

# Le modèle de représentation et de gestion hypertexte des concepts d'un domaine dans le système *CoDB-Web*

Pour élaborer des bases de concepts (BdC), nous adoptons une approche s'appuyant sur les concepts et techniques de la terminologie. Une BdC comporte trois niveaux d'abstraction : informations terminologiques, caractère définitoires et représentation. Sa mise en place est basée sur l'utilisation corrélée d'un SGBD relationnel et d'un système hypertexte. Pour son implémentation, nous avons développé le système *CoDB-Web* qui assure la gestion et l'exploitation hypertextuelle des concepts via le système hypermédia distribué *World Wide Web*. Une instanciation de BdC, *MedTrad*, a été réalisée pour capitaliser la connaissance africaine sur les plantes médicinales.

Termes clés :  
base de concepts ; terminologie ;  
information terminologique ;  
caractère définitoire ; hyperdocument ;  
médecine traditionnelle.

## 1 Introduction

**L**e projet « base de connaissance généralisée » sur lequel travaille notre équipe Ingénierie des bases de connaissances généralisées (ICOG) est né du constat de la convergence de trois domaines : bases de données (BD), bases de connaissances et hypertextes. Dans le cadre de ce projet, Nathalie Forest (1994) a proposé un modèle d'hyperdocument expert. La construction d'un tel hyperdocument expert nécessite une méthode et un outil d'aide à la structuration et à la représentation des concepts du domaine que l'on veut modéliser par une application hypertexte. Des modèles de structuration de concepts ont été proposés dans d'autres disciplines telles que : la terminotique avec D. Gouaderec (1992, 67) pour la description des concepts à l'aide de fiches terminologiques, l'ontologie avec C. Garcia (1996, 95) dans le sens restreint d'ensemble des types atomiques de concepts (« le vocabulaire de base »), les BD avec Michel Bonjour (1994, 263) pour l'intégration de schémas de bases de données, etc. Notre objectif est donc la conception d'un modèle de BdC (*cf.* M. Brou 1997) générique et réutilisable afin de fournir à toutes ces disciplines dont la structuration des concepts est un élément important, une méthode et un outil d'aide à la représentation et à la gestion des concepts d'un domaine donné. « *Nous définissons une BdC comme étant une base de données de concepts qui utilise*

*des facilités hypertextuelles pour la gestion et l'exploitation des concepts d'un domaine* ».

Dans cet article, nous proposons un modèle conceptuel pour la représentation et la gestion des concepts d'un domaine. Dans la partie 2, nous situons la notion de BdC par rapport aux travaux existants. Puis, dans la partie 3, nous décrivons le modèle de la BdC, en insistant plus particulièrement sur ses trois niveaux d'abstraction, sur son modèle hypertexte et sur son exploitation. La partie 4 aborde l'implémentation de la BdC en présentant le système *CoDB-Web* qui est le noyau d'intégration entre les différents composants de la BdC. À la partie 5, nous présentons une instanciation de la BdC à travers l'application *Medtrad* qui est une BdC sur la médecine traditionnelle africaine. La conclusion fait le bilan du travail que nous avons réalisé.

## 2 Critiques de l'existant

Le but poursuivi par la définition de la notion de BdC est à rapprocher avec celui des bases de connaissances terminologiques (BCT), à savoir l'utilisation de la terminologie dans les systèmes à base de connaissances (SBC), notamment dans l'acquisition de connaissance. Tout comme la BCT, la BdC est différente d'une base de données terminologiques (BDT). En effet, une BDT contient uniquement des termes associés à quelques informations linguistiques, spécifiques à un domaine donné. Comme le fait remarquer Nathalie

Aussenac (1995), ces BDT ne conviennent pas pour les SBC à cause de la pauvreté de la description conceptuelle, limitée à la définition des concepts donnée en langage naturel.

Une BCT intègre à la fois une base terminologique conceptuellement et sémantiquement structurée et une base de connaissance. Elles contiennent des informations conceptuelles que peuvent utiliser les cognitiens pour la construction du modèle conceptuel d'un SBC. De nombreux outils ont été développés dans le cadre des BCT. Parmi ces outils on peut citer :

- *Lexter*, Bourigault (1994) : pour l'analyse d'un corpus de texte afin de dériver une liste de termes candidats;
- *Terminae*, Brigitte Biebow (1997) : basé sur l'outil d'analyse Lexter, il permet de construire des fiches terminologiques, de garder des traces des choix de modélisation et de gérer une ontologie;
- *Macao* (Aussenac 1988) : la modélisation conceptuelle du domaine dans Macao est à rapprocher d'un réseau de concept de BDC; de plus, la méthode préconise une analyse de textes en amont pour y repérer des termes utilisés ensuite pour définir les concepts du domaine;
- *Code* (D. Skuce 1993) : est un outil complet de gestion de la connaissance. En plus de l'analyse des textes, il permet de créer un lexique de termes, de stocker, rechercher et extraire de la connaissance conceptuelle dans ces textes.

L'outil que nous avons développé pour construire une BdC et gérer efficacement les termes et les concepts d'un domaine s'appelle *CoDB-Web*. Il s'appuie à la fois sur les bases de données (qui assurent la gestion et la recherche d'informations) et les hypertextes (qui offrent une interface ergonomique d'accès aux concepts par association d'idées). Couplé à un outil d'analyse de texte, on peut le

comparer à *Terminae*; dans la gestion et la manipulation de concepts, on peut le comparer à *Macao* et à *Code*. Cependant, la BdC est caractérisée par trois niveaux d'abstraction (terminologique, définitoire et de représentation) dans la représentation des connaissances. Elle permet de construire et de consulter des réseaux conceptuels modélisant un domaine en relation avec des textes où sont décrites les informations de ces réseaux. Elle cherche à différencier la représentation des connaissances terminologiques, conceptuelles et leur implémentation; de plus, elle facilite le retour au texte d'où sont extraites ces informations.

### 3 Le modèle conceptuel de la BdC

La terminologie est la science qui étudie les termes d'un domaine en précisant les relations qui existent entre ces différents termes. Comme le fait remarquer M. Van Campenhout (1999), elle est fondée sur un modèle tripartite dont les mots clés sont : le concept, l'objet et le terme. La figure 2.1 montre les relations qui existent entre ces trois entités à l'intérieur d'un domaine.

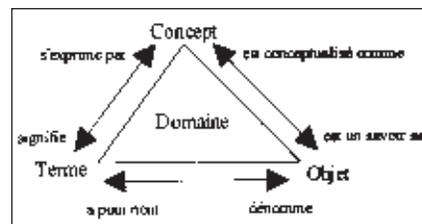


Figure 2.1  
Relations entre concept, objet et terme.

- **Domaine** : «un domaine est un ensemble de concepts reliés les uns aux autres par des relations sémantiques».
- **Concept** : «un concept est une idée abstraite donc générale pour définir un ensemble d'objets ayant des caractères ou propriétés communs». Par exemple le concept *plante* est une notion générale qui convient pour désigner toutes les plantes. Un concept peut être élémentaire ou complexe.
- **Objet** : «un objet est un élément de réalité qui peut être perçu ou conçu». Il peut être matériel (exemple : *plante*) ou immatériel (exemple : *maladie*). Un objet est une instantiation d'un concept, c'est la représentation concrète d'un concept. Par analogie au modèle objet, il peut être comparé à un objet d'une classe, c'est-à-dire une instance de classe.
- **Terme** : «un terme désigne au moyen d'une unité linguistique un concept défini dans une langue de spécialité». Il peut être constitué de un ou plusieurs mots. Exemple : *plante* et *plante ligneuse* sont deux termes qui désignent deux concepts différents.
- **Caractère** : «un caractère est une propriété d'un concept; il peut être assimilé à un attribut d'une relation du modèle relationnel ou à un attribut d'un objet du modèle objet». Par exemple le concept *plante* peut avoir les caractères suivants : famille, durée de vie, feuille.

Une BdC est donc constituée de trois niveaux d'abstraction qui sont : terminologique, caractère, et représentation. Le « niveau information terminologique » contient des informations de nature textuelle ou graphique permettant de prendre connaissance de la sémantique des concepts. Ces informations sont très peu adaptées pour effectuer des opérations de recherche pertinente sur les concepts, d'où le deuxième «niveau caractère». Quant au «niveau représentation», il

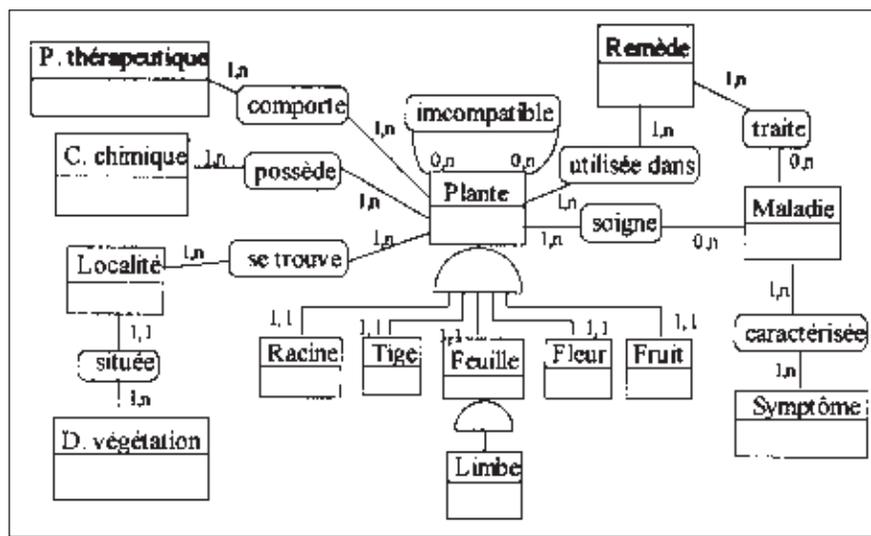


Figure 3.2  
Schéma conceptuel du domaine.

indique comment sont représentées et gérées les informations de la BdC.

Pour étayer nos propos, nous utiliserons les concepts du domaine de la phytothérapie, c'est-à-dire le traitement des maladies par les plantes (Figure 2.2).

La figure 3.2 illustre un schéma entité/association partiel simplifié des concepts manipulés dans ce domaine. Une plante est composée d'une racine, d'une tige, d'un type de feuille, d'un type de fleur et d'un type de fruit. Elle possède une composition chimique, elle peut soigner plusieurs maladies, elle peut être utilisée dans plusieurs remèdes. Un remède peut traiter plusieurs maladies.

### 3.1 Le niveau de l'information terminologique

Le niveau terminologique permet une représentation structurée mais non formelle des connaissances; la structure de concept, à ce niveau, contient des connaissances linguistiques relatives aux termes qui le désignent, et des connaissances

conceptuelles qui le situent dans une hiérarchie. On utilise la notion de fiche terminologique pour décrire les concepts à l'aide d'une structure commune. Chaque champ de cette structure est appelé information terminologique (IT), il peut contenir des informations textuelles ou graphiques permettant de prendre connaissance de la sémantique des concepts. Les différents champs de cette structure sont :

- **Terme**: un terme est la dénomination du concept, par exemple *plante*;

- **Langue**: dans un domaine, une langue est choisie et est utilisée par les experts du domaine afin d'avoir la même interprétation des concepts qu'ils utilisent; par exemple le *latin* est la langue véhiculaire en botanique;

- **Synonyme**: désignation du même concept (dans une même langue) par des termes différents; par exemple, le concept *Acacia senegal* (une plante) a pour synonyme *Acacia verek* en latin;

- **Équivalent**: désignation du même concept dans une autre langue; par exemple le concept *Acacia senegal* a pour équivalents *gommier blanc* en français et *ngobup uki* en serer (une langue du Sénégal);

- **Homonyme**: à un concept peuvent être associés un ou plusieurs homonymes c'est-à-dire des concepts ayant des termes identiques mais des valeurs de référencement différentes. Cette IT permet d'éviter les erreurs d'interprétation dues à la polysémie de certains termes. Par exemple, en botanique, le concept *oignon* désigne une plante potagère à bulbe comestible, il désigne également le bulbe de certaines plantes telles que le lis et la tulipe;

- **Concept générique**: à un concept peut être associé un concept plus général appelé concept générique. Cette IT permet de traduire la notion d'héritage entre concepts. Le concept *plante ligneuse* a pour concept

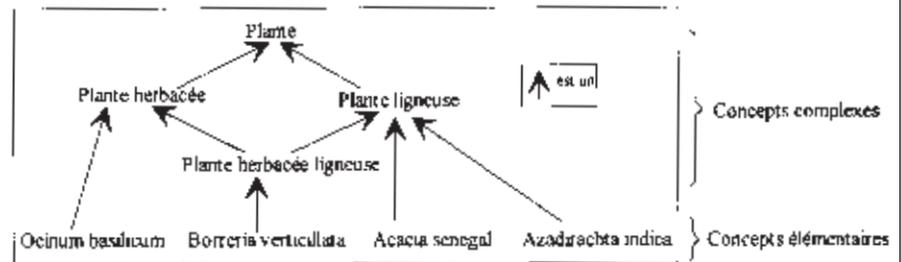


Figure 3.3  
Structuration hiérarchique du concept *plante*.

générique le concept *plante* (Fig. 2.3);

- **Sous-concept**: de même qu'à un concept peut être associé un concept générique, à un concept générique va être associé ses concepts spécifiques ou sous-concepts. Ces deux notions concept génériques et sous-concepts permettent de modéliser le domaine en hiérarchie de concepts (Fig. 3.3);
- **Concept lié**: cette IT permet de traduire toutes les relations sémantiques qui existent entre les concepts autres que la synonymie, l'homonymie et la hiérarchie. Elle peut permettre de décrire une relation d'association (par exemple *soigne, traite, est utilisée*, Fig. 3.2), d'antonymie, etc. entre concepts;
- **Définition**: la définition fournit la signification du terme qui désigne le concept; elle est donnée sous forme textuelle et peut contenir des notions faisant référence à d'autres informations de la BdC. Par exemple, dans la définition du concept *Acacia senegal* ci-dessous, les mots soulignés font référence à d'autres concepts du domaine (ex. *composées bipennées* fait référence au concept *feuille*).

[1] « L'*Acacia senegal* est un petit arbre de 4 à 7 m de haut, à épines courtes et courbes, longues de 4 à 7 mm, réunies par trois à la base du pétiole. Les feuilles sont composées bipennées...».

- **Contexte**: quant au contexte, il indique les conditions d'utilisation d'un concept. Il est de la même nature que l'information terminologique définition;
- **Image**: un concept peut être illustré par une image; cette image peut permettre d'appréhender d'autres concepts de la BdC. Il est donc intéressant d'avoir une image réactive afin d'accéder aisément à la sémantique de tous ces concepts (cf. annexe Fig. 6.2).

Le niveau information terminologique contient des informations de nature textuelle ou graphique permettant de prendre

connaissance de la sémantique des concepts. Ces informations sont très peu adaptées pour effectuer des opérations de recherche pertinente sur les concepts, d'où le deuxième niveau d'abstraction de la BdC que nous détaillons à présent.

### 3.2 Le niveau du caractère définitoire

À un concept est associé un ensemble de caractères qui sont les propriétés de leurs instances. Ce niveau permet de décrire l'ensemble de ces caractères spécifiques communs à un groupe de concepts. On peut assimiler un caractère ou caractère définitoire (CD) à la notion d'attribut d'une relation en BD. D'après D. Gouaderec (1992: 67), le terme définitoire est utilisé en terminologie pour indiquer qu'un terme contient des éléments de définition; ceci est le cas des caractères d'un concept parce qu'ils sont contenus dans leur définition. Considérons les CD de *plante* regroupés dans la table suivante:

plante	Famille	durée de vie	racine	tige	feuille	fleur	fruit
<i>Acacia senegal</i>	Mimosaceae	Pérenne	pivotante	aérienne	composée bipennée	Complète	sec déhiscent
<i>Azadirachta indica</i>	Méliaceae	Pérenne	pivotante	aérienne	composée paripennée	Complète	charnu

La première ligne contient les CD de *plante*, les autres contiennent une instanciation de ces CD pour deux instances du concept *plante*: *Acacia senegal* et *Azadirachta indica*. Les tuples de certaines tables relationnelles peuvent être également des CD des concepts; par exemple la table composition chimique suivante contient des tuples qui sont des CD des instances du concept *plante*.

### Composition chimique

partie utilisée	élément chimique	quantité
Tronc (gomme arabique)	eau	10 à 15 %
Tronc (écorce) Fruit	tanin vitamine	28 % trace

Les CD peuvent être élémentaires (famille durée de vie) ou complexes c'est-à-dire composés aussi d'autres CD (racine, composition chimique). Ils sont utilisés dans les opérations de recherche d'information dans la BdC.

Une BdC doit pouvoir répondre aux interrogations suivantes: sur la sémantique d'un CD (Qu'est ce qu'une famille?), sur la valeur d'un CD (Qu'est ce que *Mimosaceae*?), sur les concepts (Quelles sont les plantes de la famille des *Mimosaceae*?). Un modèle de requête simple, basé uniquement sur les informations terminologiques et les caractères définitoires a été proposé par Marcellin Brou (1997) pour faciliter

la recherche d'informations dans la BdC.

### 3.3 Le niveau de la représentation

Le troisième niveau indique le moyen utilisé pour la représentation et la gestion efficace des informations qui sont manipulées dans la BdC.

On peut aborder la représentation des informations de la BdC selon trois approches: hypertexte, base de données, couplage BD/hypertexte.

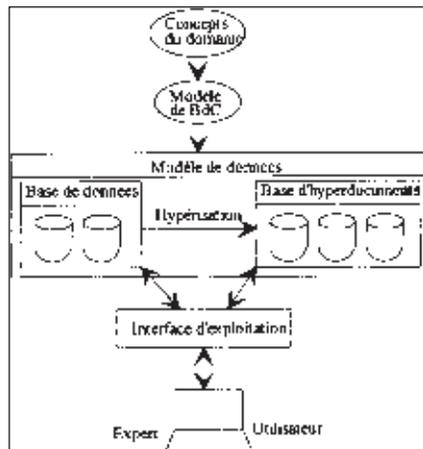


Figure 3.4  
Architecture générale d'une BdC.

La solution que nous avons retenue est fondée sur la troisième approche c'est-à-dire une BD relationnelle intégrant des facilités hypertextes qui s'inspire des solutions propriétaires O2 Web (1996), Oracle (1996) et Internet Information Server (1996). Cette approche propose non seulement une interface d'accès aux informations d'une BD, mais surtout un modèle de représentation et de gestion des informations de type hypertexte stockées dans une BD.

En effet, dans la BD nous stockons les informations de type hypertexte non pas sous forme d'un hypertexte déjà créé, mais plutôt leurs deux composants: contenu informatif et structure logique. L'hypertexte lui-même est généré par un processus d'hyperisation (fig. 3.4). D'après Mark Frisse (1988, 247) l'hyperisation est le travail qui permet de transcrire un document textuel linéaire en un réseau de nœuds hypertextes.

Cette approche facilite la maintenance de la base d'hyperdocuments (BHD), notamment dans les opérations de vérification de sa cohérence (cf. § 3.5). Elle permet également de créer un point de vue (vue partielle de la BdC modélisée par la liste des informations terminologiques et des caractères définitoires) qui permet de dériver une partie de la BHD contenant les informations qu'un utilisateur maîtrise le mieux et qu'il souhaite manipuler. Cette notion de point de vue permet de limiter la surcharge cognitive de l'utilisateur, qui est une source de désorientation, c'est-à-dire la perte de vue du but de sa recherche.

### 3.4 Le modèle conceptuel de données de la BdC

Le schéma de la figure 3.5 présente dans le formalisme Entité/Association, le modèle conceptuel de données de la BdC qui est constitué de deux parties: la partie

de droite modélise le contenu informatif de la BdC et la partie de gauche modélise sa structure logique.

- L'aspect informatif: les IT de type attribut (*terme et langue*) et de type hypertexte (*définition, contexte, image*) sont modélisées par l'entité concept. Les IT de type relation conceptuelle (*synonyme, homonyme, concept générique, sous-concept, concept lié*) par l'association «Relation\_conceptuelle». Les CD sont modélisés par les entités Cd\_simple et Cd\_type\_relation. L'entité description permet de modéliser certaines informations du domaine qui sont référencées par les IT de type hypertexte ou par les CD.
- L'aspect logique: un concept peut être stocké dans au moins une unité informative (UI) et une UI peut être composée d'autres UI; l'association «est\_composée» permet de modéliser la structure hiérarchique de la BHD. Une UI peut contenir des liens de référence; ces liens de référence sont modélisés par l'association lien. Un lien possède une ancre source et une ancre destination qui sont des UI. Un

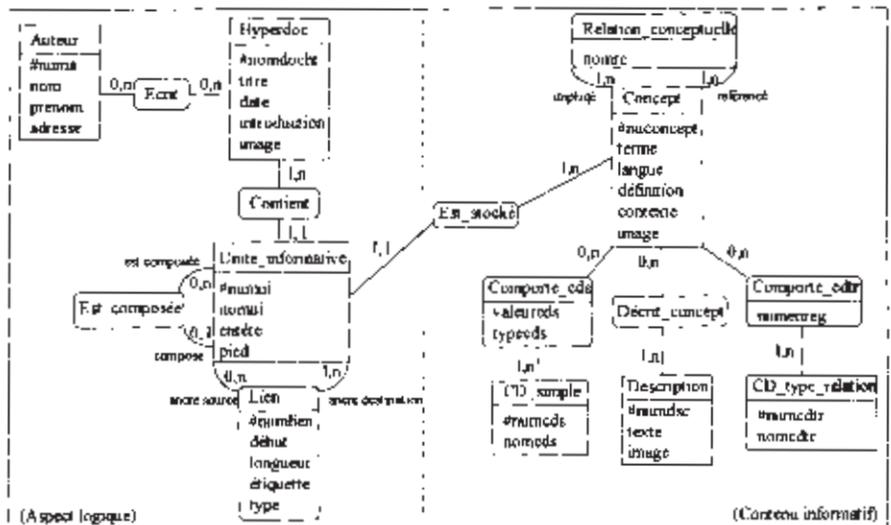


Figure 3.5  
Schéma conceptuel de données de la BdC.

hyperdocument contient au moins une UI, et il peut être l'œuvre de plusieurs auteurs.

### 3.5 Le modèle hypertexte de la BdC

Le modèle hypertexte de la BdC s'appuie sur le modèle d'hyperdocument expert défini par Nathalie Forest (1994) et Renaud Deschamps (1995). Dans ce modèle, les hyperdocuments sont décomposés en unités informatives reliées les unes aux autres par des liens hiérarchiques ou de référence.

- Les unités informatives (UI) : une UI est une extension des nœuds hypertextes classiques par l'introduction d'un ensemble de connaissances supplémentaires destinées à intégrer un aspect dynamique à leur contenu informatif; on peut la comparer aux documents actifs de Jocelyne Nanard (1993). Dans le cadre de la BdC, nous avons défini six types d'UI: UI du domaine, UI concept complexe, UI Concept élémentaire, UI informations terminologiques, UI caractères définitoires et UI description (cf. fig. 4.1).

- Les liens: on distingue deux sortes de liens: les liens hiérarchiques et les liens de référence.

Les liens hiérarchiques sont déduits des arbres conceptuels du domaine. Ils ont pour ancre les IT concept générique et sous-concept. Pratiquement, ils sont modélisés par la relation *Unité\_informative* dont nous présentons une instantiation comme suit:

#### *Unité informative*

numui	nomui	entête	numui2	nomdocht	numconcept
1	C1it	IT Plante		Medtrad	1
2	C5it	IT Acacia senegal	1	Medtrad	5
3	C5d	Définition Acacia senegal	2	Medtrad	5
5	C15i	Image Feuille		Medtrad	15
6	C5d1	Description Inflorescence	4	Medtrad	5
7	C5d2	Description épi cylindrique	4	Medtrad	5
8	C15d7	Description sommet	5	Medtrad	12

Où *numui2* est le numéro d'une UI dont l'UI de numéro *numui* et de nom *nomui* est une composante; entête est le titre de l'UI; nomdocht est le nom de la page d'accueil; numconcept est le numéro du concept qui est décrit par l'UI.

Les UI C5it et C5d sont des composantes de l'UI C1it, les UI C5d1 et C5d2 sont des composantes de l'UI C5d, elle-même composante de l'UI C5it. Cette structuration hiérarchique est réalisée grâce à l'attribut *numui2*.

Les liens de référence, dans la sont typés. On en distingue 5 types. Ces liens sont modélisés par la relation *Lien*, dont nous présentons une instantiation:

#### *Lien*

numlien	début	longueur	étiquette	type	numui	numui2
1	39	15	Caractère définitoire	définition	5	3
2	368	17		définition	5	6
3	6	0		cds	3	6
4	6	1		cds	3	7
5	360	170		image	4	8

Pour chaque type de lien, les attributs *début* et *longueur* ont une interprétation particulière:

1) Lien dont l'ancre est située dans un hypertexte, par exemple le lien n° 1: l'ancre source est située dans l'UI C5d, cette ancre commence au 39<sup>e</sup> caractère du contenu informatif de l'IT définition et a une longueur de 15 caractères.

2) Lien dont l'ancre est un CD ou la valeur d'un CD (lien n° 3 et 4). L'attribut *début* indique le rang du CD qui est l'ancre du lien. L'attribut *longueur* permet de déterminer si c'est le CD lui-même qui est l'ancre du lien (longueur = 0) ou si c'est sa valeur (longueur = 1).

3) Lien dont l'ancre est située dans une zone réactive circulaire dont le centre a pour abscisse *début* et pour ordonnée *longueur* (lien n° 5).

• Maintenance de la BHD : après les opérations de mise à jour de la BHD, elle doit rester cohérente au niveau de sa structure logique et au niveau de ses liens de référence.

Dans les hypertextes classiques, il n'y a pas de séparation entre leur contenu informatif et leur structure logique (cas des documents HTML), ceci rend leur maintenance difficile. Considérons l'hyperdocument HTML ci-dessous. Ce document contient deux liens, le premier fait référence à la partie caractères définitoires simples du document *C5cd.htm*. L'ancre de ce lien est *4 à 7 m de haut*, son activation permet d'accéder à la partie *caractère définitoire simple* du document *C5cd.htm*. Supposons que l'on supprime cette partie, pour que la BHD reste cohérente, il faut supprimer manuellement tous les liens des hyperdocuments qui font référence à cette partie.

```
<H1> Définition Acacia senegal
</H1>
L'Acacia senegal est un petit arbre de
<A HREF="C5cd.htm #Caractères
définitoires simples"> 4 à 7 m de
haut </A>, à épines courtes et
courbes, longues de 4 à 7 mm,
réunies par trois à la base du pétiole.
Les feuilles sont <A HREF=
"C15it.htm"> composées bipennées
</A>...
```

Dans la BdC, la maintenance est aisée. En effet, les opérations de mise à jour sont faites automatiquement de façon relationnelle dans les tables et le processus d'hypérisation la répercute sur la BHD.

Considérons l'UI U<sub>i</sub> contenant des liens qui pointent dans le vide (qui n'ont pas d'UI destination). Dans la BdC, la vérification de cohérence de la BHD au niveau des

liens de référence est effectuée en trois étapes :

1) la première étape consiste à rechercher les liens théoriques de cette UI dans la table lien grâce à une requête de type SQL comme suit :  
 SELECT numui, numui2, étiquette,  
 début, longueur INTO  
 Lien\_théorique  
 FROM lien  
 WHERE numui = j

#### Lien\_théorique

numui	numui2	étiquette	début	longueur
4	1	Caractère définitoire simple	40	15
4	2		167	19

2) Ensuite, grâce à un programme, nous recherchons les liens effectifs qui existent dans le code hypertexte de l'UI U<sub>j</sub>.

#### Lien\_effectif

UIS	UID	étiquette	ancre
C5d	C5cd.htm	Caractère définitoire simple	4 à 7 m de haut
C5d	C15it.htm		Composées bipennées

3) Enfin, par la comparaison entre les éléments de ces deux tables (numui, UIS), (numui2, UID), (début, longueur, ancre) nous déterminons les causes de cette incohérence.

## 4 Implémentation de la BdC

### 4.1 Le système CoDB-Web

Le schéma de la figure 3.1 présente le système *CoDB-Web* (Concept DataBase Web) qui est le système de gestion et de

représentation hypertexte de la BdC fondé sur le *World Wide Web*. C'est une interface particulière entre le SGBD qui gère la BdC et le système hypertexte qui gère la BHD. Il est constitué de quatre modules interconnectés.

Le cadre de développement de la BdC est celui d'une architecture client-serveur qui respecte les protocoles de l'Internet (adressage HTTP, langage HTML...), moyen idéal de diffusion de la connaissance stockée dans la BdC.

Le système hypertexte est donc le système hypermédia W3 qui offre une interface graphique d'accès aux ressources d'Internet. Les documents hypertextes qu'il manipule sont décrit à l'aide du langage HTML.

L'avantage de W3 est que son formalisme est bien défini et normalisé, il est mondial et disponible sur toutes les plates-formes (environnement *Windows* ou *Unix*). On peut noter cependant l'absence d'outils de conception et de maintenance des hyperdocuments, d'où notre proposition de la notion de point de vue et du système *CoDB-Web* pour gérer efficacement la BdC.

Quant au système de gestion de base de données, nous avons choisi le SGBDR *Access* qui permet de développer rapidement des applications BD dans l'environnement graphique *Windows*.

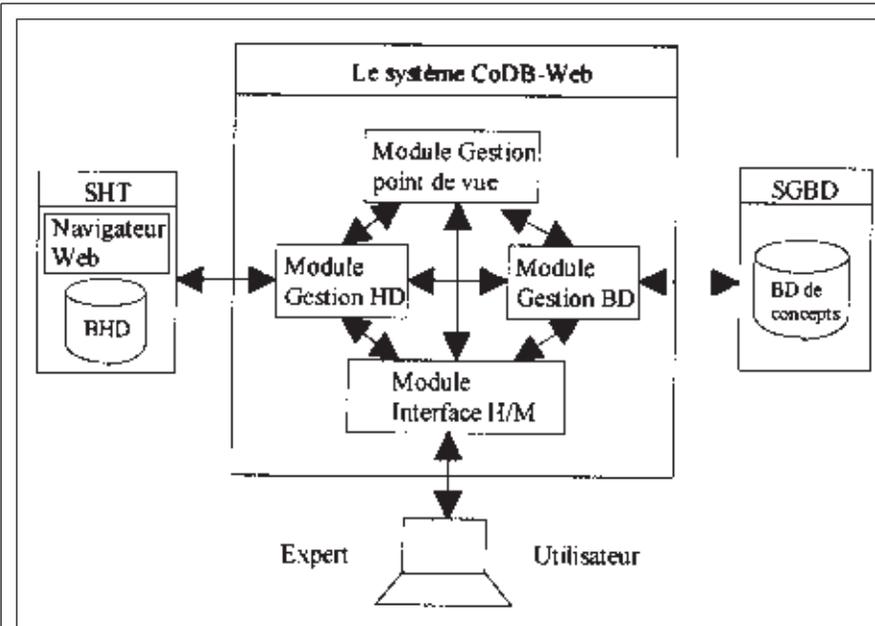


Figure 3.1  
Schéma de fonctionnement général de la BdC.

Entre le système hypertexte, le SGBD et le système *CoDB-Web*, il fallait trouver un langage fédérateur. Nous avons choisi de ce fait comme langage d'implémentation de la BdC le langage *JavaScript Intrabuilder* qui est une version enrichie de *JavaScript*, lui-même une implémentation simple de *Java*. *JavaScript Intrabuilder* permet d'intégrer un aspect dynamique aux pages Web. Sa principale qualité est son interface d'accès aux BD localisées sur les serveurs. Il faut signaler cependant sa lenteur due au fait qu'il est interprété.

## 4.2 Méthodologie d'instanciation de la BdC

Pour la construction d'une BdC, nous avons proposé une méthodologie comportant six étapes:

1) Délimitation du domaine: elle consiste à cerner l'ensemble des connaissances du domaine à étudier afin d'isoler les concepts du domaine;

2) Collecte d'informations: il s'agit de rechercher les documents de référence du domaine étudié tels que les dictionnaires, les règlements, les livres, etc. Ces documents permettront d'isoler les concepts clés du domaine et fourniront des renseignements utiles à la détermination des valeurs des informations terminologiques et des caractères définitoires;

3) Modélisation du domaine sous forme d'arbres conceptuels: il s'agit d'organiser les concepts du domaine de façon arborescente afin de faire apparaître les relations hiérarchiques qui existent entre les concepts. Ces relations permettront de trouver les concepts génériques.

4) Modélisation de l'application sous forme d'un schéma conceptuel: pour la mise en place de ce schéma, nous préconisons l'utilisation du modèle entité/association pour la simplicité de son formalisme; dans ce formalisme, nous représentons chaque concept par

une entité, et la relation sémantique entre deux concepts est traduite par une association. Le schéma conceptuel aide à la détermination des caractères définitoires des concepts (attributs des entités) et les concepts liés (entités liées par certaines associations);

5) Établissement des fiches terminologiques: une fiche représente un concept à l'aide d'un ensemble de champs textuels ou graphiques contenant les valeurs propres au concept. Il s'agit de déterminer dans un premier temps le type des informations du domaine qui vont permettre d'alimenter les champs de ces fiches, et dans un deuxième temps d'instancier ces fiches pour chaque concept. Il existe deux sortes de fiches: des fiches « informations terminologiques » et des fiches « caractères définitoires ». Ces fiches facilitent la saisie des données dans les tables relationnelles de la BdC.

6) Utilisation de la BdC: cette étape comporte 4 parties: la génération des tables relationnelles de la BdC réalisée par un programme *Access Basic*, l'instanciation de ces tables grâce à des fiches de saisie *Intrabuilder*, le lancement d'un programme *JavaScript Intrabuilder* pour générer la BHD. La dernière étape est l'utilisation proprement dite de la BdC selon deux modes d'accès: la navigation ou la recherche. L'utilisateur peut utiliser une version limitée de la BdC grâce à la version *Access* ou la version navigation qui est plus complète. Nous présentons en annexe quelques images écrans de la version navigation relatives à l'application *MedTrad*.

## 5 L'application *MedTrad*

*MedTrad* est une BdC sur la médecine traditionnelle africaine à base de plantes médicinales. La médecine traditionnelle occupe une place très importante dans les pays en

voie de développement et plus particulièrement en Afrique, mais malheureusement cette connaissance tend à disparaître d'une part à cause du modernisme, mais surtout à cause de la transmission de ce savoir qui est mal assurée.

La problématique de cette médecine traditionnelle a été abordée par Claude Frasson (1992) dans le cadre du système *Seiboga*. Ce système est basé sur une exploitation graphique nécessitant parfois une certaine compétence de l'utilisateur.

L'application *Medtrad* a aussi pour objectifs, d'une part, de capitaliser cette connaissance ancestrale africaine sur la phytothérapie qui présente des aspects rationnels et des résultats convaincants éprouvés par de nombreux chercheurs africains tels que Guy Maynard (1990), Sijelmassi (1993) et Assi (1996). D'autre part, elle doit apporter une aide aux tradipraticiens dans leur processus de diagnostic et de proposition d'une thérapie rigoureuse.

La figure 5.1 schématise la structuration de la base d'hyperdocuments de *MedTrad*. La page d'accueil (*UI domaine*) est composée d'une introduction qui décrit l'application et d'un sommaire constitué par la liste des concepts clés du domaine; ces concepts sont des ancres de liens permettant d'accéder à leurs unités informatives qui sont des *UI concept complexe*. Chaque concept est représenté par deux fiches terminologiques (UI informations terminologiques et UI caractères définitoires) comme celle du concept *Acacia senegal*.

L'application *MedTrad* permet l'apprentissage de la phytothérapie en se basant sur une interface hypertextuelle bien adaptée à une mémorisation à long terme et à une compréhension profonde (assimilation).

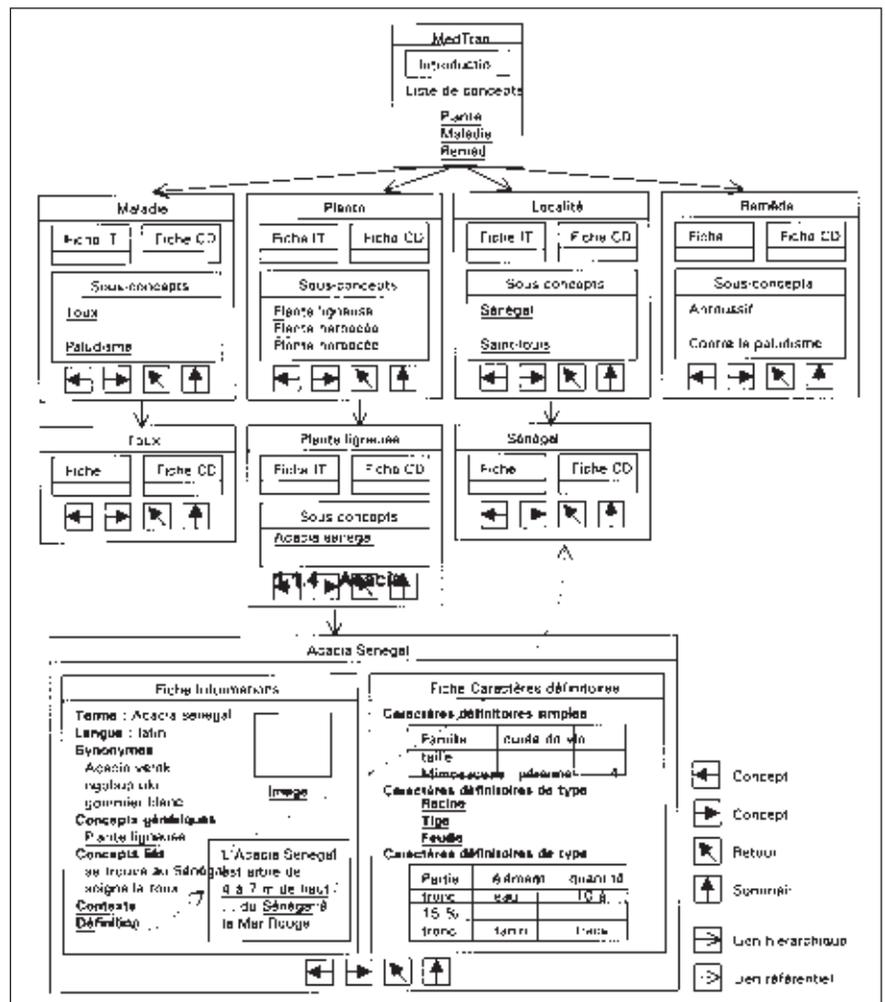


Figure 5.1  
Base d'hyperdocuments de *MedTrad*.

## 6 Conclusion

L'objectif de cet article était de proposer un modèle de représentation et de gestion des concepts d'un domaine afin de construire des ontologies facilement accessibles et réutilisables. L'outil d'exploitation de ce modèle appelé base de concepts (BdC) est *CoDB-Web*, il a été développé en utilisant les bases de données, des techniques hypertextes et des résultats sur la représentation des connaissances. Nous avons

présenté une utilisation possible de la BdC dans le cadre de l'acquisition de connaissance à travers l'application *MedTrad* dont le but est de capitaliser et de diffuser la connaissance ancestrale africaine sur les plantes médicinales. Il valide le premier point de notre objectif, à savoir la construction d'une ontologie facilement accessible. Quant au second point lié à la réutilisation d'ontologie, nous n'avons pas suffisamment d'expérience pour l'affirmer. Outre le développement

d'application, on peut citer d'autres cas d'utilisation de la BdC tels que : l'explication de la connaissance statique d'un SBC par le couplage de *CoDB-Web* avec un système expert, la création d'interface ergonomique d'accès à une base de données, l'aide au traducteur qui cherche à comprendre un texte dans un domaine dont il n'est pas le spécialiste...

L'instanciation de la BdC a travers *MedTrad* a soulevé les problèmes suivants :

- La détermination des termes et des concepts : elle est faite manuellement par l'expert du domaine, peut-on l'automatiser ?
- La classification des concepts : peut-on la faire automatiquement ;
- Le calcul de la position des ancrés dans le contenu textuel des informations terminologiques de type hypertexte ;
- Le mode d'hypérisation : dynamique ou statique ?

Dans l'avenir, nous essaierons d'apporter des éléments de réponse à ces problèmes.

A. Hocine,  
Département informatique  
Université de Pau et des  
pays de l'Adour  
France.

Konan Marcellin Brou,  
Département mathématique-  
informatique  
Institut national polytechnique  
Yamoussoukro,  
Côte d'Ivoire.

## Bibliographie

Aussenac (N.), Bourigault (D.), Condamines (A.) : «How Can Knowledge Acquisition Benefit from Terminology?», dans *Proceedings of KAW95*, Banff (Can.), fév. 1995, [http://www.irit.fr/ACTIVITES/EQ\\_SMI/PUBLI/banff95.html](http://www.irit.fr/ACTIVITES/EQ_SMI/PUBLI/banff95.html).

Assy (L.-A.), 1996 : *Plantes utilisées dans la médecine traditionnelle en Afrique de l'Ouest*, Édition Roche Basse, Suisse.

Biebow (B.), Szulman (S.), 1997 : «Avancée sur le concept de base de connaissance terminologique», 6<sup>e</sup> journée nationale du PRC-GDR intelligence artificielle.

Bonjour (M.), Falquet (G.), Léonard (M.), 1994 : «Base de concepts et intégration de bases de données», dans *actes du congrès INFORSID'94*, Aix-en-Provence, 17-20 mai 1994, p. 263-280.

Borland, 1996 : *Borland Intrabuilder pour Windows 95 & Windows NT*, manuel d'introduction, 1996.

Bourigault (D.), Lepine (P.), 1994, «Une méthode d'utilisation de Lexter en acquisition des connaissances», dans 5<sup>e</sup> Journées d'acquisition des connaissances, Strasbourg.

Brou (K.-M.), 1997 : *Base de concepts : Contribution à la représentation et à l'exploitation hypertexte de concepts – le système CoDB-Web*, thèse de doctorat de l'Université de Pau, novembre 1997.

Deschamps (R.), 1995 : *Bases de connaissances généralisées : une approche fondée sur un modèle hypertexte expert*, thèse de doctorat de l'Université Paul Sabatier de Toulouse, janvier 1995.

Forest (N.), 1994 : *Bases de connaissances généralisées : le modèle de gestion d'informations de types hypertexte*, thèse de doctorat de l'Université de Pau, janvier 94.

Frasson (C.), Houtsa (F.), Obenson (P.), 1992 : «Interface visuelle pour l'aide au diagnostic médical en médecine traditionnelle» dans *ICO*, vol. 4, n° 1 et 2, 1992, p. 17-25.

Frisse (M.), 1988 : «From text to hypertext», Byte, octobre 1988, p. 247-254.

Gouadec (D.), 1992 : «Terminologie et terminotique, outils, modèles et méthodes», actes de la première Université d'automne en terminologie, Rennes 2 - 21 au 26 sept 1992, Édition la Maison du dictionnaire, p. 67-120.

Garcia (C.), 1996 : «Construction coopérative d'ontologies dans un cadre de multi-expertise : ébauche d'un outil», dans JAC 96, Journée acquisition apprentissage, Sète 8-10 mai 1996, p. 95-107.

Internet Information Server, 1996 : *Documentation technique Windows NT Server 4.0*, 1996.

Maynard (G.), Lô (M.), Fortin (D.), 1990 : *Plantes médicinales du sahel*, Série études recherches, 1990.

Nanard (J.), Nanard (M.), Massotte (A.), Djemaa (A.), Joubert (A.), Betaille (H.), Chauché (J.), 1993 : «Integrating knowledge-based hypertext and database for task-oriented access documents», dans *actes DEXA'93*, p. 721-732.

O2 Technology, 1996 : *O2 web user Manual*, release 4.6, January 1996.

Oracle Corporation, 1996, *Guide des solutions micro et Internet*, 1996.

Sijelmassi (A.), 1996 : *Les plantes médicinales du Maroc*, Édition Le Fenec, 3<sup>e</sup> édition 1993.

Skuce (D.) : *CODE4 : A Unified Sytem for Managing Conceptual Knowledge*, document Web, <http://www.csi.uottawa.ca/~ingrid/cogniterm.html>.

Van Campenhoudt (M.) : *Abrégé de terminologie multilingue*, document Web, <http://www.refer.fr/termisti/centre.htm#bref>.

## Annexe

La figure 6.1 présente la page d'accueil du système *MedTrad*; elle comporte une introduction qui pose la problématique de la médecine traditionnelle africaine et un

sommaire constitué par les deux modes d'accès à *MedTrad* (la navigation et la recherche). La figure 6.2 présente les fiches terminologiques du concept *acacia senegal*. Cette fiche comporte deux parties: la partie gauche décrit les

informations terminologiques, et celle de droite les caractères définitoires. La figure 6.3 présente l'image du concept *feuille* (image réactive); elle permet aussi de découvrir les concepts *limbe*, *nervure*, *pétiole* et *renflement*.

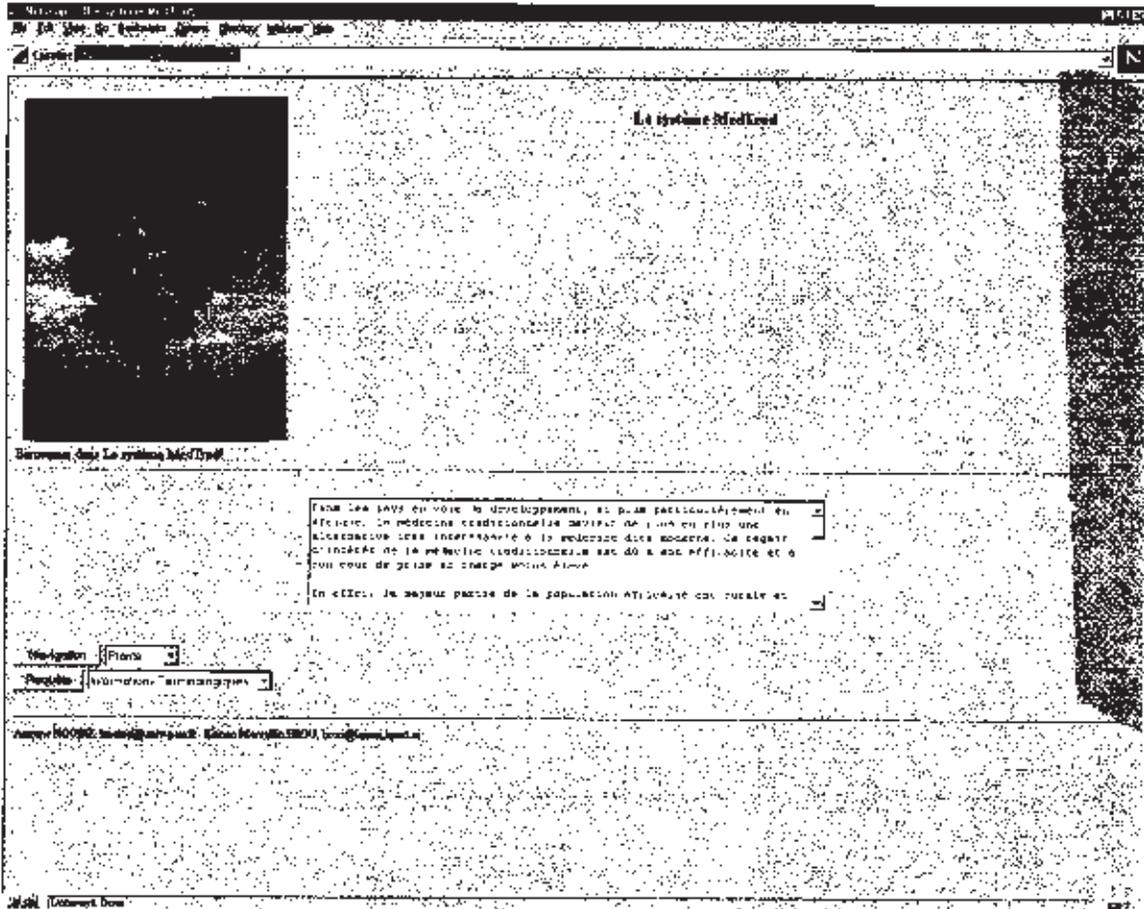


Figure 6.1  
La page d'accueil du système *MedTrad*.

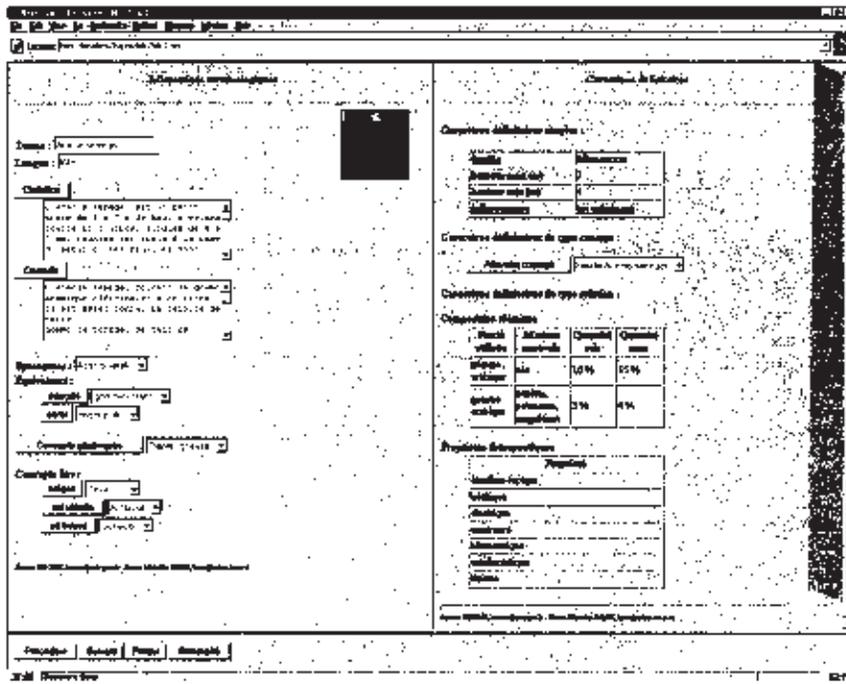


Figure 6.2  
Fiches terminologiques du concept *Acacia senegal*.

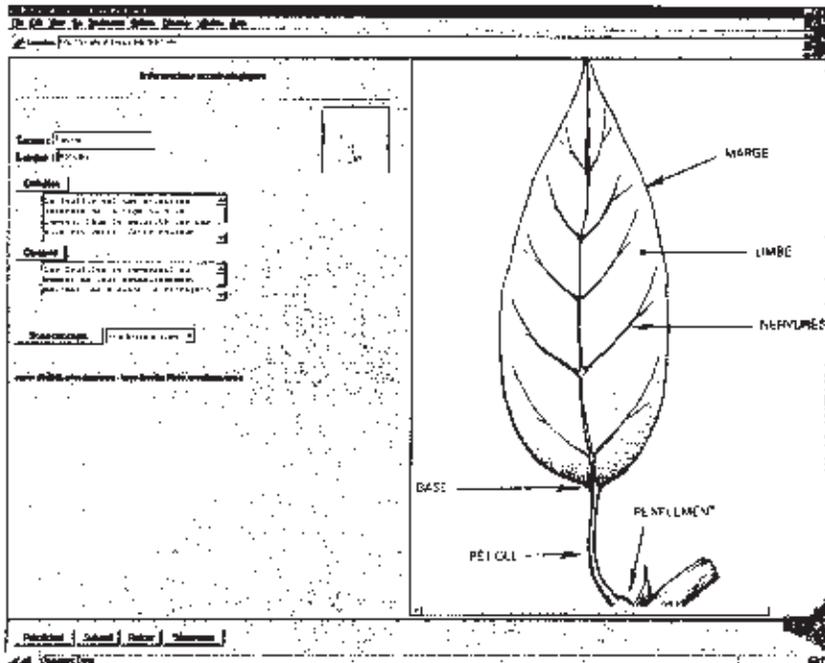


Figure 6.3  
Activation d'un lien hypertexte image.