

The project is to find the list of company names in the source data which are present in a dump of XML news articles.

Given are two source files.

1. CSV file with company names
2. News articles in XML format

The file paths are given in class `util.Filepath`. XML file path is stored in constant `"article_xml_path"` and CSV file path is stored in constant `"company_names_path"`. By default, it is pointing towards `repRisk-backend/resource` folder in which original XMLs and CSV are present.

Once the process is completed, a response message would be displayed in the console. Also, company names present in the XML dump are listed in the file path specified (`util.Filepath.companies_in_article_path`).

Assumptions :

- All company names along with their abbreviations are considered.
- All data given inside "()" are abbreviations or another name of the company {Eg: KFC (Kentucky Fried Chicken), both KFC and Kentucky Fried Chicken are considered}
- If the company name is given inside double quotes(") like "International Development Association (IDA); World Bank Group", neglect the data after semicolon (;).
- Company official names and abbreviations are taken separately and would be returned if they are present in the article.