

NFT 블록체인과 딥러닝을 활용한 신뢰성 있는 문제 거래 웹 플랫폼

류병욱¹, 권은혜¹, 신지호¹, 정동주², 이병정¹

¹서울시립대학교 컴퓨터과학부

²(주) 스마트잭

rk97989@naver.com, enehyejjang@gmail.com, sjm010529@gmail.com, lostcode7@gmail.com, bjlee@uos.ac.kr

A Reliable Problem Trading Web Platform Using NFT Blockchain and Deep Learning

Byung-Ook Ryu¹, Eun-Hye Gwon¹, Ji-Ho Shin¹,

Dong-Ju Jung³, Byung-Jeong Lee²

¹Dept. of Computer Science and Engineering, University of Seoul

²Smart Jack Co., Ltd.

요 약

본 연구에서는 다양한 문제들을 ELECTRA를 통해 문제의 유형을 분류하고, 그것을 기존의 문제들과 유사도 평가를 진행하여 문제의 표절 여부를 확인한다. 유일성이 입증된 문제는 이더리움 기반의 블록체인을 사용하여 NFT 기술을 통해 문제를 NFT로 발행하여 거래한다. 최종적으로 표절하지 않았음이 검증된 문제를 NFT를 통해 거래함으로써 문제의 저작권 및 권리를 웹 플랫폼을 통해 거래할 수 있도록 한다. 본 논문에서는 이러한 문제 거래 웹 플랫폼 설계를 기술하며, 본 플랫폼은 문제를 쉽게 거래할 수 있도록 지원하여 문제 거래 활성화에 기여할 것이다.

1. 서론

한국의 사교육 시장에서 가장 큰 규모를 차지하는 고교 수학 시장은 매년 평가원의 기조를 따라가기 위해 수많은 문제집이 출판되고 소비된다. 인터넷 강의를 제공하는 회사, 출판사에선 문제집을 발간하기 위해 자체적으로 개발한 문제를 사용할 뿐만 아니라 공모 통한 '문제 거래'를 수행한다. 하지만, 문제 거래는 문제의 최초 출제자 입증에 어렵고 문제 유사도를 따로 평가하지 않기 때문에 저작권 분쟁을 겪을 수 있다. 이러한 배경에서 본 연구는 문제를 NFT 화하여 저작권을 거래하는 방법을 제시한다. 더불어 이렇게 등록된 문제들을 효율적으로 검색하기 위해 딥러닝 기술을 활용한 문제 추천 서비스를 제공한다.

2. 관련 연구

2.1 ELECTRA

ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately)는 입력 토큰의 일부를 마스킹하는 대신, 작은 규모의 생성 네트워크로 생성한 유사한 가짜 토큰들로 대체하여 기존 문장을 훼손한다. 기존의 모델은 손상된 토큰의 실제 모습을 예측하는 것을 훈련했던 것과 달리, ELECTRA는 판별기(discriminator)가 각 토큰이 실제 토큰인지, 생성기가 생성한 가짜 토큰인지 이진 분류를 진행한다[1]. 전체 토큰의 일부가 아닌 모든 학습 데이터를 활용함으로써 적은 양의 데이터로도 높은 학습 효율성을 갖는다

2.2 문제 유사도 측정

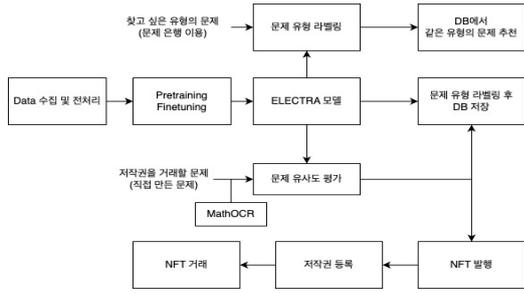
BERT 모델을 활용하여 두 자연어 문장의 유사도를 측정하는 방법이 BLEU, GLEU, TER, CHRf 등 기존에 자연 언어 처리 분야에서 활용되던 방법보다 결과가 뛰어났다[2]. 본 연구에서는 ELECTRA 모델을 활용하여 원본 문제와 비교할 문제의 임베딩을 생성한다[1]. 두 임베딩을 벡터화하여 코사인 유사도를 얻는다. 일정 기준치 이상의 유사도가 나온다면 표절로 판단한다. BERT와 더불어 ELECTRA도 문장의 문맥을 고려하기 때문에 수학 문제의 의미적 유사도를 판단한다.

2.3 NFT(Non-fungible token)

블록체인은 탈중앙화된 시스템으로, 안정성과 투명성을 띠고 있다. 이를 증명하는 증명서인 NFT를 이용한 미술품 소유권 증명 DApp(Decentralized App)이 제안됐다[3]. 이더리움(ethereum)으로 배포되는 스마트 컨트랙트(smart contract)를 이용했으며, 소유권(증명서)은 ERC-721 표준에 따라 구현되어 독립적인 소유권 증명이 가능하다. 또한, 실시간으로 경매, 입찰, 소유권을 추적할 수 있다. 미술품 거래 시, NFT의 주소가 원소유자에서 최종 구매자로 변경되고, 구매자가 입찰한 금액의 이더(ETH)가 원소유자에게 전송되게끔 하여 경매소 역할을 한다.

3. 문제 거래 웹 플랫폼

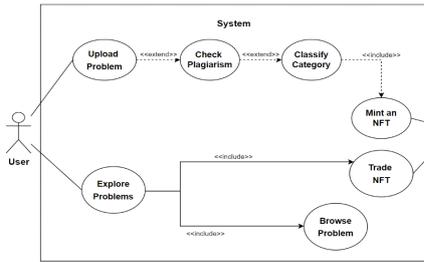
3.1 시스템 개요



(그림 1) 개념 시나리오

그림 1은 본 연구의 전체적인 개념 시나리오이다. 사용자가 저작권을 거래할 문제를 업로드하면 MathOCR를 이용해 수식을 포함한 모든 내용을 텍스트로 변환한다. 시중의 문제집과 기출문제를 학습한 ELECTRA 모델로 문제 유사도를 평가하여 표절 여부를 판단한다. 표절이 아닌 것으로 판별되면, NFT를 발행하고 거래할 수 있다. 또한, ELECTRA 모델로 문제 유형을 라벨링 하여 데이터베이스에 저장한다. 찾고 싶은 유형의 문제를 업로드하면 ELECTRA로 모델로 문제 유형을 라벨링 하여 데이터베이스에서 같은 유형의 문제를 추천한다.

3.2 유스케이스도



(그림 2) 유스케이스도

그림 2는 본 연구에서 개발하는 웹 플랫폼의 유스케이스도이다. 본 플랫폼은 사용자가 문제를 제작하여 업로드하면, 기존 문제들과의 표절 여부를 검사한다. 문제가 기존 문제를 표절하지 않았다면, 딥러닝 모델을 사용하여 문제의 유형을 분류한 후 문제를 NFT로 발행 후 플랫폼에서 거래할 수 있다. 사용자는 거래 중인 NFT를 문제 유형으로 구분하여 확인할 수 있다. 사용자는 원하는 문제 NFT를 스마트 계약을 통해 거래할 수 있다. 마지막으로, 사용자는 문제를 NFT로 거래하지 않더라도 문제 열람권을 획득한 뒤, 문제를 열람할 수 있다.

3.3 UI 설계



(그림 3) 문제 풀기 화면



(그림 4) 문제 거래 화면

그림 3과 그림 4는 본 웹 플랫폼의 UI 일부이다. 그림 3은 유형별 문제 화면으로, 직접 유형을 선택 후, 문제를 제작한다. 업로드된 문제는 AI를 이용한 자동 유형 분석도 가능하다. 그림 4는 문제 NFT 거래 화면으로, 문제를 검색하고 가격, 판매자를 확인하여 구매할 수 있다. 또한, 문제 등록 버튼으로 NFT를 발행하여 판매도 가능하다.

4. 결론

본 연구에서 제시한 웹 플랫폼을 통해, 문제 검색자는 유형별로 분류된 문제를 추천받을 수 있고, 문제 제작자는 제작한 문제를 NFT로 발행하여 등록할 수 있다. 기존에는 주로 공모전을 통해 문제가 거래되고 있었지만, 본 연구를 통해 플랫폼 기반 거래로 확장하여 더 다양한 문제가 거래될 수 있을 것을 기대한다. 또한, 등록된 문제들이 ELECTRA 모델을 통해 유형별로 분류되며, NFT로 발행되었기 때문에 문제의 유일성까지 입증할 수 있을 것이다.

사사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (NO.2022M3J6A108 4845)

참고문헌

[1] K. Clark, M. Luong, Q. Le, and C. Manning, "ELECTRA : Pre-Training Text Encoders As Discriminators Rather Than Generators," International Conference on Learning Representations (ICLR) 2020.

[2] 현중환, 최호진, "BERT 기반 한국어 문장의 유사도 측정 방법," 한국정보과학회 한글 및 한국어 정보처리 학술대회, 2019, pp. 383-387.

[3] 남부성, 유근우, 윤중배, 전정민, 홍충선, "Non-fungible Token 을 이용한 미술품 이력 관리 방안 연구," 한국정보과학회 한국컴퓨터종합학술대회, 2019, pp. 1970-1972.