****

# What are the differences between XML and HTML?

Chenxing Xie, University of Cincinnati

Revision history (ACM-018)

|  |  |  |
| --- | --- | --- |
| **Date** | **Author** | **Revision summary** |
| TBD | Chenxing | First public posting |

Both Extensible Markup Language (XML) and Hypertext Markup Language (HTML) are widely used in information design. As markup languages that illustrate information to human or machine audiences, XML and HTML employ an “easy-to-read syntax that defines the data’s structure, type, attributes, relationships between parts, display, and meaning” (AWS, n.d.). While both of them can help technical writers design documentation or web content, there are differences between XML and HTML. This article reviews and analyzes the differences between these two markup languages from a rhetorical perspective. Specifically, XML and HTML serve different rhetorical purposes in different rhetorical situations, which can be explored through the angles of exigence, audience, and constraints.

|  |  |  |
| --- | --- | --- |
| Rhetorical perspectives | XML | HTML |
| Exigence | Structured authoring | Information Presentation |
| Audience | Machine | Human |
| Constraints | Customized tags | Predefined tags |

In his foundational work *The Rhetorical Situation*, Bizter (1992) defines rhetorical situation as “*a complex of persons, events, objects, and relations presenting an actual or potential exigence which can be completely or partially removed if discourse, introduced into the situation, can so constrain human decision or action as to bring about the significant modification of the exigence*” (p.6). According to Bitzer, three key components in a rhetorical situation are exigence, audience, and constraints. A rhetorical exigence is an urgent need to address an imperfection through rhetorical discourse. The audience is the person who engages with and can be influenced by the discourse. The constraints are the factors that constrain the effectiveness of the rhetorical discourse (Bitzer, 1992). Examining XML and HTML documentations as rhetorical situations, the differences between these two documentations can be analyzed through the perspectives of exigence, audience, and constraint.

First, XML and HTML serve different rhetorical exigences. The primary rhetorical exigence of XML is the need for structured authoring. XML is designed as a universal data format for structuring, storing, and exchanging data from different applications or databases (AWS, n.d.; Keycdn, 2023). In contrast, the rhetorical exigence of HTML is to present and visualize the content in browsers (AWS, n.d.). In other words, HTML structures web content so that browsers can display textual and visual content as interactive web pages (AWS, n.d.).

With the aim to serve different rhetorical exigences, the target audience for XML and HTML is also different. XML’s target audience is machines. XML is primarily designed for machine processing through software such as OxygenXML to efficiently manage and exchange structured information between different systems (AWS, n.d., Geeksforgeeks, 2024). On the contrary, HTML is created for human audiences who view and interact with textual and visual web content (AWS, n.d.).

Another difference between XML and HTML is tags. From a rhetorical perspective, these tags can be perceived as metadiscourse. Hyland (2017) defines metadiscourse as “the commentary on a text made by its producer in the course of speaking or writing” (p. 16). Similarly, in XML and HTML, tags are not content but instead used to define “structure and type to data elements” (AWS, n.d.). Therefore, they serve as the function of metadiscourse to indicate the “beginning and end of each element of content” (AWS, n.d.). However, the flexibility of tags is different in XML and HTML, which also constrain the flexibility of documentation. While XML tags can be customized, HTML tags are predefined (Keycdn, 2023). Another distinction is that only a few HTML tags can be self-closed, while XML allows any empty tag to be written in self-closing form (AWS, n.d.).

While there are notable differences between XML and HTML, HTML can be converted to XML using specific approaches, such as Potok et al.’s (2002) ontology-based methodology. These differences influence how each markup language is used in documentation and shape their effectiveness in distinct rhetorical contexts. Therefore, it is important for technical communicators to consider the rhetorical situation when deciding whether XML or HTML is more appropriate for a particular context.

**References:**

AWS. (n.d.) What’s the difference between HTML and XML? Retrieved from: <https://aws.amazon.com/compare/the-difference-between-html-and-xml/#:~:text=The%20core%20difference%20between%20HTML,define%20as%20a%20document%20creator>.

Bitzer, L. F. (1999). The rhetorical situation. *Philosophy and Rhetoric*, *1*, 509-522.

Geeksforgeeks. (2024). HTML vs XML. Retrieved from: <https://www.geeksforgeeks.org/html-vs-xml/>.

Hyland, K. (2017). Metadiscourse: What is it and where is it going?. *Journal of pragmatics*, *113*, 16-29.

Keycdn. (2023). XML vs HTML: What's the Difference? Retrieved from: <https://www.keycdn.com/support/xml-vs-html>.

Potok, T. E., Elmore, M. T., Reed, J. W., & Samatova, N. F. (2002, January). An ontology-based HTML to XML conversion using intelligent agents. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (pp. 1220-1229). IEEE.

Acknowledgements

* Joe Golner - technical review
* Emily Gresbrink - instructional review
* TBD - editorial review

Thank You!

See <https://acm-sigdoc-structured.org> to learn more about committee activities, available resources, and volunteer opportunities.

