

Les langages d'indexation

L'indexation consiste à attribuer à un document une marque distinctive renseignant sur son contenu et permettant de le retrouver. Il s'agit donc d'une opération intellectuelle par laquelle on va codifier le contenu d'un document pour ensuite le transcrire dans un langage documentaire.

Elle se traduit par la recherche d'un symbole numérique ou nominal à partir de l'analyse du contenu du document. Ce symbole peut être :

- tiré d'une classification (indice). On parle alors d'indexation systématique.
- constitué d'un ou de plusieurs mots-clés. On parle alors d'indexation analytique ou d'indexation alphabétique matière.

Les systèmes de classification faisant l'objet d'une fiche spécifique, nous nous attacherons ici à l'indexation analytique, également appelée indexation matière.

Indexer, c'est représenter le contenu d'un document au moyen d'un langage documentaire (thésaurus/vocabulaire contrôlé) ou bien au moyen du langage naturel libre (tag).

Idéalement, les termes utilisés pour l'indexation matière doivent renvoyer précisément au contenu du document analysé, sans ambiguïté et en toute objectivité. Pour respecter la cohérence d'un fonds documentaire, les catalogueurs doivent indexer les documents sur un même sujet selon les mêmes principes.

1. Historique

Jusqu'au milieu des années 1970, la recherche documentaire dans les bibliothèques et centres de documentation français s'effectuait essentiellement à partir de fichiers sur support papier utilisant comme langage documentaire des classifications décimales (Classification Décimale Universelle ou Classification Décimale de Dewey). Ces classifications servaient à la fois d'outils d'indexation (établissement d'indices de classification visant à représenter le contenu du document concerné) et d'outils de cotation (établissement d'une cote indiquant à quel endroit ce document avait été rangé).

Un changement radical s'est opéré entre les années 1970 et 1980 avec la décision, prise en 1976, par la Bibliothèque Publique d'Information d'utiliser le Répertoire de Vedettes Matière de l'Université Laval au Québec¹ lui-même traduit de la liste de vedettes-matière de la Bibliothèque du Congrès (Library of Congress Subject Headings). Ce répertoire fonctionne en termes clés et non plus en classification décimale. La Bibliothèque nationale de France suit le même mouvement en 1980 et les premiers manuels d'indexation voient le jour, avec par exemple :

- En 1981, Le cercle de la librairie publie la première édition du « Choix de vedettes matières à l'intention des bibliothèques » ;
- En 1985, l'AFNOR publiait la norme Z47-200 définissant la "Liste française d'autorité de matières" : "LAMECH" devenue "RAMEAU" en 1987.

L'utilisation des langages d'indexation se développe jusqu'à la décision de la Bibliothèque nationale de France, prise en 1990, de demander aux bibliothèques universitaires d'adopter RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié). Constitué d'une liste fermée de mots-clés, RAMEAU a notamment pour objectif d'harmoniser la description intellectuelle des contenus afin que les références puissent se retrouver d'un catalogue à l'autre. Par ailleurs, avec le développement des catalogues informatisés, la recherche documentaire par mots-clés permet aux lecteurs d'accéder plus aisément aux documents que par la seule classification décimale. L'indexation matière permet alors d'affiner sa recherche sur les contenus, et la cotation de retrouver le document sélectionné dans les rayonnages.

2. Méthodologie de l'indexation

L'indexation comprend trois étapes : l'analyse, la sélection et la transposition.

Tout d'abord, il faut **analyser le document** (en consultant le résumé, la table des matières, la page de titre, la quatrième de couverture...) pour déterminer le ou les sujets qui y sont traités.

Ensuite, on ne va pas forcément retenir toutes les thématiques du document : **on sélectionnera certains sujets, en fonction du niveau de précision** voulu pour l'indexation. Selon les fonds documentaires et le public visé, l'indexation sera plus ou moins précise. Dans le SUDOC (Système universitaire de documentation de

¹ Pour plus d'informations sur l'adoption par la BNF du répertoire de vedettes matières de l'Université de Laval, consulter l'article de Suzanne Jouguelet, paru dans le BBF en avril 1983 : <https://bbf.enssib.fr/consulter/bbf-1983-04-0383-003> (consulté le 28/08/2022)

l'enseignement supérieur) par exemple, l'indexation est rigoureuse et détaillée pour répondre aux exigences des chercheurs.

Enfin, on procède au choix des termes représentant le sujet. Les sujets peuvent être exprimés par une indexation en langage libre (ou en langage naturel) ou bien par un langage documentaire précis et non figé comme RAMEAU ou MeSH.

3. Indexation par mots-clés ou langage documentaire construit

Les termes choisis pour décrire le contenu d'un document peuvent soit être définis librement – on parle alors d'indexation par mots-clés – soit être issus d'une liste fermée avec des règles d'utilisation – il s'agit alors d'un langage documentaire construit, comme par exemple RAMEAU.

3.1. Indexation par mots-clés

L'indexation par mots-clés utilise le langage naturel. Le terme est mis dans une zone définie de la notice de catalogage. Ce système permet de multiplier les termes (synonymes, abréviations, ...) choisis par l'auteur lui-même ou par les catalogueurs et simplement ajoutés les uns à la suite des autres.

L'absence de règles de construction et la liberté de choix dans le vocabulaire représentent à la fois un avantage et un inconvénient : si cette liberté permet un degré de précision important, elle peut aussi générer des incohérences et ambiguïtés avec le risque de multiplier les mots-clés se rapportant à un même domaine, ce qui serait alors contre-productif.

L'indexation matière en langage naturel se rencontre fréquemment sur le web et notamment sur les plateformes de dépôts de documents.

Elles peuvent être :

- institutionnelles : les bases d'archives ouvertes comme [TEL](#) et [HAL](#), gérées par le CNRS, où les chercheurs peuvent indexer leurs travaux en langage naturel. Cela leur permet d'y mettre des termes utilisés dans le cadre de leurs recherches.
- ou grand public : Flickr ou Instagram pour le partage de photos ou Youtube pour les vidéos. Les internautes ne se contentent pas de déposer leurs documents, ils les étiquettent (*to tag* en anglais) de manière tout à fait subjective et libre.

L'indexation par mots-clés est également utilisée dans le SUDOC pour les notices de thèses. Les thèses sont indexées avec RAMEAU mais aussi avec les mots clés définis au préalable par l'auteur de la thèse.

Le nuage de mots-clés (ou tag cloud) se rencontre également dans les bases de données bibliographiques. Le principe est simple : plus un terme est cité dans un article, plus le terme affiché grossit dans le nuage. Ce terme est cliquable, il renvoie aux articles le concernant et apparaît comme une aide supplémentaire et différente à la recherche documentaire. Certains catalogues de bibliothèques proposent ce système avec plus ou moins de succès.

3.2. L'indexation via un langage documentaire construit (ou indexation analytique)

Un langage documentaire est un langage contrôlé et normalisé utilisé dans les catalogues de bibliothèques ainsi que dans d'autres applications de signalement comme ORI-OAI (Outil de référencement et d'indexation pour un réseau de portails OAI-PMH) ou bien dans les bases de données bibliographiques.

« Un langage documentaire est un langage artificiel, ...constitué de notions et de relations entre ces notions... C'est un système de représentation synthétique du contenu des textes. » (Dictionnaire encyclopédique de l'information et de la documentation / Serge Cacaly, 1997).

Deux langages documentaires sont principalement utilisés en France : RAMEAU et MeSH. Notons que lorsque les catalogueurs indexent leurs documents, il est prévu dans les grilles de description de spécifier le langage utilisé. On retrouve alors dans la plupart des notices utilisant un langage d'indexation construit la mention « fmesh » ou « rameau » en fin de ligne.

4. RAMEAU

RAMEAU, Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié, est le langage d'indexation utilisé en France. La BnF, les bibliothèques universitaires qui alimentent le SUDOC, et les bibliothèques de lecture publique ou de recherche ainsi que plusieurs organismes privés comme Electre, utilisent ce langage documentaire.

Ces organismes sont réunis au sein du **réseau national RAMEAU**² depuis la signature d'une convention en 2001, convention reconduite en 2011, afin de faire vivre ce langage, de l'alimenter et d'en assurer la pertinence de façon collégiale.

Le langage RAMEAU est aussi au cœur d'un certain nombre de **coopérations internationales** :

- Un partenariat historique lie la BnF à l'Université Laval (Québec).
- RAMEAU est utilisé par le réseau suisse RERO ainsi que par de nombreux établissements documentaires belges francophones qui participent à son enrichissement par l'intermédiaire du Fichier national des propositions RAMEAU.
- Il a été adapté en polonais (langage KABA) et en roumain (langage LIVES-ro).
- Il est utilisé dans un certain nombre de pays d'Afrique du Nord et du Proche-Orient (un projet d' "arabisation" de RAMEAU est envisagé), ainsi que dans des pays francophones d'Afrique subsaharienne.

Le réseau RAMEAU est animé par le **Centre national RAMEAU**, qui dépend du Département des métadonnées de la Direction des services et des réseaux de la Bibliothèque nationale de France. Il est responsable de la gestion intellectuelle du langage au niveau national et assure via le site web RAMEAU³ la diffusion des différents outils et informations nécessaires à l'utilisation et à l'enrichissement du langage.

Le langage RAMEAU est un langage documentaire **encyclopédique car il touche à tous les domaines de la connaissance**.

Le vocabulaire RAMEAU se présente sous forme de notices d'autorité matière : noms communs, noms géographiques, subdivisions chronologiques, noms de personnes, noms de collectivités, titres uniformes. Ces autorités sont accessibles via le catalogue général de la BnF dans l'interface de recherche Autorités et sont chargées périodiquement dans [ldRef](#). ldRef (Identifiants et Référentiels) est une application Web gérée par l'ABES qui permet d'interroger, de consulter, de créer et d'enrichir des notices autorités. Elle est actuellement utilisée dans les bases SUDOC, STAR , STEP et Calames.

² <https://rameau.bnf.fr/informations/convention> (consulté le 11/07/2024)

³ <http://rameau.bnf.fr/> (consulté le 11/07/2024)

Ces autorités matières peuvent être combinées entre elles selon des règles établies pour décrire le plus précisément possible le contenu d'un document. On parle alors de vedettes-matières, ou de points d'accès construits.

On pourra par exemple, sur un sujet principal tel que la randonnée pédestre choisir une spécification géographique, et pourquoi pas une spécification de temps pour indiquer que l'on cherche un document sur les randonnées dans une région particulière à une époque précise. Le degré de spécification d'une vedette matière dépendra du contenu du document (s'il traite un sujet de manière générale ou sous un aspect particulier).

RAMEAU est un langage **contrôlé**, un terme est retenu pour un concept et des renvois sont faits vers des termes exclus (ou termes rejetés) dans les notices autorités pour faciliter la recherche par mots. Le vocabulaire est enrichi régulièrement par les catalogueurs du réseau, qui peuvent demander via **le Fichier national des propositions RAMEAU (FNPR)** la création ou la modification d'autorités RAMEAU.

RAMEAU est un langage **hiérarchisé**, il existe des relations hiérarchiques et/ou associatives entre les différentes notions, ce qui permet une navigation sémantique entre les différentes autorités RAMEAU.

RAMEAU est un langage pré-coordonné, c'est-à-dire que les **points d'accès matière** obéissent à des règles syntaxiques précises afin d'assurer une cohérence dans l'indexation des documents.

Les règles de construction sont répertoriées dans le **guide d'indexation RAMEAU** disponible sur le site RAMEAU.

En voici le modèle général :

Concept -- Lieu (subdivision géographique) - - Date (subdivision chronologique) - -
Forme du document (subdivision de forme).

Par exemple : Peinture - - Italie - - 17e siècle - - catalogues d'exposition

Il est à noter qu'il existait jusqu'à présent de nombreuses exceptions à la règle, par exemple certains noms communs n'admettaient pas de subdivision géographique.

En Unimarc, les autorités RAMEAU se trouvent dans les zones 600 (nom de personne), 601 (collectivité), 602 (nom de famille, de dynastie), 604 (Auteur. Titre), 605 (Titre sans auteur), 606 (nom commun), 607 (nom de lieu).

Focus : Réformer RAMEAU⁴

Conçu pour permettre la recherche dans un index de sujets, le langage RAMEAU doit évoluer pour répondre :

- Aux pratiques des usagers et aux nouveaux outils de recherche qui privilégient la recherche par mots.
- Au modèle IFLA-LRM.
- Au web de données.
- A l'interopérabilité des données.

En mars 2016 un groupe de travail national chargé de formuler des préconisations sur les nécessaires évolutions de la syntaxe RAMEAU est nommé. Composé de représentants des utilisateurs de Rameau (ABES, bibliothèques de lecture publique, bibliothèques universitaires, BnF), ce groupe de travail a remis, en mars 2017, un rapport⁵ comportant des préconisations. Une feuille de route⁶ l'accompagne, qui inscrit les décisions à mettre en œuvre d'ici 2022.

Ces modifications rendent en partie obsolète le Guide d'indexation RAMEAU, on trouvera les consignes de catalogage liées à ces évolutions sur le site RAMEAU, dans

⁴ <https://rameau.bnf.fr/syntaxe/introduction> (consulté le 12/07/2024) recense l'historique et l'ensemble des documents relatifs à la réforme RAMEAU.

⁵ https://rameau.bnf.fr/sites/default/files/chantier_syntaxe/pdf/rapport_final_syntaxe_rameau.pdf (consulté le 12/07/2024)

⁶ https://rameau.bnf.fr/sites/default/files/chantier_syntaxe/pdf/feuille_route_reforme.pdf (consulté le 12/07/2024)

la rubrique Réformer Rameau⁷, partie Nouvelles consignes pour l'indexation sujet RAMEAU.

5. MeSH : Medical Subject Headings

Le MeSH est un thésaurus⁸ américain de référence dans le **domaine biomédical**. Il est produit et mis à jour par la National Library of Medicine depuis 1954. L'Inserm est responsable de la traduction et de la mise à jour annuelle de la version française du MeSH, FMeSH⁹ en collaboration avec l'Inist, Institut de l'information scientifique et technique.

Contrairement à RAMEAU, ce langage d'indexation n'est pas encyclopédique mais comme RAMEAU, il s'agit d'un langage normalisé, structuré et évolutif constitué par des autorités appelées **descripteurs**. Les descripteurs sont organisés hiérarchiquement au sein d'une arborescence.

Les descripteurs peuvent être :

- **simples**, par ex : Foie
- ou **coordonnés**, ceux-ci correspondent généralement à une notion plus spécifique, par ex : Troubles de l'alimentation et des conduites alimentaires de l'enfant.

On utilise le descripteur le plus spécifique pour traduire un concept. S'il n'existe pas de descripteur spécifique, on peut ajouter au descripteur un **qualificatif** (il en existe 83) comme par ex : Foie/chirurgie pour indexer une ressource sur la chirurgie du foie.

Une **catégorisation de la population** est possible par tranche d'âges. Ces groupes sont utilisés en tant que descripteurs si le contenu du document en est le thème principal, sinon, ils s'utilisent en sous-vedettes pour préciser le descripteur.

Exemple : Enfant (de 6 à 12 ans) Adolescent (de 13 à 18 ans) Jeune adulte (de 19 ans à 24 ans)

⁷ <https://rameau.bnf.fr/syntaxe/reformer>

⁸ Répertoire alphabétique de termes normalisés pour l'analyse de contenu. Ces termes sont reliés entre eux par des relations sémantiques hiérarchisées.

⁹ Voir le site <http://mesh.inserm.fr/FrenchMesh/> (consulté le 12/07/2024)

De plus, il existe des **subdivisions de langue**, des subdivisions **géographiques** et **chronologiques** et d'autres qui portent sur le **type de publication** (ex : dissertations académiques).

L'ordre des sous-vedettes est le suivant :

1. Descripteur
2. Qualificatif ou groupe d'âge
3. Subdivision géographique
4. Subdivision de langue
5. Type de publication

MeSH est utilisé dans les bases de données de la NLM dont MEDLINE/PubMed.

[PubMed](#) n'étant indexé qu'en MeSH anglais, l'utilisation de la traduction de l'Inserm s'avère précieuse pour connaître l'équivalent du terme en anglais et pouvoir ainsi effectuer une recherche pertinente dans cette base.

La version française, FMeSH, est utilisée dans le Sudoc¹⁰.

En UNIMARC :

```
606 ##$aDescripteur $squalificatif ou groupe d'âge $ysubdivision géographique  
$2fmesh
```

Conclusion

Quel que soit le langage d'indexation utilisé, les bibliothécaires n'en oublient pas l'objectif principal : donner des clés d'accès aux documents afin que les lecteurs puissent les retrouver. C'est pourquoi, au-delà de l'expertise technique dans l'utilisation des langages d'indexation, les bibliothécaires doivent faire preuve de logique et bien connaître le fonds sur lequel ils travaillent afin d'assurer la cohérence des descriptions. L'indexation est certes un outil de signalement, mais elle permet aussi de discriminer des documents d'un même domaine les uns par rapport aux autres. Ainsi, l'expertise des bibliothécaires consiste également à définir la granularité de la description et à adapter son utilisation du langage d'indexation au fonds dont il s'occupe.

¹⁰ Voir <http://documentation.abes.fr/sudoc/regles/Indexation/FMeSH/index.htm#Generalites> (consulté le 11/07/2024)