Advanced Indexing Techniques & Flexible Query Answering in DBs

Assoc. Prof. Dr. Dang Tran Khanh
Department of Information Systems
Faculty of Computer Science and Engineering
khanh@cse.hcmut.edu.vn

Outline

- Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees
- Multidimensional Access Methods (MAMs)
- Indexing & FQAS: An Introduction
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
- Reading Suggestion
  - [1] chapters 17, 18, 27
  - V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231
  - Internet

Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees

- [1]: Chapters 17, 18
- Note: double check in your coursework/as homework

Differences between B-tree and B+-tree

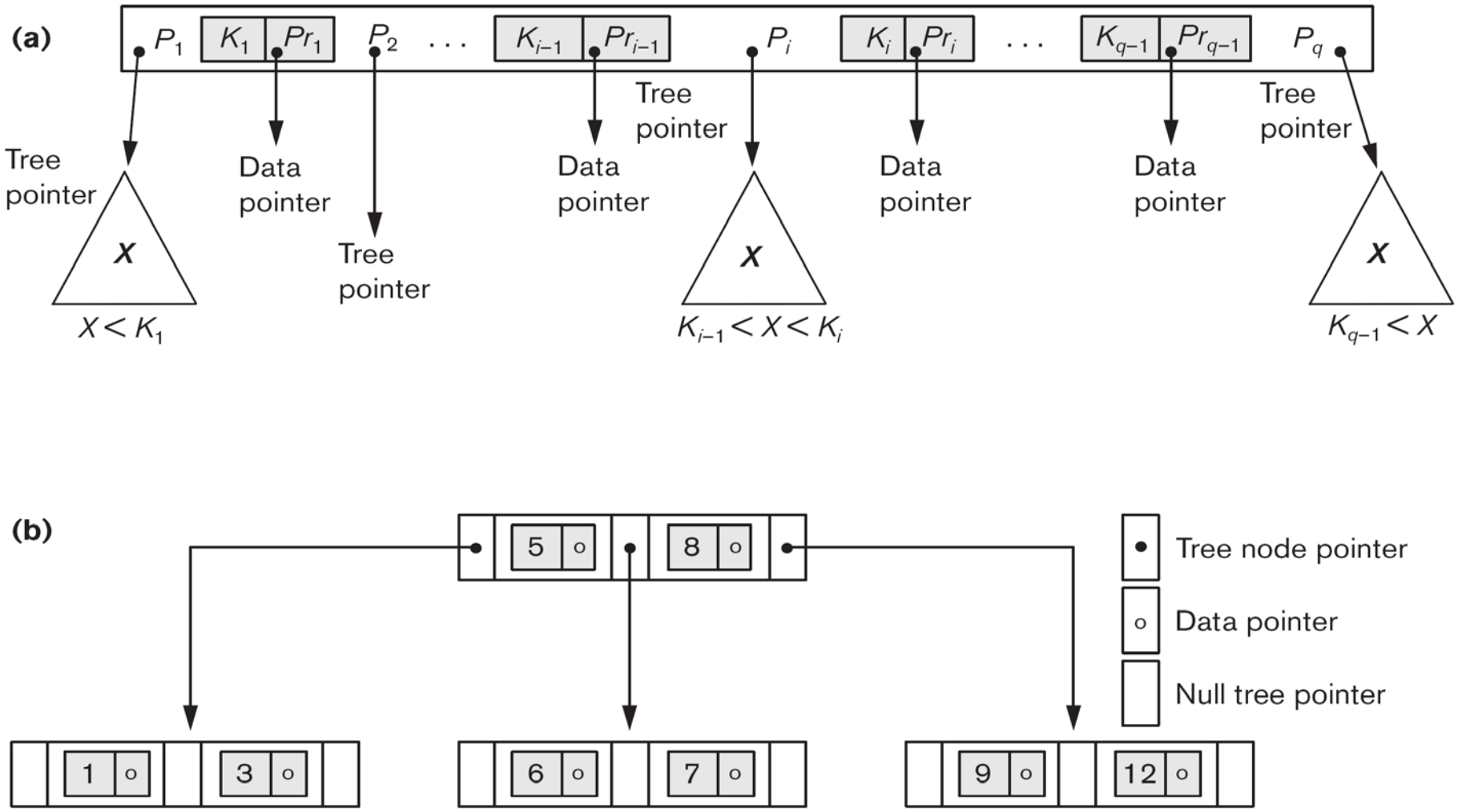
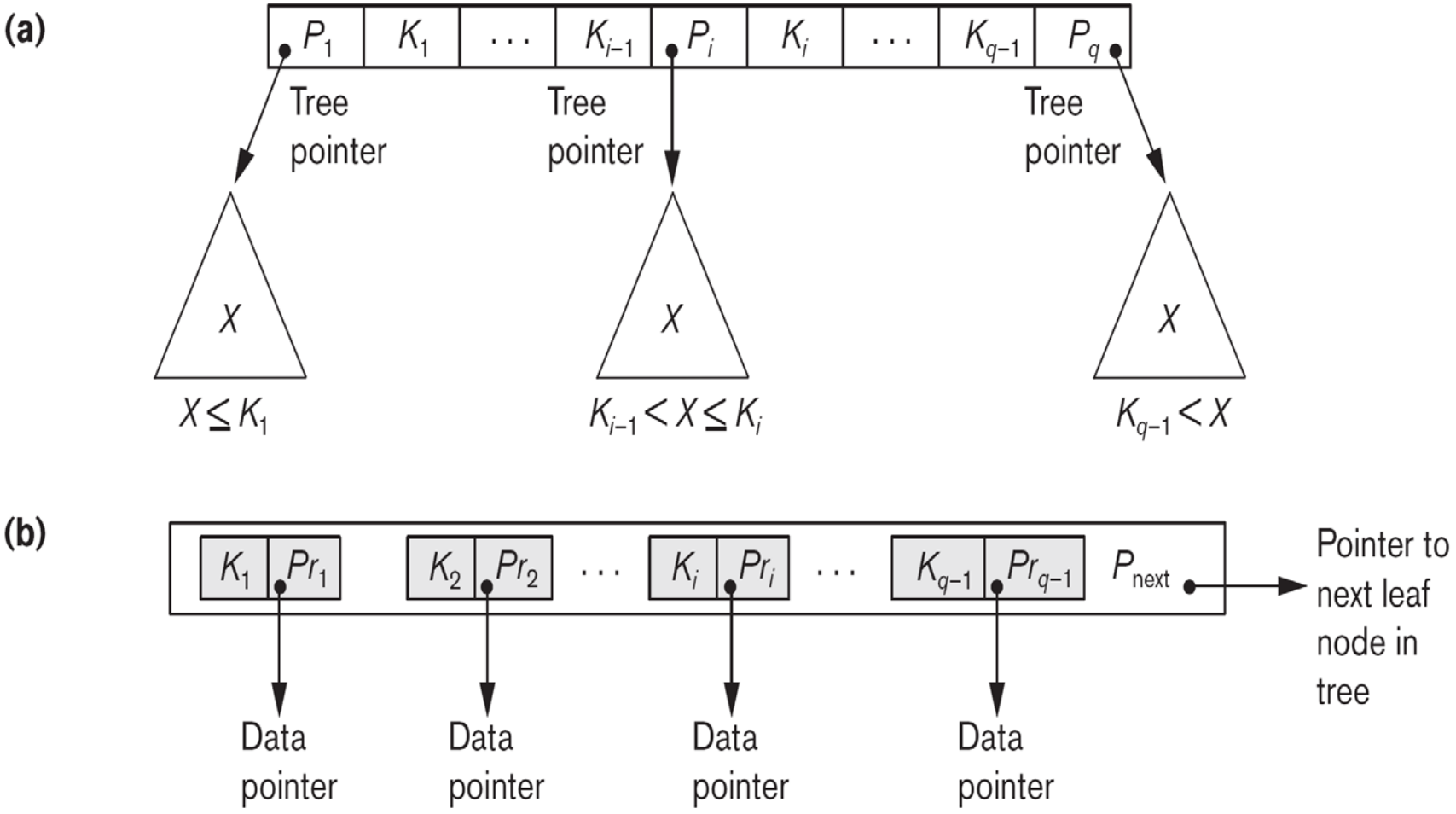


Figure 18.10
B-tree structures. (a) A node in a B-tree with $q - 1$ search values. (b) A B-tree of order $p = 3$.The values were inserted in the order 8, 5, 1, 7, 3, 12, 9, 6.

Differences between B-tree and B+-tree

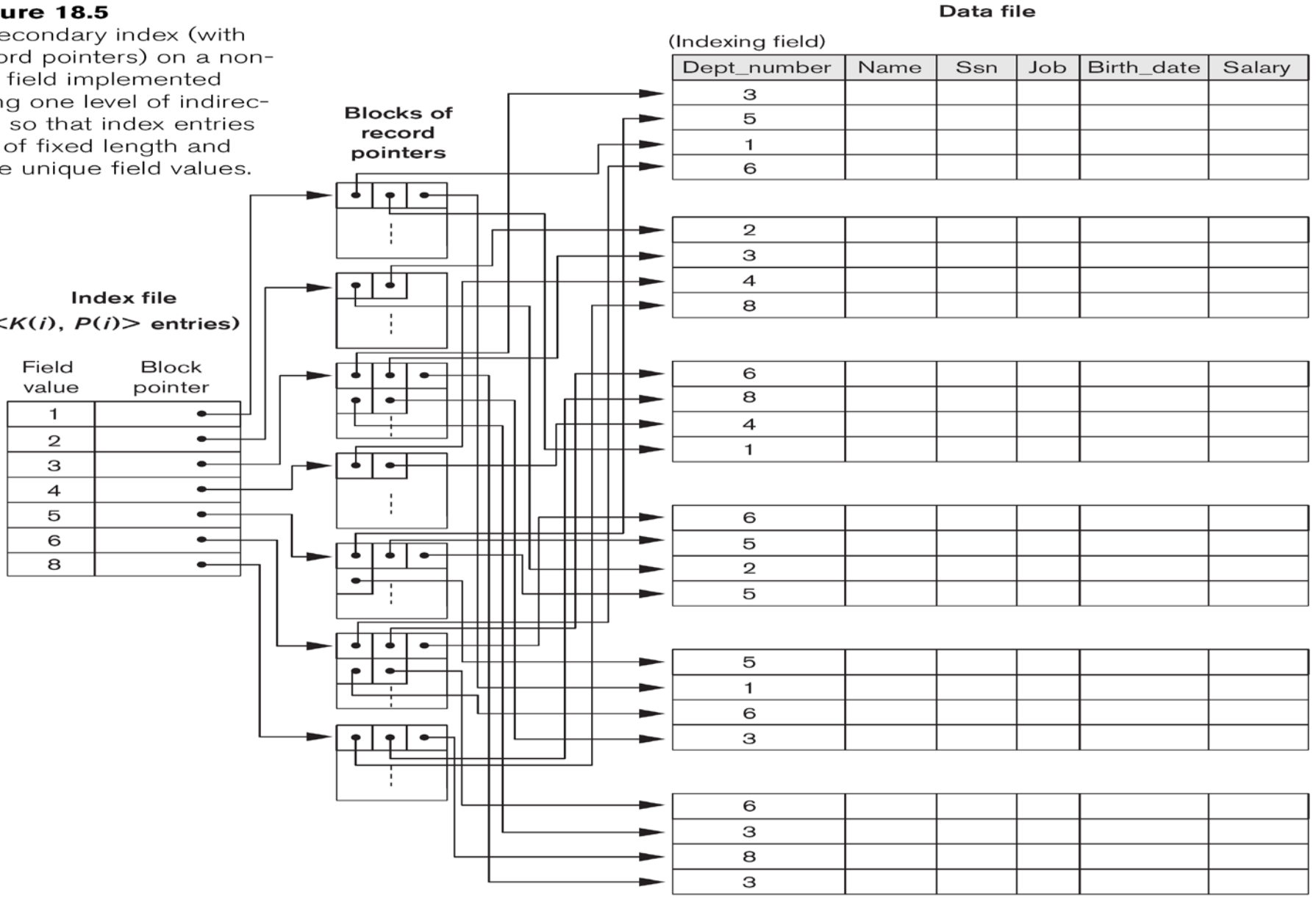
Figure 18.11

The nodes of a B^+-tree. (a) Internal node of a B^+-tree with $q - 1$ search values. (b) Leaf node of a B^+-tree with $q - 1$ search values and $q - 1$ data pointers.



Differences between B-tree and B+-tree: an option for B+-tree leaf node implementation (cf. chapter 18)

Figure 18.5
A secondary index (with record pointers) on a non-key field implemented using one level of indirection so that index entries are of fixed length and have unique field values.



Notes for B-/B$^+$-trees and indexing structure evaluation

- For all operations (search/select, delete, modify/update):
  - CPU-cost
  - I/O cost
  - Memory cost
  - Datasets: uniform or skew distribution?, etc.
- “Representative” cost(s) can be used (e.g., number of data objects obtained)

Outline

- Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees
- Multidimensional Access Methods (MAMs)
- Indexing & FQAS: An Introduction
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
- Reading Suggestion
  - [1] chapters 17, 18
  - V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231
  - Internet

Multidimensional Access Methods (MAMs)

- Students' talk
- V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231

Outline

- Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees
- Multidimensional Access Methods (MAMs)
- Indexing & FQAS: An Introduction
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
- Reading Suggestion
  - [1] chapters 17, 18, 27
  - V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231
  - Internet

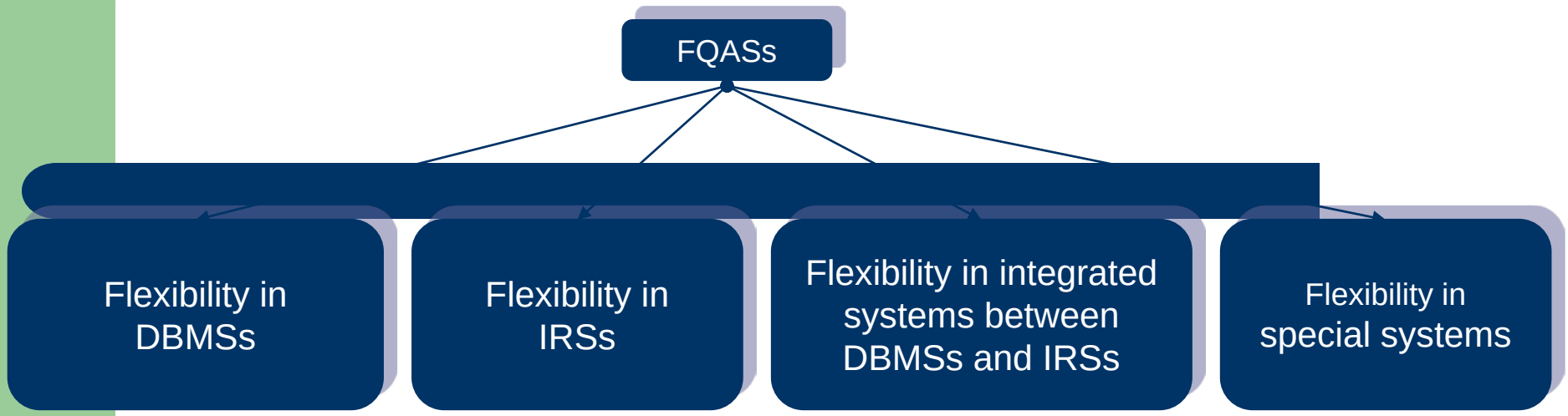
Indexing & FQAS: An Introduction

- The main problem
- Crucial research topics

Indexing & FQAS: An Introduction
The main problem

- Fundamental problem to be solved is to efficiently and flexibly address the empty result set of queries in DBMSs.
  - **Example**: Let's see the query "*Find a hotel room in the centre of HCMC with the price at 40 USD per night*". What will the user have to do more if the result set of this query is empty ??
- Generalized problem is to deal not only with the issue of *user data needs* as DBMSs usually do, but also with the issue of *user information needs* as Information Retrieval Systems (IRSs) must carry out. This is desired in lots of modern application areas (e.g., multimedia DBs, GIS & TIS, D-Lib sys)
- Flexible Query Answering Systems (FQASs) are ones capable of satisfying user information needs (at a certain level)

The flexibility in a variety of application areas



VAGUE, ARES, FLEX, CoBase, PREFERENCES, cooperative answering systems, fuzzy theory or histograms based approaches, etc.

Traditional and Modern IRSs as QBIC, MARS, PhotoBook, Virage, VisualSEEk, etc.

[SaR1990, ScP1982, VCB1996, FuR1997, Fuh1993, BMN1994], VQS, etc.

AQUA system, information sources of diverse nature and structure: [CHN+1995, GrG1997, GCG+1997, GGT1999, MWY+2001, MYL2002, etc.]

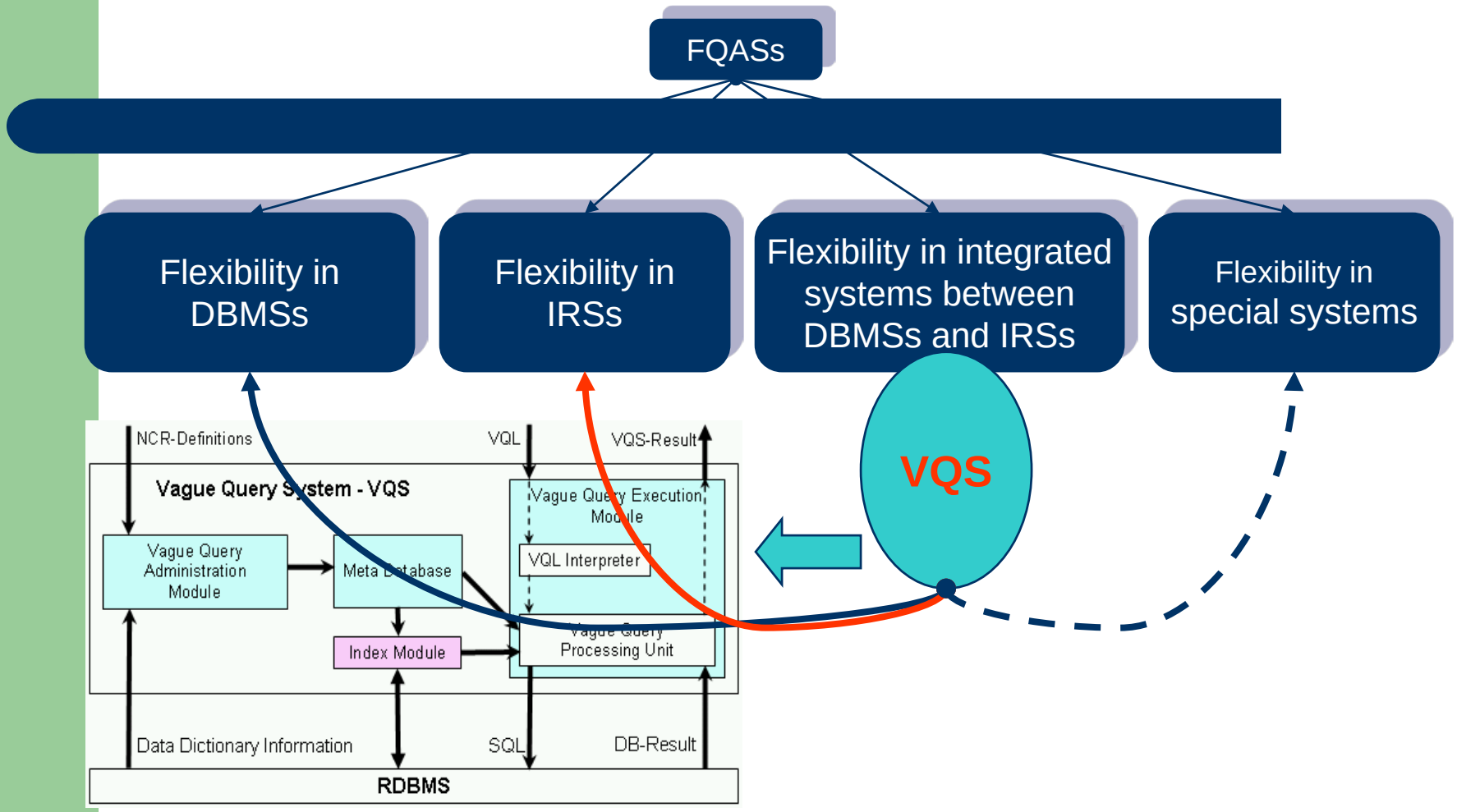
More detailed discussions: seminar

Indexing & FQAS: An Introduction
The main problem

- Although the flexibility in FQASs implies different levels of significance, in our context it can be interpreted as capabilities that provide easy, informative and intuitive access to data for every type of need. To gain this objective in the RDBMSs, we must extend and facilitate them with *flexible retrieval capabilities*, which are unavailable to most existing RDBMSs
- Main challenges:
  - Modeling the system, i.e. the FQAS: functionalities, semantics expression of stored data, possible implementation methods
  - Developing flexibilities for the system: not the flexible querying functionalities of the FQAS, but the flexibility of the FQAS itself
  - Building a transparency and guarantee system: modifications of any part of the database system should not trouble users and the correctness of returned results must be guaranteed

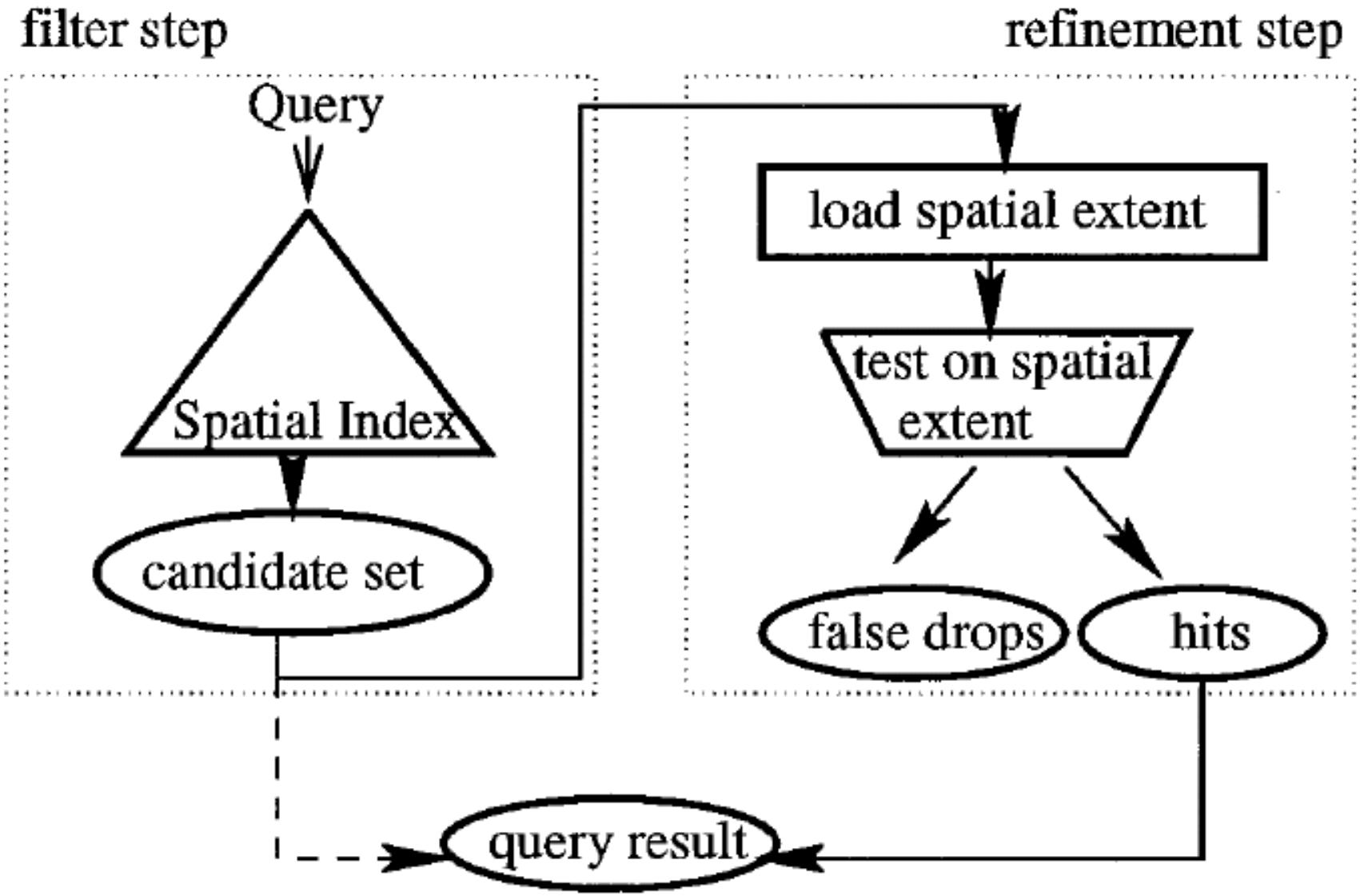
Indexing & FQAS: An Introduction

The main problem



Indexing & FQAS: An Introduction

Crucial research topics



Indexing & FQAS: An Introduction
Crucial research topics

- Multidimensional Index Structures (also MAMs – Multidimensional Access Methods)
  - Support exact as well as similarity queries (NN, k-NN, range, etc.)
- Solving complex multi-feature NN queries
- Solving approximate queries
  - S-FNN, k-NN, range, M-FNN
- Efficient processing of complex and approximate complex joins
  - Any user-defined join predicate
- Towards Integrating MAMs into DBMSs
  - E.g., Oracle with the R-tree (incurs the curse of dimensionality problem)

Outline

- Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees
- Multidimensional Access Methods (MAMs)
- Indexing & FQAS: An Introduction
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
- Reading Suggestion
  - [1] chapters 17, 18, 27
  - V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231
  - Internet

FQAS - Flexible Query Answering Systems

- Similarity Query Types
- Some Existing FQAS (developed for DBs)
  - VAGUE
  - VQS
  - QBIC (IBM)
  - …
- IR systems & Web search (see slides2)
- Emerging applications: LBS, Web service-based applications, …

Outline

- Basic Index Structures: An Overview of Hashing, B-/B$^+$-trees
- Multidimensional Access Methods (MAMs)
- Indexing & FQAS: An Introduction
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS, IRSs & Web search, …
- Reading Suggestion
  - [1] chapters 17, 18, 27
  - V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231
  - Internet

Summary

- Basic Indexing Structures: An Overview of Hashing, B-/B$^+$-trees
- Indexing & FQAS: An Introduction
- Multidimensional Access Methods (MAMs)
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
  - IR systems & Web search
  - Emerging Applications & Open Issues

Q&A

- Note: Related Open Research Directions will be revised & introduced in my last lecture (week 15)