

Word und XML

XML-Unterstützung in
Word2003 und 2007

2/8/2010

Zur Person:

Manuel Montero Pineda

Dipl.-Wirtschaftsinformatiker (FH), M.A.

Entwickler und Berater im Bereich WordML, XSL, XSL-FO.

Hauptautor von „Professionelle XML-Verarbeitung mit Word“ und „XSL-FO in der Praxis“ beides erschienen beim dpunkt-Verlag.

Basisfunktionen

1. Word als XML-Editor
2. WordML/OpenXML
3. WordML-Stylesheets
4. Smart-Documents

XML-Funktionalitäten in Word

- Benutzerführung und Validierung von XML-Dateien
- Validierung erfolgt mit Schemas

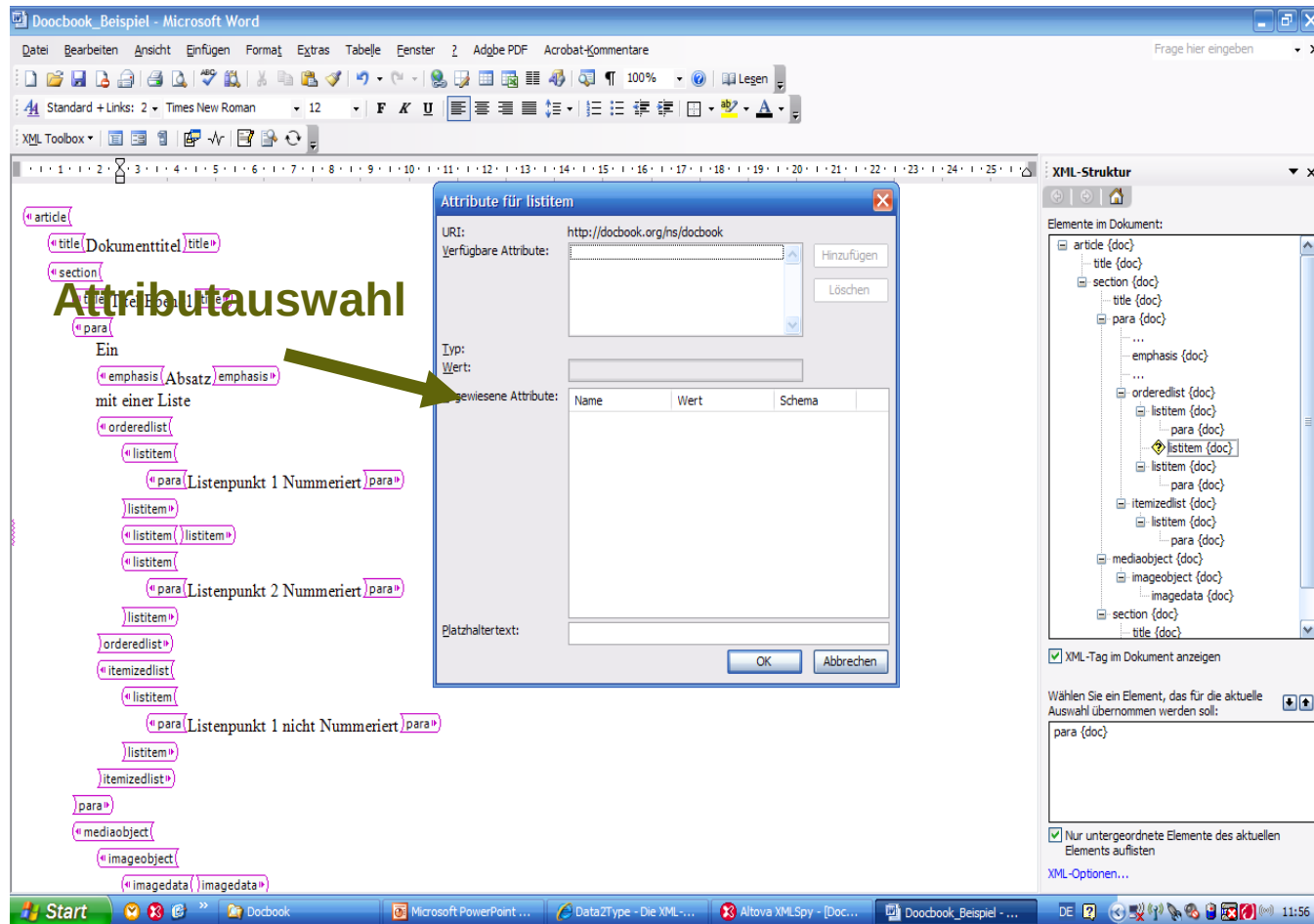
The screenshot displays the Microsoft Word 2003 interface with the document 'Dooobook_Beispiel'. The main text area shows XML tags overlaid on the document content, such as `<article>`, `<title>`, `<section>`, `<para>`, `<emph>`, `<listitem>`, `<orderedlist>`, `<itemizedlist>`, `<mediaobject>`, `<imageobject>`, and `<imagedata>`. Two green arrows point from the text 'Validierung' and 'Elementauswahl' to the 'XML-Struktur' pane on the right. The 'XML-Struktur' pane shows a hierarchical tree of the document's XML elements, including 'article (doc)', 'title (doc)', 'section (doc)', 'title (doc)', 'para (doc)', 'emph (doc)', 'orderedlist (doc)', 'listitem (doc)', 'para (doc)', 'itemizedlist (doc)', 'listitem (doc)', 'mediaobject (doc)', 'imageobject (doc)', and 'imagedata (doc)'. Below the tree, there are checkboxes for 'XML-Tag im Dokument anzeigen' and 'Nur untergeordnete Elemente des aktuellen Elements auflisten', and a section for 'XML-Optionen...'.

Validierung

Elementauswahl

XML-Funktionalitäten von Word 2003 und 2007

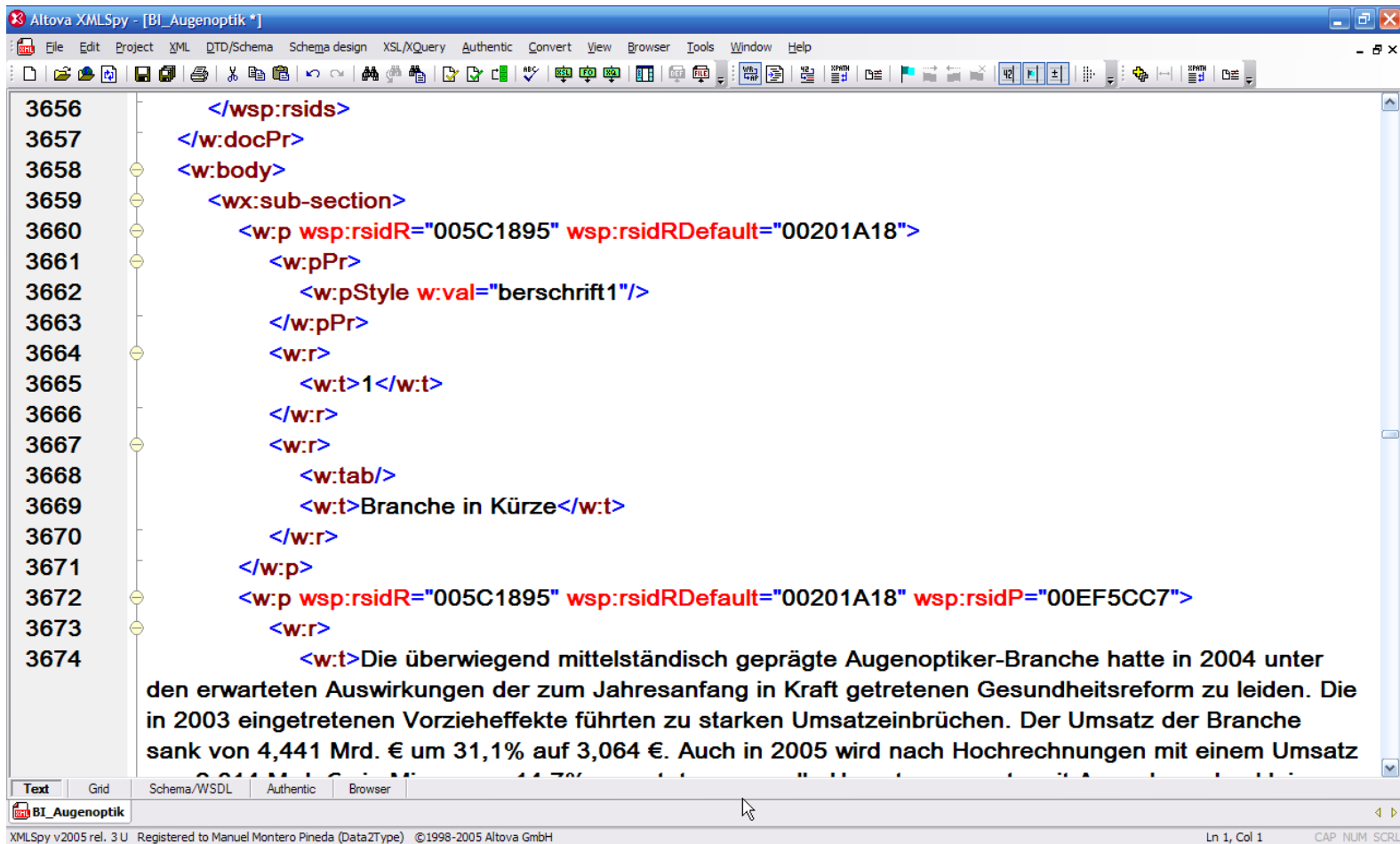
1.1 Word als XML-Editor



WordML

- WordML ist XML-Format von Microsoft und ein Teil von mehreren XML-Formaten innerhalb von Office 2003 und 2007.
- Es können alle Informationen eines Worddokumentes in WordML abgespeichert werden.

=> Migration alter Word-Dokumente (.doc) in XML (WordML-Dokumente)



The screenshot displays the Altova XMLSpy interface with the file 'BI_Augenoptik *'. The main window shows XML code for a Word document. The code is structured as follows:

```
</wsp:rsids>
</w:docPr>
<w:body>
  <wx:sub-section>
    <w:p wsp:rsidR="005C1895" wsp:rsidRDefault="00201A18">
      <w:pPr>
        <w:pStyle w:val="berschrift1"/>
      </w:pPr>
      <w:r>
        <w:t>1</w:t>
      </w:r>
      <w:r>
        <w:tab/>
        <w:t>Branche in Kürze</w:t>
      </w:r>
    </w:p>
    <w:p wsp:rsidR="005C1895" wsp:rsidRDefault="00201A18" wsp:rsidP="00EF5CC7">
      <w:r>
        <w:t>Die überwiegend mittelständisch geprägte Augenoptiker-Branche hatte in 2004 unter
den erwarteten Auswirkungen der zum Jahresanfang in Kraft getretenen Gesundheitsreform zu leiden. Die
in 2003 eingetretenen Vorzieheffekte führten zu starken Umsatzeinbrüchen. Der Umsatz der Branche
sank von 4,441 Mrd. € um 31,1% auf 3,064 €. Auch in 2005 wird nach Hochrechnungen mit einem Umsatz

The interface includes a left sidebar with line numbers (3656 to 3674), a top menu bar (File, Edit, Project, XML, etc.), and a bottom status bar indicating 'Ln 1, Col 1' and 'CAP NUM SCRL'.


```


Gesmischte Inhalte

Eigene XML-Dateien können in Word geladen werden und als WordML ohne Informations-verlust gespeichert werden.

Microsoft bezeichnet dies als Gemischten Inhalt (eigenes XML + WordML)

The screenshot shows the Altova XMLSpy interface with the file 'Dooobook_Beispiel *'. The XML content is as follows:

```
390 </w:compat>
391 <wsp:rsids>
392   <wsp:rsidRoot wsp:val="009C7A6D"/>
393   <wsp:rsid wsp:val="000C0428"/>
394   <wsp:rsid wsp:val="009C7A6D"/>
395   <wsp:rsid wsp:val="00D8448E"/>
396   <wsp:rsid wsp:val="00EF0DDD"/>
397 </wsp:rsids>
398 </w:docPr>
399 <w:body>
400   <ns0:article id="beispiel" lang="de" ns2:schemaLocation="http://docbook.org/ns/docbook docbook.xsd">
401     <w:p wsp:rsidR="009C7A6D" wsp:rsidRDefault="009C7A6D"/>
402     <ns0:title>
403       <w:p wsp:rsidR="009C7A6D" wsp:rsidRDefault="000C0428">
404         <w:pPr>
405           <w:ind w:left="360"/>
406         </w:pPr>
407         <w:r>
408           <w:t>Dokumenttitel</w:t>
409         </w:r>
410       </w:p>
411     </ns0:title>
412   </ns0:article>
413 </w:body>
```

A validation error is displayed at the bottom: "This file is not valid: Unexpected element 'w:p' in element 'ns0:article'. Expected: ns0:title". A "Revalidate" button is present next to the error message.

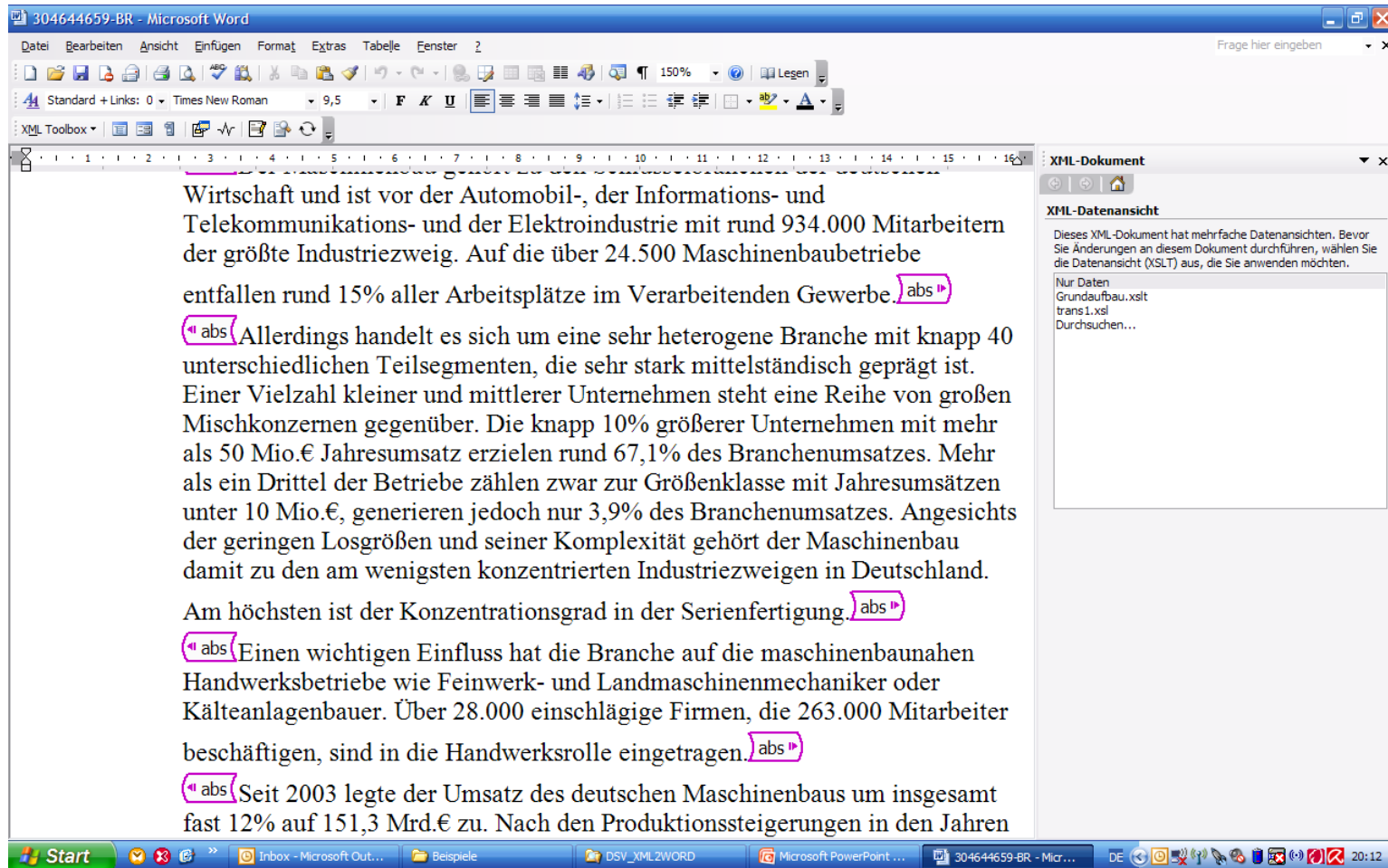
The status bar at the bottom indicates: "XMLSpy v2005 rel. 3 U Registered to Manuel Montero Pineda (Data2Type) ©1998-2005 Altova GmbH" and "Ln 349, Col 24 CAP NUM SCRL".

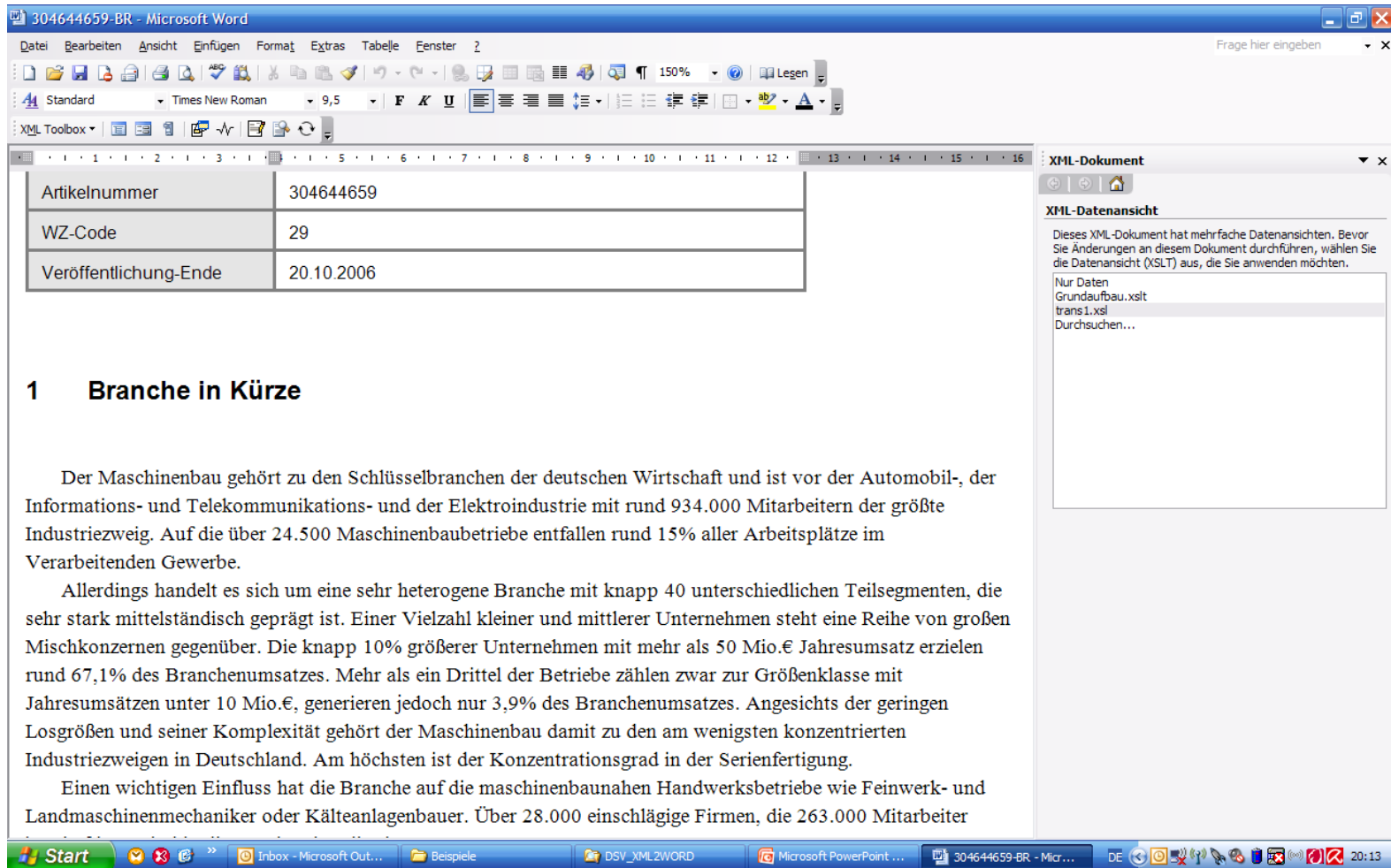
Die Speicherformate:

1. Worddokument (.doc) (Standard bis einschließlich Word 2003)
2. WordML-Dokument (.xml) (möglich seit Word 2003)
3. XML-Dokument nach eigenen Strukturen (Schema) (möglich seit Word 2003)
4. DocX unter Word 2007 (Standard seit Word 2007)

XSLT-Stylesheets

Mit Hilfe von XSLT-Stylesheets lassen sich XML-Dateien mit Layout-informationen anreichern.





The screenshot shows the Microsoft Word 2003 interface. The main document area contains a table with the following data:

Artikelnummer	304644659
WZ-Code	29
Veröffentlichung-Ende	20.10.2006

To the right of the document is the 'XML-Dokument' sidebar, which shows the 'XML-Datenansicht' (XML Data View). It contains the following text:

Dieses XML-Dokument hat mehrfache Datenansichten. Bevor Sie Änderungen an diesem Dokument durchführen, wählen Sie die Datenansicht (XSLT) aus, die Sie anwenden möchten.

Nur Daten
Grundaufbau.xslt
trans1.xsl
Durchsuchen...

Below the table, the document content begins with the section header '1 Branche in Kürze'. The text describes the machine building industry in Germany, mentioning its size, employee count, and its position within the manufacturing sector. It also discusses the industry's heterogeneity and its concentration in certain regions.

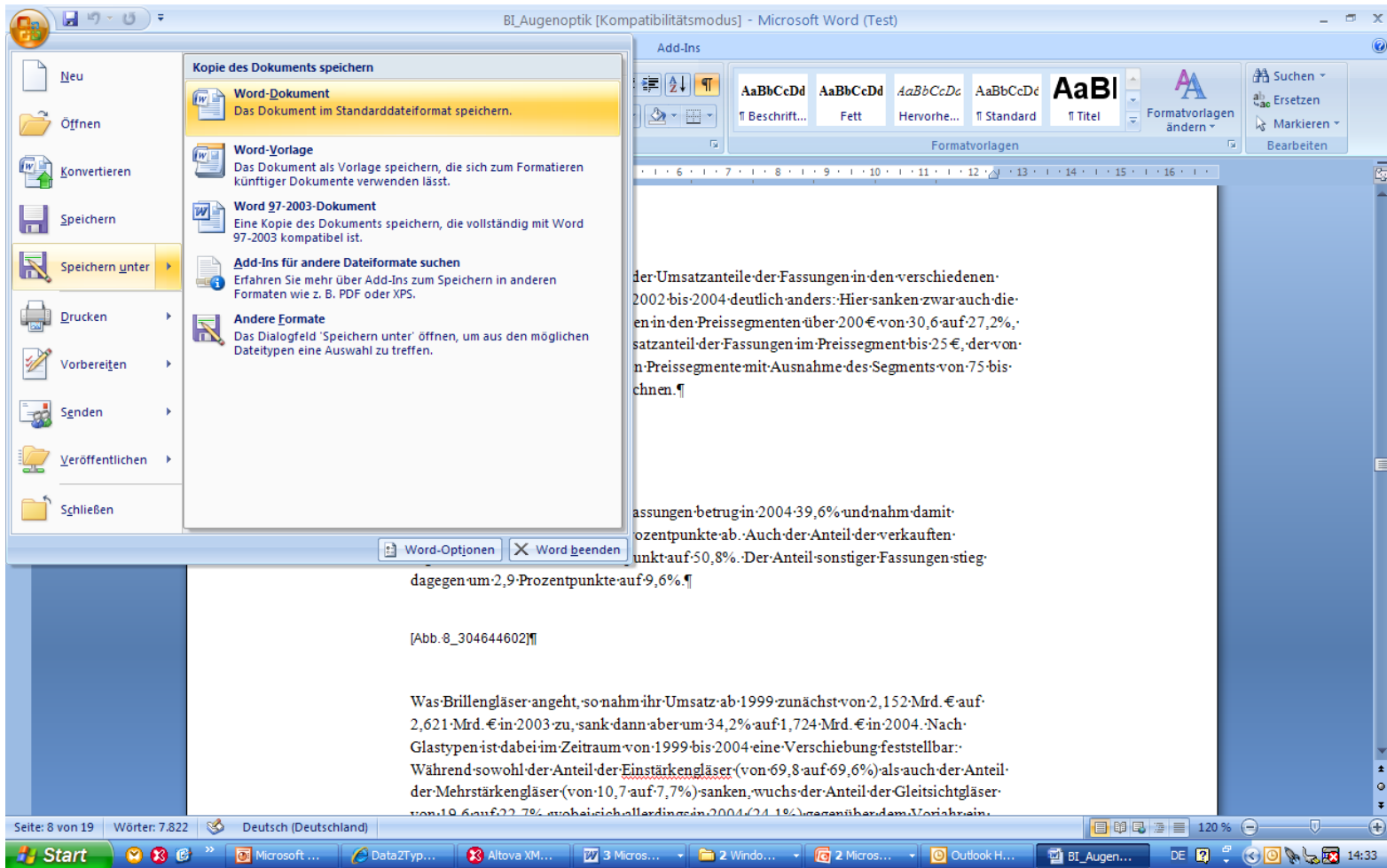
Der Maschinenbau gehört zu den Schlüsselbranchen der deutschen Wirtschaft und ist vor der Automobil-, der Informations- und Telekommunikations- und der Elektroindustrie mit rund 934.000 Mitarbeitern der größte Industriezweig. Auf die über 24.500 Maschinenbaubetriebe entfallen rund 15% aller Arbeitsplätze im Verarbeitenden Gewerbe.

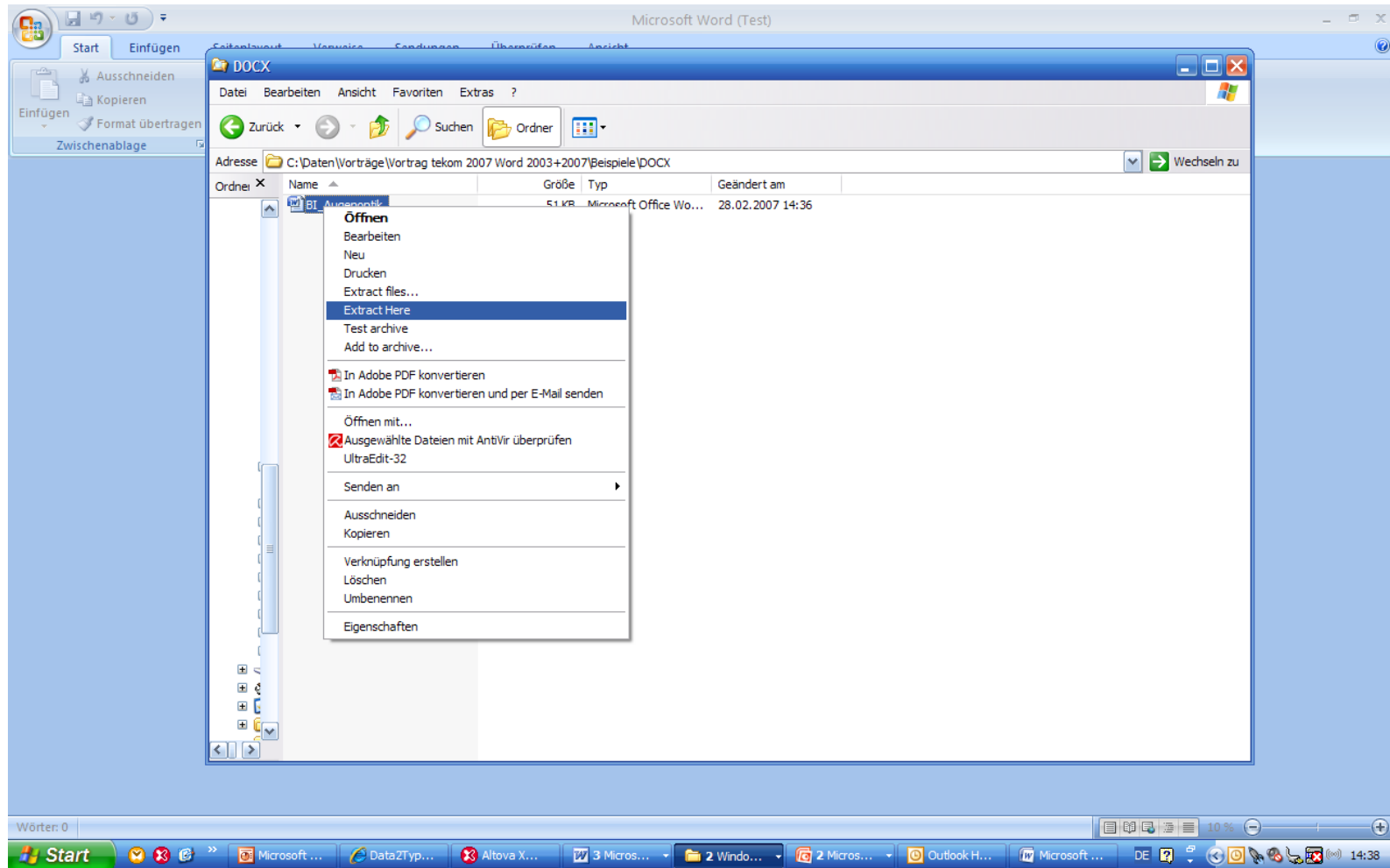
Allerdings handelt es sich um eine sehr heterogene Branche mit knapp 40 unterschiedlichen Teilsegmenten, die sehr stark mittelständisch geprägt ist. Einer Vielzahl kleiner und mittlerer Unternehmen steht eine Reihe von großen Mischkonzernen gegenüber. Die knapp 10% größerer Unternehmen mit mehr als 50 Mio.€ Jahresumsatz erzielen rund 67,1% des Branchenumsatzes. Mehr als ein Drittel der Betriebe zählen zwar zur Größenklasse mit Jahresumsätzen unter 10 Mio.€, generieren jedoch nur 3,9% des Branchenumsatzes. Angesichts der geringen Losgrößen und seiner Komplexität gehört der Maschinenbau damit zu den am wenigsten konzentrierten Industriezweigen in Deutschland. Am höchsten ist der Konzentrationsgrad in der Serienfertigung.

Einen wichtigen Einfluss hat die Branche auf die maschinenbaunahen Handwerksbetriebe wie Feinwerk- und Landmaschinenmechaniker oder Kälteanlagenbauer. Über 28.000 einschlägige Firmen, die 263.000 Mitarbeiter

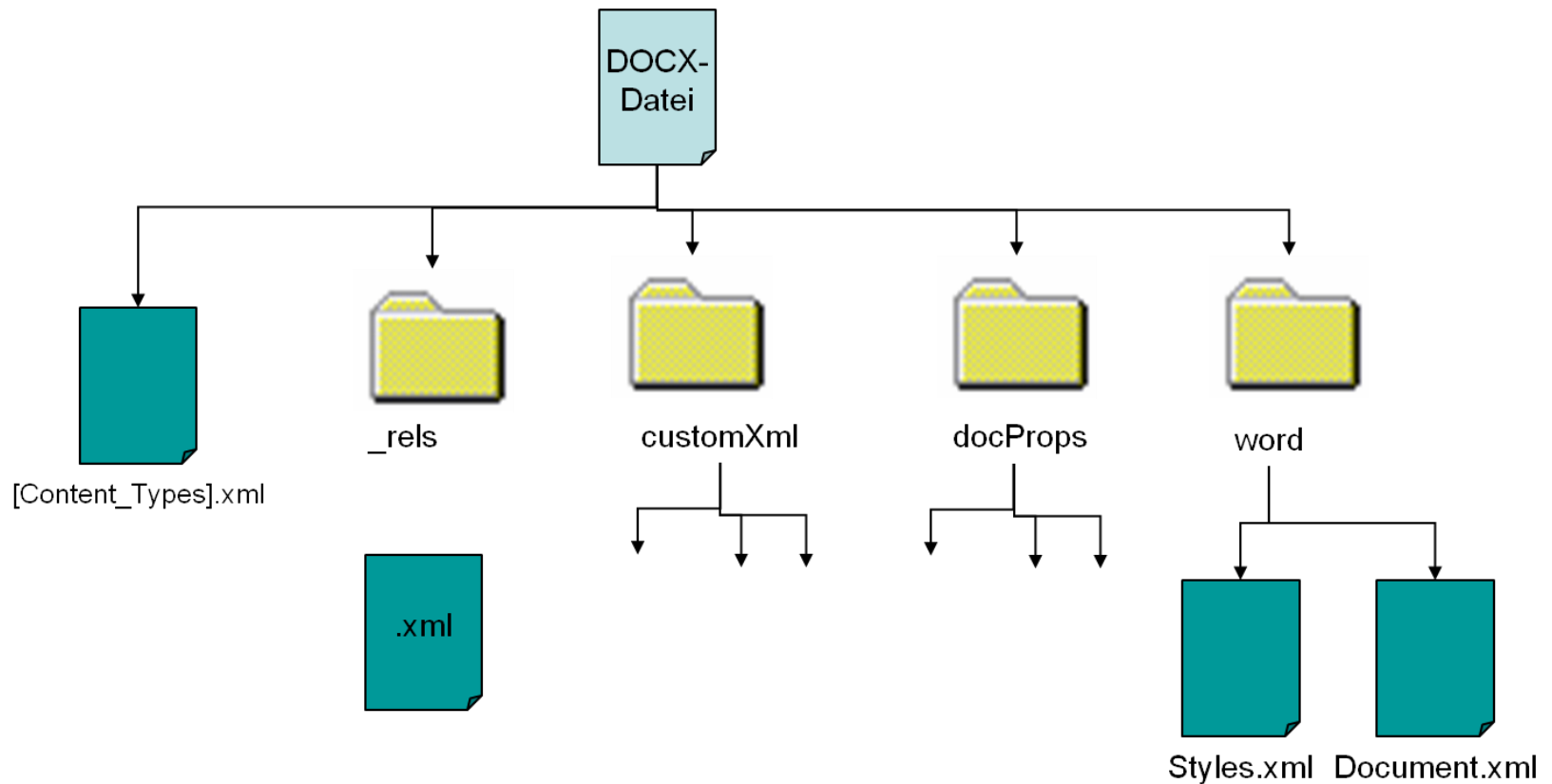
DOCX

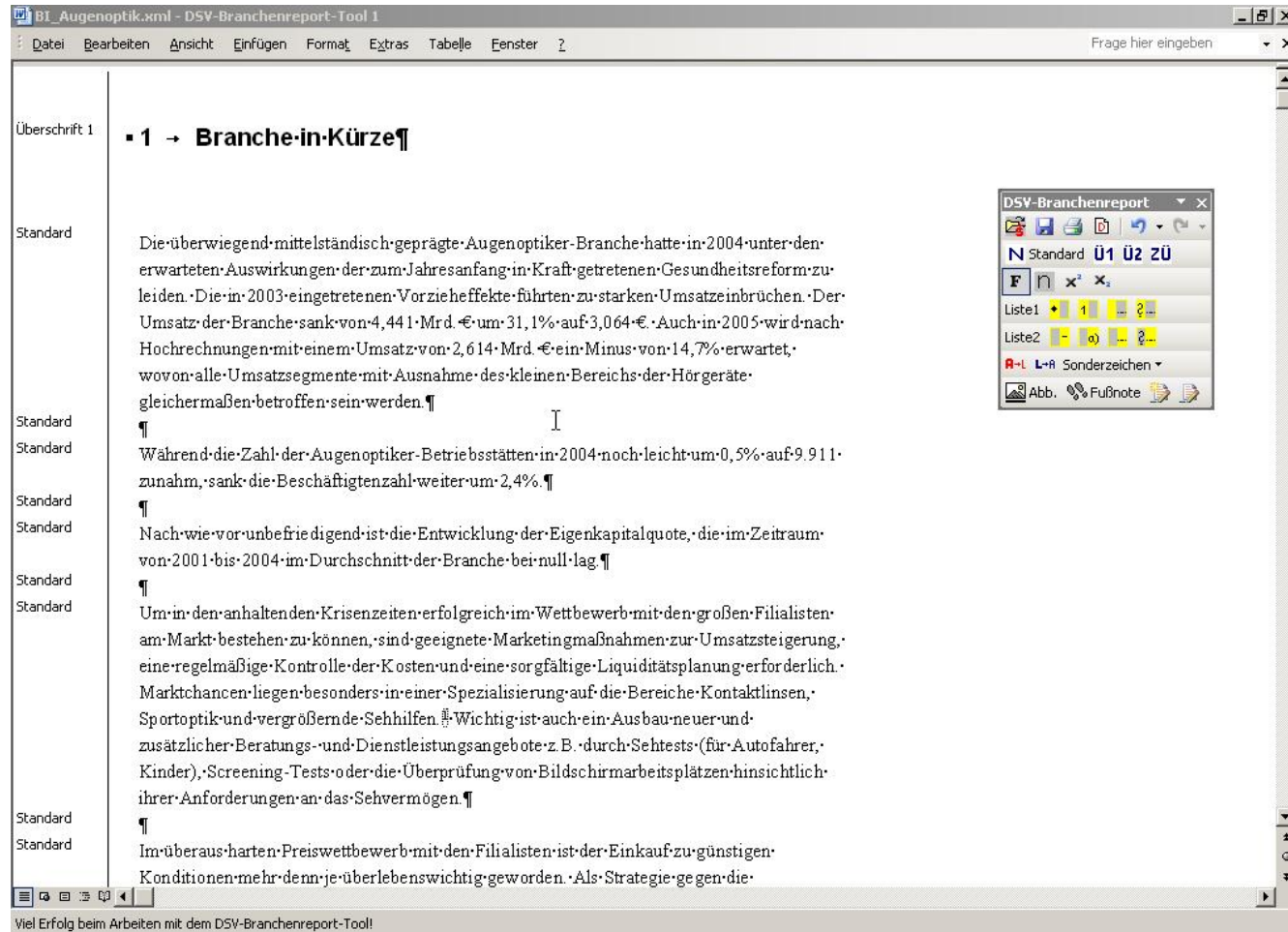
- In Word 2007 gibt es nun ein weiteres Speicherformat DOCX. Es löst das alte Speicherformat doc ab und wird als neues Standardformat eingesetzt.
- Intern wird wiederum WordML verwendet, allerdings nicht in einer Datei, sondern in verschiedenen, die gezippt werden.

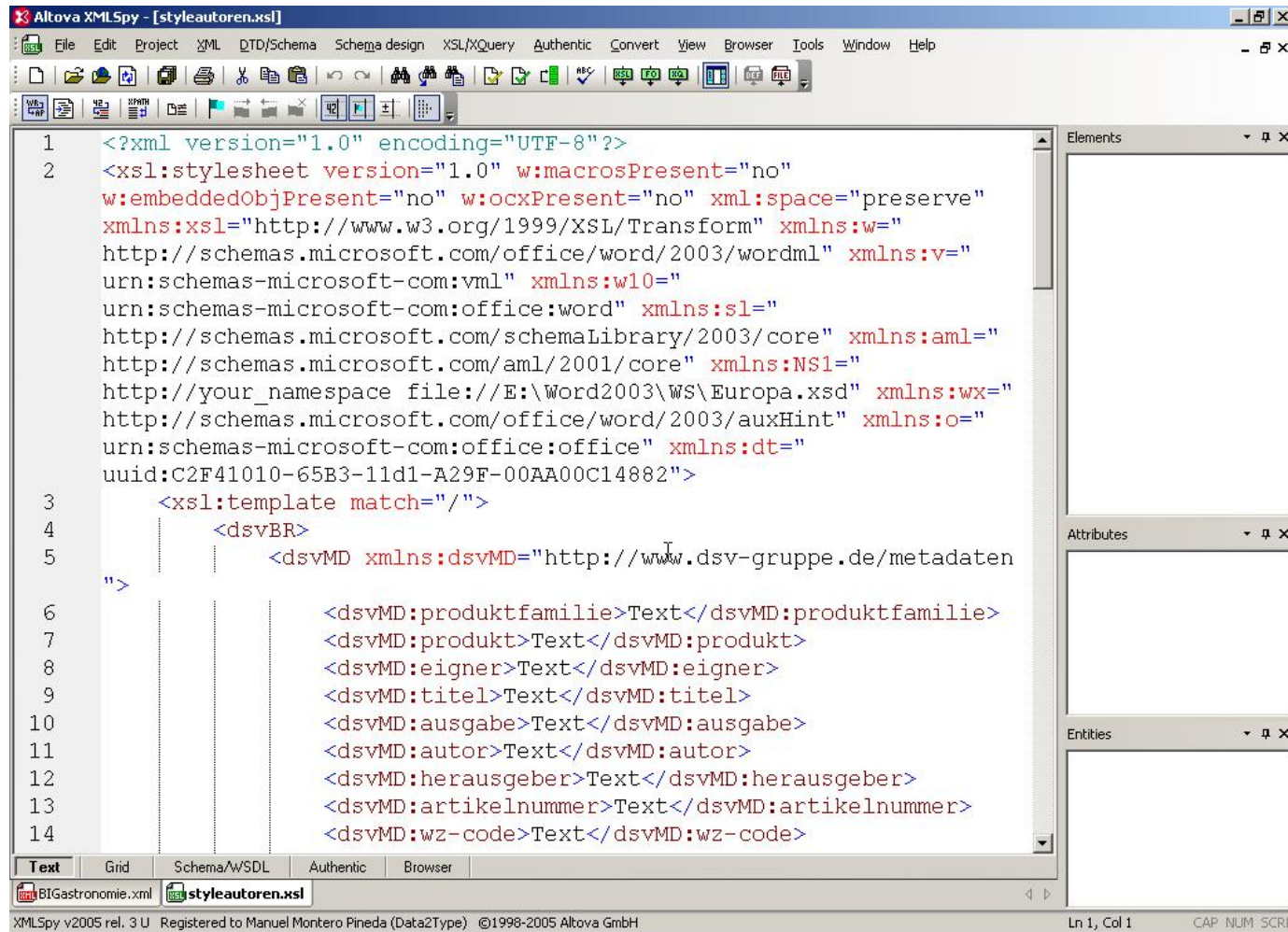




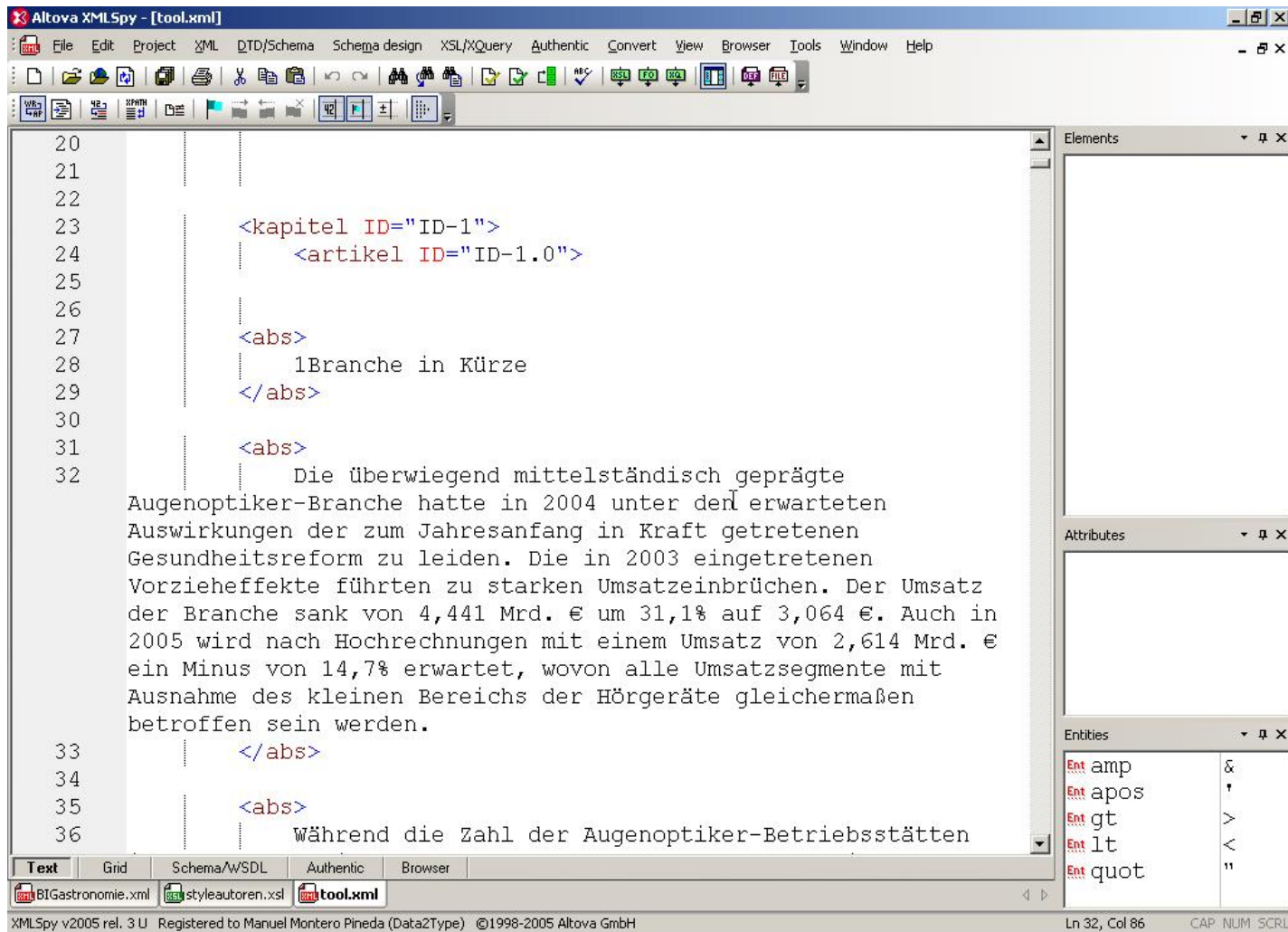
Aufbau von DOCX-Dateien:







```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet version="1.0" w:macrosPresent="no"
  w:embeddedObjPresent="no" w:ocxPresent="no" xml:space="preserve"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform" xmlns:w="
http://schemas.microsoft.com/office/word/2003/wordml" xmlns:v="
urn:schemas-microsoft-com:vml" xmlns:w10="
urn:schemas-microsoft-com:office:word" xmlns:s1="
http://schemas.microsoft.com/schemaLibrary/2003/core" xmlns:aml="
http://schemas.microsoft.com/aml/2001/core" xmlns:NS1="
http://your_namespace file:///E:/Word2003/WS/Europa.xsd" xmlns:wx="
http://schemas.microsoft.com/office/word/2003/auxHint" xmlns:o="
urn:schemas-microsoft-com:office:office" xmlns:dt="
uuid:C2F41010-65B3-11d1-A29F-00AA00C14882">
3   <xsl:template match="/">
4     <dsvBR>
5       <dsvMD xmlns:dsvMD="http://www.dsv-gruppe.de/metadaten"
6         <dsvMD:produktfamilie>Text</dsvMD:produktfamilie>
7         <dsvMD:produkt>Text</dsvMD:produkt>
8         <dsvMD:eigner>Text</dsvMD:eigner>
9         <dsvMD:titel>Text</dsvMD:titel>
10        <dsvMD:ausgabe>Text</dsvMD:ausgabe>
11        <dsvMD:autor>Text</dsvMD:autor>
12        <dsvMD:herausgeber>Text</dsvMD:herausgeber>
13        <dsvMD:artikelnummer>Text</dsvMD:artikelnummer>
14        <dsvMD:wz-code>Text</dsvMD:wz-code>
```



Dokument-Rahmen:
Daten-gestützter Absender,
Adressat etc.

Städtische Krankenhäuser Krefeld
gemeinnützige GmbH

**Kontext-sensitive
Bedienelemente zur Steuerung
des Dokuments**

Klinikum
Krefeld

Blue-Green Insurance
Herrn
Wolli Müller
Warenweg 15
45612 Erkelenz

Dokument-Körper:
Daten-gestützte Elemente,
editierbar

Hauptstraße 235, 12345 Iserlohn
Imgard Walter, Monschauer Weg 123, 45612 Erkelenz

Medizinische Klinik III
Innere Medizin
Nephrologie / Dialyse
Diabetologie / Stoffwechsel
Rheumatologie
Endokrinologie
Direktor: Prof. Dr. med. D. Bach

Station M1
Oberärztin: Dr. med. G. Schott
Fon: 02151/32-27 31
Fax: 02151/32-27 41
Email: M1_Arzt@Klinikum-Krefeld.de

Aufnahme-Nr. (bitte angeben) 1506648915 Datum 17.02.2005

Entlassungsbefehl für Frau Dr. med. Ferdinanda Buono, *15.06.1936

Sehr geehrter Herr Kollege Müller,

wir berichten über unsere Patientin, **Frau Buono, Ferdinanda**, *15.06.1936, wohnhaft Glauberweg 25, 45612 Krefeld, die sich vom 25.11.2004 – 06.12.2004 und 28.01.2005 – 08.02.2005 in unserer stationären Behandlung befand.

Diagnose: Cholangitis K83.0
Nebendiagnose: Cholangitis K83.0
Terminale Niereninsuffizienz N18.0
Prozedur: Hämodialyse: Intermittierend 8-854.0
Hämodialyse: Intermittierend 8-854.0
Aufnahmegrund: Cholangitis K83.0

Recipients
Select Address
Patient

Firma
Abteilung
Anrede - Titel FR
Vor- Nachname Terzioru Jonica
Funktion
Straße 1 Korschbroicher Str. 45
Straße 2
PLZ - Ort 41065 MAArichtengladbac

INS: AOK Rheinland
REL: Ehemann gl. Adr. Tel. 482358-0216
HA: Reinhard, Maaß, F.A.
PAT: Jonica, Terzioru

Detach XML Expansion Pack

Sec 1 1/1 At 12.2cm Ln 29 Col 13 REC TRK EXT OVR German (Ge)

Städtische Krankenhäuser Krefeld
gemeinnützige GmbH

Klinikum Krefeld

Head

Letter (Sender Clinic)
ClinicName Medizinische Klinik

Address (Company Blue-Green Insurance Company*)

Person (Salutation Herr)

FirstName Willi LastName Müller Person*

Street Warenweg 15

ZIP 45612 City Erkelenz Address*

OCList Nachrichtlich: OC (Person Title Dr. med)

FirstName Walter LastName Peters

Street Hauptstraße 235 ZIP 12345

City Iserlohn

OC (Person FirstName)

Connection Fax: 02151/32-27

Connection

Connection Email: MI_Arzt@Klinikum-Krefeld.de

Connection

Select required document:

Frame

Body

Apply

Recipients

Select Address

Verwandter

leere Adresse

Patient

Hausarzt

Versicherung

Verwandter

Vor-Nachname Ehemann g.l.A.

Funktion

Straße 1 Korschenbroicher Str. 45

Straße 2

PLZ - Ort 41065 Mönchengladbach

PAT: Jonica, Terzioru

INS: ADK Rheinland

REL: Ehemann g.l.Adr. Tel. 482358-0216

Apply

Detach XML Expansion Pack

Ser 1 1/2 At 5.2cm In 33 Col 2 REF TRK EXT OVR English (UK)

Fragen?

data<2>type

Dammstr. 18

67059 Ludwigshafen

T: ++49-(0)621-68124792

F: ++49-(0)621-68124794

E: montero@data2type.de