**Spring 2010 Presentation Schedule**

**CS591 Han Seminar: Advanced Topics on Data Mining (Spring 2010 Presentation Schedule)**

- Semester themes: (1) Information Network Analysis; and (2) data mining in cyberphysical systems
- Papers can be selected from this year or previous several year's conference proceedings or journals. We welcome students who would like to present tutorials and/or writing topic survey articles related to this theme.
- Two students per unit (20 minutes presentation and 5 minutes discussion for each research paper, i.e., two papers will be covered per class unit).

**Week 1 (1/21/10): Class organization and research plan presentation**

- Introduction
- Individual research plan presentation (one slide per person)
- Summary: Our research focus and tasks

**Week 2 (1/28/10): Two presentations**

- Ming Ji, *Graph-based Classification on Heterogeneous Information Networks*
- Sangkyum Kim - (KDD'09 Tutorial) How to do good research, get it published in SIGKDD and get it cited!

**Week 3 (2/04/10): Two presentations**

- Tianyi Wu, *Region-based Online Promotion Analysis* (EDBT'2010).

**Week 4 (2/9/10): Two presentations**

- Bo Zhao, *TextExplorer: Supporting Keyword-based Object Ranking and Exploration*

**Week 5 (2/18/10): Two presentations**

Professor Shobha Vasudevan [shobhav@crhc.illinois.edu] [slides]

GoldMine: Automatic Assertion Generation Using Data Mining and Static Analysis

Assertions are used for verification, validation and reliability tasks at different phases of the hardware design cycle.

In hardware development environments, assertion generation is an intensely manual process that is resource and time-intensive.

We present GoldMine, a methodology for generating assertions automatically. Our method involves a combination of data mining and static analysis of the Register Transfer Level (RTL) design. The RTL design is first simulated to generate data about the design's dynamic behavior. The generated data is then mined for "candidate assertions" that are likely to be invariants. These candidate assertions are then passed through a formal verification engine to filter out the spurious candidates. The assertions that are attested as true by the formal engine are system invariants.

These are then evaluated by a process of designer ranking that is provided as feedback to the data mining engine.

We present results of using GoldMine for assertion generation of the RTL of a 1000-core processor design that is still in an evolving stage. Our results show that GoldMine can generate complex, high coverage assertions in RTL, thereby minimizing human effort in this process.
Week 6 (2/25/10): Two presentations


Week 7 (3/04/10): Two presentations

Xiao Yu: Introduction to Large Scale Data Mining (slides)
Hongbo Deng: Modeling Bipartite Graph and its Application to Query Log Analysis based on Slides
Paper1: Entropy-biased Models for Query Representation on the Click Graph (Link)
Paper2: A Generalized Co-HITS Algorithm and Its Application to Bipartite Graphs (Link)

Week 8 (3/11/10): Two presentations

Yizhou Sun: ICDM’09 Tutorial “Data Mining using Matrix and Graphs” (slides1, slides2)
Xin Jin: WWW’10 Topic Initiator Detection on the World Wide Web (paper, slides)

Week 9 (3/18/10): Two presentations

Yintao Yu: Using Trees to Depict a Forest in VLDB 2009 paper slides
Bolin Ding: Differentially Private Range Count Queries: ICDE 2010 VLDB 2010

Week 10 (3/25/10): Springbreak (no class)

Week 11 (4/01/10): DASFAA conference (no class)

Week 12 (4/08/10): Two presentations

Jing Gao: Anomaly Detection: A Survey (Paper, Slides)
Peixiang Zhao: Graph Clustering Based on Structural/Attribute Similarities (VLDB 2009) (Paper Slides)

Week 13 (4/15/10): Two presentations

Zhijun Yin
LINKREC: A Unified Framework for Link Recommendation with User Attributes and Graph Structure (WWW 2010 poster) (Paper Slides)
Manish Gupta: Survey on social tagging techniques (Paper Slides)

Week 14 (4/22/10): Two presentations

Cindy Xide Lin: TwitterRank: Finding Topic-Sensitive Influential Twitterers (WSDM’10) (Paper, Slides)
Chi Wang: Social-Dimension Approach to Classification in Large-Scale Networks (SDM'09, KDD'09, CIKM'09) (KDD paper, KDD slides, CIKM slides, SocialCom Tutorial)

Week 15 (4/29/10): Two presentations

Hyungsul Kim: NDPMine: Efficiently Mining Discriminative Numerical Features for Pattern-Based Classification ppt
Week 16 (5/06/10): Data Mining Group Semester Summary