디시인사이드 갤러리 LDA 토픽 모델링을 통한 이슈 파악

-던전앤파이터 갤러리, 메이플 스토리 갤러리를 중심으로-

광고.PR.브랜딩 송찬영

1. **서론**

현대 디지털 시대에서 소비자의 관심을 사로잡고 신뢰를 구축하는 것은 기업에게 핵심적인 도전 과제이다. 이를 위해 고객의 요구를 이해하고 그에 맞춘 전략을 개발하는 것이 핵심이다. 고객의 의견과 반응을 빠르게 파악하며, 문제점을 해결하고 긍정적인 브랜드 이미지를 유지하는 것은 PR (Public Relations)의 주요 목표 중 하나이다.

디시인사이드는 이용량이 가장 큰 대표적인 커뮤니티 중 하나로, 유저들은 커뮤니티 내에서 게임에 대한 의견을 가감 없이 비추고 있다. 또한, 제목에서부터 관심을 끌어야 사람들의 반응을 이끌 수 있는 커뮤니티의 특성상 제목의 분석만으로 주요 토픽을 파악할 수 있다. 본 보고서는 디시인사이드의 ‘던전앤파이터’, ‘메이플스토리’ 각 갤러리에서 업로드되는 글의 제목을 크롤링하고, LDA(Latent Dirichlet Allocation) 토픽 모델링을 활용하여 분석하고자 한다.

이번 분석에서는 디시인사이드 커뮤니티에서의 LDA 토픽 모델링을 통해 어떤 주제들이 주요하게 등장하고 있는지를 밝히고, 이를 통해 게이머의 디시인사이드 커뮤니티 이용행태와 게임 내의 주요하게 다뤄지고 있는 이슈를 살필 것이다.

1. **데이터 수집**

데이터 수집을 위해 BeautifulSoup와 WebDriver를 사용하였다. BeautifulSoup는 웹 스크래핑을 위한 파이썬 라이브러리로, 웹 페이지에서 필요한 정보를 추출하는데 사용하였다. WebDriver는 웹 브라우저를 제어하고 웹 페이지 상의 데이터를 수집하기 위한 도구로 사용하였다.

2023년 7월 13일부터 10월 13일까지 총 3개월 동안 ‘카트라이더 드리프트’, ‘던전앤파이터’, ‘메이플스토리’ 갤러리에 올라온 게시글의 제목을 크롤링하였다. 수집된 데이터 중에서 광고, 설문, 공지 게시물은 분석 대상에서 제외하였다.

정해진 기간 동안 ‘ ‘던전앤파이터’ 갤러리에서 수집된 데이터의 수는 206,966개, ‘메이플스토리’ 갤러리에서 수집된 데이터의 수는 771,201개다.

1. **데이터 전처리**

정규표현식을 사용하여 한글과 알파벳을 제외한 문자는 제거하고, 길이가 1 이하인 단어는 제거하였다. 불용어로는 비속어, ‘ㅋㅋ’, ‘ㄱㄱ’와 같은 단순 초성 줄임말은 최대한 제거하였지만, 이용자의 커뮤니티 이용행태를 최대한 파악하기 위해 본 연구자가 판단하기에 의미 있는 용어는 제하지 않았다.

1. **분석 결과**

분석 대상인 두 갤러리 SVD\_MODEL의 n\_iter 값은 100으로 통일하였다. Algorhithm은 ‘randomized’을 사용했고, 토픽의 개수는 반복하여 LDA 토픽을 추출하였을 때, 토픽이 최대한 겹치지 않게 하는 것을 기준으로 연구자의 주관에 따라 토픽의 개수를 선정하였다.

* 1. **던전앤파이터**

다음은 던전앤파이터 갤러리에서 전체 용어 빈도에 따른 행렬 출력 값과 이에 따른 본 연구자의 분류이다. 표1의 결과는 토픽의 개수가 10개임을 가정했을 때 나온 값으로 의미 없는 비속어의 용어 행렬이나 특별한 주제가 도출되지 않는 경우엔 제외하였다.

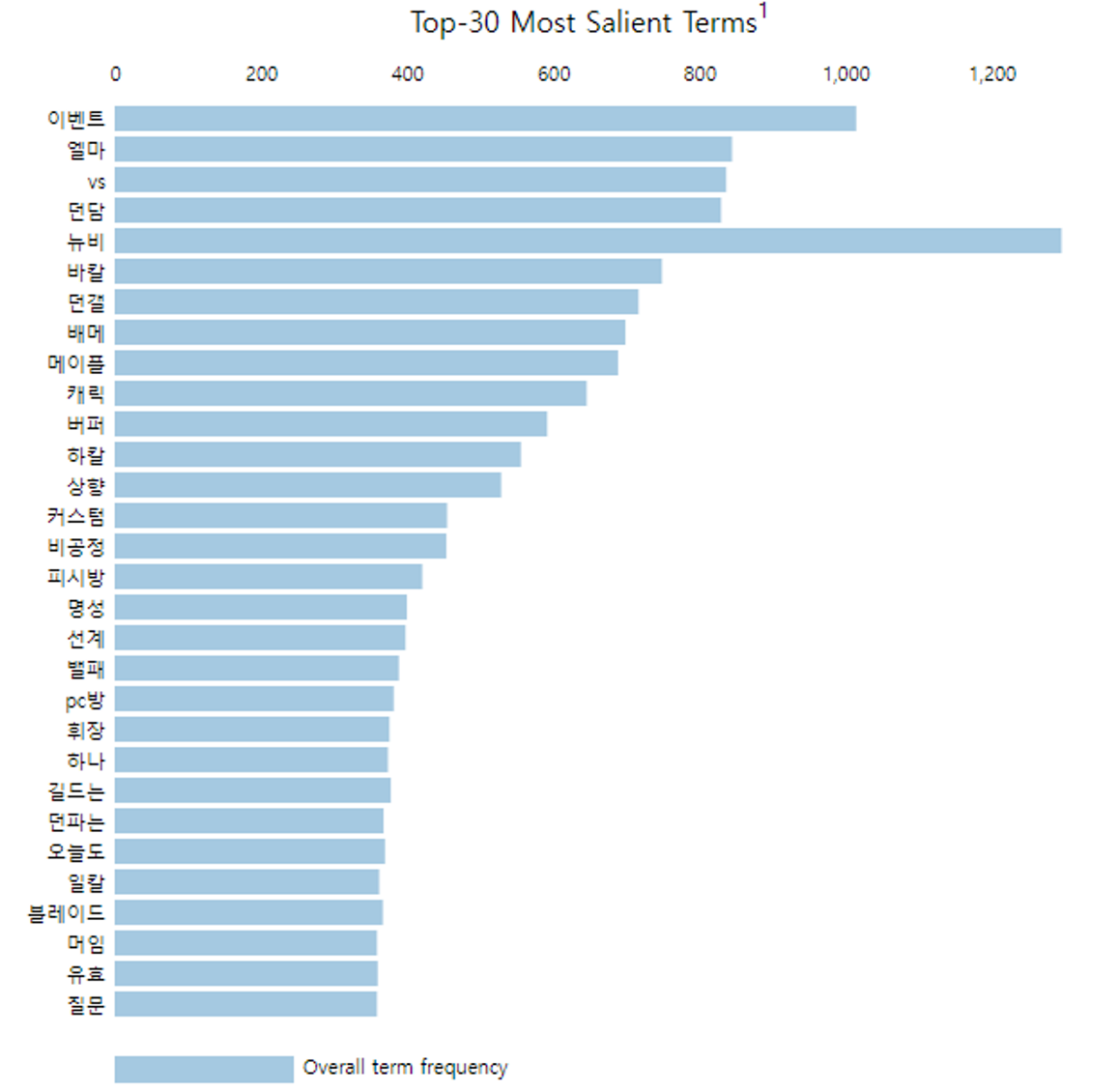
토픽 1과 토픽 2를 통해 이용자는 디시인사이드 커뮤니티를 신규 유입 시 질문용, 이벤트 시에 정보 습득용으로 활발히 활용하고 있음을 알 수 있다. 또한, 토픽3을 통해 디시인사이드 커뮤니티를 단순 정보 습득용으로 그치지 않고, 이용자들 간의 일상 대화를 나누는 등 커뮤니티 내부에서 친목을 나누고 있음을 확인할 수 있었다. 이는 던전앤파이터 이용자들이 보다 강한 이용자 간의 친밀감을 형성하고 있음을 시사한다.

표 1 던전앤파이터 토픽 행렬 출력값

|  |  |  |
| --- | --- | --- |
| 토픽 순번 | 전체 용어 빈도에 따른  토픽 용어 행렬 출력 값 | 분류 |
| 1 | 이벤트0.67023+피시방0.45068+대리0.41498+pc방0.34408+모집합니다0.1314+시간당원0.12999+새로운0.1293+뉴비0.04473+참여0.02313+질문0.01831 | 피시방 이벤트 |
| 2 | 뉴비0.93789+질문0.25915+질문좀0.1155+던갤0.07162+질문있어요0.06807+커스텀0.05345+질문점0.0479+복귀0.041+직업0.03616+캐릭0.03468 | 뉴비, 복귀 질문 |
| 3 | 던갤0.83628+잘자요0.45441+좋은아침0.28241+다들0.04192+일가자0.02894+잘자0.02673+안녕0.0238+년전0.02083+레전드0.01772+여갤러0.01671 | 일상 대화 |
| 4 | 님들아0.99552+잘자0.05566+선계0.04059+커스텀0.02343+vs0.01404+어떰0.01049+메이플0.00876+머임0.00873+던파너무재밌죠0.00692+하이영0.00692 | 선계 업데이트 |

표2는 pyLDAvis를 통해 주요 키워드 30개를 추출한 표이다. ‘이벤트’와 ‘뉴비’ 키워드가 상위 랭크에 표시된 것을 통해 표1에서 본 것과 같이 이벤트 진행시와 신규 유저가 정보를 구하는 목적으로 활용하고 있었다. 이외에도 ‘엘마’, ‘바칼’, ‘하칼’, ‘배메’ 등과 같이 게임 내의 인기 직업과 주요 보스몹의 주요 키워드로 올랐고, ‘밸패’ ‘상향’, ‘선계’ 등의 패치 내용이 언급되었음을 확인할 수 있다.

**표 2 던전앤파이터 갤러리 주요 키워드 30**



* 1. **메이플스토리**

다음은 메이플스토리 갤러리에서 전체 용어 빈도에 따른 행렬 출력 값과 이에 따른 분류이다. 표3의 결과는 토픽의 개수가 10개임을 가정했을 때 나온 값이다. 매크로 광고가 갤러리 내에서 다량 발견되었으며 이를 토픽 모델링이 반영한 것을 확인할 수 있다. **신규 유저 질문 관련 토픽의 경우가 크게 늘었다. 이를 통해 7월 13일부터 10월 13일 사이에 메이플스토리 내에 신규 유입이 많았음을 유추할 수 있다.**

또한, 던전앤파이터 갤러리와 다르게 비속어의 비율이 크게 줄었으며, **평가를 요청하는 토픽이 던전앤파이터 갤러리와 다르게 크게 많은 것을 확인하였다.** 게임 내 콘텐츠로는 리부트 월드, 에르다 조각이 주요하게 언급되어 토픽으로 올랐다.

표 3 던전앤파이터 토픽 행렬 출력값

|  |  |  |
| --- | --- | --- |
| 토픽순번 | 전체용어빈도에 따른  토픽용어 행렬 출력 값 | 분류 |
| 1 | 매크로*0.70925+자동사냥*0.70451+사냥*0.01248+mshxxmxx*0.01213+텔레그램*0.01204+판매*0.01115+채집*0.00264+아르테일*0.00212+큐브*0.00187+농장*0.00154 | 매크로 광고 |
| 2 | 리부트*0.97894+너프*0.09141+본섭*0.06609+최종뎀*0.06505+ㄴ라고*0.05653+vs*0.04218+닉네임*0.03974+생제*0.03715+메린이*0.03314+뉴비*0.03097 | 리부트  월드 |
| 3 | 닉네임*0.91743+메린이*0.20957+뉴비*0.14546+질문*0.12482+어떰*0.12223+많관부*0.08157+ㅁㅌㅊ*0.07636+만원*0.06508+팝니다*0.06503+팔리냐*0.05255 | 닉네임 판매, 평가 |
| 4 | 메린이*0.71716+질문*0.40212+뉴비*0.38532+질문좀*0.16515+유니온*0.07765+vs*0.07546+ㅁㅌㅊ*0.04749+버닝섭*0.0466+무기*0.04112+어떰*0.04089 | 신규유저 질문 |
| 5 | 뉴비*0.73497+어떰*0.39405+vs*0.15805+코디*0.04416+유니온*0.03972+버닝섭*0.03528+ㅁㅌㅊ*0.03378+추천좀*0.03305+사냥*0.02786+뭐가*0.02301 | 신규유저 질문2 |
| 6 | 어떰*0.89655+메린이*0.1561+코디*0.07668+사냥*0.02081+가격*0.02039+억에*0.01541+스카니아*0.01385+억인데*0.01162+보조*0.01137+무기*0.01093 | 신규유저 질문3 |
| 7 | vs*0.9394+뭐가*0.11284+추천좀*0.06387+누가*0.05871+불독*0.05677+좋음*0.04943+아케인*0.04693+팡이*0.04446+나로*0.0425+윈브*0.03995 | 신규유저 질문4 |
| 8 | ㅁㅌㅊ*0.89666+코디*0.37397+본인*0.09867+메붕이*0.07836+평가좀*0.06229+지듣노*0.06142+억에*0.04045+만원*0.0383+조각*0.03729+평가*0.03578 | 평가 |
| 9 | 조각*0.94447+가격*0.12993+스카니아*0.08823+에르다*0.08676+솔에르다*0.08487+루나*0.07767+크로아*0.07565+얼마임*0.07358+만원*0.06+재획*0.05935 | 에르다 조각 |
| 10 | 이벤트*0.63148+대리*0.47869+피시방*0.44696+pc방*0.37579+만원*0.06656+피케인*0.06409+모집바로시작*0.04372+피방*0.0356+모집합니다*0.03412+시간당원*0.03356 | 피시방 이벤트 |

표4는 메이플스톨리 갤러리의 주요 키워드 30개를 추출한 표이다. 리부트 월드(리부트), 에르다 조각(조각), 큐브, 나이트로드(나로) 등의 게임 내 주요 아이템과 이슈가 상위 키워드에 오른 것을 확인할 수 있다. 리부트 월드와 일반 월드 간의 갈등을 보여주는 대표적인 단어 리선족이 주요 키워드에 오른 것을 통해 여전히 유저 간의 갈등이 줄어들지 않았음을 확인할 수 있다.

표 4 메이플스토리 갤러리 주요 키워드 30

**텍스트, 스크린샷, 도표, 번호이(가) 표시된 사진

자동 생성된 설명**

1. **결론 및 제언**

던전앤파이터 갤러리와 메이플스토리 갤러리에 업로드 된 제목을 통해 커뮤니티 내 유저들의 커뮤니티 이용 목적과 주요 토픽을 확인할 수 있었다. 공통된 이용 목적으로는 신규 및 복귀 유저의 게임 방향을 묻기 위한 질문, 피시방 이벤트 관련 질문이었다. **향후 신규 및 복귀 유저를 위한 가이드라인의 보완이나 피시방 이벤트 진행 시 보다 더 자세한 가이드 제시할 수 있다.**

던전앤파이터 갤러리의 경우 선계 업데이트 내용이 토픽에 올랐다. **유저의 관심이 지목되고 있는 상황이므로 추가적인 선계 업데이트 관련 이슈 분석이 필요하다.** 또한, 던전앤파이터의 경우 특별히 유저 간의 일상대화가 주요 토픽으로 오른 것을 확인할 수 있다. **유저 간의 보다 가까운 친밀감이 형성된 것으로 향후 이벤트를 진행할 때 이런 점을 고려하여 진행할 수 있다.**

메이플스토리 갤러리의 경우 주요 토픽으로 매크로가 올랐다. 이는 매크로 불법 광고에 의한 것이었지만, 이외에도 ‘*스카니아 도서관4 매크로가 판치는데 진짜 신기하네’, ‘요즘 거탐으로 매크로 못잡는거 맞지?’* 등의 게시글이 확인되었다. 불법 광고를 포함하여 매크로 관련 게시글이 커뮤니티 내의 반복 등장한다면 이용자 사이에서 부정적인 이슈가 확산될 수 있다는 잠재적인 위험이 있다. **불법광고의 경우 사측에서 선제적으로 커뮤니티에 삭제 요청해 이슈를 관리하고 매크로 검열을 강화할 필요가 있다.**

또한, 유저 간의 코디 등의 상호 평가를 하는 문화가 토픽을 통해 확인할 수 있었다. 이를 통해 추후 **코디 공모전과 같은 이벤트를 고려할 수 있다.**

주요 키워드 30에 ‘리선족’과 ‘리부트 월드(리부트)’가 오르며 아직 리부트 월드 유저와 **일반 월드 유저 간의 갈등이 사라지지 않는 것을 확인할 수 있었다. 이또한 사측에서 주기적인 모니터링을 통해 유저의 반응을 살펴 부정적인 이슈로의 확산을 방지할 필요가 있다.**