

Softwaretechnik

Projekt „HTML-Suchmaschine“

Teil 1: Analyse und Design (Übung)

Fallbeschreibung

Ein Unternehmen mit vielen Filialen möchte die Dateiablagen der einzelnen Standorte durch eine übergreifende Suchmaschine verbinden und zugänglich machen. Die Suchmaschine soll die verschiedenen im Unternehmen genutzten Dateiformate lesen können – auch die spezifischen Formate selbst erstellter Programme.

Ihr betreibt als Jungunternehmer eine kleine Softwareentwicklungsfirma. Das Geschäft läuft eher schlecht als recht, so dass ihr sehr an diesem Auftrag interessiert seid. Bevor das Filialunternehmen den Auftrag erteilt, sollt ihr – quasi als Arbeitsprobe – einen ersten Suchmaschinenprototyp in Form einer lokalen HTML-Suchmaschine entwickeln.

Aufgaben

Aufgabe 1: Analyse

Analysiert verschiedene Webseiten und deren Quelltext.

- Wo finden sich im Quelltext für die Suche relevante Informationen?
- Welche Informationen kann/sollte man ignorieren?
- Wie lässt sich das Wesentliche vom Unwesentlichen trennen?

Untersucht relativ übersichtliche Quelltexte (bspw. die Dateien im Ordner `.\ HTML-Testseiten\`) ebenso wie komplexe Seiten (z. B. <http://www.bundesregierung.de/> oder <http://www.tagesschau.de/>). Lassen sich die von Euch gefunden Mechanismen überall anwenden oder gibt es zusätzliche Hürden?

Tipp: Der Quelltext einer HTML-Seite lässt sich in vielen Browsern mit der Tastenkombinationen `Strg + U` aufrufen.

<div data-bbox="156 219 478 302"> <p>SWR Landesschau aktuell <small>RHEINLAND-PFALZ</small></p> </div> <div data-bbox="598 212 790 246"> <p>Mobil Desktop</p> </div> <div data-bbox="156 318 518 353"> <p>Nachrichten Verkehr Wetter</p> </div> <div data-bbox="167 380 399 421"> <p>► Rheinland-Pfalz</p> </div> <div data-bbox="156 443 694 515"> <p>"VeggieDay" Ein fleischloser Tag für die Umwelt</p> </div> <div data-bbox="156 526 790 705"> <p>Im Bundestagswahlkampf hatten die Grünen einen fleischlosen Tag pro Woche in deutschen Kantinen gefordert und dafür herbe Kritik eingesteckt. Manch einer sieht in dem Vorstoß einen Grund für die Verluste der Partei. In Rheinland-Pfalz startete heute sogar offiziell eine "VeggieDay"-Kampagne.</p> </div> <div data-bbox="159 712 678 1003">  </div> <div data-bbox="156 1012 790 1227"> <p>Unter dem Motto "VeggieDay Rheinland-Pfalz" wirbt sie für einen vegetarischen Tag pro Woche in Kantinen. Damit soll ein Beitrag zum Klima-, Umwelt- und Tierschutz sowie zur Gesundheit und der globalen Ernährungssicherheit geleistet werden. Denn übermäßiger Fleischkonsum gilt als ungesund und ökologisch fragwürdig.</p> </div> <div data-bbox="156 1236 790 1361"> <p>Die Albert-Schweitzer Stiftung, der Naturschutzbund (Nabu), das Entwicklungspolitische Landesnetzwerk (ELAN) sowie die Stiftung Natur und Umwelt haben dazu aufgerufen.</p> </div>	<pre> <!DOCTYPE html PUBLIC "-//W3C//DTD HTML <html><head> <meta http-equiv="Content-Type" content= <link type="text/css" rel="stylesheet" <link type="text/css" rel="stylesheet" <title>"VeggieDay": Ein fleischloser Tag <meta content="max-age=300, no-transfor <meta content="width=320, initial-scale <meta content="320" name="MobileOptimiz <meta name="HandheldFriendly" content=" <link type="image/vnd.microsoft.icon" h <link type="image/gif" href="http://m.sw <link href="http://m.swr.de/swr/content <link href="http://m.swr.de/swr/content <link href="http://m.swr.de/swr/content <script src="%27VeggieDay%27%20%20Ein%2 </head> <body onload="load()"> <div class="page"> <div class="header"> <table border="0" cellpadding="0" cells <tbody><tr> <td><a name="top" href="http://m.swr.de <div align="right"> Mobil&nbsp; </div> </td> </tr> </tbody></table> </div> <div id="tabnavi"> <a href="h </pre>
---	---

Eine HTML-Seite (links) und deren Quelltext (rechts)¹

Aufgabe 2: Prozessmodell (Analysephase)

Erstellt in BPMN (*Business Process Model Notation*) ein Prozessmodell, welches den Ablauf des Durchsuchens eines Verzeichnisses nach einem Suchbegriff beschreibt.

Hinweise:

- Ihr werdet eine Reihe von *Modellierungsentscheidungen* treffen müssen. Bedenkt, dass häufig mehrere sinnvolle Optionen existieren.
- Denkt in *Prototypen*: Was soll der zu modellierende Demonstrationsprototyp leisten? Welche Funktionen sollte erst das fertige Programm beinhalten? Wie muss man den

¹ Quelle: <http://m.swr.de/inm/swr/kl042y2340iya>, 24.09.2013

Demonstrationsprototypen gestalten, dass sich die weiteren Funktionalitäten effizient hinzufügen lassen?

Aufgabe 3: Domänenmodell (frühe Designphase)

Erweitert nun euer Prozessmodell um Klassen. Typischerweise sucht man hierbei nach zwei Grundmustern für Klassen:

1. Klassen, die für die Durchführung bestimmter Prozessschritte („Tasks“) zuständig sind.
2. Klassen, welche die in einem Prozess genutzten oder erzeugten *Entitäten* repräsentieren – insb. genutzte Daten, beteiligte Personen oder verarbeitete/erstellte Produkte.

Ergänzt das Klassendiagramm um Assoziations-, Aggregations- und ggf. auch Spezialisierungsbeziehungen.

Aufgabe 4: Implementierungsmodell (späte Designphase)

Ergänzt euer Modell um softwarespezifische Klassen, also Klassen, die für den reibungslosen Betrieb der Software von Bedeutung sind, sich aber aus der Domäne nicht ableiten lassen – bspw. die Steuerung des Prozessablaufs oder der Umgang mit Fehlern.