

E. Paul

Gesellschaft für Kernforschung Karlsruhe

Überlegungen zu Art und Umfang eines Informationssystems
auf dem Gebiet Prozeßlenkung mit DV-Anlagen

Im Folgenden sind einige Überlegungen zusammengefaßt, die in der Zentralstelle für Atomkernenergie-Dokumentation (ZAED) im Hinblick auf ein Informationssystem Prozeßlenkung mit Datenverarbeitungs-Anlagen (PDV) angestellt wurden. Der Anstoß für diese Überlegungen kam von der PDV Projektleitung im Kernforschungszentrum Karlsruhe.

Zunächst seien einige Begriffsdefinitionen erlaubt:

"Ein Informationssystem ist die Summe aller Vorkehrungen zur Erfassung, Erschließung, Anzeige und Verfügbarmachung von Dokumenten, die es gestatten, einen Informationssuchenden innerhalb eines abgegrenzten Fachbereiches schnell an die gewünschte Information heranzuführen". Da wir uns hier zunächst nur mit der Literaturdokumentation befassen wollen, ist die Datendokumentation in dieser Definition nicht enthalten.

Ausgehend von dieser Definition ergeben sich vier Punkte, auf die ich eingehen möchte, nämlich:

1. die Festlegung des Fachbereiches;
2. die Dokumentation, also die Erfassung und Erschließung der Dokumente;
3. die Information, also der Hinweis auf die Dokumente bzw. ihre Anzeige
und
4. die Literaturversorgung, also die Verfügbarmachung der angezeigten Dokumente.

Zunächst zur Abgrenzung des Fachbereiches. Die Konzeption des Systems hängt sehr wesentlich von der zu erwartenden Menge des Materials ab, und die Menge ist ihrerseits eine Funktion der Abgrenzung des Bereichs.

Es wird daher die Menge des Materials nicht unwesentlich beeinflussen, wenn beispielsweise der Einsatz der Datenverarbeitung im medizinischen Bereich oder bei der Bibliotheksautomatisierung oder generell in Wissenschaft und Verwaltung hinzugerechnet wird oder nicht; ebenso die Entscheidung, ob solche Arbeiten, die auf den jeweiligen technischen Prozeß bezogen sind und in denen die Datenverarbeitung nur eine Nebenrolle spielt, in das System einbezogen werden sollen oder nicht.

Eine möglichst eindeutige Eingrenzung des Bereichs und damit die Abgrenzung gegen Nachbarbereiche ist nicht nur im Hinblick auf die Planung des Informationssystems wichtig, weil sie die erforderliche maschinelle Ausrüstung und den Personalbedarf beeinflußt, sondern der Benutzer des Systems muß wissen, über welchen Sachbereich genau er Information erwarten kann, wenn er das System in Anspruch nimmt, und welche Fragen auf Randgebieten nicht durch das System beantwortet werden können.

Der zweite Punkt ist die Erfassung und Erschließung der Dokumente, die wir unter dem Begriff Dokumentation zusammenfassen. Der erste Schritt ist hierbei die Überwachung des Literaturangebotes und die Auswahl der zu erfassenden Dokumente nach inhaltlichen und formalen Kriterien. Entsprechend der Abgrenzung des Fachbereiches muß festgelegt werden, welche Zeitschriften und welche Berichte-Reihen ausgewertet und ob Patentliteratur und Hochschulschriften erfaßt werden sollen.

Ferner müssen nationale und internationale Konferenzen sowie die Buchproduktion deutscher und ausländischer Verleger überwacht werden. Weiter ist zu klären, ob und auf welche Weise interne Berichte und andere Veröffentlichungen von Firmen und Verbänden beschafft und in das System aufgenommen werden können. Möglicherweise sind hier besondere Sperrklauseln für die Anzeige und Weitergabe als Auflage der Herausgeber solcher Schriften zu beachten.

Alle diese Dokumente, die entweder durch Kauf, Tausch oder kostenlose Überlassung eingegangen sind, werden nun im zweiten Arbeitsschritt bibliographisch und sachlich erschlossen, d.h. es werden die Kriterien festgehalten, die zur Wiederauffindung der Dokumente nach sachlichen und formalen Gesichtspunkten erforderlich sind.

Die bibliographischen Daten, wie Autor, Titel der Arbeit, Korporation des Autors, Quellenangabe und andere formale Kennzeichen des Dokuments werden in maschinenlesbarer Form erfaßt. Als Datenerfassungsgeräte kommen Lochstreifen-Schreibmaschinen oder solche Geräte in Frage, die während des manuellen Schreibvorganges direkt ein DV-kompatibles Magnetband erzeugen. Die letztere Geräteart wird bei uns in der ZAED bei der Datenerfassung für unsere Systeme "Informationen zur Kernforschung und Kerntechnik (IKK)" und "Internationales Nukleares Informationssystem (INIS)" eingesetzt. Die Erfassung der bibliographischen Daten geschieht nach detaillierten Arbeitsvorschriften, durch welche die Form der Autorenangabe, die Art der Aufnahme der Korporationen, die Abkürzung der Zeitschriftentitel und anderes geregelt ist.

So ein Regelwerk für die Datenaufnahme ist vom jeweiligen Fachgebiet unabhängig, und wir würden für das PDV Informationssystem das INIS-Regelwerk anwenden, welches der Datenerfassung für die Informationssysteme in der ZAED bereits zugrunde liegt.

Die sachliche Erschließung der Dokumente geschieht in der Regel in drei Ebenen, un zwar

1. es wird eine Klassifikation der Dokumente nach Sachgruppen vorgenommen
2. dem Dokument werden Schlagworte entsprechend seinen Sachverhaltsschwerpunkten zugeordnet und
3. es wird eine Zusammenfassung des Inhalts angefertigt.

Die Klassifikation besteht in der Zuweisung des Dokuments zu einer oder zu mehreren Sachgruppen, wobei das Dokument unter der ersten Sachgruppe sozusagen "abgestellt" ist während von den weiteren Gruppen auf das Dokument verwiesen wird. Die Sachgruppenliste, nach der auch das später noch zu erwähnende Informationsheft gegliedert sein wird, muß nach erfolgter Abgrenzung des Bereichs aufgestellt werden. Die Sachgruppenliste für den Bereich der Kernforschung und Kerntechnik nach INIS umfaßt beispielsweise 95 Gruppen, die in 6 Hauptgruppen zusammengefaßt sind.

Während die Klassifikation bereits ein Wiederauffinden der Dokumente nach mehr oder weniger feinen Teilbereichen ermöglicht, ist eine Suche nach Dokumenten zu einer spezifischen und engen Fragestellung nur durch Abfrage nach Schlagworten bzw. nach Schlagwortkombinationen durchführbar. Die Zuteilung solcher Schlagworte ist die zweite und wichtigste Ebene der Sacherschließung. Sie wird vorgenommen auf der Grundlage eines Thesaurus, also einer Liste von Sachbegriffen, deren Inhalt und deren Form festgelegt ist. Ein solcher Thesaurus muß für den Bereich PDV noch entwickelt werden.

Wenn man davon ausgeht, daß das PDV-Informationssystem als maschinelles System konzipiert werden soll, muß diese Begriffsliste maschinengerecht sein. Maschinengerecht bedeutet hier, daß die Prinzipien des "coordinate indexing" angewendet werden, wobei mehrere gleichberechtigte Elemente der Begriffsliste, - die in ihrem Begriffsinhalt in der Regel breiter sind als etwa die Schlagworte eines manuellen Registers-, durch ihre Kombination einen engen Sachverhaltsschwerpunkt wiederzugeben imstande sind.

Grob gesprochen kämen in einem Thesaurus für reines "coordinate indexing" keine zusammengesetzten Begriffe vor, da die einzelnen Teile untereinander zu weiteren Begriffen kombiniert werden können. Dadurch kann die Gesamtzahl der Begriffe des Thesaurus vergleichsweise niedriger sein und die Ökonomie der maschinellen Suche wird günstig beeinflusst.

Reine "coordinate-indexing"-Verfahren sind in der Praxis der Informationssysteme jedoch nicht angewendet worden, hauptsächlich wohl deswegen, um stehende, zusammengesetzte Begriffe der jeweiligen Fachsprache nicht bis zur Unkenntlichkeit verstümmeln zu müssen und dadurch auf der Eingabeseite eine künstliche Fachsprache zu schaffen. Die günstigere Maschinenökonomie scheint kein ausreichendes Äquivalent zu sein. Ein gangbarer Weg zu einem Thesaurus für das PDV-Informationssystem wäre folgender: man betrachtet einige Referateorgane, die den Bereich PDV - wenn auch nur teilweise - miterfassen und extrahiert aus den jeweiligen Registern die gebietsrelevanten Begriffe. Hier bieten sich z.B. an der "Engineering Index" und "INSPEC" (Information System in Physics, Electrotechnology and Control). Die Begriffe müßten in einer Liste unter Ausmerzung von Synonymen zusammengefaßt und in sinnvoller Weise nach den Prinzipien des "coordinate indexing" aufgespalten werden.

Die Anwendung einer solchen Liste in der Praxis würde zu notwendigen Ergänzungen und Modifikationen führen, so daß nach einer gewissen Zeit ein brauchbarer Thesaurus vorläge.

Eine andere Möglichkeit zur Gewinnung eines Thesaurus ist die Häufigkeitsanalyse von Worten aus den Titeln - oder noch besser aus den Zusammenfassungen - einiger Tausend repräsentativer Arbeiten des Gebiets. Die Begriffe mit einer Häufigkeit über einem zu definierenden Mindestwert müßten wieder der Synonym-Kontrolle und einer zweckmäßigen Reduzierung durch Aufspalten unterzogen werden und könnten in alphabetisierter Form den Thesaurus darstellen.

Ein guter Thesaurus beeinflusst ganz wesentlich die Güte des Systems; seine Erstellung ist recht arbeitsintensiv und eigentlich der wichtigste Punkt in der Systemkonzeption.

Kommen wir zum Punkt 2, zur Dokumentation, zurück und darin zur dritten Ebene der Sacherschließung. Diese dritte Ebene ist das Erstellen einer Zusammenfassung, falls eine solche nicht bereits Bestandteil der Dokumentes ist, aber auch bei Vorhandensein eines Referates ist es oft notwendig, dieses zu modifizieren oder zu ergänzen, da Autorenreferate nicht notwendigerweise eine objektive Inhaltsangabe in Bezug auf das Informationssystem darstellen, da der Verfasser gelegentlich "Lieblings-Aspekte" seiner Arbeit besonders hervorhebt und andere Sachverhalte nur kurz streift.

Auch die Ergebnisse der Sacherschließung werden mit Hilfe der vorher bereits genannten Datenerfassungsgerät in maschinenlesbare Form gebracht. Die bibliographischen Daten und die Ergebnisse der Sacherschließung werden auf Magnetband gelesen, welches in geeignet strukturierter Form die Ausgangsbasis für den anfangs genannten dritten Punkt, nämlich die Information, bildet.

Als Informationsangebot des Systems sollten zunächst drei Arten von Ausgabeprodukten vorgesehen werden:

1. ein laufender Profildienst (selective dissemination of information) (SDI) über das jeweils neue Material
2. retrospektive Recherchen über das jeweilige Gesamtmaterial
und
3. Anzeige des neuen Materials in Schnellinformationsheften mit Referaten und Registern nach Art eines Referantenblattes.

Der Profildienst besteht in regelmäßigen Listen, die dem Benutzer aufgrund seiner speziellen Fachinteressen von der Datenverarbeitungsanlage ausgedruckt werden.

Dabei sucht die Maschine nach ihr angegebenen Schlagworten bzw. Schlagwortkombinationen. Bis der Profildienst ein befriedigendes Ergebnis bringt, wird es notwendig sein, den Benutzer zu den ersten paar Lieferungen sehr genau zu befragen und die der Maschine angegebenen Schlagworte zu ergänzen oder zu reduzieren. Dann, sobald ein Profil mit gutem Ergebnis läuft, sollen die Listen - abgesehen von der sogenannten "Profilpflege", die aufgrund der eingehenden Antwort-Formblätter durchgeführt wird - vom Schnelldrucker ohne ständige inhaltliche Überprüfung durch den Fachreferenten direkt zum Versand kommen.

Es ist ein Merkmal des maschinellen Profildienstes, daß die Listen auch einen gewissen Prozentsatz nichtrelevanter Titel enthalten können; bei der geringen Gesamtmenge der Titel pro Liste ist jedoch der schnelle Überblick über die zutreffenden Titel gewährleistet.

Die Frage, ob eine angezeigte Arbeit interessant genug ist, um sich das Original zu beschaffen, wird allein vom Titel her nicht in allen Fällen eindeutig zu beantworten sein. Deshalb sollte vorgesehen werden, in den Profillisten die Zusammenfassungen der Arbeiten mit auszudrucken und die Listen dadurch informationsreicher zu gestalten. Manche Informationssysteme lehnen dies aus Kostengründen ab und unterhalten einen Mikrofische-Dienst für die Zusammenfassungen (z.B. INIS).

Eine vereinfachte Variante des Profildienstes ist der laufende Ausdruck von Listen nach Sachgruppen. Solche Listen sind eventuell sinnvoll als Abrundung eines echten Profils nach Schlagworten oder in solchen Fällen, wo sich das Interessengebiet eines Benutzers voll mit einer oder mit mehreren Sachgruppen deckt.

Bei den retrospektiven Recherchen liegen die Verhältnisse im Bezug auf die Relevanzquote anders.

Eine solche Recherche wird auf Einzelanfrage nach Schlagworten über das gesamte eingespeicherte Material oder über das Material eines vom Benutzer angegebenen Zeitraumes durchgeführt. Die von der Maschine ausgedruckten Titel werden vom zuständigen Fachreferenten sorgfältig auf Relevanz geprüft und nur die zutreffenden Hinweise werden, mit einem Kommentar versehen, dem Benutzer zugestellt. Die Zusammenfassungen sollten, wenn nicht als Computerausdruck, so doch als Rückvergrößerung oder als Kopie aus dem Informationsheft mitgeliefert werden.

Eine retrospektive Recherche wird in der Regel angefordert zur Unterrichtung auf einem neuen Gebiet oder zu Beginn einer laufenden Information nach Interessenprofilen. Es ist selbstverständlich, daß retrospektive maschinelle Recherchen in der Anfangsphase des Systems nicht sehr sinnvoll sind, da die Gesamtmenge der gespeicherten Titel noch klein ist.

Nachdem das dokumentationsseitig erstellte Magnetband mit dem neuen Material in regelmäßigen Abständen zur Verfügung steht, ist die Herausgabe eines Referateblattes mit Registern ohne großen Mehraufwand möglich. Die Titel werden entsprechend den zuvor erwähnten Sachgruppen gegliedert; die Register werden maschinell erstellt.

Im Zusammenhang mit der Nutzung des eigen-erstellten Magnetbandes für die Informationsdienste sei darauf hingewiesen, daß im Sinne eines effektiven und wirtschaftlichen Informationssystems die Frage sehr genau zu prüfen ist, ob bereits vorhandene Fremd-Magnetbänder verwandter Fachbereiche nicht auszugsweise für das eigene System nutzbar gemacht werden können. Sind relevante Teilbereiche maschineller Fremdsystem vorhanden, so könnte dadurch gegebenenfalls ein gewisser Anteil der eigenen Dokumentationsarbeit eingespart werden. Das Problem der Erfassung und Einspeicherung dieses Teilmaterials ist dann zurückgeführt auf eine maschinelle Transformation der Fremddaten in die Struktur des eigenen Bandes und auf die Erstellung einer Konkordanz der Schlagwortsysteme.

Lassen Sie mich nun noch auf den letzten Punkt, die Literaturversorgung eingehen. Wir sind der Meinung, daß die Literaturversorgung eine Fortsetzung der Informationsarbeit ist in dem Sinne, daß die Verfügbarkeit bzw. das schnelle Verfügbarmachen der in den Informationsdiensten angezeigten Literatur für den Benutzer den Wert eines Informationssystems wesentlich mitbestimmt und dessen Bestandteil sein sollte.

Von der Informationsstelle her gesehen gliedert sich das Problem in zwei Teile: nämlich in den Bereich der nicht-konventionellen Literatur und den Bereich der konventionellen Literatur. Nichtkonventionelle Literatur ist in der Regel schwer beschaffbar, denn Forschungsberichte, Konferenzvorträge, Dissertationen, Patente und Firmenschriften sind nicht im Buchhandel erhältlich. Es sollte eine vollständige Sammlung dieser Dokumente angestrebt werden, um bei Leihfragen umgehend ausleihen zu können.

Unter etwas anderem Gesichtspunkt ist der Bereich der konventionellen Literatur, also der Bücher und Zeitschriften, zu sehen. Zeitschriftenartikel werden den größten Anteil an der eingespeicherten Literaturmenge haben. Im Hinblick auf die breite fachliche Streuung des Gebietes PDV, das neben der gesamten Technik auch die Chemie, die Physik und eventuell die Medizin, das Bibliothekswesen, sowie gewisse Verwaltungsbereiche berühren wird, müßte man in der Zentrale des Systems eine große Zahl von Zeitschriften abonnieren, um sie für Ausleihfragen bereitzuhalten. Dies ist bei einem angemessenen finanziellen Aufwand nur für eine beschränkte Zahl von wichtigen, d.h. überwiegend bereichsbezogenen Zeitschriften möglich.

Für den vermutlich sehr großen Rest sollte ein Standortnachweis in Form von Zentralkatalogen aufgebaut werden, die die Informationsstelle in die Lage versetzen, Leihfragen, die aus dem Bestand heraus nicht zu befriedigen sind, gezielt an einen benutzernahen Standort weiterzuleiten und damit eine schnelle Ausleihe zu gewährleisten.

Wie ich zwischendurch bereits andeutete, haben die hier vorgetragenen Überlegungen zum großen Teil bereits einen realen Hintergrund in Form des "Internationalen Nuklearen Informationssystems (INIS)" und der "Information zur Kernforschung und Kerntechnik (IKK)", die bei uns in der ZAED schon als Informationssysteme in Betrieb sind.