UCT CSC303 2005 :: XML/IR + ToA :: Test [35 marks]

Answer Questions 1, 2 AND 3.

Question 1: XML / XSLT [10]

Answer the following questions based on this piece of XML:

```
<mdata>
    <name>CSC303 Test</name>
    <author>
        <first_name>hussein</first_name>
        <last_name>suleman</last_name>
    </author>
</mdata>
```

Assume that the **name** and **author** elements are both infinitely repeatable and optional and that **first_name** and **last_name** must both appear exactly once each.

1. Write an XSLT template to transform the XML fragment into:

```
<dublin_core>
    <title>CSC303 Test</title>
    <creator>hussein suleman</creator>
</dublin_core>
```

Assume that the values of the content of the **name** and **author** nodes may differ from one document to another. Assume the source namespace prefix is *source* and the destination prefix is *dest*. Use the following as a starting point. [5]

```
<xslt:template match="source:mdata">

. . .

</xslt:template>
```

2. Write code that uses the DOM API to access the contents of the **first_name** node within the first **author** node and store it into the *fname* (or *$fname*) variable, given that the document has been parsed and assigned to the *top* (or *$top*) variable.

Note: The sequence of commands is important, not the programming language. [2]

3. Why do we need namespaces in XML documents, such as the XSLT template created in the previous question? [1]

4. It is possible to create XML documents with multiple namespaces but no defined namespace prefixes. How? [2]

Question 2: Information Retrieval [10]

1. If we always look at only the first 20 documents in a result set, ranking of documents serves to increases precision. What is ranking and what is precision? [2]

2. Discuss 2 techniques to increase the precision of an IR system. [2]

3. When ranking documents, it is common to use logarithmic functions in the ranking formulae – what is the purpose of this non-uniform scaling of values? [2]

4. Describe the simple HITS algorithm. [4]