Spring 2009	**CSC 466:** Knowledge Discovery from Data	Alexander Dekhtyar

Design Project
Stage 1

Overview

The design project is to be performed by teams of four people each. Five teams will be formed. Team formation is left up to you. I recommend forming a team with an eye for completing both this project and the supplemental data analysis project, rather than having different teams on these projects.

The project will have the following timeline:

No.	Activity/Assignment	Dates/Due Date	Comment
1.	Initial discussion	*April 16, 2009*	brief discussion during the lab
2.	Team formation	*by April 21, 2009*	
3.	Problem description released	*April 21-23, 2009*	
4.	Stage 0 (preparation)	*April 23-28, 2009*	
5.	Discussion	*April 28, 2009*	*tentative*
6.	Stage 1 (design)	*April 30 - June 4, 2009*	
7.	Report due	*June 4, 2009*	
8.	Presentation	*June 10, 2009*	final exam slot

We may use a full lab or part of a lab session some time in May to conduct more discussions on the project.

The Problem: Design System for Supporting RFID-enhanced boarding passes

This section provides the overview of the problem for the project assignment.

Those of you who have tried to make a connection in an overcroweded airport know just how big a hassle it may be. There is a technological solution that, if implemented, can pave the way for significant reduction in airport hassle. The idea is to *embed an RFID chip into every boarding pass issued by airlines.*

Scenario. Each passenger upon flight registration receives boarding passes for each leg of their flight enhanced with RFID chips[1]. An RFID chip allows for the boarding pass to be tracked within a closed environemnt, e.g., an airport building (or collection of buildings).

Passengers continue with their routine, but because the location of their boarding passes is now available, it may be possible to improve certain experiences and services that the passengers receive during their trip.

Software Support. RFID chips embedded in boarding passes are tracked by hardware sensors spread around the airport(s): both stationary and, possibly, mobile (e.g., embedded in devices carried by airline or airport personnel). However, it falls onto *carefully designed software* to provide the appropriate functionality in support of this technological solution.

Your goal is to eventually *design the architecture for a software system* that

- Enables the use of RFID-enhanced boarding passes across multiple airlines, airports, etc...
- Provides a wide range of management and control facilities that take advantage of the features provided by RFID chips.
- Collects, stores, maintains and **analyses** information that arises in the course of running the software.

The software system may consist of multiple, possibly independent, or semi-independent sofware components, each of which has a specific role to fulfil. At the end, it is up to individual teams to determine what software components to design and how to combine them.

The role of KDD. The RFID enhancement of boarding passes will require software support that will, in turn, generate vast quantities of data. This data, which can be (potentially) combined with data from other, conventional, sources (e.g., airline passenger databases) may be analysed.

The goal of each team is to design the software support for the RFID enhancement **with an eye for information collection and analysis using KDD technologies**.

[1]Note, that this, for example, makes it impossible to print boarding passes from home... At least not in the way this is currently practiced.

The Assignment

Each team shall design of a software system supporting RFID-enhanced boarding passes in airpor environment. The system must, at the very least:

- Support communication between RFID-enhanced boarding passes and the airport sensors, i.e., support tracking of RFID-enhanced boarding passes.
- Facilitate a variety of services provided to the passengers, as well as to other parties.
- Collect information made available to the software, integrate it with any other information deemed necessary.
- Perform a number of analytical tasks on collected data.

The first and foremost goal of each team is to determine the specifics of each of the bullet points above. In particular:

Stakeholders. Each team shall start its work by determining the **stakeholders**: the parties with a stake in the process being supported by the software. The overall list of stakeholders will be outlined during the April 23 discussion.

If the overall list is sufficiently short (up to five different stakeholders identified), each team will design the software with all identified stakeholders in mind.

If the overall list is larger (especially if significantly more than five stakeholders are determined), then each team can select 4–5 stakeholders from the overall list and concentrate on software design for the selected group of stakeholders.

Within each group, at least at the intitial stage, each stakeholder's interests shall be represented by one student (if you select more than four stakeholders, one person may represent more than one stakeholder, but hopefully such stakeholders will be reasonably co-aligned).

All other decisions the teams make shall be negotiated from the positions of all stakeholders. (after all, you are designing vaporware).

Services. Once the stakeholders are established, each team must negotiate the services that the software system, or its individual components (in conjunction with the RFID hardware) will provide for the stakeholders.

Essentially, each team should consider two questions:

- What can we do with the facilities provided to us by the RFID-enhanced boarding passes?

- What should we do with these facilities?

and should make an effort to answer these questions.

The software system shall be then designed to provide support for the selected services.

Information. Each team shall determine:

- What types of information may be made available through the designed software system;
- What other types of information may be made available for analytical processing from other systems;
- How the information can be collected/stored/maintained/archived;
- What analytical questions do the stakeholders need/want answered;
- What types of intelligent data analysis/knowledge discovery/data mining can be performed with the available data in order to answer those questions.

The final design documents shall reflect each team's decisions with respect to each of the issues above.

Deliverables

Each team shall produce the following:

1. **An initial outline of the project.** A short document briefly outlining the direction in which the team plans to take the project. In particular, the document must specify

   - The stakeholders chosen for the project.
   - The interests/conflicts/desires of individual stakeholders.
   - The services to be provided by the software system.
   - The data you are planning to collect from the system.
   - The analytical tasks you are planning to perform.

   For the initial outline, only brief description of these items are necessary.

   The initial outline is **due Thursday, May 7**.

2. **Initial outline presentation.** Each team will describe its initial design ideas in a short presentation on **Thursday, May 7**. The presenations should be at most 10 mins long. Each team has to have some supporting visuals (e.g., Powerpoint slides, but other visual materials are allowed). It is not necessary for all four members to participate in the actual delivery, although it is not prohibited.

3. **Final design document.** The actual software system design, clarifying, expanding on and extending your initial outline. The exact specifications for the final design document will be made available to you around **May 7**. **The final design document is due last day of the classes.**

4. **Final presentation.** Each team will give a 15-20 min. presentation of their design in class during the finals week. The exam time is **Tuesday, June 10, 4pm.** We will use this time for presentations and followup discussion. The presentations are due **June 10**.