

Architektur-Standards für Dokumente

Zwar beinhalten rechtzeitig erhaltene Informationen häufig einen erheblichen Wettbewerbsvorteil, ihre Verarbeitung und Konsolidierung ist aufgrund der vielen Hersteller mit unterschiedlichen Systemen oft aber sehr schwierig. Die Vielfalt an Dokumentenformaten und fehlende Verarbeitungsmöglichkeiten für Raster- und Vektorgrafik bremsen den Informationsfluss in und zwischen Organisationen. Einheitliche Dokumentenstandards schaffen hier Abhilfe.

Das Dokument in elektronischer und Papierform dient als wichtigster Informationssträger. Es kann unter zwei Blickwinkeln betrachtet werden: einem logischen

* Von Dr. Richard Schlupp,
Rank Xerox AG

und einem graphischen (Layout). Die logische Struktur wird vom Autor während des Editierens oder Überarbeitens bestimmt und verknüpft den Inhalt eines Dokuments mit einer Hierarchie von logischen Objekten, wie Anhängen, Überschriften, Fussnoten und Abbildungen. Der graphische Aufbau dagegen verknüpft den Inhalt mit einer Hierarchie von Layout-Objekten. In einem "Compound Document" können verschiedene inhaltliche Grundelemente verwendet werden, wie zum Beispiel mathematische Formeln, Text, Bitmaps, Raster- oder Vektorgraphiken, digitalisierte Audio- und Videosignale und Kalkulationsblätter.

Publishing-Modell

Ein Dokument durchläuft in seiner Erstellungs- und Wiedergabephase drei wichtige Prozesse (vgl. Abb. 1). Der Editierprozess generiert den Dokumenteninhalt und den logischen Aufbau (Revisable Form Document), welcher im Layout Prozess für die Ausgabe formatiert wird (Formatted Final Form Document). Der Wiedergabeprozess (Imaging Process) bringt das Dokument in eine für den Menschen lesbare Form, z.B. auf Papier oder einen Bildschirm. Das Publishing-Modell bestimmt eindeutige Schnittstellen der Arbeitsteilung, z.B. kann der Autor eines Buches dessen Gestaltung einem Grafiker überlassen und sich völlig auf seinen Text konzentrieren.

Anforderungen

Zur Realisierung von Compound Dokumenten existieren hauptsächlich vier Voraussetzungen:

- Die Trennung von logischem und gestalterischem Dokumentenaufbau,
- Die Koexistenz von Text, Bild, Kalkulationsblatt, Graphik, Ton, etc.,
- Die Gleichartigkeit von Dokumenten verschiedener Autoren,

- Die Möglichkeit des Austausches von "Compound Documents" in heterogenen und offenen Systemwelten.

Um diesen Anforderungen gerecht zu werden, wurden verschiedene 'Document Architecture Standards' entwickelt. Der ODA/ODIF-Standard wurde für den Austausch von Dokumenten in Büroumgebungen entwickelt. Die SGML/SDIF-Standards dienen dem Dokumentenaustausch in Publishing-Umgebungen, meist zwischen einem Autor und einem Verlag. Der SPDL-Standard beschreibt ein aus Text, Bildern und geometrischen Graphiken bestehendes elektronisches Dokument in einer für den Druck passenden Form. Für den Austausch von Dokumenten aus dem kommerziellen Sektor, also Formularen, Bestellungen oder Rechnungen, besteht der EDIFACT-Standard. Weil hier Aufbau und Format bereits vorgegeben sind, wird im folgen-

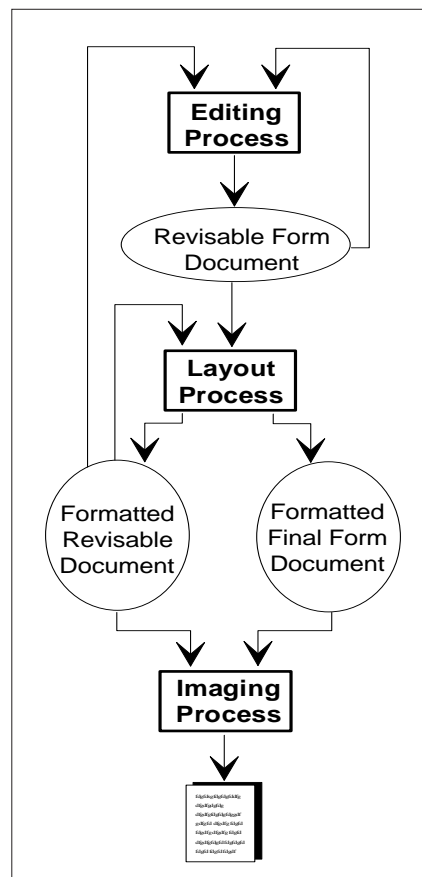


Abb. 1: Das Publishing-Modell

den genauer auf SPDL, ODA und SGML eingegangen.

SPDL

Die Verbreitung von Laserdruckern ermöglichte den kostengünstigen Druck von Dokumenten, die Text und Graphiken enthalten. Dadurch wuchs das Interesse an einer standardisierten Drucker-schnittstelle (PDL). Eine solche definiert SPDL und beschreibt damit das Format eines Final Form Document. SPDL ist eine Weiterentwicklung aus Postscript II (Adobe) und Interpress (Xerox) und für den Einsatz in einer grossen Anzahl von Druck- und Publishing-Umgebungen bestimmt. Dokumente, die von ODA-Systemen für den Druck generiert wurden können einfach und effizient beschrieben werden. Um die Ausgabe auf verschiedenen Geräten zu ermöglichen, müssen druckerspezifische Parameter (DPI) bestimmt werden. Es handelt sich zum Beispiel um Ausgabe-Position, X oder Y Schriftbildverschiebung, Trennblatt, Bindart, Simplex oder Duplex, Kopien, Ausgabemedium mit Grösse, Farbe, Gewicht, etc. SPDL ermöglicht die Integration in eine Umgebung von Geräten verschiedener Hersteller und ist Teil der Standardfamilie ISO/IEC JTC1/SC18.

Wiedergabeprozess bei SPDL

Der Wiedergabeprozess erstellt Druckseiten oder Bildschirmhalte mit der Auflösung der SPDL Page Description Language. Er besteht aus fünf Teilen:

1. Der SPDL-Dokument-Struktur, welche das Ergebnis der hierarchischen Unterteilung eines SPDL-Dokuments in immer kleinere Elemente, den Structure Elements, ist.
2. Dem Verarbeitungs-Modell, dessen Basis die Trennung des SPDL-Dokument in Dokument-Struktur und Dokument-Inhalt ist. Dabei werden die Tokensequences, die unterste Ebene der Structure Elements, identifiziert und ausgeführt, wodurch sukzessive eine leere Seite aufgebaut wird.
3. Einem Imaging- bzw. Wiedergabe Modells, einer zweidimensionalen graphischen Darstellung bei welcher Bildelemente sukzessive auf der Seite durch Aufbringen von Tinten plazierte werden (Tinte bezeichnet die Farbe und das visuelle Gefüge, das vom Ausgabegerät angezeigt wird).
4. Einem Koordinatensystem, das aus einem fest definierten kartesischen (RCS) und einem flexiblen Koordinatensystem (UCS) besteht. Ersteres bestimmt die Lage und Ausrichtung auf einem Medium (z.B. Druckträger) und letzteres die Lage auf dem Bild selbst.

5. Ein *Schriften-Modell* bzw. eine Font-Architektur. Ein Font ist eine Sammlung von Bildern desselben Grunddesigns, z.B. identifizierte graphische Symbole, alphanumerische Zeichen und wird für die Ausgabe von Text verwendet.

SPDL wurde im März 1991 entworfen und soll im November 1993 durch die ISO veröffentlicht werden.

ODA/ODIF

Der international anerkannte Standard ODA/ODIF wurde seit 1982 für den Austausch von "Compound Documents" in Büroumgebungen entwickelt. Gegenstand von ODA/ODIF ist ein Dokument-Architekturmodell und ein Format für integrierte Text- und Graphik-Dokumente, die sowohl zum Zweck des Weiterbearbeitens (Revisable Form) als auch zum Zweck der Wiedergabe in der vom Autor beabsichtigten Formatierung (Final Form) ausgetauscht werden sollen. Wegen der oft bestehenden Unterschiede zwischen Funktionsspektrum der Sender-Dokument-Applikation und dem der Empfänger-Applikation, bietet ODA die Möglichkeit, Dokumente einer bestimmten Kategorie zuzuordnen. Bislang bestehen drei hierarchisch aufgebaute Kategorien (DAP), die Funktionalitäten von reinem Text bis zum Text mit komplexer Grafik bezeichnen.

ODA-Dokumente sind anhand zweier Strukturen beschrieben: Die *logische* Struktur stellt einen Satz von Regeln dar, z.B. dass ein Dokument ein oder mehrere Kapitel besitzt. Die *graphische* Struktur enthält die zur Darstellung des Dokumentinhalts notwendigen Informationen, etwa die Seitengröße. Beide Strukturarten beinhalten generische und spezifische Elemente. Generische Strukturelemente gelten für eine Gruppe von Dokumenten, wobei jedes Dokument innerhalb einer Gruppe eine spezifische logische und grafische Struktur, die nur dieses selbst betrifft, besitzt. Zusätzlich besteht ein ODA Dokument aus einem Document Profile, den Daten über Autor, Titel, etc. und den Styles, die das Erscheinungsbild des Dokumentes regeln. Teil 5 des ODA Standard, ISO 8613, beschreibt die Spezifikationen des Austauschformats. Das Office Document Interchange Format (ODIF) ermöglicht den Austausch von ODA-Dokumenten über OSI-Netzwerke.

SGML

SGML wurde für den Einsatz in der Publishing-Branche entwickelt und 1986 als ISO 8879-Norm verabschiedet. Durch Einfügen von generalisiertem Markup kann der Autor eines Dokuments typographische oder gestalterische Überlegungen dem Publisher überlassen.

SGML enthält in einer Metasprache implementierte Regeln für die Beschreibung von Dokumenten, die ausschliesslich in Revisable Form bestehen. Die Regeln beschreiben die Dokumentart (z.B. technische Berichte, Handbücher, Zeitschriften), die logische Struktur (z.B. Titel, Kapitel, Fussnoten, Abbildungen) und Referenzen betreffend ausgegliederter Dokumentinhalte (z. B. Sonderzeichen, Graphiken, Zeichnungen, Hypertext, Audio- und Videoauszüge). Die SGML-Regeln dienen der Markierung aller Elemente, die ein Benutzer in einem Dokument identifizieren möchte. Menschliches Know-how geht durch die Verwendung von generischem Markup nicht verloren. Änderungen in Betriebssystemen, Programmen und Benutzerschnittstellen sind brennende Probleme bei den heutigen Computern. Die generische Markierung wird auch beim Wechsel des eingesetzten Textverarbeitungssystems beibehalten. Unterstützt wird SGML durch eine Parser-Software zur Analyse von Zeichenfolgen oder zur Erkennung des Dokumentinhaltes.

An SGML werden im wesentlichen folgende Anforderungen gestellt:

- Unterstützung durch möglichst viele Textverarbeitungssysteme.
- Unbegrenzte Anzahl an Zeichensätzen und Landessprachen (ISO 639).
- Unlimitierte Auswahl von Datenflussarten, Datenbanksystemen und File-Organisationen.
- Markierte Elemente können mit anderen Daten gleichzeitig vorhanden sein.
- Die Markierung ist für Mensch wie Maschine verständlich.

Der SDIF-Standard spezifiziert eine Datenstruktur, mit der sich ein SGML-Dokument, das vielleicht in mehreren getrennten Einheiten gespeichert ist, für den Austausch in einen Datenstrom verpacken lässt. Der Austausch erfolgt über OSI- und nicht-OSI-konforme Netzwerke.

ODA versus SGML

ODA ist ein komplexer Standard, weil er alle Inhaltstypen und Zusammenhänge der Struktur definiert, die für die vollständige Interpretation eines empfangenen Dokuments notwendig sind. ODA ist der Standard für Bürosysteme in einer dezentralen Büroumgebung, wo Final Form Documents benötigt werden oder Dokumente in ihrer endgültigen Form zu verteilen sind. Die Anwendungsbreite von SGML ist wie erwähnt fast unbegrenzt, wobei SGML aber nicht alle zu einem Dokument gehörigen Informationen standardisiert. Es ist vielmehr ein Mittel zur Beschreibung einer unbegrenzten Anzahl von Dokumenten, was auf verschie-

denste Weise deren einfachere Verarbeitung in verschiedenen Applikationen ermöglicht.

Ausblick

Die ODA-Architektur ist so ausgelegt, dass andere Inhaltstypen, wie Ton, Kalkulationsblätter, Formeln etc., in Erweiterungen berücksichtigt werden können. Ab 1993 soll ein Toolkit zur raschen

DAP:	Document Application Profile
DPI:	Document Production Instructions
ECMA:	European Computer Manufacturers Association
PDL:	Page Description Language
ODA/ODIF:	Office Document Architecture / Office Document Interchange Format
RCS:	Reference Coordinate System
SGML/SDIF:	Standard Generalized Markup Language / Standard Document Interchange Format
SPDL:	Standardized Page Description Language
UCS:	User Coordinate System

Abb. 2: Abkürzungsverzeichnis

Entwicklung von ODA-Anwendungen verfügbar sein. Ferner wird im Rahmen des ESPRIT-Projektes PODA-SAX die Integration von ODA in X.400 und Security-, Document Filing and Retrieval-Diensten angestrebt. Ziel eines weiteren Projektes (EWOS PT N011) ist die Realisierung eines Austausches von Dokumenten unter SGML- und ODA- Umgebungen, insbesondere von ODA nach SGML. Dieses Projekt ist ein weiterer Schritt in Richtung harmonischer Koexistenz von ODA/SGML-Dokumenten in einer Reihe von Büro- und Publishing-Umgebungen

Die Kooperation der verschiedensten Dokument-Applikationen, wie Archivierung, Retrieval, Drucken, Editieren, HyperText, ist in Zukunft nur möglich mit standardisierten, aufeinander abgestimmten, Dokument-Architekturen. Der Nutzen von Investitionen in hochwertige Dokument-Applikationen wird enorm steigen und sich besser mit den eigentlichen Bedürfnissen in Einklang bringen lassen - hin zu einer grenzenlosen Büroautomation. ■

* Dr. Richard C. Schlupp ist Vize-Direktor bei der Rank Xerox AG Schweiz und verantwortlich für das System Competence und Know-how Center.