Genetic Programming and Evolvable Machines (2020) 21:205–217
DOI:10.1007/s10710-019-09344-6

Genetic Programming and Evolvable Machines at 20

W. B. Langdon

15 Mar 2019

Abstract The journal and in particular the resource reviews have been running for twenty years. We summarise the GP literature, including top papers and authors, as seen by users of the genetic programming bibliography. Then revisit our original goals for GPEM book reviews and compare them with what has achieved.

1 Genetic Programming Publications

The GP literature continues to grow [1; 2; 3]. We start by giving details, including breaking them down by publication type and number of authors, and note the increasing numbers of co-authors involved with GP. Section 1.1 notes the place of GP in the computer science literature and the increasing use of on-line publication. Whilst Section 1.2 reflects on the changed use of the bibliography since 2009 and lists the most downloaded papers, Table 1, and their authors, Table 2. Section 1.3 dispels the myth that only recent publications are ever read. Section 1.4 reflects on what still needs to be done for the GP bibliography. Finally, Section 2, returns to the journal's 20^{th} birthday.

These statistics are derived from the GP-bibliography. The bibliography is available from the Internet in a variety of formats and locations. Also entries from it have been incorporated in other online bibliographic resources. Unfortunately, no similar effort has been undertaken for the literature on evolvable hardware, so we initially deal only with GP before returning to the journal in general in Section 2.

At the start of 2019 there were 11 906 GP entries in the GP bibliography (excluding late breaking papers, unpublished, miscellaneous, master thesis, undergraduate student reports and some short posters). This is more than

W. B. Langdon
Department of Computer Science, University College London
E-mail: W.Langdon@cs.ucl.ac.uk

twice the number in 2009. Figure 1 shows the number of each entry according to when they where published and by type. Naturally most papers were published in conference proceedings. Initially there was an exponential rise, with the number of publications doubling every year from 1988 to 1996. This has been followed by a rapid linear increase since 1997. (Figure 2 right shows that this is a typical behaviour for Computer Science bibliographies.) In the case of GP the effort available to maintain the bibliography has not increased exponentially. Hence the bibliography has lagged behind the growth in the field. Unfortunately there have been GP publications which have escaped recording in the bibliography. This leads to a bias in favour of those researchers who actively support the maintenance of the bibliography.

Figure 2 plots the number of people active in GP (i.e. according to the GP bibliography they published in a given year). Figure 2 shows that typically in recent years almost half the authors who published in a given year were new to GP. The total number of people who have published GP related papers is about 8000 (More than twice as many as in 2009.) The number of new authors per year, along with the rise and fall of publication types like PhD dissertations, gives a sense of how vibrant the GP community is. However, as GP has matured more application papers appear in biology, chemistry and other non-computer science journals. Unfortunately this probably means some recent articles are missing.

Figure 3 shows the distribution of the number of authors per GP publication for each year (excluding MSc. theses, unpublished, etc.). We can see that as the field took off in the 1990s, the publications were dominated by papers with one or two authors. However, as Steve Gustafson pointed out, the number of papers with three, four or five authors, steadily increased. Figure 3 shows that this trend is still continuing. This makes sense if we consider that as applications and analysis has matured, more collaboration is taking place resulting in multiple authors. Also, as GP is applied to other disciplines, we would expect to see more co-authors appearing on GP publications.

1.1 The Genetic Programming Literature

The GP bibliography can be searched via the collection of computer science bibliographies. The Artificial Intelligence collection of bibliographies is the 5^{th} largest computer science collection by subject. And within the AI collection, the GP bibliography is the 2^{nd} largest (up from 4^{th} in 2009).

Online electronic versions of papers, including those on publisher web pages, can be directly linked to the bibliography. There are 11809 GP publications with such links (almost three times that in 2010). In almost all cases, the paper is actually available via its links. In other words the text of 92% of GP papers are immediately available via hyperlinks in the bibliography. This is 16% percent more than nine years ago. Most of the change comes from the increased proportion of online journal articles (99% v. 92%), chapters in edited books (97% v. 46%), conference proceedings (96% v. 73% and PhD

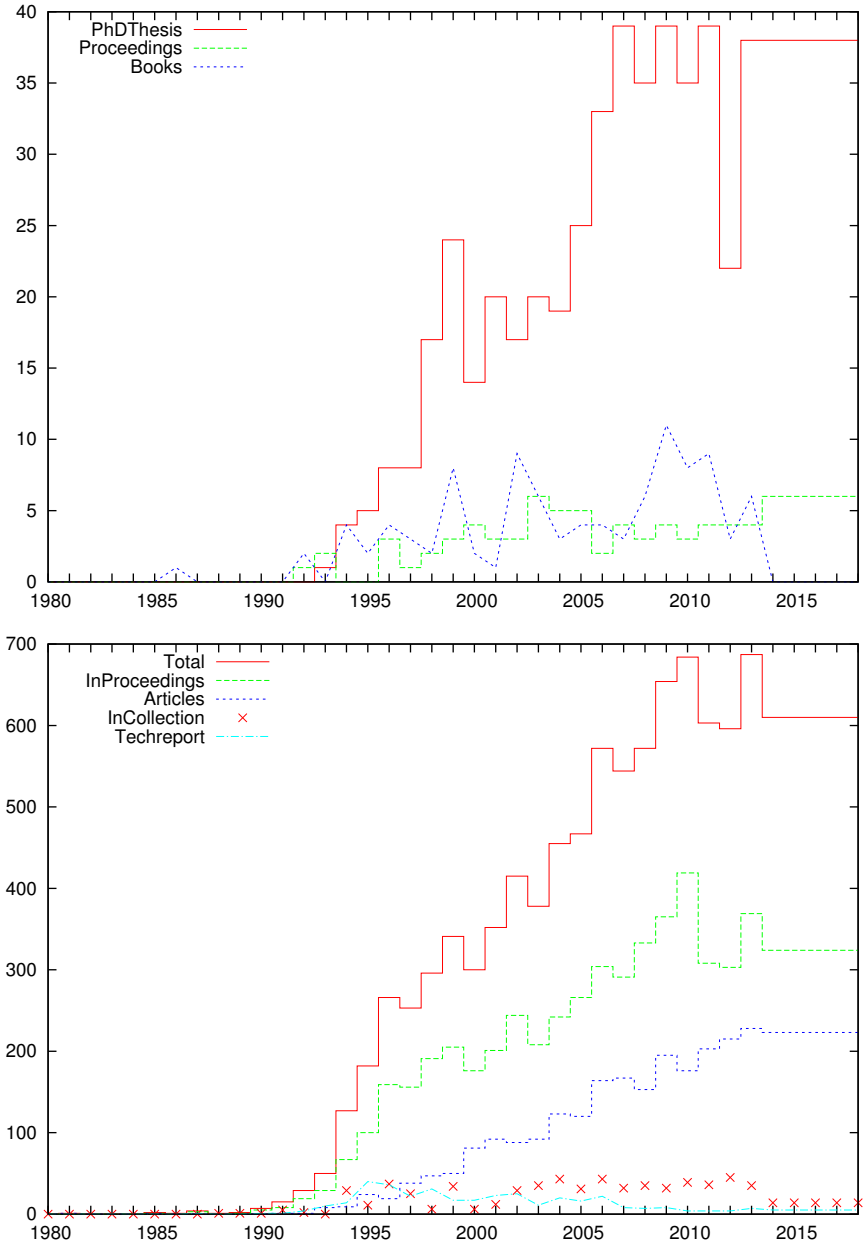

Fig. 1 Number of genetic programming entries in the GP bibliography, according to when they where published and by type. We have excluded late breaking papers, unpublished, misc, masters thesis, undergraduate students and some short posters. Note change in vertical scale between top and bottom.

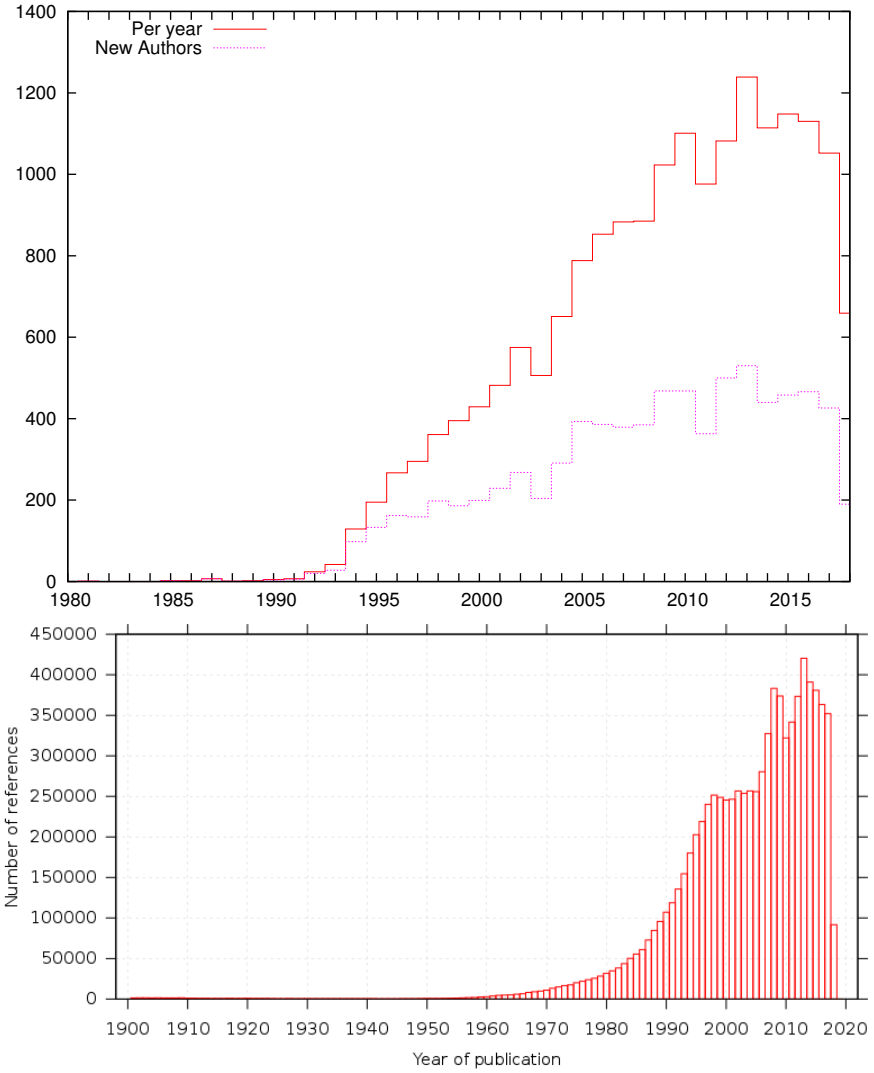

Fig. 2 Top: Number of authors of genetic programming entries in the GP bibliography (1980 – 2018). We exclude late breaking papers, unpublished, misc, masters thesis, undergraduate students and some short posters. Bottom: Distribution of publication dates across the whole of the Computer Science Bibliography Collection showing the GP distribution of publication dates is somewhat typical of computer science bibliographies.

theses (94% v. 63%). The fraction of technical reports (85% v. 87%) is little changed. This increase reflects the general scientific trend to insist on work being online. Many students have only seen the inside of a library as a place to buy coffee and would not dream of looking for interesting research in bound paper volumes anywhere let alone in a library.

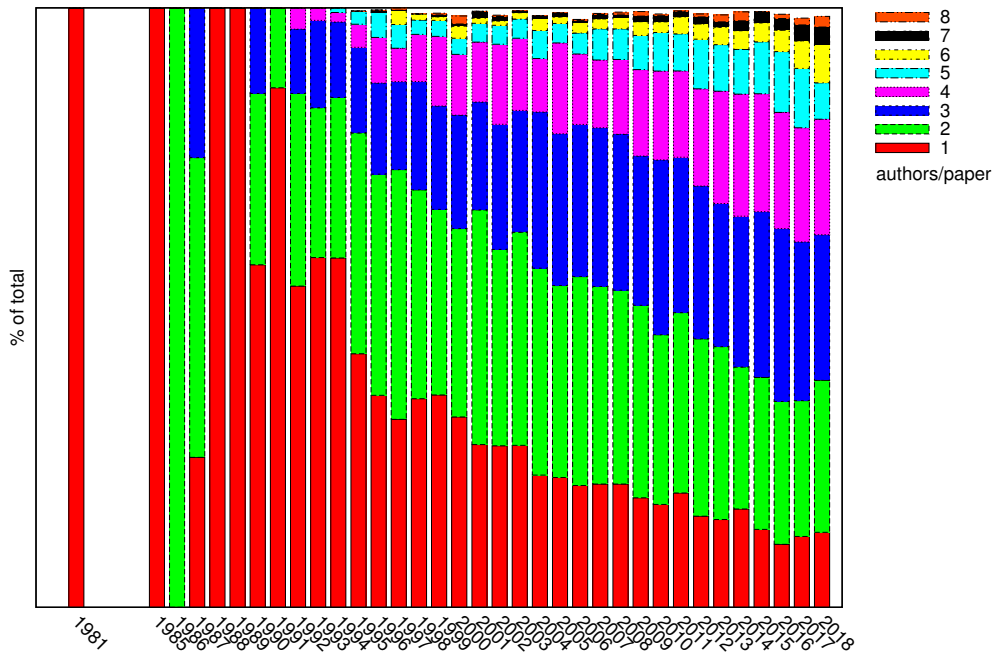

Fig. 3 Increasing proportion of multi-authored GP papers. Colours indicate the number of authors per paper.

1.2 Computer and Human Use of the GP Bibliography Papers

Figure 4 shows that almost all requests to download GP papers are automatically generated by computers interrogating other computers. (After all the Internet was built to connect computers to other computers.) However, since many robots are written to try and conceal their owners and their purposes, it is impossible to be sure that web traffic is correctly allocated. Nonetheless in the rest of this section we shall do our best to discount malicious usage.

Since 2009 the download traffic has changed quite a lot. Firstly, the volume of traffic has increased more than three fold but the the volume of real user direct downloads appears to have fallen by a factor of three. Given the overwhelming volume of computer generated traffic it is especially difficult to be sure when allocating traffic to users. Also this is traffic which reaches the server in Birmingham, other user download requests may be being satisfied by intermediate caches. Indeed in addition to CiteSeer, there are now several web system which host copies of GP papers alongside their publisher's web systems.

In 2009 the two main robots belonged to Google and Yahoo [3, Figs. 5–7]. They still scan GP papers via the bibliography but the other web bots from 2009 no longer comprise more than 1% of the download traffic (and so are omitted from Figure 4). The new main bots are listed in Figure 4.

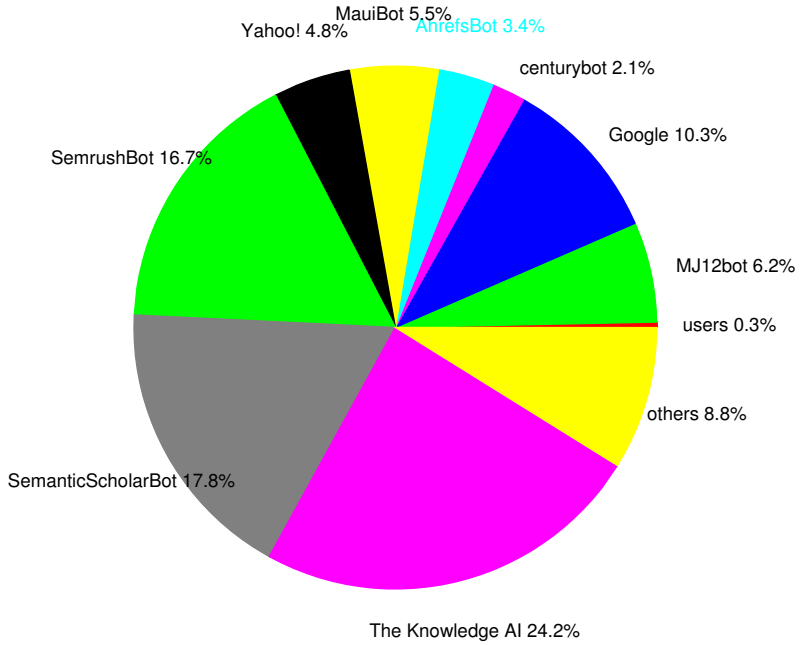

Fig. 4 Web robots used to download GP papers (1 Oct 2017 to 1 Oct 2018). Human activity, "users" 3775, is a small fraction of the total 1 143 402

Robot activity varies radically. Figure 5 suggests a vaguely log-normal distribution, perhaps due to combination of several different power law distributions. Although the download rates appear to be near the average, downloads are in fact clustered. I.e., downloads are related. The spread of download rates is far wider than that which would result from genuinely independent random events. Some 86% of the load appears to be due to robots re-reading papers.

Figure 6 indicates that the bibliography is used continuously (including Christmas). Over the whole year (2018), there are no downloads recorded for just 30 hours. Suggesting the web site was probably down for only one day and a few hours in the whole year.

It appears that on average more than 10 GP papers are downloaded via the bibliography per day by people (about half the rate in 2009).

Considering just downloads by people, over the past twelve years the most popular papers have been tutorials on GP, followed by financial applications, cf. Table 1. This is followed by surveys, user manuals for GP packages and more widely drawn applications. Naturally the most downloaded authors are those with the most online papers linked to the bibliography and authors of popular tutorials or popular finance papers, cf. Table 2. One surprise in Table 1 is how consistent it has been. Since 2009 there have only been four changes. Two were published about 2009 [4; 3] and another's PDF was added to the bibliography long after it was originally published [5].

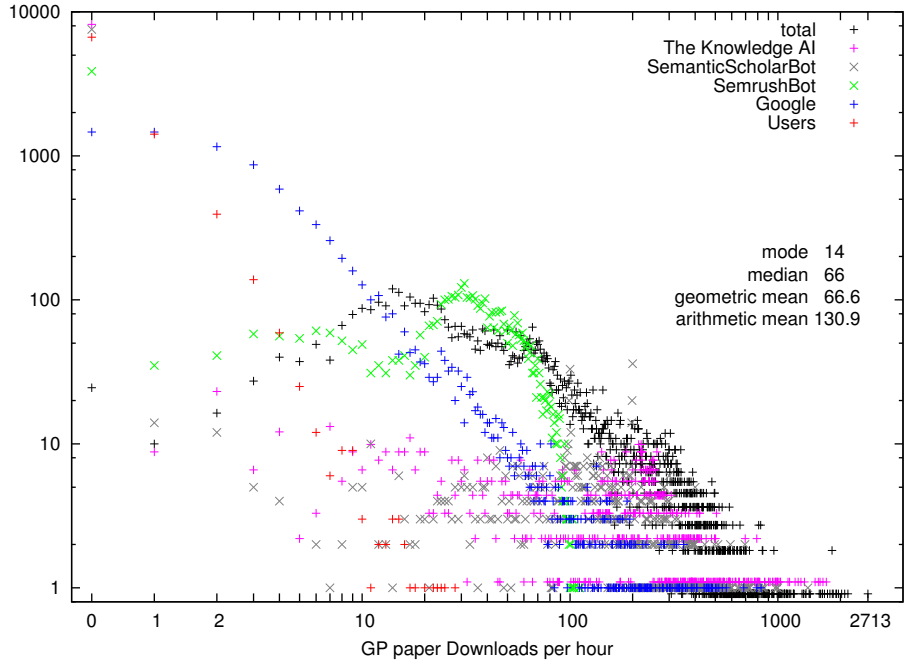

Fig. 5 GP paper downloads per hour (1 Oct 2017 to 1 Oct 2018). Note log scales. User activity + and Google bot + follow approximate power laws (x^{-3} and $x^{-3/2}$). SemanticScholarBot (grey ×) has prolonged periods when it is not active (0 7549 hours, more than ten months in total). Similarly SemrushBot (green ×) has inactive periods and indeed was off from November 2018 for three months. Whilst "The Knowledge AI" (red +) activity was concentrated into two periods May–June and September again with many inactive periods. However these three and the total have some similarity to a parabola suggesting log-normal distributions. The non-Gaussian distribution of the total downloads is emphasises by the differences between the four types of average reported.

When calculating Tables 1 and 2 we have been as scrupulous as possible to ensure we include only real personal down loads and exclude all web robots. Unfortunately this is not easy and so we have deliberately erred on the cautious side. By excluding cases where we are not sure, we will underestimate. Nonetheless the results should still give a fair indication of use of the GP bibliography by people.

GP papers downloaded by people via the bibliography are mostly accessed using commercial Internet service providers (ISPs). Even the most active university, UCD, is only sixth. This could be because universities may have subscriptions which encourage academics to search via publishers' web pages or simply because most people access the Internet via ISPs.

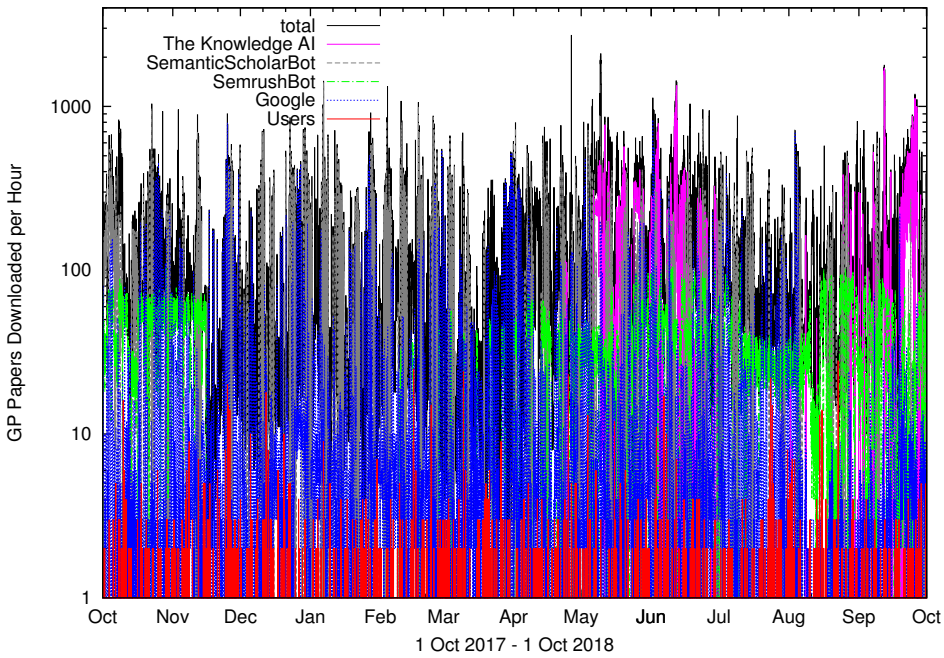

Fig. 6 A year's use of the GP bibliography links to papers (black) and broken down by major web bots and users (colour). The infrequent peaks of very high activity (max 2713) are less emphasised by the log vertical scale.

1.3 Use of the GP Papers by Year

It has been suggested that the GP community might have lost touch with its roots. That is, are people nowadays not interested in older papers? Perhaps this is a fear in every scientific community but the GP data, as far as we can interpret it, Figure 7, does not seem to support it. Instead Figure 7 shows a strong trend towards people still downloading established as well as newer work.

For several months the bibliography did have active summary data of citations, gathered from Google Scholar [24]. However the data gathering proved unstable and we fell back to providing links to Google Scholar, which allow users to do their own citation queries.

1.4 The Future of the GP Bibliography

The major work must remain keeping the bibliography as up to date as possible. Not just adding new GP publications but also ensure new entries have working hyperlinks, especially URLs or DOIs, to the publications themselves.

Although a web based tool (WBT) to enable people to add their own BibTeX to the bibliography has been available since 2000, it has fallen out of

Table 1 20 Frequently Downloaded GP Bibliography Papers. From Sep 2006 to start 2019

Reference	Title
[6]	A Genetic Programming Tutorial
[7]	An automated FX trading system using adaptive reinforcement learning
[8]	Genetic Programming with Wavelet-Based Indicators for Financial Forecasting
[9]	A real-time adaptive trading system using genetic programming
[10]	A SIMD Interpreter for Genetic Programming on GPU Graphics Cards
[4]	A field guide to genetic programming
[5]	Rule Induction in Forensic Science
[11]	Computational learning techniques for intraday FX trading using popular technical indicators
[3]	Genetic Programming and Evolvable Machines: ten years of reviews
[12]	Intraday FX Trading: An Evolutionary Reinforcement Learning Approach
[13]	Genetic Programming Software to Forecast Time Series
[14]	BEAGLE A Darwinian Approach to Pattern Recognition
[15]	The Profitability of Intra-Day FX Trading Using Technical Indicators
[16]	Data Mining using Genetic Programming: Classification and Symbolic Regression
[17]	Evolutionary reinforcement learning in FX order book and order flow analysis
[18]	lilgp 1.01 User's Manual
[19]	Solving the Graph Coloring Problem using Genetic Programming
[20]	Survey of genetic algorithms and genetic programming
[21]	Adaptive systems for foreign exchange trading
[22]	GP-Robocode: Using Genetic Programming to Evolve Robocode Players

Table 2 Twenty most Downloaded GP Bibliography Authors.

John Koza	William B Langdon	Riccardo Poli	Michael Dempster
Mahmoud Kaboudan	Wolfgang Banzhaf	Jin Li	Vasco Leemans
Richard Forsyth	Jeroen Eggermont	Conor Ryan	Michael O'Neill
Graham Bates	Chris M Jones	Ian W Evett	Maarten Keijzer
Justine W Shen	Sean Luke	Frederic Gruau	Yazann Romahi

favour with authors. A wiki based tool allowed authors to update the URL of their home pages attracted malicious users and had to be abandoned after two years. This experience, dampened enthusiasm for updating the WBT approach to use a wiki approach. However perhaps it may become time to review this decision? Alternatively perhaps a more collaborative approach, which seems to works well with open source programming code, perhaps based on GitHub or something similar might replace WBT and email.

The bibliography is primarily available in BibTeX and refer formats. BibTeX is used by LaTeX whilst there are several tools for importing refer into microsoft word. (There is also a plain text version.) Although all three versions may seem large, gp-bibliography.bib, gp-bibliography.ref and gp-bib-alpha.txt are each less than thirty megabytes. This is well within the means of most computers. (In total the bibliography occupies some 9 GB.)

Indeed, over the years, the speed of communication links to Birmingham (today over 100MB per second) and typical processing power and storage on even modest consumer computers has grown much faster than these files. Even the search facilities in a screen based editor, such as Emacs, can be readily used to find suitable references. Hence it seems reasonable to keep distributing the whole bibliography in these files. So far, there does not seem to be a great need

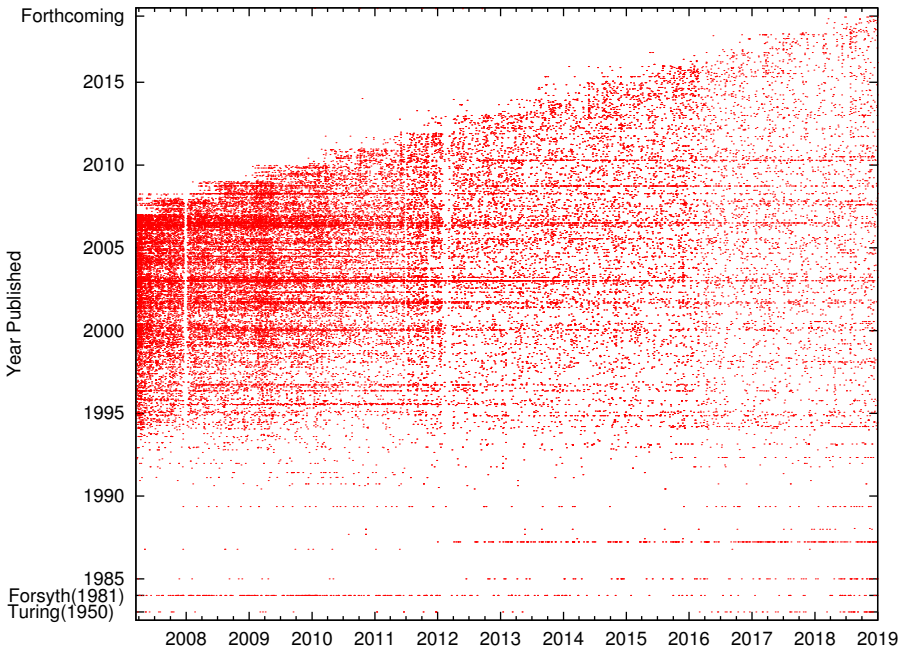

Fig. 7 User downloads via the GP bibliography. There are a few outages without logs (white vertical stripes). Clearly some papers are more popular than others (horizontal stripes). But even more than 25 years after [23] was published people are still downloading older publications as well as new ones.

to support other formats. However typical interactive web browsing patterns are not well supported by these monolithic files.

Since 2000 the bibliography has also been available via individual HTML web pages (one per GP entry and one per author). Uses can readily search these pages via Google and other commonly used Internet search engines. These are typically much more flexible than traditional library search tools, that often employ Lucene syntax or even cruder techniques. Our HTML pages were designed for desktop web browsers. Although available for eighteen years, it may be time to update them for users of small screen devices such as mobile smart phones.

Although we have retained direct link's to Google's Scholar, the previous section described the failure of our initial attempt to provide citation information directly. Undoubtedly citation information could be extracted from the GP papers themselves and provided to the community, however it seems unlikely this could be done in a straight forward way which could compete with Scholar. Nevertheless there may be other avenues or APIs to obtain GP citation network information and present it via the bibliography.

Figure 8 in [3] depicts a web based tool for displaying co-authorship links within the bibliography. Although the tool is still available, there are now more sophisticated, more interactive, tools which might be used to do this.

Since the raw data for [3, Figure 8] is distributed on-line, it may be that new tools to present GP co-author relationships such as [25] could be added to the bibliography without the need to integrate them into it or even to host them on the same web server.

The first section has been very much about the genetic programming literature and the GP bibliography. In the next section we return to the journal itself and in particular the resource reviews, which I have edited since the journal started.

2 Twenty Years of Resource Reviews

Since the journal started we have published reviews covering 104 topics. It was intended from the start that the journal's "resource reviews" would cover not just books but "resources" in the wider sense, particularly including web pages, on-line resources, packages and products. However most of the published reviews have been of books

We reviewed 79 books plus 14 edited collections/proceedings. We have also published one article on Internet-based resources, seven on software (covering nine packages) and reviewed one commercial product. Topics have included not only genetic programming (42), evolvable hardware (9), genetic algorithms (9) and evolutionary computing in general (9), but also AI (6), robotics (5), artificial development and embryos (3) particle swarm optimisation (2), neuro evolution (2), learning classifier systems, ant colony optimisation, evolutionary programming, data mining (2), evolutionary design and art, games (2) and EC in computer vision (2) and in food science. Reviews have also included books on DNA computing (2), inductive logic programming, quantum computing, cellular automata, intelligent bioinformatics and the history of artificial intelligence.

Articles have been written by authors based in the USA (28), the UK (22), Canada (6), France (6), Australia (4), Brazil (4), Argentina (3), Israel (3), Germany (2), Ireland (2), Italy (2), Mexico (2), New Zealand (2), Spain (2), Sweden (2), Chile, China, the Czech Republic, Finland, Holland, Hong Kong, Jordan, Norway, Pakistan and Singapore.

When the journal started it was intended that, excluding special issues, the last article in every issue should be a review. Since GP/EM continues to publish four issues per volume and typically one of these is a special issue, this makes an average of about three reviews per year. Nowadays the electronic versions of all reviews are freely available and often published as "online first" months before the date of the issue holding them. Therefore a continuous flow, as seen from a journal issue perspective, is less important. So in recent years some issues of GP/EM have had several book reviews and some none. Nonetheless we continue to publish at least three reviews per year. Indeed we have comfortably exceeded this target.

In keeping with readers' suggestions we have continued to try to limit book reviews to about two pages. However reviews of tools are usually longer.

3 Summary

Section 1 described in detail the evolution and use of the GP bibliography. We continue to publish reviews of interesting and relevant resources. The reviews are summarised in Section 2. For practical reasons we have concentrated on books. This trend will probably continue but with the continued inclusion of books from related areas as well as genetic programming and evolvable machines. In summary the journal continues to reflect the expansion of the most exciting field in evolutionary computation.

Acknowledgments

I would like to thank Paul Ortyl, and the many subscribers to the GP discussion list for help with maintaining the GP bibliography.

References

[1] William B. Langdon. Genetic programming and evolvable machines: Books and other resources. *Genetic Programming and Evolvable Machines*, 1(1/2):165–169, April 2000.
[2] 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.
[3] 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, September 2010. Tenth Anniversary Issue: Progress in Genetic Programming and Evolvable Machines.
[4] Riccardo Poli, William B. Langdon, and Nicholas Freitag McPhee. *A field guide to genetic programming*. Published via `http://lulu.com` and freely available at `http://www.gp-field-guide.org.uk`, 2008. (With contributions by J. R. Koza).
[5] Ian W. Evett and E. J. Spiehler. Rule induction in forensic science. In *KBS in Government*, pages 107–118. Online Publications, 1987.
[6] John R. Koza and Riccardo Poli. A genetic programming tutorial. www, 2003.
[7] M. A. H. Dempster and V. Leemans. An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications*, 30(3):543–552, April 2006. Special Issue on Financial Engineering.
[8] Jin Li, Zhu Shi, and Xiaoli Li. Genetic programming with wavelet-based indicators for financial forecasting. *Transactions of the Institute of Measurement and Control*, 28(3):285–297, August 2006.
[9] M. A. H. Dempster and C. M. Jones. A real-time adaptive trading system using genetic programming. *Quantitative Finance*, 1(4):397–413, 2001.
[10] William B. Langdon and Wolfgang Banzhaf. A SIMD interpreter for genetic programming on GPU graphics cards. In Michael O'Neill et al., editors, *Proceedings of the 11th European Conference on Genetic Programming, EuroGP 2008*, volume 4971 of *Lecture Notes in Computer Science*, pages 73–85, Naples, 26-28 March 2008. Springer.
[11] M. A. H. Dempster, Tom W. Payne, Yazann Romahi, and G. W. P. Thompson. Computational learning techniques for intraday FX trading using popular technical indicators. *IEEE Transactions on Neural Networks*, 12(4):744–754, July 2001.
[12] M. A. H. Dempster and Y. S. Romahi. Intraday FX trading: An evolutionary reinforcement learning approach. In Hujun Yin et al., editors, *Proceedings of Third International Conference on Intelligent Data Engineering and Automated Learning - IDEAL 2002*, volume 2412 of *Lecture Notes in Computer Science*, pages 347–358, Manchester, 12-14 August 2002. Springer.

[13] M. A. Kaboudan. Genetic programming software to forecast time series. In *Computing in Economics and Finance*, University of Washington, Seattle, USA, 11-13 July 2003.
[14] Richard Forsyth. BEAGLE a Darwinian approach to pattern recognition. *Kybernetes*, 10(3):159–166, 1981.
[15] M. A. H. Dempster and C. M. Jones. The profitability of intra-day FX trading using technical indicators. Working Paper 35/00, Judge Institute of Management Studies, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, 2000.
[16] Jeroen Eggermont. *Data Mining using Genetic Programming: Classification and Symbolic Regression*. PhD thesis, Institute for Programming research and Algorithmics, Leiden Institute of Advanced Computer Science, Faculty of Mathematics & Natural Sciences, Leiden University, The Netherlands, 14 September 2005.
[17] R. G. Bates, M. A. H. Dempster, and Y. S. Romahi. Evolutionary reinforcement learning in FX order book and order flow analysis. In *IEEE International Conference on Computational Intelligence for Financial Engineering*, pages 355–362, Hong Kong, 20-23 March 2003.
[18] Douglas Zongker and Bill Punch. lilgp 1.01 user's manual. Technical report, Michigan State University, USA, 26 March 1996.
[19] Justine W. Shen. Solving the graph coloring problem using genetic programming. In John R. Koza, editor, *Genetic Algorithms and Genetic Programming at Stanford 2003*, pages 187–196. Stanford Bookstore, Stanford, California, 94305-3079 USA, 4 December 2003.
[20] John R. Koza. Survey of genetic algorithms and genetic programming. In *Proceedings of 1995 WESCON Conference*, pages 589–594. IEEE, 1995.
[21] Mark P. Austin, Graham Bates, Michael A. H. Dempster, Vasco Leemans, and Stacy N. Williams. Adaptive systems for foreign exchange trading. *Quantitative Finance*, 4(4):37–45, August 2004.
[22] Yehonatan Shichel, Eran Ziserman, and Moshe Sipper. GP-Robocode: Using genetic programming to evolve robocode players. In Maarten Keijzer et al., editors, *Proceedings of the 8th European Conference on Genetic Programming*, volume 3447 of *Lecture Notes in Computer Science*, pages 143–154, Lausanne, Switzerland, 30 March - 1 April 2005. Springer.
[23] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.
[24] Yue Jia, Mark Harman, William B. Langdon, and Alexandru Marginean. Grow and serve: Growing Django citation services using SBSE. In Shin Yoo and Leandro Minku, editors, *SSBSE 2015 Challenge Track*, volume 9275 of *LNCS*, pages 269–275, Bergamo, Italy, 5-7 September 2015.
[25] Gabriela Ochoa and Nadarajen Veerapen. GECCO statistics and collaboration network. *SIGEVOlution newsletter of the ACM Special Interest Group on Genetic and Evolutionary Computation*, 10(3):3–6, September 2018.