2024

Observations: Automation

TECHNICAL NOTE

GILBERT D. BRAULT

Table of Contents

Introduction

The problem to be solved

The maintenance of elevator installations is governed by laws and decrees in France.

The control offices, mandated by the owner's representative, carry out periodic general inspections (VGP) or other visits, acceptance of modernisation work for example, etc. Their auditors then produce inspection visit reports.

These reports record, among other things, anomalies concerning the equipment visited. Regulations, in some cases, require these problems to be addressed. These anomalies are called observations or reservations. They must be followed up to resolve the problems identified.

These reports are then sent to the elevator maintenance companies, mandated by the owner's representative.

The company Ilex ascenseurs, operating on French territory, whose headquarters are in Antibes in France must therefore face this problem.

Ilex Elevators regularly receives a stream of inspection visit reports, written by various control offices mandated for the inspection of equipment in the fleet for which Ilex Elevators is a maintenance provider.

The purpose of this note is to describe the process of processing observations at Ilex elevators to show the architecture of the technologies contributing to the automation of information processing, focusing on the analysis phase.

The purpose of the analysis phase is to assign an observation to a problem-solving method with a view to assigning the work to be carried out to the Ilex elevator construction teams, and sometimes even to the customer.

The proposed solution

The solution proposed to Ilex is an integration of the solution into the existing information system with the possibility for users to:

- To visualize the tree of observations issued by a control office over a given period. One of the offices can be chosen according to the content of the pre-existing database
- Search for semantically similar observations by entering an observation text.
- Interact with the representation to define the groupings corresponding to observations sharing the same semantics by adjusting the size of the groups by a distance.
- Build the dataset needed for visualization periodically to update visualization and search

Steps for processing observations

There are three steps to processing observations

- Entry: collect all the reports and enter the observations contained therein
- Analysis: identifying the nature of observations and assigning them to resolution methods
- Processing: monitor the progress of the resolution process, when the observation has been assigned to an Ilex elevator team that must execute the resolution of the problem.

Annually, Ilex elevators must process between 20,000 and 30,000 observations. Automated processing tools are necessary to obtain the level of quality required in the elevator maintenance business to comply with regulations.

Entry

The purpose of the Entry is

- Collect all reports received by Ilex Elevators through different channels
- Codify information to facilitate further processing
- Associate observations with a device under Ilex elevator mandate

In this phase, all the observations processed by Ilex elevator are therefore found in a centralized database.

It should be noted, however, that the observation as such is not codified: it is a text in natural language, as it was entered by the auditor during his inspection visit.

Analysis

The purpose of the analysis, which is carried out by experts from the company Ilex elevator (the analyst), is to analyse the observations to define the work corresponding to the resolution of the problem posed.

The analyst has a screen in front of him where all the observations for a given device are represented in the form of a table.

He takes them one by one and fills in an assignment sheet by observation in which he characterizes the team in charge of the treatment and the nature of the work to be carried out.

The purpose of this technical note is to show how this assignment work can be automated.

Treatment

Processing is subject to a classic work follow-up to monitor the progress of the tasks assigned to the staff of the various execution teams. Ultimately, the "lifting of reservations" is carried out – with transmission to the control office – when the work is completed, or its planning is certain.

Conclusion for Automating the Analysis Phase

The analysis phase

- Requires skilled resources when done manually.
- Implies a certain dispersion of assignment results due to the manual work of classification linked to the expertise of each person and the fatigue of a routine activity.
- Imposes delays for completion due to the need for human resources.
- Is a bottleneck in the process of processing observations.

So, it's only natural to look for ways to automate this task. We will therefore focus on the automation of the analysis phase, in particular:

- Understand the problem of assigning observations to assign them to a problem-solving method. A first step is to look at the nature of the observations.
- Describe the architecture for automating this process.

Nature of the observations

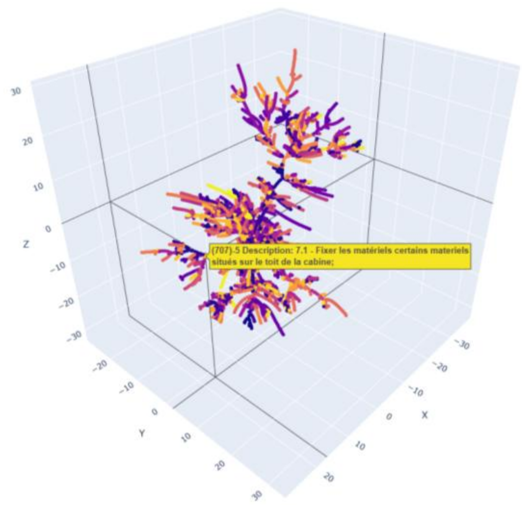
Take, for example, the observations that are issued by the AASA audit office.

Over the last two years, AASA has issued 4721 observations in the reports concerning lifts for which Ilex is mandated for maintenance.

Of these 4721 observations, 2755 are the subject of a unique description. That is, there are 2755 sentences in natural language, which describe observations in a unique way: these are sentences that have a different syntax from each other.

Let's now dive into the details of an observation to understand what this variety, related to the use of natural language, means.

Nombres -- Observations(4721|2755)|groupes=(743)



In the tree of observations, comprising 743 groupings, let's select an observation at random:

```
7.1 - Attach the equipment to the roof of the cab
```

And now let's look at the observations that are close

	index	Enregistre_le	societe_ratachement	Agence_Id	Appareil_Id	Description
1231	1231	2022-12-23 14:39:39	ILEX-06	601	22491	1 7.1 - Fixer les matériels certains materiels situés sur le toit de la cabine;
1389	1389	2023-01-04 15:35:24	ILEX-06	601	24990	2 7.1 - Fixer les matériels, situés sur le toit de la cabine. Prestation non réalisée
2444	2444	2023-02-19 23:34:49	ILEX-06	601	24502	3 7.1 - Fixer les matériels qui le nécessitent situés sur le toit de la cabine.
2894	2894	2023-03-02 09:52:20	ILEX-06	601	214	4 7.1 - Fixer les matériels rajoutés sur le toit de la cabine. Prestation non réalisée
3389	3389	2023-03-28 15:26:52	ILEX-06	601	25102	1 7.1 - Fixer les matériels qui le nécessitent situés sur le toit de la cabine.

With the "distance" that was imposed, a group of 5 observations was found.

Of these 5, 4 are unique:

```
7.1 - Attach the equipment, certain equipment, located on the roof of the cabin; 1
7.1 - Attach the equipment, located on the roof of the cabin. Service not performed 2
7.1 - Attach the equipment that requires it located on the roof of the cabin. 3
7.1 - Attach the added equipment to the roof of the cab. Service not performed 4
```

We can therefore see that a group contains observations with small variations in language, or complements.

If we increase the distance, we see that the number of groups drops to 241. Let's look at the evolution of the group from our previous observation.

We changed the distance from 1 to 1.8 and we recalculated the groups (recoloring).

	index	Enregistre_le	societe_ratachement	Agence_Id	Appareil_Id	Description
428	428	2022-05-31 14:40:32	ILEX-06	601	23829	7.1 - Fixer les différents élements situées sur le toit de la cabine, notamment les alimentations des vigiks et les différents téléphones.
824	824	2022-07-20 15:59:07	ILEX-06	601	25513	7.1 - Fixer les matériels certains materiels situés sur le toit de la cabine, notamment le téléphone.....
844	844	2022-07-20 15:59:21	ILEX-06	601	25514	7.1 - Fixer les matériels certains materiels situés sur le toit de la cabine, notamment le téléphone.....
1231	1231	2022-12-23 14:39:39	ILEX-06	601	22491	7.1 - Fixer les matériels certains materiels situés sur le toit de la cabine;
1355	1355	2023-01-04 11:19:12	ILEX-06	601	24989	7.1 - Fixer les matériels , notamment le téléphone, situés sur le toit de la cabine. Prestation réalisée
1389	1389	2023-01-04 15:35:24	ILEX-06	601	24990	7.1 - Fixer les matériels, situés sur le toit de la cabine. Prestation non réalisée
1426	1426	2023-01-04 15:35:30	ILEX-06	601	24991	7.1 - Fixer certains matériels situés sur le toit de la cabine, notamment le téléphone..... Prestation réalisée
1687	1687	2023-02-06 09:49:14	ILEX-06	601	24448	7.1 - Fixer les matériels certains materiels situés sur le toit de la cabine, notamment le téléphone..... Prestation non réalisée
2444	2444	2023-02-19 23:34:49	ILEX-06	601	24502	7.1 - Fixer les matériels qui le nécessitent situés sur le toit de la cabine.
2758	2758	2023-02-28 15:32:49	ILEX-06	601	25371	7.1 - Fixer les matériels qui le nécessitent situés sur la cabine, notamment le téléphone..
2894	2894	2023-03-02 09:52:20	ILEX-06	601	2142	7.1 - Fixer les matériels rajoutés sur le toit de la cabine. Prestation non réalisée
3389	3389	2023-03-28 15:26:52	ILEX-06	601	25102	7.1 - Fixer les matériels qui le nécessitent situés sur le toit de la cabine.

We then move on to a group of 12 observations, 10 of which are unique:

```
7.1 - Fix the various elements located on the roof of the cabin, in particular the power
supplies of the vigiks and the various telephones.
7.1 - Attach the equipment to certain equipment located on the roof of the cabin, in
particular the telephone.....
7.1 - Attach the equipment, certain equipment, located on the roof of the cabin;
7.1 - Attach the equipment, in particular the telephone, located on the roof of the cabin.
Service provided
7.1 - Attach the equipment, located on the roof of the cabin. Service not performed
7.1 - Attach certain equipment located on the roof of the cabin, in particular the
telephone..... Service provided
7.1 - Attach the equipment to certain equipment located on the roof of the cabin, in
particular the telephone..... Service not performed
7.1 - Attach the equipment that requires it located on the roof of the cabin.
7.1 - Attach the equipment that requires it located on the cabin, in particular the telephone.
7.1 - Attach the added equipment to the roof of the cab. Service not performed
```

New elements have slipped into the sentences, however, the work to be done certainly remains the same.

So that's the problem with these natural language observations:

Many observations can be grouped together because they correspond to an identical method of problem solving because they have the same semantics.

The technical problem posed to the information processing system is therefore to group together observations that are expressed in natural language when they have the same semantics.

Defining the Similar Observation Grouping Problem

What are the limitations of conventional solutions?

Relational database systems are the traditional tools for enterprise data processing.

When relational database systems are used, the processing of information is very efficient, if the information has been codified.

The codification of information is not only the translation of information into a form that can be assimilated by a machine, but there is an additional step, which is to define the information in such a way that the machine inherits part of the knowledge contained in this information.

If the observations had been the subject of sufficiently detailed coding work beforehand to unequivocally define the work to be done, the machine, with a relational database system, would have encountered no difficulty in grouping the observations!

But this is not the case, due to the multitude of stakeholders and the relative complexity of the problems to be addressed, natural language coding of observations is a necessity.

Conventional database management systems are therefore powerless to solve this problem of classification of observations!

It was therefore natural that this codification should be done manually: this is the analysis stage.

A typical database management system does not know how to group natural language sentences that have the same semantics

What are the expectations for automatic coding of observations?

The problem that needs to be solved is therefore to automatically codify the observations to solve the question of automating the analysis phase.

Before moving on to the actual resolution of the problem, the question must be asked what the needs are to arrive at a viable solution.

The needs are as follows:

1. Find an automatic way to group observations of the same semantics.
2. Represent the observations so that a business expert can confirm that the proposed classification is in line with the state of the art of the trade and make adjustments to the groups.
3. Search for a group by giving the text of an observation and locate it on the representation

The first point is self-evident and the second is the ability for the men in the field to judge the relevance of the classification and to attribute it to themselves in an operational way.

Solution to the problem of classification of observations

The necessary techniques

To solve the problem of automatic classification of observations, several techniques are necessary to meet the needs analysed in the introduction.

Here is the table giving the techniques, their descriptions and their contributions to the needs

Technique	Description	Contribution to needs
Vectorization Search	Natural language sentences must be transformed into vectors ($\mathbb{R}^n$). The projection of sentences into this vector space must ensure that • Sentences of the same semantics must be close (distance between two vectors) • We must be able to find a sentence that has been vectorized by presenting a text with similar semantics	• Prerequisite for automatic grouping • Search for groups by giving the text of an observation
Size reduction	Vectors in large spaces are not ergonomically representable and easily interpreted by humans • It is therefore necessary to have a reduction technique to project the vectors in 2D or 3D • It is desirable that this reduction makes it possible to show observations with the same semantics next to each other	• Representation of observations for use by the subject matter expert
Clustering	We have vectors in space with n dimensions that represent observations, how to automatically group observations from neighbouring semantics • Possibilities of grouping observations at a given distance by connection • Colouring Groups	• Creation of observation groups of the same semantics
3D graph representation	After reducing the size of the vectors, it must be represented on the screen	• Visualization of observation graphs

Technologies used

The required techniques are supported by available and well-maintained open-source software.

Here are the most significant components that are used for the solution

<table>
<tr><th>#</th><th>Component</th><th>Technique</th><th>Description</th></tr>
<tr><td rowspan="4">Data creation for visaulization and research</td><td>all-MiniLM-L6-v2</td><td>Vectorization</td><td>Origin: Microsoft
"Sentence Transformer" which has the ability to transform a list of sentences into vectors (dimensions: 384).
Neighboring vectors correspond to sentences of neighboring semantics</td></tr>
<tr><td>DO</td><td>Research</td><td>Origin: META (Facebook)
Vector database for searching by distance or nearest neighbors</td></tr>
<tr><td>UMAP</td><td>Size reduction</td><td>Source: Leland McInnes, Tutte Institute for Mathematics and Computing (TIMC) in Ottawa, Canada
Uniform Manifold Approximation and Projection
Dimension reduction method</td></tr>
<tr><td>igraph</td><td>Size reduction</td><td>Origin: igraph
C++ Graph Processing Library with Python Proxy
MST technique and 3D graph representation</td></tr>
<tr><td></td><td>Plotly</td><td>3D Visualization and Graphic Interaction</td><td>Origin: Plotly
3D Graphics Visualization and Interaction Library</td></tr>
<tr><td></td><td>Sklearn</td><td>Grouping by agglomeration</td><td>Origin: Sklearn
Machine Learning Library
We use the clustering algorithm</td></tr>
</table>

The solution is therefore based on solid technological foundations corresponding to man-years of research and development.

The Minimum Spaning Tree (MST) method

In the end, this is the method that was chosen for the representation of the observations. It's not the most efficient, UMAP is reputed to be faster, but it's the one that is the easiest to explain and that offers a rendering that corresponds to the problem:

1. In n-dimensional space, we have the collection of observations that have been transformed into vectors.
2. We can therefore calculate the matrix of distances between all observations using vectors.
3. But a distance matrix between vectors is also the representation of a graph whose nodes are observations and relations are distances. There are codes to create a graph from the distance matrix.
4. Then, we calculate the graph corresponding to the minimum spanning tree MST of this graph.
5. A minimum spanning tree is like a network of roads that connects several cities in the most economical way possible. Let's say you want to connect all the cities in a region with roads, but you want to spend as little money as possible to build those roads. A tree covering minimum weight shows exactly which roads to build to connect all the cities without making unnecessary detours, and with the least amount of money possible. The distance from one city to another is the smallest, hence the term minimum spanning.
6. Then, we use a code that generates the 3D coordinates of the MST nodes: this is what we use to represent the tree on the screen.
7. We don't show the edges because there are too many of them

This method allows you to represent the observations with a 3D view, each point or node of the graph corresponds to an observation.

Observations that are semantically similar are represented by long threads, which allows them to be spotted very quickly.

This graph is extremely useful for visualizing observations and their proximity.

The graph presented in the "Nature of Observations" paragraph is an example illustrating all 4,721 observations issued by the AASA audit office. We have experience with graphs of more than 15,000 nodes that work perfectly.

The creation of MSTs and their 3D coordinates is the most demanding in CPU resources and time.

For small graphs, for example for 2725 nodes, it takes about 95 seconds, but for 16061 nodes, it takes 3974 seconds. This on a powerful CPU (I7, or even I9 or Xeon, although only one core is used) and with at least 32GB of RAM.

Ilex Solution Overview

This solution is part of the integration with an existing information system: the database of observations is pre-existing.

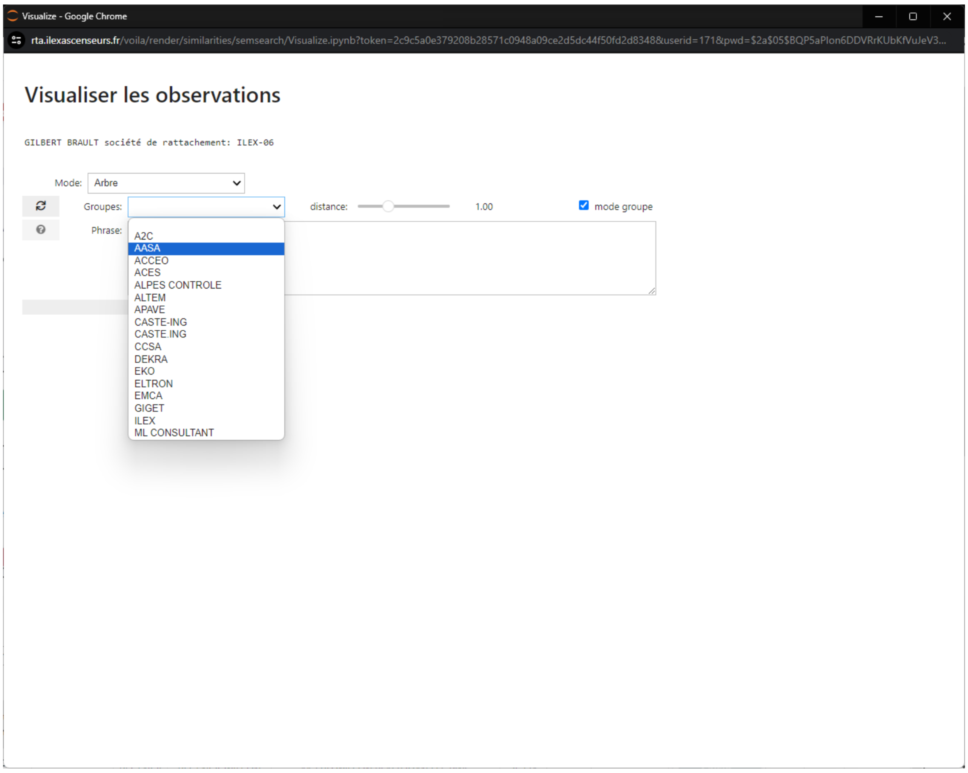
The solution implements a dedicated software server that fetches data from the database's SQL server.

Now here is an overview of the main user features

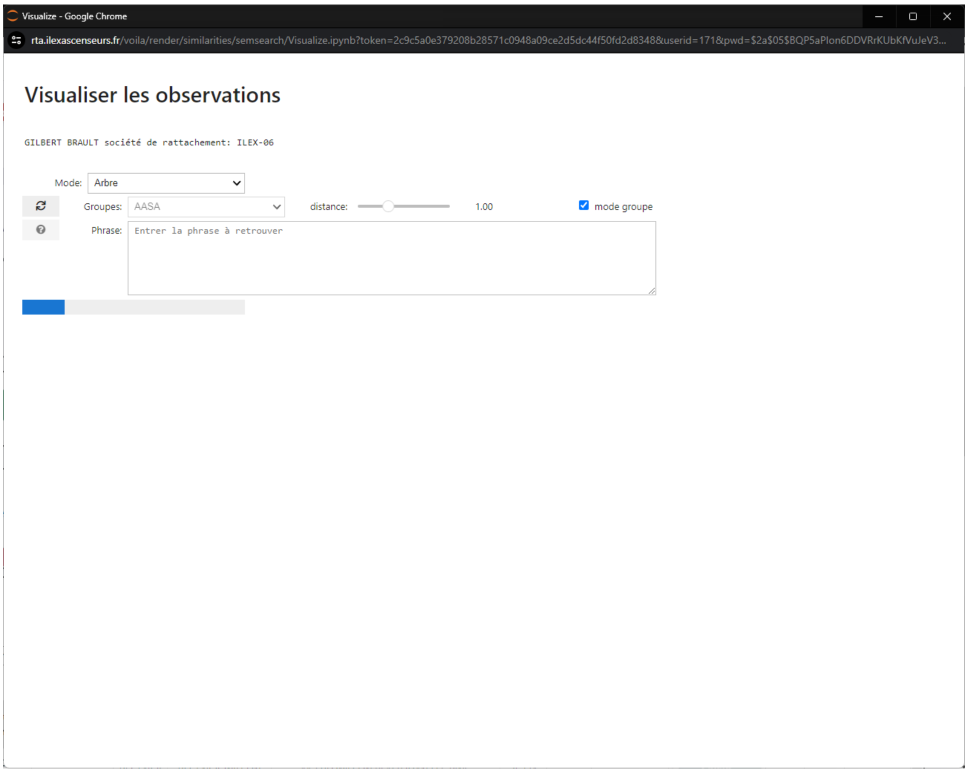
Tree Visualization

The control offices each have their own tree for representing the observations

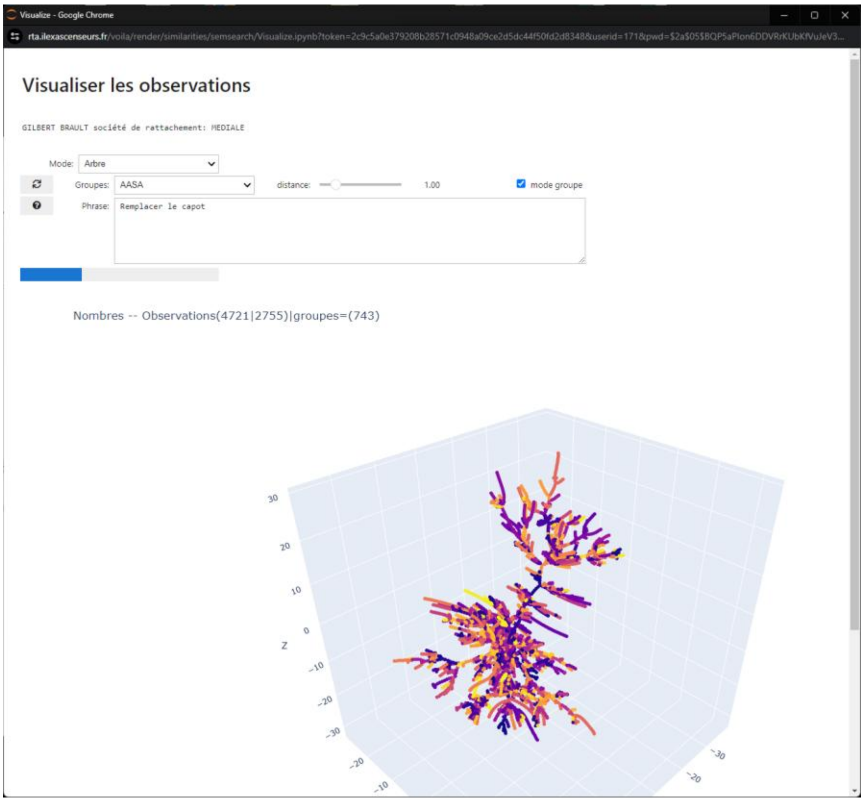
The first thing is to choose a group:



A progress bar is set up



The time it takes for the coloration calculation and the 3D representation of the tree to be carried out.

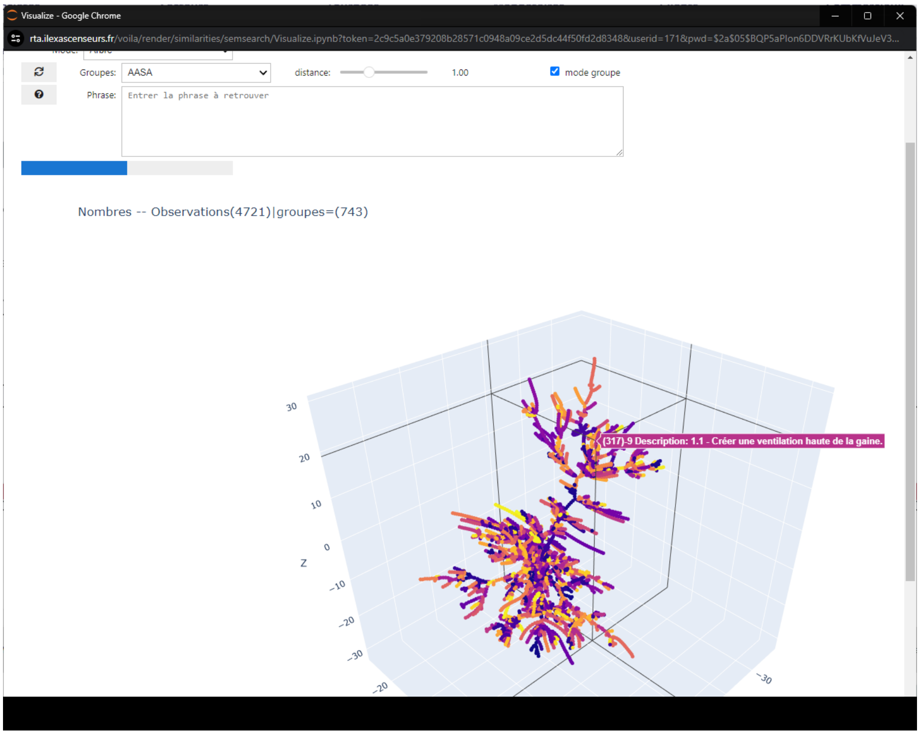


When it's finished (a few tens of seconds at most), the tree appears in 3D.

Each point in the tree is an observation issued by an office in a report.

The total number of observations, the number of unique observations, and the number of groups are shown.

When you hover the mouse over the tree, the text of the observation is displayed



The data at the beginning of observations is – a colour group number – the number of elements in the group

(317)-9 Description: 1.1 - Créer une ventilation haute de la gaine.

You also must maximize the window and when you hover the mouse over the tree to see a toolbar appear when the mouse hovers over the tree.

To activate a tool, just click on it and it turns gray a little bit more intense

"Zoom": simply operate the mouse wheel to enlarge the tree around the mouse point

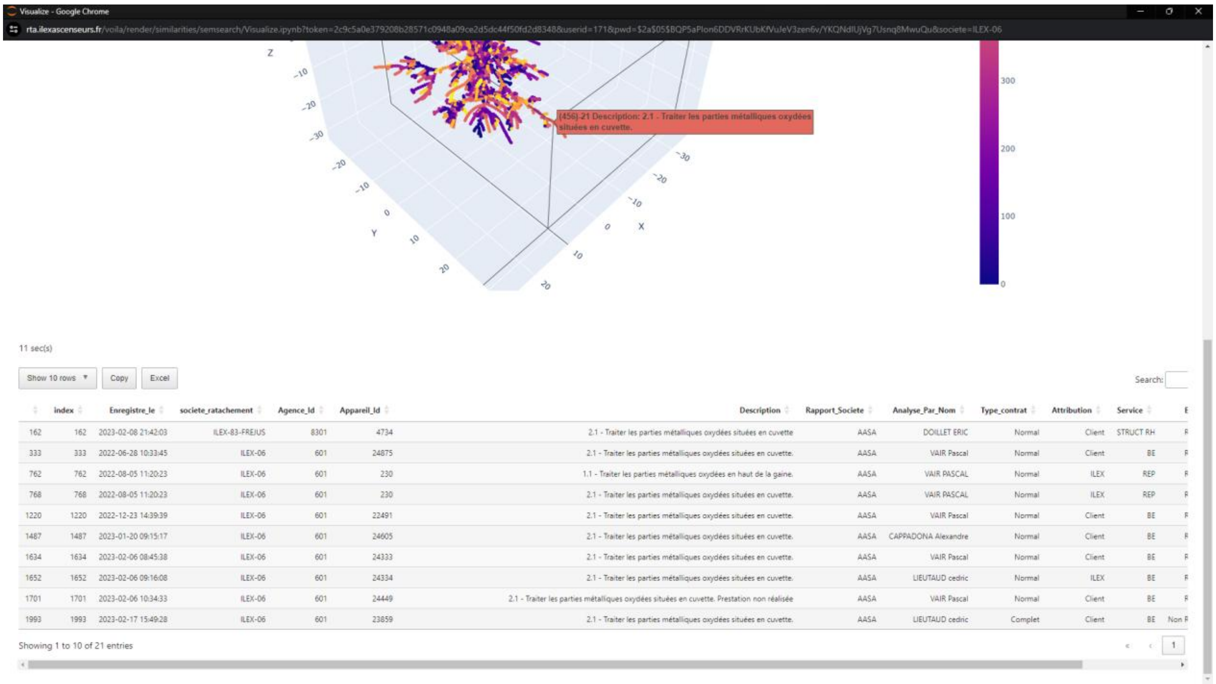
The "pan": it allows the tree to be translated. Choose a point on the tree and, while holding down the left mouse button, move it. The tree translates from the movement of the mouse

"Orbital rotation": choose a point, hold down the left mouse button, move the mouse. The tree rotates around the point

The "'turntable' rotation": same as before, but instead of rotating around a point, it is around an axis.

The "house": return to the initial positioning

When you click on a point, you can see the list of observations in the colored group



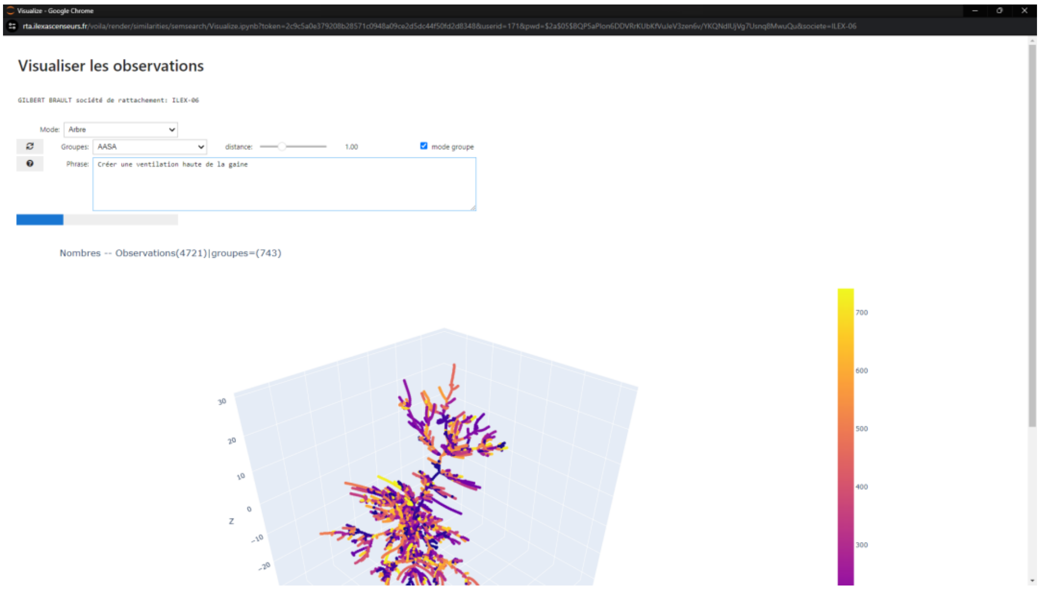
We can

- Expand the list with the "Show xx rows" tool
- Copy the table data to the clipboard
- Export the list of observations in the group to Excel
- Use pagination to see other pages in the group

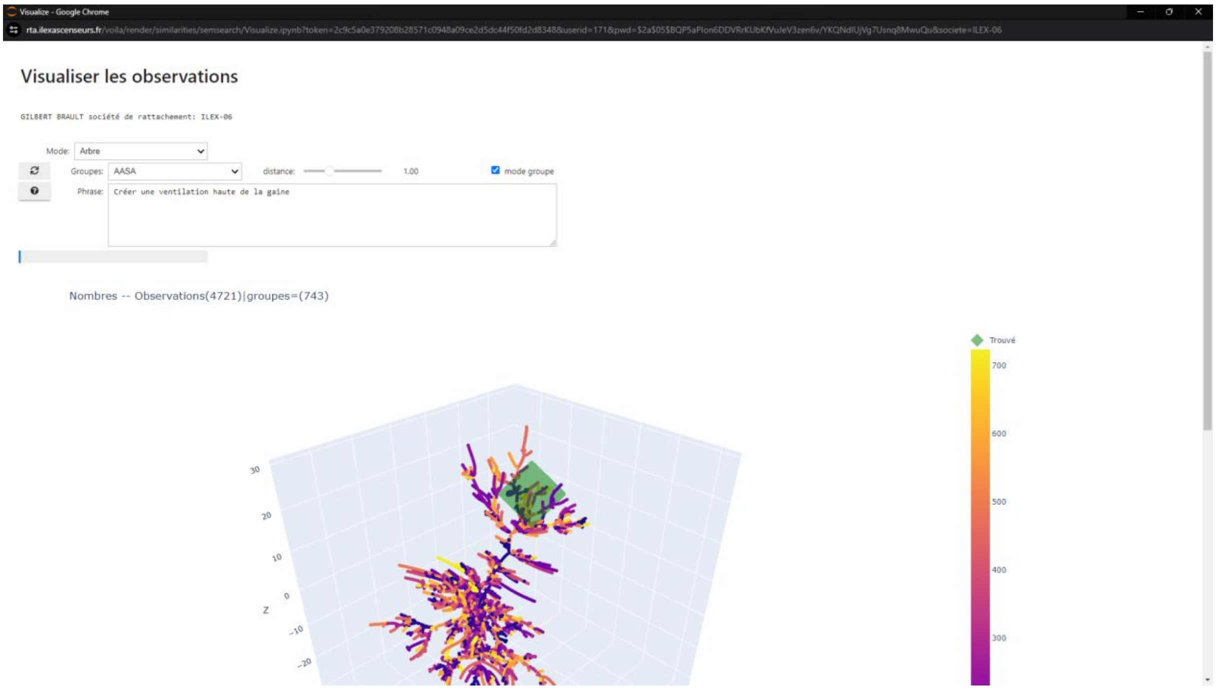
Research

To find, from a text of an observation, the place where it is in the tree

Enter text in the related box



Then click on the button? on the left.



A green paving stone, centred on the nearest existing observation, becomes visible.

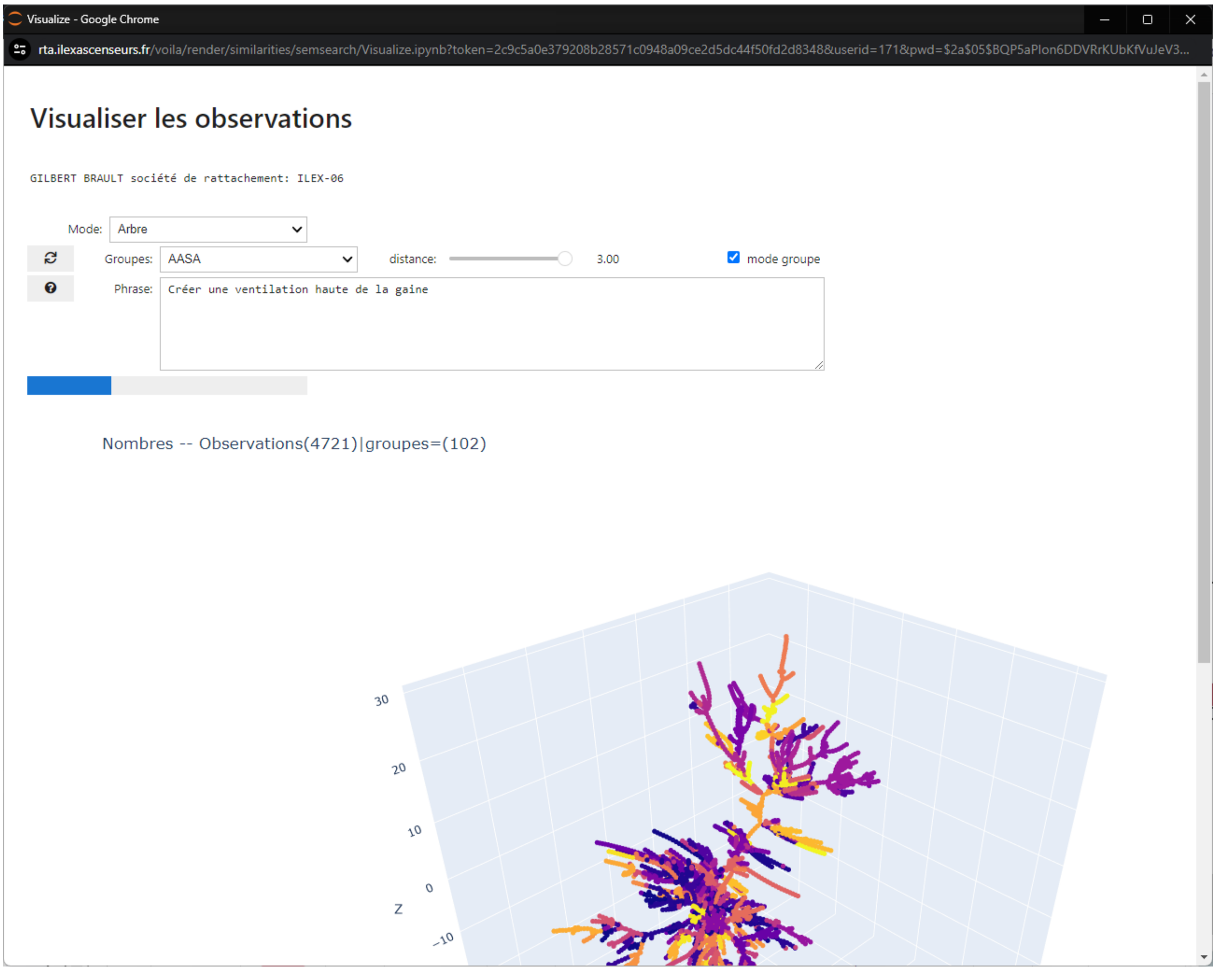
Just zoom in on the area to find the point.

Above the coloured bar we can also see the green square next to the word Trouvé. If you click on it with the mouse, the green area disappears and if you click again, it reappears.

This allows you to discover the area corresponding to the search

Choose the level of agglomeration

We move the distance cursor (in our case, it is set to 3). Click on the "Synchronization" button . The colour groups are calculated and the result is displayed by giving the number of groups and adjusting the point colours on the graph



Of course, the greater the distance, the fewer groups there are, but also the business verisimilitude is no longer there: it's just a mathematical artifact.

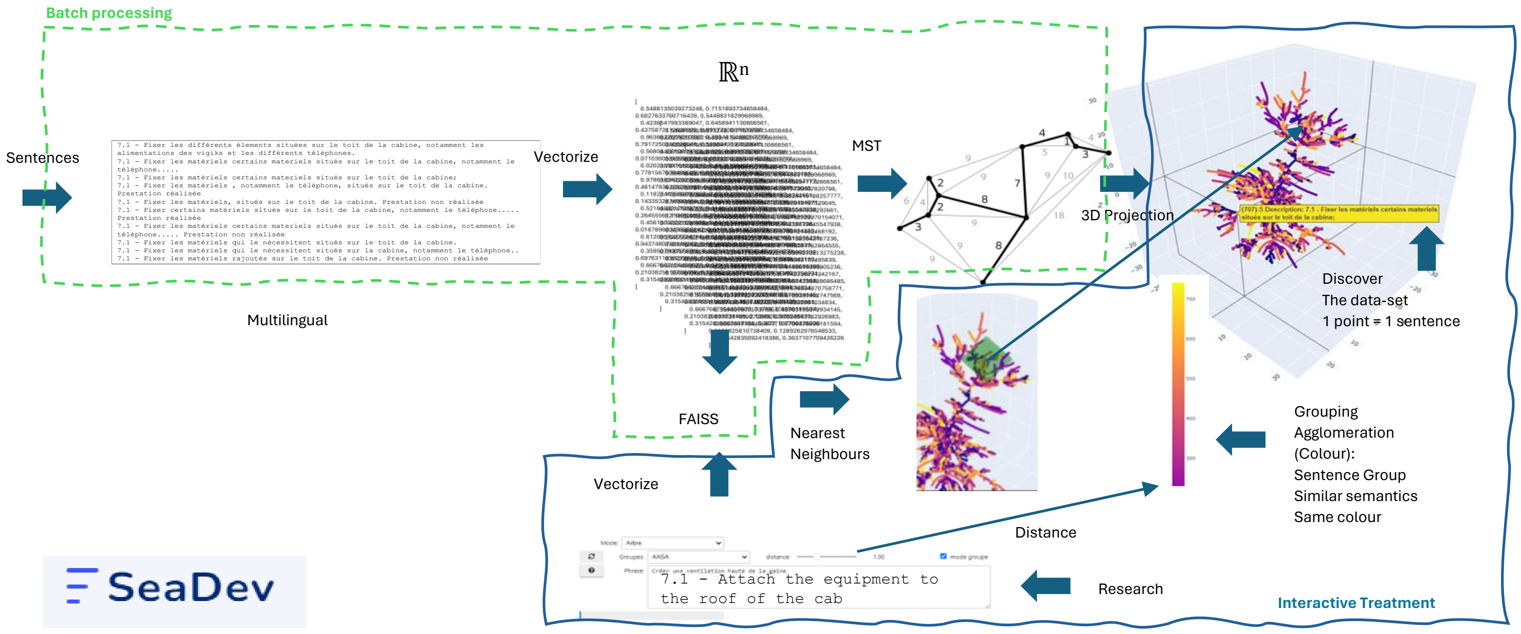
Note: the distance[1] displayed on the cursor is not the same as the distance used for the search. A ratio of the order of 3 can be estimated experimentally.

One can check by using the neighbour mode and looking at the lists that display the search distance.

[1] The search distance is calculated by the FAISS module while the color distance is calculated by the AgglomerativeClustering function.

Conclusion

Solution Architecture



- In the portion of the processing performed in <u>batch processing</u>
  - Observations are the sentences used as input to the system
  - These sentences are vectorized and the vectors are
    - Submitted to FAISS to create a vector database for research
    - Transformed into MST thanks to the distance matrix (a distance matrix is a graph)
    - MSTs are transformed into 3D coordinates (one point per observation) for later display
- In the <u>interactive part</u>
  - Discover the observations
    - The MST tree is presented in interactive 3D with a per-point observation
    - To see the text of the observation, simply hover the mouse over the point
    - When a point is selected, we obtain the list of the group obtained by agglomeration
  - Search for an observation
    - A text is entered to search for semantically similar observations
    - Thanks to FAISS, we perform the search calculation
    - The corresponding area can then be indicated on the 3D graph
  - Definition of agglomerate groupings
    - We set a distance
    - We then use an algorithm for calculating groups by agglomeration that associates all points within this distance by connexity.

Lesson learned?

- The classification and semantic search of sentences makes it possible to group observations that have the same semantics
- These groups of observations, if the distance criterion used to define them is correctly positioned, correspond to identical or similar work.
- The mathematical tools used are quite complex, but we were able to share the results with the men in the field thanks to a representation that allows us to interact with the men in the field.
    - Define the "size" of the groupings
    - To verify whether their semantic content is indeed equivalent for the skilled person by producing the list of semantically similar observations
        - Or from a point that we select on the 3D graph, and we produce all the observations of the group by agglomeration grouping.
        - Either by entering an observation text and searching for the nearest neighbours on the 3D graph yields observations that are semantically like the text entered for the search.
- We therefore have a tool that makes it possible to discover and standardize the list of tasks corresponding to the observations issued by a control office.