

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE ,

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENE (U. S. T. H. B.) ALGER

INSTITUT D'INFORMATIQUE

MEMOIRE DE FIN D'ETUDES
En vue de l'Obtention du Diplôme d'Ingénieur d'Etat
en Informatique

Option : SOFTWARE

***SYSTEME D 'AIDE A LA CONCEPTION
DE BASES DE DONNEES EN
LANGAGE NATUREL***

Présenté par :

Mr. Arezki AZOUAOU
Mr. Mohamed Tayeb DJEMAI

Promoteur :

Mr. Omar NOUALI

Devant le jury :

Mme Z. ALIMAZIGHI President,
Melle K. AKLI Examineur,
Mr A. BOUDJADJA Examineur

Organisme d'accueil :
Centre de Recherche sur l'Information Scientifique et Technique
(CERIST)

N°: 12/94

Résumé

Cette these **décrit un** systeme d'aide à la conception de bases de **données** en **langage** naturel utilisant l'approche systbme expert. Le systeme est destine a diriger **un dialogue avec** le concepteur, **déetecter** les **éventuelles incohérences**, faire en sorte que le schema conceptuel obtenu soit **complet**, et **enfin**, produire le schema **relationnel** pour le systeme de gestion de bases de données utilisé. Il s'agit alors de produire les structures conceptuelles par une analyse des phrases en **langage naturel**.

Les principaux objectifs du systeme sont :

- construire une base de **connaissance** pour la conception de bases de **données**,
- doter le systeme **d'une** interface qui **permet à** l'utilisateur de **donner** la specification de l'application à **modéliser**, en **langage** naturel, et par consequent, lui Cviter l'apprentissage dun **langage** technique.

Le systeme a une architecture **générale** (modulaire) lui permettant de s'adapter à toute extension **éventuelle**. La version courante du systeme est **écrite** en **langage** C++ et LISP, elle est **implémentée** sous le systeme d'exploitation UNIX sur une station de travail SUN-4.

Mots-clés

Conception de bases de **données**, Analyse des besoins, **Systèmes experts**, Schema conceptuel E/A, Schema relationnel, SGBD, **langage** naturel.

Sommaire

INTRODUCTION	1
--------------------	---

Chapitre I

ETUDE COMPARATIVE DE METHODES DE CONCEPTION DE BASES DE DONNEES

1. Introduction	5
2. Architecture d'un SGBD	6
2. 1. Les niveaux de l'architecture ANSI/SPARC	6
2. 2. Difficulté d'une telle approche	8
2. 3. Formalismes associés aux différents niveaux	9
2. 4. Modèles conceptuels, différentes approches	9
3. Modèles sémantiques de données versus modèles de données classiques	11
3.1. Défauts des modèles de données conventionnels	11
3.2. Objectifs et exemples de modèles sémantiques	13
3.3. Critères pour le choix d'un modèle de données	15
4. Modèle sémantique de données Entité-Association Ctendu	17
4.1. Quelques points forts du modèle E/A Ctendu	20
4.2. Exemple de schéma E/A Ctendu	21
5. Conclusion	23

Chapitre II

LA COMPREHENSION AUTOMATIQUE DU LANGAGE NATUREL

1. Introduction	24
2. Bref historique	25
3. Difficulté de la compréhension automatique du langage naturel	25
4. Les connaissances nécessaires à la compréhension	27
4.1. Les connaissances morpho-lexicales	28
4.2. Les connaissances syntaxiques	29
4.3. Les connaissances sémantiques	31
4.4. Les connaissances pragmatiques	32

5. Representation des connaissances	33
5.1. Les connaissances a considérer	33
5.2. Représentations déclaratives vs représentations procédurales	33
5.3. Panorama des modèles de représentation du sens	34
6. Les principes de l'analyse de phrases	37
6.1. Caractéristiques issues des connaissances et de la représentation inteme	38
6.2. Choix des mecanismes d'analyse	40
6.3. Quelques types de systèmes d'analyse	44
7. Conclusion	45

Chapitre III

PRESENTATION GENERALE DU SYSTEME D'AIDE A LA CONCEPTION DE BASES DE DONNES EN LANGAGE NATUREL

1. Introduction	46
2. Le choix du modele conceptuel de données	47
3. Présentation de l' interface langage naturel	47
3.1. Choix des ATN sémantiques comme outil d'analyse syntaxico- sémantique	49
3.2. Choix des connaissances adequates	50
4. Organisation globale du systeme	53
4.1. Fonctionnement inteme du systeme	53
4.2. Aspect exteme du systeme	56
5. Conclusion	57

Chapitre IV

METHODOLOGIE DE CONCEPTION DE BASES DE DONNEES

1. Introduction	58
2. Le modele Entité-Association étendu	58
2.1. Restrictions imposées par le modele	59
2.2. Représentation d'un schema conceptuel	60
3. Description de la méthodologie de conception	63
3.1. Stratégie globale de conception d'une base de données	64
3.2. Contraintes de coherence d'un schema conceptuel E/A	64
3.3. Détection et élimination des redondances	65
3.4. Complétude du schema conceptuel conformément au modele	

choisi	67
3.5. Traduction d'un schema E/A Ctendu en un schema logique relationnel	68
3 5.1. Principaux concepts et contraintes du modèle relationnel	68
3.5.2. Elimination des hierarchies de généralisation	69
3 5.3. Elimination des identifiants externes	70
3.5.4. Elimination des attributs multi-values	71
3.5.5. Traduction des associations	72
3.6. Choix des identifiants d'entités	76
3.7. Nonnalisation du schema logique relationnel obtenu	77
4. Caractéristiques du générateur de système expert utilisé	79
5. Conclusion	80

Chapitre V

L'ANALYSE MORPHO-LEXICALE

1. Introduction	81
2. Description de la méthode d'analyse morpho-lexicale conçue	81
3. Analyse de mots	83
3.1. Analyse de mots à nombre variable	85
3.2. Analyse des verbes	85
3.3. Traitements particuliers	88
4. Organisation du dictionnaire	88
5. Conclusion	90

Chapitre VI

L'ANALYSE SYNTAXICO-SEMANTIQUE

1. Introduction	91
2. Description de l'analyseur conçu	91
2.1. Types d'arcs	93
2.2. Algorithme d'analyse	96
3. Ordonnancement des arcs	96
4. Exemple d'ATN	98
5. Conclusion	99

Chapitre VII

LES GRAPHES CONCEPTUELS DE SOWA

1. Introduction	100
-----------------------	-----

2. Le formalisme des graphes conceptuels de Sowa	101
2.1. Les concepts	101
2.2. Relations conceptuelles	103
3. Les graphes conceptuels comme modèle de représentation interne du sens	104
4. Les graphes conceptuels comme outil d'analyse sémantique	107
5. Conclusion	110
 CONCLUSION ET PERSPECTIVES	 111

BIBLIOGRAPHIE

ANNEXES