Photon: Energy Market Opportunities Map

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

May 26, 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. However, how do we detect the early proof-of-concept, non-obvious opportunities with real growth potential?

This challenge led us to the identification of a relevant and broad industry sector - Energy. Energy is one of the biggest drivers for global issues, like climate change and agricultural shortage. Therefore, when starting new projects and companies, it is vital to make sure that the main problems are relevant. A potential solution can create a substantial impact and have market potential.

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

1 Introduction

Photon is an application capable of extracting and analyzing energy-related data from various sources to detect and identify real, high growth opportunities within the energy market and industry. Additionally, Photon showcases 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, and the platform's requirements to communicate and define the success criteria. As this project intends to provide an innovative solution, it is vital to research state-of-the-art art in this area, identifying similar products and related work. Given that the project heavily relies on the consumption and analysis of energy-related data, the primary sources and models will be identified and described. Finally, the concrete implementation is discussed, from the architecture and components to development details and next steps.

2 Context and Motivation

The project's central theme is tightly related to detecting and identifying high potential opportunities in a given market.

The 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 substantial subareas that can be further studied and analyzed to identify promising 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 various sources, ranging from existing news articles, publications, and patents to unstructured data present in social media platforms.

It is also essential to develop a platform that allows the showcase and presentation of these analyses to be easy to comprehend and interpret and develop an interactive visual interface.

3 Goals

The project's primary goal is to understand several non-obvious, early impact topics and market needs in the Energy sector. 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 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 Energy's "problem area".

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

In a later phase of the project, that falls 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 Existing solutions

Although no existing platforms attempt to tackle the exact challenges mentioned above, some related projects and solutions that are relevant to the subject have been explored, and listed below.

4.1 Renewable Energy (RE) Data Explorer

User-friendly geospatial analysis tool [1] developed by the USA's National Renewable Energy Laboratory (NREL) 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 particular field. It is very similar to the Sourcing phase of our project but applied to a different area. It focuses on the hydrogen research field, identifying 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 project's main goal was to provide better analytics for FEW systems, allowing 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.

5 Data Sources and Data Models

The APIs presented in this section are sources that were used to extract valuable information and data. Furthermore, these APIs are grouped into 2 different categories: **social media** and **news**. All of the referred APIs return information in the **JSON** format.

5.1 Open Energy Ontology

Open Energy Ontology OEO [9] is a domain ontology within the field of energy system modelling. With this Ontology, it was possible to define energy sectors that are the graph system's central nodes. By expanding these nodes, it is then possible to access the nodes generated with data coming from the third-party APIs.

5.2 Twitter API

The Twitter API [4] allows tapping into the public conversation to understand what is happening, discover insights, listen for events, and more. It is also possible to 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.

5.3 Reddit API

When using the Reddit API [5] it is possible to get information related to specific subreddits and search across the entire Reddit to get feeds, comments and reactions filtered by specific keywords.

5.4 Usearch API

The Usearch API [6] is a simple, easy-to-use REST API with an endpoint that returns JSON search results for current and historical news articles published by many worldwide sources. It allows the retrieval of news by keywords, phrases, country and publishers.

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.

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 compiled were the following:

- **R01:** Visualize and navigate through the graph, to be aware of the different opportunities. **(High Priority)**
- **R02:** Have information about a specific graph node (energy sub-area) to analyse a specific opportunity or area further. **(High Priority)**
- **R03:** See all of the attributes of the nodes and their relationships so that it is possible to understand the connections and what makes a specific sub-area promising or not. **(High Priority)**
- **R04:** Have filtering options based on the number of articles and posts to see only relevant nodes. **(Medium Priority)**
- **R05:** Have information about the specific articles/posts or have some links to relevant pages so that it is possible to analyse, first-hand, the posts and articles that were extracted for a certain sub-area or topic. **(Medium Priority)**
- **R06:** See distinct node sizes proportional to a sector's disruptiveness so that I can understand where the best opportunities lie. **(Low Priority)**

For the Minimum Viable Product, the planned requirements consisted of **R01**, **R02** and **R03**. All of them have been successfully implemented. Furthermore, in addition to these, the requirements **R04** and **R05** have also been accomplished, as we thought they also greatly contributed to the value that the platform offered to potential users.

7 Technologies and Tools

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

- **Node:** A Typescript variation of Node used in the backend to fetch data from the specified APIs and store it in the database.
- **React:** A Javascript frontend library to build the graph-based web interface with the ForceGraph3D library.
- **JSON:** The communication between the frontend and the backend of the application, as well as the interaction with the external

APIs, were done using the JSON format to transfer data.

- **Neo4J:** Graph-based database that efficiently represents the different entities and nodes, as well as the relationships between them. It uses Cypher, a querying language similar to SPARQL.

- **OWL:** Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. Used in the context of the ontology that provides all energy concepts and entities.

- **Protégé:** Used to tailor the OWL ontology for energy domains to Photon's needs.

8 Architecture and Development Details

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

8.1 Data Sources

Two types of data sources were used to extract information for Photon. The first one is related to the various energy sectors and concepts that the system should identify and consider. To this end, the **OEP Energy Concept Ontology** was used. This ontology presents, in a structured way, many concepts and entities related to the energy field, as well as their relationships and hierarchies. The second type of data sources is related to content on the internet that mentions or is connected to relevant energy concepts. The system performs the extraction of social media content through the **Pushshift API (allows access to Reddit data)** and the **Twitter API**. Photon also extracts information from news articles and posts, using the **Usearch API**.

8.2 Extraction Manager

This module interacts directly with the various data sources, extracting and parsing all information. It also connects with the Database Manager in order to store all data in the graph database.

8.3 Database Manager

The Database Manager acts as a middleware between other software components and the Neo4J graph database. It stores all energy concepts from the parsed ontology and all social media and news content from third-party APIs. Moreover, it helps provide the frontend with the necessary information to display to the user, including the main energy graph.

8.4 Graph Database

Represents the central database of Photon, graph-based storage that contains all relevant information extracted from the Data Sources.

8.5 Graph Controller

The Controller layer is the backend part of the Photon web application that researchers may use to help in the identification of good, non-obvious opportunities in the Energy sector. It exposes a REST API that provides relevant endpoints, allowing external components to access the energy graph. In Photon, it receives requests from the frontend and interacts with the Database Controller to extract the wanted information from Neo4J, and it offers a response containing the data.

8.6 View

The topmost layer of Photon. The View is the frontend portion of the Photon web application and is the layer that the user directly interacts with. Among other features, it shows a 3D knowledge graph containing all information extracted and processed.

A high-level diagram of the architecture can be seen in Figure 4 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 D3 submission zip.

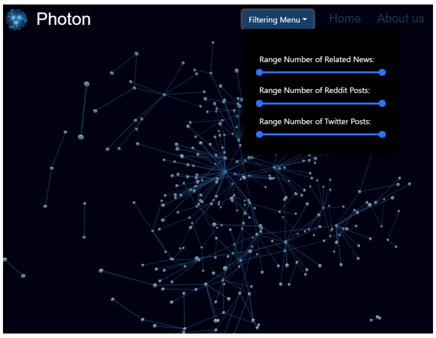

Figure 1: *Graph Visualization without nodes expansion and filtering menu.*

Figure 2: *Details present on a node from the graph.*

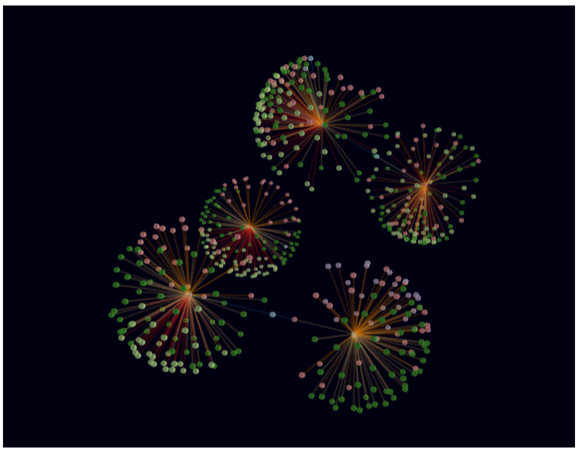

Figure 3: *Graph Visualization with nodes expansion.*

10 Conclusions and evaluation of the result

This article serves as the final report of the Photon project, which we view as interesting and valuable, related to an area that has an enormous impact on our lives and the planet. Based on our previous work and research (that includes the motivation for the project, goals, identification of existing solutions, and user requirements), we developed a successful product that satisfies its main objectives. We believe that it is an excellent basis to be used and improved by Fraunhofer AICOS in the future. Possible future features to be implemented include:

- Changing node sizes proportionally to an energy sector's popularity.
- Implementing a search functionality that allows users to search for a specific sector.
- Dynamically expanding the ontology by detecting new energy domains/concepts through Machine Learning and NLP.
- Improving metrics to define the size of the nodes (that symbolizes disruptiveness and potential), making use of sentiment analysis.
- Automatically generate a report with the topmost disruptive energy topics to view results quickly.

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

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

[6] Usearch. Usearch API.
https://nodejs.org/en/

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

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

[9] Open Energy Ontology.
https://openenergy-platform.org/
ontology/oeo/releases/oeo-physical.
owl

Appendix

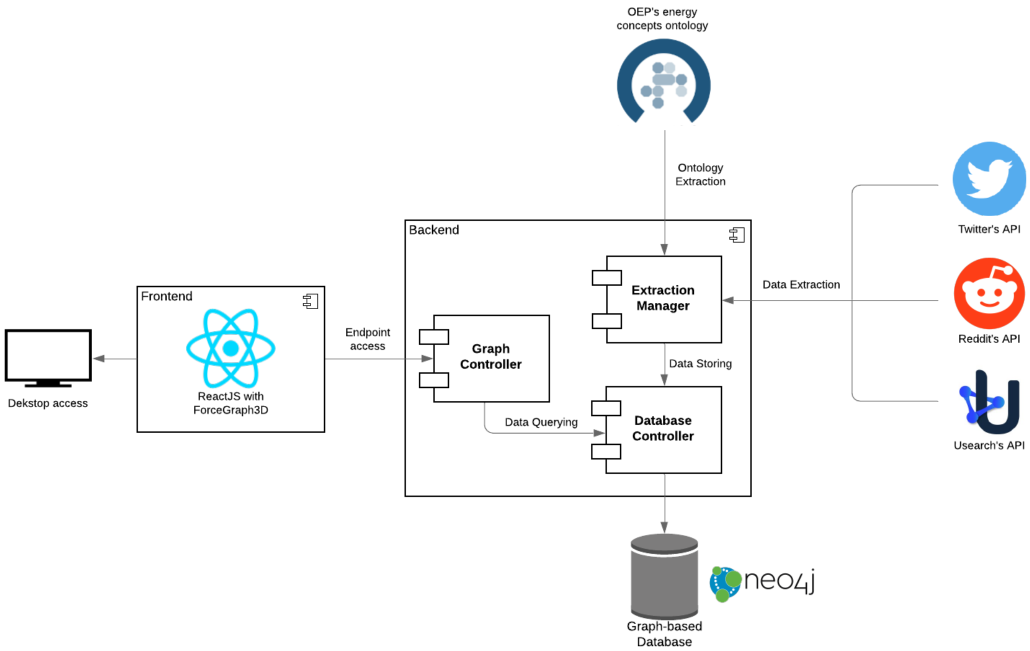

Figure 4: *Diagram describing the various modules present in a high-level view of the architecture of the system.*