

# Aufgabe 1B: XML-Verarbeitung

---

## *Automatische Datenerfassung vom Online-Speiseplan des Studentenwerks*

Das Portal *Lorbeerblatt* soll Studenten eine Plattform bieten, auf der sie sich über das aktuelle Mensaessen austauschen können. Die Benutzer sollen Kommentare, Bewertungen und eigene Bilder der verschiedenen Speisen einstellen können. Dazu muss zunächst der aktuelle Speiseplan von der Website des Studentenwerks eingelesen werden.

Ihre Aufgabe ist es, diesen Einlesevorgang zu automatisieren: Schreiben Sie ein kleines Java-Programm, das sich mit dem Webserver des Studentenwerks verbindet, die Daten der Speisen des aktuellen Tages strukturiert ausliest und als Java-Objekte zur Verfügung stellt. Zu erfassende Daten sind:

- Name der Mensa
- Name des Gerichts
- Preise, getrennt für Studenten und für Mitarbeiter
- Inhaltsstoffe unter „Infos“, als Liste (z.B. Schweinefleisch, vegetarisch, Knoblauch, ...)
- URL des zugehörigen Bildes (falls vorhanden)
- URL der Detailseite der Speise

Schreiben Sie zu Debugzwecken außerdem ein XSL-Stylesheet, welches aus dem Eingabedokument eine einfache Visualisierung in Form einer XHTML-Tabelle erzeugt.

**Tipp:** Beginnen Sie zunächst mit dem XSL-Stylesheet, um ohne aufwendige Programmierung an die für Sie relevanten Seitenelemente heranzukommen. Sie können das Ergebnis der XSL-Transformation auch als „bereinigte“ Ausgangsbasis für die Erzeugung der Java-Objekte verwenden.

## *Das Ergebnis muss folgende Anforderungen erfüllen:*

- Das Parsen des Speiseplans soll so vorgenommen werden, dass kleinere Änderungen auf der Seite nicht zum vollständigen Versagen des Programms führen bzw. der Parser leicht angepasst werden kann. Hierzu kann insbesondere eine XSL-Vorverarbeitung hilfreich sein.
- Sehen Sie eine Logausgabe beim Einlesen der Speiseplandaten vor, anhand der grob festgestellt werden kann, ob der Vorgang erfolgreich war (beispielsweise Anzahl der erfassten Gerichte).
- Verwenden Sie nur tatsächliche Daten des Mensa-Speiseplans. Sie können dazu beispielsweise die Übersichtsseite<sup>1</sup> oder den RSS-Feed<sup>2</sup> benutzen, wobei der Feed jedoch nicht alle erforderlichen Informationen bereitstellt.
- *Alle verwendeten Bestandteile müssen entweder Public Domain sein oder die erforderlichen Nutzungsrechte müssen schriftlich vorliegen.*

## *Abzugeben sind:*

- Quellen (z.B. als gepacktes Verzeichnis)

---

<sup>1</sup> <http://www.studentenwerk-dresden.de/mensen/speiseplan/>

<sup>2</sup> <http://www.studentenwerk-dresden.de/feeds/speiseplan.rss>