1. 데이터 분석

연구 목적에 따라서 전반적인 후기를 조사하기 위해서 구글 플레이 스토어(Google Play Store)의 ‘따릉이’ 어플리케이션 리뷰와 홈페이지의 시민의견수렴 게시판의 불만사항 및 문제점을 수집한다. 구글 플레이 스토어에 작성된 리뷰와 시민 의견 수렴 게시판 데이터 수집을 통해서 실제 ‘따릉이’ 서비스를 이용한 사용자의 경험을 알 수 있다. 이렇게 조사한 텍스트데이터를 텍스트마이닝 기법을 활용하여 사용자의 불만 사항에 대해 분석을 진행한다.

텍스트 데이터는 2015년 9월부터 2019년 7월까지 구글 플레이 스토어에 등록된 ‘따릉이’ 리뷰와 2016년 11월 3일부터 2019년 8월 3일까지 시민 의견 수렴 게시판에 작성된 게시글이다. 위 데이터는 파이썬의 ‘BeautifulSoup 4’와 ‘Selenium’을 사용하여 크롤링을 진행한다. 리뷰의 경우 평점, 이용자명, url 항목을 수집하였다. 수집된 리뷰의 수는 1847개, 리뷰 별점 1.6점이다. 시민 의견 수렴 게시판은 작성자 ID, 게시 날짜, 제목, 내용을 수집한다. 수집된 게시글의 갯수는 8618개 이다.

1. 데이터 전처리

이와 같이 수집된 데이터는 인터넷 상에서 개인이 작성한 데이터이기 때문에 맞춤법 오류, 고유명사의 다양한 표기 등 예외적 사용이 잦기 때문에 텍스트 분석을 진행하기 전 데이터 전처리를 진행한다. KoNPLPy Twitter과 KoNLPy의 수정된 버전인 customized KoNLPy(<https://github.com/lovit/customized_konlpy>)를 사용하여 명사, 형용사를 추출한다. 이렇게 추출한 명사와 형용사를 기반으로 문제점을 분석한다.

1. 데이터 분석

따릉이 어플리케이션에 달린 1847개의 리뷰를 전처리 과정을 진행 후, 자주 등장한 형용사와 명사 상위 30개 단어는 다음과 같다.

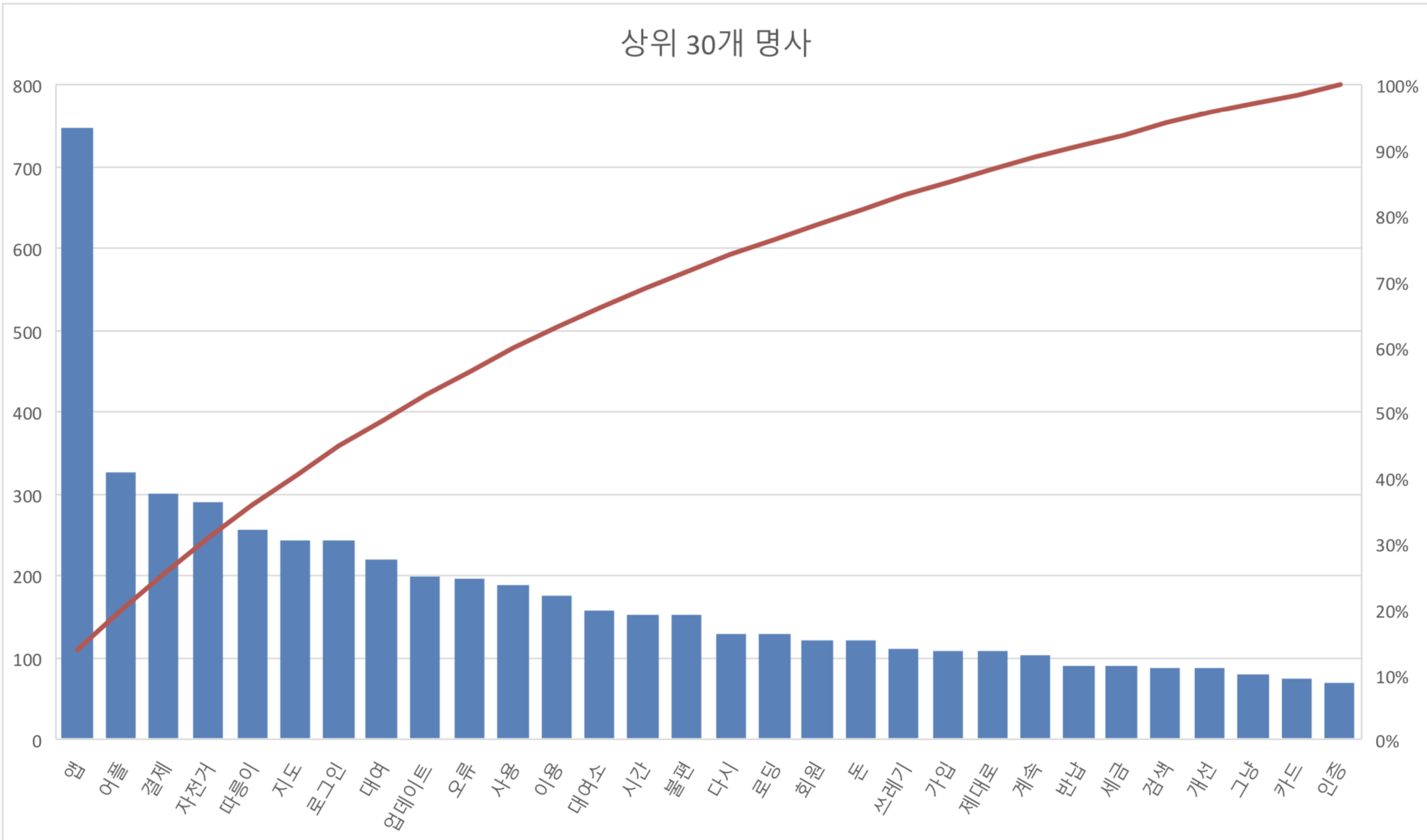
<형용사>

('안되고', 109), ('입니다', 89), ('어떻게', 86), ('안됨', 59), ('이런', 53), ('있는', 49), ('좋아요', 48), ('좋은데', 44), ('좋은', 44), ('같은', 42), ('없', 41), ('없는', 36), ('없고', 36), ('안되네요', 35), ('없네요', 33), ('같습니다', 32), ('많고', 32), ('안되서', 31), ('있습니다', 30), ('없음', 30), ('같아요', 30), ('느리고', 29), ('아닌', 26), ('있는데', 25), ('새로', 24), ('느려', 24), ('좋', 20), ('안되는', 19), ('없다', 19)

<명사>

('앱', 749), ('어플', 329), ('결제', 301), ('자전거', 293), ('따릉이', 258), ('지도', 246), ('로그인', 245), ('대여', 221), ('왜', 213), ('업데이트', 201), ('오류', 198), ('진짜', 191), ('사용', 190), ('이용', 177), ('최악', 163), ('대여소', 159), ('시간', 154), ('불편', 153), ('다시', 131), ('로딩', 131), ('회원', 123), ('돈', 122), ('쓰레기', 113), ('제대로', 109), ('가입', 109), ('계속', 105), ('정말', 96), ('반납', 92), ('세금', 92), ('검색', 90)

형용사의 경우 전반적으로 ‘안되고’ , ‘어떻게’, ‘안됨’ 과 같이 부정적이 형용사가 문장에서 많이 사용되고 있다. 명사의 경우 형용사 사용에 비추어 볼 때 ‘어플’ , ‘결제’, ‘자전거’, ‘지도’ 등 따릉이 어플리케이션의 기능과 관련된 단어가 추출 되었다. 명사를 분석한 파레토 차트는 다음과 같다. ‘앱’을 제외한 대분의 명사가 고루 분포 되어 있다.



따릉이 홈페이지 시민게시판에 작성된 8618개의 게시글을 전처리 과정을 진행 후, 자주 등장한 형용사와 명사 상위 30개 단어는 다음과 같다.

<형용사>

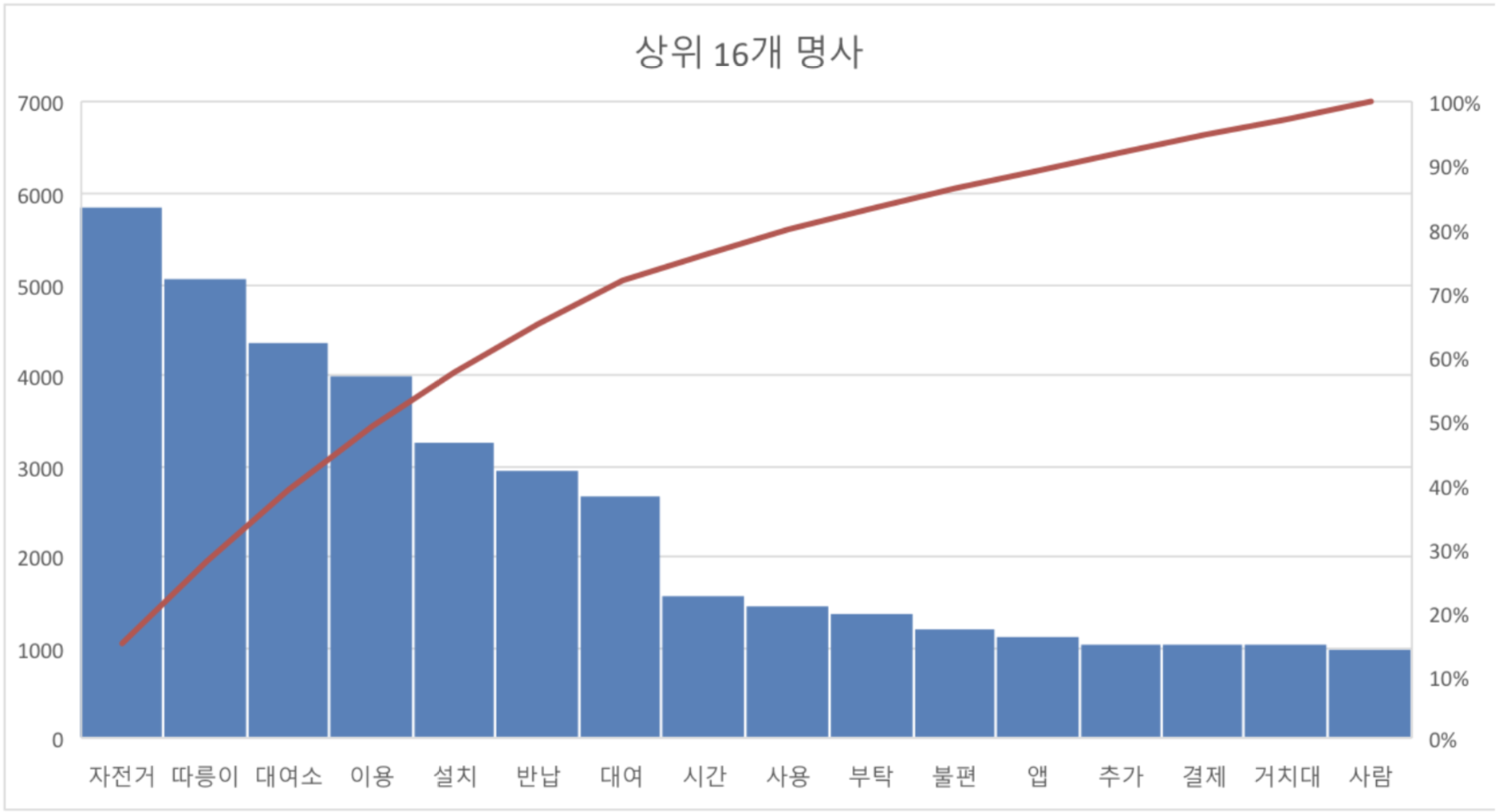
('입니다', 1789), ('있는', 1272), ('있습니다', 1101), ('있', 1091), ('없', 921), ('같습니다', 743), ('많은', 631), ('있는데', 520), ('좋겠습니다', 476), ('어떻게', 461), ('없습니다', 453), ('없는', 408), ('이런', 372), ('없네요', 354), ('같은', 340), ('같은데', 304), ('빠른', 294), ('있고', 280), ('있어', 280), ('안되고', 276), ('좋은', 273), ('많습니다', 273), ('아닌', 272), ('있으면', 270), ('없어', 259), ('없고', 250), ('같아요', 247), ('있게', 210), ('있을', 175), ('안되네요', 158), ('안됩니다', 156), ('많은데', 156), ('안되는', 156),

<명사>

('자전거', 5858), ('따릉이', 5085), ('대여소', 4369), ('이용', 3995), ('설치', 3278), ('반납', 2967), ('대여', 2677), ('시간', 1587), ('사용', 1475), ('부탁', 1379), ('불편', 1207), ('앱', 1132), ('추가', 1055), ('결제', 1048), ('거치대', 1039), ('사람', 998), ('답변', 993), ('확인', 942), ('가능', 934), ('생각', 847), ('고장', 827), ('지도', 802), ('근처', 797), ('출구', 790), ('시민', 770), ('다른', 760), ('다시', 736), ('처리', 728), ('글', 714), ('경우', 708), ('위치', 672), ('아파트', 653), ('계속', 648), ('왜', 645), ('문제', 628), ('정기권', 622), ('장애', 619), ('주변', 617), ('전화', 589), ('타고', 588), ('번호', 584), ('분', 569), ('연결', 567), ('건의', 567), ('이용자', 563), ('못', 554), ('카드', 554), ('지역', 545), ('주시', 542), ('하나', 536), ('통신', 534), ('개선', 533), ('등록', 529), ('오늘', 525), ('요청', 519), ('환불', 509), ('신고', 501), ('거리', 498), ('이용권', 495)

시민게시판에서 추출된 형용사의 경우 어플리케이션 리뷰에 비해 다소 중립적이나 ‘없’, ‘없습니다.’ , ‘없는’ ,’ 없네요’와 같은 부정적 의미의 파생어가 많이 사용되었다.

명사의 경우 어플리케이션에 관련된 단어 보다는 ‘대여소’, ‘이용’, ‘설치’ ‘반납’, ‘대여’ 등 과 같이 자전거를 이용하는 물리적인 부분을 나타내는 단어가 많이 사용되었다. 명사를 분석한 파레토 차트는 다음과 같다.



위와 같이 수집된 구글 플레이 스토어의 빈도수 상위 70개의 명사와 시민게시판의 빈도수 상위 70개 명사의 공통된 단어를 추출하였다.

'자전거', '따릉이', '대여소', '이용', '반납', '대여', '시간', '사용', '불편', '앱', '결제', '사람', '생각', '지도', '다시', '위치', '계속', '문제', '전화', '타고', '번호', '카드', '개선', '등록', '이용권', '오류', '서울시', '어플'

1. 데이터 분석 결과

구글 플레이 스토어의 리뷰 점수(1.6점) 와 자주 사용된 부정적 형용사를 미루어 보았을 때 공통적으로 추출된 명사가 어플리케이션 이용에 부정적 영향을 준다고 생각한다. 그러므로 공통 단어를 중점적으로 따릉이 어플리케이션의 프로세스 분석을 진행하였다.

<추가사항>

서울시가 조사한 만족도 대비 구글 플레이 스토어(Google play store) 50만 다운로드 기준으로 대부분의 사용자가 아래<그래프>와 같이 1점에 별점을 주며 낮은 만족도(1.9점)를 보여주었다. 또한 2,148(2019.04.28 기준)개의 리뷰가 작성되어 있으며, 많은 불만 의견들과 전반적으로 낮은 평점을 기록했다.(리뷰, 만족도 다시 조사하기)

<시민 수렴 게시판 – 형용사>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | word | | Freq | |  | | word | | Freq | |  | | word | | Freq | |
| 1 | | 입니다 | | 1792 | | 11 | | 어떻게 | | 461 | | 21 | | 안되고 | | 276 | |
| 2 | | 있는 | | 1274 | | 12 | | 없습니다 | | 453 | | 22 | | 좋은 | | 273 | |
| 3 | | 있습니다 | | 1101 | | 13 | | 없는 | | 408 | | 23 | | 많습니다 | | 273 | |
| 4 | | 있 | | 1091 | | 14 | | 이런 | | 372 | | 24 | | 아닌 | | 272 | |
| 5 | | 없 | | 921 | | 15 | | 없네요 | | 354 | | 25 | | 있으면 | | 271 | |
| 6 | | 같습니다 | | 743 | | 16 | | 같은 | | 340 | | 26 | | 없어 | | 259 | |
| 7 | | 많은 | | 631 | | 17 | | 같은데 | | 304 | | 27 | | 없고 | | 250 | |
| 8 | | 있는데 | | 520 | | 18 | | 빠른 | | 294 | | 28 | | 같아요 | | 247 | |
| 9 | | 좋겠습니다 | | 476 | | 19 | | 있고 | | 280 | | 29 | | 있게 | | 210 | |
| 10 | | 입니다 | | 1792 | | 20 | | 있어 | | 280 | | 30 | | 있을 | | 175 | |

<시민 수렴 게시판 – 명사>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | word | | Freq | |  | | word | | Freq | |  | | word | | Freq | |
| 1 | | 자전거 | | 5858 | | 11 | | 앱 | | 1132 | | 21 | | 지도 | | 802 | |
| 2 | | 따릉이 | | 5085 | | 12 | | 추가 | | 1055 | | 22 | | 근처 | | 797 | |
| 3 | | 대여소 | | 4369 | | 13 | | 결제 | | 1048 | | 23 | | 출구 | | 790 | |
| 4 | | 이용 | | 3995 | | 14 | | 거치대 | | 1039 | | 24 | | 시민 | | 770 | |
| 5 | | 설치 | | 3278 | | 15 | | 사람 | | 998 | | 25 | | 다른 | | 760 | |
| 6 | | 반납 | | 2967 | | 16 | | 답변 | | 993 | | 26 | | 다시 | | 736 | |
| 7 | | 대여 | | 2677 | | 17 | | 확인 | | 942 | | 27 | | 처리 | | 728 | |
| 8 | | 시간 | | 1587 | | 18 | | 가능 | | 934 | | 28 | | 글 | | 714 | |
| 9 | | 사용 | | 1475 | | 19 | | 생각 | | 847 | | 29 | | 경우 | | 708 | |
| 10 | | 부탁 | | 1379 | | 20 | | 고장 | | 827 | | 30 | | 위치 | | 672 | |

<구글 플레이스토어 – 형용사 >

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | word | | Freq | |  | | word | | Freq | |  | | word | | Freq | |
| 1 | | 안되고 | | 109 | | 11 | | 좋은 | | 44 | | 21 | | 있습니다 | | 30 | |
| 2 | | 입니다 | | 89 | | 12 | | 같은 | | 42 | | 22 | | 없음 | | 30 | |
| 3 | | 어떻게 | | 86 | | 13 | | 없 | | 41 | | 23 | | 같아요 | | 30 | |
| 4 | | 있 | | 72 | | 14 | | 없는 | | 36 | | 24 | | 느리고 | | 29 | |
| 5 | | 안됨 | | 59 | | 15 | | 없고 | | 36 | | 25 | | 아닌 | | 26 | |
| 6 | | 이런 | | 53 | | 16 | | 안되네요 | | 35 | | 26 | | 있는데 | | 25 | |
| 7 | | 있는 | | 49 | | 17 | | 없네요 | | 33 | | 27 | | 새로 | | 24 | |
| 8 | | 좋아요 | | 48 | | 18 | | 같습니다 | | 32 | | 28 | | 느려 | | 24 | |
| 9 | | 좋은데 | | 44 | | 19 | | 많고 | | 32 | | 29 | | 좋 | | 20 | |
| 10 | | 안되고 | | 109 | | 20 | | 안되서 | | 31 | | 30 | | 안되는 | | 19 | |

<구글 플레이스토어 – 명사 >

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | word | | Freq | |  | | word | | Freq | |  | | word | | Freq | |
| 1 | | 앱 | | 749 | | 11 | | 오류 | | 198 | | 21 | | 쓰레기 | | 113 | |
| 2 | | 어플 | | 329 | | 12 | | 사용 | | 190 | | 22 | | 가입 | | 109 | |
| 3 | | 결제 | | 301 | | 13 | | 이용 | | 177 | | 23 | | 제대로 | | 109 | |
| 4 | | 자전거 | | 293 | | 14 | | 대여소 | | 159 | | 24 | | 계속 | | 105 | |
| 5 | | 따릉이 | | 258 | | 15 | | 시간 | | 154 | | 25 | | 반납 | | 92 | |
| 6 | | 지도 | | 246 | | 16 | | 불편 | | 153 | | 26 | | 세금 | | 92 | |
| 7 | | 로그인 | | 245 | | 17 | | 다시 | | 131 | | 27 | | 검색 | | 90 | |
| 8 | | 대여 | | 221 | | 18 | | 로딩 | | 131 | | 28 | | 개선 | | 89 | |
| 9 | | 업데이트 | | 201 | | 19 | | 회원 | | 123 | | 29 | | 그냥 | | 80 | |
| 10 | | 앱 | | 749 | | 20 | | 돈 | | 122 | | 30 | | 카드 | | 75 | |