# Datasheet

The product is a piece of software processing natural language analysis.

Table of contents

Table des matières

[Datasheet 1](#_Toc479154392)

[The purpose of my project 1](#_Toc479154393)

[The context of my project 1](#_Toc479154394)

[What is needed by Quorum 1](#_Toc479154395)

[How does it work? 1](#_Toc479154396)

[The features 1](#_Toc479154397)

[The theory 2](#_Toc479154398)

[Results and perspective 2](#_Toc479154399)

# The purpose of the project

## The context

During an election campaign, some candidate supporters, or employees from poll institutes go door to door to ask for opinions. They gather information they have to process eventually which represents hundreds of answers. The questions are split into two categories: close-ended questions and open questions. Close ended questions can be easily automatically processed, but open questions are much harder to handle.

## What is needed by Quorum

The Quorum Company is working on a complete solution for people, such as mayors, who organise campaigns to know people’s needs or opinions. The solution should help their users to analyse the answers automatically. This project provides a solution by using natural language processing. The idea is to download raw text data, and to automatically process each answer to return words that matter and topics mentioned. Even though the process won't be as good as a human one, it will cut the number of people needed to do this tedious work.

# How does it work?

## The features

The project is divided into two steps : a computational part and a human interface.

* The computational part is composed of different types of analysis. A frequency analysis which returns the mostly used words for each question. An ontological analysis finds which topic is most likely to be mentioned, according to an existing ontology. An LDA analysis finds the topics people mentioned according to a probabilistic analysis.
* The human interface is not the part the company needs. In fact, they want my computational part to be perfected and inserted into their own interface. However, in order to work on the link between a computational part and a visual one, my work has an interface to allow its user to enter a link to his data and to choose the type of analysis he wants the piece of software to process.

## The theory

My project is split into two parts as mentioned before. The data analysis is processed thanks to Python language and its libraries. In fact a lot of theoretical researches about natural language processing have been conducted and implemented using this language.

Thus, the probabilistic models are implemented on a server using a python interpreter. When using the piece of software, a user interact with a C# interface that calls python functions I implemented on the server

## Results and perspective

My project’s objectives were modified during the project to be adapted to what I was succeeding in and what I was finding. A complete human interface with graphs and words clouds would have been too hard to be implemented and useless for the company. Thus I focused my work on the algorithm that could be used.

It appeared the ontology analysis using the French equivalent of the Wordnet was not relevant as the topics it returned were far from what humans found.

-          Perspective