

DPaper – Papier als digitales Speichermedium Blah blah blah blah

Sebastian Muszytowski – Lars Stoltenow

Chaosdorf

30. Juli 2010

Teil I

Barcodes

von Sebastian Muszytowski

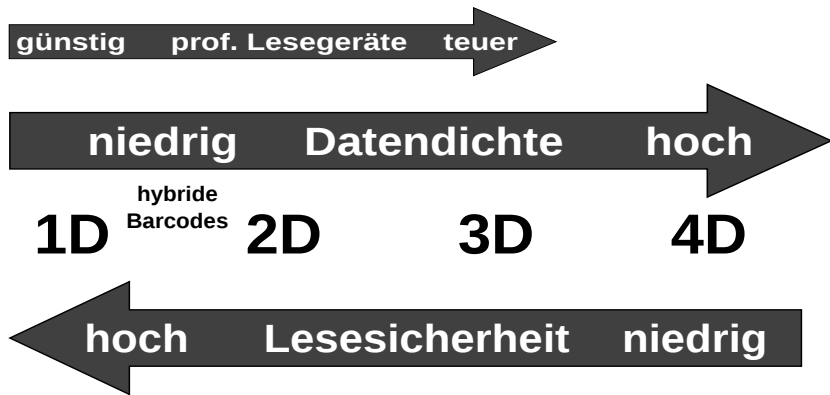
Was ist ein Barcode?

- ▶ Graphischer Datenspeicher
- ▶ Mögliche Datentypen (Nicht jeder Barcodetyp kann alles codieren):
 - ▶ Zahlen (0-9)
 - ▶ Buchstaben (a-z,A-Z)
 - ▶ Alphanumerische Zeichen (a-z,A-Z,0-9)
 - ▶ ASCII
 - ▶ Kodierte Binärdaten (Base64)
 - ▶ Binärdaten

Barcodetypen

- ▶ 1D-Codes: Strichmuster (Barcode)
- ▶ 2D-Codes: Matrix
- ▶ 3D-Codes: Matrix mit Farben
- ▶ 4D-Codes: Matrix mit Farben und Zeitachse
- ▶ Hybride Barcodes: Mischungen zwischen Barcodetypen (z.B. 1D und 2D)

Vergleich der Code-Typen



1D-Codes: Handelsstrichcodes

- ▶ Beispiel: EAN (European Article Number)
- ▶ Erlaubte Zeichen: 0-9
- ▶ Auf fast jedem Produkt zu finden
- ▶ Prüfziffer ist vorhanden



2D-Barcodes



QR-Code



DataMatrix



Cool-Data-Matrix



Semacode



000133



Trillcode



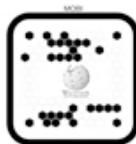
Quickmark



Shotcode



connexto



Beetagg



**CFCHESS
Qode**

2D-Barcodes



QR-Code



DataMatrix



Cool-Data-Matrix



Semacode



000133



Trillcode



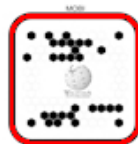
Quickmark



Shotcode



connexto



Beetagg



CFCHESS
Qode

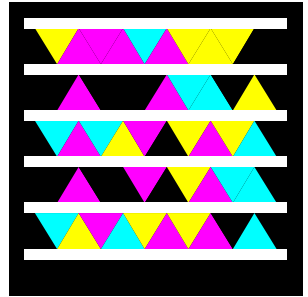
Steuerzeichen rot markiert

2D-Codes: Vor- und Nachteile

- ▶ Vorteile
 - ▶ Hohe Datendichte
 - ▶ Fehlertoleranzen bis 30%
 - ▶ Je nach Software hohe Lesesicherheit
- ▶ Nachteile
 - ▶ Je nach Software
 - ▶ Schlechte Lesesicherheit bei leicht gedrehten oder verzerrten Codes
 - ▶ Langsame Lesegeschwindigkeit
 - ▶ Lesegeräte vergleichsweise aufwendig

3D-Codes

- ▶ 2D-Codes mit zusätzlicher Dimension
- ▶ Dritte Dimension ist meist Farbe
- ▶ Beispiel: High Capacity Color Barcode (Microsoft)

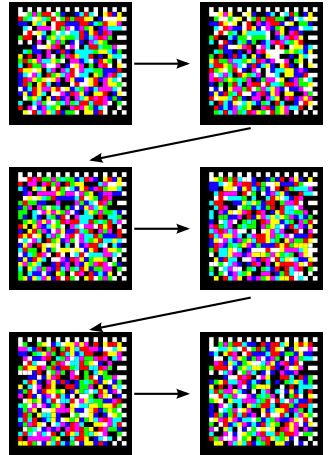


3D-Codes: Vor- und Nachteile

- ▶ Vorteile
 - ▶ Höhere Datendichte als 2D-Codes
- ▶ Nachteile
 - ▶ Noch keine Nutzung
 - ▶ Weitestgehend unbekannt
 - ▶ Alle Nachteile von 2D-Codes
 - ▶ Lesegeräte noch komplizierter und störungsanfälliger

4D-Codes

- ▶ 3D-Codes mit zusätzlicher Dimension
- ▶ Vierte Dimension ist Zeit



4D-Codes: Vor- und Nachteile

- ▶ Vorteile
 - ▶ Noch höhere Datendichte
- ▶ Nachteile
 - ▶ Noch fehleranfälliger durch das Filmen des Barcodes
 - ▶ Alle Nachteile von 3D-Codes

Hybride Barcodes

- ▶ Mischung aus 1D Barcodes und 2D Matrixcodes
- ▶ Bekanntester Vertreter PDF417



Hybride Barcodes: Vor- und Nachteile

- ▶ Vorteile

- ▶ Höhere Datendichte als 1D Barcodes
- ▶ Fehlerkorrekturverfahren vorhanden
- ▶ Kann mit 1D Barcodescannern gelesen werden

- ▶ Nachteile

- ▶ Wird von 2D Matrixcodes verdrängt

Dekodieren von Barcodes am Beispiel: Code128

- ▶ Erlaubte Zeichen: ASCII
- ▶ Drei verschiedene Subcodes, für effizientere Kodierung von bestimmtem Input
- ▶ Möglichkeit zwischen Subcodes zu wechseln



Weitere Merkmale von Code128

- ▶ Es gibt START und STOP Zeichen
- ▶ Jedes Zeichen ist mit 11 Strichen codiert
- ▶ Nur das STOP Zeichen ist mit 13 Strichen codiert
- ▶ Es gibt eine Prüfsumme

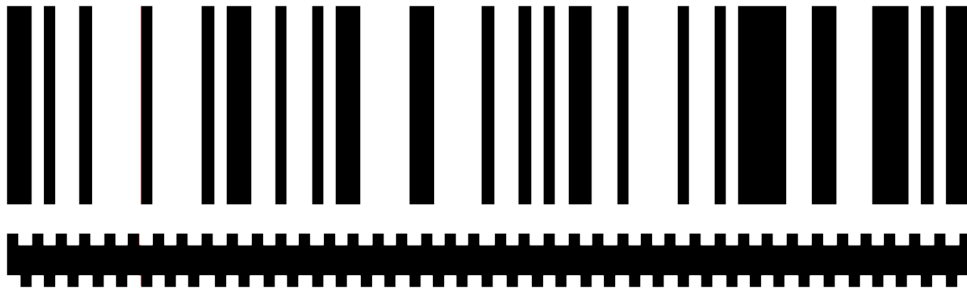
Ein Code128



Ein Code128 - Strichbreiten



Ein Code128 - Strichbreiten im Detail

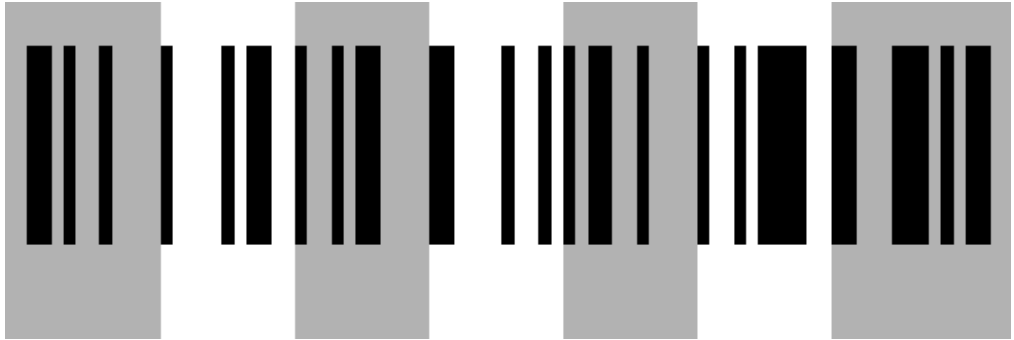


Ein Code128 - Strichbreiten im Detail



Jedes Zeichen ist 11 Striche breit (STOP-Zeichen 13)

Ein Code128 - Die Segmente im Detail



Ein Code128 - Die Segmente im Detail

Nun das Dekodiersheet zur Hand nehmen!

Ein Code128 - Erfassen der Balken und Leerzeichen Breite



Ein Code128 - Dekodieren



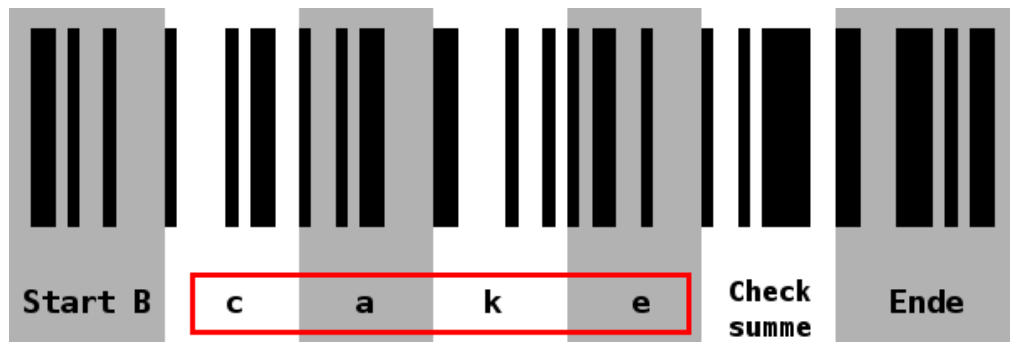
Mithilfe des Dekodiersheets kann nun jedes Zeichen dekodiert werden.

Ein Code128 - Dekodieren



Es lassen sich Steuerzeichen, Nutzdaten sowie eine Prüfsumme erkennen.

Ein Code128 - Erfassen der Daten



Ein Code128 - Erfassen der Daten



CAKE!

Berechnung der Prüfsumme

- ▶ Ziel: Prüfung der Daten auf Korrektheit
- ▶ Berechnung:
- ▶ $1 \times \text{Startz.} + 1 \times 1. \text{ Wert} + 2 \times 2. \text{ Wert} [\dots]$
- ▶ $\text{Summe} \div 103, \text{ Rest} = \text{Prüfsumme}$

Berechnung der Prüfsumme

- ▶ In unserem Beispiel:
- ▶ $104 \text{ (StartB)} + 1 \times 67 \text{ (c)} + 2 \times 65 \text{ (a)} + 3 \times 75 \text{ (k)} + 4 \times 69 \text{ (e)}$
- ▶ $= 802 \mid 802 \bmod 103 = 81 = \text{Prüfzeichen "q"}$

Lösung des Tasksheets

- ▶ Sheet 01: A_place_
- ▶ Sheet 02: and_a_ti
- ▶ Sheet 03: me,_neit
- ▶ Sheet 04: her_one_
- ▶ Sheet 05: too_far_

Lösung des Tasksheets

- ▶ Sheet 06: away. □51
- ▶ Sheet 07: .2139□6.
- ▶ Sheet 08: 7845□201
- ▶ Sheet 09: 0-08-07□
- ▶ Sheet 10: 18:00:00

A place and a time, neither one too far away.

51.2139 6.7845 2010-08-07 18:00:00

Teil II

DPaper – Papier als digitales Speichermedium von Lars Stoltenow

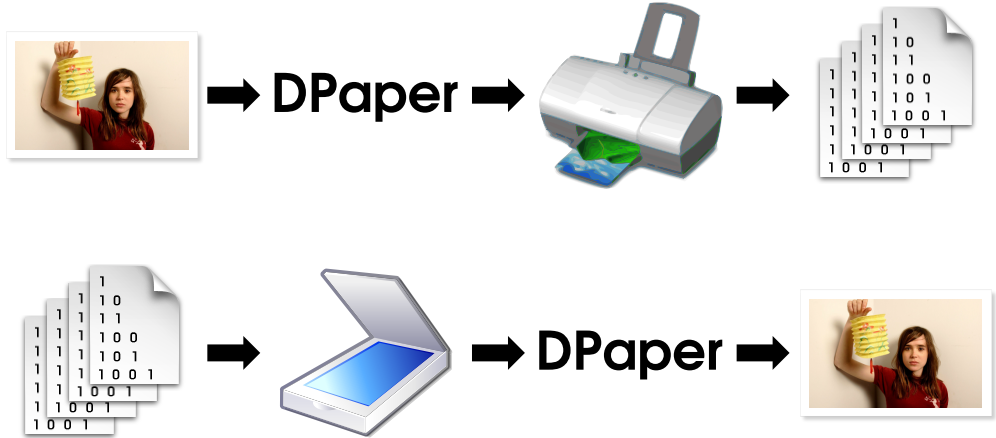
Worum geht's (nicht)?

- ▶ Vielzahl von Backupmedien
- ▶ DVDs, Festplatten, Flash, Magnetband
- ▶ Warum nicht mal Papier?
- ▶ Einfach mal Spaß haben, statt etwas Praxistaugliches zu produzieren :-)

Vorteile von Papier als Speichermedium

- ▶ Herkömmliche Speichermedien haben begrenzte Lebensdauer:
 - ▶ CD, DVD: 5-10 Jahre
 - ▶ Festplatten: 2-10 Jahre
 - ▶ Magnetband: 30 Jahre
 - ▶ Papier: *>100* Jahre
- ▶ Holz ist ein nachwachsender Rohstoff

Funktionsweise



Mögliche Speicherformate

- ▶ Text + Zeichenerkennung
- ▶ 2D-Codes (QR-Code, DataMatrix)
- ▶ Barcodes

OCR-A im Jahr 1968

- ▶ Schrift aus der Frühzeit der optischen Zeichenerkennung
- ▶ Sieht scheiße aus
- ▶ Die Kisten der 70er konnten es lesen



OCR-A im Jahr 2010

OCR-A Test Page

The alphabet, from 12 to 4.

12 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

10 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

09 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

08 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

07 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

06 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

05 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

04 Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz

Numerals, from 12 to 4.

12 0 1 2 3 4 5 6 7 8 9

11 0 1 2 3 4 5 6 7 8 9

10 0 1 2 3 4 5 6 7 8 9

9 0 1 2 3 4 5 6 7 8 9

8 0 1 2 3 4 5 6 7 8 9

7 0 1 2 3 4 5 6 7 8 9

6 0 1 2 3 4 5 6 7 8 9

5 0 1 2 3 4 5 6 7 8 9

4 0 1 2 3 4 5 6 7 8 9

_n_0pp_qRr__tu_vy__xx_yzz Nn_Dpp_qRrg_Ttuuuv__x_yyzz
_ ^__ 0 ^_,00

The al_habet1 fr0_ _2 to 4_

_2 Aa Bb Cc Dd Ee ff Gg Hh Ii Jj Kk Ll Mm
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww X_ Yy Zz

10 _a Bb Cc Dd Ee ff Gg Hh Ii Jj Kk Ll _m
Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww XK Yy Zz

09 _a Bb Cc Dd Ee ff Lg Hh Ii Jj Kk Ll _m
Nn Oo Pp dq Rr _s Tt Uu Vv Ww Xx Yy Zz

0_ _a B_ Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll rlm
Nn Oo Pp aq Rr S_ Tt Uu Vv Ww Kx Yy Zz

0_ _a Bb Cc _d Ee Ff Gg Hh Ii Jj Kk Ll _m
Nn Oo Pp _q Rr 5_ Tt Uu Vv Ww X_ Yy Zz

NU_era___ frOm _2 to 4_

12 0123456789

_ 0_23456709

10 0123456109

9 0_234567B9

_ 01234567_9

1 0123456__9

t D_2345t7_9

OCR-A im Jahr 2010

- ▶ Nicht einmal OCR-A wird richtig erkannt
- ▶ Auch nicht mit sehr großer Schrift
- ▶ Dabei war es nur ein Screenshot
- ▶ Ein Test mit Scan scheiterte komplett

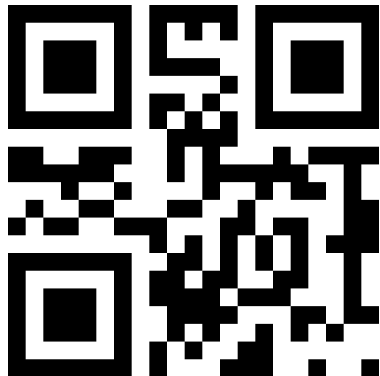
OCR-A im Jahr 2010

- ▶ 16 Zeichen können recht fehlerfrei unterschieden werden:
AaDdEeHhTtRrYy2347
- ▶ Mit kleinstmöglicher Schriftgröße
ca. 5700 Byte pro Seite
- ▶ Und 20 manuell zu behebenden
Lesefehlern



QR-Code und andere 2D-Geschichten

- ▶ Zweidimensional
- ▶ Bietet Fehlerkorrektur
- ▶ Weitere Vorteile von muzy
- ▶ Maximale Datenmenge:
 - ▶ QR-Code: ca. 3 KiB/Code
 - ▶ DataMatrix: ca. 1.5 KiB
 - ▶ Pro Seite 10 bis 20 KiB



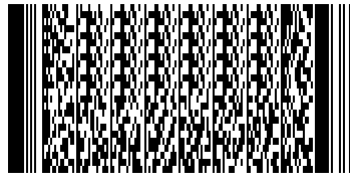
QR-Code und andere 2D-Geschichten

- ▶ Freie Dekodiertools verfügbar
- ▶ Erkennen aber größere Codes nicht richtig
- ▶ Funktionieren auch bei minimalen Bildfehlern (Scan) nicht richtig
- ▶ Zu unzuverlässig für Datenspeicherung



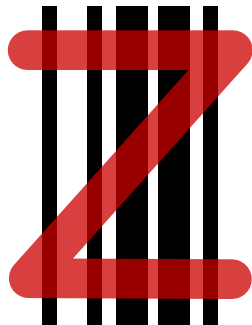
Barcodes

- ▶ Speichern vergleichsweise wenig Daten pro Code
- ▶ Lassen sich aber gut stapeln
- ▶ Beispiel: PDF417 (gestapelte 1D-Codes)
- ▶ Aber keine freie Lesesoftware
- ▶ Also normale Codes selbst stapeln



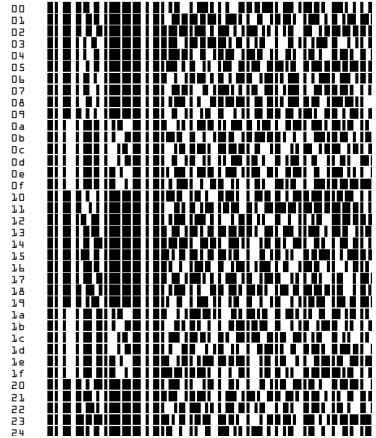
ZBAR

- ▶ Für 1D-Codes gibt es gute Software
- ▶ ZBAR `<http://zbar.sf.net>`
- ▶ Erkennt viele Barcode-Typen
- ▶ Gute Erkennungsrate auch bei schlechten Scans



Erster Versuch mit gestapelten Codes

- ▶ Rechts: Anhäufung von Code128-Barcodes
- ▶ ZBAR liest Codes in zufälliger Reihenfolge aus
- ▶ Enthält darum Nummer des Datenblocks
- ▶ 3840 Byte pro Seite



Fehlerkorrektur

- ▶ Etwa 90% aller Codes werden ohne weiteres erkannt
- ▶ Der Rest müsste aufwendig nachbearbeitet werden
- ▶ Weitere Gefahr: teilzerstörte Dokumente (Staub, Flecken, Risse, ...)
- ▶ Lösung: Vorwärtsfehlerkorrektur
- ▶ Nun *4800* Byte in *160* Zeilen
 - ▶ Aber weiterhin nur 3840 Byte Nutzdaten

Weitere Optimierungen

- ▶ Verteilung auf zwei Spalten
 - ▶ Weniger nicht erkannte Codes
- ▶ Erhöhung der Datenmenge auf 4096 Byte pro Seite
- ▶ Hinzufügen eines Seitenheaders (Name, Offset, Länge)

DPaper fid:shrine.jpg off:20480 rawlen:2337



Anwendung

```
% ./encode_file -I foo2.jpg < foo.jpg  
% lpr output.ps
```

```
% while true; do scanimage | ./decode_file; done  
retrieved data from image!
```

```
... ^C
```

```
% diff -s foo.jpg foo2.jpg
```

```
Files foo.jpg and foo2.jpg are identical
```

Nachteile von Papier als Speichermedium

- ▶ Geringe Speicherdichte
- ▶ Für größere Datenmengen zu teuer
- ▶ Und zu langsam
- ▶ Büropapier und Druckertinte haben keine lange Lebensdauer
- ▶ Warum also auf Papier speichern?
- ▶ Because we can

Fragen?

github.com/penma/dpaper