

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE D'ALGER

FACULTE DES SCIENCES HUMAINES

DEPARTEMENT DE BIBLIOTHECONOMIE

MEMOIRE POUR L'OBTENTION DU DIPLOME DE MAGISTER EN
BIBLIOTHECONOMIE ET SCIENCES DOCUMENTAIRES

**Etude de la problématique de
l'incompatibilité des langages documentaires,
dans le cadre de l'hétérogénéité des bases de
données: Cas de la base *ALGERIANA***

Présenté par : DZIRI GHOUAS Anissa

Sous la Direction du : Dr. DAHMANE Madjid

2000

Avant propos

Ce mémoire représente l'aboutissement d'un long cursus universitaire, je me permet donc de remercier tous ceux qui m'ont aidé de près ou de loin pour l'achèvement de cette présente étude.

Je remercie tout d'abord Mr. Dahmane, chef du laboratoire recherche et développement en IST au CERIST, pour son encadrement, son suivi et sa patience avec lesquels il a dirigé ce mémoire.

Je remercie Mr. Allahoum, Directeur du Département bibliothéconomie (Université d'Alger) qui a bien voulu me faire l'honneur de présider le jury de soutenance.

Je remercie vivement Mr. Aarab, pour avoir accepté de faire partie du jury.

Un grand hommage est à rendre à l'ensemble des enseignants de l'institut de bibliothéconomie ainsi qu'au personnel administratif.

Je n'oublie pas également Mr. Benhamadi, directeur du CERIST pour ses encouragements ainsi que tous les collègues du laboratoire IST et le personnel du DCTD (Département collecte, traitement et diffusion) qui m'ont aidé et m'ont accordé de leur temps pour m'apporter des compléments d'information nécessaires à mon travail.

Je tiens également à exprimer mes remerciements les plus distingués à Mr. J. Maniez (Maître de conférence en linguistique à l'IUT de Dijon) et à Madame D. Degez (Consultante en Ingénierie documentaire - Cabinet documentation et gestion, Chaville) pour leurs conseils judicieux et éclaircissements sur certaines questions d'ordre technique.

Je remercie aussi, Mr. M. Khyari de l'INIST (Nancy) ; Mr. Fournier de la bibliothèque de l'Université Laval (Canada) et Mr. O. Rousseaux (BN de France) pour m'avoir aidé par leurs précieuses collaborations et orientations.

Je termine mes remerciements à tous mes amis et surtout à ma famille proche pour m'avoir encouragé et soutenu à accomplir ce présent travail.

Dziri Ghouas, Anissa. – Etude de la problématique de l'incompatibilité des langages documentaires dans le cadre de l'hétérogénéité des bases de données : Cas de la base Algeriana. – Alger : Université d'Alger, 2000. – 180p. – bibliogr. ; ann.- mem. Magister : Département de bibliothéconomie.

Résumé

La prolifération des langages documentaires vers des structures, syntaxes et vocabulaires plus développés est une condition nécessaire pour améliorer la qualité des opérations de traitement de l'information et pour rendre plus efficace les fonctions de recherche et de repérage de cette information.

Cependant cette prolifération des outils d'indexation a en même temps engendré un phénomène paradoxal, qui leur incompatibilité au moment de leurs mise en relation. En effet, ce phénomène documentaire apparaît dans certaines conditions d'échange d'information, de fusion de données ou de récupération de références étrangères à partir d'un système d'information. Pour pallier à cette problématique, d'immenses efforts se sont concentrés sur le plan international. Vu la complexité du problème, les solutions standards n'existent pas, mais quelques voies ont été explorées vers l'uniformisation des langages ayant soit la même structure soit le même domaine d'application. A ce titre c'est dans l'intégration des thesauri et la technique de la reindexation que les résultats furent les plus encourageants.

Par ailleurs, dans le contexte national, la problématique de l'incompatibilité des langages documentaires ne se pose pas de façon très accrue. Cette dernière se démontre dans un cas particulier relatif au projet de construction des bases de données nationales. La base Algeriana en est l'exemple concret.

Pour pallier à la problématique d'hétérogénéité des langages de cette base, le langage Rameau représente la solution la plus adéquate pour l'uniformisation des descripteurs. A cet effet, un questionnaire sur les caractéristiques d'Algeriana et un test de conversion des termes d'origines vers les vedettes Rameau ont été conçus pour étudier les capacités sémantique de ce langage à convertir et par là à uniformiser l'ensemble des descripteurs d'Algeriana. Les résultats de cette expérience démontre que la langage Rameau couvre un taux élevé de descripteurs ses quelques insuffisances résident dans l'absence des termes d'indexation très spécialisés et des notions spécifiques à l'Algérie. Dans ce cas d'incompatibilité, la solution du langage de référence n'est pas totalement satisfaisante néanmoins d'autres solutions peuvent y contribuer moyennant une restructuration de la base en sous bases thématiques et un développement d'interfaces de conversion automatiques.

Mots clés : Incompatibilité des langages documentaires/ Base de données hétérogènes/ Algeriana/ Liste Rameau/ Technique d'unification

Summary

The indexing language proliferation towards more developed structures, syntaxes and vocabularies is a necessary condition to improve the quality of information processing and to make more efficient the functions of research and location of relevant information.

However the proliferation of indexing language is generating a paradoxical phenomenon, which consists on their incompatibility when they make them together.

Indeed, this documentary phenomenon appears in conditions of exchange of information, or melting heterogeneous Data Bases. To resolve this problematic, vast efforts concentrated on the international level. Because the complexity of the problem, the standards solutions don't exist, but some ways have been explored towards the unification of the languages having either the same structure or the same domain of application. Nevertheless it's in the integration of the thesauri and in the process of re indexing that the results were the most encouraging

Moreover, in the national context the problematic of the incompatibility between indexing languages don't appear at intensive manner. This problem appears in a relative particular project of developing national Data Bases. The Algeriana Data Base is a true example of it. This past one is developed from integration of various heterogeneous anglo saxons and francophone Data bases, therefore we assist at very complex problem of incompatibility . Then the more adapted solution, is standardization of indexing terms from a referential language. In this case we have opt for "Rameau" subject heading. To demonstrate a semantic capacity of Rameau to convert all indexing items of Algeriana, a questionnaire is undertaken to detect features of Algeriana and one test of conversion has been conceived to a panel of original terms towards Rameau subject headings.

The results of this experience demonstrate that Rameau covers a rate raised of indexing terms, however there are some insufficiencies such the absence of specialized indexing terms and specific notions about Algeria.

In this case of indexing language compatibility the solution of the reference language is not always the best one. Nevertheless others solutions can contribute to resolve this kind of problem. It's consists for example on correspondence tables, Switching language...etc. It is to note, that application of this type of solutions requires a restructuring of the Data Base Algeriana and developing an automatic interfaces of conversion between different indexing language.

Key words

Incompatibility of indexing language/ heterogeneous data base/ Algeriana/ Rameau/ Technique of standardization

دراسة إشكالية عدم تواعم لغات توثيقة في إطار قواعد معطيات لا متجانسة : حالة قاعدة ألبيريانا . - الجزائر: جامعة الجزائر ، 2000 . - 180 ص. - مذكرة ماجستير: قسم علم المكتبات و التوثيق .

ملخص:

كان بروز وتنامي لغات التوثيق نحو بنيات و تراكيب لغوية و إصطلاحية أكثر تطورا، وراء ظاهرة عدم تواعم هذه اللغات خاصة في سياق تبادلات المعلومات . هذه الإشكالية التوثيقية تظهر في ظروف خاصة بتبادل المعطيات، ادماج قواعد لامتجانسة و إدخال بيانات خارجية داخل أنظمة توثيقية . و لبلوغ درجة قصوى من فعالية داخل أنظمة البحث و الإسترجاع للمعلومات، ينكب خبراء الإعلام و التوثيق، و ذلك منذ الستينات (1960) على دراسة مسائل توافق مختلف التصنيف و المكانز، غير أنه لا يمكن الحديث عن حلول صالحة لكل الحالات، بل فقط في إطار إدماج مكانز متخصصة أو في إطار تحويل مفردات التشفيف من لغة إلى أخرى . دراسة هذه الظاهرة في سياق التوثيق الوطني تستمد جدواها في إطار بناء قواعد المعطيات الوطنية. نخص بالذكر قاعدة المعطيات ألبيريانا . بالفعل و بحكم تصميمها على أساس دمج مجموعات بيانات مستخرجة من مصادر معلوماتية مختلفة، فإن هذه القاعدة تشكل أرضية خصبة بحدوث عدم تجانس مفردات التشفيف. الحل المقترح في هذه الحالة هو إعادة التشفيف للمفردات عبر لغة توثيقية معيارية ألا و هي قائمة الإسناد RAMEAU . و للتحقق من مدى موائمة هذه القائمة مع قاعدة ألبيريانا تم إجراء عمليتين تقييميتين: الأولى تتعلق بالخصائص التوثيقية لقاعدة ألبيريانا و الثانية تخص قدرة القائمة RAMEAU على استيعاب كافة المفردات المستوردة من الناحية الدلالية. لذلك قمنا بإجراء استبيان بغرض جمع معطيات خاصة بألبيريانا ، كما أجرينا تجربة إعادة كشف لعينة مكونة من 50 بطاقة بيبليوغرافية . نستنتج من خلال هذه العمليات التقييمية أن قائمة RAMEAU تتوائم جزئيا مع قاعدة الجيريانا. بعض من نقائصه تتمثل أساسا في عدم قدرته على تحويل مفردات جد متخصصة أو مفردات متعلقة بالجزائر. من هنا نقترح أن يصحب التشفيف بقائمة RAMEAU جملة من التدابير و القواعد المؤدية إلى تجانس تام . كما يتبين من خلال الدراسة أن السيناريو RAMEAU ليس الحل الوحيد لإشكالية ألبيريانا . بالفعل يمكن التفكير في بدائل أخر وذلك من خلال إعادة تركيب القاعدة على أساس معرفي و الإستعانة بوسائل إصطلاحية. هذه البدائل تتوافق خاصة مع أنظمة الاسترجاع التوثيقي المعتمدة على مبدأ البحث المتعدد المصادر. نتصور في الأخير أن التكنولوجيات الجديدة للمعلومات و خاصة الأنترنت تفتح ورشات بحث عديدة سواء باتجاه عصرنة اللغات التوثيقية التقليدية أو باتجاه التجانس فيما بين أنظمة التصنيف الرئيسة التي يتم إعتماها في وصف المعارف الكامنة في المواد الإلكترونية .

الكلمات الدالة:

لغات توثيق /عدم تواعم اللغات/ قواعد معلومات مختلفة/إدماج المكانز / إعادة التشفيف /قاعدة ألبيريانا/ قائمة RAMEAU.

Sommaire	Pages
Introduction	
1/ Problématique.....	01
2/ Méthodologie	06
Chapitre I : Fondements théoriques du phénomène de la compatibilité des langages documentaires	
I.1 / Définition de la notion de " <i>Compatibilité documentaire</i> ".....	10
I.2/ Evolution des langages documentaires et origine de Leur incompatibilité.....	15
I.2.1/ Evolution sur le plan structurel.....	15
I.2.2/ Evolution sur le plan linguistique.....	20
I.2.3/ Evolution sur le plan thématique.....	21
I.3/ Les différents types d'Incompatibilité	24
I.3.1/ Compatibilité Lexicale.....	24
I.3.2/ Compatibilité thématique.....	29
I.3.3/ compatibilité structurelle.....	30
I.4/ Travaux et expériences internationales en matière de compatibilité Des langages documentaires.....	36
I.4.1/ Travaux Américains.....	43
I.4.2/ Travaux Européens	44
Chapitre II : Solutions préconisées au problème de l'incompatibilité des langages	
II.1/ Les Tables d'Equivalences	48
II.2/ Le Langage Intermédiaire (Switching Language).....	54
II.3/ Le Langage de Référence.....	60
II.4/ Fusion des thesauri.....	63
II.5/ Construction des bases de données de descripteurs	67

II.6/ Les Macrothesauri.....	70
II.7/ Synthèse des solutions.....	73
Chapitre III : Etude de l'incompatibilité des langages dans le cadre de la base de données <i>ALGERIANA</i>	
III.1/ Présentation de la base Algeriana.....	77
III.2/ Principaux facteurs favorisant l'hétérogénéité des langages de la base Algeriana.....	87
III.3/ Travaux antérieurs	110
III.4/ Apport des solutions pratiques pour l'unification des descripteurs d'Algeriana : cas du langage Rameau	124
III.4.1/ Présentation de la liste d'autorité Rameau.....	124
III.4.2/ Evaluation sur l'adaptation du langage Rameau à la base Algeriana.....	127
III. 4.2.1/ Evaluation de l'adaptation Rameau aux caractéristiques documentaires d'Algeriana.....	127
III.4.2.2/ Evaluation sémantique de Rameau par rapport au corpus lexical d'Algeriana.....	132
III.4.2.2.1/ Expérimentation d'une reindexation des descripteurs sources vers des vedettes Rameau.....	133
III.4.2.2.2/ Analyse et interprétation des résultats.....	150
III.5/ Perspectives vers d'autres solutions d'uniformisation des langages d'indexation : cas des Tables d'équivalences	165
Conclusion Générale.....	173
Bibliographie.....	177
Annexes.....	181