Electronic Publishing and Document Delivery of German Patent Information†

GERD TITTLBACH

Fachinformationszentrum Karlsruhe, D 7514 Eggenstein-Leopoldshafen 2, FRG

Received June 13, 1985

Advances in information techniques and telecommunication improve the world-wide transfer of scientific and technical information including patent information. This paper will discuss current and short-dated planned online patent information services offered via the STN Node Karlsruhe. Special attention is addressed to ongoing developments in extending the access to the German Patent Database PATDPA containing both text and graphical representations. The novel features will be the joint storage of text and drawings in one database, the conversion of digitized graphical data into vector-graphics out format, and the combined transmission of text and graphics via telecommunication networks to various types of terminals. The advantages of vector graphics vs. raster graphics for storing, transmitting, and displaying technical drawings as well as some details of graphics handling and terminal support are outlined. The project is carried out by a consortium consisting of, among others, the German Patent Office and FIZ Karlsruhe as the consortium leader. Information services are planned to be extended step by step from bibliographic searches and SDI services via retrieval and transmission of text and graphics of the front page of patent publications toward full-text searching and delivery. Online access is available via STN International including cross-file searching in the leading databases in science and technology.

OBJECTIVES AND USER DEMAND

The value of patent information as an indispensable source for scientific and technical information transfer and technology transfer is more and more recognized. Patents have a dual function, i.e., an information function and a protective function. In consequence, quick and direct access to patents and other relevant publications in science and technology is a basic prerequisite for efficient utilization of the available know-how and for technical innovation.

Patent publications consist of a structured text that may be supplemented by drawings (Figure 1). The text comprises application and publication data as well as data on the legal status of the patent, the abstract, the patent claims, the technical description, the state-of-the-art review, etc. The drawings and other graphical representations, e.g., chemical structures, provide additional information for an assessment of the applicant's protective rights and of the technical features of an invention.

In addition to the existing patent information services,[1,2] a German pilot project[3] intends to facilitate improved access to the millions of pages of patent publications by setting up a national German Patent Database containing both text and graphical representations and by making this database accessible to the professional community via computerized information services, i.e., by providing online access via the public telecommunication networks including cross-file searching in other patent databases as well as bibliographical and factual databases, by SDI services, by statistical evaluations, etc.

The dual function of patent information makes it interesting to a broad user spectrum, e.g., Patent Offices (German Patent Office, European Patent Office, other national patent offices) and the public (patent and information departments of large industrial companies, small and medium-sized enterprises, patent agents, technology-transfer agents, innovation consultants, chambers of commerce and trade, university institutes, research centers).

Market studies demonstrate the demand as follows: There is a need for setting up a German Patent Database embedded in surrounding online patent and nonpatent information services. Apart from direct access to the bibliographic data (on the application and the legal status), access to technical data, facts, contents, patent claims, major drawings, and the complete document is of particular interest. Patent documentation as far back as possible is requested, meeting the professional needs of the various user groups.

PLANS, TECHNICAL CONCEPT, AND INFORMATION SERVICES

Plans to develop and set up a German full-text database containing text and graphics—patent drawings, chemical structures, complex mathematical formulae, etc.—will be realized step by step.

First of all the machine-readable data of the German Patent Office will be made available for online access, including the bibliography of German patent publications (applications and granted patents), of German utility models, and of European and PCT applications designated to the Federal Republic of Germany, all of them back to 1973, as well as the abstracts of these publications back to 1981. File design and patent formats take into account convenient search possibilities specific to patent publications. Formats are designed in accordance with international standards and, where not yet available, along the lines of the INPADOC formats, the producer of the world's most comprehensive patent database. Access to the technical content will be made available via the International Patent Classification (IPC), via free-text searching of the title and abstract, and via an IPC file including the full-text of the IPC in German, English, and French.

Novel features to be developed by the pilot project in 1985 concern combined storage of text and graphics in one database system, digitizing the graphics and conversion of the graphical data into vector graphics storage and output format, and combined transmission of text and graphics via public telecommunication networks supporting different types of terminals.

The current developments concerning storage, transmission, and display of text and graphics have been carried out by using the information of the front page of patent publications with the most important patent drawing or other graphics embedded in the text. The technical developments achieved are planned to be used for extending the database to a full-text database including all the patent drawings. Machine-readable recording of the full text via keyboarding the text and digitizing the

† Presented at the Spring Meeting of the American Chemical Society, Division of Chemical Information, April 29, 1985, Miami Beach, FL.

(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Offenlegungsschrift**

(11) **DE 3411936 A1**

(51) Int. Cl. 3:
G 01 N 15/07
A 61 B 5/14
G 01 N 33/48

DE 3411936 A1

(21) Aktenzeichen: P 34 11 936.1
(22) Anmeldetag: 30. 3. 84
(43) Offenlegungstag: 11. 10. 84

(30) Unionspriorität: (32) (33) (31)
31.03.83 CH 1805-83

(71) Anmelder:
Meier, Johann, Brione sopra Minusio, CH;
Bretschger, Eduard, Dr.med., Stäfa, CH

(74) Vertreter:
Walter, H., Pat.-Anw., 8000 München

(72) Erfinder:
gleich Anmelder

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Teilchenzähleinrichtung und Verfahren zu deren Betrieb

Diese z. B. zum Zählen von in Blut sich befindenden Blutkörperchen dienende Teilchenzähleinrichtung ist mit einem Fördersystem (16, 17) zur Beförderung einer abgemessenen Menge an Blut von einem ersten Behälter (3) über eine in einem Meßkopf (4) vorgesehene kalibrierte Meßstrecke in einen zweiten Behälter (6), sowie mit einer elektronischen Zählvorrichtung (10, 11, 20, 21) zur Zählung der die Meßstrecke passierenden Blutkörperchen versehen.
Um einen einfachen Unterhalt der Einrichtung zu erreichen, sind der erste Behälter (3) auf seiner Austrittsseite und der zweite Behälter (6) auf seiner Eintrittsseite direkt mit dem Meßkopf (4) verbunden und nur durch diesen voneinander getrennt, und bilden zusammen eine integrierte Einheit, welche als ganzes in diese Einrichtung einsetzbar bzw. aus ihr herausnehmbar ausgebildet ist.

DE 3411936 A1

Fig 2 Fig 1

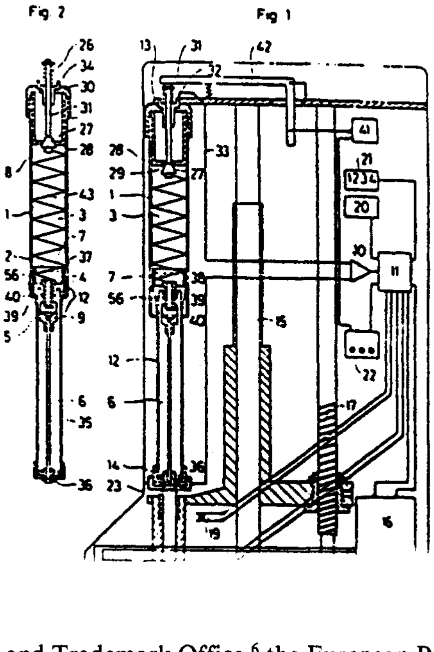

Figure 1. Front page of a German patent publication.

graphs will then be used for setting up a full-text database for retrieval and electronic delivery purposes as well as for computer-assisted composition of the printed publications.

The pilot project is funded in parts by the German Federal Government and the Commission of the European Communities[4] and it is carried out by a consortium consisting of the German Patent Office, the Satz-Rechen-Zentrum Hartmann + Heenemann, the Gesellschaft für Information und Dokumentation, and the Fachinformationszentrum Karlsruhe, which is the project leader and is responsible for the technical development.

The work of the consortium is part of the plans of the German Patent Office in improving patent information by means of electronic data processing.[5] These plans go along with the plan and the tripartite arrangements of the U.S. Patent and Trademark Office,[6] the European Patent Office, and the Japanese Patent Office. These patent offices plan to support their tasks by office automation and transition from paper-oriented work to electronic means. That means set up and use of full-text files for retrieval and display purposes and in-house facsimile files of all patent and nonpatent documents to be considered for examination. In consequence, this input also is the basis for computer-assisted information services to the public. Regarding the European situation, the plans of the European Patent Office as well as of the leading European nations have to be considered. There are some ideas at the patent offices to set up a European patent database in cooperation with the European Patent Office and the nations. A German full-text database extended to the German language area including Germany, Austria, and Switzerland would be a valuable input to a cooperative multilingual European patent database.

The German patent database will be made available to the Patent Offices and the public for online access by Fachinformationszentrum Karlsruhe via STN International, the world-wide International Information Network for Science and Technology. FIZ Karlsruhe has considerable experiences in scientific and technical information services and host services. In cooperation with other database suppliers, it offers online access to some 60 bibliographic or factual databases via INKA Online Service, and it is the partner of Chemical Abstracts Service in setting up and operating STN International. STN because of its decentralized database management and its common services is an ideal concept for a world-wide patent information service based on national or regional databases. Furthermore, it will offer online access to a broad spectrum of literature and factual databases for cross-file searching. STN is operated with Messenger, a database software developed by Chemical Abstracts Service and extended step by step to an integrated system for bibliographic, factual, and full-text databases. The specific features of the German patent information project for handling and transmitting text and graphical data will be realized within Messenger by 1986, and at the same time a convenient full-text handling and retrieval software will be available.

So, the online information services based on the German patent database will be extended as follows: from access to bibliographic data and abstracts and access to the citations of the search report, available at the end of 1985; via access to text and graphics of the front page of patent publications, available for test purposes in the beginning of 1986; via online ordering and delivery of the complete documents; to full-text searching and electronic delivery, as planned, started by a test phase in 1986.

Online patent information via STN International will be extended in 1987 by specific INPADOC services, by the WIPO Abstracts (World Intellectual Property Organization), and by the files of the European Patent Office. Cross-file searching of numerous bibliographic databases in science and technology supplements the direct access to the state of the art.

DEVELOPMENTS IN INFORMATION TECHNIQUES

Concept. In contrast to most other projects, the technical progress within this patent information project will not be based on facsimile transmission but on a novel concept for storage, transmission, and display of graphical information. The main developments in this field are conversion of the digitized graphical data from the raster images of the scanner into vector graphs, handling, storage, and transmission of both text and vectorized images, and terminal support, especially support of various types of commercially available terminals with text and graphics features promising a high user population.

A study comparing the raster and vector techniques, carried out in 1984, resulted in a decision in favor of vectorization.[7] Vectorization is feasible, but vectorization systems available had to be optimized to be fit for the purpose. Tests will run at FIZ Karlsruhe in September 1985. Developments in extending the database software for storing, processing, and transmitting drawings in vector format as well as developments in terminal support will be concluded at the end of 1985.

Raster to Vector Conversion. Like all other data stored and processed in and recalled from a computer system, drawings must be coded in bit form. Drawings are broken up into pixels of a raster image, and each element of a drawing is assigned a bit defining it as belonging to the drawing (black) or to the background (white).

Integration of binary raster images of line drawings as in the case of patent drawings in a data bank system makes the following requirements on data representation: (i) Drawings (together with the text of the patent document) must be stored and transmitted via the public data networks. The cost of storage and transmission is part of the overall system cost and should therefore be kept as low as possible, e.g., by efficient image-coding techniques. (ii) Processing of drawings involves input of drawings of different scale and output on displays and printers of different resolutions and formats. The adaptation of input and output rasters should take place in the receiving stations. There is a clear trend toward intelligent terminals with a processing function of graphical information. Input–output raster adaptation should be part of the image-decoding procedure as far as possible, and faithful reproduction of the images must be ensured.

Possible solutions in raster graphics and vector graphics meeting these requirements were analyzed and compared in detail within this project; the characteristics of system components developed by U.S. and European companies and institutions were analyzed and tested. Requirements for storing, transmitting via public telecommunication lines, and displaying the images were taken into account.

First, some general outlines concerning vector vs. raster graphics!

The primary reproduction for binary images generates bit maps of raster images. Most of the available devices for raster display of binary images (e.g., raster screens, matrix or laser printers) today use bit map memories. Modern personal computers make bit map memories accessible to the user via suitable interfaces, using the bit map memories both as refresh memories and as main memories for graphical image manipulation.

However, due to their enormous space requirements, bit maps are not suited for large-scale storage and teletransmission of binary images. For example, one DIN A4 page with a raster of 16 lines/mm covers almost 2 Mbyte in bit map representation.

Coding techniques for compression of binary images from the bit map representation have been developed for digital facsimile reproduction.[8] There are one-dimensional facsimile codes that only group image elements within one line and two-dimensional facsimile codes grouping image elements of several lines. Regarding the storage capacity, two-dimensional compression has advantages, but that means additional processing.

As an alternative to the raster compression techniques, there are techniques reducing binary images to their geometric information. The information contained in a picture is represented by broken lines following either the center lines or the outlines of the drawings. These techniques are commonly referred to as vectorization techniques as they mainly consist of position vectors of the corner points of polygon lines.

There are various vectorization procedures due to the requirements to be fulfilled. In cases where line widths are not important, e.g., for geographical maps, skeleton images using center-line vectors by binary thinning operations are sufficient. This procedure, which requires considerable processing time, can be avoided by outline vector imaging. In this technique, the outlines of image objects are directly represented by closed lines either going through the edge points of image objects or following the cracks of the image objects. Outlines as well as center lines can be represented either by chain codes or by approximation.[9-13]

Advantages of vector graphics vs. raster graphics are as follows: (i) Vectorization especially when combined with the techniques for broken line approximation in digital images[9,10] leads to a higher compression than can be achieved by the most efficient raster compression techniques; thus, about 20% less storage space is required (important for storage on floppy disks), and less data have to be transmitted; both resulting in less costs. Vectorization of a simple drawing results in approximately 4000 vectors, while a complex drawing will have about 20 000 vectors. It is expected to reduce the number of bits to about 30 Kbytes (from 2 Mbytes). (ii) Transfer of image vectors into a bit map for display is a simple operation, while with raster formats it is difficult to adapt the number of pixels transmitted from the host to the raster format of the display terminal. (iii) Rescaling is possible in the vector format without loss of quality because image vectors group image points of geometrical relevance. (iv) Graphic data can be manipulated and processed locally when in vector format, provided appropriate software is available on the micro. (v) Zooming is easier to do and without loss of information.

The systems analyzed so far resulted in a decision in favor of vectorization and in requirements to be made on an ideal vectorization system for the present project task.[7] An optimum method of raster image vectorization for storage in a data bank and output on displays and printers of different resolution should have the following features: (i) Conversion of the original image into closed polygon lines following the contours of the original image. (ii) Reduction of the number of corners of the polygon lines by subsequent line approximation. The permissible tolerance of this processing step is the only parameter of the whole vectorization method. (iii) Chain coding of the resulting polygon lines in a clear-text code in standardized format and storage in the data bank. Storage requirements should be less than those with two-dimensional facsimile codes. For a DIN A4 engineering drawing, the whole process should not take longer than 5 min on a 1-Mips computer. (iv) Outline images are to be preferred to center-line images. They avoid artifacts resulting from quantization of line widths and give a faithful representation of solid-area features. (v) Use of standards in the field of computer graphics at the host as well as at the terminal side.[14,15] The purpose is the decoupling of graphical application programs from graphical systems and hardware knowledge and the detachment of these graphical packages from their device by drivers. In the field of two-dimensional images, a suitable standard is already available by the Graphics Kernel System (GKS), developed in the Federal Republic of Germany and internationally accepted as an ISO Standard and probably in future also as an ANSI Standard in the U.S. With the introduction of so-called metafiles as a concept for reproduction of images according to graphics standards, standardized representation of graphical data structures became possible. Standardized vector image representations and standardized functional interfaces have freed the storage and distribution systems from the necessity of internal reconversion of symbolic vector images into concrete images. (vi) Transmission of metafiles by using the features polyline approximation, chaincoding, and filling, in standardized form, to all graphics work stations connected with the host computer. Metafile interpreters, which are standardized except for their scaling factors, serve to decode the input data and transmit them to resident GKS systems via suitable function references.

Systems analysis and tests with operational vectorization systems had the results that three systems are principally suitable (two in U.S., one in Germany), but all the systems have to be optimized for the purpose to automatically vectorize complex patent drawings. Test versions will be submitted to FIZ Karlsruhe until Summer 1985. On-going developments including polygon filling will be concluded at the end of this year.

Graphics Handling, Storage, and Transmission. The Graphical Data Structure System (GDS) of the STN Messenger software has to be extended for handling, storing, and transmitting the patent drawings in vector format. CAS will support the development and extend its GDS system, allowing the storage of outlines as chain codes. Each vector coordinate will be ASCII encoded. So, transmission of graphics will be done in ASCII characters like text, indicating the beginning and the ending of graphics transmission by special keywords.

A test version will be available at the end of 1985. Text and graphics will be transmitted via the public telecommunication networks. The only problems are caused by the current low transmission speeds. But the transmission times will be reduced considerably when the Integrated Services Digital Network (ISDN) at 64 kbit/s will be available in various countries within the next years.

Terminal Support. It is planned to support a large variety of intelligent terminals with graphic features, in particular those that promise a big market share.

A study on the operating systems used on micros show that MS-DOS, CPIM, and UNIX cover 85% of the market; the project will aim at supporting all micros that run under these operating systems provided they have a sufficiently high resolution of at least 720 × 350 dots. Color is not necessary. Also, terminal printers with high resolution (>300 dots/in.) will be supported.

To display the graphics, a transformation from vector structure to raster structure is necessary. This will be done locally by the user system.

Software for this process already exists. Until the end of 1985, communication software and administration software for various micros will be developed by FIZ Karlsruhe so that these can display polygons with filling.

CONCLUSIONS

Summarizing, we can state that the plans and developments in electronic publishing, dissemination, and delivery of German patent information go along with the lines of corresponding developments in leading industrialized countries and will be the basis for international cooperation. A novel technical concept for storing, transmitting, and displaying text and graphical information will be realized. It is expected that in this case electronic means will be of benefit to a broad user community for quicker, direct, and improved access to patent information and for better utilization of this important information source than in the past.

REFERENCES AND NOTES

(1) Kaback, S. M. "Online Patent Information. Patents Online: Where Do We Stand?". *World Pat. Inf.* **1983**, *5* (4), 239–240.
(2) Walker, R. "Patents as Information—An Unused Resource". *IFLA J.* **1984**, *10* (2).
(3) Tittlbach, G. "Electronic Publishing of Patent Information". "ICSU AB/CEC Joint Conference on Electronic Document Delivery and Electronic Publishing", June 1984, Luxembourg.
(4) Vernimb, C.; Leamy, C. C. "The CEC Plans for Electronic Publishing and Document Delivery". "Proceedings of International Online Information Meeting"; 1982; pp 351–360.
(5) Wittmann, A. "Die Technische Dokumentation des Deutschen Patentamts. Derzeitiger Stand, Problematik und Möglichkeiten der Weiterentwicklung". *Mitteilungen der Deutschen Patentanwälte* **1983**, *74* (3), 48–54.
(6) Bryant, J. H. "Automated Patent Searching: Preliminary Results of USPTO Studies". *World Pat. Inf.* **1983**, *5* (4), 226–229.
(7) Speck, P. T. "Integration of Drawings in a Databank System. System Planning and Conversion of Raster Images into Vector Graphs"; Fachinformationszentrum Karlsruhe: Eggenstein-Leopoldshafen, FRG, 1985; FIZ-P-1/3, Internal Report, April 1985.
(8) Hunter, R.; Robinson, A. M. "International Digital Facsimile Coding Standards". *Proc. IEEE* **1980**, *68*, 854–867.
(9) Wall, K.; Danielson, P. E. "A New Method for Polygonal Approximation of Digitized Curves". *Comput. Vision Gr. Image Process*, in press.
(10) Williams, C. M. "Bounded Straight-Line Approximation of Digitized Planar Curves and Lines". *Comput. Gr. Image. Process.* **1981**, *16*, 370–381.

(11) Speck, P. T. "Ubersetzung von Linien- und Flächenstrukturen in kombinatorisch-relationale Datenstrukturen zur automatischen Mustererkennung in Digitalbildern". Dissertation 7508, ETH Zürich, 1984.
(12) Danielsson, P. E. "An Improved Segmentation and Coding Algorithm for Binary and Nonbinary Images". *IBM J. Res. Develop.* **1982**, *26*, 698–707.
(13) Ramachandran, K. "Coding Method for Vector Representation of Engineering Drawings". *Proc. IEEE* **1980**, *68*, 813–817.
(14) "Status Report of the Graphic Standards Planning Committee". *Comput. Gr.* **1979**, *13*.
(15) Enderly, G.; Kansy, K.; Pfaff, G. "Computer Graphics Programming. GKS—the Graphic Standard"; Springer: Berlin, 1984.

The Chemical Reactions Documentation Service: Ten Years On†

ALAN F. FINCH

Derwent Publications Ltd., Rochdale House, London WC1X 8RP, England

Received August 12, 1985

The philosophical basis, historical development, and essential features of Derwent's Chemical Reactions Documentation Service are outlined with particular reference to the monthly *Journal of Synthetic Methods*, the link with Theilheimer's *Synthetic Methods of Organic Chemistry*, and the two computer-based systems used for retrieval.

During the last 4 decades, William Theilheimer's annual volumes of *Synthetic Methods of Organic Chemistry*[1] have afforded the practicing chemist direct and comprehensive access to the key organic reactions reported in the literature. With the active support of a team of technical advisors in Europe and the U.S., the editor of this renowned series guided the first 35 volumes through a period when synthetic chemistry underwent radical change and development. On Theiheimer's retirement in 1981, the responsibility for producing the yearbooks was passed to Derwent Publications Ltd., which some 5 years earlier had introduced the monthly *Journal of Synthetic Methods* (JSM)[2] as an adaptation of the yearly volumes. Today JSM provides the input for *Synthetic Methods* and is at the same time the cornerstone of Derwent's Chemical Reactions Documentation Service (CRDS). In this paper, the philosophical basis, the historical development, and essential features of this service are described with an illustration of retrieval capabilities from the current database of over 60 000 reactions.

PHILOSOPHY, SCOPE, AND COORDINATION

The growth of the scientific literature should ideally be matched by the development of efficient means of accessing the information contained therein. For retrieval of specific compounds and compound classes, this has been largely the case. For chemical reactions, however, the development of retrieval methods progressed more slowly, and it was not until Theilheimer came on the scene that a comprehensive and analytical system for collating, indexing, and presenting synthetic methods became available. That this system and the philosophy behind *Synthetic Methods* have remained basically unchanged through the years is validation in itself, and with current refinements, computer-based equivalents, and extensions in JSM, it is enjoying a new lease of life and recognition.

For both JSM and *Synthetic Methods*, selection of material is still the most critical phase, as may be appreciated from the crude representation of the chemical literature as shown in Figure 1. In this analogy with an "inverted iceberg", the bulk of synthetic organic chemistry—pertaining largely to applications of known methods and specialized fields of chemistry—is clear to view and readily accessible by conventional retrieval techniques and handbooks. Much less so, however, are the data beneath the surface: namely, the 2000–3000 new synthetic methods published each year, plus—in the darkened area—approximately the same number of additional publications describing minor modifications of known methods. It is the material in these two lower layers that is presented in JSM: the former as informative abstracts on such aspects as novel functional group and ring transformations of a general nature, new reagents and synthetic techniques, and new syntheses of key ring systems; and the minor modifications as additional data published biannually in *Supplementary Reference Indexes*.

Reactions are selected for JSM by scrutiny of the worldwide journal and patents literature and are usually published some 6–8 weeks from receipt of the primary documents. The novelty of the data is confirmed by a prior analysis, or coordination, with the current CRDS database, which comprises over 60 000 key reactions from *Synthetic Methods* Volumes 1–30 (containing information from 1942 to 1975) and JSM itself (1975 to date). Such coordination is also necessary in order to identify the original methods to which to append the supplementary references. In this way, key synthetic items are located quickly and available in a compact form ideal in size to allow rapid scanning on a monthly basis and with cumulated subject indexes for retrospective searching.

FORMAT OF THE JOURNAL OF SYNTHETIC METHODS

The structure of the journal is similar to that developed by Theilheimer for *Synthetic Methods*. The abstracts, illustrated in Figure 2, feature a Derwent-assigned title, a typical experimental procedure, and, most important for scanning purposes, a clearly expressed reaction scheme. The reactions, numbering 250 in each monthly issue, are arranged according to Theilheimer's **reaction symbol notation**, the systematic classification given at the top of the abstract, so that all the reactions of a given type are found together.

The entries in the *Subject Indexes* (Figure 3) include starting materials, products, reagents, and any other synthetic items considered relevant. Where appropriate, a dynamic term, such as C-alkylation, may be used but the indexing normally takes the form "A starting material for B" and "B from A". Systematic nomenclature is used throughout and supported by a unique cross-reference network, so critical for comprehensive retrieval. This comprises three essential elements: synonyms, cross-referencing to related concepts, and

† Presented at the 188th National Meeting of the American Chemical Society, Philadelphia, PA, Aug 30, 1984.

0095-2338/86/1626-0017$01.50/0