

Werkzeuge zum Erstellen von HTML-Dokumenten für das Tübinger Informationssystem

Das Tübinger Informationssystem und HTML

Texte, die im Tübinger Informationssystem zur Verfügung gestellt werden, müssen nach einem bestimmten System sachlicher Auszeichnungen (siehe zu diesem Begriff auch den BI-Artikel »Offene Texterfassung durch sachbezogene Auszeichnung« in BI 92/11+12, S. 22–23) markiert werden. Erst dadurch sind die beiden Programme *mosaic* und *lynx* (das sind die in Tübingen gebräuchlichen Browser zum Lesen von World Wide Web-Dokumenten) in der Lage, den Text des Dokuments am Bildschirm in ansprechender Form zu zeigen und Querverbindungen zwischen Dokumenten herzustellen. Das verwendete Auszeichnungssystem im WWW (World Wide Web), zu dem das Tübinger Informationssystem gehört (siehe dazu den BI-Artikel »Das Informationssystem der Universität Tübingen am ZDV« in BI 94/1+2, S. 6–8), ist HTML (Hyper Text Markup Language). HTML ist eine konkrete Anwendung der SGML-Richtlinien (Standard Generalized Markup Language; ISO-Norm 8879 von 1986), d. h. HTML besteht aus einer SGML-konformen Document Type Definition, aus der die Richtlinien für die Auszeichnung einzelner Dokumente und die Richtlinien zum Erstellen von Programmen, die mit HTML-Dokumenten umgehen können, entnommen werden können. HTML bietet eine Menge von ca. 70 sachlich orientierten Tags (eindeutige Zeichenfolgen) an, mit denen man ein Dokument auszeichnen kann. Als Grundregel für die Verwendung dieser Tags gilt: sie sind deskriptiv, d. h. man kann mit ihnen die inhaltliche Struktur seines Dokuments beschreiben, unabhängig davon, wie dies nachher von *mosaic* oder *lynx* weiterverarbeitet wird.

Für Informationsanbieter liegen gedruckte Einführungen in HTML im ZDV bereit. Auch im WWW finden sich reichlich Einführungen in die Verwendung von HTML als Auszeichnungssprache für Dokumente (jedoch ist nicht alles SGML-konform, was dort zu lesen ist).

Das Erstellen eines HTML-Dokumentes

Hat man einen ersten Einblick in HTML und seine Regeln (meist genügt die Kenntnis der wichtigsten Tags und das Wissen, daß die Tags grundsätzlich wie Klammern verwendet werden, die den Textbereich, den sie betreffen, umschließen), stellen sich zwei Probleme: zum einen ist das Eintippen der Tags und der Codierungen für die Umlaute (als Entities: z. B. *ä* für *ä*) sehr schreibintensiv und daher fehleranfällig, zum anderen schleichen sich logische Fehler ein, und man wünscht sich eine Konsistenzprüfung des erstellten Dokuments. Beide Probleme löst der unter TUSTEP entwickelte HTML-Editor: Zum einen reduziert er die Schreibarbeit beim Eingeben von HTML-Tags auf einen Mausklick, erlaubt die Eingabe der Umlaute direkt mit der Tastatur (oder nach TeX-Konventionen) und führt den Benutzer durch kontextsensitiv wechselnde Menüs, die jeweils nur die regelkonformen Tags anbieten. Zum anderen stehen Prüfprogramme bereit, die die Konsistenz der HTML-Tags überprüfen und detaillierte (über 100 verschiedene) Hinweise auf mögliche Inkonsistenzen geben. Auch für das Erstellen von Querverweisen (Hyperlinks) werden Hilfen gegeben. Durch die vielen Hilfsmenüs setzt die Benutzung des HTML-Editors weder HTML-Kenntnisse noch TUSTEP-Kenntnisse voraus.

Technische Voraussetzungen des HTML-Editors

Der HTML-Editor läuft auf dem Text-Server (für andere Rechner gibt es Versionen auf Anfrage) und benötigt eine Maus. D. h. er kann nur von einem X11-Fenster (das Fenster muß mindestens 25 Zeilen haben) oder einem PC mit einem entsprechenden KERMIT genutzt werden.

Bei Interesse an einer Version für ASCII-Terminals, die anstelle der Maus die Funktionstasten verwendet, wird um Rückfrage gebeten.

Aufruf des HTML-Editors

Den HTML-Editor gibt es in zwei Fassungen, die auch verschieden aufzurufen sind:

1. Eine Version für TUSTEP-Benutzer, die ihre Daten in TUSTEP-Dateien verwalten, mit TUSTEP umgehen können und den HTML-Editor innerhalb von TUSTEP aufrufen wollen.
2. Eine Version für UNIX-Benutzer, die die zu verarbeitenden Daten in UNIX-Dateien verwalten, die von TUSTEP nichts wissen, daher den Aufruf von der UNIX-Ebene aus starten wollen, und die bei der Dateneingabe Umlaute nach TeX-Konventionen eingeben wollen.

1. Aufruf der »TUSTEP«-Version

Zur Vorbereitung des Aufrufs muß *einmalig* folgender UNIX-Befehl gegeben werden:

```
% source /soft/tustep/lddv
```

In Zukunft kann dann der HTML-Editor innerhalb von TUSTEP folgendermaßen aufgerufen werden:

```
Gib Kommando >
```

```
#an, zrlldv1*htmlprog, tr=lddv
```

```
Gib Kommando >
```

```
#tu, zrlldv1*htmlprog, start
```

Alle weiteren erforderlichen Angaben, werden abgefragt.

Das zu bearbeitende Dokument wird in einer TUSTEP-Datei erwartet. Sie kann bereits Text enthalten, wobei die Umlaute und Akzentbuchstaben nach TUSTEP-Konventionen eingetragen sind. Gibt es eine solche Datei noch nicht, so wird sie neu angelegt.

2. Aufruf der »UNIX«-Version

Zur Vorbereitung des Aufrufs muß man sich *einmalig* ein Verzeichnis namens *htmledit* einrichten, das sich in einem beliebigen Pfad befinden kann.

```
% mkdir htmledit
```

Vor dem Aufruf muß man *immer* in dieses Verzeichnis *htmledit* wechseln:

```
% cd htmledit
```

Der Aufruf erfolgt nun mit folgendem Befehl:

```
% source /home/textserv/zr/htmledit
```

Alle weiteren Angaben werden abgefragt. Das zu bearbeitende Dokument wird als UNIX-Datei erwartet. Existiert sie noch nicht, so wird sie angelegt. Sie kann aber auch schon Text enthalten, in dem die Umlaute nach TeX- und/oder HTML-Konventionen eingetragen sind. Beim Laden der Datei in den Editor werden die Umlaute in jedem Fall nach TeX-Konventionen ("a für ä) konvertiert, so

daß der Text problemlos lesbar ist, beim Abspeichern werden die Umlaute in jedem Fall nach HTML-Konventionen konvertiert.

Die drei Arbeitsweisen des Programms

Beim Starten des Programms wird die gewünschte Arbeitsweise abgefragt: *Edieren*, *Pruefen* oder *Konvertieren*. *Edieren* ist die Voreinstellung (leere Eingabe), da hierin auch die anderen Funktionen enthalten sind.

Um einen Überblick über Leistung und Arbeitsweise des HTML-Editors zu geben, seien hier die wichtigsten Funktionen knapp beschrieben. Eine Beschreibung aller Funktionen findet sich in den Hilfsmenüs.

1. Das Edieren

Wird die Arbeitsweise *Edieren* gewählt, so wird das Dokument in den Editor zur Bearbeitung geladen (die Umlaute werden dabei in jedem Fall in der gewohnten Form angezeigt). Falls es die angegebene Datei noch nicht gibt, wird eine neue Datei angelegt.

In der Menüleiste erscheint *HTML_Start*. Sie klicken die Menüleiste mit der Maus an, nachdem Sie die Eingangsinformationen gelesen haben. Die Eingangsinformationen verschwinden und es erscheint eine Menüleiste mit mehreren Feldern. Hier wählt man aus, was man tun möchte, indem man mit der Maus das entsprechende Feld anklickt. Das Feld *HELP* gibt grundsätzlich kontextsensitive Hilfe und erklärt die Wirkung der übrigen Menü-Felder. Mit *UP* und *DOWN* kann grundsätzlich um einen Bildschirm rückwärts oder vorwärts geblättert werden, ggf. ausgehend von der Zeile, in der der Cursor steht. *CLEAR* ist nur im Fehlerfall zu benutzen, falls der Editor hartnäckig bei immer derselben Fehlermeldung bleibt. Unter Verlust der auf dem Bildschirm neu eingetragenen Daten ist damit ein Weiterarbeiten möglich.

Wählt man das Feld *NewDoc* oder *Append* gelangt man in den eigentlichen Arbeitsmodus zur Eingabe von neuem Text. Der Cursor steht dabei in der Regel im Textfeld, in dem die Daten eingegeben werden. Die Menüleiste bietet (außer dem Blättern *UP* und *DOWN* sowie der Hilfe *HELP* und *CLEAR*) Felder an, hinter denen sich HTML-Tags verbergen. Durch Anklicken eines solchen Feldes wird das entsprechende Tag in den Text geschrieben (Wichtig!!! Der Cursor muß beim Anklicken im Textfeld des Editors stehen!) und anschließend kontextsensitiv eine neue Menüleiste gezeigt,

die die Tags anbietet, die nun HTML-konform möglich sind. Folgt man beim sequentiellen Erstellen eines Dokumentes ausschließlich diesen Menüs zum Eintragen von Tags, so entsteht in jedem Fall ein HTML-konformes Dokument.

Einige Felder der primären Elemente *Heading*, *Lists* und *Misc* bieten zunächst ein Untermenü an. Erst durch Anklicken eines Feldes im Untermenü erfolgt die entsprechende Aktion. Wird ein Mausklick an beliebiger Stelle außerhalb des Untermenüs ausgeführt, verschwindet diese Menüleiste wieder, ohne daß eine Aktion stattgefunden hat.

Die Hauptmenüleisten enthalten das Feld *DONE*, mit dem man (nachdem der Text eingegeben ist) ins End-Menü zum Erstellen von Links, Prüfen und Konvertieren des Dokumentes gelangt.

Die Leiste des End-Menüs bietet u. a. folgende Leistungen an:

QUIT verläßt den HTML-Editor. Mit dem Untermenü *SaveAndQuit* wird die Konvertierung der Daten (v. a. Umlaute) in das HTML-Format vorgenommen (siehe unten).

Checking prüft die Konsistenz der HTML-Tags und der internen Links (siehe unten).

PreView bietet eine sehr einfache Formatierung des Dokuments am Bildschirm. Es dient dazu, einen ersten Eindruck von der typographischen Wirkung der Tags zu bekommen, ohne daß man dazu TUSTEP verlassen muß.

Das Feld *MakeLink* dient als Hilfe zur Erstellung von Links. Es erspart vor allem Schreibarbeit und bietet kleine Hilfen zur Fehlervermeidung an. Das Erstellen externer Links wird wesentlich erleichtert, wenn man sich die Uniform Resource Locators (URLs) der Dokumente, auf die man häufig verweisen will, in einer eigenen Datei verwaltet, die man beim Starten des HTML-Editors angeben kann. Man kann nun beim Erstellen von Links sich diese Datei zeigen lassen (Feld *Ex_List* im Untermenü) und durch Anklicken eines Namens diesen in den Link seines Dokumentes übernehmen.

2. Das Prüfen

Die Arbeitsweise *Pruefen* kann gleich beim Start des Programms gewählt werden oder durch das Anklicken des Feldes *Checking* in den Menüs des Editors. Beidesmal wird das

selbe Programm gestartet, das die Daten der angegebenen Datei bzw. den aktuellen Zustand der im Editor bearbeiteten Daten einer Konsistenzprüfung unterzieht. Geprüft wird zunächst die Legalität der vorkommenden Sonderzeichen und Tags. Anschließend werden die Tags geprüft, ob die dadurch ausgezeichneten Textelemente den Regeln von HTML entsprechen. Falsche Stellen im Dokument werden spezifiziert durch über 100 verschiedene Fehlermeldungen angezeigt. Es erfolgt zunächst die Fehlermeldung; dieser folgen eine oder mehrere Textstellen, die diesen Fehler haben. Bei den Textstellen ist vorne eine Nummer angegeben. Sie entspricht der Numerierung der einzelnen Records im Editor (TUSTEP-Satznummer). Mit Hilfe dieser Nummer können die Stellen im Editor zur Korrektur aufgesucht werden.

Anschließend werden die internen Links geprüft und zuletzt eine Liste aller Links ausgegeben.

Die Ausgabe des Fehlerprotokolls erfolgt wahlweise auf den Bildschirm oder auf einen Drucker. Voreinstellung für den Drucker ist *ps030.d4*. Ein anderer Drucker kann beim Start des HTML-Editors angegeben werden.

3. Das Konvertieren

Das Konvertieren, das man durch die Angabe *Konvertieren* beim Starten des HTML-Editors bzw. implizit durch das Feld *SaveAndQuit* beim Verlassen des Editors aufruft, konvertiert die Umlaute und Sonderzeichen, die man nach TeX bzw. TUSTEP-Konventionen in seinen Daten stehen hat, in die HTML-konformen Entities.

In der »TUSTEP«-Version des HTML-Editors wird das konvertierte Ergebnis in einer SDF-Datei namens *transfer.htm* im aktuellen Projekt abgelegt. Diese Datei kann auf den Informations-Server übertragen und ins Tübinger Informationssystem eingebunden werden. Bei *SaveAndQuit* werden gleichzeitig die im Editor veränderten Daten in der angegebenen TUSTEP-Datei gesichert.

In der »UNIX«-Version des HTML-Editors wird das konvertierte Ergebnis auf die angegebene Datei zurückgeschrieben.

Bei Fragen zum HTML-Editor wenden Sie sich bitte an Dr. W. Bader, Tel. 29-6973, Zi. 125 im ZDV.