Prof. Dr. Stefan Brass
Institut für Informatik
Universität Halle

3. Februar 2005

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)		8	24 min
2 (HTTP Request)		4	10 min
3 (XML)		10	20 min
4 (SGML/HTML)		3	10 min
5 (Namespaces)		4	10 min
Summe		29	74 min

Hinweise:

- Bearbeitungsdauer: 90 Minuten
- Skript, Bücher, Notizen sind erlaubt. Notebooks, PDAs, etc. dürfen nicht verwendet werden. Mobiltelefone bitte ausschalten (oder mit Aufsicht besprechen).
- Die Klausur hat 8 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 Aufgaben 1 und 5 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) 8 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) Mit welchem Befehl können Sie sich den Besitzer bzw. Ansprechpartner der Domain `oracle.com` anzeigen lassen?

- ☐ `ping`
- ☐ `traceroute`
- ☐ `whois`
- ☐ `nslookup`

b) Mit welchem Befehl können Sie den für eine Domain zuständigen Mailserver herausfinden?

- ☐ `ping`
- ☐ `traceroute`
- ☐ `whois`
- ☐ `nslookup`

c) Wer verwaltet die Vergabe der Domainnamen für den deutschen Adressraum (`.de`)?

- ☐ DENIC
- ☐ Microsoft
- ☐ RIPE NCC
- ☐ ICANN

d) Wie entstand das Internet?

- ☐ Microsoft hat es erfunden.
- ☐ Es ist eine Weiterentwicklung des Faxabrufes.
- ☐ Aus einem vom US-Verteidigungsministerium geförderten Wissenschaftsprojekt.

e) Kennt ein Webserver die IP-Adresse des Rechners, auf dem der Client (Browser) läuft?

- ☐ Ja (bzw. die IP-Nummer der Firewall, falls der Browser hinter einer Firewall ist).
- ☐ Nein. Ein solches Feld wurde beim HTTP-Protokoll leider vergessen.
- ☐ Nur noch ganz selten. Es gibt ein solches Feld im HTTP-Protokoll, aber heutige Browser setzen es aus Datenschutzgründen nicht mehr.

f) Was ist der Zweck von Cookies?

- ☐ Die Verknüpfung verschiedener Requests des gleichen Benutzers
- ☐ Der Zugriff auf Passwort-geschützte Seiten.
- ☐ Die Leistungssteigerung durch Pufferung von Daten auf dem Proxy.

g) Was ist "Chunked Encoding"?

- ☐ Ein Datenkompressionsverfahren.
- ☐ Das Verfahren, das Router beim IP-Protokoll verwenden, falls die Datenpakete zu groß sind.
- ☐ HTTP-Server verwenden das, wenn Sie die Größe der Webseite nicht kennen.
- ☐ Chunked Encoding dient dem Zugriff auf Passwort-geschützte Daten.

h) Eine XHTML-Datei enthält deutsche Umlaute im ISO Latin/1 Zeichensatz. Müssen Sie die XML-Deklaration verwenden, um den Zeichensatz explizit anzugeben (mit `encoding="ISO-8859-1"`)?

- ☐ Ja, das ist notwendig.
- ☐ Nein, Unicode/UTF-8 ist aufwärtskompatibel zu ISO Latin/1, und dies ist der Default. Daher ist jede in ISO Latin/1 codierte Datei automatisch auch in Unicode (UTF-8) codiert.

Aufgabe 2 (HTTP Request) 4 Punkte

Schreiben Sie einen HTTP/1.1 Request, der über den Proxy `www-cache.uni-halle.de` auf Port `3128` die folgende URL abruft:

`http://www.acm.org/dl/`

Geben Sie dabei noch folgendes an:

- Sie benötigen die aktuelle Version (nicht aus dem cache).
- Sie akzeptieren sowohl HTML (`text/html`) als auch reinen Text (`text/plain`), würden aber HTML vorziehen.
- Die Verbing soll für weitere Anfragen offen gehalten werden.

Aufgabe 3 (XML) 10 Punkte

Eine Bibliothek möchte ihren Buchbestand als XML-Datei ins Netz stellen. Entwerfen Sie dazu eine XML DTD und ein passendes Beispiel-Dokument mit folgenden Daten:

- Über Bücher sind Autor und Titel sowie eine eindeutige Signatur abzuspeichern (besteht aus Buchstanden und Ziffern und fängt mit einen Buchstaben an):

BUCH		
Signatur	Autor	Titel
CS1001	Erik Wilde	World Wide Web
CS1002	Lincoln D. Stein	Web Security

- Von einem Buch kann es eventuell mehrere Exemplare geben (z.B. bei Lehrbüchern). Buchexemplare werden über die Signatur des Buches sowie eine Exemplarnummer identifiziert. Selbstverständlich gibt es die Signatur immer in der `BUCH`-Tabelle. Falls das Buch(-Exemplar) gerade ausgeliehen ist, wird auch das Datum abgespeichert, zu dem das Buch wieder abgegeben werden muß. In diesem Fall ist auch wichtig, ob das Buch schon vorgemerkt ist oder nicht. Falls das Buch nicht ausgeliehen ist, kann es auch nicht vorgemerkt sein.

EXEMPLAR			
Signatur	NR	ENTLIEHEN_BIS	VORGEMERKT
CS1001	1	10.02.2005	N
CS1001	2		N
CS1002	1	14.02.2005	Y

- Schließlich werden zu jedem Buch noch Schlagworte abgespeichert, um zu einem Thema besser die passenden Bücher zu finden. Schlagworte beziehen sich natürlich auf die Bücher, nicht die Exemplare. Die Signaturen müssen in der `BUCH`-Tabelle auch vorkommen.

SCHLAGWORT	
Signatur	BEGRIFF
CS1001	WWW
CS1001	HTTP
CS1002	WWW

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) — 3 Punkte

Gegeben sei folgende stark vereinfachte SGML DTD für HTML:

```
<!ELEMENT html  O O (head, body)>
<!ELEMENT head  O O (title)>
<!ELEMENT title - - (#PCDATA)>
<!ELEMENT body  O O ((p|h1)*)>
<!ELEMENT h1    - - (#PCDATA)>
<!ELEMENT p     - O ((#PCDATA|em|a|img)*)>
<!ELEMENT em    - - (#PCDATA)>
<!ELEMENT a     - - (#PCDATA|em|img)>
<!ATTLIST a     href CDATA #REQUIRED
                rel (start|next|prev) #IMPLIED>
<!ELEMENT img   - O EMPTY>
<!ATTLIST img   src CDATA #REQUIRED>
```

Folgendes HTML-Dokument enthält (mindestens) drei Fehler. Finden Sie diese Fehler. Nehmen Sie dabei an, daß die obige DTD die HTML 4.01 strict DTD ist, was natürlich nicht stimmt. Ich habe mich bemüht, nur drei Fehler in das Dokument einzubauen. Wenn Sie weitere Fehler finden, bekommen Sie Extrapunkte. Andererseits ziehe ich Ihnen für jede falsche Angabe, die über die drei geplanten Fehler hinaus geht, einen Punkt ab. Deswegen sollten Sie mehr als drei Fehler nur angeben, wenn Sie ganz sicher sind. Bis zu drei Fehlern können Sie notfalls raten.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
        "http://www.w3.org/TR/html4/strict.dtd">
<html>
    <title>My Second HTML document!</title>
    <body>
    This is my second HTML document.
    <p><a href="hello.c">Hello</a>, world!
    <img src="brass.jpg">Picture of me</img>
    <a rel="bigversion" href="nina2.jpg">
        <img src="nina.jpg">Picture of my wife</a>
</html>
```

1. ______________________________

2. ______________________________

3. ______________________________

Aufgabe 5 (Namespaces) — 4 Punkte

Gegeben sei folgendes XML-Dokument:

```
(1)  <a xmlns="http://n1.com" xmlns:x="http://n2.com">
(2)      <b xmlns:x="http://www.n3.com"/>
(3)      <x:c xmlns="http://www.n4.com"/>
(4)      <x:d/>
(5)  </a>
```

Welches Element hat welchen Namespace?

a) Element `a` (Zeile 1)?

☐ Keinen
☐ `http://n1.com`
☐ `http://n2.com`
☐ `http://n3.com`
☐ `http://n4.com`

b) Element `b` (Zeile 2)?

☐ Keinen
☐ `http://n1.com`
☐ `http://n2.com`
☐ `http://n3.com`
☐ `http://n4.com`

c) Element `x:c` (Zeile 3)?

☐ Keinen
☐ `http://n1.com`
☐ `http://n2.com`
☐ `http://n3.com`
☐ `http://n4.com`

d) Element `x:d` (Zeile 4)?

☐ Keinen
☐ `http://n1.com`
☐ `http://n2.com`
☐ `http://n3.com`
☐ `http://n4.com`