

Business Intelligence

BENLARBI Samy - RISSER-MAROIX Olivier





Sommaire

- I. Introduction
- II. Datawarehouse : De la conception jusqu'à l'analyse
 - A. Modélisation
 - B. Analyses
- III. Datamining : Diverses méthodes pour plusieurs résultats
 - A. TSNE : t-distributed stochastic neighbor embedding
 - B. Clustering sur PCA
 - C. Règles d'association
- IV. Conclusions

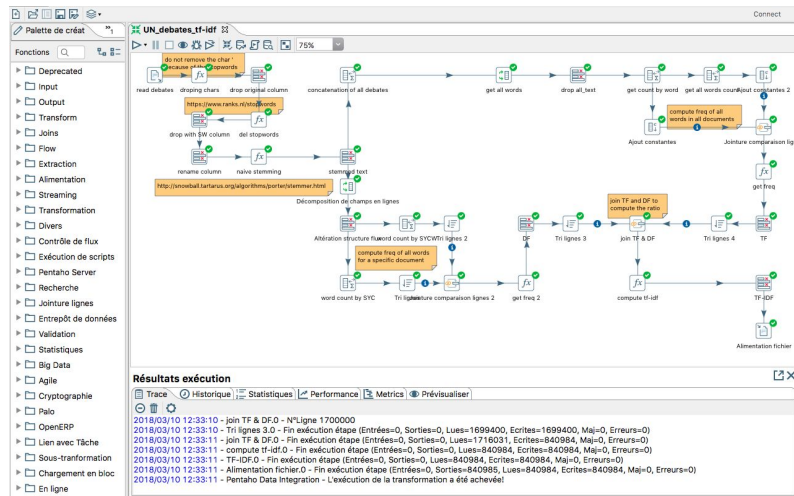
I. Introduction

- Étudier les discours de l'Organisation des Nations Unies depuis les années 1970
 - Tenter de corréler ces données à diverses informations de la base world
- Utiliser des méthodes de data mining pour observer des phénomènes au niveau des mots employés

II. Datawarehouse : de la conception jusqu'à l'analyse

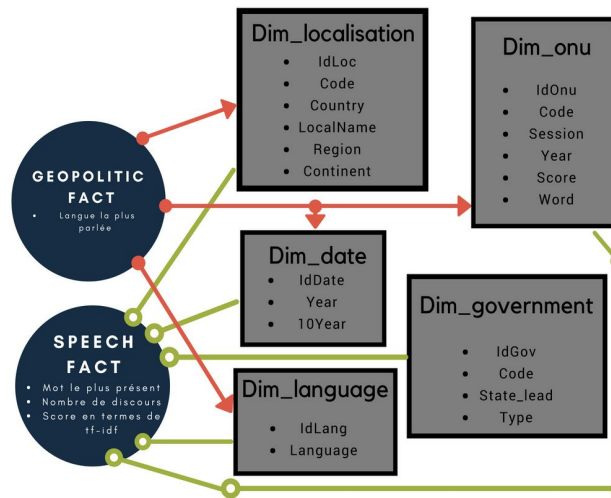
A. Modélisation

D'un schéma conceptuel jusqu'au schéma logique dénormalisé



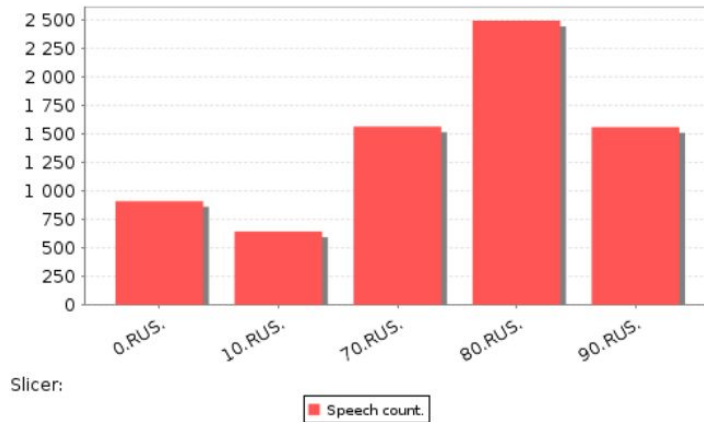
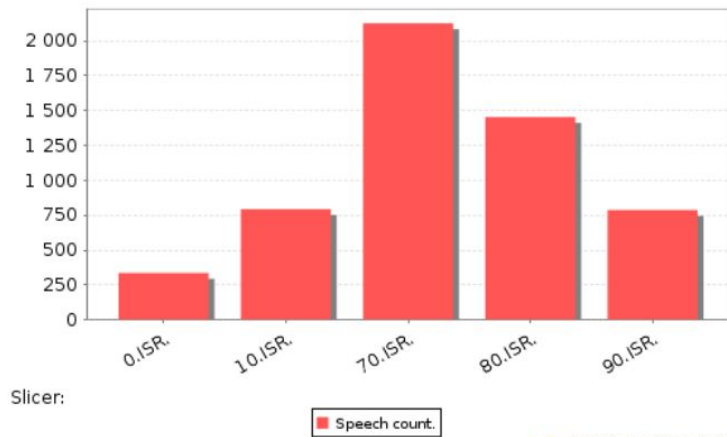
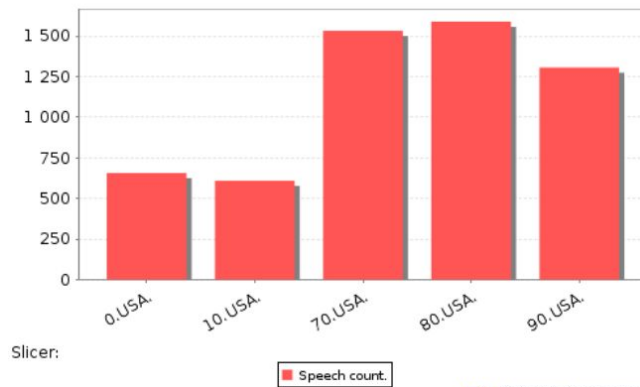
```
Dim_localisation(IdLoc,Code,Country,LocalName,Region,Continent)
Dim_language (IdLang,Language)
Dim_date (IdYear,Year,10Year)
Dim_onu (IdOnu,Code,Session,Year,Score,Word)
Dim_government (IdGov,Code,State_lead,Type)

fact_speech (#IdLoc,#IdLang,#IdYear,#IdOnu,#IdGov, Best_word, Best_score,
Speech_count)
fact_geopolitics (#IdLoc,#IdLang,#IdYear,#IdOnu, Most_talked_language)
```



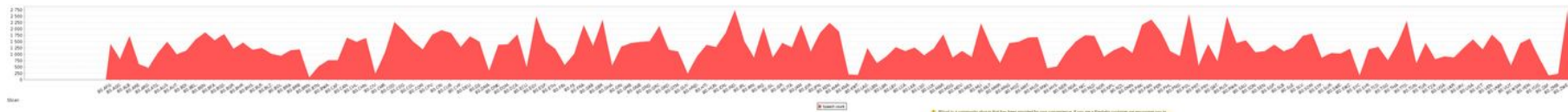


B. Analyses

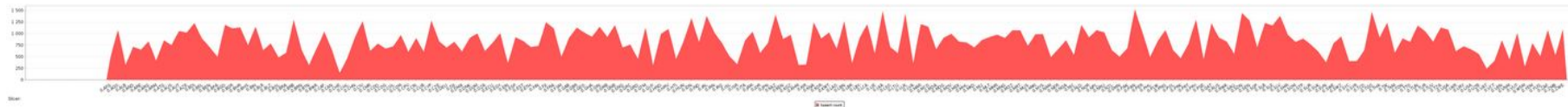




Speech count suivant le langage



1980



2000



Score et profiling politique

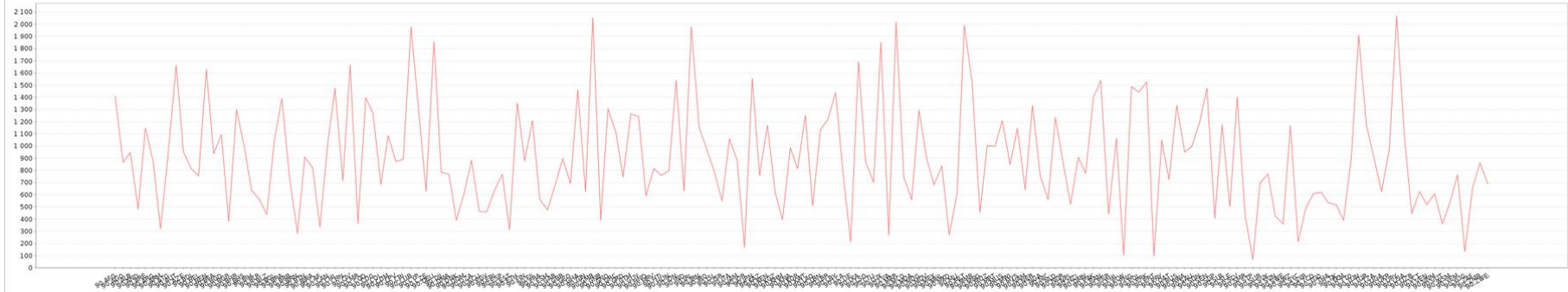
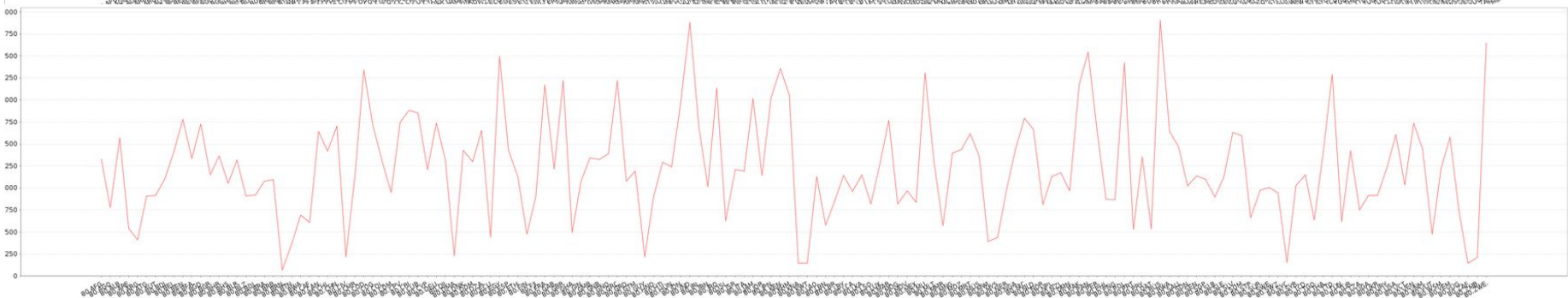
MDX Query Editor

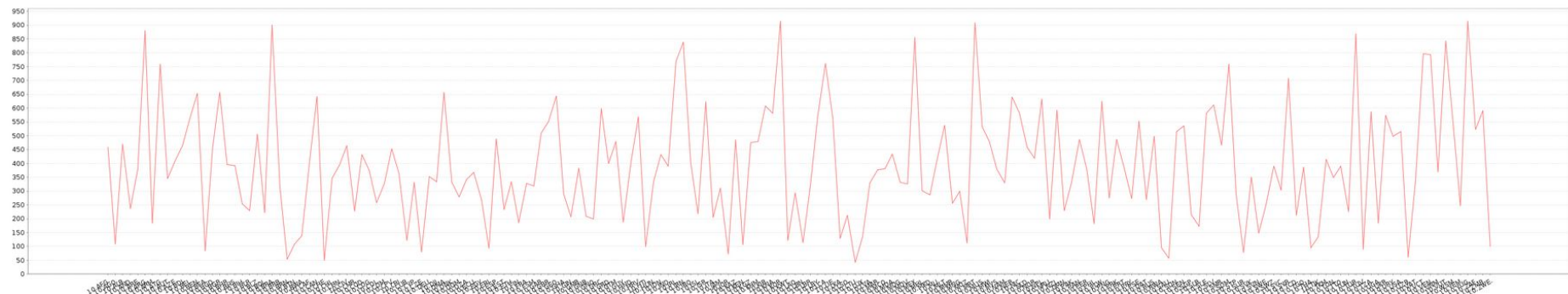
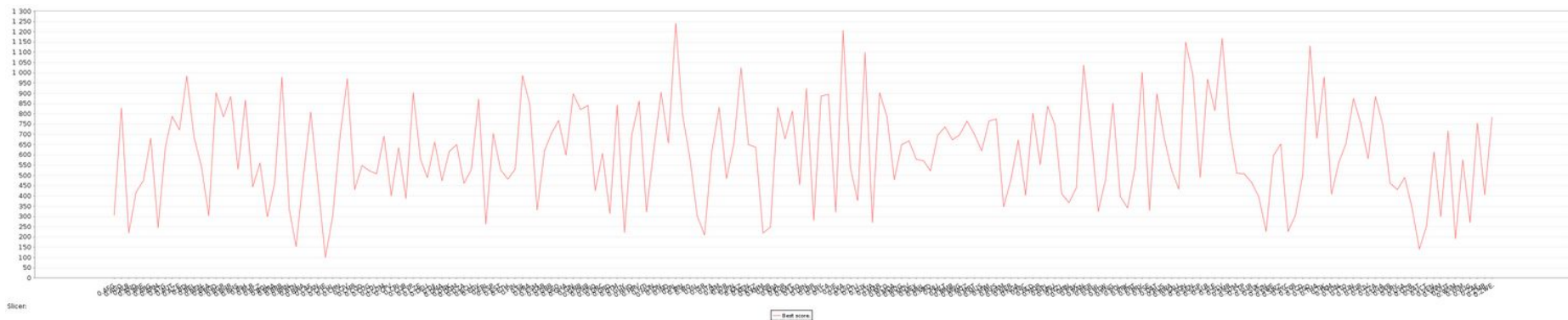
MDX Query Editor

```
select NON EMPTY {[Measures].[Best score]} ON COLUMNS,  
  NON EMPTY Hierarchy(Union(Union(Union(Union(Crossjoin({[Date.10Year].[0]}, [Geography].[All Geographys].Children),  
Crossjoin({[Date.10Year].[10]}, [Geography].[All Geographys].Children)), Crossjoin({[Date.10Year].[70]}, [Geography].[All Geographys].Children)),  
Crossjoin({[Date.10Year].[80]}, [Geography].[All Geographys].Children)), Crossjoin({[Date.10Year].[90]}, [Geography].[All Geographys].Children)))  
ON ROWS  
from [bi_speech]
```

Apply

Revert





III. Mining & Visualization

Problèmes :

+800k discours

Mesures qualitatives

Hi Olivier

I noticed that you recently ran a RapidMiner Studio process with over 10,000 rows of data. The free version of RapidMiner Studio will only use up to 10,000 data rows. If you need more data, here are a couple of options:

1. You can purchase a paid license of RapidMiner Studio. Pricing is [available on our website](#) or email me back and I can help you understand the options.
2. For each [friend or colleague you refer](#), you get an additional 10,000 data rows for free, up to a total of 50,000 rows.

Here's one additional option: I could set you up with an extended trial of RapidMiner Studio Large, but I'd want to have a quick chat first to understand your goals. Reply back to this email if you are interested.

Regards,

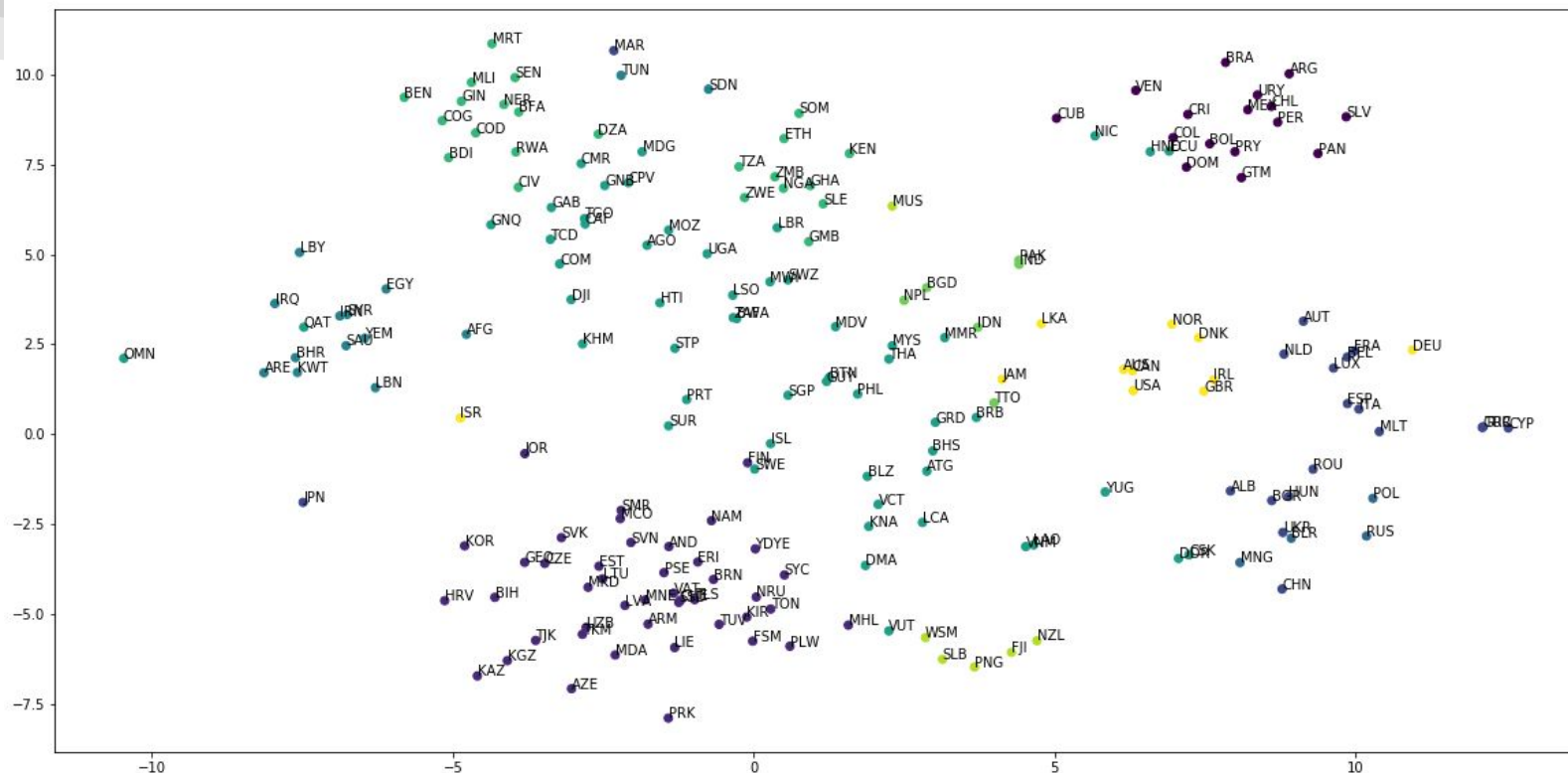
Andrew

Andrew Julio | Business Development | RapidMiner

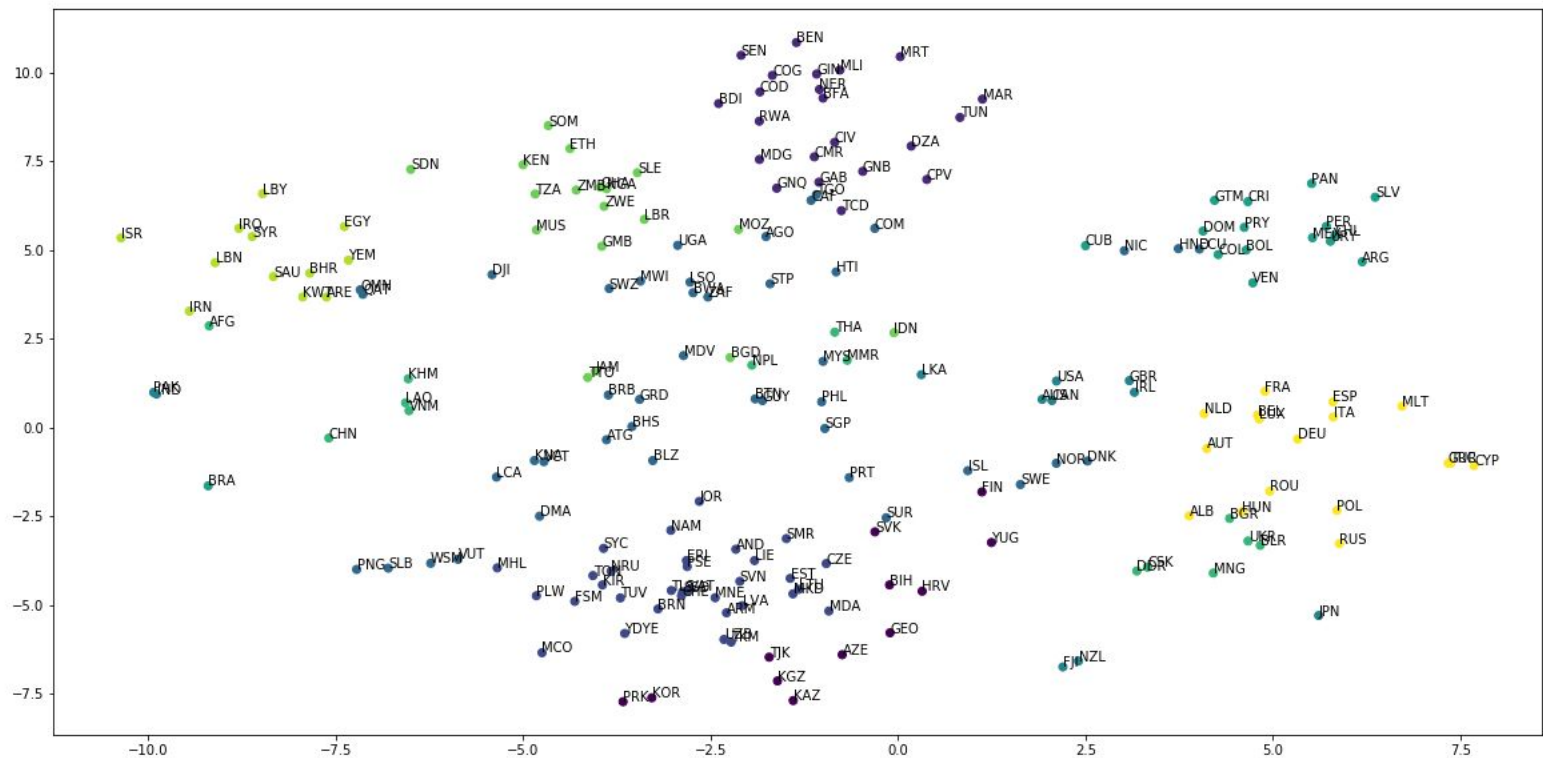
Phone: | Email: ajulio@rapidminer.com



A. TSNE : cartographie en fonction des discours



Visualisation **TSNE** sur les **embeddings** des pays créés à partir d'une **PCA** sur leur **discours**
Coloration par **KMean** sur les embeddings



B. Clustering : carte des thèmes



Word Cloud

USA - Russie

europa europa
know know
let let
grow grow
course course
european european
disarma disarm
sai sai

Amérique Latine - Pays de la Drogue

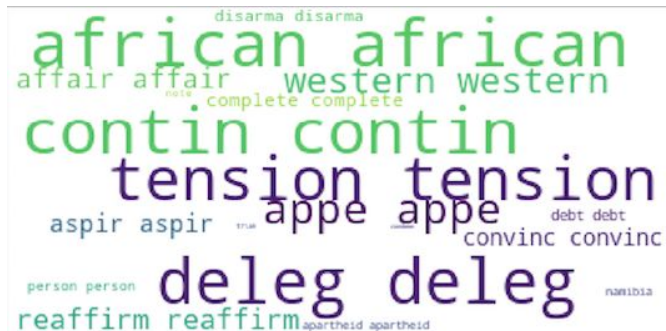
american american
true true
market market
latin latin
purpose purpose
limit limit
democraci democraci

Asie

deleg deleg
disarma disarm
asia asia
african african
iran iran
land land
return return
palestine palestine



Afrique - Europe



Nord

Ouest



Sud

Est





C. Association de concepts

will, oil	receiv, remark, report, short, signif, strongli, undertak...	0.737	∞	1.737	0.737
will, receiv	oil, remark, report, short, signif, strongli, undertake...	0.737	∞	1.737	0.737
oil, receiv	will, remark, report, short, signif, strongli, undertak...	0.737	∞	1.737	0.737
will, oil, receiv	remark, report, short, signif, strongli, undertake, zo...	0.737	∞	1.737	0.737
will, remark	oil, receiv, report, short, signif, strongli, undertake, ...	0.737	∞	1.737	0.737
oil, remark	will, receiv, report, short, signif, strongli, undertake, ...	0.737	∞	1.737	0.737
will, oil, remark	receiv, report, short, signif, strongli, undertake, zone	0.737	∞	1.737	0.737
receiv, remark	will, oil, report, short, signif, strongli, undertake, zo...	0.737	∞	1.737	0.737
will, receiv, remark	oil, report, short, signif, strongli, undertake, zone	0.737	∞	1.737	0.737
oil, receiv, remark	will, report, short, signif, strongli, undertake, zone	0.737	∞	1.737	0.737
will, oil, receiv, remark	report, short, signif, strongli, undertake, zone	0.737	∞	1.737	0.737
will, report	oil, receiv, remark, short, signif, strongli, undertake...	0.737	∞	1.737	0.737
oil, report	will, receiv, remark, short, signif, strongli, undertak...	0.737	∞	1.737	0.737

Association Rules

[act] --> [women] (confidence: 0.889)

[action] --> [women] (confidence: 0.889)

[agenda] --> [women] (confidence: 0.889)

[area] --> [women] (confidence: 0.889)

[assembli] --> [women] (confidence: 0.889)

[bas] --> [women] (confidence: 0.889)

[believe] --> [women] (confidence: 0.889)

[call] --> [women] (confidence: 0.889)

[come] --> [women] (confidence: 0.889)

[commit] --> [women] (confidence: 0.889)

[common] --> [women] (confidence: 0.889)

[commun] --> [women] (confidence: 0.889)

[condition] --> [women] (confidence: 0.889)

[consensu] --> [women] (confidence: 0.889)

[cooper] --> [women] (confidence: 0.889)

[council] --> [women] (confidence: 0.889)

[countri] --> [women] (confidence: 0.889)

[cre] --> [women] (confidence: 0.889)

[demand] --> [women] (confidence: 0.889)

[determin] --> [women] (confidence: 0.889)

[dialogue] --> [women] (confidence: 0.889)

[differ] --> [women] (confidence: 0.889)

[econom] --> [women] (confidence: 0.889)

[end] --> [women] (confidence: 0.889)

[even] --> [women] (confidence: 0.889)

[fact] --> [women] (confidence: 0.889)

[failure] --> [women] (confidence: 0.889)

[first] --> [women] (confidence: 0.889)

[framework] --> [women] (confidence: 0.889)

[g] --> [women] (confidence: 0.889)

[glob] --> [women] (confidence: 0.889)

[go] --> [women] (confidence: 0.889)

[govern] --> [women] (confidence: 0.889)

[gre] --> [women] (confidence: 0.889)

[hand] --> [women] (confidence: 0.889)

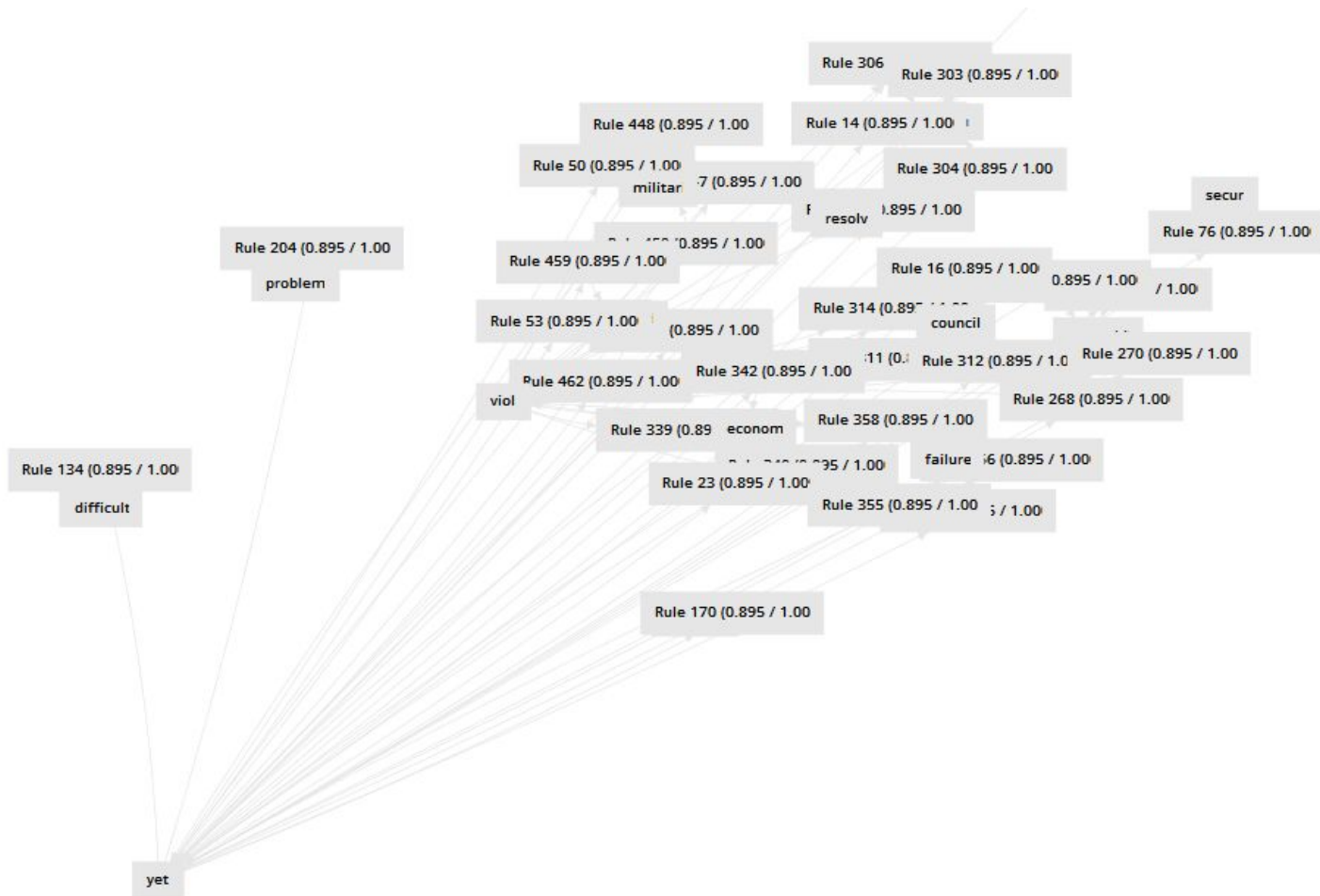
[help] --> [women] (confidence: 0.889)

[import] --> [women] (confidence: 0.889)

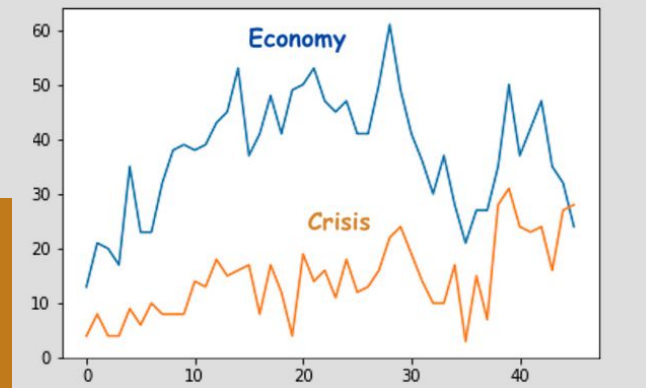
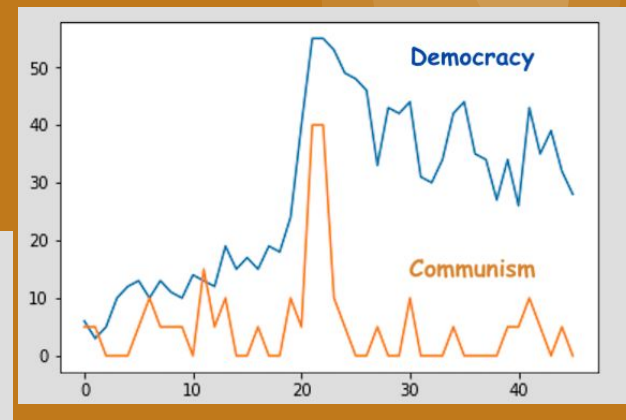
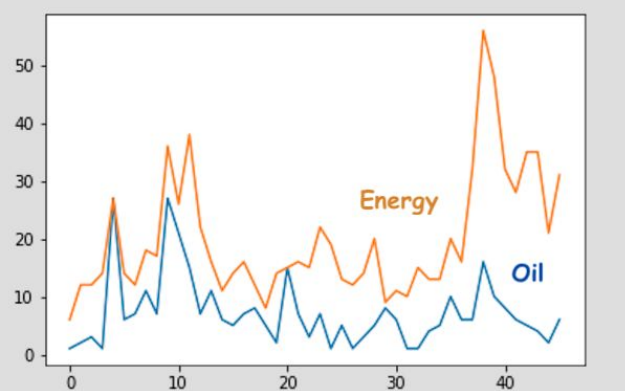
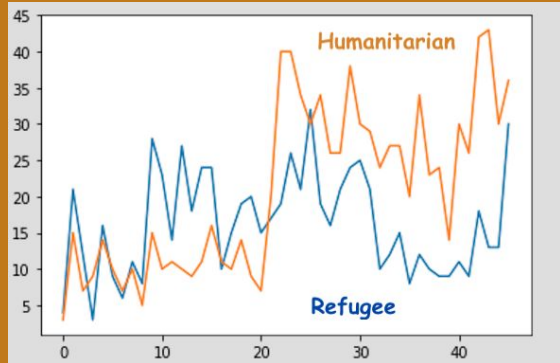
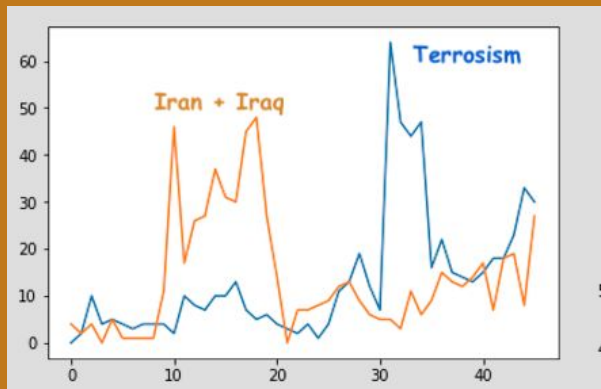
[inde] --> [women] (confidence: 0.889)

[individu] --> [women] (confidence: 0.889)

[institution] --> [women] (confidence: 0.889)



Corrélations entre Mots *



* ajouter +1970 sur les abscisses

Conclusions

LES MOTS PARLENT !

