



HAL
open science

Towards Heterogeneous Data Integration: The NeuroLang Approach

Gastón E Zanitti, Majd Abdallah, Demian Wassermann

► **To cite this version:**

Gastón E Zanitti, Majd Abdallah, Demian Wassermann. Towards Heterogeneous Data Integration: The NeuroLang Approach. OHBM 2022 - Organization for Human Brain Mapping, Jun 2022, Glasgow, United Kingdom. hal-03786717

HAL Id: hal-03786717

<https://inria.hal.science/hal-03786717>

Submitted on 23 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards Heterogeneous Data Integration: The NeuroLang Approach

Gastón E. Zanitti, Majd Abdallah, Demian Wassermann

September 23, 2022

1 Introduction

A major goal of neuroscience research is to understand the circuits and patterns of neural activity that give rise to mental processes and behavior. But a task of this complexity often requires the integration of datasets, which are usually presented under a variety of topologies –ontologies, images, text–. This makes those analyses prone to human error and time-consuming.

We believe that a tool with the ability to integrate heterogeneous datasets in a comprehensive and unified manner would alleviate the workload required for such a task. To this end, we will demonstrate how we can combine different datasets to perform an analysis that allows us to use histological knowledge of the brain and relate it to a variety of cognitive processes.

Our main proposal is that having a probabilistic language based on first-order logic able to combine data under different topologies, solve complex queries and deal with uncertainty in a single framework is a great fit for this type of analysis.

2 Methods

2.1 Data integration under the open world assumption (1352 characters with eq)

Now, we will see how we can make use of NeuroLang to solve queries that require to comply with the open-world assumption. This assumption states that if a fact is not implied from the knowledge explicitly recorded in the system, then the fact may be true or false [1]. In particular, we aim at relating the *visual awareness* cognitive topic from the Cognitive Atlas ontology (CogAt) [8] with brain areas reported activated during this process. This can't be done directly through Neurosynth [10], as the cognitive topic is not reported, but there is a tuple-generating dependency (TGD) in CogAt specifying that *spatial attention* is a sub-process of *visual awareness*. Which, expressed as a Datalog+/- rule in

CogAt’s TGD set, is:

$$\begin{aligned} \forall X \text{SpatialAttention}(X) \rightarrow \\ \exists Y \text{PartOf}(X, Y) \wedge \text{VisualAwareness}(Y) \in \Sigma_1^{\text{CogAt}}, \end{aligned} \quad (1)$$

which states that, for every element X belonging to the `SpatialAttention` class, there is at least one element Y , which is part of the `VisualAwareness` class and is also part of the former. It also has an existentially-quantified variable in the head (Y), hence representing open knowledge. By solving this query with NeuroLang, we can then make use of the open knowledge provided by the ontology.

2.2 From histology to cognitive processes, a case study

We are going to show how, combining heterogeneous datasets, we can analyze cognitive topics defined through meta-analysis and their relationship to brain areas defined using a histological atlas. We combined terms and studies from the NeuroSynth [10] database, cognitive topics from CogAt’s [8] ontology, and regions of the Julich-Brain Atlas [2] to performed a series of queries in which we calculated the probability of a term being present in an area, given that the area is activated and given that the area is not activated.

3 Results

3.1 Data integration under the open world assumption

We seamlessly combine an automatically-produced literature database like NeuroSynth with an expert-produced ontology like CogAt, to analyze activations related to *visual awareness* using NeuroLang’s built-in capabilities: we write a program (see listing 1) to obtain all studies that, while not mentioning *visual awareness*, mention terms which, according to CogAt, imply that the cognitive process is involved. The resulting activations can be seen in fig 1.

Listing 1: Open world assumption

```
Entity(t, s) :- TermInStudy(t, s)

OpenWorldStudies(s) :- PartOf(t, s), VisualAwareness(s).

ProbMap(x, y, z, PROB) :- FocusReported(x, y, z, s) //
    (SelectedStudy(s) & OpenWorldStudies(s))

ProbabilityImage(create_region_overlay(x, y, z, p)) :-
    e.ProbMap(x, y, z, p)
```

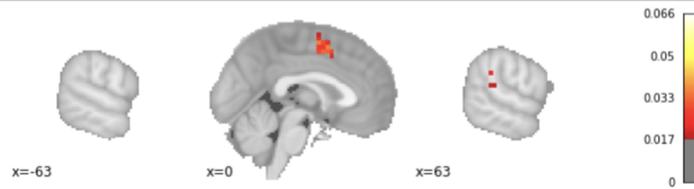


Figure 1: Resulting thresholded brain image from the NeuroLang use case showing the activations related to spatial attention obtained through the resolution of a query under the open world assumption.

Figure 1: Resulting thresholded brain image from the NeuroLang use case showing the activations related to spatial attention obtained through the resolution of a query under the open world assumption.

3.2 From histology to cognitive processes, a case study

In this example, we can see how the ability to combine heterogeneous databases allows us to perform an analysis where we combine cognitive topics with cytoarchitecture based atlases. In particular, we can see how the ACC area of Julich's atlas (fig 2) is mostly related to concepts associated to Emotion, Motivation, Perception, and Decision Making (figs. 3 & 4). At the same time, we can disaggregate these concepts individually and analyze, as in this case, the NeuroSynth terms associated with Emotion (fig 5 & 6).

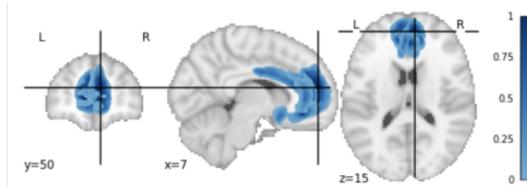


Figure 2: ACC region and its associated probabilities, obtained from Julich's atlas

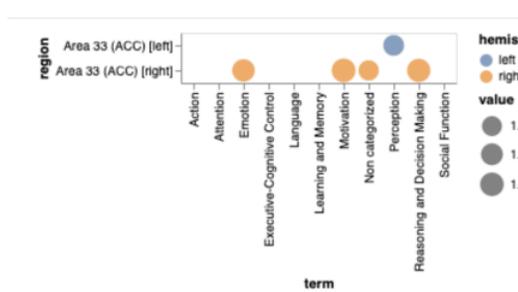


Figure 3: Bayes factor thresholded at 1, under the hypothesis that the top CogAt concepts are related to the ACC region. Values per concept and hemisphere.

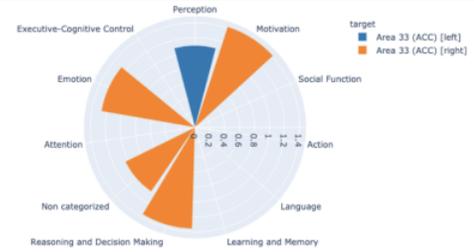


Figure 4: Bayes factor thresholded at 1, under the hypothesis that the top CogAt concepts are related to the ACC region. Values per concept.

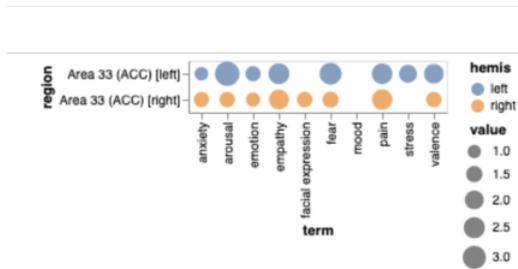


Figure 5: Bayes factor for NeuroSynth terms defined under the top concept emotion in the CogAt ontology. Values per concept and hemisphere. Threshold Threshold equal to 1

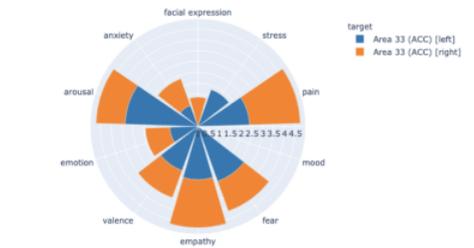


Figure 6: Bayes factor for NeuroSynth terms defined under the top concept emotion in the CogAt ontology. Values per concept. Threshold Threshold equal to 1

4 Conclusions

In this work we show how we can go beyond current meta-analyses tools whose queries are based on propositional logic [10, 7] and harness the full power of the first-order logic fragment extended with existentials and negation ($FO^{\exists, \neg}$), as well as open-world semantics, to express meta-analysis tasks in a sound and declarative manner. By using probabilistic semantics, with open-world semantics [3, 6, 4], and query resolution approaches [5, 4, 9], we have produced a language that is ready to be used in neuroimaging applications.

References

- [1] S. Abiteboul, Richard Hull, and Victor Vianu. *Foundations of databases*. Addison-Wesley, Reading, Mass, 1995.
- [2] K. Amunts, H. Mohlberg, S. Bludau, S. Caspers, S. B. Eickhoff, and P. Pieperhoff. Whole-brain parcellation of the Julich-Brain Cytoarchitectonic Atlas (v2.0), June 2020. type: dataset.
- [3] Andrea Cali, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Semant.*, 14:57–83, 2012.
- [4] İsmail İlkan Ceylan, Adnan Darwiche, and Guy Van den Broeck. Open-world probabilistic databases: Semantics, algorithms, complexity. *Artificial Intelligence*, 295:103474, June 2021.
- [5] Nilesch Dalvi and Dan Suciu. The dichotomy of probabilistic inference for unions of conjunctive queries. *Journal of the ACM*, 59(6):1–87, December 2012.
- [6] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Query Rewriting and Optimization for Ontological Databases. *ACM Transactions on Database Systems*, 39(3):1–46, October 2014.
- [7] Angela R Laird, Simon B Eickhoff, P Mickle Fox, Angela M Uecker, Kimberly L Ray, Juan J Saenz, D Reese McKay, Danilo Bzdok, Robert W Laird, Jennifer L Robinson, Jessica A Turner, Peter E Turkeltaub, Jack L Lancaster, and Peter T Fox. The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. *BMC Research Notes*, 4(1):349, December 2011.
- [8] Jessica A. Turner and Angela R. Laird. The Cognitive Paradigm Ontology: Design and Application. *Neuroinformatics*, 10(1):57–66, January 2012.
- [9] Jonas Vlasselaer, Angelika Kimmig, Anton Dries, Wannes Meert, and Luc De Raedt. Knowledge Compilation and Weighted Model Counting for Inference in Probabilistic Logic Programs. In *The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence Beyond NP*, page 6, 2016.
- [10] Tal Yarkoni, Russell A. Poldrack, Thomas E. Nichols, David C. Van Essen, and Tor D. Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665–670, June 2011.