Vapor-Liquid Equilibrium Data Bibliography—A System of Computer Documentation

IVAN WICHTERLE
Institute of Chemical Process Fundamentals, Czechoslovak Academy of Science, Praha-Suchdol, Czechoslovakia

Received November 13, 1972

Automatic processing of the bibliography on vapor-liquid equilibrium data (4500 references by 1972, and 6000 references expected by 1975) is done by means of a specially developed mnemonic coding system. The formation of codes is simple and makes easy the preparation and checking of input data.

Distillation and absorption are among the most important separation processes in the chemical industry. The data on the phase equilibrium behavior of components to be separated belong to the large variety of data needed for the reliable design of distillation, absorption, etc. units. Even though there are several data compilations, many data still remain scattered through the chemical literature. A chemical engineer may find equilibrium data by use of the collections containing direct or recalculated data.[1,2] Besides these collections, there are many surveys, which, however, usually present only a bibliography. Furthermore, the compilations mentioned above cover the literature up to about 1965 and, in addition, they include the data sets which are available in tabular form. On the other hand, in the literature we often meet individual experimental points, or data presented in diagrams, or—still more frequently—the analytical expression by means of correlation equations. The heterogeneity of these data presentations makes completely impossible their collective compilation.

When a separation unit is to be designed, any equilibrium datum is better than the value provided by more or less reliable predicting methods. If direct data are absent, one appreciates the disclosing of a certain value or at least the azeotropic point of the mixture. For these reasons, we still cannot avoid a survey of the literature.

The completed bibliography (even without any experimental data) is required to make the surveying easier and quicker. Various indexing systems have been developed,[3,4] but the only system published as a part of Hála's monograph[5] is the most comprehensive. It consists of a cross-index to the reference for binary and multi-component systems. A new improved system lists, in addition, the reference numbers arranged according to year of publication. The equilibrium system searched can be found under any of its constituents. Then the list of references provides the full bibliographic identification.

The exponentially increasing number of references is evident (from the linear plot of logarithm of number of references *vs.* time) and it can be demonstrated on the consequent editions of Hála *et al.* The figures are given in the following table.

Date	Total Refs.	Cross-Index Pages Printed
APR 1954	750	30
FEB 1957	1050	65
MAY 1965	2350	180
JAN 1972	4400	945

Obviously, the sorting of all entries into the cross-index becomes more difficult and elaborate and, at the present time, automatic data processing of bibliographic information is necessary.

The idea of providing the survey with the additional information as for the kind of data (isobaric, isothermal, relative volatility, solubility, vapor pressure, dew and bubble point, diagrams, etc.) and condition ranges had to be rejected both owing to the inconsistency of sources and to the fact that the citations from unavailable journals were taken from *Chemical Abstracts*.

The input data for the computer processing consist of three independent files:

(1) Compound cards containing compound code, formula, name, and card serial number

(2) Author cards containing full bibliographic description, and card serial number

(3) Reference cards containing total number of components followed by compound codes and author card serial number(s)

These files are stored on magnetic tape. The individual sets can be sorted—e.g., in this particular case, the author

146

C2 H4 O	ETHYLENE OXIDE		
	C4 H10 O	BUTYL ALCOHOL	1299
	C4 H10 O	ISOBUTYL ALCOHOL	1299
	C12 H26 O	1-DODECANOL	2612
	C22 H46 O6	PENTAETHYLENE GLYCOL DODECYL ETHER	2612
	C25 H44 O6	PENTAETHYLENE GLYCOL (NONYLPHENYL) ETHER	2612
	H2 O	WATER	1835,1689, 687,2301
	C H N	HYDROGEN CYANIDE	186R
	C3 H5 N O	ETHYLENE CYANOHYDRIN	
	C H N	HYDROGEN CYANIDE	1927
	C4 H4 N2	1,2-DICYANOETHANE	
C2 H4 O2	ACETIC ACID		
	C CL4	CARBON TETRACHLORIDE	1762,2672,2375,2925, 2508,2228,2229
	C H CL3	CHLOROFORM	705, 533, 527
	C H2 CL2	DICHLOROMETHANE	877
	C H2 O2	FORMIC ACID	52, 87, 705, 86, 2106,1560
	C H4 O	METHYL ALCOHOL	3181,2300,3182
	C O2	CARBON DIOXIDE	2092
	C2 H CL3	TRICHLOROETHYLENE	3747
	C2 H4 CL2	1,2-DICHLOROETHANE	2812,1561
	C2 H4 O	ACETALDEHYDE	2865,2564,1145
	C2 H5 N O	ACETAMIDE	2812
	C2 H6 O	ETHYL ALCOHOL	2300,3182
	C3 H6	CYCLOPROPANE	210

Figure 1. Illustration of a Page in the Cross-Index Printout

Table I. List of Abbreviations

Abbreviation	Prefix	Stem or Termination	Abbreviation	Prefix	Stem or Termination
D	di-, bis-, D-		EX	ethoxy-	
T	tri-, trans-, tert-	ester	PX	propoxy-	
TE	tetra-		BX	butoxy-	
ME	methyl, meth-	methane	E	epi-, ep-	ether
ET	ethyl, eth-	ethane	NI	nitro-	
PR	propyl, prop-	propane	V	vinyl-	
BU	butyl, but-	butane	AL	allyl-	
PE	pentyl, penta-	pentane	BZY	benzyl-	
HX	hexyl, hexa-	hexane			
HP	heptyl, hepta-	heptane	OL		alcohol
OC	octyl, octa-	octane	YD		aldehyde
NO	nonyl, nona-	nonane	ON		ketone
DE	decyl, deca-	decane	IL		nitrile
F	fluoro-	fluoride	TH		thiol, mercaptane
C	chloro-	chloride	IN		hydrin
B	bromo-, beta	bromide	A	alpha-	amide
I	iodo-, iso-	iodide	G	gama-	glycol
AM	amino-, (amyl-)	amine	S	sec.-	sulfide
CN	cyano-	cyanide	IM		imine
FO	formic	form	SI		silane, silicon
AC	acetic	acet, acid	OX		oxide
AK	acrylic	acryl	RID		anhydride
BE	benzoic	benzo	EN		olefine
MY		methylene	YN		acetylene
EY		ethylene	DEN		diene
PY		propylene	DON		dione
BY		butylene	HZ		hydrazine
H	hydro-, hydroxy-		AN		aniline
U	un-		PYR		pyridine
DO	do-		BZ		benzene
CY	cyclo-		TO	tolyl	toluene
P	para-, poly-, per-		XY	xylyl	xylene
M	meta-		ST		styrene
O	ortho-		NA	naphthyl-	naphthalene
MX	methoxy-		CR	cresyl	cresol

file has been sorted alphabetically according to the following keys: first author surname, first initials, second surname, second initials, ..., year of publication, name of journal. The compound formula cards have been sorted according to the Hill system as used in the *Chemical Abstracts* Formula Indexes. The second key (in the case of identical formulas) is the name of the compound. Therefore, unimportant inconsistencies can be found—e.g., C_4 aliphatic alcohols are not listed next to each other, but mixed together with C_4 ethers. The file of the reference cards has been sorted in such a way as to obtain the cross-index described below.

It turns out that an important factor was the coding of compounds (especially organic). The original idea to convert the formula into certain numerical codes (e.g., with regard to the number of individual atoms) was abandoned for the absence of an easy survey and impracticability of checking. Then, the alphanumeric and mnemonic coding system was developed possessing the following features: (1) easy to check, (2) unambiguity of codes (there are only few exceptions), (3) easy to form the new codes, and (4) maximum six-character code.

The basic abbreviations can be roughly divided according to their meaning into two groups: (1) prefixes, and (2) stems and terminations, which are summarized in Table I. The formation of most codes is possible by means of abbreviations—e.g., MECYHX = methylcyclohexane, EYDAM = ethylenediamine, OCAN = orthochloroaniline, PEDON = pentadione, ETBET = ethyl benzoate, TFACAC = trifluoroacetic acid, PYOX = propylene oxide, DMEFOA = dimethylformamide, EYGETE = ethylene glycol ethyl ether, etc. Isomers and similar com-

2922. SCATCHARD,G.,SATKIEWICZ,F.G./J.AM.CHEM.SOC.86,130(1964).
2923. SCATCHARD,G.,TICKNOR,I.B./J.AM.CHEM.SOC.74,3724(1952).
2924. SCATCHARD,G.,WILSON,G.M./J.AM.CHEM.SOC.86,133(1964).
2925. SCATCHARD,G.,WOOD,S.E.,MOCHEL,J.M./J.AM.CHEM.SOC.61,3206(1939).
2926. SCATCHARD,G.,WOOD,S.E.,MOCHEL,J.M./J.CHEM.PHYS.43,119(1939).
2927. SCATCHARD,G.,WOOD,S.E.,MOCHEL,J.M./J.AM.CHEM.SOC.62,712(1940).
2928. SCATCHARD,G.,WOOD,S.E.,MOCHEL,J.M./J.AM.CHEM.SOC.68,1957(1946).
2929. SCATCHARD,G.,WOOD,S.E.,MOCHEL,J.M./J.AM.CHEM.SOC.68,1960(1946).
2930. SCHAEFER,K.,RALL,W./Z.ELEKTROCHEM.62,1090(1958).
2931. SCHAEFER,K.,RALL,W.,WIRTH-LINDEMANN,F.C./Z.PHYSIK.CHEM.(FRANKFURT) 14,197(1958).
2932. SCHAY,G.,SZEKELY,G.,RACZ,G.,TRAPLY,G./PERIOD.POLYTECH.CHEM.ENG.2(1),1(1958).
2933. SCHAY,G.,VARSANYI,G.,BILLES,F./PERIODICA POLYTECH.1,131(1957),ROCZNIKI CHEM.32,375(1958).
2934. SCHEELINE,H.W.,GILLILAND,E.R./IND.ENG.CHEM.31,1050(1939).
2935. SCHEFFER,F.E.C./PROC.KON.AKAD.WET.AMSTERDAM 15,380(1912).

Figure 2. Illustration of Part of a Page in the Author Index

pounds are distinguished by means of number(s)—e.g., MEPYR2 and MEPYR3 for 2-methylpyridine and 3-methylpyridine, respectively.

Of course, there are complex organic names yielding a more-than-six-character code. In these cases, a reasonable omission of several characters reduces the original abbreviated form—e.g., DEGDPE instead of DEYGDPRE = diethylene glycol dipropyl ether. Fortunately, such compounds do not occur among those often found in vapor-liquid equilibrium data. As far as inorganic compounds are concerned, the codes usually consist of formula.

Based on these mnemonic codes, reference input cards were prepared. A set of programs was prepared to check the punched cards to eliminate possible formal errors. Then, the cards were recorded on magnetic tape. The computer TESLA 200 (compatible with General Electric System) was applied for the automatic processing of the survey from 1900 through January 1972, listing 4500 literature citations on vapor-liquid equilibrium data for 1450 compounds. Examples of the cross- and author-indexes are shown in Figures 1 and 2, respectively.

In addition to the two indexes, the files offer alternative classifications, including data retrieval. The whole system can be supplemented easily and kept up-to-date.

LITERATURE CITED

(1) Chu, J. C., Wang, S. L., Levy, S. L., Paul, R., "Vapor-Liquid Equilibrium Data," Ann Arbor, Mich., 1966.
(2) Kogan, V. B., Fridman, V. M., Kafarov, V. V., "Ravnovesie mezhdu zhidkostiyu i parom" (Vapor-Liquid Equilibrium), Nauka, Moscow, 1966.
(3) Stage, H., Faldix, P., "Gleichgewichtsverhalten von Ein- und Mehrstoffsystemen," *Fortschr. Verfahrenstech.*, Verlag Chemie, 1952–1967.
(4) "Bulletin of Thermodynamics and Thermochemistry," Ann Arbor, Mich.
(5) Hála, E., Pick, J., Fried, V., Vilím, O., "Vapour-Liquid Equilibrium," Pergamon Press, New York, N. Y., 1967.

A Chemical Search System for a Small Computer*

DANIEL U. WILDE** and ALBERT C. STARKE
New England Research Application Center,
University of Connecticut, Storrs, Conn.

Received November 30, 1973

The mechanization of chemical information retrieval systems until now has been limited to those organizations that have access to extensive computer facilities. Now, small, low-cost computers, such as IBM's 1130 or DEC's PDP-11, are available with input/output capacities that make them suitable for SDI and retrospective searching on any of the many commercially available data bases. Such a machine, located at the New England Research Application Center (NERAC), is described and the problems of using it for chemical information retrieval are discussed. NERAC's SDI Chemical Search System is described, and an example profile is used to illustrate its capabilities.

During the past six years, the New England Research Application Center (NERAC) has operated as a NASA Research Dissemination Center at the University of Connecticut. The purpose of the Center is to aid and promote technology transfer in the Eastern United States by helping industry locate appropriate technical information. During this period, NERAC has performed some 5000 retrospective searches, while its data base has grown to over 2,000,000 documents. Presently, this data base includes six files: National Aeronautics and Space Administration (NASA), Department of Defense (DDC), U. S. Government Reports Abstracts (GRA), Education Resources Information Center (ERIC), American Society for Metals (METADEX), and World Aluminum Abstracts (WAA). Recently, searching of *CA Condensates* tapes of *Chemical Abstracts* (CA) has been added as a current awareness or Selective Dissemination of Information (SDI) service, thereby augmenting SDI services performed on each update of the previously cited data bases.

When NERAC began, there were two choices for computer power. First, time could be purchased on a large, extensive computer system; or second, NERAC could rent its own small machine. In the first case, although the amount of machine time used would be small, the organization would still be dependent upon the services and priorities of someone else. In addition, NERAC's computer cost would be directly proportional to use. In the second case, a small machine might be slower, but NERAC would be able to schedule it so as to meet its own requirements. Here, computer costs would be fixed, and increased use would result in decreased unit costs.

NERAC decided to rent its own machine. Currently, NERAC's computer system consists of an IBM 1130 with an 8k, 16-bit word memory with a 2.2 microsecond cycle time. Search questions are read via a 1442 card/read punch, and results are printed on a 1403 line printer. Data files are read from and intermediate results are saved on two 2401 tape drives. Here, the often used input drive has a transmission rate of 60,000 characters/second, and the less used output drive transmits at 30,000. This configuration rents for approximately $3700 per month and includes a small disk for storage of a monitor, user programs, and temporary data sets.

Random access searching on inverted files is out of the

* This research was sponsored in part by the National Aeronautics and Space Administration Contract NASW-2307 and 2516

** To whom correspondence should be addressed.