

# EFFECTIVENESS OF KWIC INDEX AS AN INFORMATION RETRIEVAL TECHNIQUE FOR SOCIAL SCIENCES\*

By YOSHIRO MATSUDA\*\*

AND

SACHIKO MATSUI\*\*\*

## *Introduction*

The literature search is one of the major aspects of the research activity and deserves the researchers' serious consideration and yet has attracted least attention except that of the documentalists and librarians. But recent increase of the publications and growth of the various means of information have brought up the mechanized information processing in literature search first in the field of natural sciences and gradually in the field of social sciences.<sup>1</sup>

Recent development of the computer has made a revolutionary change to the processing of the non-numerical information. The powerfulness is overlooked or not fully realized among the researchers in the field of social sciences; it is so even among the econometricians whose daily work heavily depends on computer and even more among the historians of economic and social thoughts whose hostility against computerized information retrieval is quite strong.

Our survey, though limited in scope, shows that several important trials of computerized information retrieval in action in the field of social sciences have been ventured mostly by the social scientists of particular disciplines without or with least collaboration of librarians.<sup>2</sup> Thus, the mechanized literature search is faced with the indifference both from users and librarians. This paper aims to fill the credibility gap between the researchers of documentation and researchers of social sciences through presenting the main findings

---

\* This work was carried out by the group conducted by Yoshiro Matsuda with the aids of the grants from Isetan Scholarship Foundation for 1970-71 and the Ministry of Education for 1972-73 (Experiments for comparison of effectiveness of retrieval by KWIC index and facet classification scheme using the titles and abstracts.) The authors wish to thank to all the participants to the project for their stimulating discussions and co-operation and for their generosity to use whole results and findings freely in this paper; Professor Jun'ichi Akiba of Hitotsubashi University, Assistant Professors Shigeyuki Sato of Hokkaido University and Ryo Suzuki of Saga University and Lecturers Akira Hirota of Hosei University and Kiyonari Kidahashi of Hokkai Gakuen University. Computer processing is executed at Computing Centre of Hokkaido University.

\*\* Assistant Professor (*Jokyōju*), Documentation Centre for Japanese Economic Statistics, Institute of Economic Research, Hitotsubashi University.

\*\*\* Assistant, Department of Management Sciences, Faculty of Commerce, Otaru University of Commerce.

<sup>1</sup> Janda [10] especially Part II; Research applications, Matsuda [14].

<sup>2</sup> More exactly without the initiative of traditional librarians or bibliographers. See the Table in Appendix 1.

of our trials published in the *KWIC index series for social sciences*,<sup>3</sup> comparing them with other similar trials in the field of social sciences.

Such advanced documentation activities undertaken in a specified discipline as abstract and / or index making have often encountered with suspicion against their usefulness. Such suspicion shares the following two suppositions.

- i) In the case of social sciences, literature search is so unique and peculiar to the researcher's personal frame of reference or value premises that any information processing planned for common users is quite unpractical for the researcher. And librarians are expected to remain as a neutral interpreter to writers and users. This "librarians' neutrality" is guaranteed so far as he depends on usual devices of information control such as classified catalog or subject catalogs in a traditional form.
- ii) It is for the current awareness services and not for the retrospective search which is indispensable for social sciences depending on the search of so huge pile of volumes stored from the past that the recently developed information retrieval techniques are most suitable.

The first proposition is less supportable judging from the various surveys which show that the librarians' classified and subject catalog themselves have come to be a jargon not only for the outsider researchers in other disciplines but also for the insiders like library science specialists or information officers.<sup>4</sup>

Severe attack comes from the researchers in the interdisciplinary sciences. For "the librarians' neutrality" is supported by the bias that the librarian's or indexer's judgement is itself neutral. But such neutrality can scarcely be supported; because the descriptors used by librarians, which are so to speak subject headings, require the writer to transform his intention into librarians' vocabulary and the user of these descriptors is in turn requested to transform his retrieving vocabulary into librarians' vocabulary. These double transformations will not guarantee the equivalence of the semantics of the writer's descriptors and user's descriptors and make it difficult for librarians to remain as a neutral interpreter or translator.

It is clear that in the case of classified catalog the same problem occurs more severely. It should be well remembered that Bliss classification group and CRG inspired by Ranganathan's theory of faceted classification have gained practical success in particular subject area.<sup>5</sup> This suggests that universal classification scheme is difficult to be designed and single purpose classification scheme is much more powerful and, thus it may be implicitly supposed that librarians' non-neutrality or the homogeneity among the writers, librarians or indexers and users. From this it will be well supposed that librarians' efforts could be flourished when they commit in a particular discipline or a frame of reference. It is this kind of reasoning that stimulates the direct descriptor making tendencies represented by the KWIC or KWOC indexes now in fashion.<sup>6</sup>

<sup>3</sup> See [B-6].

<sup>4</sup> Hans (Hanan) Wellisch; "Subject retrieval in the seventies—methods problems, prospects," in Wellisch [24] p. 16.

<sup>5</sup> T.D. Wilson; "The work of the British Classification Research Group," in Wellisch [24]. They still hold the possibility of compiling the effective universal classification scheme.

<sup>6</sup> As to the development of these techniques, see Fischer [7]. Another approach of compiling descriptors eliminating the indexer's bias, see Jean Aitchison; Thesaurofacet; a new concept in subject retrieval schemes," in Wellisch [24].

However the mechanical information processing is undertaken in most cases through the collaboration of the researchers of a particular discipline and documentalists or computer specialists, excluding the librarian's participation. These facts show that a philosophical change within library science is required even when it is applied to the field of social sciences. Thus we should examine the plausibility of the second proposition.

Table 1 shows the chronological development of KWIC index applied to information retrieval in social sciences and natural sciences. The time lag of the impact of computerization is not large between them but the stage of practical application comes so soon in the field of natural sciences while that in social sciences is staggering in the experimental stage.<sup>7</sup>

But, even from this short table, it is clear that the KWIC index compilation is more favoured in the interdisciplinary sciences like sociology, cultural anthropology, political science, econometrics and computer sciences or in the field where like legal processes immediate informations are required. Thus this kind of information retrieval techniques has a bright future in the interdisciplinary sciences. The most important point is that trials in the social sciences are mostly oriented to the retrospective research and not for current awareness service. The prosperity of the cumulative indexes may be its evidence and acts also as a counter example for the first proposition.

Needless to say there remains a supporting evidence for the second proposition. The past publication of mankind is so huge that even wealthy United States cannot afford enough financial support for her RECON project; to convert her national union catalogue into machine readable form. But there are two examples qualified enough to support the retrospective search in social sciences. One is Harvard University's Widener Library Shelflist Conversion and Publication Program resulted in the publication of the *Harvard Widener Library Shelflists* which are facsimile of computer outputs having author, title and chronological lists in addition to the main list (classified shelflist)<sup>8</sup>. This series promises us the future fertility of the retrospective research brought out through the further corporation of the important libraries sharing the MARC (machine readable cataloging) program.

The second example is Pollin's *Godwin Criticism* which covers about 3,379 items published in 1783-1966.<sup>9</sup> Although it has some defects which we will discuss later, it has broken a new path for the retrospective search by machine. And we might be permitted to register our work as the third example in this field. It deals with about 686 works of Claude Henri de Saint-Simon and Saint-Simonians published in 1817-1971.

## I. Natural Language or Descriptor

Here we will analyze the possibility of neutral transformation of the contents by virtue of the librarians' descriptors called Subject Headings (hereafter often abbreviated

<sup>7</sup> The cease of publication of *Jurindex* occurred in the second year of its short life is quite symbolic and it symbolizes the present situation of the mechanical processing in the field of social sciences. A few periodical publication of the mechanized indexes are established like the Annual supplements to the *Universal Reference System*; *Political Science, Government & Public Policy: an annotated and intensively indexed compilation of significant books, pamphlets, and selected and processed by the Universal Reference System—a computerized information retrieval service in the social and behavioral science*. 1965+

<sup>8</sup> De Gannaro [5].

<sup>9</sup> Pollin [B-5].

TABLE 1. CHRONOLOGY OF THE DEVELOPMENT OF KWIC INDEX (ABRIDGED)

	A. INDEX & ABSTRACTS	B. CUMULATIVE INDEX	C. SUBJECT BIBLIOGRAPHY	
			NATURAL SCIENCES	SOCIAL SCIENCES
1959			Luhn, H.P.; Keyword-in-context Index for Technical Literature. (KWIC index). RC-127. IBN.	
1960			MIT Libraries; Science Abstracts of China.	Feldman, Julian; Selected Articles on Methodology & Research in the Social Sciences. American Political Science Review. 1960. Sep.
1961	Chemical Titles. KWIC cf. Chemical Abstracts. 1907- (American Chemical Society) Biological Abstracts. 1926-, KWIC (Union of American Biological Societies)		IBM San Jose Research Laboratory; Dissertations in Physics. 1861-1959. Mimosa Frenk Foundation for Applied Neurochemistry & IBM; Index to Neurochemistry.	
1962	Index to Legal Theses and Research Project. 1952-, KWIC. (American Bar Foundation)			
1963	Current Legal Legislation Index, (American Bar Foundation)	Straus, Marry & Graham, Suasanne C.; Marriage and Family Living, Vol. 1-24. (1939-62)	Farley, Earl; Kansas Slavic Index.	
1964		Janda, Kenneth; American Political Science Review, Vol. 1-57. (1906-1963)		Bwy, Douglas; Social Conflict; a keyword-in-context-bibliography on the literature of developing areas, with supplementary reference from Latin America.
1966		Janda, Kenneth & Cutler, Neal E.; Midwest Journal of Political Science. Vol. 1-10. in Vol. 11, May issue.		Aldous, Joan, & Hill, Reuben; International Bibliography of Research in Marriage and the Family. 1900-1064.
1970	JURINDEX, l'informa- tique appliquée à l'actualité biblio- graphique du droit française. 1971-1972. KWIC (Institut de recherche d'infor- matique juridique.)	O'Barr, Mark & Jean; Current Anthropology. Vol. 1-24. (1959-1966)	Gotterer, M.H.; KWIC Index; Computer Management.	
1971				Matsuda, Yoshiro & Matsui, Sachiko; Bibliography; Works on econometrics, reviewed in Econometrica. 1951-1971.
1973				Matsuda, Yoshiro, Sato, Shigeyuki Hirota, Akira & Matsui, Sachiko; Bibliography; Works by Saint-Simon and Saint-Simonians. 1802-1970.

1974				<b>Matsuda, Yoshiro, Sato, Shigeyuki Hirota, Akira &amp; Matsui, Sachiko;</b> Bibliography; Works on Saint-Simon and Saint-Simonians. 1831-1970. Vol. I. <b>Matsuda, Yoshiro et al.</b> Bibliography; Works on Library and Information Science. 2 vols.
------	--	--	--	--

as SH). They are expected to be called neutral to both writers and users but in fact, should be regarded as to be subject to the librarians' frame of reference or paradigm. The most important paradigm of librarian to compile SH is "public use first, specialist use second". Thus users sometimes encounter with a serious gap between technical terms which now become of their common usage and SH terms using colloquial and sometimes unsophisticated words which are seldom used in researchers' discussion. This fact must have been stemmed from the belief that the paraphrasing by daily language will help the recognition of the public easier. This belief is severely criticized by John M. Christ who takes the SH of the Library of Congress of the United States as an example of the most popular SH.<sup>10</sup>

He compared the subject headings of social sciences in the 7th edition and the Supplements of the *Library of Congress Subject Headings* (hereafter abbreviated as LCSH) with

TABLE 2. THE DESCREPAncy OF THE USAGES OF TECHNICAL TERMS  
AMONG LIBRARY SCIENTISTS AND SOCIAL SCIENTISTS\*

	Main index headings of IESS**	Interdisciplinary terms***	Sociological terms****
The terms that match with LCSH	280 (42.5%)	24 (19.50%)	163 (34.5%)
The terms similar to LCSH	145 (21.5)	24 (18.50)	53 (11.5)
The terms found in LCSH See-Also reference	84 (12.5)	19 (14.50)	33 (7.5)
The terms not listed in LCSH	156 (23.5)	61 (47.5)	225 (46.5)
Total terms examined	665 (100%)	128 (100%)	470 (100%)

\* Figures are based on J.M. Christ's survey. See Christ [3], pp. 68, 83 & 123. A miscalculation of percentage is corrected. Comparison is based on the Subject Headings of Library of Congress (LCHS). 7th edition. [22].

\*\* IESS is *International Encyclopedia of the Social Sciences*, ed. by D. L. Sills. New York, Macmillan and Free Press, 1968.

\*\*\* Interdisciplinary terms are based on the terms used in the 13 basic interdisciplinary social science texts. See Christ [3], pp. 74-5. Whole list of these 128 terms are listed in table 5 of Christ [3], pp. 76-81.

\*\*\*\* Sociological terms here defined are extracted from the indexes of Broom Leonard & Philip Selznick's *Sociology; a text with adapted readings*, Evanston, Ill., 1955 (287 terms) and Robert Faris (ed.)'s *Handbook of Modern Sociology*, Chicago, 1964 (235 terms), sharing only 41 terms in common.

<sup>10</sup> Christ [3]. As to the criticism and defence to the LCSH, see Richard S. Angell; "Library of Congress Subject Headings—Review and Forecast," in Wellisch [24].

the index headings from the *International Encyclopedia of the Social Sciences*. His comparison was carried out as to i) general social science terms, ii) interdisciplinary social science terms and iii) sociology terms and other detailed comparison of particular concept like "value". His conclusion summarized in Table 2 was that LCSH is governed by different frame of reference from that of the social scientists. A quarter of general terms cannot be traced and in the case of sub-terms having the form of compound terms untraceable ones amount to 93.3%. The LCSH prefers single word heading contrary to the tendency among the scientists to prefer corporate words or collective terms.<sup>11</sup> This is shown in Table 3.

TABLE 3. CHARACTERISTICS OF THE SUBJECT HEADINGS SHOWN IN THE NUMBER OF WORDS IN THE SUBJECT HEADINGS

	Terms without subject heading referents			Terms with subject heading referents		
	Single word	Multi or corporate words	Total terms examined	Single word	Multi or corporate words	Total words examined
1) General social science terms in the main headings of IESS (partial)	14 (27%)	38 (73%)	52 (100%)	36 (87%)	15 (13%)	41 (100%)
2) Interdisciplinary social sciences terms (partial)	22 (49%)	24 (51%)	47 (100%)	18 (75%)	6 (25%)	24 (100%)
3) Sociological terms (partial)	12 (24%)	39 (76%)	51 (100%)	43 (91%)	4 (9%)	47 (100%)
4) Management sciences terms (LCSH before & in 7th ed.)				104 (21%)	454 (79%)	558 (100%)
5) Management sciences terms (after 1966 or Suppl. to 7th ed.)				0	77 (100%)	77 (100%)
6) Sub total of 4) & 5)				104 (20%)	521 (80%)	635 (100%)

Sources: From 1) to 3) figures are obtained from Christ [3] pp. 68, 69—70, 83—85, 121—3.

But some counting errors in the original tables are corrected.

From 4) to 5) figures are obtained from Matsui [15].

To interpret his finding we must take into consideration the following: that articles or index terms in Encyclopedia treat narrow concepts like articles in academic journals or in one chapter of a book but that books on which LCSH is assigned are apt to treat much broader concepts. But his findings are still striking enough. And this must be the background of many trials of thesaurus construction.

<sup>11</sup> LCSH is also slowly catching up this process. In the 5th edition of LCSH one thirds of the LCSH are single word and the 7th edition shows increase of corporate or multi words. See Daily [4] p. 3962. And the Table 3 shows that this tendency is still going on after the 7th edition.

FIG. 1. EXAMPLES OF KWIC INDEX

## 1. From [B-6, no. 1] (LP output facsimile)

KWIC column	Item number
Y OF THE FANCTION OF THEORY+PROBABILTY AND LAW IN INS UNIVERSITY STUDIES IN HISTORICAL AND POLITICAL TICAL STATISTICS AND PROBABILITY+VOLUME 3:PHYSICAL /MATHEMATICS IN THE SOCIAL TION GENERALE+L'APPORT DES INGENIEURS FRANCAIS AUX CS+IN HONOR OF JOHAN AKERMAN+(NO.29 IN LUND SOCIAL ICS+WITH APPLICATIONS IN THE SOCIAL AND MANAGEMENT D'ETUDE DES PROBLEMES /MODEL BUILDING IN THE HUMAN /METHODOLGY OF THE SOCIAL CHARACTIC PROCESSES WITH APPLICATIONS TO THE NATURAL RPHETATION OF RESEARCH+DOCUMENTATION IN THE SOCIAL /INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL /MATHEMATICAL METHODS IN THE SOCIAL AMMING+STANFORD MATHEMATICAL STUDIES IN THE SOCIAL AMMING+STANFORD MATHEMATICAL STUDIES IN THE SOCIAL /THE UNIVERSITY TEACHING OF SOCIAL TABLES+THEIR COMPILATION AND USE+PROCEEDINGS OF A /THE LOGIC OF ILITY AND LAW IN SCIENCE+.	2758 E36- 3- 91- / 2174 E24- 1- 90 2036 L 4- 1-1228 2806 E38- 2- 376 2788 E37- 1- 157 2046 E20- 4- 676 2567 L 4- 1-1302 2719 E33- 3- 252 2781 E37- 2- 364 2428 E29- 1- 100 2660 L 8- 5- 360 2372 E28- 1- 144 2852 L 4- 1- 441 2495 L 4- 1-1369 2540 E31- 3- 611 2413 L 8- 1- 83 2176 E24- 1- 92 2368 L13- 3- 109 2412 E28- 4- 925 2174 E24- 1- 90

## 2. From [B-6, no. 2]

ERELLE-PASSE+PRESENT+FUTURE+PAR PROSPER ENFANTIN E L'ORGANISATEUR+ET D'LETTRE D'UN DISCIPLE DE LA UTEUR+L'CEISE ET A OTI+TRAITE DE POLITIQUE ET DE VE+JOURNAL DE LA DOCTRINE DE SAINT-SIMON+RELIGION- SAINT-SI/LE PRODUCTEUR+JOURNAL DE L'INDUSTRIE+DES RODUCTEUR+JOURNAL PHILOSOPHIQUE DE L'INDUSTRIE+DES RODUCTEUR+JOURNAL PHILOSOPHIQUE DE L'INDUSTRIE+DES /INTRODUCTION AUX TRAVAUX TEME INDUSTRIEL(3EME PARTIE)+TRAVUX PHILOSOPHIQUES ERE+THOIS A/MORALE+REUNION GENERALE DE LA FAMILLE ONCE A LA CHAMBRE DES DEPUTES+PAR V.LAFITTE+A LA SCIENCE DE L'HOMME+PHYSIOLOGIE RELIGIEUSE+(OEUVRES.46).+ VIE ET SCIENCE NOUVELLE+AU REGIMINAIRES+PRETENDUS SAINT-SIMONIENS+D SCIENCE SOCIALE+ANSEE 1832+THEORIE DE CHARLES FOURIER+D SCIENCE SOCIALE+PUBLIE PAR LES EXECUTEURS TESTAMENTAIRES DE L'A SCIENCE+INDUSTRIE+7E ANNEE+NO 56--7E ANNEE+NO 151+(PERIODIQUE SCIENCE+INDUSTRIE+7E ANNEE+NO 152--7E ANNEE+NO 374+(PERIODIQUE SCIENCES ET DES BEAUX-ARTS+(LE PRODUCTEUR+TOME 1-2)+(PERIODIQUE SCIENCES ET DES BEAUX-ARTS+(LE PRODUCTEUR+TOME 3-4)+(PERIODIQUE SCIENCES ET DES BEAUX-ARTS+(UN SEUL CAHIER)+(LE PRODUCTEUR+TOME SCIENCES ET DES BEAUX-ARTS+(UN SEUL CAHIER)+(LE PRODUCTEUR+TOME SCIENTIFIQUES DU XIX-NEUVIEME SIECLE+(1807-1808).+ SCIENTIFIQUES ET POETIQUES AYANT POUR OBJET DE FACILITER LA REO SEANCE DES 19 ET 21 NOVEMBRE 1831+CINQ ENSEIGNEMENTS DE NOTRE P SEANCE DU 10 FEVRIER 1817+(L'INDUSTRIE+TOME 2).+ DU DISCOURS PR	3615 D 3912 T3807 3718 T3706 3644 T3427 3909 T3791-3 3910 T3791-4 3918 T3796-1 3919 T3796-2 3920 T3796-3 3535 S12 3547 S12 3683 T3683 3562 T3593
--	--



His findings suggest us to use another kind of descriptors for subject retrieval. One extreme way is to use natural language used in the text to be retrieved. "Natural language" used here means vocabulary uncontrolled by indexers or librarians.

The most practical system of information retrieval using natural language may be the KWIC index system formulated by H.P. Luhn in 1959.<sup>12</sup> As is well known KWIC means key-word-in-context index which uses title words as keywords instead of the thesaurus or SH except some stopwords preassigned and they are listed in alphabetical order in the middle of the page, accompanying the rest of the title after and before the keyword. See illustration in Fig. 1.

To compare the KWIC index with LCSH we should use the data base consisting of the bibliographic description of the documents in book form. According to our KWIC index to management sciences summarized in Table 4, the subject headings (LCSH) whose

TABLE 4. DEGREE OF CORRESPONDENCE OF TITLE WORDS TO THE SUBJECT  
HEADINGS OF LIBRARY OF CONGRESS

	Number of titles to which LCSH are assigned	Number of total LCSH assigned	Exact correspondence to title words	Partial correspondence to title words	No correspondence to title words
LCSH before & in 7th ed.	907	1642 (100%)	441 (27.1%)	677 (41.2%)	521 (31.7%)
LCSH after 1966 or suppl. to 7th ed.	225	147 (100%)	41 (27.9%)	61 (41.5%)	45 (30.6%)
Total	1132	1789 (100%)	485 (27.1%)	738 (41.3%)	566 (31.6%)

Total different LCSH examined are 558 in 7th ed. and 77 after 7th ed. Some adjustments are carried over in matching the LCSH and title words and they amounts to three percentage of the total subject headings examined.

Source: Matsui [15] pp. 295—298.

words and / or phrases share the title words in common are only 68.4%. And D.H. Kraft's similar survey in the field of legal periodicals shows it 64.4%.<sup>13</sup> Although V. Mostecky pointed out that the subject headings should not share the words in common with the title words in order to avoid the vagueness found in the titles.<sup>14</sup> But Christ's figures shown in Table 2 indicate that through the subject headings almost 54% of the technical terms get no direct access and they must have escaped from Mostecky's blame because they are more exact than the title words and belong to the researchers' common language. Though Christ's figures are not weighted with the frequency of occurrence in the published documents, they are quite close to those of ours and Kraft's. This similarity suggests that the natural language or title words are much closer to the technical terms than LCSH in the

<sup>12</sup> His pioneering article is now easy to access through the facsimile reprint in Schultz [21].

<sup>13</sup> Kraft [11] p. 50.

<sup>14</sup> Mostecky [18] p. 303.

social sciences and that the utilization of the title words instead of the subject headings like KWIC index makes it easier to get access to the technical terms.<sup>15</sup>

Table 5 shows that the descriptors assigned to one title is about 6.2 on average in case

TABLE 5. AVERAGE NUMBER OF DESCRIPTORS TO ONE TITLE

	Library and information science	Management sciences
Kinds of different LCSH	654	635
Total number of LCSH assigned	2010	1789
Number of titles assigned with LCSH	1194	1136
Percentage to total titles in the bibliography	66.9%	
Average number of LCSH to one title	1.5	1.57
Kinds of different keywords	2878	1511
Total number of keywords in KWIC	15055	7340
Number of titles in the bibliography	1785*	1189
Average number of keywords in one title	6.2	6.3

\* Number of titles indexed in KWIC index includes 616 articles in *Encyclopedia of Library and Information Science* (vol. 1—10). Thus total number indexed amounts to 2401.

of KWIC index and about 1.6 in case of LCSH. Thus the title words will bring out more access points than the usual subject heading approach.

As KWIC index requires no librarian's work to examine the content of the book, the above comparison of natural language and LCSH might lead to a conclusion that the KWIC or mechanized information processing is superior to the subject heading approach.

But we should take the following points into consideration; the descriptiveness of the contents by the title decreases with the sophistication due to the writer's literary taste. In this case some additional descriptors are required. This compensates the labour saving character of KWIC index making. Kraft's experiment mentioned above says that 10.5% titles need this kind of correction and in K. Janda's case on *American Political Science Review* it is 33%. Other experiment like Aldous & Hill's case also reports necessity of the correction of this kind, though no statistics is reported.<sup>16</sup>

Supplementary descriptors for KWIC index are required especially for the empirical research dealing with specific time and space elements. K. Janda's work on political science, Aldous & Hill's work on family and marriage and our work on econometrics will show a good example on this point.

In most cases, except ours, numerical figures are assigned as stopwords. The date in the title are excluded from the KWIC index and only verbal expression of time or period remains. However, the pronounciations of numbers and their spellings are so different even among European languages<sup>17</sup> that diversification of the same word is too wide to be used for retrieval.

<sup>15</sup> As to the application to the natural sciences, see Ruhe [20] and Rosenberg [19].

<sup>16</sup> See Kraft [11] p. 51 and Janda [10] p. 57. Also see Aldous [B-1] p. 5.

<sup>17</sup> Except this point merger of different languages causes no serious difficulty in the case of European languages including Slavic. In case of Japanese the situation is rather different. Transcription via Roman alphabet is not applicable for Japanese, Chinese and Korean to be treated simultaneously.

Examining these previous results, we excluded numerical figures from the stopwords.<sup>18</sup> In the case of econometrics the dates shown in the title are concentrated to those in twentieth century (about 40 titles) and the rest are 13 titles on the nineteenth century (including two covering eighteenth century) and two exceptional titles: one is Gossen's biography 1810-1858 and the other is British historical analysis of 1790-1859. To obtain the information of these 55 titles occupying 105 lines, we should scan 308 lines in the KWIC index. Similar results are obtained in the cases of management sciences and others shown in Table 6.

TABLE 6. THE ROLE OF NUMERICAL FIGURES IN THE TITLE

	Management Sciences	Econometrics	On Saint-Simon & Saint-Simonians
Total lines of numerical figures in KWIC index	303	307	98
Figures for the years and century	90	115	70
i) Description of the contents	78	6	70
ii) Century	1	0	7
iii) Date of the Conference, etc.	12	109	-
Others (editions, etc.)	212	192	28
Total titles retrieved by i) & ii)	9	55	40

In the case of LCSH no exact dates are given even for the documents on history. The analysis especially on the contemporary phenomena requires exact specification of date and the figures in the title or in the sub-title are quite effective for this specification. But for this purpose, the assignment of such a date in the title is not sufficient and we should add some supplementary descriptors in figures.

As to assigning the area specification, LCSH is superior to the direct use of the title words, because in most cases those works written for writer's own countrymen seldom use his country's name in the title. See Table 7.

Thus additional descriptors for area designation are required for KWIC index used by the users aiming international comparison. Maybe for those whose primary concern rests on his native country such additions are unnecessary. For example K. Janda assigned AMERICA, AMERICANS, US as stopwords. His presupposition might be that articles in *American Political Science Review* deal mainly with American affairs.<sup>19</sup> But when a bibliography becomes more international and extensive through inclusion of the articles or books written by foreigners, this kind of omission of the keywords becomes obstacle

<sup>18</sup> Our latest trial, library and information science seen in [B-6, no. 4], is requested to make a further improvement for its data base structure includes some figures useless for retrieving due to the inclusion of the contents of the documents like *Encyclopedia of Library and Information Science*. Thus the figures after V. for volume, P. for gapination, VOL. ED., NO., etc. are assigned as stopwords to be excluded from KWIC column. And figures like nineteenth century are transcribed into 19TH CENTURY. This technique is a variant of KWIC index. Similar trials are found in Matthews [17].

<sup>19</sup> He did not make any explicit reasoning on this point. But he did not exclude UNITED STATES, U.S. and U.S.A. from keywords. We suppose that he needs UNITED, and U. to get UNITED NATIONS and U.N. as keywords. As to the case of Aldous and Hill no information on stopwords is given as mentioned before, this prevent us to evaluate this bibliography exactly. It should be remembered for the compiler of KWIC indexes that the exposition of the list of stopwords is necessary not only to show the compilation process but also to make a thorough retrieval.

for effective retrieval.

Our tentative conclusion is that KWIC index supplemented with dates and places will be a powerful substitute for the hand-made index like LCSH. And it may act as a more powerful retrieval technique with these supplementation and it will rarely meet incongruence between the librarians' conceptualization and that by the social scientists.

TABLE 7. CO-OCCURRENCE OF NAMES OF PLACE AND PERSON AND TIME IN  
TITLE WORDS AND LCSH (MANAGEMENT SCIENCES)

LCSH \ Title		Found		Not found		Total	
		Total	Different	Total	Different	Total	Different
Place	Found	29	19				
	Not found	4	4	62	60	91	79
	Total	33	23				
Time	Found	—	—	3	1	3	1
	Not found	17	17				
	Total	17	17				
Person	Found	8	7	—	—	8	7
	Not found	8	8				
	Total	16	15				
Total	Found	37	26				
	Not found	29	29	65	61	102	87
	Total	66	55				

Source: Matsui [15] p. 298.

The second point that should be taken into consideration is that, when the vocabulary is uncontrolled, the natural language of the same meaning will be scattered throughout the whole KWIC column. For example, VALUE THEORY—PRICE THEORY, INTERINDUSTRY ANALYSIS—INPUT-OUTPUT ANALYSIS—LEONTIEF ANALYSIS, etc. This problem will become more serious in the case of merging the different languages like WIRTSCHAFT—OEKONOMIE—ECONOMIE, or HISTORY—HISTOIRE—ISTORIA—GESCHICHTE, etc.

But as far as European languages are concerned our experiment shows remarkable success because their differences in spelling concentrate to the ending part showing inflexion, etc. This, however, cannot be applicable to Romanized items in Chinese, Korean and Japanese which share common letters in part.

Needless to say, it is impossible to retrieve all the relevant documents by KWIC index without thesaurus or controlled vocabulary lists. This defects cannot be driven away by means of SH approach because of the lack in hierarchical categories in SH. Thus we should make reservation for the final evaluation of classification scheme approach.<sup>20</sup>

<sup>20</sup> See "Panel discussion," in Wellisch [24] and the future possibility of information science suggested in the articles of Elias [6].

## II. *Retrospective Search and Content Analysis through KWIC Index*

As mentioned above, interdisciplinary approach may become one of the most important aspects of our day and stimulate our interests in computerised indexing. And similar approach has occurred even in the field of analysis of social and economic thoughts. It is required to give the whole picture of one person who participated in several movements in the history of thoughts or one person whose activities and thoughts created a sort of a sect of social movements with international influence. Such a "whole picture" itself changes in accordance with the change of "paradigms." Thus the change of analysis itself becomes an object of research. These kinds of research work requires much more intensive bibliography than ever compiled. Recent close international contact among librarians and researchers has made it possible to compile such a bibliography in the world wide scope.

To take for example, Anthony P. Campanella's *Giuseppe Garibaldi e la Tradizione Garbaldina; una bibliografia dal 1807 al 1970* includes about 16,141 documents written in 54 languages and dialects and Burton R. Pollin's *Godwin Criticism; a synoptic bibliography* covering 1783-1966 includes about 3,374 documents in 14 languages.<sup>21</sup> The latter is compiled in machine readable form and provided with 12 computer made indexes and statistics tables. As far as we know, this may be the first trial of computerised bibliography designed for retrospective search for social sciences. But unfortunately this work sticks too much to the traditional form of a bibliography and does not realize the potential power of the machine readable form. Even in the field of humanities the computer processing of non-numerical information has come to be a strong weapon to compile a concordance or to identify the authorship of the anonymous or pseudonymous texts through quantification of the contents of the texts. Such a technique of analysis of non-numerical information may be applicable for the bibliography in machine readable form like *Godwin Criticism* equipped with short abstracts and many descriptors evaluating the degree of relevance to Godwin and / or his works. Abbreviations used for descriptor are A; article on Godwin, R; review of a book by Godwin, P; passage on Godwin, C; short comments, interspersed, M; mention of Godwin; B; book solely on Godwin, N; necrology of Godwin, Q; quoted material from Godwin, W; work dealing in part with Godwin and combinations of them.

The chronological lists of all entries is a mere listing of item number accompanied with these abbreviations. But a cross-tabulation with languages used or place of publication could have shown Godwin's international influence quantitatively. And chronological listing of quotations showing from what material of Godwin one had quoted would have shown what aspects of Godwin had attracted the public attention. This may be a variant of the citation index often used in natural sciences.

Pollin's treatment lacks a taxonomic point of view. Although his first version before publication classified all the material into three groups i) those written in the days when Godwin was working actively, ii) in Victorian period and iii) in modern times, he suppressed all these distinctions into two sections before and after 1836 because of the smallness of the

<sup>21</sup> See Campanella [B-2] and Pollin [B-5].

amounts published in Victorian period. But the waves of the number of publication are the reflection of the fashion of thoughts of the day. Relevant classification will reveal the change of thoughts more clearly. KWIC index applied to the titles of each period might suggest the difference of thoughts more clearly. For the selection of the title words are governed by the frame of reference of the original writer. And this frame of reference is difficult to escape the fashion of the day. Thus the analysis of the difference of the distribution of keywords will supply the quantitative basis for this comparison.<sup>22</sup>

One might say that these suggestions were beyond the scope of a bibliography. But such an intensive bibliography as Pollin's one itself provides one interpretation of Godwin's thought and influence and implications in Pollin's bibliography should be fully visualized through quantification of the bibliographical contents and abstracts.

We will show this kind of quantification through our experiments on Saint-Simon and Saint-Simonians' works. Before going further it may be allowed to make some digression on late Professor Juro Tedzuka whose endeavours, as a result, had provided the basis for our experiments. Although he is known as one of the most excellent pioneers of mathematical economics in Japan, his concern was not restricted to the mere refinement of mathematical treatment of economic models but extended to the philosophical implications underlying the economic theories and the possibility of the social reform from the side of libertarian favouring P.-J. Proudhon.<sup>23</sup> And he spent his stay in Paris from 1921 to 1926 in collecting books and brochures and copying unobtainable ones at Bibliothèque Nationale for this purpose. His collection covers from Pierre Bayle, Castel de Saint-Pierre, Encyclopedistes to contemporary radicals Andre Lorulot and Max Nettlau, especially strong in the nineteenth century visionaries like Saint-Simon, Fourier, Cabet, Considerant and their followers. After his death his whole collection was donated to Otaru University of Commerce by the courtesy of Miyakichi Itaya.

It is on this collection that our experiment on Saint-Simon and Saint-Simonians' works was undertaken. What he left was not mere books but he also left us one disciple, Mr. Kiyonari Kidahashi who compiled the first catalogue of Tedzuka collection and designed a hand-made KWOC (Key-Word-Out-of-Context) like index for the section *Philo-*

---

<sup>22</sup> One may easily find this way of investigation as a variant of a quantitative symbol analysis or, more generally, a content analysis. For example, see Richard L. Merritt, *Symbols of American Community, 1735-1775*. New Haven & London, Yale University Press, 1966.

If we extend our scope beyond the bibliography making, much suggestions will be obtained through the trials in the field of the computerized information processing found in the journals like *Computers and the Humanities* edited by Queens College since 1967.

<sup>23</sup> Tamotsu Matsuura describes him as Toshiro Tedzuka not as Juro Tedzuka in his "Marginalism in Japan" in *The Marginal Revolution in Economics; interpretation and evaluation* edited by R.D.C. Coats & C.D.W. Goodwin, Duke University Press, 1973. But we followed Tedzuka's own alphabetization used in his articles like a contribution to W.L. Valk, *The Principles of Wages* appeared as a Bibliography annexed by J. Tedzuka.

As to his life and works, see an anonymous article, "Recherche d'économie politique du professeur Tedzuka et son mérite," in *Bibliographie des Idées sociales et économiques en France; Catalogue de la bibliothèque Tedzuka de l'université d'Otaru*, Redacteur Taro Sakata, Tokyo, 1966.

*sophie et science naturelle* in late forties.<sup>24</sup> His trial found no supporter or collaborator and was isolated until we found a precursor of H.P. Luhn in him. And after twenty years interval this small brochure of about one hundred pages drove us to extend his intention through computerisation.

The importance of Saint-Simon and Saint-Simonians' role in the history of social and economic thought requires no additional explanation. Our data base is composed of the bibliographical description of books and pamphlets written by and on Saint-Simon and Saint-Simonians excluding articles in periodicals. The data base is subdivided into two files; one is for the books written by Saint-Simon and Saint-Simonians themselves (hereafter called BY-SS data base) and the other is for the works written about Saint-Simon and Saint-Simonians' works and their activities (hereafter called ON-SS data base.) BY-SS data base is not complete and we are preparing the second version through international cooperations. Thus the results obtained are rather tentative. ON-SS data base is based on three major information sources; i) 70 items through direct access to the books in Tedzuka Collection and 6 items in Léon Centnell Collection now in Otaru University of Commerce as Ohnishi-Tedzuka Memorial Library,<sup>25</sup> ii) 131 items depending on the description of the secondary information source, Walch's *Bibliographie du Saint-simonisme*, and iii) 30 items from various information sources including our private collections.

Indexes obtained from these data bases are i) KWIC index, ii) chronological index, iii) title index and iv) author index. Other information produced from these data bases is vi) matrix of publication place and date, vii) matrix of language and date and viii) various statistics based on these indexes, etc.

Table 8 is the distribution of publication date obtained from BY-SS ON-SS chronological indexes. This clearly shows the interaction of BY-SS and ON-SS. The upheaval of the sect of Saint-Simonians brought out a burst of publication in 1830s and their revivals are found in BY-SS data three times; at the end of the nineteenth century, after World War I and in 1960s. In most cases the waves of ON-SS data follows BY-SS data. First

<sup>24</sup> This is an index of 888 headings to 1993 titles for the Section; Philosophie et science naturelle of Tedzuka's collection. An facsimile example is shown below. See *Bibliographie de feu prof. Juro Tedzuka: Section Philosophie. Table analytique*, [rédigée par Kiyonari Kidahashi], x, 100 p. (copie carbonée), [n.d., Otaru].

Binet, Alfred (1857-1911).  
 --. -Martin, R. Paris, 1924. p1380.

Biologie.  
 -- générale./ Eléments de --. -Rabaud, E. 1920. p908.  
 --./ Les limites de la --. -Grasset, J. 1917. p555.  
 Biologique en Europe./ Le mouvement --. -Bohne, G. p153.  
 --./ La force et le droit le prétendu droit --.  
 -Antheny, R. Paris, 1917. p501.  
 -- et sociales./ Les mathématiques dans les sciences  
 --. -Volterra, V. 1906. p1725.  
 Biologiste sur l'objet, les méthodes et les limites  
 de la psychologie./ Réflexions d'un --. -Antheny, R. p57.

Blondel, Maurice (1861-1949).  
 --./ L'intelligence d'après --. -Maritain, J. p139.

<sup>25</sup> This collection was donated by the alumni to commemorate the sixtieth anniversary of this university. As to the detail of the collection, see Matsuda [13].

TABLE 8. NUMBER OF PUBLICATIONS CLASSIFIED BY DATE

Period	BY-SS	ON-SS
1802—1820	12	0
1821—1830	32	0
1831—1840	218*	14
1841—1850	19	10
1851—1860	8	7
1861—1870	65	13
1871—1880	37	10
1881—1890	0	10
1891—1900	0	14
1901—1910	0	24
1911—1920	2	11
1921—1930	3	28
1931—1940	0	23
1941—1950	2	12
1951—1960	0	21
1961—1970	16	15
1971	1	2

\* This figure includes many brochures whose publication dates are assigned to our estimation mainly based on Fournel's *Bibliographie saint-simonienne, de 1802 du 31 décembre 1832. 1833.*

upheavals in 1830-50 is the result of the controversies and interests of the contemporaries of Saint-Simonians. The second upheaval at the end of nineteenth century was brought out by the publication of *Œuvres de Saint-Simon & d'Enfantin* which made possible to utilize many unpublished documents. It should be noticed that the KWIC index shows the increase of SOCIALISME, FEMINISME, POSITIVISME. This may be the change of frame of reference of this period. The second increase in 1920s may relate to Neo-Saint-simonism and the Russian Revolution. Increase of monographs in English in 1940s may reflect the totalitarian political movements of the day. These changes of interests of the time are seen in the KWIC index, and the difference of frame of references between Saint-Simonians and researchers is also clearly observed in the KWIC index. For example, Saint-Simonians preferred the words, ORGANISATION, ASSOCIATION, RELIGION etc. to SOCIALISME, UTOPIE etc.<sup>26</sup>

Further application of content analysis requires much more elaborate data base which at least includes additional descriptors or abstracts. Another limitation of this data base is its exclusion of other synchronic social reformers' works, which are necessary to follow their mutual influences. Tedzuka Collection and Ohnishi-Tedzuka Memorial Library are qualified enough to provide materials to compile a data base suitable for this purpose. But as our initial concern was limited to estimate the effectiveness of the automatised indexing and abstracting processes, such kind of elaboration of data base is postponed to the next stage of our experiment.

Some words should be added as to the effectiveness of a retrieval technique of this kind. In the previous section we made a reservation for the final evaluation of classification scheme approach in case of interdisciplinary approach. But in the case of

<sup>26</sup> See also Matsuda [13] and Sato's introduction to Matsuda [B-6, no. 3].

retrospective search the superiority of mechanized information retrieval over classification scheme approach is quite clear. Evolution or expansion of classification scheme is itself a result of the division of research disciplines and change of the frames of reference. This means the applicability of older classification scheme to older documents which will be retrieved in retrospective search. Unfortunately change of frames of reference differs from topics to topics. Thus the mechanical conversion or application of the classification scheme has no practical meaning.

### III. *Association of Keywords and SDI System*

Along with KWIC index system, H.P. Luhn had developed the idea of SDI (Selective Dissemination of Information) System where newly arrived data base will be screened at once through a compound list of the descriptors or keywords which each user had prescribed in order to avoid repeated scanning of the same machine readable data many times for each user. The practical applicability of SDI to social sciences might be severely limited because of the lack in massive machine readable data. But even if such data were available, there would remain another factor which will suppress the effectiveness of SDI for the social sciences. For, exact meaning of the categories used in the social sciences often depends on the context and exact specification of descriptors is quite difficult.

The same factor will, in turn, favour the KWIC index. The effectiveness of KWIC index depends on the degree of the context-dependency of the category shown by the keyword. Exact meaning of a keyword will be brought out by the association with other words surrounding the keyword. Thus the effectiveness of KWIC index and similar techniques cannot be determined uniquely. It depends on the user's insight. This fact is applied to another aspect of a KWIC index retrieval process. The effectiveness will be increased with the association of keywords<sup>27</sup> through glancing the column of keywords.

Some might say that this kind of arbitrariness is a defect of the system. But this arbitrariness should be regarded as the flexibility of the system which can create the room for learning process in retrieving process in action. For the rigid SDI system provides new information aspect only through assigning new descriptors and so learning process at the retrieving process occurs within a limit of user's memory and documents retrieved previously. On the other hand, glancing the column of keywords will supply a new frame of association of categories which might be overlooked by the researcher in the case of SDI system. Our hypothesis underlying this way of thinking is that it is impossible for the searcher of information to grasp the whole idea of what he wants to retrieve. This might sound ridiculous but is much closer to the reality. Thus the effectiveness of any retrieval technique still rests on the skill or insight of the user.

Another limitation to the information retrieval system, however, comes from the extent of the exhaustiveness of the documents to be retrieved and the degree of additions of descriptors or abstracts which are utilized for retrieval. Like production of goods, no input creates no output but inappropriate input creates no output either. Similarly no

<sup>27</sup> The association here defined is not the same as to the association found in Lauren B. Doyle's "Indexing and abstracting by association" in Elias [6] but may correspond to the abduction process of perception.

## APPENDIX 1. MAIN FEATURES OF THE

Reference no.	1 [B-4]	2 [B-1]	3 [B-5]
Organizer	Kenneth Janda	Joan Aldous & R. Hill	Burton R. Pollin
Publication	1964	1967	1967
Subject field	Political science	Marriage & family	On William Godwin
Coverage of publication date	1906-1963	1900-1964	1783-1966
Coverage of publication form; A, articles, B, books.	A	A,B	A,B
Source materials; D, Direct access or primary sources, S, secondary sources	D	S	D,S
Selection criteria; M, mechanical, S, selective	M*	S*	S
Treatment of languages other than English	not applicabel	English translation	Original language in Roman alphabet transcription with English translation
Items covered	2,614	12,850	3,379
Data base compiler	K. Janda and others	J. Aldous & R. Hill with international corporation	B.R. Pollin with international corporation
Notes	*All items in the <i>American Political Science Review</i> .	*Selection aims to select all relevant materials based on the various reference books, and not to exclude unimportant materials. This procedure is based on the efforts of the Minnesota Inventory of Published Research on Marriage & Family Behavior since 1955.	

## BIBLIOGRAPHIES DISCUSSED IN THE TEXT

4 [B-3]	5 (not yet published)	6 [B-6, no. 1]	7 [B-6, no. 33]	8 [B-6, no. 4]
Malcom H. Gotterer	Yoshiro Matsuda	Yoshiro Matsuda	Yoshiro Matsuda	Yoshiro Matsuda
1970	(1970)	1973	1973-	1974
Computer management	Management sciences	Econometrics	Saint-Simon & Saint-Simonians	Library & information science
1955-1969		1947-1969	1817-1971 (texts) 1831-1970 (criticism)	1786-1973
A,B	B	B	B*	B
n.a.	D	S	D,S	D
S	M*	M*	S	S*
not applicable	Original languages in Roman alphabet transcription	Original language in Roman alphabet transcription with English translation	Original language in Roman alphabet transcription	Original language in Roman alphabet transcription
900 above	1,136	864	455 (texts), 231 (criticism)	1785
M.H. Gotterer	Sachiko Matsui	Y. Matsuda & S. Matsui	Shigeyuki Sato, Akira, Hirota, Kiyonari Kidahashi, & Y. Matsuda	S. Matsui, Kiyonari Kidahashi & Y. Matsuda
	*All items purchased by the Department of Management Sciences. Selection for purchase was done by teaching staffs of the department.	*All items reviewed in Econometrica. Editor of this review section was Gerhard Tintner.	*Books contain some offprints from journals. As to the criticism the articles in journals will be issued.	*All items held in the public and university libraries in Hokkaido District. Selection was carried out by the corporation of the Hokkaido Librarians' Study Circle.

Reference no.	1	2	3
Master list	**Author alphabetized bibliography (listing by alphabetized authors)	**Complete reference list (listing by alphabetized authors)	**Synoptic bibliography (classified by publishing dates before and after 1837, books or periodicals)
Indexes			
Name approach			
Author	*Author-cross reference (author and coauthor)	*Author list (author and coauthor)	*Authors of books and articles
Persons mentioned	-	-	*Persons mentioned (selective)
Title	-	-	-
Titles cited	-	-	*Books mentioned (selective)
Periodical title	+	(*)Periodicals list (not index)	*(Combined with the synoptic bib. listed by alphabetized periodicals)
Subject approach			
Natural language	**Keyword listing (i.e. KWIC index)	**KWIC index	-
Descriptors	-	-	*(Publication form)
Classified scheme	-	Subject index (i.e. classified index)	-
Evaluation approach	-	-	** (Abbreviations & short abstracts)
Chronological approach	-	- (Only statistics)	*Chronological listing of all entries, Writings of Godwin
Language approach	-	- (Only statistics)	*Languages of original sources
Number of indexes	3	5	11
Number of statistics	0	5	3
Computational work			
Number of data cards	n.a.	50,000 above	20,000
Format of masterlist	B-5, 90 lines/1 column	A-4, 107 lines/2 columns	B-5, 58 lines/1 column
Pages (catalogue, illustrations)	(225,12)	(225,12)	(659, 46)
Computing facility	Northwestern Univ. Computing Center	Univ. of Minnesota Data Processing Center	New York Univ. The Institute for Computer Research in the Humanities
Computer Programming language	IBM 709	IBM 1401	IBM 360-30
Programmer	MAP, FORTRAN James S. Aagaard	n.a. Ralph J. Willard, Ernst Duffing	PL/I, FORTRAN George W. Logmann

*Notes:*

- \* Index indicating the item number.
- \*\* Index combined with short bibliographical description.
- Not tried.
- + Not applicable.

4	5	6	7	8
<b>**Bibliography</b>	<b>**Master list (classified)</b>	<b>**Master list</b>	<b>**Master list</b>	<b>**Main bibliography; classified</b>
<b>*Author index (author, coauthor)</b> - - - -	<b>*Author index (author, coauthor)</b> - <b>**Title index</b> - -	<b>**Author index (author, coauthor)</b> - <b>**Title index</b> - +	<b>**Author index (author, coauthor)</b> - <b>**Title index</b> - +	<b>**Author index (author, coauthor, corporate author)</b> - <b>**Title index</b> - +
<b>**KWIC index</b> - -	<b>**KWIC index</b> <b>**LCSH index (subject headings)</b> <b>** (Combined with master list)</b>	<b>**KWIC index</b> - -	<b>**KWIC index</b> - <b>** (Texts; publication form), (Criticism)</b>	<b>**KWIC index</b> <b>**LCSH index (subject headings)</b> <b>** (Combined with main bibliography)</b>
- - -	- <b>**Chronological index</b>  <b>*Key to the languages other than English</b>	<b>**Reviewer index</b> <b>**Chronological index</b> -	- <b>**Chronological index</b>  <b>*Key to language other than French (texts)</b>	<b>** (Source book reference)</b> <b>**Chronological index</b>  <b>*Key to languages other than English</b>
3 0	6 0	6 0	6 0	8 56
n.a. B-5, 60 lines/1 column (152,2)	n.a. A-4, 100 lines/2 column	8,520 A-4, 100 lines/2 columns (145,32)      A-4, 100 lines/2 columns (92,32; texts), (49,32 criticism)		A-4, 104 lines/2 columns (392, 19, and statistics)
Johns Hopkins Univ.	Hokkaido University Computing Center	Hokkaido University Computing Center		Hokkaido Univ. Computing Center & Kodak Micro-System Service Center
n.a. n.a. Donald Derenge	FACOM 230-60 PL/I Sachiko Matsui	FACOM 230-60 PL/I Sachiko Matsui		FACOM 230-60, KOM-90 Sachiko Matsui

information will be obtained by inappropriate data base. Thus the real problem is how to compile data base for a bibliography or catalogue.

Apart from this problem of adequacy, the exhaustiveness or thoroughness of coverage requires a close contact with various institutions in international scope in the case of the social science bibliography which, more or less, contains some historical and international aspects. This requirement will be fulfilled by the process similar to compiling a subject union catalogue. Unlike the national or aggregated union catalogue, subject union catalogue requires selection from the whole holdings of institutions participating in the union catalogue making. In this case, not a card catalogue but a check list equipped with many indexes will decrease the burden of librarians to cooperate with the program, because a book form catalogue is easy to scan. Thus, not a mere listing of the items to be searched but a bibliography fully equipped with indexes like a final draft is necessary as a check list for the holding survey. For this purpose, the machine readable cataloguing technique is the most powerful weapon for editing. For, once several indexes are prepared, often changes of the item numbers causes serious difficulty in the case of hand made catalogue. So in most cases tentative drafts are mere listing of whole items to avoid the overall re-numbering which will be caused by the adjustment of the item numbers for the new additions or deleting. But, by virtue of the computer processing, this kind of adjustment comes to be a fairly easy task and so the circulation of the drafts of bibliography is possible for any version of the drafts. While it is not quite easy for computer processing to prepare the drafts which can be circulated several times with ease, recent development of COM (computer output microfilm) device provides a good means of producing a circulating version.

Thus our final remark is that the future of the computerised bibliography making depends on the piling up of the trials of compilation of such bibliography in practice. We hope this short essay will be a stimulus for the future trials.

## APPENDIX 2. DATA BASE AND SOME TECHNICAL NOTES

The data base for [B-6] was first restricted by the compiler of PL / 1 of FACOM 230-60. The area of characters is within 256 letters, and so the title of the document is restricted within 256 letters including blanks between words. The input format is fixed field and one record is 1,000 letters in case of [B-6, no. 1-no. 3]. Extension to 1,400 letters is executed for [no. 4] to utilize note area.

The master file format is shown in the Figure 2, below. As this is a final print out format, the end of the line is adjusted by the hair space technique. Fig. 3 shows the example from [B-6, no 4] which doubly prints out the titles having articles at the top as omitting article case and including article case. As to the detail of the programming techniques, see Matsui [16].

## FIG. 2. DATA BASE

Management Sciences (not yet published, Ref. no. 5 in Appendix 1)

507 E

K 8- 1- 70

TARSKI, ALFRED

INTRODUCTION TO LOGIC AND TO THE METHODOLOGY OF  
DEDUCTIVE SCIENCES. =

1941 NEW YORK

LCSH1 MATHEMATICS--PHILOSOPHY

LCSH2 ARITHMETIC--FOUNDATIONS

[B-6, no. 1]

2323 J

OHKAWA, KAZUSHI.

NOGYO NO DOTAI BUNSEKI (DYNAMIC ANALYSIS OF  
AGRICULTURE). =

1954 TOKYO: JOSUI SHCBO.

(F-MFTR 27- 3- 497 REV. BY NAKAMURA, JAMES I.)

2324 F

ALLAIS, MAURICE.

FONDEMENTS D'UNE THEORIE POSITIVE DES CHOIX COMPORTANT  
UN RISQUE ET CRITIQUE DES POSTULATS ET AXIOMES DE  
L'ECONOMIE AMERICAINE. =

1955 53P PARIS: IMPRIMERIE NATIONALE.

(F-MFTR 27- 3- 496 REV. BY SUPPES, PATRICK.)

2325 E

L 1- 1- 427

MEADE, J. E.

TRADE AND WELFARE. (WITH MATHEMATICAL SUPPLEMENT) =

1955 616P LONDON: OXFORD UNIVERSITY PRESS.

(E-MFTR 27- 3- 500 REV. BY HARBURG, ARNOLD C.)

[B-6, no. 4]

0039 J CL: 001C

HOKKAIDO, MUSASHI JOSHI TANKI DAIGAKU. TOSHOKAN.

TOSHOKAN JOHOGAKU SHIRYO MOKUROKU. --1972 NEN 3 GATSU

GENZAI--"

1972 SAPPORO: THE LIBRARY. 77P.

LOC.: OTARU U.CM MUSASHI JC

--- 002 ENCYCLOPEDIAS, DICTIONARIES,  
GLOSSARIES, HANDBOOKS AND MANUALS ---

&lt;ENCYCLOPEDIAS&gt;

0040 E CL: 002D

D 11- 3- 143

KENT, ALLEN (ED)

LANCOUR, HAROLD (ED)

NASRI, W. Z. (AS ED)

ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE. VOL.

1: A TO ASSOCIAC. =

1968 NEW YORK: MARCEL DEKKER. 12, 676P.

LCSH1 LIBRARY SCIENCE--DICTIONARIES

LCSH2 INFORMATION SCIENCE--DICTIONARIES

LOC.: OTARU U.CM HU(CTR) HU(ENG) HU(EDC)

HU(AGR) HU(SAP) KITAMI I.T SAP.MED.C.

HOKKAIGU HOKUSEI GU SAPPORO U. MUSASHI JC

SAPPORO JC HOK.PRE.LB UC/KANAGAW SB/BELZER

&lt;01&gt; ABBREVIATIONS (ELIS V.1 P.1-12)

(STERNBERG, VIRGINIA)

&lt;02&gt; ABSTRACT CLASSIFICATION (ELIS V.1 P.12-16)

(DAILY, JAY E.)

&lt;03&gt; ABSTRACTS AND ABSTRACTING (ELIS V.1 P.16-38)

(BERNIER, CHARLES L.)

&lt;04&gt; ACADEMIC STATUS OF LIBRARIANS AND INFORMATION

SCIENTISTS (ELIS V.1 P.32-43)

(KHURSHID, ANIS +)

## [B-6, no. 2]

- 3698 F S 348  
FOURNEL, HENRI J.-M.,  
BIBLIOGRAPHIE SAINT-SIMONIENNE DE 1802 AU 31 DECEMBRE  
1832... 1833 130P PARIS.  
(T 3692)
- 3699 F E 669  
FOURNEL, HENRI J.-M.,  
DU CHEMIN DE FER DU HAVRE A MARSEILLE, PAR LA VALLEE DE  
LA MARNE. PREMIERE PUBLICATION...  
1833 31P PARIS.  
(T 3693)
- 3700 F S 317  
FOURNEL, HENRI.  
REPONSE A UNE BROCHURE INTITULEE "QUESTIONS SUR LE  
DROIT D'HEREDITE, SPECIALEMENT CONSIDERE DANS LA  
MONARCHIE ET DANS LA PAIRIE, PAR M. LE BARON MASSIAS."  
(EXTRAITS DU "GLOBE")...  
1831 30P PARIS.  
(T 3694)

## [B-6, no. 3]

- 3016 F S 1222  
FIDAO-JUSTIANT, J.-L.,  
LE DROIT DES HUMILES, ETUDES DE POLITIQUE SOCIALE: LA  
POLITIQUE SOCIALE, LES PROPHETES ET LA LOI, SAINT-SIMON.  
SAINT-SIMONIENS, LA PENSEE POLITIQUE DE LAMARTINE.  
AUGUSTE COMTE ET LA SYNTHESE SOCIOLOGIQUE, J.-B. BUCHEZ.  
PAR J.-L. FIDAO...  
1904 360P PARIS  
(T 1398)
- 3017 F S 285  
GARNIER, H.  
INVOCATION A SAINT-SIMON, ODE DEDIEE AUX MEMBRES DU  
COLLEGE SAINT-SIMONIFN...  
1831 8P PARIS  
(T 1528)

FIG. 3. TITLE INDEX

- 1965 LA PRESSE EN FRANCE: GENESE ET EVOLUTION DE SES FONCTIONS PSYCHOSOCIALES. (UNIVER... 0362  
1960 LA PRESSE MOERSE. (QUE SAIS-JE? 414)... 0360  
1948 A LABORATORY COURSE IN READING AND WRITING... (J) 0335  
1964 LANGUAGE AND INFORMATION: SELECTED ESSAYS ON THEIR THEORY AND APPLICATION. (ADINES... 1503  
1962 THE LANGUAGE OF THE FOREIGN BOOK TRADE, ABBREVIATIONS, TERMS, PHRASES... 6074  
1963 THE LARGER UNITS OF PUBLIC LIBRARY SERVICE IN CANADA. WITH PARTICULAR REFERENCE TO THE... 0406  
1941 LAWS AFFECTING SCHOOL LIBRARIES. (U. S. OFFICE OF EDUCATION, BULLETIN 1940, NO. 7)... 0726  
1972 LE LIVRE... 6303  
1923 LE LIVRE: HISTORIQUE, FABRICATION, ACHAT, CLASSEMENT, USAGE ET ENTRETIEN... 0355  
1920 LEARNING FROM REFERENCE BOOKS. (COMMONWEALTH AND INTERNATIONAL LIBRARY)... 0670  
1905 LEARNING FOR LIBRARIES... 1288  
1903 LEHRNACH DER BIBLIOTHEKSVERWALTUNG... (J) 0802  
1923 LENDING LIBRARY METHODS... 0422  
1963 LES BIBLIOTHEQUES. (QUE SAIS-JE? 944)... (J) 0196  
1970 LES CATALOGUES DU DEPARTEMENT DES IMPRIMERIES DE LA BIBLIOTHEQUE NATIONALE... 0668  
1963 LES INSECTES ENNEMIS DES LIVRES: LEURS MOEURS, MOYENS DE LES DETRUIRE... 1074  
1966 LES TRANSFORMATIONS DE LA BIBLIOTHEQUE NATIONALE ET LE DEPOT ANNEXE DE VERSAILLES... 0149  
1964 LET'S BUILD A BOOK... (J) 1032  
1962 LEXIKON DES BUCHHESEN... 0056  
1952 LEXIKON DES GESAMTEN BUCHHESEN. BD. 1. AA-GOETTERNA... (J) 0052  
1963 THE LIBRARIAN AND THE MACHINE: OBSERVATIONS ON THE APPLICATIONS OF MACHINES IN ADMINIS... 1393  
1963 THE LIBRARIAN AND THE TEACHER IN GENERAL EDUCATION: A REPORT OF LIBRARY-INSTRUCTIONAL ... 0694  
1964 THE LIBRARIAN AND THE TEACHER IN ENGLISH. (EXPERIMENTING TOGETHER)... 0666  
1964 THE LIBRARIAN'S GLOSSARY OF TERMS USED IN LIBRARIANSHIP AND THE BOOK CRAFTS AND REFERENCE... 0663  
1971 THE LIBRARIAN'S GLOSSARY OF TERMS USED IN LIBRARIANSHIP AND THE BOOK CRAFTS. 2. ED. REV... 0244  
1969 THE LIBRARIANSHIP AT A GLANCE. SCHOOL OF LIBRARIANSHIP PROSPECTUS, UNIVERSITY OF CMPE... 0236  
1970 THE LIBRARIANSHIP AT A GLANCE. WILFELD L. (ED)... 0129

## REFERENCES

## I. Bibliography cited.

- [B-1] Joan Aldous & Reuben Hill, *International Bibliography of Research in Marriage and the Family, 1900-1964*. Minneapolis, University of Minnesota, 1967. vi, 508 p.
- [B-2] Anthony P. Campanella, *Giuseppe Garibaldi e la Tradizione Garibaldina, una bibliografia dal 1807 al 1970*. Ginevra, Comitato Dell'Istituto Internazionale di Studi Garibaldini. 1971. 2 vols.
- [B-3] Malcolm H. Gotterer, *KWIC Index, A Bibliography for Computer Management*. Princeton / Brandon, System Press, 1970. viii, 152 p.
- [B-4] Kenneth Janda (ed.), *Cummulative Index to the American Political Science Review, volumes 1-57; 1906-1963*. Evanston, Northwestern University Press, 1964. xxiii, 225 p.
- [B-5] Burton R. Pollin, *Godwin Criticism, A Synoptic Bibliography*. Toronto, University of Toronto Press, 1967. xlv, 659 p.
- [B-6] Yoshiro Matsuda (ed.); *KWIC Index Series for Social Sciences*.
  - No. 1. *Bibliography; Works on Econometrics*—reviewed in *Econometrica*, 1951-1957—(Introduction by Yoshiro Matsuda & Sachiko Matsui), Otaru, 1973. vii, 24, 145 p.
  - No. 2. *Bibliography; Works by Saint-Simon and Saint-Simonians*. (Introduction by Akira Hirota and others) (to be published)
  - No. 3. *Bibliography; Works on Saint-Simon and Saint-Simonians.—1831-1971*—(Introduction by Shigeyuki Sato & Sachiko Matsui) Otaru, 1973. viii, 24, 49 p.
  - No. 4. *Bibliography; Works on Library and Information Science, based on the holdings of the libraries in Hokkaido region*. Compiled by RINET (Regional Information Network) Society with the joint auspices of Hokkaido Librarians' Study Circle. Otaru, 1974. 2 vols.

## II. Books and articles cited.

- [1] Wm. M. Adams, "Relationship of keyword in titles reference cited." *American Documentation*. Vol. 18, no. 1967. pp. 26-32.
- [2] H. Borko, "Measuring the reliability of subject classification by men and machines." *American Documentation*. Vol. 15, no. 4. 1964. pp. 268-73.
- [3] John M. Christ; *Concepts and Subject Headings; their Relation in Information Retrieval and Library Science*. Metuchen, N.J., The Scarecrow Press, Inc., 1972. 174 p.
- [4] J.E. Daily, "Many changes. No alteration; an analysis of Library of Congress Subject Headings, 7th edition." *Library Journal*, Vol. 42. Nov. 1967. p. 3962.
- [5] Richard De Gennaro, "Harvard University's Widener Library Shelflist Conversion and Publication Program. *College and Research Libraries*. Vol. 31, no. 5. 1970. pp. 318-31.
- [6] Arthur W. Elias (ed.), *Key Papers in Information Science*. American Society for Information Science. Washington D.C., 1971. 223 p.
- [7] M. Fischer, "The KWIC index concept; a retrospective view," *American Documentation*. Vol. 17, no. 2. 1966. pp. 57-70.

- [ 8 ] H.H. Goom, "A computerbased current awareness system producing both SDI output and conventional abstracts bulletin." *Aslib Proceedings*. Vol. 26. no. 3. 1974. pp. 98-108.
- [ 9 ] G. Jahoda & L.M. Stursa, "A comparison of a keyword from title indexes with a single access point per document alphabetical index." *American Documentation*. Vol. 20, no. 4. pp. 377-80.
- [10] Kenneth Janda, *Information Retrieval; application to political science*. New York, Bobbs-Merill Co. Inc., 1968. xxiii, 230 p.
- [11] D.H. Kraft, "A comparison of Keyword-in-context (KWIC) indexing of titles with a subject heading classification system." *American Documentation*. Vol. 15, no. 1. 1964. p. 50.
- [12] Yoshiro Matsuda, "No Deus ex Machina will appear in our library modernization." *Keizai Shiryo Kenkyu*, (in Japanese), no. 4. 1971. pp. 1-23.
- [13] ditto, "Retrospective search through KWIC index; Bibliography and library for social sciences." *Shogaku Tokyu* (Economic Review, in Japanese), New series vol. 24, no. 2. 1973. pp. 1-32.
- [14] ditto, "Present state of documentation in Japanese social sciences." *Joho Kanri* (Information & Documentation, in Japanese). Vol. 17, no. 6. 1974.
- [15] Sachiko Matsui, "Information retrieval for interdisciplinary science through subject headings of Library of Congress. Case study of management sciences." *Shogaku Tokyu* (Economic Review, in Japanese), New series vol. 23. no. 2 / 3. 1971. pp. 287-310.
- [16] ditto, "Programming techniques for machine readable bibliography. Review and case study in the field of social sciences." *Shogaku Tokyu*. (Economic Review, in Japanese). New series vol. 24, no. 2. 1974. pp. 33-70.
- [17] F.W. Matthews & A.D. Shillingford, "Variations on KWIC." *Aslib Proceedings*. Vol. 25. no. 4. 1973, pp. 140-152.
- [18] V. Mostecky, "Study of the See-Also reference structure in relation to the subject of international law." *American Documentation*. Vol. 7, 1956. p. 303.
- [19] K.C. Rosenberg & C.L.M. Bolcher, "A comparison of the relevance of KWIC vs. descriptor indexing terms." *American Documentation*. Vol. 19, no. 1. 1968. pp. 27-29.
- [20] M.J. Ruhe, "Chemical documents and their titles; human concept indexing vs. KWIC machine indexing." *American Documentation*. Vol. 15, no. 2. pp. 136-149.
- [21] Claire K. Schultz (ed.), *H.P. Luhn; Pioneer of Information Science; Selected works*. New York, Spartan Books, 1968.
- [22] U.S. Library of Congress, Subject Cataloging Division, Processing Department. *Subject Headings used in the Dictionary Catalog of the Library of Congress*. (ed. by M.V. Quattelbaum) 7th ed. Washington D.C., Government Printing Office, 1966. 1432 p.
- [23] B.C. Vickery, *On Retrieval System Theory*. London, Butterworths, 1961.
- [24] Hans (Hanan) Wellisch & Thomas D. Wilson, *Subject Retrieval in the Seventies; new directions; Proceedings of an international symposium held at the Center of Adult Education University of Maryland College Park, May 14 to 15, 1971*. Westport, Greenwood Publishing Co., 1972. (Contributions in Librarianship and Information Science. No. 3.)