

- (14) Padučeva, E. V.: The Language of Mathematical Logic as a Semantic Model for Natural Language. In: Social Science Information 7 (1968) pp. 27–39.
- (15) Padučeva, E. V.: Semantic Analysis of Natural Language During Translation into the Language of Mathematical Logic. (In Russian) In: Vsesojuznaja Konferencija po Informacionno – Poiskovym Sistemam i Avtomatizirovannoj Obrabotke Naučno – Techničeskoj Informacii. Moscow 1967. p. 156.
- (16) Shapiro, S. C., Woodmansee, G. H.: A Net Structure Based on Relational Question Answerer. In: Norton, Walker (Eds.): Proceedings of the International Joint Conference on Artificial Intelligence, MITRE Corporation, Bedford, Mass., 1969. pp. 325–346.
- (17) Blagden, J. F.: How Much Noise in a Role-Free and Link-Free Coordinate Indexing System? In: J. of Doc. 22 (1966) p. 203–209.
- (18) Cleverdon, C., et al.: Factors Determining the Performance of Indexing Systems. Cranfield, England: College of Aeronautics, A A L I B, Cranfield Research Project, 1966.
- (19) Costello, J. C.: Storage and retrieval of chemical research and patent information by links and roles in Du-Pont. In: Amer. Doc. 12 (1961) No. 2 p. 11.
- (20) Perry J. W., et al.: Machine Literature Searching, New York: Interscience Publ. 1956.
- (21) Costello, J. C., Wall, E.: Recent Improvements in Techniques for Storing and Retrieving Information. Wilmington, Del.: E. J. Du Pont de Nemours & Co. 1959.
- (22) Thesaurus of Engineering Terms. New York: Engineers Joint Council 1964.
- (23) Gardin, J. C.: Syntol. New Brunswick, N. Y.: Rutgers, the State University 1960.
- (24) Skorochoodko, E. F.: The Information Retrieval System at the Institute of Cybernetics, Academy of Sciences of the Ukrainian S. S. R. (In Russian). In: International Forum on Informatics, Moscow: VINITI 1969. Vol. II, p. 68–78 (A. J. 70.3.160).
- (25) Otradiskij, V. V., Stokolova, N. A.: On one Methodology of Construction of Information Retrieval Languages without Grammar. (In Russian). In: Naučno-Techničeskaja Informacija, ser. 2. (1968) No. 6, p. 14–18.
- (26) Preisler, W.: Thesaurusarten und Probleme ihrer Strukturierung. In: Informatik 20 (1973) No. 4, pp. 9–16.
- (27) Vickery, B. C.: Faceted Classification. London. A S L I B. 1960.
- (28) Lancaster, F. W.: Vocabulary Control for Information Retrieval. Washington. Information Resources Press 1972. 233 pp.
- (29) Kent, A., Perry, J. W.: Searching Metallurgical Literature. In: Casey R. S., et al.: Punched Cards: Their Application to Science and Industry. 2nd. ed., New York: Reinhold 1958, p. 248–260.
- (30) Perry J. W., Kent, A.: Tools for Machine Literature Searching Searching. New York: Interscience Publishers, Inc. 1958.
- (31) Aitchison, J., Gilchrist, A.: Thesaurus Construction. A practical manual. London: ASLIB 1972. 95 p.
- (32) Van Oot, J. G., et al.: Links and Roles in Coordinate Indexing and Searching: An Economic Study of their Use, and an Evaluation of their Effect on Relevance and Recall. In: J. of Chem. Doc. 6 (1966) p. 95–101.
- (33) Montague, B. A.: Testing, Comparison and Evaluation of Recall, Relevance and Cost of Coordinate Indexing with Links and Roles. In: Pro. Amer. Doc. Inst. 1 (1964) p. 357–367.
- (34) Sinnett, J. D.: An Evaluation of Links and Roles Used in Information Retrieval. Dayton, Ohio: Air Force Materials Laboratory, Wright-Patterson Air Force Base, 1964. A D 432 198.
- (35) Stokolova, N. A.: On a Technique of Evaluation of the Semantical Power of Information Languages with Grammar. (In Russian). Moskow. Naučnyj Sovet po Kompleksnoj probleme “Kibernetika”, Vsesojuznyj Seminar po Informacionnym Jazykam, 1971. vyp 3.
- (36) Vilenkin, N. Ya. Combinatorics. (In Russian). Moscow: Nauka 1969.
- (37) Simmons, R. F.: Natural Language Question-Answering Systems. In: Comm. ACM 13 (1970) No. 1, p. 15–29.

W.-W. Höpker  
Pathologisches Institut der Universität, Münster i. W.

## Struktur und Kompatibilität des Thesaurus der Medizin (Structure and compatibility of the Thesaurus of Medicine)

Höpker, W.-W.: Struktur und Kompatibilität des Thesaurus der Medizin. (Structure and compatibility of the Thesaurus of Medicine) (In German). In: Intern. Classificat. 3 (1976) No. 2, p. 81–84.  
The medical thesaurus described is an international, compatible thesaurus in the German language the structure of which is based upon both a hierarchical classification and a variable classification of facets. Both classification principles are shown by three digit letters and figures. The thesaurus is compatible with the Clinical Key of Diagnosis (KDS), the International Classification of Diseases (E) (ICD/E), and the Systematized Nomenclature of Pathology (SNOP). Its volume comprises 22,000 different terms.  
(Author)

### 1. Problemstellung

Die differenzierten Ansprüche, welche Benutzer an ein Dokumentationssystem stellen, finden zunächst ihren Niederschlag in einem entsprechenden Schlagwortverzeichnis, Sachkatalog oder – Thesaurus. Wir verstehen unter Thesaurus *ein Klassifikationssystem mit klartextlichem Einstieg* und legen – dem gegenwärtigen Stand automatisierter Dokumentationsvorhaben Rechnung tragend – das Schwergewicht auf Klassifikationssystem. Es zeichnet sich ab, daß sogenannte „Totallösungen“ der Dokumentationsfrage in der Medizin immer weiter in die Ferne rücken. Um so mehr muß angestrebt werden, mit den bisher verfügbaren Methoden und mit vertretbaren finanziellen Mitteln pragmatische Teillösungen zu realisieren, die dann als Ausgangspunkt für weitere Entwicklungen angesehen werden können. Wir sind Gegner von perfektionistischen Systemen, die „zu gut“ sind, als daß sie funktionieren können. Diese Systeme haben meist noch die Angewohnheit, niemals „fertig und funktionstüchtig“ zu werden. Aus diesen Gründen haben wir wenig Scheu, einen (in diesem Sinne nur halb-fertigen) Thesaurus vorzustellen, der noch nicht vollständig ausgetestet ist und noch sicherlich zahlreiche inhaltliche Fehler aufweist. Gesamtkonzept und Struktur haben sich jedoch bereits bewährt.

### 2. Voraussetzungen

Unsere Arbeitsgruppe ist von folgenden *Voraussetzungen* zur Realisierung eines international kompatiblen Thesaurus ausgegangen:

1. Ein medizinischer Thesaurus muß eine *Sammlung situationsgerechter Diagnose- und Krankheitseinheitsbegriffe mit ihrem klartextlichen Einstieg* enthalten. Die diagnostischen Begriffe beziehen sich auf die verschiedenen Ebenen der (meist unzureichend) klinischen Diagnosenfindung, welche nur in Ausnahmefällen identisch ist mit dem angestrebten Begriff der Krankheitseinheit. Über diese Diagnosebegriffe muß ein klartextlicher Einstieg realisiert sein.

2. Die *medizinische Terminologie* muß enthalten sein. Eine solche ist weder für den internationalen noch für den deutschsprachigen Bereich bindend definiert. Es ist daher anzustreben, sämtliche nur möglichen Begriffe einer solchen Terminologie implizit in einem Thesaurus darzustellen.

3. Das vermutete (wenn auch nicht immer korrekte) *Vorwissen des Benutzers* sowie dessen Verhalten gegenüber einem Ordnungssystem sind oberste Kriterien für die Codierung und das Retrieval. Ein Thesaurus, der sich in der Praxis zu bewähren hat, hat sich *nicht* in erster Linie an den Klassifikationskriterien der gängigen Hand- und Lehrbücher zu orientieren.

Im Vordergrund hat die pragmatische Intention des Benutzers zu stehen, möglichst schnell und in assoziativer Denkweise Informationen zu speichern bzw. speichern zu lassen und diese nach bestimmten Gesichtspunkten wiederzugewinnen.

4. Das *Gesamtsystem muß für den Benutzer transparent* sein. Die einzelnen Stationen der Informationsverarbeitung als auch der bedeutungsmäßigen Zergliederung der Begriffe müssen für den Benutzer nachvollziehbar sein. Auch diese Vorgänge haben nach pragmatischen, nicht vordergründig nach systematischen Gesichtspunkten zu erfolgen.

5. Der Thesaurus soll *von einem Dokumentationssystem unabhängig* sein. Dies betrifft einerseits die technischen Realisierungsmöglichkeiten, andererseits die methodischen Ansprüche des Benutzers; letztlich müssen auch mögliche zukünftige Innovationen auf zahlreichen, nicht explizit definierten Informationsebenen berücksichtigt werden.

6. Der *Gültigkeitsbereich eines Thesaurus muß fachübergreifend* angegeben werden. Selbst für eine medizinische Spezialdisziplin reicht es nicht aus, die diagnostischen Begriffe ausschließlich dieses Faches darzustellen. Es ist daher notwendig, sämtliche Fächer synoptisch abzubilden.

7. Ein Thesaurus soll nach Struktur und Umfang eine hohe *Flexibilität* aufweisen. Dies bedeutet nicht nur Anpassungsfähigkeit gegenüber möglichen Umfangsänderungen, sondern auch die Möglichkeit, die gesamte Struktur kurzfristig nach nicht vorhersehbaren Gesichtspunkten (meist inhaltlicher, aber auch intern-organisatorischer Natur) zu ändern.

8. *Kompatibilität* zu den großen internationalen Schlüsselssystemen muß gegeben sein. Dieses ist nicht nur aus Gründen einer möglichen Zusammenführung von internationalen Daten, sondern auch im Hinblick auf eine einheitliche Terminologie und Klassifikation nötig.

9. Ein Thesaurus muß nahtlos an mindestens eine *Schnittstelle der automatisierten Klassifikation* anschließbar sein. Diese Schnittstellen sind dort zu suchen, wo die kleinsten Informationseinheiten der Gliederungsstruktur anzutreffen sind. Diese Einheiten müssen so konstruiert sein, daß sie über einen längeren Zeitraum ihre Gültigkeit behalten. Sie betreffen daher die einfachsten pathologischen Vorgänge und deren Verknüpfungen untereinander.

10. Realisierung von *Anschlußmöglichkeiten an Systeme mit automatisierter Diagnosenhilfe*. Dieser Bereich ist sehr viel schwieriger zu realisieren. Insbesondere wird hier deutlich, daß mögliche Variable einer automatisierten Diagnosenhilfe zu einer anders gearteten begrifflichen Abgrenzung des resultierenden Diagnosesystems führen.

### 3. Realisierung

Der „Thesaurus der Medizin“ gliedert sich in vier Dimensionen: Topographie, Nosologie, Ätiologie, Modifikation. Jede dieser Dimension weist eine streng-hierarchische Gliederung auf, wobei die Klassen durch mnemotechnische Kürzel gekennzeichnet sind. Der Umfang der hierarchischen Klassen innerhalb der einzelnen Dimensionen schwankt zwischen vier (Modifikation) und 88 Unterklassen (Topographie) (Abb. 1). Diese Gliederung entspricht mehr oder weniger dem traditionellen Aufbau medizinischer Schlüsselssysteme. Jede dieser Klassen wird durch einen dreistelligen Buchstabenkürzel gekennzeichnet, welcher eine mnemotechnische Struktur hat (z. B. Herz: HRZ; Magen: MAG). – Bereits im Laufe der Vorarbeiten hatte sich gezeigt, daß die strenge hierarchische Klassifikation auch im Falle der Polyhierarchie der komplexen Struktur der medizinischen Diagnosesprache nicht gerecht werden kann. Es wurde daher eine zusätzliche Facettenklassifikation eingeführt. Das Ziel dieser Facettenklassifikation ist, die Intensität der einzelnen Begriffsbenennungen zu steigern und somit den semantischen Bedeutungsinhalt derselben auf Kosten der syntaktischen Information zu vergrößern. Dieser Aspekt ist insbesondere für Prozeduren der automatisierten Klassifikation von großer Bedeutung. Rein pragmatische Gesichtspunkte sowie insbesondere der Stand der gegenwärtigen Entwicklung auch komplexer Dokumentationssysteme haben mit zunehmender Konsequenz an diesem Schritt festhalten lassen. Informationstheoretische Erwägungen sowie Berechnungen der Entropie und Redundanz an medizinischen Schlüsselssystemen sind der formale Einstieg in die Struktur des TdM (Höpker, 1972; Kayser et al., 1974).

Die Facettenklassifikation wird in eine Differentialfacette sowie eine Segmentalfacette untergliedert. Das Gliederungsprinzip der Differentialfacette ist die Operationalität der abgebildeten Begriffe, dasjenige der Segmentalfacette die Darstellung eines zusätzlichen Horizontes. Unter letzterem ist ein funktioneller, systematischer, topographischer, morphologischer oder ähnlicher Aspekt zu verstehen. Beide Facetten zeigen zusätzlich eine Untergliederung in eine allgemeine und hierarchische Facette (Focus). Unter einer allgemeinen Facette werden diejenigen Aspekte zusammengefaßt, welche die hierarchische Klassifikation erweitern. Die hierarchische Facette stellt im Grunde nichts anderes dar, als die

zweite, differenziertere Ebene der hierarchischen Klassifikation, deren erste Ebene bereits als hierarchische Klassifikation der dimensional Klassen vorgestellt wurde.

Für den thesaurus-internen Informationsfluß ist von Bedeutung, daß diese beiden völlig unabhängigen Gliederungsprinzipien in enger Weise miteinander verschränkt sind. Bei der Indexierung wurde so vorgegangen, daß oberstes Kriterium für die Benutzung der Hierarchie der Klassifikation als auch der Facettengliederung die Frage des Benutzers war, so kann ich *wie* etwas ablegen, *wo* kann ich wie etwas wiederfinden! Selbstverständlich wurde ebenfalls der operationale Aspekt (und damit der intersubjektiv nachweisbare Anteil) berücksichtigt. Anhand des vorgelegten Beispiels (hepatentikuläre Degeneration, Morbus Wilson) wird demonstriert, daß die Suchstrukturen nicht identisch sind mit den Bedeutungsstrukturen eines Begriffes (Abb. 2). In jedem Einzelfall wurde davon ausgegangen, daß gleichzeitig beide Aspekte zu berücksichtigen sind.

Jedem Begriff wurde entsprechend seinem semantischen Bedeutungsinhalt eine Stringkombination zugeordnet. Diese setzt sich aus den bereits angesprochenen mnemotechnischen Kürzeln der hierarchischen Klassifikation als auch den dreistelligen Zahlenkombinationen (Facettenklassifikation) zusammen. Zusammengehörige Stringanteile werden durch einen “.” (Punkt) miteinander verbunden. Die dreistellige Zahlenkombination hinter dem mnemotechnischen Buchstabenkürzel ist demnach als Differenzierung der hierarchischen Klassifikation aufzufassen. Verwechslungsmöglichkeiten sind durch die eindeutige Stringzuordnung ausgeschlossen. Die Anzahl der Teilstrings (Zahlenkombination) ist nicht begrenzt, die

Anzahl der Facettenzuordnungen wie auch die Gesamtzahl der einem Begriffe zugeordneten Strings sind ebenfalls unbegrenzt.

Jeder Begriff erhält zusätzlich eine Identifikationsziffer, es schließt sich der Code des KDS (Klinischer Diagnoseschlüssel) als auch der der ICD/E (International Classification of Diseases/Extended) an. Danach wird der entsprechende Term der SNOP aufgeführt (Abb. 3). In denjenigen Fällen, in welchen eine Repräsentation des entsprechenden Begriffes nicht in dem zusätzlich eingeführten Schlüsselssystem vorgefunden wurde, konnten die Terms (selbstverständlich) nicht aufgeführt werden (insbesondere beim KDS als auch bei der SNOP) haben sich erhebliche Kompatibilitätslücken herausgestellt.

#### Quellen:

- (1) American Medical Association: Standard Nomenclature of Diseases and Operations (SNDO). 5 th Ed. 1961.
- (2) College of American Pathologists. Committee on Nomenclature and Classification of Diseases: Systematized Nomenclature of Pathology (SNOP). Chicago, 1965
- (3) Höpker, W.-W.: Vergleichskriterien unterschiedlicher Klassifikationssysteme. In: Meth. Inf. Med. 11 (1972) p. 144–151.
- (4) Höpker, W.-W.; Grellmann, U.; Kayser, K.; Mehracin, P.; Müller, U.; Ramisch, W.; Rothmund, E.; Seither, G.; Weimer, W.: Das Problem der Diagnose und ihre operationale Darstellung in der Medizin. – Berlin-Heidelberg-New York: Springer-Verlag 1977:
- (5) Immich, H.: Klinischer Diagnoseschlüssel (KDS). Stuttgart: Schattauer-Verlag 1966.
- (6) Kayscr, K., Höpker, W.-W.: Formal Consideration on Text Analysis in Anatomic Pathology. In: Meth. Inf. Med 12 (1973) p. 143–146.

DIMENSION	HIERARCHISCHE KLASSIFIKATION	FACETTENGLIEDERUNG			
		differential		segmental	
		allgemein	hierarchisch	allgemein	hierarchisch
TOPOGRAPHIE	.... Herz Harnblase Uterus .... Magen 88	.... rechts links proximal .... distal 26	Herz .... Spitze Septum .... Klappe 1098	.... Brust Bauch Rücken .... Becken 9	.... ZNS Endokrinum Harnorgane .... Psyche 17
NOSOLOGIE	.... Fremdkörper Mißbildungen äztl. Eingriffe .... Metastasen 29	.... Kachexie Koma Fieber .... Sucht 34	Fremdkörper .... fest flüssig .... gastörmig 375	.... Zyste Divertikel Fistel .... Stenose 60	.... Geburt Vergiftung Senium .... Gutachten 19
ÄTIOLOGIE	.... Bakterien Viren Pharmaka .... physik. Urs. 10		Viren .... Pocken Herpes .... Grippe 262		
MODIFIKATION	.... Grad Größe Diagn.-Sich. .... Zeit 4		Zeit .... frisch älter .... alt 53		Σ 2084

Abb. 1: Hierarchische sowie Facettenklassifikation des Thesaurus der Medizin (TdM). Die hierarchische Klassifikation bezieht sich auf die Dimensionen Topographie, Nosologie, Ätiologie, Modifikation. Die Facetten unterteilen sich in Differential- sowie Segmentalfacetten, wobei jeweils allgemeine und hierarchische Gliederungen angegeben werden! Die Ziffern in jeder Zelle bedeuten den Begriffsumfang, die angegebenen Termini Beispiele aus den entsprechenden Gliederungsbereichen. Anordnung und Umfang der Gliederung entsprechen informations-theoretischen Kriterien.

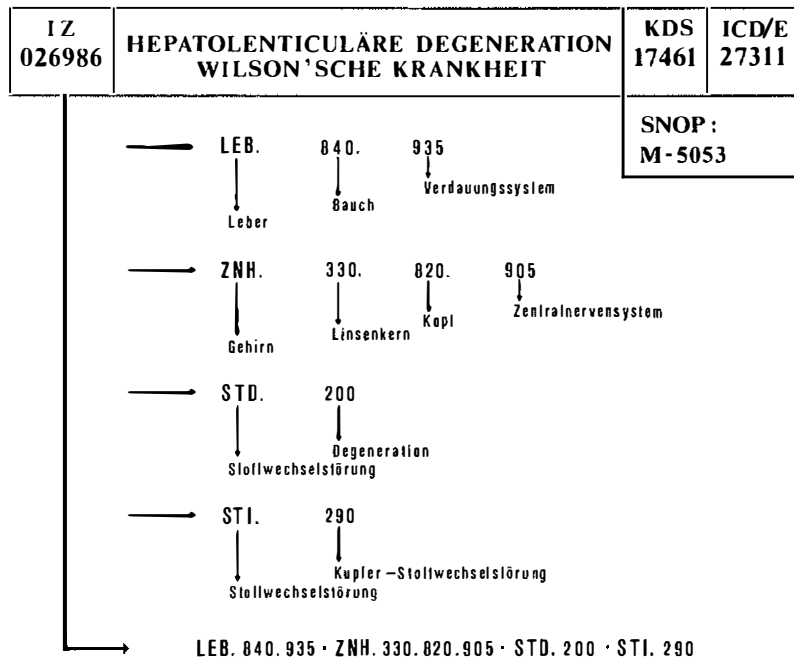


Abb. 2: Beispiel für die Codierung des Begriffes "Hepato-entliculäre Degeneration, WILSON'sche Krankheit". Wie aus dem Beispiel ersichtlich, macht es keine Schwierigkeiten, gleichzeitig verschiedene topographische als auch nosologische Positionen anzugeben. Der Terminus "Degeneration" ist ein historisch gewachsener Begriff, der ausschließlich aus Gründen der Vergrößerung der Benutzerfreundlichkeit nach Möglichkeit – im entsprechenden Sinnzusammenhang – bei sämtlichen innerhalb des TdM enthaltenen Begriffen aufgeführt wird. IZ: Identifikationsziffer; übrige Abkürzungen vgl. Text.

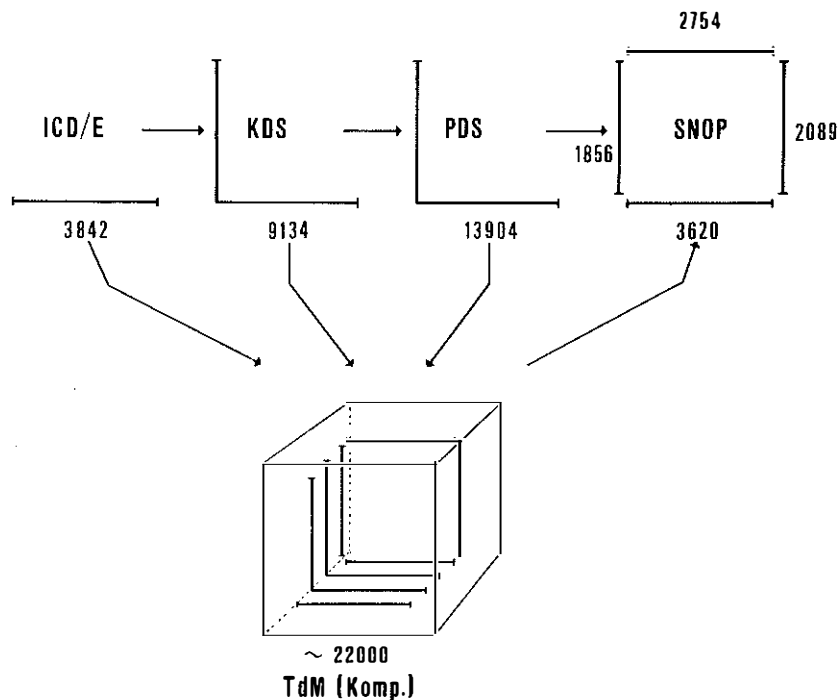


Abb. 3: Kompatibilität des Thesaurus der Medizin (TdM). Innerhalb des KDS ist die ICD/E (IMMICH) enthalten. Beide Schlüssel wurden vollständig in den PDS (Pathoanatomischer Diagnoseschlüssel) aufgenommen. Dieser bildete die Grundlage für den TdM. Für sämtliche in diesem Thesaurus aufgeführten Begriffe wurden die entsprechenden Codebegriffe aus dem SNOP dargestellt. Der Ablauf dieses Vorgehens ist entscheidend für den Benutzer des TdM. Wenn nach einer SNOP-Codierung eines Begriffes gefragt wird, darf nicht der TdM befragt werden. In diesem Falle ist auf die SNOP zurückzugreifen. Andererseits ist es durchaus möglich, Begriffe des KDS aus dem TdM zu entnehmen.