Frequency of Titles Containing "DOPA-words" in a Complete Collection of Published Documents on DOPA (3,4-Dihydroxyphenylalanine)

DONALD A. WINDSOR
The Norwich Pharmacal Co.
P. O. Box 191, Norwich, N. Y. 13815

Received July 30, 1971

Seventy-one per cent of the documents in a complete collection of publications on DOPA and dopamine in mammalian systems had "DOPA-words" in the titles.

The use of keywords in titles for the selection of pertinent documents is a commonly used practice,[3] ranging from the scanning of tables of contents in primary journals directly or in *Current Contents*[8] to the reading of titles under index terms, as in *Index Medicus*,[9] to the use of KWIC (key word in context) indexes,[4] such as in *Biological Abstracts*[17] or in *Chemical Titles*.[6]

Resnick found that titles were as effective as abstracts for alerting purposes[11] and Saracevic found that seventy-eight per cent of the time titles were just as effective as abstracts or even full texts in conveying relevance to readers[12] but in only half of Papier's cases did the readers assign the same title as the authors.[10] Tell's experiments showed that greater indexing consistency was achieved with titles than with abstracts or full texts.[14]

Titles as generated by their authors are usually not very useful without some additional manipulation. Smith increased the recall of titles from *Biological Abstracts* by using root character strings[13] while Flury and Henderson obtained successful retrieval with titles only after enriching them.[5]

After testing *Chemical Titles*, *Chemical-Biological Activities*, and *Automatic Subject Citation Alert*, Abbot, Hunter, and Simkins did not feel that a pure title search would ever be sufficient.[1] However, Tocatlian found that the titles of chemical papers are becoming more informative.[15] Garfield has wisely cautioned that the information content of titles must be judged in the context of the other citation parameters, such as the author and the journal.[7]

To evaluate the effectiveness of a keyword in the title search it is necessary to examine a collection already known to contain all of the documents on a specific subject. Fortunately, we already have a complete collection of published papers on DOPA (3,4-dihydroxyphenylalanine), and the opportunity to conduct such an investigation was seized.

MATERIALS

For over two years, we have been acquiring all published papers on DOPA and dopamine in mammalian systems to support a research program in the use of levodopa (L-DOPA) for the treatment of parkinsonism.[2] The papers are cited appropriately[16] and are indexed in depth. At the time of this study, there were 1310 documents in the collection.

METHODS

Each citation was examined visually for at least one of these "DOPA-words": DOPA; Dopamine; 3,4-Dihydroxyphenylalanine; 3,4-Dihydroxyphenethylamine; 3-Hydroxytyramine; 3-Hydroxytyrosine; L-DOPA; Levodopa; Laevodopa.

RESULTS

Of the 1310 papers in the collection, 932 had at least one "DOPA-word" in the title for a frequency of 71%.

CONCLUSION

Our collection of published papers on DOPA-dopamine in mammalian systems would be 71% complete if it were constructed only on the basis of "DOPA-words" in the titles.

Increasing the keywords to include such relevant concepts as "parkinsonism" or "catecholamine" would, undoubtedly have uncovered many of the papers in the 29% category. However, the number of false drops would have been much too high to tolerate.

Had our collection been acquired on the basis of a title search alone, most of the remaining papers would have been detected eventually from references and from secondary journals. Nevertheless, when time is important and it is essential to know about a paper as soon as it is published, there is no substitute for in-depth journal scanning. However, the results reported here clearly indicate that more reliance can be made on titles and the amount of in-depth scanning can be greatly reduced.

LITERATURE CITED

(1) Abbot, M. T. J., Hunter, P. S., and Simkins, M. A., "Current Awareness Searches on *CT*, *CBAC* and *ASCA*," *Aslib Proc.* **20**(2), 129–43 (1968).

(2) Barbeau, A., and McDowell, F. H., Ed., "L-DOPA and Parkinsonism," F. A. Davis Co., 433 pp., Philadelphia, Pa., 1970.

(3) Black, J. D., "The Keyword; Its Use in Abstracting, Indexing and Retrieving Information," *Aslib Proc.* **14**, 313–21 (1962).

(4) Fischer, M., "The KWIC Index Concept: A Retrospective View," *Amer. Doc.* **17**, 57–70 (1966).

(5) Flury, W. R., and Henderson, D. D., "A User Oriented KWIC Index; KWOC-ed, Tagged and Enriched," *Proc. Amer. Soc. Inform. Sci.* **7**, 101–5 (1970).

(6) Freeman, R. R., and Dyson, G. M., "Development and Production of *Chemical Titles*, a Current Awareness Index Publication Prepared with the Aid of a Computer," *J. Chem. Doc.* **3**, 16–20 (1963).

(7) Garfield, E., "Methods and Objectives in Judging the Information Content of Document Titles," *J. Chem. Doc.* **10**, 260 (1970).

(8) Garfield, E., "*Current Contents* receives First Information Industry Hall-Of-Fame Award—and Publishes Its 700th Consecutive Issue," *Current Contents* **14**(23), 5 (1971).

(9) National Library of Medicine, "Guide to MEDLARS Services," P. H. S. Publ. No. **1694**, 22 pp., 1967.

(10) Papier, L., "Reliability of Scientists in Supplying Titles: Implications for Permutation Indexing," *Aslib Proc.* **15**, 333-7 (1963).
(11) Resnick, A., "Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents," *Science* **134**, 1004-6 (1961).
(12) Saracevic, T., "Comparative Effects of Titles, Abstracts and Full Texts on Relevance Judgments," *Proc. Amer. Soc. Inform. Sci.* **6**, 293-9 (1969).
(13) Smith, P. P., "Increasing Recall of Keyword Indexes by Automatically Indexing Root Character Strings," *Ibid.*, **7**, 275-7 (1970).
(14) Tell, B. V., "Document Representation and Indexer Consistency; A Study of Indexing from Titles, Abstracts and Full Text using UDC and Keywords," *Ibid.*, **6**, 285-92 (1969).
(15) Tocatlian, J. J., "Are Titles of Chemical Papers Becoming More Informative?," *J. Amer. Soc. Inform. Sci.* **21**(5), 345-50 (1970).
(16) Windsor, D. A., "A Format for the Standardized Citation of Both Published and Unpublished Documents," *Ibid.*, **21**(3), 195-7 (1970).
(17) Zabriskie, K. H., and Farren, A., "The B. A. S. I. C. Index to *Biological Abstracts*," *Amer. J. Pharm. Ed.* **32**, 189-200 (1968).

Todai Scientific Information Retrieval (TSIR-1) System. I. Generation, Updating, and Listing of a Scientific Literature Data Base by Conversational Input

TAKEO YAMAMOTO,* TOMOKO KUMAI, KEN-ICHI NAKANO, CHIKATAMI IKEDA, TOSIYASU L. KUNII, HIDETOSI TAKAHASI, and SHIZUO FUJIWARA

The University of Tokyo, Hongo, Tokyo, Japan

Received August 30, 1971

A file structure for scientific literature data, STF, to be used in a scientific information retrieval system, TSIR-1, was developed by modifying the CAS Standard Distribution Format (SDF). The use of STF allows one to merge data records originating in CAS and those generated locally. A set of programs for the generation, updating, and listing of a scientific literature data base in STF from a TSS terminal in conversational mode is described.

For a scientist who wants to accumulate a data base of scientific literature for his own use and for the use of his scientific community, the CAS SDF[1, 2, 3] file structure has several attractive characteristics. It is a highly flexible format and is largely based on the natural language in its expression of the content of the literature. However, it still leaves much to be desired for generating new data records or for data records obtained from other sources than CAS as input to the data base.

We have been building a scientific information retrieval system (TSIR-1) using a HITAC 5020 TSS[4, 5, 6] of the Computer Centre, University of Tokyo. The TSIR-1 system is based on a data structure named STF (for *S*implified *T*odai *F*ormat, Todai being the abbreviated Japanese word for the University of Tokyo). STF is closely related to SDF, and its logical content is similar to that of SDF except for a number of modifications. Although more detailed description of STF will be published elsewhere, significant modifications follow.

In STF, all keywords corresponding to an article item are combined into one record element, and the element is included in the same logical record as the other data related to the item. Thus, each logical record corresponds to one article except for the case when the data are too great for one logical record (maximum record length: 3520 characters).

The text of the abstract and the full text of an article are given their own identification (ID) numbers and may be recorded in STF files.

Both Japanese (Romanized) and English entries are allowed for natural language input. The search programs and queries are expected to take care of the occurrence of the two languages in the retrieval phase.

The SERIAL NUMBER data element, ID number 0012 01, and the Temporary Abstract Number (TAN) data element, ID number 0054 01, in SDF[2] are modified to allow the use of these data elements for the file key in STF. The modification allows one to distinguish, if necessary, between the locally generated data records and records originating in CAS SDF files, after they have been merged into a unified data base. Thus, the compatibility as well as the distinguishability of the two kinds of records are obtained at the same time.

Only capital letters are used in the alphabet. This restriction was necessary because our input-output facilities do not include the lower-case alphabet.

In the rest of the present paper, a set of programs for the generation, updating, and listing of STF data as TSS disc files by conversational input will be described. The programs were written to help TSS users to generate and maintain their own literature files, without having a detailed knowledge of SDF or of STF.

The programs were written in PL/IW language,[5, 6] a subset of PL/I. They were compiled and executed by the HITAC 5020 TSS.

*To whom correspondence should be addressed at Department of Chemistry, Faculty of Science, the University of Tokyo, Hongo, Tokyo 113, Japan.