# Testdokument für XML-Konvertierung

Das folgende Dokument basiert auf der normal.dot, es wird nur in allen Formaten die Arial Unicode verwendet. Es wurden ein paar Absatzformate hinzugefügt um die Übernahme von Absatzformat-Namen zu demonstrieren.

Um das Dokument übersichtlich zu halten, werden die möglichen Inhalte mit einer kurzen Erläuterung aufgeführt. Bestimmte Eigenschaften des Textes werden grundsätzlich nicht ausgewertet. Hierzu gehört vor allem die Schriftart, da eine zuverlässige Konvertierung des Textes nur möglich ist, wenn eine einzige Schriftart, die möglichst viele Zeichen enthält, verwendet wird. Daher sind alle Absatzformate grundsätzlich in der Arial Unicode gehalten. Ebenso wenig wird der aktuelle Schriftgrad ausgewertet, da sich die Gestaltung erst später aus der Datenstruktur ergibt.

# Überschriftebenen

Der folgende Abschnitt enthält alle konvertierbaren Überschriftebenen.

## Überschrift Ebene 2

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

### Überschrift Ebene 3

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

#### Überschrift Ebene 4

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

##### Überschrift Ebene 5

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

###### Überschrift Ebene 6

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

Überschrift Ebene 7

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

Überschrift Ebene 8

In den folgenden Abschnitten befinden sich die restlichen Überschriftebenen.

Überschrift Ebene 9

Dies ist die unterste Überschriftebene. Mehr als 9 Ebenen sind in Word nicht möglich.

# Tabellen

Bei Tabellen werden primär die Struktur (Reihen, Zellen und Überspannen von Zellen) und die Hintergrundfarbe übernommen. Linien werden nicht übernommen.

**Tabelle 1**: Bevölkerungsentwicklung Göttingens 1740-1795

|  |  |  |  |
| --- | --- | --- | --- |
| **Jahr** | **Einwohner (ohne Stud.)** | **Studenten** | **Professionisten** |
| 1740 | 5562 | 313 | 670 |
| 1756 | 6900 | 600 | 698 |
| 1766 | 6306 | 695 | 702 |
| 1775 | 8107 | 840 | 804 |
| 1785 | 8674 | 838 | 826 |
| 1795 | 9078 | 674 | 833 |

Die Hintergrundfarbe von Zellen kann dazu verwendet werden, Tabellenkopf und -fuß zu kennzeichnen, falls dies benötigt wird. Dies muss aber individuell eingerichtet werden.

# Inline Elemente

Der folgende Abschnitt behandelt alle Inline-Elemente, die zurzeit übernommen werden. Dies kann theoretisch noch beliebig ausgeweitet werden.

## Zeichenformate

Die folgenden Absätze enthalten alle Zeichenformate, die bei der Konvertierung berücksichtigt werden. Andere Zeichenformate sind zwar nicht verboten, werden aber bei der Konvertierung ignoriert.

Mustertext mit Texten in **fett**, in *kursiv*, in ***fettkursiv*** und unterstrichen. Bei Unterstreichungen wird die Gestaltung der Linie nicht übernommen. Hoch- und Tiefstellung: 17 mg CO2 / m3.

Spezielle Auszeichnungen Kapitälchen, Text mit farbiger Hinterlegung, durchgestrichener ~~Text~~ und farbiger Text. Bei der Hinterlegung wird nur übernommen, dass der Text hinterlegt ist, die Textfarbe wird generell nicht übernommen.

Separat definierte Zeichenformate wie spezial werden ebenfalls übernommen, es kann hier aber zu nicht eindeutigen Situationen kommen, wenn die Zeichenformate lediglich Eigenschaften enthalten, die auch mittels einer Word-Funktion zugewiesen werden können.

## Formeln

Mathematische Formeln können mit dem integrierten Formeleditor von Word 2007 und höher erstellt und als MathML übernommen werden[[1]](#footnote-1). Formeln, die mit der Vorgängerversion des Word-Formeleditors erstellt wurden, können nicht übernommen werden.

Es ist empfehlenswert, das Absatzformat für Formeln zu verwenden, wenn abgesetzte Formeln verwendet werden.

Es gibt auch die Möglichkeit Inline-Formeln zu verwenden: unschön besser Hierbei sollten aber nur einfache Formelausdrücke verwendet werden, weil sonst die Lesbarkeit leidet.

## Fußnoten

Fußnoten, die mit der Standard-Funktionalität von Word[[2]](#footnote-2) erzeugt werden, können in XML übernommen werden. Endnoten werden im Moment nicht unterstützt, es ist aber kein Problem, diese Funktionalität hinzuzufügen.

## Querverweise

Word unterstützt zahlreiche Varianten von Querverweisen. Die wichtigste basiert auf automatisch nummerierten Elementen. Da in diesem Dokument die Überschriften nicht nummeriert sind, gibt es diese Funktionalität nicht. Ein Verweis kann daher nur den Überschrifttext enthalten. Dies sind Verweise mit Überschrifttext auf Überschriftebenen und auf Tabellen. Im Prinzip kann auch die Seitenzahl angegeben werden, so wie hier 2, dies ist aber nicht sehr praktikabel.

Darüber hinaus können Textmarken verwendet werden. Hierzu muss zunächst am Sprungziel eine Textmarke eingefügt werden, auf die dann verwiesen werden kann. Hier ein Verweis auf Überschriftebenen und auf Tabellen.

## Indexeinträge

Indexeinträge können mit der Word-Funktion für Indexeinträge eingefügt werden. Die zusätzlichen Funktionen "siehe" , fette Seitenzahlen und kursive Seitenzahlen werden ebenfalls übernommen..

# Listen

Der folgende Abschnitt enthält die wichtigsten Listentypen. **Achtung**: die Zählerstände von Listen werden nicht übernommen, d.h. eine neue Liste beginnt immer mit 1, unabhängig davon, was in Word angezeigt wird. Die Listen-Nummerierung von Word ist immer für Überraschungen gut. Es ist also durchaus denkbar, dass die Konvertierung von Listen nicht immer zum gewünschten Ergebnis führt. Das ist aber kein Problem des Konverters.

Ab hier ungezählte Listen:

* Erster Listenpunkt Ebene 1
* Zweiter Listenpunkt Ebene 1
  + Erster Listenpunkt Ebene 2
  + Zweiter Listenpunkt Ebene 2
    - Erster Listenpunkt Ebene 3
    - Zweiter Listenpunkt Ebene 3
    - Dritter Listenpunkt Ebene 3
  + Dritter Listenpunkt Ebene 2
* Dritter Listenpunkt Ebene 1
* Vierter Listenpunkt Ebene 1

Ab hier gezählte Listen:

1. Erster Listenpunkt Ebene 1
2. Zweiter Listenpunkt Ebene 1
   1. Erster Listenpunkt Ebene 2
   2. Zweiter Listenpunkt Ebene 2
      1. Erster Listenpunkt Ebene 3
      2. Zweiter Listenpunkt Ebene 3
      3. Dritter Listenpunkt Ebene 3
   3. Dritter Listenpunkt Ebene 2
   4. Vierter Listenpunkt Ebene 2
3. Dritter Listenpunkt Ebene 1
4. Vierter Listenpunkt Ebene 1

Zurzeit werden Listen, die bis maximal 3 Ebenen tief verschachtelt sind, unterstützt.

# Abbildungen

Theoretisch können die Abbildungen aus dem Word-Dokument übernommen werden. Für Print-Produkte reicht die Qualität der Word-internen Jpeg-Dateien in der Regel aber nicht aus. Daher wird dies in dieser Version des Konverters nicht gemacht. Grundsätzlich ist es aber möglich, die Word-internen Bilddateien auszulesen.

Die einfachste Möglichkeit zur Übernahme von Bildern in XML besteht darin, mit speziellen Absatzformaten, den Dateinamen und die Bildlegende zu markieren und dies in der individuellen Konvertierung entsprechend umzusetzen.

1. Dieses Format ist zwar für XML der Standard, wird aber von den wichtigsten Layoutprogrammen nur bedingt unterstützt. [↑](#footnote-ref-1)
2. Diese ist unter Verweise -> Fußnote einfügen zu finden. [↑](#footnote-ref-2)