Konvertierungsbericht zum Buch "Wissensrepräsentation und Inferenz"

Christoph Quix   Thomas List   René Soiron

30. Januar 1997

Die Konvertierung des Buches "Wissensrepräsentation und Inferenz" erfolgte in den folgenden Arbeitsschritten:

- Vorarbeiten
- Grobkonvertierung
- manuelles Nachbearbeiten

1 Vorarbeiten

1.1 Vervollständigung der Source-Files

Der ursprünglich gelieferte Satz von Source-Files, der vom Lehrstuhl des Autors zur Verfügung gestellt worden war, war nicht vollständig. Nach mehrfacher Kommunikation mit dem Lehrstuhlmitarbeiter Georg Krafczyk wurden die fehlenden Bibtex- und Style-Files identifiziert und nachgeliefert.

1.2 Analyse und Auswahl des Konvertierungswerkzeugs

Zunächst wurde aufgrund von Produktinformationen das Leistungsvermögen mehrerer Werkzeuge untersucht (z.B. latex2html, tex2rtf, Adobe File Utilities). Weiterhin wurde untersucht, ob es zu der Transformationen von mathematischen Formeln in Gif-Bilder eine Alternative gab. Der HTML Standard 3.0 enthielt zwar mathematische Elemente, diese reichten aber nicht aus. Außerdem ist dieser Standard durch den 3.2 Standard ersetzt worden, der keine mathematische Umgebungen mehr aufweist. Als Konvertierungswerkzeug wurde latex2html gewählt und installiert.

Weitere eingesetzt Werkzeuge:

rcs: Mittels rcs und make wurde eine Versionsumgebung erstellt, die das parallele Arbeiten auf den Quelldateien ermöglichte.

xv: Für die Konvertierung und Bearbeitung von Grafiken

Grafikkonverter (Mac): Für die Konvertierung von Grafiken

Scanner und zugehörige Software: Scannen der Titelseite des Original-Buches

1.3 Vorbereitung der Source-Files

Die Quelldateien wurden Kapitel- bzw. Abschnittsweise untersucht und testweise mit latex2html übersetzt, um Probleme zu sondieren. Folgende Schwierigkeiten wurden durch Änderungen des latex-Quelltexts behandelt:

- Transformation von tabular Umgebungen, z.B Tabellen mit leeren Einträgen
- Transformation des Wörterbuchs in eine Tabular-Umgebung
- Transformation von benutzerdefinierten Aufzählungsumgebungen, beipielsweise in den Übungen
- Aufspalten der Übungen in eigene Subsections, damit jeder Übungsaufgabe automatisch eine eigene Seite zugeordnet wird
- Entfernen von Zeilenumbrüchen in index-Eintraegen, da diese in der HTML-Version ansonsten fehlerhaft waren
- Ändern der Theorem-Umgebungen: Die Theorem-Umgebungen wurden in latex2html wegen der Nummerierung in ein großes GIF-Bild übersetzt. Die Makros für diese Umgebungen wurden so geändert, daß nur der Kopf mit der Nummerierung als GIF übersetzt wird. Der Rumpf einer solchen Umgebung wird dann als normaler Text behandelt. Dadurch sind auch Fußnoten und Verweise in den Theorem-Umgebungen möglich.

Während dieser Phase wurde die neueste latex2html Version 96.1 Rev. g installiert. Diese enthielt neue Features (z.B. die $BODYTEXT-Variable) aber leider auch bugs, z.B. bei der Auflösung von Referenzen. Dieser Fehler wurde durch manuelle Korrektur des Skripts behoben.

2 Konvertierung des Gesamtdokuments

Nachdem die einzelnen Tex-Dateien vorbereitet waren, wurde versucht das Gesamtdokument mittels latex2html zu konvertieren. Dieser Versuch schlug zunächst mit einem Absturz von latex2html fehl. Als Problem wurde der enorme Speicherbedarf identifiziert. Die Grobkonvertierung war in der folgenden Umgebung schliesslich erfolgreich:

- Maschine: SPARC Ultra 1
- Betriebssystem: Solaris 2.5
- Hauptspeicher: 184 MB
- Swap-Space: 1GB
- Laufzeit: etwa 3 Stunden

Den Großteil der Laufzeit nahm die Generierung der 2104 gif-Dateien in Anspruch. Die Verarbeitungsgeschwindigkeit konnte durch die Verwendung einer lokalen Platte des Rechners erhöht werden.

In der ersten Phase der Übersetzung wächst der Speicherbedarf von latex2html durch den Aufruf von Subprozessen wie *latex* oder *ghostscript* durch einen *fork* Befehl stark an. Durch diesen Befehl verdoppelt sich kurzzeitig der Speicherbedarf. Die Bereitstellung von mehr Swapspace hilft hier, da zwar viel Speicher angefordert, aber anscheinend nicht darauf zugegriffen wird, da der Rechner nur selten "swappen" muss.

Interessant wäre zu untersuchen, wodurch der Speicherbedarf entsteht (wahrscheinlich durch viele Verweise und Kapitel) und ob die in latex2html mögliche Dokument-Segmentierung die Speicherbedarf reduziert (Segmentierung funktioniert allerdings nur bei latex2e-Dokumenten).

3 Manuelle Nachbearbeitung

Die durch latex2html erzeugte HTML-Version wurde wie folgt manuell aufbereitet:

- Kontrolle der erzeugten HTML Dateien durch Vergleich mit gedrucktem Buch
- Einige Abbildungen mussten neu gezeichnet werden, da die von latex2html erzeugten GIFs nicht brauchbar waren
- Wahl der Hintergrundfarbe
- Erzeugen des Deckblatts
- Auswahl und Manipulation der Navigations-Icons
- Korrektur von Fehlern bei der latex2html-Konvertierung:
  - Backslashes, die im Quelltext am Zeilenende standen, wurden in den HTML-Text übernommen.
  - Bei Verweisen auf Kapitel landet man nicht in der ersten Zeile des HTML-Dokuments sondern in der Kapitel-Überschrift
  - Fehlende Bilder im Index

  Häufige Fehler wurden durch sed/awk/perl-Skripte korrigiert
- Nachbearbeitung von Inhaltsverzeichnis und Index
- Suche nach Homepages der Buchautoren im Internet und Aufnahme entsprechender Verweise in das Deckblatt
- Erstellung des Konvertierungsberichts

Die erzeugte HTML-Version umfaßt 235 html und 2128 gif Dateien. Der gesamte Speicherbedarf beträgt 4,6 MB.