

Transferability document

Collaborative learning using graph database

Research theme: NLP

Semester name: DDBL

Project owner: Mark de Graaf

Students: Nikita Gavrilov, Ilya Tsakunov, Alexander Vereshchagin, Victor Plesciuc, Yaniek Martens, Gergana Agorasteva

Date: 18/01/2022



# Project Description

## Context

Flooid [[1]](#footnote-2)is a start-up company that mainly functions as a consulting partner in the networking industry with platforms using natural language processing and knowledge graphs. Flooid acts as a catalyst in the shift of systems creating an impact on how people work, live and care for our environment. Their sense making indicates that many systems have reached their highest potential in relation to their context and need to shift. They do this by bringing people, knowledge and artificial intelligence together augmenting human capabilities, seamlessly integrating on- and offline worlds.

Flooid has its very roots in learning about reinventing organizations through communities. It has been the cradle for a model of community-based learning and innovation supported by a platform using the Natural Language Processing concept. This platform creates contexts for making explicit what is implicitly known, resulting in collective learning. Flooid has a list of partner universities and mindlabs that are ready to be a part of the networking environment. (See appendix)

## This project

The main idea of this network is to share partners’ knowledge and connect people to work together based on needed skills, with a help of neo4j graph data platform.  This network is also supposed to prevent time-wasting problems while two or more parties are working on similar projects apart. The network will be designed with the help of Flooid’s partner EyeOnText[[2]](#footnote-3), that provides workshops on the Wowool SDK. It is a cross-platform tool and comes with language bindings for Python, C++. This tool helps to Ingest unstructured textual data and deliver structured semantic objects, such as entities, sentiments, profiles, facts and links.

## Main research question

How to design an information network to enable its’ members to increase transparency and data flow?

That would let partners to share their knowledge and connect people to work together based on specifically required or similar information. Thus, the information is publicly available for all members of the network.

##  Sub questions

1. What kind of data should be in network?
2. What are the community's needs for the network?
3. How can Datastic make a network using the tools (web scrapping, NLP, graphs) that are chosen?
4. How can the data from network be stored?
5. How can the network be presented to the end-user?
6. How can this project be transformed to a business?

## Results

In this section, we describe all the results that are to be delivered by Datastic. This contain research documents that look into the possibility of creating a knowledge graph based on Flooid’s and the communities’ needs and the planning of how this was conducted, a user manual describing the work flow of the proof of concept, content files resulting from applying webscraping techniques from project partners and the programming code necessary. These can all be found in the Transferability folder that Datastic handovers to Flooid.

**Project Plan**

At the beginning of the project, the group created a Project plan where the most important guidelines of the project were described, including the main research question, sub-questions, the scope of the project, methodologies that were used throughout the project, and settled up initial deadlines. The project plan was reviewed several times by the client and by Datastic’s coach in order to provide quality feedback and agree on the terms of the project. When both parties were agreed on provided way of working that was documented in the project plan, it was validated.

**Research Doc**

In the research document all the decisions, research, and findings that were made during the lifespan of the project were documented. The document itself is used as a validation of the results that the Datastic have achieved throughout the project.

**Business plan**

Business plan used as validation of the main idea of the product. In this document promotional strategies, distribution strategy, SWOT and TOWS analysis, market research were discussed.

**Manual**

The Manual Document describes the way of setting up and working with the applications and technologies that were used. The main purpose of the document is to provide an explanation and understanding for the future project groups on how to work with the technology and provide for them a good start.

**Webscraped files**

Datastic performed research on techniques which allow us to retrieve content from websites that are connected to the community. Based on the research and Flooid’s advice the technique chosen is called ‘webscraping’. Datastic gathered this content and structured it into a folder containing textfiles named after the entities that provided the data.

**Code**

To reach a proof of concept/demo product there were a few coding steps necessary for Datastic to go through. Firstly, webscraping, getting data about the target companies, secondly, creating rules using the Wowwolian language to process the webscraped data and make it ready to be translated into Nodes, Relationships, Labels and Properties for the Neo4j database. Lastly, Datastic also provides code for the visualisations in Neo4j using the CYPHER query language.

**Validation and Added Value**

The project was realised using the Agile SCRUM working method. Datastic members had weekly meetings with stakeholders from both Flooid and EyeOnText. During these meetings the prototypes developed were discussed and evaluated, and new actions were created to ensure that the product developed was in line with the community's needs.

## Methodology

|  |  |  |  |
| --- | --- | --- | --- |
| Sub question  | Research strategy  | Methods | How?  |
| 1. What kind of data should be in Network? | -Workshop -Field -Lab | * Brainstorm
* Interview
* Hardware validation
 | * Weekly meetings where the group discussed the next steps of projects .
* Interview with Petra van Dijk representative of MindLabs.
* Test how well the software is running on local devices.
 |
| 2. What are the community’s needs for the network? | -Workshop-Field | * Stakeholder analysis
* Interview
 | * Business plan was created.
* Interview with Petra van Dijk representative of MindLabs.
 |
| 3. How can Datastic make a network using the tools (Web scraping, NLP, Graphs) that are chosen? | -Library -Workshop-Lab-Field-Showroom | * Expert interview
* Prototyping
* Product review
* Code review
 | * Throughout the project lifespan the several prototypes of application were created to test and gain more knowledge about the technology.
* Weekly meetings with Flooid partners evaluate the current versions of software.
* Monthly meetings and workshops with EyeOnText, experts in NLP.
 |
| 4. How can the data from the network be stored? | -Library -Workshop -Field  | * Literature study
* Exploration of available tools
 | * Extracted data from websites were stored locally on Datastic’s devises and were updated via GitHub.
 |
| 5. How can the network be presented to the end-user? | -Workshop -Showroom-Field  | * Prototype
* Observation
* Available product analysis
* Peer review
 | * The prototype presented to the end user and their interaction with it was evaluated.
* Research how we can replace Neo4j web application so it will be more intuitive for the end user.
 |
| 1. How can this project be transformed to a business
 | -Library-Field | * SWOT analysis
* Stakeholder analysis
 | * Business plan.
 |

# Handover

 

Figuur 1 One of the first version of the prototype



# Appendix

List of DR organizations involved in the Flooid Project:

Universities:

**Breda University of Applied Sciences:**

<http://www.buas.nl/>

 **Fontys:**

<https://fontys.nl/>

 **Tilburg University:**

<https://www.tilburguniversity.edu/nl>

Mindlabs:

<https://www.mind-labs.eu/>

Partners:

**Mijzo**

<https://www.mijzo.nl/>

**WPG/Zwijsen**

<http://www.zwijsen.nl/>

**Interpolis**

<https://www.interpolis.nl/>

**Thebe and De Wever**

<http://www.thebe.nl/>

<https://dewever.nl/>

Junior Partners:

**Anyware**

<https://www.any-ware.nl/>

**Castlab**

<https://castlab.nl/>

**Genius Voice**

<https://geniusvoice.nl/>

**Ivy Works**

<https://ivyworks.nl/>

**Thoughtline**

<https://thoughtline.nl/>

**Blewscreen**

<https://www.blewscreen.com/>

**Flow Concepts**

<https://flowconcepts.nl/>

**Hi Lex**

<https://hallolex.nl/>

**Purple Mountain**

<https://www.purplemountain.nl/>

DigiShape:

**Rijkswaterstraat**

<https://www.rijkswaterstaat.nl/>

<https://www.rijksoverheid.nl/ministeries/ministerie-van-infrastructuur-en-waterstaat>

**Deltares**

<https://www.deltares.nl/nl/>

**Van Oord**

<https://www.vanoord.com/>

**TU Delft**

<https://www.tudelft.nl/>

**Boskalis**

<https://boskalis.com/>

**BZ Ingenieurs & Managers**

<https://boskalis.com/>

**Periplus Consultancy**

<https://www.periplus.nl/nl/home/>

**Marin**

<https://www.marin.nl/en>

**HKV**

<https://www.hkv.nl/>

**Witteveen Bos**

<https://www.witteveenbos.com/>

**IHM**

<https://www.informatiehuismarien.nl/>

**Maris**

<https://www.maris.nl/>

**Port of Rotterdam**

<https://www.portofrotterdam.com/nl>

1. https://flooid.eu/ [↑](#footnote-ref-2)
2. https://www.eyeontext.com/ [↑](#footnote-ref-3)