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
23. Januar 2003

Proseminar

Algorithmen für Digitale Bibliotheken

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

Themenliste

1. Sprachen zur Wrapper-Generierung

   [1] Alberto H. F. Laender , Berthier A. Ribeiro-Neto , Altigran S. da Silva , Juliana S. Teixeira. A brief survey of web data extraction tools. ACM SIGMOD Record. June 2002

   [2] Valter Crescenzi, Giansalvatore Mecca. Grammars Have Exceptions. Information Systems. 1998

   [3] Gustavo O. Arocena, Alberto O. Mendelzon. WebOQL: Restructuring Documents, Databases, and Webs. Proceedings if the 14th IEEE International Conference on Data Engineering. 1998

2. HTML-verarbeitende Algorithmen

   [1] Alberto H. F. Laender , Berthier A. Ribeiro-Neto , Altigran S. da Silva , Juliana S. Teixeira. A brief survey of web data extraction tools. ACM SIGMOD Record. June 2002

   [4] World Wide Web Consortium. W3C. The Document Object Model. http://www.w3.org/DOM

   [5] Ling Liu, Calton Pu, Wei Han. XWRAP: An XML-Enabled Wrapper Construction System for Web Information Sources. Proceedings of the 16th IEEE International Conference on Data Engeneering. 2000

3. Wrapper

   [1] Alberto H. F. Laender , Berthier A. Ribeiro-Neto , Altigran S. da Silva , Juliana S. Teixeira. A brief survey of web data extraction tools. ACM SIGMOD Record. June 2002

   [6] Naveen Ashish , Craig A. Knoblock. Wrapper generation for semi-structured Internet sources. ACM SIGMOD Record December 1997.Volume 26 Issue 4

   [7] Nicholas Kushmerick. Wrapper induction: Efficiency and expressiveness. Artificial Intelligence Journal 118. 2000

4. Modell-basierte Algorithmen

[1] Alberto H. F. Laender , Berthier A. Ribeiro-Neto , Altigran S. da Silva , Juliana S. Teixeira. A brief survey of web data extraction tools. ACM SIGMOD Record. June 2002

[8] Brad Adelberg. NoDoSE A tool for Semi-Automatically Extracting Structured and Semistructured Data from Text Documents. SIGMOD Record 27. 1998

[9] Berthier Ribeiro-Neto, Alberto H.F. Laender, Altigran S. da Silva. Extracting Semi-Structured Data Through Examples. Proceedings of the 8th ACM International Conference on Information and Knowledge Management. 1999

5. Ontologie-basierte Algorithmen

[1] Alberto H. F. Laender , Berthier A. Ribeiro-Neto , Altigran S. da Silva , Juliana S. Teixeira. A brief survey of web data extraction tools. ACM SIGMOD Record. June 2002

[10] David W. Embley , Cui Tao , Stephen W. Liddle. Automatically Extracting Ontologically Specified Data from HTML Tables of Unknown Structure. Volume 2503, Issue , pp 322-337. Lecture Notes in Computer Science

[11] D.W. Embley, Y.S. Jiang, Y.-K. Ng. Record-Boundary Discovery in Web Documents. Proceedings ACM SIGMOD International Conference on Management of Data. 1999

6. Datenvalidierung mit Hilfe von Grammatiken

[12] Wegener, Ingo. Theoretische Informatik. B.G. Teubner. Stuttgart 1993

7. Datenvalidierung mit Hilfe von regulären Ausdrücken

[13] Herrmann, Eric C. Mastering Perl 5. Sybex. 1989

[14] Friedl, Jeffrey E. F. Mastering Regular Expressions. O'REILLY 2002

8. Der KMP Algorithmus (Knuth, Morris, Pratt)

[15] Dan Gusfield. Algorithms on strings, trees, and sequences. Cambridge University Press. 1997

[16] Alfred V. Aho , Margaret J. Corasick. Efficient string matching: an aid to bibliographic search. Communications of the ACM June 1975. Volume 18 Issue 6

[17] Volker Heun. Grundlegende Algorithmen. Vieweg Wiesbaden. 2000

[18] Knuth, D. E., Morris, J. H., Pratt, V.R. Fast pattern matching in strings, SIAM J. Comput. 6, 2 (1977), 323-350

9. Suffix-Trees (Ukkonen)

[15] Dan Gusfield. Algorithms on strings, trees, and sequences. Cambridge University Press. 1997

[17] Volker Heun. Grundlegende Algorithmen. Vieweg Wiesbaden. 2000

[19] Edward M. McCreight. A Space-Economical Suffix Tree Construction Algorithm. Journal of the ACM (JACM) April 1976. Volume 23 Issue 2

[20] E. Ukkonen. On-Line Construction of Suffix Trees. Algorithmica, 14, 1995

[21] P. Weiner. Linear pattern matching. IEEE 14th Annual Symposium on Switching and Automata Therory. IEEE, 1973

10. Der Boyer - Moore - Algortithmus

[15] Dan Gusfield. Algorithms on strings, trees, and sequences. Cambridge University Press. 1997

[17] Volker Heun. Grundlegende Algorithmen. Vieweg Wiesbaden. 2000

[22] Robert S. Boyer , J. Strother Moore. A fast string searching algorithm. Communications of the ACM October 1977. Volume 20 Issue 10

11. Abstandsdefinition in Texten (EDIT-Dist., Hamming-Distance, ...)

[15] Dan Gusfield. Algorithms on strings, trees, and sequences. Cambridge University Press. 1997

[23] Gonzalo Navarro. A guided tour to approximate string matching. ACM Computing Surveys (CSUR) March 2001. Volume 33 Issue 1

[24] Robert A. Wagner , Michael J. Fischer. The String-to-String Correction Problem. Journal of the ACM Volume 21, Issue 1. January 1974

[25] C. K. Wong , Ashok K. Chandra. Bounds for the String Editing Problem. Journal of the ACM Volume 23, Issue 1. January 1976

[26] Ming Li, Bin Ma. On The Closest String and Substring Problems. Journal of the ACM. 2002

12. agrep

[27] Wu, S.. and Manber. U. Agrep - A fast approximate pattern-matching tool. Usenix Winter 1992 Technical Conference (San Francisco Jan. 1992), pp. 153-162

[28] Sun Wu , Udi Manber. Fast text searching allowing errors. Communications of the ACM October 1992. Volume 35 Issue 10

[29] Ricardo Baeza-Yates , Gaston H. Gonnet. A new approach to text searching. Communications of the ACM October 1992. Volume 35 Issue 10