Bericht über den zweiten TEI-Workshop am ZDV vom 11.–13.3.96

Von 11.–13. März wurde am ZDV der Workshop zum Thema *SGML-konforme Textauszeichnung nach den Richtlinien der Text Encoding Initiative (TEI)* wiederholt, der im November 1995 erstmals abgehalten wurde. Nachdem der November-Workshop auch in einem Artikel der Computer Woche (CW 49 vom 8.12.95, S. 26) ein Echo gefunden hatte, kamen im März erneut 42 Teilnehmer/innen aus ganz Europa (Deutschland, Norwegen, Österreich, Polen, Spanien) zusammen. Die Mischung aus kommerziellen und wissenschaftlichen Anwender/innen ergab sich gegenseitig befruchtende Diskussionen: Die Projekte der Teilnehmer/innen reichten von der textkritischen Edition von Autographen bis hin zur Wörterbucharbeit, von dem Aufbau wissenschaftlicher Textarchive bis hin zur buchidentischen CD-ROM-Edition, von der wissenschaftlichen Erarbeitung ägyptischer, hebräischer, baskischer, englischer und deutscher Texte bis hin zu den Arbeitsgängen im Verlagsablauf. Geleitet wurde der Workshop von Michael Sperberg-McQueen PhD (University of Illinois at Chicago, USA) und Dr. Winfried Bader (ZDV der Universität Tübingen). Der Workshop umfaßte im wesentlichen die gleichen Themen wie im November (vgl. dazu den Bericht in BI 95/11+12, S. 11–12). Sie wurden um folgende Punkte erweitert.

Textauszeichnung als Interpretation

Die Standard Generalized Markup Language (SGML) – und damit die *Document Type Definitions* (DTD) der TEI als Anwendung dieses Standards – ist eine deklarative Auszeichnungssprache. Gegenüber einem prozeduralen Auszeichnungssystem bringt das den Vorteil der Unabhängigkeit von Hardware und Software. Das erleichtert den Datenaustausch, was eines der Ziele der TEI ist. Die wichtige Frage lautet aber: Welche Informationen sollen die Daten für den Austausch enthalten?

Zum einen kann die typographische Information codiert, also eine Seitenbeschreibung vorgenommen werden. Der Text wird aufgefaßt als eine Abfolge von einzelnen Seiten. SGML ist zu einer typographischen Textbeschreibung fähig; besser geeignet ist eine Faksimile-Wiedergabe im PostScript-Format. Doch wird damit nur ein kleiner Aspekt eines Textes berücksichtigt, seine »objektive« Erscheinung in konkreten körperlichen Seiten.

Ein Text im eigentlichen Sinn (*textus* – Gewebe) ist ein abstraktes Gebilde mit verschachtelten Strukturen und vielschichtigen Informationen. Textauszeichnung in diesem Sinne ist ein analytischer Vorgang. Es ist das Einbringen von zusätzlichen Informationen, die die Struktur, den Zustand und die Eigenschaften eines Textes beschreiben. SGML hat seine Stärken in der Unterstützung analytischer Textbeschreibung. Gegenüber der rein typographischen Seitenbeschreibung stellt diese einen Mehrwert dar. Der so codierte Text enthält qualitativ neue Information, ist angereichert mit »Intelligenz«, Hintergrund- und Verwaltungsinformationen, auf die mit Abfrageprogrammen zugegriffen werden kann. Die Daten sind damit allerdings auch »subjektiv«, denn diese Informationen hängen vom Codierenden ab.

In der gegenüber dem November-Workshop erweiterten Einführung in die Grundzüge der Richtlinien der TEI konnte dieses Nebeneinander von typographischer und analytischer Textauszeichnung herausgearbeitet werden. Die Richtlinien der TEI unterstützen die analytische Auszeichnung gerade auch dadurch, daß Mechanismen (der TEI-Header) bereitgestellt werden, die helfen, die »subjektiven« Bedingungen und Entscheidungen offenzulegen und nachvollziehbar zu machen.

Ein Beispielsatz:

Es gibt nur *eine* Antwort:
die *TEI-Guidelines*

Typographisch unterscheidet sich kursive von aufrechter Schrift, was gemäß der TEI als Hervorhebung ausgezeichnet werden kann. Betrachtet man den Satz nicht typographisch sondern analytisch, so haben die beiden kursiven Wörter allerdings sehr unterschiedliche Bedeutung: *eine* ist betont, genauer ist es die Unterscheidung von unbestimmtem Artikel und Zahlwort; *TEI-Guidelines* ist eine Art von Zitat, das als Buchtitel, Fremdwort oder Terminus technicus näher bestimmt werden kann.

Auszeichnung von Wörterbüchern

Die Basisgruppe der TEI-Auszeichnungen für Wörterbücher dient dazu, die Struktur der Artikel von ein- und mehrsprachigen gedruckten Wörterbüchern explizit zu machen, um die darin enthaltenen Informationen im maschinellen Zugriff für Weiterverarbeitungs- und Suchprogramme zu haben. Stichwort, grammatische Angabe, Nebenformen, Orthographie, Aus-

sprache, Silbentrennung, Definition, Übersetzung, Sprachgebrauch, Beispiele, Literaturangabe, Zeitangabe, Etymologie, Homographen etc. sind Teile, die im gedruckten Wörterbuch oft nur ein verstehender Leser erkennen kann, die aber als die für die Struktur konstitutiven Teile bei der elektronischen Textauszeichnung explizit kenntlich gemacht werden müssen.

Software

a) Formatter

Bei der Demonstration der Software konnte im Bereich der Formatter eine neue Möglichkeit des TUSTEP-SATZ-Programms gezeigt werden. Das TUSTEP-SATZ-Programm ist nun in der Lage, SGML-Auszeichnungen in Form von Style-Sheets zu verarbeiten. Das Besondere und Neue daran ist, daß der Text mit der Umbruchinformation und den originalen SGML-Auszeichnungen zusammen in einer Datei abgelegt wird. Das ist die Voraussetzung für die Erstellung buchidentisch zitierfähiger elektronischer Publikationen.

b) Browser

PANORAMA FREE von der Firma SoftQuad steht unter Microsoft WINDOWS als freier Browser fürs INTERNET zur Verfügung. Damit ist ein Schritt getan, neben dem wenig anspruchsvollen HTML eine wissenschaftlich verantwortete Textauszeichnung im WWW zu präsentieren. PANORAMA fürs WWW findet sich zusammen mit dafür geeigneten Beispieltexen, die nach den Richtlinien der TEI aufbereitet sind, unter dem URL:

<http://www.ceth.rutgers.edu/projects/hercproj/front.htm>

Informationsquellen

Neben den in der BI 95/11+12 genannten Informationsquellen ist die kommentierte Bibliographie zu SGML von Robin COVER zu nennen unter dem URL:

<http://www.sil.org/sgml/biblio.html>

Die im Workshop verwendeten Unterrichtsmaterialien stehen als PostScript-Files auf dem Rechner

x400link.zdv.uni-tuebingen.de

im Verzeichnis

scratch/teislides

über *anonymous ftp* zur Verfügung.

Die Files enthalten die Materialien zu den folgenden Themen (die Zahlen im Dateinamen geben den Seitenumfang an):

doc1_13.ps Dokumentenanalyse
sgml1_32.ps Einführung in SGML
teil_30.ps Einführung in TEI
int1_30.ps Fortgeschrittenes SGML und TEI-Interna
link1_26.ps Querverweise und Parallelisierung, Hypertext
anal_12.ps Einfache Analysemechanismen
fs1_11.ps Fortgeschrittene Analysemechanismen, Merkmalstrukturen
crit1_12.ps Textkritische Apparate
wb1_33.ps Wörterbücher
wsd1_16.ps Schreibsystemvereinbarungen
who1_8.ps Laufende TEI-Projekte
bi.ps BI-Artikel 95/11+12, S. 11–12
teelite.ps TEI-Lite, Eine Einführung in TEI (66 Seiten, englisch)

Winfried Bader

bader@zdv.uni-tuebingen.de