



Chapitre d'actes

2018

Accepted version

Open Access

This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

Une ontologie pour la formalisation et la visualisation des connaissances scientifiques

Daponte, Vincenzo; Falquet, Gilles

How to cite

DAPONTE, Vincenzo, FALQUET, Gilles. Une ontologie pour la formalisation et la visualisation des connaissances scientifiques. In: 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018. Sylvie Ranwez (Ed.). Nancy (France). [s.l.] : [s.n.], 2018. p. 129–136.

This publication URL: <https://archive-ouverte.unige.ch/unige:138011>

Une ontologie pour la formalisation et la visualisation des connaissances scientifiques

Vincenzo Daponte¹, Gilles Falquet¹

¹ CUI, Centre Universitaire d'informatique, Université de Genève, Genève, Suisse
Vincenzo.Daponte@unige.ch, Gilles.Falquet@unige.ch

Résumé : La construction d'une ontologie des objets de connaissance scientifique, présenté ici, s'inscrit dans le développement d'une approche orientée à la visualisation des connaissances scientifiques. Il est motivé par le fait que les concepts d'organisation de la connaissance scientifique (théorème, loi, expérience, preuve, ...) apparaissent dans des ontologies existantes mais qu'aucune de celles-ci n'est centrée sur cette thématique et n'en présente une organisation simple et facilement utilisable. Nous présentons la première version construite à partir de sources ontologiques (ontologies des objets de connaissance de certains domaines, lexicales et de niveau supérieur), de bases de connaissances spécialisées et d'interviews avec des scientifiques. Nous avons aligné cette ontologie avec certaines des sources utilisées, ce qui a permis de vérifier sa consistance par rapport à ces dernières. La validation de l'ontologie consiste à l'utiliser pour formaliser des connaissances de diverses sources, ce que nous avons commencé à faire dans le domaine de la physique.

Mots-clés : Ontologies, Connaissance scientifique, Visualisation des connaissance

1 Motivations

L'accès à la connaissance scientifique, qu'elle soit générale ou factuelle doit forcément passer par une présentation visuelle, auditive, ou autre qui fait appel à un ou plusieurs sens de l'être humain. Si on s'intéresse à la présentation visuelle des connaissances, on constate que la langue naturelle écrite y occupe une place prépondérante mais que d'autres formes graphiques (notations, formules mathématiques et chimiques, diagrammes, tableaux, formulaires, hypertextes, etc.) y jouent un rôle important pour faciliter l'accomplissement de diverses tâches intellectuelles (calcul, comparaison, déduction, etc.).

Le cadre général dans lequel s'inscrit notre travail est l'étude des techniques de visualisation des connaissances scientifiques et en particulier leur spécification formelle en vue des construire des outils de visualisation adaptés aux tâches de l'utilisateur scientifique. En effet, l'expérience montre qu'il n'existe pas *une* technique optimale de visualisation mais que l'efficacité de chaque technique dépend du contexte et des objectifs de l'utilisateur (voir, par exemple Card *et al.* (1999)).

Pour représenter formellement la notion de technique de visualisation il faut, suivant le model de référence proposé par Chi (2000), définir un modèle abstrait des données à visualiser, un modèle abstrait des objets visuels et une application du modèle de données dans le modèle visuel abstrait. Dans le cas de la visualisation de connaissances scientifiques il faut donc créer un modèle abstrait des connaissances scientifiques à visualiser. Notre but étant de fournir une formalisation de la visualisation qui s'applique à n'importe quelle science, nous avons décidé de construire une ontologie des objets de structuration de la connaissance utilisés dans les diverses sciences. Une telle ontologie rendra possible

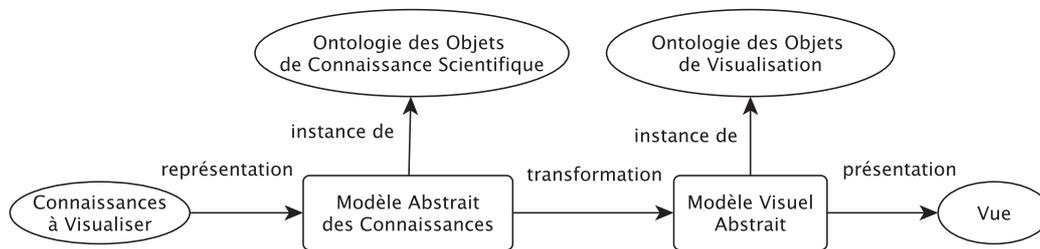


FIGURE 1 – Adaptation du modèle de Chi (2000) à la visualisation des connaissances scientifiques

la modélisation des connaissances à visualiser sous forme d'instances des classes de cette ontologie, selon le schéma de la figure 1.

Si les sciences varient en fonction de leurs objets d'étude (les fonctions différentiables, les papillons, les sociétés humains, les particules élémentaires, ...), elles possèdent également leurs propres concepts pour structurer la connaissance produite. Les mathématiques produisent des théorèmes, corollaires, lemmes, conjectures, preuves alors que la physique parle plutôt de lois, de principes, de mesures, de résultats expérimentaux où que l'anthropologie produit des observations, théories, hypothèses explicatives, méthodes. De plus, chaque science a développé un ensemble de formalismes (donc des constructions linguistiques) pour exprimer et traiter ces connaissances : formules et équations chimiques, formules mathématiques, diagrammes de flux, diagrammes d'interaction, arbres syntaxiques, etc. et un ensemble de techniques de productions de la connaissance : expérimentation, raisonnement formel, enquêtes, observations, etc. On constate cependant que le vocabulaire de classification des objets de connaissance scientifiques est restreint et que plusieurs termes recouvrent souvent des concepts similaires. Par conséquent on peut penser arriver à construire une ontologie centrale composée d'un petit nombre de classes d'objets scientifiques. Les connaissances à visualiser pourront alors être représentées comme des instances de ces classes.

Dans le reste de cet article nous commencerons par examiner les travaux centrés sur la représentation des connaissances scientifiques et ceux qui, tout en étant plutôt connexes, peuvent fournir des éléments importants. Nous présenterons ensuite la méthode suivie pour créer une première version de l'ontologie SKOO des objets de connaissance scientifique et l'ontologie obtenue. Nous présenterons ensuite les premières évaluations que nous avons réalisées. Dans la conclusion nous donnerons des perspectives sur l'utilisation pratique de cette ontologie et sur la poursuite des travaux d'évaluation et de mise au point de l'ontologie.

2 Etat de l'art

À notre connaissance il n'existe pas à l'heure actuelle d'ontologie dont le champ est la conceptualisation des objets servant à représenter ou structure la connaissance scientifique en général. Notons aussi que les travaux d'épistémologie des sciences ne s'attaquent en général pas à cette question ontologique générale mais traitent soit d'une science parti-

culière, soit d'un aspect particulier des sciences. Ce travail ontologique a par contre été réalisé en ingénierie des connaissances pour quelques domaines spécifiques.

Le noyau de l'ontologie OMDoc présenté dans (Lange (2013)) présente une modélisation de la connaissance mathématique sous forme d'éléments de connaissances (*knowledge items*) qui sont des types d'objets mathématiques, des théories ou des énoncés (*statements*), lesquels peuvent être de type *Assertion*, *Proof*, *Definition*, *Axiom*, etc. De plus, des relations lient ces types d'éléments entre eux, par exemple la relation *proves* lie *Proof* à *Assertion*. Bien que cette ontologie soit dédiée aux mathématiques, on voit qu'elle pourrait s'étendre facilement à d'autres sciences.

L'ontologie SIO (Semanticscience Integrated Ontology), Dumontier *et al.* (2014), est une ontologie de niveau supérieur qui vise essentiellement la représentation des connaissances biomédicales. Le but premier de SIO est la descriptions des objets complexes du domaine biomédical (avec des relations composant-composé) et des processus dans lesquels ils interviennent ou des procédures (expérimentales) qu'on leur applique. On trouve cependant dans la classe *description* de SIO bon nombre de concepts servant à structurer la connaissance scientifique : *argument*, *belief*, *conclusion*, *evidence*, *hypothesis*, ... Mais contrairement à OMDoc, il n'y a cependant pas de relations spécifiques entre ces concepts, ils peuvent par contre être liés aux objets qu'ils décrivent. SIO décrit également les objets linguistiques, mathématiques et les médias servant de support à la connaissance.

La notion de processus scientifique expérimental est au centre du système EXPO (Soldatova & King (2006)). Il associe l'ontologie SUMO ((Niles & Pease, 2001,)) à des ontologies d'expériences spécifiques à un sujet en formalisant les concepts génériques de conception expérimentale, de méthodologie et de représentation des résultats. EXPO vise à décrire différents domaines expérimentaux et à donner une description formelle des expériences pour l'analyse, l'annotation et le partage des résultats.

On peut également considérer ce qui a été fait dans les bases de connaissances scientifiques, telles Gene Ontology (Ashburner *et al.* (2000)), OntoMathPro (Nevzorova *et al.* (2014)) ou encore FMA (Rosse & Mejino (2007)) qui ont pour but de représenter l'état actuel de nos connaissances dans un domaine. Elles sont en général constituées d'une partie terminologique qui organise très précisément les concepts du domaine et d'une partie composée d'assertions (énoncés) qui représentent nos connaissances à propos de ces concepts. Dans Gene Ontology (GO) les énoncés sont appelés « annotations ». Ils lient typiquement un gène et un terme de l'ontologie GO (par exemple pour indiquer que le gène possède une certaine fonction). Les énoncés sont qualifiés par un type de preuve (expérimentale, inférence phytogénétique, inférence automatique, ...). Dans OntoMathPro (Nevzorova *et al.* (2014)) les niveaux terminologiques et assertions existent bel et bien mais ne sont pas structurellement séparés. Ainsi le théorème de Stokes (assertion) n'est pas une instance mais une sous-classe de la classe *Theorem*. De même, on trouve comme sous-classe de premier niveau de *Mathematical knowledge object* aussi bien *Theorem* que *Tensor*. Il y a donc agrégation des objets de description de la connaissance et des objets du domaine sur lequel on travaille.

Notons encore qu'il existe des ontologies dont le but est uniquement de répertorier et classer les objets d'étude d'un domaine ou de créer une terminologie d'une domaine (SWEET, ScienceWISE, ...). En général ces ontologies ne s'intéressent pas aux objets de structuration de la connaissance.

À l'inverse, on trouve dans une ontologie lexicale comme WordNet un grand nombre de concepts tels que théorème, loi, définition, hypothèse, corollaire. Cependant on remarque

que ces concepts ne sont pas organisés de manière directement utilisable. On a par exemple les chaînes de relation d'hyperonymie

theorem < idea < content < cognition

et

corollary < ... < process < content < cognition

alors que d'un point de vue formel un corollaire est un théorème. En d'autres termes, on ne peut pas extraire une ontologie des objets de connaissance scientifique par simple projection d'une partie de WordNet. Il en va de même pour d'autres ontologies de niveau supérieur (SUMO, CyC, ...)

3 Construction de l'ontologie SKOO

Pour construire l'ontologie des objets de connaissance scientifique (*Scientific Knowledge Objects Ontology* - SKOO) nous avons appliqué le processus suivant :

1. Nous avons collecté un ensemble de termes utilisés pour structurer les connaissances dans différents domaines scientifiques. Cette opération a été effectuée par consultation d'ouvrages (manuels, formulaires, monographies « handbooks ») en biochimie, physique, mathématiques, linguistique, sociologie ; interviews avec des scientifiques de différents domaines ; analyse du niveau terminologique de bases de connaissances et d'ontologies scientifiques (Gene Ontology, OntoMathPro, ...)
2. Pour construire le niveau supérieur de l'ontologie nous avons tout d'abord associé les termes mis en évidence à des « synsets » équivalents ou plus généraux de WordNet. Puis nous avons utilisé l'ontologie DOLCE, déjà alignée avec WordNet, pour trouver des concepts de niveau supérieur.
3. Enfin nous avons défini des relations entre les concepts du niveau supérieur à partir de relations trouvées dans des ontologies scientifiques, en particulier OMDoc, et par spécialisation de certaines relations de haut niveau de DOLCE.

La figure 2 montre le niveau supérieur de l'ontologie obtenue et ses liens avec DOLCE et WordNet. Nous décrivons ci-dessous l'interprétation de chacune de ses classes.

Sci_Knowledge_Item Les éléments de connaissance scientifique sont tous les objets qui servent à structurer l'expression de la connaissance scientifique. Il peut s'agir d'objets, tels les théorèmes, lois (physique, chimiques), modèles ou méthodes qui portent en eux-mêmes de la connaissance au sens platonicien de croyance vraie et justifiée. Mais il peut aussi s'agir d'objets « auxiliaires » tels que les définitions, exemples, preuves, hypothèses, problèmes. Ces objets correspondent aux objets de la classe *description* de l'ontologie DOLCE (Masolo *et al.* (2003)).

Sci_Information_Object Cette classe a pour but de regrouper toutes les formes d'expression des éléments de connaissance, qu'elles soit linguistique ou sous forme de diagramme, schémas, formules, etc.. Il s'agit d'une sous-classe de la classe

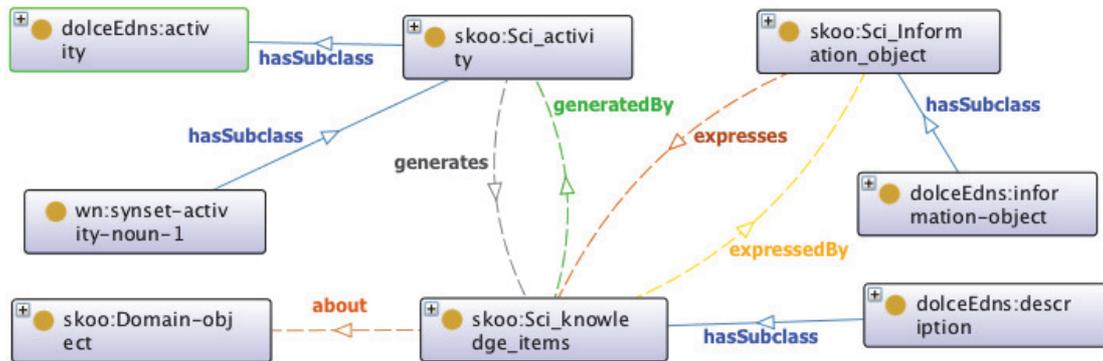


FIGURE 2 – Structure du niveau supérieur de l'ontologie SKOO

information-object de l'ontologie DOLCE (Masolo *et al.* (2003)), et sa classe principale *Sci-linguistic-object* est une sous-classe de la classe *linguistic-object* de DOLCE (Masolo *et al.* (2003)). Cette classe veut inclure toutes les formes et méthodes pour exprimer les concepts utilisés pour représenter la connaissance des disciplines considérées.

Sci_Activity Cette classe représente les activités, au sens de *activity* (hyponyme de *human activity*) dans WordNet, qui servent à engendrer des éléments de connaissances scientifique. Il peut s'agir d'activités de type expérimental (processus, expérimentation, observation), mais aussi de type empirique (mener des enquêtes) ou formelles (prouver formellement, calculer). La description précise des activités, en particulier des expérimentation, n'est pas définie dans cette ontologie car elle est déjà couverte par d'autres ontologies, telles SIO et EXPO.

Domain-object représente tous les objets à propos desquels des connaissances scientifiques sont exprimées. Cette classe sert de point d'ancrage aux classes décrivant les objets étudiés dans des domaines spécifiques. Lors de l'usage pratique de l'ontologie, le principe est d'importer une ontologie d'objets du domaine scientifique concerné et créer des axiomes de subsomption $C \sqsubseteq \text{Domain-object}$ pour ses classes de niveau supérieur.

4 Evaluation

Nous avons mené deux types d'évaluation de type consistance et capacité. En plus de la consistance interne de l'ontologie, pour donner une indication de la consistance externe (par rapport à d'autres ontologies), l'ontologie a été alignée avec les ontologies OMDoc, DOLCE et WordNet. Pour cela nous avons traduit sous forme de classes OWL les concepts de OMDoc et WordNet, puis nous avons créé des axiomes de correspondance de type owl:subClassOf et owl:EquivalentClass entre celles-ci et SKOO. La table 1 montre quelques uns de ces axiomes. Nous avons ensuite vérifié la consistance de l'ontologies

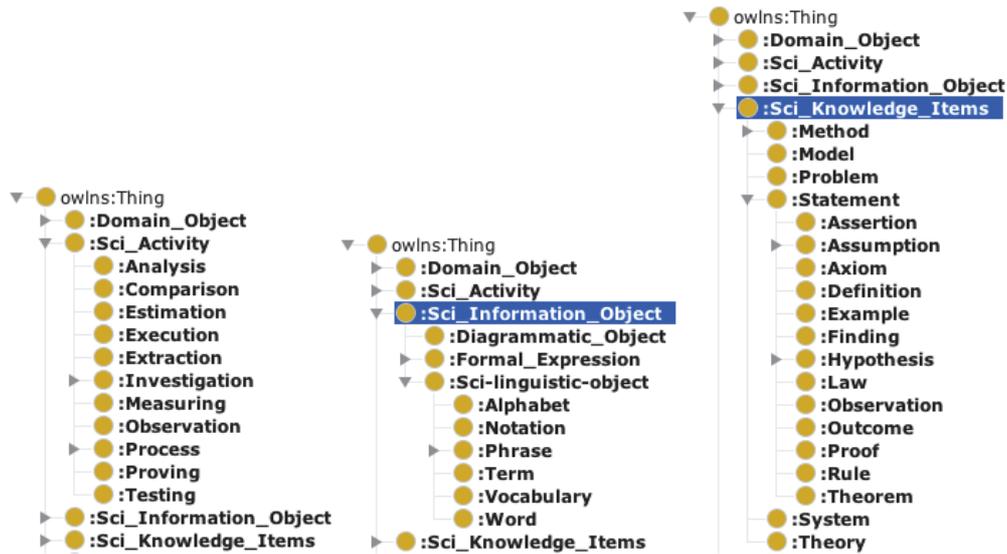


FIGURE 3 – Les niveau supérieures des sous-classes de Sci-activity, Sci-knowledge-items et Sci-information-object.

SKOO	OMDoc	DOLCE	WordNet
Sci-knowledge-items	\sqsupseteq MathKnowledgeItem	\sqsubseteq description	
Statement	\sqsupseteq Statement		\sqsubseteq statement
Theory	\sqsupseteq Theory	\sqsubseteq theory	\sqsubseteq theory
Assertion	\sqsupseteq Assertion		\sqsubseteq assertion
Axiom	\sqsupseteq Axiom		\sqsubseteq axiom
Definition	\sqsupseteq Definition		\sqsubseteq definition
Proof	\sqsupseteq Proof		\sqsubseteq proof
Sci-activity			\sqsubseteq activity
Process		\sqsupseteq activity	\sqsubseteq process
Sci-information-object		\sqsubseteq information-object	

TABLE 1 – Correspondances entre les classes SKOO et OMDoc, DOLCE, WordNet

obtenue en fusionnant SKOO, les trois ontologies et les axiomes de correspondance (mais sans correspondances entre DOLCE, WordNet et OMDoc).

Pour évaluer les capacités de l'ontologie nous devons vérifier si, étant donné un système de visualisation de connaissances scientifiques, l'ontologie permet de créer un modèle abstrait adéquat pour ces connaissances. Dans le cas des bases de connaissances structurées et homogènes, par exemple les annotations de Gene Ontology ou des formulaires mathématiques il est aisé de vérifier l'adéquation de l'ontologie. En effet ces connaissances correspondent généralement à des énoncés (*Statement*) qui peuvent être des théorèmes. Par contre le cas des connaissances exprimées dans des textes est plus complexe. Nous avons effectué un premier test en prenant comme système de visualisation une partie d'un

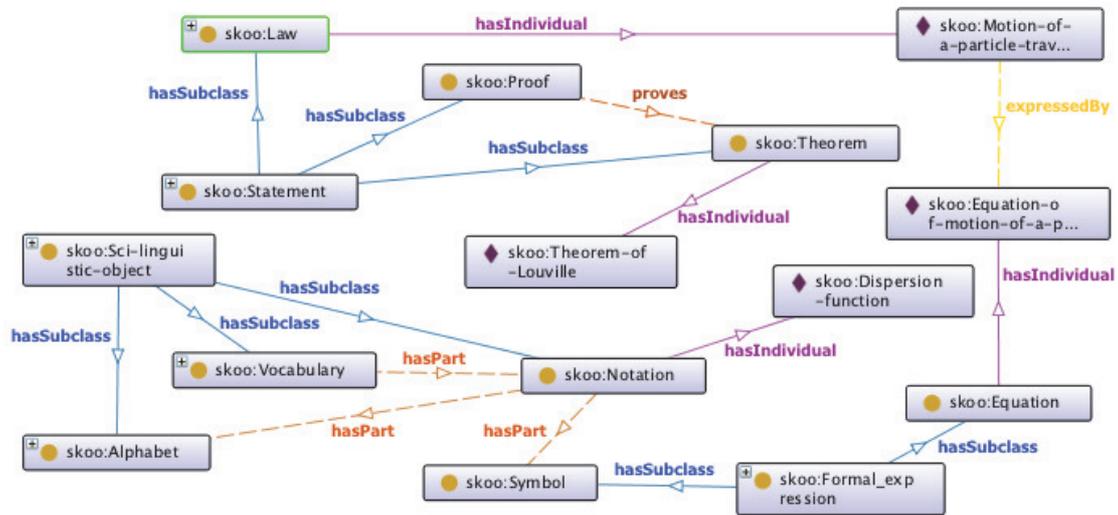


FIGURE 4 – Quelques concepts contenus dans plusieurs sections du chapitre 3 de (Wille (2000)) exprimés sous forme d’instances de l’ontologie SKOO.

ouvrage de physique des accélérateurs (Wille (2000)). Nous avons modélisé différents concepts des différentes sections du chapitre 3. En particulier, le concept principal exprimé dans la section 3.2, la section 3.6 complète de cet ouvrage (*dispersion and momentum compaction factor*) et un théorème utilisé dans la section 3.8 sous forme d’instances de classes SKOO. La figure 4 montre la modélisation d’une loi physique (instance de *Law*) représentée par une équation (instance de *Equation*) et aussi le modélisation d’un théorème (instance de *Theorem*) et d’une notation (instance de *Notation*) utilisée pour représenter un concept particulier. Il faut souligner que dans les relations montrées par cette figure, la relation *hasIndividual* est utilisée par *Protégé* pour associer le type de l’instance (la classe) à l’instance elle-même.

5 Conclusions et travaux futurs

Nous avons présenté la construction de la première version de l’ontologie SKOO dont le but est de fournir un modèle général pour la modélisation de connaissances scientifiques à visualiser (selon le schéma de Chi (2000)). Bien que les concepts représentés dans cette ontologie existent tous dans d’autres ontologies, aucune de celles-ci ne les regroupe de manière à être directement utilisables pour représenter la connaissance scientifique. D’où l’intérêt de l’ontologie SKOO. L’ontologie a été alignée avec des ontologies de référence pour vérifier à l’aide d’un raisonneur qu’elle n’entraient pas en contradictions avec celles-ci. D’autre part nous avons commencé la validation de la capacité de cette ontologie à modéliser les connaissances représentées dans des bases de connaissances existantes, ce qui ne pose pas de problème particulier, et les connaissances représentées dans des (hyper)textes, ce qui est plus difficile, surtout pour les textes de sciences humaines et sociales. Après avoir réalisé un test sur une partie d’un ouvrage de physique, nous allons entre-

prendre des tests sur des ouvrages de sciences humaines et sociales.

La prochaine étape de ce travail consistera à modéliser complètement divers systèmes de visualisation de connaissances existants. Nous utiliserons pour cela le langage SPARQL pour spécifier les transformations d'un modèle de connaissance (exprimé avec SKOO) vers un modèle d'objets de visualisation (comprenant des listes, arbres, graphes, textes, formes géométriques, etc.). Ceci permettra de valider le modèle complet de spécification de techniques de visualisation. À partir de là il sera possible de créer un système de génération de visualisations à partir de leur spécification et de l'utiliser pour créer de nouvelles techniques de visualisation et de les tester avec des utilisateurs.

Au cours de ce travail nous nous sommes aperçus que l'intérêt de cette ontologie va au-delà de la seule visualisation des connaissances. Elle est par exemple applicable dans le cadre de la recherche d'information précise ou du raisonnement automatique sur de grands ensembles de connaissances scientifiques.

Remerciements

Ce travail a été enrichi par les précieux conseils et l'expérience de Giuseppe Cosenza et Jean-Pierre Hurni, que les auteurs souhaitent remercier.

Références

- ASHBURNER M., BALL C. A., BLAKE J. A., BOTSTEIN D., BUTLER H., CHERRY J. M., DAVIS A. P., DOLINSKI K., DWIGHT S. S., EPPIG J. T. *et al.* (2000). Gene ontology : tool for the unification of biology. *Nature genetics*, **25**(1), 25.
- CARD S. K., MACKINLAY J. D. & SCHNEIDERMAN B. (1999). *Readings in Information Visualization : Using Vision to Think*. Morgan Kaufman.
- CHI E. (2000). A taxonomy of visualization techniques using the data statereference model. In *InfoVis 2000. IEEE Symposium*, p. 69–75.
- DUMONTIER M., BAKER C. J., BARAN J., CALLAHAN A., CHEPELEV L., CRUZ-TOLEDO J., DEL RIO N. R., DUCK G., FURLONG L. I., KEATH N., KLASSEN D., MCCUSKER J. P., QUERALT-ROSINACH N., SAMWALD M., VILLANUEVA-ROSALES N., WILKINSON M. D. & HOEHNDORF R. (2014). The semantic science integrated ontology (sio) for biomedical research and knowledge discovery. *Journal of Biomedical Semantics*, **5**(1), 14.
- LANGE C. (2013). Ontologies and languages for representing mathematical knowledge on the semantic web. *Semantic Web*, **4**(2), 119–158.
- MASOLO C., BORGIO S., GANGEMI A., GUARINO N. & OLTRAMARI A. (2003). *The WonderWeb library of foundational ontologies and the DOLCE ontology*. WonderWeb (EU IST project 2001-33052) deliverable D18. Rapport interne, LOA-ISTC-CNR.
- NEVZOROVA O. A., ZHILTSOV N., KIRILLOVICH A. & LIPACHEV E. (2014). Ontomath pro ontology : a linked data hub for mathematics. In *International Conference on Knowledge Engineering and the Semantic Web*, p. 105–119 : Springer.
- NILES I. & PEASE A. (2001). Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, p. 2–9 : ACM.
- ROSSE C. & MEJINO J. (2007). The foundational model of anatomy ontology. In *Anatomy Ontologies for Bioinformatics : Principles and Practice*, volume 6, p. 59–117 : Springer.
- SOLDATOVA L. N. & KING R. D. (2006). An ontology of scientific experiments. *Journal of the Royal Society Interface*, **3**(11), 795–803.
- WILLE K. (2000). *The physics of particle accelerators : an introduction*. Clarendon Press.