Seek and ye shall find

I promised I would be more upbeat this month so apart from a minor go at British Gas, (nothing personal chaps apart from whoever designed your telephone meter entry system), I thought a word or two on the humble PC and its increasingly important role in searching. This summer I have been working on a 'discovery engine', a bibliographic research project developing new algorithms based on a technology known as 'Chance Discovery' first developed in Japan, with the object of discovering unlikely but significant events amongst very large unstructured datasets, for example the web. Given the size of the data involved, this may seem an insurmountable task, however when PCs aren't spending most of their available resource driving enormous clumsy interfaces, it is positively amazing how much raw data processing they are capable of achieving.

Such specialised searches are an increasingly important part of modern day living. It might be a best deal, a lost relative, business information on a company's prospects or it may have some academic role. Search engines on the web tend to be strongly guided by the underlying links in and links out, using them to prune the amount of search and to order the results. This works very well for certain kinds of search although it does rather favour those sites which know how to work the system or who pay to be elevated in the search. If you are looking for a needle in a haystack, even a very bright one, you've generally had it. That's where discovery engines come in. They are extraordinarily versatile. Let's suppose you are a devotee of Indiana Jones and would really like to trace all phrases linked in some significant way to use of the “Ark of the Covenant”. Then off you go to the excellent Project Gutenberg, download the text of the King James Bible, and feed it into a discovery engine. Such is the power of a modern PC that just seven seconds later, you have a nice picture of all significant related phrases. I'm currently using it to analyse computer Common Vulnerability records. The main Mitre database contains 450,000 lines, (about 5 times bigger than the King James Bible). Even that took only a few minutes. Wonderful. More on www.leshatton.org.

This month's featured human interface most likely to drive you to drink is courtesy of British Gas. My recent gas bill was based on the difference of two estimates so British Gas urgently request that I use their automated telephone entry keypad system to update their records. Oh goody. First you enter your customer reference number. This it does not repeat but I was pretty sure I entered it correctly. Then it asked for my reading. I entered it and this time it repeated it correctly, went away for a moment or two and then said “Please enter again, this does not match our records.” Flipping into automatic software test mode, I duly entered exactly the same number, had it read back to me again correctly but this time it accepted it with grateful thanks. Ho-hum. I presume this feature also allows me to update anybody else's record to any number I can think of but please don't try this at home folks.

I don't look for this stuff you know, it kind of finds me.

l.hatton@kingston.ac.uk