Konvertierung PICA3 Downloads in Text mit Trennzeichen / Stand : 23.07.02

http://www.uni-rostock.de/ub/LBS/all/pica_to_text.pdf

Ausgangssituation UB Rostock :

Durch die Firma Satztec wurde der gesamte Bestand der Zentralbibliothek von 1960-1992 retro-katalogisiert und im GVK erfasst. Über zwei ABM-Maßnahmen konnten der Bestand des Freihand-magazins (FHM) und des Magazins (MAG) der ZB ab Signatur 64 A 1 ff. mit Barcode versehen werden. Dabei wurde regalweise vorgegangen. Diese Maßnahmen sind beendet.

Für uns ist es jetzt wichtig zu wissen, welche "Bücher" aus dem System keinen Barcode haben. Damit machen wir gleichzeitig eine Revision des Bestandes.

Eine weitere Auswertung erfolgt nach der RVK (Regensburger Verbundklassifikation = gleichzeitig Aufstellung) in der Kategorie 7105. Hintergrund sind die Baumaßnahmen und damit die Zusammen-legung von Beständen. Die UB wird künftig nur noch 4 Standorte haben und somit werden die Bestände FHM und MAG der ZB ab 1964 ff. auf die entsprechenden Standorte verteilt. Die UB vergibt seit 1994 die RVK. Eine Abschätzung konnte bisher nur sehr grob vorgenommen werden. So kann jetzt die Anzahl der Exemplare und damit auch der Platzbedarf am neuen Standort kalkuliert werden.

Selektion der Standorte :

Magazin (7100 !28-MAG!)
Freihandmagazin (7100 !28-FHM!)

mit den Kategorien :

3000	1. Verfasser
4000	Hauptsachtitel, Zusätze, Verfasserangabe
7100	!Standort!Signatur Numerus Currens @ Ausleihindikator
7105	RVK
7800	EPN (Exemplar Produktionsnummer)

Aufgabenstellung : PICA3 Download in Text mit Trennzeichen konvertieren.

In der ersten Zeile der Ausgabedatei sollen die Kategorien als Feldbezeichner stehen,
in den folgenden Zeilen die Werte der jeweiligen Kategorien.

Die Ausgabedatei kann dann in eine Datenbank eingelesen und diversen Auswertungen unterzogen werden, z.B. Filter und Sortierungen bestimmter Kategorien.

Die richtige Darstellung von Umlauten und Sonderzeichen ist nicht relevant.

Als Trennzeichen muss ein Zeichen verwendet werden, das nicht in den PICA3-Dateien vorkommt - z.B. Euro (€).

Realisierung : Visual-Basic Sript, Ausführung aus MS Word (97)

Jeder PICA3-Exemplarsatz wird in einen TEXT-Satz übertragen. Dabei wird vorausgesetzt, dass Beginn und Ende eines Titel und Exemplars fest definiert sind :

	Beginn	Boolsche Variable	Ende	Boolsche Variable
Titel	KAT 3000 oder 4000	sot	Leerzeile	eot
Exemplar	KAT 7100 oder 7105	soe	KAT 7800	eoe

sot / eot = start / end of title
soe / eoe = start / end of exemplar

Sonstiger Text ausserhalb eines Titel-Exemplar Konstrukts (sot-eot) wird herausgefiltert.

Durch einen Abgleich über die EPN mit dem Lokalen System (LBS - Tabelle volume) können die Exemplare ermittelt werden, die noch keinen Barcode haben bzw. nicht ordnungsgemäß vom CBS in das LBS übertragen wurden.

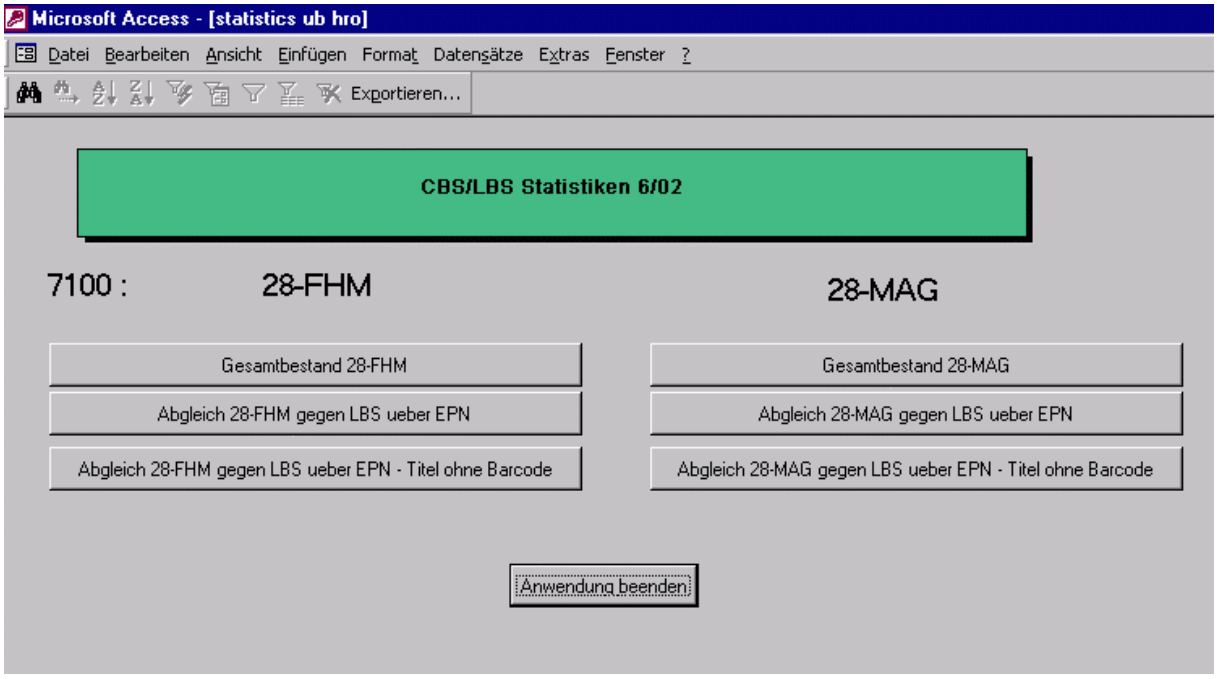
Export CBS zu PICA3 (Auszug) **:**

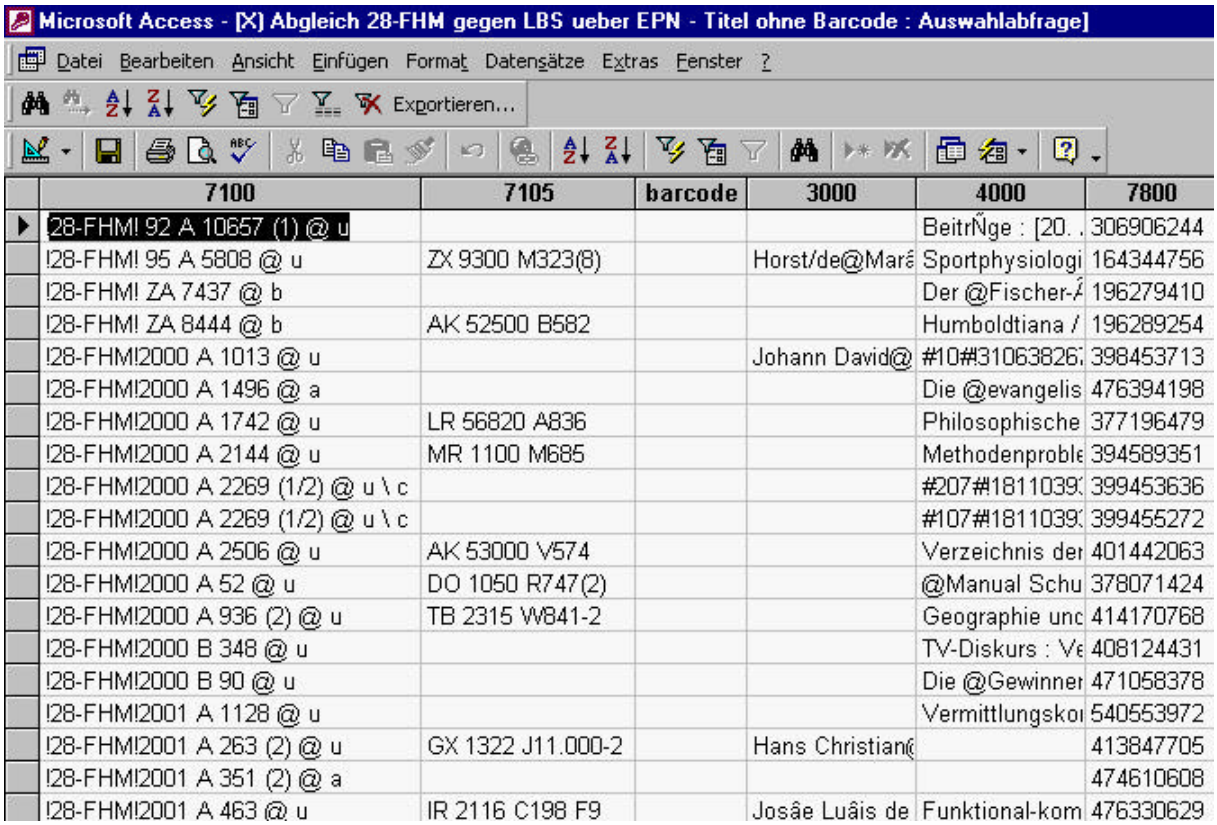
```
3000 Hilla@Jacoby
4000 Die @Juden: Gottes Volk / Hilla Jacoby; Max Jacoby
7100 !28-FHM!94 B 89 @ u
7105 BC 8780 J17
7800 093948298

4000 @Geschichten:+: Riemels un LÓÓd'snack : mecklenburgische VolksÓberlieferungen /
ges. von Richard Wossidlo. Neu hrsg. von Ulrich Bentzien
7100 !28-FHM!99 A 4191 @ u
7105 GT 1495 B478
7800 354605364

4000 @Feldfutterbau / KÑmpf:+: Robert u.a. [Mitarb.]
7100 !28-FHM!90 A 2848 @ u
7800 313728879
7100 !28/052!ZC 34000 K11(2) @ u
7800 313728887
7100 !28/052!90 A 2848/2 @ u
7800 313728895
7100 !28/052!90 A 2848/3 @ u
7800 313728909
7100 !28/052!90 A 2848/4 @ u
7800 313728917
7100 !28/052!90 A 2848/5 @ u
7800 313728925
7100 !28/052!90 A 2848/6 @ u
7800 313728933
7100 !28/052!90 A 2848/7 @ u
7800 313728941
7100 !28/052!90 A 2848/8 @ u
7800 31372895X
7100 !28/052!90 A 2848/9 @ u
7800 313728968
7100 !28/052!90 A 2848/10 @ u
7800 313728976

3000 Arthur Erwin@Imhof
4000 @GrundzÓge der nordischen Geschichte / Arthur Erwin Imhof
7100 !28-FHM!90 A 3444 @ u
7800 316310964

3000 Carl F./von@WeizsÑcker!136388175!Carl Friedrich/von@WeizsÑcker
4000 Die @Zeit drÑngt : eine Weltversammlung der Christen fÓr Gerechtigkeit:+: Frieden
und die Bewahrung der SchÒpfung / Carl Friedrich von WeizsÑcker
7100 !28-FHM!2000 A 1430 @ a
7800 505019507

3000 Eric@Zimmer
4000 Die @Frage nach dem Sinn : Sinn des Lebens / Eric Zimmer
7100 !28-FHM!94 A 4259 @ u
7105 CC 6600 Z72
7800 179014935

3000 Hans-Albert@Walter!23803030X!Hans-Albert@Walter
4000 "Im Anfang war die Tat" : Arnold Zweigs "Beil von Wandsbek" ; Roman einer Welt -
Welt eines Romans / Hans-Albert Walter
7100 !28/601!GM 7498 W232+2 @ u
7800 244979650
7100 !28-FHM!90 A 1135 @ u
7800 335460550
```

Wird konvertiert zu TEXT :

```
NR€3000€4000€7100€7105€7800
 1€Hilla@Jacoby€Die @Juden: Gottes Volk / Hilla Jacoby; Max Jacoby€!28-FHM!94 B 89 @
u€BC 8780 J17€093948298
 2€€@Geschichten, Riemels un LÓÓd'snack : mecklenburgische VolksÓberlieferungen / ges.
von Richard Wossidlo. Neu hrsg. von Ulrich Bentzien€!28-FHM!99 A 4191 @ u€GT 1495
B478€354605364
 3€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28-FHM!90 A 2848 @ u€€313728879
 4€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!ZC 34000 K11(2) @ u€€313728887
 5€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/2 @ u€€313728895
 6€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/3 @ u€€313728909
 7€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/4 @ u€€313728917
 8€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/5 @ u€€313728925
 9€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/6 @ u€€313728933
 10€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/7 @ u€€313728941
 11€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/8 @ u€€31372895X
 12€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/9 @ u€€313728968
 13€€@Feldfutterbau / KÑmpf, Robert u.a. [Mitarb.]€!28/052!90 A 2848/10 @ u€€313728976
 14€Arthur Erwin@Imhof€@GrundzÓge der nordischen Geschichte / Arthur Erwin Imhof€!28-
FHM!90 A 3444 @ u€€316310964
 15€Carl F./von@WeizsÑcker!136388175!Carl Friedrich/von@WeizsÑcker€Die @Zeit drÑngt :
eine Weltversammlung der Christen fÓr Gerechtigkeit, Frieden und die Bewahrung der
SchÒpfung / Carl Friedrich von WeizsÑcker€!28-FHM!2000 A 1430 @ a€€505019507
 16€Eric@Zimmer€Die @Frage nach dem Sinn : Sinn des Lebens / Eric Zimmer€!28-FHM!94 A
4259 @ u€CC 6600 Z72€179014935
 17€Hans-Albert@Walter!23803030X!Hans-Albert@Walter€"Im Anfang war die Tat" : Arnold
Zweigs "Beil von Wandsbek" ; Roman einer Welt - Welt eines Romans / Hans-Albert
Walter€!28/601!GM 7498 W232+2 @ u€€244979650
 18€Hans-Albert@Walter!23803030X!Hans-Albert@Walter€"Im Anfang war die Tat" : Arnold
Zweigs "Beil von Wandsbek" ; Roman einer Welt - Welt eines Romans / Hans-Albert
Walter€!28-FHM!90 A 1135 @ u€€335460550
```

Auswertung z.B. in MS Access Datenbank :





7100	7105	barcode	3000	4000	7800
28-FHM! 92 A 10657 (1) @ u				BeitrÑge : [20. .	306906244
!28-FHM! 95 A 5808 @ u	ZX 9300 M323(8)		Horst/de@Marâ	Sportphysiologi	164344756
!28-FHM! ZA 7437 @ b				Der @Fischer-Â	196279410
!28-FHM! ZA 8444 @ b	AK 52500 B582			Humboldtiana /	196289254
!28-FHM!2000 A 1013 @ u			Johann David@	#10#31063826:	398453713
!28-FHM!2000 A 1496 @ a				Die @evangelis	476394198
!28-FHM!2000 A 1742 @ u	LR 56820 A836			Philosophische	377196479
!28-FHM!2000 A 2144 @ u	MR 1100 M685			Methodenproble	394589351
!28-FHM!2000 A 2269 (1/2) @ u \ c				#207#1811039:	399453636
!28-FHM!2000 A 2269 (1/2) @ u \ c				#107#1811039:	399455272
!28-FHM!2000 A 2506 @ u	AK 53000 V574			Verzeichnis der	401442063
!28-FHM!2000 A 52 @ u	DO 1050 R747(2)			@Manual Schu	378071424
!28-FHM!2000 A 936 (2) @ u	TB 2315 W841-2			Geographie unc	414170768
!28-FHM!2000 B 348 @ u				TV-Diskurs : Ve	408124431
!28-FHM!2000 B 90 @ u				Die @Gewinner	471058378
!28-FHM!2001 A 1128 @ u				Vermittlungsko	540553972
!28-FHM!2001 A 263 (2) @ u	GX 1322 J11.000-2		Hans Christian@		413847705
!28-FHM!2001 A 351 (2) @ a					474610608
!28-FHM!2001 A 463 @ u	IR 2116 C198 F9		Josâe Luâis de	Funktional-kom	476330629

ANLAGE : Quelltext MS Visual-Basic / Ausführung aus WORD97

```
' -------------------------------------------------------
' Typdefinitionen des Moduls
' -------------------------------------------------------
Public Type PicaRecTyp
    Kat As String * 4
    KatText As String
    End Type

Public Type TxtRecTyp
    RowCount As Long    ' lfd. Nummer der Zeile
    Kat3000 As String
    Kat4000 As String
    Kat7100 As String
    Kat7105 As String
    Kat7800 As String
    RowOut As String    ' aufbereitete Ausgabezeile
End Type

Sub ConvertPicaTOText2()
' Strukturieren von PICA Download Files (PICA3-Format) in
' datenbankkonforme ascii-textfiles mit Trennzeichen

Dim InFile As String        ' Name Eingabedatei PICA3 Format
Dim OutFile As String       ' name Ausgabedatei Text mit Trennzeichen
Dim RowCountIn, RowCountOut As Long ' Zeilencounter Source und Ausgabefile
Dim unitsMoved As Integer   ' Zaehler für Navigation in Text
Dim Row As String           ' Zeile und Zeilentext
Dim PicaRecord As PicaRecTyp
Dim Kat As String * 4       ' Erste 4 Zeichen der Zeile ergeben PICA-Kategorie
Dim RowOut As TxtRecTyp     ' Ausgabezeile
Dim LastRow As Boolean      ' Zum Abchecken EOF
Dim sot, eot, soe, eoe As Boolean ' start und end of title/exemplar
Dim simpletxt As Boolean    ' sonstigen text ausfiltern
Dim Sep As String * 1       ' Separator
' sot - wenn vorher Row = Leer (eot) und dann Kat = "3000" oder "4000"
' eot - wenn vorher "7800" (eoe) und Row = Leer
' soe - wenn sot true und Kat = "7100"
' eoe - wenn soe true und Kat = "7800"

' Variablen initalisieren
InFile = "\\ubcs001\lbsreport\02-06\cbs\mag.doc"
OutFile = "\\ubcs001\lbsreport\02-06\cbs\mag-out.txt"
Row = ""
With RowOut
    .RowCount = 0
    .Kat3000 = ""
    .Kat4000 = ""
    .Kat7100 = ""
    .Kat7105 = ""
    .Kat7800 = ""
    .RowOut = ""
End With
PicaRecord.KatText = ""
PicaRecord.Kat = ""
' Sep = ";"
Sep = "€" ' Euro getestet : tritt in PICA3 File nicht auf
RowCountIn = 0
RowCountOut = 0
sot = False
eot = False
soe = False
eoe = False
simpletxt = False
LastRow = False

' Kommata im InFile ersetzen
SubstComma (InFile)

' Zu konvertierendes File zum Lesen öffnen
Open InFile For Input As #1
' Debug.Print Seek(1)

' Ausgabedatei öffnen
Open OutFile For Output As #2
' Zeile mit Feldbezeichnern einfuegen
Print #2, "NR" + Sep + "3000" + Sep + "4000" + Sep + "7100" + Sep + "7105" + Sep + "7800"

Do While Not LastRow
    ' Zeile einlesen
    Row = ReadRow2
    LastRow = (Row = "EOF")
    If Not LastRow Then
        RowCount = RowCount + 1
        RowCountIn = RowCountIn + 1
```

```
        ' Kontrollausgabe im Testfenster :
        ' Debug.Print ("Row" + Str(RowCount))
        ' Debug.Print ("-" + Row + "-")

        ' Hier jetzt Logik zur Zeilenbearbeitung
        eot = (Row = " ") Or (Row = Empty)
        simpletxt = (Len(Row) < 6) And Not eot
        If (Not eot) And (Not simpletxt) Then
            PicaRecord = SplitRow(Row)
            Select Case PicaRecord.Kat
            ' KAT 3000 eröffnet in jedem Fall einen neuen Titel
            Case "3000", "4000"
                sot = True
                ' verarb.
                If PicaRecord.Kat = "3000" Then
                    ' RowOut.Kat3000 = SubstSep(PicaRecord.KatText)
                    RowOut.Kat3000 = PicaRecord.KatText
                    RowOut.Kat4000 = ""
                Else
                    ' RowOut.Kat4000 = SubstSep(PicaRecord.KatText)
                    RowOut.Kat4000 = PicaRecord.KatText
                End If
                sot = False
            Case "7100", "7105"
                ' If Not soe Then RowCountOut = RowCountOut + 1 ' Eine Ausgabezeile je exemplar
                soe = True
                ' verarb.
                If PicaRecord.Kat = "7100" Then
                    ' RowOut.Kat7100 = SubstSep(PicaRecord.KatText)
                    RowOut.Kat7100 = PicaRecord.KatText
                    RowOut.Kat7105 = ""
                Else
                    ' RowOut.Kat7105 = SubstSep(PicaRecord.KatText)
                    RowOut.Kat7105 = PicaRecord.KatText
                End If
                soe = False
            Case "7800"
                eoe = True
                ' verarb.
                ' RowOut.Kat7800 = SubstSep(PicaRecord.KatText)
                RowOut.Kat7800 = PicaRecord.KatText
                RowCountOut = RowCountOut + 1
                With RowOut
                    .RowCount = RowCountOut
                    .RowOut = Str(.RowCount) + Sep + .Kat3000 + Sep + .Kat4000 _
                    + Sep + .Kat7100 + Sep + .Kat7105 + Sep + .Kat7800
                    WriteRow2 (.RowOut)
                    .Kat7100 = ""
                    .Kat7105 = ""
                    .Kat7800 = ""
                End With
                eoe = False
            End Select
        ElseIf eot And (Not simpletxt) Then ' if eot dann Satz initialisieren
            With RowOut
                .RowCount = 0
                .Kat3000 = ""
                .Kat4000 = ""
                .Kat7100 = ""
                .Kat7105 = ""
                .Kat7800 = ""
                .RowOut = ""
            End With
        End If

        ' Neues Format der Sätze in separates File übertragen

        ' unitsMoved = Selection.MoveDown(unit:=wdLine, Count:=1)
    End If ' Not lastrow

Loop

' PicaFile Schliessen
' PicaFile.Activate
' ActiveDocument.Close SaveChanges:=False
Close #1
' Outputfile Sichern und Schliessen
Close #2

ReSubstComma (OutFile)

End Sub


Function ReadRow2()
Dim Row As String
Dim SepPos As Integer
```

```
    If EOF(1) Then
        ReadRow2 = "EOF"
        ' Debug.Print ("EOF")
    Else
        Input #1, Row
        ReadRow2 = Row
        ' Debug.Print ("-" + Row + "-")

    End If
End Function


Function SubstSep(Row As String) As String
' Ersetzt Separatorzeichen in Row
SepPos = InStr(6, Row, ";")
Do While Not (SepPos = 0)
    ' Debug.Print (zeile)
    Row = Left(Row, SepPos - 1) + "-semikolon-" + Right(Row, Len(Row) - SepPos)
    SepPos = InStr(6, Row, ";")
Loop
SubstSep = Row
End Function


Function SplitRow(Row As String) As PicaRecTyp
Dim PicaRecord As PicaRecTyp
    PicaRecord.Kat = Left(Row, 4)
    PicaRecord.KatText = Right(Row, Len(Row) - 5)
    ' Debug.Print (PicaRecord.Kat + "," + PicaRecord.KatText)
    SplitRow = PicaRecord
End Function


Function WriteRow2(Row As String)
' Schreibt Row in Ausgabedatei
Print #2, Row
' Debug.Print (Row)
End Function


Function SubstComma(InFile As String)
' Kommata in Datei durch :+: ersetzen, damit Einlesen mit Input #Dateinummer, Zeile korrekt
' Documents.Open FileName:= "\\ubcs001\lbsreport\02-06\cbs\fhm.doc",
Documents.Open FileName:=InFile
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
    .Text = ","
    .Replacement.Text = ":+:"
    .Forward = True
    .Wrap = wdFindContinue
    .Format = False
    .MatchCase = False
    .MatchWholeWord = False
    .MatchWildcards = False
    .MatchSoundsLike = False
    .MatchAllWordForms = False
End With
Selection.Find.Execute Replace:=wdReplaceAll
ActiveDocument.Close SaveChanges:=wdSaveChanges
End Function


Function ReSubstComma(OutFile As String)
Documents.Open FileName:=OutFile, ReadOnly:=False, ConfirmConversions:=False,
AddToRecentFiles:=False, Format:=wdOpenFormatAuto
    Selection.Find.ClearFormatting
    Selection.Find.Replacement.ClearFormatting
    With Selection.Find
        .Text = ":+:"
        .Replacement.Text = ","
        .Forward = True
        .Wrap = wdFindContinue
        .Format = False
        .MatchCase = False
        .MatchWholeWord = False
        .MatchWildcards = False
        .MatchSoundsLike = False
        .MatchAllWordForms = False
    End With
Selection.Find.Execute Replace:=wdReplaceAll
ActiveDocument.Close SaveChanges:=wdSaveChanges
End Function
```