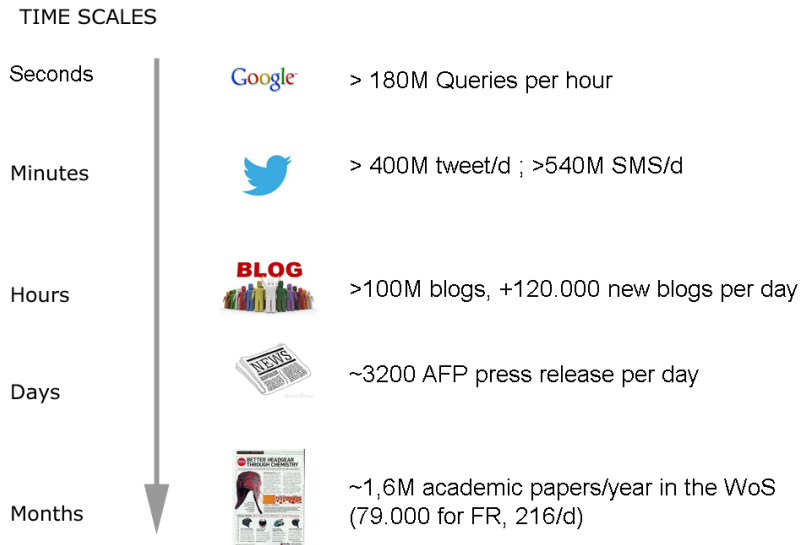




Gargantext :  
Collaborative Web Platform  
for Text-Mining  
<http://gargantext.org>

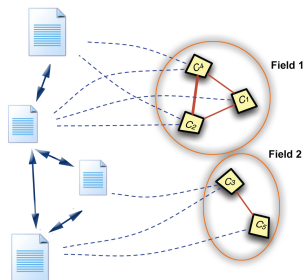
November 7, 2016

# Our digital environment



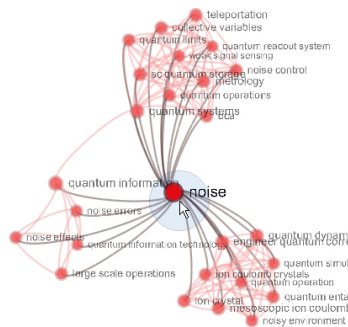
# Interactive knowledge maps

- ▶ **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).



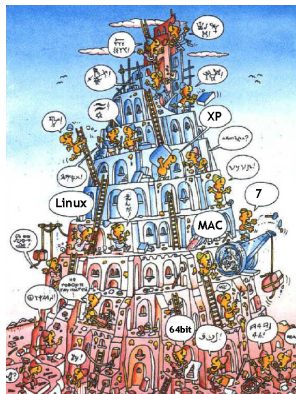
# Interactive knowledge maps

- ▶ **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).



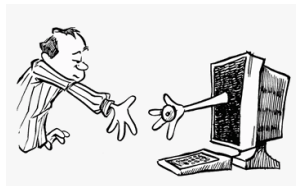
# Interactive knowledge maps

- ▶ **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

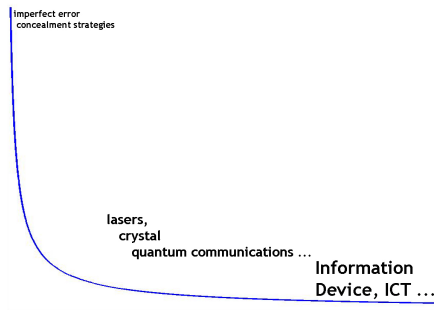
- ▶ Different views of the same complex network rather than THE map,
- ▶ Most valuable information about complex knowledge networks is obtained by collaborative map manipulations and queries,
- ▶ At any moment the user should be able to go from the aggregated level (the graphs) to the micro level (the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

# CERN publications in Web of Science (2013-2015) 3 000 documents (work in progress)



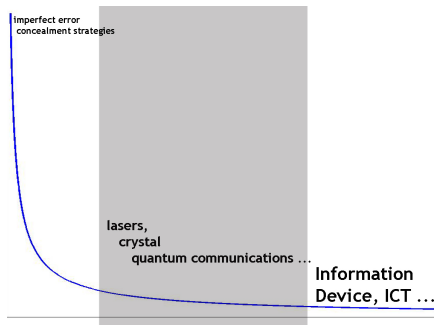
# Some notion about terms selection

- ▶ 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.



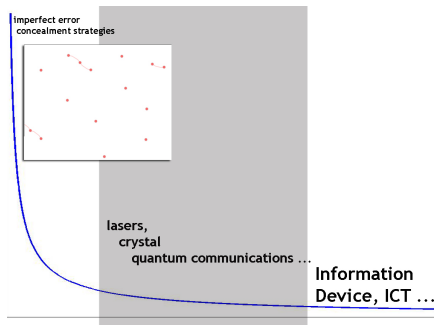
# Some notion about terms selection

- ▶ 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.



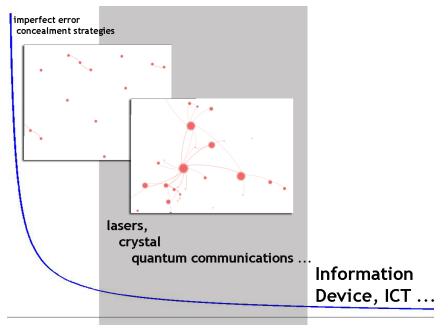
# Some notion about terms selection

- ▶ 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.



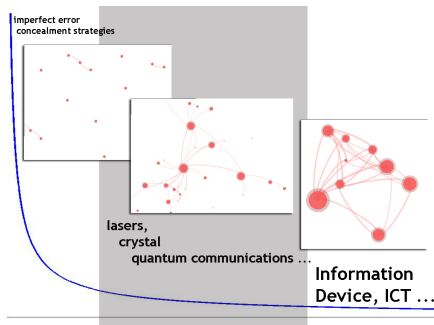
# Some notion about terms selection

- ▶ 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

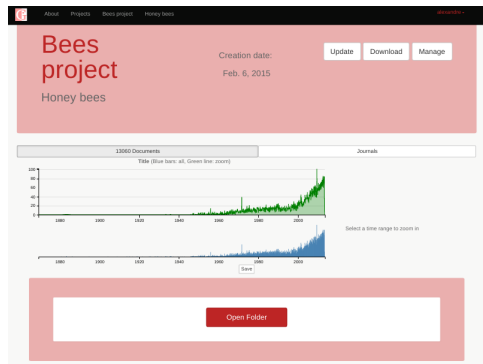
- ▶ Powerlaw distribution is frequent
- ▶ Most interesting terms for maps are those with medium range occurrences.
- ▶ Each category of terms will bring characteristic patterns to the maps.





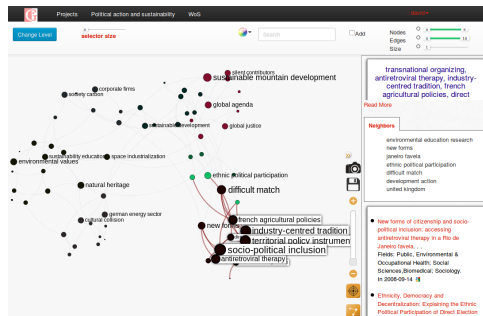
# About Gargantext

- ▶ **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.

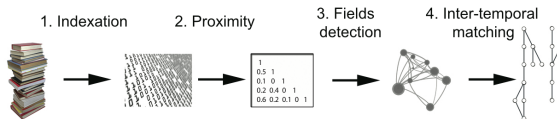


# About Gargantext

- **Graph 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. ISTEEX (CNRS, Elsevier)
9. CERN API Soon!

# 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 terms
  - 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 impacts <sup>1</sup>

\* untradable categories <sup>1</sup>

\* newer aspects <sup>1</sup>

\* japanese agriculture <sup>1</sup>

\* extensive use <sup>4</sup>

\* human health <sup>24</sup>

\* satisfactory method <sup>1</sup>

\* safer use <sup>1</sup>

\* adverse impacts <sup>2</sup>

\* appropriate procedures <sup>1</sup>

\* problem-oriented case <sup>1</sup>

\* technology assessment <sup>1</sup>

\* modern technologies <sup>1</sup>

\* effect sequences <sup>1</sup>

\* wider spectrum <sup>1</sup>

Á

À

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)

# Partners and supports



FORCAST





# Partners and supports



**FORCAST**



*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 !)