# Introduction to
# Spatial and Moving Objects Databases

(CIS 4930/6930)

*Department of Computer & Information Science & Engineering (CISE)*
*Database Systems Research & Development Center*

Course Syllabus[1] – Fall 2012

Instructor: Dr. Markus Schneider

*August 10, 2012*

## Table of Contents



[1] Note that this syllabus has a "dynamic" character. That is, it can (and will be) updated and extended at any time without prior notice. Hence, from time to time you should check the syllabus.

# 1 General Information

| | |
|---|---|
| Credits: | 3 |
| Prerequisites: | Database Systems (e.g., COP 5725) (recommended) |
| Instructor: | Dr. Markus Schneider |
| Lecture times: | Tuesdays: 8th & 9th period (3:00 pm - 3:50 pm, 4:05 pm - 4:55 pm),<br>Thursdays: 9th period (4:05 pm - 4:55 pm) |
| Location: | CSE E222 |
| Office hours: | Tuesdays: 2:00 pm to 2:45 pm,<br>Thursdays: 3:00 pm to 3:45 pm,<br>or by appointment |
| Office: | CSE Building, room E450 |
| Course web site: | http://www.cise.ufl.edu/~mschneid/<br>Teaching/CIS6930_Fall2012/CIS6930_Fall2012.html |
| Tel.: | (352) 505-1584 |
| Email: | *mschneid@cise.ufl.edu* |

# 2 Course Objectives

This is an undergraduate-level and graduate-level advanced course on the concepts and principles of spatial and moving objects databases.

## 2.1 Traditional Versus Advanced Database Management

Traditional database systems have mainly been designed for business, administrative, and financial applications. They deal exclusively with *standard*, i.e., alphanumeric, data which include numerical data like integers, floats, and dates as well as string data. *Relational*, *object-relational*, and *object-oriented* database systems have been developed as powerful tools for handling large volumes of this kind of data in an efficient manner.

But more and more, other kinds of data, called *non-standard* data, arise from *new, emerging applications* and new application areas. Examples are video data, multimedia data, image data, genomic data, medical data, temporal data, spatial data, and spatiotemporal data, to name only a few. Two essential features are that in contrast to alphanumeric data they have an internal, complex structure and that they can be of arbitrary, finite representation length. Further, domain-specific knowledge is necessary to be able to support non-standard database applications. Traditional database systems have great difficulties to cope with these kinds of data, because they have been tailored to fixed-length data of very simple internal structure.

## 2.2 Spatial and Moving Objects Database Systems

The focus of this course is on the underlying database technology to support spatial and spatio-temporal applications. Extending database technology to deal with spatial and moving objects means first, as for many other nonstandard database applications, to provide facilities in a data

model of a database management system (DBMS) to describe such entities, to extend the query language by constructs for analyzing them, and to formulate predicates about them. Second, it means that the implementation of a DBMS must be extended. The two major strategies for this are (i) to build a layer on top of an existing DBMS and so to map spatial object representations and spatial predicates to existing facilities of the DBMS, and (ii) to actually extend the DBMS by providing data structures for spatial objects, methods for evaluating operations, specialized indexes and join algorithms, and so forth. It turns out that only the second strategy leads to success.

From a theoretical and conceptual point of view, in this course, we examine the problems, principles, techniques, and mechanisms to store, retrieve, query, and manipulate spatial and spatiotemporal data in a database context. This will lead us to the state-of-the-art of *spatial database systems* for the management of *spatial* or *geometric objects* as well as *moving objects databases* for the management of *moving* or *spatiotemporal objects*. Spatial database systems form the data management foundation of so-called *Geographic Information Systems* (*GIS*) which incorporate computer-based system and analysis support for all kinds of geographic applications (for example, geography, hydrology, meteorology, mobile computing, whale-watching, forest fire management systems, to name only a few) in the broadest sense. Moving objects database systems have so far not been commercialized and are still in the development and exploration phase. Their later integration into GIS would improve their capabilities to support spatiotemporal analysis and applications.

*Spatial database systems*[2] are full-fledged database systems with additional capabilities to represent, store, retrieve, manipulate, and query *geometric* information like *points*, *lines*, *regions*, *circles*, *rectangles*, *volumes*, and *surfaces*. Estimates state that approximately 80% of all data have spatial/geometric features. This shows the importance and necessity of understanding the nature of space and how to deal with spatial data.

Consider the map of the 50 states of the USA. Each state has besides its thematic attributes like name and population also a geometry which is described by its region. Such a region is an example of a spatial or geometric object. It has an internal complexity since it can have holes (like enclaves) and consist of several components (like mainland and islands). Cities can, e.g., be represented as points, i.e., we are here interested in their location and not so much in their extent. We can then pose very interesting queries. Let us assume the two relations (tables)

states(sname: *string*, spop: *integer*, territory: *region*)
cities(cname: *string*, cpop: *integer*, loc: *point*)

Here, *point* and *region* are so-called *spatial data types*. They are in the same way attribute data types as the types *string* and *integer*. A query could ask for all pairs of cities and states where a city is located in a state. This can then be formulated as a so-called *spatial join*:

**select** cname, sname
**from** cities, states
**where** loc *inside* territory

---

[2]Spatial databases must be seen in contrast to *image databases* which deal with the management and manipulation of *images* or *raster bitmaps*. Image databases will not play a big role in this lecture.

The term *inside* is a so-called *topological predicate* testing whether a point is located inside a region. A topological predicate gives a boolean statement about the relative position of a spatial object to another spatial object. Another query orders the states by the size of their geometric extent:

**select** sname
**from** states
**order by** *area*(territory)

The operator *area* is a geometric function taking a *region* object as an argument and yielding a numerical value, namely the area of this object, as a result.

*Moving objects database systems* are full-fledged spatial database systems with additional capabilities to represent, store, retrieve, manipulate, and query the *continuous temporal evolution* (that is, *movement* (change of location), *change of shape*, *change of extent* (shrinking, growing)) of spatial objects.

Consider the following two relations (tables) *flight* and *weather*.

flight(id: *string*, from: *string*, to: *string*, route: *mpoint*)
weather(id: *string*, kind: *string*, area: *mregion*)

The table schemas show the two *spatiotemporal data types mpoint* and *mregion* which represent moving points and moving regions respectively. A moving point describes the continuous location change of a point object (here: a plane). Its extent and shape and their possible change are *not* of importance here. This is different for a moving region like a weather zone. Here, beside the location change, especially the change of shape and the change of extent are important. We consider the query when flight BA488 was within the snowstorm S16. This query could be formulated as

**select** *deftime*(*intersection*(f.route, w.area))
**from** flight **as** f, weather **as** w
**where** f.id = 'BA488' **and** w.id = 'S16'

The operation *intersection* computes the part of a moving point whenever it lies inside a moving region. The result is a moving point again. The operation *deftime* returns the set of time intervals when a moving point is defined. The result will be stored in a value of a newly introduced type *periods*.

Another query asks for any pair of airplanes which during their flight came closer to each other than 500 meters. This query can be written as

**select** f.id, g.id
**from** flight **as** f, flight **as** g
**where** f.id <> g.id **and** *min*(*distance*(f.route, g.route)) < 0.5

The operation *distance* computes the distance between two moving points at any time of their common lifetime. The result is stored in a newly introduced temporal data type *mreal* for *moving real* numbers. The function *min* determines the smallest real number kept in a moving real.

The application fields for spatial and moving objects databases are numerous and ubiquitous: all kinds of geographic applications (geography, hydrology, meteorology, hurricane research,

transportation, etc.), mobile computing, whale-watching, forest fire management systems, homeland security applications, bioterrorism, to name only a few. A very large market are *Geographical Information Systems* (*GIS*) which incorporate many of these applications and which urgently need computer scientists as database experts for spatial and spatiotemporal data management, analysis, and data mining. GIS technology incorporates and leverages spatial databases and provides various spatial analysis tools on top of these databases. Moving objects databases are currently not integrated into GIS.

From a practical point of view, we consider how spatial and moving objects database systems are implemented. For example, they provide fundamental spatial and spatiotemporal data types together with a large collection of appropriate operations and predicates. These type systems are then integrated into or coupled with the an extensible database system like Oracle, Informix, or PostgreSQL.

By providing a balanced view of "theory" and "practice", the material covered should allow the student to get a deeper insight into spatial and moving objects databases and to understand the concepts of accommodating commercially and publicly available database systems to these advanced needs.

Due to the many concepts that have to be learnt to obtain sufficient knowledge about these databases and due to the fact that no TA will be assigned to this class, we will not perform own implementation work in class.

### 2.3 General Course Objectives

This course provides students with advanced knowledge in computer and information science. Specifically, the course offers students the opportunity to study and critically examine the outcome of state-of-the-art research projects in the area of spatial and spatiotemporal data management. As part of this process, students will also obtain general skills like how to find and review research literature, present their findings in well-prepared PowerPoint presentations, write down their findings in an essay, and contribute to and lead technical discussions.

## 3 Textbooks and Readings

A number of textbooks are *recommended* but not required since we will study original research literature for the different topics. Two main textbooks and one overview article are currently available on spatial databases:

- M. de Berg, O. Cheong, M. van Krefeld, and M. Overmars. Computational Geometry: Algorithms and Applications. Springer-Verlag, 3rd Edition, 2010.
- R.H. Güting and M. Schneider. Moving Objects Databases. Morgan Kaufmann Publisher, 2005.
- R.H. Güting. An Introduction to Spatial Databases. *VLDB Journal (Special Issue on Spatial Databases),* 3(4):357-399, 1994.
- Franco P. Preparata and Michael I. Shamos. *Computational Geometry : An Introduction*. Springer-Verlag, 1993.

- P. Rigaux, M. Scholl, and A. Voisard. Spatial Databases - with Applications to GIS. Morgan Kaufmann Publisher, 2002.
- Shashi Shekar and Sanjay Chawla. *Spatial Databases: A Tour*. Prentice Hall, 2003.

In addition, the instructor will provide a number of *research publications* describing various aspects of spatial and moving objects databases. Each student's deliverables will require additional, specialized readings.

Students wishing to consult a general text on database management may use the following *optional* readings:

- A. Silberschatz, H.F. Korth, and S. Sudarshan. *Database System Concepts*. Fifth Edition, McGraw-Hill, 2006.
- R. Elmasri and S. B. Navathe. *Fundamentals of Database Systems*. Addison-Wesley, 2004.

# 4 Course Assessment

Student assessment will be based on

- an *in-class presentation on a research topic* and
- a *report on a research topic*.

There will be *no* exams, *no* homeworks, *no* projects, and *no* implementation work. All deliverables have to be provided on an *individual* basis, i.e., teamwork is *not* allowed unless announced differently.

## 4.1 In-Class Presentation

Each student is required to give an overview of an assigned research topic in a slide presentation to the class and to lead the ensuing discussion in class. Research topics are defined by the instructor. For each research topic, an important research paper is provided to the student by the instructor. The student has to extend this list of publications through an extensive literature study and to submit this list to the instructor. Based on the quality and the acceptance of the submitted publication list by the instructor, the student must study the papers, structure them, compare them, and design a slide presentation yielding the main results of the study.

Students will be evaluated based on the quality of the submitted list of publications, the organization of their slide presentation, the clarity and comprehensibility of their talk as well as on the knowledge and depth of the presented material (as demonstrated during the presentation as well as during the discussion in class).

The slide presentation will contribute 50% to your grade. In detail, the following list of criteria will be used to evaluate your slide presentation:

- Quality of the submitted list of publications (15%)
  - Have appropriate publications been selected?
  - Is the list of publications "rather complete"?
- Organization of the slides (35%)

  - Was the presentation well structured?
  - Did the presentation have a clearly defined goal/focus/message?
  - Did the presentation give the essential facts and results?
  - Were there any important aspects of the topic that were missing or only partially covered?
- Clarity and comprehensibility of the talk (15%)
  - Was the presentation easy to follow?
  - Did the presenter express his/her thoughts with a clear, loud, and expressive voice?
  - Did the presentation include a demo (this is optional)?
- Knowledge and depth of the presented material (35%)
  - Was the material covered at a depth that is adequate for the class taking also into account the preparation time and background of the presenter?
  - Was the presenter able to answer questions from the audience?
  - Was the presenter successful in involving the audience in a discussion (e.g., by preparing questions)?
  - Was the presenter knowledgeable about related literature?

The slide presentation has to be submitted to the instructor before the date of presentation at an individually announced date.

### 4.2 Research Report

Each student is also required to write and complete a semester-long report with respect to an assigned research topic. The research topic can correspond to the research topic in Section 4.1 so that the student benefits from previous work. Research topics are defined by the instructor. For each research topic, an important research paper is provided to the student by the instructor. The student has to extend this list of publications through a literature study and to submit this list to the instructor. Based on the quality and the acceptance of the submitted publication list by the instructor, the student must study the papers, structure them, compare them, and formulate the findings as a Related Work section in the report.

The research report has the following fixed structure:

Title page
Table of Contents
1 Introduction
2 Related Work
3 <Conceptual section about assigned research area (select own section heading)>
4 Conclusions
References

Additional and appropriate subheadings and subsubheadings have to be inserted by the student. The instructor expects a report of *sufficient* length (20 pages $\leq$ length of report $\leq$ 30 pages). The report should give an answer to at least the following aspects:

- Give a clear description of the research topic and the problems that you have considered and investigated. Why is it a worthwhile research topic? Why should we care about these problems? What is difficult/challenging/interesting about them?

- Give a brief summary of the work/technologies most related to your project.
- Give a description of the general solution approach in the literature if any exists (conceptual work).
- Describe the outcome of your investigation including any contributions you have made (conceptual work).

The research report must be provided electronically as a PDF file to *mschneid@cise.ufl.edu*; the filename should be of the form <author>.pdf. Formats other than PDF will not be accepted.

Students will be evaluated based on the quality of the submitted list of publications, the structure of their report, the quality of the contents, knowledge and depth of their report, the clarity and comprehensibility of their writing and language, and the spelling and grammatical correctness of their writing.

The research report will contribute 50% to your grade. In detail, the following list of criteria will be used to evaluate your research report:

- Quality of the submitted list of publications (10%)
  - Have appropriate publications been selected?
  - Is the list of publications "rather complete"?
- Structure of the report (10%)
  - Does the structure of the report make sense?
  - Is the structure of the report logically consistent?
  - Does the structure of the report clearly reflect the intended goal/focus?
- Quality of the contents, knowledge and depth of the report (60%)
  - Is the contents of the report of high quality? Is it sound and complete?
  - Has related work been studied, compared, and described in an appropriate way?
  - Has the student's own conceptual work been sufficient and has it been adequately described?
  - Has the student's own implementation work been sufficient and has it been adequately described?
  - Is the material covered at an appropriate depth?
- Clarity and comprehensibility of writing and language (10%)
  - Is the report easy to read and easy to understand?
  - Is the report written in complete English sentences? Bulleted lists are not enough.
  - Is the report clear and precise in explanations and statements as it should be in technical documents? Does it avoid colloquialisms?
- Spelling and grammatical correctness of writing (10%)

The deadline for the research report, which has to be submitted to the instructor, will be announced by him. No late submissions will be accepted.

### 4.3 Class Participation

Class participation is mandatory and will be checked each time. Absence from class without a *prior* and *justifiable* excuse will lead to a penalty of 1% from the total student's performance for

each missing lesson. For instance, if the total performance of a student is 0.9 and the student missed five lessons, the total performance is reduced to 0.85. A lesson is a period of 50 minutes. The instructor would also like to point out that he highly advocates punctuality for the lectures.

## 4.4 Grading

Your grade will be based to 50% on the in-class slide presentation and to 50% on the research report. From that, the penalty for absence from class without excuse is subtracted. Hence, the student's performance $p$ will be calculated according to the following formula:

$$p = sp \cdot 0.5 + rr \cdot 0.5 - pe$$

where the variables have the following meaning:

- *sp* performance of the slide presentation in percent
- *rr* performance of the research report in percent
- *pe* penalty for absence from class without excuse in percent

Based on the student's performance, the following grading policy will be applied:

| **Student's Performance in %** | **Grade** |
|---|---|
| >94-100 | A |
| >88-94 | A- |
| >82-88 | B+ |
| >76-82 | B |
| >70-76 | B- |
| >64-70 | C+ |
| >58-64 | C |
| >52-58 | C- |
| >46-52 | D+ |
| >40-46 | D |
| >34-40 | D- |
| 0-34 | E |

# 5 Other Important Issues

## 5.1 Late Policy

No late report and project submissions and no re-scheduling of presentations will be accepted.

## 5.2 Academic Honesty

Students are required to respect the ethical standards for academic honesty established by the Office for Student Judicial Affairs. The University of Florida student body voted in Fall 1995 to approve the following *Student Honor Code*:

> *We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity.*

Work submitted must be produced individually by each student, except for tasks explicitly assigned to a group by the instructor. All work submitted individually in the form of exams, homework, presentations, reports, software projects, etc., is subject to the following implicitly or explicitly **required** pledge:

> *On my honor, I have neither given nor received unauthorized aid in doing this assignment.*

Violations of academic honesty and integrity in this course will not be tolerated. The instructor will deal strictly with any violations. Since ethical behavior in science and engineering is equal in importance to specific knowledge, the instructor will assign a non-passing letter grade to students who violate academic honesty standards, regardless of the violator's grade performance in exams, homework, quizzes, and other assignments. Official sanctions issued by the Office of Student Judicial Affairs **will become permanently noted in the student's official transcript**.

The instructor's advice to the student is: immerse yourself in the class, learn the material, do your tasks (homework, presentation, report, implementation, etc.) yourself. The benefit and enjoyment you will receive as a result of hard work will be much more valuable than any penalty you might receive as a result of cheating.

## 5.3 Class Etiquette

The instructor would like to make this course, which will have more the character of a seminar, as stimulating and at the same time as relaxed as possible. However, a few ground rules apply. Students are expected to arrive in class a few minutes before the scheduled period so that we can start on time. Students are permitted to bring food and drinks (unless expressly prohibited by the building code) but please turn off all cell phones and pagers before you arrive so that we don't get disturbed during class.

## 5.4 Students with Disabilities

Students requesting classroom accommodation must first register with the Dean of Students Office. The Dean of Students Office will provide documentation to the student who must then provide this documentation to the Instructor when requesting accommodation.

### 5.5 Where to Get Class-Related Information

The latest class-related information is given at the beginning of each class when announcements are made. Some information related to this class will be provided on the course web site. This especially relates to the slides of the different lectures. The home page for the course is located at

http://www.cise.ufl.edu/~mschneid/Teaching/CIS6930_Fall2012/CIS6930_Fall2012.html

### 5.6 Final Advice

The students who get the most out of this class will be the ones who put in the most effort. If you want to do well, come to all the lectures, read the assigned research papers and sections of the books before coming to class, and start early on your different tasks. If you are having difficulty, you owe it to yourself to get help. The instructor sincerely wants all of you to do well. If you work hard and master the material presented in this class, you will learn some powerful, fundamental concepts of spatial and spatiotemporal database management systems, which are very useful in today's high-tech industry like geographical information systems (GIS) and mobile computing. The instructor will try his best to make the course as interesting and stimulating as possible and an enriching experience for you.

*Markus Schneider*