

Bitte bearbeiten Sie die folgenden Aufgaben in Zweier- oder Dreiergruppen. Wenn Sie Feedback zu Ihren Lösungen haben möchten, notieren Sie Ihre Lösungen in einem Textdokument und schicken Sie es mir per E-Mail (bitte keine Mehrfachabgabe identischer Lösungen).

1. Dateiformat-Konverter

Sie erhalten den Auftrag, einen Universalkonverter zu schreiben, der u. a. die folgenden Dokumentumwandlungen ermöglicht:

- a) HTML → TIFF-Bild
- b) ASCII-Text → MPEG-2-Video
- c) PNG-Bild → ASCII-Text
- d) MPEG-2-Video → PNG-Bild
- e) Word-Dokument (XML, docx) → Excel-Tabelle (XML, xlsx) mit Liste aller im Word-Dokument auftretenden Wörter und deren Häufigkeiten

Nach einem kurzen Blick auf die Liste wollen Sie die Aufgabe schon beleidigt ablehnen („Unfug!“), ein paar Minuten später denken Sie sich aber: Doch, das geht!

Geben Sie zu allen fünf gewünschten Konvertierungen an,

- warum auf den ersten Blick eine Konvertierung nicht möglich oder sinnlos ist,
- wie Sie doch eine sinnvolle Formatwandlung vornehmen können.

2. XML: Wohlgeformt und valide

Zur Erinnerung: XML-Dokumente sind **wohlgeformt**, wenn sie ein paar einfache syntaktische Regeln einhalten, im Wesentlichen:

- Jedes öffnende Tag `<xyz>` braucht ein schließendes Tag `</xyz>` oder muss selbst schon schließend sein (`<xyz />`). Das schließende Tag muss dieselbe Schreibweise wie das öffnende verwenden (`<xyz>...</XYZ>` ist falsch).
- Tags müssen in richtiger Reihenfolge geschlossen werden: `<abc><xy></abc></xy>` ist falsch.
- Wenn es Attribute gibt (`<xyz attr="123">...`), muss der Wert des Attributs immer in Anführungszeichen stehen – auch wenn es eine Zahl ist.

(Es gibt noch mehr Regeln, aber ausführlicher haben wir XML nicht besprochen.)

Mit einer Document Type Definition (DTD) kann man aus der Menge aller denkbaren wohlgeformten Dokumente zudem eine Teilmenge auswählen, die eine neue Sprache bildet. Die DTD kann Regeln erzwingen, z. B. für XHTML, dass der Dokumenttitel (`<title>...</title>`) nur im Header des Dokuments (`<head>...</head>`) definiert werden darf. Das Aussehen einer DTD haben wir nicht besprochen, aber worum es geht, dürfte klar sein. Ist ein XML-Dokument nicht nur wohlgeformt, sondern erfüllt auch die Regeln einer DTD, dann heißt es **valide**.

Sie haben eine DTD für mathematische Ausdrücke erstellt, die im wesentlichen aus folgenden Regeln besteht (die wir hier sprachlich formulieren, weil unbekannt ist, wie eine DTD auszusehen hat).

1. Ein Dokument fängt mit `<formel>` an und endet mit `</formel>`.
Dazwischen steht eine *FORMEL*.
2. Eine *FORMEL* kann z. B. eine der Ziffern 1, 2, 3, 4, 5, 6, 7, 8, 9, 0 sein.
3. Wenn A und B *FORMELN* sind, dann sind auch `<sum>A,B</sum>` und `<prod>A,B</prod>` *FORMELN*.
4. Ansonsten gibt es keine *FORMELN*.

Damit sollte klar sein, dass `<formel>1</formel>`, `<formel><sum>1,<prod>3,4</prod></sum></formel>` und `<formel><sum><sum><sum>1,2</sum>,3</sum>,4</sum></formel>` wohlgeformte und sogar valide Dokumente sind.

- Stellen Sie den mathematischen Ausdruck $3 + 4 * 2 + 5$ auf zwei unterschiedliche (!) Arten mit Hilfe unserer neuen XML-Sprache dar.
- Welches Problem ergibt sich, wenn Sie das Dokument `<formel><prod>3,<sum>4,5</sum></prod></formel>` in eine math. Formel umwandeln?
- Unsere XML-Sprache kennt keine Klammern, sondern nur Summen und Produkte. Einen Ausdruck wie $(3+4) * 5$ könnte man auch als `...<prod><klammer><sum>3,4</sum></klammer>,5</prod>...` schreiben. Entscheiden Sie: Ist es sinnvoll, diese Klammern in die Sprache aufzunehmen? Begründen Sie Ihre Antwort.

3. HTTP-Requests

Die „kleinste“ Anfrage, die ein Webbrowser an einen Webserver schicken kann, ist immerhin zwei Zeilen lang und hat – im Beispiel für die FOM-Startseite – die Form

GET / HTTP/1.1

Host: www.fom.de

- Warum schickt der Browser die zweite Zeile mit (und darf darauf auch nicht verzichten)? Schließlich geht die Anfrage doch an den richtigen Rechner (eben: www.fom.de).
- In der Antwort schickt der Webserver in den Headern (vor dem eigentlichen Inhalt) auch eine Last-Modified-Zeile der Form
Last-Modified: Fri, 14 Jan 2011 23:01:49 GMT
Dieses Datum der letzten Änderung wird nicht vom Browser angezeigt, ist aber wichtig. Wofür?
- Teil einer Antwort ist immer auch ein HTTP Status Code. Einige Beispiele sind:

200 OK	301 Permanently Moved	400 Bad Request	500 Internal Server Error
	304 Not Modified	401 Unauthorized	503 Service Unavailable
		404 Not Found	

Die Gruppierung nach führender Ziffer (200er, 300er etc.) entspricht auch einer Kategorie-zugehörigkeit. Überlegen Sie sich sinnvolle Namen für die vier Kategorien.

4. Gruppen-Kurz-Paper

Die Summe der Matrikelnummern aller Mitglieder Ihrer Gruppe (zur Erinnerung: 2er- oder 3er-Gruppen) ist entweder *gerade* oder *ungerade*.

Schreiben Sie gemeinsam (!) ein einseitiges Kurz-Paper (DIN A4, Standardeinstellungen) zum Thema

- „Schwierigkeiten von Archivprogrammen, die auf mehreren Betriebssystemen einsetzbar sein sollen“ (*gerade* Summe) oder
- „Vor- und Nachteile der Linux-Paket-Repositories gegenüber Installationsmethoden unter Windows“ (*ungerade* Summe)

Sammeln Sie zunächst gemeinsam Ideen (Stichworte notieren). Wenn Sie in bestimmten Aspekten verschiedener Meinung sind, nehmen Sie beide Positionen in Ihr Paper auf und weisen Sie darauf hin, dass hier keine Einigkeit herrschte. (Bevor Sie loslegen, sprechen Sie sich bitte kurz mit Nachbargruppen ab, damit nicht alle dasselbe Thema bearbeiten; Sie können bei zu großer Übereinstimmung das Thema wechseln.) Schicken Sie mir die Kurz-Papers als PDF-Datei per Mail (h.g.esser@gmx.de, CC: alle Gruppenmitglieder), Sie erhalten dann in den nächsten Tagen ein kurzes Feedback.

5. Linux-Paketformate

Erklären Sie für Software-Pakete die beiden Begriffe „Abhängigkeit“ und „Konflikt“ und beschreiben Sie, wie eine Repository-basierte Paketverwaltung das Problem der Abhängigkeiten behebt.