

# Ontology-Based Knowledge Discovery and Sharing in Bioinformatics and Medical Informatics: A Brief Survey

Jingshan Huang<sup>1</sup>

School of Computer and Information Sciences  
University of South Alabama  
Mobile, Alabama 36688  
huang@usouthal.edu

Dejing Dou

Computer and Information Science Department  
University of Oregon  
Eugene, Oregon 97403  
dou@cs.uoregon.edu

Lei He

National Institutes of Health  
Bethesda, Maryland 20892  
lei.he@nih.gov

Jiangbo Dang

Siemens Corporation  
Princeton, New Jersey 08540  
jiangbo.dang@siemens.com

Pat Hayes

Institute for Human and Machine Cognition  
Pensacola, Florida 32502  
phayes@ihmc.us

**Abstract**—Worldwide health scientists are producing, accessing, analyzing, integrating, and storing massive amounts of digital medical data daily, through observation, experimentation, and simulation. If we were able to effectively transfer and integrate data from all possible resources, then a deeper understanding of all these data sets and better exposed knowledge, along with appropriate insights and actions, would be granted. Unfortunately, in many cases, the data users are not the data producers, and they thus face challenges in harnessing data in unforeseen and unplanned ways. In order to obtain the ability to integrate heterogeneous data, and thereby efficiently revolutionize the traditional medical and biological research, new methodologies built upon the increasingly pervasive cyberinfrastructure are required to conceptualize traditional medical and biological data, and acquire the “deep” knowledge out of original data thereafter. As formal knowledge representation models, ontologies can render invaluable help in this regard. In this paper, we summarize the state-of-the-art research in ontological techniques and their innovative application in medical and biological areas.

**Index Terms**—bioinformatics; medical informatics; knowledge sharing; ontology matching; heterogeneous semantics; and semantic integration

## I. INTRODUCTION AND RESEARCH MOTIVATION

In medical informatics area, an abundance of digital data has promised a profound impact in both the quality and rate of discovery and innovation. Modern experimental and observational instruments generate and collect large sets of data of varying types at increasing speeds. Worldwide health scientists are producing, accessing, analyzing, integrating, and storing massive amounts of digital medical data daily, through observation, experimentation, and simulation. If we were able to effectively transfer and integrate data from all possible resources, then a deeper understanding of all these data sets

and better exposed knowledge, along with appropriate insights and actions, would be granted. Unfortunately, in many cases, the data users are not the data producers, and they thus face challenges in harnessing data in unforeseen and unplanned ways. In order to obtain the ability to integrate heterogeneous data, and thereby efficiently revolutionize the traditional medical and biological research, new methodologies built upon the increasingly pervasive cyberinfrastructure are required to conceptualize data, and acquire the “deep” knowledge out of original data thereafter. In particular, ontologies are formal, declarative knowledge representation models. They form a semantic foundation for many domains, and thus render great help to medical informatics researchers in better capturing hidden knowledge from large amounts of original data. The most renowned example of applying ontological techniques into medical and biological research is the Gene Ontology (GO) project [16].

Note that a lot of time and efforts have been spent in every search for available information in each small medical informatics subarea. This situation is further aggravated by great complexity and imprecise terminologies, which characterize typical medical and biological fields. A great deal of variety has been identified in the adoption of different biological terms, along with the relationships among all these terms. Such variety has inhibited effective information acquisition by both human and computers. Therefore, there is a need to explore innovative, cyber-enabled computing frameworks that are based on ontological techniques. Such frameworks will facilitate knowledge acquisition from existing resources, assist biologists in better understanding important biological functions at different levels, and ultimately, help clinicians in making sound decisions when treating their patients.

In the rest of this paper, we briefly summarize the state-

<sup>1</sup>Corresponding author Tel./Fax: 1-251-460-7612/1-251-460-7274.

of-the-art research in two areas, (1) ontologies and (2) the application of ontological techniques in medical and biological research. We aim to provide the reader with an introduction of major themes in these research areas, along with pointers to different research projects.

## II. ONTOLOGY RESEARCH

### A. Background

**Ontology** is a computational model of some portion or domain of the world [32]. The model describes the semantics of the terms used in the domain. Ontology is often captured in some form of a semantic network, i.e., a graph whose nodes are concepts or individual objects and whose arcs represent relationships or associations among the concepts. The semantic network is augmented by properties and attributes, constraints, functions, and rules, which govern the behavior of the concepts. In brief, an ontology consists of a finite set of concepts, along with these concepts' properties and relationships. In addition, most real-world ontologies have very few or no instances, i.e., they only have the aforementioned graphical structure (also known as "schema").

**Ontology Heterogeneity** is an inherent characteristic of ontologies developed by different parties for the same (or similar) domains. The heterogeneous semantics may occur in two ways. (1) Different ontologies could use different terminologies to describe the same conceptual model. That is, different terms could be used for the same concept, or an identical term could be adopted for different concepts. (2) Even if two ontologies use the same name for a concept, the associated properties and the relationships with other concepts are most likely to be different.

**Ontology Matching** is short for "Ontology Schema Matching", also known as "Ontology Alignment," or "Ontology Mapping." It is the process of determining correspondences between concepts from heterogeneous ontologies (often designed by distributed parties). Such correspondences include many relationships, for example, *equivalentWith*, *subclassOf*, *superClassOf*, and *siblings*.

**Machine Learning** is a scientific discipline that is concerned with the design and development of algorithms that allow computers to change behavior based on available data (also known as "training data"). A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data.

Ontological techniques have been widely applied to medical and biological research. The most successful example is the GO project [16], which is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. GO provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, as well as tools to access and process such data.

There are many research directions in ontologies, e.g., automatic ontology generation, ontology engineering, and ontology matching, etc., with ontology matching the most related one to this paper's theme.

### B. Related Work

According to the classification in [10], most ontology-matching techniques [12], [31] can be divided into two categories: rule-based approaches and learning-based approaches.

1) *Rule-Based Ontology Matching*: In [29], N.F. Noy and M.A. Musen describe PROMPT, a semiautomatic approach to ontology alignment. By performing some tasks automatically and guiding the user in performing other tasks for which intervention is required, PROMPT helps in understanding ontologies covering overlapping domains.

S. Castano et al. present H-MATCH in [5]. The authors divide the semantics of a concept into its linguistic and contextual parts. The former captures the meaning of terms used as concept names, while the latter evaluates the semantic affinity between two concepts by taking into account the affinity between their contexts, which are concept properties and relationships.

In [11], D. Dou et al. view ontology translation as ontology merging and automated reasoning, which are in turn implemented through a set of axioms. They obtain the merger of two related ontologies by taking the union of the terms and the axioms defining them, then adding bridging axioms through the terms in the merge. The language used in this approach, Web-PDDL, has the right degree of flexibility.

Similarity Flooding (SF) [28] is a matching algorithm based on a fixpoint computation that is usable across different scenarios. SF takes two graphs as input, and produces as output a mapping between corresponding nodes. This work defines several filtering strategies for pruning the immediate result of the fixpoint computation.

Cupid [26] is an algorithm for generic schema matching outside of any particular data model or application. It discovers mappings between schema elements based on their names, data types, constraints, and schema structure. Cupid has a bias toward leaf structures where much of the schema content resides.

S-Match [13], [15] views match as an operator that takes two graph-like structures and produces a mapping between the nodes of the graphs. Mappings are discovered by computing semantic relations, which are determined by analyzing the meaning that is codified in the elements and the structures of the schemas.

[14] presents structure preserving match, which preserves a set of structural properties of the graphs being matched. An approximate structure matching algorithm is described, based on a formal theory of abstraction, and build upon tree edit distance measures.

In [20] and [21], J. Huang et al. describe Puzzle to merge two source ontologies. The main idea is to relocate each concept from one ontology into another one. During this automated process, both linguistic and contextual features are considered, and the final decision relies on a set of rules that are based on domain-independent relationships and concept properties. In addition, the authors present a vector system, CVS, to manage compatibility among a number of ontologies. Compatibility vectors are stored in a center, encoding a

measure of distance (dissimilarity) from an original ontology to the center. Such vectors can be efficiently adjusted, so that to render help in choosing appropriate partners based on their compatibility.

B. Hu et al. [19] explore the ontology matching in a dynamic and distributed environment where on-the-fly alignments are needed. Their approach exploits imperfect consensus among heterogeneous data holders by collaborating the logic formalisms with Web repositories.

In [38], the authors design a procedure for mapping hierarchical ontologies populated with properly classified text documents. Through the combination of structural and instance-based approaches, the procedure reduces the terminological and conceptual ontology heterogeneity, and yields certain granularity and instantiation judgments about the inputs.

2) *Learning-Based Ontology Matching*: In [9], A. Doan et al. describe GLUE, which employs machine learning techniques to find semantic mappings between ontologies. A Metalearner is used to combine the predictions from both Content Learner and Name Learner; a similarity matrix is then built; and common knowledge and domain constraints are incorporated through a Relaxation Labeler. In addition, GLUE has been extended to find complex mappings.

A.B. Williams and C. Tsatsoulis [41] present their theory for learning ontologies among agents with diverse conceptualizations to improve group semantic concept search performance. The authors introduce recursive semantic context rule learning and unsupervised concept cluster integration to address the issue of how agents teach each other to interpret and integrate knowledge.

L.-K. Soh describes a framework for distributed ontology learning in a multiagent environment [35]. The objective is to improve communication and understanding among the agents while agent autonomy is still preserved. Each agent maintains a dictionary for its own experience and a translation table, and the concept learning and interpretation are based on a description vector.

[7], [8], and [30] are a series of work in ontology matching based on a Bayesian (BN) approach. The methodology is built on BayesOWL. The algorithm learns probabilities using the naive Bayes text classification technique; then these probabilities and original ontologies are translated into the BN structures; finally, the algorithm finds new mappings between concepts.

[1] presents a general method for agents using ontologies to teach each other concepts to improve their communication and thus cooperation abilities. An agent gets both positive and negative examples for a concept from other agents; it then makes use of one of its known concept learning methods to learn the concept in question, involving other agents again by taking votes in case of knowledge conflicts.

J. Madhavan et al. [27] use a corpus of schemas and mappings to augment the evidence about the schemas being matched. The algorithm exploits a corpus in two ways. It first increases the evidence about each element by including evidence from similar elements in the corpus; then it learns

statistics about elements and their relationships and uses them to infer constraints to prune candidate mappings.

[39] tackles the challenge of aligning multiple concepts simultaneously. Two statistically-grounded measures (Jaccard and LSA) are explored to build conversion rules that aggregate similar concepts, and different ways of learning and deploying the multi-concept alignment are evaluated.

In order to solve the problem of low precision resulted from ambiguous words, J. Gracia et al. [18] introduce techniques from Word Sense Disambiguation. They validate the mappings by exploring the semantics of the ontological terms involved in the matching process. They also discuss techniques to filter out mappings resulting from the incorrect anchoring of ambiguous terms.

P. Lambrix et al. [25] describe SVM-based algorithms to align ontologies using literature. The authors have discovered: (1) SVM-S and NB obtain similar results; (2) the combinations of TermWN with SVM-S and with SVM-P lead to a large gain in precision compared to TermWN and SVM-P, with still a high recall.

SOCCER [22] is a learning-based approach to reconcile ontologies. Unlike most of other approaches, SOCCER's learning process depends on ontology schema information alone. Due to the fact that instances usually have a lot less varieties than schemas, it is extremely challenging for the learning to rely on schema information only. SOCCER applies an Artificial Neural Network (ANN) to learning different weights for different semantic aspects, and then it adopts an agglomerative clustering algorithm to find out a set of equivalent concepts from heterogeneous ontologies. J. Huang et al. evaluate SOCCER by aligning two real-world biological ontologies (BiologicalProcess and Pathway).

### III. ONTOLOGICAL TECHNIQUES IN MEDICAL AND BIOLOGICAL RESEARCH

Ontological techniques have been widely applied to medical and biological research. The most successful example is the GO project [16], which is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. GO provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, as well as tools to access and process such data. The focus of GO is to describe how gene products behave in a cellular context. Besides, research has been carried out for ontology-based data integration in bioinformatics.

[4] discusses the issue of mapping concepts in GO to Unified Medical Language System (UMLS). Such mapping may allow for the exploitation of the UMLS semantic network to link disparate genes, through their annotation in GO, to unique clinical outcomes, potentially uncovering biological relationships. This study reveals the inherent difficulties in the integration of vocabularies created in different manners and by specialists in different fields, as well as the strengths of different techniques used to accomplish this integration.

J. Kohler, S. Philippi, and M. Lange [24] describe principles and methods used to implement SEMEDA (Semantic Meta Database). Database owners may use SEMEDA to provide semantically integrated access to their databases; they may collaboratively edit and maintain ontologies and controlled vocabularies as well. This work enables biologists to use SEMEDA to query the integrated databases in real time without having to know the structure or any technical details of the underlying databases. The authors aim to handle technical problems of database integration and issues related to semantics, e.g., the use of different terms for the same things, different names for equivalent database attributes, and missing links between relevant entries in different databases.

E.P. Sulman, P.S. White, and G.M. Brodeur [36] report a high-resolution integrated map of the region constructed (CompView) to identify all markers in the smallest region of overlapping deletion (SRO). A regional somatic cell hybrid panel is used to more precisely localize those markers identified in CompView as within or overlapping the region, and a sequence from clones is used to validate STS content by electronic PCR and to identify transcripts. The authors conclude that the annotation of a putative tumor suppressor locus provides a resource for further analysis of meningioma candidate genes.

The authors in [6] adopt global gene expression profiling combined with an evaluation of GO and pathway mapping tools as unbiased methods for identifying the molecular pathways and processes affected upon toxicant exposure. They use the acute effects caused by the non-genotoxic carcinogen and peroxisome proliferator (PP) diethylhexylphthalate (DEHP) in the mouse liver as a model system. By revealing that gene expression changes associated with additional biological functions, along with mechanisms by which non-genotoxic carcinogens control hepatocyte hypertrophy and proliferation, their work demonstrates that GO mapping can identify, in an unbiased manner, both known and novel DEHP-induced molecular changes in the mouse liver and is therefore a powerful approach for elucidating modes of toxicity based on toxicogenomic data.

V. Jakoniene and P. Lambrix [23] argue that during the process of retrieving and information integration from multiple biological data sources, approaches should be enhanced by ontological knowledge. They identify the different types of ontological knowledge that are available on the Web, based on which they propose an approach to use such knowledge to support integrated access to multiple biological data sources. Their work also shows that current ontology-based integration approaches only cover parts of their proposed approach.

A. Birkland and G. Yona [3] present a system, Biozon, to address the problems encountered in the integration of heterogeneous data types in biology domain. Biozon offers biologists a new knowledge resource to navigate through and explore by unifying multiple biological databases consisting of a variety of data types (such as DNA sequences, proteins, interactions, and cellular pathways). Biozon is different from previous efforts as it uses a single extensive and tightly

connected graph schema wrapped with hierarchical ontology of documents and relations. Beyond warehousing existing data, Biozon computes and stores novel derived data, similarity relationships and functional predictions for example. The integration of similarity data allows propagation of knowledge through inference and fuzzy searches.

The authors in [42] develop a computational approach to analyze the annotation of sets of molecules. The distance between any two proteins is measured as the graph similarity between their GO annotations. These distances are then clustered to highlight subsets of proteins sharing related GO annotation. By determining the distances between annotations, this methodology reveals trends and enrichment of proteins of particular functions within high-throughput datasets at a higher sensitivity than perusal of end-point annotations.

The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or ontologies. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates obstacles to integration. In order to overcome such problem, B. Smith et al. [33] describe a strategy, the Open Biomedical Ontologies (OBO) Foundry initiative. Their long-term goal is that the data generated through biomedical research should form a single, consistent, cumulatively expanding and algorithmically tractable whole. Efforts to realize this goal, which are still very much in the proving stage, reflect an attempt to walk the line between the flexibility that is indispensable to scientific advance and the institution of principles that is indispensable to successful coordination.

X. Zhou and Z. Su [44] present EasyGO, a Web server to perform GO-based functional interpretation on groups of genes or GeneChip probe sets. EasyGO makes a special contribution to the agronomical research community by supporting Affymetrix GeneChips of both crops and farm animals, and by providing stronger capabilities for results visualization and user interaction. EasyGO has the ability to uncover hidden knowledge by analyzing a group of probe sets with similar expression profiles.

J. Ye et al. [43] propose to integrate heterogeneous data for Alzheimer's disease (AD) prediction based on a kernel method. They further extend the kernel framework for selecting features (biomarkers) from heterogeneous data sources. The proposed method is applied to a collection of MRI data from 59 normal healthy controls and 59 AD patients. The MRI data is pre-processed using tensor factorization. In this study, the authors treat the complementary voxel-based data and region of interest (ROI) data from MRI as two data sources, and they attempt to integrate the complementary information by the proposed method. Experimental results show that the integration of multiple data sources leads to a considerable improvement in the prediction accuracy. Results also show that the proposed algorithm identifies biomarkers that play more significant roles than others in AD diagnosis.

[17] is an extension on the opening invited talk by C. Goble

given at the Health Care and Life Sciences Data Integration for the Semantic Web Workshop collocated with WWW2007. The authors believe that if the bioinformatics community could become better organized on only one topic, then it should be addressing the issue of identity and naming. Projects such as Bio2RDF are a step towards the provision of real time translation and harmonization of identifiers over bioscience datasets, but have yet to gain real traction. The failure to address identity will be the most likely obstacle that will prevent mashups, or any other technology or strategy, from becoming an effective integration mechanism. It is thus critical to grasp the nettle of identity management and to show how, using lightweight semantic techniques, the user can rapidly aggregate data just in time and just when it needs to be, by the user and for the user.

A set of high-quality electronic and manual associations (annotations) of GO terms to UniProt Knowledgebase (UniProtKB) entries are provided in the Gene Ontology Annotation (GOA) project [2]. Annotations created by the project are collated with annotations from external databases to provide an extensive, publicly available GO annotation resource. Currently covering over 160,000 taxa, with greater than 32 million annotations, GOA remains the largest and most comprehensive open-source contributor to the GO Consortium (GOC) project.

In a survey paper, G.A. Thorisson, J. Muilu, and A.J. Brookes [37] discuss the challenging problem of interconnecting large sets of related information so that they can be searched and downloaded from a single portal. They look into how this has been tackled in the past for genotype-to-phenotype (G2P) data, and they also investigate how the relevant technologies are currently being improved. This work discusses some of the technical issues surrounding database development, and the recent trend towards an increased emphasis on federated database solutions, which can link independent databases through a central portal and be married with the proven benefits of traditional central databases in which related data is stored all in one place. In addition, the authors consider even more revolutionary approaches to data integration and utilization, and discuss potential challenges that need to be addressed.

[Washington09] proposes a hypothesis that ontological annotation of disease phenotypes will facilitate the discovery of new genotype-phenotype relationships within and across species. The authors apply an Entity-Quality (EQ) methodology to annotate the phenotypes of 11 gene-linked human diseases described in Online Mendelian Inheritance in Man (OMIM). In addition, four similarity metrics are utilized to compare phenotypes, and an ontology of homologous and analogous anatomical structures is developed to compare phenotypes between species. They conclude that EQ-based annotation of phenotypes, in conjunction with a cross-species ontology and a variety of similarity metrics, can identify biologically meaningful similarities between genes by comparing phenotypes alone.

The Mammalian Phenotype Ontology (MP) is a tool presented in [34] for classifying and organizing phenotypic infor-

mation related to the mouse and other mammalian species. Use of the MP ontology allows comparisons of data from diverse sources, facilitates comparisons across mammalian species, assists in identifying appropriate experimental disease models, and aids in the discovery of candidate disease genes and molecular signaling pathways.

#### IV. CONCLUDING REMARKS

While there are many challenges, much excitement has been identified in the fields of bioinformatics and medical informatics. We believe that a systematic approach that combines a variety of data sources will help us to better understand various biological functions and features, especially those closely related to different types of disease. In order to obtain the ability to integrate heterogeneous data, and thereby efficiently revolutionize the traditional medical and biological research, new methodologies built upon the increasingly pervasive cyberinfrastructure are required to conceptualize data, and acquire the “deep” knowledge out of original data thereafter. Being formal, declarative knowledge representation models, ontologies form a semantic foundation for bioinformatics and medical informatics research, and thus render great help to medical and biological researchers in better capturing hidden knowledge from large amounts of original data. Ontological techniques have been widely applied to medical and biological research. Nevertheless, there is still a need to explore innovative, cyber-enabled computing frameworks that are based on ontological techniques. These frameworks will facilitate knowledge acquisition from existing resources, assist biologists in better understanding important biological functions at different levels, and ultimately, help clinicians in making sound decisions when treating their patients.

#### REFERENCES

- [1] M. Afsharchi, B.H. Far, and J. Denzinger, “Ontology-Guided Learning to Improve Communication between Groups of Agents,” Proc. the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 06), Hakodate, Japan, May 2006.
- [2] D. Barrell, E. Dimmer, R.P. Huntley, D. Binns, C. O’Donovan, and R. Apweiler, “The GOA database in 2009—an integrated Gene Ontology annotation resource,” *Nucleic Acids Research*, 37: D396-403, 2008.
- [3] A. Birkland and G. Yona, “BIOZON: a system for unification, management and analysis of heterogeneous biological data,” *BMC Bioinformatics*, 7:70, 2006.
- [4] M.N. Cantora, I.N. Sarkara, R. Gelmana, F. Hartelb, O. Bodenreiderc, and Y.A. Lussiera, “An Evaluation of Hybrid Methods for Matching Biomedical Terminologies: Mapping the Gene Ontology to the UMLS,” *Studies in Health Technology and Informatics*, 95:62-67, 2003.
- [5] S. Castano, A. Ferrara, and S. Montanelli, “H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems,” Proc. the first VLDB International Workshop on Semantic Web and Databases (SWDB 03), Berlin, Germany, September 2003.
- [6] R.A. Currie, et al., “Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate,” *Journal of Toxicological Sciences*, 86:453-469, 2005.
- [7] Z. Ding, Y. Peng, and R. Pan, “A Bayesian Approach to Uncertainty Modeling in OWL Ontology,” Proc. the International Conference on Advances in Intelligent Systems - Theory and Applications, Luxembourg, November 2004.

- [8] Z. Ding, Y. Peng, R. Pan, and Y. Yu, "A Bayesian Methodology towards Automatic Ontology Mapping," Proc. Technical Report WS-05-01 of AAAI Workshop on Contexts and Ontologies: Theory, Practice, and Applications, Pittsburgh, PA, July 2005.
- [9] A. Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Halevy, "Learning to match ontologies on the Semantic Web," *The VLDB Journal*, vol.12, no.4, pp. 303-319, 2003, Springer-Verlag, New York, NY, USA.
- [10] A. Doan and A.Y. Halevy, "Semantic Integration Research in the Database Community: A Brief Survey," *AI Magazine*, vol.26, no.1, pp. 83-94, Spring 2005.
- [11] D. Dou, D. McDermott, and P. Qi, "Ontology Translation on the Semantic Web," Proc. the International Conference on Ontologies, Databases, and Applications of Semantics, Lecture Notes in Computer Science, Berlin: Springer-Verlag, 2003.
- [12] J. Euzenat and P. Shvaiko, *Ontology Matching*, Springer-Verlag, Berlin Heidelberg (DE), 2007.
- [13] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, "Semantic Schema Matching," Proc. the Thirteenth International Conference on Cooperative Information Systems (CoopIS 05), Agia Napa, Cyprus, November 2005.
- [14] F. Giunchiglia, M. Yatskevich, and F. McNeill, "Structure preserving semantic matching," Proc. the Second International Workshop on Ontology Matching (OM 07), Busan, Korea, November 2007.
- [15] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, "Semantic matching," *Encyclopedia of Database Systems*, pp. 2561-2566, Springer, 2009.
- [16] Gene Ontology Website, <http://www.geneontology.org/index.shtml>, January, 2010.
- [17] C. Goble and R. Stevens, "State of the nation in data integration for bioinformatics," *Journal of Biomedical Informatics*, 41:687-693, 2008.
- [18] J. Gracia, V. Lopez, M. Aquin, M. Sabou, E. Motta, and E. Mena, "Solving semantic ambiguity to improve Semantic Web based ontology matching," Proc. the Second International Workshop on Ontology Matching (OM 07), Busan, Korea, November 2007.
- [19] B. Hu, S. Dasmahapatra, and P. Lewis, "Emerging consensus in-situ," Proc. the Second International Workshop on Ontology Matching (OM 07), Busan, Korea, November 2007.
- [20] J. Huang, J. Dang, and M.N. Huhns, "Ontology Reconciliation for Service-Oriented Computing," Proc. 2006 IEEE International Conference on Services Computing (SCC 06), pp. 3-10, Chicago, IL, September 2006.
- [21] J. Huang, J. Dang, M.N. Huhns, and Y. Shao, "Ontology Alignment as a Basis for Mobile Service Integration and Invocation," *International Journal of Pervasive Computing and Communications (JPCC)*, vol.3, no.2, pp. 138-158, Emerald, 2007.
- [22] J. Huang, J. Dang, M.N. Huhns, and W.J. Zheng, "Use Artificial Neural Network to Align Biological Ontologies," *BMC Genomics* 2008, 9 (Suppl 2): S16.
- [23] V. Jakoniene and P. Lambrix, "Ontology-based integration for bioinformatics," Proc. VLDB Workshop on Ontologies-based techniques for DataBases, Trondheim, Norway, September, 2005.
- [24] J. Kohler, S. Philippi, and M. Lange, "SEMEDA: ontology based semantic integration of biological databases," *Bioinformatics*, 19(18):pp.2420-7, 2003.
- [25] P. Lambrix, H. Tan, and W. Xu, "Literature-based alignment of ontologies," Proc. the Third International Workshop on Ontology Matching (OM 08), Karlsruhe, Germany, October 2008.
- [26] J. Madhavan, P.A. Bernstein, and E. Rahm, "Generic Schema Matching with Cupid," Proc. the Twenty-seventh VLDB Conference, Roma, Italy, 2001.
- [27] J. Madhavan, P.A. Bernstein, A. Doan, and A. Halevy, "Corpus-based Schema Matching," Proc. the Twenty-first International Conference on Data Engineering (ICDE 05), Tokyo, Japan, April 2005.
- [28] S. Melnik, H. Garcia-Molina, and E. Rahm, "Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching," Proc. the Eighteenth International Conference on Data Engineering (ICDE 02), San Jose, CA, 2002.
- [29] N.F. Noy and M.A. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment," Proc. the 17th National Conference on Artificial Intelligence (AAAI 00), AAAI Press, Menlo Park, CA, USA, 2000.
- [30] R. Pan, Z. Ding, Y. Yu, and Y. Peng, "A Bayesian Network Approach to Ontology Mapping," Proc. the International Semantic Web Conference (ISWC 05), Galway, Ireland, November 2005.
- [31] P. Shvaiko and J. Euzenat, "Ten Challenges for Ontology Matching," Proc. the 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 08), LNCS 5332, Monterrey, Mexico, November 11-13, pp. 1164-1182, 2008.
- [32] M.P. Singh and M.N. Huhns, *Service-Oriented Computing - Semantics, Processes, Agents*, first ed. Chichester, England: Wiley, 2005.
- [33] B. Smith et al., "The OBO foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature Biotechnology*, 25(11):pp.1251-5, 2007.
- [34] C.L. Smith and J.T. Eppig, "The mammalian phenotype ontology: enabling robust annotation and comparative analysis," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(3):390-9, 2009.
- [35] L.-K. Soh, "Multiagent Distributed Ontology Learning," Proc. Working Notes of the second AAMAS OAS Workshop, Bologna, Italy, July 2002.
- [36] E.P. Sulman, P.S. White, and G.M. Brodeur, "Genomic annotation of the meningioma tumor suppressor locus on chromosome 1p34," *Oncogene*, 23:1014-20, 2004.
- [37] G.A. Thorisson, J. Muilu, and A.J. Brookes, "Genotype-phenotype databases: challenges and solutions for the post-genomic era," *Nature Reviews Genetics*, 10:9-18, 2009.
- [38] K. Todorov and P. Geibel, "Ontology mapping via structural and instance-based similarity measures," Proc. the Third International Workshop on Ontology Matching (OM 08), Karlsruhe, Germany, October 2008.
- [39] S. Wang, A. Isaac, L. Meij, and S. Schlobach, "Multi-concept alignment and evaluation," Proc. the Second International Workshop on Ontology Matching (OM 07), Busan, Korea, November 2007.
- [40] 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.
- [41] A.B. Williams and C. Tsatsoulis, "Diverse Web Ontologies: What Intelligent Agents Must Teach to Each Other," Proc. AAAI Spring Symposium on Intelligent Agents in Cyberspace, Stanford University, March 1999.
- [42] C. Wolting, C.J. McGlade, and D. Tritchler, "Cluster analysis of protein array results via similarity of Gene Ontology annotation," *BMC Bioinformatics*, 7:338, 2006.
- [43] J. Ye et al., "Heterogeneous data fusion for Alzheimer's disease study," Proc. the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 08), pp.1025-33, Las Vegas, Nevada, August, 2008.
- [44] X. Zhou and Z. Su, "EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species," *BMC Genomics*, 8:246, 2007.