



ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES IZTACALA

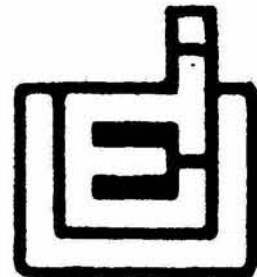
H. B. C.

Una base de datos Limnológica.

ENRIQUE GANEM CORVERA

BIOLOGIA

**E. N. E. P. IZTACALA
U. N. A. M.**



1985



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.

CONTENIDO

AGRADECIMIENTOS

INDICE DE FIGURAS

RESUMEN

I. INTRODUCCION

- I.1) Antecedentes generales.
- I.2) Antecedentes del estudio.
- I.3) Propósitos del estudio.
- I.4) Definición de términos.

REVISION DE LITERATURA

- II.1) Introducción.
- II.2) Características y estructura de una base de datos.
- II.3) Las bases de datos en Biología.
- II.4) Resumen.

III. DISEÑO DEL PAQUETE H.B.C.

- III.1) Introducción.
- III.2) Los criterios de selección.
- III.3) Selección de técnicas.
- III.4) Diagramas de flujo.
- III.5) Descripción del paquete H.B.C..
- III.6) Resumen.

IV. DISEÑO EXPERIMENTAL Y RESULTADOS

- IV.1) Introducción.
- IV.2) Puntos por verificar.
- IV.3) Ejemplo del funcionamiento de H.B.C..
- IV.4) Resultados.
- IV.5) Resumen.

V. DISCUSION

- V.1) Introducción.
- V.2) Discusión.
- V.3) Resumen.

VI. CONCLUSIONES Y RECOMENDACIONES

- VI.1) Introducción.
- VI.2) Conclusiones sobre el paquete H.B.C..
- VI.3) Recomendaciones para el usuario.
- VI.4) Recomendaciones para la adaptación de H.B.C. a otras microcomputadoras.
- VI.5) Mejoras propuestas para versiones posteriores de H.B.C.
- VI.6) Proyección al futuro de H.B.C..
- VI.7) Proyección de la computación en la Biología Moderna.
- VI.8) Resumen.

VII. ANEXOS

- VII.1) Formas codificadas.

VIII. BIBLIOGRAFIA

AGRADECIMIENTOS

Las siguientes personas contribuyeron de diversas formas en la realización del trabajo.

El M. en C. Manuel Guzmán Arroyo merece un reconocimiento especial por la dirección de la tesis. Su constante apoyo fue indispensable para el desarrollo del trabajo.

El Dr. James Curtis G. merece un reconocimiento especial por su contribución en la investigación bibliográfica y en la estructura final del presente reporte. Sus recomendaciones y apoyo son invaluable.

Las siguientes personas merecen reconocimiento por su colaboración en el desarrollo del presente trabajo.

Act. José Antonio Villar Leal

Ing. Gilberto Fernández Vincent

Ing. James P. Kelleghan Monges

La revisión del trabajo estuvo a cargo de las siguientes personas.

Dr. James B. Curtis García

Biól. Enrique Kato Miranda

Biól. Manuel Elías Gutiérrez

Biól. Jonathan Franco López

El Biól. Jesús Ponce sugirió algunos de los artículos citados.

Los dibujos presentados en el anteproyecto fueron elaborados por el Ing. Pesq. Henry Ruiz Santos. Los diagramas de flujo en la versión final fueron elaborados por el Biól. Juan Chávez Rojas.

Otras personas que de alguna manera contribuyeron con la realización del trabajo son:

Biól. Sonia Galavíz Solís

Biól. Ma. de los Angeles Aranda Macías

Y los miembros del laboratorio de Limnología.

Mi familia merece un reconocimiento especial por un apoyo que comenzó mucho antes de la realización del presente trabajo.

INDICE DE FIGURAS

| | pp. |
|--|-----|
| 1... Tabla de figuras utilizadas en diagramas de flujo. | 18 |
| 2... Diagrama de la estructura general del paquete H.B.C. | 18a |
| 3... Diagrama de la estructura general de un programa captador de datos. | 18b |
| 4... Diagrama de la estructura general de un programa selector de datos. | 18c |

RESUMEN

El laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología de la U.N.A.M. ha capturado grandes cantidades de datos de calidad desde 1977. Su volumen es tan alto (únicamente del langostino M. tenellum se tienen aproximadamente 500,000) que el proceso manual resulta casi imposible. Además los datos así almacenados son vulnerables a pérdidas o alteraciones.

Para resolver este problema fue elaborada una base de datos relacional especializada que incluye rutinas de búsqueda con los criterios de selección empleados por los investigadores del laboratorio; estas rutinas permiten localizar datos por profundidad, por localidad y otros criterios. El formato de captura de datos se basa en las formas codificadas empleadas en el laboratorio desde la década pasada aunque con pequeñas modificaciones necesarias para ampliar el rango del tipo de datos capturables y facilitar así la adaptación del presente sistema a cualquier cambio en la orientación del trabajo futuro del laboratorio.

El trabajo fue desarrollado en BASIC (Applesoft) en las microcomputadoras Apple IIplus y Franklin ACE 1000 (48 y 64 K respectivamente). Con algunas modificaciones, este paquete podría adaptarse a las necesidades de casi cualquier laboratorio de Biología.

Prácticamente no existen antecedentes del trabajo en cuestión pues, si bien las computadoras han sido utilizadas por los biólogos desde hace más de 25 años, la mayoría de los trabajos relacionados con el aquí expuesto han sido elaborados en máquinas grandes. Las limitaciones de éstas son muy diferentes a las de las microcomputadoras y por ello los programas desarrollados en las primeras no son traducibles al sistema usado en el presente trabajo.

Los listados del paquete (en BASIC Applesoft) pueden ser consultados en el laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología de la U.N.A.M. a cargo del M. en C. Manuel Guzmán Arroyo.

El formato del presente trabajo se basa en las recomendaciones del manual de la A.P.A..

CAPITULO UNO

INTRODUCCION

I.1) Antecedentes Generales.

Una de las enseñanzas más importantes para un estudiante de una carrera científica, como la Biología, es la del concepto de ciencia. Resulta imposible que un estudiante aprenda a profundidad temas tan complejos como la Zoología, la Ecología y la Bioquímica en algunos semestres, sin embargo el estudio primario de estas disciplinas, con el consiguiente contacto con el método científico, genera un criterio sobre la adquisición de información que permitirá al estudiante asimilar nuevos conocimientos en los campos de su interés.

El concepto de ciencia ha sido discutido por filósofos y científicos de todas la épocas, y las conclusiones (Einstein e Infeld, 1938; Poincaré, 1964; Poincaré, 1981) se podrían resumir de la siguiente forma: El investigador interesado en un cierto fenómeno genera un modelo que intenta explicar los mecanismos de su funcionamiento, y a partir de este modelo genera predicciones sobre el comportamiento del fenómeno bajo ciertas circunstancias. Posteriormente, el investigador observa el fenómeno en cuestión bajo las circunstancias mencionadas anteriormente, ya sea en el lugar donde ocurre o (de ser posible) en un medio cuyo ambiente pueda ser controlado; a la primera observación se le conoce como MUESTREO y a la segunda como EXPERIMENTO. El propósito de un muestreo o experimento es el de generar datos objetivos que reflejen el comportamiento del fenómeno estudiado para así poder comparar el modelo del investigador con la realidad (o para generar nuevos modelos). Posteriormente, si los datos muestrales o experimentales concuerdan con el modelo se dice que éste es consistente y es sometido a nuevas pruebas. En el caso contrario es necesario modificar el modelo, o desecharlo por completo.

De acuerdo a lo anterior, todo estudiante de una carrera científica debe aprender a obtener datos confiables y objetivos cuyo valor no dependa de otro fenómeno que no sea el estudiado; además deberá aprender a manejar esos datos de manera que su integridad sea respetada. La alteración de los datos experimentales o muestrales puede desvirtuar por completo la actividad del investigador, pues las conclusiones obtenidas por la comparación de las predicciones con los datos alterados serán incorrectas en grado impredecible en la mayoría de los casos.

Los datos pueden perder su objetividad en cualquiera de los pasos necesarios para su proceso: las técnicas empleadas para su obtención pueden ser incorrectas, el manejo y distribución de esos datos puede ser equivocado y las conclusiones obtenidas del resultado del proceso pueden ser erróneas. La dificultad para conservar la integridad de los datos muestrales o experimentales depende, entre otros factores, de su cantidad, del grado de dificultad de las técnicas empleadas para obtenerlos y del proceso utilizado para analizarlos.

La Biología es quizá la más interdisciplinaria de las ciencias, ya que requiere del auxilio de prácticamente todos los otros campos del conocimiento para comprender a fondo los fenómenos por ella estudiados. Esta situación no ocurre con la Química o la Física, pues un físico no necesita aprender Biología para explicar los detalles más finos del objeto principal de su estudio; para un profesional de estas ciencias los seres vivos sólo son casos especiales del comportamiento de la materia. Por otro lado, el biólogo interesado en explicar a fondo el comportamiento molecular del ADN debe aprender Química Molecular y debe tener cuando menos algunas nociones de Mecánica Cuántica. Un ejemplo de la aplicación de nociones avanzadas de Física en Biología es el empleo de las leyes de la Termodinámica para explicar los mecanismos del metabolismo y de la evolución (Schrödinger, 1962). Un ejemplo de la aplicación de las matemáticas en la Biología Moderna aparece en Lotka (1956).

Esta naturaleza interdisciplinaria ha tenido consecuencias muy variadas, una de ellas es que tanto el tipo como la cantidad de datos de interés para el biólogo ha crecido rápidamente en los últimos años. Durante mucho

tiempo, la Biología fue una ciencia puramente descriptiva, y bastaba con el lenguaje común para describir el trabajo de frontera realizado por los investigadores de aquellos días. En la actualidad el biólogo debe aprender nuevos lenguajes distintos a los de uso diario (Curts y Saal, 1979; Waddington, 1976) para generar modelos consistentes con el amplio cuerpo de datos experimentales y muestrales capturados en las últimas décadas.

Uno de los lenguajes más útiles para el biólogo moderno es el de la Cibernetica, la ciencia de los sistemas. La Cibernetica nació como ciencia en las primeras décadas del siglo XX y durante mucho tiempo se nutrió con ejemplos biológicos; biólogos como Ludwig von Bertalanffy generaron muchos de los conceptos cibernéticos básicos (Ackoff, 1974; Navarro, 1982). Posteriormente, con la aparición de la primera computadora electrónica en la década de 1940 los físicos, matemáticos e ingenieros comenzaron a contribuir enormemente con la Cibernetica y es por esto que en la actualidad muchas personas no relacionadas directamente con la Biología Moderna consideran que la Computación (o Cibernetica aplicada) es un campo de estudio exclusivo para ingenieros, matemáticos y otros profesionales. Como ejemplo del uso de los conceptos de la Cibernetica en la Biología Moderna se puede citar a Bojorquez (1975). Una discusión mas amplia de la relación entre la Biología Moderna y la Cibernetica se encuentra en George (1968).

Debido al gran aumento en la cantidad y tipo de datos de interés biológico descrito anteriormente, el manejo manual de ellos resulta cada vez mas difícil y peligroso, pues el investigador puede cometer errores durante la selección de aquellos datos de interés para su trabajo. Otro problema grave para el biólogo moderno que administra sus datos únicamente con la ayuda del lápiz y el papel es que existen ciertos procesos que, por su complejidad, resultan absolutamente irrealizables sin la ayuda de una computadora, como el análisis multivariado de todos los parámetros conocidos de un ecosistema. Para resolver estos problemas, los biólogos han empleado el lenguaje de la Cibernetica desde hace más de 25 años. Algunos investigadores de las Ciencias Biomédicas han generado programas como el Bio-Medical Data Processor (Dixon, Brown, Engelman, Frane, Hill, Jennrich y Toporek, 1985); la primera versión de B.M.D.P. apareció en 1960. Muchas universidades, como la de York en Inglaterra, ofrecen en la actualidad maestrías y doctorados en Computación exclusivamente a investigadores en Ciencias Biomédicas.

La incorporación de la Computación en la Biología Moderna no se realizó sin problemas, pues durante mucho tiempo los investigadores consideraron que bastaba con solicitar la ayuda de un especialista en Computación para implementar los programas necesarios para resolver un problema dado y así evadían la responsabilidad de aprender una serie de conceptos y técnicas que a primera vista parece extremadamente compleja. Estas consideraciones no tienen un fundamento sólido, pues para generar un programa es necesario conocer a fondo todas las posibles ramificaciones del problema a resolver y los problemas relacionados con la investigación biológica son más complejos que los problemas administrativos y contables a los que normalmente enfrentan los técnicos en programación. Además existen varios peligros inherentes a esta actitud, pues resulta arriesgado emplear una técnica cuyo fundamento es desconocido ya que, bajo estas circunstancias, el investigador ignora las condiciones en las que la técnica es válida. Un ejemplo de lo anterior es el caso de ANOVA (Análisis de Varianza) que ha sido ampliamente utilizado por muchos biólogos para decidir si existe alguna relación entre las variables de un sistema. Los resultados obtenidos pueden ser utilizados para generar conclusiones sobre la validez de un modelo. En un artículo relativamente reciente, Curts (1984) señala que, debido a que muchos investigadores ignoran las condiciones bajo las cuales ANOVA genera resultados confiables, muchos trabajos biológicos de alto nivel podrían tener conclusiones insuficientemente fundamentadas, e inclusive erróneas. Debido a lo anterior, la comunidad biológica internacional ha adoptado a la Computación como una rama válida en la investigación biológica, como lo hizo anteriormente con la Química y las Matemáticas. Un ejemplo de esta nueva actitud son las siguientes palabras de J.D. Spain, tomadas de su libro "BASIC Microcomputer Models in Biology" (1984):

"...En un futuro cercano, es probable que los investigadores de las Ciencias Biomédicas sean caracterizados más por su capacidad para emplear una computadora en el análisis de datos y en la simulación que por su capacidad para emplear correctamente un microscopio. La Biología es, potencialmente, la ciencia mas matemática de todas..."

Gracias a este nuevo punto de vista, el biólogo moderno no solo emplea técnicas elaboradas por otros

profesionales, sino que desarrolla las suyas propias desde hace varios años. En la introducción de la tesis doctoral de J. Benson (1977), que versa sobre el empleo de modelos de simulación para la enseñanza en Biología se discute ampliamente el papel de la computadora en la Biología Moderna. Otras fuentes son Curtis (1985a), Mc Millen y Esch (1984), Spain (1981) y Spain (1983).

Desde la década de 1960, algunos biólogos han publicado trabajos de investigación en donde se proponen nuevas técnicas de computación que permiten la captura y la selección de grandes cantidades de datos biológicos. Los programas que emplean estas técnicas reciben el nombre de BASES DE DATOS (Kruglinski, 1983) y son usados en muchos centros de investigación, empresas y departamentos gubernamentales de casi todo el mundo.

Durante más de 30 años los investigadores de cualquier ciencia han tenido que resolver muchos problemas antes de ganar acceso a una computadora, pues (hasta la década pasada) el costo de la máquina más económica era muy superior al de un microscopio electrónico, el costo de mantenimiento era elevado y la operación y programación eran muy complejas. En la década de 1970 aparecieron las microcomputadoras, que aún en aquella época ofrecían mayor capacidad de proceso que la de las grandes máquinas utilizadas en los principios del programa espacial. Las características que hacen ideales a las microcomputadoras como herramientas del investigador moderno son:

- a) La facilidad de acceso. Las microcomputadoras son relativamente baratas y por lo tanto es posible contar con una de ellas en cada laboratorio; el investigador la tendrá a su alcance en cualquier momento.
- b) El bajo costo de operación y mantenimiento. Una microcomputadora típica (la APPLE IIe) consume 50 wats y el tiempo promedio entre dos descomposturas es de unas 20,000 horas (unos 2 y medio años de operación ininterrumpida).
- c) La eficiencia en el almacenamiento y el procesamiento de datos. La microcomputadora es capaz de almacenar grandes cantidades de datos en sistemas baratos y confiables (los discos magnéticos de 5.25 pulgadas conocidos como diskettes). Además resulta sencillo y rápido (unos 10 seg. aprox.) copiar un archivo cualquiera y por lo tanto disminuye mucho la posibilidad de perderlo accidentalmente. Otra ventaja es que, una vez programadas adecuadamente, prácticamente nunca cometen errores (cuando se detecta alguno, generalmente se debe a una programación deficiente o a una falla mayor en algún circuito).
- d) Las posibilidades de expansión. Existe la posibilidad de implementar lenguajes avanzados (antes accesibles solo en máquinas grandes), aunque se requiere de un gasto regular ya que es necesario ampliar la memoria disponible; además es necesario conseguir el programa correspondiente. Existe la posibilidad de agregar sistemas periféricos, como graficadoras, impresoras y digitalizadores de imagen.
- e) La capacidad de proceso. Las microcomputadoras pueden realizar procesos muy complejos, como el cálculo de la posición de un planeta en la bóveda celeste o el cambio en la estructura de una población ideal al ser alterado un parámetro cualquiera (como el número de productores primarios).
- f) La facilidad de uso. No se requiere de una formación especial para operar una máquina adecuadamente programada.

Si bien las microcomputadoras presentan algunas desventajas, como la limitación en la memoria y en la velocidad de proceso, los puntos mencionados anteriormente las superan ampliamente. Otra desventaja de las microcomputadoras es que no siempre son directamente compatibles, aunque este problema es parcialmente soluble si se toma en cuenta que prácticamente todas las microcomputadoras tienen integrado algún dialecto de BASIC; de esta manera se facilita el transporte de un sistema de una máquina a otra. Existen además módulos, llamados emuladores, que permiten el acceso a programas generados en máquinas no compatibles con la del usuario, sin embargo estos equipos son caros.

En resumen, el aumento del conocimiento biológico ha puesto en contacto íntimo a la Biología con otras ciencias y la consecuencia es que la cantidad (y el tipo) de datos de interés biológico ha crecido hasta alcanzar proporciones ingobernables. En fechas recientes, algunos investigadores de las Ciencias Biomédicas han modificado (o creado) técnicas cibernéticas para resolver los problemas inherentes al manejo de enormes cantidades de datos.

I.2) Antecedentes del estudio.

El personal del laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología ha capturado grandes cantidades de datos de interés biológico que han generado más de 20 proyectos de investigación y más de 15 tesis de distintos niveles, desde licenciatura hasta doctorado. Estas tesis tratan temas como la fecundidad del langostino *Macrobrachium tenellum* (Negrete, 1977) o el estudio ictiológico de la laguna de Cuyutlán (Nuñez, 1982).

Los datos comenzaron a ser capturados en 1975. Los muestreos se han realizado en distintos cuerpos de agua, principalmente en los estados de Guerrero y Michoacán.

El volumen de los datos es enorme. Se calcula que en el caso de una especie de langostino (*M. tenellum*) se conservan unos 500,000 datos biométricos.

Los datos almacenados en papel (ver sección VII) son difíciles de separar para su proceso y además son vulnerables a pérdidas (los datos son irrecuperables) y a transcripciones erróneas debido al cansancio durante la selección de datos para un trabajo o a la letra del capturista, que en ocasiones es confusa (o totalmente ilegible). Este deterioro (y eventual pérdida) de los datos puede afectar la calidad de los trabajos presentados. El proceso de verificación manual de los datos empleados en un proyecto puede demorar meses.

I.3) Propósitos del estudio.

El objetivo del presente trabajo fue el de elaborar una base de datos que permitiera:

- a) Almacenar en discos magnéticos (diskettes) los datos capturados en las formas codificadas "HIDROLOGIA A1" (físicoquímicos y climáticos), "HIDROLOGIA A2" (nutrientes), "BIOMETRIA B3" (dimensiones y estadio sexual) y "CAPTURA C1" (pesquería) del laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología de la U.N.A.M..
- b) Seleccionar los datos pertinentes para cualquier trabajo de investigación de acuerdo a los criterios de selección empleados por los investigadores del laboratorio.
- c) La corrección de errores en cualquier momento.
- d) La interacción con usuarios no entrenados en el uso de microcomputadoras.

El nombre del sistema es H.B.C. pues almacena y selecciona datos Hidrológicos, Biométricos y de Capturas realizadas durante los muestreos.

El paquete fue diseñado para las microcomputadoras APPLE IIplus, FRANKLIN ACE 1000 (48 y 64 K respectivamente) y máquinas compatibles. El lenguaje usado fue el BASIC (Applesoft). Con algunas modificaciones, el paquete puede ser empleado en otras microcomputadoras.

I.4) Definición de terminos.

El objetivo de esta sección es el de servir como referencia para el resto del trabajo. La mayoría de las definiciones fueron obtenidas por consulta personal con profesionales de la Biología y la Computación. El resto fueron tomadas de muchas de las referencias citadas en la sección VIII.

-ALFANUMERICA (variable) Todas aquellas variables que no tienen un valor numérico asignado. Las variables alfanuméricas tienen el subfijo \$ (por ejemplo ARCHIVO\$). Estas variables pueden representar letras o numerales, aunque estos últimos son interpretados como caracteres y no como números.

-ALGORITMO Es el procedimiento detallado necesario para resolver un problema.

-BASIC Acrónimo de Beginners All Purpose Symbolic Instruction Code. Es un lenguaje superior de aplicación general.

-BIT Es la unidad de información más pequeña posible. El concepto de bit puede aclararse con el ejemplo del alfabeto MORSE. En MORSE cada caracter es codificado con una cadena de 3 bits que pueden tomar dos valores, el punto y la raya; no existe una unidad menor de información.

-BYTE Es una cadena de 8 bits que codifica a un caracter. En el ejemplo anterior el "byte" del alfabeto MORSE tiene 3 bits.

-CHARACTER Cualquier conjunto de bits (8 en el caso de las microcomputadoras) que se puede representar en la pantalla de la computadora o que genera instrucciones especiales. Puede ser un número, una letra, un espacio vacío o un caracter generado por la pulsación de dos teclas simultaneamente (por ejemplo CTRL-A, generado por la pulsación de las teclas Control y A).

-CHIP Microcircuito elaborado con un copo ("chip" en inglés) de silicato puro marcado con impresas conductoras que representan al circuito. El nombre se hace extensivo al paquete protector, de color negro.

-COBOL Acrónimo de Common Bussines Oriented Lenguaje. Es un lenguaje superior de uso casi exclusivamente administrativo.

-COMPILADOR Programa que convierte una cadena de instrucciones de algún lenguaje (como PASCAL) en una serie de instrucciones directamente ejecutables en lenguaje de máquina (ver LENGUAJE DE MAQUINA).

-D.O.S. Acrónimo de Disk Operating System. Es un conjunto de programas que permiten cualquier operación que involucre a una lectora de discos magnéticos, como el grabado y la recuperación de datos (o de un programa); algunos D.O.S. incluyen un sistema operativo. El D.O.S. no debe confundirse con el O.S. (ver SISTEMA OPERATIVO).

-FLOP Acrónimo de FLloating Point operation, es una unidad de velocidad de proceso. El flop equivale a una operación fundamental (suma, resta, multiplicación o división) con números de doble precisión (14 cifras significativas) por segundo.

-FORTRAN Acrónimo de FORMula TRANslation. Es un lenguaje superior muy utilizado por los científicos pues facilita el proceso de codificación de una fórmula para su posterior ejecución por la computadora. En fechas recientes ha sido reemplazado en algunos campos por el PASCAL.

-GENERACION DE MAQUINAS Las generaciones de computadoras se distinguen tanto por el tipo de tecnología empleada en su construcción como por la capacidad de proceso de la misma. La primera generación de computadoras apareció en la década de 1940. A esta generación pertenecen máquinas como el ENIAC que tenía 18,000 válvulas de vacío ("bulbos"), consumía varios kilowatts y sufría una avería cada semana (o cada vez que algún cálculo sobrepasaba la capacidad de la máquina). La segunda generación fue elaborada con transistores, consumía de 10 a

100 veces menos energía y funcionaba a varios centenares de kiloflops (la generación anterior funcionaba a unos pocos kiloflops); las máquinas de esta generación sufrían una falla cada mes de operación continua aproximadamente. La tercera generación utilizó por primera vez a los "chips", cada chip contenía una docena de transistores aproximadamente; Las máquinas de esta generación funcionaban a varios megaflops y sufrían una falla cada 3 o 5 meses. La generación actual utiliza chips con decenas de miles de transistores, algunas de ellas pueden funcionar a varios megaflops (CRAY-X1 por ejemplo) y sufren una avería cada año de operación continua, aproximadamente; Las microcomputadoras se cuentan en esta generación aunque sus características funcionales son similares a las de las máquinas de la generación anterior. La quinta generación de computadoras aparecerá en 1989, y según sus constructores, estará compuesta por máquinas con una inteligencia en algunos casos indistinguible de la humana (Lenat, 1984).

-GENERACION DE LENGUAJES Los lenguajes son códigos de instrucciones específicas que le indican a la máquina los pasos que debe ejecutar. Las generaciones de lenguajes se distinguen tanto por la facilidad con la que el programador puede resolver sus problemas como por la potencialidad que le otorga a la máquina. Los lenguajes más simples operan al nivel de las instrucciones binarias del microprocesador y por ello, a un programa así elaborado se le conoce como "programa en lenguaje de máquina". Este lenguaje puede resultar, por su simplicidad, muy difícil de utilizar, pues existen pocas instrucciones de programa, además la computadora funciona con números en base hexadecimal (base 16) y todas las operaciones fundamentales deben ser realizadas en este sistema numérico. Los lenguajes superiores permiten la codificación del algoritmo del programa en un lenguaje similar al humano aunque sin las ambigüedades inherentes a los lenguajes humanos. Los lenguajes de Ultra Alta Generación (UHLL o Ultra High Level Lenguajes) permiten que la máquina altere el programa según un patrón definido por el programador; estos lenguajes serán empleados en las máquinas de quinta generación.

-K (KILOBYTE o KILOPALABRA) Unidad utilizada para medir la capacidad de memoria de algún componente de un sistema de computación (computadora, disco magnético, terminal, etc.). Equivale a 1024 caracteres (bytes). Es importante señalar que las máquinas grandes funcionan con "bytes" de más de 8 bits (conocidos como PALABRAS); así una memoria de 48 K en un máquina grande en realidad tiene 48 Kilopalabras de 16, 32 o hasta 48 bits cada una. Una microcomputadora típica tiene de 48 a 64 K de memoria base.

-LENGUAJE DE MAQUINA Es una secuencia de instrucciones codificada en instrucciones directamente ejecutables por el microprocesador.

-LENGUAJE ESTRUCTURADO Es un lenguaje superior que permite la descomposición del algoritmo en bloques de fácil programación.

-LISP Acrónimo de LIST Processing language. Es un lenguaje de ultra alto nivel (UHLL) que le permite a la máquina alterar el programa según "aprende" como debe modificarlo para mejorar su rendimiento. Es uno de los lenguajes propuestos para las máquinas de quinta generación. El otro lenguaje propuesto es el PROLOG.

-MAQUINA MULTIUSUARIO Las computadoras grandes son capaces de ejecutar varias decenas de programas al mismo tiempo ya que disponen de grandes cantidades de memoria, microprocesadores rápidos y varias terminales. Por su arquitectura, estas máquinas funcionan de forma muy diferente a la de las microcomputadoras.

-MATRIZ Conjunto de datos ordenados en dos dimensiones. Una tabla con más de una variable es un buen ejemplo. A cada columna o fila de la matriz se le conoce como VECTOR; este término también se aplica al valor de cada columna o fila necesarios para localizar un dato. Las matrices unidimensionales se llaman VECTORES (por estar compuestas por un solo vector) y las de 3 o más se conocen como ARREGLOS.

-MICROPROCESADOR Es un circuito integrado o chip que lleva a cabo todas las funciones lógicas de la computadora. Un microprocesador moderno puede tener mas de 50,000 transistores aproximadamente en un área inferior a 1 centímetro cuadrado. Los microprocesadores pueden tener 8, 16, 32 o más líneas de dirección que permiten el acceso a las localidades de la memoria; mientras más líneas de dirección tiene el circuito, más capacidad de memoria posee la máquina. Las microcomputadoras más comunes son de 8 líneas (8 bits, ya que cada línea puede enviar un bit cada vez) y mandan 2 señales de direccionamiento (16 bits en total), así que pueden acceder hasta a la 16 (más de 65,000) localidades de memoria. Los microprocesadores funcionan con una serie de instrucciones binarias (lenguaje de máquina) y cada instrucción puede ejecutarse en unos pocos ciclos de

operación del circuito. Un procesador común (6502) opera a razón de 1,000,000 de ciclos por segundo.

-PASCAL Lenguaje de alto nivel muy utilizado por científicos. Su nombre recuerda al famoso pensador Blaise Pascal. Este es un lenguaje estructurado.

-PERIFERICOS Cualquier dispositivo conectado directamente a la computadora. Los periféricos pueden recibir datos de la computadora (periféricos de salida como el monitor o la impresora), pueden mandarlos a la máquina (periféricos de entrada, como un digitalizador de imágenes) o pueden ser compuestos (periféricos de entrada y salida, como la lectora de discos magnéticos).

-RAM Acrónimo de Random Access Memory. En este sistema de memoria el usuario puede almacenar, recobrar y borrar información.

-ROM Acrónimo de Read Only Memory. En esta memoria se encuentra la información indispensable para que la máquina funcione adecuadamente (por ejemplo, para que entienda algún lenguaje). El conjunto de esta memoria, pregrabada en la fábrica, solo es accesible para la lectura pero no para la escritura; es decir, no se puede alterar.

-SISTEMA OPERATIVO Es un conjunto de programas (generalmente almacenados en ROM) que permite que la máquina se comunique con el mundo exterior; el operativo puede incluir algún lenguaje superior (generalmente BASIC).

-SUBROUTINA Es una sección de un programa que se repite. Para evitar el desperdicio de memoria (al repetir la misma serie de instrucciones en otra parte del proceso) se define a este conjunto de instrucciones como subrutina, que puede ser llamada con una sola instrucción cuando sea necesario. Una vez ejecutada la subrutina, el procesador ejecutará la siguiente instrucción a la llamada correspondiente.

-TERMINAL Es un periférico para computadoras multiusuario que permite tanto la entrada de datos o programas como la salida de los resultados de un proceso. Generalmente consta de una pantalla y un teclado, aunque a veces cuenta con una lectora de discos magnéticos y una impresora.

-VERSION Es el equivalente a una edición. Si un programa tiene un número de versión de 4 significa que el programa ha sufrido 3 grandes modificaciones anteriores. La versión 1.0 se asigna al primer programa que cumple con los requisitos mínimos señalados por el programador. Si la modificación es pequeña, el número entero se conserva y se aumenta un decimal (versión 1.1, por ejemplo).

I.7) Resumen.

En el presente capítulo se presenta el problema del laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología, problema que resulta común a muchos laboratorios de investigación; el del proceso de grandes cantidades de datos acumulados durante varios años de trabajo. El volumen de estos datos puede ser tan grande que su manejo manual puede resultar inoperante. Para solucionar el problema se propone modificar algunas de las técnicas desarrolladas para el manejo de grandes cantidades de datos en computadoras y se propone para ello el empleo de una microcomputadora debido a su bajo costo y facilidad de operación; gracias a esto es posible adaptarla con mucha facilidad al ambiente de trabajo de un laboratorio de investigación. En este capítulo se presentan además las justificaciones del trabajo que se resumen en tres puntos; los biólogos deben aprender Computación porque a) ya existen antecedentes de trabajos de investigación en Computación presentados por biólogos, b) Debido a la gran complejidad de los sistemas estudiados por la Biología, solo con ayuda de una computadora es posible extraer información del gran caudal de datos capturados en los últimos años y c) si el investigador desea emplear el gran potencial de estas máquinas al máximo (como ha ocurrido con el microscopio), debiera conocer sus características a fondo.

En el presente capítulo se incluye además una serie de términos técnicos empleados a lo largo de este reporte.

CAPITULO DOS

REVISION DE LITERATURA

II.1) Introducción.

El objetivo del presente capítulo es el de presentar el marco teórico utilizado para resolver el problema propuesto en el capítulo anterior. Se incluye además una revisión de la literatura correspondiente con un énfasis particular en los trabajos realizados en problemas similares al tratado en el presente trabajo.

II.2) Características y estructura de una base de datos.

De acuerdo a la literatura correspondiente (Cook y Brandon, 1984; Krajewski, 1984; Kruglinski, *op. cit.*; Lesk, 1984; Martip, 1977) las bases de datos están compuestas por CAMPOS (fields) y por REGISTROS (records). Un campo es representado por una variable del archivo, mientras que el registro es el conjunto de datos por cada campo de la base. Por ejemplo:

| NOMBRE | PESO | EDAD |
|----------------------|------|------|
| <u>M. tenellum</u> | 125 | 1.2 |
| <u>M. americanus</u> | 112 | 2.0 |

En este archivo tiene 3 campos (NOMBRE, EDAD y PESO) y 2 registros (uno para M. tenellum y otro para M. americanus). A la disciplina que estudia el tipo y distribución de registros y campos en una base de datos se le conoce como ANALISIS DE ESTRUCTURA DE DATOS (Flores, 1981) y es el primer paso en la construcción de una base de datos, pues gracias al análisis se determinará un formato de captura que luego será empleado para determinar los modos de selección en el archivo.

Cada base de datos tiene una arquitectura particular (Flores, *op cit*) que estara determinada tanto por el número de tipos de campos como por las características de cada campo particular. Por ejemplo:

| FECHA | HORA | PROFUNDIDAD |
|--------|-------|-------------|
| Feb.12 | 12.00 | 1 |
| | | 2 |
| | | 3 |
| Feb.13 | 15.00 | 2 |
| | | 4 |
| | | 5 |
| Feb.14 | 16.00 | 12 |
| | | 14 |
| | | 16 |

En este archivo existen 3 campos y los registros son compuestos, pues existen varias profundidades por

registro (existen varias profundidades por fecha). La arquitectura de este archivo está determinada por la existencia de 2 tipos distintos de campos y por las características de cada campo, pues las restricciones para el campo 1 (el día no puede ser superior a 28 ni menor que 1, por ejemplo) son distintas a las del resto. La determinación de la arquitectura de los datos es el primer paso en la elaboración de un paquete como H.B.C., y es aquí en donde el criterio del biólogo resulta indispensable, pues el proceso que seguirán los datos estará determinado en buena parte por la forma en que fueron estructurados durante su captura.

Otro paso importante en el desarrollo de una base de datos es la selección del tipo de base a utilizar (Krajewski, op. cit.). Las bases de datos pueden ser de los siguientes tipos:

a) Sistemas de Administración de Datos (DataBase Management Systems o DBMS). Este tipo de sistemas son capaces de almacenar grandes cantidades de datos con un formato fijo. La desventaja de los DBMS es que no pueden combinar la información contenida en dos archivos distintos aunque ambos tengan algunos campos en común (por ejemplo, no es posible combinar dos archivos que tengan el campo FECHA). Un sistema así no cumple eficientemente con los objetivos propuestos en I.3 pues no es posible buscar una correlación entre un dato hidrológico y uno biométrico, por ejemplo. Muchos DBMS no tienen rutinas de selección de datos.

b) Bases de Datos Relacionales (Relational Databases). Permiten la combinación de dos archivos distintos siempre y cuando ambos archivos tengan cuando menos un campo en común.

c) Bases de Datos Jerárquicas (Hierarchical Databases). Los registros no están descompuestos en campos, pues los datos tienen un valor que depende de su posición en la base. Las claves de identificación son bases de este tipo, pues los datos están colocados en una estructura vertical similar a un árbol (un árbol de decisiones). Los datos son localizados siguiendo un camino de elección jerárquica (generalmente binario, p. ej. "el organismo tiene 4 pares de patas si/no").

Las bases de datos pueden ser GENERALES, cuando pueden capturar cualquier valor (sin verificar su validez) y cualquier tipo de dato o ESPECIALIZADAS cuando verifican que los valores de los datos se encuentren en los límites definidos; además este tipo de bases sólo captura datos de un cierto tipo.

El siguiente paso en la elaboración de una base de datos consiste en la selección de un formato para la captura de los datos. En el caso del segundo ejemplo (que es similar al de las tablas del laboratorio) se podrían utilizar varias alternativas:

a) Se puede construir un arreglo tridimensional; el primer vector apuntaría al número de registro, el segundo al número de campo y el tercero al subcampo (dato de profundidad). El arreglo tendría la forma $X(3,3,9)$ (existen 3 registros, 3 campos y 9 subcampos). Este método desperdicia mucha memoria, pues la máquina reserva $3 \times 3 \times 9$ espacios (81 en total) cuando en realidad solo existen 15 datos en total (3 de fecha, 3 de hora y 9 de profundidad). La ventaja que presenta es que la selección de cualquier juego de datos es simple, pues basta con buscar los vectores que caracterizan a los datos selectos para encontrarlos. Por ejemplo, para buscar datos por fecha bastaría con buscar todos los elementos del arreglo, correspondientes a la profundidad, cuyo segundo vector sea 1 (el primer campo, FECHA). A pesar de estas ventajas, el desperdicio de memoria es tan alto que no es aconsejable usarlo en una microcomputadora, pues estas máquinas tienen una reserva limitada de ella.

b) Se puede generar una matriz cuyo primer vector apunte al número de campo y el segundo vector al número de registro; la diferencia con la opción anterior consiste en la ausencia de un apuntador del subcampo. La ventaja del sistema consiste en que sigue siendo

relativamente fácil y rápido el localizar cualquier campo, pero la selección de datos de cada subcampo se hace muy laboriosa, además el desperdicio de memoria sigue siendo grande, pues la máquina reservaría en el caso del ejemplo 27 datos (3 registros y 9 para campos y subcampos), 12 de los cuales serían desperdiciados. Este sistema tampoco cumple con los requisitos de eficiencia de memoria requeridos por las microcomputadoras.

c) Es posible generar dos matrices, una de las cuales almacenaría los datos de los primeros dos campos y la otra guardaría los datos del tercer campo (PROFUNDIDAD). Con este sistema se evita el desperdicio de memoria, aunque el problema de selección se complica un poco.

Existen otras formas de almacenar datos en una computadora, como las utilizadas en algunos de los trabajos mencionados en las referencias, pero no son aplicables en H.B.C. pues las técnicas desarrolladas en esos trabajos no son eficientes en el uso de la memoria, ya que fueron desarrolladas en máquinas grandes.

Como el mayor problema de las microcomputadoras de 8 bits (como APPLE) es el del espacio de memoria disponible (de 48 a 64 K en la mayoría de los casos) la tercera opción es la más practicable. Existen, desde luego, otras posibilidades documentadas en la literatura (Bond, 1984; Flores, op. cit.; Lesko, op. cit.; Martin, op. cit.), sin embargo esta opción es la que se puede implementar con más facilidad, dadas las características del sistema.

El último punto previo a la elaboración de la base de datos consiste en decidir el lenguaje a utilizar. Las microcomputadoras actuales ofrecen una gran variedad de lenguajes superiores como PASCAL, FORTRAN, LISP y COBOL por mencionar solo algunos. Cada lenguaje requiere de programas especializados (que consumen memoria). Las ventajas de un lenguaje superior son:

- a) La velocidad. La mayoría de los lenguajes superiores son compilados, por lo que el tiempo de ejecución es muy corto.
- b) La simplificación de procesos. Los lenguajes modernos, como PASCAL, permiten elaborar programas por unidades de trabajo, y cada unidad tiene una asignación específica (por ejemplo, una sección se dedica a seleccionar los datos, otra a corregir errores, etc.). A estos lenguajes se les conoce como LENGUAJES ESTRUCTURADOS.

Las microcomputadoras comunes tienen un lenguaje superior residente en memoria ROM; generalmente se trata del BASIC. Para implementar otro lenguaje se requiere de discos con los programas adecuados, de una memoria razonable (de 48 a 64 K como mínimo, aunque lo ideal son 128 K) y, de preferencia, de 2 lectoras de disco; sólo algunas microcomputadoras cuentan con este equipo. Es por ello que BASIC es el lenguaje ideal para elaborar un sistema que debe ser fácilmente transportable a otras máquinas (siempre y cuando el tiempo de ejecución no sea un problema).

II.3) Las bases de datos en Biología.

Existen bases de datos para cualquier computadora, desde las microcomputadoras de 16 K hasta las máquinas multiusuario con procesadores múltiples de 48 bits. Las bases de datos para microcomputadoras han evolucionado desde las "hojas de trabajo", como VisiCalc, elaborado por la corporación VisiCorp hasta las verdaderas bases de datos que permiten hacer una selección aleatoria de ellos, como dBASE II de la compañía Ashton Tate. Las bases de datos comerciales para microcomputadoras tienen algunas desventajas que las hacen inadecuadas para su uso en el laboratorio de Limnología. Estas desventajas son las siguientes:

- a) Algunas han sido diseñadas únicamente para aplicaciones comerciales y caseras. Este tipo

de bases generalmente carecen de subrutinas selectoras de datos; esta carencia en particular las hace prácticamente inútiles para los propósitos del laboratorio.

b) Prácticamente todas son demasiado generales y por lo tanto requieren de muchas modificaciones para adaptarlas a los problemas del laboratorio; estas modificaciones no son siempre posibles y en cualquier caso dificultan enormemente el uso del sistema.

c) Algunas son demasiado complicadas. Para utilizar este tipo de bases se requiere de mucha experiencia en el uso de microcomputadoras.

Existen muchos ejemplos de bases de datos programadas por profesionales de las Ciencias Biológicas para resolver algún problema particular en campos como la Medicina, la Ecología y la Genética. Otros paquetes, como el SPSS (Norman et al, 1975) han sido desarrollados para resolver problemas estadísticos en Ciencias Sociales pero se adaptan con facilidad a problemas biológicos; la mayoría de estos paquetes (incluyendo al SPSS) tienen una base de datos incluida. Algunos ejemplos de paquetes adaptados para resolver problemas biomédicos son el MUMPS y el INGRES (Held, 1975) que han sido aplicados en la investigación clínica. Los paquetes antes mencionados son bases de datos generales, pues el usuario tiene que especificar el nombre de los campos y sus condiciones (en la mayoría de los casos esto último no es posible y así aumenta la posibilidad de capturar un dato erróneo). Como este tipo de paquetes tiene muchas limitaciones, la Sociedad Austriaca del Corazón (Austrian Heart Association) utiliza ahora una nueva base de datos (Birman y Joskowicz, 1982). Este nuevo sistema tiene los datos de unos 6800 pacientes (cada paciente tiene de 20 a 150 datos clínicos). Otra base similar captura y selecciona los datos obtenidos en electroencefalogramas (Birman y Joskowicz *op. cit.*).

Otro ejemplo de base de datos aplicada a problemas en Ciencias Biomédicas es el EMIS (Environmental Mutagenesis Information System), desarrollado para el EMDP (Environmental Test Development Program). El EMDP es un programa de investigación iniciado en 1978 por el Instituto Nacional de las Ciencias de la Salud Ambiental (National Institute of Environmental Health Sciences) en Carolina del Norte, Estados Unidos (Rowley et al, 1982). El programa se basó en el paquete PROPHET, una base de datos general que, como en el caso anterior, no cubría todas las necesidades de los investigadores. Es importante señalar que ambos programas han sido desarrollados para máquinas grandes (el primer programa se diseñó para una PDP 11/34 con el operativo UNIX y el EMIS reside en una DEC System-10 con el operativo TOPS-10. En ambos casos se trata de máquinas grandes).

Existen pocas bases de datos elaboradas para resolver problemas biológicos en microcomputadoras pues en los países donde se ha desarrollado más la Computación existen los recursos para comprar máquinas grandes (como la PDP o la Burroughs). Un ejemplo de un paquete para microcomputadoras elaborado por profesionales de la Biología es el MORPHO (Prothero, J. y Prothero, J., 1982), el paquete consta de varios programas elaborados en FORTRAN (SPY, DMPSPY, STATSPY) y se le utiliza para reconstruir imágenes tridimensionales a partir de secciones histológicas seriales. El paquete requiere de una microcomputadora con 64 K, el operativo CP/M (una variedad de D.O.S), tres lectores de disco de 5.25 pulgadas (de 170 K c/u), de un monitor de 512 X 256 pixeles, una impresora, una tableta digitalizadora y un graficador; todo esto está dentro del alcance de un laboratorio de investigación de nuestro país. Es importante señalar que esta es una base de datos especializada; sólo una base de este tipo emplea la memoria de la máquina con la eficiencia necesaria para formar parte de un sistema tan complejo en una microcomputadora. Otros programas elaborados por biólogos, como el CURFIT presentado por J. Curts en su tesis doctoral (Curts, *op. cit.*) no incluyen bases de datos relacionales verdaderas, aunque ocasionalmente permiten el acceso a archivos de datos selectos generados por bases de datos como PFS-File o dBASE II. El empleo de estos sistemas esta limitado a personas con conocimientos de Computación.

Las diferencias entre las máquinas grandes y las microcomputadoras son tantas que resulta casi imposible (en el mejor de los casos) traducir un programa de una máquina multiusuario a una microcomputadora, pues las máquinas grandes poseen cantidades casi ilimitadas de memoria, además la velocidad de proceso es inmensamente superior; la CRAY-1, la máquina más rápida del mundo hasta hace unos años (Feutcher, 1984), puede realizar 100 megaflops. El proceso que toma 10 horas en una microcomputadora es realizado en 2 segundos en la CRAY-1. La sucesora de esta máquina es la CRAY-X1, aún más rápida. Las máquinas grandes pueden disponer además de lenguajes superiores (UMLL) que facilitan enormemente la tarea del programador. Estos lenguajes no son implementables en su totalidad en microcomputadoras por las razones expuestas arriba.

Resulta claro que el problema del laboratorio debe ser resuelto con técnicas diferentes a las empleadas en las grandes computadoras multiusuario.

Debido a la gran diferencia en capacidad de memoria, velocidad de proceso y lenguajes disponibles que existe entre las microcomputadoras y las máquinas grandes, por un lado, y al poco tiempo transcurrido desde la introducción de las primeras al mercado, prácticamente no existen antecedentes de trabajos directamente relacionados con el aquí expuesto. Las bases de datos han sido aplicadas de muy diversas formas en problemas biológicos. En los Estados Unidos y en el Canadá existen sistemas de bases de datos, llamados STORET (U.S.A.) y NAQUADAT (Canadá) que sirven para verificar la calidad del líquido contenido en distintos cuerpos de agua de esos países. En 1981 fue desarrollado un nuevo sistema (Deininger, 1981) que permite encontrar con facilidad aquellos datos de interés tanto para el científico como para el legista encargado de decidir si se han infringido las leyes de contaminación de los países en cuestión. El sistema fue desarrollado en una computadora AMDAHL 470/V8 con una memoria central de 16 Megabytes y lectoras periféricas con capacidad de 8 Gigabytes; compárese este sistema con la FRANKLIN ACE 1000 de 64 K de memoria residente y 170 K de memoria por cada cara de un disco de 5.25 pulgadas.

El sistema sucesor de NAQUADAT captura datos hidrológicos (como los programas HIDROLOGIA A1 e HIDROLOGIA A2 de H.B.C.); sin embargo el programa es intraducible a BASIC Applesoft por las razones mencionadas en el último párrafo.

Existen otros trabajos orientados a resolver problemas similares a los del laboratorio que, como los anteriormente expuestos, fueron parcial o totalmente desarrollados por profesionales de las Ciencias Biomédicas, pero todos ellos fueron elaborados en máquinas grandes (Birman y Jaskowicz *op cit*; Rowley *et al op cit*).

El único antecedente de H.B.C. fue elaborado por el Ing. Quím. Armando Reyes en 1981. Su trabajo consistió en la preparación de una base de datos en FORTRAN (para la terminal Burroughs del Instituto); la base sólo cubría a la tabla BIOMETRIA B2 (ver tablas anexas). En el período comprendido entre 1983 y 1984, el Act. José Antonio Villar elaboró una base de datos como parte de un paquete diseñado para realizar análisis multifactorial. Por su características y enfoque, esta base de datos no puede ser considerada como antecedente directo del trabajo realizado, aunque tanto el autor, como el trabajo mismo, sugirieron algunas de las técnicas empleadas en H.B.C..

En resumen, prácticamente no existen antecedentes directos del trabajo a realizar y por lo tanto su importancia va más allá de facilitar el proceso de los datos del laboratorio de Limnología.

II.4) Resumen.

En el presente capítulo se presentan los puntos a considerar durante la elaboración de un base de datos de acuerdo con la bibliografía especializada. Se revisa además el desarrollo de este tipo de programas en el ámbito de las Ciencias Biológicas en los últimos años. Debido al poco tiempo transcurrido desde su aparición, las microcomputadoras no han tenido todo el impacto que podría esperarse de ellas debido a su capacidad y bajo costo; prácticamente no existen antecedentes de un sistema como el descrito en el presente reporte.

CAPITULO TRES

DISEÑO DEL PAQUETE H.B.C.

III.1) Introducción.

El objetivo del presente capítulo es el de desarrollar los puntos propuestos en la sección I.3. Para ello se enumeran los criterios utilizados en el laboratorio de Limnología para seleccionar datos para un trabajo de investigación y se describen las técnicas elaboradas para implementar dichos criterios en el paquete H.B.C.. Se incluyen además 3 diagramas de flujo para mostrar gráficamente el proceso seguido por los datos desde su captura hasta su selección.

III.2) Los criterios de selección.

a) Para HIDROLOGIA A1

- 1.-Selección de parámetros totales (p. ej. todos los datos de nubosidad de la tabla).
- 2.-Selección de parámetros por código de localización (p. ej. todos los datos de nubosidad para alguna laguna en particular).
- 3.-Selección de datos muestrales generales (p. ej. todos los datos de salinidad de la tabla).
- 4.-Selección de datos muestrales por profundidad (p. ej. todos los datos de salinidad tomados en el fondo en toda la tabla).
- 5.-Selección de datos muestrales por código de localización (p. ej. todos los datos de salinidad tomados en alguna laguna en particular).
- 6.-Selección de datos muestrales por profundidad y código de localización (p. ej. todos los datos de salinidad del fondo de alguna laguna en particular).

b) Para HIDROLOGIA A2

- 1.-Selección de datos muestrales generales (p. ej. todos los datos de amonio total de toda la tabla).
- 2.-Selección de datos muestrales por profundidad (p. ej. todos los datos de amonio total del fondo).
- 3.-Selección de datos muestrales por código de localización (p. ej. todos los datos de amonio total tomados en alguna laguna en particular).
- 4.-Selección de datos muestrales por profundidad y código de localización (p. ej. todos los datos de amonio total del fondo de alguna laguna en particular).

c) Para BIOMETRIA B3

- 1.-Selección de datos muestrales generales (p. ej. todos los datos de longitud de rostro de toda la tabla).
- 2.-Selección de datos muestrales por código de localización (p. ej. todos los datos de longitud de rostro tomados en alguna laguna en particular).
- 3.-Selección por sexualidad de todos los datos muestrales (p. ej. todos los datos de longitud del rostro de los machos de toda la tabla).
- 4.-Selección por sexualidad y por código de localización (p. ej. todos los datos de longitud del rostro de los machos de alguna laguna en particular).
- 5.-Selección por estadio (hembras únicamente) de todos los datos muestrales (p. ej. los datos de longitud del rostro de las hembras prerreproductoras de toda la tabla).
- 6.-Selección por estadio (hembras únicamente) y por código de localización (p. ej. todos los datos de longitud

del rostro de las hembras prerreproductoras de alguna laguna en particular).

d) Para CAPTURA C1

- 1.-Selección de datos muestrales generales (p. ej. todos los pesos promedio de todos los organismos de la tabla).
- 2.-Selección de datos muestrales por código de localización (p. ej. todos los pesos promedio de todos los organismos capturados en alguna laguna en particular).
- 3.-Selección de datos muestrales por especie (p. ej. todos los pesos promedio de M. tenellum en toda la tabla).
- 4.-Selección de datos muestrales por especie y código de localización (p. ej. todos los pesos promedio de M. tenellum de una laguna en particular).

Los criterios fueron delimitados por entrevistas con el personal del laboratorio y las justificaciones son las siguientes:

- a) La mayoría de los crustáceos estudiados son bentónicos y por lo tanto los datos hidrológicos de interés son los del fondo del cuerpo de agua estudiado.
- b) Algunos organismos planctónicos (larvas de distintos organismos, copépodos, etc.) presentan zonalidad en su distribución vertical (prefieren ciertas profundidades) y en ocasiones presentan migraciones periódicas y por lo tanto los datos obtenidos a profundidades intermedias específicas serán de gran interés para algunos investigadores, pues permitirán identificar y describir dichas migraciones y a los organismos involucrados.
- c) Muchos trabajos de clasificación se basan en estadísticas sobre pesos y longitudes particulares de organismos que viven en distintos ambientes, y de ahí la importancia de la selección por código de localización. Además muchos organismos sólo se encuentran en localidades restringidas y el criterio antes mencionado podría servir para la identificación y descripción detallada de ellos.
- d) Muchos trabajos en Ecología, Biología Poblacional, Evolución y otras disciplinas se basan en datos obtenidos durante los muestreos. Los datos de captura, manejados con técnicas numéricas especiales, pueden dar una idea adecuada tanto de la cantidad de individuos en una población como de su variabilidad, distribución por edades, etc..

III.3) Selección de técnicas.

El sistema adoptado debe ser lo más eficiente posible en el uso de la memoria de la máquina, pues como se mencionó anteriormente, la capacidad de las microcomputadoras es limitada (en comparación con la de las máquinas grandes). Las características del sistema estarán determinadas por la arquitectura de los datos de cada tabla.

Las tablas HIDROLOGIA A1 e HIDROLOGIA A2 tienen una arquitectura similar, aunque A1 resulta un poco más compleja. La distribución de los datos es la siguiente en A1; existen 14 campos con un solo dato por estación (FECHA, HORA, %NUBES, TIPO NUBES, VELOCIDAD DEL VIENTO, DIRECCION DEL VIENTO, VELOCIDAD DE LA CORRIENTE, DIRECCION DE LA CORRIENTE, OLEAJE, SECHII, TEMPERATURA DEL AIRE, OLOR, COLOR y MATERIAL FLOTANTE) y 5 campos con varios (hasta 15) datos por estación (PROFUNDIDAD, TEMPERATURA DEL AGUA, OXYGENO DISUELTU, SALINIDAD y P.H.). Existe además una variable con el nombre de cada estación. Para almacenar estos datos con la mayor eficiencia posible se preparan 8 matrices o vectores. La primera matriz almacena los parámetros ambientales de las estaciones y tiene las dimensiones P(14,B), donde B es el número de estaciones en el archivo. Para capturar los datos restantes se prepara otra matriz con los vectores D(S,E), donde E es el número de datos muestrales totales del archivo. Con esta técnica los datos muestrales de cada estación quedan apilados unos sobre otros

(VER EJEMPLO). La tercera matriz es en realidad un vector de tamaño $F(I)$, donde I representa el número de datos muestrales por estación; el vector tiene B datos de $F(I)$. Los últimos vectores tienen también B datos y almacenan, en el vector alfanumérico $ES(I)$ el nombre de cada estación. Los datos del código de localización de cada estación están almacenados en los vectores $E(I)$ y $L(I)$ y los nombres integrantes del código en los vectores $CE(I)$ y $CL(I)$.

Los datos restantes (ver ANEXOS, tabla HIDROLOGIA A1) no se incluyen por estar en función de la salinidad y el oxígeno disuelto.

El caso de A2 es similar aunque solo existe un parámetro (FECHA) y hasta 12 datos muestrales (PROFUNDIDAD, ALCALINIDAD, DUREZA, AMONIO, NITRATOS, NITRITOS, ORTOFOSFATO, FOSFORO TOTAL, CLOROFILA A y hasta 3 variables cuyo nombre es especificado por el usuario).

Las tablas BIOMETRIA B3 (una modificación de la tabla B2 que únicamente sirve para crustáceos; B3 permite la captura de datos de cualquier organismo macroscópico, y algunos microscópicos, trabajados en el laboratorio) y CAPTURA C1 tienen, de nuevo, arquitecturas similares. La tabla B3 tiene 1 dato para toda la tabla (el nombre de la especie), 3 parámetros de la estación (FECHA, ARTE DE PESCA Y MALLA) y hasta 6 datos muestrales (desde 1 hasta 3 distintos tipos de longitud y desde 1 hasta 3 tipos de peso especificados por el usuario). Existe también un vector $F(I)$ cuyo papel es similar al del mismo vector en las tablas anteriores. La tabla C1 tiene los mismos 3 parámetros y tiene 7 datos muestrales (ESPECIE, # DE INDIVIDUOS, PESO TOTAL, PESO PROMEDIO POR INDIVIDUO, HORA INICIAL DE CAPTURA, HORA FINAL DE CAPTURA, DURACION).

EJEMPLO

Se presenta un ejemplo simplificado de la tabla A1. El manejo de las matrices y vectores es parecido en el resto de las tablas, aunque el número de vectores en cada matriz puede variar.

| FECHA | HORA | PROF. | O.D. | NOMBRE EST. |
|---------|-------|-------|------|-------------|
| 1604.85 | 12.12 | 0 | .121 | 101A |
| | | .1 | .125 | |
| | | .5 | .118 | |
| 1605.85 | 15.13 | 2 | .212 | 101B |
| | | 2.1 | .210 | |
| | | 2.5 | .202 | |
| | | 3 | .200 | |

La primera estación se localiza en el estado de México, laguna Amarilla; la segunda en el estado de Puebla, laguna Verde (los datos y las lagunas son ficticios).

Este ejemplo ilustra la estructura de las ocho matrices y vectores utilizados en los captadores de H.B.C.. El primer grupo de datos, con los parámetros, quedaría almacenado en la matriz $P(2,2)$.

| | 1 | 2 |
|---|---------|-------|
| 1 | 1604.85 | 12.12 |
| 2 | 1605.85 | 15.13 |

Los datos muestrales están almacenados en una matriz $D(2,7)$.

| | 1 | 2 |
|---|-----|------|
| 1 | 0 | .121 |
| 2 | .1 | .125 |
| 3 | .5 | .118 |
| 4 | 2 | .212 |
| 5 | 2.1 | .210 |
| 6 | 2.5 | .202 |
| 7 | 3 | .200 |

Para identificar los datos de la estación I existe el vector F(I) con los siguientes valores.

| | |
|---|---|
| 1 | 3 |
| 2 | 4 |

El vector ES\$(I) identifica a las estaciones:

| | |
|---|------|
| 1 | 101A |
| 2 | 101B |

Los vectores CE\$(I) y CL\$(I), combinados con E(I) y L(I), almacenan, respectivamente, los datos del estado y de la localidad de cada estación.

| | CE\$(I) | CL\$(I) |
|---|---------|-----------------|
| 1 | Puebla | Laguna Amarilla |
| 2 | Mexico | Laguna Verde |

| | E(I) | L(I) |
|---|------|------|
| 1 | 2 | 1 |
| 2 | 1 | 2 |

Los últimos vectores indican que la primera estación está localizada en el segundo estado y en la primera localidad definidas en el código almacenado en CE\$(I) y en CL\$(I).

El primer valor indica que la amplitud del subcampo de la estación 1 (número de datos muestrales de dicha estación) es de 3; el 4 indica la longitud del subcampo de la estación 2.

Como se puede apreciar en el ejemplo, esta técnica permite optimizar al máximo el uso de la memoria de la máquina, pues no se desperdicia ningún elemento de los arreglos. Otra ventaja del sistema es que facilita la búsqueda de datos de acuerdo al criterio utilizado en el laboratorio. Al investigador le interesa localizar sus datos por profundidad, pues muchos de ellos se especializan en organismos bentónicos y por lo tanto los datos pertinentes a ellos se encontrarán en el fondo, de manera que los datos de interés se encontrarán en la matriz D; el primer apuntador corresponderá al tipo de dato deseado (oxígeno disuelto u O.D.) y los segundos

apuntadores corresponderán con el último dato de cada estación. Estos apuntadores se calcularán fácilmente a partir del vector F. Se puede incluir una rutina que verifique el código de localización de cada estación, para seleccionar sólo aquellos datos de profundidad máxima que fueron obtenidos en las localidades de interés para el investigador.

Los arreglos son alfanuméricos para incluir la posibilidad de un dato no numérico; cuando el dato no existe basta con pulsar N. De usarse un arreglo numérico, el dato N sería interpretado como una cantidad ilegal y el programa sería interrumpido.

El tipo de base de datos que corresponde mejor a esta descripción es el de BASE DE DATOS RELACIONAL ESPECIALIZADA, pues es capaz de establecer una relación entre los datos de cualquier forma codificada ya que es posible seleccionar datos por criterios comunes a cualquiera de ellas (por código de localización, por ejemplo); además este sistema está orientado a resolver un problema específico. Por el encadenamiento de los programas que lo forman, es posible definir a H.B.C. como un paquete jerárquico.

El lenguaje empleado es el BASIC pues se pretende que H.B.C. pueda ser implementado en otras microcomputadoras comunes, y el uso de un lenguaje más avanzado restringiría el uso del paquete a usuarios con el equipo (y dinero) suficientes para implementar un nuevo lenguaje.


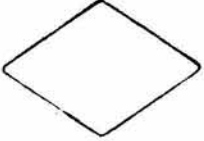
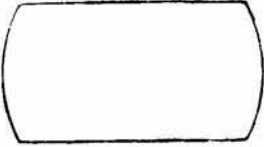






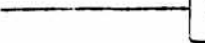
III.4) Diagramas de flujo.

El paso siguiente a la selección de técnicas consiste en la elaboración de una secuencia simbólica de pasos que tendrá que seguir el programa para resolver el problema. A esta secuencia de pasos, elaborada en papel con una serie de símbolos específica se le conoce como DIAGRAMA DE FLUJO pues permite seguir con facilidad la secuencia en la ejecución de un programa. No se incluyen listados del paquete H.B.C. debido a las diferencias existentes entre los distintos dialectos de BASIC de las microcomputadoras más comunes, sin embargo los diagramas permiten implementar el algoritmo propuesto anteriormente con facilidad. Los listados de H.B.C. en BASIC (Applesoft) pueden ser consultados en el laboratorio de Limnología del Instituto de Ciencias del Mar y Limnología a cargo de M. en C. Manuel Guzmán Arroyo.

Los diagramas de flujo pueden ser muy detallados aunque en el caso de un paquete grande, como H.B.C., un diagrama perfectamente detallado sería contraproducente, pues en lugar de aclarar el proceso del programa lo complicaría enormemente. Como la estructura de los programas captadores son similares entre sí, y como la misma situación se presenta para los selectores, se ilustran únicamente 3 diagramas, el diagrama general del sistema, uno para los captadores y otro para los selectores. Se incluye además una tabla de figuras para facilitar la interpretación de los diagramas.

En el comentario sobre PARAMETROS en el diagrama captador se incluyen los datos que aparecen sólo una vez en cada estación, como el número de estaciones, la lista de especies (C1), los tipos de longitudes y de pesos (B3), etc.,

TABLA DE FIGURAS UTILIZADAS EN DIAGRAMAS DE FLUJO

| | |
|---|---|
|  | Sección encargada de un proceso simple o de un proceso similar a otro predefinido en el diagrama. |
|  | Condicional (para elegir entre dos opciones). |
|  | Pantalla (para mostrar datos o resultados de procesos). |
|  | Entrada manual de datos (por teclado). |
|  | Programa. |
|  | Punto de unión de una línea de flujo. |
|  | Dirección del flujo. |
|  | Impresión en papel. |
|  | Operación de almacenamiento o carga de datos grabados en un dispositivo magnético. |
|  | Comentario al programa. |

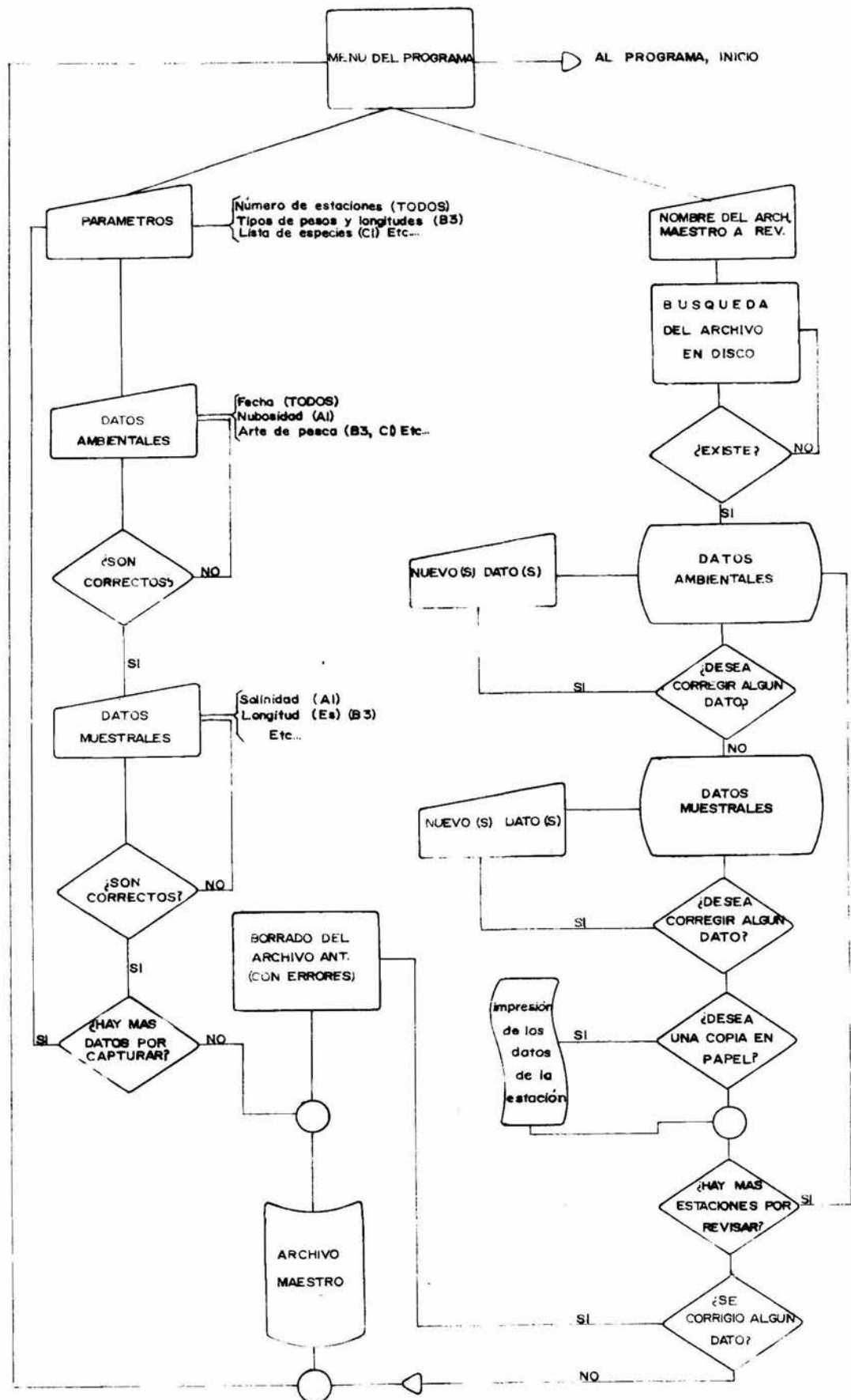


Fig. 1. ESTRUCTURA GENERAL DE UN CAPTADOR

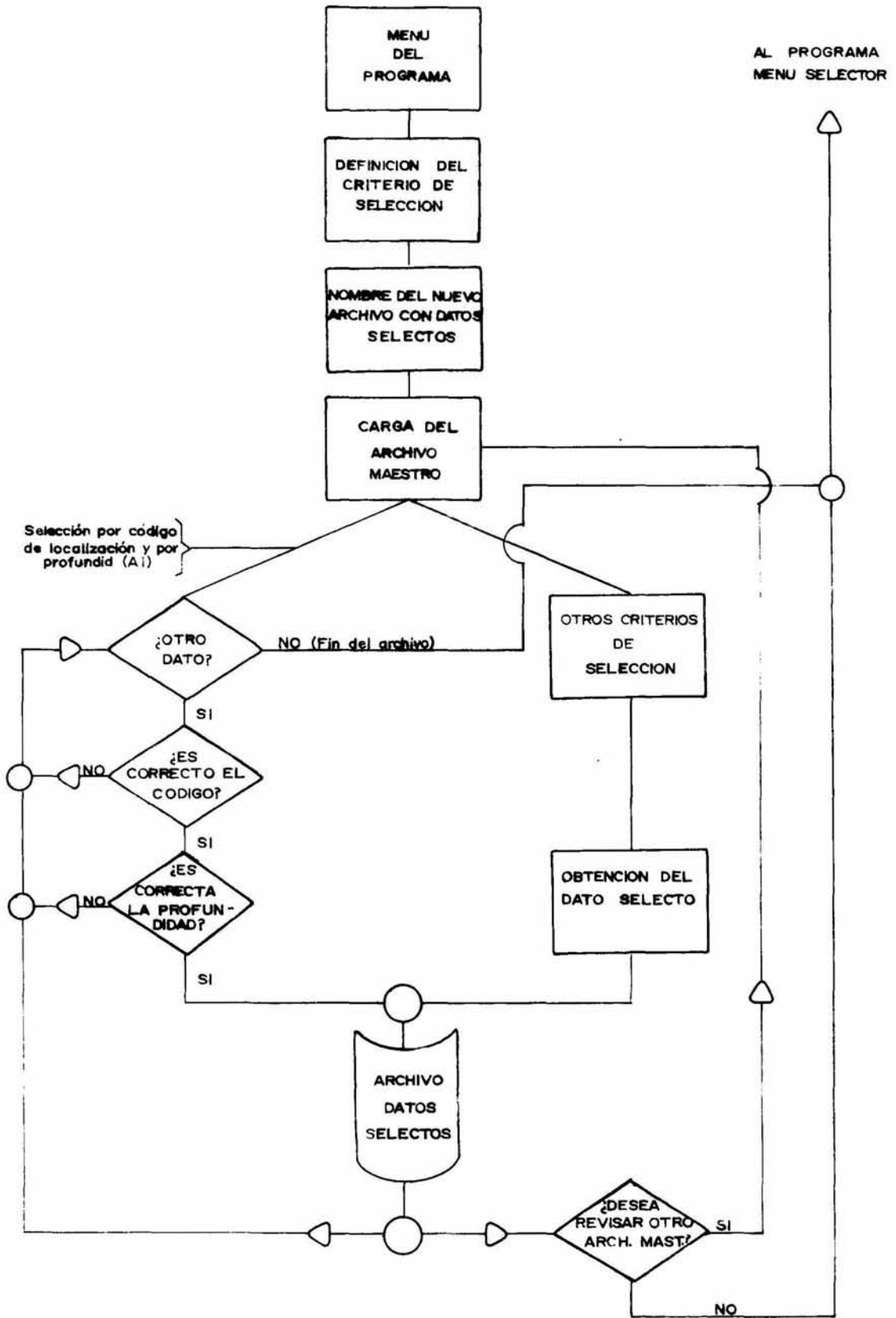


Fig. 2. ESTRUCTURA GENERAL DE UN SELECTOR

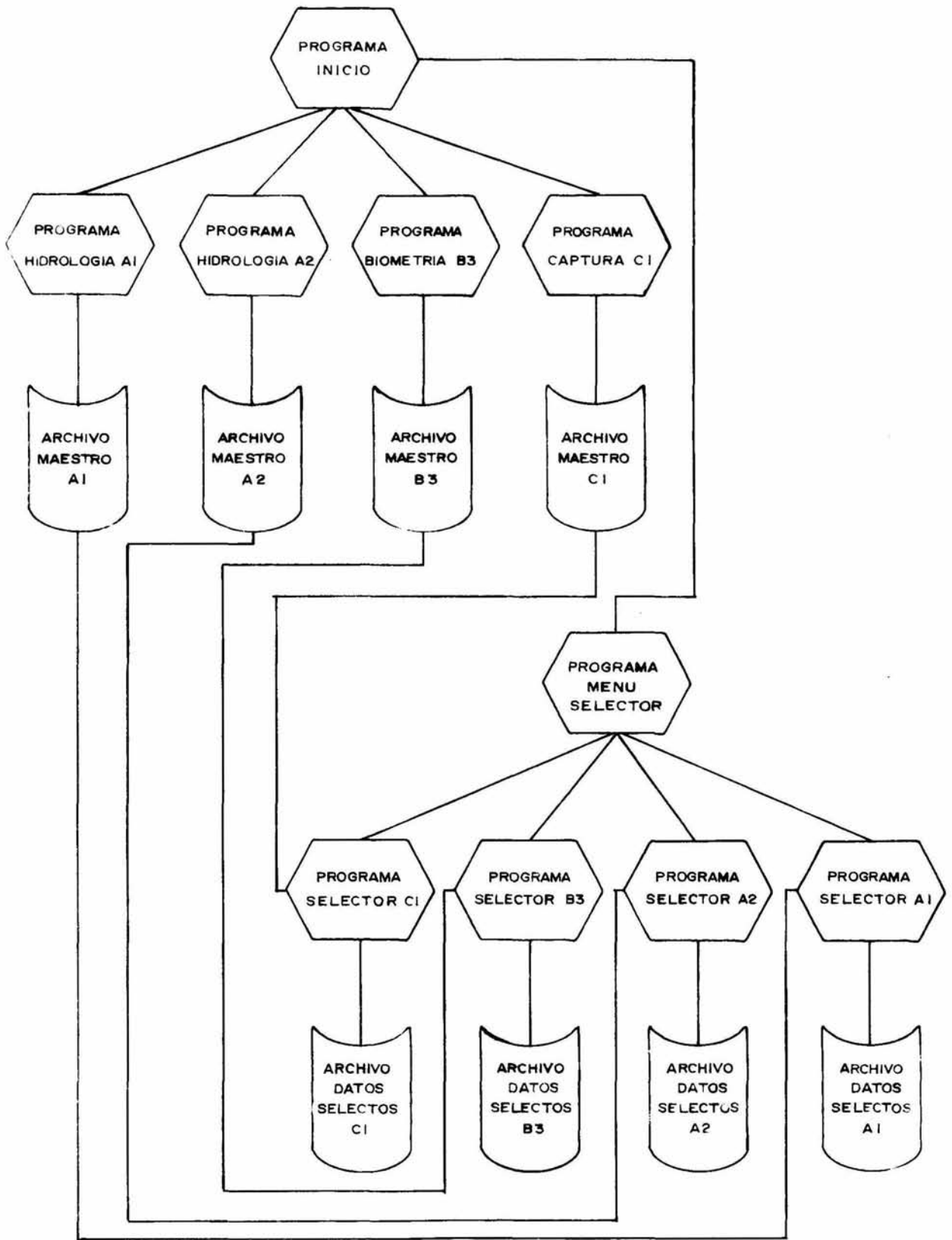


Fig. 3. ESTRUCTURA GENERAL DEL H.B.C.

III. 5) Descripción del paquete H.B.C.

El paquete H.B.C. versión 3.0 consta de 11 programas:

-INICIO Este programa hace el papel de menú general del paquete, a través de él se puede llamar a cualquier programa captador de datos o al menú de selección. Todos los demás programas captadores (y el MENU SELECTOR), llaman a este programa cuando terminan su trabajo.

-HIDROLOGIA A1, HIDROLOGIA A2, BIOMETRIA B3 y CAPTURA C1 capturan los datos muestrales.

-MENU SELECTOR Este programa se encarga de llamar a cualquier programa selector. También llama al programa encargado de imprimir los datos selectos.

-SELECTOR A1, SELECTOR A2, SELECTOR B3 y SELECTOR C1 se encargan de seleccionar los datos de un archivo maestro almacenado en disco con los criterios mencionados anteriormente. Una vez terminada su labor llaman al programa MENU SELECTOR

-IMPRESOR Se encarga de imprimir los archivos con datos selectos; la impresión se puede hacer en pantalla o en impresora.

El sistema fue realizado con el D.O.S. Pronto-Dos modificado con una rutina en lenguaje de máquina que permite elaborar el directorio de aquellos archivos relativos al programa en cuestión. El sistema Pronto-Dos fue elegido por la velocidad de acceso al disco, ya sea para lectura o escritura (este tiempo puede ser hasta 10 veces menor que el empleado por el D.O.S. original de Apple). La modificación fue realizada con una adaptación del programa CATALOG.PLUS (Costales, 1985), y consiste en la alteración de la función CATALOG que elabora un listado de los archivos contenidos en un disco magnético.

Los programas captadores constan de un menú de 3 opciones, la primera que permite la captura de datos, la segunda que permite la revisión (e impresión en papel) de un archivo en disco y la tercera que permite el retorno al menú general del programa; los archivos son identificados tanto por los captadores como por los selectores con ayuda de un prefijo (A1., A2., B3., C1., para los archivos maestros y AS. para los archivos de datos selectos).

Todos los programas pueden ser operados prácticamente con una tecla, pues incluyen opciones si/no (p. ej. "DESEA UNA COPIA EN PAPEL? (S/N)"). Además se incluyen rutinas que verifican los datos para disminuir mucho la probabilidad de error. Estas técnicas facilitan la interacción con usuarios no entrenados en el uso de microcomputadoras.

Todas las tablas fueron ligeramente modificadas en el paquete, pues existe en ellas información redundante. En el caso de A1 se capturan los datos de SALINIDAD y de CONDUCTIVIDAD, como ambas están relacionadas directamente (el agua destilada, de salinidad 0, no es conductora), se dejó fuera la columna de CONDUCTIVIDAD ya que generalmente se captura la SALINIDAD y la otra variable se calcula a partir de ésta. Otro caso similar ocurre con O.D. y SAT. de OXIGENO.

La tabla Biometría B3 merece consideración aparte, pues su antecesora, B2, únicamente captura datos pertinentes a crustáceos y con ello deja fuera a una gran gama de organismos acuáticos, como vertebrados e insectos, de gran importancia limnológica. La tabla B2 fue modificada para permitirle al usuario especificar el nombre de hasta tres tipos de longitudes y de hasta tres pesos distintos; de esta manera es posible capturar datos tanto para peces (como LONGITUD PATRON o PESO EVICERADO) como para crustáceos (como LONGITUD ROSTRO o PESO TOTAL). La técnica utilizada para elaborar este programa podría ser utilizada posteriormente para ampliar la captura de datos a organismos como los protozoos o las bacterias

La primera opción también incluye una rutina para definir un código de localización en dos niveles, ESTADO

y LOCALIDAD; cada estación estará asociada a dos números de este código. Cada estación también incluye un espacio alfanumérico para definir el NOMBRE de la estación.

La opción 2 de cada captador tiene un detector de correcciones, si el usuario corrige algún dato, la máquina automáticamente borra el antiguo archivo y genera uno nuevo con los datos correctos.

Los formatos de almacenamiento de datos son parecidos en todos los captadores; primero se graban los parámetros del archivo, como el número de estaciones (B) y luego se capturan los datos. Los archivos son exclusivos del captador en cuestión (un archivo A1 no es accesible para revisión en A2) y por ello están identificados por un prefijo (A1, A2, B3, C1, para los archivos maestros, y SA1, SA2, SB3 y SC1 para archivos con datos selectos).

Los programas selectores son manejados por un programa menú, parecido en funciones y estructura al programa INICIO. El uso de programas menú (conocidos como DRIVER) permite emplear la memoria de disco en forma eficiente, pues de esa manera los programas quedan efectivamente unidos sin tener que ser cargados a la memoria de la máquina simultáneamente.

Los programas selectores permiten el acceso a aquellos datos de interés para el investigador con los criterios de selección mencionados anteriormente. Cada criterio de selección está representado por una línea de programa. De esta manera es posible corregir cualquier problema con facilidad; además resulta sencillo agregar nuevos criterios de selección (selección por ARTE DE PESCA, por ejemplo) que actualmente no son utilizados en el laboratorio.

Los programas permiten el acceso secuencial a cada archivo maestro para así almacenar los datos selectos en un solo archivo que contenga datos selectos de varios años.

III.6) Resumen.

En este capítulo se resumen las técnicas desarrolladas para cumplir con los requisitos señalados en la sección I.3. Se eligió un sistema que consta de 8 arreglos que almacenan tanto a los datos de trabajo como a una serie de parámetros numéricos que facilitan la localización de cada dato de acuerdo a los criterios de selección, también descritos en este capítulo. Se eligió el lenguaje BASIC para elaborar la base de datos pues es un lenguaje común a otras microcomputadoras y por ello el sistema resulta transportable (con algunas modificaciones).

CAPITULO CUATRO

DISEÑO EXPERIMENTAL Y RESULTADOS

IV.1) Introducción.

El objetivo del presente capítulo es el de especificar los puntos por verificar en el funcionamiento del paquete H.B.C.; además se incluye un ejemplo del funcionamiento del paquete desde la captura de datos hasta su selección. Los datos incluidos en el ejemplo son ficticios, pero representan adecuadamente a los datos reales.

IV.2) Puntos por verificar.

Los puntos por verificar en el funcionamiento de los programas que componen al paquete H.B.C. son:

- a) Que capturen correctamente los datos almacenados en las formas codificadas del laboratorio de Limnología (sección I.3, objetivo a).
- b) Que la implementación de las rutinas de selección sea correcta (sección I.3, objetivo b).
- c) Que la corrección de errores sea sencilla y funcione correctamente (sección I.3, objetivo c).
- d) Que el paquete sea sencillo de manejar para facilitar la interacción con usuarios no especializados (sección I.3, objetivo d).

IV.3) Ejemplo del funcionamiento de H.B.C..

El paquete fue sometido a distintas pruebas para verificar el cumplimiento de los puntos señalados en la sección anterior. Los puntos a y c no son mostrados en esta sección pues su verificación solo se puede realizar con una microcomputadora; estos puntos fueron verificados tanto en el diagrama de flujo como en la versión de trabajo. El punto d fue verificado de la manera siguiente; se entregó el paquete a algunas personas del laboratorio de Limnología junto con una pequeña serie de instrucciones verbales y un juego de datos ficticios y a la media hora se realizó una encuesta entre los participantes. Los resultados de esta prueba del funcionamiento de H.B.C. entre personas no especializadas son interesantes, pues todas fueron capaces de emplear correctamente todas las funciones del paquete.

El punto b fue verificado con una serie de pruebas similares a la mostrada en esta sección. Esta función del paquete es la más utilizada pues los datos solo se capturan una vez, pero la selección de datos se realizará cada vez que el investigador necesite analizar algún aspecto del lugar (o organismo) estudiado.

En la prueba siguiente se muestra un archivo de cuatro estaciones para la forma HIDROLOGIA A1 y se muestra un ejemplo de selección, se eligieron los datos de SALINIDAD por profundidad y por código de localización (se solicitaron todos los datos de salinidad del fondo de la Laguna Amarilla localizada en el estado de México; estos datos son ficticios).

A1 PRUEBA HIDROLOGIA

ESTADO: MEXICO
MUESTRA: 1

LOCALIDAD: LAGUNA AMARILLA
ESTACION: LAE01

DATOS AMBIENTALES

FECHA: DIA 4 MES 6 AÑO 85
% DE NUBES: 100
VEL. VIENTO: 15 MTS / SEG.
VEL. CORR: 8. MTS / SEG.
OLEAJE: FUERTE
TEMPERATURA DEL AIRE: 14 CENTIGRADOS
COLOR: VERDE

HORA: 10 HRS 50 MIN.
CUMULUS/STRATUS
DIRECCION: 45 GRADOS
DIRECCION: 43 GRADOS
SECHIT: 1 CENTIMETROS
OLOR: CON OLORES

MATERIAL FLOTANTE: ESPUMA

DATOS HIDROLOGICOS

| MUESTRA | PROFUNDIDAD METROS | TEMPERATURA AGUA CENTIGRADOS | OXIGENO DISUELT O RPM | SALINIDAD O/00 | P. H |
|---------|--------------------|------------------------------|-----------------------|----------------|-------|
| 1 | 0 | 12 | 405 | 404 | 7.600 |
| 2 | 1 | 11.8 | 407 | 398 | 7.58 |
| 3 | 2 | 11.8 | 405 | 359 | 7.65 |
| 4 | 3 | N | 39 | 298 | 7.610 |
| 5 | 4 | 11.299 | 369 | 286 | 7.600 |
| 6 | 5 | 11.129 | 345 | 299 | 7.590 |

A1 PRUEBA HIDROLOGIA

ESTADO: GUANAJUATO
MUESTRA: 2

LOCALIDAD: LAGUNA ROJA
ESTACION: LRE03

DATOS AMBIENTALES

FECHA: DIA 15 MES 6 AÑO 85
% DE NUBES: 0
VEL. VIENTO: 0 MTS / SEG.
VEL. CORR: 0 MTS / SEG.
OLEAJE: NO HAY
TEMPERATURA DEL AIRE: 19 CENTIGRADOS
COLOR: INCOLORA

HORA: 14 HRS 11 MIN.
SECHIT: 15 CENTIMETROS
OLOR: INCOLORA

MATERIAL FLOTANTE: NO HAY

DATOS HIDROLOGICOS

| MUESTRA | PROFUNDIDAD METROS | TEMPERATURA AGUA CENTIGRADOS | OXIGENO DISUELT O RPM | SALINIDAD O/00 | P. H |
|---------|--------------------|------------------------------|-----------------------|----------------|-------|
| 7 | 0 | 19.5 | 39 | 30 | 7.610 |
| 8 | 1 | 19.299 | 37 | 358 | 7.600 |
| 9 | 2 | 19.1 | 359 | 344 | 7.580 |
| 10 | 3 | 19 | 344 | 345 | 7.422 |
| 11 | 3.5 | 18.7 | 398 | 375 | 7.45 |
| 12 | 3.799 | 16.899 | 193 | 4 | 7.4 |

A1. PRUEBA HIDROLOGIA

ESTADO: JALISCO
MUESTRA: 3

LOCALIDAD: LAGUNA VERDE
ESTACION: LV44

DATOS AMBIENTALES

FECHA: DIA 30 MES 3 AÑO 85
% DE NUBES: 10
VEL. VIENTO: 1.9 MTS / SEG
VEL. CORR.: 0 MTS / SEG
OLEAJE: LIGERO
TEMPERATURA DEL AIRE: 25 CENTIGRADOS
COLOR: AMARILLA

HORA: 16 HRS. 16 MIN
DIRECCION: N GRADOS
DIRECCION: N GRADOS
SECHII: 12 CENTIMETROS
OLOR: CON OLOF
MATERIAL FLOTANTE: PETROLEO

DATOS HIDROLOGICOS

| MUESTRA | PROFUNDIDAD METROS | TEMPERATURA AGUA CENTIGRADOS | OXIGENO DISUELTO PPM | SALINIDAD 0/00 | P. H |
|---------|--------------------|------------------------------|----------------------|----------------|------|
| 1 | 0 | 20 | .202 | 125 | 7.1 |
| 2 | 1 | 22 | 1 | 2 | 7.39 |
| 3 | 2 | 22.5 | .089 | 219 | 7.1 |
| 4 | 3 | 22.299 | .75 | 215 | 7.55 |

A1. PRUEBA HIDROLOGIA

ESTADO: MEXICO
MUESTRA: 4

LOCALIDAD: LAGUNA AMARILLA
ESTACION: LAE02

DATOS AMBIENTALES

FECHA: DIA 4 MES 4 AÑO 85
% DE NUBES: 96
VEL. VIENTO: 4 MTS / SEG
VEL. CORR.: 2 MTS / SEG
OLEAJE: MODERADO
TEMPERATURA DEL AIRE: 12 CENTIGRADOS
COLOR: VERDE

HORA: 18 HRS. 58 MIN.
STRATOCUMULUS
DIRECCION: 92 GRADOS
DIRECCION: 90 GRADOS
SECHII: 5 CENTIMETROS
OLOR: CON OLOF
MATERIAL FLOTANTE: ESPUMA

DATOS HIDROLOGICOS

| MUESTRA | PROFUNDIDAD METROS | TEMPERATURA AGUA CENTIGRADOS | OXIGENO DISUELTO PPM | SALINIDAD 0/00 | P. H |
|---------|--------------------|------------------------------|----------------------|----------------|-------|
| 5 | 0 | 15 | .5 | 304 | 7.599 |
| 6 | 1 | 16 | .505 | 344 | 7.49 |
| 7 | 2 | 15.6 | 408 | 391 | 7.55 |
| 8 | 3 | 15.1 | 39 | 329 | 7.599 |

SA1. PRUEBA DE SELECCION

| NUMERO | MUESTRA | FECHA | ESTACION | DATO SELECTO |
|--------|---------|---------|----------|--------------|
| 1 | 1 | 0406.85 | LAE01 | 283 |
| 2 | 4 | 0406.85 | LAE02 | 329 |

IV.4) Resultados.

El ejemplo anterior fue realizado con el programa BIOMETRIA B3; los datos fueron seleccionados con el programa SELECTOR B3 y fueron impresos con el programa IMPRESOR. Los resultados de cada punto mencionado en la sección IV.2 son los siguientes:

- a) Los datos capturados por el programa fueron comparados con los anotados en una tabla aparte. Se incluyeron todas las variantes posibles para los datos; por ejemplo, cuando el dato de SEXUALIDAD es 1 (MACHO), el dato de ESTADIO no existe. Además, cuando un dato es desconocido el programa almacena una N.
- b) Las técnicas de selección fueron verificadas una por una en todos los programas correspondientes con pruebas similares a las del ejemplo anterior. Las técnicas de selección son independientes del tipo de dato (ya sea numérico o una N).
- c) La corrección de errores fue verificada incluyendo errores intencionales en todos los campos de todos los programas captadores; además se incluyeron situaciones especiales, como el intento de grabación de un archivo en una lectora sin disco, etc..
- d) El paquete fue presentado a varias (6) personas del laboratorio junto con una serie sencilla de instrucciones verbales; en 15 minutos todos los participantes eran capaces de emplear el paquete en forma adecuada.

La idea de incluir decisiones binarias ("SON CORRECTOS SUS DATOS? (S/N)") resultó adecuada, pues reduce la interacción del usuario con la computadora a decisiones simples, seleccionadas por la pulsación de una sola tecla. Si bien 6 personas no son una muestra representativa del total de la población de investigadores del país, el hecho de que el paquete haya sido empleado correctamente por ellos en un tiempo tan corto da una idea del grado en que se cumplió este último objetivo.

IV.5) Resumen.

En esta sección se describen las pruebas a las que fue sometido el paquete H.B.C.; las pruebas fueron diseñadas para verificar todos los puntos mencionados en la sección I.3. En este capítulo se presenta la prueba del punto b (selección de datos de acuerdo a los criterios del laboratorio) pues es la opción más utilizada por el investigador ya que solo captura los datos una vez. Los otros puntos fueron verificados con la ayuda de los miembros del laboratorio de Limnología.

CAPITULO CINCO

DISCUSION

V.1) Introducción.

En este capítulo se discuten los resultados de las pruebas descritas en el capítulo anterior y se compara además la diferencia del rendimiento de la máquina con el de una persona entrenada en el uso de las formas codificadas. Se incluyen además algunos comentarios sobre las ventajas y desventajas del sistema.

V.2) Discusión.

Los programas captadores facilitan la captura pues incluyen rutinas que impiden la entrada de datos obviamente falsos (como un número de mes superior a 12); algunos datos ilógicos pueden pasar, como un día 31 en el mes 2 (Febrero), aunque estas fallas de seguridad son pequeñas y la implementación de una rutina que las elimine sería complicada y resultaría muy costosa en lo que al espacio de memoria se refiere.

Los programas selectores funcionan de acuerdo a los criterios de selección. Como cada criterio de selección está implementado en una línea de programa, resulta sencillo modificarlo. Además resulta muy fácil agregar nuevos criterios de selección, pues basta con agregar una línea al programa y modificar las líneas encargadas de seleccionar el criterio correspondiente.

El uso de la memoria en H.B.C. es eficiente pues muchos datos son almacenados en clave, por ejemplo el oleaje es almacenado como un número entero de 1 a 4 que corresponde a las cuatro categorías, desde INEXISTENTE hasta FUERTE. De esta forma es posible evitar un desperdicio grande de memoria, pues por cada dato de este tipo solo se almacena un caracter en lugar de los 5 o 10 necesarios para especificar (con letra) el oleaje correspondiente. Lo mismo ocurre con el color, el olor, el arte de pesca y otros datos. Como el funcionamiento de los programas de este sistema en particular no depende del número de datos procesados, la eficiencia en el cumplimiento de sus funciones será constante (es decir, si no comete errores con 10 datos no los cometerá con 1000). En comparación, la eficiencia del ser humano disminuirá con el tiempo, pues el cansancio alterará la capacidad de concentración del individuo. En una prueba de comparación entre el ser humano y el programa, el ser humano cometió 4 errores en la selección de datos en un archivo de 10 estaciones y 12 errores en uno de 22 estaciones mientras que la máquina no cometió un solo error; además el tiempo empleado en seleccionar datos de una sola estación fue aumentando de unos pocos segundos a casi un minuto en el caso del ser humano, y en el caso de la máquina el tiempo empleado en este proceso fue constante para cada estación analizada. Otros problemas que pueden afectar la calidad de los datos son la ambigüedad en la interpretación de la escritura de la persona que los capturó (a veces el número 6 parece 0, por ejemplo), la pérdida de una hoja del archivo, la transposición de un archivo por otro, etc.; estos errores no son cometidos por el paquete H.B.C. si es usado adecuadamente.

Algunas de las desventajas posibles de H.B.C. son:

- a) La sobreespecialización. Es necesario modificar el programa para que capture otro tipo de datos, aunque esta modificación es sencilla y puede tomar una semana de trabajo a una persona con conocimientos de BASIC.
- b) La ausencia de programas de apoyo. Esta versión de H.B.C. no cuenta con un programa copiador o un programa inicializador de discos (por ejemplo) y por ello el usuario deberá recurrir a otros programas.

c) La interacción entre el capturista y el programa. La captura de datos se lleva a cabo, en el caso del laboratorio, en el campo, y siempre existe la posibilidad de error durante la lectura del dato o durante su captura en las formas codificadas.

d) La falta de comunicación con otros programas. La versión actual de H.B.C. no es capaz de generar archivos de datos selectos accesibles a programas de proceso posterior, como paquetes estadísticos o de graficación. Esta opción no ha sido implementada en H.B.C. pues cada programa de proceso posterior acepta archivos con un formato muy particular, de manera que si se elaboran archivos accesibles a un programa de graficación, como VISIPILOT/VISITREND, algunos programas de proceso estadístico resultaran inútiles al no tener acceso a los datos selectos. Para implementar esta opción es indispensable que el investigador decida qué programas de acceso posterior desea encadenar con la base de datos; el resto de los programas recibirán los datos en forma manual, como ocurre actualmente en el laboratorio de Limnología. Es importante señalar que el volumen de los datos selectos generalmente es reducido y por ello la entrada manual de datos a un programa de proceso posterior no resulta muy difícil, sin embargo aumenta ligeramente la probabilidad de alterar algún dato durante el transporte.

V.3) Resumen.

En el presente capítulo se ha discutido la eficiencia con la que el paquete H.B.C. cumple con los objetivos señalados en la sección I.3. El paquete cumple con todos los objetivos con eficiencia, sin embargo puede sufrir aun varias mejoras que facilitarían aun más su uso. Estas medidas no fueron implementadas en esta versión pues no fueron consideradas indispensables para el cumplimiento de los objetivos señalados. Debido a que una persona con conocimientos elementales de BASIC puede modificar con facilidad al paquete, H.B.C., este podría ser adaptado con facilidad a cualquier otro laboratorio de Biología.

CAPITULO SEIS

CONCLUSIONES Y RECOMENDACIONES

VI.1) Introducción.

En el presente capítulo se presentan las conclusiones sobre la eficiencia del paquete H.B.C. y se discute la proyección tanto de las bases de datos como de las microcomputadoras en la Biología Moderna. Se incluyen además algunas recomendaciones para el usuario.

VI.2) Conclusiones sobre el paquete H.B.C.

El paquete H.B.C. comenzó a ser utilizado en el laboratorio de Limnología desde su primera versión (Enero, 1985) y a pesar del tiempo tan corto que ha transcurrido, ya se emplea en varios proyectos de investigación a distintos niveles. Algunos de estos proyectos son "Características Limnológicas del Rio Duero" y "Atlas Limnológico del Lago de Patzcuaro" (en colaboración con el Gobierno del Estado de Michoacán). El laboratorio de Química del Instituto de Ciencias del Mar y Limnología también está empleando el paquete H.B.C. (con algunas modificaciones).

La facilidad de almacenar archivos en disco con la consiguiente facilidad tanto para el acceso de los datos como para su copiado en otros discos lo hacen muy adecuado para las necesidades del laboratorio de Limnología. Las técnicas desarrolladas para la captura y la selección de datos presentan las siguientes ventajas:

- a) Rapidez en el manejo. Las técnicas desarrolladas para la captura y selección de datos en H.B.C. son lo suficientemente eficientes como para permitir la selección de los datos para un trabajo cualquiera en una jornada de trabajo (para el caso de un archivo elaborado a partir de los 500,000 datos del laboratorio). En la mayoría de los casos no será necesario consultar todos pues los trabajos generalmente tienen limitaciones (de fecha, por ejemplo) que facilita aun más el proceso de búsqueda.
- b) Facilidad en el manejo. El paquete facilita el trabajo del usuario pues no acepta errores fácilmente; además siempre es posible recuperarse de uno con facilidad.
- c) Transportabilidad. Como el paquete fue desarrollado en BASIC es fácilmente traducible a otras máquinas comunes con el mismo lenguaje.
- d) Facilidad de adaptación. Como cada criterio de selección está representado por un línea de programa, bastará con alterar estas líneas para adaptar el programa a otros ambientes.

El paquete H.B.C. no es una base de datos relacional completa pues no puede mezclar los archivos generados por el programa "HIDROLOGIA A1" con los generados por el programa "BIOMETRIA B3", por ejemplo. Si bien es posible implementar esta mezcla y generar así solo un programa captador, un solo programa selector y un solo tipo de archivos, la dificultad sería tan grande que resultaría imposible dicha implementación (en BASIC) sin perder muchas de las características del sistema, como la verificación de los datos o la posibilidad de recobrar el control del programa despues de un error de cualquier especie. Para implementar un sistema así sería necesario utilizar otro lenguaje (FASCAL o COBOL serían ideales) y conseguir una expansión de memoria para la máquina (a 128 K como mínimo).

Los programas selectores pueden aumentar de capacidad (se pueden implementar otros modos de selección) con solo agregar unas pocas líneas. Este manejo de las formas de selección no solo facilita su corrección durante la etapa de desarrollo del programa sino que facilita tanto la traducción como la adaptación del sistema a otras necesidades, pues el programador trabaja con pequeños bloques lógicos en lugar de enfrentarse al programa como conjunto; cada bloque lógico corresponde a un criterio de selección. La traducción del sistema deberá estar a cargo de la persona interesada, pues un capturista, un ingeniero en sistemas o un programador seguramente alterará los criterios de selección de manera impredecible, ya que no se encontrará familiarizado con los criterios utilizados en el laboratorio en cuestión.

Las desventajas de H.B.C., reales o percibidas fueron discutidas en el capítulo anterior. Estas desventajas son relativas, pues la primera (la sobreespecialización) en realidad resulta del cumplimiento del primer punto señalado en los propósitos del trabajo (sección I.3). La eliminación de el resto de las desventajas planteadas anteriormente dependerá del tipo de programas de proceso posterior que cada investigador decida emplear.

Desde hace varios años se ha intentado adoptar un formato universal para los archivos empleados en microcomputadoras y hacer así compatibles a todos los programas de proceso numérico; un ejemplo de estos formatos universales es el D.I.F. (Data Interchange Format). Este formato es empleado por programas graficadores como VISIPLLOT/VISITREND y por paquetes estadísticos avanzados, como DAISY. La técnica para generar (y leer) archivos D.I.F. es descrita en Miller (1982) y en Fylstra y Kling (1982). En la actualidad, H.B.C. no está conectado con programa alguno; los datos selectos son impresos en papel (con el programa IMPRESOR) y son alimentados a mano en el programa de proceso posterior elegido por el investigador. Este proceso generalmente dura poco tiempo pues en la mayoría de los casos los datos selectos no llegan a 200.

VI.3) Recomendaciones para el usuario.

Las siguientes recomendaciones son válidas para cualquier microcomputadora de 8 bits con algún dialecto de BASIC. El punto más importante de todos es el tamaño de la memoria; generalmente el mínimo es de 48 K en las máquinas antiguas y 64 K en las modernas. Para evitar la saturación de la memoria, se recomienda que los archivos tengan cuando mucho unas 30 o 40 estaciones. El número de datos debe conservarse, de ser posible, por debajo de los 4000 (aunque la configuración máxima puede almacenar hasta 6000).

Si el usuario tiene acceso a dos lectoras de disco ("drives"), se recomienda que la lectora #1 sea utilizada para programa y la #2 para el disco con los datos. Para el caso de las máquinas con una sola lectora se recomienda proteger el disco con H.B.C. y colocar otro (cuando se vaya a grabar un archivo) en su lugar; el disco con el paquete deberá ser retornado a la lectora inmediatamente después.

Se recomienda elaborar una lista de los nombres de los archivos sometidos a selección de datos para evitar repeticiones. Se recomienda además que los nombres reflejen el tipo de datos almacenados en el archivo de datos selectos ("BIOMETRIA TENELLUM MAYO 1983" por ejemplo); es importante recordar que los nombres de los archivos tienen una cantidad límite de caracteres (el valor del límite dependerá de la máquina en cuestión).

Es importante recordar que el usuario no debe hacer caso de los prefijos de los archivos, pues éstos solo le sirven a la máquina para identificar los distintos tipos de archivos disponibles; por lo tanto, cuando la máquina pregunta por el nombre de algún archivo, el usuario no deberá incluir dicho prefijo.

Se recomienda conservar cuando menos una copia de cada archivo en discos almacenados en un lugar seguro; de ser necesario estos discos deberán ser actualizados en forma regular.

Se recomienda además el anotar los nombres de los archivos maestros en donde se vá haciendo la selección de los datos para un trabajo para así localizar fácilmente el archivo fuente de un cierto dato y para evitar repeticiones por buscar en dos ocasiones en el mismo archivo.

Los usuarios de máquinas compatibles con Apple que no tienen la versión modificada del Pronto-DOS con

CATALOG-PLUS y los usuarios de máquinas no compatibles con Apple deberán modificar las líneas que generan el catálogo del disco. Estas instrucciones se encuentran en todos los programas captadores, en todos los selectores y en el impresor. De esta forma el programa correrá normalmente con cualquier D.O.S..

Es importante señalar que, en la versión empleada en el laboratorio de Limnología, el cursor se coloca automáticamente en la zona de la pantalla en donde debe entrar un cierto dato y por ello el usuario no deberá utilizar la barra espaciadora para alterar la posición del cursor.

El usuario inexperto deberá elaborar algunos archivos de prueba para familiarizarse con H.B.C. y con la máquina antes de utilizar datos reales.

Se recomienda guardar copia en papel de todos los archivos de disco (utilizando la opción 2 de cada programa captador).

VI.4) Recomendaciones para la adaptación de H.B.C. a otras microcomputadoras.

Las recomendaciones siguientes están orientadas sobre la versión en BASIC de H.B.C., disponible en el laboratorio de Limnología. Prácticamente todas las microcomputadoras comerciales tienen un dialecto de BASIC en ROM así que la mayoría de las instrucciones serán fáciles de transcribir. Los dialectos de BASIC generalmente difieren en las instrucciones que permiten el acceso a lectoras de disco, a impresoras y en las instrucciones que manejan el formato de la pantalla. En algunos casos, el dialecto no permite la existencia de arreglos alfanuméricos multidimensionales; en este caso el usuario deberá convertir los arreglos alfanuméricos en numéricos o en su defecto deberá generar tantos arreglos alfanuméricos como campos tenga su archivo. Este último punto es muy importante, aunque afortunadamente solo aparece en pocas máquinas. Otro detalle importante que debe considerar el usuario es que las instrucciones que actúan directamente sobre localidades específicas de la memoria de la máquina, como PEEK y POKE, deberán ser modificadas. En H.B.C. se usa el PEEK (222) que indica el número del error cometido y el POKE 36 que indica la columna en la que la impresora debe colocar un mensaje. Se evitó en lo posible el uso de este tipo de instrucciones en H.B.C. para facilitar su transportabilidad. El usuario interesado en transportar H.B.C. a otra máquina deberá consultar tanto a los manuales de Apple como a los manuales de la máquina en cuestión.

VI.5) Mejoras propuestas para versiones posteriores de H.B.C.

Se recomienda, para los programas captadores, un sistema de conteo para evitar algún error durante la especificación de los parámetros del archivo. Por ejemplo, si el usuario especifica que un cierto archivo contiene 5 estaciones y en realidad éste consta de 6 el error no será corregible y el proceso de captura deberá ser reiniciado. En las pruebas realizadas en el laboratorio este error nunca apareció, sin embargo la posibilidad de cometerlo existe en la versión actual.

Se recomienda optimizar los programas selectores, pues en la versión actual su análisis resulta un poco elaborado y por ello su modificación puede tomar varios días. La versión actual funciona correctamente, así que para optimizar el sistema se recomienda convertir las líneas de programa encargadas de seleccionar datos en subrutinas que serían llamadas desde una sección de menú (driver). Es importante señalar que este análisis solo tiene que ser elaborado cuando el paquete va a ser adaptado a otro ambiente distinto al del laboratorio.

Se recomienda la implementación de una rutina, para los programas selectores, que permita que el usuario especifique todos los archivos maestros en donde quiere realizar una cierta selección. En la actualidad el proceso se realiza por pasos pues al final de la selección de datos de un archivo el programa pregunta al usuario si desea continuar la selección con otro archivo. Esta opción no aceleraría el proceso de selección, pero liberaría al usuario de la necesidad de atender a la máquina mientras trabaja. Una rutina extra que debería ser considerada en la presente sugerencia debería permitir la búsqueda de datos de archivos maestros en dos lectoras de disco (el sistema actual solo busca los datos en una lectora).

Se recomienda la elaboración de una rutina que permita el encadenamiento de los archivos de datos selectos con los programas de proceso posterior más comunes.

VI.6) Proyección al futuro de H.B.C.,

El paquete H.B.C. representa la primera parte de un proyecto más ambicioso. La intención, a mediano plazo (1 o 2 años aproximadamente) es la de elaborar un sistema de procesamiento de datos que incorporará programas de proceso estadístico y de graficación. Los programas de proceso estadístico contarán tanto con las técnicas ya conocidas, por ejemplo el análisis de varianza como con las nuevas técnicas desarrolladas en los últimos años. Los programas de graficación contarán con sistemas para generar gráficas bidimensionales comunes (como gráficas de barra, histogramas y gráficas de líneas) y gráficas tridimensionales (como gráficas del perfil del fondo de la laguna estudiada o gráficas de la distribución del P.H. a cierta profundidad en toda la laguna). Si bien ya existen programas capaces de llevar a cabo algunos de estos procesos, su orientación, sus limitaciones o sus características de captura de datos los hacen parcial o totalmente incompatibles con los objetivos perseguidos en el proyecto. Con un sistema así el investigador sería capaz de realizar rápidamente un proceso que ahora toma meses y tendría más tiempo a su disposición para interpretar los resultados; además sería capaz de realizar procesos hasta ahora imposibles. La importancia de la aplicación (o diseño) de técnicas de proceso de datos es tan grande como el proceso mismo.

Otros paquetes desarrollados por investigadores han seguido un desarrollo similar, como el de Prothero (op. cit.), que comenzó con una base de datos y terminó como un sistema indispensable para la realización de ciertos tipos de trabajos.

VI.7) Proyección de la Computación en la Biología Moderna.

Existen varias ramas de investigación en la computación actual, y todas ellas pueden tener gran relevancia en el desarrollo de la Biología Moderna. Estas ramas son:

a) Las Bases de Datos. Su importancia en el laboratorio de Biología ha quedado ejemplificada tanto en el capítulo II como en el desarrollo del presente trabajo. Es importante señalar que las posibles aplicaciones de las microcomputadoras que son descritas en los apartados posteriores no son posibles de implementar sin una base de datos que facilite el acceso de los datos de interés a los programas de proceso posterior.

b) La Simulación. Las computadoras permiten elaborar modelos altamente complejos que pueden ser comparados posteriormente con la realidad. La Física de Partículas, la Ingeniería y la Aeronáutica son disciplinas que emplean la computadora para verificar modelos altamente complejos, como el impacto de un electrón con una energía superior en varios órdenes de magnitud a la generada con un acelerador moderno.

Algunos de los campos de la Biología que emplean en la actualidad modelos elaborados son: la Biología Pesquera, la Genética de Poblaciones, la Ecología y la Bioquímica (Bechgaard y Jerome, 1982).

c) La Generación de Imágenes. Este punto está muy relacionado con el anterior. La simulación gráfica es un campo nuevo, con no más de un lustro de historia, y ya rinde grandes resultados en campos como la Genética Molecular y la Bioquímica, pues es posible generar imágenes detalladas de ciertas moléculas con una actividad específica (como una enzima) y es posible comparar la geometría de la molécula, y el supuesto mecanismo de reacción de ésta, con la realidad (Dickenson, 1983; Kirkland, 1895; Ptashne et al, 1982).

Como se puede apreciar con los ejemplos mencionados anteriormente, las computadoras han conquistado un lugar importante entre el instrumental del Biólogo. La mayoría de los trabajos mencionados anteriormente han sido elaborados en máquinas grandes. Las microcomputadoras no fueron introducidas al mercado sino hasta 1970 y es por ello que no han sido explotadas al máximo.

El grado de especialización necesario para programar una máquina grande ha ido disminuyendo con el tiempo al avanzar las técnicas de programación y al desaparecer la noción de que la programación es labor de técnicos especializados. Con la introducción de las microcomputadoras esta disminución ha sido acelerada.

Por su bajo costo, su facilidad de uso y su gran potencial de procesamiento de información, la microcomputadora puede facilitar enormemente el proceso actual de los datos obtenidos durante un proyecto de investigación y además puede permitir otros procesos hasta ahora fuera del alcance del investigador.

VI.8) Resumen.

En el presente capítulo se presentan las conclusiones sobre la utilidad del paquete H.B.C. en el laboratorio de Limnología y se discute su posible papel en otros laboratorios de investigación. Además se presentan algunas sugerencias para emplear el sistema de la mejor forma posible y se discute el papel de las microcomputadoras en la Biología Moderna. Por su bajo costo, la facilidad de su uso y su gran potencial de proceso, las microcomputadoras pueden ayudar a resolver muchos problemas comunes a los laboratorios de investigación del país y pueden permitirle al investigador la realización de proyectos que, por su extrema complejidad por un lado, y la relativa escases de computadoras grandes por el otro, serían prácticamente imposibles de realizar de otra forma. Para implementar procesos complejos en cualquier máquina, el primer paso a realizar es la elaboración de una base de datos adecuada que facilite el proceso posterior de los datos.

VII. ANEXOS

VII.1) Formas codificadas.

Las formas codificadas son el resultado de varios años de experiencia en la captura de datos limnológicos. Están diseñadas para recibir la información obtenida en el campo en un formato que disminuye mucho la probabilidad de cometer un error durante la selección de los datos para un proyecto de investigación en comparación con el cuaderno de notas. Las tablas HIDROLOGIA A1 e HIDROLOGIA A2 son similares en estructura, aunque A2 permite la captura de hasta 3 variables cuyo nombre es especificado por el usuario. La tabla BIOMETRIA B3 es una modificación de B2, que capturaba únicamente datos biométricos de crustáceos. B3 permite la captura de datos de prácticamente todos los organismos estudiados en el laboratorio. La modificación fue realizada durante el desarrollo del presente trabajo.

VIII. BIBLIOGRAFIA

- Ackoff, R.L., 1974. Rediseñando el Futuro. Ed. Limusa. pp. 14-16.
- American Psychological Association Publication Manual., 1984. A.P.A., Washington. 3 rd. Ed.
- Anderson, N., 1974. A mathematical model for the growth of giant kelp. Simulation. Vol. 22.: 97-105.
- APPLE II Reference Manual. 1981. Apple Computer Inc.
- APPLESOFT II: BASIC Programming Reference Manual. 1981. Apple Computer Inc.
- Atkins, R. W. y Mazur, M. L., 1984. The DayFlo Architecture. BYTE Magazine. Vol. 9, No.11.: 155-162
- Bechgaard, K y Jerome, D., 1982. Organic Superconductors. Scientific American. Vol. 247, No. 1.: 52-61.
- Benson, J.E. 1977. The Use of Computer Simulation and Biomathematical Models for Mathematical Instruction of Undergraduate Students. Univ. of New York. School of Education, Health, Nursing and Arts Professions. Ph. D. Thesis.
- Birman, K. P. y Joskowics, G., 1982. A System for Clinical Data-Base Applications. Computers and Biomedical Research. Vol. 15.: 598-604.
- Bojorquez, C.L. 1975. Introduction to Systems Analysis in Science Courses. Progress in Cybernetics and Systems Research, Vol. 2. pp. 213-221.
- Bojorquez, C. L., Tribe, M. A. y Galvan, S., 1982. Simulations in mathematical ecology: population dynamics. Univ. of Sussex, School of Biological Sciences. Students Guide.
- Bond, G., 1984. A Database Catalog. BYTE Magazine. Vol. 9, No. 11.: 227-238.
- Cook, R. y Brandon, J., 1984. Information Management. BYTE Magazine. Vol. 9, No. 11.: 177-198.
- Costales, B., 1985. Catalog Plus. NIBBLE Magazine. Vol. 6, No. 1.: 92-95.
- Curts, J.B. 1984. Introducción al Estudio de Análisis de Residuos en Biología, Biótica. Vol. 9. pp. 271-278.
- Curts, J.B. 1985a. Teaching the Simple Linear Regression Model Using an Interactive Microcomputer Graphics Software Package. Ph. D. Thesis. Univ. of Wisconsin, Madison.
- Curts, J.B. 1985b. Aplicación de la Estadística a Problemas de Contaminación. Conferencia dictada en la Univ. Pedagógica Nacional. Enero 21-26. Memorias en Prensa.
- Curts, J.B. y Saal, A. 1979. Biomatemáticas: Una Introducción. Escuela Nacional de Estudios Profesionales Iztacala.
- Dam, A. V., 1984. Computer Software for Graphics. Scientific American. Vol. 251, No. 3.: 102-113.
- Davies, R. G., 1971. Computer programming in quantitative biology. Academic. Press, N.Y.
- Deininger, R. A., 1982. An Interactive Data Management System for River Water Quality Data. Water Research. Vol. 16.: 31-36.
- Dickerson, R. E., 1982. The DNA Helix and How it is Read. Scientific American. Vol. 249, No. 6.: 94-111.

- Dixon, W.J.; Brown, M.B.; Engelman, L.; Frane, J.W.; Hill, M.A.; Jennrich, R.I. y Toporek, J.D. 1985. B.M.D.P. Statistical Software Manual. Univ. of California at Berkeley Press.
- Dominguez, E.L. 1984. Temas para un Futuro Biólogo. U.N.A.M. pp. 551,560.
- Einstein, A. e Infeld, L. 1938. The Evolution of Physics. Ed. Simon and Schuster.
- Farina, M. V., 1970. Flowcharting. Prentice-Hall. 1 era. Ed. pp 119.
- Fee, E. J., 1969. A numerical model for the estimation of photosynthetic production, integrated over time and depth, in natural waters. Limnol. Oceanogr. Vol. 14.: 906-911.
- Feutcher, C.A., 1984. TRS-80 versus a giant brain. Sky & Telescope Magazine. Vol. 67 No. 4.: 358-359.
- Fields, C. 1973. Introducción a los Computadores. Alianza Editorial.
- Flores, I., 1981. Data Base Architecture. Van Nostrand-Reinhold. 1 era. Ed. 396 pp.
- Flystra, D. y Kling, B., 1982. VISICALC. Software Arts Technical Note SATN-18. Personal Software Inc. 1 era. Ed. 350 pp.
- Garfinkel, D., 1962. Digital computer simulation of ecological systems. Nature. Vol. 194.: 856-857.
- Garfinkel, D. y Sack, R., 1964. Digital computer simulation of an ecological system based on a modified mass action law. Ecology. Vol. 45.: 502-507.
- Garfinkel, D., Garfinkel, L., Pring, M., Green, S. B. y Chance, B., 1970. Computer applications to biochemical kinetics. Ann. Rev. Biochem. Vol. 39.: 473-498.
- George, F.H. 1968. Cibernetica y Biología. Ed. Alhambra, Espana.
- Gold, H.J. 1977. Mathematical Modeling of Biological Systems. An Introductory Guidebook. Wiley & Sons Interscience Series.
- Guzmán, A.M., Kensler, C. y Granados A.B., 1978. Los Langostinos del género *Macrobrachium* en Mexico y sus posibilidades de cultivo. II. Simp. Asoc. Latinoamer. de Acuicultura, México.
- Guzmán, A.M., y Gonzalez, L.G., 1979. Análisis del Estado del Arte de la Acuicultura en México. I Simposio Internacional de Educación y Organización Pesqueras. Cancún Q.R. México.
- Guzmán, A.M., et al., 1979. Análisis del Estado del Arte de la Acuicultura en México. Depto. de Pesca, Dir. Gral. de Acuicul., Inst. de Cienc. del Mar y Limnol., Univ. Nal. Autón. de México., México, 265 pp.
- Guzmán, A.M., Cabrera, J. y Kensler, C., 1979. Notes on *Macrobrachium* species in Mexico, in Hanson, J.A. y H.L. Goodwin (Edit.) Shrimp and Prawn Farming in the Western Hemisphere. Dowden Hutchinson & Ross, Inc. Pennsylvania, 347 pp.
- Guzmán, A.M., Rojas, J.L. y Gonzalez, L.G., 1982. Ciclo anual de maduración y reproducción del chacal *Macrobrachium tenellum* (Decapoda, Palaemonidae) y su relación con factores ambientales en las lagunas costeras de Mitla y Tres Palos, Guerrero México. An. Inst. Cienc. del Mar y Limnol., Univ. Nal. Autón. de México. 9 (1): 67-80 pp.
- Guzmán, A.M., 1983. Sistematización de la Información Sobre Investigaciones de los Recursos Acuícolas Nacionales. Sria. de Rec. Hidrául. Plan Nal. Hidrául. Seminario sobre el Desarrollo de los Recursos Hidráulicos. Mexico, D.F.

- Held, G.D., Stonebraker, M.R. y Wong, E., 1975. INGRES: A relational database system. Proceedings of the AFIPS 1975 Conference. Vol. 44.: 409-416.
- Jacob, F., 1970. La Lógica de lo Viviente. Ed. LAIA.
- Jones, R. y Hall, W.B., 1973. A simulation model for studying the population dynamics of some fish species. M.S. Bartlett y R.W. Hiorns (editores), The Mathematical Theory of the Dynamics of Biological Populations. London. Academic Press. pp. 35-59.
- Justice, K. E., 1968. The use of Computer Modeling in Biology Laboratory Instruction. Univ. of California, Irvine, Bulletin.
- Kay, A., 1984. Computer Software. Scientific American. Vol. 251, No. 3.: 40-47.
- Kidd, N. A. C., 1979. Simulation of population processes with a programmable pocket calculator. Journal of Biological Education. Vol. 13.: 284-290.
- Kirkland, E. J., 1985. Viewing Molecules with the MacIntosh. BYTE Magazine. Vol. 10, No. 2.: 251-259.
- Krajewski, R., 1984. Database Types. BYTE Magazine. Vol. 9, No. 11.: 137-142.
- Kruglinski, D., 1983. Data Base Management Systems. A Guide to Microcomputer Software. Osborne/McGraw-Hill. 1 era. Ed. 250 pp.
- Ledley, R.S. 1965. Use of Computers in Biology and Medicine. McGraw-Hill, New York.
- Lee, J. D. y Lee, T. D., 1982. Statistics and numerical methods in BASIC for biologists. Van Nostrand/Reinhold.
- Lenat, D. B., 1984. Computer Software for Intelligent Systems. Scientific American. Vol. 251, No. 3.: 152-160.
- Lesk, M., 1984. Computer Software for Information Management. Scientific American. Vol. 251, No. 3.: 114-124.
- Lesko, M., 1984. Low-Cost On-Line Databases. BYTE Magazine. Vol. 9, No. 11.: 167-174.
- Lotka, A.J. 1956. Elements of Mathematical Biology. Dover Publ. Corp.
- Macchia, D. D., 1982. A Program in BASIC for Calculation of Tissue Extracellular Space and Ion Distributions in vivo Using Morphometric and/or Extracellular Tracer Methods. Computers and Biomedical Research. Vol. 15.: 592-596.
- Martin, J., 1977. Organización de las Bases de Datos. Ed. Prentice/Hall Internacional. 1 era. Ed. (1977) en español de la 1 era. Ed. en Inglés (1977). 544 pp.
- Mc Millen, J.D. y Esch, H.E. 1984. Microcomputers for Laboratory Data Collection. The American Biology Teacher. Vol. 46. No. 3. pp. 157-161.
- Miller, D., 1982. Apple Files. Reston Publ. Company Inc. (Prentice Hall). 1 era. Ed. 450 pp.
- Navarro, F.P. 1982. La Dirección por Sistemas. Ed. Limusa, Mexico. pp. 20.
- Negrete, M., 1977. Estudio sobre la fecundidad de Macrobrachium tenellum (langostino) (Smith, 1871) (Decapoda: Palaemonidae) en la laguna de Tres Palos. Tesis de Licenciatura. (Biología). Fac. de Ciencias. Univ. Nal. Autónoma de México. pp 73.

- Núñez, M., 1982. Estudio Ictiológico de la Laguna de Cuyutlán, Colima, México. Características Ambientales y Poblacionales. Tesis de Maestría. Colegio de Ciencias y Humanidades. Unidad Académica de los ciclos Profesionales y de Posgrado. Inst. de Ciencias del Mar y Limnol. Univ. Nal. Autón. de México. pp 83.
- Norman, N. H. et al., 1975. SPSS Statistical Package for the Social Sciences. McGraw-Hill Book Company, 2 da. Ed. pp. 1-215.
- Orris, J.B. 1982. The Role of Microcomputer in Statistical Computing. Teaching Statistics and Statistical Consulting. Rustagi, J. y Wolfe (editores). Academic Press. pp. 485-495.
- Pauli, D., 1983. Tropical Stock Assessment Packages for Programmable Calculators and Microcomputers. ICLARM Newsletter, July, 1983.: 10-13.
- Peters, L. J., 1981. Software Design: Methods and Techniques. Yourdon Press, U.S.A. 234 pp.
- Poincare, H. 1981 (ed.). Filosofía de la Ciencia. CONACYT.
- Poincare, H. 1964. El Valor de la Ciencia. Colección Austral. Ed. Espasa-Calpe.
- Prothero, J. y Prothero, J., 1982. Three-Dimensional Reconstruction from Serial Sections. A Portable Microcomputer-Based Software Package in FORTRAN. Computers and Biomedical Research. Vol. 15.: 598-604.
- Ptashne, M., Johnson, A. D. y Pabo, C. O. A Genetic Switch in a Bacterial Virus. Scientific American. Vol. 247, No. 5.: 128-141.
- Randall, J. E., 1980. Microcomputers and Physiological Simulation. Reading. Addison-Wesley Publishing Co. 235 pp.
- Ribes, B., 1978. Biología y Ética. U.N.E.S.C.O. (ed.). Colección Actual.
- Richter, O. y Betz, A., 1976. Simulation of biochemical pathways and its application to biology and medicine. Levin S. (editor). Lecture Notes in Biomathematics. Vol. 11.: 181-197.
- Rowley, R. M. et al, 1982. A PROPHET-Based Laboratory Data Entry and data-Base Management System for the Environmental Mutagenesis Test Development Program. Computers and Biomedical Research. Vol. 15.: 544-553.
- Ryan, T.A.; Joiner, B.L. y Ryan, B.F. 1981. MINITAB Reference Manual. MINITAB Project, Statistics Laboratory, The Pennsylvania State Univ.
- SAS User's Guide: Statistics. 1982. Statistical Analysis System Institute Inc.
- Schrodinger, E. 1962. Qué es una ley de la Naturaleza?. Ed. Fondo de Cultura Económica.
- Shapiro, E., 1984. Data Structures and Algorithms. Scientific American. Vol. 251, No. 3: 48-57.
- Smythe, R. y Lovatt, K. F., 1979. Applications of the computer in biology teaching: computer assisted and computer managed learning. Journal of Biological Education. Vol. 13.: 207-220.
- Spain, J. D., 1973. The introduction of biological mensuration techniques through simulation. Proc. of Conf. on Computers in Undergraduate Curricula. Vol. 3.: 80-84.
- Spain, J. D., 1981. Teaching Simulation Techniques with Microcomputers. Proc. Nat. Ed. Computer Conf. Vol. 2.: 225-227.

- Spain, J. D., 1983. Microcomputers for teaching simulation to undergraduate students. Simulation and Games, Vol. 14.: 87-94.
- Spain, J. D., 1984. BASIC Microcomputer Models in Biology, Addison-Wesley Publ. Company, pp. 1-11.
- The D.O.S. Manual, Disk Operating System, 1981, Apple Computer Inc.
- Waddington, C.H. y otros. 1976. Hacia una Biología Teórica, Alianza Editorial, Madrid, pp. 17-53.
- Wirth, N., 1984. Data Structures and Algorithms. Scientific American, Vol. 251, No. 3.: 48-57.
- Wolfram, S., 1984. Computer Software in Science and Mathematics. Scientific American, Vol. 251, No. 3.: 140-151.