

# 글로벌 게임 트렌드의 빅데이터 분석을 통한 국내 게임 시장의 발전 방향성 제시

송준협<sup>○</sup>, 임민우<sup>\*</sup>, 김한수<sup>\*\*</sup>

<sup>○</sup>서원대학교 소프트웨어 학부 학생,

<sup>\*</sup>서원대학교 소프트웨어 학부 학생,

<sup>\*\*</sup>서원대학교 소프트웨어 학부 교수

e-mail: narosong@gmail.com<sup>○</sup>, immw8146@gmail.com<sup>\*</sup>, kutestar@seowon.ac.kr<sup>\*\*</sup>

## Suggestion of development for domestic game market through big data analysis of global game trend

Junhyup Song<sup>○</sup>, Minwoo Lim<sup>\*</sup>, Hansoo Kim<sup>\*\*</sup>

<sup>○</sup>Student, School of software, Seowon University,

<sup>\*</sup>Student, School of software, Seowon University,

<sup>\*\*</sup>Professor, School of software, Seowon University

### ● 요약 ●

게임 산업은 기술의 발전과 비대면 서비스 수요 증가로 해마다 발전하고 있다. 본 연구는 사용자들의 수요를 조사하기 위하여 대중성이 가장 높은 온라인 게임 플랫폼에서 이용 시간이 많은 게임 정보를 확인하였다. HTML 파싱(parsing) 라이브러리를 통해 해당 게임들의 리뷰를 크롤링하여 엑셀 파일로 데이터베이스화하였고, 자연어 처리 라이브러리를 활용하여 데이터를 정제하였다. 총 5개 장르에 대하여 분석한 결과 각 장르에 해당하는 대표적인 키워드를 확인할 수 있었다. 취득한 키워드는 범용 시각화 패키지를 활용하여 워드 클라우드 형태로 한눈에 알아볼 수 있도록 시각화하였다.

**키워드:** 크롤링(crawling), 텍스트 마이닝(text Mining), 스팀(Steam), 게임 트렌드, 워드 클라우드

## I. Introduction

코로나19 팬데믹은 게임을 온라인으로 사람들과 만날 수 있는 수단이자 집 안에서 누구나 즐길 수 있는 엔터테인먼트로 자리매김하게 했다.

청년 남성이라는 게임 이용자에 대한 스테레오 타입 이미지도 깨졌다. 이전보다 더 다양한 사람들이 게임을 즐기게 되었으며, 게임에 대한 부정적인 인식이 상당히 개선되는 효과를 거두었다[1].

게임 산업은 연마다 규모가 커지고 있다. 모바일 데이터 및 분석 플랫폼 data.ai와 게임 마케팅 인텔리전스 기업인 IDC가 함께 조사한 "게임 스포트라이트 2022년 리뷰"에 따르면 22년도 기준으로 전 세계의 게임 산업 규모가 한화 약 282조 원에 달할 것으로 예측되었다. 이 중에서 모바일 게임은 61%(1,360억 달러)로 예측됐으며, 이는 PC 게임(400억 달러) 대비 3.4배, 가정용 콘솔 게임(420억 달러) 대비 3.3배 높은 수치다[2].

한국의 경우, 문화체육관광부의 "국민여가활동조사 2021"에 따르면, 여가생활 중 8위가 게임이며(18.9%), 연령대별 여가 활동으로는 15~18세, 20대가 각각 58.0%(1위), 44.8%(2위)로 10대와 20대가

게임의 주요 소비층으로 확인되었다[3].

한국콘텐츠진흥원이 발간한 "2021 대한민국 게임백서"에 따르면 2020년 기준 한국의 게임 시장 규모는 모바일이 4위, PC가 3위 등이다. 세계 시장에서의 점유율은 6.9%로 미국(21.9%), 중국(18.1%), 일본(11.5%)에 이은 4위다[4].

본 연구는 한국 게임산업의 트렌드를 제시하기 위해 세계의 게임 산업 트렌드를 조사하였다. 온라인 게임 플랫폼 스팀(Steam)의 이용 시간 상위 100개의 게임 정보를 크롤링하여 정보 데이터셋을 구축하였고, 해당 게임들의 정보를 통해 사용자 리뷰를 크롤링하여 리뷰 데이터를 수집하였다. 그 결과 5개의 장르에 대하여 장르별로 상위 50개의 키워드를 추출하여 시각화 작업을 하였다.

## II. Preliminaries

### 2.1. Related works

#### 2.1.1 Crawling

크롤링(crawling)은 웹 크롤러에서 출발한 말로 인터넷상의 페이지 (HTML, 문서 등)를 수집해서 분류하고 저장하며 쉽게 찾아볼 수 있는 역할을 하는 일종의 로봇이다. 웹 크롤러의 행위는 복합적 정책들의 산물이다[5].

크롤링 정책에는 페이지의 다운로드를 언급하는 선택정책과 문서의 변경 사항을 언제 검사할지 언급하는 재방문 정책, 이 정책은 변경 빈도에 상관없이 동일한 빈도로 컬렉션의 모든 페이지를 다시 방문하는 동일 정책과 자주 변경되는 페이지를 더 자주 방문하는 것으로 포함하는 비례정책 이 두 가지 정책으로 나누어진다. 두 경우 모두 반복되는 페이지의 크롤링 순서를 임의 순서 또는 고정 순서로 수행할 수 있다. 그리고, 웹사이트의 과부하를 막기 위한 언급하는 공손성 정책과 분산 웹 크롤러를 어떻게 조율할지 언급하는 병렬화 정책이 있다[5].

웹 크롤러는 대체로 모든 페이지의 복사본을 생성하며, HTML 코드 검증과 링크를 주기적으로 체크하며 웹사이트의 자동 유지 관리 작업 등 웹 페이지의 특정 형태의 정보를 수집하는 데도 사용된다. 인터넷에 공개된 웹페이지의 HTML을 수집하기 위해 사용한다. 웹사이트 데이터를 크롤링하기 위해서는 웹사이트의 구조를 파악해야 한다[6].

#### 2.1.2 Text Mining

빅데이터를 활용한 분석은 “텍스트 마이닝, 데이터마이닝, 평판 분석, 소셜 분석, 클러스터 분석, 현실 마이닝” 등 크게 6가지로 나눌 수 있다.

그중 텍스트 마이닝(text mining)은 비정형화된 텍스트 자료를 대량 수집해 의미 있는 패턴과 관계성 등을 추출하는 분석법을 의미한다. 특히, 사용자의 댓글이나 게시 글에서 파편적인 키워드를 수집하는 수준이 아니라 맥락을 분석하고 데이터 속에 숨겨진 패턴을 발견해 가치 있는 결론을 도출할 수 있다[7].

방대한 텍스트 데이터에서 의미 있는 정보를 추출하기 위해서는 자동화된 텍스트 분석 기법의 활용이 요구된다. 자동화 텍스트 분석은 자연어 처리 기술을 활용해 사람들의 텍스트 언어를 컴퓨터가 인식할 수 있는 형태로 전환하고, 이를 텍스트 마이닝으로 분석해 의미 있는 데이터로 추출한다.

텍스트 마이닝은 다양한 연구에서 활용되고 있는 분석법으로, 선행 연구를 통해 소비자의 인지 및 동기, 행동 분류, 감정 등을 탐구할 수 있어 가치가 크다[7].

## III. The Proposed Scheme

### 3.1. Building a Dataset

#### 3.1.1 Divide by genre

스팀의 인기 게임 정보를 확인하기 위하여 스팀에서 제공하는 “SteamDB”[8]의 6월 1일 기준 플레이한 시간이 많은 상위 100개 게임의 게임제목, 장르, 출시 년도, 게임번호 데이터를 크롤링하여 데이터셋을 구축하였다. 게임 장르 분류 기준은 사용자 평가에 따라 분류하였다.

Table 1. Game genre

Genre	Count
simulation	17
strategy	15
RPG	13
FPS	11
survival	10

게임 분류를 장르별로 나누었을 때 시뮬레이션(simulation) 장르가 17%로 1위이며, 전략(strategy) 장르가 15%로 2위, 역할수행게임 (RPG) 장르가 13%로 3위를 기록하였다. Table 1과 같이 게임 장르 선호도를 확인하였다.

상위 5개 장르의 게임 리뷰를 크롤링하기 위하여 HTML 파싱 라이브러리인 BeautifulSoup 라이브러리를 사용하여 html을 총 66개의 게임에 대하여 파싱하였다.

#### 3.1.2 Keyword Analysis

스팀에서 크롤링한 리뷰 데이터는 엑셀 파일형태로 데이터베이스화 하였다. 게임마다 리뷰의 개수가 상이하기 때문에, 리뷰 개수를 게임마다 10,000개로 제한하여 수집하였다.

수집된 리뷰 데이터는 자연어 처리 라이브러리인 nltk 패키지로 정제 과정을 거쳐 이모티콘과 불용어 등을 제거하였고, 유의미한 키워드를 얻기 위하여 “game”, “play”, “fun” 등과 같은 키워드는 게임 트렌드 분석에 유의미한 결과를 주지 못하여 제외하였다. 정제 과정을 마친 데이터는 wordcloud 패키지로 시각화 과정을 마친 후 분석을 완료하였다.

### 3.2 Analysis Results

Table 2. Word frequency for Simulation

Word	Frequency	Word	Frequency
click	12245	cooky	6496
factory	8202	even	6742
love	7749	much	6559
mod	7366	year	6178
cookie	7062	still	6151

시뮬레이션 장르 17개의 리뷰데이터에 대하여 텍스트 빈도 분석을 실행하였다. Table 2는 빈도수가 높은 상위 10개의 키워드를 뜻하며, 시뮬레이션 장르에서는 “factory”, “city”, “grow”, “zombie” 등과 같은 키워드들이 주요 키워드로 분석되었다.

“click”, “cookie”는 “Cookie Clicker”라는 게임이며, 클릭어(Clicker) 게임 종류에 해당하였다. (Fig. 1. 참조)

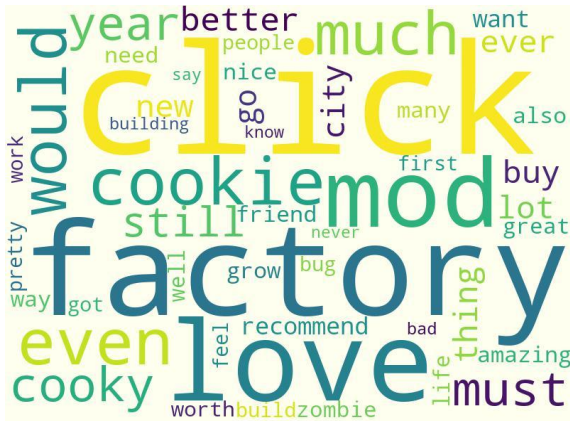


Fig. 1. simulation wordcloud

Table 3. Word frequency for Strategy

Word	Frequency	Word	Frequency
war	12289	even	9319
dlc	12207	much	9227
strategy	9710	love	8838
make	9319	new	8713
would	9302	total	8280

Table 3은 전략 장르 키워드 중 빈도수가 높은 상위 10개의 키워드를 뜻한다. 전략 장르는 전쟁과 관련된 키워드가 주로 등장하며 “total”, “civ”와 키워드는 게임 제목 각각 “Total war”, “Sid meier’s Civilization”을 뜻하며 “war”, “make” 등의 키워드와 같이 무엇인가를 만들어 전쟁하는 종류의 게임이다. (Fig. 2. 참조)



Fig. 2. strategy wordcloud

Table 4. Word frequency for RPG

Word	Frequency	Word	Frequency
monster	12782	character	7808
world	12259	MMO	7626
story	12239	combat	6870
feel	8161	also	6844
quest	7997	amazing	6493

Table 4는 롤 플레이 게임 키워드 중 빈도수가 높은 키워드 10개에 해당한다. 롤 플레이 게임은 “monster”, “world”, “story”, “quest”, “character”, “MMO” 등과 같은 키워드가 주요 키워드로 분석되었다. (Fig. 3. 참조)

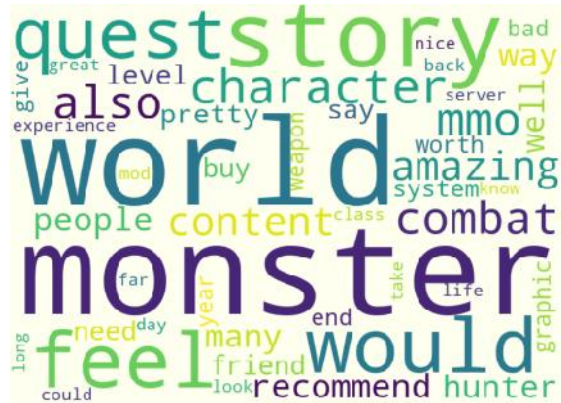


Fig. 3. RPG wordcloud

Table 5. Word frequency for FPS

Word	Frequency	Word	Frequency
friend	3916	squad	3011
people	3519	team	2872
go	3400	community	2614
make	3305	old	2484
ever	3222	want	2614

Table 5는 1인칭 슈팅 게임(FPS) 장르의 키워드 중 빈도수가 높은 상위 10개의 키워드를 뜻한다. 1인칭 슈팅 게임에선 “friend”, “people”, “squad”, “team”, “community” 등과 같은 협동의 의미를 갖는 키워드가 주로 등장하였다. (Fig. 4. 참조)

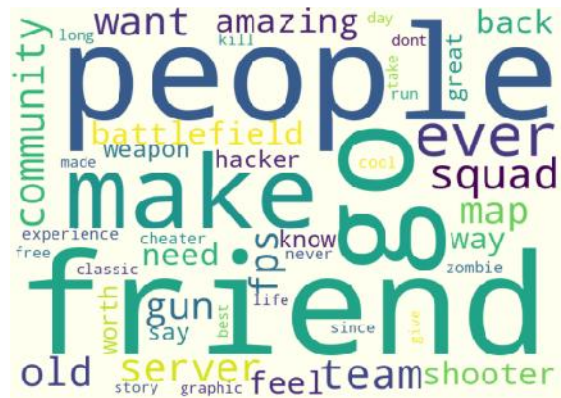


Fig. 4. FPS wordcloud

Table 6. Word frequency for Survival

Word	Frequency	Word	Frequency
friend	5316	survival	3814
would	4554	people	3534
mod	4483	thing	3264
server	3964	bad	3232
much	3850	want	3121

Table 6은 생존 장르 키워드 중 빈도수가 높은 상위 10개의 키워드를 뜻한다. “friend”, “people” 등은 협동의 의미를 가지며 다른 주요 키워드로는 “killer”, “base” 등의 키워드가 있다. (Fig. 5. 참조)

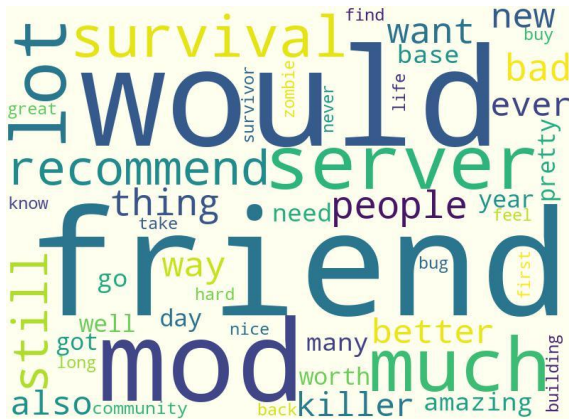


Fig. 5. survival wordcloud

#### IV. Conclusions

본 논문은 한국의 게임 발전 방향을 제시하기 위하여 스팀의 인기 게임에 대하여 장르별로 분류 후 사용자 리뷰를 크롤링하여 분석하였다. 상위 5개의 장르는 시뮬레이션, 전략, 롤 플레이 게임, 1인칭 슈팅 게임, 서바이벌로 구성되어 있었으며, 롤 플레이, 슈팅, 생존 장르는 싱글 플레이보다 사람들과 같이하는 멀티플레이가 선호되고, 전략과 시뮬레이션 장르의 게임은 상대적으로 싱글 플레이 게임이 많은 것을 확인하였다. 장르별로 살펴보면, 시뮬레이션 장르는 사물이나 건물, 도시 등을 육성하는 유형의 게임과 클릭 게임이 상위권을 차지하였고, 전략 장르는 턴제와 실시간 전략 게임 유형이 상위권을 차지하였다. RPG는 스토리, 퀘스트, 전투 등을 중요하게 여겼으며 FPS와 서바이벌 장르는 사람들과 멀티플레이를 중점으로 보았다.

본 논문은 향후 게임 사업 관계자들이 게임 산업의 흐름을 파악하고 전략 수립 과정에서 실무적인 운영에 활용될 수 있다는 점에서 의미가 있다.

#### ACKNOWLEDGEMENT

본 논문은 과학기술정보통신부와 한국지능정보사회진흥원 주관으로 소방청 컨소시엄에서 수행하는 “소방안전 빅데이터 플랫폼 및 센터 구축 사업”(2020~2022)의 지원을 받았습니다.

본 논문은 과학기술정보통신부 정보통신 창의 인재 양성사업의 지원을 통해 수행한 2022년도 ICT 멘토링 프로젝트[22\_PG068] 결과물입니다.

#### REFERENCES

- [1] Strabase, “Global Game Industry Trend (January + February 2022)”, Korea Creative Content Agency, KOCCA
- [2] “Game Spotlight 2022 Review with IDC”, <https://www.data.ai/kr/>
- [3] “2021 National Leisure Activities Survey”, Ministry of Culture, Sports and Tourism (South Korea)
- [4] “WHITE PAPER ON KOREAN GAMES”, Korea Creative Content Agency, KOCCA
- [5] “RobotsDarabase”, <http://robotstxt.org/db.html>
- [6] “Web crawler” Wikipedia, [http://en.wikipedia.org/wiki/Web\\_crawler](http://en.wikipedia.org/wiki/Web_crawler)
- [7] Ji Young Na, “Analysis of VR Game Trends using Text Mining and Word Cloud -Focusing on STEAM review data-”, Journal of Korea Game Society Vol. 22 No.1: 87-98. 2022 Feb
- [8] “SteamDB”, <https://steamdb.info/graph/>