

Capítulo 3.

Modelo de bases de datos botánicas

A lo largo de este capítulo se presentará un breve análisis de los modelos de bases de datos botánicas, una comparación entre BIOTICA versión 1 y BIOTICA versión 2 y se finalizará con tablas comparativas entre todos estos modelos. Cabe mencionar que no se analizó en su totalidad cada modelo sino que simplemente se tomó la parte de cada modelo que para fines de la tesis era relevante analizar.

3.1 Modelo E-R de bases de datos botánicas

Al inicio de este documento se planteó que para la construcción de una biblioteca digital era necesaria la creación de grandes depósitos de información.

Crear una base de datos implica que se partió de un modelo de datos de la misma. En este caso nuestro modelo de datos se enfoca a colecciones botánicas.

En los siguientes apartados se presenta el análisis de algunos modelos de datos para la colecciones botánicas existentes hoy en día. Estos son comparados con el modelo de datos BIOTICA versión 1, del que se tomó un subesquema aplicándolo al modelo de datos de Xanath. De ahí la importancia de analizar y comparar con otros modelos para comprender cuales aspectos en cuanto a datos fueron importantes durante la elaboración del esquema de base de datos de Xanath.

Antes de dar inicio con el análisis explicaremos cual es la diferencia que existe entre un modelo de bases de datos botánicas aplicado a una flora y el que se aplica a un herbario. Cuando se habla de información de una flora se refiere a que el nivel de abstracción de la información botánica es más alto que el manejado en un herbario, es decir, al hablar de una flora se hace referencia a toda la información disponible de la clasificación taxonómica botánica existente. La clasificación taxonómica se encuentra en niveles como son: familia, género, especie, sin llegar a nivel de ejemplares. En cambio cuando se habla de la información que se manipula en un herbario, se refiere a la información disponible de cada ejemplar en específico, el cual se encuentre dentro de sus colecciones.

Como consecuencia, una diferencia entre el manejo de información botánica de un herbario y la de una flora es que, para un herbario es importante almacenar información sobre datos que se refieren al día en que se colectó un ejemplar, los datos del recolector, las coordenadas del sitio de colecta mientras que estos datos no forman parte de la información de una flora.

Un ejemplo que muestra las diferencias entre la información de un herbario y la de la flora es: al hablar de la distribución geográfica de la flora es un agregado de todos los especímenes conocidos; mientras que esta información para un herbario hace referencia al lugar en que se recolectó un ejemplar.

Una vez aclaradas las diferencias en el manejo de información de estos modelos botánicos, establecemos que el criterio de evaluación fue principalmente considerar aquellos datos que proporcionan información sobre los especímenes

botánicos, referente a:

como hacer la consulta de un ejemplar dado cierto nivel taxonómico.

como se manejan la imágenes de los especímenes

cómo se manejan las ilustraciones de los especímenes

cómo poder hacer consultas de un espécimen basándose en sus características

cómo obtener la distribución geográfica de un espécimen, según sea el nivel de detalle al que se aplique el sistema.

3.1.1 FDL.

La biblioteca digital florística (FDL, Floristic Digital Library) es un proyecto de investigación y desarrollo, llevando a cabo por el jardín botánico de Missouri (MBG) en colaboración con otras instituciones y con la participación del Laboratorio de Tecnologías Interactivas y Cooperativas (ICT) de la Universidad de las Américas-Puebla. En este proyecto se contempla la creación de un depósito de los documentos sobre información botánica y que se encuentran accesibles vía WWW.

Una parte importante de su construcción es la creación de una base de datos, la cual funciona con el manejador de base de datos Informix Universal Server.

Con respecto a la parte del modelo de datos que en ella se maneja, la figura 3.1 muestra una versión simplificada de su diagrama entidad-relación. Este modelo de datos botánicos se encuentra enfocado a una flora.

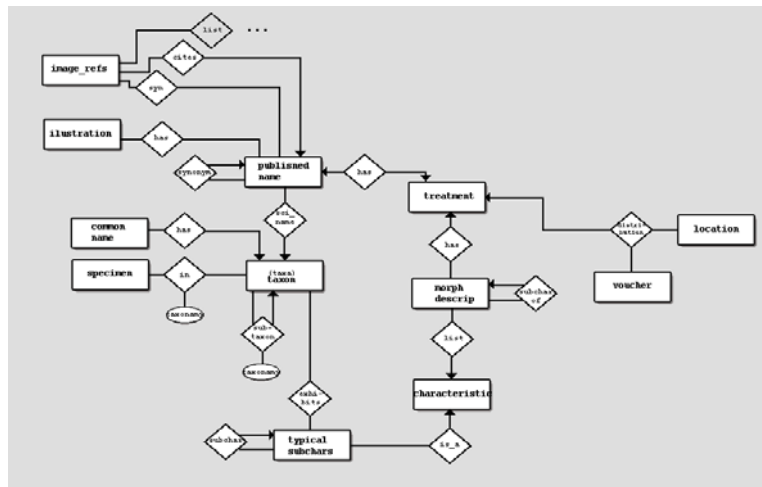


Figura 3.1: Diagrama Entidad-Relación de FDL (se omiten atributos por simplicidad)

En el modelo para el conjunto de entidades *taxón* podemos observar que existe recursividad, esta se aprecia al tener el conjunto de relaciones *subtaxón* la cual se vuelve cíclica. Con esta representación lo que se establece es que un ejemplar dado tiene un rango taxonómico pero este rango puede ser desde el más infraespecífico hasta el más general. Utilizar este modelo de datos facilita la consulta por información taxonómica de la flora. Es decir puede un nivel taxonómico encontrarse arriba de otros niveles y encontrarse debajo de otros niveles, por ello para obtener la información del nivel que se desea es necesario recorrer en forma recursiva el conjunto de relaciones *subtaxon*.

En cuanto a ilustraciones, con el conjunto de entidades *illustration* muestra que en el modelo se manejan éstas. Esto establece que a partir de un taxón

publicado él deberá poseer o tener una ilustración o más que lo representan gráficamente. De la misma forma ocurre con la tabla o conjunto de entidades *image_refs* el cual se refiere a las imágenes que representan el nivel taxonómico. Es necesario mencionar que tanto las imágenes como las ilustraciones son introducidas directamente a la base de datos.

Cabe mencionar que generalmente las ilustraciones o imágenes son propias del espécimen tipo, ejemplar más representativo de la especie o nivel taxonómico al que pertenece.

Las características de un nivel taxonómico son aquellas que se refieren a un atributo del nivel con un valor dado. En el caso de las plantas una de sus características puede a la vez tener otras características y así sucesivamente. La forma como se visualiza esta descripción en el modelo entidad relación de FDL en el conjunto de entidades *taxa* muestra mediante el conjunto de relaciones *exhibits* el enlace hacia el conjunto de entidades *typical_subchars*, éstas a su vez pueden tener otras subcaracterísticas por lo que la llamada al conjunto de relaciones *subchar* se vuelve recursiva. Además un *typical_subchars* puede ser una *characteristic* la cual a su vez puede encontrarse en la descripción morfológica del nivel taxonómico en el conjunto de entidades *morp_descrip* dentro de esa descripción puede a la vez haber otras subcaracterísticas de esa descripción, esto se representa con el conjunto de relaciones *subchars_of* la cual por esa razón también es recursiva. La distribución se puede obtener a partir del conjunto de relaciones *distribution* la cual posee como llave primaria el "id_taxon" llave de la tabla "taxón", y la abreviaron del nombre de la localización en donde se ubica la misma.

Como se puede apreciar este modelo de datos de la base de datos botánicas posee los cinco criterios de evaluación que en este capítulo se analizan en cada modelo propuesto.

3.1.2. SMASCH

SMASCH, sistema para el manejo de especies del herbario de California, es un trabajo para el desarrollo y la implementación de bases de datos distribuidas que pueden almacenar y manejar el acceso a la información sobre plantas vasculares recolectadas en California. Con esta base de datos se busca reforzar el conocimiento sobre el estado de los recursos de las plantas vasculares y ayudar a la preservación de las mismas especies.

El modelo de datos SMASCH es aplicable a colecciones de tipo herbario. el diagrama entidad relación que muestra una parte del modelo completo de SMASCH es el siguiente, en este diagrama, al igual que en el de FDL apartado 3.1.1, se escogen aquellos datos de relevancia en este análisis. Cabe mencionar que sobre este modelo de datos existe muy poca información disponible por ello el análisis que aquí se presenta no es tan detallado (ver <http://ucjeps.herb.berkeley.edu/smasch.html>).

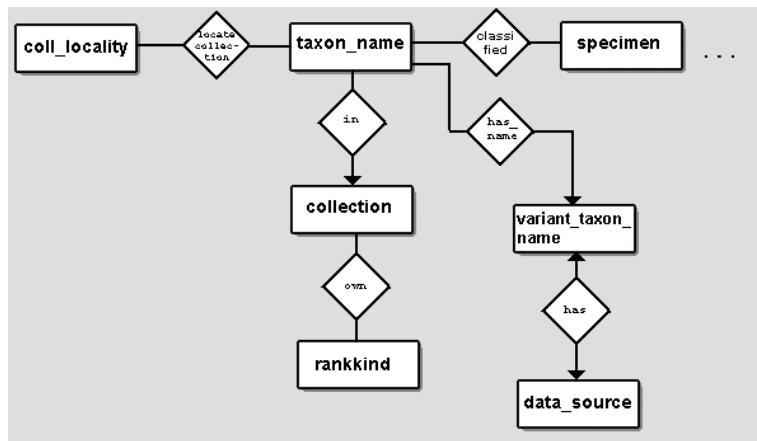


Figura 3.2 : Diagrama entidad-relación SMASCH (se omiten atributos por simplicidad)

La información sobre el taxón es obtenida fácilmente a través del conjunto de entidades *taxon_name*. La búsqueda por niveles taxonómicos de un ejemplar no se hace en forma recursiva. Para buscar por nivel taxonómico se logra a través del conjunto de entidades *rankkind*, sin embargo antes de ello es necesario saber en que colección se encuentra el ejemplar y para ello se busca en el conjunto de entidades *collection*. En caso de que se quiera acceder directamente la información del ejemplar se hace la consulta en el conjunto de entidades *specimen*.

Para el caso de la distribución de un espécimen o de cualquier nivel taxonómico en un nivel dado, se tiene que en el conjunto de entidades *coll_locality* se puede saber el nombre del país, estado, condado en el que se puede localizar el ejemplar. Los datos como imágenes son manejados como recursos de datos por lo que con el "id_taxon" se accesa directamente el conjunto de entidades *data_source*. En este caso no es muy claro si la imágenes se insertan directamente a la base de datos o son solamente referencias.

Aquí no se manejan ilustraciones, ni tampoco se consideran las características de un ejemplar como tal.

Este modelo de datos se implementó utilizando Sybase como manejador de base de datos.

3.1.3 BIOTICA

En el capítulo 1 se dijo que la CONABIO en su preocupación por la preservación del conocimiento de la biodiversidad estableció REMIB. La CONABIO creó también un estándar en el manejo de información sobre la biodiversidad denominado BIOTICA para el registro y manejo de datos curatoriales, nomenclaturales y bibliográficos, entre otros. Al referirnos a BIOTICA se habla tanto del sistema de información como del modelo de su base de datos que se maneja internamente en ese sistema. Es decir BIOTICA se compone de modelo de datos y un sistema de información. Nuestro enfoque se establece en el modelo de datos de BIOTICA y como descripción general del mismo tenemos: modelo de datos relacional para el manejo de información sobre biodiversidad, es decir este modelo maneja información referente tanto a colecciones botánicas como zoológicas.

Del modelo de datos de BIOTICA existen dos versiones hoy en día: la versión 1, la cual es utilizada por el Herbario XAL, y la versión 2 que es la más reciente. Ya que ambos modelos son de importancia para el análisis de modelos de bases de datos sobre biodiversidad, a cose presenta una comparación entre ambos modelos.

Cabe mencionar que al pertenecer el herbario XAL a la fase 1 de REMIB ellos utilizan BIOTICA 1, por ello, se hará un mayor énfasis en el mismo.

La parte del modelo entidad relación que formo parte del enfoque principal en BIOTICA se muestra en la figura 3.3, mientras que el de BIOTICA versión 2 se puede apreciar en la figura 3.4.

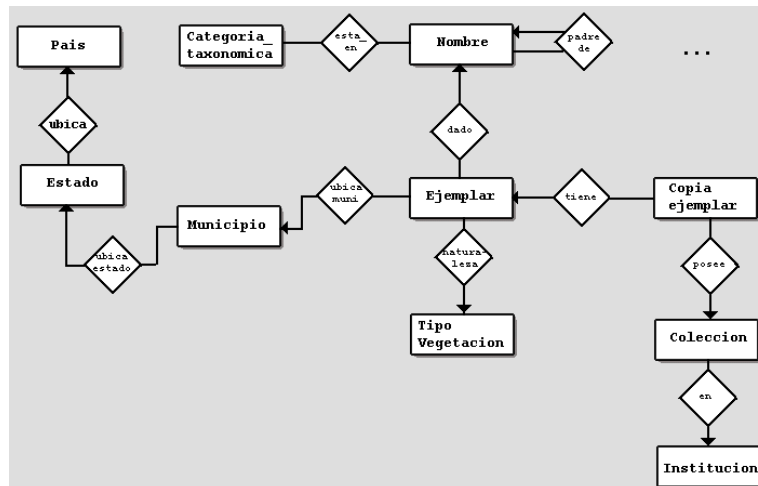


Figura 3.3: Submodelo de BIOTICA versión 1 (se omiten atributos por simplicidad)

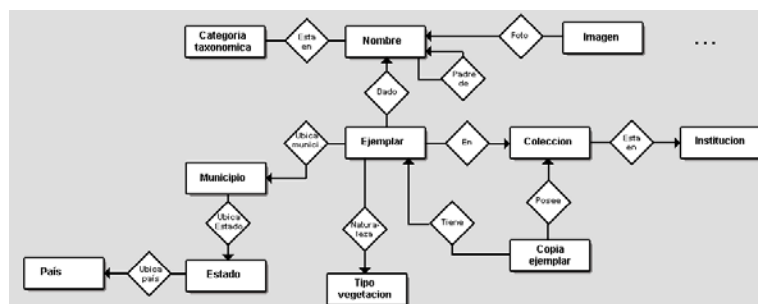


Figura 3.4: Submodelo de BIOTICA versión 2 (se omiten atributos por simplicidad)

De manera general tenemos que BIOTICA es un modelo de datos de colección botánicas y zoológicas, por ello la cantidad de atributos que manejan es un tanto extensa. Sin embargo en lugar de convertirse esta característica en una ventaja podría convertirse en desventaja. La razón de esto, radica en que existen muchos atributos que se refieren únicamente a colecciones botánicas o zoológicas.

A la parte del modelo BIOTICA que estamos analizando se le está dando un enfoque botánico haciendo de lado al zoológico. Esto se debe a que el objetivo de este capítulo es analizar los modelos de bases de datos botánicas.

Basándonos en los criterios de evaluación que al principio del capítulo se plantearon, tenemos:

En BIOTICA tanto versión 1 como 2, para obtener información de ejemplares a cierto nivel taxonómico, se accede al conjunto de entidades *ejemplar*, llevándonos éste al conjunto de entidades *nombre*. En este conjunto de entidades es donde se maneja la recursividad, a través del conjunto de relaciones *padre de*, para obtener el nivel taxonómico del mismo ejemplar, después de ello se accede al conjunto de entidades *Categoría taxonomica*, que devuelve el nombre del nivel. En ambas versiones de BIOTICA el tipo de información de distribución geográfica realmente sólo se representa en relación al lugar que se colectan los ejemplares y no en función de los lugares en que

existen dichos ejemplares. Una ventaja en la forma como manejan la información de ubicación del lugar de colecta es la distribución se da a nivel de municipio, estado, país.

En cuanto a características de ejemplares en ambos modelos no se manejan como tal, sólo existe un conjunto de entidades *Tipo_vegetación* en el que la descripción de la vegetación habla de las características de cómo es ésta. Este modelo como tal no muestra un conjunto de entidades que proporcione las descripciones morfológicas o las características de los ejemplares, tal vez una razón se fundamente en el hecho de que en el modelo se manipula información de colecciones botánicas y zoológicas. Si existieran este tipo de tablas o entidades haría demasiado complejo el modelo ya que las características de un ejemplar botánico son completamente diferentes de las de uno zoológico.

En el manejo de imágenes e ilustraciones, se tiene que sólo BIOTICA versión 2 ya hace uso de imágenes, es decir BIOTICA 1 carece de este conjunto de entidades, en el caso de ilustraciones ninguno de los dos modelos manejan este tipo de información.

Es importante mencionar que en el modelo BIOTICA 2 la forma como accesan las imágenes en su base de datos es a través de una ruta, es decir el conjunto de entidades *imagen* tiene un atributo denominado "ruta_imagen" . Este atributo indica que la consulta de imágenes a su base de datos la hacen recuperando las rutas o URLs que se encuentran almacenados en este campo de la tabla imagen. El manejar las imágenes de esta forma tiene sus ventajas y desventajas.

Las desventajas son que se pueden generar inconsistencias en la información de la base de datos si algunas o de estas imágenes fueran borradas del disco en que han sido almacenadas en y que en su base de datos no han sido borradas las rutas.

Las ventajas son que al no almacenar las imágenes en la base de datos el espacio necesario para almacenar esta información es mínimo y en caso contrario, es decir si se almacenaran las imágenes en la base de datos habría que reservar un mayor espacio.

En cuanto al manejador de base de datos que utilizan es MS-ACCESS.

De manera general el análisis de BIOTICA se hace aplicando los criterios presentados al inicio del mismo capítulo.

3.1.4. Diferencias entre BIOTICA versión 1 y versión 2.

En realidad son pocas las diferencia que existen entre ambos modelos. Las figuras 3.5 y 3.6 únicamente muestran aquellas diferencias que son visibles en la región del modelo BIOTICA que ha sido tomado como análisis en este capítulo.

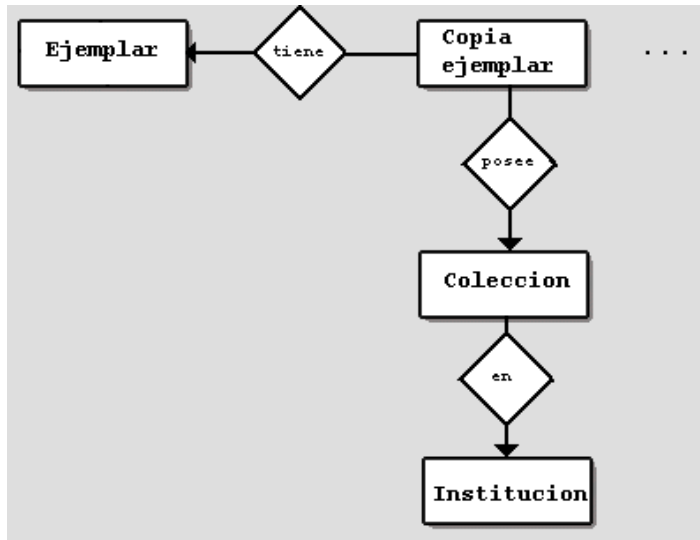


Figura 3.5: Región del submodelo de BIOTICA versión 1 (se omiten atributos por simplicidad)

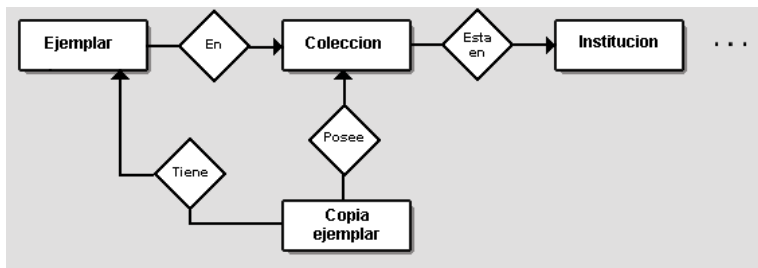


Figura 3.6: Región del submodelo BIOTICA versión 2 (se omiten atributos por simplicidad)

En la figura 3.7 podemos apreciar que el conjunto de entidades *Nombre* y el conjunto de relaciones *padre de* se refieren a la información sobre copias de ejemplares botánicos existentes en otras instituciones.

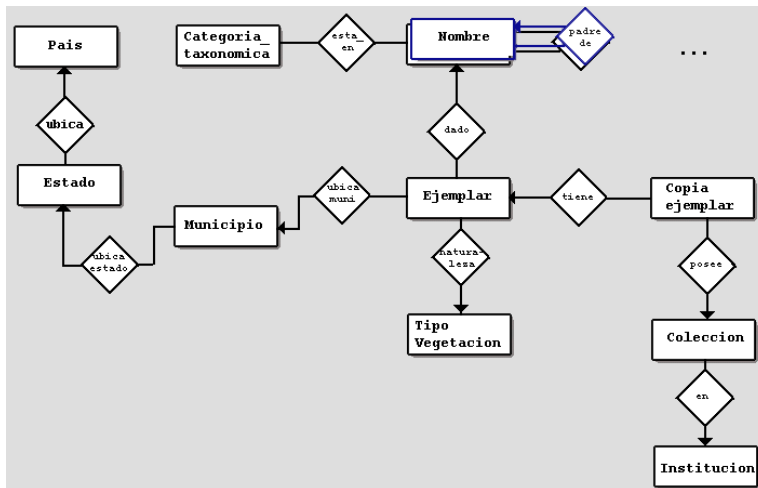


Figura 3.7: Parte del diagrama E-R de BIOTICA versión 1.

En el modelo de BIOTICA versión 2 el manejo de imágenes se establece por el conjunto de relaciones *Foto* que existe entre los conjuntos de entidades *Nombre* e *Imagen* (ver figura 3.8).

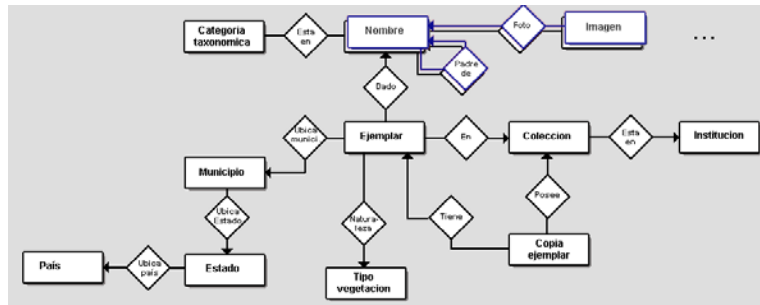


Figura 3.8 : Parte del diagrama E-R de BIOTICA 2, en donde se manejan imágenes.

3.2 Tabla comparativa por criterios de evaluación

Modelo de datos	Nivel Taxonómico	Imágenes	Ilustraciones	Características	Distribución
FDL	Consultas en forma recursiva, no incluyen ejemplares	Se almacenan en la base de datos.	Se almacenan en la base de datos.	Se manejan con recursividad al momento de hacer la consulta de ellas.	La información representa la distribución geográfica de los especímenes.
SMASCH	No manejan recursividad en la consulta por niveles taxonómicos; lo que utilizan es una tabla que presenta el atributo de rango por colección.	Se almacenan en su base de datos.	Carecen de ellas.	No cuentan con esa información.	Dan la distribución de los ejemplares existentes en la flora de California.
BIOTICA 1	Consultas en forma recursiva.	No manejan ese tipo de información.	Carecen de esa información.	Esa información no se maneja en forma independiente dentro de una sola tabla, la manipulan por breves descripciones.	La distribución la presentan de acuerdo a los lugares de colecta del ejemplar
BIOTICA 2	Consultas en forma recursiva.	Utilizan una ruta de acceso a las imágenes, es decir éstas no son almacenadas en su base de datos.	No manejan ese tipo de información.	Lo mismo que en BIOTICA versión 1	De la misma forma que en BIOTICA versión 1

3.3 Comparación técnica entre las implementaciones de los modelos de datos

Modelo de datos	Colecciones de ejemplares	Manejador de base de datos	Flora
FDL	El modelo de FDL tiene un enfoque hacia la flora.	Informix Universal Server.	Flora de Norteamérica, Flora de China y Flora de Mesoamérica.
SMASCH	Colección botánica.	Sybase.	Flora de California.
BIOTICA 1	Colección botánica/ Colección Zoológica	MS-ACCESS	Flora y Fauna de México.
BIOTICA 2	Colección botánica/ Colección Zoológica	MS-ACCESS	Flora y Fauna de México.

índice resumen 1 2 3 4 5 A B referencias

Gómez Hernández, M. C. 1998. *Xanath: Componentes Visuales para la Biblioteca Digital de un Herbario*. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Diciembre.
Derechos Reservados © 1998, Universidad de las Américas-Puebla.