

제4기 지오빅데이터 오픈플랫폼 모니터링단 최종 보고서

지오빅데이터 오픈플랫폼: 지질자원의 새로운 패러다임

김동희(인하대학교)

들어가는 말

세상은 AI를 바탕으로 매우 빠른 속도로 변화하고 있다. 지질 및 자원 분야에서 인공지능은 유전 탐사 및 개발의 많은 단계에서 사용되고 있다.¹⁾ 인공지능을 이용하기 위해서 데이터의 양과 질이 중요하다.²⁾

지오빅데이터 오픈플랫폼이란 한국지질자원연구원이 개발한 대한민국의 대표적인 지질정보 플랫폼으로, 방대한 양의 지질 데이터를 국민 누구나 쉽게 활용할 수 있도록 신뢰성 높은 지질자원 정보를 제공하는 대한민국의 데이터 허브이다.

한국지질자원연구원은 1918년부터 대한민국의 지질과 자원관련 연구를 진행해 현재 국토우주지질연구, 광물자원연구, 해저지질에너지연구, 지진재해연구, 자원활용연구, 기후변화대응연구를 진행한다. 한국지질자원연구원의 2024년 현재까지의 논문은 5065건이며 과제는 1339건으로 활발한 연구가 진행됨을 알 수 있다. 한국지질자원연구원이 개발한 지오빅데이터 오픈플랫폼은 한국데이터산업진흥원의 데이터 품질인증(DQC-V) 심사에서 최고 등급인 ‘플래티넘 클래스’를 획득해 데이터의 품질이 훌륭함을 증명했다.³⁾

I. 오픈플랫폼의 검색 활성화 방안 제안

1. 지오빅데이터 오픈플랫폼의 검색 기능 개선 사항 및 활성화 방안 제시

(1) 검색 중복 기능 선택적 사용

지오빅데이터 오픈플랫폼은 데이터의 양과 질이 좋다는 장점 이외에도 많은 장점이 있지만, 그 중 국민 누구나 이용할 수 있다는 정보의 공유성을 뽑고 싶다. 많은 정보 속 본인이 원하는 정보를 이용하기 위해서는 검색이 필수이다. 하지만, 오픈플랫폼의 검색 기능은 처음 접해보는 사람이 겪기에는 여러 불편함이 있다. 아래 그림은 검색어를 캐나다로만 적용한 결과이다. 이렇게 캐나다만 검색한 결과 38건의 결과가 나온다. 하지만, 이전에 내가 다른 자료를 찾고 캐나다가 검색한 결과 그림 2를 보면, 0건임을 확인할 수 있다. 사람들이 많이 사용하는 플랫폼인 google이나 naver에서는 검색 중복 기능을 기본적으로 지원하지 않기 때문에 처음 오픈플랫폼을 접하는 사람들에게는 혼란스러움을 줄 수 있다. 따라서, 검색어의 중복 기능을 기본적으로 제거하는 것을 개선 사항으로 제안한다.

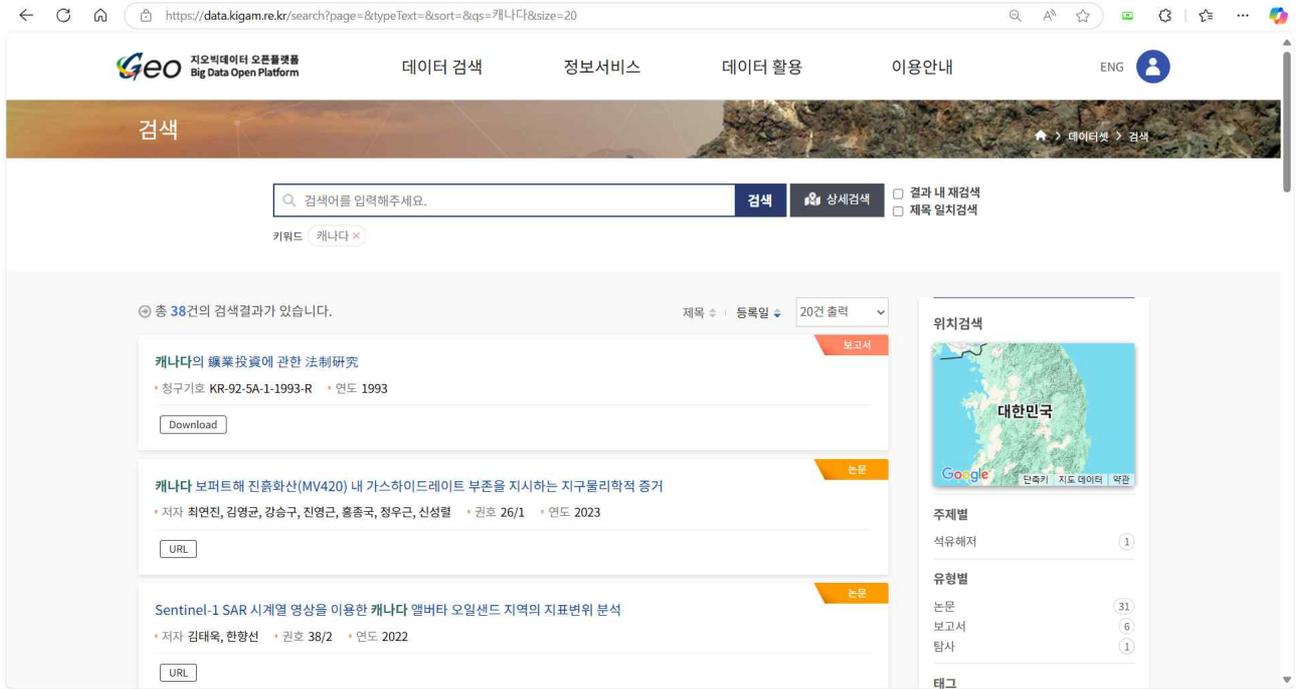


그림 1 : 통합검색에서 캐나다를 검색한 결과

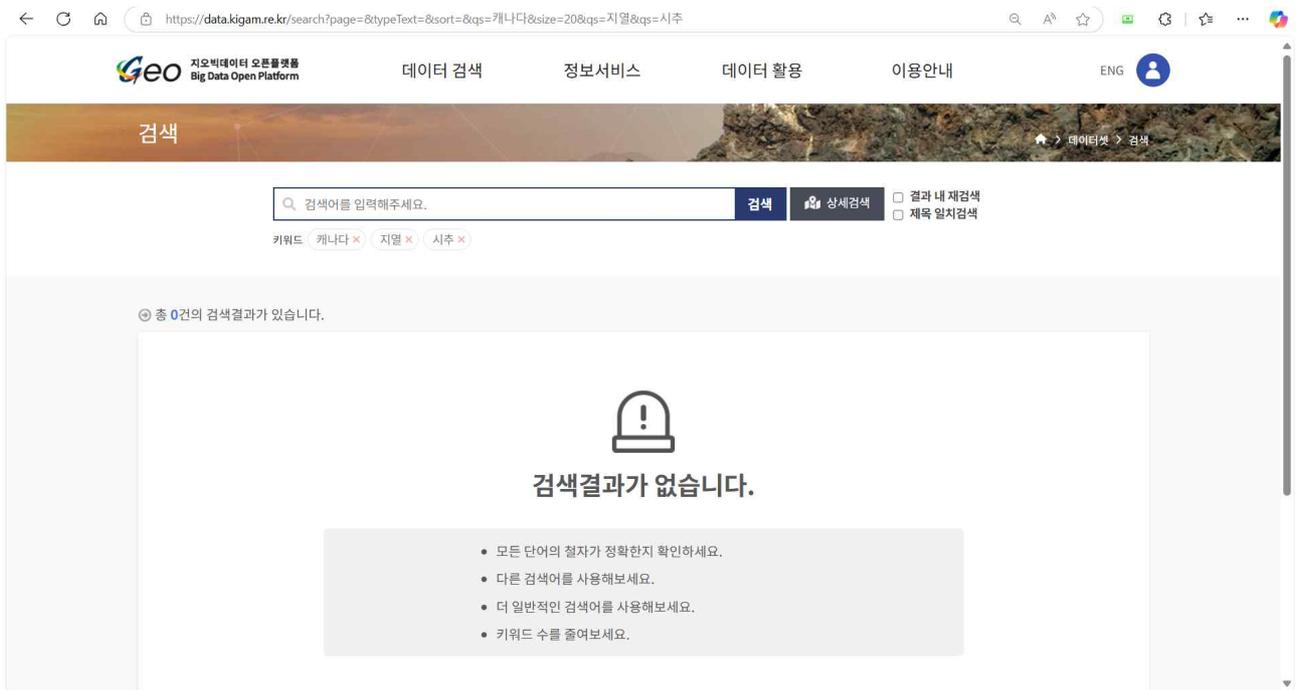


그림 2 : 검색어 순서를 시추, 지열, 캐나다 순서로 검색한 결과

(2) 등록일 기준 검색이 아닌 자료의 생산일 기준 정렬 기본적인 기능으로 추가 제안

사용자는 자료의 등록일 기준이 아닌 자료가 생산된 날짜가 중요하다고 생각한다. 예를 들어, 내가 2024년의 전국의 지열 자료가 필요해서 찾고 있다면 2024년에 생산된 자료 중 검색으로 찾는 게 편하다. 하지만, 오픈플랫폼에서의 검색 기능은 제목과 등록일이 기본적인 기능으로 존재한다. 등록일을 기준으로 내림차순 정렬을 해보면 그림 3을 통해 제일 처음으로 지열 자원 조사 연구가 나온다. 지열 자원 조사 연구를 들어가면 그림 4를 통해 등록일이 2099년 3월 30일임을 확인할 수 있다. 등록일이 정확하지 않아 등록일 기준의 자료 검색도 의미가 없는 것을 볼 수 있어 등록일이 아닌 자료의 생산일 기준 정렬을 기본적으로 검색 기능에 추가함을 제안한다.

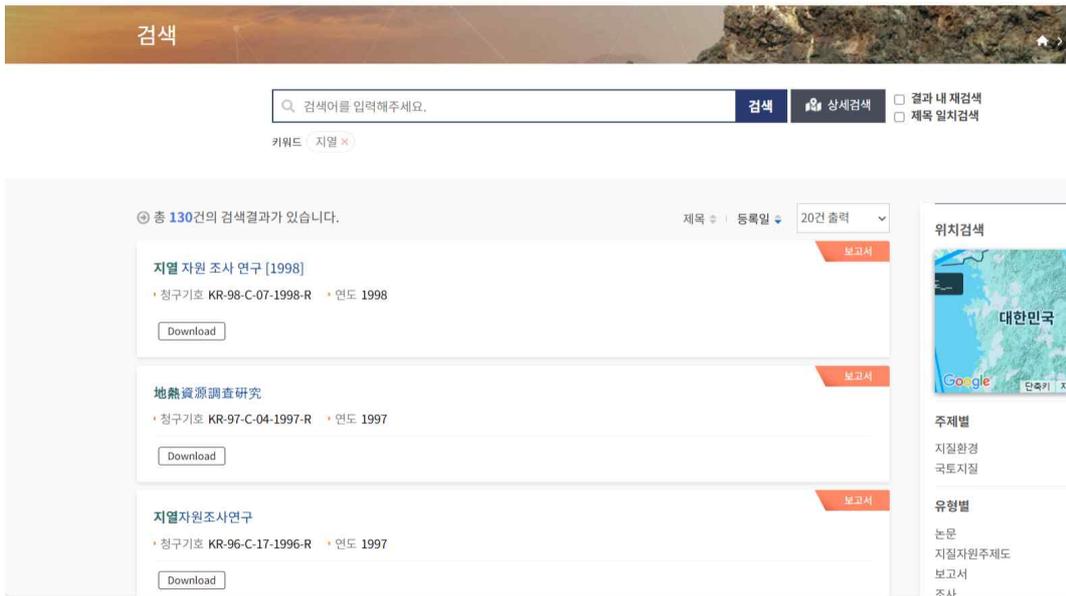


그림 3 : 지열 검색 후 등록일 내림차순 정렬한 결과

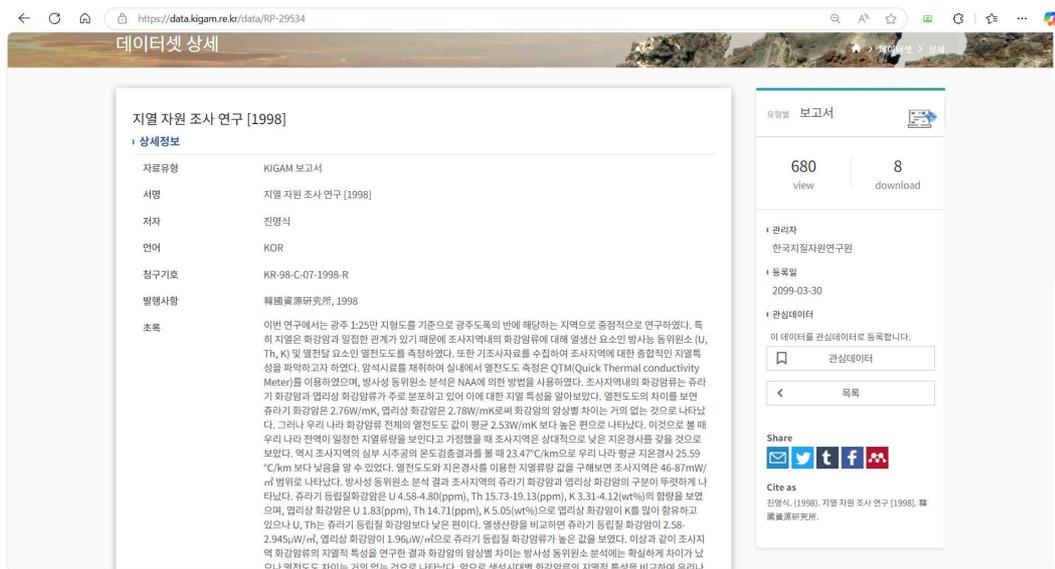


그림 4 : 지열 자원 조사 연구 데이터셋 상세 페이지

ii. 오픈플랫폼의 홍보 활성화 방안 제안

1. 지오빅데이터 오픈플랫폼의 홍보 방안 제안

(1) 학생 서포터즈 또는 홍보팀 신설

지질 또는 자원공학을 전공하는 학생으로서 전공 분야에 관심을 가지고 여러 대외활동을 찾는다. 하지만, 국내에서 해당 전공을 활용하는 대외활동은 타 분야와 비교했을 때 수가 많지 않다. 전공 대학생을 활용한 학생 서포터즈나 홍보팀을 신설해 오픈플랫폼을 홍보하는 내용물을 제작한다. 전공 대학생들을 모집해 홍보팀을 만든다면, 홍보물을 제작하면서 관련 전공 지식 함양 및 전공자들끼리의 의사소통 능력 향상뿐만 아니라 지오빅데이터 오픈플랫폼의 홍보 효과 또한 발달할 것으로 판단한다.

iii. 오픈플랫폼의 데이터 오류 검출

1. 지오빅데이터 오픈플랫폼의 데이터 오류 검출

(1) 충청도의 지열 데이터 오류 검출

충청도의 지열 데이터의 데이터셋 상세의 파일 데이터를 확인한 결과, 충청도의 자료가 아닌 제주도의 자료임을 확인했다. 아래 그림 5의 경우 엑셀 파일은 제주도의 주소를 나타내고 있으며, 워드 파일의 제목이 Geothermal Data of jeju island in South Korea로 제주도의 지열 자료를 나타낸다. 제주도의 지열 자료를 표현하고 싶다면 충청도의 지열 데이터가 아닌 제주도의 지열 데이터로 표시하거나 충청도의 지열 데이터를 표현하고 싶은 경우, 자료를 충청도의 지열 데이터로 재업로드가 필요하다.

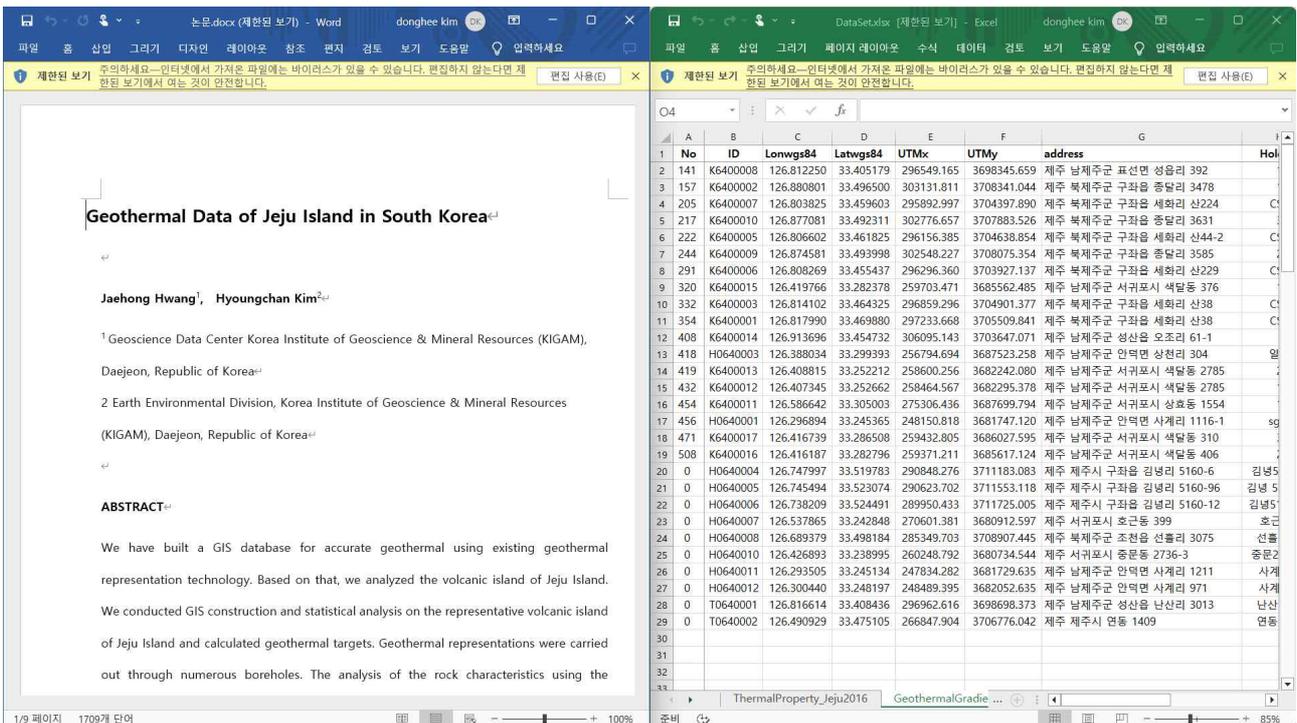


그림 5 : 왼쪽 - 제주도의 지열 데이터 논문, 오른쪽 - 제주도의 지열 데이터 자료

(2) 논문 데이터 오류 검출

총 2건의 논문 데이터와 관련된 오류를 검출했다. 논문의 제목은 1) 물리탐사 분야에서의 딥러닝

기술: 현황, 도전 과제 및 미래 방향 2) 현풍(북동일대)지열의 지하수 분포 이다. 각 논문들은 파일이 존재하지 않는다는 안내가 나온다. 해당 문제는 해당 자료에 맞는 논문을 업로드해 해결할 수 있다.

iv. 오픈플랫폼의 홈페이지 오류 검출

1. 지오빅데이터 오픈플랫폼의 홈페이지 오류 검출

(1) 오픈플랫폼의 홈페이지 오류와 보안

지오빅데이터 오픈플랫폼을 활용하면서 느낀 불편한 점 중 하나는 페이지 문제이다. 1) 페이지의 정보 누출이 정상적으로 진행되지 않은 경우 2) 페이지의 보안이 확실하지 않은 경우 등이 있는데, 이 두 문제가 매우 지속적으로 나타난다.

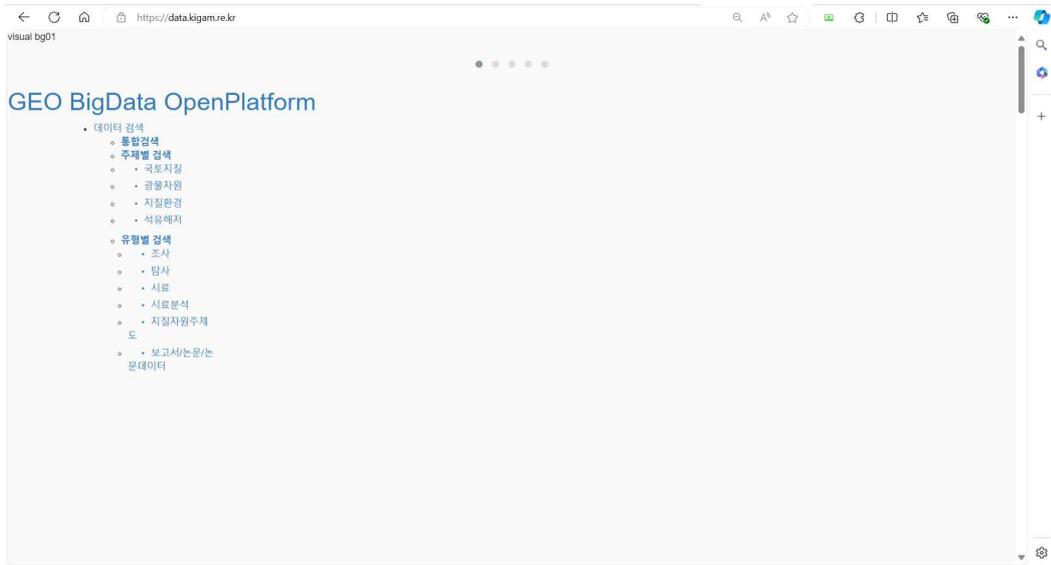


그림 6 : 페이지의 정보 누출이 정상적으로 진행되지 않은 경우

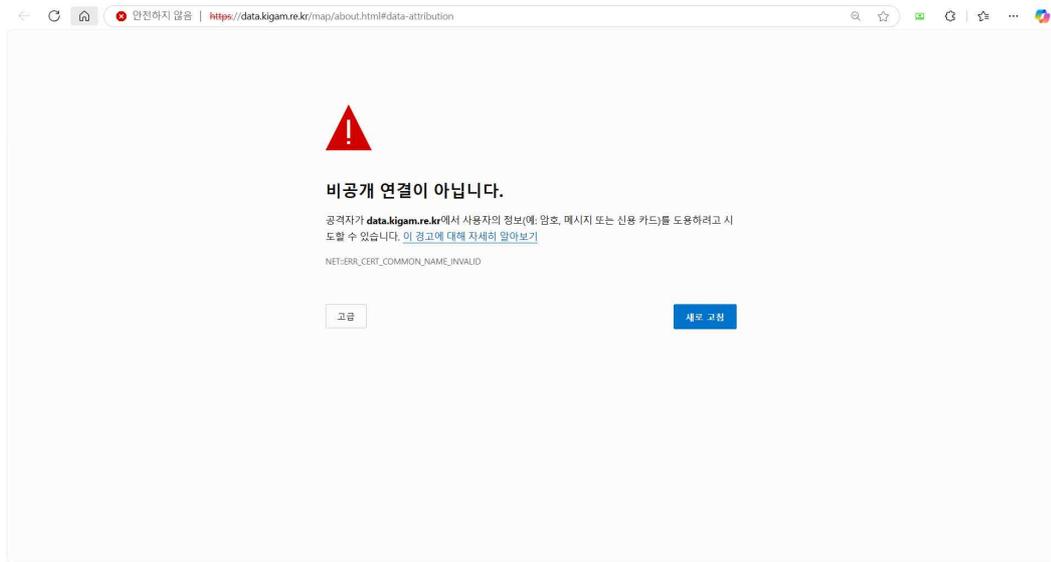


그림 7 : 페이지의 보안 문제가 확실하지 않은 경우

v. 오픈플랫폼의 인공지능 역량 활성화 방안 제안

1. 지오빅데이터 오픈플랫폼을 통한 인공지능 역량 활성화 방안 제안

(1) 인공지능 활성화 방안으로 엑셀 또는 이미지 자료 제안

많은 사람들이 코딩에 사용하는 언어인 python에서는 인공지능을 활용하기 위한 많은 라이브러리가 있다. 대표적으로 scikit-learn과 tensorflow가 있으며, 이외에도 많은 라이브러리가 존재해 코딩을 통해 인공지능을 활용한다. 인공지능을 활용하기 위해서는 데이터가 필수적으로 필요한데, 데이터는 엑셀이나 이미지 데이터를 통해서 얻는 것이 편리하다. 오픈플랫폼에서는 주로 논문이나 보고서 형태의 자료가 많고 엑셀이나 이미지로 되어있는 자료는 흔하지 않아 인공지능과 연계하기에는 힘들다고 생각한다. 인공지능 역량을 강화하는데 편리한 엑셀 자료나 이미지 자료를 지오빅데이터 오픈플랫폼에서 공개 자료로 공개한다면, 많은 전공자들이 자료를 활용해 인공지능 역량을 발전시킬 수 있을 것이다.

나오는 말

총 5개의 내용으로 ‘I. 오픈플랫폼의 검색 활성화 방안 제안, ii. 오픈플랫폼의 홍보 활성화 방안 제안, iii. 오픈플랫폼의 데이터 오류 검출, iv. 오픈플랫폼의 홈페이지 오류 검출, v. 오픈플랫폼의 인공지능 역량 활성화 방안 제안’ 하고 앞으로의 발전 방향에 대하여 확인했다.

앞에서 설명한 5가지의 내용은 모두 지오빅데이터 오픈플랫폼이 앞으로 지질자원 분야의 새로운 패러다임으로 발전하기 위해 개선되어야하거나 새로운 제안들이다. 기존의 데이터의 체계적인 관리를 바탕으로 데이터의 높은 신뢰도를 이끌어내고 이후 오픈플랫폼의 장점인 데이터의 확장성을 이용하면 새로운 인공지능 시대에서 지질자원 오픈플랫폼으로써 혁신이 될 수 있음을 굳게 믿는다.

<작성요령>

- 본인이 영향력 및 파급력이 높다고 생각하는 제안사항 및 아이디어 5개를 선별하여 작성하며, 오픈플랫폼 활용 사례 및 활성화 제안 분류의 아이디어 1개는 필수로 포함되어야 함
- 기존 제출한 월별 보고서를 활용하여 작성 가능하며, 새로운 내용으로 작성해도 무방함
- 서식 변경 가능, 서술식으로 작성 (분량 3장 이상)
- 글꼴준수, 용지여백은 위·아래 15, 머리·꼬리말 15, 좌·우 20
- 파일명은 이름_지오빅데이터 오픈플랫폼 모니터링단 최종 보고서.hwp로 저장하여 활동게시판에 제출
- 아래의 예시를 참고하여 작성

<참고문헌>

- 1) Di, S., Cheng, S., Cao, N., Gao, C., & Miao, L. (2021). AI-based geo-engineering integration in unconventional oil and gas. *Journal of King Saud University-Science*, 33(6), 101542.
- 2) 인공지능 전문 매체, [칼럼] 데이터 중심 AI와 소프트웨어 2.0, 참조.
(2024년 5월 28일 검색)
(<https://www.newstheai.com/news/articleView.html?idxno=5546>)
- 3) 한국지질자원연구원 보도자료
(2022년 11월 9일 검색)
(https://www.kigam.re.kr/board.es?mid=a10703040000&bid=0025&act=view&list_no=52957))