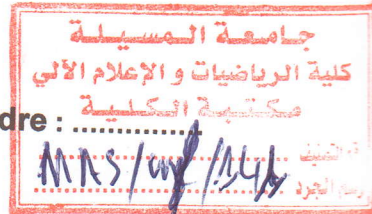




N° d'ordre :



UNIVERSITE DE M'SILA
FACULTE DES MATHÉMATIQUES ET DE L'INFORMATIQUE

Département des Sciences des Technologies de l'information
et de la Communication

MEMOIRE de fin d'étude

Présenté pour l'obtention du diplôme de **MASTER**

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Technologie de l'information et de la communication

Par: **BENKHELIFA Nassima**

SUJET

**Conception et réalisation d'un système Questions-Réponses
en langage naturel pour l'exploitation d'une base de
données relationnelle**

Soutenu publiquement le : .../06/2015 devant le jury composé de :

.....
Mr. MAHDJOURI Roussafi

Université de M'sila

Université de M'sila

Université de M'sila

Université de M'sila

Président

Rapporteur

Examineur

Examineur

Promotion : 2014 / 2015

Table des matières

Introduction générale.....	1
Chapitre I : Les systèmes Questions-Réponses	
1. Introduction	3
2. Comparaison entre moteur de recherche et système Question-Réponse.....	4
3. Évolution des systèmes Question-Réponse	5
4. Caractéristiques des systèmes Question-Réponse	7
5. Architecture des systèmes Question-Réponse	7
5.1. Analyse des questions	9
5.1.1. Le type de la question	9
5.1.2. Les entités nommées de la question	10
5.1.3. Le type de la réponse attendue	10
5.1.4. Le focus de la question	11
5.2. Recherche des documents	11
5.3. Extraction des réponses	13
5.3.1 Patrons d'extraction	13
5.3.2. Utiliser des critères	14
5.3.3. Par les relations syntaxiques	14
5.3.4. Utilisation de la logique	15
5. Domaines des systèmes Question-Réponse	15
6. Présentation de quelques systèmes de réponses aux questions	15
6.1. Le système QALC	16
6.2 Le système QRISTAL.....	16
6.3 Le système PIQUANT	16
6.4 Le système WEBCOOP	16
7. Conclusion	17
Chapitre II : L'interrogation des bases de données en langage naturel	
1. Introduction	18
2. L'Interface en langage naturel pour les bases de données NLIDB (Natural Language Interface for Database).....	19
3. Avantages et inconvénients des NLIDB	19

3.1. Avantages des NLIDB (Natural Language Interface for Database)	19
3.2. inconvénients des NLIDB (Natural Language Interface for Database) ...	20
4. Présentation de quelques systèmes NLIDB	20
5. L'architecture d'une NLIDB (Natural Language Interface for Database)	21
6. Traitement automatique des langues naturelles	23
6.1. Le niveau morphologique	24
6.2. Le niveau syntaxique	25
6.3. Le niveau sémantique	25
7. Conclusion	26
Chapitre III: Conception	
1. Introduction	27
2. Objectif de travail	27
3. Présentation du SGBD (<i>Système de Gestion de Bases de Données</i>)	28
3.1. Définition et principes de fonctionnement.....	28
3.2. Objectifs	28
3.3. La bases de données	29
4. Fonctionnalités du système	34
4.1. Installe du dictionnaire grammatical	35
4.2. Analyse est phrase étiqueté par une grammaire	35
5. Système « NATURAL LANGUAGE TO SQL».....	35
6. conception de traducteur de langage naturel	36
6.1. traduction de langage naturel	39
7. Le principe de fonctionnement des systèmes	40
8. Conclusion	40
Chapitre IV : Implimentation et Exploitation	
1. Introduction	41
2. Présentation des outils utilisée	41
2.1. La langage Perl	42
2.2. Utilisation de modules	43
2.2.1. Principaux les modules utilisés	44
2.2.2. Exécuter des programmes Perl	49
2.3. Outils Perl Padre	49
2.4. MySQL.....	50

3. Description d'interface du système	
NATURAL LANGUAGE TO SQL (NL TO SQL).....	51
3.1. La Configuration de la base de données.....	51
3.2. Identifiant	51
3.3. Fenêtre principale	52
3.4. Structure de base de données	52
3.5. Fenêtre de relations ou jointeur de Tableau	53
3.6. Fenêtre de relations ou jointeur avec condition de Tableau	53
3. Présentation de quelques interfaces	54
Conclusion générale.....	55
Bibliographie.....	56

Introduction générale

L'expansion constante du nombre de documents électroniques, notamment grâce à Internet, a rendu l'accès à l'information plus aisée et rapide. De nos jours, rechercher une information ou un document sur le Web est devenu une activité quotidienne et prépondérante pour les internautes. Cette explosion du nombre de documents s'accompagne d'un accroissement du nombre de données et d'utilisateurs interrogeant les différents moteurs de recherche devenus très populaires.

La recherche d'information (RI) permet à un utilisateur de formuler un besoin d'information, à l'aide d'une requête pour obtenir une réponse issue d'un ensemble de documents. L'idée est apparue dès la naissance des premiers ordinateurs. Elle s'est développée dans les années 1960 avec la possibilité de stocker et d'analyser des masses importantes de données. Depuis, avec la gigantesque augmentation des connaissances produites et conservées numériquement, elle est devenue un secteur stratégique pour beaucoup d'applications par l'intermédiaire de l'internet. Dans l'histoire de la recherche d'information, les systèmes sont basés sur plusieurs modèles visant à réaliser l'appariement entre la représentation des documents et celle des requêtes.

Cependant, l'exploitation de cette grande quantité de données rend la recherche d'une information précise complexe et coûteuse en termes de temps. Cette difficulté a motivé le développement de nouveaux outils de recherche adaptés, comme les systèmes de question-réponse.

En effet, ce type de système permet à un utilisateur de poser une question en langage naturel et de retourner une réponse précise à sa requête au lieu d'un ensemble de documents jugés pertinents, comme c'est le cas des moteurs de recherche où plusieurs moteurs d'inférences peuvent être intégrés au système afin d'interroger la base de connaissances. Racer et Pellet sont des exemples de moteurs d'inférence reposant sur la logique de description. Afin d'interroger ces moteurs, plusieurs langages d'interrogation ont été définis. Ces langages fournissent des mécanismes permettant d'exprimer des requêtes complexes. La requête est alors exécutée sur la connaissance ou la base de données et les instances qui satisfont la requête sont donc retournées.

Cependant, interroger en langage naturel une base de données n'est pas une tâche simple. Cela engendre deux obstacles majeurs :

Introduction générale

(i) L'ambiguïté de la langue naturelle : Ceci génère de nombreuses difficultés se présentant à tous les niveaux de l'analyse linguistique des requêtes (les problèmes peuvent être d'ordre morphologique, syntaxique et sémantique).

(ii) La formalisation des requêtes en langage d'interrogation de bases de données : même après l'analyse linguistique des requêtes en langage naturel, beaucoup de défis restent à relever dans la conversion des requêtes en langages formels et la puissance du langage d'interrogation utilisé.

Dans ce projet, nous proposons l'architecture d'un système de conversion de requêtes formulées en langage naturel vers des requêtes SQL (Structured Query Language) à l'aide d'une interface à travers laquelle, l'utilisateur peut interroger le système en langage naturel comme l'anglais, pour obtenir les résultats recherchés. La recherche dans ce cas ne repose pas sur la notion de mots clés mais sur celle de grammaire du langage. Le système NLIDB est proposé en tant que solution au problème de l'accès aisé aux informations au profit de tout utilisateur même inexpérimenté.

Pour présenter ce travail, nous avons organisé le mémoire comme suit:

Le premier chapitre : donne une vue générale sur les systèmes de questions-réponses en présentant quelques-uns d'entre eux.

Le deuxième chapitre : présente l'idée de l'interface en langage naturel à base de données (NLIDB). Nous étudions ensuite, le traitement automatique des langues naturelles (TALN) et son influence sur ce domaine. Finalement, nous citons les principaux modèles existants qui visent à réaliser l'appariement entre la représentation des données et celle des requêtes.

Le troisième chapitre : Conception d'un système de conversion des requêtes en langage naturel vers SQL.

Le quatrième chapitre : présente les outils de développement de notre système, ainsi que la présentation de quelques-unes de ses interfaces.

Enfin, une conclusion clôturera ce mémoire pour expliciter l'intérêt et le fruit de ce travail.

Conclusion générale

Les approches de la recherche d'information actuelles ne saisissent pas formellement la signification explicite d'une requête en langue naturelle ; mais fournissent une voie confortable pour l'utilisateur qui spécifie ces besoins en informations sur la base de données .

Nous avons abordé dans ce mémoire, la problématique de la construction des requêtes formelles et nous avons présenté une nouvelle approche pour la résoudre. Notre système vise à traduire les requêtes de langage naturel vers des requêtes SQL, Nous nous sommes fixés au départ, la réalisation d'un système qui permettrait à quiconque de poser des questions en langage naturel pour obtenir en contrepartie, une réponse et afficher les résultats et sans que l'utilisateur ne soit obligé d'apprendre un langage de requêtes spécifique. Notre approche peut se résumer ainsi : Une fois que l'utilisateur entre une requête en langage naturel, la première étape réalisée est l'analyse linguistique des requêtes utilise la grammaire, la deuxième étape est la conversion des requêtes en langage naturel vers SQL, Puis, l'étape de la génération de requêtes SQL. Enfin, la requête SQL sera remise à l'utilisateur final sous forme de résultat.

La plupart des objectifs ont été mis en application avec succès. Le programme peut traduire des questions en langage naturel simples vers SQL. Il peut traduire vers différents types de questions SELECT, qui incluent recherche des données dans une ou deux tables.

L'objection visé au départ est plus ou moins atteint. Le système n'est pas complètement efficace comme il se doit car il ne peut utiliser pour le moment, que le nom des tables et des champs. ainsi qu'un nombre limité de mots.

Pour ce qui nous concerne, ce travail nous a permis non seulement de mettre en pratique les notions théoriques acquises mais aussi d'en découvrir beaucoup de moyens pour répondre aux besoins de logiciel, ces moyens sont multiples allant des configurations et installation d'un serveur de données MYSQL et langage Perl, TK pour les interfaces graphique.

Intégrer l'ensemble de ces outils était une grande découverte pour nous et une très riche expérience que nous espérons mettre au profit de notre pays.

BIBLIOGRAPHIES

- [1] Mehdi Embarek, Un système de question-réponse dans le domaine médical : le système Esculape, Thèse pour obtenir le grade de Docteur de l'Université Paris-Est, 2008.
- [2] Luc Bélanger, Architecture question-réponse pour l'automatisation des services d'information, Thèse présentée à la Faculté des études supérieures en vue de l'obtention du grade de Philosophie Doctor (Ph. D.) en informatique, Université de Montréal, 2006.
- [3] Green, Jr., Bert F., Alice K. Wolf, Carol Chomsky et Kenneth Laughery. « Baseball : an automatic question answerer ». Dans *Computers & thought*. McGraw-Hill,
- [4] Woods, Williams A., Richard M. Kaplan et Bonnie Lynn Nash-Webber. *The Lunar Sciences Natural Language Information System : Final Report*. BBN Report 2378, Bolt Beranek and Newman Inc., Cambridge, MA, juin 1972.
- [5] J. MAZLACK, J; A. FEINAUER, R: Surface analysis of queries directed toward a database. North Holland Publishing Company 1982
- [6]. Androutsopoulos, G.D. Ritchie, and P. Thanisch, Natural Language Interfaces to Databases - An Introduction, *Journal of Natural Language Engineering* 1 Part 1 (1995)
- [7] ZWEIGENBAUM, P ; POIBEAU, T ; NAZARNKO, A. Traitement automatique des langues pour les systèmes de question/réponse. Document de travail élaboré dans le cadre de l'action spécifique RIP-WEB, 2003
- [8] KNOWLES, S; MITROVIC, T. A natural language database interface for SQL tutors. 5th November, 1999
- [9] Androutsopoulos I., Ritchie G., Thanisch, P. 1993. "An Efficient and Portable Natural Language Query Interface for Relational Databases". *Proceedings of the 6th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems*, Edinburgh, U.K., pages 327-330. Gordon and Breach Publishers Inc., Langhorne, PA, U.S.A.
- [10] Androutsopoulos, I. 1993. "Interfacing a Natural Language Front-End to a Relational Database (MSc thesis)". Technical paper 11, Dept. of AI, Univ. of Edinburgh.
- [11] Filipe, P. 1999, "Sistema de Interrogações em Língua Natural para Bases de Dados: Modelo Conceptual, Aquisição de Vocabulário e Tradução", M.Sc. Dissertation. Instituto Superior Técnico, Universidade Técnica de Lisboa.
- [12] A. Dister. "Réflexion sur l'homographie et la désambiguïsation des formes les plus fréquentes". In *Actes des Journées Internationales d'Analyse Statistique des données Textuelles (JADT 2000)*, 2000. Université de Liège -1b, Quai Roosevelt, Liège - Belgique.

BIBLIOGRAPHIES

- [13] P. Zweigenbaum, N. Grabar , S. Darmoni. "L'apport de connaissances morphologiques pour la projection de requêtes sur une terminologie normalisée".TALN 2001, Tours, 2-5 juillet 2001.
- [14] C. Jacquemin, P.Zweigenbaumu. Chapitre quatrième : " Traitement automatique des langues pour l'accès au contenu des documents? " Le document en sciences du traitement de l'information, chapitre 4, pages 71-109. Cepadues, Toulouse, 2000 .
- [15] J.Kamps. "Visualizing WordNet Structure". Institute for logic, language and computation university of amsterdam nieuwe achtergracht 166 1018wv amsterdam the netherlands 2002.
- [16] WARREN, H.D; FERNANDO, C. N. An Efficient Easily Adaptable System for Interpreting Natural Language Queries American Journal of Computational Linguistics, 1982, Volume 8, Number 3-4, July-December.
- [17] GROSZ, J. TEAM: a transportable natural language interface system. in Proceedings of the 1st Conference on Applied Natural Language Processing, Santa Monica, California.
- [18] ANDROUTSOPOULOS, I; RITCHIE, G.D; TANISH, P Natural Language Interfaces to Databases An Introduction. Natural Language Engineering, vol 1, 1995

منذ بداية عصر الإعلام الآلي كان أهم إشغالات المستخدم تخزين كميات هائلة من المعطيات و إستعمالها وأبضا إستعراضها بطريقة منتظمة .

في حين أن مستخدم قواعد البيانات غالبا ما يكون غير متخصص في الإعلام الآلي , مما يجعل الحوار بين المستخدم والنظام صعبا. لذلك كانت فكرة هذه المنكرة هو تطوير نظام سؤال -جواب يسمح للمستخدم بطرح سؤال باللغة الطبيعية ويحصل على الإجابة من قاعدة البيانات, إن إنشاء نظام لإستجواب قواعد البيانات باللغة البشرية , عملية معقدة نتيجة لتتوع المصادر المعارف المستخدمة من طرف النظام والحاجة الى تبادل المعلومات بين مختلف عناصره .

بحثنا هذا يندرج ضمن مشروع المسائلة عن قاعدة البيانات ويتمثل في إقتراح هندسة لنظام يسمح بإستخدام اللغة البشرية. يسمح النظام للمستخدم بطرح السؤال باللغة الطبيعية ثم يقوم بترجمتها الى عبارة بلغة SQL ويعرض النتيجة للمستخدم ومن أجل إنشاء تصور لبنية هذا النظام نحتاج , Mysql لإنشاء قاعدة بيانات و لغة برمجة perl واستعمال واجهة TK الكلمات المفتاح :نظام سؤال-جواب ,لغة طبيعية قواعد تحليل الاسئلة ,قواعد البيانات

Résumé

Dés les débuts de l'informatique, l'un des soucis majeurs de l'utilisateur fut de pouvoir stocker massivement des données et de pouvoir en disposer régulièrement afin d'en extraire de nouvelles.

Cependant, l'utilisateur de bases de données est bien souvent un non informaticien. Un dialogue homme-machine trop rigide syntaxiquement, l'interrogation et la génération de réponse est très compliquée. Rend parfois difficile l'emploi d'un tel système. Ceci a généralement pour effet de limiter le nombre d'utilisateur potentiels,

Notre travail s'inscrit dans le projet interrogation en langage naturel d'une base de données il consiste à proposer une architecture pour un système qui permet d'utiliser le langage naturel pour exploiter une base de données. Il reçoit en entrée une question en langue naturelle formulée par l'utilisateur du système qui la traduit en une requête SQL.

Pour réaliser notre système, on a utilisé le langage Perl et le SGBD (MYSQL) pour le développement de la base de données. Ainsi que TK pour les interfaces graphique.

Mots clés : système question-Réponse(SQR), Langage Naturel(LN),SQL(Structured Query Language),Base de données.

Summary

Since the beginning of computer science, one of the major concerns of the user was able to store massive amounts of data and to be able to have periodically in order to extract new. However, the user of databases is often a non-computer scientist. A man-machine dialog too rigid syntactically, the query and the generation of answer is very complicated. Sometimes makes it difficult to use such a system. This generally has the effect of limiting the number of potential user, our work is part of the project query in natural language of a database it is to proposed an architecture for a system that allows you to use the natural language to exploit a database. It receives in input a question in natural language formulated by the user of the system that translated into a SQL query For realized our system, it has used the Perl language and the DBMS (MYSQL) for the development of the data base. As well which TK for the graphical interfaces.

Keywords: system question-answer(QAS) ,Natural Language(LN) ,SQL(Structured Query Language) ,Database.