Photon: Energy Market Opportunities Map

Eduardo Ribeiro up201705421
Martim Silva up201705205
Miguel Pinto up201706156
Nuno Cardoso up201706162

April 21, 2021

Abstract

Every day, we learn about emerging technologies and developments that have the potential to impact our lives, solve real business problems and exploit new open opportunities. But how do we detect the early proof-of-concept, non-obvious opportunities with real growth potential?

This challenge led us to identify a relevant and broad topic, such as the Energy sector. Energy is one of the biggest drivers for global issues, like climate change, and when starting new projects and companies, it is important to make sure the main problems approached are relevant enough so that a potential solution can not only create a big and positive impact but also have market potential.

The research, extraction and analysis of energy related data, from the right sources, can help identifying high potential and growth opportunities, and detect the most promising, emerging and non-obvious problems that need to be solved.

1 Introduction

The **Photon** project is an application capable of extracting and analysing energy related data from various sources, in order to detect and identify real, high growth opportunities within the energy market and industry. Additionally, Photon will showcase that information to the user in an easy to use visual interface.

This article is separated into various components. After a brief abstract, the context and motivation are discussed, presenting the main issue and needs that the project wishes to solve. Moreover, the goals are identified and explained, in order to communicate what is the success criteria for the project. As this project intends to provide a real innovative solution, it is important to do a research related to the state of the art in this area, identifying similar products and related work. Because the project heavily relies on the consumption and analysis of energy-related data, the main sources and models are going to be identified and described. Finally, we will discuss the planned project calendar, with all the different phases and deadlines imposed, and present a brief conclusion to the article.

2 Context and Motivation

The project's main theme is tightly related with the detection and identification of high potential opportunities in a given market.

Our decision to focus on the energy market was based on its importance and influence in many relevant worldwide issues, like climate change and global warming, transportation, overall quality of life, among many others. It is a broad field with many subareas that can be further studied and analyzed in order to identify good opportunities for projects or companies. Due to this, it is apparent the relevance and necessity of analyzing and detecting promising opportunities within the energy field, with real potential, based on information from various sources, ranging from already existing publications and patents, to unstructured data present in social media platforms.

It is also important to develop a platform that allows the showcase and presentation of these analyses in a way that is easy to comprehend and interpret, and to develop an interactive visual interface for that purpose.

3 Goals

The main goal of the project is to understand in detail the non-obvious, early impact topics and market problems in the Energy sector and where there is a need for new solutions. The main topics that describe the success criteria for the project are the following:

- **Data Extraction Component** — Development of web crawlers and modules that communicate with external APIs and sources in order to extract relevant, energy-related data.
- **Knowledge Graph** — Parsing and interpretation of the collected data, in order to build a knowledge graph consisting of the various energy sub-areas and topics, and the connections and relations between them;
- **Visual Interface** — Creation of a visual interface, that will present the knowledge graph to the user, to have a sound source of insight into the "problem area" of Energy.

This can be described as the **Sourcing** phase of the project and application.

In a later phase of the project, that may fall out of scope with the development done within the Markup Languages and Document Processing (LAPD) subject, the following topics should be considered:

- **ML Based Prediction** — Machine Learning based screening and prediction of opportunity growth potential.

This last phase can be described as the **Screening** phase.

4 State of the art

Although there are no existing platforms that attempt to tackle the aforementioned challenges, there is still relevant research and tools that can be applied and studied.

4.1 Renewable Energy (RE) Data Explorer

User-friendly geospatial analysis tool [1] developed by the USA's National Renewable Energy Laboratory (NREL) that performs visualization and analysis of renewable energy potential that can be customized for different scenarios.

4.2 Hydro-graph

Hydro-graph [2] aims to provide researchers in the field of hydrogen research with a knowledge graph of relevant fields. It can help researchers quickly understand the development status, research hot spots and research trends of a certain field. It is very similar to the Sourcing phase of our project, but applied to a different area, as it is focused on the hydrogen research field, being able to identify relevant opportunities and trends within it.

4.3 Building a Knowledge Graph for Food, Energy, and Water Systems

A thesis in Computer Science on knowledge graphs for the Food, Energy and Water (FEW) systems, written by Mohamed Gharibi [3]. The main goal of the project was to provide better analytics for FEW systems, giving the opportunity to the domain experts to conduct data-driven research. In order to build the knowledge graph, Semantic Web technologies were employed, such as the Resource Description Framework (RDF), the Web Ontology Language (OWL) and SPARQL.

4.4 Desarquivo

Project developed by Miguel Ramalho, a FEUP alumni participating in the Prémio Arquivo.pt [4]. This application extracts news data and information from various portuguese journals, present in the Arquivo.pt records. It then aggregates and groups together related data and entities, in order to create a dense network of concepts that is then displayed in a web application through a visual interface. We can conclude this project is very similar to the Sourcing phase of Photon (which is our main focus), as it comprises its three main steps: the extraction and collection of data, the clustering and grouping of that data and the concepts and entities within it, and the presentation of the

knowledge graph results in a visual interface.

Other tools that were identified and could be helpful on the development of the project:

4.5 Dandelion API

Dandelion [5] can be described as "Semantic Text Analysis as a service". It is an API that allows to extract meaning from unstructured data and text. It has several use cases and features, like entity extraction, sentiment analysis, extraction of keywords and concepts, and more. It can prove to be a useful tool, as we will need to interpret unstructured data from various sources (social media, news, etc), and group together different concepts and entities.

5 Data Sources and Data Models

The APIs presented in this section are possible sources that can be used to extract valuable information and data, to be used in later phases of the project. Furthermore, these APIs are grouped into 3 different categories: **social media**, **news** and **patents**. All the referred APIs return information in the **JSON** format.

5.1 Twitter API

With the Twitter API [6], you can tap into the public conversation to understand what's happening, discover insights, listen for events, and more. With this API we can gather different metrics data and tweets performance, search for specific topics or keywords to analyze the related conversations, surface tweets in real-time and get recent popular searches across the Twitter platform.

5.2 Reddit API

Using Reddit API [7] it's possible to get information related to specific subreddits and search across the entire Reddit to get feeds, comments and reactions filtered by specific keywords.

5.3 Usearch API

The Usearch API [8] is a simple, easy-to-use REST API with an endpoint that returns JSON search results for current and historic news articles published by many worldwide sources. With this API we can retrieve news by keywords, phrases, country, publishers, and since it has a very diverse set of sources it will be extremely useful.

5.4 Lens API

API that provides access to a large database containing academic work and patents from around the world [9]. Like other APIs specified in this document, Lens API also allows to search desired results via a keyword or query string, allowing us to easily filter the information that is useful to us.

6 User Requirements

In order to decide and prioritize the various features and functionalities of Photon, the team has researched and discussed different User Requirements and User Stories. The items are identified by an ID and a short description of what is intended and how it adds value to the user. They are also classified using a priority scale (High, Medium, Low). Furthermore, the requirements that are necessary for the Minimum Viable Product are also identified.

The results were compiled in a table (Table 1), that can be seen in the Appendix.

7 Tech Stack & Tools

This section describes the main frameworks and tools of the tech stack used to develop this project.

7.1 React

A Javascript front-end library to build the graph-based web interface [10].

7.2 Express

NodeJS web application framework, present in the back-end of the application. Allows the creation of API endpoints that can then be accessed by the front-end to extract relevant information to output to the user [11].

7.3 Node

A variation of Node using Typescript will be used to create the Photon back-end, to make the layer that will access the database and implement the business logic of the accessible back-end endpoints [12]. Node will also be used to create the scripts that will fetch data from the specified APIs and insert it into the database.

7.4 Neo4J

Graph-based database that is able to efficiently capture and represent the different entities and nodes, as well as the data relationships between them [13].

7.5 Protégé

Used to tailor the ontology for energy domains to Photon's needs [14].

8 Architecture

Our application can be seen as a layered system, at the bottom being the modules that are related to the extraction and processing of relevant data, and at the top being the layers that directly interact with the user. From bottom to top, Photon can be described through the following layers:

8.1 Data APIs

At the bottom, we have the various third-party APIs that the system will access, that allow the extraction of energy-related data through social media posts, news articles, patents, among other sources.

8.2 Data Extractor

This module directly accesses the various data APIs and extracts useful information and data for the application, storing it on the Graph Database.

8.3 Graph Database

Represents the main database of Photon, a graph-based storage that will contain all relevant information extracted from the Data Extractor layer from the third-party APIs. To be accessed by the Controller layer.

8.4 Controller

The Controller layer is the back-end part of the Photon web application, that researchers may use to help in the identification of good, non-obvious opportunities in the Energy sector. The Controller will access the database and will provide relevant endpoints for the View layer to use.

8.5 View

The topmost layer of Photon. The View is the front-end portion of the Photon web application, and is the layer that the user directly interacts with. Among other features, it will show a 3D knowledge graph that contains all information that was extracted and processed.

A high-level diagram of the architecture can be seen in Figure 1 of the Appendix.

9 Prototype

A prototype was developed for the proposed web application. The code and instructions for installation can be found in the D2 submission zip.

10 Project Calendar

In this section we present a calendar with the planning for the project's tasks. See Figure 2 for more information.

11 Conclusions

This article serves as a proposal and introduction of a project that we view as interesting and useful, related to an area that has an enormous impact on our lives and on the planet. We were able to describe the main goals and motivation for the project, as well as collect information about similar and related applications, and identify possible data sources and models to extract information related to possible problems and opportunities in the energy department. All these topics are relevant for the project's success, which we are optimistic about. Furthermore, we identified the main User Stories and defined requirements for the Minimum Viable Product. We also described the projected architecture of the system, and the tech stack and tools that will be used to implement the application. Finally, we presented an early prototype that shows the overall look and functionalities of the project.

References

[1] Renewable Energy (RE) Data Explorer.
https://www.climatelinks.org/resources/renewable-energy-re-data-explorer

[2] Hydro-graph.
https://www.sciencedirect.com/science/article/abs/pii/S0360319919345367

[3] Building a Knowledge Graph for Food, Energy, and Water Systems.
https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/62663/Thesis_2017_Gharibi.pdf?sequence=1&isAllowed=y

[4] Desarquivo.
https://msramalho.github.io/desarquivo/

[5] Dandelion. Dandelion API.
https://dandelion.eu/#dandelion

[6] Twitter. Twitter API.
https://developer.twitter.com/en/docs/twitter-api

[7] Reddit. Reddit API.
https://www.reddit.com/wiki/api

[8] Usearch. Usearch API.
https://usearch.com/web-search-api

[9] LENS. Lens API.
https://www.lens.org/lens/user/subscriptions

[10] React.
https://reactjs.org/

[11] Express.
https://expressjs.com/

[12] Node.js.
https://nodejs.org/en/

[13] Neo4j.
https://neo4j.com/

[14] Protégé.
https://protege.stanford.edu/

Appendix

ID	Description	Priority	MVP
US01	As a user, I want to visualize and navigate through the graph, to be aware of the different opportunities.	High	X
US02	As a user, I want to have more information about a specific graph node (energy sub-area), so I can make a further analysis to a specific opportunity or area.	High	X
US03	As a user, I want to see all the attributes of the nodes and their relationships, so I can understand the connections and what makes a certain sub-area promising or not.	High	-
US04	As a user, I want to see distinct node sizes, so that I can understand where the best opportunities lie.	Medium	-
US05	As a user, I want to have information about the specific articles/posts or have some links to pertinent pages, so I can analyse in first-hand the posts and articles that were extracted for a certain sub-area or topic.	Medium	-
US06	As a user, I want to have filtering options based on certain properties, to see only relevant nodes.	Low	-

Table 1: *User requirements.*

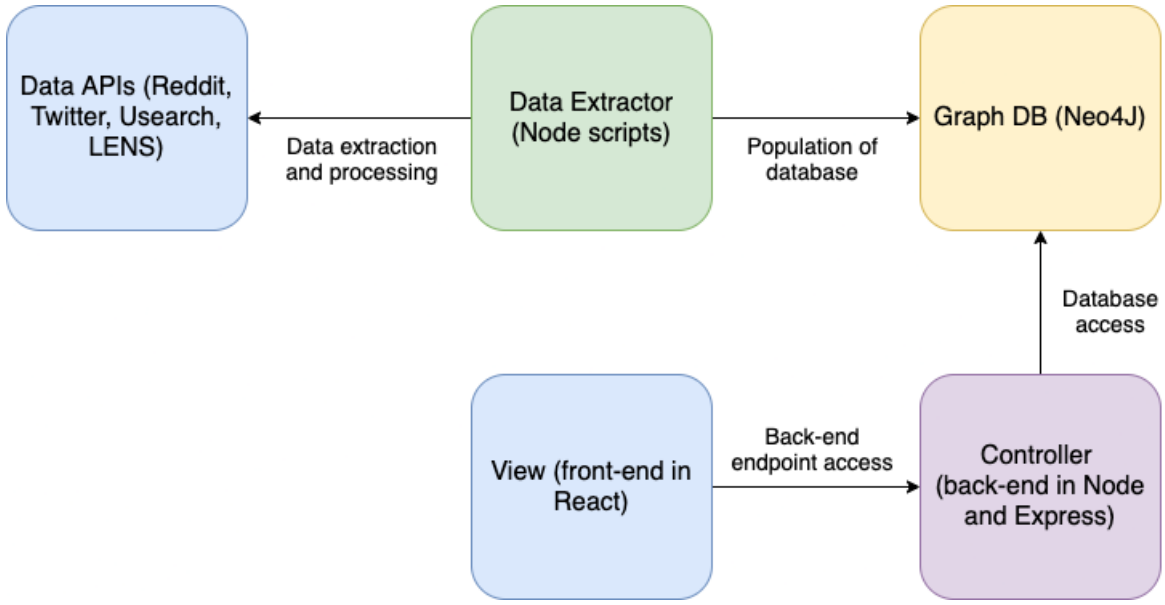

Figure 1: *Diagram describing the various modules present in a high-level view of the architecture of the system.*

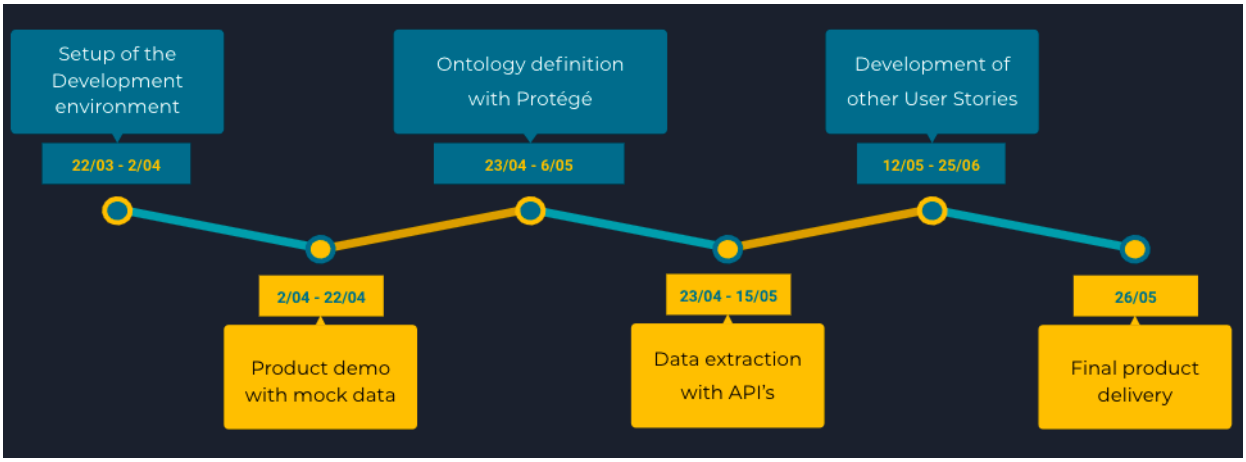

Figure 2: *Roadmap showcasing the projected calendar for the project*