

BERICHT ÜBER DIE TEI-WORKSHOPS AM ZENTRUM FÜR DATENVERARBEITUNG DER UNIVERSITÄT TÜBINGEN

Winfried Bader

Vom 15.-17. November fand am ZDV der Universität Tübingen ein Workshop zum Thema *SGML-konformen Textauszeichnung nach den Richtlinien der Text Encoding Initiative (TEI)* statt, der infolge des großen Andrangs vom 11.-13. März 1996 wiederholt wurde. Insgesamt 80 Teilnehmer/innen aus ganz Europa, die elektronische Texte einsetzen für kritische Edition von Einzeltexten und Textsammlungen, linguistische Analyse, erzählanalytische Textbeschreibung, Wörterbücher und Lexika, Bibliothekswesen, Sammlung von Einzelartikeln, Bereitstellung von Lehrmitteln, besuchten die beiden Veranstaltungen.

Sie wurde gehalten von Michael Sperberg-McQueen PhD (Chicago, IL, USA), der als leitender Herausgeber die TEI-Richtlinien maßgeblich mitgestaltet hat, und Dr. Winfried Bader (ZDV), Mitglied des Advisory Boards der TEI. Das Thema SGML erweckt erst seit der weiten Verbreitung von HTML im WWW allgemeines Interesse; die ISO-Norm selbst existiert aber bereits seit 1986. Auch die Problemstellung, SGML für den Austausch und die elektronische Publikation geisteswissenschaftlicher Texte und ihrer wissenschaftlichen Analyse zu verwenden, beginnt jetzt bei einer größeren Öffentlichkeit Interesse zu finden.

Der folgende Bericht über die Themen der Workshops gibt gleichzeitig einen ersten Überblick über die wichtigsten Ziele und Eigenschaften der TEI:

1 Dokumentenanalyse

Bei diesem Schritt ging es darum, die Phasen der Arbeit mit elektronischen Texten zu zeigen. Wichtig ist, vor jeglicher elektronischer

Arbeit sich klar zu machen, daß Textauszeichnung heißt, eine oder mehrere Interpretationen eines Textes explizit zu machen. Die Dokumentenanalyse klärt die inhaltlichen und wissenschaftlichen Ziele, aber auch die Rahmenbedingungen (Geld, Zeit, Ziel), für die Erfassung elektronischer Texte. Als Beispiele dienten eine Akkadische Rationenliste, Autographen eines modernen Autors, ein- und mehrsprachige Wörterbücher, Enzyklopädien, Gedichtanalysen.

2 Einführung in SGML

Die ISO-Norm 8879:1986 Standard Generalized Markup Language ist eine Metasprache für Textauszeichnungssprachen. Ihre Syntax besteht aus Element-, Attribut- und Entity-Vereinbarungen, die konstitutive Bestandteile einer Document Type Definition (DTD) sind. In einer Demonstration wurde das Erstellen einer DTD und eines dazugehörigen Dokumentes gezeigt, das anschließend mit dem Parser sgmls auf SGML-Konformität überprüft wurde.

3 Einführung in TEI

Die Text Encoding Initiative hat Richtlinien entwickelt, in der die Problematik der Auszeichnung geisteswissenschaftlicher Texte aus allen Fachrichtungen berücksichtigt ist, und die gleichzeitig SGML-konforme DTDs bereitstellt, die in jeder SGML-Software verwendet werden können. Eine Übersicht über das Grundinventar der TEI-Auszeichnungen

wurde gegeben und die modulare Struktur der TEI-DTDs samt ihrer SGML-technischen Details erklärt. Auf einer Kerngruppe von Auszeichnungen, die bei allen Textsorten gebraucht werden, bauen Basisgruppen auf, die jeweils für einen bestimmten Texttyp (Prosa, Poesie, Drama, Wörterbuch, ...) Auszeichnungen bereitstellen. Frei kombinierbar kommen die Zusatzgruppen hinzu, die jeweils einer bestimmten Art der Textuntersuchung (Verkettung und Verweise, Analyse, Textkritik, ...) dienen. Mit dem Author/Editor, der dank der Erlaubnis der Firma SoftQuad auch für eigenes Üben im PC-Pool zur Verfügung stand, wurde der praktische Einsatz der TEI-DTD demonstriert.

4 TEI-Auszeichnungen für "Verkettung und Parallelisierung"

Diese Zusatzgruppe der TEI-Auszeichnungen stellt zahlreiche Mechanismen für Querverweise, zur Verkettung von Textstellen, für die Verweismechanismen zum Anbinden von Analysen, zur Parallelisierung verschiedener Dokumente und für Hyperlinks bereit. Sie sind flexibler und umfassender als die Verweismechanismen von HTML (Hyper Text Markup Language).

5 TEI-Auszeichnungen für "Einfache Analyse Mechanismen"

Die Auszeichnungen dieser Zusatzgruppe dienen vor allem Analysen aus dem linguistischen Bereich. Außerdem bietet diese Gruppe interessante Möglichkeiten, Einzelanalysen zu bestimmten Textstellen über Verweismechanismen an getrennter Stelle im Dokument zu verwalten, wodurch eine übersichtlichere Anordnung ermöglicht wird. Es erlaubt auch, schon vorhandene Dokumente nachträglich mit Analysen zu versehen, ohne das Dokument selbst verändern zu müssen.

6 SGML-Software in der Praxis

a) Editoren:

SGML-Editoren sind in der Lage, eine beliebige DTD zu lesen, um dann bei der Erstellung des Dokumentes kontextsensitiv nur die Auswahl der an dieser Stelle entsprechend der DTD legalen Auszeichnungen zu ermöglichen, und sie bequem per Mausclick einzutragen. Demonstriert wurde psgml, ein Zusatz zum emacs (unter LINUX), Author/Editor und RulesBuilder (zum Kompilieren von DTDs) von SoftQuad (beide unter WINDOWS), und TUSTEP (unter AIX).

b) Parser:

Parser prüfen, ob ein bestimmtes Dokument zu einer gegebenen DTD konform ist. Vorgestellt wurde sgmls (unter LINUX), der neben der Prüfung als zusätzlichen Output eine Umformung des Dokumentes bietet, die die Weiterverarbeitung erleichtert, die Parsing-Funktion von Author/Editor und eine TUSTEP-Kommandofolge zum Prüfen von HTML-Dokumenten.

c) Browser:

Browser dienen dazu, SGML-Dokumente am Bildschirm in ansprechender Form anzuzeigen, das Blättern innerhalb des Dokumentes zu erleichtern, das Verfolgen von Verweisen (Hyperlinks) sowie eine Volltextsuche zu ermöglichen. Vorgestellt wurde Panorama von SoftQuad, dem allerdings die Suchmöglichkeiten fehlen, DynaText und DynaWeb von EBT. Letzteres bietet die Möglichkeit, beliebige SGML-Dokumente im WWW mit HTML zu präsentieren, trotzdem aber im Hintergrund die volle SGML-Struktur zu haben, so daß diese für Suchen und Verweisen genutzt werden kann.

d) Retrieval:

Neben den Browsern gibt es auch spezielle Programme, die der Suche in SGML-Dokumenten bzw. weitergefaßt der analytischen Verarbeitung dieser Dokumente unter Auswertung ihrer Struktur dienen. Vorgestellt wurden Beispiele, die mit TUSTEP realisiert waren.

e) Formatter:

Diese Programme bringen SGML-Dokumente zum Druck, ohne daß typographische Handarbeit notwendig ist. Vorgestellt wurde in dieser Funktion TUSTEP.

7 Schreibsystemvereinbarung

Von den verschiedenen Spezialthemen der TEI wurde die Writing Systems Declaration behandelt. Hier geht es darum, die Verwendung von nicht-lateinischen Schreibsystemen in leicht handhabbarer Transkription eindeutig zu dokumentieren, zu jedem Transkriptionszeichen die Bedeutung, den UNICODE-Character und das entsprechende Glyph aus dem Glyphregister zuzuordnen. Die Schreibsystemvereinbarung der TEI erfolgt in einer Form, die auch von (zukünftiger) Software ausgewertet werden kann.

8 Informationsquellen

Informationen über SGML und TEI erhält man im WWW unter folgenden URLs. Homepage der Text Encoding Initiative:

<http://www-tei.uic.edu/orgs/tei>

Die Richtlinien der TEI im Zugriff mit dem DynaWeb-Browser, der auch eine Volltextsuche erlaubt:

<http://etext.virginia.edu/TEI.html>

Webpage, in der alles Wissenswerte zu SGML gesammelt ist: <http://www.sil.org/sgml/sgml.hun1>
Steve Peppers *Whirlwind Guide to SGMLTools*, der ausführlich über SGML-Software informiert, ist via ftp erhältlich:

<ftp://falch.no/pub/SGML-Tools/sgmltool.Ut>

Eine E-mail-Diskussionsgruppe gibt es unter der Adresse:

tei-l@uicvm.bitnet

In einer der folgenden Hefte des LDV-Forums wird die TEI genauer vorgestellt.

*Winfried Bader
Universität Tübingen Zentrum
für Datenverarbeitung
Brunnenstraße 27
D-72074 Tübingen
bader@zdv.uni-tuebingen.de*