



CEPAL

COMISION ECONOMICA PARA
AMERICA LATINA

CENTRO LATINOAMERICANO DE DOCUMENTACION
ECONOMICA Y SOCIAL

CLADES

RESTRINGIDA

E/CEPAL/CLADES/R.5
Julio de 1980

ORIGINAL: ESPAÑOL



FORMACION DE BASES DE DATOS
(Apuntes)

RESTRINGIDA

E/CEPAL/CLADES/R.5

Julio de 1980

ORIGINAL: ESPAÑOL

FORMACION DE BASES DE DATOS *

(Apuntes)

* Este documento fue preparado por el señor Abel Henríquez, funcionario de la Comisión Económica para América Latina, como un aporte al Curso-seminario sobre la Estructura y Funcionamiento del Sistema INFOPLAN, realizado entre el 14 de julio y el 1º de agosto de 1980

Santiago de Chile, 1980

S U M A R I O

	Pág.
FORMACION DE BASES DE DATOS	
Introducción	1
El Sistema ISIS	5
Establecimiento de una base de datos ISIS	10

FORMACION DE BASES DE DATOS

Objetivo: Presentar los aspectos técnicos y prácticos que deben considerarse al elegir manejar la información mediante computadora.

Introducción

La computadora

- es una máquina programable
- pueden diferir en tamaño y capacidad, pero siempre son capaces solo de:
 - sumar, restar, multiplicar, dividir
 - comparar
 - mover
 - leer, grabar o transmitir información
- realizan una cantidad limitada de operaciones elementales, pero a una velocidad mucho mayor que la humana y con mayor capacidad
- cada operación elemental es una instrucción de máquina. Cada computadora de una línea determinada tiene su propio juego de instrucciones de máquina
- una secuencia de instrucciones de máquina es un programa o rutina

Partes físicas de la computadora

Hardware

- CPU o UCP: controla y/o ejecuta todas las operaciones que se realizan en la máquina "cerebro"
- Memoria principal: contiene las instrucciones y datos de uno o varios programas "cerebro"
- Canales de E/S: transmiten la información hacia y desde la memoria a los diferentes dispositivos
- Dispositivos de E/S: entregan, reciben y/o almacenan la información. Ejemplo: cinta, disco, tarjetas, impresora, terminales, etc.

Ejecución de una instrucción

- CPU toma una instrucción de la memoria y la analiza leer una tarjeta
- CPU da la orden al canal
- el canal transmite la orden al dispositivo lectora de tarjetas está "ready"?
- el dispositivo inicia la lectura y transmite la información a la memoria, a través del canal se lee la tarjeta columna tras columna físicamente

/el dispositivo

- el dispositivo termina la operación y lo indica al canal
- el canal lo indica a la CPU
- la CPU verifica la operación y toma la siguiente instrucción

Partes lógicas de la computadora

Software

- Archivos: contienen la información
- Programas: contienen las instrucciones
- Sistema operativo:
 - controla el correcto funcionamiento de la computadora
 - realiza una serie de operaciones rutinarias: E/S, control de recursos
 - provee una serie de programas corrientes: SORT, compiladores
- Consola del operador: dispositivo para manejar el computador

Ejecución de una orden

- el operador escribe la orden en la consola y la envía al sistema (operativo)
- el sistema operativo analiza la orden e indica al programa respectivo que comience a trabajar
- el programa realiza su actividad y va requiriendo el acceso a los archivos (E o S) a través del sistema operativo
- el programa termina su función y lo indica al sistema operativo
- el sistema operativo toma nota y lo indica al operador a través de una respuesta en la consola

Tipos de procesamiento

- Procesamiento por grupos o lotes (batch)

- Etapas:
- recepción
 - registro
 - digitación
 - envío
 - entrada a la cola
 - proceso
 - salida a la cola
 - devolución
 - llegada y verificación

/Muchas etapas

Muchas etapas significan múltiples errores potenciales

Un solo error implica repetir una serie de etapas

La recuperación de un error es simple

Un solo usuario o una cantidad limitada

Sistemas pequeños o grandes

- Procesamiento interactivo (online)

Etapas: - recepción
- registro
- digitación
- respuesta y verificación
- (acumulación?)

Pocas etapas significan pocos errores potenciales

Un error simple implica corrección inmediata

Recuperación de un error no suele ser simple

Puede haber interacción con otros usuarios, implica proteger la información

Cantidad ilimitada de usuarios

Por costo: solo sistemas grandes

Manejo de información (enfoque bottom-up)

- Dato: es la información más elemental
se registra o almacena en un campo
usualmente es una cadena de caracteres
- Campo: unidad elemental de información
contiene un dato
tiene identificación o rótulo y
atributos: alfabético
numérico
largo
posición
- Registro: unidad para transferir información hacia o desde un
dispositivo
contiene una serie de campos
tiene una identificación y
estructura: con clave
sin clave
fijo
variable

- Archivo: conjunto de registros, todos de la misma estructura
tiene un nombre y organización: secuencial
directo
indexado
- Base de datos: conjunto integrado de archivos que contienen toda
la información pertinente a una aplicación
tiene un nombre y una estructura para el
acceso: relacional
red
arbol

Procesamiento de información (def. "batch")

- Programa: es una secuencia de declaraciones e instrucciones ordenadas lógicamente que señalan la manera como debe ser manejada la información de entrada para obtener la información de salida
- Proceso: constituye la ejecución de un programa, con un grupo de datos determinado, para obtener un resultado
- Listado: físicamente es una pila de hojas; es el registro que hace el programa de su actividad, imprimiendo con una presentación determinada la información requerida

EL SISTEMA ISIS

ISIS Integrated Set of Information Systems

Fue desarrollado por la OIT en Ginebra para el manejo computarizado de su biblioteca.

Lo tiene también la IDRC de Canadá con los mismos fines.

Está en CEPAL desde 1976, gracias a un convenio con la IDRC.

Tenemos la versión DOS del ISIS.

ISIS es un conjunto de programas, escritos en lenguaje Assembler IBM/360, operados en un sistema "batch"; manejan un conjunto de archivos similares a una base de datos

Implica: que depende de computadoras IBM, muy fuertemente por el lenguaje de máquinas; el sistema "on-line" es primitivo; el sistema "batch" es bueno; los archivos no se actualizan automáticamente

Características

- diseño modular: cada programa de ISIS cumple una tarea específica y rígida, y pueden combinarse para realizar una amplia gama de tareas diferentes
implica: flexibilidad
amplitud de aplicación
- No está orientado al usuario: cada nueva aplicación o modificación debe ser realizada mediante personal técnico capacitado
- Permite el uso de textos libres: los archivos pueden almacenar textos en lenguaje natural, y no es necesario tener definidos a priori las palabras claves. ISIS provee un mecanismo de separación y comparación para seleccionar palabras o frases significativas
- Tiene un lenguaje de impresión codificado: que simplifica bastante la tarea de especificar formatos para la presentación visual o impresa de la información
- Permite el uso de formularios especiales: se pueden emplear formularios pre-impresos y/o de tamaño especial para informes, listados especiales o etiquetas
- Permite establecer archivos de consulta (look-up files): que facilitan el ingreso de información mediante códigos, los que pueden ser reemplazados por una traducción al momento de imprimir
- Prevé el uso de terminales para el procesamiento interactivo: que permite la revisión y corrección inmediata, y la recuperación de información mediante consulta directa

/-No prevé

- No prevé la posibilidad de que la información sea compartida por distintos registros del mismo archivo: lo que implica repetir esta información en cada registro
- Amplia libertad para almacenar la información: en forma textual, utilizando descriptores, abreviada o codificada, numérica
- Amplia libertad para recuperar la información e indizarla: la información en un campo puede seleccionarse como un todo, por trozos fijos, traduciendo códigos, por frases, descriptores o palabras, traduciendo códigos, con la posibilidad de seleccionar por comparación antes de indizar

Aplicaciones del sistema ISIS

- Bibliográficas

Objetivos: mantener información acerca de documentos y publicaciones, producir catálogos periódicos para diseminación de la información, permitir la recuperación directa o diferida (búsquedas)

- CLADBIB

Recoge los documentos publicados por el sistema CEPAL.
Se produce el CLADINDEX en forma periódica.
Se hacen búsquedas seleccionadas a pedido de investigadores e instituciones. 1700 ítems

- TRANBIB

Recoge documentos y publicaciones en el área de Transporte y Facilitación.
Es usado sólo por la División 300 ítems

- CLAPLAN y ahora INFOPLAN

Recoge información bibliográfica sobre planificación.
Se produce el PLANINDEX en forma periódica.
Pueden hacerse búsquedas 400 ítems

- BIBLOS

Recoge todos los libros, documentos y publicaciones que recibe la Biblioteca CEPAL/ILPES.
Se publica un Boletín Mensual con la última información ingresada para distribución interna y externa.
Se preparan bibliografías seleccionadas a pedido de investigadores y otras instituciones 4100 ítems

- DOCPAL

Recoge documentación bibliográfica sobre población en América Latina.

Es manejado por CELADE.

Se publica una Revista en forma periódica

Otras aplicaciones

- ADDRESS

Contiene nombres y direcciones de personas y organismos con las cuales la CEPAL mantiene vinculaciones.

Objetivo: Preparar en forma automática etiquetas para distribución de documentos y publicaciones de la CEPAL o de sus divisiones, y mantener la información ya mencionada.

Se producen etiquetas para (casi) todas las publicaciones periódicas de la CEPAL y varias de las publicaciones de divisiones, como Revista, Cuaderno, Boletín FAL, Boletín de la Biblioteca, Micro-noticias, etc.

7200 items

- VOYCHAR

Contiene datos acerca de las condiciones de fletes marítimos y volúmenes de carga hacia y desde puertos de América Latina. Sirve de base para seleccionar información para usar en programas estadísticos

2000 items

- SIGLAS

Contiene siglas de organismos con sus respectivos significados y traducciones para producir catálogos alfabéticos

2000 items

- CLADIR

Recoge información acerca de unidades de documentación en América Latina.

Se publica el Directorio de unidades de información

Características técnicas

- Sistema ISIS/DOS (original) procesos "batch"

Hardware: computadoras IBM/360 /370
40K de memoria
lectora de tarjetas
impresora
unidad de cinta
unidades de disco

/Software

Software: sistema operativo DOS
compilador assembler

- Sistema ISIS/CMS/VSAM (modificado por CEPAL)

Hardware: IBM/370 145 o más ó /333x
512K de memoria real o más (+)
periféricos standard

Software: sistema operativo CMS
método de acceso VSAM (Access method service)
compiladores assembler y PL/1

Tipos de campos

- elementos del campo fijo

identificación: - letra
- posición, largo
alineación: - a la izquierda
- a la derecha
contenido: usualmente códigos o números

campos variables

identificación: - rótulo de dos dígitos
largo variable
simple o repetible o dividido
contenido: usualmente información alfabética
resúmenes
descriptores

Estos son los diferentes campos que puede contener el archivo principal y se describen en la TDC.

Archivos

- TDC Tabla de definición de campos org. secuencial
Contiene la descripción de los diferentes campos que puede tener el archivo principal, más un formato de presentación de la información para correcciones
- MASTER Archivo principal org. indexado
Contiene la información de cada una de las entradas del sistema.
Cada registro tiene el siguiente esquema:

Identificación	Información
00 1.5	Nº ISIS, por ej.: 00155
01	
02	
03	
...	
...	/XREF

- XREF Archivo de indicadores directo
Contiene la ubicación física de cada registro ISIS del archivo principal.
No se usa en VSAM.

- ELEMENT Archivo de elementos secuencial
Contiene la información de campos seleccionados del archivo principal más el número ISIS
Estructura:

1-5 N° ISIS
6-x información del campo primario
x+1-y " de campos secundarios
...
...
...

- LOOKUP Archivo de consultas Indexado
Contiene la información de reemplazo para cada uno de los códigos de un conjunto.
Estructura:

1-10 código
11-110 traducción

- ACCESS Archivo de accesos Indexado
Contiene las diferentes claves por las cuales se puede tener acceso a los registros.
Estructura:

1-35 clave
36- cantidad de registros anotados
 nexa al archivo INDEX

- INDEX Archivo de índices Directo
Contiene números ISIS
Estructura:
N° ISIS

- presentación inicial de la información.

El programa de ingreso presenta la información en un formato determinado en la TDC.

En CLADES se usa presentar la información en el mismo orden que la hoja, indicando el campo a que pertenece cada trozo de información.

III Entrada de los datos

- Preparación de los datos en la hoja
 - en tarjetas o similar para sistema "batch"
 - entrada "on-line"
 - listados para revisión
- Preparación de los datos en la hoja: la documentalista anota los datos en la hoja de acuerdo a las normas estipuladas de antemano en un manual de procedimiento. Es recomendable la revisión por otro documentalista.
- Ingreso a la computadora:
 - batch: digitación en tarjeta por la perforista
envío de tarjeta a la máquina
 - on-line: digitación en el terminal
la información se digita tal cual aparece en la hoja
- listados para revisión

En forma periódica se preparan listados para revisión de la última información ingresada.
Contraste de un término con otro.
Se detectan errores de digitación
" " " " procedimiento

IV Corrección de los datos

- preparación de correcciones
- correcciones en tarjeta
- " on-line
- Preparación de correcciones on-line

Es simple, se va accediendo a cada registro y se corrige directamente o mediante funciones (órdenes).
- Preparación de correcciones batch
- Funciones
 - agregar, borrar, reemplazar un campo variable o elemento de campo fijo

/- cambiar parte

- cambiar parte de la información de un campo variable
- extender un campo variable
- cambiar el rótulo de un campo variable
- Procedimiento
 - identificar el registro
 - identificar la función
 - identificar el campo (ojo con repetibles)
 - anotar la información adicional info1 # info2 #
- Revisión del listado de correcciones

V Producción de información

- preparación de listados
- " del archivo de elementos
- producción de listados
- Diseño de listados.

Se decide con el documentalista el formato de presentación de la información: ordenamiento, encabezamiento, distribución de los diferentes campos en la presentación: ubicación, sangría, información fija y variable.

- Preparación del archivo de elementos
 - generación de campos primarios y secundarios, principales y alternativos
 - procedimiento de selección: textual, random, rango
 - procedimiento de generación: traducción de códigos, selección de trozos, selección de palabras y/o descriptores, colocación de prefijos e indicadores
 - con el fin de usarlos como encabezamiento y como campos de ordenamiento
Ordenamiento = selección de elementos: uso de stop-words y su ubicación en el registro de elementos
- Preparación de listados
 - especificación de formatos
formato corriente y alternativo
presentación: ubicación, sangría, prefijos y sufijos
saltar espacios y líneas: opcional y obligatorio
 - especificación de encabezamientos
ubicación y largo
presentación: con/sin subrayar, etc.
primer encabezamiento y continuación

/-especificaciones

- especificaciones del listado
ancho de la línea, margen
tamaño de la página
numeración de páginas
título inicial
cambio de páginas
índice de páginas y encabezamiento
- especificaciones para índices
ubicación de los números ISIS: derecha y debajo
números por línea
margen

VI Recuperación de información

- Preparación de un archivo (ACCESS/INDEX) invertido
- recuperación con archivo (" ") "
- Preparación del archivo invertido
Se basa en un archivo de elementos
Elección del campo a invertir
Selección de palabras y/o descriptores, colocación de indicadores
Ordenamiento y selección
Creación de los archivos ACCESS e INDEX
 - controlada: basado en uno existente
 - no controlada: se crean ambos a la vezesta es la estructura de recuperación
- Recuperación
Especificación de la búsqueda

operación AND	+	restringe la búsqueda
" OR	*	amplía " "
" Excluir	-	restringe

Restricción textual selección textual
Archivo ANY: nombre común para un grupo de descriptores
presentación del resultado: mediante un formato
rapidez: la búsqueda se realiza fuera del maestro

VII Evaluación del sistema

- Importancia del diseño preliminar y de la flexibilidad del sistema
Si el diseño preliminar era bueno, el resultado es ahora exitoso
pero: todo cambia: los requerimientos y las posibilidades del
sistema
La experiencia adquirida permite una revisión y una afinación de la
estrategia llevada a cabo.

Tal vez vemos nuevas posibilidades, que con ligeras modificaciones se pueden llevar a cabo.

Tal vez hay nuevos requerimientos, que se pueden realizar.

Todo cambio requiere flexibilidad e ISIS la ofrece ampliamente; podemos:

- agregar nuevos campos
- modificar los ya existentes
- eliminar campos inútiles

sin que el sistema se resienta, o agregar nuevos procedimientos, en particular sobre estos archivos de elementos y obtener otros resultados.

