Technische Universität München
Fakultät für Informatik
Lehrstuhl für effiziente Algorithmen
Prof. Dr. E.W. Mayr, Prof. Dr. A. Steger

Proseminar

Algorithmen für Digitale Bibliotheken

Sommersemester 2003
Prof. Dr. E.W. Mayr, Prof. Dr. A. Steger

1 Motivation

Das Internet bietet ein Fülle an Informationen. Springer, IEEE und ACM sind nur einige Beispiele, die Publikationen per Internet zur Verfügung stellen.

Der Lehrstuhl für Effiziente Algorithmen betreibt ebenfalls eine Literaturdatenbank (LEA-BIB) und arbeitet am Projekt FIS-I (Fachinformationssystem Informatik) mit. Ziel des Projektes ist es, die Literaturdatenbanken mehrerer Institute zusammenzuführen.

Die Pflege und der Betrieb dieser digitalen Bibliotheken stellen hohe Anforderungen an die zugrundeliegenden Softwaresysteme.
Einerseits ist es nötig, laufend die Datenbank mit neu erschienen Literaturdaten zu aktualisieren, wobei die Erfassung der Daten möglichst automatisch erfolgen sollte. Aus diesem Grund ist es aber nötig, die gesammelten Daten vor dem endgültigen Einfügen in die Literaturdatenbank nochmals mit geeigneten Verfahren auf Richtigkeit zu überprüfen (Seitenzahlen, Bandnummern, Jahreszahlen, etc.).
Andererseits muß eine Möglichkeit geschaffen werden, die Daten in der Literaturdatenbank schnell und effizient abfragen zu können. Dabei soll eine Suche nach bestimmten Schlagwörtern, Autoren, Titeln, etc. mittels exakter oder fehlertoleranter Suche möglich sein.

In diesem Proseminar sollen algorithmische Verfahren erarbeitet werden, die die Pflege der Datenbestände erleichtern und den Nutzen der Literaturdatenbanken steigern.

2 Themen

1. Sprachen zur Wrapper-Generierung
2. HTML-verarbeitende Algorithmen
3. Wrapper
4. Modell-basierte Algorithmen
5. Ontologie-basierte Algorithmen

6. Datenvalidierung mit Hilfe von Grammatiken
7. Datenvalidierung mit Hilfe von regulären Ausdrücken
8. Der KMP Algorithmus (Knuth, Morris, Pratt)
9. Suffix-Trees (Ukkonen)
10. Der Boyer-Moore-Algortithmus
11. Abstandsdefinition in Texten (EDIT-Distance, Hamming-Distance, etc.)
12. agrep

3 Schein

Ein Schein für die erfolgreiche Teilnahme am Proseminar wird vergeben, wenn folgende Leistungen erbracht worden sind (die Gesamtnote setzt sich aus gewichteten Einzelnoten zusammen):

Probevortrag (ohne Bewertung)	Der Probevortrag erfolgt spätestens zum angegebenen Termin beim Betreuer. Vorzulegen sind dabei die fertig ausgearbeiteten Folien oder ähnliche Präsentationshilfsmittel und die Erstfassung der Seminararbeit. Vereinbaren Sie für den Probevortrag rechtzeitig einen Termin beim Betreuer (spätestens eine Woche vor dem anvisierten Termin).
Seminarvortrag (in mindestens zufriedenstellender Qualität)	Der Seminarvortrag ist zum festgelegten Termin zu halten und dauert 45 (+/-5) Minuten. Tafelvorträge werden nicht akzeptiert. Nach dem Vortrag muss auf Fragen aus dem Publikum eingegangen werden.
Seminararbeit (in mindestens zufriedenstellender Qualität)	Die Endfassung der Seminararbeit ist spätestens 2 Wochen nach dem Votrag als TEX-Datei und Postscript-Datei abzugeben. Der Umfang der Seminararbeit beträgt 10 (+/- 1) Seiten (ohne Literaturverzeichnis) unter LATEX.
Anwesenheit bei den Vorträgen	Bei den einzelnen Vorträgen wird eine Anwesenheitsliste geführt.

4 Ablauf der Vorbereitung

Der Vortrag und die Ausarbeitung müssen mit dem Betreuer abgesprochen werden. Hierzu sind nachfolgende Terminvorgaben bindend (soweit nicht anders mit dem Betreuer abgestimmt). Werden die Termine nicht eingehalten, führt dies zur Streichung des Vortrags und zum Nichtbestehen des Seminars.

bis 5 Wochen vor dem Vortrag	erstes Treffen mit dem Betreuer (vor dem Treffen ist die Literatur bereits zu lesen); der genaue Termin ist bei den Vortragsterminen angegeben;
bis 3 Wochen vor dem Vortrag	Gliederung des Vortrags und der Ausarbeitung mit dem Betreuer besprechen;
bis 1 Woche vor dem Vortrag	Probevortrag vor den Betreuer; fertige Folien und vollständige erste Version der Ausarbeitung mit dem Betreuer abstimmen;
bis 2 Wochen nach dem Vortrag	fertige Ausarbeitung abgeben;

5 Weiteres

Weitere Informationen über das **Anfertigung der Ausarbeitung** und **Hinweise zur Gestaltung der Vorträge** finden Sie unter
http://www14.in.tum.de/lehre/2003SS/proseminar