



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

May 18, 2015

Sponsors



Core Team

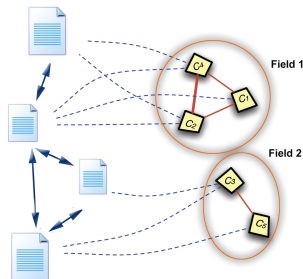
images/coreteam.pdf

Community

- Mailing-lists:
 - User mailing-list: (soon)
 - Devel mailing-list: (soon)
- Interactive forum:
 - IRC: #gargantext channel on OFTC
- Code access:
 - Licence: AGPL
 - GIT access (soon)

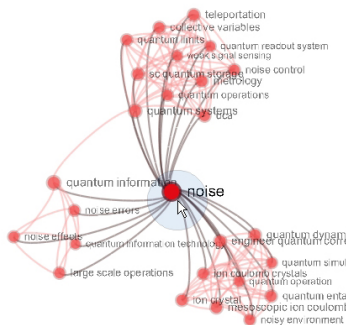
Interactive mapping of project portfolio

- **Handle unstructured textual datasets** extract implicit relations (co-word analysis) through text-mining methods (no specific relation required, only some textual content).
- **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- **Easy-to-use & user-friendly** novices can play immediatly, become quickly confirmed users.



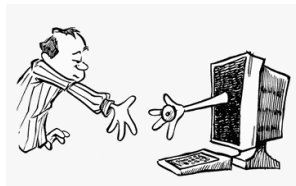
Interactive mapping of project portfolio

- **Handle unstructured textual datasets** extract implicit relations (co-word analysis) through text-mining methods (no specific relation required, only some textual content).
- **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- **Easy-to-use & user-friendly** novices can play immediatly, become quickly confirmed users.



Interactive mapping of project portfolio

- **Handle unstructured textual datasets** extract implicit relations (co-word analysis) through text-mining methods (no specific relation required, only some textual content).
- **Interactivity** Tools for real-time interaction with corpora. **Make your own idea** of what is there. Cumulative outcomes.
- **Easy-to-use & user-friendly** novices can play immediatly, become quickly confirmed users.



Mapping philosophy

- Different views of the same complex network rather than THE map,
- Most valuable information about complex networks is obtained by 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 networks is obtained by 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 networks is obtained by 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.

Gargantext is Database agnostic

Custom parser for each database:

- Pubmed (XML format and direct queries)
- Web Of Science (ISI Format)
- Scopus (RIS Format)
- Europress (HTML Format)
- Zotero (RIS Format)
- Jstor (RIS Format)

Main views

1. Global view of the corpus to clean the corpus
2. Advanced charts to explore the corpus
3. Explorer to interact with the graphs
4. Document view to read and annotate.

Maps building blocks

1. The corpora (csv format, etc.) *e.g.* the project portfolio
2. The white list and stop words that defines the thematic perimeter (cumulative expertise with the master white list),
3. Proximity measures between terms and documents

“corp_id”, “doc_id”, “title”, “abstract”

“FET OPEN”, “59”, “Gargantext”, “Gargantext is a support action aiming to develop tools for the analysis and visualisation of the relationships between scientific research projects and proposals.”

Maps building blocks

1. The corpora (csv format, etc.) *e.g.* the project portfolio
2. The white list and stop words that defines the thematic perimeter (cumulative expertise with the master white list),
3. Proximity measures between terms and documents

status

label

forms

w	<u>bio-sensors</u>	<u>bio-sensors</u> *** <u>bio-sensing</u>
w	<u>biocompatible nanophotonic dev</u>	<u>biocompatible nanophotonic dev</u>
		<u>biocomputational model</u> *** <u>bio-computational model</u> ***
w	<u>biocomputational model</u>	<u>biocomputational models</u> *** <u>bio-computational models</u>
w	<u>biodosimeter</u>	<u>biodosimeter</u>
w	<u>bioelectric signals</u>	<u>bioelectric signals</u>
w	<u>biohazard identification</u>	<u>biohazard identification</u>
w	<u>bio-hybrid</u>	<u>biohybrid</u> *** <u>bio-hybrid</u> *** <u>bio hybrid</u>
w	<u>biohybrid systems</u>	<u>biohybrid systems</u>
w	<u>bioimpedance</u>	<u>bioimpedance</u>
w	<u>bioinformatics</u>	<u>bioinformatics</u> *** <u>bio-informatics</u>
w	<u>bioinspired information processing</u>	<u>bioinspired information processing</u>

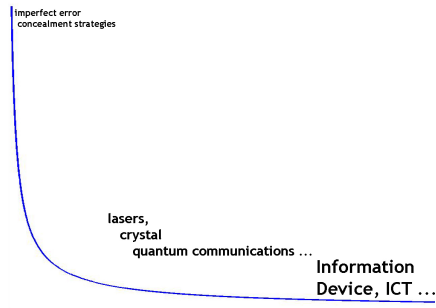
Maps building blocks

1. The corpora (csv format, etc.) *e.g.* the project portfolio
2. The white list and stop words that defines the thematic perimeter (cumulative expertise with the master white list),
3. Proximity measures between terms and documents

- $P_1 = \frac{n_{ij}}{\min(n_i, n_j)}$
- $P_2 = \frac{n_{ij}^2}{n_i \cdot n_j}$
- $P_3 = ((\frac{n_{ij}^T}{n_i^T})^\alpha (\frac{n_{ij}^T}{n_j^T})^{1/\alpha})^{\min(\alpha, \frac{1}{\alpha})}$.

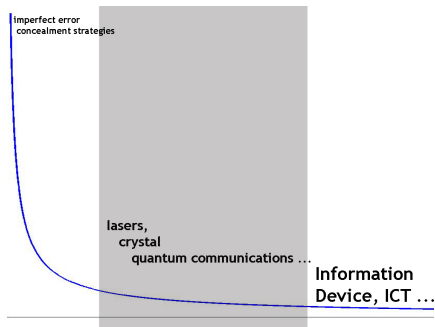
Some notion about terms selection

- 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.



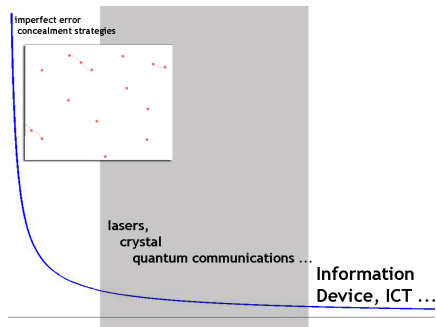
Some notion about terms selection

- 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.



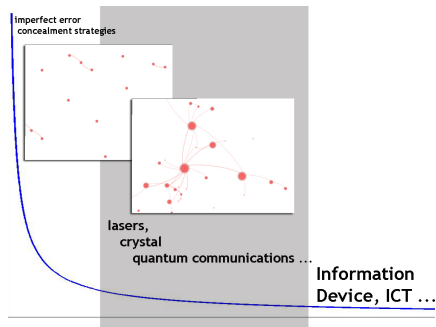
Some notion about terms selection

- 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.



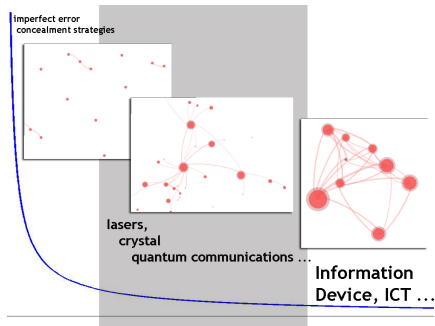
Some notion about terms selection

- 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.



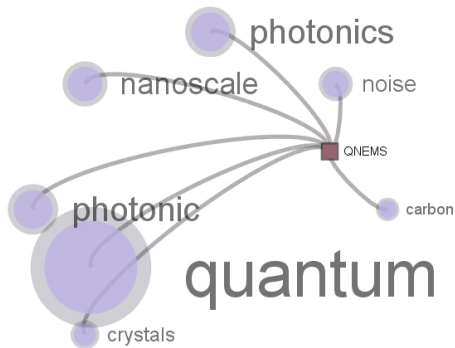
Some basic notions about bipartite graphs

- **Project neighborhood**

A project can have several keywords

- **Term neighborhood**

A term can be mentioned by several projects.



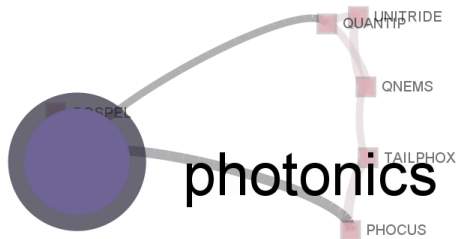
Some basic notions about bipartite graphs

- **Project neighborhood**

A project can have several keywords

- **Term neighborhood**

A term can be mentioned by several projects.



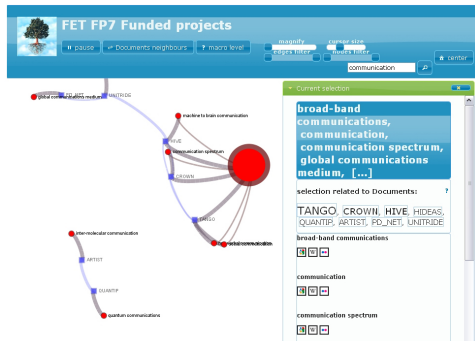
The Gargantext software suite

- **Tinasoft desktop** 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.



The Gargantext software suite

- **Explorer** is a stand alone interface for the exploration of projects maps created by Tinasoft. It can be ran on a personal computer or online.



Mapping projects with Gargantext



1. Indexation



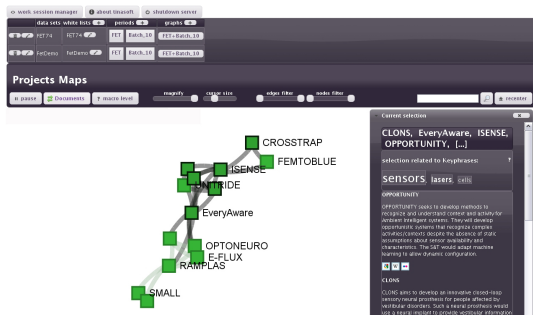
2. Proximity



3. Fields detection



4. Inter-temporal matching



Some basic operators

- **Search** for a string
- Select one or several nodes,
- Switch to the other bipartite component,
- Change level from global view to neighborhood view,
- Adjust filter Unfold the graph removing edges and nodes



Some basic operators

- **Search** for a string
- **Select** one or several nodes,
- Switch to the other bipartite component,
- Change level from global view to neighborhood view,
- Adjust filter Unfold the graph removing edges and nodes



Some basic operators

- **Search** for a string
- **Select** one or several nodes,
- **Switch** to the other bipartite component,
- **Change level** from global view to neighborhood view,
- **Adjust filter** Unfold the graph removing edges and nodes



Some basic operators

- **Search** for a string
- **Select** one or several nodes,
- **Switch** to the other bipartite component,
- **Change level** from global view to neighborhood view,
- **Adjust filter** Unfold the graph removing edges and nodes



Some basic operators

- **Search** for a string
- **Select** one or several nodes,
- **Switch** to the other bipartite component,
- **Change level** from global view to neighborhood view,
- **Adjust filter** Unfold the graph removing edges and nodes



Queries examples

- Find the core/periphery of the project network or the semantic network



- Map the semantic background of a part of the portfolio



- Target a specific domain



- Find the semantic and project background of an expression



- Find a path between two projects or terms



Queries examples

- Find the core/periphery of the project network or the semantic network



- Map the semantic background of a part of the portfolio



- Target a specific domain



- Find the semantic and project background of an expression



- Find a path between two projects or terms



Queries examples

- Find the core/periphery of the project network or the semantic network



- Map the semantic background of a part of the portfolio



- Target a specific domain



- Find the semantic and project background of an expression



- Find a path between two projects or terms



Queries examples

- Find the core/periphery of the project network or the semantic network



- Map the semantic background of a part of the portfolio



- Target a specific domain



- Find the semantic and project background of an expression



- Find a path between two projects or terms



Queries examples

- Find the core/periphery of the project network or the semantic network



- Map the semantic background of a part of the portfolio



- Target a specific domain



- Find the semantic and project background of an expression



- Find a path between two projects or terms



Perspectives

- **Miam—Stop list—words manager**
- Multipartite heterogeneous network : still simple and reactive
- Dynamics: Phylogenies landscapes

Perspectives

- **Miam—Stop list—words manager**
- **Multipartite heterogeneous network** : still simple and reactive
- **Dynamics: Phylogenies landscapes**

Perspectives

- **Miam—Stop list—words manager**
- **Multipartite heterogeneous network** : still simple and reactive
- **Dynamics**: Phylogenies landscapes