

September 23, 2007

King Algorithm

An Oracle for Our Time, Part Man, Part Machine

By GEORGE JOHNSON

IN the 12th century A.D., when the Arabic treatise “On the Hindu Art of Reckoning” was translated into Latin, the modern decimal system was bestowed on the Western world — an advance that can best be appreciated by trying to do long division with Roman numerals. The name of the author, the Baghdad scholar Muhammad ibn Musa al-Khwarizmi, was Latinized as Algoritmi, which mutated somehow into algorismus and, in English, algorithm — meaning nothing more than a recipe for solving problems step by step.

It was the Internet that stripped the word of its innocence. Algorithms, as closely guarded as state secrets, buy and sell stocks and mortgage-backed securities, sometimes with a dispassionate zeal that crashes markets. Algorithms promise to find the news that fits you, and even your perfect mate. You can’t visit Amazon.com without being confronted with a list of books and other products that the Great Algoritmi recommends.

Its intuitions, of course, are just calculations — given enough time they could be carried out with stones. But when so much data is processed so rapidly, the effect is oracular and almost opaque. Even with a peek at the cybernetic trade secrets, you probably couldn’t unwind the computations. As you sit with your eHarmony spouse watching the movies Netflix prescribes, you might as well be an avatar in Second Life. You have been absorbed into the operating system.

Last week, when executives at MySpace told of new algorithms that will mine the information on users’ personal pages and summon targeted ads, the news hardly caused a stir. The idea of automating what used to be called judgment has gone from radical to commonplace.

What is spreading through the Web is not exactly artificial intelligence. For all the research that has gone into cognitive and computer science, the brain’s most formidable algorithms — those used to recognize images or sounds or understand language — have eluded simulation. The alternative has been to incorporate people, with their special skills, as components of the Net.

Go to Google Image Labeler ([images.google.com/imagelabeler](http://images.google.com/imagelabeler)) and you are randomly matched with another bored Web surfer — in Korea, maybe, or Omaha — who has agreed

to play a game. Google shows you both a series of pictures peeled from the Web — the sun setting over the ocean or a comet streaking through space — and you earn points by typing as many descriptive words as you can. The results are stored and analyzed, and through this human-machine symbiosis, Google's image-searching algorithms are incrementally refined.

The project is still experimental. But the concept is not so different from what happens routinely during a Google search. The network of computers answering your query pays attention to which results you choose to read. You're gathering data from the network while the network is gathering data about you. The result is a statistical accretion of what people — those beings who clack away at the keys — are looking for, a rough sense of what their language means.

In the 1950s William Ross Ashby, a British psychiatrist and cyberneticist, anticipated something like this merger when he wrote about intelligence amplification — human thinking leveraged by machines. But it is both kinds of intelligence, biological and electronic, that are being amplified. Unlike the grinning cyborgs envisioned by science fiction, the splicing is not between hardware and wetware but between software running on two different platforms.

Several years ago, SETI@home became a vehicle for computer owners to donate their spare processing cycles for the intense number-crunching needed to sift radio-telescope data for signs of extraterrestrial life. Now a site run by Amazon.com, the Mechanical Turk ([www.mturk.com](http://www.mturk.com)), asks you to lend your brain. Named for an 18th-century chess-playing automaton that turned out to have a human hidden inside, the Mechanical Turk offers volunteers a chance to search for the missing aviator Steve Fossett by examining satellite photos. Or you can earn a few pennies at a time by performing other chores that flummox computers: categorizing Web sites (“sexually explicit,” “arts and entertainment,” “automotive”), identifying objects in video frames, summarizing or paraphrasing snippets of text, transcribing audio recordings — specialties at which neural algorithms excel.

(Not all of these Human Intelligence Tasks, or HITs, as Amazon calls them, involve serving as a chip in some entrepreneur's machine. Hoping to draw more traffic to their sites, bloggers are using the Mechanical Turk to solicit comments for their online postings. In some cases you get precisely 2 cents for your opinion.)

In his 1950 paper “Computing Machinery and Intelligence,” Alan Turing foresaw a day when it would be hard to tell the difference between the responses of a computer and a human being. What he may not have envisioned is how thoroughly the boundary would blur.

How do you categorize Wikipedia, a constantly buzzing mechanism with replaceable human parts? Submit an article or change one and a swarm of warm- and sometimes hot-blooded proofreading routines go to work making corrections and corrections to the corrections.

Or maybe the mercurial encyclopedia is more like an organism with an immune system of human leukocytes guarding its integrity. (Biology too is algorithmic, beginning with the genetic code.) When the objectivity of Wikipedia was threatened by tweaking from special interests — a kind of autoimmune disease — another level of protection evolved: a Web site called WikiScanner that reports the Internet address of the offender. Someone at PepsiCo, for example, removed references about the health effects of its flagship soft drink. With enough computing power the monitoring could be semiautomated — scanning the database constantly and flagging suspicious edits for humans to inspect.

No one but a utopian would have predicted how readily people will work for free. We're cheaper than hardware — a good thing considering how hard we are to duplicate.