



Aufwand nicht, weil alle diese „Formeln“ nur einmal verwendet werden.

**Ergebnis**

Das Ergebnis eines Spiels wird meist in zwei Zeilen angegeben. Die erste Zeile ist das Endergebnis, die zweite ist der eingeklammerte Halbzeitstand.

```
Suchen (\d+)\:(\d+)\r\n\((\d+)\:(\d+)\)\r\n
Ersetzen \r\n#Ergebnis#\1:\2\r\n#ErgebnisH#\3:\4\r\n
```

Aus 2:3(1:1) wird #Ergebnis#2:3 #ErgebnisH#1:1

**Art**

Suchbegriff (Freundschaftsspiel|Weltmeisterschaft|Europameisterschaft|WM-Qualifikation|EM-Qualifikation|EC der Nationen|Olympia)

```
Ersetzungsbeispiel \r\n#Art#\1\r\n
```

Der senkrechte Strich ist die Trennung zwischen Textalternativen. Ganz egal, welcher der Texte gefunden wird, immer wird das Feld Art generiert.

Der vorangestellte Zeilenumbruch beim Ersetzungsbeispiel garantiert, dass die Feldbezeichnung jedenfalls in einer neuen Zeile beginnt.

**Länderspiel**

Jedes Spiel enthält eine Zeile, die etwa so aussieht:

633. Länderspiel, Innsbruck (Tivoli-NEU), 28.04.2004 Schiedsrichter: Damir Skomina, SLO

Aus diesen Zeilen werden die Felder **Spiel**, **Datum**, **Schiedsrichter** und **Stadion** extrahiert.

Leider ist der Aufbau des Dokuments nicht ganz gleich, daher muss man in Einzelschritten vorgehen und kann sich nicht darauf verlassen, diese Zeile in einem einzigen Schritt umwandeln zu können.

```
Suchen (\d+)\. Länderspiel(.* )
Ersetzen \r\n#Spiel#\1\r\n#Details#\2\r\n
```

Das Wort „Länderspiel“ dient und lediglich zum Auffinden der richtigen Zeilen. Zuerst trennen wir die Spielnummer (\d+) vom Rest der Zeile (.\*) . \d steht für eine Ziffer \d+ steht für mindestens eine Ziffer. Der Punkt . steht für irgendein Zeichen, der Stern \* bedeutet, dass davor stehende Zeichen beliebig oft und auch gar nicht vorkommen kann. Damit erkennen wir auch Zeilen, die überhaupt nur mit dem Wort Länderspiel enden.

**Datum**

```
Suchen (\d+)\. (\d+)\. (\d\d\d\d)
Ersetzen \r\n#Datum#\1. \2. \3\r\n
```

Es gibt auch Datumsangaben mit ausgeschriebenen Monatsangaben und die müssen in einem weiteren, etwas komplizierteren Schritt verarbeitet werden.

**Schiedsrichter**

```
Suchen Schiedsrichter\:(\w+)
Ersetzen \r\n#Schiedsrichter#\1\r\n
```

Im Feld **Detail** bleiben nach diesen Umformungen Angaben zum Stadion stehen.

Abgeschlossen werden diese Umformungen durch Entfernen mehrfacher Zwischenräume:

```
Suchen \s+
Ersetzen " " (ohne Anführungszeichen)
```

Und mehrfacher Zeilenumbrüche, die sich während der Ersetzungen durch Voranstellen von \r\n\ angehäuft haben.

```
Suchen (\r\n)+
Ersetzen \r\n
```

Diese Ersetzungen sind eine grobe Skizze, wie man von dem Rohtext alles-original.txt zu einem strukturierten Text alles.txt mit Feldangaben kommen kann.

In der Praxis ist die Zahl der erforderlichen Umwandlungsschritte deutlich höher und zwischen durch durch manuelle Textkorrekturen begleitet, die dadurch erforderlich werden, dass die Originaldateien Tippfehler enthalten aber auch dadurch, dass die Umwandlungen versehentlich eine Ersetzung an einem Ort ausführen, an dem das gar nicht gewünscht ist.

**Import in Datenbank**

Es gibt jetzt also eine strukturierte Textdatei alles.txt. Jede Zeile enthält ein Datenfeld der Form #Feldname#Inhalt. Ein Datensatz ist vom nächsten durch eine Zeile ##### getrennt. Die Reihenfolge der Datenfelder ist unerheblich. Fehlen Datenfelder, bleibt das entsprechende Datenbankfeld eben leer.

**Programmwahl**

Man könnte den Import der Texte mit dem in Access eingebauten Visual Basic ausführen. Man kann aber auch die ungemein flexible **PowerShell** verwenden, die alles, was das DotNet-Framework anbietet, auch nutzen kann. Wenn man daher routinemäßig mit C# arbeitet, ist die Verwendung der PowerShell naheliegender.

Man kann die PowerShell entweder mit dem in Windows vorhandenen **Power Shell ISE (Integrated Software Environment)** oder auch mit dem kostenlosen Tool **Power Gui Script Editor** verwenden. Mit gefällt **Power Gui** besser aber für den Anfang ist das egal.

Eine kurze Darstellung der Skriptsprache PowerShell zeigt die folgende Seite, eine etwas längere Darstellung ist bei der Webversion dieses Beitrags verfügbar.

Der Import der Daten erfolgt in mehreren Schritten:

- Tabelle **Spiel** durch sequentielles Auslesen der Text-Datei alles.txt füllen. (TextDateiLesen.ps1)  
Das Feld Spiel (in der Textdatei alles.txt wurde es ID genannt) dient als eindeutige Datensatzkennzeichnung und Verknüpfungsfeld zu den Tabellen Spieler-Spiel und Tore.
- Tabelle **Spieler-Spiel** aus dem Feld **Aufstellung** der Tabelle **Spiel** erstellen (SpielerTabelleErstellen.ps1)
- Die Tabelle **Spieler-Spiel** enthält zunächst noch sowohl den **Verein** als auch den **Spieler**, beides in vielfacher Anzahl. Durch Gruppierung nach dem **Spieler** und nach dem **Verein** wird daraus die Tabellen **Spieler-Verein** und aus dieser dann die Tabellen **Spieler** und **Verein** hergestellt. Löscht man danach die nicht mehr benötigten Felder sind die genauen Schritte nicht mehr erkennbar.

- Tabelle **Tore** aus dem Feld **Tore** ermitteln (ToreTabelleErstellen.ps1)
- Tabelle **Spieler-Tausch** aus dem Feld **Austausch** erstellen und dann die Spielminuten der bereits eingetragenen Feldspieler korrigieren (SpielerTauschTabelleErstellen.ps1). Die neuen Eintauschspieler an die Tabelle **Spieler** anfügen und die Spielminuten eintragen. Dieser relativ komplizierte Vorgang ist notwendig, weil die ÖFB-Daten zwischen der Aufstellung (die 11 Mann, die das Spiel beginnen) und den Wechselspielern unterscheidet. Für die Zählung der Zahl der Einsätze ist diese Unterscheidung eher unpraktisch wie man sieht. Daher müssen eben die Austauschspieler zu der Spielertabelle hinzugefügt werden und dabei die Beginn- und Endminute der neuen Spieler als auch jener, die ausgetauscht werden, bestimmt werden.

**Datenquellen**

- <http://www.oefb.at> (Spiel 1..716)
- <http://www.weltfussball.at> (Spiel 717-727)

**Verwendete Programme**

- Microsoft Access 2013
- Microsoft Excel 2013
- Notepad++
- Power GUI Script Editor
- Windows Power Shell ISE

**Download**

**Webversion "Vermessung des Runden Leders"**  
<http://d.pcnews.at/ins/pcn/139/000500/>

- Österreich.accdb Datenbank
- Österreich.xlsx Grafische Auswertungen
- Serien.docx Textversion aller Serien

**Webversion "Datamining"**  
<http://d.pcnews.at/ins/pcn/139/002000/>

- Alles-Original.pdf
- Alles-Original.txt
- Alles.txt
- TextDateiLesen.ps1
- SpielerTabelleErstellen.ps1
- SpielerTauschTabelleErstellen.ps1
- ToreTabelleErstellen.ps1

**Links**

Einführung in Access von Karel Štípek in PCNEWS-103

<http://d.pcnews.at/ins/pcn/103/002700/main.htm>

Und alle anderen Artikel von Karel Štípek  
<http://pcnews.at/?Show=AutorenArtikel&n=17484&id=13725&detail=%C5%A0>