

BEARBEITUNG, VERWALTUNG UND VERSAND VON DOKUMENTEN

H. P. Godbersen; G. Köhler; Ch. Lewin; J. Seib; G. Zschoche

Nixdorf NME, Berliner Straße 66, D1000 Berlin 27

erschienen in: GI-Jahrestagung, Informatik Fachberichte Bd. 126, 1986 p. 118-133

Zusammenfassung

Die Ziele eines integrierten Dokument-Ansatzes werden diskutiert. Eine im Rahmen eines ESPRIT Projekts entwickelte Prototyp-Implementierung wird vorgestellt, die sowohl die Bearbeitung und Verwaltung als auch den Versand von Büro- Dokumenten unterstützt.

1 Einleitung

Das Gebiet der computergestützten Dokumentbearbeitung ist zu einem Feld ständiger Forschung und Entwicklung geworden. Dabei hat der Begriff *Dokument* als auch die Funktionalität, die der Begriff *Dokumentbearbeitung* beinhaltet, in der Vergangenheit eine erhebliche Erweiterung erfahren. Während sich bislang die computergestützte Dokumentbearbeitung auf "reinen" Text beschränkte, schließt der moderne Dokumentbegriff zusätzlich die Integration von graphischen Informationen, Bildern (Faximile) und digitalisierter Sprache mit ein.

Gleichzeitig erweitert sich auch das Verständnis von den zu dem Bereich der Dokumentbearbeitung gehörenden Dienstleistungen. Ein wesentliches Merkmal ist die zukünftige Einbettung der Dokumentbearbeitung in ein umfassendes Bürosystem. Diese Integration zeichnet sich durch die Berücksichtigung aller Arten von Aktivitäten der Dokumenthandhabung aus. Neben den bislang im Mittelpunkt stehenden Aufgaben der lokalen *Dokumentbearbeitung* (Erfassen, Ändern, Formatieren, Darstellen, Drucken usw.) treten heute auch die Aufgaben der *Dokumentverwaltung* und vor allem des *Dokumentaustausches* stärker in den Vordergrund. Dazu gehören die Funktionalitäten zum Einordnen und Wiederfinden, sowie die des Verteilens von Dokumenten (Adressieren, Zustellen, Identifizieren, Überwachen und Abfragen von Zustellungen usw.).

Die Entwicklung von Dokumentbearbeitungssystemen (und deren Teilkomponenten), die in der Lage sind, die angesprochenen Funktionalitäten zu realisieren, setzt einen einheitlichen Aufbau von Dokumenten voraus. Um die angesprochene Verteilfunktionalität auch zwischen heterogenen Systemen unterschiedlicher Hersteller zu gewährleisten, ist die Einhaltung internationaler Konventionen unbedingt erforderlich. Dementsprechend beschäftigen sich internationale Normungsgremien, wie ISO, CCITT und ECMA seit einiger Zeit mit der Erstellung geeigneter Normen.

Allen diesen Normierungsaktivitäten ist die Bemühung gemeinsam, die den zu bearbeitenden Dokumenten innewohnende Struktur stärker als bisher zu nutzen und so eine strukturorientierte Sichtweise der Dokumentbearbeitung einzuführen (s. /1/). Richtungsweisend sind dabei der Standard ECMA 101 /2/, der hierfür den Begriff der Dokumentarchitektur (*Office Document Architecture*) geprägt hat und der ISO Standard SGML *Standard Generalized Markup Language* /3/. In engem Zusammenhang damit ist die T.73 Empfehlung der CCITT zu sehen /4/, die sich vor allem mit der Problematik des Austausches von Dokumenten befaßt.

Bei Nixdorf wird derzeit ein Prototyp für die Bearbeitung, Verwaltung und den Versand von Dokumenten im Rahmen eines von ESPRIT geförderten Projekts **An Integrated Network Architecture für Office Communications** entwickelt (s. /5/). Darin werden die bereits angesprochenen Anforderungen berücksichtigt, um Erfahrungen mit der Integration unterschiedlicher Informationsarten zu sammeln. Weitere Ziele sind die Weiterverwendung der in Unternehmen bereits eingeführten SubEditoren und die ökonomische, redundanzfreie Speicherung und Übertragung der Dokumente.

Abbildung 1 zeigt die beabsichtigte Integration unterschiedlicher Informationsarten in einem Dokument. Kapitel zwei beschreibt die gewählte Systemstruktur des Prototyps. In Kapitel drei wird die Dokumentbearbeitung erläutert. Die folgenden Abschnitte behandeln schließlich die Themen Verwaltung und Versand.

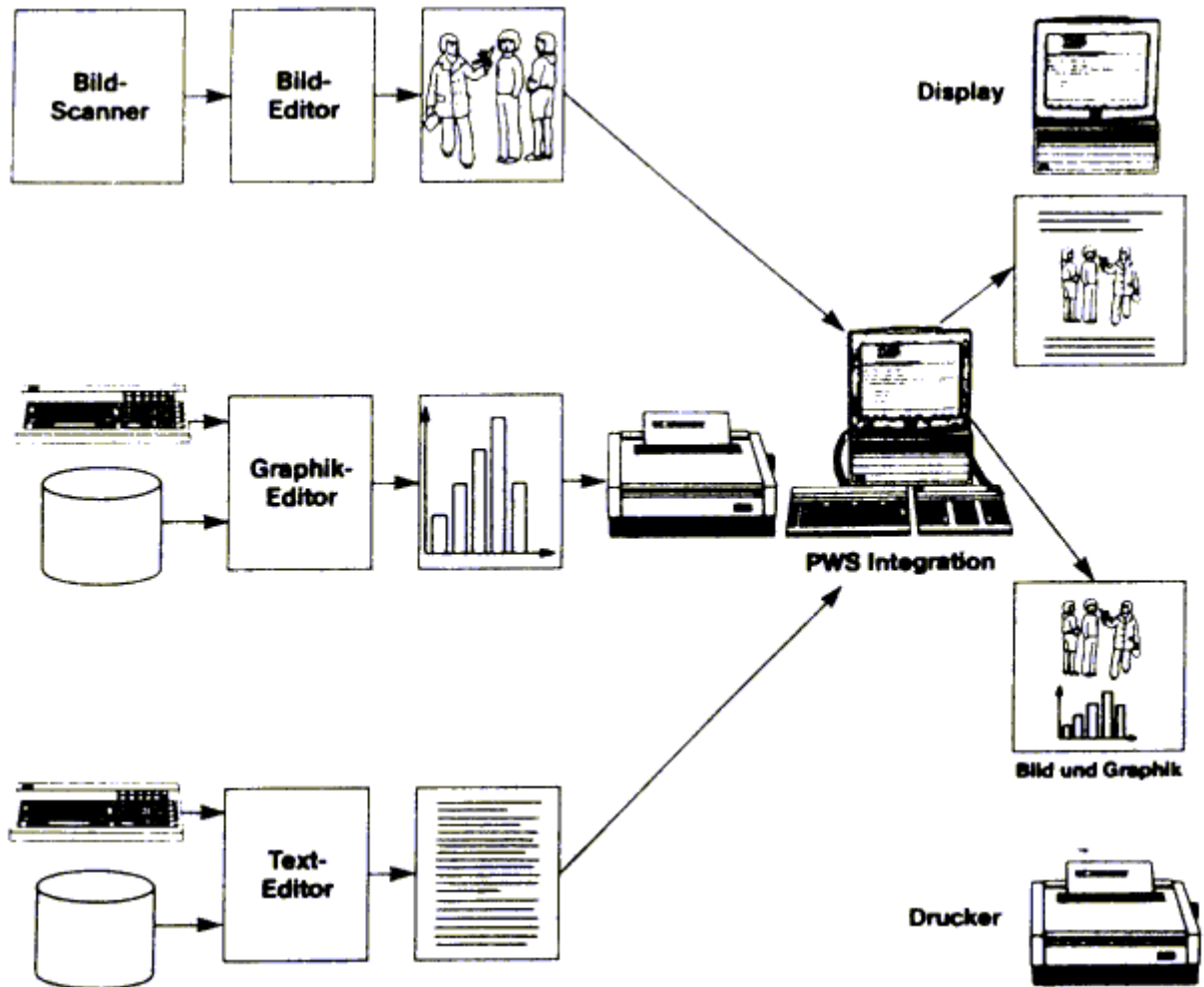


Abb. 1: Integration von Text, Bild und Graphik

2 Die Systemarchitektur

Die Konzepte werden im Rahmen einer Client/Server Architektur entwickelt. Mehrere *TARGON Server* und *Professional Workstations (PWS)* sind über ein Kommunikationsnetz miteinander gekoppelt. Die Benutzer können in der Regel frei zwischen den PWS wählen. Diese logische Strukturierung kann auf unterschiedliche Weise in physikalische Geräte abgebildet werden, insbesondere lassen sich unterschiedliche Server Funktionalitäten in einem Gerät zusammenfassen. Abbildung 2 soll die Komponenten in der Workstation grob skizzieren.

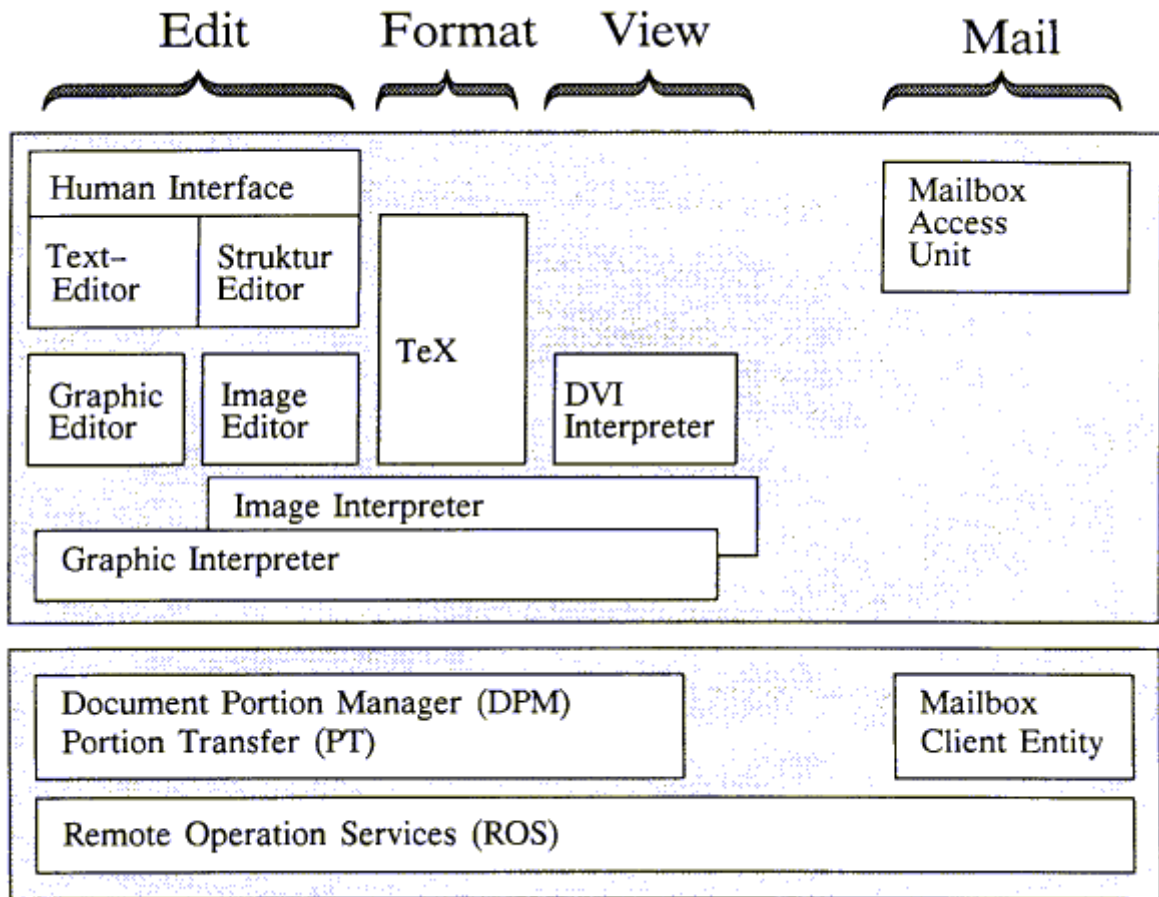


Abb. 2: Komponenten in der PWS

In der PWS-Software sind unterhalb der einheitlichen Benutzeroberfläche die Applikationen Bearbeitung (Edit, Format, View) und Versand (Mail) angesiedelt. Diese benötigen neben allgemeinen Verwaltungshilfen (Ablage/Archiv) eine spezielle Verwaltungskomponente, deren Aufgabe darin besteht, die Portionen eines Dokuments (wie Text, Image, Graphik) zusammenzuhalten. Die Arbeitsteilung zwischen Applikationen und 'Portionsverwaltung' besteht darin, daß die Applikationen im wesentlichen innerhalb der Portionen arbeiten, die Verwaltung aber die Portionen als kleinste Einheiten bereitstellt und das gesamte Dokument mit verwaltet. Zur Abwicklung der Kommunikationsaufgaben mit den Dokument-Servern steht der Dienst Portion Transfer zur Verfügung.

Die Versand-Applikation auf der PWS (Mailbox Access Unit) kommuniziert mit allem Mail-Server mittels eines Mailbox Access Protokolls. Der Mail-Server versendet die Post, benutzt dazu Directory-Dienste und enthält die Postfächer der Benutzer. Die Server enthalten neben den Transfer- und Remote-Operation-Moduln (s. /6/) zusätzliche Verwaltungskomponenten, die die Zugriffs- und Besitzrechte auf labiles bzw. Dokument-Portionen regeln..

3 Die Dokumentbearbeitung

3.1 Der Texteditor

Zur Aufgabe des Texteditors gehört es, dem Benutzer ein Werkzeug zur Verfügung zu stellen mit dessen Hilfe er den textuellen Inhalt eines Dokuments bearbeiten kann. Bearbeiten heißt der Benutzer muß Text eingeben und modifizieren (löschen, kopieren, verschieben, ändern können. Neben dieser Basisfunktionalität benötigt der Benutzer zusätzliche Hilfsmittel, die ihn seine Arbeit erleichtern (Travelling-Funktionen usw.). Moderne Editoren zeichnen sich dadurch aus, daß sie darüber hinaus Funktionen zur Textgestaltung bereitstellen. Nach der Mächtigkeit dieser Funktionalität (Hervorheben von Text, automatischer Zeilenumbruch mit Silbentrennung und Randausgleich usw.) unter Berücksichtigung der Prämisse *What You See, Is What You Get* werden heute Texteditoren klassifiziert und beurteilt.

3.2 Der Struktureditor

Struktureditoren sind bislang kaum in Erscheinung getreten. Die grundsätzliche Aufgabe eines Struktureditors ist, den Benutzer bei der Erstellung, Bearbeitung und Gestaltung strukturierter Dokumente zu unterstützen. Fehlt dieses Hilfsmittel (dies ist die heutige Situation vieler Texteditoren), so ist es die Aufgabe des Benutzers, vorhandene, u.U. streng gefaßte, Strukturierungs- und Gestaltungsrichtlinien für Dokumente (z.B. DIN-Norm für Handelsbriefe) "intellektuell" einzuhalten. Geradezu notwendig wird ein Hilfsmittel dann, wenn ein Dokument von einer anderen Person (an einem anderen Ort) weiterbearbeitet werden soll, ohne daß die oben erwähnten Richtlinien verletzt werden.

Für die Strukturierung von Dokumenten sind zwei Gesichtspunkte von Bedeutung: die logische Sichtweise (die Einteilung in Kapitel, Absätze usw.) und die Betrachtung des Erscheinungsbildes (Layout). Unter Letzterem ist dabei zum einen die Einteilung des Dokuments in Seiten, Spalten, Blöcke usw. zu verstehen und zum anderen die Darstellung der logischen Einheiten (man spricht hier auch von logischen Objekten).

Die Aufgabe eines Struktureditors besteht nun darin, anhand einer gegebenen Strukturierungsrichtlinie (man spricht hier von einer Dokumenttypdefinition) den Benutzer bei der Erstellung und Bearbeitung der logischen Texteinheiten zu unterstützen. Die Erstellung des äußeren Erscheinungsbildes kann ein System genau dann automatisch leisten, wenn die logische Strukturierung bekannt ist (alle Kapitel erster Ordnung beginnen auf einer neuen Seite usw.).

Der Prototyp-Struktureditor basiert auf dem oben erwähnten SGML-Standard.

3.3 Der Graphikeditor

Mit Hilfe des Graphikeditors ist es möglich, Flußdiagramme, Blockdiagramme, Organigramme, Firmenlogos und Business Diagramme (Balken, Torten, Histogramme) zu erstellen.

Um eine Unabhängigkeit von den Darstellungsmedien (Bildschirm, Drucker) zu gewährleisten, basiert die derzeit eingesetzte Software auf dem Standard VDI *Virtual Device Interface* /7/, der Graphik-Primitive geräteunabhängig beschreibt.

Die erzeugten Diagramme können auf zwei Arten abgespeichert werden: als Arbeitsfile (editoreigenes Format, editierbar) und als Metafile (nicht editierbar). Unter einem Metafile versteht man eine Datei, die Graphikinformation in einem Standard Format VDM *Virtual Device Metafile* /8/ enthält. Die von dem Graphik Editor erzeugten Graphik-Befehle werden über die VDI-Schnittstelle mittels eines speziellen Treibers, dem Metafile Generator, in das File geschrieben. Soll ein Metafile verändert werden, wird das zugehörige Arbeitsfile editiert und danach ein neues Metafile erzeugt. Ein Metafile enthält in der vorliegenden Implementation genau ein Bild.

Zur Darstellung werden das Metafile gelesen, der Inhalt interpretiert, und die Graphik-Befehle über das VDI auf einem Ausgabegerät angezeigt. Hierbei wird das gesamte Bild in ein vorher definiertes Fenster auf der Gerätedarstellungsfläche skaliert (s. Abb. 5).

3.4 Der Bildeditor

Für die Bearbeitung und Darstellung photographischer Informationen ist ein entsprechender Bildeditor vorgesehen. Die Bildverarbeitung beschäftigt sich, anders als die objektorientierte Graphik mit reinen Pixeldaten (Pixel = picture element). Insbesondere lassen sich dabei auch Unterschriften in ein Dokument einbringen.

Um Bilder einlesen zu können, wird ein Scanner eingesetzt. Kostengünstige Geräte haben heute eine Auflösung von 200300 Linien pro Inch (Ipi) und besitzen einen eigenen Prozessor zur Datenkompression, Erzeugung von PseudoGrauwerten usw.

Um Bilder manipulieren und darstellen zu können, muß ein Bildeditor folgende Funktionen bereitstellen:

- Einfügen bzw. Löschen eines Bildausschnitts. Hierbei kann der herheinzukopierende Bildteil transparent oder überdeckend eingefügt werden. Darüber hinaus ist es möglich den Bild ausschnitt zu drehen, zu spiegeln oder aber auch zu stauchen oder zu strecken.
- Umfangreiche Gestaltungsmöglichkeiten bieten die unterschiedlichen Arten, Teile eines Bildes einzufärben. Hierunter fallen das Löschen von Bereichen, "Einsprühen" mit verschiedenen Mustern und das Eintragen von Freihandzeichnungen.

Ein wesentlicher Punkt bei der Darstellung eines Bildes ist die Anpassung an die gewünschte, Ausmaße. Hierbei soll möglichst wenig wesentliche Bildinformation verlorengehen und der

optische Eindruck ansprechend sein. In dem häufig auftretenden Fall der Verkleinerung von Schwarzer textähnlicher Information auf weißem Grund, führt dies zu einer verstärkten Berücksichtigung der schwarzen Bildpunkte (Black Pale Preservation) und der Abbildung gleich breiter vertikaler Linien auf wiederum gleich breite Linien (Line Width Equality Preservation).

Die anfallenden sehr großen Datenmengen verlangen eine verdichtete Darstellung. Als erste Stufe hierzu faßt man zusammenhängende horizontale Folgen von Punkten gleicher Farbe zu Lauflängen zusammen. In einer zweiten Bearbeitungsstufe werden die Lauflängen selber komprimiert dargestellt. Abbildung 3 stellt den Kompressionsgrad für eine voll beschriebene A4 Seite dar, aufgenommen mit einer Auflösung von 300 lpi. Der Kompressionsgrad hängt jedoch stark von dem Inhalt der Vorlage ab.

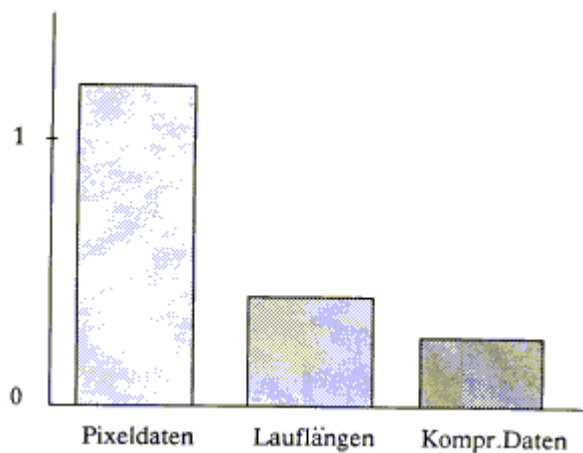


Abb. 3: Kompressionsgrad einer A4Seite (Datenvolumen in MB)

Die Darstellung in Pixelform und durch Lauflängen erlauben (im Gegensatz zur komprimierten) eine Manipulation des Bildinhaltes. Die somit auftretenden Datenströme veranschaulicht die Abbildung 4.

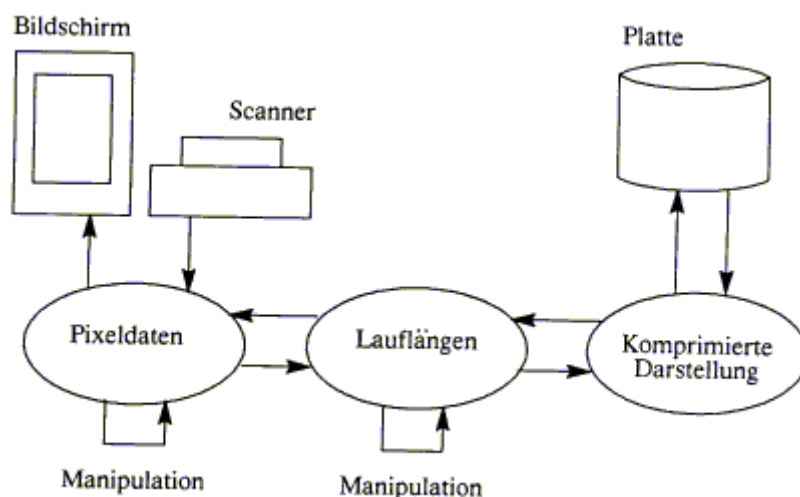


Abb. 4: Datenströme bei der Bildverarbeitung

3.5 Der Formatierer

Die Erstellung des endgültigen äußeren Erscheinungsbildes eines Dokumentes wird von der Formatierkomponente übernommen. Im wesentlichen wird der Formatierer dabei durch Strukturierungsmerkmale (siehe 3.2) gesteuert. Aufgabe des Formatierers ist es den endgültigen Zeilenumbruch und Seitenumbruch festzulegen. Darüber hinaus legt der Formatierer die Position der einzelnen Zeichen, der Graphiken und Bilder fest. Er übernimmt die Numerierung von Seiten, Kapiteln, Fußnoten usw.

Wesentliches Merkmal für einen modernen Formatierer ist seine Flexibilität hinsichtlich der Gestaltung eines Dokumentes (Mehrspaltigkeit, Kopf und Fußzeilen) und die Fähigkeit die Möglichkeiten eines hochauflösenden Laserdruckers bzw. Pixelbildschirms voll zu nutzen.

Interaktive Formatiersysteme bieten einen hohen Bedienungskomfort, besitzen jedoch eine begrenzte Leistungsfähigkeit. Batchformatierer sind dagegen sehr vielseitig, haben aber eine komplizierte Kommandosprache. Wir entschieden uns für den Batchformatierer TeX in Verbindung mit der komfortablen Benutzeroberfläche des Text/Struktureditors.

3.6 Der Darsteller

Aufgabe der Darstellungskomponente (Viewing) ist es, das geräteunabhängige Ausgabeformat des Formatierers (DVI), des Graphikeditors (VDI) und des Bildeditors die jeweilig spezifischen Ausgabegeräte (Bildschirm, Drucker) aufzubereiten, wobei die unterschiedlichen Auflösungsmöglichkeiten berücksichtigt werden.

Abbildung 5 soll die Integration von Graphik und Bild in Text veranschaulichen: Für die Integration eines Bildes in ein Dokument wird vom DVI Interpreter eine sogenannte Graphik Box erzeugt. Er übergibt deren Position, geometrische Abmessungen sowie den Namen der gewünschten Metadatei an den entsprechenden Interpreter, der das Bild dann, auf die passende Größe skaliert, vollständig in dieser Box darstellt.

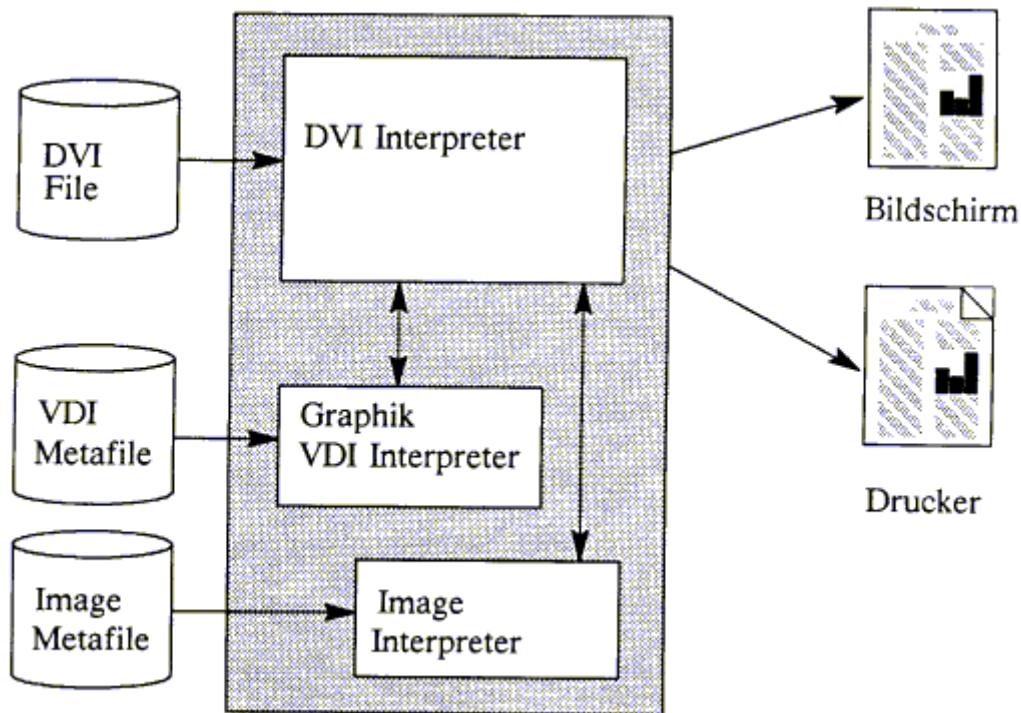


Abb. 5: Darstellen eines Dokuments

4 Die Dokumentverwaltung

In den heutigen Bürosystemen stellt das Dokument die kleinste adressierbare Einheit aus der Sicht der klassischen Verwaltungssysteme (Ablage, Archiv) dar. Um die bereits vorhandenen Software-Applikationen auch in komplexen Dokumenten weitgehend unverändert einsetzen zu können, ist die Bereitstellung eines Verwaltungssystems für Dokumentportionen nötig. Abbildung 6 zeigt beispielhaft die Partitionierung des Dokuments "Einladung" in eine Ansammlung von Portionen im "Processable Format", die nach einer Formatierung ein fertiges, zusammenhängendes Dokument ergeben.

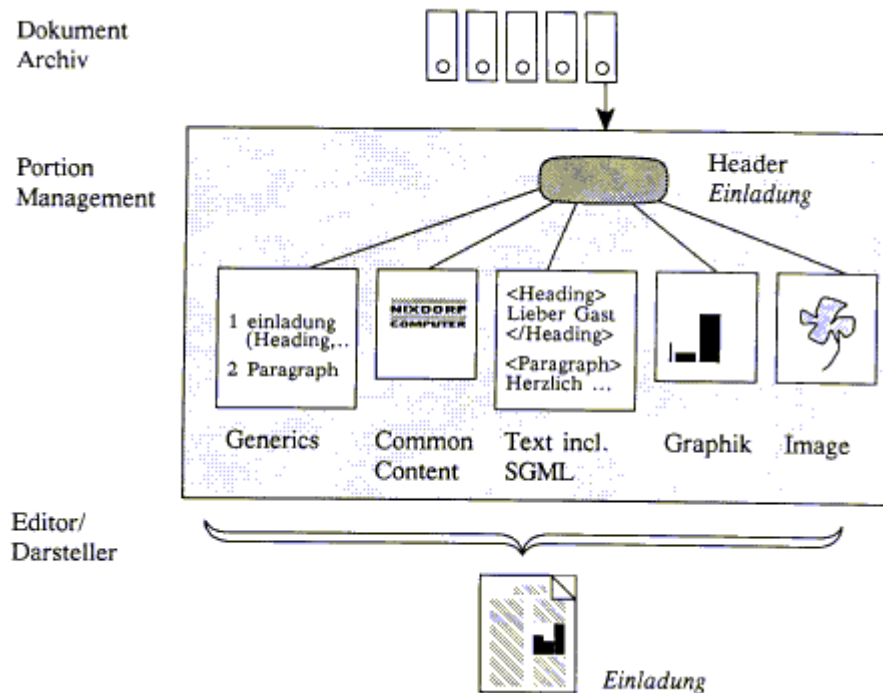


Abb. 6: Partitionierung von Dokumenten

Ein typisches Dokument kann beispielsweise folgende Portionen enthalten:

- *Dokument Header* (DH) für Besitz und Verwaltungsattribute des Dokuments, "Document-Profile".
- *Dokument Generics* für generische Logische/Layout-Strukturdefinitionen.
- *Common Content* (z.B. Logos).
- *Text* (einschließlich der SGML-Annotation zur spezifischen logischen Strukturierung des Documents).
- Pro *Graphik* ein Metafile und ggf. ein Arbeitsfile.
- Pro *Bild* ein File.

Die Partitionierung der Dokumente ermöglicht eine effiziente Nutzung von Kommunikationswegen und Speicherplatz, da redundantes Abspeichern und Übertragen von Portionen vermieden wird, ohne dabei die Robustheit des Systems einzuschränken.

Aus der Sicht der Applikation ist die Partitionierung sinnvoll, um etwa die unterschiedlichen SubEditoren problemlos in das integrierte Bürosystem einzubinden. Aus der Sicht der Verwaltung eröffnet die Einführung von Portionen neue Möglichkeiten:

- Einheitliche, systemunabhängige logische Service Schnittstelle zu den Applikationen.
- Redundanzfreies Speichern und Übertragen durch:
 - Multiple Dokumentversionen:* Nur die jeweils veränderten Portionen und getrennte DH müssen extra gespeichert werden, alle restlichen Portionen bleiben gemeinsamer Besitz. Hier eröffnet sich auch die Möglichkeit, die Veränderung in Form von Delta-Files zu speichern.
 - Multi-Besitz:* Mehrere Benutzer können ein Dokument (mit unterschiedlichen Zugriffsrechten) gleichzeitig besitzen. Für jeden Eigner wird dabei ein neuer DH angelegt.

- Ausnutzung unterschiedlicher Kommunikationswege für unterschiedliche Informationen.
- Örtlich/ organisatorisch getrennte Speicherung der Portionen, etwa aus: Sicherheitsanforderungen, Migration in ein Langzeit-Archiv, wobei ggf. nur der DH schnell erreichbar ist, Ausnutzung unterschiedlicher Preis/ Leistungsverhältnisse von Speichermedien (Harddisk, Optische Platte, etc.), Backup, und Systemerweiterung.
- Unterstützung beim Versand (s.u.) und der Vorgangsbearbeitung (Trennung von Masken und Inhalt).

Die Vorteile des Einsatzes unserer Konzeption lassen sich an einigen Zahlen verdeutlichen: Während jeder DH mit einer Größe von 1 KB zu Buche schlägt, ist für ein Dokument im Mittel von 21 KB (die Angaben schwanken zwischen 2 und 200 KB) anzusetzen. Geht man weiter davon aus, daß in einer Organisation von jedem Dokument sechs Kopien entstehen und der Versand zur Hälfte organisationsintern abgewickelt wird, so ergeben sich beträchtliche Einsparungen an Speicherbedarf und Kommunikationsaufwand, wenn die Portionsverwaltung eingesetzt wird. Durch die zentrale Bereithaltung von generischen Informationen und Common Content ergeben sich weitere Vorteile.

ECMA beginnt derzeit, Standardisierungen auch auf diesem Gebiet im Rahmen eines *Distributed Services Framework* voranzutreiben /9/.

5 Der Dokumentversand

Durch die Portionsverwaltung wird es möglich, beim (lokalen) Versand von Dokumenten von einem aufwendigen Kopieren der Inhalte Abstand zu nehmen, indem neben dem Mailheader (Umschlag in Sinne von X.400) nur der Dokumentheader (DH) an den/die Empfänger übermittelt wird. Jeder Empfänger besorgt sich die dazugehörigen Portionen dann selbst mit Hilfe der Portionsverwaltung.

Der Versand eines Dokuments beginnt, in dem der Absender mit Hilfe der Portionsverwaltung einen zweiten DH erstellt, der in der Versand-Applikation einen der X.400 Empfehlung entsprechenden Umschlag erhält. Das (Rumpf) Dokument wird dann an den Mail-Server übermittelt, von wo es dem jeweiligen Empfänger zur Verfügung gestellt wird. Dabei erfolgt sowohl auf der Absender als auch auf der Empfängerseite ein Eintrag ins Logbuch. Der Empfänger erhält zunächst den DH ausgehändigt und kann den gesamten Inhalt des Dokuments über seine lokale Portionsverwaltung anfordern.

6 Literatur

1. Blumenfeld, M.: **Dokumenten Architektur und Austauschformate**. Büroautomation 85, Berichte German Chapter of the ACM 25, S.4061
2. ECMA 101 **Office Document Architecture**. Sept. 1985
3. ISO TC97/SC18/WG8 N8 **Dokument Markup Metalanguage**. Sept. 1984
4. T.73 **Dokument Interchange Protocol for the Telematic Services**. CCITT Red Book, VII.3, Geneva 1985
5. Godbersen,H.P.; Malpeli,F.: **First Results on the Integration of Communication and Application Services in the Office**. Proc. ESPRIT Technical Week 85, North Holland Publ. Co.
6. ECMA TR/31 **Remote Operations. Concepts, Notation and ConnectionOriented Mappings**. Dec. 1985
7. ANSI X3H3 **Virtual Device Interface**. Sept. 1982
8. DIN 66293 **Datei für die Speicherung und Übertragung von Bildinformationen**. Juni 1985
9. ECMA/TC32/85/275 **Distributed Services Framework**. Oct. 1985