You Can Look It Up— Or Maybe Not

Stan Kelly-Bootle, Author

Many are said to have said, “If I can’t take it with me, I’m not going!” I’ve just said it, but that hardly counts. Who, we demand, said or wrote it first? It’s what I call (and claim first rights on) a FUQ (frequently unanswerable question, pronounced *fook* to avoid ambiguity and altercation). Yogi Berra’s famous advice was “You can look it up,” meaning, in fact, “Take my word on this.” He knew quite well that few had the means or patience to wade through the records. Nowadays, of course, as we quip in Unix, it’s easier done than sed.¹ The portmanteau *wep* for grepping the Web, now realized and refined in countless search engines, lets us take up Yogi’s challenge at face value with a few simplistic keystrokes and mouse-clicks. Yet, as I aim to indicate, life—at least the life of serious scholarship—is not a bowl of Googles.

Seeking earliest attributions has serious applications for etymologists and sociolinguists. Related questions of establishing priorities also loom large, litigable, and expensive in patents and intellectual property disputes. Indeed, it has emerged as part of a newish science called citationology, of which more anon. Take the proper noun *Google* and its derived parts of speech: verbal (“As I was a-Googling”); adjectival (“Google morality”); improper noun (“There’s nothing like a nice Google.”). You may think, with good reason, that the origins of *Google* have something to do with the incredibly large number googol (the 10¹⁰⁰ spurious matches planted by advertisers), or with the act of goggling (the bulging or squinting [Middle English] eyes as the spurious matches scroll by). My own preferred etymology is the googly, a sneakily-bowled cricket ball that spins in the opposite direction to that expected by the batter.² All of which hints at the perceived dangers of naively scanning the Web. Even with enlightened Boolean modifiers, literal `strcmp()`-type char-string matches are, indeed, too damned literal. As Dr. Walter Martin (the original Bible Answer Man) used to say, “Text without context is pretext.” And context cannot yet be conveniently, decently parsed and automated.

Further, the commercial search engines (show no surprise) have commercial agendas, meta-matches, and sponsored, hyped hyperlinks. These ploys can be overt and reasonable (as someone said, “No free lunch!”); others

Chasing citations

THROUGH ENDLESS,
MISLABELED NODES

are hidden and insidious. “Galileo” may take you to Amazon’s books on the guy (sort of acceptable) or to some dubious agent offer-

ing cheap flights to Pisa (sort of unacceptable, especially if it’s the first match displayed). Browsing, formerly the idle nibbling of grass or the leisurely reading of books, is now the frantic chaining through endless, mislabeled nodes.

Citations have been tracked and dissected for centuries by dusty, manual dictionary-makers—recall Samuel Johnson’s self-definition of lexicographer as a “harmless drudge”? The classic case is the agonizing (1879-1928) compilation of the first *OED* (*Oxford English Dictionary*), started by Sir James Murray (1837-1915), who strove to supply the best-available and first-known citations for each headword as an essential supplement to the proposed etymologies and shifting semantics. Murray’s way-out outsourced army of readers, flooding the *OED* HQ with millions of hand-written 4 x 6 “dictionary cards,” included some truly paranoid word-tracers. There was the convicted “Lambeth murderer,” American-born Dr. William Chester Minor, who supplied 20 years of wondrous citations from his library-cell in the Broadmoor Asylum for the Criminally Insane. And you must love Herbert Coleridge, another of Murray’s fanatical helpers, “whose last words, at the age of thirty-one, were ‘I must begin Sanskrit tomorrow.’”³

The citation chase involves years of bleary-eyed rummaging through rare manuscripts, books of all kinds—banned and available—newspapers, glyphs, tombstones, ostracons, and buried genizoth (Hebrew plural of *genizah*, a repository of discarded, damaged holy books). The introduction of computers and the gradual digitization of many of these sources clearly increased accuracy and reduced the drudgery, yet the nirvana of “beating” the *OED* by predating any of its “first” citations is still more likely to be attained by old-fashioned grunt work than by electronic searches. Anyone with minimal skills and a modem can scan the available databases. But the news-making predatings are those made from locating and

Continued on page 54

Continued from page 56

painstakingly reading the very rare sources not yet online. Perhaps the record is the pushing back overnight, in one fell swoop, of 76 *OED* botanical citations when in 1976 French lexicographer Claude Boisson discovered a 17th-century Spanish herbal. I'll drink to that!⁴

So, yes, there are professional citationologists who earn their daily-donnish, inchworm bread seeking authorial precedences. When asked by their kids, "Mummy/Daddy, what do you do for a living?" they might reply, "I believe that question was first posed by F. Josephus Jr., circa 67 CE, although there are good scholars who place it much earlier in the Shang dynasty...why aren't you in bed?" The JobSpec template is, informally: Who *first* said <T>, how, when, where, and (holiday bonus), why?

Finding the first usage of *citationology* requires, somewhat recursively, its own methods. Appending the Grecian formulaic *-ology* to indicate the scientific study of the suffixed stem is not in itself a shattering neological achievement. But establishing an accurate space-time-stamp for the emergence of a particular *-ology* can be useful to science historians. I seem to have been the first to use *citationeering* (*Unix Review*, March 1989—but submitted January 1989!) as a rather derogatory dig at dilettante citation hunters (I have since dubbed them, unfairly, Wikipedophiles). The first appearance of *citationology* that I can find is an authoritative claim by Eugene Garfield (publisher of *The Scientist* magazine): "I offer the term *citationology* as the theory and practice of citation, including its derivative disciplines, citation analysis and bibliometrics," submitted April 9, 1998.⁵ It's rather exciting to find the expert practitioner naming an extension to his own domain, and with strong evidence ("I offer the term...") of his priority. Nevertheless, priority always remains an open case.

Search engines such as Web of Science, Google Scholar, and Scopus permit endless refinements in the citation game. Careers, tenures, and funding can depend on not only establishing publishing priorities but also comparing "citedness scores" (who has been quoting or referencing your papers) and deducing an author's or paper's "impact." Note, of course, that "I love Chomsky" and "I hate Chomsky" both register a match for "Chomsky".⁶

What distinguishes the citationologist from the citationeer is not only a mature, finely honed skepticism about the commercial search-engine distortions mentioned earlier, but also a balanced cynicism about the uneven quality of Web *content*. We must be equally suspicious of all our sources, of course, whether in bound-paper books and journals, or on disks and screens. But it

is often more difficult to authenticate the volatile media even though most of our information actually originates and is exchanged electronically. Our Web sites and mailboxes include peer-reviewed repositories of mankind's accumulated knowledge intermingled with suppositories of willful or accidental disinformation. As one tech writer's prologue warned: "Believe only the true sections. Ignore the rest."

AUTHOR! AUTHOR!

Let's not be too grumpy. There's a growing citation and pop-grammarians industry to amuse the lay wordsmithy. I've counted six books with *Lost for Words* in their titles. Anthologies of "familiar quotations," pioneered by John Bartlett, abide in print and online. Our very own ACM Web site maintains a list of computer-related sayings and witticisms. Kevin G. Barkes keeps an unusually varied collection at <http://www.goodquotations.com>. In the BBC Radio quiz "Quote/Unquote," the strategy of contestants who cannot immediately identify a given quotation seems to be a plausible guess based on genre. The Bible, Shakespeare, or Churchill are statistically promising for somber and majestic pronouncements, while for modern

As one tech writer's prologue warned,
"Believe only the true sections, ignore the rest."

wisecracks, try Oscar Wilde, Mark Twain, or Will Rogers, with a side bet on Dorothy Parker. The wildcard, catchall "no match" is the ageless polymath Anon, accidentally tenured as Professor Anon. Thereby hangs a warning tale of search-engine hazards.

The "Professor Anon," whom I thought I had invented as the obvious putative source of all otherwise unasccribed wisdom, turns up more than 70 matches when Googled. Ignoring the blogger who signs off as Anonymous and is then addressed sarcastically as "Professor Anonymous," and overlooking a Dr A. N. Other and Prof Anon at the (possibly?) mock University of Anon, Scotland, we do finally find a real live don. He is referenced as Professor Anon Monshouwer, a leading educologist (you can look that up—it's not quite an educationalist) at the University of Nijmegen, The Netherlands. Funny, methinks. Further research solves the mystery: He is, in fact, Anton Mon-

shower. A spelling error in a Web paper on the history of educology spawns other references to my supposed phantom expert. You will all no doubt have your own pet encounters with the Web's high noise-signal ratio.

Ideally, we should examine all sources: glyph, manuscript, print, and audiovisual in every available language (not to mention the Latin motto tattooed on Angelina Jolie's torso).⁷ Matching sounds and pictures is, we are promised, "on the way!" We must also allow for semantic variations; seek the gist rather than matching exact phrases. Aye, there's a major rub.

My lead-in quotation, for example, was assigned by Charlie Zimmerman to comedian Jack Benny, who actually quipped, "If I can't take it with me, I refuse to go," which is a close enough match for humans even if it presents a challenge to automated search engines. But, pause, which, if any, of Jack Benny's many scriptwriters might have prior claim? Likewise, who penned Fred Allen's less-famous response, "If you could take it with you, it would melt!" Naïve grepping for literal string matches, however, would not fully resolve or explicate the Jack Benny quote. What is also needed is the wider context, starting, say, with the earlier post-Depression catchphrase, "You can't take it with you," popularized by George S. Kaufman's 1937 novel and Frank Capra's movie version. This is complex context, spelling hedonism to some: Live it up while you can, with a hint that you may not be "going" anywhere. Compare the cynical Union hymn, "Work and pray, live on hay; you'll get pie in the sky when you die."

Others, though, can read it as a warning against consumerism and as supporting Bill Gates's welcomed philanthropy, which derives from the aforementioned consumerism. I can live with this paradox, having visited the wonderful new (2001) William Gates Computer Laboratory, Cambridge, that replaces my 1950s EDSAC math lab haunts on the old Cavendish site. The Christian message is that such good deeds are rewarded: "You can't take it with you, but you can mail it ahead."

One can also find personal-computing interpretations of the Jack Benny gag. There are echoes of Adam Osbourne's first "portable" PC: "You *can* take it with you!" (well, with the help of Hernia, the Goddess of Weightlifters). With the advent of the mobile-phone-PDA-MP3-camera-Web-terminal, we move on to the portability of the credit card: "Don't leave home without it." We know who holds the copyright, but who said it first? ☺

REFERENCES

1. When I first used this pun (*Unix Review*, September 1985), it was edited as "easier said than done." My

jokey "IT laxicon" also suffers textual harassment, restored to "lexicon" as recently as my July/August 2006 Curmudgeon column in *ACM Queue*!

2. Like Yogi's baseball, with its curves, knuckle, and spitballs, cricket has a complex taxonomy of wicked deliveries (off-spin, leg-spin, yorkers, doozies) that has fascinated even pure mathematicians such as G. H. Hardy.
3. Winchester, S. 1998. *The Surgeon of Crowthorne*, London, Viking. The U.S. version was more enticingly titled *The Professor and the Madman*, HarperCollins, 1998. The *OED* still enlists help from outside readers. Details of the North American Reading Program: <http://www.rice.edu/oed/readers.html>.
4. Boisson, C. Earlier quotations for Amerindian loanwords in English. *OUP: International Journal of Lexicography* 1(4). The star in Boisson's list is *totora*, a South American plant that will be more familiar to you as the *Typha domingensis*. The *OED*'s 1936 citation for *totora* gets rolled back all the way to 1604!
5. Garfield, E. 1998. Random thoughts on citationology, its theory and practice. *Scientometrics* 43(1).
6. The infamous "selected quote" in show biz is relevant here, and worth a mention. The billboard says, "Incredible!—NY Times"; the critic has written, "I find it incredible that any sane person would pay money to see this show." And "Don't Miss It!" was extracted from "If the show's running late and you have a train to catch, don't miss it."
7. Subject to many palimpsestuous modifications: *Quod me nutrit me destruit* (What nourishes me also destroys).

LOVE IT, HATE IT? LET US KNOW

feedback@acmqueue.com or www.acmqueue.com/forums

STAN KELLY-BOOTLE (<http://www.feniks.com/skb/>; <http://www.sarcheck.com>), born in Liverpool, England, read pure mathematics at Cambridge in the 1950s before tackling the impurities of computer science on the pioneering EDSAC I. His many books include *The Devil's DP Dictionary* (McGraw-Hill, 1981), *Understanding Unix* (Sybex, 1994), and the recent e-book *Computer Language—The Stan Kelly-Bootle Reader* (<http://tinyurl.com/ab68>). *Software Development Magazine* has named him as the first recipient of the new annual Stan Kelly-Bootle ElecTech Award for his "lifetime achievements in technology and letters." Neither Nobel nor Turing achieved such prized eponymous recognition. Under his nom-de-folk, Stan Kelly, he has enjoyed a parallel career as a singer and songwriter.

© 2006 ACM 1542-7730/06/1000 \$5.00