

TELDOK

TELDOK 440 (Telefunken Dokumentationssystem auf der Großrechenanlage TR 440, ist ein allgemeines Informationsbereitstellungssystem (Information Storage and Retrieval System). Mit Hilfe dieses Systems wurde ein wirtschaftliches, sehr leistungsfähiges Informationssystem geschaffen. Die Anwendungsgebiete von TELDOK 440 unterliegen keinerlei Beschränkung, da die Daten, mit denen TELDOK arbeitet – Dokumente oder Auszüge aus Dokumenten – nichts anderes darstellen, als die verbale Beschreibung eines beliebigen Gegenstandes.

Von der **Konzeption** her ist TELDOK 440 vollkommen modular aufgebaut. Es kann daher leicht an die speziellen Bedürfnisse verschiedener Anwender angepaßt werden. Bei Erweiterungen des Systems werden einfach weitere Moduln hinzugefügt.

Durch diesen modularen Aufbau wird der Rechner optimal ausgelastet, da durch Segmentierung der Kernspeicher nur von benötigten Moduln belegt wird und durch andere Programme ausgenutzt werden kann.

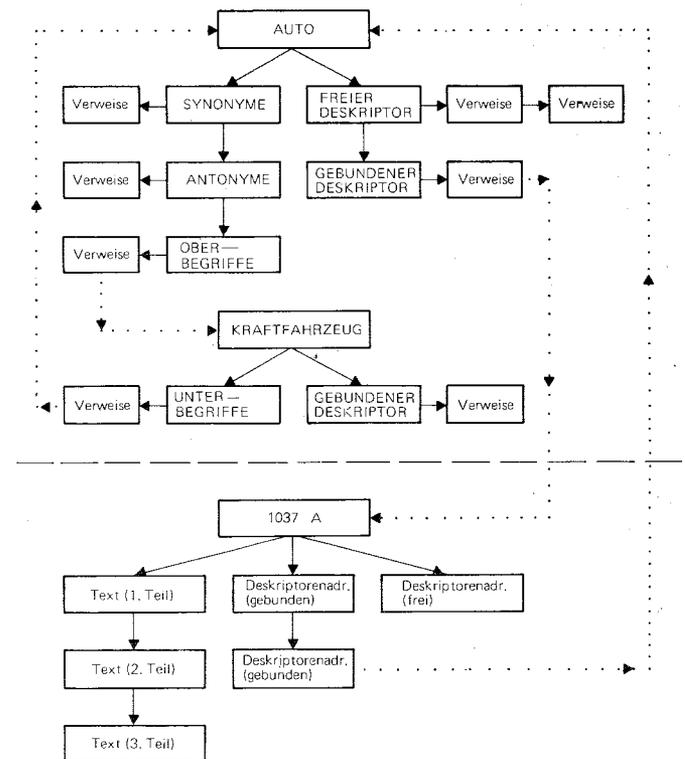
TELDOK 440 ist ein System, das möglichst **rechnerunabhängig** gehalten worden ist. Es ist größtenteils in der maschinenunabhängigen Sprache ALGOL 60 programmiert worden. Die von ALGOL 60 gebotene komfortable Unterprogrammtechnik, Tabellen dynamisch aufzubauen, und die bequem zu formulierende Entscheidungslogik werden vorteilhaft ausgenutzt. Dynamisch aufgebaute Tabellen ermöglichen es, die Größe der Arbeitsspeicher und Bereiche an die jeweils vorhandene Datenmenge anzupassen.

Die **Benutzerfreundlichkeit** von TELDOK 440 ist durch die Verwendung von übersichtlichen Terminals (Sichtgeräte/Fernschreiber) und die Verwendung einer einfachen und leicht merkbaren Kommandosprache stark betont worden, so daß langwierige Schulungen für Benutzer nicht erforderlich sind.

TELDOK 440 gehört zu den **textverarbeitenden** Informationssystemen, die auch nicht formatierte Daten verarbeiten können. Dies ist von besonderer Bedeutung, da jedes zu verarbeitende Dokument einen anderen Aufbau hat.

Bei formatierten Daten haben Felder bestimmte Formate und Bedeutung bezüglich ihres Inhalts. Dies ist bei unformatierten Daten nicht der Fall. Hier werden die Daten nicht mehr einzeln einem bestimmten Feld zugeordnet, sondern erhalten erst dann Sinn und Bedeutung, wenn sie im Zusammenhang (mit ihrem Kontext) betrachtet werden.

Hier werden die Daten nicht mehr einzeln einem bestimmten Feld zugeordnet, sondern erhalten erst dann Sinn und Bedeutung, wenn sie im Zusammenhang (mit ihrem Kontext) betrachtet werden.



Zusammenhang Thesaurusteil-Dokumententeil

So sagt uns z. B. das Wort „Schloß“ für sich allein nicht viel, wenn nicht aus dem Begleittext hervorgeht, ob hier ein „Türschloß“ oder eine „Burg“ gemeint ist.

Das Problem der **Datenerfassung** liegt darin, die Dokumente selbst abzuspeichern und sie weiterhin so zu verarbeiten, daß sie jederzeit mit Hilfe von z. B. Deskriptoren (Schlagwörter und Stichwörter) wieder aufzufinden sind.

Die Datenerfassung bleibt dem Benutzer selbst überlassen, um ein flexibles System für jeden Anwender anbieten zu können. Es wird eine Reihe von generellen Anwendungsprozeduren geboten, die dem Benutzer eine einfache Datenerfassung und einen leichten Datenbankaufbau ermöglichen. Ein Standard-Datenerfassungsprogramm steht zusätzlich zur Verfügung.

Um mit TELDOK 440 arbeiten zu können, werden mit Hilfe von Ausgangsdaten, die auf beliebigen Datenträgern vorliegen können, vier Bereiche erstellt:

- Thesaurusbereich
- Dokumentenbereich
- Invertierter Bereich
- Direkter Bereich

Der Thesaurusbereich (gr. Thesaurus = Schatzkammer, hier Sammlung von Deskriptoren) enthält die Deskriptoren und die semantischen Verknüpfungen zwischen den Deskriptoren.

Deskriptoren können sein: Eigennamen, Städtenamen, Datumsangaben, signifikante Wörter eines Textes oder Zitate.

Der Invertierte Bereich enthält die Verweise der Deskriptoren auf die Dokumente, während der Direkte Bereich die Rückverweise der Dokumente auf die Deskriptoren aufnimmt.

Beispiel (ohne semantische Verknüpfung):

Dokumentenbereich:

Dok. Adr.	Dokumente
1	Tante Berta und Sabine fahren mit dem Auto ins Kino
2	Sabine hat eine Tante, die im Kino arbeitet
3	Tante Berta sieht sich das Auto an

Thesaurusbereich:

Deskr. Adr.	Deskriptoren
1	Auto
2	Berta
3	Kino
4	Sabine
5	Tante

Invertierter Bereich:

Deskr. Adr.	Dok. Adr.
1	1, 3
2	1, 3
3	1, 2
4	1, 2
5	1, 2, 3

Direkter Bereich:

Dok. Adr.	Deskr. Adr.
1	1, 2, 3, 4, 5
2	3, 4, 5
3	1, 2, 5

Dokumente und Deskriptoren können nach dem Willen des Benutzers in beliebiger Form auf beliebigen Datenträgern an TELDOK 440 übergeben werden. Die Daten selbst werden mit Hilfe von DBS 440 (Datenbanksystem 440) abgespeichert und zu einer Datenbank aufgebaut.

Der Thesaurusbereich mit den semantischen Verknüpfungen und der Invertierte Bereich werden zum Thesaurusteil zusammengefaßt, während der Dokumententeil aus dem Dokumentenbereich und dem Direkten Bereich hervorgeht.

Einen Überblick über die wesentlichen Zusammenhänge gibt das Bild auf der Vorderseite.

Mit Hilfe eines **Retrievals** (es wird darunter ein Vorgang verstanden, um gespeicherte Information wieder aufzufinden) kann der Benutzer bei seinen Recherchen wählen, ob es ihm bei einer Suche nach Dokumenten mehr auf Vollständigkeit oder auf Präzision ankommt. Da beides aber nur selten gleichzeitig erreicht werden kann, bietet TELDOK 440 die Möglichkeit, Ergebnisse einzukreisen. Der Benutzer erreicht damit bei hoher Vollständigkeit ein gewisses Maß an Präzision.

Um der „persönlichen Note“ eines Benutzers gerecht zu werden, muß ein geeignetes Retrieval vorausgesetzt werden, verbunden mit einer entsprechenden Strukturierung der Daten. Diese Strukturierung in Verbindung mit dem Aufbau des Thesaurus macht es erst möglich, das Suchen nach bestimmten Dokumenten sinnvoll und optimal zu gestalten.

In TELDOK 440 sind alle allgemein bekannten **Datenstrukturierungen** möglich. Dazu gehören:

- Synonyme (verschiedene Wörter mit gleicher Bedeutung)
- Homonyme (gleiche Wörter mit verschiedener Bedeutung)
- Antonyme (verschiedene Wörter mit gegensätzlicher Bedeutung)
- Semantische Felder (assoziativ zusammengehörige Deskriptoren)
- Polyhierarchische Begriffe (z. B. Baum, Tannenbaum)

Diese Strukturen können in eine Suchanfrage mit einbezogen werden, und es kann damit differenziert angegeben werden, wie weit ein bestimmter Retrieval-Vorgang eingeschaltet werden soll.

Suchanfragen an das System setzen einen Suchvorgang in Bewegung. Sie stellen einen logisch zusammengehörigen Fragenkomplex dar und werden mit Buchstaben (A–Z) gekennzeichnet (numeriert).

Diese „Zeilennummern“ können als Parameter in Kommandos oder als Suchzeilenelemente auftreten. Suchzeilenelemente – Deskriptoren und/oder vorher bereits definierte „Suchzeilennummern“ – können logisch oder gewichtet verknüpft werden.

Man unterscheidet Suchanfragen wie folgt:

- Logische Suchanfragen
- Gewichtete Suchanfragen
- Suchanfragen als natürlicher Satz

Bei einer logischen Suchanfrage werden die Suchzeilenelemente logisch miteinander verknüpft und von links nach rechts abgearbeitet. Bei einer Verknüpfung steht:

- + für logisch 'UND'
- für logisch 'UNDNICHT'
- / für logisch 'ODER'

Eine logische Anfrage wird benutzt, wenn eine möglichst hohe Präzision des Ergebnisses gefordert wird, d. h., daß nur Dokumente gefunden werden, die genau allen Bedingungen der angegebenen Suchzeile genügen (Volltreffer).

Beispiel:

A = AUTO + EWG — ENGLAND

Bei einer gewichteten Suchanfrage werden die Suchzeilenelemente durch Kommata getrennt. Jedes Element darf als „Gewicht“ eine Bewertung erhalten.

```
KDC/SUCH A =H:*GIBTAB(EIN)H.
KDC/SUCH A =H:AUSSCHUSS,5KRAFTFAHRZEUG,INDUSTRIE,3BENZINH.

SUCHZEILE A ENTHAELT FOLGENDE HOMONYME:
0010 AUSSCHUSS
MEINEN SIE :
0011 ABFALL
0012 KOMITEE
ICH MEINE H:0012H.

KLASSIFIKATIONSTABELLE

      ANZAHL      BEWERTUNG      KLASSE
      1          10(VOLLTREFFER)  1
      2          8                2

KDC/SUCH B =H:*GIBDOK(A,1,1)

EIN AUSSCHUSS DER KRAFTFAHRZEUG-INDUSTRIE BERIET AM 20.1.1972
UEBER PROBLEME, ENTSTANDEN DURCH DIE HERABSETZUNG DES BLEIGE-
HALTES IM BENZIN.

KDC/SUCH B =H:*GIBDOK(A,2,ALLE)

DOKUMENT 1:
ES WURDE BESCHLOSSEN, DASS BEI DER ENTWICKLUNG VON KRAFTFAHR-
ZEUG-MOTOREN DIE NEUE BENZINQUALITAET BERUECKSICHTIGT WERDEN
SOLLTE.

DOKUMENT 2:
BENZIN WIRD AUCH NOCH IM NAECHSTEN JAHRZEHT DER VORHERRSCHEN-
DE TREIBSTOFF FUER KRAFTFAHRZEUGE BLEIBEN.

KDC/SUCH B =H:*ENDEH.
```

Diese ist als ganze Zahl b den Elementen voranzustellen, mit $b \neq 0$ und $-127 \leq b \leq +127$. Wird keine Bewertung explizit angegeben, so wird implizit $b = 1$ eingestellt.

Beispiel:

B = PERSON, 10 NARBE, 5 STIRN

Eine Suchanfrage als natürlicher Satz erlaubt dem Benutzer, seine Anfrage in einem freiformulierten deutschen Satz oder auch in Stichworten an TELDOK 440 zu richten.

Sollen Anfragen in natürlicher Sprache an das System gerichtet werden, so ist ihm dies durch ein Kommando mitzuteilen.

Beispiel:

A = WIE KANN MAN DIE VERSCHMUTZUNG DES BODENSEES VERHINDERN

Dabei werden Füllwörter eliminiert und Substantive auf Nominativ Singular gewandelt.

Kommandos bewirken die Ausführung von exakt festgelegten Dienstleistungen und Funktionen. Kommandos können bei Eingabebereitschaft des Systems beliebig eingegeben werden. Es wird von TELDOK 440 nur geprüft, ob die Form syntaktisch richtig und die Ausführung des Kommandos z. Zt. möglich ist.

Kommandos bestehen aus einem mit * beginnenden Namen, gefolgt von in Klammern eingeschlossenen Parametern.

* Name (p₁, p₂, ...)

Im folgenden sind einige Kommandos aufgeführt, ihre Wirkung wird jedoch nicht beschrieben:

* ERWEITERN	* LOEHID
* GIBTAB	* SAMHID
* GIBSEM	* GIBDOK
* GIBHID	* HILFE

Der nebenstehende Protokollauszug zeigt ein Teilgespräch mit Angabe von Kommandos, einer Suchanfrage und Ausgabe von Ergebnissen.

Als **Zusammenfassung** folgen noch einmal die Leistungen von TELDOK 440 in Stichworten:

- Wahlweise Datenerfassung unter Kontrolle von TELDOK-Erfassungsmodulen oder mit einem Standarderfassungsprogramm
- Ausführlicher Dialogverkehr mit weitgehender Unterstützung des Rechners
- Benutzerfreundliche Reaktionen
- Variieren von Retrieval-Techniken
- Steuerung von Präzision (precision) und Vollständigkeit (recall)
- Hervorheben von bedeutungsvollen Deskriptoren in der Suchanfrage
- Formulierungsmöglichkeit von Suchanfragen in natürlicher Sprache
- „Blättern“ im Thesaurusbereich
- Datenstrukturierung: Synonyme, Homonyme, polyhierarchische Beziehungen etc.
- Datenkorrektur (updating)
- Einkreisen von Ergebnissen
- Relevance Feedback

Durch Suchanfragen und Kommandos kann der Benutzer seinen Informationswunsch sowohl im Dialogbetrieb als auch im Stapel- und Stapelfernbetrieb spezifizieren; dabei liefern ihm die oben aufgeführten Retrieval-Methoden umfangreiche Möglichkeiten.