

The scientometric portrait of Eugene Garfield through the free ResearchID service from the Web of Science Core Collection of 67 million master records and 1.3 billion references

Peter Jacso¹

Received: 10 December 2017 / Published online: 3 January 2018
© Akadémiai Kiadó, Budapest, Hungary 2018

Abstract Eugene Garfield came up with the idea of citation-based searching in the early 1950s, and followed it by releasing three unique databases, for the Sciences, Social Sciences, and Arts & Humanities, as well as a yearly Journal Citations Report for the Sciences and the Social Sciences. It took more than four decades before other database publishers started to add the cited references to their indexing/abstracting records. The Google Scholar database has been built on Garfield's original idea, and broadly idolized as a free database. Garfield's ultimate response -among others- was to release a database to allow the users to look up by author names or identification code. This article paints a scientometric portrait of Garfield as a tribute, demonstrating and commenting on how many times his oeuvre was cited by, in which sources (processed for the Citation Indexes), when and where, from which countries and institutions, in which format.

Keywords Eugene Garfield · Citation analysis · Citation-based searching · Scientometrics · Scholarly publishing · Research assessment · Freeware

Introduction

Some of my best moments as an author have been when I was invited by journal editors to write editorials, introductions, short essays on the 80th, 85th, and 90th birthdays of Eugene, or a short essay for the 50th birthday of his first paper in *Science* about the “Science Citation Index—A new dimension in indexing” www.garfield.library.upenn.edu/essays/v7p525y1984.pdf. It was not merely a new dimension, but a revolutionary one.

For his 80th birthday I was asked for a short piece for a special booklet published by ISI. I made it a short one, kind of a divertimento because I could deep-link it paragraph by paragraph to 38 of his scholarly and funny essays <http://www2.hawaii.edu/~jacso/extra/>

✉ Peter Jacso
Jacso@hawaii.edu

¹ Department of Information and Computer Sciences, University of Hawaii, Honolulu, HI, USA

80/. It is also a source of joy that I received this invitation for the Garfield Memorial special issue of *Scientometrics*, edited by the same 3 researchers who taught me bibliometrics and scientometrics more than 30 years ago.

He often came to Hawaii because he loved it, and spent much time here with one of his earlier wives who was of Hawaiian origin, and his old-time friends. He came also to conferences, and he was kind enough to come and do guest talks to our students and alumni and faculty. Often he just made a one night stopover on his way to and from Japan, Korea, and in Southeast Asia. We went with him for lunch, brunch or dinner, or just to swim on the splendid beaches, and “talking story” as the locals so well put it. It was always a pleasure to read him, to talk with him, and especially listen to him, even when he passionately criticized me for some comments in my articles, columns and conference presentations about WoS. In the following I present his scientometric profile, using the now free ResearcherID service, and the WoS Core Collection (WoS)—reasonably restricted to looking up authors and getting bibliographic records with bibliometric indicators.

The “grandfather of Google”

The bibliographic references were the functional equivalents of the contemporary links of the World Wide Web. The idea was developed by Garfield, 50 years before Google Scholar was launched, making the founders of Google consider Garfield “the grandfather of Google, and especially of Google Scholar”. I was not able to find an acknowledgement of and reference to Garfield from its developer. The concept and technology were first used in print format since the early 1950s in the *Genetics Index*, followed by volumes of *Citation Indexes for the Sciences, Social Sciences, and Arts and Humanities* fields in the next few years.

These were later, in the 1970s transformed, converted into a family of databases (digital collections), at a time when computers—in their physical breadth, not in their capacity—were a few thousand times larger in all dimensions than today’s tablets and palm-top computers. Garfield worked on the *Genetics Index* like Gaudi on the *Catedral de la Sagrada Familia* in Barcelona, building it brick by brick, manually without a break, and without access to twenty-first century construction and technical gizmos, let alone free brunch-lunch-and coffee breaks as the employees in the GooglePlex Diner and in the many cafes do on the Google campus, as a plus to the very pretty paycheck.

Garfield used an ENIAC computer, worked with punch card readers and their limitations in the early 1950s for describing on 80 positions the bibliographic records which must have had (at least part of) the name of the cited author, the title of the paper, its publication year, the classification code, and some additional data elements. He often may have spent 25 h a day in that computer center, but finally could have the first volume of the *Science Citation Index* printed. After lifting the very large and heavy reference book for packaging or just for a quick look-up he must have decided to make the content digital and sell it to libraries. He was a very qualified librarian, with an MLS degree from Columbia University to round out his degree in Chemistry, and doctoral title in Linguistics.

ResearcherID

The WoS databases have been very expensive resources to license, but there came a very important change. Garfield acquired the original ResearcherID software about 10 years ago to make the WoS system easier to use. The biggest change came when the ResearcherID software became free for any users merely requiring registration for a researcher identifier (RID), which allows browsing several indexes of metadata of nearly 67.5 million master records created from the WoS since 1900, display/print/download/save the result lists for free. This free service offers high quality, very well structured traditional bibliographic records and nearly 210 million references created, compiled curated by competent catalogers, human indexers, and bibliophile programmers from the 1950s under the direction and in the Renaissance spirit of Eugene Garfield. He passed away earlier this year at age 91, after a phenomenally productive and innovative life of more than 60 years of labor of love. He will be remembered for his innovations and inventions for helping researchers which should be called the Garfield Factors and Indicators (GARFI, if I may recommend it) as they have contributed as significantly to computerizing and smartening library automation, knowledge discovery, retrieval, management and dissemination of information as the MARC records of Madam Avram's Remarkable Contribution.

The WoS Core Collection (WoS) with the ResearcherID software could become the most appropriate, hopefully unbiased metric tool to create reproducible results, metric indicators, using metric tools and skills for measuring the authors' scholarly impact/influence through analyzing the citations they received. The 10 subcomponents of WoS include the 3 most traditional ones (Sciences, Social Sciences, Arts and Humanities), which make up 65.3, 11.6 and 6.3% of WoS. The relatively new Book Citation Index, the 3 Conference Proceeding Indexes, and the Emerging Sources Citation Index represent a modest 6%, and the two chemical databases account for the rest.

In this case I did this for the comprehensive set of Gene's publications from the mid-1950s to the end of 2017, focusing on how many of them were cited, by whom, from which countries, institutions, in which years, journals/conference proceedings, subject categories, and how many times. There will be additional citations even before the end of 2017 (there is already one from 2018 as this tribute is written). They will increase his citation counts, h-index, and citations/paper indicator for a long time, even if his publications count will remain the same in the impressive 1500-item range of articles, conference papers, reviews, editorials, and commentaries.

The size of the WoS database with the master records and the precious metadata elements of nearly 210 million cited references for the 1900–2017 time span, will also keep growing just as our joy when we shall see highly pertinent hits from citation searching on top of thesaurus-based queries, as "Gene told us so".

Eugene Garfield's scientometric portrait

The searches were done not by his name, but by his ResearcherID.¹ Gene has a relatively simple name for spelling correctly and consistently. He has a middle name, Elli, but he did not use it as middle name or as middle initial in business and in publications. He has three name formats/contents: Eugene Garfield; Garfield, Eugene; and Garfield, E. Each would

¹ <http://www.researcherid.com/rid/A-1009-2008>.

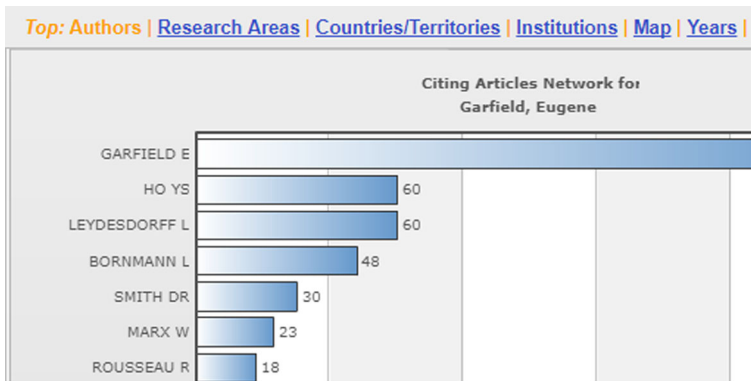


Fig. 1 Top authors citing Garfield (including himself)

retrieve several other names, even if the first name is not truncated, such as garfield, e*, which would retrieve many other researchers. Putting the name in between double quotes would not be sufficient, either. The easiest strategy could be to shrug our shoulder and check only the appropriate entries in the Author facet in the sidebar, when the first results are shown. The faceting is very well done in ResearcherID, just as in the subscription-based WoS version. This doesn't apply to journal names which cannot be directly searched through the search template of ResearcherID.

However, from the result lists a very appealing menu can be activated, showing a variety of Top-20s, such as the journals/proceedings which published articles, conference papers that cited the target persons' publications, the names of the citing authors, their co-authors, institutional affiliations and countries, the research disciplinary areas where she or he was cited from, and the number of citations received by years (some examples are shown in Figs. 1, 2 and discussed below). These graphics provide highly informative, at-a-glance bar charts with the counts and the names of the citing authors, journals/conference proceedings, monographic series, the WoS categories and the research areas, as well as the distribution of publications and citations across the years.

Not all the names may sound familiar to all the readers among the 20 shown in Fig. 1, even if they have been citing Eugene the most, because they may not signed up for ResearcherID, or they may represent the next generation of scientometricians. This may be the case with Lutz Bornmann who is not only exceptionally productive, but also very talented, after his debut paper some 10 years ago he became one of the very best scientometricians.

The institutional affiliations of the authors who have cited Garfield's works the most often are listed in Fig. 2. Enhanced in the name is also informative by itself. Clarivate Analytics is the new owner of Garfield's company and this could be indicated parenthetically as (Thomson ISI or Thomson Reuters), for the benefit of the new owner. The Organizations-Enhanced facet in addition to Organizations without a qualifier, helps the institutions to distinguish the system-level versus the campus level indicators, as is the case with the Indiana University and the Indiana University Bloomington in the Top 20 list below. A "see also" note would help in clarifying this. The check-boxes allow the ad-hoc combinations of, say, the many campus-departments of University of California.

The country/territory affiliations of the authors could also be highly useful, for example when doing a survey to find out who researched, published, coordinated and could qualify

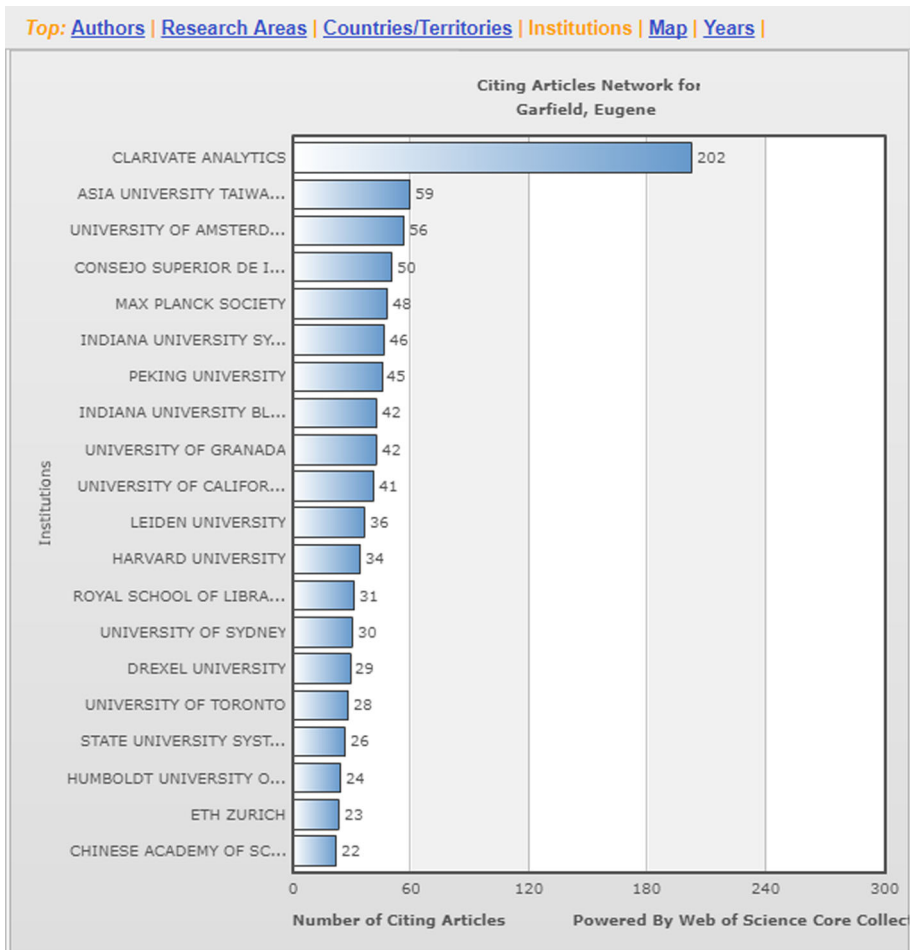


Fig. 2 TOP institutions’ researchers citing Garfield

the best to lead a reconstruction project in Puerto Rico once the very basic health-care, alimentation, nursery and elderly home services are finally restored. At the time of this writing some of the faceting features were available only very irregularly and unpredictably.

The variety of infographics for Garfield paint an impressive picture, a multifaceted and highly informative portrait about his publishing oeuvre, and the quality of his works and publications. It would be useful to add the h-index, g-index and e-index or the rank position of the citing journals, persons, institutions (especially of universities) or their percentile ranks.

The process is easy and the bar charts are instantly informative. It would be good to have an option where values could be displayed as percentages for better visibility and at-a-glance comparison. Mixed casing instead of all uppercasing (reminiscent of the limits of the punch card era) would make the labels more legible. Offering an option to have more

than 20 scrollable facet elements displayed, and showing the “soft and round” bar charts for two facets side by side would be welcome.

Software errors

There are two very important problems in the ResearcherID package, large scale of errors of commissions and errors of omissions. The set of the 1538 items consists of works that “our Garfield” authored, or co-authored, in—among others—the series and subseries of *Current Contents*, his popular, short, but scientifically enlightening and/or entertaining essays in addition to those highly academic interdisciplinary super-journals, as *Nature*, *Science*, *NEJM*, and subject oriented, discipline-focused ones (*Scientometrics*, *Journal of Information Science*, *Journal of Documentation*, *Journal of Chemical Documentation*, *CORTEX*, *Journal of Chemistry*, *Journal of Informetrics*), *The Library Quarterly*, *Information Processing and Management*, *International Journal of Epidemiology*, *International Microbiology*, and dozens of others.

Many academics who were (and still are) envious of his productivity, recognition, impact, bonheur, and courtmanship, look down at these essays, even if they themselves are the navel-gazing-philosophy types. Eugene has been the ultimate scientist in action, doing the talk and walking the walk, even when he was down financially, living in a then highly rundown part of lower Manhattan, driving a cab, and playing his sax when thinking about citation indexing. In addition, he published more than a thousand such essays. There is nothing wrong in combining these in assessing his scientific cultural, and educational heritage, humanity and generosity, but two problems must be corrected. One of the problems is that for 1127 publications of Eugene, the citation counts and indicators are wrongly calculated. It is triggered by the *Current Contents* journal of ISI. It came up in a rather routine step of creating a set for all of Garfield’s works in ResearcherID. The total number of papers for his name was 1556. With a small clean-up re-run it went down to 1538, the same number when searching by his ResearchID code. Out of these there were 1127 related to records for papers published in Garfield’s *Current Content* series (see Fig. 3).

ResearcherID reported a hit count of 1127 for this subset, and—in the LIS field—an unlikely high h-index of 138. It turned out that each item in *Current Contents* which was cited at least once is credited with the highest hit count in each year and these were than aggregated. This is (partially) illustrated in Fig. 4, which shows that each of the first 50 items were cited 170 times, from item #51 each papers were credited by 154 citations, and so on, and on, and on. I alerted the folks about the problem, who were very prompt, responsive and helpful in the earlier phase, but they could not provide—yet—a solution for it.

The other problem is that there are missing years in the cited years bar chart index of Garfield, as if his papers had not been cited at all in 1989, and from 1991 to 2000 (see Fig. 5). Gene would be rolling in his grave seeing this absurdity shown in Fig. 5. It should be top priority to fix this problem as it may undercut the credibility of the important, and excellent idea of celebrating the life and work of Eugene—among others—with releasing the otherwise very useful ResearcherID freeware.

Facet elements in ReseacherID

The unique and even today often incompletely and inconsistently implemented metadata elements such as the imprint, the parallel, former, successor titles are well designed in ReseacherID, so are the facet elements, which can be sorted alphabetically or by hit counts, although limited to 100 entries. These—along with colors, and good menu formats, with collapsible and expandable lists of author names, journal names, descriptor words,

Set	Results	Save History / Create Alert	Open Saved History
# 10	1,127 #7 AND #9 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=1970-2017</i>		
# 9	6,340 SO=(CURRENT CONTENTS OR CURRENT CONTENTS AGRICULTURE BIOLOGY ENVIRONMENTAL SCIENCES OR CURRENT CONTENTS ARTS HUMANITIES OR CURRENT CONTENTS CLINICAL MEDICINE OR CURRENT CONTENTS CLINICAL PRACTICE OR CURRENT CONTENTS CLINICAL PRACTICE OR CURRENT CONTENTS ENGINEERING TECHNOLOGY APPLIED SCIENCES OR CURRENT CONTENTS LIFE SCIENCES OR CURRENT CONTENTS PHYSICAL CHEMICAL EARTH SCIENCES OR CURRENT CONTENTS SOCIAL BEHAVIORAL SCIENCES) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years</i>		
# 8	6,340 SO=(CURRENT CONTENTS OR CURRENT CONTENTS AGRICULTURE BIOLOGY ENVIRONMENTAL SCIENCES OR CURRENT CONTENTS ARTS HUMANITIES OR CURRENT CONTENTS CLINICAL MEDICINE OR CURRENT CONTENTS CLINICAL PRACTICE OR CURRENT CONTENTS CLINICAL PRACTICE OR CURRENT CONTENTS ENGINEERING TECHNOLOGY APPLIED SCIENCES OR CURRENT CONTENTS LIFE SCIENCES OR CURRENT CONTENTS PHYSICAL CHEMICAL EARTH SCIENCES OR CURRENT CONTENTS SOCIAL BEHAVIORAL SCIENCES) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years</i>		
# 7	1,556 AU=(GARFIELD E) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years</i>		

Fig. 3 Query details of searching for Garfield’s papers published in Current Contents

Results: 1,127
(from Web of Science Core Collection)

You searched for: #7 AND #8
...More

Create Alert

Refine Results

Search within results for...

Publication Years

- 1985 (58)
- 1978 (55)
- 1979 (54)
- 1990 (54)
- 1983 (53)

more options / values... Refine

Refine Results

Search within results for...

Publication Years

- 1985 (58)
- 1978 (55)
- 1979 (54)
- 1990 (54)
- 1983 (53)

more options / values... Refine

Sort by: Times Cited -- highest to lowest

Page 1 of 113

Select Page 5K Save to Other File Formats Add to Marked List Create Citation Report Analyze Results

- HIGHLY CITED ARTICLES .35. BIOCHEMISTRY PAPERS PUBLISHED IN 1940S**
By: GARFIELD, E
CURRENT CONTENTS Issue: 8 Pages: 5-11 Published: 1977
Times Cited: 170 (from Web of Science Core Collection)
Usage Count
- HIGHLY CITED ARTICLES .33. ARTICLES FROM ITALIAN JOURNALS AND FROM ITALIAN LABORATORIES**
By: GARFIELD, E
CURRENT CONTENTS Issue: 6 Pages: 5-12 Published: 1977
Times Cited: 170 (from Web of Science Core Collection)
Usage Count
- NEW YEARS GREETINGS - AND OTHER CORRESPONDENCE - KEEP SPIRIT BRIGHT ALL YEAR LONG**
By: GARFIELD, E
CURRENT CONTENTS Issue: 5 Pages: 5-11 Published: 1977
Times Cited: 170 (from Web of Science Core Collection)
Usage Count
- ...
- VEGETARIAN ALTERNATIVE**
By: GARFIELD, E
CURRENT CONTENTS Issue: 11 Pages: 5-13 Published: 1977
Times Cited: 170 (from Web of Science Core Collection)
Usage Count
- THE 1982 NOBEL-PRIZE IN PHYSICS**
By: GARFIELD, E
CURRENT CONTENTS Issue: 50 Pages: 5-13 Published: 1983
Times Cited: 154 (from Web of Science Core Collection)
Usage Count
- ARTIFICIAL-INTELLIGENCE - USING COMPUTERS TO THINK ABOUT THINKING .1. REPRESENTING KNOWLEDGE**
By: GARFIELD, E
CURRENT CONTENTS Issue: 49 Pages: 5-13 Published: 1983
Times Cited: 154 (from Web of Science Core Collection)
Usage Count

Fig. 4 Aggregated citation counts of the articles in the same year in Current Contents

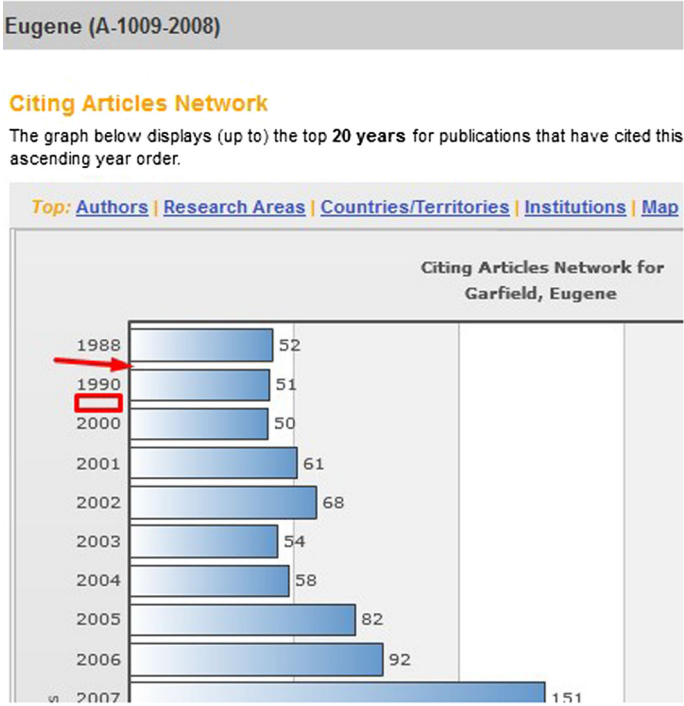


Fig. 5 According to this graph citations received by any of Garfield publications in 1989 and between 1991 and 1999—Really?

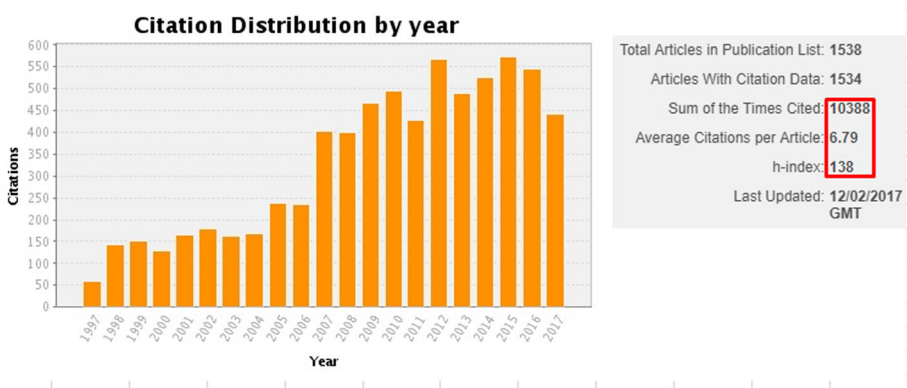


Fig. 6 Garfield’s citations count and h-index—according to ResearcherID

document types, subject area names much facilitate the browsing and searching process for finding scholarly articles, conference papers, books, book chapters.

Very importantly, the above metadata elements show their frequency in the set retrieved by the simple query of the searchers, alerting them to narrow the search, when it becomes clear, that there are way too many items for the search terms, several authors with the same last name, first name, but different middle initial(s), journal names with a variety of

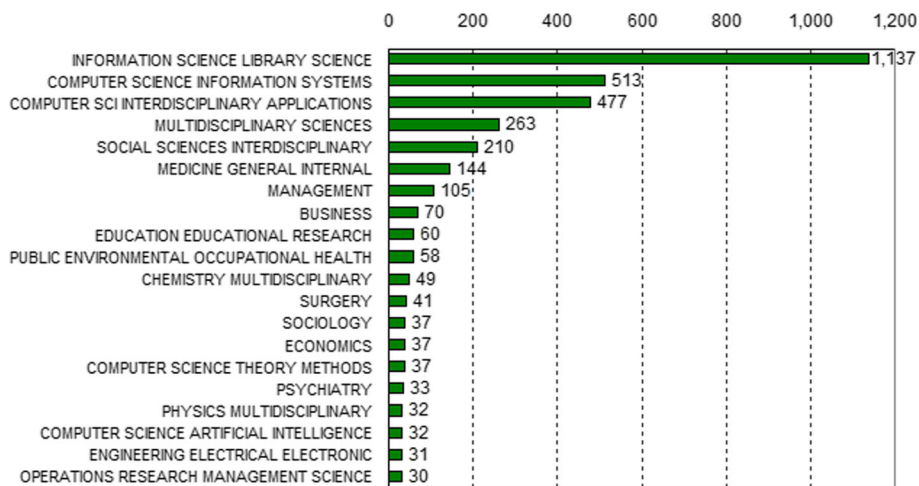


Fig. 7 Research areas/categories most citing Garfield

abbreviations, punctuations, spacing, and handling of accented characters, as is the case with one of the editors and this author, often make them losing citation credit points, especially if the variations are not adjacent in the list.

Garfield’s extraordinary inventions, hardware and software designs and developments deserve to say farewell to him by painting his citation portrait (see Fig. 6), showing his research productivity (the quantity), and acknowledgement via citations received in scholarly and trade publications, from well-known researchers working in universities, research institutions around the world, in several languages, and different publications formats. My calculations suggest that Gene’s h-index is 45, which is among the highest in the LIS field.

There are very well-chosen facets (sources, document types, WoS subject categories, Research areas, as shown in Fig. 7) to allow the users to have search results customized as the hand-tailored shirts/blouses, and bespoke dresses/suites are for their preferences at Savile Row. (Garfield loved splendid suites, but felt very comfortable in his very Hawaiian shirts, even at a meeting in the Hungarian Academy of Sciences as I heard).

These inform the searchers how many items there are at the current search phase by year, topic, language, document type. This could be improved only by offering the option to show (perhaps optionally) the distribution of sets in percentage, because the 6–7 digit numbers make some of the facet lists too crowded (see Fig. 8). It would also help to display the hit counts percentages without decimals as thin horizontal bar-lines.

The faceted, field-specific index entries show for each entry the number of records which have that term, such as 454,660 for the subject category of Information Science and Library Science as shown in Fig. 9. While it could be better to use a more compact term, such as Information and Library Science, the variety of indexes are excellent, simple to navigate and choose several related terms, such as Computer Science Information Systems, and to retain the original format with &, /, -, symbols. The index entries can be sorted by frequency or alphabetically.

Fig. 8 Options for smartly refining results subsets based on a variety of metadata along with volume information for each

Publication Years ▼

- 2016 (2,889,470)
- 2015 (2,792,515)
- 2014 (2,540,912)
- 2013 (2,450,585)
- 2012 (2,365,614)

[more options / values...](#)

Refine

Web of Science Categories ▼

- ENGINEERING ELECTRICAL ELECTRONIC (3,115,607)
- BIOCHEMISTRY MOLECULAR BIOLOGY (2,592,259)
- MEDICINE GENERAL INTERNAL (2,417,039)
- CHEMISTRY MULTIDISCIPLINARY (2,353,111)
- MATERIALS SCIENCE MULTIDISCIPLINARY (1,943,577)

[more options / values...](#)

Refine

Document Types ▼

- ARTICLE (40,187,966)
- PROCEEDINGS PAPER (7,661,124)
- MEETING ABSTRACT (7,349,867)
- BOOK REVIEW (4,117,645)
- EDITORIAL MATERIAL (2,701,700)

[more options / values...](#)

Concluding remarks

The ResearcherID service is very good except for the software glitches. It helps students, teaching and research faculty to discover and learn about highly relevant topics free of charge in the most expensive databases, with an intuitive, and attractive, menu-driven software (presuming the fixing of its two significant software deficiencies). Librarians, educators will be in a much better situation to demonstrate and train students how to take advantage of following the route of cited references which lead from one good article to several and even better: current conference papers, or go backward to find to trace the origin and development of new, better, cheaper alternatives for solving problems in every subject areas. They will realize that “garfielding” a topic can be more effective than just googling it. Of course, finding and reading the most cited research articles by the best journals, requires access to the full text. Not surprisingly, Gene digitized practically all of his writing, and made them available online—for free. Once the author names, journal and

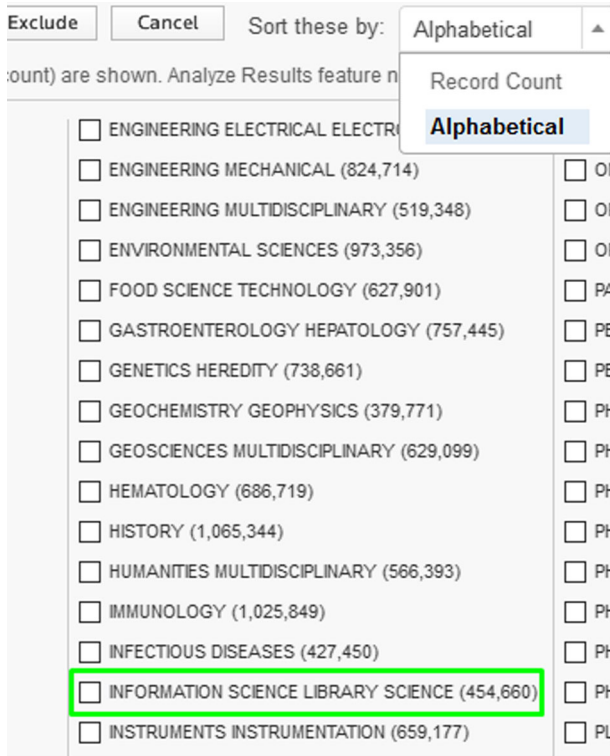


Fig. 9 Field-specific index entries of the analyze results software feature

article titles identify the most promising primary documents, there is still one more step of retrieving the information, which are likely to be better by their high citations/year relative indicators, than the single citation counts on their own.

Linking is everything, and now that the ORCID database will add more than 3 million bibliographic/bibliometric records to ResearcherID will be massively upgraded with metadata elements, such as the eISSNs, the published URLs are enhancing the digital networks, and increasing the chances of instant gratification.

Garfield was not only well-born (Eugenius), but also very well-learned, very well-lived. He was a discoverer, he practiced glasnost and preached perestroika well before Reagan and Gorbachov, shook hands repeatedly in Reykjavik, traveled North and South, made business and worked with researchers, East and West, including Novosibirsk, not exactly a tourist destination. He was donatious, which he told me was not an English word, but he instantly understood and liked it. And I liked him very much, especially after Michael Koenig, my longest lasting mentor at Columbia University introduced me to him, and Carol Tenopir who offered me to do guest sessions about Citation-Based Searching, in her classic “Online Searching” course.

I hope Eugene, who did so much and so well, finally will rest in peace.

Acknowledgements I much appreciate the support and advise I received from Professor Glänzel in writing this eulogy.