시각화/데이터분석 수정사항 (2/7)

# 전문가 토픽모델

## ‘토픽모델 인터페이스’

### ‘토픽저장’

#### 토픽 ‘카테고리 자동제시’ – 구성한 ‘토픽모델 노드 모든 항목’을 과학기술표준분류 API에 입력하여 ‘과학기술표준분류코드’를 도출하고 ‘토픽 카테고리’를 자동 제시

##### ‘저장할 카테고리를 선택해주세요’칸 밖에 ‘카테고리 자동선정’ 버튼 추가

##### ‘카테고리 자동선정’ 버튼을 누르면 카테고리 자동 선정

##### 과제분류 API 연계 요청 또는 과제분류 사전학습모델 생성 필요

#### 자동제시 ‘카테고리’ 수정

##### 토픽모델이 변경되면 자동분류 API 재적용

##### ‘저장하기’ 누를 때마다 API 활용 분류 업데이트 필요

##### 자동제시 ‘카테고리’가 적절하지 않을 때

###### 현재 ‘비활성’상태를 ‘활성’으로 변경

###### 사용자가 적절한 카테고리를 고르거나 ‘검색’할 수 있게 변경 🡪 ‘토픽 카테고리 like 검색’ 참조

#### 토픽 카테고리 Like검색 – 현재 콤보상자를 열면 목록이 보이는 화면에서 ‘카테고리’명을 입력하면 관련 카테고리만 보이게 처리

#### 토픽저장 후, ‘토픽저장’모달에 기존 ‘카테고리’와 ‘토픽설명’이 남아있는 경우가 있음

##### 토픽베이스 검색이 실패 후, 다른 토픽베이스를 검색하여 저장하려 할 때

### 툴팁 추가 - 노드크기보다 긴 용어는 ‘툴팁’으로 모든 글자가 보이게 처리

### 추천 키워드 정렬 –가나다순으로 정렬하여 사용자가 용어 검색 편의성 향상

#### ‘중요도’와 ‘가나다’순으로 번갈아 선택 가능하게 구현

##### ‘중요도’순 – 현재 보이는 방식

##### ‘가나다’순 – 키워드 리스트는 정렬하여 제시

#### ‘추천 키워드’제목 우측에 ‘중요도순’, ‘가나다순’ 버튼을 추가

### 토픽모델 구성화면 단계별 버튼 추가

#### 1, 2, 3단계 버튼을 각각 추가하여 동시에 여닫기

### 토픽내 ‘워드’입력 편의성 개선

#### ‘워드’입력 후, 워드 연속 입력 기능 개선

##### ‘워드’입력 후, 같은 토픽내 워드 추가입력창 자동제시

##### 추가입력이 없을 때, ‘esc’로 빠져나오기

#### ‘워드’입력 할 때, 동시 입력

##### ‘워드’등을 콤마( , )로 분리하여 입력하면

##### 콤마로 분리한 ‘워드’ 등을 동시에 여러 노드입력

#### ‘워드’입력 할 때, 동의어 사전 도입

##### ‘워드’입력 할 때, 동의어 사전상 동의어 자동 제안 🡪 엔터만으로 자동 노드 추가

##### 차차 토픽모델내 동의어를 동의어 사전으로 자동 증보

### 토픽, 워드 선택 순서 변경

#### 현재 – 제외, 포함, 미포함

#### 개선 – 포함, 제외, 미포함

### ‘토픽재료’ 검색시 ‘강제 검색’

#### 용어사전 추가 등 변경이 있을 때, 캐시를 지우고 재검색해야 할 필요

#### 현재 –기존 캐시가 있을 때, 캐시 데이터(추천키워드, 워드클라우드, 소시오그램, 토픽베이스) 가져옴

#### 개선 – ‘강제 재검색’ 옵션을 검색란 우측 밖에 reset 아이콘으로 선택가능 🡪 캐시데이터 지우고 다시 검색

### 토픽모델 편집화면 빠져나갈 때, 경고알림 추가

#### 모델 편집중 빠져나갈 때, 자동저장 또는

#### 빠져나갈 때, 경고알림(모달)

### 토픽 삭제

#### 토픽 삭제 불가

#### ‘토픽’ 삭제할 때, 알림창

##### 현재 - ‘최소 한 개이상의 토픽을 선택 되어있어야 합니다.’

##### 개선 – “토픽을 한 개 이상 선택해야 합니다.”

## 토픽 모델링 인터페이스 구조

### 개선 필요성과 방향

#### ‘토픽재료찾기’ – 재료 찾기에 필요한 ‘키워드 입력란’만 필요, 아래 추천키워드 등 세 가지 재료 상자는 비어있으며 현재 표출할 필요가 없음

#### 사용자 사고흐름을 연결하는 표출 필요 – 사용자가 ‘키워드’를 입력하여 찾았던 ‘토픽베이스’를 내림차순으로 보여줄 필요

#### 개선방향

##### ‘키워드 입력 전’ - 세 가지 재료 먼저 보여줄 필요가 없으므로 제외하고, 기존에 찾았던 ‘토픽베이스’를 제시할 필요 (현재 이력이 없다면 ‘토픽베이스 생성’이력 테이블을 만들필요

##### ‘키워드 입력 후’ – 현재 검색 결과처럼 세 가지 토픽재료 들을 보여줌

##### ‘토픽재료 찾기’ 클릭시 – ‘키워드 입력 전’ 화면처럼 다시 기존 ‘토픽베이스’ 이력을 제시

### AS-IS – 개선 전

#### 왼쪽 - 토픽재료와 토픽모델 찾기

##### 토픽재료 찾기

###### 입력

키워드 입력란

###### 출력

추천키워드

워드클라우드

소시오그램

##### 전문가 토픽 불러오기

###### 입력

검색 토픽명 입력 (like검색)

카테고리 선택

###### 출력

검색 전

없음

검색 후

‘검색어’와 ‘카테고리’ 검색 결과 ‘전문가 토픽모델’을 내림차순으로 표출

### TO-BE – 개선 후

#### 왼쪽 - 토픽재료와 토픽모델 찾기

##### 토픽재료 찾기

###### 입력

키워드 입력란 🡪 토픽모델명으로 활용

장문 입력란 🡪 연구제목, 연구내용, 키워드 등을 함께 입력 (입력란내에 구분 없음)

###### 처리

입력받은 ‘키워드’를 토픽모델명으로 적용

LDA분석 - ‘키워드입력란’ + ‘장문입력란’ 내용을 함께 모아서 API로 전달

###### 출력

검색 전

기존 ‘토픽베이스’ 검색 내역 내림차순 표출

출력사항 없음

검색 후

추천키워드

워드클라우드

소시오그램

##### 전문가 토픽 불러오기

###### 입력

검색 토픽명 입력 (like검색)

카테고리 선택

카테고리 선택 – 기존 방식 포함

카테고리 검색 추가

카테고리 선택란에 검색어로써 카테고리 3분류에 LIKE검색 하여 ‘카테고리 리스트’ 선제시

‘카테고리 리스트’에서 ‘카테고리’선택하면 해당 카테고리 입력

###### 출력

검색 전

기존 ‘전문가토픽모델’ 생성 내역 내림차순 표출

검색 후

‘검색어’와 ‘카테고리’ 검색 결과 ‘전문가 토픽모델’을 내림차순으로 표출

## 검색식

### 검색식 토글

#### ‘검색식’ 버튼명 변경 – ‘검색식’ 🡪 ‘토픽모델로 검색’ 명칭 변경

#### ‘과제 검색하기’ 누르면

##### AS-IS – “생성된 검색식으로 과제를 검색하시겠습니까? 저장되지 않은 토픽은 소실될 수 있습니다.”

##### TO-BE – “토픽모델 검색식으로 과제를 검색하시겠습니까? 검색전에 토픽모델을 자동저장합니다.”

###### 1단계 - 토픽모델을 자동저장

###### 2단계 – 검색 실행

# 전문가 토픽모델링

## 검색 멀티프로세싱 처리

### 워드클라우드, 소시오그램, 토픽베이스, 추천키워드 생성을 멀티프로세싱

## 용어 소시오그램 형성

### 네트워크 분석에 ‘검색어’는 제외 – ‘서비스로봇’ 소시오그램에 ‘서비스검색’노드 포함

## ‘토픽재료 찾기’ 검색어 ‘공백’ 조합 개선

### 필요성과 방향 – 검색어에 ‘공백’을 포함할 때, 검색결과는 같지 않다면 개선할 필요

### 현재 이슈

#### 검색불가 유형

##### ‘웨어러블 디바이스’ 검색 불가 🡪 ‘웨어러블디바이스’ 검색 가능

##### ‘맞춤형 바이오 치료제’ 검색불가 🡪 ‘바이오치료제’ 검색 가능

#### ‘공백’을 포함하였을 때 다른 토픽베이스 산출

##### 사례: ‘인공 지능’, ‘인공지능’의 토픽베이스 다름

### 개선 방향

#### 검색시 공백 처리

##### 공백 처리 조합하여 입력

###### 공백 포함 검색어

###### 공백 제거 검색어

##### 두 가지 공백 처리 방식 병행

###### ‘인공 지능’을 입력하면

###### ‘인공 지능, 인공지능’을 API 에 입력

#### ‘공백’을 포함한 검색어는 ‘공백 제거 검색어’로

## 흔한 단어

### ‘스마트산업기계’ 검색불가

## 토픽별 중복되는 단어 처리 필요

### ‘설계’, ‘분석’, ‘시스템’, ‘사용자’, ‘정보’ 등 토픽별로 자주 등장하는 ‘워드’

### 자주 등장 ‘워드’는 제거하거나 다른 ‘워드’로 대체 필요.

## 모델링 편의성

### 토픽모델에 동의어 자동 제시

#### 워드 추가할 때, 자동 제시

#### 키워드(상단 후보 ‘워드’) 드래그 추가할 때, 자동 제시

#### 토픽베이스에 자동 생성한 ‘워드’의 동의어 관련어 생성

### 토픽모델 내 ‘워드’노드에 ‘동의어/관련어’ 생성 아이콘 추가

#### ‘토픽베이스’ – 생성과정

##### 자동 생성한 ‘토픽베이스’내 ‘워드’노드에 아이콘 추가

##### 자동 생성과정에서 ‘토픽베이스’내 ‘워드’노드에 ‘동의어/관련어’ 자동 추가

#### ‘토픽 모델’ – 모델링 과정

##### 토픽모델링 과정에 ‘아이콘’을 눌러서 자동 추가

### 토픽구조 붙여넣기

#### ‘토픽’수준 노드 추가시

##### 기존 ‘토픽모델명’을 서치(일치하는 최근 토픽모델명)

##### 검색한 토픽모델의 ‘토픽들’을 불러오기

##### 불러온 ‘토픽들’을 ‘모델에 입력한 토픽수준’의 하위 ‘워드’로 붙여넣기

## 토픽모델링 동작 개선

### 노드 활성화 개선

#### ‘전기자동차’ – ‘hydrogen’ 동의어 활성화 불가

### 노드 이동과 중복

#### 하위 레벨에서 상위레벨로 이동시 ‘중복 메시지’

##### 하위 노드가 삭제, 상위 노드를 생성 🡪 ‘중복’아니며 노드‘이동’

### ‘토픽모델’ 검색 후, 상태 유지

#### 검색 후 ‘ < ’눌러 되돌아온 상태 개선

##### 현재 기존 ‘추천키워드’, ‘워드클라우드’, ‘소시오그램’은 남아있음

##### 검색에 사용한 ‘토픽모델’이 보이지 않음 🡪 저장한 토픽모델이 아닌 ‘토픽베이스’가 보임

##### 개선방향: 검색했던 토픽모델이 보여야함.

## 포함어 제외어 처리 안내 (가이드라인 외에도 캔버스에 표출)

### 토픽재료 3종에 제외어가 있을 때, 토픽모델에 ‘제외어 처리’

#### 제외어란

##### ‘검색어’에는 부합하지만

##### 사용자가 모델링하는 토픽주제에는 적절하지 않은 용어 🡪 제외어

#### 제외어 사례

##### ‘인공지능’ 제외어 – ‘조류독감’

##### ‘모빌리티 서비스’ 제외어 – ‘유전자’

### 제외어 처리 방식을 어딘가에 설명할 필요

#### 토글 이나 도움말 처리

##### 위치 – 우측 토픽모델링 캔버스

##### 토글 설명

###### “왼쪽 세 가지 토픽 재료의 단어들을 클릭하고”

###### “캔버스에 나타난 단어를 토픽모델 적당한 위치에 드래그합니다.”

###### ‘포함어’나 ‘제외어’로 설정합니다.

## 토픽재료 검색 개선

### ‘검색어’가 재료로 생성에 부적절할 때

#### 현상

##### 입력한 검색어가 아닌 기존 검색어의 재료와 토픽베이스가 그대로 나타남

#### 원인

##### 검색어가 EdgelistTerm 색인에 존재하지 않을 때

##### 검색어 관련 문서나 용어가 충분하지 않을 때

#### 조치

##### 검색 직전에 기존 검색 결과, ‘재료’와 ‘토픽베이스’를 비움

##### 검색결과가 없을 때, 메시지를 표출

#### 재료생성 불가능할 때, 메시지 표출

##### 추천키워드 메시지 – “검색어를 달리하여 입력하거나 복합어 또는 문장으로 입력해보세요”

##### 워드클라우드 메시지 - “검색어를 달리하여 입력하거나 복합어 또는 문장으로 입력해보세요”

##### 소시오그램 메시지 - “검색어를 달리하여 입력하거나 복합어 또는 문장으로 입력해보세요”

#### 토픽베이스

##### 기존 검색어 토픽베이스가 유지

##### 메시지 표출 – “검색어를 달리하여 입력하거나 복합어 또는 문장으로 입력해보세요”

### 노드 일괄 처리

#### 일괄 삭제 – 불필요한 노드를 한꺼번에 삭제

##### ‘포함어’, ‘제외어’를 선택 후,

##### ‘미포함’(비활성 회색노드)를 한 번에 지우는 기능

#### 일괄 ‘포함어’ 또는 ‘제외어’ 처리

#### 모든 ‘워드’ 동의어 관련어 붙이기

## ‘토픽재료 찾기’ 개선

### ‘토픽재료 찾기’ 입력값 개선 필요

#### AS-IS: ‘단어’, ‘복합어’, ‘단일 문장’ 수준으로 ‘토픽베이스’ 완성도가 낮음

#### TO-BE: 연구내용 등 장문을 ‘토픽베이스’ LDA분석에 활용

##### ‘키워드 입력란’ – ‘토픽모델’의 명칭으로 활용

##### ‘장문 입력란’ 추가 – ‘연구내용’ 등을 한 칸에 추가 입력하여 보다 풍부한 ‘토픽베이스’ 생성가능

###### 과제명

###### 연구목표, 연구내용, 키워드

### ‘토픽재료’ 검색이력

#### 기존 토픽재료 검색이력 표시 🡪 클릭으로 즉각 캐시 확인

#### ‘토픽재료 검색이력’의 중복 최근 검색했던 이력은 중복하지 않기.

## 과제입력으로 토픽모델링

### 단일 과제 토픽베이스 형성 - 과제 한 건을 입력 토픽모델 형성

### 단일 과제 유사과제 검색 - 과제 한 건으로 생성한 토픽모델로써 유사과제 검색

### 과제 토픽베이스 캐싱

#### 과제 토픽베이스 캐싱 🡪 과제 80만건 토픽베이스 형성 자동화

#### 과제 토픽베이스 형성과정

##### 입력

###### 토픽모델명 – 과제명

###### 토픽베이스 생성 입력 – 과제명, 연구내용, 연구목적, 키워드

##### 처리 – LDA 기존 로직

##### 출력

###### 과제 토픽베이스

###### 과제 토픽재료 – 추천 키워드, 워드 클라우드, 소시오그램

# 키워드 검색

## 워드클라우드 수량과 시각화지표 - 조정 개선 필요

### 워드클라우드 시각화 지표

#### 4~12는 4, 8, 12 위주로 나타남 🡪 데이터팀 확인

### 워드클라우드 시각화

#### 워드 수량 – 50개~70개

#### 워드 폰트 – 디자이너와 상의

#### 클라우드 모양 – 현재 정방형에서 원형 가능한지 라이브러리 파라미터 확인 필요

### 워드클라우드 잔상

#### 키워드 검색 후, 다른 검색 시도하면 기존 검색 결과 ‘워드클라우드’가 희미하게 렌더링 함.

## ‘검색 결과 화면’

### 검색 입력 ‘키워드’나 ‘문장’ 표출 필요 – 입력한 내용을 알 수 없으므로.

#### ‘검색결과 X건’ 옆 등 적절한 곳에 표출

## 검색 안되는 케이스 검토

### 검색결과가 나타나지 않음 – 이유는 ‘/’를 포함하고 있기 때문임을 추정

#### “100V/600A 급 MOSFET 모듈 개발”

#### “생체신호의 초감도 감지를 위한 저차원 소재기반 플렉시블/웨어러블 기기 개발”

#### ‘/’을 제외하고 입력하면, 두 가지 항목 검색 가능

## 키워드 검색 ‘문장’ 검색이력

### ‘문장’ 검색 후 이력 – ‘검색 이력’에 이력(‘키워드’, ‘문장’)이 나타나게 처리.

# 과제 검색

## 검색 결과화면

### 검색결과 화면에 ‘검색에 입력한 내용’을 표현 (결과는 있으나 무엇으로 검색했는지 모름)

# 과제 선택(다차원 큐레이션)

## ‘과제필터링 도구’(‘과제필터하여 선택하기’)

### ‘과제/용어 네트워크’(표준분류)

#### 과제 네트워크 (paper) – ‘날짜 설정’만 활성화 (‘네트워크지표 설정’ 비활성화)

#### 용어 네트워크 (word) – ‘네트워크지표 설정’만 활성화 ( ‘날짜 설정’ 비활성화)

##### 주요 용어 10~20개 용어 표기

##### 하단에 ‘용어목록’ 슬라이드 – 소시오그램에 주요 용어를 표출 또는 소시오그램 아래에 용어목록을 확인할 수 있는 슬라이드(기본 숨김 형태)

## 사업내 다수 과제 인터페이스

### 콤보박스 형태 시인성 개선 – 적절한 기능 구현했지만, 작고 직관적이지 않음

### 선택 단계(phase) 도입

#### 단계 – ‘사업내 과제목록’ > ‘유사과제’ > ‘선택과제’

#### 인터페이스 – ‘사업내 과제목록’을 ‘필터도구’위치 등에 ‘목록화’하여 시인성 확보

## 필터 도구와 과제목록간 데이터 정합

### 소시오그램 과제노드 선택항목 과제목록 존재여부 🡪 데이터팀과 상의

### 워드클라우드 ‘용어’ 선택항목 과제목록 존재여부 🡪 데이터팀과 상의

## 필터도구 ‘노드’, ‘용어’ 선택 다중 선택

### 워드클라우드 선택한 용어(2개이상) – 동시만족 과제 선택 기능

## 유사과제 선택하기

### ‘과제상세 보기’

#### 과제 목록 각각 ‘∨’는 없어도 과제 ‘클릭’만으로 볼 수 있음

##### ‘∨’ 업애기

##### ‘∨’ 자리에 ‘ > ’추가하여 체크박스 체크가 없이도 ‘선택과제’로 개별 이동 가능하게 처리

#### 펼침목록 – ‘연구목표’, ‘연구내용’, ‘기대효과’는 ‘더보기’ 버튼 제거

##### ‘더보기’ 버튼 제거 후, ‘더보기’기능을 ‘연구목표’등 해당 셀(내용)을 눌렀을 때 펼치기

##### ‘상세보기’ 버튼은 우측으로 붙이고, ‘과제 조사분석 상세’로 변경, 모달창 제목도 같이 변경

### 정렬 도구

#### 현재 – ‘키워드 관련성’, ‘과제영향력’, ‘과제 매개력’, ‘과제 연결성’, ‘최선성’

#### 개선 – ‘체크리스트’를 ‘on/off’버튼으로 추가

##### 정렬도구 창에서 체크한 과제가 멀리 떨어져있을 때, ‘체크리스트’만 시인성 확보 차원

##### 참고

###### ‘체크리스트’ on 상태에서 체크한 과제 중 한 과제 체크를 없애면 해당 과제는 체크리스트에서 제외되어야 함.

###### ‘체크리스트’를 off하면 체크 항목을 유지하면서 제크하지 않은 과제 전체가 보여야 함.

#### 선택 키워드 입력란 개선

##### 현재 – 한 단어만 입력

##### 개선 – 두 단어 이상을 콤마 분리하여 입력하면 여러 단어를 처리할 수 있는지 확인 필요

### 필터링 도구

#### 소시오그램에 과제 상태 표현

##### ‘선택과제’ – 선택완료한 과제 🡪 진청

##### ‘체크리스트’ – 선택하기 위해 체크한 과제 🡪 파랑

##### Mouse rollover 과제 – 과제목록에서 마우스를 롤오버한 과제 🡪 적색

#### 소시오그램에 주요과제/주요용어 레이블 추가

##### 과제 네트워크 – 상위 10과제 레이블 추가

###### 10개를 집계하기 어렵다면 degree나 weight를 이용

##### 용어 네트워크 – 주요 10용어 레이블 추가

###### 10개를 집계하기 어렵다면 degree나 weight를 이용

#### 소시오그램 – 용어네트워크 노드선택 상태 표시

##### 워드클라우드의 용어선택을 차용하여, 하단에 선택한 용어 노드가 나타나게 함

#### 두 가지 필터링 도구 크기와 하단 처리

##### 과제/용어 네트워크

###### 소시오그램 캔버스 – 소시오그램 조절도구들을 하단에 밀착하고 최대한 캔버스 확보

##### 워드클라우드

###### 워드클라우드 캔버스 – 아래 조절도구를 하단에 ‘선택용어’창 위로 밀착 캔버스 사이즈 확보

#### 워드클라우드 용어조합 (기정원 프로젝트 적용)

##### 현재 – 용어를 포함한 문서 ‘합집합’ 처리

##### 개선 – 용어를 포함한 문서 ‘합집합’ 또는 ‘교집합’ 두 가지 처리

##### 표시방식 – 워드클라우드 ‘선택용어 창’ 우측에 ‘합집합’ 벤다이어그램, ‘교집합’ 벤다이어그램 아이콘을 넣어서 활성/비활성 할 수 있게 처리

##### 필터적용

###### 현재 – 두 개 이상 용어를 선택하면, 두 개 용어중 하나라도 포함하는 과제를 체크함

###### 개선 – 두 개 이상 용어를 선택하면,

‘합집합’ 활성화 – 두 개 용어중 하나라도 포함하는 과제 체크

‘교집합’ 활성화 – 두 개 용어 모두 포함하는 과제 체크

### 유사과제 선택상태 표시 (KISTEP 적용)

#### 왼쪽 ‘유사과제 선택창’에서 체크한 항목의 개수를 나타냄

#### 표시형식 - 선택목록 해더 ‘과제명’ 옆에 ‘( 23건/500건)’ 🡪 과제명 (23건/500건)

## ‘선택 과제’ 처리

### ‘과제 규모산정’이동 방식

#### 현재 – ‘과제규모산정’ 눌러서 ‘과제규모산정’으로 바로이동

#### 개선 – ‘과제규모산정’ 누르면 자동저장하고 ‘과제규모산정’화면으로 이동

### 선택과제

#### 정보제공 – 현재 ‘과제명’툴팁에 ‘연구비’와 ‘당해년도 연구기간’을 추가하여 제공

#### 정렬 및 집계 – 유사과제를 집계하나 각 과제의 연구비 구성을 알 수 없는 상태

##### 정렬기준 – 연구비(내림차순), ‘당해년도 연구기간’ 종료일(내림차순), 수행기관(오름차순)

##### 집계 – 하단 테이블 footer에 최저, 최고, 평균, 중위수를 자동계산하여 제시

# 시각화

## 시각화 데이터 구성

### 키워드별 과제 수행 집계(건 수)

#### 초기(용어 미선택시) – 과제 전체 수를 1개 곡선으로 년도별로 표시

#### 용어 선택 시 – 해당 용어 관련 과제를 용어 수만큼 곡선으로 년도별 표시

### 논문, 특허, 과제 수행 집계(건 수)

#### 초기(용어 미선택시) – 전체 특허, 논문, 과제 수를 각각 곡선으로 년도별로 제시

#### 용어 선택시 – 해당 용어 관련 특허, 논문, 과제 수를 용어 수만큼 곡선으로 년도별 표시

### 1억당 논문/특허 집계(건 수)

#### 초기(용어 미선택시) – 전체 과제의 1억당 특허, 논문 수를 각각 곡선으로 년도별로 제시

#### 용어 선택시 – 해당 용어 1억당 특허, 논문, 과제 수를 용어 수만큼 곡선으로 년도별 표시

## 시각화 범례(레이블) 구성

### 세 가지 곡선형 그래프 ‘전체’와 ‘용어선택’을 구분

#### 초기(용어 미선택시) – ‘v전체’라는 범례를 생성 현재 그래프가 용어전체의 집계임을 나타냄

#### 용어 선택할 때 – 초기 ‘v전체’라는 범례를 비활성화하고, 해당 용어 범례가 추가하여 활성화함

#### 용어 선택을 모두 해제할 때 – ‘v전체’범례를 활성화하여 용어전체의 집계를 나타냄.

## 시각화 페이지 상단에 검색 주제 표시

### 키워드, 과제, 토픽 검색 주제 표시

#### 키워드 – “검색한 주제를 표현하는 곳” 🡪 키워드 - ‘인공지능 자율주행 자동차’

#### 과제 – “검색한 주제를 표현하는 곳” 🡪 과제 - ‘소재부품산업 미래성장동력’

#### 토픽 – “검색한 주제를 표현하는 곳” 🡪 토픽 - ‘스마트 에너지 관리 시스템’

### 시각화 레이블 변경

#### 소시오그램 박스 레이블 변경

##### 현재 – “과제연관 최근 10년 현황입니다”

##### 개선 – “검색한 모든 과제와 수행기관 네트워크입니다. 색깔이 있는 노드는 선택한 과제입니다.”

#### 워드클라우드 박스 레이블 변경

##### 현재 – “과제연관 최근 10년 현황입니다”

##### 개선 – “아래 키워드를 1개이상 선택하면, 관련 과제와 특허, 논문 흐름을 알 수 있습니다.”

#### 논문, 특허, 과제 박스 레이블 변경

##### 현재 – “설명 없음”

##### 개선 – “선택한 키워드 관련 과제, 특허, 논문의 건수와 1억당 집계를 나타냅니다.”

#### 주요키워드 추이 박스 레이블 변경

##### 현재 – “설명 없음”

##### 개선 – “아래 차트를 플레이하면 키워드 출현증가율을 확인할 수 있습니다.”

# 사업내 과제별 제안규모 산정

## 타이틀과 안내문구 변경

### ‘사업별 과제 규모 산정’ 🡪 ‘사업 과제별 적정 규모 산정’

### 안내문구

#### 기존 – “유사과제 규모 추정 내역을 참조 적정금액 산정란에 값을 입력하면, 조정금액이 집계됩니다.”

#### 변경 – “과제별로 ‘유사과제 집계’를 참조하여 ‘적정금액’ 입력하세요.”

## 구성 개선

### **컬럼 구성**

#### 표를 크게 3영역으로 구분할 필요 – 배경색으로 구분

##### 1영역 – No**.**, 제안과제명, 제안기관, 제안기간, 제안규모 🡪 ‘제안 과제’

##### 2영역 – 적정규모, 규모조정 🡪 ‘적정규모 산정’

##### 3영역 – 최저, 평균, 중위수, 최고, 유사과제 수 🡪 ‘유사과제 집계’

###### 3영역은 Colspan으로 4개 컬럼 위에 ‘유사과제 집계’ 행을 추가

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 유사과제 집계 | | | | |
| 최저 | 평균 | 중위수 | 최고 | 유사과제 수 |

##### 컬럼 너비

###### 숫자 컬럼 – 동일 너비로 고정하고 셀 값을 우측정렬

금액 - 10억 단위

년도 – 열제목 너비

###### 문자 컬럼 – 상대 넓이로 1행이 넘을 때 ‘…’생략처리, 과제명 전체 토글처리

‘과제명’ – 상대 넓이 전체 명칭 토글처리

‘제안기관’ – 고정 넓이 전체 명칭 토글처리

### 행 구성

#### 테이블 하단 footer ‘집계’ 행 개선 – 컬럼 제목(header)제목처럼 배경색과 글씨체 강조 처리

### 값 조작 및 처리

#### 적정금액 입력 산정

##### 탭(Tab), 엔터(Enter)로 값이력과 함께 ‘규모조정’컬럼에 값 처리

#### 적정금액 자동 입력

##### 과제별 유사과제 집계 컬럼을 클릭했을 때, ‘적정금액’에 해당 셀 값 자동 입력

##### 예시: No. 1의 유사과제 ‘평균’ 값을 누르면, 해당 금액이 No.1 ‘적정금액’에 자동 입력

# 데이터 조회, 가공

## 과제 데이터 연차 데이터

### 1년차 데이터만 표출 또는 연차를 함께 표출가능성 확인

#### “과제명 (1년차)” 형식

## 키워드, 과제, 토픽 검색 결과 한정

### 현재 문서 수가 많아서 표현이나 시연에 적절하지 않음

#### (시연) 문서 검색결과 10% 기준을 임시로 300건으로 한정

##### 소시오그램 – 150건 내외 (현재설정 유지)

##### 워드클라우드 – 100건 이하 50~70건 추천

##### 추천단어 - 100여건

## 용어사전

### 불용어 추가

#### 상, 가능해, ~적(선택적, 경제적, 정량적, 화학적, 기술적), ~의(최적의)

### 일반용어 제외(분석시점 또는 시각화 시점에 제외)

#### 공통어(일반용어) - 특성, 구조, 기술, 시스템, 분석, 설계, 기능, 측정, 구현, 처리, 장점, 조성, 조사, 산업, 발전, 방식, 모델

#### 기술용어 – 빅데이터, 사물인터넷, 자율주행 등

### 용어지표를 도입한 불용어 추가(기술용어 강화)

#### 기존 BM25의 한계 – 분류 공통어가 부각되어 나타나는 현상

#### 용어지표 적용대상 – 분류 공통어에 해당하는 용어를 불용어 처리

##### 분류(카테고리)별 적용이 불가하여 용어 전체를 대상으로 불용어 추출

##### Z-Score를 활용한 처리 (처리완료)

#### Edgelist\_term 테이블상 불용어 추가 – 불용어 사전 아닌 ‘분석 제외어’

##### C4.7 수준으로 ‘분석 제외어’ 추가하고, ‘분석 제외어’에서 예외처리한 용어는 분석에서 제외하지 말것

### 대표어 적용 (고려)

#### 같은 의미의 용어를 대표어로 표현

##### 워드클라우드

##### 소시오그램

##### 추천어

#### 예시

##### Tablet, 타블렛, 태블랫, 테블릿

##### 이차전지, 2차전지

#### 동의어 전체를 대표어로 지정

##### 동의어는 서로 의미가 일치하는 용어로 대표어로 먼저 지정

##### 동의어 중 대표어 선택 🡪 현재 동의어 설정한 방식으로 지정

## 엘라스틱 서치 방식 확정

### Simple query? & Query string 방식

#### Query string으로 검색 토큰 범위 한정 – 검색식 포함 토큰 수 3000

### 불용어 처리 방식

#### 동의어 관련어가 포함하는 불용어 처리

##### 토픽모델 검색식 – 동의어 필터 제외, 토픽모델에 동의어 관련어 처리

##### 단어, 문장 검색 – 동의어 필터 적용

## 과제 조회 시간 집계와 튜닝

### 조회 과제 수에 따른 과제목록 표출시간 집계

### 과제 조회 방식 튜닝

## 토픽모델링 대상 변경 시도

### 과제 전체 – 기존 분석 방식

### 검색 대상 과제 1개 - 단장 요구

#### 기존 분석방식과 같으나

#### 토픽베이스 모델을 생성하는 대상 (유사)과제 적정 수 한정 필요

## 토픽모델 강화

### 토픽베이스가 포함하는 ‘워드’를 주제(검색어) 삼아 LDA를 재실행

#### 주제어(검색어)로써 토픽을 추출 🡪 토픽으로 토픽베이스 재구성 🡪 워드 생성

#### 🡺 ‘토픽모델링(수작업)’후에 해당 토픽으로 토픽베이스를 재구성하여 ‘워드’생성에 사용하면 보다 토픽모델을 강화할 수 있음.

## 플레이어 네트워크

### 플레이어 네트워크 JSON 테스트

### 플레이어 네트워크에 문서 리스트 포함 여부 확인

#### 플레이어(수행기관)가 수행한 과제가 플레이어 노드에 ‘문서 리스트’를 포함하여야 함

#### 노드를 선택했을 때 해당 과제 리스트 선택하여 과제 상세를 보여줄 필요

## 시각화 데이터 구성

### 키워드별 과제 수행 집계(건 수)

#### 초기(용어 미선택시) – 과제 전체 수를 1개 곡선으로 년도별로 표시

#### 용어 선택 시 – 해당 용어 관련 과제를 용어 수만큼 곡선으로 년도별 표시

### 논문, 특허, 과제 수행 집계(건 수)

#### 초기(용어 미선택시) – 전체 특허, 논문, 과제 수를 각각 곡선으로 년도별로 제시

#### 용어 선택시 – 해당 용어 관련 특허, 논문, 과제 수를 용어 수만큼 곡선으로 년도별 표시

### 1억당 논문/특허 집계(건 수)

#### 초기(용어 미선택시) – 전체 과제의 1억당 특허, 논문 수를 각각 곡선으로 년도별로 제시

#### 용어 선택시 – 해당 용어 1억당 특허, 논문, 과제 수를 용어 수만큼 곡선으로 년도별 표시

## 소시오그램과 유사과제 목록간 데이터 일관성 확인

### 소시오그램내 과제노드 수와 유사과제 목록 노드 수 일치여부 확인

#### 단순한 숫자(개수)일치

### 네트워크 지표 정렬 확인

#### 영향력, 매개력, 연결성이 적절하게 전달되었는지 확인

#### 정렬도구를 조작하면 적절하게 정렬하는지 확인

## 시각화 데이터 구성 라이브러리

### 워드 히스토리 – 검색결과 나타난 과제 ‘워드’의 히스토리를 표현할 수 있도록 데이터 가공

#### 참조 - <https://observablehq.com/@tezzutezzu/world-history-timeline>

### 워드별 과제 포함구조 – 워드별 과제 포함 여부를 집합관계로 표현

#### 참조 - <https://observablehq.com/@d3/zoomable-circle-packing>

### 과제 계층구조 – 과학기술분류를 이용한 과제의 계층구성

#### 참조 - <https://observablehq.com/@d3/tree-of-life>

# 시각화

## 소시오그램 – 상위 2단(1~2단) 🡪 1~2단을 탭으로 합치거나 2단으로 표현할지 선택

### 플레이어 네트워크 – 과제를 수행한 ‘수행기관’ 네트워크

#### 인터페이스 구성

##### 플레이어 네트워크 소시오그램 - 왼쪽 2/3

##### 플레이어 랭킹 – 오른쪽 1/3

###### 수행과제 수로 랭킹

#### 조절

##### 연도 조절

##### 네트워크 지표 조절

### 과제 네트워크 – 유사과제 네트워크

#### 인터페이스 구성

##### 과제 네트워크 소시오그램 - 왼쪽 2/3

##### 과제 랭킹 – 오늘쪽 1/3

###### 과제 score로 내림차순 리스트 제시.

#### 조절 – 연도만 조절 가능할 수도.

##### 연도 조절

##### 네트워크 지표 조절

## 워드클라우드와 워드별 과제 시계열 흐름 – 1단 (3단)

### 주요키워드 클라우드 – 왼쪽 1/2

#### 인터페이스 구성

##### 워드클라우드로 검색결과 워드를 표현

#### 조절

##### 워드 표시수준 조절 – 프로그레스바 조절

##### 워드 선택/제거 – 오른쪽 그래프에 ‘년도별 키워드별 과제 수’를 표시할 ‘워드’를 선택 제거

###### 프로그레스바로 한 번에 선택(현재 워드 클라우드 선택) – 추가/제거 키워드 칸으로 이동

###### 클라우드 내 ‘워드’를 ‘제거’, ‘선택’ – 추가/제거 키워드 칸으로 이동

### 년도별 키워드별 과제 수 – 오른쪽 1/2

#### 인터페이스 구성

##### 다중곡선형 그래프 – 한 개 이상 워드를 하나의 선으로 연도별 과제수로 표현

##### 범례 – 선택한 ‘워드’를 범례로 표시

#### 조절

##### 범례 선택/제거 – 그래프 내 ‘범례’선택으로 그래프(선) 표현/제거

## 논문과 특허 시계열 흐름 – 1단 (4단)

### 년도별 논문 특허 수, 과제 수 다중곡선그래프– 워드클라우드에서 선택한 키워드 관련 과제의 특허와 논문 수를 집계하여 년도별로 시계열로 표현 (과제 건당 특허와 논문이 아닌 과제 전체 특허와 논문 집계를 의미)

#### 인터페이스 구성

##### 논문/특허/과제 수 다중곡선형 그래프 - 왼쪽 1/2

##### 범례 – 논문, 특허, 과제 세 가지중 선택하여 표현

### 1억당 논문 수와 특허 수를 다중곡선그래프 – 워드클라우드에서 선택한 키워드 관련 과제 1억당 논문 수와 특허 수를 집계하여 시계열로 표현 (과제 전체 집계에서 1억당 집계를 그래프로 제시)

#### 인터페이스 구성

##### 논문 수와 특허 수를 다중곡선형 그래프 - 오른쪽 1/2

## 키워드 추이 – 1단(5단)

### 주요키워드 추이 – 왼쪽 1/2

#### 주요 키워드 20개를 연도별로 버블차트로 표현

### 분야별 과제 추이 – 오른쪽 1/2

#### 과제 전체를 ‘분야별’연도별 과제 추이 적층형바차트로 표현