

Fundamentos Recuperación en Internet

En este primer módulo veremos conceptos básicos de recuperación documental en la Web. La Web es actualmente el principal medio de recuperación de información, si bien los principios que se pueden ver en la misma son aplicables en cualquier red. La Web no obstante presenta unas características especiales: obsolescencia de localización y contenido, alto grado de repetición de contenidos, público y contenidos heterogéneos, sitios inseguros, etc.

Inicialmente nos centraremos en los distintos tipos de buscadores Web y su tipología. Todos los tipos de buscadores mencionados tienen una aplicación actual, si bien algunos como los directorios tuvieron su apogeo en los años 90. El interés de utilizar un buscador u otro depende básicamente de la importancia testimonial del documento, de su naturaleza única y de la audiencia a la que va dirigido.

La Web más ampliamente utilizada tiene, no obstante, deficiencias. No todos los contenidos son accesibles por los buscadores, por lo que se han desarrollado buscadores especiales. Así veremos los buscadores especializados en contenido multimedia. Estos buscadores por su naturaleza no están adaptados a los buscadores tradicionales. Tampoco lo está la llamada web invisible. Es decir recursos que, por estar tras un formulario o por no tener texto, no pueden ser accedidos automáticamente por una aplicación. En este caso también existen estrategias y buscadores especializados en estos contenidos.

La Web también evoluciona hacia una mayor capacidad semántica, una importancia creciente de los datos sobre los documentos, importancia de la personalización y la ubicuidad. En el futuro, estos desarrollos darán lugar a una Web no basada en páginas web solamente.

