

All topics are discussed in such a way that it is possible to receive benefit from the reading of the essays without too much of theoretical background.

Rolf G. Henzler

SEELBACH, Dieter: **Computerlinguistik und Dokumentation**. Key-Phrases in Dokumentationsprozessen. = UTB Uni-Taschenbücher 468. München: Verlag Dokumentation 1975. 151 p. ISBN 3-7940-2643-8

Gibt es hier eine Antwort auf die Frage, ob Ergebnisse linguistischer Forschungen gegenwärtig eine Chance haben, in Anwendungsbereiche der elektronischen Datenverarbeitung einzudringen, also über ein experimentelles Stadium hinauszukommen? Die Dokumentation, hier vor allem der Problemkreis der Dokumenterschließung (Indexing), ist bestimmt durch einen hohen Zeit- und Kostenaufwand für die intellektuelle/manuelle Dokumentdeskription. Linguistische Verfahren sollen daher eine Automatisierung unterstützen oder gar erst ermöglichen. (Ob allerdings die Kosten sich reduzieren lassen, ist noch offen).

Der Verf. stellt ein Verfahren vor, durch das die noch wenig zureichenden bisherigen Methoden – meist auf der Vergabe von Einzel-Schlüsselwörtern (= Key-Words) beruhend – durch die automatische Zuweisung mehrgliedriger Nominalsyntaxen (= Key-Phrases) präzisiert und damit die Retrieval-Ergebnisse verbessert werden sollen. Maschinell erkennbare Text- (Satz-) Segmente werden automatisch in standardisierte Paraphrasen transformiert, d. h. in die Form komplexer, i. a. mehrere Präpositionen und Nomina enthaltender Nominalgruppen (z. B. CONSUMPTION OF BEER IN YUGOSLAVIA) überführt. Die Verwendung der expliziteren Präpositionalgruppenform soll dabei Mehrdeutigkeiten vermeiden helfen:

WATER TREATMENT → TREATMENT OF/WITH WATER
Trotz der vielfältigen Analyseprozeduren – durch eine Reihe von Tabellen und Listen ist die große Praxisnähe der Untersuchung dokumentiert – wird eine günstige Kosten-Nutzen-Relation angestrebt. Gemessen an der geringen Tiefe linguistischer Operationen (es wird mit Wortbildungsregeln und syntaktischen Oberflächen-Transformationen/Permutationen Harris'scher Art gearbeitet) erscheint die Effizienz beachtlich.

Angestrebt ist neben einer Verbesserung der (automatischen) Deskriptorenvergabe (Schlüsselwortvergabe) zu Dokumenten mithilfe der syntagmatischen Verknüpfung von Einzeldeskriptoren die Herstellung benutzerfreundlicher (gedruckter) Register (Verbesserung von KWIC- und KWOC-Verfahren). Dabei werden im Text vorkommende Begriffe nach einer morphologischen Suffix-Analyse auch semantisch grob nach „Vorgang“ und „Eigenschaft“ zu klassifizieren versucht. Auf eine Auseinandersetzung mit ähnlichen, wenn auch eher intellektuell orientierten Verfahren in der Dokumentation (und entsprechenden Literatur) wird aber weitgehend verzichtet. In einem Ausblick wird auf eine mögliche Anwendung des vorgestellten Regelsystems auf die maschinelle Übersetzung von Key-Phrases hingewiesen, die die noch halbintellektuellen Abstract-Übersetzungen nach Art des TITUS-Systems im Bereich der Key-Phrases ablösen helfen sollen.

Wenn auch das spezifische Thema nur als Exemplum für die Problematik der Automatisierung von Dokumentationsprozessen gelten kann, wenn auch der mittlerweile weitgespannte Bereich der Computerlinguistik (von der Wortformenreduktion bis zu sprachlichen Aspekten der künstlichen Intelligenz) hier nur ausschnitthaft (vorwiegend syntaxbezogen) angesprochen ist: die Erläuterungen zu den verschiedensten Bereichen in Linguistik, Datenverarbeitung und Dokumentation in einer gewissen Zusammenschau **rechtfertigen** vielleicht den allgemeinen Obertitel, wenn es auch nicht beabsichtigt erscheint, durch diese Hinweise eine grundlegende Einführung in den Bereich der Computerlinguistik und Dokumentation zu ersetzen.

Fazit: Ausführliches (Tabellen, Listen, Regelsysteme) Material, praxisnahe Resultate, weiterführende Vorschläge, die weitgehend realistisch, d. h. anwendbar scheinen. Doch noch keine schlüssige Antwort auf die Ausgangsfrage: es fehlen die „harten“ Zahlen einer Effizienz-Untersuchung. Welcher Anwender nutzt dies Möglichkeiten aus?
Harald Zimmermann

INTERNATIONAL FEDERATION FOR DOCUMENTATION: **Cumulated UDC Supplement 1965–1975**. Vols. I–V. The Hague, Netherlands: FID 1976. 548 p., hfl. 250,- (each vol. hfl. 50) = FID 543; ISBN 92-66-00543-6 (for the 5-volume set).

In addition to the P-notes, containing the proposals for new or revised UDC-notations, and issued in about 30–40 notes each year since 1931, the FID Classification Department issues “Extensions and Corrections to the UDC” ever since 1951 in a semi-annual mode and cumulates these every three years so that the final issue of a three-year-period replaces the foregoing 5 ones.

The cumulation under review contains all the alterations of the last 10 years, that is the cumulations from the Series 6, 7, 8 and No. 1 of Series 9 of the Extensions and Corrections to the UDC. In 1967 a similar ‘Cumulated UDC Supplement’ had been issued in 6 volumes, 795 pp., covering the changes from 1954 until 1964. Thus, the present volumes are the natural continuation of this practice.

The coverage is as follows: Vol. I: Auxiliaries + Classes 0/3; Vol. II: Classes 5; Vol. III: Classes 6/62; Vol. IV: Classes 63–676 and Vol. V: Classes 677/9.

Each volume can be obtained separately. The language is either English or German or French, depending on the language of the proposals accepted. Thus, very much effort has still to be put into the translation of the texts if new full editions in different languages are to be developed from this basis. From 1976 on the Extensions and Corrections to the UDC will only be published once a year.

It seems to be impossible to review any of the contents of the new tables, although the new numbers for “Classification and Indexing” (p. 70 of vol. I) (025.4.01/.49) which include also “Information retrieval techniques” (025.4.03) could be a very adequate “victim”.

Users of the UDC must welcome this new publication which does away with the looking-up of changes in previously 4 publications. For the rest this new cumulation may provide (1) an orientation on the enormous development of the UDC revision activities in the last years, (2) evidence of the new areas of interest and (3) also of dissatisfaction with the existing classes. A history of the additions to the UDC has still to be written. It would become a history of the UDC-users capacity to "mentally digest" the scientific and technical developments of the last 100 years and of their sometimes tantalizing struggle to relate the new insights to the knowledge of yesterday.

I. Dahlberg

HUTCHINS, W. J.: Languages of Indexing and Classification: A Linguistic Study of Structures and Functions. – Stevenage, Herts.: Peregrinus 1975. VIII, 148 p., ISBN 0 901223 689 = Librarianship and Information Studies, Vol. 3

Theoretical bases for classification and indexing have been drawn from several disciplines: the traditional library classifications were based on philosophical and logical grounds, later on mathematical and psychological bases were sought. In recent years the development of schemes for classification and indexing found source in general systems theory and in integrative level theory; now it appears that the theoretical developments in modern linguistics provide a basis for the structural and semantic aspects of classificatory and indexing languages.

Hutchins' book is an attempt to study the linguistic features of indexing and classification languages. He names them Documentary Languages (DL). The following aspects are covered: The place of DL within the context of communication systems (semiotic systems) in general and identification of the characteristic features of DL being different from natural and other artificial languages (Chap. 2). Within this framework, the formal structure (Chap. 3), sememic aspects for both, paradigmatic and syntagmatic relations (Chap. 4 and 5) and pragmatic factors (Chap. 6) influencing DL are discussed. The linguistic framework for the investigation is based on the work of Lamb (1966), Lockwood (1972), Halliday (1970, 1973), Chafe (1970), Mel'čuk and Žolzkovski (1970) and Hutchins (1971). It makes use of models such as transformational grammar wherever appropriate especially also of Fillmore's case grammar (Chap. 5). The linguistic processes involved in indexing documents (Chap. 7) and in searching document descriptions in index files (Chap. 8) are treated as well. Chapter 9 contains hypotheses on deep structures of DL and their relations to the universals of natural languages.

Every class number or index tag is a message carrier. It communicates the ranking of the subject embodied in a document in an ensemble of subjects in a collection of documents. The index tag acts as a surrogate through which a user will be able to interact with a collection in a library or with an information system. Hutchins identifies "the description of DL structures and functions as analogues and parallels to those of NL in its informative functions and in a broader context, as features of semiotic systems on three basic levels, formal, semantic and pragmatic".

The formal structure of a classificatory language is guided primarily by two principles in order to arrive at a sequence helpful to a majority of intellectuals: the 'Wall Picture Principle' (principle of dependence) and the 'Postulate of Decreasing Concreteness'. The facet syntax of class numbers may be compared with the case categories of Fillmore in a sentence structure. The semi-syntactic nature of a DL is given by its syntagmatic relation (in general – a relation between an element in a string and other elements appearing in the same string) and by its paradigmatic relation (in general – a relation between an element in a string and other elements not appearing in the string but which could have done so in the place of that element). Paradigmatic relations in index languages cover inclusion and/or associative relations. Syntagmatic relations are given in subject descriptions comprising one or more descriptor phrases which may have a single descriptor or more than one. In the latter case, the relation may be indicated by role indicators, relators, facet indicators or conjunction signs.

The devices have their analogues in Fillmore's case categories derived from the following eleven categories: patient, factive, agent, instrument, experience, beneficiary, source, goal, location, time and purpose. Each of them has a particular case relation to the 'verbal' elements: state, process, action and action-process. Hutchins identifies the similarities of categories in several documentary languages such as SYNTOL, EJC roles, Farradane's nine relators, faceted classification systems. He discusses the value and helpfulness of a DL with deep (sememic) case relations. The case categories may, however, also be seen to have their analogues in the Standard Citation Order postulated by Ranganathan and also by Vickery. Ranganathan called it 'Absolute Syntax' and he meant the sequence in which the component ideas of subjects in a certain subject-field arrange themselves in the minds of a majority of normal intellectuals, for instance, when they communicate about such subjects. The hypothesis of 'Absolute Syntax' may be regarded as a move in the search for universals for documentary languages. The developments in indexing languages such as POPSI and PRECIS, the recent trends towards faceted classifications and the growing demand for a deep structure for interconvertibility of indexing languages in the context of the evolution of global information networks have given scope for research in evolving an 'Absolute Syntax'. Therefore, research programs are being promoted to test the various models of Absolute Syntax on the basis of developments in the fields of communication science and educational psychology. Hutchins' book gives some penetrating insights into the deeper aspects of the semisyntaxics of indexing and classification.

The book is meant for specialists in indexing and classification particularly the designers of the same with no linguistic background. However, the linguistic jargon in the text complicates somewhat the understanding. In spite of this, Hutchins' book brings home the linguistic basis of classification and indexing with greater clarity than hitherto achieved.

M. A. Gopinath