

DESY-Bibliothek

13. SEP. 1971

Interner Bericht
DESY F58-71/3
Juli 1971

T Y P S E T

Ein Programm zur Formatierung von Texten
für die PDP-8

G. Hochweller

Inhaltsverzeichnis

=====

- 1 Einführung
- 2 Laden des Programms
- 3 Typewriter Input
- 4 Typesetting
 - 4.1 Eingabe der benötigten Parameter
 - 4.2 Randausgleich und Worttrennung
 - 4.3 Leerzeilen
 - 4.4 Leerstellen innerhalb von Worten
 - 4.5 Beschreibung der Steuerkommandos
 - 4.5.1 Neue Seite
 - 4.5.2 Neue Zeile
 - 4.5.3 Verschiebung des linken Randes
 - 4.5.4 Verschiebung des rechten Randes
 - 4.5.5 Unterstreichen
 - 4.5.6 Setzen von Tabulatorwerten
 - 4.5.7 Tabulator
 - 4.5.8 Sprung auf eine angegebene Position
- 5 Anhang

1 EINFÜHRUNG

Das Programm 'TYPSET' dient dazu, in freier Form eingegebene Texte so zu formatieren, daß ein druckfähiger Satz entsteht. Die Steuerung der Formatierung erfolgt im wesentlichen über eine Reihe von Steuerkommandos, die mit dem Text selbst zusammen eingegeben werden und als Bestandteil des Textes auf dem Eingabeband abgespeichert werden. Die Darstellung der Texte erfolgt im standardmäßigen Editortextformat (siehe Beschreibung EDITORPROGRAMM SUPEDT, F58-69/1); ausgedruckt werden die formatierten Texte mit dem normalen SUPEDT-Programm.

Um die Eingabe nicht auf dem Teletype vornehmen zu müssen (die Eingabe über Schreibmaschine ist im SUPEDT-Programm nicht unterstützt), enthält das Programm 'TYPSET' einen eigenen Eingabeteil, der es ermöglicht, beliebige Texte unter Display-Kontrolle über die Schreibmaschine (IBM 73) einzugeben. Die Berücksichtigung der Groß/Klein-Umschaltung erfolgt dabei automatisch.

2 LADEN DES PROGRAMMS

=====

- a. Systemband, das das Programm 'TYPSET' enthält, auf Einheit 8 montieren
- b. Einheit 8 auf "REMOTE" und "WRITE LOCK" (Kippschalter an der Bändeinheit) schalten
- c. Falls die Nummer 8 noch für eine andere Bändeinheit gewählt ist, diese ausschalten oder eine andere Nummer einstellen (Drehschalter an der Bändeinheit)
- d. "DATA FIELD" und "INST FIELD" (Schalter an der PDP) beide auf 0 einstellen
- e. "SWITCH REGISTER" (Schalter an der PDP) auf 7600 stellen
- f. "LOAD ADD" (Taste an der PDP) drücken
- g. "START" (Taste an der PDP) drücken
- h. Nachdem der Teletype ein Geräusch von sich gegeben hat, TYPSET schreiben (auf dem Teletype)
- i. "RETURN" (Taste am Teletype) drücken

Das Programm 'TYPSET' wird nun vom Systemband in den Speicher geladen. Auf dem Display erscheint nach Beendigung des Ladevorgangs der Text

ENTER WANTED MODE

'I' TYPEWRITER INPUT
'T' TYPESETTING
'S' RETURN TO SUPEDT

Durch Drücken der entsprechenden Taste am Teletype wird der gewünschte Programmteil gestartet.

3 TYPEWRITER INPUT

Dieser Programmteil ermöglicht es, beliebige Texte über die Schreibmaschine einzugeben; sie werden im standardmäßigen Editortextformat auf DECTape abgespeichert. Erforderlich ist ein Band auf Einheit 6 (WRITE ENABLED) als Ausgabeeinheit.

Vor Beginn der Texteingabe müssen noch einige Parameter über den Teletype eingegeben werden. Zunächst erscheint auf dem Display die Frage nach der Nummer des ersten Editorblockes, auf dem der Text abgespeichert werden soll:

FIRST EDITOR BLOCK (DEC):

Einzugeben ist eine Zahl (dezimal), die zwischen 0 und 91 (einschließlich) liegen muß; abgeschlossen wird die Eingabe der Zahl durch Drücken der Taste "RETURN". Ist der eingegebene Text länger als ein Editorblock, so werden automatisch die nachfolgenden Blöcke benutzt.

Da die Möglichkeit besteht, daß von einer früheren Eingabe her auf dem ersten Editorblock schon Text steht, der nicht überschrieben, sondern fortgesetzt werden soll, erscheint als nächstes die Frage

DO YOU WANT TO CONTINUE THE TEXT
ON THIS EDITOR BLOCK ?

PRESS Y(ES) OR N(O)

Drückt man nun die Taste 'Y', so wird der Inhalt des angegebenen Editorblockes in den Textspeicher eingelesen. Da bei Beendigung der Eingabe vom TYPSET-Programm automatisch noch eine Leerzeile generiert wird, wird hier nun die letzte Zeile gelöscht. Der nachfolgend eingegebene Text wird dann an den Inhalt des

Textspeichers angehängt. Drückt man dagegen die Taste 'N', so geht auf dem gewählten ersten Editorblock eventuell schon vorhandener Text verloren; er wird durch den neu eingegebenen Text überschrieben.

Nun kann die Eingabe des Textes beginnen. Auf dem Display wird während der Eingabe folgendes angezeigt:

CURRENT EDITOR BLOCK: x

LINE: y CC = z

Bisheriger Inhalt der laufenden Zeile

Dabei bedeuten 'x' die Nummer des Editorblocks, auf dem der zur Zeit eingegebene Text abgespeichert werden wird, 'y' die Nummer der laufenden Zeile in diesem Editorblock und 'z' die bisherige Anzahl von Zeichen in dieser Zeile. Darunter wird die Zeile selbst unter Berücksichtigung der Groß/Klein-Umschaltung angezeigt. Der Buchstabe 'ß', der auf unserem Display nicht darstellbar ist, erscheint dabei als '#'.
(

Bei der Eingabe der Texte kann im wesentlichen wie beim normalen Schreiben auf einer Schreibmaschine vorgegangen werden; es bestehen nur die folgenden kleinen Unterschiede:

- a. Beim Beginn einer neuen Zeile wird immer davon ausgegangen, daß sich die Schreibmaschine in der Stellung 'KLEIN' befindet.
- b. Das Drücken der Taste 'BACKSPACE' an der Schreibmaschine führt zum Löschen des zuletzt eingegebenen Zeichens einschließlich eventuell davor oder dahinter stehender Umschaltzeichen. Durch wiederholtes Betätigen dieser Taste läßt sich die laufende Zeile von rechts her beliebig weit löschen.

Soll die Eingabe beendet werden, so muß auf dem Teletype die Taste 'S' gedrückt werden. Dadurch wird erreicht, daß eine

zusätzliche Leerzeile generiert und der Rest des Textes, der sich noch im Textspeicher befindet, auf das Band geschrieben wird. Auf dem Teletype erscheint dann die Meldung:

LAST EDITOR BLOCK (DEC): x

'x' ist dabei die Nummer des letzten für die Speicherung des eingegebenen Textes benutzten Editorblocks. Anschließend springt das Programm wieder an seinen Ausgangspunkt zurück.

4 TYPESETTING

Dieser Programmteil dient zur eigentlichen Formatierung des Textes. Der unformatierte Text wird dabei auf einem DECTape auf Einheit 6 (WRITE LOCK) erwartet; der formatierte Text wird auf ein weiteres DECTape auf Einheit 5 (WRITE ENABLED) ausgegeben. Begonnen wird auf dem Ausgabeband mit Editorblock 1.

4.1 Eingabe der benötigten Parameter

Für diesen Programmteil werden eine Reihe von Parametern benötigt; sie müssen über den Teletype eingegeben werden. Sämtliche Zahlen sind dabei durch das Drücken der Taste "RETURN" abzuschließen.

Zunächst wird gefragt, wo der zu formatierende Text auf dem Eingabeband steht. Auf dem Display erscheint

FIRST EDITOR BLOCK (DEC):

Einzugeben ist die Nummer des ersten Editorblockes, auf dem der Text abgespeichert ist ($0 \leq x \leq 91$). Dann folgt

LAST EDITOR BLOCK (DEC):

Hier ist die Nummer des letzten für die Speicherung des zu formatierenden Textes benutzten Editorblockes auf dem Eingabeband einzugeben ($0 \leq x \leq 91$); sie darf natürlich nicht kleiner als die Nummer des ersten sein.

Ein weiterer Parameter, der für die Formatierung wichtig ist, ist der Zeilenabstand. Es sind drei verschiedene Zeilenabstände möglich: einzeilig, anderthalbzeilig und zweizeilig (siehe Tab. 1). Es wird gefragt

DO YOU WANT 1.5FOLD LINE SPACING ?

PRESS Y(ES) OR N(O)

Wird die Taste 'N' gedrückt, so erscheint

DO YOU WANT DOUBLE LINE SPACING ?

PRESS Y(ES) OR N(O)

Wird wiederum die Taste 'N' gedrückt, so wird einfacher Zeilenabstand angenommen. Ist der gewünschte Zeilenabstand gewählt, so hat sich der Benutzer zu entscheiden, ob er die Seiten durchnumeriert haben möchte oder nicht.

DO YOU WANT PAGE NUMBERS ?

PRESS Y(ES) OR N(O)

Falls Numerierung der Seiten gewünscht wird, ist nun die Nummer der ersten Seite anzugeben.

ENTER FIRST PAGE NUMBER:

Erwartet wird hier eine Zahl zwischen 1 und 999; die Seitennummer 0 ist nicht zulässig. Das TYPSET-Programm sorgt dafür, daß die Seitennummer möglichst in der Mitte des zum Druck benutzten Papierteils sitzt. Zur Berechnung dieser Position muß jedoch die Gesamtzahl von Zeichen pro Zeile bekannt sein. Es erscheint also

die Anweisung

ENTER NUMBER OF CHARACTER LOCATIONS PER LINE
(USED FOR CALCULATION OF PAGE NUMBER
POSITION ONLY!)

Die hier eingegebene Zahl hat keine Bedeutung für die eigentliche Textformatierung; sie wird ausschließlich für den angegebenen Zweck benutzt.

Seitenformatierung	Zeilenzahl					
	Leer	Nummer	Leer	Text	Leer	
ein- zeilig	ohne mit	- 3	- 1	5 3	60 58	1 1
anderthalb- zeilig	ohne mit	- 2	- 1	4 2	39 38	1 1
zwei- zeilig	ohne mit	- 3	- 1	5 3	60 58	1 1

Tab. 1: Seitenaufteilung bei den verschiedenen Formaten

Nachdem nun alle Parameter, die das Programm benötigt, bekannt sind, beginnt die eigentliche Formatierung. Da an sich jeder beliebige Text legal ist, können Fehler nur in den Steuerkommandos auftreten. In einem solchen Fall erscheint auf

dem Teletype die Meldung

```
COMMAND ERROR IN BLOCK x, LINE y
```

Die Zeile 'y' des Editorblockes 'x' des Eingabebandes enthält ein illegales Steuerkommando. Das Programm kehrt dann zu seinem Ausgangspunkt zurück, da eine weitere Verarbeitung des Textes unter diesen Umständen nicht mehr sinnvoll ist.

4.2 Randausgleich und Worttrennung

Eine der Hauptaufgaben des TYPSET-Programms ist es, für einen sauberen Randausgleich zu sorgen. Hierzu wird der eingegebene Text wortweise gelesen; nach jedem Wort wird untersucht, ob es noch in die laufende Zeile paßt oder nicht. Sobald die Stelle erreicht ist, bei der die laufende Zeile bei der vorgegebenen Länge das folgende Wort nicht mehr aufnehmen kann, werden die noch freien Positionen in Form von Leerstellen auf die einzelnen Wortzwischenräume in der Zeile verteilt. Die Zeile wird also so 'gespreizt', daß das rechte Ende des letzten Wortes immer auf derselben Position landet, der rechte Rand also ausgeglichen wird.

Dabei kann es vorkommen, daß durch eine Aufeinanderfolge ungünstiger Wortlängen so viele Leerstellen eingeschoben werden müßten, daß ein unschönes Schriftbild entstehen würde. In einem solchen Fall wird dem Benutzer vom TYPSET-Programm die Möglichkeit gegeben, das nachfolgende Wort zu trennen. Dazu erscheint beispielsweise folgender Display:

```
| b e i s p i e l s w | l e i s e
```

'SPACE' TO SHIFT RIGHT

'L' TO SHIFT LEFT

'RET' TO CONTINUE

Oben steht das Wort, das zum Trennen angeboten wird (hier also 'beispielsweise'). Der Doppelstrich zeigt an, welcher Teil des Wortes allerhöchstens noch in die laufende Zeile paßt (unter Berücksichtigung des zu generierenden Trennungsstriches). Der einfache Strich läßt sich durch Drücken der Tasten 'SPACE' bzw. 'L' nach rechts bzw. links verschieben und gibt die Stelle an, an der getrennt werden soll. Der Benutzer hat nun diesen Strich auf die von ihm gewünschte Position zu setzen und dann die Taste 'RETURN' zu drücken. Wird der Strich ganz links stehengelassen, so unterbleibt das Trennen.

Soll ein Wort, das einen Bindestrich enthält, an dieser Stelle getrennt werden, so ist der Markierungsstrich auf die rechte Seite des Bindestrichs zu schieben; in diesem Fall wird kein zusätzlicher Trennungsstrich erzeugt.

Normalerweise wird ein Wort zum Trennen angeboten, wenn sämtliche Zwischenräume in einer Zeile aus mehr als zwei Leerstellen bestehen würden. Diese Zahl kann vom Benutzer geändert werden, indem er vor der Eingabe der Parameter das 'SWITCH REGISTER' an der PDP auf einen Wert zwischen 1 und 5 (einschließlich) einstellt; der eingestellte Wert wird dann als Schwelle für das Trennen benutzt. Liegt der Wert außerhalb dieses Bereiches (normalerweise 7600 vom Starten des Programms her), so bleibt es bei der Grenze von zwei Leerstellen.

4.3 Leerzeilen

Jede Leerzelle auf dem Eingabeband führt auch zu einer Leerzeile im formatierten Text. Beim Auftreten einer Leerzeile wird die laufende Zeile beendet und (unformatiert) abgespeichert; anschließend wird eine Leerzeile generiert. Befindet man sich jedoch gerade am Anfang einer neuen Zeile, so unterbleibt das Abspeichern dieser 'leeren' Zeile, d.h. es wird ebenfalls nur eine Leerzeile erzeugt.

4.4 Leerstellen innerhalb von Worten

Für das Formatierungsprogramm bedeutet jede Leerstelle das Ende eines Wortes. Da sämtliche Wortzwischenräume jedoch zur Formatierung bzw. zum Ausgleich des rechten Randes herangezogen werden können, ist es nicht möglich, bestimmte ausgewählte Leerstellen aus dem Formatierungsvorgang herauszuhalten. Da man für bestimmte Fälle jedoch die genaue Anzahl von Leerstellen zwischen zwei Worten schon bei der Eingabe festlegen und sie nicht mehr verändert bekommen möchte, wurde ein besonderes Zeichen, das normalerweise nicht in den zu formatierenden Texten vorkommt, zum zusätzlichen 'SPACE' umdefiniert. Dieses Sonderzeichen ist, je nach Maschinentastatur, entweder der '\$' oder der '§' (ASCII-Code 44). Taucht dieses Zeichen im Eingabetext auf, so wird es im formatierten Text als 'SPACE' ausgegeben, ohne jedoch als Wortgrenze zu zählen.

Benutzt werden muß dieses Zeichen auf jeden Fall immer dann, wenn irgendwo mehr als eine Leerstelle erzwungen werden soll.

4.5 Beschreibung der Steuerkommandos

Da die Steuerkommandos als normaler Bestandteil des zu formatierenden Textes mit auf dem Eingabeband stehen, mußte eine besondere Markierung eingeführt werden, um die Steuerkommandos vom Text selbst unterscheiden zu können. Zu diesem Zweck wurde ein bestimmtes Zeichen als Kommandozeichen definiert, das jeweils direkt vor und nach Steuerinformation eingegeben werden muß und diese somit aus dem Text heraushebt. Als Kommandozeichen dient das '&'; es darf somit im zu formatierenden Text an anderen Stellen nicht verwendet werden.

Steht zwischen zwei Kommandozeichen irgendetwas, das hinsichtlich Format, Größe der Parameter etc. kein legales Steuerkommando darstellt, so führt dies zu dem oben schon erwähnten COMMAND ERROR.

Die Steuerkommandos müssen immer am Anfang einer Zeile oder direkt vor einem Wort stehen; falls dies nicht eingehalten wird erscheint ebenfalls die Meldung COMMAND ERROR. Eine Ausnahme

bildet lediglich das Unterstreichungskommando, das überall im Text stehen darf.

Zwischen zwei Kommandozeichen dürfen beliebig viele Steuersymbole stehen; die einzelnen Kommandos müssen durch Kommas voneinander getrennt werden. Nicht mit anderen Kommandos kombiniert werden dürfen lediglich die Tabulatoranweisungen (S und T).

4.5.1 Neue Seite

Format: &P&

Die laufende Zeile wird (unformatiert) beendet; mit dem nachfolgenden Text wird eine neue Seite begonnen.

Befindet man sich dabei zufällig am Anfang einer Seite, so bleibt das Kommando &P& wirkungslos, es wird keine Leerseite erzeugt.

4.5.2 Neue Zeile

Format: &Ln&

Hierbei ist 'n' eine Zahl mit $n \geq 0$. Die laufende Zeile wird (unformatiert) beendet; mit dem nachfolgenden Text wird eine neue Zeile begonnen.

Ist die Anzahl der auf der laufenden Seite verbleibenden freien Zeilen kleiner als 'n', so wird eine neue Seite angefangen; für diesen Fall ist das Steuerkommando identisch mit &P&.

Befindet man sich gerade am Anfang einer Zeile, so wird keine Leerzeile erzeugt. Fehlt der Parameter 'n' ganz, so wird für n der Wert 0 angenommen, d.h. eine neue Zeile auf der laufenden Seite begonnen.

4.5.3 Verschiebung des linken Randes

Format: &Mn3

Hierbei ist 'n' eine Zahl mit $n \geq 0$. Ab der nächsten Zeile wird der linke Rand um n Positionen eingerückt.

Fehlt der Parameter 'n' ganz, so wird der Wert 0 angenommen, die nächste Zeile beginnt also wieder ganz links.

4.5.4 Verschiebung des rechten Randes

Format: &Nn&

Hierbei ist 'n' eine Zahl mit $n > 10$. Ab der nächsten Zeile beträgt die Gesamtbreite des formatierten Textes n Zeichenpositionen. Dabei zählen die Positionen, um die der linke Rand eingerückt ist, mit. Falls am Anfang des zu formatierenden Textes kein anderer Wert angegeben wird, beträgt die Anzahl der Zeichen pro Zeile 65.

Bei der Festlegung des linken und des rechten Randes ist darauf zu achten, daß die daraus resultierende Breite des Textes noch mindestens 10 Zeichenpositionen betragen muß.

4.5.5 Unterstreichen

Format: &U&

Durch dieses Steuerkommando werden Anfang und Ende eines Textabschnittes markiert, der einfach unterstrichen werden soll. Das bedeutet, daß von der ersten Anweisung &U& an der gesamte folgende Text bis zum erneuten Auftreten der Anweisung &U& unterstrichen wird.

4.5.6 Setzen von Tabulatorwerten

Format: &Sn1:n2:n3:n4:n5&

Hierbei muß erfüllt sein: $0 < n1 < n2 < n3 < n4 < n5 < \text{Zeichen/Zeile}$. Die eingegebenen Werte 'n1' bis 'n5' werden als Tabulatorwerte abgespeichert. Bis zu fünf Tabulatorwerte können definiert werden, es dürfen aber auch weniger angegeben werden.

4.5.7 Tabulator

Format: &T&

Die laufende Zeile wird unformatiert abgespeichert und bis zur nächstgrößeren Tabulatorposition (ausschließlich) mit Leerstellen aufgefüllt; das erste Zeichen des nachfolgenden Wortes steht dann also auf dieser durch den Tabulator angegebenen Position. Ist die Anweisung aus irgendeinem Grund nicht ausführbar (keine Tabulatorwerte definiert, kein Tabulatorwert mehr rechts von der momentanen Position, etc.), so erfolgt die Meldung COMMAND ERROR.

4.5.8 Sprung auf eine angegebene Position

Format: &Tn&

Die laufende Zeile wird unformatiert abgespeichert und bis zur Position 'n' (ausschließlich) mit Leerstellen aufgefüllt; das erste Zeichen des nächsten Wortes steht dann also auf der Position 'n'. Dies Kommando ist unabhängig von eventuell gesetzten Tabulatorwerten. Ist das Kommando nicht ausführbar, so erfolgt ebenfalls die Meldung COMMAND ERROR.

5 ANHANG
=====

A. getrimmter ASCII-Code

Code	Zeichen	Sonder- bedeutung	Code	Zeichen	Sonder- bedeutung
00	@	'CR'	40	'SPACE'	
01	A		41	!	
02	B		42	"	
03	C		43	#	ß
04	D		44	\$	'NP'
05	E		45	%	
06	F		46	&	'UNDER'
07	G		47	'	
10	H		50	(
11	I		51)	
12	J		52	*	'STOP'
13	K		53	+	
14	L		54	,	
15	M		55	-	
16	N		56	.	
17	O		57	/	
20	P		60	0	
21	Q		61	1	
22	R		62	2	
23	S		63	3	
24	T		64	4	
25	U		65	5	
26	V		66	6	
27	W		67	7	
30	X		70	8	
31	Y		71	9	
32	Z		72	:	
33	[Ä	73	;	
34	\	ö	74	<	'BACK'
35]	U	75	=	
36	↑	'UC'	76	>	-
37	←	'LC'	77	?	

ASCII = American Standard Code for Information Interchange

Die Sonderbedeutung der ASCII-Zeichen gilt für Ausgabe auf der Schreibmaschine und Benutzung des Programms 'TYPSET'.
Es bedeuten:

- 'CR' = Wagenrücklauf, Zeilenvorschub
- 'UC' = Umschaltung auf große Buchstaben
- 'LC' = Umschaltung auf kleine Buchstaben
- 'NP' = Beginn einer neuen Seite
- 'UNDER' = Anfang oder Ende der Unterstreichung
- 'STOP' = Stoppen der Ausgabe bis 'CTRL+G'
- 'BACK' = Rückschritt

