Introduction :

Après un détaille sur le web sémantique et les ontologies, nous visons dans ce chapitre à présenter un état de l'art spécifié aux différentes approches existantes représentant les travaux connexes du domaine, avec la citation de quelques modèles de moteurs de recherche sémantique connus par les experts.

2.1 .Description :

Un problème important des utilisateurs des bases de données, en général, est la formulation et la soumission des requêtes aux systèmes de gestion des données pour rechercher des réponses à leurs besoins. Pour formuler leurs requêtes, les utilisateurs ne disposent que d'une connaissance limitée des données à interroger. Ce problème devient encore plus complexe lorsque les utilisateurs se trouvent dans un environnement de bases de données distribuées avec des sources distantes et hétérogènes, dont les accès sont, très souvent, règlementés.

De plus, pour faciliter la communication et l'intégration des données entre les sources hétérogènes, des ontologies ont vu le jour. L'intégration de données est réalisée à partir des schémas des données représentés dans des ontologies. Cette évolution accroit également la complexité de l'interrogation et de l'exploitation de ces bases de données. Dans ce contexte de données volumineuses, distribuées, hétérogènes, dynamiques et liées, rend la tache de satisfaire l'utilisateur plus complexe.

Les systèmes d'interrogation existants offrent principalement aux utilisateurs les trois services suivants :

- ✓ l'interrogation avec des requêtes dans les langages tels que SQL ou SPARQL ;
- ✓ l'interrogation des données par des applications dédiées ;
- ✓ l'interrogation des données via des interfaces graphiques telles que des formulaires ou des langages graphiques de composition de requêtes qui sont plus intuitifs.

Les algorithmes d'interrogation tiennent compte des différents niveaux d'indexation offerts et du niveau de connaissances de l'utilisateur sur les ontologies gérées par la base d'ontologies. L'interrogation peut se faire de deux manières :

- **Interrogation en mode simple :** permet de faire des interrogations sans se soucier des champs sur lesquels va porter la recherche. Par défaut, le système reformule la requête de l'utilisateur en la propageant sur les différents champs de recherche disponibles.

- **Interrogation en mode complexe :** Ce mode d'interrogation est conçu pour une utilisation avec des connaissances a priori sur le contenu de la base d'ontologies. On spécifie de façon explicite les champs de recherche (termes, attributs, relations).

2.2. Détail des approches existantes :

2.2.1. Interrogation en mode simple (Approches Coopératives) :

L'objectif de ces approches est d'assister les utilisateurs dans l'interrogation des données. Ils ont été intégrées ensuite dans des systèmes d'interrogation tels que les systèmes de recommandation.

Ces approches permettent d'alerter les utilisateurs en cas de problème pendant l'exécution. Ces problèmes peuvent être :

- aucune réponse n'est retournée, cas des requêtes à réponse vide (empty answer query).
- un utilisateur, familier de la base de données et aux données, n'est pas satisfait des réponses retournées par la requête. On parle alors de réponses insatisfaisantes ;
- un utilisateur qui ne comprend pas pourquoi il a obtenu des réponses mais pas celles qu'il attendait, c'est un autre aspect des réponses insatisfaisantes ;
- un utilisateur est partiellement satisfait par les réponses retournées : réponses incomplètes.

Nous pouvant citer dans cette catégorie les approches suivantes :

a .Suggestion de requêtes :

La suggestion de requêtes aide les utilisateurs à formuler leurs besoins sous forme de requêtes syntaxiquement et sémantiquement correctes. Elle a initialement été proposée pour éviter aux utilisateurs de formuler des requêtes qui ne pourront pas être satisfaisantes à cause des fausses suppositions. Deux types de fausses suppositions peuvent conduire à l'insatisfaction des utilisateurs. Le premier type est l'ensemble des suppositions faites par les utilisateurs sur le schéma des données, on parle de présupposition. Le deuxième type est l'ensemble des suppositions sur les données que les utilisateurs pensent exister ou pas dans la base de données.

b. Interrogation par l'exemple :

Malgré l'aide apportée par les techniques de suggestion de requêtes dans la formulation des requêtes, il est montré que la conversion des besoins en requêtes est une tâche complexe. En effet, pour les utilisateurs n'ayant qu'une connaissance imprécise de leurs besoins, la formulation des requêtes correspondantes à ces besoins est très complexe. L'interrogation par l'exemple (Query by Example, QBE) permet d'alléger cette difficulté. L'approche d'interrogation par l'exemple a été initialement conçue par IBM comme un langage d'interrogation pour les données relationnelles. Le concepteur de cette approche, a eu l'idée d'un langage d'interrogation pour les non-programmeurs, afin de permettre aux utilisateurs de formuler leurs requêtes avec précision et d'éviter toutes fausses présuppositions. Cette nouvelle approche d'interrogation été conçu en primo pour les données relationnelles, et par la suite, elle été adaptée pour les autres types de données, notamment RDF et XML.

c. Explications sur les réponses d'une requête :

Les approches présentées dans cette section ont pour but de fournir aux utilisateurs des explications aux questions qu'ils peuvent se poser devant les réponses retournées par leurs requêtes. Ces approches étaient, initialement, destinées à traiter des questions provenant de l'insatisfaction des utilisateurs. Cette insatisfaction se traduisait par la question : (pourquoi n'ai-je pas obtenu la réponse X que j'attendais ?), plus connue sous le terme (Why-Not Anwsers) . Les utilisateurs se posent cette question lorsque l'ensemble des réponses retournées ne contient pas celles qu'ils attendaient. Ainsi, l'explication consiste à fournir aux utilisateurs les raisons pour lesquelles les réponses attendues ne se trouvent pas dans l'ensemble des réponses retournées. Par la suite, d'autres types d'explications de réponses de requête ont été développés. Il s'agit de l'explication de la présence de certaines réponses, inattendues, dans l'ensemble de réponses retournées (WhyAnswers ?), de l'explication des réponses vides (Why Empty Answers ?) et de l'explication des réponses pléthoriques (Why Too Many Answers ?). Cette section présente ces différentes approches d'explication des réponses d'une requête.

d. Raffinement de la requête par l'utilisateur (Users refinement) :

L'approche de raffinement de la requête par l'utilisateur (Users refinement) permet avant tout la réparation de la requête. Elle permet aux utilisateurs de redéfinir leurs requêtes de façon itérative jusqu'à être satisfait par les réponses retournées. Nous rappelons que

l'ensemble des réponses peuvent ne pas satisfaire les utilisateurs parce qu'il ne contient pas les réponses attendues ou parce qu'il contient des réponses indésirables. Les utilisateurs peuvent également être insatisfaits du nombre de réponses retournées.

Dans tous ces cas, la requête doit être réparée et le " Users refinement " peut le faire de deux façons :

- les techniques interactives font intervenir les utilisateurs et exploitent leurs retours.
- les techniques automatiques sont basées sur les requêtes, les données et une modélisation de l'utilisateur.

e. Redéfinition des requêtes :

La redéfinition d'une requête reformule les contraintes de cette requête afin de retourner des réponses satisfaisantes. Il existe trois types de reformulation de requête, ou modification de requête :

- ✓ la restriction de la requête : la requête issue de cette modification est plus restrictive que la requête initiale. Toutes ses réponses sont contenues dans celles de la requête initiale ;
- ✓ la relaxation de la requête : la requête issue de cette modification est moins contraignante que la requête initiale. Ses réponses incluent toutes celles de la requête initiale et en rajoutent ;
- ✓ l'approximation de la requête : la requête issue de cette modification est une approximation de la requête initiale. Elle contient une partie des réponses de la requête initiale et rajoute d'autres réponses qui ne sont pas inclues dans celles de la requête initiale.

Remarque**:** Il existe une relation entre l'approximation et la relaxation de requête. Certains travaux existants exploitent l'approximation d'une requête pour réaliser la relaxation de cette requête.

2.2.2. Interrogation en mode complexe (Approches basées sur la structure des ressources) :

Avec le développement des entrepôts de triplets RDF, un besoin d'interfaçage des moteurs SPARQL a émergé, puisqu'il est impossible pour un utilisateur final de maîtriser la complexité des schémas de ces connaissances : pour pouvoir exprimer une requête valide sur les connaissances du web sémantique, l'utilisateur doit connaître le langage SPARQL, les ontologies utilisées pour exprimer les triplets qu'il veut interroger ainsi que la "forme" des graphes RDF considérés.

Dans ce contexte, nous pouvons citer les approches suivantes :

- Le travail de Bhalotia, G., Hulgeri, A., Nakhe, C., Chakrabarti, S., Sudarshan [BHA, HUL, NAK 02] démarche la recherche de chemin le plus court sur chaque élément correspondant afin de trouver une arborescence de jointure.
- L'approche DBExplorer de l'équipe de Chaudhuri [CHA, DAS, NAR 02] et l'approche Discover de l'équipe de Hristidis [HRI PAP 02], profitent du schéma de données pour calculer un ensemble de réseaux de jointure possibles.
- Les travaux de N.Athanasis et VC.Kotzinos [ATH VC KOT 04], proposent une recherche basée sur des graphiques ils ont construit des requêtes graphiques a travers la navigation et la sélection sur l'ontologie.
- Les auteurs comme COMPAROT C., HAEMMERLÉ O. & HERNANDEZ N. [COM, HAE, HER 10] ont proposé un moyen de construire des requêtes exprimées en graphes conceptuels à partir d'une requête utilisateur exprimée sous forme de mots-clés, puis ils ont étendu le système afin qu'il prenne en considération les relations exprimées par l'utilisateur entre les mots-clés utilisés dans sa requête, ensuite ils ont présenté de manière précise la notion de patron de requêtes.

2.3. Les approche selon le degré de formalité :

Depuis quelques années, plusieurs travaux se sont intéressés à l'utilisation des liens sémantiques entre les termes afin d'enrichir les requêtes. Ces travaux utilisent un vocabulaire structuré allant d'une simple taxonomie à l'utilisation d'ontologies complexes.

- ✓ D'après E. M. Voorhees [VOO 94], les méthodes statistiques ont moins de succès que les méthodes utilisant les relations sémantiques et linguistiques. Il propose une méthode d'enrichissement de requêtes basée sur les relations sémantiques entre les termes (synsets) dans Wordnet .
- ✓ Selon Zhiguo Gong, Chan Wa Cheang, and Leong Hou U. [GON, CHA, LEO 05] qui utilisent les relations hyperonymie, hyponymie et les relations de synonymie de Wordnet pour l'enrichissement de la requête.
- ✓ Selon Shuguang Wang and Milos Hauskrecht [WAN, MIL 10], qui proposent une nouvelle mesure pour calculer la similarité entre termes afin d'enrichir les requêtes. Une mesure de similarité est calculée en interprétant le graphe de termes comme un réseau électrique.
- ✓ Le travail de P. Sebillot V. Claveau [SEB, CLA 04] décrit comment les liens sémantiques entre les noms et verbes peuvent améliorer les résultats des systèmes de

recherche d'information, la requête étendue se compose donc des termes de la requête originale et des verbes qualia. Un lien qualia relie un nom et un verbe relies sémantiquement (par exemple couper est la fonction du terme couteau).

- ✓ Le travail de R. Navigli and P. Velardi [NAV,VEL 03] propose une approche qui consiste a extraire des concepts et relations d'une ontologie et les ajouter dans la requête initiale en se basant sur les liens entre les termes.
- ✓ Les chercheurs dans Aime X., Furst F., Kuntz P., and Trichet F [AIM,FUR,KUN,TRI 10] ont utilisé une ontologie personnalisée pour une recherche d'information sémantique en offrant a l'utilisateur une interrogation par concepts avec l'ensemble des termes décrivant le concept.
- ✓ Afin d'améliorer la recherche dans le Web sémantique et permettre a l'utilisateur de créer des requêtes simples, le système SPARK proposé par Zhou, Q., Wang, C., Xiong, M., Wang, H., Yu [ZHO,CHO,MIA,HAO,YON 07] traduit des mots clés d'une requête en requêtes Sparql.

La classification des approches proposée par KAUFMANN E. & BERNSTEIN A. [KAU, BER 10] (voir tableau 2.1) , selon un continuum du plus formel au plus permissif, en commençant par les langage de requêtes eux-mêmes :

Approches	Exemples	Observations
Les langages formels de requêtes graphes	SPARQL SeREQL	Inadaptés aux utilisateurs finals
Des moteurs qui assistent intégralement l'utilisateur pendant la formulation de la requête. (implémentent de simples fonctionnalités comme la coloration syntaxique et l'auto-complétion)	Flint SparQLed	Inadaptés aux utilisateurs finals
Approches qui reposent sur des interfaces graphiques (contribue également, en construisant des requêtes graphiques a travers la navigation et la sélection sur l'ontologie.)	specific-to-general pour SPARQL	Ne dépasse pas les limites d'utilisabilité des langages de requête formels.
Introduit la recherche par facette à base de requêtes (un paradigme combinant la recherche par facette et le requêtage)	Opérateur PICO dans le secteur de santé.	Facile à exploiter par la plus part des utilisateurs
Génération automatique des requêtes avec retour aux utilisateurs pour plus de raffinement	Moteur de recherche Evi	Plus adaptés aux utilisateurs finals

Tableau 2.1 : Classification des approches par KAUFMANNE. & BERNSTEIN A.

2.4. Modèles de Moteurs sémantique existants :

On peut citer quelques applications représentatives de la recherche sémantique à base des mots-clés :

- ✓ SemSearch [SEM 06]: a une petite interface structurée de requête mots-clés pour cacher la complexité de la recherche sémantique.
- ✓ Avatar SemanticSearch [SEM 06]: est un prototype de moteur de recherche qui exploite les annotations dans le contexte de la recherche par mot-clé classique.
- ✓ CORESE : Ce moteur est dédié à des applications de web sémantique communautaire ou d'entreprise. Le moteur Corese permet la recherche d'information basée sur des modélisations de connaissances, des modèles conceptuels ou des référentiels métier et organisationnel (voir figure 2.1).



Figure 2.1 : L'interface du moteur Corese [CHA, NAD, OLI,WIL 04]

- ✓ Evi: de la société Amazon, l'utilisateur peut enrichir l'outil avec ses propres connaissances, liens vers DBPedia, Wikipedia, Freebase, ... (voir figure 2.2)

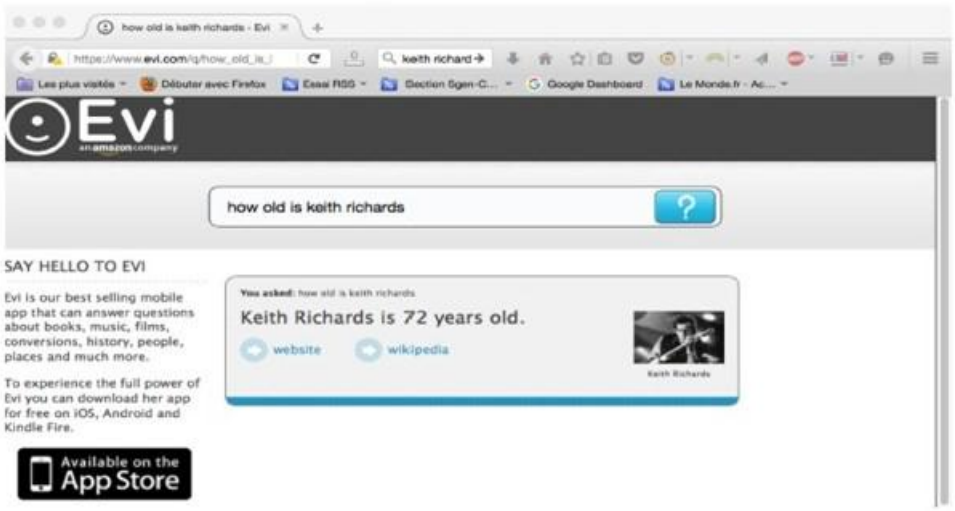


Figure 2.2 : L'interface du moteur Evi [08]

- ✓ Wolframalpha [09] : Contrairement à Google, Yahoo et consorts, WolframAlpha ne restitue pas des informations piochées sur le Web. Son rôle : compiler, croiser et présenter synthétiquement les données issues de bases de connaissances, dont celle de Mathematica, et d'autres sources communiquées à chaque résultat de recherche. Il fournit alors des « fiches pratiques » aux utilisateurs (voir figure 2.3)



Figure 2.3 : L'interface du moteur wolframalpha [09].

2.5 . Conclusion :

Nous avons vu au sein de ce chapitre un balayage sur les différentes approches représentées par des systèmes qui offrent aux utilisateurs des services remarquables afin de réussir à interroger les ontologies sur le web.

On a distingué deux modes d'interrogation par mots-clés (simple et complexe) suivi des travaux connexes justifiés par quelques exemples des moteurs de recherches spécialement sémantique .