



UNIVERSITY OF LEEDS

This is a repository copy of *Erroneous error correction*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/939/>

Article:

Smith, A. orcid.org/0000-0001-6224-3700 (1983) *Erroneous error correction*. *New Library World*, 84 (2002). p. 198. ISSN 0307-4803

Reuse

See Attached

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



White Rose
university consortium
Universities of Leeds, Sheffield & York

White Rose Consortium ePrints Repository

<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **New Library World**.
This paper may not include final publisher proof-corrections or journal pagination.

White Rose Repository URL for this paper:
<http://eprints.whiterose.ac.uk/archive/00000939/>

Citation for the published paper

Smith, A. (1983) *Erroneous error correction*. *New Library World*, 84 (2002), 198.

Citation for this paper

To refer to the repository paper, the following format may be used:

Smith, A. (1983) *Erroneous error correction*. Author manuscript available at:
<http://eprints.whiterose.ac.uk/archive/00000939/> [Accessed: *date*].

Published in final edited form as:

Smith, A. (1983) *Erroneous error correction*. *New Library World*, 84 (2002), 198.

New library world
Vol. 84
Number 2002
Page 198

December 1983
ISSN 0307 4803

Erroneous error correction

Adrian Smith, Leeds University Library, Leeds LS2 9JT
email a.smith@leeds.ac.uk

Libraries have been using computers for years now. Many librarians are not mystified by mega-byte-sized jargon and take MOP-buckets in their stride. Nonetheless those black boxes can still come up with a surprise or two. I know a library which has just installed a new issue system (it hardly matters which brand, since this is just a cautionary tale). This computer has been told to capitalise each word in the book-title (machines don't have to write in CAPITALS now), which looked a little odd when it came to Richard Iii. It was also given a list of stop-words (The, A, Le, La ...) which are not searchable, frustrating anyone looking for titles like *A B C of ...* or *A E Houseman* or books by Mr Das or Ms Lo. On one occasion the index 'slipped' overnight and to look up SMITH one actually had to enter TNJUI.

Naturally, bigger computers allow you to make bigger mistakes. Online searching, rather like fishing in the dark, gives you every opportunity to go in over your neck. In one search, seed-germination of hardy annuals, I retrieved everything apparently published in annual reports or annual reviews; in another, on lead-pollution, we pulled out anything with the verb 'to lead'. In developmental psychology, looking for infant or infants, I inadvertently called up the infantry (they warn you about it in the manual, but I didn't think they were still

developing). And then I discovered *Science citation index*^{*}, and I think they've made a mistake all on their own.

Large databases (like SCI) are, naturally, likely to contain large numbers of errors, even where these are small proportionately. Automatic procedures for validating terms are therefore attractive to database producers (and publishers of printed sources). They have been used to detect and correct errors in *Chemical abstracts*. Librarians know all about what Duncan Blanchard called 'references and unreferences' (*Science*, vol.185, page 1003, Sept 1974 for those who like footnotes), and so do the editors at ISI (the Institute for Scientific Information). In 1981 they introduced 'new computer programs for correction and unification of citation data'. Unfortunately, I believe that part of this verification procedure can actually multiply errors, which are now appearing in *Science citation index*. It seems that the machine checks 'new' references which are given to it, against those it already 'knows' in the database. Where a match occurs in some particulars this *deus ex machina* obligingly glosses over human frailty and changes the new citation to conform with the established one. A match of volume, page numbers and title may be enough for the machine to change the author's name. Where the mistake was actually in the earlier reference, the error is thus multiplied.

My suspicion was roused when scanning the SCI entries for H U Bergmeyer's *Methods of enzymatic analysis*. This in itself requires some concentration as there are various German and English-language editions. Connoisseurs of unreferences will not be surprised to know that this also involved looking up BERGMAYER, BERGMEIER, BERGMEIJER and BERGMEYR, and checking 20 variants of initials (transpositions and transcription errors) from A U to W V. You can see why ISI would love to have a computer tidy up these junk references. However I was surprised to find an erroneous 'MVC' Bergmeyer (a specific reference to page 783 of the 1963 edition) in the 1980

SCI reappearing in the computer-corrected 1981 SCI; seven authors, in nine papers, had apparently discovered the relevance of page 783 of 'MVC' Bergmeyer's non-book to their own work. My first thought was that these later authors had picked up the reference and, taking it as read, cited it as it stood (a common enough ploy). A quick check of two of the papers in our own Library showed, however, that while the page number was correct, the initials 'MVC' had been introduced at ISI and had not been present in the original papers. The machine had rules; number were perhaps more plausible, or inherently more accurate than names and the credulous thing determined that 783 must be right and H U Bergmeyer should be changed to fit the database. At least that's my theory; I'm still waiting to hear from Dr Garfield.

All of which is not to deride ISI, the computer, or new technology. I would not be without the citation index, and if issue systems had not been automated, I would be filing issue slips until 9pm. Just remember that machines that work fast, can also make more mistakes per second than the human 'factor', who made it 'go'.

* *Science citation index*, SCI, Institute for Scientific Information and ISI are all registered names.

New Library World, Vol.84, Page 198 (December 1983)
NLW Journals Ltd, 16 Pembridge Road, London W11 3HL