

Question 4: 1: (1) <option> Air Bag </option> (2) <color>blue</color> (3) <model> Voxy </option>

<color> blue </color>

Tourist </special>

<color> red </color>

</model>

2. <!DOCTYPE cars [

<ELEMENT cars (car)*>

<!ELEMENT car (maker*, year*, model*, option*, color*)>

<!ELEMENT maker>

<!ELEMENT year>

<!ELEMENT model (special*)>

<!ELEMENT special (option*)>

<!ELEMENT option>

<!ELEMENT color>

<!ELEMENT grade (#CDATA #REQUIRED>

3.

4.

XML allows user to define new collections of tags that can be used to structure any type of data or
information that user wishes to transmit.

However, XML does not prevent a user from designing tags that enable the display of the
data in Web browser.

Information Organization Final Examination (January 25th, 2012)

Notice

- Write your name and student number to all the answer sheets.
- You can bring printed materials (slides, notes, memos, etc). But NO electronic device, such as cell phone, laptop, electronic dictionary, or calculator is allowed.
- You can write answers either in English or Japanese.

Question 1.

1. Suppose that user John executed the following SQL commands.

John: CREATE TABLE lunch(name, price);
REVOKE ALL PRIVILEGES ON lunch FROM PUBLIC;
GRANT SELECT, INSERT ON lunch TO Tom WITH GRANT OPTION;
GRANT SELECT, DELETE ON lunch TO Keiko;

Then, the following commands (1)-(4) were executed in this order, by the users indicated at the front of each command line.

- (1) Tom: GRANT SELECT ON lunch TO Hanako;
- (2) Keiko: INSERT INTO lunch(name, price) VALUES ('bento', '500');
- (3) John: REVOKE SELECT ON lunch FROM Tom;
- (4) Hanako: SELECT * FROM lunch;

For each of (1)-(4), answer whether each command was granted or denied by the database. Also, describe reasons why.

2. What is the name of the access control model in which users can have privileges on objects and give privileges to other users?

Question 2.

Let us consider multi-granularity lock modes consisting of IS, IX, S, and X. Here IS is intension shared lock, IX is intension exclusive lock, S is shared lock, and X is exclusive lock. Multi-granularity locking is applied on a hierarchy of nodes having containment relationships. Examples of nodes are table, index, and tuples.

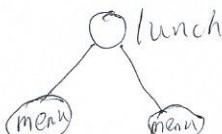
1. To lock a tuple in X mode, what kinds of locks are necessary before locking the tuple? Also explain why.
2. To unlock a tuple locked in X mode, what kind of unlock order must be followed? Also explain why.

Suppose that transaction T1 uses the index to read only part of table R. For simplicity, assume that the index height is one. To access R, T1 uses the following locks: T1 gets an IS lock on R, gets an IS lock on an index page, and repeatedly gets an S lock on tuples of R.

3. Describe an appropriate lock schedule for transaction T2: T2 updates all the tuples of R.
4. Argue whether executions of T1 and T2 can be overlapped.
5. Describe an appropriate lock schedule for transaction T3: T3 reads a part of R using the index, and updates a few of them.
6. Argue whether executions of T1 and T3 can be overlapped.

Question 3. Consider the following XML document d_1 :

```
<lunch>
  <menu name="ramen" type="noodle">
    <price> 500 </price>
    <options>
      <option price="100"> large_noodle </option>
      <option price="200"> extra_meat </option>
    </options>
  </menu>
  <menu name="udon" type="noodle">
    <price>300</price>
    <options>
      <option price="150">large_noodle</option>
    </options>
  </menu>
</lunch>
```



index

1/2

/lunch/*[text()='ramen']

1. Write the answers to the following XPath queries applied on d_1 .

属性， $\textcircled{2}$ = ']
A ; B 带有 B 属性的 A.

- (a) `//*[@price > "100"]/text()`
 - (b) `options/ancestor::*/price`
 - (c) `*[preceding-sibling::*/attribute::price]`
2. Write an XPath query which returns the following:
- (a) Find 'ramen' elements such that all of their 'option' descendants have 'price' larger than 100.
 - (b) Find elements such that whose name is either 'menu' or 'option' and whose attribute value or text value is 'noodle'.
 - (c) Find elements that have a sibling which has an attribute named 'special'.

3. Write a DTD e such that document d_1 is valid against e .
4. Write a DTD f such that document d_1 is also valid against f , and there is another document d_2 such that d_2 is valid against e but not valid against f . Show such DTD f and document d_2 .

Question 4. Explain the following notions.

- 1. User authentication for web sites.
- 2. RW-, WR-, and WW-conflicts.
- 3. Optimistic concurrency control.

d_2 支持所有
属性节点。

d_2 支持 ATTLIST 语句。

1 *

Information Organization Final Examination (January 25th, 2012)

Notice

- Write your name and student number to all the answer sheets.
- You can bring printed materials (slides, notes, memos, etc). But NO electronic device, such as cell phone, laptop, electronic dictionary, or calculator is allowed.
- You can write answers either in English or Japanese.

Question 1. 25'

1. Suppose that user John executed the following SQL commands:

```
John: CREATE TABLE lunch(name, price),  
      REVOKE ALL PRIVILEGES ON lunch FROM PUBLIC;  
      GRANT SELECT, INSERT ON lunch TO Tom WITH GRANT OPTION;  
      GRANT SELECT, DELETE ON lunch TO Keiko;
```

Then, the following commands (1)-(4) were executed in this order, by the users indicated at the front of each command line.

- (1) Tom: GRANT SELECT ON lunch TO Hanako;
- (2) Keiko: INSERT INTO lunch(name, price) VALUES('bento', '500');
- (3) John: REVOKE SELECT ON lunch FROM Tom;
- (4) Hanako: SELECT * FROM lunch;

For each of (1)-(4), answer whether each command was granted or denied by the database. Also, describe reasons why.

2. What is the name of the access control model in which users can have privileges on objects and give privileges to other users?

Question 2. 30'

Let us consider multi-granularity lock modes consisting of IS, IX, S, and X. Here IS is intension shared lock, IX is intension-exclusive lock, S is shared lock, and X is exclusive lock. Multi-granularity locking is applied on a hierarchy of nodes having containment relationships. Examples of nodes are table, index, and tuples.

1. To lock a tuple in X mode, what kinds of locks are necessary before locking the tuple? Also explain why.

2. To unlock a tuple locked in X mode, what kind of unlock order must be followed? Also explain why.

3. Suppose that transaction T1 uses the index to read only part of table R. For simplicity, assume that the index height is one. To access R, T1 uses the following locks: T1 gets an IS lock on R, gets an IS lock on an index page, and repeatedly gets an S lock on tuples of R.

IS(R) IS(index) S(tuple)

IX(R) X(tuple)

4. Describe an appropriate lock schedule for transaction T2: T2 updates all the tuples of R.

X → S conflicts

IX(R)

5. Argue whether executions of T1 and T2 can be overlapped.

6. Describe an appropriate lock schedule for transaction T3: T3 reads a part of R using the index, and updates a few of them.

7. Argue whether executions of T1 and T3 can be overlapped.

SIX(R) SIX(index) S(tuple) X(tuple)

Question 3. Consider the following XML document d1: 36

```
<lunch>  
  <menu name="ramen" type="noodle">  
    <price> 500 </price>  
    <options>  
      <option><option price="100"> large_noodle </option>  
      <option><option price="200"> extra_meat </option>  
    </options>  
  </menu>  
  <menu name="udon" type="noodle">  
    <price>300</price>  
    <options>  
      <option><option price="150">large_noodles </option>  
    </options>  
  </menu>  
</lunch>
```

table R

index

/+ C

/lunch/*[text(),
]

1. Write the answers to the following XPath queries applied on d1.

[]

(a) `//*[@@price > "100"]/text()`

(b) `options/ancestor::*[price]`

(c) `*[preceding-sibling::*[attribute::price]]`

2. Write an XPath query which returns the following:

(a) Find 'ramen' elements such that **all** of their 'option' descendants have 'price' larger than 100.

(b) Find elements such that whose name is either 'menu' or 'option' and whose attribute value or text value is 'noodle'.

(c) Find elements that have a sibling which has an attribute named 'special'.

3. Write a DTD e such that document d_1 is valid against e .

4. Write a DTD f such that document d_1 is also valid against f , and there is another document d_2 such that d_2 is valid against e but not valid against f . Show such DTD f and document d_2 .

Question 4. Explain the following notions. $5 \times 3 = 15'$

1. User authentication for web sites.

2. RW-, WR-, and WW-conflicts.

3. Optimistic concurrency control.

(1 *

解 答 用 紙

年 月 日 提出

科目/Subject	4	4	1	1	5	3	5	-	3	Score
担当教員/Lecturer	$1\frac{1}{2}$ 游									

Q1.

- (1) Granted. Because John granted Tom with the "WITH GRANT OPTION" on select on lunch, thus Tom is able to grant select on lunch to others
- (2) Denied. John only gave SELECT and DELETE privilege to Keiko, so she is not allowed to insert.

(3) Granted. John is the creator of the table, he can revoke freely.

(4) Denied. Hanako received SELECT privilege from Tom, whose privilege is revoked in (3), this removes Hanako's (solely) privilege

- Discretionary access control

Q2.

- Must have IX or SIX on parent node. Conflicts will happen when other transactions try to read/write that parent node if there is no lock held on it.
- From specific to general (i.e. bottom-up). Similar to the previous question, if we release the locks from up to bottom, problems will occur on ancestor nodes that are not locked.

1. must move in or six on parent node. Conflicts will happen when other transactions

try to read/write that parent node if there is no lock held on it.
2. From specific to general (i.e. bottom-up). Similar to the previous question, if we release the locks from up to bottom, problems will occur on ancestor nodes that are not locked.

3. T2 gets an IX on R then repeatedly gets an X lock on tuples of R

and update the tuple Get X on R
4. ~~cannot overlap~~ Cannot overlap because lock conflicts on R.

5. T3 gets SIX on R and SIX on index, then repeatedly get an S on tuple and occasionally upgrades to X on some tuples.

b. Can overlap. The locks are ~~not~~ conflict on parents.

Q3

1. (a) large ~~noodle~~

(b) no result

(c) no result

extra meat

(not starting from "/")

large noodle preceding siblings

2. (a) ~~/* []~~ @name=ramen [/options / option[@price>100]]
(b)
(c) /* [] sibling::*/ /attribute::special]

SIX(R)
SIX(index)
S (tuples)
X(tuples)

13

3. ex. <!ELEMENT lunch (menu*)>

<!ELEMENT menu (price, options+)> *price?*

<!ELEMENT price EMPTY> *EMPTY*

<!ELEMENT options (option*)> *3*

<!ELEMENT option EMPTY> *EMPTY*

d₁ is valid against the previous DTD.

4. f: <!ELEMENT lunch ANY>

<ATTLIST

<!ELEMENT menu ANY>

d₁ valid against f.

d₂ <lunch>

<menu> *name*

<price> 123 </price> *f*

<options>

<option> large </options> price = "100" > large </option>

</options>

</menu>

</lunch>

d₂ is valid against e but not f.

</options>

</menu>

</lunch>

d_2 is valid against e but not f.

Q4.

1. A website can ask user to use password ~~authentication~~. This is done after SSL is used to establish a session key so that the transmission of password is secure. But this is still at risk if the user's information is leaked.

Digital signatures can also be used. the user ~~will~~ encrypts the request using his private key and ~~the~~ encrypt using websites' public key. The website uses his private key and user's public key to get the original request. This makes it impossible to forge a fake request.

2. WR~~W~~ - conflicts : reading uncommitted date, also called "dirty reads"

RW - conflicts : unrepeatable reads, two reads ~~on~~ on the same object at different times return different results, thus ~~ever~~ called "unrepeatable"

WW - conflicts : overwriting uncommitted data, two transactions wrote the data on the same object while the other contains uncommitted data.

3. This is based on the assumption that conflicts are rare. Instead of using complex locking policies, we check for conflicts before transaction commits.

15

Information Organization

Exercises Part 2 (security and transaction, XML)

from the textbook by Ramakrishnan and Gehrke + α

Ppt Pg. P16 S-1 What is the main idea behind discretionary access control? What is the idea behind mandatory access control? What are the relative merits of these two approaches?

Ppt Pg. P17 S-2 What is a Trojan Horse attack and how can it compromise discretionary access control? Explain how mandatory access control protects against Trojan horse attacks.

Ppt Pg. S-3 Describe the privileges recognized in SQL. In particular, describe privileges regarding SELECT, INSERT, UPDATE, DELETE, and REFERENCES. For each privilege, indicate who acquires it automatically on a given table. Create.

Pt Pg. S-4 What is the difference between symmetric and public-key encryption? Give examples of well-known encryption algorithms of both kinds. What is the main weakness of symmetric encryption and how is this addressed in public key encryption?

Pt Pg. S-5 Explain the need for each of the following limits in a statistical database system:

1. A maximum on the number of queries a user can pose.
2. A minimum on the number of tuples involved in answering a query.
3. A maximum on the intersection of two queries (i.e., on the number of tuples that both queries examine).

T-1 Define these terms: atomicity, consistency, isolation, durability, schedule, dirty read, unrepeatable read, serializable schedule, cascading abort.

T-2 Describe Strict 2PL.

T-3 What is the phantom problem? Can it occur in a database where the set of database objects is fixed and only the values of objects can be changed?

T-4 Consider the following actions taken by transaction T1 on database objects X and Y:

$$R(X), W(X), R(Y), W(Y)$$

1. Give an example of another transaction T2 that, if run concurrently to transaction T1 without some form of concurrency control, could interfere with T1.

并发取消

T4

- 1) If the transaction T_2 performed $w(Y)$ before T_1 performed $R(X)$, and then T_2 aborted, the value read by T_1 would be invalid and the abort would be cascaded to T_1 .

T_1	T_2
$R(X)$	
$w(X)$	$w(Y)$
$R(Y)$	
$w(Y)$	

aborted

- 2) Strict 2PL would require T_2 to obtain an exclusive lock on Y before writing to it. This lock would have to be held until T_2 committed or aborted; this would block T_1 from reading Y until T_2 was finished, thus there would be no interference.
- 3) Considers only 'Safe' interleaving of transactions so that transactions are recoverable, avoid cascading aborts.

- ② is very simple and easy to implement. The lock manager only needs to provide a look up for exclusive locks and an atomic locking mechanism (such as with a semaphore).

T-6. 1: Locking thrashing occurs when the database system reaches to a point where adding another new active transaction actually reduces throughput due to competition for locking.

- Explain how the use of Strict 2PL would prevent interference between the two transactions. Use symbol $s(X)$ for a shared-lock on X and $x(X)$ for an exclusive lock on X .
- Strict 2PL is used in many database systems. Give two reasons for its popularity.

① recoverable ② easy to implement

PPT 16.3 T-5 Consider a database with objects X and Y and assume that there are two transactions T1 and T2. Transaction T1 reads objects X and Y and then writes object X . Transaction T2 reads objects X and Y and then writes objects X and Y .

- Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-read conflict.

<u>T₁</u>	<u>R(X) R(Y) W(X)</u>
<u>T₂</u>	

- Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a read-write conflict.

<u>T₂</u>	<u>R(X) R(Y) W(X)</u>
<u>T₁</u>	

- Give an example schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-write conflict.

<u>T₁</u>	<u>R(X) R(Y) W(X)</u>
<u>T₂</u>	<u>R(X) R(Y) W(X)</u>

- For each of the three schedules, show that Strict 2PL disallows the schedule.

(a). T_2 Exclusive(X), (b) T_2 exclusive(X) or share(X)

T-6 We call a transaction (that only reads database object) a read-only transaction, otherwise the transaction is called a read-write transaction. Give brief answers to the following questions:

P16.2

- What is lock thrashing and when does it occur?

- What happens to the database system throughput if the number of read-write transactions is increased?

- What happens to the database system throughput if the number of read-only transactions is increased?

- Describe three ways of tuning your system to increase transaction throughput.

T-7 In a recoverable schedule, transactions commit only after all transactions (whose changes they read) commit. Below is an unrecoverable schedule:

$T_1 : R(A), T_1 : W(A), T_2 : R(A), T_2 : W(A), T_2 : R(B), T_2 : W(B), T_2 : Commit, T_1 : Abort$

T_2 has read a value for A that should never have been there (aborted transactions' efforts are not supposed to be visible to other transactions.) If T_2 had not yet committed, we could deal with the situation by cascading the abort of T_1 and also aborting T_2 ; this process recursively aborts any transaction that read data written by T_2 , and so on. But T_2 has already committed, and so we cannot undo its actions.

<u>T₁</u>	<u>T₂</u>
<u>R(A)</u>	
<u>W(A)</u>	<u>R(A) W(A)</u>
	<u>R(B) W(B)</u>

读取未提交的数据，读别人提交才提交，所以读

1.	T ₁	T ₂	T ₁	T ₂	T ₁	T ₂	T ₃
	R(X)	R(X)		R(Y)	R(X)	R(Y)	
	W(X)	W(X)		R(X)	R(X)	R(X)	W(X)
						R(Y)	

未读被更改

not serializable.

	serializable	conflict	view	recoverable	avoids	safe
1	(abort T ₂) ✓ not decided	X	X	V	✓	X
2	✓	✓	✓	not decided	not decided	X
3	✓	✓	✓	not decided	X	not decided
4	(above T ₂ , T ₃) not decided	X	not decided	not decided	X	X
5	✓	✓	✓	✓	✓	X
6	X	X	X	✓	✓	X
7	✓	✓	✓	X	✓	✓
8	V ₀	X	X	X	X	X
9	✓	✓	✓	X	X	X
10	✓	✓	✓	✓	✓	✓
11	X	X	X	✓	✓	✓
12	X	X	X	✓	X	X

4.	T ₁	T ₂	T ₃	5.	T ₁	T ₂	6.	T ₁	T ₂	7.	T ₁	T ₂
	R(X)	R(X)			R(X)	W(X)		R(X)	W(X)		W(X)	R(X)
	R(Y)				W(X)			W(X)			W(X)	Abort
	W(X)					Abort			Commit			Commit
		R(Y)										

8.	T ₁	T ₂	9.	T ₁	T ₂	10.	T ₁	T ₂	T ₃	11.	T ₁	T ₂	T ₃
	W(X)	R(X)		W(X)	R(X)		W(X)	R(X)	W(X)		R(X)	W(X)	
	W(X)	R(X)		W(X)	Commit		W(Y)	W(Z)	Commit		W(X)	Commit	
		Commit						R(Y)				R(X)	Commit

静态 database = 冲突可串行化 \rightarrow 可串行化

serializable = commit, 每个事务一样
abort

Conflict-serializable = 冲突 (RW, WR, WW) 部分相同

View-serializable = 开始、读来读去一样、中间读一样

读提交后公数据 / Abort 后你也读 abort

If transactions read only the changes of committed transactions, not only is the schedule recoverable, but also aborting a transaction can be accomplished without cascading the abort to other transactions. Such a schedule is said to avoid cascading aborts. \rightarrow 加了这个

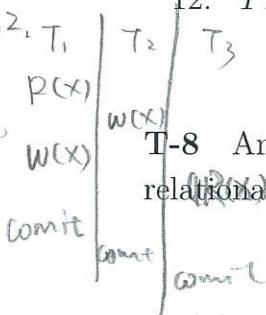
A schedule is said to be strict if a value written by a transaction T is not read or overwritten by other transactions until T either aborts or commits. Strict schedules are recoverable, do not require cascading aborts, and actions of aborted transactions can be undone by restoring the original values of modified objects.

Now, consider the following classes of schedules: serializable, conflict-serializable, view-serializable, recoverable, avoids-cascading-aborts, and strict. For each of the following schedules, state which of the preceding classes it belongs to. If you cannot decide whether a schedule belongs in a certain class based on the listed actions, explain briefly. \rightarrow 不一定要 commit <

The actions are listed in the order they are scheduled and prefixed with the transaction name. If a commit or abort is not shown, the schedule is incomplete; assume that abort or commit must follow all the listed actions.

- 看 $R(X)$ 是否有 $W(X)$
若有, 则 $W(X)$ 必须先提交
- 看 $read(X)$ 前如果有 $W(X)$, $W(X)$ 必须在 $read(X)$ 前提交
- 看 $W(X)$ 是否有操作
若有, 要在操作之间 $W(X)$ 已提交
- 冲突 \neq serializable 的充分条件
长女制 view
1. $T_1 : R(X), T_2 : R(X), T_1 : W(X), T_2 : W(X)$ not
 2. $T_1 : W(X), T_2 : R(Y), T_1 : R(Y), T_2 : R(X)$ not \Rightarrow strict \Rightarrow recoverable
 3. $T_1 : R(X), T_2 : R(Y), T_3 : W(X), T_2 : R(X), T_1 : R(Y)$ 看 R 的操作, 无论怎样
 4. $T_1 : R(X), T_1 : R(Y), T_1 : W(X), T_2 : R(Y), T_3 : W(Y), T_1 : W(X), T_2 : R(Y)$
 5. $T_1 : R(X), T_2 : W(X), T_1 : W(X), T_2 : Abort, T_1 : Commit$ not recoverable \Rightarrow not avoid, not strict
 6. $T_1 : R(X), T_2 : W(X), T_1 : W(X), T_2 : Commit, T_1 : Commit$
 7. $T_1 : W(X), T_2 : R(X), T_1 : W(X), T_2 : Abort, T_1 : Commit$
 8. $T_1 : W(X), T_2 : R(X), T_1 : W(X), T_2 : Commit, T_1 : Commit$
 9. $T_1 : W(X), T_2 : R(X), T_1 : W(X), T_2 : Commit, T_1 : Abort$
 10. $T_2 : R(X), T_3 : W(X), T_3 : Commit, T_1 : W(Y), T_1 : Commit, T_2 : R(Y), T_2 : W(Z), T_2 : Commit$
 11. $T_1 : R(X), T_2 : W(X), T_2 : Commit, T_1 : W(X), T_1 : Commit, T_3 : R(X), T_3 : Commit$
 12. $T_1 : R(X), T_2 : W(X), T_1 : W(X), T_3 : R(X), T_1 : Commit, T_2 : Commit, T_3 : Commit$

T-8 Answer each of the following questions briefly. The questions are based on the following relational schema:



Emp(eid : integer, $ename$: string, age : integer, $salary$: real, did : integer)
Dept(did : integer, $dname$: string, $floor$: integer)

and on the following update command:

S → X

replace (salary = 1.1 * EMP.salary) where EMP.ename = 'Santa'

$$\text{Salary} = \text{Emp.Salary} - 100$$

- Give an example of a query that would conflict with this command (in a concurrency control sense) if both were run at the same time. Explain what could go wrong, and how locking tuples would solve the problem.

保持一致性 ~~和并发性~~

- Give an example of a query or a command that would conflict with this command, such that the conflict could not be resolved by just locking individual tuples or pages but requires index locking.

$T_1 : \text{where } T_2 : \text{Insert}(A), \text{Insert}(B)$, ~~where~~ $\text{Insert}(A) \times \text{Insert}(B)$. 和 $T_1.A, T_2.B \neq A, B$ 不同

- Explain what index locking is and how it resolves the preceding conflict.

应该锁住索引页，包含数据条目 $\text{EMP.ename} = \text{'Santa'}$.

- ~~DB Unit T-9~~ Consider a database organized in terms of the following hierarchy of objects: The database itself is an object (D), and it contains two files (F_1 and F_2), each of which contains 1000 pages ($P_1 \dots P_{1000}$ and $P_{1001} \dots P_{2000}$, respectively). Each page contains 100 records, and records are identified as $p : i$, where p is the page identifier and i is the slot of the record on that page.

Multiple-granularity locking is used, with S, X, IS, IX and SIX locks, and database-level, file-level, page-level and record-level locking. For each of the following operations, indicate the sequence of lock requests that must be generated by a transaction that wants to carry out (just) these operations:

- F_1 F_2
- Read record $P_{1200}:5$. $IS(D), IS(F_2), IS(P_{1200}), S(P_{1200}:5)$
 - Read records $P_{1200}:98$ through $P_{1205}:2$. $IS(D), IS(F_2), IS(P_{1200}), S$ on P_{1201} through P_{1204} , IS on P_{1205} , S on $P_{1200}:98/2$
 - Read all (records on all) pages in file F_1 . $IS(D), S(F_1)$ S on $P_{1205}:1/2$.
 - Read pages P_{500} through P_{520} . $IS(D), IS(F_1), S$ on P_{500} through P_{520}
 - Read pages P_{10} through P_{980} . $IS(D), S$ on F_1 .
 - Read all pages in F_1 and (based on the values read) modify 10 pages. IS and IX on D , SIX on F_1
 - Delete record $P_{1200}:98$. (This is a blind write.) $IX(D), IX(F_2), X(P_{1200})$
 - Delete the first record from each page. (Again, these are blind writes.) $IX(D), X(F_1), X(F_2)$.
 - Delete all records. $IX(D), X(F_1)$ and $X(F_2)$

X-1 What is the main concepts of HTML? What are the short coming of HTML, and how does XML address them? *Unit 7. P226. 228*

X-2 Explain the following terms and describe what they are used for: HTML, URL, XML, XPath, DTD, cookie, HTTP, HTTPS.

<?xml version='1.0'?>

well-formed \Rightarrow valid

有结构

X-3 Give an example of a DTD. Also give an XML document that is valid against the DTD but not well-formed, and vice versa. 反过来也一样

Unit 1, PPT, 28TR

说明

教材

X-4 The following pages contain XPath language specification and tutorial:

<http://www.w3.org/TR/xpath>

<http://www.w3schools.com/Xpath/>

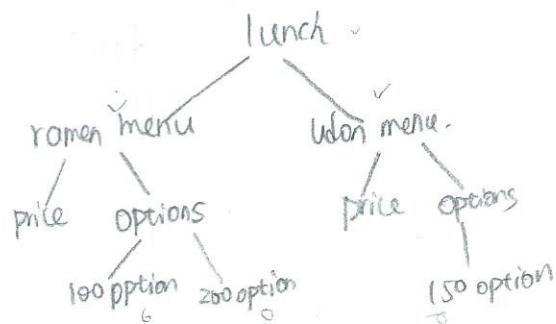
Construct a sample XML document, and write XPath query examples that use the XPath language constructs.

X-5

1. Design an algorithm for testing whether given XML documents are well-formed.
① type, name, domain ② subelement.
2. Design an algorithm for testing whether the DOM tree of a given well-formed XML document is valid against a given DTD. Try using nondeterministic finite automata (NFA).
3. Design an algorithm answering XPath queries over given XML documents. You may restrict the language constructs of XPath to a subset (for example, just considering a subset of document axis, and excluding predicates, etc).

X6 Consider the following XML document d_1 :

```
<lunch>
  <menu name="ramen" type="noodle">
    <price> 500 </price>
    <options>
      <option price="100"> large_noodle </option>
      <option price="200"> extra_meet </option>
    </options>
  </menu>
  <menu name="udon" type="noodle">
    <price>300</price>
    <options>
      <option price="150">large_noodle</option>
    </options>
  </menu>
</lunch>
```



1. Write the answers to the following XPath queries applied on d_1 .

(a) $//*[@@price > "100"]/\text()$

(a) extra_meet
large_noodle

(b) options/ancestor::* / price

(b) <price>500</price>
<price>300</price>

(c) * [preceding-sibling::* / attribute::price]

(c) <option price="200">extra_meet</option>

前面优先节点

2. Write an XPath query which returns the following:

- (a) Find 'ramen' elements such that all of their 'option' descendants have 'price' larger than 100.
- (b) Find elements such that whose name is either 'menu' or 'option' and whose attribute value or text value is 'noodle'.
- (c) Find elements that have a sibling which has an attribute named 'special'.

3. Write a DTD e such that document d_1 is valid against e .

在 DTD 里 4. Write a DTD f such that document d_1 is also valid against f , and there is another document d_2 such that d_2 is valid against e but not valid against f . Show such DTD f and document d_2 .

2. (a) $*[@name='ramen'][descendant::option[@price > 100]]$
- (b) $*[@name='menu|option'][@attribute::*='noodle'] | [text()='noodle']$
- (c) $*[Sibling::*[@attribute::*='special']]$

④. $<!DOCTYPE lunch [$

```

<!ELEMENT lunch (menu)*>
<!ELEMENT menu (price*, options*)>
<!ELEMENT price #PCDATA>
<!ELEMENT options(option*)>
<!ELEMENT option>
<!ATTLIST option price (#CDATA) #REQUIRED>
<!ATTLIST menu name (#CDATA) #REQUIRED>
<!ATTLIST menu type (#CDATA) #REQUIRED>
```

4. I>

3. 去掉 REQUIRED.

4. 2. $d_2 = \text{无元素}$.

