CSIT6000P Spatial and Multimedia Databases

2022 Spring

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Course Introduction

Prof Xiaofang Zhou

The Teaching Team

- Course Coordinator & Lecturer
  - Prof Xiaofang Zhou
  - Email: zxf@ust.hk
  - Room: 3533
  - Phone: 2538 8340
  - Consultation: by appointment
- Teaching Assistant
  - Ms Yehong XU
  - Email: yxudi@connect.ust.hk

+ Assumed Background

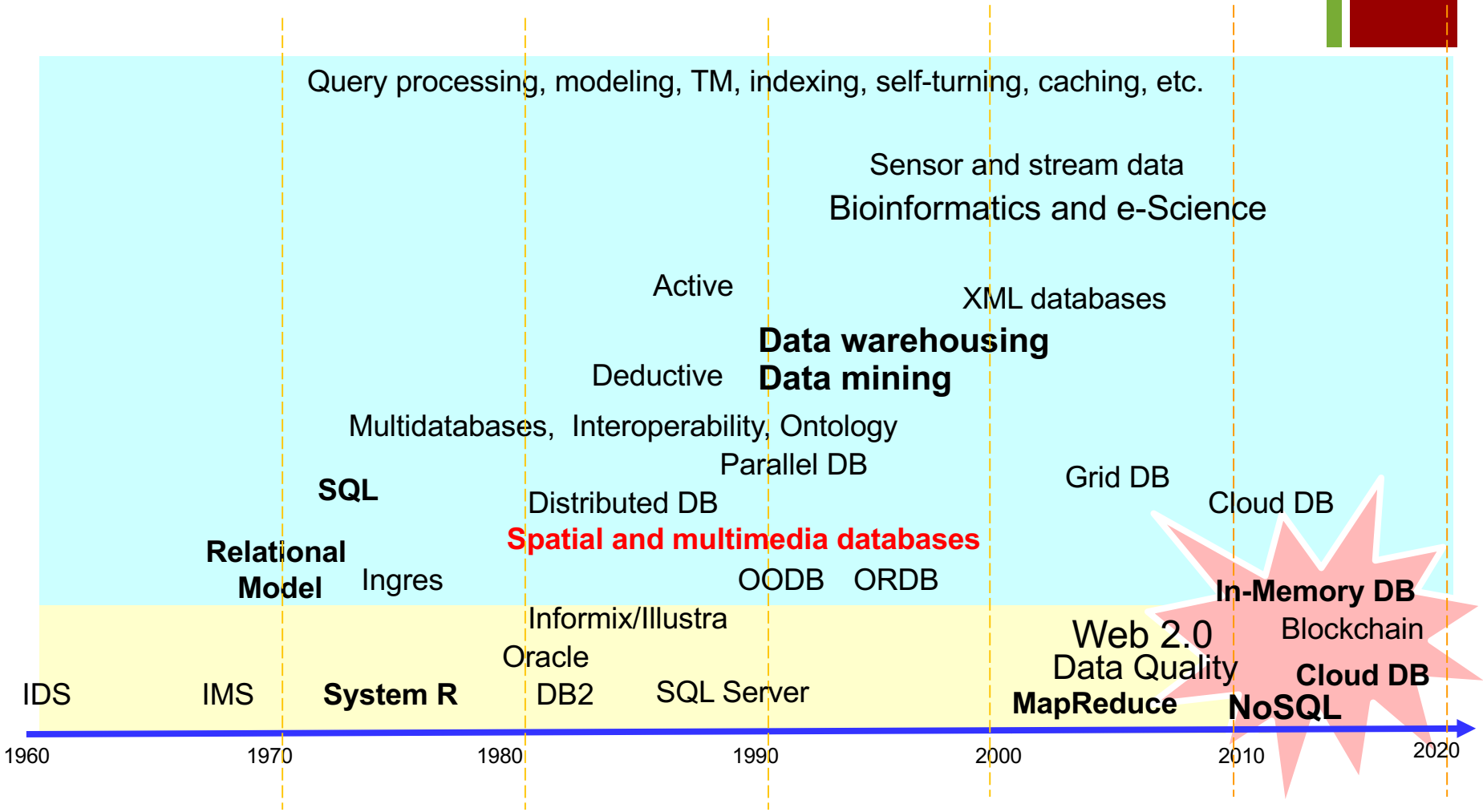
- Database knowledge
  - Database management system concepts
  - Relational database design and SQL
  - Database indexing
  - Query processing
- Programming skills
  - Java, C/C++, Python

We will review RDBMS first in this course

+ Relational DBMS

- What is RDBMS?
  - The relational model and database design
  - SQL
  - Database indexing
  - Query processing
  - ACID/Transaction Management/Recovery
  - Access control
- RDBMS products?
  - Oracle, Microsoft SQL Server, IBM DB2, MySQL, Postgres…

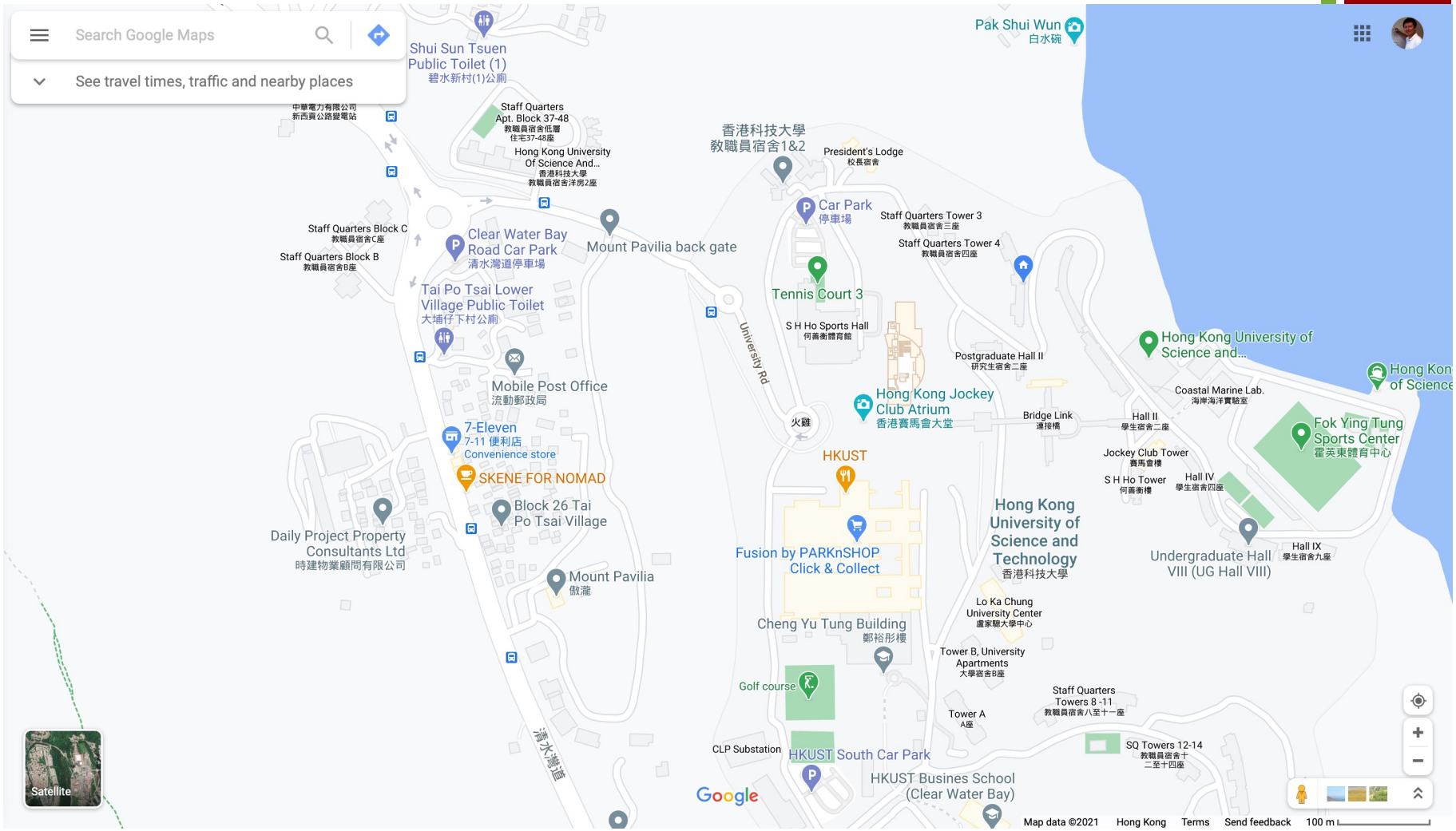
+ A Brief History of DBMS



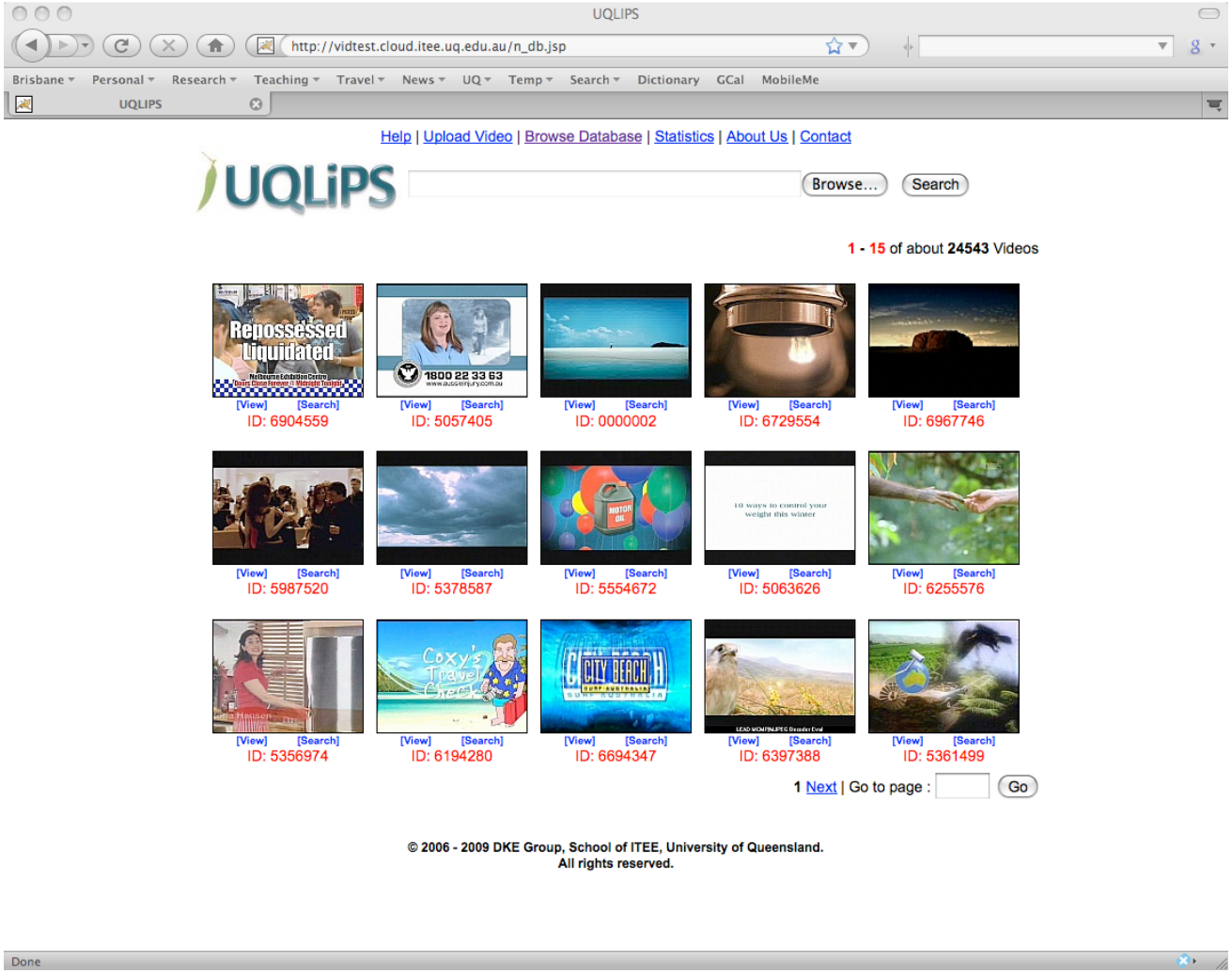
+ Beyond RDBMS

- History of RDBMS
  - The 60s:
  - The 70s:
  - The 80s:
  - The 90s:
  - New millennium:
  - Now and future:
- Beyond RDBMS
  - Spatial and temporal attributes are ubiquitous
    - But they are poorly support by RDBMS…
  - One fundamental change:
    - Number of dimensions increases from 1, to 2, to many…

+ One Example


Search Google Maps
See travel times, traffic and nearby places
Shui Sun Tsuen Public Toilet (1) 碧水新村(1)公廁
Staff Quarters Apt. Block 37-48 教職員宿舍低層 住宅37-48座
Hong Kong University Of Science And... 香港科技大學 教職員宿舍洋房2座
香港科技大學 教職員宿舍1&2
President's Lodge 校長宿舍
Pak Shui Wun 白水碗
Car Park 停車場
Staff Quarters Tower 3 教職員宿舍三座
Staff Quarters Tower 4 教職員宿舍四座
Staff Quarters Block C 教職員宿舍C座
Staff Quarters Block B 教職員宿舍B座
Clear Water Bay Road Car Park 清水灣道停車場
Mount Pavilia back gate
Tennis Court 3
S H Ho Sports Hall 何善衡體育館
Tai Po Tsai Lower Village Public Toilet 大埔仔下村公廁
University Rd
Hong Kong University of Science and...
Hong Kon of Science
Postgraduate Hall II 研究生宿舍二座
Coastal Marine Lab. 海岸海洋實驗室
Mobile Post Office 流動郵政局
Hong Kong Jockey Club Atrium 香港賽馬會大堂
Bridge Link 連接橋
Hall II 學生宿舍二座
Fok Ying Tung Sports Center 霍英東體育中心
7-Eleven 7-11 便利店 Convenience store
火雞
HKUST
Jockey Club Tower 賽馬會樓
S H Ho Tower 何善衡樓
Hall IV 學生宿舍四座
SKENE FOR NOMAD
Block 26 Tai Po Tsai Village
Daily Project Property Consultants Ltd 時建物業顧問有限公司
Fusion by PARKnSHOP Click & Collect
Hong Kong University of Science and Technology 香港科技大學
Undergraduate Hall VIII (UG Hall VIII)
Hall IX 學生宿舍九座
Mount Pavilia 傲瀧
Lo Ka Chung University Center 盧家聰大學中心
Cheng Yu Tung Building 鄭裕彤樓
Tower B, University Apartments 大學宿舍B座
Golf course
Staff Quarters Towers 8 -11 教職員宿舍八至十一座
Tower A A座
CLP Substation
HKUST South Car Park
SQ Towers 12-14 教職員宿舍十二至十四座
HKUST Busines School (Clear Water Bay)
Satellite
Google
Map data ©2021 Hong Kong Terms Send feedback 100 m

+ Another Example



+ Modern Database Applications

- Very large datasets
  - Many with streaming nature too
- Complex data structures
  - Often represented as high dimensional data
  - Many with spatial and temporal attributes
  - Search is typically based on similarity, rathe than equality
  - One application often needs to integrate heterogeneous data from many sources
- Data analytics queries
  - Databases, data warehouses, data mining, machine learning…

Descriptive, predictive and prescriptive analytics

The Three Vs of Big Data

- Volume, Velocity and Variety
  - YouTube:
    - Serves around 2 billion videos per day
  - GIS/Geospatial:
    - Sales of software and data grow dramatically
  - Location-based Services (LBS):
    - Global market about US$30 billion
  - Sensor Data
    - (Radio Frequency ID) RFID, On-Board Diagnostics (OBD)…

+ In This Course…

- **What** will we learn?
  - The latest in large-scale data management and processing: **complex data types** and **similarity-based queries**
  - How to **represent** and **manage** complex data to enable efficient **processing** of advanced queries
  - Applicable to a wide range of **applications**, including spatial and multimedia data
  - Challenges and solutions for high dimensional data
- **How** will we learn?
  - Attending lectures
  - Studying selected research papers
  - Doing individual assignments
  - Writing reports and making presentations

+ Multidimensional Data

- An object is ***k*-dimensional** means that the object is described by *k* attributes.
- Examples:
  - point(x, y) - 2D: geographic data (GIS, LIS…)
  - point(x, y, z) - 3D: the universe, brain, molecule structure…
  - point(x, y, t) – 3D: spatiotemporal
  - person(age, weight, height) – 3D
  - color(c1, c2, … c128) – 128D
  - image(texture, shape, colour) – high dimension!
  - video – a sequence of image frames

+ Why Spatial Databases?

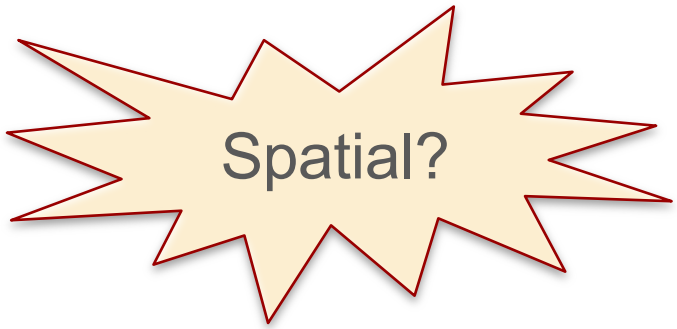
- One of the most widely used multidimensional databases
  - Foundation of understanding other high-D data
- Many important application domains have **spatial data** and **queries**. Some Examples:
  - **Insurance Risk Manager:** Which homes are most likely to be affected in the next great flood on the Brisbane river?
  - **Molecular Biologist:** Is the topology of the amino acid biosynthesis gene in the genome found in any other sequence feature map in the database?
  - **Medical Doctor:** Based on this patient's MRI, have we treated somebody with a similar condition?

How About Relational DBMS?

- Highly successful
  - Data Independence
    - The relational model
    - Non-procedural query
  - Efficiency & Performance
    - Indexing and query optimization
    - Allows concurrent access to data
  - Reliability
    - Persistence across failures
  - ACID?

+ Spatial Queries

- Queries
  - Which courses are meeting in Jockey Club Tower this semester?
  - Where is Jockey Club Tower?
  - Find all ocean-front buildings in HKUST.
  - Find all teaching theatres which are 150 meters from a restaurant.



What's Special about Spatial Queries?

- Retrieval & update of spatial data is based on the **spatial location** of a data object
  - *(vs. alphanumeric in RDBMS)*
- Fast execution of **geometric operations**
  - *(vs. simple comparison in RDBMS)*
- RDBMS obvious limitations
  - Limited data types – no support for multidimensional data!
  - Limited query types

Spatial DBMS (SDBMS)

- A SDBMS is a software module that
  - can work with an underlying DBMS
  - supports spatial data models, spatial abstract data types (ADTs)
  - supports a spatial query language
  - supports spatial indexing, efficient algorithms for processing spatial operations, and domain specific query optimization

- Examples:
  - Oracle Spatial and Graph in 12c (formally Oracle Spatial Extension), now separately licensed
  - Microsoft SQL Server supports spatial types since 2008
  - PostgreSQL DBMS uses the spatial extension PostGIS

From Spatial to Multimedia

- Very different on the surface, but many similarities fundamentally
  - Data represented as **multidimensional vectors**
- Keywords-based vs content-based search
  - **Keywords-based:** using text annotations
  - **Content-based:** using automatically extracted features such as colors, textures and shapes
  - Both have advantages and disadvantages
- Applications of content-based multimedia search

+ Some Projects I Participated

- QUEST (1998): delivering map-based Govt services
- UQLIPS (2006): near duplicate video search
- SharkDB (2014): an in-memory spatiotemporal database
- Vehicle trajectory mining
- Route planning (including real-time information)
- Telecom token analysis
- Bluetooth data at road intersections
- Social media recommendation for next POI
- Processing and annotation for biomedical images
- Entity linking and location privacy protection…

What to be Covered

1. Review of RDBMS
2. Spatial databases introduction
3. Spatial indexing mechanisms
4. Spatial algorithms and query processing
5. Spatiotemporal data management
6. High-dimensional indexing, search and dimensionality reduction
7. Multimedia databases
8. Other high dimensional applications

+ Learning Modules

Week	Date	Lecture	Assessment
1	5/2	Course Introduction; Review of RDBMS	
2	12/2	Indexing & Query Optimization in RDBMS	Group project out
3	19/2	Introduction to Spatial Databases	
4	26/2	Spatial Data Indexing (1)	
5	5/3	Spatial Data Indexing (2)	
6	12/3	Spatial Query Processing (1)	
7	19/3	Spatial Query Processing (2)	
8	26/3	Spatiotemporal Data Management	Assignment 1 (30%)
9	2/4	High Dimensional Indexing and Search	
10	9/4	Dimensionality Reduction and Approximate Search	
11	23/4	Multimedia Databases	
12	30/4	Trends; Course Review	Assignment 2 (30%)
13	7/5	Student presentations	Group project (40%)

+ Lectures

- 3-Hour lecture session per week
  - Real-time online initially (via Zoom through Canvas)
  - Lecture room: Rm 2303, Lift 17-18 (if in mix-mode teaching)
  - Time: 10:00-12:50 pm on Saturdays
- Lecture slides will be available online as PDF files
  - Likely to be updated before and after each lecture
- Recommended readings
  - No textbook
  - Selected papers will be available online when possible

+ Assessment

- Two assessment components
  - Two assignments (individual): 30% + 30%
    - Marking criteria can be found in the assignment specs
  - Final report and presentation (group): 40%
    - Marking criteria can be found in the project spec
- No exams
- Your final score for this course is the sum of the above components

+ Why Study This Course?

- For an CS graduates from HKUST
  - What you have learnt about relational DB may be not applicable to some applications, but…
- For prospect researchers
  - Spatial and multimedia databases are still a major research area with many open problems, and the knowledge we will learn in this course are applicable to many other types of high dimensional data
- For job seekers
  - Complex data are typically dimensional
  - High-dimensional data management and analytics evolve rapidly from a specialised area to a “commodity” skill with great demand

+ Some Questions…

- Let's try to identify some research questions and applications for spatial and multimedia databases…

+ Your Exercise…

- Take one application area that you are familiar with, or you are interested in
  - What types of spatial or spatiotemporal data in the application?
  - How such data have been and could be used, for what applications?
  - If RDBMS is the technology to use to manage such data, do you see any problems?
- Consider another high-D application area, and repeat the above
- Who are the main technology providers and users for spatial databases? Name a few products