

Auswertung von Serverzugriffen

Bei diesem Dokument handelt es sich um eine Aufgabe für Kandidaten der UFirst GmbH im Bereich Software-Entwicklung. Sie ist bewusst relativ umfangreich gestaltet. Es kann aber durchaus nur der Backend-Teil oder nur der Frontend-Teil gelöst werden. Im letzteren Fall werden wir die JSON-Dateien zu Verfügung stellen.

Die Hauptsprachen bei UFirst sind PHP und Javascript, Sie sind aber bei der Umsetzung frei in der Wahl der Programmiersprache. Sie sind ausserdem frei in der Wahl des Setups; Ihre Lösung darf auf einem Webserver laufen, sie darf aber auch komplett ohne Webserver (Kommandozeilen-Scripts und statische HTML/JS-Seiten) auskommen. Sie dürfen und sollen Libraries und Open-Source-Module einsetzen, wo immer Sie diese benötigen.

Zustellung Ihrer Lösung

Bitte lassen sie uns Ihre Lösung innerhalb von 7 Tage Zeit zukommen. Sollte Ihre Lösung von weiteren Systemkomponenten (z.B. Webserver) abhängen, bitten wir Sie, uns die funktionierende Applikation auf dem Web zugänglich zu machen.

Die Aufgabe

Ihre Aufgabe besteht darin, die Access-Logs der EPA vom Jahre 1995 einzulesen, strukturell umzuwandeln und als graphische Auswertung darzustellen.

Vorbereitung

Der folgende Link führt zu einer Seite mit Links zu Access-Logs der EPA. Die Seite beschreibt ausserdem, welche Daten im Access-Logfile enthalten sind und wie die Daten strukturiert sind. Laden Sie die Datei herunter und entpacken Sie sie (Das Entpacken soll nicht Teil Ihres Codes sein.)

- <http://ita.ee.lbl.gov/html/contrib/EPA-HTTP.html>

Aufgabenstellung

1. Schreiben Sie ein Script, welches die Datei einliest und eine **neue** Datei generiert, welche alle Daten der Access-Log-Datei als JSON-Array enthält. (Siehe Beispiel am Ende des Dokuments)
2. Erstellen Sie eine oder mehrere HTML-Dateien, welche mit Hilfe von Javascript-Code die JSON-Files einliest die folgenden Auswertungen graphisch anhand von Diagrammen darstellt.
 - Requests pro Minute über die gesamte Zeit
 - Verhältnis der Methode (GET,POST,HEAD,...)
 - Verhältnis der HTTP-Antwortcodes (200, 404, 302,...)
 - Verteilung der Antwortgrösse zwischen aller Requests mit Antwortcode 200 und Antwortgrösse < 1000B

Hinweise

- Achten Sie darauf, dass ALLE Datensätze eingelesen werden.
- Die Daten sind aus dem Jahre 1995 und enthalten unbekannte Textzeichen - sie sollten in Schritt 1 bereinigt werden.
- Der Ordner "template" kann als Basis für den zweiten Teil verwendet werden.
- Wählen Sie bei den Auswertungen das aus Ihrer Sicht jeweils sinnvollste Diagramm (Liniendiagramm, Balkendiagramm, Kuchendiagramm,...)
- Die Applikation muss nicht von mehreren Browsern unterstützt sein. Geben Sie uns an, in welchem Browser sie läuft.

Beispiel

Beispiel für die Ausgabe-JSON-Datei aus Aufgabe 1 (zur Veranschaulichung wurden im Nachhinein die Zeilenumbrüche und Einrückungen schön formatiert):

```
[
  {
    "host": "141.243.1.172",
    "datetime": {
      "day": "29",
      "hour": "23",
      "minute": "53",
      "second": "25"
    },
    "request": {
      "method": "GET",
      "url": "/Software.html",
      "protocol": "HTTP",
      "protocol_version": "1.0"
    },
    "response_code": "200",
    "document_size": "1497"
  },
  ...weitere Datensätze...
]
```