Control and Elimination of Errors in ISI Services*

IRVING H. SHER, EUGENE GARFIELD, and ARTHUR W. ELIAS
Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106
Received May 16, 1966

The Institute for Scientific Information produces several indexes and abstracting and alerting services which contain error-controlling features. Current Contents, Index Chemicus have different and unique mechanisms for finding and correcting errors that have appeared in the primary literature and those generated during their production of these secondary publications. Aspects of error control in these publications will be discussed.

Current Contents, Index Chemicus (IC), Science Citation Index (SCI), and ASCA (Automatic Subject Citation Alert) each have different and unique mechanisms for finding and correcting errors that have appeared in the primary literature and those generated during the production of these secondary publications.

It is obviously only possible to state very briefly some of the problems that exist, as well as some of the methods that are used to overcome them. This paper illustrates some types of errors involving people, other types involving machines, and, finally, general or composite errors.

The error-checking methods vary in complexity and extent according to the circumstances. For example, in the production of the Index Chemicus, not only are the articles indexed and graphically abstracted, but the molecular formulas of the individual compounds described by each author are recalculated by trained chemists. These abstracts are then sent to the original author for approval, giving him the opportunity of confirming corrected errors or making additional changes. Thus, the data found in Index Chemicus are sometimes more accurate than in the original article from which the abstract was prepared.

There is another important aspect to the whole problem of errors in the literature. Consider the perpetuation of a mistaken method or data that can go on being used for years without knowledge of subsequent modification. It is one of the unique capabilities of the Science Citation Index system that a straightforward check of the indexes will reveal such “corrections” (Figure 1).

An example of what one error in Current Contents can do is the case of an article by D. M. Baron of London, D.G. Chapman, ANN. MATH. ST. 36, 1583 (1965).

**CORRECTION TO**

"A COMPARATIVE STUDY OF SEVERAL ONE-SIDED GOODNESS-OF-FIT TESTS”

BY D. G. CHAPMAN

In the paper cited above (Ann. Math. Statist. 29 (1958) 655-674), it is stated that “my monotone test is admissible.” This is in reference to the hypothesis $P = F$ against the alternative $P < F$. R. Doksum has pointed out that this test is not a counter-example to this assertion which should therefore be deleted.

1983

Figure 1. Actual correction note by Chapman citing original 1958 paper.
England. We intended, with Professor Baron's consent, to present his address in a slightly modified form in the address directory of Current Contents. However, the address for Professor Baron was listed in Current Contents as Athens, Greece, instead of London, England. By coincidence there happened to be a Dr. D. Baronos in Athens, Greece, who eventually received 17 reprint requests. Dr. Baronos was kind enough to forward these to Professor Baron and enable us to complete the experiment.

An example of a mechanical aid that can be helpful in checking for errors is the HydroBond Computer, a circular slide rule invented at ISI, and used by the chemist-typewriter are used to check for errors in the molecular formulas of newly reported compounds (1). This hand calculator provides a simple checking procedure that can also reveal when an index may have inadvertently selected the wrong Markush group for a derivative (2).

In addition, the electronic computer and automatic typewriter are used to check for errors in the molecular formulas which might be introduced during keypunching— for errors in the spacing between chemical elements.

The large-scale data processing associated with the preparation of the Science Citation Index provides numerous opportunities for correcting errors that have appeared in the literature. For example, in a paper by C. L. Standley published in the Journal of Applied Physics (3), reference was made to an article by C. A. Mead (4) which had appeared earlier in the Physical Review. However, in Standley's article, this reference is erroneously cited as having appeared in the Journal of Applied Physics. This brought to our attention through a phone call from a researcher who was trying to find the elusive article by Mead. As shown in Figure 2, the error was immediately exposed by examination of the citation index entries for C. A. Mead (5). The erroneous citation preceded five correct references to the same paper. The true identity of the incorrect reference can be deduced because of the coincidence of the cited author, year, volume, and page. In this particular instance, the automatic computer correction routine we employ leaves the error unchanged. On the other hand, the logic of the computer programs we use actually does modify and correct most literature errors of this type in such a way that the user of the Science Citation Index would never know that Standley had, in fact, cited the paper incorrectly. Thus, citation errors rarely result in a complete loss of information.

Figure 3 illustrates an occurrence of an error in the literature which is automatically corrected by the computer during preparation of the Science Citation Index (6). The article by Hirata (7) misspells C. H. Altshuler's name when citing the 1949 work from the American Journal of Pathology (8). When other correct citations to the same work are brought together with the error, the correction procedure modifies the incorrect spelling so that all the citing works appear properly together under a single header for Altshuler, spelled correctly. This unification procedure selects the most correct version of the cited author's name and cited publication title independently from each cited work. The unification does not extend across different cited works.

MEAD CA---------------61-J APPL PHYS---------32 646
CHAPMAN RA J APPL PHYS 64 35 2832
CHRISTY R W J APPL PHYS 64 35 2179
COREN J J APPL PHYS N 64 35 3056
FORKEN PI MUD CEMENT 64 31 1246
HARRIS LA J APPL PHYS 64 35 268
HARTMAN TE PHYS REV A 64 134 1094
MANN HT J APPL PHYS 64 35 2173
RHODERIC RH JR J APPL PHYS 64 65 1613
STANDLEY CL J APPL PHYS 64 35 130

STANDLEY CL J APPL PHYS 64 35 1268
STANDLEY CL J APPL PHYS 64 35 1503
STANDLEY CL J APPL PHYS 64 35 1530

62-62-PHYS REV LETTERS---------8 56
BERGLUND CN PHYS REV A 64 136 1044
HARRIS LA J APPL PHYS 64 35 268
MOTIZUKI K J PHYS JAP 64 44 1968

62-62-PHYS REV LETTERS---------9 46
MOTIZUKI K J PHYS JAP 64 19 468
MOTIZUKI K J PHYS JAP 64 44 1968
MOTIZUKI K J PHYS JAP 64 19 468
MOTIZUKI K J PHYS JAP 64 44 1968

62-62-PHYS REV---------------126 2088
BASHEA MM IEEE COMP P 64 CP11 4
BASHER NM J APPL PHYS 64 35 3498
KU HY J APPL PHYS 64 35 265
POLLACK SR J APPL PHYS 64 35 1503

62-62-PHYS REV LETTERS---------9 46
POLLACK SR J APPL PHYS 64 35 1503
POLLACK SR J APPL PHYS 64 35 1503
POLLACK SR J APPL PHYS 64 35 1503

62-62-PHYS REV LETTERS---------8 56
BASHEA MM IEEE COMP P 64 CP11 4
BASHER NM J APPL PHYS 64 35 3498
KU HY J APPL PHYS 64 35 265
POLLACK SR J APPL PHYS 64 35 1503

62-62-PHYS REV LETTERS---------8 56
HARRIS LA J APPL PHYS 64 35 268
HARTMAN TE PHYS REV A 64 136 1044
HICKMOTT TW J APPL PHYS 64 35 2118
VANDERZEL J RES NBS D M 64 D 68 5660

Figure 2. Page from SCI showing Standley's erroneous citation of reference article.

Figure 3. Example of automatic computer correction of spelling error in citation of Altshuler's work.

In producing our early experimental citation index files we were not verifying the punched card file containing the cited publication title. We encountered a case where one keypuncher consistently made a transposition finger error which changed "diary" to "diary". The weight of her cumulated mistakes actually overrode the correct version in some instances when our unification procedure was applied. This resulted in printouts containing citations to a nonexistent Journal of "Diary" Science. Needless
to say, this was partially instrumental in our changing to full verification. This was not a trivial decision insofar as it involves more than 3 million cards per year.

In the preparation of the IBM cards for the Science Citation Index, all keypunched data are verified by a second operator or reader. The IBM 557 interpreter with proof feature is used to interpret the cards and edit for erroneous multipunch characters and a preliminary field edit. After all the errors detected by the verifier have been corrected, an IBM 360 computer is used to edit and identify any further errors of the type which can be checked for by algorithm. This very elaborate computer procedure confirms card sequencing, field positioning, and identify any further errors of the type which can be found.

ASCA, the Automatic Subject Citation Alert, is a personalized selective dissemination system designed to alert any subscriber rapidly to the appearance of published material relevant to his interests. ASCA utilizes an extremely complex system to ensure that errors or variations in literature presentation do not interfere with the proper servicing of an individual's specific profile of interest. This control complex extends all the way from preliminary examination of each journal issue through assured receipt of reports in the mail. We try, for instance, to ensure the ASCA subscriber against an error of omission (whether ISI's or the postal system's) by sending him a report every week whether or not any information was detected. Figure 4 illustrates a report that is sent when a negative search has resulted for a particular week. The error we are ensuring against here is that the customer may have missed a report which did, in fact, contain information. As long as he knows that he must receive a weekly report, he then knows that the failure to receive such a report means that it has gone astray somewhere. For the rare occasions when that does happen, we maintain a file copy of every ASCA report so that it can be retrieved promptly for the customer who files a claim for a missing report.

Like any other large library, we must be careful about the receipt of our journals in order to guarantee prompt and complete coverage. We not only receive several copies of each journal, each of which may be received from different sources, but also use a rather elaborate journal inventory control system. Among other things, this system checks for the arrival of journals. As shown in Figure 5, the system also alerts each ISI product manager to those journals which have arrived—which are to be covered by his product—but have not yet arrived in his department.

Another very important and vital aspect of CC, SCI, and ASCA processing of journals is page-by-page indexing in contrast to the selective indexing done by most other services. Since ISI supplies well-defined outputs based on complete coverage of all the journals we process, we and the users know specifically what data should be found in the publications. This knowledge greatly facilitates the detection of errors of omission.

The inventory of journal issues indexed appears in every issue of CC, IC, and SCI, giving the user precise information exactly what is, and thereby what is not, covered.
SCI coverage is identical to a cumulation of the weekly ASCA coverage.

The total error control which we employ is, in practice, inseparable from the total system of production. We consider it an important guiding principle that humans and machines be coordinated in error-checking systems—the two used in complementary capacities.

LITERATURE CITED


(2) Elias, A. W., Garfield, E., Foeman, G. H., Revesz, G., "Mechanization of Chemical Information Publications and Services;" paper presented at the 149th National Meeting of the American Chemical Society, Division of Chemical Literature, 149th Meeting, ACS, Detroit, April 1965.


(5) Garfield, E., Sher, I. H., Science Citation Index 1964, Annual Cumulation, Citation Index, Part 3, Column 8001, Institute for Scientific Information, Philadelphia, Pa., 1965.

(6) Ibid., 1965, Annual Cumulation, Citation Index, Part 1, Column 314, Institute for Scientific Information, Philadelphia, Pa., 1966.
