

Data Fellowship Project 2024

The London Data Company – Data Accessibility Tool

Organisation and Team

The London Data Company – Consultancy
Team

Application Criteria

Essential:

- Python
- Collaborative Mindset

Desirable:

- Pair Coding Experience

Practical Considerations

The opportunity can be home working or hybrid with work taking place in the London office

Selection Method

Virtual on Teams

Project Outline

Through this project LDCo aims to enhance its internal knowledge management and efficiency by creating an internal database of open-source data sets commonly used in our projects. By the end of the project, we anticipate delivering a functional proof of concept internal database of open-source data sets, complete with a thematic organisation scheme, search functionality with near neighbour suggestions, and a user-friendly python interface. This database will enable our company to better document and utilise internal knowledge, leading to increased efficiency and informed decision-making.

Project Objectives:

- Compile a comprehensive list of open-source data sets commonly used within the company's operations.
- Develop a method for organising the data sets into thematic categories based on their content and relevance to the company's activities.
- Implement a search functionality allowing users to query the database for specific data sets based on keywords or themes.
- Incorporate near neighbour suggestions to recommend related data sets based on user queries, enhancing discoverability and usability.
- Build a user-friendly method for accessing and browsing the database, with intuitive features.
- Document the database structure, search algorithms, and user interface design to facilitate future maintenance and expansion.



Data Fellowship Project 2024

Project Tasks

Proposed Tasks:

1. Identify and compile a list of open-source data sets commonly used within the company's operations.
2. Gather metadata for each data set, including title, description, source, and relevant keywords.
3. Develop a method for theming the data sets based on their content, purpose, and relevance to the company.
4. Establish a hierarchical taxonomy or tagging system to categorize the data sets into thematic clusters.
5. Implement a search engine using Python and appropriate libraries (e.g., Elasticsearch) to enable keyword-based searches across the database.
6. Develop algorithms for ranking search results based on relevance and similarity to user queries.
7. Apply data clustering or similarity measures to suggest related data sets based on user queries.
8. Implement algorithms for near neighbour suggestions to enhance the discoverability of relevant data sets.
9. Design and develop a user-friendly method within Python for accessing and browsing the database.
10. Incorporate features for filtering, sorting, and visualizing search results to enhance usability.
11. Document the database structure, search algorithms, and user interface design for future reference and maintenance.
12. Conduct thorough testing of the database functionality, ensuring reliability, performance, and usability.

Supporting Information

Established in 2021 and headquartered in the City of London, The London Data Company (LDCo) are a small team of Data experts from a blended mix of Consultancy and Academic backgrounds, utilising expertise in Data for the betterment of issues facing organisations today.

We deliver the full scale of data services, from Data Architecture and Engineering to Data Science and AI. At the heart of our mission is the commitment to enabling trusted data-driven decisions, guiding businesses toward a future where actionable insights pave the way for success in a sustainable way.

LDCo consciously consider the ethical implication of all of our solutions, adopting the appropriate Ethical framework to apply specific to the data and the data challenge being addressed.

We believe strongly in building capability in the organisations we work with, the aim of capability building is to achieve a point where our services are no longer required. Our processes and methods have been specially developed to incorporate this key pillar and supporting behaviours from the very outset of an engagement. LDCo are highly alert to the impacts of data on Sustainability.