The Next Generation of Similarity Measures that fully explore the Semantics in Biomedical Ontologies

Francisco M. Couto

Corresponding author

Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa

fcouto@di.fc.ul.pt

H. Sofia Pinto

INESC-ID, Departamento de Engenharia Informática, Instituto Superior Técnico

sofia@inesc-id.pt

Abstract

There is a prominent trend to augment and improve the formality of biomedical ontologies. For example, this is shown by the current effort on adding description logic axioms, such as disjointness. One of the key ontology applications that can take advantage of this effort is the conceptual (functional) similarity measurement. The presence of description logic axioms in biomedical ontologies make the current structural or extensional approaches weaker and further away from providing sound semantics-based similarity measures. Although beneficial in small ontologies, the exploration of description logic axioms by semantics-based similarity measures is computational expensive. This limitation is critical for biomedical ontologies that normally contain thousands of concepts. Thus in the process of gaining their rightful place, biomedical functional similarity measures have to take the journey of finding how this rich and powerful knowledge can be fully explored while keeping feasible computational costs. This manuscript aims at promoting and guiding the development of compelling tools that deliver what the biomedical community will require in a near future: a next-generation of biomedical similarity measures that efficiently and fully explore the semantics present in biomedical ontologies.

Keywords: Ontologies; Semantic Similarity; Functional Similarity; Description Logics

Electronic version of an article published as Journal of Bioinformatics and Computational Biology, 11, 5, 1371001, 2013, dx.doi.org/10.1142/S0219720013710017

Why Functional Similarity?

We have a natural tendency to compare biomedical entities based on their "looks", i.e. based on the digital representation of their primary structure. For example, sequence similarity measurement, like BLAST [1], is a common step to almost all molecular biology studies that compare genes and proteins. This happens because sequences can be treated efficiently by computational methods and because sequences are accurate and common digital representations of the primary structure of genes and proteins. Sequence similarity requires only the information on the primary structure (the sequence itself), but limits the analysis to proteins that share a similar sequence, independently of their biological role. The function of a protein derives from its structure, but not always a high structural similarity corresponds to a high functional similarity and vice versa [22, 31]. For example, the Human protein hIL-10 and the Epstein-Barr virus protein vIL-10 share a high structural similarity, but they have a clear distinct physiological profile [56].

Divergence of function with sequence conservation is an exception rather than a rule, so in general sequence similarity remains as a reliable technique to determine the functional similarity of proteins. However, this gap between structural and functional similarity creates an opportunity to develop similarity measures that can juxtapose, combine and/or complement structural similarity measures with a degree of shared functional characteristics. For example, when searching for proteins with an oxidoreductase activity, we may also be interested in proteins with similar activities, such as monooxygenase activity, independently of their structural similarity. This analysis of similar activities has become computationally possible due to the prevailing usage of ontologies to functionally characterize biomedical entities.

Ontologies and Similarity Measures

Ontologies can be loosely defined as "a vocabulary of terms and some specification of their meaning" [27, 52].

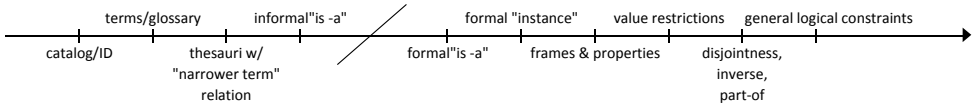

Figure 1: The Ontology Spectrum from the ontology panel at AAAI'99 [30]. The panelists: Lehman, McGuinness, Ushold, and Welty.

Figure 1 shows a well-known classification attempt to describe how ontologies cover a wide spectrum [30,50], going from their most basic form of a catalog ID, Glossary, Thesauri; to one where concepts are formally expressed and struc-

tured through logically defined *is-a* relations; to one where formal instances are explicitly connected to their defining concepts; and finally to one where general logical constraints are used to express the definition of those concepts in a formally defined logic. Going up in the formality&complexity scale, ontologies get more accurate and expressive.

Recently, conceptual similarity was defined as a **function** that, given two ontology terms or two sets of terms annotating two entities, returns a numerical value reflecting the closeness in **meaning** between them [38]. The formality&complexity of the ontology defines the type of function and meaning that a conceptual similarity measure can implement. Thus, most of the approaches can be classified into the following categories [18]:

Terminological approaches focus on the names of the classes. For example, the term Cats is a morphological derivation of Cat, and thus classes bearing such names are likely equivalent. These approaches can be complemented with thesauri and dictionaries, to explore lexical relations such as synonymy.

Structural approaches explore the structure of the classes, i.e. their relations to other classes. The sub-class and super-class relations provide a taxonomic backbone that can be explored using graph matching techniques. Ontology-specific relations, along with their properties (or facets), such as domain, range and cardinality, can also be explored.

Extensional approaches can only be applied when there is a large set of instances. The intuition behind these approaches is that the more instances two classes share, the more likely they have a high similarity. In case there are no shared instances between ontologies, distance metrics between individuals can be computed.

Semantics-based approaches are sensitive to the semantics of the logical formalism in which the ontologies are formalized, and are thus enabled to resort to inference techniques. The goal of this type of measures is to fully explore the available description logic axioms.

Note that ontology similarity, or global similarity [18], is the task of assessing the similarity of ontologies as wholes. This is usually done by flooding or aggregating conceptual similarity values.

Going up in the formality&complexity scale has consequences. Although there are many terminological, structural and/or extensional measures (including hybrid ones), there are only a few semantics-based similarity measures [3, 20, 24]. The lack of semantics-based similarity measures can be explained by the following factors:

Computational complexity: semantics-based measures usually require some sort of deduction, which usually requires exponential time, even in less expressive formalisms.

Incomplete knowledge: classes in an ontology do not need to be, and indeed rarely are, completely defined.

The incomplete knowledge problem can only be effectively addressed if we improve the quality of the process of ontology development. For example, the omission of simple and obvious relationships in the January 2003 release of SNOMED-CT®, did not allow the inference of "uterus" as part-of "female genital tract", nor the inference that "uterus" as a role in "pregnancy" [9]. However, design patterns and best practices in ontology specification are gaining popularity, which may contribute to minimize the impact of this problem. For example, the OBO (Open Biomedical Ontology) foundry has been successfully promoting the correct application of a set of principles to the task of ontology development to support biomedical data integration [48].

The computational complexity problem is due to the large size of biomedical ontologies. They usually contain thousands of concepts that are used to annotate an even larger set of entities. For example, studies show that consistency checking using description logics is EXPTIME-Complete [17, 44] Note that, EXPTIME-Complete are thought to be the hardest problems in EXPTIME, which is the set of all decision problems solvable by a deterministic Turing machine in $\mathcal{O}(2^{p(n)})$ time, where $p(n)$ is a polynomial function of n. Most biomedical high-throughput studies analyze large numbers of entities simultaneously involving computationally intensive similarity calculations. Thus, the computational cost of current semantics-based similarity measures poses a major bottleneck to their use in biomedical ontologies.

Biomedical Ontologies

Etymologiae was one of the first attempts to systematize medicine knowledge [33, Book IV: Medicine], however only in the last decades the biomedical community engaged on a tremendous and noble effort of developing and using ontologies. These ontologies normally serve as controlled vocabularies to annotate the vast amount of biomedical entities being discovered with their functional characteristics [43]. For example, SNOMED-CT® (Systematized Nomenclature Of Medicine-Clinical Terms) is considered to be the most comprehensive collection of medical terms, which is used as a standard terminology to represent clinically relevant information and enable the ontological annotation of electronic health records. In 2012, SNOMED-CT® included more than 300,000 unique concepts. Another example is GO (Gene Ontology), an extensively used ontology to annotate proteins with concepts describing their molecular function, biological process and cellular component [4]. These ontological annotations enabled the interoperability and automatization of these semantic characterizations through different communities. GO is part of OBO and in 2012 it included more than 38,000 unique concepts fully defined. Another ontology part of OBO is ChEBI (Chemical Entities of Biological Interest), a prominent ontology of molecular entities focused on small chemical compounds [16]. In 2013, ChEBI included more than 32,000 fully annotated entities.

As the number of people developing and using ontologies continues to grow, their size will rise too. However, to maintain the quality of ontologies we have now to pay special attention to important features of ontology languages that were neglected in their initial specification, such as the lack of description logical axioms in GO [54] and SNOMED-CT® [9]. Design patterns and best practices in ontology specification are being disseminated and their application stimulated [7, 49]. Description logical axioms, such as disjointness, have been recently included in these large and popular biomedical ontologies, for example in GO [10] and ChEBI [21]. Roadmaps for overall improvement of SNOMED-CT® by addressing both logical and ontological issues are also being considered [40, 46]. Although the primary purpose of adding these axioms is to automatically identify inconsistencies and misannotations in the ontologies, they can also have an important role for enhancing the way conceptual similarity is calculated [13]. Therefore, the inclusion of the description logical axioms being added to biomedical ontologies in similarity measures is much required.

Current Biomedical Functional Similarity

A successful application of terminological approaches to biomedical ontologies has been to find the most similar ontological concepts to the terms recognized in biomedical literature by text mining methods. For example, in BioCreAtIvE 2004 [23] a measure based on the textual descriptors of GO concepts was used to resolve references to GO in biomedical literature [12]. One of the first structural approaches successfully applied to a biomedical ontology was a path distance measure that demonstrated the advantages of using the hierarchical relations of the Medical Subject Headings (MeSH) [39]. More recently, structural approaches based on graph-based algorithms have been proposed [35].

In Molecular Biology, as in other biomedical areas, mainstream methods for functional characterization of genes and proteins are based on ontological annotation, for example using GO. This enabled the successful development and application of conceptual similarity measures based on extensional approaches. These measures compare two proteins according to the amount of ontological information their annotations share. The shared ontological information can be inferred from the most informative common ancestor of the annotated concepts, or from all the disjunctive common ancestors [11].

Inspired on Tversky's contrast model [51] and Jaccard similarity measure [25], conceptual similarity between proteins has been presented as a good predictor of functional similarity. The seminal work based on Resnik's measure [41] identified a correlation between structural and conceptual similarity [34], just by defining the information content of each GO concept as inversely proportional to its number of annotated proteins. Conceptual similarity has become a popular approach to compare biomedical entities based on their functional characterizations (annotations). For example, similarity measures have already been proven to be useful in many biomedical studies, such as: in information retrieval [45], in discovering novel relationships [2,32,36,53], in clustering entities [55], in clinical

diagnostics [29], and in classification problems [19].

Nowadays, many conceptual similarity measures use structural or extensional approaches and are usually designated as semantic similarity measures [37, 38]. However, these approaches do not include the description logic axioms that ontologists are starting to provide. In the remainder of this manuscript, we will explain the limitations of these structural or extensional approaches and how semantics-based similarity measures may overcome this problem.

Meaning and Function

As presented previously, conceptual similarity measures can be defined in terms of the notions of **meaning** and **function**. We assume **meaning** to be the role or activity that biomedical entities have in living systems, which is denoted by their ontological annotations as opposed to their primary structure. For example, the meaning of a protein can be unambiguously described by GO annotations in terms of its (1) molecular function; (2) biological process or (3) cellular component. Even by using ontological annotations to define the **meaning** some issues should be taken in account to avoid common pitfalls [42]. For example the following issues may have a significant impact on the calculation of conceptual similarity:

Negligibility: many biomedical entities do not have ontological annotations or are too general to be of any value, in opposition to structure that is almost always available.

Inaccuracy: unsound ontological annotations that are inferred directly from automated methods, for example by improper application of structural similarity to propagate annotations.

Subjectivity: different communities have different perspectives of the meaning of a biomedical entity. For example, the activity of a protein may vary according to different species.

These issues will be mitigated as the community that collaborates in the annotations refinement continues to grow. For example, the quality of computationally inferred annotations is now very high in some popular ontologies, such as in GO [47]. This enhancement effort enabled the successful application of conceptual similarity in many biomedical studies, as described in the previous section. Additionally, an effective application of the set of good principles and practices proposed by OBO could also be a way to address some of these issues.

A mathematical definition of a similarity **function** is available in [6]. **Function** is defined as a measure that receives as input two concepts and returns an element of a totally ordered set (usually $[0, 1]$), However, not all the conceptual similarity measures applied to biomedical ontologies satisfy the properties of positiveness ($sim(x, y) \geq 0$), reflexibility ($sim(x, x) = 1 \wedge sim(x, x) \geq sim(x, y)$) and symmetry ($sim(x, y) = sim(y, x)$). For example, the seminal Resnik's measure does not satisfy the reflexibility property, since it returns the information

of the concept ($sim(x, x) = IC(x)$) and not 1 [34, 41]. Moreover, conceptual similarity measures usually do not include disjoint axioms that may contain relevant information for the similarity calculation. For example, if two classes are disjoint, i.e. an individual cannot simultaneously be an instance of both classes, then this may suggest a lower similarity between these disjoint classes.

According to D'Amato et al. [14] the measures in both approaches (Structural or Extensional) fail to consider (equivalence) soundness and disjointness incompatibility:

Equivalence Soundness: if $x \equiv y \Rightarrow sim(x, z) = sim(y, z)$

Disjointness Incompatibility: if x and y are disjoint $\not\Rightarrow sim(x, y) = 0$

In biomedical ontologies equivalence and disjoint specifications were in most cases non-existent, so both notions were implicitly considered to be true for all conceptual similarity measures applied to biomedical ontologies. However, this is quickly changing since the community is now realizing the power of using formal conceptualizations to automatically check logical consistency and to enhance the reasoning capabilities [5,28]. Thus, developing proper conceptual similarity measures that fully explore the domain semantics formally expressed in the ontology will not only improve their effectiveness but also further motivate the ongoing process of adding description logic axioms to biomedical ontologies.

A fictitious example can be used to elucidate the usefulness of description logic axioms for calculating biomedical functional similarity. Let us assume that we have three drugs α, β and γ, each including in their composition a metal. Assuming that we know a priori that the metal of α is *ferrous*, of β is *noble* and of γ is *precious*, and assuming that the class *ferrous* metals is disjoint from the overlapping classes *noble* and *precious* metals, then we can infer that α cannot be composed by the same metal as β and γ, unlike β and γ that can be composed by the same material, e.g. *gold*. Thus, when comparing the drugs α, β and γ the disjointness would give us a useful insight into the similarity calculation.

Similarity using Description Logics

The current trend of improving the formality of biomedical ontologies will boost the availability and quality of their description logic axioms. Thus, conceptual similarity measures have to be prepared to include this rich knowledge that will improve their effectiveness. The advantages of developing well-defined conceptual similarity measures based on description logic axioms have already been demonstrated [8, 14, 26]. However, the exploration of description logic axioms by semantics-based similarity measures is computational expensive, i.e. their direct application to the large biomedical ontologies is now cumbersome. Thus, novel semantics-based similarity measures less computationally demanding are much required.

The theoretical foundations of semantics-based similarity may serve as the basis for developing novel measures. D'Amato et al. developed an efficient

method for resource retrieval that explores description logics to cluster the resources according to their conceptual similarity [15]. Another approach is developing add-ons for existing conceptual similarity measures in order to satisfy the equivalence soundness and the disjointness incompatibility properties. For example, add-ons for conceptual similarity measures were already successfully applied in finding common disjunctive ancestors [11].

One could argue that the description logic axioms present today in biomedical ontologies are still too scarce, vague and insignificant to represent a valid semantic source to enhance similarity results. However, this manuscript presents the benefits that would result from using these axioms for calculating functional similarity, even if they do not represent yet a perfect and comprehensive set. And as more and more description logic axioms become available the benefits will increase dramatically, making the usage of this powerful knowledge imperative. Thus, time is ripe to propose and develop the next-generation of similarity measures that take advantage of the valuable effort of ontology and description logics specialists that are currently working on the enhancement of the formal and logical aspects of biomedical ontologies.

Acknowledgments

The authors thank the Portuguese Fundação para a Ciência e Tecnologia through the financial support of the SOMER project (PTDC/EIA-EIA/119119/2010) and INESC-ID funding under project PEst-OE/EEI/LA0021/2013 and through the LASIGE multi annual support.

References

[1] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17):3389–3402, 1997.

[2] C. Andronis, A. Sharma, V. Virvilis, S. Deftereos, and A. Persidis. Literature mining, ontologies and information visualization for drug repurposing. *Briefings in bioinformatics*, 12(4):357–368, 2011.

[3] R. Araújo and H.S. Pinto. Towards semantics-based ontology similarity. In *Proceedings of the Workshop on Ontology Matching (OM2007) at ISWC/ASWC2007, Busan, South Korea*. Citeseer, 2007.

[4] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, et al. Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1):25, 2000.

[5] W. Blondé, V. Mironov, A. Venkatesan, E. Antezana, B. De Baets, and M. Kuiper. Reasoning with bio-ontologies: using relational closure rules to enable practical querying. *Bioinformatics*, 27(11):1562–1568, 2011.

[6] H.H. Bock and E. Diday. *Analysis of symbolic data: exploratory methods for extracting statistical information from complex data*. Springer Verlag, 2000.

[7] Martin Boeker, Daniel Schober, Djamila Raufie, Niels Grewe, Johannes Röhl, Ludger Jansen, and Stefan Schulz. Teaching Good Biomedical Ontology Design. In *Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012), KR-MED Series*, 2012.

[8] A. Borgida, T. Walsh, and H. Hirsh. Towards measuring similarity in description logics. In *Proceedings of the 2005 International Workshop on Description Logics (DL2005)*, volume 147, 2005.

[9] Werner Ceusters, Barry Smith, Anand Kumar, and Christoffel Dhaen. Ontology-based error detection in SNOMED-CT. *Proceedings of MEDINFO*, 2004:482–6, 2004.

[10] J. Chan, R. Kishore, P. Sternberg, and K. Van Auken. The Gene Ontology: enhancements for 2011. *Nucleic Acids Research*, 40(D1):D559–D564, 2012.

[11] F.M. Couto and M.J. Silva. Disjunctive shared information between ontology concepts: application to Gene Ontology. *Journal of Biomedical Semantics*, 2:5, 2011.

[12] F.M. Couto, M.J. Silva, and P.M. Coutinho. Finding genomic ontology terms in text using evidence content. *BMC bioinformatics*, 6(Suppl 1):S21, 2005.

[13] C. d'Amato, N. Fanizzi, and F. Esposito. A semantic similarity measure for expressive description logics. In *Proceedings of Convegno Italiano di Logica Computazionale, CILC05*, 2005.

[14] C. d'Amato, S. Staab, and N. Fanizzi. On the influence of description logics ontologies on conceptual similarity. *Knowledge Engineering: Practice and Patterns*, pages 48–63, 2008.

[15] Claudia d'Amato, Steffen Staab, Nicola Fanizzi, and Floriana Esposito. Efficient discovery of services specified in description logics languages. *Proc. of SMR2*, 243:15–29, 2007.

[16] Kirill Degtyarenko, Paula de Matos, Marcus Ennis, Janna Hastings, Martin Zbinden, Alan McNaught, Rafael Alcántara, Michael Darsow, Mickaël Guedj, and Michael Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research*, 36(suppl 1):D344–D350, 2008.

[17] F.M. Donini and F. Massacci. EXPTIME tableaux for ALC. *Artificial Intelligence*, 124(1):87–138, 2000.

[18] J. Euzenat and P. Shvaiko. *Ontology matching*, volume 18. Springer Berlin, 2007.

[19] J.D. Ferreira and F.M. Couto. Semantic similarity for automatic classification of chemical compounds. *PLoS computational biology*, 6(9):e1000937, 2010.

[20] F. GIUNCHIGLIA and P. SHVAIKO. Semantic matching. *The Knowledge Engineering Review*, 18(03):265–280, 2003.

[21] J. Hastings, P. de Matos, A. Dekker, M. Ennis, V. Muthukrishnan, S. Turner, G. Owen, and C. Steinbeck. Modular Extensions to the ChEBI Ontology. *International Conference on Biomedical Ontologies (ICBO)*, 2012.

[22] Troy Hawkins and Daisuke Kihara. Function prediction of uncharacterized proteins. *Journal of bioinformatics and computational biology*, 5(01):1–30, 2007.

[23] L. Hirschman, A. Yeh, C. Blaschke, and A. Valencia. Overview of BioCreAtIvE: critical assessment of information extraction for biology. *BMC bioinformatics*, 6(Suppl 1):S1, 2005.

[24] B. Hu, Y. Kalfoglou, H. Alani, D. Dupplaw, P. Lewis, and N. Shadbolt. Semantic metrics. *Managing Knowledge in a World of Networks*, pages 166–181, 2006.

[25] P. Jaccard. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz, 1901.

[26] K. Janowicz and M. Wilkes. SIM-DL A: A novel semantic similarity measure for description logics reducing inter-concept to inter-instance similarity. *The Semantic Web: Research and Applications*, pages 353–367, 2009.

[27] R. Jasper, M. Uschold, et al. A framework for understanding and classifying ontology applications. In *Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling, and Management KAW*, volume 99, pages 16–21, 1999.

[28] S. Jupp, R. Stevens, and R. Hoehndorf. Logical Gene Ontology Annotations (GOAL): exploring gene ontology annotations with OWL. *Journal of biomedical semantics*, 3(Suppl 1):S3, 2012.

[29] S. Köhler, M.H. Schulz, P. Krawitz, S. Bauer, S. Dölken, C.E. Ott, C. Mundlos, D. Horn, S. Mundlos, and P.N. Robinson. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *The American Journal of Human Genetics*, 85(4):457–464, 2009.

[30] O. Lassila and D. McGuinness. The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science*, 6(5):2001, 2001.

[31] D. Lee, O. Redfern, and C. Orengo. Predicting protein function from sequence and structure. *Nature Reviews Molecular Cell Biology*, 8(12):995–1005, 2007.

[32] JG Lees, JK Heriche, I. Morilla, JA Ranea, and CA Orengo. Systematic computational prediction of protein interaction networks. *Physical biology*, 8(3):035008, 2011.

[33] W.M. Lindsay. The Editing of Isidore Etymologiae. *The Classical Quarterly*, 5(1):42–53, 1911.

[34] P.W. Lord, R.D. Stevens, A. Brass, and C.A. Goble. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics*, 19(10):1275–1283, 2003.

[35] A Alvarez Marco and Changhui Yan. A graph-based semantic similarity measure for the gene ontology. *Journal of Bioinformatics and Computational Biology*, 9(06):681–695, 2011.

[36] G. Pandey, C.L. Myers, and V. Kumar. Incorporating functional interrelationships into protein function prediction algorithms. *BMC bioinformatics*, 10(1):142, 2009.

[37] T. Pedersen, S.V.S. Pakhomov, S. Patwardhan, and C.G. Chute. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of biomedical informatics*, 40(3):288–299, 2007.

[38] C. Pesquita, D. Faria, A.O. Falcao, P. Lord, and F.M. Couto. Semantic similarity in biomedical ontologies. *PLoS computational biology*, 5(7):e1000443, 2009.

[39] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(1):17–30, 1989.

[40] Alan Rector, Luigi Iannone, and Robert Stevens. Quality assurance of the content of a large DL-based terminology using mixed lexical and semantic criteria: experience with SNOMED CT. In *6th International Conference on Knowledge Capture; 26–29 June 2011*, pages 57–64, 2011.

[41] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995.

[42] S.Y. Rhee, V. Wood, K. Dolinski, and S. Draghici. Use and misuse of the gene ontology annotations. *Nature Reviews Genetics*, 9(7):509–515, 2008.

[43] D.L. Rubin, N.H. Shah, and N.F. Noy. Biomedical ontologies: a functional perspective. *Briefings in bioinformatics*, 9(1):75–90, 2008.

[44] K. Schild. *A correspondence theory for terminological logics: Preliminary report*. TU, Fachbereich 20, Projektgruppe KIT, 1991.

[45] A. Schlicker, T. Lengauer, and M. Albrecht. Improving disease gene prioritization using the semantic similarity of Gene Ontology terms. *Bioinformatics*, 26(18):i561–i567, 2010.

[46] Stefan Schulz, Boontawee Suntisrivaraporn, Franz Baader, and Martin Boeker. SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International journal of medical informatics*, 78:S86–S94, 2009.

[47] Nives Škunca, Adrian Altenhoff, and Christophe Dessimoz. Quality of computationally inferred gene ontology annotations. *PLoS computational biology*, 8(5):e1002533, 2012.

[48] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255, 2007.

[49] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A.L. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome biology*, 6(5):R46, 2005.

[50] B. Smith and C. Welty. Ontology: Towards a new synthesis. In *Formal Ontology in Information Systems*, pages 3–9. ACM Press, USA, pp. iii-x, 2001.

[51] A. Tversky. Features of similarity. *Psychological review*, 84(4):327, 1977.

[52] M. Uschold, M. Gruninger, et al. Ontologies: Principles, methods and applications. *Knowledge engineering review*, 11(2):93–136, 1996.

[53] N.L. Washington, M.A. Haendel, C.J. Mungall, M. Ashburner, M. Westerfield, and S.E. Lewis. Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS biology*, 7(11):e1000247, 2009.

[54] CJ Wroe, R. Stevens, CA Goble, and M. Ashburner. A methodology to migrate the gene ontology to a description logic environment using. In *Pacific Symposium on Biocomputing*, volume 8, pages 624–635, 2003.

[55] T. Xu, J.L. Gu, Y. Zhou, and L.F. Du. Improving detection of differentially expressed gene sets by applying cluster enrichment analysis to Gene Ontology. *BMC bioinformatics*, 10(1):240, 2009.

[56] S.I. Yoon, B.C. Jones, N.J. Logsdon, and M.R. Walter. Same Structure, Different Function: Crystal Structure of the Epstein-Barr Virus IL-10Bound to the Soluble IL-10R1 Chain. *Structure*, 13(4):551–564, 2005.