Distributed Information Systems
Lecture WS 2007/08

SSC and SI Masters

Time and Place
Lecture: Tuesday 8-10 Room INM 200
Exercise: Tuesday 10-11 Room INM 200

Karl Aberer
Distributed Information Systems Laboratory
Goals of the Course

• What is a "Distributed Information System"?
  - e.g. Web search engines, Web data management, mobile data management etc.

• Which are key problems studied for DIS?
  - e.g. efficient search, abstraction and modelling, optimization of resource usage etc.

• What are typical techniques used to solve these problems?
  - e.g. XML storage and querying, vector space retrieval, association rule mining etc.

• How to apply these techniques?

©2007/8, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis
Focus of the Course

- **Models and Algorithms** for representing, storing and processing information
  - Related systems aspect covered in "Advanced Databases", "Middleware", "Advanced computer networks and distributed systems"

- **Models and algorithms** for representing, storing and processing information on the **Web**
  - Relational data model covered in "Relational Databases" and "Introduction to Information Systems"
The Course - Lecture

- Lecture
  - standard ex cathedra lecture
  - but feel free to interrupt, ask questions ...
    ... even if it is early in the morning and cold outside
  - I also will ask questions once in a while ...

- A first question
  - (how many) paper copies of course notes?
The Course - Exercises

- Exercises
  - can be solved by hand or using programming
  - some are much easier solved by programming :-) (java, matlab, mathematica, excel)
  - we will provide input/output for selected algorithmic problems
  - no correction
  - solutions presented and discussed during exercise hour
  - exercises from previous years will be made available as well
The Exam

- Two midterm exams and one final exam (written)
  - midterms contribute 25% each to final grade, if improvement

- Conceptual questions and practical problems
  - will assume you attended the lecture
  - will assume you did the programming exercises
  - examples from earlier years (exercises, exams) provided for preparation

- Support: Lecture Slides + Exercises + Handwritten Notes
### Time Schedule (indicative)

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Lectures</th>
<th>Programming exercises and exams</th>
<th>Assistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1 2007.Sep.18</td>
<td>Organisational Info</td>
<td>Recap XML</td>
<td></td>
</tr>
</tbody>
</table>

#### Semi-structured Data Management
- Week 3 2007.Oct.2: Graph Databases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 7 2007.Oct.30</td>
<td>Data Broadcasting in Mobile Networks</td>
<td>Implementation of the band energy algorithm</td>
<td>All</td>
</tr>
<tr>
<td>Week 8 2007.Nov.6</td>
<td>Peer-2-Peer Systems (general)</td>
<td>Broadcast disks</td>
<td>All</td>
</tr>
<tr>
<td>Week 9 2007.Nov.13</td>
<td>Structured Overlay Networks</td>
<td>Peer-2-Peer Systems</td>
<td>Hung</td>
</tr>
</tbody>
</table>

#### Information Retrieval and Data Mining
- Week 10 2007.Nov.20: Vector Space Model
- Week 12 2007.Dec.4: Inverted Files and Link-based Ranking
- Week 13 2007.Dec.11: Data Mining (Association rules)
- Week 14 2007.Dec.18: Data Mining (Clustering, Classification)

<table>
<thead>
<tr>
<th></th>
<th>Week 17 2008.Jan.15</th>
<th>Final exam (covers the entire lecture)</th>
</tr>
</thead>
</table>

©2007/8, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis
### Organizational Info

- **Web site**
  - [http://lsirwww.epfl.ch](http://lsirwww.epfl.ch) *(menu item Students)*

<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Email</th>
<th>Office</th>
<th>Office hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof. Karl Aberer</td>
<td><a href="mailto:karl.aberer@epfl.ch">karl.aberer@epfl.ch</a></td>
<td>SC 180</td>
<td></td>
</tr>
<tr>
<td>Dr. Yongxuan Zhou</td>
<td><a href="mailto:yongxuan.zhou@epfl.ch">yongxuan.zhou@epfl.ch</a></td>
<td>SC 146</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assistants</th>
<th>Email</th>
<th>Office</th>
<th>Office hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adriane Budure</td>
<td><a href="mailto:adriane.budure@epfl.ch">adriane.budure@epfl.ch</a></td>
<td>SC 130</td>
<td>Wed. 14-15h</td>
</tr>
<tr>
<td>Ali Salehi</td>
<td><a href="mailto:ali.salehi@epfl.ch">ali.salehi@epfl.ch</a></td>
<td>SC 143</td>
<td>Fri. 14-15h</td>
</tr>
<tr>
<td>Le-Hung Vu</td>
<td><a href="mailto:lehung.vu@epfl.ch">lehung.vu@epfl.ch</a></td>
<td>SC 142</td>
<td>Thurs. 14-15h</td>
</tr>
<tr>
<td>Wojciech Galuba</td>
<td><a href="mailto:wojciech.galuba@epfl.ch">wojciech.galuba@epfl.ch</a></td>
<td>SC 143</td>
<td>Fri. 14-15h</td>
</tr>
</tbody>
</table>
References

- Parts of the course are based on the following text books
  - Jiawei Han, Data Mining: concepts and techniques, Morgan Kaufman, 2000.

- Further references to the research literature will be given during the lecture