1.2. Los sistemas de recuperación en internet

Las bases del éxito en la búsqueda y recuperación de información en internet son el conocimiento de los principios básicos de la recuperación de información y de los sistemas que la hacen posible, y de las características propias de los documentos existentes en internet. Las herramientas de búsqueda en internet aplican los principios sobre tratamiento y recuperación de información textual que se han revisado en el apartado anterior, y los usuarios disponen de similares prestaciones para la recuperación, y para su consulta y filtrado. Por lo tanto, **resulta crucial que el usuario conozca los tipos de información, la variabilidad de formatos y las diferentes presentaciones que puede adoptar la información en internet.** Ello le dotará de una mayor capacidad para conocer y valorar los resultados obtenidos durante el proceso de búsqueda.

Si bien un sistema de recuperación, en su formulación clásica, trabajaba sobre corpus documentales bastante homogéneos, no puede decirse lo mismo de los sistemas de recuperación en internet. Al tratarse de un entorno abierto y cambiante, las herramientas de búsqueda ofrecen listados de resultados, que dirigen al usuario hacia el documento original. Los cambios que se producen, por la propia dinámica del web, hacen que en ocasiones esa redirección no ofrezca los resultados esperados, y que haya que completar la búsqueda mediante procesos de exploración basados en la navegación. El usuario siempre debe pensar que no es suficiente, en recuperación de información en internet, con seguir los resultados obtenidos de un motor de búsqueda: hay que explorarlos, analizarlos, valorarlos, y seleccionarlos como adecuados, o desecharlos como no pertinentes. Los sistemas de recuperación de información en el web son un medio más, una fase intermedia, no un fin.

Una cuestión que debe tenerse en cuenta cuando se busca información en internet es que, contra la extendida creencia, **no todo está disponible a través de los motores de búsqueda**, ni en Wikipedia. La puesta en línea a través de internet,desde la década de 2000, de un gran número de fuentes y recursos de información, no supuso que su contenido fuese automáticamente incorporado al contenido procesado por los motores de búsqueda. Diferentes intereses comerciales y/o limitaciones técnicas excluyen enormes volúmenes de información de la vigilancia de los motores, configurando lo que se ha dado en llamar la “***internet invisible***”.

Copyright 2025

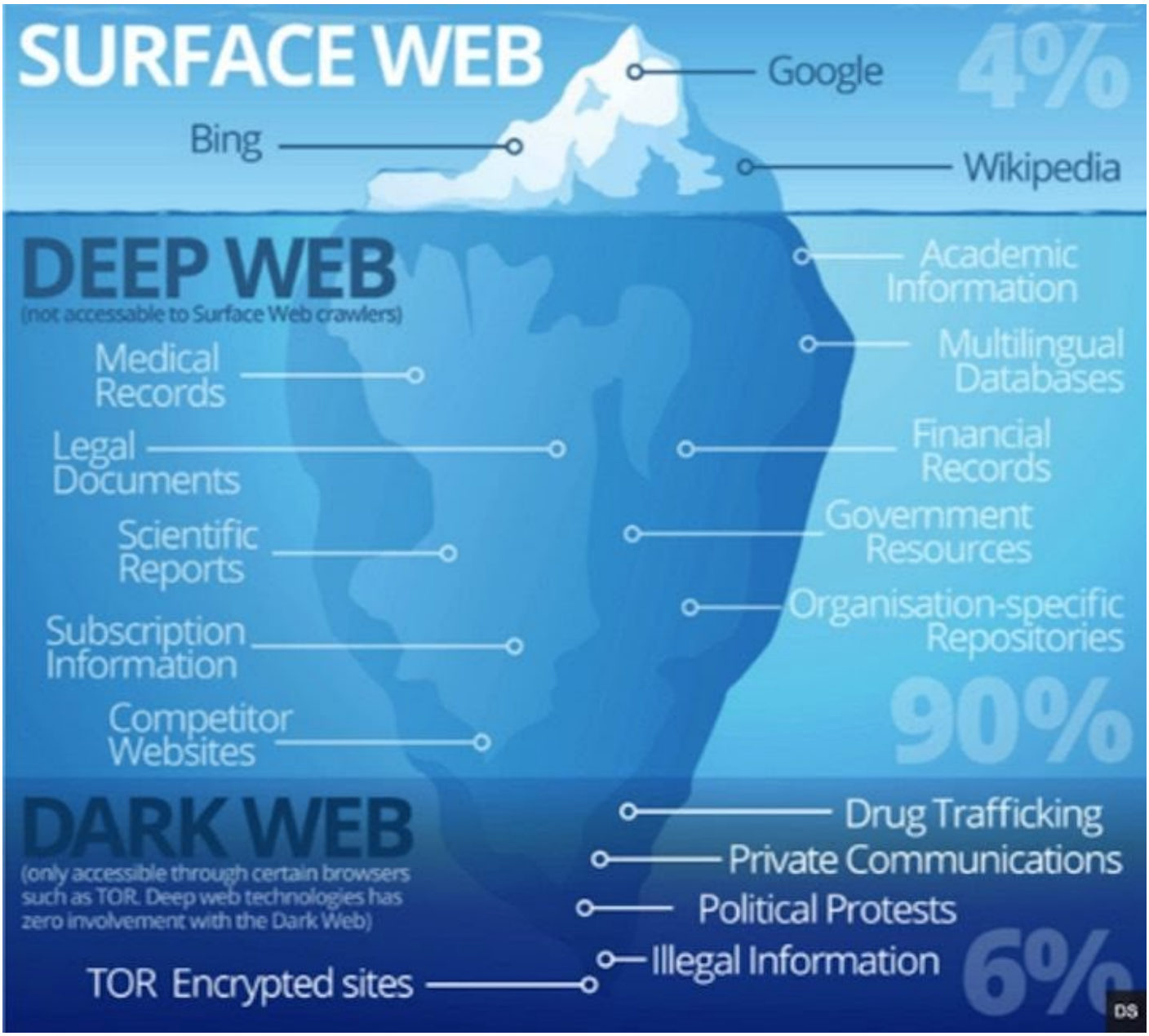

Fig. 2. El clásico iceberg de internet (múltiples fuentes)

En realidad, estos contenidos no son invisibles para el usuario: lo son para los motores. La noción de internet invisible se asocia a la presencia en la red de recursos de información, cuyo contenido sólo está disponible a través de los sistemas de recuperación que ofrecen los propios recursos. Esto es debido precisamente a que, a su vez, esta internet invisible se encuentra recogida en bases de datos que sólo muestran su contenido cuando son interrogadas, generando páginas web dinámicas, que evidentemente no pueden ser descubiertas y analizadas por los robots que utilizan los buscadores tradicionales. Dentro de la esta área invisible se engloban los directorios y las bases de datos especializadas, los catálogos de bibliotecas, archivos y museos, las bases de datos de prensa, etc. La conclusión lógica que se deriva de ello es que **el usuario debería conocer aquellos recursos de información especializada que resulten más adecuados para sus**

necesidades. Una aproximación común es comenzar la búsqueda en un motor generalista, para completarla en recursos especializados en una segunda fase.

Material complementario

- Búsqueda y recuperación de información en la web: qué ha pasado y qué podemos esperar en el futuro (2011)

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
GOBIERNO DE ARAGON

Revision #14
Created 27 October 2022 18:31:12 by Jesús Tramullas
Updated 6 November 2023 14:02:24 by Javier Anzano