

# Search strategies the answer

SIR — Koreaki Ito (*Nature* 382, 666; 1996) recommends the abandonment of British in favour/favor of US spellings in order to make more efficient the automated searching of on-line databases. Dangerous waters indeed!

Some years ago, as scientific editor of *The Canadian Entomologist*, one of the oldest scientific journals in North America, I made a similar suggestion. Ambiguously and awkwardly stuck here between the Empire and the rebel Colony, we Canadians had compromised in that oh-so-typical Canadian fashion — the journal permits either spelling convention. That not only made for a lot of work for the copy editor, it was also an insurmountable challenge for many authors, who illogically used both spelling conventions in single manuscripts. However, my suggestion to adopt a single (American) spelling convention in order to reduce this confusion was met with howls and screams of protest. I was called some pretty nasty names and I beat a hasty retreat. Evidently, even scientists cannot be counted upon to treat the mechanics and symbols of language objectively — and I would predict that UK editors who might recommend a similar change would receive pretty much the vituperation I endured.

My personal databases use a single spelling convention — American. But the solution, surely, for the public databases, is simply more intelligent programming. There aren't a lot of words that have dual spellings and it would be a simple matter to have a look-up table to ensure that a search for 'behavior\*', for example, would also automatically search for 'behaviour\*'. Such a scheme, intelligently constructed, could probably automatically find up to 90% of cases of alternative spellings.

**Stephen M. Smith**

Department of Biology,  
University of Waterloo,  
Waterloo, Ontario, Canada N2L 3G1  
e-mail: bugman@sciborg.uwaterloo.ca

SIR — Some weeks ago I searched for information about Münchhausen's syndrome in the Medline Database 1991–95. The entry 'Münchhausen' is not possible because the system does not accept the German letter 'ü'. The correct transliteration to 'Muenchhausen' yielded no results, but the version 'Munchhausen' led to 26 entries. After omitting the second 'h', however, I found 283 entries under 'Munchausen'. But in this form the name is badly mutilated for those who speak German and are familiar with Baron von Münchhausen's name and tales. I cannot understand why not even personal names are spelt correctly in such a database.

This is not the only example. The name of Gottfried Ritter von Rittershain (paediatrician 1820–83) is cruelly shortened in Ritter's syndrome, and that of Georges E. A. B.

Gilles de la Tourette (internist, 1857–1904) in Tourette syndrome, or sometimes Gilles-de-la-Tourette syndrome. Moreover, I found the name of Kurt Gödel (mathematician, 1906–78) shown as 'Godel' in *The Lancet* (346, 1172; 1995).

Such discrepancies are also found in scientific names. The name of the genus *Necrophorus* (burying beetles) was first published as *Nicrophorus* because of a typographical error. Although the author, Fabricius, corrected the error as early as 1801, both forms appear in the *Biological Abstracts* database.

It is unlikely that authors and editors of English-language journals will agree to uniform spellings. The best solution is a more intelligent searching system for databases that will offer alternative spellings where they exist.

**Achim Th. Schäfer**

Institut für Rechtsmedizin,  
52057 Aachen, Germany

SIR — While I have the greatest sympathy for Koreaki Ito when faced with the vagaries of spelling in the English language, I feel compelled to comment on the techniques used in his searches.

The method used was a textword search in an individual field. In general, this is the most inefficient way of searching any database. The subject thesaurus system adopted by most text databases was designed to alleviate the very problems caused by differing use of terminology and variations in spelling. A subject search using the Medline index term 'disulfides' would have found all articles on this subject regardless of the spelling convention used.

If a searcher insists on looking for a specific word in any one field, most databases will accept some form of truncation symbol. Thus, in some Medline systems, a search for the term 'disul-' will detect any words beginning with that series of letters. In some of the better quality databases, truncation is also possible in the middle of words, so a search on 'col\*r' would detect both 'colour' and 'color'.

Indeed, there are many pitfalls associated with textword searching. In many databases, the search run on the word 'disulfide' would not find the plural 'disulfides'. In addition, no amount of regulation would do away with variation in spelling due to typographical errors, which can slip through the editorial net both in the original journal and at a later stage when the details are entered onto the database. The subject search method can help to minimize the effect of human error.

This whole scenario highlights one of the problems created by the brave new world of the Internet. The information on any database is of value only when it can be accessed efficiently and effectively. In order to do

that, people must have adequate training in search techniques appropriate for each database accessed. Most Internet sources offer little or no training in the art of designing and carrying out a high-quality search strategy.

Those considering 'solo' literature searching as their main information-seeking strategy would be well advised to contact their local academic librarian, who should be able to offer help and training in professional search syntax and techniques.

**Bernadette Coles**

Violet Hughes Library,  
Velindre Hospital,  
Cardiff CF4 7XL, UK  
e-mail: wlbmc@thor.cf.ac.uk

SIR — Ito writes about the necessity to search for both 'sulphide' and 'sulfide' when using a CD-ROM. This problem is worse in some specialities than others. Child health is spelt 'pediatrics' in North America and 'paediatrics' in the United Kingdom, and 'anaesthesia' in Britain is 'anesthesia' in North America. 'Fetus' has almost completely replaced 'foetus', although I see there was a 3:1 split in this spelling in papers in translation from Japan last year; this bears out the experimental observation that a language mutates fastest at its point of origin. All this can make life difficult for a paediatric anaesthetist. For complete information, one has to use the wild card in search algorithms.

Ito's proposed solution to standardize spelling by jiggling a vowel or two would not solve the additional problem that words themselves may be different. I am an anaesthetist in Europe but an anesthesiologist in America, just as orthopods are otherwise orthopedic surgeons or even orthopedists. His solution, in short, would not combat the underlying problem that languages change — which is not a problem but a necessary accompaniment to describe technological innovations and advances that are yet to come.

**Oliver Dearlove**

Department of Paediatric Anaesthesia,  
Royal Manchester Children's Hospital,  
Manchester M25 2HA, UK  
e-mail: oliver.dearlove@man.ac.uk

## Never say never?

SIR — In his News and Views report about superconductors (*Nature* 382, 760; 1996), David Bishop misattributes to Yogi Berra a quotation by Dan Cook, who said: "The opera ain't over till the fat lady sings". Yogi Berra said: "It ain't over until it's over." Of course, those of us in the academic world know that its never over.

**Michael Katz**

March of Dimes Birth  
Defects Foundation,  
1275 Mamaroneck Avenue,  
White Plains, New York 10605, USA  
e-mail: katzmi@nyc.pipeline.com