

XML file validated using xml schema:

1. **xmllint** tool was used to validate ,islamqa_final_1500.xml'. **Command:**
xmllint --schema islamqa_final_1500.xsd islamqa_final_1500.xml
2. XML Schema grammar: It defines three complex types: EntryType, EmbeddingType, and data.
 - EntryType has elements for "url," "title," "question," "answer," "ada_embedding," and "gpt4all_embeddings." Each of these elements has specific data types defined, such as xs:anyURI for "url" and xs:string for "title," "question," "answer," "ada_embedding," and "gpt4all_embeddings."
 - EmbeddingType extends from a base type of xs:string. It has an attribute named "type" of type xs:decimal. This attribute can hold decimal values.
 - Finally, the root element "data" also has a complex type defined, which includes a sequence of "entry" elements. The "data" element can have multiple "entry" elements (maxOccurs="unbounded")

XPath queries on XML file to get the references and theologians:

1. To count 6 most authentic Hadith books mentioned in the answers, one example: **xpath -e 'count(/data/entry[contains(answer, "Sahih Muslim")]/answer)' islamqa_final_1500.xml** **Outputs:** Bukhari: 112, Tirmidhi: 186, Nasai: 29, Dawood: 136, Sahih Muslim: 20, Majah: 31
2. To count how many times different theologians are mentioned in the answer: **xpath -e 'count(/data/entry[contains(answer, "Nawawi")]/answer)' islamqa_final_1500.xml** **Outputs:** Nawawi: 108, Taymiyah: 72, Albani: 35, Baz: 28, Uthaymin: 23, Hanbal: 38