

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE Mathématiques, Informatique,
Physique Théorique et Ingénierie des Systèmes (MIPTIS)
Laboratoire d'Informatique Fondamentale d'Orléans (LIFO)

THÈSE présentée par :

Nicolas HIOT

soutenue le : XX mois en lettres 202x

pour obtenir le grade de : Docteur de l'Université d'Orléans

Discipline/ Spécialité : Informatique

TITRE

[Sous titre éventuel]

THÈSE dirigée par :

Mme Mirian HALFELD FERRARI ALVES

Professeur, Université d'Orléans

RAPPORTEURS :

[Civilité NOM Prénom]
[Civilité NOM Prénom]Titre, établissement
Titre, établissement

JURY :

[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]
[Civilité NOM Prénom]Titre, établissement, Président du jury
Titre, établissement
Titre, établissement
Titre, établissement
Titre, établissement
Titre, établissement
Titre, établissement
Titre, établissement

Remerciements

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Table des matières

Liste des figures	iv
Liste des tables	v
I Introduction	1
II Maintenance et cohérence d’une base de donnée incomplète	3
1 Introduction	5
2 État de l’art	7
3 Maintien de la cohérence	9
4 Mise à jour incrémentale	11
4.1 Préliminaires	11
4.1.1 Logique du premier ordre	11
4.1.2 Base de données	11
4.1.3 Base de données graphes	12
4.2 Graph Database Design in Neo4J	12
5 Conclusion	13
III Intégration de données textuelles	15
6 Introduction	17
7 Structuration d’informations	19
8 Extraction d’informations	21
9 Conclusion	23
IV Conclusion et perspectives	25
A Annexe 1	27
B Annexe 2	29

Bibliographie	31
Glossaire	31
Acronymes	31

Table des figures

Liste des tableaux

Première partie

Introduction

Deuxième partie

Maintenance et cohérence d'une base de donnée incomplète

Chapitre 1

Introduction

Chapitre 2

État de l'art

Chapitre 3

Maintien de la cohérence

Chapitre 4

Mise à jour incrémentale

4.1 Préliminaires

4.1.1 Logique du premier ordre

En assumant une certaine familiarité avec la logique du premier ordre, on utilise comme alphabet :

- un ensemble de prédicats P, Q, R, \dots
- un ensemble de variables x, y, z, \dots
- un ensemble de constantes a, b, c, \dots
- un ensemble de valeurs nulles marquées N_1, N_2, N_3, \dots
- le quantificateur universel \forall (quel que soit)
- le quantificateur existentiel \exists (il existe)

Avec cet alphabet, on définit un atome $P(t_1, \dots, t_n)$ où P est le symbole de prédicat d'arité n et t_1, \dots, t_n sont les termes. Les termes peuvent être une variable, une constante ou une valeur nulles. Un fait est un atome composé uniquement de constantes et un atome instancié est un atome qui ne contient aucunes variables. Étant donné un atome instancié A , $null(A)$ représente l'ensemble des valeurs nulles de A .

Un homomorphisme h entre deux ensembles d'atomes A et B est une application surjective sur les termes de A vers les termes de B telle que : (i) si, $h(t) = t$, alors t est une constante et (ii) si, $P(t_1, \dots, t_n) \in A$, alors $P(h(t_1), \dots, h(t_n)) \in B$. S'il existe un homomorphisme inverse h^{-1} tel que $h^{-1}(h(t)) = t$ alors on dit que A est *isomorphe* à B .

Φ est l'ensemble de toutes les formules ϕ , existentiellement quantifiées, de la forme $(\exists X)(\varphi_1(X_1) \wedge \dots \wedge \varphi_n(X_n))$ où X est l'ensemble des variables et $\varphi_i(X_i)$ est une formule atomique dont $X_i \subset X$ représente l'ensemble des variables libres. $atoms(\phi)$ est l'ensemble des formules atomiques de ϕ . Un modèle M d'une formule ϕ est l'ensemble des faits qui ont un homomorphisme de $atoms(\phi)$ vers M . Étant donné deux formules ϕ_1 et ϕ_2 : $\phi_1 \implies \phi_2$ si tout modèle de ϕ_1 est un modèle de ϕ_2 . ϕ_1 et ϕ_2 sont *équivalents*, noté $\phi_1 \iff \phi_2$, si et seulement si $\phi_1 \implies \phi_2$ et $\phi_2 \implies \phi_1$. ϕ_1 est *plus simple que* ϕ_2 , noté $\phi_1 \sqsubseteq \phi_2$ si $\phi_1 \iff \phi_2$ et $atoms(\phi_1) \subseteq atoms(\phi_2)$. Une simplification ϕ_1 de ϕ_2 est *minimale* si $\phi_1 \sqsubseteq \phi_2$ et qu'il n'existe pas de ϕ'_1 tel que $\phi'_1 \sqsubset \phi_1$. Par exemple, étant donné la formule $\phi = (\exists x, y)(P(a, x) \wedge P(a, y))$, les formules $(\exists x)(P(a, x))$ et $(\exists y)(P(a, y))$ sont équivalentes et sont toutes deux une simplification de ϕ . De plus, [?] démontre que si ϕ_1 et ϕ_2 sont deux simplifications minimales d'une même formule ϕ alors les ensembles $atoms(\phi_1)$ et $atoms(\phi_2)$ sont isomorphes. La simplification minimale d'une formule ϕ est appelé le core et s'écrit $core(\phi)$.

4.1.2 Base de données

L'*instance* d'une base de donnée peut être décrite comme la Skolémisation d'une formule ϕ correspondant à un ensemble minimal d'atomes instanciés tel que $core(\phi) = \phi$. La Skolémisation d'une formule ϕ noté $Sk(\phi)$ consiste à instancier l'ensemble des variables par des constantes ou des valeurs nulles et à éliminer toutes les occurrences de quantificateur existentiel.

Une *contrainte* (ou règle) sur une base de donnée est une dépendance génératrice de n-uplets (ou Tuple-generating dependency, TGD en anglais) de la forme $(\forall X, Y)(body(X, Y) \rightarrow (\exists Z)head(X, Z))$.

Plus simplement écrit $body(X, Y) \rightarrow head(X, Z)$, X , Y et Z sont des ensembles de variables, $bod(X, Y)$ est un ensemble d'atomes et $head(X, Z)$ est un atome. Une contrainte c est dite satisfaite par un ensemble I d'atomes instanciés, noté $I \models c$ si pour tout homomorphisme h tel que $h(body(c)) \subseteq I$ il existe un homomorphisme h' tel que $h'(head(c)) \subseteq I$. Intuitivement, cela signifie que pour qu'une contrainte soit satisfaite, s'il existe une instance du corps dans l'instance de la base de données, la tête correspondante doit aussi être présente. Une instance I d'une base de données est dite *cohérente* par rapport à un ensemble de contraintes \mathcal{C} , noté $I \models \mathcal{C}$ si pour toute contrainte c dans \mathcal{C} , $I \models c$.

4.1.3 Base de données graphes

Quand on parle de graphe on peut distinguer deux grandes familles :

Les bases de triplets principalement le Resource Description Framework (RDF) porté par le World Wide Web Consortium (W3C). Ce type de base de données enregistre des triplets de la forme $\langle x, P, y \rangle$ qui sont des prédicats binaires où P est un symbole de prédicat, x , le sujet, est l'URI d'une ressource (un noeud) et y , l'objet est un URI ou un littéral (une valeur de propriétés).

Les Graphes de Propriétés , en anglais Labeled Property Graph (LPG)

4.2 Graph Database Design in Neo4J

La proposition de mise à jour incrémentale sur des bases de données graphes est implémentée sur le gestionnaire de base de données graphes Neo4J.

Chapitre 5

Conclusion

Troisième partie

Intégration de données textuelles

Chapitre 6

Introduction

Chapitre 7

Structuration d'informations

Chapitre 8

Extraction d'informations

Chapitre 9

Conclusion

Quatrième partie

Conclusion et perspectives

Annexe A

Annexe 1

Annexe B

Annexe 2

Acronymes

LPG Labeled Property Graph. 12

RDF Resource Description Framework. 12

W3C World Wide Web Consortium. 12

Nicolas HIOT

[TITRE de la thèse] (en français)

Résumé : (4000 caractères max.)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin volutpat ipsum id purus ultrices lobortis. Maecenas ornare enim quis eros. Nunc eget mauris ut quam malesuada mattis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Integer vel tellus. Nam rutrum, purus non sodales rhoncus, quam magna imperdiet eros, sit amet euismod justo metus atorci. Suspendisse neque turpis, feugiat interdum, faucibus vel, aliquet quis, risus. Etiam est elit, eleifend a, consequat sit amet, scelerisque nec, odio. Quisque id odio quis libero iaculis tincidunt Sed non mi. Morbi aliquam commodo nibh. Integer justo purus, pulvinar a, suscipit vel, iaculis a, justo. Morbi ut orci. Maecenas fringilla orci. Phasellus auctor, enim vitae tempus egestas, justo mic cursus sem, vel blandit leo turpis vitae quam. Etiam sit amet felis vitae eros ornare porttitor. Proin orci ligula, vehicula non, ultrices at, ultrices ut, massa. Vestibulum ac est. Curabitur at erat. Mauris gravida. Praesent vestibulum. Curabitur eget orci ac massa cursus condimentum. Integer sapien dui, ultricies eget, dapibus a, dapibus id, mauris. Curabitur felis velit, aliquam at, aliquet in, iaculis vitae, velit. Nunc lobortis magna id ligula. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Integer congue ultrices mi. Isdem diebus Apollinaris Domitiani gener, paulo ante agens palatii Caesaris curam, ad Mesopotamiam missus a socero permilitares numeros immodice scrutabatur, an quaedam altiora meditantis iam Galli secreta susceperint scripta, qui conpertis Antiochiae gestis per minorem Armeniam lapsus Constantinopolim petit.

Mots clés : mot 1, mot 2, mot3, ...

[TITRE de la thèse] (en anglais)

Abstract : (4000 caractères max.)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin volutpat ipsum id purus ultrices lobortis. Maecenas ornare enim quis eros. Nunc eget mauris ut quam malesuada mattis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Integer vel tellus. Nam rutrum, purus non sodales rhoncus, quam magna imperdiet eros, sit amet euismod justo metus atorci. Suspendisse neque turpis, feugiat interdum, faucibus vel, aliquet quis, risus. Etiam est elit, eleifend a, consequat sit amet, scelerisque nec, odio. Quisque id odio quis libero iaculis tincidunt Sed non mi. Morbi aliquam commodo nibh. Integer justo purus, pulvinar a, suscipit vel, iaculis a, justo. Morbi ut orci. Maecenas fringilla orci. Phasellus auctor, enim vitae tempus egestas, justo mic cursus sem, vel blandit leo turpis vitae quam. Etiam sit amet felis vitae eros ornare porttitor. Proin orci ligula, vehicula non, ultrices at, ultrices ut, massa. Vestibulum ac est. Curabitur at erat. Mauris gravida. Praesent vestibulum. Curabitur eget orci ac massa cursus condimentum. Integer sapien dui, ultricies eget, dapibus a, dapibus id, mauris. Curabitur felis velit, aliquam at, aliquet in, iaculis vitae, velit. Nunc lobortis magna id ligula. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Integer congue ultrices mi. Isdem diebus Apollinaris Domitiani gener, paulo ante agens palatii Caesaris curam, ad Mesopotamiam missus a socero permilitares numeros immodice scrutabatur, an quaedam altiora meditantis iam Galli secreta susceperint scripta, qui conpertis Antiochiae gestis per minorem Armeniam lapsus Constantinopolim petit.

Keywords : word 1, word 2, word 3, ...