3) 전국 지자체 제공 데이터 질적 검증

-Stopwords: 시, 군, 구, 동 등 지역 이름, 현황, 목록 등 주제와 연관 없는 명사



<전국 지자체 제공 데이터 파일명 빈도수 상위100>

-폐기물, 배출 상대적으로 출현 빈도수 낮음(폐기물 319, 배출 308 빈도수 8,9위)

-복지, 장애인, 노인 등 복지와 관련된 단어 빈도수 높음



<전국 지자체 제공 이상치 데이터 파일명 빈도수 상위 100>

-폐기물, 배출 등 환경 관련 키워드의 출현빈도 높음 (폐기물161, 배출138로 빈도수 2,3위)

-전체 제공 데이터와 이상치 데이터의 명사가 69% 일치

\*파일 데이터 이용의 ‘최근’ 동향과 관련이 있는지?

\*환경 분야의 데이터 제공을 증가시켜야 한다고 볼 수 있는지?

\*공공데이터 활용 인기데이터와 비교 언급할 수 있는지?

\*자치행정기관 필터링..?

<공공데이터 활용 인기데이터>

1. 67호

1) 도로교통공단\_교통사고 통계

2) 상가(상권)정보

3) 기상청\_단기예보(월간)

4) 기상청\_초단기실황(월간)

5) 중기예보자료(일간)

지역별 공공데이터포털의 데이터 공유 현황 분석

1. 서론

2. 관련 연구

2.1 공공데이터

2.2 공공데이터 활용

2.3 공공데이터의 지역별 데이터 품질평가

3. 연구방법

3.1 데이터 수집

3.2 데이터 분류 (분류체계)

3.3 탐색적 분석 (Exploratory Data Analysis)

4. 결과

4.1 데이터 현황 (분류체계(표), 다운로드/조회수, 업로드 등)

4.2 지역별 데이터 특징 (지역별로 설명)

4.3 요약 (데이터 분석 결과에 대한 요약)

5. 시사점과 토의 사항

5.1 공공데이터포털에 대한 시사점

5.2 연구방법의 한계

6. 결론

5. 시사점과 토의사항

5.1 공공데이터포털에 대한 시사점

-검색어 vs 제공기관 태그 차이

2.5 연구방법의 한계

-데이터 수집: 크롤링을 통한 데이터 수집->공공데이터포털에서 정식으로 제공하는 데이터가 아님(제공기관 데이터의 누락)

-상관관계: 제공기관별 필터링…??

6. 결론

-선행연구에 없던 새로운 점을 발견했다. (선행연구 topic 무엇?)

-행정기관별 제공 데이터의 유의미한 질적/양적 탐색(가장 특이했던 결과 질적, 양적 1가지씩 요약)

-향후 행정기관별 데이터 제공 방식 제안

-> 태그(널값 없게)

-> 태깅 방식의 통일(--시 vs –행정부서)

-> 이용률을 높일 수 있도록 이용자 니즈 파악한 데이터 제공…?(현재 69퍼센트면 상당히 높은편..?)

-> 분류체계별 제공