La minute du PSIR

1er juin 2022 - # 50

Différence entre métadonnées et vocabulaire contrôlé

NIVEAU DE DIFFICULTÉ

Services/outils

Alyssa Turgis

PERSONNES CONCERNÉES

Les métadonnées sont des données permettant de définir ou de décrire une autre donnée. Elles sont à la base du web sémantique (voir minute #21). Alors que le vocabulaire contrôlé correspond à une liste de termes prédéfinis permettant d'organiser les connaissances et d'optimiser la recherche d'information, mais également d'uniformiser les connaissances (voir minute #16). Ces deux modes d'indexation peuvent être complémentaires, mais ne s'utilisent pas de la même manière. Voici quelques lignes pouvant guider le choix.

La différence entre les deux modèles

Un vocabulaire contrôlé est structuré, hiérarchisé, et respecte des normes précises, alors que les métadonnées sont des données subjectives (définies par chaque utilisateur) : il existe autant de métadonnées que d'utilisateurs, *a contrario* un thésaurus a pour objectif de limiter la création de concepts identiques. L'autre grande différence entre ces deux modèles est l'utilisation d'identifiants pérennes dans un thésaurus : ils permettent d'identifier des données sur le long terme (voir minute #15).

Les avantages et les inconvénients du modèle de métadonnées

- **Avantages :**
  - Clarifier les documents.
  - Facile à mettre en œuvre.
  - Facilite la recherche dans des documents hétérogènes.
- **Inconvénients :**
  - Les métadonnées sont subjectives et dépendent du contexte dans lequel le document est créé, ce qui ne facilite pas leur exploitation.
  - Cela peut être chronophage lorsqu'on doit traduire les métadonnées en plusieurs langues dans le cadre d'un document multilingue.

Les avantages et les inconvénients du modèle de vocabulaire contrôlé

- **Avantages :**
  - Vocabulaire commun facilitant la compréhension entre utilisateurs et les échanges de données (interopérabilité, voir minute #22).
  - Existence d'outils permettant de faciliter la création et l'enrichissement des thésaurus, comme Opentheso.
  - Améliore la recherche documentaire, comme dans le catalogue Frantiq.
  - Possibilité de chercher par terme associé ou par rebond vers les génériques et les spécifiques.
  - Possibilité de rechercher un terme dans une langue : la recherche se fait automatiquement sur les autres langues disponibles dans le thésaurus.
- **Inconvénients :**
  - Nécessite des connaissances de base sur la structuration d'un thésaurus.
  - Demande un effort de réflexion sur la définition des termes, et leur utilisation.
  - Difficile de trouver une infrastructure ayant un serveur d'hébergement : à la MOM, il existe un serveur maintenu et disponible pour héberger vos thésaurus.

Pour aller plus loin :

- Carnet Hypothèses d'Opentheso, gestionnaire de thésaurus multilingue et multi-hiérarchique
- Page Wikipédia sur les métadonnées
- Page de l'ENSSIB sur les thésaurus