Photon: Energy Market Opportunities Map

Eduardo Ribeiro
up201705421

Martim Silva
up201705205

Miguel Pinto
up201706156

Nuno Cardoso
up201706162

March 10, 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 lead 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 identify 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, and 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 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.
- **Clustering of Subareas and Fields** — Parsing and interpretation of the collected data (possibly done with NLP proximity analysis), in order to build clusters and groups;
- **Visual Interface** — Creation of a graph-based visual interface, that will represent a knowledge graph, 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 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

4.1   Renewable Energy (RE) Data Explorer

User-friendly geospatial analysis tool 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 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 hotspots 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. 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. 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 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 to use in the Clustering phase of our project, where 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. The chosen sources are not fully decided yet, however these are strong candidates that are worth mentioning. 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, 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 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 NewsAPI API

News API is a simple, easy-to-use REST API that returns JSON search results for current and historic news articles published by over 75,000 worldwide sources. With this API we can retrive news by keywords, phrases, country, publishers,

and since it has a very diverse set of sources it will be extremely useful.

5.4 Newscatcher API

Newscatcher API is another news API we will use. Even though it doesn't have the same amount of information and sources as NewsAPI, it has some interesting tools regarding Media Monitoring and Market Research.

5.5 PatSnap API

PatSnap API aggregates data points and provides information from patents, licensing, litigation and companies.

5.6 PatentsView API

The PatentsViewAPI intends to inspire the exploration and enhanced understanding of US intellectual property (IP) and innovation systems. It allows to discover people and companies and to visualize trends and patterns across the US innovation landscape.

6 Project Calendar

In this section we present a calendar with the planning for the project's tasks. See Figure1 for more information.

7 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.

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] NewsAPI. NewsAPI API. https://newsapi.org

[9] Newscatcher. Newscatcher API. https://newscatcherapi.com

[10] PatSnap. PatSnap API. https://developer.patsnap.com

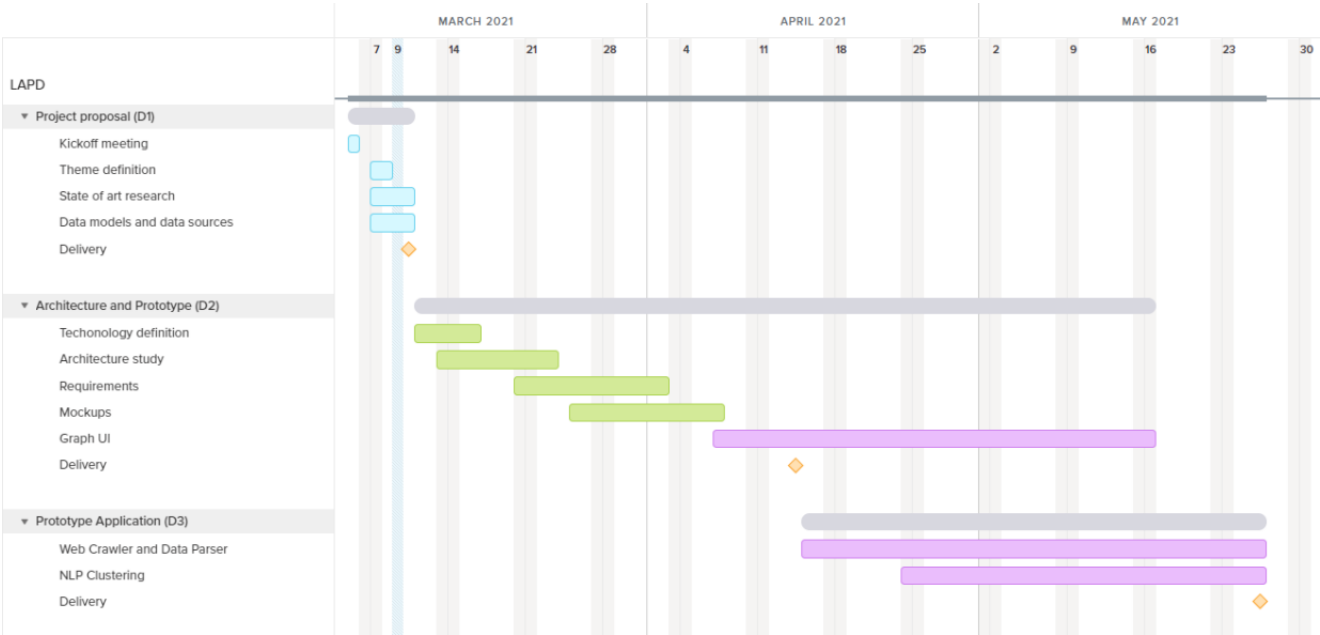[11] PatentsView. PatentsView API. https://api.patentsview.org/query-language.html

Appendix

Figure 1: *Gantt Chart showcasing the projected calendar for the project*