Describing Photoelectric Conversion by Co-Word Analysis

Radosvet Todorov[*] *and Matthias Winterhager*[**]

Introduction

Developments of software in host systems for pure retrieval purposes offer new possibilities for counting and arranging of different items from a set of document records.(1) Lists of frequently appearing document types, research characteristics, controlled terms, classification codes, journal titles, institutes, etc. as well as their distribution give an overview of some features which are relevant to bibliometric studies and more specifically to further structure analysis of a preselected subfield of science. Presentation of such structures is performed usually for social studies of science and science policy purposes. It is based mostly on data of appearance and co-appearance of document related elements such as citations, (2-3) words (selected from different parts of the text or the document record) (4-8), or their combination (9) and subject headings.(10) Co-word analysis has been viewed as a convenient (or alternative) approach (11) as compared to co-citation analysis to describe subject structure by selecting highly appearing and co-appearing controlled terms. For this purpose, appropriate similarity measures are used to show closeness of frequently used terms in two dimensions.(12-13)

* Center for Science Studies, Bulgarian Academy of Sciences, Serdika St No 4, 1000 Sofia, Bulgaria

** Center for Science Studies, University of Bielefeld, P.O. Box 8640, 4800 Bielefeld 1, Federal Republic of Germany

The aim of this paper is to describe the subfield of photoelectric conversion:

1) by on-line bibliometrics, i.e. by collecting and (internal) processing of data from a bibliographic file using only the software options included in the host system, and
2) by co-word analysis using an appropriate software for counting frequency of appearance and co-appearance of controlled terms of a preselected set of document records.

The subfield of photoelectric conversion has been selected because of some recent scientific and technological advances (especially in condensed matter physics and electronics) in this area.

Data and Method

For the purpose of the study all documents containing any classification code for photoelectric conversion, namely SOLAR CELLS AND ARRAYS or PHOTOELECTRIC DEVICES have been retrieved from the INSPEC bibliographic file for the publication years 1980–1989. Data on paper output in the subfield, document types, research characteristics (theoretical, experimental, applied, etc.), publication language, etc. have been then obtained by using the retrieval language and implemented software of the host system STN. More specifically, the items under consideration have been counted and ranked in the preselected set of document records. Thus, different distributions of publications are derived which are used to describe the scope and characteristics of the subfield as well as the publication activity of researchers.

Further, all record fields containing the Controlled Terms (CT) of 1988 documents have been extracted (downloaded) and processed by an external software to obtain the absolute frequency of appearance as well as their co-appearance. (Two Controlled Terms C_i and C_j are said to co-appear if there is at least one record containing both C_i and C_j.)

The resulting matrix $C = \{C_{ij}\}$ is used for calculating an appropriate relative measure of similarity. A diagonal element C_{ii} of C is the absolute frequency of appearance of C_i and an off-diagonal element C_{ij} represents the co-occurrence of C_i and C_j in the set of records under consideration. Here, the inclusion index $I_{ij} = C_{ij} / \min (C_{ii}, C_{jj})$ is selected as a measure of similarity.(12) The I_{ij} expresses the degree to which two CT are related. It is a measure of similarity as the highest values correspond to pairs that most co-appear in the given set of document records. The I_{ij} offers the possibility to highlight the main terms or "central poles" in photoelectric conversion papers and to reveal their relations to other "peripheral" terms.

To display the position of terms in two dimensions (according to the selected measure of similarity) a multidimensional scaling (ALSCAL program) is used. Since such a representation does not display the configuration correctly, relation strength has been additionally shown by connecting frequently co – appearing terms (from the lower to the higher frequency of appearance) with lines according to the inclusion index values. Furthermore, the results of a hierarchical clustering are included to restore information lost in the display process: terms that belong to the same cluster have been surrounded (enclosed in boxes) to show clearly their stronger links in a complementary and independent way.

Results

Figures 1 to 6 as well as Tables 1 and 2 give the results of the on – line bibliometric analysis. Publication activity in photoelectric conversion is Figure 1). Then, distribution of publications over different research characteristics are shown: experimental, theoretical, and applied (Figure 2), practical and applied (Figure 3), new developments and economic aspects (Figure 4). Figures 5 and 6 give the distributions of documents by language of publication. The main sources (journals) and producers (research units) of publications are listed in Tables 1 and 2. The most used controlled terms are ranked according to their absolute frequency in Table 3. Links between these terms are shown in a two – dimensional ALSCAL configuration: first by connecting only those of them that frequently co – appear (see Figure 7), then by adding the results of cluster analysis, i.e. by grouping related terms (see Figure 8) and at last by representing additional links of low co – appearance (see Figure 9).

Discussion

Collection of data directly from a bibliographic file and their further on – line processing depend on record structure and software options incorporated in the host system. Results of such quick bibliometric analyses are generally displayed in the form of ranked lists of appearance of preselected field items in a given set of document records. Lists of extracted items are arranged alphabetically or according to the absolute frequency of occurrence. Here, different software options of STN are used to count the number of documents and their characteristics in each publication year between 1980 and 1989 as well as most "productive" journals and (first) author addresses in the subdivisions SOLAR CELLS AND ARRAYS or PHOTOELECTRIC DEVICES for the year 1988.

The variation in publication output during the last ten years is due exclusively to the considerable fluctuation of conference papers (see Figure 1). The total number of documents in the selected subdivisions is slightly decreasing. As many documents are classified in more than one type, the total number is not equal to the sum of the different types. The subfield of photoelectric conversion is represented mostly by articles and conference papers, and by a negligible number of reviews and bibliographies. The same is true for the research characteristics. Publications describing experimental methods, observations or results as compared to those being of a theoretical nature dominate the subfield. The number of applied–oriented documents is relatively low (see Figure 2). "Application" expresses here the actual or potential use of an instrument, device, method or technique. A difference is made between the two research characteristics "applied" and "practical" (Figure 3). Publications including practical results are intended to be of direct practical use and are likely to be of interest particularly to engineering and development staff.

A rather small number of publications make claims of novelty (in the patent sense). They are called by INSPEC "new development" documents. Their number is represented in Figure 4 together with the number of publications describing economic aspects of solar cells use.

Most publications in the subfield under consideration are published in English (see Figure 5). Documents in the languages Japanese and German increase (as compared to French) in the last five years (see Figure 6).

Journal articles in the subfield of photoelectric conversion are rather scattered over titles (see Table 1): half of the articles are published in ten journals which represent 9% of all journals. This could be related to the predominance of experimental and practical papers. The productive journals are published primarily in Switzerland, USA (including AIP CONFERENCE PROCEEDINGS), and the Netherlands.

Aside from the USA, main research units, i.e. those generating more than 3 publications in 1988, are concentrated in West Germany, Italy, and Japan (see Table 2).

The structure of photoelectric conversion is displayed here by co–word analysis because of some advantages such as explicit content of the terms and simplified selection of the relevant records for the analysis. Some limitations linked to "indexer" effect should be mentioned as well.(11,14)

Using an appropriate software a matrix of (controlled) terms co–appearance is constructed. On the diagonal the frequencies of appearance of the terms are placed in descending order (from Table 3). The full size of the matrix is equal to the number of different terms occurring in the selected records. With a view to statistical validity and better representation, the initial size is reduced by introducing a threshold. The practical criteria for its determination are the amount of information lost and the number of terms (and their links) to be displayed.(3) If a given term of low appearance is not on the list (not included in the matrix) an additional detailed representation could be generated by taking only those terms that strongly co – appear with the term in consideration. Thus, a "zoom" map could reveal links of any term of interest.

Once the similarity measure is selected, multidimensional scaling has been used to display the structure of photoelectric conversion. In Figure 7 a two – dimensional solution from ALSCAL is shown. On this spatial representation two terms should be further apart the smaller their inclusion index value is. Since two dimensions are, in general, not enough to reflect adequately all relations (the residual departure from monotonicity or 'stress' is large) those terms whose similarity exceeds the threshold value of 0.75 are linked by lines. Except SOLAR CELL ARRAYS and PHOTOVOLTAIC POWER SYSTEMS all topics are strongly related to the "central pole", i.e. SOLAR CELLS (see Figure 7). Other relationships form separated groups which are also confirmed by hierarchical cluster analysis. This analysis is used here as a complementary method. The cluster solution is shown in Figure 8. The terms within the clusters are connected with each other (by lines which are drawn from terms with lower to such with higher absolute frequency of appearance), and not connected with those outside the clusters. This absence of crossing lines (based on the proximities data) indicates an agreement between closeness in the data and closeness in the space. If the threshold value is lowered, some relations between clusters appear. They are indicated by dashed lines on Figure 9.

The map of photoelectric conversion (see Figure 7 and 8) displays one main topic (SOLAR CELLS) and related secondary poles or items: SILICON, III – V SEMICONDUCTORS, II – VI SEMICONDUCTORS, SEMICONDUCTOR THIN FILMS, and TERNARY SEMICONDUCTORS. The apparent "splitting" on the map in a left and right part which could be called, respectively, THIN FILM and SILICON sections, reflects the development of two different technologies (see Figure 9). Amorphous silicon (expressed by the link between SILICON and AMORPHOUS SEMICONDUCTORS) is in a leading position (high absolute frequency of appearance and co – appearance) in the SILICON group. As main thin film materials, used in solar cells, appear GALLIUM ARSENIDE (from the III – V SEMICONDUCTORS group), CADMIUM COMPOUNDS (from the II – VI SEMICON – DUCTORS group), and COPPER and INDIUM COMPOUNDS (from the TERNARY SEMICONDUCTORS group).

Summary and Conclusions

The study has been designed to investigate capabilities of host software options for preliminary data selection and quick processing. The results of such on – line bibliometrics are used to describe different publication – related characteristics of the subfield of photoelectric conversion. Additionally, co–word analysis is applied to display main subject topics and their links. Both quantitative representations are performed without any substantive knowledge and could be utilized (in conjunction with other quantitative methods or expert opinion) in science and science policy studies. The structure map is based on a co–occurrence model which is introduced here outside of a conceptual framework (11) as well as without a separate analysis of "indexer effect" (14) and technical limitations.(3)

References

1. O. PERSSON, Online Bibliometrics. A Research Tool for Every Man, *Scientometrics, 10* (1986) 69–75.
2. J. J. FRANKLIN and R. JOHNSTON, Co–Citation Bibliometric Modeling as a Tool for S&T Policy and R&D Management: Issues, Applications, and Developments, In: A.F.J. Van Raan (Ed.) *Handbook of Quantitative Studies of Science and Technology*, North–Holland, Amsterdam (1988), 325–389.
3. J. E. J. OBERSKI, Some Statistical Aspects of Co–Citation Cluster Analysis and a Judgement by Physicists, In: A.F.J. Van Raan (Ed.) *Handbook of Quantitative Studies of Science and Technology*, North–Holland, Amsterdam (1988), 431–462.
4. M. CALLON, et al., From Translation to Problematic Networks: An Introduction to Co–Word Analysis, *Social Science Information, 22* (1983) 191–235.
5. A. RIP and J. P. COURTIAL, Co–Word Maps of Biotechnology: An Example of Cognitive Scientometrics, *Scientometrics, 6* (1984) 381–400.
6. J.–P. COURTIAL and J. LAW, A Co–Word Study of Artificial Intelligence, *Social Studies of Science, 19* (1989) 301–311.
7. L. LEYDESDORFF, Words and Co–Words as Indicators of Intellectual Organization, *Research Policy, 18* (1989) 209–224.
8. J. WHITTAKER, J. LAW, Creativity and Conformity in Science – Titles, Keywords and Co–word Analysis, *Social Studies of Science, 19* (1989) 473–496.
9. R. R. BRAAM, H. F. MOED, and A. F. J. VAN RAAN, Comparison and Combination of Co–Citation and Co–Word Clustering, In: A. F. J. Van Raan, A. J. Nederhof, H. F. Moed (Eds.) *Science and Technology Indicators*, DSWO Press, Leiden (1989), 307–337.
10. R. TODOROV and M. WINTERHAGER, Mapping Australian Geophysics: A Co–Heading Analysis, *Scientometrics, 19* (1990) 35–56.
11. A. RIP, Mapping of Science: Possibilities and Limitations, In: A. F. J. Van Raan (Ed.) *Handbook of Quantitative Studies of Science and Technology*, North–Holland, Amsterdam (1988), 253–273.

12. J. – P. COURTIAL, Technical Issues and Developments in Methodology, In: M. Callon, J. Law, A. Rip (Eds.) *Mapping the dynamics of science and technology*, London, Macmillan (1986), 189 – 210.
13. R. J. W. TIJSSEN and A. F. J. VAN RAAN, Mapping Co – Word Structures: A Comparison of Multidimensional Scaling and Leximappe, *Scientometrics, 15* (1989) 283 – 295.
14. J. – P. COURTIAL, et al., Is Indexing Trustworthy? Classification of Articles through Co – Word Analysis, *Journal of Information Science, 9* (1984) 47 – 56.

Fig.1 Number of publications and main document types

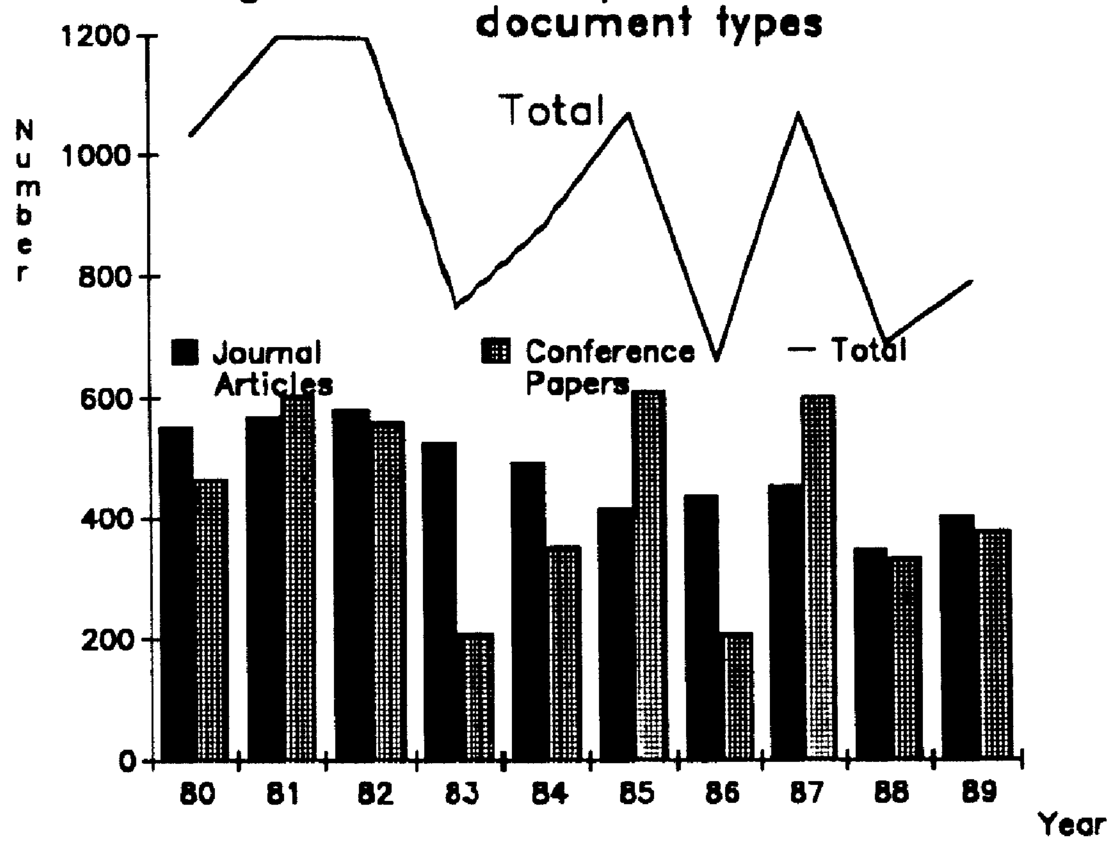


Fig.2 Research characteristics of publications (1)

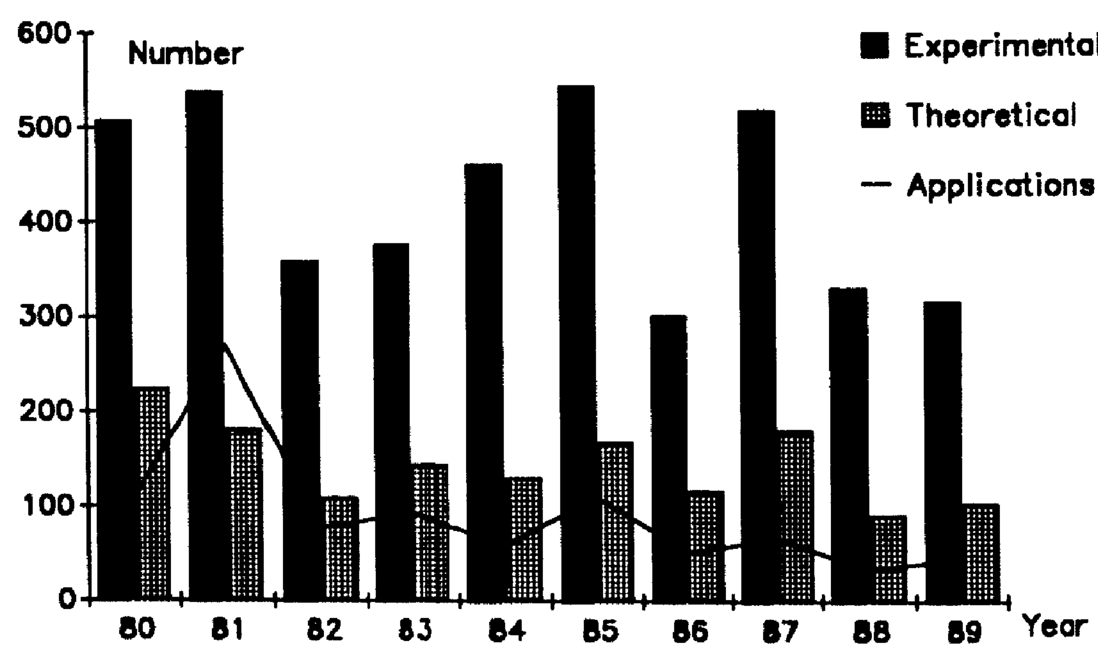


Fig.3 Research characteristics of publications (2)

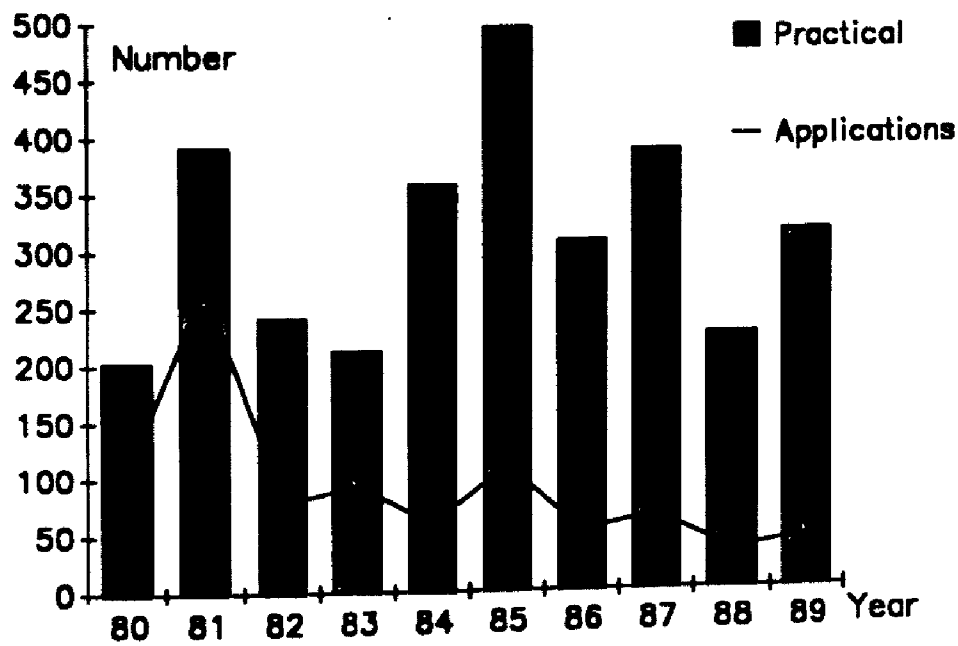


Fig.4 Research characteristics of publications (3)

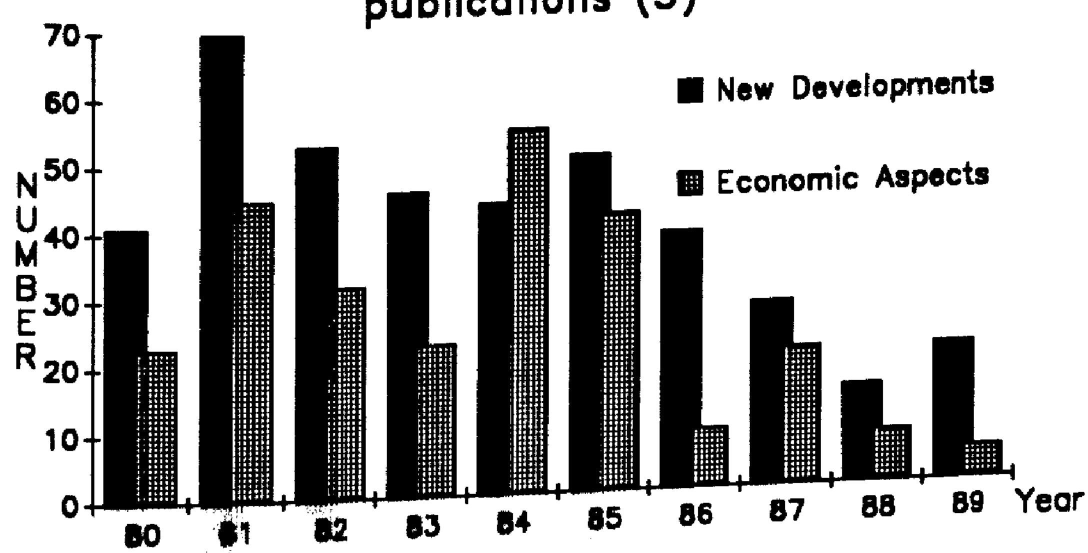


Fig.5 Distribution of documents by language of publications

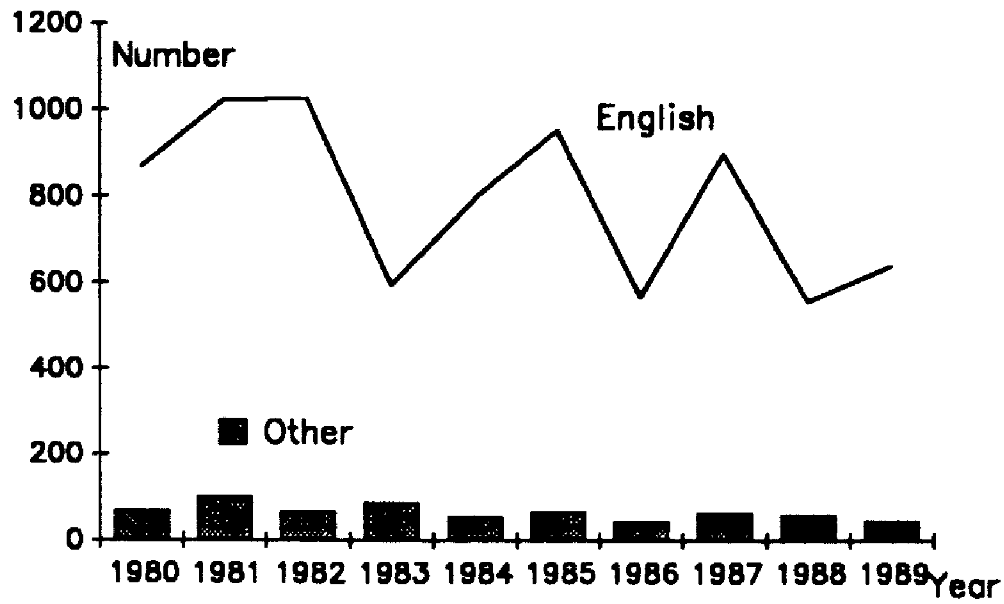


Fig.6 Number of publications in other languages

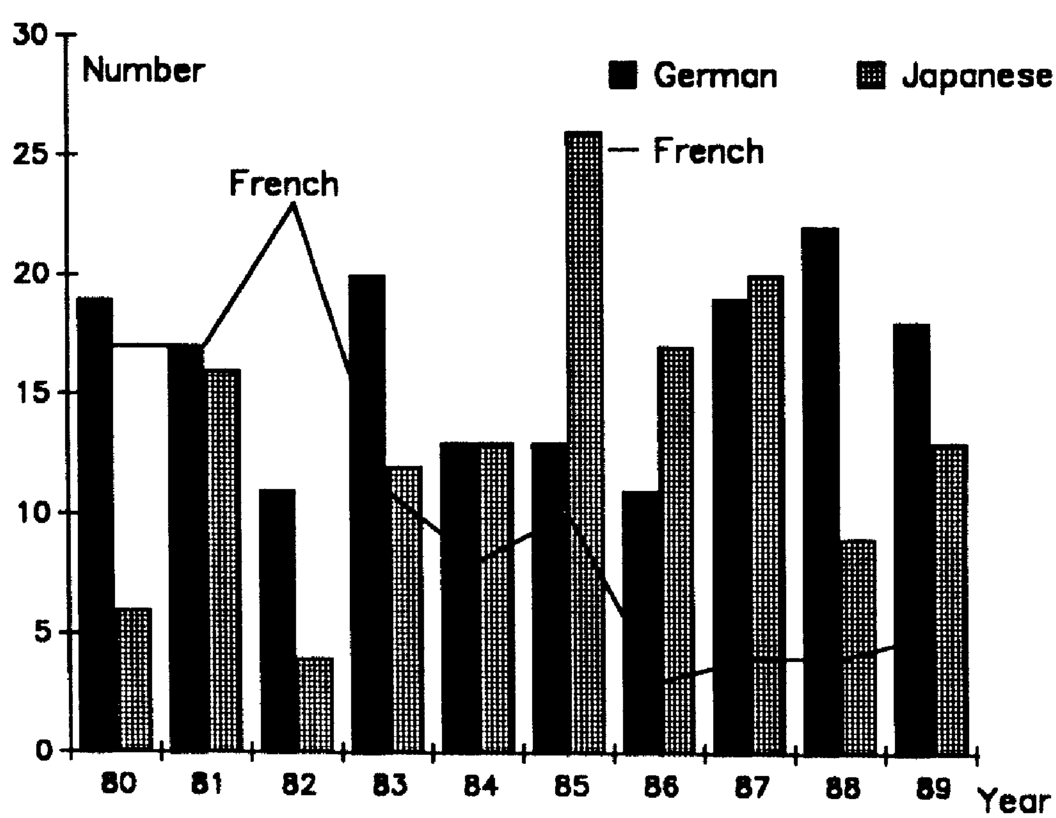
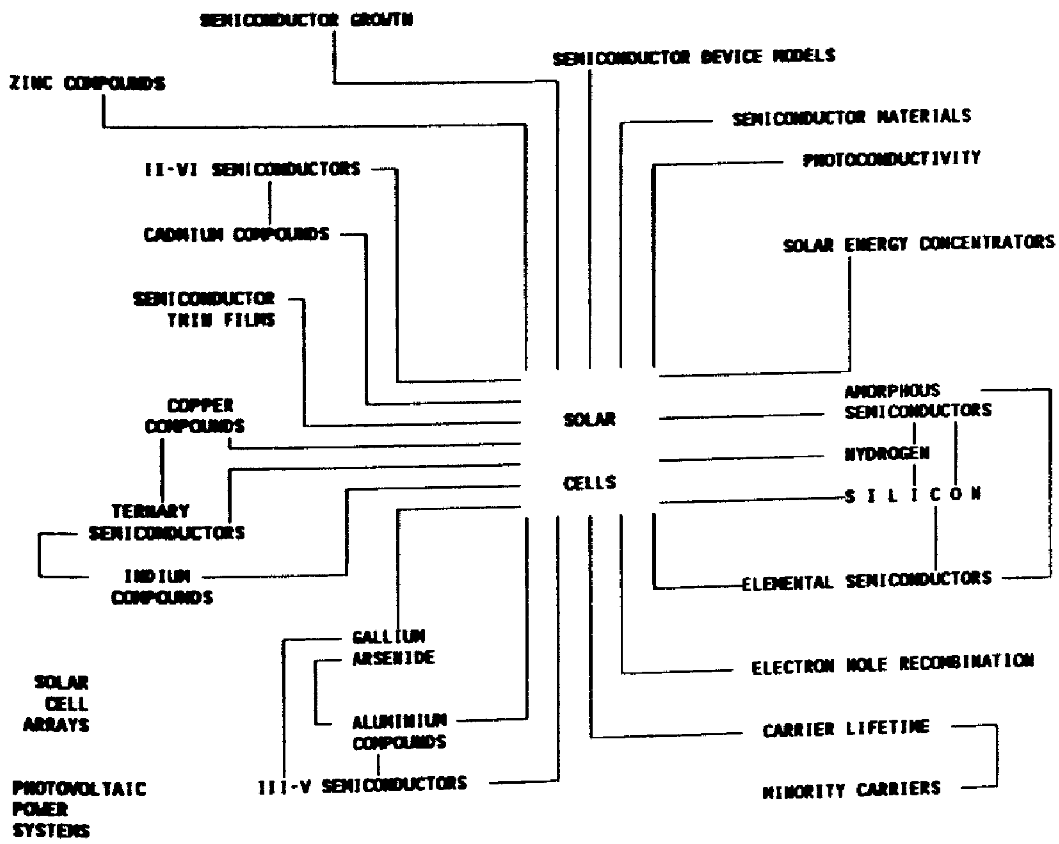
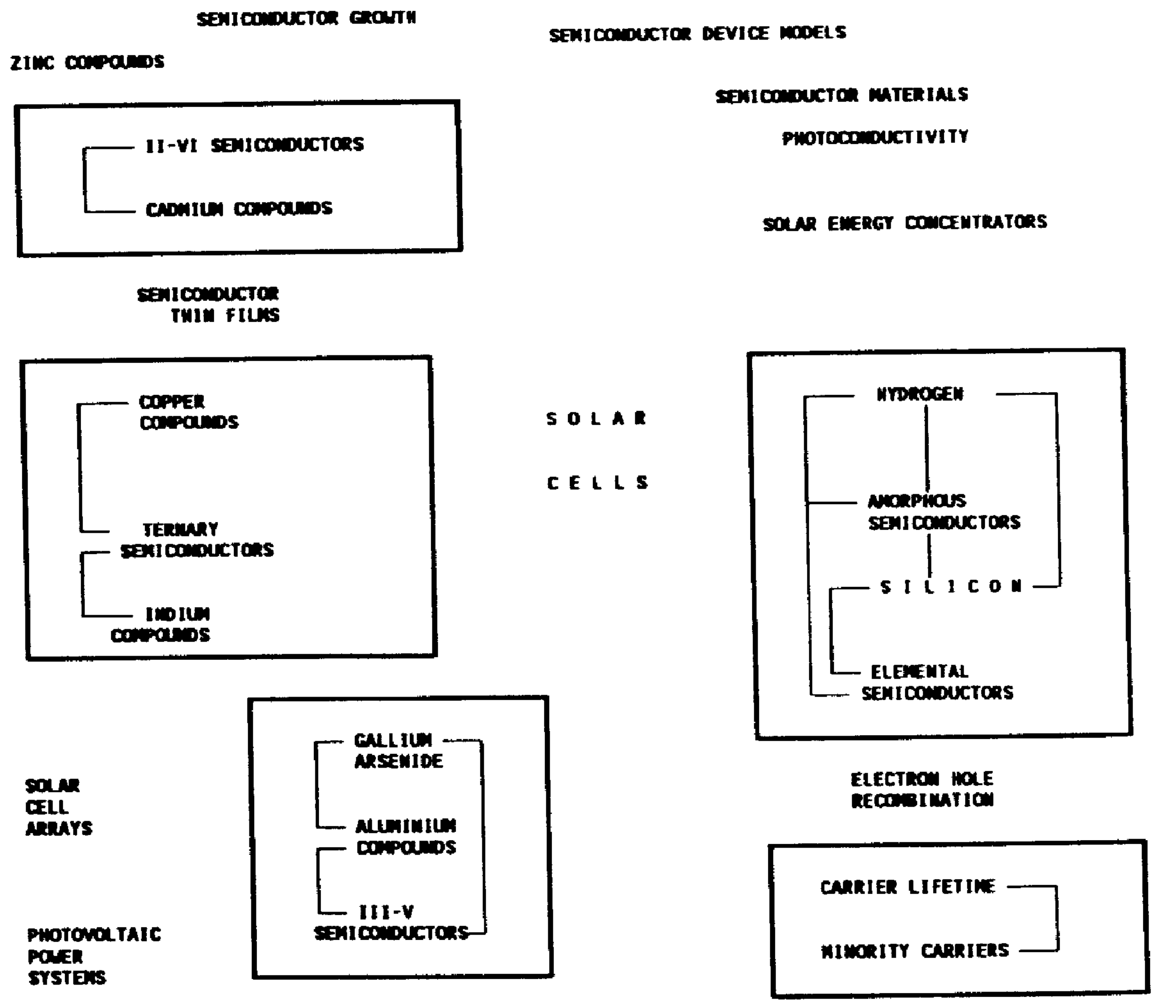


Fig. 7 Main Links between highly appearing Controlled Terms in 1988 Photoelectric-Conversion-Papers*



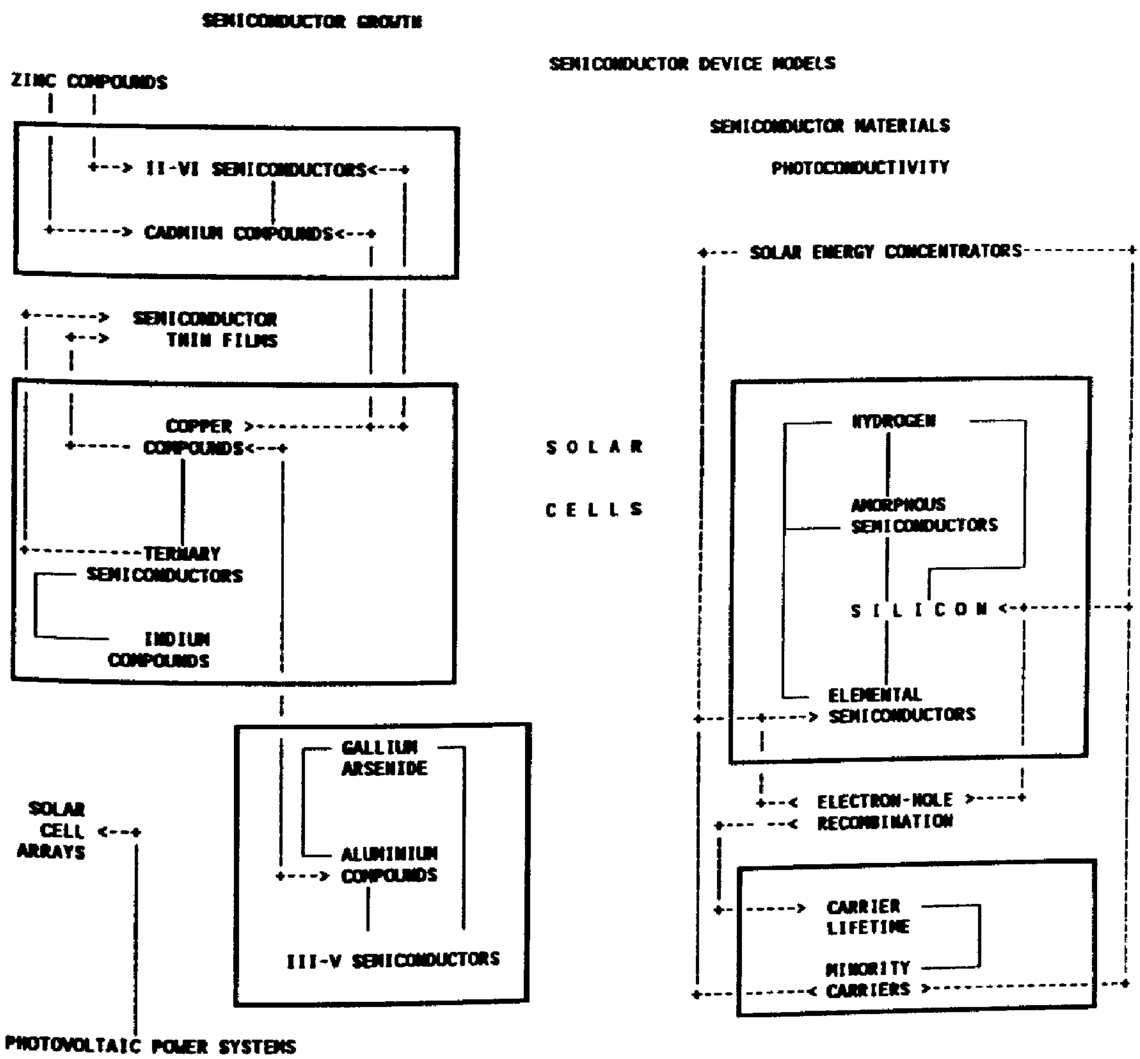
* Lines express relation strength which corresponds to an Inclusion Index value ≥ 0.75

Fig. 8 Grouping of Controlled Terms based on their co-appearance in 1988 Photoelectric Conversion Documents*



----
* Links between SOLAR CELLS and the other terms as shown in Figure 7 are omitted.
——— Inclusion Index ≥ 0.75
——— Cluster border

Fig. 9 Additional links between highly appearing Controlled Terms in 1988 Photoelectric-Conversion-Papers



——— Inclusion Index ≥ 0.75
------- Inclusion Index ≥ 0.50
——— Cluster border

Table 1:

Distribution of Photoelectric – Conversion – Articles by Journal Titles

Nb	%	Journal Title (Country)
65	18	SOLAR CELLS (SWITZERLAND)
22	6	JOURNAL OF APPLIED PHYSICS (USA)
18	5	SOLAR ENERGY MATERIALS (NETHERLANDS)
16	4	APPLIED PHYSICS LETTERS (USA)
15	4	AIP CONFERENCE PROCEEDINGS (USA)
12	3	APPLIED SOLAR ENERGY
12	3	GELIOTEKHNIKA (USSR)
10	3	INTERNATIONAL JOURNAL OF SOLAR ENERGY
9	2	SOLID-STATE ELECTRONICS
9	2	THIN SOLID FILMS (SWITZERLAND)
7	2	JAPANESE JOURNAL OF APPLIED PHYSICS, PART 1
7	2	SOLAR WIND TECHNOLOGY (UK)
6	2	BULLETIN OF MATERIALS SCIENCE
5	1	BULLETIN DER SCHWEIZ. ELEKTROTECH. (SWITZERLAND)
5	1	IEEE TRANSACTIONS ELECTRON DEVICES (USA)
5	1	PHYSICA STATUS SOLIDI A
4	1	ELECTROTEHNICA, ELECTRONICA SI AUTOMATICA.
4	1	IEEE ELECTRON DEVICE LETTERS (USA)
4	1	IEEE TRANSACTIONS ENERGY CONVERSION (USA)
4	1	INDIAN JOURNAL OF PURE AND APPLIED PHYSICS
3	1	ENERGY CONVERSION MANAGEMENT (UK)
3	1	JOURNAL OF MATERIAL SCIENCE (UK)
3	1	JOURNAL OF MATERIAL SCIENCE LETTERS (UK)
3	1	JOURNAL OF THE ELECTROCHEMICAL SOCIETY
3	1	MODELLING, SIMULATION & CONTROL A
3	1	PIS'MA V ZHURNAL TEKHNICHESKOI FIZIKA
3	1	SANYO TECHNICAL REVIEW
3	1	SOVIET TECHNICAL PHYSICS LETTERS
3	1	SPACE POWER
2	10	Other 18 JOURNALS PUBLISHING 2 ARTICLES EACH
1	18	Other 65 JOURNALS PUBLISHING 1 ARTICLE EACH

Table 2:

Distribution of Photoelectric – Conversion – Papers by most productive[*] institutes (according to the first author's address only)

Nb	Institute/Country
21	SOLAR ENERGY RES. INST., GOLDEN, CO, USA
13	FRAUNHOFER-INST. FUR SOLARE ENERGIESYST., FREIBURG, WEST GERMANY
13	SIEMENS RES. LABS., MUNICH, WEST GERMANY
12	INST. DE ENERGIA SOLAR, UNIV. POLITECNICA DE MADRID, SPAIN
8	SOLAREX THIN FILM DIV., NEWTOWN, PA, USA
7	ARCO SOLAR INC., CHATSWORTH, CA, USA
6	DEPT. PHYS., KONSTANZ UNIV., WEST GERMANY
6	INST. FUR PHYS. ELEKTRONIK, STUTTGART UNIV., WEST GERMANY
6	JRC, CEC, ISPRA ESTABL., ITALY
6	SCH. OF ELECTR. ENG., PURDUE UNIV., WEST LAFAYETTE, IN, USA
6	A. F. IOFFE PHYS. TECH. INST., LENINGRAD, USSR
5	DEPT. PHYS., KALYANI UNIV., WEST BENGAL, INDIA
5	ENIRICERCHE SPA, ROMA, ITALY
5	SANYO ELECTR. CO. LTD., OSAKA, JAPAN
5	SCH. OF ENG. & APPL. SCI., DURHAM UNIV., UK
5	PHOTOVOLTAIC TECHNOL. DIV., SANDIA NAT. LAB., ALBUQUERQUE, NM, USA
5	ALL UNION SCI. RES. INST. OF CURRENT SOURCES, KRASNODARSK, USSR
4	AEG AG, WEDEL, WEST GERMANY
4	HELIOTRONIC GMBH, BURGHAUSEN, WEST GERMANY
4	DEPT. DE ELECTRON., UNIV. POLITECNICA DE CATALUNA, BARCELONA, SPAIN
4	AMETEK APPL. MATER. LAB., HARLEYSVILLE, PA, USA
4	CHRONAR CORP., PRINCETON, NJ, USA
4	DEPT. OF ELECTR. ENG., FLORIDA UNIV., GAINESVILLE, FL, USA
4	JET PROPULSION LAB., CALIFORNIA INST. OF TECHNOL., PASADENA, CA, USA
4	SPIRE CORP., BEDFORD, MA, USA
4	VARIAN RES. CENTER, PALO ALTO, CA, USA
3	SOLAREX CORP., FREDERICK, MD, USA
3	LAB. DE PHYS. DU SOLIDE ET ENERGIE SOLAIRE, CNRS, VALBONNE, FRANCE
3	BATTELLE INST. EV, FRANKFURT, WEST GERMANY
3	INST. WERKSTOFFE DER ELEKTROTECH., TECH. UNIV. BERLIN, WEST GERMANY
3	INST. WERKSTOFFWISSENSCHAFTEN, ERLANGEN UNIV., WEST GERMANY
3	SIEMENS SOLAR GMBH, BERGISCH GLADBACH, WEST GERMANY
3	DEPT. PHYS., ROME UNIV., ITALY
3	DIPARTIMENTO DI FISICA, PARMA UNIV., ITALY
3	DEPT. OF CHEM., OSAKA UNIV., JAPAN

Table 2 (continued):

Nb	Institute/Country
3	FAC. OF ENG. SCI., OSAKA UNIV., JAPAN
3	DEPT. DE FISICA APLICADA I ELECTRON., BARCELONA UNIV., SPAIN
3	NEWCASTLE POLYTECH., UK
3	SOLAPAK LTD, HIGH WYCOMBE, UK
3	BIOMED.& ENVIRON. ASSESS. DIV., BROOKHAVEN NAT. LAB., UPTON, NY, USA
3	DEPT. OF PHYSIOL., MICHIGAN STATE UNIV., EAST LANSING, MI, USA
3	GEORGIA INST. OF TECHNOL., ATLANTA, GA, USA
3	GLASSTECH SOLAR INC., WHEAT RIDGE, CO, USA
3	INST. OF ENERGY CONVERSION, DELAWARE UNIV., NEWARK, DE, USA

* publishing more than 3 papers in 1988 (data from the INSPEC file)

Table 3:

Frequency of appearance of Controlled Terms in bibliographic records of Photoelectric – Conversion – Papers*

Rank	All	Selected**	Controlled Terms
1	508	469	SOLAR CELLS
2	241	241	SILICON
3	240	240	ELEMENTAL SEMICONDUCTORS
4	126	125	AMORPHOUS SEMICONDUCTORS
5	87	87	HYDROGEN
6	75	75	SEMICONDUCTOR THIN FILMS
7	66	66	III-V SEMICONDUCTORS
8	65	65	CADMIUM COMPOUNDS
9	62	62	II-VI SEMICONDUCTORS
10	56	56	GALLIUM ARSENIDE
11	51	51	INDIUM COMPOUNDS
12	80	48	SOLAR CELL ARRAYS
13	42	42	COPPER COMPOUNDS
14	42	42	CARRIER LIFETIME
15	38	38	PHOTOCONDUCTIVITY
16	38	38	SEMICONDUCTOR GROWTH
17	39	37	PHOTOVOLTAIC POWER SYSTEMS
18	29	29	TERNARY SEMICONDUCTORS
19	29	29	ALUMINIUM COMPOUNDS
20	28	28	ELECTRON-HOLE RECOMBINATION
21	28	28	MINORITY CARRIERS
22	27	27	SEMICONDUCTOR MATERIALS
23	25	25	ZINC COMPOUNDS
24	24	24	SEMICONDUCTOR DEVICE MODELS
25	24	24	SOLAR ENERGY CONCENTRATORS
..	..	..	....

* data are extracted from the 1988 INSPEC file
** appearing in records that contain two or more Controlled Terms