CSE 344 Introduction to Data Management

Section 7: Conceptual Design

TA: Aloka Krishnan (alokak@uw.edu)

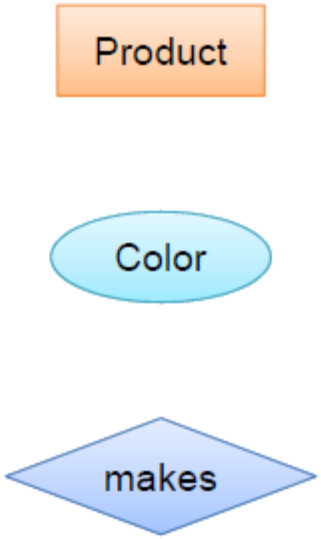
- Announcements
  - HW5 due tonight at 11.59pm
  - Please check your available late days
  - HW6 will be posted tonight
- Today
  - XML to E/R Diagram
  - E/R Diagram to Relation
  - BCNF

CONTEXT-SPECIFIC (DTD-BASED) MAPPING

```
<!DOCTYPE bib [
    <!ELEMENT bib  (book* )>
    <!ELEMENT book  (title,  (author+ | editor+ ), publisher?, price )>
    <!ATTLIST book  year CDATA  #REQUIRED >
    <!ELEMENT author  (last, first )>
    <!ELEMENT editor  (last, first, affiliation )>
    <!ELEMENT title  (#PCDATA )>
    <!ELEMENT last  (#PCDATA )>
    <!ELEMENT first  (#PCDATA )>
    <!ELEMENT affiliation  (#PCDATA )>
    <!ELEMENT publisher  (#PCDATA )>
    <!ELEMENT price  (#PCDATA )>
]>
```

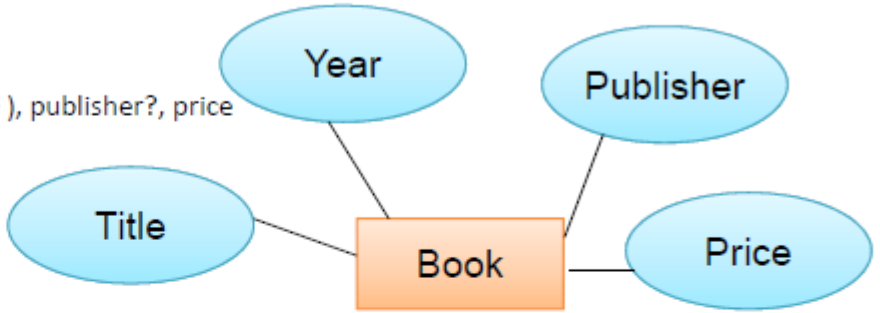
(Recall) Entity / Relationship Diagrams

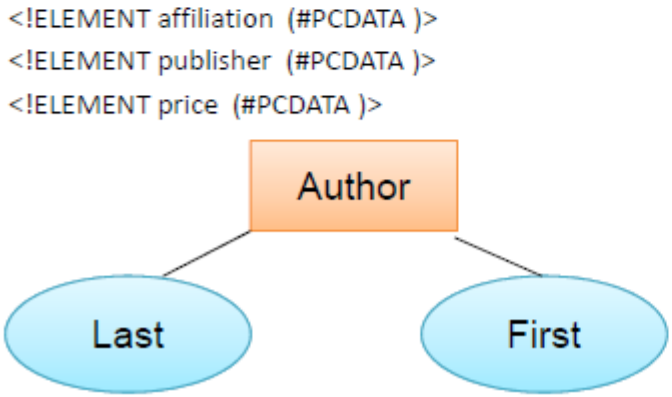
- Entity set = a class
  - An entity = an object
- Attribute
- Relationship

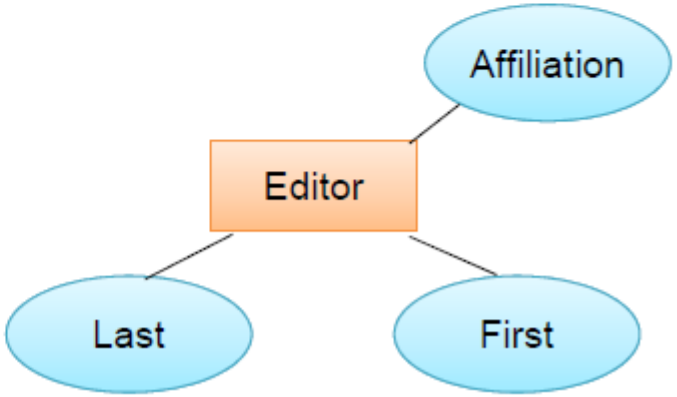


What Kinds of Entities?

```
<!DOCTYPE bib [
        <!ELEMENT bib  (book* )>
        <!ELEMENT book  (title,  (author+ | editor+ ), publisher?, price
)>
        <!ATTLIST book  year CDATA  #REQUIRED >
        <!ELEMENT author  (last, first )>
        <!ELEMENT editor  (last, first, affiliation )>
        <!ELEMENT title  (#PCDATA )>
        <!ELEMENT last  (#PCDATA )>
        <!ELEMENT first  (#PCDATA )>
        <!ELEMENT affiliation  (#PCDATA )>
        <!ELEMENT publisher  (#PCDATA )>
        <!ELEMENT price  (#PCDATA )>
]>
```

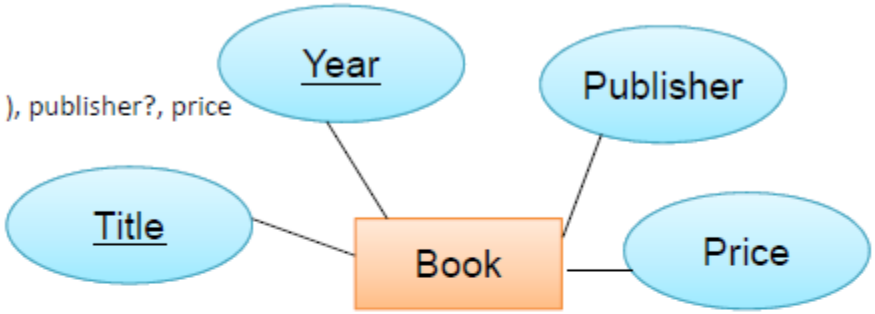


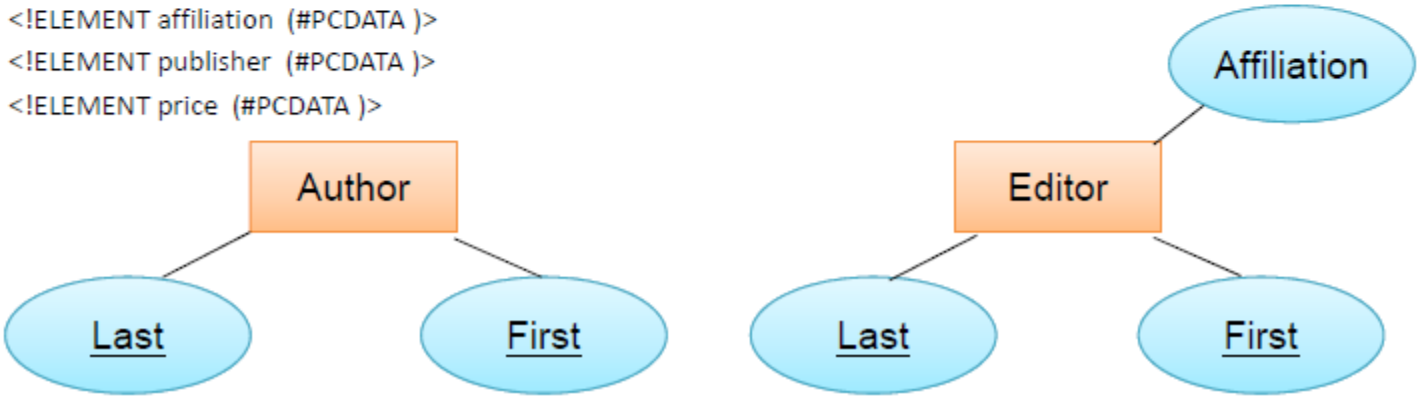




Primary Key in each Entity

```
<!DOCTYPE bib [
        <!ELEMENT bib  (book* )>
        <!ELEMENT book  (title,  (author+ | editor+ ), publisher?, price
)>
        <!ATTLIST book  year CDATA  #REQUIRED >
        <!ELEMENT author  (last, first )>
        <!ELEMENT editor  (last, first, affiliation )>
        <!ELEMENT title  (#PCDATA )>
        <!ELEMENT last  (#PCDATA )>
        <!ELEMENT first  (#PCDATA )>
        <!ELEMENT affiliation  (#PCDATA )>
        <!ELEMENT publisher  (#PCDATA )>
        <!ELEMENT price  (#PCDATA )>
]>
```



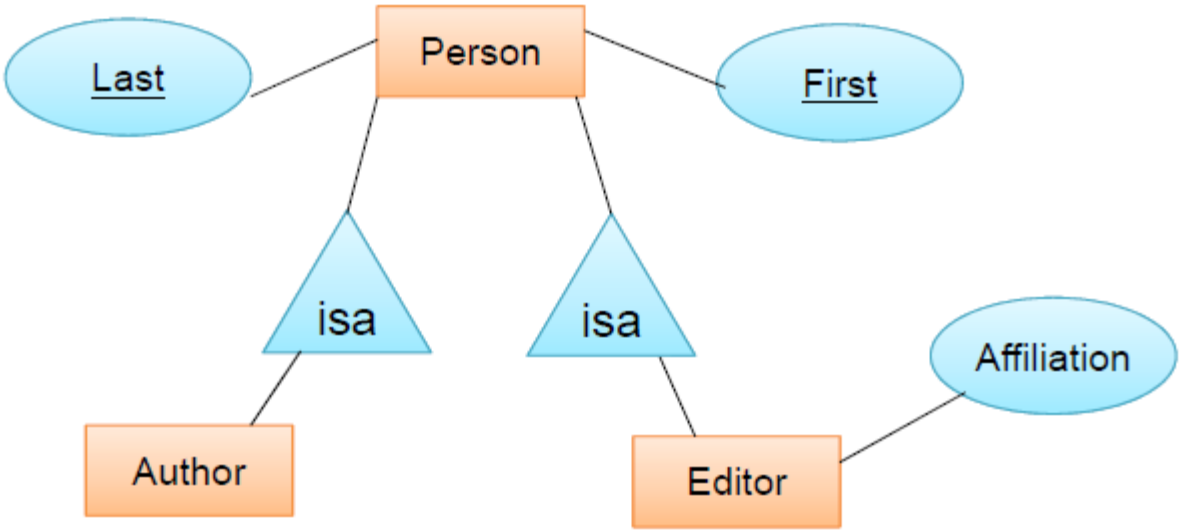


What's wrong with this model?

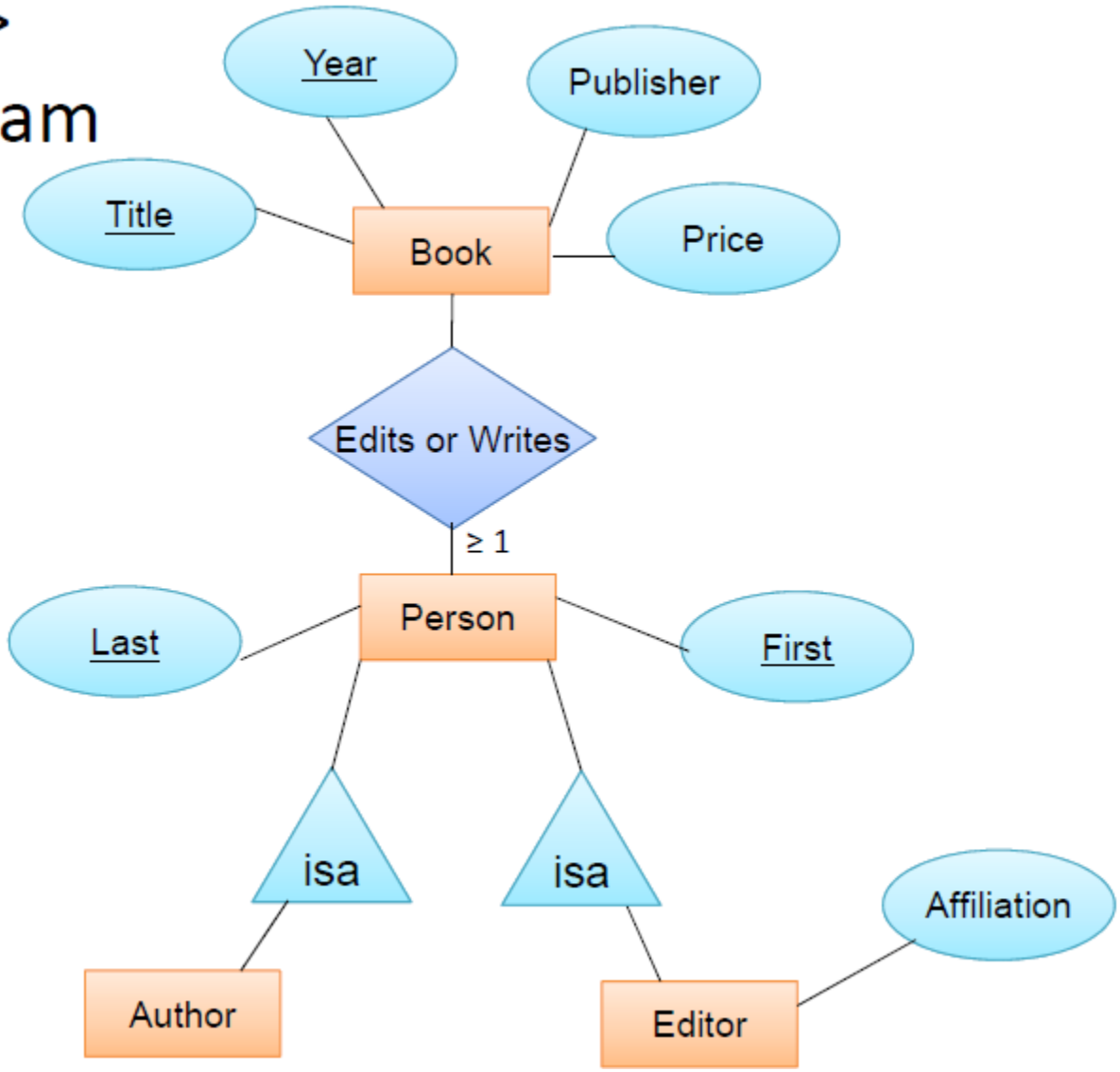
What's wrong with this model?

- In the DTD, a book can have authors, or editors, but not both.
- Perhaps, we can we combine Author and Editor.
  - Define an entity set Person with attributes last and first.
  - Add is-a relationship such that Author is-a Person, and Editor is-a Person.

Modeling UnionTypes With Subclasses



XML -> E/R Diagram



E/R Diagram -> Relation

```
CREATE TABLE Person(
        first CHAR(20),
        last CHAR(20),
 PRIMARY KEY (first, last)
)
```

```
CREATE TABLE Book(
        title CHAR(30),
        year INT,
        publisher CHAR(30),
        price FLOAT,
PRIMARY KEY (title, year)
)
```

```
CREATE TABLE Author(
        first CHAR(20),
        last CHAR(20),
PRIMARY KEY (first, last),
FOREIGN KEY (first, last)
        REFERENCES  Person
)
```

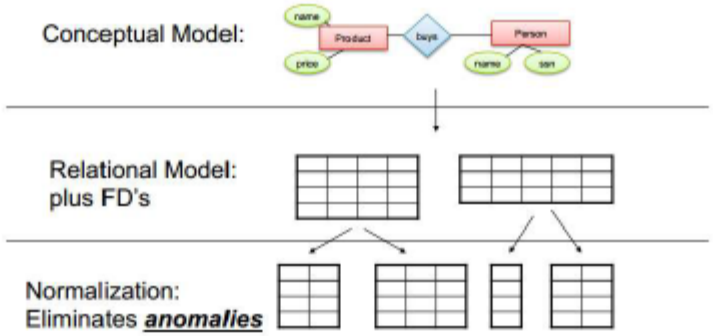
```
CREATE TABLE EditsOrWrites(
        first CHAR(20),
        last CHAR(20),
        title CHAR(30),
        year INT,
PRIMARY KEY (first, last, title, year),
FOREIGN KEY (first, last)
        REFERENCES  Person,
FOREIGN KEY (title, year)
        REFERENCES  Book
)
```

```
CREATE TABLE Editor(
        first CHAR(20),
        last CHAR(20),
        affiliation CHAR(20),
PRIMARY KEY (first, last),
FOREIGN KEY (first, last)
        REFERENCES  Person
)
```

Conceptual Design

Normal forms and functional dependencies:

- Anomalies(redundancy, update/deletion anomalies), functional dependencies, attribute closures, BCNF decomposition



- The BCNF (Boyce-Codd Normal Form) ---- A relation R is in BCNF if every set of attributes is either a superkey or its closure is the same set.

Example 1

Relation R(A,B,C,D,E,F) and functional dependencies:

A → BC and D → AF

Decompose R into BCNF.

Example 1 -- Solution

Relation R(A,B,C,D,E,F) and FD's A → BC and D → AF

A→BC violates BCNF since A+ = ABC ≠ ABCDEF. So we split R into R1(ABC) and R2(ADEF).

The only non-trivial FD in R1 is A→BC, and A+ = ABC, so R1 is in BCNF.

R2 has a non-trivial dependency D→AF that violates BCNF because D+ = ADF ≠ ADEF. So we split R2 into R21(DAF) and R22(DE). Both of these are in BCNF since they have no non-trivial dependencies that are not superkeys.

Example 2

The relation is R (A, B, C, D, E) and the FDs :

A -> E, BC -> A, and DE -> B

Decompose R into BCNF.

Example 2 – solution 1

The relation is R (A, B, C, D, E) and the FDs :

A -> E, BC -> A, and DE -> B

Notice that {A}+ = {A,E}, violating the BCNF condition. We split R to R_1(A,E) and R_2(A,B,C,D).

R_1 satisfies BCNF now, but R_2 not because of: {B,C}+ = {B,C,A}. Notice that the fd D E -> B has now disappeared and we don't need to consider it! Split R_2 to: R_2A(B,C,A) and R_2B(B,C,D).

Example 2 – solution 2

The relation is R (A, B, C, D, E) and the FDs :

A -> E, BC -> A, and DE -> B

Can we split differently? Let's try with the violation {B,C}+ = {B,C,A,E}. We initially split to R_1(B,C,A,E) and R_2(B,C,D). Now we need to resolve for R_1 the violation {A}+ = {A,E}. So we split again R_1 to R_1A(A,E) and R_1B(A,B,C). The same!

We can also start splitting by considering the BCNF violation {D,E}+ = {D,E,B}. Which is the resulting BCNF decomposition in this case? (it will be a different one)