République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira de Bejaia
Faculté des Sciences Exactes
Département d'Informatique

ECOLE DOCTORALE RESEAUX ET SYSTEMES DISTRIBUES

Mémoire de Magistère

En Informatique

Option : Réseaux et Systèmes Distribués

Thème

Les services de la bibliothèque numérique « Digital Library » et leur interopérabilité basés sur la technologie GRID

Présenté par

Amel BOUFENISSA

Devant le jury composé de :

Président	Abdelkamel TARI	MC/A	Université de Bejaïa
Rapporteur	Nadjib BADACHE	Professeur	CERIST
Examinateur	Omar NOUALI	DR	CERIST
Examinateur	Makhlouf ALIOUAT	MC/A	Université de Sétif
Invité	Aouaouache EL MAOUHAB	CR	CERIST

Promotion : 2009/2010

REMERCIEMENTS

C'est avec un réel plaisir et un grand enthousiasme que je me livre à la rédaction de cette page. Bien plus que le point final du manuscrit, cette page constitue l'opportunité de m'accorder une réflexion sur une période de ma vie très riche en émotions.

Je remercie en premier notre grand Dieu pour m'avoir donné le courage et la volonté pour terminer ce modeste travail.

Un grand merci à mon directeur de thèse, le Professeur Nadjib BADACHE, directeur du CERIST. Je lui suis reconnaissante de m'avoir donné l'opportunité de préparer mon diplôme de magistère.

Je tiens à exprimer ma très vive reconnaissance envers Madame Aouaouache EL-MAOUHAB, chargée de recherche au CERIST, pour la qualité de son encadrement. Ses compétences scientifiques, mais aussi ses qualités humaines ont été des éléments précieux pour l'avancement de mon travail. Je lui suis reconnaissante pour sa patience, ses encouragements, sa sympathie et sa disponibilité à tout moment.

Je suis vivement reconnaissante à Monsieur Abdelkamel TARI d'avoir accepté la charge de président du jury.

Je tiens à exprimer mes sincères remerciements aux Messieurs Omar NOUALI et Makhlouf ALIOUAT, qui m'ont fait l'honneur d'évaluer ce travail et qui, par la qualité et la pertinence de leurs remarques, permettront d'améliorer la rédaction de cette thèse. Je les remercie sincèrement de m'avoir permis de profiter de leurs compétences et de leurs connaissances, et l'honneur qu'ils m'ont accordé en acceptant la charge de rapporter mon travail.

J'exprime également toute ma gratitude à ma très chère amie Faiza, pour son aide précieux, ses encouragements et sa disponibilité durant toute la durée de préparation de mon mémoire.

Je remercie très particulièrement ma chère maman et mon mari pour m'avoir soutenu et encouragé pendant toute la durée de la préparation de cette thèse. Leur soutien et amour sans limites ont été essentiels. Cette thèse est aussi le fruit de leur effort.

Un grand MERCI à Mon Papa, Ma sœur et Mes frères, Mes tantes et Mes oncles pour leurs encouragements afin que ce travail arrive à sa fin.

Je remercie aussi mes beaux-parents pour m'avoir encouragé.

Je ne pourrais terminer ces remerciements sans y associé mes amis du CERIST, en particulier, Lamia, Houda et Sahar, et tant d'autre sans leurs soutien je n'aurai pu entreprendre ces études.

Et pour être sûr de n'oublier personne, que tous ceux, qui de près ou de loin, ont contribué par leurs conseils, leurs encouragements ou leur amitié, à l'aboutissement de ce modeste travail, trouvent ici l'expression de ma profonde reconnaissance.

RÉSUMÉ

Plusieurs bibliothèques numériques utilisent le modèle de la recherche distribuée. Sur un certain nombre de sites, les modèles restent satisfaisants. A partir d'un nombre plus important, le modèle montre ses limites et d'autres modèles s'y prêtent mieux. Le plus adapté reste le modèle « harvest metadata » qui est basé sur l'Open Archive Infrastructure / PMH. Dans un contexte d'une collection de bibliothèques numériques, les performances d'une telle infrastructure qui repose sur la découverte de ressources informationnelles et sur l'exécution d'indexation nécessitent de considérer une base architecturale basée sur la haute performance.

La haute performance est une contrainte qui devient de plus en plus solvable sur des architectures de grilles de calcul et de stockage. Dans des environnements dynamiques et hétérogènes, les architectures GRID offre des fonctionnalités de partage de ressources, de stockage et de gestion des données distribuées sur ces clusters GRID.

A cet effet, l'objet de ce travail est de montrer la faisabilité de l'intégration de la technologie des OAI avec leur interopérabilité dans un contexte large avec la technologie GRID. Pour cela, nous avons donc étudié les technologies associées à cette thématique et validé l'approche d'intégration par une implémentation sur un environnement GRID.

MOTS-CLÉS : Bibliothèques numériques, archives ouvertes, interopérabilité, OAI-PMH, grille de calcul.

ABSTRACT

Several digital libraries use the model of distributed research. On a certain number of sites, the model still satisfied. From a larger number, the model shows its limits and other models are better suited. The most suitable model is the "harvest metadata" that is based on the Open Archive Infrastructure / PMH. In the context of a digital library collection, the performance of such infrastructure based on the discovery of informational resources and the execution of indexation requires considering a basic architecture based on high performance.

High performance is a constraint becoming more creditworthy in grid computing and storage architectures. In dynamic environments and heterogeneous, GRID architectures provides functionality for resource sharing, storage and management of data distributed over these GRID clusters.

For this purpose, the purpose of this work is to demonstrate the feasibility of integrating OAI technology with their interoperability in a broad context with the GRID technology. For this, we have studied the technologies associated with this theme and validated the approach of integration with an implementation on a GRID environment.

KEYWORDS: Digital libraries, open archives, interoperability, OAI-PMH, grid computing.

Table des matières