La DATA

dans tous ses états ...

Data quoi ?

Pourquoi faire

Enjeux de la data

Comment s'en servir

Exemples et cas d'**usage**

[Questions & Atelier]

nom de code:

c24b

Constance **de Quatreb**arbes

geek girl, 32 ans

- Ingénieuse, datascientiste *(7 ans)*
- Formatrice *(HackinScience, HackYourPhD, Master NUMI)*
- Chercheuse *(SHS, STS, SIC, DH,IT)*

DATA quoi ?

#Type

#Taille

#Source

#Format

Une donnée est une unité minimale d'information, figée, transmissible
Elle varie selon son format, sa source, son mode de stockage

Sauriez vous m'en citer?

Pourquoi faire

#Profiling

#ROI

#Targeting

#Optimisation

Le traitement et l'analyse de données intervient dans tous les secteurs
et modifient de nombreux métiers (veille, communication, marketing, ventes)

Quelques examples de dataintelligence

Comment faire ?

Data Mining

collecter

Data Management

stocker

Data Sciences

calculer

Data Visualisation

voir

Data Mining

***collecter des données** depuis des **sources** multiples*

(blogs, websites, forums, social networks, sensors,applications, API, Databases, ...)

avec des outils différents:

- ETL
- Convertors
- Parsers
- Extractors
- WebCrawler
- WebScrapper

Using python libraries, web & markup language, software tools for data types

HTML	DOM	CSS	img	txt
Xpath	XML	JS	sound	video
requests	spynner	datatypes		
scrapy	splinter	file I/O		
crawtext	pyquery	regex	pdf2text	
beautifulsoup	lxml	pdfminer	sed	
pytesseract	scikit-image	scikit-video		

*pour construire des **#datasets***

Data Management

organiser, stocker et indexer les données

en utilisant :

- Database management system
- Database Query Langages
- Files structured with MarkupLanguages

such as:

- MySQL
- PosgreSQL
- SQLite
- MongoDB
- ElasticSearch
- Redis
- SparQL
- JSON
- CSV
- XML

pour fouiller, qualifier et requeter les ***#datasets***

Data Sciences

experimenter des hypotheses,
traiter les datasets, appliquer des #algorithmes

such as :

- Classification
- Ranking
- Clustering
- Machine Learning
- Natural Language Processing

using **python libraries**

pandas	matplotlib		numpy
scipy	scikit-learn		networkx
pattern	nltk	difflib	textblob
maths	itertools		...

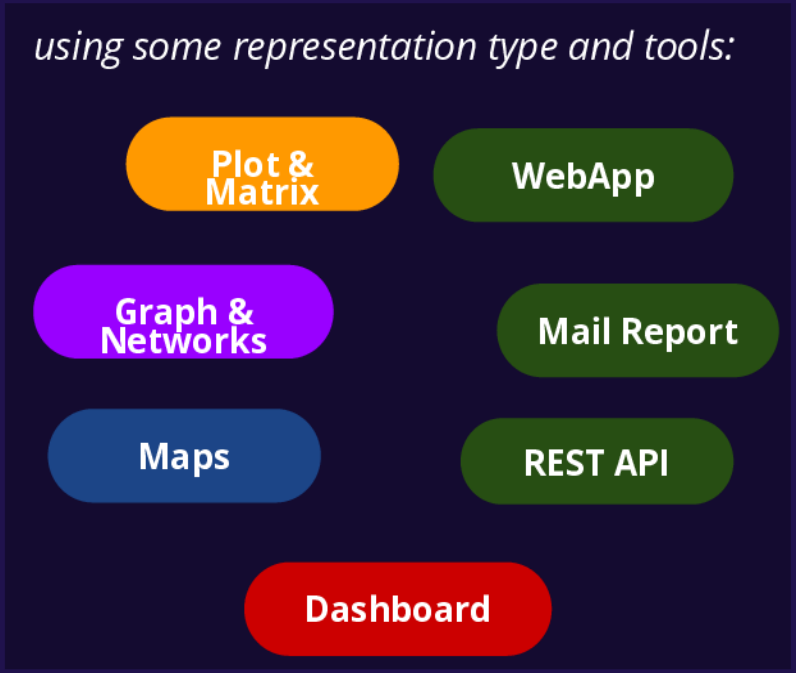
pour transformer les data en

#information

Data Visualisation

*présenter des **vues**, montrer des **resultats**,*
*cartographier l'**information***

using some representation type and tools:

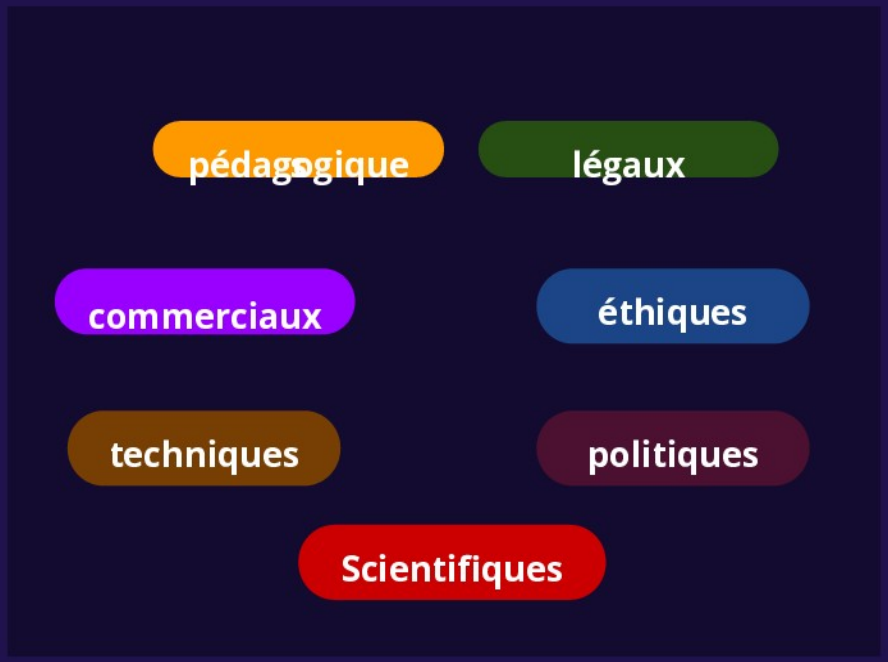


using multiples librairies:

pyplot pylab jupyter-nb pandas
d3.js gephi pygraphviz sigma.js
leaflet.js folium openlayers jinja2
boostrap angular bottle wordpress
django kibana graphite caravel

pour comprendre l' #information

Quels Enjeux ?



un changement de #paradigme

Exemples

- E-veille E-marketing
- Etudes de marché
- Système de recommandation
- Open Data Sciences

E-Veille E-marketing

Consulting and web agencies needs data processing for monitoring online reputation of brand

Filtering and extracting pertinent informations from

Online videos

A search engine for DailyMotion and Youtube

Online images

A search engine for Instagram and FlickR

Social networks

A search engine for Twitter, Facebook, Google+

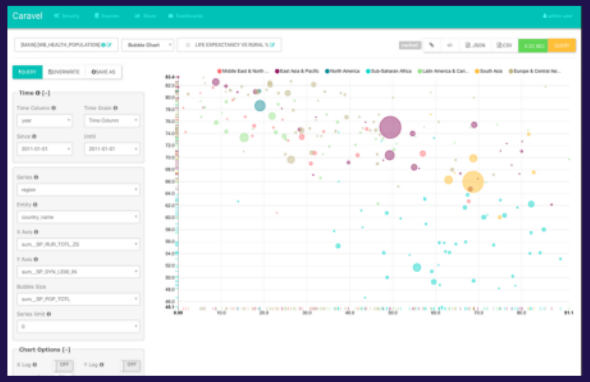
Websites, blogs and forums

A target oriented web-crawler for html content

Building qualified datasets and integrate it to internal monitoring systems

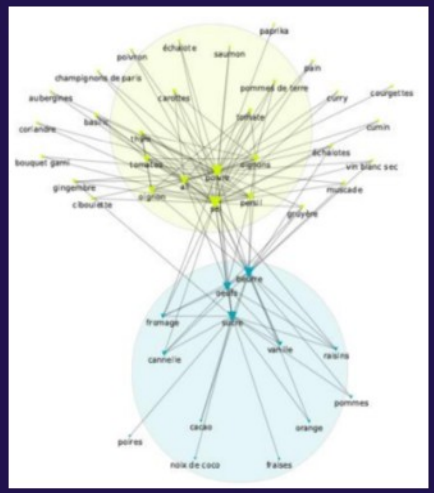
Develop automatic classification systems and ranking according to KPI

Send automatic report and alerts



Conseil et études de marchés

Consulting and web agencies needs data analysis for market and strategic studies

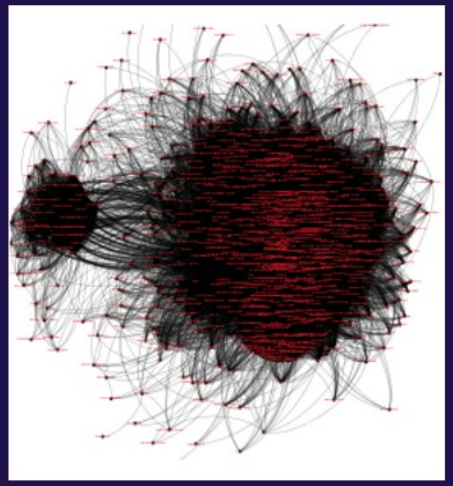


Cooking receipes

What is the specificity of french cooking, is there "families of taste"? For this study, we collected 28 000 receipes from the most popular cooking website for France.

Health preoccupations

By monitoring public discussions during 3 monthes on the main french medical forum, we enlighted the growing proportion of homeopathy recommandation for specific symptoma.



Political participations

How to qualify the debate on open access that took place during the online consultation on the ***law project for a digital republique****?*

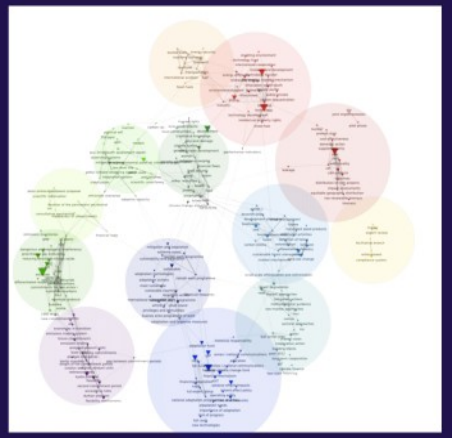
Open Data Sciences

Researcher needs tools to collect and analyse scientific production,

I develop tools to:

Build a dataset from multiple sources and multiple extraction methods:

- scientific publications
- press articles
- public mail archives

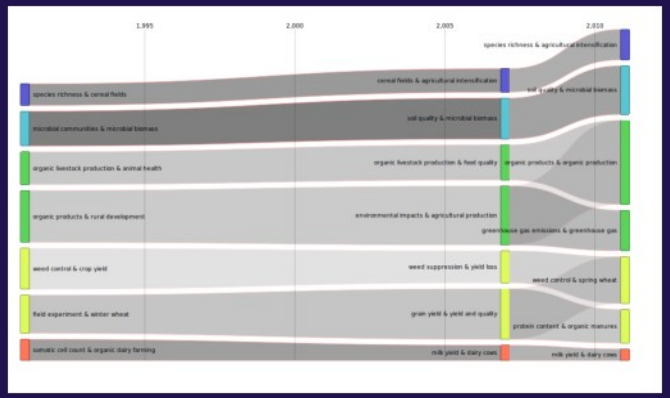


Organize it to access to multiple facets:

- authors references
- questions/answers
- publication date
- topic, title, tags, categories

Analyse the domain area:

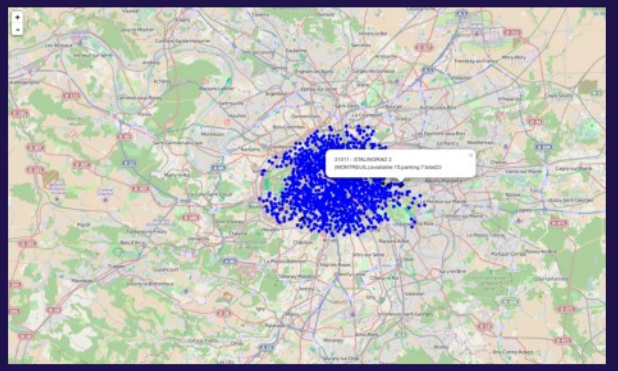
- topics into domain area
- clusters of authorship
- clusters of domain vs actors
- key field evolutions



Systèmes de Recommandation

Mapping building permits in Paris:

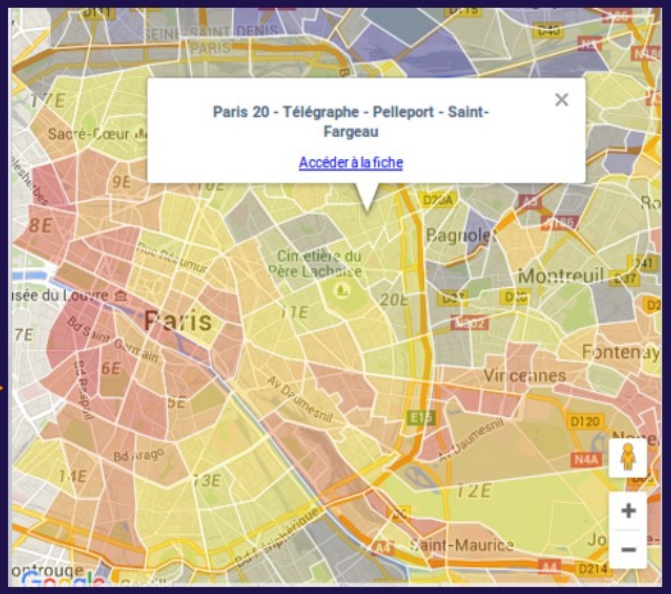
Showing the exact location of construction, modification or destructions of any building in Paris requiring permits from 2006 to 2016.



Semantic search engine API:

Detect the district and caracteristics of the goods to qualify the district and the offer caracteristics

Métro Pyrénées - Rue du Transvaal Au 1er étage d'un immeuble ancien, cet appartement d'une surface de 42m² comprend: une entrée, une séjour, une cuisine aménagée, une chambre, une salle de douche et WC séparé. Plein sud, parquet et moulures.



C'est à VOUS !


Petit tour de présentation


Vos questions


Vos projets


Atelier