

## 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,

### 1965 SCIENCE CITATION INDEX

CHAPMAN DG	*****	*IN PRESS	22	191
MURPHY GI	J FISH RES	65	27	37
DRING SV	49-CANAD J RES	65	120	45
GREENGAR O	P SOC EXP M	65	30	194
BLISCHKE WR	50-SCI AGR	65	147	1571
TUKEY JW	51-ANN MATH STATIST	65	22	581
HALL WL	ANN MATH ST	65	36	352
PELLETIE O	P NAS US	65	53	127
KNIGHT W	51-ANAL CHEM	65	23	1113
MOORJANI MN	HALL WL	65	48	1236
LIBBY DA	PELLETIE O	65	47	401
MORRISON AB	52-BIOMETRICS	65	8	286
HARTER HL	ANN MATH ST	65	36	1494
MORRISON AB	54-CAN J BIOCHEM PHYSIO	65	37	679
BARTON DE	MOORJANI MN	65	19	212
CAPON J	54-J PHARMACEUTICAL SCI	65	43	297
CHAPMAN DG	LIBBY DA	65	48	981
CSORGO M	MORRISON AB	65	54	1
DAVID HT	HARTER HL	65	27	498
QUADE D	MORRISON AB	65	7	639
BRESSANI R	56-ANN MATH STAT	65	45	374
KM	TECHNOM	65	54	1
ONE EE	58-ANN MATH STATIST	65	29	655
ONE EE	BARTON DE	65	36	236
JANSEN GR	CAPON J	65	60	843
KRISHNAN K	CHAPMAN DG	65	36	1583
KRISHNAN MA	CSORGO M	65	36	322
MORRISON AB	DAVID HT	65	36	897
	QUADE D	65	36	1000
	59-CANAD J BIOCHEM PHYS	65	37	679
	NUTR DIET	65	7	161
	BJC J NUTR	65	19	125
	AM J CLIN N	65	16	315
	AM J CLIN N	65	16	321
	FOOD TECHN	65	19	1439
	L J EX BIOL	65	43	168
	CAN J BIOCH	65		
	FOOD TECHN	65		
	CAN	65		

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 "any monotone test is admissible." This is in reference to the hypothesis  $F = F_0$  against the alternative  $F < F_0$ . K. Doksum has pointed out that the test  $\phi = \alpha$  is a counter-example to this assertion which should therefore be deleted.

1583

Figure 1. Actual correction note by Chapman citing original 1958 paper.

\* Presented before the Division of Chemical Literature, Symposium on Error Control in the Chemical Literature, 151st National Meeting of the American Chemical Society, Pittsburgh, Pa., March 23, 1966.

England. We intended, with Professor Baron's consent, to present his address in a slightly modified form in the address directory of *Current Contents*. This coding was to enable him to recognize those reprint requests which came to him through users of the author address directory in *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-indexers when recalculating molecular formulas of newly reported compounds (1). This hand calculator provides a simple checking procedure that can also reveal when an indexer 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 example, 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 was 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 precedes 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 RW J APPL PHYS	64	35 2179
COHEN J J APPL PHYS N 64	35	3056
FORLANI F NUOV CIMENT	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.EH BR J A PHYS	64	15 613
STANDLEY CL J APPL PHYS	64	35 1530
-----61-PHYS REV LETTERS-----	6	545
BASHARA NM IEEE COMP P	64	CP11 4
BASHARA NM J APPL PHYS	64	35 3498
KU HY J APPL PHYS	64	35 265
POLLACK SR J APPL PHYS	64	35 1503
-----62-J APPL PHYS-----	128	2088
STANDLEY CL J APPL PHYS	64	35 1530
-----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	19 486
-----62-PHYS REV LETTERS-----	9	46
MOTIZUKI K J PHYS JAP	64	19 486
-----62-PHYS REV-----	128	2088
BASHARA NM IEEE COMP P	64	CP11 4
HARRIS LA J APPL PHYS	64	35 268
HARTMAN TE PHYS REV A	64	134 1094
HICKMOTT TW J APPL PHYS	64	35 2118
VANDERZI.A J RES NBS D	M 64	D 68 660

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

36

Mataru Hirata

## ACKNOWLEDGEMENT

It is a great pleasure to acknowledge gratefully the guidance I have received from Prof. Uemura. I would also like to thank Assistant Prof. Iwamori (Department of Surgery, Research Institute for Nuclear Medicine and Biology) for his advice, and Assistant Prof. Okuda (Department of Biochemistry) for his helpful criticism.

## REFERENCES

- Altshuler, C. H., and Angevine, D. M.: Histochemical Studies on the Pathogenesis of Fibrinoid, *Am. J. Path.*, New York, 25, 1061-1076, 1949.
- Boas, N. F., and Ludwig, A. W.: The Mechanism of Estrogen Inhibition of Comb Growth in the Cockerel, with Histologic Observations, *Endocrinol.*, Springfield, 46, 299-306, 1950.
- Boas, N. F., and Peterman, A. F.: Effects of Age, Food Intake, and Stress on Plasma Hexamine Levels in the Rat, *Proc. Soc. Exp. Biol.*, New York, 82, 19-24, 1955.

ALTSHULER CH-----49-AMER J PATH-----	25	1061
RR MED J	E 65	1 877
HIRATA M HIROS J MED	64	13 29
HORN RG AM J PATH	65	46 197
SCHUEER G ACT HISTOCH	65	20 91
-----50-AMER J PATH-----	65	24 693
MOROZOVA MM FED PROC	65	24 T515
-----51-AMER J PATH-----	27	41
BRAUNFAL.D DEUT MED WO	65	90 2269
-----63-ARCH PATH-----	75	206
SEPPALA P SC J CL INV	64	S 16 9
ALTSHULER E-----64-IEEE T ANTENNAS PROPAP	8	526
IIZUKA K IEEE ANTENN	N 65	AP13 469
ALTSHULER EE-----61-IEEE T ANTENNAS PROPAP	9	324
WU TT IEEE ANTENN	65	AP13 369
ALTSHULER KZ-----57-EUGENICS QUARTERLY--	4	
HANHART I ACT GENET M	65	14 13
-----62-EXPANDING GOALS GENE	65	275
RAINER JD NY ST J MED	M 65	65 1651

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 field containing the cited publication title. We encountered a case where one keypuncher consistently made a transposition finger error which changed "dairy" 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 acceptability of the types of data allowed in the various field positions.

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.

**ascas**  
AUTOMATIC SUBJECT CITATION ALERT

a service of the INSTITUTE FOR SCIENTIFIC INFORMATION

DR. CHARLES VOYTKO  
CHEMIOLOGY INSTITUTE  
DEPT. OF RESEARCH  
NEW YORK, NY 32516

D6679 ACCOUNT NUMBER  
49 UNITS USED  
1 UNITS REMAINING

REPORT FOR 22 OCT 65

68,448 citations from current scientific literature and  
current patents were processed for ASCA this week

NO HITS THIS WEEK. USE BACK TO ADD NEW ITEMS TO YOUR PROFILE.  
ACCT NO 11

Figure 4. ASCA weekly report sent to subscriber when no new information has been detected in search.

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 on exactly what is, and thereby what is not, covered.

JOURNAL TITLE	VOLUME	ISSUE	MONTH	YEAR	DIST	RECEIVED ISI	CC	STILL REQUIRED BY SCI
ZH FIZ KHIM	39	12	DEC	1965	S I	01-06-66		
BULL SOC ITAL BIOL SPER	41	19	OCT 15	1965	PK	01-13-66		REQ.
DIS COLON RECTUM	9	1	JANFEB	1966	PK	02-14-66	P	REQ.
NEPHRON	2	4		1965	PK	02-12-66	P	REQ.
PHYSICA STATUS SOLIDI	13	1		1966	SK	02-02-66		REQ.
SOVIET PHYS-JETP ENGL TRANS	22	1	4AN	1966	SK	02-02-66		REQ.
INORG CHEM	5	1	JAN	1966	SKI	01-07-66		
DOKL AKAD NAUK SSSR	165	6		1965	CKI	01-10-66		REQ.
ZH ORG KHIM	1	12		1965	C I	01-10-66		
BULL CHEM SOC JAPAN	38	12	DEC	1965	CKI	01-11-66		
ACTA CHEM SCAND	19	8		1965	CKI	01-12-66		
J ORGANOMETALLIC CHEM	4	6	DEC	1965	SKI	01-12-66		
ARCH PHARM	298	12	DEC	1965	PKI	01-13-66		
ISRAEL J CHEM	3	4	DEC	1965	CKI	02-12-66	S&P	REQ.
SOVIET PHYS-USPEKHI ENGL TRAN	8	3	NOVDEC	1965	SK	02-14-66	S	REQ.
ANGEW CHEM	78	1	JAN	1966	C I	01-15-66		
J AM CHEM SOC	88	1	JAN 5	1966	CKI	01-15-66		
ANAL CHIM ACTA	34	1	JAN	1966	CKI	01-17-66		
BIOCHEM J	98	1	JAN	1966	PKI	01-17-66		
CHEM PHARM BULL TOKYO	13	12	DEC	1965	PKI	01-17-66		
J PHARM SOC JAPAN	85	12	DEC	1965	P I	01-17-66		
DOKL AKAD NAUK SSSR	165	1		1965	CKI	01-18-66		
J PHARM SCI	55	1	JAN	1966	PKI	01-18-66		

Figure 5. ISI journal inventory system showing arrival dates and departmental requirement alert.

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

- (1) Sher, I. H., Foeman, G. H., Baus, E. H., "A Slide Rule for Calculation of the Number of Double Bonds and Hydrogen Atoms;" paper presented at the 145th National Meeting of the American Chemical Society, Division of Chemical Literature, New York, September 1963.
- (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.
- (3) Standley, C. L., *J. Appl. Phys.* **35**, 1530 (1964).
- (4) Mead, C. A., *Phys. Rev.* **128**, 2088 (1962).
- (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.
- (7) Hirata, M., *Hiroshima J. Med. Sci.* **13**(1), 29 (1964).
- (8) Altshuler, C. H., Angevine, D. M., *Am. J. Pathol.* **25**, 1061 (1949).