



HAL
open science

Interfaces de consultation de bibliothèques numériques : exemple du site du CNUM

Jean-Paul Cheung

► **To cite this version:**

Jean-Paul Cheung. Interfaces de consultation de bibliothèques numériques: exemple du site du CNUM. Interface homme-machine [cs.HC]. 2015. dumas-01535555

HAL Id: dumas-01535555

<https://dumas.ccsd.cnrs.fr/dumas-01535555>

Submitted on 9 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

PARIS

MÉMOIRE

Présenté en vue d'obtenir le

DIPLÔME D'INGÉNIEUR C.N.A.M

SPÉCIALITÉ : Informatique

OPTION : Réseaux, Systèmes et Multimédia (IRSM)

par

Jean-Paul Cheung

INTERFACES DE CONSULTATION DE BIBLIOTHÈQUES NUMÉRIQUES :
EXEMPLE DU SITE DU CNUM

Soutenu le 10 juillet 2015

JURY

PRÉSIDENT : M. Eric Gressier-Soudan, Professeur des universités, *C.N.A.M.*
MEMBRES : M. Thierry Cherpitel, Chef de projet, *agence web Eclhydre.*
M. Pierre Cubaud, Professeur des universités, *C.N.A.M.*
Mme Emmanuelle Minault-Richomme, Conservateur
en chef des bibliothèques, *C.N.A.M.*
M. Serge Rosmorduc, Maître de conférences, *C.N.A.M.*

Remerciements

Je tiens à remercier tout d'abord Madame Marie-Annick Cazaux, directeur du Service commun de la documentation du Conservatoire national des Arts et Métiers, Madame Emmanuelle Minault-Richomme, directrice adjointe et conservateur en chef de bibliothèque en charge du Pôle patrimoine, Monsieur le professeur d'informatique Pierre-Henri Cubaud et la précédente directrice de la bibliothèque, Mme Mireille Le Van Ho, qui m'ont fait confiance pour la réalisation de ce mémoire.

J'adresse également mes remerciements à toutes les personnes de la bibliothèque centrale, plus particulièrement à mes collègues du Conservatoire numérique pour leur soutien et leurs encouragements : Caroline Aubert, Françoise Chevalier, Florence Desnoyers, Karine Raczynski. J'ai suivi avec attention les conseils bienveillants des membres de l'équipe du laboratoire CEDRIC, Rodrigo Andrade Botelho de Almeida, Joël Berthelin et Alexandre Topol qui nous a malheureusement quitté en 2013.

Je remercie chaleureusement pour sa relecture assidue Françoise Bercovici, une des mes anciennes collègues, professeur de français du temps où j'enseignais l'informatique. De même, je remercie Judith Ducourtieux pour sa relecture.

Et enfin, je remercie avec affection Valérie qui m'a accompagné avec sagesse.

Je dédie ce mémoire à mes parents et en particulier à ma mère, qui bien que me souhaitant une carrière ecclésiastique, m'a toujours soutenu et encouragé dans mes choix. Je l'imagine rassurée de savoir que je travaille aujourd'hui dans une ancienne abbaye.

La roue à livres

« Celle-ci est une belle & artificielle machine, laquelle est fort utile & commode à toute personne qui se délecte à l'étude, [...] car avec cette sorte de machine un homme peut voir & lire une grande quantité de livres, sans se mouvoir d'un lieu : outre, elle porte avec soi une belle commodité, qui est de tenir & occuper peu de place, au lieu où on la met, comme tout homme d'entendement peut bien comprendre par son dessin. »¹

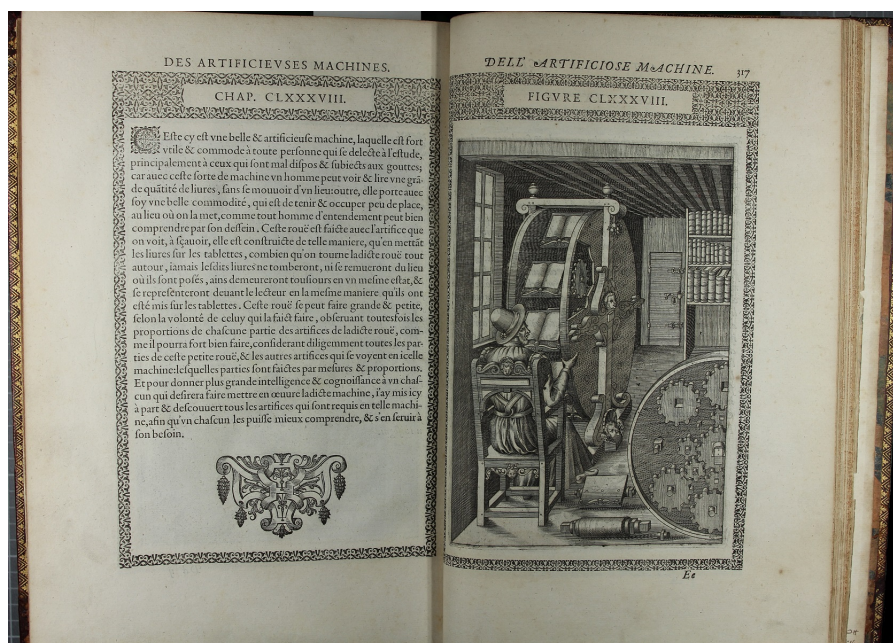


FIGURE 1 – La roue à livres de l'ingénieur Agostino Ramelli

1. Agostino Ramelli capitaine ingénieur militaire pour le roi. Texte p. 316v et planche 188 dans « *Le diverse et artificiose machine* » In Conservatoire numérique des Arts et Métiers [en ligne]. Ramelli, Agostino, 1588, [consulté le 18/10/2014]. Disponible sur <http://cnum.cnam.fr/CGI/fpage.cgi?fDY3/649/100/689/0/0> et <http://cnum.cnam.fr/CGI/fpage.cgi?fDY3/650/100/689/0/0>

Table des matières

1	Introduction	8
1.1	Préambule	8
1.2	La bibliothèque numérique du Cnum	9
1.3	Problématique	10
1.4	Plan du mémoire	13
2	Environnement administratif du Cnum	14
2.1	Présentation des différents partenaires Cnam	14
2.1.1	Le CÉDRIC	14
2.1.2	La bibliothèque centrale du service commun de la documenta- tion du CNAM	15
2.1.3	Le CDHTE	16
2.2	Le service Cnum du pôle patrimoine	17
2.3	Fonctions et missions de l'ingénieur	18
3	Présentation de l'existant du Cnum	20
3.1	Architecture technique interne	20
3.1.1	Architecture matérielle	20
3.1.2	Structure logique	20
3.2	Architecture technique externe	23
3.2.1	Organisation logicielle	24
3.2.2	Constats et solutions	24
4	Analyse des besoins et étude d'usage	27
4.1	Besoins internes des professionnels des bibliothèques	27
4.1.1	Analyse des enquêtes d'usage	29
4.2	Besoins externes de l'utilisateur	31
4.2.1	Public actuel	31

TABLE DES MATIÈRES

4.3	Étude des statistiques d'accès	32
4.3.1	L'analyseur AWStats et logs d'Apache	32
4.3.2	Site multilingue	43
5	État de l'art	45
5.1	Les bibliothèques numériques	45
5.1.1	Introduction	45
5.1.2	Les premières du web	46
5.1.3	Bibliothèque numérique d'initiative privée	47
5.1.4	Bibliothèques numériques institutionnelles	48
5.1.5	Bibliothèques numériques dites de niche	50
5.1.6	Les bibliothèques numériques en tout genre et sites divers	51
5.2	Les logiciels de bibliothèques numériques	53
5.2.1	Tableau comparatif	53
5.3	Les moteurs de recherche	55
5.3.1	Apache Lucene Solr	55
5.3.2	Tableau comparatif	57
5.4	Les interfaces de visualisation	58
6	Refonte du site	61
6.1	Design participatif	61
6.1.1	Parcours de navigation	62
6.2	Mise en œuvre	64
6.2.1	Planning	64
6.2.2	Point de vue informatique	65
6.2.3	Modélisation conceptuelle de la base de données	67
6.2.4	Moteur de recherche Solr	68
6.3	Prototypes	73
6.3.1	Généralités sur l'ensemble des pages	73
6.3.2	Page d'accueil	74
6.3.3	Page catalogue	76
6.3.4	Page synoptique	79
6.3.5	Page feuilleteur	84
6.3.6	Page de recherche avancée	86
6.3.7	Page de résultat	87
6.3.8	Outils de découverte	87

TABLE DES MATIÈRES

6.4	Phase de test	89
6.4.1	Page d'accueil	90
6.4.2	Page catalogue	91
6.4.3	Page synoptique	92
6.4.4	Page feuilleteur	94
6.4.5	Page de recherche avancée	95
6.4.6	Page de résultat	97
6.5	Visualisation d'information - Infovis	97
6.5.1	Visualisation Expérimentale	97
7	Bilan et perspectives	103
8	Conclusion	109
	Bibliographie	111
	Annexes	116
A	Planning de développement	116
B	Capture d'écran de Cnum 1	118
C	Capture d'écran de Cnum 2 Wireframe	128
D	Capture d'écran de Cnum 2	139
D.1	Page d'accueil	139
D.2	Page thématique	141
D.3	Tri dans le catalogue général	143
D.4	Pages synoptiques	149
D.5	Pages du feuilleteur	160
D.6	Recherche avancée	166
D.7	Pages de résultat	168
D.8	Divers	173
D.9	Cnum2.1	179
E	Aide à la recherche en ligne	182
F	Récurrence des mots de la table des matières BSPI.9	186

TABLE DES MATIÈRES

G Occurrence des mots peu courants	188
H Modèle conceptuel de données	190
I Commandes phpMyAdmin de création d'une vue ou table virtuelle	191
Glossaire	194
Index terminologique	195
Table des figures	196
Liste des tableaux	199

Chapitre 1

Introduction

1.1 Préambule

Selon le rapport d'étude mené par la BNF « *Usages des bibliothèques numériques en sciences et techniques* » [MA11, p. 46], « *L'usage quotidien de la bibliothèque comme lieu de travail semble cependant en perte de vitesse.* ». En effet, actuellement la richesse d'une bibliothèque se définit de plus en plus en relation avec le nombre de documents numériques qu'elle offre à ses usagers, que ces documents soient numérisés ou « nativement numériques ».

Le fameux phénomène « Bookless library » décrit dans le livre de [Pol11], explique la disparition progressive des étagères de bibliothèques en raison de la place prise par le livre numérique. Ce phénomène touche aussi les éditeurs de journaux, revues et autres encyclopédies, qui abandonnent progressivement le format papier au profit du numérique. Mais ce modèle économique payant n'est pas le seul : en parallèle s'est développée toute une offre numérique de téléchargement illégal de livres sur Internet. L'étude [CM09, p. 7] citant le rapport Gaymard pour le marché français affirme que « *le piratage de livres concerne un peu plus de 20% des droits d'auteur servis chaque année* ». En effet, nous remarquons que dans le milieu universitaire, les étudiants ne se cantonnent plus aux simples références bibliographiques d'ouvrages citées par leurs professeurs. La plupart auront à leur disposition au prochain cours - si ça ne l'est déjà - le livrel (livre numérique) de cette référence, au pire celui de l'édition précédente.

Ainsi les bibliothèques numériques présentent d'énormes défis et enjeux sociaux, politiques, économiques et culturels. Le livre numérique, qui est au centre de ces enjeux, ouvre de nouvelles perspectives aux lecteurs en les transportant vers un univers qui leur était encore inimaginable il y a quelques années.

Après deux décennies d'existence, les bibliothèques numériques - à différencier des librairies numériques - semblent se répartir entre deux grands types : d'une part

les bibliothèques généralistes qui s'adressent à différents publics : chercheurs, étudiants ou simples amateurs passionnés... Les grands acteurs de ce domaine, qui ont opté pour la numérisation de masse, sont notamment Google Books pour les bibliothèques d'initiative privée ou la Bibliothèque du Congrès et la Bibliothèque nationale de France avec leurs projets respectifs *American Memory* et *Gallica* pour celles d'initiative publique. D'autre part, un autre type de bibliothèque numérique dite « de niche », dont les publics, donc les contenus, sont plus restreints. Le même rapport d'étude cité plus haut insiste sur le fait que les chercheurs apprécient particulièrement ce type de bibliothèques numériques, « *face à la masse attractive des documents en ligne (Google Books), la bibliothèque de niche représente le segment rassurant de l'information organisée dans lequel on sait ce que l'on va trouver et la forme sous laquelle on va l'y trouver; ceci d'autant plus que les chercheurs enquêtés sont, le plus souvent, associés à la sélection documentaire.* », et de plus, elles « *influencent des directions de recherche et renouvellent leurs méthodes d'exploration des contenus* » [MA11, p. 2-3] .

Parmi ces bibliothèques de niche, l'une des plus anciennes est le Conservatoire Numérique (Cnum) bibliothèque numérique du Conservatoire national des Arts et Métiers (Cnam).

1.2 La bibliothèque numérique du Cnum

Le Cnum (<http://cnum.cnam.fr>) est une bibliothèque numérique consacrée à l'histoire des sciences et des techniques qui s'appuie sur les collections du Centre de documentation des arts et métiers et surtout sur le fonds patrimonial de la bibliothèque centrale du Cnam. Ce dernier est principalement constitué par les saisies révolutionnaires des abbés Grégoire et Gruvel portant sur les préoccupations scientifiques de l'époque des Lumières et de l'Encyclopédie.

Lancé en 1998 et inauguré en 2000 [CD03], le Cnum offre gratuitement à la consultation sur son site (cf. infra chap. 3) plus de 2000 ouvrages numérisés. Au fil des années, après une première numérisation consacrée à l'électricité et le magnétisme au XVIII^e siècle, il s'est enrichi de nouveaux corpus comprenant des ouvrages à la typologie différente : périodiques, livres illustrés, manuscrits ou de nouvelles thématiques. La masse grandissante de documents, l'évolution des technologies web et l'apparition des nouveaux dispositifs de lecture associée au développement de nouveaux usages ont renforcé la nécessité de mener à bien le projet de refonte de son interface de consultation.

En prenant en compte ces différentes contraintes, nous analyserons dans ce mémoire la problématique de la construction d'interfaces de bibliothèques numériques, en nous efforçant de proposer un nouvel environnement stable et pérenne.

En vue d'obtenir le diplôme d'ingénieur du CNAM en informatique, ce mémoire

présente le travail de la refonte du site du Cnum.

1.3 Problématique

Les bibliothèques numériques sont à la frontière des deux grandes disciplines que sont l'informatique et la bibliothéconomie. Leur conception n'est pas l'apanage de l'une ou de l'autre mais des deux. En effet, un ouvrage numérisé n'est pas qu'un simple *fac simile* d'un ouvrage papier. Les métadonnées qui lui sont associées enrichissent son utilisation par rapport à celle de l'original.

Dans ce contexte, les interfaces de visualisation doivent permettre une interaction homme-machine globale et fluide, tout en mettant en valeur, aussi bien la richesse des informations contenues dans ces métadonnées que l'apparence sensible du document.

Dans le champ spécifique de l'informatique, les bibliothèques numériques sont à la croisée de plusieurs domaines. Nous en discernons trois à savoir : la recherche d'informations, les bases de données, le world wide web. A ces domaines bien définis s'ajoutent des phénomènes plus difficiles à quantifier relevant de la psychologie, des sciences cognitives ou de la culture. Nous avons représenté ces inter-dépendances dans le schéma ci-dessous :

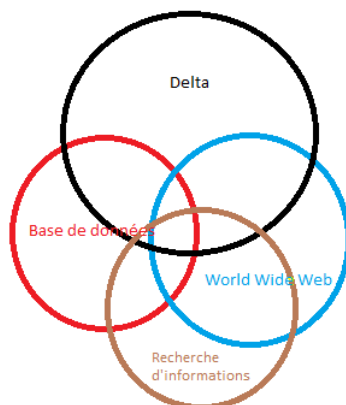


FIGURE 1.1 – Domaines de l'informatique des bibliothèques numériques

A la complexité de ce sujet d'étude s'ajoute la difficulté d'appréhender des connaissances et des pratiques en perpétuelle évolution. Ainsi, depuis le démarrage du projet jusqu'à la soutenance du mémoire, de très nombreuses évolutions techniques sont apparues. De même, un état de l'art exhaustif est difficile à garantir.

Ainsi, la refonte du site du web du Cnum devait prendre en compte plusieurs

1.3. PROBLÉMATIQUE

paramètres :

Réingénierie. La refonte du site va de pair avec une étude préalable de l'existant. La remise à plat du système d'information, du modèle conceptuel de la base de données, de l'architecture du site et des développements informatiques revêtent une importance capitale dans notre réflexion.

Evolution de l'infrastructure matérielle. Les contraintes techniques du début du Cnum en 1998, en particulier, pour le web un débit réseau assez faible, et pour les serveurs une faible puissance de calcul, obligeaient à limiter le traitement des pages html (pages statiques) et le poids des images (binarisées) pour en accélérer le transfert. Ces contraintes sont en parties levées par les évolutions techniques : augmentation de la vitesse de traitement du processeur et des capacités de stockage, amélioration du débit.

Qualité des données et métadonnées. Les métadonnées ont un impact sur la visualisation, l'architecture, la navigation interne et le positionnement des pages dans les résultats des moteurs de recherche. Ainsi, la correction et l'enrichissement des données sont fondamentales pour offrir une interface satisfaisante. Ce travail important ne peut pas être effectué par des machines mais nécessite l'intervention de professionnels de l'information.

Variété des dispositifs et des interfaces de consultation. Le site du Cnum, développé en 1998, n'était pas placé à l'origine sous la contrainte des variétés de supports et d'interfaces. Depuis lors se sont développés de nouveaux modes de lecture, particulièrement l'Internet mobile. Ainsi, les tablettes tactiles et les téléphones portables posent des problèmes de visualisation et d'interaction. Actuellement, un usager s'attend à pouvoir consulter avec le même confort les sites web sur un éventail de type d'appareils : moniteurs d'ordinateur, smartphones, tablettes, écran de télévision, etc.

Les nouvelles perspectives de l'information visualisation (InfoVis). Ce domaine est d'après [Vui10, chap. 2.1.2] « *la représentation visuelle de données organisées et principalement abstraites, afin d'optimiser les capacités visuelles et cognitives de l'humain* ».

Il n'existe à l'heure actuelle aucune méthode de visualisation universelle compatible avec toutes les interfaces. D'autre part, de nombreuses questions restent encore non résolues, telles que la vision globale, la découverte accidentelle ou sérendipité. Comment accéder au document par ces nouvelles fonctionnalités ? Comment permettre la localisation rapide d'information en exploitant des indices visuels et spatiaux ? Comment mettre en regard les deux entités image et texte, comme par exemple les

1.3. PROBLÉMATIQUE

fac-similés et leurs vignettes (entité image), et les tables des matières et leur nuage de mots clés (entité texte)? Comment offrir un autre type de visualisation que celle des vignettes dans le but d’appréhender rapidement un livre? De même, comment approcher un lot massif et hétérogène d’images? Comment proposer une visualisation qui compenserait l’absence des actes suivants : ouvrir un livre, le feuilleter, le remettre en place et en prendre un autre...?

Lecture d’une collection et numérisation. Dans notre cas, la visualisation est plus que la simple lecture cursive d’un texte. On ne regarde pas un graphique et une illustration de la même façon qu’on lit un texte ou une table des matières. L’appréhension d’un document virtuel, et à plus forte raison d’une collection est fondamentalement différente de celle d’un document réel et d’une bibliothèque physique. D’autant plus que les documents du Cnum sont de nature très variée. Par ailleurs, l’auteur [Roe13, p. 16] pointe la mauvaise qualité des images proposées dans plusieurs bibliothèques numériques et la gêne qui en résulte pour l’usager. Une amélioration de la qualité numérique des images est souvent nécessaire avant d’en proposer la visualisation quelle que soit l’interface.

Recherche d’informations interne et référencement par les moteurs externes. En interne, le moteur de recherche est l’outil principal pour accéder à une page précise d’un ouvrage. Sans celui-ci, la navigation devient très vite fastidieuse. En externe, l’optimisation pour les moteurs de recherche ou SEO en anglais pour « Search engine optimization » permettent aux robots d’indexation de bien comprendre le contenu d’une page web, afin de la classer parmi les premiers résultats de la recherche.

Par ailleurs, comment consulter et interroger le Cnum via d’autres interfaces de consultation, d’autres sites? Et plus généralement, comment en assurer l’interopérabilité?

Web design. La conception d’un site web demande des compétences variées dans le domaine du développement informatique, de l’interaction homme-machine, du graphisme... Nous devons en parallèle nous conformer à plusieurs critères : charte graphique de l’établissement, normes du W3C, accessibilité (Référentiel Général d’Accessibilité pour les Administrations¹)... Le graphisme ne relevant pas de notre compétence, il nous a semblé utile de faire appel à un prestataire extérieur.

Ainsi, la refonte du site web s’inscrit dans un contexte en pleine évolution et doit prendre en compte de nombreux paramètres.

1. <https://references.modernisation.gouv.fr/rgaa-accessibilite>

1.4 Plan du mémoire

Compte tenu des différentes caractéristiques de ce sujet, il nous est apparu opportun d'articuler ce mémoire en trois grandes parties :

- tout d'abord l'analyse de l'existant et du contexte institutionnel du Cnum
- ensuite, l'étude des pratiques des usagers et l'évolution des bibliothèques numérique à partir de l'état de l'art
- enfin la présentation de la refonte du site.

Chapitre 2

Environnement administratif du Cnum

Ce chapitre présente le contexte de réalisation du mémoire. Dans une première partie, nous présenterons les différentes entités du Cnam associées au Cnum, ensuite nous détaillerons le cadre de travail. Le Cnum est intrinsèquement lié à ses différents partenaires, chacun apportant à l'un et l'autre, selon ses ressources : les publications, les compétences et les moyens matériels et financiers.

2.1 Présentation des différents partenaires Cnam

Le Conservatoire national des arts et métiers est un grand établissement public d'enseignement supérieur et de recherche qui a pour missions : la formation continue professionnelle, la recherche technologique et l'innovation ainsi que la diffusion de la culture scientifique et technique.

2.1.1 Le CÉDRIC

Le CEDRIC (Centre d'Études et De Recherche en Informatique du CNAM) fondé en 1988 regroupe l'ensemble des activités de recherche en informatique, mathématiques appliquées et électronique menées au Conservatoire national des arts et métiers.

Les recherches du laboratoire tournent autour de sept axes :

- Systèmes sûrs
- Ingénierie des Systèmes d'Information et de Décision (ISID)
- ElecTronIque, TélécommunIcations et Automatique (LAETITIA)
- Médias Interactifs et Mobilité (MIM)

2.1. PRÉSENTATION DES DIFFÉRENTS PARTENAIRES CNAM

- Méthodes Statistiques de Data Mining et Apprentissage (MSDMA)
- Optimisation Combinatoire (OC)
- Bases de Données Avancées (Vertigo)

Au 1er novembre 2012, le laboratoire compte environ 135 membres dont 70 enseignants-chercheurs permanents et 3 administratifs affectés. Le budget annuel (hors salaires des permanents, allocations et bourses étrangères, CIFRE), est de l'ordre de 800 k€HT dont 15% alloués par le ministère, le reste provenant de contrats ou subventions obtenus par les chercheurs.

Le CEDRIC est reconnu par le ministère de la Recherche, de la Technologie et de l'Enseignement Supérieur comme Equipe d'Accueil (EA 1395 d'abord, actuellement EA 4629) depuis sa création. Sa taille et sa maturité actuelle en font le laboratoire de recherche en informatique le plus important du Cnam et le plus reconnu de la région parisienne.

Le groupe « Interactivité pour lire et jouer » de l'équipe MIM poursuit depuis plusieurs années divers axes de recherche sur les interfaces. Elle a initié un important travail de recherche sur la visualisation d'informations dans les bibliothèques numériques. Cf. les mémoires d'Alexandre Topol [Top01, chap. : Interfaces 3D pour les bibliothèques numériques] et [Top02, chap. : 2.4.1 Les bibliothèques numériques], Jérôme Dupire [Dup06, chap. 2.2], Rodrigo Almeida [dAR09, chap. 6 et 7] et l'HDR de Pierre Cubaud [Cub06].

M. Pierre Cubaud, élu depuis le mois de mai 2014 directeur du laboratoire Cédric est un des fondateurs du Cnum. Il est intervenu tout au long de ce vaste projet comme consultant, et référent technique et bibliothéconomique. Il a fourni les ressources humaines informatiques nécessaires à son maintien quotidien.

Le département informatique¹ du Cnam, structure parallèle au Cédric par le biais de ses ingénieurs systèmes, joue également un rôle essentiel en tant qu'administrateur de l'architecture matérielle et logicielle du serveur Cnum (voir plus de détails dans le chap 3.1.1).

2.1.2 La bibliothèque centrale du service commun de la documentation du CNAM

Bibliothèque de grand établissement rattachée au Ministère de l'enseignement supérieur, située au centre de Paris, la bibliothèque centrale du CNAM est une bibliothèque d'étude, d'enseignement et de recherche destinée aux enseignements du CNAM et réservée prioritairement aux enseignants, chercheurs et élèves de l'éta-

1. <http://deptinfo.cnam.fr>

2.1. PRÉSENTATION DES DIFFÉRENTS PARTENAIRES CNAM

blissement. Elle accueille en priorité les 30 000 auditeurs du centre d'enseignement parisien, mais a vocation à proposer certaines ressources électroniques, pour l'ensemble des bibliothèques du réseau CNAM en France et à l'étranger. Le fonds de la bibliothèque est riche de 185 000 titres dans les domaines des sciences et des techniques, dont un fonds ancien sur l'histoire des techniques constitué de quelque 8 000 volumes (XV^e-XVIII^e siècle), 23 000 mémoires d'ingénieurs et thèses soutenus au CNAM, 3 900 revues papier dont 500 titres vivants, 21 000 revues électroniques en sciences, techniques et économie, 21 bases de données en ligne. Elle numérise une partie de ses fonds anciens, portant sur l'histoire des techniques du XVI^e au XX^e siècle pour le diffuser au travers du Conservatoire numérique.

La réserve contient 6 incunables, 600 ouvrages du XVI^e siècle, près de 2000 du XVII^e et plus de 4000 du XVIII^e.

2.1.3 Le CDHTE

Le CDHTE (Centre d'histoire des techniques et de l'environnement) est un laboratoire de recherche en histoire des sciences et des techniques fondé en 1960. Il est centre commun au Conservatoire national des arts et métiers (Cnam) et à l'École des Hautes Etudes en Sciences Sociales (EHESS).

Ses missions sont :

- la documentation en histoire des techniques et en environnement
- la recherche fondamentale et appliquée en histoire des techniques et du patrimoine industriel
- l'enseignement par la recherche

Les domaines de recherche du CDHTE sont orientés autour de 6 axes :

- Histoire de l'Environnement
- Histoire de l'Innovation et des Savoirs Techniques
- Muséologie des techniques
- Techniques, Territoire et Architecture
- Patrimoine Documentaire
- Techniques et Vie Quotidienne

Le CDHTE est une Equipe Accueil de Doctorants (EAD) du Ministère de l'Éducation Nationale, de la recherche et de la technologie qui publie régulièrement dans la revue « Documents pour l'histoire des techniques »² (DHT). Il travaille en partenariat avec le Conservatoire Numérique (CNUM) et le Musée des arts et métiers en

2. <http://dht.revues.org/>

particulier sur la sélection des ouvrages à numériser.

La chaire d'histoire des techniques du Cnam, au sein du Pôle Economie et gestion, offre des cours diversifiés en histoire des techniques réalisés par des enseignants-chercheurs compétents et motivés.

Depuis septembre 2012, le CDHTE a successivement pris le nom de HTTP (Histoire, technique, technologie et patrimoine) puis HT2S (Histoire des technosciences en société).

2.2 Le service Cnum du pôle patrimoine

Le Cnum est un service du pôle patrimoine du Service commun de la documentation. Il est né d'un projet commun de la Bibliothèque centrale, du Centre d'histoire des techniques et de l'environnement et du Centre d'étude et de recherche en informatique et communications (CEDRIC) du CNAM. Depuis 2005, le Cnam et la BnF³ ont signé une « *convention de coopération documentaire dans le cadre du réseau des pôles associés* » pour la numérisation des expositions universelles. Le musée des arts et métiers⁴, par l'intermédiaire du centre de documentation fait aussi partie intégrante de la politique de numérisation.

Le budget numérisation annuel est en hausse d'environ 20% depuis 2009 à 2011 et une partie est prise en charge par la BNF et le programme pluriformation du MESR.

L'équipe du Cnum est composée d'un conservateur en chef des bibliothèques, de trois bibliothécaires (dont une travaillant au centre de documentation du musée), d'un professeur et d'un ingénieur d'études en informatique. Chacune de ces personnes assure également en parallèle d'autres travaux et tâches administratives en dehors du Cnum.

La gestion informatique du Cnum a été successivement assurée à temps partiel par M. Pierre Cubaud puis par deux doctorants avec un sujet de thèse ad hoc (cf. page 14). J'ai repris à temps plein la suite de cette activité à partir du 15 avril 2009 jusqu'à aujourd'hui. Chacun d'entre nous a apporté sa pierre à l'édifice à partir des demandes des usagers, avec de nouveaux services sur l'application de « front » et de « back office ».

Voici sommairement indiqué dans un tableau récapitulatif les travaux et le nombre de titre mis en ligne par chacun durant toute cette période :

3. Bibliothèque nationale de France www.bnf.fr

4. <http://www.arts-et-metiers.net>

2.3. FONCTIONS ET MISSIONS DE L'INGÉNIEUR

<u>Ingénieur du Cnum</u>	<u>Période</u>	<u>Nouveaux services</u>	<u>Nombre de titre</u>
Pierre Cubaud	1998-2002	système d'information, modélisations et développements de Cnum 1	86
Jérôme Dupire	2002-2006	base de données, coordination passerelle OAI	493
Rodrigo Almeida	2006-2009	back office et contrôle qualité, visualisation	311
Jean-Paul Cheung	2009-2015	cnum 2	1261

TABLE 2.1 – Nom des ingénieurs du Cnum

2.3 Fonctions et missions de l'ingénieur

J'ai été recruté en avril 2009 sur un poste d'ingénieur d'études en développement et déploiement d'applications à la bibliothèque centrale, à la suite de l'obtention de mon Master 2 de recherche au CNAM en Conception d'Applications Multimédia.

Mes missions concernent la gestion et la maintenance du site de production (interface Cnumdev) et de diffusion de la bibliothèque numérique (interface Cnum), hébergé sur un système GNU/Linux. Outre l'administration courante de la base de données (MySQL), à chaque niveau de la chaîne de traitement, je mets en œuvre des processus qui optimisent le déroulement du workflow : le contrôle qualité, les échanges de notices via la passerelle OAI-PMH⁵, l'analyse de statistiques. J'assure le traitement des lots à leur retour de numérisation : contrôle des images et des données textes via des interfaces de visualisation et des systèmes de correction développés en langage de script (PHP, Perl, Shell Unix), à adapter selon les lots. L'ensemble des sources du projet est sous CVS et je mets régulièrement à jour pour l'équipe des bibliothécaires la documentation utilisateur et développeur sur un wiki libre. Dans le cadre du partenariat « Pôle associé numérisation » avec la Bibliothèque nationale de France (BnF), j'ai finalisé une application d'échanges de métadonnées s'appuyant sur le protocole OAI-PMH développée en Java et Php/MySQL.

Je travaille également en étroite collaboration avec les ingénieurs systèmes du laboratoire Cédric chargés de la gestion applicative et matérielle du serveur.

En relation avec les prestataires de numérisation, je veille au bon suivi du traitement des lots. Je collabore avec les bibliothécaire pour l'élaboration du cahier des charges de numérisation. Mon rôle est celui de référent et d'encadrant technique.

Dans le même temps, je travaille sur d'autres projets :

5. http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html

2.3. FONCTIONS ET MISSIONS DE L'INGÉNIEUR

- en collaboration avec les bibliothécaires et la direction des systèmes d'information, nous avons mis en place un extranet qui s'intégrera dans le futur guichet unique (ENT), un des projets phare du schéma directeur numérique de l'établissement (SDN). Celui-ci propose la consultation des annales (environ une cinquantaine de sujets par semestre) et d'une partie des mémoires d'ingénieurs (« communicables non diffusables » et « non communicables ») dont le chiffre s'élève à 800 par an dont 105 DPE (voir l'interface <http://memoires.cnam.fr/ann.php?id=2012.ANN.SITI.RSX.07.pdf> qui a été basculée sur <https://edocuments-scd.cnam.fr/ann.php?id=1.2012.ANN.SITI.RSX.07.pdf>). Dans ce projet, j'ai défini tout le workflow de traitement de ces documents numériques.
- en parallèle, j'ai également développé en langage Php sur une période de six mois avec les bibliothécaires et un stagiaire [Ron11] une exposition virtuelle portant sur la thématique des « Expositions universelles » accessible à cette adresse http://cnum.cnam.fr/expo_virtuelle.
- membre du jury pour l'étude des dossiers des prestataires pendant deux semaines du projet CNAM-Rémut (le réseau des musées et collections techniques de France⁶)
- j'ai suivi le démarrage du projet Moodle du Cnam, servant à réaliser un ensemble d'outil numérique d'aide au tutorat de mémoire de fin d'étude

6. <http://www.remut.fr>

Chapitre 3

Présentation de l'existant du Cnum

Nous nous attelons dans ce chapitre à présenter l'architecture interne et externe du Cnum du point de vue de la structure physique et logique. Selon les besoins requis, cette description déterminera notre marge de manœuvre sur l'extension du projet de refonte du site.

3.1 Architecture technique interne

3.1.1 Architecture matérielle

Le Cédric héberge dans ses locaux le serveur du Cnum, géré sous la responsabilité du directeur technique et ingénieur de recherche en informatique M. Joël Berthelin. Il est assisté par deux ingénieurs systèmes et deux techniciens.

Open Suse Linux est le système d'exploitation libre installé sur ce serveur. Les disques sont montés en RAID 5 avec 2 quadri-processeurs de 2.83Ghz, la capacité totale du disque est d'environ 4 téraoctets (dédiée en grande partie au stockage des images en haute définition), il a été porté à 8 téraoctets en avril 2012. A partir du mois de mars 2013, le Cnum s'est doté d'un nouveau serveur d'une capacité mémoire de 11To et d'un CPU avec 6 cœurs tournant à 3Ghz.

3.1.2 Structure logique

3.1.2.1 Typologie des documents et regroupement

Les documents numérisés peuvent être de nature différente : texte, illustrations (planches, figures) en noir et blanc ou en couleur, des photographies, des planches simples, dépliantes, doubles, parfois de grand format. Il existe aussi des livres à systèmes, c'est à dire avec des parties mobiles comme des volvelles, des pages contenant des tissus et des planches en braille.

3.1. ARCHITECTURE TECHNIQUE INTERNE

Ces documents sont regroupés selon trois types : les périodiques ou revues et leurs volumes, les grandes monographies (en plusieurs volumes également), et les monographies ou ouvrages. Ces derniers contiennent parfois plusieurs tomes et sont appelés dans notre jargon interne les « quanta virtuels ». Dans ce type de regroupement particulier, seule la notice mère de ces tomes dispose d'une description bibliographique, mais c'est tout l'ensemble qui est sauvegardé comme un unique enregistrement dans une table de la base de données (cf. chap. 6.2.3).

3.1.2.2 Système d'information

L'application de « Back office » du Cnum appelé « Cnumdev » représente la partie non visible au public d'une copie du site. Il est accessible uniquement à notre équipe dans l'enceinte du Cnam. Avant la mise en ligne des documents numérisés, nous effectuons par l'interface « Cnumdev » un enrichissement et des corrections bibliographiques au niveau des notices, des tables de matières et des planches, ainsi qu'un contrôle qualité de l'indexation et de la numérisation.

A la genèse du Cnum, les informations bibliographiques de chaque titre étaient sauvegardées dans des fichiers texte avec les noms d'extension « des » (revues et grandes monographies), « deq » (volumes ou tomes) et « deo » (monographies). Elles ont été basculées à partir de 2002 dans une base de données MySQL (cf. le chapitre 6.2.3) suite à la création de l'entrepôt OAI-PMH.

Afin d'éviter le travail de « multi catalogage », nous nous sommes posés la question de centraliser les informations bibliographiques du Cnum dans le SIGB de la bibliothèque centrale. Nous pourrions par ce biais récupérer ces données par un webservice ou via un export, puis une conversion au format sql et une intégration dans la base de données du Cnum. Des requêtes via le protocole en mode synchrone Z39.50 seraient une des solutions, contrairement au protocole en mode asynchrone OAI-PMH. Or, ce dernier n'étant pas assez riche au niveau des métadonnées ne pourrait donc résoudre notre problème. Ce changement de système d'information n'est pas encore d'actualité et doit faire l'objet d'une étude plus approfondie.

3.1.2.3 Structure des tables

Suivant les lots envoyés en numérisation, les tables des matières et des planches sont selon le cas, saisies en interne ou par le prestataire. Elles sont enregistrées au format texte en ISO-Latin avec les extensions : tdm, tdp, som¹. Chaque ligne de ces tables est construite sur quatre champs séparés par un croisillon : « 0#titre de chapitre#numéro de page ou planche#numéro image incrémenté ». Le premier représente le niveau hiérarchique, le deuxième le titre de chapitre ou la légende, le troisième le numéro de la page ou de la planche, et le dernier le numéro image

1. Une table des sommaires a été envisagée à l'époque

3.1. ARCHITECTURE TECHNIQUE INTERNE

(chiffre incrémenté).

L'interface Cnumdev effectue automatiquement la correspondance entre le numéro de page de la ligne de table et le numéro image du fichier image.

La typologie des tables des matières est variable d'après [ABC09] :

- « *Les tables chronologiques suivent l'ordre des pages et peuvent présenter des constructions hiérarchiques plus ou moins complexes : section, sous-section, chapitre, paragraphe. Autant de segmentations qui organisent la pensée et les propos.*
- *Les tables index proposent des entrées par ordre alphabétique, comme un dictionnaire, qui soulignent ainsi la variété et le grand nombre des termes scientifiques. Si elles offrent un riche vocabulaire, elles ne rendent pas compte de la structure de l'ouvrage.*
- *Les tables des matières des revues suivent l'ordre de parution des numéros toutefois, au fil des années, elles peuvent modifier leur présentation, changer la structure des rubriques. »*

3.1.2.4 Format des fichiers image

Les images sont stockées suivant 4 formats :

- images en haute définition numérisées à 400 dpi en double page pour la conservation (brut de scan)
- images en haute définition pour la diffusion découpées en simple page : il a été prévu de proposer ces images dans l'interface feuilleteur afin de pallier à certains défauts de numérisation des images de diffusion.
- images en noir et blanc ou « binarisées » (simple page) stockées au format TIFF et compressées suivant la norme CCITT groupe 4
- images pour la diffusion en simple page : elles ont été dégradées au format gif à 96 DPI et 8 niveaux de gris [CT99] et [Cub06, p. 24] à partir des images « binarisées ». Ce choix de format s'explique par une meilleure lisibilité des lettres sur des textes à contraste négatif de ceux à contraste positif [SN02].

3.1.2.5 Codage et développement

Nous présentons ici la liste des principaux scripts développés en Perl et shell Unix :

- génération des pages statiques du site
- génération des images de diffusion web au format gifs
- génération des vignettes
- génération des documents pdf
- intégration des notices dans la base de données en mode insertion et modification
- scripts de « cnumdev »

3.2. ARCHITECTURE TECHNIQUE EXTERNE

- scripts de consultation des pages html du site
- scripts de recherche

L'ensemble représente environ 301 454 caractères de codage soit environ 9 298 lignes.

3.2 Architecture technique externe

La navigation dans le corpus s'effectue au travers du site de diffusion, encore appelé application de « front office », accessible à l'adresse <http://cnum.cnam.fr>. Elle représente l'interface de consultation du grand public. L'accès aux fac-similés se fait soit par le catalogue général² avec la liste des titres qui rebondissent sur les tables, soit par le moteur de recherche. Ce dernier, développé en langage Perl a été spécifiquement mis en place par des membres de l'équipe ILJ du laboratoire Cédric. L'interrogation en mode texte s'effectue sur les notices bibliographiques, les tables des matières et des illustrations qui sont les véritables points d'entrée des documents. L'interface permet de combiner les critères de recherche : possibilité de choisir le type de document (monographie ou périodique), les types de table (matière ou illustration) pour rechercher un ou plusieurs mots, une expression ou une partie seulement d'un mot.

Durant la période allant de juillet à septembre 2005, la bibliothèque centrale a engagé M. Niels Henner³ pour réaliser un moissonneur et un entrepôt OAI-PMH⁴. Le système est basé sur une solution open source écrit en JAVA⁵ et PHP⁶. A l'heure actuelle, seules les notices bibliographiques sont moissonnables : auteur, titre, sujet.

Mais depuis janvier 2009, le Cnum moissonne et intègre les métadonnées de Gallica. Par la suite, l'équipe travaillera à l'identification et au repérage précis des catégories du catalogue Gallica qui correspondent aux thématiques du Cnum. Ce filtrage permettra d'affiner les notices importées pour que les résultats de recherche correspondent davantage aux besoins des lecteurs, dont les modes d'interrogation ont fait l'objet d'une première étude statistiques en 2008 par [Cla08]. De même, [Ism08] a étudié la consultation des résultats de ce moissonnage par différents systèmes de visualisation.

A partir de 2010, les bibliothécaires ont souhaité étendre l'entrepôt OAI des notices aux tables des matières, illustrations et index. Mais ceci n'est réalisable qu'en

2. Il existe aussi un autre accès via le catalogue de la Bibliothèque centrale

3. <http://ca.viadeo.com/fr/profile/nils.henner> [référéncedu24/07/2012]

4. Open Archives Initiative's Protocol for Metadata Harvesting

5. oclc.org/research/activities/past/orprojects/harvester2/harvester2.htm[référéncedu24/07/2012]

6. physnet.uni-oldenburg.de/oai [référéncedu24/07/2012]

reformalisant le modèle conceptuel de la base de données (cf. 6.2.3).

3.2.1 Organisation logicielle

3.2.2 Constats et solutions

Page catalogue. Certains utilisateurs ont remis en cause l'utilité de la page catalogue. Or, il serait pertinent de proposer d'autres modes de parcours pour accéder au document. En effet, la plupart des bibliothèques ne proposent pas ou plus d'accès par leur catalogue, préférant les résultats de leurs moteurs de recherche. Dans notre cas, l'analyse des logs de consultation du catalogue du Cnum a démontré l'importance capitale de cette page. En moyenne, deux pages de recherche (34172) sont consultées contre une du catalogue général (17369). voir le tableau 4.11

Scories dans les données. Les informations de la base de données contiennent des scories comme certains indices servant de point de repère (signe arobase et barre oblique) dans le champ titre « TI » destiné au titre court « TC ». Ce dernier a été spécialement créé à l'époque pour un affichage personnalisé des informations. En effet, nous nous sommes rendu compte par la suite que le titre court pouvait être légèrement différent du début du titre « long »⁷ (Champ « TI » contenant le signe de séparation « @ »). Mais dans notre projet de conception de la nouvelle interface Cnum, ce titre court demande surtout à être retraité au niveau du nombre de caractères maximum.

En résumé, le champ « TI » est dépendant des normes de catalogage bibliothéconomique, et le champ « TC » a été créé pour s'adapter à l'interface et à l'interaction homme machine.

Suite à ces constats, les bibliothécaires en ont profité pour effectuer un important travail de corrections et d'enrichissement de plusieurs champs de la notice. Nous avons par là même développé des scripts spécifiques d'intégration de ces corrections, et aussi dans le but de corriger automatiquement les sauts de lignes et les apostrophes curvilignes générées par le logiciel de traitement de texte Word.

Modèle conceptuel des tables. Il existe actuellement une problématique importante des entrées multiples dans les tables de matières⁸. En effet, pour éviter la répétition des titres de chapitres, nous avons pris l'habitude d'insérer le terme « Idem »⁹. Celui-ci a du sens sur un document physique mais plus dans le cas du virtuel. Du côté de l'affichage, la non répétition du mot apporte un certain confort

7. ex. : <http://cnum.cnam.fr/redir?8CA121-1>, <http://cnum.cnam.fr/redir?8XAE376>,
<http://cnum.cnam.fr/redir?8KU54>

8. ex. : <http://cnum.cnam.fr/CGI/fpage.cgi?4KY28.99/636/100/636/0/0>

9. ex. : <http://cnum.cnam.fr/CGI/redird.cgi?FOLXAE10.1>

3.2. ARCHITECTURE TECHNIQUE EXTERNE

de lecture mais éclipse une partie des résultats des moteurs de recherche et en relèguent son positionnement. Le problème est identique avec les tirets indiquant un même mot répété inutilement dans une énumération¹⁰. Par là même, nous pourrions aussi redéfinir le modèle conceptuel des tables, en répétant uniquement les numéros de pages au lieu des titres de chapitres.

Dans ce contexte, un utilisateur nous a fait remarquer que les tables ne lui semblaient pas bien structurées hiérarchiquement. Il a souhaité que les chercheurs en sciences et techniques du Cnam - maîtrisant bien le corpus - soient sollicités dans la redéfinition de cette structuration.

Moteur de recherche. Les temps de réponse du moteur de recherche interne se sont allongés dramatiquement. D'après nos tests, le problème ne semble pas provenir de la virtualisation de l'architecture, ni d'un conflit de mise à jour de la bibliothèque Perl, ni même de la croissance du corpus. En effet, ces tests ont été menés sur un autre serveur avec un volume de notices et de tables identiques et nous avons obtenu le même résultat. Sous le shell Unix, le temps de recherche est de moins de 1 milliseconde sur la ligne de commande « *time grep oiseau /home/cnum2/httpd/htdocs/tables/*.tdm* », alors qu'il prend environ 8 secondes via l'interface de recherche du Cnum.

Par ailleurs, un autre point faible du moteur est qu'il n'offre aucun traitement linguistique comme la sensibilité à la casse, la lemmatisation, la flexion, la racinisation, etc. Cette contrainte pénalisant l'internaute, ce dernier doit reformuler en permanence ses requêtes sous de multiples variantes.

Nous remarquons que les critères de recherche, « Début du mot » et « Fin du mot » dans le moteur du Cnum actuel, présente comme utilité d'extraire tous les titres dont les mots commencent ou finissent d'une certaine manière. Par exemple, les mots se terminant par « cycles » peuvent être tricycles, motocycles... Le groupe de travail trouve ce critère pertinent, et souhaite le remplacer par des jokers comme « * », « ? », ... Ce type de requête nous oblige aussi à mettre en place de systèmes de sécurité, pour éviter un problème de déni de service ou que l'on aspire tous les résultats de la recherche.

IHM. Au niveau de l'interaction homme machine, nous pouvons améliorer l'interface gérant les cas d'erreurs liés au CGI ou aux pages inexistantes. Nous présenterons une page comportant le bandeau principal supérieur, accompagné d'un message d'erreur et une proposition de rebond sur d'autres pages (voir ici un exemple de cas d'erreur : http://cnum.cnam.fr/CGI/redir.cgi?identifiant_inexistant cf. Annexe D.28)

Nous remarquons aussi que la plupart des boutons de navigation sont totalement dénués d'interactivité ou d'infobulles. Cependant, d'après un utilisateur, la dernière fonctionnalité manquante faisait partie d'une demande volontaire des bibliothécaires.

10. ex. <http://cnum.cnam.fr/CGI/fpage.cgi?8KE304.6/416/100/426/62/84>

3.2. ARCHITECTURE TECHNIQUE EXTERNE

SEO. Dans la problématique du référencement par les moteurs de recherche, plusieurs éléments du site posent actuellement problème. Par exemple, les cadres ou « frames » et l'absence de feuille de style « css » ne sont pas recommandés par le W3C, car ces implémentations rallongent le temps d'indexation et en réduisent l'efficacité. De plus, nous avons remarqué que les liens résultats de ces moteurs renvoient parfois vers des pages sans élément cadre.

Les liens hypertextes renvoyant vers la page numérisée ont été intégrés aux numéros de pages au lieu du titre de chapitre. Nous supposons que cela pourrait nuire au classement des résultats et que la meilleure méthode serait de l'intégrer au titre de chapitre. De plus, la zone interactive d'un numéro de page étant moins étendue que celle d'un titre de chapitre, demande une plus grande précision de clic avec le pointeur de la souris.

La recherche dans le moteur Google avec les mots « *Traité des aimans artificiels ; contenant une méthode courte et aisée* » montre que le lien <http://cnum.cnam.fr/redirect?12SAR12> n'apparaît pas en première page. Il semble que le classement change lorsqu'une partie du titre de l'ouvrage est mise en caractère gras, à moins que le problème ne vienne du cadre « frame ».

Développement statique ou dynamique. Au début du développement du Cnum, la puissance de calcul des ordinateurs ne favorisait pas le développement de sites dynamiques, ce qui explique son développement en version statique. Nous avons mené un débat sur la poursuite de ce type de développement ou bien le passage à une version complètement dynamique. Il s'avère que les versions en plusieurs langues du site pourraient poser problème de par son développement entièrement statique. Ainsi, il faudrait créer autant d'instances de pages statiques que de langues, et de plus, les adresses pérennes ne seraient plus identiques. L'autre solution serait d'incorporer toutes les traductions possibles dans la page avec un basculement en Javascript.

Chapitre 4

Analyse des besoins et étude d'usage

Nous traiterons dans un premier temps de l'analyse des besoins et nous terminerons par une étude des statistiques d'accès.

4.1 Besoins internes des professionnels des bibliothèques

Restructuration du corpus et accès par un catalogue. Le corpus du Cnum grandissant, un besoin urgent s'est fait sentir de le structurer de manière plus réduite. La classification par thématique et typologie s'est donc imposée naturellement. Encore faut-il bien déterminer les différentes thématiques pour être le plus exhaustif possible, et prévoir celles qui pourraient émerger à l'avenir.

Une première solution envisagée a été de diviser encore plus finement les thématiques en sous-thématiques. Or, cette structure pourrait poser des problèmes d'accès au document, contredisant la règle des 3 clics [NBL11, chap. 5.1.2]. L'utilisateur peut ressentir une frustration et quitter le site s'il n'obtient pas rapidement l'information qu'il recherche. Mais cette règle n'est plus tout à fait vérifiée et elle est remise en cause dans cet article [Por03]. Évidemment, ce problème ne se pose pas si l'on passe par les résultats du moteur de recherche. Ici, nous inscrivons notre raisonnement uniquement dans une navigation par le catalogue.

Finalement, le corpus a été divisé en neuf thématiques : Catalogues de Constructeurs, Construction, Énergie, Expositions universelles, Généralités et vulgarisation scientifique, Histoire du Cnam, Machines et instrumentation, Technologies de l'information, Transports ; trois types de documents : Revues, ouvrages (regroupant les grandes et petites monographies) et manuscrits et dessins (voir le chapitre 6.2.3.1). Nous avons préféré respectivement les termes « revues » et « ouvrages » à « monographie » et « périodiques » car plus cohérents sémantiquement dans cet ensemble, et aussi plus accessibles à un public de non chercheurs. Nous veillons ici à satisfaire tous les profils d'utilisateurs.

Liste des fonctionnalités. Le groupe de travail a souhaité la mise en place de plusieurs fonctionnalités. Nous les regroupons ici sous forme d'une liste d'item :

- tri par lieu d'édition dans le catalogue (demande des chercheurs)
- exportation de la notice bibliographique sous différents formats : Zotero, BibTeX, Unimarc, ePub
- exportation et impression des résultats de recherche dans un document pdf
- affichage d'une page au « hasard » sur la page d'accueil
- enregistrement par les internautes de leurs paramètres de navigation par un système d'authentification dans la base de données
- passage de la dernière page d'un volume à la première du suivant dans le feuilleteur. Ce problème est d'autant plus critique pour les ouvrages constitués en « quanta virtuels ».
- mécanisme de partage et d'abonnement sur les réseaux sociaux. Le web 2.0 a vu l'émergence de ces nouveaux types de médias du web, il était important de les exploiter.
- un bouton permettant la rotation des images dans le visualiseur, dans le sens horaire et antihoraire, serait utile pour les pages mal orientées ou dont la surface d'affichage est de sens diverse
- visualisation des documents lus au même moment par les autres internautes
- intégration d'un « vrai feuilleteur » permettant de lire, tourner et faire défiler les pages comme dans le monde réel (mécanisme efficace sur interface tactile)
- navigation avec l'aide d'un fil d'Ariane
- impression d'une partie de l'ouvrage
- lecture exportable du feuilleteur, il peut s'avérer utile pour les personnes souhaitant consulter les documents sur un autre site web, blog ou sites de réseaux sociaux

Varia. Lors de la première réunion de notre groupe de travail, nous avons relevé cette demande d'une chercheuse qui est la récupération dans un document pdf de tous les articles du même auteur. Cette demande légitime des chercheurs faciliterait grandement leurs travaux d'inventaire mais ne pourra pas être satisfaite car le Cnum ne procède pas à un dépouillement à l'article.

Dans un autre registre, une des bibliothécaires nous a expliqué la méthodologie de recherche en texte intégral par l'un des chargés de collection du musée. Ce dernier récupère les références de pagination ou d'extrait du livre dans le moteur de recherche « Google livres », et grâce à ces informations, il retrouve la même page sur le Cnum, car celle de Google est parfois mal numérisée ou seul un court extrait est proposé. En effet, d'après [MA11, p. 2] « *Les fonctionnalités (recherche, visualisation, etc.) offertes par les bibliothèques SCT¹ comptent beaucoup moins que les fonds proposés, même en mode image. Cette bonne acceptation du mode image révèle une défiance vis-à-vis d'un mode texte qui fragmente et dissémine*

1. Sciences et Techniques

4.1. BESOINS INTERNES DES PROFESSIONNELS DES BIBLIOTHÈQUES

les contenus. Les fonctionnalités techniques propres aux bibliothèques numériques sont d'autant moins attendues que le moteur de recherche de Google, presque systématiquement utilisé pour explorer les collections numériques, donne toute satisfaction en matière d'accès aux contenus des textes numérisés ».

Une autre demande concernait la mise en place d'un formulaire de recherche simplifié sur la page d'accueil avec un lien sur la recherche avancée. En effet, les internautes n'appréhendent pas bien le moteur de recherche avancée du Cnum, plutôt destiné à un public de spécialistes en bibliothéconomie. Il se trouve de plus que les usages des moteurs commerciaux actuels influencent énormément les perceptions et les attentes des internautes (voir cette étude de cas sur des étudiants [Gri05]).

Les demandes restantes portaient sur le changement du plan de navigation du site. En effet, certains usagers souhaiteraient, à partir de la page catalogue, un accès direct aux fac-similés sans passer par la page synoptique. Or, la structure actuelle est depuis une dizaine d'années bien ancrée chez les habitués du Cnum, et son changement pourrait quelque peu les perturber.

Dans la foulée, nous avons remis à plat le système d'information et la structure d'arborescence des tables. Faut-il également fusionner les « quanta virtuels » ? Cette appellation propre au Cnum désigne les documents constitués par un regroupement de plusieurs tomes, ce qui représente bibliothéconomiquement un ouvrage compris entre une grande et une petite monographie (voir un exemple dans « *L'art des expériences de Nollet Jean-Antoine* »²). En effet, ce genre de regroupement pose plusieurs problèmes en terme d'affichage, de modèle conceptuel des données, et des nouveaux services que nous souhaitons offrir par l'entrepôt de moissonnage OAI.

Enfin, il apparaît un important problème dans le moteur de recherche. Le temps d'affichage des résultats est devenu de plus en plus long d'après notre étude dans le chapitre 3.2.2. Nous avons résolu indirectement ce problème en renseignant le fichier « sitemap.xml » décrivant les principaux fichiers à analyser par les robots d'indexation.

4.1.1 Analyse des enquêtes d'usage

Nous présentons dans ce chapitre, les demandes des chercheurs en histoire des sciences et techniques, dans l'enquête d'usages [MA11, p. 15] menée dans le cadre du pôle associé avec la BnF. Cette enquête révèle le besoin d'un accès en mode texte intégral, bien qu'il ne fasse pas l'unanimité chez les chercheurs interviewés. En effet, deux sur quatre rejettent le mode OCR par crainte du bruit généré (« *je pense que ça ferait pas mal de bruit. C'est un peu à double tranchant. On a toujours*

2. <http://cnum.cnam.fr/redirect?12C15>

4.1. BESOINS INTERNES DES PROFESSIONNELS DES BIBLIOTHÈQUES

la tentation d'en demander plus, mais ça ne serait pas forcément très sage. »).

De même, un des utilisateurs de notre groupe de travail ne considère pas la mise en place de cette fonctionnalité comme étant une plus-value. A la place, le groupe a proposé une océrisation à la page, lancée par l'internaute grâce à un bouton dans le feuilleter. Or, le besoin de cette reconnaissance de caractères doit être placé en amont et non en aval du processus (voir la procédure utilisée par un chargé de collection du musée au chapitre 4.1.1). Le besoin n'est pas tant de récupérer une partie du document en mode texte mais plutôt de rechercher dans ce mode.

La reconnaissance de caractères sur du livre ancien en sciences et techniques posera des problèmes sur les caractères non latins, les formules chimiques et mathématiques, et les manuscrits... Dans le dernier cas, une piste peut être exploitée via le système de reconnaissance intelligente de caractères, désigné en anglais par l'expression « *character recognition* » ICR [BH99]. Mais d'autres problèmes vont se poser sur le traitement de cette énorme quantité de données, surtout par rapport au temps d'indexation, au classement des résultats et au bruit généré.

La mise en place de cette nouvelle fonctionnalité pour pallier ce besoin de recherche en mode texte, est très dépendante de la qualité obtenue par le logiciel de reconnaissance de caractères. Ainsi, nous avons mené des tests d'océrisation avec le logiciel libre Tesseract[Smi07], mais les résultats ne se sont pas avérés très convaincants. A l'opposé de la version libre, il existe un logiciel commercial nommé ABBYY, géré en ligne de commandes sur Linux³. Cependant, le prix du kit de développement (sdk) est dépendant du nombre de pages à numériser, sachant que le Cnum contient environ 650 000 pages.

La reconnaissance de caractères est un domaine de recherche très actif en informatique et « *Le Cnum attend beaucoup des développements en cours au Centre d'Études Supérieures de la Renaissance de Tours (Agora) ou du projet européen Impact ((IMProving ACcess to Text) entre autres. »* [MA11, p. 17].

Il est tout aussi important d'indiquer pour notre étude que d'après [MA11, p. 3], les chercheurs sont beaucoup moins intéressés par « *les dispositifs techniques d'accès ou d'exploration des documents numériques* » que par les contenus. Ils souhaitent par contre « *encore plus de sources, de tout type, rares ou difficiles à trouver de préférence* ».

3. <http://france.abbyy.com/Default.aspx?DN=26dcdcca-da97-425b-83ae-91f1dfe13406>

4.2 Besoins externes de l'utilisateur

4.2.1 Public actuel

D'après l'analyse des courriels qui couvre la période 2003-2010, les critiques sur le site ne sont pas nombreuses. Une des remarques que nous avons relevée résumant bien les besoins de l'internaute est la suivante : « *Le système n'est pas simple d'emploi, un peu laborieux et archaïque d'un certain point de vue et assez difficile d'accès, ressemblant à du vieux Gallica!* ». Néanmoins, d'autres usagers précisent que les sources sont très propres au niveau de la qualité de l'image. Ils soulignent le fait que la consultation donne l'impression d'avoir l'ouvrage devant soi.

Mais la demande principale et récurrente des utilisateurs concerne le téléchargement de la totalité ou d'une partie de l'ouvrage, comme le montrent bien ces extraits : « *Existe-il un moyen de télécharger des fragments plus importants, volumes ou totalité, des ouvrages?* », « *Il serait agréable, voire très utile de pouvoir télécharger les fichiers numérisés sur la même page car la consultation écran est parfois assez fastidieuse* ».

Ces remarques dénotent bien le besoin d'appropriation, d'impression et de consultation hors connexion. En dehors de ces considérations, les usagers ont peut-être besoin de consulter le document dans son aspect global, avec les tables des matières et les fac-similé mis en regard.

Dans un autre registre, un internaute nous a envoyé un courriel nous demandant une liste des mises en ligne successives.

Sur le web, nous avons relevé sur ce site⁴ des évaluations de bibliothèques numériques, la remarque d'un internaute qui décrit à lui seul les principaux problèmes du Cnum : « *L'ouverture d'une notice n'ouvre aucune fenêtre. La consultation du texte se fait page par page, aucune possibilité de visionner l'ensemble des pages en même temps. Il n'y a pas d'information sur les auteurs... C'est un site assez triste où il y a peu d'images et les couleurs sont mal choisies. Beaucoup de texte écrit au kilomètre. Trop de scrolling vertical. Arborescence pas très claire, l'internaute est vite perdu. Manque d'informations sur les autorités et les mises à jour. L'image représentant des livres sur la page d'accueil est affreuse, elle est mal choisie et mal disposée. C'est dommage qu'un site qui a la vocation d'être une bibliothèque numérique soit aussi mal présenté. L'internaute lambda ne restera pas sur le site. Seul l'internaute connaisseur pourra comprendre le site.* ».

Enfin, le reste des critiques des autres usagers se porte sur l'aspect démodé du graphisme, et surtout sur l'absence de texte intégral.

Il apparaît aussi un autre besoin sous-jacent que nous avons relevé dans le mémoire

4. <http://biblionum.pbworks.com/w/page/12576609/CNUM>

de [Cla08, page 11]. Il est indiqué que le mot « carte » est un des mots-clés les plus utilisés dans le moteur de recherche interne. Bien que cette étude soit cantonnée au périodique « La Nature », on peut supposer qu'elle peut être généralisée à l'ensemble du corpus.

4.3 Étude des statistiques d'accès

Les demandes différant considérablement d'un utilisateur à l'autre, seule une analyse des journaux d'accès peut nous fournir une réponse définitive sur les réels besoins à mettre en place. Nous abordons dans un premier temps l'étude des statistiques de consultation du site par l'outil AWStats, installé dès la première mise en ligne du Cnum. Puis dans un deuxième temps, nous analysons les logs d'accès au serveur HTTP Apache. L'ensemble de ces informations nous orientera sur les décisions à prendre pour le nouveau site.

4.3.1 L'analyseur AWStats et logs d'Apache

Nous décrivons sur l'ensemble de l'année 2011 le détail des visites, de visiteurs uniques, de pages, de hits, de transfert de fichiers, de navigateur. . .

Nous détaillons au préalable les différents titres de colonnes provenant de la documentation de l'outil Awstats :

- Visiteurs différents : nombre d'hôtes ou d'adresses IP différents pour accéder au site et voir au moins une page. Ce chiffre reflète le nombre de personnes physiques (visiteurs) différentes ayant accédé au site.
- Visites : on considère une nouvelle visite pour chaque consultation d'une page par un visiteur ne s'étant pas connecté dans la dernière heure.
- Pages : nombre de fois qu'une page du site est vue (cumul de tout visiteur, toute visite). Ce compteur diffère des « hits » car il ne comptabilise que les pages HTML et non les images ou autres fichiers.
- Hits : nombre de fois qu'une page, image, fichier du site est vu ou téléchargé par un visiteur.
- Bande passante : nombre d'octets téléchargés lors des visites du site. Il s'agit aussi bien du volume de données dû au chargement des pages et images que des fichiers téléchargés.

L'historique annuel :

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Mois	Visiteurs différents	Visites	Pages	Hits	Bande passante
Janvier	10869	15877	1033060	1463858	188.89 Go
Février	9557	13398	1050475	1440752	187.61 Go
Mars	9744	14269	860700	1250961	204.69 Go
Avril	8722	12618	756511	1097719	142.10 Go
Mai	9428	13068	632316	952685	144.31 Go
Juin	8803	12615	660035	968191	185.93 Go
Juillet	7871	11190	589222	893551	210.69 Go
Août	8169	11408	693087	985139	130.15 Go
Septembre	8843	12572	607136	910639	152.55 Go
Octobre	10125	14397	651656	957443	131.28 Go
Novembre	10615	15650	1073135	1474353	198.05 Go
Décembre	10132	15215	946949	1327058	150.62 Go
Total	112878	162277	9554282	13722349	2026.87 Go

TABLE 4.1 – Historique mensuel de l'année 2011

Nous voyons que le nombre de visiteurs annuels reste assez stable, sauf pendant la période estivale. Une baisse conséquente de la consultation nous pousserait à analyser plus précisément ce creux, et de pallier ce problème sur la nouvelle interface du Cnum, mais ce n'est pas encore le cas.

Système matériel :

Les statistiques de cette catégorie étant absentes dans l'outil AWStats, nous avons analysé à la place les logs d'accès du serveur Apache. Au préalable, nous avons vérifié que la requête de Google « *site :cnum.cnam.fr Mobile* » ne contient que peu de résultat (un seul au final) avec le mot « Mobile » en majuscule. On peut considérer que l'erreur n'est pas significative sur ce mot et qu'il ne risque pas d'interférer avec ceux figurant dans les tables des matières. Le champ en-tête « User-Agent » de la requête HTTP ne peut être exploité dans les lignes des fichiers journaux car sa position n'est pas identique.

Ainsi, la requête « *cat /home/cnum/httpd/logs/access_log | grep Mobile | grep iPad | awk -F " " ' print \$1 ' | sort | uniq -c | sort -rn | wc -l* » donne environ 800 utilisateurs différents, hélas nous ne pouvons regrouper les adresses IP dynamiques d'un même utilisateur. De même que la requête « *cat /home/cnum/httpd/logs/access_log | grep BlackBerry | grep -v google | awk -F " " ' print \$1 ' | sort | uniq -c | sort -g | wc -l* » donne à peu près une quinzaine d'utilisateurs différents sous Blackberry. Enfin, la commande « *cat /home/cnum/httpd/logs/access_log | grep MozBraille* »⁵ ne donne aucun résultat.

5. MozBraille est une extension pour transformer le navigateur Internet Firefox en une version accessible conçu pour les utilisateurs aveugles ou malvoyants.

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Nous voyons qu'un grand nombre d'utilisateurs emploie une interface tactile comme l'Ipod, mais l'utilisation du BlackBerry reste encore très minoritaire.

Top 10 des systèmes d'exploitation :

Système d'exploitation	Hits	%
Windows	11412522	83.1%
Macintosh	1661808	12.1%
Linux	410900	2.9%
Inconnu	236109	1.7%
Symbian OS	437	0%
Unknown Unix system	189	0%
BSD	158	0%
Sony PlayStation	145	0%
Nintendo Wii	51	0%
HP UX	29	0%
Autres	1	0%

TABLE 4.2 – Classement des systèmes d'exploitation

Top 10 des navigateurs Internet :

Navigateur	Aspirateur	Hits	%
MS Internet Explorer	Non	5 767 371	42 %
Firefox	Non	5 076 364	36.9 %
Google Chrome	Non	1 259 313	9.1 %
Safari	Non	1 178 227	8.5 %
Opera	Non	181 315	1.3 %
Mozilla	Non	79 371	0.5 %
WebCopier	Oui	76 057	0.5 %
Inconnu	?	74 851	0.5 %
Netscape	Non	14 170	0.1 %
FDM Free Download Manager	Oui	5 258	0 %

TABLE 4.3 – Classement des navigateurs Internet

La déclinaison du nouveau site doit être compatible sur ces principaux navigateurs.

Provenance géographique des visiteurs :

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

<u>Domaines/Pays</u>	<u>Pages</u>	<u>Hits</u>	<u>Bande passante</u>
France	3 673 284	5 256 254	749.62 Go
Network	3 017 105	4 228 846	534.26 Go
Inconnu Inconnu	959 095	1 477 448	319.08 Go
Commercial	298 412	396 888	65.80 Go
Belgium	246 115	395 335	43.01 Go
Germany	208 608	311 453	27.47 Go
Italy	189 202	300 547	43.97 Go
Austria	159 440	194 872	10.20 Go
Switzerland	118 572	166 787	44.09 Go
Canada	86 732	135 135	54.92 Go
Netherlands	64 231	90 778	5.61 Go
Poland	59 506	81 824	8.48 Go
Portugal	44 681	62 691	9.44 Go
Japan	41 058	65 153	7.37 Go
Spain	36 892	53 900	8.09 Go
Brazil	30 154	45 951	10.92 Go
Chile	26 751	31 603	1.55 Go
USA Educational	25 037	34 123	4.69 Go
Czech Republic	23 822	30 132	3.55 Go
Colombia	20 799	27 006	964.93 Mo
Greece	16 732	22 520	1.28 Go
Russian Federation	16 483	26 190	5.30 Go
Mexico	15 290	22 105	2.73 Go
Romania	15 144	19 868	3.64 Go
United Kingdom	14 711	22 133	1.42 Go
Argentina	14 092	20 045	12.77 Go
Non-Profit Organizations	11 342	17 673	8.16 Go
San Marino	10 264	12 835	675.51 Mo
Denmark	9 929	12 802	5.68 Go
Australia	8 483	13 890	1.40 Go

TABLE 4.4 – Domaines/pays visiteurs

Les visiteurs du site proviennent essentiellement des pays européens puis des États-Unis, ensuite de divers pays de par le monde. Ces chiffres pourraient justifier la déclinaison de l'interface en plusieurs langues. Mais, l'étude des logs dans 4.3.2 nous démontre que les internautes n'utilisent pas systématiquement l'outil de traduction de langue proposé gratuitement par certains sites. De plus, nous recevons régulièrement par courriel des demandes de reproduction d'images de l'étranger, ce qui montre bien que le contenu du site rédigé en langue française n'apparaît pas comme un obstacle à sa consultation pour les utilisateurs non francophones.

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Phrases clés recherchés :

Ce tableau détaille la liste des phrases ou mots clés les plus souvent utilisés, pour retrouver et accéder au site du Cnum depuis la plupart des moteurs de recherche Web.

<u>40220 phrases clé différentes</u>	<u>Hits</u>	<u>%</u>
cnum	3424	5.6 %
cnum cnam	753	1.2 %
cnam	425	0.6 %
cnam cnum	365	0.5 %
cnum.cnam.fr	358	0.5 %
cnum la nature	225	0.3 %
cnum.cnam	156	0.2 %
dictionnaire technologique	103	0.1 %
structure et fonctionnement des ordinateurs	101	0.1 %
cnam bibliothèque numérique	100	0.1 %
web	91	0.1 %
conservatoire des arts et métiers	85	0.1 %
turgan les grandes usines	84	0.1 %
magneto rb	81	0.1 %
conservatoire numerique des arts et metiers	79	0.1 %
abbé nollet	75	0.1 %
houe a bras	69	0.1 %
exposition universelle 1867	64	0.1 %
conservatoire numerique	63	0.1 %
turgan	62	0.1 %
cnumcnam	56	0 %
la lumiere electrique 1879	55	0 %
barometre enregistreur richard	55	0 %
site cnum.cnam.fr tsf	53	0 %
conservatoire numérique des arts et métiers	53	0 %
julien turgan	53	0 %
cnam la nature	52	0 %
Autres phrases	55068	90.1 %

TABLE 4.5 – Classement des phrases clés recherchés

Mots-clés recherchés :

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

<u>Mot-clé</u>	<u>Nombre</u>	<u>%</u>
cnum	5967	2.4 %
exposition	3937	1.6 %
cnam	3219	1.3 %
universelle	2988	1.2 %
paris	1866	0.7 %
1900	1402	0.5 %
cnum.cnam.fr	1249	0.5 %
arts	1183	0.4 %
nature	1085	0.4 %
catalogue	1074	0.4 %
site	1010	0.4 %
machine	986	0.4 %
1889	823	0.3 %
instruments	815	0.3 %
electrique	747	0.3 %
pdf	742	0.3 %
richard	694	0.2 %
métiers	680	0.2 %
industrie	675	0.2 %
appareil	668	0.2 %
microscope	662	0.2 %
usines	618	0.2 %
conservatoire	618	0.2 %
Autres mots	196645	80.4 %

TABLE 4.6 – Top 10 des mots-clés recherchés

Ces phrases et mots-clés déterminent ceux à intégrer éventuellement dans les balises « meta keywords ». Les expositions universelles étant une thématique bien spécifique de notre corpus, nous pourrions intégrer ces mêmes mots-clés à ces documents.

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Lien depuis un moteur de recherche Internet :

<u>Moteurs de recherche</u>	<u>Pages</u>	<u>%</u>	<u>Hits</u>	<u>%</u>
Google	59085	93.7 %	59 571	93.7 %
Yahoo!	1676	2.6 %	1 686	2.6 %
Unknown search engines	1052	1.6 %	1 052	1.6 %
Ask	240	0.3 %	240	0.3 %
Voila	221	0.3 %	221	0.3 %
Microsoft Windows Live	206	0.3 %	212	0.3 %
AOL (fr)	143	0.2 %	143	0.2 %
Index	86	0.1 %	86	0.1 %
Gazeta.pl	51	0 %	51	0 %
Yandex	48	0 %	54	0 %
ix quick	28	0 %	28	0 %
MyWebSearch	27	0 %	27	0 %
AltaVista	16	0 %	16	0 %
Toile du Quebec	14	0 %	14	0 %
AOL	13	0 %	13	0 %
DMOZ	12	0 %	12	0 %
Lycos	11	0 %	11	0 %
Scroogle	10	0 %	10	0 %
Virgilio	6	0 %	6	0 %
Web.de	5	0 %	5	0 %
Seznam	5	0 %	5	0 %
Stumbleupon	4	0 %	4	0 %
Baidu	4	0 %	7	0 %
Excite	4	0 %	4	0 %

TABLE 4.7 – Lien sur le Cnum depuis un moteur de recherche Internet

Le moteur de recherche Google reste sans contexte le plus utilisé par les internautes. Nous veillerons à bien déclarer le fichier sitemap.xml servant à l'indexation automatique des pages par les robots. Cette technologie développée à l'origine par Google a été ensuite progressivement adoptée par les autres moteurs.

Lien depuis une page externe (autres sites, hors moteurs) :

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

6 568 pages différentes	Pages	%	Hits	%
http://translate.googleusercontent.com/translate_c	13 954	16.8 %	25 662	24.1 %
http://gallica.bnf.fr/Search	8 792	10.6 %	8 792	8.2 %
http://abu.cnam.fr	1 594	1.9 %	1 594	1.4 %
http://www.bing.com/search	1 576	1.9 %	1 576	1.4 %
http://webcache.googleusercontent.com/search	1 375	1.6 %	1 963	1.8 %
http://www.lrpresse.fr/trains/viewtopic.php	1 249	1.5 %	1 276	1.2 %
http://bibliotheque.cnam.fr	1 214	1.4 %	1 214	1.1 %
http://bibliotheque.cnam.fr/jsp/fiche_pagelibre.jsp	943	1.1 %	943	0.8 %
http://www.arts-et-metiers.net/musee.php	864	1 %	864	0.8 %
http://ng263.prod.exalead.com	842	1 %	842	0.7 %
http://cafcom.free.fr/piano/pianos.html	830	1 %	830	0.7 %
http://blog.bnf.fr/gallica/	796	0.9 %	796	0.7 %
http://www.bloooo.fr/forums/viewtopic.php	784	0.9 %	889	0.8 %
http://seaus.free.fr/spip.php	576	0.6 %	576	0.5 %
http://en.wikipedia.org/wiki/Giant_of_Castelnau	498	0.6 %	498	0.4 %
http://abu.cnam.fr/index.html	419	0.5 %	419	0.3 %
http://www.nikolatesla.fr/documents.htm	408	0.4 %	408	0.3 %
http://www.gratisnette.com	383	0.4 %	383	0.3 %
wikipedia Exposition universelle	381	0.4 %	381	0.3 %
http://fr.wikipedia.org/wiki/Compteur_d'eau	352	0.4 %	352	0.3 %
http://stella.cnam.fr	351	0.4 %	351	0.3 %
http://retro-forum.com/viewtopic.php	320	0.3 %	320	0.3 %
Exposition Universelle 1855	317	0.3 %	317	0.2 %
http://fr.wikipedia.org/wiki/La_Nature	298	0.3 %	298	0.2 %
List of tallest people	286	0.3 %	286	0.2 %
http://www.sfr.fr/recherche/	245	0.2 %	245	0.2 %
http://fr.wikipedia.org/wiki/Compteur_d%27eau	244	0.2 %	244	0.2 %
http://www.industrientionale.fr	241	0.2 %	241	0.2 %
http://www.facebook.com/l.php	234	0.2 %	234	0.2 %
http://fr.wikipedia.org/wiki/Barbe	231	0.2 %	231	0.2 %
http://fr.wikipedia.org/wiki/Henri_Giffard	230	0.2 %	230	0.2 %
http://www.blooo.fr/forums/viewtopic.php	228	0.2 %	399	0.3 %

TABLE 4.8 – Lien sur le Cnum depuis une page externe

Le premier lien depuis une page externe provient du service de traduction de Google. Nous ne nous attardons pas plus longtemps sur ce lien particulier car la langue de conversion a été étudiée dans le chapitre 4.3.2.

Par ailleurs, nous remarquons que le site de Wikipédia référence d'importants liens de rebonds vers le Cnum. Ceci nous alerte sur les précautions à prendre quant à d'éventuels changements d'urls internes.

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Durée des visites :

<u>Durée</u>	<u>Visites</u>	<u>%</u>
0s-30s	101 957	62.8 %
30s-2mn	16 178	9.9 %
2mn-5mn	11 987	7.3 %
5mn-15mn	13 000	8 %
15mn-30mn	7 214	4.4 %
30mn-1h	6 422	3.9 %
1h+	5 516	3.3 %
Inconnu	3	0 %

TABLE 4.9 – Durée des visites

La grande majorité des visites dont la durée n'excède pas les 30 secondes nous indique peut-être un problème au niveau de l'interface. Est-ce que l'internaute n'a pas su trouver l'information adéquate, l'arborescence du site est-elle compliquée, est-ce un problème d'ergonomie, est-ce lié à un taux de rebond ⁶ ?

Fichiers téléchargés :

Ce tableau indique le nombre de fois qu'une page, image, fichier du site est vu ou téléchargé par un visiteur.

6. http://fr.wikipedia.org/wiki/Taux_de_rebond

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Types de fichiers	Hits	%	Bande passante	%
cgi	9 007 759	65.6 %	7.14 Go	0.3 %
gif	4 015 750	29.2 %	170.49 Go	8.4 %
html	502 241	3.6 %	16.37 Go	0.8 %
jpg	150 976	1.1 %	4.76 Go	0.2 %
pdf	41 306	0.3 %	1827.32 Go	90.1 %
php	2 891	0 %	117.98 Mo	0 %
css	1 240	0 %	1.10 Mo	0 %
js	57	0 %	10.26 Ko	0 %
pl	33	0 %	21.59 Ko	0 %
zip	26	0 %	476.71 Mo	0 %
xml	24	0 %	1.92 Mo	0 %
png	20	0 %	198.07 Mo	0 %
Inconnu	14	0 %	786.68 Ko	0 %
tdm	7	0 %	272.47 Ko	0 %
tdp	2	0 %	59.58 Ko	0 %
fr	2	0 %	19.48 Ko	0 %
gz	1	0 %	16.36 Mo	0 %

TABLE 4.10 – Classement des fichiers téléchargés

Nous voyons que 41306 fichiers pdf ont été téléchargés du site. Les principaux sont les ouvrages numérisés, et le reste les documents d'informations (mémoire, articles, thèmes, glossaires, bibliographies...) représentent une infime partie. Le nombre d'ouvrages téléchargés au format pdf est important, malgré le fait que ces documents n'ont pas été générés en totalité avant cette période. Comme le montre bien le chapitre de l'analyse des besoins, l'export du document pdf est une des demandes principales des internautes.

L'auteur [Cub06, p. 34, Fig. 1.8] montre dans son analyse des logs que le nombre croissant de documents pdf proposé sur la période 2000-2005 n'a pas pour autant réduit le nombre de visites et par conséquent l'acte de feuilletage. L'interface feuilletage reste encore primordiale pour les utilisateurs. Par ailleurs, nous constatons que certains sites de bibliothèques numériques comme Numdam⁷, proposent uniquement le téléchargement du document pdf des ouvrages numérisés à la place d'une interface feuilletage.

7. <http://www.numdam.org>, site de recherche et téléchargement d'archives de revues mathématiques numérisées

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Pages-URL (Top 10) :

<u>6 536 pages différentes</u>	<u>Pages vues</u>	<u>Taille moyenne</u>	<u>Entrée</u>	<u>Sortie</u>
/CGI/menug.cgi	2 309 150	857 Octets	3 150	8 945
/CGI/page.cgi	2 135 704	236 Octets	1 181	8 185
/CGI/menud.cgi	2 130 813	1.40 Ko	930	47 104
/CGI/fpage.cgi	2 087 260	560 Octets	25 384	2 937
/ (page accueil)	114 822	9.25 Ko	32 074	21 710
/CGI/redirect.cgi	113 227	1.13 Ko	11 686	861
/CGI/rechdetser.cgi	55 066	2.26 Ko	85	702
/CGI/gplanche.cgi	50 630	564 Octets	576	73
/CGI/sresrech.cgi	41 308	563 Octets	1 897	129
/CGI/rechglob.cgi	34 172	12.23 Ko	190	2 215
/CGI/rechser.cgi	24 060	6.90 Ko	62	271
/CGI/gpage.cgi	20 217	527 Octets	1 035	195
/RUB/cata.html	17 369	32.12 Ko	959	3 236
/RUB/fcata.html	16 386	344 Octets	3 560	397
/RUB/frech.html	12 709	337 Octets	1 282	168
/RUB/rech.html	12 678	3.93 Ko	158	719
/SER/4KY28.html	9 287	26.22 Ko	95	765
/DET/4KY28.63.html	9 255	142.79 Ko	3 906	3 904
/RUB/cata_expo.html	5 136	45.14 Ko	99	697
/RUB/fcata_expo.html	5 103	369 Octets	826	32
/CGI/rechmono.cgi	4 883	2.81 Ko	8	93
/RUB/fbiblio.html	4 145	382 Octets	285	238
/RUB/biblio.html	3 735	5.74 Ko	28	308
/RUB/f_musee_conservatoire.html	3 566	406 Octets	561	28
/RUB/musee_conservatoire.html	3 424	3.86 Ko	31	420
/RUB/infoplus.html	2 882	72.16 Ko	2 031	1 819
/RUB/finfo.html	2 818	389 Octets	241	29
/RUB/fcata_construct.html	2 803	364 Octets	1 505	285
/SER/BSPI.html	2 734	27.07 Ko	265	358
/RUB/info.html	2 705	12.90 Ko	30	294
Autres	482 130	3.92 Mo	85 120	69 409

TABLE 4.11 – Classement des pages urls

Nous constatons que le feuilletage restreint aux pages de planches et illustrations (script CGI gplanche.cgi) est une fonctionnalité bien plus utilisée que celle de la recherche avancée.

Codes Status HTTP :

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Code	Message	Hits	%	Bande passante
302	Moved temporarily (redirect)	256 427	67.6 %	70.10 Mo
206	Partial Content	95 900	25.2 %	1199.96 Go
404	Document Not Found	22 075	5.8 %	24.16 Mo
405	Method not allowed	1 648	0.4 %	1.50 Mo
403	Forbidden	1 370	0.3 %	1.37 Mo
400	Bad Request	829	0.2 %	476.58 Ko
416	Requested range not valid	326	0 %	126.07 Ko
406	Document not acceptable to client	249	0 %	0 Ko
301	Moved permanently (redirect)	205	0 %	62.17 Ko
414	Requested filename too long	30	0 %	9.73 Ko
500	Internal server Error	6	0 %	4.56 Ko

TABLE 4.12 – Codes Status HTTP

La liste des codes du protocole HTTP détaillée dans rfc2616⁸ nous permet, dans les messages associés aux codes, de détecter les erreurs éventuelles. Nous remarquons dans ce tableau que les deux erreurs principales sont liées au document déplacé et à la transmission partielle de la requête.

De même, l'analyse des erreurs des logs de connexion dont la taille du fichier dépasse les 50Go, nous témoigne de l'existence d'énormes « bugs » informatiques à corriger. Nous voyons ainsi que la plus importante concerne les résultats de recherche, qui génère automatiquement à elle seule, pour une requête cinq lignes d'erreurs dans les logs.

4.3.2 Site multilingue

Nous avons analysé dans nos logs la construction des requêtes du service gratuit Google traduction, dont voici une des lignes : <http://translate.google.fr/translate?hl=fr&sl=fr&tl=en&u=http%3A%2F%2Fnum.cnam.fr%2FDET%2F8XAE348.6.html>. Elle déclare les variables et les données « sl=fr » et « tl=en », comme étant la langue source et cible. Suite à ce constat, nous avons lancé sur les fichiers logs la ligne de commande Unix « `cat /home/cnum/httpd/logs/access_log | grep translate.google. | grep 'tl= | awk -F " " ' print $11 ' | awk -F "tl=" ' print $2 ' | awk -F "'" ' print $1 ' | sort | uniq -c | sort -gr` ». Nous avons regroupé les résultats (triés par ordre décroissant du nombre de pages traduites) dans le tableau suivant :

8. <http://tools.ietf.org/html/rfc2616>

4.3. ÉTUDE DES STATISTIQUES D'ACCÈS

Langue cible	Nombre de pages traduits
en : anglais	1387
pl : polonais	542
it : italien	476
de : allemand	109
ru : russe	102
ja : japonais	63
es : espagnol	46
nl : néerlandais	20
pt : portugais	16
zh-TW : Taiwanais Mandarin	1

TABLE 4.13 – Tableau du nombre de pages traduites dans une autre langue

Parfois, certains utilisateurs ne traduisent qu'une seule page : la page synoptique. On peut en tirer comme conclusion que leur besoin concerne uniquement la traduction des données et non des métadonnées. Autrement dit, les gens sont intéressés par la traduction des mots des tables des matières et non par les termes de navigation. De par ce constat, y-a-t-il un réel intérêt à développer un site en anglais ou dans d'autres langues ?

Chapitre 5

État de l'art

Ce chapitre présente l'état de l'art des bibliothèques numériques. Nous décrivons la structure, les interfaces et les fonctionnalités de certains sites, puis nous présentons quelques logiciels de ce domaine. Nous terminons par la présentation de quelques interfaces de visualisation.

5.1 Les bibliothèques numériques

Nous veillerons tout d'abord à définir différentes typologies de bibliothèques numériques avec leurs fonctionnalités. En effet, cette catégorisation nous semble nécessaire afin de couvrir l'exhaustivité de cette étude.

5.1.1 Introduction

Les bibliothèques numériques ou bibliothèques en ligne, électroniques ou virtuelles gèrent une collection organisée de documents numériques consultables (en général gratuitement) via un accès distant (Internet). Ces documents sont généralement des livres, des articles de journaux, des périodiques, des enregistrements sonores (cassette et disques), des enregistrements vidéo (cassette et DVD), des cartes et plans, des partitions, des documents électroniques (natifs numériques).

Nous nous intéressons ici essentiellement aux bibliothèques de livres numérisés telles que définies dans [Cub06, page 8]. Ces documents numériques qui ne le sont pas originellement, sont issus d'un processus de numérisation d'un média « analogique ». Ainsi, nous ferons une différence entre les librairies en ligne telles qu'Amazon¹, ACM² Digital Library, Ieeexplore³, Google Scholar⁴...

1. amazon.fr

2. Association for Computing Machinery dl.acm.org

3. ieeexplore.ieee.org

4. scholar.google.fr

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

Cependant, à l'heure actuelle, cette définition peut être remise en cause avec les systèmes de liseuses électroniques (e-Reader, Amazon kindle,...), la puissance de calcul et la capacité mémoire de ces appareils permettant de contenir à la fois des documents numériques et numérisés. On voit ainsi que le champ des bibliothèques numériques a rapidement évolué avec le temps.

Nous ne prétendons pas être exhaustif dans notre recensement des bibliothèques numériques. Nous décrivons tout d'abord les premières bibliothèques numériques du web puis les plus importantes en terme de structure, ensuite celles plus spécifiques à notre domaine, à savoir les bibliothèques de niche et enfin divers types de bibliothèque. Nous pensons que les bibliothèques numériques de grandes envergures disposent de ressources et de moyens financiers importants, leur permettant de proposer des interfaces et des fonctionnalités peut-être moins accessibles à d'autres.

5.1.2 Les premières du web

Le projet Gutenberg et ABU semblent être les premières bibliothèques numériques du Web. Même si elles ne présentent actuellement que peu d'intérêts techniques, elles étaient réellement innovantes à l'époque, c'est pourquoi il nous est apparu intéressant de les mentionner étant des précurseurs de ce domaine.

5.1.2.1 Le projet Gutenberg

[Leb09, page 5-13] décrit ce projet comme étant « *Fondé par Michael Hart en juillet 1971 alors qu'il était étudiant à l'Université d'Illinois (États-Unis), le Projet Gutenberg a pour but de diffuser gratuitement par voie électronique le plus grand nombre possible d'oeuvres littéraires. (...) Lorsque l'utilisation du web se généralise dans les années 1990, le Projet Gutenberg trouve un second souffle et un rayonnement international. Au fil des ans, des centaines d'oeuvres sont patiemment numérisées en mode texte par des milliers de volontaires.* ».

Le site propose une navigation dans le catalogue par catégorie, nom d'auteur, titre, langue et derniers documents mis en ligne. Les documents sont consultables et téléchargeables sous divers formats : html, epub (avec ou sans les images), Kindle, pdf, texte brut en UTF-8. L'interface a l'avantage de proposer le document complet en texte brut sur une seule page html avec une navigation facilitée par des liens ancrés sur la table des matières. Certaines illustrations sont intégrées au mode texte.

Le site dispose d'un moteur de recherche interne pointant sur les champs des métadonnées et le plein texte mais exploite aussi celui de Google et Yahoo via des requêtes spécifiques du type « *site :gutenberg.org mots-clés* ».

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

5.1.2.2 Association des bibliophiles universels (ABU)

ABU⁵ a été créée en octobre 1993 par M. Pierre Cubaud [Cub06, page 18], n'ayant pas trouvé d'antécédent sur le web hormis le projet Gutenberg, nous pouvons ainsi affirmer qu'elle est la première bibliothèque numérique française du web. ABU diffuse des ouvrages classiques du domaine public francophone numérisés en mode texte grâce au travail collaboratif des membres de cette association.

Les fonctionnalités sont quasi identiques à celles du site Gutenberg. ABU propose un catalogue classé par auteurs, textes et aussi plusieurs liste de mots : mots communs, prénoms, nom de cités française, nom de pays, difficultés de la langue. Les documents sont consultables à la fois en texte complet ou page à page. L'interface feuilleteur contient comme fonctionnalité de recherche limitée au document consulté, une liste de mots classés par fréquence décroissante (ancêtre des nuages de mots-clés).

5.1.3 Bibliothèque numérique d'initiative privée

5.1.3.1 Google Books

La bibliothèque numérique la plus connue dans ce domaine est probablement Google Books. Elle fut lancée en 2004 et propose des ouvrages libres de droits, ainsi que des courts extraits de livres sous droits au travers d'une librairie en ligne en lien avec des éditeurs et distributeurs commerciaux.

La page d'accueil propose un catalogue classé par thématique et à l'identique de toutes les interfaces de Google, le service Google Books propose sur sa page d'accueil un formulaire de recherche simple. Il y est également proposé un lien vers la bibliothèque personnelle de l'utilisateur protégée par son identifiant et mot de passe. Elle donne la possibilité de créer et consulter des listes de documents achetés, consultés, évalués, préférés...

Pour une meilleur analyse et comparaison d'interface, nous prenons le cas d'un ouvrage numérisé à la fois par le Cnum et par Google Books, comme par exemple les « *Annales du Conservatoire impérial des arts et métiers, Volume 9* »⁶. Sur la première page de présentation de l'ouvrage sont indiqués le nom d'auteur, le nom d'éditeur et l'année d'édition suivi d'un formulaire de recherche en plein-texte. Ensuite viennent dans l'ordre les rubriques avis des internautes, quelques pages sélectionnées, les tables des matières (pas systématiquement), les expressions et termes fréquents dans l'ouvrage sous forme de nuage de mots-clés, les phrases de l'ouvrage fréquemment cités dans d'autres ouvrages, les informations bibliographiques, l'export de la notice au format Bibtex, EndNote et RefMan et enfin un Code QR d'identification (code-barres en deux dimensions).

5. abu.cnam.fr

6. <http://books.google.fr/books?id=qfMOAAAAMAAJ&dq=annales%20du%20conservatoire>

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

L'interface feuilleteur présente de nombreuses fonctionnalités comme un système de zoom, une consultation en simple et double page, une interface vignette, le téléchargement du document au format pdf. L'internaute peut sélectionner une partie de l'image et la partager sur un réseau social ou la citer dans une page web. Via un menu déroulant nous pouvons naviguer dans le livre par ses chapitres. De plus, la navigation dans l'ouvrage se fait en mode continue ou double page.

La recherche plein texte et son système de présentation constituent la réelle plus-value de cette interface. En effet, nous pouvons consulter les résultats de la recherche, soit au travers d'une liste d'extraits de pages contenant les mots recherchés qui sont géolocalisés et mis en surbrillance, soit à partir d'une heuristique visuelle représentée par une barre verticale placée à droite de l'interface.

Nous ne décrivons pas la page de recherche avancée qui nous paraît très classique et donc ne présente pas d'intérêt particulier. Toute la puissance du moteur et l'attractivité de ce type d'interface est situé au niveau de la recherche simple. Nous pouvons critiquer l'absence d'une table d'illustration⁷ et d'une « notice mère » dans le cas d'un volume de périodique. La page catalogue est très restreinte au niveau de la consultation, elle propose à l'utilisateur plus un aperçu qu'une réelle exploration exhaustive du fonds.

5.1.4 Bibliothèques numériques institutionnelles

5.1.4.1 Gallica

Nous présentons la bibliothèque numérique Gallica⁸, elle propose des documents numérisés par la Bibliothèque nationale de France et par les bibliothèques partenaires, mais aussi des contenus numériques soumis au droit d'auteur en offre légale. Dans ce dernier cas, la recherche s'effectue sur les métadonnées et le texte intégral mais seuls des courts extraits sont proposés. L'internaute est renvoyé par rebond sur les sites des e-distributeurs s'il souhaite consulter l'ouvrage au complet.

L'accès aux ouvrages s'effectue suivant différentes rubriques comme les thématiques, les sélections, les actualités, les réseaux sociaux, les flux rss, les documents les plus consultés, les nouveaux documents, les derniers billets du blog et le moteur de recherche. L'interface feuilleteur propose à peu près les mêmes fonctionnalités que celles de Google Books, hormis le fait que la plupart des fonctionnalités sont

7. Google Books ne fait pas de création ex nihilo sauf de la notice.

8. <http://web.archive.org/web/19990117013434/http://gallica.bnf.fr/MetaPrincipal.htm> (version de 1999), <http://gallica.bnf.fr/?lang=FR> (version de 2014), <http://gallicalabs.bnf.fr> (version de 2015)

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

concentrées sur la même interface (en dehors du mode plein écran). Il y a un système de recherche plein texte, un lecteur exportable, des permaliens, un système de signalement d'anomalies classées par catégorie, des systèmes de partage vers les réseaux sociaux. De plus, il existe deux modes feuilleteurs, un réduit et l'autre plein écran qui offre plusieurs fonctionnalités dont la navigation en page simple, double et continue. Il y a également une barre de défilement supérieure pour naviguer rapidement dans l'ouvrage (système identique sur le site Archive⁹).

D'après les informations fournies sur cette page d'aide à la recherche et un blog de Gallica¹⁰, il est indiqué que le moteur de recherche¹¹ est basé sur un logiciel produit par la Société Exalead. Les fonctionnalités offertes par la recherche avancée sont très riches. Lors de la saisie dans le formulaire, des mots-clés suggérés sont proposés au fur et à mesure pour éviter de les ressaisir complètement. Nous pouvons rechercher en croisant les champs avec différents opérateurs booléens (et, ou et sauf), tout en activant au choix la recherche exacte (chaîne de caractère précise), et ceux à partir des champs classiques de la notice bibliographique, par année d'édition, type de documents, langue, collections numériques... Le moteur permet surtout d'effectuer une recherche en texte intégral dans le cas où le document a été « océrisé ». Les données et les champs sélectionnés sont conservés en mémoire sur la dernière requête pour faciliter éventuellement la redéfinition de la recherche.

Les résultats de la recherche peuvent être classés par pertinence (pourcentage de classement calculé par le moteur), auteur, titre, date... avec un nombre de résultat par page configurable. Chaque document est présenté avec la vignette de sa page de titre et quelques éléments de la notice. Selon le cas sur chaque document, la mention « *Livre en mode image et en mode texte, recherche plein texte disponible* » fournit une information importante quant au type de consultation et de recherche à entreprendre. Dans le cas d'une recherche plein texte, quelques extraits de pages en mode texte sont présentés avec les termes recherchés mis en surbrillance.

Dans le cas d'un volume de périodique, un regroupement est proposé pour rebondir ou relancer la recherche sur tous les autres volumes de ce périodique.

L'affinage de la recherche dans les pages de résultats appelée recherche par facettes est une fonctionnalité utile pour réduire le nombre de documents à consulter.

Dans le feuilleteur, nous pouvons critiquer le mode de navigation double page qui reconstitue virtuellement les pages avec reliure, et crée indirectement un artefact au centre de la page biaisant d'une certaine manière le principe de lecture de ce mode. La navigation dans divers volumes de revues n'est pas aisée, elle est seulement

9. <https://archive.org/stream/devisesembleme001afeu#page/n7/mode/2up>

10. <http://gallica.bnf.fr/html/und/comment-faire-une-recherche>, <http://blog.bnf.fr/gallica/index.php/2013/09/12/un-nouveau-moteur-de-recherche-pour-gallica>

11. La troisième version de Gallica datant de 2007 utilisait le moteur de recherche Lucene d'après <http://fr.wikipedia.org/wiki/Gallica>

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

permise sur le numéro suivant ou précédent du volume consulté. Les boutons de navigation permettant de passer à la page suivante et précédente, placés d'un côté et de l'autre de l'interface oblige l'utilisateur à déplacer le pointeur sur une plus grande distance. Mais ces deux problèmes ont été résolus sur la future version de Gallica (interface proposée en consultation test).

Nous relevons comme point positif la recherche en plein texte et le mode d'affichage en texte brut, qui couplé au module de synthèse vocale peut s'avérer fort utile pour les personnes malvoyantes. Par ailleurs au niveau de l'affichage, la page est contenue dans un cadre qui fixe d'emblée la taille de l'image et évite de défiler horizontalement. Ce système laisse de la place libre en haut et bas de page ainsi que sur les côtés pour placer éventuellement plus d'informations ou d'outils. La consultation détaillée de la page est possible en réglant le niveau de zoom dans le mode feuilletage plein écran, et ceci sans défilement avec les barres mais uniquement en déplaçant la page. Une vignette contenant un petit encadré permet de situer la zone zoomée sur la page. L'interface propose au choix le téléchargement d'une sélection ou le document entier sous plusieurs formats.

Enfin, le système de moissonnage OAI permet aux bibliothèques partenaires d'offrir une plus grande visibilité sur ses collections. D'ailleurs, celles qui n'ont pas les moyens de mettre en place une bibliothèque numérique peuvent adhérer au système « *Gallica Marque Blanche* »¹². Elles peuvent ainsi bénéficier d'un site clé en main avec toutes les fonctionnalités de Gallica avec en plus la possibilité de le personnaliser et de l'exporter.

Il existe de nombreuses bibliothèques numériques du même type comme la British library ou la bibliothèque du Congrès... nous les avons parcourues en relevant les points essentiels et nous les avons décrit dans les chapitres suivants. Dès à présent, nous passons à l'étude des bibliothèques numériques dites de niche.

5.1.5 Bibliothèques numériques dites de niche

5.1.5.1 Le Cnum

Le Cnum est un décalque d'ABU, son objectif n'est alors plus de diffuser du texte en mode intégral mais des fac-similés de livres essentiellement en mode image. La première version du Cnum a été développée en trois ans par M. Pierre Cubaud, dont une partie du moteur de recherche a été développée par M. Alexandre Topol.

Nous ne nous attardons pas plus longtemps sur ce paragraphe, car nous l'avons largement présenté dans d'autres chapitres et de plus nous décrivons en détail tout l'environnement technique dans le chapitre 3.

12. http://www.bnf.fr/documents/JPAC2010_MARTIN_GODET.pdf (p. 31), <http://e-corpus.hypotheses.org/521>

Et bien d'autres... Il existe encore de nombreuses bibliothèques numériques de niche. Celles dans le domaines des sciences et techniques sont notamment la bibliothèque patrimoniale virtuelle de la bibliothèque interuniversitaire de santé (Medic@¹³), la bibliothèque numérique patrimoniale de l'UPMC (la Jubilothèque¹⁴), la bibliothèque de recherche et téléchargement d'archives de revues mathématiques numérisées (Numdam¹⁵). Au lieu d'en décrire plus et dans le détail, nous avons préféré lister d'autres types de bibliothèques dans notre état de l'art.

5.1.6 Les bibliothèques numériques en tout genre et sites divers

Le site bibliopedia¹⁶ référence un nombre très important de bibliothèques numériques de tout type. De même, en témoigne le nombre de réponse fournit par les résultats du moteur de recherche Google. En effet, ce dernier permet la recherche de pages similaires à une URL via par exemple les requêtes <https://www.google.fr/#q=related:cnum.cnam.fr>, <https://www.google.fr/#q=related:gallica.bnf.fr> ou <https://www.google.fr/#q=related:biblioth%C3%A8que+num%C3%A9rique>. Nous avons inventorié à partir de différents sites et leurs interfaces toutes les fonctionnalités utiles pour le Cnum en analysant également les aspects graphiques. Nous avons sélectionné au hasard des sites de bibliothèques avec des documents nativement numériques et numérisés sans différenciation de typologie. Par ailleurs, nous n'avons pas consulté uniquement des sites de bibliothèques numériques mais aussi d'autres sites s'y rapprochant comme des sites de musées, des collections, des portails, des sites patrimoniaux, des sites culturels...

Voici la liste de quelques sites que nous avons consultés :

- Site du musée d'Orsay (hors bibliothèque numérique) : <http://www.musee-orsay.fr>
- Site du musée du quai Branly (hors bibliothèque numérique) : <http://collections.quaibranly.fr/>
- Site du musée Guimet (hors bibliothèque numérique) : <http://www.guimet.fr/fr>
- Site du Musée Carnavalet (hors bibliothèque numérique) : <http://www.carnavalet.paris.fr>
- Centre Pompidou, établissement polyculturel (hors bibliothèque numérique) : <https://www.centrepompidou.fr>
- Bibliothèque-numérique de l'Enssib (bibliothèque de fichiers numériques) : <http://www.enssib.fr/bibliotheque-numerique>
- Bibliothèque municipale de Lyon : <http://www.bm-lyon.fr/trouver/>

13. <http://www.biusante.parisdescartes.fr/histmed/medica.htm>

14. <http://jubilotheque.upmc.fr>

15. <http://www.numdam.org>

16. http://www.bibliopedia.fr/wiki/Biblioth%C3%A8ques_num%C3%A9riques

5.1. LES BIBLIOTHÈQUES NUMÉRIQUES

- [basesdedonnees/Collections-numeriques.htm](#)
- Bibliothèque numérique patrimoniale de Cujas : <http://cujasweb.univ-paris1.fr>
 - Patrimoine numérique : <http://www.numerique.culture.fr/pub-fr/index.html>
 - Bibliothèque littéraire Jacques Doucet : <http://www.bljd.sorbonne.fr>
 - Archives du Collège des Irlandais : <http://archives.centreculturelirlandais.com>
 - La Bibliothèque Départementale André Desguine : <http://bibliotheque-desguine.hauts-de-seine.net/desguine/?lang=fr>
 - Site du Ministère de la culture et de la communication. Département de l'information et de la communication : <http://collections.culture.fr>
 - Patrimoine numérique de la bibliothèque de Valenciennes : <https://bibiguana-01.ville-valenciennes.fr>
 - La Bibliothèque numérique Scientifica : <http://www.cite-sciences.fr/cs/Satellite?c=Page&cid=1195217022070&pagename=Portail%2FMED%2FPortailLayout&pid=1195217022149>
 - La bibliothèque numérique de l'INHA : <http://bibliotheque-numerique.inha.fr/>
 - Wikisource, une bibliothèque de textes libres et gratuits : http://fr.wikisource.org/wiki/Cat%C3%A9gorie:Ouvrages_issus_du_partenariat_avec_la_BnF
 - Edition des manuscrits de Madame Bovary de Flaubert : <http://www.bovary.fr>
 - Portail de bibliothèques numériques : <http://www.hathitrust.org/>
 - Patrimoine numérisé de Besançon : <http://memoirevive.besancon.fr>
 - Bibliothèque nationale du Royaume-Uni : <http://www.bl.uk>

Au niveau du graphisme et des fonctionnalités, les points positifs qui se dégagent de certains de ces sites sont :

- le côté simple, épuré avec des dégradés de couleur et de la transparence
- l'accès à la structure hiérarchique de l'ouvrage dans le feuilleteur, le système de zoom avec un aperçu global
- les boutons de saut de 5,10 et 50 pages¹⁷
- la simplicité, la sobriété et couleurs gais au niveau du graphisme.

Les points négatifs à éviter sont :

- les zones vides car gaspillage de l'espace disponible
- le trop grand éparpillement et espacement des accès
- la taille réduite de la police de caractère entraînant une difficulté de lecture
- le trop grand nombre d'objets mouvants sur la page (animation flash)

17. <http://daten.digitale-sammlungen.de/~db/0003/bsb00034024/images/index.html?id=00034024&fip=yztssdasxdxydxsdydqrseayaqrsewqsdas&no=&seite=11>

- l’absence de présentation de la bibliothèque
- la densité graphique
- l’architecture compliquée
- la charge informationnelle importante : trop d’informations rapides, page trop chargée et peu lisible.

5.2 Les logiciels de bibliothèques numériques

Nous étudions dans cette partie un ensemble de logiciels pour bibliothèques numériques. Nous ne serons certainement pas exhaustif dans notre étude, car d’après [AcAM12], le nombre de logiciels existants s’élevant à plus d’une dizaine et les moyens de tous les tester ne nous sont pas accessibles. De plus, de par notre budget financier limité, nous avons exclu de notre étude les logiciels payants et étudié uniquement des logiciels libres. Nos critères de choix sont non seulement basés sur la gratuité des licences nous donnant une indépendance et une liberté envers un quelconque organisme, mais les logiciels sélectionnés doivent aussi être des projets maintenus par une large communauté de développeurs nous garantissant sa pérennité.

5.2.1 Tableau comparatif

Dans la réalisation de ce tableau, nous nous sommes basés sur les logiciels étudiés dans l’excellent ouvrage très complet de [AcAM12, p. 42]. Nous avons sélectionnés les critères qui nous paraissaient les plus pertinents à notre projet en laissant de côté d’autres, comme par exemple les formats de fichier supportés, le système de gestion de base de données (car migration relativement aisée), le serveur d’application. . . Le tableau suivant est extrait des pages de [AcAM12, p. 43,44,49,55,56,58,60 et 61]

5.2. LES LOGICIELS DE BIBLIOTHÈQUES NUMÉRIQUES

Logiciel	Moteur d'indexation et de recherche	Accessibilité RGAA	Service en ligne d'auto-inscription des usagers	Mise en valeur des documents par regroupement	Fonctions de manipulation de documents numériques	Gestion des droits d'accès au document	Sollicitation de toutes les données en recherche	Export bibliographiques
Invenio	Interne. Solr possible	A et probablement AA	Oui	Oui	Oui	Oui	Oui	Oui
Greenstone	Lucene ou interne	?	Non	Oui	Oui	Non	Oui	Non
Omeka	Solr	?	Oui	Oui	Oui	Oui	Non	Non
EPrints	Eprints Indexer et Xapian	?	Oui	Non	Oui	Oui	Oui	Oui
Suite ORLOAI	Solr	A probable	Non	Non	Non	Non	Oui	En cours
DSpace	Lucene	Tous les A au choix	Oui	Non	Oui	?	Oui	Oui

TABLE 5.1 – Tableau comparatif des logiciels libres de bibliothèques numériques

Nous voyons d'après ce tableau qu'une fonctionnalité manquante sur un logiciel est parfois disponible sur un autre et vice versa. Il s'agit désormais d'en sélectionner un dont les fonctionnalités manquantes soient les moins contraignantes à notre projet. Ainsi, les logiciels Invenio ou Omeka semblent les plus convenir, mais si nous rentrons précisément dans les détails de ces logiciels, nous nous rendons compte qu'ils ne sont pas complets car ils ne répondent pas aux besoins formulés dans le chapitre 4. Il s'avère que l'utilisation d'un logiciel libre du marché ne convient pas au Cnum. A la place, nous avons besoin d'une implémentation sur mesure, voir peut-être d'un développement « from scratch », autrement dit à partir de rien. Nous pouvons cependant entrevoir une éventuelle porte de sortie au travers de la réadaptation du framework Cnum actuel.

5.3 Les moteurs de recherche

La navigation par le catalogue reste assez limitée dès lors que le corpus atteint une taille importante. Ainsi, le recours à un moteur de recherche est des plus indispensables. Une solution logicielle pour bibliothèque numérique n'étant pas complètement satisfaisante, nous avons dissocié la partie gestion des documents de la partie recherche des informations. Ainsi, nous étudions dans ce chapitre différents moteurs de recherche libres. Nous ne nous sommes pas orientés vers un moteur de recherche visuel et cartographique car ils sont moins utilisés par le grand public d'après [FL06, p. 111-125]. D'ailleurs la plupart d'entre eux ont disparus du web comme Kartoo, Ujiko... Nous décrivons en premier lieu celui qui nous est apparu comme le plus répandu et connu de tous, à savoir le projet d'Apache : Lucene Solr.

5.3.1 Apache Lucene Solr

Le logiciel Solr¹⁸ est une déclinaison web du moteur Apache Lucene. Ce dernier est une bibliothèque libre développée en Java permettant d'indexer et de rechercher du texte. Il offre de multiples possibilités comme la recherche à facettes, des traitements linguistiques (lemmatisation, racinisation, suppression de points, mots au pluriel, acronymes)... Solr structure les données dans un document avec différents champs ou balises contenant par exemple la notice bibliographique d'un titre et sa table des matières et illustration (cf. un exemple dans le chapitre 6.2.4.3). Chacun de ses champs est indexé et représenté sous de multiples variantes lexicales.

A chaque requête, Solr classe les documents suivant un score basé sur un calcul de pertinence. Ce calcul prend en compte certains critères comme la fréquence d'apparition du terme dans les champs du document, le nombre de document contenant le terme... Ce lien¹⁹ fournit une explication détaillée du système de classement opéré par Solr. D'ailleurs, il existe une importante documentation sur

18. <http://lucene.apache.org/solr>

19. https://wiki.apache.org/solr/SolrRelevancyFAQ#How_are_documents_scored

5.3. LES MOTEURS DE RECHERCHE

Internet dont en particulier les guides de référence de [Fou13] et [Ros10].

Nous présentons une liste de fonctionnalités du moteur Solr :

- recherche plein texte avec la librairie Java Lucene : segmentation du texte et indexation des mots suivant la casse typographique, les accents, gestion des ligatures, des élisions et des signes diacritiques. . .
- recherche par lexèmes : racinisation, flexion
- intégration de fichiers json, csv, pdf, xml, bases de données via connecteurs
- recherche avec les algorithmes phonétiques d'indexation : Soundex, Metaphone, double Metaphone
- système de correction orthographique
- console d'administration ou configuration par des fichiers xml
- recherche par facettes et proximité
- mise en surbrillance des résultats
- gestion des formats de sorties en xml, php, Json, Python, Ruby
- mécanisme de réplication des données vers d'autres serveurs Solr
- recherche distribuée
- affectation d'un poids spécifique à chaque champs
- mise en cache des requêtes
- recherche géospatiale
- regroupement et limitation des résultats
- gestion des langues suivant disponibilité de l'algorithme
- correction orthographique
- gestion des synonymes
- groupement des clauses et formulation des sous-requêtes avec les parenthèses donnant un ordre de priorité
- définition des champs et des types de données
- recherche avec les opérateurs booléens « et », « ou », « sauf », « + », « - » dans les documents et les champs
- recherche par jokers (Wildcards) avec les opérateurs de troncature « ? » ou « * »
- recherche floue avec l'opérateur « ~ »
- recherche par intervalle pour les dates en particulier
- extension du projet : clustering avec Carrot, système cloud avec Elastic search, crawler avec Apache Nutch

Il existe bien d'autres moteurs de recherche libres mais aucun ne rivalise encore avec Solr sur leurs nombres de fonctionnalités. Au lieu d'en décrire d'autres, nous préférons établir une étude comparative de différents moteurs suivant certains critères.

5.3.2 Tableau comparatif

De multiples critères techniques entrent en ligne de compte dans la comparaison des moteurs de recherche tels que la gestion de la taille de l'index, le temps d'indexation, le temps de recherche, le format des documents et le type de données. Certaines fonctionnalités comme les recherches booléennes, par facette, par réplication, la recherche distribuée, l'affinement des résultats (facettes), la clusterisation, le type d'analyse linguistique, la gestion des langues, une console d'administration sont également pertinents selon les besoins. D'autres critères plus subjectifs comme la profusion de la documentation et de forums sur Internet, une plus grande dynamique de développement du projet sont aussi importants.

Nous présentons un tableau comparatif des moteurs de recherche libres extrait de [MBY07, p. 42]. Or, le résultat du classement de la pertinence de certains moteurs n'étant pas indiqué, le calcul final s'en trouve quelque peu biaisé :

Moteurs / Critères	Temps d'indexation	Taille de l'index	Temps de la recherche	Pertinence	Total
htDig	7	10	6	-	-
Indri	4	9	2	2	17
IXE	8	4	2	5	19
Lucene	10	2	4	-	-
MG4J	3	8	5	4	20
SwishE	5	5	8	-	-
Swish++	6	3	10	-	-
Terrier	9	7	9	3	28
XMLSearch	2	1	1	-	-
Zettair	1	6	6	1	14

TABLE 5.2 – Classement des moteurs de recherche par critères

Il en ressort que le moteur de recherche Zettair est le plus complet grâce à sa capacité à traiter une large collection de données dans un temps réduit. Ce dernier est talonné par XMLSearch, IXE, Indri, Lucene... Mais sur le site officiel de Zettair²⁰, il semble que le projet ne soit plus maintenu au vu de la dernière mise à jour datant du 15 octobre 2009.

Cependant, selon les auteurs de cet article, dans le cas où les données ne changent pas souvent, les autres solutions telles que MG4J, Terrier or Lucene sont également viables.

Un autre auteur [Sin09] a également comparé certaines de ces différentes solutions,

20. <http://www.seg.rmit.edu.au/zettair>

mais il a obtenu des résultats légèrement différents du précédent que nous avons synthétisé dans le tableau suivant :

Moteurs / Critères	Temps d'indexation	Taille de l'index	Temps de la recherche	Pertinence	Total
Lucene	4	1	2	1	8
Zettair	1	2	4	3	10
Sphinx	2	3	1	4	10
Sqlite	3	5	5	5	18
Xapian	5	4	3	2	14

TABLE 5.3 – Autre classement des moteurs de recherche par critères

D'après ces résultats, en plus de ceux fournis sur le web (en particulier le nombre de réponses dans les résultats des moteurs de recherche) et recueillis sur le terrain, l'auteur de cet article préconise le moteur Lucene.

L'échantillon de données choisi par les auteurs de ces deux articles ainsi que le type de test n'étant pas les mêmes, il est très difficile d'opter pour un choix définitif. Les deux critères que sont le temps d'indexation et la taille de l'index jouent un rôle peu significatif dans notre projet. Au final, seul compte réellement le temps de la recherche et la pertinence, et au vu des résultats qui sont tous de l'ordre de la milliseconde pour le temps de recherche et un score assez proche au niveau de la pertinence, tous ces moteurs sont à égalité au niveau de ces critères.

Nous en concluons que le moteur Lucene de la fondation Apache semble plus favorable à notre projet, car il dispose de nombreuses fonctionnalités, d'une large communauté de développeurs ainsi que d'une importante documentation. La solution cloud de Lucene Solr n'a pas été sélectionnée de par notre configuration matérielle limitée à un seul serveur (cf. chap 3.1.1).

5.4 Les interfaces de visualisation

Nous présentons ici un bref état de l'art sur la visualisation d'informations pour les bibliothèques numériques. Il existe de très nombreuses interfaces de visualisation. En effet, le nombre d'ouvrages et de conférence sur ce sujet (IEEE VIS²¹) ainsi que le nombre de « toolkits » libres en ligne ont régulièrement augmenté²².

21. <http://ieevis.org>

22. <http://www.computerworld.com/article/2506820/business-intelligence/business-intelligence-chart-and-image-gallery-30-free-tools-for-data-visualization-and-analysis.html>, <http://philogb.github.io/jit/demos.html>, <http://d3js.org>

5.4. LES INTERFACES DE VISUALISATION

D'après [FL06, p. 19], les solutions de visualisation de collection sont de deux types : la visualisation d'une collection de documents et intra-document. Les auteurs de l'article mettent en avant l'importance capitale de la structuration de la collection pour la visualisation.

Il existe par exemple le système DocCube de [MCDA03], il regroupe les documents du fonds de manière structurée dans des sphères de taille variable, réparties dans l'espace suivant des axes selon certains critères.

Un autre auteur [CTT98] représente tous les documents d'une collection au travers d'un environnement 3D suivant plusieurs métaphores de visualisation (voir fig. 5.1). La collection est placée en arrière plan et l'utilisateur sélectionne un ouvrage qui se place automatiquement au premier plan. Selon le choix de l'utilisateur, la collection peut être représentée sous forme d'étagère aplatie, cylindrique, conique, sphérique... Cette dernière représentation combine à la fois les deux types de visualisation énoncés plus haut.



FIGURE 5.1 – La métaphore du cockpit de [Top02, p. 175]

D'autres représentations appelées « focus-contexte » présentent les informations d'un point donné de manière précise avec d'autres informations en contexte mais sous forme dégradée. C'est le cas notamment de la métaphore tridimensionnelle du « *Perspective Wall* » de [MRC91], ces auteurs déforment un mur en perspective afin de mettre en avant les informations principales dans la partie centrale tout en conservant les informations secondaires sur les côtés mais sous une échelle réduite.

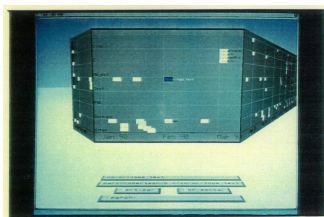


FIGURE 5.2 – Perspective Wall de [MRC91, Plate 3]

5.4. LES INTERFACES DE VISUALISATION

Il existe aussi des représentations permettant de mettre en valeur une structuration implicite de la collection comme par exemple les solutions sous forme de treillis ou en hiérarchies. Les travaux de [Tol96] et de [LRP95] ont étudié ces domaines, ils ont implémenté des systèmes de vue sous forme de nœuds avec des liens ou sous forme hyperbolique.

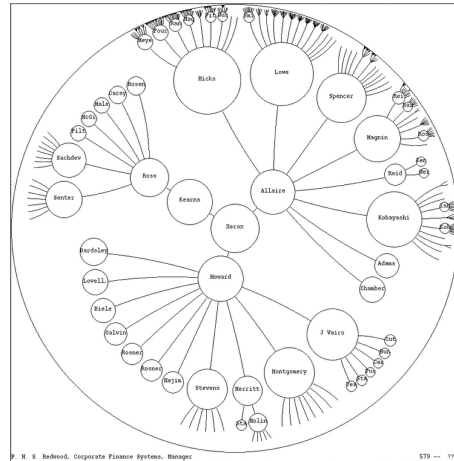


FIGURE 5.3 – Régions avec des nœuds de [LRP95, Figure 7]

Nous ne nous attardons pas plus longtemps sur ce sous-chapitre et nous passons au chapitre de la refonte du site.

Chapitre 6

Refonte du site

6.1 Design participatif

Depuis septembre 2010, un groupe de travail composé tour à tour de sept bibliothécaires : Caroline Aubert (arrivé en 2013), Françoise Chevalier, Florence Desnoyers, Nathalie Hû, Emmanuelle Minault-Richomme, Karine Raczinski, Mireille Le Van Ho, deux stagiaires : Ronan German et Emilie Trompille, deux informaticiens¹ (un professeur et un ingénieur d'études) M. Pierre Cubaud et moi-même se réunit deux heures par mois environ pour concevoir le wireframe du site. La participation de toutes ces personnes issues d'horizons professionnels différents apporte une incontestable plus-value à ces sessions de travail.

Nous avons organisé nos activités sur ces différents blocs IHM : page accueil, catalogue, synoptique, visualiseur ou feuilleteur et enfin recherche et résultat. Un premier groupe de discussion s'est constitué autour de la page d'accueil, puis un deuxième sur la page catalogue, synoptique et feuilleteur. L'analyse et la conception a suivi un fil conducteur car nous souhaitons une cohérence globale du site. En effet, tous les éléments doivent s'accorder et non se contredire tant au niveau de l'architecture, de la hiérarchie, du message, des couleurs, du contenu, de la lisibilité, de la typographie...

Dès le départ, nous souhaitons faire appel aux services d'un graphiste professionnel car c'est une compétence indispensable qui manquait cruellement à notre équipe. Mais par manque de budget, ce travail a été relégué provisoirement vers des ressources en interne. C'est ainsi que durant l'année 2011, deux bibliothécaires ont pris en charge le design graphique de la page d'accueil.

A partir de l'année 2012, la bibliothèque centrale a lancé un appel d'offre avec mise en concurrence de plusieurs prestataires. Le marché a été remporté par l'agence web

1. Il faut aussi compter parmi les informaticiens M. Rodrigo Almeida, docteur en informatique au laboratoire Cédric qui est intervenu quelques temps au démarrage du projet.

Eclydre² qui répondait à notre besoin concernant le web design et l'intégration. Cette équipe est composée d'un chef de projet, de deux intégrateurs Web (dont un expert Accessiweb) et d'un Webdesigner.

6.1.1 Parcours de navigation

Nous sommes passés par plusieurs phases de modélisation. Nous avons d'abord réfléchi à la pertinence de conserver la structure actuelle à trois niveaux (page catalogue, synoptique et feuilleteur). Sachant que pour chacun d'entre eux, il fallait mettre en place des éléments de navigation permettant de basculer facilement de l'un à l'autre. Le système à fil d'Ariane s'est alors imposé naturellement.

Au départ, nous avons souhaité un accès direct à l'ouvrage par la page catalogue sans passer par la page synoptique. Mais selon un utilisateur, cela ne semble pas répondre à une demande des lecteurs et serait source de complications, rendant par là-même le site beaucoup moins intuitif. L'accès direct supposerait que l'on supprime l'interface synoptique au profit d'une interface à « tableau de bord » où toutes les fonctionnalités seraient intégrées dans la même page comme c'est le cas actuellement sur Gallica. Or, nous partons du principe de la limitation des choix de parcours pour éviter que l'internaute ne s'égaré dans le site.

Finalement, la solution à « tableau de bord » n'a pas été retenue. En effet, elle n'est pas pertinente car la longueur des titres de chapitres ainsi que celle des tables sont extrêmement variables d'un ouvrage à l'autre. Nous allons perdre en lisibilité car nous ne pouvons pas exploiter toute la largeur de l'interface. Des lignes trop longues ou trop courtes induisent une fatigue et une gêne de l'œil du lecteur. Les auteurs de [NB05] définissent une longueur maximale de 70 caractères. Au delà de cette limite survient un problème de ralentissement de la lecture et de compréhension. L'œil a du mal à se repositionner sur la ligne suivante. Cette valeur dépend aussi de la typographie, du type de police et de la taille, de la ligne et de l'espacement des caractères ainsi que des différentes structures de l'écran, comme la variation du nombre de colonnes et de leurs largeurs.

De plus, nous avons souhaité réduire le nombre de choix des fonctionnalités sur chacune des interfaces afin de réduire le temps de décision de l'utilisateur d'après la loi de Hick [Hic52].

Nous présentons un schéma de l'architecture du site. Les flèches indiquent la possibilité de passer d'une interface à l'autre.

2. <http://www.eclydre.fr>

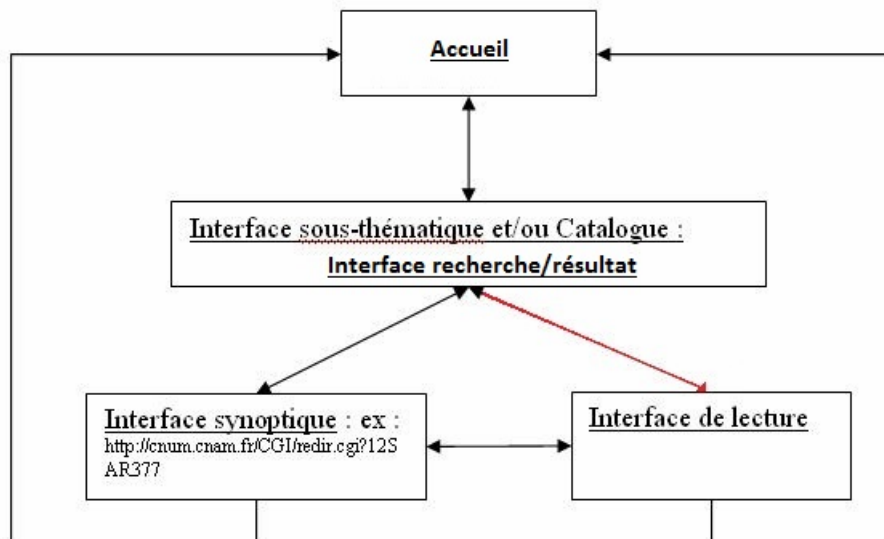


FIGURE 6.1 – Architecture du site

Le fil d'Ariane. Nous courrons le risque de perdre le principe statique du site car la restructuration en sous-thématiques induit plusieurs chemins possibles vers le document final. La mise en place d'un système de gestion de sessions (sans authentification) a été nécessaire pour mémoriser les préférences de l'utilisateur. Une autre solution pourrait consister à créer les mêmes pages statiques pour chacune des sous thématiques, nous obligeant à programmer des liens différents vers le même ouvrage. Or, cela entrerait en contradiction avec le choix d'une adresse pérenne unique pour chaque ouvrage (ex. <http://cnum.cnam.fr/redirect?P5200>), sans oublier que la génération des pages aurait demandé un temps plus long.

Le titre court du fil d'Ariane est préféré à celui de l'identifiant de l'ouvrage. Ainsi, l'utilisateur aura une meilleure mémorisation du contexte par le titre court (colonne « TC » dans la base de données) que par l'identifiant de l'ouvrage (colonne « ID »). Mais encore faudrait-il définir un nombre maximum de caractères pour ne pas déborder de l'écran. Finalement, nous avons opté pour une taille maximum de 90 caractères. Il sera automatiquement coupé et suivi de trois points dans le cas d'une taille supérieure.

Un codage particulier du fil d'Ariane a été appliqué sur les titres de « quanta virtuels », le numéro du tome n'apparaissant pas au final. Nous avons également rendu tous les liens du fil interactifs à part le dernier.

Quant à l'emplacement de ce fil, l'article [RC03] préconise de le positionner au niveau des yeux et à proximité des autres liens. Mais d'autres auteurs remettent en cause cette fonctionnalité, car déjà présente différemment dans le navigateur avec la fonction « page précédente ou suivante » et l'historique.

6.2 Mise en œuvre

6.2.1 Planning

De notre première réunion datant du 24/09/2010 jusqu'à la dernière du 24/01/2014, nous avons organisé cinquante-deux réunions plus huit avec les prestataires graphiques, auxquelles s'ajoute une réunion pour l'établissement d'une grille d'évaluation technique et financière, et deux autres pour l'évaluation des candidats. Des réunions supplémentaires ont été organisées entre les bibliothécaires pour l'harmonisation et l'enrichissement des notices de données bibliographiques.

Au final, le nouveau site Cnum 2 a été officiellement mis en ligne le 10/06/2014.

Voir le diagramme de planification Gantt [A.1](#) et [A.2](#) en annexe.

6.2.2 Point de vue informatique

Le sous-domaine `cnumdev` étant visible uniquement en interne (dans l'enceinte de la bibliothèque), il était essentiel de travailler sur un autre qui soit visible de l'extérieur (mise en place d'un extranet). Il serait ainsi possible d'une part, de valider le code html par le validateur du W3C, et d'autre part, d'accéder au site depuis l'extérieur, en particulier par l'intégrateur et le graphiste. L'environnement de développement du « staging » a ainsi été placé sous un nouveau sous-domaine que nous avons appelé `cnum2.cnam.fr`.

La conception du site s'est déroulée en deux phases : développement du contenu et présentation à l'équipe. Nous avons développé le site en nous attachant d'abord aux aspects architecturaux, organisationnels et fonctionnels. La demande initiale du groupe de travail était de développer le site sans différenciation typographique (style, corps et graisse identique) avec comme seules couleurs du noir et du blanc. De cette manière, le graphiste ne serait pas influencé par tous ces éléments perturbateurs.

Conservation des urls existants. Les bibliothécaires ont souhaité que les adresses actuelles du Cnum soient conservées car elles sont enregistrées sur plusieurs sites et bases catalographiques : SIGB de la bibliothèque centrale, bases du musée³ en particulier des dossiers numériques (ex. du Pendule de Foucault⁴), Sudo, Calames⁵, exposition virtuelle du Cnum. Une mise à jour de ces données leur aurait demandé un travail supplémentaire non négligeable, car le remplacement de toutes ces urls est difficilement réalisable dans les conditions actuelles, pour ne pas dire impossible.

Il en est de même des liens enregistrés sur d'autres sites. Le moteur de recherche Google propose dans ses fonctionnalités une requête de recherche spécifique sur les pages reliées à une URL donnée, par exemple la requête « [linked:http://cnum.cnam.fr/CGI/](#) » donne 3 500 réponses au 19/11/2012 correspondant à des sites de wikis, forums, blogs, réseaux sociaux... (cf. tableau 4.8) Or, tous les internautes redirigés vers notre site par ces liens représentent une part importante de notre trafic.

Sauvegarde - SVN. Le risque de perte des données de codage correspondant à des heures de développement étant bien réel, nous nous prémunissons de ce problème par des sauvegardes régulières d'abord en local (outil Auto save) puis à distance (SVN).

3. <http://catalogue.cnam38.bibliobre.com>

4. <http://catalogue.cnam38.bibliobre.com/cgi-bin/koha/opac-detail.pl?biblionumber=17884>

5. <http://www.calames.abes.fr/pub/cnam.aspx#details?id=FileId-424>

Sécurité. L’affichage du site, en partie en statique, nous protège dans une certaine mesure des attaques informatiques comme l’injection Sql.

L’accès aux différentes pages descriptives de l’ouvrage s’établit au moyen d’un identifiant qui est vérifié par un script. En cas de non validité, l’internaute est redirigé automatiquement vers la page d’erreur.

L’accès en extranet à travers le protocole « http » et « ssh » est protégé par le pare feu de la DSI.

Codage - développement. Nous avons étendu les scripts de génération et d’accès aux pages de Cnum 1 codés dans le langage Perl. Il y a eu un important travail de création de fonctions et de factorisation du code PHP/Perl pour optimiser le développement. Nous avons veillé à ce que les informations bibliographiques des pages web soient exportables au format Zotero. Les balises « Dublin Core metadata », élément meta de l’entête html head ont été renseignées à cet effet : author, keywords, language, creator, date. . .

Nous avons utilisé également les fonctions Perl et Php de conversion des caractères texte en entités html (« use HTML : :Entities » et « htmlentities ») pour nous conformer au langage Xhtml et au nouveau jeu d’encodage des caractères du charset UTF-8.

Une des demandes du cahier des charges était de conserver la propriété statique du site. Les pages devaient être affichées le « moins dynamiquement » possible, alors nous nous sommes limités dans le codage à de simples inclusions de fichiers.

Choix du langage d’intégration web. Une question importante concernait le choix du langage de développement des pages : Flash, Xhtml ou HTML5 ? Il s’avère que le langage Flash pose encore des problèmes d’indexation et de référencement même s’il s’est un peu amélioré. Quant au langage HTML5, il présente l’avantage d’associer des métadonnées à des éléments HTML mais tous les navigateurs n’implémentent pas encore leurs nouvelles fonctionnalités. Donc, nous avons finalement choisi le langage de balisage Xhtml successeur d’html. Celui-ci permet d’associer une structure logique au document avec une meilleure qualité et un affichage rapide du code. On y a intégré une feuille de style CSS3 servant à dissocier le fond de la forme.

Accessibilité et S.E.O. ⁶ La question importante de l’accessibilité et du positionnement des pages par les moteurs de recherche a été gérée en partie par le prestataire graphique qui disposait dans son équipe d’une expertise en AccessiWeb en Evaluation (AEA) et optimisation du référencement (Ranking Metrics).

6. en anglais : Search engine optimization ou Optimisation pour les moteurs de recherche

6.2.3 Modélisation conceptuelle de la base de données

Le nouveau site du Cnum et les nouveaux besoins demandent une redéfinition du modèle conceptuel de la base de données.

6.2.3.1 Refonte de la base de données

Les bibliothécaires ont souhaité intégrer les tables des matières et des planches dans l'entrepôt OAI. Or, ce type de service ne peut être fourni avec la modélisation actuelle, les notices des volumes de série et de périodiques n'étant incluses que partiellement dans l'OAI par le champ « dc :relation » de la notice mère. L'association des tables des matières avec les notices n'est pas réalisable dans ce cas de figure, idem dans le cas des « quanta virtuels ».

Les deux autres problèmes sont situés au niveau du rattachement des notices aux thématiques (éventuellement les sous thématiques) et des traitements informatisés des données qui demandent la reformulation des requêtes conditionnelles sur les deux tables existantes.

Nos premières discussions sur ce sujet se sont d'abord orientées sur le changement conceptuel des tables `dei_chapeau` et `dei_quantum` et un basculement des enregistrements vers les tables suivantes :

- unité de catalogage (UC), entité de plus haut niveau : notices de périodiques, grandes monographies et « quanta virtuels »
- unité documentaire (UD), entité de plus bas niveau : notices de monographies, volumes de périodiques, tomes de grandes monographies et volume de « quanta virtuels »

Mais ce nouveau modèle n'allait pas sans un important travail de modification de tous les scripts de l'application de front et de back office. L'idée nous est alors venue de conserver cette modélisation par le biais des vues. La création de ces nouvelles tables virtuelles résout parfaitement tous les problèmes mentionnés. Par là même, il était alors possible de rajouter la table supplémentaire unité de catalogage et documentaire (UCD) constituée de l'union des tables UC et UD et ainsi rattacher le numéro d'image et le titre de chapitre de la tables des matières et illustrations.

Les deux tables virtuelles `Cle_lieu` et `Cle_nom` permettent de répéter autant de fois le lieu d'édition et le le nom d'éditeur dans l'interface catalogue (cf. 6.3.3).

Finalement, nous avons conservé la structure actuelle et nous avons rajouté dans la base de données les tables suivantes : `thematiques`, `typologies`, `appartenir_thematiques`, `appartenir_typologies` et les attributs suivants dans la table `dei_chapeau` : `largeur`, `hauteur`, `cle_nom`, `cle_date`, `cle_lieu`, `cle_titre`, `sous_droit`, `pdf_ocr_bool` et `ID_note_presentation`. Nous avons fait de même sur la table `dei_quantum` avec les attributs `largeur`, `hauteur` et `cle_date`. Ce dernier attribut

provient d'un courriel d'un usager ne trouvant pas les informations qu'ils recherchaient. Les colonnes « `cle_titre` » (est un indicateur de saut de caractère en catalogage) et « `cle_date` » sont respectivement destinées à des fins de classement alphabétique et chronologique.

Certaines données comme la largeur et la hauteur du livre sont manquantes et le travail rétrospectif de cette description bibliographique est difficilement réalisable par l'équipe. Malgré ce problème, nous avons tenté de remplir ce champ de manière semi-automatisée en récupérant la hauteur par déconcaténation du champ collation (CO) et avec cette donnée, en déduire la largeur par comparaison entre la résolution et la taille de l'image en pixel. Hélas, les tests n'ont pas été très concluants (voir le chapitre 6.3.3)

Voir en annexe le modèle conceptuel de données H.1 et les commandes phpMyAdmin de création des vues I.

6.2.4 Moteur de recherche Solr

Le développement du moteur de recherche a représenté une phase importante du projet, elle s'est étalée sur une période d'environ cinq mois.

6.2.4.1 Installation et configuration de Solr

Nous avons déclaré dans le fichier `default-server.conf` de Linux, la directive « *Proxy-Pass /solr http ://localhost :8983/* », pour protéger et faciliter l'accès en interne à la console d'administration de Solr. Après téléchargement et décompactage du logiciel Solr (avec son serveur Jetty intégré) via son site⁷, nous avons lancé la machine virtuelle Java.

6.2.4.2 Fichiers de configuration du noyau

Le noyau représente les données enregistrées dans des champs spécifiques avec les fichiers de déclaration des types de données (`schema.xml`) et de configuration (`solr-config.xml`) pour les indexer.

Nous avons défini deux noyaux identiques `core_notices1` et `core_notices2`. Nous pouvons ainsi, suivant la mise à jour de l'index, basculer de manière transparente de l'un à l'autre sans interrompre ce service à l'utilisateur.

Fichier `schema.xml`.

Nous détaillons ici les principaux types de données des champs avec les filtres Solr que nous avons sélectionnés.

7. <http://lucene.apache.org/solr>

6.2. MISE EN ŒUVRE

```
<!-- Types de base -->
<fieldType name="string" class="solr.StrField" sortMissingLast="true" omit-
Norms="true"/>

<!-- Types numériques nombre entier -->
<fieldType name="tint" class="solr.TrieIntField" precisionStep="8" omit-
Norms="true" positionIncrementGap="0"/>

<!-- Types numériques optimisés pour les requêtes sur des dates : cle_date,
date_mise et dateoai -->
<fieldType name="tdate" class="solr.TrieDateField" precisionStep="6" omit-
Norms="true" positionIncrementGap="0"/>

<!-- Type texte pour l'affichage avec très peu de traitements -->
<fieldType name="rawText" class="solr.TextField">
<analyzer>
<charFilter class="solr.HTMLStripCharFilterFactory"/>
<tokenizer class="solr.KeywordTokenizerFactory"/>
</analyzer>
</fieldType>

<!-- Traitement des notices -->
<fieldType name="stemmedFrenchText" class="solr.TextField" positionIncre-
mentGap="100">
<analyzer>

<!-- suppression d'éventuelles balises HTML-->
<filter class="solr.LowerCaseFilterFactory"/>

<!-- gestion des synonymes comme microscope et mikroskop en allemand --
>
<filter class="solr.SynonymFilterFactory" synonyms="synonyms.txt" ignore-
Case="true" expand="true"/>

<!-- suppression des mots classiques comme les pronoms personnels : je, tu,
avec -->
<filter class="solr.StopFilterFactory" ignoreCase="true"
words="stopwords_fr.txt" enablePositionIncrements="true"/>

<!-- lemmatisation (pluriels,...) -->
<filter class="solr.SnowballPorterFilterFactory" language="French"
protected="protwords_fr.txt"/>
```

```
<!-- suppression des doublons éventuels -->
<filter class="solr.RemoveDuplicatesTokenFilterFactory"/>

<!-- suppression des élisions (l', qu',...) -->
<filter class="solr.ElisionFilterFactory" articles="elisionwords.txt"/>

<!-- normalisation des accents, cédilles, les ligatures comme l'e dans l'o,...-
>
<filter class="solr.ASCIIFoldingFilterFactory"/>

</analyze>
</fieldType>
```

La commande « `copyField` » de Solr nous permet de copier tous les champs de la notice dans un champ réservé à la recherche par défaut (`stAllFr` et `stAllFrStrict`).

Nous avons aussi défini un type spécial pour les tables, dont les spécifications sont identiques à celles listées ci-dessus avec en plus, une non indexation du caractère « `##00000000+numéro image` » correspondant au séparateur de champ dans les titres de chapitre des tables. Deux autres types ont été rajoutés pour la recherche exacte sur les notices et les tables, par contre ils ne contiennent aucun traitement linguistique (lemmatisation, racinisation...) ou sur la casse typographique.

Fichier `solrconfig.xml`. Le plugin « *ExtendedDismaxQParserPlugin* » ou « *edismax* » nous permet d'appliquer un poids spécifique à certains champs. Nous avons configuré l'algorithme avec un coefficient élevé sur le champ titre de l'ouvrage puis le nom d'auteur : « `stTI^ 5.0 stAU^ 2.0` ».

6.2.4.3 Génération et indexation des documents Solr

Grâce à un script codé en langage Perl, nous convertissons les tables des matières et illustrations du format Non-ISO, ASCII et ISO-8859 au format UTF-8. Nous remplaçons automatiquement quelques caractères spéciaux comme les tirets longs ou cadratins, guillemets français ouvrants et fermants, apostrophes curvilignes, et certaines ligatures par le caractère adéquat. Ensuite, un deuxième script intègre ces données avec les notices de la base de données Mysql dans un fichier xml, vu comme un document dans le moteur Solr. Chaque titre de chapitre est indexé à la fin de la ligne par le caractère séparateur de champ « `##00000000` », suivi du numéro image à quatre chiffres pour conserver le renvoi vers le fac-similé.

6.2. MISE EN ŒUVRE

Exemple de structure xml correspondant à l'identifiant de l'ouvrage M1789⁸ :

```
< ?xml version="1.0" encoding="UTF-8" ?>
<add>
<doc>
< !- Identifiant ->
<field name="id">M1789</field>
< !- Type de l'ouvrage (revue, monographie ou volume) ->
<field name="rwST">PMONO</field>
< !- Auteur ->
<field name="rwAU">Compagnie des freins et signaux Westinghouse</field>
< !- Auteur catalogue ->
<field name="rwAC">Compagnie des freins et signaux Westinghouse</field>

< !- <field name="rwChiffreIncreVolume">0000</field> ->

< !- Titre ->
<field name="rwTI">Triple valves types
Lu-V et Lu-R : notice n° 1521-34</field>
< !- Titre court ->
<field name="rwTC">Triple valves types Lu-V et Lu-R</field>
< !- Mots-clés ->
<field name="rwSU">Freins - France ; Transports ferroviaires -
Appareils et matériel</field>

< !- Nom de l'éditeur ->
<field name="rwCle_nom">Marc All</field>
< !- Lieu d'édition ->
<field name="rwCle_lieu">Paris</field>
< !- Cote ->
<field name="rwCT">CNAM-MUSEE TR0.5-WES</field>

< !- Date d'édition ->
<field name="cle_date">1939-01-01T00 :00 :00Z</field>
< !- Date de mise en ligne ->
<field name="date_mise">2015-01-20T00 :00 :00Z</field>
< !- Date de mise en ligne pour l'OAI ->
<field name="dateoai">2015-01-20T00 :00 :00Z</field>

< !- Nombre de volumes ->
<field name="rwNQ">1</field>
```

8. <http://cnum.cnam.fr/CGI/redir.cgi?M1789>

6.2. MISE EN ŒUVRE

```
<!-- Numéro de la thématique (chiffre compris entre 1 et 9) -->
<field name="numero_thematiques">7</field>
<field name="numero_thematiques">9</field>

<!-- Numéro de la typologie (chiffre compris entre 1 et 3) -->
<field name="numero_typologies">2</field>

<!-- Numéro de la langue (chiffre compris entre 1 et 4) -->
<field name="numero_langues">1</field>

<!-- Table des matières -->
<field name="rwTdm">Triples valves types
Lu-V et Lu-R Westinghouse $#000000000002</field>
<field name="rwTdm">Triples valves types
Lu-V et Lu-R Westinghouse : fonctionnement $#000000000003</field>
<field name="rwTdm">Triples valves types
Lu-V et Lu-R Westinghouse : entretien $#000000000006</field>

<!-- Table des illustrations -->
<field name="rwTdp">Planche 1 - Triples valves types
Lu-V et Lu-R Westinghouse $#000000000004</field>
<field name="rwTdp">Planche 2 - Triples valves types
Lu-V et Lu-R Westinghouse $#000000000007</field>

</doc>
</add>
```

La suppression et l'indexation des documents Solr s'effectue via les commandes Shell Unix suivantes :

```
« url=http ://localhost :8983/solr/core_notices1/update
chemin_des_fichiers_xml=/home/solr/xml/core_notices1/conf
chemin_du_noyau=/home/solr/core/core_notices1/docs
java -Ddata=args -Dcommit=yes -Durl=$url -jar $chemin_du_noyau/post.jar
"<delete><query>id :*</query></delete>"
java -Durl=$url -jar $chemin_du_noyau/post.jar $chemin_des_fichiers_xml/*.xml »
```

Au niveau de l'interface de recherche, le lancement des requêtes vers le moteur Solr à partir d'une page codée en langage Php nous est facilité par la librairie open source Apache Solr PHP Client⁹.

9. <https://code.google.com/p/solr-php-client>

6.3 Prototypes

6.3.1 Généralités sur l'ensemble des pages

Nous avons appliqué sur l'ensemble des pages du site quelques règles de bon sens ergonomiques comme :

- éviter les pages culs-de-sac, c'est-à-dire celles ne menant nulle part sans liens de rebond sur d'autres pages du site
- intégrer des infobulles permettant de guider rapidement l'utilisateur et le dispenser de la lecture fastidieuse d'un long texte explicatif en ligne
- proposer un titre de page pertinent (balise *title* du code source html). Il permet de repérer rapidement dans son historique de navigation les titres que l'on a consultés. On y a inclus le titre court de l'ouvrage et le numéro de page.
- proposer une organisation identique sur toutes les pages
- placer les éléments principaux au dessus de la ligne de flottaison
- afficher si possible toutes les informations sur la même page
- placer une barre supérieure principale présente sur toutes les pages. Elle contiendra les éléments principaux permettant à l'utilisateur de rebondir sur d'autres pages du site.
- personnaliser le favicon
- personnaliser la page de gestion des cas d'erreurs
- éviter la multiplicité des barres de défilement internes car elles incommode terriblement la navigation
- rester simple dans la conception, ce qui est loin d'être une tâche aisée
- éviter la surcharge d'informations
- supprimer tous les éléments superflus
- suivre les conventions d'usage de ce qui se fait fréquemment sur le web. Le respect de certaines pratiques connues des internautes facilitent grandement leur navigation.

Nous avons placé et verrouillé horizontalement le menu de navigation en haut de toutes les pages du site (il était initialement placé à gauche dans Cnum 1). Cela présente l'avantage d'afficher en permanence toutes les informations de navigation, même lors d'un défilement vertical, et d'offrir à l'internaute la possibilité de rebondir rapidement sur les autres pages principales du site. Le verrouillage de ce menu implique de manière logique une limitation de sa hauteur pour laisser un maximum de place disponible et visible en dessous.

Ce menu supérieur verrouillé contenant un formulaire de recherche est en même temps conforme aux critères d'accessibilité. En effet, ce champ sera accessible facilement aux utilisateurs naviguant avec le seul périphérique du clavier.

De même, une autre recommandation sur l'accessibilité concerne l'intégration d'un outil d'agrandissement de la police de caractère. Or, le redimensionnement entraîne de nombreuses contraintes comme le masquage d'information et les superpositions de conteneurs. Nous avons préféré exploiter les nombreuses fonctionnalités déjà présentes par défaut dans la majorité des navigateurs web et des claviers. Ainsi, nous n'avons pas intégré le bouton d'agrandissement de cette police et le suivi automatique de retour en haut de page. Ils sont remplacés par les boutons « zoomer » et « dézoomer » du navigateur et les touches claviers « début » ou « origine ».

La question s'est posée quant à l'exploitation de toute la largeur de la fenêtre du navigateur. Ce genre d'implémentation, appelée communément site élastique, était le parti pris du premier site Cnum. Ainsi, nous l'avons conservé sur l'ensemble des pages mais sur les conseils du prestataire Eclydre, nous ne l'avons appliqué au final que sur la seule interface feuilleter.

6.3.2 Page d'accueil

La page d'accueil représente la porte d'entrée du site, elle doit guider les visiteurs vers les pages qui les intéressent. Sa conception plus ou moins bonne incite ou non l'internaute à poursuivre sa consultation et aussi à y revenir par la suite, ce qui en fait la page la plus importante du site.

De cette page découlent également tous les principaux éléments graphiques de l'ensemble du site, à la construction duquel elle sert de base et de structure.

Ainsi, nous avons pris soin de bien la hiérarchiser, en affichant d'abord les informations les plus importantes tout en haut de page et en reléguant progressivement le reste vers le bas.

On y trouve les logos du Cnam et du Cnum, une présentation succincte du site, des indications chiffrées, un formulaire de recherche simple, un lien vers une recherche avancée, les dernières mises en ligne, le catalogue général et les différentes thématiques et typologies. Nous avons placé un encart spécial à droite destiné aux expositions virtuelles, réseaux sociaux et nouveautés. En pied de page, nous trouvons différents liens tels que l'adresse du site de la bibliothèque centrale, les coordonnées de contact, les différents partenaires du Cnam, les conditions d'utilisation, les flux rss...

Nous veillons à ce que les noms des thématiques et des catégories soient visibles entièrement sur la page d'accueil. En particulier, nous limitons le défilement d'écran ou tout autre fonction d'ouverture d'onglet pour que l'internaute puisse visualiser au premier coup d'œil toute l'information à l'écran. Ceci risque de poser problème à l'avenir si d'autres thématiques émergent.

Le premier prototype de la page d'accueil a été conçu et développé initialement par un membre du groupe de travail. Nous l'avons repris, corrigé et converti en un code

XHTML valide.

A chaque typologie et thématique est associée une représentation iconographique. Il était aussi important de donner une idée du nombre de titres de chacune des thématiques, typologies et du catalogue général. Cependant, dans cette même optique, une autre question nous interpelle, à savoir le nombre d'images de ces regroupements, car le nombre restreint de titres comme celui par exemple de la typologie « Revues » par rapport à celles des « Ouvrages » n'est pas significatif de la différence entre leur nombre d'images. Comme ces chiffres l'attestent, au 21/12/2012, il y a 352 278 images et 821 titres de petites monographies plus 349 titres de grandes monographies pour les ouvrages, contre 387 777 images et 15 titres pour les revues. Cependant, ce calcul tient compte des tomes des grandes monographies comme étant de simples monographies.

Le choix de la représentation iconographique de la thématique peut s'implémenter de deux manières différentes : soit en visualisation mode « macro », où l'image remplit la totalité de l'espace de la vignette, soit en mode « micro », où l'on voit une représentation miniaturisée d'un élément. Dans le cas particulier de la vignette « manuscrits et dessins », une présentation combinant à la fois un dessin et une écriture manuscrite seraient des plus pertinents à cet affichage.

La vignette nouveauté posait un problème sur la cible du lien hypertexte. En effet, comment afficher la liste des dernières mises en ligne ? Dans ce cas, nous pouvions présenter une nouvelle page identique à celle des pages catalogues mais en y ajoutant une nouvelle colonne « date de mise en ligne ». Ce système, bien que judicieux n'est pas cohérent avec le reste des pages qui ne présentent pas cette information.

Au final, le graphiste nous a conseillé d'insérer deux hyperliens, dont un rebondissant sur la page de résultats de la recherche des dernières mises en ligne et l'autre sur une page listant tous les lots mis en ligne successivement.

La disposition des thématiques peut être présentée de plusieurs manières : nous pouvons exploiter les treemaps, newsmap¹⁰, nuages de mots-clés¹¹, histogramme, barres, secteurs ou camemberts, radars, sachant que la représentation à double axe nous donnera une idée à la fois de cette double volumétrie que constituent le nombre de titres et le nombre d'images.

Ces différentes représentations sont certes beaucoup moins esthétiques que la représentation iconographique mais présentent l'avantage d'une appréhension rapide et exhaustive du fonds. Elles sont décrites dans le chapitre 6.5.

Nous avons effectué quelques essais avec un fond de page (Code css : body

10. <http://newsmap.jp>

11. <http://www.wordle.net>

background-image :url('fond-de-page.jpg');). Ce fond présente l'avantage de réduire la charge cognitive du lecteur, ce qui évite d'insérer des cadres autour des images. En revanche, son inconvénient majeur est de ralentir le chargement de la page, si la taille de l'image est assez élevée, et de consommer plus d'encre des cartouches d'imprimante. Mais l'acte d'impression d'une page d'accueil d'un site n'est peut-être pas une pratique courante.

Cf. les figures C.1 et C.2.

6.3.3 Page catalogue

Certains utilisateurs se sont interrogés sur l'utilité d'un catalogue (considéré plutôt comme une liste sous l'angle bibliothéconomique) par rapport à un moteur de recherche. En effet, de par la taille grandissante du corpus, la navigation dans cette longue page pourrait s'avérer fastidieuse. Ces utilisateurs vont même jusqu'à dire que la mise en place du catalogue était justifiée pour compenser la défaillance du temps allongé de l'affichage des résultats du moteur de recherche de Cnum 1. Mais cela a été contredit par un autre utilisateur, qui nous a confié que la conception de cette page était liée essentiellement à une demande des internautes.

La navigation page à page sur des portions plus réduites n'a pas été retenue, d'ailleurs le problème engendré par cette unique et longue liste a été résolu en intégrant des liens ancrés dans le sous-menu supérieur.

Nous avons placé en haut de chaque page la même vignette iconographique que celle de la page d'accueil, pour en rappeler le contexte. On y a développé un contenu éditorial identique aux parcours thématiques proposés sur Gallica¹². Ainsi, les chercheurs en sciences et techniques du Cnam ont été sollicités pour rédiger quelques paragraphes de présentation des thématiques.

Certaines fonctionnalités qui ont été proposées sur le tard, servant en particulier au panier virtuel, n'ont pas été retenues. Par exemple, nous avons eu des suggestions d'affichage telles que des cases à cocher et des numéros d'ordre au niveau du catalogue et des pages de résultats du moteur de recherche. Ce besoin n'ayant pas été relevé dans l'enquête d'usage peut cependant être comblé via le plugin Zotéro.

Nous avons tenté d'intégrer la vignette de la page de titre des ouvrages dans la page catalogue, mais celle-ci n'est pas considérée par un utilisateur comme un indice visuel suffisant pour appréhender un ouvrage. Il y a d'autres paramètres tout aussi importants, comme les dimensions (largeur, hauteur, profondeur), le format (in-plano, in-folio, in-4o, in-8o, in-16...), l'ampleur de l'ouvrage (nombre de pages) et la

12. <http://gallica.bnf.fr/FromHomeToDossiers?lang=FR>

volumétrie (nombre de volumes auxquels il est rattaché, cas notamment des séries et des périodiques). Dans cette perspective, nous avons mené deux tentatives de prototypage :

Première tentative. Une heuristique visuelle avec un système de bâtonnets représentant le nombre de volumes a été proposée sur d’anciens prototypes du Cnum développés par [Lau07]. Nous avons essayé d’élargir ce prototype pour appréhender les dimensions physiques du livre. Il était possible d’exploiter la largeur, la hauteur et la profondeur du livre afin de les comparer par leurs dimensions. Or, les tests ont montré qu’à une échelle réduite (dimension d’un écran d’ordinateur classique), le facteur de réduction ne permet pas de discriminer correctement les différentes dimensions de l’ouvrage. A la place, nous avons proposé un affichage classique de tri par nombre de volumes, décrit dans les paragraphes qui suivent.

Cf. la figure C.3.

Deuxième tentative. Les titres des ouvrages du catalogue sont classés par ordre alphabétique ou numérique (croissant) dans différents onglets par auteurs, dates, titres, lieu d’édition, nom d’éditeur, nombre d’images et nombre de volumes avec un accès aux titres par des liens ancrés de type alphanumérique.

Le basculement de tri par ordre croissant ou décroissant nous semble futile, en effet la lecture de l’alphabet se fait essentiellement de la lettre A vers la lettre Z. De même, les liens ancrés de début et de fin ne sont pas pertinents car le sens de la lecture se fait toujours de haut en bas.

Cf. la figure C.4.

6.3.3.1 Fonctionnalité de chaque onglet

Les liens ancrés ont été intégrés dans le bandeau verrouillé principal et non dans la partie contenant les titres des ouvrages comme c’est le cas sur la première interface « Gallica Classique »¹³.

Nous n’avons pas intégré d’élément graphique de retour en haut de page qui suit en permanence le défilement (implémentable avec la bibliothèque javascript jquery). Cette fonctionnalité nous paraît futile car elle pourrait perturber la lecture à l’écran et de plus, elle est déjà indirectement intégrée dans le menu supérieur verrouillé contenant les liens ancrés et dans les boutons claviers.

Tri par nom d’auteur. Le nom d’auteur des revues étant inexistant, il a été comblé par le terme « Revue ». En réalité, les volumes de revue ont parfois été rédigés par des auteurs différents, idem pour les articles de certains volumes. Ce classement

13. <http://gallicadossiers.bnf.fr/Classique>

n'est pas tout à fait correct car on fait d'abord un premier classement par type de document (revue et monographie) ensuite par auteurs, et de plus, les auteurs des volumes, et les auteurs secondaires (parfois plus de deux) n'apparaissent pas dans cet affichage. La non intégration de ces auteurs particuliers remet en cause la définition du catalogue et implique des changements dans la base de données. D'ailleurs, le choix de la présence de ces titres de revues sur cette page a fait débat.

L'intégration des auteurs secondaires était envisageable suivant trois solutions : auteur1 et auteur2 dans le même champ, ou dans des colonnes séparées, ou la répétition des titres autant de fois qu'il y a d'auteurs. Cela ne va pas sans une remodelisation de la base de données car les auteurs secondaires, contenant plus de deux auteurs, sont enregistrés dans la même colonne. De plus, l'ajout de colonnes supplémentaires impliquera peut-être un dépassement de la taille (fixe) de la fenêtre, et donc un défilement horizontal que cette interface ne facilite malheureusement pas.

Tri par titre et par date. Ces deux classements sont identiques au premier tri. Seuls, les liens ancrés du classement par date changent, il y a deux listes comprises dans des intervalles de siècle et de demi-siècle.

Les tranches de date ont été calquées sur celles choisies par le musée des arts et métiers et sont cohérentes avec l'histoire des techniques. Ces deux tranches offriront une navigation plus fine avec des sauts de cinquante et cent ans. On s'est rendu compte par la suite que les siècles débutent par les années finissant par « 01 » et non « 00 ». Malgré cela, le point d'ancrage a été volontairement interverti par respect bibliothéconomique.

Tri par lieu d'édition. Ce tri pouvait faire l'objet d'un test d'implémentation d'une interface cartographique indiquant les lieux géographiques comme c'est déjà le cas dans le plugin « Geolocation » de l'outil « open source » Omeka.

Tri par nom d'éditeur. Le nombre de caractères du nom de l'éditeur pouvant être assez long décalait l'affichage de la première colonne car sa largeur était celle de la longueur du nom d'éditeur le plus long. Nous avons résolu ce problème en découpant la phrase au bout de trente caractères mais sans couper les mots. Une autre solution aurait été de basculer cette fonctionnalité dans le moteur de recherche. Dans les deux précédents tris, les titres se répèteront autant de fois s'ils ont été édités par plusieurs éditeurs ou lieux différents.

Tri par nombre d'images et de volumes. Nous avons également mené des tests d'affichage sur d'autres systèmes de tri tel que le nombre d'images et le nombre de volumes. Ce genre de tri pouvait donner une idée de l'ampleur physique et bibliothéconomique du titre consulté. Mais cette idée n'a pas été retenue par le groupe de travail.

Vignettes. Les vignettes des pages de titre - initialement supprimées dans notre modèle - ont été rajoutées par la suite par le graphiste les jugeant pertinentes pour les personnes plus portées sur le côté visuel. Le graphiste pense que son emplacement est plus viable en dernière colonne plutôt qu'en première, contrairement à la plupart des sites Internet.

L'ajout de toutes les métadonnées sur chacune des pages de tri risquerait de les alourdir et influence grandement le graphisme.

6.3.4 Page synoptique

La page « synoptique » est un système présentant des perspectives de vue différente de l'ouvrage, elle contient les onglets suivants, placés graphiquement selon leurs ordres d'importance : notice bibliographique, tables des matières, tables des illustrations, vignettes, export du document et des informations bibliographiques, parfois note de présentation, page de recherche restreinte au document consulté. La séparation de la notice et de ses tables, initialement fusionnées dans Cnum 1 (cf. chapitre B.3), permettra d'augmenter la zone de lecture.

La barre supérieure qui est affichée en permanence (verrouillée avec le code css `position :fixed`) empêche une exploitation totale de l'espace de lecture. Le premier prototype implémentait une barre de défilement interne, qui était propre à un bloc et non à l'interface du navigateur. Nous avons conservé cette fonctionnalité de la barre verrouillée supérieure en supprimant simplement la barre de défilement interne.

La proposition d'insertion de la vignette de la page de titre consultée dans le coin supérieur gauche n'a pas été validée. Elle aurait permis de réduire le nombre de clics pour accéder à la page du feuilleté. Au final, le graphiste l'a placée dans la page catalogue.

Un système de gestion des sessions avec cookies a été mis en place pour mémoriser les informations de parcours dans le fil d'Ariane. En effet, le site est en grande partie statique au niveau des pages bibliographiques (catalogues, notices, tables des matières et des planches).

L'article 32 II de la loi informatique et libertés demande expressément de renseigner l'internaute afin d'obtenir son consentement avant le dépôt préalable du cookie. Cependant, cette règle ne s'applique pas pour « *les cookies permettant d'enregistrer la langue parlée par l'utilisateur (pour les sites traduits en plusieurs langues) ou autres préférences nécessaires à la fourniture du service demandé* ». ¹⁴

14. <http://www.cnil.fr/en-savoir-plus/fiches-pratiques/fiche/article/ce-que-le-paquet-telecom-change-pour-les-cookies/>

Cf. la figure C.5.

6.3.4.1 Onglet notice bibliographique

En plus des informations bibliographiques, il est indiqué un permalien du titre consulté. Il est formulé de la manière suivante : « <http://cnum.cnam.fr/redir.cgi?> » suivi de l'identifiant de l'ouvrage, p. ex. <http://cnum.cnam.fr/redir?12KY1>.

Cf. la figure C.6.

6.3.4.2 Onglet liste des volumes

Dans le cas des revues et grandes monographies, cet onglet présente la liste des volumes dont les titres sont répartis dans deux colonnes.

6.3.4.3 Onglet tables des matières et d'illustrations

Plusieurs pistes d'affichages des titres de chapitre ont été explorées comme une présentation sur deux colonnes, ou l'insertion des vignettes à côté de la ligne de table, ou bien celle d'un affichage sélectif au curseur du niveau hiérarchique mais toutes ces solutions n'ont pas été retenues.

Ce dernier affichage nous aurait permis de fusionner les tables des matières intégrales et abrégées et de présenter les niveaux de manière sélective et plus fine.

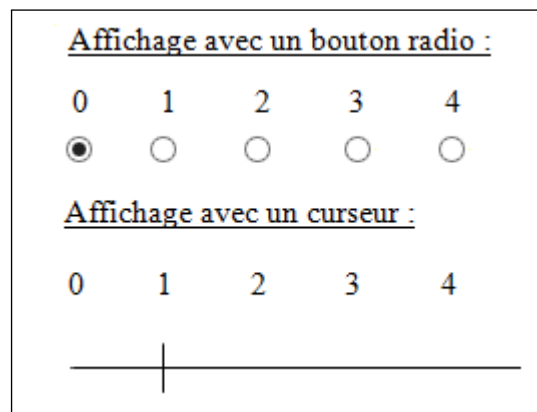


FIGURE 6.2 – Activation de certains niveaux hiérarchiques des tables

Le lien hypertexte du titre de chapitre renvoyant au fac-similé sera placé sur ce dernier au lieu du numéro de page. Les titres de chapitre contenant un plus grand nombre de caractères que les numéros de page offriront une plus grande zone de pointage. Par ce changement, nous pensons aussi que l'algorithme de Google classerait la page en meilleure position par rapport à d'autres pages de sites contenant les mêmes mots-clés. Nous relevons dans les consignes de Google aux webmasters¹⁵ et aussi sur les sites de conseils en SEO (Search engine optimization) de porter une attention particulière au contenu des attributs techniques des liens tels que « Title » et « Alt », mots-clés inclus dans le texte du lien.

Pour des raisons de confort de lecture, nous avons désactivé le soulignement par défaut des liens hypertextes des titres de chapitres mais il redevient actif au survol du pointeur.

Le système de pagination multiple, nommage particulier des numéros de pages, est un système propre au Cnum. Il permet de différencier les numéros de pages lors des reprises de numérotation dans l'ouvrage. Un débat a porté sur la suppression ou le placement des numéros de pages simples et multiples, soit sur le côté gauche, soit sur le côté droit du titre de chapitre. Ces deux éléments, texte et nombre sont perçus par certains comme des éléments antagonistes et allogènes, sauf à trouver une présentation différente en jouant sur la ponctuation, la typographie ou le graphisme.

Le placement à gauche - déjà sur Cnum 1 - contrairement à celui de droite présente comme avantage d'aligner tous les chiffres dans une colonne unique.

Une balise de style a été placée spécifiquement sur ces numéros de pages mais n'a pas été exploité graphiquement.

Un retrait de paragraphe au niveau des titres de chapitre n'est pas suffisant dans la distinction entre les différents niveaux hiérarchiques, surtout sur plusieurs lignes de même niveau. Ainsi, nous avons étendu le système des barres de couleur (ressemblant à du timeline) de Cnum 1 en les dissociant de la même colonne.

6.3.4.4 Onglet vignettes

Cette interface présente toutes les pages de l'ouvrage sous forme réduite au format vignette. A l'origine, conçu pour l'interface « cnumdev » qui est l'application « back office » développée par [dAR09] dans le cadre de sa thèse, elle est désormais intégrée dans le nouveau site. Par ailleurs, cette interface faisait partie d'une problématique de visualisation du contrôle qualité de lots d'images numérisées [dAATC09].

Elle sert aussi à localiser rapidement les pages en couleur et celles contenant des illustrations. D'ailleurs, un bibliothécaire se sert de cette interface pour détecter les

15. <https://support.google.com/webmasters/answer/35769?hl=fr>

Pagination	
-	Première image
-	PAGE DE TITRE
r1x1	AVANT-PROPOS (p.r1x1)
1x1	De la découverte des longitudes en m
r2x1	INTRODUCTION Au Supplément du Trai
r2x9	TABLE DES CHAPITRES Du Supplément
2x1	PREMIERE PARTIE. Des moyens propres
2x1	CHAPITRE I. Disposition à donner aux h
2x1	Compensateur isochrone (p.2x1)
2x3	De la manière de régler l'isochronism
2x3	De la manière de régler l'isochronism
2x3	De la manière de régler l'isochronism
2x5	CHAPITRE II. Des moyens propres à obt
2x5	Table composée des arcs et de la tem

FIGURE 6.3 – Mise en valeur du niveau hiérarchique des titres de chapitre

images en couleur afin de le renseigner dans le Sudoc.

Le script de génération des vignettes a été développé en 2008 et n'a pas été appliqué rétroactivement sur le reste du lot. En effet, le nommage des fichiers images binarisés ayant changé d'une période à l'autre, il était impossible d'appliquer ce script sans générer d'erreurs. Nous l'avons modifié pour l'adapter aux images gifs dont le système de nommage est resté identique. Par ailleurs, nous avons fait de même sur la génération des documents pdf dont le problème était similaire.

Nous avons limité l'épaisseur de la marge entre les vignettes pour augmenter la fluidité du basculement entre la vignette complète et rognée lors du survol du pointeur. Dans le cas particulier des titres de revue, les vignettes correspondent aux pages de titres des volumes dont le lien hypertexte pointe sur ces dernières.

Les deux pistes suivantes n'ont pas été exploitées : nous pouvions différencier graphiquement les vignettes contenant des planches de celles contenant du texte, en exploitant la métadonnée enregistrée dans le nommage du fichier image. Une autre solution proposée dans le chapitre des tests sera plus pertinente car les vignettes d'illustration sont déjà repérables visuellement. De même, nous pouvions étendre les fonctionnalités de cette interface en affichant les titres des chapitres et les légendes au survol du pointeur sur les vignettes correspondantes (infobulle en css3).

Cf. la figure C.7.

6.3.4.5 Onglet export

L'export du document complet correspond à une demande primordiale des internautes, d'après l'étude faite dans le chapitre 4.1.1. Le téléchargement du document

6.3. PROTOTYPES

par les internautes n'est pas seulement lié à un besoin d'appropriation, en effet, certains appréhendent son retrait ou du moins la restriction à son accès par un système d'authentification suite à diverses raisons : plaintes, non respect du droit d'auteur, erreurs, changement de politique documentaire. . . Ils redoutent également une consultation par un système d'abonnement payant ou réservé à une certaine catégorie d'utilisateurs dont ils seraient exclus : étudiants, personnels, chercheurs de l'établissement où se trouve la bibliothèque.

L'export permet aussi d'imprimer tout ou une partie du document, car il n'est pas donné à tout le monde de lire aisément sur écran.

Nous avons bloqué l'indexation de cet onglet par les robots d'indexation en déclarant le fichier robots.txt qui est placé à la racine du site. Nous limitons ainsi les montées en charge du serveur suite au téléchargement de ces documents pdf. Par contre, nous avons autorisé l'indexation dans le fichier sitemap.xml de deux documents pdf numérisés en texte intégral, car ce mode n'est pas encore consultable sur une interface dédiée à cet effet. En effet, la bibliothèque a mené en 2008 des expériences de test d'OCR sur un ouvrage de « Pierre Le Muet » et de « Brame, Edouard »¹⁶ dans l'optique d'évaluer ce traitement et d'offrir peut-être dans l'avenir, tous les documents du Cnum en texte intégral.

Nous pensons proposer par la suite d'autres formats d'export comme Bibtex, Epub et insérer un système de téléchargement avec les conditions d'utilisation et une validation par un captcha. Il était possible de proposer en export les notices Unimarc du SIGB¹⁷, mais celles-ci ne sont pas normalisées car le caractère « \$ », représentant le séparateur de champ est manquant.

A noter que les bibliothécaires souhaitaient une intégration des notes de présentation de certaines revues ou thématiques dans le document pdf.

6.3.4.6 Onglet Rechercher dans ce titre

Cette fonctionnalité s'avère utile dans le cas d'une recherche dans des séries ou des périodiques et surtout, en prévision d'une recherche en texte intégral. En effet, la restriction de la recherche permet de réduire le temps d'affichage des résultats.

Sur les conseils du graphiste, l'ouverture de cet onglet nous fait basculer sur la page de recherche avancée et pré-remplit automatiquement le champ identifiant en plaçant le curseur dans le deuxième champ. Ce contournement semble plus judicieux que celui de l'intégration dans ce même onglet d'une nouvelle « mini » page de recherche avancée. Cet onglet particulier a été bien mis en valeur par le graphiste pour le

16. <http://cnum.cnam.fr/redir?FOLK05>, <http://cnum.cnam.fr/redir?8LE144>

17. http://orion.cnam.fr/search~S8*frf?/.b1132393/.b1132393/1%2C1%2C1%2CB/marc~b1132393

différencier des autres.

6.3.4.7 Nuage de mots-clés

Le problème de cette fonctionnalité « tags » est que l'on ne dispose pas à l'heure actuelle de l'algorithme optimal. D'ailleurs, cela mériterait de faire l'objet d'un autre sujet d'étude. Il a été traité en partie par un élève du Cnam de master 2 CAM lors de son stage [ABD10]. Nous le développons plus en détails dans le chapitre 7. Nous pouvons cependant remarquer que les nuages de mots-clés pourraient s'avérer très pertinents sur des documents de type série et périodique car le bruit (i.e. les réponses non-pertinentes) est relativement acceptable. En guise d'exemple, nous présentons en annexe la liste de tous les mots des tables de matières de l'ouvrage BSPI.9 F classés par occurrence décroissante.

Cf. la figure D.34.

6.3.5 Page feuilleteur

Cette page nommée également visualiseur permet de naviguer dans tous les fac-similés. Le document numérisé doit recouvrir au maximum tout l'espace pour en faciliter la lecture.

Les pages étant souvent orientées en mode portrait, il était plus pertinent de placer les barres de navigation en position verticale pour couvrir le moins possible la zone de lecture ; de plus, les droitiers étant en nombre majoritaire, nous l'avons placée par défaut à droite. Or, le système devant être accessible pour n'importe quel type d'interface et aussi par les personnes avec une latéralité différente (les gauchers), nous avons offert la possibilité aux internautes de déplacer cette barre de navigation suivant trois modes : gauche, droite et haut (mais pas le mode bas).

Nous nous sommes posé alors la question de la pertinence de la création d'une barre de navigation horizontale, sachant qu'elle cacherait une partie de la zone de lecture, et que la plupart des pages se lisent dans le sens de la largeur. De plus, la barre placée au niveau supérieur et non inférieur posera des contraintes de pointage sur les interfaces tactiles si on accède au bouton par le bas de l'écran car la main masque la zone de lecture. Le sens de feuilletage avec la barre de navigation placée à l'horizontal peut paraître plus naturel de par l'analogie avec le feuilletage dans le monde réel. Nous pourrions valider toutes ces interprétations par l'étude des logs de connexions des internautes sur le nouveau site.

La navigation fluide sur les interfaces tactiles devant être prise en compte, nous avons veillé à ce que les boutons de cette barre soient assez larges et bien agencés lors de leur activation avec les doigts.

6.3. PROTOTYPES

Dans un autre registre, nous nous sommes interrogés quant à l'intégration des tables des matières dans l'interface feuilleteur, idem pour le cas des vignettes. En effet, la longueur de certains titres de chapitres et l'espace réduit de cette deuxième barre de navigation ne facilite pas la lecture. Cependant, cette fonctionnalité s'avère utile lorsque nous mettons en regard le fac-similé avec son titre de chapitre ou sa légende d'illustration.

Une question d'un autre ordre a porté sur la pertinence de la navigation dans l'ouvrage avec un affichage en mode page unique ou continu (pratique pour les interfaces tactiles) en exploitant au choix : la barre de défilement du navigateur, la molette ou le pointeur (équivalent de l'outil main dans les lecteurs pdf) de la souris. Nous n'avons pas opté pour ce choix et nous avons comblé ce manque en proposant le système de navigation par vignettes qui a pour avantage d'offrir une plus grande appréhension du document.

Certaines fonctionnalités comme la consultation des images en haute définition ont été modélisées dès le début sur le feuilleteur. En effet, nous avons défini dans le cahier des charges de 2010 une demande de numérisation des images en haute définition (400 dpi) pour la diffusion. Cependant, les bibliothécaires ne souhaitent pas diffuser ces images car elles pourraient être détournées à des fins commerciales comme la vente de cartes postales ou d'affiches...

De même, le bouton permettant de lire en mode texte intégral - ou d'océriser - suivant l'état d'avancement du développement a été lui aussi supprimé au final.

Il était possible de fournir une navigation par pagination en passant par la liste déroulante comme c'est le cas sur Gallica ou sur la bibliothèque numérique d'histoire régionale du Nord Pas-de-Calais¹⁸, mais nous avons provisoirement laissé de côté cette fonctionnalité.

Un bouton servant à l'impression papier (contenant par la suite les conditions d'utilisation) faisait partie des fonctionnalités demandées par le groupe de travail. En effet sur l'ancien feuilleteur, l'impression à partir du menu du navigateur nous sortait le cadre de navigation (frame) et non le fac-similé. Cependant, il est à noter qu'aucune plainte à ce sujet ne nous a été envoyée par courriel.

La fonctionnalité rotation des pages du feuilleteur peut être mise à contribution. Tout en respectant la loi « UE » sur les cookies, l'étude de l'historique des sessions utilisateurs nous permettrait d'effectuer un contrôle qualité sur les images. Nous pourrions par exemple, déterminer les pages qui ont été le plus souvent retournées et par là même corriger si besoin leurs surfaces d'affichage, pour les orienter dans le sens de la largeur ou de la hauteur.

18. http://nordnum.univ-lille3.fr/sdx/nordnum/consultation.xsp?b=B590092101&d=00090.844_001

Nous avons conservé au final la fonctionnalité défilement des illustrations, malgré un certain double emploi avec le menu vignette. Nous avons aussi fait de même pour le mode zoom qui permet de voir la page entièrement ou plus en détail car les images sont de taille variée.

Cette page a été modélisée en prévision de l'OCR, le menu de gauche avec les différents onglets de navigation a été prévu à cet effet.

Cf. la figure [C.8](#).

6.3.6 Page de recherche avancée

Nous souhaitons conserver la recherche dans les différentes tables et les notices bibliographiques à l'identique de Cnum 1 (idem pour la page de résultat), avec des croisements supplémentaires d'option par les opérateurs logiques « Et », « Ou » et « Sauf ». En partant de ces besoins, nous avons posé les briques de base de ce prototype en nous inspirant principalement de la page de recherche de Gallica, car elle est complète et connue de la plupart des gens. Nous avons veillé à présenter tous les éléments en limitant au maximum le masquage d'information.

Nous proposons dans une liste déroulante les champs de recherche suivants : tables des matières et illustration (les deux en même temps ou séparées), auteur, titre, date, lieu d'édition, nom d'éditeur, sujet, cote et identifiant. La première liste déroulante a été remplacée par la liste à choix multiples, conformément à notre souhait de présenter toutes les informations visibles sur la page, également parce qu'elle est pratique sur tablettes tactiles.

La recherche dans les champs « type de documents » et « thématiques » sera faite au travers de cases à cocher car les titres peuvent être inclus dans un ou plusieurs éléments de ces critères. Par contre, la date d'édition et de mise en ligne seront accessibles via deux boutons radio car il n'y a qu'un seul choix possible entre les deux.

La recherche par langue a été rajoutée en détournant dans la base de données l'attribut dédié à l'OAI-PMH. Les différents noms du menu de recherche par langue sont affichés par occurrence décroissante des titres. Ainsi au 18/11/2012, il y avait deux mille quatre cent quatre-vingt quatorze ouvrages en français, trente et un en anglais, trente et un en allemand et un en italien, le classement des éléments des deux autres filtres se faisant par ordre alphabétique et de granularité.

Nous prévoyons une éventuelle recherche en mode texte intégral. Dans ce cas, une case à cocher « Recherche en texte intégral » sera rajoutée à condition que toutes les pages du Cnum soient complètement « océrisées ».

Le système de tri des résultats suivant l’auteur n’avait pas de sens pour le groupe de travail. Le cas des volumes de revues n’ayant parfois aucun auteur exprime clairement l’erreur de ce type de classement. Ce problème est exactement identique à celui du classement de la page catalogue.

Cf. la figure [C.9](#).

6.3.7 Page de résultat

Nous structurons cette longue liste des résultats de recherche suivant quatre modes de présentation : abrégé, texte, vignette-texte et vignette. Le mode par défaut est l’abrégé, mais c’est le dernier sélectionné qui prévaut sur la présentation des prochains résultats.

Dans le mode vignette, le premier prototype prévoyait d’insérer les titres de chapitre et les légendes sous les vignettes. Mais la place prise par la longueur de certaines légendes nous a finalement poussé à intégrer cette information dans l’infobulle. Nous avons ainsi conservé la cohérence d’affichage de ce mode avec celui de l’interface synoptique.

Par ailleurs, ce mode pourrait être exploité comme outil de repérage de cartes (cf. chap. [4.2](#)).

La présentation du mode abrégé regroupe le nombre d’occurrences par titre comme c’est le cas sur Cnum 1. Le groupe de travail souhaite un système d’ouverture et de fermeture de chacun de ces regroupements contenant les titres de chapitre des titres de volume. Les trois autres modes obligent à présenter tous les résultats du mode abrégé sur la même page, c’est-à-dire avec la liste complète des volumes de revues et les monographies, contrairement à l’affichage Cnum 1.

D’autres fonctionnalités ont été rajoutées, comme un système de recherche à facettes [[Hea06](#)] pour affiner les résultats, et aussi, un bouton « modifier la recherche » qui permettra de revenir sur la page de recherche avancée, avec les mêmes champs et valeurs saisis au départ par l’utilisateur.

6.3.8 Outils de découverte

Nous avons eu quelques difficultés à trouver un emplacement adéquat menant vers les nouveaux outils et services complémentaires que l’on souhaitait offrir. La barre de navigation supérieure, présente sur toutes les pages, en est finalement devenue le lieu idéal. L’icône a été privilégiée par rapport au texte en hyperlien pour rationaliser l’espace de lecture qui servira à l’avenir à placer d’autres éléments.

L'icône représentant une roue dentée nous est apparue comme la métaphore de visualisation la plus opportune à ce service en lien avec l'histoire des techniques, l'idée sous-jacente de la roue d'Agostino Ramelli et la notion d'aléatoire du jeu de la roue de loterie.

Ainsi, nous avons rassemblé dans une page spéciale nommée « Outils de découverte » les futurs services d'exploration du Cnum. Nous en proposons actuellement deux : le premier favorise la sérendipité par un défilement aléatoire de pages spécifiques et le deuxième est un index alphanumérique des tables des matières et des illustrations.

6.3.8.1 Découvrir des pages au hasard

La notion de sérendipité est le fait de découvrir une chose par hasard. Ce concept, qui est d'ailleurs à l'origine de nombreuses découvertes et inventions scientifiques, l'auteur [MA11, p. 37], décrit les caractéristiques des « *usagers intensifs de corpus numérisés* » : « *La facilité et la rapidité d'accès encouragent en effet la "rigueur" scientifique et la "curiosité" intellectuelle. De façon quasi unanime, le groupe interrogé souligne l'impact positif de la "découverte hasardeuse" que permet la recherche sur les corpus numériques (...) ça ouvre des pistes qu'on ne connaît pas* ».

Cet accès « Cnum-roulette » était initialement prévu sur le bouton de validation du formulaire de la recherche simple. Nous l'avons étendu en exploitant la variabilité des typologie des documents (monographie, périodique et série) et des onglets de l'interface synoptique. Il offre ainsi à tout un chacun un choix de découverte plus exclusif et adapté à sa convenance. Nous les présentons dans la liste suivante :

- N'importe quelle page : <http://cnum.cnam.fr/CGI/random.cgi?tout>
- Une notice : <http://cnum.cnam.fr/CGI/random.cgi?n>
- Une notice de série : <http://cnum.cnam.fr/CGI/random.cgi?npg>
- Une table des matières abrégée : <http://cnum.cnam.fr/CGI/random.cgi?a>
- Une table des matières intégrale : <http://cnum.cnam.fr/CGI/random.cgi?d>
- Une table des illustrations : <http://cnum.cnam.fr/CGI/random.cgi?i>
- Une liste des volumes : <http://cnum.cnam.fr/CGI/random.cgi?lv>
- Vignettes d'une série : <http://cnum.cnam.fr/CGI/random.cgi?t>
- Vignettes détaillées d'une série : <http://cnum.cnam.fr/CGI/random.cgi?x>
- Vignettes : <http://cnum.cnam.fr/CGI/random.cgi?v>
- Vignettes détaillées : <http://cnum.cnam.fr/CGI/random.cgi?u>
- Une page de titre : <http://cnum.cnam.fr/CGI/random.cgi?ft>
- Une page d'illustration : <http://cnum.cnam.fr/CGI/random.cgi?fi>
- Une page d'un ouvrage : <http://cnum.cnam.fr/CGI/random.cgi?f>
- Une note de présentation : <http://cnum.cnam.fr/CGI/random.cgi?np>

Dans le cas du défilement aléatoire des images, nous avons opté pour l'interface de

consultation courant (feuilleter avec le fil d'Ariane) plutôt que pour un affichage brut (image seule). L'image brute serait certes plus rapide à l'affichage mais nous empêcherait d'en connaître la provenance.

Nous pouvons aussi pousser plus loin ce concept, en positionnant aléatoirement la barre de défilement verticale, qui reste actuellement par défaut en haut de la page, et restreint d'une certaine manière la découverte aléatoire.

Cf. la figure [D.29](#).

6.3.8.2 L'index alphanumérique

La difficulté principale de ce type d'interface réside dans l'affichage exhaustif des données. La longueur variable de certaines listes de mots regroupés par lettre alphabétique pose un problème au niveau du temps d'affichage des résultats et de la navigation. Mais nous avons résolu ce problème en proposant trois listes alphabétiques dont la dernière est basée sur la navigation par liens ancrés. La compatibilité d'affichage et de navigation avec le reste du site a aussi été respectée. La page catalogue avec ses différentes colonnes de tri et ses liens ancrés ont servi de base à la conception de l'interface index. La première, deuxième et troisième colonne (ou liste alphabétique) servent respectivement à afficher la liste des mots commençant par la première, deuxième et troisième lettre de l'alphabet.

Cette interface est utile à double titre, d'abord comme outil de découverte pour l'utilisateur, puis comme outil de correction des tables pour les bibliothécaires. En effet, les erreurs étant d'origine variables, nous pouvons trouver des lettres mal ou non reconnues, des mots collés, etc. Ceux-ci proviennent majoritairement des erreurs d'océrisation des tables des matières.

Nous prévoyons également de proposer par la suite un index sujet classé par ordre alphabétique et d'occurrence. Ce type d'interface sert aussi à présenter sous un autre angle que celui de la page d'accueil, de manière exhaustive et détaillée, le contenu du fonds documentaire.

Voir la capture d'écran à l'annexe [D.30](#)

6.4 Phase de test

Au fur et à mesure des réunions, les différents prototypes étaient présentés et soumis à des tests utilisateurs et à une validation en interne. Nous avons confronté le site à différents profils d'utilisateurs. Il y a eu des allers retours entre la spécification et l'implémentation.

Avec la variabilité des supports électroniques, la phase de tests des résolutions d'écran peut s'avérer très coûteuse financièrement. Cependant, plusieurs outils gratuits en ligne existent comme Testsize¹⁹ ou Screenfly²⁰. Ils nous permettent de visualiser et détecter rapidement les problèmes d'affichage de notre site sur les différents supports et résolutions écrans.

Le site fonctionne sur les dernières versions de Chrome, Internet Explorer, Firefox mais reste Safari et Opera. Il sera également soumis aux tests sur tablettes tactiles (Ipad et Surface).

Un test de débit sera aussi effectué, selon que l'internaute disposera d'une connexion bas débit ou bien sera situé très loin du serveur. Un simple paramétrage au niveau du serveur suffit à mener ces tests, sans nul besoin de se déplacer sur une zone éloignée.

6.4.1 Page d'accueil

Au niveau du codage, un fichier php récupère dynamiquement tous les chiffres du nombre de volumes, de pages et de titres de la base de données avant d'être converti en un fichier html (site statique oblige).

La proposition du graphiste concernant le défilement automatique de l'image principale (appelé carrousel), situé tout en haut de page n'a pas été retenue par l'équipe, privilégiant de préférence les informations statiques.

Au départ, ces thématiques et catégories étaient disposées de manière aléatoire ou en nuage de mots-clés. Les utilisateurs ont souhaité à la place un classement en colonne et par ordre alphabétique comme c'est le cas dans les dictionnaires, avec néanmoins une séparation marquée entre les deux ensembles. Suite à ce constat, nous avons intégré une séparation via un menu déroulant en JavaScript B.1. Ce type d'affichage présente l'avantage de contenir l'ensemble des informations dans un espace réduit et d'intégrer facilement de nouvelles thématiques. A l'inverse, son inconvénient majeur est la non visibilité de toutes les informations dans l'espace utile. Nous sommes finalement revenus à la représentation iconographique.

L'extension de la zone cliquable du texte indiquant le nombre de titres de la thématique à celle de sa vignette est malheureusement impossible. En effet, ce texte est placé au premier plan alors que la vignette est incrustée en « background » par le css. Cette combinaison oblige à changer tout le codage d'intégration de la page.

Certains utilisateurs trouvaient cette page trop chargée en terme visuel et informationnel, en particulier sur les données de volumétrie. Or, fait étrange, il s'est avéré bien plus tard, qu'un de ces mêmes utilisateurs m'a demandé pour des calculs statistiques, les mêmes chiffres qu'il trouvait inutiles.

19. <http://testsize.com>

20. <http://quirktools.com/screenfly>

Un autre utilisateur n'étant pas satisfait du système de classement par thématique et typologie a émis le souhait d'une navigation avec un plan de classement ou un thésaurus comme c'est le cas au CNHI²¹. Ce système d'affichage structuré en exploitant par exemple les mots sujets dont le lien hypertexte lancerait le moteur de recherche serait peut-être une solution à notre problème de représentation exhaustive du fonds documentaire.

Cf. la figure D.1.

6.4.2 Page catalogue

Nous pourrions remplacer certains pictogrammes au format gif par du SVG, réduisant ainsi le temps de téléchargement des pages. Or, certaines versions de navigateurs ne le prennent pas en charge. Néanmoins, les tests ont montré que selon l'image de la vignette, la taille entre les deux fichiers était sensiblement identique.

La liste des liens ancrés a été basculée au dessus des onglets de tri. Nous avons privilégié dans ce cas le graphisme plutôt que la logique de navigation.

Lors de l'intégration, le texte de présentation des thématiques est masqué automatiquement lors d'un défilement pour augmenter l'espace de lecture des titres d'ouvrages.

Un utilisateur trouvait anormal la présence des titres de revues dans le tri par auteur, considérant sa suppression plus pertinente à ce classement, sauf peut-être à remplacer le terme « Revue » par « Auteurs collectifs de revues ».

Un utilisateur nous a demandé de répéter autant de fois les titres contenant plusieurs lieux ou plusieurs noms d'éditeurs. Ainsi, nous avons créé spécialement une table avec vue regroupant tous les attributs `cle_lieu1`, `cle_lieu2` et `cle_lieu3`, idem pour `cle_nom1`, `cle_nom2` et `cle_nom3`.

Nous avons amélioré la lisibilité en supprimant l'information superflue du nombre de volumes « (1 vol.) », celle-ci est supposée être comprise implicitement. De même, nous pouvons différencier les titres de grandes monographies de ceux des « quanta virtuels ». Ces derniers étant regroupés parfois en tomes ou en volumes²² impliquent de présenter ces informations conformément aux règles de catalogage (ex. 2 tomes en 1 vol.), à cette fin un nouvel attribut « nombre de tomes » doit être rajouté dans la table de la base de données.

21. Cité nationale de l'histoire de l'immigration http://ressources.histoire-immigration.fr/ex1-php/cadcgp.php?CMD=CHERCHE&VUE=cnhi_recherche_thematique&MODELE=vues/cnhi_recherche_thematique/home.html&query=1%0A%09

22. ex. <http://cnum.cnam.fr/CGI/redir.cgi?8XAE17>, http://cnum.cnam.fr/CGI/redir.cgi?M9977_2, <http://cnum.cnam.fr/CGI/redir.cgi?8SAR12>

Les premiers essais avec des vignettes blanches des titres soumis aux droits d'auteurs ont été remplacées par des vignettes avec une croix. Le verrouillage du dossier vignette par un fichier « .htaccess » nous oblige à créer un autre dossier destiné spécifiquement à cette présentation.

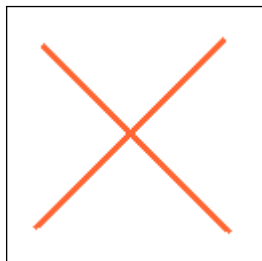


FIGURE 6.4 – Vignette de la page de titre soumis au droit d'auteur

Nous avons désactivé les vignettes du catalogue général pour alléger la lecture. De même, les colonnes « lieu édition » et « éditeur » du catalogue des manuscrits n'ayant pas de sens bibliothéconomique ont été désactivées de ce regroupement.

D'autres pistes restent encore à explorer comme l'amélioration de la lecture du tableau en appliquant un ombrage du fond de ligne en alternance. Certaines données se répétant d'une ligne à l'autre pourraient être supprimées. De même, on pourrait raccourcir certaines lignes de séparation des titres.

Cf. les figures [D.2](#), [D.3](#), [D.4](#), [D.5](#), [D.6](#), [D.7](#).

6.4.3 Page synoptique

6.4.3.1 Onglet Notice bibliographique

A la suite des tests utilisateurs, nous avons intégré un saut de ligne après chaque groupe de mot matière. Dans le cas des ouvrages soumis au droit d'auteur, nous avons inséré une mention spéciale dans la zone de note pour les retrouver plus facilement par le moteur de recherche.

Au vu du nombre restreint de champs, l'alternance des couleurs lors du survol du pointeur a été désactivée.

Cf. la figure [D.8](#).

6.4.3.2 Onglet Tables des matières abrégée et intégrale

Le groupe de travail s'est demandé s'il était pertinent de toujours conserver la « table abrégée » au profit de la « table intégrale ». Pour effectuer ce choix, nous

nous sommes basés sur l'étude des logs en calculant le nombre de fois où ces deux pages ont été consultées. La commande Unix permettant d'effectuer ce calcul est la suivante : « `cat access_log | grep /cnum/SYN/ | grep -v google | wc -l` » et « `cat access_log | grep /cnum/DET/ | grep -v google | wc -l` » retournant chacune 93202 et 39200 occurrences. Il s'avère donc que $39200 \times 100 \times 93202 = 42\%$ des internautes ont ouvert la « table des matières intégrale » après la « table des matières abrégée ». En conséquence, nous avons conservé les deux onglets.

De même, dans le cas plus précis où l'une de ces tables n'existerait pas, faudrait-il laisser ou non l'onglet actif (mis seulement en grisé)? Un utilisateur préfère le supprimer, malgré parfois un décalage de l'emplacement d'un titre à l'autre, mais cela n'a pas été pénalisant lors des tests.

Cf. les figures [D.9](#), [D.10](#) et [D.11](#).

6.4.3.3 Onglet vignettes

Les vignettes des monographies contenant plusieurs volumes, appelés dans notre jargon interne « quanta virtuels » sont disposées les unes à la suite des autres sans aucune séparation de tomes.

Nous avons remarqué un problème de navigation si la page est très longue. Un système de lien ancré pourrait s'avérer fort utile pour sauter directement sur une vignette précise de la page courante. Ce point d'ancrage pourrait être soit celui des vignettes dont les titres de chapitre sont de niveau 0, soit toutes les pages de dizaines ou de centaines. Mais le problème se situe au niveau du nombre de lien ancré qui est proportionnel au nombre de point d'ancrage. Ces liens vont rapidement saturer toute la zone supérieure, il faudra alors trouver un compromis entre le nombre de lien ancré et l'écart des points d'ancrage pour couvrir toutes les zones de l'interface.

Dans l'onglet « table intégrale », chaque titre de chapitre est associé à un niveau hiérarchique et une couleur spécifique (cf. fig. [6.3](#)). Cette dernière pouvait être liée à celle de la bordure de la vignette, voir sur l'exemple [D.33](#) en annexe. Ce type de représentation valorise l'affichage en mettant en avant les vignettes les plus pertinentes. Cependant, nous n'avons pas poursuivi cette expérimentation, qui pourtant aurait été bénéfique pour le contrôle qualité des tables des illustrations.

Cf. la figure [D.12](#).

6.4.3.4 Onglet Export

Dans le cas particulier des périodiques, nous avons souhaité l'intégration de cases à cocher pour faciliter le téléchargement de tout l'ensemble ou de plusieurs

volumes à la fois. Cependant, la crainte d'une attaque par déni de service nous a fait rapidement abandonner cette fonctionnalité. Or, la limitation du nombre de téléchargements serait la voie de résolution de ce problème.

Cf. la figure [D.13](#).

6.4.3.5 Note de présentation

La création de cet onglet provient d'une harmonisation des données du champ « note » et des pages de présentation de certains titres au format html ou pdf, l'exemple des « Annales du Conservatoire des arts et métiers »²³ ou de l'« Étude sur les signaux de chemins de fer à double voie »²⁴.

Il a fallu rajouter dans la base de données le nouveau champ « ID_note_presentation » contenant comme enregistrement le nom du fichier html à ouvrir, pour pallier le problème des codes html saisis dans le champ Note (NB). Ce dernier étant converti automatiquement en code html par la fonction php « htmlentities », rend le document final illisible.

Cf. la figure [D.14](#).

6.4.4 Page feuilleteur

Un internaute nous a indiqué qu'il n'a pas repéré le bouton de basculement du menu de navigation. Ceci est intéressant et nous pousse à analyser les logs sur la réelle pertinence de cette fonctionnalité.

L'onglet liste des volumes des périodiques et des grandes monographies a été modélisé pour la première fois sur la page synoptique des volumes, puis basculé sur le feuilleteur.

De multiples améliorations sont possibles, comme l'insertion d'un onglet supplémentaire dans le menu gauche permettant un affichage dynamique du titre de chapitre. En effet, le feuilletage page à page ou via un défilement aléatoire ne permet pas de connaître le titre de chapitre, si ce n'est en lisant dans la page même. Il existe bien la solution de l'infobulle, mais elle est au prix d'un parasitage régulier de l'image lors du survol du pointeur et donc de la lecture. L'autre possibilité est celle de la mise en valeur de la ligne de table, moyennant un changement complet du codage pour passer de la version statique à dynamique.

Nous pouvons aussi intégrer le défilement des pages via des raccourcis clavier (touche flèche droite ou gauche), mais pas avec la molette de la souris, car ce système entre en conflit avec un défilement vers le bas de la même page du mode

23. <http://cnum.cnam.fr/CGI/redirect.cgi?8KU54>

24. <http://cnum.cnam.fr/CGI/redirect.cgi?8LE144>

page à page mais non continu.

Les boutons de la barre de navigation et en particulier ceux permettant de passer à la page suivante et précédente pose un problème d'interaction sur des pages dont la hauteur est supérieure à l'écran²⁵. En effet, si on défile verticalement vers le bas, les boutons de navigations ne sont plus accessibles et nous obligent à re-défiler dans le sens opposé pour y accéder. Actuellement, trois solutions permettent de régler ce problème :

- faire coulisser automatiquement vers le bas le menu latéral de navigation (notamment adoptée sur le feuilleteur de Gallica)
- intégrer des raccourcis clavier de basculement de page
- basculer la barre de navigation au niveau supérieur pour conserver en permanence sa visibilité

Nous avons adopté la solution du verrouillage du menu de navigation lors du défilement.

Un autre problème du même ordre est celui des pages au « format paysage » qui oblige d'abord à défiler verticalement vers le bas, ensuite horizontalement vers la partie non visible de droite²⁶. Nous réglons ce problème en proposant une solution extérieure à l'interface qui est celle de cliquer sur la molette de la souris.

Tous ces problèmes proviennent du fait que l'on ne souhaitait pas intégrer de barre de défilement interne dans l'interface.

De plus, nous relevons quelques dysfonctionnements comme le clignotement du menu gauche en position fermée lors du défilement des pages, en raison peut-être du débit réseau. De même, la fonctionnalité « zoom » n'est pas complètement opérationnelle. Nous pouvons l'améliorer en récupérant la taille en pixel de la fenêtre du navigateur. Enfin, la navigation en double page a été suspendue provisoirement, faute d'un code de développement adéquat. Elle sera intégrée dans un second temps.

Un système asynchrone en Ajax permettant un chargement différé des images (idem dans le mode vignette de la page de résultat) reste la fonctionnalité principale à mettre en place. En effet, le temps d'affichage de la page est très ralenti par le chargement de toutes les vignettes.

Cf. les figures [D.18](#), [D.19](#), [D.20](#), [D.21](#) et [D.22](#).

6.4.5 Page de recherche avancée

Une confusion est apparue lors de l'utilisation des opérateurs booléens dans le même champ de saisie. En effet, les espaces sont considérés dans Solr comme des

25. Ex. <http://cnum.cnam.fr/CGI/fpage.cgi?8LE73/415/0/416/0415/0415>

26. Ex. <http://cnum.cnam.fr/CGI/fpage.cgi?FOLXAE12.1/39/100/146/0011/0145>

opérateurs logiques « Ou » et non « Et ». La recherche des mots contenus dans la même ligne de table doit être formulée avec des guillemets (voir le détail dans l'annexe E). Ainsi, une deuxième liste déroulante contenant entre autres les champs « contient », « est exactement » permettrait de régler le problème des opérateurs booléens dans les champs de saisies²⁷.

Nous avons aussi remarqué d'après les logs que l'aide en ligne du moteur de recherche n'a pas été beaucoup consultée (voir l'annexe E). Cela est très révélateur de l'usage de ce genre d'outil par les internautes. De par ce constat, nous avons placé un texte explicatif très succinct dans l'infobulle du champ de saisie. Il guidera rapidement les internautes pressés, ne souhaitant pas s'attarder à la lecture d'un long texte d'aide en ligne.

Plusieurs corrections ont été apportées sur cette page, que nous décrivons suivant une liste d'item par exemple :

- le traitement linguistique ne convenant pas dans le cas d'une recherche précise, nous avons rajouté une case à cocher « Recherche exacte » pour désactiver l'insensibilité à la casse, la racinisation et la lemmatisation.
- la balise « optgroup » améliore la lisibilité dans liste déroulante. Ce codage html sépare par catégories les champs des tables de ceux des notices.
- dans le cas particulier de la recherche par année, si le bouton radio « En » est activé, le 2ème champ est automatiquement grisé donc désactivé de la saisie
- le format des dates d'année-mois-jour a été simplifié en année uniquement
- le bruit engendré par les mots vides nous a obligé à les supprimer de l'indexation
- le groupe de travail n'a pas souhaité la mise en place d'outil de suggestion de mots-clés ni de correction orthographique.

Au niveau de la sécurité, certains caractères comme le croisillon ou signe typographique « # » sont bannis de la recherche car ils n'existent pas dans le titre de chapitre (voir l'explication de la construction des tables dans 3.1). De même, le nombre de caractères a été limité à cinq cents. Ce chiffre a été fixé par rapport au nombre maximal de la somme des caractères des identifiants des volumes d'un périodique.

La recherche dans les notices moissonnées de Gallica a été provisoirement différée à cause d'un temps de développement trop important. A la place, nous avons créé un lien de rebond sur le site de Gallica.

Cf. la figure D.23.

27. Voir un exemple sur le lien vers ce diaporama à la page 15 <http://fr.slideshare.net/BUPMCformM/archives-ouvertes-et-thses-en-ligne-hal-et-tel>

6.4.6 Page de résultat

Le groupe de travail souhaitait une séparation entre les notices et les tables dans le mode texte. Nous l'avons mis en place de manière implicite par un retrait à gauche de la marge sur chacune de ces catégories.

La présentation restreinte des résultats, actuellement de 10 par pages, s'est imposée obligatoirement de par le retard d'affichage sur un nombre élevé de réponse. Or, ce genre d'implémentation pourrait rentrer en conflit avec le mode abrégé.

Si l'internaute est situé en dehors de l'enceinte de la bibliothèque centrale, les vignettes soumis aux droits d'auteur n'apparaissent pas à l'affichage, laissant place à une vignette dotée d'une croix (fig. 6.4).

Cf. les figures [D.24](#), [D.25](#), [D.26](#), [D.27](#).

6.5 Visualisation d'information - Infovis

En visualisation d'information, il est important de stimuler la mémoire spatiale de l'utilisateur. Le système doit être rapide, perceptif et non cognitif pour réduire au maximum la charge cognitive.

A l'heure actuelle, ces différentes expérimentations ne sont restées qu'à la seule phase de modélisation.

6.5.1 Visualisation Expérimentale

Page d'accueil. Nous présentons dans ce paragraphe différents graphiques de répartition du nombre de titres et d'images dans les thèmes de la page d'accueil. Mais ce type de représentation ne correspond peut-être pas à un besoin spécifique des utilisateurs. En effet, certains testeurs ont trouvé que l'indication volumétrique était sans intérêt. De plus, ils pensaient que les différents titres étaient répartis exclusivement dans une seule thématique.

Cette dernière remarque nous invite à proposer une interface avec des ensembles disjoints et conjoints de surface proportionnels au nombre de titres. La difficulté de représentation de cette interface se situe au niveau des intersections et de la forme des ensembles. La projection en 3D de cette représentation se rapproche du système Lighthouse de [LA00] et du diagramme de Venn [RW97]. Nous représentons ce dernier dans les figures²⁸ ci-dessous.

28. Diagramme réalisé avec le site <http://bioinformatics.psb.ugent.be/webtools/Venn>

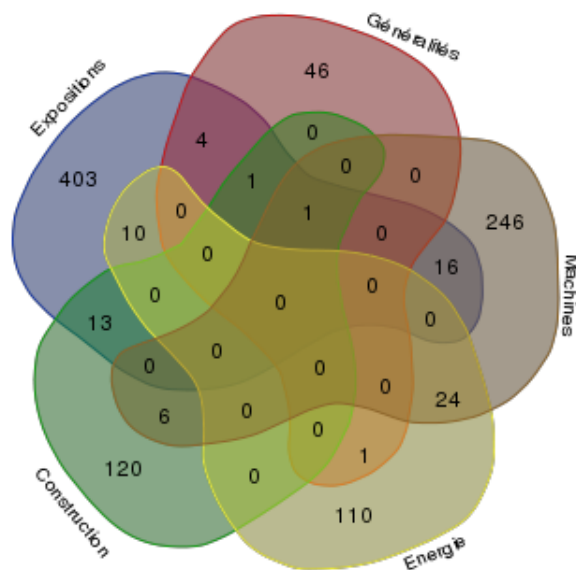


FIGURE 6.5 – Répartition des titres dans les thématiques 1

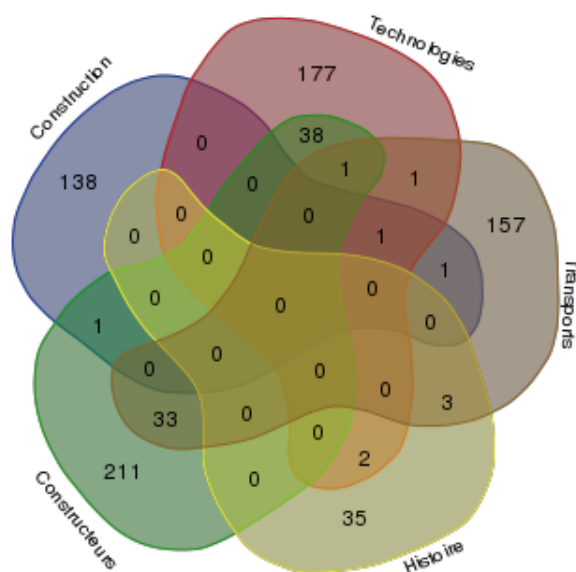


FIGURE 6.6 – Répartition des titres dans les thématiques 2

6.5. VISUALISATION D'INFORMATION - INFOVIS

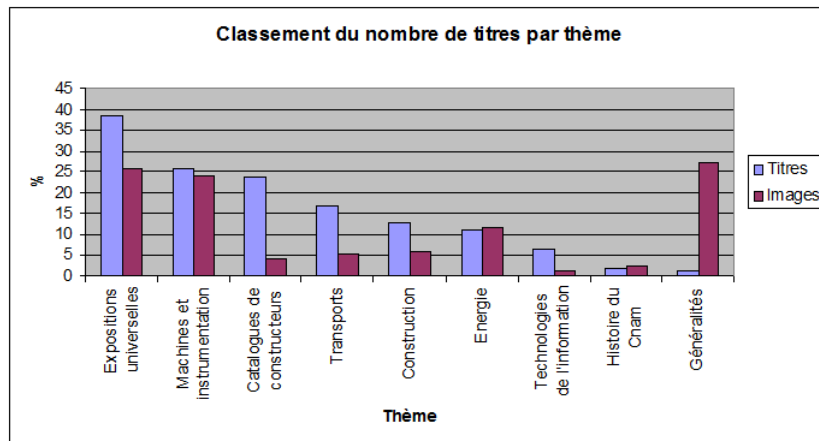


FIGURE 6.7 – Histogramme 1

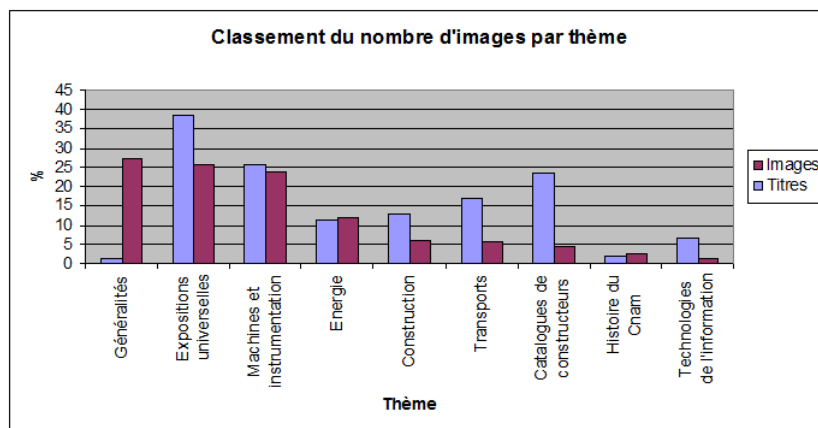


FIGURE 6.8 – Histogramme 2

Ces figures peuvent être aussi implémentées dans des navigateurs web en javascript.²⁹

29. Sortable Bar Chart <http://bl.ocks.org/mbostock/3885705> et <http://bl.ocks.org/mbostock/3885304>

6.5. VISUALISATION D'INFORMATION - INFOVIS

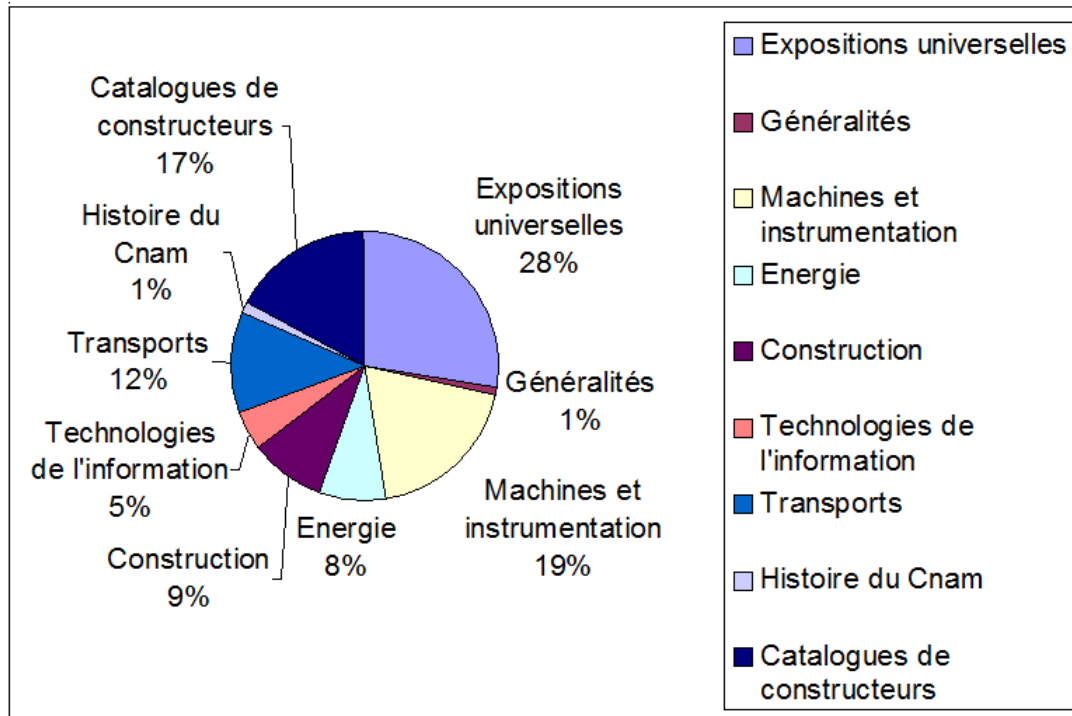


FIGURE 6.9 – Graphique en secteurs

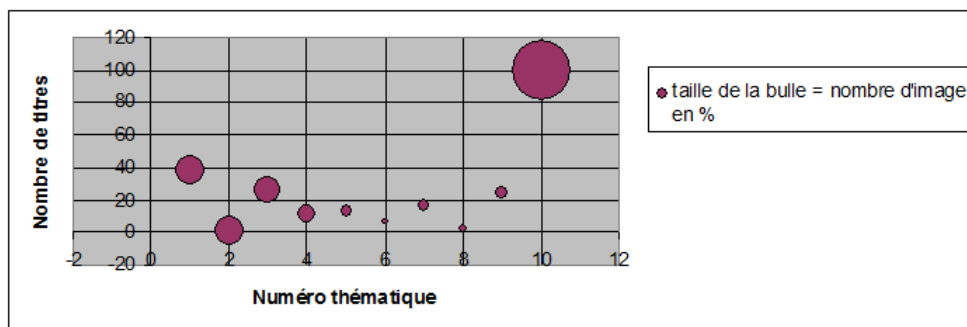


FIGURE 6.10 – Graphique en bulles

Mais selon la vision de [TGM83], il n'y pas de raison que des barres ou des points soient reliés entre eux. Si on pousse plus loin le raisonnement, les axes eux aussi sont inutiles. Au final, on peut représenter toutes ces informations suivant une simple heuristique visuelle, ex. :



FIGURE 6.11 – Heuristique visuelle du nombre de titres et nombre d'images

Visualisation par réduction extrême des vignettes. Lors des premières modélisations, nous avons pensé intégrer un onglet visualisation expérimentale dans les pages catalogue et revues. De cette manière nous pouvons offrir un système permettant de consulter intégralement toutes les vignettes d'un lot ou d'une revue et pas seulement un ouvrage. Ce système, déjà modélisé par [dAC09, chap. 3] pourrait s'implémenter en Ajax avec la bibliothèque JQuery.

La navigation serait identique à celle de l'interface Google map avec un système de zoom interactif.

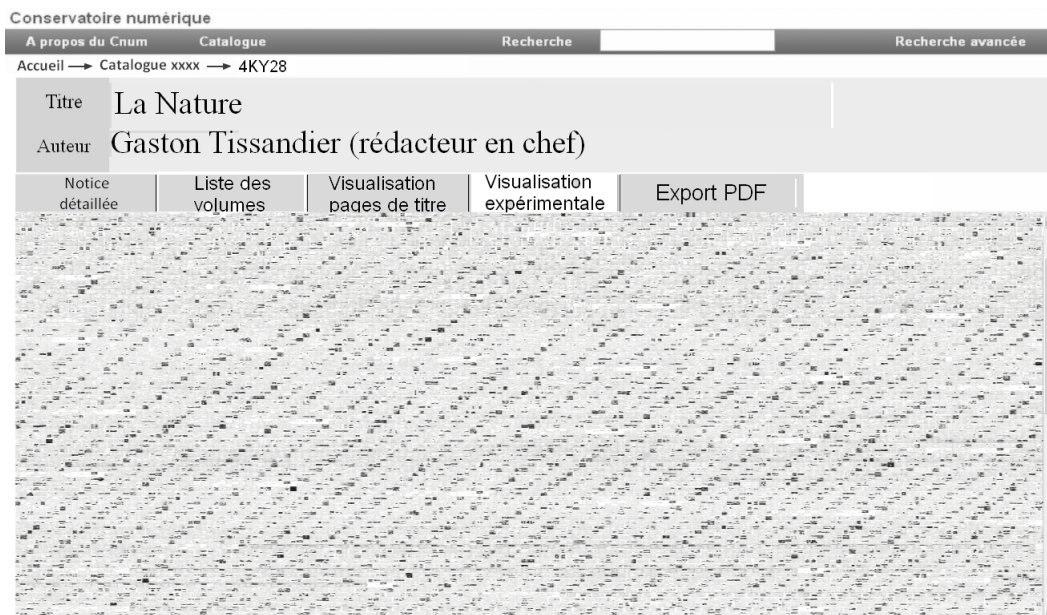


FIGURE 6.12 – Visualisation par réduction extrême

Interface 3D. Des expérimentations en VRML ont déjà été effectuées par [Top02, chap. 5.1.2]. Nous pouvons mener les mêmes expérimentations en 3D avec le langage HTML5. Cependant, nous ne disposons pas de certaines informations dans la base de données telles que la hauteur et la largeur du livre. (voir 6.2.3).

Visualisation cartographique et géoréférencement. D'après l'étude des logs, le mot-clé « carte » fait partie de celui le plus utilisé par les usagers (voir 4.1.1). Une recherche simple dans les tables nous renvoie environ 1 300 occurrences trouvées dans 36 titres. En partant de ce constat, nous pourrions géolocaliser chaque carte avec les métadonnées suivantes : latitude, longitude, orientation magnétique, hauteur, échelle, date. La clusterisation des données géolocalisées (API Google Map MarkerClusterer) serait une solution d'affichage des cartes adjacentes. Ce service peut également être étendu vers le web de données, modèle identique à Geonames³⁰. Ainsi la mise à disposition sur le web de données structurées générera d'autres applications composites (ou « mashups ») et systèmes de visualisations plus riches.

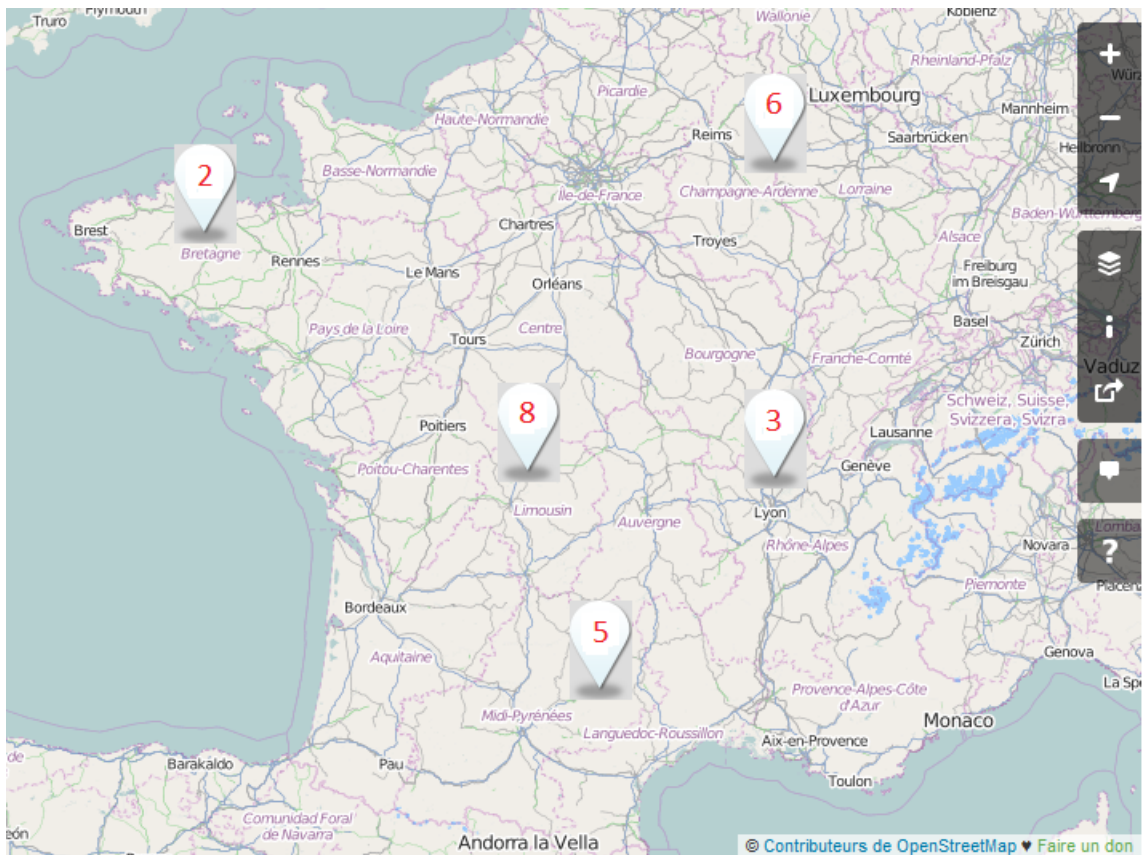


FIGURE 6.13 – Carte géolocalisée et clusterisation

30. <http://fr.wikipedia.org/wiki/GeoNames>

Chapitre 7

Bilan et perspectives

Ce chapitre fait le bilan des contributions apportées. Nous étudions également quelques pistes intéressantes pour la suite de notre activité. En effet, les nouveaux services Cnum sont perfectibles et peuvent trouver ici de nouvelles issues.

Cnum 2 projet libre. Le site entièrement développé sur une brique statique présente comme avantage d'être exportable rapidement sur tout environnement. Un simple « copier-coller » suffit à le faire tourner sur n'importe quel serveur LAMP. De plus, le site dispose désormais depuis sa première version de 2001 de multiples fonctionnalités supplémentaires. Actuellement, il incorpore :

- une version multilingue
- un accès par thématique et typologie
- un fil d'Ariane
- un feuilleteur avec menu basculant
- une interface de visualisation de vignettes à doubles facettes
- un export de la notice bibliographique au format (BibTeX, Zotero)
- un moteur de recherche issu d'un logiciel libre
- un outil de découverte
- un index alphanumérique

Le souhait du laboratoire Cédric serait de faire évoluer la brique logicielle actuelle du Cnum vers un projet libre de développement collaboratif déposé sur une forge. Une initiative de ce type serait une opportunité d'évolution rapide du logiciel Cnum, au vu des ressources limitées mises à notre disposition. Elle apporterait aussi aux professionnels des bibliothèques numériques de nouveaux outils innovants et l'occasion d'un travail interdisciplinaire dans les domaines de la visualisation d'informations, de la fouille de données, du traitement d'image et de l'interaction 3D.

La mise en place de ce projet open source demande une étude sérieuse sur le plan technique, juridique et administratif [bS09] sans oublier les ressources humaines re-

quises pour sa parfaite réalisation.

Nuage de mots-clés. Le groupe de travail a lancé une réflexion sur l'intérêt de cet outil et du pointage final de ses hyperliens. En effet, quelle est la pertinence de ce type de représentation visuelle sur des mots récurrents et de moindre importance, tels que « planche », « chapitre », « nom », « Société », etc ? De même, dans quelle mesure certains articles, mots, élisions, signes diacritiques doivent-ils être épurés de cette liste ? Doit-on respecter la casse typographique ? Est-ce que la problématique de la création de ce nuage de mots-clés, serait la même sur certains type de données et regroupement ? Par exemple, les mots des tables des matières ou le texte intégral, le regroupement en ouvrage, revue, thématique, ou le corpus complet ?

Une solution possible, quant au pointage de ces hyperliens serait de présenter la page de résultat du moteur de recherche, en fonction du mot-clé sélectionné. Une autre possibilité serait de positionner des heuristiques visuels sur les pages de l'interface feuilleteur, avec comme conséquence non négligeable une redéfinition de son architecture initiale. Dans le cas où nous appliquons ce nuage sur le texte intégral, nous pourrions fournir comme résultat des extraits de pages, comme c'est le cas sur l'interface Google Livres.

Le nombre de mots du nuage peut être aussi considérablement réduit par la lemmatisation. C'est en tout cas la piste envisagée par [ABD10] à l'aide de l'outil libre Treetagger. L'exploitation de l'index inversé du moteur de recherche Solr offre également une autre voie de résolution.

Les nuages de mots-clés mettent en valeur les mots les plus occurants en délaissant les autres. Or, le Cnum renferme dans ses tables des matières des mots dotés d'une richesse vocable incomparable. Il apparaît en surabondance, des mots peu courants (dans le sens moins connus du grand public) et des néologismes n'appartenant pas à des mots du dictionnaire tels que « monaul, phragmitetum, borolime,... ». En partant de ce constat, nous avons essayé de construire une interface ad hoc. Nous avons d'abord réalisé une extraction par l'interpréteur de commandes Shell Unix des occurrences de tous les mots utilisés dans les tables des planches que l'on a trié ensuite par ordre croissant (voir une partie des résultats en annexe G). Ensuite nous avons supposé que plus un mot est récurrent, plus il est connu (et inversement) même si cela ne se vérifie pas forcément dans tous les cas de figure. Compte tenu de ces résultats, nous avons imaginé une interface de visualisation regroupant toutes ces données et triée de manière alphabétique ou sur le nombre d'occurrence, mais celle-ci ne nous a pas apporté de résultats probants. Alors, l'idée nous est venue de constituer un index alphanumérique (cf. D.30) du corpus dont les liens hypertextes des mots de toutes les tables rebondissent sur les résultats du moteur de recherche.

Le célèbre jeu de la grille de mots mêlés, cher aux cruciverbistes serait peut-être une piste de résolution de ce problème, la règle étant de placer tous ces mots dans un

espace réduit, dont la mise en valeur serait dépendante d'une action de l'utilisateur. Le croisement de certains mots d'un même titre de chapitre est peut-être à privilégier par rapport à d'autres représentations. Le passage de cette interface en 3D offrirait certes une interaction plus avantageuse. Mais l'étude de cette solution nous mènerait vers des problèmes d'optimisation combinatoire et d'interaction homme-machine, nous écartant trop de notre sujet de départ.

Adaptation au langage du Web sémantique. Toute cette partie a été intégrée dans le Cnum d'une manière sommaire, en exploitant l'élément meta de l'entête html head. Le renseignement de ces balises servait au départ à l'exportation des métadonnées au format Zotero. Ces balises sont sujettes à discussion quant à leur réelle exploitation par les moteurs de recherche¹, bien que la requête « allintitle :mécanique site :cnum.cnam.fr » lancée dans Google filtre bien les résultats sur les titres du Cnum.

Nous pouvons aussi envisager que les pages html du CNUM, soient complétées par un balisage de description (comme RDF ou OWL) de la structure des fonds, permettant une description plus fine des informations.

Interface vignette. L'évaluation de cette interface est une problématique importante à résoudre. Actuellement, l'interface vignette contient pour chaque image, deux vignettes dont une correspondante à l'image réduite, et l'autre au centre de l'image rognée sous forme de carré. Nous voyons bien que ce critère n'est pas significatif dans le cas d'une page contenant de multiples figures, en particulier si elles sont situées hors du centre. De même, quid de la vignette d'une page aux caractères imprimés ? La taille actuelle des vignettes est de dimension 128x128 pixels [dAR09, chap. 7.4]. Est-il possible de définir une taille idéale de ces vignettes pour une lecture fluide ? De même, est-ce que d'autres formes sont à privilégier : carré, rectangle, triangle, losange, cercle, ellipse, oval... ? Ou encore, faut-il adapter la taille et la forme proportionnellement à celle de l'image d'origine et au nombre de pages de l'ouvrage ? Nous entrevoyons ici une possibilité d'affichage à la newsmag [OCSZ05]. De plus, les tailles et les formes variables poseront des problèmes d'empilement et d'occupation optimale de l'espace. C'est le problème classique du « bin packing »² en recherche opérationnelle. De même, le questionnement portera aussi sur le respect de l'ordre d'apparition des vignettes suivant l'ordre de pagination de l'ouvrage.

En conséquence la taille et la forme de la vignette, sa zone de rognage et son niveau de zoom sont des données à redéfinir, si nous souhaitons une interface vignette plus pertinente à la visualisation.

Deux pistes s'offrent à nous pour mener à bien nos investigations : d'une part, le partenariat initié en 2012 avec l'équipe QGAR du laboratoire du Loria³ de Nancy

-
1. <http://www.alsacreation.com/article/lire/628-balises-meta.html>
 2. http://fr.wikipedia.org/wiki/Probl%C3%A8me_de_bin_packing
 3. <http://www.loria.fr>

(professeur Salvatore-Antoine Tabbone), pour une recherche par similarité et détection des planches ; d'autre part, les API Open CV et OCRopus de Google dispose entre autres dans leurs fonctionnalités d'un détecteur de figures (Document layout analysis).

Une fois cette étape franchie, il nous restera à faire une étude et une analyse par traitement d'image de ces zones, en faisant appel à ces grands domaines de l'informatique que sont la vision par ordinateur et la visualisation d'information.

Nous voyons par ailleurs tout l'intérêt et l'avantage d'un site statique, en particulier pour ce type d'interface qui ne pourrait s'afficher en temps réel.

Interface feuilleteur. Les fonctionnalités proposées dans les lecteurs de document pdf, tel que « Adobe Acrobat Reader » ou les visionneuses d'images comme « XnView » favoriserait grandement la visualisation de cette interface. Il existe notamment des fonctionnalités telles qu'une lecture du document en double page, une pagination unique ou continue, un affichage de la liste des numéros de pages redimensionnable manuellement. . .

L'interface feuilleteur est construite de manière à proposer un système de lecture en texte intégral. Un bouton de bascule a été initialement prévu à cet effet sur le menu de navigation, pour passer du mode image au mode texte. Des fonctionnalités connexes peuvent aussi y être associées comme un système de lecture vocale et une recherche plein texte.

Interface de résultats. Le mode abrégé a été provisoirement désactivé, car non implémenté, et demandant un remaniement complet du codage des autres modes. Son absence a d'ailleurs dérouté plusieurs utilisateurs qui étaient habitués à cette fonctionnalité sur l'ancien Cnum, et ont donc vivement souhaité son rétablissement. La solution pour pallier ce problème consistant à envoyer une deuxième requête au serveur ne sera pas sans conséquence sur son temps de traitement. De plus, les résultats ne seront pas identiques aux trois autres modes d'affichage. Les deux seules solutions possibles sont la mise en place d'un tri obligatoire ou l'implémentation de la fonctionnalité de clusterisation de Solr mais aux dépens du classement par pertinence.

Deux nouvelles fonctionnalités peuvent aussi être mises en place : d'une part, en activant la recherche phonétique qui n'a pas été exploitée dans Solr et d'autre part, du côté de l'affichage, la liste des résultats pourrait apparaître dans un nouvel onglet du menu gauche de l'interface feuilleteur.

Par ailleurs, les chercheurs en histoire des sciences et des techniques pourraient être sollicités pour définir les synonymes de certains mots dans le moteur Solr.

Numérisation en mode texte. Ce mode de numérisation n'est pas exempt d'erreurs, car les logiciels d'OCR ont un taux de reconnaissance des caractères variant entre 90 et 97%. Le passage d'un document numérisé en mode image, à

une version en mode texte est loin d'être fidèle à l'original. Le seul moyen d'y parvenir, est de passer par une phase de post correction manuelle pour éliminer les erreurs résiduelles, telles que les caractères mal et non reconnus. Hélas, ce type de traitement n'est pas réalisable, de par les ressources réduites de notre petite équipe. Ce problème est un frein important à ce nouveau service sur lequel s'ajoute le prix de ces logiciels.

Cependant, des raisons d'espérer existent grâce à la folksonomie [PM13]. Depuis l'avènement du Web 2.0, les internautes deviennent des participants actifs du web, et travaillent désormais de manière collaborative sur des projets communs. Par exemple, la plateforme CORRECT du projet de recherche FUI12 Ozalid⁴ de la BNF et Wikisource⁵ de la fondation Wikimedia, élaboré par des contributeurs bénévoles offrent une solution intéressante.

D'autres systèmes de correction de l'OCR existent, comme un système de validation par captcha visuel. Ce système peut être implémenté sur la fonctionnalité téléchargement du pdf 6.3.4.5 et impression de la page du feuilleter. Par respect à l'accessibilité, un captcha auditif devrait être également proposé sur cette page. Il pourrait être étendu vers une procédure de vérification des textes « océrisés », telle que Google l'a implémentée dans [LvAMB08]. Dans ce système, deux mots servent de test, dont l'un provient d'une page d'un livre numérisé qui n'a pas été correctement reconnu. Le second mot ne sera validé comme correct qu'après de nombreuses réponses identiques fournies par les utilisateurs. Cependant, des questions de droits et de brevets entreront peut-être en ligne de compte.

L'amélioration du mode texte des documents représente un enjeu de taille, car la recherche sera d'autant plus efficace si la qualité de ce mode est élevée. Toutefois, ce genre de système de correction collaboratif ne semble pas totalement convaincre les bibliothécaires du Cnum.

Cependant, un autre type de reconnaissance de caractères peut être étudié. En effet, nous pensons aussi que la recherche de motifs de caractères images à la place des mots en caractères, est peut-être une autre piste d'étude plus pertinente que celle de l'OCR.

Interface 3D. Nous n'avons pas implémenté d'interface de visualisation 3D. De nombreux travaux tels que [SCL⁺99], démontrent que les systèmes de visualisation d'une grande complexité, et en particulier en 3D, n'est pas recommandé pour les utilisateurs novices. En effet, elles demandent un temps d'apprentissage et de repérage de l'information plus important en comparaison aux interfaces 2D

4. blog.bnf.fr/gallica/?p=7038, <http://www.reseau-correct.fr>

5. <http://fr.wikisource.org/wiki/Wikisource:Accueil>

actuelles.

Le développement de cette nouvelle interface 3D, qui résoudrait cette problématique est un fascinant défi à relever dans les années à venir. Les travaux de [Top01] offrent une perspective d'étude intéressante en la matière en passant néanmoins du VRML (code en vigueur à l'époque) au HTML5. Ce nouveau format généralisera sans nul doute l'appréhension de cette nouvelle interface, de par son déploiement généralisé due à la compatibilité de portage du code sur tous les navigateurs.

Chapitre 8

Conclusion

Comme nous avons pu le constater, la refonte du site web du Cnum a été un problème complexe à résoudre. Il a fallu prendre en compte le contexte et l'histoire d'une des bibliothèques numériques les plus anciennes dans sa thématique et imaginer comment l'adapter aux usages émergents. Il a fallu également tenir compte du caractère hétérogène d'une bibliothèque numérique composée non seulement de documents mais aussi de métadonnées. Enfin, il a fallu adapter le site à la disparité des écrans et aux nouvelles exigences d'accessibilité.

Comme nous l'avons montré dans notre mémoire, l'interface de consultation d'une bibliothèque numérique idéale n'existe pas. Son appréciation est largement subjective. Cependant, nous avons essayé de proposer une interface de consultation qui soit la plus satisfaisante possible.

Cette refonte du site web s'est appuyée sur une démarche qui comporte les éléments suivants :

- la division du corpus en thématique et typologie
- l'enrichissement des notices qui a entraîné des modifications dans la base de données (thématique, typologie)
- la mise en place de nouvelles fonctionnalités (fil d'Ariane, interface vignette, rotation des pages, navigation inter volumes, amélioration à l'accès des titres dans les listes...)
- un nouveau moteur de recherche avec des outils linguistiques, bien qu'il reste encore à résoudre le problème de la clusterisation des données.
- la refonte du graphisme du site : dans un premier temps, nous avons développé le « wireframe », qui a ensuite été soumis à la validation de tous les membres du groupe de travail constitué pour ce projet. Ensuite, j'ai développé à partir du framework « Cnum 1 » avec le langage de programmation Php et Perl tout le prototype fonctionnel. Ce premier jeu d'essai a donné de la matière aux webdesigners de la société Eclandre, qui ont été en mesure de nous proposer la couche graphique la plus adaptée. De plus, cette société a été force de propositions

pour le graphisme, la conception et l'ergonomie des interfaces. Dans un deuxième temps, l'intégration s'est faite en développement collaboratif avec les deux intégrateurs. Le site répond désormais aux normes du W3C et aux nouveaux critères d'accessibilité définis dans les guides du RGAA et d'Accessiweb. Et dans un troisième temps, j'ai mis en place le moteur de recherche et développé la page dynamique de présentation des résultats.

Les retours d'expérience des utilisateurs sur cette nouvelle version du Cnum sont positifs, ils la trouvent plus esthétique, ergonomique, intuitive et bien adaptée au besoin d'un public de niche. Dans ce projet de refonte du site web du Cnum, nous nous sommes attachés au souci du détail tout en conservant la vision globale du problème. La majorité des besoins ont été pris en compte, et la plupart des fonctionnalités souhaitées ont été mises en place.

Ainsi le site du Cnum a rattrapé son retard et est en conformité avec les normes et les usages actuels. Cependant, dans un domaine où les évolutions sont rapides, il faut veiller à être toujours en adéquation avec les nouvelles pratiques et les évolutions technologiques. Il est difficile de prévoir les nouveaux dispositifs de lecture qui vont émerger à l'avenir comme par exemple les lunettes avec une réalité augmentée ou les montres connectées.

Une autre piste d'amélioration possible serait de faire évoluer le Cnum vers une numérisation en mode texte. En effet, le Cnum proposant uniquement la numérisation en mode image et non en mode texte, pourrait en pâtir à terme. Bien que la technologie actuelle de l'océrisation montre ses limites dans la reconnaissance de caractères sur les ouvrages antérieurs au XIX^e siècle, on obtient des résultats relativement satisfaisants sur des ouvrages postérieurs à cette époque. Nous pouvons espérer que l'évolution des techniques permettra des résultats satisfaisants dans la reconnaissance des caractères d'imprimerie plus anciens ou de type manuscrit.

En conclusion, j'espère avoir rempli ma mission en développant, avec l'accompagnement des équipes du Cnum et de l'agence web Eclydre, une nouvelle sorte de « *belle & artificielle machine* [...] » permettant à toute personne de « *voir & lire une grande quantité de livres, sans se mouvoir d'un lieu* »¹, selon une vision moderne proche de celle de l'ingénieur Agostino Ramelli.

1. Agostino Ramelli, « *Le diverse et artificiose machine* » Paris - 1588, p.316v . Source : Cnum - Conservatoire numérique des Arts et Métiers - Disponible sur <http://cnum.cnam.fr/CGI/fpage.cgi?fDY3/649/100/689/0/0>

Bibliographie

- [ABC09] Almeida, Cheung Bernardoni, and Raczynski Chevalier, Le Van Ho Cubaud. *Le Conservatoire numérique : enjeux et perspectives de la numérisation documentaire*, volume 5, pages 61–72. Documents pour l’histoire des techniques, 2009.
- [ABD10] Ali BEN ABDALLAH. Applications des nuages de mots aux bibliothèques numériques. mise en oeuvre et évaluation. Master’s thesis, CEDRIC Laboratory, Paris, France, 2010.
- [AcAM12] M. Andro, Tosca consultants, E. Asselin, and M. Maisonneuve. *Bibliothèques numériques : logiciels et plateformes*. Sciences et techniques de l’information. ADBS, 2012.
- [BH99] G.A. Baraghimian and G.E. Host. Apparatus and method of fusing the outputs of multiple intelligent character recognition (icr) systems to reduce error rate, oct 19 1999. US Patent 5,970,171.
- [bS09] Livre blanc Smile. *Introduction à l’open source et au logiciel libre*. Open source solution France, 2009.
- [CD03] P. Cubaud and G. Deblock. Le conservatoire numérique des arts et métiers. historique du projet et organisation du site, January 2003. <http://cedric.cnam.fr/fichiers/RC736.pdf>.
- [Cla08] Bernardoni Claire. Étude des parcours individuels de consultation à partir des logs du cnum, 2008. http://cnum.cnam.fr/RUB/usages_cnum.pdf.
- [CM09] Mathias Daval Cécile Moscovitz. *EbookZ ? Etude sur l’offre numérique illégale des livres français sur Internet en 2009*. Enssib, 2009.
- [CT99] P. Cubaud and A. Topol. A www-based digital library for antiquarian collections. Technical Report CEDRIC-99-208, CEDRIC laboratory, CNAM-Paris, France, 1999.
- [CTT98] P. Cubaud, S. Thiria, and A. Topol. Experimenting a 3D Interface for the Access to a Digital Library. In *ACM DL, Pittsburgh, PA, USA*, January 1998. note.
- [Cub06] P. Cubaud. *Du texte au volume : contributions aux bibliothèques numériques*. PhD thesis, Laboratoire Cédric CNAM Paris, CEDRIC La-

- boratory, Paris, France, 2006. Post-Doctoral Degree (Phd students supervisor) Université Paris 8. Jury : M. Saleh, Imad M. Beaudoin-Lafon, Michel M. Le Coadic, Yves M. Quint, Vincent M. Natkin, Stéphane Mme Pélachaud, Catherine M. Salaün, Jean-Michel M. Scholl, Michel <http://cedric.cnam.fr/PUBLIS/RC1093.pdf>.
- [dAATC09] R. Andrade Botelho de Almeida, P. Alessio, A. Topol, and P. Cubaud. A visualisation technique for quality control of massive digitization programs. In *ECDL'09, 13th europ. conf. on res. and adv. tech. for Digital Libraries, Corfu*, LNCS 5714, pages 150–155, January 2009.
- [dAC09] R. Andrade Botelho de Almeida and P. Cubaud. Visualisations de contrôle pour la numérisation massive. In *Actes du douzième colloque international sur le document électronique (CIDE.12), Montréal - Canada*, pages 171–188, January 2009.
- [dAR09] Andrade Botelho de Almeida Rodrigo. *Contribution aux techniques pour enrichir l'espace moteur et l'espace visuel des dispositifs d'interaction bureautique*. PhD thesis, Laboratoire Cédric CNAM Paris, CEDRIC Laboratory, Paris, France, 2009. <http://cedric.cnam.fr/fichiers/RC1883.pdf>.
- [Dup06] J. Dupire. *Interaction 3D pour les documents mis en contexte*. PhD thesis, Laboratoire Cédric CNAM Paris, CEDRIC Laboratory, Paris, France, 2006. <http://www.interaction3d.fr/uploads/Main/theseDupire.pdf>.
- [FL06] Jean-Daniel Fekete and Eric Lecolinet. *Visualisation pour les bibliothèques numériques*, volume 9 of *2006/2*. Lavoisier, 2006.
- [Fou13] Apache Software Foundation. *Apache Solr Reference Guide. Covering Apache Solr 4.6*, 2013.
- [Gri05] Peter Griffiths, Jillian R. ; Brophy. Student searching behavior and the web : Use of academic resources and google. Technical report, illinois digital environment for access to learning and scholarship, 2005.
- [Hea06] Marti Hearst. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*, pages 1–5. Seattle, WA, 2006.
- [Hic52] William E Hick. *On the rate of gain of information*, volume 4, chapter 4, pages 11–26. Quarterly Journal of Experimental Psychology, 1952, <http://dx.doi.org/10.1080/17470215208416600>.
- [Ism08] Ahmada Ismaila. *Moissonneur oai et visualisation de l'information*. Master's thesis, CEDRIC Laboratory, Paris, France, 2008.
- [LA00] Anton Leuski and James Allan. Lighthouse : Showing the way to relevant information. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 125–129. IEEE, 2000.

BIBLIOGRAPHIE

- [Lau07] Guggenheim Laure. Prototype cnum, 2007. <http://projetcnum.free.fr>.
- [Leb09] Marie Lebert. *Une courte histoire de l'ebook*. Net des études françaises), 2009.
- [LRP95] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 401–408, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [LvAMB08] Maurer Luis von Ahn, Abraham McMillen, and Blum. recaptcha : Human-based character recognition via web security measures. Technical report, Science magazine, 2008.
- [MA11] Cécile Touitou Muriel Amar. Enquête d'usages des bibliothèques numériques en sciences et techniques. rapport d'étude. Technical report, Bibliothèque nationale de France, 2011.
- [MBY07] Christian Middleton and Ricardo Baeza-Yates. A comparison of open source search engines. *Journal article*, 2007.
- [MCDA03] Josiane Mothe, Claude Chrisment, Bernard Dousset, and Joel Alaux. Doccube : Multi-dimensional visualisation and exploration of large document sets. *J. Am. Soc. Inf. Sci. Technol.*, 54(7) :650–659, May 2003.
- [MRC91] Jock D. Mackinlay, George G. Robertson, and Stuart K. Card. The perspective wall : Detail and context smoothly integrated. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 173–176, New York, NY, USA, 1991. ACM.
- [NB05] Anuj A Nanavati and Randolph G Bias. Optimal line length in reading—a literature review. *Visible Language*, 39(2) :121–145, 2005.
- [NBL11] J.F. Nogier, T. Bouillot, and J. Leclerc. *Ergonomie des interfaces : Guide pratique pour la conception des applications web, logicielles, mobiles et tactiles*. Dunod, 2011.
- [OCSZ05] Thian-Huat Ong, Hsinchun Chen, Wai-ki Sung, and Bin Zhu. Newsmap : a knowledge map for online news. *Decision Support Systems*, 39(4) :583–597, 2005.
- [PM13] Isabelle Josse Pauline Moirez, Jean Philippe Moreux. *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques*. bnf, 2013.
- [Pol11] S. Polanka. *No Shelf Required : E-books in Libraries*. ALA editions. American Library Association, 2011.
- [Por03] Joshua Porter. Testing the three-click rule. *User Interface Engineering*, 2003.
- [RC03] Bonnie Lida Rogers and Barbara Chaparro. Breadcrumb navigation : Further investigation of usage. *Usability News*, 5(2) :1–7, 2003.

BIBLIOGRAPHIE

- [Roe13] Denis Roegel. La numérisation durable, 2013. <http://locomat.loria.fr/other/2012numerisation-durable.pdf>.
- [Ron11] German Ronan. La valorisation des fonds patrimoniaux de la bibliothèque centrale du conservatoire national des arts et métiers : conception et réalisation d'expositions virtuelles, 2011. <http://www.doyoubuzz.com/ronan-john-german>.
- [Ros10] Guillaume Rossolini. *La recherche Full Text avec Solr. Configurer un moteur de recherche performant à l'aide d'Apache Lucene/Solr et Apache Tomcat*, 2010.
- [RW97] Frank Ruskey and Mark Weston. A survey of venn diagrams. *Electronic Journal of Combinatorics*, 4, 1997.
- [SCL+99] Marc M. Sebrechts, John V. Cugini, Sharon J. Laskowski, Joanna Vasilakis, and Michael S. Miller. Visualization of search results : a comparative evaluation of text, 2d, and 3d interfaces. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 3–10, New York, NY, USA, 1999. ACM.
- [Sin09] Vik Singh. A comparison of open source search engines, 2009.
- [Smi07] Ray Smith. An overview of the tesseract ocr engine. In *ICDAR*, volume 7, pages 629–633, 2007.
- [SN02] Lauren FV Scharff and TX Nacogdoches. Letter identification performance is better for negative contrast than positive contrast. *Vision Sciences Society Annual Meeting*, 2002.
- [TGM83] Edward R Tufte and PR Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.
- [Tol96] Ioannis G. Tollis. Graph drawing and information visualization. *ACM Comput. Surv.*, 28(4es), December 1996.
- [Top01] Alexandre Topol. Vtml : étude, mise en oeuvre et applications. Master's thesis, CEDRIC Laboratory, Paris, France, 2001.
- [Top02] Alexandre Topol. *Interactions 3D pour les paysages informationnels : application aux bibliothèques numériques et à la spécification de comportements*. PhD thesis, Laboratoire Cédric CNAM Paris, CEDRIC Laboratory, Paris, France, 2002. <http://cedric.cnam.fr/fichiers/RC409.pdf>.
- [Vui10] R. Vuillemot. *Un cadre de conception pour la Visualisation d'Information Interactive*. PhD thesis, Université Lyon, laboratoire LIRIS, Lyon, France, 2010. <http://liris.cnrs.fr/Documents/Liris-4820.pdf>.

Annexes

Annexe A

Planning de développement

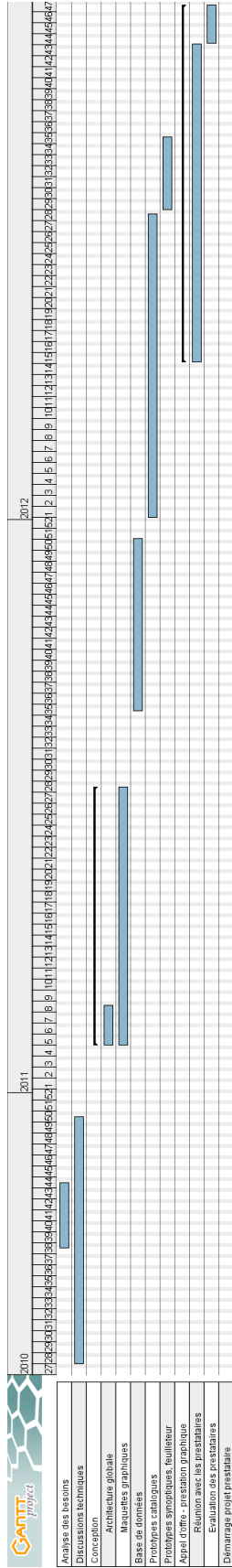


FIGURE A.1 – Planning de développement - Phase 1

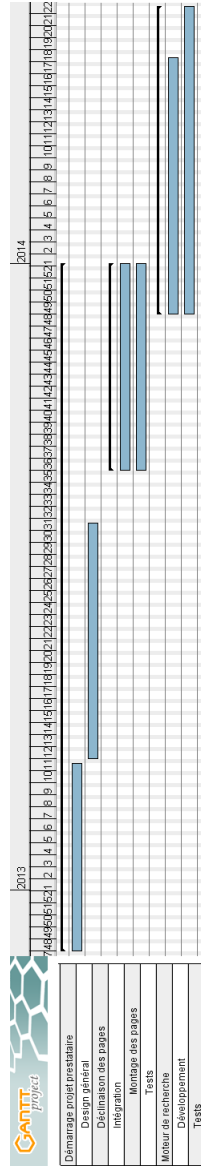


FIGURE A.2 – Planning de développement - Phase 2

Annexe B

Capture d'écran de Cnum 1

Cnum

Le Conservatoire numérique des Arts & Métiers

Une bibliothèque numérique consacrée à l'histoire des sciences et des techniques, constituée à partir du fonds ancien de la bibliothèque du CNAM.

Le CNUM est un projet commun de la [Bibliothèque](#), du [Centre d'histoire des techniques et de l'environnement](#) et du [Centre d'étude et de recherche en informatique](#) du CNAM.

Le CNUM est référencé dans [NUMES](#), inventaire en ligne des corpus numérisés et des projets de numérisation des établissements et organismes de l'Enseignement supérieur et de la Recherche.

TOUTES LES MISES EN LIGNE

Prochainement :
Mercredi 11 juin
2014 en fin
d'après-midi
**Mise en ligne du
nouveau site du
Cnum**

Les dernières mises en ligne

Juin 2013
[Mise en ligne des
Catalogues de
constructeurs et
documents soumis aux
droits d'auteur](#)

17 Novembre 2012
 [FILS RSS](#)

Octobre 2012
Ouvrages sur les
Transports (146 vol.)


Juin 2012
Ouvrages sur la
Construction (82 vol.)


Décembre 2011
[Publications des Expositions
nationales et universelles](#)


Février 2011
[Catalogues des collections
du musée du Conservatoire
\(14 vol.\)](#)


[Milano Città delle Scienze
présente Mondes aquatiques
et marins aux expositions
universelles du 19e siècle](#)


Aide
? Configuration conseillée,
organisation du serveur, droits
d'usage.

Informations
i Partenariats, thèmes, historique et
projets.

**Outils bibliographiques
et compléments éditoriaux**
👉

Recherche plein-texte
🔑 dans les notices des ouvrages et
leurs tables des matières et
d'illustrations

Catalogue
📖 accès aux ouvrages numérisés (par
nom d'auteurs)

FIGURE B.1 – Page d'accueil

Catalogue des ouvrages (par auteurs)

L'icône permet d'accéder la notice descriptive et aux tables d'une monographie. Vous pourrez ensuite accéder aux pages références dans les tables.

L'icône permet d'accéder aux périodiques. L'icône permet d'accéder aux dictionnaires, aux grands traités et aux séries. Vous pourrez ensuite choisir le volume à consulter.

A terme, d'autres catalogues seront disponibles (thèmes, auteurs secondaires).

- Annales du CNAM (1861-1864, 25 vol.)
- Bulletin de la Société d'Encouragement pour l'Industrie Nationale (1802-1943, 166 vol.)
- Cahiers d'histoire du CNAM (1992-1996, 4 vol.)
- Catalogues de constructeurs (21 vol.)
- Catalogues de constructeurs du fonds Bovis (29 vol.)
- Catalogues de constructeurs du fonds Brieux (182 vol.)
- Catalogues des collections du musée du Conservatoire (14 vol.)
- Publications des Expositions nationales et universelles (611 vol.)
- Le Génie industriel (1851-1871, 41 vol.)
- La Lumière électrique (1879-1894, 89 vol.)
- Mémoires de la Société des ingénieurs civils (1848-1925, 114 vol.)

FIGURE B.2 – Page Catalogue

The screenshot shows the CNUM website interface. At the top, there is a navigation bar with icons for home, search, help, and other functions. Below this, the document's metadata is displayed in a light blue box. The metadata includes the author (Terrot et Cie), title (Cycles, motocyclettes, voitures : saison 1914), address (Lyon; Paris), collation (48 pages, 18 cm), and code (CNAM-MUSEE TR0.5-TER). Below the metadata, there is a link to download the PDF version of the document. To the right of the metadata, there are two buttons: 'Table des matières intégrale' and 'Table des illustrations'. Below these buttons, the title 'Table des matières abrégée' is displayed. The table of contents is listed below the title, showing page numbers and corresponding topics.

Auteur(s)	Terrot et Cie (Etablissements)
Titre	Cycles, motocyclettes, voitures : saison 1914
Adresse	Lyon; Paris : Impr. B. Arnaud, 1914
Collation	48 p. : ill. ; 18 cm
Cote	CNAM-MUSEE TR0.5-TER
Mot(s) matière(s)	Motocyclettes*France ; Voitures*France ; Cycles

[Télécharger la version PDF de ce volume \(2.5Mo\)](#)

[Table des matières intégrale](#) [Table des illustrations](#)

Table des matières abrégée

- [PAGE DE TITRE \(Première image\)](#)
- [48](#) Répertoire
- [2](#) Avertissement
- [3](#) A propos des motorettes
- [4](#) Palmarès de la moterette Terrot 1913
- [20](#) Bicyclettes
- [37](#) Changement de vitesse n° 2 (moyeu 2 vitesses)
- [37](#) Changement de vitesse n° 3 (moyeu 3 vitesses)
- [8](#) Changement de vitesse pour moterette
- [9](#) Changement de vitesse pour moterette
- [43](#) Conseils et table de développements
- [7](#) Débrayage moterette
- [6](#) Fourche élastique
- 38 Freins

FIGURE B.3 – Page synoptique d'une monographie, d'un volume de périodique ou d'une grande monographie

The screenshot shows the CNUM website interface. At the top, there is a navigation bar with icons for home, search, help, and other functions. The main content area is divided into two sections: metadata and a table of contents.

Metadata Section:

- Auteur(s):** Terrot et Cie (Etablissements)
- Titre:** Cycles, motocyclettes, voitures : saison 1914
- Adresse:** Lyon; Paris : Impr. B. Arnaud, 1914
- Collation:** 48 p. : ill. ; 18 cm
- Cote:** CNAM-MUSEE TR0.5-TER
- Mot(s) matière(s):** Motocyclettes*France ; Voitures*France ; Cycles

Below the metadata, there is a link to download the PDF version: [Télécharger la version PDF de ce volume \(2.5Mo\)](#).

Table of Contents Section:

The table of contents is titled "Table des matières abrégée" and lists the following items:

- [PAGE DE TITRE \(Première image\)](#)
- [48](#) Répertoire
- [2](#) Avertissement
- [3](#) A propos des motorettes
- [4](#) Palmarès de la moterette Terrot 1913
- [20](#) Bicyclettes
- [37](#) Changement de vitesse n° 2 (moyeu 2 vitesses)
- [37](#) Changement de vitesse n° 3 (moyeu 3 vitesses)
- [8](#) Changement de vitesse pour moterette
- [9](#) Changement de vitesse pour moterette
- [43](#) Conseils et table de développements
- [7](#) Débrayage moterette
- [6](#) Fourche élastique
- [38](#) Freins

At the bottom of the page, there are two buttons: "Table des matières intégrale" and "Table des illustrations".

FIGURE B.4 – Page synoptique d'une monographie avec plusieurs tomes

Auteur(s) secondaire(s)	Académie des sciences (France), éd. Bertrand, Jean-Elie, éd.
Titre	Descriptions des arts et métiers, faites ou approuvées par Messieurs de l'Académie royale des sciences de Paris
Adresse	A Neuchâtel : de l'imprimerie de la Société typographique, 1771-1783 ; à Paris : chez calixte Volland, an VII [1798]
Edition	Nouvelle édition publiée avec des observations, & augmentée de tout ce qui a été écrit de mieux sur ces matières, an Allemagne, en Angleterre, en Suisse, en Italie
Description	20 vol. : ill. gr. s. c. ; 25 cm
Cote	CNAM 4° Ky 58
Mot(s) matière(s)	Métiers*18e siècle ; Artisanat*18e siècle

Liste des volumes

- ☞ Tome I. L'art du meunier, du boulanger, du vermicellier.
- ☞ Tome II. Les quatre premières sections sur les fers, et l'art du charbonnier.
- ☞ Tome III. L'art du tanneur, du chamoiseur, du mégissier, du corroyeur, du parcheminier, de l'hongroyeur, du maroquinier, de travailler les cuirs dorés & argentés, du cordonnier, du paumier-raquetier [sic raquetier] & de la paume.
- ☞ Tome IV. L'art du tuilier & du briquetier, l'art de tirer des carrières la pierre d'ardoise, de la fendre de la tailler, l'art du couvreur, l'art du chaudfourmier, l'art de faire le papier, l'art du cartonnier, et l'art du cartier.
- ☞ Tome V. Les trois premières sections du traité des pêches & l'histoire des poissons.
- ☞ Tome VI. L'art du serrurier, l'art du chandelier, & l'art d'exploiter les mines de charbon de terre.
- ☞ Tome VII. L'art de la draperie, l'art de friser ou ratatiner les étoffes de laine, l'art de faire les tapis, façon de Turquie, l'art du chapelier, l'art du tonnelier, l'art de convertir le cuivre en laiton, & l'art de l'épinglier.
- ☞ Tome VIII. L'art de l'indigotier, l'art de la porcelaine, l'art du potier de terre, l'art de faire les pipes, l'art de faire les colles, fabrique de l'amidon, l'art du savonnier, & l'art du relieur.
- ☞ Tome IX. Les six premières parties de l'art du fabricant d'étoffes de soie.
- ☞ Tome X. Les deux premières sections de la seconde partie du traité des pêches.
- ☞ Tome XI. La troisième section de la seconde partie du traité des pêches.
- ☞ Tome XII. L'art du distillateur d'eaux-fortes, l'art du distillateur liquoriste, & l'art du vinaigrier ; avec des notes & des additions.
- ☞ Tome XIII. L'art de la peintures sur verre & de la vitreie ; l'art du plombier & fontainier.

FIGURE B.5 – Page synoptique d'une grande monographie



Titre	La Nature. Revue des sciences et de leurs applications aux arts et à l'industrie. Suivi de : Bulletin météorologique de La Nature, Boîte aux lettres, Nouvelles scientifiques
Mention(s) de responsabilités	TISSANDIER, Gaston , réd. en chef ; PARVILLE, Henri de, dir.
Adresse	Paris : Masson, 1873-1962
Collation	153 vol. + 3 vol. de tables + 1 vol. de couvertures (année 1916): ill. ; 27 cm
Cote	CNAM 4° Ky 28
Mot(s) matière(s)	Sciences * Vulgarisation ; Culture scientifique et technique ; Presse scientifique
Note	A pour suppléments : Bulletin météorologique de La Nature (1889-1891) Boîte aux lettres (1889-1891) Nouvelles scientifiques (1891-1905)- Tables cumulatives 1873-1882 - Tables cumulatives 1883-1892 - Tables cumulatives 1893-1902

Liste des volumes

- 1873 : Première année : n°1 à 26
- 1874 : Deuxième année, premier semestre : n° 27 à 52
- 1874 : Deuxième année, deuxième semestre : n° 53 à 78
- 1875 : Troisième année, premier semestre : n° 79 à 104
- 1875 : Troisième année, deuxième semestre : n° 105 à 130
- 1876 : Quatrième année, premier semestre : n° 131 à 156
- 1876 : Quatrième année, deuxième semestre : n° 157 à 182
- 1877 : Cinquième année, premier semestre : n° 183 à 208
- 1877 : Cinquième année, deuxième semestre : n° 209 à 234
- 1878 : Sixième année, premier semestre : n°235 à 260
- 1878 : Sixième année, deuxième semestre : n°261 à 287
- 1879 : Septième année, premier semestre : n° 288 à 313
- 1879 : Septième année, deuxième semestre : n°314 à 339
- 1880 : Huitième année, premier semestre : n°340 à 365
- 1880 : Huitième année, deuxième semestre : n°366 à 391
- 1881 : Neuvième année, premier semestre : n°392 à 417

FIGURE B.6 – Page synoptique d'un périodique

NOS BICYCLETTES

.....

Nos cadres sont établis pour quatre tailles différentes, avec des hauteurs de 52, 57, 62 et 67 centimètres, comptés du centre de l'axe du pédalier à l'orifice du tube qui reçoit la tige de selle. Tous nos cadres sont renforcés.

Les JANTES en forme W, émaillées noir, nickelées ou demi-nickelées. Nous livrons aussi la jante bois Kuntz, la meilleure des jantes en bois. (Suppléments, voir page 42). Les ROULEMENTS de tous nos modèles sont bien clos et à l'abri de la poussière. Les GUIDONS peuvent être choisis à volonté d'après notre tableau page 40. Le serrage du guidon est intérieur. La CHAÎNE est à petits rouleaux avec pignons de 54, 50 ou 46 X 16, 18, 20, 22 et 24 dents, suivant le modèle.

.....

NOS PNEUMATIQUES

.....

Tous à tringles, sont de première qualité comme gomme, toile et facture. Ils sont munis de bonnes chambres à air. Les modèles de route sont montés avec pneus dits de 38 %, mais nous les livrons aussi avec les grosseurs dites 35, 42 et 45 %. — Les

FIGURE B.7 – Page feuilleteur

Recherche

Vous avez ici la possibilité de localiser des occurrences de mots dans l'ensemble de toutes les tables de matières, tables d'illustrations et index des ouvrages du CNUM, ainsi que dans leurs notices bibliographiques. Merci de noter qu'il ne s'agit pas du texte des ouvrages eux-mêmes : nous ne disposons malheureusement pas de ces données.

Avant d'effectuer une recherche, consultez peut-être le [catalogue](#) du Conservatoire Numérique pour avoir une idée du fonds (uniquement consacré à l'histoire des sciences et des techniques).

Rechercher :

dans :

- un de ces mots
- mots à la suite
- mots à proximité
- le début de mot
- la fin de mot

pour :

- les titres du CNUM
- les monographies du CNUM
- les périodiques du CNUM
- les titres de *Gallica* (uniquement dans les notices)

LANCER LA RECHERCHE

La recherche s'effectue sans tenir compte de la casse des caractères. Rechercher `galvani`, `Galvani` ou `GALVANI` revient au même.

En revanche, la recherche tient compte de l'accentuation des caractères, ce qui n'est pas très adapté à la recherche dans des textes anciens. Aussi, proposez plusieurs variantes d'accentuation pour les mots recherchés (option "un de ces mots").

L'orthographe ancienne a été maintenue. Nous avons établi un [glossaire](#) des variantes de noms propres et communs rencontrés lors de la saisie des tables de matières. N'hésitez pas à le consulter avant de formuler vos requêtes.

Une alternative consiste à utiliser un moteur externe, en limitant la recherche au CNUM :

Rechercher dans le [Dictionnaire Prosopographique des Inventeurs de France](#)

FIGURE B.8 – Page de recherche avancée

Recherche - résultats

Revenir à la [recherche plein-texte](#)

Le lien [+] permet d'approfondir la recherche.

L'icône [] permet d'accéder à la notice d'une monographie.

L'icône [] permet d'accéder à la notice d'une série.

L'icône [] permet d'accéder à la notice d'un périodique.

Recherche de avion dans toutes les tables des matières

212 occurrences trouvées dans 9 titres :

- 169 [\[+\]](#) [\[\]](#) La Nature
- 26 [\[+\]](#) [\[\]](#) Bulletin de la Société d'Encouragement pour l'Industrie Nationale
- 7 [\[+\]](#) [\[\]](#) Toussaint, Albert-Charles-Louis
- L'aviation actuelle
- 4 [\[+\]](#) [\[\]](#) Lainé, André
- Dictionnaire de l'aviation
- 2 [\[+\]](#) [\[\]](#) Boileau, Charles
- Le moteur à essence adapté à l'automobile et à l'aviation
- 1 [\[+\]](#) [\[\]](#) Percheron, Maurice (18...-19... ; ingénieur)
- Guide du mécanicien d'aviation. Tome I. Le moteur
- 1 [\[+\]](#) [\[\]](#) Rives, Gustave (1858-1926)
- Rapport sur le premier salon de l'aéronautique : Grand Palais, Paris, décembre 1908
- 1 [\[+\]](#) [\[\]](#) Percheron, Maurice (18...-19...)
- Nos avions. L'essor et l'atterrissage (ou l'amerrissage)
- 1 [\[+\]](#) [\[\]](#) Bréguet, Louis (1880-1955)
- Stabilité longitudinale des avions

La recherche a pris 15 seconde(s)

FIGURE B.9 – Page de résultat

Annexe C

Capture d'écran de Cnum 2 Wireframe

Cnum
Conservatoire numérique des Arts et Métiers
A propos du Cnum Catalogue général

Rechercher dans le Cnum OK Recherche avancée

Il y a 177/2012 dans le Cnum 734 titres, 1 737 volumes, 700 075 pages, 1 542 ouvrages numérisés téléchargeables

Bibliothèque numérique en histoire des sciences et des techniques

Thématiques	Liens	Titres
Catalogues de constructeurs	Expositions universelles	Technologies de l'information
179 titres	347 titres	45 titres
Construction	Histoire du Cnum	Transports
115 titres	17 titres	20 titres
Énergie	Machines et instrumentation	Généralités scientifiques et vulgarisation
75 titres	213 titres	13 titres
Typologies	Ouvrages	Revue
Manuscrits et dessins	719 titres	15 titres
3 titres		

Nouveautés

- [Catalogues de constructeurs](#)
- [Publications des expositions nationales et universelles](#)
- [LA NATURE](#)
- [La Nature \(1873-1962\)](#)

Expositions virtuelles

- [Palais, pavillons et galeries](#)
- [Vélocipèdes, bicyclettes...](#)
- [Expos.3](#)
- [Le Cnum sur Facebook](#)
- [Flux RSS](#)

Le CNUM est un projet commun de la Bibliothèque, du Laboratoire Histoire, technique, technologie et patrimoine et du Centre d'étude et de recherche en informatique et communications du CNAM.

[Contact](#) [Plan du site](#) [Bibliothèque du Cnum](#) [Musée des Arts et Métiers](#) [Cédric Hitt](#) [Hitt](#) [Nunes](#) [Patrimoine numérique](#)

FIGURE C.1 – Page d'accueil





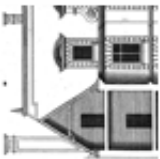









Thématique et type de documents	
 <p>Généralités 29 titres</p>	 <p>Énergie 74 titres</p>
 <p>Transports 19 titres</p>	 <p>Technologies de l'information 45 titres</p>
 <p>Construction 43 titres</p>	 <p>Machines et instrumentation 208 titres</p>
 <p>Expositions universelles 347 titres</p>	 <p>Revue 15 titres</p>
 <p>Catalogues de constructeurs 179 titres</p>	 <p>Manuscrits et dessins 3 titres</p>
 <p>Catalogue des Revues 15 titres</p>	 <p>Logo catalogue des Monographies 645 titres</p>
 <p>Logo catalogue des Manuscrits et dessins 3 titres</p>	 <p>Manuscrits et dessins 3 titres</p>

FIGURE C.2 – Page d'accueil

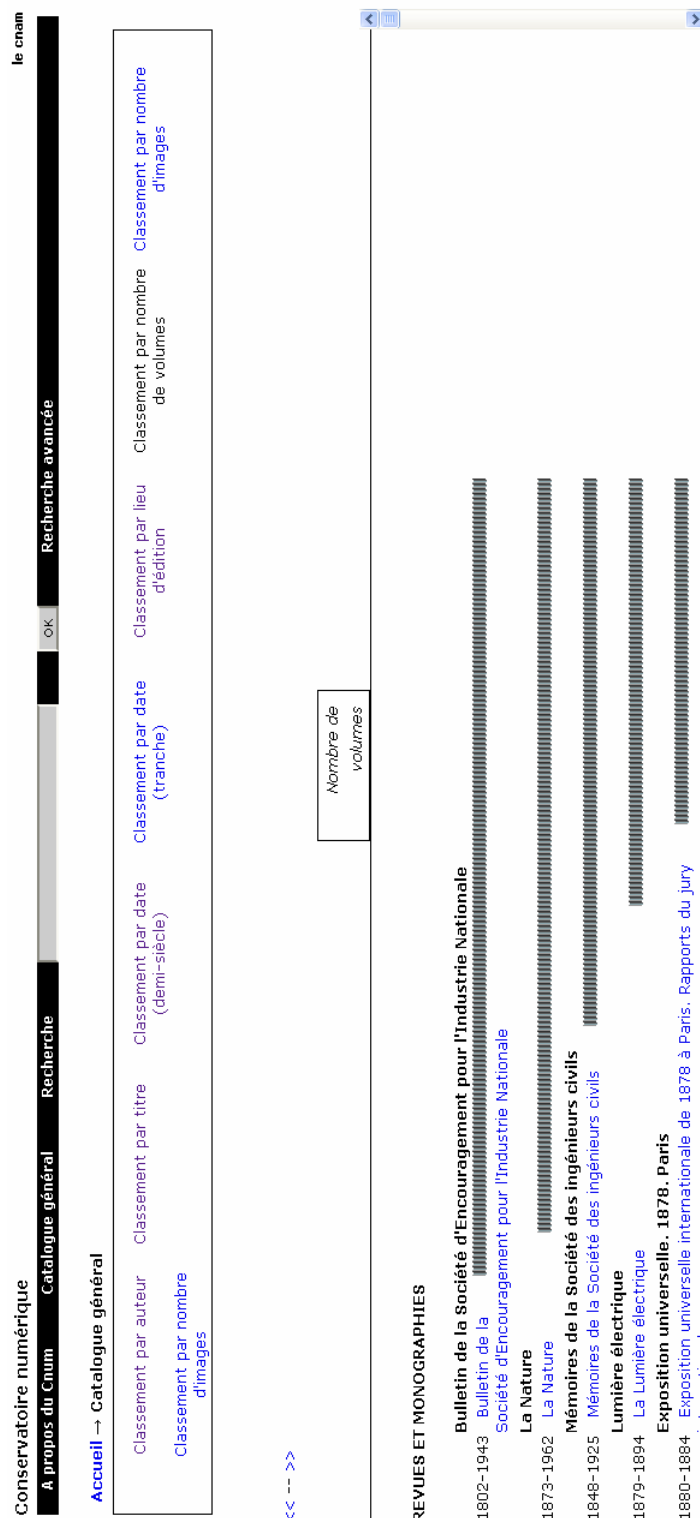


FIGURE C.3 – Page catalogue avec les volumes

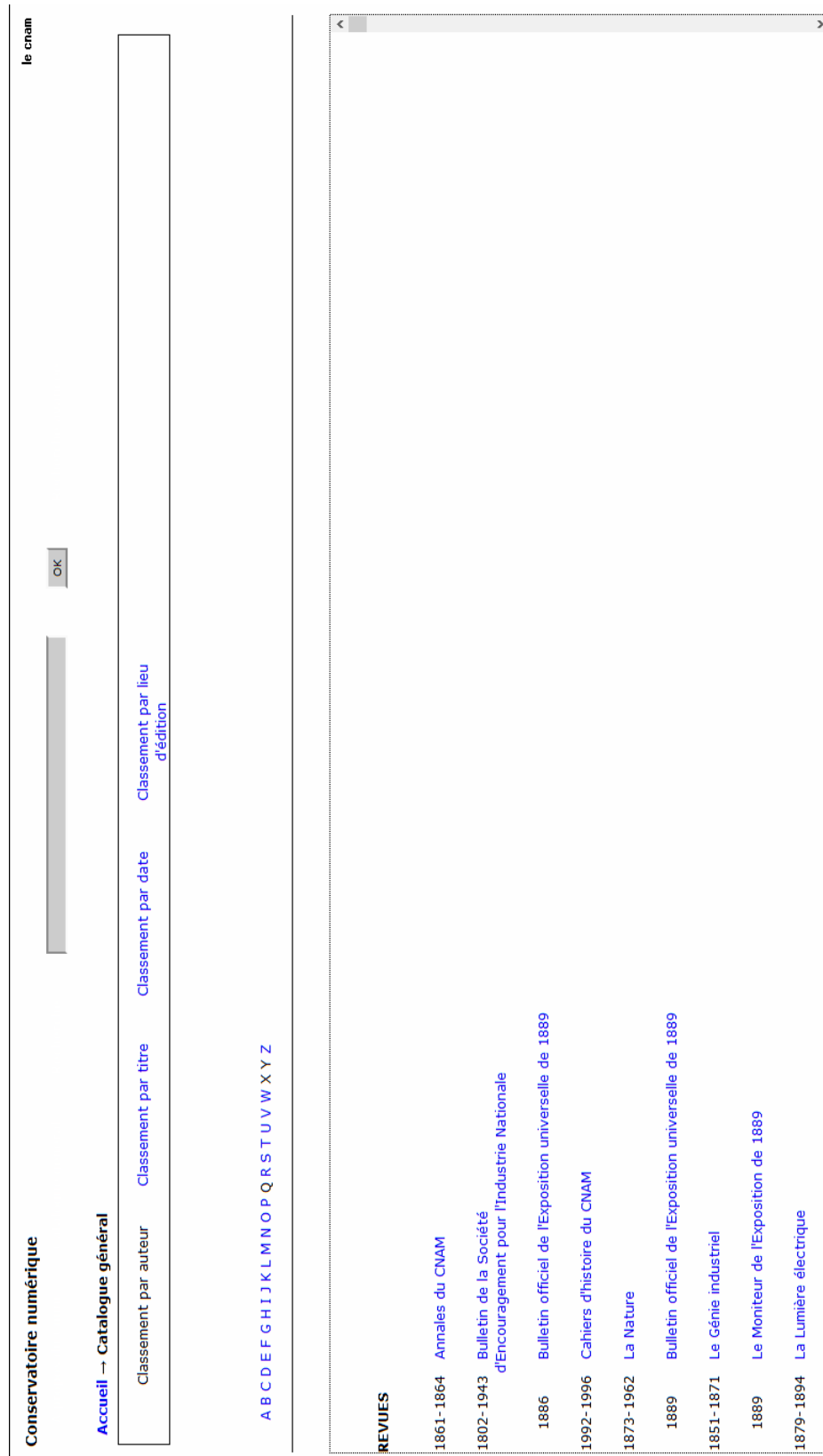


FIGURE C.4 – Page catalogue

le cnum
Recherche avancée
OK

A propos du Cnum
Catalogue général
Recherche

Accueil → **Catalogue général** → **L'Architecture pratique qui comprend le détail du toisé**

Titre L'Architecture pratique qui comprend le détail du toisé, & du devis des ouvrages de maçonnerie [maçonnerie], charpenterie, menuiserie, serrurerie, plomberie, vitrerie, ardoise, tuile, pavé de grès & impression. Avec une explication de la coutume sur le titre des servitudes & rapports qui regardent les Bâtimens [bâtimens]. Ouvrage très nécessaire aux Architectes, aux Experts, & à tous ceux qui veulent bâtir par M. Bullet, Architecte du Roy, & de l'Académie Royale d'Architecture

Auteur Bullet, Pierre

Notice détaillée
[Table des matières](#)
[Table des illustrations](#)
[Exportation du document](#)
[Rechercher dans ce titre](#)

Auteur(s) Bullet, Pierre

Auteur(s) secondaire(s) Levesque, ill.

Titre L'Architecture pratique qui comprend le détail du toisé, & du devis des ouvrages de maçonnerie [maçonnerie], charpenterie, menuiserie, serrurerie, plomberie, vitrerie, ardoise, tuile, pavé de grès & impression. Avec une explication de la coutume sur le titre des servitudes & rapports qui regardent les Bâtimens [bâtimens]. Ouvrage très nécessaire aux Architectes, aux Experts, & à tous ceux qui veulent bâtir par M. Bullet, Architecte du Roy, & de l'Académie Royale d'Architecture

Adresse A Paris : chez J. B. Delespine et Jean-Th. Herissant, 1741

Collation [16]-392-[18] p. : ill., frontispice (grs.b. et s.c.),in-8

Cote CNUM-8B 8 L 1 Res

Sujet(s) Habitations * Conceptions et construction ; Architecture * France * 18e siècle

Url <http://cnum.cnam.fr/redir?8L1>

permanent

FIGURE C.5 – Page synoptique 1

le cnam

Conservatoire numérique
Catalogue
Recherche
Recherche avancée

A propos du Cnum
Recherche
Recherche avancée

Accueil
→ Construction
→ Manière de bastir pour toutes sortes de personnes

Titre : Manière de bastir pour toutes sortes de personnes

Auteur : Pierre Le Muet

Notice détaillée

Table des matières

Table des matières détaillée

Table des illustrations

Vignettes

Exportation du document

Rechercher dans ce titre

Auteur(s)	Le Muet, Pierre (1591-1669)
Titre	Manière de bastir pour toutes sortes de personnes par Pierre Le Muet, architecte ordinaire du Roy et conducteur des desseins des fortifications en la province de Picardie dédiée au Roy
Adresse	A Paris : chez Melchior Tavernier, 1623
Edition	1ère éd.
Collation	[4]-114 p. : fig. gr. ; in fol.
Cote	CNAM Fol Ko 5
Mots matière	Construction - France - 17e siècle Architecture domestique - France - 17e siècle
Note	Relié parchemin. Ex libris ms "Collegi Paris.Soc. Jesu". Melchior Tavernier, graveur imprimeur du Roy pour les tailles douces demeurant en Liste du Palais sur le quay qui regarde la Mégisserie au coing de la rue du Harlay à la Rose Rouge
Uri permanente	http://cnum.cnam.fr/redir?FOLKO5

FIGURE C.6 – Page synoptique 2

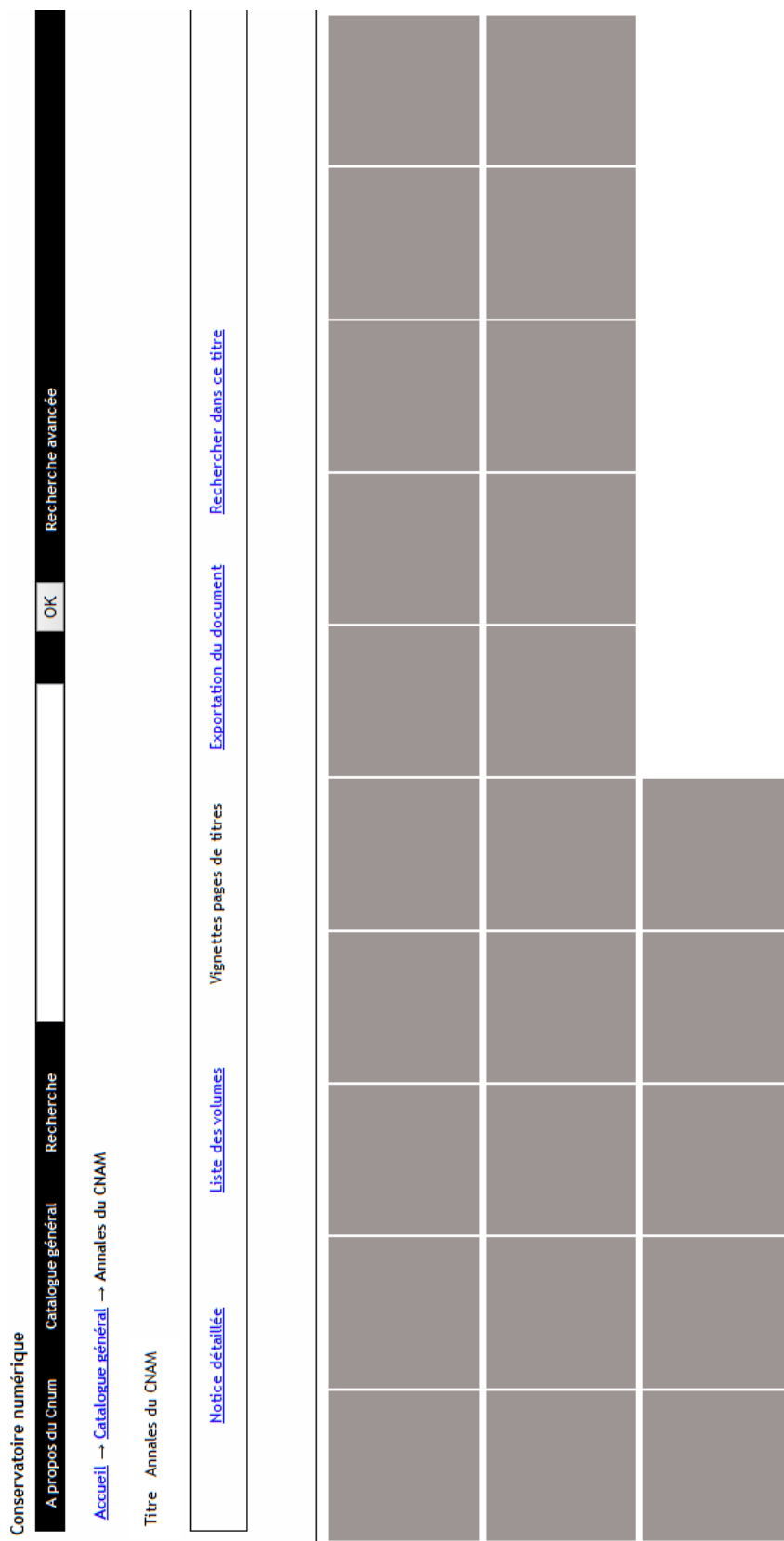


FIGURE C.7 – Page vignettes

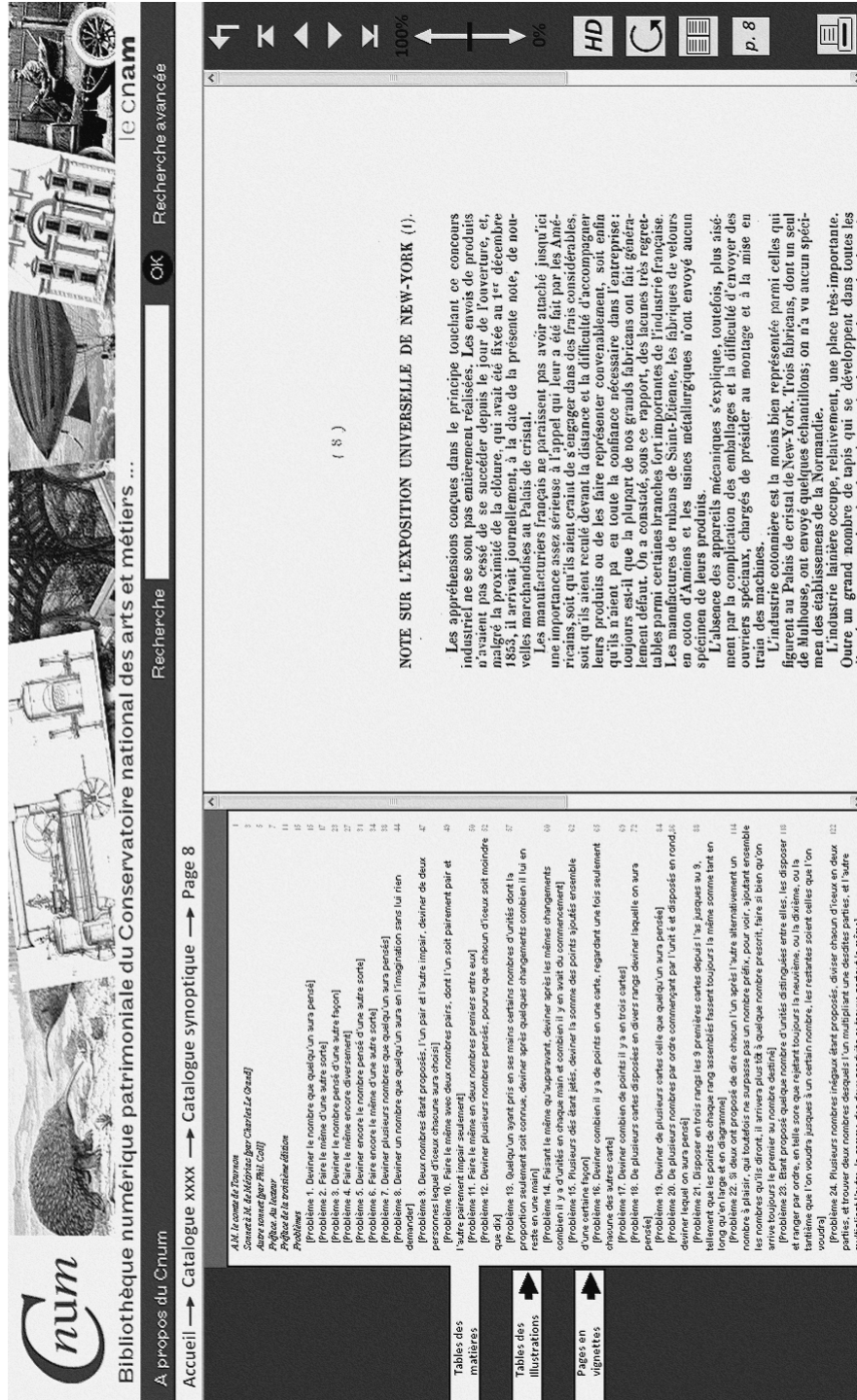
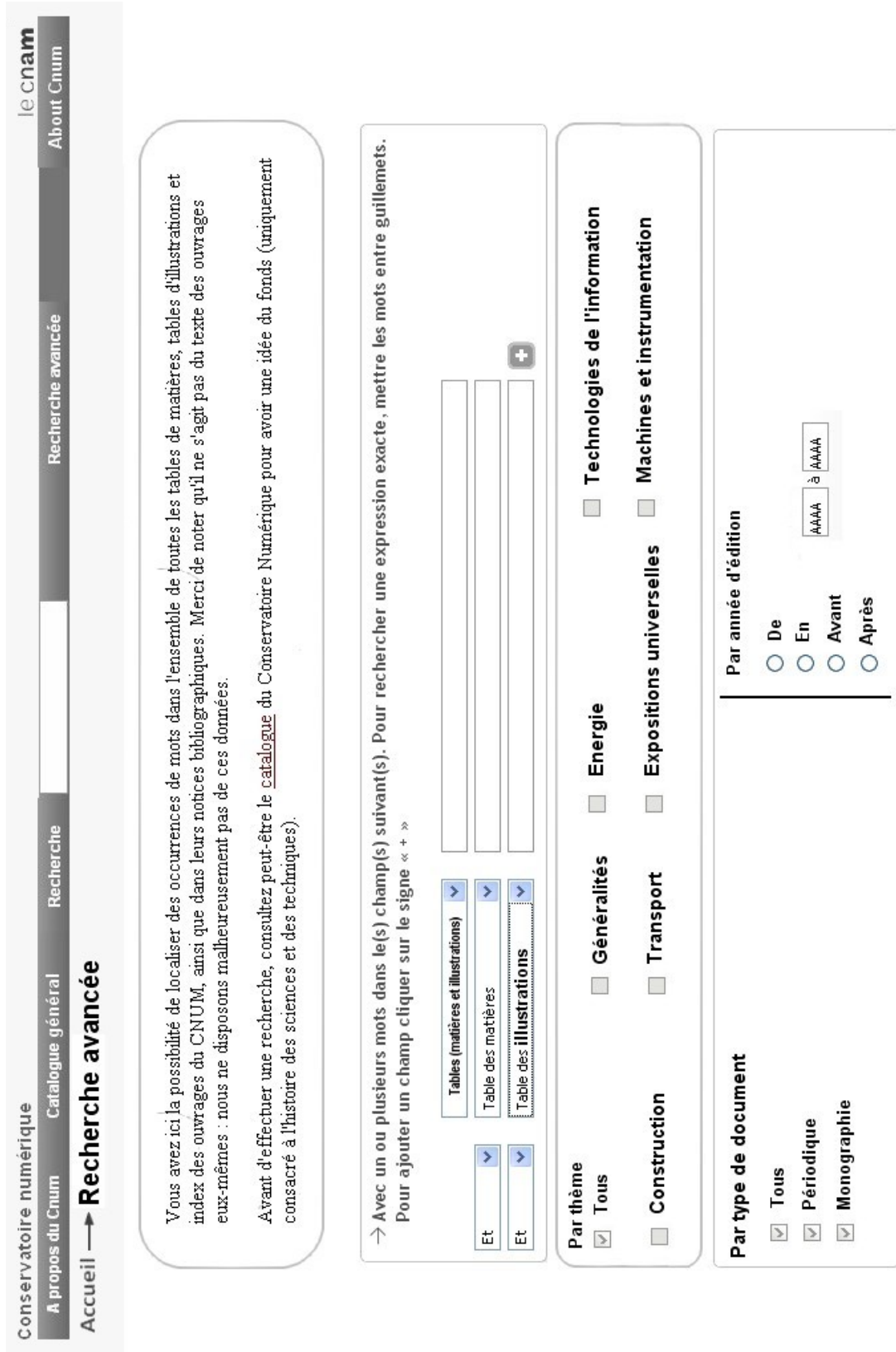


FIGURE C.8 – Page feuilleteur - maquette graphique



LANCER Effacer

FIGURE C.9 – Page recherche

Recherche - résultats (fin)

[Revenir à la recherche plein-texte](#)

Titre examiné :

■ La Nature (4KY28)

Recherche de oiseau dans les tables des matières

4KY28.1 [72](#) L'**oiseau** moqueur

4KY28.2 [194](#) L'**oiseau** à dents de l'argile de Londres (E. OUSTALET)

4KY28.11 [209](#) Nouvel **oiseau** fossile

4KY28.16 [39](#) L'architecture des oiseaux. Le nid du Lorient de Ballinore, le nid du Glaucis hirsuta, le nid de l'**Oiseau**-mouche huppécol (E. OUSTALET)

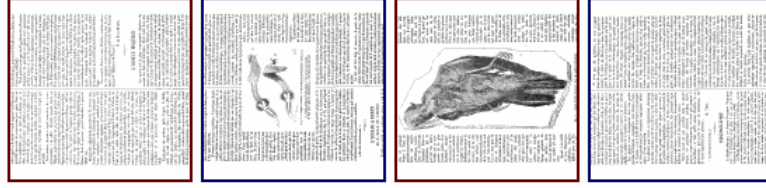


FIGURE C.10 – Page résultats vignette

Annexe D

Capture d'écran de Cnum 2

D.1 Page d'accueil

D.1. PAGE D'ACCUEIL

leCnum
Conservatoire numérique des Arts et Métiers

Bibliothèque numérique en histoire des sciences et des techniques
Au 30/06/2014, le Cnum contient 989 titres, 2063 volumes, 692 098 pages,
2038 unités téléchargeables

À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Le site du Cnum fait peau neuve ! Merci d'adresser vos remarques et questions à cnum@cnam.fr

Thématiques

<p>Catalogues de constructeurs 217 titres</p>	<p>Expositions universelles 406 titres</p>	<p>Technologies de l'information et de la communication 52 titres</p>
<p>Construction 117 titres</p>	<p>Histoire du Cnam 17 titres</p>	<p>Transports 105 titres</p>
<p>Energie 102 titres</p>	<p>Machines et instrumentation scientifique 106 titres</p>	<p>Généralités scientifiques et vulgarisation 13 titres</p>

Nouveautés

JUN 2014

Mise en ligne du musée centennial et des expositions universelles

Expositions

PROCHAINEMENT

Palais, pavillons et galeries : les bâtiments des expositions universelles en représentation (1775-1900)

Typologies

<p>Manuscrits 3 titres</p>	<p>Ouvrages 973 titres</p>	<p>Revue 16 titres</p>
--------------------------------	--------------------------------	----------------------------

leCnum Le Cnum est né d'un projet commun de la Bibliothèque Centrale, du Centre d'histoire des techniques et de l'environnement et du Centre d'étude et de recherche en informatique et communications du Cnam. Le centre de documentation du musée des arts et métiers est également associé à la politique de numérisation. le cnam

FIGURE D.1 – Page d'accueil

D.2 Page thématique

leCnum
À PROPOS
CATALOGUE GÉNÉRAL
RECHERCHER DANS LE CNUM
RECHERCHE AVANCÉE

Accueil > Technologies de l'information et de la communication

Technologies de l'information et de la communication

Le XIXe siècle est une phase d'innovation très soutenue dans le domaine des techniques de communication, grâce à la maîtrise conjointe de la mécanique, de l'électricité et de la chimie. Les plus emblématiques en sont la photographie, le téléphone, le cinéma, la radio, après le télégraphe ou encore la presse à grande diffusion. Communiquer à distance instantanément, saisir l'instant en un cliché, diffuser une information à l'échelle mondiale, tels sont les enjeux. Ces techniques se perfectionnent tout au long du XXe siècle pour converger avec l'informatique et le numérique aujourd'hui. Au fil de ces mutations, les techniques de l'information et de la communication ont transformé notre monde, jusqu'à notre quotidien.

Les ouvrages numérisés dans le Cnum donnent une juste idée du rythme de l'innovation au XIXe siècle, des difficultés et des impasses dans le foisonnement d'inventions, des enjeux du monde contemporain attenants à ces techniques.

Revue **A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**

AUTEUR	TITRE	DATE	LIEU D'ÉDITION	ÉDITEUR
Ancel, Louis	Appareils & instruments de précision pour les sciences de l'industrie : catalogue général illustré n° 1 (1 vol.)	1979	Fontenay-Le-Comte	Impr. Commerciale BC
Anthony, E. & H. T. & Co	Illustrated catalogue of photographic equipments and materials for amateurs (1 vol.)	1891	New York	Polhemus, John
Bréguet, Louis-François-Clément	Télégraphie électrique : son avenir, poste aux lettres électrique, journaux électriques (1 vol.)	1849	Paris	Mathias, Augustin (Librairie scientifique industrielle)
Calmels, H. (Maison)	Fournitures pour les procédés photomécaniques, la photographie industrielle, scientifique (1 vol.)	1907	Versailles	Société anonyme des imprimeries Gérardin







FIGURE D.2 – Une des pages thématique

D.3 Tri dans le catalogue général

D.3. TRI DANS LE CATALOGUE GÉNÉRAL

leCnum		À PROPOS		CATALOGUE GÉNÉRAL		RECHERCHER DANS LE CNUM		RECHERCHE AVANCÉE			
Accueil > Catalogue général		A B C D E F G H I J K L M N O P Q R S T U V W X Y Z		TITRE		DATE		LIEU D'ÉDITION		ÉDITEUR	
	Revue	Annales du Conservatoire des arts et métiers (27 vol.)		1861-[193.]	Paris	Lacroix, Eugène (Librairie scientifique, industrielle et agricole)					
	Revue	Bulletin de la Société d'Encouragement pour l'Industrie Nationale (166 vol.)		1802-1943	Paris	Société d'encouragement pour l'industrie nationale (France)					
	Revue	Bulletin officiel de l'exposition universelle de 1889 (3 vol.)		1886	Paris	[s.n.]					
	Revue	Les Cahiers d'histoire du CNAM (4 vol.)		1992-1996	Paris	Conservatoire national des arts et métiers (France)					
	Revue	La Lumière électrique (89 vol.)		1879-1894	Paris	Union technique de l'électricité (France)					
	Revue	La Nature (157 vol.)		1873-1962	Paris	Masson					
	Revue	Bulletin officiel de l'exposition universelle de 1889 (2 vol.)		1889	Paris	[s.n.]					
	Revue	Le Génie industriel (41 vol.)		1851-1871	Paris	Armengaud, Eugène (1810-1891)					
	Revue	Le Moniteur de l'exposition de 1889 (5 vol.)		1889	Paris	Le Moniteur de l'exposition					


FIGURE D.3 – Catalogue général - classement par auteur

D.3. TRI DANS LE CATALOGUE GÉNÉRAL

AUTEUR	TITRE	DATE	LIEU D'ÉDITION	ÉDITEUR
Carlès, F.	Les accessoires de l'automobile (1 vol.)	1913	Paris	Dunod, Henri (1865-1946)
Carafoli, Elie	Aérodynamique des ailes d'avion : théorie et applications (1 vol.)	1928	Paris	Chiron, Étienne (Librairie aéronautique)
Gaston, Raymond de	Les avions de 1910 (1 vol.)	[1911]	Paris	Librairie aéronautique
Nansouty, Max de	Aérostation. Aviation (1 vol.)	1911	Paris	Boivin et Cie
Grilleau, B. de.	Les aérostats dirigeables : leur passé, leur présent, leur avenir. Le ballon de Meudon et les progrès les plus récents de l'aéronautique (1 vol.)	1884	Paris	Dentu, Edouard (1830-1884)
Société Lumière	Agenda Lumière 1938 (1 vol.)	1933	Paris	Gauthier-Villars (Librairie)
Schlumberger, Victor	Aide-mémoire pratique de tissage mécanique et en particulier du tissage du coton (1 vol.)	[1896]	Paris	Baudry (Librairie polytechnique)
Georges, O. (Maison)	Album des accessoires pour chaudières, machines et chauffages à vapeur (1 vol.)	1909	Paris	Céas, Jules et fils (Imprimerie)
École Centrale d'architecture (Paris)	Lamphithéâtre en 1865-66. Leçons d'ouverture (1 vol.)	[1857]	Paris	Morel, A.
Leitz, Ernst	Anleitung zum Gebrauche der Mikroskope aus der optischen Werkstatt (1 vol.)	[1892]	Limburg	Goerlach (Gebrüder)

FIGURE D.4 – Catalogue général - classement par titre

D.3. TRI DANS LE CATALOGUE GÉNÉRAL



[À PROPOS](#) [CATALOGUE GÉNÉRAL](#)

[RECHERCHER DANS LE CNUM](#)

[RECHERCHE AVANCÉE](#)

Accueil > Catalogue général

XVIIe XVIIIe XIXe XXe Avant 1750 1750-1850 1850-1950 Après 1950

AUTEUR	TITRE	DATE	LIEU D'ÉDITION	ÉDITEUR
Vitruve	M. Vitruvius per Jocundum solito castigatior factus cum figuris et tabula ut jam legi et intelligi possit (1 vol.)	1511	Venise	Tridino, J. de
Végèce (0383?-0450?)	De re militari (1 vol.)	1535	Paris	Wechel, Chrétien (1495-1554)
Valturio, Roberto	Les douze livres touchant la discipline militaire (1 vol.)	1555	Paris	Périer, Charles (152.-1572?)
Vitruve	M. Vitruvii Pollionis De architectura libri decem, cum commentariis Danielis Barbari (1 vol.)	1567	Venise	De Franceschi, Francesco (1500-1599)
[S.n.]	[Le « Recueil des Arts et Métiers » : 128 dessins au crayon légendés ; avec inventaire manuscrit daté du 8 mars 1876 et notes manuscrites non datées] (1 vol.)	[157.]	[S.l.]	[s.n.]
Vinet, Elie	L'Arpenterie. Livre de geometrie, enseignant à mezzurer les champs, et pluzieurs autres chozes (1 vol.)	1577	Bordeaux	Millanges, Simon (1540?-1623)
Chauvet, Jacques	Instruction et usage du cosmometre (1 vol.)	1585	Paris	[s.n.]
Chauvet, Jacques	La pratique universelle de geometrie (1 vol.)	1585	Paris	Thierry, Henri (15.-1585)
Chauvet, Jacques	La pratique universelle de l'arpenterie (1 vol.)	1585	Paris	Thierry, Henri (15.-1585)

FIGURE D.5 – Catalogue général - classement par date

D.3. TRI DANS LE CATALOGUE GÉNÉRAL

AUTEUR	TITRE	DATE	LIEU D'ÉDITION	ÉDITEUR
Riquier, Jean-Charles Jumel	Traité d'économie-pratique ou moyens de diriger par économie différentes constructions, réparations ou entretiens (1 vol.)	1780	Amiens	Mastin
Dögen, Matthias	L'architecture militaire moderne, ou fortification (1 vol.)	1648	Amsterdam	Elzevier, Lodewijk (1604?-1670)
Fournier, Georges	Traité des fortifications, ou architecture militaire, tirée des places les plus estimées de ce temps, pour leurs fortifications (1 vol.)	1668	Amsterdam	Janssonius, Johannes (1588-1664)
Marolois, Samuel	Fortification ou architecture militaire tant offensive que deffensive (1 vol.)	1627	Amsterdam	Janssonius, Johannes (1588-1664)
Sigaud de La Fond, Joseph-Aignan	Lettre sur l'électricité médicale (1 vol.)	1771	Amsterdam	Desventes de La Doué, Antoine (1737-17..?)
Radiguet et Massiot	Prix-courant illustré des appareils électriques, instruments de démonstration... : 1898-1899 (1 vol.)	1899	Anzin	Ricouart-Dugour (Impr.)
Vène, Antoine	Essai sur une nouvelle théorie de l'électricité (1 vol.)	1817	Arras	Bocquet (Imprimerie)
Martinage, Louis	Outils français et américains : catalogue n° 24 d'outils et fourniture pour le découpage (1 vol.)	1928	Bar-Le-Duc	Comte-Jacquet et Chuquet (Impr. réunies)
Bourdets	Traité pratique des digues le long des fleuves et des rivières (1 vol.)	1773	Berlin	Jombert, Charles-Antoine (1712-1784)

FIGURE D.6 – Catalogue général - classement par lieu d'édition

D.3. TRI DANS LE CATALOGUE GÉNÉRAL

AUTEUR	TITRE	DATE	LIEU/DÉDITION	ÉDITEUR
The Autotype Cie	Papiers au charbon : papiers simple et double transfert, procédé Carbro... (1 vol.)	1939	Paris	Acker et Cie (Impr.)
Arnold and Sons	A catalogue of surgical instruments manufactured by Arnold and sons (1 vol.)	1879	Londres	Adlard, J. E.
Adolf Bleichert & Cie. (Leipzig, Allemagne)	Installations modernes de transport et de transbordement pour tous genres de matériaux (1 vol.)	1910	Paris	Adolf Bleichert & Cie. (Leipzig, Allemagne)
Adolf Bleichert & Cie. (Leipzig, Allemagne)	Transport dans les mines de charbon (1 vol.)	1913	Paris	Adolf Bleichert & Cie. (Leipzig, Allemagne)
Eiffel, Gustave	Conférence faite à l'Aéro-club de France sur la résistance de l'air et l'aviation (1 vol.)	1911	Paris	Aéro-club de France
Thornthwaite, W. H.	Hints on reflecting and refracting telescopes and their accessories (1 vol.)	1878	Londres	Alabaster and Passmore (Impr.)
Exposition universelle. 1900. Paris	Congrès international de l'enseignement des sciences sociales (1 vol.)	1900	Paris	Alcan, Félix (1841-1925)
Exposition universelle. 1900. Paris	Le Premier congrès de l'enseignement des sciences sociales (1 vol.)	1901	Paris	Alcan, Félix (1841-1925)
Toussaint, Albert-Charles-Louis	L'aviation actuelle. Étude aérodynamique et essais des avions. L'aviation actuelle et la sécurité (1 vol.)	1928	Paris	Alcan, Félix (Librairie)
Gautier Henri	Traité des fortifications contenant l'a	1485	Lyon	Amoulay Thomas

FIGURE D.7 – Catalogue général - classement par éditeur

D.4 Pages synoptiques

le **Cnum** À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Accueil > Catalogue général > L'Automobile théorique et pratique. II. Voitures à pétrole

L'Automobile théorique et pratique. II. Voitures à pétrole

NOTICE DÉTAILLÉE TABLE DES MATIÈRES ABRÉGÉE TABLE DES MATIÈRES INTÉGRALE VIGNETTES EXPORTS RECHERCHER DANS CE TITRE

Auteur(s) Baudry de Saunier, Louis (1865-1938)

Titre L'Automobile théorique et pratique. Traité élémentaire de locomotion à moteur mécanique. II. Voitures à pétrole

Adresse Neuilly-Levallois : L. Baudry de Saunier, 1901

Collation 1 vol. (512 p.) : 21 cm

Cote CNUM-BIB 8 De 187 (2)

Sujet(s) Automobiles -- Moteurs -- France -- 19e siècle

Url permanent <http://cnum.cnam.fr/redir?8DE:187.2>

FIGURE D.8 – Notice bibliographique

À PROPOS

CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE

Accueil > Catalogue général > L'Automobile théorique et pratique. II. Voitures à pétrole

L'Automobile théorique et pratique. II. Voitures à pétrole

NOTICE DÉTAILLÉE

TABLE DES MATIÈRES ABRÉGÉE

TABLE DES MATIÈRES INTÉGRALE

TABLE DES ILLUSTRATIONS

VIGNETTES

EXPORTS

RECHERCHER DANS CE TITRE

PAGE DE TITRE (Première image)

Avant-propos (p.5)

Chapitre premier. Généralités sur les Voitures Automobiles (p.7)

Chapitre deuxième. Les Pères de l'Automobile à Pétrole (p.56)

Les voitures à moteurs verticaux (p.79)

Chapitre troisième. La Voiture de Dion-Bouton (vertical et 1 cylindre) (p.80)

Chapitre quatrième. La Voiture Panhard et Levassor (vertical, 2 et 4 cylindres) (p.107)

Chapitre cinquième. La Voiture Mors (vertical, 2 et 4 cylindres) (p.170)

Chapitre sixième. La Voiture Rochet (vertical, 2 et 4 cylindres) (p.213)

Chapitre septième. La Voiture Hautier (vertical, 1 cylindre) (p.227)

Les voitures à moteurs horizontaux (p.243)

Chapitre huitième. La Voiture Rochet et Schneider (horizontal, 1 cylindre) (p.243)

Chapitre neuvième. La Voiture Peugeot (horizontal, 2 cylindres) (p.259)

Chapitre dixième. La Voiture Georges Richard (horizontal, 2 et 4 cylindres) (p.313)

Chapitre onzième. La Voiture Delahaye (horizontal, 2 et 4 cylindres) (p.335)

Chapitre douzième. La Voiture de Dietrich, système Amédée Bollée (horizontal, 2 et 4 cylindres (p.349)

Chapitre treizième. La Voiture Bolide (horizontal, 2 et 4 cylindres) (p.377)

Chapitre quatorzième. Les principaux accessoires (p.393)

Chapitre quinzième. L'utilisation d'un Voiture (p.419)

Chapitre seizième. Les pannes principales et leurs remèdes (p.465)

FIGURE D.9 – Table abrégée

À PROPOS

CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE

Accueil > Catalogue général > L'Automobile théorique et pratique. II. Voitures à pétrole

L'Automobile théorique et pratique. II. Voitures à pétrole

NOTICE DÉTAILLÉE

TABLE DES MATIÈRES ABRÉGÉE

TABLE DES MATIÈRES INTÉGRALE

TABLE DES ILLUSTRATIONS

VIGNETTES

EXPORTS

RECHERCHER DANS CE TITRE

PAGE DE TITRE (Première image)
 Avant-propos (p.5)
Chapitre premier. Généralités sur les Voitures Automobiles (p.7)
 I. Le confort (p.7)
 II. Éléments de carrosserie (p.9)
 III. Observations sur le plan général d'une automobile (p.21)
 IV. La carrosserie automobile proprement dite (p.32)
 V. Points communs à toutes les automobiles à pétrole (p.40)
 VI. Observations sur le refroidissement par eau (p.50)
Chapitre deuxième. Les Pères de l'Automobile à Pétrole (p.56)
 Lenoir, Benz et Daimler (p.56)
 Le moteur Daimler en V (p.58)
 I. Dispositions générales (p.61)
 II. Le régulateur (p.65)
 III. La distribution (p.68)
 Les voitures à moteurs verticaux (p.79)
Chapitre troisième. La Voiture de Dion-Bouton (vertical et 1 cylindre) (p.80)
 I. Le plan général (p.82)
 II. Le moteur, la carburation, l'allumage (p.88)
 III. La transmission (p.98)
 IV. L'utilisation de la voiture (n.105)

FIGURE D.10 – Table détaillée

À PROPOS
CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHER DANS CE TITRE

RECHERCHE AVANCÉE

NOTICE DÉTAILLÉE
TABLE DES MATIÈRES ABRÉGÉE
TABLES DES MATIÈRES INTÉGRALE
TABLE DES ILLUSTRATIONS
VIGNETTES
EXPORTS

L'Automobile théorique et pratique. II. Voitures à pétrole

Accueil > Catalogue général > L'Automobile théorique et pratique. II. Voitures à pétrole

PAGE DE TITRE (Première image)

Fig. 1. Voiture du type « Mylord » (p.10)

Fig. 2. Un essieu (p.11)

Fig. 3. Roue de carrosserie (type Hannover) (p.12)

Fig. 4. Schéma de l'Ecuateur et de Devers (p.13)

Fig. 5. Moyeu patent et ses principaux éléments (p.14)

Fig. 6. Moyeu demi-patent (p.14)

Fig. 7. Rapport de la caisse et de l'essieu (p.16)

Fig. 8. Principaux type de ressorts de voiture (p.19)

Fig. 9. La caisse séparée du châssis par des ressorts (p.22)

Fig. 10. La caisse sur le châssis et le tout sur les ressorts (p.23)

Fig. 11. Schéma exagéré des déplacements du châssis par rapport au train moteur (p.25)

Fig. 12. Roue à bossages (p.28)

Fig. 13. Moyeu à joues (p.29)

Fig. 14. Rai à douille (p.29)

Fig. 15. Schéma de l'avant-train directeur d'une voiture attelée et d'une automobile (p.31)

Fig. 16. Principaux modèles de carrosserie automobile (p.34)

Fig. 17. Coupe de la caisse en partinium d'une voiture de cours type Paris-Bordeaux 1900 (p.37)

Fig. 18. Tonneau de luxe en partinium, et sa coupe (p.38)

Fig. 19. Coupe de la voiture électrique de course en partinium « La Jamais Contente » système Jenatzy (p.39)

FIGURE D.11 – Table des illustrations

le num
À PROPOS
CATALOGUE GÉNÉRAL
RECHERCHER DANS LE CNUM
RECHERCHER AVANCÉE

RECHERCHER DANS CE TITRE

NOTICE DÉTAILLÉE

TABLE DES MATIÈRES ABRÉGÉE

TABLE DES MATIÈRES INTÉGRALE

TABLES DES ILLUSTRATIONS

EXPORTS

VIGNETTES

MONSIEUR LE DAUPHIN.

MONSIEUR,

Je vous prie de m'excuser de vous avoir écrit si tard, & de vous avoir écrit si peu. Je suis en ce moment à la campagne, & j'ai beaucoup de choses à faire. Je vous envoie ci-joint un petit ouvrage que j'ai écrit pour vous, & que j'espère que vous voudrez bien lire. Je vous prie de m'écrire quand vous en aurez l'occasion.

VOTRE DÉVOUÉ

RECHERCHER DANS CE TITRE

Accueil > Catalogue général > L'art des expériences, ou avis aux amateurs de la physique

L'art des expériences, ou avis aux amateurs de la physique

FIGURE D.12 – Interface vignettes

le **Cnum** À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Accueil > Catalogue général > L'art des expériences, ou avis aux amateurs de la physique

L'art des expériences, ou avis aux amateurs de la physique

NOTICE DÉTAILLÉE TABLE DES MATIÈRES ABRÉGÉE TABLE DES MATIÈRES INTÉGRALE TABLE DES ILLUSTRATIONS VIGNETTES EXPORTS RECHERCHER DANS CE TITRE

PDF (12Mo)	Tome 1 : télécharger ce document au format pdf
PDF (13Mo)	Tome 2 : télécharger ce document au format pdf
PDF (11Mo)	Tome 3 : télécharger ce document au format pdf
BiBtex (6880)	Tome 1 : télécharger la notice bibliographique de ce document au format BiBtex
BiBtex (6880)	Tome 2 : télécharger la notice bibliographique de ce document au format BiBtex
BiBtex (6880)	Tome 3 : télécharger la notice bibliographique de ce document au format BiBtex

FIGURE D.13 – Export des documents

le num À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHER DANS CE TITRE

Accueil > Catalogue général > Catalogue des instruments de chimie et physique...illustré de 400 gravures dans le texte

Catalogue des instruments de chimie et physique...illustré de 400 gravures dans le texte

NOTICE DÉTAILLÉE TABLE DES MATIÈRES ABRÉGÉE TABLE DES MATIÈRES INTÉGRALE VIGNETTES EXPORTS NOTE DE PRÉSENTATION RECHERCHER DANS CE TITRE

Note de présentation des catalogues de constructeurs du fonds Brioux

Le musée des arts et métiers a acquis en 2004 une partie de la bibliothèque personnelle d'Alain Brioux, libraire, expert, spécialiste des instruments scientifiques, dans le cadre de la vente publique : « Instruments de sciences et médecine de l'antiquité à nos jours : deux siècles de catalogues de fabricants français et étrangers (1739-1939) & Livres sur les instruments de collection de la fin du XIXe au début du XXe siècle ».

Un ensemble de 186 catalogues de constructeurs a ainsi rejoint la collection du [centre de documentation du musée](#). Le fonds Brioux, incontournable pour les historiens des techniques est aussi un outil précieux pour les chargés de collection du musée puisqu'il documente les objets concernant les domaines de la communication, de l'énergie, de l'instrumentation scientifique, de la mécanique. Hormis le catalogue de Van Musschenbroek datant de 1739, l'ensemble couvre la période 1815-1970. Y figurent des constructeurs français (Adnet, Alvergniat, Brewer, Charrière, Chauvin, Deleuil, Lerebours, Nacet...), anglais (Amadio, Arnold & sons, Beck, Becker...) et allemands (Brückner, Geissler Nachfolger, Schleicher, Zeiss...). Ces documents se présentent sous forme de catalogues détaillés, notices techniques ou publicitaires. Riches en contenu, abondamment illustrés, ils sont le plus souvent accompagnés de tarifs, parfois de courriers.

Le choix de numériser ces catalogues s'est imposé de lui-même, de par l'homogénéité de ce lot et surtout devant l'intérêt historique qu'il représente : 182 volumes fibres de droit sont désormais disponibles en ligne.

Notons que des catalogues de ce type, numérisés par d'autres institutions et accessibles sur Internet, sont recensés par [The Scientific Instrument Commission](#).

FIGURE D.14 – Note de présentation

le **cnum** À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHER AVANCÉE


Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale

Bulletin de la Société d'Encouragement pour l'Industrie Nationale

NOTICE DÉTAILLÉE LISTE DES VOLUMES VIGNETTES EXPORTATION DU DOCUMENT NOTE DE PRÉSENTATION RECHERCHER DANS CE TITRE

1802. An XI [1802]. 1e année
 1803. An XI-XII [1803]. 2e année
 1804. Messidor An XII [1804]- Prairial An XIII [1805]. 3e année. N. 1-12
 1805-1806. Messidor an XIII [1805]- Juin 1806. 4e année. N. 13-24
 1806-1807. Juin 1806-juin 1807. 5e année. N. 25-36
 1807. Juillet-décembre 1807. 6e année. N. 37-42
 1808. 7e année. N. 43-54
 1809. 8e année. N. 55-66
 1810. 9e année. N. 67-78
 1811. 10e année. N. 79-90
 1812. 11e année. N. 91-102
 1813. 12e année. N. 103-114
 1814. 13e année. N. 115-126
 1815. 14e année. N. 127-138
 1816. 15e année. N. 139-150
 1817. 16e année. N. 151-162
 1818. 17e année. N. 163-174
 1819. 18e année. N. 175-186
 1820. 19e année. N. 187-198
 1821. 20e année. N. 199-210
 1822. 21e année. N. 211-222
 1823. 22e année. N. 223-234

FIGURE D.15 – Liste des volumes



[À PROPOS](#)

[CATALOGUE GÉNÉRAL](#)

[RECHERCHER DANS LE CNUM](#)

[RECHERCHER AVANCÉE](#)

Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale

Bulletin de la Société d'Encouragement pour l'Industrie Nationale

NOTICE DÉTAILLÉE

LISTE DES VOLUMES

VIGNETTES

EXPORTATION DU DOCUMENT

NOTE DE PRÉSENTATION

RECHERCHER DANS CE TITRE









							
SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME I. 1828	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME II. 1829	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME III. 1830	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME IV. 1831	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME V. 1832	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME VI. 1833	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME VII. 1834	SOCIÉTÉ D'ENCOURAGEMENT L'INDUSTRIE NATIONALE TOME VIII. 1835

FIGURE D.16 – Liste des vignettes de la page de titre des volumes


À PROPOS
CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM
RECHERCHER AVANCÉE




Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale

Bulletin de la Société d'Encouragement pour l'Industrie Nationale

NOTICE DÉTAILLÉE
LISTE DES VOLUMES
VIGNETTES
EXPORTATION DU DOCUMENT
NOTE DE PRÉSENTATION
RECHERCHER DANS CE TITRE

PDF	(26Mo)	Tome 1 : télécharger ce document au format pdf
PDF	(24Mo)	Tome 2 : télécharger ce document au format pdf
PDF	(35Mo)	Tome 3 : télécharger ce document au format pdf
PDF	(36Mo)	Tome 4 : télécharger ce document au format pdf
PDF	(45Mo)	Tome 5 : télécharger ce document au format pdf
PDF	(21Mo)	Tome 6 : télécharger ce document au format pdf
PDF	(38Mo)	Tome 7 : télécharger ce document au format pdf
PDF	(51Mo)	Tome 8 : télécharger ce document au format pdf
PDF	(43Mo)	Tome 9 : télécharger ce document au format pdf
PDF	(40Mo)	Tome 10 : télécharger ce document au format pdf
PDF	(36Mo)	Tome 11 : télécharger ce document au format pdf
PDF	(40Mo)	Tome 12 : télécharger ce document au format pdf
PDF	(41Mo)	Tome 13 : télécharger ce document au format pdf
PDF	(49Mo)	Tome 14 : télécharger ce document au format pdf
PDF	(38Mo)	Tome 15 : télécharger ce document au format pdf
PDF	(38Mo)	Tome 16 : télécharger ce document au format pdf
PDF	(52Mo)	Tome 17 : télécharger ce document au format pdf
PDF	(49Mo)	Tome 18 : télécharger ce document au format pdf
PDF	(49Mo)	Tome 19 : télécharger ce document au format pdf
PDF	(64Mo)	Tome 20 : télécharger ce document au format pdf
PDF	(53Mo)	Tome 21 : télécharger ce document au format pdf

FIGURE D.17 – Export des volumes

D.5 Pages du feuilleteur

Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale > 1803. An XI-XII [1803]. 2e année > p.19

le Cnum À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

VIGNETTES TABLE DES MATIÈRES LISTE DES VOLUMES

(19)

aux plus beaux bleus de l'étranger, qu'ils en ont fabriqué en présence des commissaires de votre conseil d'administration, et qu'ainsi vous devez leur accorder le prix que vous avez proposé en l'an X.

Votre conseil est également d'avis de faire une mention très-honorable du mémoire cité dans ce rapport et qui a pour épigraphe : *Nata est ars ab experimento.*

Adopté en séance, le 2 messidor an XI.

ANALYSE comparée des ocres de Pourrain et de Saint-Amand, par le C. Méral-Guillot.

L'ocre est une substance minérale d'un jaune plus ou moins foncé, tirant quelquefois sur le rouge. Cette matière, après avoir subi quelques préparations, est employée dans les arts, particulièrement dans la peinture.

Le ci-devant Berry (département du Cher) était autrefois le seul pays de la France où l'on exploitait une mine d'ocre; on assure que les Hollandais l'achetaient presque toute (1), et que, après l'avoir fait calciner, ils nous la revendaient dix fois plus cher sous le nom de *rouge de Prusse* ou *d'Anapolitano*.

FIGURE D.18 – Feuilleteur en menu ouvert avec onglet vignette actif

le Cnum À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale > 1803. An XI-XII [1803]. 2e année > p.19

VIGNETTES TABLE DES MATIÈRES LISTE DES VOLUMES

1802. An XI [1802]. 1e année
 1803. An XI-XII [1803]. 2e année
 1804. Messidor An XII [1804]- Prairial/An XIII [1805]. 3e année. N. 1-12
 1805-1806. Messidor an XIII [1805]-Juin 1806. 4e année. N. 13-24
 1806-1807. Juin 1806-juin 1807. 5e année. N. 25-36
 1807. Juillet-décembre 1807. 6e année. N. 37-42
 1808. 7e année. N. 43-54
 1809. 8e année. N. 55-66
 1810. 9e année. N. 67-78
 1811. 10e année. N. 79-90
 1812. 11e année. N. 91-102
 1813. 12e année. N. 103-114
 1814. 13e année. N. 115-126
 1815. 14e année. N. 127-138
 1816. 15e année. N. 139-150
 1817. 16e année. N. 151-162
 1818. 17e année. N. 163-174
 1819. 18e année. N. 175-186
 1820. 19e année. N. 187-198
 1821. 20e année. N. 199-210
 1822. 21e année. N. 211-222
 1823. 22e année. N. 223-234

(19)

aux plus beaux bleus de l'étranger, qu'ils en ont fabriqué en présence des commissaires de votre conseil d'administration, et qu'ainsi vous devez leur accorder le prix que vous avez proposé en l'an X.

Votre conseil est également d'avis de faire une mention très-honorable du mémoire cité dans ce rapport et qui a pour épigraphe : *Nata est ars ab experimento.*

Adopté en séance, le 2 messidor an XI.

ANALYSE comparée des ocres de Pourrain et de Saint-Amand, par le C. Méral-Guillot.

L'ocre est une substance minérale d'un jaune plus ou moins foncé, tirant quelquefois sur le rouge. Cette matière, après avoir subi quelques préparations, est employée dans les arts, particulièrement dans la peinture.

Le ci-devant Berry (département du Cher) était autrefois le seul pays de la France où l'on exploitait une mine d'ocre; on assure que les Hollandais l'achetaient presque toute (1), et que, après l'avoir fait calciner, ils nous la revendaient dix fois plus cher sous le nom de *rouge de Prusse* ou *d'Anapolitano*.

FIGURE D.19 – Feuilleteur en menu ouvert avec onglet liste des volumes actif

le Cnum À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale > 1803. An XI-XII [1803]. 2e année > p.19

(19)

aux plus beaux bleus de l'étranger, qu'ils en ont fabriqué en présence des commissaires de votre conseil d'administration, et qu'ainsi vous devez leur accorder le prix que vous avez proposé en l'an X.

Votre conseil est également d'avis de faire une mention très-honorable du mémoire cité dans ce rapport et qui a pour épigraphe : *Nata est ars ab experimento.*

Adopté en séance, le 2 messidor an XI.

ANALYSE comparée des ocres de Pourrain et de Saint-Amend,
par le C. Merat-Guillot.

L'ocre est une substance minérale d'un jaune plus ou moins foncé, tirant quelquefois sur le rouge. Cette matière, après avoir subi quelques préparations, est employée dans les arts, particulièrement dans la peinture.

Le ci-devant Berry (département du Cher) était autrefois le seul pays de la France où l'on exploitait une mine d'ocre; on assure que les Hollandais l'achetaient presque toute (1), et que, après l'avoir fait calciner, ils nous la revendaient dix fois plus cher sous le nom de *rouge de Prusse* ou *d'Ancolatovo*

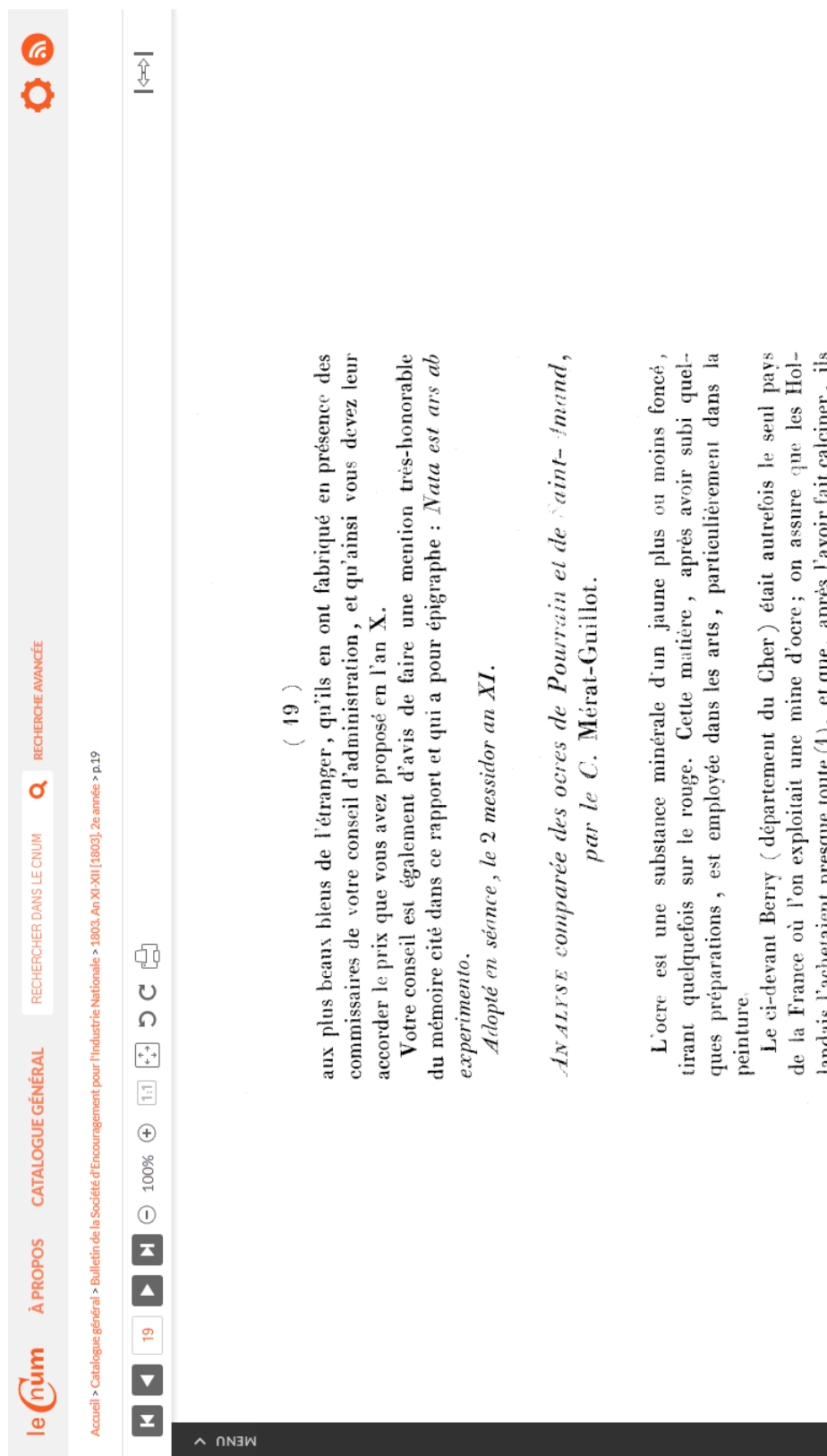
MENU

19

100%

↑ ↓

FIGURE D.20 – Feuilleteur en menu fermé avec barre de navigation à droite



le cum

À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHE AVANCÉE

Accueil > Catalogue général > Bulletin de la Société d'Encouragement pour l'Industrie Nationale > 1803. An XI-XII [1803]. 2e année > p.19

19

100%

MENU

(19)

aux plus beaux bleus de l'étranger, qu'ils en ont fabriqué en présence des commissaires de votre conseil d'administration, et qu'ainsi vous devez leur accorder le prix que vous avez proposé en l'an X.

Votre conseil est également d'avis de faire une mention très-honorable du mémoire cité dans ce rapport et qui a pour épigraphe : *Nata est ars ab experimento.*

Adopté en séance, le 2 messidor an XI.

ANALYSE comparée des ocres de Pourrain et de Saint-Amand,
par le C. Mérat-Guillot.

L'ocre est une substance minérale d'un jaune plus ou moins foncé, tirant quelquefois sur le rouge. Cette matière, après avoir subi quelques préparations, est employée dans les arts, particulièrement dans la peinture.

Le ci-devant Berry (département du Cher) était autrefois le seul pays de la France où l'on exploitait une mine d'ocre; on assure que les Hollandais l'achetaient presque toute (1), et que, après l'avoir fait calciner, ils

FIGURE D.21 – Feuilleteur en menu fermé avec barre de navigation en haut

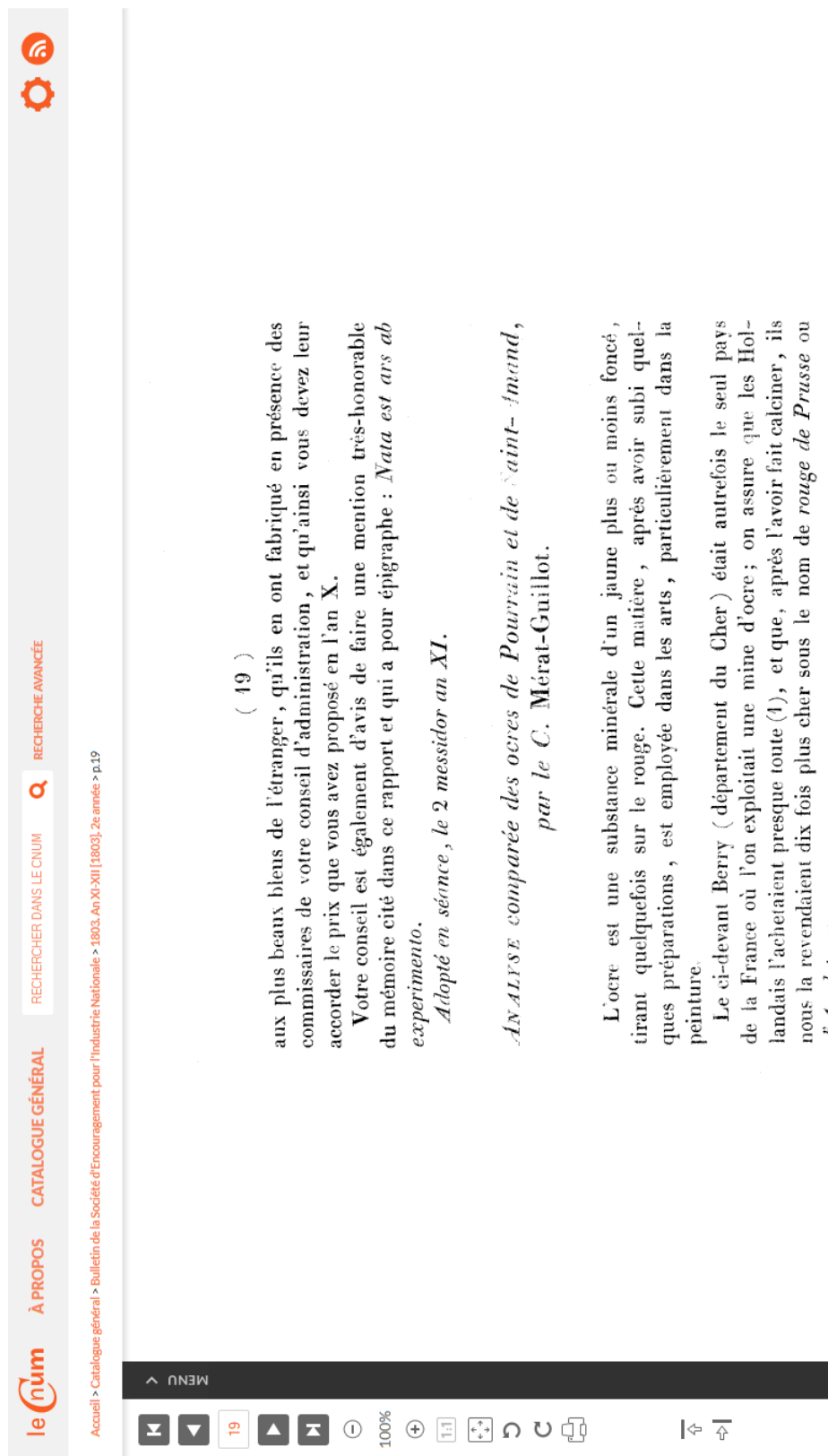





FIGURE D.22 – Feuilleter en menu fermé avec barre de navigation à gauche

D.6 Recherche avancée


[À PROPOS](#)
[CATALOGUE GÉNÉRAL](#)
[RECHERCHER DANS LE CNUM](#)
[RECHERCHE AVANCÉE](#)

Accueil > Recherche avancée

Tables des matières

- Table des matières et illustrations
- Table des matières
- Table des illustrations
- Notice bibliographique**
- Tous les mots de la notice
- Mot(s) auteur
- Mot(s) du titre
- Mot(s) sujet
- Mot(s) éditeur
- Lieu d'édition
- Cote
- Mot(s) des notes
- Identifiant Cnum

Et

Et

+

Par année d'édition

Par année de mise en ligne

De à

Avant

Après

X EFFACER LA RECHERCHE

LANCER LA RECHERCHE

Recherche exacte

? AIDE À LA RECHERCHE

FILTRES PAR :

THÈME

- Tous
- Catalogues de constructeurs
- Construction
- Expositions universelles
- Énergie
- Généralités scientifiques et vulgarisation
- Histoire du Cnum
- Machines et instrumentation
- Technologies de l'information
- Transports

TYPE DE DOCUMENTS

- Tous
- Revue
- Ouvrage
- Manuscrit

LANGUE

- Tous
- Français
- Allemand
- Anglais
- Italien

Recherche sur Gallica

FIGURE D.23 – Recherche avancée

D.7 Pages de résultat



À PROPOS CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE




Accueil > Recherche avancée > Liste des résultats

Le champ "Table des matières et des illustrations" contient tous les mots **avlon**

MODIFIER LA RECHERCHE

RÉSULTAT(S) PAR :

TYPE DE DOCUMENTS
 Revues (83)
 Ouvrages (11)

LANGUE
 Français (94)

94 documents en 2,204 secondes. Documents 1 à 10.

0 1 2 3 4 5 6 7 8 9 Page suivante ▶▶

REVUES

- + **183 occurrences trouvées dans 3 documents**
 - **169 occurrences dans 53 volumes de** **La Nature**
 - + **14 occurrences dans le titre** 1926 : Cinquante quatrième année, deuxième semestre : n° 2726-2751
 - **10 occurrences dans le titre** 1924 : Cinquante deuxième année, premier semestre : n° 2596-2621
 - Avlon** automatique et sa direction par T. S. F. (P. HÉMARDINQUER) (p.99)
 - Avlon** : de Paris au Japon (p.153)
- + **26 occurrences dans 9 volumes du** **Bulletin de la Société d'Encouragement pour l'Industrie Nationale**

OUVRAGES

- **X occurrences dans** **Ader, Clément - La première étape de l'aviation militaire en France (1907)**
 - [Avlon d'Ader]**
 - Croquis N°2. Position de l'**avlon**, avec ses roues droites, pendant l'embarquée par rapport à la direction du vent et à celle de la distance parcourue
- **X occurrences dans** **Toussaint, Albert-Charles-Louis - L'aviation actuelle (1928)**
 - Chapitre I. Étude aérodynamique de l'**avlon**

FIGURE D.24 – Résultat mode abrégé

À PROPOS

CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE

Accueil > Recherche avancée > Liste des résultats

Le champ "Table des matières et des illustrations" contient tous les mots **avlon**

94 documents en 2,194 secondes. Documents 1 à 10.

0 1 2 3 4 5 6 7 8 9 Page suivante ▶▶

MODIFIER LA RECHERCHE

RÉSULTAT(S) PAR :

TYPE DE DOCUMENTS	
Revues (83)	
Ouvrages (11)	
LANGUE	
Français (94)	

Ader, Clément - La première étape de l'aviation militaire en France (1907)
[Avlon d'Ader]

Croquis N°2. Position de l'**avlon**, avec ses roues droites, pendant l'embarquée par rapport à la direction du vent et à celle de la distance parcourue

Toussaint, Albert-Charles-Louis - L'aviation actuelle (1928)
Chapitre I. Étude aérodynamique de l'**avlon**

d) Caractéristiques aérodynamiques des principales résistances passives d'un **avlon**

c) **Avlons** à surface variable

d) **Avlon** à surface et courbure variables

d) Les organes de stabilité longitudinale à leur emplacement sur l'**avlon**

Chapitre II. Étude des ensembles d'éléments d'**avlons**

3. Étude de l'**avlon** complet

Chapitre III. Les essais de l'**avlon** au Laboratoire aérodynamique et en vol

1. Les essais de l'**avlon** au Laboratoire aérodynamique
2. Les essais en vol des **avlons**
1. Sécurité de construction des **avlons**
5. Les **avlons** à voilures sustentatrices tournantes

Fig. 8. Résistances aérodynamiques d'une plaque rectangulaire, d'un cylindre circulaire et d'un montant fuselé d'**avlons**

Fig. 14. Roue d'**avlon**. Influence de diverses dispositions d'entoilage

Fig. 59. Quelques formes de l'empennage horizontal d'un **avlon**

Fig. 64. Vue de face d'un **avlon** dont la voilure a un " dièdre latéral "

Fig. 65. Quelques formes de l'empennage vertical des **avlons**

FIGURE D.25 – Résultat en mode texte

À PROPOS
CATALOGUE GÉNÉRAL

Q

RECHERCHER DANS LE CNUM
RECHERCHE AVANCÉE

Accueil > Recherche avancée > Liste des résultats

Le champ "Table des matières et des illustrations" contient tous les mots **avion**

94 documents en 2,194 secondes. Documents 1 à 10.

[0](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[6](#)
[7](#)
[8](#)
[9](#)
[Page suivante ▶](#)

MODIFIER LA RECHERCHE Q

☰
☰

RÉSULTAT(S) PAR :

TYPE DE DOCUMENTS

Revue **(83)**

Ouvrages **(11)**

LANGUE

Français **(94)**

[Avion d'Ader]

Croquis N°2. Position de l'**avion**, avec ses roues droites, pendant l'embarquée par rapport à la direction du vent et à celle de la distance parcourue

Chapitre I. Étude aérodynamique de l'avion

d) Caractéristiques aérodynamiques des principales résistances passives d'un **avion**

c) **Avions** à surface variable

d) **Avion** à surface et courbure variables

d) Les organes de stabilité longitudinale à leur emplacement sur l'**avion**

FIGURE D.26 – Résultat en mode vignette et texte

D.7. PAGES DE RÉSULTAT

À PROPOS

CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE

Accueil > Recherche avancée > Liste des résultats

Le champ "Table des matières et des illustrations" contient tous les mots avion

RECHERCHER LA RECHERCHE

0 1 2 3 4 5 6 7 8 9 Page suivante ▶

94 documents en 2,194 secondes. Documents 1 à 10.

RÉSULTAT(S) PAR :

TYPE DE DOCUMENTS
Revue (83)
Ouvrages (11)

LANGUE
Français (94)

	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>
<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>
<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>
<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>	<p>LES BRAS DE L'AVION DE L'AVION. ET DE L'AVION.</p>

FIGURE D.27 – Résultat en mode vignette

D.8 Divers

leCnum

A PROPOS CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

RECHERCHE AVANCÉE

S'abonner

Accueil > Erreur

La page demandée P84.389 est introuvable. L'ouvrage est sans doute en cours de numérisation...

Nous vous invitons à consulter la page d'accueil ou le moteur de recherche.

Contact Plan du site Crédits Mention légale

© Cnum 2013

FIGURE D.28 – Gestion des erreurs

The image shows a horizontal navigation bar for 'leCnum' with the following elements from left to right: the 'leCnum' logo, 'À PROPOS', 'CATALOGUE GÉNÉRAL', a search box with 'RECHERCHER DANS LE CNUM' and a magnifying glass icon, 'RECHERCHE AVANCÉE', a gear icon, and a Wi-Fi icon. Below the navigation bar, the text reads: 'Accueil > Outils de découverte'.

Découvrir des pages au hasard
 En hommage à l'ingénieur Agostino Ramelli pour qui « ... cette sorte de machine un homme peut voir et lire une grande quantité de livres sans se mouvoir d'un lieu », nous vous proposons une fonctionnalité permettant d'afficher une page du Cnum au hasard.

Une fois que vous avez cliqué sur un de ces liens, réactualiser autant de fois que vous le souhaitez l'affichage de la page. Dans certains cas, placez aléatoirement la barre de défilement verticale de votre navigateur avant de réactualiser.

- N'importe quelle page
- Une notice au hasard
- Une notice de série au hasard
- Une table des matières abrégée au hasard
- Une table des matières intégrale au hasard
- Une table des illustrations au hasard
- Une liste des volumes au hasard
- Vignettes d'une série au hasard
- Vignettes détaillées d'une série au hasard
- Vignettes au hasard
- Vignettes détaillées au hasard
- Une page de titre au hasard
- Une page d'illustration au hasard
- Une page d'un ouvrage au hasard
- Une note de présentation au hasard

Index alphabétique (prochainement)
 L'index alphabétique contient l'ensemble des mots des tables des matières et illustrations.

FIGURE D.29 – Outils de découverte

À PROPOS
CATALOGUE GÉNÉRAL

RECHERCHER DANS LE CNUM

Q

RECHERCHE AVANCÉE

Accueil > Index alphabétique

Premier caractère : A B C D E F **G** H I J K L M N O P Q R S T U V W X Y Z 123

Deuxième caractère : A B C D E F G H I J K L M N **O** P Q R S T U V W X Y Z 123

Troisième caractère : Début A B C D E F G H I J K L M N O P Q R S T U V W X Y Z fin

Premier caractère

Deuxième caractère

Troisième caractère

Mots


g	o	k					goitrigènes
			Gokasho				
			Gokel				
			Gokstad				
			Gokstad.				
			Gokteik				
g	o	l					
			GOLA				
			Golay				
			Golay,				
			GOLAY,				
			Golaz				
			GOLAZ				
			Golaz-Sénac.				
			Gold				

FIGURE D.30 – Index alphabétique

le Cnum
À PROPOS
CATALOGUE GÉNÉRAL
RECHERCHER DANS LE CNUM
RECHERCHE AVANCÉE

Accueil > A propos
RECHERCHER
RECHERCHER

À propos



Le Conservatoire numérique des Arts et Métiers (Cnum) est une bibliothèque numérique consacrée à l'histoire des sciences et des techniques. Conçu en 2000, le Cnum est alors un projet commun de la [bibliothèque centrale](#), du Centre d'histoire des techniques et de l'environnement (CDHTE) et du [Centre d'études et de recherche en informatique et communications \(CEDRIC\)](#) du Cnam.

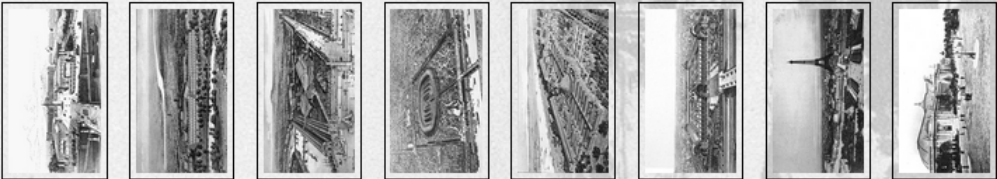
Constitué initialement à partir des fonds patrimoniaux de la bibliothèque centrale, le Cnum propose la consultation en ligne d'une collection d'ouvrages et de périodiques francophones, traitant des sciences et des techniques. Le Conservatoire numérique s'adresse aux chercheurs de ces domaines, en offrant des textes et une documentation iconographique spécifiques ; le Cnum est aussi un outil de vulgarisation scientifique qui remplit l'une des missions fondamentales du Cnam : la diffusion du savoir et la reconnaissance du patrimoine scientifique et technique francophone. C'est à ce titre que les collections imprimées du [musée des arts et métiers](#) alimentent depuis 2008 la bibliothèque numérique.

Le programme de numérisation des ouvrages se décline selon neuf thématiques majeures :

- [Catalogues de constructeurs](#)
- [Construction](#)
- [Énergie](#)
- [Expositions universelles](#)
- [Généralités scientifiques et vulgarisation](#)
- [Histoire du Cnam](#)
- [Machines et instrumentation scientifique](#)
- [Technologies de l'information et de la communication](#)
- [Transports](#)

Contact
Conditions d'utilisation
© Cnum 2014

FIGURE D.31 – A propos



Palais, Pavillons et Galeries

Les bâtiments des expositions universelles en représentation (1798-1900)

Exposition : sujet de délire du 19^e siècle. Ainsi Flaubert définissait-il dans son facétieux *Dictionnaire des idées reçues* ces grand-messes à la gloire du progrès que furent les expositions universelles. Temple des arts et des techniques, mise en scène du développement industriel, instrument au service des nationalismes, apologie du colonialisme : l'exposition universelle résume le 19^e siècle occidental. Sa géographie épouse d'ailleurs largement celle du développement industriel de l'Angleterre, en passant par la France, avant de gagner de nouveaux territoires en Europe et aux États-Unis. Elle imprime sa marque sur l'espace urbain qu'elle remodelé durablement. L'espace dédié aux exposants ne cesse de s'étendre, passant d'une dizaine d'hectares à Londres en 1851 à plus de deux cents pour Vienne en 1873 ; et de se multiplier, du chapiteau unique de l'exposition de Paris de 1798 à la multitude des palais et pavillons de l'exposition de 1900. Au gré de véritables prouesses architecturales, la mise en scène du Progrès transforme le paysage urbain en une véritable « féerie géographique ». [[Lire la suite...](#)]

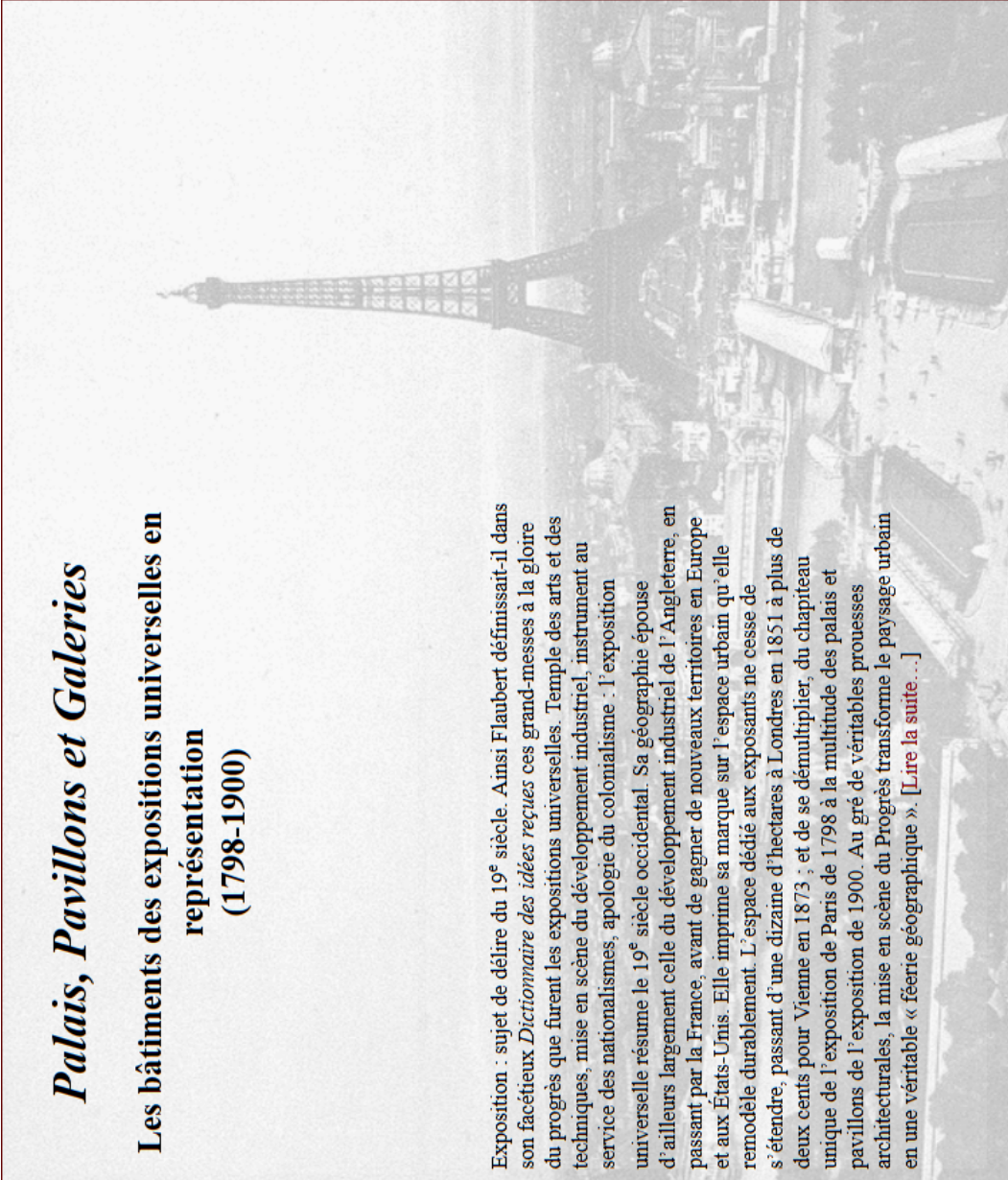


FIGURE D.32 – Exposition virtuelle

D.9 Cnum2.1

leCnum À PROPOS CATALOGUE GÉNÉRAL RECHERCHER DANS LE CNUM RECHERCHER AVANCÉE

Accueil > Catalogue général > Essai mathématique sur les prix de revient des transports par chemins de fer

Essai mathématique sur les prix de revient des transports par chemins de fer

NOTICE DÉTAILLÉE TABLE DES MATIÈRES ABRÉGÉE TABLE DES MATIÈRES INTÉGRALE VIGNETTES EXPORTS RECHERCHER DANS CE TITRE

NOTICE DÉTAILLÉE

TABLE DES MATIÈRES ABRÉGÉE

TABLE DES MATIÈRES INTÉGRALE

VIGNETTES

EXPORTS

RECHERCHER DANS CE TITRE

EMITA

FIGURE D.33 – Vignettes 2



FIGURE D.34 – Nuage de mots-clé

Annexe E

Aide à la recherche en ligne

Introduction

Vous avez la possibilité de localiser des occurrences de mots dans toutes les tables de matières, tables d'illustrations et index des ouvrages du CNUM, ainsi que dans leurs notices bibliographiques. Merci de noter qu'il ne s'agit pas du texte des ouvrages eux-mêmes : nous ne disposons malheureusement pas de ces données.

Avant d'effectuer une recherche, consultez le [Catalogue général](#) du Conservatoire Numérique pour avoir une idée du fonds (uniquement consacré à l'histoire des sciences et des techniques).

L'orthographe ancienne a été maintenue. Nous avons établi un [glossaire des variantes](#) de noms propres et communs rencontrés lors de la saisie des tables de matières. N'hésitez pas à le consulter avant de formuler vos requêtes.

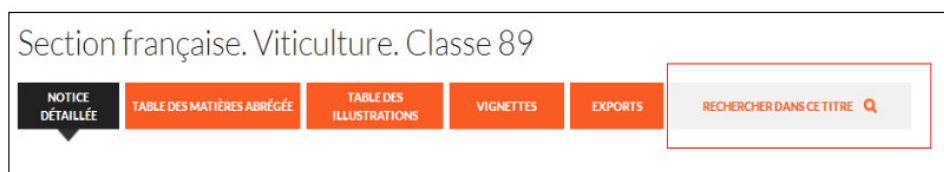
Interface de recherche

La recherche simple

La recherche simple est accessible par le champ de saisie placée en haut de toutes les pages. Il vous permet de rechercher les mots à la fois dans les notices d'ouvrages et les tables de matières et illustrations.

La recherche dans un titre d'ouvrage spécifique

Chaque titre d'ouvrage contient un onglet de restriction de recherche sur cet unique titre. Le lien ouvre la page de recherche avancée avec présélection de la métadonnée identifiant et pré-remplissage automatique du premier champ. Voir l'exemple sur l'ouvrage [Section française. Viticulture. Classe 89](#)



Capture d'écran de cette [page](#)

La recherche avancée

Cette recherche permet de combiner vos mots-clés sur différents champs en utilisant les opérateurs logiques (ET, OU et SAUF) et de limiter les résultats à une période d'édition et de mise en ligne, à un thème, à un type de document ou une langue. Avec ces différentes combinaisons, vous pourrez en même temps activer la case à cocher de la "Recherche exacte".

La recherche sur texte intégral

Actuellement, la recherche en plein texte par le moteur n'est pas intégré. Seul le téléchargement du document pdf de ces deux titres vous offre cette possibilité : Le Muet, Pierre - [Manière de bastir pour toutes sortes de personnes](#) et Brame, Edouard - [Etude sur les signaux de chemins de fer à double voie](#)

Interface de résultat

Dans cette interface, nous avons limité le nombre maximum de document par page à 10 et de ligne de la table des matières à 100 dans le but de réduire le temps d'affichage des résultats.

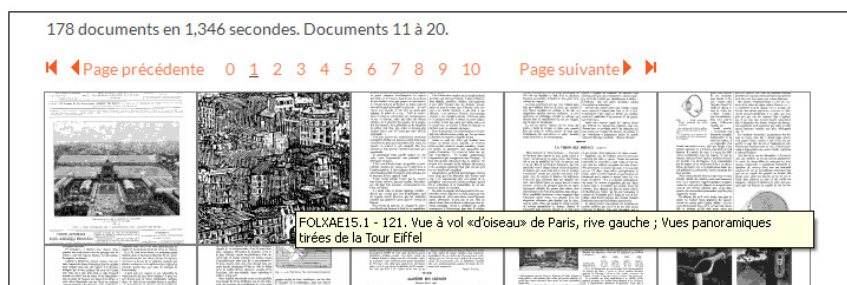
Les chevrons de début et de fin  vous aideront à passer directement soit à la première ou la dernière page de la liste des résultats.

Vous avez le choix entre 3 modes de visualisation pour consulter vos résultats de recherche : « Texte », « Texte et vignette » et « Vignette ». Le dernier mode sélectionné sera celui qui apparaîtra dans la recherche suivante. Veuillez noter que les deux derniers modes n'afficheront aucun résultat dans le cas d'une recherche restreinte aux champs de la notice.



Capture d'écran des trois modes de visualisation

Dans le mode vignette, la ligne de la table des matières sera visible dans l'infobulle. Celle-ci apparaîtra en laissant immobile quelques secondes le pointeur de souris sur la vignette. Voir un exemple sur l'image suivante :



Capture d'écran du mode vignette

Le nombre de document, de thématique, de typologie et de langue est égale à la somme des titres de revue plus ses propres volumes. Cette somme explique la différence de valeurs entre les chiffres de la page d'accueil et ceux de la page des résultats.

Les titres de regroupement sans renvoi vers un fac-similé, ouvriront directement la notice du titre. Ex : [Aéronautique \(intitulé avec renvoi sur la notice, voir l'aide\)](#). Ensuite vous devrez rechercher ce mot sur la page d'une des tables proposées.

Aéronautique.

La vitesse moyenne en aérostation (HENRI DE PARVILLE) .	18
Un voyage aérostatique au long cours (G. ESPITALIER) .	85
Le ballon dirigeable Sève (G. ESPITALIER)	587
Les ascensions en cerf-volant (J. LEBORGNE)	410

Capture d'écran d'un titre de regroupement

Fonctionnement du moteur de recherche

Le Cnum utilise le moteur de recherche libre **Apache Lucene Solr**. Il offre de multiples possibilités comme la recherche à facettes, la lemmatisation, la racinisation, suppression de points, mots au pluriel, acronymes...

Le moteur Solr structure les données dans un document avec différents champs contenant la notice bibliographique d'un titre et sa table des matières et illustration. Chacun de ses champs est indexé et représenté sous de multiples variantes lexicales. Veuillez prendre en compte que ce type de représentation n'est plus appliqué lors de la recherche exacte.

Classement et tri des résultats

Le moteur affecte un score à chaque document basé sur un calcul de pertinence. Ce calcul prend en compte certains critères comme la fréquence d'apparition du terme dans le document, le nombre de document contenant le terme... Vous trouverez sur ce [lien](#) une explication détaillée du système de classement opéré par Solr.

Dans la page de résultat final, les documents sont classés par ordre décroissant de leur score de pertinence.

Une recherche sur le champ date active automatiquement le tri sur ce champ par ordre croissant. En dehors de cette recherche, le tri est fondé sur la pertinence du document.

Opérateurs booléens dans le champ de recherche

Les opérateurs booléens « ET » et « OU » dans le champ de saisie sont reconnus par le moteur de recherche en jouant sur les guillemets. Par exemple, la recherche dans les tables de matières et illustrations des mots "machine turbine" (entre guillemets) renvoie comme résultat les lignes de la table des matières contenant ces 2 mots à la fois. Par contre machine turbine (sans les guillemets) renvoie les lignes de la table des matières contenant l'un ou l'autre de ces 2 mots.

Recherche dans les tables des matières	Résultats à l'affichage
Entre guillemets dans le champ 1. Ex.: Champ 1 : « mot1 mot2 »	Le titre de chapitre du document contient mot1 ET mot2
Sans les guillemets dans le champ 1. Ex.: Champ 1 : mot1 mot2	Le titre de chapitre du document contient mot1 OU mot2
Champ1 ET Champ2 Ex.: Champ 1 : mot 1 ET champ 2 : mot2	Le titre de chapitre du document contient mot1 OU mot2 mais les 2 mots ont été trouvés au moins une fois dans cet ensemble

Vous pouvez définir une distance de séparation entre les mots en mettant entre guillemets les termes suivi du signe tilde « ~ » (Taper Alt Gr + chiffre 2 pour le saisir sous Windows) et d'un chiffre (valeur 50 par défaut). Par exemple :

- "récents électricité Hughes"~15 veut dire que chaque mot est séparé de l'un de l'autre d'au plus 15 mots.
- "transport charbon"~4 renverra les lignes des tables des matières contenant à la fois les mots transport et charbon séparés l'un de l'autre par 4 mots au maximum.
- "Téléphonie+sans+fil" est équivalent à "Téléphonie sans fil"
- "Téléphonie sans fil"~0 veut dire 0 mot d'écart entre les mots
- "commerce industrie" est équivalent à "industrie commerce"
- "transport charbon" est équivalent à "transport charbon"~50 car la valeur 50 est défini par défaut dans les paramètres du moteur

« Booster » un terme

Vous pouvez attribuer plus de poids à certains champs en affectant à la fin des termes recherchés le signe « ^ » suivi d'un chiffre. Par exemple, ces deux exemples de recherche sur les tables dont les coefficients ont été intervertis avion^10 et atlantique^2 et avion^2 atlantique^10 afficheront les mêmes résultats mais seront classés différemment.

Sensibilité à la casse

La recherche s'effectue sans tenir compte de la casse des caractères. Rechercher galvaní, Galvani ou GALVANI revient au même, de plus la lemmatisation renverra aussi les résultats de galvanisme, galvanisation... Mais tout cela ne s'applique plus lors d'une recherche exacte.

Opérateur de troncature

Le caractère « * » prend en compte la répétition (de 0 à n fois) d'un caractère dans une chaîne recherchée. Ex : une recherche du terme archi* retournera toutes les occurrences de « architecture », « archipel », « Archimède »,... A noter que l'opérateur de troncature à gauche n'est pas intégré.

Caractères spéciaux

Ligatures

La plupart des ligatures (comme æ, œ, Æ, Œ) ont été remplacées automatiquement par leurs lettres non accolées. Vous pouvez néanmoins les saisir manuellement dans le formulaire de recherche. Par exemple la ligature « œ » sur le système d'exploitation Windows s'écrit en tenant la touche Alt enfoncé puis tapant les chiffres 0156. Nous vous invitons à [consulter cette page](#) si vous souhaitez saisir d'autres caractères spéciaux.

Exemple de recherche : [kiær](#)

Signes diacritiques

Certains signes diacritiques ont été remplacés par leur équivalent sans le signe comme « Ÿ » par « Y » : TAYLOR, Foley.

Elisions

Les élisions « c' », « d' », « j' », « l' », « m' », « n' », « qu' », « s' », « t' » ont été pris en compte pendant l'indexation. Ainsi, une recherche sur les mots oiseau et l'oiseau renverront les mêmes occurrences.

Exemples de formulation de requêtes

Nous vous proposons ici quelques exemples de recherche dont les combinaisons et mots-clés pourront vous servir :

- Les auteurs du 18ème siècle
- Les éditeurs du 20ème siècle
- Les titres sans nom d'éditeurs
- Les titres sans lieu d'édition
- Les titres contenant une zone de notes
- Les titres soumis aux droits d'auteur
- Les titres de la collection du musée des arts et métiers
- Les titres de la collection de la bibliothèque centrale des arts et métiers
- Les titres qui n'appartiennent ni à la collection de la bibliothèque centrale ni à celle du musée des arts et métiers
- Les titres des Catalogues de constructeurs du fonds Bovis
- Les titres des Catalogues de constructeurs du fonds Brioux

- [Les titres des catalogues historiques du Musée des arts et métiers](#)
- [Les titres avec la notice complète](#)

La recherche approximative correspond aux recherches sur des mots avec une orthographe assez proche. Exemple :

- [tecture](#)
- [tecture~0.5](#) suivi d'un coefficient de précision, le chiffre zéro étant le mot sans déformation.

Vous avez la possibilité de rechercher dans un intervalle de lettre ou de chiffre. Bien que les résultats ne soient pas tout à fait corrects, ils restent néanmoins exploitables. En effet, les champs n'ont pas été indexés pour ce type de requête. Exemple :

- [\[1803 TO 1807\]](#) renvoie les lignes de tables contenant les chiffres (correspondant à des années le plus souvent) de 1803 à 1807
- [\[2.7 TO 2.8\]](#) renvoie les lignes de tables contenant les chiffres compris dans l' intervalle 2.7 et 2.8
- [\[3.1 TO 3.4\]](#) renvoie les lignes de tables contenant les chiffres compris dans l' intervalle 3.1 et 3.4
- [\[T TO W\]](#) renvoie le champ auteur contenant les mots dont la première lettre est soit T, U, V ou W.

Moteurs alternatifs

Moteurs de recherche du Web

Les moteurs de recherche du Web peuvent restreindre la requête sur un site prédéfini. Nous vous proposons cette possibilité dans le champ de saisie suivant :



Moteurs des fournisseurs de services

Depuis 2006, les notices du Cnum peuvent être moissonnées par l'entrepôt OAI-PMH. Ainsi, certains moissonneurs comme [Gallica](#) ou [Isidore](#) vous offrent ce service spécifique de recherche sur notre entrepôt de données.

Annexe F

Récurrence des mots de la table des matières BSPI.9

12 est	31 leur
12 [Machine	32 prix
12 nouvelle	33 sa
13 Arts	34 fait
13 aux	37 moyen
13 bois	40 Prix
13 celles	47 dans
13 d'une	53 en
13 Est	58 au
13 [Pastel	67 ,
13 proposé	67 (cf.
14 cette	67 note)]
14 machine	67 probable
15 ib	74 le
15 laine	82 [Idem]
16 Conseil	93 sur
16 son	102 du
17 culture	122 et
17 un	134 les
19 avec	137 des
19 MM.	148 par
19 (progr.),	154 pour
21 concours	173 M.
23 d'un	231 la
23 [Machines	233 ...]
26 ce	521 de
27 fabrication	747
27 Société	
28 RAPPORT	
30 nom	
31 De	

Annexe G

Occurrence des mots peu courants

1 cetoehilus	.
1 etone	.
1 etorhinus	.
1 etoses	1457 l'appareil
1 -ette	1467 montrant
1 ceuta	1510 i
1 cevedale	1520 type
1 cevenol	1542 qui
1 cevin	1665 e
1 ceylon	1684 est
1 ceysac	1710 aux
1 cezallier	1713 son
1 cezembre	1797 mm
1 cfcl	1815 paris
1 cfd	1841 droite
1 c°-forme	1872 c
1 cfpa	1882 gauche
1 cga	1914 n"
1 cgg	2113 fer
1 cgo	2137 b
1 cgv	2196 moteur
1 ch'	2383 ou
1 cha	2408 systeme
1 chaalis	2421 deux
1 chabannais	2501 schema
1 chabelardi	2553 plan
1 chabiague	2868 appareil
1 chabin	2901 photographie
1 chaboche	3024 electrique
1 chabraque	3819 planche
1 chaclacayo	4066 machine
1 chadouf	
1 chadoufe	

Annexe H

Modèle conceptuel de données

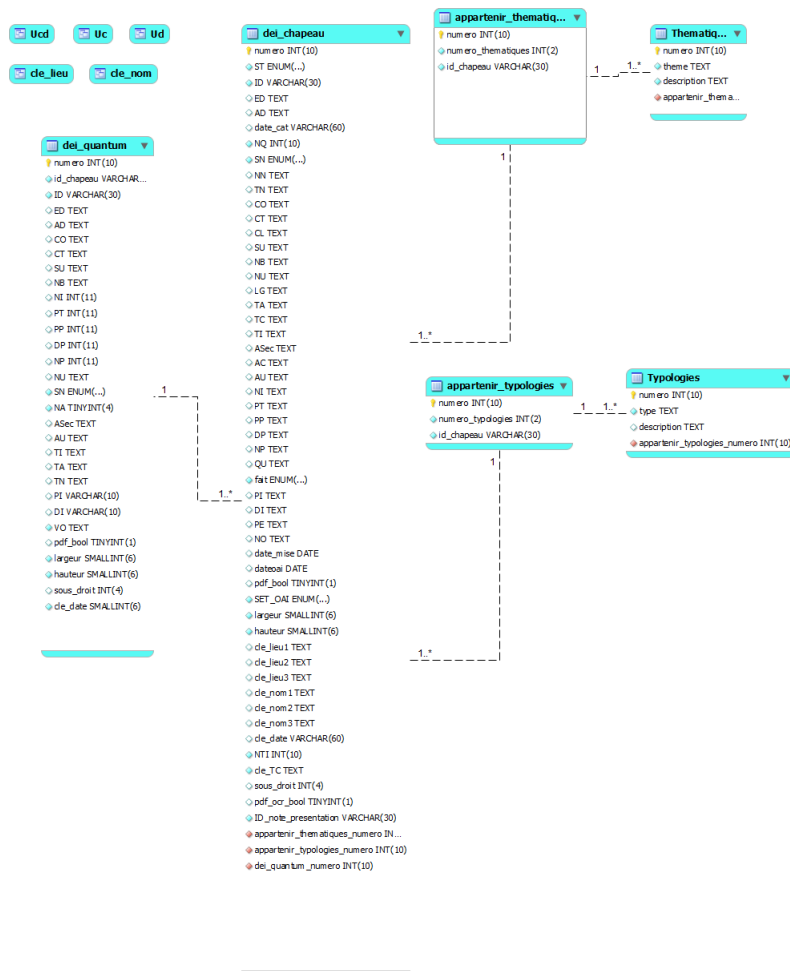


FIGURE H.1 – Le modèle conceptuel de données

Annexe I

Commandes phpMyAdmin de création d'une vue ou table virtuelle

```
# Vue Ud : Unité documentaire
CREATE OR REPLACE
ALGORITHM = UNDEFINED
VIEW 'Ud' AS
SELECT
numero,ID,ED,AD,SN,TN,CO,CT,SU,NB,NU,TA,
TI,ASec,AU,NI,PT,PP,DP,NP,PI,DI,pdf_bool FROM 'dei_chapeau'
WHERE 'ST' = 'PMONO'
UNION
SELECT
numero,ID,ED,AD,SN,TN,CO,CT,SU,NB,NU,TA,
TI,ASec,AU,NI,PT,PP,DP,NP,PI,DI,pdf_bool
FROM 'dei_quantum'
ORDER BY 'numero' ASC
```

```
# Vue Uc : Unité de catalogage
CREATE OR REPLACE
ALGORITHM = UNDEFINED
VIEW 'Uc' AS
SELECT *
FROM 'dei_chapeau'
WHERE 'ST' != 'PMONO'
```

```
# Vue Ucd : Unité de catalogage et documentaire
CREATE OR REPLACE
ALGORITHM = UNDEFINED
VIEW 'Ucd' AS
SELECT
```

```

numero,ID,ED,AD,SN,TN,CO,CT,SU,NB,NU,TA,
TI,ASec,AU,NI,PT,PP,DP,NP,PI,DI,pdf_bool
FROM 'dei_chapeau'
UNION
SELECT
numero,ID,ED,AD,SN,TN,CO,CT,SU,NB,NU,TA,
TI,ASec,AU,NI,PT,PP,DP,NP,PI,DI,pdf_bool
FROM 'dei_quantum'
ORDER BY 'numero' ASC

```

```

# Vue cle_lieu
CREATE OR REPLACE
ALGORITHM = UNDEFINED
VIEW 'cle_lieu' AS
SELECT
numero,ST,ID,ED,AD,date_cat, NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,
fait,PI,DI,PE,NO,date_mise,dateoai,pdf_bool,SET_OAI,largeur,hauteur,
cle_lieu1,cle_nom1,cle_nom2,cle_nom3,cle_date,NTI,
cle_TC,sous_droit FROM 'dei_chapeau'
WHERE cle_lieu1 !=''
UNION
SELECT
numero,ST,ID,ED,AD,dat
e_cat,NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,
fait,PI,DI,PE,NO,date_mise,dateoai,pdf_bool,SET_OAI,largeur,
hauteur,cle_lieu2,cle_nom1,cle_nom2,cle_nom3,cle_date,NTI,cle_TC,sous_droit
FROM 'dei_chapeau'
WHERE cle_lieu2 !=''
UNION
SELECT
numero,ST,ID,ED,AD,date_cat,NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,
fait,PI,DI,PE,NO,date_mise,dateoai,pdf_bool,SET_OAI,largeur,
hauteur,cle_lieu3,cle_nom1,cle_nom2,cle_nom3,cle_date,NTI,cle_TC,sous_droit
FROM 'dei_chapeau'
WHERE cle_lieu3 !=''
ORDER BY 'numero' ASC

```

```

# Vue cle_nom
CREATE OR REPLACE

```

```

ALGORITHM = UNDEFINED
VIEW 'cle_nom' AS
SELECT
numero,ST,ID,ED,AD,date_cat,NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,
fait,PI,DI,PE,NO,date_mise,dateoai,pdf_bool,SET_OAI,largeur,
hauteur,cle_lieu1,cle_lieu2,cle_lieu3,cle_nom1,cle_date,NTI,cle_TC,sous_droit
FROM 'dei_chapeau'
WHERE cle_nom1 !=''
UNION
SELECT
numero,ST,ID,ED,AD,date_cat,NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,
fait,PI,DI,PE,NO,date_mise,dateoai,pdf_bool,SET_OAI,largeur,
hauteur,cle_lieu1,cle_lieu2,cle_lieu3,cle_nom2,cle_date,NTI,cle_TC,sous_droit
FROM 'dei_chapeau'
WHERE cle_nom2 !=''
UNION
SELECT
numero,ST,ID,ED,AD,date_cat,NQ,SN,NN,TN,CO,CT,CL,SU,NB,NU,LG,TA,
TC,TI,ASec,AC,AU,NI,PT,PP,DP,NP,QU,fait,PI,DI,PE,NO,date_mise,documentaire
dateoai,pdf_bool,SET_OAI,largeur,
hauteur,cle_lieu1,cle_lieu2,cle_lieu3,cle_nom3,cle_date,NTI,cle_TC,sous_droit
FROM 'dei_chapeau'
WHERE cle_nom3 !=''
ORDER BY 'numero' ASC

```

Glossaire

- *API* : Application programming interface
- *CEDRIC* : Centre d'étude et de recherche en informatique et communications
- *CNAM* : Conservatoire national des arts et métiers
- *CNUM* : Conservatoire numérique
- *Infovis* : Information visualization
- *IHM* : Interaction humain-machine
- *OCR* : Optical character recognition
- *RAID* : Redundant array of independent disks
- *SEO* : Search engine optimization
- *SIGB* : Système intégré de gestion de bibliothèque
- *SVG* : Scalable Vector Graphics
- *WAI* : Web accessibility initiative
- *W3C* : World wide web consortium
- *WCAG* : Web content accessibility guidelines

Index

Association des bibliophiles universels, [47](#) W3C, [64](#)
Web 2.0, [107](#)
workflow, [19](#)

Back office, [21](#)

Bibliothèque nationale de France, [17](#), [29](#)

Cédric, [14](#)

Conservatoire national des arts et métiers, [14](#), [18](#)

Conservatoire numérique, [17](#)

CVS, [18](#)

DSI, [19](#)

ENT, [19](#)

folksonomie, [107](#)

Front office, [23](#)

Gallica, [31](#), [62](#)

Google, [80](#)

ICR, [30](#)

MySQL, [18](#), [21](#)

OAI-PMH, [18](#), [21](#), [23](#), [29](#), [66](#), [86](#)

OCR, [82](#)

Page synoptique, [61](#), [62](#), [78](#)

Perl, [22](#)

RAID, [20](#)

Schéma directeur numérique, [19](#)

SEO, [80](#)

SIGB, [21](#)

Staging, [64](#)

SVG, [90](#)

Table des figures

1	La roue à livres de l'ingénieur Agostino Ramelli	3
1.1	Domaines de l'informatique des bibliothèques numériques	10
5.1	La métaphore du cockpit de [Top02, p. 175]	59
5.2	Perspective Wall de [MRC91, Plate 3]	59
5.3	Régions avec des nœuds de [LRP95, Figure 7]	60
6.1	Architecture du site	63
6.2	Activation de certains niveaux hiérarchiques des tables	80
6.3	Mise en valeur du niveau hiérarchique des titres de chapitre	82
6.4	Vignette de la page de titre soumis au droit d'auteur	92
6.5	Répartition des titres dans les thématiques 1	98
6.6	Répartition des titres dans les thématiques 2	98
6.7	Histogramme 1	99
6.8	Histogramme 2	99
6.9	Graphique en secteurs	100
6.10	Graphique en bulles	100
6.11	Heuristique visuelle du nombre de titres et nombre d'images	101
6.12	Visualisation par réduction extrême	101
6.13	Carte géolocalisée et clusterisation	102
A.1	Planning de développement - Phase 1	117
A.2	Planning de développement - Phase 2	117
B.1	Page d'accueil	119
B.2	Page Catalogue	120
B.3	Page synoptique d'une monographie, d'un volume de périodique ou d'une grande monographie	121
B.4	Page synoptique d'une monographie avec plusieurs tomes	122
B.5	Page synoptique d'une grande monographie	123
B.6	Page synoptique d'un périodique	124
B.7	Page feuilleteur	125
B.8	Page de recherche avancée	126
B.9	Page de résultat	127

TABLE DES FIGURES

C.1	Page d'accueil	129
C.2	Page d'accueil	130
C.3	Page catalogue avec les volumes	131
C.4	Page catalogue	132
C.5	Page synoptique 1	133
C.6	Page synoptique 2	134
C.7	Page vignettes	135
C.8	Page feuilleteur - maquette graphique	136
C.9	Page recherche	137
C.10	Page resultats vignette	138
D.1	Page d'accueil	140
D.2	Une des pages thématique	142
D.3	Catalogue général - classement par auteur	144
D.4	Catalogue général - classement par titre	145
D.6	Catalogue général - classement par lieu d'édition	147
D.7	Catalogue général - classement par éditeur	148
D.8	Notice bibliographique	150
D.9	Table abrégée	151
D.10	Table détaillée	152
D.11	Table des illustrations	153
D.12	Interface vignettes	154
D.13	Export des documents	155
D.14	Note de présentation	156
D.15	Liste des volumes	157
D.16	Liste des vignettes de la page de titre des volumes	158
D.17	Export des volumes	159
D.18	Feuilleteur en menu ouvert avec onglet vignette actif	161
D.19	Feuilleteur en menu ouvert avec onglet liste des volumes actif	162
D.20	Feuilleteur en menu fermé avec barre de navigation à droite	163
D.21	Feuilleteur en menu fermé avec barre de navigation en haut	164
D.22	Feuilleteur en menu fermé avec barre de navigation à gauche	165
D.23	Recherche avancée	167
D.24	Résultat mode abrégé	169
D.25	Résultat en mode texte	170
D.26	Résultat en mode vignette et texte	171
D.27	Résultat en mode vignette	172
D.28	Gestion des erreurs	174
D.29	Outils de découverte	175
D.30	Index alphabétique	176
D.31	A propos	177
D.32	Exposition virtuelle	178
D.33	Vignettes 2	180

TABLE DES FIGURES

D.34 Nuage de mots-clé	181
H.1 Le modèle conceptuel de données	190

Liste des tableaux

2.1	Nom des ingénieurs du Cnum	18
4.1	Historique mensuel de l'année 2011	33
4.2	Classement des systèmes d'exploitation	34
4.3	Classement des navigateurs Internet	34
4.4	Domaines/pays visiteurs	35
4.5	Classement des phrases clés recherchés	36
4.6	Top 10 des mots-clés recherchés	37
4.7	Lien sur le Cnum depuis un moteur de recherche Internet	38
4.8	Lien sur le Cnum depuis une page externe	39
4.9	Durée des visites	40
4.10	Classement des fichiers téléchargés	41
4.11	Classement des pages urls	42
4.12	Codes Status HTTP	43
4.13	Tableau du nombre de pages traduites dans une autre langue	44
5.1	Tableau comparatif des logiciels libres de bibliothèques numériques	54
5.2	Classement des moteurs de recherche par critères	57
5.3	Autre classement des moteurs de recherche par critères	58

INTERFACES DE CONSULTATION DE BIBLIOTHÈQUES NUMÉRIQUES : EXEMPLE DU SITE DU CNUM

Mémoire d'ingénieur CNAM en informatique. Paris, le 10 juillet 2015

RÉSUMÉ : La refonte du site web du Conservatoire numérique (<http://cnum.cnam.fr>), bibliothèque numérique en histoire des sciences et des techniques du Conservatoire national des Arts et Métiers est un projet de longue date. Nous avons réalisé une analyse de l'existant, recueilli les besoins des usagers et étudié divers logiciels de gestion de collections numériques afin de proposer la meilleure interface de recherche et de consultation. Avec l'aide d'un prestataire en webdesign, la conception du nouveau site s'est déroulée au travers des phases de maquettage, prototypage, test, développement et intégration suivies d'une concertation régulière avec tous les membres de l'équipe.

Mots-clés : Bibliothèque numérique, bibliothèque numérisée, Conservatoire numérique, Cnum, collection numérique, document numérique, interface de consultation, feuilleteur, visionneuse, recherche d'informations, base de données, interaction homme-machine (IHM), visualisation d'information, Internet, world wide web, Cnam, moteur de recherche libre, moteur Apache Lucene Solr, catalogue.

Abstract : The redesign of the « Conservatoire numérique » website (<http://cnum.cnam.fr>), digital library about history of science and technology of the « Conservatoire national des Arts et Métiers » is a longstanding project. We have achieved an analysis of the existing, collected user requirements and studied various digital collections management software to provide the best search interface and consultation. With the help of a service provider in web design, design of the new site took place through the modeling phase, prototyping, testing, development and integration followed by regular meetings with all members of the team.

Keywords : Digital repository, digital library, digital collection, digital document, search interface, image viewer, image browser, information search, database, human-computer interaction (HCI), infovis, Internet, world wide web, Cnam, open source search engine, Apache Lucene Solr, catalog.