

Para facilitar el intercambio y el funcionamiento de datos, desde hace mucho tiempo la gente tiene la motivación de estructurar archivos en un formato estándar. IBM creó GML (lenguaje de marcado generalizado) en la década de 1960 para implementar esta necesidad dentro de su sistema de publicación. IBM utiliza GML para mantener libros, informes y otros archivos a partir de un único tipo de archivo fuente.

SGML (lenguaje de marcado generalizado estándar, lenguaje de marcado generalizado estándar) evolucionó a partir del GML de IBM y SGML se convirtió en el medio interno de IBM para formatear y mantener archivos legales. Posteriormente, SGML se amplió y modificó como un estándar de información integral para adaptarse a una amplia gama de aplicaciones en toda la industria. Pero no fue hasta 1986 que SGML se convirtió en un estándar ISO. Aunque SGML es muy poderoso, no se ha popularizado durante mucho tiempo porque es muy complejo y requiere una gran cantidad de software costoso para ejecutarse.

En 1989, Tim Berners-Lee y Anders Berglund, investigadores del Laboratorio Europeo de Física de Partículas (CERN), crearon un lenguaje basado en marcado para marcar artículos compartidos en Internet y nació HTML. HTML puede considerarse como una aplicación simplificada de SGML. Su nacimiento ha promovido el gran desarrollo de la industria de la información.

HTML originalmente solo proporcionaba un método para expresar información de texto estático, pero obviamente no puede satisfacer las crecientes necesidades. Posteriormente, nacieron cada vez más etiquetas, como `` para describir imágenes y `<script>` para agregar soporte de secuencias de comandos a las páginas web para proporcionar contenido web dinámico. Los dos principales fabricantes de navegadores, Microsoft y Netscape, incluso crearon sus propias etiquetas de compatibilidad de productos. Como resultado, HTML se convirtió en un monstruo grande e inapropiado que perdió su estructura rigurosa.

En 1996, el W3C buscó una manera de aplicar la flexibilidad y el poder de SGML en la WEB, lo que llevó al lanzamiento de la especificación XML 1.0 en febrero de 1998. XML (eXtensible Markup Language, lenguaje de marcado extensible) tiene las características principales de SGML y es muy conciso. El contenido de la especificación XML es incluso menos de una décima parte de SGML. La simplicidad de XML es evidente. La aplicación de XML se multiplicó de inmediato y se la consideró la estrella en ascenso de la TI.

La diferencia entre SGML y HTML y entre XML y HTML es enorme. HTML puede considerarse como una aplicación específica de SGML/XML, mientras que SGML/XML como metalenguaje puede definir innumerables lenguajes de marcado nuevos, como DocBook. La diferencia entre SGML y XML es sutil. Como subconjunto de SGML, XML hereda las ventajas de SGML: extensibilidad,

estructura y eficacia.