



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO
FACULTAD DE INGENIERIA

DIRECTRICES A SEGUIR PARA LA
STANDARIZACION DE BASES DE DATOS
DE BIBLIOTECAS UNIVERSITARIAS

TESIS QUE PARA OBTENER EL GRADO DE ING. EN
COMPUTACION PRESENTA:

VICTOR MANUEL MONTIEL SOLIS

DIRECTOR: MTRA. LUZ MARINA QUIROGA

MEXICO D.F., NOVIEMBRE DE 1990

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TESIS CON FALLA DE ORIGEN

INDICE

I	INTRODUCCION -----	1
	I.1.-Objetivo -----	1
	I.2.-Definición de Base de Datos (B.D.) -----	1
	I.3.-Necesidad de un standard en las B.D. de bibliotecas de la U.N.A.M. -----	2
	I.4.-Explicación de la estructura de la tesis -----	3
II	SELECCION DE BIBLIOTECAS Y VARIABLES -----	4
	II.1.-Selección de bibliotecas representativas del medio universitario -----	4
	II.2.-Selección de variables a manejar -----	5
III	BIBLIOTECA NACIONAL -----	9
	III.1.-introducción -----	9
	III.2.-Antecedentes de la Automatización -----	10
	III.3.-Descripción de la B.D. -----	11
	III.3.1.-Diseño -----	11
	III.3.2.-Software -----	15
	III.3.3.-Hardware -----	21
	III.3.4.-Humanware -----	22
IV	DIRECCION GENERAL DE BIBLIOTECAS -----	24
	IV.1.-introducción -----	24
	IV.2.-Antecedentes de la Automatización -----	25
	IV.3.-Descripción de la B.D. -----	27
	IV.3.1.-Diseño -----	27
	IV.3.2.-Software -----	39
	IV.3.3.-Hardware -----	48
	IV.3.4.-Humanware -----	49

V	CENTRO UNIVERSITARIO DE INVESTIGACIONES BIBLIOTECOLOGICAS	51
	V.1.-introducción	51
	V.2.-Antecedentes de la Automatización	52
	V.3.-Descripción de la B.D.	53
	V.3.1.-Diseño	53
	V.3.2.-Software	55
	V.3.3.-Hardware	60
	V.3.4.-Humanware	64
VI	INSTITUTO DE INVESTIGACIONES EN MATEMATICAS	
	APLICADAS Y SISTEMAS	66
	VI.1.-introducción	66
	VI.2.-Antecedentes de la Automatización	67
	VI.3.-Descripción de la B.D.	68
	VI.3.1.-Diseño	68
	VI.3.2.-Software	72
	VI.3.3.-Hardware	78
	VI.3.4.-Humanware	82
VII	CONCEPTOS PARA LA STANDARIZACION DE LAS	
	DIFERENTES B. D.	83
	VII.1.-Problemas a solucionar	83
	VII.2.-Directrices a seguir para una	
	standarización de B.D. de	
	bibliotecas universitarias	86
VIII	CONCLUSIONES	97
	APENDICE: El Modelo de Datos Relacional de la IDM 500	101
	GLOSARIO DE TERMINOS	113
	BIBLIOGRAFIA	117

I.1 Objetivo

Se persigue en esta tesis exponer el estado en que se encuentran las diferentes Bases de Datos (B.D.) de cuatro bibliotecas de la U.N.A.M., para obtener conocimiento suficiente que permita formular directrices para una mayor compatibilidad (standarización) entre las B.D. de las bibliotecas universitarias. Para ello los productos de la tesis serán una serie de variables representativas de cada B.D. con las cuales se puedan realizar comparaciones, así como las conclusiones que se desprendan del estudio realizado que establezcan las bases de las directrices mencionadas.

I.2 Definición de Base de Datos

Para adentrarnos en el tema de la presente tesis es indispensable desde un principio contar con el concepto de base de datos.

Podemos definirla como un depósito de datos que se interrelacionan entre sí y que se almacenan agrupados.

Para C. J. Date es "un repositorio de datos almacenados, y en general es tanto integrada como compartida". (C. J. Date Introducción a los sistemas de base de datos, Addison-Wesley Iberoamericana, 1986). Por integrada se entiende que la base de datos elimina cualquier redundancia entre los datos que la conforman, esto indica que se eliminan las repeticiones innecesarias. Por compartida se indica que a una misma zona de la base de datos pueden tener acceso varios usuarios distintos.

I.3 Necesidad de un standard en las bases de datos de bibliotecas de la U.N.A.M.

En las dos últimas décadas la cantidad de información publicada ha ido en creciente aumento a tal grado que los procesos manuales para el manejo de los materiales bibliográficos han generado grandes rezagos en muchas bibliotecas. La automatización ha sido una de las medidas que más demanda ha tenido para la solución de los problemas que acarrea la manipulación de grandes volúmenes de documentos. A raíz de ello las bibliotecas más poderosas de la UNAM han generado sistemas computacionales para enfrentar esta situación, pero también han salido al mercado paquetes y sistemas orientados a procesos bibliotecarios que pueden montarse incluso en microcomputadoras.

Se estima que el diseño y creación de las bases de datos (B.D.) bibliográficas se realiza aisladamente en cada uno de los diversos centros de la UNAM como se ha observado, lo cual duplica esfuerzos y mantiene un desconocimiento general sobre la tarea que los demás desarrollan en este campo, sin contar con la ayuda que daría un documento que indicara sobre las características, capacidades, resultados y limitaciones que poseen las b.d. de las instancias más relevantes en el materia dentro de la Universidad. Este es el motivo que inclina al tesista a realizar el presente trabajo, quien cuenta con varios años de laborar en la Dirección General de Bibliotecas dentro de la Subdirección de Informática en calidad de analista - programador.

I.4 Explicación de la estructura de la Tesis

En esta introducción se indican las razones que impulsaron al tesista a realizar el presente trabajo. También se define el significado de base de datos que es el tema que se tratará a lo largo de la tesis.

El capítulo II nos muestra las razones que se utilizaron para escoger a las bibliotecas del estudio, así como los parámetros que se utilizarán como método para que el diseñador pueda valorar las bases de datos de las bibliotecas propuestas.

Del capítulo III al VI se describen las diferentes bases de las bibliotecas escogidas.

El capítulo VII muestra finalmente la proposición de las directrices tomadas una vez que se analizaron las distintas descripciones de las b.d.

Por último en el capítulo VIII se encuentran las conclusiones obtenidas al final del trabajo realizado.

Al término de esta tesis se encuentra un apéndice que trata sobre el modelo relacional orientado a la IDM 500 que espero será de utilidad para los interesados en este trabajo.

II

SELECCION DE BIBLIOTECAS Y VARIABLES

II.1 Selección de Bibliotecas representativas del medio bibliotecario.

Se eligió la muestra buscando que:

- Incluyera las bibliotecas más representativas de los acervos de la U.N.A.M.
- Se Usaran diferentes manejadores de B.D.
- Se tomaran en cuenta sólo las B.D. catalográficas debido a que son más rigurosas y completas en su contenido.

Las bibliotecas seleccionadas fueron:

- Biblioteca Nacional (BN)
- Dirección General de Bibliotecas (DGB)
- Centro Universitario de Investigaciones Bibliotecológicas (CUJIB)
- Instituto de Investigaciones en Matemáticas Avanzadas y Sistemas (IIMAS)

Biblioteca Nacional (BN) se eligió debido al importante lugar que ocupa dentro de la Universidad, ya que es la depositaria de todo el material que se imprime en el país, así como al gran volumen de su acervo. La B.D. que se maneja en esta institución se basa en el sistema MINISIS

La Dirección General de Bibliotecas (DGB) fue escogida porque

rige a 214 bibliotecas pertenecientes a Preparatorias, Facultades, Posgrados, Direcciones, Centros e Institutos. La B.D. se manipula con una máquina especializada IDMS00.

El Centro Universitario de Investigaciones Bibliotecológicas (CUIB) fue seleccionado por tratarse del centro donde se realiza investigación bibliotecológica que posteriormente incide en el desarrollo de las bibliotecas universitarias. La B.D. a analizar se opera con el sistema LOGICAT basado en DBase III+.

Instituto de Investigaciones en Matemáticas Avanzadas y Sistemas (IIMAS). Se optó por este instituto debido a su acervo especializado guardado en una B.D. que se explota con el sistema MICROISIS.

II.2 Selección de variables a manejar

La selección de variables estuvo fuertemente influenciada por la configuración de los paquetes MINISIS y MICROISIS debido principalmente a tres motivos.

--Son paquetes altamente especializados en el manejo de procesos bibliotecarios, lo que ahorra programación e implementación y cubren casi todas las aspiraciones de las bibliotecas.

--Son populares en los medios bibliotecarios universitarios, lo que ha permitido la formación de grupos de usuarios tanto nacionales como foráneos, con los cuales se comparten experiencias y dudas.

--Son sistemas que se distribuyen gratuitamente a las

Bibliotecas universitarias y dependencias gubernamentales, lo que explica también su popularidad pues no es necesario desviar recursos monetarios de las partidas presupuestales para su obtención.

Esto hace que la selección de variables se hayan basado fundamentalmente en las características de MINISIS y MICROISIS pues para que tengan un valor práctico y útil al diseñador de b.d. bibliotecarias, se utilizará como regla a los sistemas más populares.

A continuación se enlistan las variables escogidas, agrupadas según sus características:

1.--SOFTWARE

a.--Base de Datos

1).--tipo

2).--capacidad

- # regs en b.d.
- tamaño máximo de reg.
- máximo # de campos por reg.
- máximo # de subcampos por campo.
- longitud máxima de una clave

3).--características generales

- posibilidad de tratar un subconj. de regs como una b.d.
- facilidad de uniones
- conjunto de caracteres en difs. idiomas
- ordenamiento de acuerdo a inf. adicional
- campos de long. fija
- campos de long variable
- subcampos repetibles
- campos obligatorios

b.--DBMS

1).--tipo

2).--lenguaje

3).--mantenimiento

- bitácora de transacciones para recobrase de caídas del sistema
- bitácora de transacciones para contabilidad
- verificación de tipo de dato
- rutinas de validación propias del usuario
- verificación de campos con valores únicos

- tipo de archivos invertidos
- tipos de extracción
- eliminación de palabras no significativas
- actualización de archivos invertidos

4).-expansión

- lenguajes
- generación de aplicaciones completas
- contribuciones de usuarios
- relaciones con otros manejadores
- mejoras previstas

5).-aplicaciones

- aplicaciones típicas
- aplicaciones futuras

6).-ordenamiento

- máximo número de niveles
- ordenamiento de caracteres españoles
- remoción de palabras stop
- conservación de etiquetas de campo
- selección de campos alternativos
- procesamiento de índices kwic
- manejo de sufijos/prefijos
- especificación de # máximo de claves que se extraerán

7).-características generales

- identificación de campos
- # de archivos invertidos
- búsqueda en varias bases de datos
- búsqueda libre de textos
- operaciones permitidas en cálculo numérico

8).-portabilidad

c.-Paquetes, programas de explotación.

1).-lenguaje

2).-captura datos

- definición de pantalla de captura
- facilidad para edición
- dandado a campos que se están modificando
- manejo de valores default
- operadores

3).-productos resultantes

- creación automática de un formato de impresión
- truncamiento
- función dot
- manejo de tesauros
- despliegue de resultados de búsqueda
- páginas lógicas
- tabulados/columnas
- máximo # de regs. por pag.
- corte a medio reg.

- líneas de encabezado
- encab. que sea parte de los datos
- selección de campo alternativo si uno no existe
- especificación de cual ocurrencia imprimir
- márgenes, sangrías
- supresión de campo si el reg. previo tiene igual contenido
- ordenamiento alfabético de las ocurrencias de un campo repetible
- impresión de literales

4).-portabilidad

d.-Obtención y costos del Software

- 1).-proveedor
- 2).-costo
- 3).-equipo
- 4).-sistema operativo
- 5).-facilidad de obtención

2.-HARDWARE

- a.-Tipo
- b.-Características Generales
- c.-Obtención y Costos

3.-HUMANWARE

- a.-Capacitación de personal
 - 1).-responsables
 - 2).-material didáctico
 - 3).-capacitación a instructores
- b.-Documentación de los sistemas
 - 1).-documento introductorio
 - 2).-calidad de manuales
 - 3).-traducción al español de documentación
- c.-Grupos de usuarios
 - 1).-objetivos
 - 2).-# de participantes
 - 3).-actividades
 - 4).-vínculo con centro responsable
- d.-soporte del proveedor (centros responsables)

III.1 Introducción

Para comprender las razones que implicaron la automatización, es necesario conocer la importancia que tiene la Bibliografía mexicana.

La Bibliografía mexicana muestra el conjunto de documentos de todo tipo, que indican el pulso nacional en un momento determinado de la historia patria.

Debido a ello la Bibliografía mexicana se ha convertido en una herramienta de gran utilidad para investigadores, eruditos y público en general, como el contenedor de información que permite estudiar, analizar y en cierto grado predecir los movimientos, opiniones y corrientes sociales, políticos, científicos y culturales, así como todo tipo de ideas, convicciones y gustos que afectan a la población del país.

Ello ha desembocado en la compilación de la Bibliografía Nacional que comenzó a publicarse anualmente, con el nombre de Anuario Bibliográfico", a partir del año de 1958, registrando en siete años un promedio de 5000 fichas de tipo bibliográfico descriptivo correspondiendo en su mayoría a obras monográficas publicadas en el Distrito Federal. En 1967 la publicación tomó su nombre actual "Bibliografía Mexicana" y comenzó a aparecer bimestralmente, logrando compendiar hasta 4000 fichas anuales. Desde 1979 se

modificó profundamente la estructura de compilación de cada fascículo, pues las fichas son ahora de tipo catalográfico, y se basan en las reglas de Catalogación Angloamericanas, ampliando su campo de acción al abarcar, aparte de obras monográficas, nuevos materiales tales como mapas, discos, partituras, publicaciones periódicas, tesis, patentes, etc., y cambiando su frecuencia de aparición en 1981 siendo hasta nuestros días una publicación mensual, conteniendo en cada fascículo un promedio de 750 fichas.

III.2 Antecedentes de la Automatización

En 1979 un grupo de trabajo presentó al coordinador de Humanidades el documento "El Sistema BIBA (Banco de Información Bibliográfica Automatizada)" junto con el "Proyecto de diseño de publicación de la Bibliografía Mexicana Automatizada" donde se establecieron los objetivos del sistema que a continuación se listan:

- Realizar la catalogación y consulta en línea
- Evitar duplicación de tareas tal como la transcripción de datos en borradores, matrices, tarjetas, relaciones de envío, etc., con la consecuente agilización del procesamiento técnico del material.
- Establecer programas de intercambio de información, transferida ya sea a través de cintas en formato MARC (Machine Readable Cataloging) o a través de redes de telecomunicación.
- Obtener subproductos tales como bibliografías

especializadas, índices acumulados, tarjetas catalográficas y el fascículo mensual de la Bibliografía Mexicana.

-Obtener estadísticas y evaluaciones

En 1985 se organiza formalmente el equipo de cómputo seleccionándose para la descripción bibliográfica el formato MARC, dentro del ámbito del software se optó por el manejador de base de datos MINISIS debido a su alto grado de especialización en el manejo de datos bibliográficos, y en hardware se eligió una computadora HF3000/42 con su unidad de cinta, unidad de disco, 2 terminales exclusivas para el uso de Bibliografía Mexicana, una impresora de 300 líneas por minuto, una microcomputadora PC y una impresora láser de ocho páginas por minuto en la cual se imprime el fascículo.

III.3 Descripción de la Base de Datos

III.3.1 Diseño de la base de datos monográfica

El diseño quedó formado por una base de datos de especificaciones (diccionario), y varias bases de datos que contienen la información bibliográfica.

La parte principal del análisis realizado para el diseño se basó en el estudio del formato MARC para definir los elementos de información que integrarían la base de datos, la cual en una primera fase abarcaría obras monográficas. Esta base (llamada BIELB) está formada por 98 campos agrupados en:

- a) Campos para datos de control administrativo del trabajo como son las claves de quien investiga, cataloga y captura, lo

mismo que las fechas en que se hizo este trabajo.

b) Campos para datos bibliográficos que requieren investigación a fin de ser normalizados y que se consignan en una hoja de codificación antes de ingresar a la base. Entre estos datos algunos tienen longitud fija tales como códigos de país de publicación, idioma, contenido, audiencia a que está dirigida la obra, tipo de obra, etc. Otros tienen longitud variable, ejemplo de ellos son la clasificación y los asientos principales y secundarios de autor, título, temas y series.

c) Campos para datos bibliográficos que no requieren investigación (catalogación descriptiva)

Para manejar eficazmente las características de cada elemento, se construyó asimismo una base de datos con información sobre los datos (diccionario) en la que se guarda información sobre: Nombre del campo de información, longitud máxima de ese campo, si es numérico o alfabético, si es repetible o no, que subcampos permite y cuales son los valores que acepta con fines de normalización, se puede almacenar una descripción sobre el uso que se dará a ese campo y finalmente este diccionario permite guardar información sobre si ese campo se utiliza en otro formato tal como MARC/DGB (formato usado por la base de datos de DGB).

Con este diccionario es posible generar las bases de datos bibliográficas, por lo que las modificaciones hechas a las mencionadas bases se realizan sólo en el diccionario, el cual mediante un programa actualiza las definiciones de las bases de datos.

En la siguiente página se listan los campos de la base BIBLE:

DATA BASE: BIBLB

FIELD	NAME	MNEMONIC	TAG	T	H	C	O L N D R S P						
							F	E	U	E	U	R	
							S	N	M	C	P	B	O
							E	G	E	E	F	M	
							E	T	R	P	A	I	P
							T	H	I	L	T	E	T
													L
													D
CODIGO DE CASO		W000	CASO	W000	-1	2							Y
SALA A LA QUE SE ENVIA ESTE M		W050	SALA	W050	-1	1							Y
NOMBRE DE QUIEN INVESTIGA DAT		W100	INVEST	W100	-1	3							Y
FECHA DE INVESTIGACION		W150	FECHIN	W150	-1	10							Y
NOMBRE DE QUIEN DA V. B. INVE		W200	VBINV	W200	-1	3							Y
FECHA DE V. B. DE INVESTIGACI		W250	FUBINV	W250	-1	10							Y
NOMBRE DE QUIEN HACE CATALOGA		W300	CATALO	W300	-1	3							Y
FECHA DE CATALOGACION		W350	FECHCA	W350	-1	10							Y
NOMBRE DE QUIEN REVISA CAPTUR		W400	REVCAP	W400	-1	10							Y
FECHA DE REVISION DE CAPTURA		W450	FREVCA	W450	-1	10							Y
NOMBRE DE QUIEN CORRIGE CAPTU		W500	CORCAP	W500	-1	10							Y
FECHA DE CORRECCION DE CAPTURA		W550	FCORCA	W550	-1	10							Y
NOMBRE DE QUIEN DA V.B. FINAL		W600	VOBO	W600	-1	10							Y
FECHA DE VISTO BUENO FINAL		W650	FVOBO	W650	-1	10							Y
CAMPOS FIJOS PARA MONOGRAFIAS		M080	M080	M080	-1	40							Y
CODIGO DE IDIOMA		B040	IDIOMA	B040	-1	30							Y
CODIGO DE AREA GEOGRAFICA		B050	GEOGRA	B050	-1	50						Y	Y
COD. CRONOLOGICO FECHA/TIEMPO		B060	CRONO	B060	-1	50					Y		Y
NUMERO DE CLASIFICACION DEWEY		B180	DEWEY	B080	-1	50							Y
PRINCIPAL/AUTOR PERSONAL		B100	PPERSO	B100	-1	1000							Y
PRINCIPAL/AUTOR CORPORATIVO		B110	PCORP	B110	-1	1000							Y
PRINCIPAL/CONFERENCIA		B120	PCONFE	B120	-1	1000							Y
PRINCIPAL/TIT. UNIF.(ANONIMO)		B130	PUNIF	B130	-1	1000							Y
TIT. UNIFORME C/ASIENTO DE AU		B200	TUNIF	B200	-1	1000							Y
TITULO UNIFORME COLECTIVO		B230	TUNIFC	B230	-1	1000							Y
TITULO		B240	TITULO	B240	-1	1000							Y
LITERAL NUM. ROMANO P/TITULO		B280	ROMTIT	B280	-1	5							Y
NOTA/TESIS		B520	TESIS	B520	-1	1000					Y		Y
SEC. TEMA/AUTOR PERSONAL		B600	TPERSO	B600	-1	1000					Y		Y
SEC. TEMA/AUTOR CORPORATIVO		B610	TCORPO	B610	-1	1000					Y		Y
SEC. TEMA/CONFERENCIA-REUNION		B620	TCONFE	B620	-1	1000					Y		Y
SEC. TEMA/TITULO UNIFORME		B630	TTUNIF	B630	-1	1000					Y		Y
SEC. TEMA/ASPECTO TEMATICO		B650	TEMA	B650	-1	1000					Y		Y
SEC. TEMA/NOMBRE GEOGRAFICO		B660	TGEOGR	B660	-1	1000					Y		Y
SEC./AUTOR PERSONAL		B700	SPERSO	B700	-1	1000					Y		Y
SEC./AUTOR CORPORATIVO		B710	SCORPO	B710	-1	1000					Y		Y
SEC./CONFERENCIA O REUNION		B720	SCONFE	B720	-1	1000					Y		Y
SEC./TITULO UNIFORME(ANONIMO)		B730	SUNIF	B730	-1	500					Y		Y
SEC./TITULO DIFERENTE		B740	STDIFE	B740	-1	1000					Y		Y
NUMERO DE TARJETA DE BIB. MEX		B010	REGMEX	B010	-1	50							Y
ISBN		B020	ISBN	B020	-1	100					Y		Y

DATA BASE: BIBLB

FIELD NAME	MNEMONIC TAG	T	H	I	L	T	E	T	R	S	P	D	
												C	L
EDICION	B250 EDICIO	B250	-1	1000									Y
PIE DE IMPRENTA	B260 PIEIMP	B260	-1	1000									Y
DESCRIPCION FISICA	B300 FISICA	B300	-1	1000									Y
SEC. DE SERIES/TITULO	B430 SSTIT	B430	-1	100					Y		Y		Y
NOTAS GENERALES	B500 GENERA	B500	-1	1000					Y		Y		Y
NOTA/"ENCUADERNADO CON"	B510 ENCUAD	B510	-1	100					Y		Y		Y
NOTA/BIBLIOGRAFIA-DISCOGRAFIA	B540 BIBLIO	B540	-1	1000					Y		Y		Y
NOTA/CONTENIDO	B550 CONTEN	B550	-1	1000									Y
NOTA DE REPRODUCCION	B580 REPRD	B580	-1	500					Y		Y		Y
NOTA VERSION ORIGINAL	B590 ORIGIN	B590	-1	1000					Y		Y		Y
SEC. SERIE/PERSONA-TITULO	B800 SSPERS	B800	-1	1000					Y		Y		Y
SEC. SERIE/CORPORATIVO-TITULO	B810 SSCORP	B810	-1	1000					Y		Y		Y
SEC. SERIE/CONFERENCIA-REUNIO	B820 SSCONF	B820	-1	1000					Y		Y		Y
SEC. SERIE/TITULO UNIF.	B830 SSUNIF	B830	-1	1000					Y		Y		Y
LITERAL NUM. ROMANO P/SERIES	B840 ROMSER	B840	-1	6									Y
TITULO ABREVIADO	H200 TABREV	H200	-1	1000									
CAMPO USADO PARA RENUM DE MINISIS	RNUM	W900	-1	10									
NUMERO DE EJEMPLARES	W930 NUNEJ	W930	-1	2		Y	0						
VARIANTE/AUTOR PERSONAL	B900 XPERSON	B900	-1	2000					Y		Y		
VARIANTE /AUTOR PERSONAL	B901 VPERSON	B901	-1	1000									Y
ENC. CORRECTO/AUTOR PERSONAL	B902 CFPERSON	B902	-1	1000									Y
VARIANTE/AUTOR CORPORATIVO	B910 XCORP	B910	-1	2000					Y		Y		
VARIANTE/AUTOR CORPORATIVO	B911 VCORP	B911	-1	1000									Y
ENC. CORRECTO/AUTOR CORPORATI	B912 CCORP	B912	-1	1000									Y
VARIANTE/CONFERENCIA	B920 XCONF	B920	-1	2000					Y		Y		
VARIANTE CONFERENCIA	B921 VCONF	B921	-1	1000									Y
ENC. CORRECTO/CONFERENCIA	B922 CCONF	B922	-1	1000									Y
VARIANTE/TITULO UNIFORME	B930 XUNIF	B930	-1	2000					Y		Y		
VARIANTE/TITULO UNIFORME	B931 VUNIF	B931	-1	1000									Y
ENC. CORRECTO/TITULO UNIFORME	B932 CUNIF	B932	-1	1000									Y

III.3.2 Software

a.-Base de Datos

1).-Tipo

Tenemos que esta base de datos es tipo relacional

2).-Capacidad

-Número máximo de registros que pueden almacenarse

Sólo está limitado a la capacidad en disco del Sistema.

-Tamaño máximo del registro

Hasta 4,000 bytes

-Número máximo de campos por registro

Se puede tener hasta 256 campos por registro

-Máximo número de subcampos por campo

Hasta nueve con MINISIS normal y sin límite con la interfase UNIMARC

-Máxima longitud de una clave

La clave puede alcanzar un máximo de 256 bytes

3).-Características generales

-Posibilidad de tratar un subconjunto de regs. como una B.D.

Por medio de un rango, "HITFILE" y definiendo una restricción inicial

-Facilidad de uniones

La unión de Base de Datos se puede dar por medio del ISN (Internal Sequence Number), de la clave de un KSAM o de cualquier campo común (invertido).

-Permite almacenar conjunto de caracteres en diferentes idiomas / alfabetos

Se puede guardar tablas para manejo de acentos, Ñ, ç, ll

-Permite el ordenamiento de acuerdo a información adicional

Con la interfase UNIMARC a través de indicadores del número de caracteres que se deben ignorar en el ordenamiento, y a través de un archivo KSAM con un campo para anotar la equivalencia entre lo que se imprime y lo que se ordena

-Se permiten campos de longitud variable

Si esta permitido

-Se permiten campos de longitud fija

Sólo con interfase UNIMARC

-Se permiten la existencia de subcampos repetibles

Sólo con la interfase UNIMARC

-Existen campos obligatorios

Si existen

b.-Data Base Manager System (DBMS)

1).-Lenguaje en el que se montó el DBMS

SQL

2).-Mantenimiento

-Bitácora de transacciones para recobrase de caídas del sistema

Si

-Bitácora de transacciones para contabilidad

Si

-Verificación de tipo de dato

numérico, alfanumérico, de acuerdo archivo de validación.

-Rutinas de validación propias del usuario

Si

-Verificación de campos con valores únicos

SI

-Tipo de archivos invertidos

KSAM, BTREE

-Tipos de extracción

Por palabra, término, campo completo, según rutina del usuario

-Eliminación de palabras stop (no significativas)

De varios tipos: sólo al inicio, en cualquier parte del texto, o según indicadores si se usa UNIMARC

-Actualización de los archivos invertidos

Al entrar cada campo a la Base.

4). -Expansión del Manejador

-Lenguaje

Permite que se pueda interfasar con SPL, PASCAL, COBOL

-Generación de aplicaciones completas

Por medio de intrínsecos de MINISIS

-Contribuciones de usuarios

Existe UCL "User Contributed Library", distribuida con cada versión

-Relaciones con otros manejadores de base de datos

IMAGE, y con MICROISIS

-Mejoras

Lenguaje 4GL, Diccionario mejorado, interfases gráficas, fotocomposición, correo electrónico, BD textos

5). -Aplicaciones

-Aplicaciones típicas

Procesos Técnicos; recuperación bibliográfica; control de proyectos; tesauros; E.D. legales

-Aplicaciones futuras

CD/RDM; automatización de oficinas; préstamo; tipografía y fotocomposición; estadísticas

6).-Ordenamiento

-Máximo número de niveles de ordenamiento

Hasta 5

-Ordenamiento de caracteres españoles/ acento, ch, ll

Si

-Remoción de palabras stop

Hasta 9 archivos stop

-Conservación de etiquetas del campo

Si, en HITFILE

-Selección de campos alternativos

Si

-Procesamiento de índices KWIC

Si

-Manejo de sufijos / prefijos

Si

-Especificación de número máximo de claves que se extraerán

Si

7).-Características generales

-Identificación de campos

Por etiqueta y por nemónico

-Número de archivos invertidos

Varios, pueden varios campos compartir un archivo

invertido

-Búsqueda en varias Bases de Datos

No

-Búsqueda libre de texto

Si, incluso puede forzarse en campos invertidos

-Operaciones permitidas en cálculo numérico

Suma, Max, Min, Prom

8).-Portabilidad

Únicamente puede implantarse en la computadora HP3000,
por lo que su portabilidad es deficiente

c.-Paquetes Programas de explotación

1).-Lenguaje

En PASCAL se han realizado los programas de apoyo, como
son los de generación de tarjetas de continuación en la
obtención de juegos de tarjetas (por autor, título, tema,
etc.) de los datos almacenados.

2).-Captura de datos

-Definición de pantalla de captura

No existe

-Facilidad para edición

Difícil, a través de varios comandos y funciones

-Candado a campos que se están modificando

Si

-Manejo de valores default

No

-Operadores usados

se utilizan los operadores AND, OR, NOT, IGN, ADJ

3).-Productos resultantes

-Creación automática de un formato de impresión implícito

Si

-Truncamiento

Por la derecha o por la izquierda

-Manejo de tesauros

Si, pero sólo para consulta, no para creación

-Despliegue de resultado de búsqueda

Deficiente

--Páginas lógicas

Si

--Tabulados / columnas

Si

-Máximo número de registros por página

Si

-Corte a medio registro

Si

-Lineas de encabezado

Hasta diez

-Encabezado que sea parte de los datos

Si, limitado

-Selección de campo alternativo si uno no existe

Si

-Especificación de cual ocurrencia imprimir

No

-Márgenes, sangrías

Si

-Supresión de campos si el registro previo tiene igual contenido

Supresión de entrada idéntica

-Ordenamiento alfabético de las ocurrencias de un campo repetible

SI

4).-Portabilidad

Como MINISIS es un sistema integrado que incluye además del manejador, estos paquetes, esto indica que sólo pueden exportarse a máquinas parecidas, o sea únicamente a computadoras HP3000

d).-Obtención y costos del Software

1).-Proveedor

El proveedor del sistema MINISIS es el International Development Research Center (IDRC)

2).-Costo

Es sin cargo para entidades del gobierno y centros de educación

3).-Equipo

Es necesario contar con una HP3000 y 256K de memoria para montar el sistema MINISIS

4).-Sistema operativo

MPE

III.3.3. Hardware

a.-Tipo

Minicomputadora HP3000

b.-Características generales

Capacidad de memoria principal: 4 Mbytes

Capacidad almacenamiento disco duro: 3.5 Gigabytes

Longitud de palabra: 16 bits

Puertos que soporta: 64

Voltaje de entrada: 220 Volts

c.-Obtención y costos

Se obtiene estos sistemas en Hewelt Packard de México,

III.3.4 Humanware

a.-Capacitación de personal

1).-Responsables

Está el IDRC; y aquí en México ciertos Consultores ofrecen cursos con la instalación, cursos intermedios y cursos para avanzados

2).-Material didáctico

Deficiente en español

3).-Capacitación a instructores

En IDRC mediante convenios

b.-Documentación de los sistemas

1).-Documento introductorio o folleto

Bueno

2).-Calidad de los manuales

Buena, vienen completos

3).-Traducción al español de documentación

En elaboración

c.-Grupos de usuarios

1).-Objetivos

Existen grupos a nivel nacional e internacional y su meta

es principalmente compartir experiencias

2).-Actividades que realizan

Se realizan reuniones, cursos, asesorías, y un congreso internacional

3).-Vínculo con el centro responsable

A través del coordinador del grupo

IV.1. Introducción

Sin lugar a dudas la UNAM es uno de los centros de educación superior e investigación más importantes del país ya que se imparte cátedra a alrededor de 300,000 estudiantes y en ella se desarrolla el 50% de la investigación nacional.

En las mismas proporciones, su sistema bibliotecario, el cual cuenta con 216 bibliotecas ubicadas en Facultades, Institutos de Investigación, Escuelas, Preparatorias, Colegios de Ciencias y Humanidades, Coordinaciones, Centros, en diferentes estados de la República e incluso a bordo del buque de investigación oceanológica "Puma"; es asimismo enorme y cuenta con alrededor de 3,000,000 de volúmenes que se incrementan con aproximadamente 120,000 volúmenes anuales.

La creación de la Dirección General de Bibliotecas fue necesaria para dirigir y coordinar el Sistema Bibliotecario Universitario y lograr de esta manera la standardización y optimización de los recursos técnicos, administrativos y patrimoniales y así establecer lineamientos generales del Sistema Bibliotecario y dirigir los proyectos, programas, procedimientos e investigaciones que se desarrollan y sustentan los servicios que las bibliotecas prestan a los usuarios universitarios.

En general las bibliotecas están organizadas en:

-Servicios Técnicos: "Tienen la responsabilidad de adquirir, organizar y controlar el material bibliográfico; es decir, todos los procesos diseñados para hacer el material más accesible al usuario, incluyen: catalogación, clasificacicación, elaboración de índices y resúmenes y colocación del material en los estantes"(Lbrunam: Sistema Automatizado para Bibliotecas, Charlotte Bronsolier et al., 1979, DGB)

-Servicios al público: "Permiten al usuario localizar el material con la mayor precisión y el menor esfuerzo, gracias a un sistema adecuado de información" (idem)

El problema que enfrentó la Dirección General de Bibliotecas con los libros seleccionados por las bibliotecas y enviados a esta Dirección para que se realizara su registro, pago al proveedor y el proceso técnico del libro, funciones que realiza en forma en forma centralizada la DGB, se debió al impresionante incremento en la adquisición de libros que de 1974 a 1979 aumentó a 400%. Para 1977 el rezago en los procesos técnicos indicaba la suma de 66,000 volúmenes detenidos que no era posible poner en servicio y que causaba que a veces pasara más de un año desde la selección del material hasta su puesta en el estante en la biblioteca.

IV.2 Antecedentes de la Automatización

En 1973 se inició la reorganización del sistema bibliotecario y entre sus proyectos se contempló la creación de un banco de datos que agilizara el proceso técnico del libro y brindara facilidades al usuario para la consulta del material bibliográfico.

Para 1975 se había realizado una tesis profesional que abarcaba el diseño y programación de un sistema automatizado de adquisiciones y proceso técnico de los libros. sin embargo la rigidez y particularidad del sistema creado no permitió la inversión necesaria para su creación.

En 1976 se inició una investigación más profunda en la que participaron la Subdirección de la DGB, los departamentos de Planeación, Adquisición, Procesos Técnicos y el personal del nuevo Departamento de Cómputo. Esta interrelación dió por resultado el diseño y programación del sistema LIBRUNAM, y que para 1979 estuviera ya en funciones, utilizando la computadora Burroughs 6800 del entonces Centro de Servicios de Cómputo, estaba programado en lenguaje ALGOL, y utilizaba funciones HASH en vez de árboles para la búsqueda, recuperación y grabación de datos.

Para 1984, la vida del sistema había llegado a su fin, por lo que el Departamento de Cómputo comenzó la búsqueda de nuevos paquetes manejadores de bases de datos, encontrando finalmente lo que se buscaba en una máquina especializada IDMS00, de la firma Critton Lee, la cual cuenta con firmware dirigido al manejo de bases de datos, la cual mostraba enorme potencia, versatilidad y capacidad dado su precio (mucho menor a la Burroghs que se pensaba adquirir); por lo que para 1985 se inició la reprogramación, ahora en lenguaje ALPHABASIC, de los programas de explotación en una minicomputadora ALPHAMICRO conectada a la IDMS00. La nueva adquisición permitió ampliar los servicios a los usuarios, incluyendo servicio de préstamo y consulta al público en general (la velocidad de respuesta promedio en un día normal de trabajo

en la búsqueda de una ficha de información, ya sea por título, autor o tema, tarda alrededor de 1350 milisegundos) a un nivel razonable, así como la obtención de tarjetas catalográficas en línea, y la conexión de esta base, vía telefónica, a otras dependencias, ampliándose sus horizontes.

A la fecha el equipo con el que se cuenta es la IDMS00 con varios discos que dan una capacidad de 1500 Mbytes, a la cual se conectan dos minicomputadoras ALPHAMICRO, se tienen 3 impresoras de 800 líneas por minuto, 600 líneas por min. y una lenta respectivamente, así como un no-break de varios KVA y un transformador de voltaje y corriente, varias decenas de terminales, algunas de las cuales tienen anexo lápices ópticos, varios modems para comunicaciones, alrededor de 11 microcomputadoras entre PCs y ATs que comparten varias impresoras, entre ellas dos laser-jet, así como sistemas CD-ROM para la utilización de discos ópticos.

IV.3 Descripción de la Base de Datos libros

IV.3.1 Diseño

Debido a que la IDMS00 maneja base relacionales rigurosas, el argot cambia, por lo que en vez de archivo se le llamará relación, en vez de registro se le nombrará tuple y en vez de campo se le denominará atributo (ver el apéndice).

La base de datos, llamada "libros", quedó formada por 6 relaciones o tablas cuyos nombres son: fichas, inventario, tablainv, numerounico, lista3, dispersion.

los tuples de la relación fichas se basan en el formato MARC para monografías y a continuación se describe el formato MARC/DGB para mostrar cómo se guarda la información en la rel. "fichas":

Cada tuple de "fichas", guarda una etiqueta que indica de que trata la información que se almacena en ese tuple.

ETIQUETA CAMPO

#008 En este campo, llamado "campo fijo", se guarda una serie de valores en los que su posición es importante pues se almacenan y recuperan a través de una máscara; los valores son:

valor	pos.	observaciones
fecha	1	S - fecha conocida C - Copyrigh M - fecha múltiple Q - fecha incompleta H - otras fechas
1a fecha	2-5	son 4 dígitos
2a fecha	6-9	son 4 dígitos
origen de catalogación	10-12	NUC (National Union Catalog) DGB (Dir Gral de Bibs) UBI (Unidad Bibs Inv Cientf)
clase de ilustraciones	13-16	A - ilustraciones B - mapas C - láminas O - otros
presentación física	17	A - micropellicula B - microficha D - fotocopia Z - otros
forma de contenido de mat.	18-21	A - resumen analítico B - bibliografías C - catálogos D - diccionarios E - enciclopedias

H - manuales
 I - índices
 O - Biografías
 P - libros de texto program.
 R - directorios
 S - estadísticas
 Y - anuarios

tipos de publicaciones 22

G - origen nacional
 E - origen extranjero

idioma 23-25

ENG - inglés
 FRE - francés
 ITA - italiano
 GER - alemán
 POR - portugués
 OTR - otro
 español son tres blancos

Asiento Principal parte del título 26

1 - Si forma parte
 blanco - no forma parte

código de catalogación 27

blanco - 1a edición de reglas de catalogación angloamericanas
 2 - 2a edición de reglas de catalogación angloamericanas

ETIQUETA CAMPO

\$020 ISBN
 \$050 Clasificación

Asiento principal

\$100 Autor personal
 \$110 Autor corporativo
 \$111 Congreso o reunión
 \$130 Asiento bajo título uniforme

Título

\$240 Título uniforme
 \$245 Título propiamente dicho
 \$250 Edición
 \$260 Pie de imprenta
 \$300 Descripción física del libro.
 \$430 Serie
 \$500 Notas generales
 \$505 Notas de contenido

Encabezamientos de materia

\$600	Autor como materia
\$610	Organismos Corporativos como materia
\$611	Congreso como materia
\$630	Titulo uniforme como materia
\$650	Encabezamientos de materias generales
	Asientos secundarios
\$700	Autores personales
\$710	Autores corporativos
\$711	Congresos
\$730	Titulos uniformes
\$740	Titulos abreviados u otros titulos
	Controles internos
\$900	Clave de la dependencia
\$901	Numero(s) de adquisición
\$902	Claves de Clasificador/Supervisor/Capturista

Para una mayor comprensión vea el diagrama que indica los atributos de cada tabla y las relaciones que existen entre ellas (diag. 1).

En seguida se describe la base:

relación fichas

En esta relación se guarda la información bibliográfica, teniendo cada tupla la etiqueta correspondiente al formato MARC/DGB y el texto de información referente a ella.

atributos:

matriz: Contiene un número que se da secuencialmente a cada nueva ficha que entra al banco, y el cual es la llave (está indexado) para encontrar rápidamente esa ficha. Debido a que una ficha puede tener un número dinámico de tuples, según las etiquetas presentes, todos ellos llevan el mismo número de matriz y de esta manera forman un

conjunto definido que puede ser recuperado y que ampara la información de toda la ficha .

tipo=i4(entero de 4 bytes), indexado (cluster-index).

etiqueta: Aquí se guarda la etiqueta correspondiente al formato MARC/DSB y que indica que clase de información se guarda en este tuple. etiqueta: tipo= string de 3 char, no indexado.

tarjeta: Debido a que la longitud de cada atributo no puede ser mayor a 256 bytes, se tendría la limitante de que la información estuviera constreñida a este tamaño, como esto no puede ser posible so pena de mutilar la información bibliográfica, el atributo tarjeta guarda un número secuencial que liga la continuación de la información en otro tuple que tiene las mismas características del anterior y que solo se diferencia en este atributo. Esto permite simular un atributo dinámico de longitud limitada al número máximo que se pueda guardar en el atributo tarjeta multiplicado por 256.

tipo= i2 (entero de 2 bytes), no indexado.

idc: En este atributo cuyo nombre significa identificador de campo, el primer caracter sirve de control para manejar peculiaridades de ciertas etiquetas, y el segundo indica cuantos caracteres se deben saltar del texto para ordenar la información.

tipo= string de 2 char, no indexado.

texto: Aquí se guarda la información de la ficha bibliográfica relacionado al tipo de etiqueta correspondiente en ese tuple.

tipo= string de 256 char, no indexado.

relación dispersión

Esta relación es la primera con la que se entra en contacto desde el mundo exterior. El usuario mete la cadena de búsqueda, la cual es separada en tokens por un programa que realiza la función de scanner, y las palabras son pasadas al manejador de bases de datos el cual por medio de una búsqueda que utiliza B-trees (en vez de HASH como era en el antiguo sistema), encuentra las palabras en la relación.

atributos:

llave: Aquí se guarda el conjunto de palabras existentes y reconocidas por el sistema para el acceso de información (una palabra por tuple).

tipo = string de 30 char, indexado (cluster-index).

numuni: Aquí se guarda un número único asociado exclusivamente a esa palabra, y que nos servirá, una vez que se multiplique por cien, como llave para entrar a la relación lista3.

tipo = i4, no indexado

existen: Indica cuantas fichas bibliográficas están

reunidas bajo esa palabra. En el caso de que el atributo llave contenga la clasificación de la ficha en vez de palabras de búsqueda (ej. HC23.4/Q2), entonces en este atributo se guarda directamente el número de matriz con la cual se recupera esa ficha, ya que la clasificación es única para cada ficha bibliográfica.

tipo= 14, no indexado

relación lista3

En esta relación se tienen las múltiples correspondencias a las fichas, dadas las entradas hechas por las palabras de búsqueda del usuario, estas correspondencias las tiene agrupadas por palabra y a éstas por autor, título y tema; obteniéndose los números de matriz con los cuales se obtienen las fichas y las bibliotecas donde se localiza el material.

atributos:

numfam: En este atributo se guarda un número formado por el valor del atributo numuni de la relación de dispersión, multiplicado por 100, numfam = numuni * 100. La razón de multiplicar por cien se debe a que los valores de numuni son números secuenciales, y la llave de acceso a esta relación está formada por numfam + tipo, si no se multiplicara por cien, al sumar el tipo se tendría

un número que colisionaría con otro numuni, y esta multiplicación permite un rango de 0 a 99 en el cual al sumar el tipo cae dentro de este dominio sin afectar el espacio del siguiente numuni. Tenemos entonces que el número de numfam abarca el ámbito de todas las fichas donde se localiza esa palabra bajo autor, título o tema.

tipo= 14, indexado(junto con tipo, o sea llave= numfam + tipo) cluster-index.

tipo: Es un número que indica si la palabra buscada pertenece a autor, título o tema. Forma parte del índice de esta relación (llave= numfam + tipo) y permite realizar la búsqueda en el conjunto de autores, títulos o temas según se especifique, si el usuario desea búsqueda libre, entonces la llave se arma con tipo=cero. El tipo indica incluso si se trata, en el caso de autor, de autor corporativo, etc..., igual en caso de título y tema para búsquedas más precisas.

tipo= 12, indexado (con el atributo numfam), cluster-index.

matriz:

con este atributo es posible buscar en la relación fichas la información catalográfica, y en la relación inventario la biblioteca donde se localiza el material.

tipo= 14, no indexado.

relación inventario

En esta relación se maneja el patrimonio bibliográfico, se indica el status no sólo de un libro (amparado por una ficha), sino de cada uno de los volúmenes adquiridos del mismo libro.

atributos:

dep: Aquí se guarda un número que indica que biblioteca es la que tiene físicamente el volumen en su poder. No se tiene indexada esta relación debido a que cuenta sólo con 216 tuples (uno para cada biblioteca).

tipo= i2, indexado junto con numadq; llave= dep+numadq, (cluster-index).

numadq: En este atributo se tiene el número de adquisición del volumen, este número es único dentro del conjunto de tuples que tienen la misma dependencia.

tipo= i4, indexado junto con dep; llave= dep + numadq, (cluster-index).

numat: Se tiene el número de matriz del volumen y relaciona a inventario con fichas.

tipo= i4, indexado (non cluster-index)

fecha: Se guarda la fecha de entrada (mes y año) de la adquisición.

tipo= string de 4 char, no indexado.

tipo: Indica si la entrada fue por alta (nuevo libro que

no existía antes en el banco) o por cargo
(búsqueda en el banco de datos)

tipo= i2, no indexado.

colec: atributo que se utiliza para identificar grupos de
volumenes que se manejan de manera especial.

tipo= string 1 char, no indexado.

relación tablainv:

En esta relación se guarda información relativa
a las bibliotecas departamentales.

atributos:

dep: Con este número es posible relacionar a inventario
con tablainv.

tipo= i2, no indexado

clave: Es un conjunto de caracteres que identifican a
cada biblioteca (ej. Facultad de Ing. es FG), esta
clave aparece en las tarjetas catalográficas.

tipo= string de 6 char, no indexado

nombre: Viene el nombre completo de cada dependencia

tipo= string de 150 char, no indexado

numadq: Se tiene el último número de adquisición de esa
biblioteca, y se utiliza para las nuevas
adquisiciones.

tipo= i4, no indexado

relación numerounico

esta relación consta únicamente de tres tuples,
guarda información relacionada con los números

últimos utilizados por el sistema. Cuando ingresan nuevas fichas al banco, estos números son leídos para asignarlos a los nuevos tuples, incrementándose sus valores en esta relación.

atributos:

llave: En el primer tuple se guarda la palabra matriz.
En el segundo tuple se guarda la palabra libros.
En el tercer tuple se guarda la palabra titulo.
Estas palabras sirven de acceso para leer la inf.
tipo= string de 6 char, no indexado

numuni: En el primer tuple se tiene el número último de matrices en el banco, que para la fecha de diciembre de 1989 era de 468,370 esto implica que se han usado 468,369 números de matrices.

En el segundo tuple se tiene el valor de 887,207 lo que indica que en el banco se tienen registradas (en dispersión) 887,206 palabras diferentes en dic. de 89.

En el tercer tuple se guarda la cantidad de 371,449 que nos muestra que el número total de fichas catalográficas existentes en el banco alcanza la cifra de 371,448 para el mes de diciembre de 1989.

tipo= i4, no indexado.

IV.3.2 Software

a.-Base de Datos

1).-Tipo

Tenemos que esta base de datos es tipo relacional

2).-Capacidad

-Número máximo de registros que pueden almacenarse

Sólo está limitado a la capacidad en disco del Sistema.

-Tamaño máximo del registro

Hasta 2,000 bytes

-Número máximo de campos por registro

Se puede tener hasta 180 campos por registro

-Máximo número de subcampos por campo

No existe el concepto de subcampos en esta base, pero el DBMS tiene operadores que permiten el manejo y localización de substrings utilizando para ello comodines (wild cards) refinados.

-Máxima longitud de una clave

La clave puede alcanzar un máximo de 255 bytes

3).-Características generales

-Posibilidad de tratar un subconjunto de regs. como una B.D.

Se puede generar nuevas relaciones físicas creando un comando que incluya instrucciones de "retrieve into", o pueden ser virtuales con comandos que incluyan "create view".

-Facilidad de uniones

El concepto de Base de datos que maneja la IDM500

considera que aquella representa el universo completo donde existen todos los conjuntos (tablas) necesarios y encierra el espacio del producto cartesiano de todas las relaciones (ligas) posibles entre sus dominios (atributos) por lo que las bases de datos son tomadas como entidades absolutas independientes unas de otras. Esto implica que no existe comandos ni medios en esta computadora para unir distintas bases de datos. Para realizar algo parecido se deben de "bajar" estas bases a la computadora host y manejarlas con programas contruidos por los usuarios para tal fin.

-Permite almacenar conjunto de caracteres en diferentes idiomas / alfabetos

No

-Permite el ordenamiento de acuerdo a información adicional
Si es posible pero complicado, y el usuario debe de crear los comandos de manera que utilicen relaciones (tablas) de equivalencias.

-Se permiten campos de longitud variable

No está permitido, pero usando la técnica descrita en la relación fichas es posible simularlo bastante bien.

-Se permiten campos de longitud fija

No sólo se permiten, son obligados.

-Se permiten la existencia de subcampos repetibles

No existe ni el concepto de subcampo.

-Existen campos obligatorios

No existen

b.-Data Base Manager System (DBMS)

1).-Lenguaje en el que se montó el DBMS

Lenguaje C

2).-Mantenimiento

-Bitácora de transacciones para recobrase de caídas del sistema

Se tiene un Transaction Management, que asegura que las transacciones se lleven a cabo completamente o no actualiza nada (back out), para el caso de una falla del sistema. Además se tiene también un sistema de auditoría "Audit Logging" con el que la computadora mantiene checkpoints de transacciones de logs para auditoría, backup y recobro de databases. El sistema es eficiente pues el tesista ha comprobado que incluso en casos graves (excepto daño en hardware), La máquina, aunque ha tardado días, recupera la base, sin ser necesario bajar la información desde cinta.

-Bitácora de transacciones para contabilidad

Se cuenta con un Integrated Data Dictionary propio de la máquina que mantiene automáticamente la información sobre las relaciones de los datos, sus tipos, accesos privilegiados, y otros indicadores guardados en este diccionario. Así como también cuenta con el Audit logging

-Verificación de tipo de dato

numérico, alfanumérico, y por medio de comandos creados por el usuario para validaciones más sofisticadas.

-Rutinas de validación propias del usuario

Si, creando comandos (los cuales están sustentados en cálculo relacional) que pueden ser lanzados desde programas en la computadora host.

-Verificación de campos con valores únicos

Si

-Tipo de archivos invertidos

BTREE clustered y no-clustered

-Tipos de extracción

Por palabra, término, campo completo, según rutina del usuario

-Eliminación de palabras stop (no significativas)

Construyendo comandos que utilicen la función "substring" para la eliminación de palabras dentro de la información a guardar.

-Actualización de los archivos invertidos

Al momento de dar "append" (función que se utiliza para dar de alta un tuple en la Base) o "replace" (función que permite modificar un tuple de la Base) o "delete" (función que elimina un tuple de la Base).

4).-Expansión del Manejador

-Lenguaje

El Sistema está hecho de tal forma que permite interactuar con cualquier lenguaje nativo de la computadora host de acuerdo al protocolo de comunicaciones entre las dos computadoras. Actualmente existen interfases en software que son transparentes al usuario respecto al protocolo, estas interfase están

hechas para diversas computadoras hosto tales como Vax, AlphaMicro, etc..., que soportan lenguajes como 'C', PASCAL, BASIC, FORTRAN, etc...

-Geenración de aplicaciones completas

Mediante "Store Commands"

-Contribuciones de usuarios

Hasta el momento no se tiene noticia en México de grupos de usuarios debido a que la DGB fue la primera institución que manejó este tipo de Sistema en la república, sin embargo a últimas fecha la Universidad de Guadalajara y la Comisión Federal de Electricidad han adquirido este tipo de equipos. Lo que a futuro permitiría la formación de tales grupos.

-Relaciones con otros manejadores de base de datos

No se tiene noticia.

-Mejoras

Un help en línea para la mejor utilización del lenguaje del manejador que indica con mayor precisión errores al teclear funciones erróneas y su correcta aplicación.

5).-Aplicaciones

-Aplicaciones típicas

Sistemas Bancarios, Sistemas basados en gráficos, Sistemas Expertos, Bases de datos Sísmicas, Automatización Industrial, control de documentación.

-Aplicaciones futuras

Inteligencia, Soporte en Sistemas Militares, Sistemas Policiacos.

6).-Ordenamiento

-Máximo número de niveles de ordenamiento

Se permiten varios niveles

-Ordenamiento de caracteres españoles/ acento, ch, ll

No implícitamente, el usuario debe crear comandos que manejen esas situaciones.

-Remoción de palabras stop

Utilizando comandos que realicen esa función.

-Conservación de etiquetas del campo

Si, si el usuario deja un atributo para ello.

-Selección de campos alternativos

No es fácil de manejar esto.

-Procesamiento de índices KWIC

No

-Manejo de sufijos / prefijos

El usuario debe de construir las rutinas manejadoras

-Especificación de número máximo de claves que se extraerán

El usuario lo define en sus rutinas.

7).-Características generales

-Identificación de campos

Por su nombre.

-Número de archivos invertidos

Hasta 256 índices es posible tener para una sola relación, con un máximo de 15 índices sobre un mismo atributo.

-Búsqueda en varias Bases de Datos

No

-Búsqueda libre de texto

Si

-Operaciones permitidas en cálculo numérico

sum, max, min, avg, count, *, /

B).-Portabilidad

Unicamente existe este manejador en la máquina IDMS00, por lo que su portabilidad es pésima.

c.-Paquetes Programas de explotación

1).-Lenguaje

El paquete de explotación de las Bases (llamado FREEFORM) que está montado en 'C', es parecido a DBASE III+, pero más poderoso en sus habilidades entre las que cuenta una facilísima comunicación con IDMS00, y también versatilidad, permite programar en un pseudoPASCAL.

Los programas de explotación que el usuario produce, son realizados en el lenguaje nativo de la computadora host que esté comunicada a la IDMS00 y son más rígidos en sus prestaciones pero más orientados a lo que el usuario desea.

2).-Captura de datos

-Definición de pantalla de captura

Si, con FREEFORM

-Facilidad para edición

Fácil, con FREEFORM

-Candado a campos que se están modificando

Si (FREEFORM), y progs. del usuario que usan las funciones de "begin transaction" y "end transaction"

-Manejo de valores default

Se puede en FREEFORM y con programación del usuario

-Operadores usados

se utilizan los operadores AND, OR, NOT

3).-Productos resultantes

-Creación automática de un formato de impresión implícito

SI (FREEFORM), y con los progs. del usuario "ATASCA" y "TECNIC"

-Truncamiento

Por la derecha o por la izquierda (Tanto con FREEFORM como con ciertos progs. del usuario)

-Manejo de tesauros

No

-Despliegue de resultado de búsqueda

Bueno (con FREEFORM), excelente (progs. del usuario)

-Páginas lógicas

SI (FREEFORM)

-Tabulados / columnas

SI (FREEFORM) y progs. usuario

-Máximo número de registros por página

SI (FREEFORM)

-Corte a medio registro

SI (FREEFORM)

-Lineas de encabezado

Permite un manejo versatil de encabezados pues pueden programarse para que aparezcan los que se desean según las circunstancias (FREEFORM)

-Encabezado que sea parte de los datos

SI, buen manejo y versatil (FREEFORM)

-Selección de campo alternativo si uno no existe

Si (FREEFORM)

-Especificación de cual ocurrencia imprimir

Si (FREEFORM)

-Márgenes, sangrías

Buen manejo

-Supresión de campos si el registro previo tiene igual contenido

Si, si así se desea (FREEFORM)

-Ordenamiento alfabético de las ocurrencias de un campo repetible

Si

4).-Portabilidad

Para FREEFORM es excelente pues igual puede montarse en VAX y AlphaMicro entre otras máquinas minis que en PC's que tengan sistema operativo MS-DOS, PC-DOS y Z-DOS.

Para los programas creados por el usuario la portabilidad está restringida a las máquinas AlphaMicro

d).-Obtención y costos del Software

1).-Proveedor

El proveedor del DBMS es Britton Lee Inc, el cual no comercializa por separado el manejador ya que parte del mismo esta en firmware dentro de la máquina IDMS00.

El proveedor de FREEFORM es Dimension Software Systems, Inc.

El proveedor del lenguaje AlphaBasic y el Sistema Operativo AMOS/L es Industrias Digitales S.A.

En México quien distribuye los anteriores productos es la
compañia Industrias Digitales S.A.

2).-Costo

El lenguaje Alpha Basic y el Sistema Operativo vienen ya
incluidos en la compra de la máquina Alpha Micro, el
lenguaje IDL viene también incluido con la compra de la
IDM.

3).-Equipo

El DBMS necesita forzoso una IDM500.

Los progs. del usuario necesitan una AlphaMicro

Para FREEFORM es necesario contar con una máquina que
maneje UNIX o AMOS o MS-DOS o PC-DOS o Z-DOS, entre otras
y que tenga un mínimo de 300K de memoria, siendo
preferible 640k

4).-Sistema operativo

El DBMS forma parte del Sistema Operativo el cual a su
vez forma parte inseparable del hardware de la IDM500.

MS-DOS, PC-DOS, Z-DOS, UNIX, AMOS, etc... (para FREEFORM).

Para los progs. del usuarios el Sist. op. es AMOS

IV.3.3. Hardware

a.-Tipo

El Hardware necesario y único es una IDM500 para el DBMS.

b.-Características generales

IDM500:

Memoria Principal: hasta de 6 Mb
Memoria Disco: hasta 8000 Mb
Unidad de Cinta: soporta hasta 8

Peso de la Máquina: 77.1 Kg
Altura de la máquina: 44.5 cm
Ancho de la máquina: 48.3 cm
Profundidad de la máquina: 66.0 cm
Consumo eléctrico: 10 Amp max. a 208 Volts +20% - 10%, 47 a 63 HZ AC.
Temperatura operación: 10 a 35 grados centígrados
Humedad relativa: 20% a 80%
16 slots para tarjeta.
Comunicación con otras computadoras: hasta 64 conectadas al mismo tiempo a la IDMS500, entre las cuales es posible conectar a: DEC VAX/VMS, IBM 370/VMS-CMS, IBM 43XX/VM-CMS, IBM 308X/VM-CMS, IBM 303X/VM-CMS, DEC VAX/UNIX, DEC PDP/R SX y RSTS, ATT 3/5 y 20, PRIME 2250/PRIMOS, HONEYWELL, DATAPOINT, HARRIS, GOULD-SEL, ALPHA MICRO, UNIVAC, CONTROL DATA, IBM PC, DEC RAINBOW, WICAT, COMPAQ, SPERRY PC, APOLLO, NETWORKS, ETHERNET, 3BNET, ARCNET.

c.-Obtención y costos

Se obtiene en México en la compañía Industrias Digitales.

Su costo es alrededor de 100,000 dólares.

IV.3.4 Humanware

a.-Capacitación de personal

1).-Responsables

En México se tiene a la compañía Industrias Digitales como la responsable de la capacitación de personal, para cursos más avanzados es posible contratar los servicios de la empresa Britton Lee, para el envío de un instructor.

2).-Material didáctico

No existe en español

3).-Capacitación a instructores

En Industrias Digitales y Britton Lee

b.-Documentación de los sistemas

1).-Documento introductorio o folleto

Bueno

2).--Calidad de los manuales

Buena, vienen completos

3).--Traducción al español de documentación

Nula

c.-Grupos de usuarios

No existen.

V.1 Introducción

La bibliotecología, como ciencia, también se nutre de la investigación básica y aplicada, y siendo los momentos actuales de rápidos progresos que urgen la búsqueda de nuevas soluciones a viejos problemas, los cuales sean atacados con la aplicación de métodos rigurosamente científicos que enriquezcan el conocimiento bibliotecológico, através de la experimentación, la encuesta, la investigación histórica, la investigación de operaciones y la aplicación de métodos cuantitativos.

El campo de estudio es muy amplio dado que involucra investigación en aspectos teóricos y problemas prácticos, propios de las bibliotecas y de los centros de información, como son adquisición, catalogación, almacenamiento y disseminación de la información; asimismo del cómo, del porqué y para quién se genera la información, su comportamiento y medios de comunicación.

Ante esta situación la Universidad Nacional Autónoma de México consideró que era necesario que la bibliotecología contara con un espacio y personalidad propios dentro de un programa de investigación. Así que el 14 de diciembre de 1981 se creó el Centro Universitario de Investigaciones Bibliotecológicas, el cual realiza las siguientes funciones:

- Realizar investigaciones teóricas y metodológicas relacionadas con todos los aspectos de las ciencias bibliotecológicas prioritariamente.

-diseñar modelos alternativos de organización bibliotecaria, de catalogación, clasificación y de automatización de la información, así como la diseminación de los conocimientos pertinentes.

-Atender programas de investigación que contemplen la formación del personal académico de alto nivel.

Prestar al rector de la UNAM, al Secretario General, al Director General de Bibliotecas y a las dependencias universitarias que así lo requieran, el asesoramiento necesario en materia de bibliotecología.

V.2 Antecedentes de la Automatización

El BD INFOBILA (Información y Bibliotecología Latinoamericana) se originó dentro del proyecto de formación de investigadores en bibliotecología para América Latina, que se desarrolló a partir del convenio que el CUIB tiene firmado con la Organización de los Estados Americanos (OEA) desde 1984.

INFOBILA es una BD que incluye materiales latinoamericanos: monografías, artículos de publicaciones periódicas, tesis, capítulos de monografías, ponencias y documentos diversos en las áreas de bibliotecología e información. de esta manera se ofrece información bibliotecológica sobre Latinoamérica producida por latinoamericanos que es fácilmente accesible.

Para manejar la información se decidió adquirir un paquete comercial ahorrando así tiempo y dinero en la implantación del proyecto tras evaluar varios paquetes (Micro-Questel y Micro-BRS) se optó por LOGICAT, paquete de software desarrollado en México y

que ofrece la posibilidad de registrar diversos materiales conforme al formato MARC y a que al estar el paquete orientado a bibliotecas ofrece facilidades a las mismas como la impresión de resultados en forma de ficha catalográfica y/o analítica y a que los datos se pueden manejar externamente (LOGICAT crea archivos en formato DBASE IIIp).

V.3 Descripción de la B.D.

V.3.1 Diseño

En este tipo de sistema, no es posible diseñar la base de datos, pues esta, debido a sus características, no es posible modificarla y siempre se tiene la misma filosofía de diseño en cualquier sistema LOGICAT. Únicamente se instala y se corre. El usuario puede aumentar los campos presentados para satisfacer sus necesidades, dentro de un rango limitado, pero no puede modificar la construcción de la base de datos. LOGICAT está montado en DBase III plus, y maneja tres archivos principales con el mismo nombre raíz y un sufijo (con extensión .DBF) donde en el primero se guarda los campos que definen (según formato MARC) la ficha catalográfica, en el segundo (con sufijo VAR "variables") se guardan dos campos y en el tercero (sufijo NOT "notas") campos que el usuario puede crear. Estos archivos están cada uno con su archivo de índices (extensión .NDX) para aumentar la eficiencia del sistema.

A continuación se exponen las estructuras de cada archivo .DBF;

ARCHIVO:

Field Name	Type	Width	Dec
1 NUM_FICHA	CHARACTER	6	
2 TIPO_MAT	CHARACTER	2	
3 TIPO_REP	CHARACTER	1	
4 COD_FECHA	CHARACTER	1	
5 FECHA_1	CHARACTER	4	
6 FECHA_2	CHARACTER	4	
7 ILUSTRADO	CHARACTER	3	
8 IDIOMA	CHARACTER	3	
9 EDICION	CHARACTER	3	
10 ISBN_ISSN	CHARACTER	14	
11 NUM_CLASF	CHARACTER	47	
12 E10	CHARACTER	3	
13 AUTOR	CHARACTER	60	
14 ARTICULO	CHARACTER	4	
15 TITULO	CHARACTER	79	
16 PAIS	CHARACTER	25	
17 EDITORIAL	CHARACTER	35	
18 E61	CHARACTER	3	
19 TEMA_1	CHARACTER	60	
20 E62	CHARACTER	3	
21 TEMA_2	CHARACTER	60	
22 E63	CHARACTER	3	
23 TEMA_3	CHARACTER	60	
24 PAGINAE	CHARACTER	25	
25 TIPO_COL.	CHARACTER	4	
26 MAS_FIC	CHARACTER	1	
27 SUCUR	CHARACTER	1	
28 INX	CHARACTER	6	
29 VECESP	NUMERIC	5	0

ARCHIVO:

Field Name	Type	Width	Dec
1 FICHA	CHARACTER	9	
2 CAMPO	CHARACTER	60	

V.3.2 Software

a. -Base de Datos

1). -Tipo

Tenemos que esta base de datos es tipo relacional

2). -Capacidad

-Número máximo de registros que pueden almacenarse

Se permiten 1,000,000,000 de registros.

-Tamaño máximo del registro

hasta 4,000 bytes

-Número máximo de campos por registro

Hasta 128 por registro

-Máximo número de subcampos por campo

En DBASE III no existe ese concepto

-Máxima longitud de una clave

La clave puede alcanzar un máximo de 100 bytes

3). -Características generales

-Posibilidad de tratar un subconjunto de regs. como una S.D.

Mediante funciones tales como "COPY TO"

-Facilidad de uniones

Mediante funciones tales como "SET RELATION" o "JOIN"

-Permite almacenar conjunto de caracteres en diferentes idiomas / alfabetos

No

-Permite el ordenamiento de acuerdo a información adicional

No

-Se permiten campos de longitud variable

No

--Se permiten campos de longitud fija

Si

--Se permiten la existencia de subcampos repetibles

No

--Existen campos obligatorios

No existen

b.-Data Base Manager System (DBMS)

1).-Lenguaje en el que se montó el DBMS

"C"

2).-Mantenimiento

--Bitàcora de transacciones para recobrase de caidas del sistema

No

--Bitàcora de transacciones para contabilidad

Si, limitado, (función HISTORY)

--Verificación de tipo de dato

Si

--Rutinas de validación propias del usuario

Si, mediante programación propia del usuario

--Tipo de archivos invertidos

Arboles Binarios

--Tipos de extracción

Por palabra, término, campo completo, según rutina del usuario

--Eliminación de palabras stop (no significativas)

Si, utilizando rutinas del usuario

-Actualización de los archivos invertidos

Si, en línea con "SET INDEX" o en batch con "REINDEX"

4).-Expansión del Manejador

-Lenguaje

Utilizando el lenguaje propio de DBASE III

-Generación de aplicaciones completas

Utilizando el lenguaje propio de DBASE III y comandos intrínsecos del mismo

-Contribuciones de usuarios

Son recibidas por ASHTON - TATE

-Relaciones con otros manejadores de base de datos

MICROISIS

-Mejoras

Hacerlo multiusuario, utilizando lenguaje en SQL

5).-Aplicaciones

-Aplicaciones típicas

Archivonomía, inventarios, contabilidad, administración de empresas, etc...

-Aplicaciones futuras

Redes de cómputo, empresas intermedias

6).-Ordenamiento

-Máximo número de niveles de ordenamiento

Sin límite

-Ordenamiento de caracteres españoles/ acento, ch, ll

No

-Remoción de palabras stop

Utilizando programación del usuario

-Conservación de etiquetas del campo

No utiliza etiquetas

-Selección de campos alternativos

Depende de la programación del usuario

-Procesamiento de Indices KWIC

No, el usuario puede programarlo

-Manejo de sufijos / prefijos

No, el usuario define en sus rutinas de programación

-Especificación de número máximo de claves que se extraerán

El usuario lo define

7). Características generales

-Identificación de campos

Por su nombre

-Número de archivos invertidos

Sin límite, pero simultáneamente se permiten hasta 7

-Búsqueda en varias Bases de Datos

Si, con "SET RELATION"

-Búsqueda libre de texto

Si

-Operaciones permitidas en cálculo numérico

Si, suma, resta, multiplicación, división y potenciación

8). Portabilidad

La portabilidad es excelente, pero es a nivel de microcomputadoras tipo PC XT y AT o clones

c. Paquetes Programas de explotación

1). -Lenguaje

Se pueden crear programas con el lenguaje propio de DBASE

o se pueden correr programas con extensión EXE o COM

2).-Captura de datos

-Definición de pantalla de captura

Si existe

-Facilidad para edición

fácil de editar

-Candado a campos que se están modificando

No es necesario en versión monousuario

-Manejo de valores default

Si, mediante "SET CARRY ON"

-Operadores usados

se utilizan los operadores AND, OR, NOT

3).-Productos resultantes

-Creación automática de un formato de impresión implícito

Si

-Truncamiento

Si

-Manejo de tesauros

No

-Despliegue de resultado de búsqueda

versatil

-Lineas de encabezado

Hasta 4

-Encabezado que sea parte de los datos

Si, versatil

-Selección de campo alternativo si uno no existe

No

-Especificación de cual ocurrencia imprimir

Si

-Márgenes, sangrias

Bueno

-Ordenamiento alfabético de las ocurrencias de un campo repetible

Si

4).-Portabilidad

Los programas de explotación pueden exportarse a sistemas PC XT y AT

d).-Obtención y costos del Software

1).-Proveedor

Aston-Tate, o con cualquier distribuidor de software

2).-Costo

alrededor de 495 dólares

3).-Equipo

Una PC XT o AT con mínimo de 256K

4).-Sistema operativo

MS-DOS

V.3.3. Hardware

a.-Tipo

Microcomputadora

b.-Características generales

Computadora:

-Procesador 8088, 8086, 80286, 80386

-Velocidades desde 4 MHz hasta 33 MHz (aceptable)

- de 8 a 12 MHertz)
- Memoria principal mínima de 256 Kb, preferente 640 Kb
- Dos drives para diskette (mínimamente de 5 1/4", de 360 Kb o mayor capacidad: 1.2 Mb). O en su defecto, y mejor aun, un disco duro (mínimo 20 Mb) y drive para floppy disk
- Puerto paralelo tipo CENTRONICS, para impresora, y puerto serial tipo RS-232C para comunicaciones.
- Monitor a color o monocromático con tarjeta CGA o VGA de preferencia en vez de tecnología TTL
- Teclado básico de 80 teclas, de preferencia de 101 teclas y 12 funciones.
- Fuente de poder mínima de 150 Watts (depende del Hardware).
- Gabinete para un mínimo de 3 unidades de media altura (para el caso de necesitar expansión de memoria masiva).
- Al menos tres ranuras (slots) de expansión para montar tarjetas adicionales.
- De preferencia que el sistema cuente con llave de seguridad bancaria para protección y control del uso de la máquina.
- Coprocesador matemático no necesario, pero aumenta la velocidad de proceso.
- De preferencia con reloj calendario de tiempo real, muy útil en el control de procesos en bibliotecas.
- Sistema con botón de cambio de velocidad a turbo, e

indicador lumínico de cambio.

-Sistema con indicador de acceso a disco duro.

Impresora:

-Impresión por medio de matriz de punto (lo que permite que se puedan imprimir gráficas).

Las impresoras Laser no son de importancia en la biblioteca debido a que la calidad de impresión de una Laser es útil fundamentalmente en los casos de utilizar paquetes de edición especializados en manejo de textos o en funciones administrativas que implican papelería de calidad y no de cantidad. En la biblioteca saldría caro utilizar este tipo de equipo, la tinta cuesta alrededor de 80 dólares, y su mantenimiento, aunado al enorme costo de una impresora de este tipo las hace onerosas. Las impresoras de matriz son más que suficientes.

-Cabeza de impresión de al menos 7 puntos en la matriz para una calidad aceptable de impresión.

-Capacidad de imprimir letra de calidad (modo NLD "Nice Letter Quality").

-Velocidad de impresión de mínimo 190 caracteres por segundo en modo Draft.

-Velocidad de impresión de mínimo 40 caracteres por segundo en modo NLD.

-Memoria de Buffer mínima de 1 Kb (1.6 páginas carta escritas a máquina en promedio), mejor de 8 Kb (3.2 páginas) lo que permite que la computadora se libere

con anticipación aceptable del proceso de impresión.

-Carro de mínimo 10'' (hojas de 80 columnas), aunque es conveniente un carro de 15'' lo que permite hojas de 132 columnas.

-Permita se tengan al menos dos tipos de letra distinta (por ej. Elite y Pica) y también exista modo condensado para imprimir línea de 132 caracteres en hoja de 80 columnas.

-La interfase debe ser paralela compatible con CENTRONICS, como opcional se tendría serial (RS-232C)

-Se pueda emular el modo de la impresora IBM Proprinter principalmente, y/o la de la EPSON, las cuales son standard para muchos paquetes, no sólo de bases de datos, y así ser compatibles con el software de la computadora y poder ser aprovechadas al máximo.

a.-Obtención y costos

Hoy en día las microcomputadoras se pueden obtener con diferentes distribuidores y de muy varias marcas. Por lo que su adquisición no es ningún problema.

Los costos varían extraordinariamente de acuerdo al fabricante en cuestión, siendo las máquinas de IBM, Hewlett Packard, Olivetti y Unisys las más caras del mercado, aunque sus prestaciones son confiables y el servicio que ofrecen muy bueno.

Una máquina se puede conseguir desde 500 y pico de dólares hasta varios miles. A continuación damos una pequeña tabla

con diversos precios para las computadoras e impresoras:

Computadora XT procesador 8088 o NEC V20; 4-10 o 4-12 MHz; un drive floppy disk y hard disk de 30 Mb; monitor TTL o CGA; 640 Kb RAM. Costo en dólares.

IBM: Sistema Personal/2 Modelo 25 (drive diskette de 3.5" de 720Kb, disco duro de 20Mb y monitor con MCGA). Costo de \$2,029.

Hewlett Packard: Sistema VECTRA 286/12-20 (procesador 286, 1Mb de RAM, drive para floppy de 5.25" de 1.2Mb con disco duro de 20Mb, y monitor con tarjeta VGA). Costo de \$2,488.

OLIVETTI: Computadora OLIVETTI PC386 (con monitor y tarjeta VGA, disco duro de 20Mb, floppy de 3.5" de 720Kb). Costo de \$1,831.

ACER: Computadora ACER 500+II (con tarjeta de video triple para MDA, CGA, Hercules). Costo de \$1,390.

TELEVIDEO: Computadora TELEVIDEO XT (monitor TTL). Costo de \$1,330.

SAMA: Computadora SAMA 88 (monitor TTL). Costo \$1,150.

ONYX: Computadora ONYX XT-10 (reloj tiempo real, monitor TTL). Costo de \$1,241.

W.3.4 Humanware

a.-Capacitación de personal

1).-Responsables

Distribuidores y centros educativos

2).-Material didáctico

Existe gran cantidad de material y en español, para todos los niveles de aprendizaje

3).-Capacitación a instructores

Con los distribuidores y en centros educativos

b.-Documentación de los sistemas

1).-Documento introductorio o folleto

Existen en gran cantidad

2).-Calidad de los manuales

Variable, dada la gran cantidad de información impresa sobre el tema

3).-Traducción al español de documentación

Existe en gran cantidad

c.-Grupos de usuarios

1).-Objetivos

Más que grupos de usuarios, son instituciones de enseñanza las que enseñan el manejo, básico y avanzado, del DBASEIII, debido al gran interés que despierta en los usuarios este manejador.

2).-Actividades que realizan

Las instituciones educativas abocadas a la rama de computación, son quienes tienen la mayor actividad en la enseñanza y difusión de este popular paquete.

3).-Vínculo con el centro responsable

A través de los distribuidores

VI INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y SISTEMAS

VI.1 Introducción

Atendiendo a la necesidad de crear lugares apropiados para la investigación en matemáticas en conjunción con las incipientes ciencias computacionales, fue creado el Instituto de Investigaciones en Matemáticas y Sistemas en el año de 1958, siendo el primer lugar no sólo de la UNAM sino del país e incluso de Latinoamérica en convertirse en la punta de lanza de la computación teniendo a su cargo la primera computadora que se compraba para investigación en México. Mucho tiempo mantuvo el liderazgo en este campo y alcanzó fama reputada en la investigación avanzada en un área que en la década de los 50's era inexplorado en Latinoamérica. Sopesando acertadamente la importancia que pronto adquiriría la computación, las administraciones dieron gran impulso al Instituto, y este a su biblioteca quien contó con la fortuna de comenzar desde cero, ya que no existía acervo relacionado con la materia que se estudiaba, ello permitió evitar desviar recursos en inventariar y poner en orden adquisiciones anteriores, como era el caso de la mayoría de las demás bibliotecas, y así todos los recursos económicos bibliotecarios se utilizaron en la compra de material bibliográfico. Esto propició un crecimiento importante en el acervo de esta biblioteca a tal grado que en pocos años alcanzó y

superó las colecciones de bibliotecas más antiguas, pues el IMAS cuenta con más de 15,000 volúmenes y 307 títulos de revistas, mientras que el Instituto de Física (creado en 1938) quien tiene 3 bibliotecas, entre todas juntas llegan sólo a 12,955 volúmenes y 258 títulos de revistas, o el caso del Instituto de Astronomía (formado en 1929) quien llega nada más a 4,700 volúmenes y 125 títulos de revistas.

Los anteriores datos indican la importancia de este Instituto, a quien se le presupuestó en 1989 aproximadamente la partida de 361,790,000 pesos para la compra de material bibliográfico.

VI.2 Antecedentes de la Automatización

La Dirección General de bibliotecas es quien suministra el soporte automatizado a las bibliotecas del Sistema Universitario, esto se traduce en servicios centralizados tales como la entrega de juegos de tarjetas catalográficas que amparan los libros adquiridos por cada biblioteca, y ofrece a usuarios remotos comunicación en línea, vía modem, con la b.d. para consultas sobre los libros adquiridos por todas las bibliotecas de la UNAM. La DGB ha enfocado estos esfuerzos principalmente a la creación de una b.d. monográfica, la cual es la que se tiene en uso para la comunidad.

Para ciertas bibliotecas especializadas como la del IMAS, se necesitan otros servicios automatizados que la DGB todavía no cubre como son el poder abarcar todo el material adquirido por esta Biblioteca, que no solo se circunscribe a libros (monografías), y una falta de flexibilidad para almacenar en la

B.D. de la DGB claves que indiquen arreglos especiales en el acervo de bibliotecas como la del IIMAS.

El IIMAS se encontró con el problema de que, además de libros, adquiere asimismo folletos (impresos de no más de 49 hojas) de los cuales se tienen alrededor de 5000, revistas que ocupan el 50% del acervo total, tesis nacionales y extranjeras que aproximadamente alcanzan 2000 títulos y por último microfichas y programas para computadora. Por otra parte los usuarios de esta biblioteca especializada consultan ciertos grupos de libros que manejan como una unidad, y que no desean tener físicamente dispersada por temas entre los estantes del acervo. tal caso se presenta con manuales de equipos y lenguajes computacionales, memorias de congresos y otros.

Estas causas incidieron para que la biblioteca del IIMAS, sin duplicar funciones de la DGB, pusiera en marcha un proyecto de automatización en su biblioteca. basándose en el sistema MICROISIS, y generara dos b.d. las cuales fueron: "COOPER", en la que se archiva folletos, impresos, las adquisiciones, etc... y "KARDEX" en la cual se guarda información referente a revistas.

VI.3 Descripción de la B.D.

VI.3.1. Diseño

Se eligió MICROISIS debido a la facilidad que encierra para el diseño, manejo y mantenimiento de bases de datos bibliográficas. Para ello la biblioteca cuenta con una microcomputadora compatible con la IBM que tiene 640K de memoria, y una impresora de 180 caracteres por segundo, también tiene terminales remotas

conectadas a la b.d. "libros" de DBB. Después de analizar el problema se consideró crear una base de datos.

La B.D. llamada KARDEX se abocaría al almacenamiento de los títulos de revistas, y así solucionar este problema. Para ahorrar tiempo y esfuerzo, se contactó al CONACYT quien tiene un banco de datos sobre revistas, y se transportaron a KARDEX los registros con los títulos que maneja el IIMAS, con lo que se logró que el conjunto de revistas a clasificar que no se encontraron en la base fuera menor y más manejable.

La base se encuentra actualmente en proceso de captura.

En las siguientes hojas se muestran las tablas de definición de campos que muestran al lector el diseño de la base:

Field definition Table (FDT)

Data Base: KARDEX

Tag:	Name	Len:	Typ:	Rep:	Delm/Pattern
10	FOLIO DE TITULO/CONACYT (K100)	6	X		
12	FOLIO DE TITULO/CICH	5	X		
22	ISSN (C100)	26	X		
30	CODEN (KB00)	6	X		
82	DEWEY (K700)	8	X		
85	CODIGO PAIS DE LA PUB. (K200)	4	X		
86	CODIGO DEL IDIOMA (K500)	4	X		
87	CODIGO DE SITUACION (S200)	1	X		
90	FECHA INICIO (S300)	4	X		
91	FECHA FIN (S400)	4	X		
92	TIPO DE PUB (S500)	1	X		
94	TIPO DE MATERIAL (S700)	1	X		
95	FRECUENCIA (C300)	20	X	R	
220	TITULO/IIMAS	300	X		
222	TITULO CLAVE (M100)	300	X		ABC
210	TITULO CLAVE ABREV. (M200)	300	X		ABC
245	TITULO PROPIO (T100)	750	X		ASU
246	TITULO VARIA (T200)	750	X	R	ASU
510	COBERTURA POR INDICES (S100)	750	X	R	TX
550	ORGANISMO RESPONSABLE (CS50)	300	X	R	
560	CIUDAD (H200)	100	X		
565	ESTADO (H300)	100	X		
570	NUMERO DE EDITORIAL (I100)	6	X		
580	NOMBRE DE LA EDITORIAL (E200)	200	X	R	
590	MENTION RESPONSABILIDAD (C900)	30	X		
759	ES ED. EN OTRO IDIOMA (M300)	750	X	R	TX
760	ES SUBSERIE DE: (M300)	750	X	R	TX
762	TIENE SUBSERIES: (M400)	750	X	R	TX
769	TIENE ED. EN OTRO IDIOMA (M700)	750	X	R	TX
779	ES INSERTO Y/O SUPLEM. (M500)	750	X	R	TX
780	TITULO ANTERIOR (T300)	750	X	R	TX
785	TITULO POSTERIOR (T400)	750	X	R	TX
787	ES TITULO RELAC. CON (M900)	750	X	R	TX
789	TIENE INSERTO Y/O SUPLEM. (M600)	750	X	R	TX
805	SOLICITUDES PERSONALES	30	X	R	
810	RECOMENDACION CC	5	X		
811	RECOMENDACION BISTAC	5	X		
812	RECOMENDACION BSD	5	X		
813	RECOMENDACION MMS	5	X		
814	RECOMENDACION PEIC	5	X		
815	RECOMENDACION MMN	5	X		
817	RECOMENDACION MYM	5	X		
818	RECOMENDACION BIS	5	X		
820	FORMA DE ADQUISICION	4	X		
833	90: POSIBLE DISTR./PROVEEDOR	500	X	R	ABC
834	90: PROVEEDOR ACEPTADO	100	X		
836	90: ESTADO DE LA ADD.	4	X		
837	90: COBERTURA SUSCRIPCION	100	X		
840	90: PRECIO (MONEDA EXT.)	10	X		
841	90: MONEDA	10	X		

Field definition Table (FDT)

Data Base: KARDEX

Tag	Name	Len	Typ	Rep	Delm/Pattern
842 90:	TIPO CAMBIO	10	X		
843 90:	PRECIO (MONEDA NAL.)	12	X		
846 90:	NRO DE PEDIDO/IIMAS	8	X		
847 90:	FECHA DE PEDIDO/IIMAS	8	X		
850 90:	NRO DE RECEPCION/CICH	20	X		
851 90:	FECHA DE RECEP./CICH	8	X		
854 90:	NRO DE GIRO/CHEQUE	20	X		
855 90:	FECHA DEL GIRO	8	X		
858 90:	FECHA RECIBO GIRO/IIMAS	8	X		
859 90:	FECHA ENVIO GIRO AL PROVEE	8	X		
860 90:	FECHA ACUSE O RECEPCION	8	X		
863 89:	POSIBLE DISTR./PROVEEDOR	500	X	R	ABCD
864 89:	PROVEEDOR ACEPTADO	100	X		
866 89:	ESTADO DE LA ADO.	4	X		
867 89:	COBERTURA SUSCRIPCION	100	X		
870 89:	PRECIO (MONEDA EXT.)	10	X		
871 89:	MONEDA	4	X		
872 89:	TIPO DE CAMBIO	10	X		
873 89:	PRECIO (MONEDA NAL.)	12	X		
876 89:	NRO DE PEDIDO/IIMAS	8	X		
877 89:	FECHA DE PEDIDO	8	X		
880 89:	NRO DE RECEPCION/CICH	20	X		
881 89:	FECHA DE RECEP./CICH	8	X		
884 89:	NRO DE GIRO/CHEQUE	20	X		
885 89:	FECHA DEL GIRO	8	X		
888 89:	FECHA RECIBO GIRO/IIMAS	8	X		
889 89:	FECHA ENVIO/GIRO AL PROVEE	8	X		
890 89:	FECHA ACUSE O RECEPCION	8	X		
900	OTRAS BTECAS: (D100/SIGLAS)	1650	X	R	
910	ACERVO IIMAS/CONACYT (D200)	1650	X	R	ABCDE
915	ACERVO (LOCAL)	1650	X	R	ABCDE
917	ENFACADO	100	X		
960	PROPIETARIO (IIMAS/CUERNAVACA)	20	X		
962	COLOR Y MATERIAL	20	X		
963	DATOS X REMESA	1000	X	R	ABCDEFGHIZ
965	DETALLE DE LA REMESA	1000	X	R	ABCDEFGHIZ
999	OBSERVACIONES	1000	X	R	

VI.3.2 Software

a.-Base de Datos

1).-Tipo

Tenemos que esta base de datos es tipo relacional

2).-Capacidad

-Número máximo de registros que pueden almacenarse

Se permite hasta 16 millones de registros en un disco de 500Mb.

-Tamaño máximo del registro

Hasta 8,000 bytes

-Número máximo de campos por registro

Hasta 99 por registro

-Máximo número de subcampos por campo

Sin límite

-Máxima longitud de una clave

La clave puede alcanzar un máximo de 30 bytes

3).-Características generales

-Posibilidad de tratar un subconjunto de regs. como una B.D.

Por medio de un rango o "RANGEFILE"

-Facilidad de uniones

Limitado a imprimir tablas de códigos

-Permite almacenar conjunto de caracteres en diferentes

idiomas / alfabetos

No

-Permite el ordenamiento de acuerdo a información adicional

Colocando la información adicional entre los símbolos <>, por ejemplo: Los 4 <cuatro> jinetes del Apocalipsis.

-Se permiten campos de longitud variable

Si está permitido

-Se permiten campos de longitud fija

Si está permitido

-Se permiten la existencia de subcampos repetibles

Si está permitido

-Existen campos obligatorios

No existen

b.-Data Base Manager System (DBMS)

1).-Lenguaje en el que se montó el DBMS

PASCAL

2).-Mantenimiento

-Bitácora de transacciones para recobrase de caídas del sistema

No

-Bitácora de transacciones para contabilidad

No

-Verificación de tipo de dato

numérico, alfanumérico, de acuerdo a un patrón

-Rutinas de validación propias del usuario

No

-Verificación de campos con valores únicos

No

-Tipo de archivos invertidos

Parece ser que son de tipo BTREE

-Tipos de extracción

Por palabra, término, campo completo

-Eliminación de palabras stop (no significativas)

Si

-Actualización de los archivos invertidos

Si, en línea

4).-Expansión del Manejador

-Lenguaje

Permite que se pueda interfazar con ISISPAS

-Generación de aplicaciones completas

Utilizando el lenguaje ISISPAS

-Contribuciones de usuarios

Se ha propuesto

-Relaciones con otros manejadores de base de datos

DBASE III, LOTUS y REFLEX

-Mejoras

Hacerlo multiusuario

5).-Aplicaciones

-Aplicaciones típicas

Recuperación Bibliográfica, Manejo de personal, Historias médicas, Estudios bibliométricos, Informes de actividades

-Aplicaciones futuras

Utilización de la tecnología CD-ROM, Tesauros, Tipografía, Estadística

6).-Ordenamiento

-Máximo número de niveles de ordenamiento

Hasta 4

-Ordenamiento de caracteres españoles/ acento, ch, ll

No

-Remoción de palabras stop

Un archivo stop

-Conservación de etiquetas del campo

No

-Selección de campos alternativos

No

-Procesamiento de índices KWIC

No

-Manejo de sufijos / prefijos

No

-Especificación de número máximo de claves que se extraerán

No

7).-Características generales

-Identificación de campos

Por etiqueta

-Número de archivos invertidos

Uno

-Búsqueda en varias Bases de Datos

Sólo hay una

-Búsqueda libre de texto

No, sólo sobre campos invertidos

-Operaciones permitidas en cálculo numérico

No

8).-Portabilidad

La portabilidad es a nivel de microcomputadoras y en las

PC XT y AT o clones

c.-Paquetes Programas de explotación

1).-Lenguaje

Los programas de explotación se realizan en PASCAL

2).-Captura de datos

-Definición de pantalla de captura

Si existe

-Facilidad para edición

Bastante fácil de editar

-Candado a campos que se están modificando

No es necesario en versión monousuario

-Manejo de valores default

Si, temporales o permanentes

-Operadores usados

Se utilizan los operadores AND, OR, NOT

3).-Productos resultantes

-Creación automática de un formato de impresión implícito

No

-Truncamiento

Si

-Manejo de tesoros

No

-Despliegue de resultado de búsqueda

versatil

-Páginas lógicas

Si hasta 4

-Tabulados / columnas

No

-Máximo número de registros por página

No

-Corte a medio registro

No

-Lineas de encabezado

Hasta 3

-Encabezado que sea parte de los datos

No

-Selección de campo alternativo si uno no existe

No

-Especificación de cual ocurrencia imprimir

No

-Márgenes, sangrias

Limitado

-Supresión de campos si el registro previo tiene igual contenido

Si, para campo ordenado

-Ordenamiento alfabético de las ocurrencias de un campo repetible

No

4).-Portabilidad

Los programas de explotación pueden exportarse a sistemas

PC XT y AT

d).-Obtención y costos del Software

1).-Proveedor

La UNESCO, y en México el CONACYT

2).-Costo

Sin costo

3).-Equipo

Una PC con mínimo de 256K o 512K para v.2

4).-Sistema operativo

MS-DOS

VI.3.3. Hardware

a.-Tipo

Microcomputadora

b.-Características generales

Computadora:

- Procesador 8088, 8086, 80286, 80386
- Velocidades desde 4 MHertz hasta 32 MHertz (aceptable de 8 a 12 MHertz)
- Memoria principal mínima de 256 Kb, preferente 640 Kb
- Dos drives para diskette (minimamente de 5 1/4", de 360 Kb o mayor capacidad: 1.2 Mb). O en su defecto, y mejor aun, un disco duro (mínimo 20 MB) y drive para floppy disk
- Puerto paralelo tipo CENTRONICS, para impresora, y puerto serial tipo RS-232C para comunicaciones.
- Monitor a color o monocromático con tarjeta CGA o VGA de preferencia en vez de tecnología TTL.
- Teclado básico de 80 teclas, de preferencia de 101 teclas y 12 funciones.
- Fuente de poder mínima de 150 Watts (depende del Hardware).
- Gabinete para un mínimo de 3 unidades de media altura (para el caso de necesitar expansión de memoria masiva).

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

- Al menos tres ranuras para expansión para montar tarjetas adicionales.
- De preferencia que el sistema cuente con llave de seguridad bancaria para protección y control del uso de la máquina.
- Coprocesador matemático no necesario, pero aumenta la velocidad de proceso.
- De preferencia con reloj calendario de tiempo real, muy útil en el control de procesos en bibliotecas.
- Sistema con botón de cambio de velocidad a turbo, e indicador luminoso de cambio.
- Sistema con indicador de acceso a disco duro.

Impresoras:

- Impresión por medio de matriz de punto (lo que permite que se puedan imprimir gráficas).

Las impresoras Laser no son de importancia en la biblioteca debido a que la calidad de impresión de una Laser es útil fundamentalmente en los casos de utilizar paquetes de edición especializados en manejo de textos o en funciones administrativas que implican papelería de calidad y no de cantidad. En la biblioteca saldría caro utilizar este tipo de equipo, la tinta cuesta alrededor de 80 dólares, y su mantenimiento, aunado al enorme costo de una impresora de este tipo las hace onerosas. Las impresoras de matriz son más que suficientes.

- Cabeza de impresión de al menos 9 puntos en la matriz

para una calidad aceptable de impresión.

-Capacidad de imprimir letra de calidad (modo NLQ "Nice Letter Quality").

-Velocidad de impresión de mínimo 180 caracteres por segundo en modo Draft.

-Velocidad de impresión de mínimo 40 caracteres por segundo en modo NLQ.

-Memoria de Buffer mínima de 4 Kb (1.6 páginas carta escritas a máquina en promedio), mejor de 8 Kb (3.2 páginas) lo que permite que la computadora se libere con anticipación aceptable del proceso de impresión.

-Carro de mínimo 10" (hojas de 80 columnas), aunque es conveniente un carro de 15" lo que permite hojas de 132 columnas.

-Permita se tengan al menos dos tipos de letra distinta (por ej. Elite y Pica) y también exista modo condensado para imprimir línea de 132 caracteres en hoja de 80 columnas.

La interfase debe ser paralela compatible con CENTRONICS, como opcional se tendría serial (RS-232C)

-Se pueda emular el modo de la impresora IBM Proprinter principalmente, y/o la de la EPSON, las cuales son standard para muchos paquetes, no sólo de bases de datos, y así ser compatibles con el software de la computadora y poder ser aprovechadas al máximo.

c.-Obtención de costos

Hoy en día las microcomputadoras se pueden obtener con

diferentes distribuidores y de muy varias marcas. Por lo que su adquisición no es ningún problema.

Los costos varían extraordinariamente de acuerdo al fabricante en cuestión, siendo las máquinas de IBM, Hewlett Packard, Olivetti y Unisys las más caras del mercado, aunque sus prestaciones son confiables y el servicio que ofrecen muy bueno.

Una máquina se puede conseguir desde 600 y pico de dólares hasta varios miles. A continuación damos una pequeña tabla con diversos precios para las computadoras e impresoras:

Computadora XT procesador 8088 o NEC V20; 4-10 o 4-12 MHz; un drive floppy disk y hard disk de 20 Mb; monitor TTL o CGA; 640 Kb RAM. Costo en dólares.

IBM: Sistema Personal/2 Modelo 25 (drive diskette de 3.5" de 720Kb, disco duro de 20Mb y monitor con MCGA). Costo de \$2,029.

Hewlett Packard: Sistema VECTRA 286/12-20 (procesador 286, 1Mb de RAM, drive para floppy de 5.25" de 1.2Mb con disco duro de 20Mb, y monitor con tarjeta VGA). Costo de \$2,486.

OLIVETTI: Computadora OLIVETTI PC586 (con monitor y tarjeta VGA, disco duro de 20Mb, floppy de 3.5" de 720Kb). Costo de \$1,831.

ACER: Computadora ACER 500+II (con tarjeta de video triple para MDA, CGA, Hercules). Costo de \$1,390.

TELEVIDEO: Computadora TELEVIDEO XT (monitor TTL). Costo de \$1,330.

GAMA: Computadora GAMA 88 (monitor TTL). Costo \$1,150.

ONYX: Computadora ONYX XT-10 (reloj tiempo real, monitor TTL). Costo de \$1,241.

VI.3.4 Humanware

a.-Capacitación de personal

1).-Responsables

Grupos de usuarios

2).-Material didáctico

Existe en español para la v. 1

3).-Capacitación a instructores

No se tiene noticia

b.-Documentación de los sistemas

1).-Documento introductorio o folleto

Próximo

2).-Calidad de los manuales

Complejo

3).-Traducción al español de documentación

Existe

c.-Grupos de usuarios

1).-Objetivos

Principalmente compartir experiencias

2).-Actividades que realizan

Se realizan reuniones, cursos, y un Congreso Nacional

3).-Vínculo con el centro responsable

A través del CONACYT

VII CONCEPTOS PARA LA STANDARIZACION DE LAS DIFERENTES B.D.

VII.1 Problemas a solucionar

Dado el enfoque con el que se ha abordado la presente tesis, dividiremos los problemas del mismo modo en que se ha dividido los capitulos que tratan de las diferentes instancias estudiadas.

HARDWARE: En el caso de que se desee transportar totalmente las B.D. de un sistema a otro, nos encontramos con el obstáculo principal representado por la capacidad de almacenamiento en disco duro. De las instituciones contempladas, dos de ellas cuentan con microcomputadoras (IIMAS y CUIB) y las otras dos con minicomputadoras (DGB y BN). Los DBMS de IIMAS y CUIB corren en versiones de sistema operativo MS-DOS menores a la 4.0, por lo que sólo se puede manejar particiones de hasta 32 Mb, lo cual impide que una B.D. como la de DGB pueda transportarse ya que ésta ocupa alrededor de 200 Mb. de memoria en disco. Además de que el tiempo de búsqueda sería lo suficientemente largo como para hacer incosteable la utilización de esta B.D., además los backup necesarios serían onerosos si tomamos en cuenta que las PCs tienen drives de 360 Kb. Con BN sucede lo mismo. Sólo es posible la transportación de la B.D. entre CUIB e IIMAS ya que cuentan con hardware de parecidas prestaciones. Entre BN y DGB sí es posible el transporte de B.D. pues ambas cuentan con hardware que permite el almacenamiento de B.D. grandes.

SOFTWARE: Aquí se presentan varias situaciones que el DBMS debe

poder controlar para que las B.D. puedan ser utilizadas en otros sistemas distintos.

-Capacidad del DBMS para soportar el total de registros de la B.D. En este caso tenemos que con los DBMS de BN y DGB no existe problema con ello. En el caso de DBASE III, aunque en teoría permite soportar un gran número de registros, en la práctica, al cabo de unas decenas de miles de regs. el DBMS comienza a dar problemas, por lo que el CUIB no podría soportar las bases de BN y DGB. Con MICRODISIS es posible tener una B.D. más grande, pero los recursos del DBMS serían inadecuados para controlar bases de gran tamaño.

-Capacidad del DBMS para manejar el número total de campos por registro. Aquí no existe problema, ya que ninguna B.D. pasa del máximo número de campos permitidos por cada DBMS.

-Longitud máxima del registro que puede manejar el DBMS. Todos los DBMS no tienen problema con ello excepto el DBMS de IDMSOO (DGB), ya que el máximo tamaño se localiza en la B.D. de BN la cual tiene regis. de 2000 bytes, y el DBMS de DGB puede manejar máximo hasta 2000 bytes por reg., lo cual conlleva algunos problemas si se desea meter números de control en ese registro. Sin embargo es posible partir ese reg. en varios con un código de control lo cual permite (y de hecho así se realiza) manejar grandes trozos de información.

-Posibilidad de manejar subcampos. Si bien los DBMS de DGB y CUIB no pueden manejar formalmente este concepto, pueden manejarlo utilizando funciones especiales para campos string. Por lo que este no sería un problema insalvable.

-Máxima longitud de clave. En este rubro, quien sale con problemas es el DBMS de Microisís (IMAS) ya que sólo alcanza un máximo de 30 caracteres a diferencia de los otros DBMS que sus índices son mínimo de 100 bytes. Aun así la mayoría de las situaciones pueden resolverse con 30 caracteres.

-Tipo de B.D. No existe dificultad en el tipo de B.D. pues todas ellas tienen estructura relacional.

HUMANWARE: Finalmente uno de los principales obstáculos para lograr una mejor standarización de las B.D. se encuentra en las personas que están a cargo, no sólo de las B.D., sino de las políticas y proyectos a seguir en este ramo.

-Falta de conocimiento y coordinación entre los diferentes proyectos de B.D. A pesar que la UNAM engloba una serie de actividades y recursos para un objetivo común, no existe todavía un consenso respecto a las opciones y desarrollos que se crean en el campo de las B.D. bibliotecarias. Ello lleva a un desconocimiento sobre los desarrollos realizados en el Área y que no permiten la creación de B.D. que pueden ser movilizadas de un ambiente a otro. Los recursos que se entregan en estos proyectos pueden optimizarse de conocerse la situación general que presenta el panorama de las B.D. catalográficas de la UNAM.

-Falta de políticas y recursos integracionistas. Las diferentes dependencias que tienen a su cargo las B.D. catalográficas no tienen reuniones entre ellas para definir políticas en este campo. Lo que da por resultado que estas instituciones se comporten como islas sin contacto estrecho sobre los movimientos

a realizar en este campo. Lo que resulta que se dupliquen esfuerzos en vez de compartirlos. Si bien la DGB, gracias al avance de la tecnología, puede ahora, mediante el disco CD-ROM, exportar su B.D. a microcomputadoras, únicamente se tiene comunicación en un sólo sentido, ya que no existe canales de "feedback" que permitan a la DGB compartir experiencias en este campo. Las demás bibliotecas estudiadas no se encuentran en posibilidades de realizar algo parecido.

VII.2. Directrices a seguir para una standarización de B.D. de Bibliotecas Universitarias.

HARDWARE: En este campo no es mucho lo que se puede modificar ya que el hardware no tiene flexibilidad para adaptarse fácilmente a nuevas situaciones. De hecho las B.D. de CUIB e IIMAS son B.D. locales y pequeñas que fueron creadas para cubrir sólo los requerimientos de las instituciones a las que sirven, esto ocasiona que el hardware, y los fondos económicos destinados a $\text{C}\ddot{\text{a}}\text{mputo}$, se orienten a microcomputadoras. A estas dependencias les sería más útil poder tener, y modificar en sus máquinas, subconjuntos de las B.D. de DGB y BN. Los sistemas CD-ROM (Compact Disk Read Only Memories, de tecnología láser) que la DGB está entregando a las bibliotecas que dependen de ella, pueden servir en ese caso, ya que en cada Compact Disk se encuentra almacenada toda la B.D. de DGB, sin embargo hasta el momento sólo sirven como fuente de consulta, sin incorporación de su información a otras B.D. Finalmente el hardware no representa el

principal escollo a superar para una standarización de las B.D.

SOFTWARE: El software tampoco es un obstáculo importante, pues las B.D. serían exportadas en código ASCII o EBCDIC que son los que tienen mayor popularidad en el medio computacional, y teniendo la programación adecuada y específica a cada sistema, es posible ajustar la B.D. foránea al ambiente receptor

HUMANWARE: Aquí reside la principal dificultad para lograr la standarización. Dado que no existe conocimiento de un grupo administrador de una de las B.D. sobre cualquier otra de las B.D. es necesario:

-Conocer la estructura de las distintas B.D. de las instituciones mencionadas y las prestaciones de sus DBMS para discutir y obtener la normalización de las B.D. que se exporten logrando así un formato standard que permitiría al transportar una B.D. que:

-El impacto sobre el hardware de la B.D. transportada, o un subconjunto de ella, fuera mínimo al no traer en la estructura información que no pueda almacenarse fácilmente (longs. de regs y longs de campo adecuados, inexistencia de subcampos, etc.)

-Que el software necesario para manejar y gustar las B.D. foráneas o subconjuntos de ellas, sea el más sencillo en cada caso. Para lograrlo la estructura de la B.D. importada no debe diferir en gran medida de la local.

-Existan políticas de cooperación a nivel de departamentos de Cómputo para la realización de reuniones en las que exista

contacto entre el personal de las distintas instituciones para el intercambio de ideas, problemas y conocimientos que enriquecerán el bagaje teórico-práctico de los diseñadores, programadores y administradores de las B.D. Así como contemplar la posibilidad de lograr reglas específicas para el intercambio de información de sus respectivas B.D.

Para el manejo de la información catalográfica se propone el uso del formato MARC, o sea el formato que permite clasificar la información no sólo de libros sino de revistas y otras entidades que son necesario guardar. Esto también se considera atinado porque el DBMS MINISIS de BN tiene salidas a formato MARC, DGB maneja una versión recortada de ese formato, LOGICAT (CUIB) asimismo tiene su información almacenada en un formato recortado del MARC, y el IHAS tiene la B.D. Kardex orientada a este mismo formato pero para revistas.

El transporte de la B.D. standard, vendría en un formato basado en una versión más sencilla del propuesto por Stephen Faunce en su libro: "MARCAL. Manual de Automatización de las Reglas Catalográficas para América Latina". Ya que ocupa menos espacio y es más sencillo de utilizar, además de que cumple con las necesidades estudiadas.

La estructura del formato de transporte propuesto constaría de un campo numérico de control, campo de continuación, un directorio del registro y los campos de información.

El campo de control o guía, sería de 6 bytes y contendría un número secuencial y único para cada registro, el cual serviría

para los manejos y transformaciones que se debieran hacer con la información para introducirla en la E.D. receptora.

- El campo de continuación (un byte) indica con "0" si la ficha catalográfica está contenida en su totalidad en ese registro. En cambio un "1" informa que la continuación de la ficha se encuentra en el siguiente registro, el cual contendría el número de control y el campo de continuación, seguido de los campos con información, pues ya no se necesitaría el directorio.

- El directorio del registro contendría una serie de entradas de longitud fija cada una, donde se tendría el número de etiquetas que contiene esa ficha (2 bytes), el identificador de campo (3 bytes), la longitud de ese campo (4 bytes), la posición de inicio de ese campo (4 bytes) y un campo (1 byte) que indica desde que posición, relativa al campo de información (algo así como un offset), se debe tomar la cadena de datos para fines de alfabetización o manejo de sorts. Se repiten este grupo de campos tantas veces como etiquetas tenga la ficha, excepto la entrada de número total de etiquetas.

- Los campos de información son unidos formando una cadena continua, pues en el directorio se tiene la máscara para partirlos.

Todos los registros vendrían del mismo tamaño físico aunque la cantidad de información relativa a cada ficha bibliográfica sería variable. En caso de sobrar espacio en el registro, se llenaría el sobrante con blancos; en caso de no haber la información de una ficha en un solo registro, se tendría en el campo de continuación una bandera que indicaría que el sig. registro es continuación del anterior y así sucesivamente. Se

propone que el tamaño del registro sea de 512 bytes que es el tamaño de sector de las microcomputadoras, y así en ellas no se desperdiciaría mucho espacio de almacenamiento.

En la gráfica (2) se observa la estructura propuesta para el transporte.

Debido a que la B.D. de BN tiene el conjunto de etiquetas de MARC más completo, faltando sólo la etiqueta \$050 de clasificación LC que maneja DGB, y que a la vez contiene todas las etiquetas que desarrolla LOGICAT (CUIB), será entonces el formato de BN el que se utilice como standard para la estructura que permita mayor compatibilidad entre las B.D.

El caso del LINAS es aparte ya que al estudiar esta biblioteca, se encontró que su B.D. está orientada a la explotación de revistas, la cual también es un importante filón de información de la Universidad. La DGB cuenta asimismo con otro B.D. llamado SERIUNAM, que almacena publicaciones periódicas. Afortunadamente el formato MARC también se ha utilizado en estos casos, por lo que también se mostrará una estructura que podría ser standard para las B.D. universitarias que se diseñen para el manejo y explotación de publicaciones periódicas, complementando así a las B.D. que manejan información monográfica.

Finalmente a continuación se muestra las estructuras de B.D. tentativas que pudieran servir como norma para el intercambio de información entre las dependencias tomadas para este estudio.

FORMATO PROPUESTO PARA LA BASE STANDARDIZADA

NUMERO DE CONTROL	CAMPO DE CONTINUACION	NUMERO DE ETIQUETAS	IDENTIFICADOR DE 1a ETQA	LONGITUD CAMPO 1o	POSICION CAMPO 1o	CAMPO ALFAB 1o
-------------------	-----------------------	---------------------	--------------------------	-------------------	-------------------	----------------

IDENTIFICADOR DE 2a ETQA	LONGITUD CAMPO 2o	POSICION CAMPO 2o	CAMPO ALFAB 2o
--------------------------	-------------------	-------------------	----------------

IDENTIFICADOR DE n ETQA	LONGITUD CAMPO n	POSICION CAMPO n	CAMPO ALFAB n	CAMPO 1o INFORMACION	CAMPO 2o INFORMACION
-------------------------	------------------	------------------	---------------	----------------------	----------------------

CAMPO n INFORMACION

- NUMERO DE CONTROL: Este es un numero entero y es unico para cada registro.
- CAMPO DE CONTINUACION: Indica con una bandera si es continuacion del anterior registro.
- TOTAL DE ETIQUETAS: Nos muestra cuantas etiquetas tiene esa ficha catalografica.
- IDENTIFICADOR: Tiene el numero de la etiqueta del campo.
- LONGITUD CAMPO: Se guarda la longitud que tiene ese campo.
- POSICION CAMPO: Se almacena la posicion de inicio del campo.
- CAMPO INFORMACION: Se guarda la informacion perteneciente al identificador.

La B.D. standard contendría las siguientes etiquetas en formato

MARC II para monografías:

ETIQUETA CAMPO

008 En este campo, llamado "campo fijo", se guarda una serie de valores en los que su posición es importante pues se almacenan y recuperan a través de una máscara; los valores son:

valor	pos.	observaciones
fecha	1	S - fecha conocida C - Copyrigh M - fecha múltiple Q - fecha incompleta H - otras fechas
1a fecha	2-5	son 4 dígitos
2a fecha	6-9	son 4 dígitos
origen de catalogación	10-12	iniciales de la biblioteca.
clase de ilustraciones	13-16	A - ilustraciones B - mapas C - láminas O - otros
presentación física	17	A - micropelícula E - microficha D - fotocopia Z - otros
forma de contenido de mat.	18-21	A - resumen analítico B - bibliografías C - catálogos D - diccionarios E - enciclopedias H - manuales I - índices Q - Biografías P - libros de texto program. R - directorios S - estadísticas Y - anuarios
tipos de publicaciones	22	G - origen nacional

E - origen extranjero

idioma 23-25 ENG - inglés
 FRE - francés
 ITA - italiano
 GER - alemán
 POR - portugués
 OTR - otro
 español son tres blancos

Asiento Principal parte
del título 26 1 - Si forma parte
 blanco - no forma parte

código de catalogación 27 blanco - 1a edición de reglas
 de catalogación angloamericanas
 2 - 2a edición de reglas de
 catalogación angloamericanas

área geográfica 28-38

ETIQUETA CAMPO

020 ISBN
050 Clasificación LC (Library Congress)
082 Clasificación DEWEY (Decimal)

Asiento principal
100 Autor personal
110 Autor corporativo
111 Congreso o reunión
130 Asiento bajo título uniforme

Título

240 Título uniforme
245 Título propiamente dicho

250 Edición
260 Pie de imprenta
300 Descripción física del libro.
430 Serie
500 Notas generales
501 Notas que indican más de una obra en una encuadernación
502 Notas de tesis
505 Notas de contenido

Encabezamientos de materia

600 Autor como materia
610 Organismos Corporativos como materia
611 Congreso como materia
630 Título uniforme como materia
650 Encabezamientos de materias generales

Asientos secundarios
 700 Autores personales
 710 Autores corporativos
 711 Congresos
 730 Títulos uniformes
 740 Títulos abreviados u otros títulos

Asiento secundario de serie

800 Asiento secundario de serie nombre personal/título
 810 Asiento secundario de serie nombre corporativo/título
 811 Asiento secundario de serie nombre conferencia/título
 830 Asiento secundario de serie encabezamiento tit. convencional
 840 Asiento secundario de serie-título

----- * -----

La B.D. standard contendría las siguientes etiquetas en formato

MARC II para publicaciones periódicas:

ETIQUETA CAMPO

008	Campos Fijos	valor	pos.	observaciones
	código de situación		1	e = publicada en la actualidad d = terminada u = desconocida
	fecha inicio de publicación		2-5	
	fecha terminación publ.		6-9	
	código de país de publ.		10-12	
	frecuencia		13	blanco = sin especificar d = diario c = bisemanal w = semanal e = quincenal m = mensual b = bimestral q = trimestral t = cuatrimestral f = semestral

a = anual
 g = bienal
 h = trienal
 z = otra
 u = desconocida

señalador de
 tipo de publ. 14

blanco = ninguno de los
 mencionados.

p = revista
 m = serie monográfica
 n = periódico

tipo de
 material 15

blanco = sin especificar

b = bibliografías
 c = catálogos
 i = índices
 a = resúmenes analíticos
 d = diccionarios
 e = enciclopedias
 r = directorios
 s = estadísticas
 h = biografías
 o = reseñas de libros
 y = anuarios
 l = leyes
 w = compendios legales
 g = estudios legales
 n = casos y notas legales

idioma 16-18

ENG - inglés
 FRE - francés
 ITA - italiano
 GER - alemán
 POR - portugués
 DTR - otro
 español con tres blancos

ETIQUETA CAMPO

022 ISSN
 030 Designación CODEN (abreviación asignada a cada título por el Franklin Institute)
 050 Clasificación LC
 082 Clasificación DEWEY
 210 Asiento o título abreviado
 222 Título clave
 245 Título completo
 246 Formas variantes del título
 260 Pie de imprenta: Editorial, Ciudad y Estado
 510 Notas de inclusión en índices o compendios
 550 Notas de organismos o grupos responsables
 570 Notas referentes al "Editor"

760 Indica que se trata de una subserie de...
762 Indica que existen subseries
770 Indica si tiene inserto o suplemento
777 Indica si es inserto o suplemento en otra serie
780 Nombre del título precursor de la serie
785 Nombre del título posterior de la serie
787 Asiento de relación sin especificar

Diversas adaptaciones del formato MARC han sido las que se han implementado en las bibliotecas universitarias que han automatizado sus catálogos. Esto implica que el proceso de standarización para lograr un formato único, que permita amplia compatibilidad entre las mismas, sea una versión del mismo formato, que incluya las etiquetas que sirvan a todas las B.D. estudiadas.

BN: por ser una institución a nivel nacional, tiene por fuerza que ser más rigurosa en el asentamiento de la información que almacena, por lo que su B.D. tiene mayor cantidad de etiquetas MARC que las demás. Esto hace que la información de las otras B.D. no le sea muy útil ya que no cumplen con la rigurosidad que necesita para lograr su cometido. Pero en cambio puede exportar sus datos, los cuales cumplirían las especificaciones de las demás B.D., aumentando los servicios que puede prestar su sistema computacional a la comunidad, no sólo universitaria.

DGB: esta dependencia es la que sigue en importancia, su B.D. es bastante completa respecto a la cantidad de información guardada por ficha catalográfica. De hecho existe ya la exportación de su B.D. completa en discos compactos para tecnología láser, estos discos ya se están entregando a las dependencias del Sistema Bibliotecario Universitario junto con el equipo necesario para manejar los CD-ROM. Esto puede convertir al formato MARC/DGB en un standard para las bibliotecas que forman parte del Sistema

universitaria de bibliotecas, pero el formato propuesto por el tesista es más universal ya que permite que la B.D. que se desee pueda tener mayor compatibilidad.

CUIB: Teniendo la biblioteca de este Centro su B.D. en microcomputadora, aumenta la popularidad de la misma para otras dependencias de tamaño medio que tienen centros de información que sólo pueden costearse este tipo de hardware, y que desean diseñar sus B.D. o importar datos de otras. Como sería el caso de instituciones que tengan en su haber el sistema SABE. Asimismo el CUIB tendría la posibilidad de incorporar fichas catalográficas de otras bibliotecas.

IIMAS: Este caso es especial ya que su B.D. versa sobre el almacenamiento de información de publicaciones periódicas, por lo que no es de gran interés para el caso de bibliotecas que tienen B.D. orientadas a publicaciones monográficas. Sin embargo es un campo en el cual el desarrollo es incipiente y tiene un horizonte prometedor. DGB cuenta con la B.D. SERIUNAM, el CICH cuenta con centros de información sobre la misma cuestión, y el CONACYT es un importante proveedor de este tipo de datos. Debido a ello se consideró conveniente que existiera también una standardización sobre la estructura de esta clase de B.D. ya que prometen ser de las más ágiles, dinámicas y consultadas dentro del ámbito bibliotecario.

Respecto al hardware, se encuentra que BN y DGB tienen facilidades para el intercambio de información entre ellas ya que pueden usar cintas para ello. En cambio CUIB e IIMAS, debido al equipo que

tienen, no les es posible usar otro medio de transporte mas que los diskettes, los cuales no son herramientas que manejen las anteriores bibliotecas. Una solución para ello caso seria la comunicación en línea para el transporte de datos. Los problemas a solucionar son que seria muy costoso la trasmisión de datos ya que estos deberian viajar por el sistema telefónico el cual tiene un ancho de banda reducido y los trenes de pulsos viajarían a 1200 bauds lo cual implicaría mucho tiempo para obtener un conjunto aceptable de información, además mientras DGB cuenta con medios para la conexión por modem, BN no tiene aun implementado un sistema de este tipo.

No obstante lo anterior, el impacto, cada vez mayor, que ocasionan las computadoras dentro del ámbito de las bibliotecas llevará a que estas necesariamente homogenicen su información para que pueda ser compartida por todas ellas, y ahorren esfuerzos, principalmente en software y en la captura de datos. La Universidad de Guadalajara se ha dado cuenta de ello y ha firmado un convenio con la DGB para implementar en su biblioteca un sistema parecido al de esta institución. El tesisista diseñó y creó programas de soporte y menús, y asimismo modificó sistemas existentes para adaptarlos a las necesidades de la biblioteca tapatia; lo cual le ha ahorrado a esta mucho tiempo y dinero, cosa que no sucederia si hubiera comenzado desde cero, en software y captura de fichas catalográficas, pues cuentan ya con la B.D. de la DGB. Ahora sólo tienen que adecuar y eliminar lo que no les es necesario. Si esto se hace con instituciones foráneas, en la UNAM con mayor razón debe existir un modelo standard para nuestras propias bibliotecas que evite dupliquemos

esfuerzos y gastemos recursos extra, que nuestra Casa de Estudios
podria emplear en nuevas inversiones.

APENDICE

EL MODELO DE DATOS RELACIONAL DE LA IDM 500

1. Definición de el Modelo de Datos

La definición está dada por la combinación de tres componentes:

- a.- Una colección de tipos de estructura de datos (o sea los bloques de construcción de cualquier B.D.)
- b.- Una colección de operadores o reglas de inferencia, las cuales pueden aplicarse a cualquier ejemplo válido de los tipos de datos mostrados en (a.-), para recuperar o derivar datos de cualquier parte de esas estructuras en la combinación que se desee.
- c.- Una colección de reglas de integridad, las cuales implícita o explícitamente definen la consistencia de la B.D. y/o los cambios de estado de esa B.D. (estas reglas se conocen también como insert-update-delete rules).

Los operadores relacionales de (b.-) y las reglas de integridad de (c.-) están fuertemente influidos por el "implementation - dependent", o sea por la computadora y Sistema Operativo bajo el cual se asentará la B.D.

2. La estructura de datos en el Modelo Relacional

La estructura básica en el modelo de datos relacional es la relación. La cual se ha descrito como "tabla", "matriz", "flat file" (literalmente "archivo plano"). A continuación se define el

término:

Sea D un conjunto ordenado de conjuntos D_i para $1 \leq i \leq n$
esto es:

$$D = (D_i), 1 \leq i \leq n$$

Una relación en D se define como cualquier subconjunto del Producto Cartesiano de D , y los conjuntos D_i se llaman dominios de la relación.

En otras palabras, R es una relación en D si y sólo si :

- (1) R es un conjunto
- (2) Cada miembro de R es un conjunto ordenado

$$(x_i : x_i \text{ es un elemento de } D_i \text{ para } 1 \leq i \leq n)$$

Cada elemento de R se conoce como tupla (nombre recortado de "n-tuple", el cual es un conjunto ordenado de n elementos), al valor de n se le llama el grado de R ; en particular, si $n = 2$, entonces a R se le nombra relación binaria. El conjunto consistente del i elemento de cada tupla de R se le llama el atributo de R , y se dice que está definido sobre el dominio D_i .

Existen varias distinciones entre una relación tal como se ha definido, y los files, tablas, o mátrices.

Tradicionalmente los archivos (files) contienen registros, y el concepto de registro siempre incluye una correlación con su representación física. En particular la correspondencia lógica y física que existe en los campos adyacentes en un registro, es algo comúnmente explotado por los programadores. En el modelo de datos relacional, no obstante, la adyacencia de atributos no es únicamente lógica. El implementador de un DBMS relacional es

libre de usar cualquier representación física. En la implementación de Britton Lee los atributos de una relación son mencionados por sus nombres más que por su secuencia en la relación.

Un flat file es considerado usualmente como un archivo secuencial, con conceptos de "primer", "último", "previo", "siguiente" registro. Una relación es un conjunto, y los conjuntos por definición no son ordenados; por lo que los conceptos de correspondencia no son sintácticos sino más bien semánticos para las relaciones.

Similarmnte, las representaciones físicas de tablas y matrices utilizados en memoria son explotadas generalmente por motivos de eficiencia utilizando subíndices y otros esquemas de direccionamiento. Esta manera de procesar las tablas y matrices es posible porque:

- (I) Las entradas de una columna de una tabla son típicamente de longitud constante.
- (II) Los elementos de una matriz son típicamente de longitud constante.
- (III) Los elementos de una fila o columna son generalmente adyacentes en memoria, como los campos sucesivos en un registro de un archivo.

No existen tales restricciones en la representación física de una relación. Tenemos entonces que si bien los atributos de una relación son lógicamente análogos a las columnas de una tabla o de una matriz de dos dimensiones, no son lo mismo.

3. Operadores Relacionales

Varios tipos de operadores se utilizan para generar nuevas relaciones a partir de las ya existentes. Estos caen dentro de dos clases básicas: El álgebra relacional y el cálculo relacional. Ambos pueden verse como sistemas notacionales que describen nuevas relaciones derivadas de otras existentes. En cada uno de ellos una relación derivada se representa como una expresión simple en la cual las variables son relaciones existentes y sus atributos. Ambos el álgebra y el cálculo comparten importantes ventajas sobre otros lenguajes usados para acceder a B.D., ellos expresan los resultados en términos de conjuntos completos de datos así que en términos de operaciones sobre registros individuales. El cálculo, como se verá después, tiene una ventaja adicional: es menos procedural que el álgebra, y los lenguajes no-procedurales se consideran herramientas más productivas para los programadores que los tradicionales lenguajes más procedurales.

3.1. Álgebra Relacional

Una expresión en el álgebra define una secuencia específica de operaciones para generar una relación de resultado. Cada operador del álgebra es unario o binario (tome uno o dos operandos). Las operaciones algebraicas caen dentro de dos clases: el tradicional lenguaje de operaciones y las operaciones relacionales especiales, como sigue:

Juego de Operaciones

Operaciones Relacionales

unión

proyección

selección

asociación

intersección

producto

diferencia

Nótese que para los operadores unión, intersección y diferencia, los operandos deben ser del mismo grado y sus correspondientes atributos deben estar definidos en dominios comparables.

Unión

A UNION B = $\{a, x \mid x \text{ es un tuplo tanto de A o B o ambos}\}$

Intersección

A INTERSECT B = $\{a, x \mid x \text{ es un tuplo tanto de A como de B}\}$

Diferencia

A DIFFER B = $\{a, x \mid x \text{ es un tuplo de A, pero no es un tuplo de B}\}$

Concatenación

A TIMES B = $\{a, (x, y) \mid x \text{ es la concatenación de un tuplo de A y un tuplo de B, en ese orden}\}$

Obsérvese que TIMES tal como se ha definido es no conmutativo; esto es, A TIMES B, no es igual a B TIMES A, en general. A TIMES B es conocido también como el producto Cartesiano de A y B, y usualmente el producto cruz de A y B, este operador es análogo al producto Cartesiano de los conjuntos.

Selección

El operador SELECT - WHERE genera una relación la cual es un subconjunto del operando.

SELECT A WHERE <calificación> = {x : x es un tuple de A y <calificación> es verdad para x}

Donde <calificación> es una función booleana sobre los atributos de x. Por ejemplo, si A es una relación cuyos tuples representan empleados, y un atributo se llama "salario", entonces:

```
SELECT A WHERE salario > 10000
```

entiéndose entonces una relación resultante que consista de aquellos tuples de A donde el atributo "salario" sea mayor de 10000.

Proyección

El operador PROJECT-OVER genera una relación cuyos atributos constituyen un subconjunto de los atributos del operando:

```
PROJECT-OVER lista de atributos
```

Esto obtiene la relación en la que se descartan todos aquellos atributos que no se encuentran en la lista de atributos. Por ejemplo, si A es una relación cuyos tuples representan empleados, y dos de sus atributos son "numero" y "salario", entonces:

```
PROJECT A OVER numero, salario
```

da por resultado la relación A con todos sus atributos, excepto "numero" y "salario" descartados.

Asociación

JOIN A AND B OVER <calificación> =

SELECT (A TIMES B) WHERE <calificación>

donde <calificación> especifica que un atributo de A y un atributo de B, definidos sobre dominios comparables, son iguales.

2.3 Cálculo Relacional

El cálculo relacional, al igual que el álgebra relacional, es un sistema notacional para describir nuevas relaciones resultantes derivadas de relaciones ya existentes. La principal distinción entre las dos notaciones está en que no se especifica secuencia de operaciones. En otras palabras, es un lenguaje no procedural. La forma general de una expresión en el cálculo es:

<lista-objetivo> where <calificación>

donde <lista-objetivo> es una lista especificaciones de dominio para expresar la relación resultante en términos de los atributos de las relaciones existentes, y <calificación> es un expresión booleana dada en términos similares. La sintaxis semántica de las expresiones del cálculo relacional, son detalladas a nivel de implementación; para nuestro caso la versión de cálculo utilizado por Oracle que se amplían en el "IBM Software Reference Manual" y no se examinarán a profundidad en este apéndice. Unos cuantos ejemplos mostrarán las características básicas del cálculo al final del apéndice.

4. Integridad

Las reglas de integridad que gobiernan la consistencia de la B.D. también son detalladas a nivel de implementación. Ellos son de

crucial importancia para los usuarios, y deben de estar explícitamente definidas en cualquier DBMS relacional.

En la IBM 500, el mecanismo de integridad más básico es el sistema de cerradura (locking system), ya que se permiten múltiples operaciones de lectura concurrentes en una relación, mientras que cualquier operación de escritura bloquea todos los accesos concurrentes.

La mayoría de las funciones del mecanismo de cerradura son automáticas e invisibles para el usuario, pero algunas están bajo control explícito. Por ejemplo, un grupo de comandos con los que se desea realizar una transacción en la B.D., pueden conjuntarse en un bloque para realizar esa transacción con las instrucciones "begin" y "end", y si sucediera algo como falla de luz, etc., durante el proceso, todos los efectos que se hubieran realizado durante la transacción son anulados. Manteniéndose así la consistencia de la B.D., especialmente durante las actualizaciones de la misma, también se evita que las lecturas a la B.D. obtengan datos incorrectos en los casos de que sean partes actualizadas de la B.D. y partes que todavía no se han licenciado a actualizar.

Los eventos de seguridad e integridad se superlapan significativamente, y cualquier sistema práctico debe proveer de ambos controles automático y explícito sobre el acceso a la información. Mayor información sobre los mecanismos de seguridad e integridad de la IBM 500 se discute en el "IBM Software Reference Manual".

5. DBMS Relacional

Para un programador de B.D. el poder concebir el acceso a la información en un alto nivel de abstracción, simplifica el diseño de aplicación y su implementación, todas las relaciones lógicas para manejar el medio de almacenamiento físico y de creación, de direccionamiento, y de mantenimiento del esquema físico, se encuentran dentro de las funciones del DBMS. Por lo tanto el código de aplicación es dependiente de la aplicación, no es dependiente de la B.D. Así que todo el manejo de almacenamiento y control del desarrollo se realiza aislado del código de aplicación.

Una gran aplicación de este tipo de implementación es que la aplicación lógica no es sensible al desempeño de los eventos. Por ejemplo, una aplicación que corre despacio, puede incrementarse grandemente con la creación de índices. Esta operación no causa impacto sobre el código de aplicación. Esto significa que un prototipo de un sistema completo puede construirse y analizarse, y la configuración final en disco y su estructura de índices pueden determinarse más tarde en el desempeño del producto sin causar impacto sobre el código de aplicación.

Ejemplos:

Se utilizará el lenguaje IDL (basado en cálculo relacional) para manejar las operaciones básicas del Álgebra relacional.

Para entrar a este Ambiente, es necesario recibir desde el Sistema Operativo de la computadora huesped la orden: "IDL"

Ella nos introduce al ambiente de la IDM 500. El prompt de IDL es un número seguido del signo ">", y para indicar que se envía a ejecutar la instrucción es necesario teclear la palabra "go".

1) open mbasededat go

1) 14 > UNION B

2) no existe ningún tipo de recuperación que permita recuperar

3) la union de dos relaciones. Se debe unir de A (append) a B

4) union debe generarse en una tercera relación ?

5)

6) range of a is a

7) range of b is b

8) retrieve (a,all) go

campos color

0 amarillo

1 azul

2 blanco

3 verde

4 naranja

10 rojo

100 verde

1 tuples affected

11 retrieve (b,all) go

campos

1

2

3

4

5

1 tuples affected

14 > INTERSECT B

2)

11 retrieve (a,campus) where a=campus & b=campus go

campoa

1
3
4

3 tuples affected

1) /* a MINUS b */

2)

3) retrieve (a.campoa) where

4) count(b.campoa by a.campoa where b.campob = a.campoa) >= 0 go

campoa

0
2
10
100

4 tuples affected

1) /* SELECT a WHERE (condicion de seleccion) */

2)

3) retrieve (a.campoa) where a.campoa = 1 or a.campoa = 2 go

campoa

0
1
1
1
10
100

5 tuples affected

1) /* PROJECT -- OVER -- (atributo 1, atributo 2)

2)

3) retrieve (a.campoa, a.campo) go

campoa color

0 amarillo
1 azul
2 blanco
3 rojo
4 naranja
5 verde
10
100 verde

7 tuples affected

```
1) /* JOIN a AND b WHERE <condicion del join> */
2)
3) retrieve( a.all,b.all) where a.campos = b.campos go
```

campos	color	campos
1	azul	1
3	cafe	3
4	naranja	4

3 tuples affected

GLOSARIO

Archivo

Conjunto de juegos de datos para una aplicación determinada

Archivo Índice

Un archivo índice o indexado, tiene apuntadores a bloques de datos y si realiza ordenamiento lógico sobre la B.D.

Archivo Invertido

Un archivo invertido es un árbol binario construido con todas las llaves de una B.D. y que no realiza ordenamiento lógico sobre la B.D.

Atributo

La definición exacta de atributo se localiza en el apéndice

Árbol B

Árbol B, es una forma particular de índices de niveles múltiples en estructura de árbol. Según lo describe Knuth, el índice se compone de dos partes: el conjunto secuencial y el conjunto índice. el conjunto secuencial consiste en un índice denso (índice denso significa que el índice contiene una entrada para cada ocurrencia de registro almacenado en el archivo indexado) de un solo nivel para los datos reales; las entradas en el índice se agrupan en bloques, y estos se encadenan entre sí de modo que el ordenamiento del archivo de datos representado por el índice se obtiene al tomar las entradas (en orden físico) del primer bloque de la cadena, seguidas por las entradas (en orden físico) del segundo bloque de la misma, y así sucesivamente. De este modo el conjunto secuencial proporciona acceso secuencial rápido a los datos, el conjunto índice, a su vez, proporciona acceso directo rápido al conjunto secuencial (y por tanto a todos los datos). El conjunto índice es en realidad un índice en estructura de árbol para el conjunto secuencial. No tenemos estruobos, el conjunto índice es el árbol B real.

COMPROM

Compact Disk Read Only Memory, dispositivo que utiliza tecnología láser para la lectura de discos compactos de 5.25" que permiten almacenar hasta 300 Kb.

Clasificación LC

Clasificación Library Congress, este tipo de clasificación se hace en los siguientes elementos:

Nombre (autoridad) con nota

Cuando aparece la clasificación del tema sobre el que versa el libro, la clasificación del tema es obligatoria.

La siguiente parte (nota) no es obligatoria y se le puede

como decimales, sirven para ajustar con mayor precisión el tema.

Sigue al cutter de tema Ann, indica sobre que tema específico trata la obra. No es obligatorio.

Sigue después el cutter de autor Ann, que nos especifica que autor o cual obra de él, se tiene en el material bibliográfico. Es obligatorio este elemento.

Finalmente viene la fecha Ann, que nos diferencia una edición de obra, para el caso de reimpresiones. No es obligatorio.

Ejemplo: 2699.4M2/F38 Es la clasificación de el libro: MARCEL: Manual de automatización de las Reglas Bibliotecarias para América Latina Faunce Stephen, 1a ed. 2699.4 nos dice que la obra trata sobre bibliotecología.

M2 indica que el tema específico es automatización con formato MARC.

F38 versa sobre el autor Faunce.

Como no existe 2a ed. no se tiene campo de fecha.

Copy To

Instrucción de REASIII. Copia todo o parte de una B.D. a un archivo.

Hitfile

El Hitfile (archivo de éxitos) es el archivo resultante que almacena todos los registros que cumplen con la condición especificada en el proceso de búsqueda. En MINISIS y MICROISIS.

ISISPAS

Es un subconjunto del lenguaje de programación PASCAL, que existe en Microisis, y que contiene además herramientas para manejar la B.D.

ISAM

Keyed sequential access method. Sea método de acceso secuencial por claves. Distribución de archivos y grupo de archivos se organizó que permitan, en conjunto, al usuario tener la siguiente estructura para archivos, mediante el contenido del campo de claves y en el orden secuencial de acuerdo con el orden de los contenidos del campo de claves.

IRIC

Keyword-in-context (contexto de palabras clave). Índice de palabras interconectados, ordenados según el uso de palabras claves, sea exista una entrada con índice para cada palabra clave significativa del título.

MARC

Machine readable Cataloging. El formato MARC, desarrollado por la Biblioteca del Congreso de los EE.UU. provee el intercambio de información bibliográfica en forma mecánicamente legible entre los centros interesados.

Páginas Lógicas

Instrucción de MINISIS que permite colocar en columnas la información dentro de una página física, (tres págs. lógicas indican tres columnas en la hoja de impresión).

Palabra Stop

Cuando en MINISIS se invierte por palabras, para evitar almacenar en el archivo invertido palabras inútiles tales como los artículos, conjunciones, etc. se utiliza un archivo stop (word stop file) donde se guardan esas palabras; al invertir el sistema busca la palabra en ese archivo y si la encuentra, la ignora.

Proceso Técnico

Cuando un libro se adquiere en la biblioteca, el departamento técnico debe procesarlo para analizar el material bibliográfico en detalle para poder extraer todos los elementos importantes que servirán para que el usuario pueda aprovechar al máximo el catálogo de la biblioteca. Además el proceso técnico servirá para poder asignarle, al libro, un lugar en los estantes que sea coherente con el material ahí almacenado.

Registro

Se trata de un conjunto de campos; la información que se relaciona con una área de actividad en una explotación de procesamiento de datos (por ejemplo, toda la información sobre un artículo de inventario). A veces se denomina elemento de datos o ítem. Más información se encuentra en el apéndice.

Reindex

Instrucción de DBaseIII. Actualiza los índices de una B.D.

Retrieve Info

Al recuperar la información de una búsqueda, se almacenan los resultados en otra relación. En lenguaje SQL.

Set Carry

Instrucción de DBaseIII. Permite que los datos se acumulen. Cada nuevo registro que ingresará a la B.D.

Set History

Instruc. de DBaseIII. Convierte el sistema de comandos en comandos incluidos.

Set Index

Instrucción de DBaseIII. Sobre los archivos asociados de una B.D.

Set Relation

Instrucción de DBaseIII. Directa por B.D. mediante una llave.

Store Command

Es un conjunto de instrucciones en lenguaje IDL que es creado por el usuario y que se guardan en un archivo en la IDM. El archivo con comandos se ejecuta al llamarse por su nombre. De ese modo store command es un macro y pueden pasarsele parámetros.

Tuple

La definición exacta de tuple se localiza en el Apéndice.

Union

Instrucción de HIRIS, que permite unir registros dos o más I.D. a través de un campo común.

View

En el lenguaje IDL, la instrucción create view crea una relación virtual que el usuario puede consultar. La información de otra manera, sin necesidad de crear relaciones físicas.

El Almacén (DBE)

Una definición especializada donde se realiza almacenamiento para catalogar la información de las producciones a nivel de país se puede manejar.

BIBLIOGRAFIA

- Deaton Lee Inc. IBM Sistema de Recursos Humanos, Deaton Lee Inc. 1984.
- Francisco, Charlotte et al. **LIBRARIAS: Sistema Automatizado para Bibliotecas**, Dirección General de Bibliotecas UNAM, 1979.
- Dyck, Robert B. **Introducción a los Sistemas de Datos con IBM 311** Eng. McGraw Hill, 1967.
- Instituto Mexicano de Investigación Bibliotecaria. **Libros de Referencia 1982**, Instituto Mexicano de Investigación Bibliotecaria, 1982.
- Osby, W. J., **Introducción a los Sistemas de Datos de Datos**, Addison - Wesley Interamericana, 1984.
- Dirección General de Bibliotecas, **Libros de Referencia 1983**, Dirección Gen. de Bibliotecas UNAM, 1983.
- Shannon, Stephen. **LIBROS a través de automatización de los Datos Bibliotecarios para las Bibliotecas de Referencia de las Bibliotecas de Referencia**, Instituto de Estudios de la UNAM, 1981.
- Tomkins, Peter. **Libros de Referencia en las Bibliotecas de Referencia en la América Latina**, Instituto de Estudios de la UNAM, 1981.
- Ward, Joseph. **Automatización de Bibliotecas para Bibliotecarios**, Universidad de las Américas Mexicanas de Bibliotecología de la UNAM en San Juan de los Rios, Veracruz, Veracruz, de México, 1983.
- Ward, Joseph. **Los Libros de Referencia en la América de la Bibliotecología Mexicana y de América Latina**, Instituto de Estudios de la UNAM, San Juan de los Rios, Veracruz, de México, 1983.

Guiroga, Luz Marina, Evaluación de Software para Sistemas Bibliográficos Microfilmados, Notas inéditas de la conferencia dictada por la Mtra. Luz Marina Guiroga el 27 de noviembre de 1987 en el Centro de Investigación Científica y Humanística.

Guiroga, Luz Marina, Documentación para el acceso de tóms con Minisig, basada en el Sistema de Biblioteca Nacional, Instituto de Investigaciones Bibliográficas, Biblioteca Nacional INAM, 1987.

Siipi, Charles J., Microcomputadoras y Vinculación de Términos, segunda edición, McGraw - Hill, 1985