

Languages, Translation (Compilers) and Stringology Group

The Prague Stringology Club

Head: Prof. Ing. Bořivoj Melichar, DrSc.

Contact person: Ing. Jan Holub, Ph.D.

about 15 members

<http://www.stringology.org>



Topics

- **Stringology (algorithms on strings)**
 - string processing applied in musicology, DNA processing, etc
- **Data compression**
 - comparison of data compression algorithms, design of specialized compression algorithms
- **Compiler Construction**
 - compilers and parsing; effective, incremental, and parallel algorithms
- **Arbology (algorithms on trees)**
 - tree pattern matching, tree indexing

Stringology

contact person: Jan Holub, Ph.D.; holub@fel.cvut.cz

- Algorithms on strings
- Complex patterns searching algorithms
- Selected Publications:
 - 2008: Holub J., Smyth W. F., Wang S. **Fast Pattern-Matching on Indeterminate Strings.** Journal of Discrete Algorithms. 2008, vol. 6, no. 1, p. 37-50. ISSN 1570-8667.
 - 2008: Holub, J. **Factor automata and Their Implementations.** Habilitation thesis. CTU FEE, Prague, 2008.

- Selected citations:

- Lecroq T. **Fast exact string matching algorithms** Information Processing Letters 102 (6): 229-235 Jun 15 2007, ISSN 0020-0190, SCI Expanded, 2007
- Epifanio C., Mignosi F., Shallit J., Venturin I. **On Sturmian graphs.** Discrete Applied Mathematics 155 (8): 1014-1030, ISSN 0166-218X, SCI, 2007
- Lemstrom K., Pienimaki A. **On comparing edit distance and geometric frameworks in content-based retrieval of symbolically encoded polyphonic music.** Musicae Scientiae 135-152 Sp. Iss. SI, 2007, ISSN 1029-8649, 2007
- etc...

Teaching:

- text algorithms, automata and languages, ATHENS Courses (Advanced Technology Higher Education Network, Socrates)

- **Projects:**

- **Text Processing and Analysis** (Czech Science Foundation project No. GA201/06/1039)

- like exact and approximate string matching,
- finding nontrivial exact and approximate patterns like repeats, palindromes, seeds, covers.
- Those problems came from bioinformatics, musicology, data compression and other research fields.

- **Indexing Automata**

An efficient implementation of indexing automata over a given text

They allows to find a pattern in time linear to the length of pattern regardless the length of the text.

Data compression

contact person: Jan Holub, Ph.D; holub@fel.cvut.cz

- Specialized data compression algorithms
- Selected publications:
 - 2008: Lahoda J., Melichar B., Žďárek J. **Pattern Matching in DCA Coded Text**. In Proceedings of the 13th International Conference on Implementation and Application of Automata. Heidelberg: Springer, 2008, p. 151-160. ISBN 978-3-540-70843-8.
 - 2008: Fiala M., Holub J. **DCA using Suffix Arrays** . In Data Compression Conference 2008. Los Alamitos: IEEE Computer Society, 2008, p. 516. ISBN 0-7695-3121-0.

- **Projects:**

- Data Compression in Memory Databases (Sitronics Center)

The project aims to design compression algorithm to be used in memory databases which has very strong emphasis to the speed of compression and decompression.

Teaching:

- Data compression,
- ATHENS Courses (Advanced Technology Higher Education Network, Socrates)

Compiler Construction

contact person: Jan Janoušek, Ph.D.;
janousej@fel.cvut.cz

- Parsing, translation, and virtual machines.
- Incremental and parallel algorithms.
- Selected Publications:
 - Aycok , J., Horspool, N., Janoušek, J., Melichar, B.: **Even Faster Generalized LR Parsing**. Acta Informatica. Springer, 2001, vol. 37, no. 9, p. 633-651 . ISSN 0001-5903.
 - Ladislav Vagner, L., Melichar, B.: **Parallel LL Parsing** Acta Informatica. Springer, 2007, vol. 44, no. 1, p. 1-21. ISSN 0001-5903.

- **Selected Citations:**

- Dick Grune and Criel J. H. Jacobs **Parsing Techniques. A Practical Guide. 2nd edition.** Springer New York, ISBN 978-0-387-20248-8, ISSN , 2008
- Scott, E., Johnstone, A. **Reducing non-determinism in right nulled GLR parsers**
ACTA INFORMATICA 40 (6-7), ISSN 0001-5903, SCI, 2004
- Scott, E., Johnstone, A., Economopoulos, Rob **BRNGLR: a cubic Tomita-style GLR parsing algorithm.** Acta Informatica, Volume 44, Number 6, October 2007 , pp. 427-461(35), ISSN 0001-5903, SCI, 2007
- B.A. Malloy, T.H. Gibbs, J.F. Power **Decorating tokens to facilitate recognition of ambiguous language constructs.** Software-Practice & Experience, ISSN 0038-0644, SCI Expanded, 2003
- McPeak S, Necula GC **Elkhound: A fast, practical GLR parser generator**
LECTURE NOTES IN COMPUTER SCIENCE, ISSN 0302-9743, SCI Expanded, 2004.
- etc...

- Commercial Projects:

- 2002-2007 Verilog-AMS and MAST compilers (Lynguent)

This project deals with compilers from and to analog hardware description languages Verilog-AMS and MAST. Also, the corresponding IDE tool is a part of this project.

About 15 students participating during the years.

- Since 2007: IEC 61131 compliant structured text development and run-time system (Energocentrum)

The goal of this project is to design and implement a complex development environment and portable run-time environment for industrial PLC controllers. The project comprises

IDE tools for IEC 61131 sources,
a compiler, linker and a virtual machine.

- Teaching:

- Compiler construction, Operation systems, Programming: C++, Java, Lisp

Arbology

contact person: Jan Janoušek, Ph.D.;
janousej@fel.cvut.cz

- tree matching algorithms, tree indexing
- since 2008, first papers submitted
- Possible applications:
 - compiler construction: code optimization
 - implementing interpreters
 - searching in trees