

AI Project

Meeting started: Nov 12, 2024, 4:57:56 PM

Meeting duration: 125 minutes

Meeting participants: Agita Ferstere, Diāna Koržaviha, Natalja Krjučkova, Sergejs Kopils

Quick summary

During this meeting, Sergejs, Natalja, Diāna, and Agita discussed the progress and challenges of their project, focusing on data scraping, company risk assessment, and the use of APIs. They explored different methods for gathering and analyzing information about companies, including the use of Google searches and potential API access. The team also considered the categorization of companies based on risk levels and planned future steps to refine their approach.

Summary of Meeting

- **Current Progress:**
 - Discussed the use of Google API and ChatGPT for data scraping.
 - Evaluated the effectiveness of different scripts for scraping company information.
 - Identified issues with accessing certain news sites like Reuters and Washington Post.
 - Discussed the challenges of working with different website structures and tags.
 - Reviewed the process of identifying companies with potential risks based on keywords.
 - Discussed the limitations of using the Belgian company registry (KBO) and the need for a broader approach.
- **Key Findings:**
 - Successfully scraped data for a company under sanctions using specific keywords.
 - Identified the need for a more structured approach to categorizing risk levels.
 - Recognized the importance of having a clear algorithm for assigning risk scores.

Next Tasks

- **For Sergejs:**
 - Modify the script to handle a CSV file with company names.
 - Test the script to determine if Google IP usage incurs costs.
 - Implement a labeling system for risk assessment based on search results.
- **For Diāna:**
 - Finalize the script for checking company status (active/inactive) using the Belgian registry.
 - Ensure the script can handle variations in company names.
- **For Agita:**

- Compile a list of trusted news sources (Trast sources) for more accurate data scraping.
 - Review the scraped data for potential inclusion in the risk assessment process.
- **For Natalja:**
 - Provide a detailed description or algorithm for assigning risk scores.
 - Investigate the possibility of accessing a paid API for more comprehensive data scraping.
- **General:**
 - Share all scripts and findings in the GitHub repository for collaboration.

Conclusion

The team agreed to continue refining their approach, with specific tasks assigned to each member, and planned to reconvene on November 19 2024 at 5 PM to review progress and further develop their methodology.