Week 4: Well-Formedness

Big Data Exercise Session

- CSV, JSON ,XML
- JSON building blocks:
  - String "abc"
  - Number 0.5
  - Null null
  - Boolean true/false
  - Object {"a": 1, "b": 2}
  - Array [1, null, "abc"]
- XML building blocks:
  - Element <a> (<a/>)
  - Attribute
  - Text <a> abcdefg
  - Comment
  - Text declaration
    <?xml version="1.0" encoding="utf-8"?>
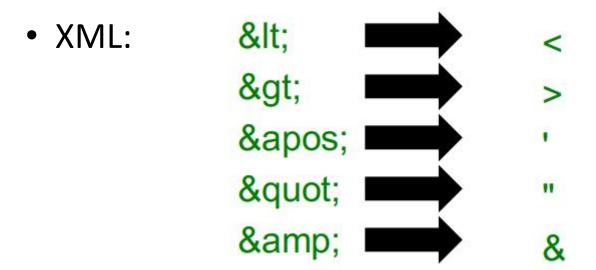   <!DOCTYPE catalog>

Rules for naming

- JSON key names:
  - must be a string (double quoted)
  - cannot contain unescaped " or \
  - no duplication keys within an object
- XML element/ attribute names:
  - no duplication attributes for one element
  - duplication elements allowed

Rules for escaping

- JSON:

Escape	Meaning
\\	\
\n	new line
\r	carriage return
\t	tabulation
\u followed by four hexadecimal digits	any character

- XML:

Entity	Character
<	<
>	>
'	'
"	"
&	&

XML namespaces

- Prefix binding

<m:math xmlns:m="http://www.w3.org/1998/Math/MathML"/>

- Default namespace

<math xmlns:"http://www.w3.org/1998/Math/MathML"/>

- Unprefixed attributes are in no namespace even if there is a default namespace in scope

- In practice, declare all namespaces at the top level element.

Four "clean" kinds of documents

```xml
<?xml version "1.0"?>
<bar>
  <foo/>
  <foobar>
    <foo/>
  </foobar>
</bar>
```

1. No namespaces at all.

```xml
<?xml version "1.0"?>
<bar xmlns="http://example.com/a">
  <foo/>
  <foobar>
    <foo/>
  </foobar>
</bar>
```

2. One default namespace.

```xml
<?xml version "1.0"?>
<a:bar xmlns:a="http://example.com/a">
  <a:foo/>
  <a:foobar>
    <a:foo/>
  </a:foobar>
</a:bar>
```

3. One prefixed namespace.

```xml
<?xml version "1.0"?>
<a:bar
  xmlns:a="http://example.com/a"
  xmlns:b="http://example.com/b"
  xmlns:c="http://example.com/c"
  xmlns:d="http://example.com/d">
  <b:foo/>
  <c:bar>
    <d:foo/>
    <a:foobar/>
  </c:bar>
</a:bar>
```

4. Many prefixed namespaces.