Google's Free Trash Delivery Service

Nous sommes les suceuses de pus, les mouches,
nous partagerons tout avec toi, … nous t'escorterons jusqu'à la tombe….[1]

*H*ave you ever heard of Ghostscript or Ghostview? These are software programs related to the placement of hidden print commands in electronic documents. The commands are visible to web search engines, but not to the naked eye. This hidden text can be used as a means to deliver electronic garbage to another's doorstep with no obvious connection between the receiver of the garbage and its sender.

According to a recent article[2] in the online edition of the New York Times the Google search engine accounts for "55 percent of all searches on the World Wide Web". I do not know how this figure was derived, but it suggests that Google is a good place to start when seeking to learn how others think about another's website, or when simply looking for a website whose internet address you do not know.

Say, for example, that you have heard that EARTH is hosting a project on English language educational reform in Hong Kong. Assume that you are interested in this topic and want to learn more about it. If you did not know EARTH's internet address or did not have the email address of EARTH's creator, you might begin your search by typing in keywords closely associated with EARTH and English language reform. Simply typing in the keyword EARTH would likely result in millions of hits -- far too many for a very rewarding initial search. Other keywords or phrases such as "Hong Kong", and "language reform" entered along with EARTH's name, would probably be your best bet.

Now let us assume that you found EARTH's website, looked around, and were further interested, but wanted to see what other people had to say about EARTH, its projects, and its people before exploring further. So, you type in the word HKLNA, the name of EARTH's principle research project, and find a long list of garbage links after the website's name. What would you be inclined to think?

When I first posted the HKLNA-Project on EARTH's website a simple Google search using the keyword HKLNA resulted in a full page of hits dedicated to nothing but the HKLNA-Project. No other listings were visible. So clean was the ledger that I remember a friend jokingly remarking that my company's project had "poll position" for the name HKLNA. Indeed, it was a good feeling to discover that my company's principal project could be so easily found on the world wide web. After several weeks had passed and many new additions to the website had been posted, I made another Google search with the keyword HKLNA, but found something very puzzling. Below the HKLNA-Project's entry and several other pages related to the project I found a long list of other entries whose names appeared as unintelligible ASCII script. My natural inclination was to test with a few different character formats and see, if the documents were not written in a language that my web browser could capture.

[1] Jean Paul Sartre. 1976. Les mouches, Act 2. Choeur des Erinnyes. Collection Folio. SaintAmand: L'imprimerie Bussière
[2] Geoffrey Nunberg. 2003 (May 18). As Google Goes, So Goes the Nation. New York Times. [online document].

No change. The unintelligible, garbled ASCII script remained unintellible. Upon closer investigation of the pages' content I discovered warning notices such as "Page not found", PDF documents with no visible reference to the HKLNA-Project, pseudo links to nonexistant websites, and Google queries asking me, if I were not looking for a different webpage. It was all very disturbing. What had once been a unique list of direct links to my company's research, was now a garbage dump for Google's webcrawler. Further web searches related to my company, my research, and me led to very much the same conclusion. My own and my company's web presence was being sabotaged by *electronic garbage flys*!

If my company were not an e-business, I could simply sweep away the trash with a broom every morning, as shopkeepers the world over sweep away fallen leaves, road dirt, cigarette butts, gum wrappers, soda pop cans, plastic bags, and whatever other litter is wont to accumulate at anyone's doorstep in a cosmopolitan commercial setting. Unfortunately this was not the case; I had no control over this trash. Google was picking it up during its regularly scheduled webcrawls and delivering it to my doorstep. Obviously it was not Google that was placing my name, my company's name, and my company's project names on someone else's garbage; rather, it was others, who knew that Google would deliver it. I am hardly a computer hack, but more than five years of experience on the net had taught me enough to figure out how it could work.

Webpages listed by Google searches must either be frequently visited, or less seldom visited by people from heavily used ISP addresses. This meant that the probable sources of the garbage were people affiliated with large institutions and people who were adept at computer programming and had created algorithms generating large numbers of repeated visit to the garbage. My subsequent investigation led me to several large research institutions in Israel, Argentina, and New Zealand among other likely such places whose identity I was unable to pinpoint. In all cases but one, identifying the institutions was insufficient to determine the actual signatories of the garbage. The one potential *garbage fly* that I was able to identify was an associate professor and software design engineer named James Noble at Victoria University of Wellington in New Zealand. So, I wrote him an email, told him about the document containing hidden script that was cluttering Google's listing of my company's research project, and asked him what he knew about it. He never responded. Several days later the garbage disappeared.

In contrast most of the other garbage is well beyond my humble technical reach, and I have little choice but to live with it. Who these people are and why they have selected my websites as a place to dump their garbage, one can only wonder. Of course, surely that must be the idea. Afterall, who wants to visit a website that is associated with so much garbage.

EARTH prides itself in being frank for the sake of most. Others pride themselves in hurting others without detection. Take a careful look at EARTH's website and see who its potential enemys are. After that, you might just better understand why EARTH will continue to write what it does.

R. A. Stegemann
May 31, 2003
word count: 1108