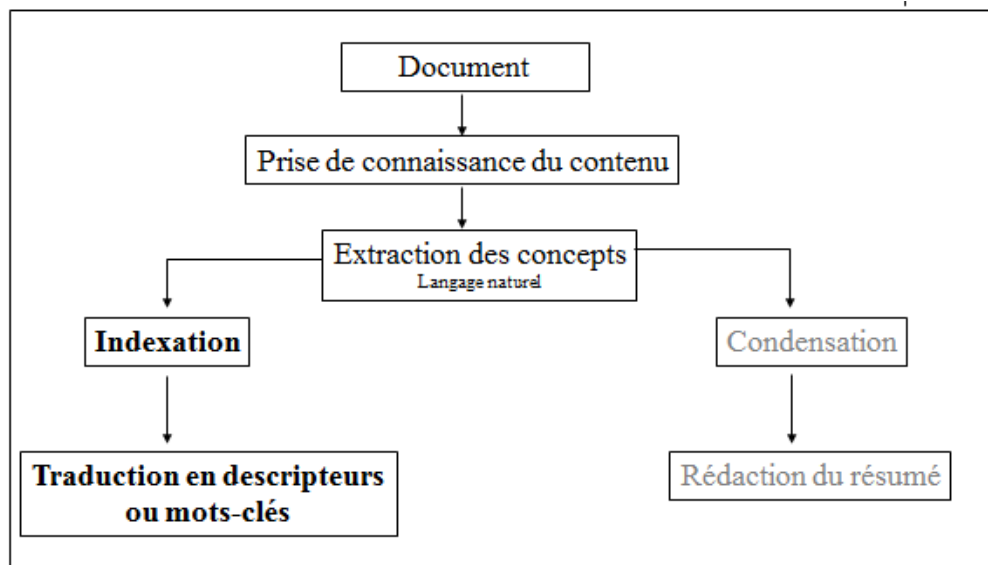


Repères pour l'indexation – Documentation Ehesp

Définition et rôle de l'indexation

L'indexation est l'une des formes de la représentation du contenu des documents.

- Elle vise à rendre compte sous une forme concise des informations stockées dans un fonds documentaire.
- Elle permet de faire le lien entre une offre d'informations (le fonds documentaire) et une demande d'informations (les questions des utilisateurs).
- Elle permet d'élaborer des produits documentaires : bibliographies, bulletins, newsletters, catalogues...
- Elle est le résultat de l'analyse documentaire, au même titre que le résumé :



Indexer donc consiste à **décrire le contenu du document** (descripteurs, carte d'identité du document) en vue d'une **exploitation** par un **utilisateur**.
L'indexation doit **faciliter les recherches** et **garantir la pertinence des réponses**.

Le thesaurus

Le thesaurus est un **langage contrôlé** et dynamique de termes ayant entre eux des relations sémantiques et génériques replacés dans leur champ sémantique.

Le **champ sémantique** est un groupe de mots qui par leur proximité permettent de replacer un concept dans un domaine bien précis (ex. Organisation des soins).

Le descripteur doit être **univoque** (un descripteur désigne un seul concept) ; cette univocité est assurée grâce aux différentes fonctions du thesaurus :

- la synonymie qui est réglée par les relations d'équivalence (EM /EP)
- la polysémie, qui est réglée par trois dispositifs (l'appartenance au groupe sémantique (MT), le niveau hiérarchique (TG / TS), la possibilité de choisir une expression comme descripteur (ex. Promotion de la possibilité d'ajouter une note explicative au descripteur).

- Les termes équivalents, appelés aussi non-descripteurs, sont les synonymes, les quasi-synonymes et les termes voisins, liés par une relation d'équivalence sémantique avec un descripteur du thesaurus.

Les règles d'indexation

1. Étapes de l'indexation

Une indexation performante est le résultat d'une analyse de contenu du document qui doit se faire à trois niveaux :

- Niveau conceptuel

La reconnaissance des concepts suppose une prise de connaissance globale du document. Pour un document imprimé, on s'appuiera sur :

le titre, s'il est significatif, le sommaire, les sous-titres s'ils existent, l'introduction, la conclusion du document, les notions mises en valeur par certains artifices typographiques.

L'indexeur répondra à un certain nombre de questions, celles que poserait un utilisateur :

- de qui, de quoi parle le document ?
- de quel point de vue ?
- quelle zone géographique concerne-t-elle ?

Par exemple, un document peut traiter : des enfants handicapés (= qui ?), de leur cursus scolaire (= de quel point de vue ?), en France (= où ?).

- Niveau du langage naturel

Les concepts ainsi définis seront inscrits dans un premier temps en langage naturels, on peut noter également les synonymes d'un même terme (voir grille d'analyse QQQCP).

- Niveau du langage documentaire

La traduction des concepts représentés par des mots-clés en descripteurs du thesaurus est l'opération d'indexation proprement dite.

2. Choix des concepts

Le choix des concepts à représenter repose sur deux règles qui peuvent paraître contradictoires : l'exhaustivité, la sélectivité. Au nom de l'exhaustivité, tous les concepts pour lesquels le document apporte une information significative doivent être retenus. Au nom de la sélectivité :

- les seuls concepts à retenir sont ceux susceptibles d'intéresser l'utilisateur et de lui apporter une information significative sur des centres d'intérêt actuels ou prévisibles ;
- les sujets peu représentés dans le centre de documentation ou marginaux par rapport aux préoccupations de la majorité des usagers pourraient être analysés moins finement que les autres.

3. Profondeur de l'indexation

La profondeur de l'indexation, c'est-à-dire le nombre de descripteurs affectés à un document, ne doit pas être limitée arbitrairement, elle doit dépendre :

- de la quantité et de la pertinence de l'information apportée par ce document,
- de la nécessité de prendre en compte deux paramètres quelquefois difficiles à concilier : le traitement de son fonds propre parfois très limité, le partage des tâches en réseau avec des centres de documentation dont le fonds est plus riche,
- des contraintes imposées par le logiciel documentaire utilisé (certains systèmes informatiques n'autorisent que cinq descripteurs).

1. Indexation de la question

Comme l'indexation d'un document, une recherche documentaire se prépare.

Le processus d'analyse de contenu est le même (cf. grille QQQQCP remise aux étudiants lors de la formation à la recherche documentaire).

Dans tous les cas, il est préférable de connaître non seulement le langage documentaire utilisé pour savoir formuler sa question, mais également les caractéristiques propres du système dans lequel on recherche (choix du catalogue ou de la base de données interrogée).

Outil de recherche hiérarchique et thématique, le thésaurus, peut fournir une aide fructueuse dans l'accès au contenu d'une base documentaire. En explorant les relations de synonymie, de hiérarchie et d'association qu'entretiennent entre eux les concepts du thésaurus, on peut définir, délimiter ou élargir le sujet de la recherche.