Koi

Incentives for an Open Data Economy

Abstract

The Koi Network provides incentives for participants to seek out the best content on the web, archive it, and receive dividends when the archive is viewed in the future. Through the use of the Arweave Permaweb [Williams, 2018], Koi participants will soon be able to receive lifelong dividends for submitting content that receives a large amount of attention.

Anyone with an internet connection can install the node app to act as a witness to this process, and receive a share of the rewards for helping us to ensure that the archive is a valid and accurate representation of the source content. Additionally, we provide an access layer to decentralized data marketplaces such as Ocean [McConaghy, 2020] to ensure that this archive can be repackaged by any member of the community and sold as index data for interested parties.

We expect that as this technology comes to maturity it will stimulate a new growth in decentralized systems, especially in the areas of search, content moderation, and artificial intelligence.

As the network grows, fees and earnings will be reinvested through the Koi DAO to ensure that new technologies are incorporated, and the system can continue to expand.

For more information, please visit openkoi.com, or send us an email at hello@openkoi.com.

Contents

1 Incentivising an Honest Archive of the Web

Accountability is hard on the internet. The events of 2020 have made it clear that there is a vacuum for reliable discourse and a real threat that censorship may result in a loss of knowledge, quite possibly forever. From the protests in Hong Kong to the turmoil in US and European elections, it has become clear that the internet in its current state is often no longer a reliable source of information.

Koi provides a solution to these problems by incentivizing the creation of an open knowledge archive. We apply free-market principles with deliberate mechanism design to ensure competition and accountability, and are working to build the technology infrastructure that allows for reliable mapping of content, which allows for transparent discourse.

1.1 Game Model

The Koi Network engages in a content-finding game, and participants are rewarded proportionately to the value of the data they create. Every piece of content that is submitted is stored for a minimum of 100 years and, during that time, pays a dividend to the miner who submitted it.

The Koi Network is made up of two parties, with an optional third party providing added incentives. The **content miners** gather information they believe to be valuable from across the web, and archive it to the permaweb (see 3.1). Once archived, the content is then verified by **peer witnesses** automatically to ensure that it has not been falsified or tampered with. Finally, a **customer** can incentivize the collection of particular data by setting a bounty.

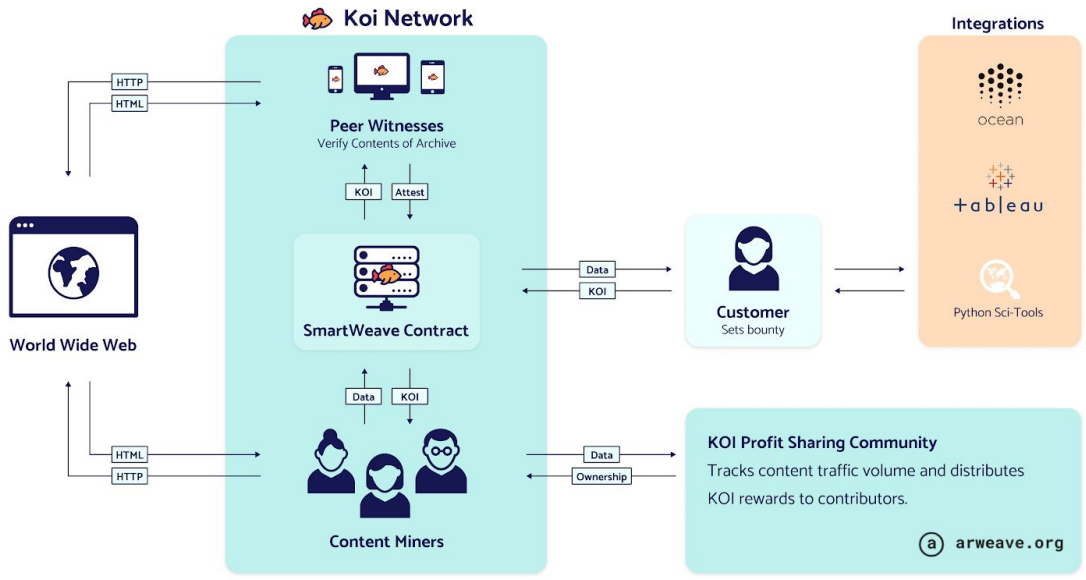

Figure 1: The Koi Network gathers data from the web and curates a high-quality permanent archive.

1.1.1 Summary

During each round, the network distributes new KOI to the owner of the data assets which have received the most attention during the game period. The process is as follows:

1. **Content Miners** crawl the web for useful content, and encode it into a common format
2. **Content Miners** submit standardized data sets to the **KOI Profit Sharing Community**, storing them for a minimum of 100 years
3. **Content Miners burn KOI** requesting the peer network to validate that the content matches the submitted source URL
4. **Peer Witness Nodes** compete to validate the data, and **stake to submit attestations**
5. After the period has ended, any **Miner** can call the **Distribution Function** within the smartweave contract, and new KOI are released and distributed to miners proportional to the share of traffic their content generated during the period

Note: At any time, any user can add an additional KOI bounty to a page or other web resource, incentivising miners to index it and add it to the archive.

1.1.2 Participants

1.1.2.1 Peer Witnesses

In order to verify that the content submitted by miners is accurate and matches the true content of the web resource in question, peer devices are incentivized to verify and attest to the validity of the permaweb archive. This accomplishes two things: first, it ensures a transparent chain of accountability for the archived data, and second, it removes Koi tokens from circulation in the form of staking, and distributes rewards to any peer that participates. Peer witnesses do not need to be technologically sophisticated, and can simply install a 'node' client to begin participating and earn Koi.

1.1.2.2 Content Miners

Web crawling is a mature technology in which a programmatic agent navigates through the internet from page to page, seeking content. While this is the basis of a majority of online content gathering systems (search engines, analytics tools, and more) there is currently no open marketplace for these businesses to compete. Koi provides incentives for these existing players to expose some portion of their data to the public and to contribute to our archive.

1.1.2.3 Data Curators / Customers

Whether for personal use, or on a larger scale as a business, a wide range of use cases exist for gathering online data. A customer of the Koi Network can easily set a bounty to request certain data types be collected from the web, and this will incentivize miners to focus on those areas to reduce their costs.

1.2 Mechanism Design

The purpose of this network is to establish an archive of the most accurate, relevant, and useful information online. This is accomplished by compensating **content miners** for the volume of attention that their archived content receives over the long term.

Every day, the Koi Smartweave Contract will release a fixed amount of KOI tokens as rewards. The rewards are proportionally distributed to the owners of the content which received the most attention during that period. Attention is easily tracked through Arweave (see 3.1), and the contract is able to verify the creation, existence, and popularity of the content at each interval.

Initially, rewards will be quite substantial as there will be a limited quantity of content within the archive. As the network grows, we expect greater competition among miners, which should make rewards increasingly valuable.

2 Technology

The Koi stack combines mature application-layer technologies such as HTML parsing and headless browsing with more modern tools like the Arweave permaweb and token-based economics ("tokenomics") to provide a solution which can provide highly compatible services in a fully decentralized way.

2.1 The Arweave Permaweb

The Arweave network has established a permanent archival tool referred to as the permaweb [Williams, 2018] which is structured similar to Bitcoin [Nakomoto, 2008] or Ethereum [Buterin, 2015] but makes use of a more scalable architecture, which is designed specifically to reward users for storing data, and which can support a substantially higher transaction throughput via its blockweave structure.

2.1.1 Profit-Sharing Communities

While profit-sharing is not a new concept, the decentralized version of this model provides many new advantages. In particular, permaweb profit sharing communities support a new kind of engagement through direct economic representation and incentives. [Community.xyz, 2020]

Profit Sharing Tokens are minted when new content is added, and dividends are distributed when revenue is generated.

2.1.2 Attention Tracking

Previous blockchain projects such as Steemit [Steemit, 2018] and the Basic Attention Token [Brave, 2018] have established the feasibility of tracking 'reads' of information through nodes of the network. On the Arweave Permaweb, this feature is built into the SmartWeave contract format, which supports not only attention tracking but also the direct verification and reading of on-chain content archives.

2.1.3 Throughput

Unlike typical proof-of-work blockchains, Arweave utilizes a blockweave that can be expanded asynchronously, supports a much higher rate of content creation, and has much lower per-transaction costs. This is required not only for the volume of content that Koi will create, but also for the feasibility of micropayments, which is not economically viable on Ethereum or other proof-of-work blockchains.

2.2 Koi Network

The Koi Network exists to retrieve, verify, and archive the content of the web. This process follows three key steps: Encode, Crawl, and Verify; more details on these below.

2.2.1 Web Encoding

As a matter of necessity, the Koi Network creates a structured representation of each page that is added to the archive, and maps the resource to identify the useful parts of the data. Because only the important parts of the page are archived, the network is able to reach consensus on otherwise complex content pages, and peer witnesses can easily verify that the content matches the archive.

When a bounty is set, either by the DAO or a customer, specific parts of the web page are identified for archival purposes. Miners have the freedom to archive additional information if they believe it may become valuable in the future.

2.2.2 Profitable Web Crawling

Many organizations already engage in web crawling. The Koi Network gives them a way to monetize their insights by receiving dividends for submitting content to the permaweb archive.

2.2.3 Low-cost 'witness' mining

Because peer nodes are not storing any data or performing costly operations like hashing, the Koi witness node can operate on a wide range of devices with minimal cost to the end user.

3 Tokenomics

The KOI token is used to interact with the network and is designed to reward contributors to the archive as the network grows over time. Most importantly, the day-to-day function of the network will reduce circulating supply through required burning and staking. The core value of the network as a web crawler and scraper will ensure demand for the tokens as they are required to access these services.

3.1 Applications & Utility

The primary function of the Koi Network is to provide distributed web scraping capabilities to support businesses and individuals gathering and curating web-data. The Koi DAO will gradually build up a number of Profit Sharing Tokens for key resources, then distribute dividends received towards continuing reinvestment in the protocol and supporting technologies.

In 2021, the market for machine learning training data is expected to top USD $10B [McKinsey, 2017], and a wide range of businesses from e-commerce companies to airlines use online data to drive their decision-making processes.

The following are just a few examples of the utility of this information, and how it can be commoditized and monetized by the Koi process:

- Content Archives (historical reference data is stored on Arweave and other archives)
- Machine Learning training data (provides images and text, labelled consistently)
- Public Record Data in machine-friendly format (e.g. licensing or financial data)
- Product catalogues, comparative pricing, and general e-commerce data
- Market Data (flight prices, commodity information, weather, etc.)

As the network grows, the DAO will provide grants and funding to integrate KOI as a data mining tool with existing data analysis suites such as Python Scitools and Tableau, so that business analysts and other customers can easily take advantage of our network of on-demand web scraping tools and nodes.

3.2 Long Term Value Creation

As the network grows, the main driver of token value will be the desire to add to the archive and receive dividends. Anyone can upload content by burning KOI, which creates an open contest among creators, search engines, and anyone else with an internet connection, which can produce valuable data assets with long term dividends, ultimately enabling a new form of wealth creation by helping to write history.

3.2.1 Implicit Content Appraisal

By creating an open marketplace for content archiving, we also establish the future value of current content, which might provide useful information for content moderators or those seeking to build search engines and other technologies that provide content hierarchies.

3.2.2 Indexing and Search

By keeping the attention data of this network public, Koi will create a new paradigm under which any internet user can deploy a personal intelligent agent to search on their behalf, without the need to consult a formal authority to retrieve this information.

3.2.3 Data Marketplaces

The notion of data marketplaces is quite old, but they've recently become quite popular. Anyone is welcome to repackage the Koi data sets, process them, clean them, and sell them through data marketplaces. Ultimately, this practice will further the consensus of the network and add to the attention calculus, leading to a stronger overall network, and a more tightly integrated permaweb.

3.3 Token Buyback

Finally, as the Koi DAO builds a larger and larger store of valuable Profit Sharing Tokens, there's the possibility that the proceeds can be used to repurchase Koi tokens and redistribute them through a community grants pool to improve the existing technology stack and expand integrations.

References

1. Berners-Lee, T. Semantic Web Road map. http://www.w3.org/DesignIssues/Semantic.html (last modified Oct. 14, 1998)

2. Brave Software, "Basic Attention Token", https://basicattentiontoken.org/wp-content/uploads/2017/05/BasicAttentionTokenWhitePaper-4.pdf, 2018

3. Community.xyz Documentation, Arweave, https://github.com/CommunityXYZ/community-js/blob/master/docs/classes/_community_.community.md, 2020

4. Marshall, C.C. and Shipman, F.M. n.d. "Which Semantic Web." [Online]. Available WWW: http://www.csdl.tamu.edu/~marshall/ht03-sw-4.pdf. 2008

5. Mary L. Gray and Siddharth Suri, "Ghost Work" 2019

6. McKinsey & Company, "Artificial Intelligence, the Next Digital Frontier", https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx, 2017

7. Salih Ismail1 and Talal Shaikh2, "A Literature Review on Semantic Web - Understanding the Pioneers' Perspective" Sixth International Conference on Computer Science, Engineering & Applications DOI: 10.5121/csit.2016.61102 05 Sep 2016

8. S. Williams, V. Diordiiev, L. Berman, I. Raybould, I. Uemlianin, "Arweave: A Protocol for Economically Sustainable Information Permanence" https://www.arweave.org/yellow-paper.pdf, 2018

9. Steemit, "Steem", https://steem.com/steem-whitepaper.pdf, 2018

10. Trent McConaghy, "Ocean Protocol: Tools for the Web3 Data Economy", https://oceanprotocol.com/tech-whitepaper.pdf 2020

11. Yuji Roh, Geon Heo, Steven Euijong Whang, Senior Member, IEEE, "A Survey on Data Collection for Machine Learning" arXiv:1811.03402v2 [cs.LG] 12 Aug 2019

If this project interests you, run a node now and get rewarded for being and early adopter.

Earn KOI, and help us preserve human knowledge.

You can make a difference.