

Konvertierung von ASCII-Textdateien nach WinWord

Walter Riemer

Die Konvertierungsmethoden, die WinWord 6.0 anbietet, sind für ASCII-Textdateien (wie sie jeder ASCII-Editor erstellt) unbefriedigend.

In einer ASCII-Textdatei ist jede Zeile mit CR LF (0D 0A) abgeschlossen; Absätze werden durch eine Leerzeile begrenzt. Außerdem sind Umlaute und scharfes S anders codiert als im von Word verwendeten ANSI-Code.

Word bietet folgende Konvertierungen an

Nur Text unbrauchbar, weil ANSI-Zeichensatz vorausgesetzt. Zeilenende wird zu Absatzende konvertiert, daher wird aus jeder Zeile ein Absatz.

Nur MS-DOS-Text besser, weil ASCII-Zeichensatz vorausgesetzt ist. Das Absatzproblem ist das gleiche.

Nur Text oder Nur MS-DOS-Text, jedoch mit Zeilenwechsel im Prinzip gleiche Arbeitsweise.

Nur Text oder Nur MS-DOS-Text, jedoch mit Layout Das Absatzproblem besteht nicht, jedoch ist kein Unterschied zwischen Text mit Layout und MS-DOS-Text mit Layout hinsichtlich des ANSI-Codes: beide Konvertierungen setzen offensichtlich ANSI-Code voraus, was wohl ein Fehler ist.

Es bieten sich daher drei Möglichkeiten an

- mit einem Makro in Word die Umsetzung durchführen
- mit einem C-Programm den ASCII-Code in ANSI-Code umwandeln und dann mit „Text mit Layout“ konvertieren
- mit einem C-Programm die Zeilenwechsel entsprechend umsetzen und dann mit „MS-DOS-Text“ konvertieren.

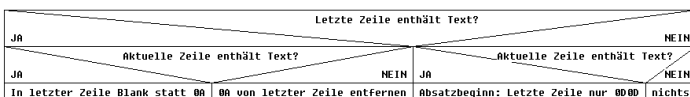
Die letzte Möglichkeit wurde gewählt.

Das Programm setzt ein CR-LF am Zeilenende in ein Blank um, da Word einen Absatz gewissermaßen als eine einzige Zeile, welche den gesamten Absatz enthält, speichert.

Falls aber auf ein CR-LF ein weiteres folgt (Kennzeichnung eines Absatzendes), wird das erste entfernt und anstelle des zweiten die Word-Absatzende-Markierung CR-CR eingesetzt.

In weiterer Ausbaustufe des Programms erwies es sich als zweckmäßig, auch noch die Umsetzung in ANSI-Code auszuführen, sodaß es dann gleichgültig ist, ob in Word mit „Nur Text“ oder mit „Nur MS-DOS-Text“ konvertiert wird.

Von Vorteil ist es, wenn in der Quelldatei keine abgeteilten Wörter vorkommen, da unser Programm natürlich nicht sicher sein kann, ob es sich um einen Abteil-Strich (der entfernt gehört) oder einen Bindestrich handelt.



Das Struktogramm stellt nur den kritischen Teil dar, nämlich das Behandeln der Zeilenenden in der ASCII-Datei (Zeilen 27 bis 36). Zu beachten ist dabei, daß aufgrund der C-Konventionen nicht auf 0D0A, sondern nur auf 0A abzufragen ist.

```

1 /* AscWord.C .
   Wandel t ASCII-Textdatei in MSDOS-Textdatei ohne 0AH um.
2   ODOA wird durch Blank ersetzt,
   ODOAODOA (Leerzeile als Absatzende) durch ODD.
3   Umlaute und á werden in ANSI-Code umgesetzt */
4 #include <stdio.h>
5 #include <string.h>
6 #include <conio.h>
7 #include <stdlib.h>
8
9 char QuellDateiName[80], ZielDateiName[80],
10 AktZeile[128], LetztZeile[128]=" ";
11 int ZeilenNr=0;
12 FILE *QDatei, *ZDatei;
13
14 void Abort(void);
15 void LetztZeileSchreiben(void);
16
17 void BearbeiteZeile(void)
18 { int LetztLaenge;
19   LetztLaenge=strlen(LetzZeile);
20   ZeilenNr++;
21   puts("");
22   gotoxy(1, wherey()-1);
23   printf("Zeile Nr. %d", ZeilenNr);
24   if (kbhit())
25     Abort();
26   if (LetztZeile[0] != 0x0A)
27   { if (AktZeile[0] != 0x0A)
28     LetztZeile[LetztLaenge-1]=' ';
29     else
30     LetztZeile[LetztLaenge-1]=0;
31   }
32   else
33     if (AktZeile[0] != 0x0A)
34     LetztZeile[0]=0;
35   LetztZeileSchreiben();
36 }
37
38 void Abort(void)
39 { printf("\n\nAbbruch bei Zeile Nr. %d:\n%s\n", ZeilenNr, LetztZeile);
40   exit(3);
41 }
42
43 void LetztZeileSchreiben(void)
44 { char *Ptr;
45   Ptr=LetztZeile;
46   while (*Ptr != 0)
47   { switch (*Ptr)
48     { case ' ': *Ptr=0xC4; break;
49       case '": *Ptr=0xD6; break;
50       case 's': *Ptr=0xDC; break;
51       case ',': *Ptr=0xE4; break;
52       case '*': *Ptr=0xF6; break;
53       case 'á': *Ptr=0xFC; break;
54       case 'á': *Ptr=0xDF;
55     }
56     fputc(*Ptr, ZDatei);
57     Ptr++;
58   }
59   if (*(Ptr-1) != ' ')
60     fputc(0x0D, ZDatei);
61 }
62 main()
63 { char *EOFPtr;
64   puts("\nQuelldatei ? ");
65   gets(QuellDateiName);
66   QDatei=fopen(QuellDateiName, "r");
67   puts("Zieldatei ? ");
68   gets(ZielDateiName);
69   ZDatei=fopen(ZielDateiName, "w");
70   EOFPtr=fgets(AktZeile, 129, QDatei);
71   do
72   { strcpy(LetzZeile, AktZeile);
73     EOFPtr=fgets(AktZeile, 129, QDatei);
74     BearbeiteZeile();
75   }
76   while (EOFPtr != NULL);
77   printf("\nUmwandlung beendet.\n");
78   fclose(QDatei);
79 } /* AscWord.C . Wandel t ASCII-Textdatei
   in MSDOS-Textdatei ohne 0AH um.

```