10-2015

ezDI's Semantics-Enhanced Linguistic, NLP, and ML Approach for Health Informatics

Raxit Goswami

Neil Shah

Amit P. Sheth
Wright State University - Main Campus, amit@sc.edu

ezDI's semantics-enhanced linguistic, NLP, and ML approach for health informatics

Raxit Goswami*, Neil Shah* and Amit Sheth*,**
ezDI Inc, Louisville, KY and Ahmedabad, India. ** Kno.e.sis-Wright State Univ.

Abstract

ezDI uses large and extensive knowledge graph to enhance linguistics, NLP and ML techniques to improve structured data extraction from millions of EMR records. It then normalizes it, and maps it with various computer-processable nomenclature such as SNOMED-CT, RxNorm, ICD-9, ICD-10, CPT, and LOINC. Furthermore, it applies advanced reasoning that exploited domain-specific and hierarchical relationships among entities in the knowledge graph to make the data actionable. These capabilities are part of its highly scalable AWS deployed heath intelligence platform that support healthcare informatics applications, including Computer Assisted Coding (CAC), Computerized Document Improvement (CDI), compliance and audit, and core measures and utilization, as well as support improved decision making that involve identification of patients at risk, patterns in diseases, outcome prediction, etc. This paper focuses on the key role of its semantic approach and techniques.

Keywords: knowledge graph, graph reasoning, health informatics applications, semantics-enhanced NLP

Introduction and Background

There is a radical shift in healthcare with healthcare reforms. The pressure to deliver transparent quality of care using evidence-based medicine has created even more challenges for healthcare providers to meet government and legislative demands to reduce healthcare costs. Health systems and providers are buried under enormous amounts of data generated by a variety of healthcare platforms such as Electronic Health Record systems, patient portals, medical device data, and much more—data that is extremely difficult to classify and manage, let alone analyze. Traditional data analytics tools such as Data Warehousing and Business Intelligence software are limited to work only on structured data. Furthermore, the frustration mounts while investment in new technologies like EHR or analytics products, structured around 20% of the healthcare's data, fail to provide comprehensive, credible, and actionable information combining data from disparate systems (structured and unstructured) to meet with hospitals' strategic goals for better, safer care at lower costs.

So, where is the solution to make clinical data easily accessible, meaningful, and actionable with real-time interventions to support value-based care and actually improving patient outcomes? ezDI, Inc. [ezDI] was founded with a mission to improve patient care and reduce overall healthcare cost using the latest research and technologies. We believe the key is to abstract only meaningful and actionable clinical data from an ever-increasingly unstructured healthcare dataset, and to make it easily accessible and useful for healthcare professionals.

This means developing the technology that can convert the unstructured data that makes up significant portions of clinical records into structured data/knowledge, normalize it, and map it with various computer-processable nomenclature such as SNOMED-CT, RxNorm, ICD-9, ICD-10, CPT, and LOINC, and then apply advanced reasoning to make the data actionable. This involves use of various clinical informatics applications, including Computer Assisted Coding (CAC), Computerized Document Improvement (CDI), compliance and audit, core measures and utilization, identification of patients at risk and patterns in diseases, outcome prediction, etc.

At ezDI, it begins by extracting meaningful and actionable data from tens of millions of patient datasets at scale. The technology that has traditionally been used for this is that of natural language processing (NLP). While NLP does generate clinical insights, the quality of extraction in converting unstructured data into structured data/knowledge and its ability to reason, leading to deeper insights and decision support, is very limited. That is why ezDI developed Knowledge Graphs that can provide meaning to the abstracted data. The scale is equally important for processing large amounts of data and providing real-time insights; ezDI built on AWS HIPAA Compliant Cloud platform and makes integration and scalability a no brainer for health systems resulting in faster deployment at significantly lower costs than traditional legacy platforms. We also understand that extensibility is the key to responding to rapidly changing business priorities. The ezDI health intelligence platform provides comprehensive APIs to enable integration with a variety of platforms such as healthcare content service providers.

Tackling the healthcare data storm gets harder everyday; the challenges of providing greater quality of care in a more efficient and cost-effective model is a common goal across all healthcare delivery organizations. Our ability to provide real-time clinical insights at scale to help healthcare providers make better clinical decisions is revolutionary and a stepping stone in the Healthcare Data Mining and Analytics space. We successfully demonstrated our abilities to improve overall patient care and reduce healthcare costs, aiding major hospitals from a 500-bed academic university hospital to a 100-bed critical access hospital, and are being promoted to more than 500,000 healthcare providers by key healthcare alliances in Texas, New York, and North Carolina.

Challenge: Healthcare data size and complexity

Healthcare data is rapidly evolving and is locked in disparate systems with different permissions and in different formats. More than 80% of medical data is unstructured in terms of dictation-transcription reports, such as office visits, discharge summaries, progress notes, scanned reports, and much more. This data and the information it contains is difficult to classify and manage, let alone analyze.

Traditional data analytics software works reasonably well with structured data. The processes to abstract meaningful and actionable data from unstructured data is tedious, retrospective, resource-intensive, and time-consuming, and as a result, very expensive.
Furthermore, the pressure to deliver transparent, quality healthcare using evidence-based medicine has created additional challenges for healthcare providers to meet government and

legislative demands to reduce healthcare costs. Hospitals and providers are buried under the enormous amounts of data generated to meet these initiatives, and making sense of the data is a monumental task.

Step forward in organizing medical data to improve outcome

We believe the key to solving healthcare data challenges is to organize and abstract meaningful and actionable medical data while weeding out the noise. This means developing the technology that can convert unstructured data into structured data, normalize it, and map it with various computer-processable nomenclature such as SNOMED-CT, RxNorm, ICD-9, ICD-10, CPT, LOINC, and much more. The complexity of the involved tasks has led ezDI industry leading solutions consisting of:

- Development of extensive background knowledge, termed ezKG (for ezDI's very large knowledge graph), that uses existing ontologies, medical knowledge, as well as techniques for semi-automatic knowledge extraction and curation to continue expansion and refreshing of extensive knowledge used for subsequent semantics-enhanced tasks.
- Exploiting state-of-the-art linguistic, NLP, and machine learning (ML) techniques, and using the semantic approach of employing extensive domain knowledge to improve them and to enable natural language understanding (NLU) and structured semantic data extraction.
- Using rule-based, statistical, and graph-based reasoning to support advanced analysis and decision support.
- Modular applications and solutions developed ground up for scalable, cloud-based deployment to meet the market needs.

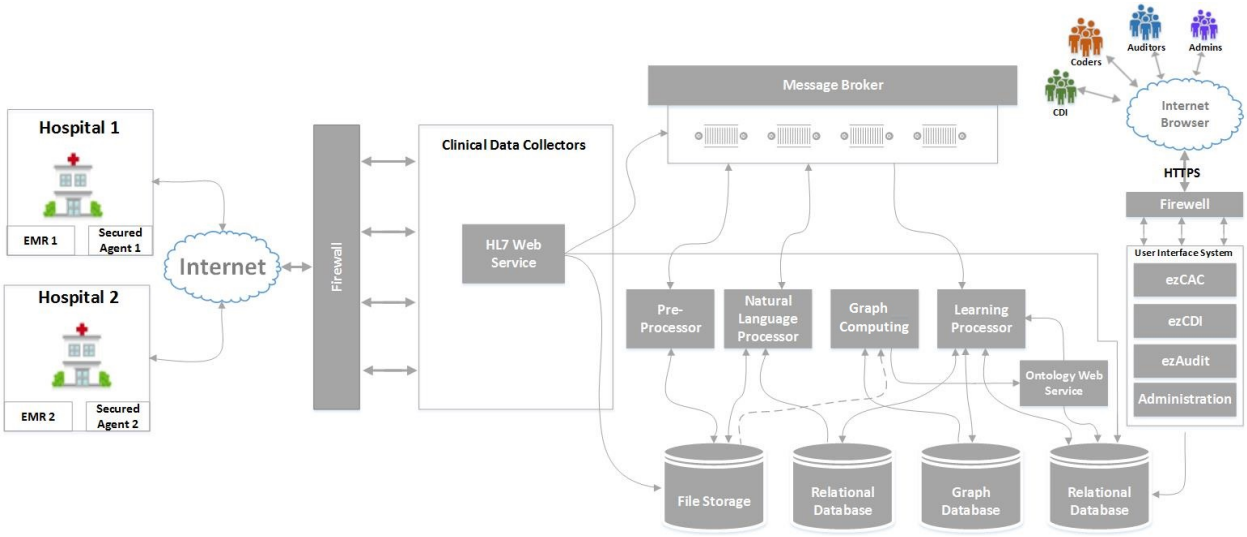

Figure 1 shows software architecture of AWS deployed ezDI's health intelligence platform. ezKB is part of Graph Database, and Semantic Computing is generically captured as part of Graph Computing to minimize education required for most enterprise customers.

As part of the presentations (and a live demo, if desired), we will focus on two objectives: (a) illustrating the unique role of the semantic approach in ezDI solutions, (b) our ability to scale computations in which semantic techniques for extraction and reasoning plays a key role in supporting the needs of ezDI's typical customers who are a few hundred-bed hospitals. Key examples include:

a. How does ezDI leverage the semantics of the large amount of EMR data and the existing knowledge graph to continue to expand and keep ezKG updated [Perera et al 2012, Perera et al 2014]
b. How can we improve upon the current state-of-the-art in NLP for the quality of core tasks such as named entity recognition, coreference resolution, and implicit entity recognition [Perera et al 2015], to ultimately improve the quality of clinical text understanding and NLU, with the help of a large domain-specific knowledge (embodied in ezKG) [Perera et al 2013]
c. What have been pros and cons of using RDF in our technology
d. How have we exploited domain-specific and hierarchical relationships among entities in advanced graph-based reasoning tasks, such as identification and ranking of ICD-10-PCS codes, for achieving high quality for CAC and CDI. This is crucial since ICD-10-PCS represents over an order of magnitude higher complexity than the current ICD-9-CM coding. In the process, we will share insights on why non-semantic solutions, such as current approaches for ICD-9 CAC based on rule-based NLP or statistical machine learning are unlikely to give the accuracy needed as evidenced in part from the failure of competing solutions in the market. A key take away is that use of a very comprehensive medical knowledge base helps to avoid very large training and deployment times needed by current solutions that do not use such a knowledge base.

Conclusion

We conclude with the following observations:

- There is unprecedented market need to exploit unstructured data in healthcare as part of applications such as CAC and CDI.
- This requires deep clinical text understanding including extraction of structured knowledge.
- State-of-the-art rule-based NLP and Machine Learning technologies do not give high quality solutions, take long deployment times and do not scale. ezDI has found use of semantics supported by large and growing medical knowledge indispensable to enhance the state-of-the-art technologies, and has also been able to deploy the solution entirely on a cloud platform to make it highly scalable.

- Customers find it easier to understand and appreciate terms graph processing and knowledge graph, which is what we used for communication instead of semantics, semantic web, and reasoning.
- Meeting the needs of hospitals with hundreds of beds is now a reality, not a future.

About the Authors: Raxit is the lead semantic engineer, Neil is the VP, Product Engineering, and Amit is the primary research & technology advisor for ezDI. Kno.e.sis center, led by Prof. Sheth is a research collaborator of ezDI since the founding of ezDI.

Acknowledgements: Many ezDI researchers/developers and collaborating researchers deserve credit for extensive technology covered in this paper. Special mentions go to Sujan Perera, Parth Pathak and Suhas Nair.

References

[ezDI] ezDI Healthcare Data Intelligence Solutions, including ezCAC, ezDCI and ezAudit, accessible at http://www.ezdi.us/

[Perera et al 2012] Sujan Perera, Cory Henson, Krishnaprasad Thirunarayan, Amit Sheth, Suhas Nair, 'Data driven knowledge acquisition method for domain knowledge enrichment in the healthcare', IEEE International Conference Bioinformatics and Biomedicine 2012 (BIBM '12), pp.1-8, 4-7 Oct. 2012

[Perera et al 2013] Sujan Perera, Amit Sheth, Krishnaprasad Thirunarayan, Suhas Nair and Neil Shah, 'Challenges in Understanding Clinical Notes: Why NLP Engines Fall Short and Where Background Knowledge Can Help', International Workshop on Data management & Analytics for healthcaRE (DARE) at ACM Conference of Information and Knowledge Management (CIKM), pp. 21-26, Burlingame, USA, Nov 1, 2013, DOI: 10.1145/2512410.2512427

[Perera et al 2014] Sujan Perera, Cory Henson, Krishnaprasad Thirunarayan, Amit Sheth, Suhas Nair, 'Semantics Driven Approach for Knowledge Acquisition From EMRs', IEEE Journal of Biomedical and Health Informatics, vol.18, no.2, pp.515-524, March 2014, doi: 10.1109/JBHI.2013.2282125, PMID: 24058038

[Perera et al 2015] Sujan Perera, Pablo Mendes, Amit Sheth, Krishnaprasad Thirunarayan, Adarsh Alex, Christopher Heid, Greg Mott, 'Implicit Entity Recognition in Clinical Documents', In Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM), June 2015.