

Livre blanc

www.afnor.org

ISO 25964-1 - Thésaurus pour la recherche documentaire

Edition Janvier 2013

Sylvie Dalbin, Nathalie Yakovleff, Hélène Zysman.
Avec la collaboration de François Feyler,
Roselyne Bloch, Michèle Lénard, Katell Briatte.



**DOCUMENT PROTÉGÉ
PAR LE DROIT D'AUTEUR**

Droits de reproduction réservés. Sauf prescription différente, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans accord formel.

Contacteur :

AFNOR – Norm'Info

11, rue Francis de Pressensé

93571 La Plaine Saint-Denis Cedex

Tél : 01 41 62 76 44

Fax : 01 49 17 92 02

E-mail : norminfo@afnor.org

afnor

Livre blanc

ISO 25964-1 – Thésaurus pour la recherche documentaire

Sommaire

Préambule	2
1 - Enjeux et contexte	3
1.1. Évolution des usages des thésaurus	3
1.2. Pourquoi normaliser sur ce domaine ? Pourquoi une nouvelle norme ?	3
1.3. Périmètre et contexte de la norme ISO 25964	4
1.4. Une norme en deux volets	5
2 - Présentation générale de la Partie 1 de la norme ISO 25964	6
2.1. Termes, définitions et symboles	6
2.2. Sémantique métier : construire et maintenir un vocabulaire structuré.....	7
2.3. Sémantique technique : modèle de données, schéma XML et documentation technique	7
3 - Composants d'un thésaurus	8
3.1. Thésaurus.....	9
3.2. Concepts	10
3.3. Termes.....	12
3.4. Relations.....	14
3.5. Regroupement de concepts : thèmes et facettes.....	16
3.6. Multiculturalisme et multilinguisme.....	20
4 - Usages du thésaurus et de ses composants	21
4.1. Présentation des thésaurus pour consultation	21
4.2. Intégration des thésaurus à des applications.....	23
4.3. Formats d'échange et protocoles	23
4.4. Conception, gestion et maintenance.....	24
5 - Conclusion : Que faire ? Comment faire ?	25
Annexes	26
Annexe 1 : Lexique Anglais-Français du vocabulaire de la norme	26
Annexe 2 : Tableau partiel des symboles et sigles.....	31
Annexe 3 : Modélisation entités-relations – quelques repères	32
Annexe 4 : Ressources bibliographiques.....	35
Annexe 5 : Organisation de la production de la norme ISO 25964.....	36

Préambule

Ce Livre Blanc propose à la communauté francophone une présentation générale de la Partie 1 de la norme **ISO 25964-1 - Thésaurus pour la recherche d'information** parue officiellement, dans sa version anglaise, en août 2011.

La norme ISO 25964-1:2011

La nouvelle norme ISO 25964-1:2011 annule les deux anciennes normes internationales ISO 2786:1986 (thésaurus monolingue) et ISO 5964:1985 (thésaurus multilingue), ainsi que les normes françaises NF Z47-100:1981 (thésaurus monolingue) et NF Z47-101:1990 (thésaurus multilingue).

La norme ISO 25964-1:2011 fournit des directives et des recommandations applicables à la conception et la gestion des thésaurus monolingues et multilingues ainsi qu'à leur mise en forme afin d'optimiser leur exploitation dans des applications et sur les réseaux.

Comment lire ce Livre Blanc

Ce Livre Blanc expose les grandes lignes et les nouveautés d'un guide métier de 180 pages. Son objectif est de fournir avant tout une vue synthétique de ce document de norme à l'attention de professionnels ayant déjà une connaissance des outils d'organisation que sont les thésaurus, taxonomies, classification, vocabulaires contrôlés en général : professionnels de l'information-documentation, informaticiens en charge de dispositifs d'accès à l'information, architectes de l'information,

Ce n'est pas un guide pratique ou méthodologique de création ou de mise à jour de thésaurus. La sélection des directives et recommandations de la norme, exposées dans ce Livre Blanc aussi bien que l'organisation et le format de ce document, visent plutôt à une prise de conscience par les acteurs les plus impliqués dans les systèmes documentaires, des enjeux sémantiques et techniques actuels et à leur fournir des pistes sur les actions à mener dans le moyen terme.

Motiver la communauté pour porter un projet de traduction en français de la norme est le deuxième objectif de ce Livre Blanc.

Le Livre Blanc de la norme ISO 25964-1 met en exergue les points saillants de la norme et précise lorsque cela s'avère important pour la communauté francophone, leur caractère obligatoire. Chaque chapitre de ce Livre Blanc fait référence à des parties précises de la norme et en expose les points-clés. Les caractéristiques plus techniques (appelées dans le Livre Blanc « sémantique technique ») sont exposées dans un sous-chapitre dédié « > *Modèle de données ISO 25964-1* ».

1 - Enjeux et contexte

1.1. Évolution des usages des thésaurus

L'histoire des premières normes pour le développement des thésaurus est étroitement liée à une utilisation humaine de thésaurus sur support papier. Les thésaurus sur support papier ont progressivement cédé la place aux thésaurus actuels exploités essentiellement dans des dispositifs informatiques pour l'indexation et la recherche, ou comme ressource terminologique en consultation. Élaborés et maintenus à partir de logiciels ou modules spécialisés, les thésaurus sont exploités par d'autres logiciels ou applications tels que des moteurs de recherche ou des systèmes de gestion de contenu.

Conçus initialement pour des professionnels formés à l'indexation et à la recherche ou pour des spécialistes ayant une expertise dans le domaine de connaissances couvert, les thésaurus doivent aujourd'hui répondre de plus en plus aux besoins et pratiques d'utilisateurs non entraînés et/ou ayant des connaissances moins affirmées dans le domaine traité. La recherche fédérée conduit quant à elle à manipuler des ressources d'origines et de natures diverses pouvant être traitées selon des protocoles et/ou des vocabulaires différents. Plus récemment, à la suite du développement des technologies du Web, les thésaurus sont mis à profit dans les systèmes terminologiques permettant à des machines d'effectuer des inférences.

Les conséquences de ces évolutions portent aussi bien sur les caractéristiques conceptuelles, terminologiques ou structurelles des thésaurus eux-mêmes, sur leur place par rapport à d'autres outils sémantiques telles que les classifications ou les ontologies, ainsi que sur les modalités de conception et d'intégration dans les systèmes documentaires.

1.2. Pourquoi normaliser sur ce domaine ? Pourquoi une nouvelle norme ?

L'application des normes dans différents domaines de l'activité humaine comme l'énergie, le transport ou la construction, est considérée comme essentielle et jamais remise en question. Mais il semble plus difficile de percevoir l'intérêt de standardiser à un niveau international, des composantes non informatisées des systèmes d'information¹ ou qui apparaissent comme telles. Pourtant, depuis une quinzaine d'années se sont développées des pratiques de normalisation dans les activités de management ou sur des aspects méthodologiques des systèmes d'information considérés comme des éléments-clés de leur performance telles que l'ergonomie ou les démarches qualité. A ce critère lié au périmètre même des normes, s'ajoute la prise en compte d'une situation nouvelle : la propagation du phénomène de numérisation et le déploiement en masse d'informatisation numérique auxquelles nous assistons conduit à la multiplication des échanges et des flux d'information sur les réseaux et entre les systèmes d'information. Ce nouveau contexte rend indispensable le développement de règles communes sous la forme de normes et de standards² adaptés à la représentation, au transfert, à l'accès et à la réutilisation des connaissances. C'est dans ce contexte que se place cette norme sur les thésaurus.

¹ Voir Michèle Hudon, ISKO 2011

² Le terme « norme » ou « norme technique » est souvent réservé en particulier en France, à une "Spécification technique approuvée par un organisme reconnu à activité normative pour application répétée ou continue, dont l'observation n'est pas obligatoire" [au sens de la directive 2004/18/CE du Parlement européen et du Conseil (et directive 2004/17/CE). Il se distingue alors du terme « standard », dont les spécifications s'imposent par le marché.

Le thésaurus faisait déjà l'objet de documents normatifs depuis les années 1970 ([voir Annexe 4](#)). Mais alors même que les usages et les technologies évoluaient, on ne pouvait que constater l'absence de rigueur dans la maintenance de ces normes, en particulier au niveau international. Initié par la Grande-Bretagne rejointe en 2007 par plusieurs pays au sein du groupe de travail TC46/SC9/WG8³ de l'ISO, ce travail de refonte des normes sur les thésaurus mono- et multilingues devenait indispensable.

La norme ISO 25964:2011 offre la transition nécessaire au monde de la communication numérique de l'information. Tout en restant en harmonie avec les normes et les systèmes plus anciens, elle offre un cadre normatif pour faire évoluer l'existant et pour développer des outils adaptés aux situations actuelles.

1.3. Périmètre et contexte de la norme ISO 25964

1. Périmètre et contenu de la norme ISO 25964

La norme ISO 25964 s'applique :

- à des vocabulaires composés de **termes** représentant des concepts et/ou des sujets ;
- à des vocabulaires **structurés** organisant ces ensembles de concepts et de termes entre eux ;
- à des vocabulaires utilisables lors de processus de **recherche d'information** ;
- quel que soit le type de **ressources d'information** et quel que soit le **support** utilisé (texte, son, image fixe ou animée, objet physique ou multimédia) ; y compris les bases de connaissances et les portails, les bases de données bibliographiques, les bases de textes, les collections de musées ou les collections multimédias, et pour retrouver des éléments au sein de ces collections.

2. Contexte de production de la nouvelle norme

Les fondements de ces vocabulaires (répertoires de concepts, termes représentant les concepts, hiérarchie entre concepts,...) et leur développement tels qu'exposés dans les documents normatifs des années 1980 restent valables. Mais l'utilisation des thésaurus comme outils de repérage et d'accès à l'information pour des collections aujourd'hui numériques, la variété de leurs contextes d'application et des contraintes qui en découlent, en particulier l'exigence d'interopérabilité entre plusieurs applications en réseau, ainsi que les fonctionnalités requises pour répondre à cette diversité d'applications ont fortement évolué, rendant ces documents normatifs difficiles d'exploitation.

Ce constat a poussé le groupe ISO en charge des travaux normatifs à procéder à une révision en profondeur des anciennes normes du domaine, tout en s'appuyant sur celles-ci.

Ces anciennes normes, matière première de ce travail international et francophone, sont issues de quatre sources principales :

- Les normes internationales : ISO 2788:1986 (monolingue) et ISO 5964:1985 (multilingue) ;
- Les normes françaises NF Z47-100:1981 (monolingue) et NF Z47-101:1990 (multilingue) ;
- La norme BS 8723 développée plus récemment entre 2000 et 2007, qui a servi de base au projet de norme dans le processus ISO ;
- Enfin la norme ANSI/NISO Z39.50:2005.

Les travaux initiés en 2008 ont été réalisés au sein du Groupe de travail ISO WG8 du TC46/SC9, piloté par Stella Dextre Clarke.

Le groupe ISO s'est donné un cadre de travail précis combinant :

³ L'Annexe 5 fournit des précisions sur l'organisation des travaux réalisés au niveau ISO et l'implication du Groupe de travail français rattaché à la commission Afnor/GC46/CN357 « Modélisation, production et accès aux documents.

- Une grande ouverture sur les usages et les technologies les plus récentes dont le Web de données, d'où un modèle de données (normatif) associé à une documentation technique à destination des développeurs (informatif).
- Une filiation visible avec les normes plus anciennes.

Le choix s'est porté sur un texte à la fois normatif et explicatif ou plus descriptif. Ainsi la norme distingue les règles qui doivent (« must »), peuvent (« should ») ou pourraient (« could ») être appliquées.

1.4. Une norme en deux volets

La norme ISO 25964 est constituée de deux parties distinctes mais interdépendantes.

La Partie 1 - *Thésaurus et recherche d'information* offre des directives et recommandations pour le développement et la gestion de vocabulaires structurés utilisables pour la recherche d'information.

Cette partie 1 - qui fait l'objet du présent Livre Blanc - a été officiellement publiée en août 2011.

La Partie 2 - *Interopérabilité avec les autres vocabulaires*⁴ propose des solutions pour l'utilisation simultanée de plusieurs langages contrôlés pour accéder à de très larges collections de ressources distribuées sur de multiples réseaux, rapidement, efficacement, et dans la langue choisie par le chercheur d'information.

Le projet de norme de cette partie 2 devrait être publiée début 2013.

Le processus de mise à jour des normes ISO prévoit une révision tous les 5 ans.

⁴ Sur le site de l'ISO :
http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658&commid=48836

2 - Présentation générale de la Partie 1 de la norme ISO 25964

La Partie 1 de la norme ISO 25964 publiée en août 2011 est composée de

- **18 chapitres** appelés Clauses⁵ dans le document normatif,
- 2 annexes informatives contenant des exemples d'extraits de thésaurus ainsi que le schéma XML,
- une bibliographie de 69 références et un index⁶.

Ces 18 chapitres peuvent être scindés en deux grands volets que nous appellerons dans ce Livre Blanc : « **sémantique métier** » et « **sémantique technique** ».

- La **sémantique métier** renvoie à la construction du thésaurus en tant qu'outil sémantique et terminologique d'un système de recherche d'information : pourquoi (*contrôler le langage humain pour optimiser la recherche d'information*), quoi (*choisir et caractériser des concepts, des termes et leurs relations*), comment (*présenter et gérer ces réseaux sémantiques et terminologiques qui composent un thésaurus*).
- La **sémantique technique** renvoie quant à elle à la modélisation des systèmes d'information (modèle de données et diagramme UML, schéma XML et documentation technique).

Remarques : Dans d'autres secteurs comme la terminologie TC37, ce découpage correspond à deux séries de normes distinctes, regroupées ici dans un seul document normatif.

2.1. Termes, définitions et symboles

ISO 25964-1, Clause 2 – Terms and Definitions ; p.1-12

ISO 25964-1, Clause 3 - Symbols, abbreviated terms and other conventions, p.12-15

69 termes utilisés dans la norme sont définis et intégrés dans la Clause 2, chapitre traditionnel des normes intitulé « Termes et Définitions ». Le lexique bilingue proposé en Annexe 1 de ce Livre Blanc reprend l'ensemble de ce vocabulaire proposé uniquement en anglais dans la norme.

Un important travail a été réalisé au niveau national et international pour sélectionner les concepts à intégrer à cette liste, s'assurer que leur signification était bien partagée, choisir le ou les termes les représentant, et rédiger des définitions qui tiennent compte du contenu de la norme et d'un ensemble étendu de pratiques et cultures.

Le groupe francophone a travaillé sur la traduction française des concepts clés - ainsi que sur les exemples de la norme - dès le début du projet et tout au long de la rédaction de cette première partie, ceci en vue de faciliter la production de ce lexique et de diminuer les difficultés de traduction ultérieure.

Concernant les symboles et abréviations utilisés pour représenter des éléments descriptifs (i.e. Note explicative = NE), des relations (i.e. Terme spécifique d'instance = TSI) ou certains codes comme les catégories de groupes de concepts, la norme propose un tableau complet des symboles utiles en langue anglaise. Ce tableau est complété pour certains de ces symboles par d'autres versions linguistiques dont le français (voir Annexe 2). Notons que le groupe ISO en charge de ces travaux n'a pas souhaité procéder à trop de modifications par rapport aux pratiques existantes et a conservé, par

⁵ « Clause » : Disposition particulière d'un acte (<http://www.cnrtl.fr/definition/clause>). Nous conserverons dans ce Livre Blanc la dénomination de Clause pour tous les renvois à la norme ISO 25964 ; le terme chapitre sera réservé à la structure du Livre Blanc.

⁶ Complétée par des ressources en ligne à cette adresse – <http://www.niso.org/schemas/iso25964/>

exemple pour les relations entre concepts et termes, les codes déjà connus tel que le code « TS » (Terme spécifique) pour désigner la relation hiérarchique de subordination entre deux concepts (et de fait pour désigner la relation entre les deux termes les représentant).

2.2. Sémantique métier : construire et maintenir un vocabulaire structuré

La structure et le contenu de la norme **en termes de sémantique métier** restent très proches du contenu des normes ISO 2788:1986 et 5964:1995, et assez proches des normes françaises. Ainsi, aucune modification majeure n'est à signaler dans les clauses touchant aux concepts représentés et à leurs traitements, aux termes (sélection, forme, méthodes de désambiguïsation des termes homographes) ou encore à la structure relationnelle conceptuelle et terminologique.

Par contre, des précisions et des enrichissements ont été apportés, et plusieurs parties nouvelles sont proposées. Le document a également été enrichi d'exemples. Quant aux nouvelles clauses, celles-ci portent sur : les équivalences conceptuelles dans les thésaurus multilingues (Clause 9), l'analyse par facettes (Clause 11), la présentation et l'affichage des thésaurus (Clause 12), la gestion des thésaurus (Clause 13) et des recommandations concernant les logiciels pour la gestion des thésaurus (Clause 14).

2.3. Sémantique technique : modèle de données, schéma XML et documentation technique

La plus grande nouveauté vient de l'intégration de plusieurs chapitres ou supports techniques en vue de faciliter et stimuler l'utilisation et le développement des applications informatiques et en réseau.

Cette matière nouvelle se présente sous la forme de trois parties spécifiques et d'un certain nombre d'outils techniques à destination des concepteurs et développeurs d'applications et de logiciels :

- Trois **nouvelles Clauses** (normatives) portant sur l'intégration de thésaurus dans les applications (Clause 16), les formats d'échange (Clause 17) et les protocoles (Clause 18) ont été ajoutées. Celles-ci rendent compte de l'importance de la place prise par les thésaurus dans les systèmes de recherche d'information.
- Un **modèle de données** (normatif), présenté sous la forme d'un diagramme UML et de tableaux détaillés pour chacune des classes du modèle, fait l'objet de la Clause 15. Suivant les règles des diagrammes de classe UML⁷, le modèle de données s'exprime à travers un ensemble de classes, chacune représentant un des composants-clés du thésaurus : *Thesaurus*, *ThesaurusConcept*, *ThesaurusTerm*, *ConceptGroup*, *ThesaurusArray*, *Note*. Chaque classe est enrichie d'attributs et de relations avec d'autres classes. Les éléments liés au modèle de données sont répartis dans chaque chapitre de ce Livre Blanc dans une section dédiée (**> Modèle de données ISO 25964-1**). (Pour des repères sur UML, voir l'Annexe 3 de ce Livre Blanc).
- Un **schéma XML** (informatif) – le choix s'est porté sur un schéma XML à plat (plutôt qu'un schéma XML imbriqué). Le schéma peut être étendu en fonction de l'intérêt de la communauté concernée suivant deux axes : l'ajouts de contraintes d'intégrité référentielle dans le cas de l'échange d'un thésaurus complet et l'ajout d'attributs personnalisés pour pouvoir effectuer des échanges partiels.
- La **documentation** (informatif) du schéma au format HTML et un **document test** sont associés au schéma XML.

Les outils complémentaires au document normatif sont disponibles sur le site du NISO⁸.

Cette documentation à caractère technique devrait faciliter le déploiement d'applications conformes aux directives de la norme.

⁷ Diagramme UML. Voir Annexe 3 et http://fr.wikipedia.org/wiki/Diagramme_de_classes

⁸ Rubrique « ISO25964 » sur le site du NISO - <http://www.niso.org/schemas/iso25964/schema-intro>

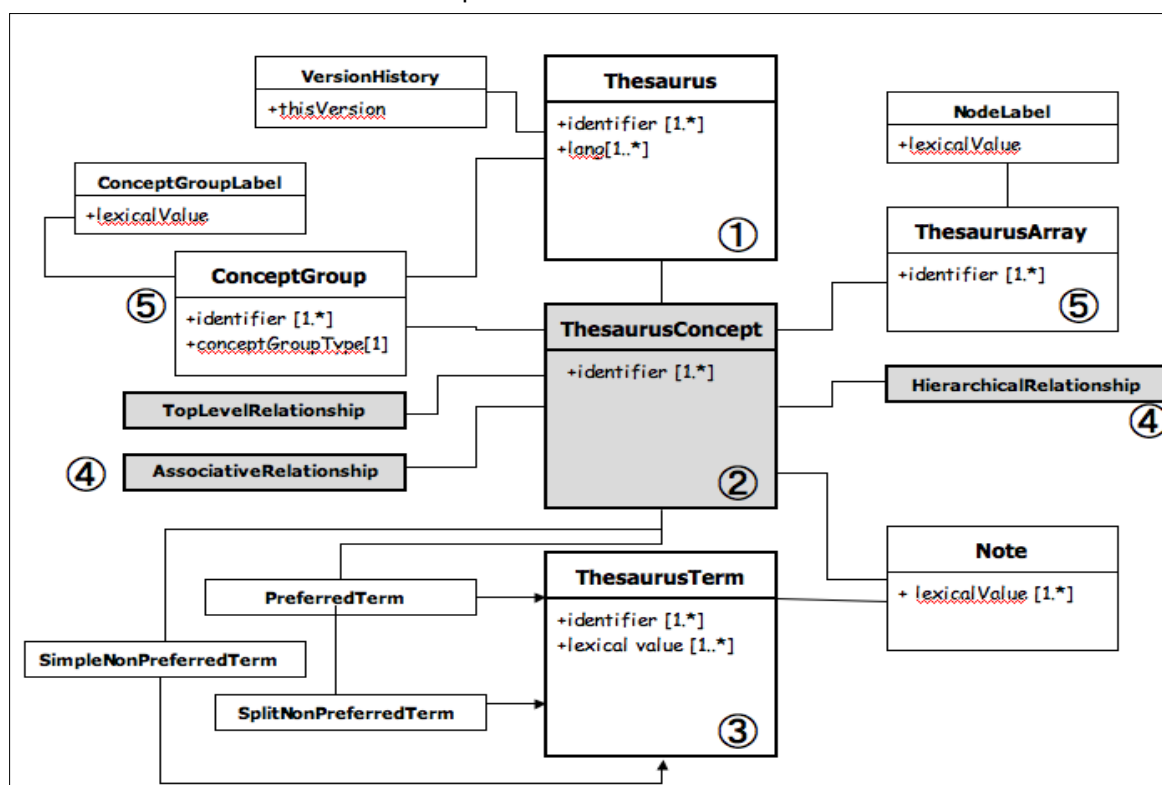
3 - Composants d'un thésaurus

La conception et la construction d'un thésaurus en tant que structure organisée de concepts et de termes sont exposées dans huit (8) clauses de la norme ISO 25964-1.

Le graphique présenté dans la **Figure 1** est un schéma très simplifié du modèle de données proposé dans la norme⁹. Il offre une vision globale des **composants principaux d'un thésaurus** qui font chacun l'objet d'une section particulière de ce Chapitre 3.

- Le Thésaurus (Chapitre 3.1.)
- Les Concepts (Chapitre 3.2.)
- Les Termes (Chapitre 3.3.)
- Les Relations entre concepts et entre concepts et termes (Chapitre 3.4.)
- Les Regroupements de concepts thématiques ou par facettes (Chapitre 3.5)
- Nous ajouterons à ces composants une section sur les aspects multiculturels et linguistiques présents dans un thésaurus (Chapitre 3.6)

FIGURE 1 – Schéma simplifié du modèle de données ISO 25964-1:2011



⁹ Lien vers le Schéma XML (<http://www.niso.org/schemas/iso25964/schema-intro/>) et le modèle de données (http://www.niso.org/schemas/iso25964/Model_2011-06-02.jpg).

3.1.Thésaurus

ISO 25964-1, Table 4 — Attributes and associations of Thesaurus, p.110

La norme ISO 25964-1 fournit des directives et recommandations pour traiter le thésaurus comme un objet à part entière. Le thésaurus dispose ainsi dans la nouvelle norme d'un ensemble d'attributs, pour la plupart les éléments du format normalisé de description du Dublin Core, permettant de le décrire, de le gérer et de le positionner dans un ensemble d'autres thésaurus. D'autres éléments (non obligatoires) permettent de tracer son évolution dans le temps.

Exemple (non présent dans la norme ISO 25964-1)

Titre	Thésaurus « Santé publique »
Identifiant	URL:http://asp.bdsp.ehesp.fr/Thesaurus/
Contributeur	Réseau BDSP
Editeur	École des hautes études en santé publique (EHESP) Banque de données en santé publique (BDSP)
Date	2007
Date (création)	1992
Date (modification)	2007
Version	v4
Langue	fra
Type	thésaurus
Périmètre	614-Santé publique (Dewey)
Sujet	santé ; politique publique ; système de soin ; pathologie ; épidémiologie ;
Description	Thésaurus exploité dans le cadre de la banque de données bibliographique BDSP développée par le réseau BDSP (France). La version 2007 comporte 12 825 entrées terminologiques organisés en 57 micro-thésaurus.
Droits	Droit d'usage limité à une utilisation strictement non commerciale
Format	pdf, txt, html

> Modèle de données ISO 25964-1

L'objet « Thésaurus » est lui-même considéré comme une classe dans le modèle de données.

La classe « *Thesaurus* » reprend l'ensemble des 15 attributs du Dublin Core (version 1.1. de 2006) : Identifiant (obligatoire) ; Titre ; Langue (obligatoire) ; Date (permet de gérer les mises à jour) ; Droits ; Domaine d'application...

Cette liste est ouverte à la personnalisation : il est tout à fait possible d'ajouter un attribut qui ne soit pas listé dans la norme, par exemple une relation vers un autre vocabulaire auquel ce thésaurus serait relié, ou encore la liste des systèmes ou bases de ressources dans lesquels ce thésaurus est exploité.

La classe *Thesaurus* est elle-même associée à plusieurs autres classes :

- la classe *Historique des versions* (« *VersionHistory* »), une nouvelle version correspondant à un changement dans la structure conceptuelle du thésaurus ;
- la classe des Concepts du thésaurus (« *ThesaurusConcept* ») – voir 3.2
- la classe de Groupes de concepts (« *ConceptGroup* ») – 3.5 1.
- la classe des Regroupements par facettes (« *ThesaurusArray* ») – voir 3.5.2.

Ces trois dernières classes (concepts, groupes de concepts et facettes) sont des parties du Thesaurus (relation isPartOf/contains).

Le modèle de données impose a minima : un identifiant, la/les langues couvertes par le thésaurus, ainsi que des concepts de la classe *ThesaurusConcept*.

Ces éléments (identifiant, langue, concepts) correspondent à la représentation minimale d'un thésaurus.

3.2. Concepts

Une des évolutions fondamentales de la norme ISO 25964 porte sur la distinction formalisée entre d'une part les concepts et d'autre part les termes les représentant. Concepts et termes font l'objet de recommandations ciblées décrites dans des Clauses particulières (Clause 5 – Concepts et Clause 6 – Terms) et d'un formalisme particulier dans le modèle de données.

3.2.1. Le thésaurus : une sélection de concepts

ISO 25964-1, Clause 5 — Concepts - p. 18-21

ISO 25964-1, Table 6 — Attributes and associations of Thesaurus Concept, p.112-113

ISO 25964, Table 9 — Attributes and associations of Note, p.115

Le thésaurus regroupe une sélection de concepts d'un domaine (de spécialité ou d'activités) permettant de formuler les sujets représentés dans des ressources ou des questions, relevant de ce domaine de spécialité.

Les caractéristiques principales des **concepts** reprennent celles énoncées dans les versions précédentes de la norme sur les thésaurus :

- Les catégories générales des concepts présents dans un thésaurus peuvent être : des objets et leurs caractéristiques physiques, des matériaux, activités et processus, des événements et faits, des caractéristiques (propriétés) de personnes, objets, matériaux ou actions, des disciplines ou domaines de spécialité (*subject field*), des unités de mesure, types de personnes et organismes ainsi que des entités individuelles telles que des noms propres (lieux, objets spécifiques, nom de personnes ou d'organismes) ;
- La « note explicative »¹⁰ définit ou clarifie le périmètre sémantique du concept tel que défini dans le vocabulaire contrôlé. Elle se distingue donc d'une définition. D'autres types de note paramétrables sont proposées (définition, note éditoriale, ...).

La norme insiste sur la nécessité de prendre en compte les particularités linguistiques ou culturelles de l'environnement d'utilisation du thésaurus (voir chapitre 3.6). Toutefois, l'organisation conceptuelle du thésaurus prise en charge dans la Partie 1 de la norme ISO 25964-1 repose sur une structure *symétrique* quelles que soient les langues ou les cultures. Dans le cas où le thésaurus admet des relations non-symétriques entre langues ou cultures, la partie 1 de la norme renvoie à la partie 2¹¹ pour un traitement par le biais de méthodes d'alignement entre vocabulaires (Clause 10.2.6.).

Les concepts au coeur du thésaurus sont reliés à :

- *d'autres concepts* par des *relations hiérarchiques et associatives*, formant ainsi un réseau de concepts (voir chapitre 3.4.)
- *des termes* : un seul terme de *type préférentiel* et autant de *termes non préférentiels* que souhaités (voir chapitre 3.3) ;
- *d'autres concepts par des relations d'une autre nature* que les relations hiérarchiques ou associatives : les *relations thématiques ou par facettes* (voir chapitre 3.5.).

> **Modèle de données ISO 25964-1**

Le modèle de données propose une classe spécifique pour représenter les concepts, distincte de la

¹⁰ Le terme « note explicative » était utilisé dans la norme ISO 2788 en français pour ce qui se nommait dans la norme Afnor NF47-100, « note d'application » (NA). « Note explicative » (NE) est le terme conservé dans cette norme de 2011. Mais d'autres désignations sont toutefois possibles : note d'emploi, note d'application.

¹¹ La publication de la partie 2 de la norme est prévue pour fin 2012. Une enquête publique a été lancée entre janvier et avril 2012.

représentation du réseau des termes : la classe « *ThesaurusConcept* » caractérisée par une sélection de propriétés - identifiant, date de création et modification, statut (candidat, approuvé, ...), notation...

Les seuls éléments imposés par la norme sont : un identifiant unique par concept et au minimum un terme préférentiel par langue. Ces éléments correspondent à la représentation minimale d'un thésaurus comme réseau de concepts représentés par des termes (voir Chapitre 3.3. Termes de ce Livre Blanc).

3.2.2. Traitement des concepts complexes

ISO 25964-1, Clause 7 – Concepts complexes – p.37-44

ISO 25964-1, Clause 8.5. Representation of complex concepts by a combination of terms – p.49-50

Un thésaurus peut intégrer des concepts dont le niveau de complexité peut varier. On distingue ainsi les concepts dits « simples » (par exemple « hôpital » ou « logiciel ») des concepts dits « complexes » dans le sens où ils articulent en leur sein des concepts élémentaires : « hôpital pour enfants », « logiciel de gestion de thésaurus » ou « biodégradabilité ».

Une des questions récurrentes lors de la conception et la maintenance des thésaurus porte sur le degré de précision ou de généralité à déployer. Par exemple, le thésaurus devra-t-il intégrer tous les types de logiciels sujets des ressources documentaires traitées ou ceux faisant l'objet d'une question d'utilisateur, ou bien conserver uniquement les concepts élémentaires « logiciel », « gestion » et « thésaurus », ou encore « logiciel de gestion » et « thésaurus » - le concept « logiciel de gestion de thésaurus » étant alors formé au moment de l'indexation ou de la recherche par la combinaison des concepts élémentaires.

La Clause 7 entièrement dédiée à cette question, introduit la problématique et fournit des directives et des recommandations pour l'intégration ou le rejet d'un concept complexe, établies selon certains critères précis (Clause 7.3.2) ou dans des cas d'utilisation particuliers (Clause 7.3.3. et 7.3.4.).

> *Modèle de données ISO 25964-1*

Le mécanisme du renvoi d'un concept vers plusieurs autres concepts vise à assister un utilisateur posant sa question ou cherchant à indexer un concept, ce dernier n'ayant pas été sélectionné comme concept du thésaurus. Ce principe de renvoi existait dans les anciennes normes mais n'était visible que dans les impressions papier des thésaurus utilisés alors par les indexeurs, ce mécanisme n'étant à peu près jamais implanté dans les applications de recherche documentaire.

La solution proposée dans la nouvelle norme reste dans la continuité des anciennes normes. Elle repose sur la création d'une relation d'équivalence combinée au niveau des termes (voir chapitre 3.3), et non des concepts. Ainsi « logiciel de gestion de thésaurus » n'est pas un concept du Thésaurus, mais un type particulier de terme : un « terme non préférentiel scindé » (*SplitNonPreferredTerm*) relié à deux entrées du thésaurus par une relation d'équivalence particulière appelée « équivalence composée » (*CompoundEquivalence*).

Exemples - Thésaurus d'éthique des sciences de la vie¹²

Logiciel de gestion de thésaurus = EM Logiciel de gestion + Thésaurus

ou : logiciel de gestion de thésaurus : EM+ Logiciel de gestion, Thésaurus

Dans le thésaurus en ligne « Ethique des Sciences de la vie »

cancer des reins **USE** rein **AND** cancer
empreinte écologique **USE** développement durable **AND** méthode de recherche
greffe de rein **USE** transplantation d'organe **AND** rein
transplantation de rein **USE** rein **AND** transplantation d'organe

¹² <http://www2.drze.de/BELIT/thesaurus/suchergebnis.html?la=fr&begriff=rein>

3.3. Termes

ISO 25964-1, Clause 6 – Thesaurus terms – p.21-37

ISO 25964-1, Clause 8 The equivalence relationship, in a monolingual context, pp.44-50

ISO 25964-1, Table 8 — Attributes and associations of ThesaurusTerm, p.114

ISO 25964, Table 9 — Attributes and associations of Note, p.115

Les termes comme expressions linguistiques d'une langue, représentent les concepts et sont reliés à ces derniers par une relation appropriée. Les directives ou recommandations concernant les *termes*, *leur sélection*, *leur forme* ou *les méthodes de désambiguïsation* reprennent celles énoncées dans les versions précédentes de la norme :

- Un terme peut être un mot ou un groupe de mots ;
- Des règles de désambiguïsation (qualificateur pour les homonymes, ajout de notes explicatives,...), ainsi que des règles pour choisir la forme des termes doivent être appliquées de façon cohérente sur l'ensemble du thésaurus ;
- Un terme préférentiel par concept et par langue doit être sélectionné ;
- Une relation d'équivalence est établie entre des termes non préférentiels et un terme préférentiel.

Quelques nouveautés ou précisions ont été introduites pour tenir compte de la diversité des contextes de développement des thésaurus :

- Une définition ainsi qu'une note historique gardant la trace des changements peuvent être attribuées à un terme, qu'il soit préférentiel ou non. D'autres types de notes paramétrables (éditoriales, ...) sont proposés (Clause 15.2.14)
- La norme insiste sur la nécessité de prendre en compte les particularités linguistiques ou culturelles, en acceptant par exemple le singulier ou le pluriel, les adjectifs et les verbes,... (Clause 6.5).
- Les noms propres sont acceptés et bénéficient ainsi des mêmes possibilités de mise en relation que tout autre concept. Néanmoins s'ils sont en grand nombre ou contrôlés dans leur forme par une liste d'autorité externe ou un ensemble de règles, la norme laisse le choix de les conserver ou de les retirer du thésaurus.
- Le traitement des concepts complexes est réalisé par le biais des termes les désignant (*voir chapitre 2.2.2. Traitement des concepts complexes*).

> **Modèle de données ISO 25964-1**

Le reproche fait traditionnellement aux thésaurus de ne pas être « conceptuels » mais « terminologiques » porte sur l'absence de distinction formelle entre concepts et termes : les concepts n'étaient alors manipulables qu'à travers le terme préférentiel. Mais pour faciliter l'interopérabilité entre les thésaurus¹³ et prendre en charge les évolutions dans le temps, cette distinction concept/terme devient indispensable dans le contexte des systèmes de recherche d'information en réseau.

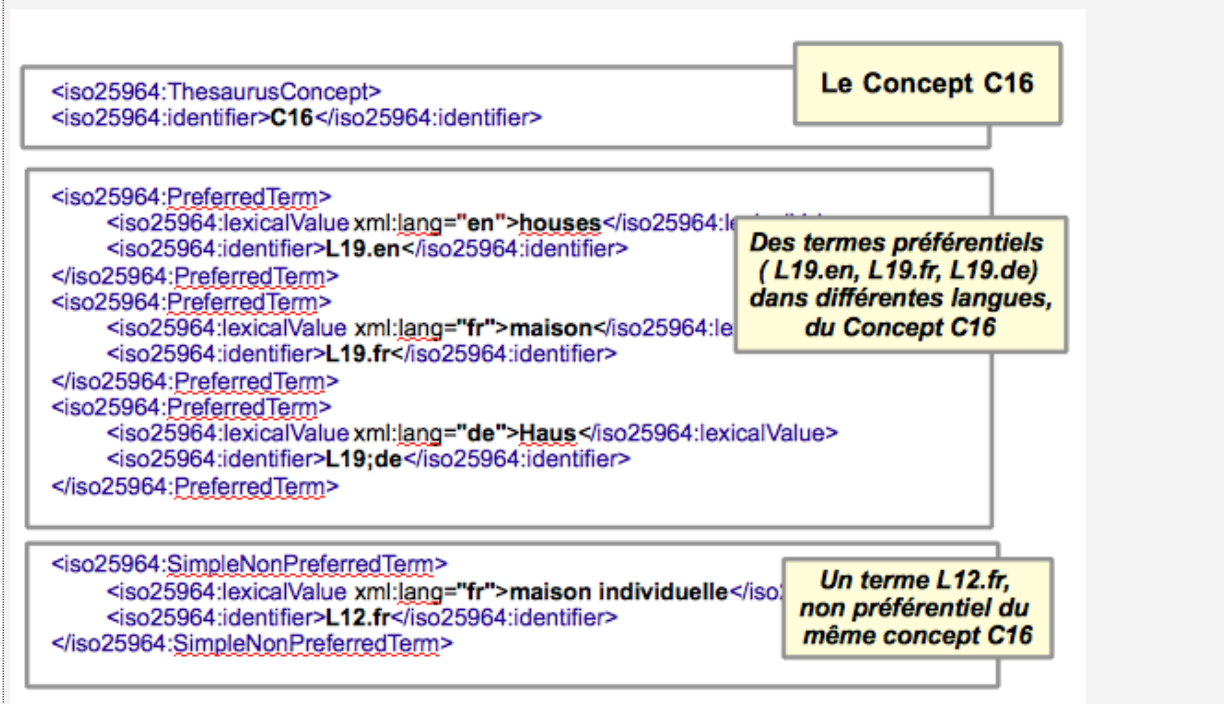
Dans le modèle de données, la nouvelle norme distingue formellement les termes des concepts au moyen d'une classe dédiée « *ThesaurusTerm* » et de deux relations entre termes et concepts : « *isPreferredLabel* » / « *hasPreferredLabel* » et « *isNonPreferredLabel* » / « *hasNonPreferredLabel* ».

- Les termes du thésaurus sont donc gérés au sein d'une classe spécifique « *ThesaurusTerm* », enrichie par un ensemble d'attributs permettant d'assurer la gestion dans le temps du lexique. Parmi ces attributs, la valeur lexicale et l'identifiant du terme sont les seuls qui soient obligatoires, les autres, comme la langue, les dates de création et modification, la source ou le statut (candidat, approuvé,...) sont recommandés, mais restent facultatifs.

¹³ Chaque dispositif peut avoir opté pour des termes préférentiels différents pour un même concept, par exemple.

- Les termes préférentiels et non préférentiels sont reliés *aux concepts* par une relation spécifique « *isPreferredLabelFor* » ou « *isNonPreferredLabelFor* ». Seule la relation entre un concept et un terme préférentiel est obligatoire.
- Les termes préférentiels et non préférentiels sont reliés entre eux par une relation d'équivalence établie entre les deux sous-classes « *PreferredTerm* » et « *NonPreferredTerms* ». Les abréviations utilisées en français pour représenter ces relations restent identiques : EM (employer) et EP (employé pour) (Clause 3 - Table 2).
- La nature de la relation entre un terme préférentiel et un terme non préférentiel (sigle / développé, nom d'emprunt / néologisme, synonyme, quasi-synonyme, antonyme, etc.) pourrait être précisé en utilisant l'attribut « *role* » proposé par le modèle de donnée de la norme.
- Comme dans la recommandation SKOS (voir paragraphe sur SKOS dans le Chapitre 4.3), il est possible d'ajouter un attribut au terme non préférentiel pour indiquer que celui-ci est caché (hidden), ce qui permet d'intégrer des écritures fautives utilisées pour la recherche mais non visibles par les utilisateurs.
- Il est également possible d'enrichir le thésaurus de différents types de notes rattachées aux termes. Elles sont distinctes des notes attachées aux concepts. Trois classes (*HistoryNote*, *Definition*, *EditorialNote*) facultatives sont ainsi associées à la classe *ThesaurusTerm*.

Exemples : Extrait du schéma XML de la norme ISO 25964-1.



3.4. Relations

ISO 25964-1, Clause 10 — Relationships between concepts, pp.57-67

ISO 25964-1, Clause 14.3 — Relationships between terms and between concepts

ISO 25964-1, Table 6 — Attributes and associations of ThesaurusConcept, pp.112-113

L'ensemble des types de relations conceptuelles ou terminologiques proposées dans cette norme, existait dans les versions précédentes des normes : relations d'équivalence entre termes (Clause 8) relations hiérarchiques (Clause 10.2) et relations associatives, relations entre concepts (Clause 10.3). La norme française Afnor n'était toutefois pas aussi précise en ce qui concerne le typage des relations hiérarchiques.

Les nouveautés de la norme ISO 25964 portent d'une part sur les possibilités de spécifier plus avant les types de relations et d'autre part sur le formalisme imposé par le modèle de données.

Les possibilités de spécifier et de formaliser les types de relations permettent de prendre en compte des caractéristiques propres à certains domaines, activités ou catégories d'utilisateurs. Ainsi :

- Une relation « appartient à la discipline » / « est un objet de la discipline » permet de typer plus finement une relation d'association pouvant exister entre Oiseau et Ornithologie ;
- Une relation hiérarchique d'instance entre les noms d'artistes, molécules, produits, ... et leurs catégories d'appartenance permet de spécifier des relations associatives ou polyhiérarchiques ;
- Une relation d'appartenance à une facette « origine animale » peut être transformée en une relation hiérarchique ou associative d'un type particulier (par exemple « origine du produit »),...

Ces quelques exemples montrent que la spécialisation ou typage des relations a un impact sur la construction d'un thésaurus et sur les types de relations existants. Il convient donc de prendre du recul et d'étudier l'ensemble des relations formant le réseau conceptuel du thésaurus sur un plan plus global. C'est pourquoi nous faisons ici une présentation d'ensemble de ce réseau de concepts et des relations qui le structurent, même si la norme aborde les spécificités de chacune d'entre elles dans une Clause particulière.

Nous nous sommes limités dans ce chapitre 4.4. aux relations établies entre les concepts du thésaurus, et abordons dans un autre chapitre les relations entre ces concepts et leur regroupement soit par domaines (*ConceptGroups*) soit par facettes (voir chapitre 4.5).

Relation hiérarchique (Clause 10.2)

La relation hiérarchique générale TG (générique) /TS (spécifique), présentée dans la Clause 10.2., peut être spécifiée suivant trois axes : Espèce / Genre, Tout / Partie, Instance¹⁴.

Par exemple :

« Oiseau TSG Moineau » et « Moineau TGG Oiseau » pour une relation hiérarchique générique ;

« Vaisseau sanguin TSP Veine » et « Veine TGP Vaisseau sanguin » pour une relation hiérarchique partitive ;

« Montagne TSI Alpes » et « Alpes TGI Montagne » pour une relation hiérarchique d'instance, associée à une relation d'équivalence entre Montagne EP Massif montagneux (ou l'inverse suivant les choix des concepteurs du thésaurus)

Comme dans les précédentes normes, il n'est pas obligatoire de préciser le type de relation, et la mention de la seule relation hiérarchique sans plus de précision est bien sûr autorisée.

Par exemple : « Moineau TG Oiseau » et « Oiseau TS Moineau ».

¹⁴ La relation d'instance ne figurait pas dans la norme Afnor 47-100, mais existait dans la norme ISO de 1986 ainsi que dans les normes plus récentes BS 8723:2008 et ANSI/NISO Z39.19:2005

Relations avec le concept de tête de la branche hiérarchique (TopConcept, Clause 10.2.4.)

L'affichage de la relation directe entre un concept et le concept situé en tête de la structure hiérarchique dans laquelle il se trouve peut être très utile dans le cas de thésaurus fortement hiérarchisés ou dans le cas de thésaurus autorisant la polyhiérarchie. Déduite généralement par le système de gestion du thésaurus, cette possibilité reste toutefois facultative.

Cette relation de tête de hiérarchie, dont le code est TT, se distingue formellement de la relation d'appartenance d'un concept à un groupe de concepts (voir chapitre 3.5.).

Exemples

Thésaurus de l'Unesco (p.130 de la norme)

Tension mentale TT Comportement (TT – concept de tête de hiérarchie du concept Tension mentale)

Tension mentale CS 4.10 Psychologie (CS¹⁵ - catégorie sémantique - appartenance à un domaine)

GEMET (p.133 de la norme)

emploi TT Travail (concept de tête de hiérarchie du concept emploi)

emploi CS SOCIETE (appartenance à un domaine)

Relation polyhiérarchique (Clause 10.2.5)

La relation polyhiérarchique permet de caractériser plus finement un concept et d'en proposer plusieurs clés d'accès en établissant des relations hiérarchiques avec plus d'un concept appartenant à des catégories différentes.

Exemple : une *caméra sous-marine* est à la fois une *caméra* à usage spécifique et un *matériel de plongée*
caméra sous-marine TG caméra à usage spécifique
caméra sous-marine TG matériel de plongée

Relation associative (Clause 10.3)

Les relations associatives couvrent, comme dans les normes précédentes, les relations autres que hiérarchiques génériques/spécifiques entre concepts. Elles forment un réseau sémantique complémentaire au réseau hiérarchique et enrichissent ainsi les possibilités de navigation entre concepts.

Par exemple « Ornithologie TA Oiseau » ou encore « Poison TA Toxicité ».

Différents cas de relations associatives comme les relations cause/effet ou opération/agent, sont exposés dans la norme (Clause 10.3.3.).

> Modèle de données ISO 25964-1

Le réseau conceptuel du thésaurus est modélisé autour de trois classes créées pour chacun des types de relations entre concepts : « *HierarchicalRelationship* », « *AssociativeRelationship* », « *Top LevelRelationship* ».

Les classes « *HierarchicalRelationship* » et « *AssociativeRelationship* » possèdent chacune un attribut « **role** » qui permet de paramétrer des types de relations particulières. Mais si la norme traite de différents cas de relations, aucun vocabulaire autre qu'« instance » et « tout-partie » pour les relations hiérarchiques n'est proposé pour coder ces types de relations avec l'attribut « role ».

Exemple

Une relation peut être établie entre une action (Courtage) et l'agent de cette action (Courtier) agent dénommé souvent par un nom de métier.

Agriculture [TA « action-agent »] Agriculteur

Courtier [TA « action-agent »] Courtage

¹⁵ L'étiquette CS n'est pas normalisée. C'est la traduction de l'étiquette en anglais : SC – semantic category.

Dans le modèle de base, les relations d'association sont réciproques. Ainsi l'établissement d'une relation associative entre Oiseau et Ornithologie est symétrique et valable dans les deux sens. Il est recommandé de pouvoir contraindre et contrôler les applications en charge de la gestion d'un thésaurus comme précisé dans la Clause 14.3 (Relationships between terms and between concepts).

Toutefois pour certains types de relation plus spécialisée, telle la relation « Cause/Effet » ou une relation « historique », non symétrique, il peut être nécessaire de modifier cette contrainte en offrant une relation inverse non réciproque.

Le modèle de données de la norme autorise la polyhiérarchie, chaque concept pouvant être relié à plusieurs autres concepts par une relation hiérarchique, sans restriction.

Exemple

Orgue (instrument de musique) TG Instrument à vent
Orgue (instrument de musique) TG Instrument à clavier

crâne TG os
crâne TG tête

Ou

crâne TGG bones (*générique générique*)
crâne TGP tête (*générique partitif*)

3.5. Regroupement de concepts : thèmes et facettes

La structure conceptuelle de base d'un thésaurus est constituée par le réseau de concepts créé par les relations hiérarchiques et associatives entre concepts. Mais la norme permet également d'autres types de regroupements de concepts, non obligatoires mais qui enrichissent les accès et la navigation au sein de ce réservoir de concepts.

Deux types fonctionnellement distincts de regroupement de concepts sont identifiés dans la norme :

- Des regroupements à partir d'une classification - les « Groupes de Concepts » (voir chapitre 3.5.1).
- Des regroupements par Facettes (voir chapitre 3.5.2).

Les dispositifs sur les Intranets ou l'Internet profitent pleinement de ce type de système d'organisation des connaissances.

3.5.1. Groupe de concepts – regroupements par domaines ou thématiques

ISO 25964-1, Clause 12.2.5 — Classified display, p. 76-78

ISO 25964-1, Clause 15.2.18 — Concept groups, p.108

ISO 25964-1, Table 7 — Attributes and associations of ConceptGroup, pp.118

Parallèlement aux réseaux hiérarchiques et associatifs des concepts, les thésaurus autorisent le regroupement des concepts à partir d'une structure classificatoire prédéfinie¹⁶. Cette double organisation (conceptuelle et classificatoire) qui fait la spécificité de certains thésaurus, bien que facultative, est de plus en plus fréquente dans les dispositifs d'accès à l'information sur les réseaux Web ou Intranet.

- Les bases utilisées pour regrouper des concepts sont de nature variée : disciplines, domaines thématiques ou domaines d'activité métier... On peut également citer l'organisation chronologique ou géographique des concepts,.... Ces groupes de concepts sont parfois dénommés « catégories »

¹⁶ Afnor Z47-100, 5.7.1. Regroupement thématique, p. 14

ou « catégories de sujets », « thèmes » ou « thématiques », « domaines », « groupes », « microthésaurus » ou « champs sémantiques ».

- Cette structure classificatoire peut elle-même être hiérarchisée sur 2 ou plus rarement 3 niveaux.
- Chaque groupe de concepts peut posséder un nom ou label (« *ConceptGroupLabel* ») dans une langue donnée et une notation.
- L'appartenance d'un concept à un regroupement n'implique pas que tous les concepts ayant une relation hiérarchique entre eux dans le thésaurus, appartiennent au même « groupe de concepts » tel que défini ici.

Exemples

GEMET (p.136 de la norme)

Ce thésaurus propose deux systèmes distincts de regroupement de concepts :

- 42 thèmes¹⁷ ;

- 32 Groupes¹⁸ ou champs sémantiques, regroupés en 4 grands domaines auxquels sont rattachés toute la hiérarchie de concepts en partant des termes de tête.

On trouve par exemple dans le thésaurus Gemet, cette hiérarchie de concepts :

Agglomération

Zone industrielle

Zone militaire

« Agglomération » appartient au Thème « Bâtiments ». tandis que son terme spécifique « zone militaire » est sous le Thème « aspects militaires ». Ils appartiennent tous deux au même Groupe.

> **Modèle de données ISO 25964-1**

La modélisation de ces groupes de concepts repose sur 2 classes, « *ConceptGroup* » et « *ConceptGroupeLabel* », reliées entre elles. Chaque concept du thésaurus (« *ThesaurusConcept* ») est ainsi relié par une relation « *isMemberOfGroup* » / « *hasAsMember* » à la classe « *ConceptGroup* ».

Concernant les classes « *ConceptGroup* » et « *ConceptGroupeLabel* » :

- La classe « *ConceptGroupeLabel* » permet de représenter chaque groupe de concepts par une valeur lexicale.
- Les groupes de concepts de la classe « *ConceptGroup* » (et les sous-groupes dans le cas d'une classification hiérarchique) ont obligatoirement des identifiants. De plus le type de groupe de concepts (Disciplines, Microthésaurus, Thèmes, Domaines, ...) peut être spécifié grâce à l'attribut « *conceptGroupType* ».
- Ce mode d'organisation classificatoire des concepts reste facultatif.

Rappel : La tête de hiérarchie du concept (voir chapitre 4.4.3) est formellement distincte du/des groupes de concepts évoqués ici.

3.5.2. Regroupement par facettes

ISO 25964-1, Chapitre 11, p. 68-69

ISO 25964-1, Table 5 — Attributes and associations of ThesaurusArray, p.111

Une façon différente de regrouper des concepts d'un thésaurus repose sur l'**analyse par facettes** (*facet analysis*). La notion de « facettes » est beaucoup plus abstraite et donc moins directement accessible que la notion de thèmes ou domaines, mais elle offre des possibilités variées d'organisation et de filtrage dans les systèmes de recherche d'information.

L'analyse par facettes aboutit à la constitution de groupes de concepts appelés concepts frères ou de

¹⁷ GEMET - Listes thématiques - http://www.eionet.europa.eu/gemet/index_html?langcode=fr

¹⁸ GEMET – Groups - <http://www.eionet.europa.eu/gemet/groups?langcode=fr>

même rang (« sibling concepts») autour d'une facette particulière. Le choix des facettes peut varier en fonction du domaine, mais il est fréquent d'utiliser des catégories fondamentales telles que : objets, matériaux, agents, actions, lieux, temps, etc. Plus le thésaurus est spécialisé, plus ces facettes seront elles-mêmes spécialisées. Ainsi de la facette « par origine animale » dans un thésaurus agricole, ou de la structuration d'une liste de noms de colorant en fonction de leur nature artificielle ou naturelle.

Le relais virtuel (*node label*) est un terme choisi pour désigner cette facette : « acteur », « produit », « lait par source animale » ou « lait par type de traitement » dans l'exemple suivant.

Exemple – regroupement par facettes

```

.....industries agricoles
.....(acteur)
.....cultivateur-exploitant
.....fermier
.....berger
.....(produit)
.....produit cérééalier
.....produit laitier
.....beurre
.....crème
.....fromage
.....lait
.....(lait par matière grasse)
.....lait entier
.....lait demi-écrémé
.....lait écrémé
.....(lait par source animale)
.....lait de bufflonne
.....lait de vache
.....lait de chèvre
.....lait de brebis
.....(lait par type de traitement)
.....lait condensé
.....lait évaporé
.....lait homogénéisé
.....lait pasteurisé
.....lait stérilisé
.....industries mécaniques
.....(acteur)
.....ingénieur
.....(produit)
.....roue
.....etc

```

Cette organisation des concepts par facettes est indépendante de leur organisation hiérarchique.

Il est important de noter la configuration des relations hiérarchiques des concepts autour des relais virtuels. Lorsque le relais virtuel représente une caractéristique de division du concept, tous les termes de cette série sont alors considérés comme des termes spécifiques du terme supérieur. Ainsi, dans la figure ci-dessus les concepts « lait entier », « lait de bufflonne » et « lait stérilisé » sont tous des concepts spécifiques de « lait », provenant de différents critères de subdivision respectivement « matière grasse », « origine animale » et « type de traitement ». En revanche, lorsque le relais virtuel introduit une nouvelle facette (acteur, produit, ...), les concepts frères ne sont généralement pas des concepts spécifiques du concept supérieur. Ainsi « cultivateur-exploitant » (facette « acteur ») et « produits céréaliers » (facette produit) ne sont pas des termes spécifiques d'« Industries agricoles ».

Au sein d'une séquence, les concepts du thésaurus peuvent être ordonnancés suivant l'ordre alphabétique des termes préférentiels ou bien de façon systématique.

> **Modèle de données ISO 25964-1**

Dans les applications informatiques, l'analyse par facettes peut être prise en compte de différentes manières.

Le formalisme adopté dans le modèle de données de la norme est facultatif et suit principalement les règles fonctionnelles énoncées en 1986 en visant essentiellement une fonction de présentation (visuelle) d'un thésaurus. Ainsi les relais virtuels ne sont ni des concepts (« ThesaurusConcept ») ni des termes (« ThesaurusTerm ») du thésaurus, et ne sont donc pas en l'état, exploitables en recherche. Ils ne sont utiles qu'à la présentation du thésaurus, ce qui en limite la portée¹⁹.

L'organisation par facettes est représentée principalement par :

- Deux classes spécifiques : « *ThesaurusArray* » (série de concepts) et « *NodeLabel* » (valeur lexicale du relais virtuel) ;
- Une relation « *isMemberOfArray* » (concepts membres de la série) entre la classe « *ThesaurusConcep* » et la classe « *ThesaurusArray* ».

Les facettes peuvent éventuellement être décomposées en sous-facettes - grâce à la relation « *hasSuperOrdinateArray* », jusqu'au niveau de précision requis. Par exemple une facette « Action » peut être subdivisée en différents types d'action tels que « réparation » ou « détérioration » ; une facette « Acteur » peut être subdivisée en « Acteur par métier » ou « Acteur par rôle ».

L'ordonnancement des concepts sous une facette par ordre alphabétique ou de façon systématique s'appuie sur l'attribut « *ordered* »..

¹⁹ Il serait possible d'utiliser l'attribut « *role* » et un attribut « *relais-virtuel* » pour des concepts particuliers pour formaliser ce type de regroupement de concepts.

3.6. Multiculturalisme et multilinguisme

ISO 25964, Clause 9, Équivalence entre langues naturelles, p. 50-57

La norme ISO 25964 sur les thésaurus, intègre les caractéristiques des thésaurus multilingues considérant ceux-ci comme des thésaurus dont une des caractéristiques est de représenter des concepts par des termes issus de différentes langues naturelles.

Le traitement du multilinguisme est abordé dans les différents chapitres concernés par cette problématique sous la forme d'ajustements, de précisions et d'exemples s'ajoutant aux recommandations générales. Ainsi le choix des relations hiérarchiques (Clause 10.2.6) ou associatives (Clause 10.3.4), la présentation du thésaurus (Clause 12) ou la problématique de l'équivalence entre différentes langues naturelles (Clause 9) permettent d'intégrer les particularités des thésaurus multilingues au contenu général de la norme sur les thésaurus, au lieu de produire une norme spécifique.

Comme dans la norme internationale de 1985 sur les thésaurus multilingues, la question de l'équivalence entre termes de langues naturelles différentes représentant un même concept fait l'objet de diverses recommandations. La norme propose quatre degrés d'équivalence possibles entre termes de langues naturelles différentes.

Exemple

Équivalence exacte	fr: physique / en: physics / es: física / de: Physik
Équivalence quasi-exacte ou proche	fr: jeu didactique / en: learning games
Équivalence partielle	
ou équivalence générique/spécifique	fr: sécurité / de: Sicherheit / en: safety, security
Non équivalence	

La deuxième partie de la Clause 9 de la norme propose des solutions aux problèmes d'équivalence partielle ou proche : transformer le terme préférentiel du langage source, modifier le langage source, importer le terme d'une langue source comme terme d'emprunt, créer un terme forgé équivalent...

> **Modèle de données ISO 25964-1**

Chaque version linguistique d'un thésaurus multilingue partage le même réseau conceptuel et la langue devient, dans ce modèle de donnée, un attribut de la classe « ThesaurusTerm ».

Les langues du thésaurus sont déclarées dans la classe « Thésaurus ».

Le degré d'équivalence entre termes dans différentes langues naturelles, s'il est expliqué (Clause 9.2.) pour la conception d'un thésaurus, n'est toutefois pas modélisé dans le modèle de données.

Le nombre de synonymes ou de quasi-synonymes d'un concept varie généralement d'une langue à l'autre, et il n'est pas nécessaire dans un thésaurus multilingue à des fins de recherche documentaire, de présenter les équivalences inter-langues pour des termes non préférentiels. Cette relation entre termes n'est donc pas prise en charge par la norme ISO 25964-1. Toutefois le modèle de données, en formalisant les termes au sein d'une classe spécifique (*ThesaurusTerm*) et en les caractérisant par les attributs Langue (*lang*) et par la relation « *isNonPreferredLabelFor* » permettra l'exploitation de cette association inter-langue entre des termes non préférentiels.

4 - Usages du thésaurus et de ses composants

Nous venons de passer en revue dans le chapitre 3, les composants principaux du thésaurus tels qu'exposés dans la Partie 1 de la norme ISO 25964:2011.

Plusieurs clauses de la norme, d'ampleur différente, concernent quant à elles les usages et services possibles autour des thésaurus. Ces services sont exposés dans ce chapitre 4 du Livre Blanc:

- La présentation des thésaurus pour consultation (chapitre 4.1);
- Les formats et protocole d'échanges (chapitre 4.2) ;
- La méthode de conception et de maintenance des thésaurus (chapitre 4.3) ;
- Les logiciels de gestion de thésaurus (chapitre 4.4).

Il ne s'agit pas dans ces clauses de traiter la question des usages d'un thésaurus pour l'indexation ou à recherche, mais d'étudier les aspects techniques et fonctionnels pour présenter ou porter un thésaurus dans des applications.

4.1. Présentation des thésaurus pour consultation

ISO 25964, Clause 12, p. 70-88

La norme (Clause 12) ne propose pas de nouveauté en la matière. Différentes formes de présentation, qui se complètent les unes les autres, sont possibles. Des exemples de présentation de ces différents formats sont fournis. Qu'il soit présenté sur un écran ou sous une forme imprimée, le thésaurus peut être présenté sous les formes suivantes:

- Présentation unitaire d'un concept du thésaurus;
- Liste alphabétique des termes présents dans le thésaurus ;
- Classement hiérarchique des concepts représentés dans le thésaurus ;
- Présentation systématique des concepts représentés dans le thésaurus (domaines ou facettes) ;
- Table de correspondance pour les thésaurus multilingues.

Sans en préciser le contenu, la norme cite deux autres formes de présentation :

- Présentation sous forme de liste alphabétique permutée permettant de trouver ou de repérer dans une liste alphabétique les mots inclus dans les termes multi-mots;
- Présentation graphique du thésaurus.

Précisons qu'aucune de ces présentations n'est considérée comme étant la meilleure pour toutes les situations : leur pertinence varie en fonction des contextes d'application. Il est cependant primordial qu'au moins une des présentations listées ci-dessus soit mise à disposition des utilisateurs en complément de la présentation alphabétique.

Présentation unitaire d'un concept du thésaurus

Le concept est représenté par le terme préférentiel et son environnement sémantique : les termes non préférentiels et, de manière facultative, les relations qu'il entretient avec d'autres concepts, notes etc. La norme recommande un ordre de présentation de ces attributs.

Liste alphabétique des termes présents dans le thésaurus

Ce type de liste permet aux utilisateurs d'accéder rapidement à l'ensemble des termes, préférentiels, non préférentiels et non préférentiels scindés, depuis le mot utilisé par l'utilisateur. Cette liste peut également être utilisée comme un index.

Classement hiérarchique des concepts représentés dans un thésaurus

Ce classement est basé sur la relation hiérarchique TG/TS entre les concepts. Il est recommandé d'utiliser cette présentation en complément de la liste alphabétique des termes du thésaurus.

Présentation systématique des concepts représentés dans un thésaurus

Cette présentation se concentre sur les relations entre les concepts et leur domaine d'appartenance.

Deux choix de présentation sont possibles :

- les termes du thésaurus sont ordonnés par groupes de concepts (domaines) (voir chapitre 4.5.1)
- les termes du thésaurus sont classés par facette à l'intérieur d'un domaine (voir chapitre 4.5.2)

Il est recommandé d'utiliser cette présentation en complément de la liste alphabétique des termes du thésaurus. Dans la pratique, cette approche peut-être combinée avec le classement hiérarchique.

Liste permutée

Cette présentation permet de retrouver facilement un concept à partir de l'un des mots qui composent le terme préférentiel voire non préférentiel.

Présentation graphique d'un thésaurus

C'est la représentation visuelle des concepts via leurs termes préférentiels et de leurs relations. Citons la représentation par schéma fléché très utilisée dans les années 1970-80. Ce format de présentation d'un thésaurus peut bénéficier aujourd'hui des outils de cartographie.

Cas particuliers

Cas des relations polyhiérarchiques

Le terme représentant un concept apparaît ici à chacun de ses emplacements dans l'arborescence conceptuelle du thésaurus.

Dans le cas des thésaurus imprimés, on peut choisir un emplacement principal dans lequel sont logés les termes spécifiques, notes, synonymes etc. Dans le cas d'affichage sur écran, on doit pouvoir voir l'ensemble des relations quel que soit l'emplacement du terme polyhiérarchique.

Présentation et affichage des thésaurus multilingues.

Toutes les langues doivent être considérées d'une façon égale.

On mettra à disposition des utilisateurs une liste alphabétique des termes du thésaurus dans chaque langue. Des équivalents interlinguistiques doivent être disponibles pour les termes préférentiels de chaque version. La norme apporte des conseils quant aux présentations unitaires, alphabétiques, systématiques et hiérarchiques du thésaurus dans ce cas de figure.

Des recommandations sont faites pour gérer les problématiques techniques ou humaines de codage des caractères : les caractères doivent toujours être représentés dans une forme lisible pour les natifs de la langue et l'usage de la norme UCS (Universal Character Set) est fortement recommandé. La question des critères de tri fait l'objet de recommandations (Clause 12.4.3). Les systèmes de traitement et gestion de vocabulaires doivent être capables de produire des séquences alphabétiques selon les conventions de chacune des langues. A défaut, la notice explicative de chaque version doit clairement exposer l'ordre de classement retenu.

> Modèle de données ISO 25964-1

L'exploitation du modèle de données du thésaurus proposé dans la norme ISO 25964-1 permet la production de ces différentes modalités de présentation des thésaurus.

L'ajout d'attribut sur les entités (concepts, termes, ...) ou sur les relations permet également de

prendre en compte les besoins et les usages particuliers des utilisateurs finaux ou des gestionnaires.

4.2. Intégration des thésaurus à des applications

ISO 25964,-1 Clause 16 - , p. 115-118

Les situations où un thésaurus est exploité au sein d'un système de recherche ou de gestion d'information se sont diversifiées et rendent beaucoup plus fréquentes les activités d'échange de données entre plusieurs systèmes. Dans ce contexte, la Clause 16 de la norme présente :

- Les caractéristiques et fonctions qui permettent d'assurer l'interopérabilité entre thésaurus à travers les systèmes et les réseaux, ainsi que l'import/export - partiel ou total - de thésaurus, selon des formats ou des protocoles standards.
- Les caractéristiques et fonctions nécessaires pour exploiter un thésaurus dans une application d'indexation ou de recherche : l'installation du thésaurus dans le système, la question du codage des caractères, les fonctionnalités de navigation, recherche et sélection au sein du thésaurus, de façon autonome ou en conjonction avec un moteur de recherche.
- Les caractéristiques et fonctions requises pour les applications qui utilisent deux thésaurus ou plus seront traitées dans la Partie 2 de la norme ISO 25964.

4.3. Formats d'échange et protocoles

ISO 25964-1, Clause 17 - p. 118-119

ISO 25964-1, Clause 18 - , p.119-121

Les Clauses 17 et 18 exposent brièvement :

- Les 4 formats d'échange les plus courants : SKOS ; MARC21, zThes, et DD 8723-5 (BSI)
- Des cas d'utilisation de protocoles spécifiques pour l'échange de thésaurus, dont SKOS ou ADL Thesaurus protocole, ou des protocoles plus généraux comme Search Web Services d'OASIS ou SPARQL dans le monde du Web sémantique.

En raison de changements fréquents dans le domaine des technologies de l'information, ces chapitres introduisent la problématique et ne proposent que les pistes les plus évidentes en 2012.

Note concernant SKOS (information non incluse dans la norme)

SKOS (Simple Knowledge Organization System / Système simple d'organisation des connaissances) cité dans les deux clauses 17 et 18 de la norme ISO, est une famille de langages formels permettant une représentation standardisée des thésaurus, classifications ou de tout autre type de vocabulaire contrôlé et structuré sur le Web. SKOS est, depuis le 18 août 2009, une recommandation du W3C.

Ce standard du W3C s'est appuyé dès le démarrage du projet sur les normes existantes, en particulier les normes ISO 2788 or BS 8723-2:2005²⁰. Il est ainsi possible d'intégrer un thésaurus à un système d'information reposant sur les technologies du Web, tout en le rendant exploitable par des humains et par une machine. Cette recommandation du W3C ne fournit aucune directive ou recommandation pour la construction et la maintenance dans le temps du thésaurus lui-même, ni pour les spécifications destinées aux outils de gestion des vocabulaires. Pour cela il est recommandé de se reporter aux Clauses 5 à 14 de la norme ISO 25964-1.

²⁰ ISO 2788:1986 - http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=7776

BSI:8723-2:2005 : <http://shop.bsigroup.com/ProductDetail/?pid=00000000030094114>

Les communautés de la norme ISO 25964 et de la recommandation W3C SKOS²¹ proposent un [tableau de mise en correspondance](#)²² entre les éléments du modèle de données de la norme et ceux de [SKOS](#) et SKOS XL²³, une extension de SKOS. Ce tableau de correspondance, disponible sur le site dédié à la norme ISO 25964, remplace à compter du 12 décembre 2012²⁴, l'ancien tableau.

4.4. Conception, gestion et maintenance

ISO 25964-1, Clause 13 - Managing thesaurus construction and maintenance, p.88-98

ISO 25964-1, Clause 14 - Guidelines for thesaurus management software , p. 98-103

La norme s'est considérablement étoffée sur les aspects liés à l'administration et la maintenance du thésaurus :

- La Clause 13 traite spécifiquement du processus de création d'un thésaurus. Des recommandations complètent la présentation de chacune des étapes, et une section est dédiée à la mise à jour et à la nature des changements qui peuvent intervenir dans un thésaurus ;
- La Clause 14 fournit des spécifications fonctionnelles attendues pour un logiciel de gestion d'un thésaurus qui soit capable de gérer, importer et exporter les données d'un thésaurus telles que présentées dans la norme. Ces spécifications portent sur les relations entre termes et entre concepts, les notes appliquées aux concepts ou aux termes, les codes et les notations, les relais virtuels, le statut des langues, l'import/export de données, les fonctions éditoriales, la sécurité des données et du système et les outils d'administration.

²¹ Recommandation SKOS, 2009 - <http://www.w3.org/2004/02/skos/core#>

²² Tableau d'alignement entre ISO 25964 et SKOS (intégré à SKOS fin novembre 2012)
<http://www.niso.org/schemas/iso25964/correspondencesSKOS/>

²³ Recommandation SKOS XL, 2009 - <http://www.w3.org/2008/05/skos-xl#>

²⁴ Annonce officielle sur le site de SKOS le 12 décembre 2012 - <http://www.w3.org/2004/02/skos/>.

5 - Conclusion : Que faire ? Comment faire ?

Muni de cette norme, que peut-on faire aujourd'hui ?

Trois cas de figure se présentent aux gestionnaires de vocabulaires et aux concepteurs ou développeurs d'application :

« On met tout à la poubelle »

Certains prennent ce chemin, avec beaucoup de perte tant sur le plan des compétences, des outils sémantiques eux-mêmes, des contenus (banques de données) que sur le plan humain.

« On bascule »

Cette bascule peut être attaquée sous trois angles :

- Un changement de système technique conduit à se poser la question de l'usage d'un thésaurus, des modalités de sa gestion et diffusion ;
- La nécessité de diffuser la ressource sémantique dans le Web pour être exploitée plus largement²⁵ ;
- Le projet de développement d'un système plus complexe à base de règles (type ontologie de recherche) conduit à repositionner le modèle « thésaurus » en tant qu'outil de représentation des concepts. La norme se prête tout à fait à ce type de cas.

« On se prépare à basculer »

Dans le cas où il est impossible d'envisager aujourd'hui les changements portés par la norme ISO (distinction entre le réseau conceptuel et terminologique, spécialisation de relations, ...), le travail de fond sur le thésaurus lui-même (*sémantique métier*) peut toutefois être initié : version linguistique pour se préparer au multilinguisme, travail de fond sur les concepts et les termes, typage de relation en particulier pour les relations d'instance (TSI/TGI) ou développement des équivalences de type synonyme ou sigle/développé, facettes pour préparer à des accès métiers par facettes ou taxonomies sur des portails, établissement de liens vers d'autres ressources terminologiques pour optimiser des systèmes de recherches fédérées....

Cette préparation peut s'effectuer soit hors du système documentaire, soit avec ce dernier si celui-ci offre la possibilité d'ajouter des attributs exploitables dans une phase ultérieure du projet.

²⁵ Citons comme développements réalisés en France entrant dans cette cible :

- Le thésaurus utilisé par pour l'ensemble des services d'archives territoriales en France - <http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/thesaurus/>

- Le vocabulaire d'autorité du Sudoc sur le site IdRef - <http://www.idref.fr/autorites/autorites.html>

Annexes

Annexe 1 : Lexique Anglais-Français du vocabulaire de la norme

La terminologie de la norme ISO 25964-1 (Clause 3 – Terms and Definition) est intégrée à la plateforme dédiée de l'ISO (<http://www.iso.org/obp/ui/>)

Par domaines, puis par ordre alphabétique des termes en français

Remarque : La traduction en français et la numérotation associée aux termes en français proposées ici ne sont pas normalisées.

numFR	numEN	Terme FR	Terme EN
RESEAU CONCEPTUEL - CONCEPTS ET RELATIONS			
fr2.06	2.11	concept	concept
fr2.04	2.04	caractéristique de division	characteristic of division
fr2.07	2.52	concept de même rang, concept frère	sibling concept
fr2.11	2.08	équivalence composée	compound equivalence
fr2.13	2.20	facette	facet
fr2.26	2.50	note explicative, note d'emploi, note d'application	scope note
fr2.34	2.02	relation associative	associative relationship
fr2.35	2.18	relation d'équivalence	equivalence relationship
fr2.36	2.23	relation hiérarchique	hierarchical relationship
fr2.37	2.41	relation paradigmatique, relation a priori	a paradigmatic relationship, a priori relationship
fr2.38	2.59	relation syntagmatique, relation a posteriori	a syntagmatic relationship, a posteriori relationship
fr2.43	2.34	structure monohiérarchique	monohierarchical structure
fr2.44	2.42	structure polyhiérarchique	polyhierarchical structure
REPRESENTATION DE CONCEPTS : RESEAU DE TERMES			
fr2.15	2.24	homographe/ie	homograph
fr2.18	2.22	indicateur de facettes	facet indicator
fr2.25	2.40	notation, code ou numéro de classification, indice ou cote de classement	notation, class code, class number, classmark
fr2.31	2.47	quasi-synonyme, synonyme proche	quasi-synonym, near-synonym
fr2.33	2.38	étiquette de nœud, relais virtuel	node label
fr2.41	2.01	série (liste, regroupement)	array

fr2.45	2.58	synonyme	synonym
fr2.46	2.61	terme	term
fr2.47	2.48	terme associé	related term
fr2.48	2.09	terme composé	compound term
fr2.49	2.30	terme d'emprunt	loan term
fr2.51	2.53	terme de même rang, terme frère	sibling term
fr2.53	2.16	terme de recherche, terme rejeté	entry term - lead-in term
fr2.54	2.63	terme de tête (de hiérarchie)	top term
fr2.55	2.07	terme forgé	coined term
fr2.56	2.03	terme générique	broader term
fr2.57	2.36	terme multi-mots	multi-word term
fr2.58	2.39	terme non-préférentiel, non-descripteur	non-preferred term, non-descriptor
fr2.59	2.45	terme préférentiel, descripteur	preferred term, descriptor
fr2.60	2.37	terme spécifique	narrower term
SYSTEMES D'ORGANISATION DE CONCEPTS			
fr2.39	2.06	schéma de classification	classification scheme
fr2.40	2.57	schéma de vedettes matières, langage de vedettes matières, Liste de vedettes matières	subject heading scheme, subject heading language, subject heading list, SHL
fr2.61	2.62	thésaurus	thesaurus
fr2.62	2.35	thésaurus multilingue	multilingual thesaurus
fr2.63	2.12	vocabulaire contrôlé	controlled vocabulary
fr2.64	2.56	vocabulaire structuré	structured vocabulary
ALIGNEMENT (MAPPING)			
fr2.09	2.17	correspondance équivalente	equivalence mapping
fr2.12	2.13	équivalence interlangue	cross-language equivalence
fr2.19	2.29	interopérabilité	interoperability
fr2.20	2.60	langage cible	target language
fr2.22	2.54	langage source	source language
SYSTEME TECHNIQUE & INFORMATIQUE			
fr2.02	2.10	application informatique	computer application
fr2.03	2.31	balisage	markup
fr2.14	2.19	format d'échange	exchange format
fr2.16	2.25	identifiant	identifier
fr2.21	2.32	langage de balisage, de balises	markup language
fr2.24	2.14	modèle de données	data model

fr2.30	2.46	protocole	protocol
CONTEXTE GENERAL			
fr2.01	2.21	analyse par facettes	facet analysis
fr2.05	2.05	classification	classification, classifying
fr2.08	2.64	contrôle du vocabulaire	vocabulary control
fr2.10	2.15	document	document
fr2.17	2.27	indexation	indexing
fr2.23	2.33	métadonnée	metadata
fr2.27	2.49	plan, liste (répertoire du vocabulaire)	schedule
fr2.28	2.43	post-coordination	post-coordinate
fr2.29	2.44	pré-coordination	pre-coordinate
fr2.32	2.28	recherche d'information	information retrieval
fr2.42	2.55	spécificité	specificity
fr2.50	2.26	terme d'indexation (indexat)	index term
fr2.52	2.51	terme de recherche	search term

Par domaines, puis par ordre alphabétique des entrées en anglais

numFR	numEN	Terme FR	Terme EN
RESEAU CONCEPTUEL - CONCEPTS ET RELATIONS			
fr2.34	2.02	relation associative	associative relationship
fr2.04	2.04	caractéristique de division	characteristic of division
fr2.11	2.08	équivalence composée	compound equivalence
fr2.06	2.11	concept	concept
fr2.35	2.18	relation d'équivalence	equivalence relationship
fr2.13	2.20	facette	facet
fr2.36	2.23	relation hiérarchique	hierarchical relationship
fr2.43	2.34	structure monohiérarchique	monohierarchical structure
fr2.37	2.41	relation paradigmatique, relation a priori	paradigmatic relationship, a priori relationship
fr2.44	2.42	structure polyhiérarchique	polyhierarchical structure
fr2.26	2.50	note explicative, note d'emploi, note d'application	scope note
fr2.07	2.52	concept de même rang, concept frère	sibling concept
fr2.38	2.59	relation syntagmatique, relation a posteriori	syntagmatic relationship, a posteriori relationship
REPRESENTATION DE CONCEPTS : RESEAU DE TERMES			
fr2.41	2.01	série (liste, regroupement)	array

fr2.56	2.03	terme générique	broader term
fr2.55	2.07	terme forgé	coined term
fr2.48	2.09	terme composé	compound term
fr2.53	2.16	terme de recherche, terme rejeté	entry term - lead-in term
fr2.18	2.22	indicateur de facettes	facet indicator
fr2.15	2.24	homographe/ie	homograph
fr2.49	2.30	terme d'emprunt	loan term
fr2.57	2.36	terme multi-mots	multi-word term
fr2.60	2.37	terme spécifique	narrower term
fr2.33	2.38	étiquette de nœud, relais-virtuel	node label
fr2.58	2.39	terme non-préférentiel, non-descripteur	non-preferred term, non-descriptor
fr2.25	2.40	notation, code ou numéro de classification, indice ou cote de classement	notation, class code, class number, classmark
fr2.59	2.45	terme préférentiel, descripteur	preferred term, descriptor
fr2.31	2.47	quasi-synonyme, synonyme proche	quasi-synonym, near-synonym
fr2.47	2.48	terme associé	related term
fr2.51	2.53	terme de même rang, terme frère	sibling term
fr2.45	2.58	synonyme	synonym
fr2.46	2.61	terme	term
fr2.54	2.63	terme de tête (de hiérarchie)	top term
SYSTEMES D'ORGANISATION DE CONCEPTS			
fr2.39	2.06	schéma de classification	classification scheme
fr2.63	2.12	vocabulaire contrôlé	controlled vocabulary
fr2.62	2.35	thésaurus multilingue	multilingual thesaurus
fr2.64	2.56	vocabulaire structuré	structured vocabulary
fr2.40	2.57	schéma de vedettes matières, langage de vedettes matières, Liste de vedettes matières	subject heading scheme, subject heading language, subject heading list, SHL
fr2.61	2.62	thésaurus	thesaurus
ALIGNEMENT (MAPPING)			
fr2.12	2.13	équivalence interlangue	cross-language equivalence
fr2.09	2.17	correspondance équivalente	equivalence mapping
fr2.19	2.29	interopérabilité	interoperability
fr2.22	2.54	langage source	source language
fr2.20	2.60	langage cible	target language
SYSTEME TECHNIQUE & INFORMATIQUE			

fr2.02	2.10	application informatique	computer application
fr2.24	2.14	modèle de données	data model
fr2.14	2.19	format d'échange	exchange format
fr2.16	2.25	Identifiant	identifier
fr2.03	2.31	Balisateur	markup
fr2.21	2.32	langage de balisage, de balises	markup language
fr2.30	2.46	Protocole	protocol
CONTEXTE GENERAL			
fr2.05	2.05	Classification	classification, classifying
fr2.10	2.15	Document	document
fr2.01	2.21	analyse par facettes	facet analysis
fr2.50	2.26	terme d'indexation (indexat)	index term
fr2.17	2.27	Indexation	indexing
fr2.32	2.28	recherche d'information	information retrieval
fr2.23	2.33	Métadonnée	metadata
fr2.28	2.43	post-coordination	post-coordinate
fr2.29	2.44	pré-coordination	pre-coordinate
fr2.27	2.49	plan, liste (répertoire du vocabulaire)	schedule
fr2.52	2.51	terme de recherche	search term
fr2.42	2.55	Spécificité	specificity
fr2.08	2.64	contrôle du vocabulaire	vocabulary control

Annexe 2 : Tableau partiel des symboles et sigles

Description	Symbole	Etiquette EN	Etiquette FR	Signification
<i>Eléments descriptifs</i>		SN	NE	Note explicative <i>Une note signale l'usage de « note d'application » dans la norme Afnor Z47-100)</i>
		DEF	DEF	Définition
		HN	NH	Note historique
<i>Codes</i>		SC	_26	Catégorie sémantique ; un code ou une notation, appliqué à un groupe de concepts relatif à un sujet particulier.
		CC		Code de concept or notation
<i>Relations</i>	→	USE	EM	Employer ; le terme qui suit cette étiquette est le terme préférentiel devant être utilisé à la place du terme préférentiel qui précède l'étiquette.
	=	UF	EP	Employer pour ou Employé pour : le terme qui suit cette étiquette est un terme non préférentiel pour lequel doit être utilisé le terme préférentiel précédant l'étiquette.
		USE...+	EM... +	Les deux termes préférentiels ou plus qui suivent cette étiquette doivent être utilisés ensemble pour représenter le concept désigné.
		UF+	EM+	Le terme non préférentiel qui suit (cette étiquette) doit être représenté par la combinaison des termes préférentiels précédant l'étiquette.
		TT	_27	Terme de tête : le terme préférentiel qui suit cette étiquette représente le concept le plus haut dans la hiérarchie à laquelle appartient ce concept précis.
	<	BT	TG	Terme générique ; le terme préférentiel qui suit cette étiquette représente un concept ayant un sens plus large/générique dans la hiérarchie à laquelle appartient le concept spécifique.
		BTG	TGG	Terme générique (générique)
		BTI	TGI	Terme générique (instance)
	—<	BTP	TGP	Terme générique (partitif)
	>	NT	TS	Terme spécifique ; le terme qui suit cette étiquette renvoie à un concept ayant une sens plus spécifique.
		NTG	TSG	Terme spécifique (générique)
		NTI	TSI	Terme spécifique (instance)
	>—	NTP	TSP	Terme spécifique (partitif)
	—	RT	TA	Terme associé ; le terme qui suit cette étiquette est en relation d'association, mais n'est pas un synonyme, quasi-synonyme, terme générique ou terme spécifique.

²⁶ Il est possible de proposer pour la prochaine version de la norme (révision prévisible à partir de 2016) l'usage de CS pour catégorie sémantique. Beaucoup de pays ont tendance à utiliser les codes en anglais.

²⁷ Il est possible de proposer pour la prochaine version de la norme (révision 2016) l'usage de TT pour terme de tête (de la hiérarchie).

Annexe 3 : Modélisation entités-relations – quelques repères

Cette annexe a pour seule ambition d'apporter quelques éléments de compréhension pour ce Livre Blanc. Nous renvoyons les lecteurs intéressés à d'autres sources bibliographiques plus détaillées²⁸.

Pour représenter un thésaurus et sa structure conceptuelle et terminologique, les éditeurs de la norme ont choisi le langage UML et en particulier le vocabulaire et formalisme des diagrammes de classes, en raison de son important déploiement dans un grand nombre de pays dans différentes régions du monde.

UML est un langage graphique de modélisation des données et des traitements. Il propose une représentation graphique de ce modèle métier. Outil technique des développeurs, ces diagrammes constituent également des outils de communication avec l'ensemble des acteurs impliqués dans le développement des systèmes, moyennant une formation adaptée pour les intervenants n'ayant pas été familiarisés avec ce formalisme.

Toutefois il existe quelques limites. La norme ISO 25964-1 est ouverte et laisse un grand nombre de possibilités à l'initiative du concepteur de thésaurus. Le diagramme UML proposé dans la norme ne couvre donc pas toutes ces possibilités. Une note de spécifications est nécessaire pour chaque application. De plus, toutes les possibilités offertes dans la norme ne sont pas formalisables avec le langage UML et ne sont donc pas visibles sur le diagramme UML du modèle ISO 25964 proposé. Cette représentation graphique n'est donc pas autosuffisante pour spécifier et développer des applications. Les spécifications d'une application doivent donc s'appuyer précisément sur les besoins, le contenu de la norme et sur les tableaux de présentation du modèle de données proposés dans la Clause 15.

Vocabulaire UML

Seul le diagramme de classe est utilisé dans la norme ISO 25964-1.

UML s'appuie sur des **Classes** possédant des **attributs** qui les caractérisent et des **associations** entre classes.

- **Classes**

Déclare des caractéristiques communes à un ensemble d'objets, c'est-à-dire des attributs représentant l'état des objets

Exemple : classe « ThesaurusConcept », «ThesaurusTerm », ...

- **Association**

Connexion sémantique entre deux classes

Exemple : « *hasTopConcept* » entre concepts, ou « *isPreferredLabelFor* » entre les classes ThesaurusConcept et ThesaurusTerme,...

- **Attributs et propriétés**

Les attributs sont des éléments de description des classes ou des associations.

Chaque attribut a des propriétés, caractéristiques particulières qui précisent l'attribut : obligatoire ou optionnel, répétable ou non, type de données.

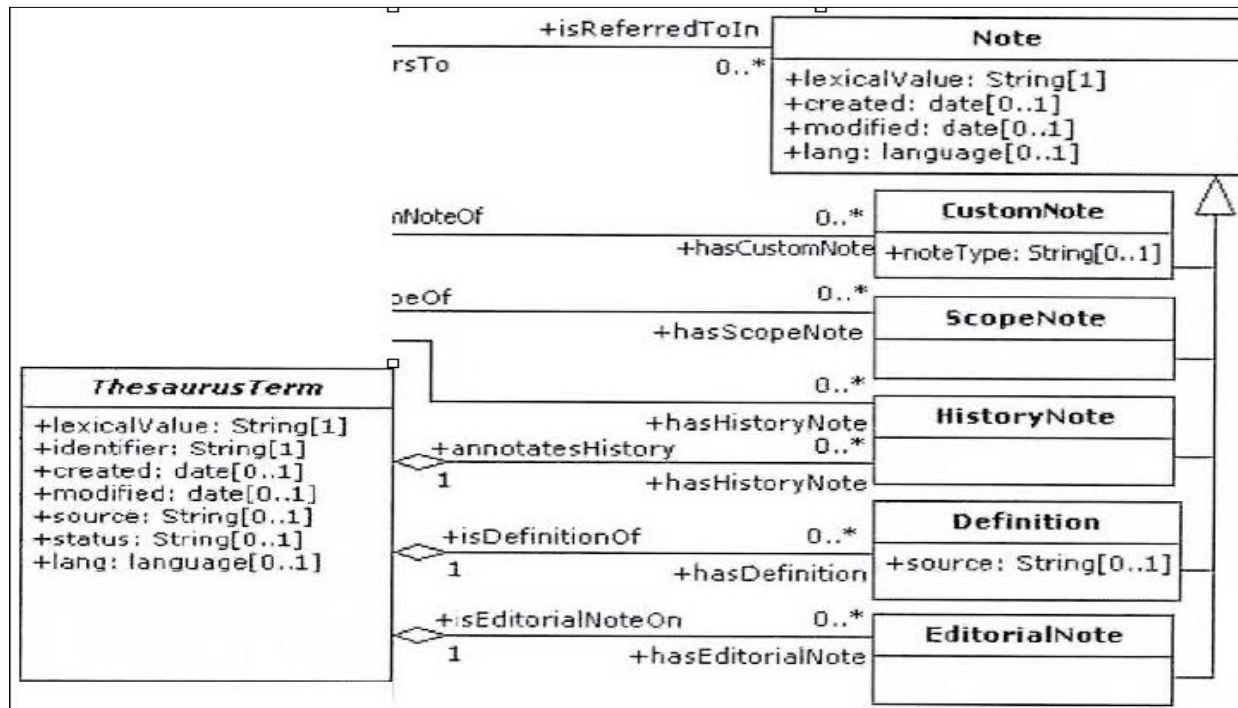
Exemple d'attributs pour la classe « ThesaurusConcept » : identifiant(*identifier*), date de création (*created*), de modification (*modified*) ; statut(*status*) ; notation (« *notation* ») ; concept de tête (« *topConcept* »).

Exemple de propriété d'un attribut : les identifiants sont de type de données « string » et sont

²⁸ Consulté le 10 juillet 2012 - UML (http://fr.wikipedia.org/wiki/Unified_Modeling_Language) et Diagramme de classes (http://fr.wikipedia.org/wiki/Diagramme_de_classes).

uniques.

Comment comprendre le diagramme UML ?



Clause 15, p.103 et 109

a) Caractéristiques des attributs

Obligatoire ou optionnel, et/ou répétable ou non

- 0..1 : optionnel et non répétable (zéro ou un)
- 0..* : optionnel et répétable (zéro ou plus)
- 1 : obligatoire, non répétable (un seulement)
- 1..* : obligatoire et répétable (un et plus)

Types données

- string (chaîne de caractère) ; boolean (true/false) ; date ; language ; class ; association class.

Exemples

+lexicalValue : String[1] : l'attribut Valeur lexicale (« lexicalvalue ») d'un terme est de type chaîne de caractères (« string ») ; il existe obligatoirement une valeur et est non répétable (1)

+lang:language[0..1] : l'attribut langue (« lang») d'un terme est de type langue («language»), c'est-à-dire que les valeurs sont choisies dans la liste des codes pays ISO 639-1 (p.106 de la norme) ; cet attribut est optionnel (0) ou bien s'il existe, cet attribut n'est pas répétable (1).

b) Caractéristiques des associations

Les associations entre classes sont représentées soit par trait simple lorsqu'elle sont de même importance, soit par un trait possédant à une extrémité un triangle ou un losange vide. Dans ces deux

derniers cas, la sémantique de cette association est la suivante :

- **is-a** : « **généralisation** » ; « est un type de » - association entre deux classes dont l'une est générale (classe de base ou classe parent) et l'autre spécialisée (ou sous-classe) ; cela se traduit par le concept d'héritage, l'enfant héritant des attributs du parent – *formalisme du triangle*.
Par exemple : les sous-classes « customNote », « scopeNote », etc héritent des attributs de la classe "Note ».
- **has-a** : « **agrégation** » ; relation partie/tout, d'inclusion - une classe constitue un tout composé de parties. L'agrégation est transitive – *formalisme du losange vide du côté de l'agrégat* .
Par exemple : +isDefinitionOf[0..*] : l'association « est une définition de » (« isDefinitionOf ») avec le losange vers la classe « ThesaurusTerm », indique que la classe « *Definition* » fait partie de la classe « *ThesaurusTerm* ».

Annexe 4 : Ressources bibliographiques

Où acheter la norme

ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: International Organization for Standards, 2011.

Site Afnor

(http://www.boutique.afnor.org/NEL5DetailNormeEnLigne.aspx?CLE_ART=XS120446&nivCtx=NELZNELZ1A10A101A107&ts=2801922)

Site ISO

(http://www.iso.org/iso/fr/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=01&ics2=140&ics3=20&csnnumber=53657)

Site pour la documentation technique (NISO) - <http://www.niso.org/schemas/iso25964/>

Ressources complémentaires

ISO 25964 : Vers une nouvelle norme pour l'organisation et l'accès à l'information et aux connaissances, Michèle Hudon, ISKO-France, 2011 ([Référence en ligne](#)).

Résumé – <http://conferences.isko-france.asso.fr/fr/abstracts/sa1a.htm>. Les actes sont sortis en mai 2012 chez Hermes Publication - <http://www.amazon.fr/Lorganisation-connaissances-dynamisme-techniques-linformation/dp/2746232278>

Thésaurus documentaire, Michèle HUDON, Techniques de l'Ingénieur, Thésaurus documentaire, Référence H7250, Date de publication : 10 nov. 2012

(<http://www.techniques-ingenieur.fr/base-documentaire/technologies-de-l-information-th9/gestion-de-contenus-numeriques-42311210/thesaurus-documentaire-h7250/>)

Les schémas de concepts et le Web sémantique : la norme sur les thésaurus ISO 25964 et le Web sémantique par Hélène Rabaut (Lexis Nexis) et Hélène Zysman (Smile - Open Source Solutions), Afnor/Bnf, 2011 ([Présentation en ligne](#))

Une nouvelle norme pour le thésaurus : modèles et perspectives à l'ère du web sémantique, Hélène Rabaut (Lexis Nexis) et Hélène Zysman (Smile - Open Source Solutions), ADBS, 2011 ([Présentation en ligne](#))

The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus, Leonard Will, ASIS, 2012 ([Texte en ligne](#)).

From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling, Stella G. Dextre Clarke et Marcia Lei Zeng, In Information Standards Quarterly, Winter 2012, v.24, no. 1 ([Texte pdf en ligne](#)).

Guide pratique pour l'élaboration d'un thésaurus documentaire, Michèle Hudon, avec la collaboration de Danièle Dégez et Dominique Ménillet. – Montréal : Les Éditions ASTED (diff. en France : ADBS), 2009. – 274 p. – ISBN 978-2-923563-17-6.

Annexe 5 – Organisation de la production de la norme ISO 25964

La norme ISO 25964, tout en restant en harmonie avec les normes et les systèmes plus anciens, offre un cadre normatif renouvelé. Pour conduire ce travail « entre l'ancien et le nouveau », de nombreuses personnes d'horizons différents ont été impliquées entre 2007 et 2012.

Au sein de l'ISO

17 pays ont participé aux différents votes au sein de l'ISO.

Le Groupe de travail dédié à la réalisation de ces travaux, TC46/SC9/WG8 - *Structured vocabularies*²⁹, piloté par Stella Dextre Clarke (UK), regroupe des experts de différents pays : Sylvie Dalbin (France) ; Johan De Smedt (Belgium) ; F. Javier García Marco (Spain) ; Michèle Hudon (Canada) ; Daniel Kless (Germany) ; Traugott Koch (Germany) ; Richard Light (U.K.) ; Jutta Lindenthal (Germany) ; Marianne Lykke (Denmark) ; Esther Scheven (Germany) ; Douglas Tudhope (U.K.) ; Leonard Will (U.K.) ; Marcia Zeng (U.S.).³⁰

Au sein de l'Afnor

Entre 2007 et décembre 2011, la France avait mis en place au sein de l'Afnor un Groupe de travail dédié (CG46/CN357/GE10) rattaché à la commission GC46/CN357 « Modélisation, production et accès aux documents »³¹ de l'Afnor, commission pilotée par Françoise Bourdon de la Bnf.

Le Groupe GE10 pour la production de la norme ISO 25964 était composé de : Dominique Chichereau ; Sylvie Dalbin** ; Alina Deniau ; François Feyler** ; Thierry Guillotin** ; Michèle Lénart ; Anne Martel ; Dominique Ménillet ; Mingam Michel ; Hélène Rabault** ; Virginie Triboulin ; Bernard Vatant** ; Nathalie Yakovieff** ; Hélène Zysman**.

** Participants à l'élaboration de la Partie 2 de la norme ISO 25964³², en cours d'édition (2013).

²⁹ ISO TC46/SC9 - http://www.nexus.ao/RS/ISO/www.iso.org/iso/iso_technical_committee6fac.html?commid=48836

³⁰ « Project Organization » sur le site dédié à la norme - <http://www.niso.org/schemas/iso25964/>

³¹ Afnor CG46/CN357 - http://www2.afnor.org/espace_normalisation/structure.aspx?commid=6152

³² Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies - http://www.iso.org/iso/catalogue_detail.htm?csnumber=53658