



HAL
open science

Organisation des connaissances médicales : principes et recherche récentes

Marcin Trzmielewski, Claudio Gnoli

► **To cite this version:**

Marcin Trzmielewski, Claudio Gnoli. Organisation des connaissances médicales : principes et recherche récentes. ISTE. Pratiques d'information et connaissances en santé, pp.173-210, 2022. hal-03767360

HAL Id: hal-03767360

<https://hal.science/hal-03767360v1>

Submitted on 1 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Organisation des connaissances médicales : principes et recherches récentes

8.1. Introduction

L'objectif de cette contribution est d'évaluer les dynamiques et les systèmes d'organisation des connaissances médicales développés par les spécialistes de l'information et de la documentation et d'expliquer les tendances et les évolutions spécifiques de ce domaine depuis les années 1960. L'organisation des connaissances est un champ de recherche réunissant des chercheurs issus de disciplines telles que les sciences de l'information et de la documentation, la linguistique, la philosophie et l'informatique [GNO 20, ZAC 10]. Après la clarification des concepts-clés et de la méthodologie d'étude, nous procéderons à une revue de littérature, combinée avec l'analyse d'un corpus de systèmes d'organisation des connaissances en médecine. Un bilan, soulignant les directions actuelles et expliquant le positionnement de la discipline, terminera la réflexion.

8.2. Sciences de l'information et de la documentation et organisation des connaissances

Les sciences de l'information, de la documentation et des bibliothèques (avec ses équivalents disciplinaires anglo-saxons *Documentation et Library and Information Science*) forment un champ de recherche interdisciplinaire rassemblant la

1. Marcin TRZMIELEWSKI et Claudio GNOLI

bibliothéconomie (*Librarianship*) et les sciences de l'information (*Information Science*), qui se situe entre les sciences humaines et sociales (SHS) et l'informatique [HJØ 18]. Par ailleurs, la tendance actuelle est d'utiliser les termes *Library and Information Science* et *Information Science* comme synonymes [HJØ 18]. Les réflexions sur les notions théoriques et pratiques, développées par les spécialistes de ce domaine, c'est-à-dire les enseignants-chercheurs et les professionnels de l'information-documentation (les bibliothécaires, documentalistes, catalogueurs, taxonomistes, etc.) se concentrent autour de « l'accès au savoir et à la culture » et du « savoir enregistré » [MET 06, p. 43]. L'un des concepts fondamentaux et l'une des missions principales des sciences de l'information et de la documentation est d'étudier et d'envisager la *médiation documentaire*. Comme le souligne Gérard Régimbeau, « la médiation documentaire est une médiation des savoirs » [RÉG 11, p. 108]. Elle se manifeste par les interactions entre le professionnel de l'information-documentation et l'utilisateur d'un système d'information et « nécessite la mise en place d'un système de communication adapté à l'utilisateur » [*Ibid.*, p. 108-109]. L'enjeu de ce processus est de « permettre à l'utilisateur de trouver l'information dont il a besoin » afin qu'il « bâtisse ses propres connaissances en vue de se constituer un savoir » [*Ibid.*, p. 106-109]. Les interactions entre le producteur de l'information, le médiateur et l'utilisateur sont mises en œuvre au sein des *dispositifs documentaires*, qui sont des « composantes instrumentales » au sein des systèmes d'information [MET 06, p. 54], responsables de la diffusion des connaissances entre ces acteurs, dans le but de l'appropriation de l'information par l'utilisateur comme contenu permettant d'agir [COU 11]. Les dispositifs sont dotés de trois dimensions : technique, socio-organisationnelle et socio-économique, qui sont indissociables et interdépendantes [MET 06]. Pour permettre l'accès, la consultation et la gestion de l'information dans les systèmes d'information, et pour mettre en place la médiation au sein des dispositifs documentaires, les professionnels, les utilisateurs et les programmes informatiques ont recours à l'*organisation des connaissances* [COU 11, GNO 20, HJØ 16, RÉG 11]. Les deux aspects principaux de cette activité sont : les processus d'organisation des connaissances (catalogage, indexation, marquage social, classement, ...) et les systèmes d'organisation des connaissances (règles et formats de description bibliographique, taxonomies, folksonomies, classifications, thésaurus, réseaux sémantiques, ontologies, ...) [HJØ 16, MAZ 18]. Les *systèmes d'organisation des connaissances* (SOC) sont des schémas qui organisent les concepts (représentés par les termes qui proviennent d'un usage courant des mots) et les relations entre ces concepts pour refléter toutes les connaissances humaines ou relatives à un domaine de connaissance [GNO 16, p. 403]. Par *domaine de connaissance*, nous désignons « un groupe avec une base ontologique qui révèle une téléologie sous-jacente, un ensemble d'hypothèses communes, un consensus épistémologique sur les approches méthodologiques et une sémantique sociale » (nous traduisons) [SMI 12, p. 114]. Les spécialités médicales, telles que

l'allergologie ou la pédiatrie, peuvent être considérées comme de tels domaines. La Classification internationale des maladies (CIM) et le thésaurus Medical Subject Headings (MeSH) sont des exemples de systèmes d'organisation des connaissances médicales (SOCM), qui répertorient et organisent les concepts relevant de domaines médicaux.

8.3. Méthodologie

La recherche documentaire d'articles originaux et d'actes de colloques, évalués par les pairs, a été effectuée le 25 mars 2020 et constitue notre corpus. Nous avons d'abord parcouru manuellement l'archive de la revue *Knowledge Organization* (1974-2019) et les actes de conférences *Advances in Knowledge Organization* (1990-2018), publiés par l'International Society for Knowledge Organization (ISKO). En raison d'un nombre modeste d'articles pertinents trouvés dans l'archive (34 travaux retenus), nous avons étendu la recherche en consultant la base de données *Knowledge organization literature*. Le moteur de recherche dans la base est disponible sur le site officiel de l'ISKO². La base contient une bibliographie relative à l'organisation des connaissances, provenant de revues en sciences de l'information, en linguistique et en informatique. Pour interroger la base, nous utilisons les termes « health » et « medicine ». Les requêtes répondent avec 135 résultats. Après avoir enlevé les doublons et les résultats non pertinents à notre démarche, nous retenons 25 travaux. Pour compléter et varier notre échantillon, nous nous sommes rendus sur *LISTA (Library, Information Science & Technology Abstracts)*, une base de résumés des publications en sciences de l'information et de la technologie, éditée par EBSCO. Les termes de la requête utilisés pour interroger le champ « title », sont composés d'adjectifs « health » et « medical », combinés avec les noms de différents types de SOC [MAZ 18] : « (TI health OR TI medical) AND (TI vocabular* OR TI gazetter+ OR TI dictionnar* OR TI glossar* OR TI terminolog* OR TI nomenclature* OR TI taxonom* OR TI classification+ OR TI subject headings OR TI thesaur* OR TI folksonom* OR TI conceptual scheme+ OR TI ontolog* OR TI semantic network+ OR TI knowledge graph+ OR TI knowledge organization) ». Nous avons filtré les résultats par articles de revues scientifiques évalués par les pairs et nous obtenons 122 publications. Après la sélection des travaux pertinents et l'élimination des doublons, nous avons retenu 10 articles. Pour compléter notre échantillon, nous avons intégré également 2 articles obtenus lors de

2. ISKO, « KO literature », disponible à l'adresse : <https://www.isko.org/lit.html>, 26 juin 2020.

notre recherche documentaire dans un moteur de recherche généraliste. Le corpus final se monte à 71 publications (cf. Annexe).

Nous avons analysé, évalué et synthétisé le contenu des documents du corpus, en utilisant deux grilles de lecture, établies à la suite des premières lectures exploratoires : la première est dédiée aux approches théoriques et aux évaluations, et la seconde porte sur les travaux de conception de SOCM. La lecture progressive du corpus selon une approche herméneutique a permis d'organiser les données, les informations et les connaissances en catégories et de les connecter entre elles pour découvrir des faits, des régularités, des relations et des phénomènes. Nous avons repéré les SOCM abordés dans les publications et avons collecté des informations les concernant. L'analyse des SOCM, avec l'aide de la seconde grille, a permis d'illustrer certains propos avec des exemples concrets. La consultation de quelques ressources supplémentaires a complété le rapport.

8.4. Présentation du corpus de publications entre 1960 et 2019

Parmi les 71 articles publiés en sciences de l'information et de la documentation entre 1968 et 2018, nous avons observé 11 publications parues entre les années 1960 et 1999 et 60 entre 2000 et 2019. L'augmentation significative des publications à partir de 2000 peut s'expliquer par le développement de l'Internet médical et par l'informatisation du secteur de la santé à grande échelle, qui ont eu lieu à cette époque [CAM 14, GON 07, p. 14-55 ; THO 04] et qui ont attiré l'attention des spécialistes de la discipline.

8.4.1. Les spécialistes travaillant sur l'organisation des connaissances en médecine

Parmi les auteurs de notre corpus, 70 % sont des chercheurs alors que 30 % proviennent du secteur professionnel. D'abord, nous recensons des spécialistes de l'information et de la documentation. En ce qui concerne les scientifiques, ils sont affiliés à des écoles, facultés, *colleges* ainsi qu'aux centres, départements et programmes, au sein d'universités spécialisées en *Archivistics*, *Library & Information Science*, *Information Science*, *Biblioteconomía y Documentación*, *Information and Documentation*, *Information Management*, *Communication & Arts*, *Communication Studies*, *Information and Media Studies* et en Sciences de l'information et de la communication. Du côté des professionnels, ce sont souvent des bibliothécaires, des documentalistes et des catalogueurs. Ils travaillent dans des bibliothèques nationales de santé ou dans des bibliothèques et centres de ressources spécialisés, qui s'adressent au grand public, à la communauté académique (instituts

supérieurs de santé, universités) ou aux professionnels des hôpitaux. Nous avons également trouvé des articles publiés par des salariés qui travaillent dans des sociétés privées fournissant des services de documentation. Ensuite, nous avons recensé des co-auteurs qui participent aux projets menés par les spécialistes de l'information et de la documentation. Ce sont des chercheurs attachés à des écoles, à des centres et départements d'informatique au sein des universités, ou affiliés aux facultés de médecine et aux instituts de recherche en épidémiologie et santé publique. Nous avons également identifié des informaticiens et des médecins qui travaillent dans des hôpitaux universitaires et dans des centres de soins.

La majorité des auteurs (58 %) provient des pays des deux Amériques : les Etats-Unis sont en tête, puis le Canada, le Brésil, le Mexique, et le Puerto Rico. La minorité (27 %) concerne les pays européens, en particulier les pays d'Europe de l'Ouest (France, Espagne, Royaume-Uni, Italie, Allemagne, Suisse) et d'Europe du Nord (Suède, Finlande, Norvège, Danemark). Ils sont suivis par les spécialistes des pays asiatiques (Inde, Japon, Singapour, Israël, Chine, Corée du Sud) (14 %) et de l'Australie (1 %). Aucun travail d'un pays africain n'a été recensé.

8.4.2. Les champs socio-documentaires étudiés

Les recherches portent d'abord sur les techniques et les SOCM pour l'organisation des ressources dans les organismes documentaires traditionnels : les bibliothèques nationales et publiques, les bibliothèques universitaires générales, les bibliothèques de santé et les centres de documentation [Textes 11, 13, 16, 37, 46].

A la suite de l'informatisation des catalogues de bibliothèques et de la constitution des bibliographies médicales dans les années 1970 [NAT 19], puis au développement du Web à partir de la deuxième moitié des années 1990 [GON 07], les spécialistes du domaine se sont intéressés à l'évaluation et à la conception des SOCM et des méthodes d'indexation dans ces environnements [Textes 3, 4, 6, 10, 19, 20, 21, 23, 25, 35, 41, 52, 55, 66].

Après l'informatisation des systèmes de santé dans les pays occidentaux, au tournant des années 1990 et 2000 [CAM 14, GON 07, p. 14-55 ; THO 04], l'attention des chercheurs s'est portée sur les établissements de soins. Il s'agit de réflexions sur l'indexation et sur la représentation des documents d'activité (notamment des dossiers patients) et des ressources informationnelles (des articles scientifiques) ainsi que de la conception d'ontologies pour le développement de bases de connaissances, essentielles pour les pratiques des professionnels de santé [Textes 2, 5, 8, 15, 17, 31, 48, 49].

La généralisation de l'information de santé sur l'Internet médical, au tournant des années 2000 et 2010 [ROM 08], a contribué à l'extension du terrain d'investigation à la webosphère. Désormais, les spécialistes étudient les méthodes d'organisation des connaissances médicales et les activités info-communicationnelles des usagers sur les sites web et sur les médias sociaux [1, 7, 9, 24, 28, 29, 30, 32, 34, 35, 43, 44, 47, 51, 56, 55, 61, 69, 70].

Récemment, les sciences de l'information et de la documentation se sont investies dans la recherche en santé publique et en ingénierie des connaissances médicales, en essayant d'extraire des connaissances à partir des informations massivement accessibles sur Internet (*big data*). Il s'agit d'études qui appliquent les techniques de découverte des connaissances basée sur la littérature scientifique (*literature-based discovery*), d'analyse des sentiments (*sentiment analysis*) et de visualisation de l'information (*dataviz*) dans le traitement des données provenant des médias sociaux, de la presse en ligne ou des bases de données bibliographiques [Textes 33, 39, 59, 61].

8.5. Les systèmes d'organisation des connaissances médicales

Les systèmes d'organisation des connaissances médicales (SOCM) sont des schémas constitués de concepts et de relations entre les concepts, représentant les connaissances relatives aux différents domaines de la médecine et aux domaines associés. Dans les SOCM, les concepts correspondent à des termes qui proviennent de la terminologie spécialisée (p. ex. « Rhinite allergique saisonnière ») et/ou profane (p. ex. « Rhume des foins »). Par ailleurs, comme l'atteste l'analyse des SOCM de notre corpus, les maladies sont classées souvent par référence au système du corps humain, par exemple : système cardiovasculaire, gastro-intestinal, respiratoire ou immunitaire. Les sciences de l'information et de la documentation interviennent principalement dans l'organisation des connaissances en médecine occidentale. Néanmoins, comme la médecine complémentaire et alternative gagne en popularité auprès du grand public, elle fait aussi l'objet de réflexions. Sont évoquées par exemple la médecine Ayurveda, la médecine Siddha, la médecine traditionnelle chinoise, l'aromathérapie et l'herboristerie [Textes 30, 31, 64, 67, 68].

8.5.1. La typologie et les usages

En prenant en considération le niveau de spécialisation des thèmes couverts par les SOCM et la visée des dispositifs pour laquelle ils sont implémentés (public général ou spécifique), nous pouvons distinguer d'une part les systèmes généraux, et d'autre part, les systèmes spécialisés. Les SOCM généraux ont pour objectif de

répertorier tous les savoirs humains, et consacrent une partie de leurs schémas aux concepts représentant les connaissances médicales. Ils sont utilisés plutôt dans les services et dispositifs à caractère multidisciplinaire, destinés à un large public, souvent non spécialisé. Les SOCM spécialisés ont vu le jour au moment où la croissance rapide du volume des connaissances médicales, l'évolution constante de la médecine et l'émergence de nouvelles spécialités, au tournant du XX^e siècle, ont nécessité le développement d'outils spécialisés, suffisamment détaillés pour satisfaire les besoins informationnels des usagers de l'information scientifique et technique [Textes 46, 50, 54]. Les systèmes spécialisés sont dédiés principalement aux phénomènes relevant de domaines médicaux, mais comprennent aussi des disciplines associées. Ils sont mobilisés dans les organisations, dans les systèmes et dans les services de documentation spécialisés en médecine et en santé, s'adressant principalement aux professionnels du secteur, mais aussi au grand public. Leur typologie n'a pas changé depuis les années 1960 [Texte 46] – il s'agit soit de systèmes spécialisés très généraux qui comprennent tous les domaines de sciences médicales, comme la National Library of Medicine (NLM) Classification et le Medical Subject Headings (MeSH), soit de SOCM dédiés à des domaines limités tels que le Thesaurus of Psychological Index Terms et la Colon Cancer Treatment Ontology. Les SOCM de domaine portent souvent sur des spécialités médicales, par exemple : obstétrique et néonatalogie [Texte 2], pharmacie [Texte 37], neurochirurgie [Texte 49], soins infirmiers [Texte 23], mais aussi sur des pathologies particulières : tumeur au cerveau [Texte 17], cancer du côlon [Textes 33, 45], maladies infectieuses et parasitaires [Texte 52], SIDA [Texte 27]. L'intérêt des concepteurs s'est également porté sur des thématiques importantes pour la santé publique, comme l'assurance maladie [Texte 11, 60], la santé et vieillissement [Texte 55] ou la nutrition [Texte 71].

L'analyse de notre corpus met en évidence l'usage de différents types de SOCM, parmi lesquels : les classifications générales, les classifications médicales, les nomenclatures médicales, les répertoires de vedettes-matière, les thésaurus médicaux, les réseaux sémantiques, les ontologies médicales, les glossaires, dictionnaires et répertoires de terminologies médicales, les schémas de catégorisation, les taxonomies médicales, les folksonomies et les vocabulaires dédiés aux consommateurs de soins de santé.

Les classifications générales peuvent être divisées en classifications énumératives (Classification décimale de Dewey, Classification décimale universelle, Classification de la Bibliothèque du Congrès) et en classifications à facettes (Bliss Bibliographic Classification, Colon Classification). Elles sont développées depuis le XIX^e siècle, pour classer les collections en médecine et en santé, dans les bibliothèques publiques, dans les bibliothèques universitaires

générales et aussi, mais plus rarement, dans les bibliothèques universitaires de santé qui adoptent le système de la bibliothèque centrale [Textes 11, 16, 37, 65].

Comme le soulignent les textes 13, 16, 37 et 65, les classifications bibliographiques médicales sont utilisées dans les bibliothèques de santé, situées dans les facultés de médecine ou dans les hôpitaux (NLM Classification), et dans les sections de santé, créées dans les bibliothèques publiques (Plantree Classification), pour le classement de documents en accès libre. Il est intéressant d'observer qu'entre les années 1950 et 1970, il existait une grande variété de classifications médicales dans les bibliothèques de santé (Classification Cunningham, Classification de Boston, Classification de Barnard, Army Medical Library Classification, ...) [Texte 46] alors qu'à présent, la NLM Classification et les classifications générales dominent significativement le secteur [Texte 65]. D'autres classifications médicales (Classification internationale des maladies – CIM, Manuel diagnostique et statistique des troubles mentaux, Classification commune des actes médicaux) et les nomenclatures (SNOMED-CT, RxNorm) sont mobilisées pour indexer les ressources contenues dans le dossier patient informatisé [Textes 2, 15, 48]. La CIM est par exemple mobilisée pour le codage de la morbidité permettant de générer les statistiques épidémiologiques à l'échelle mondiale, alors que la Classification commune des actes médicaux est utilisée en France pour les fins médico-économiques, notamment pour le codage des gestes médicotéchniques pratiqués par les professionnels de santé. Par ailleurs, les classifications médicales et les nomenclatures sont considérées comme des standards terminologiques qui assurent l'interopérabilité sémantique entre les entrepôts de données de santé. Elles deviennent par conséquent des composantes du dossier médical partagé, qui permet aux patients de partager les données les concernant avec les professionnels de santé³ [ODE 16, Texte 15].

Les répertoires de vedettes-matière (Library of Congress Subject Headings, RAMEAU, Répertoire de vedettes-matière de l'Université de Laval), conçus entre la fin du XIX^e et la seconde moitié du XX^e siècle, sont utilisés par les professionnels de la documentation pour décrire et indexer le contenu des documents dans les catalogues de bibliothèques publiques, universitaires générales et de santé [Textes 10, 41, 65].

Les thésaurus médicaux (MeSH, Emtree, Thesaurus of Psychological Index Terms), les réseaux sémantiques (UMLS), et plus récemment, les ontologies

3. WHO, « Classifications », disponible à l'adresse : <https://www.who.int/classifications/icd/en/>, 2020.

(OntoDip), sont développés et utilisés par les professionnels de l'information-documentation et par les programmes informatiques pour décrire et indexer le contenu des documents dans les catalogues de bibliothèques en ligne, dans les bases de données bibliographiques (p. ex. MEDLINE/PubMed) et dans les dispositifs de veille informationnelle [Textes 3, 4, 6, 10, 20, 23, 25, 32, 37, 51, 52, 53, 65, 66].

De plus, les ontologies, qui permettent de « modéliser conceptuellement un domaine et définir un vocabulaire partagé » ainsi que de « simuler le raisonnement humain » [WAL 18, p. 163], agissent comme des bases de connaissances, et sont utilisées pour la recherche d'information sémantique et pour l'extraction des connaissances à partir des dossiers électroniques des patients et de la littérature scientifique [Textes 17, 31, 33, 39, 49]. Comme l'attestent les travaux en informatique, les ontologies sont également mobilisées pour la gestion et la recherche de l'imagerie médicale [MAR 11].

Les spécialistes de l'information et de la documentation conçoivent également des glossaires, des dictionnaires (lexiques) et des répertoires de terminologies médicales [Textes 3, 44]. Ce sont des schémas, organisés souvent par ordre alphabétique, qui recensent des vocabulaires spécialisés ou profanes utilisés dans un domaine médical, qui peuvent servir de ressources terminologiques dans les systèmes de traitement de la langue fondés sur une approche linguistique ou pour la création d'ontologies.

Enfin, certains SOCM (schémas de catégorisation, taxonomies, vocabulaires dédiés aux consommateurs de soins de santé, ontologies) sont utilisés pour l'organisation et la recherche d'information spécialisée et grand public sur le Web. Un schéma de catégorisation est utilisé sur le site web du National Center for Complementary and Integrative Health⁴. La taxonomie à facettes Health Topics organisant des thèmes relatifs aux maladies et aux affections courantes médiatise les contenus sur MedlinePlus [Texte 51]. Les folksonomies sont des tags, mots-clés et mots-dièse créés par les utilisateurs des médias sociaux via le processus de marquage social (*social tagging*) [Textes 1, 18, 34, 35]. Les vocabulaires dédiés aux consommateurs de soins de santé (*Consumer Health Vocabularies – CHV*) sont des schémas utilisés pour représenter la terminologie non spécialisée, en prenant la structure de différents SOC (d'un répertoire terminologique, d'une classification, d'un thésaurus ou d'une ontologie), et sont utilisés par exemple dans les moteurs de recherche de services destinés au grand public [Textes 9, 44, 56]. Comme le

4. National Institutes of Health, « National Center for Complementary and Integrative Health », disponible à l'adresse : <https://www.nccih.nih.gov/>, 2020.

soulignent les travaux en informatique et en biostatistique, les *CHV* sont également utilisés dans les outils de fouille de texte lors des traitements automatiques des corpus de données issues des médias sociaux [NZA 15].

8.5.2. Les sources de la description

Les SOCM visent à représenter au premier chef le contenu des documents par l'expression de leur sujet à travers un concept, représenté par un terme, parfois accompagné des symboles alphanumériques (indices) ou graphiques (images). Le sujet peut concerner une discipline à laquelle le document se réfère : *L9F:4* « Gynecology » (Colon Classification), ou un phénomène discuté dans le document, par exemple une maladie : *EA80* « Atopic Eczema » (CIM-11) ou un traitement : *QW 900* « Desensitization, Immunologic » (NLM Classification). Nous pouvons donc les considérer comme des *modèles du contenu des documents* [ZAC 10, p. 148]. De plus, avec certains systèmes spécialisés, il est possible de modéliser des *situations extradocumentaires* [Ibid., p. 148]. Dans le MeSH, édité par la NLM, nous trouvons par exemple le descripteur « Caractéristiques d'une publication » [V] qui permet de décrire la forme de documents, notamment les « Catégories de publications » (p. ex. « Article de périodique », « Acte de congrès », « Dissertation universitaire »), les « Caractéristiques des études » (p. ex. « Etude clinique », « Méta-analyse », « Etude multicentrique ») et les « Composantes d'une publication » (p. ex. « Commentaire », « Annonce publicitaire », « Erratum publié »). La classification et le tri des publications en médecine selon ces critères a de l'importance pour la pratique de la médecine factuelle, où le type de publication et de méthodologie d'étude conditionne le niveau de preuve des travaux [HAS 13, HJØ 16]. Par ailleurs, dans Siddha Ontology, modélisée par Hemalata Iyer et K. S. Raghavan, nous relevons des concepts portant sur les institutions, les spécialistes du domaine ainsi que sur les types de collections et de documents : les ouvrages de référence, les revues, les articles et les manuscrits du domaine [Texte 31]. De plus, dans les schémas de SOCM, les usagers peuvent être regroupés en catégories sociologiques en fonction du type d'information ou de collection susceptibles de satisfaire leurs besoins et leurs pratiques info-communicationnelles. Par exemple, le schéma de catégorisation qui est utilisé sur le site web du National Center for Complementary and Integrative Health contient les catégories « Be an informed consumer » et « For Health Care Professionals » qui s'adressent aux types d'utilisateurs particuliers auxquels le service est dédié.

8.5.3. La conception de nouveaux SOCM

Notre corpus comporte 22 articles originaux portant sur la conception de SOCM. Les systèmes les plus souvent développés sont les ontologies (10 travaux) [Textes 2, 9, 17, 31, 33, 45, 49, 52, 60, 71] dans le cadre de travaux de chercheurs, puis les thésaurus (7 travaux) [6, 10, 23, 25, 27, 55, 67] et enfin les classifications (3 travaux) [11, 40, 50], développées et mises à jour plutôt par les professionnels de la documentation. Nous notons aussi la conception d'un cadre conceptuel de la médecine complémentaire et alternative pour établir des passerelles entre la tradition de la médecine occidentale et la tradition de la médecine orientale [Texte 68]. De plus, un travail propose le développement d'un répertoire de terminologie médicale profane [Texte 44].

8.5.3.1. Les concepteurs

Comme l'attestent les textes de notre corpus, les chercheurs en sciences de l'information et de la documentation développent le plus souvent des modèles conceptuels [Textes 2, 9, 17, 31, 33, 49, 50, 60, 67, 71]. Le développement de systèmes organisationnels prêts à l'implantation dans les systèmes infodocumentaires (par exemple dans un système intégré de gestion de bibliothèque ou un système de signalisation d'ouvrages dans les rayonnages de bibliothèque) est plutôt l'apanage des professionnels de l'information-documentation [Textes 6, 10, 13, 23, 25, 55].

Certains SOCM sont développés en dehors de la communauté de spécialistes de l'information et de la documentation. La CIM est, par exemple, révisée et maintenue par l'Organisation mondiale de la Santé en collaboration avec les experts de spécialités médicales. La SNOMED-CT est développée par une organisation à but non lucratif, SNOMED International, qui rassemble des spécialistes en technologies de l'information et de la communication, en architecture de l'information et en médecine. Par ailleurs, comme le témoignent les programmes de manifestations scientifiques autour de l'intelligence artificielle et de l'ingénierie des connaissances organisés en France et à l'international, notamment les Journées francophones d'Ingénierie des Connaissances⁵ (Atelier IA et Santé), la Plateforme de l'Intelligence Artificielle⁶ (Journée Santé & IA) et le Congress on Medical and Health Informatics [OHN 19], les laboratoires, les sociétés savantes et les consortiums d'informatique

5. IC 2016, « Programme – 6 juin : IA et Santé 2016 », disponible à l'adresse : <https://ic2016.sciencesconf.org/resource/page/id/27.html>, 2016.

6. PFIA 2020, « Journée Santé et IA », disponible à l'adresse : <http://pfia2020.fr/journeesthematiques/sante-et-ia/>, 2020.

et d'intelligence artificielle semblent s'installer durablement dans le champs de l'organisation des connaissances médicales. En observant les différentes sources de financement de ces évènements (les gouvernements, les communautés territoriales, les sociétés savantes, les universités, les organismes de recherche, les consortiums scientifiques, les sociétés privées et géants du marché des TIC), listées sur les sites web et mentionnées dans les travaux publiés dans les actes, il semble que l'informatique et l'intelligence artificielle ont trouvé un terrain d'application lucratif, auquel s'intéressent à la fois les acteurs privés et publics, puisque les enjeux économiques et sociétaux sont très importants. Ces financements leur permettent de conduire et de présenter les travaux de recherche sur les ontologies, les terminologies et les services associés. Ils s'appuient sur l'exploitation de corpus volumineux de textes et recourent à des techniques de traitement automatique des langues naturelles (TALN), en faisant l'hypothèse que ces corpus « fournissent les éléments stables, consensuels et partagés d'un domaine » [AUS 12, p. 252]. Probablement, avec la même question de financement de la recherche, depuis une dizaine d'années, nous pouvons noter une croissance significative de l'intérêt accordé aux ontologies médicales dans le champ de l'information et de la documentation. Cette tendance est intéressante puisque les ontologies, bien qu'elles aient déjà été conceptualisées en informatique par Gruber en 1993 [GRU 93], ne sont même pas mentionnées dans le *Dictionnaire encyclopédique de l'information et de la documentation*, publié en 1997 par Serge Cacaly [CAC 97]. Comme le montrent les publications de notre corpus, les spécialistes de l'information et de la documentation sont à présent très investis dans le développement et la recherche sur les ontologies médicales [Textes 2, 9, 17, 31, 33, 45, 49, 52, 60, 71]. Par ailleurs, les textes 2, 9, 52 et 60 confirment le financement de projets d'ontologies menés par des chercheurs en sciences de l'information, par les ministères de l'éducation, par les organisations de santé publique, par les agences du développement scientifique et les associations scientifiques.

8.5.3.2. Les méthodes de conception

L'élaboration des SOCM consiste à repérer différents éléments pertinents pour un domaine : les concepts, les termes, les situations, les objectifs et les enjeux, via l'analyse de différents corpus textuels. Pour cela, différents types de documents sont utilisés : ouvrages, résumés, articles scientifiques ou dédiés au public général, documents cliniques, messages de médias sociaux, sites Web de sociétés savantes, historiques des évènements, descripteurs de documents indexés, SOCM et données d'enquêtes [Textes 2, 9, 10, 13, 17, 23, 25, 27, 31, 33, 40, 44, 45, 49, 50, 55, 56, 60, 67, 68]. Récemment, des auteurs ont également eu recours à la méthode de production participative par des non-experts [Texte 71]. Par ailleurs, comme le soulignent les textes 49 et 56, des questions éthiques doivent être posées lors de la conception de SOCM, notamment lorsqu'il s'agit d'extraire des termes à partir des

contenus de dossiers patients informatisés et de médias sociaux, comprenant des données à caractère personnel. Après la collecte des éléments pertinents pour un domaine donné, les concepts et les termes qui les expriment sont ensuite normalisés et organisés par les professionnels ou par la machine en systèmes de relations sémantiques.

Les méthodes de conception de SOCM par les spécialistes de l'information et de la documentation peuvent être attribuées à cinq différentes perspectives épistémologiques : pragmatisme, historicisme et herméneutique, socioconstructivisme, rationalisme et empirisme [GNO 20, p. 39-40 ; HJØ 03, p. 107 ; HUB 98, WEI 16].

8.5.3.2.1. La méthode pragmatique

L'analyse des objectifs, des valeurs et des orientations applicatives d'une communauté, qui orientent la production et l'usage des genres de discours dans un domaine (comptes rendus d'hospitalisation, articles scientifiques, messages sur les médias sociaux, ...), est la première méthode, attachée au pragmatisme [GNO 20, p. 39]. Elle consiste en l'étude d'usages de l'information de santé, vus comme des pratiques sociales, communicationnelles et culturelles [Texte 58]. Il s'agit le plus souvent d'utiliser les catégories de recherche d'information, dégagées via des études du comportement informationnel des usagers [Texte 57]. Hemalata Iyer conduit par exemple des entretiens avec des praticiens et des éducateurs de la médecine Siddha pour identifier les principes, les pratiques professionnelles et informationnelles, les modèles de communication et les structures de connaissances développées pour transposer ensuite ces informations dans un modèle d'ontologie de domaine [Texte 31].

8.5.3.2.2. La méthode historiciste et herméneutique

La deuxième méthode s'inscrit dans la perspective historiciste ou herméneutique. Il s'agit d'étudier le contexte et le développement des connaissances et de la division sociale et scientifique du travail et de catégoriser les concepts d'un point de vue culturel [GNO 20, p. 39 ; HJØ 03, p. 107]. La Classification décimale universelle, la Bliss Bibliographic Classification et la NLM Classification sont ainsi conçues et révisées par les bibliothécaires et les documentalistes d'Europe et d'Amérique, qui organisent les connaissances médicales du point de vue de la médecine occidentale. En conséquence, les domaines de la médecine complémentaire et alternative, qui relèvent souvent de traditions orientales, sont pauvrement représentés [Textes 30, 64]. Il semble par ailleurs que les SOCM construits dans les cultures orientales sont également influencés par la médecine occidentale. La Colon Classification par

exemple, qui est révisée en Inde, n'est pas suffisamment développée pour couvrir les collections sur la médecine indienne Ayurveda [Texte 30].

8.5.3.2.3. La méthode socioconstructiviste

La troisième méthode, relevant du socioconstructivisme, consiste en la modélisation des schémas de classification qui reflètent les connaissances influencées par les constructions sociales, développées via des actions mutuelles entre les individus et leur environnement [HUB 98, WEI 16]. Ces constructions, qui dérivent des expériences individuelles et des perspectives sociales (formées par exemple par des idéologies sociales ou des intérêts politiques), stimuleraient la représentation d'une réalité, qui peut être parfois déconnectée de raisonnement scientifique [HUB 98, p. 195-196]. Dans le domaine du VIH/SIDA, par exemple, les maladies et les phénomènes associés coexistent ensemble dans une telle réalité, car la catégorisation, la marginalisation et la stigmatisation des populations affectées par le VIH (les gays, les afro-américains, les toxicomanes, ...) ou de leurs comportements sexuels (considérés dans certains discours comme déviants ou immoraux) sont influencées par des stimuli sociaux [*Ibid.*, p. 195-201]. Selon une approche socioconstructiviste, les SOCM doivent refléter la complexité des réalités marquées par la coexistence des constructions sociales à côté des facteurs historiques et culturels. Pour développer un vocabulaire-médiateur du dialogue dans la communauté VIH/SIDA, étant une tentative d'accueillir une telle complexité, Jeffrey F. Huber et Mary L. Gillaspay s'intéressent aux différents modèles de communication et de production des connaissances (à la fois traditionnels et non traditionnels) ainsi qu'aux différents vocabulaires vernaculaires développés par divers acteurs (les cliniciens, les chercheurs, les patients, etc.). Compte tenu de la nature multidisciplinaire, interdisciplinaire et transdisciplinaire de ce champ, Huber et Gillaspay reposent leur vocabulaire sur les différents domaines : « [...] Epidemiology and Transmission; Education and Prevention; [...] Treatments, Therapies, and Medical Management of HIV Disease; Psychosocial and Religious Issues, Case Management; Legal, Ethical, Economic, and Political Aspects; [...] Fine Arts; Belles Lettres and Non-Fiction » auxquelles ils attribuent les concepts pertinents [Texte 27, p. 297]. Ils ont également recours aux facettes communes : « Age Ranges, Sexual Orientation, Gender, Stages of Infection, Ethnic Groups, Geographic Names, At-Risk Populations, Religious Faiths, Signs and Symptoms, and Special Populations », dotées d'un rôle syntaxique qui leur permet de s'associer aux catégories principales, et les compléter, si nécessaire [HUB 98, p. 203]. Il est intéressant d'observer que, bien que le travail relève d'une perspective socioconstructiviste, les concepteurs recourent finalement à l'analyse par facettes, ce qui permet de constater que leur méthode s'inscrit également dans une approche rationaliste, une autre méthode de classification qui sera présentée ci-après.

8.5.3.2.4. La méthode rationaliste

La quatrième méthode consiste à créer des divisions logiques considérées comme universelles [GNO 20, p. 39 ; HJØ 03, p. 107]. Cette méthode d'organisation des concepts s'appuie souvent sur la théorie d'analyse par facettes. Elle est considérée par Birger Hjørland, comme une expression du rationalisme [HJØ 14], même si d'autres auteurs, comme par exemple Joseph T. Tennis [TEN 08], y voient aussi des composantes du pragmatisme, car les catégories fondamentales utilisées pour organiser les connaissances (expliquées plus loin) se révèlent utiles pour le processus de classification, et les classifications qui s'appuient sur ces catégories permettent de gagner du temps pour le lecteur. Par ailleurs, même si la confusion entre le pragmatisme et le caractère pratique de cette méthode (*practicalism*) peut être reprochée, il nous semble que c'est avant tout l'analyse par facettes à partir des contenus échangés et utilisés dans un domaine (articles scientifiques, livres, dossiers patients, ...) pour une application concrète, qui permet de constater que celle-ci a des influences pragmatistes. Les principes de cette approche s'appuient sur une théorie générale de la classification conçue par le bibliothécaire indien Shiyali Ramamrita Ranganathan, et ces principes ont été introduits en Europe dans les années 1950 par le Classification Research Group [VIC 60]. A partir de l'examen d'un échantillon de littérature dans un domaine spécialisé, les concepts et les termes sont regroupés dans des facettes, qui remplissent généralement différents rôles sémantiques et syntaxiques. Dans les classifications à facettes, par exemple, les concepts sont attribués aux facettes de la médecine. Comme le souligne Vanda Broughton, dans la classe *H* « Physical anthropology, human biology, health sciences » de la deuxième édition de la Bliss Bibliographic Classification, les termes sont attribués aux différentes catégories :

- (Thing) Human beings ;
- (Kind) Women, children, old people, etc. ;
- (Part) Head, legs, muscles, bones, heart, brain, lungs, etc. ;
- (Process) Respiration, digestion, reproduction, disease, etc. ;
- (Operation) Surgery, drug therapy, physiotherapy, etc. ;
- (Agent) Doctors, nurses, equipment, buildings, etc. [BRO 15, p. 308].

Ainsi, les termes de chaque catégorie constituent une facette d'un sujet. De cette façon, dans la Bliss, tous les termes de la catégorie « Part » deviennent par exemple la facette de la classe *HDP* « Anatomy » ou de la *HPT* « PARTS, ORGANS, SYSTEMS OF THE BODY », alors que les termes de la catégorie « Operation » deviennent la facette de la classe *HNP* « Treatment, therapy » [*Ibid.*]. Dans la même

approche, K. S. Raghavan et C. Sajana proposent des facettes du domaine, regroupées en trois facettes majeures (qui font référence aux catégories de Ranganathan) pour développer une ontologie de la neurochirurgie dédiée aux professionnels de santé : « Personality (Patients, Body Organs, Health Care personnel, Institutions, Hospital Managers), Matter-Property (Diseases, Symptoms, After effects), Energy (Diagnosis, Treatment) » [Texte 49, p. 210]. Comme l'ontologie a pour mission d'assurer un accès rapide aux documents cliniques, Raghavan et Sajana élaborent ces catégories et leur associent les concepts à partir des dossiers patients produits et utilisés par les professionnels dans leurs pratiques professionnelles. Le fait de mobiliser les contenus informationnels, utilisés par les acteurs dans leur pratique quotidienne, pour concevoir un SOCM, permet de constater que cette approche a, en effet, des aspirations pragmatistes.

8.5.3.2.5. La méthode empirique

La cinquième méthode, relevant de l'empirisme, se manifeste par des généralisations statistiques, basées sur la « similarité » d'objets informationnels qui partagent un grand nombre de propriétés en commun [GNO 20, p. 39 ; HJØ 03, p. 107]. Comme l'attestent les textes 2, 9, 44, la méthode empirique consiste souvent à analyser des corpus textuels très volumineux via des techniques de TALN pour établir des relations sémantiques entre les termes des ontologies, ou via des techniques d'apprentissage automatique (*machine learning*) pour établir des liens entre les termes et les définitions de terminologies. Une étude exemplaire de cette approche est le travail de Yunseon Choi, qui porte sur la construction d'une ontologie dédiée aux consommateurs de soins de santé. Elle repère un vocabulaire non expert via une analyse sémantique latente des tags sociaux, attribués par les utilisateurs de Del.icio.us (un site de marque-page social) pour indexer et échanger les documents [Texte 9]. Il est intéressant d'observer que même si la neutralité de la technique statistique émane de l'empirisme, la prise en compte de la terminologie profane pour que l'ontologie satisfasse au mieux les besoins des usagers, liés à la recherche de ressources informationnelles pertinentes, relève selon nous d'une influence pragmatiste. Par ailleurs, le fait de développer une ontologie représentant une vue non experte sur le domaine, à partir d'un langage profane construit par une communauté en ligne pour indexer les documents, témoigne également d'aspirations socioconstructivistes.

Les auteurs déclarent rarement les cadres théoriques sur lesquels reposent leur travail. Une explication réside dans le caractère récent de l'assise épistémologique des SOC, développée avec la formalisation du champ de l'organisation des connaissances [GNO 20, HJØ 18].

8.5.3.3. Les méthodes d'essai, de validation et d'évaluation

Les organisations de connaissances, proposées par les spécialistes de l'information et de la documentation, sont testées, validées et évaluées de façon qualitative et quantitative. Dans la méthode qualitative, qui est le plus souvent pratiquée, il s'agit d'évaluer l'utilité et l'efficacité des SOCM pour représenter les termes relevant des domaines ciblés ou de vérifier leur efficacité et leur potentialité pour intégrer les systèmes d'information auxquels ils sont dédiés [Textes 2, 9, 17, 31, 49, 50]. Les concepteurs valident les SOCM conçus en conduisant les entretiens semi-directifs avec les individus représentant les catégories d'utilisateurs futurs ou potentiels : les consommateurs de soins de santé [Texte 9] ou les spécialistes d'un domaine [Texte 2]. On leur demande, par exemple, d'indiquer dans un document les informations dont ils souhaiteraient disposer dans une situation précise et on vérifie si ces éléments sont correctement représentés dans le système évalué. Les SOCM sont également validés via les analyses des textes, effectuées par les experts, les professionnels de l'information-documentation ou par les machines [Textes 2, 50]. Pour valider l'ontologie *OntoNeo*, Mauricio Barcellos Almeida et Fernanda Farinelli utilisent les techniques de TALN pour extraire automatiquement la terminologie de l'obstétrique et néonatalogie à partir des dossiers cliniques, ce qui leur permet de vérifier si les termes les plus fréquents, obtenus à partir de l'extraction, sont présents dans l'ontologie [Texte 2]. Dans d'autres travaux, les concepteurs posent des requêtes pour tester l'efficacité et les fonctionnalités potentielles des systèmes d'information, basés sur les SOCM [Textes 17, 31, 49]. Les auteurs de l'ontologie du domaine de la médecine *Siddha* effectuent par exemple la recherche d'information sémantique dans la base de connaissances qui repose sur cette ontologie [Texte 31]. Les concepteurs de l'ontologie du domaine de la neurochirurgie, sur laquelle s'appuie un système d'aide à la décision basé sur l'indexation sémantique des dossiers des patients, posent à ce système des requêtes précises, comme par exemple : « identify & retrieve records of patients: age => 40 with astrocytoma » [Texte 49, p. 214]. L'évaluation quantitative, quant à elle, a pour but de mesurer le nombre de documents retrouvés via une requête posée au système d'information. À l'issue de cette opération, les taux de rappel et de précision sont calculés [Texte 60]. Par ailleurs, pour évaluer l'exactitude des déclarations dans l'ontologie de la nutrition, les sujets recrutés évaluent le taux d'exactitude des relations établies par les concepteurs [Texte 71].

8.5.3.4. Les nouveaux défis pour la conception de SOCM

Les classifications générales et spécialisées, dans lesquelles les connaissances sont organisées par discipline, se confrontent aujourd'hui au défi de l'interdisciplinarité des connaissances. Or, certains concepts interdisciplinaires, relevant de sciences médicales, peuvent apparaître dans plusieurs classes de ces

classifications, puisqu'ils sont étudiés par plusieurs disciplines. Dans la NLM Classification, par exemple, le concept de « tumeur » se trouve dans différentes disciplines : pathologie, ophtalmologie, gynécologie, soins infirmiers, etc. Une telle méthode de classification conduit à la dispersion des ressources qui est illustrée dans la localisation des documents (portant sur des sujets interdisciplinaires) dans plusieurs espaces du plan de classement d'une bibliothèque. En conséquence, on risque un éparpillement des documents sur différentes étagères d'une bibliothèque, et de cette façon, les usagers peuvent perdre une partie des ressources susceptibles de les intéresser. L'étude un peu datée (1996), mais avec une problématique tout à fait actuelle, de Lynn Silipigni Connaway et de MaryEllen C. Sievert, montre magistralement ce problème : à cause de la nature interdisciplinaire du concept d'« assurance maladie », les ressources sur ce thème, cataloguées dans des bases de données bibliographiques, sont dispersées dans 13 classes différentes de la NLM Classification et de la Classification décimale de Dewey ainsi que dans 23 classes de la Classification de la Bibliothèque du Congrès [Texte 11]. Nous verrons plus loin qu'une telle inconsistance dans les classifications peut créer des contraintes pour la représentation des connaissances dans les standards du Web sémantique.

Désormais, les spécialistes de l'organisation des connaissances recommandent de déplacer les SOC vers des applications en ligne afin d'optimiser l'indexation et la recherche d'information dans cet environnement [BIN 20]. Une telle volonté conduit à représenter l'organisation traditionnelle des concepts dans la syntaxe des technologies du Web sémantique. Alors que les ontologies médicales sont nativement créées dans ces standards, les SOC, dont les origines remontent au XIX^e siècle et aux prémices de la transformation numérique, doivent subir la conversion en systèmes d'organisation des connaissances en ligne (SOCL) – *Networked Knowledge Organization Systems (NKOS)*. En suivant ces recommandations, la Bibliothèque du Congrès et la NLM publient respectivement les Library of Congress Subject Headings et le MeSH en format *SKOS (Simple Knowledge Organization System)*, qui est un standard et un ensemble de spécifications développées pour supporter l'usage des SOC dans le cadre du Web sémantique [Texte 6, W3C 09]. Toutefois, à notre avis, faire reposer les classifications générales et spécialisées en médecine sur les disciplines peut devenir une contrainte technique considérable pour la conversion de ces classifications en SOCL. Or, pour la publication des classifications en format *SKOS*, le World Wide Web Consortium (W3C) recommande d'identifier les concepts de SOC par des *URI (Uniform Resource Identifiers)*, qui permettent d'identifier les concepts par une adresse unique pour leur description et leur mise en relation, en utilisant les langages *Resource Description Framework (RDF)* ou *RDF Schema (RDFS)*, basés sur le modèle de triplet « sujet – prédicat – objet » [W3C 09]. Cette spécification semble difficile à réaliser avec les classifications disciplinaires où certains concepts interdisciplinaires peuvent

apparaître plusieurs fois dans un schéma, et chaque fois dans une perspective différente. Un groupe des chercheurs de différents pays, qui travaille sur l'Integrative Levels Classification (ILC), une classification interdisciplinaire à facettes, propose une approche alternative : au lieu d'intégrer les concepts interdisciplinaires dans des classes disciplinaires différentes, il s'agirait de les représenter à travers des phénomènes, qui sont des unités basiques de la connaissance, indépendantes de toute discipline [GNO 16]. Le dernier travail de cette équipe, portant sur la conversion de l'ILC en format *SKOS*, souligne que l'approche de classification par phénomènes est tout à fait adaptée pour la représentation des classifications dans les standards du Web sémantique [BIN 20].

D'autres textes présentent les réflexions sur la conception et l'évaluation de SOCM et de formats d'échanges de données bibliographiques en vue d'améliorer les techniques de découverte des connaissances fondées sur la littérature scientifique [Textes 33, 39, 59]. Ils s'inscrivent dans la continuité des travaux initiés par Don R. Swanson sur l'établissement de nouvelles relations entre les connaissances contenues dans les articles publiés dans les domaines biomédicaux à travers des traitements engageant l'usage des technologies sémantiques. Cette idée peut être rapprochée de la philosophie du Web de données (*linked data*) qui est une infrastructure de métadonnées développée pour l'interopérabilité de référentiels afin de « regrouper l'information de manière utile, comme dans une gigantesque base de données où tout est écrit en langage structurée » dans le but de développer les différents services de recherche, de consultation, d'échange, d'annotation et d'indexation des contenus [WAL 18, p. 159]. Pour décrire les connaissances sur les ressources documentaires et les relations entre ces ressources, il faut avoir recours aux langages du Web sémantique que nous venons de mentionner (*RDF*, *RDFS*) et construire des magasins de triplets (*triplestores*), interrogeables avec le langage *SPARQL* [*Ibid.*, W3C 09].

Enfin, les concepteurs de SOCM sont actuellement amenés à organiser les informations dans un nouveau type de dispositifs médicaux, notamment dans les « organisations d'interface du secteur de la santé » [Texte 5]. Comme exemples français de tels dispositifs, nous pouvons évoquer le Dossier Médical Partagé, la Plateforme Territoriale d'appui ou les outils de télémédecine. Le but de leur développement est d'associer les secteurs public et privé, y compris les patients et leurs familles, pour qu'ils deviennent des espaces de coopération autour de connaissances partagées. Comme le souligne Christian Bourret dans le texte 5, de tels dispositifs ont besoin d'une nouvelle réflexion sur l'organisation et la représentation des connaissances, adaptées aux besoins et habiletés informationnels ainsi qu'aux capacités cognitives des communautés de savants, experts,

professionnels et profanes, dans l'esprit de construction d'une intelligence collective issue des connaissances de tous leurs membres.

8.6. Les nouveaux défis pour l'indexation et la recherche d'information

Une partie des travaux porte sur les nouveaux défis pour l'indexation et la recherche d'information médicale et de santé dans le cadre du Web social, de l'Internet médical et des bases de données bibliographiques.

La première catégorie de publications porte sur les pratiques d'indexation sociale [Textes 1, 9, 18, 28, 32, 34, 35], « permettant aux usagers non spécialistes d'indexer des ressources documentaires » sur le Web 2.0 [WAL 18, p. 161]. Dans la démarche netnographique, les spécialistes collectent et analysent les données avec les techniques qualitatives (observation, entretien, analyse par facettes, comparaison, mise en équivalence, TALN) et quantitatives (statistique descriptive, mesures infométriques et de similarité). Les résultats de ces travaux mettent en évidence une nature subjective de l'indexation sociale, effectuée par différentes catégories d'usagers sur des services en ligne. L'étude d'Hemalata Iyer et de Lucy Bungo montre, par exemple, que moins d'1 % des tags concernant la médecine complémentaire et alternative, posés par les utilisateurs d'une application web de catalogage social, correspondent à des termes du réseau sémantique UMLS [Texte 28]. Les données collectées par Margaret E. I. Kipp sur CiteULike, une application Web de catalogage bibliographique, comparées avec celles recueillies dans PubMed, soulignent qu'il existe des différences importantes dans les critères d'indexation de la littérature biomédicale entre les chercheurs en sciences biomédicales et les catalogueurs professionnels [Texte 34]. Marit Kristine Ådland et Marianne Lykke mettent en évidence que les utilisateurs et les éditeurs d'un site Web dédié au cancer ne sont pas d'accord à la fois sur le but des tags et sur ce qui caractérise un bon tag [Texte 1].

Un grand nombre de SOCM ont été conçus pour les professionnels de l'information-documentation afin de simplifier l'accès à l'information. Les travaux susmentionnés soulignent toutefois que la terminologie médicale utilisée par les usagers ainsi que leurs critères d'indexation peuvent être différents des terminologies et des critères pratiqués par les professionnels. Pertti Vakkari évalue par exemple dans quelle mesure le vocabulaire utilisé par les usagers des services numériques question-réponse sur le Web, concernant la santé, la nutrition et les services sociaux, est couvert par les systèmes généraux et spécialisés, développés par les professionnels de l'information-documentation. Les résultats varient entre 15 % et 65 % de convergence terminologique, ce qui montre une fracture sémantique entre les SOCM professionnels et le vocabulaire utilisé par les usagers

non spécialisés sur le Web [Texte 63]. Ces résultats font réfléchir quant aux outils d'organisation des connaissances développés dans une approche *top-down*, qui peuvent être incompatibles avec les comportements et les pratiques des usagers, ce qui peut avoir des conséquences négatives sur la démarche de recherche d'information de l'utilisateur et l'éloigner de l'information dont il a besoin. Des solutions pour remédier à ces divergences terminologiques et indexatoires sont le plus souvent cherchées dans la construction de ressources terminologiques et de vocabulaires dédiés aux consommateurs de soins de santé [Textes 9, 44, 56] et dans la prise en compte des préférences des usagers en matière d'indexation des documents et des critères de pertinence d'information [1, 14, 29, 34, 35]. Ce genre de connaissances sur les usagers peut être acquis, par exemple, via l'analyse de domaine (*domain analysis*), qui est un paradigme s'inspirant d'une approche sociologique de l'organisation des connaissances, selon laquelle l'indexation doit être considérée avec une perspective socio-culturelle, c'est-à-dire liée aux pratiques sociales des individus d'une communauté de discours [HJØ 16]. L'analyse de domaine rassemble ainsi aux deux approches herméneutique et pragmatiste [GNO 20].

La deuxième catégorie des travaux se nourrit d'une approche cognitive de l'organisation des connaissances [GNO 20] qui cherche à prendre en compte les capacités cognitives de l'utilisateur (situé dans une culture ou une communauté) lors de la recherche d'information [Textes 24, 29, 47, 51, 56]. Dans cette approche, Dagobert Soergel, Tony Tse et Laura Slaughter s'inspirent des travaux sur la cognition humaine menés en informatique biomédicale, qui soulignent que la compréhension des pathologies par les patients et les médecins peut être résumée en deux modèles distincts : *illness model*, manifesté chez les patients, et *disease model*, caractérisant la perception de médecins [Texte 56]. L'équipe de Soergel repose leur travail sur ce modèle et développe un *interpretive layer*, qui a pour objectif d'établir la médiation de l'information de santé entre la perspective des spécialistes et la perspective des non-experts. Le modèle s'attache à la conception d'un ensemble d'outils : une ontologie éducative (avec des définitions et terminologies non expertes), une base de connaissances biomédicales validées, une base d'explication des fausses croyances, et une interface d'affichage visuel et de navigation dans une classification. Tous ces outils forment une « couche interprétative » permettant à l'utilisateur non spécialisé de trouver le document en lien avec son besoin d'information et de comprendre son contenu. De plus, pour améliorer l'accès à l'information, certains concepteurs cherchent à représenter les connaissances par un design approprié aux capacités cognitives, associées à la culture informationnelle des groupes particuliers. Les travaux de Lisa M. Given *et al.*, et de Hemalata Iyer et Mary Guadrón, tentent par exemple d'identifier les stratégies d'organisation et de visualisation des connaissances sur les sites web de santé, destinées aux personnes

âgées [Textes 24, 29]. Kyong Eun Oh, Soohyung Joo et Eun-Ja Jeong recueillent les facettes qui sont préférées, faciles à utiliser, et utiles au public général pour accéder à l'information de santé en ligne [Texte 47].

La troisième catégorie d'études traite du phénomène de l'organisation des connaissances d'une manière participative (*crowdsourcing*) [Texte 66]. L'étude de James W. Woods *et al.*, fournit un retour d'expérience instructif de cette pratique, à l'exemple de l'indexation des images de lésions de peau, provenant de la collection de la NLM, effectuée par les dermatologues [Texte 66]. 9 experts ont passé en moyenne 30 minutes à décrire les images via un formulaire mis à leur disposition sur le Web. La comparaison du vocabulaire utilisé pour l'indexation, avec les termes d'UMLS montre que 300 termes de ce SOCM sur les 346 termes uniques soumis par les praticiens ont une équivalence exacte [Texte 66]. Cet exemple souligne un grand potentiel de la production participative, qui pourrait être utilisée pour l'indexation et l'annotation de grands corpus de documents d'activité clinique ou de ressources informationnelles en santé (p. ex. des fonds patrimoniaux). Le *crowdsourcing* pourrait constituer également une alternative intéressante aux techniques automatiques d'indexation et d'annotation.

La quatrième catégorie de travaux montre que le contrôle du vocabulaire et la structuration rigide de la terminologie dans les SOCM construits dans une approche *top-down*, par exemple des thésaurus médicaux, sont encore utiles par la recherche d'information scientifique. Ying-Hsang Liu et Nina Wacholder soulignent que les termes du MeSH sont indispensables pour mesurer le taux de précision pour les experts du domaine, même pour les étudiants diplômés dans les domaines biomédicaux qui n'ont pas suivi une formation spécifique en recherche d'information [Texte 36]. De plus, Xiangming Mu, Kun Lu et Hohyon Ryu, ayant conduit une étude auprès des étudiants non expérimentés en recherche d'information, inscrits au premier et deuxième cycle d'études supérieures, mettent en évidence qu'un système permettant la recherche par terme du MeSH et la navigation dans son schéma sont plus efficaces en matière de recherche d'information qu'un système proposant une interface de recherche simple [Texte 43]. Enfin, comme en témoignent de nombreuses thèses et articles en médecine, le MeSH est actuellement largement sollicité à la fois par les médecins-chercheurs et les étudiants en médecine pour interroger PubMed afin d'effectuer des revues de littérature systématiques.

8.7. Les enjeux sociaux, culturels, politiques et économiques

A côté des problématiques liées à l'indexation et la recherche d'information, l'organisation des connaissances médicales soulève également les enjeux sociaux, culturels, politiques et économiques.

Comme l'attestent les textes 12 et 26 de notre corpus, le fait d'ajouter et de supprimer les concepts dans un SOCM repose sur une évolution socio-politique des connaissances, et soulève des questions sensibles et même des tensions politiques, suscitées par des enjeux normatifs. En 1968, l'Association américaine de psychiatrie enregistre l'homosexualité dans le Manuel diagnostique et statistique des troubles mentaux (une classification des maladies mentales). Cet ajout provoque des protestations de militants homosexuels. Après un long débat, en 1973, le terme est supprimé de la classification [Textes 12, 26]. De plus, dans la neuvième version de la Classification internationale des maladies (CIM-9), utilisée jusqu'à la fin des années 1980, le concept d'homosexualité (302.0 « Homosexuality ») figure dans le schéma, en tant que la première subdivision de la classe 302 « Sexual deviations and disorders », et appartient au chapitre 300-316 « NEUROTIC DISORDERS, PERSONALITY DISORDERS, AND OTHER NONPSYCHOTIC MENTAL DISORDERS ». Le concept est accompagné d'une note d'usage à destination des professionnels de santé : « Codez l'homosexualité ici, que ce soit ou non considéré comme un trouble mental » (nous traduisons). Même si la classification n'a jamais explicitement déclaré que l'homosexualité devrait être considérée comme une maladie, l'analyse des concepts génériques, permettant de dévoiler le contexte dans les classifications [ZAC 10], a conduit à laisser courir des interprétations du discours médiatisé comme discriminant et erratique [DRI 94]. Dans la dixième version de la classification, publiée en 1990 et remplaçant la CIM-9, le terme n'apparaît plus en tant que motif de contact sexuel [DRI 94]. Les exemples évoqués ci-dessus illustrent clairement les enjeux normatifs des SOCM qui reflètent un pouvoir de certains acteurs de catégoriser ce qui relève du normal et ce qui relève du pathologique [WAL 18, p. 158]. La catégorisation des individus a un lien direct avec leur représentation dans une société. Si bien que l'on peut comprendre que cela devienne un outil de révolte des communautés qui ne s'identifient pas avec certaines catégorisations proposées par des instances normatives.

De plus, le principe de *cultural warrant*, qui invite à concevoir des SOC appropriés et utiles aux individus d'une culture, peut conduire à négliger les intérêts, les valeurs, les droits et les croyances des autres, les privant du droit d'accès à l'information [BEG 02]. Nous l'avons déjà souligné dans la partie consacrée aux méthodes de conception de SOCM que les domaines de médecine complémentaire et alternative sont faiblement représentés dans les classifications révisées par les

professionnels de l'information-documentation occidentaux [Textes 30, 64]. Les ressources documentaires portant sur les thèmes de la médecine orientale peuvent donc être moins visibles dans les collections d'une bibliothèque localisée en Europe. En conséquence, les adeptes de la médecine Siddha, par exemple, qui fréquentent cette bibliothèque, ont moins de chances de trouver l'information dont ils ont besoin.

Enfin, comme le soulignent les travaux en sciences de l'information et de la communication et en médecine, la conception de classifications médicales de référence soulève aussi des enjeux économiques, liés aux décisions politiques prises par des instances gouvernementales et des organisations intergouvernementales, responsables de leur révision [TAN 16, WAL 18, p. 158]. L'inscription ou l'exclusion de certaines maladies des classifications, telles que la CIM ou la Classification Commune des Actes Médicaux, connaît, par exemple, des répercussions sur la mise sur le marché de nouveaux médicaments par l'industrie pharmaceutique et sur le remboursement des actes médicaux par les caisses nationales de l'assurance maladie, les mutuelles de santé et les compagnies d'assurance. Comme illustration d'un tel pouvoir, nous pouvons évoquer un cas reporté par les spécialistes en allergologie, une spécialité récente qui est parfois incorrectement représentée dans les classifications médicales. L'absence de classe spécifique à l'anaphylaxie dans la CIM-10, due à la confusion de cette maladie avec le choc anaphylactique dans la classe *T78.2* « Choc anaphylactique, sans précision » (les termes « anaphylaxie » et « choc anaphylactique » ne représentent pas le même concept), et la dégradation du concept d'anaphylaxie à la classe *T78* « Effets indésirables, non classés ailleurs », a conduit à la sous-notification de la mortalité par anaphylaxie par les professionnels de santé au Brésil lors du codage des certificats de décès. Le manque de données épidémiologiques dans la base de données nationale brésilienne a considérablement affecté la prise de décision par les pouvoirs publics et les laboratoires pharmaceutiques, et en conséquence la non-mise sur le marché d'adrénaline auto-injectable pour le traitement de l'anaphylaxie [TAN 16].

Les exemples susmentionnés soulignent clairement que les SOCM représentent et imposent une certaine vision du monde à un instant donné et dans un contexte particulier [DUB 20, p. 12]. Ils peuvent également orienter les connaissances, donner de la visibilité à une pathologie, devenir des instruments de pouvoir, de révolte et d'accès à l'information ainsi que permettre ou empêcher la circulation des biens dans l'économie de la santé.

8.8. Discussion

Notre analyse nous amène au constat que les théories et les techniques d'organisation des connaissances médicales, développées par les spécialistes de l'information et de la documentation, ont été profondément affectées par les transformations numériques du champ de l'information de santé. Du côté technique, la croissance constante du volume des connaissances dans ce domaine a créé, par exemple, le besoin de SOCM spécialisés. Le paradigme de la médecine factuelle a influencé, quant à lui, les critères d'indexation de la littérature médicale. Le développement du Web 2.0 a contribué à l'émergence de la pratique d'indexation sociale et a créé la nécessité d'adapter les outils d'organisation des connaissances à cet environnement. De plus, le Web sémantique et l'idée du Web de données ont orienté la représentation des SOCM dans les nouveaux standards et langages informatiques. Enfin, les données massives ont attiré l'attention des chercheurs qui tentent d'en extraire des connaissances. Nous observons, en parallèle de ces aspects techniques, le développement de réflexions théoriques. Les apports épistémologiques et méthodologiques ainsi que les études ethnographiques et netnographiques sont la manifestation de cette nouvelle orientation, qui invite à mettre l'utilisateur au cœur de la réflexion. Il semble que ce sursaut soit une conséquence du caractère participatif et social du Web 2.0, qui a facilité l'obtention de données comportementales, dont l'analyse a conduit à la prise de conscience des besoins des usagers. D'abord, l'importance est accordée à l'utilisateur non spécialisé. Les terminologies médicales et les techniques d'indexation traditionnelles sont remises en question. Elles sont confrontées avec le vocabulaire, les préférences d'indexation et de navigation ainsi qu'avec les capacités cognitives des usagers. Ensuite, d'autres travaux se focalisent sur l'utilisateur spécialisé situé dans son contexte de travail. En témoignent les méthodes de conception de SOCM, qui prennent en compte différents vocabulaires vernaculaires construits au sein des domaines médicaux, ainsi que les besoins, les comportements et les pratiques info-communicationnelles des professionnels de santé. Enfin, l'intérêt des spécialistes se porte sur les différents enjeux de l'organisation des connaissances, une préoccupation qui est propre aux SHS.

Malgré la constitution d'un corpus issu de trois sources de données seulement, les tendances techniques et théoriques s'inscrivent dans les mutations globales qui affectent le champ de l'organisation des connaissances [SOS 19, p. 31-43 ; WAL 18, p. 155-164]. La conception à large échelle d'ontologies médicales pour les applications en ligne et pour les processus d'organisation, de gestion et d'extraction des connaissances est un exemple phare. Les concepteurs font évoluer les systèmes traditionnels pour qu'ils soient conformes aux standards du Web sémantique. La production participative se répand aussi, petit-à-petit, que ce soit pour la

construction de SOCM et pour l'indexation de ressources documentaires sur les thématiques médicales. De plus, les techniques de TALN, de clustering et de visualisation d'informations, sont de plus en plus pratiquées, ce qui peut traduire un rapprochement des champs de l'ingénierie des connaissances et de l'intelligence artificielle, compte tenu de la nature interdisciplinaire des sciences de l'information qui oscillent entre les SHS et l'informatique, mais aussi, tout simplement, comme un pas vers l'adaptation des techniques d'organisation des connaissances aux nouvelles tendances et exigences technologiques.

Par ailleurs, les techniques statistiques de classification, comme celle de l'apprentissage automatique (AA), sont d'ores et déjà utilisées pour développer des modèles prédictifs (algorithmes de classification), par exemple pour le diagnostic du cancer, sans recours aux outils classiques de l'organisation des connaissances [SID 19]. En conséquence, la pertinence des SOCM, de la description, de la classification et de l'indexation documentaire semblent être mise en question vis-à-vis des techniques d'AA telles que l'apprentissage profond, le réseau de neurones artificiels, etc. En revanche, Andrew MacFarlane, Sondess Missaoui et Sylwia Frankowska-Takhari soulignent que les outils d'organisation des connaissances ne devraient pas être négligés dans cette tâche [MAC 20]. En s'appuyant sur l'exemple des documents multimédia, les auteurs mettent en évidence que la mise en place des projets engageant les techniques d'apprentissage supervisées nécessite d'abord des connaissances sur les besoins et les pratiques informationnelles dans les domaines d'investigation (qui peuvent être médiatisés par les SOCM) afin d'identifier les éléments pertinents pour les acteurs dans les contenus de documents à « apprendre » par les machines. Ensuite, les algorithmes utilisés dans ce genre de travaux nécessitent des documents d'entraînement, dotés de métadonnées décrivant leurs caractéristiques « générales » et « contextuelles » (p. ex. les objets, les styles, les genres, les émotions, les endroits), qui sont difficiles à détecter par les algorithmes [MAC 20]. Selon le modèle de MacFarlane *et al.*, les corpus d'entraînement pour l'apprentissage automatique peuvent être constitués de documents et de métadonnées sujet y associées (contrôlées par les SOC), provenant par exemple de banques d'images. Ils peuvent être fusionnés par la suite avec des corpus d'images, dont les caractéristiques peuvent être extraites via les techniques de l'AA, pour constituer des index nécessaires pour la classification automatique. La pertinence de ce modèle pour la classification de l'information dans le domaine médical reste à évaluer.

Le positionnement des sciences de l'information et de la documentation vis-à-vis de l'organisation des connaissances médicales peut être expliqué par la prise en considération des SOC généraux et spécialisés et par l'analyse de différentes facettes des domaines médicaux. Cette approche permet de saisir les objets d'étude dans une

perspective holistique. Comme nous venons de le voir, les dimensions de la connaissance qui sont souvent investigués sont : la réalité, l'ontologie, l'épistémologie, les documents, les collections, les besoins d'information (connectés aux comportements informationnels et aux pratiques info-communicationnelles) et les capacités cognitives d'individus. Ces « étapes entre le monde et la classification », qui peuvent être identifiées dans les contenus des SOCM, sont mises en exergue par Claudio Gnoli à partir d'un schéma initial élaboré par Brian Vickery [GNO 20]. Elles peuvent être appliquées dans les domaines médicaux comme suit :

- α réalité (la totalité indifférenciée en amont de la compréhension humaine),
- β phénomène, p. ex. l'hypersensibilité allergique,
- γ perspective, p. ex. une approche intégrée,
- δ document, p. ex. un article dans une revue spécialisée qui en discute,
- ϵ collection, p. ex. une bibliothèque numérique qui contient l'article,
- ζ besoin d'information, p. ex. une recherche par thèse de spécialisation,
- η personnes, p. ex. un interne dans un hôpital universitaire,
- θ cognition, p. ex. ses capacités naturelles d'examiner des listes.

Les travaux inspirés par une approche pragmatiste, analysés dans cette étude, soulignent également que certains concepteurs de SOCM ressentent à présent le besoin de se dégager d'une approche technocentrée, qui fait la part belle aux outils implémentés dans le secteur de la santé [GRO 07], et de construire les dispositifs et les services numériques, en lien avec les pratiques des usagers. L'étude des domaines médicaux dans une approche holistique, propre aux sciences de l'information et de la documentation, paraît donc très favorable. Cette démarche peut déboucher sur la proposition de modèles de SOCM pour la conception de dispositifs de médiation, traduits par la suite en réalisations concrètes par les ingénieurs. Nous, en tant que chercheur en sciences de l'information et de la communication et professionnel de l'information-documentation, encourageons les spécialistes de la discipline à monter des projets pluridisciplinaires en humanités digitales, en partenariat avec le secteur de la santé, et en coopération avec les spécialistes en informatique.

8.9. Conclusion

Les travaux en sciences de l'information et de la documentation, portant sur l'organisation des connaissances médicales, se divisent en deux axes. Le premier invite à concevoir et évaluer les SOCM. Il se caractérise par des analyses de corpus

textuels, l'usage de techniques documentaires, de technologies sémantiques et par des études d'utilisateurs, de leurs activités info-communicationnelles et de leurs capacités cognitives. Le deuxième englobe des réflexions propres à une science des mentefacts⁷, notamment sur la théorie et l'épistémologie de l'organisation des connaissances et sur les enjeux liés à la classification. Dans le premier axe, il s'agit de développer les SOCM pour la conception de systèmes permettant aux utilisateurs la gestion, l'accès et la consultation de l'information. Dans le deuxième axe, le discours médical occupe une place prépondérante. Les spécialistes analysent le discours, porté par les ressources informationnelles et par les SOCM, et soulignent leur rôle médiateur des phénomènes politiques, scientifiques, sociaux, culturels et économiques.

Cette étude met en évidence que l'organisation des connaissances couvre un vaste spectre d'applications dans le domaine de la médecine et de la santé. Elle permet également d'expliquer pourquoi les SOCM, malgré leur nature artificielle et souvent contrôlée, sont capables d'attirer l'attention de la société sur le caractère sensible de la santé dans la population et sont des instruments politiques puissants.

8.10. Bibliographie

- [AUS 12] AUSSENAC-GILLES N., CHARLET J., REYNAUD C., « Les enjeux de l'Ingénierie des Connaissances », dans Sèdes F., Ogier J.-M., Marquis P. (dir.), *Information-Interaction-Intelligence : le point sur le I3*, p. 244-266, Editions Cépaduès, Toulouse, 2012.
- [BEG 02] BEGHTOL C., « A proposed ethical warrant for global knowledge representation and organization systems », *Journal of Documentation*, vol. 58, n° 5, p. 507-532, 2002.
- [BIN 20] BINDING C., GNOLI C., MERLI D., TRZMIELEWSKI M., TUDHOPE D., « Integrative Levels Classification as a networked KOS: a SKOS representation of ILC2 », *Advances in Knowledge Organization, Vol. 17 (2020) – Knowledge organization at the interface – Proceedings of the Seventeenth International ISKO Conference 6-8 July 2020 Aalborg, Denmark*, Ergon, Baden-Baden, 2020. [in press]
- [BRO 15] BROUGHTON V., *Essential Classification – Second Edition*, Facet Publishing, London, 2015.
- [CAC 97] CACALY S. (DIR.), *Dictionnaire encyclopédique de l'information et de la documentation*, Editions Nathan, Paris, 1997.

7. Par le concept de « mentefacts » nous désignons les produits intellectuels de la créativité (culture) humaine : les théories, les SOCM, les ouvrages, la pensée critique, etc. [GNO 19].

- [CAM 14] CAMPION-AWWAD O., HAYTON A., SMITH L., VUARAN M., *The National Programme for IT in the NHS – A Case History*, University of Cambridge, Cambridge, 2014.
- [COU 11] COUZINET V., « Les dispositifs : question documentaire », dans GARDIES C. (DIR.), *Approche de l'information documentaire : concepts fondateurs*, p. 117-130, Cepaduès, Toulouse, 2011.
- [DRI 94] van DRIMMELEN-KRABBE J.J., USTUN T.B., THOMPSON D.H., L'HOURS A., ORLEY J., SARTORIUS N., « Homosexuality in the International Classification of Diseases: A Clarification », *JAMA*, vol. 272, n° 21, p. 1660, 1994.
- [DUB 00] DUBRIDGE G., *Lost Books of Medieval China*, The British Library, London, 2000.
- [GNO 16] GNOLI C., « Classifying Phenomena, Part 1: Dimensions », *Knowledge Organization*, vol. 43, n° 6, p. 403-415, 2016.
- [GNO 19] GNOLI C., « Levels of information and library and information science as a science of mentefacts », *Information Research – Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia, June 16-19, 2019*, vol. 24, n° 4, 2019.
- [GNO 20] GNOLI C., *Introduction to knowledge organization*, Facet, London, 2020.
- [GON 07] GONOD BOISSIN F., L'usage de l'information numérique en médecine générale : étude exploratoire en Rhône-Alpes, Thèse de doctorat, Université Claude Bernard Lyon 1, 2007.
- [GRO 07] GROSJEAN S., BONNEVILLE L., « Logiques d'implantation des TIC dans le secteur de la santé », *Revue française de gestion*, vol. 3, n° 172, p. 145-157, 2007
- [GRU 93] GRUBER T.R., « Towards Principles for the Design of Ontologies Used for Knowledge Sharing », *International Journal Human-Computer Studies*, vol. 43, p. 907-928, 1993.
- [HAS 13] HAS – HAUTE AUTORITE DE SANTE, « Niveau de preuve et gradations des recommandations de bonne pratique – Etat des lieux », disponible à l'adresse : https://www.has-sante.fr/jcms/c_1600564/fr/niveau-de-preuve-et-gradation-des-recommandations-de-bonne-pratique-etat-des-lieux/, 14 juin 2013.
- [HJØ 03] HJØRLAND B., « Fundamentals for Knowledge Organization », *Knowledge Organization*, vol. 30, n° 2, p. 87-111, 2003.
- [HJØ 14] HJØRLAND B., « Is Facet Analysis Based on Rationalism? A Discussion of Satija (1992), Tennis (2008), Herre (2013), Mazzocchi (2013b), and Dousa & Ibekwe-SanJuan (2014) », *Knowledge Organization*, vol. 41, n° 5, p. 369-376, 2014.
- [HJØ 16] HJØRLAND B., « Knowledge organization », *Knowledge Organization*, vol. 43, n° 6, p. 475-484, 2016.
- [HJØ 18] HJØRLAND B., « Library and Information Science (LIS). Part 2 », *Knowledge Organization*, vol. 45, n° 4, p. 319-338, 2018.

- [HUB 98] HUBER J.T., GILLASPY M.L., « Social Constructs and Disease: Implications for a Controlled Vocabulary for HIV / AIDS », *Library trends*, vol. 47, n° 2, p. 190-208, 1998.
- [MAC 20] MACFARLANE A., MISSAOUI S., FRANKOWSKA-TAKHARI S., « On Machine Learning and Knowledge Organization in Multimedia Information Retrieval », *Knowledge Organization*, vol. 47, n° 1, p. 45-55, 2020.
- [MAR 11] MARAGOUDAKIS M., MAGLOGIANNIS I., « A medical ontology for intelligent web-based skin lesions image retrieval », *Health Informatics Journal*, vol. 17, n° 2, p. 140-157, 2011.
- [MAZ 18] MAZZOCCHI F., « Knowledge organization system (KOS) », *Knowledge Organization*, vol. 45, n° 1, p. 54-78, 2018.
- [MET 06] METZGER J.-P., « L'information-documentation », dans OLIVESI S. (DIR.), *Sciences de l'information et de la communication - Objets, savoirs, discipline*, p. 43-62, Presses universitaires de Grenoble, Grenoble, 2006.
- [NAT 19] NATIONAL LIBRARY OF MEDICINE, « A Brief History of NLM », disponible à l'adresse : <https://www.nlm.nih.gov/about/briefhistory.html/>, 20 mars 2019.
- [NZA 15] NZALI M.D.T., BRINGAY S., LAVERGNE C., OPITZ T., AZE J., MOLLEVI C., « Construction d'un vocabulaire patient/médecin dédié au cancer du sein à partir des médias sociaux », *26èmes Journées francophones d'Ingénierie des Connaissances*, Rennes, France, 1-3 juillet 2015, disponible à l'adresse : <https://hal.inria.fr/IC-2015/>, 2015.
- [ODE 16] ODEH S., « Du dossier médical personnel au dossier partagé : Vers un dispositif de médiation documentaire », *Les Cahiers du numérique*, vol. 12, n° 1, p. 31-50, 2016.
- [OHN 19] OHNO-MACHADO L., SÉROUSSI B. (DIR.), *MEDINFO 2019: Health and Wellbeing e-networks for All – Proceedings of the 17th World Congress on Medical and Health Informatics*, IOS Press BV, Amsterdam, 2019.
- [RÉG 11] REGIMBEAU G., « 3-Médiation », dans GARDIES C. (DIR.), *Approche de l'information documentation – Concepts fondateurs*, p. 75-114, Cépaduès, Toulouse, 2011.
- [ROM 08] ROMEYER H., « TIC et santé : entre information médicale et information de santé », *Tic en santé*, vol. 2, n° 1. Disponible à l'adresse : <https://journals.openedition.org/ticetsociete/365>, 2008.
- [SID 19] SIDEY-GIBBONS J., SIDEY-GIBBONS C., « Machine learning in medicine: a practical introduction », *BMC Medical Research Methodology*, vol. 19, 2019.
- [SMI 12] SMIRAGLIA R., « Epistemology of Domain Analysis », dans SMIRAGLIA R., LEE H. (DIR.), *Cultural Frames of Knowledge*, p. 111-124, Ergon, Würzburg, 2012.
- [SOS 19] SOSIŃSKA-KALATA B., « Organizacja wiedzy w cyfrowym świecie: wpływ rewolucji cyfrowej na ewolucję metod i narzędzi organizacji wiedzy », *Zagadnienia Informatyki Naukowej*, vol. 57, n° 1, p. 26-46, 2019.

- [TAN 16] TANNO L.K, CALDERON M.A., DEMOLY P., « New Allergic and Hypersensitivity Conditions Section in the International Classification of Diseases-11 », *Allergy Asthma & Immunology Research*, vol. 8, n° 4, p. 383-388, 2016.
- [TEN 08] TENNIS J.T., « Epistemology, theory, and methodology in knowledge organization: toward a classification, metatheory and research framework », *Knowledge Organization*, vol. 35, n° 2-3, p. 102-112, 2008.
- [THO 04] THOMPSON T.G., BRAILER D.J., *The Decade of Health Information Technology: Delivering Consumer-centric and Information-rich Health Care – Framework for Strategic Action*, U.S. Dept. of Health and Human Services, Office for the National Coordinator for Health Information Technology, Washington, D.C., 2004.
- [WEI 16] WEISS L.C., BRÄSCHER M., VIANNA W.B., « Pragmatism, Constructivism and Knowledge Organization », dans CHAVES GUIMARÃES J.A., OLIVIERA MILANI S., DODEBEI V. (DIR.), *Advances in Knowledge Organization, Vol. 15 (2016) – Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society – Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil*, 211-218, Ergon, Würzburg, 2016.
- [W3C 09] W3C, « SKOS: Simple Knowledge Organization System Reference », disponible à l'adresse : <http://www.w3.org/TR/skos-reference/>, 18 août 2009.
- [WAL 18] WALTER J., DOUYERE D., BOUILLON J.-L., OLLIVIER-YANIV C. (DIR.), *Dynamiques des recherches en sciences de l'information et de la communication – Conférence permanente des directeurs-trices des unités de recherche en sciences de l'information et de la communication (CPDirSIC)*, 2018. Disponible à l'adresse : <https://pastel.archives-ouvertes.fr/CREM-UL/hal-01885229v2>, 27 juin 2019.
- [VIC 60] VICKERY B., *Faceted classification: a guide to the construction and use of special schemes*. ASLIB, London, 1960.
- [ZAC 10] ZACKLAD M., « Evaluation des systèmes d'organisation des connaissances », *Les cahiers du numérique*, vol. 6, n° 3, p. 133-166, 2010.

8.11. Annexe 1 : le corpus d'articles de revues et d'actes de colloques

- Texte 1 – ÅDLAND M.K., LYKKE M., « Tags on healthcare information websites », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization, Vol. 16 (2018) – Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 684-692, Ergon, Baden-Baden, 2018.
- Texte 2 – ALMEIDA BARCELLOS M., FARINELLI F., « Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology », *Journal of the American Society for Information Science and Technology*, vol. 68, n° 11, 2529-2542, 2017.

- Texte 3** – DE ANDRADE J., LOPEZ GINES DE LARA M., « Interoperability and Mapping Between Knowledge Organization Systems: Metathesaurus—Unified Medical Language System of the National Library of Medicine », *Knowledge Organization*, vol. 43, n° 2, 107-112, 2016.
- Texte 4** – BOEHR D., BUSHMAN B., « Preparing for the Future: National Library of Medicine's® Project to Add MeSH® RDF URIs to its Bibliographic and Authority Records », *Cataloging & Classification Quarterly*, vol. 56, n° 2-3, 262-272, 2018.
- Texte 5** – BOURRET C., « New ways to produce shared knowledge to improve cooperation in overcoming societal challenges in healthcare: the lever of innovative interface organizations in France », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization, Vol. 16 (2018) – Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 871-879, Ergon, Baden-Baden, 2018.
- Texte 6** – BUSHMAN B., ANDERSON D., FU G., « Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of MeSH in RDF », *Journal of Library Metadata*, vol. 15, n° 3-4, 157-176, 2015.
- Texte 7** – CAMPBELL D.G., « Farradane's Relational Indexing and its Relationship to Hyperlinking in Alzheimer's Information », dans NEELAMEGHAN A., RAGHAVAN K.S. (DIR.), *Advances in Knowledge Organization, Vol. 12 (2012) – Categories, Contexts, and Relations in Knowledge Organization – Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*, 195-200, Ergon-Verlag, Würzburg, 2012.
- Texte 8** – CAMPBELL D.G., « The Human Life as Warrant: A Facet Analysis of Protocols for Dealing with Responsive Behaviours in Dementia Patients », dans BABIK W. (DIR.), *Advances in Knowledge Organization, vol. 14 (2014) – Knowledge organization in the 21st century: between historical patterns and future prospects – Proceedings of the Thirteenth International ISKO Conference 19-22 May 2013 Kraków, Poland*, 121-128, Ergon-Verlag, Würzburg, 2014.
- Texte 9** – CHOI Y., « Supporting better treatments for meeting health consumers' needs: extracting semantics in social data for representing a consumer health ontology », *Information Research*, vol. 21, n° 4, 2016.
- Texte 10** – COGNETTI G., POPPI G., « The problems of indexing in Italian biomedical libraries », dans NEGRINI G., FARNESI T., BENEDIKTSSON D. (DIR.), *Advances in Knowledge Organization, vol. 3 (1991) – Documentary Languages and Databases –Papers from the Rome Conference, Dec. 3-4, 1990*, 245-249, INDEKS Verlag, Frankfurt/Main, 1991.
- Texte 11** – CONNOWAY L.S., SIEVERT M.C., « Comparison of Three Classification Systems for Information on Health Insurance », *Cataloging & Classification Quarterly*, vol. 23, n° 2, 89-104, 1996.
- Texte 12** – COOPER R., « Diagnostic and Statistical Manual of Mental Disorders (DSM) », *Knowledge Organization*, vol. 44, n° 8, p. 668-676, 2017.

- Texte 13 – COSGROVE T.L., « Planetree health information services: public access to the health information people want », *Bulletin of the Medical Library Association*, vol. 82, n° 1, 57-63, 1994.
- Texte 14 – CRYSTAL A., GREENBERG J., « Relevance criteria identified by health information users during Web searches », *Journal of the American Society for Information Science and Technology*, vol. 57, n° 10, 1368-1382, 2006.
- Texte 15 – CUENCA G.M., SALVADOR OLIVÁN J.A., « Representación del conocimiento en historia clínica electrónica interoperable: el caso de la Historia Clínica Digital del Sistema Nacional de Salud de España », *Scire: Representación y organización del conocimiento*, vol. 23, n° 1, 25-38, 2017.
- Texte 16 – DALI K., DILEVKO J., « Towards Improved Collections in Medical Humanities: Fiction in Academic Health Libraries », *The Journal of Academic Librarianship*, vol. 32, n° 3, 259-273, 2006.
- Texte 17 – DAS S., ROY S., « Faceted Ontological Model for Brain Tumour Study », *Knowledge Organization*, vol. 43, n° 1, 3-12, 2016.
- Texte 18 – DEFRANCIS SUN B., « Folksonomy and health information access: how can social bookmarking assist seekers of online medical information? », *Journal of Hospital Librarianship*, vol. 8, n° 1, 119-126, 2008.
- Texte 19 – DOUYÈRE M., SOUALMIA L.F., NÉVÉOL A., ROGOZAN A., DAHAMNA B., LEROY J.-P., THIRION B., DARMONI S.J., « Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway », *Health Information and Libraries Journal*, vol. 21, n° 4, 253-261, 2004.
- Texte 20 – EFTHIMIADIS E.N., AFIFI M., « Population groups: indexing, coverage, and retrieval effectiveness of ethnically related health care issues in health sciences databases », *Bulletin of the Medical Library Association*, vol. 84, n° 3, 386-396, 1996.
- Texte 21 – FERRAIOLI L., « An exploratory study of metadata creation in a health care agency », *Cataloging & Classification Quarterly*, vol. 40, n° 3-4, 75-102, 2005.
- Texte 22 – FERNÁNDEZ L.M.M., IZQUIERDO ALONSO M., MAURANDI LÓPEZ A., VALLÉS VALENZUELA J., « Consistency between indexers in the LILAC database (Latin American and Caribbean Health Science Literature) », *Information Research*, vol. 18, n° 4, 2013.
- Texte 23 – FLOR P., JAKOBSSON A., MOGSET I., TAYLOR S., EASEN S.E., « A controlled vocabulary for nursing and allied health in Norway », *Health Information and Libraries Journal*, vol. 18, n° 1, 10-19, 2001.
- Texte 24 – GIVEN L.M., RUECKER S., SIMPSON H., SADLER E.B., RUSKIN A., « Inclusive interface design for seniors: image-browsing for a health information context », *Journal of the American Society for Information Science and Technology*, vol. 58, n° 11, 1610-1617, 2007.
- Texte 25 – HAMADA M., SAKUMA S., MISAWA K., « “Thesaurus for Medical and Health related Terms” and its relation to Medical Subject Headings », *Journal of Information Processing and Management*, vol. 51, n° 9, 642-652, 2008.

Texte 26 – HUBER J.T., GILLASPY M.L., « An Examination of the Discourse of Homosexuality As Reflected in Medical Vocabularies, Classificatory Structures, and Information Resources », dans BEGHTOL C., HOWARTH L.C., WILLIAMSON N.J. (DIR.), *Advances in Knowledge Organization*, vol. 7 (2000) – *Dynamism and Stability in Knowledge Organization – Proceedings of the Sixth International ISKO Conference 10-13 July 2000 Toronto, Canada*, 219-223, ERGON Verlag, Würzburg, 2000.

Texte 27 – HUBER J.T., GILLASPY M.L., « Controlled Vocabulary for HIV/AIDS: An Evolving Nosological: Record of a Diseased Body of Knowledge », dans GREEN R. (DIR.), *Advances in Knowledge Organization*, Vol. 5 (1996) – *Knowledge Organization and Change – Proceedings of the Fourth International ISKO Conference 15-18 July 1996 Washington, DC, USA*, 293-301, INDEKS Verlag, Frankfurt/Main, 1996.

Texte 28 – IYER H., BUNGO L., « An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine », *Information Research*, vol. 16, n° 3, 2011.

Texte 29 – IYER H., GUADRÓN M., « Older Adults and eHealth Literacy: Challenges to Knowledge Representation », dans BABIK W. (DIR.), *Advances in Knowledge Organization*, vol. 14 (2014) – *Knowledge organization in the 21st century: between historical patterns and future prospects – Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014 Kraków, Poland*, 429-436, Ergon, Würzburg, 2014.

Texte 30 – IYER H., « Alternative System of Medicine, Ayurveda: Challenges to Knowledge Organization and Representation », dans CHAVES GUIMARÃES J.A., OLIVIERA MILANI S., DODEBEI V. (DIR.), *Advances in Knowledge Organization*, Vol. 15 (2016) – *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society – Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil*, 384-391, Ergon, Würzburg, 2016.

Texte 31 – IYER H., RAGHAVAN K.S., « Medical ontology: Siddha System of Medicine », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization*, Vol. 16 (2018) – *Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 347-355, Ergon, Baden-Baden, 2018.

Texte 32 – JORENTE M.J.V., « Information Design as knowledge technology in the organization of digital information environments on the Zika Virus and its effects », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization*, Vol. 16 (2018) – *Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 820-829, Ergon, Baden-Baden, 2018.

Texte 33 – KHOO C.S.G., NA J.-C., WANG V.W., CHAN S., « Developing an Ontology for Encoding Disease Treatment Information in Medical Abstracts », *DESIDOC Journal of Library & Information Technology*, vol. 31, n° 31, 103-115, 2011.

- Texte 34 – KIPP M.E.I., « Tagging of Biomedical Articles on CiteULike: A Comparison of User, Author and Professional Indexing », *Knowledge Organization*, vol. 38, n° 3, 245-261, 2011.
- Texte 35 – LEE D.H., SCHLEYER T., « Social tagging is no substitute for controlled indexing: A comparison of Medical Subject Headings and CiteULike tags assigned to 231,388 papers », *Journal of the American Society for Information Science and Technology*, vol. 63, n° 9, 1747-1757, 2012.
- Texte 36 – LIU Y.-H., WACHOLDER N., « Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers », *Information Processing & Management*, vol. 53, n° 4, 851-870, 2017.
- Texte 37 – LÓPEZ-MERTZ E.M., « The Adequacy of the Structure of the National Library of Medicine Classification Scheme for Organizing Pharmacy Literature », *Library Resources & Technical Services*, vol. 41, n° 2, 123-135, 1997.
- Texte 38 – MARTELETO R.M., DOS SANTOS CARVALHO L., « Health as a Knowledge Domain and Social Field: Dialogues with Birger Hjørland and Pierre Bourdieu », *Knowledge Organization*, vol. 42, n° 8, 581-590, 2015.
- Texte 39 – MARCONDES C.H., DA COSTA L.C., « A Model to Represent and Process Scientific Knowledge in Biomedical Articles with Semantic Web Technologies », *Knowledge Organization*, vol. 43, n° 2, 86-101 2016.
- Texte 40 – MCLWAIN I.C., WILLIAMSON N.J., « Medicine and the UDC: The Process of Restructuring », dans ARSENAULT C., TENNIS J.T. (DIR.), *Advances in Knowledge Organization, Vol. 11 (2008) – Culture and Identity in Knowledge Organization – Proceedings of the Tenth International ISKO Conference 5-8 August 2008 Montréal, Canada*, 50-55, Ergon, Würzburg, 2008.
- Texte 41 – MCTAVISH J.R., FORTIER A., « A domain-analytic perspective on sexual health in LCSH and RVM », dans BEGHTOL C., SMIRAGLIA R. (DIR.), *Proceedings from North American Symposium on Knowledge Organization*, vol. 3, 83-93, Toronto, Canada, June 16-17, 2011.
- Texte 42 – MCTAVISH J.R., NEAL D.R., WATHEN C.N., « Is What You See What You Get? Medical Subject Headings and their Organizing Work in the Violence Against Women Research Literature », *Knowledge Organization*, vol. 38, n° 5, 381-397, 2011.
- Texte 43 – MU X., LU K., RYU H., « Explicitly integrating MeSH thesaurus help into health information retrieval systems: an empirical user study », *Information Processing & Management*, vol. 50, n° 1, 24-40, 2014.
- Texte 44 – MURESAN S., KLAVANS J.L., « Inducing terminologies from text: A case study for the consumer health domain », *Journal of the American Society for Information Science and Technology*, vol. 64, n° 4, 727-744, 2013.
- Texte 45 – NA J.-H., LENG NEOH H., « Effectiveness of UMLS semantic network as a seed ontology for building a medical domain ontology », *Aslib Proceedings*, vol. 60, n° 1, 32-46, 2008.

- Texte 46 – NICOLE-GENTY G., « Les classifications médicales », *Bulletin des bibliothèques de France*, vol. 13, n° 7, p. 289-313, 1968.
- Texte 47 – OH K., JOO S., JEONG E.-J., « Online consumer health information organization: users' perspectives on faceted navigation », *Knowledge Organization*, vol. 42, n° 3, 176-186, 2015.
- Texte 48 – PINTO V., RABELO C., GIRÃO I., « SNOMED-CT as Standard Language for Organization and Representation of the Information in Patient Records », *Knowledge Organization*, vol. 41, n° 4, 311-318, 2014.
- Texte 49 – RAGHAVAN K.S., SAJANA C., « NeurOn. Modeling ontology for neurosurgery », dans GNOLI C., MAZZOCCHI F. (DIR.), *Advances in Knowledge Organization, vol. 12 (2010) – Paradigms and conceptual systems in knowledge organization – Proceedings of the Eleventh International ISKO Conference 23-26 February 2010 Rome, Italy*, 208-215, Ergon, Würzburg, 2010.
- Texte 50 – RAJ P., « Depth Colon Classification Schedule for Anaesthesiology », *International Classification*, vol. 14, n° 2, 77-83, 1987.
- Texte 51 – REDALEN A., MILLER N., « Evaluating Website Modifications at the National Library of Medicine through Search Log Analysis », *D-Lib Magazine*, vol. 6, n° 1, 2000.
- Texte 52 – RIBEIRO C.J.S., DA SILVA D.R., « Knowledge Organization in the health field: an ontology project to improve the information retrieval process », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization, Vol. 16 (2018) – Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 330-338, Ergon, Baden-Baden, 2018.
- Texte 53 – SCHOPEN M., « Cross File Searching of Biomedical Databases at DIMDI », dans FUGMANN R. (DIR.), *Advances in Knowledge Organization, vol. 1 (1990) – Tools for Knowledge Organization and the Human Interface – Proceedings 1st International ISKO-Conference Darmstadt, Germany, 14-17 August 1990*, 231-237, INDEKS Verlag, Frankfurt/Main, 1990.
- Texte 54 – SEETHARAMA S., « Compatibility among Medical Classification Systems: A Case Study in the Classification of Cardiovascular Diseases », *International Classification*, vol. 12, n° 2, 80-86, 1985.
- Texte 55 – SMITH J.B., DEACON P., « Precision with ease: refining thesaurus support for quality health information searching on HealthInsite », *The Australian Library Journal*, vol. 58, n° 1, 47-62, 2009.
- Texte 56 – SOERGEL D., TSE T., SLAUGHTER L., « Helping Healthcare Consumers Understand: An “Interpretative Layer” for Finding and Making Sense of Medical Information », dans FIESCHI M., COIERA E., LI Y.-C. (DIR.), *MEDINFO Proceedings of the 11th World Congress on Medical Informatics*, 931-935, IOS Press, Amsterdam, 2004.

- Texte 57** – SOLOMON P., « On the Use of Research Categorizations as the Basis for Organizing Knowledge: A Test in the Domain of Information Behavior in Health Care », dans MUSTAFA EL-HADI W., MANIEZ J., POLLITT S.A. (DIR.), *Advances in Knowledge Organization, Vol. 6 (1998) – Structures and Relations in Knowledge Organization – Proceedings of the Fifth International ISKO Conference 25-29 August 1998 Lille, France*, 293-301, Ergon, Würzburg, 1998.
- Texte 58** – SUNDIN O., « Towards an Understanding of Symbolic Aspects of Professional Information: An Analysis of the Nursing Knowledge Domain », *Knowledge Organization*, vol. 30, n° 3/4, 170-181, 2003.
- Texte 59** – SWANSON D.R., SMALHEISER N.R., TORVIK V.I., « Ranking indirect connections in literature-based discovery: The role of medical subject headings », *Journal of the American Society for Information Science and Technology*, vol. 57, n° 11, 1427-1439, 2006.
- Texte 60** – TANG X., WEI W., LIU G., ZHU J., « An Inference Model of Medical Insurance Fraud Detection: Based on Ontology and SWRL », *Knowledge Organization*, vol. 44, n° 2, 84-96, 2017.
- Texte 61** – TANTI M., ROUX P., CARRIERI M.P., SPIRE B., « Exploiting the Knowledge Organization of Health 2.0 to Create Strategic Value in Public Health – An Example of Application to the Problem of Drug Consumption Rooms in France », dans CHAVES GUIMARÃES J.A., OLIVIERA MILANI S., DODEBEI V. (DIR.), *Advances in Knowledge Organization, Vol. 15 (2016) – Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society – Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil*, 530-537, Ergon, Würzburg, 2016.
- Texte 62** – TANTI M., « The French military documentary system to anticipate health risk: content and information classification », dans RIBEIRO F., CERVEIRA M.E. (DIR.), *Advances in Knowledge Organization, Vol. 16 (2018) – Challenges and Opportunities for Knowledge Organization in the Digital Age – Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, 940-944, Ergon, Baden-Baden, 2018.
- Texte 63** – VAKKARI P., « How specific thesauri and a general thesaurus cover lay persons' vocabularies concerning health, nutrition and social services », dans GNOLI C., MAZZOCCHI F. (DIR.), *Advances in Knowledge Organization, vol. 12 (2010) – Paradigms and conceptual systems in knowledge organization – Proceedings of the Eleventh International ISKO Conference 23-26 February 2010 Rome, Italy*, 299-307, Ergon, Würzburg, 2010.
- Texte 64** – WILLIAMSON N.J., « Complementary and alternative medicine: its place in the reorganized medical sciences in the Universal Decimal Classification », dans MCLWAIN I.C. (DIR.), *Advances in Knowledge Organization, vol. 9 (2004) – Global Organization and the Global Society – Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK*, 257-262, Ergon, Würzburg, 2004.

- Texte 65 – WOMACK K.R., « Conformity for Conformity's Sake? The Choice of a Classification System and a Subject Heading System in Academic Health Sciences Libraries », *Cataloging & Classification Quarterly*, vol. 42, n° 1, 93-115, 2006.
- Texte 66 – WOODS J.W., SNEIDERMAN C.A., HAMEED K., ACKERMAN M.J., HATTON C., « Using UMLS Metathesaurus Concepts to Describe Medical Images: Dermatology Vocabulary », *Computer in Biology and Medicine*, vol. 36, n° 1, 89-100, 2006.
- Texte 67 – YITZHAKI M., SHAHAR T., « A draft version of a consolidated thesaurus for the rapidly - growing field of alternative medicine », *66th IFLA Council and General Conference*, Jerusalem, Israel, 13-18 August 2000.
- Texte 68 – ZENG M.L., KRONENBERG F., MOLHOLT P., « Toward a Conceptual Framework for Complementary and Alternative Medicine: Challenges and Issues », *Knowledge Organization*, vol. 29, n° 1, 27-40, 2001.
- Texte 69 – ZHANG J., WOLFRAM D., WANG P., HONG Y., GILLIS R., « Visualization of health-subject analysis based on query term co-occurrences », *Journal of the American Society for Information Science and Technology*, vol. 59, n° 12, 1933-1947, 2008.
- Texte 70 – ZHANG J., WOLFRAM D., « Visual analysis of obesity-related query terms on HealthLink », *Online Information Review*, vol. 33, n° 1, 43-57, 2009.
- Texte 71 – ZHITOMIRSKY-GEFFET M., EREZ E.S., JUDIT B.-I., « Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health », *Journal of the American Society for Information Science and Technology*, vol. 68, n° 3, 681-694, 2017.

8.12. Annexe 2 : le corpus de SOCM

- Bliss Bibliographic Classification, 2nd ed. – BLISS CLASSIFICATION ASSOCIATION, « The Bliss Bibliographic Classification : schedules », disponible à l'adresse : <https://www.blissclassification.org.uk/bcsched.shtml/>, 28 décembre 2016.
- Classification Commune des Actes Médicaux – ASSURANCE MALADIE, « Classification Commune des Actes Médicaux », disponible à l'adresse : <https://www.ameli.fr/accueil-de-la-ccam/index.php>, 2020.
- Classification Internationale des maladies-10 – WHO, « CIM-10 Version :2008 », disponible à l'adresse : <https://icd.who.int/browse10/2008/fr>, 2008.
- Colon Classification – ISKO ITALIA, « Colon Classification - L Medicine », disponible à l'adresse : <http://www.iskoi.org/doc/colon.htm#L/>, novembre 2006.
- Dewey Decimal Classification – OCLC, « Dewey Services – Resources », disponible à l'adresse : <https://www.oclc.org/en/dewey/resources.html>, 2020.
- Health Topics – MEDLINEPLUS, « Health Topics », disponible à l'adresse : <https://medlineplus.gov/healthtopics.html/>, 1 février 2017.

International Classification of Diseases-9 – CDC – NATIONAL CENTER FOR HEALTH STATISTICS, « International Classification of Diseases, Ninth Revision (ICD-9) », disponible à l'adresse : <https://www.cdc.gov/nchs/icd/icd9.htm/>, 6 novembre 2015.

International Classification of Diseases-11 – WHO, « ICD-11 for Mortality and Morbidity Statistics (Version 04 / 2019) », disponible à l'adresse : <https://icd.who.int/browse11/l-m/en/>, 2019.

Library of Congress Subject Headings – Library of Congress, « Linked Data Service - Library of Congress Subject Headings », disponible à l'adresse : <http://id.loc.gov/authorities/subjects.html>, 2019.

MeSH bilingue anglais-français – INSERM, « Le MeSH bilingue anglais-français », disponible à l'adresse : <http://mesh.inserm.fr/FrenchMesh/>, 2019.

NLM Classification – NATIONAL LIBRARY OF MEDICINE, « NLM Classification 2020 Winter Edition », disponible à l'adresse : <https://www.nlm.nih.gov/class/index.html/>, 2020.

SNOMED – SNOMED INTERNATIONAL, « Leading healthcare terminology, worldwide », disponible à l'adresse : <http://www.snomed.org/>, 2020.

Universal Decimal Classification – UDC CONSORTIUM, « UDC Summary », disponible à l'adresse : <http://www.udcsummary.info/php/index.php>, 2020.