



EDI

Electronic
Data
Interchange
(Elektronischer Datenaustausch)



“Exkurse”

Prüfsummenverfahren
RFID, GTAG
Efficient Consumer Response (ECR):
Überblick



Prüfsummenverfahren

Eingabefehler
Die Verhoeff'sche Fehlerstatistik
Prüfsummenverfahren
Vergleichende Bewertung



Eingabefehler



- Manuelle Eingabefehler
 - Bereits bei 6-stelligen Zahlen relevant
 - Manuelle EAN-Erfassung (12/13 Stellen)?
- Scannerfehler
 - Barcodeleser arbeiten nicht fehlerfrei
 - Barcodes können beschädigt werden
- OCR-Verfahren
 - Neuartige Erfassungsfehler mit eigener Charakteristik
 - Typ "Einzelfehler", ca. 1%



Klassische Arbeit zum Thema:

J. Verhoeff, Error-Correcting Decimal Codes, Mathematical Centre Tracts, Vol. 29, Mathematisch Centrum, Amsterdam 1969

J. Verhoeff untersuchte die Art und relative Häufigkeit von Fehlertypen bei der Eingabe beliebiger sechsstelliger Zahlen

Er veröffentlichte eine Fehlerstatistik und ersann Prüfziffer-Verfahren zur Minimierung der von ihm ermittelten häufigsten Fehlerarten

Analoga zu Prüfziffer-Verfahren in der Informatik:
CRC-, Parity-, ECC-Verfahren der Digitaltechnik.



Fehlerart	Symbol	Häufigkeit/%
Einzelfehler	$x \Leftrightarrow y$	79,0
Nachbarvertauschung	$xy \Leftrightarrow yx$	10,2
Sprungtransposition	$xzy \Leftrightarrow yzx$	0,8
Zwillingsvertauschung	$xx \Leftrightarrow yy$	0,6
phonetische Fehler (Bsp: 50 \Leftrightarrow 15)	$a0 \Leftrightarrow 1a$	0,5
Sprung-Zwillingsfehler	$xzx \Leftrightarrow yzy$	0,3
sonstige Fehler	-	8,6



- **Erster Ansatz**
 - Prüfziffer hinzu, so dass $\sum_{i=0,n} d_i = 0$
 - Bemerkungen
 - Prüfziffer darf "irgendwo" stehen.
 - Einfache Begrenzung auf Ziffernwert durch Rechnung "modulo 10"
 - Theoretischer Hintergrund
 - Mathematischer Begriff des Körpers (Menge K , Abb. $+$, Abb. x),
 - Endlicher Körper, Minimal-Beispiel: $K_2 = (\{0, 1\}, +, x)$, mit Definition der Abbildung per 2x2-Matrizen
 - schließlich: K_{10}
 - Problem: **Immun gegen Nachbarvertauschungen!**



- **Zweiter Ansatz**
 - Gewichte hinzu, so dass $\sum_{i=0,n} (w_i * d_i) = 0$
 - Problem: Wie Gewichte so wählen, dass Verhoeff-Fehler minimiert werden?
 - Modulo-10 Verfahren sind begrenzt:
 - **Einzelfehler:**
Sämtliche Gewichte müssen teilerfremd zur 10 sein
 - Frage: Gewichte sind Ziffern - warum?
 - Frage: Welche Gewichte sind teilerfremd zur 10?
 - Antwort: 1, 3, 7, 9
 - **Nachbarvertauschungen:**
Differenzen benachbarter Gewichte müssen teilerfremd zur 10 sein.
 - **Beide Bedingungen schließen sich gegenseitig aus!**



- Prüfzifferermittlung beim EAN/UCC-System
 - Gewichte: 1 - 3 - 1 - 3 - ... von rechts nach links (!)
 - Beispiel EAN:
 - EAN = 4 2 1 0 2 0 1 1 4 2 2 7 0
 - Gew. = 1 3 1 3 1 3 1 3 1 3 1 3 1
 - Prod. = 4 6 1 0 2 0 1 3 4 6 2 1 0 (modulo 10)
 - Summe = 30 = 0 (mod 10)
 - Beispiel NVE, Prüfziffer x gesucht:
 - NVE = 3 4 2 1 0 2 0 1 1 2 3 4 5 6 7 8 9 x
 - Gew. = 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1
 - Prod. = 9 4 6 1 0 2 0 1 3 2 9 4 5 6 1 8 7 x (modulo 10)
 - Summe = 68 + x (mod 10) = 8 + x (mod 10)
 - Wähle x so, dass $8 + x = 0 \pmod{10} \implies x = 2$



- Ausweg
 - Teilerfremdheit trivial erfüllt, wenn Basis eine Primzahl ist
 - Praktisch: 11 ist schon eine
- Optimaler Fall:
 - $w_i := 2^i \pmod{11}$
 - jeder Wert von 1 bis 10 (0) wird durch 2^i genau einmal angenommen!
- Problem der Praxis: Prüfziffer kann $\{0, \dots, 9\}$ verlassen
 - Ausweg A:
 - Derartige Nummern schon bei der Vergabe ausschließen
 - Ausweg B:
 - Nichtnumerische "Ziffer" einführen. Beispiel: "X" bei der ISBN



- 10 Ziffern für vier Teile, mit Bindestrichen oder Leerzeichen getrennt
 1. 1-5 Länderziffern (0-9)
 - 0 und 1 / 2 / 3 / 4: Englischer / Französischer / Deutscher / Japanischer Sprachraum, usw.
 - 99937: Macau
 2. 2-7 Verlagsziffern (0-9)
 3. 1-6 Artikelziffern (0-9)
 4. 1 Prüfziffer (0-9 sowie X für 10)
- Gewichte:
 - $w_i := i$ (von links nach rechts)
 - jeder Wert von 1 bis 10 (X) wird genau einmal angenommen, und zwar der Reihe nach



- Beispiele
 - Addison-Wesley, "XML in der Praxis":
 - ISBN=3 8273 1330 9, mit EAN=9783827313300
 - Addison-Wesley, "Programming Ruby":
 - ISBN=0-201-71089-7, mit EAN=9780201710892
 - Addison-Wesley, "LaTeX..."
 - ISBN=0-201-15790-X (ohne EAN, Existenz von "X"!)
 - Wrox Press, "XSLT" Programmer's Reference 2nd ed."
 - ISBN=1-861005-06-7, mit UCC=676623050670
- ISBN und EAN/UCC
 - EAN-Präfix "978" und "979" ist für ISBN reserviert
 - Addison-Wesley erzeugte die EAN sinnvoll
 - Wrox's Schema ergibt Probleme bei Prüfziffer "X" !



Fehlererkennung der Verfahren



Verfahren	$x \Leftrightarrow y$	$xy - yx$	$xzy - yzx$	$xx - yy$	$a0 - 1a$	$xzx - yzy$	sonst. (Schätz.)
mod 10, 1,1,1,1	100,0	0,0	0,0	88,9	100,0	88,9	90,0
mod 10, 1,2,1,2	94,4	100,0	0,0	100,0	87,5	88,9	90,0
mod 10, 1,3,7,9	100,0	88,9	88,9	44,4	100,0	88,9	90,0
mod 11, ISBN	100,0	100,0	100,0	80,0	88,9	100,0	90,9
mod 11, 2^i	100,0	100,0	100,0	100,0	100,0	100,0	90,9

Nach: J. Michael, Mit Sicherheit, c't 7/1996, p. 264-266.