

Validation methods

External DTD :

The external DTD is stored in a separate file and referenced by the XML document.

- The External DTD defines the structure and rules for the XML document.
- It states that the document must have a root element called "data" that can contain multiple "entry" elements.
- Each "entry" element should have four child elements: "url," "title," "question," and "answer."
- The "url," "title," "question," and "answer" elements can only contain text content (PCDATA).
- When validating the XML file using the external DTD, two additional lines were added in the xml file:
 - `<?xml version="1.0" encoding="UTF-8" standalone="no"?>`: This line specifies the XML version (1.0) and the character encoding (UTF-8) used in the document. The "standalone" attribute is set to "no," indicating that the XML document relies on external resources, such as the referenced DTD.
 - `<!DOCTYPE data SYSTEM "ext_islamqa_10thousand_500.dtd">`: This line declares the Document Type Definition (DTD) for the XML document. It specifies that the DTD is an external DTD located in a separate file called "ext_islamqa_10thousand_500.dtd."

Validation command for external dtd: `xmllint --dtdvalid ext_islamqa_10thousand_500.dtd islamqa_10thousand_500.xml`

Internal DTD:

Likewise, the XML document with an internal DTD provides a structure for representing data related to questions and answers. The DTD defines the elements and their ordering, ensuring the document adheres to a specific structure and rules. Unlike external DTD, the internal DTD is included directly within the XML document itself, after the XML declaration and before the root element.

- `<?xml version="1.0" encoding="UTF-8"?>`: This line specifies the XML version (1.0) and the character encoding (UTF-8) used in the document.

- `<!DOCTYPE data [...]>`: This line declares the Document Type Definition (DTD) for the XML document. It defines the structure and rules for the elements in the document.
- The rest of the structure and rules adhere to the external DTD.

Validation command for internal dtd: `xmllint --valid int_islamqa_10thousand_500.xml`

XML Schema:

The XML Schema defines the structure and constraints for an XML document. The schema specifies the following:

- `<?xml version="1.0" encoding="UTF-8"?>`: Like the internal DTD, this line specifies the XML version (1.0) and the character encoding (UTF-8) used in the document.
- `<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">`: This line declares the XML Schema namespace (xmlns) with the prefix "xs" and associates it with the XML Schema Definition (XSD) namespace (["http://www.w3.org/2001/XMLSchema"](http://www.w3.org/2001/XMLSchema)).
- The root element of the XML document should be named "data." The "data" element has a complex type, meaning it can contain multiple occurrences of the "entry" element. The child elements of the "data" element must appear in a specific sequence.
- The "entry" element also has a complex type, meaning it can contain child elements. Each "entry" element should have child elements in the following sequence: "url," "title," "question," and "answer."
- The "url" element should contain a valid URI.
- The "title," "question," and "answer" elements should contain string values.
- Two additional lines were added in the xml file so that the XML document associates an XML Schema file ("schema.xsd") with the document and provides information for the XML parser to locate and validate the XML against the specified schema rules.

Validation command for XML Schema: `xmllint --schema islamqa_10thousand_500.xsd schema_islamqa_10thousand_500.xml`