Prof. Dr. Stefan Brass
Institut für Informatik
Universität Halle

12. Juli 2006

Vorlesung “Grundlagen des World Wide Web” — Klausur —

Name: ______________________________

Studiengang: ______________________________

Matrikelnummer: ______________________________

(Diese Daten werden zur Ausstellung des Leistungsnachweises benötigt.)

Aufgabe	Punkte	Max. Punkte	Zeit
1 (Internet, DNS, HTTP, Zeichensätze)		10	30 min
2 (HTTP Request)		4	10 min
3 (XML)		10	20 min
4 (SGML, HTML, XHTML)		6	15 min
Summe		29	75 min

Hinweise:

- Bearbeitungsdauer: 75 Minuten
- Skript, Bücher, Notizen sind erlaubt. Notebooks, PDAs, etc. dürfen nicht verwendet werden. Mobiltelefone bitte ausschalten (oder mit Aufsicht besprechen).
- Die Klausur hat 9 Seiten. Bitte prüfen Sie die Vollständigkeit.
- Bitte benutzen Sie den vorgegebenen Platz. Wenn Sie auf die Rückseite ausweichen müssen, markieren Sie klar, daß es eine Fortsetzung gibt.
- Tauschen Sie keinesfalls irgendwelche Dinge mit den Nachbarn aus. Notfalls rufen Sie eine Aufsichtsperson zur Kontrolle.
- Bei Aufgabe 1 zum Ankreuzen sollten Sie wenigstens raten, wenn Sie die richtige Lösung nicht wissen.
- Fragen Sie, wenn Ihnen eine Aufgabe nicht klar ist!
- Es wird eine Umfrage zu dieser Vorlesung ins Netz gestellt (voraussichtlich Freitag nachmittag), Teilnehmer der Umfrage werden 1% im Gesamtergebnis hochgestuft.

Aufgabe 1 (Internet, DNS, HTTP, Zeichensätze) 10 Punkte

Bitte kreuzen Sie jeweils die korrekte Antwort an. Pro Teilaufgabe sollte nur genau eine Antwort korrekt sein. Falls die die korrekte Antwort nicht wissen, sollten Sie wenigstens raten: Bei keinem Kreuzchen bekommen Sie ebenso null Punkte wie bei einem falschen Kreuzchen.

a) Zu welchem Host und Port stellt ein Browser (mit aktivierten Prowy-Einstellungen: “`www-cache.uni-halle.de:3128`”) eine Verbindung her, nachdem Sie die folgende URL eingegeben haben: “`http://www.sun.com:8080/sparc/t1/`”?

- ☐ Host: `www.sun.com`, Port: 80
- ☐ Host: `www.sun.com`, Port: 8080
- ☐ Host: `www.sun.com`, Port: 3128
- ☐ Host: `www-cache.uni-halle.de`, Port: 80
- ☐ Host: `www-cache.uni-halle.de`, Port: 8080
- ☐ Host: `www-cache.uni-halle.de`, Port: 3128

b) Welches Protokoll verwendet der Browser, um obige Verbindung herzustellen?

- ☐ `UDP`
- ☐ `TCP/IP`
- ☐ `SMTP`
- ☐ `FTP`

c) Mit welchem Programm können Sie die Hardwareadresse (MAC) eines anderen Rechners im gleichen Subnetz herausfinden?

- ☐ `ping`
- ☐ `whois`
- ☐ `arp`
- ☐ `netstat`
- ☐ `nslookup`

d) Welches Programm würden Sie verwenden, um den/die zuständigen Mailserver (MX) zu einer bestimmten Domain zu bestimmen?

- ☐ `ping`
- ☐ `whois`
- ☐ `arp`
- ☐ `netstat`
- ☐ `nslookup`

e) Verweisen die beiden URLs

- `http://www.informatik.uni-halle.de/dbs/download/`
- `http://www.informatik.uni-halle.de/dbs/download/index.html`

auf dasselbe Dokument?

☐ Ja, immer.
☐ Das hängt von der Konfiguration des Web-Servers ab.
☐ Nein: Verschiedene URLs liefern auch verschiedene Dokumente.

f) Wozu sind URNs gut?

☐ U.a. zur automatischen Wahl des nächsten Servers bei gespiegelten Resourcen.
☐ Eindeutige Namen für Resolver.
☐ Eine einheitliche Sprache für das Retrieval von Informationen aus dem Web.

g) Kann es gleichzeit zwei verschiedene Netzwerk-Verbindungen zum gleichen Port auf dem gleichen Rechner geben?

☐ Ja, wenn die andere Seite der Verbindung sich unterscheidet.
☐ Nein. Der Port muß erst wieder frei werden, bevor er von einer anderen Verbindung genutzt werden kann.

h) Angenommen, Sie verwenden nur ASCII-Zeichen und die üblichen deutschen Umlaute. Macht es dann einen Unterschied, ob Ihr Text in ISO Latin/1 oder Unicode mit der UTF-8 Codierung codiert ist?

☐ Die deutschen Umlaute haben in ISO Latin/1 und Unicode unterschiedliche Codenummern.
☐ ISO Latin/1 und Unicode haben zwar die gleichen Codenummern für die Umlaute, aber in UTF-8 sind sie mit zwei Byte codiert, in ISO Latin/1 nur mit einem Byte.
☐ Nein, das ist kein Unterschied: Solange man nur die deutschen Umlaute verwendet, sind ISO Latin/1 und Unicode UTF-8 vollständig kompatibel.

i) Angenommen, Sie haben für den Server `www.xyz.de` per `nslookup` die IP-Adresse `123.45.67.89` ermittelt. Können Sie dann statt `http://www.xyz.de/` auch einfach die URL `http://123.45.67.89/` abrufen?

☐ Ja, das funktioniert ohne Unterschied.
☐ Bei manchen Servern steckt im Domainnamen auch eine wichtige Information: Ohne den Namen ist die auszuliefernde Seite dann nicht eindeutig bestimmt.
☐ Das Web funktioniert grundsätzlich nur mit den Domainnamen, und nicht mit den IP-Nummern.

j) Sie wollen an einem heien Sommertag mit Ihrem Notebook gerne von Ihrem überhitzten Büro in einen klimatisierten Raum wechseln, welcher über einen Internet-Anschluss an das selbe Subnetz wie Ihr Büro verfügt. Müssen Sie nun, um die Verbindung zum Internet wiederherzustellen, die IP-Adresse Ihres Notebooks ändern? + Ja. + Nein. + Das hngt davon ab, ob das Notebook mit der IP-Adresse im DNS registriert ist.

☐ Ja.
☐ Nein.
☐ Das hängt davon ab, ob das Notebook mit der IP-Adresse im DNS registriert ist.

Aufgabe 2 (Proxy Server) 4 Punkte

Ein Proxy Server empfängt den folgenden Request:

```
GET http://dbs.informatik.uni-halle.de/ HTTP/1.1
Host: dbs.informatik.uni-halle.de
Accept: text/plain, text/html, image/gif
Accept-Language: en
```

Da er keine passende Antwort in seinem Cache hat, schickt er den Request weiter an `dbs.informatik.uni-halle.de`. Er bekommt die folgende Antwort:

```
HTTP/1.1 200 OK
Date: Mon, 10 Jul 2006 15:30:00 GMT
Last-Modified: Wed, 28 Jun 2006 09:05:32 GMT
Expires: Tue, 11 Jul 2006 15:29:59 GMT
ETag: "123456"
Vary: Accept-Language
Content-Length: 2345
Connection: close
Content-Type: text/html

...
```

Angenommen, es ist jetzt “`Wed, 12 Jul 2006, 11:00:00`”. Der Proxy bekommt folgenden Request:

GET http://dbs.informatik.uni-halle.de/ HTTP/1.1
Host: dbs.informatik.uni-halle.de
<u>Accept: text/*, image/gif, image/jpeg, application/msword</u>
<u>User-Agent: Mozilla/4.0 (compatible; MSIE 5.5)</u>
Accept-Language: en

Die Zeilen, in denen sich dieser Request vom ursprünglichen Request unterscheidet, sind unterstrichen.

Was sollte der Proxy jetzt tun (um sich korrekt zu verhalten und Bandbreite zu sparen)? Wenn er einen Request an `dbs.informatik.uni-halle.de` schicken muss, schreiben Sie diesen Request bitte auf:

Aufgabe 3 (XML) 10 Punkte

Sie wollen eine Datenbank Ihrer Liebingsfilme aufbauen, zunächst in Form einer XML-Datei. Entwerfen Sie dazu eine DTD und ein passendes Beispiel-Dokument mit folgenden Daten:

- Über Filme sind der Titel, das Produktionsjahr und die Information, ob der Film schon in Ihrer Sammlung enthalten ist (Spalte `Besitz`) abzuspeichern. In der Spalte “Besitz” sollen die folgenden drei Werte möglich sein: “`V`”: Videokassette, “`D`”: DVD, “`N`”: nein. Außerdem hat jeder Film eine eindeutige ID (besteht aus Buchstaben und Ziffern und fängt mit einen Buchstaben an). Diese ID ist auch außerhalb der Datenbank wichtig (sie haben sie auf die bereits erfassten DVDs und Videokassetten geklebt).

FILM			
<u>ID</u>	Titel	Jahr	Besitz
F001	Genie und Schnauze	1988	V
F002	Die tollen Abenteuer des Monsieur L.	1965	N
F003	Die Götter müssen verrückt sein	1980	D

- Zu den Filmen wollen Sie auch die wichtigsten Schauspieler festhalten:

SCHAUSPIELER			
<u>FILMID</u>	<u>NR</u>	NAME	ROLLE
F001	1	Micheal Cain	Sherlock Holmes
F001	2	Ben Kingsley	Dr. Watson
F002	1	Jean-Paul Belmondo	Arthur Lempereur
F002	2	Ursula Andress	Alexandrine Pinardel
F003	1	N!xau	Xixo

  Die Spalte `NR` dient nur zur Festlegung der Reihenfolge bei der Ausgabe. Sie wird immer von 1 an laufend durchnumeriert (pro Film).

- Außerdem wollen Sie noch die Art/Gattung jedes Films abspeichern (Genre). Filme gehören häufiger zu mehreren Arten:

GENRE	
<u>FILMID</u>	<u>ART</u>
F001	Komödie
F001	Krimi
F002	Komödie
F002	Abenteuer
F003	Komödie
F003	Natur

Es ist Platz für die Lösung auf der nächsten Seite. Geben Sie DTD und Beispieldokument an, sowie ggf. Bedingungen, die oben genannt sind, die Ihre DTD aber nicht sicherstellt.

Platz zur Lösung von Aufgabe 3

Aufgabe 4 (SGML, HTML, XHTML) 6 Punkte

Sie haben folgendes Dokument in HTML verfasst:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
            "http://www.w3.org/TR/html4/strict.dtd">
<META http-equiv="Content-Type"
            content="text/html; charset=iso-8859-1">
<TITLE>Klausuraufgabe</TITLE>
<LINK rel=next href="solution.html">
<H1>Eine Überschrift (mit Umlaut)</h1>
<P>Text
<UL><LI>Erstens.
    <LI>Zweitens.
</UL>
```

Dieses Dokument ist syntaktisch korrekt. Übersetzen Sie das Dokument nach XHTML. Hier eine (stark vereinfachte) DTD für (den relevanten Ausschnitt von) XHTML:

```
<!ELEMENT html        (head, body)>
<!ENTITY  % head.misc "(meta|link)*">
<!ENTITY  % block     "p|h1|ul">
<!ENTITY  % inline    "#PCDATA|em">
<!ELEMENT head        (%head.misc;, title, %head.misc;)>
<!ELEMENT title       (#PCDATA)>
<!ELEMENT meta        EMPTY>
<!ATTLIST meta        http-equiv NMTOKEN #IMPLIED
                      name NMTOKEN #IMPLIED
                      content CDATA #REQUIRED>
<!ELEMENT link        EMPTY>
<!ATTLIST link        href CDATA #REQUIRED
                      rel (start|next|prev) #IMPLIED>
<!ELEMENT body        (%block;)+)>
<!ELEMENT h1          (#PCDATA)>
<!ELEMENT p           (%inline;)*>
<!ELEMENT em          (%inline;)*>
<!ELEMENT ul          (li)+>
<!ELEMENT li          (%inline;|%block;)*>
```

Deklarieren Sie im html-Element den Namespace

```
http://www.w3.org/1999/xhtml
```

Der Public Identifier für XHTML ist

```
"-//W3C//DTD XHTML 1.0 Strict//EN"
```

Der System Identifier ist

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"
```

Es ist Platz für die Lösung auf der nächsten Seite. Dort ist das Eingabedokument noch einmal angegeben.

Lösung zu Aufgabe 4

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
            "http://www.w3.org/TR/html4/strict.dtd">
<META http-equiv="Content-Type"
            content="text/html; charset=iso-8859-1">
<TITLE>Klausuraufgabe</TITLE>
<LINK rel=next href="solution.html">
<H1>Eine Überschrift (mit Umlaut)</h1>
<P>Text
<UL><LI>Erstens.
    <LI>Zweitens.
</UL>
```