



Unit 7

Improving the Quality of a Database Schema

本單元目的

根據下列特性, 重建 schema, 以改善 schema 之品質.

- ✦ **Completeness** 完整性(即完整表達 AP domain 之特徵)
- ✦ **Correctness** 正確性(正確使用 ER model 的語法和語意)
- ✦ **Minimality** 極小性
- ✦ **Expressiveness** 表達性高
- ✦ **Readability** 可讀性
- ✦ **Self-explanation** 自我表達性
- ✦ **Extensibility** 有擴充性
- ✦ **Normality** 正規化

Outline

- ★ Qualities of a Database Schema
- ★ Schema Transformations
- ★ Transformations for Achieving Minimality
- ★ Transformations for Achieving Expressiveness and Self-Explanation
- ★ Transformations for Achieving Normalization
- ★ An Example of Schema Restructuring

Minimality

每個概念在 schema 中只出現一次, minimal schema 中任何概念都不能刪除, 否則會失去某些欲表達的資訊。

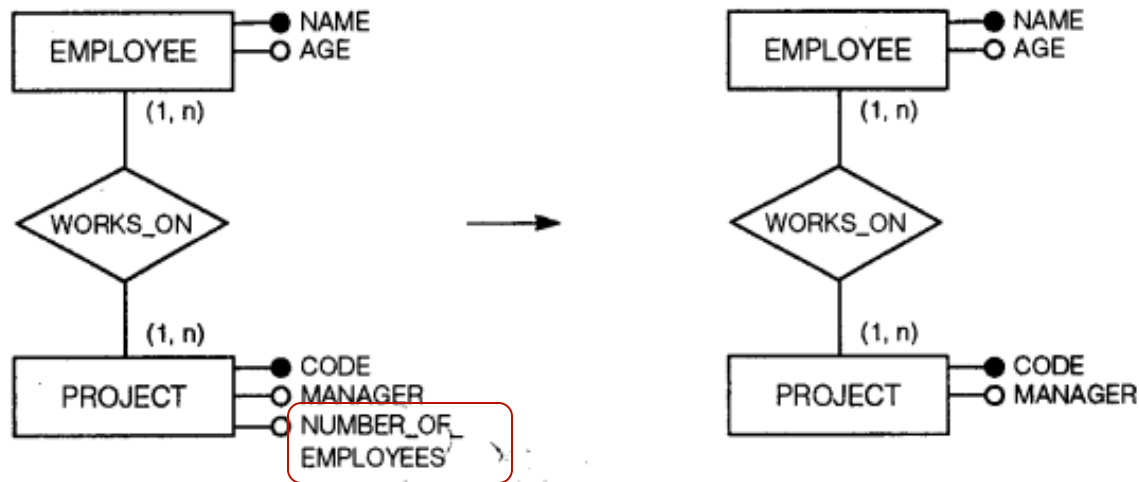
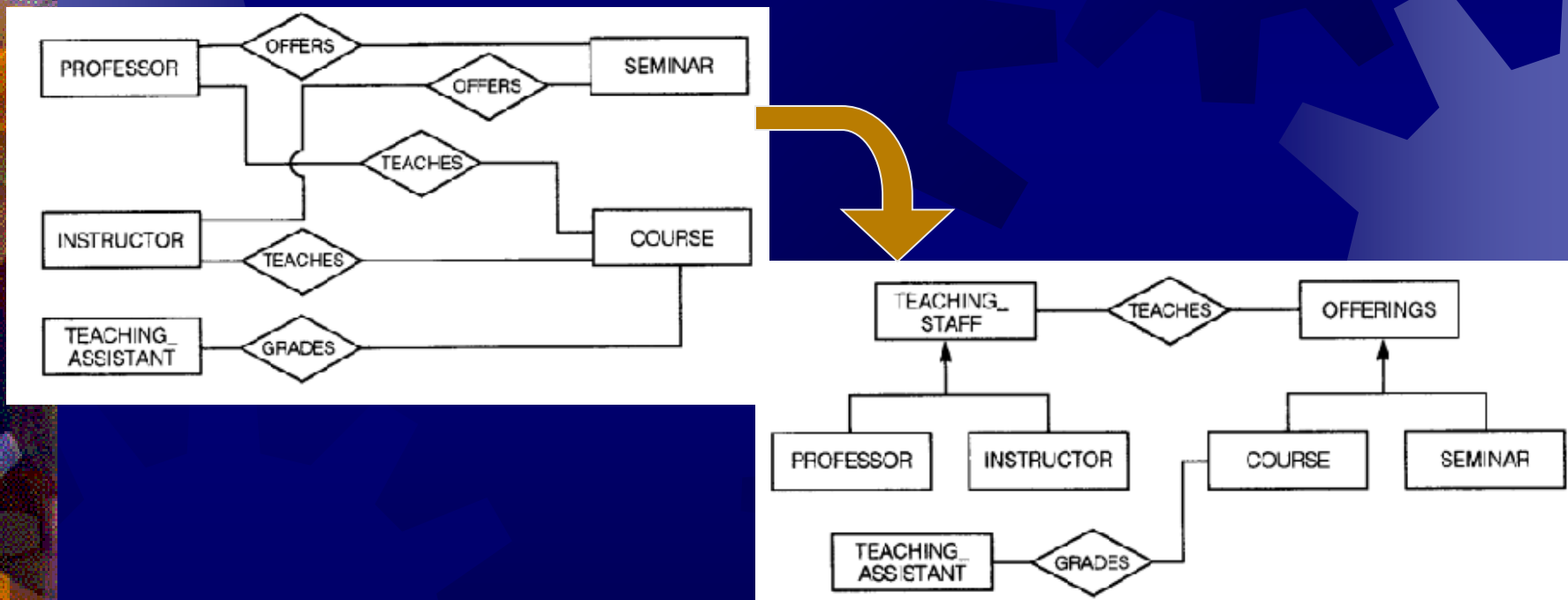


Figure 6.1 A redundant schema

可由 WORKS_ON 的關係
找到該 attribute 可以去掉

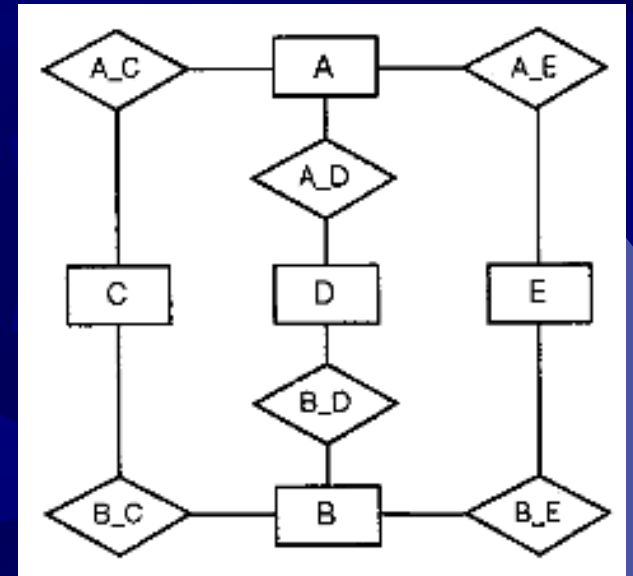
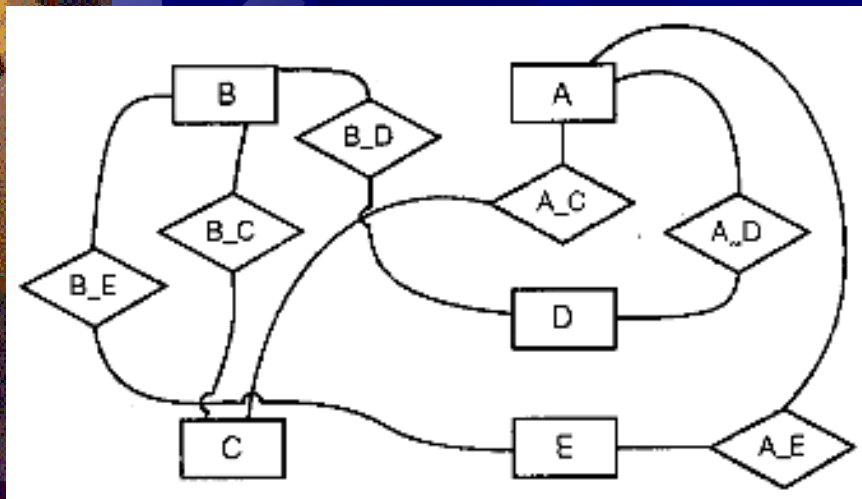
Expressiveness

直接從 schema 即能明瞭欲表達的概念，不需要再另外的解釋或文字說明。



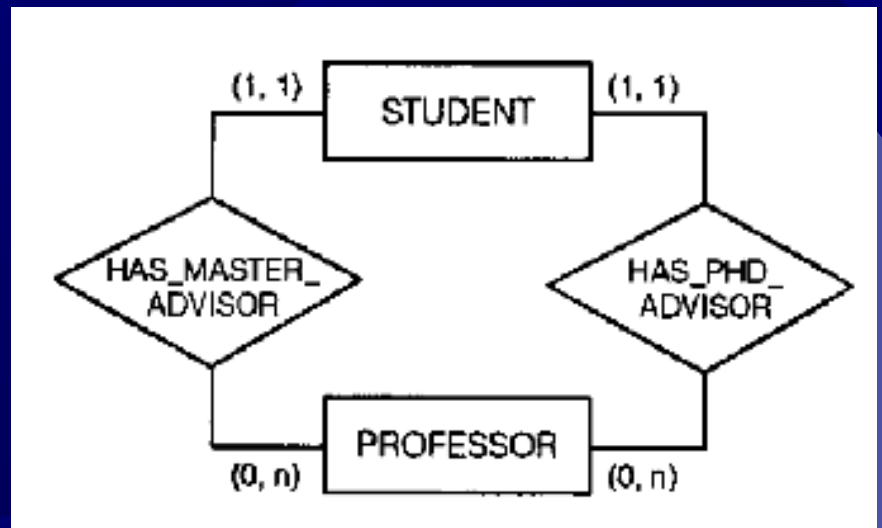
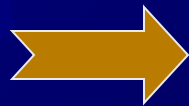
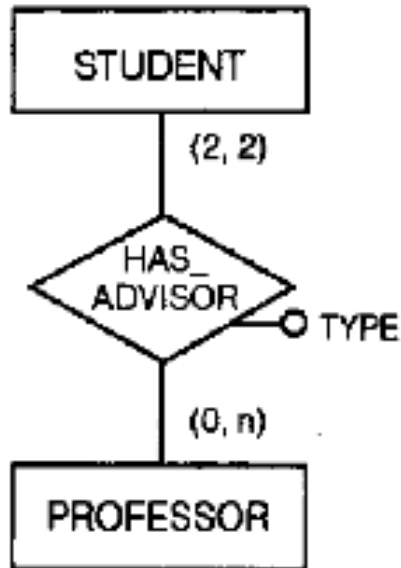
Readability

畫出的圖形要可讀性高



- Entity 及 relationship 圖形大小一致,
- 連線用平行或垂直線, 盡量減少線條交叉的個數

Self-explanation



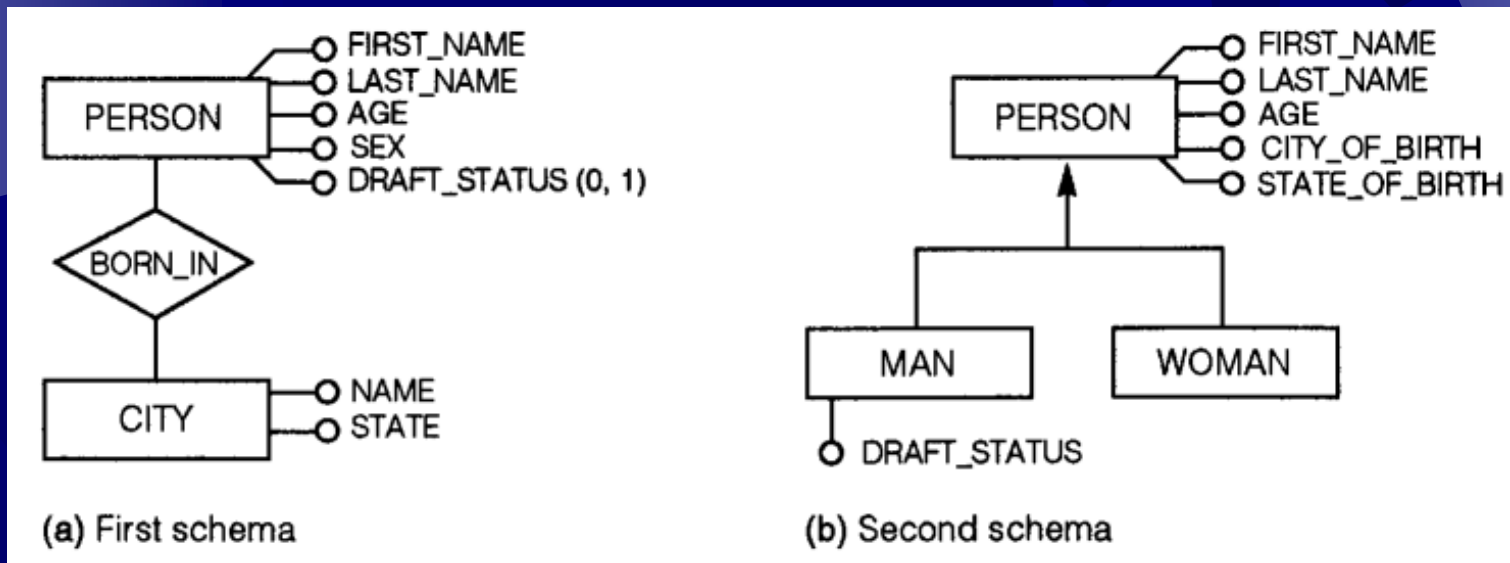
Schema Transformations

- ★ After transformation, the results maybe
 - Information-preserving transformation (不改變)
 - Information-changing transformation
 - Augmenting (擴大)
 - Reducing (縮小)
 - Noncompatible (不相容)
- ★ 為改善 schema 品質而重建之 schema 應是 information-preserving transformation

Schema Transformations

★ Same information content for two schema

若 schema S1 和 S2 是相等(equivalent) 則表示在 S1 中的任何查詢 Q 可以在 S2 中找到一查詢 Q', 使得 Q 與 Q' 查詢結果是相同.



Two schemas with the same information content

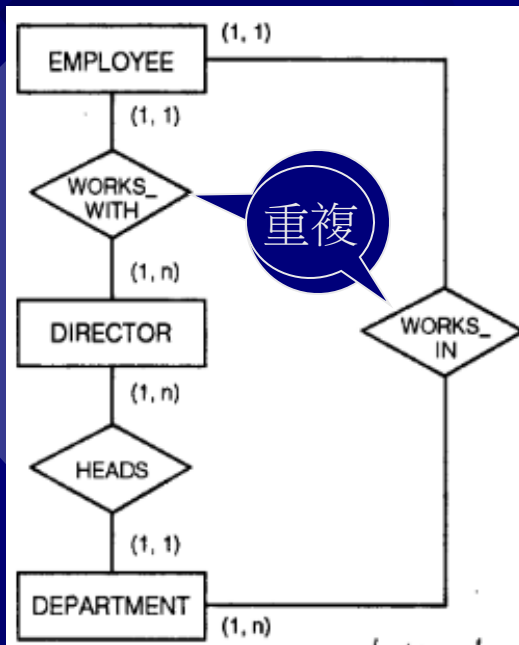
Transformations for Achieving Minimality

合併或轉換後形成的重複關係包括

- **Cycles of relationship**
形成循環的現象
- **Containment constraints**
某個 relationship 包含在另一個 relationship 中
- **Derived attributes**
某個 attribute 可由其他的 attributes 計算出來
(見 page 4, Fig. 6.1)
- **Implicit subsets**
某個 subset 隱含另一原有的 subset

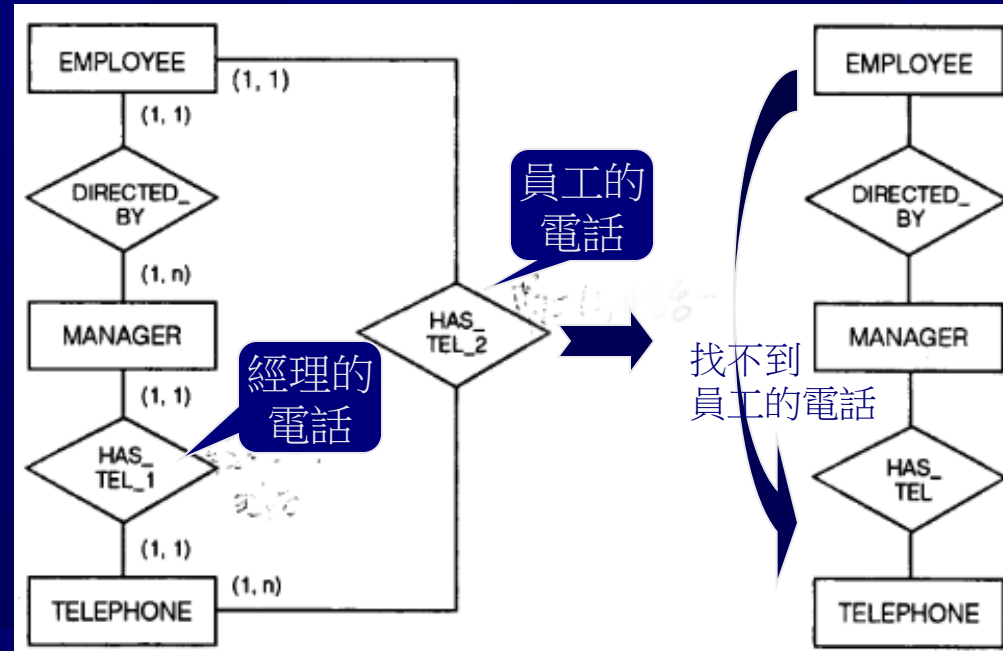
將重複關係從轉換後的 **schema** 中找出來並刪去, 以達到最小化 (**minimality**) 的目的

Cycles of relationship



發生循環且重複

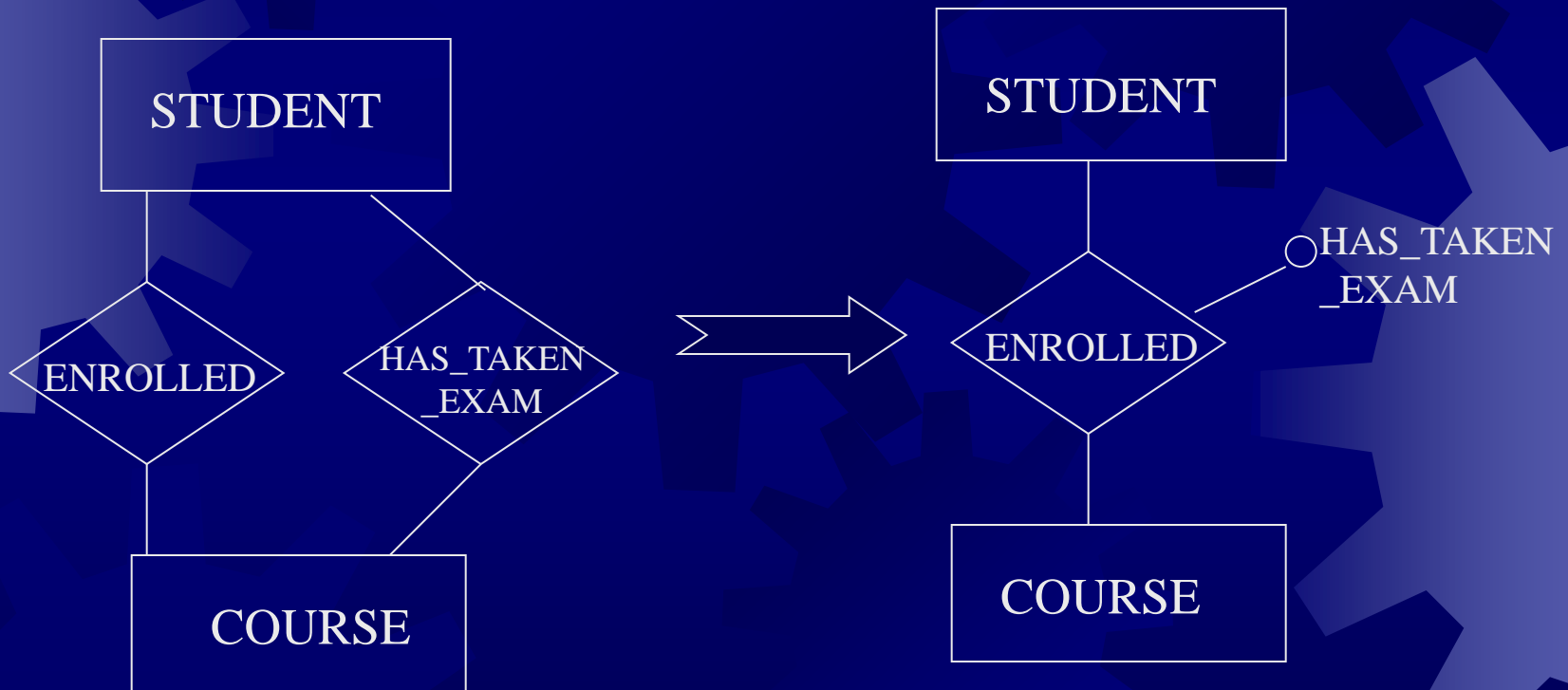
是否發生重複的現象可由 cardinality 看出對等關係.



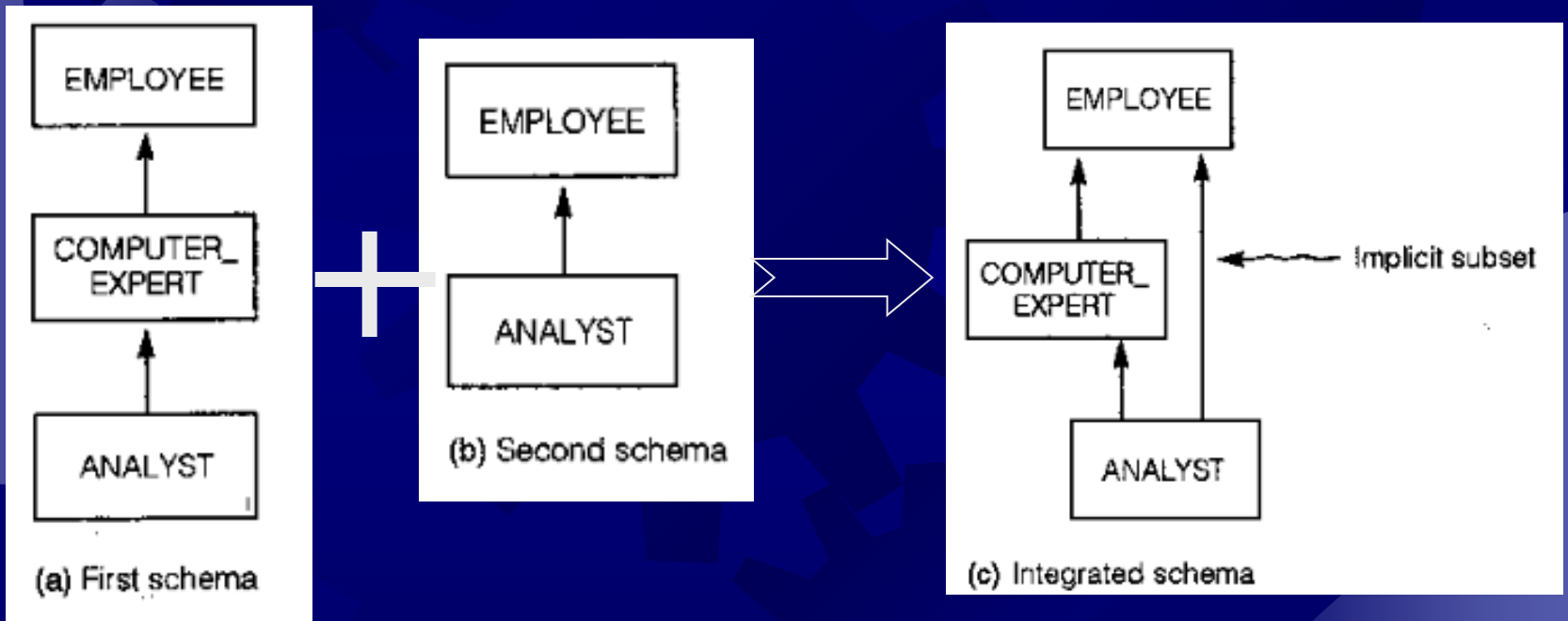
Relationship 雖然形成 cycle 但不重複
把 HAS_TEL_2 去掉則意義不同

Containment constraints

某個 relationship 包含在另一個 relationship 中



Implicit subsets

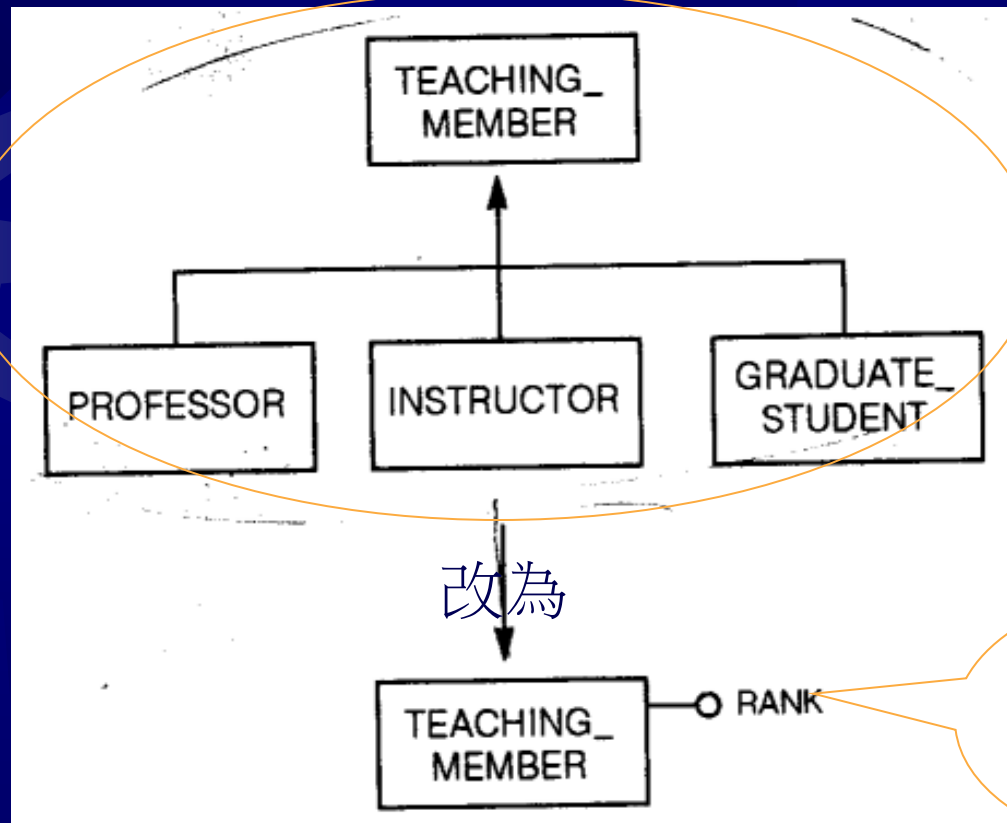


Transformation for Achieving Expressiveness & Self-Explanation

轉換之後做簡化, 以提昇schema 表達性及自我解釋的能力

- ✱ Elimination of dangling subentities in generalization hierarchies
 - ✱ Elimination of dangling entities
- } 把 attributes 很少的 entity 合併到其他的 entity
- ✱ Creation of generalization
 - ✱ Creation of a new subset
- } 為充分表達其特定之含義

Elimination of dangling subentities in generalization hierarchies.



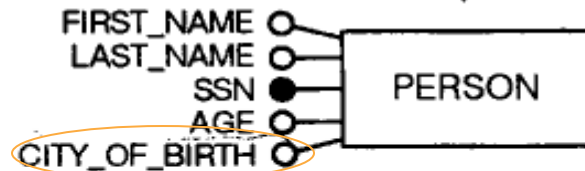
subentities
的特性差異不大

全部合併為唯
一個 entity,
用 RANK 表示
不同的 subentities

Elimination of dangling entities



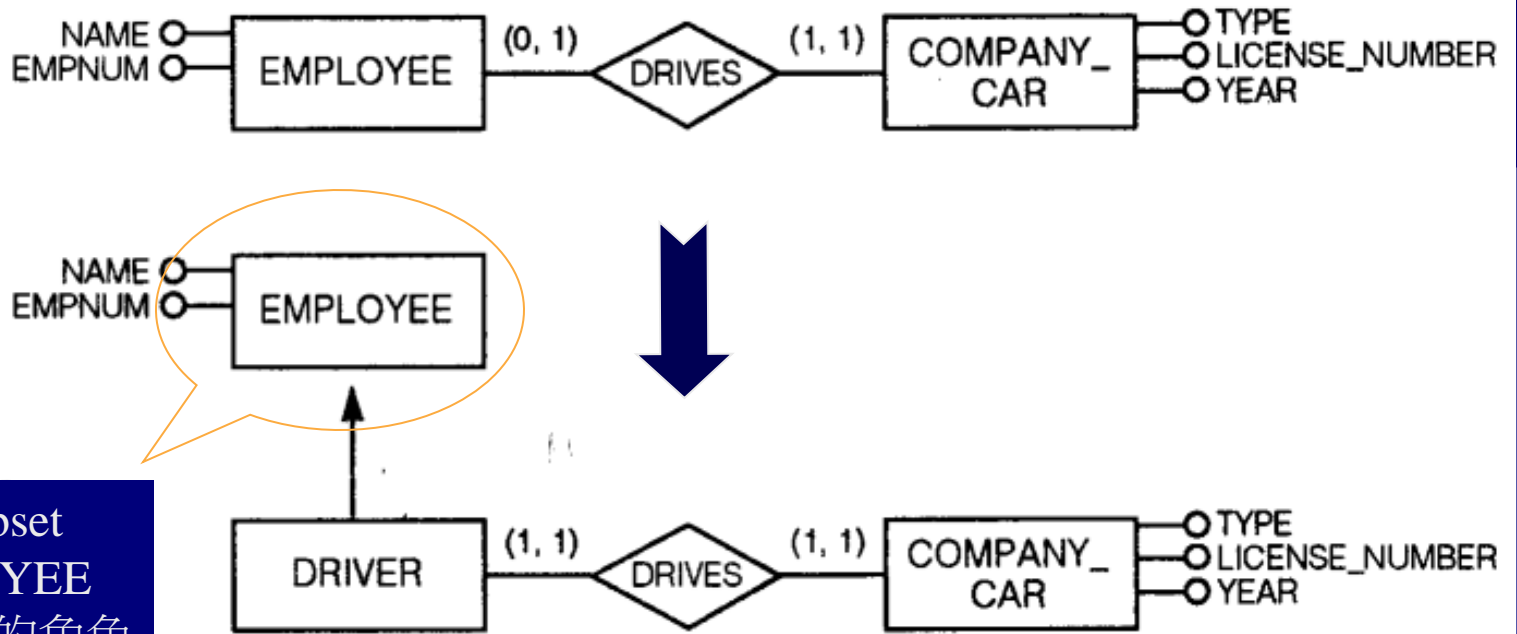
改為



只有一個 attribute 是 dangling entity

Figure 6.10 Elimination of a dangling entity

Creation of a new subset



產生新的 subset
強調EMPLOYEE
中 DRIVER 的角色

Normalization

Normalization is the analysis of functional dependences between attributes.

★ Functional dependences (功能相依)

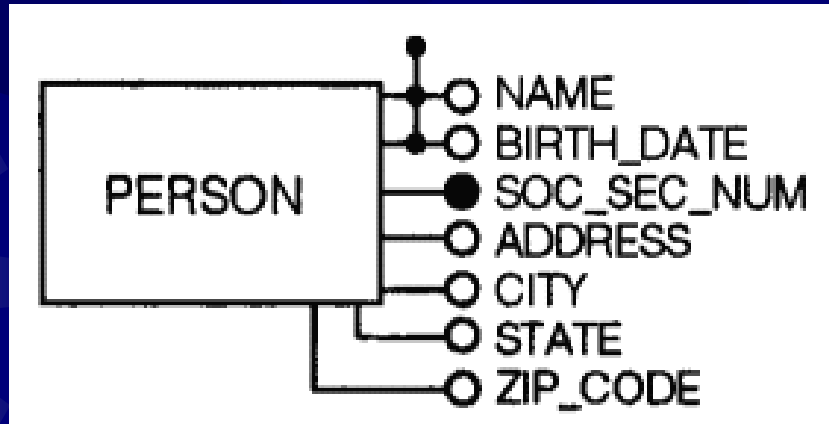
$$A_1 \longrightarrow A_2$$

A_1 and A_2 are attributes.

A_1 的值可以決定 A_2 的值, 則稱 A_1 和 A_2 是功能相依.

A_1 is a determinant.

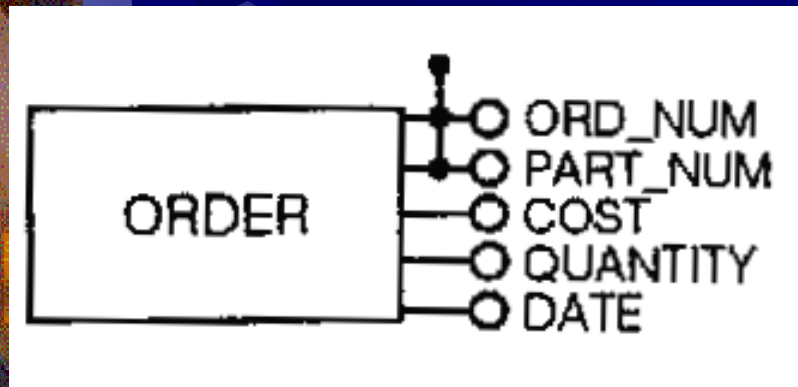
在正確的 schema 中，identifier 可決定其他所有 attribute 的值。



$SOC_SEC_NUM \Rightarrow NAME, BIRTHDAY, ADDRESS, CITY, STATE, ZIP_CODE$

$NAME+BIRTHDAY \Rightarrow NAME, SOC_SEC_NUM, ADDRESS, CITY, STATE, ZIP_CODE$

Update Anomalies (更新異常現象)



ORD_NUM+PART_NUM
→ COST, QUANTITY,
DATE

PART_NUM → COST

★ Instances of the entity ORDER

1518, PEN, 1, 12, 3-8-90
1518, PENCIL, 0.5, 15, 3-8-90
1521, PENCIL, 0.5, 18, 2-9-89
1407, PEN, 1, 15, 2-6-89
1407, ERASER, 0.2, 28, 2-6-89
1407, BOARD, 5, 3, 2-6-89
1729, ERASER, 0.2, 1, 3-1-90
1729, DISKETTE, 2, 10, 3-1-90
1729, PENCIL, 0.5, 15, 3-1-90

Update Anomalies

Insertion anomaly

有 order 才可能加入某零件的成本

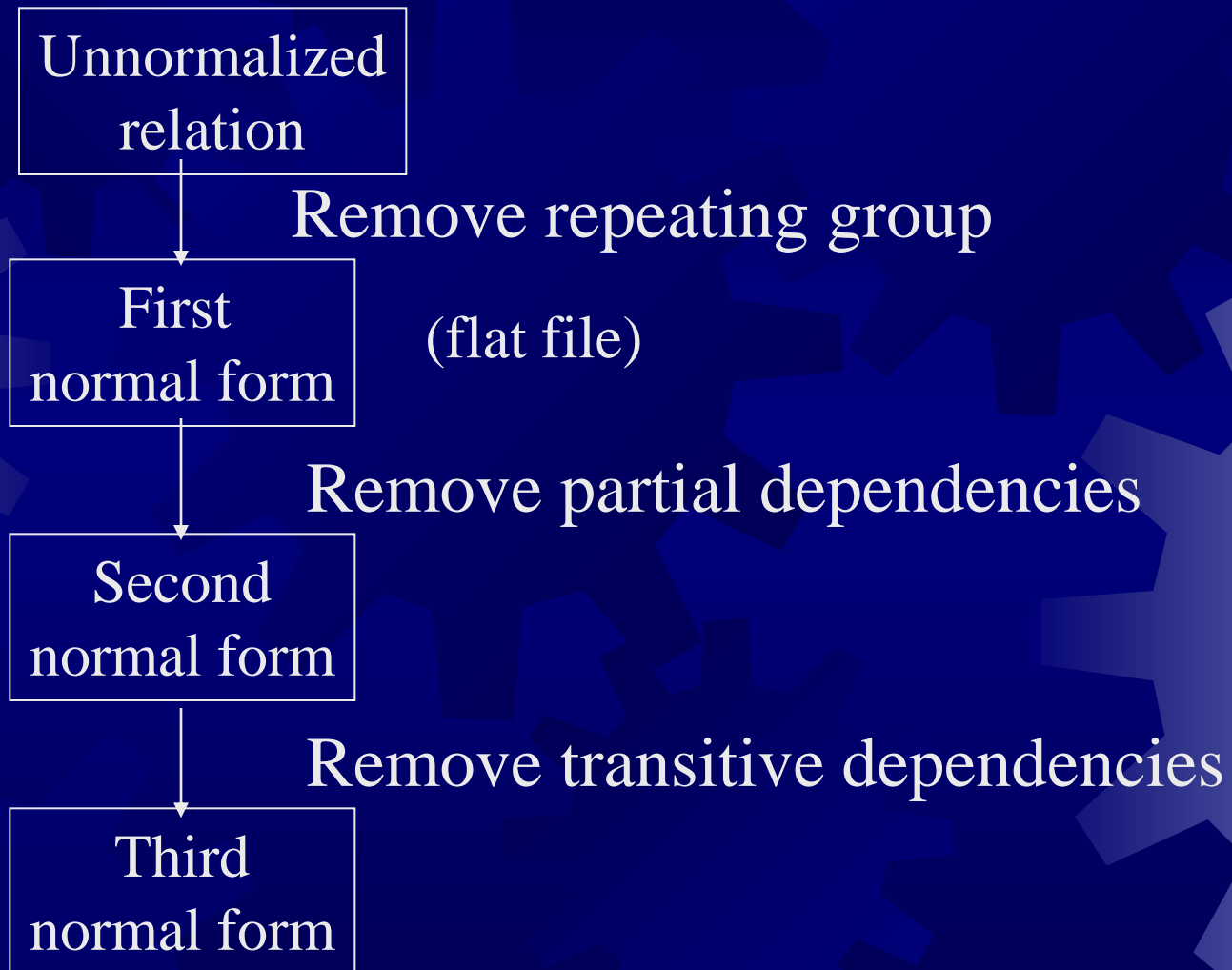
Deletion anomaly

刪掉某一 order 會把某零件的成本跟著刪去.

Update anomaly

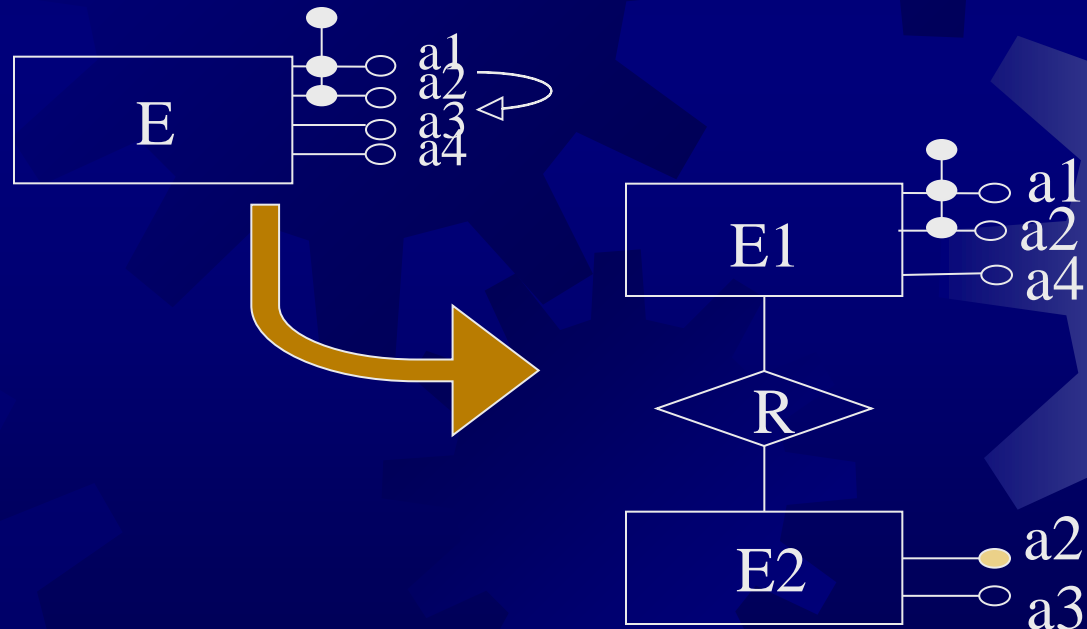
修改某零件成本時, 在 order entity 中會有數個相同的 data instance 要跟著修改.

Steps in Normalization

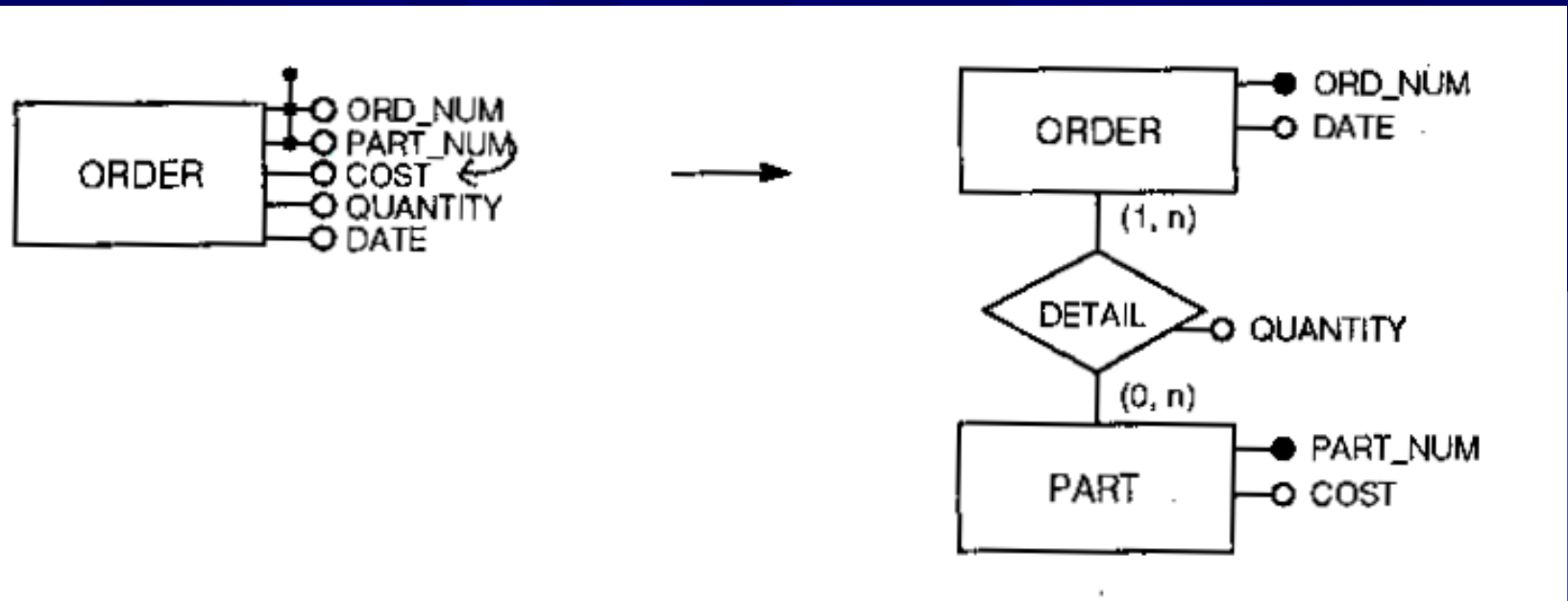


Second normal form

- ✦ It is in first normal form
- ✦ No partial dependencies
 - There exists nonprime attribute depends on part of identifier.
 - Nonprime attribute is the attribute which does not belong to identifier.

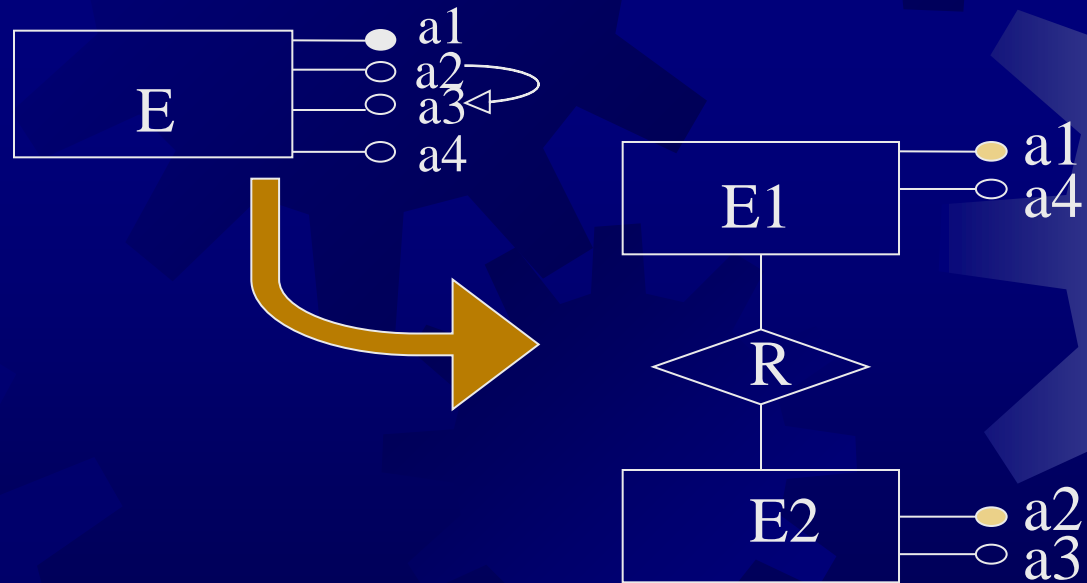


An example of achieving second normal form

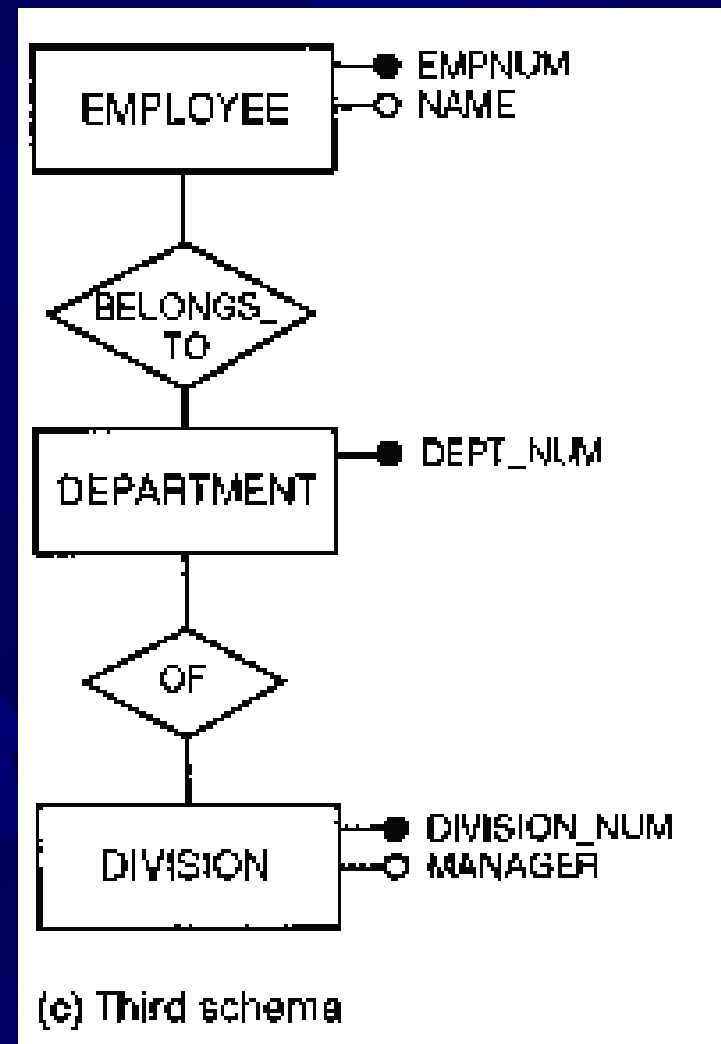
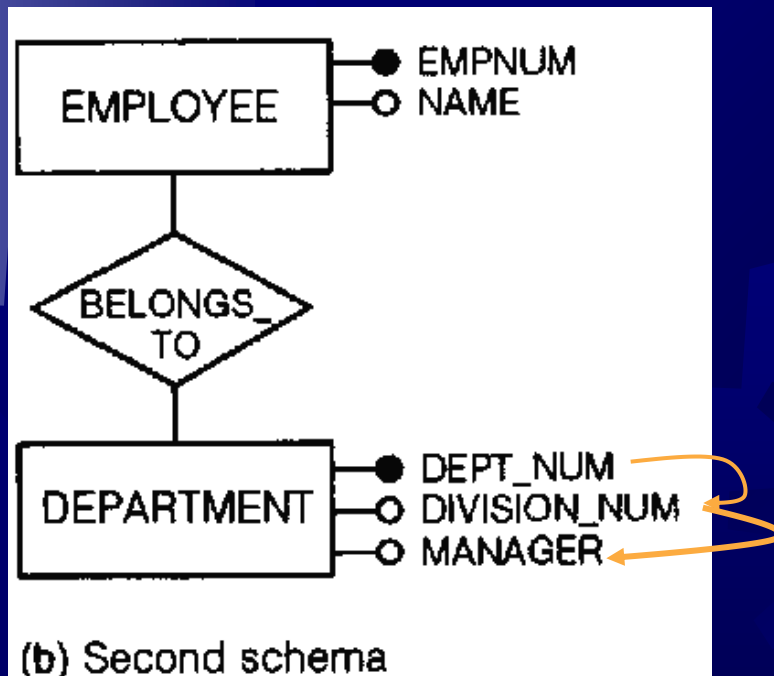
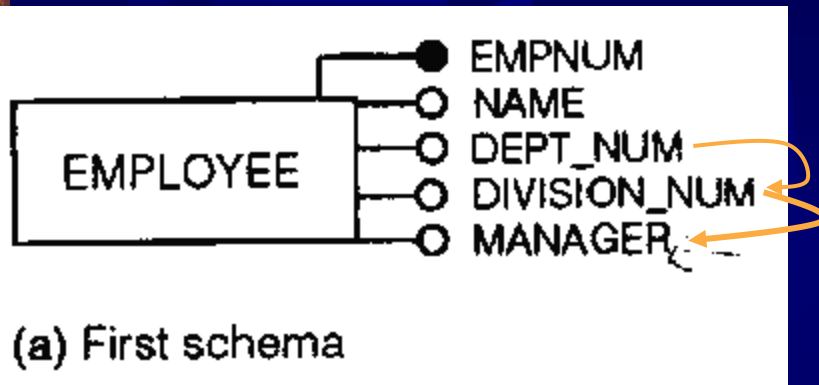


Third normal form

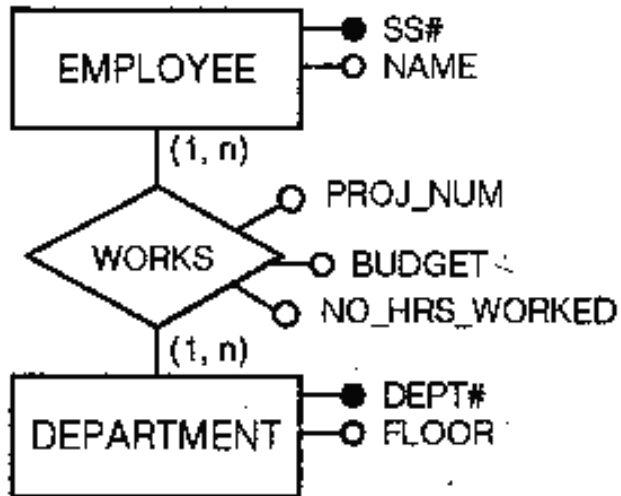
- ✦ It is in second normal form
- ✦ No transitive dependencies
 - Transitive dependency 指 nonprime 之間有功能相依的關係存在。



An example of achieving third normal form

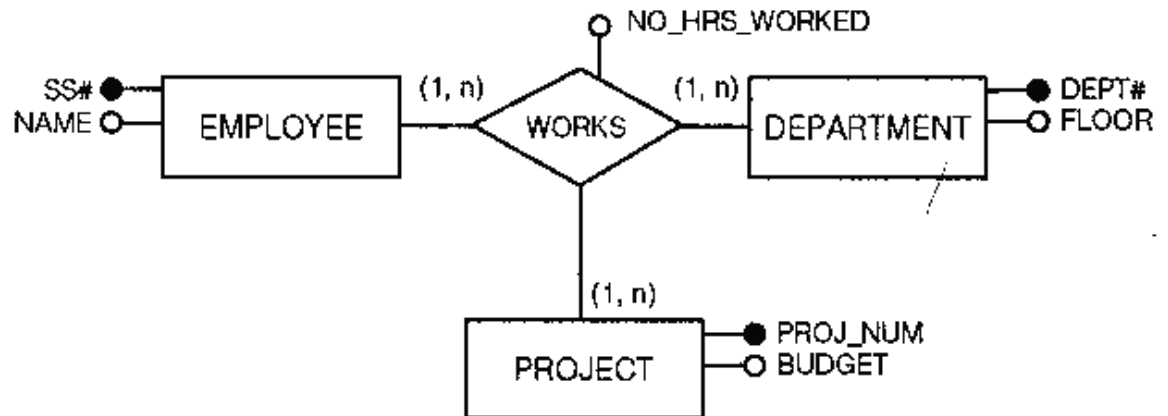


Another example of achieving third normal form



(a) First schema

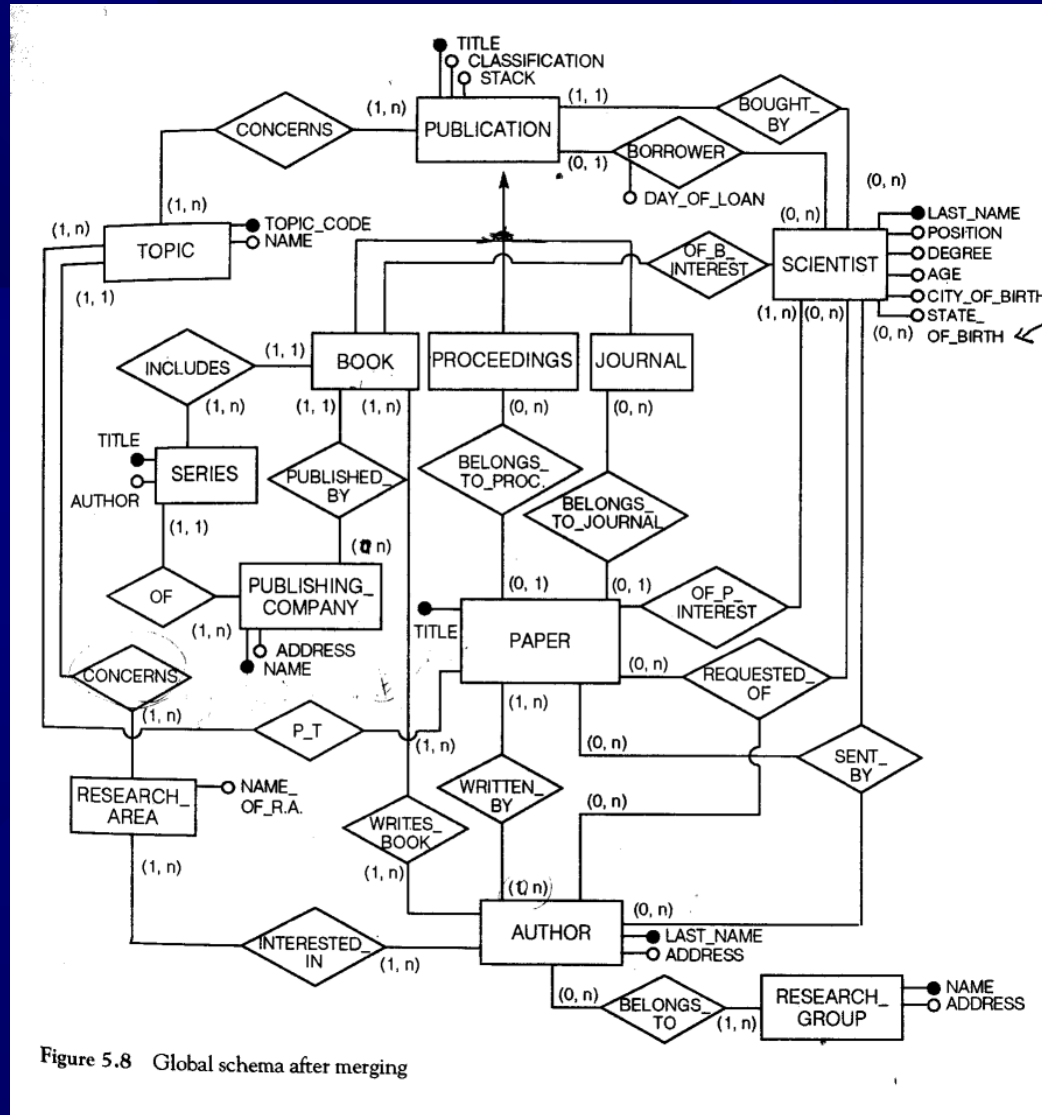
SS#, DEPT# → BUDGET
SS#, DEPT# → PROJ_NUM
SS#, DEPT# → NO_HRS_WORKED
PROJ_NUM → BUDGET



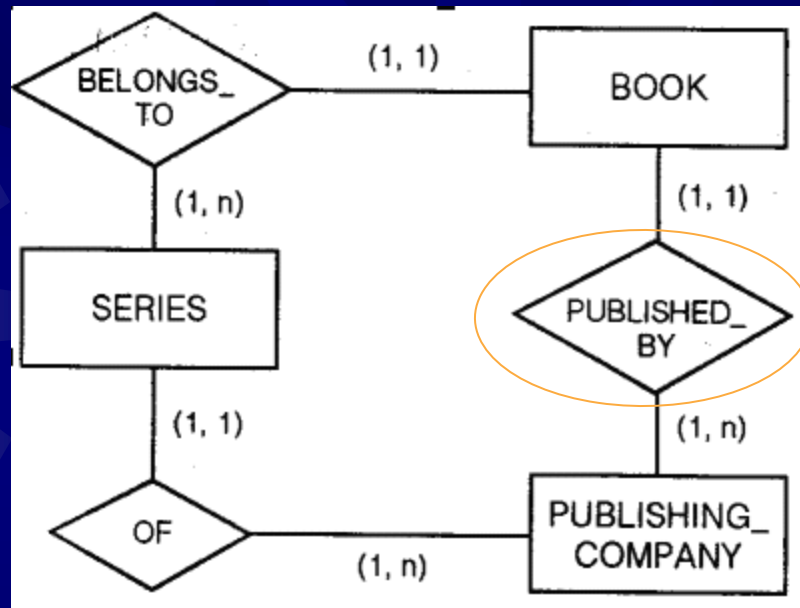
(b) Second schema

An Example of Schema Restructuring

Starting from the global schema in unit 6 page 34



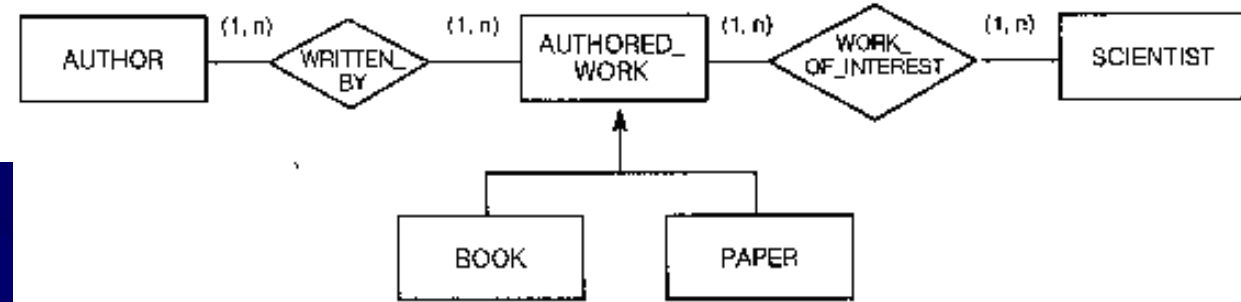
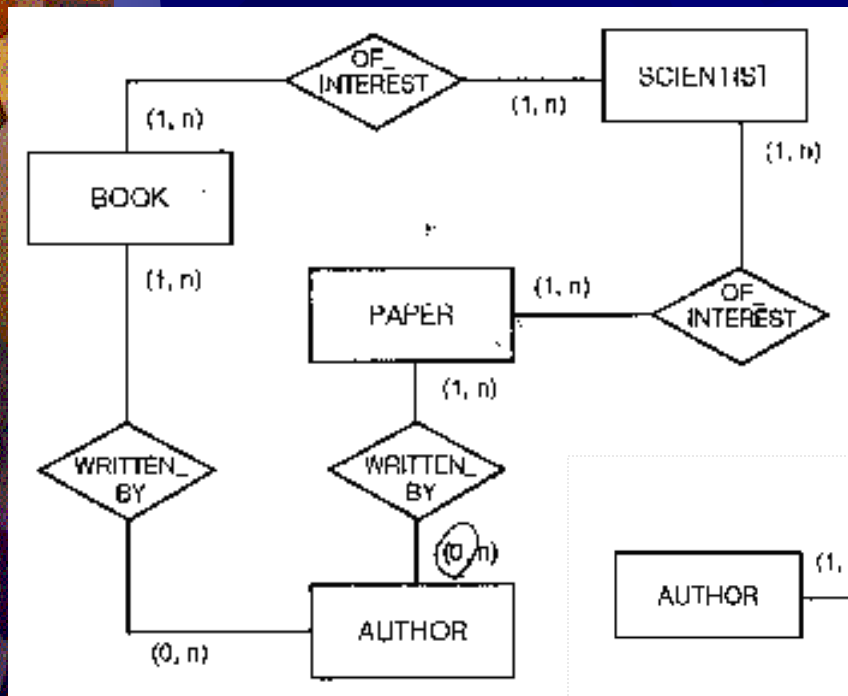
(1) Minimality



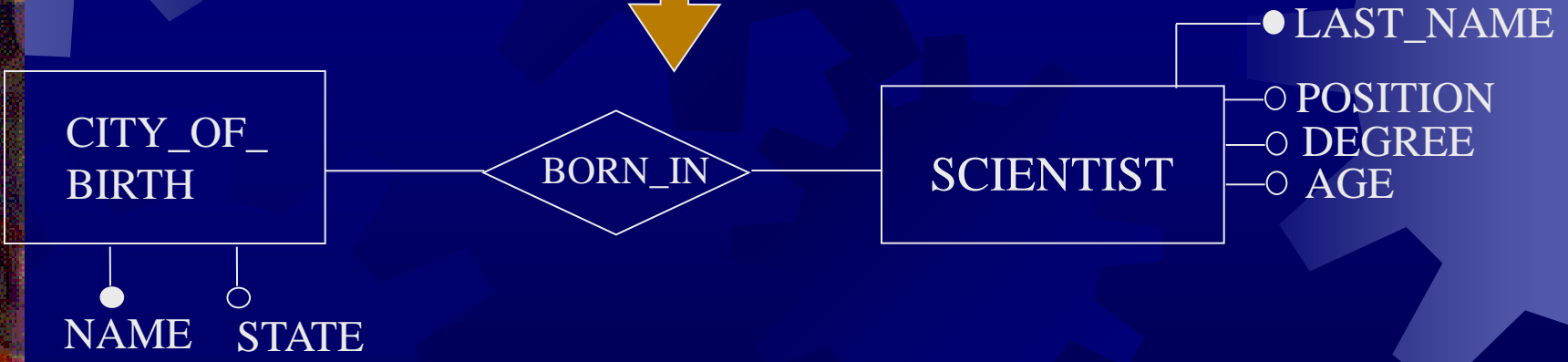
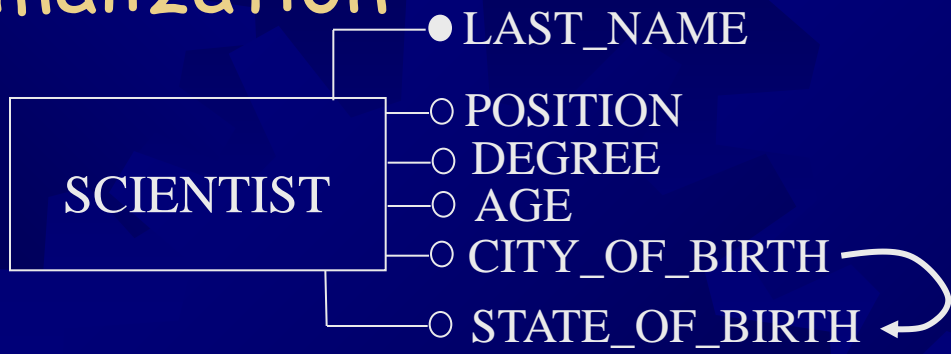
redundancy

- ✦ 把 PUBLISHING_BY 去掉, 會除去 redundancy, 從 BELONGS_TO 和 OF 找到.
- ✦ 但是若經常要找 BOOK 是哪一家書店出版的就較麻煩
- ✦ 延到 logical design 再決定.

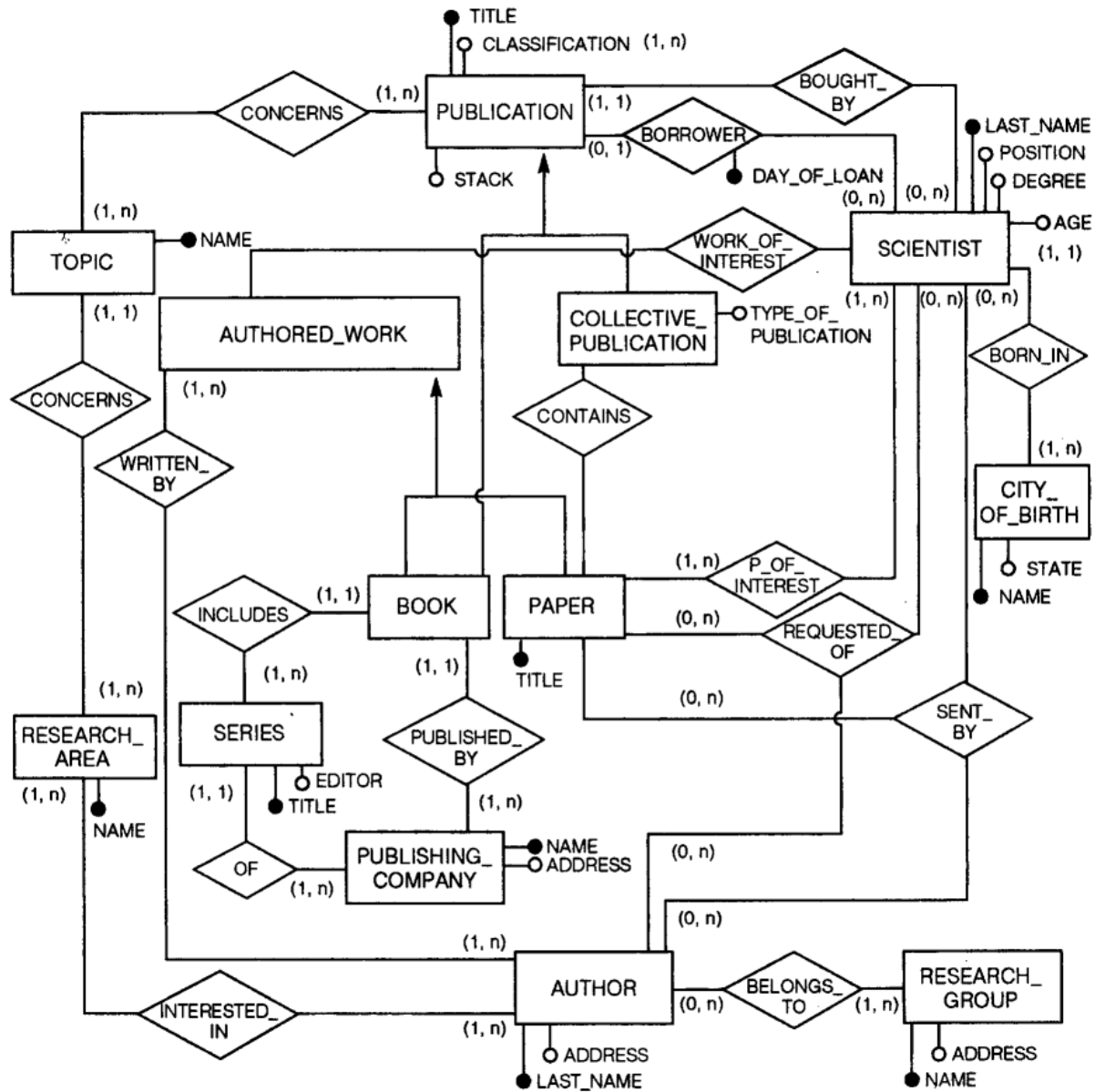
(2) Expressive and Self-explanation



(3) Normalization

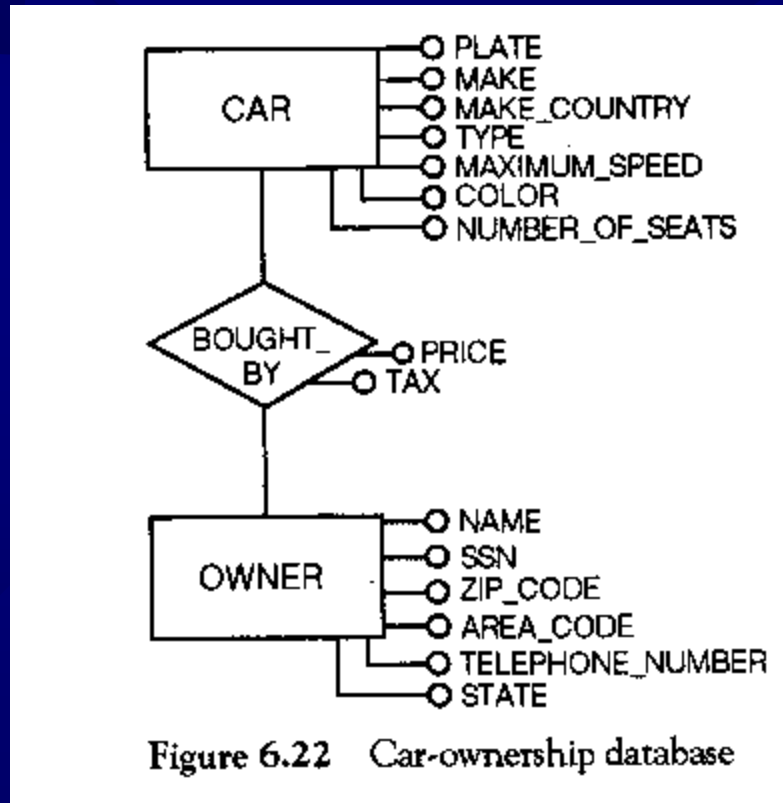


Global schema after reconstructing



Exercises

1. Check the normality of the following schema. Make any reasonable assumptions you need to make and state them.



- Check the redundancy and normality of the following schema. Make any reasonable assumptions you need to make and state them.

