1.1. Los sistemas de recuperación de información

El tratamiento y recuperación de la información ha sido una preocupación, y una actividad, que han llevado a cabo todas las sociedades. Desde una perspectiva académica, los fundamentos de esta disciplina se establecen progresivamente desde mediados del siglo XIX, cuando se establecen y se formalizan las enseñanzas superiores sobre bibliotecas, archivos y museos. Sin embargo, es necesario esperar hasta mediados del siglo XX para que se conforme la denominada *Information Science*, que traslada el centro de la actividad científica y profesional hacia diferentes sistemas de tratamiento y recuperación de información, mediados tecnológicamente, y cuyo objetivo final es satisfacer las necesidades de información de los usuarios.

La multiplicación del volumen de información científica y técnica, derivada del esfuerzo de la Segunda Guerra Mundial, trajo como consecuencia la necesidad de introducir máquinas que facilitasen el tratamiento, almacenamiento y recuperación de la información. El desarrollo del sistema SMART (*System for the Mechanical Analysis and Retrieval of Text*) por el equipo de G. Salton en la Cornell University, en la década de 1960, estableció los fundamentos de los modernos sistemas de recuperación de información, de los que los motores de búsqueda en internet son herederos. Las décadas entre 1970 y 1990 vieron un desarrollo progresivo de este tipo de sistemas de recuperación, así como la aparición de sistemas comerciales de pago, accesibles a través de redes de comunicaciones, que ofrecían principalmente acceso a bases de datos documentales de contenido científico, económico o financiero. Cuando se populariza el acceso a internet, a mediados de la década de 1990, ya existía un mercado previo y grupos de usuarios especializados en la búsqueda de información. Los motores de búsqueda en internet de la primera generación (*AltaVista*, *Lycos...*) pusieron al alcance de cualquier usuario capacidades hasta entonces limitadas a entornos cerrados y especializados. La aparición de *Google* en 1998 supuso la disponibilidad de una herramienta de fácil comprensión y uso para que cualquier usuario pudiese buscar y localizar información en internet.

Los sistemas de recuperación de información son aquellos que ofrecen al usuario funcionalidades para acceder a fuentes y recursos de información en entornos digitales, y consultar, recuperar y extraer de los mismos aquellos documentos cuyo contenido sea capaz de dar respuesta una cuestión planteada por el usuario. En muchas ocasiones, estos sistemas se presentan integrados en otros, por ejemplo como componentes de sistemas de gestión de documentos o sistemas de gestión de contenidos.

Copyright 2026

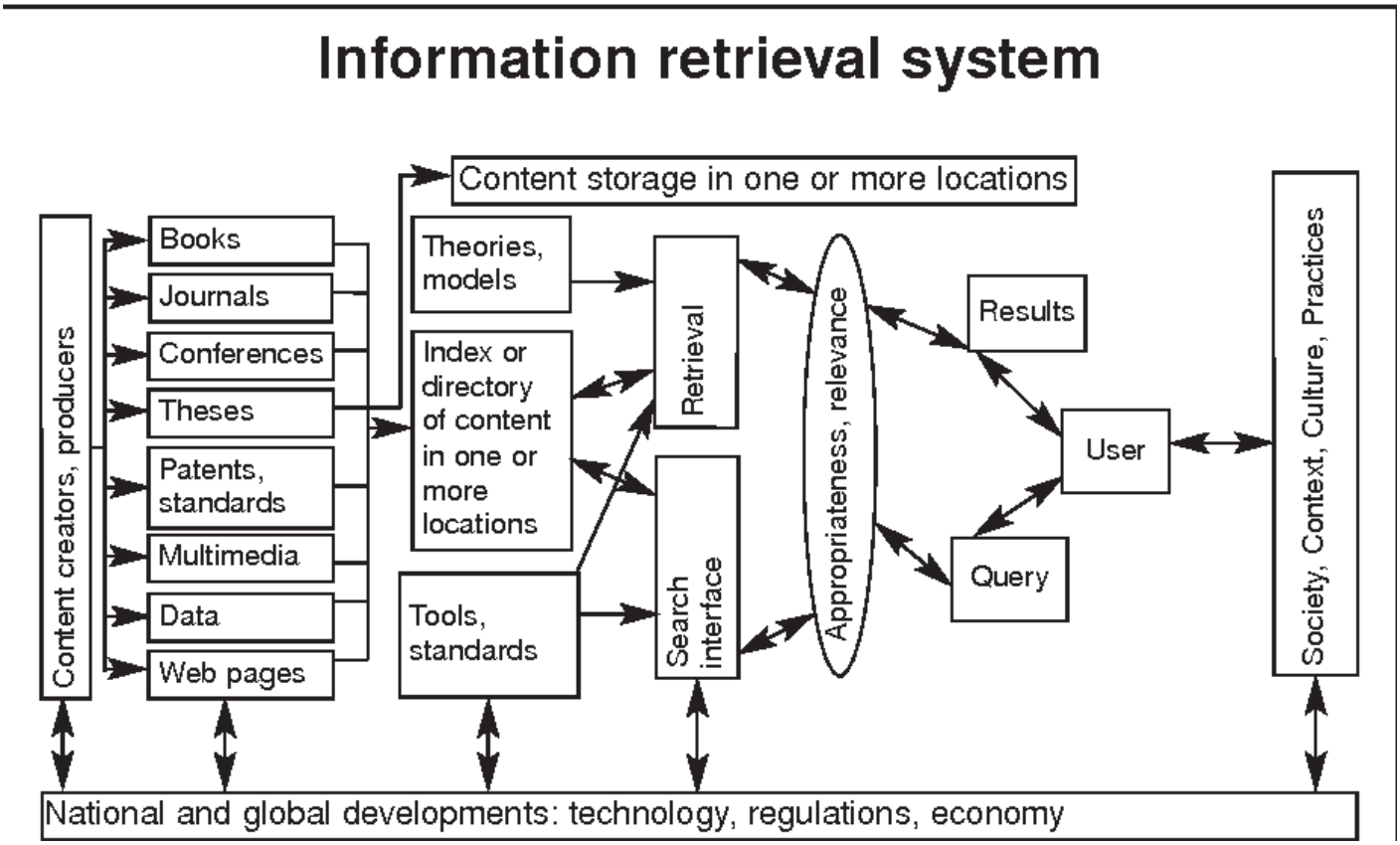


Fig. 1. Arquitectura de un sistema de recuperación de información (fuente original)

El factor clave que identifica a estos sistemas es su capacidad para procesar información textual, tanto en lo que se refiere a su adquisición y tratamiento, como en lo que respecta a la recuperación de la información contenida en el mismo. Esta información textual se recoge en documentos, que no suelen tener una estructura claramente formalizada (a diferencia de una hoja de cálculo o una tabla de base de datos, por ejemplo), y que pueden ofrecer múltiples combinaciones de contenido. La estructura funcional de un sistema de recuperación de información responde a:

- **Captura de información**: funcionalidades para capturar y almacenar documentos en diferentes formatos, para formar un corpus documental.
- **Procesamiento de información:** funcionalidades y algoritmos para generar representaciones de los documentos originales, según diferentes aproximaciones y criterios.
- **Recuperación de información**: funcionalidades y prestaciones para la formulación de expresiones o ecuaciones de búsqueda complejas, con una sintaxis propia y definida, y para la ejecución de esas expresiones contra el conjunto de representaciones resultado del procesamiento previo.
- **Salida de información:** las funcionalidades de presentación de resultados y de manipulación (filtrado, etc.), que se ofrecen al usuario con los resultados obtenidos de la ejecución de las expresiones de búsqueda.

Copyright 2026

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



Revision #6
Created 2022-10-27 18:29:01 CEST by Jesús Tramullas
Updated 2023-01-17 16:19:54 CET by Equipo CATEDU

Copyright 2026