An Overview of Methods and Tools for Biomedical Ontology Matching

Aditya Laghari

February 2, 2023

An Overview of Methods and Tools for Biomedical Ontology Matching

Aditya Laghari

Delhi Technological University (India)

Abstract
This paper presents a comprehensive analysis of current methods for biomedical ontology matching. Biomedical ontologies are an essential tool for organizing and representing biomedical knowledge, but their effectiveness depends on the ability to match concepts across different ontologies. We review several state-of-the-art methods for ontology matching, including lexical-based, structural-based, and semantic-based approaches. The analysis indicate that semantic-based methods perform better than lexical or structural-based methods. Machine learning can further improve the performance of biomedical ontology matching. The paper also discusses the challenges and future directions in biomedical ontology-matching research.

Keywords
Ontology Matching, Ontology Alignment, Biomedical ontologies

1. Introduction

Ontology matching is the process of identifying and aligning equivalent concepts from different ontologies. On the one hand, an ontology is a model of knowledge that allows an efficient representation of the facts of the real world. Representing the facts of the real world is a great help. On the other hand, ontology matching is an important part of knowledge management. It is a way for organizations to model their knowledge without the constraints of a particular standard. The majority of organizations are not interested in working with a standard to model their knowledge for two reasons: First, it is very difficult or expensive for many organizations to agree on a common standard, and second, these standards often do not meet the specific needs of all participants in the standardization process [1].

One of the most common challenges faced by data management researchers is ontology reconciliation. This is because ontologies have many applications across many industries. Finding overlapping elements in two independently designed and constructed ontologies is a great way to implement new strategies to improve current methods for data interoperability and query extension, among others. This can be accomplished by exploiting the potential for discovering overlapping entities in two independently created and evolved ontologies.

Due to the wide interest in this scientific problem, many publications have presented potential solutions to this problem. Over the last decade, several strategies and approaches have been developed [2]. All of them claim to solve the problem once and for all. The problem is that there still needs to be a way to solve the ontology matching problem. If we ask several different experts

---

Preprint under review

CEUR Workshop Proceedings (CEUR-WS.org)

about a particular use case, we may get as many different answers as there are different experts. This suggests that this is a problem that anyone can tackle. As with all challenges, finding a solution that works well in all circumstances and can be adapted to everyone's preferences will take work.

In the field of biomedicine, particular biomedical ontologies, such as SNOMED[1], the National Cancer Institute Thesaurus (NCI[2]), and the Foundational Model of Anatomy (FMA[3]), have grown in popularity and are frequently implemented in a variety of different solutions and systems. However, most approaches to matching biological ontologies have concentrated on feature engineering. There are many features, such as terminological, structural, extensional (examples of a particular notion), and external resources.

The contribution of this work is based on an extended discussion of the problem of biomedical ontology matching and a thorough review of existing approaches and methods. The rest of this work presents related works (Section 2), an overview of methods and tools (Section 3), the concluding remarks, and possible future research lines.

2. Related Works

Biomedical ontologies are formal representations of concepts and relationships in a specific biomedical science domain, such as anatomy, diseases, or drugs [3]. At the same time, biomedical ontology matching aims to enable the integration and comparison of data from different ontologies represented as electronic health records, clinical trials, and bioinformatics databases. This can support various applications, such as data integration, knowledge discovery, and decision support in biomedical research and health care. It considered being a complex task involving issues such as concept representation, representation of relationships, the granularity of concepts, and ontology structure. Different matching methods can be used, such as string-based, structural, semantic similarity, and machine learning, to improve the accuracy of the matching process. Below is an overview of the popular data types, methods, and tools in this domain.

2.1. Data

Regarding the data, it is necessary to note that there are many biomedical ontologies available; some examples are:

- **The Chemical Entities of Biological Interest (ChEBI)**[4]: an ontology that represents the names, definitions, and relationships of small chemical compounds, including drugs, metabolites, and natural products.
- **The Disease Ontology (DO)**[5]: an ontology that represents the names, definitions, and relationships of diseases and disorders, including genetic, infectious, and environmental causes.

---

[1] https://www.snomed.org/

[2] https://ncithesaurus.nci.nih.gov/ncitbrowser/

[3] https://bioportal.bioontology.org/ontologies/FMA

[4] https://www.ebi.ac.uk/chebi/

[5] https://disease-ontology.org/

- **The Foundational Model of Anatomy (FMA)**[6]: an ontology that represents the structure and organization of the human body, including the relationships between different body parts and systems.
- **Gene Ontology (GO)**[7]: an ontology representing the functions, processes, and cellular components of genes and proteins across different organisms.
- **The Human Phenotype Ontology (HPO)**[8]: An ontology that represents the different human phenotypic abnormalities in a structured manner, including signs and symptoms.
- **The National Cancer Institute's Thesaurus (NCI-THES)**[9]: an ontology that represents the names, definitions, and relationships of cancer-related concepts, including cancer types, treatments, and clinical trials.
- **SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms)**[10] is a comprehensive, multilingual, standardized clinical terminology system. It is used to consistently and precisely encode the meaning of clinical concepts, including diagnoses, symptoms, procedures, and medications.

Most of the mentioned ontologies are widely used in the healthcare industry, particularly in electronic health records and clinical decision support systems. They are also used for research, quality improvement, and public health reporting.

2.2. Methods

One of the pioneer works to address the matching problem was [4], where it was talked about how matching may be planned out methodically. These authors were the first to suggest utilizing machine learning techniques to build the most effective and efficient ensemble of matchers rather than a matching algorithm (also known as a matcher).. From this seminal work, research turned to the proposal of machine learning techniques to aggregate the basic matchers, GAOM [5] and GOAL [6], two of the earliest genetic algorithm works. The fundamental concept was that it could maximize either the precision value, recall value, or f-measure, which was a mix of both. Later research has suggested strategies from other academic fields that are very well suited to the nature of the issue [7, 8, 9, 10].

In [11], a survey on ontology meta-matching was presented with emphasis on the aggregation technologies used by the different proposals existing to date. From these initial works, solutions based on the most diverse machine learning techniques began to appear [12], ranging from heuristic techniques [13] or formal concept analysis [14] to techniques based on memetic computation [15], passing through Bayes learning [16], automated weighted aggregation [17] or even fuzzy aggregation [18]. Even particle swarm optimization [19], or logics [20]. In this way, the proposal of new matchers and individual similarity measures ceased to be the focus [21], so the community relied on the design of methods that would allow their aggregation in the best possible way. For example, the use of neural networks is gaining much traction [22]. Other aspects, such as scalability [23] and querying [24], have also been considered.

---

[6] https://bioportal.bioontology.org/ontologies/FMA

[7] http://geneontology.org/

[8] https://hpo.jax.org/app/

[9] https://ncithesaurus.nci.nih.gov/ncitbrowser/

[10] https://www.snomed.org/

2.3. Tools

In addition, there is a selection of tools at one's disposal with which to carry out this task. This list is not meant to be exhaustive; nonetheless, some of the most popular tools are as follows:

- **OBO-Edit** [25]: This is a graphical ontology editor and tool for matching ontologies. It supports various ontology matching techniques and allows users to perform manual and automated matching using string-based and semantic-based methods.
- **OntoMatch** [26]: This is a semantic matching tool designed specifically for biomedical ontologies. It uses graph-based matching and semantic similarity measures to identify correspondences between ontologies.
- **LogMap** [20]: This is a semantic matching tool that supports the alignment of large-scale ontologies. It uses string-based, semantic-based, and machine learning-based methods to identify correspondences between ontologies.

From now on, we are going to see the different alternatives that exist to carry out a matching of biomedical ontologies in an effective and efficient way.

3. Overview

Finding semantic correspondences between ontologies that pertain to the same domain but have been established by different people or teams can, in principle, be accomplished using a wide variety of techniques. There are many different approaches to matching biomedical ontologies, including the following:

- **String-based matching**: This method uses string-matching algorithms to compare the labels of concepts in different ontologies.
- **Structural matching**: This method compares the relationships between concepts in different ontologies, such as parent-child or sibling relationships.
- **Semantic similarity**: This method calculates the similarity between concepts based on their definitions, using techniques such as Latent Semantic Analysis or WordNet-based measures.
- **Hybrid matching**: This method combines multiple matching techniques, such as string-based and semantic similarity, to improve the accuracy of the matching process.
- **Machine Learning approaches**: This method uses machine learning algorithms to learn the matching patterns from the sample data and then uses the learned model for matching new ontologies.

If preferred, there are even strategies to go beyond the classic simple techniques. For example, it is possible to build ensembles using some stacking techniques. Stacking is a machine learning technique that can be used to combine the predictions of multiple semantic similarity methods. The idea behind stacking is to train a second-level model, called the meta-model, to make a final prediction based on the predictions of the base models. Here are some examples of how stacking can be applied:

- **Multi-view stacking**: In this approach, multiple methods are applied to different views of the concepts, such as their definitions, labels, and relationships. The meta-model then combines the predictions of the base models to make a final prediction [27].
- **Multi-source stacking**: In this approach, multiple methods are applied to different sources of information, such as WordNet, Wikipedia, UMLS, or any other resource of this kind. The predictions of the base models are then combined with the meta-model to make a final prediction.
- **Hybrid stacking**: In this approach, multiple methods are applied, some based on the structure of the ontologies and others based on the semantic content of the concepts. The predictions of the base models are then combined with the meta-model to make a final prediction.

It is important to note that stacking methods can increase the accuracy of the ontology matching process, but it also increases the complexity of the model. Therefore, careful evaluation and selection of the base models and meta-model are essential for the success of the stacking approach.

Last, we can see matcher aggregation as a technique used to combine the results of multiple matchers and produce a single overall score. This technique is complementary to the ones mentioned above and is used when there are multiple ways to calculate the likeness of two biomedical concepts. In situations where it could be very complicated to clarify which measure is best, one possible strategy would be based on the basic idea behind the matcher aggregation. This idea is based on taking the individual scores produced by different methods and combining them to produce a single overall score. There are several ways to aggregate the different scores, such as:

- **Averaging:** The most simple approach takes the average individual similarity scores.
- **Weighted averaging:** Similar to averaging, each similarity measure is assigned a weight, representing its importance or reliability, which is multiplied by the similarity score before the averaging.
- **Rank-based aggregation:** This approach ranks the individual similarity scores and then combines them to produce a single overall score.
- **Machine learning-based aggregation:** This approach uses machine learning algorithms to learn how to combine the individual similarity scores based on a set of labeled examples.
- **Dempster-Shafer theory-based aggregation:** This approach uses the Dempster-Shafer theory to combine the individual similarity scores into a single overall score [28].
- **Fuzzy logic-based aggregation:** This approach uses fuzzy logic to combine the individual similarity scores into a single overall score.

It is widely acknowledged that matcher aggregation can improve the performance of biomedical ontology matching by considering the strengths and weaknesses of different measures and combining them to produce a more accurate overall score. However, caution should be exercised because such techniques are computationally costly. This is because training and deploying an optimal configuration demand a substantial amount of both time and resources.

4. Conclusions

This paper has discussed the problem of biomedical ontology matching and presented a review of existing approaches and methods. We have highlighted the importance of ontology matching in the biomedical domain and the challenges that arise in the process. We have seen how biomedical ontology matching is essential for integrating different biomedical ontologies. The methods and tools discussed above can be used to identify correspondences between ontologies. However, the final choice depends on the requirements of the matching task.

We have also identified some limitations in current approaches and highlighted the need for further research. We have seen how state-of-the-art methods combine machine learning and domain knowledge with improving the performance of ontology matching. Several approaches have the potential to be applied in a variety of biomedical ontology matching tasks.

As future work, it is necessary to do additional research in order to enhance the performance of ontology matching methods and overcome the limitations of existing approaches. In subsequent work, one of our goals is to investigate the many assessment criteria that may be used to measure the effectiveness of ontology matching algorithms that are already in place.

References

[1] J. Martinez-Gil, I. Navas-Delgado, J. F. Aldana-Montes, Maf: An ontology matching framework, J. Univers. Comput. Sci. 18 (2012) 194–217. URL: https://doi.org/10.3217/jucs-018-02-0194. doi:10.3217/jucs-018-02-0194.

[2] A. Annane, Z. Bellahsene, F. Azouaou, C. Jonquet, Selection and combination of heterogeneous mappings to enhance biomedical ontology matching, in: E. Blomqvist, P. Ciancarini, F. Poggi, F. Vitali (Eds.), Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, volume 10024 of *Lecture Notes in Computer Science*, 2016, pp. 19–33. URL: https://doi.org/10.1007/978-3-319-49004-5_2. doi:10.1007/978-3-319-49004-5_2.

[3] S. de Coronado, M. W. Haber, N. Sioutos, M. S. Tuttle, L. W. Wright, NCI thesaurus: Using science-based terminology to integrate cancer research results, in: M. Fieschi, E. W. Coiera, J. Y. Li (Eds.), MEDINFO 2004 - Proceedings of the 11th World Congress on Medical Informatics, San Francisco, California, USA, September 7-11, 2004, volume 107 of *Studies in Health Technology and Informatics*, IOS Press, 2004, pp. 33–37. URL: https://doi.org/10.3233/978-1-60750-949-3-33. doi:10.3233/978-1-60750-949-3-33.

[4] I. F. Cruz, C. Stroe, C. Pesquita, F. M. Couto, V. Cross, Biomedical ontology matching using the agreementmaker system, in: O. Bodenreider, M. E. Martone, A. Ruttenberg (Eds.), Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, NY, USA, July 26-30, 2011, volume 833 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2011. URL: http://ceur-ws.org/Vol-833/paper51.pdf.

[5] J. Wang, Z. Ding, C. Jiang, GAOM: genetic algorithm based ontology matching, in: Proceedings of The 1st IEEE Asia-Pacific Services Computing Conference, APSCC 2006, December 12-15, 2006, Guangzhou, China, IEEE Computer Society, 2006, pp. 617–620. URL: https://doi.org/10.1109/APSCC.2006.59. doi:10.1109/APSCC.2006.59.

[6] J. Martinez-Gil, E. Alba, J. F. Aldana-Montes, Optimizing ontology alignments by using genetic algorithms, in: C. Guéret, P. Hitzler, S. Schlobach (Eds.), Proceedings of the First International Workshop on Nature Inspired Reasoning for the Semantic Web, Karlsruhe, Germany, October 27, 2008, volume 419 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2008. URL: http://ceur-ws.org/Vol-419/paper2.pdf.

[7] X. Xue, Y. Wang, W. Hao, Using MOEA/D for optimizing ontology alignments, Soft Comput. 18 (2014) 1589–1601. URL: https://doi.org/10.1007/s00500-013-1165-9. doi:`10.1007/s00500-013-1165-9`.

[8] X. Xue, Y. Wang, W. Hao, Using MOEA/D for optimizing ontology alignments, Soft Comput. 18 (2014) 1589–1601. URL: https://doi.org/10.1007/s00500-013-1165-9. doi:`10.1007/s00500-013-1165-9`.

[9] X. Xue, Y. Wang, A. Ren, Optimizing ontology alignment through memetic algorithm based on partial reference alignment, Expert Syst. Appl. 41 (2014) 3213–3222. URL: https://doi.org/10.1016/j.eswa.2013.11.021. doi:`10.1016/j.eswa.2013.11.021`.

[10] X. Xue, Y. Wang, Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio, Artif. Intell. 223 (2015) 65–81. URL: https://doi.org/10.1016/j.artint.2015.03.001. doi:`10.1016/j.artint.2015.03.001`.

[11] J. Martinez-Gil, J. F. Aldana-Montes, An overview of current ontology meta-matching solutions, Knowl. Eng. Rev. 27 (2012) 393–412. URL: https://doi.org/10.1017/S0269888912000288. doi:`10.1017/S0269888912000288`.

[12] D. Ritze, H. Paulheim, Towards an automatic parameterization of ontology matching tools based on example mappings, in: P. Shvaiko, J. Euzenat, T. Heath, C. Quix, M. Mao, I. F. Cruz (Eds.), Proceedings of the 6th International Workshop on Ontology Matching, Bonn, Germany, October 24, 2011, volume 814 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2011. URL: http://ceur-ws.org/Vol-814/om2011_Tpaper4.pdf.

[13] J. Martinez-Gil, J. F. Aldana-Montes, Evaluation of two heuristic approaches to solve the ontology meta-matching problem, Knowl. Inf. Syst. 26 (2011) 225–247. URL: https://doi.org/10.1007/s10115-009-0277-0. doi:`10.1007/s10115-009-0277-0`.

[14] M. Zhao, S. Zhang, W. Li, G. Chen, Matching biomedical ontologies based on formal concept analysis, J. Biomed. Semant. 9 (2018) 11:1–11:27. URL: https://doi.org/10.1186/s13326-018-0178-9. doi:`10.1186/s13326-018-0178-9`.

[15] G. Acampora, P. Avella, V. Loia, S. Salerno, A. Vitiello, Improving ontology alignment through memetic algorithms, in: FUZZ-IEEE 2011, IEEE International Conference on Fuzzy Systems, Taipei, Taiwan, 27-30 June, 2011, Proceedings, IEEE, 2011, pp. 1783–1790. URL: https://doi.org/10.1109/FUZZY.2011.6007517. doi:`10.1109/FUZZY.2011.6007517`.

[16] O. Sváb, V. Svátek, Combining ontology mapping methods using bayesian networks, in: P. Shvaiko, J. Euzenat, N. F. Noy, H. Stuckenschmidt, V. R. Benjamins, M. Uschold (Eds.), Proceedings of the 1st International Workshop on Ontology Matching (OM-2006) Collocated with the 5th International Semantic Web Conference (ISWC-2006), Athens, Georgia, USA, November 5, 2006, volume 225 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2006. URL: http://ceur-ws.org/Vol-225/paper21.pdf.

[17] M. Gulic, B. Vrdoljak, M. Banek, Cromatcher: An ontology matching system based on automated weighted aggregation and iterative final alignment, J. Web Semant. 41 (2016) 50–71. URL: https://doi.org/10.1016/j.websem.2016.09.001. doi:`10.1016/j.websem.2016.`

09.001.

[18] J. Martinez-Gil, Coto: A novel approach for fuzzy aggregation of semantic similarity measures, Cogn. Syst. Res. 40 (2016) 8–17. URL: https://doi.org/10.1016/j.cogsys.2016.01.001. doi:10.1016/j.cogsys.2016.01.001.

[19] J. Bock, J. Hettenhausen, Discrete particle swarm optimisation for ontology alignment, Inf. Sci. 192 (2012) 152–173. URL: https://doi.org/10.1016/j.ins.2010.08.013. doi:10.1016/j.ins.2010.08.013.

[20] E. Jiménez-Ruiz, B. C. Grau, Logmap: Logic-based and scalable ontology matching, in: L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. F. Noy, E. Blomqvist (Eds.), The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I, volume 7031 of *Lecture Notes in Computer Science*, Springer, 2011, pp. 273–288. URL: https://doi.org/10.1007/978-3-642-25073-6_18. doi:10.1007/978-3-642-25073-6_18.

[21] P. Kolyvakis, A. Kalousis, D. Kiritsis, Deepalignment: Unsupervised ontology matching with refined word vectors, in: M. A. Walker, H. Ji, A. Stent (Eds.), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers), Association for Computational Linguistics, 2018, pp. 787–798. URL: https://doi.org/10.18653/v1/n18-1072. doi:10.18653/v1/n18-1072.

[22] P. Kolyvakis, A. Kalousis, B. Smith, D. Kiritsis, Biomedical ontology alignment: an approach based on representation learning, J. Biomedical Semantics 9 (2018) 21:1–21:20. URL: https://doi.org/10.1186/s13326-018-0187-8. doi:10.1186/s13326-018-0187-8.

[23] G. Diallo, An effective method of large scale ontology matching, J. Biomed. Semant. 5 (2014) 44. URL: https://doi.org/10.1186/2041-1480-5-44. doi:10.1186/2041-1480-5-44.

[24] A. L. Paoletti, J. Martinez-Gil, K. Schewe, Top-k matching queries for filter-based profile matching in knowledge bases, in: S. Hartmann, H. Ma (Eds.), Database and Expert Systems Applications - 27th International Conference, DEXA 2016, Porto, Portugal, September 5-8, 2016, Proceedings, Part II, volume 9828 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 295–302. URL: https://doi.org/10.1007/978-3-319-44406-2_23. doi:10.1007/978-3-319-44406-2_23.

[25] J. Day-Richter, M. A. Harris, M. Haendel, G. O. O.-E. W. Group, S. Lewis, Obo-edit—an ontology editor for biologists, Bioinformatics 23 (2007) 2198–2200.

[26] A. Bhattacharjee, H. Jamil, Ontomatch: A monotonically improving schema matching system for autonomous data integration, in: 2009 IEEE International Conference on Information Reuse & Integration, IEEE, 2009, pp. 318–323.

[27] W. Li, X. Duan, M. Wang, X. Zhang, G. Qi, Multi-view embedding for biomedical ontology matching, in: P. Shvaiko, J. Euzenat, E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn (Eds.), Proceedings of the 14th International Workshop on Ontology Matching co-located with the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, October 26, 2019, volume 2536 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 13–24. URL: http://ceur-ws.org/Vol-2536/om2019_LTpaper2.pdf.

[28] P. Besana, A framework for combining ontology and schema matchers with dempster-shafer, in: Proceedings of the 1st International Conference on Ontology Matching-Volume 225, 2006, pp. 196–200.