

XML i nowoczesne technologie zarządzania treścią

Egzamin, 2 lutego 2010 r.

Imię i nazwisko: _____

Pytania testowe są wielokrotnego wyboru, prawdziwa jest co najmniej jedna odpowiedź. Każde pytanie – 1 punkt. W przypadku pytań opisowych możliwe są także oceny ułamkowe.

Tym kolorem oznaczono podpunkty traktowane ulgowo – pytanie z niepoprawną odpowiedzią na taki podpunkt, ale z poprawnymi pozostałymi, było oceniane na 0,5 punktu.

1. Zaznacz stwierdzenia prawdziwe:

- a) Każdy dokument XML-owy musi być poprawny strukturalnie (ang. *valid*).
- b) Element pusty nie może zawierać atrybutów.
- c) Jeśli wartość atrybutu nie zawiera spacji, można pominąć otaczające ją cudzysłowy lub apostrofy.
- d) Dokument XML-owy nie musi być zapisany w kodowaniu UTF-8.

2. Które deklaracje elementów mogą być poprawne?

- a) `<!ELEMENT test (#PCDATA)>`
- b) `<!ELEMENT test CDATA>`
- c) `<!ELEMENT test EMPTY>`
- d) `<!ELEMENT test (ANY)>` - oznacza podelement o nazwie ANY

3. Wskaż i objaśnij trzy błędy składni XML-owej w poniższym dokumencie. Zakładamy, że dokument jest zapisany w kodowaniu UTF-8.

```
<?xml version="1.0" ?>
<!-- $Id: -- egz-2010-02-02, v 1.2 <mog> Exp $ -->
<html><p>Ten dokument
<div align='center'>
<b>nie jest, <i><br>
<_tag-własny_>niestety</_tag-własny_>, <b/> </i></b>,
<i>poprawny składniowo</i>.</div></p>
<hr noshade>
</html>
```

- -- w komentarzu
-
 nie zamknięte
- noshade - atrybut bez wartości lub brak znacznika > po hr

4. Które stwierdzenia są prawdziwe?

- a) Normalizacja dla napisów typu `normalizedString` polega na usunięciu białych znaków z początku i końca napisu oraz zastąpieniu ciągów białych znaków przez pojedynczą spację.
tylko zamiana innych białych znaków na spację
- b) Element o wartości nieokreślonej (ang. *nil value*) musi mieć zawartość pustą niezależnie od definicji jego modelu zawartości w schemacie, natomiast jego atrybuty muszą być zawsze zgodne z modelem.
- c) Jedną z metod zapewniania odporności na zmianę struktury dokumentów w czasie jest parametryzacja schematem – użycie atrybutów opcjonalnych do przechowania etykiet pól formularza, odwzorowania elementów na tabele i pola w bazie danych itp.
nie opcjonalnych, tylko ustalonych (*fixed*)
- d) W składni XPointer napis `dokument.xml#element(test)` oznacza odwołanie do wszystkich elementów o nazwie `test` znajdujących się w pliku `dokument.xml`.
do elementu o ID równym `test`

5. Które stwierdzenia nt. przestrzeni nazw w XML-u są prawdziwe?

- a) Nazwy kwalifikowane muszą być poprzedzone prefiksem skojarzonym z przestrzenią nazw, do której należą.
- b) W definicji schematu URI używane w deklaracji docelowej przestrzeni nazw nie musi wskazywać miejsca lokalizacji schematu.
- c) Nazwa atrybutu nie poprzedzona prefiksem przestrzeni nazw należy do tej samej przestrzeni nazw, co nazwa elementu, w którym atrybut się znajduje.
- d) Nazwy encji mogą używać niestandardowych prefiksów przestrzeni nazw.
zgodnie ze standardem przestrzeni nazw nazwy encji nie mogą zawierać dwukropka

6. Zapisz dany fragment DTD w składni XML Schema (jako fragment pełnego schematu – zdefiniuj globalne elementy **test** i **pierwszy** wraz z atrybutami, pomini definicje elementów **drugi**, **trzeci** i **czwarty**, ale odwołaj się do nich, jak gdyby były zdefiniowane globalnie).

Każdy element za 0,5 pkt.

```
<!ELEMENT test (pierwszy?, (drugi | trzeci+), czwarty*)>
```

```
<!ELEMENT pierwszy (#PCDATA | drugi)*>
```

```
<xsd:element name="pierwszy">
  <xsd:complexType mixed="true">
    <xsd:choice>
      <xsd:element ref="drugi"/>
    </xsd:choice>
  </xsd:complexType>
</xsd:element>
```

```
<xsd:element name="test">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="pierwszy" minOccurs="0"/>
      <xsd:choice>
        <xsd:element ref="drugi"/>
        <xsd:element ref="trzeci" maxOccurs="unbounded"/>
      </xsd:choice>
      <xsd:element ref="czwarty" minOccurs="0" maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

7. XML Schema pozwala na:
- a) ograniczenie zawartości elementu do postaci listy 5 liczb naturalnych z przedziału 1-10 oddzielanych dowolną liczbą białych znaków,
 - b) określenie, że wymienione podelementy danego elementu mogą wystąpić w dowolnej kolejności,
 - c) określenie, że wartość tekstowa dwóch atrybutów danego elementu ma być identyczna, tak, poprzez klucze i referencje
 - d) określenie, że element może zawierać dowolne atrybuty.
8. Które stwierdzenia nt. standardu XLink są prawdziwe?
- a) XLinki mogą łączyć nie więcej niż 2 zasoby.
 - b) XLinki mogą występować wyłącznie w dokumentach poprawnych strukturalnie.
 - c) XLinki umożliwiają nawigację w wielu kierunkach.
 - d) XLinki nie muszą być zagnieżdżane w treści linkowanych dokumentów.
9. Zaznacz prawdziwe stwierdzenia dotyczące dokumentów XSL Formatting Objects.
- a) Dokument XSL-FO pełni rolę arkusza stylu.
arkusz XSLT pełni rolę arkusza stylu, a XSL-FO jest wynikiem przekształcenia, wizualną wersją 1 dokumentu
 - b) Formatowanie elementów danego rodzaju (np. tabel) opisuje się w jednym miejscu dokumentu.
 - c) Formatowanie elementów zawierających treść opisuje się w ich atrybutach.
 - d) Można raz podać zawartość, która będzie powtarzana na każdej stronie (np. w nagłówku).
10. Poniższy fragment arkusza XSLT
- ```
<xsl:variable name="x" select="/test/*"/>
<xsl:if test="$x">
 <xsl:variable name="x" select="/test/*[1]"/>
</xsl:if>
<xsl:value-of select="$x"/>
```
- przy dokumencie wejściowym
- ```
<test><a>jeden</a><b>dwa</b><c>trzy</c></test>
```
- może (w XSLT 1.0 lub 2.0) spowodować wypisanie na wyjście (przy metodzie serializacji xml):
- a) <a>jeden
 - b) <a>jedendwa<c>trzy</c>
 - c) jeden
 - d) jeden dwa trzy
11. Zaznacz prawdziwe stwierdzenia dotyczące XQuery 1.0.
- a) Zapytanie XQuery jest dokumentem XML.
 - b) Efektywna wartość logiczna sekwencji składającej się z jednego węzła tekstowego to prawda, niezależnie od tekstu zawartego w węźle.
węzeł (z drzewa dokumentu) jest zawsze rzutowany na prawdę, nawet jeśli zawiera pusty napis, węzeł tekstowy to nie to samo co wartość tekstowa (napis)
 - c) Wynik zapytania może zawierać elementy znajdujące się w dokumencie źródłowym.
 - d) Wynik zapytania może zawierać elementy nie znajdujące się w dokumencie źródłowym.

12. Pewien dokument XML zawiera wiele elementów **klucz** z zawartością tekstową. Napisz ścieżkę XPath 1.0, która zwraca liczbę różnych wartości tekstowych elementów **klucz**.

```
count(//klucz[not(preceding::klucz = .)])
```

13. Zaznacz prawdziwe stwierdzenia dotyczące korzystania z SAX2.

- a) Parsowanie nie wymaga wczytywania całego dokumentu na raz do pamięci.
- b) W oparciu o schemat dokumentu typy zawartości elementów i atrybutów są odzwierciedlane na odpowiednie typy w języku programowania.
- c) Aby rozpoznawać identyfikatory przestrzeni nazw elementów i atrybutów, należy zapamiętywać historię deklaracji przestrzeni nazw.
- d) Filtry mogą przekazywać do dalszego przetwarzania zdarzenia nie występujące na wejściu.

14. Podaj zalety i wady XML jako formatu elektronicznej wymiany danych w stosunku do tradycyjnego EDI (np. EDIFACT).

- + większa elastyczność, możliwość rozszerzania formatów o elementy prywatne, specyficzne dla branży itp.
- + możliwość zmiany formatu z zachowaniem kompatybilności z istniejącymi aplikacjami (które ignorują nieznanne elementy)
- + możliwość wykorzystania istniejącego wsparcia dla XML (parsery, walidacja, XSLT, XPath)
- + prostsza integracja z WebSerwisami i technologiami WWW (np. eksport do HTML przez XSLT)
- większa objętość dokumentów XML w stosunku do np. EDIFACT
- problem biznesowy: systemy i procesy biznesowe działające w oparciu o „stare” standardy wystarczyło podać dwa sensowne plusy i jeden minus

15. Wymień trzy czynności, które podsystem wyszukiwarki przetwarzający tekst zapytania powinien wykonać zanim zacznie dopasowywać go do zawartości indeksu.

- ujednoczenie wielkości liter
- segmentacja/tokenizacja tekstu
- lematyzacja segmentów
- usunięcie słów nieznaczących

16. W jaki sposób reprezentowane są zdania logiczne w standardzie RDF (na poziomie idei, nie składni)? Podaj przykład z objaśnieniem.

Za pomocą trójki wartości: podmiot/orzeczenie/dopełnienie, z których pierwsza i trzecia reprezentują zasób, a druga - własność (która też może być zasobem).

Np. (student X.Y., zarejestrowany na, „XML i nowoczesne...”)