

Global Scientific Positioning System *for* CDS

Philippe Caillou & Jean-Daniel Fekete
TAO-AVIZ, CNRS – INRIA – LRI, Univ. Paris-Sud

Goal

Automatically build a description of a SON (Scientific Organism or Network) expertise, including 3 functionalities: i) a zoomable map, graphically displaying the strengths/weaknesses of the SON w.r.t. the worldwide state of the art; ii) a historical perspective, showing the evolution of the SON focus along the years; iii) a search operator: given a query (bag of words), the SON members/articles most relevant to the query are displayed and the query is interactively answered.

Proposal

The GSPS proposal builds upon the Cartolabe project, a 2-year engineering project achieving the visual representation of the scientific production of a SON, funded by INRIA and exploiting the abstract or full-text material in the HAL database¹. Two interdependent challenges remain to be addressed:

Firstly, several indicators measuring the quality of a scientific landscape need be defined, to support the automatic calibration (hyper-parameter setting) of the GSPS. Quality criteria could include for instance the dispersion of team members, the dispersion of keyword-based clusters, or the alignment w.r.t. the authorship graph-based distance. Discrepancies between the author-graph and the full text-based distances will be analyzed.

Secondly, the GSPS cartography must be tailored to different user profiles (e.g., scientific manager, industrial manager looking for scientific partners, jury members looking for competent reviewers with no conflict of interest). Interactive optimization will be deployed to allow the expert user to shape the map according to her prior knowledge; among the related scientific challenges is the representation of the user's preferences and the trade-off between the interface flexibility and its comfort (visual persistence).

GSPS will be complementary to its main competitors in terms of scale and high-level interaction: DBLP-Arizona² offers a visualization reflecting the research of a single researcher, or the contents of a single conference, based on the titles of papers from DBLP; Gargantext³ relies on the extensive interaction with the user (providing the corpus, calibrating the distance, the word filtering, the thresholds); Academic landscape⁴ likewise relies on the extensive expert's support for the intelligibility of the landscape.

Expected results

The GSPS system will provide a sophisticated and operational representation of the CDS activity, supporting its visibility in particular w.r.t. international and industrial partners. It will also act as a demonstrator of the CDS know-how within the Université Paris-Saclay, to which further fundings will be asked to ensure a better coverage of the scientific literature using WoS or Scopus data.

Financial proposal

The GSPS proposal asks for one year post-doc (50k). The GSPS software will be released as open source software.

¹cartolabe.lri.fr

²<http://mocs.cs.arizona.edu/>

³<http://gargantext.org/>, ISCPIF, David Chavalarias

⁴Pr. Sakata, Policy Alternative Research Institute, Tokyo University; collaboration Elsevier