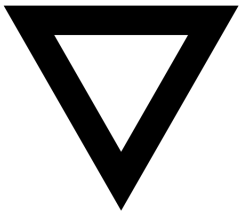

Whitepaper

{Nick Zhu}@tethys.ai
Last Updated: June 6th, 2018
Version 2.1

Abstract

Our dependence on the internet has grown exponentially over the last three decades. Google alone processes 4 billion web searches per day. However, with competing standards and haphazard leaps in technology it is a far cry from the efficient information sharing network it's early developers envisioned. Moreover, the need to efficiently search the internet for semantic information has been growing exponentially along with rate of data generation.

As Artificial Intelligence becomes ubiquitous with our internet behavior, we will need to create tools that facilitate the transition, and enable humans to interact with computers and the internet in a natural way. Simultaneously, Blockchain technology enables the creation of a tamper-proof, time-proof, and decentralized database ideal elements for the semantic web. Simply put, if we asked our search engine "Which bank offers the best small business loan rates?" it should be able to understand its meaning, compare various service, and give us a list of prices by various providers. Semantics is about understanding not only words, but the contextual meaning behind them.

This paper introduces our research and development into the creation of the first successful semantic web engine and marketplace for AI specializing in gathering and training semantic web data through Blockchain and Virtual Machines.

Table of Contents

The New Internet **4**
- Introduction 4
- Development 4
- How websites are made 5
- How we find websites 6
- Limitations of the current web 8
  - Text based, context unaware 8
  - Untracked Changes 8
  - Unverified Information 9
- How Semantic Web solves these problems 10
- Implementation of Semantic Web 10

The Tethys Network **11**
- Overview 11
- The Tethys Foundation 12
  - Introduction 12
  - Mission 12
  - Responsibility and long term development 12
- The Future and the Semantic Web 12
- Artificial Intelligence and the Semantic Web 13
- Why Blockchain? 13

Tethys Marketplace **14**
- Introduction 14
- The Ecosystem 14
- The Token 16
- Agents and Incentives 17
  - Contract Issuer 17
  - Full Nodes 17
  - VM Operators 17
  - End User 17

Technology **18**
- Blockchain Design 18
  - Split Ledger 19

High Throughput 19
Multi-Channel Architecture 20
Contracts and Consensus 21
Reputation 23
Virtual Machines 24
ML Models 24

Launching Partner: Yroo 25
Point System 26

Token Sale 26
Token Minting and Distribution 26
Discount 27

Opportunities 27
E-Commerce 27
Self-Service 28
Online Training 28
Auto Industry 29

Potential Problems 30
Blockchain limitations 30
Blockchain size and throughput 30
Data Storage and bandwidth 31

Roadmap 31
Traction 31
Future Plans and challenges 32

Conclusion 33

Team 33
Core Team 33
Advisors 33

The New Internet

Introduction

The World Wide Web was invented by English scientist Tim Berners-Lee in 1989, it was created as a side project for the world's most ambitious experiment: CERN. Its original purpose was to enable scientists to send large amounts of data from experimental results, through a network of computers. As we all know, the invention outgrew its purpose and it is now an integral part of society.

20 years after its creation, the internet demonstrated its value to society. And in 2001, its author Tim Berners-Lee came forward and revealed his idea of a new internet. The new idea was to design the internet in such a way that computers could understand context and meaning. And therefore enable humans to interact with machines like we do with each other, through actions and meaning. His idea was the Semantic Web:

"The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users." [1]

In short, the Semantic Web is a network of actionable information derived from data, with "meaning" through interpretation by a semantic system. It has been 15 Years since Tim Berners-Lee described the Semantic Web, this paper will describe the challenges and advances in technology that enable it to be a reality today.

Development

Early attempts on the semantic web have focused primarily on tagging through eXtensible Markup Language (XML), Resource Description Framework (RDF), and more recently with W3C Microdata standard, however, tagging on web-scale has proven to be challenging in practice.

University of Mannheim's director of Computer Science, Christian Bizer, observed in his quantitative analysis on semantic web that microformat data, excluding organizational contact information, has **less than 5% coverage**.[2] More recent research has shown promising new

[1] Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 34-43.
[2] Bizer, Christian, et al. "Deployment of rdfa, microdata, and microformats on the web–a quantitative analysis." *International Semantic Web Conference*. Springer, Berlin, Heidelberg, 2013.

paradigms in semantic information retrieval with both noisy web textual content[3] as well as imagery content[4]. As similarly stated in Semantic Web Revisited[5], we also believe the future of the Semantic Web is in developing and understand distributed information systems, systems of humans and machines, operating on a global scale with AI playing a central contributing role.

On the other hand, since its inception blockchain technology has been widely seen as transformative in its most natural application - value transfer using decentralized digital currency. However, it has incredible transformative power in information science as well. A general purpose blockchain data structure that is capable of recording arbitrary chronological state transition is an excellent tamper resistant choice for providing a long-term memory for the World Wide Web. This ability to record and recall information with point-in-time accuracy is a crucial component of a Semantic Web since a large percentage of semantic information are time sensitive for example product stock level and hotel price.

How websites are made

In order to understand the benefits of a semantic approach to retrieving information from the web, we must first describe how we currently browse the internet. As it is expected, there are two types of agents that interact with files and websites on the internet: People and Machines.

When an user loads a website, the browser loads a series of files that are presented to the user as text, audio, video, and other media. The data displayed on a website is **human-readable** information. However, a computer has to be able to know how to render, load, and display to you that data. This meta data is known as **machine-readable** information.

Websites follow the World Wide Web are written in the Hypertext Markup Language (HTML) convention, which combines machine and human readable data. HTML is the blueprint of a website, and it describes to the browser what type of information to render on the website, the behavior of that information, and lastly it includes keywords and metadata that enable search engines to categorize the type of information that is in that website.

For example, if someone created a website about tuna, the HTML could look like this:

```
<meta name = "description" content =  "tuna, fish, animals"/>
```

[3] Zhang, Xiang, Junbo Zhao, and Yann LeCun. "Character-level convolutional networks for text classification." *Advances in neural information processing systems*. 2015.
[4] Huang, Changqin, et al. "Large-scale semantic web image retrieval using bimodal deep learning techniques." *Information Sciences* 430 (2018): 331-348.
[5] Shadbolt, Nigel, Tim Berners-Lee, and Wendy Hall. "The semantic web revisited." *IEEE intelligent systems* 21.3 (2006): 96-101.

In this case, the programmer would have to specifically describe the keyword “description” in the metadata, in order for the browser to interpret that the information inside of the content would be a description of that website. While the above example is purposely simple it demonstrates the problem well - unless websites specifically optimize information for crawlers it is very hard for them to understand what information is being presented.

This method is both a primitive and inefficient way to extract meaning from a website. Browsers do not inherently understand what you are looking at and neither do web crawlers.

How we find websites

Of course, in order to even load a website, we have to first find it. Typically by utilizing a search engine. Search engines inherit the same limitations of HTML, and also add their own, including the fact that their architecture is centralized by nature.

The following diagram illustrates how web crawling is conducted in a centralized architecture.

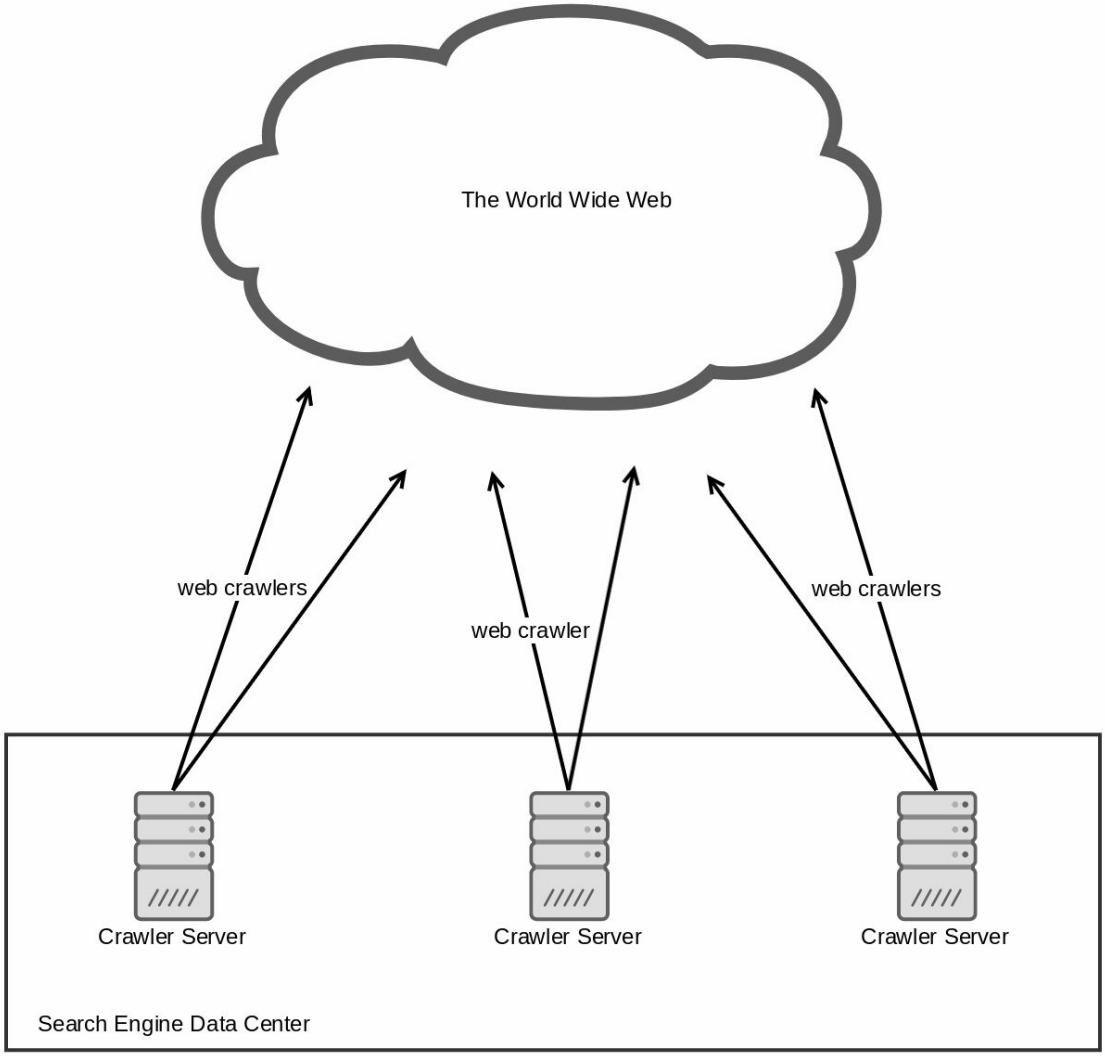

Figure 1: *Centralized web crawling architecture deployed by the search engines*

The search engines send out web crawlers, specialized bot agent programmed to systematically visit the World Wide Web, from servers operated within their data centers. Obviously, this illustration is much simplified, a large search engine will deploy dozens or more types of specialized web crawlers for different kinds of assets or even different verticals across many geographic regions from multiple data centers, however, the centralized architectural theme remains unchanged. This centralized approach works well with semi-static and non-semantic information in the shallow web yet facing severe challenges when dealing temporal semantic web information.

Limitations of the current web

Text based, context unaware

The web content of today is a lot more semantic than just merely textual. A number on a particular web page might be a price, discount rate, mortgage rate or payback period for example. The only way to discover what it means requires a comprehensive understanding of the context and the semantic meaning of other critical elements on the page. Current crawling infrastructures lack this ability. Taking a specific product as an example, this is what a search engine displays on its results page:

Figure 2: *An example of the results currently given by search engines*

In this example, other than the product ratings, none of the important semantic information of a product (price, discount, color, image, availability, shipping cost, etc,) is available from the search engine. Additionally, if we examine another example, mortgage rate, the result is equally textual centric and lack of semantics.

Figure 3: *Clear example of data formatting issues from naive web crawling*

Web users today are trained to ignore this incomprehensible string of numbers. However, this paper proposes a better, semantic, context aware search mechanism. In the case of the examples above, users should be presented with information like this:

Untracked Changes

The centralized crawling infrastructure used by search engines today is ideally suited to index static information. However, for time-sensitive information, this level of crawling frequency becomes unacceptable. Data on insurance rates, product prices, flight tickets, and many other verticals require frequent updates to be valuable. Today's search engines also lack the ability to preserve and retrieve historical information on a subject. This historical information could be invaluable for consumers - knowing the past price history on a particular product or past mortgage rate from a bank could be instrumental in decision making. The inability to store and access such historical information severely limits the usefulness of the search engines and the web at large.

Unverified Information

Search engines today try to establish truth through a single observation, which doesn't take into account websites malfunctioning. Some websites even use this to their advantage - by showing crawlers a lower price they get potential customers to navigate to their page and then charge a higher price for their product. Current search engines are unable to detect any misinformation in such case. As an example, the following is a flight results from Google Flights:

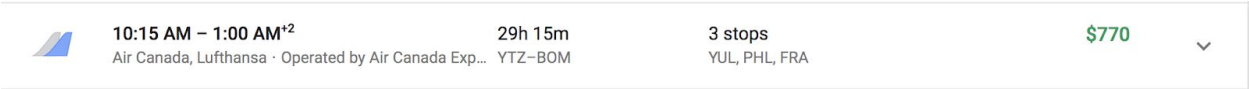

Figure 4: *Inaccuracies in search engines*

The displayed price for this itinerary is $770. However, once we click the result, a different result is shown:

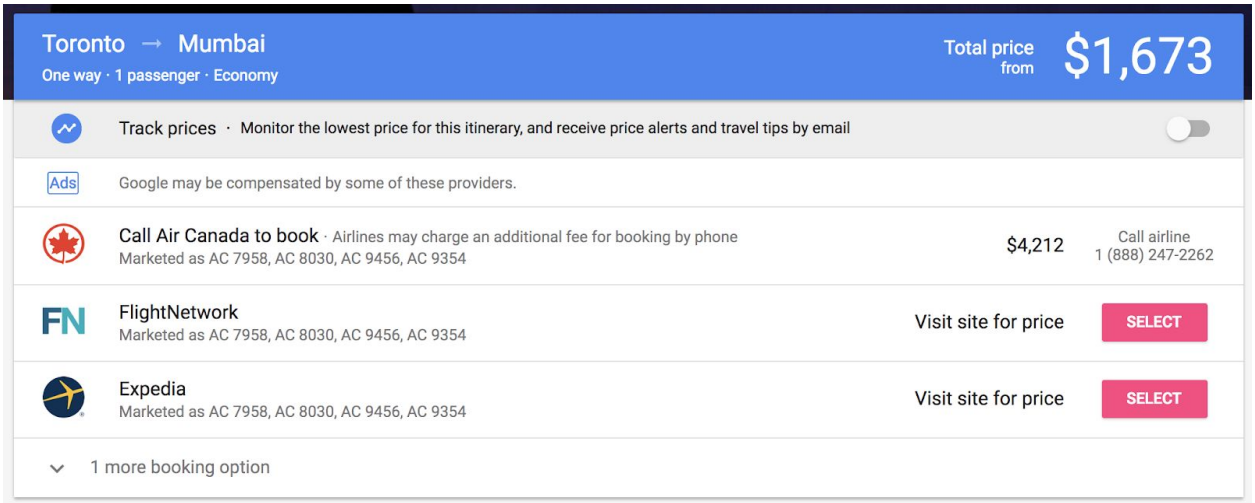

Figure 5: *Inaccuracies in search engines, continued*

In fact, the customer is shown two different prices over here, both of which are much higher than the price originally quoted. This illustrates the problem - internet users are trained to place a huge amount of trust on these information middlemen despite the fact that this information is unverified.

How Semantic Web solves these problems

The Semantic Web can solve all of these problems by introducing the following properties:

- **Context awareness:** by design, semantic web enables its users to request information by taking a context-based approach. This means that asking a question or requesting an action will take into consideration the users intentions, desires, and the quality of the expected outcome.
- **Verifiable information**: The utilization of Blockchain enables the creation of a verifiable database that goes back in time since the inception of the first block. By parallelizing this database we can have multiple Blockchains for different sites and relationships that are both faster to load than a single Blockchain, and much more verifiable than centralized architectures through a single and agreed-upon history
- **Accuracy:** Semantic information generates a series of relationships between data. By doing so, it builds a strong and accurate representation of reality with the web. The semantic web can give **honest** and more **accurate** results when an agent requests information, and therefore better representations of reality

Implementation of Semantic Web

In order to achieve truly semantic web, there has to be a massive amount of data mining and metadata written into the current web. This can only be achieved practically by creating monetary incentives for players to do the required work and activities to index, categorize, and write metadata.

Fortunately, the value of this activity is immense for a multitude of players. By creating an ecosystem that rewards users for crawling the web and extracting data, we can give the semantic web a realistic chance of being globally implemented.

The creation of such an ecosystem, would also enable the participation of semantic-based Artificial Intelligence, which will also bring a high degree of value back to users, in the form of tools and products that will improve our ability to interact with computers seamlessly.

The Tethys Network

Overview

The Tethys Network is the culmination of our research on semantic web, and the development of the first decentralized semantic crawling architecture. In order to create the new semantic web it is imperative to crawl the internet utilizing semantic searches and data storage, as less than 5% of the current internet has been searched semantically.

Tethys' decentralized crawling architecture leverages a unique blend of blockchain, artificial intelligence, and human intelligence to solve the throughput, temporal, and depth of extraction challenges that the centralized crawling architecture is facing today. With this in mind, the Tethys Network is designed to index and record captured semantic information in a temporal state transition format, using a special purpose, high-throughput blockchain implementation. Finally, the network will supply a native currency, the Tethys Token, which will become the sole currency used in its ecosystem for contract payment and to incentivize end-users' participation.

The Tethys Network introduces four fundamental concepts and innovations:

- **The Tethys network** is a permissioned partitioned blockchain implementation utilizing Practical Byzantine Fault Tolerance (PBFT) to achieve very high transaction throughput required for web-scale information collection.
- **A non-turing complete scripting language** along with open VM design specifications to allow for consumer devices to join the network as data collectors.
- **Pre-built deep learning models** designed and trained to make classification, recognition, and extraction of semantic information in both textual and image form relatively easy, with unstructured noisy web content.
- **A statistical consensus model** based on reputation that is capable of establishing statistical truth through inductive reasoning for information collected, thus rewarding users accordingly.

The Tethys Foundation

Introduction

The Tethys Foundation is a non for profit organization created by a group of individuals specialized in Artificial Intelligence, Commerce, Data Science, and Web crawling. It is clear that the Internet is one of humanity's most powerful inventions. However, the internet was not designed to grow at the level it has, and it is therefore not optimized to be simple to use and interact with the real world.

Mission

Our mission is to bring the semantic web to the world, in order to make people connect with the internet in a completely seamless way that is both natural, and simple. In order to achieve this goal, we have created an ecosystem that combines Artificial Intelligence, Semantic web crawlers, and agents that are rewarded for creating this ecosystem, and interacting with its players. A platform that combines human and machine intelligence to improve significantly over the centralized shallow web crawling methods employed today.

Responsibility and long term development

Some of the activities that the foundation will be in charge of include:

- The creation of the Tethys Marketplace
- Providing technical architecture, APIs, support, key partnerships, and open source software that will enable the proliferation of the semantic web
- Growing the network of partners, nodes, and miners

The Future and the Semantic Web

Besides the current limitations of the web, the future of the semantic web offers a series of potential benefits that will bring incredible performance in human-machine applications. Here we explore some of the potential applications for the future of the internet with the utilization of the Semantic Web and Blockchain technology.

Contextual Actions

With a fully-semantic web, users could interact with the internet in almost the same way that we interact with other humans.

Hiring Crawlers

The reality is that in order to create a semantic web, it is necessary to re-crawl a majority of the web, and therefore it is necessary to incentivize agents to do so. For example, if a large Bank wanted to gather information on the credit card interest rate of other Banks, it could choose to pay crawlers to look this information up, and then pay users to analyze it or train datasets with it.

Training Datasets
One of the most important parts about developing an AI system, is the quality of the labeled data you use to train your algorithms. The Tethys platform will enable organizations to pay to have humans labeling data in order to create accurate and honest datasets. Essentially enabling a hybrid model that combines machine-readable and human-readable data into one semantic network.

Artificial Intelligence and the Semantic Web

The semantic web offers to bridge the communication gap between humans and machines. Consequentially this would enable humans to request actions and information from machines and databases in a natural way.

The Tethys network simplifies the process to produce AI for semantic web in two different ways:

1. By creating a semantic database, incentivized by enabling agents to pay devices to crawl the semantic web, and individuals to train datasets in semantic web
2. Through a marketplace of AI that any individual or organization can tap into, and which enables its participants to offer services in exchange of the Tethys (TETH) token

The Tethys platform will provide the necessary tools for the formation of an ecosystem of semantic web that enables people to pay for services and data gathered by both humans and machines, culminating in semantic AI.

More information can be found on our Marketplace section.

Why Blockchain?

While the semantic web is all about linked data, Blockchain is all about **immutable**, and **transparent** data.

At its most essential level, Blockchain technology is a shared database in a network. In order to add to this database, it is required for the participants of the network to agree to what will be changed, which makes it completely **transparent**. Furthermore Blockchain technology keeps track of every change in the database since its inception, this is called **immutability**.

It is the combination of these properties, transparency and immutability, that make Blockchain perfect for the semantic web. In short, it enables the participants of the Tethys network to create applications that are guaranteed to be honest and accurate.

For example, if I asked “What is the cheapest this product has ever been?” the answer given by the network would be completely verifiable and correct, because there is no opportunity to ever change the answer to this question, since it has been answered before.

This enables a series of interesting solutions that can be found on our Technology section.

Tethys Marketplace

Introduction

In order to create an efficient system that produces, distributes, and trains semantic web data, we have created the Tethys Marketplace. The marketplace’s purpose is to create a self sustained decentralized ecosystem that combines Users, Data Miners, AI Semantic Nodes, and other agents. Fundamentally, the marketplace functions as a way for participants to request certain information to be mined, or datasets to be trained. The Tethys Token (TETH) acts as the fundamental method of incentivizing the process of data exchange, mining, and dataset training for the Tethys platform.

The marketplace is at the core of the concept of the Tethys network. It is the environment that keeps the network alive by providing value to those looking to generate, store, train, or retrieve semantic information. The marketplace utilizes the Tethys token (TETH) as its currency and enables transaction and smart contract development on the network.

The Ecosystem

At its most fundamental level, the ecosystem consists of individuals, organizations, and AI systems. This enables organizations to request users to mine or train datasets, and the AI systems to be formed a result of the training.

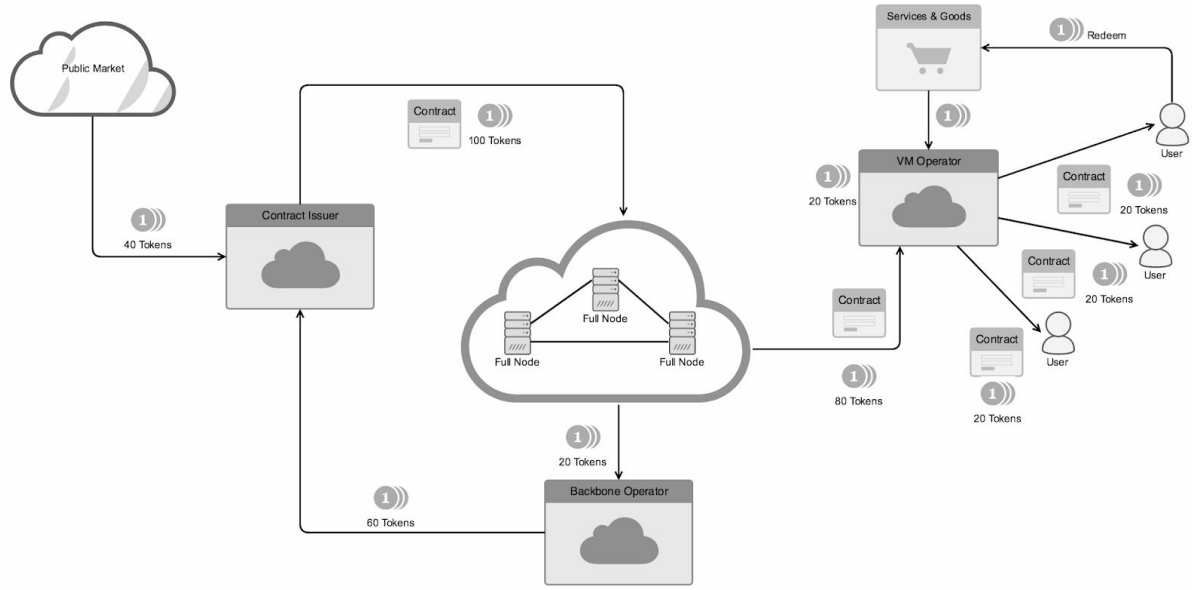

Figure 7: *Tethys token ecosystem*

This illustration demonstrates how contracts and tokens flow through Tethys ecosystem.

The ecosystem has the following players:

- **Users:** These are individuals that interact with the Tethys network. This could be any person with a device connected to the internet. Users can be paid tokens by VM operators in exchange of their work training datasets.
- **The Public Market**: This refers to the supply and demand of products and services that can be offered in the tethys network. The public market can request a certain job to be created, for example utilizing a specific AI algorithm, and creating a smart contract to deploy it
- **Nodes:** Nodes are the networked devices that hold the Tethys network. In short, the network **is** the nodes running the Tethys software and achieving consensus
- **Contract Issuer**: This means the creator of the smart contract that would facilitate a transaction between the public market, and other nodes and operators
- **VM Operators**: The VM operators are devices that are running the smart contracts and are paid tokens for the power required to do so
- **Partners:** Partners refers to companies that develop AI solutions for the Tethys network and are therefore rewarded for helping develop the ecosystem
- **Tethys Foundation:** The job of the Tethys foundation is to provide a service layer and to create a self sustainable ecosystem, including: creating the overall architecture, finding partners, incentivizing users, forming relationships, and fixing software along with the community bounty system

- **Bounty System:** Bounties are generated as a way to incentivize the community for helping with the identification and solution to bugs, problems on the network, and possible attacks

The Token

All of the relationships between each agent and user are mediated through the use of the ERC-20 compliant Tethys Token (TETH). The token serves multiple purposes, as an economic incentive to run the network, as well as creating a fair compensation and payment structure for everyone that creates value for the Tethys network. Furthermore the utilization of the ERC-20 standard enables a transparent and proven technology that benefits from software and hardware that is compliant with it.

Utilizing a fiat token for sending small amounts of money is extremely expensive, the Tethys token is both fundamental to the network, and also the simplest way of having a unified monetary system. Furthermore users can **earn** tokens by doing activities including:

- Crawling the web
- Training datasets
- Storing data
- Participating in our bounty program
- Becoming a partner by creating valuable semantic algorithms

The TETH tokens are necessary to interact with any part of the ecosystem. However, there are obvious limitations to utilizing a cryptocurrency, and less technically-savvy users should not be punished by the barrier of entry associated with having a cryptocurrency wallet. It is for this reason that we have also created a Point System, unique to those users that do not want or are able to utilize a cryptocurrency wallet. This system is however a more centralized approach, and therefore not recommended if the user has the technical ability to formalize the use of our cryptocurrency.

Details on the minting, distribution, and utilization of the token can be found under the "Token Creation and Distribution" section of this document.

Agents and Incentives

Contract Issuer

The contract issuers are Tethys' primary customers and represent the demand side of the marketplace. They are businesses, academics, and researchers who want semantic data from the internet for various reasons. The Issuers create contracts using Tethys script that leverage ML models created by themselves or by others to extract semantic data, and pay the associated costs for the execution of these contracts. The contract issuers also determine what the end result of the actions on the network are - whether raw data or semantic processed data are collected, whether the same is stored or not, and where it is private or public. Issuers can choose to store their semantic data on the Tethys network itself where it will be entered into the ledger and monitored over time, creating the temporal data store of the future. Additionally, by storing their data on a public channel, they allow their customers to validate the veracity of their information.

Full Nodes

The full nodes are the backbone of Tethys. They are responsible for establishing consensus and for ordering the executed transactions. They parse Tethys script and create a pool of tasks for VM Operators. Additionally, the receive data from the end users and run the ML models (or train them), reach consensus, and store the data. In return, they receive 20% of the contract price as payment.

VM Operators

VM operators create platforms that can run on users devices and parse Tethys script in order to crawl the internet and/or receive human input for tasks. In the return, they receive 80% of the contract fee, which they can distribute to their users as per the specific agreements of their platform. At launch, Yroo will be the first VM Operator, and they will transfer 75% of their received funds directly to their users in the form of points.

End User

The end users are the chief enablers of the share-economy facilitated by Tethys network. In essence, end users can place unused computing power and bandwidth to provide the decentralized web crawling architecture. Along with their human intelligence, they provide these resources on public market in exchange for services or goods as dictated by their agreements with the VM operators.

Technology

In this paper, we propose the Tethys Network, the first practical and functionally rich decentralized semantic information collection network that addresses all the issues mentioned above. Tethys consists of the following fundamental concepts and building blocks

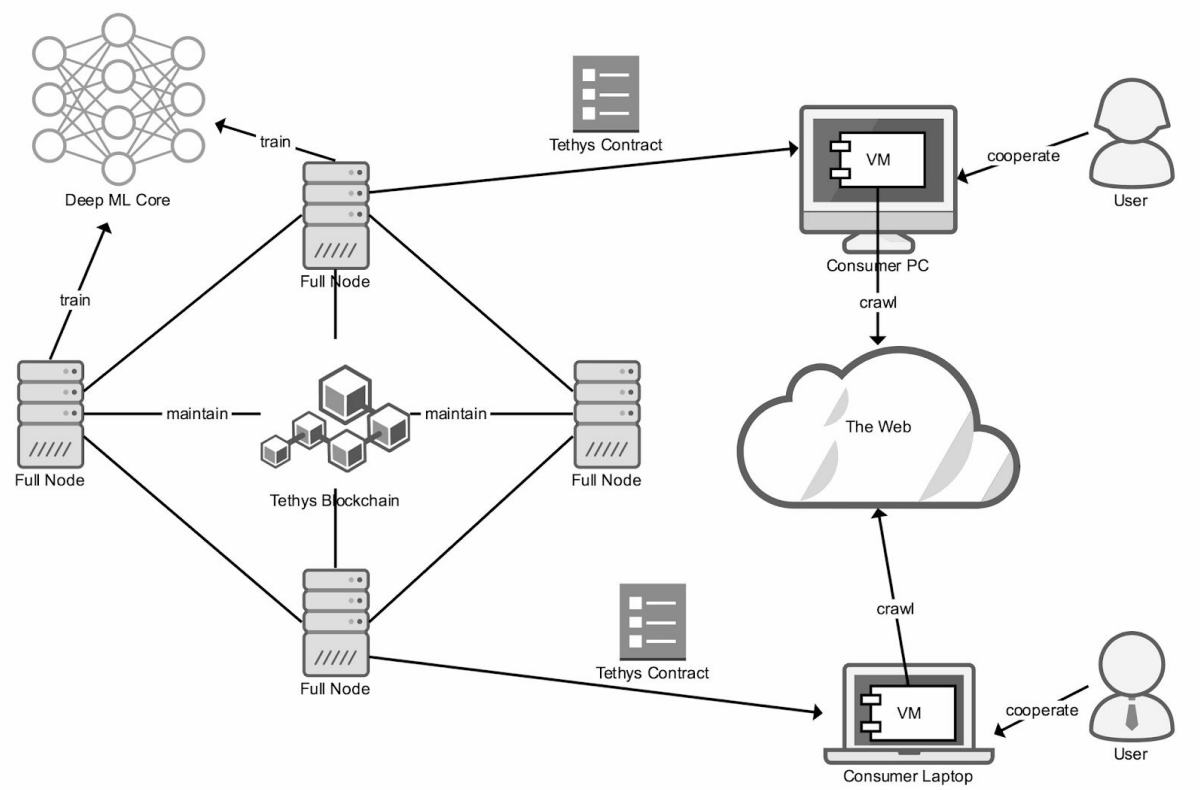

Figure 8: *Technical Architecture of the Tethys platform*

Blockchain Design

Tethys blockchain is a sequenced, tamper-resistant, and temporal record of all state transitions of semantic data It is built on Hyperledger Fabric, using Byzantine Fault Tolerance[6]. Some interesting features of Tethys blockchain are explained below

[6] Castro, M., & Liskov, B. (n.d.). Practical Byzantine Fault Tolerance. *Proceedings International Conference on Dependable Systems and Networks*. doi:10.1109/dsn.2001.941437

Split Ledger

Each channel on the blockchain has its own ledger. Each peer (full node) maintains a copy of the ledger for every channel it is a part of. Every ledger is split into two data structures - the **transaction log** and the **world state**.

The **transaction log** is a blockchain consisting of hash-linked blocks, where each block consists of a set of transactions. Once a block is added to the chain it is immutable and the chain itself is tamper proof, since the hash of the current block represents every transaction before it. The actual information stored in the blocks is a set of key-value pairs, representing the semantic value of a piece of information at a particular time.

While the transaction log represents an immutable record of all values at various times, consumers are most often interested in the value of given key at the current instance. In order to speed up the lookup time, the blockchain maintains a world state. The **world state** maintains the latest value for every piece of information ever stored on the channel. This state database is basically an indexed view of the transaction - it is possible, in fact, the build an accurate world state from scratch by parsing through the transaction log.

High Throughput

Tethys blockchain has to offer scalable throughput to handle information collection and index at web scale. We designed the Tethys blockchain to be a permissioned blockchain meaning full nodes are by invitation only. However, permissioned blockchain architecture as mentioned previously alone is not enough to offer unlimited scalability in throughput. Unlike most currently popular blockchains, Fabric, and by extension Tethys, separates the Execution, Ordering, and Validation steps of each transaction.[7]

In fact, separate nodes handle the ordering of transactions. What this means in practical terms is that peers can process multiple transactions simultaneously, and then pass on the results to the ordering service. Lightweight client nodes are always going to be processing nodes, capable of submitting queries and processing transactions. Consensus and authorization are handled by the orderer's, and all processors have to do is establish trust with an ordering service and process the tasks - leaving the full nodes to order and validate the transactions.

[7] "Transaction Flow - Hyperledger Fabric - Read the Docs." http://hyperledger-fabric.readthedocs.io/en/release-1.0/txflow.html. Accessed 7 Jun. 2018.

This works due to the fact that Tethys blockchain is designed to store chronological state transition instead of transaction ledgers. For financial ledgers, strong sequence and consistent guaranteed is required. For example in a case where A transfers 10 coins to B, then B transfers 10 coins to C these two transactions have to be recorded and verified consistently in the right order regardless who is mining the block or which block eventually gets to be appended to the chain. However, in Tethys' case state transitions can tolerate temporary inconsistency, in other words, we believe temporarily it is acceptable if certain transitions is not immediately consistent or visible from a particular partition.

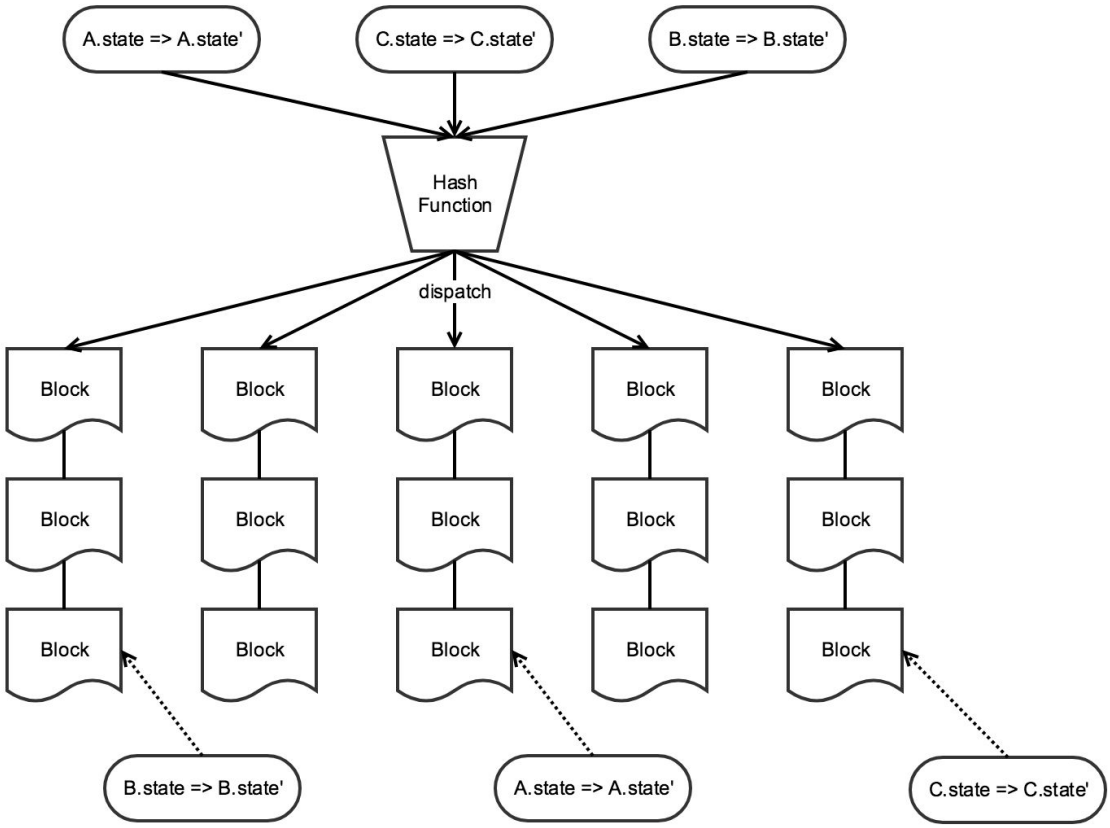

Figure 9: *Tethys parallel multi-chain blockchain architecture*

Multi-Channel Architecture

Every organization that stores data or ML models on the Tethys blockchain does so on it's own segregated channel. Tethys supports both public and private channels. Data stored on a public channel can be queried by any client while private channel data is only accessible by the authorized clients. Programmatically, each channel is it's own namespace - data and models can be imported from these channels as long as the user has the requisite permissions. On the blockchain, each channel has its own ordering service which is responsible for authorization.

This split allows users to compare data from various models or organizations in the case of public channels, and allows organizations to protect their data from competitors or share with specific collaborators using private channels. Querying publicly available information is also rate limited - the goal here is to provide information to consumers or researchers but prevent competitors from stealing an organizations competitive advantage.

Contracts and Consensus

Tethys Contracts are task blueprints written in **Tethys Script** language. The network supports two different kinds of contracts - automated and manual. Automated contracts can be executed by peers and VM's without human intervention, while manual contracts require explicit human participation. Each contract has a specific amount of tokens associated as a reward to incentivize end-user participation, and specifies what data is to be collected, which models are to be run, and how the data is to be stored. Contracts are written and entered into a pool, where they can be picked up by VM operators. This creates an open market that rewards consumers based on complexity of tasks performed.

Tethys Script is a specialized programming language designed to automate web crawling tasks. Script interpretation is conducted in a sandbox environment with strong runtime safety policies to make sure the execution is safe on end-user's device. We have also intentionally designed the language to be Turing incomplete with a set of simple and unsurprising constructs for added safety.[8] Finally being Turing incomplete Tethys VM environment is capable of providing guarantees with bounded memory and computation time.

Contract creators are able to determine what consensus means to them. The following diagram illustrates at a high level how statistical truth is established with Tethys consensus layer and quorum system.

[8] Pike, Lee. "Hints for High-Assurance Cyber-Physical System Design." *Cybersecurity Development (SecDev), IEEE*. IEEE, 2016.

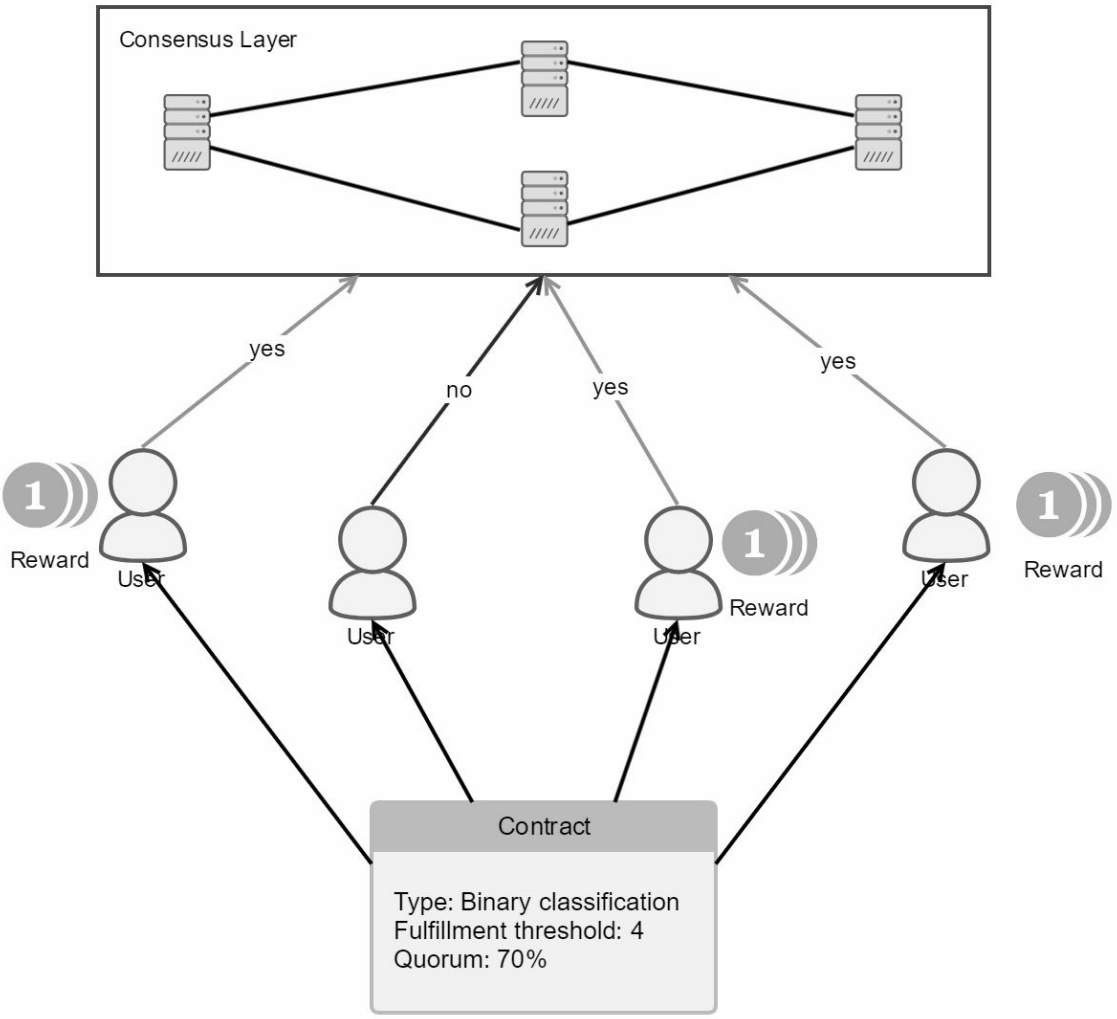

Figure 6: *Tethys Consensus Protocol*

Let's go over a few essential points worth highlighting in this simple example.

1. **Built-in Tethys Contract Support**. Tethys contracts have built-in constructs to allow the issuer to specify consensus parameters - specifically, the fulfillment threshold, quorum threshold, and reputation threshold. The fulfillment threshold is the minimum number of successful executions by edge nodes required to fulfill the contract.The quorum threshold is minimum percentage of nodes that have to agree on a certain answer being true. The reputation threshold (described in more detail below) takes into account past performance of the workers. In this example, this particular contract is considered successfully fulfilled since four nodes executed the contract and a quorum of 75% has reached

consensus. For more complex contracts that generate multiple results, an iterative consensus process is required to go through each result in sequence.

2. **Consensus-based Bounty System.** As illustrated by this example only results that conform to the consensus are counted as valid and receive the reward.
3. **Cryptographic Checksum**. Each Tethys contract is embedded with a unique cryptographic puzzle. Every node needs to solve the cryptographic puzzle and append the answer to the contract result before it can submit the result back to the chain. Since verifying the puzzle solution is computationally trivial the Tethys consensus layer can perform a preliminary checksum verification before processing the contract result, hence protecting the network from DDoS attack.

Higher thresholds result in a higher confidence level and a higher execution cost. This allows contract issuers to control how much statistical verification they require for various tasks.

Reputation

The informational consensus protocol built into the contracts described above is not sufficient in guarding the network against misinformation. Specifically, if a malicious actor on the network were to specifically try to spread information and had control of majority of the workers (like a Sybil attack), it would be possible for them to subvert the system. To defend against this kind of attack, the following three strategies are proposed:

- **Reputation tracking:** Tethys dedicates a public channel on the blockchain to track and record every participating node's reputation score based on their past performance, taking into account number of contracts fulfilled, number of correct results produced, etc. With this global reputation repository, Tethys backbone can allow contract creators to set minimum reputation scores for tasks based on sensitivity.
- **Honeypots**: Tracking reputations still has a fundamental problem - on any given task, the system assumes that consensus is equivalent to truth. However, in the case of malicious actors, this would not be true. In order to prevent this outcome, Tethys will occasionally bury honeypots into contracts. These are tasks for which the correct outcome is known - information is made available on websites administered by Tethys. A harsh penalty is imposed on workers that fail these tests - and their results are voided.
- **Enforced waiting period:** One of the avenues that attackers could use to spread misinformation is to create many workers that bombard the system with false answers. In order to make this economically expensive for attackers', all new workers in the system will have a waiting period during which they have to be connected to the network before

they can start participating. These times will be variable, and will be dependant on risk factors - geography, VM operators, etc.

Virtual Machines

Tethys Virtual Machines (VM) are software agents deployed on end-users' computing device, e.g. desktop, browser, mobile phone, etc. Tethys VM's are execution environments for Tethys Script, as well as lightweight clients to the blockchain, converting Tethys code to chaincode understandable by the Fabric-based blockchain.

Open Standard is at the core of its design. Tethys' VM protocol and specification will be published publicly and can be implemented by any general purpose programming language in various embedded environments for example browser extension, mobile app, desktop software or even directly on a web page. Open source reference implementations will be made available in Phase I. The goal is to have many different operators creating VM's that can be run on various environments, optimised for different tasks. The VM operators are ultimately responsible to selecting contracts to execute from the available pool, and for transferring value to the end users of the ecosystem.

These Virtual Machines are responsible for being the interface between Tethys blockchain and the internet. While executing contracts they interact and collect data from the internet or from humans. They pass this data back to the full nodes that run ML models convert that raw information into semantic data.

ML Models

Tethys aims to maintain a collection of machine learning models built and tuned to handle context-aware semantic information extraction from unstructured noisy web data. The initial release of ML Core will include the following models built by launching partner Yroo:

- **Classification Model**: specifically designed to answer a binary question whether a particular page fits a specific type. A sample prediction could answer whether a given page is about a mortgage product or a credit card from a bank.
- **Entity Resolution Model**: designed to leverage syntactic and semantic context information to cluster identical elements together. A typical prediction of this kind of model can answer is whether a given number on a page is about a price of a product for example.

- **Object Detection Model**: convolutional neural network designed to segment images and classifies regions on a given image. Predictions made by this model typically focuses on recognizing a given image or a region of a given image. For example, this model can answer the question if a given image is about a credit card or a laptop.

Businesses can develop and train sub-models based on the core models generated with the crawled data generated by Tethys' network to handle specific niche and vertical. For instance, a business can leverage the classification model to develop a sub-model that identifies whether a web page is about a small business loan product. Additionally, future releases will also allow businesses to bring their own models to the network so proprietary machine learning models can be leveraged.

Apart from using existing models, Tethys Network also makes it possible for organizations to pay users to do analysis on human-readable data. This would work by presenting the user with information and asking them to take an action on this information, akin to how Google's Captcha captures information by asking users to identify elements on a picture, or words and numbers after heavy distortion. The potential to label and train data would increase the rate of adoption of the semantic web exponentially, as it would create a monetary incentive for users to train datasets, and for organizations to keep improving their Machine Learning and semantic data systems.

Launching Partner: Yroo

Yroo is a big data startup specialized in shopping vertical currently based in Ireland with 11 million USD raised to date.[9] Today Yroo already indexes over 100-million products daily using direct data feed and API integration offered by e-retailers and marketplaces; however they are planning to upgrade a significant part of their pipeline with Tethys.

Many of the architectural choices and design innovations came as a result of Yroo's internal prototype - decentralized semantic crawling PoC project. As a launch partner, Yroo has committed a significant amount of resources and technical expertise in partnership with Tethys Network to launch Phase I of the implementation through their browser extension and mobile app environment.

The critical objective for Yroo in this partnership is the ability to automatically verify information through Tethys contracts, as well as the ability to control information retrieval frequency hence making sure their information is always accurate and up-to-date. Finally having

[9] Yroo, The World is Your Store. http://www.nibletz.com/events/collision/yroo

Tethys information blockchain to record and store historical prices in an immutable way allows Yroo to offer historical prices with confidence and transparency to its consumers. Yroo believes that Tethys tokens could be an powerful incentivization scheme for its users thus helping fueling Yroo's growth in user acquisition and engagement. As a result, Yroo is also planning to launch world's first online store where Tethys tokens are accepted in exchange for digital goods.

Point System

We understand that it is not practical for all users to utilize a cryptocurrency, due to the natural limitations of ERC-20 tokens, for example, if a user loses its private key, the amount in the public wallet would be lost forever. Other users could struggle to get a hold of a cryptocurrency or even to utilize it for purchasing goods in the physical world. It is for this reason that Yroo has decided to create a secondary point system, for those users that do not want to utilize Tethys token.

Yroo will be the first VM operator on Tethys network. As such, they will enable users to convert their tokens into points, and then enable users to receive loyalty points, and gift cards in order to utilize their points with physical products in the real world.

Token Sale

Token Minting and Distribution

In order to finance all the operations required to research, develop, and expand operations. The Tethys foundation will sell its ERC-20 Tethys (TETH) tokens, which are the only unit of value that is used to pay for all transactions in the network.

A total of 8 billion tokens will be minted, and no more will produce, as we do not believe in artificially changing the amount through deflation or artificial inflations. The tokens will be distributed in the following way:

- **Token Sale:** 50% of the tokens or 4 billion tokens
- **Partnership Program:** 25% will be locked for 3 months and spent only to incentivize partner organizations to join the network
- **Tethys Foundation:** 20% will be locked for 5 years, and distributed at a rate of 4% per year to pay for research and operations

- **Advisors:** 5% will be distributed to the experts that enabled the creation of our intellectual property and research

Discount

As a reward early investors, the Tethys Foundation has created an incentive and discount system. This system can be found on our website: https://tethys.ai. We have decided to keep all the information regarding discounts on our website as a way of transparently showing how much money has been raised and how much each user will receive immediately. Since the foundation has a limited amount of tokens allocated to each agent, we want to guarantee all the information regarding our token distribution is both transparent and up-to-date.

Opportunities

The range of opportunities for the semantic web is extremely wide. At its core the semantic web will enable people to easily access information in the same way that humans talk to each other and request actions and information. For this reason some of the examples detailed below are extremely marginal and small compared to the full potential of what people will be able to do in such platform. That said, the following are some potential uses of the semantic web:

E-Commerce

Global retail e-commerce sale is projected to reach 4.48 trillion US dollars in 2021.[10] About 65% of online shoppers compare prices online before making a purchase.[11] More than 55% of US online shoppers now start their search on Amazon instead of Google.[12]

Tethys can offer access to real-time verified semantic product information. It is conceivable a specialized semantic search engine with access to all e-commerce sites, and their products can be built to offer consumer instant insight on price, shipping, ratings, and all other relevant information. In fact, our launching partner Yroo is a company with precisely this mission in mind. The details about Yroo are described on the Partnership section of this paper.

---

[10] Global retail e-commerce sales 2014-2021.
https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/
[11] PEW Research Center Online Shopping & Ecommerce Survey.
http://www.pewinternet.org/2016/12/19/online-shopping-and-e-commerce/
[12] BloomReach.
https://www.bloomberg.com/news/articles/2016-09-27/more-than-50-of-shoppers-turn-first-to-amazon-in-product-search

Self-Service

A combination of the semantic web and artificial intelligence, is the ability to have conversations with search engines. Instead of asking a singular question, the semantic web could create a series of them to find the right answer.

One of the most utilized and frustrating experiences about search engines, is self-help. Simply put, sometimes a user wants to solve a specific problem in their lives and they search for something such as “What bicycle should I buy in Canada?”. Unfortunately, a search as simple as that would not bring any meaningful results. Most of the answers to this question would simply return some variation “Guide on how to buy a bicycle”, but not an answer to this question, as it is both subjective, and very specialized. So therefore search engines cannot directly continue a query of the answer, and instead are forced to display blogs and guides, and expect users to select the right one, with no knowledge to which one is the most common or best answer.

The semantic web could enable people to quickly find solutions to problems by simply asking the way they would ask a friend “What bicycle do you recommend I buy?”. An intelligent system could even start asking further questions to the users such as “What budget do you have?” or “Do you want to do mountain biking, or road biking?” So instead of looking at multiple guides, a user could be guided to the right answer by having a conversation, just how we typically find solutions between each other.

This could go much further than making selections, it could help people with questions such as “How do I fix my car?”, “How do I make crème brûlée?”, and an endless amount of other self-serving questions.

Online Training

The semantic web, combined with AI could also revolutionize the ways we learn. When we currently find information about a specific topic, we typically have to rely on searching something and then reading a series of articles on websites and encyclopedias related to that topic. The semantic web could make the process of learning information much easier, by essentially utilizing its database to quizz users, and score their answers.

For example, we could use the semantic search in reverse, and utilize the same search engine to ask a user a question, for example: “How do you boil an egg?” and enable a user to give an answer, the semantic web can then rate this answer, and improve its accuracy. It can also be used for educational purposes, where a user can take an online test and be scored.

Auto Industry

In 2017 US auto industry sold over 17 million vehicles.[13] On average consumers spent **14 hours** through the purchase journey with more than 50% of this time spent on researching and shopping online.[14] This is completely understandable for anyone who has purchased a car recently. First of all car buying is a major purchase for most people and the potential saving could easily amount to hundreds or even thousands of dollars.

To make matter worse, the amount of information and variables to consider for a car nowadays is incredibly large. The following are some of the major categories of variables the consumers need to consider:

- Maker
- Model
- Trim
- Packages
- Financing
- Etc.

Finally to make this research project even more challenging is that auto dealers and automakers run different promotions and special offers in different geographic regions for different periods. The main reason why this research process is so time-consuming is due to the fact that all information above buried in the deep web.

The consumers need to manually go to different automaker's website to find out the exact quote for the desired configuration while knowing that the quote might change in a matter of weeks or even days. Facing this kind of time-consuming challenge a specialized search engine or comparison engine, leveraging Tethys Network's unique capabilities, can aggregate all semantic information from the deep web, e.g. from different automakers, dealers, and aggregators website simplifying this research process significantly. Additionally this auto comparison engine will also be capable of recording all of these variable in a chronological order using Tethys blockchain providing an extra dimension of information to consumers that is not currently available; finally, answering the proverbial question, once and for all, "When is the best time to buy a car?" with statistics and hard data.

---

[13]Vehicle sales in the United States 1977-2017,
https://www.statista.com/statistics/199983/us-vehicle-sales-since-1951/
[14] AutoTrader 2016 Car Buyer Journey Report.
https://b2b.autotrader.com//agame/pdf/2016-car-buyer-journey.pdf

Potential Problems

Blockchain limitations

Blockchain size and throughput

The main characteristic of a Blockchain is being an immutable database, this also means that every single transaction in the history of the network is recorded since its genesis. This is done in order to verify that every single transaction is correct and traceable to its inception.

Transaction approval size on Bitcoin is limited by its block size. Essentially there is a limit on the amount of transactions that can be stored on a block, and there is also a limit on how many blocks can be approved per unit of time, at 6 per hour, through a system called Proof of Work (PoW).

The biggest problem with Blockchain size, is that the amount of people with access to computers of increasingly large storage capabilities is small. The larger the Blockchain, the smaller the amount of people who can store its contents, and therefore the more centralized it becomes.

While Bitcoin limits this transactions through mining and hashpower, Ethereum took a different approach with Virtual Machines and instead created a bidding system where people pay to have their smart contracts processed in exchange of a currency called gas.

The overall problem is that as more transactions enter the Ethereum Network, the higher the price of gas, and also the smaller the proportion of transactions that are approved from the storage for pending transactions (mempool) and into an authorized block.

Simultaneously, as the network becomes more popular, an increasingly number of people create applications on the network, significantly increasing the size of its directory. The full unpruned size of the current Ethereum directory has reached over 1TB and it is increasing exponentially.

This provides a series of challenges that the Ethereum foundation is trying to mitigate. The more serious of which is the fact that very few people will soon be able to run full Ethereum nodes, and therefore there’s a risk of decentralization. The ethereum network is also trying to improve the transaction speed from about 15 per second to thousands to the incorporation of Proof of Stake (PoS).

If a Blockchain network becomes centralized, it would lose a lot of its fundamental positive effects, namely that an attacker could choose how the network behaves if it controls the majority of it, which could cause all kinds of negative consequences, and losing the ability to keep transparency and immutability, which are the main reasons for a Blockchain implementation.

The Tethys network will utilize ERC20 tokens for its economic system. This is an Ethereum protocol and it is therefore sensitive to the downsides of Ethereum. In the case that Ethereum is deemed unsafe, the Tethys foundation will work with its community to find a better system based on consensus.

Data Storage and bandwidth

As the Tethys network continues to grow, it will become apparent that the storage and retrieval of semantic data will steadily accumulate. There are two ways in which the movement of large amounts of data can pose a problem. The first is the storage itself, and the second is the amount of internet bandwidth that will be required to move this data.

Currently the most likely scenario is that we will decentralize this data, and incentivize those that hold it by paying them for the storage and retrieval of data. This method is similar to the one described by Filecoin. At its core, organizations and users can be pay miners for storing and retrieving information.

This is an area of ongoing research, and therefore the most updated information will be available on our technical wiki.

Roadmap

Traction

The Tethys foundation has been able to gather all the necessary parts to kickstart the product, as well as developing the infrastructure, tools, and information required to verify all the information contained in this document.

The necessary parts to launch this product include:

- **Foundation Genesis:** We have already successfully incorporated, structured, and developed the foundation. Documentation and legal structure have been completed.

- **Team Formation:** A wide variety of experts in the required disciplines has already joined the Foundation. The biographies and details can be found at the end of this document.
- **Investment and Fundraising:** An initial seed fund has been raised, this fund was used to gather the team, generate the core intellectual property, and create and distribute the network
- **Advisors and Partners:** The platform will launch with established advisors and partners in the industry. Information can be found in this document, as well as our website
- **Smart Contract Development:** Initial smart contracts have been written and deployed, more information can be found on our wiki
- **Project Wiki:** As a way to maintain transparency and show progress, we have created a wiki that details all the technical aspects of the project
- **Intellectual Property:** All of our technology, algorithms and tests can be freely available through our wiki

Future Plans and challenges

The next 3 years will be the most critical for the platform. The following explains our year-to-year plans and the challenges and solutions we expect to develop.

Year 1: The Tethys foundation will launch its platform by Q3, 2018. We already have a working prototype and the purpose of the first year is to move from prototype stage into a fully working product. To stay up to date with our technology please visit our website, https://tethys.ai/. Furthermore the foundation is looking to partner with at least 3 organizations looking producing semantic web algorithms.

Year 2: This year will be strongly focused on scaling. There are a series of well known problems in blockchain-based solutions that we acknowledge openly through this document. Storage and throughput are the biggest of those and we will focusing on tackling this, as well as increasing our partnership program. We will also be launching our user training program this year. This will enable humans to train semantic dataset through their devices and be rewarded with our token.

Year 3: At this time, the platform would focus on parallelizing ML algorithms. Essentially if a business or individual wanted a very complex action to occur, and no partner alone could solve the problem, we would be able to utilize multiple models in parallel in order to arrive to a solution. At this point we expect the platform to be self-sustainable, with a constant amount of agents creating value and exchanging the token.

Conclusion

The internet started as a scientific tool and quickly grew into an economical behemoth. The reason the internet grew to its current stage was the development of an infrastructure that enabled commerce through the structuring and sale of data.

Analogously, the Tethys Foundation has created a series of technical, social, and economical tools to make the semantic web a reality. We understand that the primary incentive for people to create the semantic web will me monetary incentives.

For this reason, our main mission is to guarantee our network to be focused on practical uses of the semantic web, and therefore, guarantee that the information collected by the network focuses on creating value to those using the network.

As a result of this platform, we will have an internet that enables people to communicate with machines in the same way that we communicate with each other, with simple words that rely on meaning and context.

Team

Core Team

Advisors