# 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)

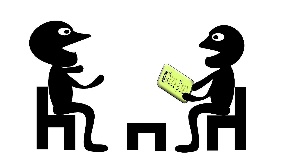
[The features 1](#_Toc479154397)

[The theory 2](#_Toc479154398)

[Results and perspective 2](#_Toc479154399)

# The purpose of the project

## The context

[](https://www.google.fr/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiB7qPoq8TTAhUJQBoKHREeBWoQjRwIBw&url=https%3A%2F%2Fwww.cap-coherence.fr%2Fblog%2F2016%2F3%2F21%2Fchanger-metier-bon-choix&psig=AFQjCNFbm7uXh47xhK2bNTHxp-9BCkgWHw&ust=1493372344845749)

During an election campaign, some candidate supporters, or employees from poll institutes go door to door to ask for opinions. They gather information they process eventually which are 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 analyse.

## What is needed by Quorum

[](https://www.google.fr/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiw04mtq8TTAhVFvBoKHc3WCG0QjRwIBw&url=https%3A%2F%2Ffr.linkedin.com%2Fcompany%2Fquorum-apps&psig=AFQjCNGrOfAdc-g7u_FzEiFXVxKzYyS3HQ&ust=1493372257856009)The Quorum Company is working on a complete solution for its cleints, such as mayors, who organise campaigns to know 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 present words that matter and topics mentioned. Even though the process will not be as good as a human one, it will cut the number of employees to do this tedious work.

# How does it work?

## The features

The project is divided into five features:

[](https://www.google.fr/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjr4ce1rMTTAhUIWBoKHXpZAGcQjRwIBw&url=http%3A%2F%2Fwww.busitelce.com%2Fdata-visualisation%2F30-word-cloud-of-big-data&psig=AFQjCNEHzD3VKvx8m1hTrDFURQhRyAjE6Q&ust=1493372548230463)It is composed of different types of analysis. A question analysis provides questions which are opened-questions, a frequency analysis which presents the most employed words for each question. An ontological analysis provides which topic is most likely to be mentioned, according to an existing ontology. An LDA analysis presents the topics mentioned according to a probabilistic analysis. An HAC analysis gathers answers according to similarities.

## The theory

Data analysis is processed thanks to Python language and its libraries. In fact theoretical research 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. The piece of software will be integrated into the Quorum interface which will call function from the server.

## Results and perspective

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

[](https://www.google.fr/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwj9_bmRrcTTAhXFvBoKHblZD2oQjRwIBw&url=https%3A%2F%2Fwww.quest.com%2Fcommunity%2Fproducts%2Fstatistica%2Ff%2Fforum%2F7383%2Flooking-forward-to-version-13-of-statistica&psig=AFQjCNGGckAT3E3kYQgwAlm4YFLBEhvuRg&ust=1493372714787693)The ontology analysis and the HAC ontology did not appear to be  as relevant as expected but were important to analyse.

A complete interface using the functions and an improvement of the ontology analysis and the HAC analysis are still possible.