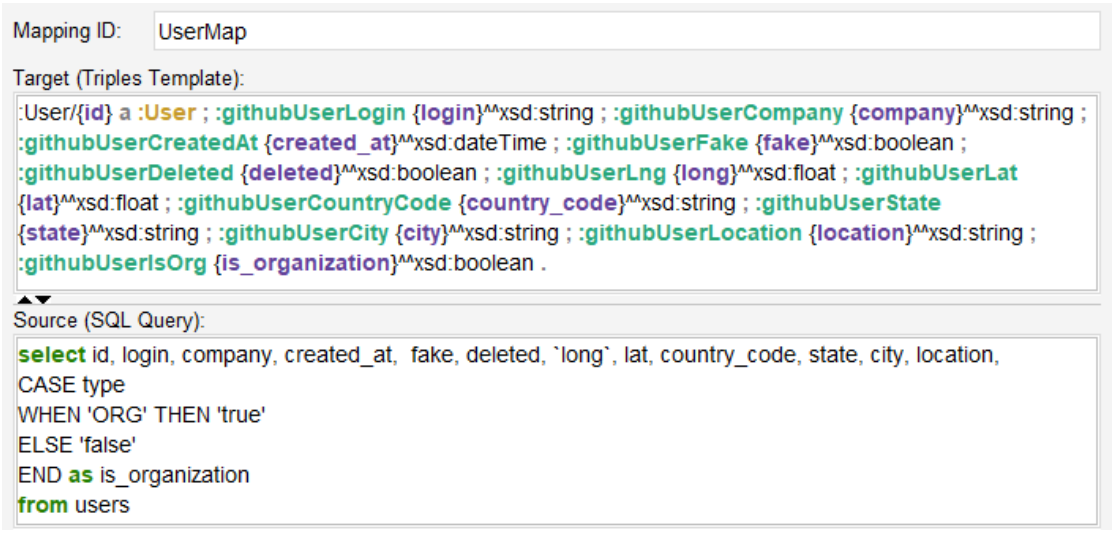

FIGURE B.19: User map

FIGURE B.20: User commit map

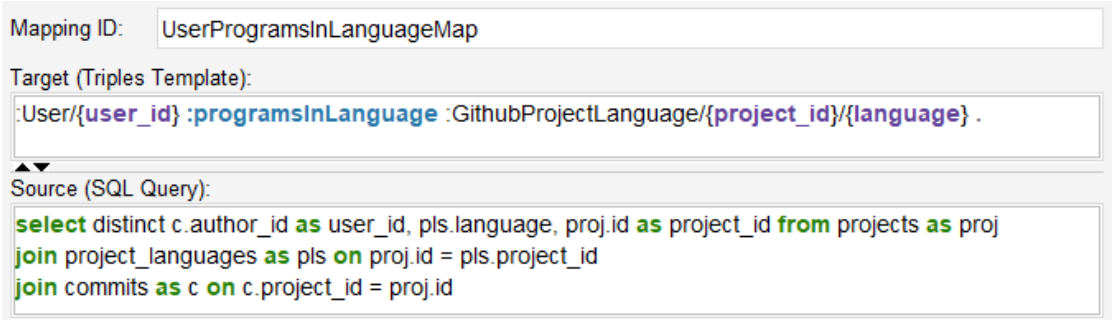

FIGURE B.21: User programming languages map

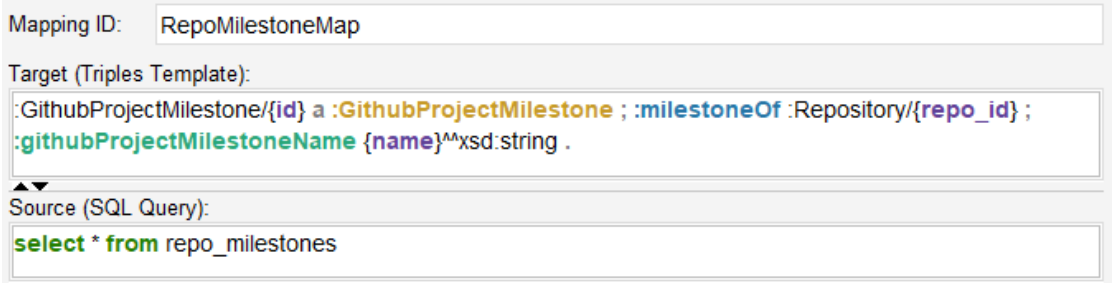

FIGURE B.22: Repository milestone map

FIGURE B.23: Watcher map

Appendix C

Ontology

This appendix contains a summary of the extended SemanGit ontology described in chapter 4.

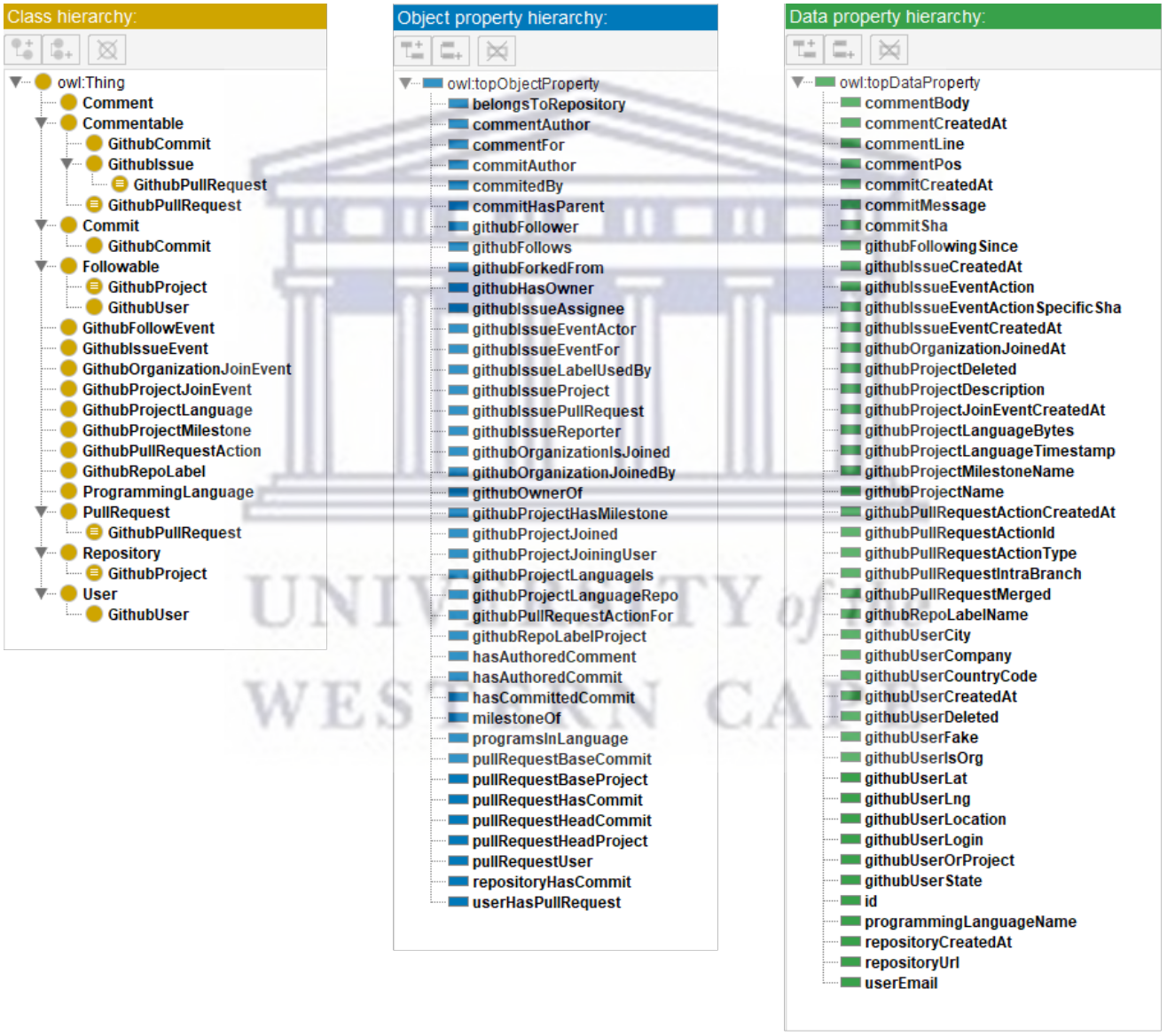

FIGURE C.1: SemanGit Ontology summary (classes and properties)

Ontology metrics:

Metrics

Axiom	413
Logical axiom count	202
Declaration axioms count	108
Class count	21
Object property count	40
Data property count	44
Individual count	0
Annotation Property count	5

Class axioms

SubClassOf	7
EquivalentClasses	2
DisjointClasses	1
GCI count	0
Hidden GCI Count	2

Object property axioms

SubObjectPropertyOf	0
EquivalentObjectProperties	0
InverseObjectProperties	6
DisjointObjectProperties	0
FunctionalObjectProperty	12
InverseFunctionalObjectProperty	4
TransitiveObjectProperty	0
SymmetricObjectProperty	0
AsymmetricObjectProperty	1
ReflexiveObjectProperty	0
IrrefexiveObjectProperty	1
ObjectPropertyDomain	40
ObjectPropertyRange	40
SubPropertyChainOf	0

Data property axioms

SubDataPropertyOf	0
EquivalentDataProperties	0
DisjointDataProperties	0
FunctionalDataProperty	0
DataPropertyDomain	44
DataPropertyRange	44

Individual axioms

ClassAssertion	0
ObjectPropertyAssertion	0
DataPropertyAssertion	0
NegativeObjectPropertyAssertion	0
NegativeDataPropertyAssertion	0
SameIndividual	0
DifferentIndividuals	0

Annotation axioms

AnnotationAssertion	103
AnnotationPropertyDomain	0
AnnotationPropertyRangeOf	0

FIGURE C.2: SemanGit Ontology metrics

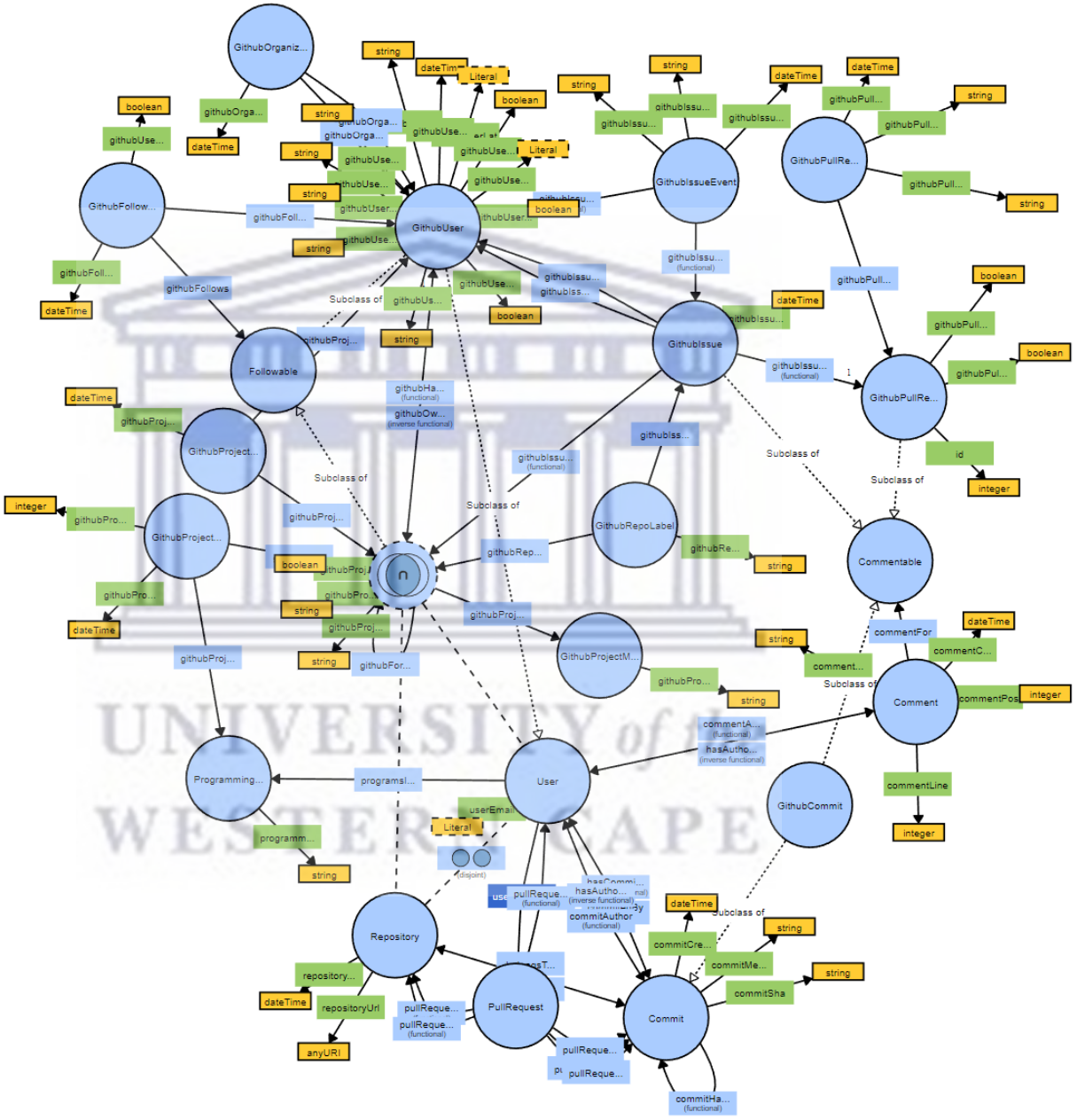

FIGURE C.3: Ontology visualisation using WebVOWL [74].

Bibliography

[1] D. Abadi, A. Ailamaki, D. Andersen, P. Bailis, M. Balazinska, P. Bernstein, P. Boncz, S. Chaudhuri, A. Cheung, A. Doan, L. Dong, M. J. Franklin, J. Freire, A. Halevy, J. M. Hellerstein, S. Idreos, D. Kossmann, T. Kraska, S. Krishnamurthy, V. Markl, S. Melnik, T. Milo, C. Mohan, T. Neumann, B. Chin Ooi, F. Ozcan, J. Patel, A. Pavlo, R. Popa, R. Ramakrishnan, C. Ré, M. Stonebraker, and D. Suciu, "The Seattle Report on Database Research," *SIGMOD Rec.*, vol. 48, no. 4, p. 44–53, feb 2020. [Online]. Available: https://doi.org/10.1145/3385658.3385668

[2] M. AlMarzouq, A. AlZaidan, and J. AlDallal, "Mining GitHub for research and education: challenges and opportunities," *International Journal of Web Information Systems*, vol. 16, no. 4, pp. 451–473, Jan 2020. [Online]. Available: https://doi.org/10.1108/IJWIS-03-2020-0016

[3] J. Angele, M. Kifer, and G. Lausen, "Ontologies in F-Logic," in *Handbook on Ontologies*, ser. International Handbooks on Information Systems, S. Staab and R. Studer, Eds. Springer, June 2009, pp. 45–70. [Online]. Available: https://ideas.repec.org/h/spr/ihichp/978-3-540-92673-3_2.html

[4] G. Antoniou and F. van Harmelen, *Web Ontology Language: OWL*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 67–92. [Online]. Available: https://doi.org/10.1007/978-3-540-24750-0_4

[5] F. Baader, I. Horrocks, C. Lutz, and U. Sattler, *Introduction to description logic*. Cambridge University Press, 2017. [Online]. Available: https://doi.org/10.1017/9781139025355

[6] B. Ben Mahria, I. Chaker, and A. Zahi, "A novel approach for learning ontology from relational database: from the construction to the evaluation," *Journal of Big Data*, vol. 8, no. 1, p. 25, Jan 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00412-2

[7] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 34–43, 2001.

[8] C. Bizer and R. Cyganiak, "D2r server-publishing relational databases on the semantic web," in *Poster at the 5th international semantic web conference*, vol. 175, 2006.

[9] C. Bizer and A. Schultz, "The berlin sparql benchmark," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 2, pp. 1–24, 2009.

[10] P. Buneman, S. Khanna, and T. Wang-Chiew, "Why and where: A characterization of data provenance," in *Database Theory — ICDT 2001*, J. Van den Bussche and V. Vianu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 316–330.

[11] J.-P. Calbimonte, O. Corcho, and A. J. Gray, "Enabling ontology-based access to streaming data sources," in *The Semantic Web–ISWC 2010: 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I 9*. Springer, 2010, pp. 96–111. [Online]. Available: https://doi.org/10.1007/978-3-642-17746-0_7

[12] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, and G. Xiao, "Ontop: Answering SPARQL queries over relational databases," *Semantic Web*, vol. 8, no. 3, pp. 471–487, 2017. [Online]. Available: https://doi.org/10.3233/SW-160217

[13] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriguez-Muro, R. Rosati, M. Ruzzi, and D. F. Savo, "The MASTRO system for ontology-based data access," *Semantic Web*, vol. 2, no. 1, pp. 43–53, 2011. [Online]. Available: https://doi.org/10.3233/SW-2011-0029

[14] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Ontology-based Database Access." in *SEBD*, 2007, pp. 324–331.

[15] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, "Data complexity of query answering in description logics," *Artificial Intelligence*, vol. 195, pp. 335–360, 2013. [Online]. Available: https://doi.org/10.1016/j.artint.2012.10.003

[16] D. Calvanese, A. Gal, N. Haba, D. Lanti, M. Montali, A. Mosca, and R. Shraga, "ADaMaP: Automatic Alignment of Relational Data Sources Using Mapping Patterns," in *International Conference on Advanced Information Systems Engineering*, Springer. Springer International Publishing, 2021, pp. 193–209. [Online]. Available: https://doi.org/10.1007/978-3-030-79382-1_12

[17] D. Calvanese, T. E. Kalayci, M. Montali, and S. Tinella, "Ontology-Based Data Access for Extracting Event Logs from Legacy Data: The onprom Tool and Methodology," in *Business Information Systems: 20th International Conference, BIS 2017, Poznan, Poland, June 28–30, 2017, Proceedings 20*, Springer. Springer International Publishing, 2017, pp. 220–236. [Online]. Available: https://doi.org/10.1007/978-3-319-59336-4_16

[18] D. Calvanese, D. Lanti, T. M. De Farias, A. Mosca, and G. Xiao, "Accessing scientific data through knowledge graphs with Ontop," *Patterns*, vol. 2, no. 10, p. 100346, 2021.

[19] S. Chacon and B. Straub, *Pro git*. Springer Nature, 2014. [Online]. Available: https://doi.org/10.1007/978-1-4302-1834-0

[20] Y.-H. Chen, E. J.-L. Lu, and T.-A. Ou, "Intelligent SPARQL query generation for natural language processing systems," *IEEE Access*, vol. 9, pp. 158 638–158 650, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3130667

[21] S. Cluet, C. Delobel, J. Siméon, and K. Smaga, "Your mediators need data conversion!" in *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, 1998, pp. 177–188. [Online]. Available: https://doi.org/10.1145/276305.276321

[22] E. F. Codd, "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, no. 6, pp. 377–387, 1970. [Online]. Available: https://doi.org/10.1145/362384.362685

[23] J. Coelho, M. T. Valente, L. L. Silva, and E. Shihab, "Identifying unmaintained projects in github," in *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2018, pp. 1–10. [Online]. Available: https://doi.org/10.1145/3239235.3240501

[24] V. Cosentino, J. L. C. Izquierdo, and J. Cabot, "A Systematic Mapping Study of Software Development With GitHub," *IEEE Access*, vol. 5, pp. 7173–7192, 2017. [Online]. Available: https://doi.org/10.1109/ACCESS.2017.2682323

[25] O. Dabic, E. Aghajani, and G. Bavota, "Sampling Projects in GitHub for MSR Studies," in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, 2021, pp. 560–564. [Online]. Available: https://doi.org/10.1109/MSR52588.2021.00074

[26] G. De Giacomo and M. Lenzerini, "TBox and ABox reasoning in expressive description logics." *KR*, vol. 96, no. 316-327, p. 10, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:182613

[27] E. Della Valle and S. Ceri, "Querying the semantic web: SPARQL," in *Handbook of Semantic Web Technologies*, 2011. [Online]. Available: https://doi.org/10.1007/978-3-540-92913-0_8

[28] L. Di-Jorio, S. Bringay, C. Fiot, A. Laurent, and M. Teisseire, "Sequential patterns for maintaining ontologies over time," in *On the Move to Meaningful Internet Systems: OTM 2008*, R. Meersman and Z. Tari, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1385–1403.

[29] C. Franke, S. Morin, A. Chebotko, J. Abraham, and P. Brazier, "Distributed semantic web data management in HBase and MySQL cluster," in *2011 IEEE 4th International Conference on Cloud Computing*. IEEE, 2011, pp. 105–112. [Online]. Available: https://doi.org/10.1109/CLOUD.2011.19

[30] M. R. Genesereth and N. J. Nilsson, "CHAPTER 2 - Declarative Knowledge," in *Logical Foundations of Artificial Intelligence*, M. R. Genesereth and N. J. Nilsson, Eds. San Francisco (CA): Morgan Kaufmann, 1987, pp. 9–44. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780934613316500082

[31] G. D. Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Using Ontologies for Semantic Data Integration ," in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*. Springer, 2018, pp. 187–202. [Online]. Available: https://doi.org/10.1007/978-3-319-61893-7_11

[32] B. Glavic, "Big data provenance: Challenges and implications for benchmarking," in *Specifying Big Data Benchmarks*, T. Rabl, M. Poess, C. Baru, and H.-A. Jacobsen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 72–80.

[33] M. Golzadeh, A. Decan, D. Legay, and T. Mens, "A ground-truth dataset and classification model for detecting bots in GitHub issue and PR comments," *Journal of Systems and Software*, vol. 175, p. 110911, 2021. [Online]. Available: https://doi.org/10.1016/j.jss.2021.110911

[34] H. Gorskis, L. Aleksejeva, and I. Polaka, "Ontology-Based System Development for Medical Database Access," in *ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference*, vol. 2, 2017, pp. 24–29. [Online]. Available: https://doi.org/10.17770/etr2017vol2.2572

[35] G. Gousios, "The GHTorent dataset and tool suite," in *2013 10th Working Conference on Mining Software Repositories (MSR)*. IEEE, 2013, pp. 233–236. [Online]. Available: https://doi.org/10.1109/MSR.2013.6624034

[36] G. Gousios and D. Spinellis, "GHTorrent: GitHub's data from a firehose," in *2012 9th IEEE Working Conference on Mining Software Repositories (MSR)*. IEEE, 2012, pp. 12–21. [Online]. Available: https://doi.org/10.1109/MSR.2012.6224294

[37] G. Gousios, B. Vasilescu, A. Serebrenik, and A. Zaidman, "Lean GHTorrent: GitHub data on demand," in *Proceedings of the 11th working conference on mining software repositories*, 2014, pp. 384–387. [Online]. Available: https://doi.org/10.1145/2597073.2597126

[38] T. J. Green, *Bag Semantics*. Boston, MA: Springer US, 2009, pp. 201–206. [Online]. Available: https://doi.org/10.1007/978-0-387-39940-9_979

[39] T. Gruber, "Ontology," *Encyclopedia of Database Systems*, 2008. [Online]. Available: https://cir.nii.ac.jp/crid/1570009751340313600

[40] N. Guarino, *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. IOS press, 1998, vol. 46. [Online]. Available: https://dl.acm.org/doi/10.5555/521669

[41] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" in *Handbook on ontologies*, S. Staab and R. Studer, Eds. Springer Berlin Heidelberg, 2009, pp. 1–17. [Online]. Available: https://doi.org/10.1007/978-3-540-92673-3_0

[42] A. Gusenkov, N. Bukharaev, and E. Birialtsev, "On ontology based data integration: problems and solutions," vol. 1203, no. 1. IOP Publishing, apr 2019, p. 012059. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1203/1/012059

[43] S. Hasan, E. A. Fox, K. Bisset, and M. V. Marathe, "EpiK: A Knowledge Base for Epidemiological Modeling and Analytics of Infectious Diseases," *Journal of Healthcare Informatics Research*, vol. 1, no. 2, pp. 260–303, Dec 2017. [Online]. Available: https://doi.org/10.1007/s41666-017-0010-9

[44] J. Heflin, "OWL Web Ontology Language Use Cases and Requirements," *W3C Recommendation*, vol. 10, no. 10, pp. 1–12, 2004. [Online]. Available: https://www.w3.org/TR/webont-req/

[45] S. Heymans, L. Ma, D. Anicic, Z. Ma, N. Steinmetz, Y. Pan, J. Mei, A. Fokoue, A. Kalyanpur, A. Kershenbaum *et al.*, "Ontology Reasoning with Large Data Repositories," in *Ontology Management*. Springer US, 2008, pp. 89–128. [Online]. Available: https://doi.org/10.1007/978-0-387-69900-4_4

[46] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, S. Rudolph *et al.*, "OWL 2 Web Ontology Language Primer (Second Edition)," *W3C recommendation*, 2012. [Online]. Available: https://www.w3.org/TR/owl2-primer/

[47] M. Horridge, S. Jupp, G. Moulton, A. Rector, R. Stevens, and C. Wroe, "A practical guide to building "owl" ontologies using protégé 4 and co-ode tools edition1. 2," *The university of Manchester*, vol. 107, 2009.

[48] I. Horrocks, P. F. Patel-Schneider, and F. Van Harmelen, "From SHIQ and RDF to OWL: The making of a web ontology language," *Journal of Web Semantics*, vol. 1, no. 1, pp. 7–26, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826803000027

[49] N. Hu, Y. Wu, G. Qi, D. Min, J. Chen, J. Z. Pan, and Z. Ali, "An empirical study of pre-trained language models in simple knowledge graph question answering," *World Wide Web*, pp. 1–32, May 2023. [Online]. Available: https://doi.org/10.1007/s11280-023-01166-y

[50] J. Huang, D. J. Abadi, and K. Ren, "Scalable SPARQL querying of large RDF graphs," *Proceedings of the VLDB Endowment*, vol. 4, no. 11, pp. 1123–1134, aug 2011. [Online]. Available: https://doi.org/10.14778/3402707.3402747

[51] N. Imtiaz, J. Middleton, J. Chakraborty, N. Robson, G. Bai, and E. Murphy-Hill, "Investigating the effects of gender bias on GitHub," in *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*, ser. ICSE '19. IEEE, 2019, pp. 700–711. [Online]. Available: https://doi.org/10.1109/ICSE.2019.00079

[52] Y. Jafta, L. Leenen, and T. Meyer, "Investigating Ontology-Based Data Access with GitHub," in *The Semantic Web*, C. Pesquita, E. Jimenez-Ruiz, J. McCusker, D. Faria, M. Dragoni, A. Dimou, R. Troncy, and S. Hertling, Eds. Cham: Springer Nature Switzerland, 2023, pp. 644–660. [Online]. Available: https://doi.org/10.1007/978-3-031-33455-9_38

[53] E. G. Kalaycı, I. Grangel González, F. Lösch, G. Xiao, A. ul Mehdi, E. Kharlamov, and D. Calvanese, "Semantic Integration of Bosch Manufacturing Data Using Virtual Knowledge Graphs," in *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19*. Springer, 2020, pp. 464–481. [Online]. Available: https://doi.org/10.1007/978-3-030-62466-8_29

[54] E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian, "The promises and perils of mining GitHub," in *Proceedings of the 11th working conference on mining software repositories*, ser. MSR 2014, 2014, pp. 92–101. [Online]. Available: https://doi.org/10.1145/2597073.2597074

[55] R. Kallis, A. Di Sorbo, G. Canfora, and S. Panichella, "Predicting issue types on GitHub," *Science of Computer Programming*, vol. 205, p. 102598,

2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167642320302069

[56] Y.-B. Kang, S. Krishnaswamy, W. Sawangphol, L. Gao, and Y.-F. Li, "Understanding and improving ontology reasoning efficiency through learning and ranking," *Information Systems*, vol. 87, p. 101412, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437917306476

[57] C. M. Keet, *An introduction to ontology engineering*. University of Cape Town, 2018.

[58] E. Kharlamov, D. Hovland, E. Jiménez-Ruiz, D. Lanti, H. Lie, C. Pinkel, M. Rezk, M. G. Skjæveland, E. Thorstensen, G. Xiao *et al.*, "Ontology Based Access to Exploration Data at Statoil," in *The Semantic Web-ISWC 2015: 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II 14*. Springer International Publishing, 2015, pp. 93–112. [Online]. Available: https://doi.org/10.1007/978-3-319-25010-6_6

[59] E. Kharlamov, D. Hovland, M. G. Skjæveland, D. Bilidas, E. Jiménez-Ruiz, G. Xiao, A. Soylu, D. Lanti, M. Rezk, D. Zheleznyakov *et al.*, "Ontology Based Data Access in Statoil," *Journal of Web Semantics*, vol. 44, pp. 3–36, 2017, Industry and In-use Applications of Semantic Technologies. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826817300276

[60] E. Kharlamov, Y. Kotidis, T. Mailis, C. Neuenstadt, C. Nikolaou, Ö. Özçep, C. Svingos, D. Zheleznyakov, Y. Ioannidis, S. Lamparter *et al.*, "An ontology-mediated analytics-aware approach to support monitoring and diagnostics of static and streaming data," *Journal of Web Semantics*, vol. 56, pp. 30–55, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826819300010

[61] E. Kharlamov, T. Mailis, G. Mehdi, C. Neuenstadt, Ö. Özçep, M. Roshchin, N. Solomakhina, A. Soylu, C. Svingos, S. Brandt *et al.*, "Semantic access to streaming and static data at Siemens," *Journal of Web Semantics*, vol. 44, pp. 54–74, 2017, industry and In-use Applications of Semantic Technologies. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826817300124

[62] E. Kharlamov, O. Savković, M. Ringsquandl, G. Xiao, G. Mehdi, E. G. Kalayc, W. Nutt, M. Roshchin, I. Horrocks, and T. Runkler, "Diagnostics of Trains with Semantic Diagnostics Rules," in *Inductive Logic Programming:*

28th International Conference, ILP 2018, Ferrara, Italy, September 2–4, 2018, Proceedings 28. Springer, 2018, pp. 54–71. [Online]. Available: https://doi.org/10.1007/978-3-319-99960-9_4

[63] T. Kinsman, M. Wessel, M. A. Gerosa, and C. Treude, "How do software developers use GitHub Actions to automate their workflows?" in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, 2021, pp. 420–431. [Online]. Available: https://doi.org/10.1109/MSR52588.2021.00054

[64] K. I. Kotis, G. A. Vouros, and D. Spiliotopoulos, "Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations," *The Knowledge Engineering Review*, vol. 35, p. e4, 2020.

[65] M. Krötzsch, F. Simancik, and I. Horrocks, "A Description Logic Primer," *arXiv preprint arXiv:1201.4089*, vol. abs/1201.4089, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:1221862

[66] D. O. Kubitza, M. Böckmann, and D. Graux, "SemanGit: A Linked Dataset from git," in *The Semantic Web – ISWC 2019*. Springer, 2019, pp. 215–228. [Online]. Available: https://doi.org/10.1007/978-3-030-30796-7_14

[67] B. Lakzaei and M. Shamsfard, "Ontology learning from relational databases," *Information Sciences*, vol. 577, pp. 280–297, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025521006654

[68] M. Lenzerini and C. Daraio, *Challenges, Approaches and Solutions in Data Integration for Research and Innovation*. Cham: Springer International Publishing, 2019, pp. 397–420. [Online]. Available: https://doi.org/10.1007/978-3-030-02511-3_15

[69] S. Liang, K. Stockinger, T. M. de Farias, M. Anisimova, and M. Gil, "Querying knowledge graphs in natural language," *Journal of big data*, vol. 8, pp. 1–23, 2021. [Online]. Available: https://doi.org/10.1186/s40537-020-00383-w

[70] C.-h. Liao, Y.-f. Wu, and G.-h. King, "Research on Learning OWL Ontology from Relational Database," in *Journal of Physics: Conference Series*, vol. 1176, no. 2. IOP Publishing, 2019, p. 022031. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1176/2/022031

[71] Z. Liao, D. He, Z. Chen, X. Fan, Y. Zhang, and S. Liu, "Exploring the characteristics of issue-related behaviors in github using visualization techniques," *IEEE Access*, vol. 6, pp. 24 003–24 015, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2810295

[72] J. Liu, J. Li, and L. He, "A Comparative Study of the Effects of Pull Request on GitHub Projects," in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. IEEE, 2016, pp. 313–322. [Online]. Available: https://doi.org/10.1109/COMPSAC.2016.27

[73] J. Loeliger and M. McCullough, *Version Control with Git: Powerful tools and techniques for collaborative software development.* " O'Reilly Media, Inc.", 2012.

[74] S. Lohmann, V. Link, E. Marbach, and S. Negru, "WebVOWL: Web-based visualization of ontologies," in *Knowledge Engineering and Knowledge Management: EKAW 2014 Satellite Events, VISUAL, EKM1, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers. 19.* Springer, 2015, pp. 154–158.

[75] M. Luczak-Rösch, "Towards agile ontology maintenance," in *The Semantic Web - ISWC 2009*, A. Bernstein, D. R. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, and K. Thirunarayan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 965–972.

[76] C. Ma and B. Molnr, "Ontology learning from relational database: Opportunities for semantic information integration," *Vietnam Journal of Computer Science*, vol. 9, no. 01, pp. 31–57, 2022. [Online]. Available: https://doi.org/10.1142/S219688882150024X

[77] C. Ma and B. Molnár, "Use of ontology learning in information system integration: a literature survey," in *Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings 12.* Springer, 2020, pp. 342–353. [Online]. Available: https://api.semanticscholar.org/CorpusID:212564890

[78] M. Madsen and O. Lhoták, "Fixpoints for the masses: programming with first-class Datalog constraints," *Proceedings of the ACM on Programming Languages*, vol. 4, no. OOPSLA, pp. 1–28, 2020. [Online]. Available: https://doi.org/10.1145/3428193

[79] M. N. Mami, D. Graux, S. Scerri, H. Jabeen, S. Auer, and J. Lehmann, "Squerall: Virtual Ontology-Based Access to Heterogeneous and Large Data Sources," in *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18.* Springer, 2019, pp. 229–245. [Online]. Available: https://api.semanticscholar.org/CorpusID:204754566

[80] F. Manola, E. Miller, B. McBride *et al.*, "RDF primer," *W3C recommendation*, vol. 10, no. 1-107, p. 6, 2004. [Online]. Available: https://www.w3.org/TR/rdf-primer/

[81] M. A. Martínez-Prieto, M. Arias Gallego, and J. D. Fernández, "Exchange and Consumption of Huge RDF Data," in *Extended Semantic Web Conference*. Springer Berlin Heidelberg, 2012, pp. 437–452. [Online]. Available: https://doi.org/10.1007/978-3-642-30284-8_36

[82] H. E. Massari, S. Mhammedi, N. Gherabi, and M. Nasri, "Virtual OBDA Mechanism Ontop for Answering SPARQL Queries Over Couchbase," in *International Conference on Advanced Technologies for Humanity*. Springer, 2021, pp. 193–205. [Online]. Available: https://doi.org/10.1007/978-3-030-94188-8_19

[83] N. Matentzoglu, D. Goutte-Gattat, S. Z. K. Tan, J. P. Balhoff, S. Carbon, A. R. Caron, W. D. Duncan, J. E. Flack, M. Haendel, N. L. Harris, W. R. Hogan, C. T. Hoyt, R. C. Jackson, H. Kim, H. Kir, M. Larralde, J. A. McMurry, J. A. Overton, B. Peters, C. Pilgrim, R. Stefancsik, S. M. Robb, S. Toro, N. A. Vasilevsky, R. Walls, C. J. Mungall, and D. Osumi-Sutherland, "Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies," *Database*, vol. 2022, p. baac087, 10 2022. [Online]. Available: https://doi.org/10.1093/database/baac087

[84] B. McBride, "The resource description framework (RDF) and its vocabulary description language RDFS," in *Handbook on ontologies*. Springer Berlin Heidelberg, 2004, pp. 51–65. [Online]. Available: https://doi.org/10.1007/978-3-540-24750-0_3

[85] K. McGlinn, M. A. Rutherford, K. Gisslander, L. Hederman, M. A. Little, and D. O'Sullivan, "FAIRVASC: A semantic web approach to rare disease registry integration," *Computers in Biology and Medicine*, vol. 145, p. 105313, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522001056

[86] C. Messaoudi, R. Fissoune, and H. Badir, "IPDS: A semantic mediator-based system using Spark for the integration of heterogeneous proteomics data sources," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 1, p. e5814, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:219513264

[87] R. D. Morris, "Web 3.0: Implications for online learning," 2011. [Online]. Available: https://doi.org/10.1007/s11528-011-0469-9

[88] C. Nikolaou, E. V. Kostylev, G. Konstantinidis, M. Kaminski, B. C. Grau, and I. Horrocks, "The Bag Semantics of Ontology-Based Data Access," *arXiv preprint arXiv:1705.07105*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1705.07105

[89] N. F. Noy, D. L. McGuinness *et al.*, "Ontology development 101: A guide to creating your first ontology," 2001. [Online]. Available: https://api.semanticscholar.org/CorpusID:500106

[90] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007. [Online]. Available: https://doi.org/10.2753/MIS0742-1222240302

[91] N. Petersen, L. Halilaj, I. Grangel-González, S. Lohmann, C. Lange, and S. Auer, "Realizing an RDF-Based Information Model for a Manufacturing Company – A Case Study," in *The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II 16*. Springer, 2017, pp. 350–366. [Online]. Available: https://doi.org/10.1007/978-3-319-68204-4_31

[92] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati, "Linking Data to Ontologies," in *Journal on data semantics X*. Springer, Berlin, Heidelberg, 2008, pp. 133–173. [Online]. Available: https://doi.org/10.1007/978-3-540-77688-8_5

[93] F. Priyatna, R. Alonso-Calvo, S. Paraiso-Medina, and O. Corcho, "Querying clinical data in HL7 RIM based relational model with morph-RDB," *Journal of biomedical semantics*, vol. 8, no. 1, pp. 1–12, 2017. [Online]. Available: https://doi.org/10.1186/s13326-017-0155-8

[94] F. Priyatna, O. Corcho, and J. Sequeda, "Formalisation and experiences of R2RML-based SPARQL to SQL query translation using Morph," in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 479–490. [Online]. Available: https://doi.org/10.1145/2566486.2567981

[95] A. Rastogi, N. Nagappan, G. Gousios, and A. van der Hoek, "Relationship between geographical location and evaluation of developer contributions in github," in *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2018, pp. 1–8. [Online]. Available: https://doi.org/10.1145/3239235.3240504

[96] R. Rudman and R. Bruwer, "Defining web 3.0: opportunities and challenges," *The Electronic Library*, vol. 34, 2016. [Online]. Available: https://doi.org/10.1108/EL-08-2014-0140

[97] G. Santipantakis, K. Kotis, and G. A. Vouros, "OBDAIR: Ontology-Based Distributed framework for Accessing, Integrating and Reasoning with data in disparate data sources," *Expert Systems with Applications*, vol. 90, pp. 464–483, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417305705

[98] O. Savković, E. Kharlamov, M. Ringsquandl, G. Xiao, G. Mehdi, E. G. Kalayc, W. Nutt, and I. Horrocks, "Semantic diagnostics of smart factories," in *Semantic Technology: 8th Joint International Conference, JIST 2018, Awaji, Japan, November 26–28, 2018, Proceedings 8*. Springer, 2018, pp. 277–294. [Online]. Available: https://doi.org/10.1007/978-3-030-04284-4_19

[99] T. Schneider and M. Šimkus, "Ontologies and Data Management: A Brief Survey," *KI-Künstliche Intelligenz*, vol. 34, no. 3, pp. 329–353, 2020. [Online]. Available: https://doi.org/10.1007/s13218-020-00686-3

[100] G. Sejdiu, D. Graux, I. Khan, I. Lytra, H. Jabeen, and J. Lehmann, "Towards a Scalable Semantic-Based Distributed Approach for SPARQL Query Evaluation," in *International Conference on Semantic Systems*. Springer, Cham, 2019, pp. 295–309. [Online]. Available: https://doi.org/10.1007/978-3-030-33220-4_22

[101] J. F. Sequeda and D. P. Miranker, "Ultrawrap: SPARQL execution on relational data," *Journal of Web Semantics*, vol. 22, pp. 19–39, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570826813000383

[102] A. P. Sheth and J. A. Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases," *ACM Computing Surveys (CSUR)*, vol. 22, no. 3, pp. 183–236, 1990. [Online]. Available: https://doi.org/10.1145/96602.96604

[103] A. P. Sheth and C. Ramakrishnan, "Semantic (Web) Technology in Action: Ontology Driven Information Systems for Search, Integration, and Analysis," *IEEE Data Engineering Bulletin*, vol. 26, no. 4, p. 40, 2003.

[104] A. C. Sima, T. Mendes de Farias, E. Zbinden, M. Anisimova, M. Gil, H. Stockinger, K. Stockinger, M. Robinson-Rechavi, and C. Dessimoz, "Enabling semantic queries across federated bioinformatics databases," *Database*, vol. 2019, 11 2019. [Online]. Available: https://doi.org/10.1093/database/baz106

[105] G. Singh, S. Bhatia, and R. Mutharaju, "OWL2Bench: A Benchmark for OWL 2 Reasoners," in *International semantic web conference*. Springer International Publishing, 2020, pp. 81–96. [Online]. Available: https://doi.org/10.1007/978-3-030-62466-8_6

[106] D. Spinellis, "Git," *IEEE software*, vol. 29, no. 3, pp. 100–101, 2012. [Online]. Available: https://doi.org/10.1109/MS.2012.61

[107] S. Staab and R. Studer, *Handbook on ontologies*. Springer Science & Business Media, 2010. [Online]. Available: https://doi.org/10.1007/978-3-540-92673-3

[108] C. Stadler, G. Sejdiu, D. Graux, and J. Lehmann, "Sparklify: A Scalable Software Component for Efficient Evaluation of SPARQL Queries over Distributed RDF Datasets," in *International Semantic Web Conference*. Springer, 2019, pp. 293–308. [Online]. Available: https://doi.org/10.1007/978-3-030-30796-7_19

[109] F. Taglino, F. Cumbo, G. Antognoli, I. Arisi, M. D'Onofrio, F. Perazzoni, R. Voyat, G. Fiscon, F. Conte, M. Canevelli *et al.*, "An ontology-based approach for modelling and querying Alzheimer's disease data," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, pp. 1–15, 2023. [Online]. Available: https://doi.org/10.1186/s12911-023-02211-6

[110] R. Thirumahal, G. Sudha Sadasivam, and P. Shruti, "Semantic Integration of Heterogeneous Data Sources Using Ontology-Based Domain Knowledge Modeling for Early Detection of COVID-19," *SN Computer Science*, vol. 3, no. 6, p. 428, 2022. [Online]. Available: https://doi.org/10.1007/s42979-022-01298-4

[111] W. Van Der Aalst, *Process mining: Data Science in Action*. Springer Berlin, Heidelberg, 2016, vol. 2. [Online]. Available: https://doi.org/10.1007/978-3-662-49851-4

[112] G. Vega-Gorgojo, L. Slaughter, and M. Giese, "Seeing the whole picture: integrated pre-surgery reports with PreOptique," *Journal of Biomedical Semantics*, vol. 10, pp. 1–15, 2019. [Online]. Available: https://doi.org/10.1186/s13326-019-0197-1

[113] P. S. Vikas Trikha and S. Kothari, "Managing Data Provenance in the Semantic Web," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT)*, vol. 8, no. 3, 2019.

[114] J. Vom Brocke, A. Hevner, and A. Maedche, "Introduction to Design Science Research," *Design science research. Cases*, pp. 1–13, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-46781-4_1

[115] J. Wachs, M. Nitecki, W. Schueller, and A. Polleres, "The Geography of Open Source Software: Evidence from GitHub," *Technological Forecasting and Social Change*, vol. 176, p. 121478, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162522000105

[116] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, and M. Zakharyaschev, "Ontology-based data access: A survey." International Joint Conferences on Artificial Intelligence, 7 2018, pp. 5511–5519. [Online]. Available: https://doi.org/10.24963/ijcai.2018/777

[117] G. Xiao, L. Ding, B. Cogrel, and D. Calvanese, "Virtual knowledge graphs: An overview of systems and use cases," *Data Intelligence*, vol. 1, no. 3, pp. 201–223, 2019. [Online]. Available: https://doi.org/10.1162/dint_a_00011

[118] G. Xiao, R. Kontchakov, B. Cogrel, D. Calvanese, and E. Botoeva, "Efficient handling of SPARQL optional for OBDA," in *International Semantic Web Conference*. Springer, Cham, 2018, pp. 354–373. [Online]. Available: https://doi.org/10.1007/978-3-030-00671-6_21

[119] G. Xiao, D. Lanti, R. Kontchakov, S. Komla-Ebri, E. Güzel-Kalaycı, L. Ding, J. Corman, B. Cogrel, D. Calvanese, and E. Botoeva, "The Virtual Knowledge Graph System Ontop," in *International Semantic Web Conference*. Springer, Cham, 2020, pp. 259–277. [Online]. Available: https://doi.org/10.1007/978-3-030-62466-8_17

[120] H. Zhang, Y. Guo, Q. Li, T. J. George, E. Shenkman, F. Modave, and J. Bian, "An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival," *BMC medical informatics and decision making*, vol. 18, no. 2, pp. 129–147, 2018. [Online]. Available: https://doi.org/10.1186/s12911-018-0636-4

[121] H. Zhang, Y. Guo, Q. Li, T. J. George, E. A. Shenkman, and J. Bian, "Data Integration through Ontology-Based Data Access to Support Integrative Data Analysis: A Case Study of Cancer Survival," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 1300–1303. [Online]. Available: https://doi.org/10.1109/BIBM.2017.8217849

[122] X. Zhang, M. Zhang, P. Peng, J. Song, Z. Feng, and L. Zou, "A Scalable Sparse Matrix-Based Join for SPARQL Query Processing," in *Database Systems for Advanced Applications*, G. Li, J. Yang, J. Gama, J. Natwichai, and Y. Tong, Eds. Cham: Springer International Publishing, 2019, pp. 510–514. [Online]. Available: https://doi.org/10.1007/978-3-030-18590-9_77