

**ISO 25964 : pour le développement, la gestion et
l'interopérabilité des langages documentaires**
**ISO 25964: The Development, the Management and the
Interoperability of Documentary Languages**
**ISO 25964: para el desarrollo, la gestión y la interoperabilidad
de los lenguajes documentales**

Michèle Hudon

Volume 58, Number 3, July–September 2012

Normalisation et normes pour la gestion de l'information et des documents

URI: <https://id.erudit.org/iderudit/1028903ar>

DOI: <https://doi.org/10.7202/1028903ar>

[See table of contents](#)

Article abstract

The context in which the ISO 25964 standard (Thesauri and Interoperability with other Vocabularies) was developed is outlined in this article. The structure of the standard is described as well as its potential impact on the semantic interoperability of systems and information retrieval.

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Hudon, M. (2012). ISO 25964 : pour le développement, la gestion et l'interopérabilité des langages documentaires. *Documentation et bibliothèques*, 58(3), 130–140. <https://doi.org/10.7202/1028903ar>

ISO 25964 : pour le développement, la gestion et l'interopérabilité des langages documentaires

MICHÈLE HUDON

EBSI, Université de Montréal
michele.hudon@umontreal.ca

RÉSUMÉ | ABSTRACTS | RESUME

Description du contexte au sein duquel la norme ISO 25964 Thésaurus et interopérabilité avec les autres vocabulaires a été conçue et élaborée. Présentation de la structure de la norme. Évo- cation de l'impact potentiel de cette norme sur l'interopérabilité sémantique des systèmes et sur la recherche d'information.*

ISO 25964 : The Development, the Management and the Interoperability of Documentary Languages

*The context in which the ISO 25964 standard (Thesauri and Interoperability with other Vocabularies**) was developed is outlined in this article. The structure of the standard is described as well as its potential impact on the semantic interoperability of systems and information retrieval.*

ISO 25964 : para el desarrollo, la gestión y la interoperabilidad de los lenguajes documentales

*Descripción del contexto en el cual la norma ISO 25964 Tesau- ro e interoperabilidad con otros vocabularios ha sido concebida y elaborada**. Presentación de la estructura de la norma. Evo- cación del impacto potencial de esta norma en la interoperabilidad semántica de los sistemas y en la búsqueda de información.*

ON RECONNAÎT LE CARACTÈRE ESSENTIEL de l'existence et de la stricte application de normes dans plusieurs domaines de l'activité humaine : la construction résidentielle, le transport aérien, les essais cliniques et la mise en marché de médicaments, etc. ; on admet également l'importance de la normalisation dans le développement des technologies. Il semble nettement plus difficile de percevoir l'intérêt de normaliser les composantes « non mécaniques » des systèmes de recherche d'information, les structures d'organisation des sujets et les langages de représentation et de recherche par exemple. C'est ainsi que certaines directives ayant trait au traitement de la documentation émises par l'Organisation internationale de normalisation (ISO), l'Association française de normalisation (AFNOR) et la British Standard Institution (BSI), entre autres, tendent à excéder de beaucoup la durée de leur vie utile et continuent d'être appliquées bien qu'on les sache désuètes.

Le cas qui nous intéresse est celui des directives pour l'élaboration et la gestion des thésaurus monolingues et multilingues. Rédigées dans le cadre du programme UNISIST de l'UNESCO au début des années 1970, ces directives furent approuvées et adoptées par l'ISO qui en proposa une deuxième édition à la communauté internationale en 1985 (ISO 5964 Thésaurus multilingues) et 1986 (ISO 2788 Thésaurus monolingues). Pourtant soumises à évaluation à tous les cinq ans, ces directives n'avaient jamais été remises en question avant les pressions exercées en 2007 par des spécialistes britanniques ayant pris l'initiative de réécrire leurs propres normes. Considérant la profondeur des changements qui ont radicalement transformé le champ du traitement et de la recherche documentaire au cours des trois dernières décennies, on comprend mal que ces directives n'aient pas été réévaluées pendant cette période.

Qu'on se rassure : les fondements philosophiques, sémantiques et terminologiques des processus de création et d'utilisation des langages documentaires restent bien évidemment valides. Mais avouons que le contexte d'utilisation des outils langagiers d'aide à la navigation et au repérage d'information, la diversité de leurs applications potentielles et la complexité des systèmes au sein desquels ils sont utilisés n'ont plus grand-chose à voir avec la réalité des années 1980. Rédigées dans un tout autre environnement, celui des collections tradi-

* Les titres et titres de section sont une traduction de l'auteur. Aucune traduction officielle en français de la norme ISO 25964 n'est encore disponible.

** The titles and section titles were translated by the author. An official French translation of the ISO 25964 standard is not yet available.

*** Los títulos y títulos de sección fueron traducidos por la autora. Todavía no existe traducción oficial en francés de la norma ISO 25964.

Cet article a pour objectif de présenter le contexte au sein duquel la norme ISO 25964 *Thésaurus et interopérabilité avec les autres vocabulaires* a été conçue, structurée et rédigée, ainsi que la structure et l'essentiel du contenu de la norme elle-même.

De l'importance des normes applicables aux langages documentaires

Plusieurs doutent encore de la nécessité de réécrire une norme consacrée aux langages documentaires, à l'époque des moteurs de recherche puissants et de la domination du vocabulaire libre et de la syntaxe naturelle pour la représentation et la recherche d'information. Pourtant, thésaurus, vedettes-matières et schémas de classification sont encore bien présents dans les systèmes d'information, bien qu'ils en aient quitté l'avant-scène pour mieux assumer leur rôle de soutien sémantique à la recherche. Ces moteurs de recherche qui donnent l'illusion de « réfléchir » et ces suggestions si « logiques » pour élargir ou restreindre la portée d'une recherche sont très dépendants des structures et des relations fournies par les langages contrôlés présents en arrière-plan. De plus, il semble raisonnable de croire que l'implantation annoncée du Web dit « sémantique » accroît, au lieu de la réduire, la nécessité d'un contrôle non intrusif des composantes langagières permettant d'accéder aux immenses réservoirs de connaissances constitués par l'ensemble des bases de données disponibles et d'autres types de ressources numériques réseautées.

Dans un environnement réseauté et décentralisé, les normes techniques sont certes essentielles, mais elles ne suffisent pas à régler tous les problèmes d'accès à la diversité des ressources disponibles. Les langues humaines, qui restent l'instrument privilégié pour la

Pourtant, thésaurus, vedettes-matières et schémas de classification sont encore bien présents dans les systèmes d'information, bien qu'ils en aient quitté l'avant-scène pour mieux assumer leur rôle de soutien sémantique à la recherche.

recherche d'information, résistent à toute tentative d'uniformisation; on peut au mieux les harmoniser et établir entre elles des passerelles, en conservant à chacune un statut égalitaire dans l'univers de la transmission d'information. Des méthodes de contrôle langagier éprouvées, résultant de l'application combinée de théories et de pratiques issues de la linguistique, de la terminologie et de la bibliothéconomie, servent encore à faciliter les échanges d'information. Aujourd'hui, indexeurs, chercheurs et moteurs de recherche utilisent les mêmes langages documentaires pour représenter et repérer le contenu thématique des documents: les fonctions, les structures et les modes de gestion et d'exploitation de ces langages contrôlés gagnent à être normalisés.

Par ailleurs, nombre de langages documentaires traditionnels sont toujours d'usage courant. Les systèmes de vedettes-matière et les schémas de classification donnent accès à de volumineuses collections de ressources disponibles sous forme imprimée ou numérisée et ces langages contrôlés, conçus en fonction des besoins et des fonctionnalités de systèmes d'information créés aux XIX^e et XX^e siècles, doivent pouvoir être exploités au XXI^e. Même s'ils poursuivent des objectifs distincts, ces outils d'accès traditionnels peuvent désormais être utilisés en combinaison pour améliorer les résultats de la recherche et la localisation de sources pertinentes. La mise à disposition de directives claires sur les processus d'alignement et d'établissement de passerelles entre les différentes formes de représentation d'un même concept (par exemple sous forme de vedette-matière, de descripteur et d'indice de classification) favorise l'interopérabilité sémantique des systèmes.

L'interopérabilité est la capacité qu'ont certains agents, services, systèmes et applications d'échanger des données, de l'information et des connaissances en préservant l'intégrité et la pleine signification de celles-ci. L'interopérabilité sémantique concerne plus spécifiquement le langage et le vocabulaire utilisés dans un processus de communication (entre systèmes informatiques ou entre humains et systèmes informatiques); elle facilite pour l'utilisateur le repérage et le partage d'information, peu importe le langage et le vocabulaire utilisés. L'interopérabilité sémantique passe par le contrôle plus ou moins strict des significations, contrôle associé aux systèmes d'organisation des

connaissances (SOC) de types verbal ou symbolique. Or, les SOC, et plus particulièrement les thésaurus, se sont multipliés avec l'intégration d'outils technologiques dans les systèmes de recherche d'information. Cette multiplication des langages documentaires a de beaucoup diminué la capacité des systèmes à communiquer et à échanger des données, tout en compliquant la vie des usagers, forcés d'apprendre un langage différent à chaque fois qu'ils interrogeaient une nouvelle base de données. Dans les réseaux d'information contemporains, les problèmes liés à la multiplicité et à la diversité des langages de recherche sont devenus critiques : comment peut-on, par exemple, assurer l'exhaustivité ou la précision du repérage dans un système où cohabitent les indices de la classification Dewey, des vedettes-matière et les termes d'un thésaurus spécialisé ? Le chercheur d'information ne peut que bénéficier de tout effort qui lui permette d'explorer simultanément plusieurs réservoirs d'information à l'aide du langage documentaire qui lui est le plus familier, sans égard à la diversité des langages utilisés pour représenter les concepts et les sujets au sein de chacun d'entre eux.

Enfin, normes et directives sont d'autant plus utiles aujourd'hui que les opérations de représentation et de recherche ne sont plus l'apanage de spécialistes formés à l'analyse documentaire et à la recherche d'information. Dans les milieux autres que les milieux traditionnels de la documentation, la gestion et la recherche d'information et de documents sont la responsabilité de non-spécialistes (de l'information), de plus en plus souvent le créateur même de la ressource ou l'individu qui en a un besoin immédiat.

Normes pour l'élaboration et la gestion des thésaurus

Devant l'indifférence apparente des autorités compétentes face à la nécessité de revoir la portée et le contenu des normes ISO 2788 et ISO 5964, quelques spécialistes britanniques du développement et de la gestion des langages documentaires décident en 2000 de s'attaquer à la révision des normes britanniques équivalentes (BS 5723 et BS 6723), avec l'intention claire de proposer ensuite une nouvelle norme à la communauté internationale (Dextre Clarke, Gilchrist et Will 2004). Sept ans plus tard, la cinquième et dernière section de la nouvelle norme BS 8723 *Structured vocabularies for information retrieval* est publiée et l'importance accordée à l'interopérabilité sur les réseaux numériques mondiaux y est évidente.

Les sections 1 et 2 de la norme britannique témoignent d'une attitude plutôt conservatrice chez ses rédacteurs ; ces sections reprennent en effet la quasi-totalité du contenu de la norme ISO 2788-1986. Nos collègues britanniques innoveront toutefois dans la préparation de trois sections inédites. La section 3 « *Vocabularies other than thesauri* » décrit le contenu et la

structure des schémas de classification, des répertoires de vedettes-matière, des taxinomies, des ontologies et des fichiers d'autorités, dans le but d'aider les utilisateurs de la norme, de plus en plus diversifiés, à percevoir les différences qui existent entre ces outils. La section 4 « *Interoperability between vocabularies* » énonce des directives pour l'établissement de passerelles dans le but de favoriser l'interopérabilité sémantique des systèmes d'information ; c'est ici que l'on trouve l'essentiel du contenu portant sur les thésaurus multilingues, considérés comme cas complexe dans un processus de mise en équivalence. La cinquième et dernière section témoigne du caractère essentiel de la technologie dans l'environnement théssaural en détaillant modèles, protocoles et formats pour le traitement et l'échange des données.

Les Britanniques ne sont pas les seuls à devancer l'ISO dans ces travaux d'envergure devenus essentiels. Au début des années 2000, un groupe de travail s'attelle aux États-Unis à la restructuration de la norme pour le développement des thésaurus monolingues. La norme révisée NISO/ANSI Z 39.19 est mise à disposition en 2005. Les nouvelles directives ne s'appliquent pas seulement aux thésaurus mais également à d'autres types de vocabulaires contrôlés, tels les anneaux de synonymes et les taxinomies. Contrairement à la norme britannique cependant, la norme Z 39.19-2005 ne couvre ni les vocabulaires multilingues, ni la mise en correspondance de langages documentaires distincts.

Les langages documentaires multilingues sont par contre l'unique préoccupation d'un groupe de travail de l'IFLA, formé en 2003 à l'instigation du néerlandais Gerhard Riesthuis. Une série de directives s'appliquant au développement des thésaurus multilingues et destinées à être utilisées en complément à la norme ISO 5964 est publiée en 2009 (IFLA 2009).

En France, le groupe informel Langages documentaires de l'Association des documentalistes et bibliothécaires spécialisés (ADBS) se donne pour tâche en 2005 d'examiner les normes NISO Z 39.19-2005 et BS 8723 récemment publiées. Le groupe oriente sa réflexion autour de plusieurs questions : comment ces nouvelles normes posent-elles la problématique des langages documentaires ? En quoi sont-elles novatrices ? Comment les taxinomies et les ontologies peuvent-elles s'articuler aux thésaurus ? (Chichereau *et al.* 2007). Le groupe reconnaît les caractéristiques essentielles de l'environnement actuel d'utilisation des normes : un public plus large, des interfaces souples et faciles d'accès, une gestion des recherches moins sophistiquée et non exclusivement destinée à des professionnels médiateurs, des innovations technologiques constantes et la multiplication de nouveaux outils d'organisation de l'information et des connaissances.

Naissance et développement d'une norme ISO

La majorité des normes ISO sont soumises à évaluation de façon régulière, soit à tous les trois ou cinq ans. Les pays membres du comité technique responsable de la norme à évaluer sont alors requis de se prononcer sur la nécessité de réviser, de retirer ou de reconduire celle-ci pour une nouvelle période à durée déterminée. S'il est décidé à la majorité des membres que des changements sont requis, l'opération de révision (*re-editing*) est lancée¹.

La première étape de l'opération est la « proposition de nouveau projet ». À ce stade, un document de travail proposant la restructuration et la réécriture de la norme est préparé par un ou plusieurs membres du comité technique. Si la proposition est acceptée, un groupe de travail constitué d'experts est formé, avec pour premier mandat de préparer un « avant-projet de norme » (*committee draft*), lequel sera soumis à évaluation et commentaires. Le travail requis est plus ou moins lourd, selon la longueur et la complexité de la norme à réviser, la période écoulée depuis sa dernière révision, le nombre et l'importance des changements à apporter, la disponibilité de documents de référence, le nombre de délégués auprès desquels il faut rechercher un consensus, la compétence du chef de projet, etc. Le calendrier des travaux prévoit normalement la publication de la norme révisée au plus tard 36 mois après acceptation de la proposition de nouveau projet.

Le groupe de travail emprunte ou rédige un premier texte, à partir duquel il prépare son avant-projet de norme. L'avant-projet est soumis pour commentaires à l'ensemble des membres du comité technique responsable de la norme. Il n'est pas rare que plusieurs avant-projets successifs soient soumis pour commentaires dans un processus itératif qui vise l'atteinte d'un consensus sur le contenu technique du document et sur sa présentation.

Une fois le consensus obtenu, une version définitive du texte devenu « projet de norme » (*draft international standard*) est expédiée à tous les comités techniques de l'ISO et soumise au vote. Les membres votants disposent de cinq mois pour examiner le texte de la norme et à ce stade il leur est encore possible de suggérer des ajouts. Un projet final de norme est soumis à l'ensemble des comités techniques de l'ISO deux mois avant publication ; seules des modifications mineures, d'ordre stylistique ou typographique, peuvent alors y être apportées.

On notera qu'une proportion importante des 36 mois alloués au développement d'une norme est en fait consacrée à la relecture et à l'évaluation par les délégués des pays membres de l'ISO. Les périodes consa-

crées à la relecture pourraient sembler inutilement longues, mais elles sont nécessaires. Lorsqu'il reçoit un projet de norme, l'organisme responsable dans chaque pays de la normalisation dans un domaine particulier doit recruter des experts qui accepteront de lire très attentivement et de commenter le contenu d'un ou même de plusieurs documents et ce, sans aucune compensation financière ou autre. La lecture d'une norme en développement est un travail ardu exigeant du temps et une énergie que plusieurs praticiens ne sont pas toujours en mesure de fournir. Au Canada, pour examiner les normes de la documentation par exemple, Bibliothèque et Archives Canada (BAC) doit solliciter des évaluateurs provenant des deux groupes linguistiques officiels, de diverses régions du pays, de diverses communautés de pratique (université, recherche, gouvernement, organisme à but non-lucratif, entreprise privée), etc. Ces spécialistes ne sont pas nécessairement largement connus et un travail de réseautage considérable peut devoir être effectué en un temps record.

Le projet de norme ISO 25964

Le Comité technique 46 (TC46) « Information et documentation » est chargé à l'ISO des normes touchant les pratiques relatives aux bibliothèques, aux centres de documentation, à l'édition, aux archives, à la gestion des documents administratifs, à la documentation muséographique, aux services d'indexation et de résumés et aux sciences de l'information plus généralement ; une quarantaine de normes sont sous sa responsabilité directe et une centaine d'autres sont produites par d'autres Comités techniques avec lesquels TC46 maintient des liens (citons en exemple les Comités techniques 37 « Terminologie et autres ressources langagières et ressources de contenu », et 154 « Processus, éléments d'information et documents dans le commerce, l'industrie et l'administration »). Trente-cinq pays participent activement aux travaux du Comité technique 46 (dont l'Australie, le Canada, la Chine, les États-Unis et la France) et 36 autres y ont un statut d'observateur.

Le Sous-comité 9 (SC9) du Comité technique 46, « Identification et description », est responsable à lui seul d'une vingtaine de normes. On y dénombre une dizaine de groupes de travail œuvrant à la préparation de nouvelles normes ou à la réécriture de normes publiées.

En mai 2007, lorsque les normes ISO 2788 et ISO 5964 sont une nouvelle fois soumises à évaluation, c'est au Comité technique 46 que les représentants du Royaume-Uni soumettent la norme BS 8723-2005 aux fins de comparaison et d'évaluation par la communauté internationale. Leur proposition souligne qu'une mise à jour des normes ISO 2788 et ISO 5964 sera bénéfique à tous ceux qui font usage de volumineux réservoirs d'information dans les secteurs publics, privés et sans but

1. On peut trouver une description complète du processus de préparation d'une norme ISO sur le site de l'Organisation à <http://www.iso.org/iso/fr/standards_development.htm>.

lucratif. Le projet de norme ISO 25964 est mis sur pied lorsque le Comité technique 46 rejette (enfin !) le *statu quo* et recommande l'utilisation de la norme britannique comme document de référence pour le remplacement des directives pour le développement et la gestion des thésaurus en vigueur depuis plus de 20 ans.

Le projet de norme ISO 25964 est officialisé en octobre de la même année par la création du groupe de travail TC46/SC9/WG8 « Vocabulaires structurés », formé de spécialistes représentant une quinzaine de pays et plusieurs groupes linguistiques². Le mandat du groupe de travail est clair : « *développer une nouvelle norme ISO portant sur les vocabulaires structurés pour remplacer les normes existantes, i.e. ISO 5964 publiée en 1985 et ISO 2788 publiée en 1986* ». Le WG8 est dirigé par Stella Dextre Clarke, spécialiste du développement de thésaurus et de taxinomies et corédactrice de la norme BS 8723-2005.

Les membres du groupe de travail se rencontrent virtuellement lors de conférences téléphoniques tenues le premier lundi de chaque mois. Un ordre du jour serré permet d'entendre les points de vue de chacun sur les questions litigieuses et de faire progresser la rédaction des directives et la validation des exemples. La présidente du groupe attribue à chacun des mandats et des tâches précises, filtre les commentaires, intègre les ajouts et corrections sur lesquels il y a consensus, applique les directives de l'ISO pour tout ce qui concerne l'utilisation d'une terminologie et d'un style de rédaction normalisés ainsi que pour la mise en forme des textes. Les membres du WG8 et quelques consultants externes ont accès à l'ensemble des ordres du jour, procès-verbaux, brouillons, commentaires, textes et messages échangés sous forme électronique dans un espace de travail virtuel créé à leur intention par la National Information Standards Organization (NISO). Les membres du groupe de travail ont également tiré avantage de quelques séances de travail en mode présentiel, dans les bureaux londoniens du British Standards Institute (BSI) ; de nombreuses décisions ont pu alors être finalisées et des avancées importantes ont été réalisées.

Deux sous-groupes de travail ont été actifs en 2008-2009. Un premier sous-groupe s'est consacré à la préparation des clauses concernant les thésaurus multilingues et le deuxième a structuré les clauses portant sur les modèles de données. Un troisième sous-groupe a pris forme en 2010 pour convenir de la place à accorder aux ontologies et proposer les contenus appropriés.

2. Les pays représentés dans le groupe de travail original étaient l'Afrique du Sud, l'Allemagne, la Bulgarie, le Canada, la Chine, le Danemark, l'Espagne, les États-Unis, la Finlande, la France, la Nouvelle-Zélande, le Royaume-Uni et la Suède. Des représentants de la Belgique et de l'Australie ont intégré le groupe après le début des travaux.

Structure et contenu de la norme ISO 25964

D'entrée de jeu, il fut convenu entre les parties intéressées que la portée de la norme ISO 25964 serait plus large que celle des normes ISO 2788 et ISO 5964 qu'elle remplace, et qu'elle traiterait de l'établissement de passerelles entre diverses formes de langages documentaires et d'interopérabilité sémantique des systèmes d'information. Il parut évident que la nouvelle norme devrait refléter clairement l'importance du numérique et des réseaux et reposer sur des bases dont la validité ne serait pas liée à des technologies, des logiciels ou des architectures de systèmes particuliers.

Les cinq sections de la norme britannique BS 8723 furent soumises en janvier 2008 à l'ensemble des membres du groupe de travail pour lecture et évaluation. Cette première opération a généré des centaines de commentaires, certains très critiques, et mené à la première décision importante du groupe, celle de structurer la norme ISO 25964 en deux sections distinctes plutôt qu'en cinq sections, comme le faisait BS 8723. La première section traiterait uniquement des thésaurus et s'inspirerait des sections 1, 2 et 5 de la norme britannique, la seconde traiterait de passerelles et d'interopérabilité en intégrant le contenu des sections 3 et 4.

Objet et domaine d'application

La norme ISO 25964 offre des recommandations pour le développement et la gestion de vocabulaires structurés de termes représentant des sujets et des concepts utilisables pour la recherche d'information. Elle propose également des directives pour l'échange de données entre vocabulaires contrôlés et entre un vocabulaire contrôlé et l'application au sein de laquelle il est utilisé³.

La norme ISO 25964 s'applique aux vocabulaires utilisés pour le repérage d'information dans des sources textuelles telles que les bases de connaissances, les bases de données bibliographiques et les collections de documents, mais aussi dans les catalogues d'images ou d'objets comme ceux qui décrivent les collections muséales par exemple. Les vocabulaires structurés sont utilisés dans des systèmes d'information post-coordonnés, dans les répertoires hiérarchisés, les index pré-coordonnés et les systèmes de classification⁴.

3. « ISO 25964 gives recommendations for the development and maintenance of structured vocabularies of subject terms and concepts for information retrieval. It also gives guidance on the exchange of data between multiple vocabularies and between a vocabulary and the software with which it interacts ». (ISO 25964 Part 1).

4. « ISO 25964 applies to vocabularies used for retrieving information from text resources, such as knowledge bases, bibliographic databases or collections of full text documents, and also catalogues of images or artefacts, such as those in a museum collection. Structured vocabularies are used in post-coordinate retrieval systems, hierarchical directories, pre-coordinate indexes and classification systems ». (ISO 25964 Part 1).

Tableau 1.

Comparaison des clauses principales dans les normes ISO 25964, 2788 et 5964.

ISO 25964	ISO 2788 ET ISO 5964
4. Le thésaurus et ses fonctions	5. Le contrôle du vocabulaire
5. Types de concepts représentés	6. Les termes d'indexation
6. Sélection et forme des termes dans un thésaurus	6.2 La forme des termes d'indexation 6.5 Le choix des termes
7. Traitement des concepts complexes et des termes composés	7. Les termes composés
8. L'équivalence en contexte monolingue	8. Les relations dans un thésaurus 8.2 La relation d'équivalence
9. L'équivalence entre langues naturelles	9. Problèmes langagiers dans les thésaurus multilingues (ISO 5964)
10. Relations entre concepts 10.2 Relation hiérarchique 10.3 Relation associative 10.4 Autres relations	8.3 La relation hiérarchique 8.4 La relation associative
11. L'analyse par facettes	--
12. Présentation et affichage des données	9. L'affichage des termes et des relations (ISO 2788)
13. Construction et gestion du thésaurus	10. Construction du thésaurus (ISO 2788) 15. Organisation du travail (ISO 5964)
14. Fonctionnalités requises des logiciels dédiés à la gestion des thésaurus	--
15. Modèle de données	--
16. Intégration du thésaurus au SRI	--
17. Formats d'échange des données	--
18. Protocoles	--

Structure

Lorsqu'on compare la norme ISO 25964 à celles qu'elle remplace, on remarque immédiatement qu'elle propose un volume considérable de textes explicatifs ou descriptifs plutôt que strictement normatifs. Les normes ISO 2788 et ISO 5964 s'adressaient à des spécialistes et misaient sur la présence de nombreux exemples pour assurer la compréhension. Ici, le parti de l'explication a été pris dans le but évident de favoriser l'utilisation de la norme par les développeurs de langages ou de structures pour la représentation et la recherche d'information qui ne sont pas familiers avec les règles du contrôle du vocabulaire.

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

Partie 1 Thésaurus et recherche d'information (Thesauri for information retrieval)

C'est dans la première section de la norme que l'on trouve les principes généraux et les recommandations applicables à la conception et à la gestion des thésaurus monolingues et multilingues. Le texte des dix premières clauses de la section 1 reste très près de celui que l'on

trouvait déjà dans les normes ISO 2788 et ISO 5964 (Tableau 1). Ainsi, aucune modification majeure n'est à signaler dans les clauses touchant aux types de concepts représentés dans les thésaurus (Clause 5), à la forme des termes et aux méthodes de désambiguïsation des homographes (Clause 6), aux recommandations concernant le traitement des concepts et des termes complexes (Clause 7). La structure relationnelle du thésaurus, telle que présentée dans les clauses 8 à 10, ne s'éloigne pas non plus du cadre traditionnel. Les quelques paragraphes (Clause 10.4) qui abordent la question de l'utilité potentielle de relations additionnelles, plus explicites, ressemblent davantage à une mise en garde sur la difficulté de l'opération qu'à un encouragement à enrichir le contenu du thésaurus.

La partie 1 de la norme propose par ailleurs plusieurs directives nouvelles. La clause 11 réaffirme l'utilité de l'analyse par facettes, un processus auquel on s'intéresse davantage aujourd'hui qu'au moment de la rédaction des normes ISO 2788 et 5964 ; la clause synthétise des éléments d'information disséminés auparavant dans les clauses portant sur le développement de la structure et la présentation des données. Les clauses 12 « Présentation et affichage », 13 « Gestion (de la création à la mise à jour) » et 14 « Logiciels pour la

Tableau 2.

ISO 25964-2 Structure et contenu.

ISO 25964-2 SECTION 1	ISO 25964-2 SECTION 2
5. Objectifs	17. Schémas de classification
6. Modèles structuraux pour l'établissement de passerelles	18. Plans de classement pour la gestion des documents
7. Types de « mapping »	19. Taxinomies
8. Équivalence	20. Systèmes de vedettes-matière
9. Hiérarchie	21. Ontologies
10. Association	22. Terminologies
11. Équivalence exacte, inexacte et partielle	23. Liste d'autorité des noms
12. « Mapping » et recherche d'information	24. Anneaux de synonymes
13. La pré-coordination	
14. Identification des candidats	
15. Gestion et échange des données	
16. Affichage des données	

gestion des thésaurus » ont été révisées et augmentées pour rendre compte de la réalité du contexte actuel de gestion et d'exploitation des thésaurus. Quant au modèle de données et au schéma XML dérivé détaillés à la clause 15, ils résultent des discussions et décisions d'un sous-groupe de spécialistes membres du WG8 et d'experts externes appelés en renfort. Il faut souligner que, bien que les deux modèles ne soient pas identiques, le modèle de données proposé par ISO 25964 est compatible dans ses grandes lignes avec le modèle SKOS (Simple Knowledge Organization System) du W3C (Miles et Bechhofer 2009) ; la publication et l'exploitation sur le Web de thésaurus développés dans le respect de la norme ISO 25964 devraient en être grandement facilitées. Les clauses 16 à 18, sur les protocoles d'application, témoignent de la volonté de présenter le thésaurus comme composante essentielle d'un système de recherche d'information.

Partie 2 Interopérabilité avec les autres vocabulaires (Interoperability with other vocabularies)

La Partie 2 de la norme ISO 25964 est consacrée à l'établissement de passerelles (*mapping*) entre langages documentaires pour faciliter, d'une part, l'atteinte de l'objectif d'interopérabilité des langages eux-mêmes et, d'autre part, d'interopérabilité sur le plan sémantique des systèmes d'information qui en font usage. ISO 25964-2 propose des solutions fonctionnelles aux problèmes engendrés par l'utilisation simultanée de plusieurs langages documentaires pour accéder rapidement à de larges collections de ressources distribuées sur de multiples réseaux. L'existence de passerelles entre langages documentaires permet la conversion automatique de termes d'indexation ou d'expressions de recherche établis à l'aide d'un langage documentaire

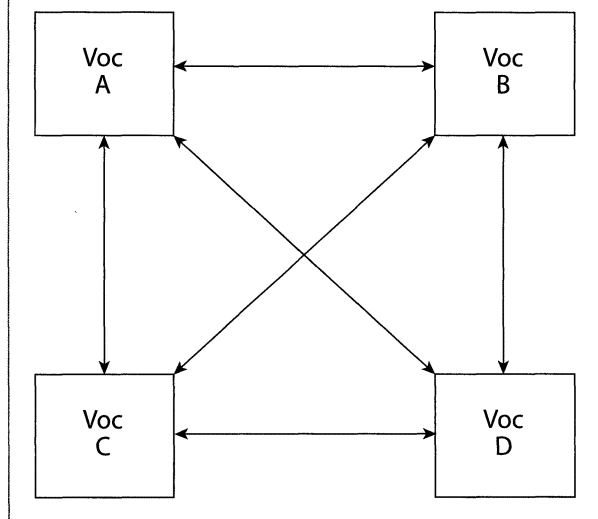
X en termes d'indexation ou expressions de recherche établis à l'aide d'autres langages documentaires de structure identique ou différente.

On distingue deux sous-sections dans la partie 2 de la norme (Tableau 2).

Les 16 premières clauses présentent le processus de mise en équivalence. Les clauses 17 à 24 traitent de l'établissement de passerelles entre le thésaurus, objet de la norme, et une variété de langages documentaires traditionnels (thésaurus, système de vedettes-matière, schéma de classification, liste d'autorité des noms, terminologie) ou d'implantation récente (taxinomie, ontologie, anneau de synonymes). L'objectif n'est pas de guider le développement et la gestion de ces langages documentaires, et le minimum d'information offert sur leur contenu et leur structure ne sera d'aucune assistance aux spécialistes qui les connaissent déjà. Les clauses 17 à 24 sont plutôt rédigées à l'intention des professionnels dont la tâche est de rendre interopérables des langages à structures identiques, comparables ou complètement différentes, et elles soulignent les caractéristiques de chaque langage qui pourraient avoir un impact sur le processus. La mise en équivalence est toujours directionnelle et seul l'établissement d'une passerelle qui part du thésaurus pour aller vers un autre langage est traité, jamais l'inverse. Les cas et les difficultés les plus courants sont abordés ; les cas d'exception, qu'on sait très nombreux, ne sont pas présentés.

La norme ISO 25964-2 définit le processus de mise en correspondance et son résultat comme l'établissement d'une relation entre un concept représenté dans un langage documentaire et un ou plusieurs concepts représentés dans un autre langage documentaire. La relation la plus importante et la plus intéressante est évidemment l'équivalence, mais il peut être utile de reconnaître, à l'occasion, une relation hiérarchique ou

Figure 1.
Mise en équivalence directe



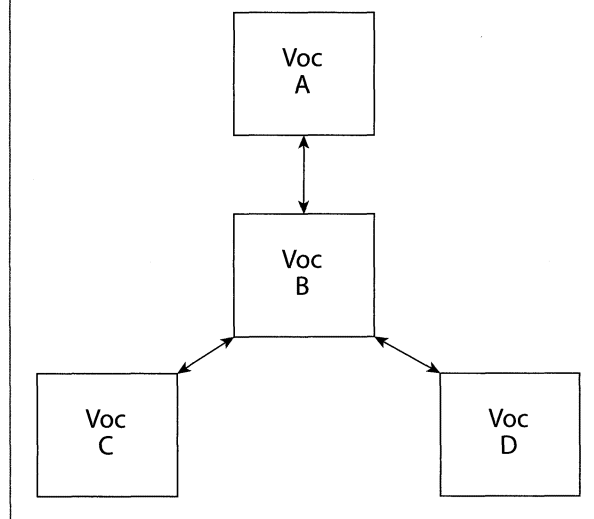
une relation d'association entre deux concepts, comme on le fait déjà dans un thésaurus monolingue.

L'utilité des deux parties de la norme est incontestable mais la partie 2 est l'apport le plus significatif au fonctionnement des systèmes et à l'efficacité de la recherche d'information. Par ses directives concrètes sur l'établissement de passerelles, la norme ISO 25964-2 contribue à l'atteinte de l'objectif d'interopérabilité sémantique des systèmes d'information en facilitant la recherche simultanée dans plusieurs bases de données à partir d'un seul et même langage, ainsi que l'accès aux connaissances représentées dans plusieurs langues naturelles. Elle favorise aussi la démocratisation de l'accès à d'énormes réservoirs de données et de ressources jusque-là disponibles à quelques individus seulement et la fin de l'isolement dans lequel ont longtemps existé les bases de données et les collections documentaires. Elle facilite enfin l'ouverture des systèmes d'information à des usagers non professionnels, pour la recherche d'abord, mais également pour la représentation des concepts et des sujets, sur le Web participatif notamment.

Modèles de mise en équivalence

Lorsqu'il s'agit de mettre en correspondance plusieurs langages de représentation et de recherche, deux façons de faire paraissent efficaces. Dans le cas d'une opération d'établissement de passerelles entre plusieurs langages documentaires, l'équivalence directe implique l'établissement de passerelles distinctes entre chaque concept représenté dans le langage A et ses équivalents dans chacun des langages B, C, D...n, entre chaque concept représenté dans le langage B et ses équivalents dans chacun des langages A, C, D...n, etc. (Figure 1). Le processus est efficace et précis, mais coûteux en temps et en ressources.

Figure 2.
Mise en équivalence par commutation



La mise en équivalence par commutation est plus économique mais présente le désavantage de n'être pas aussi exacte que la mise en équivalence directe. Lorsque ce modèle est adopté, chacun des concepts d'un langage A est lié à un ou plusieurs équivalents présents dans un langage de commutation X, chacun des concepts du langage B est à son tour lié à ses équivalents dans le langage de commutation X, et ainsi de suite (Figure 2). Une seule opération d'établissement de relations est ici requise pour chacun des langages à rendre interopérables, alors que le nombre d'opérations de mise en relation nécessaires est équivalent au nombre de langages documentaires concernés dans le modèle direct.

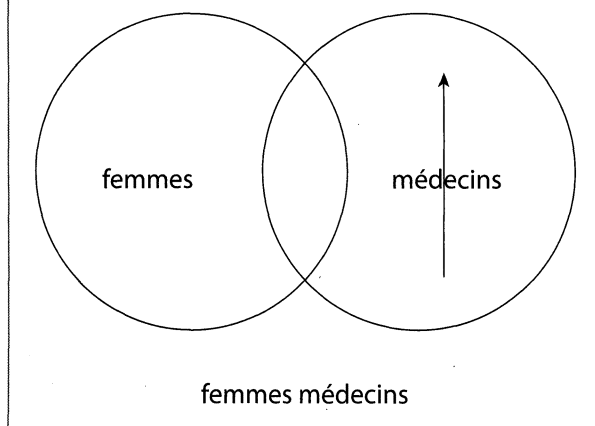
Dans la figure 2, c'est le langage B qui joue le rôle de langage de commutation et qui permet de lier indirectement les langages A, C et D. Le langage de commutation peut être un langage verbal, un thésaurus ou un répertoire de vedette-matières par exemple, ou un langage symbolique comme un schéma de classification. On choisit comme langage de commutation celui qui énumère déjà la totalité, ou à tout le moins le plus grand nombre, des concepts représentés dans tous les vocabulaires à réconcilier. Si un tel langage n'est pas disponible, il peut être avantageux de construire un langage de commutation.

Équivalence exacte et inexacte

La relation d'équivalence entre concepts est essentielle à l'établissement de passerelles entre langages documentaires couvrant en totalité ou en partie le même domaine. ISO 25964-2 reconnaît l'existence de deux types d'équivalence, l'équivalence exacte et l'équivalence inexacte.

La création du lien d'équivalence entre concepts représentés dans des langages documentaires distincts,

Figure 3.
Exemple d'équivalence composée (par intersection)

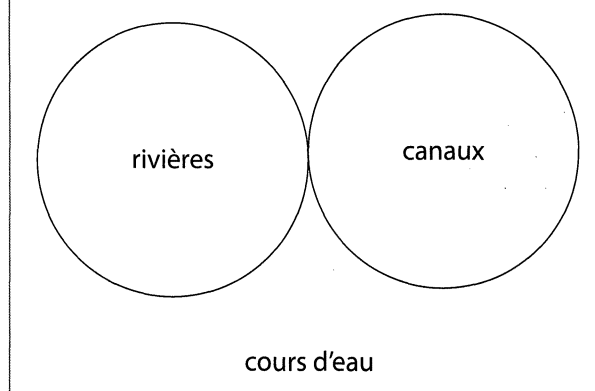


peu importe que ces derniers soient de structure similaire ou différente, n'est pas un processus aussi simple qu'on peut l'imaginer au premier abord. Même dans des langages documentaires qui couvrent un même domaine (deux thésaurus dans le domaine de l'éducation ou encore un thésaurus en éducation et la division 370 de la Classification décimale de Dewey par exemple), tous les concepts ne seront pas nécessairement représentés et ceux qui le sont ne sont pas nécessairement représentés de la même façon. C'est donc l'équivalence inexacte qui se rencontrera le plus souvent ; on en trouve la description dans les clauses de la norme qui portent sur l'équivalence dite « composée » (ou encore équivalence un-à-plusieurs). Une équivalence composée existe lorsqu'un seul concept d'un langage A a deux équivalents ou plus dans un langage B. On distingue alors l'équivalence par intersection (Figure 3) de l'équivalence par cumul (Figure 4).

Dans l'exemple de la Figure 3, le concept « femmes médecins » est représenté sous forme pré-coordonnée dans le langage documentaire A, mais pas dans le langage documentaire B. Pour obtenir une équivalence conceptuelle, il faut utiliser deux concepts constitutifs heureusement présents dans le langage B (« femmes » et « médecins »), à la jonction desquels se trouve l'aire sémantique couverte par le concept « femmes médecins » et le terme qui le représente.

Dans l'exemple de la Figure 4, le concept « cours d'eau » du langage documentaire A ne peut être représenté de manière équivalente que par le cumul des significations recouvertes par les concepts distincts « rivières » et « canaux » représentés dans le langage documentaire B.

Figure 4.
Exemple d'équivalence composée (par cumul)



Les thésaurus multilingues

La norme ISO 25964 accorde une attention considérable aux questions touchant le multilinguisme et le multiculturalisme dans les systèmes d'organisation de l'information et dans les langages documentaires. Le tableau qui présente les codes et symboles utilisés dans un thésaurus, par exemple, énumère déjà ceux-ci en une dizaine de langues naturelles distinctes et il est prévu que d'autres s'ajouteront au fil de l'adoption de la norme sur le plan international.

Mais c'est surtout la disparition de la norme distincte traitant des thésaurus multilingues (ISO 5964-1985) qui doit retenir l'attention. Quelques ajustements aux directives générales et l'ajout d'une clause dédiée à la relation d'équivalence entre différentes langues naturelles ont permis d'intégrer les particularités du thésaurus multilingue à la nouvelle norme. Ces modifications ont ainsi permis de légitimer le thésaurus multilingue, considéré jusque-là comme un langage documentaire distinct (et complexe) plutôt que comme un type particulier de thésaurus dont la caractéristique essentielle est de représenter les concepts d'un domaine ou d'une discipline au moyen de termes issus de diverses langues naturelles. C'est à la clause 9 de la Partie 1, « Équivalence entre langues naturelles », que sont présentés les quatre degrés d'équivalence possibles entre représentations d'un même concept dans deux langues naturelles (exacte, inexacte, partielle, non équivalence) ainsi que les solutions les plus efficaces au problème de non-équivalence. Les directives concernant les relations hiérarchiques et associatives dans les thésaurus multilingues, ainsi que l'affichage des termes et de leurs relations, ont été aisément intégrés aux clauses générales de la Partie 1 qui traitent de ces sujets.

On se souviendra que la norme britannique décrivait le thésaurus multilingue dans la section consacrée à l'interopérabilité ; dans la norme ISO 25964, on ne s'étonne donc pas de constater que la clause 9 de la

Partie 1 maintient de nombreux renvois vers la Partie 2 de la nouvelle norme.

Défis

Comme c'est le cas pour tout projet de cette envergure, la préparation de la norme ISO 25964 présente plusieurs défis. La première catégorie de difficultés à surmonter est liée à la gestion générale du projet. La langue des délibérations et de la rédaction est l'anglais, deuxième ou même troisième langue pour la majorité des membres du groupe de travail ; il faut reconnaître que cela complique la poursuite d'une « conversation » productive par courrier électronique ou par téléphone et les échanges sur un site de travail collaboratif. La gestion des nombreuses versions de chacun des documents de travail n'est pas simple ; il arrive trop régulièrement que plusieurs versions annotées d'un même document circulent simultanément. Il faut intégrer les suggestions, parfois contradictoires, venant des membres du groupe de travail, tous détenteurs d'une excellente connaissance et d'une expérience pratique considérable dans le domaine. Il faut réagir aux commentaires, parfois non pertinents ou inapplicables en contexte, des délégués votants de l'ISO ; la tâche est délicate, sachant le temps et l'énergie qu'ils ont consacré à la mise en forme de suggestions qui seront éventuellement rejetées. Et tout ce travail doit être accompli dans le respect d'échéanciers serrés.

La deuxième catégorie de difficultés et de défis, la plus importante, touche au contenu même de la norme, lequel doit être cohérent et témoigner clairement d'un consensus. Les décisions quant aux termes utilisés et à leur définition sont cruciales puisque ceux-ci seront réutilisés pendant plusieurs années dans les discussions et les écrits des spécialistes et des chercheurs. Une fois sélectionné et défini, le terme spécialisé doit toujours être utilisé avec la même signification. Il faut établir ce qui doit être conservé des normes ISO 2788 et ISO 5964 et ce qui ne sera pas repris dans la norme ISO 25964. Il faut constamment s'interroger sur la portée de la norme et développer en conséquence les éléments de contenus appropriés ; lorsque la décision a été prise d'intégrer un élément de contenu, il faut s'assurer de son insertion à l'endroit le plus approprié et déterminer le niveau de détail nécessaire à sa compréhension et à son utilisation. Il faut choisir et valider les exemples. Il faut revenir constamment sur les questions de multilinguisme et de multiculturalisme, sur lesquelles les spécialistes anglophones peuvent avoir des opinions divergentes de celles qu'expriment leurs collègues allemands, français, canadiens, belges et scandinaves pour qui l'égalité des langues en présence ne signifie pas l'identité de toutes les versions linguistiques d'un langage bilingue ou multilingue.

Conclusion

La Partie 1, Thésaurus, de la norme ISO 25964 a été approuvée et publiée en août 2011. La publication de la Partie 2, Interopérabilité avec les autres vocabulaires, est annoncée pour la fin de l'année 2012.

Pour le développement, la gestion et l'utilisation des langages documentaires, la norme ISO 25964 représente une combinaison fonctionnelle du traditionnel et du contemporain. Soutenue par les principes établis et les conventions dictées par les normes existantes, elle introduit des concepts et des directives qui la rendront utilisable en environnement réseauté.

Les normes applicables à la gestion et de la recherche d'information ne font pas l'objet de règles législatives au même titre que les normes sur la sécurité aérienne ou la mise en marché de médicaments par exemple, et leur adoption reste dépendante de leur qualité intrinsèque et de la volonté des intervenants concernés. Les normes dont l'application n'est pas obligatoire auront davantage de chances d'être adoptées si elles sont simples, répondent à un besoin réel, proposent des directives déjà acceptées par les communautés de pratique, peuvent s'appliquer sans réinvestissements majeurs, notamment au plan technologique, sont mises à disposition très largement et si possible gratuitement.

On doit déplorer que le travail remarquable réalisé au sein des organismes nationaux et internationaux de normalisation par des dizaines de chefs de projets dévoués et des centaines d'experts qui ne comptent pas leurs heures soit mal publicisé et peu connu en dehors des cercles d'initiés. Il est dommage que les projets et documents de travail soient protégés par des clauses de confidentialité et de propriété intellectuelle très strictes et que les normes elles-mêmes soient vendues à des prix qui n'en favorisent pas l'utilisation à grande échelle. Ce modèle de diffusion des normes, décrié récemment par Lynch (2011), doit être remis en question face à l'éclosion des normes du W3C, par exemple, mises à disposition très largement et gratuitement.

Les membres du groupe de travail TC46/SC9/WG8 sont conscients que la nouvelle norme ISO 25964 ne sera pas nécessairement accessible à l'ensemble des institutions documentaires et des petits organismes et entreprises qui en auraient pourtant besoin pour gérer plus efficacement leurs collections et exploiter les réservoirs d'information dont ils ne sont pas propriétaires. Le WG8, convaincu que les langages documentaires construits ou mis en relation dans le respect de la norme ISO 25964 seront plus utiles, plus adaptables et plus performants, a entrepris de publiciser la norme le plus largement possible et espère que des modalités d'accès mieux adaptées à ses utilisateurs potentiels en permettront l'adoption au cours des années qui viennent. ☉

Sources consultées

- Chichereau, Dominique, et al. 2007. Les normes de conception, gestion et maintenance de thésaurus : évolutions récentes et perspectives. *Documentaliste – Sciences de l'information* 44 (1) : 66-74.
- Dextre Clarke, Stella. 2008. ISO 2788 + ISO 5964 + Much Energy = ISO 25964. *Bulletin of the American Society for Information Science* 35 (1) : 31-33.
- Dextre Clarke, Stella, Alan Gilchrist and Leonard Will. 2004. Revision and extension of thesaurus standards. In *Knowledge organization and the global information society : Proceedings of the 8th International ISKO Conference, 13-16 July 2004, London UK*. Würzburg : Ergon. 215-219.
- Lynch, Clifford A. 2011. The case for new economic models to support standardization efforts [Page web]. Accessible à <http://www.niso.org/publications/white_papers/wp_lynch/>.
- Miles, A. et S. Bechhofer, éd. 2009. *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation [Page web] <<http://www.w3.org/TR/skos-reference>>.

Normes citées

- Association française de normalisation. 1981. *NF Z47-100 Règles d'établissement des thésaurus monolingues*. Paris : AFNOR.
- Association française de normalisation. 1991. *NF Z47-101 Règles d'établissement des thésaurus multilingues*. Paris : AFNOR.
- IFLA. Section on Classification and Indexing. Working Group on Guidelines for Multilingual Thesauri. *Guidelines for multilingual thesauri*. The Hague : IFLA, 2009. <<http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>>.
- National Information Standards Organisation. 2005. *Z 39-19 Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Baltimore, MD : NISO Press.
- Organisation internationale de normalisation. 1986. *ISO 2788 Documentation. Principes directeurs pour le développement et l'établissement de thésaurus monolingues*. Genève : ISO.
- Organisation internationale de normalisation. 1985. *ISO 5964 Documentation. Principes directeurs pour le développement et l'établissement de thésaurus multilingues*. Genève : ISO.