

Historia de las bases de datos en Ciencia de la Información

De WikiJaveriana, la enciclopedia libre.

Objetivo General

Mostrar una evolución histórica de las bases de datos haciendo un énfasis en aquellas desarrolladas en Ciencia de la Información, que sirva como herramienta de consulta por los estudiantes, profesores y profesionales de tal manera que se tenga una conciencia de las diferentes bases de datos existentes y el uso que podamos hacer.

Concepto de bases de datos

Las bases de datos son una herramienta de vital importancia para el desarrollo de la actividad profesional. Existen dos colectivos que a menudo son el mismo especialmente beneficiados por las bases de datos. Por una parte, los investigadores y, por otra, los profesionales. Se podría considerar como base de datos cualquier recopilación organizada de información sobre la que haya habido análisis documental y que disponga de un sistema de búsqueda específica. Partiendo de esta idea, se podrían incluir los catálogos bibliotecarios. Sin embargo, en este trabajo se ha optado por emplear como criterio el seleccionar sólo aquellas bases formadas principalmente por descripciones de documentos impresos o electrónicos de libre acceso procedentes de revistas y artículos en Ciencia de la Información.

Historia de los sistemas de bases de datos

El uso de sistemas de bases de datos automatizadas, se desarrollo a partir de la necesidad de almacenar grandes cantidades de datos, para su posterior consulta, producidas por las nuevas industrias que creaban gran cantidad de información.

Herman Hollerit (1860-1929) fue denominado el primer ingeniero estadístico de la historia, ya que invento una computadora llamada "Máquina Automática Perforadora de Tarjetas". Para hacer el censo de Estados Unidos en 1880 se tardaron 7 años para obtener resultados, pero Herman Hollerit en 1884 creo la máquina perforadora, con la cual, en el censo de 1890 dio resultados en 2 años y medio, donde se podía obtener datos importantes como número de nacimientos, población infantil y número de familias. La máquina uso sistemas mecánicos para procesar la información de las tarjetas y para tabular los resultados.

Máquina_perforadora Herman_Hollerit(1860-1929) Tarjetas perforadoras

A diferencia con la maquina de Babbage, que utilizaba unas tarjetas similares, estas se centraban en dar instrucciones a la máquina. En el invento de Herman Hollerit, cada perforación en las tarjetas representaba un número y cada dos perforaciones una letra, cada tarjeta tenia capacidad para 80 variables. La máquina estaba compuesta por una perforadora automática y una lectora, la cual por medio de un sistema eléctrico leía los orificios de las tarjetas, esta tenía unas agujas que buscaban los orificios y al tocar el plano inferior de mercurio enviaba por medio del contacto eléctrico los datos a la unidad.

Este invento disparo el desarrollo de la tecnología, la industria de los computadores, abriendo así nuevas perspectivas y posibilidades hacia el futuro.

Década de 1950

En este lapso de tiempo se da origen a las cintas magnéticas, las cuales sirvieron para

suplir las necesidades de información de las nuevas industrias. Por medio de este mecanismo se empezó a automatizar la información de las nóminas, como por ejemplo el aumento de salario. Consistía en leer una cinta o más y pasar los datos a otra, y también se podían pasar desde las tarjetas perforadas. Simulando un sistema de Backup, que consiste en hacer una copia de seguridad o copia de respaldo, para guardar en un medio extraíble la información importante. La nueva cinta a la que se transfiere la información pasa a ser una cinta maestra. Estas cintas solo se podían leer secuencial y ordenadamente.

Década de 1960

El uso de los discos en ese momento fue un adelanto muy efectivo, ya que por medio de este soporte se podía consultar la información directamente, esto ayudo a ahorrar tiempo. No era necesario saber exactamente donde estaban los datos en los discos, ya que en milisegundos era recuperable la información. A diferencia de las cintas magnéticas, ya no era necesaria la secuencialidad, y este tipo de soporte empieza a ser ambiguo.

Los discos dieron inicio a las Bases de Datos, de red y jerárquicas, pues los programadores con su habilidad de manipulación de estructuras junto con las ventajas de los discos era posible guardar estructuras de datos como listas y árboles.

Década de 1970

Edgar Frank Codd (23 de agosto de 1923 – 18 de abril de 2003), en un artículo "Un modelo relacional de datos para grandes bancos de datos compartidos" ("A Relational Model of Data for Large Shared Data Banks") en 1970, definió el modelo relacional y publicó una serie de reglas para la evaluación de administradores de sistemas de datos relacionales y así nacieron las bases de datos relacionales.

A partir de los aportes de Codd el multimillonario Larry Ellison desarrollo la base de datos Oracle, el cual es un sistema de administración de base de datos, que se destaca por sus transacciones, estabilidad, escalabilidad y multiplataforma.

Inicialmente no se uso el modelo relacional debido a que tenía inconvenientes por el rendimiento, ya que no podían ser competitivas con las bases de datos jerárquicas y de red. Ésta tendencia cambio por un proyecto de IBM el cual desarrolló técnicas para la construcción de un sistema de bases de datos relacionales eficientes, llamado System R.

Edgar_Frank_Codd

IBM_Corporation

Larry Ellison

Década de 1980

Las bases de datos relacionales con su sistema de tablas, filas y columnas, pudieron competir con las bases de datos jerárquicas y de red, ya que su nivel de programación era bajo y su uso muy sencillo.

En esta década el modelo relacional ha conseguido posicionarse del mercado de las bases de datos. Y también en este tiempo se iniciaron grandes investigaciones paralelas y distribuidas, como las bases de datos orientadas a objetos.

Principios década de los 90

Para la toma de decisiones se crea el lenguaje SQL, que es un lenguaje programado para consultas. El programa de alto nivel SQL es un lenguaje de consulta estructurado que analiza grandes cantidades de información el cual permite especificar diversos tipos de

operaciones frente a la misma información, a diferencia de las bases de datos de los 80 que eran diseñadas para las aplicaciones de procesamiento de transacciones. Los grandes distribuidores de bases de datos incursionaron con la venta de bases de datos orientada a objetos.

Finales de la década de los 90

El boom de esta década fue la aparición de la WWW "Word Wide Web" ya que por éste medio se facilitaba la consulta de las bases de datos. Actualmente tienen una amplia capacidad de almacenamiento de información, también una de las ventajas es el servicio de siete días a la semana las veinticuatro horas del día, sin interrupciones a menos que haya planificaciones de mantenimiento de las plataformas o el software.

Siglo XXI

En la actualidad existe gran cantidad de alternativas en línea que permiten hacer búsquedas orientadas a necesidades específicas de los usuarios, una de las tendencias más amplias son las bases de datos que cumplan con el protocolo Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) los cuales permiten el almacenamiento de gran cantidad de artículos que permiten una mayor visibilidad y acceso en el ámbito científico y general.

Historia y evolución de los distribuidores de bases de datos con énfasis en Ciencia de la Información

Los comienzos de la industria online se remontan a principios de los años 70 en Estados Unidos, con tres organizaciones como pioneras: Lockheed Missiles and Space Company, System Development Corporation y National Library of Medicine. Con el paso de los años, se han ido incorporando nuevos e importantes distribuidores a esta industria, produciéndose cambios significativos en cuanto a nombres, fusiones y adquisiciones, sobre todo de algunos históricos (como BRS, Infoline, Data-Star) . Presentaremos algunos de los distribuidores más importantes que comercializan bases de datos en el área de Ciencia de la Información:

Dialog

La historia de Dialog está íntimamente ligada a la figura de Roger Kent Summit, considerado como el "padre" de la industria online, y que trabajó para la Lockheed Missiles and Space Company. En 1966, esta empresa obtuvo un contrato con la NASA (National Aeronautics and Space Administration) para desarrollar un software de recuperación que permitiera buscar de manera rápida y eficaz en sus bases de datos al que se le llamó llamado RECON (Remote Console). En 1969-1970, por encargo de la Atomic Energy Commission, comenzó a gestionar la colección de resúmenes de Nuclear Science.

En 1972, la Lockheed decide explotar su sistema de recuperación de información instalando dos grandes ordenadores que unidos a las principales redes de transmisión existentes (Telenet y Tymnet), dan servicio público tanto en USA como en Europa, convirtiéndose en el primer servicio comercial online del mundo. El sistema se denominó desde entonces Dialog, teniendo su sede central en Palo Alto (California). En agosto de 1988, Dialog es adquirido por Knight-Ridder Inc.; esta misma empresa compra en 1993 el servicio Data-Star, que era el distribuidor europeo más utilizado en 1985 y le estaba haciendo una gran competencia sobre todo en el mercado europeo. En 1995, Dialog

Information Services pasa a denominarse Knight- Ridder Information Inc., que en 1997 es comprada por M.A.I.D. plc, pasando a denominarse la fusión de las dos compañías The Dialog Corporation; y ya por último, en Abril del 2000, The Thomson Corporation compra la División de Servicios de Información de The Dialog Corporation, entre los que se encuentran Dialog y Data-Star.

Ovid

La historia de Ovid comienza en 1984, cuando su presidente y fundador Mark Nelson desarrolló el software original. En 1985 la compañía pasó a denominarse Online Research Systems, que Nelson cambió a CDP Plus en 1988, y a CDP Technologies Inc. en 1993. En 1994, la compañía adquiere BRS Online Services, y en 1995 cambia su nombre pasando a llamarse Ovid Technologies Inc. OCLC

En 1967, los presidentes de los colegios y universidades del Estado de Ohio fundaron OCLC (Ohio College Library Center) con el objetivo de desarrollar un sistema automatizado para que las bibliotecas de las instituciones académicas de Ohio pudieran compartir sus recursos y reducir costes. En 1977, se permitió que otras bibliotecas de fuera de Ohio se convirtieran en miembros de OCLC, pasando a denominarse OCLC Inc. En 1981, el nombre legal de la corporación pasó a ser el de OCLC Online Computer Library Center Inc. En la actualidad, OCLC sirve a más de 30.000 bibliotecas de USA y de otros 65 países. H.W. Wilson En 1984, H.W. Wilson Company introdujo su servicio online por primera vez, proporcionando acceso a sus propias bases de datos, convirtiéndose en un competidor importante de Dialog y BRS.

STN

En 1983, se crea STN (Scientific & Technical Information Network), como consecuencia de la operación conjunta de Chemical Abstracts Service (perteneciente a la American Chemical Society), FIZ Karlsruhe (de Alemania) y Japan Science and Technology Corporation, Information Center for Science and Technology (JICST). Ofrece acceso a bases de datos científicas, técnicas, de negocios y de patentes .

Historia de las bases de datos en Ciencia de la Información

Para el desarrollo de éste estudio se tuvieron en cuenta las bases de datos en Ciencia de la Información de acuerdo a los siguientes aspectos:

- Que las bases de datos sean de libre acceso y pública.
- Que recojan información de todo tipo de textos, especialmente de revistas tanto electrónicas como en formato impreso.
- Una cobertura y características suficientes como para dar respuesta a las peticiones de búsquedas más comunes.
- Se ha hecho especial hincapié en recopilar todas aquellas que tengan contenidos en español.

A continuación, se exponen brevemente los inicios y la evolución de algunos de los bases de datos importantes:

LISA - Library and Information Science Abstracts

LISA tiene una perspectiva internacional con una audiencia diversa incluyendo investigadores, estudiantes y profesionales de la información. Desde su inicio en 1969, LISA ha desarrollado una amplia gama de información relacionada con la bibliotecología, las teorías y prácticas de la Ciencia de la Información (LIS) Library Information Science,

cubriendo gran variedad de publicaciones.

METABASE

Desde 1973 se inician las gestiones para la creación del Colegio de Bibliotecarios de Costa Rica. La bibliotecaria Nelly Kopper asistió a un curso de la Universidad de Panamá donde fue designada, en una reunión conjunta entre los bibliotecarios centroamericanos y los profesores de dicho curso, para fundar una asociación de bibliotecarios costarricenses.

La Fundación Acceso, formuló un proyecto para consolidar la biblioteca electrónica especializada en medio ambiente y desarrollo sostenible para Centroamérica denominada MetaBase. Esta iniciativa fue sometida a consideración del programa INFODEV del Banco Mundial en el año 1996-1997. El proyecto fue aprobado en los siguientes dos años y su ejecución inició en el mes de julio del año 2000.

MetaBase, al ser una base de datos en Internet que contiene el conjunto total de registros bibliográficos de múltiples centros de información, le permite al usuario final realizar búsquedas de material bibliográfico, en forma simultánea, en todas las bases de datos de los centros participantes. El resultado de la búsqueda es una lista de los materiales relevantes y la referencia bibliográfica completa de cada uno de ellos

Es una poderosa herramienta de investigación que permite a los usuarios finales acceder y ubicar con facilidad recursos bibliográficos disponibles en diversas bibliotecas y centros de documentación en Centroamérica.

BEDOC

Elisa García-Morales y Carlota Bustelo Ruesta iniciaron junto con otros profesionales en el año 1984 el Gabinete de Asesores Documentalistas, empresa dedicada a prestar servicios en el área de la documentación. En 1996 crean Inforárea como parte de una reorientación estratégica que surge de las necesidades del mercado. Es así como dentro de sus productos internos crean BEDOC como una base de datos bibliográfica con toda la información referente a los temas que consultan. Cada trimestre se actualiza la información que tienen en su base de datos y permiten el acceso libre a los registros que contienen.

DOAJ - Directory of Open Acces Jounals

El Directorio de Revistas Open Access (Directory of Open Acces Jounals, DOAJ), fundado por el Open Society Institute - Budapest, es el directorio más amplio existente en Internet de revistas open access. Reside en las bibliotecas de la Universidad de Lund y está financiado por la coalición SPARC, SPARC Europa, BIBSAM, Axiell. DOAJ se ha puesto en ejecución en dos fases. Durante la primera, se estableció el directorio en sí mismo. La fase dos en el 2003 comenzó desarrollado un sistema comprensivo en la búsqueda de contenidos a nivel de artículos, de esta manera el sistema de búsqueda tuvo que proveerse de meta datos para la localización más descriptiva de la información. En el área de Ciencia de la Información (library Information Science) se encuentran indexadas 60 revistas de libre acceso, las cuales en su mayoría son de universidades, asociaciones y profesionales alrededor del mundo. El objetivo del Directorio de revistas Open Access (DOAJ) es incrementar la visibilidad y fomentar el uso de la literatura científica a través de las revistas científicas y académicas. La iniciativa Open Acces se define como un modelo en el que el acceso a la literatura científica de las revistas pertenecientes al DOAJ es gratuito tanto para los usuarios como para sus organizaciones.

ISOC - Biblioteconomía y Documentación

En la década de los 70, y como consecuencia del informe de la OCDE (Organización de Cooperación y Desarrollo Económico) sobre la política española en materia de información y documentación, se crea un órgano coordinador del Plan Nacional de Información Científica y Técnica como rector de todas estas actividades, con el nombre de Centro Nacional de Información y Documentación Científica (CENIDOC). El CENIDOC coordinaba tres Institutos orientados sobre grandes áreas del conocimiento: Ciencia y Tecnología, Biomedicina y Ciencias Sociales y Humanidades. De esta forma, en 1975, el CID se convierte en el Instituto de Información y Documentación en Ciencia y Tecnología (ICYT). En el mismo año, se crea, en el CSIC, el Instituto de Información y Documentación en Ciencias Sociales y Humanidades (ISOC), a partir del Departamento de Información Científica y Técnica del Instituto Bibliográfico Hispánico del Ministerio de Cultura. De igual forma se crea el de Biomedicina, sobre el Centro de Documentación e Informática Médica de Valencia.

La actuación de estos tres Institutos se basa en tres elementos fundamentales: investigación, docencia y servicios en el área de la Información y la Documentación Científica. Durante el año de 1975 se instalan en el ICYT y el ISOC, los primeros terminales para el acceso en línea a los grandes distribuidores de información: DIALOG, QUESTEL, ORBIT, etc. El ISOC publica por primera vez, en 1976, el Índice Español de Humanidades y el Índice Español de Ciencias Sociales. El ICYT inicia, en 1979, la edición del Índice Español en Ciencia y Tecnología. Estos tres Índices recogen, en forma de referencia bibliográfica y en sus respectivas áreas, los artículos publicados en las revistas científicas españolas y darán lugar a las Bases de Datos ISOC e ICYT respectivamente. En 1989 las Bases de Datos ICYT, de Ciencia y Tecnología, e ISOC, en Ciencias Sociales y Humanidades, empiezan a distribuirse en línea desde el Centro Técnico de Informática del CSIC. Las Bases de Datos del CSIC fueron, en 1990, el primer producto de información bibliográfica editado en CD-ROM. En 1992 se crea el Centro de Información y Documentación Científica (CINDOC) como resultado de la fusión del Instituto de Información y Documentación en Ciencia y Tecnología (ICYT) y el Instituto de Información y Documentación en Ciencias Sociales y Humanidades (ISOC) que asume de forma integradora los objetivos de ambos para potenciar la información científica de alta calidad en todos los campos del conocimiento.

E – LIS - E- Prints in Library and Information Science

E-LIS es el archivo abierto con más documentos en Biblioteconomía y Ciencias de la Información. Su propósito es poner a disposición de la comunidad científica el texto completo de documentos para que sean más visibles, accesibles, recuperables y útiles para cualquier usuario potencial con acceso a Internet. Además este servicio anima a los profesionales a que depositen sus documentos (publicados o no) en E-LIS para su acceso libre en Internet.

E-LIS es diferente de otras iniciativas similares en el sentido que está basado en el trabajo voluntario de profesionales en Biblioteconomía y Ciencias de la Información y Documentación y, a su vez, tiene una orientación no comercial. Se basa en el espíritu del open source initiative en informática en el que la gente trabaja conjuntamente en la construcción y desarrollo de un software que está en el dominio público.

En este sentido, E-LIS recoge esta filosofía y la aplica en la construcción de una biblioteca digital gratuita. E-LIS es administrado por un equipo internacional de voluntarios

compuestos de profesionales de países como: España, Italia, Estados Unidos de América, Bosnia y Herzegovina, Cuba, Croacia, Sudáfrica, India, Indonesia, Peru, Alemania, Nigeria, Brasil, Grecia, el Reino Unido o Serbia y Montenegro. Este fue formado en 2003 para el depósito de documentos en el área de las Ciencias de la Información (Library Information Science).

Es el primer e-servidor internacional en este tema y resultante del proyecto DoIS (documentos de RCLIS investigación en computación, bibliotecas y Ciencias de la Información, promovidos por el Ministerio Español de Cultura y recibidos por el equipo de AEPIC en los servidores del Conzorcio italiano Interuniversitario Lombardo por Elaborazione Automatica (CILEA). Actualmente se encuentra activa y se desarrolla como uno de las bases de datos de Open Acces de más acceso por los profesionales de la información.

DoIS - Documents and Information Science

El origen de DoIS se remonta hacia el año 2002 cuando un grupo interdisciplinario de bibliotecarios, economistas, diseñadores, libreros... de diversas partes del mundo, tuvieron la iniciativa de recoger, organizar y difundir literatura gris en el ámbito de la Ciencia de la Información. DoIS es un recurso bibliográfico gratuito de textos científicos especializados en el área de la Documentación Científica, que usa un sistema facilitado por RCLIS6 (Research in Computing, Library and Information Science), basado en una arquitectura distribuida, en la cual el trabajo de descripción y organización de documentos se realiza igualmente de manera distribuida entre los participantes del sistema.

Uno de los esfuerzos fundamentales de DoIS ha sido lograr la mayor visibilidad de la información. Por ello y para ello se incluyó en el directorio de recursos de calidad de Google, especializado en Library and Information Science. Ésta se ve además favorecida por la utilización del metadato ReDIF (Research Documents and Information Format). Se trata de un formato flexible que puede ser fácilmente modificado a medida que vayan surgiendo nuevas necesidades. Por otro lado es lo suficientemente simple como para que pueda ser utilizado por personal no especializado.

El tipo de material que recopila DoIS son artículos de revista, procedentes de las revistas de mayor impacto internacional en el área de la Documentación: Journal of the American Society for Information Science and Technology, Journal of Information Science, Journal of Documentation, Library and Information Science Research... , así como las más importantes en el área de la documentación en lengua hispana (Revista Española de Documentación Científica, Boletín de la ANABAD, Boletín de la Asociación Andaluza de Bibliotecarios, Anales de la Documentación, El profesional de la Información, Biblos, Ciencia da Informação... También se incluyen los trabajos presentados a los más importantes congresos celebrados en todo el mundo sobre nuestra disciplina (IFLA, Northumbria, ...). Y otros documentos de interés científico que se encuentran accesibles en la red, tales como normas, monografías, informes técnicos... También es importante destacar la destreza y rapidez en la actualización de la información contenida, respecto a otras bases de datos del mismo ámbito.

La sede central de DoIS10 se ubica en el servidor MIMAS de Manchester, donde se cargan los datos a través del servicio FTP (File Transport Protocol). DoIS utiliza el metadato ReDIF11 Resource Description Information Format), basado en el modelo IAFA, y el protocolo Guildford.

Actualmente DoIS proporciona información referencial y acceso público mediante enlaces

a texto completo, a un total de 1202 artículos y 3662 documentos de congresos, de los cuales 8529 son directamente accesibles. .

CONCLUSIONES

- Si bien es cierto que las bases de datos, son uno de los recursos más importantes en el para el desarrollo cualquier área del saber, el contexto histórico ha demostrado que son de una gran influencia el progreso de los países a nivel mundial, permitido un almacenamiento y acceso confiable, eficiente y practico en el uso de la información que se produce.
- Las bases de datos no solo han servido como repositorios de información, son a través de la historia una de las herramientas más importantes en el campo de la difusión científica.
- A nivel mundial existen muchas iniciativas propuestas por profesionales de la información, que a partir de un trabajo colaborativo e interdisciplinario, han dispuesto de herramientas para el uso y desarrollo de aquellos interesados en el área de la Ciencia de la Información.
- Si bien es sabido que la producción científica en Colombia en el área de Ciencia de la Información es escasa, debemos ser concientes que la responsabilidad de producir información que permita el desarrollo de nuestra ciencia, es de nosotros mismos como estudiantes y profesionales de la información, ya que como lo hemos visto, existen muchas bases de datos especializadas en nuestra área, en muchos casos de acceso libre, no teniendo excusas para la falta de producción científica.