

# 블록체인을 이용한 웹 사이트 이용 기록 관리 시스템

연성화, 유성현, 김흠, 주뢰, 염철민, 원유재\*

충남대학교 컴퓨터공학과

{yeonseonghwa, yoursaint, 201601994, 201602080, cjfals18, yjwon}@cnu.ac.kr ,

\*Corresponding author

## Management System Using Blockchain for Web Site Usage History

Seonghwa Yeon, Sunghyun Yu, Heum Kem, Zhu Lei, Cheolmin Yeon, Yoojae Won\*

Dept. of Computer Science, Chung-Nam University

### 요약

빅데이터와 AI의 인기에 힘입어 데이터 확보가 중요해지면서 대량의 개인정보가 무차별적으로 수집되고 있다. 그로 인하여 마이 데이터(Mydata)에 대한 관심 또한 대두되고 있다. 오늘날 데이터 수집의 주체들은 정보 수집시에 불필요하거나 대량의 웹 트래픽을 발생시켜 웹 자원을 소모하고 있다. 또한 그들은 정보의 소유자가 아닌 데도 불구하고 그를 이용하여 이득을 취하고 있지만 정보 소유자는 보상을 받지 못하고 있다. 정보의 활용에 대한 정확한 기록과 보상을 제공하고 웹에서의 불필요한 자동화 프로그램의 사용을 지양하기 위해 본 연구에서는, 블록체인(Blockchain)을 활용하여 사용자의 정보를 제어하는 시스템을 구축하고 기록된 정보를 바탕으로 구축된 웹 사이트 검색엔진을 제안한다.

### 1. 서론

빅데이터와 AI의 인기에 힘입어 마이 데이터에 대한 관심 또한 대두되고 있다. 마이 데이터란 뜻 그대로 내 데이터를 총칭하는 말이기도 하지만, 핵심은 정보 사용 및 제공의 주체를 기업에서 개인으로 바꾼다는 데에 있다는 점에서 데이터를 넘어서 그 데이터의 주권을 담고 있다[1]. 하지만 실제 많은 개인 맞춤형 서비스나 빅데이터를 이용한 서비스들은 데이터 주권을 무시하고 수집해 사용하고 있다. 특히 웹 검색엔진은 대부분 사용자의 정보를 가지고 있으며 사용자의 행위를 수집해 맞춤 서비스를 제공한다. 이 과정에서 사용자는 자신의 데이터가 어떻게 수집되고 사용되는지 알 수 없지만 결과적으로 서비스 제공자에게 이익이 돌아가게 된다. 이러한 문제점을 해결하기 위해서 사용자가 자신의 데이터 흐름을 관리하거나 확인할 수 있어야 하며 제 3 자가 자신의 데이터를 사용할 경우 본인의 확인과 데이터 제공에 대한 보상을 받는 환경이 구축되어야 한다.

기존의 웹 검색엔진은 사용자 데이터의 주권을 침해하는 것 말고도 트래픽에 관한 문제점이 있다. 검색엔진은 사용자에게 검색서비스를 제공하기 위해 봇과 웹 크롤러 등의 자동화 프로그램을 사용해 수많은 웹 트래픽을 발생시키고 있으며 이는 전체 웹 트래픽의 절반 정도를 차지하고 있다[2]. 이렇게 발생하는 트래픽은 여러가지 피해를 유발한다. 특히 자원이 한정되어 있는 저 사양 서버에서는 불필요한 자원 소모를 발생시키면 서버가 마비되어 서비스를 중단되는 등의 심각한 피해를 끼친다. 때문에 불필요한 트래픽의 발생을 개선할 방법에 대한 연구가 필요하다.

본 연구는 데이터 주권 침해와 검색엔진의 트래픽 문제점을 해결하기 위해 블록체인의 Smart Contract를 활용하는 시스템을 구축한다. 또한 활용 방법으로 기록된 정보를 바탕으로 구축된 웹 사이트 검색엔진을 제시한다. 블록체인을 사용해 사용자의 웹 사이트 방문 기록을 저장하게 되면 데이터의 변조를 방지할 수 있는 시스템을 제공할 수 있으며[3], Smart Contract를 통해 제공자와 사용자 사이의 비용지불 등의 처리를 자동화해 효율성을 늘릴 수 있다[4]. 이렇게 저장된 정보로 웹 사이트 검색엔진을 구축해 불필요한 자동화 프로그램의 사용을 줄여 더 나은 웹 생태계를 만들 수 있다.

### 2. 관련 연구

#### 2.1 개인정보 취급 범위

개인정보의 판단기준으로는 개인을 알아볼 수 있는 정보, 살아있는 개인에 관한 정보, 특정 개인과의 관련성, 식별 가능성 등등이 있다[5].

#### 2.2 블록체인을 활용한 개인정보 관리 기법

프라이빗 블록체인을 활용하여 사용자는 개인정보를 블록체인 관리자의 Key-Value 데이터베이스에 저장하고 개인정보 사용 내역을 분산원장으로 관리한다. 서비스 제공자는 사용자의 개인정보를 직접 저장하지 않고 key 값만 저장한다. 필요에 따라 key 값을 이용하여 블록체인 관리자에게 고객의 개인정보를 요청하여 열람한다[6].

### 2.3 개인정보 보호의 필요성과 문제점

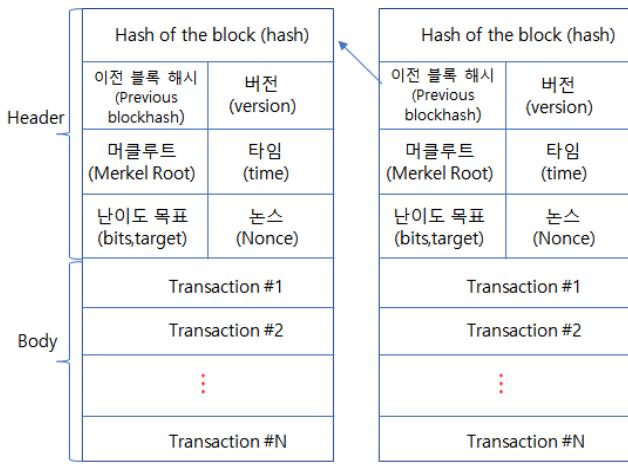
빅데이터 환경에서 개인정보의 침해 가능성이 증폭되고 있다. 개인에 대한 행위추적이나 해위패턴 추측해서 불이익을 줄 수 있다. 사용자들은 자신의 정보가 어떻게 생성되고 처리되는지 파악하는 것에 어려움을 겪고 있다. 개인정보 보호법 제3조를 따르면 개인정보 보호에 대한 정책이 충분하지만 직접적으로 구속하지 않기 때문에 원칙을 준수하지 않을 유인이 상당히 존재한다[7]. 그렇기 때문에 법적 정책으로는 무분별 수집을 막는 것이 어렵다.

## 3. 배경지식

### 3.1 Block Chain

대표적인 가상화폐인 비트코인은 온라인 가상화폐로 발행기관의 통제없이 사람과 사람간의 P2P 기술로 통해 인터넷상에서 익명으로 거래되는 화폐이다[8].

블록체인은 데이터 분산 처리 기술을 사용하여 네트워크에 참여하는 모든 사용자가 모든 거래 내역 등의 데이터를 분산, 저장하는 기술이다. 중앙서버 없이 여러 명이 데이터를 저장하기 때문에 참여자 과반수 동시에 공격하지 않는 이상 공격이 불가능하다.



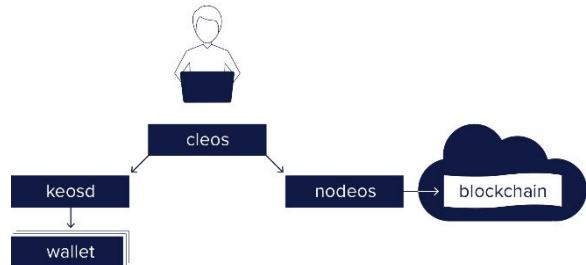
(그림 1) 블록체인의 블록 구조

블록체인은 (그림 1)처럼 서로 연결된 블록으로 만들어진다. 블록은 크게 블록헤더와 트랜잭션 두 부분으로 구성된다. 블록헤더에는 이전 헤더의 해시값, 버전, 시간 등등 6개의 정보로 구성된다. 블록이 기본 정보를 담고 있다면 트랜잭션은 주요정보를 담는다[9].

### 3.2 EOS

EOS는 Delegated Proof Of Stake (DPoS) 방식을 사용하는 제3세대 암호화폐이며 블록체인 기술에 기반한 탈중앙화 플랫폼이다. EOS는 분산 애플리케이션을 지원하도록 설계되어 있으며 Smart Contract 개발이 가능한 동시에 수수료의 부담이 없고 약 200배의 빠른 처리속도를 가지고 있다. 사용성 측면에서는 복잡한 주소를 알아보기 쉽게 계정이라는 개념으로 바꾸었고 Decentralized applications(DApp) 측면에서는 다른 블록체인에 비해 빠른 처리속도를 가지고 있다.

EOS는 nodeos, cleos, keosd 모듈로 구성되어 있다.



(그림 2) EOS의 모듈 관계도

Nodeos는 EOSIO를 설치하는 서버에서 블록을 생성하게 하는 파일이다. 여러 플러그인을 결합해 실행할 수 있고, 초당 2개의 빠른 블록 생성 속도를 가지고 있다.

Cleos는 체인과 로컬 지갑과 상호 작용하기 위한 명령어 인터페이스를 제공한다.

Keosd는 지갑을 관리하는 모듈이다. 지갑을 만들고 그 지갑에 키들을 저장하고 관리한다.

### 3.3 MongoDB

MongoDB는 관계형 데이터베이스의 한계를 극복하기 위해 만들어진 새로운 형태의 데이터베이스이다. 기존 데이터베이스와 달리 스키마가 자유롭고 JSON 형식의 문서에 데이터를 저장한다. 데이터를 저장할 때 컬렉션을 사용해 유사한 문서들을 그룹으로 묶어준다. 따라서 관계형 데이터베이스에 비해 뛰어난 조회 효율을 가지고 있다.

EOS에서 사용하는 데이터는 대부분 JSON 형식으로 되어 있어 mongodb를 사용하기에는 아주 적합하다. mongo\_db\_plugin을 지원해주고 있어 블록체인 데이터를 MongoDB에 보관할 수 있다. mongodb 없이 블록 내용을 조회할 수 있지만 mongodb를 활용하게 되면 쿼리를 이용해서 원하는 데이터를 쉽게 조회할 수 있다.

### 3.4 Elasticsearch

Elasticsearch를 사용하면 방대한 양의 데이터를 신속하게 저장, 검색, 분석할 수 있다[11]. 관계형 데이터베이스를 사용해서 조회할 경우 모든 테이블을 스캔해서 조회하는 내용을 매핑한다. 따라서 다큐먼트가 많을수록 좋지 않는 성능이 보인다. 실제로 만개만 초과해도 조회가 엄청 느리다는 단점이 존재하기 때문에 검색엔진으로 사용하기에는 효과가 좋지 않다.

&lt;표 1&gt; keyword 와 Document 의 관계 표

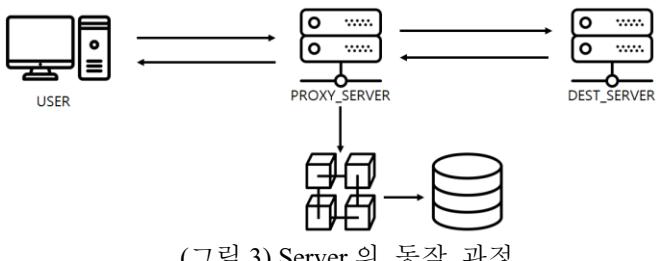
Keyword	Document
Tomato	A,B,C
Apple	C,D,E
Orange	F,G,H

<표 1>를 살펴보면 tomato라는 키워드를 사용해서 검색할 경우 이미 tomato가 어느 document에 존재하는지 미리 알고 있기 때문에 해당하는 document 즉 A, B, C가 반환되게 된다.

사용자가 “인천터미널”이라는 키워드로 조회할 경우 관계형 데이터베이스에서는 “인천터미널”이라는 내용만 조회하고 “인천종합터미널”에 대한 정보는 조회하지 못하게 된다. 하지만 Elasticsearch의 경우 “인천터미널”을 “인천”, “터미널” 두개의 키워드로 분할해서 조회하게 되고 두 키워드를 포함하는 모든 문장을 조회하게 된다.

Elasticsearch는 전체 조회와 관련성 평점 시스템을 지원한다. 따라서 조회된 결과는 점수에 따라 내림차순으로 나열해서 반환하게 된다. 사용자가 “인천터미널”을 조회할 때 “인천터미널”을 포함하는 document는 “인천” 혹은 “터미널”을 포함하는 document 보다 앞서 있다.

#### 4. 구현 원리



블록체인 네트워크는 위임지분증명 방식을 사용하는 플랫폼을 사용한다(EOS). 또한 모든 기능은 웹으로 사용할 수 있도록 구현된다. 사용자의 정보 보호와 별도의 프로그램 없이 웹 기록을 체인에 기록하기 위하여 프록시 서버가 사용된다. 모든 데이터는 Smart Contract를 통해 블록에 저장되며 데이터의 원활한 활용을 위해 데이터베이스가 사용되고 사용자의 더 나은 검색결과 경험을 위해 별도의 검색엔진을 사용한다.

이용자는 블록체인 네트워크의 여러 노드 중 선택하여 서비스를 이용할 수 있다. 더 좋은 경험을 제공하는 노드를 선택하여 이용할 수 있으며 토큰을 보유하고 있다면 해당 토큰의 영향력만큼 내가 원하는 노드에 투표하여 해당 노드의 영향력을 증가시키거나 Block Produce(BP)로 선출하여 메인 서비스 제공자로 승격시킬 수 있다.

이용자는 투표로 선출된 블록체인 네트워크의 노드 중 하나의 검색 엔진 사이트에 접속한다. 노드의 사이트는 검색 엔진에 쿼리 전송과 프록시 서버를 담당한다. 블록체인 네트워크의 노드는 contract를 이용하여 정보를 블록체인에 기록하며 이를 효율적으로 활용하기 위해 데이터베이스에 블록 데이터를 기록한다. 데이터베이스에 저장된 정보는 검색엔진에 전송되어 검색 엔진 사이트에 활용된다.

웹 기록은 본 서버의 프록시 서버를 통해 자연스럽게 수집된다. 프록시 서버를 이용하기 때문에 사용자가 방문하는 사이트에는 사용자의 정보가 전송되지 않다.



수집된 웹 기록은 Smart Contract를 통해 블록체인 네트워크로 전송된다. 전송된 기록은 블록체인 내부의 multi index 데이터베이스에 저장되며 이는 일정한 시간동안 저장되어 사용자의 개인 맞춤형 서비스 제공에 활용된다. 개인 맞춤형 서비스를 위해 나의 개인정보가 활용된다면 이 또한 Smart Contract를 통해 P2P 거래를 통해 지갑에 직접 보상이 전달된다.

블록체인 네트워크로 전송된 기록 즉 블록에 저장된 기록은 활용을 위해 데이터베이스와 연동된다. 데이터베이스에 연동된 정보를 가공하여 검색엔진에 전송하고 이를 활용하여 블록체인과 마이 데이터를 활용한 검색엔진을 구축하며 이 검색엔진은 사용자에게 직접 사이트에 대한 피드백을 받아 기록하고 이를 활용하여 사이트에 대한 선호도 점수를 측정하고 이를 활용하여 검색 결과 표시시에 표시 순서를 정하거나 관련 웹페이지를 추천하는데 사용된다.

검색 엔진은 데이터베이스에 저장된 정보의 양이 많을수록 더 많고 정확한 정보를 사용자에게 표시할 수 있으며 데이터베이스는 블록에 기록된 데이터의 양이 많을수록 정보가 증가한다. 블록체인은 Smart Contract를 위한 transaction의 전송이 많을수록 블록에 많은 정보가 저장되며, 이는 사용자의 웹 기록이 많이 발생할수록 증가한다.

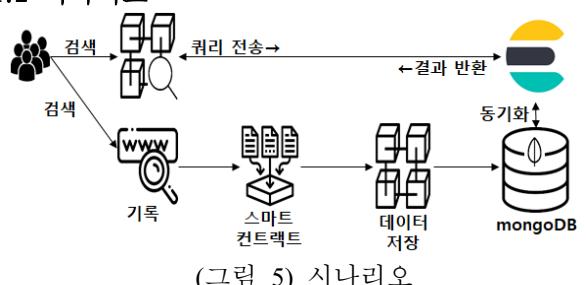
따라서 이용자들이 많을수록, 해당 서비스를 이용한 웹 기록이 많아 질수록 transaction이 많이 발생하게 되고 그 결과로 본 서비스는 발전하게 된다. 또한 우리의 웹 서핑 활동이 현재 검색 엔진 사이트들의 봇과 크롤러 보다 효율적이고 의미 있는 링크를 클릭하기 때문에 보다 의미 있는 데이터를 바탕으로 검색 엔진을 구축한다.

#### 5. 실험

##### 4.1 실험 환경

Ubuntu16.04에서 EOS 블록체인을 구축하고 수집한 데이터를 저장 및 관리하며 Windows10 환경에서 검색 서비스를 제공한다. 블록체인을 구성하기 위해서 EOS는 1.7.4 버전을 사용하였고 Smart Contract를 작성하기 위해서 EOSIO.CDT는 1.6.1 버전을 사용하였다.

##### 4.2 시나리오



(그림 5)처럼 검색엔진을 사용하듯이 사용자가 원하는 사이트를 방문하면 지갑 이름, 검색 키워드, 사이트 주소, 좋아요 등을 수집하여 Smart Contract를 통해서 블록체인에 각 정보를 저장한다. transaction이 발생하는 동시에 각 데이터는 mongodb와 Elasticsearch에 동기화한다. 만약 사용자가 원하는 정보를 블록체인에서 검색한다면 Elasticsearch에 데이터가 동기화되어 있기 때문에 Elasticsearch에서 사용자가 원하는 정보를 검색하여 추천이 높은 순으로 반환하여 보여준다.

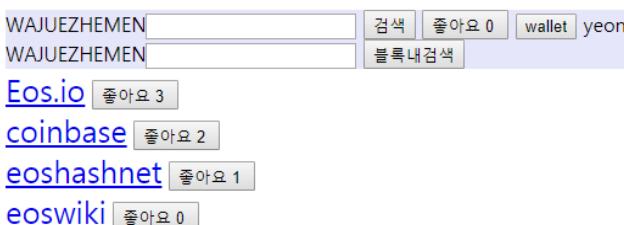
#### 4.3 실험 과정

사용자는 자신의 지갑 계정을 만들 수 있으며 지갑 계정에 이름, 코인정보가 담겨 있다. 지갑을 생성하고 검색엔진을 사용할 경우 검색엔진은 사용자의 정보를 수집하고 검색 결과에 반영하며 사용자에게는 보상을 제공한다. 반대로 정보 제공을 원하지 않을 경우에는 지갑 생성하지 않고 사용하면 된다. 검색엔진은 검색 기능을 제공하며 사용자가 주로 검색하는 내용을 수집하여 사용자의 지갑 이름, 검색 키워드, 사이트 주소를 수집해 블록에 저장한다. 추가로 유용하다고 생각하는 페이지에 좋아요를 선택할 수 있으며 좋아요의 개수도 블록체인에 저장하고 이를 이용해 신뢰도 또는 유용성을 결정한다. 블록체인내에서 검색을 할 때는 블록체인 내에 저장된 정보의 좋아요에 따라 내림차순으로 보여준다.

<표 2> 수집된 정보

지갑 이름	검색 키워드	웹 주소(예시)	좋아요
A,C,D	Eos	Eos.io	3
A,C	Eos	coinbase	2
A	Eos	eoshashnet	1
B	Eos	eoswiki	0

블록체인내 검색할 때 블록체인에 저장된 각 정보들이 <표 2>의 순서로 웹에 표시된다. 사용자가 Eos에 대해 검색을 하면 좋아요 내림차순으로 Eos.io, coinbase, eoshashnet, eoswiki 가 웹에 표시되어 많이 채택되는 문서를 먼저 읽게 된다. 표시된 문서에 좋아요를 하게 되면 해당 문서를 제공하는 계정 즉 지갑에게 일정한 보수가 들어가게 된다.



(그림 6) 블록내 검색 내용

제안하는 검색엔진은 기존 서비스와 달리 자신의 활동 정보를 이용해 사용자도 이의 창출이 가능하며 블록체인을 기반으로 한 정보 저장관리 시스템이기 때문에 정보의 안전성과 신뢰도가 보장된다.

#### 5. 결론

본 논문은 블록체인을 이용하여 사용자의 마이데이터를 기록하는 시스템과 그를 활용하는 시스템을 제안한다. 이를 통해 정보를 관리하고 정보의 활용에 대한 보상을 합당하게 지불하는 방법과 자동화 프로그램의 불필요한 웹 트래픽을 최소화하는 방법을 서술하고 있다. 이 연구를 통해서 마이데이터는 보다 더 투명하게 관리되고 그 내용을 사용자가 직접 확인하고 개입할 수 있으며 데이터 주권 의식을 가지게 될 것이다. 이때 주권 의식을 가진 사용자들의 합당한 보상을 요구한다면 필요한 효율적인 방안 또한 제공하여 데이터 주권 의식의 향상에 기여할 것이다.

활용 방법으로 제시된 웹 사이트 검색엔진은 사용자 웹 사용 기록을 바탕으로 구축되어 웹 기록 수집을 위해 자동화 프로그램에 의존하여 무분별한 트래픽을 발생시키는 고전적인 방법에서 벗어나 더 나은 웹 생태계 환경을 만드는데 기여할 것이다. 서비스 유지를 위한 자원의 소모는 줄이고, 사용자의 정보로 구축되는 만큼 보다 더 사용자 친화적인 서비스를 제공함으로써 사용자의 더 나은 경험에 기여할 것이다.

앞으로 해당 내용을 더 발전시켜 데이터 주권 시대를 여는데 기여하고 더 나은 웹 환경을 만드는 데 목표를 둘 것이다.

#### [Acknowledgement]

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음"(IITP-2019-2016-0-00304\*)

"본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음 "(2015-0-00930)

#### 참고문헌

- [1] 김상목. “마이 데이터 그리고 잊혀질 권리”
- [2] GlobalDots “2019 bad bot report”
- [3] Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system
- [4] Idelberger, Florian, et al. “Evaluation of logic-based smart contracts for blockchain systems.”
- [5] 김솔왕. “제한능력자의 개인정보보호 및 후견인의 개인정보 취급 범위에 관한 연구”
- [6] 지용선, “블록체인을 활용한 개인정보 관리 기법”
- [7] 정한신 “빅데이터 환경에서 개인정보 보호에 관한 헌법적 연구”
- [8] 소프트웨어 정책연구소 “비트코인의 기반 기술 블록체인의 원리”
- [9] 이기영. “A Study on Application of Record Management System Block Chain Technology”
- [10] EOSIO Developer Portal. “<https://developers.eos.io>”
- [11] Elasticsearch doc. <https://www.elastic.co/kr>