

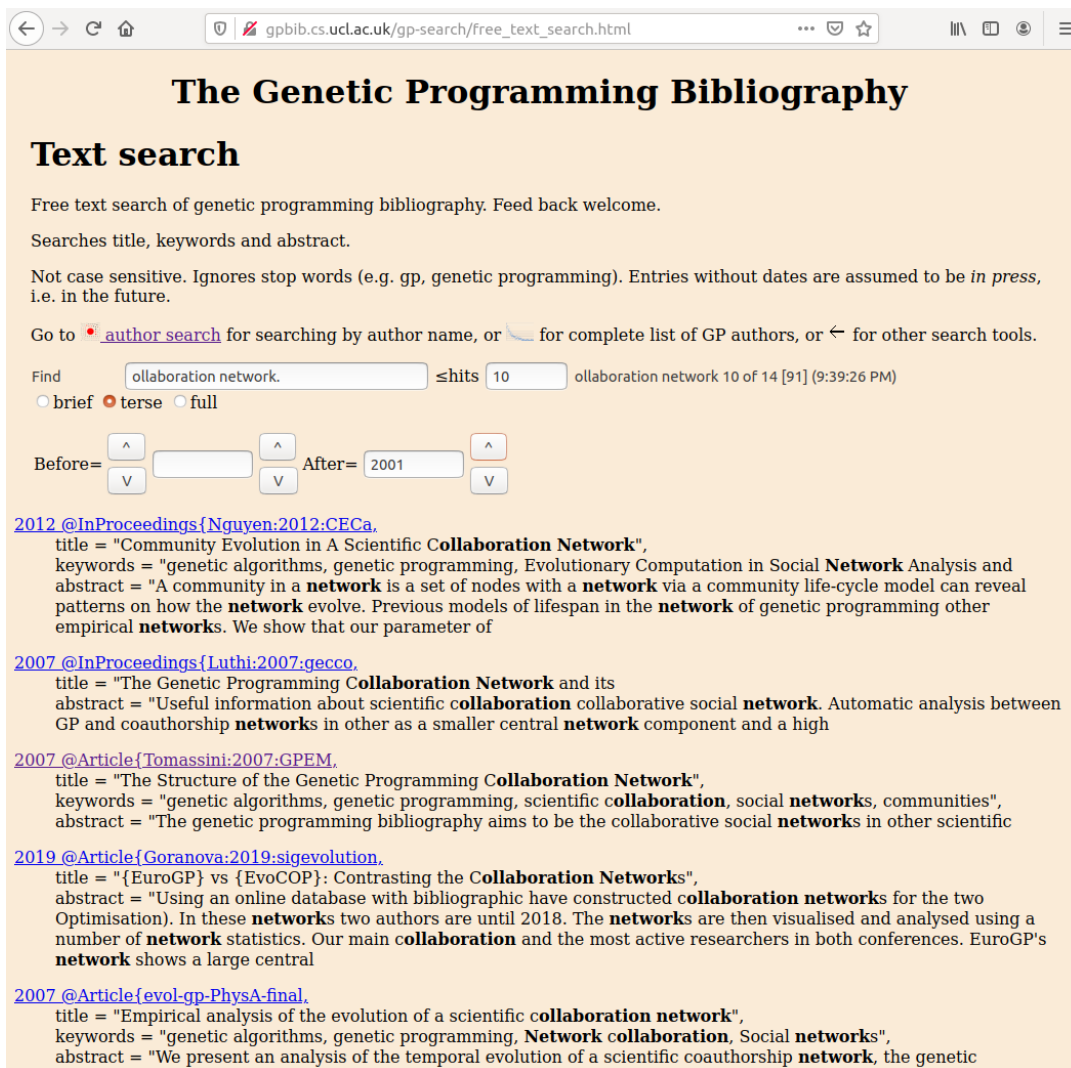
Searching the Genetic Programming Bibliography

W. B. Langdon

29 February 2024

Abstract

Since February 2024 the genetic programming bibliography has supported free text queries to find GP papers via <http://gpbib.cs.ucl.ac.uk/gp-search/>



The screenshot shows a web browser window with the URL gpbib.cs.ucl.ac.uk/gp-search/free_text_search.html. The page title is "The Genetic Programming Bibliography" and the section is "Text search".

Free text search of genetic programming bibliography. Feed back welcome.
Searches title, keywords and abstract.
Not case sensitive. Ignores stop words (e.g. gp, genetic programming). Entries without dates are assumed to be *in press*, i.e. in the future.

Go to [author search](#) for searching by author name, or [home](#) for complete list of GP authors, or [back](#) for other search tools.

Find ollaboration network 10 of 14 [91] (9:39:26 PM)

brief terse full

Before= After=

[2012 @InProceedings{Nguyen:2012:CECa.}](#)
title = "Community Evolution in A Scientific **Collaboration Network**",
keywords = "genetic algorithms, genetic programming, Evolutionary Computation in Social **Network** Analysis and
abstract = "A community in a **network** is a set of nodes with a **network** via a community life-cycle model can reveal patterns on how the **network** evolve. Previous models of lifespan in the **network** of genetic programming other empirical **networks**. We show that our parameter of

[2007 @InProceedings{Luthi:2007:gecco.}](#)
title = "The Genetic Programming **Collaboration Network** and its
abstract = "Useful information about scientific **collaboration** collaborative social **network**. Automatic analysis between GP and coauthorship **networks** in other as a smaller central **network** component and a high

[2007 @Article{Tomassini:2007:GPPEM.}](#)
title = "The Structure of the Genetic Programming **Collaboration Network**",
keywords = "genetic algorithms, genetic programming, scientific **collaboration**, social **networks**, communities",
abstract = "The genetic programming bibliography aims to be the collaborative social **networks** in other scientific

[2019 @Article{Goranova:2019:sigevolution.}](#)
title = "{EuroGP} vs {EvoCOP}: Contrasting the **Collaboration Networks**",
abstract = "Using an online database with bibliographic have constructed **collaboration networks** for the two Optimisation). In these **networks** two authors are until 2018. The **networks** are then visualised and analysed using a number of **network** statistics. Our main **collaboration** and the most active researchers in both conferences. EuroGP's **network** shows a large central

[2007 @Article{evol-gp-PhysA-final.}](#)
title = "Empirical analysis of the evolution of a scientific **collaboration network**",
keywords = "genetic algorithms, genetic programming, **Network collaboration**, Social **networks**",
abstract = "We present an analysis of the temporal evolution of a scientific coauthorship **network**, the genetic

Figure 1: GP Bibliography Web Browser Keyword Query Interface

1 Background

The Genetic Programming bibliography aims to cover all papers, books, PhD thesis, etc. on genetic programming. It was started by John Koza and first published in 1994 as Appendix F of his second GP book [1]. I took on the bibliography, and was invited by Pete Angeline to provide it as an appendix to the second “Advances in Genetic Programming” book, which he co-edited with Kim Kinnear (published in 1996 by MIT Press [2]). Since then the bibliography has been available online, firstly by FTP and now via http. Whilst Steve Gustafson was based in the UK, he assisted in the maintenance and growth of the bibliography, including designing its home page. For almost all of the first twenty years it was hosted by The Birmingham University School of Computer Science. Then in the summer of 2019 it moved to its new host <http://gpbib.cs.ucl.ac.uk>, back in the Computer Science department of University College, London. Since the autumn of 2019, with the support of Jason Moore, the Perelman School of Medicine in the University of Pennsylvania has hosted a mirror the bibliography and supporting web pages: <http://gpbib.pmacs.upenn.edu>. The ACM’s Special Interest Group on Evolutionary Computation (SIGEVO)’s newsletter SIGEVolution [4, 5] and the Genetic Programming and Evolvable Machines and other journals have published articles on the GP bibliography [3, 7, 8, 6] or analysis of data derived from it [9, 11, 10].

2 Online Data, Web Searches and Unavailable Traditional Queries

The whole bibliography is held in a single bibtex `gp-bibliography.bib`. At the time of writing, it contains 17921 entries (16696 directly related to genetic programming), 38 MBytes. Originally the intention was `gp-bibliography.bib` would just be used as a bibtex file with references being extracted when people wrote papers in L^AT_EX. However uses of the bibliography have expanded. Although many bibliography tools support bibtex, copies in refer format `gp-bibliography.ref` and as plain text `gp-bib-alpha.txt` are also available.

Perhaps the most widely used interface is that each GP paper has its own HTML web page (e.g. http://gpbib.cs.ucl.ac.uk/gp-html/koza_2003_gpt.html). Almost all have hyper text links to the paper itself, its publisher and to its Google Scholar citations. Since these pages are freely available, they make attractive targets for web browsing and web searches and so the GP bibliography is easily searchable via Google and other web search engines.

For thirty years (1993–2023) The Collection of Computer Science Bibliographies (CCSB)¹ hosted hundreds of bibliographies, including the genetic programming bibliography. It provided a number of search interfaces, including allowing online searches using Lucerne syntax. Since the GP bibliography was part of CCSB, it could be searched in a sophisticated way using the CCSB’s web interfaces. However CCSB was retired last summer and when it was switch off, its search interface for the GP bibliography was also removed.

3 The New Free Text Search Interface

As mentioned in the previous section, the CCSB sophisticated search interface was removed in July 2023, the goal of its replacement <http://gpbib.cs.ucl.ac.uk/gp-search/> is to provide something light weight, which is interactive, fast, immediate and without a learning curve, Figure 1.

Following the success of the co-author search interface <http://gpbib.cs.ucl.ac.uk/gp-coauthors>, the new free text search interface is also based in the user’s smart phone, using JavaScript running in a web browser.

¹See Wikipedia entry “Collection of Computer Science Bibliographies” https://en.wikipedia.org/wiki/Collection_of_Computer_Science_Bibliographies

3.1 Single Web Page

The interface consists of a single HTML web page² containing a web form with various input and output forms and buttons, see Figure 1.

3.1.1 Find keywords

The first input field allows the user to enter keywords to be search for.

Stop words are ignored, e.g. words of one letter, “in”, “the”, “from” and “model”. Also since every paper will automatically match “genetic” “algorithms” and “programming” (after all this is a bibliography dedicated to Genetic Programming) these and “gp” are ignored.

To avoid overly long time consuming searches, queries that match more than 1000 papers are cut short.

3.1.2 Limiting the number of matches

By default, up to ten hits are displayed. The “≤hits” input field allows the number of papers displayed to be changed (up to a maximum of 100).

3.1.3 3 types of output: brief terse full radio buttons

brief The brief display requires no further interaction from the host (or mirror) and simply displays the bibtex key of the matching GP paper and gives a hyper link to its web page within the GP bibliography.

terse The terse output format shows fragments of the matching paper’s entry with the matching parts highlighted. There is also a hyper link to the corresponding full web page. This is the default.

full shows the complete bibtex of the matching GP paper. This often includes links to the paper itself, the publisher of the article and to its Google Scholar citations.

3.1.4 Matching by year

The lower two input areas allow the user to specify the years they are interested in. Years are four digit numbers.

The up buttons \wedge increase the year by one. The down buttons \vee decrease it by one. To allow easy sweeps of papers published in a range of years, the central \wedge/\vee buttons increment or decrement both before and after years by the same amount.

3.1.5 Search summary statistics

The output display reflects the query (without stop words) in the order the query’s words are used, the number of matching papers displayed and the total number of matches. If (e.g. for speed) the search is truncated, so that the number of matches is not known, a + is appended. (The total number of searches is given in square brackets [] but it is often only useful for debugging.)

3.2 Ordering Results

The interface attempts to display the best matches at the top of the list of entries. This ordering is based on a heuristic which weights hits with matches that occur together more highly, matches on titles above those on keywords, above those on abstracts. Matches in long abstracts carry less weight but matches against exact words or matches which match case exactly are up weighted.

²http://gpbib.cs.ucl.ac.uk/gp-search/free_text_search.html

3.3 Feedback

It would be great to hear of people's experiences with the new interface, bug reports, errors and future ideas for the search tool or for the GP bibliography in general.

I asked a number of authors of recent computer science surveys in various fields unrelated to Genetic Programming how they used other online search interfaces, such as those provided by DBLP, IEEE Xplore, the ACM Digital Library. The most useful feature seemed to be the ability to search by keywords. This is what the new interface provides. Its underlying data structures are public and so interested parties could build additional features on top of them. Main stream digital libraries and online bibliography include other tools, which the authors of surveys found useful but which are not included in the GP bibliography or only indirectly supported by it.

Perhaps the SIGEVO newsletter could be a venue to enable everyone to hear of people's experiences with the new search tool and to know of their thoughts on it or possible extensions.

4 Adding Your GP Papers

The Genetic Programming bibliography has long had a web interface allowing authors to provide missing GP papers, however in practice the easiest way is simply to email a bibtex description of missing GP works to me.

Acknowledgements

I would like to thank Una-May O'Reilly who first highlighted the need for a new search tool, Aymeric Blot who highlighted the problem of scaling existing web browser search interfaces from 200 papers to 20000, Carmen Meinson who suggested stylish improvements and Farooq Karimi Zadeh, Justyna Petke, Carol Hanna, and UCL students, for testing it.

References

- [1] J. R. Koza. *Genetic Programming II: Automatic Discovery of Reusable Programs*. MIT Press, Cambridge Massachusetts, May 1994.
- [2] W. B. Langdon. A bibliography for genetic programming. In P. J. Angeline and K. E. Kinnear, Jr., editors, *Advances in Genetic Programming 2*, appendix B, pages 507–531. MIT Press, Cambridge, MA, USA, 1996.
- [3] W. B. Langdon. Genetic programming and evolvable machines: Books and other resources. *Genetic Programming and Evolvable Machines*, 1(1/2):165–169, Apr. 2000.
- [4] W. B. Langdon. Web usage of the GP bibliography. *SIGEVolution newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation*, 1(4):16–21, Dec. 2006.
- [5] W. B. Langdon. News of the GP bibliography. *SIGEVolution newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation*, 6(3-4):12–16, July 2014.
- [6] W. B. Langdon. Genetic programming and evolvable machines at 20. *Genetic Programming and Evolvable Machines*, 21(1-2):205–217, June 2020. Twentieth Anniversary Issue.
- [7] W. B. Langdon and S. Gustafson. Genetic programming and evolvable machines: Five years of reviews. *Genetic Programming and Evolvable Machines*, 6(2):221–228, June 2005.

- [8] W. B. Langdon and S. M. Gustafson. Genetic programming and evolvable machines: ten years of reviews. *Genetic Programming and Evolvable Machines*, 11(3/4):321–338, Sept. 2010. Tenth Anniversary Issue: Progress in Genetic Programming and Evolvable Machines.
- [9] W. B. Langdon, R. Poli, and W. Banzhaf. An eigen analysis of the GP community. *Genetic Programming and Evolvable Machines*, 9(3):171–182, Sept. 2008.
- [10] M. Tomassini and L. Luthi. Empirical analysis of the evolution of a scientific collaboration network. *Physica A*, 385:750–764, 2007. Available online 25 July 2007.
- [11] M. Tomassini, L. Luthi, M. Giacobini, and W. B. Langdon. The structure of the genetic programming collaboration network. *Genetic Programming and Evolvable Machines*, 8(1):97–103, Mar. 2007.