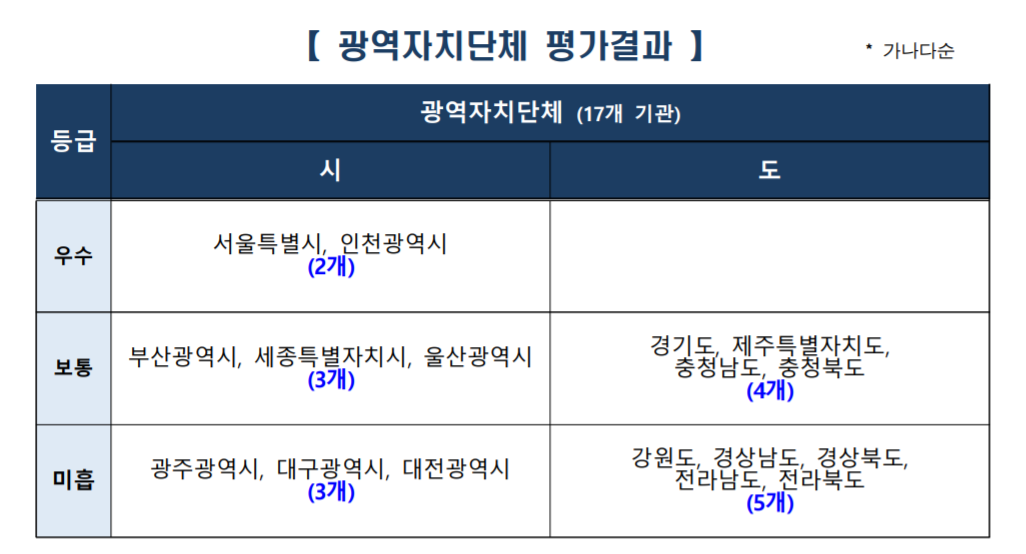
0726 회의록

저번주 to do list

* 앞으로의 방향성, 목차궝
* 지역별 특성 및 문제점 -> 지역별로 분석할 문제 정의(가능하면 지방자치 단체별 공공데이터 관련 지침?같은거 시청이나 이런데서 찾아보면 좋을 것 같습니다.)
* 관련 논문 읽어보고 의견있으면 제시하기.
* 웹 버전 공유 어떻게 할지.

<http://knconsulting.co.kr/knowledge/?q=YToxOntzOjEyOiJrZXl3b3JkX3R5cGUiO3M6MzoiYWxsIjt9&bmode=view&idx=3121806&t=board> 제 3차 공공데이터 제공 및 이용활성화계획



위 사진은 2019년 공공데이터 제공 운영실태 평가결과임(<http://www.odsc.go.kr/user/nd91584.do?View&pageLS=10&pageST=SUBJECT&pageSV=&page=1&pageSC=SORT_ORDER&pageSO=DESC&dmlType=&boardNo=00003063&dmlType=#attachdown>)

방향성 및 목차 관련 의견 -> 위 파일처럼 공공데이터 포털 관련 기준을 우리가 자체적으로 만들고 평가하는 식도 좋을 것 같습니다.

채은 - 제 3차 공공데이터 제공 및 이용활성화 기본계획을 분석 및 파악하고, 지역별로

공공데이터 관련 계획/지침이 있는지와 있다면 계획이 무엇인지 , 지역간 편차가 얼마나 있는지, 그리고 그 계획에 맞춰서 잘 운영되고있는지 등등. 최종적으로는 공공데이터포털에 잘 연계가 되어있는지를 중심으로 목차개요짜기

예) 0. 목적 ( 지역간 공공데이터 개방 현황을 분석하고, 격차없는 사회를 만들자..)

1. 지역별 공공데이터 ( 2020년 계획에 공공데이터 관련 계획있는지)

2. 지역별 공공데이터포털 현황

2.1 장단점 이나 특이점 분석

3. 정부 공공데이터 포털과의 차이점 및 우리가 했던 분석과 지역별 공공데이터포털의 평가 비교.

4. 결론 및 제언

송채은(부산, 대구, 울산, 경남, 경북)

* 부산 : 기본적으로 자체 공공데이터 포털이 잘 운영되고 있으며, 공지사항도 1달에 1번이상은 업데이트 되고 있었고 공공데이터 공모전, 수요조사 등 공공데이터 개방에 적극적으로 참여하고 있음.
* <http://www.busan.go.kr/nbtnewsBU/1378041?curPage=139&srchBeginDt=&srchEndDt=&srchKey=sj&srchText=> 계획
* <https://www.busan.go.kr/nbtnewsBU/1438063?curPage=&srchBeginDt=2020-06-05&srchEndDt=2020-06-12&srchKey=&srchText=>
  + 개방데이터 현황(7/27) : 파일 1852, api, 168, link 20.
  + 관련 지침 : <https://data.busan.go.kr/index.nm?contentId=17>

부산시에서는 ‘19년부터 구군별 다른 형식으로 공개하거나 일부 구청만 공개하고 있는 공공데이터 중 유용한 데이터를 선별하여 공통된 포맷으로 표준화하여 개방합니다.

부산시 표준데이터는 부산시 16개구의 통합하여 오픈 API로 형식으로 개방됩니다.

향후 지속적으로 대상 발굴하여 확대 개방해 가겠습니다.

* 부산시 표준데이터 : 제설함 배치 정보, 현수막 지정게시대 현황, 공개공지 현황 등 11종(‘19년 1월 기준)
* 부산시 표준 데이터 개방 절차 : 구군 공통 포맷 협의확정-데이터취합-db구축-api자동변환-개방
* 책임관 : 총 6명, 기획관, 통계빅데이터담당관, 공공데이터팀장, 주무관3명
* 부산광역시는 공공데이터의 민간 활성화 및 수요자 관점에서 유용한 공공데이터가 제공 및 이용되기 위해 아래와 같이 장기적인(2016년부터 추진) 서비스 모델을 구축 할 예정으로, 부산광역시에서 직접 개발한 서비스 모델을 민간 수요자에게 공개하여 사업화에 필요한 초기 아이디어를 제공하고, 이를 통해 공공데이터의 민간 활용도를 높여 일자리 등 부가가치를 창출할 수 있도록 하겠습니다. <https://data.busan.go.kr/index.nm?contentId=25>
  + 비콘활용, 민원 상담 db, 반려동물 종합정보 시스템 구축 등 15가지의 목표(서비스모델)을 설정해놓았으며 각각의 구체적인 계획도 첨부함.
* 대구 = 일단 대구시청에서 공공데이터개방을 들어가면 공공데이터포털과 연계하여 서비스함.

구별로 검색어 조정 안됨.

관련 담당자 : 총 3명. 책임관 1면, 담당관 1명, 주무관 1명. 각 구별로 담당자가 있긴 한듯함

<http://data.daegu.go.kr/open/main.do> 대구 빅데이터 서비스 포털

시청 홈페이지의 공공데이터와 실제 대구공공데이터포털인 D-데이터허브가 연결되어있지않음.

공지사항도 2019-06-04 이날 처음 포털 생긴거같은데, 이후 공지사항은 관리 안함.

다행히? 데이터들은 최신 데이터들이 올라옴 - 파일 1216, api 30, 그리고, 구별로도 검색가능

그리고 관련 사이트들은 많고, 데이터 업데이트는 많은듯함. 극히 대구 통합 통계포털이 잘 되엉ㅆ음.

장점 : D-데이터허브의 데이터들은 최신데이터와 구별검색가능, 관려사이트 제공으로 잘되어있음

단점 : 대구시청에서의 공공데이터는 홍보가 될 되어있지 않고, 오히려 별로임...대구만의 공공데이터 활용 지침이 보이지 않음.

계획관련 기사 : <http://news.imaeil.com/EconomyAll/2019100316282539282>

**대구시 공공데이터 공개범위 2022년까지 2.4배 확대**

[**http://news.imaeil.com/EconomyAll/2019100316282539282**](http://news.imaeil.com/EconomyAll/2019100316282539282)

* 울산 <http://data.ulsan.go.kr/user/apimng/dataset/totalList.ulsan?menuCd=DOM_000000101001000000&orderField=REGISTER_DATE&orderSort=DESC>

담당자 3명. 공지사항은 업데이트 주기가 약 6개월인듯함. 자주 하는 것은 아닌듯.

구별, 분야별 검색가능함.

파일 541, 시트91, 파트 4, 링크 12, 맵 5, api98

특이점 : 문화관광의 파일이 가장 많음.

<https://www.umca.co.kr/info/info02_0105_03.do?bmidx=19709&bid=BD000332&exec=view&currentPage=1&pagePerCount=10&searchTitle=SUBJECT&searchKey=>

->??????????? 다운로드 했더니 내용없음임…

관련 인력 3명

공공데이터 소개가 대구랑 똑같음.

* 경상북도 : <http://data.gb.go.kr/>

데이터 현황 : 시트204, 파일 29, 맵 127, api 75, 링크 28, chart 5

타지역에 비해 맵이 많음.

파인 업데이트는 정확히 파악은 안되는데, 7/27일 기준으로 27일 업데이트 자료 3개가 있었고 주기적으로 갱신되는 데이터가 많은 듯함. 그리고 파일에 연관 데이터 추천시스템이 있어서 좋았음.

* 경상남도 : <http://open.gyeongnam.go.kr/board/list.gyeong?boardId=BBS_0000001&menuCd=DOM_000000103001000000&contentsSid=28>

현황 : 파일 458, API 33

파일데이터는 2019-09-27이후로 옵데이트 안되고있음.

API 도 2018이후로 갱신 안됨.

공공공데이터 관련 계획도 없는듯함. 젤 별루…

총평 : 부산 > 울산 > = 대구,경북 > 경남.

200727 공공데이터포털 조사 (김동준)

1.공공데이터포털(data.go.kr) 內 문제 발생 지점 (지역 무관) (발견된 문제들을 수치화할 방법을 찾아서 지역별로 분석해보면 각 지역마다 두드러지는 문제점이 발견될 수 있을 것이라고 생각합니다.)

1-1) 남도축제영상

검색결과 리스트상에 표시된 확장자명 : MP4

메타데이터로서 기록된 데이터형 : 소리

실제 데이터 형식 : 영상

제공되고 있는 형식 : 링크 (링크 타고 들어간 웹사이트의 나머지 동영상들은 공공데이터포털에 공유 안돼있음. 그리고 제공년도도 2014년으로 오래됨)

==>> 업로드된 데이터에 대한 부가 정보를 메타데이터가 바르게 설명하고 있지 않았음. + 이 페이지로부터 얻을 수 있는 정보는 실질적으로 ‘전라남도 축제 영상을 시청할 수 있는 정보원’이었음에도 리스트상에서는 단일 영상자료만 제공하고 있는 것처럼 표시됨. 때문에 데이터가 지닌 의미가 크고 작고를 검색리스트 상에서 제대로 보여주지 못하고 있음.

<http://jntv.go.kr/kr/sub/sub.php?bid=vision04&category=5&menuid=1&page=1>

<https://www.data.go.kr/data/3036045/fileData.do>

1-2) 조회수가 0인데도 다운로드수가 찍혀있는 경우가 다수 존재.

가령 조회수가 0이고 다운로드수는 2인 경우들. 검색리스트상에서 다운받고서, 세부페이지를 보지는 않는 경우 때문에 이런 게 아닐까 싶었지만, 실험해보니까 그렇지도 않았음.

문제는 조회수와 다운로드수가 정량적으로 공공데이터포털을 평가할 수 있는 가장 대표적이고도 중요한 단서이기 때문. 조회수, 다운로드수의 산정방식에 문제가 있다면, 이를 이용하여 분석을 하는 것 역시도 결론을 도출하는 데에 오류를 발생시킬 것으로 예상됨.

따라서 조회수와 다운로드수가 어떠한 원리로 산정되는 것인지 알아보고서, 보다 정확한 수치를 알기 위해 다른 분석방법을 찾아야 할 듯함.

<https://www.data.go.kr/data/15060737/fileData.do>

<https://www.data.go.kr/data/15061579/fileData.do>

1-3) 차기등록예정일 vs 수정일

데이터 세부 페이지로 들어가서 나오는 여러가지 메타데이터들 중 ‘차기등록예정일’과 ‘수정일’을 볼 수 있는데, 차기등록예정일은 말 그대로 나중에 또 업데이트를 할 것인지, 그리고 ‘수정일’은 언제 수정이 되었는지임.

여기서 현재시각 차기등록예정일을 훌쩍 넘겼음에도 데이터가 갱신되지 않은 경우들이 있음. 대표적으로 아래 두 경우에서 이런 문제가 발생할 수 있을 듯함.

* 담당자가 일을 안해서 데이터를 업데이트 하지 않은 경우 (검색 리스트의 다른 곳에서도 동일한 주제의, 다른 날짜의 데이터를 발견할 수 없는 경우)
* ‘주기성데이터’로서 데이터를 추가한 게 아닌, 아예 다른 페이지에 데이터를 공유해놓은 경우 (검색리스트의 다른 곳에서 동일 주제의, 다른 날짜의 데이터를 발견할 수 있는 경우)

ex ) <https://www.data.go.kr/data/15044479/fileData.do>

<https://www.data.go.kr/data/15044480/fileData.do>

즉, 차기등록예정일과 수정일은 ‘공공데이터포털의 최신 데이터 업데이트가 적절히 진행되고 있는가?’라는 질문에 대해 정확한 지표가 될 수 있음.

<https://www.data.go.kr/data/3036045/fileData.do>

<https://www.data.go.kr/data/15050619/fileData.do>

<https://www.data.go.kr/data/3063772/fileData.do>

<https://www.data.go.kr/data/3082101/fileData.do>

<https://www.data.go.kr/data/3075702/fileData.do>

1-4) 파일명이 데이터의 특징을 똑바로 설명해주지 못하고 있음.

인천교통공사\_철도사고등 발생현황(2019.12.31. 기준) (<https://www.data.go.kr/data/15051222/fileData.do>)

까보면은 철도사고보다도 철도 운행 정보에 가까움. 이는 공공데이터포털 이용자들에게 필요한 데이터를 제공하는 데에 장애가 됨.

1-5) 메타데이터의 입력 기준 부재 (키워드도 제각각, 분류체계도 제각각, ...)

아래의 두 가지 데이터 모두 철도사고 발생상황을 보여주는 데이터지만 기입된 메타데이터가 매우 상이함.

분류체계 : 공공질서 및 안전 vs 교통 및 물류

키워드 : 부산도시철도,안전사고,부산지하철 vs 철도사고,운행장애,장애피해

<https://www.data.go.kr/data/15061229/fileData.do>

<https://www.data.go.kr/data/15061858/fileData.do>

이들에 대해 분석할 만한 요소를 찾으라면, 각 지역에서 제공된 데이터들에 대해 하나의 검색어(“철도사고”)에 따른 검색결과의 파일들이 어떤 메타데이터 분포를 나타내고 있는지(철도사고의 분류체계가 ‘교통’에 몰려있거나 ‘안전’에 몰려있거나. 또는 철도사고 데이터들의 키워드가 워드클라우드상에서 어떤 단어를 가장 많이 담고 있는지)를 확인하여 시각화가 가능.

만약 이러한 분석의 결과로, 해당 데이터가 가진 메타데이터가 사용자들의 직관과 맞지 않고, 해당 데이터를 찾는 데에 방해요소가 될 듯하면, 그 지역에 대해 개선이 필요한 사항으로서 이야기해줄 수 있을 듯함.

또는 논문의 결론이나 제언 부분에서 복합주제의 데이터를 복수 분류에 병행하여 표시해주도록 해야 한다고 귀띔하는 것도 괜찮은 방향으로 보임.

1-6) 공공데이터포털에서 특수하게 마련해놓은 분류서비스(이슈데이터 등)에 정작 데이터 제공이 필요한 지역의 것이 보이질 않음.

<https://www.data.go.kr/tcs/eds/idt/selectIssueData.do?searchCondition1=Y&issueId=343250&id=584&dType=&pageIndex=1>

위 데이터는 고령화 관련 이슈데이터를 모아둔 것. 등재된 지역은 경상북도와 제주특별자치도 데이터밖에 없고, 경상북도 중에서도 영양군 뿐임. 제주특별자치도에서도 서귀포시 단 하나.

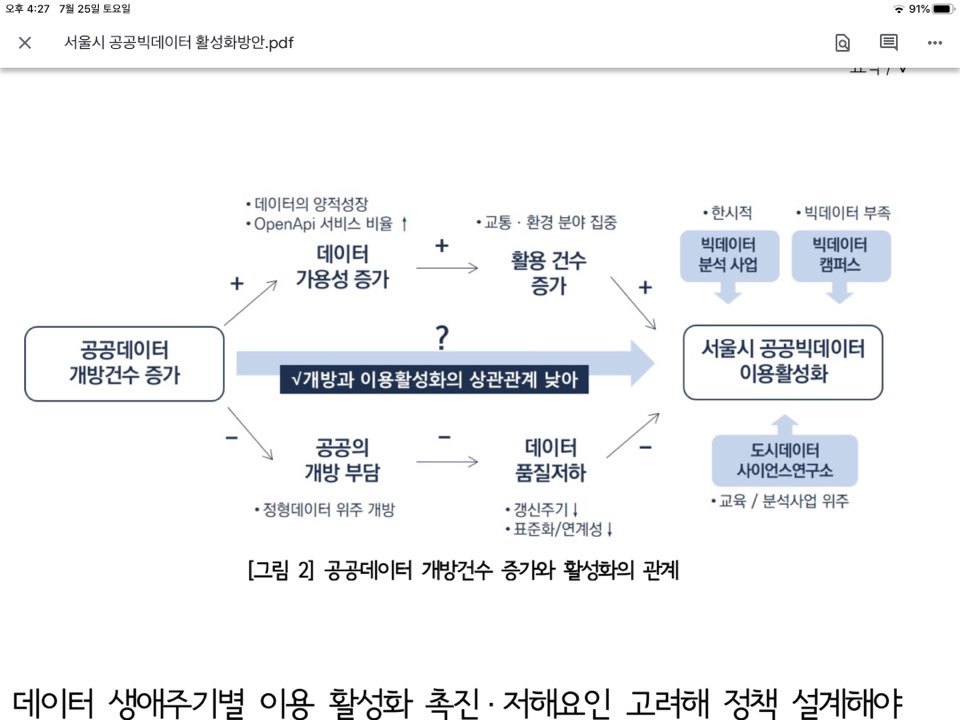
그러니 국토 행정구분에 있어서 2개의 시군밖에 고령화 정보가 제공되지 않은 것. 고령화 지수가 심각하게 나오는 경남, 경북, 전남의 추가적인 고령화 관련 지역데이터는 어디로 갔을지 의문. 그리고 경남, 경북에 비해 제주도는 비교적 덜 심각한 고령화를 보이고 있음.

즉, 정작 필요한 데이터는 제공해주고 있지 않고, 오히려 엉뚱한 정보만 고령화라고 갖다 놓고 있는 모습을 보이고 있는 것. 이는 같은 페이지 내의 또 다른 분류인 출산율도 마찬가지.

<https://images.app.goo.gl/NtUe9RSfceptakmp7>

<https://www.data.go.kr/tcs/eds/idt/selectIssueData.do?searchCondition1=Y&issueId=343250&id=585&dType=&pageIndex=1>

2. 선행연구 보완 및 이번 작업의 분석방향 고찰



위 그림은 공유폴더의 공공데이터 관련 논문 문서 중에서 ⸢서울시 공공빅데이터 활성화방안⸥에 포함된 내용으로, pdf 기준 p.9에서 찾아볼 수 있음.

⸢서울시 공공빅데이터 활성화방안⸥은 제목 그대로 서울시의 공공데이터 이용을 활성화할 수 있는 방안들에 대해 다각도적으로 접근하여 현황을 분석하고, 개선이 필요한 지점들을 도출하는 문서임.

위 그림은 공공데이터의 개방건수를 증가시킨다고 하여 그것이 반드시 공공데이터 제공 웹사이트의 이용량 증가로 이어지지는 않는다고 말하는 내용.

그림의 윗부분은 ‘제공량 증가->공공데이터 활성화’로 이어지는 과정을 말하는 것이고, 아랫부분은 ‘제공량 증가’->이용량 감소’를 말하는 것임.

윗부분의 인과가 더 크게 작용한다면 ‘개방->이용활성화’의 과정이 적용될 것이고, 아랫부분의 인과가 더 크다면 ‘개방->이용저해’ 문제가 발생될 것.

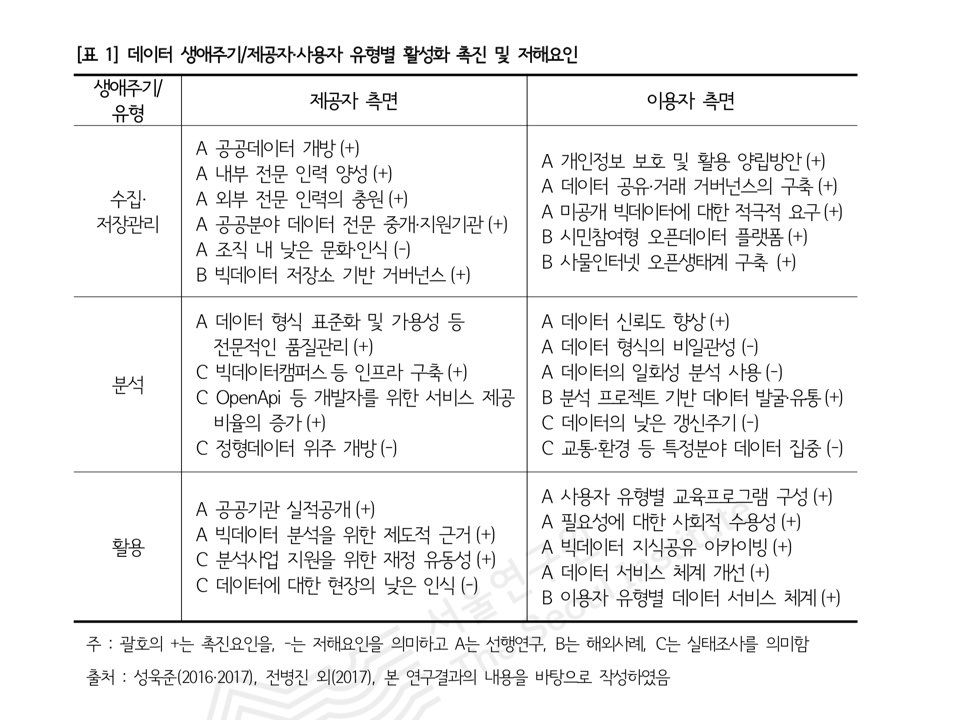
이 선행연구의 문제는 지역적으로는 대상이 ‘서울시’로 한정돼있다는 점, 그리고 시간상으로는 2018년 말경을 기준으로 분석이 이루어졌다는 점이 있을 것.

공공데이터 개방 정책이 마련된지 10년이 채 안된다는 사실을 감안하면 2년의 차이는 매우 큰 변화가 있을 것으로 예상되므로, 당시의 결과와 지금의 분석결과를 비교하여 공공데이터 개방의 각 세부 현황에 대한 개선과 후퇴를 말하는 것도 의미가 있을 듯함.

또한 전국적인 지역 공공데이터 개방 현황을 분석하고자 하는 방향성이 타 논문들과도 차별점을 만들어낼 수 있을 듯함.

공공데이터포털을 분석하기에 앞서 ‘어떠한 것을 분석할것인가?’, ‘무엇을 시각화해야 하는가?’, ‘어디서 문제가 발생했고, 이를 어떻게 효과적으로, 논리적으로 드러낼 것인가?’ 하는 질문이 생김.

이에 대해 해당 논문(서울시 공공빅데이터 활성화방안)에서 제공된 여러가지 활성화 촉진-저해 요인을 지표로 하여 전사적인 조사를 진행하고, 그에 따라 나타난 각 지역의 장점과 단점을 비교하여 특이사항을 도출하고, 그에 따른 개선방안을 제공하는 방향이 떠오름. 이러면 바로 윗 문단에서 말한 여러가지 질문에 대해 명확한 답을 제공받을 수 있다...



위 표는 동일 논문(서울시 공공빅데이터 활성화방안) p.125에서 찾아볼 수 있음.

앞서 먼저 제시한 그림은 인과관계가 너무 단순하다는 단점이 있었는데, 대신 데이터 생애주기(수집->분석->활용)에 기준을 두어 세부적으로 나누어 정리한 것이 이 표임.

이들 중 ‘개인정보 보호 및 활용 양립방안’과 ‘조직 내 낮은 문화-인식’, ‘필요성에 대한 사회적 수용성’ 등은 분석의 어려움과 주제 불일치 등 때문에 소재로 쓰이기 어렵다는 생각이 들음.

그렇더라도 ‘데이터의 낮은 갱신주기’, ‘교통-환경 등 특정분야 데이터 집중’ 등은 우리 팀에서 기말과제를 위한 분석을 진행했을 때 일정부분 진행했던 것들을 조금만 보완하면 되므로 분석이 충분히 가능할 것으로 보임. 이외에도 각 촉진-저해 요인들에 대해 데이터 분석이 필요한 것들은 분석해보고, 아닌 것들은 문장으로써 서술하여 논문에 추가할 수 있음.

김현지

* 서울열린데이터광장이 진짜 잘 되어 있음

이유 1. UI가 깔끔하고 세련

이유 2. 소장 데이터에 대한 다양한 분석을 진행

이유 3. 데이터 공개 형태의 표준화!

> 대부분의 데이터가 ‘서울시’ 전체의 데이터

> 간혹 구, 동별 데이터가 올라와 있다? 그러면 해당 데이터 제목에서 구, 동의 이름만 빼고 검색하면 다른 구, 동의 동일 데이터가 같은 형식의 제목으로 게시. 뿐만 아니라 해당 데이터의 하단에 ‘연관 데이터’에 같은 데이터 다른 구,동의 데이터 전체가 나열되어 있음. 굉장히 편리. 혼란도 없고, 이용자가 쉽게 데이터를 탐색할 수 있음

* 공공데이터 포털에 대한 분노
* 검색엔진 진짜 알다가도 모르겠음
* 뒤로가기 기능을 만든 사람이 뒤로간다는 개념을 전혀 모르는 것 같음.
* 지역별 검색기능 왜 사라진 걸까?
* 좀 묶어서 제공하면 안 되나? 제목이라도 통일하던가.. ..

전국의 대기오염정도를 제공하는 서비스를 개발하려고 할 경우, 그냥 전국의 지역명을 넣어서 검색해서 찾아와야 한다. 게다가 PK가 될 수 있는 컬럼도 다 다름.

강원도 : 시/ 동 =/= 충청남도 : 측정소코드

딱히 활용성이 높지 않아 보임. 대기오염과 같은 주기적으로, 전국에서 통일하여 업로드할 수 있는 데이터는 표준 형식을 정했으면 좋겠음. 아니면 동일 카테고리로 묶어서 한번에 목록을 찾아볼 수 있도록 하던가..

* **공공데이터 포털의 역할? 방향성?**

얘네 일을 어떻게 하는지를 좀 알아야 할 것 같음… 데이터를 지자체로부터 받아서 올리는지..아니면 따로 올리시는지...누가 관리하는지 등등

1. **가장 상위를 지역으로 하되, 지역적 특성을 갖는 데이터의 정리, 분류, 표준화에 대한 이야기를 하는 건 어떨까.. 싶음.**

> 기존의 분류체계에 지역별 분류 체계를 더하기

> 기존의 중구난방식 지역 데이터 형식, 메타데이터에 대한 분노

사실 활용도를 높이고 싶다거나, 제대로 된 분류를 만들고 싶다면 기준을 만들어 배포하는 식으로 표준화를 이루는 게 맞잖아. 엄청 시급진짜.

>> 하지만 이거를 어떻게 드러낼 것이냐… 큰 줄기를 ‘공공데이터포털의 지역 데이터 활용도 분석’으로 잡고 제언에 들어갈만한 이야기 같기도… ㅜㅜ

1. **지역별 활용/제공 양상 분석**

‘빅데이터 활성화를 위한 공공데이터포털 현황분석 및 개선방안’에는 지역별 데이터 개방 순위, 파일형태, 활용(다운로드)현황, 분류체계 등으로 전개.

우리는 아예 지역 전체(위의 내용) 훑고, 지역의 상위, 하위 분류체계별로 타이틀과 메타데이터 전체를 분석

> 워드클라우드를 통해 타이틀과 키워드, 설명 등에 가장 많이 나타난 단어를 찾아서 관광 문화에서도 특히 ‘자전거관광’에 대한 활용도가 높았다. 수요가 많았다 등등. 세부적으로 파고들어서 지역별로 어떤 데이터를 제공해야 하는지, 지역민 또는 지역별 이용자들이 어떤 부분에 관심이 많은지를 파악하는 것.

> 특정 년도에 유독 많은 데이터가 업로드, 적은 데이터가 업로드 되었다면 이유를 찾기(예를 들어 정권 교체, 지자체장의 교체가 있었을 때 데이터의 제공량이 들쑥날쑥하다면 일정한 공개 기준을 정해두어야 한다! 고 주장하기, 자율에 맡겨서는 안된다.) 또는 특정 년도의 데이터가 유독 많이 이용(다운로드)되었다면 그 이유를 찾기(2019년에 광주광역시에서 가로수 위치 데이터를 CSV파일로 제공하게되면서 이용량이 늘었다! 또는 2020년 코로나로 인하여 병원위치관련 데이터 활용도가 높아졌다..이런식 > 근데 이게 지역별로 더 파고들면 좋겠죠.. 예를 들어 포항 지진이 있었던 2017년 경상북도의 지진대피소 관련 데이터 활용률이 늘었다)

>> 리터럴리 현상 분석 > 의미를 끌어내는 것은 데이터 분석 이후에 가능함.

>> 문제정의라면, 지역별 데이터 제공 양상과 실제 활용이 일치하는지를 분석해본다는 점에서 의미가 있음. 이후 데이터 기반 행정 또는 시민들의 데이터 활용도를 높이기 위해서 필요.

1. **공공데이터포털에서 수요조사 한 걸로 아는데 이거 알아낼 수 있는지..?교수님.?**

**양지우**

1. **공공데이터포털 지역별 데이터**

1. 충북

2. 충남

---충청도 크롤링의 문제 해결방안 찾기!

3. 세종

4. 대전

--지역별로 살펴봤을때는 데사개 과제 결과와 유사

**\*공공데이터포털 공통점**

-법률, 통일외교안보 분야의 데이터가 가장 적음

-대체 표준데이터는 무엇인가?

-2019년부터 데이터 양이 증폭

-검색연산자는 왜이럴까?

-지역적 특성과 데이터 분야의 특성 연결 가능..?

1. **지자체 포털**

**\*지자체 포털 공통점**

-데이터만을 다루는 별도의 포털은 존재하지 않음

-사이트로 링크 연결 혹은 사이트 내에 연결해놨음

-공공데이터 이외의 지자체 데이터(주로 행정,정책 관련 데이터)

-업데이트 주기는 짧은 편임

-한글, 워드, 피디에프같은 파일을 주로 업로드하고 있음

-데이터를 형식별로 필터링할 수 없음

-정보공개와 데이터의 차이..?

-정보공개의 분류 상이

**>>지역의 ‘분석’ 보다는 분석에 기반한 공공데이터 포털의 지역별 데이터 ‘평가’ 쪽에 비중을 두는 것이 어떨까…**

**>>문제점&추후 개선점 제안**

**>>기존 논문과의 차이점…? ㅎㅠㅠ**

**>>데이터 관계 모델링을 해야할까…? 링크드 데이터?분석/시각화 말고…?(현재 공공데이터포털과 지자체 포털 데이터의 연결문제 해결/공공데이터포털 내의 데이터 중복, 분산 문제 해결)->동준님의 의문점 해결할수있을수도..?(공통된 메타데이터 표준 부재)->현지의 업청시급진짜 문제점도 해결할수있을수도…**

**>>다빈출 단어를 통한 데이터 모델링/시각화/클러스터링**

**>>하지만 노가다 작업을 할수없다….**

**>>그럼 텍스트마이닝…연결관계 시각화…**

**>>왜 우리가 데이터 표준화를 해야하는것인가…**

1. **공공데이터포털 메타데이터 표준화 제안**

**지식그래프^\_^**

1. **지역별 데이터 현황 분석**

* **지역별 포털 분석**

**충북충남세종대전은 빼줘,....없는거나 마찬가지야...**

1. **관련 정보**

**[큰그림을 그려보았다.-그냥 내 생각….(양지우)]**

**공공데이터포털의 데이터 표준화와 의미 연결을 위한 방안 연구**

**0. 서론. 논문 2개 언급 이렇다~. 여기서**

1. **<공공데이터포털 데이터 분석>**

**크게 = 지역별 포털이 편차가 심함. (논문 2개 언급 이렇다~. 여기서)**

**문제의식 : 구체적 = 동준님 분석 유니크 위주로 문제점 제안 전체적으로 안좋은예시 말로 정리**

1. **<공공데이터포털 메타데이터 / 분류체계 표준화 제안> = 교수님의견 전적으로 받든다. 교수님 논문 발전시켜봐요.**

**데사개에서 했던 분석을 통해 지역별 데이터 및 공공데이터 포털 전체 데이터의 문제점 발견**

**-> 지역별/분야별(공공데이터포털)/활용별(활용도 높은거?) = (?)단어추출 및 클러스터링 -> 연결 관계 분석/모델링 -> 시각화(지식그래프) = 교수님 설명 들어야겠음.**

**===> 키워드, 주제명 표목을 표준화할 수 있는 지표가 될지도.**

**지역별 키워드 색인어 만들 수 있을 듯.**

1. **<지역별 데이터 포털과 공공데이터포털의 연계성 분석>**

**지역별 포털-공공데이터포털과 데이터 연결성 분석**

* 개별 아이템 : 깃허브 파요~~~~~~~포크해가요~~~~~~

> 특정 년도에 유독 많은 데이터가 업로드, 적은 데이터가 업로드 되었다면 이유를 찾기(예를 들어 정권 교체, 지자체장의 교체가 있었을 때 데이터의 제공량이 들쑥날쑥하다면 일정한 공개 기준을 정해두어야 한다! 고 주장하기, 자율에 맡겨서는 안된다.) 또는 특정 년도의 데이터가 유독 많이 이용(다운로드)되었다면 그 이유를 찾기(2019년에 광주광역시에서 가로수 위치 데이터를 CSV파일로 제공하게되면서 이용량이 늘었다! 또는 2020년 코로나로 인하여 병원위치관련 데이터 활용도가 높아졌다..이런식 > 근데 이게 지역별로 더 파고들면 좋겠죠.. 예를 들어 포항 지진이 있었던 2017년 경상북도의 지진대피소 관련 데이터 활용률이 늘었다)

* ㅇ기

**[교수님한테 물어볼 것]**

1. **지방자치단체 데이터 납품 어떻게?**
2. **앞으로 다 흡수? 아니면 각자도생으로 잘살기? 그럼 겹치는거 어떻게?**
3. **앞으로 각 단계 데드라인!!!!!!!**