



Introduction

Inngot has developed a product called Hallmarq to allow companies to input their Intellectual Properties data and receive a detailed report on its strengths and weaknesses. With this they can take a loan against their IP against a bank as collateral.

An issue Hallmarq faces, relates to the factors. A third of the product is not automated, meaning valuable human resources have to be taken up to resolve this issue. This project was made to expand the scope of automation and check whether that Patent Claims text analysis can be automated.

Aims and Objectives

To resolve this issue, the aim of the project is to implement a process that:

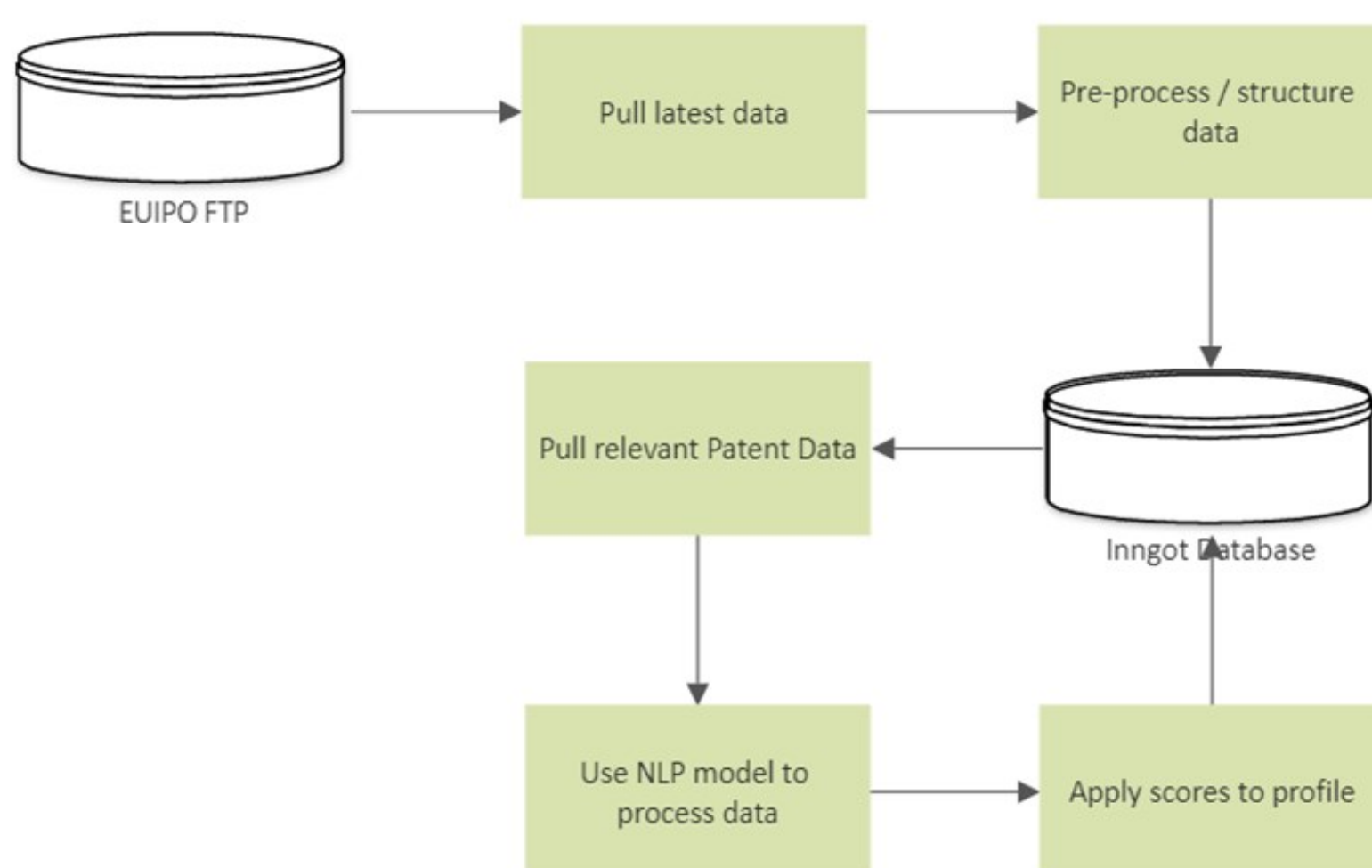
- Downloads data from an external supplier
- Process and format data to be stored internally
- Retrieve stored data to compare against clients inputs

To do this the objectives were laid out:

- Research technologies that allow the development of each stage
- Compare and contrast each technology
- Find a conclusion on which technology to use
- Design the features based around the researched technologies resolving around best standards and scenario dependence.
- Implement the up-to-standard design and experiment with Patent data
- Conduct conclusion with processed data to finalise project

Methodology

- Priorities the research to find research and white papers on technologies that have relevance to project
- Find the advantages and disadvantages of each technology found and how they would benefit the project
- Conclude resources and finalise solution
- Design system around best standards based on research



Design

It was decided that splitting up the project into three parts would be best for the design process.

Data retrieval

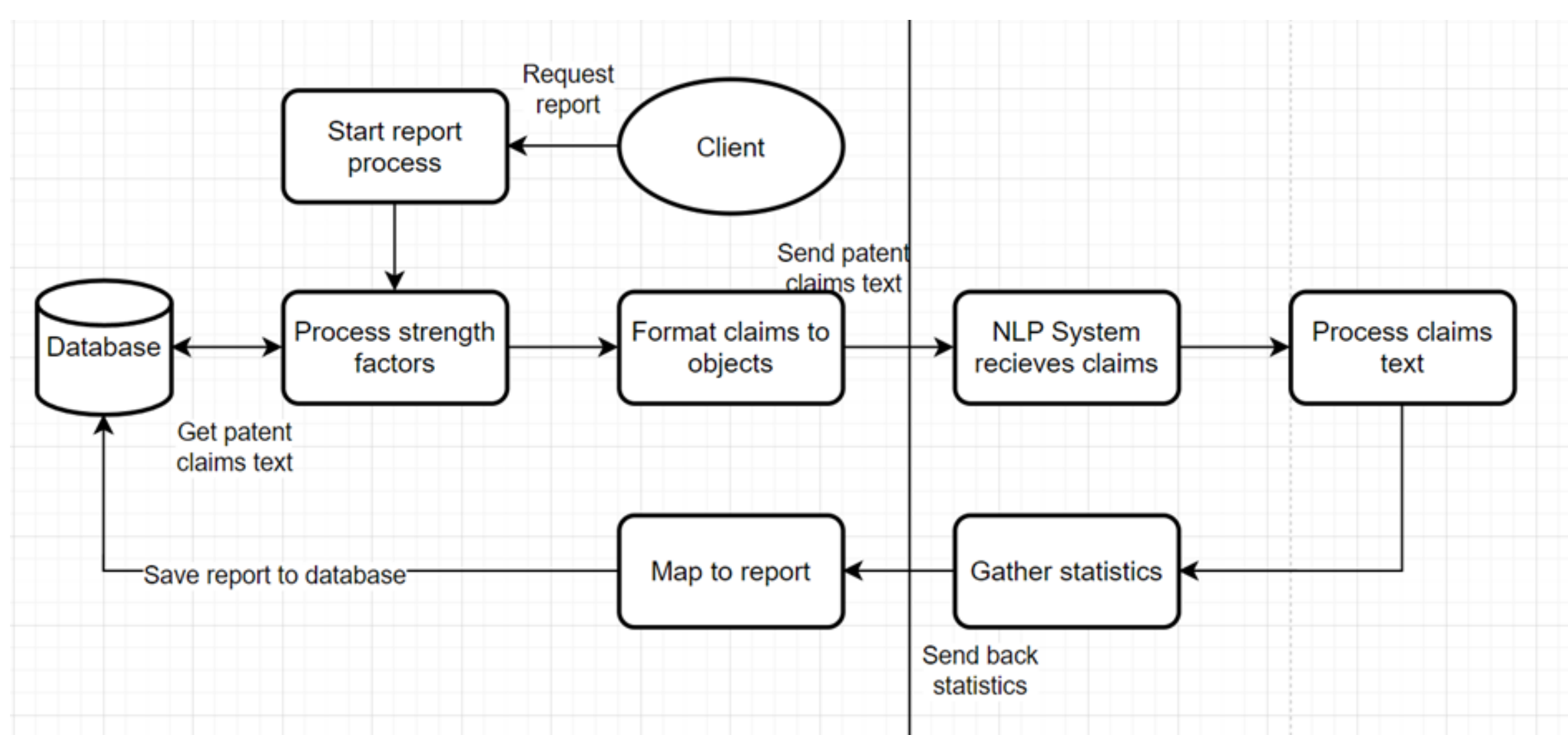
- Find the best way to utilise scripting languages within UNIX to retrieve information from a FTP

Data processing

- Design how the process will be linked up with the scripts to pull the files in and process them with the selected libraries.

NLP Model

- NLP model inputs and outputs
- Example patent was created for later experimentation and comparison



Implementation

Data retrieval

- Separate script files with their own functions were developed to allow modularity and ease of maintenance.
- Uses UNIX in built system “Cronjob” to schedule tasks to download data daily
- A file structure was developed to easily organise and process the XML files

Data processing

- Script file executes a Java executable which takes in the selected XML file and processes it with the SAX and JAXB2 libraries.
- Javas inbuilt SQL library is then utilised to store information into a database

NLP Model

- JAVA was utilised here as well with the usage of Apache’s OpenNLP library to create a NLP model to begin experimentation
- Unit tests were developed to test if the NLP model was actually fully functional and outputted the correct data.

Outcome of results

Junit was used within Java to test the corresponding outputs:

- Independent Claims Amount
- Dependent Claims Amount
- Least Amount of Sub-clauses
- Identification of Certain Claim Types

From the unit tests it was found that the NLP model was fully functional and working.