








Gargantext :
Collaborative Web Platform
for Text-Mining
<http://gargantext.org>

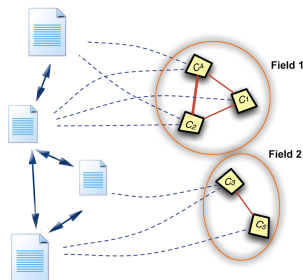
July 7, 2015

Our digital environment

TIME SCALES		
Seconds		> 180M Queries per hour
Minutes		> 400M tweet/d ; >540M SMS/d
Hours		>100M blogs, +120.000 new blogs per day
Days		~3200 AFP press release per day
Months		~1,6M academic papers/year in the WoS (79.000 for FR, 216/d)

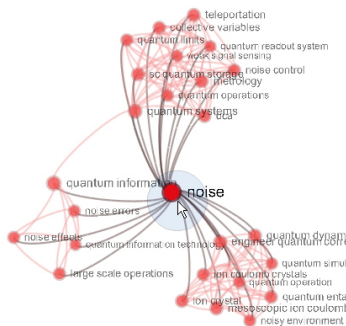
Interactive knowledge maps

- ▶ **Handle unstructured textual datasets** extract implicit relations through state-of-the-art text-mining and complex systems methods (no specific document structure is required, only some textual content).
- ▶ **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- ▶ **Portable** Run as server application with client on PC, Mac, Linux ...
- ▶ **Easy-to-use & user-friendly** novices can play immediately, become quickly confirmed users,
- ▶ **Compatibility with other softwares and database** several import and export formats, open software licence (GPL3).



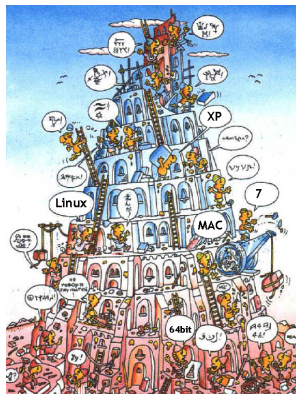
Interactive knowledge maps

- ▶ **Handle unstructured textual datasets** extract implicit relations through state-of-the-art text-mining and complex systems methods (no specific document structure is required, only some textual content).
- ▶ **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- ▶ **Portable** Run as server application with client on PC, Mac, Linux ...
- ▶ **Easy-to-use & user-friendly** novices can play immediately, become quickly confirmed users,
- ▶ **Compatibility with other softwares and database** several import and export formats, open software licence (GPL3).



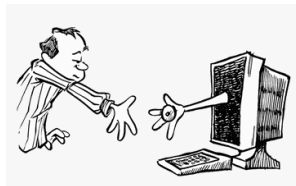
Interactive knowledge maps

- ▶ **Handle unstructured textual datasets** extract implicit relations through state-of-the-art text-mining and complex systems methods (no specific document structure is required, only some textual content).
- ▶ **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- ▶ **Portable** Run as server application with client on PC, Mac, Linux ...
- ▶ **Easy-to-use & user-friendly** novices can play immediately, become quickly confirmed users,
- ▶ **Compatibility with other softwares and database** several import and export formats, open software licence (GPL3).



Interactive knowledge maps

- ▶ **Handle unstructured textual datasets** extract implicit relations through state-of-the-art text-mining and complex systems methods (no specific document structure is required, only some textual content).
- ▶ **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- ▶ **Portable** Run as server application with client on PC, Mac, Linux ...
- ▶ **Easy-to-use & user-friendly** novices can play immediately, become quickly confirmed users,
- ▶ **Compatibility with other softwares and database** several import and export formats, open software licence (GPL3).



Interactive knowledge maps

- ▶ **Handle unstructured textual datasets** extract implicit relations through state-of-the art text-mining and complex systems methods (no specific document structure is required, only some textual content).
- ▶ **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- ▶ **Portable** Run as server application with client on PC, Mac, Linux ...
- ▶ **Easy-to-use & user-friendly** novices can play immediatly, become quickly confirmed users,
- ▶ **Compatibility with other softwares and database** several import and export formats, open software licence (GPL3).



Mapping philosophy

- ▶ Different views of the same complex network rather than THE map,
- ▶ Most valuable information about complex knowledge networks is obtained by collaborative map manipulations and queries,
- ▶ At any moment the user should be able to go from the aggregated level (the graphs) to the micro level (the documents/Ngrams selection) and back.

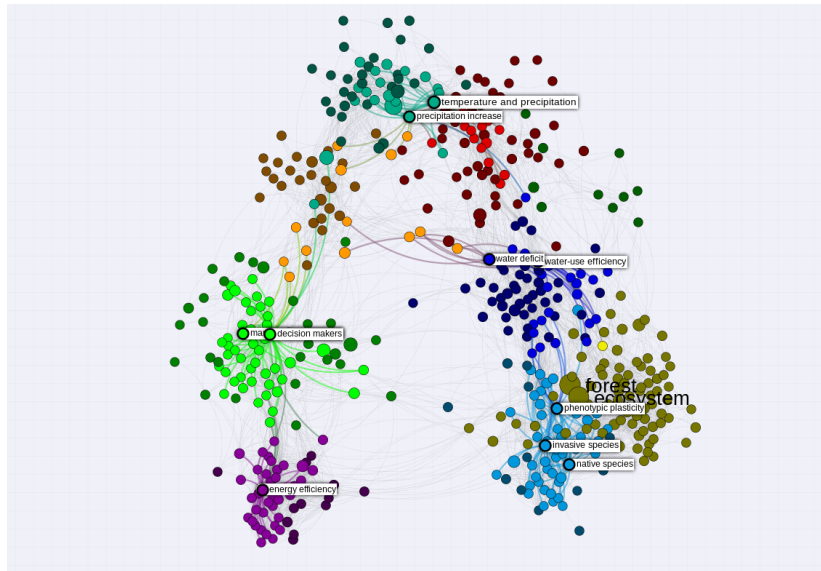
Mapping philosophy

- ▶ Different views of the same complex network rather than THE map,
- ▶ Most valuable information about complex knowledge networks is obtained by collaborative map manipulations and queries,
- ▶ At any moment the user should be able to go from the aggregated level (the graphs) to the micro level (the documents/Ngrams selection) and back.

Mapping philosophy

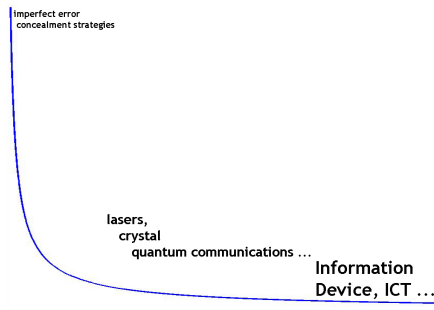
- ▶ Different views of the same complex network rather than THE map,
- ▶ Most valuable information about complex knowledge networks is obtained by collaborative map manipulations and queries,
- ▶ At any moment the user should be able to go from the aggregated level (the graphs) to the micro level (the documents/Ngrams selection) and back.

Climate Change



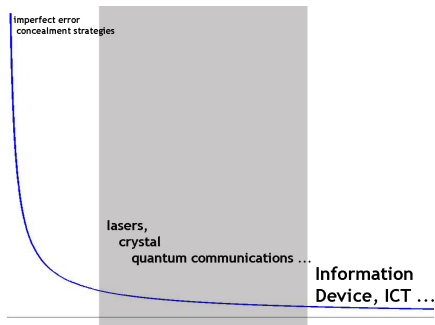
Some notion about terms selection

- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.



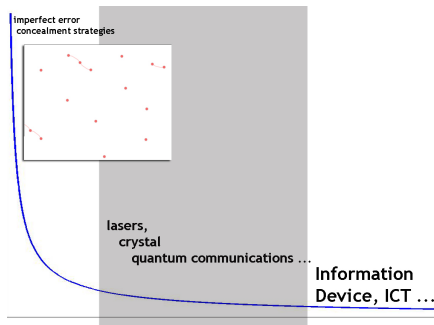
Some notion about terms selection

- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.



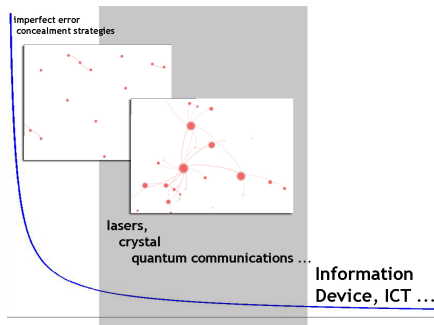
Some notion about terms selection

- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.



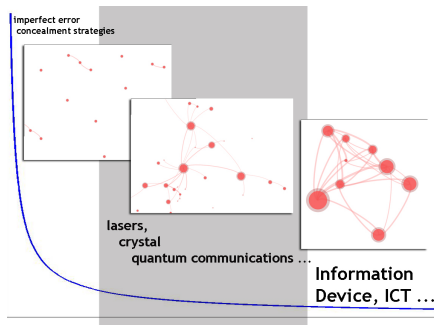
Some notion about terms selection

- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.



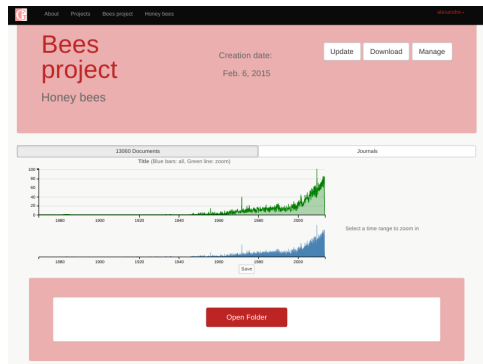
Some notion about terms selection

- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.



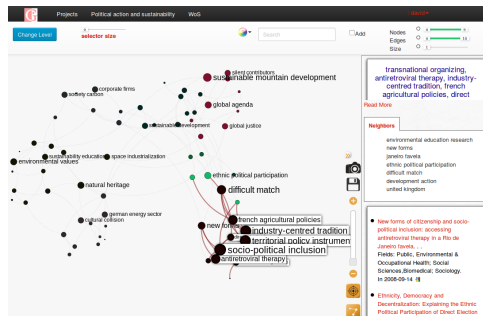
About Gargantext

- ▶ **Gargantext** is a software for the production, the exploration and annotation of projects maps. It includes text-mining and natural language processing technologies, reconstruction methods of thematic landscape and visualisation tools.

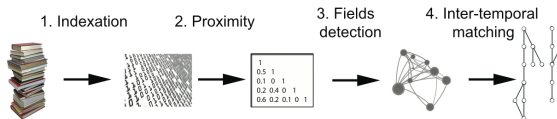


About Gargantext

- **Tina explorer** is a stand alone interface for the exploration of projects maps included in Gargantext. This explorer can also be embedded online to present Gargantext outputs.



Mapping projects with Gargantext



Gargantext is Database agnostic

Custom parser for each database:

1. Pubmed (XML format and direct queries)
2. Web Of Science (ISI Format)
3. Scopus (RIS Format)
4. Europress (HTML Format)
5. Zotero (RIS Format)
6. Jstor (RIS Format)
7. CSV files
8. ISTEK (CNRS, Elsevier)

Gargantext is Language agnostic

- ▶ English
- ▶ Français
- ▶ Spanish
- ▶ Italian
- ▶ Deutsch
- ▶ Dutch
- ▶ Português
- ▶ Polish
- ▶ Persian

Main views

1. Global view of the corpus to clean/filter it:
 - 1.1 Document by Document (or range of documents)
 - 1.2 By Journals
 - 1.3 By terms (ngrams count or TFIDF)
2. Document view to read and annotate.
 - 2.1 Chose your Miam—Stop words
 - 2.2 Manage your lists
3. Advanced charts to explore the corpus
 - 3.1 Compare corpora easily
 - 3.2 Comparison with many fields (Journal, ngrams...)
 - 3.3 Custom your charts
4. Explorer to interact with the graphs
 - 4.1 Zoom in / Zoom out
 - 4.2 Focus on clusters

Text-Mining

Miamwords

Stopwords

* untradable adverse impacts¹

* untradable categories¹

* newer aspects¹

* japanese agriculture¹

* extensive use⁴

* human health²⁴

* satisfactory method¹

* safer use¹

* adverse impacts²

* appropriate procedures¹

* problem-oriented case¹

* technology assessment¹

* modern technologies¹

* effect sequences¹

* wider spectrum¹

A

A

Add

Technology assessment on the use of pesticides.

H Ishikura / Environmental quality and safety

/ 26/04/1

Abstract

The **extensive use** of **pesticides** in **Japanese agriculture** was studied as a **problem-oriented case** of **technology assessment** an **aim** to contribute to the **establishment** of **appropriate procedures** of assessing real and **potential impacts** which **modern technologies** have or may have on **health, industries, economy, society** and on the **environment** and to obtain **clues** to the **development** of **safer use** of **pesticides**. Direct and **indirect impacts**, both real and potential, favourable and adverse, were intended to identify and evaluate systematically and comprehensively as their **cause** and **effect sequences** were studied. **Adverse impacts** were divided into tradable and **untradable categories**, **untradable adverse impacts** were related mostly to **human health**. As to the **evaluation** of **impacts**, it was suggested that the **size** of **area** and **population** affected, **irreversibility** **controllability** of the **impact** be considered. It was recognized as urgent and requisite to develop and establish a more efficient and **satisfactory method** of testing the **safety** of **pesticides** and their **metabolites** over a **wider spectrum** of **organisms** and with respect to **newer aspects** of **toxicology** as **mutagenesis, teratogenesis** and **cancerogenesis**.

No full text

Parsers and taggers implemented in Gargantext:

- ▶ TreeTagger (Perl)
- ▶ NLTK (Python)
- ▶ Turbo Parser (C++)
- ▶ Melt (Python and Perl)

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ Keep simple and modular
 - ▶ **Minimal** - **Drop that** - **words manager** (available)
 - ▶ **Minimize** **equally** **resources**
- ▶ **Collaborative**
 - ▶ **Platform** **collaborative** **workspace**
 - ▶ **Share** **resources** **collaborative** **workspace**
 - ▶ **Protect** **resources** **collaborative** **workspace**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Inter-objects and multi-view**
 - ▶ **Matrix - Data link - words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform collaborative workspace**
 - ▶ **Share resources collaborative workspace**
 - ▶ **Project resources collaborative workspace**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Learn, share and reuse** :
 - ▶ **Matrix** : Data lake - matrix manager (available)
 - ▶ **Network** : easy resources
- ▶ **Collaborative**
 - ▶ **Platform collaborative workspace**
 - ▶ **Matrix collaborative workspace**
 - ▶ **Network collaborative workspace**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ Make the data available (data available)
 - ▶ Manage easily resources
- ▶ **Collaborative**
 - ▶ Platform collaborative workspace
 - ▶ Allow researchers collaborating on data
 - ▶ Shared resources and analysis tools

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ Keep simple and reactive
 - ▶ Miam—Stop list—words manager (available)
 - ▶ Manage easily resources
- ▶ **Collaborative**
 - ▶ Federated collaborative workplace
 - ▶ Cloud collaborative workplace
 - ▶ Virtual collaborative workplace

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Phylogenies and landscapes**
 - ▶ **Phylogenies and landscapes**
 - ▶ **Phylogenies and landscapes**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Share easily resources**
 - ▶ **Share easily knowledge**
 - ▶ **Share easily data**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform: collaborative workplace**
 - ▶ **Platform: collaborative research**
 - ▶ **Platform: collaborative business**

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform**: collaborative workplace
 - ▶ **Share ressources**: collaborative workplace
 - ▶ **Protect ressources**: collaborative workplace

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform**: collaborative workplace
 - ▶ **Share ressources**: collaborative workplace
 - ▶ **Protect ressources**: collaborative workplace

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform**: collaborative workplace
 - ▶ **Share ressources**: collaborative workplace
 - ▶ **Protect ressources**: collaborative workplace

Roadmap

- ▶ **Fiability**
 - ▶ **Matrix view** : Adjacency matrix management
 - ▶ **Network view** : Multipartite heterogeneous
 - ▶ **Dynamics and temporal view**: Phylogenies landscapes
- ▶ **Agility**
 - ▶ **Keep simple and reactive**
 - ▶ **Miam—Stop list—words manager (available)**
 - ▶ **Manage easily resources**
- ▶ **Collaborative**
 - ▶ **Platform**: collaborative workplace
 - ▶ **Share ressources**: collaborative workplace
 - ▶ **Protect ressources**: collaborative workplace

Sponsors



Sponsors



Synergies with CorTexT, the digital platform of IFRIS

Core Team



Alexandre Delanoë

principal investigator, developer

[Mail](#) [Website](#)



David Chavalarias

principal investigator

[Mail](#) [Website](#)



Samuel Castillo J.

developer

[Mail](#) [Website](#)



Mathieu Rodic

developer

[Website](#)



Elias Showk

developer

[Website](#)

Community

- ▶ Mailing-lists:
 - ▶ User mailing-list: (soon)
 - ▶ Devel mailing-list: (soon)
- ▶ Interactive forum:
 - ▶ IRC: #gargantext channel on OFTC
- ▶ Code access:
 - ▶ Licence: AGPL,
 - ▶ GIT access (with the first publication on Gargantext, follow us @ISC-PIF !)