



ZUSE KG · BAD HERSFELD

Elektronische Rechenanlagen

Sortierverfahren und Probleme
des Magnetbandsortierens

Sortierverfahren und Probleme des Magnetbandsortierens

Inhaltsübersicht

	<u>Seite</u>
1.1 Einführung	
1.2 Zweck und Aufgabe der Sortierung	
1.3 Vorgang des Sortierens	
1.4 Vorbereitende Sortierung	
2. <u>Wichtige Faktoren des Sortierens</u>	2
2.1 Zahl der Magnetbandeinheiten	
2.2 Leistung der Magnetbandeinheiten	
2.3 Größe des Kernspeichers	
2.4 Rechengeschwindigkeit	
2.5 Größe der Informationseinheit	
2.6 Länge des Sortierbegriffs	
2.7 Zahl der Kanäle	
3. <u>Die Sortierverfahren</u>	4
3.1 Die Auswahl-sortierung	
3.2 Das Zähl-sortierverfahren	
3.3. Das Stellensortierverfahren	
3.4 Das Mischsortierverfahren	
3.5 Sortieren nach Adressen	
3.6 Die Einschaltungs- Sortiermethode	
3.7 Sortieren durch systemat. Adressenänderung	
3.8 Zusammenstellung der untersuchten Sortierverfahren	
4. <u>Vergleichende Betrachtungen</u>	23
4.1 Allgemeine Überlegungen	
4.2 Vergleich einiger Sortierverfahren	
5. <u>Zusammenfassung</u>	26
6. Literaturverzeichnis	28

1.1. Einführung

Mit folgenden Ausführungen soll ein Überblick über den Zweck, die Aufgabe und den allgemeinen Vorgang des Sortierens, sowie eine Übersicht über die wichtigsten Sortierverfahren gegeben werden.

Die wichtigsten Sortierverfahren werden jeweils durch ein Beispiel erläutert. Hierfür wurde eine Beispielfolge gewählt, mit der jeder Ablauf der einzelnen Sortierverfahren durchgeführt wurde.

1.2. Zweck und Aufgabe der Sortierung

In den verschiedensten Gebieten der Wirtschaft und Wissenschaft müssen oft Daten nach verschiedenen Ordnungsbegriffen geordnet werden.

Bei der Sortierung werden Datenmengen in eine solche Anordnung gebracht, daß sie danach in einer bestimmten Reihenfolge abgelesen werden können, um eine möglichst kurze Zugriffszeit für weitere Rechnungen zu erhalten.

1.2. Vorgang des Sortierens

Der Sortiervorgang umfaßt zwei Phasen:

- a) Es wird sortiert indem eine Reihe von Vergleichen durchgeführt wird.
- b) Die zu sortierenden Informationseinheiten werden entsprechend den Ergebnissen der Vergleiche umgeordnet.

Die so erhaltene Reihenfolge kann auf-, oder absteigend sein und eine Datenfolge ist dann geordnet, wenn sie keine Umkehrungen enthält (d.h. kein größeres Element vor einem kleineren bei aufsteigender Anordnung bzw. umgekehrt bei absteigender Anordnung).

1.4. Die vorbereitende Sortierung

Wenn die zu sortierenden Informationseinheiten zur Verarbeitung in die Rechenanlage eingelesen werden, hat es sich nach N. CHAPIN (1962) als zweckmäßig erwiesen, mit einer vorbereitenden Sortierung sogenannte "erste Ketten" herzustellen (worunter hier eine sortierte Gruppe von drei oder mehr Informationseinheiten verstanden wird). Solche ersten Ketten werden im 1. Durchgang hergestellt. In den nachfolgenden Durchgängen werden die Ketten in ihrer Länge vergrößert, zahlenmäßig aber vermindert, bis nur eine einzige Kette (=die vollständig sortierte Reihe der Informationseinheiten vorhanden ist.)

Nach K. KREUZER (1961) kann die erste Sortierung auf Lochkarten von konventionellen Sortiermaschinen vorgenommen werden, aber bei weiteren Auswertungen ergibt sich die Notwendigkeit, die Daten auf Magnetbänder zu speichern und mit elektronischen Rechenanlagen zu sortieren.

Nach N. CHAPIN (1962) können für einen 1. Durchgang (= vorbereitende Sortierung) die Sortierverfahren 3.1. bis 3.6. wirtschaftlich angewandt werden; für die weiteren Durchgänge empfiehlt sich (auch nach W. LAUTZ, 1963 und W. ZOBEBIER, 1960) die Anwendung des Misch-, und des Stellensortierverfahrens.

2.0. Einige wichtige Faktoren des Sortierens

Bei der Wahl eines bestimmten Sortierverfahrens für die Anwendung in einer elektronischen Rechenanlage sind mehrere Faktoren zu berücksichtigen, welche die Geschwindigkeit, Wirtschaftlichkeit und die Kosten des Sortierens wesentlich beeinflussen. Größere Datenmengen werden in der Praxis meist unter Zuhilfenahme von Magnetbändern sortiert.

2.1. Zahl der Magnetbandeinheiten

Die Sortierzeit sinkt mit der zunehmenden Zahl von Magnetbandeinheiten.

2.2. Leistung der Magnetbandeinheiten

- a) Ein Zeitgewinn durch mehr Magnetbandeinheiten wirkt sich erst bei größeren Sortiermengen entscheidend aus.
- b) Die Verwendung von Magnetbandeinheiten lohnt sich, wenn oft und viel sortiert werden muß.
- c) Man kann auf Grund der Übertragungsleistungen der Bandeinheiten keinen allgemein gültigen "prozentualen Gewinn" angeben, da die Verhältnisse je nach Sortierprogramm und Umfang der Arbeit verschieden sind.

2.3. Größe des Kernspeichers (innerer Speicher)

Ein größerer Kernspeicher erlaubt eine höhere Gruppierung der Sätze. Abgesehen von programmiertechnischen Vorteilen bietet ein größerer Kernspeicher beim Sortieren nur dann eine größere Steigerung der Geschwindigkeit, wenn ein ganzer Durchlauf durch verstärkte innere Sortierung erspart wird.

2.4. Rechengeschwindigkeit

Die Notwendigkeit einer hohen Rechengeschwindigkeit ergibt sich:

- a) bei der Verwendung von schnellen Magnetbandeinheiten und 2 Kanälen, da bei überlappendem Lesen, Verarbeiten und Schreiben die zur Verfügung stehende Programmzeit relativ kleiner geworden ist, und
- b) bei einer höheren Gruppierung, da mehr Sätze intern sortiert und damit mehr Vergleiche abgewickelt werden müssen.

2.1. Größe der Informationseinheit

Eine Informationseinheit ist eine Reihe von Worten, die auf Grund irgend welcher zwischen ihnen bestehenden Beziehungen als eine zusammengehörige Einheit behandelt werden.

Je länger die zu sortierenden Informationseinheiten sind, desto interessanter werden schnelle Magnetbandeinheiten und große Kernspeicher, da sehr viele größere Informationsmengen durch den Rechner bewegt werden müssen.

2.2. Länge des Sortierbegriffes

Der Sortierbegriff (auch Schlüssel genannt) ist ein Kennzeichen, das an einer bestimmten Stelle der Informationseinheit steht.

Die Sortierverfahren sind nur zum Teil von der Länge des Sortierbegriffes abhängig. Während beim Stellensortieren die Arbeitsmenge fast direkt proportional mit der Stellenzahl ansteigt, fällt dieser Faktor beim Bandsortieren (Mischsortieren) fast kaum ins Gewicht, da die Sortiermerkmale als Begriff verglichen werden.

2.7. Zahl der Kanäle

Wenn eine Maschine mit 2 Datenkanälen ausgerüstet werden kann, so ist das -- abgesehen von den normalen Arbeiten -- auch für das Bandsortieren von Vorteil, da das Einlesen eines neuen Blockes gleichzeitig mit dem Herausschreiben des in sich geordneten vorhergehenden ablaufen kann.

Unabhängig von der Gesamtzahl und der Anordnung der Informationseinheiten in einer Folge ermöglichen die elektronischen Rechanlagen ein rasches Sortieren.

Im Interesse einer raschen Sortierung muß es möglich sein, die Sortierbegriffe von der gesamten Informationseinheit zu lösen, die Sortierung allein mit den Begriffen vorzunehmen und nach erfolgter Sortierung den einzelnen Sortiermerkmalen die zugehörigen Informationseinheiten wieder zuzuordnen.

3.0. Die Sortierverfahren

3.1. DIE AUSWAHLSORTIERUNG

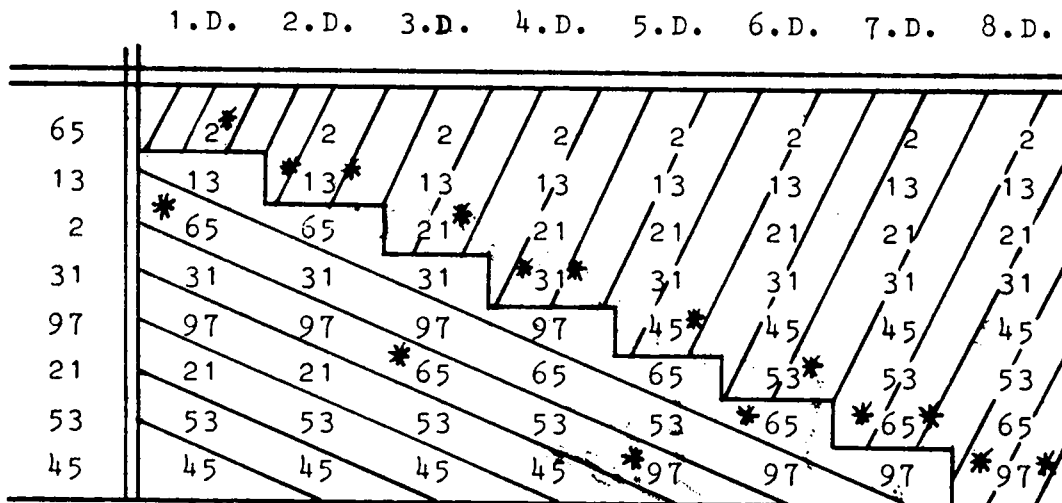
Nach N. CHAPIN (1962), O. Suppes (1963); Verfahren der Auswahl-sortierung, B. THÜRING (1957), Ermittlung einer bestimmten Informationseinheit unter vielen, W. ZOBEBIER (1960): Auswahl des kleinsten Sortiermerkmals.

3.1.1. Bei der einfachen Auswahl-sortierung werden zunächst alle zu sortierenden Elemente einer Folge miteinander verglichen, um in diesem 1. Durchgang die -- je nach Zweck und Ziel der Sortierung -- größte oder kleinste Zahl der Folge zu ermitteln. Dieser gefundene größte/kleinste Wert wird anschließend weggespeichert und von weiteren Vergleichen ausgeschlossen.

Aus der verbleibenden Restfolge wird dann in den nachfolgenden Durchgängen wieder die jeweils größte/kleinste Zahl ermittelt, bis die Sortierung beendet ist. Das Ergebnis ist dann die sortierte Folge.

Mit unserer Beispielfolge ergibt sich demnach folgender Sortierungsablauf:

Unsortierte Folge: 65,13,2,31,97,21,53,45.



die mit *() und ()* bezeichneten Zahlen tauschen jeweils die Plätze.



Sortierter Teil
der Folge



Unsortierter
Teil der
Folge

Zur praktischen Durchführung dieses Verfahrens verwendet W.ZOBERBIER (1960) 3 Magnetbänder:

- 1 Band mit unsortierten Werten (Eingabeband = EMB)
- 1 Band mit noch zu sortierenden Werten
(Sortiermagnetband = SMB)
- 1 Band mit sortierten Werten (Ausgabemagnetband = AMB)

Die Kennzahlen der einzelnen Informationseinheiten werden mit der bisher kleinsten (das ist am Anfang die Kennzahl der ersten Informationseinheit) verglichen. Ist die nächste vom EMB abgelesene Kennzahl (a_n) größer als die bisher kleinste (a_k), dann wird a_n auf SMB notiert; ist a_n kleiner als a_k , dann wird a_n zur "bisher kleinsten" und a_k wird auf SMB notiert.

Dieser Vorgang wird so lange wiederholt, bis alle Informationseinheiten vom EMB eingelesen sind.

Damit steht dann a_k (als kleinster Wert des jeweiligen Durchganges) fest, wird auf das AMB gespeichert und damit vom weiteren Vergleich ausgeschlossen.

Somit ist ein Durchlauf beendet und ein neuer kann beginnen.

Nach N. CHAPIN (1962) ergeben sich bei der praktischen Durchführung der Auswahl-sortierung mehrere Möglichkeiten und dementsprechend auch einige Arten dieses Verfahrens.

1) Sortieren mit Ersatz durch Null:

Diese weniger Zeit in Anspruch nehmende Methode ersetzt den Platz des bei jedem Durchlauf ermittelten größten/kleinsten Wertes in der zu sortierenden Folge durch Null. Dadurch wird aber die unsortierte Anordnung der Kennzahlen schrittweise zerstört, was manchmal unerwünscht ist.

3.1.2. Eine Methode der Auswahl-sortierung, die den übrigen Arten dieser Gruppe im allgemeinen überlegen ist, ist die "Auswahl höheren Grades".

Bei diesem Sortierverfahren wird die Folge der unsortierten Informationseinheiten in eine Anzahl willkürlicher (erster) Gruppen aufgeteilt, wobei die Anzahl der Informationseinheiten in jeder Gruppe möglichst gleich sein soll.

Der Grad dieser Methode ist gleich der Anzahl der Gruppierungen aus denen dann die Auswahl getroffen wird.

Die Programme werden mit dem Grad der angewandten Auswahlmethode

komplizierter; es hat sich daher die Auswahl 2. Grades als sehr vorteilhaft erwiesen:

Die Folge der Informationseinheiten wird in 2 erste Gruppen eingeteilt. Es wird dann eine einfache Auswahl aus den Schlüsselzahlen jeder ersten Gruppe getroffen, um das Element mit der größten (oder kleinsten) Zahl innerhalb jeder (ersten) Gruppe zu ermitteln. Diese ausgewählten Schlüsselzahlen bilden dann eine neue (zweite) Gruppe, aus der wieder eine gewöhnliche Auswahl getroffen wird. Der nun ausgewählte Schlüssel wird in die sortierte Liste übertragen und der freigewordene Platz in der zweiten Gruppe wird durch ein Element aufgefüllt, das durch eine einfache Auswahl aus jener (ersten) Gruppe gefunden wird, aus der das auf die sortierte Liste übertragene Element stammt. Diese aufgefüllte (zweite) Gruppe wird wieder einer einfachen Auswahl unterzogen, der Zyklus wiederholt sich.

Mit der Beispielfolge 65, 13, 2, 31, 97, 21, 53, 45
ergibt sich dann folgendes Sortieren 2. bzw. 3. Grades.

Es wird nach absteigender Reihenfolge sortiert.

a) Sortierung 2. Grades:

vergleiche	65, 13, 2, 31	(= 1.Gruppe) →	65	} → <u>97</u>
"	97, 21, 53, 45	(= 2.Gruppe) →	97	

vergleiche			65	} → <u>65</u>
"	21, 53, 45		+ 53	

"			31	} → <u>45</u>
"	21, 45		+ 45	

"			31	} → <u>31</u>
"	21		+ 21	

"			21	} → <u>21</u>
"	13, 2		+ 13	

"			13	} → 13
"	2		+ 2	

+ 2

b) Sortierung 3. Grades:

Gruppen der unsortierten Folge		1. Durchgang 1. Ausw. 2. Aus.	2. Durchgang 1. Aus. 2. Aus.	3. Durchgang 1. Aus. 2. Aus.
Gruppe	Sort. Begr.			
1	65 13 2	65	65	13
2	31 97 21	97 97	31 65	31 53
3	53 45 0	53	53	53

Gruppe	4. Durchgang 1. Aus. 2. Aus.	5. Durchgang 1. Aus. 2. Aus.	6. Durchgang 1. Aus. 2. Aus.	7. Durchgang 1. Aus. 2. Aus.
1	13	13	13	13
2	31 45	31 31	21 21	0 13
3	45	0	0	0

Gruppe	8. Durchgang 1. Aus. 2. Aus.
1	2
2	0
3	0

97, 65, 53, 45, 31, 21, 13, 2
=====

3.2. DAS ZÄHLSORTIERVERFAHREN
=====

(nach N. CHAPIN, 1962)

Bei diesem Sortierverfahren wird jede Schlüsselzahl der Informationseinheit mit sich selbst und allen anderen der Kennzahlenfolge verglichen. Jedem zu sortierenden Element wird ein Zähler zugeordnet. Zu Beginn des Verfahrens werden alle Zähler auf Null gestellt.

Bei jedem Vergleich wird dann zum entsprechenden Zähler eine "1" addiert wenn die dazugehörige Schlüsselzahl größer oder gleich der anderen ist.

Mit unserer eingangs angegebenen Beispielfolge von 8 Elementen ergibt sich dann folgendes Schema:

Die unsortierte Folge der Schlüsselzahlen:

65, 13, 2, 31, 97, 21, 53, 45.

Schlüssel	65	13	2	31	97	21	53	45	Aufsteigende Rangordnung
65	1	1	1	1		1	1	1	7
13		1	1						2
2			1						1
31		1	1	1		1			4
97	1	1	1	1	1	1	1	1	8
21		1	1			1			3
53		1	1	1		1	1	1	6
45		1	1	1		1		1	5
Absteigende Rangordnung	2	7	8	5	1	6	3	4	

Nach beendeter Zählung erhält man dann die sortierte Folge --
je nach Zweck und Wahl -- in auf-, oder absteigender Reihen-
folge:

	<u>Aufsteigende Reihenfolge</u>							
Ziffer des Zählers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2	13	21	31	45	53	65	97

Ziffer des Zählers	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
	<u>Absteigende Reihenfolge</u>							

3.3. DAS STELLENSORTIERVERFAHREN
=====

(= BASIC-SORTING)

Die verschiedenen Bezeichnungen dieses Verfahrens
in der Literatur:

W. DE BEAUCLAIR (1961) : Sortieren durch Verteilen
(nach Dezimalstellen) : Mehrfaches Vergleichen:
Umlaufsortieren.

K. KREUZER (1961) : Verfahren der Stellensortierung
oder Verteilsortierung.

W. LAUTZ (1963) : Stellensortierverfahren

B. THÜRING (1957) : Ziffernsortiermethode

W. ZOBEBIER (1960) : Schubfachsortieren - Spurensortieren

Dieses Verfahren sortiert die Informationsbegriffe nach den Stellenwerten. Beginnend mit der niedrigsten Ziffer des niedrigsten Stellenwertes werden die Sortierbegriffe dementsprechend von 0, 1, ... 8,9 (bzw. 9, 8, ... 1, 0) geordnet und in dieser Reihenfolge umgespeichert. Dann erfolgt der gleiche Vorgang mit den so im ersten Durchgang vorsortierten Begriffen nach den Ziffern des nächsthöheren Stellenwertes, usw., bis der höchste Stellenwert erreicht ist.

Die Beispielfolge 65,13,2,97,21,53,45 soll nach aufsteigender Reihenfolge geordnet werden.

Zunächst wird nach den Ziffern der Einerstelle sortiert:

0	1	2	3	4	5	6	7	8	9
	31	02	13		65		97		
	21		53		45				

Nach dem ersten Durchgang sieht die gespeicherte Folge wie folgt aus:

31, 21, 2, 13, 53, 65, 45, 97

Jetzt erfolgt die Verarbeitung nach den Ziffern 0 bis 9 der Zehnerstelle:

0	1	2	3	4	5	6	7	8	9
02	13	21	31	45	53	65			97

Damit steht als Ergebnis die sortierte Folge 2,13,21,31,45,53,65,97 fest.

ALLGEMEIN:
=====

Ist der Sortierbegriff k-stellig, dann sind k Durchläufe nötig um die Daten zu sortieren; die Anzahl der Durchläufe ist also unabhängig von der Datenmenge.

Das Stellensortieren nach Dezimalziffern wird hauptsächlich bei der Lochkartensortierung angewendet. Beim Sortieren mit Rechenanlagen ist es günstiger, nach dem Dualsystem zu sortieren, da hier nur die Ziffern 1 und 0 auftreten. Ziffernsortieren im Dualcode hat darüber hinaus noch den Vorteil, daß jede einzelne Dualziffer für die Abfrage durch Intersektion aus dem ganzen Sortierbegriff herausgeschnitten werden kann.

Das Stellensortierverfahren arbeitet optimal, wenn jeder Ziffer ein eigenes Ausgabemagnetband zugeordnet werden kann; zuzüglich dem Eingabemagnetband ergibt sich dann im Dezimalcode eine Anordnung von 21 (ohne Umspulen) bzw. 11 (mit Umspulen) Bändern. Dementsprechend im Dualcode 5 bzw. 3 Bänder.

In der praktischen Anwendung werden im Dezimalcode meist 6 Magnetbänder (1 Eingabe-, 5 Ausgabebänder) verwendet, wobei 2 Sortierziffern auf ein Ausgabeband kommen.

3.4. DAS MISCHSORTIERVERFAHREN

(TAPE - MERGING)

Literaturangaben und verschiedene Bezeichnungen des Mischsortierverfahrens:

N. CHAPIN (1962), W. DE BEAUCLAIRE (1961),
K. KREUZER (1961), W. LAUTZ (1963),
O. SUPPES (1963), B. THÜRING (1957 : 2-, bzw. 3-Listen-Collation)
W. ZOBEBIER (1960 : Gruppensortieren oder Mischen),
Lit. Verz. Nr. 13 und 14.

Beim Mischsortierverfahren wird die gewünschte Reihenfolge durch wiederholtes "Mischen" der Bänder hergestellt.

Ablauf des Mischsortierens:

Vor dem ersten Durchgang werden die unsortierten Werte alternierend auf 2 Bänder aufgeteilt (wenn sie nicht schon in dieser Form vorliegen).

Im 1. Durchgang werden jeweils 2 Werte miteinander verglichen und kommen in der richtigen Reihenfolge (auf-, oder absteigend) als 2er-Gruppen alternierend auf 2 Ausgabebänder.

Im 2. Durchgang werden jeweils 2 dieser 2er-Gruppen miteinander verglichen und zu einer 4er-Gruppe "gemischt" (d.h. geordnet) und wie oben auf die Ausgabebänder gespeichert.

Im 3. Durchgang entsteht aus zwei 4er-Gruppen eine 8er-Gruppe usf., bis das Ergebnis in einer geordneten Folge vorliegt.

Mit der Beispielfolge 65, 13, 2, 31, 97, 21, 53, 45 ergibt sich dann folgendes Bild:

<u>1. Durchgang:</u>	Band III	Bd IV	<u>2. Durchgang:</u>	BdI	BdII
(Vergleich	65 ↔	13	(Vergleich	13	} ↔ { 2
liefert 2.	2 ↔	31	liefert 3.	65	
Durchgang)	97 ↔	21	Durchgang)	21	} ↔ { 45
	53 ↔	45		97	

3. Durchgang

BdIII		Bd IV
2	} ↔	21
13		45
31		53
65		97

Nach dem Vergleich im 3. Durchgang werden die Werte in der richtigen Reihenfolge 2,13,21,31,45,53,65,97 auf Band I ausgegeben.

Für das Mischsortierverfahren sind mindestens 4 Magnetbandgeräte erforderlich (2-Weg-Mischsortieren).

Zwei dieser vier Bandgeräte enthalten die Bänder mit den unsortierten Daten, die anderen zwei dienen als Aufnahmegerät für die Ergebnisse des 1. Durchganges. Für die nachfolgenden Durchgänge erfolgt dann wechselseitig die Umkehrung dieses Vorganges.

Zur Verminderung der Anzahl der Durchläufe kann man auch 6, 8, usw. Magnetbandeinheiten verwenden und spricht dann von 3-, 4-Weg-Sortieren, usw.

Das 2-, 3-Weg-Mischsortieren entspricht der 2-, 3-Listen-Collation (nach B.THÜRING, 1957).

Werden mehr als 4 Geräte benützt, dann werden die Regeln für den Vergleich immer komplizierter, ein Gewinn ergibt sich aber für die Geschwindigkeit des Sortierens.

Ist k ...Anzahl der Magnetbandeinheiten
N... " " Sortierbegriffe,

dann sind $n = (k \log N) + 1$ Durchläufe nötig um die N Werte zu sortieren; das heißt die Anzahl der Durchläufe ist nur von der Anzahl der Sortierbegriffe abhängig und nicht von deren Länge (es werden ja immer die ganzen Sortierbegriffe verglichen, nicht die einzelnen Stellen der Worte).

Daraus ergibt sich eine gute Eignung für das Sortieren von alphanumerischen Kennworten.

3.5. SORTIEREN NACH ADRESSEN
=====

(=FUNCTION TABLE SORT)

(nach B. THÜRING, 1957)

Sind die Informationseinheiten klein, d.h. nur wenige Worte umfassend, dann kann das Sortieren nach Adressen angewendet werden, wenn folgende Voraussetzungen erfüllt sind:

1. Der Sortierbegriff (die Kennzahl) muß rein numerisch sein,
2. es dürfen keine Informationseinheiten mit gleicher Kennzahl vorhanden sein.

Wenn Z_i die Kennzahl (der Sortierbegriff),
 n die Anzahl der Worte einer Informationseinheit,
 m_0 (irgend) eine Anfangsadresse, und
 m_i die gesuchte Sortieradresse

ist, dann gilt:

$$m_i = m_0 + n \cdot Z_i$$

Für die Beispielfolge 65,13,2,31,97,21,53,45 ergibt sich folgender Ablauf:

$m_0 = 2000$			
$n = 1$		(ein Wort je Informationseinheit)	
$m_1 = 2000+65 = 2065$	dann ist:	2002	= 2
$m_2 = 2000+13 = 2013$		2013	=13
$m_3 = 2000+ 2 = 2002$		2021	=21
$m_4 = 2000+31 = 2031$		2031	=31
$m_5 = 2000+97 = 2097$		2045	=45
$m_6 = 2000+21 = 2021$		2053	=53
$m_7 = 2000+53 = 2053$		2065	=65
$m_8 = 2000+45 = 2045$		2097	=97

3.6. DIE EINSCHALTUNGS - SORTIERMETHODE

=====

(nach N. CHAPIN, 1962)

Für jeden Schlüssel (=Sortierbegriff der Informationseinheit) wird das Verhältnis dieses Schlüssels zum größten auftretenden Zahlenwert in der zu sortierenden Folge der Sortierbegriffe berechnet. Mit diesem Verhältnis wird dann die Lage des Schlüssels innerhalb der Speichereinheit bestimmt. Wenn die durch das so errechnete Verhältnis bezeichnete Adresse bereits besetzt ist, werden die Elemente so lange verschoben, bis das neue Element eingeschaltet werden kann. Das Ergebnis ist eine Anordnung in sortierter Reihenfolge.

Für unsere zu sortierenden Schlüssel ergibt sich folgende Anordnung:

a) Annahme:

<u>Unsortierte Folge</u>	65, 13, 2, 31, 97, 21, 53, 45
<u>Länge der Kette</u>	8 Informationseinheiten
<u>Zahlenbereich des Schlüssels</u>	0 bis 99
<u>Verfügbare Adressen</u>	2000 bis 2019 (= 20)

Berechnung:

$$\frac{65}{100} = 0,65; \quad \frac{13}{100} = 0,13; \quad \frac{2}{100} = 0,02; \quad \frac{31}{100} = 0,31;$$
$$\frac{97}{100} = 0,97; \quad \frac{21}{100} = 0,21; \quad \frac{53}{100} = 0,53; \quad \frac{45}{100} = 0,45;$$

	Einschaltung
(0,65) · (20) = 13,0 2000 + 13	65 in die Adresse 2013
(0,13) · (20) = 2,6 2000 + 3	13 " " " 2003
(0,02) · (20) = 0,4 2000 + 0	2 " " " 2000
(0,31) · (20) = 6,2 2000 + 6	31 " " " 2006

$(0,97) \cdot (20) = 19,4$	
2000+19	97 in die Adresse 2019
$(0,21) \cdot (20) = 4,2$	
2000+4	21 " " " 2004
$(0,53) \cdot (20) = 10,6$	
2000+11	53 " " " 2011
$(0,45) \cdot (20) = 9,0$	
2000+9	45 " " " 2009

3.7. SORTIEREN DURCH SYSTEMATISCHE ADRESSENÄNDERUNG
 =====
(nach B. THÜRING, 1957)

Diese Art der Sortierung eignet sich für Informationseinheiten, die zwar aus vielen Worten bestehen aber in relativ geringer Menge vorhanden sind.

Besteht jede Informationseinheit aus m Worten und stehen die Kennzahlen der einzelnen Informationseinheiten auf den Adressen A_1, A_2, \dots , dann soll $A_2 = A_1 + m, A_3 = A_2 + m, \dots$ sein (d.h. die Informationseinheiten sind lückenlos gespeichert.)

Die nach Kennzahlen geordneten Informationseinheiten sollen auf die Adressen $B_1, B_2=B_1+m, B_3=B_2+m \dots$ -- die im Laufe der Rechnung noch verändert werden -- gelangen. Die Adresse B_1 kann innerhalb des verfügbaren Speicherraumes gewählt werden. Die Adressen B_i ($i = 1, 2, \dots n$) werden abhängig von einem systematischen Größenvergleich der Kennzahlen folgendermaßen umgeformt:

(wobei $\langle A_i \rangle =$ "Inhalt von A_i " ist.)

1. Durchlauf:

$\langle A_2 \rangle < \langle A_1 \rangle ?$	<u>ja:</u> $B_1 + m \Rightarrow B_1; B_2 - m \Rightarrow B_2$
		<u>nein:</u> keine Änderung von B_1, B_2
$\langle A_3 \rangle < \langle A_1 \rangle ?$		<u>ja:</u> $B_1 + m \Rightarrow B_1; B_3 - m \Rightarrow B_3$
		<u>nein:</u> keine Änderung von B_1, B_3
$\langle A_4 \rangle < \langle A_1 \rangle ?$	usw.
	.	
	.	
	.	
	.	

A_1 ist der Vergleichspartner für alle folgenden A_i ($i=2,3,\dots,n$)

2. Durchlauf:

$\langle A_3 \rangle < \langle A_2 \rangle$? ja: $B_2 + m \Rightarrow B_2$; $B_3 - m \Rightarrow B_3$
nein: keine Änderung von B_2, B_3

$\langle A_4 \rangle < \langle A_2 \rangle$? usw.

A_2 ist der Vergleichspartner für alle folgenden A_i

($i = 3, 4, \dots, n$)

usw.

Es wird jeweils die dem pro Durchgang konstanten Vergleichspartner (1. Durchgang: A_1 , 2. Durchgang: A_2 , usw., bis $(n-1)$). Durchgang : $A(n-1)$ zugeordnete Adresse B um m erhöht, wenn der im jeweiligen Durchgang variable Vergleichspartner (1. Durchgang: A_2, A_3, \dots 2. Durchgang: $A_3, A_4, \dots, (n-1)$. Durchgang : A_n) kleiner ist als der konstante. Gleichzeitig wird die Adresse B des variablen Partners um m vermindert. Ist der variable Partner größer als der konstante, dann werden beide B-Adressen unverändert gelassen.

Für die Beispielfolge 65, 13, 2, 31, 97, 21, 53, 45
gelangt man zu folgenden Schema: (wobei $m=1$ ist)

65	in	Adresse	A_1	soll	sortiert	nach	B_1	(mit	Anfangsadresse	501)
13	"	"	A_2	"	"	"	B_2	("	502)
2	"	"	A_3	"	"	"	B_3	("	503)
31	"	"	A_4	"	"	"	B_4	("	504)
97	"	"	A_5	"	"	"	B_5	("	505)
21	"	"	A_6	"	"	"	B_6	("	506)
53	"	"	A_7	"	"	"	B_7	("	507)
45	"	"	A_8	"	"	"	B_8	("	508)

1. Durchlauf:

13 < 65	B_1	=	$501 + 1 = 502$,	B_2	=	$502 - 1 = 501$
2 < 65	B_1	=	503,	B_3	=	502
31 < 65	B_1	=	504,	B_4	=	503
97 > 65		--			--	
21 < 65	B_1	=	505,	B_6	=	505
53 < 65	B_1	=	506,	B_7	=	506
45 < 65	B_1	=	507,	B_8	=	507

2. Durchlauf:

2 < 13	B_2	=	$501 + 1 = 502$,	B_3	=	$502 - 1 = 501$
31 > 13		--			--	
97 > 13		--			--	
21 > 13		--			--	
53 > 13		--			--	
45 > 13		--			--	

3. Durchlauf:

31 > 2		--			--
97 > 2		--			--
21 > 2		--			--
53 > 2		--			--
45 > 2		--			--

4. Durchlauf:

97 > 31 -- --
21 < 31
53 > 31 $B_4 = 503 + 1 = 504, B_6 = 505 - 1 = 504$
45 > 31 -- --

5. Durchlauf:

21 < 97 $B_5 = 505 + 1 = 506, B_6 = 504 - 1 = 503$
53 < 97 $B_5 = 507, B_7 = 505$
45 < 97 $B_6 = 508, B_8 = 506$

6. Durchlauf:

53 > 21 -- --
45 > 21 -- --

7. Durchlauf:

45 < 53 $B_7 = 505 + 1 = 506, B_8 = 506 - 1 = 505$

A₁ kommt nach B₁ = 507
A₂ " " B₂ = 502
A₃ " " B₃ = 501
A₄ " " B₄ = 504
A₅ " " B₅ = 508
A₆ " " B₆ = 503
A₇ " " B₇ = 506
A₈ " " B₈ = 505

d.h. in den Speichern 501 bis 508 steht unsere sortierte Folge
2, 13, 21, 31, 45, 53, 65, 97.

3.8. ZUSAMMENSTELLUNG DER UNTERSUCHTEN SORTIERVERFAHREN
=====

- 3. ① AUSWAHLSORTIEREN
- 3. ② ZÄHLSORTIERVERFAHREN
- 3. ③ STELLENSORTIERVERFAHREN
- 3. ④ MISCHSORTIERVERFAHREN
- 3. ⑤ SORTIEREN NACH ADRESSEN
- 3. ⑥ EINSCHALTUNGS + SORTIERMETHODE
- 3. ⑦ SORTIEREN DURCH SYSTEMAT.-ADRESS.-ÄNDG.

Auswahlsortieren

Einschaltungssortiermethode

Gruppensortieren	siehe bzw. =	Mischsortierverf.
2-,3-, ... Listen-Collation	" "	"
Mehrfaches Vergleichen	" "	Stellensortierverf.

Mischsortierverfahren

Schubfachsortieren siehe bzw. = Stellensortierverf.

Sortieren nach Adressen

Sortieren durch systemat. Adr.-Änderg.

Stellensortierverfahren

Umlaufsortieren	siehe bzw. =	Stellensortierverf.
Verf. d. Stellensortierens	" "	"
Verteilsortierg.	" "	"

Zählsortierverfahren

Ziffernsortiermethode siehe bzw. = Stellensortierverf.

4. Vergleichende Betrachtungen

4.1. Allgemeine Überlegungen

Wenn (nach W.DE BEAUCLAIR, 1961) große Mengen von Daten in elektronischen Anlagen verarbeitet werden sollen, so lassen sie sich am zweckmäßigsten auf Magnetbänder speichern. Magnetbänder stehen wirtschaftlich zwischen den Lochkarten und den Kernspeichern; erstere sind im allgemeinen zu langsam, letztere bei genügender Kapazität zu teuer.

Auch nach K. KREUZER (1961) ergibt sich die Notwendigkeit, die auf Magnetband gespeicherten Daten zu sortieren. Ein Herausstanzen in Lochkarten und ein Sortieren derselben mit Lochkartensorten ist nicht nur meistens unwirtschaftlich und zeitraubend, sondern unter Umständen auch gar nicht möglich, wenn die 80 Stellen der Lochkarte überschritten werden.

Ein weiteres wichtiges Argument für die Bandsortierung ist (nach K.KREUZER, 1961) die große Zuverlässigkeit der Magnetbänder. Bei einer sehr großen Anzahl von Lochkarten können bei längerer Bearbeitung leichter Bedienungsfehler auftreten.

Nach W.ZOBERBIER (1960) berücksichtigt man bei der Wahl eines Verfahrens zum Sortieren von Informationen mittels elektronischer Rechenanlagen wesentlich die hohe Rechengeschwindigkeit solcher Anlagen. Für diesen Zweck sind vielfach die klassischen Verfahren des Lochkartensortierens ungünstig, und man verwendet besser eines der neueren Vergleichsverfahren. Ist jedoch eine große Anzahl von Informationen zu verarbeiten, für die Speicher sehr großer Kapazität benötigt werden, so wählt man zweckmäßigerweise Magnetbänder als Speicher.

4.2. Vergleich einiger Sortierverfahren

Zu den Auswahlverfahren höheren Grades ergibt sich der Hinweis, daß die Anzahl der Durchgänge, die für das Sortieren einer Kette bestimmter Länge erforderlich ist, gleich der Länge dieser Kette ist, genau wie im Fall der einfachen Auswahl. Die Anzahl der Vergleiche pro Durchgang ist aber nur im 1. Durchgang die gleiche wie bei der einfachen Auswahl. In allen anderen Durchgängen ist die Anzahl der Vergleiche wesentlich reduziert, da die gewöhnliche Auswahl ja nur auf jeweils eine der "ersten" Gruppen und dann immer auf die "zweite" Gruppe angewendet wird. Das Ergebnis ist also eine raschere Sortierung als beim ersten Durchgang.

Für die Zähl-sortiermethode gilt (nach N. CHAPIN, 1962) : sie ist äußerst zweckmäßig, wenn die Anzahl der Elemente klein ist.

Nach Ansicht der Verfasser ist die Einschaltungs-Sortiermethode-- im Gegensatz zu N. CHAPIN, 1962 -- nicht nur äußerst kompliziert, sondern es ist überhaupt fraglich, ob sie bei einer großen Zahl von Daten reibungslos arbeitet.

Für das Sortieren nach Adressen, bzw. das Sortieren durch systematische Adressenänderung ergaben sich aus der Literatur keine Vergleichsmöglichkeiten.

Das Mischsortierverfahren wird hauptsächlich dann mit Erfolg angewandt, wenn eine große Menge von Daten zu sortieren ist. Dieses Verfahren verlangt allerdings auch eine leistungsfähige elektronische Rechanlage deren Innenspeicherkapazität für die Sortiergeschwindigkeit mitbestimmend ist.

Die Wirtschaftlichkeit des Mischsortierens zeigt sich um so stärker, je länger die zu sortierenden Folgen sind. Aus diesem Grund wird die Mischsortierung meist nicht für einen ersten Durchgang verwendet, sondern mehr für die nachfolgenden Durchgänge, wenn schon längere Ketten (vorsortierte Folgen, siehe auch 1.4.) vorhanden sind. Sind mehr als 4 Ein-, und Ausgabe-geräte vorhanden (siehe auch 3.4.) dann wird wohl das Ablaufprogramm komplizierter, aber es ergibt sich ein Gewinn für die Sortiergeschwindigkeit.

Die Dauer des Sortiervorganges ist beim Mischsortieren unabhängig von der Anzahl der Ziffern eines Sortierbegriffes. Es entscheidet nur die Anzahl der ganzen Sortierbegriffe, die miteinander verglichen werden. Es kann daher diese Art der Sortierung auch für ein Ordnen von alphanumerischem Text verwendet werden.

Das Stellensortierverfahren ist (nach W. LAUTZ, 1963) das auf Magnetbänder übernommene Prinzip des (mechanischen) Lochkartensortierens.

Wird jeder Stelle ein Magnetband zugeordnet, dann arbeitet dieses Verfahren optimal.

Entscheidend für die Dauer des Sortiervorganges ist nur die Stellenzahl des Sortierbegriffes, und nicht (wie beim Mischsortieren) die Gesamtzahl der Begriffe.

Nach W. ZOBEBIER, (1960) kann man nach folgendem Kriterium entscheiden ob das Misch-, oder besser das Stellensortierverfahren anzuwenden ist:

ist n die Anzahl der Durchläufe beim Stellensortieren,
 k die Anzahl der Magnetbandeinheiten
 N die Zahl der Sortierbegriffe,

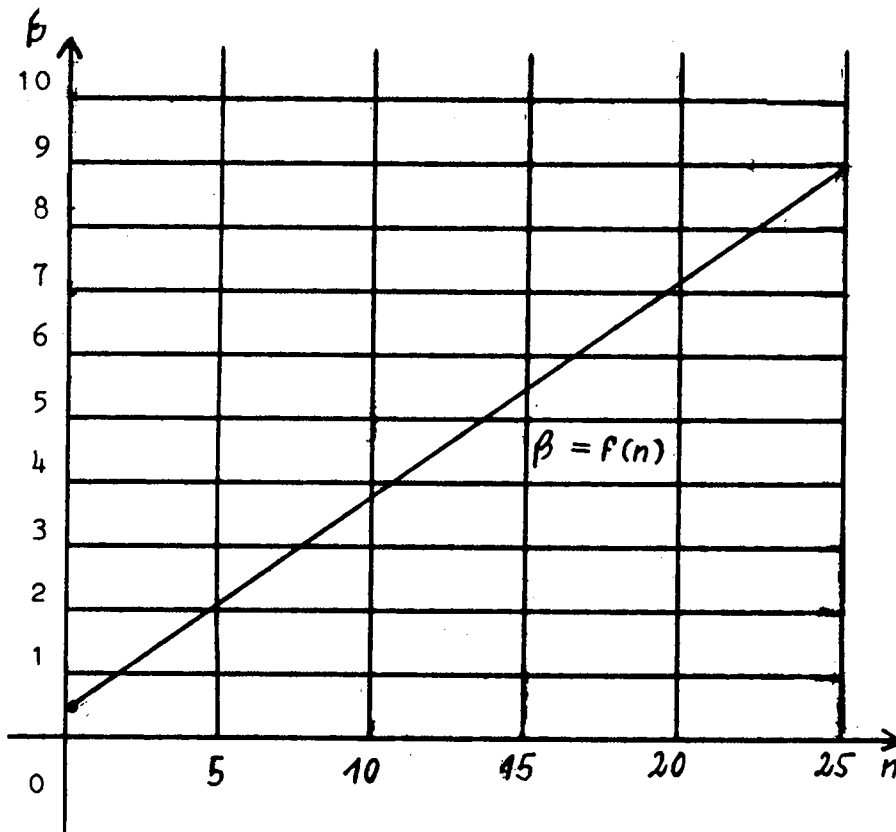
dann gilt für n folgende Beziehung:

$$n = (k \log N) + 1 .$$

Für $\beta = f(n)$ (wobei β = maximale Stellenzahl des Sortiermerkmals) hat W. ZOBEBIER (1960) das folgende Diagramm entworfen.

Hieraus kann man das jeweils günstigste Sortierverfahren (Misch-, oder Stellensortieren) auswählen und geht dabei wie folgt vor:

zunächst wird der durch die Werte β (ist bekannt) und n (wird nach obiger Formel ausgerechnet) festgelegte Punkt $P(n/\beta)$ eingetragen. Liegt dieser unterhalb der Geraden $\beta = f(n)$, dann ist das Stellensortieren wirtschaftlicher; liegt aber der Punkt auf oder über der Geraden, dann verwendet man besser das in diesem Fall günstigere Mischsortieren.



5. Zusammenfassung

In der vorliegenden Arbeit geben die Verfasser -- zunächst vorwiegend nach eingehendem Literaturstudium -- einen Überblick über die Sortierverfahren und ihre Probleme.

Nach einer Einführung in das Sortieren werden die untersuchten Sortiermethoden beschrieben und jeweils durch ein Beispiel erläutert.

Abschließend folgen dann vergleichende Betrachtungen zu allgemeinen und speziellen Problemen des Sortierens bzw. der Sortierverfahren.

Hieraus ergibt sich, daß das Sortieren mit elektronischen Rechenanlagen, besonders unter Verwendung von Magnetbandgeräten, einer Sortierung mit Lochkarten vorzuziehen ist.

Je nach Größe des Sortiermerkmals und der Anzahl der zu verarbeitenden Daten empfiehlt sich die Anwendung eines der drei folgenden Sortierverfahren:

Die Auswahl-sortierung 2. Grades
das Stellensortierverfahren (Basic-Sorting)
das Mischsortierverfahren (TAPE-Merging).

Für die beiden letzten Sortierverfahren ergeben sich folgende Gesetzmäßigkeiten:

Bei größerer Stellenzahl des Sortiermerkmals und einer kleineren Anzahl von notwendigen Durchläufen eignet sich das Mischsortieren besser, bei kleinerer Stellenzahl und größerer Durchlaufzahl wendet man besser das Stellensortierverfahren an.

6.0. LITERATURVERZEICHNIS

=====

- CHAPIN, N (1962): Einführung in die elektronische Datenverarbeitung, Abschn. 7.7., s. 109 ff.
Verlag R. Oldenbourg, Wien, 1962
①
- COX, B. und GOLDBERG, J.(1956): A Magnetic Sorting System, in
IRE National Convention Record 4,
Teil 4 (1956), s 101 ff.
②
- DAVIES, D.W. (1956):Sorting of Data on an Electronic Computer, in
Proc. Instr. Electr. Engrs. 103
Past B Suppe (1956), H I, s. 87 ff.
③
- DE BEAUCLAIR (1961):Das Sortieren von Magnetband-Daten an einfachen Buchungsanlagen, in
Elektr. Rechenanlagen (1961), s. 75 ff,
④
- FRIEND, E.H. (1956):Sorting on Electronic Computer Systems, in
Journal ACM3 (Juli 1956) H.3, s. 134 ff.
⑤
- HOSKEN, J. C.(1955):Evaluation of Sorting Methods, in
Proc. Eastern Joint Comp. Conf.,
Boston (1955), s. 39 ff.
⑥
- ISAAC, E.J. und SINGEETON, RC. (1956): Sorting by Adress Calculation, in Journal FICM 3 (Juli 1956),
H3, s. 169 ff.
⑦
- KREUZER, (1961): Sortieren mit datenverarbeitenden Anlagen,
in Elektronische Rechenanlagen, (1961),
s. 7 ff.
⑧
- LAUTZ, W. (1963): Sortierverfahren für technische Dual-Computer, in Elektron. Datenverarbeitung,
H. 2 (1963) s. 69 ff. (Teil 1) u.
H. 3 (1963) s. 133 ff. (Teil 2).
⑨
- SUPPES, O (1963): Kleiner Programmierungskursus,
(3. Folge)
in ZUSE-Forum, Nr. 4
⑩

THÜRING, B. (1957): Methoden der Programmierg. Kfm. v. Wiss.Probl. f. elektron. Rechenanlagen.-1.Teil: Logik der Programmierung.
Göller-Verlag, Baden-Baden, (1957).

⑪

ZOBERBIER, W.(1960): Vergleichende Betrachtungen zum Magnetband-sortieren, in Elektron. Datenverarbeitung, 1960, F. 5, s. 28 ff.

⑫

Ohne Verfasser-
Angabe Sorting methods and two commercial applications of the EMIDEC 1100, in Elektron. Datenverarbeitung, Nr. 3, 1962, s. 121 ff.

⑬

Was bedeutet Mischsortieren ? in BTA (Zeitschr. f. Büro- u. Verwaltungs.Auto-matisierung), 4. J., H I, (1963) s. 1

⑭

ANMERKUNGEN ZUM LITERATURVERZEICHNIS
=====

Die mit gekennzeichneten Arbeiten haben den Verfassern vorgelegen.

Die mit gekennzeichneten Arbeiten konnten in der Kürze der Zeit nicht beschafft werden (Ausleihe aus der Landesbibliothek). Sie sind aber in ZOBERBIER, W. (1960) verarbeitet worden.