



Universiteit
Leiden
The Netherlands

Healthcare information system engineering: AI technologies and open source approaches

Shen, Z.

Citation

Shen, Z. (2025, December 3). *Healthcare information system engineering: AI technologies and open source approaches*. Retrieved from <https://hdl.handle.net/1887/4284431>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4284431>

Note: To cite this publication please use the final published version (if applicable).

Bibliography

- A Seoane, J., Aguiar-Pulido, V., R Munteanu, C., Rivero, D., R Rabunal, J., Dorado, J., and Pazos, A. (2013). Biomedical data integration in computational drug design and bioinformatics. *Current computer-aided drug design*, 9(1):108–117.
- Abdallah, Z. S., Carman, M., and Haffari, G. (2017). Multi-domain evaluation framework for named entity recognition tools. *Computer Speech & Language*, 43:34–55.
- Akowuah, F., Lake, J., Yuan, X., Nuakoh, E., and Yu, H. (2015). Testing the security vulnerabilities of openemr 4.1. 1: a case study. *Journal of Computing Sciences in Colleges*, 30(3):26–35.
- Albahri, A. S., Hamid, R. A., Alwan, J. K., Al-Qays, Z., Zaidan, A., Zaidan, B., Albahri, A., AlAmoodi, A. H., Khlaf, J. M., Almahdi, E., et al. (2020). Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (covid-19): a systematic review. *Journal of medical systems*, 44:1–11.
- Alharthi, A., Krotov, V., and Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*.
- Ansper, A., Buldas, A., Freudenthal, M., and Willemson, J. (2013). Protecting a federated database infrastructure against denial-of-service attacks. In *Critical Information Infrastructures Security: 8th International Workshop, CRITIS 2013, Amsterdam, The Netherlands, September 16-18, 2013, Revised Selected Papers 8*, pages 26–37. Springer.
- Anthes, G. (2016). Open source software no longer optional. *Communications of the ACM*, 59(8):15–17.
- APIs, N. (2017). Api-based cms buyer’s guide. Accessed: 2024-08-22.
- Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., and Buyya, R. (2015). Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79-80:3 – 15. Special Issue on Scalable Systems for Big Data Management and Analytics.
- Athey, B. D., Braxenthaler, M., Haas, M., and Guo, Y. (2013). transmart: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Summits on Translational Science Proceedings*, 2013:6.

Bibliography

- Badgeley, M. A., Shameer, K., Glicksberg, B. S., Tomlinson, M. S., Levin, M. A., McCormick, P. J., Kasarskis, A., Reich, D. L., and Dudley, J. T. (2016). Ehdviz: clinical dashboard development using open-source technologies. *BMJ open*, 6(3):e010579.
- Bahrami, M. and Singhal, M. (2015). *The Role of Cloud Computing Architecture in Big Data*, pages 275–295. Springer International Publishing, Cham.
- Banda, J. M., Evans, L., Vanguri, R. S., Tatonetti, N. P., Ryan, P. B., and Shah, N. H. (2016). A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific data*, 3(1):1–11.
- Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., et al. (2017). Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7.
- Batini, C., Lenzerini, M., and Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. *ACM computing surveys (CSUR)*, 18(4):323–364.
- Beynon-Davies, P., Tudhope, D., and Mackay, H. (1999). Information systems prototyping in practice. *Journal of information technology*, 14(1):107–120.
- Björkman, I. K., Fastbom, J., Schmidt, I. K., Bernsten, C. B., and of the Elderly in Europe Research (PEER) Group, P. C. (2002). Drug—drug interactions in the elderly. *Annals of Pharmacotherapy*, 36(11):1675–1681.
- Blaya, J. A., Fraser, H. S., and Holt, B. (2010). E-health technologies show promise in developing countries. *Health Affairs*, 29(2):244–251.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Blum, M. R., Sallevelt, B. T. G. M., Spinewine, A., O’Mahony, D., Moutzouri, E., Feller, M., Baumgartner, C., Roumet, M., Jungo, K. T., Schwab, N., Bretagne, L., Beglinger, S., Aubert, C. E., Wilting, I., Thevelin, S., Murphy, K., Huibers, C. J. A., Drenth-van Maanen, A. C., Boland, B., Crowley, E., Eichenberger, A., Meulendijk, M., Jennings, E., Adam, L., Roos, M. J., Gleeson, L., Shen, Z., Marien, S., Meinders, A.-J., Baretella, O., Netzer, S., de Montmollin, M., Fournier, A., Mouzon, A., O’Mahony, C., Aujesky, D., Mavridis, D., Byrne, S., Jansen, P. A. F., Schwenkglenks, M., Spruit, M., Dalleur, O., Knol, W., Trelle, S., and Rodondi, N. (2021a). Optimizing therapy to prevent avoidable hospital admissions in multimorbid older adults (operam): cluster randomised controlled trial. *BMJ*, 374. <https://www.bmj.com/content/374/bmj.n1585>.
- Blum, M. R., Sallevelt, B. T. G. M., Spinewine, A., O’Mahony, D., Moutzouri, E., Feller, M., Baumgartner, C., Roumet, M., Jungo, K. T., Schwab, N., Bretagne, L., Beglinger,

- S., Aubert, C. E., Wilting, I., Thevelin, S., Murphy, K., Huibers, C. J. A., Drenth-van Maanen, A. C., Boland, B., Crowley, E., Eichenberger, A., Meulendijk, M., Jennings, E., Adam, L., Roos, M. J., Gleeson, L., **Shen, Z.**, Marien, S., Meinders, A.-J., Baretella, O., Netzer, S., de Montmollin, M., Fournier, A., Mouzon, A., O'Mahony, C., Aujesky, D., Mavridis, D., Byrne, S., Jansen, P. A. F., Schwenkglenks, M., **Spruit, M.**, Dalleur, O., Knol, W., Trelle, S., and Rodondi, N. (2021b). Optimizing therapy to prevent avoidable hospital admissions in multimorbid older adults (operam): cluster randomised controlled trial. *BMJ*, 374.
- Bodenreider, O. (2004). The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270.
- Boettiger, C. (2014). An introduction to docker for reproducible research, with examples from the r environment. *ACM SIGOPS Oper. Syst. Rev.*, 49.
- Boettiger, C. (2015). An introduction to docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1):71–79.
- Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1):75–90.
- Bradshaw, S., Brazil, E., and Chodorow, K. (2019). *MongoDB: the definitive guide: powerful and scalable data storage*. O'Reilly Media.
- Branson, A., Hauer, T., McClatchey, R., Rogulin, D., and Shamdasani, J. (2008). A data model for integrating heterogeneous medical data in the health-e-child project. *arXiv preprint arXiv:0812.2874*.
- Brazhnik, O. and Jones, J. F. (2007). Anatomy of data integration. *Journal of biomedical informatics*, 40(3):252–269.
- Bretthauer, D. (2002). Open source software: A history. *UConn Libraries Published Works*, 21.
- Cai, T., Giannopoulos, A. A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K. K., Rybicki, F. J., and Mitsouras, D. (2016). Natural language processing technologies in radiology research and clinical applications. *Radiographics*, 36(1):176–191.
- Calì, A., Calvanese, D., De Giacomo, G., and Lenzerini, M. (2002). Data integration under integrity constraints. In *Advanced Information Systems Engineering: 14th International Conference, CAiSE 2002 Toronto, Canada, May 27–31, 2002 Proceedings 14*, pages 262–279. Springer.
- Canonica, M. and De Russis, L. (2018). A comparison and critique of natural language understanding tools. In *Cloud Computing 2018*, pages 110–115. CLOUD COMPUTING 2018: The Ninth International Conference on Cloud Computing, GRIDs, and Virtualization ; Conference date: 01-01-2018.

Bibliography

- Carchiolo, V., Longheu, A., Reitano, G., and Zagarella, L. (2019). Medical prescription classification: a nlp-based approach. In *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 605–609. IEEE.
- Cardinal, R. N. (2017). Clinical records anonymisation and text extraction (crate): an open-source software system. *BMC medical informatics and decision making*, 17:1–12.
- Carrell, D. (2011). A strategy for deploying secure cloud-based natural language processing systems for applied research involving clinical text. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–11. IEEE.
- Cavelaars, M., Rousseau, J., Parlayan, C., de Ridder, S., Verburg, A., Ross, R., Visser, G. R., Rotte, A., Azevedo, R., Boiten, J.-W., et al. (2015). Openclinica. In *Journal of clinical bioinformatics*, volume 5-1, pages 1–2. BioMed Central.
- Cañada, A., Capella-Gutierrez, S., Oyarzabal, J., Rabal, O., Valencia, A., and Krallinger, M. (2017). LimTox: a web tool for applied text mining of adverse event and toxicity associations of compounds, drugs and genes. *Nucleic Acids Research*, 45(W1):W484–W489.
- Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). Evaluation of negation phrases in narrative clinical reports. In *Proceedings of the AMIA Symposium*, page 105. American Medical Informatics Association.
- Chapman, W. W., Cooper, G. F., Hanbury, P., Chapman, B. E., Harrison, L. H., and Wagner, M. M. (2003). Creating a text classifier to detect radiology reports describing mediastinal findings associated with inhalational anthrax and other disorders. *Journal of the American Medical Informatics Association*, 10(5):494–503.
- Chard, K., Russell, M., Lussier, Y. A., Mendonça, E. A., and Silverstein, J. C. (2011). A cloud-based approach to medical nlp. In *AMIA Annual Symposium Proceedings*, volume 2011, page 207. American Medical Informatics Association.
- Che, D., Safran, M., and Peng, Z. (2013). From big data to big data mining: challenges, issues, and opportunities. In *Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18*, pages 1–15. Springer.
- Chiang, J.-H., Lin, J.-W., and Yang, C.-W. (2010). Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using medical language extraction and encoding system (medlee). *Journal of the American Medical Informatics Association*, 17(3):245–252.
- Chyu, M.-C., Austin, T., Calisir, F., Chanjaplammoetil, S., Davis, M. J., Favela, J., Gan, H., Gefen, A., Haddas, R., Hahn-Goldberg, S., et al. (2015). Healthcare engineering defined: a white paper. *Journal of healthcare engineering*, 6(4):635–648.
- Claxton, A. J., Cramer, J., and Pierce, C. (2001). A systematic review of the associations between dose regimens and medication compliance. *Clinical therapeutics*, 23(8):1296–1310.

- Clifton, C., Housman, E., and Rosenthal, A. (1998). Experience with a combined approach to attribute-matching across heterogeneous databases. In *Data Mining and Reverse Engineering: Searching for semantics. IFIP TC2 WG2. 6 IFIP Seventh Conference on Database Semantics (DS-7) 7–10 October 1997, Leysin, Switzerland*, pages 428–451. Springer.
- Creswell, J. W., Fetters, M. D., and Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *The Annals of Family Medicine*, 2(1):7–12.
- Crowley, E. K., Sallevelt, B. T., Huibers, C. J., Murphy, K. D., **Spruit, M.**, **Shen, Z.**, Boland, B., Spinewine, A., Dalleur, O., Moutzouri, E., et al. (2020). Intervention protocol: Optimising therapy to prevent avoidable hospital admission in the multi-morbid elderly (operam): a structured medication review with support of a computerised decision support system. *BMC health services research*, 20(1):1–12. <https://doi.org/10.1186/s12913-020-5056-3>.
- Cunningham, H., Tablan, V., Roberts, A., and Bontcheva, K. (2013). Getting more out of biomedical documents with gate’s full lifecycle open source text analytics. *PLoS computational biology*, 9(2):e1002854.
- Currie, W. L., Desai, B., and Khan, N. (2004). Customer evaluation of application services provisioning in five vertical sectors. *Journal of Information Technology*, 19(1):39–58.
- Dale, R. (2015). Nlp meets the cloud. *Natural language engineering*, 21(4):653–659.
- Davis, L. (2012). *Basic methods in molecular biology*. Elsevier.
- de Abajo, B. S. and Ballester, A. L. (2012). Overview of the most important open source software: analysis of the benefits of openmrs, openemr, and vista. In *Telemedicine and e-health services, policies, and applications: Advancements and developments*, pages 315–346. IGI Global.
- De Castilho, R. E. and Gurevych, I. (2014). A broad-coverage collection of portable nlp components for building shareable analysis pipelines. In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT*, pages 1–11.
- de Oliveira, M. S. (2015). *On the use of visualization for supporting software reuse*. PhD thesis, Universidade Federal do Rio de Janeiro Rio de Janeiro, Brazil.
- Dean, J. and Ghemawat, S. (2004). Mapreduce: Simplified data processing on large clusters. In *OSDI’04: Sixth Symposium on Operating System Design and Implementation*, pages 137–150, San Francisco, CA.
- Demner-Fushman, D., Shooshan, S. E., Rodriguez, L., Aronson, A. R., Lang, F., Rogers, W., Roberts, K., and Topping, J. (2018). A dataset of 200 structured product labels annotated for adverse drug reactions. *Scientific data*, 5(1):1–8.

Bibliography

- Dingsøyr, T. and Lassenius, C. (2016). Emerging themes in agile software development: Introduction to the special section on continuous value delivery. *Information and software technology*, 77:56–60.
- Dodge, M., McDerby, M., and Turner, M. (2011). *Geographic visualization: concepts, tools and applications*. John Wiley & Sons.
- Dunkerley, M. (2013). *Information visualization: perception for design*. Waltham, MA: Morgan Kaufmann Publishers.
- Electronic Medicines Compendium (EMC) (2020). Home—electronic medicines compendium (emc). Accessed: 2020-10-14.
- Evans, R. S. (2016). Electronic health records: then, now, and in the future. *Yearbook of medical informatics*, 25(S 01):S48–S61.
- Ferrucci, D. and Lally, A. (2004). Uima: an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, 10(3-4):327–348.
- Fichman, R. G., Kohli, R., and Krishnan, R. (2011). Editorial overview—the role of information systems in healthcare: current research and future trends. *Information systems research*, 22(3):419–428.
- Frakes, W. B. and Fox, C. J. (1996). Quality improvement using a software reuse failure modes model. *IEEE Transactions on Software Engineering*, 22(4):274–279.
- Fung, K. W., Jao, C. S., and Demner-Fushman, D. (2013). Extracting drug indication information from structured product labels using natural language processing. *Journal of the American Medical Informatics Association*, 20(3):482–488.
- GAM in Python (2018). Generalized additive models in python.
- Garcia, R., Treude, C., and La, W. (2023). Towards understanding the open source interest in gender-related github projects. In *2023 IEEE/ACM 16th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 68–79. IEEE.
- Genism (2018). Genism: Topic modeling for humans.
- GeoVis (2018). Geographic visualization of the number of github repositories by country.
- Gerner, M., Sarafraz, F., M Bergman, C., and Nenadic, G. (2012). Biocontext: An integrated text mining system for large-scale extraction and contextualization of biomolecular events. *Bioinformatics (Oxford, England)*, 28:2154–61.
- Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., et al. (2018). Niftynet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, 158:113–122.

- GitHub (2018). Github api v3: Github developer guide.
- Goldberg, Y. and Levy, O. (2014). word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.
- GoogleCloudPlatform (2019). Kubernetes operator for managing the lifecycle of apache spark applications on kubernetes.
- Gregor, S. and Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, pages 337–355.
- Guaitoli, P. R., Jansma, E. P., de Vet, H. C., et al. (2014). Nutrition screening tools: does one size fit all? a systematic review of screening tools for the hospital setting. *Clinical nutrition*, 33(1):39–58.
- Haak, D., Page, C.-E., and Deserno, T. M. (2016). Electronic data capture and dicom data management in multi-center clinical trials. In *Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations*, volume 9789, pages 118–123. SPIE.
- Haffari, G., Tran, T., and Carman, M. (2017). Efficient benchmarking of nlp apis using multi-armed bandits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 408–416.
- Hansen, M. S. and Sørensen, T. S. (2013). Gadgetron: an open source framework for medical image reconstruction. *Magnetic resonance in medicine*, 69(6):1768–1776.
- Hellmann, S., Lehmann, J., Auer, S., and Brümmer, M. (2013). Integrating nlp using linked data. In *The Semantic Web–ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21–25, 2013, Proceedings, Part II* 12, pages 98–113. Springer.
- Hendrickson, S., Sturdevant, S., Harter, T., Venkataramani, V., Arpaci-Dusseau, A. C., and Arpaci-Dusseau, R. H. (2016). Serverless computation with openlambda. In *8th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 16)*, Denver, CO. USENIX Association.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly*, 28(1):6.
- IBM (2021a). Watson natural language understanding. Accessed: 2021-02-27.
- IBM (2021b). Watson natural language understanding api demo. Accessed: 2021-02-27.
- IBM Watson NLU (2018). Ibm watson natural language understanding.
- Ide, N., Suderman, K., and Kim, J.-D. (2018). Mining biomedical publications with the lapps grid. In *LREC*.

Bibliography

- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78:15169–15211.
- Jiang, M., Machiraju, R., and Thompson, D. (2005). Detection and visualization of vortices. the visualization handbook.
- Kalliamvakou, E., Damian, D., Blincoe, K., Singer, L., and German, D. M. (2015). Open source-style collaborative development practices in commercial projects using github. In *2015 IEEE/ACM 37th IEEE international conference on software engineering*, volume 1, pages 574–585. IEEE.
- Karasavvas, K., Baldock, R., and Burger, A. (2004). Bioinformatics integration and agent technology. *Journal of biomedical informatics*, 37(3):205–219.
- Karopka, T., Schmuhl, H., and Demski, H. (2014). Free/libre open source software in health care: a review. *Healthcare informatics research*, 20(1):11–22.
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., and Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, 15:104–116.
- Kiah, M. L. M., Haiqi, A., Zaidan, B., and Zaidan, A. (2014). Open source emr software: Profiling, insights and hands-on analysis. *Computer methods and programs in biomedicine*, 117(2):360–382.
- Koopman, B., Zuccon, G., Nguyen, A., Bergheim, A., and Grayson, N. (2015). Automatic icd-10 classification of cancers from free-text death certificates. *International journal of medical informatics*, 84(11):956–965.
- Korolev, V. and Joshi, A. (2014). PROB: A tool for Tracking Provenance and Reproducibility of Big Data Experiments. In *Reproduce '14. HPCA 2014*.
- Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., Forshee, R., Walderhaug, M., and Botsis, T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *Journal of biomedical informatics*, 73:14–29.
- Kreuzthaler, M. and Schulz, S. (2015). Detection of sentence boundaries and abbreviations in clinical narratives. In *BMC medical informatics and decision making*, volume 15, pages 1–13. Springer.
- Kuhn, M., Letunic, I., Jensen, L. J., and Bork, P. (2016). The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079.
- Labropoulou, P., Galanis, D., Lempesis, A., Greenwood, M., Knoth, P., Eckart de Castilho, R., Sachtouris, S., Georgantopoulos, B., Anastasiou, L., Martziou, S., Katekina, G., Manola, N., and Piperidis, S. (2018). Openminted: A platform facilitating text mining of scholarly content. In *WOSP 2018 Workshop Proceedings*, Luxemburg.

- European Language Resources Association (ELRA). Held as a workshop at the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).
- Lamurias, A. and Couto, F. (2019). *Text Mining for Bioinformatics Using Biomedical Literature*, page 602–611.
- Li, X., Ng, S.-K., and Wang, J. T. (2013). *Biological data mining and its applications in healthcare*, volume 8. World scientific.
- Liao, X., Xiao, L., Yang, C., and Lu, Y. (2014). Milkyway-2 supercomputer: system and application. *Frontiers of Computer Science*, 8:345–356.
- Ljubicic, V., Ketikidis, P. H., and Lazuras, L. (2020). Drivers of intentions to use health-care information systems among health and care professionals. *Health Informatics Journal*, 26(1):56–71.
- Louie, B., Mork, P., Martin-Sanchez, F., Halevy, A., and Tarczy-Hornoch, P. (2007). Data integration and genomic medicine. *Journal of biomedical informatics*, 40(1):5–16.
- Luna, D., Almerares, A., Mayan, J. C., de Quirós, F. G. B., and Otero, C. (2014). Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. *Healthcare informatics research*, 20(1):3–10.
- Luo, J., Wu, M., Gopukumar, D., and Zhao, Y. (2016). Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights*, 8:1.
- Ly, T., Pamer, C., Dang, O., Brajovic, S., Haider, S., Botsis, T., Milward, D., Winter, A., Lu, S., and Ball, R. (2018). Evaluation of natural language processing (nlp) systems to annotate drug product labeling with meddra terminology. *Journal of biomedical informatics*, 83:73–86.
- Mann, P. S. (2010). *Introductory statistics*. John Wiley & Sons.
- Marafino, B. J., Boscardin, W. J., and Dudley, R. A. (2015). Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to icu risk stratification from nursing notes. *Journal of biomedical informatics*, 54:114–120.
- Marien, S., Krug, B., and Spinewine, A. (2017). Electronic tools to support medication reconciliation: a systematic review. *Journal of the American Medical Informatics Association*, 24(1):227–240.
- Martínez, P., Martínez, J. L., Segura-Bedmar, I., Moreno-Schneider, J., Luna, A., and Revert, R. (2016). Turning user generated health-related content into actionable knowledge through text analytics services. *Computers in Industry*, 78:43–56.

Bibliography

- McColl, R. C., Ediger, D., Poovey, J., Campbell, D., and Bader, D. A. (2014). A performance evaluation of open source graph databases. In *Proceedings of the first workshop on Parallel programming for analytics applications*, pages 11–18.
- McCormick, M., Liu, X., Jomier, J., Marion, C., and Ibanez, L. (2014). Itk: enabling reproducible research and open science. *Frontiers in neuroinformatics*, 8:13.
- McDonald, C. J., Schadow, G., Barnes, M., Dexter, P., Overhage, J. M., Mamlin, B., and McCoy, J. M. (2003). Open source software in medical informatics—why, how and what. *International journal of medical informatics*, 69(2-3):175–184.
- McIlroy, M. D., Buxton, J., Naur, P., and Randell, B. (1968). Mass-produced software components. In *Proceedings of the 1st international conference on software engineering, Garmisch Pattenkirchen, Germany*, pages 88–98.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., et al. (2016). Point of view: How open science helps researchers succeed. *Elife*, 5:e16800.
- MedDRA (2020). Meddrahierarchy. Accessed: 2020-10-16.
- Mehrabi, S., Krishnan, A., Sohn, S., Roch, A. M., Schmidt, H., Kesterson, J., Beesley, C., Dexter, P., Schmidt, C. M., Liu, H., et al. (2015). Deepen: A negation detection system for clinical text incorporating dependency relation into negex. *Journal of biomedical informatics*, 54:213–219.
- Menger, V., Scheepers, F., and Spruit, M. (2018a). Comparing deep learning and classical machine learning approaches for predicting inpatient violence incidents from clinical text. *Applied Sciences*, 8(6):981.
- Menger, V., Scheepers, F., van Wijk, L. M., and Spruit, M. (2018b). Deduce: A pattern matching method for automatic de-identification of dutch medical text. *Telematics and Informatics*, 35(4):727–736.
- Menger, V., Spruit, M., Van Est, R., Nap, E., and Scheepers, F. (2019). Machine learning approach to inpatient violence risk assessment using routinely collected clinical notes in electronic health records. *JAMA network open*, 2(7):e196709–e196709.
- Meulendijk, M. C., Spruit, M. R., Drenth-van Maanen, A. C., Numans, M. E., Brinkkemper, S., Jansen, P. A., and Knol, W. (2015a). Computerized decision support improves medication review effectiveness: an experiment evaluating the strip assistant’s usability. *Drugs & aging*, 32:495–503.
- Meulendijk, M. C., Spruit, M. R., Jansen, P. A., Numans, M. E., and Brinkkemper, S. (2015b). Stripa: A rule-based decision support system for medication reviews in primary care. In *ECIS*.
- Meulendijk, M. C., Spruit, M. R., Willeboordse, F., Numans, M. E., Brinkkemper, S., Knol, W., Jansen, P. A., and Askari, M. (2016). Efficiency of clinical decision support systems improves with experience. *Journal of medical systems*, 40:1–7.

- Meystre, S. and Haug, P. J. (2006). Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *Journal of biomedical informatics*, 39(6):589–599.
- Midha, V. and Palvia, P. (2012). Factors affecting the success of open source software. *Journal of Systems and Software*, 85(4):895–905.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mitchell, K. J., Becich, M. J., Berman, J. J., Chapman, W. W., Gilbertson, J., Gupta, D., Harrison, J., Legowski, E., and Crowley, R. S. (2004). Implementation and evaluation of a negation tagger in a pipeline-based system for information extraction from pathology reports. In *MEDINFO 2004*, pages 663–667. IOS Press.
- Mozzicato, P. (2007). Standardised meddra queries: their role in signal detection. *Drug safety*, 30:617–619.
- Muilu, J., Peltonen, L., and Litton, J.-E. (2007). The federated database—a basis for biobank-based post-genome studies, integrating phenome and genome data from 600 000 twin pairs in europe. *European Journal of Human Genetics*, 15(7):718–723.
- Müller, H.-M., Auken, K. V., Li, Y., and Sternberg, P. W. (2018). Textpresso central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. In *BMC Bioinformatics*.
- Munger, M. A. (2010). Polypharmacy and combination therapy in the management of hypertension in elderly patients with co-morbid diabetes mellitus. *Drugs & aging*, 27:871–883.
- Ngafeeson, M. N. (2015). Healthcare information systems opportunities and challenges. *Encyclopedia of Information Science and Technology, Third Edition*, pages 3387–3395.
- Ni, Y., Kennebeck, S., Dexheimer, J. W., McAneney, C. M., Tang, H., Lingren, T., Li, Q., Zhai, H., and Solti, I. (2015). Automated clinical trial eligibility prescreening: increasing the efficiency of patient identification for clinical trials in the emergency department. *Journal of the American Medical Informatics Association*, 22(1):166–178.
- Nolden, M., Zelzer, S., Seitel, A., Wald, D., Müller, M., Franz, A. M., Maleike, D., Fangerau, M., Baumhauer, M., Maier-Hein, L., et al. (2013). The medical imaging interaction toolkit: challenges and advances: 10 years of open-source development. *International journal of computer assisted radiology and surgery*, 8:607–620.
- Octoverse (2018). The state of octoverse.
- O’Leary, D. (1988). Expert system prototyping as a research tool. In *Applied expert systems*, pages 17–32. North-Holland Amsterdam.

Bibliography

- Omta, W. A., van Heesbeen, R. G., Pagliero, R. J., van der Velden, L. M., Lelieveld, D., Nellen, M., Kramer, M., Yeong, M., Saeidi, A. M., Medema, R. H., et al. (2016). Hc stratominer: A web-based tool for the rapid analysis of high-content datasets. *Assay and drug development technologies*, 14(8):439–452.
- Omta, W. A., van Heesbeen, R. G., **Shen, Z.**, Feelders, A. J., Brinkhuis, M., Egan, D. A., and **Spruit, M.** (2020a). Purifyr: An r package for highly automated, reproducible variable extraction and standardization. *Systems Medicine*, 3(1):1–7. <https://api.semanticscholar.org/CorpusID:215898718>.
- Omta, W. A., van Heesbeen, R. G., **Z.Shen**, de Nobel, J., Robers, D., van der Velden, L. M., Medema, R. H., Siebes, A. P., Feelders, A. J., **S. Brinkkemper**, Klumperman, J. S., **M. Spruit**, Brinkhuis, M. J., and Egan, D. A. (2020b). Combining supervised and unsupervised machine learning methods for phenotypic functional genomics screening. *SLAS Discovery*, 25(6):655–664. <https://doi.org/10.1177/2472555220919345>.
- Open Source Clinical Software (2018). Source codes of open source clinical software.
- Osheroff, J. A., Teich, J. M., Levick, D., Saldana, L., Velasco, F., Sittig, D. F., Rogers, K. M., and Jenders, R. A. (2012). *Improving outcomes with clinical decision support: an implementer’s guide*. CRC Press.
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M. R., Zschocke, J., and Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in bioinformatics*, 15(2):256–278.
- Pachidi, S., Spruit, M., and Van De Weerd, I. (2014). Understanding users’ behavior with software operation data mining. *Computers in Human Behavior*, 30:583–594.
- Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., Oortwijn, W., MacGillivray, S., et al. (2005). What is ehealth (4): a scoping exercise to map the field. *Journal of medical Internet research*, 7(1):e391.
- Pandey, A., Kreimeyer, K., Foster, M., Dang, O., Ly, T., Wang, W., Forshee, R., and Botsis, T. (2019). Adverse event extraction from structured product labels using the event-based text-mining of health electronic records (ether) system. *Health informatics journal*, 25(4):1232–1243.
- Passi1, K., Lane, L., Madria, S., Sakamuri, B. C., Mohania, M., and Bhowmick, S. (2002). A model for xml schema integration. In *E-Commerce and Web Technologies: Third International Conference, EC-Web 2002 Aix-en-Provence, France, September 2–6, 2002 Proceedings 3*, pages 193–202. Springer.
- Patrick, J. and Li, M. (2010). High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *Journal of the American Medical Informatics Association*, 17(5):524–527.

- Peffers, 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.
- Pestian, J. P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., Cohen, K. B., Hurdle, J., and Brew, C. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical informatics insights*, 5:BII–S9042.
- Pletscher-Frankild, S., Pallegà, A., Tsafo, K., Binder, J. X., and Jensen, L. J. (2015). Diseases: Text mining and data integration of disease–gene associations. *Methods*, 74:83 – 89. Text mining of biomedical literature.
- plotly visualization (2018). Modern analytics apps for the enterprise.
- Pop, D.-P. and Altar, A. (2014). Designing an mvc model for rapid web application development. *Procedia Engineering*, 69:1172–1179.
- Przybyła, P., Shardlow, M., Aubin, S., Bossy, R., Eckart de Castilho, R., Piperidis, S., McNaught, J., and Ananiadou, S. (2016). Text mining resources for the life sciences. *Database*, 2016:baw145.
- Pubmed Central (2019). Pubmed central (pmc).
- Pubmed Open Access (2019). Pubmed open access subset.
- Purkayastha, S., Allam, R., Maity, P., and Gichoya, J. W. (2019). Comparison of open-source electronic health record systems based on functional and user performance criteria. *Healthcare informatics research*, 25(2):89–98.
- pyLDAvis (2018). pyldavis: Python library for interactive topic model visualization.
- Raghupathi, W. and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2:1–10.
- Rago, A., Ramos, F. M., Velez, J. I., Díaz Pace, J. A., and Marcos, C. (2016). Textract: a web-based tool for building nlp-enabled applications. In *Simposio Argentino de Ingeniería de Software (ASSE 2016)-JAIIO 45 (Tres de Febrero, 2016)*.
- Rajdho, A. and Biba, M. (2013). *Plugging Text Processing and Mining in a Cloud Computing Framework*, pages 369–390. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Ramesh, V. and Ram, S. (1995). A methodology for interschema relationship identification in heterogeneous databases. In *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, volume 3, pages 263–272. IEEE.
- Rehim, S. A., DeMoor, S., Olmsted, R., Dent, D. L., and Parker-Raley, J. (2017). Tools for assessment of communication skills of hospital action teams: a systematic review. *Journal of surgical education*, 74(2):341–351.

Bibliography

- Rehurek, R. and Sojka, P. (2010). Software framework for topic modelling with large corpora. *Irec 2010 workshop on new challenges for nlp frameworks. University of Malta, Msida, Malta*, pages 45–50.
- Reynolds, C. J. and Wyatt, J. C. (2011). Open source, open standards, and health care information systems. *Journal of medical Internet research*, 13(1):e1521.
- Rizzo, G. and Troncy, R. (2012). Nerd: a framework for unifying named entity recognition and disambiguation extraction tools. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 73–76.
- Roberts, K., Demner-Fushman, D., and Topping, J. M. (2017). Overview of the tac 2017 adverse reaction extraction from drug labels track. In *TAC*.
- Russell, P. H., Johnson, R. L., Ananthan, S., Harnke, B., and Carlson, N. E. (2018). A large-scale analysis of bioinformatics code on github. *PloS one*, 13(10):e0205898.
- Russo, E., Sittig, D. F., Murphy, D. R., and Singh, H. (2016). Challenges in patient safety improvement research in the era of electronic health records. In *Healthcare*, volume 4-4, pages 285–290. Elsevier.
- Sallevelt, B., Huibers, L., Op Heij, J., Egberts, T., Puijenbroek, E., **Shen, Z., Spruit, M.**, Jungo, K., Rodondi, N., Dalleur, O., Spinewine, A., Jennings, E., O’Mahony, D., Wilting, I., and Knol, W. (2021). Frequency and acceptance of clinical decision support system-generated stopp/start signals for hospitalised older patients with polypharmacy and multimorbidity. *Drugs & Aging*, 39. <https://doi.org/10.1007/s40266-021-00904-z>.
- Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., and Chute, C. G. (2010). Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513.
- Schots, M. (2014). On the use of visualization for supporting software reuse. In *Companion Proceedings of the 36th International Conference on Software Engineering*, pages 694–697.
- Shekhani, R., Steinacher, L., Swen, J. J., and Ingelman-Sundberg, M. (2020). Evaluation of current regulation and guidelines of pharmacogenomic drug labels: opportunities for improvements. *Clinical Pharmacology & Therapeutics*, 107(5):1240–1255.
- Shen, Z., Meulendijk, M., Knol, W., Huibers, L., Wilting, I., Jansen, P., and Spruit, M. (2016a). Stripa investigational medical device dossier (imdd). Technical report, UU/UMCU. <https://marcospruit.nl/pub/2016%20-%20Shen%20et%20al%20-%20IMDD.pdf>.

- Shen, Z., Meulendijk, M., and Spruit, M. (2016b). A federated information architecture for multinational clinical trials: Stripa revisited. *24th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2016_prototypes/2.
- Shen, Z. and Spruit, M. (2019). Locate: A web application to link open-source clinical software with literature. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019) - HEALTHINF*, pages 294–301. INSTICC, SciTePress. <https://doi.org/10.5220/0007378702940301>.
- Shen, Z. and Spruit, M. (2019). A systematic review of open source clinical software on github for improving software reuse in smart healthcare. *Applied Sciences*, 9(1). <https://www.mdpi.com/2076-3417/9/1/150>.
- Shen, Z. and Spruit, M. (2021). Automatic extraction of adverse drug reactions from summary of product characteristics. *Applied Sciences*, 11(6). <https://www.mdpi.com/2076-3417/11/6/2663>.
- Shen, Z., van Krimpen, H., and Spruit, M. (2019a). A lightweight api-based approach for building flexible clinical nlp systems. *Journal of Healthcare Engineering*, 2019(1):3435609. <https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/3435609>.
- Shen, Z. and Wang, X. (2019). Source code.
- Shen, Z., Wang, X., and Spruit, M. (2019b). Big data framework for scalable and efficient biomedical literature mining in the cloud. In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval, NLPPIR '19*, page 80–86, New York, NY, USA. Association for Computing Machinery. <https://doi.org/10.1145/3342827.3342843>.
- Shvachko, K., Kuang, H., Radia, S., and Chansler, R. (2010). The hadoop distributed file system. In *2010 IEEE 26th symposium on mass storage systems and technologies (MSST)*, pages 1–10. Ieee.
- Sievert, C. and Shirley, K. (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.
- Spruit, M. and Lytras, M. (2018). Applied data science in patient-centric healthcare: Adaptive analytic systems for empowering physicians and patients.
- Stack Overflow Survey (2018). 2018 stack overflow developers survey.
- Steinmacher, I., Silva, M. A. G., Gerosa, M. A., and Redmiles, D. F. (2015). A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*, 59:67–85.

Bibliography

- Steinman, M. A., Seth Landefeld, C., Rosenthal, G. E., Berthenthal, D., Sen, S., and Kaboli, P. J. (2006). Polypharmacy and prescribing quality in older people. *Journal of the American Geriatrics Society*, 54(10):1516–1523.
- Stewart, K. (2016). Studies of success in open source software projects 1. In *Successful OSS Project Design and Implementation*, pages 131–148. Routledge.
- Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., and Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1):17.
- Tafti, A. P., Badger, J. C., LaRose, E. R., Shirzadi, E., Mahnke, A. N., Mayer, J., Ye, Z., Page, D., and Peissig, P. L. (2017). Adverse drug event discovery using biomedical literature: A big data neural network adventure. In *JMIR medical informatics*.
- Tanaka, Y., Chen, H. Y., Belloni, P., Gisladdottir, U., Kefeli, J., Patterson, J., Srinivasan, A., Zeitz, M., Sirdeshmukh, G., Berkowitz, J., et al. (2024). Onsites (on-label side effects resource) database: Extracting adverse drug events from drug labels using natural language processing models. *medRxiv*, pages 2024–03.
- Towns, J., Cockerill, T., Dahan, M., Foster, I. T., Gaither, K. P., Grimshaw, A. S., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., and Wilkins-Diehr, N. (2014). Xsede: Accelerating scientific discovery. *Computing in Science & Engineering*, 16:62–74.
- Uzuner, Ö. (2009). Recognizing obesity and comorbidities in sparse data. *Journal of the American Medical Informatics Association*, 16(4):561–570.
- Uzuner, Ö., Solti, I., and Cadag, E. (2010). Extracting medication information from clinical text. *Journal of the American Medical Informatics Association*, 17(5):514–518.
- van de Weerd, I. and Brinkkemper, S. (2009). Meta-modeling for situational analysis and design methods. In *Handbook of research on modern systems analysis and design technologies and applications*, pages 35–54. IGI Global.
- van Tuijl, G., Leenen, W., **Z. Shen**, van de Weerd, I., and **S. Brinkkemper** (2011). Prioritizing requirements: An experiment to test the perceived reliability, usability and time consumption of bubblesort and the analytical hierarchy process. In *Proceedings of the International Requirements Engineering Efficiency Workshop (REEW 2011)*, pages 37–48. <https://api.semanticscholar.org/CorpusID:53900285>.
- Wager, K. A., Lee, F. W., and Glaser, J. P. (2021). *Health care information systems: a practical approach for health care management*. John Wiley & Sons.
- Wang, W., Bleakley, B., Ju, C., Kyi, V., Tan, P., Choi, H., Huang, X., Zhou, Y., Wood, J., Wang, D., et al. (2017). Aztec: A platform to render biomedical software findable, accessible, interoperable, and reusable. *arXiv preprint arXiv:1706.06087*.
- Wikipedia (2018). Comparison of source code hosting facilities.

- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.
- Wright, R. M., Sloane, R., Pieper, C. F., Ruby-Scelsi, C., Twersky, J., Schmader, K. E., and Hanlon, J. T. (2009). Underuse of indicated medications among physically frail older us veterans at the time of hospital discharge: results of a cross-sectional analysis of data from the geriatric evaluation and management drug study. *The American journal of geriatric pharmacotherapy*, 7(5):271–280.
- Wu, L., Ingle, T., Liu, Z., Zhao-Wong, A., Harris, S., Thakkar, S., Zhou, G., Yang, J., Xu, J., Mehta, D., et al. (2019). Study of serious adverse drug reactions using fda-approved drug labeling and meddra. *BMC bioinformatics*, 20:129–139.
- Xing, Y., Wu, C., Yang, X., Wang, W., Zhu, E., and Yin, J. (2018). Parabtm: A parallel processing framework for biomedical text mining on supercomputers. In *Molecules*.
- Yang, C., Yu, M., Hu, F., Jiang, Y., and Li, Y. (2017a). Utilizing cloud computing to address big geospatial data challenges. *Computers, Environment and Urban Systems*, 61:120 – 128. Geospatial Cloud Computing and Big Data.
- Yang, C. P., Huang, Q., Li, Z., Liu, K., and Hu, F. (2017b). Big data and cloud computing: innovation opportunities and challenges. *Int. J. Digital Earth*, 10:13–53.
- Yang, Y. C., Islam, S. U., Noor, A., Khan, S., Afsar, W., and Nazir, S. (2021). Influential usage of big data and artificial intelligence in healthcare. *Computational and Mathematical Methods in Medicine*, 2021.
- Ye, Z., Tafti, A. P., He, K. Y., Wang, K., and He, M. M. (2016). Sparktext: Biomedical text mining on big data framework. *PloS one*, 11(9):e0162721.
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., and Stoica, I. (2016). Apache spark: A unified engine for big data processing. *Commun. ACM*, 59(11):56–65.
- Zaimi, A., Ampatzoglou, A., Triantafyllidou, N., Chatzigeorgiou, A., Mavridis, A., Chaikalis, T., Deligiannis, I., Sfetsos, P., and Stamelos, I. (2015). An empirical study on the reuse of third-party libraries in open-source software development. In *Proceedings of the 7th Balkan Conference on Informatics Conference*, pages 1–8.
- Zelkowitz, M. V. and Wallace, D. R. (1998). Experimental models for validating technology. *Computer*, 31(5):23–31.
- Zettinig, O., Shah, A., Hennemersperger, C., Eiber, M., Kroll, C., Kübler, H., Maurer, T., Milletari, F., Rakerseder, J., Schulte zu Berge, C., et al. (2015). Multimodal image-guided prostate fusion biopsy based on automatic deformable registration. *International journal of computer assisted radiology and surgery*, 10:1997–2007.

Bibliography

- Zhang, M. W. and Ho, R. (2017). Personalized reminiscence therapy m-health application for patients living with dementia: Innovating using open source code repository. *Technology and Health Care*, 25(1):153–156.
- Zhang, Y., Qiu, M., Tsai, C., Hassan, M. M., and Alamri, A. (2017). Health-cps: Health-care cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1):88–95.
- Zhao, H. and Ram, S. (2007). Combining schema and instance information for integrating heterogeneous data sources. *Data & Knowledge Engineering*, 61(2):281–303.
- Ziegler, P. and Dittrich, K. R. (2004). Three decades of data integration—all problems solved? In *Building the Information Society: IFIP 18th World Computer Congress Topical Sessions 22–27 August 2004 Toulouse, France*, pages 3–12. Springer.