

# Aufbau eines integrierten Dokumentationssystems im EMD : das fachtechnische Konzept (Kurzfassung)

Autor(en): **Willi, J.-J.**

Objektyp: **Article**

Zeitschrift: **Nachrichten / Vereinigung Schweizerischer Bibliothekare,  
Schweizerische Vereinigung für Dokumentation = Nouvelles /  
Association des Bibliothécaires Suisses, Association Suisse de  
Documentation**

Band (Jahr): **48 (1972)**

Heft 2

PDF erstellt am: **21.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-770959>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# AUFBAU EINES INTEGRIERTEN DOKUMENTATIONSSYSTEMS IM EMD: DAS FACHTECHNISCHE KONZEPT (Kurzfassung)

von J.-J. Willi, Sektionschef beim EMD

## 1. Der Auftrag

Die Projektaufgabe bezieht sich auf die «Schaffung eines integrierten Dokumentationssystems, die trotz der dezentralen Beschaffung, der Auswertung und der Lagerung des Dokumentationsgutes durch die Fachdokumentationsdienste, einen integrierten zentralen Such- und Auskunftsdienst zu gewährleisten hat» (Art. 4b der Departementsverfügung vom 11. 2. 70).

## 2. Definitionen

Zum besseren Verständnis der folgenden Ausführungen sind einige Begriffserklärungen notwendig.

*Information* (als Gegenstand) = Wissenswertes, Information entsteht, wenn Wissenswertes zum Ausdruck gebracht wird; (als Tätigkeit) = informieren = weitergeben von Wissenswertem.

*Dokumente* = Materielle Informationsträger aller Art.

*Dokumentation* (als Gegenstand) = Sammlung von Dokumenten; (als Tätigkeit) = dokumentieren = inhaltlicher und formaler Nachweis von Dokumenten.

In diesem Zusammenhang muß auch auf den Unterschied zwischen Informations- und Dokumentationssystem hingewiesen werden:

*Informationssysteme* = Informationen sind vom Benutzer *direkt* verwendbar;

*Dokumentationssysteme* = Es wird auf Informationsquellen hingewiesen, die ausgeschöpft und vom Benutzer interpretiert werden müssen.

Der Betrieb eines Dokumentationssystems, ob manuell oder maschinell, erfordert ein besonderes Ordnungssystem, damit die Informationen gespeichert und nach Bedarf wieder gefunden werden können, denn es geht primär nicht darum, Dokumente als solche zu sammeln, sondern vielmehr die darin enthaltenen *Informationen zu verwerten und greifbar zu machen*.

## 3. Aktionsplan, Problemkreise und Teilprobleme

Die interne Organisation und die Tätigkeit des Fachausschusses für Dokumentationsfragen richten sich nach dem Arbeitsgruppen- und Referenten-

system. Der Aktionsplan weist im wesentlichen die folgenden Arbeitsstufen auf, deren Übergänge sachlich und zeitlich fließend sind:

- Überprüfung des IST-Zustandes
- Bearbeitung des Gesamtkonzepts
- SOLL/IST-Vergleich
- Zusammenfassung der Koordinationsmaßnahmen

sowie die Aufbaubedürfnisse und die Bearbeitung der technischen Grundlagen für den Betrieb durch eine Spezialistengruppe.

Dieses Arbeitsstufen-System wird so gehandhabt, daß aufgrund periodischer Synthesen gewisse Anträge nach Prioritäten und unter Berücksichtigung der Voranschlagstermine formuliert werden.

### 3.1 Organisatorisches

Die Ermittlung der eigentlichen *fachdienstlichen Anforderungen* ist für die Gestaltung der Zusammenarbeit in einem integrierten Dokumentationssystem von grundlegender Bedeutung. Diese Anforderungen ergeben sich weitgehend aus dem *Umfang der Dienstleistungen*, der seinerseits durch Menge und Art der zu berücksichtigenden *Informations- und Dokumentationsbedürfnisse* beeinflusst wird.

Um sich die benötigten Unterlagen für eine *Informationsbedarfsanalyse* (quantitative und qualitative Abgrenzung und Unterteilung des zu bearbeitenden *Dokumentationsspektrums*) und für die Erstellung von sogenannten «*Interessenprofilen*» zu verschaffen, wurde ein sogenannter *Primär-Erhebungsbogen* ausgearbeitet. Die darin enthaltenen *zehn Fragen* wurden allen Dienststellen des EMD bis und mit Sektionsstufe bzw. selbständige Dienste zur Beantwortung vorgelegt. Fünf Fragen bezogen sich je auf die eigentlichen *Informationsbedürfnisse* und auf die *Möglichkeit, bei der dezentralen Erfassung der dokumentationswürdigen Unterlagen mitzuwirken*. Die Ergebnisse dieser Umfrage lassen sich wie folgt zusammenfassen:

- Es wurden *107 Interessenprofile* erstellt; die Anzahl Themenbereiche schwankt zwischen 1 bis 159 pro Dienststelle. Der *EMD-Gesamtdurchschnitt liegt bei 23 Themenbereichen* je Interessenprofil.
- Die Möglichkeit, bei der dezentralen Erfassung der dokumentationswürdigen Unterlagen mitzuwirken, ist bei 63% der beantwortenden Stellen vorhanden.

Das sich aus der Summe aller Themenbereiche ergebende Dokumentationsspektrum zeigt deutlich, daß eine gezielte Informationserschließung und -vermittlung nicht nur im eigentlichen militärischen Bereich, sondern auch im Rahmen der Gesamtverteidigung einem echten Bedürfnis entspricht.

## 3.2 Dokumentationstechnik

Die Schaffung eines integrierten Dokumentationssystems, das trotz dezentralisierter Erschließung des Dokumentationsgutes einen zentralen Such- und Auskunftsdienst zu gewährleisten hat, bedingt in fachtechnischer Hinsicht die Berücksichtigung der folgenden Erfordernisse:

- das Erschließungs- und Suchsystem soll von einzelnen Personen möglichst unabhängig sein;
- die entsprechenden Vorgänge sollen weitgehend unabhängig von der Gedächtniskapazität einzelner Sachbearbeiter erfolgen können;
- die Kompatibilität der Erschließungsmethoden mit andern Systemen inner- und außerhalb des EMD soll gewährleistet sein;
- die Erschließungs- und Wiedergabevorgänge sollen unverschlüsselt erfolgen können;
- mögliche Weiterentwicklungen im Zuge der fortschreitenden Dokumentationstechnik dürfen durch die Wahl ungeeigneter Mittel nicht verbaut werden.

Diese Hauptforderungen verdichten sich schließlich auf eine nähere Prüfung von Dokumentationssystemen, die mit dem Funktionsprinzip von Datenverarbeitungsanlagen vereinbar sind. Im Fall des EMD sind in erster Linie bestehende, mit der in unserem Rechenzentrum heute *vorhandenen* und künftig verfügbaren *Maschinenkonfiguration* kompatible Dokumentationssysteme in Betracht zu ziehen. Auf «Eigenentwicklungen» wird aus naheliegenden Gründen verzichtet.

### 3.2.1 Verfahrenstechnik

Dokumentationsprogramme lassen sich zunächst grob in *Retrieval-Systeme* und *Katalog-Programme* einteilen. Retrieval-Systeme haben die Aufgabe, aus einer großen Anzahl von gespeicherten Informationen die zu besondern Problemen maßgebenden auszusondern. Katalog-Programme werden eingesetzt, um Datenbestände nach bestimmten Kategorien zu ordnen. Mit andern Worten: *Retrieval-Systeme selektieren, Katalog-Programme sortieren*. Diese Unterscheidung ist jedoch nicht zwingend und die Übergänge sind fließend. Für die Lösung der sich aus unserem Auftrag ergebenden Aufgaben kommen Retrieval-Systeme am ehesten in Frage. Es wurden daher verschiedene in der Schweiz und im Ausland in Betrieb stehende Retrieval-Systeme einer näheren Prüfung unterzogen. Auf Einzelheiten dieser Systeme wird im folgenden hingewiesen.



### 3.2.2 Instrumentarium

Das *Zentralproblem aller Dokumentationsverfahren ist das Ordnungssystem*. Hierzu einige grundsätzliche Überlegungen: Die Benutzer von Fachbibliotheken und Dokumentationsstellen der Verwaltung sind Dienststellen und Sachbearbeiter. Sie sind sowohl an Einzel- als auch an Übersichtsinformationen interessiert. Es nützt ihnen wenig, wenn sie auf Anfragen eine große Anzahl von Hinweisen erhalten, die lediglich durch Oberbegriffe charakterisiert sind. Andererseits kann es ihnen durchaus wünschenswert erscheinen, einmal eine größere Literaturzusammenstellung unter einem Oberbegriff zu erhalten. Das *Ordnungssystem* ist als Instrument zum Sichten, Ordnen, Speichern, Suchen und Verfügungstellen von Informationen das *Bindeglied zwischen dem Benutzer und der Dokumentation einerseits und der Dokumentation und dem Speichern andererseits*.

Für den Aufbau dieses «Instruments» ergeben sich einige Notwendigkeiten, die berücksichtigt werden müssen:

- die *Informationsansprüche der Benutzer* (Interessenprofile)
- die *Forderungen des Dokumentalisten als Indexier*
- die *technischen Auflagen* aus der gewählten Speicherform
- die *Postulate der Abfragelogik*.

Die Hauptaufgabe lautet: wie weiß der Abfragende, unter welchen Beschreibungselementen die ihn interessierenden Informationen zu finden sind, d. h. welche Beschreibungselemente welchen Informationen zugeordnet sind. Das Problem ist sprachlicher Natur. Letzten Endes funktioniert eine elektronische Datenverarbeitungsanlage nach eindeutigen, logischen Regeln. Das «Instrument der Umgangssprache» jedoch, das dem Menschen für die Formulierung von Sachverhalten zur Verfügung steht, funktioniert nicht eindeutig, nicht logisch, sondern in redundanter Weise, kaum begrenzt in der Auswahl syntaktischer Formen, oder es funktioniert bestenfalls nach Regeln, die voller Ausnahmen und daher im Erscheinungsbild unlogisch erscheinen (Redundanz = Verlust an Informationswert wegen der Mehrdeutigkeit der Wörter).

Der Computer «denkt» nicht assoziativ, sondern vergleicht nur das Ja/Nein-Prinzip. Deshalb müssen ihm genormte Wörter, d. h. *Deskriptoren* zur Verfügung stehen. Es geht hier ganz einfach um die *Regelung der Benennung von Gedanken*; dieses Problem kann nur außerhalb des Computers gelöst werden.

Es handelt sich nicht um ein allein morphologisches, lexikalisches oder syntaktisches Problem, sondern um Fragen, die alle drei Arten der Sprachbetrachtung umfassen.

Wilhelm von Humboldt hat einmal gesagt, Sprache sei ein «Umschaffen der Welt in das Eigentum des Geistes». Dies geschehe in einer «gedanklichen

*Zwischenwelt*». Diese gedankliche Zwischenwelt muß erfaßt werden. Hierum geht es bei der eindeutigen Darstellung von Sachverhalten.

Programmsysteme können nur zu befriedigenden Ergebnissen führen, wenn der Sprachgebrauch des zu dokumentierenden Gebiets einigermaßen systematisch und einheitlich geregelt ist. Hierzu braucht es einen *Thesaurus*.

Ein Thesaurus im Bereich der Dokumentation ist eine *geordnete Menge von Bezeichnungen*, die ein *offenes System* zur fach- und/oder problemorientierten Indexierung und Ordnung von Begriffen bilden; als Indexierungssystem strebt er die *umkehrbar eindeutige Zuordnung von Bezeichnungen zu Begriffen* an und als Ordnungssystem die *Kennzeichnung von Relationen zwischen den Bezeichnungen*. Er wird in technisch ausgereiften Dokumentationssystemen für die selektive Verbreitung von Informationen an die Benutzer (Interessenprofile) und für den Aufbau der auf präzisen Suchbegriffen gestützten Abfrage-logik benötigt. Die *Erarbeitung eines Thesaurus* mit dem Ziel, das im System zum Speichern und Wiederauffinden von Informationen zu verwendende Wortmaterial in den Begriff zu bekommen, *erfordert unabdingbar, daß bestimmte Entscheidungen getroffen werden*, damit die innere Struktur des Thesaurus einheitlich und logisch ist. Für die Erarbeitung von Thesauri bestehen Regeln und Konventionen, auf deren Einzelheiten hier nicht in allen Punkten eingegangen werden kann. Für die Erarbeitung des EMD-Grundthesaurus haben wir uns auf die Regeln gestützt, die für den Thesaurus des US-Verteidigungsministeriums maßgebend sind. Diese Regeln lassen sich kurz wie folgt zusammenfassen.

In der systeminternen Anwendung versteht man unter Thesauruswort jedes Wort, das in dem Thesaurus aufgenommen wird. Die Thesauruswörter gliedern sich in:

- Deskriptoren = zum Ordnungswort erhobenes Sachwort, mit Angabe von Beziehungen zu anderen Deskriptoren.
- Nicht-Deskriptoren = Thesauruswort, dessen Verwendung als Deskriptor ausgeschlossen ist und von dem auf Deskriptoren verwiesen wird.

Es sei daran erinnert, daß für die Erschließung von Dokumenten mit dem Ziel der Speicherung und Wiederauffindung von Informationen zwei Arten von Angaben verwendet werden:

- die bibliographischen Angaben (formale Angaben: Verfasser, Titel, Quellen usw.) und
- die inhaltlichen Angaben (Beschreibung der Sachverhalte).

Die im Thesaurus aufgenommenen Deskriptoren beziehen sich auf die inhaltliche Beschreibung der Dokumente. Die Thesaurusregeln gliedern sich in drei Arten:

- Regeln für Begriffsbenennungen
- Regeln für Kreuzverweisungen
- Alphanumerierungsregeln.

*Maßgebend für die Wahl der Deskriptoren und ihre Aufnahme in den Thesaurus ist ihre Brauchbarkeit für Zwecke der Verständigung, der Indexierung und der Recherche.* Im allgemeinen kann diese Brauchbarkeit aufgrund folgender Überlegungen beurteilt werden:

- *relative Häufigkeit des Vorkommens* in der Literatur
- *relative Häufigkeit der Verwendung* in bestehenden Dokumentationssystemen
- *Verwandtschaft mit Deskriptoren, die bereits aufgenommen sind,* und
- *Genauigkeit und Annehmbarkeit als wissenschaftlich-technischer Terminus.*

Da diese Kriterien miteinander eng verbunden sind, sollen sie bei der Wahl von Deskriptoren gemeinsam betrachtet werden. Die Frage der Annehmbarkeit einer Benennung als Deskriptor kann anhand von Wortlisten aller Art aus dem betreffenden Sachgebiet und durch *Beurteilung des Fachmanns* entschieden werden.

In Zusammenarbeit mit Fachleuten aus verschiedenen Diensten des Militärdepartements und der Bundesverwaltung und nach den hierfür geltenden Prinzipien und Regeln ist ein Grund-Thesaurus für das EMD mit beinahe 3000 Deskriptoren entstanden. Dieser Grund-Thesaurus wurde auf der Datenverarbeitungsanlage des Rechenzentrums EMD mit dem Programmsystem IRMS (Information Retrieval and Management System) nach verschiedenen Parametern kompiliert. Er kann für die inhaltliche Erfassung von Dokumenten sowohl bei einem manuellen als auch bei dem zur praktischen Erprobung in Aussicht genommenen maschinellen System verwendet werden.

Die Art der Datenerfassung für die Eingabe im künftigen Dokumentationssystem (Lochstreifen oder Magnetband) wird gegenwärtig mit dem Rechenzentrum geprüft.

#### 4. Aufbewahrung der Originaldokumente

Es ist vorgesehen, die zu den gespeicherten inhaltlichen und bibliographischen Angaben gehörenden Volltexte, je nach Umfang entweder im Original (Bibliothek) oder als *Mikrofilm* aufzubewahren, mit der Möglichkeit, kurzfristig Rückvergrößerungskopien herzustellen.

Die sich beim Aufbau eines integrierten Dokumentationssystems *stellenden Probleme* sind mannigfaltig. Sie können vom einzelnen Sachbearbeiter



allein nicht mehr gelöst werden. Arbeitsteilung und Gruppenarbeit sind unabdingbare Voraussetzungen für das Gelingen derartiger Vorhaben. Die Dokumentalisten können nicht einfach Forderungen aufstellen und die Lösungen den EDV- bzw. Mikro-Film-Fachleuten überlassen. Man muß von Anfang an zusammenarbeiten und fortwährend miteinander sprechen. Eine Ausbildung der Dokumentalisten in Richtung EDV und vermehrtes Verständnis der Informatiker für die Bedürfnisse der Dokumentation werden unumgänglich. Ein Erfahrungsaustausch im Sinn des heutigen Seminars liefert daher eine willkommene Gelegenheit, dieses gegenseitige Verständnis zu fördern.

## PROBLEMES DE L'ANALYSE AUTOMATIQUE DE TEXTES

par *J.-M. Margot*, IBM, Berne

### *Introduction*

Dans un système documentaire, manuel ou mécanisé, un document (livre, article de périodique, brevet, carte ou plan, rapport, etc.) est toujours représenté par deux éléments. La description physique du document correspond à la notice catalographique où figurent les coordonnées (titre, auteur, éditeur, date, nombre de pages, etc.) permettant de retrouver de manière univoque le document ainsi signalé. La deuxième partie sert à permettre la recherche documentaire. Il s'agit de l'image du contenu du document, un ou plusieurs codes de classification (Classification Décimale Universelle par exemple), des mots-clés, un résumé ou même un texte complet du document si celui-ci n'est pas trop long. Les codes de classification et les mots-clés ont fait l'objet des deux précédentes conférences et mon propos est de présenter quelques problèmes à résoudre si l'on désire utiliser un ordinateur pour analyser et traiter des résumés ou le texte complet d'un document. L'analyse automatique de texte a pour but de:

- sélectionner certains éléments d'un texte sans les changer,
- transformer la forme ou le contenu du texte sans en modifier le sens.

Des exemples vont mettre en évidence quelques-uns des problèmes susceptibles de surgir lors de l'analyse automatique de textes.

Les problèmes présentés peuvent être résolus et l'ont été, mais nous n'avons pas l'intention, dans le cadre de ce modeste exposé, de montrer les solutions existantes, car cela exigerait plus de temps et ferait appel à des notions d'informatique spécialisée.