

제4기 지오빅데이터 오픈플랫폼 모니터링단 최종 보고서

기능 개선 제안 및 오픈플랫폼 활성화 제안

변건우(경북대학교)

< 목 차 >

ㄱ. 머리말	2
I. 기능 개선 제안	2
1. 실시간 인기 검색어	2
(1) 필요 이유	2
(2) 구축 방안	3
2. 토론창 기능 활성화	3
(1) 필요 이유	3
(2) 구축 방안	4
3. 연관된 보고서 더보기 기능	4
(1) 필요 이유	4
(2) 구축 방안	5
4. 좋아요 기능	7
(1) 필요 이유	7
(2) 구축 방안	7
II. 오픈플랫폼 활성화 제안	8
1. 바뀌야 될 정보 찾기 포상 제도	8
(1) 필요 이유	8
(2) 오픈플랫폼 활성화 방안	9
ㄴ. 맺는말	10
<참고문헌>	10

7. 머리말

“정보의 바다가 홍수처럼 밀려오고 있다.” 우리가 사는 21세기를 대표할 수 있는 문장이다. 우리가 매일 사용하는 IT 전자기기에 담겨진 제타바이트급의 데이터는 현대 문명이 급속도로 발전할 수 있었던 원동력이다. 결국 현대 사회에서는 이 데이터를 다룰 수 있는 힘이 곧 국가의 발전에 기여할 수 있는 큰 요소이다.

우리는 많은 데이터의 집합체를 빅데이터(Big Data)라고 부른다. 빅데이터에서 빅(Big)은 단순히 데이터의 양이 많아서 붙여진 것만은 아니다. 수 없이 많은 데이터의 다양성과 복잡성, 그리고 데이터 간의 연관성으로 인해 붙여진 것이다. 데이터 간에는 복잡한 관계를 가지며 유기적으로 얽혀있다. 우리는 데이터의 진정한 가치를 뽑아내기 위해서 데이터를 체계적이고 단순하게 정리할 필요가 있다.

수 십년간 쌓여온 지질학적 데이터베이스를 체계적으로 정리해 둔 ‘지오빅데이터 오픈플랫폼’도 이러한 시대적 필요에 따른 역할을 수행하는 것이다. 이미 오픈플랫폼의 관계자와, 모니터링단 1기부터 3기까지의 단원들로 인해 많은 개선을 이뤄왔다. 하지만 데이터의 누적속도가 훨씬 빠르기 때문에 이러한 오픈플랫폼 모니터링의 필요가 여전히 재고되고 있을 것이다. 필자 역시 **데이터의 활용도를 높이고 접근성을 향상시키기 위한 방안**에 대해 고민하였다.

필자는 4가지의 기능 개선 제안과 1가지의 오픈플랫폼 활성화 제안에 대해 말하고자 한다. 기능 개선 제안사항으로는 현재 가장 관심을 받고있는 검색어를 알려주는 ‘**실시간 인기 검색어 창**’ 표시, 사용자 간의 질의 응답이 가능한 ‘**토론창 기능 활성화**’, 데이터 간의 접근성을 향상시킬 수 있는 ‘**연관된 보고서 더보기 기능**’, 사용자로 하여금 데이터의 효용성을 평가받는 ‘**좋아요 기능**’이 있으며, 오픈플랫폼 활성화 제안 사항으로는 ‘**바뀌야 될 정보 찾기 포상 제도**’를 제안한다.

I. 기능 개선 제안

1. 실시간 인기 검색어

(1) 필요 이유

현재 지오빅데이터 오픈플랫폼에는 사용자가 실시간으로 인기 검색어를 볼 수 있는 창(窓)이 없다. 만약 이러한 창이 생긴다면 얻을 수 있는 장점은 다음과 같다.

첫째, **연구 트렌드를 파악하는데 용이하다.** 인기 검색어가 표시된다면 현재 지질학과 관련된 연구자나 기업 관계자들이 어떤 주제, 분야에 관심을 가지는지 실시간으로 파악할 수 있다. 실제로 유명 포털사이트들이 실시간 인기 검색어 기능을 지원하는 것 역시 비슷한 의미일 것이라 생각된다. 따라서 사용자들에게 관련 분야의 시장 수요, 연구 개발 등 도움을 줄 수 있는 인기 검색어 키워드 창이 필요하다고 생각한다.

둘째, **데이터의 접근성 증대에 효과를 줄 수 있다.** 현재 가장 각광받고 있는 분야에 대한 정보를 손쉽게 파악할 수 있으므로 사용자가 빠르게 접근할 수 있다는 장점이 있다.

셋째, **사용자 참여를 유도할 수 있다.** 기존 방식대로 필요한 정보를 검색하여 그에 대한 정보를 얻는 것 뿐만 아니라, 현재 가장 각광받고 있는 키워드를 보며 관심을 가질 수 있으며 플랫폼에 체류 시간이 증대되어 오픈플랫폼의 활용도가 높아질 수 있다.

추가로, 한국언론진흥재단이 2018년에 실시한 설문조사에서 응답자의 69.5%가 포털을 이용 시 ‘실시간 검색어’를 확인한다고 밝혔다. 이상우 연세대 교수가 2019년에 실시한 설문조사에 따르면 실시간 검색어 서비스에 대해 평균 3.08점(5점 만점)으로 만족한다는 응답이 많았다. 응답자들은 긍정적인 면으로 ‘적시성’(최신 정보 제공), ‘유용성’(정보를 빠르고 쉽게 탐색), ‘즐거움’(검색 과정의 즐거움), ‘신뢰성’(믿을 만한 정보의 제공) 등을 꼽았다. 이러한 점으로 미루어볼 때, 실시간 인기 검색어 도입 필요의 이유는 충분하다.

(2) 구축 방안

먼저 실시간 인기 검색어 키워드창을 구축하기 위해선 사용자의 검색 로그를 분석하는 시스템이 구축되어야 한다. 일정 시간(24시간 또는 일주일 등) 동안의 검색 빈도가 높은 키워드를 기반으로 하여 실시간 인기 검색어 순위를 업데이트 해야 한다. 또한 오픈플랫폼 홈페이지와 잘 어울리는 UI, UX를 고려하여 사용자가 편리하게 접근할 수 있도록 실시간 인기 검색어 창을 배치해야 한다. ([그림 1] 참고) 또한 실시간 인기 키워드가 업데이트 되었을 때, 알림 구독을 설정한 사용자에게 업데이트 알림 서비스를 제공한다.

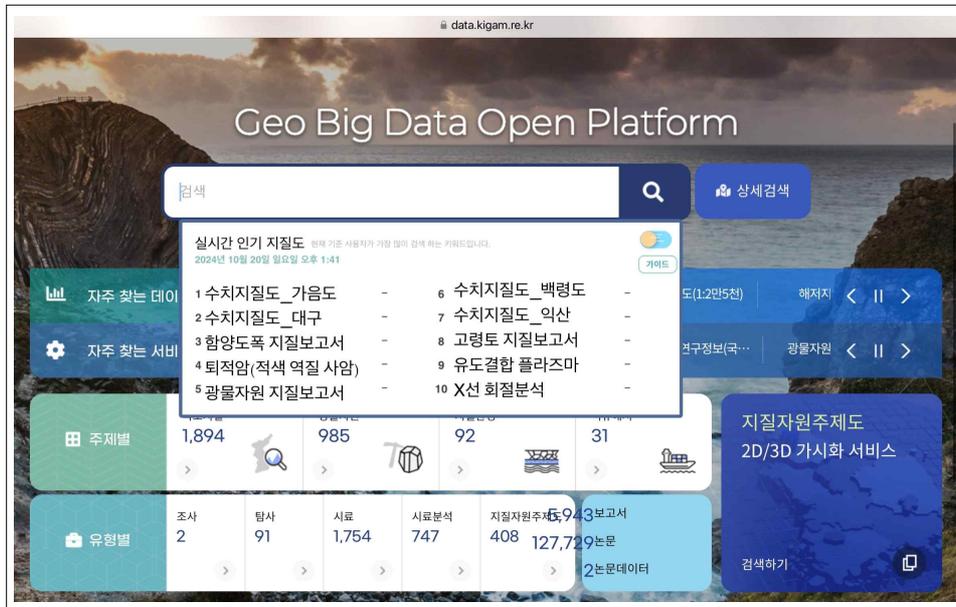


그림 1 실시간 인기 검색어 창을 구현한 모습.

2. 토론창 기능 활성화

(1) 필요 이유

현재 오픈플랫폼 내에는 사용자 간의 자유로운 소통과 토론의 공간이 마련되어있지 않다. 필자는 대구 연안산의 필드 조사를 하며 해당 지역의 지질도를 찾아보는 과정에서 이해가 되지 않는 부분이 많았고, 해당 지질 구조에 대해 나와 비슷한 사용자들과 함께 토론하고 싶다는 생각이 자주 들었다. 이러한 이유로 오픈플랫폼 내에 ‘토론창’ 기능이 생긴다면 사용자 간 지식 공유가 가능할 것이며 커뮤니티로서 기능을 수행할 수 있을 것이라고 생각했다.

(2) 구축 방안

‘토론창’이라는 이름의 창을 오픈플랫폼 내에 추가하여, **사용자 간의 대화가 가능하도록** 형식을 설정한다. 부적절한 언행이나 오픈플랫폼 취지에 부합하지 않는 댓글을 방지하기 위해 주기적인 모니터링을 하고, 신고 기능을 삽입한다. 또한 자주 올라오는 질문이나 유용한 논의에 대해서는 추가적인 게시판을 제작하여 올려두면 사용자들에게 더욱 양질의 데이터를 제공할 수 있을 것이다.([그림 2]참고)

마지막으로, 토론의 활성화를 위해 일정 기간(일주일, 한달 등)마다 **오픈플랫폼이 자체적으로 화두가 될만한 주제를 선정하여 토론창에 토론 주제를 제시**한다. 이러한 주제에 대해 사용자들이 토론을 할 수 있으며 우수한 의견을 제시한 경우 경품을 제공하거나, 수상을 한다.

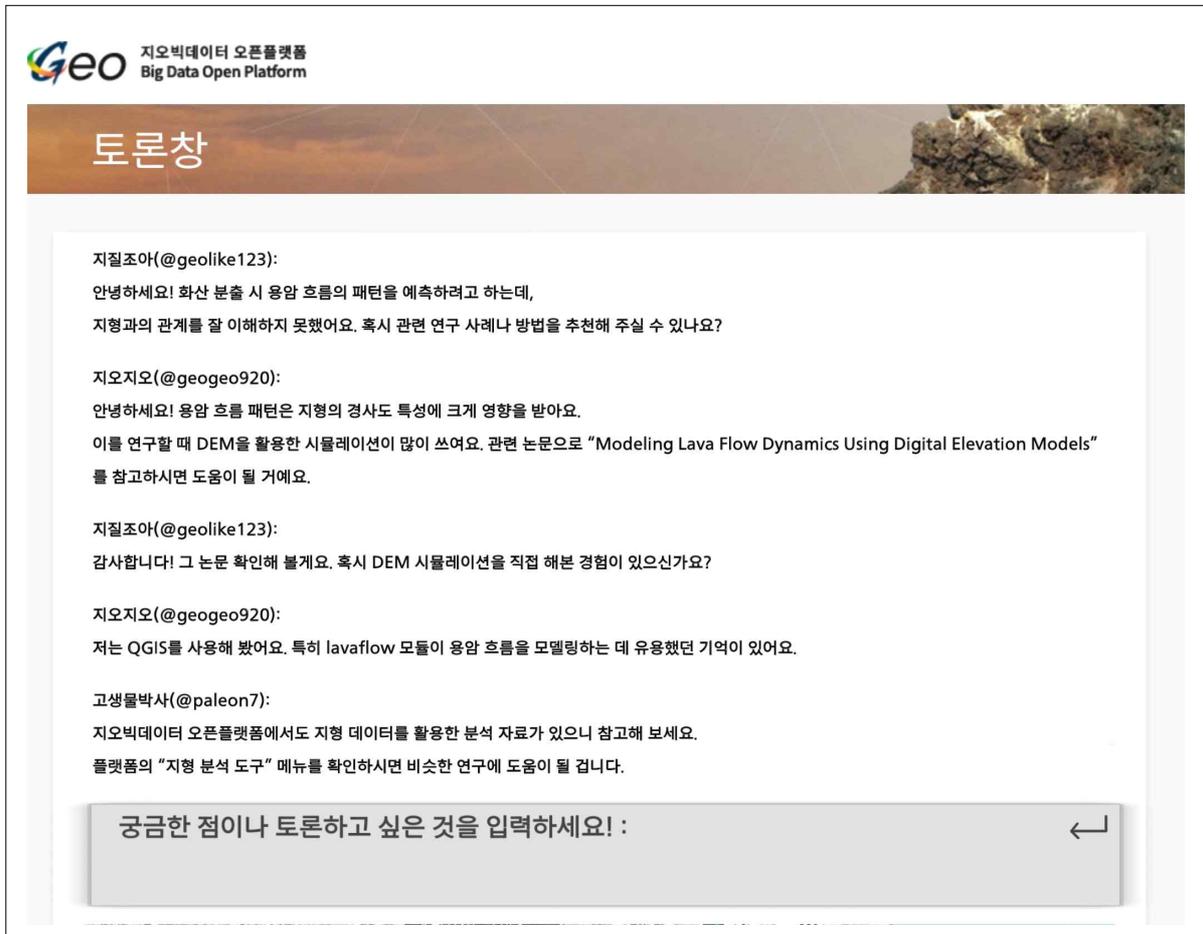


그림 2 토론창을 구현한 모습.

3. 연관된 보고서 더보기 기능

(1) 필요 이유

사용자가 모종의 이유로 자료를 검색하여 열람하는 경우에, 단 하나의 자료만 살펴보는 경우는 흔치 않다. 대부분의 사용자는 찾고자 하는 키워드를 포함하는 자료에 대해 수 - 수 십개의 자료를 열람하며 정보를 발췌한다. 하지만 현재 **오픈플랫폼은 자료를 검색하여 하나하나 눌러봐야 하는 상황**이므로 불편함이 있을 것이라 생각한다. 따라서 사용자가 자료를 열람 중일 때 화면의 우측 배너 (banner)에 **“연관된 보고서 더보기” 창을 삽입**하여, 관련있는 자료를 빠르게 더 볼 수 있도록 한다.

이러한 기능을 통해 **사용자는 연관된 자료를 자동으로 추천받게 되어 탐색 시간을 절약할 수 있으므로 연구 효율성이 증대된다.** 또한 플랫폼 내의 방대한 양의 데이터를 효율적으로 사용자에게 제시함으로써 데이터의 활용도를 올리고, 자료 간의 관계를 쉽게 이해할 수 있다. ([그림 3] 참고)



그림 3 연관있는 보고서 더보기 기능을 구현한 모습(우측).

(2) 구축 방안

‘연관된 보고서 더보기’ 기능을 구축하기 위한 가장 기본적인 방법은 **‘키워드’의 활용**이다. 대부분의 보고서에는 그 보고서를 나타내는 키워드가 태그(tag)되어있다. 이러한 키워드가 현재 사용자가 열람 중인 보고서와 2개 이상 중첩되는 자료를 사용자에게 추천하는 것이 가장 단순하고 효율적인 방법이라고 생각한다.

현재 오픈플랫폼 내의 다수의 자료들에 태그되어있는 키워드가 통일되지 않게 적혀있는 것이 다수 식별되었다. 예를 들어 똑같은 ‘망간’이란 키워드가 어떤 자료에서는 ‘대한민국 망간단괴 광구’라고 적혀있고, 어떤 자료에서는 영문인 ‘Manganese rock’으로 적혀있다. ([그림 4] 참고) 이러한 **키워드의 작성 방법에 대한 규정이 통일되지 않아 발생한 문제들**로 인해 추천 알고리즘에 방해가 될 수 있다. 따라서 **오픈플랫폼은 데이터 등록 시에 키워드 작성법을 명확히 규정하여 키워드의 통일성을 확보**해야 한다. (키워드 최소 등록 개수 증대, 한글과 영문 모두 표기 등)

또한 AI 기반 알고리즘을 통해, ‘망간’이라는 단어와 유사하거나 같은 의미를 지닌 단어들을 색출하여([그림 5] 참고) 이러한 키워드를 **‘태그 추천 알고리즘’**을 통하여 추출한 뒤 데이터베이스(DB)화하여 매칭 시스템을 구축시킨다. 이때 태그 추천 알고리즘 중 **자연어 처리 기반(Natural Language Processing) 알고리즘**은 문장의 문맥을 이해하여 중요 키워드를 추출하는 방식이고, 태그들의 동시 출현 빈도만을 사용하여 적은 태그 데이터의 저장만으로 **빠른 처리와 더 정확히 핵심 키워드를 추출하는 것이 유리**하므로 자연어 처리 기반 알고리즘 방식을 선택한다.

이현복, 김원년, 고영탁, 김종욱, 지상범, 박정기	Emirates
2012	B0412003010
45/5	Edwin Gnos, Tjerk Peters
한국해양과학기술원	1995
477 ~ 486	120/03
대한민국 망간단괴 광구, 정밀수심, 금속함량	372 ~ 377
원문링크	Tephroite,hausmannite,galaxite,granulite-facies,manganese rock,United Arab Emirates,metamorphic rock
	원문링크

그림 4 키워드가 통일되지 않고 쓰여진 모습.

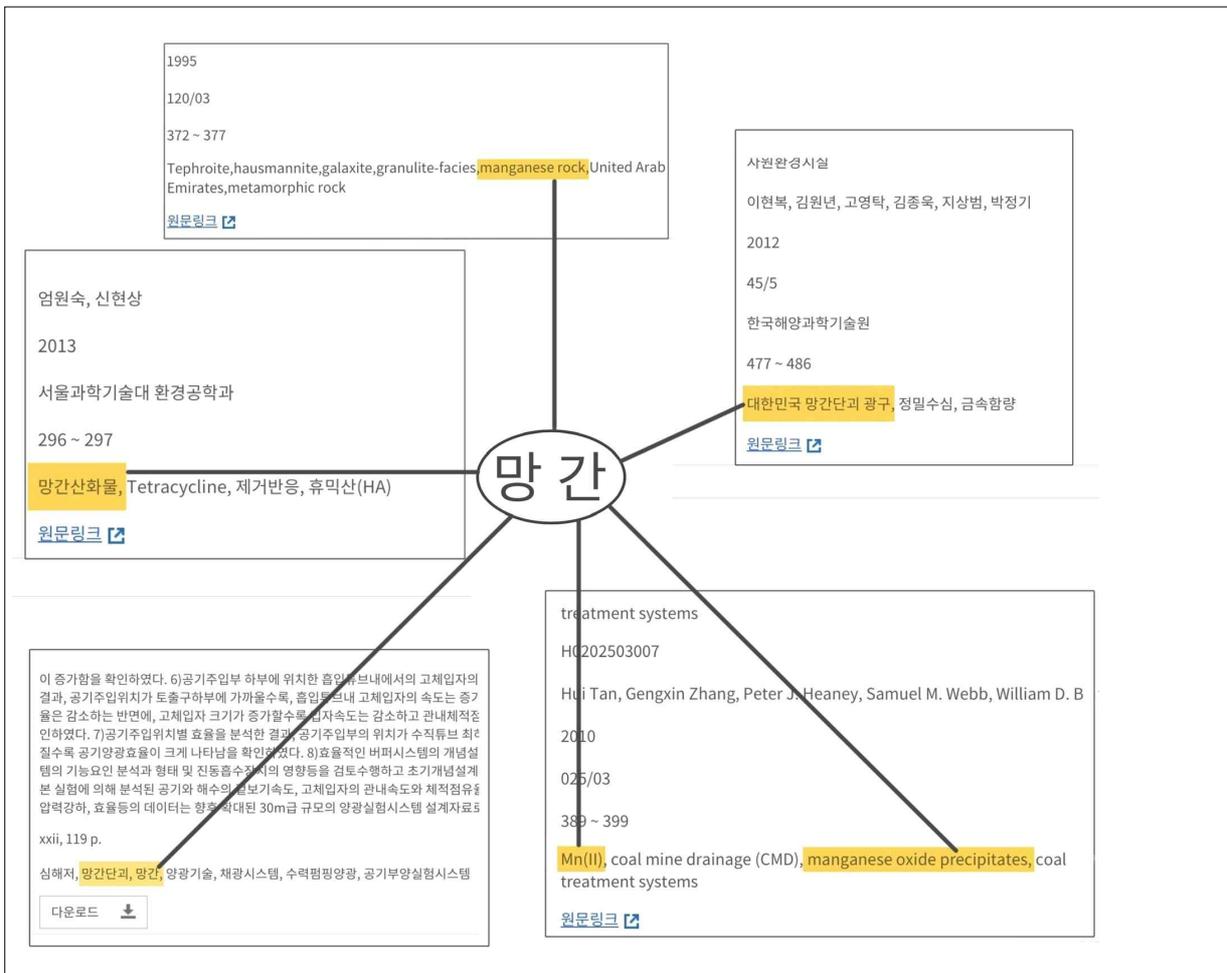


그림 5 '망간' 키워드와 유사한 단어를 색출하여 연관된 보고서를 찾는 메커니즘의 모습.

4. 좋아요 기능

(1) 필요 이유

Instagram, Facebook 등 다양한 SNS에서는 좋아요 기능을 제공하고 있다. 좋아요 기능은 그 게시물이 얼마나 **유용한 정보인지 또는 참신한 정보인지를 나타내는 하나의 지표**로서 생각해볼 수 있다. 이러한 ‘좋아요’ 기능을 오픈플랫폼에도 적용을 해보았다. 모든 데이터가 가치가 있고 유용하지만, 그 데이터가 사용자들에게 실질적으로 주는 영향력이 모두 동일하진 않다. 이러한 **데이터의 영향력은 그 데이터를 사용하는 사용자로 하여금 정해진다**. 사용자가 해당 지질 보고서, 논문의 유용함에 대하여 ‘좋아요’ 버튼을 눌러 평가를 매기고 이러한 지표를 통해 데이터의 중요도를 일부 판단할 수 있다면 훨씬 실질적인 도움이 되는 데이터를 빠르게 찾아볼 수 있다.

(2) 구축 방안

현재 오픈플랫폼 보고서의 우측 배너(banner)에는 view, download, citation의 수가 표시되어 있다. 필자는 이 섹션 바로 아래에 ‘좋아요’ 버튼의 UI 섹션을 구현하고, 그 수를 표시한다. ([그림 6] 참고) 이러한 지표를 활용하여 **‘좋아요’ 많은 순으로 데이터를 볼 수 있는 정렬법**([그림 7] 참고)을 추가로 구현한다. 따라서 **사용자들은 사용자로 하여금 매겨진 더욱 신뢰성있고 양질의 정보로 평가받은 데이터를 우선적으로 볼 수 있는 것이다**.

추가로, 한 명의 사용자가 ‘좋아요’ 버튼을 중복으로 클릭하는 것을 방지하기 위해, 로그인을 한 상태에서만 좋아요 버튼을 누를 수 있도록하며 하나의 데이터에는 한 번만 클릭할 수 있도록 설정한다.



그림 6 좋아요 기능을 구현한 모습.

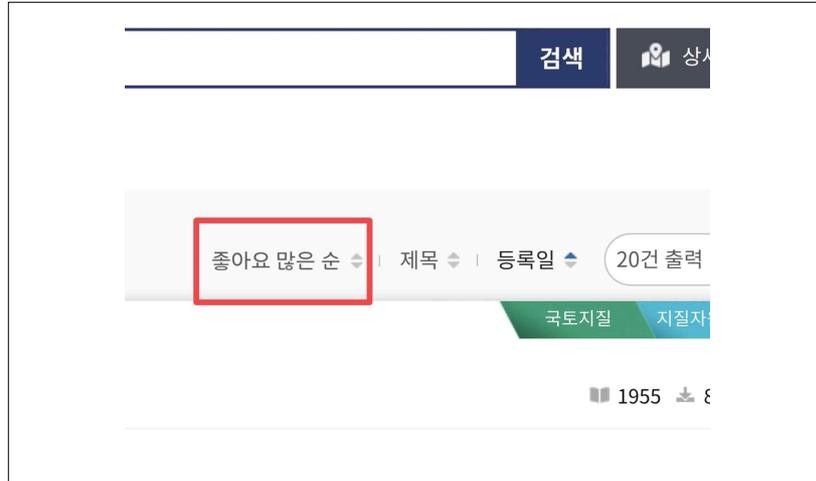


그림 7 좋아요 많은 순으로 정렬법을 추가한 모습

II. 오픈플랫폼 활성화 제안

1. 바뀌야 될 정보 찾기 포상 제도

(1) 필요 이유

올해 초 교내 ‘지질도학 및 야외지질학’ 과목을 수강하던 중 대구광역시 연암공원(연암산)의 지질 분석을 하는 팀 프로젝트를 진행하였다. 직접 필드 조사를 하며 지오빅데이터의 [수치지질도_5만축적_대구]를 비롯한 다양한 자료를 많이 활용하였다. 큰 도움이 되었지만 약간의 아쉬운 점도 있었다. 여러 데이터를 살펴보던 중 지질계통표의 세부적인 명칭이 자료, 데이터마다 각기 달라서 정확히 어떤 지층을 지시하는지 파악하는데 어려움을 겪었기 때문이다.

예를들어 Tateiwa(1929)가 정의한 경상누층군(Gyeong-Sang System)의 신라층군에 속하는 대구층(Daegu Fm.)과 Chang Ki Hong(1975)이 정의한 경상누층군(Gyeong-Sang Super Group)의 하양층군 중 함안층(Haman Fm.) 실제론 동일한 Formation을 지시하고 있었지만, 전혀 다른 이름으로 적혀있어서 이 두 명칭이 같은 지층을 지시하는지 알 수 없었다.([표 1] 참고)

지질계통표의 명칭이 제각기 다른 이유는 아마 수 년에서 수 십년간 연구가 누적되어지며 층서의 구분 방식 변화나, 단위의 명칭 변화, 새로운 층의 발견 등 다양한 요인이 작용하였기 때문일 것이다. 하지만 데이터를 활용하는 사용자의 입장에서 이러한 용어의 불일치함은 혼동을 야기하며, 정보를 얻는데 불편한 요소가 될 것이다. 따라서 용어나 표현법이 변했다면 이러한 것이 변했다는 것을 사용자가 명확히 인지할 수 있도록 해야 한다.

표 1 경상 분지의 층서 구분

Tateiwa(1929)			Chang Ki Hong(1975)		
Gyeong-sang System	Bulguksa Intrusive Rocks		Gyeong-sang Super group	Bulguksa Intrusive Rocks	
	Yucheon Group			Yucheon Group	
	Silla Series	Geoncheonri Fm.		Hayang Group	Geoncheonri Fm.
		Chaeyaksen Fm.			Chaeyaksen Fm.
		Daegu Fm.			Banyaweol Fm.
		Hakbong Fm.			Haman Fm.
	Nakdong Series	Silla Fm.			Hakbong Volc.
		Chilgok Fm.			Silla Fm.
		Jinju Fm.			Chilgok Fm.
		Hasandong Fm.			Jinju Fm.
		Nakdong Fm.		Hasandong Fm.	
				Sindong Group	Nakdong Fm.

(2) 오픈플랫폼 활성화 방안

오픈플랫폼에서 현재까지 등록된 수 많은 지질도를 플랫폼이 자체적으로 하나하나 최신화시키는 것에는 한계가 있다. 어떤 것이 어떻게 변했는지, 무엇이 빠졌는지 등은 직접 사람이 보고 인지하는 것 외에는 별다른 방법이 없기 때문이다. 따라서 오픈플랫폼에서 **“바뀌야 될 정보 찾기 포상 제도”**를 시행하여 사용자로 하여금 자발적인 최신화를 유도하고, 오픈플랫폼을 활성화시킬 수 있도록 하는 방안을 제안한다.

앞서 언급한 예시와 같이 지질 데이터가 너무 옛날의 정보를 담고있거나 변경되었을 때, 이를 발견한 사용자가 **“바뀌야 될 정보 찾기 포상 제도”** 창에 데이터 최신화를 요청 하고, 이에 대한 보상으로 **오픈플랫폼은 사용자에게 포인트를 제공한다.** 이러한 포인트를 일정량 이상 누적 시 지질자원연구원의 굿즈 상품과 교환할 수 있는 포인트샵을 구축한다. 이렇게 한다면 **사용자는 단순히 오래된 지질도를 최신화시키는 것을 넘어서, 포인트를 모으는 재미를 느낄 수도 있다.**

필자는 **CAPTCHA** 기술에서 이러한 아이디어의 영감을 받았는데, 우리가 휴대폰을 보다보면 가끔 **“로봇이 아닙니다.”**라는 문구와 함께 여러 사진 중에서 ‘신호등’이 있는 사진을 고르라는 형식의 창을 보았을 것이다. 이것이 바로 CAPTCHA 기술이다. 이 기술은 인간이 아닌 **봇(bot)**이 무차별적으로 **웹사이트를 공격하거나 해킹을 하는 등의 악의적인 것을 막기위해** 만들어졌다. 하지만 더 나아가 미국 카네기멜론대학의 연구원들은 이 기술을 **고문서를 디지털화하는 작업에** 이용했다. 많은 사람들에게 2-3초 남짓한 짧은 시간을 사용하도록 유도하여, 이 사람들이 짧은 시간동안 사진이나 글과 같은 데이터를 해독해주면 이를 이용해 수 많은 데이터를 해독하고 수집하는데 이용한다는 것이다. **사람들의 작은 시간, 노력을 모아 수 없이 많은 데이터를 최신화할 수 있다는** 점에서 필자는 오픈플랫폼 **“바뀌야 될 정보 찾기 포상 제도”**라는 발상을 떠올리게 됐다.

이러한 제도를 통해 각 대학교의 지질학 연관 학과와 고등학교 등과 협력하여 포인트를 많이 모으는 **경진대회**를 개최하고 또한 연간 최다 포인트 득점자에게는 **상장 등의 포상**을 주어 의욕을 고취시킨다면 사용자들로 하여금 더욱 양질의 지질자원 데이터를 만들 수 있고, 또한 그러한 데이터를 사용자가 이용할 수 있다. 더불어 지오빅데이터 오픈플랫폼은 더욱 활성화될 수 있다.

나. 맺음말

이번 오픈플랫폼 모니터링단 최종 보고서에서 기능 개선 사항과 오픈플랫폼 활성화 방안에 대해 **‘실시간 인기 검색어’, ‘토론창 기능 활성화’, ‘연관된 보고서 더보기 기능’, ‘좋아요 기능’, ‘바뀌야 될 정보 찾기 포상 제도’** 총 5가지의 안건에 대해 제안했다.

머리글에서 말했듯 빅데이터는 현대 과학의 급진적인 성장의 원동력이었고, 이러한 데이터의 진정한 가치를 얻기 위해서 데이터를 체계적이고 단순하게 정리하여 **접근성과 활용성**을 극대화해야 한다. 필자가 이번 보고서에서 제안한 안건들 역시, 사용자가 오픈플랫폼의 빅데이터를 활용할 때 **접근성과 활용성**을 증진시킬 수 있도록 최대한 고려한 것이다. 아마도 오픈플랫폼에서 우리 모니터링단을 모집하게 된 이유 역시 이러한 발전에 힘을 얻고자 함이었을 것이다.

이미 우리에게 누적된 데이터는 수 없이 많고, 현재도 데이터는 지수적인 성장을 보이고 있다. 이러한 빅데이터를 활용하여 지질학의 끊임없는 발전을 위하여는 지질자원연구원 오픈플랫폼과 모니터링단이 힘을 합쳐, 보다 단순하고 보다 체계적인 플랫폼을 구축하도록 꾸준히 바뀌어나가야 할 것이다. 지오빅데이터 오픈플랫폼 모니터링단의 중요성을 다시금 인식할 수 있는 기회였고, 추후에 더욱 모니터링단의 역할이 크게 자리매김하여 플랫폼의 발전에 도움이 되기를 희망한다. [끝]

<참고문헌>

- 김준경. 2020. “실시간 검색어 서비스, 무엇을 남겼나”. *언론중재위원회*. 68-79
- 김현우. 2010. “태깅 시스템의 태그 추천 알고리즘”. *정보과학회논문지*. 931-933
- 박정규. 2021. “경북 경산시 하양읍에 분포하는 백악기 함안층과 반야월층의 생흔화석층”. 16-18