Brainomics - A management system for exploring and merging heterogeneous brain mapping data

V. Michel\textsuperscript{1}, Y. Schwartz\textsuperscript{2,4}, P. Pinel\textsuperscript{3}, O. Cayrol\textsuperscript{1}, A. Moreno\textsuperscript{3}, JB. Poline\textsuperscript{4}, V. Frouin\textsuperscript{4}, D. Papadopoulos Orfanos\textsuperscript{4}

\textsuperscript{1}Logilab, Paris, France, \textsuperscript{2}INRIA Saclay Parietal Team, Paris, France, \textsuperscript{3}Unicog, INSERM-CEA, Neurospin, Gif-sur-Yvette, France, \textsuperscript{4}CEA, Neurospin, Gif-sur-Yvette, France

Introduction

- Number of large datasets for brain mapping have been released \cite{1,2}.
- Neuroimaging datasets more routinely include clinical data or genetics data.
- Exploitation requires
  - An efficient organization to integrate all the measures
  - An easy access to the relevant information.
- Neuroimaging \cite{3} and genomics \cite{4} databases are dedicated to their own field of research.

Data model

- Described in Python, using reusable modules called “cubes”.
- Modelisation of Scans, Questionnaires, Genomics results, Behavioral results, Subjects and Studies information.
- Data model optimized for large volumes (> 2000 subjects).
- Tested with several publicly available datasets \cite{1,2}.

Conclusion

- Open source solution to manage brain imaging datasets and associated meta data.
- Powerful querying and reporting tool, customized for emerging imaging-genetics field.

This work was supported by grants from the French National Research Agency (ANR GENIM; ANR-10-BLAN-0128) and (ANR IA BRAINOMICS; ANR-10-BINF-04).

Contact: brainomics@logilab.fr

Introduction

- Number of large datasets for brain mapping have been released \cite{1,2}.
- Neuroimaging datasets more routinely include clinical data or genetics data.
- Exploitation requires
  - An efficient organization to integrate all the measures
  - An easy access to the relevant information.
- Neuroimaging \cite{3} and genomics \cite{4} databases are dedicated to their own field of research.

Data model

- Described in Python, using reusable modules called “cubes”.
- Modelisation of Scans, Questionnaires, Genomics results, Behavioral results, Subjects and Studies information.
- Data model optimized for large volumes (> 2000 subjects).
- Tested with several publicly available datasets \cite{1,2}.

Conclusion

- Open source solution to manage brain imaging datasets and associated meta data.
- Powerful querying and reporting tool, customized for emerging imaging-genetics field.

This work was supported by grants from the French National Research Agency (ANR GENIM; ANR-10-BLAN-0128) and (ANR IA BRAINOMICS; ANR-10-BINF-04).

Contact: brainomics@logilab.fr

http://www.brainomics.net/demo/

- Brings together brain imaging and genetics data.
- Relies on a high-level query language (RQL).
- Solution based on CubicWeb, a semantic framework.
- Visualizing / exporting data in several formats.

CubicWeb

- Data management framework, 10 years of industrial uses (e.g. \cite{5}).
- Well established core technologies: SQL, Python, HTML5, Javascript.
- Licensed under LGPL since 2008.
- Used in production environments since 2005.
- Fine-grained security system coupled to the data model definition.
- Migration mechanisms controls model version / ensures data integrity.

Query using RQL

- Similar to the W3C’s SPARQL \cite{6}.
- Supports the basic operations (selection, insertion, etc.).
- Subquerying, ordering, counting, ...

Query all the female subjects of the database, with an age greater than 30:

\begin{verbatim}
Any S WHERE S is Subject, S age > 30, S gender "female"
\end{verbatim}

Query all the Cmap scans of subjects with an age greater than 25, and that have a score greater than 4.0 for the “algebre” question:

\begin{verbatim}
Any SA WHERE S is Subject, S age > 25, X is QuestionnaireRun, X concerns S, A is Answer, A questionnaire_run X, A question Q, Q text "algebre", A value > 4, SA is Scan, SA concerns S, SA type "c map"
\end{verbatim}

Views

- Each query result can be seen using different views.
- HTML pages, ZIP files, spreadsheets, XCEDE XML, ...
- May include processing (stat. maps computed on the fly).