Technologische verdieping databases

Dr.Ir. J.H. ter Bekke

Faculteit Informatietechnologie en Systemen, TU Delft

http://is.twi.tudelft.nl/dbs/terBekke.html

- Modern Database Technology

  Elmasri/Navathe, Fundamentals of Database Systems.

  Addison Wesley Longman Inc., Third edition (2000).

- Trends in Database Technology

  The Asilomar Report on Database Research (September 1998).

  http://www.research.microsoft.com/~gray/Asilomar_DB_98.htm

The Impact of Databases

Databases play a critical role in almost all areas where computers are used, including business, engineering, medicine, law, education, library science, etc.

A database is a collection of related data.

- it represents some aspect of the real world
- it is a logically coherent
- it serves a specific purpose
- can be of any size and of varying complexity

Traditional Examples:

- bank accounts
- hotel reservations
- library catalog
- magazine subscriptions
- purchasing in a supermarket

Figure 3.2 ER schema diagram for the company database.

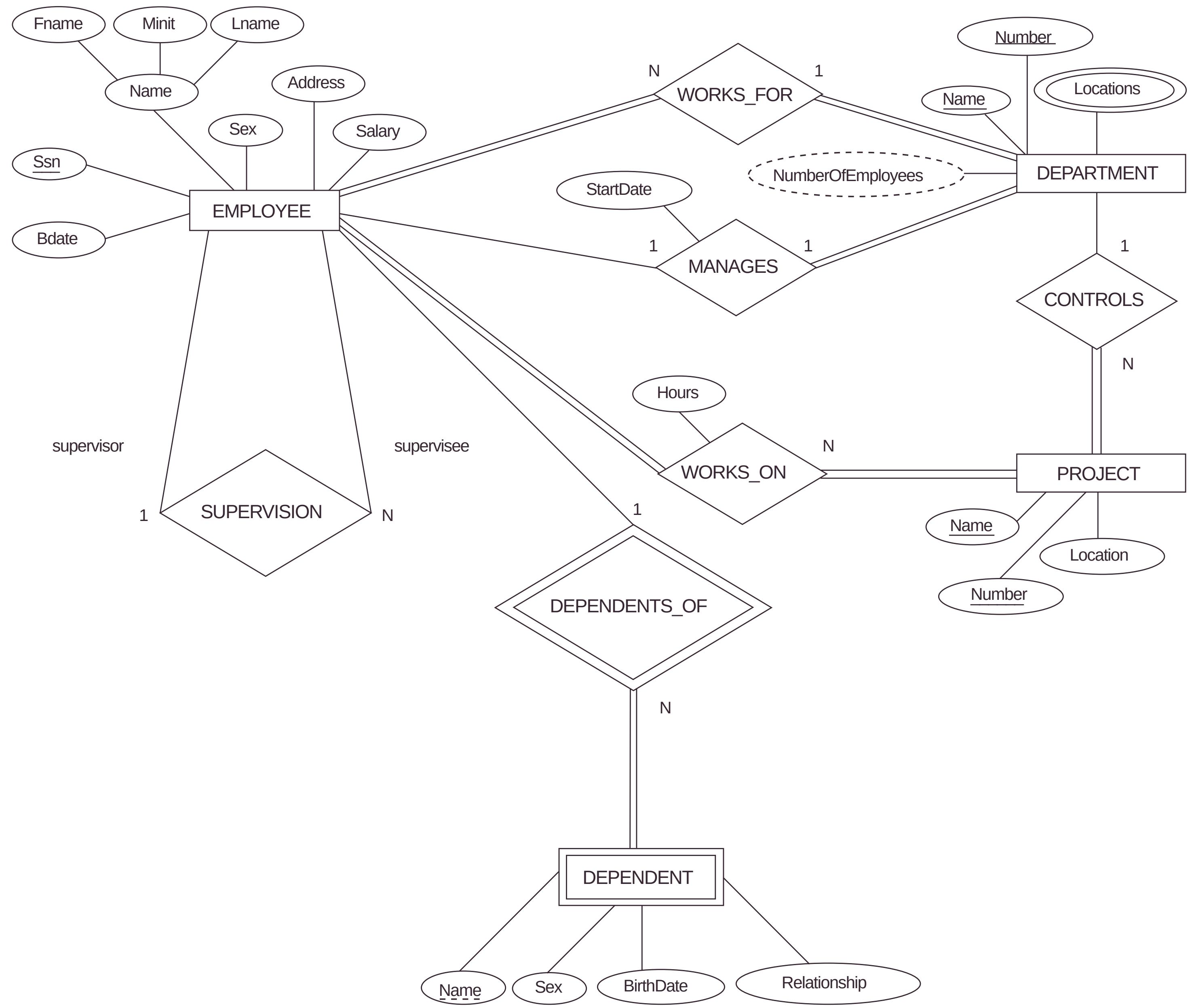
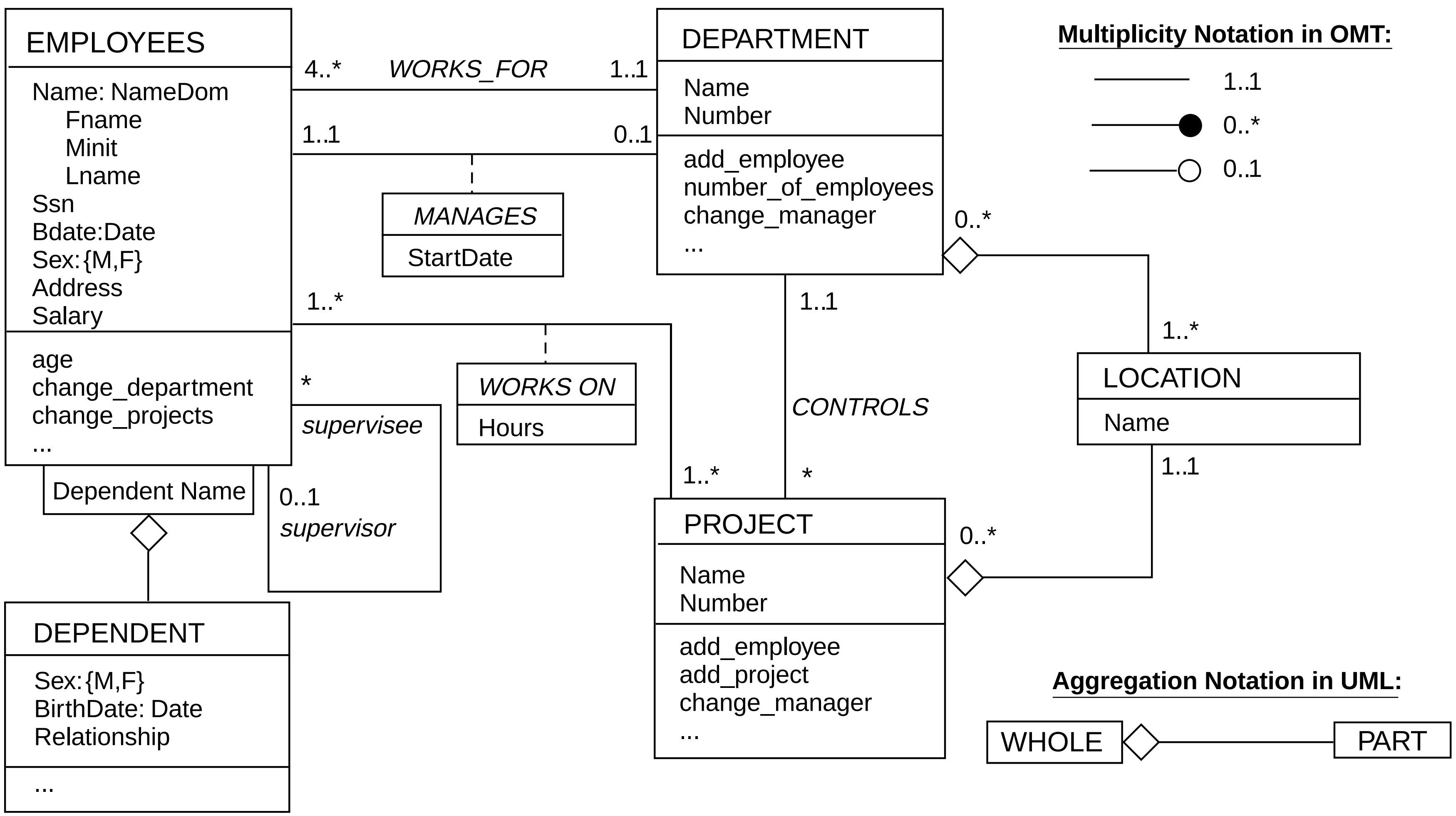


Figure 4.11 The UML conceptual schema for the COMPANY database in Figure 3.15.



Modern examples:

- engineering design and manufacturing
- software design and engineering
- telecommunications
- images and graphics
- multimedia databases
- geographic information systems
- data mining and data warehouses
- real-time databases
- internet databases

Elmasri/Navathe:

These types of databases have more complex requirements than do the more traditional applications. To represent these requirements as accurately and clearly as possible, designers of database applications must use additional semantic data modeling concepts.

Various semantic data models have been proposed in the literature.

Current approaches:

- RELATIONAL MODEL (Codd, 1970)
based on mathematical concepts, is used in modern DBMSs; standardized languages for definition and manipulation.

- ENTITY-RELATIONSHIP APPROACH (Chen 1976)
pragmatic approach, digrams are often used in design tools for DBMSs.

- SEMANTIC HIERARCHY MODEL (Smith & Smith 1977)
abstractions: classification, aggregation, generalization; no data manipulation.

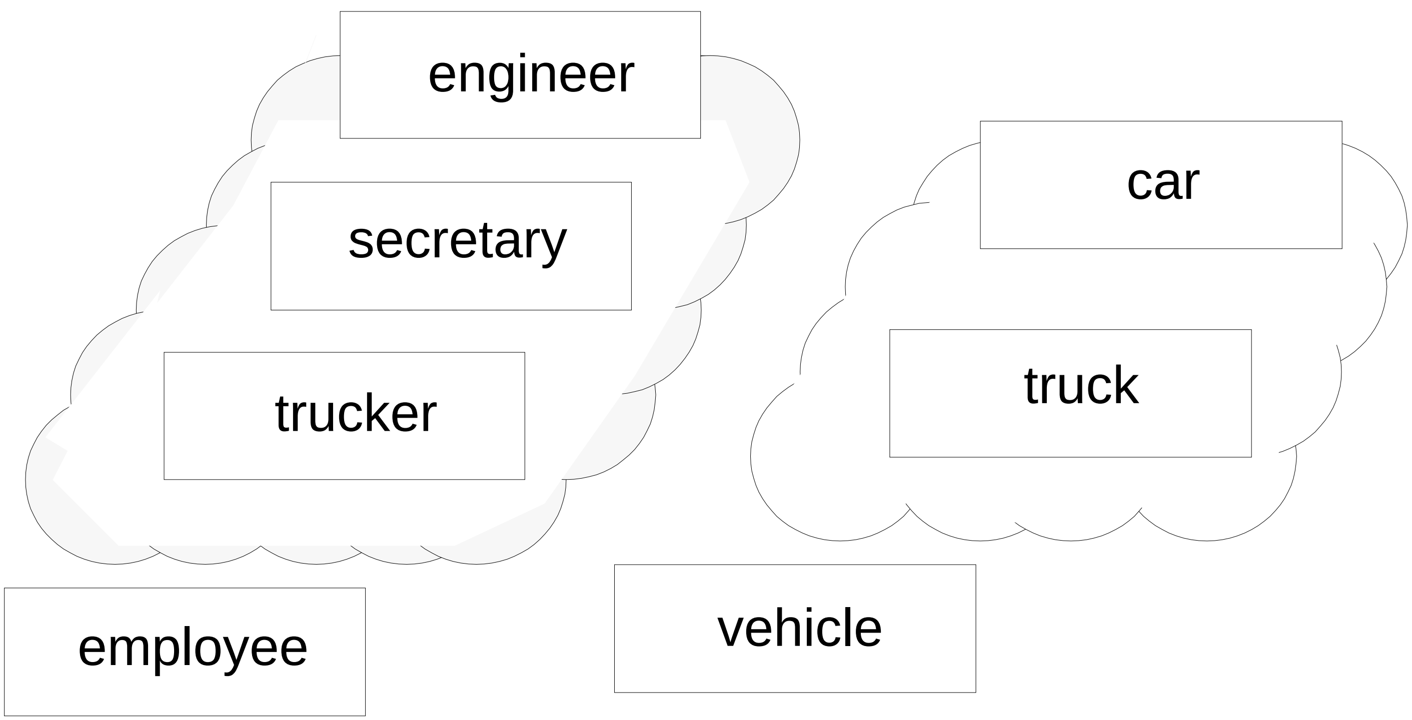
Modern approaches: an overview

- SEMANTIC HIERARCHY MODEL (Smith & Smith 1977)
  abstractions: classification, aggregation, generalization; no data manipulation.
- EXTENDED RELATIONAL APPROACH (Codd 1979)
  theoretic extensions to the basic relational approach to capture more meaning.
- SEMANTIC NETWORKS (Brachman 1979)
  artificial intelligence: generalizations and inheritance play a dominant role.
- ENHANCED ENTITY-RELATIONSHIP APPROACH (> 1980)
  extensions to the original model proposed by Chen (1976)
- FUNCTIONAL DATA MODEL (Shipman 1981)
  model based on functions, with an emphasis on inheritance.
- ACTIVE/PASSIVE COMPONENT MODELING (Brodie 1982)
  abstractions: classification, aggregation, generalization, association; integrity control by procedures instead of structures.
- OBJECT ORIENTED incl. object modeling technique (Rumbaugh 1991)
  concepts: method, encapsulation, inheritance, polymorphism.
- SEMANTIC DATA MODELING (ter Bekke 1992)
  abstractions: classification, aggregation, generalization and object relativity, includes a complete data manipulation language.

Example: Semantic Hierarchy Model
(John Smith & Diana Smith)

Introduced abstractions:

- classification
- aggregation
- generalization



Example: Semantic Hierarchy Model

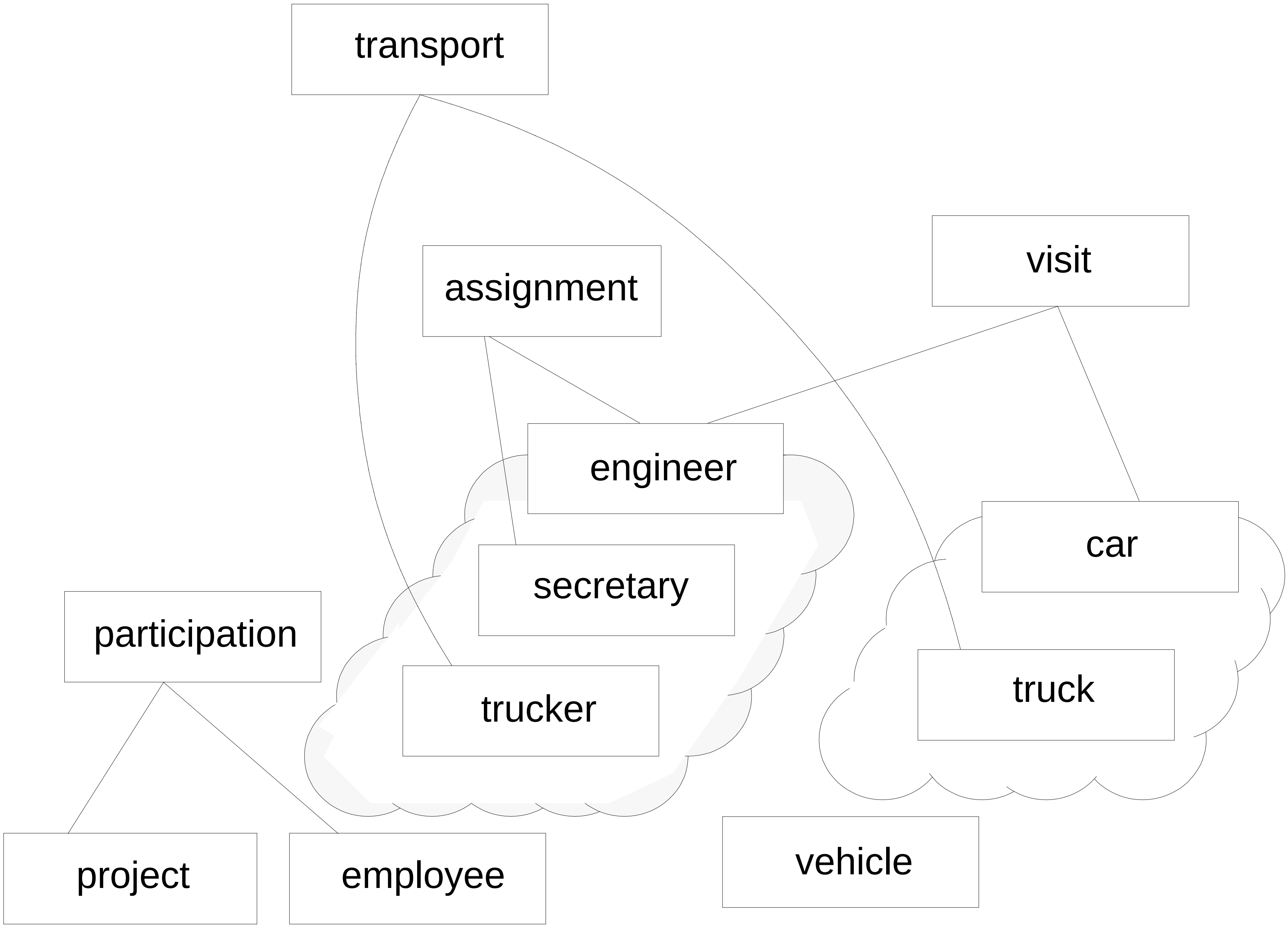


Figure 4.7 A specialization lattice (with multiple inheritance) for a UNIVERSITY database.

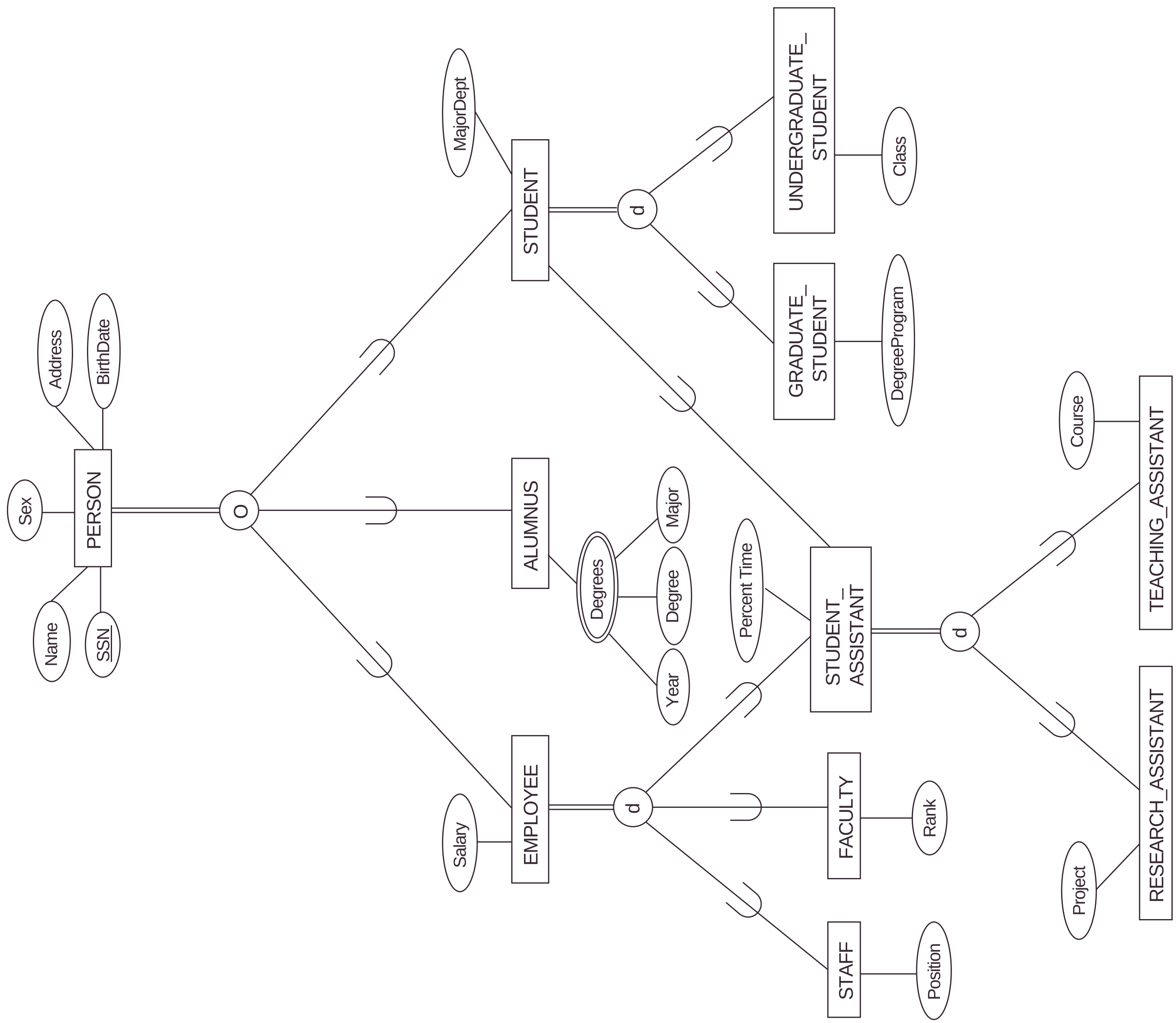
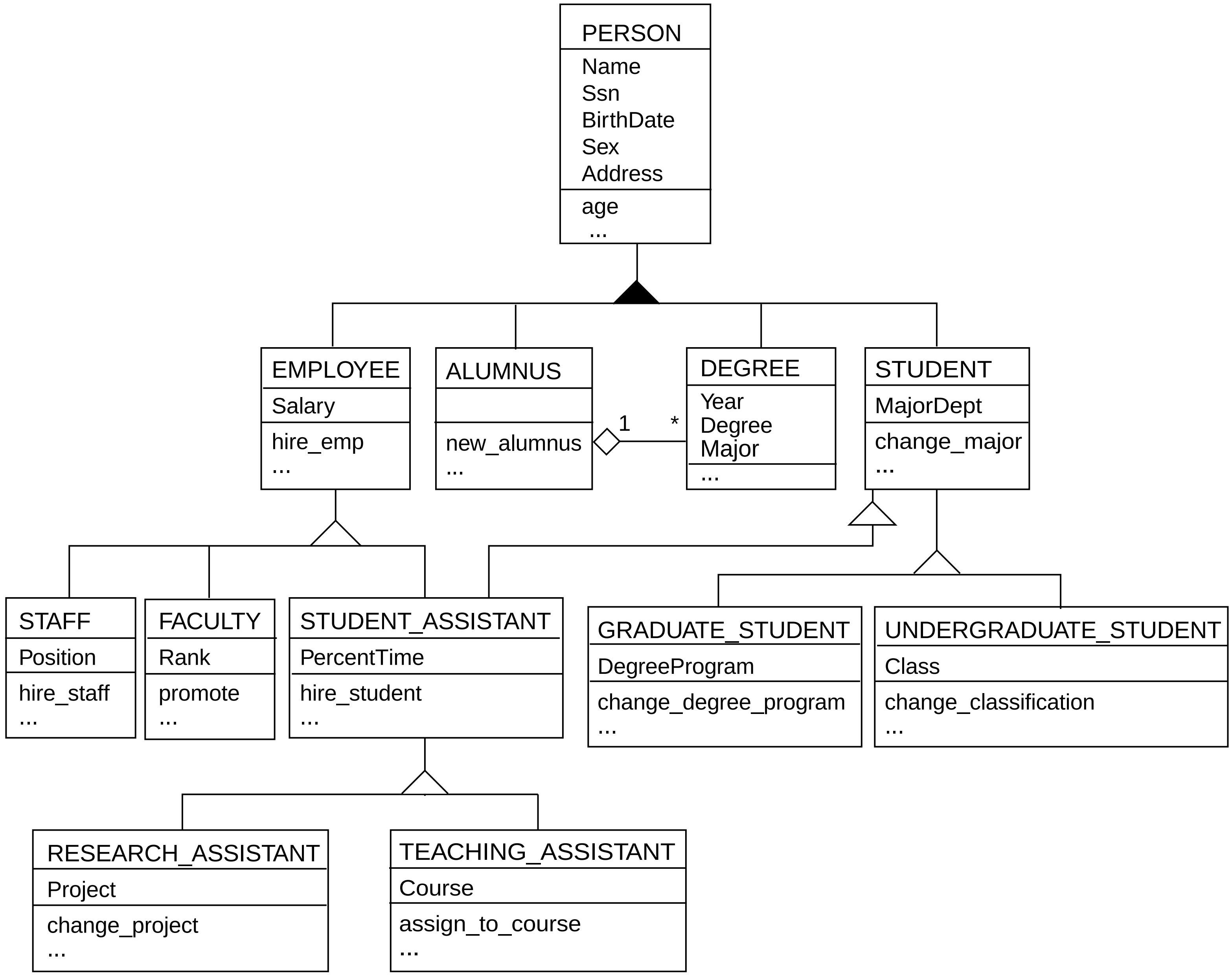
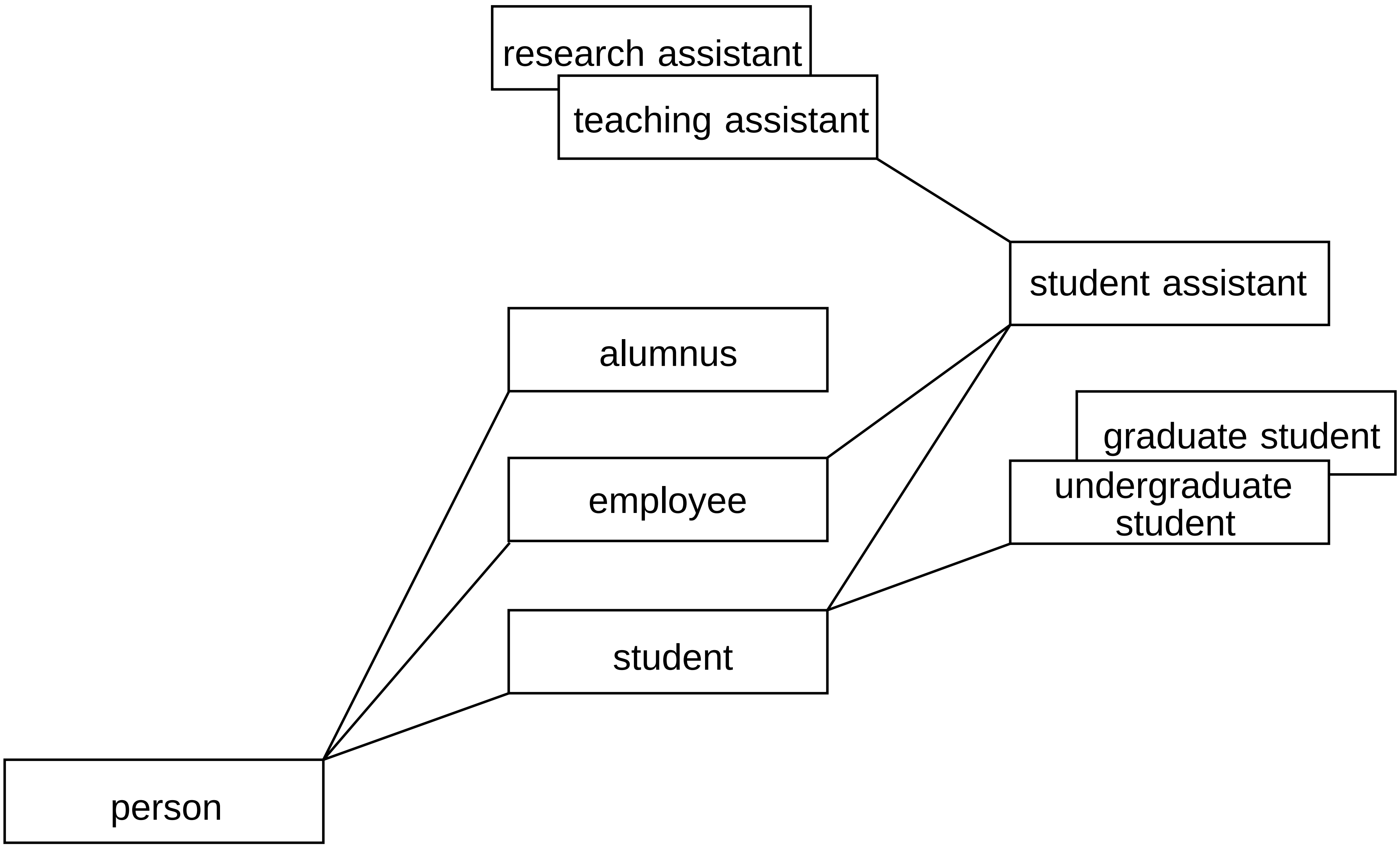


Figure 4.12 Specialization/generalization notation in UML shown by a class diagram corresponding to the EER diagram in Figure 4.7.

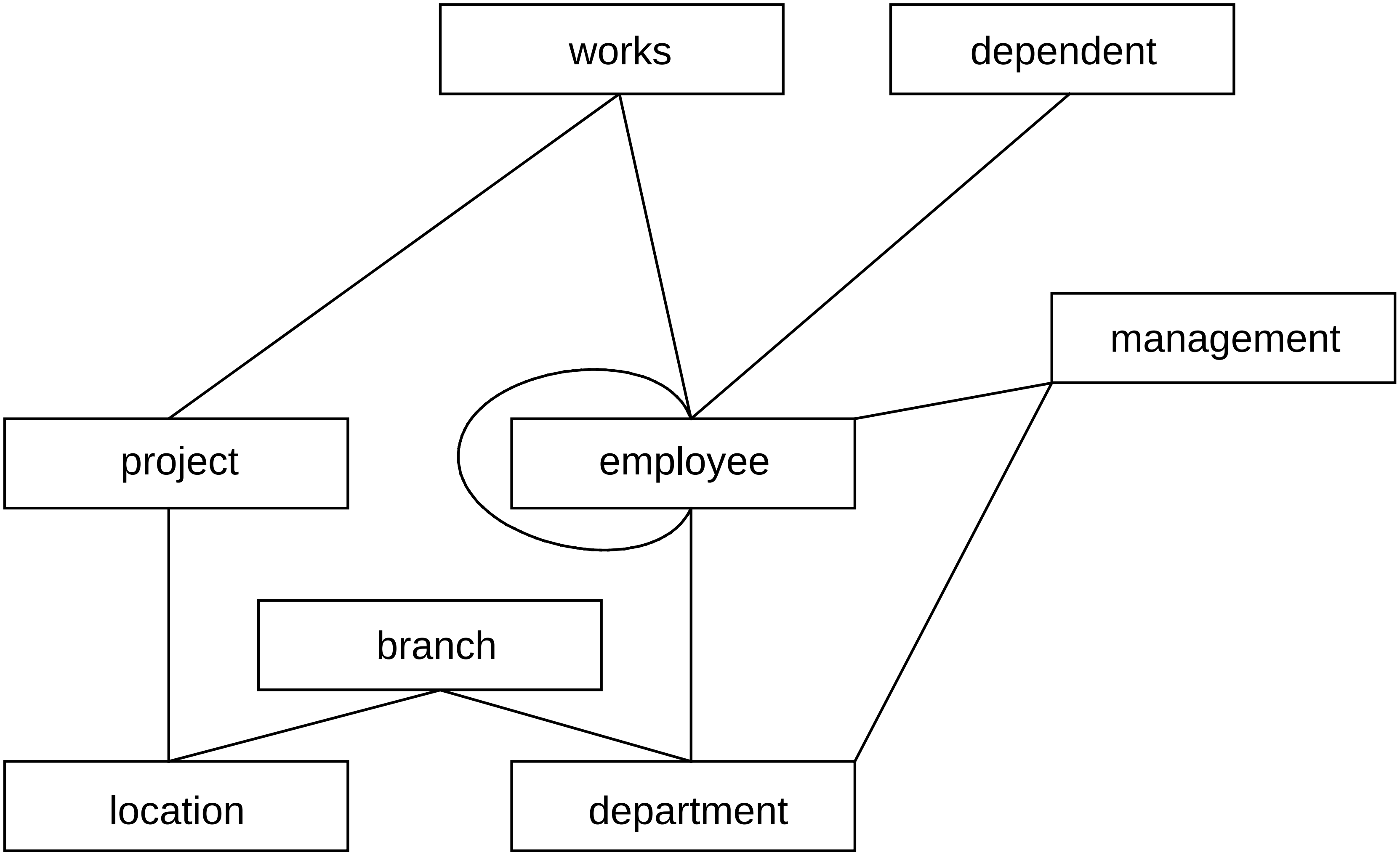


Semantic Solution: Xplain Hierarchy



According: J.H. ter Bekke, Semantic Data Modeling, Prentice Hall 1992

Semantic Solution: Xplain Hierarchy



Driving forces for database research:

- Internet Publishing Using HTML and XML;
- Complex Application Environments;
- Hardware Advances.

However some obstacles appear:

- XML is a language that can be used to describe structured data, however: its query language is reminiscent of the procedural query processing languages prevalent 25 years ago.
- XML is also driving the development of client-side data caches that will support updates, which is leading the XML designers into a morass of distributed transaction issues.

First requirement: Flexible structuring.
Cf. Data modeling developments:

- HIERARCHICAL: data represented in a tree structure.
  limited structuring capabilities, low level data manipulation.
- NETWORK: data represented in a network structure.
  more structuring capabilities, complex data manipulation by navigation.
- RELATIONAL: data represented in tables.
  simple data model, query language, wide range of hardware, limited query facilities, pitfalls, limited integrity control, relative low performance.
- SEMANTIC: data represented in tables, emphasis on abstractions.
  structuring capabilities, query language, reliability, flexibility, high performance.

Second requirement: Reliable query processing.

Pitfall: Select items with descending sales figures.

```
SELECT I.ITEM#
FROM SALES S , ITEMS I
WHERE S.ITEM# = I.ITEM#
AND S.WEEK# = 2
GROUP BY I.ITEM#
HAVING SUM (S.QTY) <
        (SELECT SUM (S.QTY)
        FROM SALES S
        WHERE S.ITEM# = I.ITEM#
        AND S.WEEK# = 1)
```

Computer systems a required to be AUTO-EVERYTHING:

- autoinstalling;
- automanaging;
- autohealing;
- autoprogramming.

In ten years time:

- Billions of people will be connected to the Web;
- Trillion "gizmos" will also be connected to the Web. Examples are computers in smart-cards, telephones, portable organizers. Each gizmo being candidate for database system technology.

Consequences:

- Gizmos will lead to an explosion in the size and scale of data clients and servers - trillions of gizmos will need billions of servers. Current software architectures are unsuitable for supporting such devices.

- Most gizmos will not have a user interface and cannot have a database administrator - they must be self-managing, very secure, and very reliable.

Proposed research agenda:

- Plug and play database management systems: no knobs operation self-tuning and adaptable database, no human-settable parameters.
- Billions of web clients will accessing millions of databases. A first aspect is query optimization, a second aspect is one of semantics (imprecise information) and execution of queries.

Ten-year goal: make it easy for everyone to store, organize, access, and analyze the majority of human information online. The majority of human information will be on the Web in ten years. It will be an exabyte (10^{18} bytes) spread around the planet in many formats.