

ICT RESEARCH AT Charles university for open access to scientific information

Click to edit Master subtitle style

Martin Nečaský, Ph.D.

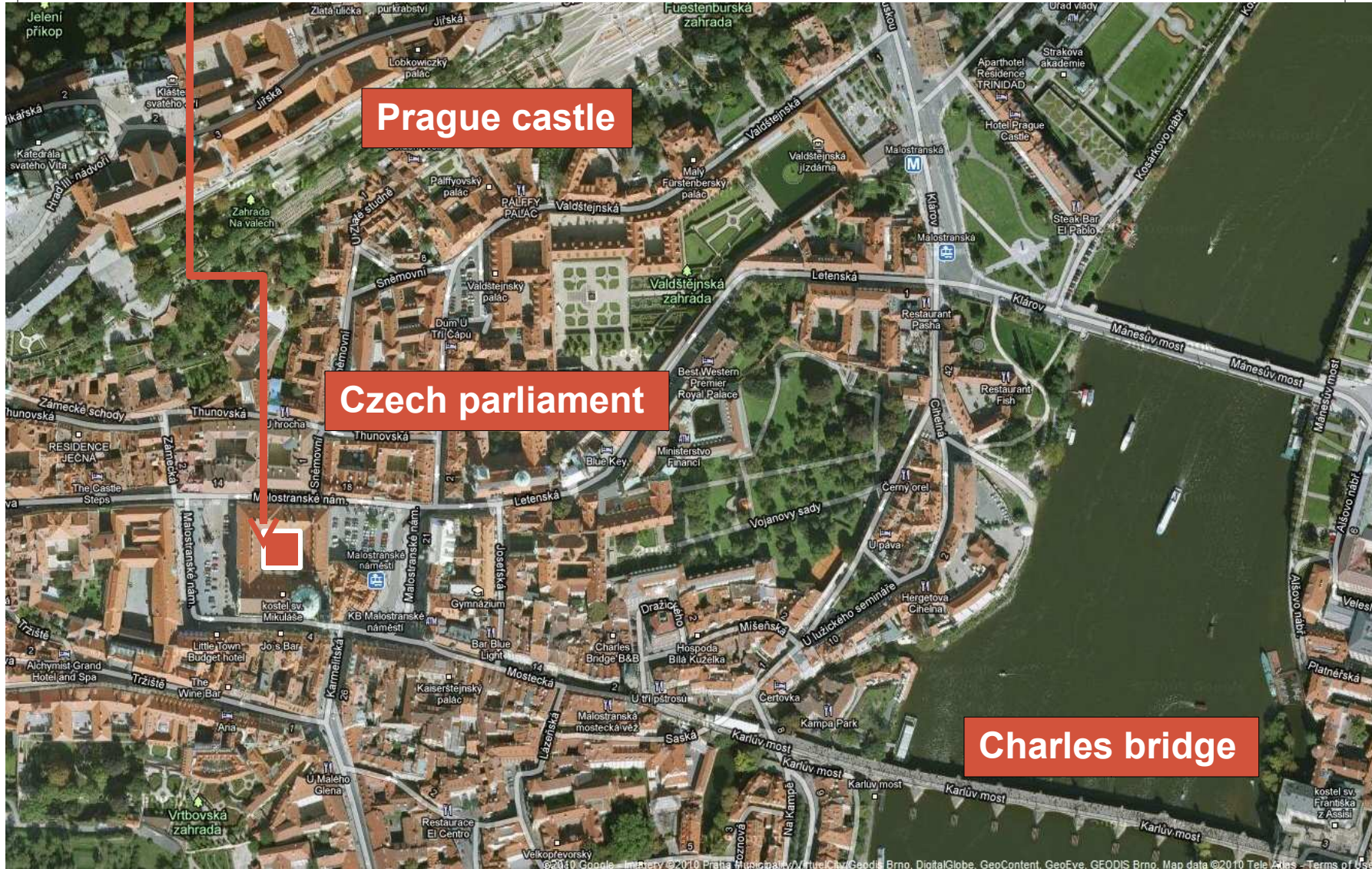
Department of Software Engineering
Faculty of Mathematics and Physics
Charles University in Prague, Czech Republic

Brussels, 28 May 2010

Our Institution

- **Charles University in Prague**
 - founded in 1348
 - 17 faculties
 - over 42.000 students
 - <http://www.cuni.cz>
- **Department of Software Engineering**
 - one of 7 computer science departments at Faculty of Mathematics and Physics
 - 12 academic staff employees and 14 Ph.D. students
 - <http://www.ksi.mff.cuni.cz>

Our Institution on Google Maps



Our International Cooperation

- **Research**

- ERCIS member
- EU
 - Fraunhofer ISST in Berlin, Technische Universität Wien, Johannes Kepler Universität Linz, University of Twente, Aristotele University in Thessaloniki, Universität Rostock, ...
- non-EU
 - Victoria University of Wellington in New Zealand, University of Chile, ...

- **Education**

- ERASMUS, SOCRATES

- **Organized Conferences and Workshops**

Our Basic Research

- our basic research areas related to
“ICT for open access to scientific information”
 - **web semantization**
 - annotation + extraction + similarity search + user preferences
 - **complex systems management**
 - modeling + integration + evolution + programming model
 - **processing large data volumes**
 - storage + indexing + querying + parallelization
 - **robotics**

Web Semantization

- **Problem:** Current Web (i.e. linked documents and multimedia) is designed to be consumed by humans but not machines.
 - *semantics* hidden between lines and hardly interpretable by machines
 - current *Web of Documents* vs. Future *Web of Data*
- **Contribution:** Enrichment of *Web of Documents* with *Web of Data* elements
 - semantic classification and annotation
 - extraction of semantic relationships
 - searching for similarities
 - modeling user preferences

Complex Systems Management

- **Problem:** Future information systems will be complex added-value systems built on top of existing third-party services and data sources operating on Web of Data.
 - many heterogeneous services and data sources which unexpectedly change in time
- **Contribution:** Methods for modeling, integration, evolution and programming of such complex systems
 - integrating heterogeneous components via their common conceptual schema
 - propagation of changes between components
 - services with user and problem oriented interfaces
 - implementation models for Web of Data applications

Processing Large Data Volumes

- **Problem:** Future Web of Data will contain huge amount of interlinked and more or less structured data which will be processed in a database-like way.
 - Web of Data ~ graph → classical relational database model will not be sufficient
 - Web of Data as *“cloud” database*
- **Contribution:** New database methods for Web of Data
 - new query languages
 - non-relational native repository
 - native indexing methods
 - parallelization and distribution

Our Basic Research - Projects

- *EU:*
 - *DaemonX - A Framework for Efficient Design, Adaptation and Evolution Management of a Native XML System (in preparation for FP7)*
 - *Eurobot: Robotic Educational Events to Promote a Dissemination of Science and Technology among Young People in Europe (FP6)*
- *Czech Republic*
 - *Methods and Tools for Web of Data (submitted)*
 - *Handling XML Data in Heterogeneous and Dynamic Environments*
 - started in 2010
 - *Web Semantization*

Our Applied and Experimental Research Projects

- *SoSIReCR* (EU)
 - information support for social network of computer specialists (researchers, practitioners, students) in Czech republic
 - <http://www.sosirecr.cz>
- *LINK* (Technological Agency of Czech Republic)
 - Intelligent digital library of legislative documents in Czech republic and European Union
 - official start in Jan 2011

Applied and Experimental Research

- complete list in <http://www.ksi.mff.cuni.cz/sw/>
- *XCase*
 - Experimental tool for XML modeling, integration and evolution
 - <http://www.ksi.mff.cuni.cz/sw/xcase.html>
- *Analyzer*
 - Tool for batch file analysis
 - <http://www.ksi.mff.cuni.cz/sw/flexbench.html>
- *AgentMat*
 - Framework for data scraping and semantization
 - <http://www.ksi.mff.cuni.cz/sw/agentmat.html>

Summary and Research Directions

- “*open access to (scientific) information*” depends on ability to
 - extract machine-interpretable semantics from current digital resources intended to be interpreted by humans
 - process huge amount of extracted information in database-like way
 - build innovative information systems on top of heterogeneous services and data sources (“scientific mash-ups”)
 - present information in user-friendly as well as machine-friendly way
- our department currently covers most of these research areas
 - ranging from basic to experimental research

QUESTIONS?