

Teorias e Ferramentas de apoio a Reparação de Alinhamento de Ontologias

Miriam O. dos Santos¹, Carlos Eduardo R. Mello¹, Tadeu M. de Classe¹

¹ Programa de Pós-Graduação em Informática
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro – Brasil

{miriam.santos, mello, tadeu.classe}@uniriotec.br

Abstract. *Information sharing and interoperability are crucial in several knowledge areas. A useful tool for this is ontologies, which allow aligning terms from the same domain, aiming at information interoperability. However, automatic alignment processes do not always produce satisfactory results, requiring domain specialists to correct inconsistencies. This work aims to support specialists in the process of repairing and reviewing the alignment of ontologies, through information visualization techniques, in a collaborative environment. This research uses Design Science Research Methodology to conceive an information system that is able to support specialists in the process of repair and review of ontologies, allowing to analyze how and which visualization techniques were more efficient. The evaluation of the proposal will be made through qualitative and quantitative approaches, collecting metrics for repairing ontologies and users' navigation log, in addition to a survey with the participants' response.*

Resumo. *O compartilhamento e a interoperabilidade de informações são cruciais em várias áreas de conhecimento. Uma ferramenta útil para isso são as ontologias, as quais permitem alinhar termos de um mesmo domínio, visando interoperabilidade de informação. Contudo, processos de alinhamento automáticos nem sempre produzem resultados satisfatórios, necessitando de especialistas de domínio para corrigir as inconsistências. Este trabalho objetiva apoiar os especialistas no processo de reparação e revisão de alinhamento de ontologias, através de técnicas de visualização de informação, em um ambiente colaborativo. Esta pesquisa usa Design Science Research Methodology para conceber um sistema de informação que seja capaz de apoiar os especialistas no processo de reparação e revisão de ontologias, permitindo analisar como e quais técnicas de visualização foram mais eficientes. A avaliação da proposta será feita através de abordagens quali-quantitativas, coletando métricas de reparação de ontologias e log de navegação dos usuários, além de um survey com a resposta dos participantes.*

1. INTRODUÇÃO

Ontologias são comumente empregadas em Sistemas de Informação para gerenciamento e compartilhamento de conhecimento [Nguyen et al. 2019]. Ontologias descrevem um domínio de interesse (área de conhecimento), a especificação de significado dos termos que compõe este domínio e como estes termos se relacionam,

sendo o fator principal em sua estrutura a organização hierárquica entre seus conceitos [Shvaiko and Euzenat 2011][Guarino et al. 2009][Euzenat et al. 2007].

Ontologias são utilizadas na integração, busca e análise de dados em diferentes fontes de dados, sendo comum existir ontologias de um mesmo domínio modeladas de formas distintas [Harrow et al. 2019][Banouar and Raghay 2016], desta forma, há a necessidade de realizar um processo de alinhamento entre estas ontologias [Dragisic et al. 2017]. O Alinhamento de Ontologias (*Ontology Matching*), é o processo de apontar relacionamentos ou correspondências entre entidades de diferentes ontologias, cujo objetivo é a redução da heterogeneidade semântica [da Silva et al. 2016][Euzenat et al. 2007]. A saída deste processo é um alinhamento que é composto por um conjunto de correspondências ou mapeamentos entre duas ou mais ontologias [Euzenat et al. 2007].

Atualmente existem diversas iniciativas suportadas por grandes laboratórios, como o *European Molecular Biology Laboratory EMBL-EBI*¹, *Alliance of Genome Resources*², *Monarch Initiative*³, *Pistoia Alliance*⁴, que visam a integração de dados clínicos e dados de pesquisas de diferentes partes do mundo, iniciativas estas que colaboram para difusão da descoberta de novas doenças e medicamentos [Harrow et al. 2017][Schofield et al. 2010] e evidenciam a importância do emprego de ontologias a problemas reais [Harrow et al. 2019] e [Banouar and Raghay 2016].

É fato atestado por diversos trabalhos que um alinhamento totalmente automático ainda não é viável [Li et al. 2019] [Dragisic et al. 2016] e pode ser ainda mais desafiador em determinados domínios de conhecimento onde alguns tipos de estratégias de alinhamento não possuem a mesma eficácia [Dragisic et al. 2017], problema ocasionado por exemplo, por sistemas de alinhamento que utilizam estratégias baseadas em *String* que buscam a similaridade entre termos pela semelhança entre as suas cadeias de texto [Cheatham and Hitzler 2013] ou ainda por possuírem ontologias muito grandes e complexas como é o caso das ontologias biomédicas [Faria et al. 2013].

Quando há necessidade de intervenção manual pelo especialista do domínio, este tem a tarefa de reparar o alinhamento gerado confirmando ou removendo correspondências existentes ou incluindo novas correspondências, a fim de garantir uma melhor precisão e cobertura nos mapeamentos apontados [Faria et al. 2013] [Meilicke 2011].

Podem ser citadas algumas iniciativas existentes na literatura que embora sejam referência utilizando estratégias automáticas ou semi-automáticas, não oferecem suporte adequado ao especialista do domínio na etapa de reparação, pois basicamente se limitam a exibir os mapeamentos gerados. São elas: *COMA- Combination of schema matching approaches* [Massmann et al. 2011][Saruladha et al. 2011][Euzenat et al. 2007]; *ASMOV - Automated Semantic Matching of Ontologies with Semantic validation* [Meilicke 2011]; *AML-Agreement Maker Light* [Faria et al. 2013][Cruz et al. 2009].

Desta forma, este trabalho visa apoiar o especialista do domínio em um processo posterior ao alinhamento automático, denominado reparação de alinhamento de ontolo-

¹EMBL-EBI, <https://www.ebi.ac.uk/spot/oxo/search>

²The Alliance, <https://www.alliancegenome.org/>

³Monarch, <https://monarchinitiative.org>

⁴Pistoia Alliance, <https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping/>

gias disponibilizando múltiplas visões dos mapeamentos e ontologias através de diferentes recursos visuais.

O restante deste artigo está organizado como se segue. Na Seção 2, é apresentado um breve referencial teórico sobre os conceitos utilizados nesta pesquisa e a identificação do problema de pesquisa. Na Seção 3, Design da Pesquisa e Metodologia, é apresentada a proposta de solução e avaliação a partir da metodologia *Design Science Research Methodology*. Na Seção 4, são apresentadas as atividades realizadas e a Seção 5 apresenta uma breve conclusão e contribuições pretendidas.

2. APRESENTAÇÃO DO PROBLEMA

Muitas abordagens estão sendo propostas para encontrar um alinhamento preciso e completo, porém a inexistência de um processo sem falhas incorre na necessidade da execução de um processo posterior chamado Reparação de Alinhamento de Ontologias (RAO), do inglês, *Ontology Alignment Repair* [Pesquita et al. 2013].

RAO é a tarefa de aperfeiçoar um alinhamento gerado, incluindo ou removendo correspondências ou ainda enfraquecendo relações inicialmente apontadas como uma relação de equivalência [Euzenat 2014]. Outras abordagens buscam identificar classes que levam a inconsistência lógica, nestes casos ocorre a remoção destas classes para que o alinhamento final se torne coerente [Pesquita et al. 2013]. A implementação dessas ações pode ser realizada seguindo uma estratégia automática, semi-automática ou manual [Meilicke 2011].

Vários trabalhos apontam que o processo de alinhamento de ontologias é uma tarefa complexa que não pode ser totalmente automatizada. Considera-se que é necessária a intervenção humana para verificar ou ajustar mapeamentos produzidos pelos algoritmos considerando que a combinação da validação do usuário com os métodos automáticos levam a melhores resultados do que os métodos puramente automáticos [Li et al. 2015][Meilicke et al. 2008][Falconer et al. 2006].

Em [Falconer et al. 2006] é enfatizado que muitos trabalhos acadêmicos se concentram no algoritmo em si, ignorando amplamente a questão do suporte cognitivo para usuários na tarefa de analisar mapeamentos propostos pelos algoritmos e na criação de novos mapeamentos. Deve-se levar em conta um conjunto de requisitos, sendo o design da interface para apoio ao especialista do domínio apontado como ponto crucial na eficiência destes sistemas [Granitzer et al. 2010].

3. DESIGN DA PESQUISA E METODOLOGIA

Para um melhor entendimento e validação das reais necessidades, problemas e avanços no processo de reparação de alinhamento de ontologias, optou-se por utilizar *Design Science Research Methodology* (DSRM). A DSRM é uma metodologia baseada na *Design Science Research* que considera o uso de artefatos como proposta de solução para resolução de problemas, sendo projetados com base em conjecturas teóricas e requisitos de design. A avaliação do produto realimenta tais conjecturas e contribuem para o crescimento do conhecimento científico-tecnológico [Peffers et al. 2007]. A DSRM vem sendo empregada por pesquisadores de Sistemas de Informação [Nguyen et al. 2019][Peffers et al. 2007], os quais enxergam nesta metodologia um grande potencial para aplicação em solução

de problemas do mundo real como em pesquisas que envolvem o uso de ontologias [Nguyen et al. 2019].

Para projetar sistemas eficazes para RAO é necessário uma compreensão mais profunda da interação entre a ferramenta, o usuário e o processo [Falconer et al. 2007]. Corroborando com esta proposta de solução identificou-se trabalhos que buscam entender o processo, ações ou sequências de ações executadas para realizar um mapeamento [Falconer et al. 2007] e trabalhos que enfatizam a importância de serem despendidos esforços para a melhoria de técnicas de visualização [Fu et al. 2013].

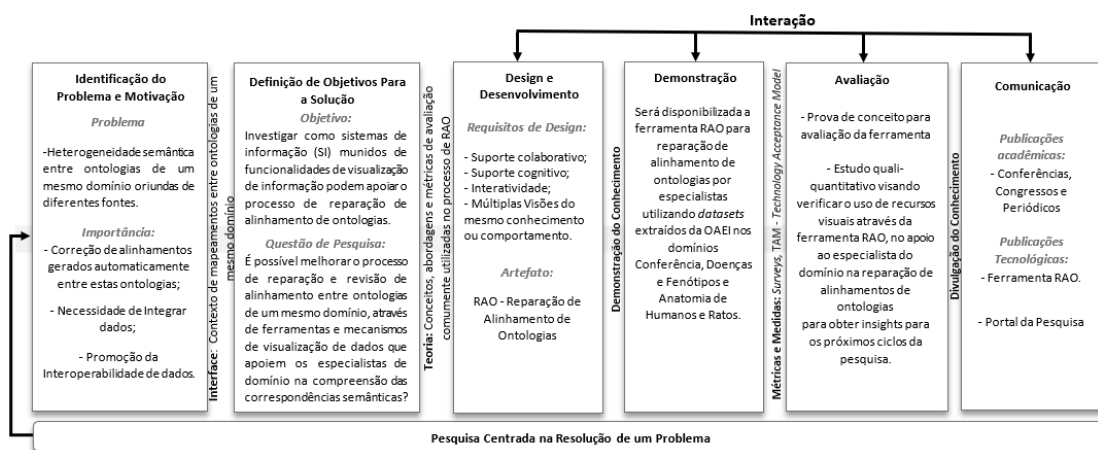


Figura 1. Design Science Research Methodology (Adaptado [Peffers et al. 2007])

Desta maneira, as etapas metodológicas desta pesquisa estão retratadas na Figura 1, e são resumidamente explicadas a partir de nossa proposta de solução:

3.1. Identificação do problema e motivação

Ontologias são modeladas de diferentes formas e possuem diferentes origens, o que ocasiona heterogeneidade semântica mesmo em ontologias de um mesmo domínio. Desta forma, Sistemas de Informação para alinhamento de ontologias buscam gerar automaticamente (ou semi-automaticamente) correspondências entre conceitos de ontologias distintas. Por outro lado, há a necessidade de execução de uma etapa manual pelo especialista do domínio para corrigir os alinhamentos gerados, desta forma, identificou-se a necessidade de investigar como sistemas de informação podem apoiar o especialista do domínio no processo de RAO [Granitzer et al. 2010].

Espera-se que as análises advindas dos resultados do estudo de caso proposto na pesquisa contribuam para melhorar o processo de reparação e revisão de alinhamento de ontologias que impactará no aumento de interoperabilidade de dados de diversos domínios possibilitando a integração de dados de pesquisas de diferentes partes do mundo.

3.2. Definição de objetivos para a solução

Esta pesquisa tem por objetivo investigar como sistemas de informação (SI) munidos de funcionalidades de visualização de informação podem apoiar o processo de reparação de alinhamento de ontologias. Sendo a questão de pesquisa endereçada neste trabalho: "É possível melhorar o processo de reparação e revisão de alinhamento entre ontologias de um mesmo domínio, através de ferramentas e mecanismos de visualização de dados que apoiem os especialistas de domínio na compreensão das correspondências semânticas?".

3.3. Design e Desenvolvimento

Este trabalho se apoia em técnicas de visualização de informação, em requisitos colaborativos, cognitivos e de interatividade para tentar melhorar o processo de reparação e revisão de alinhamento de ontologias pelo especialista do domínio, estes requisitos foram inspirados em trabalhos como [Li et al. 2015][Fu et al. 2013][Granitzer et al. 2010][Falconer et al. 2006].

Como artefato foi desenvolvido e disponibilizado o sistema RAO (Reparação de Alinhamento de Ontologias), com acesso através da web⁵. Através desta ferramenta especialistas no domínio podem reparar alinhamentos apontados automaticamente e ao fim deste ciclo espera-se que eles tenham um melhor entendimento do processo e ações realizadas, como também investigado em [Falconer et al. 2007]. Como requisito de design, estão sendo disponibilizadas múltiplas visões através de diferentes recursos visuais, [Falconer et al. 2006] apud [Richer and Clancey 1985] reforça este entendimento e diz que fornecer múltiplas visões de um mesmo conhecimento ou comportamento pode ajudar o usuário a entender um sistema complexo.

3.4. Demonstração

Dois estudos de caso estão previstos para buscar validar a questão de pesquisa proposta; Em ambos os estudos o único requisito para participação é que os voluntários estejam familiarizados com o domínio de conhecimento das ontologias a serem mapeadas, não havendo necessidade de conhecimento prévio sobre ontologias, alinhamento ou reparação de ontologias. Desta forma, haverá um primeiro *survey* onde o usuário indicará o nível de experiência no domínio a ser explorado, o que possibilitará, posteriormente, identificar se resultados de usuários com mais experiência, difere dos resultados de usuários com menos experiência.

O Primeiro estudo de caso será no domínio **Congressos e Conferências**, escolhido por ser um contexto mais abrangente e com isto podendo atingir um público de professores, pesquisadores e estudantes de pós-graduação, por exemplo. O segundo estudo de caso será aplicado no contexto de biomedicina, nos domínios **Doenças e Fenótipos e Anatomia Humana e de Ratos** o qual irá requerer um público mais específico. As bases de dados dos estudos (ontologias) foram extraídas da *OAEI - Ontology Alignment Evaluation Initiative* ⁶.

3.5. Avaliação

O Artefato RAO será avaliado por meio de uma prova de conceito usando o modelo de aceitação tecnológica **TAM - Technology Acceptance Model** [Classe et al. 2018] apud [Davis 1986] com a participação de voluntários que farão o papel do especialista do domínio.

A Estratégia proposta será avaliada a partir de uma Metodologia **quali-quantitativa (mista)** [Recker 2012]. Durante a utilização da ferramenta RAO dois tipos de análises quantitativas serão geradas, utilizando métricas de avaliação tradicionais como: **precisão, cobertura e medida-f** que são baseadas na comparação do resultado do alinhamento gerado com o alinhamento de referência [Euzenat et al. 2007], utilizando outras

⁵RAO, <http://dbemapp.uniriotec.br>

⁶OAEI, <http://oaei.ontologymatching.org/2019/>

métricas como *taxa de sucesso e taxa de erro* [Fu et al. 2013] que leva em consideração apenas o que foi respondido e não todos os resultados possíveis. Também será considerado o *tempo gasto na execução da tarefa* [Dimara et al. 2017].

A partir da coleta automática de logs gerados e da navegação na ferramenta será realizada *Análise Exploratória de Dados (EDA)* [Komorowski et al. 2016] seguida da aplicação de *Técnicas de Mineração de Dados*, como: clusterização, classificação e regras de associação [Goldschmidt and Passos 2005].

Para análise qualitativa, os voluntários serão convidados a responder um novo *survey* com foco na experiência do usuário a partir da utilização do artefato com objetivo de investigar se os requisitos implementados atenderam ao objetivo proposto e para análise de requisitos qualitativos de *satisfação, confiança e facilidade de uso* [Dimara et al. 2017].

3.6. Comunicação

Os resultados do estudo serão divulgados através de Publicações acadêmicas em Conferências, Congressos e Periódicos e Publicações Tecnológicas, além do próprio portal da Pesquisa ⁷.

4. ATIVIDADES REALIZADAS

Revisão sistemática da literatura envolvendo temas como Reparação de Alinhamento de Ontologias, DSRM, IHC, Visualização de Dados e Sistemas Colaborativos; Criação do portal para divulgação do experimento e construção da ferramenta RAO.

5. CONCLUSÃO

Este trabalho visa investigar o uso de técnicas de visualização para que estas apoiem o processo de reparação e revisão de alinhamento de ontologias, buscando melhorar a compreensão das ontologias envolvidas no processo, possibilitando identificação de conceitos correspondentes, promovendo uma rápida revisão e reparação de correspondências geradas, por especialistas do domínio individualmente ou em grupos de trabalho.

Para tratar o problema de pesquisa é sugerida uma abordagem baseada em DSRM, a partir do design e disponibilização de um artefato denominado RAO a fim de apoiar o processo de reparação e revisão de alinhamento de ontologias pelo especialista do domínio por meio de técnicas de visualização de informação.

Com este trabalho é esperado que as seguintes contribuições sejam alcançadas i) criação de modelos e ferramentas que sejam úteis para construção, elaboração e qualidade no processo de reparação de alinhamento de ontologias possibilitando interoperabilidade de dados em diversos domínios aplicados a problemas reais. ii) Compreensão do processo realizado pelo especialista do domínio até a tomada de decisão de quais mapeamentos deve-se manter, remover ou incluir. iii) Associação de outras áreas como IHC, Visualização de dados e Sistemas Colaborativos para melhoria de requisitos aplicados a este problema.

⁷Portal RAO, <http://dbemapp.uniriotec.br>

Referências

- Banouar, O. and Raghay, S. (2016). Interoperability of information systems through ontologies: State of art. *International Journal of Computer Science and Information Security*, 14(8):392.
- Cheatham, M. and Hitzler, P. (2013). String similarity metrics for ontology alignment. In *International semantic web conference*, pages 294–309. Springer.
- Classe, T., Araujo, R., and Xexéo, G. B. (2018). Process model game design: Uma ferramenta para apoio a sistematização de design de jogos digitais baseados em processos de negócio. *English title: Process Model Game Design: A Tool to Support the Systematization of Digital Games Based on Business Process*. In: *XVII Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames 2018)*.
- Cruz, I. F., Antonelli, F. P., and Stroe, C. (2009). Agreementmaker: efficient matching for large real-world schemas and ontologies. *Proceedings of the VLDB Endowment*, 2(2):1586–1589.
- da Silva, J., Baião, F. A., and Revoredo, K. (2016). Alinhamento interativo de ontologias usando anti-padrões de alinhamento: Um primeiro experimento. In *Anais do XII Simpósio Brasileiro de Sistemas de Informação*, pages 208–215. SBC.
- Davis, F. (1986). *A technology acceptance model for empirically testing new end-user information systems: theory and results*. PhD thesis, MIT Sloan School of Management, MIT Sloan School of Management.
- Dimara, E., Bezerianos, A., and Dragicevic, P. (2017). Conceptual and methodological issues in evaluating multidimensional visualizations for decision support. *IEEE transactions on visualization and computer graphics*, 24(1):749–759.
- Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jiménez-Ruiz, E., and Pesquita, C. (2016). User validation in ontology alignment. In *International Semantic Web Conference*, pages 200–217. Springer.
- Dragisic, Z., Ivanova, V., Li, H., and Lambrix, P. (2017). Experiences from the anatomy track in the ontology alignment evaluation initiative. *Journal of biomedical semantics*, 8(1):56.
- Euzenat, J. (2014). First experiments in cultural alignment repair (extended version). In *European Semantic Web Conference*, pages 115–130. Springer.
- Euzenat, J., Shvaiko, P., et al. (2007). *Ontology matching*, volume 18. Springer.
- Falconer, S. M., Noy, N. F., and Storey, M.-A. D. (2006). Towards understanding the needs of cognitive support for ontology mapping. *Ontology Matching*, 225.
- Falconer, S. M., Noy, N. F., and Storey, M.-A. D. (2007). Ontology mapping-a user survey. In *OM*. Citeseer.
- Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., and Couto, F. M. (2013). The agreementmakerlight ontology matching system. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, pages 527–541. Springer.

- Fu, B., Noy, N. F., and Storey, M.-A. (2013). Indented tree or graph? a usability study of ontology visualization techniques in the context of class mapping evaluation. In *International Semantic Web Conference*, pages 117–134. Springer.
- Goldschmidt, R. and Passos, E. (2005). *Data mining: um guia prático*. Gulf Professional Publishing.
- Granitzer, M., Sabol, V., Onn, K. W., Lukose, D., and Tochtermann, K. (2010). Ontology alignment—a survey with focus on visually supported semi-automatic techniques. *Future Internet*, 2(3):238–258.
- Guarino, N., Oberle, D., and Staab, S. (2009). What is an ontology?. in *handbook on ontologies* (pp. 1-17).
- Harrow, I., Balakrishnan, R., Jimenez-Ruiz, E., Jupp, S., Lomax, J., Reed, J., Romacker, M., Senger, C., Splendiani, A., Wilson, J., and Woollard, P. (2019). Ontology mapping for semantically enabled applications. *Drug discovery today*, 24:2068–2075.
- Harrow, I., Jiménez-Ruiz, E., Splendiani, A., Romacker, M., Woollard, P., Markel, S., Alam-Faruque, Y., Koch, M., Malone, J., and Waaler, A. (2017). Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *Journal of biomedical semantics*, 8(1):55.
- Komorowski, M., Marshall, D., Saliccioli, J., and Crutain, Y. (2016). *Exploratory Data Analysis*, pages 185–203.
- Li, H., Dragisic, Z., Faria, D., Ivanova, V., Jiménez-Ruiz, E., Lambrix, P., and Pesquita, C. (2019). User validation in ontology alignment: functional assessment and impact. *The Knowledge Engineering Review*, 34.
- Li, Y., Stroe, C., and Cruz, I. F. (2015). Interactive visualization of large ontology matching results. In *VOILA@ ISWC*, page 37.
- Massmann, S., Raunich, S., Aumüller, D., Arnold, P., and Rahm, E. (2011). Evolution of the coma match system. In *Proceedings of the 6th International Conference on Ontology Matching-Volume 814*, pages 49–60. CEUR-WS. org.
- Meilicke, C. (2011). *Alignment incoherence in ontology matching*. PhD thesis, Universität Mannheim.
- Meilicke, C., Stuckenschmidt, H., and Tamin, A. (2008). Supporting manual mapping revision using logical reasoning. In *AAAI*, pages 1213–1218.
- Nguyen, A., Gardner, L., and Sheridan, D. (2019). Towards ontology-based design science research for knowledge accumulation and evolution. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.
- Pesquita, C., Faria, D., Santos, E., and Couto, F. M. (2013). To repair or not to repair: reconciling correctness and coherence in ontology reference alignments. In *Proc. 8th ISWC ontology matching workshop (OM), Sydney (AU), page this volume*.

- Recker, J. (2012). *Scientific research in information systems: a beginner's guide*. Springer Science & Business Media.
- Richer, M. H. and Clancey, W. J. (1985). Guidon-watch: A graphic interface for viewing a knowledge-based system. Technical report, Stanford University Department of Computer Science.
- Saruladha, K., Aghila, G., and Sathiya, B. (2011). A comparative analysis of ontology and schema matching systems. *International Journal of Computer Applications*, 34(8):14–21.
- Schofield, P. N., Gkoutos, G. V., Gruenberger, M., Sundberg, J. P., and Hancock, J. M. (2010). Phenotype ontologies for mouse and man: bridging the semantic gap. *Disease models & mechanisms*, 3(5-6):281–289.
- Shvaiko, P. and Euzenat, J. (2011). Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering*, 25(1):158–176.