

CSDS 433 Database Systems

Assignment 2

Spring 2022

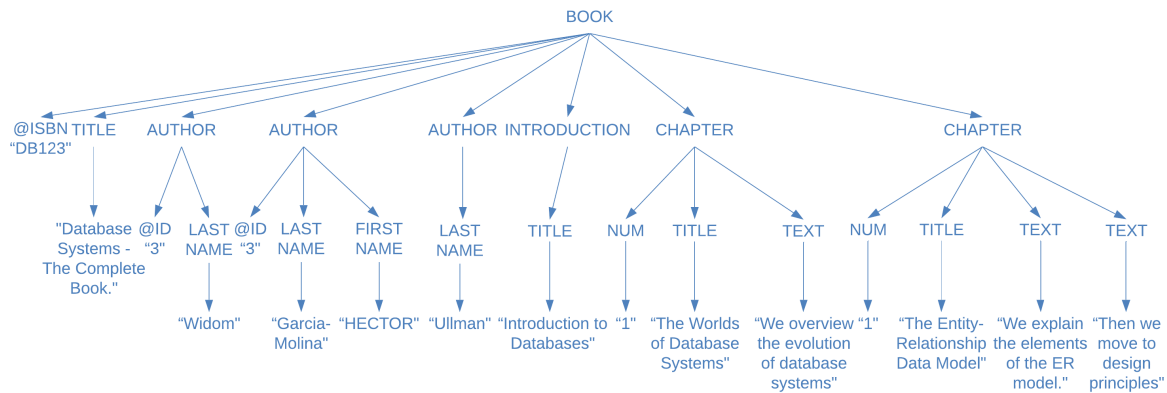
1. [XML] (30) Consider the following snippet of an XML document.

```
<BOOK ISBN="DB123">
<TITLE>Database Systems - The Complete Book.</TITLE>
<AUTHOR ID="3"><LASTNAME>Widom</LASTNAME></AUTHOR>
<AUTHOR ID="1">
  <LASTNAME>Garcia-Molina</LASTNAME>
  <FIRSTNAME>HECTOR</FIRSTNAME>
</AUTHOR>
<AUTHOR>
  <LASTNAME>Ullman</LASTNAME>
</AUTHOR>
<INTRODUCTION>
  <TITLE>Introduction to Databases</TITLE>
</INTRODUCTION>
<CHAPTER>
  <NUM>1</NUM>
  <TITLE>The Worlds of Database Systems </TITLE>
  <TEXT>We overview the evolution of database systems. </TEXT>
</CHAPTER>
<CHAPTER>
  <NUM>1</NUM>
  <TITLE>The Entity-Relationship Data Model</TITLE>
  <TEXT>We explain the elements of the ER model.</TEXT>
  <TEXT>Then we move to design principles</TEXT>
</CHAPTER>
</BOOK>
```

- (a) Is this fraction of the XML document well-formed (e.g., tags are properly nested)? If not, Identify the mistakes and change them to make the XML well-formed.

- line 3: Opening and ending tag mismatch: LASTNAME
- line 4: Attribute Value: id should be unique, and " or ' expected
- line 5: Opening and ending tag mismatch: LASTNAME
- line 15: / expected: NUM
- line 17: Opening and ending tag mismatch: TEXT

- (b) Draw the XML tree representation of the modified, well-formed XML from (a).



(c) Consider the DTD on the right side.

Test if your modified XML from part (a) is valid for the given DTD? If not, describe how to **minimally** change the XML to fit the DTD.

```
<!DOCTYPE BOOK [
<ELEMENT BOOK (TITLE, AUTHOR?, INTRODUCTION?,CHAPTER+)>
<ELEMENT AUTHOR (LASTNAME?, FIRSTNAME?)>
<!ATTLIST BOOK ISBN CDATA #REQUIRED>
<!ATTLIST AUTHOR ID #REQUIRED>
<ELEMENT TITLE (#PCDATA)>
<ELEMENT INTRODUCTION (TITLE, TEXT)>
<ELEMENT CHAPTER (NUM, TITLE, TEXT)>
<ELEMENT NUM (#PCDATA)>
<ELEMENT TEXT (#PCDATA)>
<ELEMENT LASTNAME (#PCDATA)>
<ELEMENT FIRSTNAME (#PCDATA)>]>
```

- Delete the AUTHOR sub element with LASTNAME of Ullman and one of the 2 authors left. (Line 2: "AUTHOR?" and Line 5: "AUTHOR ID #REQUIRED")
- Add a TEXT sub element of INTRODUCTION sub-element with title of "Introduction to Database" after line 12. (Line 7: "INTRODUCTION (TITLE, TEXT)")
- Delete one of the TEXT in CHAPTER whose TITLE is The Entity-Relationship Data Model. (Line 8: "CHAPTER (NUM, TITLE, TEXT)")

2. [RDF] (40) Consider the following RDF document using the XML syntax. Draw the equivalent graph representation, *for the RDF data* (in bold). The blue part is the RDF schema for your reference. You may simplify the node and edge labels (no need to put the entire URI) and add RDF nodes/edges as needed.

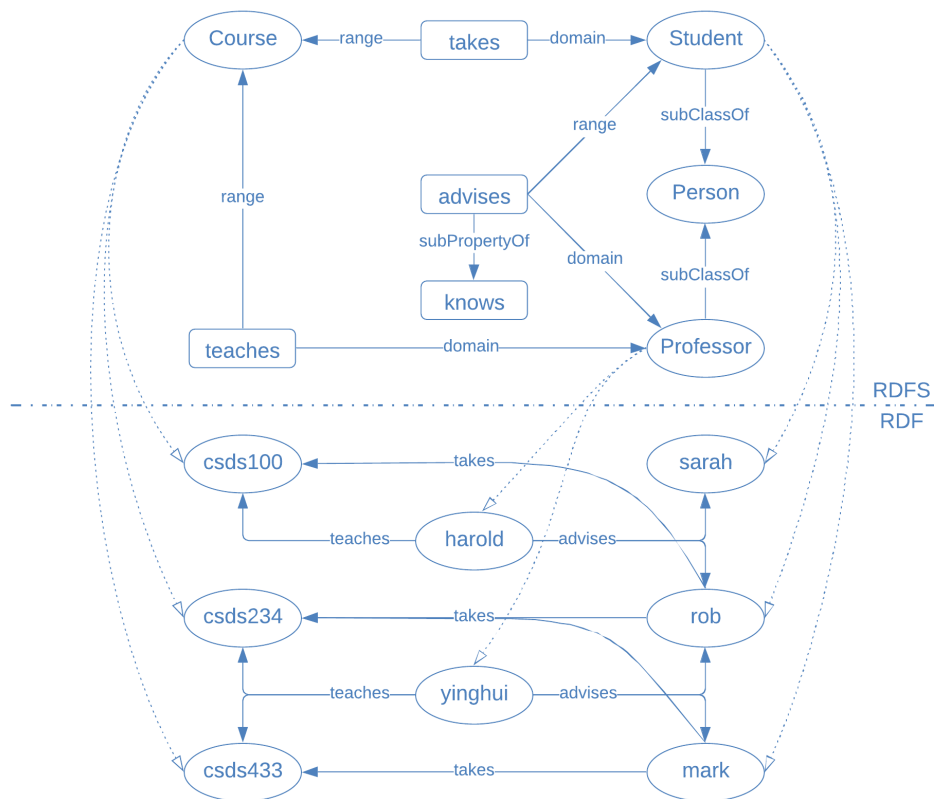
```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:u="http://www.example.org/uni#" xml:base="http://www.example.org/uni">
  <rdfs:Class rdf:ID="Person" />
  <rdfs:Class rdf:ID="Student">
    <rdfs:subClassOf="#Person" />
  </rdfs:Class>
  <rdfs:Class rdf:ID="Professor">
    <rdfs:subClassOf="#Person" />
  </rdfs:Class>
  <rdfs:Class rdf:ID="Course" />
  <rdf:Property rdf:ID="advises">
    <rdfs:domain rdf:resource="#Professor" />
    <rdfs:range rdf:resource="#Student" />
    <rdfs:subPropertyOf="#knows">
  </rdf:Property>
  <rdf:Property rdf:ID="takes">
    <rdfs:domain rdf:resource="#Student" />
    <rdfs:range rdf:resource="#Course" />
  </rdf:Property>
  <rdf:Property rdf:ID="teaches">
    <rdfs:domain rdf:resource="#Professor" />
    <rdfs:range rdf:resource="#Course" />
  </rdf:Property>
  <rdf:Property rdf:ID="knows" />

  <u:Professor rdf:ID="harold">
    <u:teaches rdf:resource="#csds100" />
    <u:advises rdf:resource="#rob" />
    <u:advises rdf:resource="#sarah" />
  </u:Professor>

  <u:Professor rdf:ID="yinghui">
    <u:teaches rdf:resource="#csds433" />
    <u:teaches rdf:resource="#csds234" />
    <u:advises rdf:resource="#mark" />
    <u:advises rdf:resource="#rob" />
  </u:Professor>

  <u:Student rdf:ID="rob">
    <u:takes rdf:resource="#csds100" />
    <u:takes rdf:resource="#csds234" />
  </u:Student>

  <u:Student rdf:ID="mark">
    <u:takes rdf:resource="#csds433" />
    <u:takes rdf:resource="#csds234" />
  </u:Student>
  ...
</rdf:RDF>
```



(For this question, a complete RDF part is enough.)

3. [Record Storage] (30)

(a) Consider the following schema:

Purchase(*Custname*, *amount*, *date*, *time*) where *Custname*, *amount*, *date* and *time* are character strings of length 15, integers of 2 bytes, 7 byte date type, and 8 byte time type, respectively. Assume we represent a tuple as fixed-length record. How many bytes does the record take if: (i) fields can start at any byte; (ii) fields must start at a byte that is a multiple of 4, and (iii) fields must start at a byte that is a multiple of 8?

(i) $15 + 2 + 7 + 8 = 32$ bytes

(ii) $16 + 4 + 8 + 8 = 36$ bytes

(The size of each field should be a multiple of 4: $15 \rightarrow 16$, $2 \rightarrow 4$, $7 \rightarrow 8$, $8 \rightarrow 8$)

(iii) $16 + 8 + 8 + 8 = 40$ bytes

(The size of each field should be a multiple of 8: $15 \rightarrow 16$, $2 \rightarrow 8$, $7 \rightarrow 8$, $8 \rightarrow 8$)

(b) Now we extend the Purchase schema (containing the four fixed-length field) with three variable-length fields: *tid*, *address*, and *product*. If pointers within a record require 4 bytes, a record length is a 4-byte integer, and we need to use a record header for record length and pointers, how many bytes in total, exclusive of the space needed for the variable-length fields, are needed for the record? (Assume fields

can start at any byte).

(i) $32 + 4 \times 2 + 4 = 44$ bytes

(Add a record header: 2 pointers and record length.)

(c) Recompute the bytes needed for the extended record in (b) under the conditions (ii) and (iii).

(ii) $36 + 4 \times 2 + 4 = 48$ bytes

(Add a record header: 2 pointers and record length, the storage size for each pointer and record length should be a multiple of 4.)

(iii) $40 + 8 \times 2 + 8 = 64$ bytes

(Add a record header: 2 pointers and record length, the storage size for each pointer and record length should be a multiple of 8 ($4 \rightarrow 8$).)
