

# Computerbenutzung und Programmieren

## Aufgabe 3: Datenformate

Ein spezielle Situation, in der Markup benutzt wird, ist das Speichern strukturierter Datensätze. Denken Sie dabei an Tabellen oder Adressenbücher. In Prinzip können Sie (fast) jede der vielen Markup-Ansätze benutzen um so einen Datensatz zu speichern (und oft wird eine Form von XML benutzt). In der Praxis haben sich aber eine Reihe von speziellen Formate herausgebildet, die viel benutzt werden um Daten zu strukturieren; vor allem, wenn es keine rohe Textdaten sind, sondern Listen oder Tabellen mit Informationen.

(Es ist natürlich schon fast eine philosophische Diskussion in wie weit Texte etwas anderes sind als eine Art Liste mit Informationen. In Theorie gibt es wohl eher keinen Unterschied, aber in der Praxis schon.)

**Aufgabe 1:** Formatiere die unterstehende Adressinformationen (d.h. die Infos auf den Webseiten, wenn Sie unterstehende Links nachgehen) in folgende Dateiformate, und reichen Sie diese ein. Versuchen Sie dabei, soviel wie möglich Informationen zu strukturieren.

- CSV
- JSON
- YAML

(unten sind noch einige Tipps angefügt, aber lesen Sie gerne auch noch mal bei Wikipedia nach was es mit diesen auf sich hat.)

<https://www.uni-marburg.de/de/fb09/dsa/einrichtung/personen/wissenschaftler/harlos>  
<https://www.uni-marburg.de/de/fb09/dsa/einrichtung/personen/cysouw/prof-dr-michael-cysouw>

---

**Aufgabe 2:** Ein wichtiger Aspekt solcher Datenformate ist, dass sie genau angewendet werden sollen. Ein Komma oder eine Klammer an der falschen Stelle führt dazu, dass die Daten eine komplett andere Bedeutung kriegen. Deshalb ist es wichtig Ihre Daten zu **validieren**, z.B. auf Webseiten wie diese:

<https://codebeautify.org/jsonvalidator>  
<https://onlineyamltools.com/validate-yaml>  
<http://www.convertcsv.com/csv-viewer-editor.htm>

**Anmerkung:** Es gibt viele Webseiten, die Anbieten zwischen solche Formate hin-und-her zu konvertieren. Versuchen Sie aber erst einmal selber “händisch” die Formate zu erstellen, und schauen Sie danach was passiert, wenn Sie Ihre Dateien automatisch konvertieren lassen.

---

## Tipps

**CSV** steht für “Comma Separated Values”, also “durch Kommas getrennte Werte”. Dieses Format wird am ehesten genutzt um ‘rechteckige’ Tabellen zu speichern, typischerweise mit ‘Entitäten’ als Reihen und ‘Merkmale’ als Spalten. In der ersten Zeile stehen oft die Namen der Spalten.

Die Extension "Rainbow CSV" innerhalb Visual Studio Code ist sehr hilfreich bei der Bearbeitung von CSV Dateien.

Es gibt drei verschiedene Zeichen, die in diesem Format eine spezielle Bedeutung haben:

- Zeilentrenner (typisch: Zeilenumbruch LF oder CR+LF)
- Spaltentrenner (typische: Komma, Tab oder Semicolon)
- Feldbegrenzung (typisch: doppelte Anführungszeichen. Wird meistens nur benutzt, wenn der Spaltentrenner auch in den eigentlichen Daten vorkommen)

Kleinere Datensätze in CSV-Format lassen sich noch gut händisch in einem Texteditor bearbeiten. Aber es lohnt sich oft ein Tabellenprogramm wie Microsoft Excel oder Open Office Calc zu benutzen. Erstellen Sie z.B. eine Tabelle mit den Daten aus der Aufgabe in Excel und exportieren Sie diese Daten als CSV. Öffnen Sie diese CSV-Datei in einem Texteditor, und schauen Sie die genaue Struktur an.

**JSON** steht für "Javascript Object Notation". Es ist ein Format, das im Rahmen des Austausches von Daten zwischen Rechner im Internet beliebt geworden ist. Es ist viel flexibler als CSV, aber auch viel aufwändiger um händisch zu bearbeiten. Es wird am ehesten automatisch generiert. Heute sollen Sie aber versuchen JSON trotzdem händisch zu erstellen.

JSON besteht im Prinzip aus Aufzählung der Form **{"Merkmal": "Wert"}**, getrennt durch Kommas. Die geschweiften Klammern, Doppelpunkte und Kommas sind zentral, Zeilenumbrüche und Leerzeichen ausserhalb der Anführungszeichen sind nur für die Lesbarkeit wichtig, sie haben keine Bedeutung und können weggelassen werden. Werte können selber auch wieder Aufzählungen sein, wodurch eine komplexe hierarchische Struktur entstehen kann.

**YAML** steht für "Yet Another Markup Language" (oder manchmal auch "YAML Ain't Markup Language": das ist Informatiker-Humor...) und ist ein Versuch eine Alternative zu JSON zu sein, die aber einfacher händisch zu bearbeiten ist. YAML besteht im Prinzip aus eine Aufzählung der Form **Merkmal: Wert** getrennt durch einen Zeilenumbruch. Auch hier können die Werte selber wieder eine Aufzählung sein, wofür dann ein Einzug mit Leerzeichen oder Tabs benutzt wird. In Gegensatz zu JSON sind bei YAML also Zeilenumbrüche und Leerzeichen wichtig, dafür werden keine Klammern oder Kommas benutzt.