

Data Science Day

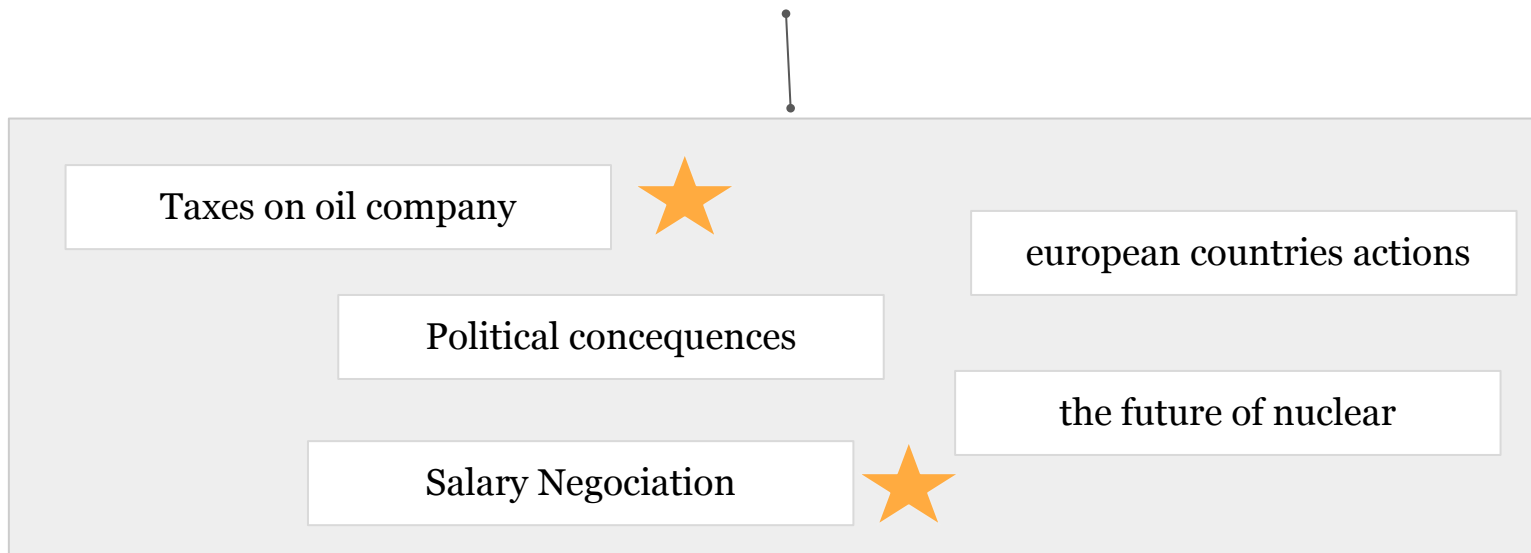
Semantic analysis at Le Figaro

21-10-2022

The challenge



Topic : **Energy crisis**

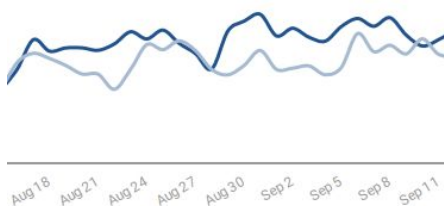


More qualifications for better actions

Having this information available allow :

DATA ANALYST

Produce more **granular analysis**



JOURNALISTS

Change **the way the information is displayed** on site

Make better article **recommendations**

ÉDITION SPÉCIALE UKRAÏNE

Guerre en Ukraine : tous nos articles pour comprendre

Quelle est la situation actuelle en Ukraine ? Comment avancent les pourparlers diplomatiques ? Pourquoi la Russie de Vladimir Poutine a-t-elle décidé d'envahir l'Ukraine le 24 février dernier ? Dernières nouvelles, cartes, reportages, décryptages, analyses, interviews... Toutes nos explications.



Le président ukrainien Volodymyr Zelensky et le président russe Vladimir Poutine. L'figaro

La rédaction vous conseille

- Carburants: la CGT reconduit la grève sur les 5 sites de TotalEnergies
- Crise des carburants: minoritaire, la CGT peut-elle malgré tout maintenir le blocage des raffineries? 🗳
- Grève TotalEnergies: un compromis trouvé avec des syndicats majoritaires, sans la CGT
- Après les raffineries, la grève s'étend dans le nucléaire chez EDF 🗳

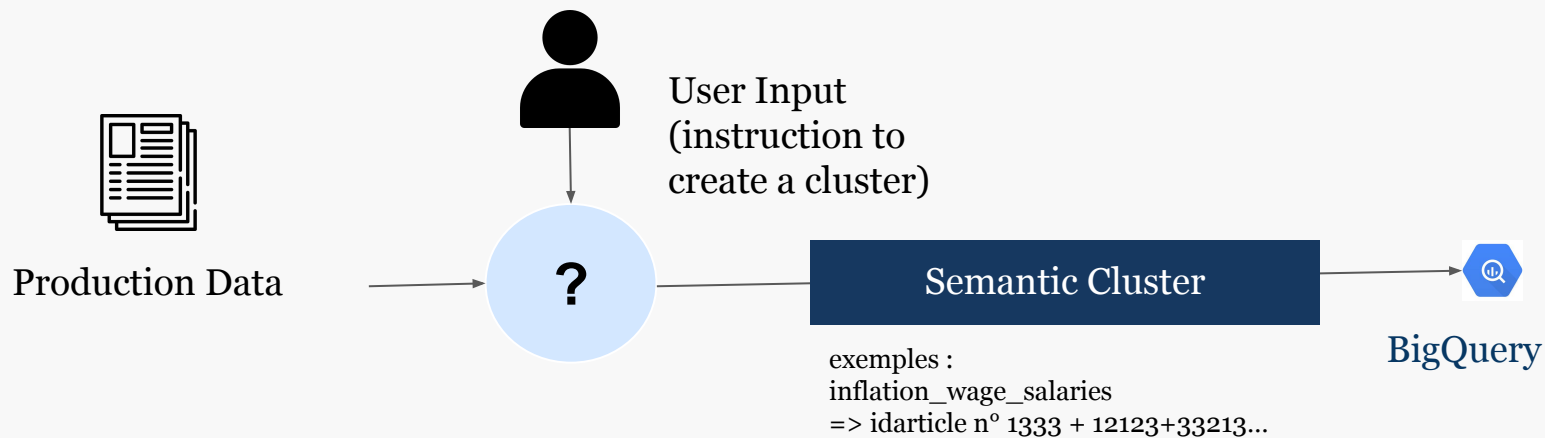
|The right level of informations is not always available

Exemple : creation of a clusters on **SALARY NEGOTIATION**



Creation of semantic clusters can be complex and knowledge is not shared

| A missing tool is needed !

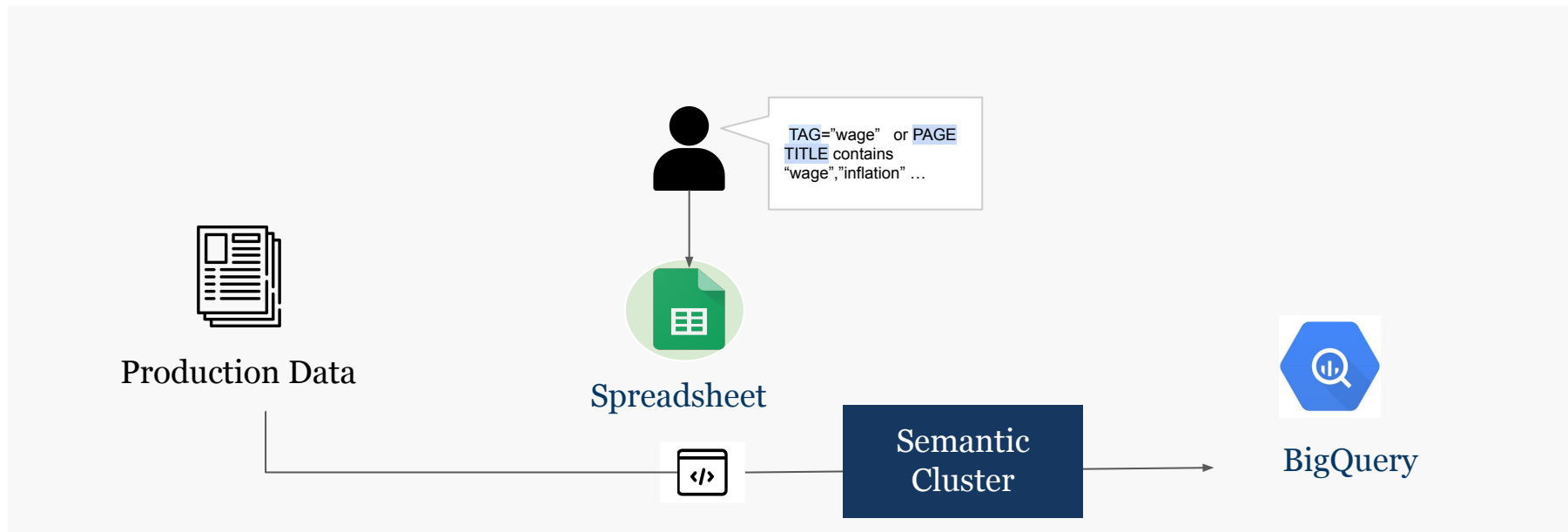


- Create cluster **thanks to user input**
- **Push the result to BI tool** (to make the cluster available as filter, dimension...)

A first approach



A first attempt at creating semantics clusters



| Result

- + A common “source of truth” share among BI Tool

 Looker Studio

pagetitle	clusters sem.	Visites ▾
Halte à la tyrannie de la bienveillance!	Psycho	20 522
Isabelle Filliozat : «Notre monde est trop hygiénique, trop prudent, trop protecteur»	Famille - Parentalité	6 082
«Le marché du "love care", c'est 7 milliards d'euros», comment entreprises et associations veulent sauver le couple français	Famille - Parentalité	6 039

- Rules can be **too complex** to write and to **maintain**
- The user need to **know the architecture** of the production system

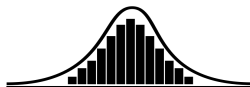
1. Design like a **search engine**

End user start with an intuitive /
simple request



2. Return **insights**

End user start with an intuitive /
simple request



3. Allow an **iterative process**

Users can easily refine results of the search by filtering,
selecting subset

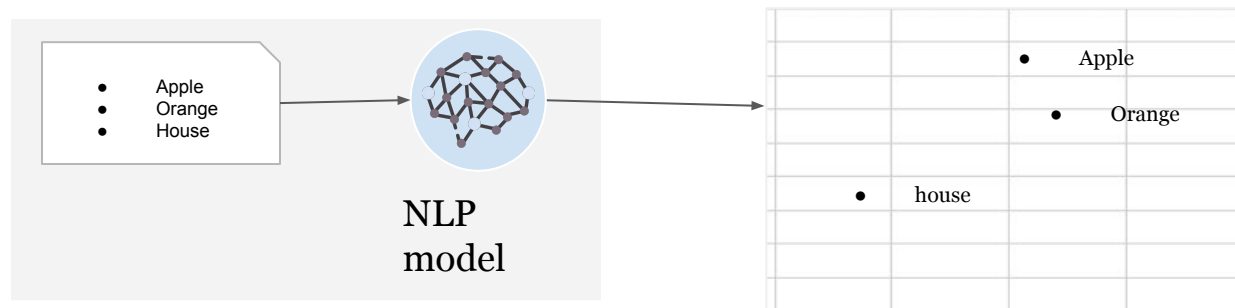


A focus on embedding



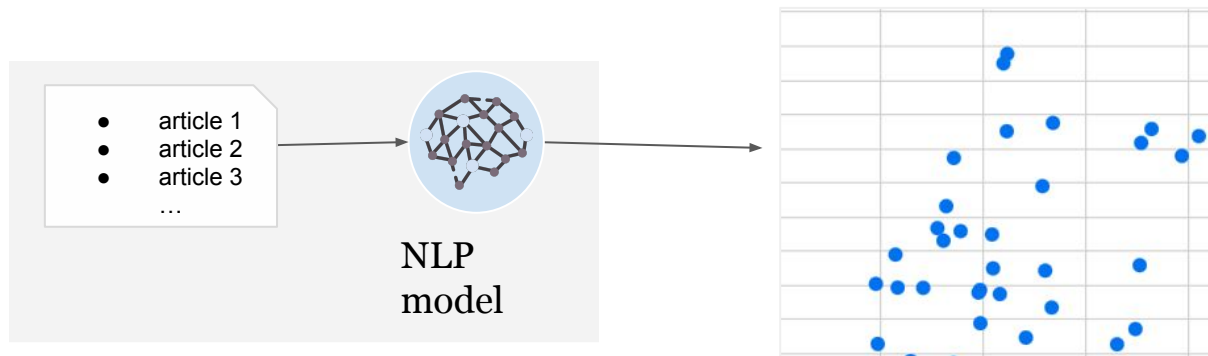
Introduction

Input words get a **numerical value** that reflect their semantic values



| How to adapt it to Le Figaro context

A semantic interpretation of the production is created



| How the model is train ?

We use a model train on huge french dataset
and **finetuned with proprietary data**



The open source part of the model :

Trained on huge data set (wikipedia)
on french language

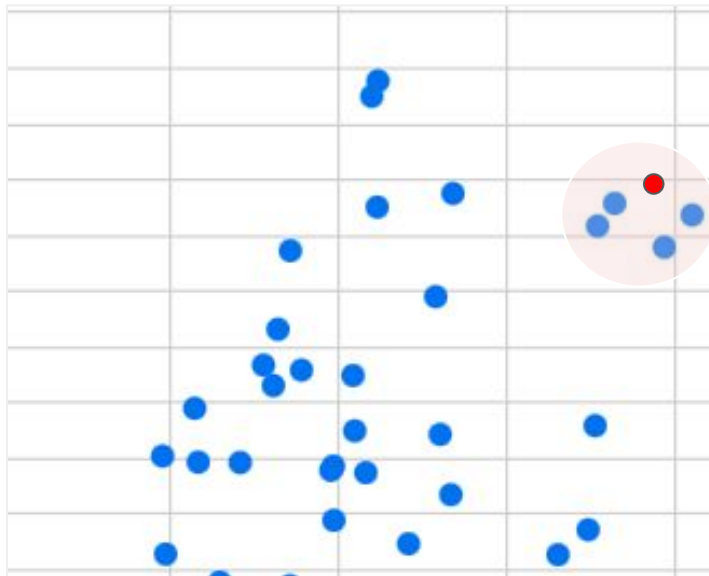
La rédaction vous conseille

- Carburants: la CGT reconduit la grève sur les 5 sites de TotalEnergies
- Crise des carburants: minoritaire, la CGT peut-elle malgré tout maintenir le blocage des raffineries? 🏹
- Grève TotalEnergies: un compromis trouvé avec des syndicats majoritaires, sans la CGT
- Après les raffineries, la grève s'étend dans le nucléaire chez EDF 🏹

The proprietary part of the model :

Trained on pair of recommended
article + summaries

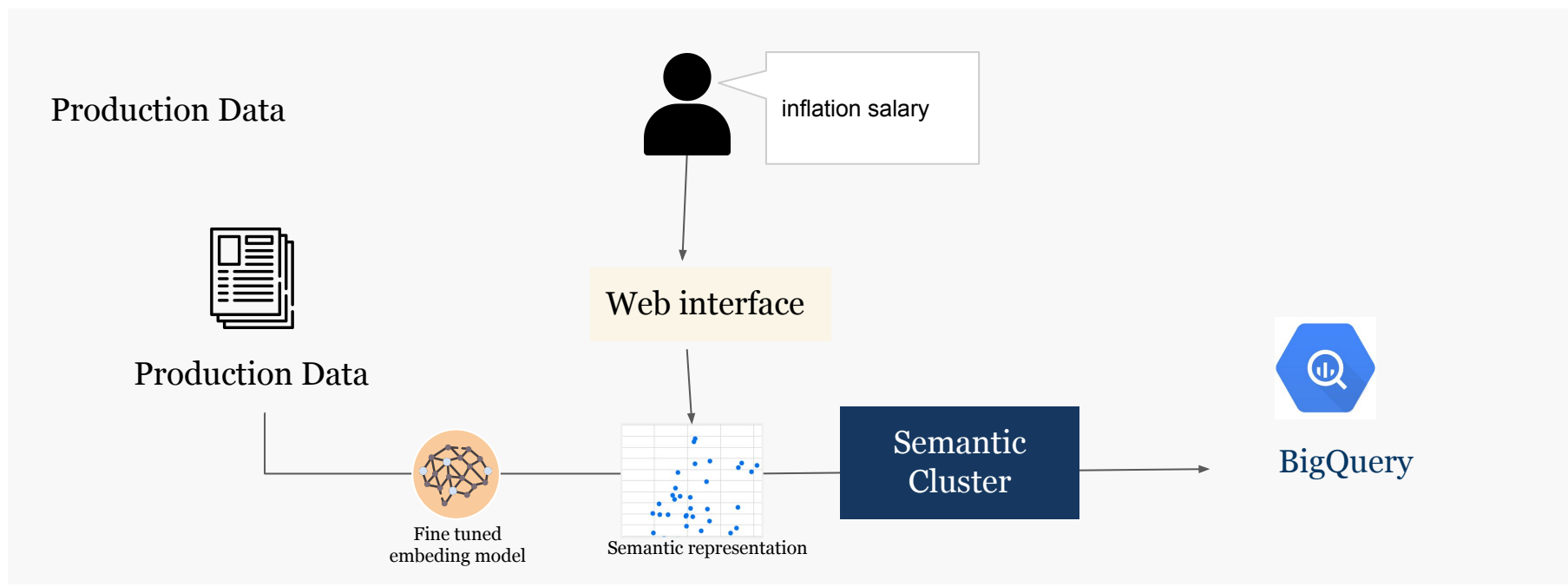
| A basis for a search engine



● keyword = “inflation”

● articles

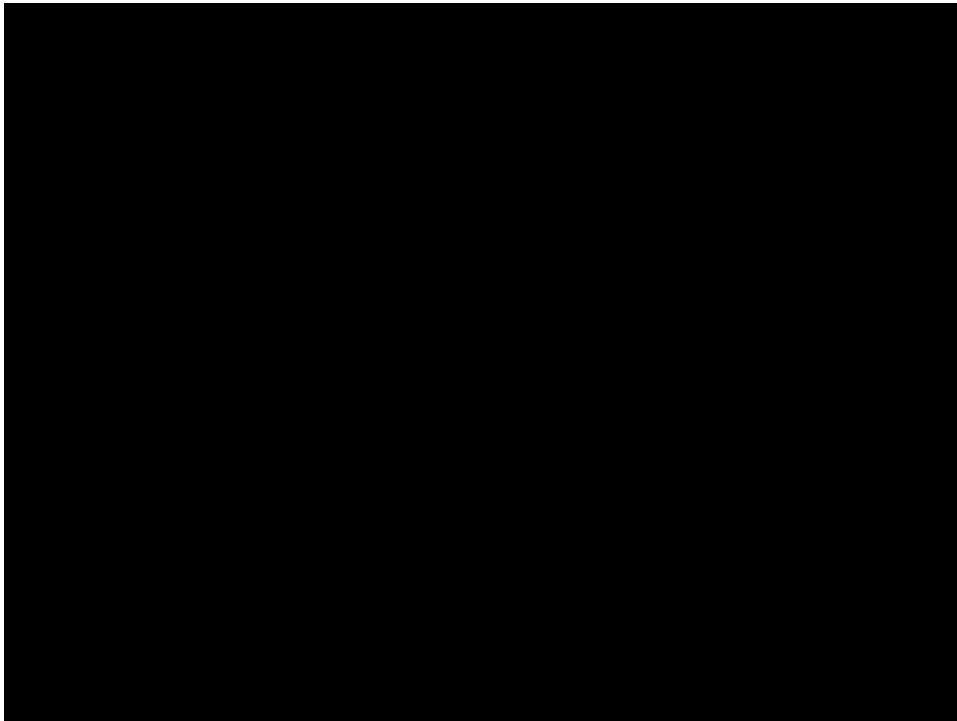
| The new schema



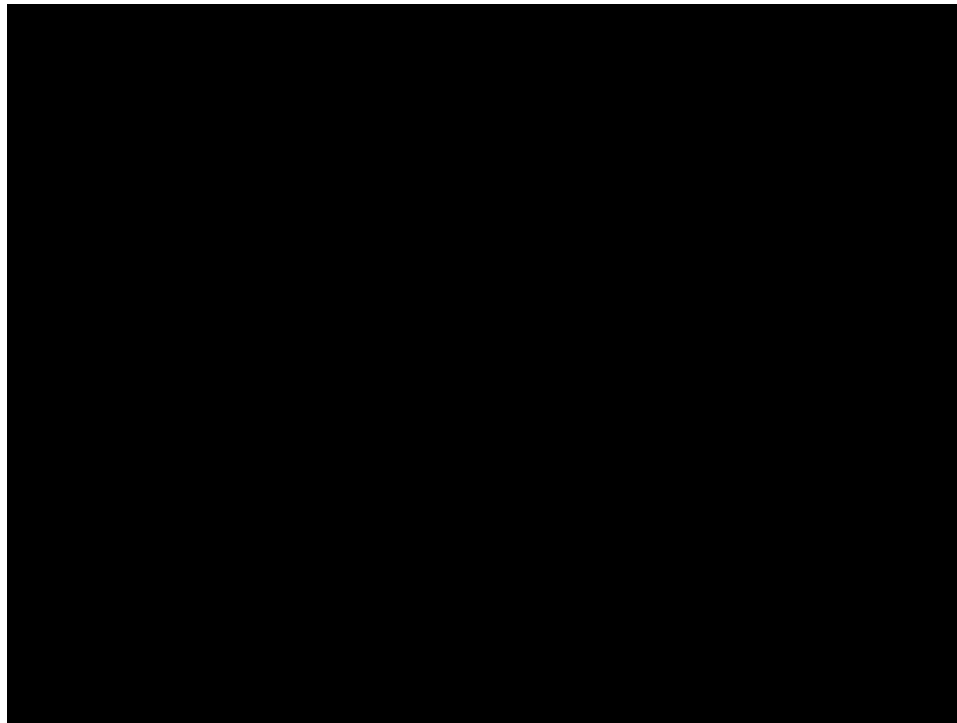
Demo



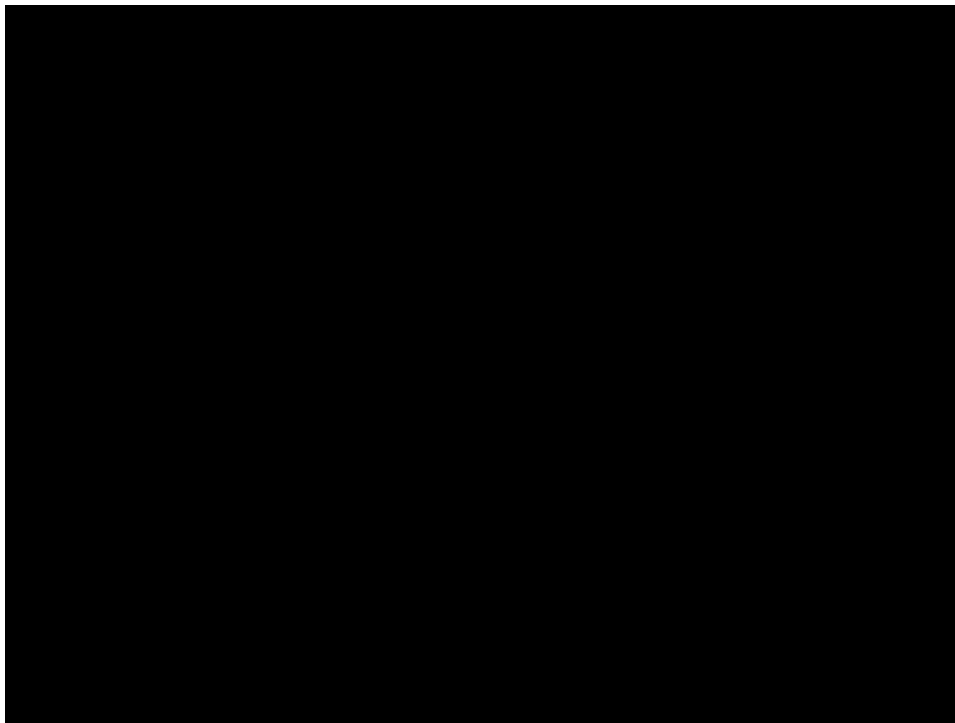
| The configuration



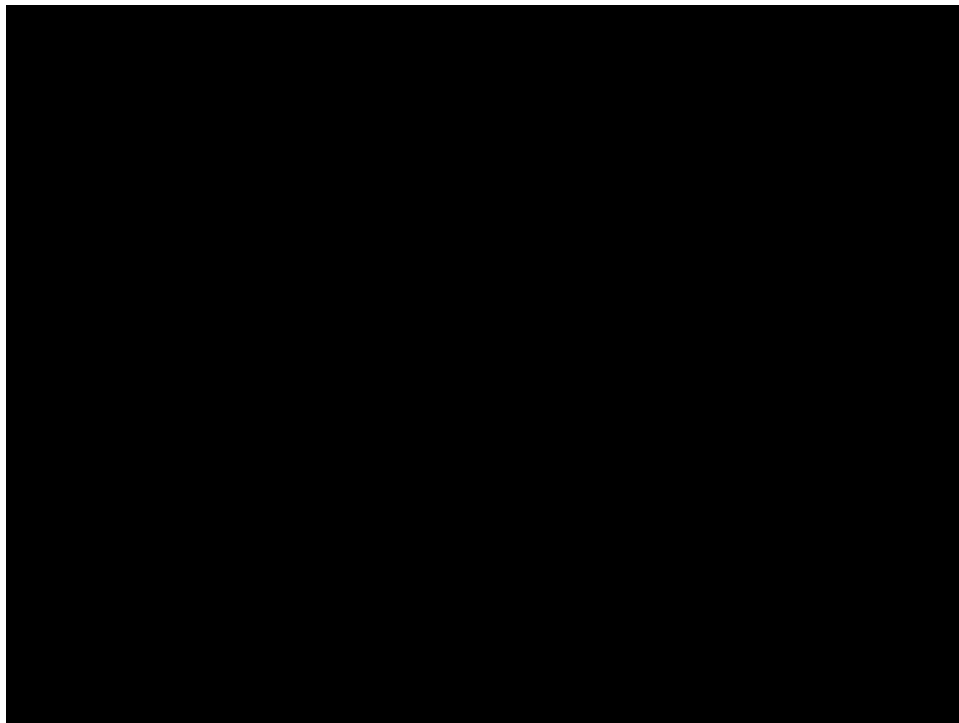
| The analysis



| The clusters



| Semantic cluster creation



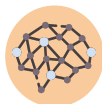
Next steps



| Next challenges

Update automatically the semantic clusters

A new article is produced



classification
model



prob.
“inflation_salaries_negociation”

prob.
“inflation_corporate_taxes”

prob.
“taxes_increase_local”

0.0

0.1

0.9

Management and governance

Hypothesis : create 2 roles ADMINISTRATOR and VIEWER

Thanks you

