

Homework #1: XML

The following exercises are due at the beginning of class on February 2. This will count for 5% of your overall grade. You are reminded that this is an individual assignment.

1. [50 points] Determine which of the following XML documents is well-formed. If a document is well-formed, show its corresponding tree model. If it is not well-formed, explain why not. You may assume that each of these documents are intended to be XML 1.0 documents, and thus the declaration is not needed.

- a) `<a><a>`
- b) `<a>foo<a>bar`
- c) `<a>foobar`
- d) `<a><c>foobar</c>`
- e) `<a/>foobar`
- f) `<a><c>foo</c>baz<c>bar</c>`
- g) `foo`
- h) `<b x="2">foo`
- i) `foo`
- j) `<a><b x="3"><c y="1">foo</c><c>bar</c>`

2. [25 points] Consider the following DTD.

```
<!ELEMENT p (q, r?, s*)>
<!ELEMENT q (u+ | v)>
<!ELEMENT r (#PCDATA)>
<!ELEMENT s EMPTY>
<!ELEMENT u EMPTY>
<!ELEMENT v EMPTY>
```

Which of the following documents are valid with respect to this DTD? For each document that is not valid, explain your answer.

- a) `<p><q><u/></q><s/></p>`
- b) `<p><q><v/></q><r>hi</r><s/></p>`
- c) `<p><q><u>it</u></q><s/><s/></p>`
- d) `<p><q><v></v></q></p>`
- e) `<p><q><u/></q><r>blah</r><r>ack</r></p>`

3. [10 points] In the books e-mail DTD on p. 36, we specified the body of an e-mail to contain exactly one text and a number of attachments. Modify the DTD to allow for an arbitrary number of texts and attachments in any order. You only need to show the lines that you would change.

4. [15 points] Consider the following XML document.

```
<A >
  <B foo="3">one</B>
  <B foo="2">
    <C>two</C>
  </B>
  <B foo="1">
    <C>three</C>
  </B>
</A>
```

For each of the XPath expressions below, describe what nodes of the tree will be selected. For clarity, you should give the complete text of each node. For example, if an element is selected, show its start tag, contents and end tag.

- a) /A/B[2]
- b) /A/B[@foo="1"]
- c) //C