Advanced Indexing Techniques & Flexible Query Answering in DBs

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
- Indexing & FQAS: An Introduction
- Multidimensional Access Methods (MAMs)
- FQAS - Flexible Query Answering Systems
  - Similarity Query Types
  - Some Existing FQAS
- Reading Suggestion
  - [1] chapters 13, 14
  - 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 13, 14
- Note: double check in your coursework
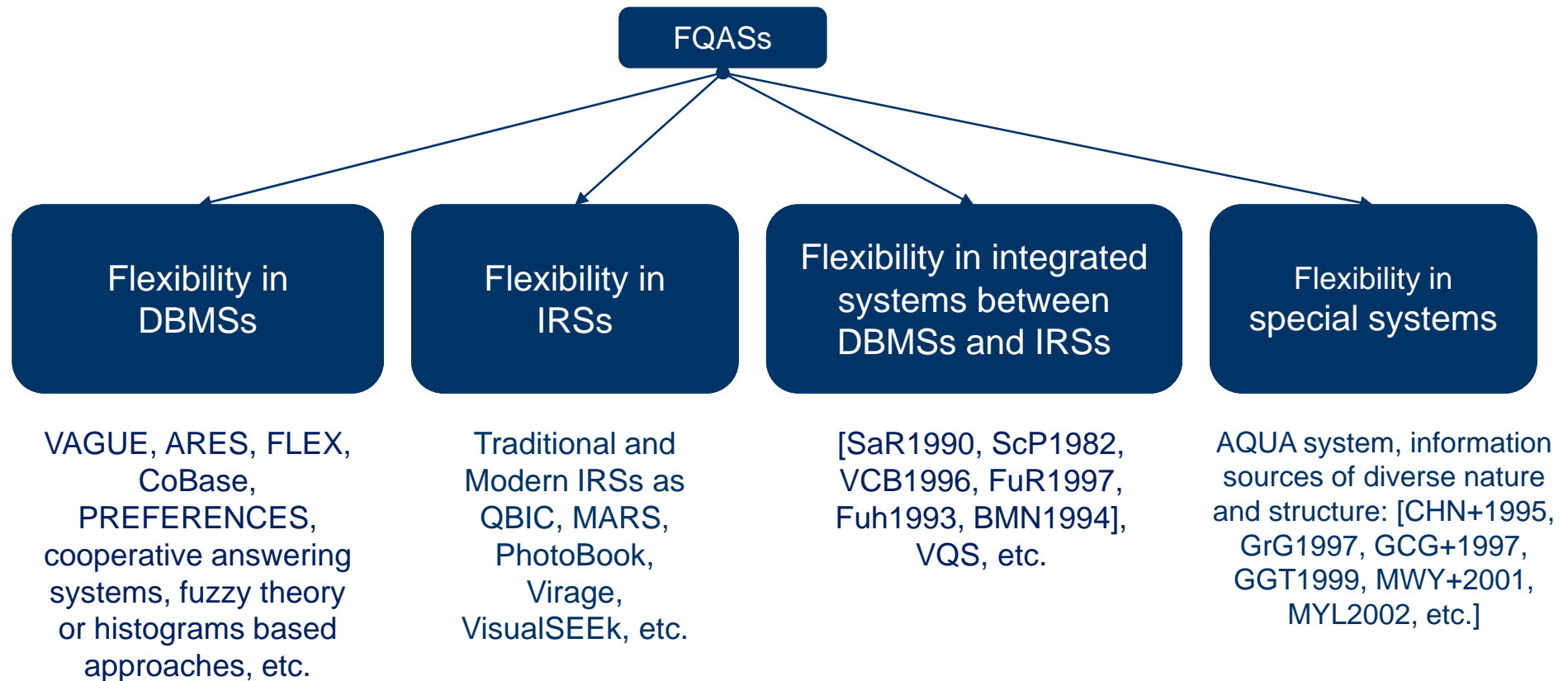
Indexing & FQAS: An Introduction

- 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

```
                          FQASs
```

Flexibility in DBMSs

Flexibility in IRSs

Flexibility in integrated systems between DBMSs and IRSs

Flexibility in special systems

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
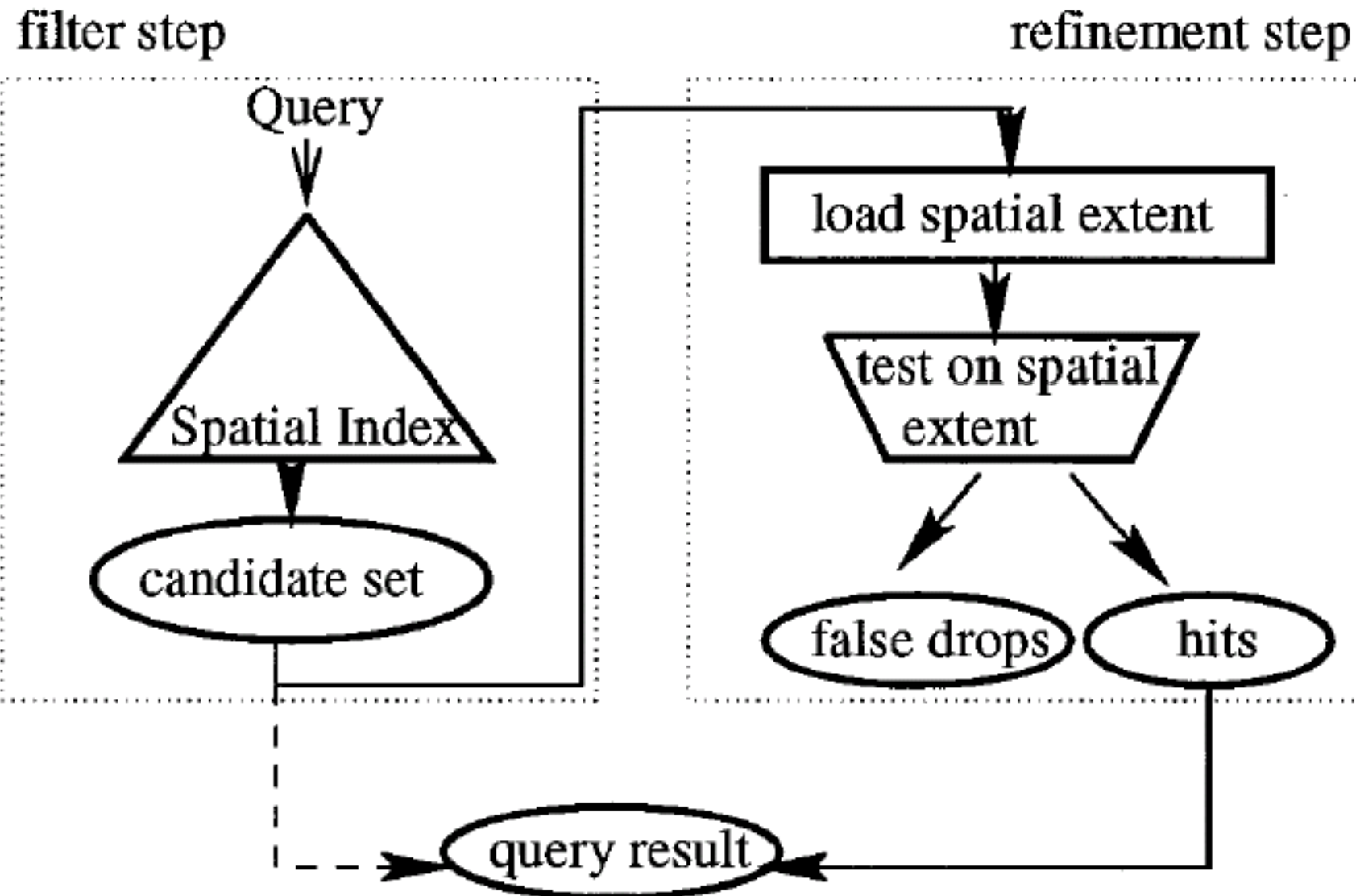
Indexing & FQAS: An Introduction
The main problem
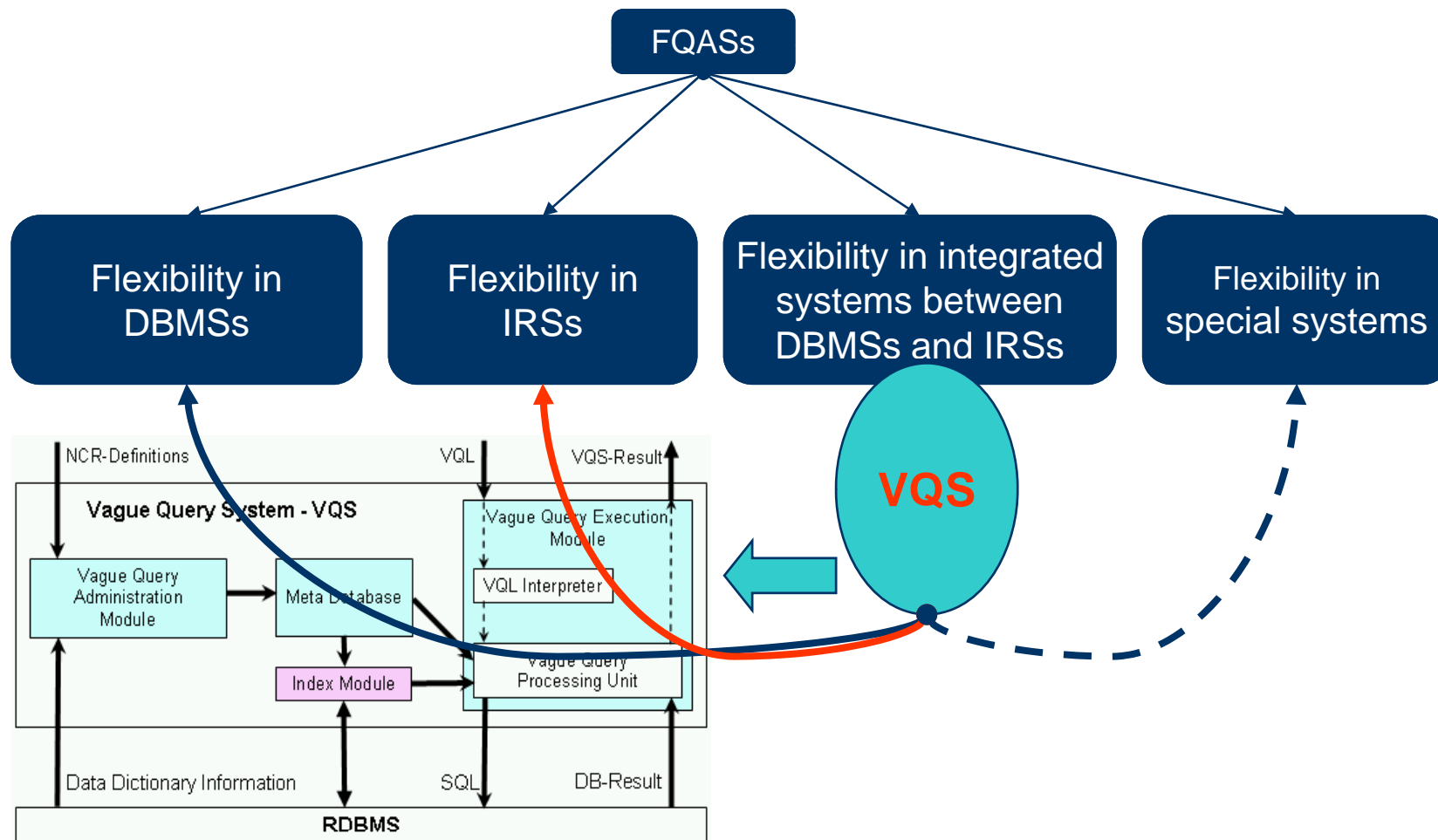
Indexing & FQAS: An Introduction
Crucial research topics

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 <u>approximate</u> 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)

Multidimensional Access Methods (MAMs)

- Talks: R-/R*-trees
- V. Gaede, O. Günther: "*Multidimensional Access Methods*", ACM Computing Surveys (CSUR), 30(2), June 1998, 170 – 231

FQAS - Flexible Query Answering Systems (*next week*)

- Similarity Query Types
- Some Existing FQAS (developed for DBs)
  - VAGUE
  - VQS
  - QBIC (IBM)

  - …

- Emerging applications: LBS, Web service-based applications, …

Summary

- Basic Indexing Structures: An Overview of Hashing, B-/B$^+$-trees
- Indexing & FQAS: An Introduction
- Multidimensional Access Methods (MAMs)
- FQAS - Flexible Query Answering Systems (*next week*)
    - Similarity Query Types
    - Some Existing FQAS
    - Emerging Applications & Open Issues
- Next Week: FQAS
    - Students' talks + group discussion
    - Reading materials: internet + (see my homepage)

Q&A

Questions ??

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