

ENSSIBEcole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques**Université
Claude Bernard - Lyon 1****DESS en INFORMATIQUE DOCUMENTAIRE****Rapport de Stage**Recouvrement des catalogues de la Bibliothèque du CERN et de la
Bibliothèque Universitaire Sciences de Grenoble

Nathalie Besson

Sous la direction de
Ingrid Geretschläger

CERN

BIBLIOTHEQUE DE L'ENSSIB



811510F

1997

ENSSIB

Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques

**Université
Claude Bernard - Lyon 1**

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de stage

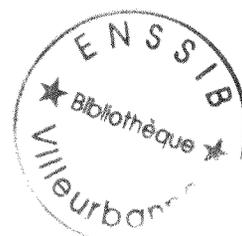
Recouvrement des catalogues de la Bibliothèque du CERN et de la
Bibliothèque Universitaire Sciences de Grenoble

Nathalie Besson

Sous la direction de
Ingrid Geretschläger

CERN

1997



1997
11)ST
03

Recouvrement des catalogues de la Bibliothèque du CERN et de la Bibliothèque Universitaire Sciences de Grenoble

RESUME

Afin d'évaluer le recouvrement d'une partie des catalogues de la Bibliothèque du CERN et de la Bibliothèque Universitaire Sciences de Grenoble, une étude préliminaire des deux catalogues a été réalisée. Une méthode de comparaison a ensuite été élaborée et mise en oeuvre. Elle repose sur un ensemble de programmes permettant d'obtenir les listes des ouvrages communs aux deux bibliothèques et celles des ouvrages présents dans une seule bibliothèque.

DESCRIPTEURS

comparaison, catalogue, bibliothèque, fonds, programmation

ABSTRACT

The aim of my training was to compare a part of the CERN Library catalogue with the Grenoble University Science Library catalogue. To this end, I studied both catalogues and implemented programs that can compare records and create lists of books common to both libraries as well as lists of books that belong to only one library.

KEY-WORDS

comparison, catalogue, library, stock, programming

REMERCIEMENTS

Je souhaiterais tout d'abord remercier les personnes qui m'ont aidée pendant ce stage :

- **Ingrid Geretschläger**, conservateur au Service d'Information Scientifique du CERN et responsable de stage

- **Corrado Pettenati**, Responsable du Service d'Information Scientifique du CERN

- **Jean-Yves Le Meur**, Ingénieur Informaticien au CERN

- **Béatrice Estéoule**, conservateur à la Bibliothèque Universitaire Sciences de Grenoble

- **Christine Musso**, conservateur à la Bibliothèque Universitaire de Grenoble et chargée du système GEAC

- toutes les autres personnes du Service qui ont eu la gentillesse de répondre à mes questions

TABLE DES MATIERES

INTRODUCTION	1
I - CONTEXTE DU PROJET	2
I - 1 - Le CERN	2
I - 2 - Le Service d'Information Scientifique (AS-SI)	2
II - DEMARRAGE DU PROJET	4
III - PLAN DE TRAVAIL	6
IV - ETUDES PREALABLES A LA REALISATION DU PROGRAMME	7
IV - I - Les bibliothèques, leur fonds (monographies)	7
IV - I - 1 - La bibliothèque du CERN	7
IV - I - 2 - La Bibliothèque Universitaire Sciences de Grenoble	13
IV - I - 3 - Bilan	14
IV - II - Etude des deux catalogues	15
IV - II - 1 - Systèmes de gestion de bibliothèque	15
IV - II - 2 - Accessibilités et modes d'interrogation des catalogues	15
IV - II - 3 - Etude des références dans les deux catalogues	17
IV - II - 3 - 1 - Le format	17
IV - II - 3 - 2 - Etude des champs possibles de comparaison	18
IV - II - 3 - 3 - Validité d'une comparaison CDU par CDU?	21
IV - II - 3 - 4 - Autres études réalisées sur les notices	25
IV - II - 3 - 5 - Conclusion de cette étude	26
IV - III - Méthodologie de comparaison des catalogues	27
IV - IV - Dernières études préparant à la programmation	32
V - LA PROGRAMMATION	33
V - I - Résumé des objectifs du programme	33
V - II - Choix du langage de programmation	33
VI - III - Ecriture et débogage des programmes	34
VI - IV - Les programmes	34
VI - IV - 1 - Programme numéro 1	34
VI - IV - 2 - Programme numéro 2	35
VI - IV - 3 - Programme numéro 3	36
VI - IV - 4 - Programme numéro 4	36
VI - V - Test final des programmes et évaluation des erreurs	37
VII - RESULTATS DE LA COMPARAISON	43
VIII - APPLICATIONS POSSIBLES	45
CONCLUSION	46
BIBLIOGRAPHIE	47

INTRODUCTION

Mon stage s'est déroulé au Service d'Information Scientifique du CERN, Laboratoire Européen pour la physique des particules, à Genève.

Ce Service travaille autour d'une bibliothèque spécialisée dans les domaines de la physique théorique, physique nucléaire, atomique et moléculaire.

La mission qu'on m'a demandé de remplir au sein de ce Service est d'évaluer une partie de la politique d'acquisition passée de la bibliothèque grâce à la comparaison de son fonds avec celui de la Bibliothèque Universitaire de Grenoble. Cette dernière étant CADIST¹ en physique, elle est en effet censée avoir une politique d'acquisition très sérieuse dans cette spécialité.

Cette étude doit aussi, éventuellement, constituer un tremplin pour une future collaboration avec la Bibliothèque Universitaire de Grenoble.

Afin de réaliser la comparaison des fonds, j'ai conçu un programme permettant d'une part d'évaluer le pourcentage d'ouvrages communs aux deux bibliothèques par rapport à leur fonds total et permettant d'autre part d'obtenir des listes d'ouvrages : la liste des ouvrages communs aux deux bibliothèques, la liste des ouvrages présents au CERN mais absents du fonds de Grenoble et la liste des ouvrages présents à Grenoble mais absents du fonds du CERN.

¹Centre d'Acquisition et de Diffusion de l'Information Scientifique et Technique . Ce centre a la responsabilité d'acheter et de diffuser la documentation de recherche dans des disciplines spécialisées.

I - CONTEXTE DU PROJET

I - 1 - Le CERN [1]

Le CERN est le plus grand laboratoire européen de recherche en physique des particules. Fondé en 1954, sa Convention constitutive avait été signée à l'origine par 12 pays; il compte aujourd'hui 19 Etats Membres² et 6 Etats Observateurs³.

Environ 6500 scientifiques utilisent les installations du CERN. Ils représentent environ 540 Instituts et Universités et plus de 80 nationalités.

Le CERN a pour raison d'être la recherche fondamentale. Ce laboratoire fournit aux physiciens des accélérateurs de particules (5 aujourd'hui) qui permettent d'obtenir des faisceaux de particules de haute énergie capables de sonder le coeur de la matière.

Le Budget du CERN, en 1996, était de 900 millions de Francs Suisse (environ 3600 millions de Francs Français).

I - 2 - Le Service d'Information Scientifique (AS-SI)

Au sein du CERN, l'AS-SI a pour mission de collecter, classer, diffuser les ouvrages et la littérature grise (pré-tirages, livres, périodiques, comptes rendus, rapports, etc...) et fournir certaines informations utiles aux employés et visiteurs du CERN.

Ce service gère une bibliothèque centrale, 5 bibliothèques satellites et les archives du CERN. La bibliothèque centrale est ouverte 7 jours sur 7 et 24 heures sur 24, et le service aux utilisateurs est fourni du lundi au vendredi de 9 heures à 20 heures. Le personnel est constitué de 17 fonctionnaires européens assistés d'étudiants et de stagiaires. Les services fournis par le Service d'Information Scientifique sont en principe gratuits.

Le Service compte 10 000 utilisateurs potentiels, chercheurs, ingénieurs ou étudiants du 3ème cycle.

² Autriche, Belgique, République Tchèque, Danemark, Finlande, France, Allemagne, Grèce, Hongrie, Italie, Pays-Bas, Norvège, Pologne, Portugal, République Slovaque, Espagne, Suède, Suisse et Royaume Uni

³ Israël, Japon, Russie, Turquie, Commission Européenne et UNESCO

La collection gérée par le Service est antérieure à 1954. Elle est constituée de 38 000 monographies et comptes-rendus, 200 000 pré-tirages et rapports, environ 700 périodiques (600 vivants et 120 électroniques). Ses ouvrages sont majoritairement de la littérature grise (80%), en anglais (80%).

5000 prêts par an sont effectués dans ces bibliothèques et 2500 prêts entre bibliothèques.

Le budget 1997 du Service est de 950 KCHF (environ 4000 KF).

Aujourd'hui les principales activités vers lesquelles s'oriente le Service d'Information Scientifique sont:

- la gestion des bases de données permettant un accès rapide et facile aux ouvrages de la collection du CERN mais aussi d'un réseau international de centres fournisseurs
- l'offre de commande en ligne d'ouvrages
- l'offre d'une aide experte pour la recherche à partir des bases de données HEP⁴ et des bases de données commerciales dans tous les domaines intéressant la R & D
- l'offre d'accès à des documents électroniques à partir de serveurs sur l'Internet

⁴High Energy Physics

II - DEMARRAGE DU PROJET

En 1995 un inventaire est réalisé pour la première fois dans l'ensemble des bibliothèques du CERN (C ⁵, LEP ⁶, PS ⁷, SPS ⁸, CS ⁹). Il s'avère que sur les 35000 ouvrages, achetés entre 1954 et 1995, 13000 ont disparu, soit environ 40% des fonds. Les subventions étant limitées, une commission, le Working Group for Acquisitions, est chargée de choisir les ouvrages qu'il est nécessaire de racheter. Les conclusions de la commission sont les suivantes : sur les 13 000 ouvrages disparus, seuls 500-600 doivent être rachetés.

La faiblesse de ce chiffre peut s'expliquer au moins en partie :

Il n'est pas prévu, par exemple, de racheter les ouvrages trop anciens qui seront plutôt remplacés par des ouvrages récents. Les ouvrages dont il reste des exemplaires ne sont pas repris non plus...

Il n'en reste pas moins qu'une question se pose au sein du Scientific Information Policy Group : la politique d'acquisition de la bibliothèque était-elle valable dans le passé ?

Afin de répondre à cette question, on propose de comparer le catalogue du CERN avec le catalogue d'une bibliothèque dont la valeur des fonds est reconnue : la Bibliothèque Universitaire Sciences de Grenoble, CADIST Physique depuis 1982, est retenue.

Une autre raison pousse à faire cette comparaison : étant donné la diminution du budget attribué au Service d'Information Scientifique, il peut être intéressant de collaborer avec une bibliothèque au niveau de l'acquisition et de la diffusion de l'information.

⁵Central Library

⁶Large Electron Positron Library

⁷Proton Synchrotron Library

⁸Super Proton Synchrotron Library

⁹Computer Science Library

Fin 1996, Ingrid Geretschläger, conservateur au Service d'Information Scientifique du CERN, prend contact avec Béatrice Estéoule, conservateur à la Bibliothèque de Grenoble.

Une visite de la Bibliothèque de Grenoble permet de s'assurer de la faisabilité de la comparaison. On note en effet que :

- les deux catalogues sont gérés par des Systèmes de gestion de bibliothèque avancés (GEAC à Grenoble, ALEPH au CERN)
- ils sont accessibles tous les deux par Telnet
- la classification des monographies se fait essentiellement par CDU ¹⁰, influence OCLC ¹¹ pour la Bibliothèque Universitaire et influence OCLC et ETHICS pour le CERN.

A partir de ses observations, les premières idées sont les suivantes :

La comparaison pourrait se faire sur les fonds de monographies.

Il serait sans doute possible de sélectionner les ouvrages intéressants des deux bibliothèques en ne gardant que ceux caractérisés par des CDU précis.

On pourrait aussi diviser le fonds choisi en plusieurs sous-fonds correspondant à des grands domaines scientifiques; ce sont ces sous fonds que l'on comparerait.

Afin de tester le futur programme, Mme Geretschläger propose de prendre tout d'abord des échantillons représentatifs :

- CDU=530.14? (physique théorique) et année de publication=1970

- CDU=539.2? (physique de l'état solide) et année de publication=1980

Ma mission a été d'élaborer une méthode de comparaison de ces catalogues et de la mettre en oeuvre.

¹⁰ Classification Décimale Universelle

¹¹ On Line Computer Library Center

III - PLAN DE TRAVAIL

A partir des besoins de mes commanditaires, j'ai tout d'abord conçu un premier plan de travail, très général, définissant les étapes par lesquelles mon étude devrait passer.

Les étapes que j'ai retenues sont les suivantes :

- 1 - Etudier les deux catalogues. Noter les points communs et les différences.
- 2 - Déterminer les champs qui permettent de faire la comparaison des notices des deux catalogues
- 3 - Choisir un "échantillon" ou plusieurs petits "échantillons" de catalogue intéressants à partir desquels on pourra ensuite tester le programme de comparaison et avoir une première vue du recouvrement
- 4 - Faire un premier bilan des difficultés rencontrées
- 5 - Elaborer une technique de comparaison (support des références, matériel utilisé, langage choisi...)
- 6 - Déterminer les informations que l'on veut obtenir après comparaison
- 7 - Elaborer le programme de comparaison
 - Lire les programmes existant proches de celui à réaliser; se renseigner sur certaines commandes
 - Ecrire ce programme sachant qu'il devra
 - sélectionner les champs voulus
 - harmoniser la présentation des champs dans les deux catalogues
 - comparer les références des deux catalogues
 - présenter les informations voulues
 - Tester le programme et le corriger si besoin est
- 8 - Faire tourner le programme sur les échantillons décidés, analyser ces premiers résultats, les présenter
- 9 - Etendre si possible cette étude à une grande partie du catalogue
- 10 - Analyser ces résultats et les présenter en soulignant l'intérêt ou non d'une collaboration avec la BU Sciences Grenoble

IV - ETUDES PREALABLES A LA REALISATION DU PROGRAMME

IV - I - Les bibliothèques, leur fonds (monographies)

IV - I - 1 - La bibliothèque du CERN

La bibliothèque du CERN contient 38000 monographies et comptes rendus.

Elle est essentiellement spécialisée dans les domaines de la Physique théorique, physique nucléaire, atomique et moléculaire, du fait de son appartenance au CERN, mais elle a aussi un fonds conséquent dans les domaines des mathématiques, de l'informatique, des autres physiques et notamment de la physique macromoléculaire, des matériaux et de la mécanique, comme le montrent les tableaux de la page suivante¹².

¹²Notons que dans ce tableau ainsi que dans les tableaux suivants ne sont comptées que les monographies dont l'indice CDU a été noté dans la base de donnée du CERN. Les comptes-rendus de conférences n'ayant plus d'indice CDU depuis 1984, ils ne sont pas tous comptés dans ces tableaux.

INDICE CDU [2]	SUJET	FONDS (nombre d'ouvrages)
0? ¹³	Généralités	1063
1?	Philosophie	169
2?	Religion / Théologie	146
3?	Sociologie générale / Sociographie	241
4?	Linguistique / Philologie	465
5?	Mathématiques / Sciences Naturelles	20 487
6?	Sciences Appliquées / Médecine / Technique	10 336
7?	Art / Architecture / Photographie / Musique / Divertissements / Jeux / Sports	246
8?	Littérature / Belles-Lettres	131
9?	Monographies régionales / Géographie / Biographie / Histoire	246

Tableau 1: Répartition du fonds des monographies du CERN par grand domaine de la Classification Décimale Universelle (juin 97)

^{13?} représente une troncature de l'indice CDU

INDICE CDU	SUJET	FONDS (nombre d'ouvrages)
51? ¹⁴	Mathématiques	9037
52?	Astronomie / Géodésie	918
53?	Physique	9764
530?	Physique - Généralités	2794
531?	Mécanique rationnelle / Mécaniques des solides, corps rigides	854
532?	Mécanique des fluides- hydraulique	192
533?	Théorie des gaz- aérodynamique	228
534?	Vibrations / Acoustique	113
535?	Optique / Lumière	403
536?	Chaleur / Thermodynamique	352
537?	Electricité	1126

Tableau 2: Répartition du fonds de monographies du CERN dans les différents domaines des Mathématiques et Sciences Naturelles (juin 97)

¹⁴? représente une troncature de l'indice CDU

INDICE CDU	SUJET	FONDS (nombre d'ouvrages)
538? ¹⁵	Magnétisme / Electromagnétisme	312
539?	Nature Physique de la matière	4697
54?	Chimie	535
55?	Géologie et disciplines connexes / Météorologie	70
56?	Paléontologie / Fossiles	1
57?	Sciences biologiques	160
58?	Botanique	2
59	Zoologie	0

Tableau 2 (suite): Répartition du fonds de monographies du CERN dans les différents domaines des Mathématiques et Sciences Naturelles (juin 97)

¹⁵? représente une troncature de l'indice CDU

INDICE CDU	SUJET	FONDS (nombre d'ouvrages)
60? ¹⁶	Généralités sur les Sciences Appliquées	45
61?	Sciences Médicales	349
62?	Art de l'Ingénieur / Technique et Industrie en général	5880
620?	Essais matériaux / Connaissance des produits fabriqués et marchandises / Installations énergétiques	514
621?	Mécanique générale et électrotechnique	5263
622?	Industrie minière	0
623?	Art de l'ingénieur militaire et naval	22
624?	Génie civil	37
625?	Voies ferrées / Routes / Voiries	0
626?	Construction hydraulique/ Canaux / Irrigation et drainage	1
627?	Cours d'eau intérieurs naturels / travaux fluviaux / Ports	2

Tableau 3: Répartition du fonds de monographies du CERN dans les différents domaines de la Technique et des Sciences Appliquées (juin 97)

^{16?} représente une troncature de l'indice CDU

INDICE CDU	SUJET	FONDS (nombre d'ouvrages)
628? ¹⁷	Technologie sanitaire	31
629?	Techniques des moyens de transport (sauf rails)	53
63?	Agriculture / Sylviculture / Zootechnie / Chasse / Pêche	7
64?	Science ménagère	1
65?	Organisation / Administration / Technique du commerce et de l'industrie	486
66?	Industries Chimiques et apparentées	308
67?	Industries et activités diverses / Technologie mécanique	164
68?	Industries et professions diverses (suite)	3039
69?	Construction	57

Tableau 3 (suite): Répartition du fonds de monographies du CERN dans les différents domaines de la Technique et des Sciences Appliquées (juin 97)

¹⁷? représente une troncature de l'indice CDU

IV - I - 2 - La Bibliothèque Universitaire Sciences de Grenoble

La bibliothèque Universitaire Sciences de Grenoble a un fonds de 70 000 titres. Elle couvre tous les domaines scientifiques des mathématiques aux sciences de l'ingénieur mais aussi certains sujets annexes (sport par exemple). Comme dans toute Bibliothèque Universitaire, les responsables sont en effet tenus d'acheter ce qui correspond aux enseignements de l' Université¹⁸.

En ce qui concerne les acquisitions de niveau recherche, le fonds est très orienté "physique et physique appliquée", du fait de la spécificité de la Bibliothèque (CADIST).

Les domaines couverts par le CADIST sont nombreux (acoustique, aérodynamique, cryophysique, composants électroniques, cristallographie, diélectriques, électronique, électromagnétisme, énergétique, génie des matériaux, génie électrique, génie mécanique, génie physique, physique de la matière, physique quantique, physique nucléaire, physique atomique et moléculaire,...). Le CADIST représente ainsi 10 000 ouvrages de recherche à 90% étrangers avec un accroissement annuel de 800 ouvrages.

Les ouvrages qui traitent de domaines comme les mathématiques ou l'informatique se trouvent essentiellement dans les bibliothèques d'Instituts.

Les premières acquisitions de la Bibliothèque Universitaire de Grenoble datent de 1967 mais elle possède aussi les collections d'Universités et d'Ecoles d'Ingénieurs avec des ouvrages datant du milieu du XIXème siècle.

¹⁸du 1er cycle aux DEA, DESS et Thèses

IV - I - 3 - Bilan

Dans le cadre de mon projet c'est-à-dire d'une comparaison avec une bibliothèque CADIST Physique, nous ne nous intéresserons pas au domaine des mathématiques dans lequel la Bibliothèque Universitaire de Grenoble n'est pas spécialisée. Nous nous contenterons donc d'effectuer une comparaison des deux fonds dans certains domaines de la Physique ou de la Technique.

Après une étude des fonds "Physique" et "Technique" du CERN les indices CDU intéressants pour la comparaison ont été choisis ¹⁹:

INDICE CDU	SUJET	FONDS DU CERN	FONDS DE LA BIBLIOTHEQUE UNIVERSITAIRE DE GRENOBLE
530.1?	Physique Théorique	2631	697
539.1?	Physique nucléaire, atomique et moléculaire	4269	802
539.2?	Physique de l'état solide	321	229
620?	Caractérisation des matériaux	514	835
621?	Mécanique générale et électrotechnique	5259	2215
678?	Industrie des matières macromoléculaires-caoutchouc-matières plastiques)	160	64

¹⁹Ces chiffres correspondent au fonds de monographies et comptes-rendus de conférence

IV - II - Etude des deux catalogues

La comparaison des fonds des deux bibliothèques nécessite l'interrogation de leur deux catalogues.

Une étude approfondie de ces catalogues (leur contenu, leur accessibilité, les champs interrogeables, les formats des références,...) a donc été indispensable.

IV - II - 1 - Systèmes de gestion de bibliothèque

La bibliothèque du CERN travaille avec le système en ligne de gestion de base de données et de bibliothèque ALEPH développé par ALEPH-VISSUM (Jérusalem).

Ce système, qui fonctionne sur UNIX, intègre à la fois un OPAC (Online Public Access Catalogs), et un langage de commande (Common Command Language, CCL) pour la recherche des documents.

ALEPH gère tout un ensemble de fichiers tels que des fichiers d'enregistrements contenant par exemple les différentes caractéristiques d'un ouvrage (auteur, titre, date de publication, numéro de système...), des fichiers d'autorité, des fichiers d'index, des fichiers de mots, ainsi que des fichiers spécifiques pour la gestion des acquisitions, le rangement et la circulation des ouvrages notamment.

La Bibliothèque Universitaire de Grenoble travaille depuis 1993 avec le système intégré de gestion des bibliothèques GEAC-ADVANCE fondé sur un système d'exploitation UNIX et l'usage d'un SGBD associé (UNIVERSE dans le cas de cette Bibliothèque).

Ce système permet l'utilisation de trois types de notices : des notices bibliographiques, des notices d'autorité et des notices d'exemplaire ²⁰.

²⁰servent pour le prêt et la localisation des ouvrages

IV - II - 2 - Accessibilités et modes d'interrogation des catalogues

Comme déjà précisé, le catalogue de la Bibliothèque Universitaire de Grenoble est accessible par session Telnet. Il constitue une sous-partie du catalogue informatisé RUGBIS ²¹.

Le Réseaux RUGBIS est un ensemble documentaire composé de 8 bibliothèques spécialisées d'enseignement et de recherche, implantées dans les différents bâtiments de l'Université Joseph Fourier (Université de Grenoble) et de l'Institut National Polytechnique de Grenoble.

Au sein de cet ensemble documentaire, les bibliothèques gèrent tout ou une partie de leurs documents en choisissant d'utiliser les modules du logiciel qui les intéressent (acquisition, prêt, recherche...), par contre l'utilisateur ne consulte qu'un seul et même catalogue pour toutes ces bibliothèques.

Une interrogation sous session Telnet de ce catalogue peut être lancée à partir de l'Internet à l'adresse :

<http://melpomene.upmf-grenoble.fr/POLE/REDOC/aide/rugbis.html>

Le catalogue permet de faire des recherches dans l'ensemble du réseau RUGBIS ou de sélectionner un établissement.

En ce qui concerne la recherche de documents, il est possible de faire des recherches par titre, auteur, collection, sujet, cote, ISBN/ISSN ainsi que par mots du titre, de l'auteur, de la collection, du sujet, de tous les index, des notes. Certaines recherches combinées sont aussi possibles grâce à des opérateurs booléens.

Les troncatures sont aussi possibles.

Le catalogue de la bibliothèque du CERN est accessible par session Telnet depuis 1983.

Il est aussi interrogeable directement sur le Web depuis 1996 à l'adresse :

<http://alice.cern.ch/>

Des recherches très complexes sont possibles dans ce catalogue. Il permet notamment de faire des recherches dans différentes sous-bases de données (bases des périodiques, monographies, pré-tirages, etc.). L'essentiel des champs est interrogeable (champs date, auteur, titre, édition, indice CDU, mot-clé,...) et ces recherches peuvent être combinées.

Plusieurs types de troncatures sont utilisables.

²¹Réseau Universitaire Grenoblois des Bibliothèques d'Information Scientifique

IV - II - 3 - Etude des références dans les deux catalogues (Format, présentation,...)

L'étude du format des références dans les deux catalogues a pour but:

- de déterminer les champs sur lesquels la comparaison des références est possible et de connaître les différences de présentation de ces champs d'un catalogue à l'autre ainsi qu'au sein d'un même catalogue
- de choisir les champs qui seront intéressants à présenter si on réalise des listes d'ouvrages²² et de connaître les différentes présentations de ces champs
- de vérifier si une comparaison des ouvrages CDU par CDU est possible

IV - II - 3 - 1 - Le format

En 1986 la Bibliothèque Universitaire de Grenoble a d'abord intégré le réseau SIBIL²³ qui se présente aujourd'hui comme un logiciel assurant les fonctions de base d'un serveur bibliographique, destiné à supporter des réseaux de bibliothèques. Ce réseau travaille avec le format de catalogage InterMarc.

Vers 1992 la Bibliothèque est passée du réseau SIBIL au réseau OCLC au sein duquel les bibliothèques peuvent importer dans leur catalogue des notices en format LCMARC.

La Bibliothèque Universitaire est donc passé du format InterMarc au format LC MARC.

Outre ce format de catalogage, plusieurs autres formats plus adaptés aux utilisateurs sont disponibles. Ainsi le catalogue RUGBIS propose trois types de formats :

- le format abrégé
- le format complet
- le format L.C. MARC

En interne, il est aussi possible d'obtenir divers types de références constitués des champs que l'on souhaite faire apparaître.

Le système GEAC permet aussi d'exporter des références en formats ISO 27 09.

²²listes des ouvrages communs aux deux bibliothèques, listes d'ouvrages présents seulement dans une bibliothèque...

²³Système Informatisé pour Bibliothèques

Globalement notons que, même s'il existe différents formats, le contenu des champs a toujours la même présentation (le contenu d'un champ n'est entré qu'une seule fois par le catalogueur, en format LC MARC).

Pensant tout d'abord obtenir toutes les informations nécessaires à mon projet à partir de RUGBIS, c'est le format complet de RUGBIS que j'ai étudié. Les exemples donnés par la suite correspondent donc à ce format.

En ce qui concerne le catalogage des monographies, la bibliothèque du CERN n'importe pas ses notices. Celles-ci sont remplies par des personnes du service. Jusqu'à 99 formats sont disponibles avec différentes combinaisons de champs. Les exemples donnés par la suite correspondent au format 1 du catalogue²⁴.

IV - II - 3 - 2 - Etude des champs possibles de comparaison

IV - II - 3 - 2 - 1 - Choix des champs de comparaison

Afin de déterminer si une monographie est présente ou non dans les deux catalogues, il est nécessaire de choisir des champs qui permettent de caractériser de façon précise cette monographie, c'est-à-dire qu'une combinaison de valeurs de ces champs doit indiquer une monographie et une seule.

Ces champs doivent, en outre, être obligatoirement remplis dans les deux catalogues et pour toutes les monographies.

Afin de rechercher ces champs de comparaison, j'ai dépouillé de nombreuses notices du catalogue du CERN et du catalogue de Grenoble. J'ai choisi d'étudier les notices dans un format le plus complet possible²⁵.

Ce dépouillement a montré que les champs obligatoirement remplis sont très rares.

²⁴format qui affiche tous les champs

²⁵format COMPLET pour Grenoble, format numéro 1 pour le CERN

Il s'est notamment avéré que l'ISBN des monographies, numéro caractéristique d'une monographie donc à priori très intéressant pour ce genre de comparaison, n'est pas indiqué pour tous les ouvrages. Dans le cas d'ouvrages non européens, surtout, cette information n'est pas indiquée. Le champ ISBN n'a donc pas pu être choisi comme champ de comparaison.

De la même façon, la date n'apparaît pas dans tous les enregistrements. Elle varie aussi avec les différentes éditions d'une même oeuvre.

Seuls le titre et l'auteur (principal ou secondaire) de l'ouvrage sont presque systématiquement indiqués dans les deux catalogues lorsqu'il s'agit de monographies. C'est pour cette raison que j'ai choisi ces deux champs comme champs de comparaison.

IV - II - 3 - 2 - 2 - Etude de la présentation des champs de comparaison

Ce choix fait, j'ai ensuite étudié les différentes présentations possibles de ces deux champs au sein des deux catalogues. Cette étude doit permettre d'une part de reconnaître la localisation des deux champs dans une référence et d'autre part de connaître toutes les différences que le programme devra éliminer avant de faire une comparaison.

Voici l'essentiel des informations que j'ai tirées de cette étude ainsi que des explications de Mme Geretschläger:

Les titres :

Dans le catalogue du CERN, j'ai pu noter que les titres sont indiqués dans les champs intitulés TI ou TI1 ou TI2.

Le champ TI correspond à un titre principal.

Le champ TI2 correspond à un titre secondaire comme un titre parallèle, un titre abrégé ou un titre traduit.

Le champ TI1 est utilisé lorsque le titre est choisi comme entrée principale. Les ouvrages ayant un champ TI1 seront donc classés, sur les étagères sous leur titre.

Dans le catalogue de la B.U. Sciences de Grenoble, les titres sont indiqués dans le champ TITRE.

Des exemples de présentations des titres sont indiqués dans le tableau suivant (ici le titre de l'ouvrage est: "Les sciences physiques") :

CATALOGUE DE LA B.U. SCIENCES DE GRENOBLE	
TITRE	: Les sciences physiques : Le titre secondaire
TITRE	: Les sciences physiques /edited by ...
TITRE	: Les sciences physiques / notes taken by...
TITRE	: Les sciences physiques;...
TITRE	: Les sciences physiques.
TITRE	: Les sciences physiques=
CATALOGUE DU CERN	
TI :	Les sciences physiques\$\$\$ sous-titre
TI :	Les sciences physiques
TI :	<<Les>> sciences physiques\$\$...

On remarque donc généralement d'assez grandes variations dans la présentation du titre au sein du catalogue de La Bibliothèque Universitaire de Grenoble. Dans le catalogue du CERN, l'unité est meilleure. La difficulté essentielle réside dans la présence des signes "<< " et ">>" autour des mots vides présents en début de certains titres. Ces signes, obligatoires au CERN dans une version antérieure du système intégré ne sont plus ajoutés actuellement.

Les auteurs :

Dans le catalogue du CERN, les auteurs apparaissent dans les champs intitulés AU, AU1, AU2 ou AU3. En ce qui concerne les monographies, seuls les champs AU1 et AU3 sont employés.

AU1 correspond à la vedette auteur (l'auteur principal) et AU3 correspond aux auteurs secondaires.

Dans le catalogue de la B.U. de Grenoble, les auteurs sont indiqués dans le champ AUTEUR quand il s'agit d'auteurs principaux ou dans le champ TITRE après le titre quand il s'agit d'auteurs secondaires.

Des exemples de présentations des auteurs sont indiqués dans le tableau suivant (ici l'auteur est soit John Antony Desanto soit Louis de Broglie; le second auteur, quand il y a en a un, est John Edward George Farina) :

CATALOGUE DE LA B.U. SCIENCES DE GRENOBLE	
Dans le champ auteur:	
AUTEUR	: Desanto, John Antony,
AUTEUR	: Desanto, John Antony.
AUTEUR	: Desanto, John A.,
AUTEUR	: Desanto, J. A.
AUTEUR	: Broglie, Louis de
Dans le champ titre :	
TITRE	: Les sciences physiques /cedited by J. A. D esanto
TITRE	: Les sciences physiques/ J. A. Desanto
TITRE	: Les sciences physiques /by J. A. Desanto
TITRE	: Les sciences physiques /John A. Desanto.
TITRE	: Les sciences physiques /J. A. Desanto, J. E. G. Farina
TITRE	: Les sciences physiques / J. A. Desanto,...J. E. G. Farina...
CATALOGUE DU CERN	
AU	: Desanto, J A
AU	: Desanto, John Antony (A chaque fois que les prénoms sont connus)
AU	: De Broglie, L

En ce qui concerne la présentation des auteurs on remarque là encore que le catalogage pratiqué à la B.U. Sciences de Grenoble est moins rigoureux qu'au CERN où une unique écriture de l'auteur est tolérée.

Par contre le CERN ne respecte pas les règles françaises de catalogage pour les noms à particules alors que la Bibliothèque de Grenoble les respecte.

Autres points :

Outre ces différences de catalogage j'ai aussi noté des différences dans l'affichage à l'écran des accents.

Par exemple le "è" apparaît sur mon écran comme " e" dans le catalogue RUGBIS et " " dans le catalogue de la bibliothèque du CERN.

IV - II - 3 - 3 - Validité d'une comparaison CDU par CDU?

IV - II - 3 - 3 - 1 - Etude des indices CDU

Afin de vérifier si une comparaison des fonds, sujet par sujet, est réalisable, j'ai été amenée à étudier les champs dans lesquels sont indiqués les indices CDU, au sein des deux catalogues. Dans ce contexte les questions que j'ai dû me poser sont les suivantes : Quel est l'intitulé des champs dans lesquels se trouve l'information sur l'indice CDU? Ces champs sont-ils toujours remplis? Ces champs sont-ils facilement interrogeables? Est-il possible de faire une sélection de tous les ouvrages qui ont un même CDU?

Dans le catalogue du CERN, l'indice CDU est indiqué dans un champ UD. Il est facilement interrogeable en ligne: pour obtenir la liste des ouvrages ayant un indice commençant par 530.1, par exemple, il suffit de taper la commande " f 530.1?/ud ".

Le problème principal rencontré dans le catalogue du CERN est l'absence d'indice pour certains comptes-rendus de conférences. En effet, au CERN, depuis 1984, le CDU a été abandonné au profit de mots libres pour le catalogue des comptes rendus de conférence.

Ces ouvrages présentant moins d'intérêt que les autres dans le cadre de cette comparaison, nous avons alors pris la décision de réduire notre étude en éliminant les compte-rendus.

En ce qui concerne le catalogue de Grenoble, les problèmes rencontrés sur les indices CDU se sont avérés beaucoup plus complexes. Effectivement, l'interrogation du catalogue et l'étude de ses références m'ont permis de remarquer l'absence régulière de ce type d'information.

Les bibliothécaires de Grenoble m'ont confirmé cette observation:

A la Bibliothèque Universitaire, seules les monographies en accès libre (soit 60% du fond total) ont obligatoirement une cote CDU interrogeable à partir du champ "cote" dans le catalogue.

Il est bien évident que ne pouvoir faire un travail de comparaison que sur 60% des fonds de la Bibliothèque de Grenoble n'était pas envisageable. Cette nouvelle a donc momentanément remis en cause l'idée de sélectionner les ouvrages d'un sujet donné grâce à l'indice CDU.

J'ai cependant essayé d'obtenir plus d'informations sur le classement des autres ouvrages c'est-à-dire les ouvrages localisés dans le magasin de la Bibliothèque:

Dans le magasin les monographies sont caractérisées par diverses cotes :

- Les monographies les plus anciennes étaient classées dans le magasin par numéro d'inventaire, ce numéro est constitué d'une cote de domaine (Ca pour la physique, F pour la technique, Cb pour la chimie...) et d'un numéro d'arrivée.

- Certaines de ces monographies ayant été auparavant en accès libre elles avaient aussi une cote CDU.

- Depuis un an un nouveau système de classement a été adopté dans le magasin et les nouvelles monographies sont, comme celles en accès libre, classées par CDU.

Lorsque la Bibliothèque a informatisé sa gestion du catalogue ²⁶ les notices des ouvrages (depuis 1970) ont été introduites dans le catalogue.

Les cotes CDU des ouvrages en accès libre ont été placées dans un champ "cote".

Les numéros d'inventaire des ouvrages en magasin (avant la nouvelle classification) ont été placés dans un champ libre. Pour certains ouvrages du magasin, la cote CDU et le numéro d'inventaire ont été notés dans cette zone de champ libre; pour d'autres, seul le numéro d'inventaire a été conservé.

Depuis l'extérieur²⁷ il n'est pas possible de faire une interrogation sur cette zone de champ libre, par contre en interne les bibliothécaires peuvent réaliser ce genre d'interrogations.

Une interrogation simultanée du champ "cote" et du champ libre (correspondant au numéro d'inventaire) est alors apparue comme la solution à notre problème. Elle doit en effet permettre de couvrir une grande partie des fonds de monographies (plus de 80%) pour un domaine donné. C'est donc finalement cette méthode qu'il a été convenu d'utiliser.

²⁶systeme GEAC

²⁷OPAC : open access catalogue (réservé au public)

IV - II - 3 - 3 - 2 - Autres méthodes de comparaison envisagées

La sélection des ouvrages scientifiques au moyen des indices CDU s'étant au départ avérée délicate, d'autres solutions avaient été envisagées par Mme Geretschläger et moi-même. En voici un aperçu :

Les mots clés RAMEAU [3]

Une autre façon de cibler un sujet scientifique donné aurait pu être de déterminer les mots clés RAMEAU²⁸ qui caractérisent ce sujet.

Hélas, cette classification n'ayant été adoptée que depuis 1986²⁹, cette solution n'est, là encore, pas valable pour l'ensemble des fonds de la Bibliothèque.

D'autre part, le CERN n'utilise pas les mots RAMEAU mais des mots clés libres et par conséquent très nombreux. Par ailleurs, lorsqu'un mot significatif apparaît dans le titre (champ TI), celui-ci n'est pas répété en mot-clé (champ KW).

Cette méthode est donc à rejeter.

Les ouvrages CADIST

La deuxième solution, proposée par Mme Geretschläger a été la sélection des ouvrages de physique de la Bibliothèque Universitaire au moyen de leur notation "CADIST". Dans l'hypothèse où cette sélection est possible, on peut en effet envisager de comparer l'ensemble de ces ouvrages à l'ensemble du fonds "physique" du CERN. J'ai donc cherché à savoir si des informations sur le budget d'acquisition spécial CADIST se reflétaient dans les notices.

²⁸Répertoire d'Autorité Matière Encyclopédique et Alphabétique Unifié

Ces mots clés sont nés d'une recherche de cohérence dans le vocabulaire d'indexation, de façon à limiter les dispersions des notices sous des rubriques trop nombreuses. Depuis 1980, les services d'indexation-matière de la Bibliothèque Nationale Française constituent donc, au fur et à mesure des besoins du catalogage courant, une liste d'autorité encyclopédique de matière, adaptée du Répertoire de vedettes-matière de l'Université Laval à Québec, lui-même dérivé de la liste de vedettes-matière de la Bibliothèque du Congrès de Washington : Library of Congress Subject Headings (LCSH).

La liste d'autorité de matière de la Bibliothèque Nationale a été informatisée en 1980, elle a été chargée dans la base BN-OPALE en 1985. La Bibliothèque Universitaire de Grenoble l'utilise depuis 1986.

²⁹la Bibliothèque Universitaire utilisait auparavant des mots libres

En effet, cette information est bien indiquée dans une zone de champ libre non interrogeable en externe par RUGBIS mais interrogeable en interne.

Par contre, le programme des CADIST n'existe que depuis 1982; les ouvrages antérieurs ne peuvent donc pas être sélectionnés de cette manière.

Par ailleurs, quand en 1993 la bibliothèque a adopté le système de gestion intégré ADVANCE GEAC les notices datant de SIBIL ont vu leur notation "CADIST" effacée.

Pour finir, en ne sélectionnant que les ouvrages CADIST, on ne couvre pas tous les ouvrages de Physique³⁰.

Cette solution a donc été rejetée.

IV - II - 3 - 4 - Autres études réalisées sur les notices

Le but du projet étant d'obtenir des listes d'ouvrages, il est nécessaire de choisir les informations que l'on va présenter dans ces listes. Afin de faire ce choix il faut se poser les questions suivantes : dans quel but réalise-t-on ces listes? quelle personnes les liront?

Ces listes doivent présenter, d'une part les ouvrages communs aux deux bibliothèques et d'autre part les ouvrages qui ne sont présents que dans une bibliothèque. Ce deuxième type d'information est particulièrement intéressant pour chacune des bibliothèques. Les listes contenant cette information vont en effet permettre de savoir si des ouvrages intéressants ne manquent pas à une bibliothèque; elles permettront peut-être aussi de choisir des ouvrages à acquérir.

Dans cette perspective, j'ai choisi de présenter sur ces listes les champs importants pour faire une demande d'acquisition et permettant aussi de juger de la qualité de l'ouvrage soit : le titre, l'auteur, l'édition (lieu, maison d'édition, date) et la collection.

³⁰En effet la BU bénéficie de trois budgets : les subventions du Ministère (pour les étudiants et la recherche), le budget BNF (pour des ouvrages de physique théorique et appliquée dans les domaines de l'optique, l'électronique et les supraconducteurs) et le budget CADIST (pour des ouvrages de physique et de physique appliquée). Des ouvrages de Physique sont donc achetés avec d'autres budgets.

IV - II - 3 - 5 - Conclusion de cette étude

L'étude des notices s'est avérée très complexe du fait des changements de politique de catalogage ayant eu lieu au sein des deux bibliothèques. Ces évolutions se retrouvent dans la plupart des bibliothèques et sont souvent étroitement liées aux évolutions des moyens de gestion des bibliothèques.

Parfois aussi, certaines politiques de catalogages, jugées inutilement rigoureuses et engendrant ainsi des pertes de temps, ont été assouplies.

Ces règles abandonnées ainsi que certaines normes jamais adoptées n'ont pas de répercussions négatives sur la gestion courante des bibliothèques. Elles peuvent, par contre, devenir des obstacles pour certaines applications exceptionnelles comme l'étude que j'ai réalisée.

Si les bibliothèques souhaitent s'ouvrir vers l'extérieur, importer et exporter facilement des informations, il leur sera peut-être nécessaire d'adopter dans l'avenir une politique de catalogage plus rigoureuse et plus unifiée.

IV - III - Méthodologie de comparaison des catalogues

La méthode de comparaison qui a été appliquée est née après avoir envisagé plusieurs autres méthodes. Ces méthodes ont été abandonnées les unes après les autres du fait peut-être de leur caractère peu réaliste pour certaines (on ne maîtrise pas toutes les données du fonctionnement d'un catalogue aisément !) mais aussi du fait des limites des catalogues par exemple.

Je me contenterai ici de citer les principales méthodes que j'ai étudiées et parfois mis en oeuvre.

Afin de comparer une partie du fonds du CERN avec une partie du fonds de la Bibliothèque Universitaire de Grenoble, ma première idée était de travailler sur des fichiers textes contenant les listes des ouvrages de la bibliothèque du CERN et de la Bibliothèque Universitaire de Grenoble. Le contenu de ces fichiers aurait pu être comparé à l'aide d'un programme.

Il aurait par exemple été intéressant de travailler avec plusieurs listes, une liste contenant tous les ouvrages d'une bibliothèque ayant un indice CDU donné et une date de publication donnée.

Obtention des notices provenant de la Bibliothèque Universitaire de Grenoble

Pour obtenir les listes des ouvrages de Grenoble je pensais d'abord interroger le catalogue de Grenoble en session Telnet à partir du catalogue RUGBIS puis sélectionner les références ayant un CDU et une année fixée et les télécharger sur un fichier texte.

L'interrogation de ce catalogue a malheureusement montré qu'il est impossible de faire une sélection par cote (CDU) et par année simultanément. Une simple sélection par cote n'est pas non plus possible, le catalogue ne propose en effet qu'un index des cotes.

D'autre part, un téléchargement des références n'est pas réalisable de façon simple puisqu'il faut télécharger une à une ces références auxquelles s'ajoutent de nombreuses informations inutiles.

J'ai alors cherché à obtenir ces listes non plus de l'extérieur par le catalogue RUGBIS mais de l'intérieur par le système GEAC. Ne pouvant pas faire moi-même cette sélection, j'ai demandé à ce que ces listes me soient envoyées de la Bibliothèque Universitaire de Grenoble par courrier électronique.

Christine Musso, responsable du système de gestion intégré des bibliothèques GEAC, m'a confirmé l'extractibilité de ces listes. Par contre, après demande auprès de ses interlocuteurs GEAC, elle m'a annoncé l'impossibilité d'exporter ces listes, disponibles alors uniquement sur support papier.

A défaut de mieux j'ai donc essayé de numériser l'image de ces listes grâce à un scanner puis d'utiliser un logiciel de reconnaissance optique des caractères (Text Bridge Professionnal Edition 3.0) pour pouvoir disposer de listes traitables par un programme.

Les listes obtenues à la suite de ces opérations se sont avérées présenter très peu d'erreurs mais, pour faire un travail rigoureux, une correction de ces quelques erreurs était nécessaire et imposait alors une relecture de toutes les références après traitement OCR ce qui aurait pu être très long pour de grandes listes.

Cette solution n'était donc pas très efficace.

C'est pourtant celle qui a en premier lieu été mise en oeuvre faute de mieux.

Les premières listes reçues se présentaient de la façon suivante :

Tit/ Aut.....	Dislocations et deformations Groh, Pierre
Cote.....	539.219 ECO
Editeur et date...	Diff. les Ed. de physique, 1980
Tit/ Aut.....	Aerosol microphysics Marlow, W.
Cote.....	539.24 AER
Editeur et date...	Springer, 1980

Un premier programme a été réalisé pour traiter ces listes tandis qu'en parallèle nous cherchions toujours comment obtenir des listes sur support électronique.

Plus tard un module du système GEAC permettant l'exportation d'enregistrements en format ISO 2709 ³¹ a été activé. Nous avons donc reçu les listes souhaitées par courrier électronique.

³¹format pour l'échange d'information - voir annexe

Par contre, pour des raisons de temps, seules 6 listes ont été demandées chacune correspondant à un indice CDU (le découpage année par année n'a pas été réalisé).

C'est donc avec ces 6 listes que j'ai finalement travaillé.

Voici, pour exemple le début d'une de ces listes³² :

01396namKaL200325 45000300060000000800410000603500170004704000180006402000150
00820360013000970410016001100920015001260990016001410490009001571000021001662450
32100187260003100508300003000539440005500569500007100624650002100695650002700716
74000460074374000360078974000580082574000340088374000500091774000460096785200570
10130CoLC960105r19941925fr 000 0 fre d aocm31618187 aFASbfreqFAS
a2876471558 a101311081 afrehgereng a530.1fEIN a530.12fEIN aFPHS1 aEinstein,
Albert10aSur l',electrodynamique des corps en mouvement; L',ether et la th,eorie
de la relativit,e; La g,eom,etrie et l'exp,erience; Quatre conf,erences sur la
th,eorie de la relativit,e; Sur le probl-eme cosmologique; Th,eorie relativiste
du champ non sym,etrique; Th,eorie de la gravitation g,en,eralis,ee/cAlbert Eins
tein aSceaux :bJ. Gabay,cc1994 aPag. multiple;c18x25 cm. 4aLes grands classiqu
es Gauthier-Villars,x0989-0602 aFac-sim. des ,ed. de Paris : Gauthier-Villars,
1925,1953,1960,1951 7aElectrodynamique 7aRelativit,e (physique)21aL',ether et la
th,eorie de la relativit,e31aLa g,eom,etrie et l'exp,erience01aQuatre conf,eren
ces sur la th,eorie de la relativit,e01aSur le probl-eme cosmologique01aTh,eorie
relativiste du champ non sym,etrique01aTh,eorie de la gravitation g,en,eralis,e
e aBUSbRECHERCHE2AcLIVPBUSkPh530.12 EINp052070606101417pam aL200313 450 00300
06000000080041000060350017000470400018000640200015000820820017000970990024001140
49000900138245007500147260003700222300003500259440004800294500007000342500014800
41250400510056065000380061165000570064965000380070665000450074465000370078970000
15008267100087008417110096009288520079010240CoLC950912s1993 nyua b 101
0 eng aocm28254715 aDLCcDLCdUKM a030644534400a530.1/43220 a530.19:530.14
5fINT aFPHC00aIntegrable quantum field theories /cedited by L. Bonora ... [et a
l.]. aNew York :bPlenum Press,c1993. aviii, 333 p. :bill. ;c26 cm. 0aNATO ASI
series.nSeries B,pPhysics ;v310 a"Published in cooperation with NATO Scientific
Affairs Division." aProceedings of a NATO Advanced Research Workshop on Integr
able Quantum Field Theories, held September 14-19, 1992, in Como, Italy"--T.p. v
erso. aIncludes bibliographical references and index. 0aQuantum field theoryxCo
ngresses. 0aIntegral solutionsxNumerical solutionsxCongresses. 0aMathematical ph
ysicsxCongresses. 7aChamps, Th,eorie quantique desxcongr-es 7aPhysique math,emat
iquexcongr-es10aBonora, L.2 aOrganisation du trait,e de l'Atlantique nord.bDivis
ion des affaires scientifiques2 aNATO Advanced Research Workshop on Integrable Q

³²Notons que les accents ne se présentent pas, sur l'écran de l'ordinateur, comme sur cette impression. Le "é" apparaît comme " e" à l'écran et comme ",e" à l'impression

Comparaison des notices provenant de la Bibliothèque Universitaire avec les notices du catalogue du CERN

Première méthode :

A partir du catalogue informatisé de la bibliothèque du CERN, il est possible de sélectionner tous les ouvrages caractérisés par un indice CDU et de se faire envoyer par courrier électronique les notices de ses ouvrages. Le système permet aussi de choisir le format des notices que l'on veut recevoir.

A partir de ces nouvelles listes et des listes envoyées par Grenoble il était donc possible de réaliser un programme de comparaison de ces fichiers texte.

Deuxième méthode :

Cependant on m'a indiqué qu'une autre méthode de comparaison était aussi réalisable : le catalogue du CERN est en effet interrogeable à partir de scripts en Shell ou Perl grâce à des fonctions spéciales prédéfinies. Des listes de notices des ouvrages du CERN ne sont alors pas indispensables pour réaliser cette comparaison.

La deuxième méthode possible était donc la suivante :

1 - Faire un programme qui sélectionne l'auteur et le titre de chaque notice de la liste de Grenoble (caractérisée par un CDU)

2 - interroger le catalogue de la bibliothèque du CERN (grâce aux fonctions spéciales) pour savoir s'il y existe un ouvrage de même titre et de même auteur. Si c'est le cas la notice est placée dans un fichier "match" qui contiendra alors toutes les notices des ouvrages communs aux deux bibliothèques , si ce n'est pas le cas la notice est placée dans un fichier "nomatch1" qui contiendra alors toutes les notices des ouvrages présents à Grenoble mais absents du CERN.

3 - Interroger le catalogue de la bibliothèque du CERN (grâce aux fonctions spéciales) pour sélectionner toutes les monographies caractérisées par le même CDU que précédemment.

4 - Pour chaque notice du catalogue de la bibliothèque du CERN, sélectionner (toujours grâce aux fonctions spéciales) le titre et l'auteur

5 - Chercher dans la liste provenant de Grenoble s'il y a un ouvrage de même titre et de même auteur. Si ce n'est pas le cas la notice est placée dans un fichier "nomatch2" qui contiendra alors toutes les notices des ouvrages présents au CERN mais absents de la Bibliothèque Universitaire de Grenoble.

L'ensemble des 3ème, 4ème et 5ème parties de cette méthode se sont avérées trop longues. J'ai donc choisi pour finir une troisième méthode qui concilie les deux premières et n'est pas trop lente.

Troisième méthode (Méthode réellement mise en oeuvre)³³ :

Elle utilise des listes de notices provenant du CERN et caractérisées par un CDU. Ces listes sont obtenues comme le précise la première méthode. Le format choisi pour les notices est le format 58 se présentant de la façon suivante :

<p>Sysno = 0251192 Main h = Mostepanenko, V M Title = The casimir effect and its applications Author = Mostepanenko, V M; Trunov, N N ed.; Znajek, R L ed. Impr.Pag= Oxford : Clarendon Press, 1997. - 199 p. Holdings= C 530.14 MOS</p> <p>Sysno = 0247799 Main h = Guidry, Mike Title = Gauge field theories : an introduction with applications Author = Guidry, Mike Impr.Pag= New York : Wiley, 1991. - 605 p. Holdings= C 530.14 GUI</p>
--

³³Un schéma récapitulatif de cette méthode est consultable en annexe

Dans cette méthode les deux premières étapes de la méthode précédente sont conservées.

Les étapes suivantes sont :

3 - Inscrire les numéros de système CERN de toutes les notices placées en fichier "match" dans un fichier ".ListSysno".

4 - Sélectionner le numéro de système de chaque notice présente sur la liste CERN

5 - Rechercher si ce même numéro de système est contenu dans la liste ".ListSysno". Si ce n'est pas le cas la notice est placée dans un fichier "nomatch2" qui contiendra alors toutes les notices des ouvrages présents au CERN mais absents de la Bibliothèque Universitaire de Grenoble.

IV - IV - Dernières études préparant à la programmation

Nous avons parlé auparavant de l'étude des notices dans le catalogue. Une fois la méthode de comparaison établie, les véritables formats des notices à traiter ont été eux aussi fixés. Une étude des particularités de ces formats a donc été nécessaire. Cette étude étant proche de celle faite pour le catalogue je ne ferai que citer les points les plus importants lorsque je traiterai de la programmation en elle-même..

V - LA PROGRAMMATION

V - I - Résumé des objectifs du programme

Le programme que j'ai écrit avait pour but de mettre en oeuvre la méthode de comparaison choisie afin d'obtenir les listes d'ouvrages désirées.

Ce programme devait aussi fournir certains résultats chiffrés comme le nombre d'ouvrages formant le fonds (monographies) de chaque établissement et le nombre d'ouvrages présents dans chacune des listes obtenues.

Pour finir il devait aussi permettre de consulter aisément les listes et les résultats chiffrés.

V - II - Choix du langage de programmation [4,5,6]

Au début de mon stage, Mme Geretschläger m'a informé qu'un autre programme de comparaison était déjà réalisé au CERN par Jean-Yves Le Meur, ingénieur informaticien chargé du catalogue de la bibliothèque du CERN. Ce programme permet notamment de savoir si les pré-tirages nouvellement inclus dans le catalogue de SLAC³⁴ sont présents dans le catalogue du CERN. Dans ce but, une comparaison est effectuée entre le numéro de rapport et le premier auteur des pré-tirages de SLAC et ceux des pré-tirages de la bibliothèque du CERN.

Ce programme interprétable par le shell UNIX (langage shell) était un modèle intéressant pour mon propre programme. J'ai donc repris son langage qui a en outre l'avantage d'être portable sur toute station (UNIX) du réseau.

D'autre part, le langage shell propose des commandes très intéressantes dans le cadre de mon travail sur des notices puisqu'il permet, par exemple :

- de sélectionner facilement une ligne commençant par "AUTEUR:", ou n'importe quelle autre expression, grâce à la commande "awk"

- de transformer une chaîne de caractère en une autre, grâce à la commande "sed" ce qui peut être utile pour le reformatage des auteurs et des titres et notamment pour la transformation des lettres accentuées en lettres non accentuées

³⁴Stanford Linear Accelerator Center

Ce langage permet aussi d'utiliser les fonctions spéciales propres à l'interrogation du catalogue de la bibliothèque du CERN.

Ayant eu à utiliser des variables de type tableau, j'ai plus précisément travaillé en Kornshell, langage qui permet de gérer ce type de variables.

VI - III - Ecriture et débogage des programmes

Les programmes que je vais décrire par la suite sont composés de nombreuses petites étapes qui ont été testées les unes après les autres dans de petits programmes.

Cette méthode m'a permis d'éliminer au fur et à mesure toutes les erreurs de programmation.

VI - IV - Les programmes³⁵

Voici plus précisément les opérations effectuées par les programmes que j'ai écrits et les différents fichiers qu'ils créent.

VI - IV - 1 - Programme numéro 1

Ce programme permet de faire un premier reformatage des listes fournies par Grenoble.

Grâce à ce reformatage :

- le fichier initialement constitué d'une seule ligne est transformé en un fichier de plusieurs lignes c'est-à-dire que des retours à la ligne sont insérés à chaque début de champ. Une ligne correspond alors (plus ou moins) à un champ

- les lignes correspondant aux champs de l'auteur principal, l'auteur secondaire, le titre et la collection sont signalées par "AUTEUR : ", "AUTEUR2 : ", "TITRE : ", "COLLECTION :"³⁶

- les fin de références sont signalées par "FIN_REF"

³⁵voir annexe

³⁶Notons que ce reformatage n'est pas parfait. Certaines lignes signalées par "AUTEUR:", par exemple, ne contiennent pas du tout d'informations sur les auteurs. Ce type d'erreur est corrigé par le programme suivant

VI - IV - 2 - Programme numéro 2

Ce programme lit la liste provenant de Grenoble. Il reconnaît le début et la fin de chaque notice, et, pour chaque notice, il répète les mêmes opérations :

L'élimination des conférences

Il cherche dans la notice si les mots "proceedings", "conference", "workshop", "compte rendu", "symposium", "congres", "lecture", "cours", "course" sont présents. Si c'est le cas les notices ne sont pas traitées.

Dans le cas contraire les notices sont comptées par un compteur et sont traitées comme suit.

La sélection

- le programme sélectionne les lignes commençant par "TITRE : " et vérifie qu'elles correspondent bien à des titres (ces lignes ne doivent pas contenir d'information sur l'indice CDU par exemple)

Si ces lignes correspondent à des titres ceux-ci sont transformés afin de ressembler aux titres présents dans le catalogue de la bibliothèque du CERN (les accents sont transformés, certaines parties de la ligne correspondant par exemple aux sous-titres ou à l'entête "TITRE : " sont éliminées,...)³⁷. Ces titres sont en plus transformés pour permettre une interrogation plus aisée du catalogue du CERN; ainsi les titres trop longs sont coupés.

Dans ces mêmes lignes sont présentes des informations sur l'édition de l'ouvrage que le programme sélectionne par ailleurs

- il sélectionne les lignes commençant par "AUTEUR : " et "AUTEUR2" et vérifie qu'elles correspondent bien à des auteurs . Si c'est le cas il ne conserve que l'auteur le plus important. Là encore un reformatage du champ auteur est effectué : seul le nom de famille est conservé.

- il sélectionne les lignes commençant par "COLLECTION", vérifie qu'elles correspondent bien à des collections et les formate

L'interrogation du catalogue de la bibliothèque du CERN

Grâce aux fonctions spéciales, il cherche dans la base BOOKS³⁸ du catalogue du CERN s'il existe des notices possédant les mêmes mots de titre et les mêmes mots d'auteur.

Selon la réponse il écrit les champs "AUTEUR : ", "TITRE", "EDITION :", "COLLECTION : " des notices dans un fichier texte intitulé .MATCH ou .NOMATCH1 (étape 2 de la méthode).

³⁷Les transformations effectuées sur les champs sont très nombreuses

³⁸Base correspondant aux monographies et compte-rendus de conférence

Il compte au fur et à mesure les notices insérées dans ces fichiers.
Il sélectionne aussi les numéros de système des notices retrouvées au CERN et les inscrit dans le fichier .LISTSYSNO (étape 3 de la méthode).

Fin du programme

Lorsque toutes les notices ont été traitées, le programme inscrit tous les chiffres comptés dans un fichier .RESULTAT

VI - IV - 3 - Programme numéro 3

Il lit les listes de notices provenant du CERN et reconnaît le début et la fin de chaque notice.

Pour chaque notice il répète les opérations suivantes :

- comptage
- sélection des lignes commençant par "Author", "Title", "Sysno" (étape 4 de la méthode), "Imprint"
- élimination de ces entêtes
- comparaison du numéro de système de la notice avec tous les numéros de systèmes présents dans le fichier .LISTSYSNO
- si ce numéro n'est pas trouvé dans la liste, les champs renommés "AUTEUR : ", "TITRE : " et "EDITION : " de la notice sont inscrits dans le fichier .NOMATCH2 (étape 5 de la méthode) et la notice est comptée

En fin de programme les chiffres comptés sont inscrits dans un fichier .RESULTAT2

VI - IV - 4 - Programme numéro 4

Il permet à l'utilisateur de visualiser les résultats ou les listes qu'il souhaite grâce à une interface faite de différents menus.

VI - V - Test final des programmes et évaluation des erreurs

Comme je l'ai annoncé au départ, nous pensions tout d'abord tester le programme sur de petites listes :

- CDU=530.14? (physique théorique) et année de publication =1970
- CDU=539.2? (physique de l'état solide) et année de publication = 1980

Finalement je n'ai demandé à Grenoble que les 6 listes précédemment décrites et j'ai testé mes programmes sur la première liste (CDU=530.1?).

Le test effectué consistait à faire tourner le programme sur cette liste puis à vérifier notamment que les premiers ouvrages présents dans la liste .MATCH appartenait bien au catalogue du CERN, que les premiers ouvrages présents dans la liste .NOMATCH1 étaient effectivement absents du catalogue du CERN et que les premiers ouvrages présents dans la liste .NOMATCH2 étaient effectivement absents du catalogue de Grenoble. J'ai aussi comparé le nombre de notices effectivement traitées par le programme par rapport au nombre total de notices reçues.

Plusieurs corrections ont été nécessaires pour que les résultats soient pertinents.

Ainsi des différences entre les deux catalogues qui n'avaient pas été remarquées auparavant, me sont apparues seulement après avoir fait tourner les programmes sur cette liste.

Je me suis, par exemple, aperçue que, dans le catalogue de la Bibliothèque du CERN, des espaces sont insérés systématiquement après les apostrophes ce qui n'est pas le cas dans le catalogue de la Bibliothèque Universitaire de Grenoble.

CERN :	L'etoile
GRENOBLE :	L'etoile

A cause de cette différence le programme considère que ces titres sont différents. J'ai donc dû améliorer le reformatage des titres.

Beaucoup d'autres erreurs ont été corrigées avant de parvenir à la version finale de mon programme.

Voici un résumé des principales erreurs qui persistent encore et une estimation de leur importance :

- Erreurs liées à l'impossibilité d'obtenir tous les ouvrages concernant un sujet à partir du CDU

Comme je l'ai déjà signalé, il manque aux listes qui nous sont fournies par Grenoble toutes les notices dans lesquelles l'indice CDU n'apparaît pas. Le programme est donc réalisé sur des listes incomplètes.

- Problèmes de faux doublons

Les faux doublons proviennent du fait que la technique de comparaison adoptée ne permet pas de faire la différence entre les différents volumes d'un même ouvrage .

- Erreurs dans l'élimination des conférences

Les notices contenant des mots caractéristiques des compte-rendus de conférence n'ont pas été traitées par ce programme. Il est possible qu'en voulant éliminer ces comptes-rendus, des monographies aient été en même temps éliminées de la comparaison. Je n'ai cependant pas trouvé cette erreur dans la cinquantaine de notices étudiées.

Par contre quelques rares conférences n'ont pas été éliminées.

- Erreurs dues à des classifications différentes au sein des deux catalogues

Il est possible que les deux bibliothèques aient parfois affecté à un même ouvrage un indice CDU différent ce qui fausserait les résultats. Cependant les CDU choisis pour cette étude caractérisent des sujets assez vastes. Ce type d'erreur devrait donc être négligeable.

- Erreurs liées à la nécessité d'avoir un titre et un auteur pour chaque ouvrage

La comparaison n'a été réalisée que sur des notices contenant à la fois les champs correspondant à l'auteur et au titre. Toutes les autres notices ne sont pas traitées par ce programme d'où là encore une certaine perte d'information. Cependant on peut considérer que presque toutes les monographies ont ces deux indications dans leur notice.

- Erreurs dans la liste des ouvrages présents seulement à Grenoble (.NOMATCH1)

- Erreurs dues à des différences de catalogage

Certains noms d'auteurs (russes par exemple) n'ont pas été écrits de la même façon dans les deux catalogues. Dans ce cas mon programme ne peut pas retrouver les ouvrages de ces auteurs.

exemples : Lifsic / Lifshitz
Braginskii/Braginsky

Les fautes de frappe au moment du catalogage empêchent aussi de retrouver un ouvrage.

Certains titres ayant été traduits, il arrive que les deux catalogues ne présentent pas le titre dans la même langue, les ouvrages ne sont, là encore, pas retrouvés.

La présence ou l'absence de guillemets dans le titre peut aussi engendrer des erreurs...

- Erreurs liées au programme

Même après un reformatage très rigoureux, il subsiste parfois, avec le titre, des informations qui ne devraient pas apparaître

exemples : Fuel-clad contract conductance cby J
Cle des schemas electriques, betude logique

Il arrive aussi que le titre soit mal coupé.

exemple : Layout of E

Dans tous ces cas, les ouvrages apparaîtront dans la liste .NOMATCH1 alors qu'ils ne le devraient pas.

- Erreurs dans la liste des ouvrages communs aux deux bibliothèques (.MATCH)

Si certains ouvrages sont absents de cette liste, se sera la répercussion des erreurs précédemment décrites.

En ce qui concerne l'insertion, dans cette liste, d'ouvrages qui ne devraient pas s'y trouver, les causes en sont les suivantes :

Afin de rechercher si un titre présent à Grenoble est aussi présent au CERN, le programme recherche dans le catalogue du CERN un titre composé des mêmes mots. Un titre composé de ces mots dans un autre sens ou composé de mots supplémentaires conviendra donc. (ces erreurs ont plus de chances de se produire pour des titres courts)

exemple : à partir de "Atomic physics and human knowledge" de Bohr on peut trouver ce même titre mais aussi "Essays 1958-1962 on atomic physics and human knowledge" de Bohr

- Erreurs dans la liste des ouvrages présents seulement au CERN (.NOMATCH2)

Ces erreurs sont la répercussion des erreurs précédentes.

En effet si un ouvrage du CERN est par erreur considéré identique à un ouvrage de Grenoble, il sera éliminé de la liste des ouvrages présents seulement au CERN.

TABLEAU RECAPITULATIF DES ERREURS :

TYPE D'ERREUR	POURCENTAGE
Sélection incomplète par le CDU	15 % des ouvrages manquent (peut-être)

Sur le total des notices sélectionnées :

monographies éliminées en même temps que les conférences	négligeable
--	-------------

Sur le total des notices restant après élimination des conférences :

conférences restantes	4%
monographies n'ayant pas à la fois titre et auteur	négligeable
plusieurs titres trouvés à partir d'un seul	< 5% des cas
traduction différente du nom de l'auteur	3%
titres dans des langues différentes	< 1%
autres différences d'écritures	négligeable
erreurs liées à une mauvaise sélection du titre	1%

Ces pourcentages ne peuvent pas s'additionner pour donner un pourcentage total d'erreurs mais l'étude du début des listes .MATCH et .NOMATCH (pour l'indice CDU=530.1?) permet d'aboutir au bilan suivant:

LISTE	POURCENTAGE D'ERREURS
liste .NOMATCH1 (ouvrages uniquement à Grenoble)	environ 10 % des ouvrages de cette liste sont en réalité aussi présents au CERN
liste .NOMATCH2 (ouvrages uniquement au CERN)	environ 10% des ouvrages de cette liste sont en réalité aussi présents à la B.U. Sciences de Grenoble
liste .MATCH (ouvrages présents dans les deux bibliothèques)	pourcentage négligeable d'ouvrages en réalité absents du CERN ou de la B.U. Sciences de Grenoble (0 trouvés sur 50 regardés)

VII - RESULTATS DE LA COMPARAISON³⁹

CDU	SUJET	FONDS TOTAL B.U. SCIENCES DE GRENOBLE	FONDS TOTAL CERN	Ouvrages présents dans les deux biblio- thèques	Ouvrages propres à la B.U. Sciences de Grenoble	Ouvrages propres au CERN
530.1?	Physique Théorique	380	1529	212	168	1333
539.1?	Physique nucléaire, atomique et moléculaire	393	2204	187	206	2042
539.2?	Physique de l'état solide	118	227	18	100	209
620?	Caractéri- sation des matériaux	207	435	40	167	423
621?	mécanique générale et électro- technique	1210	4309	252	958	4080
678?	Industrie des matières macro- moléculaires- caoutchouc- matières plastiques	41 ⁴⁰	143	2	39	145

³⁹Voir aussi graphiques de ces résultats en annexe

⁴⁰Dans le domaine de l'Industrie des matières macromoléculaires, la Bibliothèque Universitaire dispose en réalité d'un fonds plus important mais il est classé sous un autre CDU représentant la chimie macromoléculaire.

On constate que quel que soit le sujet concerné, la bibliothèque du CERN possède beaucoup plus de monographies que la Bibliothèque Universitaire de Grenoble. En physique théorique et en Physique nucléaire, atomique et moléculaire, surtout, le fonds du CERN est bien plus conséquent que celui de la Bibliothèque Universitaire ce qui s'explique bien entendu par la spécialisation du laboratoire du CERN.

Dans les autres domaines, il est probable que la bibliothèque du CERN dispose d'ouvrages très spécialisés alors que la Bibliothèque Universitaire couvre l'ensemble du domaine.

Dans les sujets les plus importants pour la Bibliothèque du CERN (CDU=530.1? et CDU= 539.1?) on constate qu'environ 50% des ouvrages disponibles à la Bibliothèque Universitaire le sont aussi au CERN ce qui valide en partie la politique d'acquisition passée du CERN.

Il reste cependant 50% des ouvrages de la Bibliothèque Universitaire qui sont absents du CERN.

Mme Geretschläger a étudié les listes de ses ouvrages et a pu noter qu'au moins 50% des ouvrages listés sont l'oeuvre d'auteurs réputés. Ils seraient donc très intéressants pour la bibliothèque du CERN.

VIII - APPLICATIONS POSSIBLES

Les résultats que je viens de présenter et les listes obtenues vont être présentés au Working Group for Acquisitions du CERN. Ils devraient être à l'origine de nouvelles acquisitions.

Une collaboration avec la Bibliothèque Universitaire Sciences de Grenoble pourrait aussi être envisagée puisque les deux établissements disposent d'ouvrages différents dans des sujets qui les intéressent tous les deux au plus haut point.

Cette étude devrait ainsi permettre d'améliorer le fonds de la Bibliothèque du CERN et de mieux satisfaire les utilisateurs de cette bibliothèque. Afin que cette amélioration ne se limite pas aux fonds antérieurs à 1997 et qu'elle puisse être faite aussi dans l'avenir, je travaille actuellement sur de nouveaux programmes.

Ceux-ci devraient permettre à Mme Geretschläger de comparer aisément, chaque année, les acquisitions des deux bibliothèques.⁴¹

⁴¹Ces programmes sont bien entendus très proches des programmes précédemment décrits

CONCLUSION

Ce projet de stage a été riche en apprentissages puisque j'ai été amenée à étudier notamment les fonctionnements de catalogues de bibliothèques et les règles de catalogage. J'ai aussi appris un langage de programmation que je ne maîtrisais pas à l'origine.

Au fur et à mesure des difficultés rencontrées, j'ai par ailleurs pu cerner le fonctionnement des bibliothèques.

La recherche d'une méthode de sélection de tous les ouvrages traitant d'un sujet donné, par exemple, a demandé de connaître l'historique des politiques de catalogage et des classifications des ouvrages au sein des bibliothèques.

De même les obstacles rencontrés dans le choix d'une méthodologie de comparaison ont conduit à rechercher différentes solutions. J'ai ainsi eu un aperçu des diverses manières permettant d'obtenir des informations à partir d'un système de gestion de bases de données.

Des méthodes plus efficaces auraient sans doute pu être adoptées dans certains cas. Peut-être, par exemple, aurait-il été intéressant d'importer les listes, reçues de Grenoble en format ISO 2709, dans une base de données, plutôt que de les traiter comme je l'ai fait dans mon programme. Mais ce programme n'était que la correction d'un autre programme réalisé sur ces listes, obtenues tout d'abord, en format "texte".

Il est aussi probable, que, débutant en langage Korn-Shell, je n'aie pas toujours choisi les commandes les plus adaptées aux traitements à effectuer.

En ce qui concerne le dépouillement des listes obtenues après traitement, mon analyse aurait pu être plus approfondie. Evaluer l'intérêt de ces listes sera la mission du Working Group for Acquisitions. Au delà de cette évaluation, il pourrait, je pense, être intéressant d'étudier les dates de publication des ouvrages compris dans ces listes. Cela permettrait de juger de la politique d'acquisition de manière historique.

Une extension de cette étude à d'autres sujets importants pour le CERN (mathématiques, informatique) serait aussi très intéressante. Dans ce cadre une comparaison avec d'autres organismes comme ETH (Zurich), ou un CADIST Mathématiques pourraient être envisagée..

Afin de mieux évaluer la qualité des fonds de la Bibliothèque d'autres méthodes pourraient aussi être utilisées comme les méthodes décrites dans un rapport de Martine Loiselet que m'a confié Béatrice Estéoule. Une interrogation ciblée de la base d'OCLC, par exemple, pourrait ainsi permettre d'obtenir des listes de monographies pertinentes sur des thèmes donnés.

BIBLIOGRAPHIE

- 1) **CERN**. (Page consulté le 2 septembre 1997). *CERN -- General info*, [En ligne]. Adresse URL : <http://www.cern.ch/CERN/GeneralInfo.html#z65>
- 2) **British Standard Institution**. *Universal decimal classification*. 3ème édition. London, 1961. 254 p.
- 3) **BEAUDIQUEZ Marcelle**. Les fichiers d'autorité BN-Opale. *Bulletins d'informations de l'Association des Bibliothécaires Français*, 1990, Numéro148. p30-44.
- 4) **DOUGHERTY, Dale; O'REILLY Tim**. *UNIX Text Processing*. Indianapolis, Ind. : Hayden, 1987. 665p.
- 5) **ARTHUR, Lowell Jay; BURNS, Edward N**. *UNIX shell programming*. 3ème édition. New York : Wiley, 1994. 462 p.
- 6) **RODGERS, Ulka**. *UNIX database management systems*. Englewood Cliffs, N.J. : Yourdon, 1990. 338 p.

ANNEXES

TABLE DES MATIERES

ANNEXE 1	
NORME ISO 27 09	1
ANNEXE 2	
Méthodologie de comparaison des deux catalogues - Schémas récapitulatifs	7
ANNEXE 3	
Programmes	10
PROGRAMME NUMERO 1	11
PROGRAMME NUMERO 2	13
PROGRAMME NUMERO 3	22
PROGRAMME NUMERO 4	27
ANNEXE 4	
Résultats de la comparaison - Graphiques	30

ANNEXE 1 : NORME ISO 27 09

Information and documentation — Format for information exchange

1 Scope

This International Standard specifies the requirements for a generalized exchange format which will hold records describing all forms of material capable of bibliographic description as well as other types of records. It does not define the length or the content of individual records and does not assign any meaning to tags, indicators or identifiers, these specifications being the functions of an implementation format.

This International Standard describes a generalized structure, a framework designed specially for communications between data processing systems and not for use as a processing format within systems.

2 Normative references

The following standards contains provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*.

ISO/IEC 10646-1:1993, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*.

3 Definitions

For the purposes of this International Standard the following definitions apply. They are listed in the order corresponding to figure 2.

3.1 record: Collection of fields, including a record label, a directory and data.

NOTE 1 If required, linking of records and their division into subrecords is allowed, which should be implemented as specified in the explicit exchange format.

3.2 field: Variable length portion of the record containing a particular category of data, following the directory and associated with one entry of the directory.

NOTE 2 A field may contain one or more subfields.

3.3 (subfield) identifier: Data element of one or more characters immediately preceding and identifying a subfield.

3.4 indicator: First data element, if present, associated with a field supplying further information about the contents of the field, about the relationship between the field and other fields in the record, or about the action required in certain data manipulation processes.

3.5 directory: Index to the location of the fields within a record.

3.6 record label: Field occurring at the beginning of each record providing parameters for the processing of the record.

3.7 directory map: Set of parameters specifying the structure of the entries in the directory.

3.8 separating character: Control character used to separate and qualify units of data logically, and in some cases hierarchically.

3.9 subfield: Part of a field containing a defined unit of information.

3.10 subrecord: Group of fields within a record which may be treated as an entity.

3.11 structure: Arrangement of the parts constituting a record.

3.12 (field) tag: Three characters associated with a field and used to identify it.

4 Structure of communication format for record

The general structure of a record is shown schematically in figure 1. A more detailed structure is shown schematically in figure 2, which includes four alternatives for the data fields.

Record label
Directory
Fields
Record separator

Figure 1 — General structure

A record contains the following fixed and variable length fields in the sequence shown in figure 2:

- a record label (fixed length);
- a directory (variable length);
- record identifier (variable length);
- reference fields (variable length);
- data fields (variable length);
- field separator(s), i.e. separator IS2 of ISO/IEC 646 or ISO/IEC 10646;
- record separator, i.e. separator IS3 of ISO/IEC 646 or ISO/IEC 10646.

The directory, record identifier, reference fields and data fields is terminated by a field separator. The record is terminated by the record separator.

4.1 Record label

The record label shown in figure 2 is fixed in length and defined as follows.

4.1.1 Record length (character positions 0 to 4)

The number of character positions in the record including the record label, directory, fields, and the record separator. The length is a 5-digit decimal number, right-aligned with zero fill if necessary.

3

NOTE 3 The record length described here is a logical record length. For practical reasons relating to machine processing of data in the magnetic tape environment, it can be necessary to divide the information into blocks.

4.1.2 Record status (character position 5)

A single character, to be defined in an implementation International Standard, describing the status of a record, for example, new or amended.

In the absence of an International Standard, special agreement shall be reached between the interchange partners.

4.1.3 Implementation codes (character positions 6 to 9)

The codes are not defined in this International Standard. Special agreement shall be reached between the interchange partners.

4.1.4 Indicator length (character position 10)

One decimal digit giving the number of character positions of the indicators.

If indicators are not used, the indicator length is set to zero.

4.1.5 Identifier length (character position 11)

One decimal digit giving the number of character positions of the identifier. The first or only character of this identifier shall always be IS1 of ISO/IEC 646 or ISO/IEC 10646.

If the identifier is not used, the identifier length is set to zero.

4.1.6 Base address of data (character position 12 to 16)

Five decimal digits, right-aligned with zero fill if necessary, equal to the combined length in characters of the record label and the directory including the field separator at the end of the directory.

4.1.7 Defined by user systems (character positions 17 to 19)

These positions are defined by user systems.

4.1.8 Directory map

Character position 20: One decimal digit equal to the length in characters of the length of field part of each entry in the directory.

Character position 21: One decimal digit equal to the length in characters of the starting character position part of each entry in the directory.

Character position 22: One decimal digit equal to the length in characters of the implementation-defined part of each entry in the directory.

Character position 23: Reserved for future use.

4.2 Directory

The directory consists of a variable number of entries each corresponding to its respective field (record identifier, reference and data fields). The directory ends with a field separator.

4.2.1 Directory entry

An entry consists of the following parts in the given order:

- a) a tag;
- b) the length of field;
- c) starting character position;
- d) implementation-defined part.

The length of the tag is three characters. No part of the entry shall exceed nine characters in length. All entries in a directory shall have the same structure.

4.2.2 Tag

Three characters which specify, according to definition in an implementation International Standard, the name of any associated field.

In the absence of an International Standard, special agreement shall be reached between the interchange partners.

4.2.3 Length of field

This length is either:

- a) the total number of characters [including indicator(s) and field separator] in the field indicated by the preceding tag; or

- b) zero, implying that the directory entry refers to a field whose total length is greater than the largest decimal number (n) which can be stored in the "length" of a directory entry. In this case, the field is regarded as being divided into a number of parts of which all but the last are of equal length (n). Each part has a corresponding directory entry containing the tag for the field and the starting character position of the part to which the directory entry refers. A length of zero indicates that the directory entry refers to a part of the field which is not the final part and that the length of this part is to be taken as (n); or
- c) the number of characters (including field separator) in the final part of a field which has been treated as described in b).

In the cases described in b) and c), all directory entries which refer to parts of the same field shall be adjacent and in sequence.

4.2.4 Starting character position

A decimal number giving the position of the first character of the field identified by the preceding tag, relative to the base address of data [i.e. the starting character position of the first field following the directory is 0 (zero)].

4.2.5 Implementation-defined part

The implementation-defined part of the entry, if present, contains control information relative to the field referenced by the entry.

4.3 Fields

All fields shall end with a field separator.

There are three types of field:

- a) record identifier field: tag 001¹⁾;
- b) reference fields: tags 002 to 009 and 0AA to 00Z²⁾ as required;
- c) data fields: tags 010 to 999 and 0AA to ZZZ as required.^{2) 3)}

1) 0 signifies zero.

2) For alphabetic characters, use either capital or small letters.

3) Any combination of numeric and alphabetic characters is allowed. When alphanumeric tags are used, they must not start with 00 since only reference fields start with two zeros.

4.3.1 Record identifier field

Characters identifying the record and assigned by the organization creating the record.

NOTE 4 The record identifier field does not contain indicators or identifiers.

4.3.2 Reference fields

A reference field supplies data which may be required for the processing of the record.

NOTE 5 Reference fields do not contain indicators or identifiers.

4.3.3 Data fields

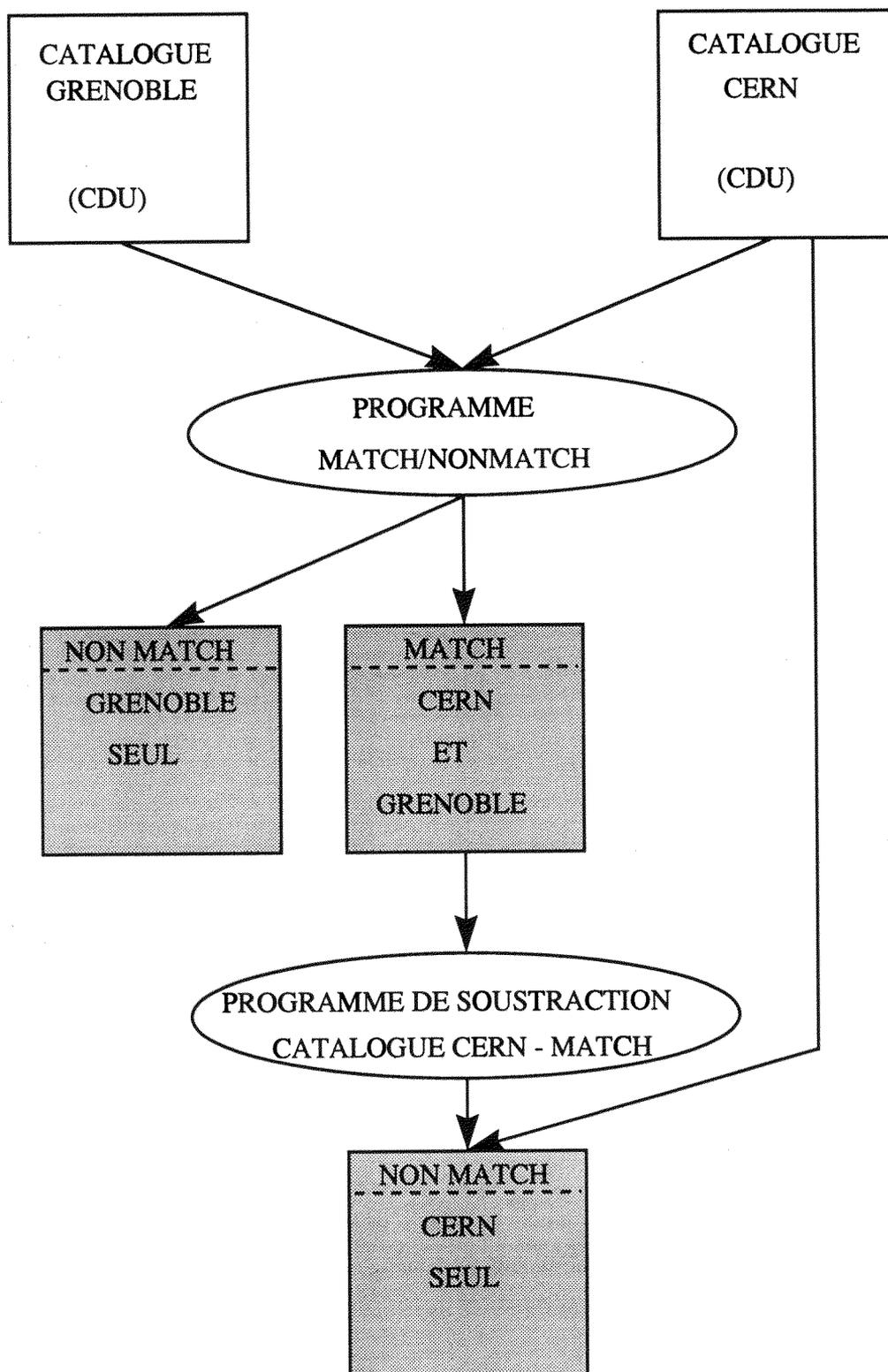
Each field consists of indicator(s) (optional), identifier(s) (optional), data and a field separator. The presence and length of the indicator(s) or identifier(s) are determined by the indicator length and identifier length as defined in the record label and shall be used consistently within each data field of the record.

Each data field in a record shall therefore be constructed according to one of the following alternatives.

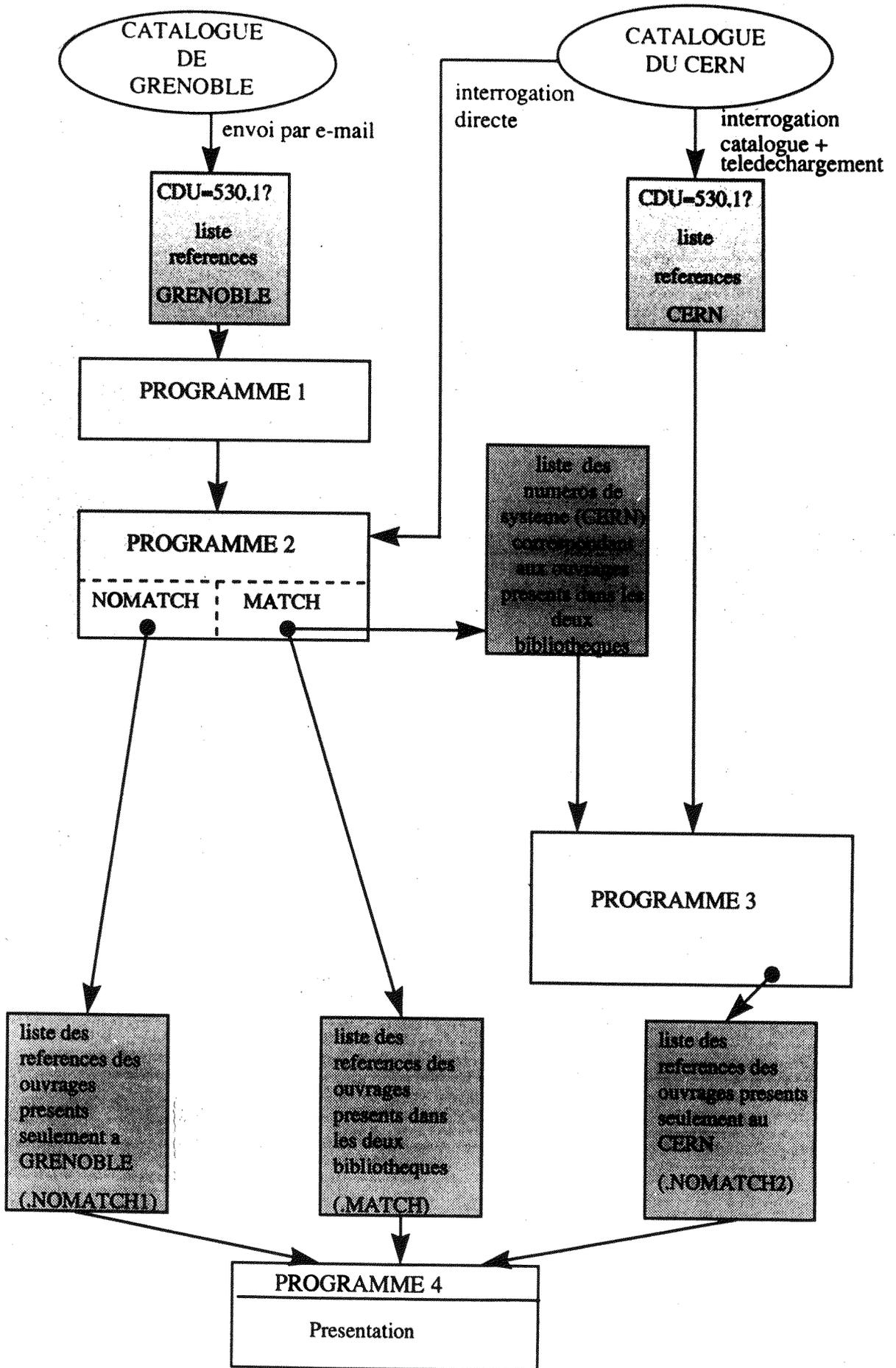
- a) **data:** In this case, the indicator length and the identifier length in the record label are set to zero. See figure 2, alternative 1.
- b) **identifier and data:** In this case, the indicator length in the record label is set to zero and the identifier length is set to one or more. See figure 2, alternative 2.
- c) **indicator and data:** In this case, the indicator length in the record label is set to one or more and the identifier length is set to zero. See figure 2, alternative 3.
- d) **indicator, identifier and data:** In this case, the indicator length and the identifier length in the record label are set to one or more. See figure 2, alternative 4.

**ANNEXE 2 : Méthodologie de comparaison des deux catalogues
Schémas récapitulatifs**

COMPARAISON DES FONDS DU CERN ET DE LA
BIBLIOTHEQUE UNIVERSITAIRE SCIENCES DE GRENOBLE



SCHEMA DES ENTREES ET SORTIES DES PROGRAMMES



ANNEXE 3 : Programmes

PROGRAMME NUMERO 1

#Ce programme effectue un preformatage des listes envoyees par Grenoble

```
echo "Nom du fichier contenant la liste de Grenoble : "
read input
echo "udc: "
read udc
cat $input > CERNtest.1
tr -s 'b''\012' < CERNtest.1 > CERNtest.2
```

```
#Definition de certaines zones TITRES:
sed 's/10.a/$TITRE : /g' CERNtest.2 > CERNtest.1
sed 's/00.a/$TITRE : /g' CERNtest.1 > CERNtest.2
sed 's/14.a/$TITRE : /g' CERNtest.2 > CERNtest.1
sed 's/04.a/$TITRE : /g' CERNtest.1 > CERNtest.2
```

```
#Definition de certaines zones AUTEUR:
sed 's/1 .a/$AUTEUR : /g' CERNtest.2 > CERNtest.1
sed 's/1B.a/$AUTEUR2 : /g' CERNtest.1 > CERNtest.2
```

```
#Definition de certaines zones COLLECTION:
sed 's/0.a/$COLLECTION : /g' CERNtest.2 > CERNtest.1
sed 's/0 .a/$COLLECTION : /g' CERNtest.1 > CERNtest.2
sed 's/4.a/$COLLECTION : /g' CERNtest.2 > CERNtest.1
```

```
#Marquage de la fin des references par l'inscription FIN_REF
sed 's/450 /$FIN_REF$/g' CERNtest.1 > CERNtest.2
```

```
#Insertion de "$" qui seront transformes plus tard en saut a la ligne
sed 's/03.a/$/g' CERNtest.2 > CERNtest.1
sed 's/ .aBUS$/g' CERNtest.1 > CERNtest.2
sed 's/2B.a/$/g' CERNtest.2 > CERNtest.1
sed 's/21.a/$: /g' CERNtest.1 > CERNtest.2
sed 's/31.a/$/g' CERNtest.2 > CERNtest.1
sed 's/01.a/$/g' CERNtest.1 > CERNtest.2
sed 's/7.a/$/g' CERNtest.2 > CERNtest.1
sed 's/2 .a/$/g' CERNtest.1 > CERNtest.2
sed 's/41.a/$/g' CERNtest.2 > CERNtest.1
```

```
#Ajout de l'inscription FIN_REF en fin de fichier
echo "$FIN_REF" >> CERNtest.1
```

```
#Suppression de sauts de lignes
tr -s '\012''b' < CERNtest.1 > CERNtest.2
```

```
#Transformation des "$" en saut de ligne
tr -s '$''\012' < CERNtest.2 > CERNtest.1
```

#Suppression de caracteres inutiles (lies a des problemes d'accents)

```
sed 's/Ê//g' CERNtest.1 > CERNtest.2
```

```
sed 's/Ê//g' CERNtest.2 > CERNtest.1
```

```
sed 's/,//g' CERNtest.1 > CERNtest.2
```

```
sed 's/â//g' CERNtest.2 > CERNtest.1
```

```
sed 's/Î//g' CERNtest.1 > CERNtest.2
```

```
sed 's/Ï//g' CERNtest.2 > CERNtest.1
```

```
sed 's/Ë//g' CERNtest.1 > CERNtest.2
```

```
sed 's/,//g' CERNtest.2 > CERNtest.1
```

```
sed 's/-//g' CERNtest.1 > CERNtest.2
```

```
sed 's/°//g' CERNtest.2 > GEAC2.$udc
```

#Suppression des fichiers temporaires crees par le programme

```
rm CERNtest.1
```

```
rm CERNtest.2
```

PROGRAMME NUMERO 2

```

# Ce programme permet de comparer une liste de references
#provenant du catalogue de la BU Sciences de Grenoble avec les
#references du catalogue du CERN
#2 fichiers sont crees :
# - 1 fichier contenant les references presentes dans les deux catalogues
#(input.match)
# - 1 fichier contenant les references des monographies presentes a
#Grenoble mais pas au CERN (input.nomatch1)

echo "Taper l'udc : "
read udc
input=~/GEAC2.$udc
ALEPHBIN=/export/home/libteam/aleph/query/bin
ud=`echo $udc | sed "s/\./\//g"`

#Creation des fichiers dans lesquels seront stockees differentes listes de references

echo "Liste des monographies communes aux deux catalogues (CDU = " $udc) : " >
$input.match
echo "" >> $input.match
echo "Liste des monographies presentes dans le catalogue de la BU Science de Grenoble" >
$input.nomatch1
echo "mais absentes du catalogue du CERN (CDU = " $udc ":" >> $input.nomatch1
echo "" >> $input.nomatch1

#Initialisation des differents compteurs de references
CompteurA=0
CompteurB=0
Compteur=0

#Lecture de la liste envoyee par Grenoble

cat $input | while read line
do
    linetest=`echo $line | sed 's/\./\//g'`
    i=1
    j=1
    k=1
    au1[1]="0_Auteur1"
    au2[1]="0_Auteur2"
    au1[2]="0_Auteur1"
    au2[2]="0_Auteur2"
    au1[3]="0_Auteur1"
    au2[3]="0_Auteur2"
    col[1]="0_col"
    ti="0_Titre"
    conf=faux

```

```

while [ $linetest != FIN_REF ]
do
  #Test pour savoir si la reference correspond a une conference
  echo $line > fich1
  awk '/[pP]roceedings /' fich1 > fichconf
  awk '/[cC]onference/' fich1 >> fichconf
  awk '/[wW]orkshop/' fich1 >> fichconf
  awk '/[cC]ompte [rR]endu/' fich1 >> fichconf
  awk '/[sS]ymposium/' fich1 >> fichconf
  awk '/[cC]ongres /' fich1 >> fichconf
  awk '/[IL]ecture/' fich1 >> fichconf
  awk '/[cC]ours /' fich1 >> fichconf
  awk '/[cC]ourse/' fich1 >> fichconf

  if [ -s fichconf ]
  then
    conf=vrai
  else
    #Traitement de tout ce qui n'est pas une conference

    #Selection des auteurs principaux
    awk '/^AUTEUR : [ a-zA-Z][ a-z ]*, [ A-Z ]/' fich1 > fich2
    awk '/^AUTEUR : [ a-zA-Z][ a-z ]* [ A-Z ][ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^AUTEUR : [ a-zA-Z][ a-z ]* [ A-Z ][ a-z ]* [ A-Z ][ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^AUTEUR : [ a-zA-Z][ a-z ]*- [ a-z ]*, [ A-Z ]/' fich1 >> fich2

    if [ -s fich2 ]
    then
      au1[$i]=$line
      au1[$i]=`echo ${au1[$i]}|sed "s/AUTEUR : //g"`
      au1[$i]=`echo ${au1[$i]}|sed "s/,.*//g"`
      let i=i+1
    fi

    #Selection des auteurs secondaires
    awk '/^AUTEUR2 : [ a-zA-Z][ a-z ]* [ A-Z ][ a-z ]*, [ A-Z ]/' fich1 > fich2
    awk '/^AUTEUR2 : [ a-zA-Z][ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^AUTEUR2 : [ a-zA-Z][ a-z ]*- [ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^TITRE : [ a-zA-Z][ a-z ]* [ A-Z ][ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^TITRE : [ a-zA-Z][ a-z ]*, [ A-Z ]/' fich1 >> fich2
    awk '/^TITRE : [ a-zA-Z][ a-z ]*- [ a-z ]*, [ A-Z ]/' fich1 >> fich2

    if [ -s fich2 ]
    then
      au2[$j]=$line
      au2[$j]=`echo ${au2[$j]}|sed "s/AUTEUR2 : //g"`
      au2[$j]=`echo ${au2[$j]}|sed "s/TITRE : //g"`
      au2[$j]=`echo ${au2[$j]}|sed "s/,.*//g"`
      let j=j+1
    fi

    #Selection du titre
    awk '/^TITRE/' fich1 > fich2
    if [ -s fich2 ]
    then

```

```

awk '/$ud/' fich2 > fich3
awk '/[ A-Z ][ 0-9 ][ 0-9 ][ 0-9 ][ 0-9 ][ 0-9 ]/' fich2 >> fich3
awk '/^TITRE : [ a-zA-Z][ a-z ]* [ A-Z ][ a-z ]*, [ A-Z ]/' fich2 >> fich3
awk '/^TITRE : [ a-zA-Z][ a-z ]*, [ A-Z ]/' fich2 >> fich3
awk '/^TITRE : [ a-zA-Z][ a-z ]*- [ a-z ]*, [ A-Z ]/' fich2 >> fich3
if [ -s fich3 ]
then ChaineTitre=faux
else
  ChaineTitre=vrai
  ti=$line
  ti=`echo ${ti} | sed "s/TITRE : //g"`
  ti=`echo ${ti} | sed "s/,.*//g"`
  ti=`echo ${ti} | sed "s/√.*//g"`
  ti=`echo ${ti} | sed "s/√.*//g"`
  ti=`echo ${ti} | sed "s/∧.*//g"`
  ti=`echo ${ti} | sed "s/:b.*//g"`
  ti=`echo ${ti} | sed "s/ :.*//g"`
  ti=`echo ${ti} | sed "s/∧.*//g"`
  ti=`echo ${ti} | sed "s/ =.*//g"`
  ti=`echo ${ti} | sed "s/aN.*//g"`
  ti=`echo ${ti} | sed "s/aC.*//g"`
  ti=`echo ${ti} | sed "s/b[ A-Z ].*//g"`
  ti=`echo ${ti} | sed "s/bg.*//g"`
  ti=`echo ${ti} | sed "s/btext.*//g"`
  ti=`echo ${ti} | sed "s/bwi.*//g"`
  ti=`echo ${ti} | sed "s/bpar.*//g"`
  ti=`echo ${ti} | sed "s/b(.*/g"`
  ti=`echo ${ti} | sed "s/,.cp.*//g"`
  ti=`echo ${ti} | sed "s/,.cb.*//g"`
  ti=`echo ${ti} | sed "s/,.ct.*//g"`
  ti=`echo ${ti} | sed "s/cvon.*//g"`
  ti=`echo ${ti} | sed "s/cpar.*//g"`
  ti=`echo ${ti} | sed "s/cedited.*//g"`
  ti=`echo ${ti} | sed "s/(.*//g"`
  ti=`echo ${ti} | sed "s/--.*//g"`
  ti=`echo ${ti} | sed "s/OB.*//g"`
  ti=`echo ${ti} | sed "s/Bc.*//g"`
  ti=`echo ${ti} | sed "s/√English.*//g"`
  #Transformation du titre pour preparer a l'interrogation d'ALICE
  ti2=`echo $ti | sed "s/^The //"`
  ti2=`echo $ti2 | sed "s/^A //g"`
  ti2=`echo $ti2 | sed "s/^An //g"`
  ti2=`echo $ti2 | sed "s/^La //g"`
  ti2=`echo $ti2 | sed "s/^Le //g"`
  ti2=`echo $ti2 | sed "s/^Les //g"`
  ti2=`echo $ti2 | sed "s/^Un //g"`
  ti2=`echo $ti2 | sed "s/^Une //g"`
  ti2=`echo $ti2 | sed "s/^Die //g"`
  ti2=`echo $ti2 | sed "s/^Der //g"`
  ti2=`echo $ti2 | sed "s/^Das //g"`
  ti2=`echo $ti2 | sed "s/^Ein //g"`
  ti2=`echo $ti2 | sed "s/^Eine //g"`
  ti2=`echo $ti2 | sed "s/^El //g"`
  ti2=`echo $ti2 | sed "s/^Los //g"`

```

```
ti2=`echo $ti2|sed "s/^Una //g"`
ti2=`echo $ti2|sed "s/^/ /g"
```

```
ti2=`echo $ti2|sed "s/to / AND /g"`
longueur=`echo $ti2|wc -C`
if [ $longueur -gt 46 ]
then
ti2=`echo ${ti2}|sed "s/^\{46\}\).*^1/g"`
ti2=`echo ${ti2}|sed "s/[^]*$/g"`
ti2=`echo ${ti2}|sed "s/ AND $/g"`
ti2=`echo ${ti2}|sed "s/ AND$/g"`
fi
ti3=`echo ${ti}|sed "s/c$/g"`
ti4=`echo ${ti2}|sed "s/c$/g"
```

```
#Selection de l'imprint
awk '/c1...\./' fich2 > fich3
if [ -s fich3 ]
then
datetemp=`echo $lin|sed "s/.*c1/1/g"`
datetemp=`echo $datetempl|sed "s/.*$/g"`
echo $datetemp > fich4
awk '/CM/' fich4 > fich3
if [ -s fich3 ]
then
echo "cm"
else
date=$datetemp
imprint=`echo $lin|sed "s/TITRE : $ti//g"`

imprint=`echo $imprint|sed "s/C1.*//g"`
fi
fi
```

```
awk '/c[1...]/' fich2 > fich3
if [ -s fich3 ]
then
date=`echo $lin|sed "s/.*c[1/1/g]"`
date=`echo $date|sed "s/].*//g]"`
imprint=`echo $lin|sed "s/TITRE : $ti//g]"`
```

```
imprint=`echo $imprint|sed "s/C[1.*//g]"`
fi
awk '/c[c1...]/' fich2 > fich3
if [ -s fich3 ]
then
date=`echo $lin|sed "s/.*c[c1/1/g]"`
date=`echo $date|sed "s/].*//g]"`
imprint=`echo $lin|sed "s/TITRE : $ti//g]"`
```

```
imprint=`echo $imprint|sed "s/C[c1.*//g]"`
```

```
fi
```

```

awk '/[cC][cC][oO][pP]/ fich2 > fich3
if [ -s fich3 ]
then
date=`echo $linelstd "s/*[cC][cC][oO][pP]\ //g"`
date=`echo $datelstd "s/\.*//g"`

imprint=`echo $linelstd "s/TITRE : $ti//g"`

imprint=`echo $imprintlstd "s/[cC][cC][oO][pP].*//g"`
fi

imprint=`echo $imprintlstd "s/,c$/,/g"`
imprint=`echo $imprintlstd "s/,.c$/,/g"`
imprint=`echo $imprintlstd "s/*\.. .a//g"`
imprint=`echo $imprintlstd "s/*.:.a//g"`
imprint=`echo $imprintlstd "s/*\..a//g"`
imprint=`echo $imprintlstd "s/*\..a//g"`

imprint=`echo $imprintlstd "s/*\]. .a//g"`
imprint=`echo $imprintlstd "s/[ ,: ].b/: /g"`
imprint=`echo $imprintlstd "s/\].b/: /g"`

imprint=`echo $imprintlstd "s/:.a/, /g"`
imprint=`echo $imprintlstd "s/^a//g"`
imprint=`echo $imprintlstd "s/^b//g"`

fi

fi

#Selection de la collection
awk '/^COLLECTION/ fich1 > fich2
if [ -s fich2 ]
then
awk '/$ud/ fich2 > fich3
awk '/539/ fich2 >> fich3
awk '/537/ fich2 >> fich3
awk '/530/ fich2 >> fich3
awk '/531/ fich2 >> fich3
awk '/547\./ fich2 >> fich3
awk '/548\./ fich2 >> fich3
awk '/516\./ fich2 >> fich3
awk '/621\./ fich2 >> fich3
awk '/duteng/ fich2 >> fich3

awk '/^spa/ fich2 >> fich3

awk '/freeng/ fich2 >> fich3
awk '/^eng/ fich2 >> fich3
awk '/engfre/ fich2 >> fich3

```

```
awk '/^fre/' fich2 >> fich3
```

```
awk '/^ger/' fich2 >> fich3
```

```
awk '/pprox/' fich2 >> fich3
```

```
awk '/corganise/' fich2 >> fich3
```

```
awk '/x[ A-Z ]/' fich2 >> fich3
```

```
if [ -s fich3 ]
```

```
then ChaineCollection=faux
```

```
else
```

```
    ChaineCollection=vrai
```

```
    col[$k]=$line
```

```
    col[$k]=`echo ${col[$k]}|sed "s/COLLECTION : //g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/,.*//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/^[ a-z ].*//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/^[^ ]*//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/?.*?//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/;.*//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/a[ A-Z ].*//g"`
```

```
    col[$k]=`echo ${col[$k]}|sed "s/a^[ A-Z ].*//g"`
```

```
    let k=k+1
```

```
fi
```

```
fi
```

```
fi
```

```
read line
```

```
linetest=`echo $line|sed "s/.*//g"`
```

```
done
```

```
if [ $conf = faux ]
```

```
then
```

```
    au1test=`echo ${au1[1]}|sed "s/.*//g"`
```

```
    au2test=`echo ${au2[1]}|sed "s/.*//g"`
```

```
    titretest=`echo ${ti}|sed "s/.*//g"`
```

```
    testcol=`echo ${col[1]}|sed "s/.*//g"`
```

```
    testdate=`echo ${date}|sed "s/.*//g"`
```

```
    testimprint=`echo ${imprint}|sed "s/.*//g"`
```

```
if [ $titretest != "0_Titre" ]
```

```
then
```

```
    if [ ${au1test} != "0_Auteur1" -o ${au2test} != "0_Auteur2" ]
```

```
    then
```

```
        if [ ${au1test} != "0_Auteur1" ]
```

```
        then
```

```
            echo $ti2
```

```

        auteur=${au1[1]}
        echo ${au1[1]}
        echo $imprint $date
        echo ${col[1]}
    else

        echo $ti2
        auteur=${au2[1]}
        echo ${au2[1]}
        echo $imprint $date
        echo ${col[1]}
    fi
    let Compteur=Compteur+1

```

#Comparaison des references fournies par Grenoble avec celles d'ALICE

```

#Comparaison avec le premier type de titre
$ALEPHBIN/SelectScope BOOKS "wti=$ti4 AND wau=$auteur" > temp
set_number=`awk '{print $1}' temp`
doc_number=`awk '{print $2}' temp`
echo $set_number
echo $doc_number

    if [ $doc_number != "0" ]
    then
        #si une Reference identique est trouvee dans ALICE alors elle
est placee dans un fichier .match

        let CompteurA=CompteurA+1
        echo "TITRE    :" ${ti3} >> $input.match
        echo "AUTEUR   :" ${auteur} >> $input.match
        if [ $testimprint != "" -o $testdate != "" ]
        then
            echo "EDITION  :" ${imprint} ${date} >> $input.match
        fi
        if [ $testcol != "0_col" ]
        then
            echo "COLLECTION:" ${col[1]} >> $input.match
        fi

        echo "" >> $input.match
        #les System number des references d'ALICE trouvees sont places
dans un fichier .ListSysno

        i=1
        while [ $i -le doc_number ]
        do
            $ALEPHBIN/MultiFetch CERCER $set_number $i $i > enreg

            cernno=`$ALEPHBIN/GetField "\`cat enreg\`" sysno`
            echo $cernno >> $input.ListSysno
            let i=i+1
        done

```

```

else
    #Comparaison avec le deuxieme type de titre
    $ALEPHBIN/SelectScope BOOKS "wti=$ti2 AND wau=$auteur" >
temp
    set_number=`awk '{print $1}' temp`
    doc_number=`awk '{print $2}' temp`
    echo $set_number
    echo $doc_number

    if [ $doc_number != "0" ]
    then
        #si une Reference identique est trouvee dans ALICE alors elle
est placee dans un fichier .match
        let CompteurA=CompteurA+1
        echo "TITRE      :" ${ti3} >> $input.match
        echo "AUTEUR     :" ${auteur} >> $input.match
        if [ $testimprint != "" -o $testdate != "" ]
        then
            echo "EDITION  :" ${imprint} ${date} >> $input.match
        fi
        if [ $testcol != "0_col" ]
        then
            echo "COLLECTION:" ${col[1]} >> $input.match
        fi

        echo "" >> $input.match
        #les System number des references d'ALICE trouvees sont places
dans un fichier .ListSysno
        i=1
        while [ $i -le doc_number ]
        do
            $ALEPHBIN/MultiFetch CERCER $set_number $i $i > enreg

            cernno=`$ALEPHBIN/GetField "\`cat enreg\`" sysno`
            echo $cernno >> $input.ListSysno
            let i=i+1
        done

    else

        let CompteurB=CompteurB+1

        #Reference non trouvee dans ALICE placee dans un
fichier.nomatch
        echo "TITRE      :" ${ti3} >> $input.nomatch1
        echo "AUTEUR     :" ${auteur} >> $input.nomatch1
        if [ $testimprint != "" -o $testdate != "" ]
        then
            echo "EDITION  :" ${imprint} ${date} >> $input.nomatch1
        fi
        if [ $testcol != "0_col" ]
        then
            echo "COLLECTION:" ${col[1]} >> $input.nomatch1
        fi
    fi

```

```
        echo "" >> $input.nomatch1
    fi
fi
fi
fi

done

#Resultats chiffres places dans un fichier .resultat
echo "RESULTATS POUR UDC = " $udc > $input.resultat
echo " Nombres d ouvrages de la BU Sciences de Grenoble : " $Compteur >> $input.resultat
echo " Nombres d ouvrages communs aux deux bibliotheques : " $CompteurA >> $input.resultat
echo " Nombres d ouvrages presents a la BU Sciences de Grenoble mais absents de la
Bibliotheque du CERN : " $CompteurB >> $input.resultat
echo "" >> $input.resultat

#Suppression des fichiers temporaires crees par le programme
rm fich1
rm fich2
rm fich3
rm fich4
rm temp
rm enreg
```

PROGRAMME NUMERO 3

i# Ce programme lit une liste de references (format 58) provenant du catalogue du CERN (cette liste a ete envoyee par mail) et cherche si les numeros de systeme de ces references appartiennent a une liste contenue dans un fichier .Sysno

Le fichier .Sysno contient la liste des numeros de systemes des monographies presentent a la fois au CERN et a la bibliotheque de Grenoble

Ce programme selectionne les references des ouvrages dont le numero de systeme n'appartient pas a la liste .Sysno et les place dans un fichier .nomatch2 (il y place donc toutes les references des ouvrages presents uniquement au CERN)

```
echo "udc : "
read udc
Compteur=0
CompteurB=0
input=GEAC2.$udc
echo "Liste des monographies presentes dans le catalogue du CERN" > $input.nomatch2
echo "mais absentes du catalogue de la BU Sciences de Grenoble : " >> $input.nomatch2
echo "" >> $input.nomatch2
echo "SysnoFin" >> mail/AL.$udc
```

#Lecture du fichier contenant la liste de references CERN jusqu'au debut de la premiere reference

```
cat mail/AL.$udc | while read line
do
    debut=faux
    while [ $debut = faux ]
    do
        read line
        echo $line
        echo $line > enr
        awk '/Sysno/' enr > enrsysno
        if [ -s enrsysno ]
        then
            debut=true
            echo "debut"
        fi
    done
    #Selection du premier numero de systeme
    sysno=`awk -F= '{print $2}' enrsysno`
    echo $sysno
    testline=`echo $line | sed 's/ .*//g'`
```

```
read line
```

```
while [ $testline != SysnoFin ]
do
    debut=faux
    trouve=faux
```

```
#Recherche du numero de systeme de chaque reference dans la liste .LISTSYSNO
```

```

cat $input.ListSysno|while read line
do

    if [ $trouve = faux ]
    then
        if [ $sysno = `echo $line` ]
        then
            trouve=vrai
            echo "ligne " $line
            echo "trouve"
        fi
    fi

fi

done

```

#Pour toutes les references ayant un numero de systeme n'appartenant pas a la liste .LISTSYSNO, le titre, l'auteur et l'edition sont selectionnes et places dans le fichier .NOMATCH2

```

if [ $trouve = faux ]
then
let CompteurB=CompteurB+1
echo "ComptB" $CompteurB

while [ $debut = faux ]
do

echo $line > enr

#Selection du titre
awk '/Title/' enr > enrtitre

if [ -s enrtitre ]
then
titre=`awk -F= '{print $2}' enrtitre`
read line
echo $line > enr2
awk '/^=/' enr2 > enrtitre

if [ -s enrtitre ]
then
titre2=""
else
titre2=`echo $line`
read line
fi

titre=`echo $titre $titre2`
titre=`echo $titre|sed 's/^/o/g'`
titre=`echo $titre|sed 's/È/c/g'`
titre=`echo $titre|sed 's/†/a/g'`
titre=`echo $titre|sed 's/É/c/g'`
titre=`echo $titre|sed 's/_/u/g'`

```

```

titre=`echo $titre | sed 's/°o/a/g'`
titre=`echo $titre | sed 's/Û/o/g'`
titre=`echo $titre | sed 's/Ï/i/g'`
titre=`echo $titre | sed 's/Á/c/g'`
titre=`echo $titre | sed 's/Í/e/g'`
titre=`echo $titre | sed 's/Ò/n/g'`
titre=`echo $titre | sed 's/,/a/g'`
titre=`echo $titre | sed 's/Û/o/g'`
titre=`echo $titre | sed 's/,/u/g'`
titre=`echo $titre | sed 's/^/o/g'`
echo "TITRE : " $titre >> $input.nomatch2
echo "titre:" $titre
else

#Selection de l'auteur
awk '/Author/' enr > enrauteur
awk '/Corp /' enr >> enrauteur
if [ -s enrauteur ]
then
auteur=`awk -F= '{print $2}' enrauteur`
auteur=`echo $auteur | sed 's/È/e/g'`
auteur=`echo $auteur | sed 's/±/a/g'`
auteur=`echo $auteur | sed 's/Ë/e/g'`
auteur=`echo $auteur | sed 's/,/u/g'`
auteur=`echo $auteur | sed 's/°o/a/g'`
auteur=`echo $auteur | sed 's/Û/o/g'`
auteur=`echo $auteur | sed 's/Ï/i/g'`
auteur=`echo $auteur | sed 's/Á/c/g'`
auteur=`echo $auteur | sed 's/Í/e/g'`
auteur=`echo $auteur | sed 's/Ò/n/g'`
auteur=`echo $auteur | sed 's/,/a/g'`
auteur=`echo $auteur | sed 's/Û/o/g'`
auteur=`echo $auteur | sed 's/,/u/g'`
auteur=`echo $auteur | sed 's/^/o/g'`
echo "AUTEUR : " $auteur >> $input.nomatch2
echo "auteur : " $auteur
read line
else

#Selection de l'edition
awk '/Impr./' enr > enredition
if [ -s enredition ]
then
edition=`awk -F= '{print $2}' enredition`
read line
echo $line > enr2
awk '/^=' enr2 > enredition

if [ -s enredition ]

```

```

then
edition2=""
else
edition2=`echo $line`
read line
fi

edition=`echo $edition $edition2`
edition=`echo $edition|sed 's/È/c/g'`
edition=`echo $edition|sed 's/Ë/a/g'`
edition=`echo $edition|sed 's/Ê/c/g'`
edition=`echo $edition|sed 's/Û/u/g'`
edition=`echo $edition|sed 's/Ö/a/g'`
edition=`echo $edition|sed 's/Û/o/g'`
edition=`echo $edition|sed 's/Ï/i/g'`
edition=`echo $edition|sed 's/Ä/c/g'`
edition=`echo $edition|sed 's/Í/e/g'`
edition=`echo $edition|sed 's/Ô/n/g'`
edition=`echo $edition|sed 's/,/a/g'`
edition=`echo $edition|sed 's/Û/o/g'`
edition=`echo $edition|sed 's/,/u/g'`
edition=`echo $edition|sed 's/^/o/g'`
echo "EDITION : " $edition >> $input.nomatch2
echo "Edition : "$edition
read line
else

#Selection du numero de systeme de la reference suivante
awk '/Sysno/ enr > enrsysno
if [ -s enrsysno ]
then
    debut=true
    sysno=`awk -F= '{print $2}' enrsysno`
    echo $sysno
    testline=`echo $line|sed 's/ .*//g'`
    read line

else
    read line
fi
fi
fi
fi

done
echo "" >> $input.nomatch2

fi

```

```
#Les references dont le numero de systeme est trouve dans le fichier .ListSysno sont juste lues
```

```

    if [ $trouve = vrai ]
    then

        while [ $debut = faux ]
        do

            echo $line
            echo $line > enr

            awk '/Sysno/' enr > enrsysno
            if [ -s enrsysno ]
            then
                debut=true
                sysno=`awk -F= '{print $2}' enrsysno`
                echo $sysno

                testline=`echo $line | sed 's/ .*//g'`
                read line

            else
                read line
            fi
        done
    fi

    let Compteur=Compteur+1

    echo "compt" $Compteur
    echo ""
    sysno=`awk -F= '{print $2}' enrsysno`
    echo $sysno
done

#Les fichiers temporaires crees par ce programme sont effaces
rm enr
rm enr2
rm enrtitre
rm enrauteur
rm enrsysno
rm enredition

#Resultats chiffres places dans le fichier .RESULTAT2
echo " RESULTATS POUR UDC = " $udc > $input.resultat2
echo " Nombres d ouvrages du CERN : " $Compteur >> $input.resultat2
echo " Nombres d ouvrages presents uniquement au CERN : " $CompteurB >> $input.resultat2
echo "" >> $input.resultat2
done
```

PROGRAMME NUMERO 4

(Ce programme de présentation est en réalité composé de deux programmes :
PROGR49 et PROGR50)

PROGR49:

```
rep=pasrep
while [ $rep != 1 -a $rep != 2 -a $rep != 3 -a $rep != 4 -a $rep != 5 -a $rep != 6 -a $rep != q ]
do
clear

echo "Resultats de la comparaison des Catalogues (monographies)"
echo "du CERN et de la Bibliotheque Universitaire de Grenoble"
echo ""
echo "-- Pour visaliser les resultats obtenus pour le CDU= 530.1? TAPER 1"
echo "-- Pour visaliser les resultats obtenus pour le CDU= 539.1? TAPER 2"
echo "-- Pour visaliser les resultats obtenus pour le CDU= 539.2? TAPER 3"
echo "-- Pour visaliser les resultats obtenus pour le CDU= 620? TAPER 4"
echo "-- Pour visaliser les resultats obtenus pour le CDU= 621? TAPER 5"
echo "-- Pour visaliser les resultats obtenus pour le CDU= 678? TAPER 6"
echo "-- Pour quitter le programme TAPER q"
read rep
done

case $rep in
  1)
    progr50 530.1;;
  2)
    progr50 539.1;;
  3)
    progr50 539.2;;
  4)
    progr50 620;;
  5)
    progr50 621;;
  6)
    progr50 678;;
  q)
    clear
    echo "Aurevoir"
esac
```

PROGR50:

```

clear
fichier=GEAC.$1?.resultat
echo $fichier
cat $fichier
echo "-----"
rep=rien
while [ $rep != 1 -a $rep != 2 -a $rep != 3 -a $rep != m -a $rep != q ]
do

echo "-- Pour Visualiser la liste des ouvrages communs aux deux catalogues TAPER 1"
echo "-- Pour visualiser la liste des ouvrages presents uniquement au CERN TAPER 2"
echo "-- Pour visualiser la liste des ouvrages presents uniquement a Grenoble TAPER 3"
echo "-- Pour revenir au menu precedent TAPER m"
echo "-- Pour quitter le programme TAPER q"
read rep
clear
done

case $rep in
1)
echo "-- POUR LIRE LA SUITE APPUYER SUR RETURN"
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"
cat GEAC.$1?.matchlmore
echo ""
echo ""
echo ""
echo ""
echo ""
echo ""
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"
read q
progr50 $1;;
2)
echo "-- POUR LIRE LA SUITE APPUYER SUR RETURN"
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"
cat GEAC.$1?.nomatch2lmore
echo ""
echo ""
echo ""
echo ""
echo ""
echo ""
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"
read q
progr50 $1;;
3)
echo "-- POUR LIRE LA SUITE APPUYER SUR RETURN"
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"
cat GEAC.$1?.nomatch1lmore
echo ""
echo ""
echo ""

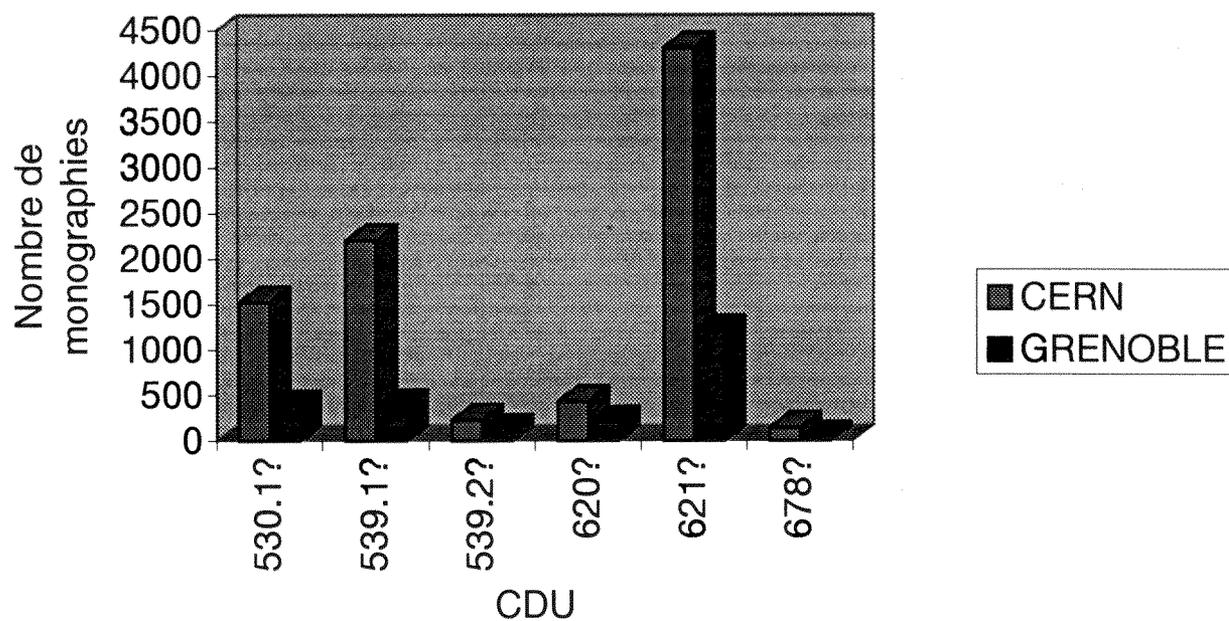
```

```
echo ""  
echo ""  
echo ""  
echo "-- POUR QUITTER LA LISTE ET REVENIR AU MENU TAPER q"  
read q  
progr50 $1;;  
q)  
echo "Aurevoir";;  
m)  
progr49;;
```

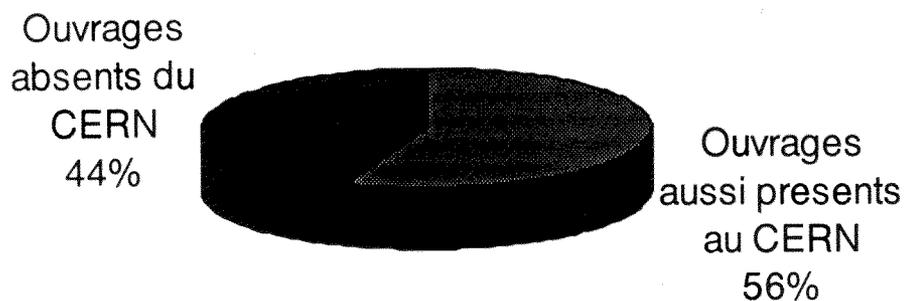
```
esac
```

**ANNEXE 4 : Résultats de la comparaison
Graphiques**

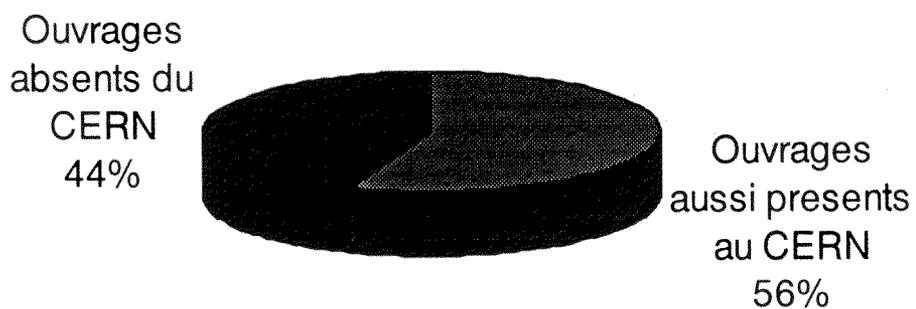
Comparaison des Fonds (monographies) de la Bibliotheque du CERN et de la Bibliotheque Universitaire Sciences de Grenoble



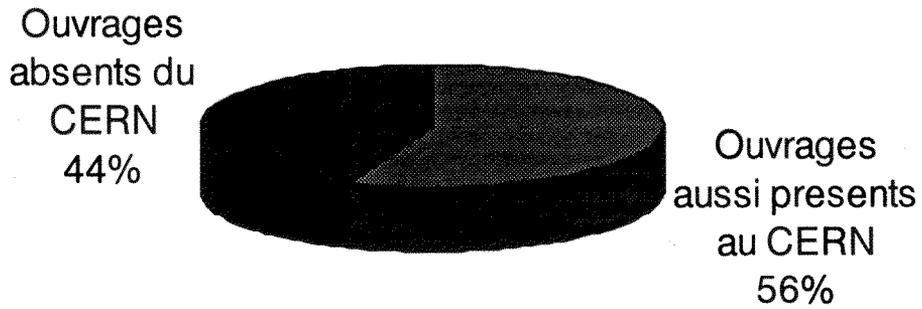
Fonds de Grenoble (CDU=530.1?)



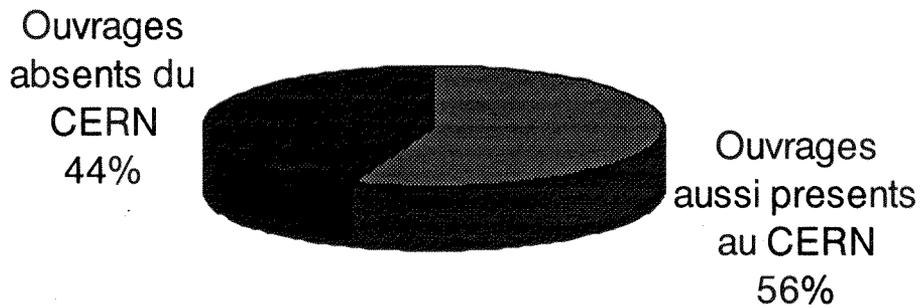
Fonds de Grenoble (CDU=539.1?)



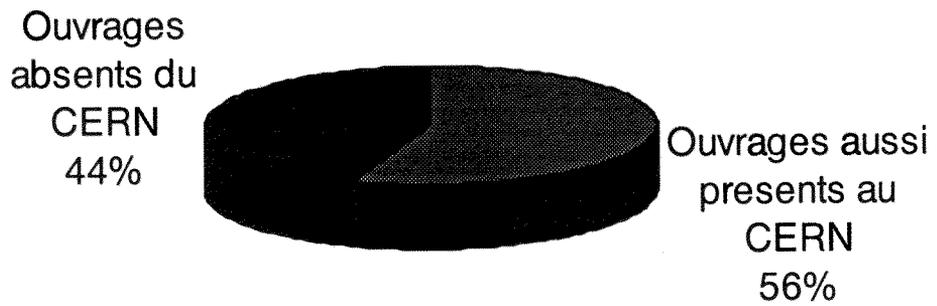
Fonds de Grenoble (CDU=539.2?)



Fonds de Grenoble (CDU=620?)



Fonds de Grenoble (CDU=621?)



Fonds de Grenoble (CDU=678?)

