



**HAL**  
open science

# RDFMiner : Grammar-Based Genetic Programming for Mining OWL 2 Axioms

Rémi Felin, Andrea G. B. Tettamanzi

► **To cite this version:**

Rémi Felin, Andrea G. B. Tettamanzi. RDFMiner : Grammar-Based Genetic Programming for Mining OWL 2 Axioms. 3IA Annual Scientific Meeting, Nov 2021, Biot, France. hal-03517059

**HAL Id: hal-03517059**

**<https://hal.science/hal-03517059>**

Submitted on 7 Jan 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.

# RDFMiner: Grammar-Based Genetic Programming for Mining OWL 2 Axioms.

Research axis of the 3IA: **Fundamental AI**

Rémi FELIN and Andrea G. B. Tettamanzi

Université Côte d'Azur, Inria, CNRS, I3S - remi.felin@etu.univ-cotedazur.fr - andrea.tettamanzi@univ-cotedazur.fr

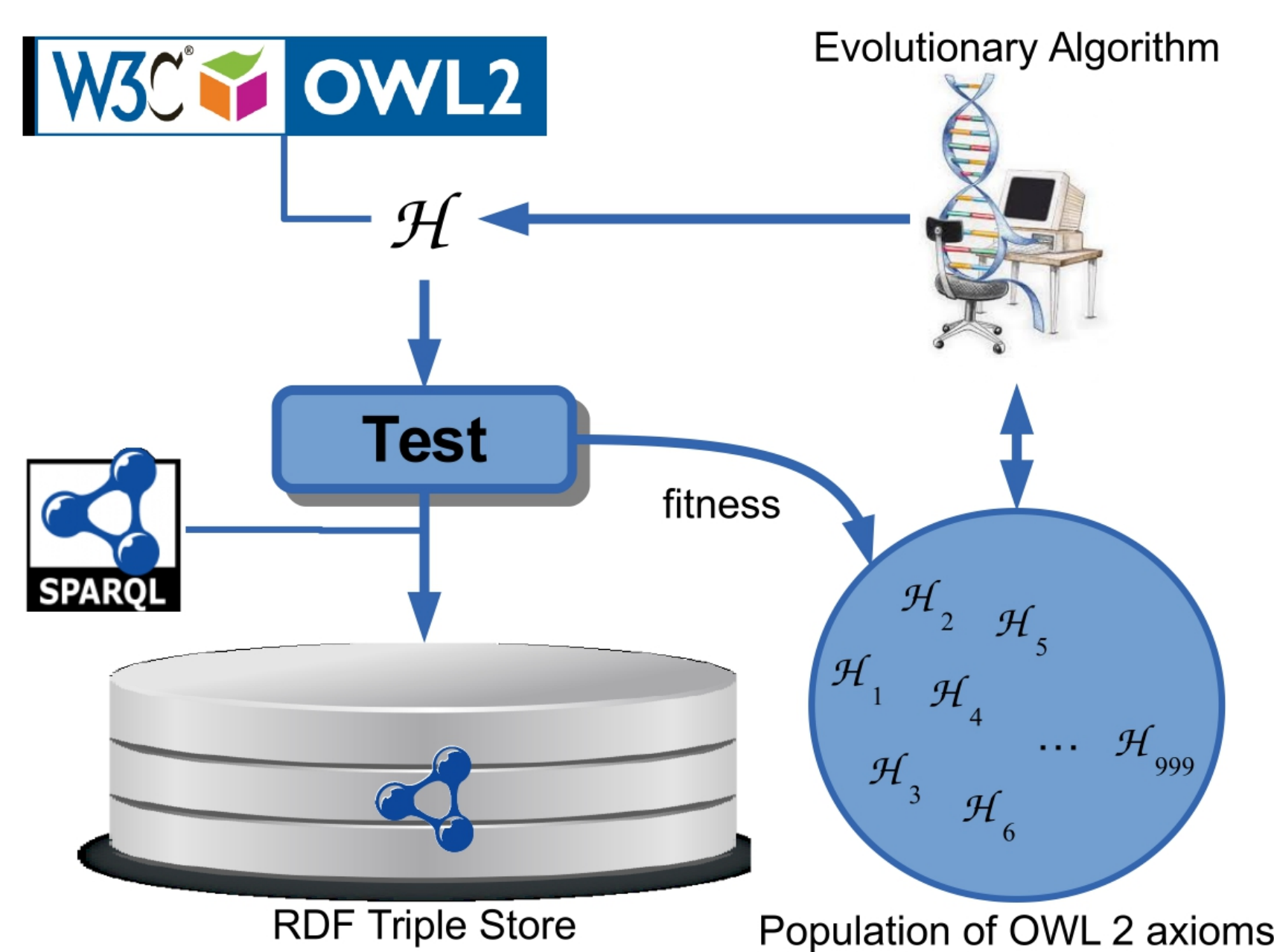


## 1. Introduction

Ontology enrichment is a key task in the area of the Semantic Web. It allows directly to enrich the links between entities of the Semantic Web and thus adding information. Our research area is part of this objective with the search for axioms using an evolutionary process. In this poster, we propose an adaptation resulting from the coupling between evolutionary algorithms and the theory of possibilities in order to allow the research of axioms of subsumption composed of complex classes.[1]

## 2. Objectives

We use an evolutionary approach, namely **Grammatical Evolution** [2]. Using a predefined grammar in *BNF format*, we can generate well-formed candidate axioms. Of course, this simple process is not sufficient to obtain axioms that is meaningful. For this purpose, we use an evolutionary process based on this grammar which considers axioms as individuals in a population. Each individuals are evaluated to define its *fitness* value, which we aim to maximise. The research of **examples** and **counter-examples** in a knowledge graph (using SPARQL) is used in our fitness calculation function. To do this, we apply the possibility theory [3].



## 3. Possibilistic OWL Axioms Scoring

In order to assess the possibility and necessity of an axiom, let us consider  $v_{\phi}^{+}$  the confirmations observed among the elements of  $v_{\phi}$  and  $v_{\phi}^{-}$  the counterexamples observed. We define the possibility  $\Pi(\phi)$  and necessity  $N(\phi)$  of an axiom as follows:

$$\Pi(\phi) = 1 - \sqrt{1 - \left(\frac{v_{\phi} - v_{\phi}^{-}}{v_{\phi}}\right)^2},$$

$$N(\phi) = \sqrt{1 - \left(\frac{v_{\phi} - v_{\phi}^{+}}{v_{\phi}}\right)^2},$$

if  $\Pi(\phi) = 1$ , 0 otherwise.

The fitness is evaluated using the possibility and necessity for a given axiom, that we maximize at each iteration. We apply the formulas such as:

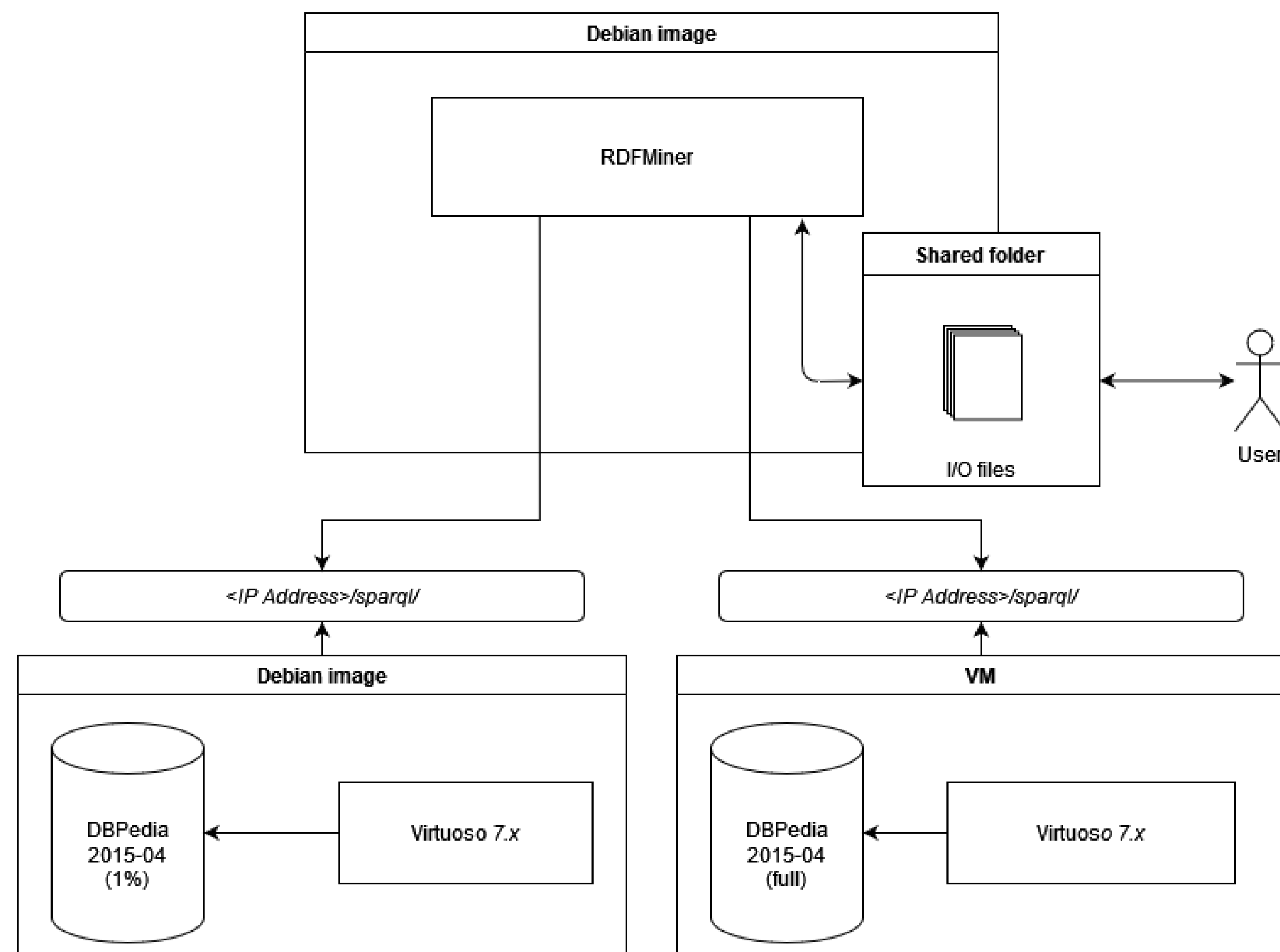
$$f(\phi) = v_{\phi} \times \frac{\Pi(\phi) + N(\phi)}{2}.$$

## 7. References

- [1] Rémi Felin and Andrea G.B. Tettamanzi. Using grammar-based genetic programming for mining subsumption axioms involving complex class expressions. Melbourne, Australia, December 2021. IEEE/WIC/ACM International Conference on Web Intelligence.
- [2] Thu Huong Nguyen and Andrea G.B. Tettamanzi. An evolutionary approach to class disjointness axiom discovery. In *IEEE/WIC/ACM International Conference on Web Intelligence, WI '19*, page 68–75, New York, NY, USA, 2019. Association for Computing Machinery.
- [3] Zadeh L.A. Fuzzy sets as a basis for a theory of possibility. In *Fuzzy Sets and Systems 1*, pages 3–28, 1978.
- [4] Andrea G. B. Tettamanzi, Thu Huong Nguyen, and Rémi Felin. Rdfmining project. <https://github.com/RemiFELIN/RDFMining>.

## 4. RDFMiner Application

The source code of the project is available here [4].



## 5. Experiments

We ran our algorithm **10 times** with populations equal to **100, 200, and 500**, respectively, with the idea of comparing the results and observing which parameters are optimal in order to obtain credible axioms. We set a time limit of **30 seconds** which seems to represent a good compromise between a fair evaluation of the axioms and obtaining results in an acceptable time. The experiments provided us with a collection of axioms evaluated on the *training* and on the *full dataset* of **DBPedia 2015-04** after **30 generations** of our algorithm, thus providing us with two distinct sets of measures.

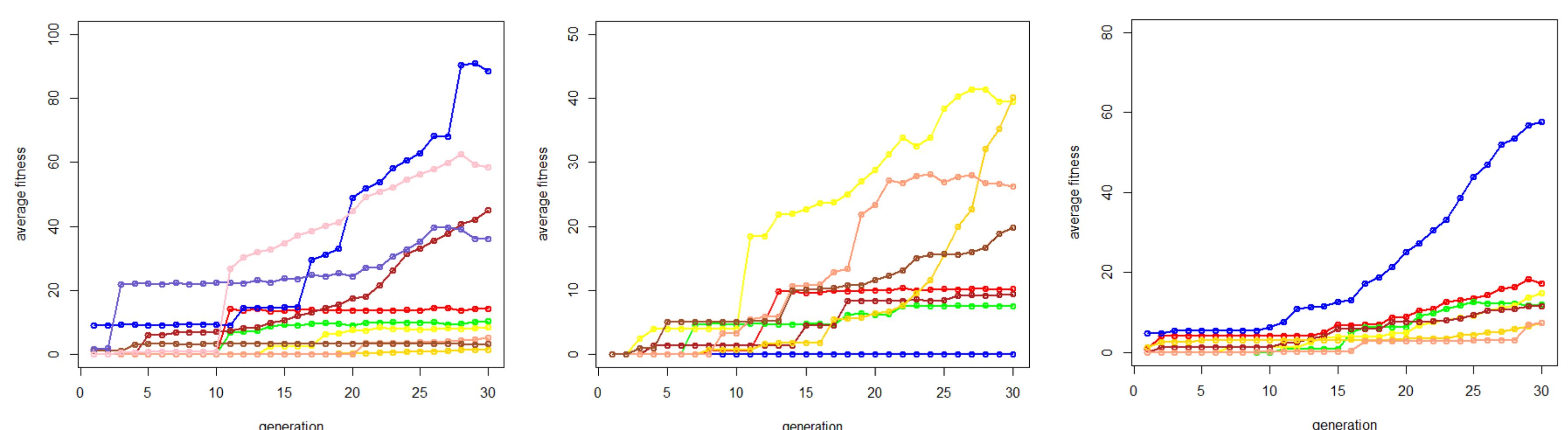
Parameter	Values
popSize	100 ; 200 ; 500
Total effort: k	3000 ; 6000 ; 15000
initLenChrom	6
pCross	80%
pMut	1%
time-cap	30 seconds

## 6. Results

We notice that a large part of the axioms found and evaluated from our training dataset, keep a non-zero possibility in our full dataset, which confirms the suitability of this approach.

Parameter	Number of axioms	
	with training dataset	with full dataset
100	476	358
200	525	525
500	1911	1888

We obtain quite good performances regarding the global evaluation with our evolutionary algorithm, where we see that the curves are quite high overall over 30 generations with notable variations and some rare runs where the fitness is quite low.



**Evolution of average fitness over 10 executions with the same parameters (those that have been successfully completed) for a population of, respectively, 100, 200 and 500 axioms.**