

	Informationstechnik Verfahren der automatischen Identifikation und Datenerfassung Spezifikationen für Strichcodesymbologien PDF417 (ISO/IEC 15438:2001) Deutsche Fassung EN ISO/IEC 15438:2003	DIN EN ISO/IEC 15438
--	--	--------------------------------

ICS 35.040

Information technology —
Automatic identification and data capture techniques —
Bar code symbology specifications —
PDF417 (ISO/IEC 15438:2001);
German version EN ISO/IEC 15438:2003

Technologies de l'information —
Techniques automatiques d'identification et de capture des données —
Spécifications pour les symboles de codes à barres —
PDF417 (ISO/IEC 15438:2001);
Version allemande EN ISO/IEC 15438:2003

Die Europäische Norm EN ISO/IEC 15438:2003 hat den Status einer Deutschen Norm.

Nationales Vorwort

Diese Europäische Norm wurde im Europäischen Komitee für Normung vom Technischen Komitee „Strichcodierung“ erarbeitet.

Das zuständige deutsche Normungsgremium ist der Normenausschuss Daten- und Warenverkehr in der Konsumgüterwirtschaft (NDWK) im DIN.

Fortsetzung 101 Seiten EN

Normenausschuss Daten- und Warenverkehr in der Konsumgüterwirtschaft (NDWK)
im DIN Deutsches Institut für Normung e. V.

— Leerseite —

Deutsche Fassung

Informationstechnik
Vefahren der automatischen Identifikation und Datenerfassung
Spezifikationen für Strichcodesymbologien
PDF417 (ISO/IEC 15438:2001)

Information technology —
Automatic identification and data capture techniques —
Bar code symbology specifications —
PDF417 (ISO/IEC 15438:2001)

Technologies de l'information —
Techniques automatiques d'identification et de
capture des données —
Spécifications pour les symboles de codes à barres —
PDF417 (ISO/IEC 15438:2001)

Diese Europäische Norm wurde vom CEN am 9. Januar 2003 angenommen.

Die CEN-Mitglieder sind gehalten, die CEN/CENELEC-Geschäftsordnung zu erfüllen, in der die Bedingungen festgelegt sind, unter denen dieser Europäischen Norm ohne jede Änderung der Status einer nationalen Norm zu geben ist. Auf dem letzten Stand befindliche Listen dieser nationalen Normen mit ihren bibliographischen Angaben sind beim Management-Zentrum oder bei jedem CEN-Mitglied auf Anfrage erhältlich.

Diese Europäische Norm besteht in drei offiziellen Fassungen (Deutsch, Englisch, Französisch). Eine Fassung in einer anderen Sprache, die von einem CEN-Mitglied in eigener Verantwortung durch Übersetzung in seine Landessprache gemacht und dem Management-Zentrum mitgeteilt worden ist, hat den gleichen Status wie die offiziellen Fassungen.

CEN-Mitglieder sind die nationalen Normungsinstitute von Belgien, Dänemark, Deutschland, Finnland, Frankreich, Griechenland, Irland, Island, Italien, Luxemburg, Malta, Niederlande, Norwegen, Österreich, Portugal, Schweden, Schweiz, der Slowakei, Spanien, der Tschechischen Republik, Ungarn und dem Vereinigten Königreich.



EUROPÄISCHES KOMITEE FÜR NORMUNG
EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION

Management-Zentrum: rue de Stassart, 36 B-1050 Brüssel

Inhalt

	Seite
Vorwort.....	4
Einleitung	4
1 Anwendungsbereich	5
2 Normative Verweisungen	5
3 Begriffe, mathematische Symbole und Abkürzungen.....	5
3.1 Begriffe.....	5
3.2 Mathematische Symbole und Operationen	7
3.3 Abkürzungen	8
4 Anforderungen	8
4.1 Eigenschaften der Symbologie.....	8
4.1.1 Haupteigenschaften.....	8
4.1.2 Zusammenfassung weiterer Merkmale.....	9
4.2 Symbolstruktur.....	10
4.2.1 Symbolparameter von PDF417	10
4.2.2 Zeilenparameter.....	10
4.2.3 Codewortfolge	11
4.3 Grundcodierung	12
4.3.1 Symbolzeichenstruktur	12
4.3.2 Start- und Stoppzeichen.....	12
4.4 Höhere (Daten-)Codierung	13
4.4.1 Funktions-Codewörter	13
4.4.2 Textverdichtungsmodus	16
4.4.3 Byte-Verdichtungsmodus	20
4.4.4 Numerischer Verdichtungsmodus	22
4.4.5 Hinweis zur Auswahl des geeigneten Verdichtungsmodus	23
4.4.6 Behandlung der reservierten Codewörter in PDF417.....	24
4.5 Erweiterte Kanalübersetzung.....	24
4.5.1 Codierung der ECI-Zuordnungsnummern.....	25
4.5.2 Vorher zugeordnete erweiterte Kanalübersetzungen	26
4.5.3 Codierung von ECI-Folgen innerhalb der Verdichtungsmodi	26
4.5.4 Nachdecodierungsprotokoll	28
4.6 Bestimmung der Codewortfolge.....	28
4.7 Fehlererkennung und -korrektur	28
4.7.1 Fehlerkorrekturgrad.....	28
4.7.2 Fehlerkorrekturfähigkeit.....	29
4.7.3 Festlegen der Fehlerkorrektur-Codewörter.....	30
4.8 Maße	30
4.8.1 Mindestlänge eines Moduls (X).....	30
4.8.2 Zeilenhöhe (Y).....	30
4.8.3 Hellzonen	31
4.9 Festlegen des Symbolformats	31
4.9.1 Festlegen des Seitenverhältnisses des Moduls	31
4.9.2 Festlegen der Zeilen und Spalten der Symbolmatrix	31
4.10 Erzeugen der Fehlerkorrektur-Codewörter	32
4.11 Codierung auf niedriger Ebene.....	34
4.11.1 Cluster	34
4.11.2 Festlegen der Symbolmatrix	35
4.11.3 Festlegen der Werte für linke und rechte Zeilenindikatoren	35
4.11.4 Zeilencodierung.....	35
4.12 Verdichtetes PDF417	36
4.13 Macro-PDF417	36
4.13.1 Verdichtungsmodi und Macro-PDF417.....	36
4.13.2 ECIs und Macro-PDF417.....	36

4.14	Anwenderrichtlinien.....	36
4.14.1	Klarschriftübersetzung.....	36
4.14.2	Fähigkeit der automatische Unterscheidung (Autodiskrimination)	37
4.14.3	Anwenderdefinierte Parameter.....	37
4.14.4	PDF417-Symbolqualität.....	37
4.15	Referenz-Decodieralgorithmus.....	38
4.16	Fehlererkennung und Fehlerkorrekturverfahren	38
4.17	Übertragene Daten	38
4.17.1	Übertragene Daten in der Grund- (Standard-) Übersetzung.....	38
4.17.2	Übertragungsprotokoll für die erweiterte Kanalübersetzung	38
4.17.3	Übertragene Daten für Macro-PDF417	39
4.17.4	Übertragung von reservierten Codewörtern unter Verwendung des ECI-Protokolls	39
4.17.5	Symbologie-Identifikator	40
4.17.6	Übertragung mit älteren Protokollen	40
Anhang A (normativ)	Codierungs-/Decodierungstabelle für Strich-Lücken-Folgen für PDF417-Symbolzeichen	41
Anhang B (normativ)	Der Standard-Zeichensatz für den Byte-Verdichtungsmodus	54
Anhang C (normativ)	Codierungsalgorithmus für den Byte-Verdichtungsmodus	55
Anhang D (normativ)	Codierungsalgorithmus für den numerischen Verdichtungsmodus	57
Anhang E (normativ)	Anwenderdefinierte Fehlerkorrektur.....	59
Anhang F (normativ)	Koeffiziententabellen für die Berechnung von Fehlerkorrektur-Codewörtern für PDF417	60
Anhang G (normativ)	Verdichtetes PDF417.....	65
Anhang H (normativ)	Macro-PDF417.....	66
Anhang J (normativ)	Prüfung der Qualität von PDF417-Symbolen	74
Anhang K (normativ)	Referenz-Decodieralgorithmus für PDF417	76
Anhang L (normativ)	Verfahren der Fehlerkorrektur	80
Anhang M (normativ)	Symbologie-Identifikator	82
Anhang N (normativ)	Übertragungsprotokoll für Decoder nach älteren PDF417-Normen	83
Anhang P (informativ)	Algorithmus zur Verringerung der Anzahl der Codewörter.....	89
Anhang Q (normativ)	Richtlinien zur Ermittlung der Symbolmatrix	90
Anhang R (informativ)	Berechnung der Koeffizienten zur Erzeugung der Fehlerkorrektur-Codewörter – Rechenbeispiel.....	93
Anhang S (informativ)	Erzeugung der Fehlerkorrektur-Codewörter – Rechenbeispiel	94
Anhang T (informativ)	Divisionsschaltung zur Erzeugung der Fehlerkorrektur-Codewörter	98
Anhang U (informativ)	Kompatibilität der Autodiskrimination	99
Literaturhinweise		100
Anhang ZA (normativ)	Normative Verweisungen auf internationale Publikationen mit ihren entsprechenden europäischen Publikationen	101

Vorwort

Der Text von ISO/IEC 15438:2001 wurde vom Technischen Komitee ISO/IEC/JTC 1 „Informationstechnologie“ der Internationalen Organisation für Normung (ISO) und der Internationalen Elektrotechnischen Kommission (IEC) erarbeitet und als EN ISO/IEC 15438:2003 durch das Technische Komitee CEN/TC 225 „Strichcodierung“ übernommen, dessen Sekretariat vom NEN gehalten wird.

Diese Europäische Norm muss den Status einer nationalen Norm erhalten, entweder durch Veröffentlichung eines identischen Textes oder durch Anerkennung bis November 2003, und etwaige entgegenstehende nationale Normen müssen bis November 2003 zurückgezogen werden.

Dieses Dokument ersetzt ENV 12925:1997.

Entsprechend der CEN/CENELEC-Geschäftsordnung sind die nationalen Normungsinstitute der folgenden Länder gehalten, diese Europäische Norm zu übernehmen: Belgien, Dänemark, Deutschland, Finnland, Frankreich, Griechenland, Irland, Island, Italien, Luxemburg, Malta, Niederlande, Norwegen, Österreich, Portugal, Schweden, Schweiz, Slowakei, Spanien, Tschechische Republik, Ungarn und Vereinigtes Königreich.

Anerkennungsnotiz

Der Text von ISO/IEC 15438:2001 wurde vom CEN als EN ISO/IEC 15438:2003 ohne irgendeine Abänderung genehmigt.

ANMERKUNG Die normativen Verweisungen auf Internationale Normen sind im Anhang ZA (normativ) aufgeführt.

Einleitung

Die Technik der Strichcodierung beruht auf der Erkennung von Mustern aus Strichen und Lücken mit festgelegten Maßen. Es gibt verschiedene Verfahren für die Codierung von Informationen im Strichcodeformat, die als Symbolgien bezeichnet werden. Die Regeln zur Übersetzung der Zeichen in Strich- und Lückenmuster und weitere wichtige Eigenschaften sind die Symbolgiespezifikation.

Hersteller von Strichcodegeräten und Anwender der Strichcodetechnik benötigen öffentlich verfügbare Standard-Symbolgiespezifikationen, auf die sie sich bei der Entwicklung von Geräten und Anwendungsstandards beziehen können. Es ist Ziel und Vorstellung von ISO/IEC, dass die in dieser Norm vorgestellte Symbolgie allgemein zugänglich und frei von jeglichen Anwendereinschränkungen, Lizenzen und Gebühren ist.

1 Anwendungsbereich

Diese Internationale Norm legt Anforderungen für die Strichcodesymbologie PDF417 fest. Es werden Eigenschaften, Datenzeichencodierung, Symbolformate, Maße, Regeln für Fehlerkorrekturen und Decodieralgorithmen der Symbologie PDF417 sowie eine Anzahl von Anwendungsparametern festgelegt.

2 Normative Verweisungen

Die folgenden zitierten Dokumente sind für die Anwendung dieses Dokuments erforderlich. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information exchange*.

ISO 8859-1:1998, *Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1*.

ISO/IEC 15416, *Information technology — Automatic identification and data capture techniques — Bar code print quality test specification — Linear symbols*.

AIM International Technical Specification, *Extended Channel Interpretations — Part 1: Identification Schemes and Protocol*.

AIM International Procedure Standard, *Extended Channel Interpretations — Part 2: Registration of Coded Character Sets and Other Data Formats*.

EN 796, *Bar coding — Symbology Identifiers*.

EN 1556, *Strichcodierung — Terminologie*.

3 Begriffe, mathematische Symbole und Abkürzungen

3.1 Begriffe

Für die Anwendung dieser Internationalen Norm gelten die in EN 1556 angegebenen und die folgenden Begriffe.

Algorithmus, Anwendungsstandard, ASCII, Autodiskrimination, bi-direktional, binär, Bit, Codewort, codieren, Datenfeld, Datenverdichtung, Datenzeichen, Decoder, Decodieralgorithmus, Element, Fehlerkorrekturgrad, führende Nullen, Hellzone, Klarschriftzeichen, kontinuierlicher Code, lineare Symbologie, Lücke/Zwischenraum, mehrzeilige Strichcodesymbologie, Modul, Modulo, (n,k) Symbologie, numerisch, Pad-Zeichen, Pad-Codewort, Referenz-decodieralgorithmus, Scanner, selbstüberprüfend, Spalte, Startzeichen, Stoppzeichen, Strich, Strichcode, Symbol-Seitenverhältnis, Symbollänge, Symbologie, Symbologie-Identifikator, Symbolzeichen, Überhang, verdichtetes Codewort, X-Modul, Y-Modul, Zeile

Darüber hinaus gelten folgende Definitionen für diese Internationale Norm.

3.1.1

Grundkanalmodell

Standardsystem für die Codierung und Übertragung von Strichcode-Daten, bei dem vom Decoder nur Datennachrichten-Bytes aber keine Steuerinformationen über die übertragene Nachricht ausgegeben werden. Ein Decoder, der diesem Modell entspricht, arbeitet im Grundkanalmodus

3.1.2

Strich-Lücke-Folge

Folge, die die Modullänge der Elemente eines Symbolzeichens darstellt

3.1.3

Cluster

Eine der drei Teilmengen von PDF417-Symbolzeichen, die sich gegenseitig ausschließen. Die Symbolzeichen in einem vergebenen Cluster entsprechen den besonderen Strukturregeln, die bei der Decodierung der Symbologie verwendet werden

3.1.4

Verdichtungsmodus

Bezeichnung für einen der drei Datenverdichtungsalgorithmen in PDF417 und zwar den Text-, numerischen und Byte-Verdichtungsmodus. Diese Modi bilden 8-Bit-Datenbytes auf PDF417-Codewörter ab

3.1.5

Fehlerkorrektur-Codewort

Codewort in einem Symbol, das einen Wert nach dem Fehlerkorrektur-Codewort-Algorithmus codiert, um Decodierungsfehler aufzufinden und je nach Fehlerkorrekturgrad zu korrigieren

3.1.6

Erweiterte Kanalübersetzung (ECI)

Verfahren in einigen Symbologien wie auch PDF417, mit dem die Standardübersetzung durch eine andere Übersetzung zuverlässig ersetzt wird. Die Übersetzung, die der Erstellung des Symbols vorausgehen soll, kann nach Decodierung des gelesenen Symbols abgerufen und die Datennachricht in ihr ursprüngliches Format zurückverwandelt werden

3.1.7

Erweitertes Kanalmodell

System zur Codierung und Übertragung von Datennachrichten-Bytes sowie Steuerinformationen über die Nachricht. Ein Decoder nach diesem Modell arbeitet im erweiterten Kanalmodus. Die Steuerinformation wird unter Verwendung der Escape-Folgen der erweiterten Kanalübersetzung übertragen

3.1.8

Funktions-Codewort

Codewort in einer Symbologie, das einen bestimmten Ablauf innerhalb der Symbologie auslöst, z. B. Umschaltung zwischen Datencodierungssätzen, um ein Verdichtungsschema aufzurufen, das Lesegerät zu programmieren oder die erweiterten Kanalübersetzungen aufzurufen

3.1.9

Universeller Etikettenidentifikator (GLI)

Verfahren in der Symbologie PDF417, das sich gleich der erweiterten Kanalübersetzung verhält. Das GLI-System ist der von der Symbologie abhängige Vorgänger des von der Symbologie unabhängigen ECI-Systems

3.1.10

Macro-PDF417

Verfahren in der Symbologie PDF417, um Daten logisch von einer Computerdatei auf eine Anzahl zugehöriger PDF417-Symbole zu verteilen. Das Verfahren erhöht die Datenkapazität beträchtlich über das eines einzelnen Symbols hinaus. Dieses Verfahren entspricht der strukturierten Verkettung in anderen Symbologien

3.1.11

Moduswechsel-Codewort

Codewort, das verwendet wird, um von einem Modus zum anderen zu wechseln, der solange gültig bleibt, bis ein weiteres Wechsel- oder Umschalt-Codewort indirekt oder direkt eingebracht wird oder bis das Ende des Etiketts erreicht ist

3.1.12

Modusumschalt-Codewort

Codewort, das verwendet wird, um zur Darstellung eines Codewortes von einem Modus in einen anderen umzuschalten, danach kehrt die Codierung wieder in den ursprünglichen Modus zurück

3.1.13

Zeilenindikator-Codewort

PDF417-Codewort neben dem Start- oder Stoppzeichen einer Zeile, das Informationen über die Struktur des PDF417-Symbols hinsichtlich Zeilenerkennung, Gesamtanzahl von Zeilen und Spalten und Fehlerkorrekturgrad enthält

3.1.14**Symbollängendescriptor**

Codewort in einem PDF417-Symbol, das die gesamte Anzahl von verdichteten Codewörtern im Symbol codiert. Der Symbollängendescriptor muss immer das erste Codewort in einem PDF417-Symbol sein

3.2 Mathematische Symbole und Operationen

Für die Anwendung dieser Norm gelten folgende mathematische Symbole. In einigen Fällen wurden die nachfolgenden Symbole in den Gleichungen unterschiedlich verwendet. Dies dient der Übereinstimmung mit einer allgemeineren Anwendung des Zeichensystems und ist im Text deutlich angegeben.

- A Symbol-Seitenverhältnis (Höhe zu Breite) eines PDF417-Symbols
- b Elementlänge in einem Symbolzeichen
- c Anzahl von Spalten im Datenfeld des Symbols (ohne Start-, Stopp- und Zeilenindikator-Codewörter)
- d verdichtetes Codewort einschließlich aller Funktions-Codewörter
- E Fehlerkorrektur-Codewort
- e Abstand von einem Rand zu einem gleichen Rand in einem Symbolzeichen
- F Zeilennummer
- f Anzahl ersetzter Fehler
- H Höhe des Symbols einschließlich Hellzone
- K Nummer des Clusters
- k Anzahl der Fehlerkorrektur-Codewörter
- L linker Zeilenindikator
- l Anzahl von Löschungen
- m Anzahl von Quelldaten-Codewörtern vor dem Zusatz des Symbollängendescriptors und von Pad-Codewörtern
- n Gesamtanzahl von verdichteten Codewörtern einschließlich Symbollängendescriptor und Pad-Codewörtern
- p Pitch oder Breite eines Symbolzeichens
- Q_H horizontale Hellzone
- Q_V vertikale Hellzone
- R rechter Zeilenindikator
- r Zeilenanzahl im Symbol
- s Fehlerkorrekturgrad
- W Länge des Symbols einschließlich Hellzone
- X X-Modul oder Modullänge
- Y Modulhöhe (auch Zeilenhöhe)

Für die Anwendung dieser Norm gelten folgende mathematische Funktionen:

- div** ist das Operationszeichen für die ganzzahlige Division, abgerundet
- INT** ist ein ganzzahliger Wert, d. h., eine Zahl wird auf ihre ganze Zahlenkomponente abgerundet, wobei Dezimalstellen vernachlässigt werden
- mod** ist der positive ganzzahlige Restwert nach der Teilung. Wenn der Rest negativ ist, wird der Wert des Divisors addiert, um ihn positiv zu machen. Beispiel: Der Rest der Gleichung $-29\ 160$ geteilt durch 929 ist -361 . Bei einer Addition dieses Wertes zu 929 ergibt das 568 .

3.3 Abkürzungen

Für die Anwendung dieser Norm gelten folgende Abkürzungen:

- ECI** (Extended Channel Interpretation) erweiterte Kanalübersetzung
- GLI** (Global Label Identifier) universeller Etikettenidentifikator

4 Anforderungen

4.1 Eigenschaften der Symbologie

4.1.1 Haupteigenschaften

PDF417 ist eine Strichcodesymbologie mit folgenden Grundeigenschaften:

- a) Codierbarer Zeichensatz:
- 1) Der Textverdichtungsmodus (siehe Abschnitt 4.4.2) ermöglicht die Codierung aller druckbaren ASCII-Zeichen, d. h. einschließlich der Werte von 32 bis 126 nach ISO/IEC 646 sowie ausgewählter Steuerzeichen.
 - 2) Der Byte-Verdichtungsmodus (siehe Abschnitt 4.4.3) ermöglicht die Codierung aller 256 möglichen 8-Bit-Byte-Werte. Das sind alle ASCII-Zeichenwert von 0 bis einschließlich 127 und bietet Unterstützung internationaler Zeichensätze.
 - 3) Der numerischer Verdichtungsmodus (siehe Abschnitt 4.4.4) ermöglicht die wirksame Codierung numerischer Datenfolgen.
 - 4) Bis zu 811 800 verschiedene Zeichensätze oder Datenübersetzungen.
 - 5) Verschiedene Funktions-Codewörter für Steuerzwecke.
- b) Symbol-Zeichenstruktur: (n, k, m) Zeichen mit 17 Modulen (n), 4 Strich- und 4 Lückenelementen (k), wobei das größte Element eine Länge von 6 Modulen hat (m).
- c) Größtmögliche Anzahl von Datenzeichen je Symbol (bei Fehlerkorrekturgrad 0): 925 verdichtete Codewörter, die codieren können:
- 1) Textverdichtungsmodus: 1 850 Zeichen (mit 2 Datenzeichen je Codewort).
 - 2) Byte-Verdichtungsmodus: 1 108 Zeichen (mit 1,2 Datenzeichen je Codewort).
 - 3) Numerischer Verdichtungsmodus: 2 710 Zeichen (mit 2,93 Datenzeichen je Codewort).

Beim mindestens empfohlenen Fehlerkorrekturgrad gibt es 863 verdichtete Codewörter, die codieren können:

- 1) Textverdichtungsmodus: 1 726 Zeichen (mit 2 Datenzeichen je Codewort).
- 2) Byte-Verdichtungsmodus: 1 033 Zeichen (mit 1,2 Datenzeichen je Codewort).
- 3) Numerischer Verdichtungsmodus: 2 528 Zeichen (mit 2,93 Datenzeichen je Codewort).

d) Symbolgröße:

- 1) Anzahl der Zeilen: 3 bis 90.
- 2) Anzahl der Spalten: 1 bis 30.
- 3) Länge in Modulen: 90X bis 538X einschließlich Hellzonen.
- 4) Größte Kapazität an Codewörtern: 928 Codewörter.
- 5) Größte Kapazität an verdichteten Codewörtern: 925 Codewörter.

Da die Anzahl von Zeilen und Spalten ausgewählt werden kann, darf das Seitenverhältnis eines PDF417-Symbols beim Drucken verändert werden, damit den räumlichen Anforderungen der Anwendung entsprochen werden kann.

e) Auswählbare Fehlerkorrektur: 2 bis 512 Codewörter je Symbol (siehe Abschnitt 4.7).

f) Überhang der Hilfsdaten:

- 1) Je Zeile: 73 Module einschließlich Hellzonen.
- 2) Je Symbol: mindestens 3 zusätzlichen Codewörter als Symbolzeichen dargestellt.

g) Codetypen: kontinuierlich, mehrzeilig, 2-dimensional.

h) Zeichen-Selbstprüfung: ja.

i) Bi-direktional decodierbar: ja.

4.1.2 Zusammenfassung weiterer Merkmale

Die folgende Zusammenfassung beschreibt weitere Merkmale, die in PDF417 entweder verankert oder optional sind.

- a) **Datenverdichtung** (verankert): Diese Schemen wurden definiert, um eine Anzahl von Datenzeichen in Codewörtern zu verdichten. Im Allgemeinen werden Daten nicht direkt auf der Basis ein Zeichen für ein Codewort dargestellt (siehe Abschnitte 4.4.2 bis 4.4.4).
- b) **Erweiterte Kanalübersetzungen** (optional): Mit diesen Mechanismen können bis zu 811 800 verschiedener Datenzeichensätze oder Übersetzungen codiert werden (siehe Abschnitt 4.5).
- c) **Macro-PDF417** (optional): Dieser Mechanismus ermöglicht die logische und aufeinanderfolgende Darstellung von Dateien in einer Anzahl von PDF417-Symbolen. Bis zu 99 999 verschiedene PDF417-Symbole können so verknüpft oder verkettet und in einer beliebigen Folge gescannt werden, damit die ursprüngliche Datei richtig wiederhergestellt werden kann (siehe Abschnitt 4.13).
- d) **Rand-zu-Rand-Decodierung** (verankert): PDF417 kann durch die Messung der Elemente von einem Rand zu einem gleichen Rand decodiert werden (siehe Abschnitt 4.3.1).
- e) **Scannen mit Zeilenkreuzung** (verankert): Die Kombination der folgenden drei Eigenschaften ermöglicht bei PDF417 das Scannen mit Zeilenkreuzung:
 - horizontale Synchronisation oder Selbsttaktung
 - Zeilenerkennung
 - vertikale Synchronisation durch Verwendung des Clusterwertes, um eine lokale Zeilenunterscheidung zu erzielen

Durch diese Kombination werden mit einem einzigen linearen Scanvorgang eine Anzahl von Zeilen überstrichen und eine teilweise Decodierung der Daten ermöglicht, solange mindestens ein vollständiges Symbolzeichen je Zeile in sein Codewort decodiert wurde. Der Decodieralgorithmus kann die einzelnen Codewörter dann in einer sinnvollen Matrix anordnen.

- f) **Fehlerkorrektur** (verankert): Der Anwender darf einen von neun Fehlerkorrekturgraden festlegen. Alle Grade außer Grad 0 erkennen nicht nur Fehler, sondern können auch falsch decodierte oder fehlende Codewörter korrigieren (siehe Abschnitt 4.7).
- g) **Verdichtetes PDF417** (optional): In relativ "reinen" Umgebungen ist es möglich, einen Teil des Zeilenüberhangs zu kürzen, um die Symboldichte zu erhöhen (siehe Abschnitt 4.12).

ANMERKUNG Frühere Spezifikationen von PDF417 wurden als "Verkürztes" PDF417 bezeichnet. Verdichtetes PDF417 ist die bevorzugte Bezeichnung, um Unklarheiten bei der allgemeineren Bedeutung von "verkürzt" zu vermeiden.

4.2 Symbolstruktur

4.2.1 Symbolparameter von PDF417

Jedes PDF417-Symbol enthält einen Stapel von mindestens 3 vertikal ausgerichteten Zeilen (höchstens 90 Zeilen). Jede Zeile muss mindestens 1 Symbolzeichen (höchstens 30 Symbolzeichen) enthalten, ohne Start-, Stopp- und Zeilenindikatorspalten. Das Symbol muss auf allen vier Seiten eine Hellzone enthalten. Bild 1 zeigt ein PDF417-Symbol, das folgenden Text codiert: PDF417-Symbologienorm.

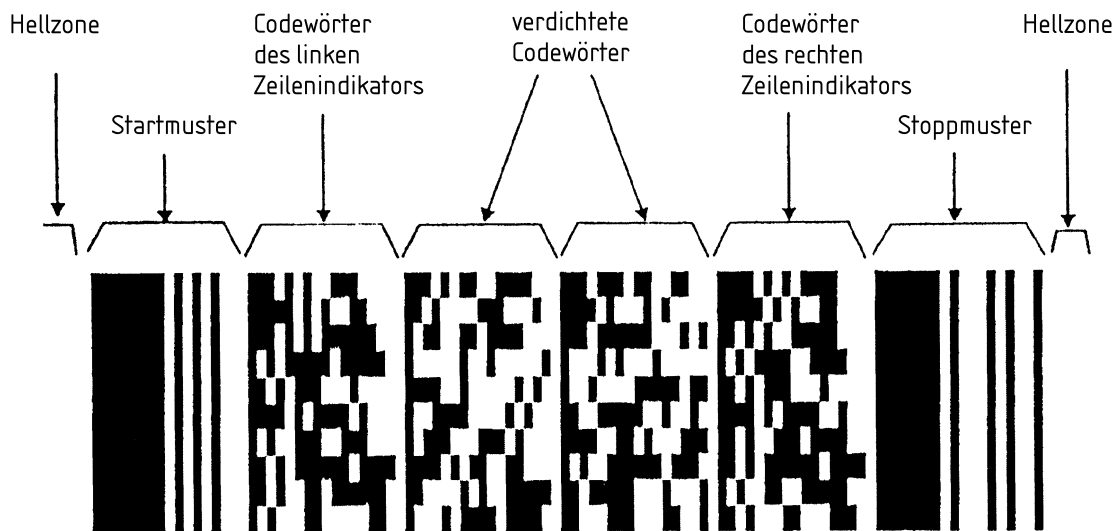


Bild 1 — Symbolstruktur in PDF417

4.2.2 Zeilenparameter

Jede PDF417-Zeile muss Folgendes enthalten:

- a) Anführende Hellzone
- b) Startzeichen
- c) Symbolzeichen für den linken Zeilenindikator
- d) 1 bis 30 Symbolzeichen
- e) Symbolzeichen für den rechten Zeilenindikator
- f) Stoppzeichen
- g) Abschließende Hellzone

ANMERKUNG Die Anzahl von Symbolzeichen (oder Codewörtern), die in Punkt d) angegeben ist, entspricht der Anzahl von Datenspalten.

4.2.3 Codewortfolge

Ein PDF417-Symbol kann bis zu 928 Symbolzeichen oder Codewörter enthalten. Symbolzeichen ist der passendere Begriff für die gedruckten Strich/Lückenmuster; Codewort hingegen ist mehr für die Bezeichnung des numerischen Wertes des Symbolzeichens geeignet. Die Codewörter müssen dieser Folge entsprechen:

- Das erste Codewort, der Symbollängendeskriptor, muss immer die gesamte Anzahl von verdichteten Codewörtern im Symbol codieren, einschließlich des Symbollängendeskriptors selbst, der verdichteten Codewörter und der Pad-Codewörter aber ohne die Anzahl von Fehlerkorrektur-Codewörtern.
- Die verdichteten Codewörter müssen beginnend mit dem höchstwertigen codierbaren Zeichen folgen. Zur Datenverdichtung dürfen Funktions-Codewörter eingeschoben werden.
- Pad-Codewörter dienen dazu, die Codewort-Folge in einer rechteckigen Matrix darzustellen. Pad-Codewörter dürfen auch dazu verwendet werden, zusätzliche Abschlusszeilen zu füllen, um ein von der Anwendung gewünschtes oder vorgeschriebenes Seitenverhältnis zu erzielen.
- Ein optionaler Macro-PDF417-Steuerblock.
- Fehlerkorrektur-Codewörter zur Fehlererkennung und -korrektur.

Die Codewörter werden so angeordnet, dass das höchstwertige Codewort neben dem Symbollängendeskriptor steht und die Codierung erfolgt von links oben nach rechts unten. Bild 2 zeigt im Layoutformat die Folge für das in Bild 1 dargestellte Symbol. In Bild 2 ist der Fehlerkorrekturgrad 1 verwendet worden und es war ein Pad-Zeichen nötig, um die Symbolmatrix zu füllen.

S T A R T	L ₁	d ₁₅	d ₁₄	R ₁	S T O P
	L ₂	d ₁₃	d ₁₂	R ₂	
	L ₃	d ₁₁	d ₁₀	R ₃	
	L ₄	d ₉	d ₈	R ₄	
	L ₅	d ₇	d ₆	R ₅	
	L ₆	d ₅	d ₄	R ₆	
	L ₇	d ₃	d ₂	R ₇	
	L ₈	d ₁	d ₀	R ₈	
	L ₉	E ₃	E ₂	R ₉	
	L ₁₀	E ₁	E ₀	R ₁₀	

Bild 2 — Beispiel eines schematischen Symbolentwurfs in PDF417

Dabei sind:

L, R, d und E wie in Abschnitt 3.2 definiert;

d₁₅ ist der Symbollängendeskriptor (in diesem Beispiel mit einem Wert von 16);

d₁₄ bis d₁ ist die codierte Darstellung der Daten;

d₀ ist das Pad-Codewort.

Die Regeln und Ratschläge zur Strukturierung der Matrix sind in Abschnitt 4.9 aufgeführt.

4.3 Grundcodierung

4.3.1 Symbolzeichenstruktur

Jedes PDF417-Symbolzeichen muss aus 4 Strich- und 4 Lückenelementen bestehen, wovon jedes ein bis sechs Module lang sein kann. Die 4 Striche und die 4 Lücken müssen insgesamt 17 Module umfassen. PDF417-Symbolzeichen können durch Messung des Abstandes e innerhalb des Zeichens decodiert werden.

Jedes Symbolzeichen wird durch eine 8-stellige Strich-Lücken-Folge definiert, die die Modullängen der 8 Elemente dieses Symbolzeichens darstellen. Bild 3 zeigt ein Symbolzeichen mit der Strich-Lücken-Folge 51111125.

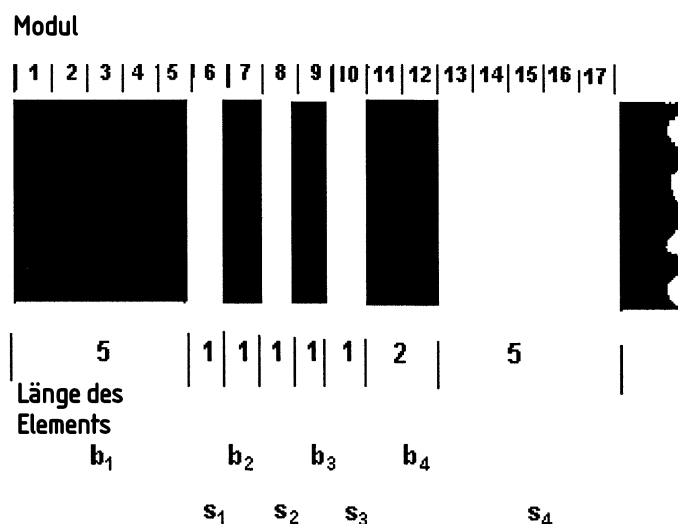


Bild 3 — PDF417-Symbolzeichen

Es gibt 929 definierte Symbolzeichenwerte (Codewörter), die von 0 bis 928 nummeriert werden.

Die Codewörter werden durch drei sich gegenseitig ausschließende Symbolzeichensätze oder Cluster dargestellt. Jedes Cluster codiert die in PDF417 929 verfügbaren Codewörter in verschiedenen Strich-Lücken-Mustern, so dass ein Cluster sich vom anderen unterscheidet. Die Clusternummern sind 0, 3, 6. Die Cluster-Definition gilt für alle PDF417-Symbolzeichen außer für die Start- und Stoppzeichen.

Die Clusternummer K wird durch folgende Gleichung ermittelt:

$$K = (b_1 - b_2 + b_3 - b_4 + 9) \bmod 9$$

wobei b_1, b_2, b_3 bzw. b_4 die Längen der vier Strichelemente in Modulen darstellen.

Die Clusternummer K für das Symbolzeichen in Bild 3 ist:

$$K = (5 - 1 + 1 - 2 + 9) \bmod 9 = 3$$

Die Codewörter und die Strich-Lücken-Folge für jedes Cluster der Symbolzeichen sind in Anhang A aufgeführt.

4.3.2 Start- und Stoppzeichen

Die Start- und Stoppzeichen müssen so zusammengesetzt werden, wie in Tabelle 1 festgelegt und in Bild 4 dargestellt:

Tabelle 1 — Strich-Lücken-Folge für Start- und Stoppzeichen

Zeichen	Strich-Lücken-Folge								
	B	S	B	S	B	S	B	S	B
Start	8	1	1	1	1	1	1	3	
Stopp	7	1	1	3	1	1	1	2	1

ANMERKUNG 1 Die Stopp- und Startzeichen in PDF417 sind eindeutig, da sie Elemente mit einer Länge von mehr als 6 Modulen haben.

ANMERKUNG 2 Das Stoppzeichen hat ein zusätzliches Strichelement aus einem einzelnen Modul.

Die Start- und Stoppzeichen müssen in allen Zeilen die gleiche Strich-Lücken-Folge aufweisen.



Bild 4 — Start- und Stoppzeichen in PDF417

4.4 Höhere (Daten-)Codierung

Mit der höheren Codierung werden die Datenzeichen in ihre entsprechenden Codewörter umgesetzt.

Für die höhere Codierung müssen Datenverdichtungsprogramme benutzt werden. Nachfolgend werden drei Modi festgelegt, von denen jeder eine besonders wirksame Abbildung von anwenderdefinierten Daten auf Codewort-Folgen ermöglicht:

- Textverdichtungsmodus (siehe Abschnitt 4.4.2);
- Byte-Verdichtungsmodus (siehe Abschnitt 4.4.3);
- Numerischer Verdichtungsmodus (siehe Abschnitt 4.4.4).

Eine gegebene Zeichenfolge von Datenbytes kann durch verschiedene Codewort-Folgen dargestellt werden, abhängig davon, wie der Codierer zwischen den Verdichtungsmodi und -untermodi umschaltet. Die Codierung von Daten in ein PDF417-Symbol erfolgt nicht nur auf eine einzige Art und Weise.

In jedem Modus sind für die Datencodierung und andere Funktionen innerhalb des Modus 900 Codewörter verfügbar. Die verbleibenden 29 Codewörter werden unabhängig vom gerade benutzten Verdichtungsmodus besonderen Funktionen zugeordnet (siehe Abschnitt 4.4.1).

PDF417 unterstützt auch das System der erweiterten Kanalübersetzung, das für die genaue Codierung im Symbol unterschiedliche Übersetzungen von Daten zulässt (siehe Abschnitt 4.5).

4.4.1 Funktions-Codewörter

Die Codewörter 900 bis 928 werden wie folgt als Funktions-Codewörter zugeordnet:

- um zwischen den Modi umzuschalten (siehe Abschnitt 4.4.1.1);
- für erweiterte Anwendungen unter Verwendung der erweiterten Kanalübersetzungen (siehe Abschnitt 4.4.1.2);
- für andere erweiterte Anwendungen (siehe Abschnitte 4.4.1.3 und 4.4.1.4).

Zum jetzigen Zeitpunkt sind die Codewörter 903 bis 912 und 914 bis 920 reserviert. Tabelle 2 zeigt die vollständige Liste der zugeordneten und reservierten Funktions-Codewörter. Die Funktionen werden in den Abschnitten 4.4.1.1 bis 4.4.1.4 beschrieben. Behandlung von reservierten Codewörtern siehe Abschnitt 4.4.6.

Tabelle 2 — Zuordnung von Funktions-Codewörtern in PDF417

Codewort	Funktion	Siehe Abschnitt
900	Moduswechsel zum Textverdichtungsmodus	4.4.1.1
901	Moduswechsel zum Byte-Verdichtungsmodus	4.4.1.1
902	Moduswechsel zum numerischen Verdichtungsmodus	4.4.1.1
903 bis 912	Reserviert	
913	Modusumschaltung zum Byte-Verdichtungsmodus	4.4.1.1
914 bis 920	reserviert	
921	Initialisierung des Lesegeräts	4.4.1.4
922	Abschluss-Codewort für den Macro-PDF-Steuerblock	4.13
923	Identifizierungskennzeichen der Folge zur Kennzeichnung des Anfangs von Optionsfeldern im Macro-PDF-Steuerblock	4.13
924	Moduswechsel zum Byte-Verdichtungsmodus (andere Anwendung als bei 901)	4.4.1.1
925	Identifikator für eine anwenderdefinierte erweiterte Kanalübersetzung	4.5
926	Identifikator eines allgemeinen ECI-Formats	4.5
927	Identifikator für eine ECI eines Zeichensatzes oder einer Codeseite	4.5
928	Codewort zur Macrokennzeichnung, um den Anfang eines Macro-PDF-Steuerblockes anzuzeigen	4.13

4.4.1.1 Funktionscodewörter für die Modusumschaltung

In einem PDF417-Symbol ist es möglich, zwischen den Modi so oft wie gewünscht umzuschalten. Hinweise zur Auswahl der zutreffenden Modi siehe Abschnitt 4.4.5.

Ein Moduswechsel-Codewort kann verwendet werden, um vom aktuellen Modus zum angezeigten Zielmodus umzuschalten, der solange wirksam bleibt, bis eine andere Modusumschaltung ausdrücklich zur Anwendung kommt. Dieser Funktion werden die Codewörter 900 bis 902 und 924 zugeordnet. Tabelle 3 gibt die Funktionen an.

Das Modusumschalt-Codewort 913 bewirkt eine vorübergehende Umschaltung vom Textverdichtungsmodus zum Byte-Verdichtungsmodus. Diese Umschaltung bleibt nur für das nächsten Codewort aktiv, danach muss der Modus zum herrschenden Untermodus der Textverdichtung zurückkehren. Codewort 913 ist nur im Textverdichtungsmodus verfügbar, seine Anwendung wird in Abschnitt 4.4.2.4 beschrieben.

Tabelle 3 — Modusdefinition und Modusumschalt-Codewörter

Zielmodus	Moduswechsel	Modusumschaltung
Textverdichtung	900	
Byte-Verdichtung	901/924	913
Numerische Verdichtung	902	

ANMERKUNG Die Tabelle definiert die Codewörter, die zur Umschaltung zum vorgegebenen Modus verwendet werden müssen.

Die Umschaltregeln zwischen den drei Modi sind in Tabelle 4 angegeben und in Bild 5 schematisch dargestellt.

Tabelle 4 — Modusübergangstabelle mit Codewörtern und deren Funktion

Ursprünglicher Modus	Zielmodus		
	Text	Byte	numerisch
Text	Moduswechsel 900	Modusumschaltung 913 Moduswechsel 901 Moduswechsel 924	Moduswechsel 902
Byte	Moduswechsel 900	Moduswechsel 901 Moduswechsel 924	Moduswechsel 902
numerisch	Moduswechsel 900	Moduswechsel 901 Moduswechsel 924	Moduswechsel 902

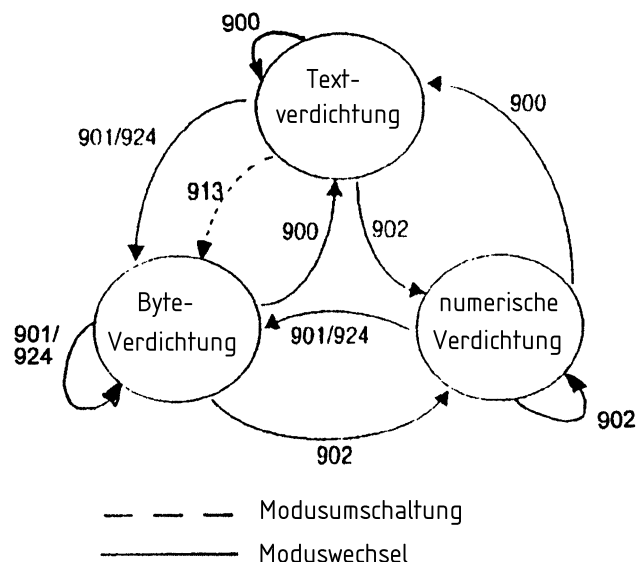


Bild 5 — Verfügbare Modusumschaltungen

Die Umschaltregeln zum Byte-Verdichtungsmodus sind in Abschnitt 4.4.3.1 umfassend festgelegt.

4.4.1.2 Funktions-Codewörter für die Umschaltung zur erweiterten Kanalübersetzung

Ein ECI-Codewort kann verwendet werden, um zu einer bestimmten Übersetzung umzuschalten, die solange aktiv bleibt, bis ein anderes ECI-Codewort ausdrücklich zur Anwendung kommt oder bis zum Datenende. Dieser Funktion werden die Codewörter 925 bis 927 zugeordnet (siehe Abschnitt 4.5).

4.4.1.3 Funktions-Codewörter für Macro-PDF417

Macro-PDF417-Symbole (siehe Abschnitt 4.13) müssen am Anfang des Macro-PDF417-Steuerblockes das Codewort 928 verwenden. Die Codewörter 922 und 923 werden für Sonderfunktionen im Macro-PDF417 verwendet.

4.4.1.4 Funktions-Codewörter zur Initialisierung des Lesegeräts

Das Codewort 921 muss verwendet werden, um das Lesegerät anzuweisen, die im Symbol enthaltenen Daten zur Programmierung für die Initialisierung des Lesegeräts zu interpretieren. Codewort 921 muss das erste Codewort nach dem Symbolhängendeskriptor sein. Bei einer Macro-PDF417-Initialisierungsfolge muss das Codewort 921 in jedem Symbol auftreten.

Die in einem Initialisierungssymbol oder einer Symbolfolge enthaltenen Daten dürfen vom Lesegerät nicht übertragen werden.

4.4.2 Textverdichtungsmodus

Der Textverdichtungsmodus umfasst alle druckbaren ASCII-Zeichen (d. h. Werte von 32 bis 126) und drei ASCII-Steuerzeichen: HT oder Horizontaltabulator (ASCII-Wert 9), LF oder Zeilenvorschub (ASCII-Wert 10) und CR oder Wagenrücklauf (ASCII-Wert 13). Der Textverdichtungsmodus umfasst auch verschiedene Wechsel- und Umschaltzeichen, die ausschließlich für diesen Modus verwendet werden.

Der Textverdichtungsmodus codiert bis zu 2 Zeichen pro Codewort. Die Verdichtungsregeln zur Umwandlung von Daten in PDF417-Codewörter sind in Abschnitt 4.4.2.2 festgelegt. Die Umschaltungen in Untermodi sind in Abschnitt 4.4.2.3 festgelegt.

4.4.2.1 Untermodi der Textverdichtung

Der Textverdichtungsmodus hat vier Untermodi:

- Alpha (großgeschriebenes Alphabet);
- Kleinschreibung (kleingeschriebenes Alphabet);
- Gemischt (numerische Zeichen und einige Interpunktionszeichen);
- Interpunktion.

Jeder Untermodus umfasst 30 Zeichen, einschließlich Wechsel- und Umschaltzeichen für Untermodi.

Der Standard-Verdichtungsmodus für PDF417, der am Anfang jedes Symbols aktiv ist, muss immer der Textverdichtungsmodus, Untermodus Alpha (großgeschriebenes Alphabet), sein. Ein Wechsel-Codewort von einem anderen Modus in den Textverdichtungsmodus muss immer in den Textverdichtungsmodus, Untermodus Alpha, umschalten.

Alle Zeichen und ihre Werte sind in Tabelle 5 angegeben.

Tabelle 5 — Definition der Untermodi der Textverdichtung

30 Grundwerte	Untermodi der Textverdichtung								
	Alpha		Kleinschreibung		Gemischt		Interpunktion		
	Zeichen	ASCII	Zeichen	ASCII	Zeichen	ASCII	Zeichen	ASCII	
0	A	65	a	97	0	48	;	59	
1	B	66	b	98	1	49	<	60	
2	C	67	c	99	2	50	>	62	
3	D	68	d	100	3	51	@	64	
4	E	69	e	101	4	52	[	91	
5	F	70	f	102	5	53	\	92	
6	G	71	g	103	6	54	]	93	
7	H	72	h	104	7	55	_	95	
8	I	73	i	105	8	56	'	96	
9	J	74	j	106	9	57	~	126	
10	K	75	k	107	&	38	!	33	
11	L	76	l	108	CR	13	CR	13	
12	M	77	m	109	HT	9	HT	9	
13	N	78	n	110	,	44	,	44	
14	O	79	o	111	:	58	:	58	
15	P	80	p	112	#	35	LF	10	
16	Q	81	q	113	-	45	-	45	
17	R	82	r	114	.	46	.	46	
18	S	83	s	115	\$	36	\$	36	
19	T	84	t	116	/	47	/	47	
20	U	85	u	117	+	43	“	34	
21	V	86	v	118	%	37		124	
22	W	87	w	119	*	42	*	42	
23	X	88	x	120	=	61	(	40	
24	Y	89	y	121	^	94	)	41	
25	Z	90	z	122	pl		?	63	
26	Leer- zeichen	32	Leer- zeichen	32	Leerzeichen	32	{	123	
27	ll		As		ll		}	125	
28	ml		Ml		al		'	39	
29	ps		Ps		ps		al		
al	Wechsel zu Alpha				ml	Wechsel zu gemischt			
as	Umschaltung zu Alpha				pl	Wechsel zu Interpunktion			
ll	Wechsel zu Kleinschreibung				ps	Umschaltung zu Interpunktion			

ANMERKUNG Die Zeichen-Spalten zeigen die Standard-Übersetzung von GLI 0 und ECI 000002 der in den angrenzenden ASCII-Spalten gezeigten Byte-Werte. Jeder Tabelleneintrag stellt die Hälfte eines Codewortes dar, d. h. einen Wert zwischen 0 und 29 (siehe Abschnitt 4.4.2.2).

4.4.2.2 Verdichtungsregeln für die Codierung im Textverdichtungsmodus

Im Textverdichtungsmodus werden Paare von Datenzeichen durch ein einziges Codewort dargestellt. Die den Datenzeichen zugeordneten Werte liegen zwischen 0 und 29 (d. h. 30 Grundwerte) und sind in Tabelle 5 angegeben. Für jedes Paar der 30 Grundwerte muss der erste oder linke Wert den höheren Wert h bezeichnen, der andere den niedrigeren Wert l .

Das codierte PDF417-Codewort wird mit folgender Gleichung festgelegt:

$$d = h \times 30 + l$$

Dabei entspricht d der Definition in Abschnitt 3.2.

Die Gleichung gilt auch für Werte der 30 Grundwerte für Umschaltungen und Wechsel in den Textverdichtungsmodus. Zwischen den Untermodi müssen geeignete Wechsel- und Umschaltwerte verwendet werden. Falls die Codierung der Zeichenfolge keine gerade Anzahl aus den 30 Grundwerten ergibt, siehe Abschnitt 4.4.2.4 für besondere Anwendungsmechanismen.

Das folgende Beispiel zeigt, wie im Textverdichtungsmodus eine Verdichtung erzielt wird.

BEISPIEL Zu codierende Daten: PDF417

Tabelle 6 — Beispiel für die Codierung mit Textverdichtung

Zeichenpaare	h	l	$h \times 30 + l$	Codewort
P D	15	3	$15 \times 30 + 3$	453
F ml	5	28	$5 \times 30 + 28$	178
4 1	4	1	$4 \times 30 + 1$	121
7 ps	7	29	$7 \times 30 + 29$	239

ANMERKUNG 1 ml (Wechsel zum Untermodus gemischt) wird verwendet, um zur Codierung der numerischen Zeichen umzuschalten.

ANMERKUNG 2 ps wird in diesem Beispiel als Füllwert verwendet, andere Umschalt- und Wechselwerte können verwendet werden (siehe Abschnitt 4.4.2.4).

Die Daten des Beispiels PDF417 wird durch die Codewörter 453, 178, 121, 239 dargestellt.

4.4.2.3 Umschaltung in die Untermodi der Textverdichtung : Wechsel- und Umschaltfunktion

Die Umschaltung von einem Untermodus zu einem anderen innerhalb des Textverdichtungsmodus muss durch die Wechsel- und Umschaltwerte erfolgen, die für den aktiven Untermodus vor der Umschaltung festgelegt sind.

Eine Umschaltung in den Untermodus wird verwendet, um für nur ein Datenzeichen von einem Untermodus der Textverdichtung zu einem anderen umzuschalten. Die nachfolgenden Codewörter kehren sofort nach dieser Umschaltung zum vorher verwendeten Untermodus zurück (es sei denn, es wird ps als Pad-Zeichen verwendet, siehe hierzu Abschnitt 4.4.2.4). Die Umschaltfunktionen sind folgende:

ps = Umschaltung in den Untermodus Interpunktion

as = Umschaltung in den Untermodus großgeschriebenes Alphabet

Eine Untermoduswechsel wird verwendet, um von einem Untermodus der Textverdichtung zu einem anderen umzuschalten, der solange aktiv bleibt, bis ein anderer Wechsel oder eine andere Umschaltung ausdrücklich zur Anwendung kommt. Die Wechselfunktionen sind folgende:

al = Wechsel in den Untermodus großgeschriebenes Alphabet

ll = Wechsel in den Untermodus kleingeschriebenes Alphabet

ml = Wechsel in den Untermodus gemischt (numerische Zeichen oder einige Interpunktionszeichen)

pl = Wechsel in den Untermodus Interpunktion

Innerhalb jedes Untermodus der Textverdichtung ist ein begrenzter Satz von Wechsel- und Umschaltfunktionen verfügbar. Die verfügbaren Funktionen sind in Tabelle 5 aufgeführt. Tabelle 7 zeigt die Übergangstabelle zwischen Untermodi der Textverdichtung; Bild 6 ist eine schematische Darstellung.

ANMERKUNG Auf einen Wechsel in einen Untermodus darf ein weiterer Wechsel in einen Untermodus oder eine Umschaltung in einen Untermodus folgen; aber auf eine Umschaltung in einen Untermodus darf weder eine Umschaltung noch ein Wechsel in einen Untermodus folgen.

Tabelle 7 — Übergangstabelle zu den Untermodi der Textverdichtung

ursprünglicher Untermodus	Ziel-Untermodus			
	Alpha	Kleinschreibung	Gemischt	Interpunktion
Alpha		ll	ml	Ps
Kleinschreibung	as		ml	Ps
Gemischt	al	ll		Ps Pl
Interpunktion	al			

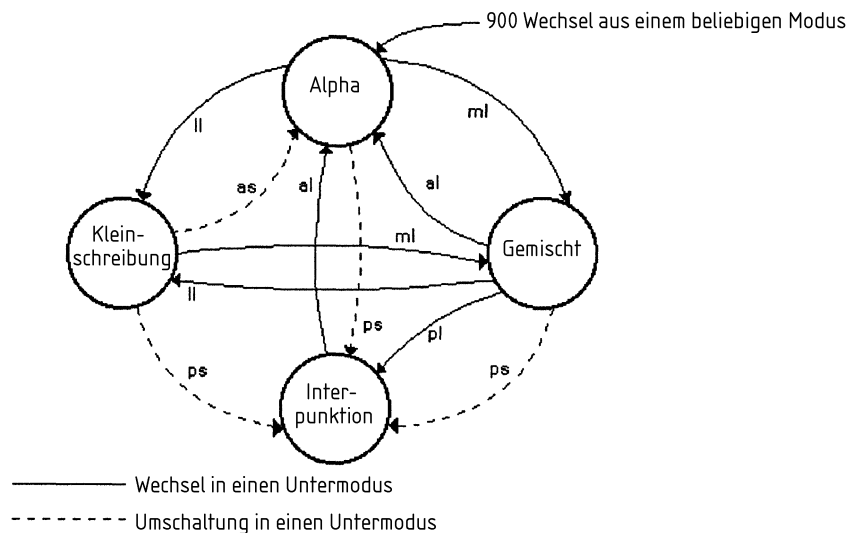


Bild 6 — Umschaltung in einen Untermodus der Textverdichtung

4.4.2.4 Mechanismus für die Verwendung eines Pad-Zeichens im Textverdichtungsmodus

Wenn die Zeichenfolge der Textverdichtung keine gerade Anzahl aus 30 Grundwerten ergibt, muss am Ende der Zeichenfolge ein Pad-Zeichen hinzugefügt werden. Ein Beispiel hierfür wird in Tabelle 6 angegeben. Da es im Textverdichtungsmodus keine besonderen Nullfunktionen gibt, müssen Umschaltung und Wechsel in die Untermodi nach den Mechanismen für die folgenden Fälle verwendet werden.

- Wenn sich die Zeichenfolge bis zum Ende der Daten fortsetzt oder der Zeichenfolge des Textverdichtungsmodus ein Wechsel zu einem anderen Verdichtungsmodus folgt, dann kann das Pad-Zeichen ein beliebiges Zeichen für Umschaltungen oder Wechsel in den Untermodus sein.

- b) Wenn der Zeichenfolge des Textverdichtungsmodus eine Byte-Umschaltung (Codewort 913) zur Codierung eines einzelnen Byte-Verdichtungsmoduszeichens folgt, können vor der Umschaltung zur Byte-Verdichtung abhängig vom Untermodus der Textverdichtung zwei Mechanismen angewendet werden:
- 1) Wenn der Untermodus der Textverdichtung nicht die Interpunktion ist, dann sollte der Wert 29 (ps) aus den 30 Grundwerten benutzt werden, wenn die Codierung in den selben Untermodus der Textverdichtung zurückkehren soll. Der Decoder muss ein (ps), dass dem Codewort 913 direkt vorangeht, ignorieren.
 - 2) Wenn der Untermodus der Textverdichtung die Interpunktion ist, dann muss der Wert 29 (al) aus den 30 Grundwerten benutzt werden. Der Decoder darf das (al) nicht ignorieren und wird daher in den Untermodus Alpha zurückkehren.

4.4.2.5 Umschaltung vom Textverdichtungsmodus

Der Textverdichtungsmodus kann durch das Symbolende oder jedes der folgenden Codewörter abgeschlossen werden:

- 900 (Wechsel in den Textverdichtungsmodus)
- 901 (Wechsel in den Byte-Verdichtungsmodus)
- 902 (Wechsel in den numerischen Verdichtungsmodus)
- 924 (Wechsel in den Byte-Verdichtungsmodus)
- 928 (Beginn des Macro-PDF417-Steuerblockes)
- 923 (Beginn des Macro-PDF417-Optionsfeldes)
- 922 (Macro-PDF417-Terminator)

Die letzten drei Codewörter treten nur innerhalb des Macro-PDF417-Steuerblockes eines Macro-PDF417-Symbols auf (siehe Abschnitt 4.13.1). Der Textverdichtungsmodus wird auch durch das Vorhandensein eines reservierten Codewortes beeinflusst (siehe Abschnitt 4.4.6).

Wenn der Decoder im Textverdichtungsmodus arbeitet und auf das Codewort 913 (Umschaltung in den Byte-Verdichtungsmodus) trifft, decodiert es das Codewort nach Codewort 913 als ein einzelnes binäres Byte und kehrt dann in den Textverdichtungsmodus zurück. Der Untermodus, in den der Decoder zurückkehrt, ist der Untermodus, in den zuletzt gewechselt wurde und der vor Codewort 913 aktiv war; ein ps-Wert für die Untermodus-Umschaltung unmittelbar vor Codewort 913 wird ignoriert.

Wenn der Decoder im Textverdichtungsmodus arbeitet und auf das Codewort 900 (Wechsel in den Textverdichtungsmodus) trifft, richtet der Decoder erneut Untermodus Alpha ein.

4.4.3 Byte-Verdichtungsmodus

Der Byte-Verdichtungsmodus ermöglicht die Codierung einer Folge von 8-Bit-Bytes in eine Folge von Codewörtern. Diese Codierung wird durch die Umwandlung von 256 Grundwerten in 900 Grundwerte durchgeführt, womit ein Verdichtungsverhältnis von 6 Bytes in 5 Codewörter (1,2:1) erreicht wird.

Alle Zeichen und deren Werte (0 bis 255) sind in Anhang B festgelegt. Diese müssen als Standardübersetzungen für die Graphik- und Steuerzeichen angesehen werden. Wenn ECIs aufgerufen werden (siehe Abschnitt 4.5), kann diese Übersetzung entweder als ECI 000000 oder ECI 000002 (siehe Abschnitt 4.5.2) definiert werden.

4.4.3.1 Umschaltung in den Byte-Verdichtungsmodus

Zum Umschalten in den Byte-Verdichtungsmodus aus dem Text- oder numerischen Verdichtungsmodus muss eins der folgenden Codewörter benutzt werden:

- Moduswechsel 924 muss verwendet werden, wenn die gesamte Anzahl von zu codierenden Byte-Verdichtungszeichen ein ganzzahliges Vielfaches von 6 ist.
- Moduswechsel 901 muss verwendet werden, wenn die gesamte Anzahl von zu codierenden Byte-Verdichtungszeichen kein Vielfaches von 6 ist.
- Modusumschaltung 913 kann anstatt des Codewortes 901 verwendet werden, wenn nur ein einzelnes Byte-Verdichtungszeichen zu codieren ist.

4.4.3.2 Verdichtungsregeln für die Codierung eines einzelnen Byte-Verdichtungszeichens (unter Verwendung der Modusumschaltung 913)

Um ein einzelnes Byte-Verdichtungszeichen zu codieren, muss das Codewort der Dezimalwert (0 bis 255) des in Anhang B definierten Zeichen sein.

4.4.3.3 Verdichtungsregeln für die Codierung längerer Byte-Verdichtungszeichenfolgen (unter Verwendung des Moduswechsels 924 oder 901)

Folgendes Verfahren muss zur Codierung von Byte-Verdichtungsdaten verwendet werden:

- 1) Zusammenstellen der Gesamtzahl der Byte-Verdichtungszeichen.
- 2) Bei einem ganzzahligen Vielfachen von 6 muss der Moduswechsel 924 verwendet werden; ansonsten Moduswechsel 901.
- 3) Unterteilen der Anzahl von Byte-Verdichtungszeichen in eine Folge von 6 Zeichen von links nach rechts (von höchstwertigen zum niederwertigsten Zeichen). Bei weniger als 6 Zeichen wird mit Schritt 7) weitergemacht.
- 4) Zuordnen der Dezimalwerte der 6 Datenbytes, die im Byte-Verdichtungsmodus als b_5 bis b_0 codiert werden sollen (dabei ist b_5 das erste Datenbyte).
- 5) Umwandlung der 256 Grundwerte in 900 Grundwerte und Erzeugung einer Folge von 5 Codewörtern. Anhang C definiert einen Algorithmus und liefert ein Beispiel.
- 6) Wiederholung ab Schritt 3, falls notwendig.
- 7) Für die verbleibenden Byte-Verdichtungszeichen bei der Verwendung des Moduswechsels 901 (d. h. wenn die letzte Gruppe weniger als 6 Byte-Verdichtungszeichen hat), muss das Codewort der Dezimalwert (müssen die Codewörter die Dezimalwerte) (0 bis 255) des Zeichens (der Zeichen) sein, wie in Anhang B definiert, vom höchstwertigen zum niederwertigsten Zeichen.

ANMERKUNG Der Byte-Verdichtungsmodus mit dem Modewechsel 901 setzt voraus, dass die Gesamtanzahl der zu codierenden Bytes kein Vielfaches von sechs beträgt. Wenn die Anzahl der im Byte-Verdichtungsmodus zu codierenden Bytes ein ganzzahliges Vielfaches von sechs ist, muss entweder ein Wechsel zum Byte-Verdichtungsmodus 901 oder 924 codiert werden und an einer Stelle im Symbol angeordnet werden, die eine richtige Codierung nach diesen Codierungsregeln erzeugen würde. Beispielsweise würde ein Codewort 924 als erstes oder zweites Codewort den folgenden Strom der Codewörter des Byte-Verdichtungsmodus als Codierung einer Byteanzahl erkennen, die ein Vielfaches von sechs ist. 901 dagegen könnte an jeder Stelle innerhalb des Codewortstroms des Byte-Verdichtungsmodus angeordnet werden, die den Strom in zwei Teile aufteilen würde, von denen keiner eine Anzahl codiert, die ein Vielfaches von sechs ist.

Falls eine zusätzliche Codierung im Textverdichtungsmodus oder numerischen Verdichtungsmodus erforderlich ist, muss das entsprechende Wechselzeichen verwendet werden (siehe Abschnitt 4.4.1.1).

4.4.3.4 Umschaltung aus dem Byte-Verdichtungsmodus

Der Byte-Verdichtungsmodus kann durch das Symbolende oder jedes der folgenden Codewörter abgeschlossen werden:

- 900 (Wechsel in den Textverdichtungsmodus)
- 901 (Wechsel in den Byte-Verdichtungsmodus)
- 902 (Wechsel in den numerischen Verdichtungsmodus)
- 924 (Wechsel in den Byte-Verdichtungsmodus)
- 928 (Beginn des Macro-PDF417-Steuerblockes)
- 923 (Beginn des Macro-PDF417-Optionsfeldes)
- 922 (Macro-PDF417-Terminator)

Die letzten drei Codewörter treten nur innerhalb des Macro-PDF417-Steuerblockes eines Macro-PDF417-Symbols auf (siehe Abschnitt 4.13.1). Der Byte-Verdichtungsmodus wird auch durch das Vorhandensein eines reservierten Codewortes beeinflusst (siehe Abschnitt 4.4.6).

Der Wiederaufruf des Byte-Verdichtungsmodus (durch Verwendung des Codewortes 901 oder 924 aus dem Byte-Verdichtungsmodus) beendet den vorhergehenden Byte-Verdichtungsmodus, in dem Gruppen von 6 Byte-Verdichtungszeichen erstellt werden, wie in Abschnitt 4.4.3.3 beschrieben, und dann beginnt eine neue Gruppierung. Dieses Verfahren kann notwendig sein, wenn eine ECI-Zuordnungsnummer codiert werden muss (siehe Abschnitt 4.5.3.2).

Während des Decodierungsprozesses für den Byte-Verdichtungsmodus hängt die Behandlung der letzten Gruppe von Codewörtern davon ab, ob der Byte-Verdichtungsmodus mit den Codewörtern 901 oder 924 aufgerufen wird.

Wenn der Byte-Verdichtungsmodus mit dem Codewort 924 aufgerufen wird, muss die Gesamtzahl von Codewörtern innerhalb des Verdichtungsmodus ein Vielfaches von fünf sein. Ist das nicht der Fall, ist das Symbol ungültig. Alle Gruppen mit 5 Codewörtern werden in Gruppen mit 6 Bytes codiert.

Wenn der Byte-Verdichtungsmodus mit dem Codewort 901 aufgerufen wird, wird die letzte Gruppe von Codewörtern direkt mit einem Byte je Codewort ohne Verdichtung interpretiert. Wenn also die letzte Gruppe aus fünf Codewörtern besteht, wird die Gruppe mit 5 Bytes interpretiert anstatt mit 6 Bytes.

4.4.4 Numerischer Verdichtungsmodus

Der numerische Verdichtungsmodus ist eine Methode zur Datenverdichtung von 10 Grundwerten in 900 Grundwerte und sollte zur Codierung langer Zeichenfolgen aufeinanderfolgender Ziffern verwendet werden. Der numerische Verdichtungsmodus codiert bis zu 2,93 Ziffern je Codewort.

4.4.4.1 Wechsel in den numerischen Verdichtungsmodus

Der numerische Verdichtungsmodus darf aus dem Textverdichtungsmodus oder dem Byte-Verdichtungsmodus mit dem Moduswechsel 902 aufgerufen werden.

4.4.4.2 Verdichtungsregeln für die Codierung von langen Zeichenfolgen aufeinanderfolgender Ziffern

Folgendes Verfahren muss zur Verdichtung numerischer Daten verwendet werden:

- 1) Teilen der Ziffernfolge in Gruppen von 44 Ziffern, bis auf die letzte Gruppe, die weniger enthalten darf.
- 2) Einfügen der Ziffer 1 an der höchstwertigen Position jeder Gruppe, um den Verlust von führende Nullen zu vermeiden.

BEISPIEL	ursprüngliche Daten	00246812345678
	nach Schritt 2)	1 00246812345678

ANMERKUNG Die Führungsziffer 1 wird im Decodieralgorithmus entfernt.

- 3) Umwandlung der 10 Grundwerte in 900 Grundwerte. Anhang D legt hierfür einen Algorithmus fest und zeigt ein Beispiel.
- 4) Wiederholung ab Schritt 2), falls notwendig.

Die folgenden Regeln können verwendet werden, um die genaue Anzahl von Codewörtern im numerischen Verdichtungsmodus zu bestimmen:

- Gruppen von 44 Ziffern werden auf 15 Codewörter verdichtet.
- Für Gruppen mit kürzeren Ziffernfolgen kann die Anzahl von Codewörtern wie folgt errechnet werden:

Codewörter = $\text{INT}(\text{Anzahl von Ziffern}/3) + 1$

BEISPIEL Für eine 28-stellige Ziffernfolge

$\text{INT}(28/3) + 1$

$= 9 + 1$

$= 10$ Codewörter

4.4.4.3 Umschaltung aus dem numerischen Verdichtungsmodus

Der numerische Verdichtungsmodus kann durch das Symbolende oder jedes der folgenden Codewörter abgeschlossen werden:

- 900 (Wechsel in den Textverdichtungsmodus)
- 901 (Wechsel in den Byte-Verdichtungsmodus)
- 902 (Wechsel in den numerischen Verdichtungsmodus)
- 924 (Wechsel in den Byte-Verdichtungsmodus)
- 928 (Beginn des Macro-PDF417-Steuerblockes)
- 923 (Beginn des Macro-PDF417-Optionsfeldes)
- 922 (Macro-PDF417-Terminator)

Die letzten drei Codewörter treten nur innerhalb des Macro-PDF417-Steuerblockes eines Macro-PDF417-Symbols auf (siehe Abschnitt 4.13.1). Der numerische Verdichtungsmodus wird auch durch das Vorhandensein eines reservierten Codewortes beeinflusst (siehe Abschnitt 4.4.6).

Der Wiederaufruf des numerischen Verdichtungsmodus (durch Verwendung des Codewortes 902 aus dem numerischen Verdichtungsmodus) beendet den aktuellen numerischen Verdichtungsmodus, in dem Gruppen nach Abschnitt 4.4.4.2 erstellt werden, und dann beginnt eine neue Gruppierung. Dieses Verfahren kann notwendig sein, wenn eine ECI-Zuordnungsnummer codiert werden muss (siehe Abschnitt 4.5.3.4).

Während des Decodierungsprozesses für den numerischen Verdichtungsmodus muss das Ergebnis der Umwandlung der 900 Grundwerte in 10 Grundwerte eine Zahl ergeben, deren höchstwertige Ziffer eine '1' ist. Wenn die Umwandlung der 900 Grundwerte in 10 Grundwerte keine Zahl ergibt, die mit '1' beginnt, muss das Symbol als ungültig behandelt werden. Die führende '1' wird entfernt, um die ursprüngliche Zahl zu erhalten.

4.4.5 Hinweis zur Auswahl des geeigneten Verdichtungsmodus

Alle Grundimplementierungen für das Drucken und Scannen von PDF417-Symbolen müssen die folgenden drei Modi unterstützen: Textverdichtung, Byte-Verdichtung und numerische Verdichtung. Der Standard-Zeichensatz für die Textverdichtung ist in Tabelle 5 festgelegt, der für die Byte-Verdichtung wie in Anhang B angegeben. Für die Codierung von Standard-ASCII-Textdateien ist der Textverdichtungsmodus aufgrund der besseren Verdichtung der ASCII-Zeichenwerte 9, 10, 13 und 32 bis 126 normalerweise wirksamer als der Byte-Verdichtungsmodus.

Der numerische Verdichtungsmodus sollte für lange numerische Zeichenfolgen verwendet werden.

Hinweise zum Umschalten zwischen den Modi zur Verringerung der Anzahl der Codewörter sind im Anhang P als Algorithmus angegeben.

4.4.6 Behandlung der reservierten Codewörter in PDF417

4.4.6.1 Überblick

PDF417-Symbole, die zur Verwendung in offenen Systemen vorgesehen sind, dürfen keines der Codewörter enthalten, die in einer veröffentlichten Ausgabe dieser Norm als reserviert angegeben sind (siehe Anhang 4.4.1). Dennoch sollten Decodierungsgeräte die Übertragung von reservierten Codewörtern unter Verwendung von Escape-Folgen nach Abschnitt 4.17.4 unterstützen. Decoder können auch eine Option unterstützen, nach der solche Symbole als ungültig behandelt werden, wie es beim Betrieb im Grundkanalmodus der Fall ist.

Empfangssysteme sollten Daten verwerfen, die Escape-Folgen mit reservierten Codewörtern enthalten, bis das System die neue Definition eines vorher reservierten Codewortes kennt.

4.4.6.2 Zukünftige Verwendung von reservierten Codewörtern

Jedes neue Funktions-Codewort, das in zukünftigen Überarbeitungen dieser Norm festgelegt wird, muss festgelegte Codierungsregeln haben, um Abwärtskompatibilität mit bereits bestehenden Geräten sicherzustellen. Vor allem gilt:

- Wenn ein neues Signal-Codewort (im Gegensatz zu einem neuen Codewort des Verdichtungsmodus) codiert wird, muss darauf sofort ein entsprechender Wechsel in den Verdichtungsmodus folgen, so dass die nachfolgenden verdichteten Codewörter als Bytestrom und nicht als Reihe umgeschalteter, nicht übersetzter Codewörter übersetzt und übertragen werden. Mit diesem Ansatz werden mit Decodierungsgeräten, die sowohl der ursprünglichen Norm als auch dieser PDF417-Norm entsprechen, die gewünschten Ergebnisse erreicht, unabhängig davon, ob das Gerät das ursprüngliche oder das neue Übertragungsprotokoll verwendet.
- Im Empfangssystem wird der ECI-Decoder die Steuer-ECIs (d. h. Macro-Steuerblöcke und umgeschaltete, nicht übersetzte Codewörter) vor den übersetzenden ECIs (wie z.B. Codierungsschemen und Zeichensätze) verarbeiten. Daher sollte der Codierer die folgende Reihenfolge der Operationen beachten:
 - 1) Die Macro-Steuerblock-ECIs, falls vorhanden, werden verwendet, um den gesamten Bytestrom in der richtigen Reihenfolge anzuordnen.
 - 2) Die umgeschalteten verdichteten Codewörter werden vom ECI-Decoder nach den Regeln des neuen Verdichtungsmodus oder der Signal-ECI übersetzt und die daraus resultierenden Datenbytes werden an den passenden Platz im Bytestrom eingefügt.
 - 3) Zuletzt werden die Zeichensätze und andere Übersetzer-ECIs auf den sich daraus ergebenden Bytestrom angewendet.

4.5 Erweiterte Kanalübersetzung

Das erweiterte Kanalübersetzungsprotokoll ermöglicht andere Übersetzungen des Ausgangsdatenstroms als den Standard-Zeichensatz. Das ECI-Protokoll gilt einheitlich für eine Anzahl von Symbolologien einschließlich PDF417.

ANMERKUNG Ursprünglich wurde für PDF417 ein symbolologiespezifisches Schema mit der Bezeichnung 'universeller Etikettenidentifikator' (GLIs) definiert. Codierung und Decodierung von ECIs sind identisch zu den früheren Spezifikationen der GLIs in PDF417. Das Übertragungsprotokoll für decodierte Nachrichten nach den früheren PDF417-Spezifikationen für GLIs ist jedoch anders als das Übertragungsprotokoll für ECIs. Es gibt auch Unterschiede hinsichtlich der Verwendung von übersetzenden ECIs mit Macro-PDF417. Diese Norm erlaubt die Verwendung von früheren und aktuellen Protokollen in der Weise, dass alte und neue Geräte weiter nebeneinander bestehen können.

Fünf Typen von Übersetzungen werden in PDF417 unterstützt:

- a) Zeichensätze (oder Codeseiten)
- b) Übersetzungen für allgemeine Zwecke wie Datenverschlüsselung und Datenkompression (im Unterschied zu den Verdichtungsmodi der Symbologie)
- c) anwenderdefinierte Übersetzungen für geschlossene Systeme

- d) Übertragung von Steuerinformationen für Macro-PDF417
- e) Übertragung von nicht übersetzten PDF417-Codewörtern

Die Übertragung des erweiterten Kanalübersetzungsprotokolls ist vollständig im Dokument von AIM International „Extended Channel Interpretations — Part 1: Identification schemes and protocol“ festgelegt. Das Protokoll bietet eine logische Methode zur Festlegung besonderer Übersetzungen von Byte-Werten vor dem Ausdrucken und nach der Decodierung.

Die erweiterte Kanalübersetzung (ECI) wird durch eine 6-stellige Zahl gekennzeichnet, die im PDF417-Symbol durch eines von drei bestimmten Codewörtern codiert wird, gefolgt von einem oder zwei Codewörtern (siehe Abschnitt 4.5.1). Eine bestimmte ECI kann überall in der codierten Nachricht aufgerufen werden, die den Regeln der Verdichtungsmodi unterliegt (siehe Abschnitt 4.5.3).

Das ECI-Protokoll kann nur von Decodern verwendet werden, die für die Übertragung des Symbologie-Identifikators freigegeben sind (siehe Abschnitt 4.17.5). Decoder, die nicht für die Übertragung des Symbologie-Identifikators freigegeben sind, können die Escape-Folgen eines Symbols mit einem ECI nicht zuverlässig befördern.

4.5.1 Codierung der ECI-Zuordnungsnummern

Ein ECI kann überall im Datenstrom aufgerufen werden, unterliegt dabei aber den Bedingungen, die in Abschnitt 4.5.3 aufgeführt sind. Wenn ein ECI aufgerufen wurde, kann zwischen beliebigen Verdichtungsmodi umgeschaltet werden. Der verwendete Verdichtungsmodus wird genau durch die 8-Bit-Datenwerte bestimmt, die codiert wurden, und hängt nicht von der aktiven ECI ab. Ein Folge von Werten im Bereich 48 bis 57 (dezimal) wird z. B. am wirksamsten im numerischen Verdichtungsmodus codiert, auch wenn die Folge nicht mit Zahlen übersetzt wird.

Die ECI-Zuordnungsnummer wird in eine von drei ECI-Codewort-Folgen codiert, die mit den Codewörtern 927, 926 und 925 beginnen. Ein oder zwei zusätzliche Codewörter werden zur Codierung der ECI-Zuordnungsnummer verwendet. Die Codierungsregeln sind in Tabelle 8 festgelegt.

Tabelle 8 — Codierung der ECI-Zuordnungsnummern

ECI-Zuordnungsnummer	Codewort-Folge	Codewörter	Bereiche
000000 bis 000899	C_0 C_1	927 ECI_no	$C_1 = (0 \text{ bis } 899)$
000900 bis 810899	C_0 C_1 C_2	926 ECI_no div 900 – 1 ECI_no mod 900	$C_1 = (0 \text{ bis } 899)$ $C_2 = (0 \text{ bis } 899)$
810900 bis 811799	C_0 C_1	925 ECI_no – 810900	$C_1 = (0 \text{ bis } 899)$

In PDF417 sind 811 800 ECI-Zuordnungsnummern möglich.

ANMERKUNG Die Codierungsmethode entspricht dem GLI-Schema, das von der ursprünglichen PDF417-Spezifikation unterstützt wird und in die Spezifikationen AIM USA (1994) und AIM Europe (1994) aufgenommen wurde.

Das folgende Beispiel zeigt die Codierung:

ECI = 013579

Codewörter: [926] [(13579 div 900) – 1] [13579 mod 900]

= [926] [15 – 1] [79]

= [926] [14] [79]

4.5.2 Vorher zugeordnete erweiterte Kanalübersetzungen

Für die Abwärtskompatibilität mit vorhandenen Symbolologiespezifikationen einschließlich PDF417 wurden folgende ECIs vorher zugeordnet.

ECI 000000 (entspricht dem ursprünglichen GLI 0) stellt das Standard-Codierungsschema der Codierer dar, die der ursprünglichen PDF417-Norm entsprechen. Der Zeichensatz ist in Anhang B dargestellt.

ECI 000001 (entspricht dem ursprünglichen GLI 1) stellt das GLI-Codierungsschema einer Anzahl von Symbolologien mit den Zeichen 0 bis 127 nach ISO/IEC 646:1991 (entspricht ANSI X3.4) und den Zeichen 128 bis 255 nach den Werten von ISO 8859-1 dar.

ANMERKUNG ECI 000000 (entspricht GLI 0) und ECI 000001 (entspricht GLI 1) erfordern am Anfang jedes codierten Symbols eines Macro-PDF417-Symbolsatzes eine Verfahrenslogik mit einer Rückkehr zu GLI 0. Dieses Protokoll wird nicht von anderen erweiterten Kanalübersetzungen übernommen.

ECI 000002 hat die gleiche Codetabelle wie ECI 000000 (siehe Anhang B) ohne die Verfahrenslogik mit einer Rückkehr zu GLI 0. Es handelt sich hierbei um das Standard-Codierungsschema für Codierer, die vollkommen dieser Norm entsprechen.

ECI 000003 hat die gleiche Codetabelle wie ECI 000001 ohne die Verfahrenslogik mit einer Rückkehr zu GLI 0.

ECI 000000 und ECI 000001 dürfen mit Ausnahme von anwenderdefinierten ECIs nicht im gleichen PDF417-Symbol oder Macro-PDF417-Symbolsatz wie anderen ECIs codiert werden. ECI 000002 bzw. ECI 000003 sind die kompatiblen Alternativen zu ECI 000000 bzw. ECI 000001 und werden bevorzugt in neuen Anwendungen benutzt.

Nach den im Dokument AIM International "Extended Channel Interpretations – Part 2: Registration of coded character sets and other data formats" festgelegten Regeln werden weitere ECIs zugeordnet.

4.5.3 Codierung von ECI-Folgen innerhalb der Verdichtungsmodi

Das allgemeine Codierungsprinzip ist, dass ECIs auf den Quelldatenbytestrom angewendet werden (um verschiedene Übersetzungen anzuzeigen) und somit einen veränderten Bytestrom erzeugen, der unter Anwendung des Verdichtungsmodus der Symbologie in PDF417-Symbole codiert wird. Die ECI-Codierung und die symbologie-spezifische Verdichtung bilden zwei logisch unabhängige Schichten des Prozesses.

Obwohl ECI-Zuordnungen und Verdichtungsmodi im Allgemeinen vermischt werden dürfen, können einige Kombinationen ein unlogisches oder zweideutiges Verhalten hervorrufen. Die folgenden Abschnitte definieren, wie ECIs ohne Zweideutigkeiten durch Festlegung gültiger Anordnungen von ECI-Escape-Folgen eingefügt werden können.

4.5.3.1 ECIs und Textverdichtungsmodus

Eine ECI-Escape-Folge kann irgendwo im Textverdichtungsmodus angeordnet werden. Der Untermodus, der unmittelbar vor der ECI-Escape-Folge aufgerufen wurde, ist für die unmittelbar danach folgende Codierung reserviert. Somit werden Wechsel und Umschaltungen in den Untermodus über eine ECI-Escape-Folge aufrechterhalten und daher wird eine Umschaltung in einen Untermodus vor einer ECI-Escape-Folge beachtet.

4.5.3.2 ECIs und Byte-Verdichtungsmodus mit dem Moduswechsel 924 und 901

Bei einer Codierung im Byte-Verdichtungsmodus unter Verwendung des Moduswechsels 924 darf von einem Codierer eine ECI-Escape-Folge unmittelbar nach dem Codewort 924 oder an jeder Grenze nach 5 Codewörtern angeordnet werden. Dies ist erforderlich, damit der Decoder zur Anordnung der Escape-Folge eine eindeutige Position in decodierten Bytestrom erhält.

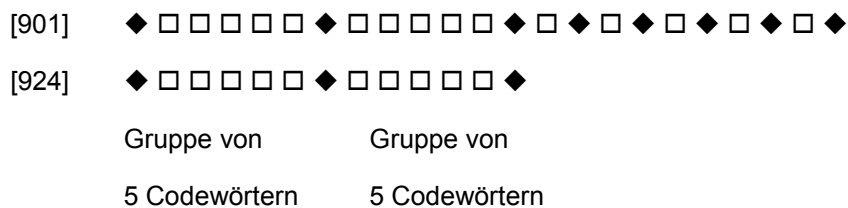
Wenn der Decoder in der Version 924 des Byte-Verdichtungsmodus arbeitet und eine ECI-Escape-Folge nach einer Gruppe von 5 Codewörtern findet, muss er die sechs Datenbytes, die mit diesen Codewörtern verbunden sind, vor der Escape-Folge ausgeben, dann die Escape-Folge und anschließend fortfahren, Codewörter für die Decodierung im Byte-Verdichtungsmodus zu erfassen. Wenn der Decoder an einer anderen als der oben beschriebenen Stelle auf eine ECI-Escape-Folge trifft, muss er dieses Symbol als ungültig behandeln.

Bei der Codierung im Byte-Verdichtungsmodus unter Verwendung des Moduswechsels 901 kann eine ECI-Escape-Folge wie folgt angeordnet werden:

- unmittelbar nach dem Codewort 901
- unmittelbar nach einem beliebigen Satz von 5 Codewörtern, in sechs Bytes codiert
- unmittelbar nach einem beliebigen der aus einem Byte bestehenden abschließenden Codewörter am Ende der Folge

ANMERKUNG Der Decoder kann nicht aufgrund dessen voraussetzen, dass die ECI-Escape-Folge nach einem Satz von 5 Codewörtern folgt, dass die fünf Codewörter in 6 Bytes codiert sind, da ein Eingabestrom mit einer Länge $6N + 5$ (dabei ist N eine ganze Zahl) einen Endsatz von fünf Codewörtern hat, die nur 5 Bytes codieren und zwar 1 Byte je Codewort. Der Decoder muss daher das Symbol nach der ECI-Escape-Folge weiterscannen, um festzustellen, wo der Modus 901 endet, wie in Abschnitt 4.4.3.4 festgelegt ist. Basierend auf dieser Information kann er dann ermitteln, wie die Gruppe von 5 Codewörtern codiert wurde.

Bild 7 zeigt gültige Stellen für ECI-Escape-Folgen bei der Codierung im Byte-Verdichtungsmodus. Wenn der Decoder auf eine ECI-Escape-Folge innerhalb einer Gruppe von 5 Codewörtern trifft, muss er das Symbol als ungültig behandeln.



dabei ist: □ = Codewort im Byte-Verdichtungsmodus
 ◆ = Gültige Stelle für eine ECI-Escape-Folge

Bild 7 — Gültige Stellen für ECI-Escape-Folgen im Byte-Verdichtungsmodus

4.5.3.3 ECIs und Byte-Verdichtung mit der Modusumschaltung 913

Bei der Codierung im Byte-Verdichtungsmodus mit der Modusumschaltung 913 kann eine ECI-Escape-Folge wie folgt angeordnet werden:

- unmittelbar vor Codewort 913
- unmittelbar nach Codewort 913
- unmittelbar nach dem Codewort nach Codewort 913

In den ersten beiden Fällen wird die ECI-Escape-Folge vor dem codierten Byte ausgegeben, während hingegen im letzteren Fall die Escape-Folge nach dem codierten Byte ausgegeben wird.

4.5.3.4 ECIs und numerischer Verdichtungsmodus

Eine ECI-Escape-Folge darf nicht innerhalb einer Gruppe von Codewörtern angeordnet werden, die mit Umwandlung der 10 Grundwerte in 900 Grundwerte nach Abschnitt 4.4.4.2 verarbeitet werden. Sie darf nur innerhalb eines Bereichs im numerischen Verdichtungsmodus an einer Grenze zwischen (den üblichen) Gruppen von 15 Codewörtern angeordnet werden. Dies ist erforderlich, damit der Decoder zur Anordnung der Escape-Folge eine eindeutige Position in decodierten Ziffernstrom erhält.

Somit kann eine ECI-Escape-Folge nur wie folgt angeordnet werden:

- unmittelbar nach Codewort 902
- nach dem 15. Codewort
- nach dem 30. Codewort
- usw.

Wenn der Codierer eine ECI-Escape-Folge an einer Stelle anordnen muss, die kein Vielfaches von 15 Codewörtern ist, muss er den numerischen Block vor dem ECI als vollständiges Objekt behandeln, wie in Abschnitt 4.4.4.2, Schritt 2, festgelegt. Er muss den numerischen Verdichtungsmodus durch Anordnung eines weiteren Codewortes 920 im Datenstrom gefolgt von der ECI-Escape-Folge neu aufrufen.

Wenn der Decoder eine ECI-Escape-Folge am einem der oben angegebenen Grenzpunkte findet, muss er die mit den Codewörtern vor der ECI-Escape-Folge verbundenen Datenbytes (falls vorhanden) senden, dann die Escape-Folge und anschließend mit dem Erfassen von Codewörtern für die Decodierung im numerischen Verdichtungsmodus fortfahren. Wenn der Decoder an einer anderen Stelle als angegeben auf eine ECI-Escape-Folge trifft, muss er das Symbol als ungültig behandeln.

4.5.3.5 Kombination von ECIs

Zwei oder mehr ECI-Escape-Folgen (z. B. Zuordnungsnummern) können an einer beliebigen Stelle angeordnet werden, wenn ein ECI sich an einer gültigen Stelle befindet. Voraussetzung ist, dass keine Codewörter zwischen ihnen angeordnet werden außer solchen, die zur Codierung der ECI-Escape-Folge verwendet werden.

4.5.4 Nachdecodierungsprotokoll

Das Protokoll zur Übertragung von ECI-Daten muss Abschnitt 4.17.2 entsprechen. Bei der Übertragung von ECIs müssen die Symbolgie-Identifikatoren (siehe Abschnitt 4.17.5) vollständig implementiert sein und der entsprechende Symbolgie-Identifikator muss als Präambel übertragen werden.

4.6 Bestimmung der Codewortfolge

Der Codierungsvorgang erzeugt eine Folge von Codewörtern mit folgender Festlegung:

$$d_{n-1} \dots d_0$$

Dabei ist:

- d das verdichtete Codewort einschließlich des Symbolhängendescriptors und aller Funktions-Codewörter
- n die Gesamtanzahl von verdichteten Codewörtern einschließlich des Symbolhängendescriptors aber ohne Fehlerkorrektur-Codewörter.

Der Symbolhängendescriptor muss das erste verdichtete Codewort sein und mit d_{n-1} bezeichnet werden. Sein Wert muss der Gesamtanzahl von verdichteten Codewörtern n entsprechen; diese Summe muss den Symbolhängendescriptor selbst beinhalten und daher im Bereich zwischen 1 und 926 liegen.

Während des Codierungsvorgangs werden Folgen von Codewörtern aufgestellt. Wie bei den ursprünglichen Daten müssen die höchstwertigen Daten wie beispielsweise Text- und numerische Daten zuerst erscheinen und zwar von links nach rechts gelesen. In der Folge der Codewörter muss das höchstwertige verdichtete Codewort, das codierte Daten enthält, mit d_{n-2} bezeichnet werden. Das letzte verdichtete Codewort wird mit d_0 bezeichnet.

Das zur Ermittlung der Symbolmatrix mit Zeilen und Spalten festgelegte Verfahren (siehe Abschnitt 4.9.2) kann den Zusatz von abschließenden Pad-Codewörtern am Ende der verdichteten Codewortfolge erfordern.

4.7 Fehlererkennung und -korrektur

Jedes PDF417-Symbol enthält zumindest zwei Fehlerkorrektur-Codewörter. Mit den Fehlerkorrektur-Codewörtern können Fehler sowohl erkannt als auch korrigiert werden.

4.7.1 Fehlerkorrekturgrad

Der Fehlerkorrekturgrad für ein PDF417-Symbol ist bei der Erzeugung des Symbols wählbar. Tabelle 9 zeigt die Anzahl von Fehlerkorrektur-Codewörtern für jeden Fehlerkorrekturgrad.

Tabelle 9 — Fehlerkorrekturgrad und Fehlerkorrektur-Codewörter

Fehlerkorrekturgrad	Gesamtanzahl von Fehlerkorrektur-Codewörtern
0	2
1	4
2	8
3	16
4	32
5	64
6	128
7	256
8	512

4.7.2 Fehlerkorrekturfähigkeit

Die Fehlerkorrektur kann verwendet werden, um Beschädigungen an Etiketten und schlecht lesbare Zeichen während der Decodierungsphase auszugleichen. Für jeden gegebenen Fehlerkorrekturgrad ist eine bestimmte Anzahl von Fehlerkorrektur-Codewörtern im PDF417-Symbol enthalten. Der verwendete Algorithmus zur Fehlerkorrektur an Codewörtern ermöglicht die Korrektur von zwei Fehlerarten:

- einer Löschung, d. h. eines fehlenden oder nicht decodierbaren Codewortes an einer bekannten Stelle;
- eines Ersetzungsfehlers, d. h. eines falsch decodierten Codewortes an einer unbekannten Stelle.

Das Fehlerkorrekturschema erfordert ein Fehlerkorrektur-Codewort, um eine Löschung zu korrigieren, und zwei Codewörter, um einen Ersetzungsfehler zu korrigieren. Somit kann ein gegebener Fehlerkorrekturgrad jede Kombination von Ersetzungsfehlern und Löschungen korrigieren, die folgende Gleichung erfüllt:

$$l + 2f \leq 2^{s+1} - 2$$

Dabei entsprechen l , f und s den Definitionen in Abschnitt 3.2.

Wenn jedoch der größte Teil der Fehlerkorrekturfähigkeit dazu verwendet wird, um Löschungen zu korrigieren, steigt die Anzahl nicht erkannter Fehler. Aus diesem Grund sollte bei weniger als 4 korrigierten Fehlern (außer wenn $s = 0$ ist) die Fehlerkorrekturfähigkeit folgendermaßen verringert werden:

$$l + 2f \leq 2^{s+1} - 3$$

Dabei entsprechen l , f und s den Definitionen in Abschnitt 3.2.

BEISPIEL Ein PDF417-Symbol mit dem Fehlerkorrekturgrad 3 besitzt 16 Fehlerkorrektur-Codewörter, von denen 14 zur Korrektur von Fehlern und Löschungen benutzt werden können. Es können bis zu 13 Löschungen oder 7 Ersetzungsfehler korrigiert werden oder jede Kombination von l Löschungen und f Ersetzungsfehlern nach den oben angegebenen Gleichungen. Tabelle 10 gibt die möglichen Kombinationen an.