XSL Formatting Objects

Sebastian Rahtz
March 2003

Core principles of XSL FO

- Conceptual compatibility with DSSSL
- Compatibility with CSS properties
- Screen properties as well as print
- No compromises on internationalization
- Closely linked to XSLT transformation language

It provides an *abstract formatting language* to describe pages.

The FO output tree

The input is transformed into an output tree consisting of:

- **page masters**, which define named styles of page layout; and
- **page sequences**, which reference a named page layout and contain a **flow** of text. Within that **flow**, text is assigned to one of five (rectangular) **regions** (the page body, areas at the top, bottom, left and right)

We also have allowance for floating objects (at the top of the page), and footnotes (at the bottom), and the model covers writing in left/right and/or top/bottom modes.

Inside regions and blocks

Within a region of text, we find one or more

- **blocks**
- **tables**
- **lists** and
- **floats**

and within a block, we find

- **inline sequences**
- **characters**
- **links**
- **footnotes** and
- **graphics**

Object properties

- aural properties
- borders, spacing and padding
- breaking
- colors
- font properties (family, size, shape, weight etc)
- hyphenation
- positioning
- special table properties
- special list properties

BUT supporting all of them is not mandatory.

Simple example

Take this bit of TEI input

```
<p>The <gi>corr</gi> element marks
<corr sic="a mistake">correction</corr></p>
```

and transform it, as an interpretation of what the `<corr>` element is supposed to do, with the following XSLT template.

Sample XSLT template

```
<xsl:template match="corr">
 <xsl:text>[</xsl:text><xsl:apply-templates/>
 <xsl:text>]</xsl:text>
 <fo:footnote>
   <fo:footnote-citation>
    <fo:inline font-size="8pt" vertical-align="super">
      <xsl:number format="a" level="any" count="corr"/>
    </fo:inline>
   </fo:footnote-citation>
  <fo:footnote-body>
   <fo:block>
    <fo:inline font-size="8pt" vertical-align="super">
      <xsl:number format="a" level="any" count="corr"/>
    </fo:inline>
    <fo:inline font-family="Helvetica" font-size="10pt">
         <xsl:value-of select="@sic"/>
    </fo:inline>
   </fo:block>
  </fo:footnote-body>
 </fo:footnote>
</xsl:template>
```

Sample FO output

```
<fo:block font-size="12pt" text-align="justify"
  text-indent="1em" space-before="0pt">
The <fo:inline color="green" font-family="Courier">corr
</fo:inline> element marks [correction]
<fo:footnote>
<fo:inline font-size="8pt"
  vertical-align="super">a</fo:inline>
<fo:footnote-body>
<fo:block>
<fo:inline font-size="8pt"
  vertical-align="super">a</fo:inline>
<fo:inline font-family="Helvetica"
  font-size="10pt">a mistake</fo:inline>
</fo:block>
</fo:footnote-body>
</fo:footnote></fo:block>
```

Sample rendition

The `corr` element marks [correction] [a]

---

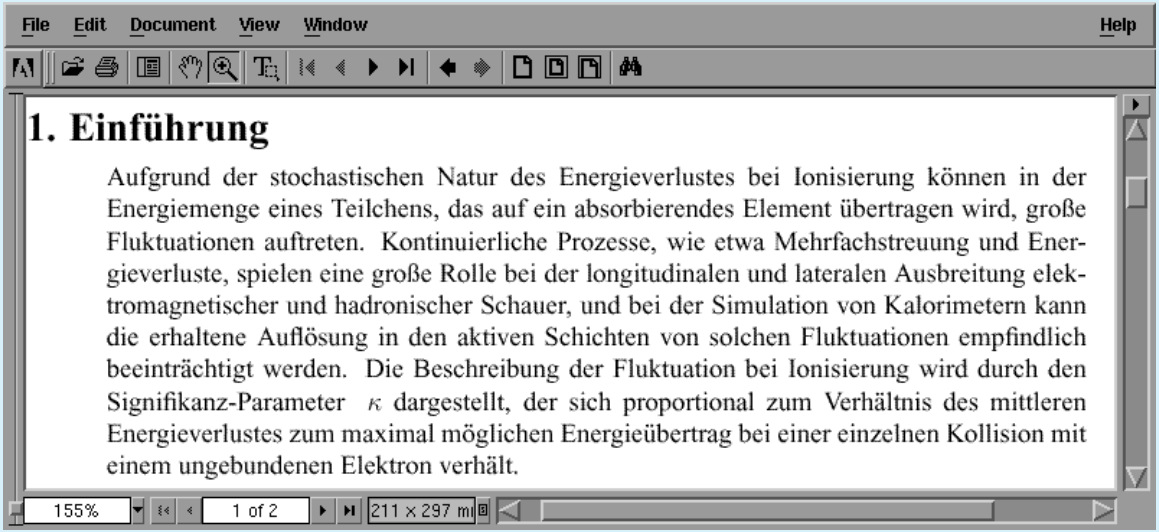
[a] a mistake

FO implementations

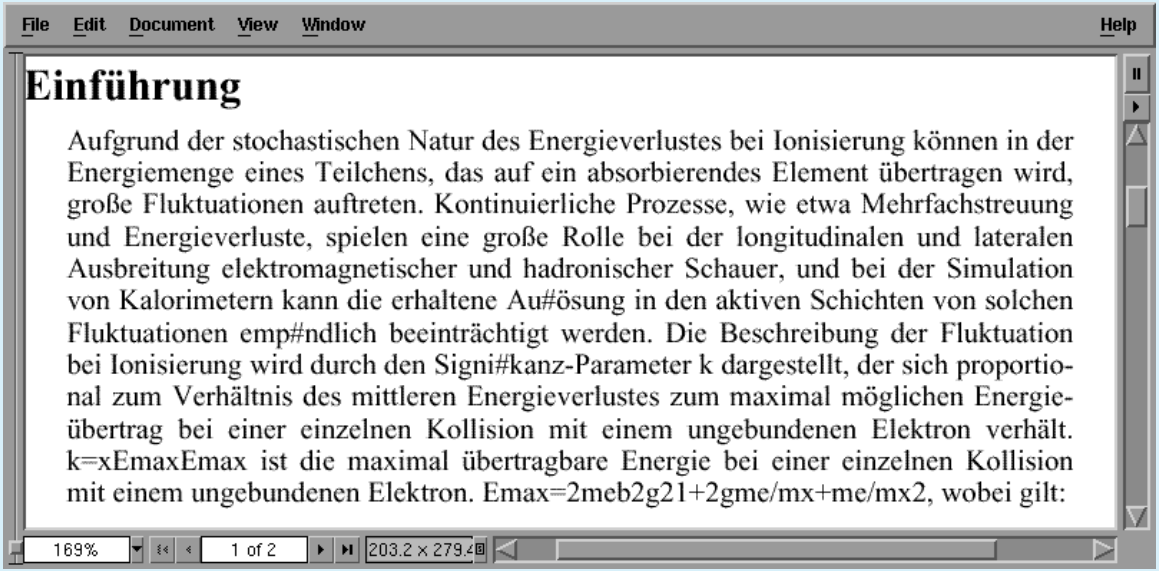
There are six XSLFO implementations to consider:

1. RenderX's *XEP*, probably the most complete and reliable
2. Apache's *FOP* open source Java, creating PDF or AWT, with a large amount of the XSLFO specification implemented.
3. Antenna House's *XSL Formatter* is different in that it formats for the screen, using Windows GDI calls. It has good coverage of the specification (albeit not complete), and is robust.
4. Sebastian Rahtz' *PassiveTeX*, a library of TeX macros which interpret XSLFO;
5. Sun's xmlroff, which implements only a subset of XSL FO so far, but is under active open-source development
6. Arbortext's *Epic* is not yet in widespread use and its coverage of the specification is unknown

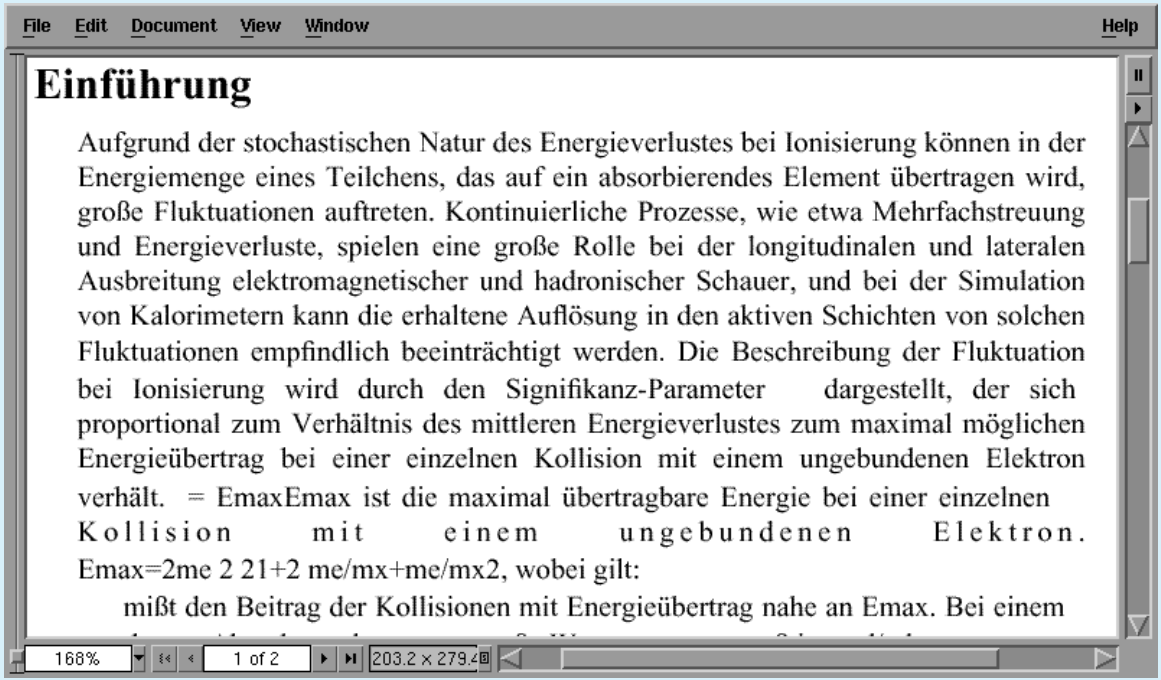
Sample 1: PassiveTeX



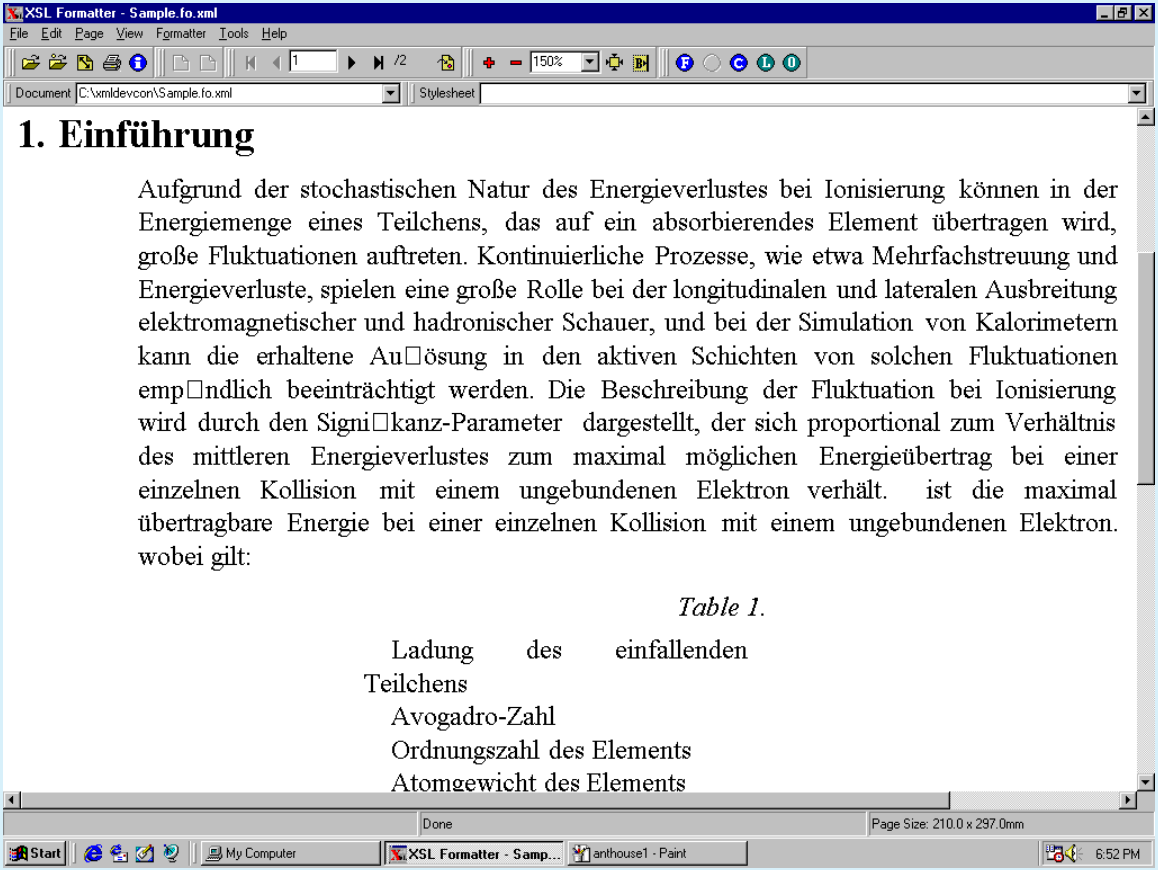
Sample 1: FOP


File Edit Document View Window Help
Einführung
Aufgrund der stochastischen Natur des Energieverlustes bei Ionisierung können in der Energiemenge eines Teilchens, das auf ein absorbierendes Element übertragen wird, große Fluktuationen auftreten. Kontinuierliche Prozesse, wie etwa Mehrfachstreuung und Energieverluste, spielen eine große Rolle bei der longitudinalen und lateralen Ausbreitung elektromagnetischer und hadronischer Schauer, und bei der Simulation von Kalorimetern kann die erhaltene Au#ösung in den aktiven Schichten von solchen Fluktuationen emp#ndlich beeinträchtigt werden. Die Beschreibung der Fluktuation bei Ionisierung wird durch den Signi#kanz-Parameter k dargestellt, der sich proportional zum Verhältnis des mittleren Energieverlustes zum maximal möglichen Energieübertrag bei einer einzelnen Kollision mit einem ungebundenen Elektron verhält. k=xEmaxEmax ist die maximal übertragbare Energie bei einer einzelnen Kollision mit einem ungebundenen Elektron. Emax=2meb2g21+2gme/mx+me/mx2, wobei gilt:
169%
1 of 2
203.2 x 279.4

Sample 1: XEP


File Edit Document View Window Help
Einführung
Aufgrund der stochastischen Natur des Energieverlustes bei Ionisierung können in der Energiemenge eines Teilchens, das auf ein absorbierendes Element übertragen wird, große Fluktuationen auftreten. Kontinuierliche Prozesse, wie etwa Mehrfachstreuung und Energieverluste, spielen eine große Rolle bei der longitudinalen und lateralen Ausbreitung elektromagnetischer und hadronischer Schauer, und bei der Simulation von Kalorimetern kann die erhaltene Auflösung in den aktiven Schichten von solchen Fluktuationen empfindlich beeinträchtigt werden. Die Beschreibung der Fluktuation bei Ionisierung wird durch den Signifikanz-Parameter dargestellt, der sich proportional zum Verhältnis des mittleren Energieverlustes zum maximal möglichen Energieübertrag bei einer einzelnen Kollision mit einem ungebundenen Elektron verhält. = EmaxEmax ist die maximal übertragbare Energie bei einer einzelnen Kollision mit einem ungebundenen Elektron. Emax=2me 2 21+2 me/mx+me/mx2, wobei gilt:
mißt den Beitrag der Kollisionen mit Energieübertrag nahe an Emax. Bei einem
168%
1 of 2
203.2 x 279.4

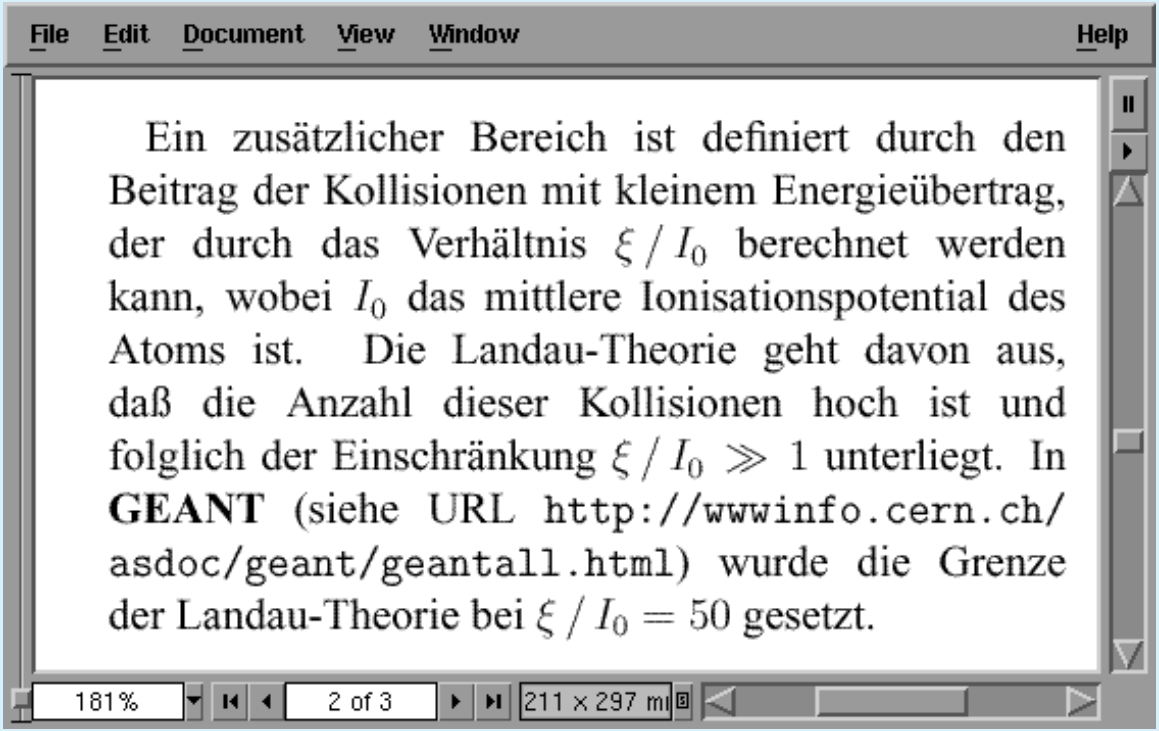
Sample 1: Antenna House



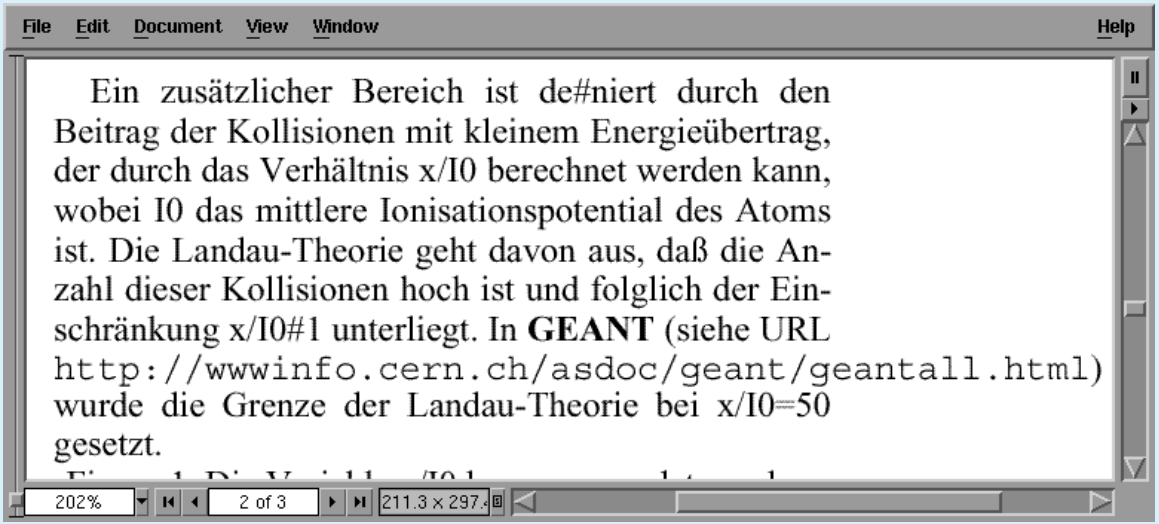
Discussion 1

The results are within fairly acceptable limits, though XEP's inability to hyphenate shows clearly how vital a good H&J engine is - the word and letter spacing on the 3rd line from the end is clearly unacceptable. Let us now decrease the page width, and look at a harder paragraph.

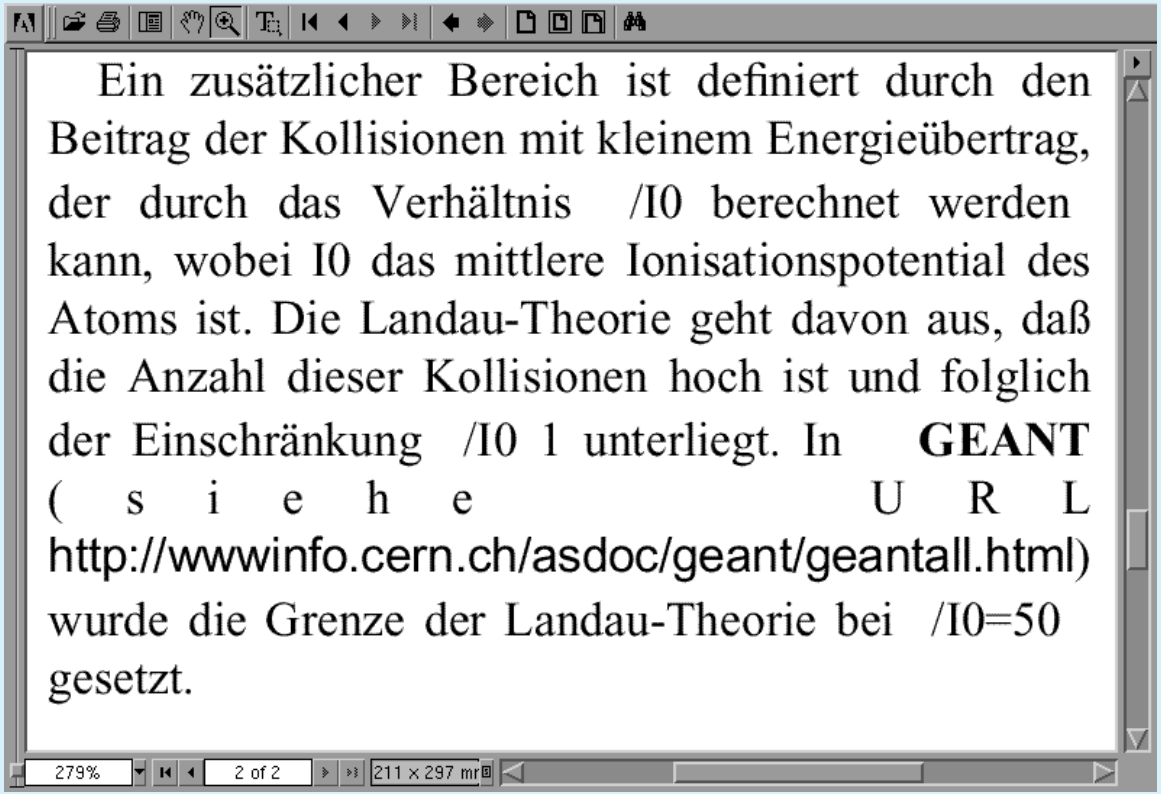
Sample 2: PassiveTeX



Sample 2: FOP


File Edit Document View Window Help
Ein zusätzlicher Bereich ist de#niert durch den Beitrag der Kollisionen mit kleinem Energieübertrag, der durch das Verhältnis x/I0 berechnet werden kann, wobei I0 das mittlere Ionisationspotential des Atoms ist. Die Landau-Theorie geht davon aus, daß die Anzahl dieser Kollisionen hoch ist und folglich der Einschränkung x/I0#1 unterliegt. In GEANT (siehe URL http://wwwinfo.cern.ch/asdoc/geant/geantall.html) wurde die Grenze der Landau-Theorie bei x/I0=50 gesetzt.
202%
2 of 3
211.3 x 297.

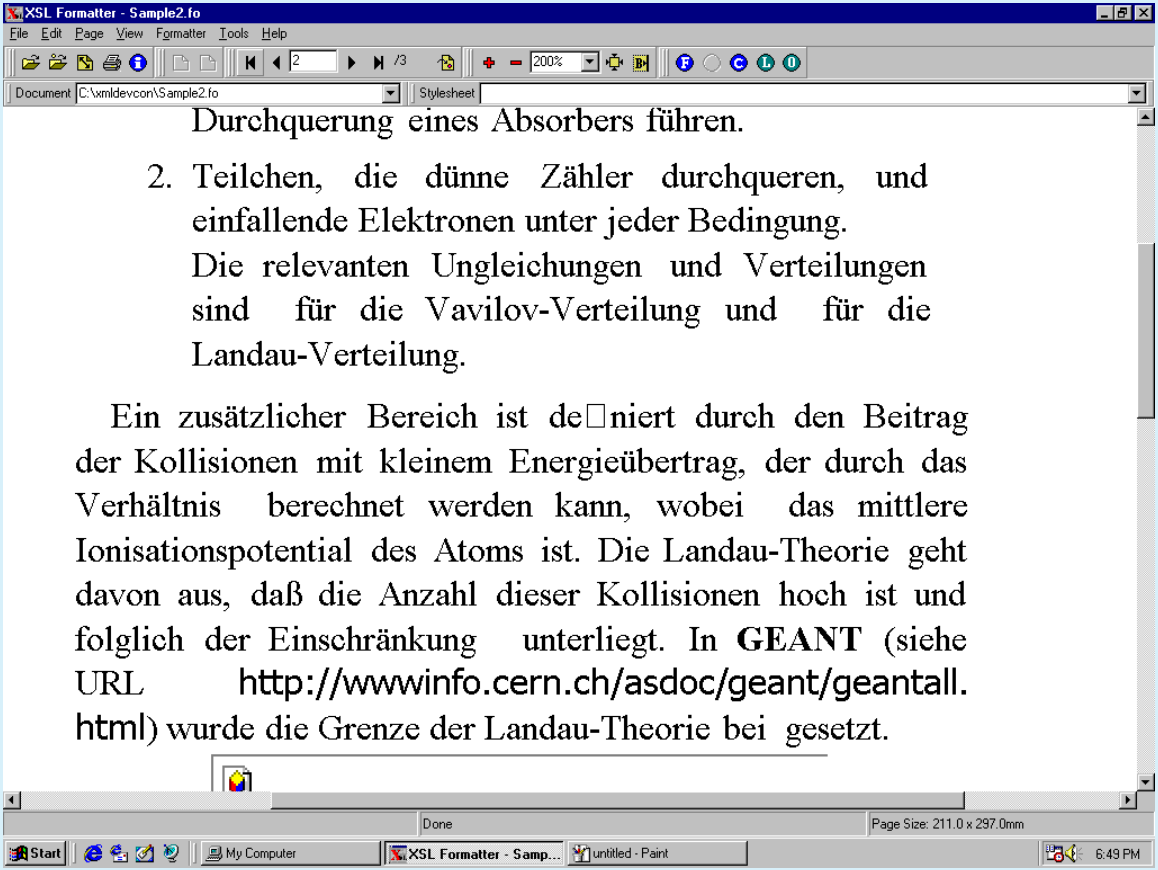
Sample 2: XEP



Ein zusätzlicher Bereich ist definiert durch den Beitrag der Kollisionen mit kleinem Energieübertrag, der durch das Verhältnis /I0 berechnet werden kann, wobei I0 das mittlere Ionisationspotential des Atoms ist. Die Landau-Theorie geht davon aus, daß die Anzahl dieser Kollisionen hoch ist und folglich der Einschränkung /I0 1 unterliegt. In **GEANT** (s i e h e U R L http://wwwinfo.cern.ch/asdoc/geant/geantall.html) wurde die Grenze der Landau-Theorie bei /I0=50 gesetzt.

279% 2 of 2 211 x 297 mr

Sample 2: Antenna House



Discussion 2

The results show two rather different formatting decisions taking place in FOP and XEP. The former opts to abandon sensible line breaking, and let the URL stick out into the margin, but XEP forces itself to extreme lengths to justify with letter spacing, while Antenna House makes an arbitrary break.