확하게 토픽을 선정해 줄 수 없는 문제점을 갖고 있다. 따라서 본 논문에서는 혼란도 점수와 일관성 점수를 정확도를 예측하는 방법으로 이용하였는데, 이는 ‘정확도가 높다’의 의미이지 ‘정확하다’를 의미하지 않는다. 따라서 토픽 모델링의

정확도를 높이면 더 정확한 분석을 할 수 있다.

세 번째, LDA 모델의 토픽 분석 결과를 다른 분석자가 분석하였을 경우 토픽의 내용이 달라질 수 있는 가능성이 있으므로 숙련되고 객관성을 지닌 분석자를 다수 투입함으로써 분석 결과를 비교하여 보다 정확한 분석을 할 수 있다.

향후 연구 과제로써 유의미한 단어 추출을 위한 기법을 연구하고, 다른 분석 모델을 통한 후속 분석이 이루어 질 수 있기를 기대한다. 또한 RPG에 대해서만 연구가 진행되었기 때문에 다른 장르의 게임들에도 본 연구를 확대할 수 있을 것으로 보인다. 본 연구가 RPG 개발에 있어서 게임 개발 기업들에게 활용되기를 바란다.

## 참고문헌

- [1] 정규만. "모바일 게임 산업 발전에 관한 연구." 한국컴퓨터게임학회 논문지 2.20, pp135-140, 2010.
- [2] "Google Play", 구글 플레이, 2020년 7월 18일 접속, <https://play.google.com>
- [3] 변현수. "예술적 창의성, 과학적 창의성, 일반적 창의성이 지각된 즐거움과 재이용의도에 미치는 영향: 롤플레이팅 게임 이용자들을 대상으로." 한국게임학회 논문지 11.1, pp59-67, 2011.
- [4] "Role-playing game", 위키피디아, 2020년 7월 18일 접속, [https://en.wikipedia.org/wiki/Role-playing\\_game](https://en.wikipedia.org/wiki/Role-playing_game)
- [5] Kim, Giuhyoung, and Nam-yong Lee. "The Quality Evaluation Model for Mobile RPG." Proceedings of the Korean Institute of Information and Commucation Sciences Conference. The Korea Institute of Information and Commucation Engineering, pp457-460, 2014.
- [6] "HowLongToBeat", HowLongToBeat, 2020년 7월 18일 접속, <https://howlongtobeat.com/>
- [7] Fields T.V., "Game Industry Metrics Terminology and Analytics Case Study". In: Seif El-Nasr M., Drachen A., Canossa A. (eds) Game Analytics. Springer, London, pp.53-71, 2013.
- [8] "Daily Active User (DAU)", Baremetrics, 2020년 7월 18일 접

- 속, <https://baremetrics.com/academy/daily-active-user>
- [9] 임준, 최성욱, “모바일 게임 시장에서 DAU와 게임 구매간의 불균형성을 발생시키는 요인에 대한 고찰 : 구글 플레이 무료게임을 중심으로”, 디지털융복합연구, vol. 12, no. 4, pp. 293 - 303, Apr. 2014.
- [10] “Average Revenue Per User (ARPU)”, Investopedia, 2020년 7월 18일 접속, <https://url.kr/1Cp2gJ>
- [11] “2019년 한국 모바일 게임 시장 객단가 분석”, App Ape Lab, 2020년 7월 18일 접속, <https://url.kr/2ZjqHz>
- [12] “한국콘텐츠진흥원”, 한국콘텐츠진흥원, 2020년 7월 18일 접속, <http://www.kocca.kr>
- [13] “컬럼 - 중국산 양산형 게임, 왜 문제인가?”, Acrofan, 2020년 7월 18일 접속, <https://url.kr/qt6WZw>
- [14] 천영준, 광규태. “집단 감성과 모바일 게임 사용경험: 카카오텔 사례연구.” 한국게임학회 논문지 15.4, pp145-155, 2015.
- [15] 기대현, 박창훈. “소셜 네트워크 서비스 (SNS) 와 소셜 네트워크 게임 (SNG) 사용자들 사이의 양방향 상호작용.” 예술인문사회융합멀티미디어논문지 9, pp1321-1329, 2019.
- [16] 정성화, et al. “MMORPG 온라인게임의 변화와 성장을 위한 디자인 설계 TIME FLOW 시나리오 연구.” 한국게임학회 논문지 15.4, pp79-91, 2015.
- [17] Ernkvist, Mirko. “Down many times, but still playing the game: Creative destruction and industry crashes in the

- early video game industry 1971-1986.", diva-portal.org, pp161-191, 2008.
- [18] Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis. "Deriving the pricing power of product features by mining consumer reviews." *Management science* 57.8, pp1485-1509, 2011.
- [19] 김주영, and 김동수. "텍스트 마이닝 기반의 온라인 상품 리뷰 추출을 통한 목적별 맞춤화 정보 도출 방법론 연구." *한국전자거래학회지* 21.2, pp151-161, 2016.
- [20] 송두현. "수집형 모바일 RPG 사용자 플레이 동기 분석." *한국정보통신학회 종합학술대회 논문집* 23.2, pp559-562, 2019.
- [21] 이규복. "모바일 RPG 게임의 자동전투 시스템 연구". Diss. 한양대학교, 2018.
- [22] 김승언, 김영식. "모바일 RPG 게임의 인 앱 결제 서비스 이용자에 대한 사용성 평가: '리니지 M'을 중심으로." *한국게임학회 논문지* 18.3, pp27-37, 2018.
- [23] 전상현, 박규원. "모바일 게임의 캐릭터 스타일 분석 연구." *브랜드디자인학연구* 14.1, pp137-148, 2016.
- [24] 전지훈, et al. "로그 데이터의 유형분석 및 인지심리적 속성 추출을 이용한 모바일 게이머 유형화 연구." *정보과학회논문지* 45.3, pp234-241, 2018.
- [25] De Mauro, Andrea, Marco Greco, and Michele Grimaldi. "What is big data? A consensual definition and a review of

- key research topics." AIP conference proceedings. Vol. 1644. No. 1. American Institute of Physics, pp97-104, 2015.
- [26] 권영옥. "빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구." 한국지능정보시스템학회 학술대회논문집, pp90-96, 2013.
- [27] 송태민. "우리나라 보건복지 빅데이터 동향 및 활용 방안." 과학기술정책 192, pp56-73, 2013.
- [28] Anshari, Muhammad, et al. "Customer relationship management and big data enabled: Personalization & customization of services." Applied Computing and Informatics 15.2, pp94-101, 2019.
- [29] Russom, Philip. "Big data analytics.", TDWI best practices report, fourth quarter 19.4 : pp1-34, 2011.
- [30] 전성해. "통계적 텍스트 마이닝을 이용한 빅 데이터 전처리." 한국지능시스템학회 논문지 25.5, 470-476, 2015.
- [31] Dhir, Amandeep, Gina M. Chen, and Sufen Chen. "Why do we tag photographs on Facebook? Proposing a new gratifications scale." new media & society 19.4, pp502-521, 2017.
- [32] Pesce, João Paulo, et al. "Privacy attacks in social media using photo tagging networks: a case study with Facebook." Proceedings of the 1st Workshop on Privacy and Security in Online Social Media. 2012.

- [33] Trattner, Christoph, and Frank Kappe. "Social stream marketing on Facebook: a case study." *International Journal of Social and Humanistic Computing* 2.1-2, pp86-103, 2013.
- [34] Rauschnabel, Philipp A., Sandra Praxmarer, and Bjorn S. Ivens. "Social media marketing: How design features influence interactions with brand postings on Facebook." *Advances in Advertising Research* (Vol. III). Gabler Verlag, Wiesbaden, pp153-161, 2012.
- [35] "Some Amazon customers will now be able to get orders delivered in five hours or less", CNBC, 2020년 7월 18일 접속, <https://www.cnn.com/2020/03/03/amazon-expands-same-day-delivery-after-building-fulfillment-centers.html>
- [36] Cavallo, Alberto. "More Amazon effects: online competition and pricing behaviors". No. w25138. National Bureau of Economic Research, 2018.
- [37] "Profitero Price Intelligence: Amazon makes more than 2.5 million daily price changes", Profitero, 2020년 7월 18일 접속, <https://url.kr/OBWRII>
- [38] "Amazon Fraud Detector", AWS, 2020년 7월 18일 접속, <https://aws.amazon.com/ko/fraud-detector/>
- [39] "Amazon Personalize", AWS, 2020년 7월 18일 접속, <https://aws.amazon.com/ko/personalize/>
- [40] "With the promise of faster speed, inventory moves closer



to consumers”, Supplychaindive, 2020년 7월 18일 접속, <http://url.kr/mXbUoP>

- [41] Kim, Joo Young, and Dong soo Kim. "A study on the method for extracting the purpose-specific customized information from online product reviews based on text mining." *Journal of society for e-business studies* 21.2, 2017.
- [42] O'Mara-Eves, Alison, et al. "Using text mining for study identification in systematic reviews: a systematic review of current approaches." *Systematic reviews* 4.1 5, 2015.
- [43] Liu, Wei, and Wilson Wong. "Web service clustering using text mining techniques." *International Journal of Agent-Oriented Software Engineering* 3.1, pp6-26, 2009.
- [44] Salloum, Said A., et al. "Using text mining techniques for extracting information from research articles." *Intelligent natural language processing: trends and applications*. Springer, Cham, pp373-397, 2018.
- [45] Salloum, Said A., et al. "Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques." *International Journal of Information Technology and Language Studies* 1.2, pp8-17, 2017.
- [46] Navathe, Shamkant B., and Ramez A. Elmasri. "Fundamentals of Database Systems" with Cdrom and Book.

- Addison-Wesley Longman Publishing Co., Inc., 2001.
- [47] 홍원의, et al. "사례 기반 지능형 수출통제 시스템: 설계와 평가." 지능정보연구 20.3, pp109-131, 2014.
  - [48] 김근형, 오성열. "온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론." 한국콘텐츠학회논문지 9.8, pp272-284, 2009.
  - [49] 강범일, 송민, 조화순. "토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구." 한국문헌정보학회지 47.4, pp315-334, 2013.
  - [50] Musat, Claudiu Cristian, et al. "Improving topic evaluation using conceptual knowledge." Twenty-Second International Joint Conference on Artificial Intelligence. 2011.
  - [51] 배정환, 손지은, 송민. "텍스트 마이닝을 이용한 2012 년 한국대선 관련 트위터 분석." 지능정보연구 19.3, pp141-156, 2013.
  - [52] 김태경, 최희련, 이홍철. "토픽 모델링을 이용한 핀테크 기술 동향 분석." 한국산학기술학회 논문지 17.11, pp670-681, 2016.
  - [53] Wang, Chong, and David M. Blei. "Collaborative topic modeling for recommending scientific articles." Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011.
  - [54] Calheiros, Ana Catarina, Sérgio Moro, and Paulo Rita. "Sentiment classification of consumer-generated online reviews using topic modeling." Journal of Hospitality

- Marketing & Management 26.7, pp675-693, 2017.
- [55] Sun, Lijun, and Yafeng Yin. "Discovering themes and trends in transportation research using topic modeling." *Transportation Research Part C: Emerging Technologies* 77, pp49-66, 2017.
- [56] Pingclasai, Natthakul, Hideaki Hata, and Ken-ichi Matsumoto. "Classifying bug reports to bugs and other requests using topic modeling." 2013 20Th asia-pacific software engineering conference (APSEC). Vol. 2. IEEE, 2013.
- [57] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3,Jan, pp993-1022, 2003.
- [58] Jelodar, Hamed, et al. "Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey." *Multimedia Tools and Applications* 78.11, pp15169-15211, 2019.
- [59] Tian, Kai, Meghan Revelle, and Denys Poshyvanyk. "Using latent dirichlet allocation for automatic categorization of software." 2009 6th IEEE International Working Conference on Mining Software Repositories. IEEE, pp163-166, 2009.
- [60] Tirunillai, Seshadri, and Gerard J. Tellis. "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation." *Journal of Marketing*

Research 51.4, pp463-479, 2014.

- [61] Bolelli, Levent, Şeyda Ertekin, and C. Lee Giles. "Topic and trend detection in text collections using latent dirichlet allocation." European Conference on Information Retrieval. Springer, Berlin, Heidelberg, pp776-780, 2009.
- [62] Somasundaram, Kalyanasundaram, and Gail C. Murphy. "Automatic categorization of bug reports using latent dirichlet allocation." Proceedings of the 5th India software engineering conference. pp125-130, 2012.
- [63] 박자현, 송민. "토픽모델링을 활용한 국내 문헌정보학 연구동향 분석." 정보관리학회지 30.1, pp7-32, 2013.
- [64] Blei, David M. "Probabilistic topic models." Communications of the ACM 55.4, pp77-84, 2012.
- [65] Newman, David, et al. "Evaluating topic models for digital libraries." Proceedings of the 10th annual joint conference on Digital libraries, pp215-224. 2010.
- [66] Wallach, Hanna M., et al. "Evaluation methods for topic models." Proceedings of the 26th annual international conference on machine learning. 2009.
- [67] Newman, David, et al. "Automatic evaluation of topic coherence." Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, pp100-108, 2010.

- [68] Chang, Jonathan, et al. "Reading tea leaves: How humans interpret topic models." *Advances in neural information processing systems*, pp288-296, 2009.
- [69] Ali, Syed Mohd, et al. "Big data visualization: Tools and challenges." 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), IEEE, pp656-660, 2016.
- [70] Gorodov, Evgeniy Yur'evich, and Vasiliy Vasil'evich Gubarev. "Analytical review of data visualization methods in application to big data." *Journal of Electrical and Computer Engineering* 2013, 2013.
- [71] "2020년 오픈 소스 데이터 시각화 툴 TOP 10", Finereport, 2020년 7월 18일 접속, <https://url.kr/Vb6AL4>
- [72] Olshannikova, Ekaterina, et al. "Visualizing Big Data with augmented and virtual reality: challenges and research agenda." *Journal of Big Data* 2.1: 22, 2015.
- [73] Ozgur, Ceyhun, et al. "MatLab vs. Python vs. R." *Journal of Data Science* 15.3, pp355-372, 2017.
- [74] Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.
- [75] Olston, Christopher, and Marc Najork. "Web crawling", *Now Publishers Inc*, 2010.

- [76] Kim, Sung-min, and Young-guk Ha. "Automated discovery of small business domain knowledge using web crawling and data mining." 2016 International Conference on Big Data and Smart Computing (BigComp). IEEE, 2016.
- [77] Linares, Ricardo, et al. "Prediction of tourist traffic to Peru by using sentiment analysis in Twitter social network." 2015 Latin American Computing Conference (CLEI). IEEE, 2015.
- [78] Musale, Deepali V., et al. "Job recommendation system using profile matching and web-crawling." Int. J. Adv. Sci. Res. Eng. Trends, 2016.
- [79] Castillo, Carlos. "Effective web crawling." Acm sigir forum. Vol. 39. No. 1. New York, NY, USA: Acm, 2005.
- [80] Chung, Kyoung-Rock, et al. "News Abusing Inference Model Using Web Crawling." Proceedings of the Korean Society of Computer Information Conference. Korean Society of Computer Information, pp175-176, 2018.
- [81] "BeautifulSoup", Crummy, 2020년 7월 18일 접속, <https://www.crummy.com/software/BeautifulSoup/>
- [82] "Selenium with Python", Selenium Python, 2020년 7월 18일 접속, <https://selenium-python.readthedocs.io/>
- [83] "Lexical analysis", Wikipedia, 2020년 7월 18일 접속, [https://en.wikipedia.org/wiki/Lexical\\_analysis#Tokenization](https://en.wikipedia.org/wiki/Lexical_analysis#Tokenization)

- [84] “텍스트 토큰화(Text Tokenization)”, 신백현 블로그, 2020년 7월 18일 접속, <https://url.kr/nZJKua>
- [85] 남기심, 이정민, 이홍배. 언어학 개론. 탑출판사, 1983.
- [86] 김덕봉, 최기선, 강제우. “한국어 형태소 처리와 사전-접속정보를 이용한 한글 철자 및 띄어쓰기 검사기.” 어학연구, 1990.
- [87] Cerda, Patricio, Gaël Varoquaux, and Balázs Kégl. “Similarity encoding for learning with dirty categorical variables.” Machine Learning 107.8-10, pp1477-1494, 2018.
- [88] Granda, Jarosław M., et al. “Controlling an organic synthesis robot with machine learning to search for new reactivity.” Nature 559.7714, pp377-381, 2018.
- [89] Kim, Kwang Ho, Byunghoon Chang, and Hwang Kyu Choi. “Deep Learning Based Short-Term Electric Load Forecasting Models using One-Hot Encoding.” Journal of IKEEE 23.3, pp852-857, 2019.
- [90] “KoNLPy: 파이썬 한국어 NLP”, KoNLPy, 2020년 7월 18일 접속, <https://konlpy-ko.readthedocs.io/ko/v0.4.3/>
- [91] 박은정, 조성준. “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지.” 제 26 회 한글 및 한국어 정보처리 학술대회 논문집, pp 1-4, 2014.
- [92] “형태소 분석 및 품사 태깅”, KoNLPy, 2020년 7월 18일 접속, <https://url.kr/LGra6k>
- [93] Park, Eunjeong L., and Sungzoon Cho. “KoNLPy: Korean

- natural language processing in Python.” Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology, Vol. 6, pp133-136, 2014.
- [94] “pyLDAvis 2.1.2”, Pypi, 2020년 7월 18일 접속, <https://pypi.org/project/pyLDAvis/>
- [95] “검은사막 홈페이지”, Pearl Abyss, 2020년 7월 18일 접속, [http://www.blackdesertm.com/Intro/Hashashin\\_Update](http://www.blackdesertm.com/Intro/Hashashin_Update)
- [96] “검은사막 구글 플레이”, 구글 플레이, 2020년 7월 18일 접속, <https://play.google.com/store/apps/details?id=com.pearlabyss.blackdesertm>
- [97] “리니지M 홈페이지”, NC Soft, 2020년 7월 18일 접속, <https://lineagem.plaync.com/event/p2020/trinity/index>
- [98] “리니지M 구글 플레이”, 구글 플레이, 2020년 7월 18일 접속, <https://play.google.com/store/apps/details?id=com.ncsoft.lineagem19>
- [99] “에픽세븐 홈페이지”, STOVE, 2020년 7월 18일 접속, <https://page.onstove.com/epicseven/kr>
- [100] “에픽세븐 구글 플레이”, 구글 플레이, 2020년 7월 18일 접속, <https://play.google.com/store/apps/details?id=com.stove.epic7.google>



## 부록

### <부록 1> Python을 이용한 웹크롤링 소스코드

```
1. ##crawling.py
2. ##염동현
3. ##웹사이트에서 데이터를 크롤링 하는 과정
4. ##필요한 데이터만 가져옴
5.
6. import sys
7. import time
8. import re
9. from bs4 import BeautifulSoup
10. from selenium import webdriver
11.
12. url = 'https:xxxxxxx'
13.
14. driver = webdriver.Chrome(executable_path='chromedriver.exe')
15. driver.get(url)
16. driver.implicitly_wait(3)
17.
18. ##더보기 횟수 저장
19. more = 0
20. max_more = 100
21.
22. while(True):
23.     driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
24.     time.sleep(0.5)
25.
26.     if (more == max_more):
27.         break
28.
29.     try:
30.         ## 더보기
31.         element = driver.find_element_by_xpath('//div[@jsname="xxxxx"]')
32.         if (element is not None):
33.             more += 1
34.             print(more)
35.             element.click()
36.
```

```

37.     except Exception:
38.         continue
39.
40. html = driver.page_source
41. driver.quit()
42.
43. soup = BeautifulSoup(html, 'html.parser')
44.
45. ##사용자 리뷰만 뽑아오기. 기본 리뷰
46. reviews1 = soup.find_all("span", {"jsname": "xxxxx"})
47. ##전체 리뷰 버튼이 있을 때
48. reviews2 = soup.find_all("span", {"jsname": "xxxxx"})
49.
50. print(len(reviews2))
51.
52. ##파일에 쓰기
53. sys.stdout = open('documents.txt', 'w', -1, "utf-8")
54.
55. count = 0
56.
57. for n in reviews2:
58.     if len(n.text) == 0:
59.         ##contents = reviews1[count].text.strip()
60.         contents = reviews1[count].get_text()
61.     else:
62.         ##contents = n.text.strip()
63.         contents = n.get_text()
64.
65.     ##특수 문자 제거
66.     contents = re.sub('[^가~힉0-9a-zA-Z ]', '', contents)
67.
68.     contents.encode('utf-8')
69.     print(contents)
70.
71.     count = count + 1

```

## <부록 2> Python을 이용한 토크닝 소스코드

```
1. ##tokening.py
2. ##염동현
3. ##크롤링 데이터에서 불필요한 단어들을 제외하는 과정
4.
5. import sys
6. from konlpy.tag import Okt
7. from collections import Counter
8.
9. stop_file_name = "stopwords.txt"
10. text_file_name = "documents.txt"
11. output_file_name = "corpus.txt"
12.
13. ## 불용어 리스트
14. stop_text_file = open(stop_file_name, 'r', -1, "utf-8")
15. stop_words = stop_text_file.read()
16. stop_text_file.close()
17.
18. stop_words=stop_words.split(' ')
19.
20.
21. open_text_file = open(text_file_name, 'r', -1, "utf-8")
22. sys.stdout = open(output_file_name, 'w', -1, "utf-8")
23.
24. while(True):
25.     line = open_text_file.readline()
26.     if not line:
27.         break
28.
29.     spliter = Okt()
30.     ## pos = part of speech
31.     token = spliter.pos(line)
32.
33.     tags = ['Noun', 'Verb']
34.
35.     newtext = ''
36.     for n in token:
37.         ## 원하는 형태소만 골라내기
38.         if n[1] in tags:
39.             ## 불용어 제거
```

```
40.         if n[0] not in stop_words:
41.             ## 불용어에 해당 하지 않은 경우에만 추가
42.             newtext = newtext + ' ' + n[0]
43.
44.     print(newtext)
45.
46. open_text_file.close()
```

### <부록 3> Python을 이용한 모델 평가하기 Passes 이용 소스코드

```
1. ##lda_test_p.py
2. ##염동현
3. ##lda 모델 평가하기 - Passes 개수에 의한
4.
5. import gensim
6. from collections import Counter
7. import pyLDAvis
8. import pyLDAvis.gensim as gensimvis
9. import sys
10.
11. corpus_file_name = "corpus.txt"
12.
13. open_text_file = open(corpus_file_name, 'r', -1, "utf-8")
14.
15. documents = []
16.
17. while(True):
18.     line = open_text_file.readline()
19.     if not line:
20.         break
21.
22.     documents.append(line.strip().split())
23.
24. open_text_file.close()
25.
26. dictionary = gensim.corpora.Dictionary(documents)
27. print('dictionary size : %d' % len(dictionary))
28.
29. corpus = [dictionary.doc2bow(text) for text in documents]
30.
31. ## 인덱스를 문자열로 변환
32. corpus2 = [[(dictionary[id], freq) for id, freq in cp] for cp in corpus[:1]]
33.
34. lda_test_file_name = "lda_test_p.csv"
35. sys.stdout = open(lda_test_file_name, 'w', -1, "utf-8")
36.
37.
38. iterations = 400
39. num_topics = 10
```

```

40.
41. print("\nNumber of Topics : ', num_topics)
42. print("\nPasses Perplexity Coherence')
43.
44.
45. for i in range(0, 11):
46.
47.     num_passes = i*5
48.     if num_passes == 0 : num_passes = 1
49.
50.     ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=num_topics,
iterations=iterations, passes=num_passes)
51.
52.     topic = ldamodel.print_topics()
53.
54.     coherence_model_lda = gensim.models.CoherenceModel(model=ldamodel,
texts=documents, dictionary=dictionary, coherence='u_mass')
55.     coherence_lda = coherence_model_lda.get_coherence()
56.
57.     print(num_passes, ' ', ldamodel.log_perplexity(corpus), ' ',coherence_lda)

```

#### <부록 4> Python을 이용한 모델 평가하기 Topic 개수 이용 소스코드

```
1. ##lda_test_t.py
2. ##염동현
3. ##lda 모델 평가하기 - Topic 개수에 의한
4.
5. import gensim
6. from collections import Counter
7. import pyLDAvis
8. import pyLDAvis.gensim as gensimvis
9. import sys
10.
11. corpus_file_name = "corpus.txt"
12.
13. open_text_file = open(corpus_file_name, 'r', -1, "utf-8")
14.
15. documents = []
16.
17. while(True):
18.     line = open_text_file.readline()
19.     if not line:
20.         break
21.
22.     documents.append(line.strip().split())
23.
24. open_text_file.close()
25.
26.
27. dictionary = gensim.corpora.Dictionary(documents)
28. print('dictionary size : %d' % len(dictionary))
29.
30. corpus = [dictionary.doc2bow(text) for text in documents]
31.
32. ## 인덱스를 문자열로 변환
33. corpus2 = [[(dictionary[id], freq) for id, freq in cp] for cp in corpus[:1]]
34.
35. lda_test_file_name = "lda_test_t.csv"
36. sys.stdout = open(lda_test_file_name, 'w', -1, "utf-8")
37.
38.
39. iterations = 400
```

```

40. ## lda_test_p.py를 먼저 실행해서 적합한 passes의 수를 구해야 한다.
41. num_passes = 35
42.
43.
44. print("\nNumber of Passes : ", num_passes)
45. print("\nnum_topics Perplexity Coherence")
46.
47.
48. for i in range(1, 11):
49.
50.     ## 토픽의 단위는 2개
51.     num_topics = i*2
52.
53.     ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=num_topics,
        iterations=iterations, passes=num_passes)
54.
55.     topic = ldamodel.print_topics()
56.
57.     coherence_model_lda = gensim.models.CoherenceModel(model=ldamodel,
        texts=documents, dictionary=dictionary, coherence='u_mass')
58.     coherence_lda = coherence_model_lda.get_coherence()
59.
60.     print(num_topics, ' ', ldamodel.log_perplexity(corpus), ' ', coherence_lda)

```



## <부록 5> Python을 이용한 Visualization 소스코드

```
1. ##lda.py
2. ##염동현
3. ##단어주머니(bow)를 만들고 이를 이용해 LDA 토피모델링을 하고 비주얼화
   하는 과정
4.
5. import gensim
6. from collections import Counter
7. import pyLDAvis
8. import pyLDAvis.gensim as gensimvis
9.
10. corpus_file_name = "corpus.txt"
11.
12. open_text_file = open(corpus_file_name, 'r', -1, "utf-8")
13.
14. documents = []
15.
16. while(True):
17.     line = open_text_file.readline()
18.     if not line:
19.         break
20.
21.     documents.append(line.strip().split())
22.
23. open_text_file.close()
24.
25. dictionary = gensim.corpora.Dictionary(documents)
26. print('dictionary size : %d' % len(dictionary))
27.
28. corpus = [dictionary.doc2bow(text) for text in documents]
29.
30. ## 인덱스를 문자열로 변환
31. corpus2 = [[(dictionary[id], freq) for id, freq in cp] for cp in corpus[:1]]
32.
33. iterations = 400
34. num_topics = 10
35. num_passes = 1
36.
37. ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=num_topics,
        iterations=iterations, passes=num_passes)
```

```
38.  
39.topic = ldamodel.print_topics()  
40.  
41.prepared_data = gensimvis.prepare(ldamodel, corpus, dictionary)  
42.pyLDavis.save_html(prepared_data, "topic.html")
```

# **ABSTRACT**

## **Research on improvement of mobile RPG using big data analysis**

**-Focusing on high rank three RPG in Korean Google Play-**

**Youm, Dong Hyun**

**Advised by Prof. Kim, Jung Yoon**

**Dept. of IT Convergence**

**Engineering**

**Graduate School of**

**Gachon University**

This paper studied the improvement plan of mobile RPG (Role Playing Game) through the big data analysis method. To this end, we collected the top 3 mobile RPG user reviews from Korean Google Play and processed them to extract meaningful information.

In order to extract meaningful information from big data, user reviews were collected through crawling in the Google Play store, and meaningful text data was extracted through a tokenizing process. Results were visualized through LDA (Latent Dirichlet Allocation) Topic Modeling, and based on the visualization results, the final analysis was performed to find the main topic.

Three games of different personalities were selected according to DAU (Daily Active Users) and sales, and a case analysis was conducted directly. The topics of each game were analyzed and direction of each game was presented. Also, we analyzed and compared the topics of each game to find out what the common topics of mobile RPG are, and it was found that the current development trend and direction of mobile RPG have excellent graphics and content that can be charged. Users have confirmed that they are more interested in the game's graphic elements, rewards, and reasonable charging than the game's inherent gameability.

In conclusion, as a result of big data analysis, it is the main factor necessary for the production of mobile RPG. First, graphic (visual) has a great influence on the loyalty of the game. Second, users are interested in character-related content. Third, the balance of billing is the most sensitive to users. Lastly, it was pointed out that the desire for combat content was not strong, and based on this, it was proposed to improve the development of mobile RPG.

Further improvements were made to improve the sophistication of crawling when collecting big data, to increase the accuracy of LDA topic modeling, and to compare through multiple analysts.

**Keywords: big data, mobile RPG, topic modeling, LDA**

빅  
데  
이  
터  
분  
석  
을  
통  
한  
모  
바  
일  
R  
P  
G  
의  
개  
선  
안  
연  
구

염  
동  
현