

Câu 1: (3đ)

Hãy soạn một tài liệu DTD và một tài liệu XSD để tài liệu XML **books.xml** sau là hợp lệ:

```
<?xml version="1.0"?>
<catalog>
  <book id="bk101">
    <author>Gambardella, Matthew</author>
    <title>XML Developer's Guide</title>
    <genre>Computer</genre>
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at creating applications with XML.</description>
  </book>
  <book id="bk102">
    <author>Ralls, Kim</author>
    <title>Midnight Rain</title>
    <genre>Fantasy</genre>
    <price>5.95</price>
    <publish_date>2000-12-16</publish_date>
    <description>A former architect battles corporate zombies.</description>
  </book>
</catalog>
```

Câu 2: (2đ)

Cho file **books.xml** như câu 1, hãy viết các biểu thức đường dẫn XPath để:

- Chọn phần tử **title** là con của phần tử **book** cuối cùng.
- Lấy ra phần tử **book** có phần tử **price** với giá trị lớn hơn **25.00**

Câu 3: (3đ)

Cho một tập các xâu $S = \{s_1, s_2, s_3\}$ với $s_1 = \text{"Nguyễn Anh Tuấn"}$, $s_2 = \text{"Nguyễn Đức Tuấn"}$ và $s_3 = \text{"Nguyễn Tuấn Trí"}$.

Trình bày cách tính *điểm tương đồng* TF/IDF giữa hai chuỗi s_1 và s_2

Câu 4: (2đ)

Hãy trình bày các lí do khiến cho tích hợp dữ liệu là một vấn đề khó.

-----HẾT-----

Soạn đề thi

Duyệt đề thi