Computer-assisted instruction courses on basic ophthalmology and glaucoma have been programmed and are available at the computer terminal. At present, the student must enter the instructional mode at the beginning of the course he has chosen and proceed sequentially through it. However, thesaurus code numbers have been assigned to frame sequences of the programmed instruction which deal with specific subjects. When necessary programs are implemented, it will be possible for the student to request instruction on a specific subject of interest, for example, on the technique of ophthalmoscopy. The frames of the basic ophthalmology course which explain ophthalmoscopy will then be presented. In this manner, a user can readily receive both instruction and a bibliography on the subject he has selected from the thesaurus.

Additional search capabilities are being planned which will be incorporated into the VIC system. These include search by author, language, and date of publication as well as by journal title. The latter capability provides the potential of generating the annual index for any given journal. It is anticipated that use of the present VIC data base by members of the scientific community will provide feedback and critical suggestions which will guide the Vision Information Center in its future development plans.

LITERATURE CITED

(1) U. S. Department of Health, Education, and Welfare, Public Health Service, National Library of Medicine, "Medical Subject Headings," Vol. **8**, U. S. Government Printing Office, Washington, D. C., 1967.

(2) Reinecke, Robert D., "Vision Information Center—Direct User Access via Computer-Assisted Instruction," *Proc. Am. Soc. Inform. Sci.*, Vol. **5**, "Information Transfer," Greenwood Publishing Corp., New York, 1968.

(3) Roth, Emanuel M., Charles W. Sargent, and Marjorie M. Benson, "Aerospace and Environmental Medicine Information System Thesaurus," Department of Aerospace Medicine and Bioastronautics, Lovelace Foundation for Medical Education and Research, Albuquerque, N. M., 1966.

(4) Schappert-Kimmijser, J., A. Colenbrander, and S. Franken, "Coding System for Disorders of the Eye," S. Karger AG, Basel, Switzerland, 1968.

(5) Breed, Robert S., E. G. D. Murray, and Nathan R. Smith, Eds., "Bergey's Manual of Determinative Bacteriology," 7th ed., Williams and Wilkens Co., Baltimore, Md., 1957.

(6) Goodman, Louis S., and Alfred Gilman, "The Pharmacological Basis of Therapeutics," 3rd ed., MacMillan, New York, 1965.

(7) Florkin, Marcel, and Elmer H. Stotz, Eds., "Comprehensive Biochemistry," Vol. **13**, 2nd ed., Elsevier, New York, 1965.

Supported by DHEW Contract PH **43-66-911** from the National Institute of Neurological Diseases and Stroke.

Compatibility in Chemical Information Systems

DAVID P. JACOBUS
Walter Reed Army Institute of Research, Washington, D.C. 20012

KENNETH H. ZABRISKIE
BIOSIS, Philadelphia, Pa.

MAXWELL GORDON*
Smith Kline and French Laboratories, Philadelphia, Pa. 19101

Received January 31, 1969

As increasing amounts of data are processed for computer storage and retrieval, it becomes important to amortize the cost of this operation more rapidly, by a wider exchange of data. Obviously, exchange of machine-processed data introduces problems of compatibility and convertibility. It is urged that planning of information operations take these problems into account at an early stage, since seemingly minor differences between data systems can have disastrous effects on data exchange efforts. The compatibility and convertibility of chemical data, as special cases of the foregoing, have their own special problems. A survey of these problems is presented, together with some suggested solutions to compatibility and convertibility in this area.

The problem of compatibility among information systems becomes increasingly important as systems become more and more mechanized. It is our feeling that, inasmuch as the cost of input of scientific information is great, maximum use of this information will be possible only if it can be exchanged among a large community of users to reduce the unit cost and optimize the use of our intellectual resources. It has been suggested that the cost of information processing and transmission will one day exceed the cost of input. However, compatibility considerations have relevance to information processing and transmission, as well as to input costs.

For the past several years, the Committee on Chemical Information of the National Academy of Sciences/

* To whom questions should be directed.

National Research Council, through its subcommittee on Economics and Evaluation, has been studying the question of compatibility among information systems, with particular reference to compatibility among chemical typewriters in their role as input devices to computer-based systems for handling chemical information. The results of these deliberations are described in this article. The authors of this paper are members of the subcommittee on Economics and Evaluation, and each is active in the field of implementing computer-based systems of handling chemical information.

In the course of these discussions, it also became apparent that much information is keyboarded in duplicate throughout the country, in addition to earlier keyboarding by the author's secretary and then by primary journals. In particular, citations and headings from journals in the chemical-biomedical areas are processed by a number of agencies, such as Chemical Abstracts, Biological Abstracts, National Library of Medicine, Institute for Scientific Information, and others. Since our goal was to explore ways of achieving compatibility, it occurred to us that if compatibility (or at least convertibility) among the forms of heading information generated by these several organizations could be encouraged, then some exchange of title and citation tapes might be possible and could yield considerable savings. Since much of this work is directly or indirectly paid for by the Federal Government, we feel this subcommittee can perform a useful public service role in trying to effect savings of time, human effort, and dollars in the input of titles and citations to computer systems.

As a result of discussions in this area, we would like to postulate the concept of an "International Keyboard Pool" which would serve as the mechanism for an exchange of title and citation tapes from among the parties concerned. A precedent for this pool is available in the operation of the U.S.-Swiss-German Documentation Ring. Member organizations coded information from chemical-biomedical journals (apportioned among the member firms), entered data onto punched cards, and the punched cards were exchanged among the parties to this cooperative effort, for the benefit of all.

We have analyzed the journal coverage of the *Science Citation Index, CBAC, Medlars, Ringdoc, Biological Abstracts, and Chemical Abstracts,* and find very substantial areas of overlap in their coverage. We make the tentative estimate that a keyboard pool such as we suggest would effect a saving of about $300,000 a year in key punch costs for these publications, and greatly improve the speed and efficiency of exchanging information. This estimate is based on an actual count of the key strokes involved. Although the citation formats of the various publications differ, they are sufficiently standardized so that a minimum programming effort could render them convertible for use by any of the services mentioned.

REVIEW OF LITERATURE

In the past, we have discussed level of compatibility in terms of economic yardsticks, such as the one cited above, to assess the importance and difficulties associated with systems that may be compatible at different levels, if not identical. It should be noted that exchange of information even between "identical" systems may raise some very difficult problems, primarily as a result of slight differences in hardware or difference in "maintenance" routines. Hence, we cannot state too strongly that even minor incompatibilities may have a relatively major effect in preventing the integration or exchange of information by large systems. A resolution of other compatibility questions will have an even more profound effect on the economics of sharing chemical structural information. We also believe that, for compatibility to be achieved, appropriate considerations need to be introduced as early as possible into system design (preferably at the time of first publication, if possible, or before).

The National Science Foundation and the Office of Education have granted a total of $1,012,890 to the libraries at Columbia, Stanford, and the University of Chicago to fund the design and implementation of a computer-based library system that will permit coordination, and possibly integration of processing functions in the three libraries.

An interesting review entitled "Cooperation, Convertibility, and Compatibility Among Information Systems" has appeared.[1] "Cooperation" is described as a collaborative effort or sharing between organizations in actual processing of material for information systems, or an exchange of the products of such processing. Systems are considered to be compatible when the results of processing in one system are immediately and directly usable by other organizations having similar but not necessarily identical systems. When results and products of processing in one system are usable in another system, but not immediately or directly, the systems may or may not be convertible one to the other. Compatibility and convertibility may apply either mutually or in only one direction between and among information systems. The increasing use of machines in the compilation and preparation of printed indexes, as well as catalogs, announcements, bulletins, bibliographies and the like, will, as we pointed out above, require increased concern for compatibility and convertibility between systems. As the use of machines increases, more material will be in machine formats and codes, therefore increasing the importance of direct use or easy conversion.

Sherwin, formerly of the Department of Commerce,[2] has also directed his attention to the standardization of information-reporting systems. He makes the important point that automatic typesetting machinery is producing substantial quantities of full-digital text as a by-product, and this could serve as a useful input to an exchange system. Kuney *et al.*[5] have been moving in this direction, and the speculations of Gordon[3] are also pertinent.

The American Standards Association work entitled "American Standards for Periodical Title Abbreviations" should serve as a useful vehicle for the standardization of abbreviations used in citations. Compilation of abbreviations by the American Standards Association was carried out from lists provided by *Index Medicus, Biological Abstracts, Bibliography of Agriculture, and Chemical Abstracts.*[4]

Sherwin[2] has identified certain levels of compatibility. These are:

1. The use of a single standard code to convert from the alphanumeric to the binary representation. At the present time, several codes are used for paper tape punching. These

include 5-channel paper tape codes, 7-channel (binary coded decimal), and the US ASCII codes.
2. The use of standard data interchange media, on which the binary symbols are written or recorded.
3. The use of standard-format marker symbols which mark the beginning of the data package and its identifying number, and the beginning of the data element and its identifying number.
4. The use of certain system-wide codes; these may include the establishment of codes for the identification of individuals, organizations, geographical regions, etc. Examples of the existing codes of this type are social security numbers and postal zip codes.

Another enumeration of levels of standardization has been made by Henderson *et al.*,[1] who list standardization at the levels of:

1. Code or representation of characters in machine language, the number of characters accommodated by the code, and which characters are accommodated by the code.
2. Characters acceptable as input: size of allowable vocabulary; what characters; provision for other characters as required for multiple coding, shift keys, etc.
3. Characters acceptable as output.
4. Indexing vocabulary or documentary language.
5. Aids—e.g., word lists, thesauri, dictionaries, and the like.
6. Internal processing of character strings: fixed *vs.* variable word length, fixed *vs.* variable field length, fixed *vs.* variable record length, and fixed *vs.* variable file length.

Further discussion of the American Standard code for Information Exchange (US ASCII) is in order.[4] The US ASCII Standard specifies the correspondence of seven- or eight-bit patterns with 128 graphics and control codes. The graphics are for printable characters and the controls are to initiate appropriate equipment functions. The 128-character code set covers 36 controls, 28 special characters, 10 numerics and alphabetic assignments, leaving 28 bit-positions unassigned. A comparison of US ASCII and the coded binary decimal codes are shown on Charts V, VI, and VII in Appendix 2. Copies of US ASCII standards can be obtained from the USA Standards Institute, 10 E. 40th Street, New York, N. Y, 10017 (NBS Technical News Bulletin, August 1968, p. 173).

Henderson *et al.*[1] have called attention to the great disparity between the very large character sets used in printing and the character sets available on typical typewriter or key punch keyboards. This problem can ultimately be solved by the addition of typesetting instructions on paper or magnetic tapes at the page composition stage.[5] It is necessary only to have an escape code to make access to non-keyboard characters possible, as has been pointed out by Kuney *et al.*[5]

CHEMICAL TYPEWRITER COMPATIBILITY

Compatibility of chemical typewriters has been discussed in detail by this subcommittee for some time. Present or past members of this subcommittee have been associated with all of the modern chemical typewriter designs (D. Jacobus–Walter Reed Army Institute of Research–Mergenthaler and Teletype; J. Mullen–Shell Development–Dura and Camwil; M. Gordon–Smith Kline and French–Friden and Teletype). Whereas in earlier discussions (about 1965) it appeared that compatibility problems were rapidly increasing, we now find that a tentative stage of synthesis is being approached, so that we can feel hopeful that the problem can be attacked in a systematic fashion, with a reasonable prospect of solution. Several developments appear to support this optimism. In this paper we describe an algorithm for converting SK & F–Friden typewriter[3] output to Shell–Dura[10] input (Appendix 1). We also exemplify some rules for chemical typewriter structure input in Appendix 3.

Additional causes for optimism in the areas of compatibility or convertibility of chemical typewriters lie in the development of paper tape to card and paper tape to magnetic tape converters, programs for converting typewriter input to connection tables,[6, 17] conversion of connection tables to line notations and the reverse, connection tables to graphic structures and line notations to graphic structures.[7, 8, 9]

It must be evident that compatibility of chemical typewriters embraces many elements, including conventions for typing structures illustrated in Appendix 3, conventions for paper tape codes illustrated in Appendix 2, conventions for putting characters on magnetic tape, conventions for connection table generation, etc.

An actual example of a chain of incompatibilities that had to be dealt with in writing a conversion program for SK & F–Friden output to high-speed Photon output illustrates most of the problems that can arise.

The output from the SK & F–Friden typewriter is an 8-channel paper tape in binary coded decimal (see Appendix 2, Chart V) plus a hard copy of organic chemical structure useful for proofreading, but not of graphics quality. The task was to write an IBM 360/40 program for converting this paper tape to a 7-channel, 556-bit density magnetic tape containing Photon code equivalents for all of the characters, to give structures suitable for printing. This program is a relatively simple one, since there is an exact one-to-one correspondence between SK & F–Friden characters and the Photon matrix characters.

However, since there was no paper tape reader on the IBM 360, it was proposed to convert paper tape to tab cards using a Friden 2316 card punch control. This conversion appeared to operate smoothly until it was realized that the punch control was wired to read a typewriter carriage return code as a tab card transport code, and thus was automatically deleted from the punched card deck. To circumvent this problem, a deck of tab cards was punched with carriage return codes (7-, 9-punch) and these cards were inserted manually at the appropriate points of the tab card deck.

When this tedious task was nearly completed, the question was raised as to whether there might not be other codes which were being incorrectly translated. Closer inspection revealed that the code for our "right slant" was missing from the punch control. This was a vital key stroke, since it occurs twice in every benzene ring, and the task of inserting this code hundreds of times manually made us abandon this effort and seek an easier way of translating paper tape.

An SDS 925 computer, which had a paper tape reader, had first been dismissed from consideration because it was capable of reading only seven channels. However, the only use made of the eighth channel was to signal carriage returns, so the paper tapes were run back through the Flexowriter and a manual control code was added after each untranslatable (8th channel) carriage return.

A short program was then written for the SDS 925 to read the modified Flexowriter paper tapes and to produce a magnetic tape containing the chemical text in binary tape code. This translation was accomplished after several attempts, but when the SDS magnetic tapes were read on the IBM 360, it was found that at output many spurious zeros were printed. It developed that the SDS tape had to have two characters preceding each tape data code. Hence, the SDS program had to be rewritten to insert a 77 ahead of each data character, and the IBM Photon conversion program was modified to read the 77 as an invalid code so it would not print. Thus, the typewriter-to-computer-to-computer incompatibility was solved.

The above saga recounts about half of the incompatibility problems encountered in this small project, but it adequately illustrates the importance of the various considerations in this article.

CONNECTION TABLE COMPATIBILITY

Intellectually, it has appeared to us that connection table compatibility of structures was an attainable goal. We have isolated five levels of compatibility:

1. Identical machine language has been used in the two systems. Information can be exchanged in the form of a "processed" connection table.
2. Different codes are used to represent specific elements, bonds, etc., but the connection tables are otherwise identical. Here an interface may be useful; this may be a translation routine written for the computer or a specific piece of hardware.
3. Either the first or second situation above exists, but the connection table has been permuted—i.e., processed to a canonical form. Compatibility is possible, provided the user is prepared to pay the cost of processing the permuted connection table to the format used by him, or vice versa. This situation is likely to be a common one.
4. Noncompatible tables. For a variety of reasons, connection tables in one system can vary appreciably from those used in another. One table may contain much more information than another—e.g., some systems provide unusual isotope identification, unusual valences, or isomeric information. Conversion from a more complete to a less complete system is possible, but the reverse is not true.
5. Different graphic representation of the same chemical structure (synonomy of structure).

The subcommittee has examined how various systems use connection tables. Connection tables are (or could profitably be) used by all chemical structure information systems, at least during the initial processing of the structural information. Many systems do not use the connection table after the initial processing has been completed. It may never be economically feasible to process the entire connection table for routine searches where shorter, derived records will serve equally well.

Two main types of connection table use are:

To generate screens or fragment codes. A screen or fragment code denotes a substructure portion, larger than an atom, but smaller than the total structure; such codes are useful per se in classifying or screening compounds. These screens require much shorter magnetic tape records than the total connection tables.

To generate notations. A notation is here defined as a symbolic representation of a compound, usually linear, other than a connection table or pictorial representation of the structure. Notations may be compact or canonical or they may represent graphs, Polish notation, Dendral,[11] etc.

Because connection tables frequently are used only in this limited manner, the subcommittee believes that a connection table that had not been preprocessed would have to be processed if it were to be used by a receiving system. This preprocessing time will vary according to the system doing the preprocessing and is a cost item in the economics of compatibility. Probably, most systems will provide some processing of a connection table in order to reduce the mutual cost of systems which will have some degree of compatibility, thereby capitalizing on their existence. Unless the systems are identical, the receiving system will probably have to process the information for questions of identity with their existing files.

If the connection table is processed to develop a compacted notation or if this notation is produced beforehand, as in the Dow system,[14] compatibility at the level of a compact hierarchical notation is complex. Because of these complexities, the subcommittee feels that conversion between notations could be achieved more reliably by conversion through a compatible connection table. This level of conversion avoids the problems of symbol definition which have plagued workers in the field for many years. The economics of each conversion is, therefore, more easily determined. The control of each conversion is probably also more reliable and more easily monitored than a system involving the development of special rules.

Two other areas which the subcommittee has considered are compatibility between screens, which is felt to be relatively easy to achieve, and therefore of potential economic interest, and compatibility between the chemistry of various systems. The subcommittee has also considered compatibility of chemistry—i.e., valence tables, definitions of atoms and bonds, definition of aromaticity of keto-enol relationships, and definitions of structural diagrams.

Based on the present state-of-the-art, the subcommittee has reached the conclusion that for any two systems desiring to exchange compatible information, two separate programs must exist to permit the exchange of information in the two areas commonly considered important for establishing overall communication between systems:

1. If compatibility of high speed output is desired, translation programs involving the technical considerations discussed in the earlier part of this paper need to be available. At the present time, there is no working program which will create good quality diagrams from digital input, unless the input process produced diagrams as a by-product. If and when such programs become available, compatibility at this level will no longer be essential.
2. If compatibility between chemical identity is needed, a program involving conversion at the level of connection tables can be visualized. The subcommittee believes that compatibility at the level of the connection table would be economically desirable, even if structural diagram input was achieved through compatible systems, since the preprocessing associated with the generation of the connection tables would be saved. Such preprocessing costs would be a function of the receiving system. If the connection tables to be exchanged were relatively incompatible, but the structural diagrams were compatible, exchange at the level of the typewriter input would probably be desirable. Exchange between systems at the connection table level would be necessary for those systems having as a major input manual connection tables or an input such as the Wiswesser Notation.

APPENDIX 1

A brief discussion of the algorithm to convert SK & F–Friden typewriter output to the Shell/IBM dot-bond representation or to high-speed printer output is given here. Figure 1 represents a cyclohexane ring with x and y coordinates assigned. This assignment can be made by the computer from the historical paper tape record. Similarly, the dot-bond form of the ring is also shown with x and y coordinates, which can be deleted by computer after conversion (Figure 2).

This algorithm will also convert random typewriter input to line-at-a-time output, as well as permit conversion of SK & F–Friden input to a connection table.

It might be useful to point out that the SK & F–Friden font provides the most elegant structures likely to be accomplished on typewriters (and possibly also on photocomposition equipment) with minimum keystroking, but at some cost in the number of symbols required and in the difficulty of adaptation to high-speed printers. A similar font has been used by Kuney *et al.*[5] in their experiments. The basic Shell characters are minimal in number and adaptable on a one-for-one basis to high-speed printers, but they require more keystrokes at input and provide reduced elegance in the printed structure. The respective virtues make their interconversion mutually attractive.

In Chart I we have given the x and y coordinates for cyclohexane in Column 1, and the corresponding coordinates for the dot-bond representation are given in Column 2.

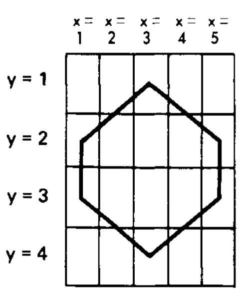


Figure 1. Flexowriter font

Column 1 Character	Column 1 Coordinates	Column 2 Character	Column 2 Coordinates	
/	x = 2, y = 2	•	x = 1, y = 3	
		/	x = 2, y = 2	
		•	x = 3, y = 1	pair
\	x = 4, y = 2	•	x = 3, y = 1	pair
		\	x = 4, y = 2	
		•	x = 5, y = 3	
|	x = 1, y = 3	•	x = 1, y = 3	pair
		|	x = 1, y = 4	
		•	x = 1, y = 5	
|	x = 5, y = 3	•	x = 5, y = 3	pair
		|	x = 5, y = 4	
		•	x = 5, y = 5	
\	x = 2, y = 4	•	x = 1, y = 5	
		\	x = 2, y = 6	
		•	x = 3, y = 7	pair
/	x = 4, y = 4	•	x = 3, y = 7	pair
		/	x = 4, y = 6	
		•	x = 5, y = 5	

Chart I. First step in translation of Friden to Shell font

Column 1 Character	Column 1 Coordinates	Column 2 Character	Column 2 Coordinates
/	x = 2, y = 2	°	x-1, y+1
		/	x,y
		°	x+1, y-1
\	x = 4, y = 2	°	x-1, y-1
		\	x,y
		°	x+1, y-1
|	x = 1, y = 3	°	x,y
		|	x,y+1
		°	x,y+2
|	x = 5, y = 3	°	x,y
		|	x,y+1
		°	x,y+2
\	x = 2, y = 4	°	x-1, y+1
		\	x,y+2
		°	x+1, y+3
/	x = 4, y = 4	°	x-1, y+3
		/	x,y+2
		°	x,1, y+1

Chart II. Second step in translation of Friden to Shell font

˘ (y=1, x-3) ° (y=1, x=3)
/ (y=2, x=2) \ (y=2, x=4)
° (y=3, x=1) ° (y=3, x=1) ° (y=3, x=5) ° (y=3, x=5)
| (y=4, x=1) | (y=4, x=5)
° (y=5, x=1) ° (y=5, x=1) ° (y=5, x=5) ° (y=5, x=5)
\ (y=6, x=2) / (y=6, x=4)
° (y=7, x=3) ° (y=7, x=3)

Chart III. Third step in translation of Friden to Shell font

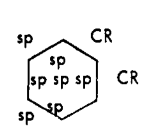
° (y=1, x=3)
/ (y=2, x=2) \ (y=2, x=4)
° (y=3, x=1) ° (y=3, x=5)
| (y=4, x=1) | (y=4, x=5)
° (y=5, x=1) ° (y=5, x=5)
\ (y=6, x=2) / (y=6, x=4)
° (y=7, x=3)

Chart IVa. Fourth step in translation of Friden of Shell font

(y=1)	line 1	sp	sp	°	CR		
	line 2	sp	/	sp	\	CR	
	line 3	°	sp	sp	sp	°	CR
	line 4	|	sp	sp	sp	|	CR
	line 5	°	sp	sp	sp	°	CR
	line 6	sp	\	sp	/	CR	
(y=7)	line 7	sp	sp	°			

= 35 keystrokes
(12 print strokes)

In comparison on SK&F Font



= 15 keystrokes
(6 print strokes)

Chart IVb. Last step in translation of Friden to Shell font.

AUTOMATIC WRITING MACHINE CODE CHART

Tab Card Punching Positions													Tape Channels											
12	11	0	1	2	3	4	5	6	7	8	9	CARD	8	7	6	5	4	3	2	1	2201	2301	2302	2303/2304
									1		1	EL	8								CR	CR	CR	CR
												SPACE				5					SPACE	SPACE	SPACE	SPACE
		1						1		1		SKIP			6	5	4	3	2		TAB	TAB	TAB	TAB
		1		1						1		PI-3			6		4		2		BACK SP		BACK SP	BACK SP
		1										0			6						) 0	) 0	) 0	0)
			1									1								1	! 1	! 1	! 1	1 '
				1								2							2		@ 2	@ 2	@ 2	2 @
					1							3				5			2	1	# 3	# 3	# 3	3 #
						1						4						3			$ 4	$ 4	$ 4	4 $
							1					5				5		3		1	= 5	= 5	= 5	5 =
								1				6				5		3	2		¢ 6	¢ 6	¢ 6	6 ¢
									1			7						3	2	1	? 7	? 7	? 7	7 ?
										1		8					4				* 8	* 8	* 8	8 *
											1	9				5	4			1	(9	(9	(9	9 (
1			1									A		7	6					1	a A	a A	a A	a A
1				1								B		7	6				2		b B	b B	b B	b B
1					1							C		7	6	5			2	1	c C	c C	c C	c C
1						1						D		7	6			3			d D	d D	d D	d D
1							1					E		7	6	5		3		1	e E	e E	e E	e E
1								1				F		7	6	5		3	2		f F	f F	f F	f F
1									1			G		7	6			3	2	1	g G	g G	g G	g G
1										1		H		7	6		4				h H	h H	h H	h H
1											1	I		7	6	5	4			1	i I	i I	i I	i I
	1		1									J		7		5				1	j J	j J	j J	j J
	1			1								K		7		5			2		k K	k K	k K	k K
	1				1							L		7					2	1	l L	l L	l L	l L
	1					1						M		7		5		3			m M	m M	m M	m M
	1						1					N		7				3		1	n N	n N	n N	n N
	1							1				O		7				3	2		o O	o O	o O	o O
	1								1			P		7		5		3	2	1	p P	p P	p P	p P
	1									1		Q		7		5	4				q Q	q Q	q Q	q Q
	1										1	R		7			4			1	r R	r R	r R	r R
		1		1								S			6	5			2		s S	s S	s S	s S
		1			1							T			6				2	1	t T	t T	t T	t T
		1				1						U			6	5		3			u U	u U	u U	u U
		1					1					V			6			3		1	v V	v V	v V	v V
		1						1				W			6			3	2		w W	w W	w W	w W
		1							1			X			6	5		3	2	1	x X	x X	x X	x X
		1								1		Y			6	5	4				y Y	y Y	y Y	y Y
		1									1	Z			6		4			1	z Z	z Z	z Z	z Z
					1					1		#					4		2	1	F1 MNP STOP	STOP	STOP	STOP
						1				1		@				5	4	3			F2	NP (AUX SP)	NP (AUX SP)	
		1				1				1		%			6		4	3			F3 PR RES	PR (AUX 0)	PR (AUX 0)	] [
	1					1				1		*		7			4	3			F4	ON 1	PUNCH ON	
1								1		1		SP2		7	6		4	3	2		F5	ON 2		
		1							1	1		E.C.2			6		4	3	2	1	F6	OFF	OFF	
	1							1		1		CR		7		5	4	3	2		F7	FC ON		3 UNITS
							1			1		PI-7					4	3		1	F8	TSR (AUX 1)	PI-7 (AUX 1)	= !
								1		1		E.C.1					4	3	2		F9	SW (AUX 2)	SW (AUX 2)	
									1	1		COR.				5	4	3	2	1	F10	DS (AUX 3)	DS (AUX 3)	
	1								1	1		ERROR		7			4	3	2	1	F11	FF (AUX L)	FF (AUX L)	FF (AUX L)
	1						1			1		PI-6		7		5	4	3		1	F12	AID (AUX J)	AID (AUX J)	
		1					1			1		PI-4			6	5	4	3		1	F13	DUP (AUX /)		
1						1				1		□		7	6	5	4	3			UPPER CASE	UPPER CASE	UPPER CASE	UPPER CASE
1				1						1		SP1		7	6	5	4		2		LOWER CASE	LOWER CASE	LOWER CASE	LOWER CASE
1					1					1		.		7	6		4		2	1	. .	. .	. .	. .
		1			1					1		,			6	5	4		2	1	, ,	, ,	, ,	, ,
	1											-		7							" -	" -	" -	- "
		1	1									/			6	5				1	: /	: /	: /	/ ?
	1				1					1		$		7		5	4		2	1	_ %	_ %	_ %	
1												&		7	6	5					; &	; &	; &	; :
				1						1		PI-1				5	4		2		± +	PI-1 (AUX 8)	PI-1 (AUX 8)	
1							1			1		PI-5		7	6		4	3		1	¼ ½	PI-5 (AUX A)	PI-5 (AUX A)	½ ¼
	1			1						1		PI-2		7			4		2		' °	PI-2 (2, 8, -)	± +	± +
1									1	1		FEED		7	6	5	4	3	2	1	TAPE FEED	TAPE FEED	TAPE FEED	TAPE FEED

Chart V. Coded binary decimal coding for paper tape and Hollerith card equivalents (Friden, Inc.)

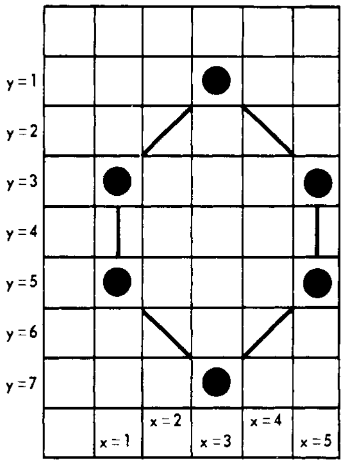


Figure 2. Organization suggested by H. Peter Luhn and devised by James Mullen provides flexibility in typing carbon atoms while reducing the number of characters required, thus facilitating input and output on machines having limited fonts

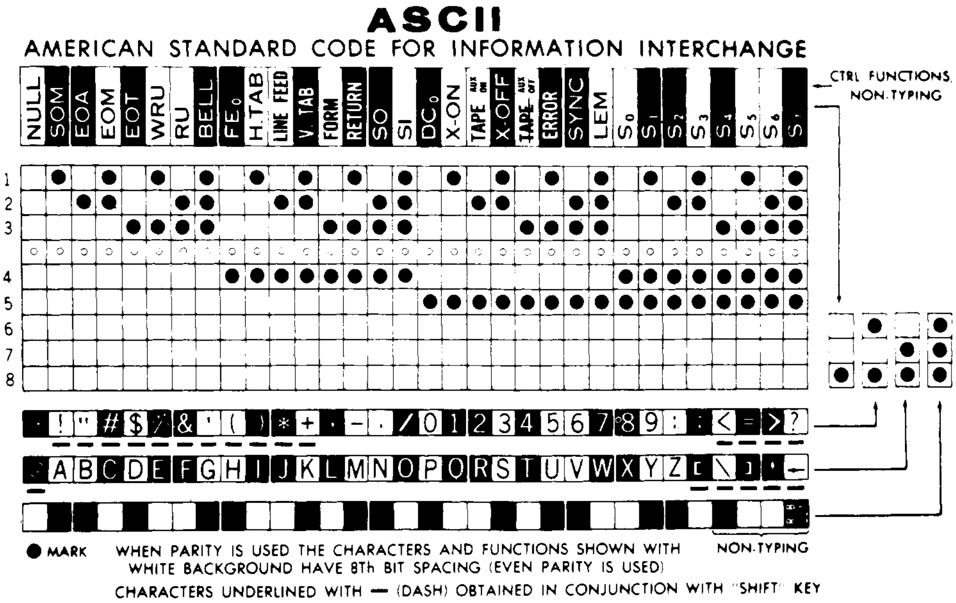


Chart VI. ASCII (Teletype) eight-channel paper tape code

Note that count is kept of y values, and $y = 2$ is added in Column 2 each time y in Column 1 increases by 1.

Column 2 contains all of the elements of a connection table (atom and bond identification); all atoms occurring in pairs are part of a ring or chain. Atoms occurring alone would be chain termini. This matching can establish the presence of rings and can also delete duplicates (circled and connected in Column 2) for print purposes.

If the coordinates in Column 2 of Chart I are subtracted from those in Column 1, we get a simplified notation which represents the topological difference between the two graphs, and hence the exact algorithm for their interconversion (Chart II). For example, in Chart I, if we subtract the first set of coordinates in Column 1 from the first set in Column 2,

$$\begin{array}{cc} x = 1 & y = 3 \\ \underline{-x = 2} & \underline{y = 2} \\ x = -1 & y = +1 \end{array}$$

we get $x = -1$ and $y = +1$, or the first coordinates in Column 2 of Chart II.

All alphanumeric characters have plus one added to their y-coordinates in order to achieve the correct positioning (cf. ref. 3 for a description of the relative levels of alphanumeric and structural elements).

Sorting on Column 2 by x and y yields Chart III which, after deletion of the duplicate members of the pairs (Chart IVa), represents (after addition of spacing and carriage returns) the line at a time output (Chart IVb) for the dot-bond typewriter or high-speed printer to give Figure 2, from the typewriter or high-speed printer.

APPENDIX 2

Appendix 2 gives US ASCII and BCD paper tape codes for reference purposes. It is important to remember, as we have recounted, that a 64-character tab card key punch cannot be directly compatible with an 88-character typewriter keyboard.

USA Standard Code for Information Interchange

1. Scope

This coded character set is to be used for the general interchange of information among information processing systems, communication systems, and associated equipment.

2. Standard Code

b_7 b_6 b_5					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b_4	b_3	b_2	b_1	COLUMN / ROW	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[	k	[illegible]
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	–	=	M	]	m	[illegible]
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US		?	O	___	o	DEL

Chart VII. ASCII seven-channel paper tape code

APPENDIX 3

Some typical conventions for typing chemical structures are listed in Chart VIII for Walter Reed Army Institute of Research[12] and for Chemical Abstracts.[13] The WRAIR system will also handle the options in the CA column, except for Ph and t-Bu.

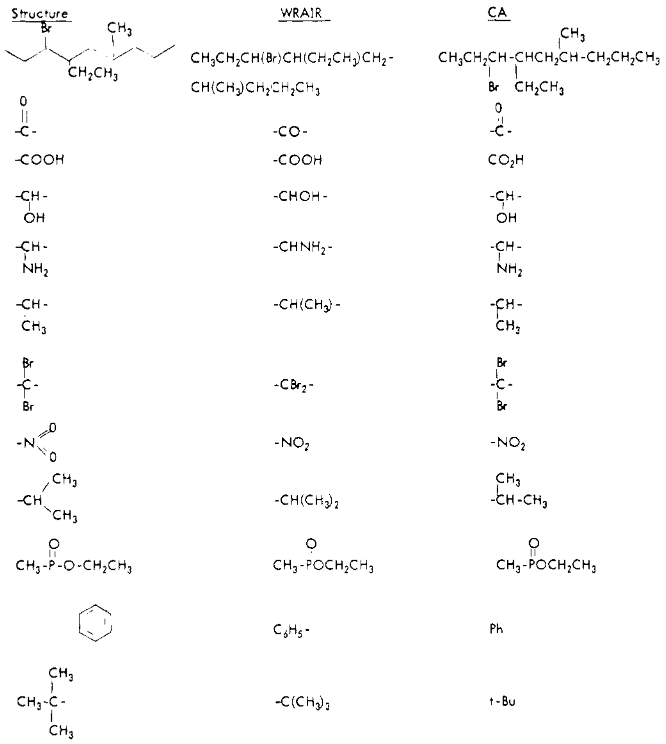


Chart VIII. Example of conventions for typing chemical structures

Among the problem areas of structure representation for computer input are the possible meanings of combinations like "CO" and "OC." There are two basic ways of handling this type of potential ambiguity. One is to set up a series of structure-drawing conventions at input, and the other is to apply rules relating to surrounding valence requirements. The actual methods chosen will depend on the circumstances surrounding the input process. In the ideal case, it is probable that the input conventions such as are noted in the above table will be applied, and, in addition, certain valence requirement checks on the input will be made by a computer edit routine.

ACKNOWLEDGMENT

We thank Aaron Addelston, C. M. Bowman, B. E. Holm, Stanley Kirschner, H. R. Koller, J. H. Kuney, H. K. Livingston, H. S. McDonald, J. M. Mullen, M. A. Paul, C. N. Rice, and E. G. Smith for helpful comments on this manuscript.

LITERATURE CITED

(1) Henderson, M. M., J. W. Moats, M. E. Stevens, and S. M. Newman, "Cooperation, Convertibility, and Compatibility Among Information Systems: A Literature Review," National Bureau of Standards, Misc. Pub. **276**, June 16, 1966.
(2) Sherwin, C. W., "A Standard Basis for a National Information System for Short Reports," U.S. Department of Commerce, February 24, 1967.
(3) Gordon, M., *Pharm. Ind.* **28**, 893 (1966).
(4) Mikhailov, A. I., *Rev. Int. Doc.* **29**, 49 (1962).
(5) Kuney, J. H., and B. G. Lazorchak, Fall Joint Computer Conference, San Francisco, Calif., Nov. 8–10, 1966; *Am. Doc.* **14**, 238 (1963); J. CHEM. DOC. **1**, 44 (1961).
(6) Lefkovitz, D., Report to National Science Foundation, March 15, 1967.
(7) Lefkovitz, D., J. CHEM. DOC. **7**, 186, 192 (1967).
(8) Hyde, E., F. W. Matthews, L. H. Thomson, and W. J. Wiswesser, J. CHEM. DOC. **7**, 200 (1967).
(9) Thomson, L. H., E. Hyde, and F. W. Matthews, J. CHEM. DOC. **7**, 204 (1967).
(10) Mullen, J. M., Abstracts, 150th Meeting, ACS, Atlantic City, N.J., September, 1965; J. CHEM. DOC. **7**, 88 (1967).
(11) Lederberg, J., Interim Report to NASA, March 13, 1968.
(12) Evans, C. J., Herner and Co., Instruction Manual for Chemical Typewriter Operators, March 6, 1967.
(13) "Structure Conventions," Chemical Abstracts Service, March 6, 1967.
(14) Bowman, C. M., F. A. Landee, and M. H. Reslock, J. CHEM. DOC. **7**, 43 (1967).
(15) DeMott, A., "Interpretation of Organic Chemical Formulae by Computer," p. 61, Spring Joint Computer Conference, 1968.

A Reverse Coordinate Concept System for Retrieving Engineering Site and Building Drawings*

ALAN R. McGARVEY
Research and Development Center, Armstrong Cork Co., Lancaster, Pa. 17604

Received October 30, 1968

A retrieval system is described for engineering and architectural drawings of buildings and the plot or site drawings showing all service lines. Using a "reverse coordinate concept," building drawings are located by specific bays, and site drawings by their physical location on a master grid or matrix with X and Y coordinates from 0-99. All drawings are indexed on cards bearing the address of a bay in a particular building or coordinates on the site grid. Approximately 5000 drawings are involved for 18 buildings on a 650-acre site.

One of the problems the engineer encounters in new construction, additions, or alterations to buildings or areas located in an industrial park is knowing what is already there. More than one company has been embarrassed by having a back-hoe operator pull out a water main that no one knew was there. Steam, water, gas, oil, sewer lines, and a myriad of other miscellaneous service lines may be buried under an area planned for excavation. Usually, large plot plans, often outdated, exist, showing each particular service line or combination of

*Presented at the 4th ACS Middle Atlantic Regional Meeting, Washington, D.C., February 14, 1969.