

Structured Data Retrieval and Analysis of HL7 v2 Messages with Elasticsearch

Ka Yung CHENG^{a,1}, Santiago PAZMINO^a, Hao QIAN^a, Björn BERGH^a and Björn SCHREIWEIS^a

^a*Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany*

Abstract. The era of the electronic health record (EHR) requires lots of semantic interoperability for data sharing and reusability. We select HL7 v2 messages as the most common structured data type in hospital information systems, to investigate the plausibility of using Elasticsearch (ES) as a healthcare search engine and data analytics tool. Due to the facts, Elasticsearch can be integrated as a powerful searchable database for practical healthcare applications, to analyze structured healthcare data from various locations. It allows easy and efficient searching for complex query tasks.

Keywords. Healthcare data retrieval, elasticsearch, structured data

1. Introduction

The German Medical Informatics Initiative, a long-term strategic program funded by the German Federal Ministry of Education and Research, aims to establish a national infrastructure for integrating, sharing and reusing clinical data in patient care and biomedical research. At the University Hospital Schleswig-Holstein Medical Data Integration Center (MeDIC), we are committed to integrating routine care data as well as research data [1]. For this purpose, we retrieve and analyze data to harmonize the integration at the local level and gain valuable insights into patient health and treatment. Therefore, we need a scalable search engine to query large volumes of data which we integrate from our Hospital Information System (HIS) ORBIS. Knowing the kind of data involved in the MeDIC, but also the differences between structured (exact value) and unstructured (full text) searches, is important. The basis for standardized communication within the HIS usually is Health Level 7 (HL7). Hence, we focus in this study on how we can integrate and make use of structured HL7 v.2 messages.

2. Methods

HL7 v.2 ADT (Admission, Discharge or Transfer) messages are used to communicate patient demographics. We store the received messages from the HIS into Elasticsearch (ES) using Apache Nifi [2]. We divide our problems into two categories to develop analytic strategies for 632,056 ADT messages in October 2022: a) search phrase to find

¹ Corresponding Author: Ka Yung Cheng, email: KaYung.Cheng@uksh.de.

the distribution of ADT message types using count API, b) extract expressions applied in specific HL7 v2 message segments.

3. Results

We successfully use ES to execute healthcare data analysis. For HL7, we verify the distribution of ADT messages, such as if type A04 (Register) is less than the number of A01, or the sum of A01 and A05 is less than A08 (Update). We discover in PV1.2 an unexpected expression, "V". Furthermore, we find some unexpected study case changes, which could be due to information losses

4. Discussion

There are multiple ways to store and query data. We faced difficulties in choosing the appropriate query strategies and needed to test many possible implementations. We initially attempted to store HL7 messages as plain text with their original pipe pattern, and then add "running fields" to existing documents without re-indexing them or understanding their structure. But this "SCHEMA-ON-READ" process slows down the overall query process, especially with large-scale data.

5. Conclusions

With Elasticsearch, we can execute powerful queries on our structured medical data, e.g. HL7 v2 messages. The shown queries can also be easily extended to other elements (e.g. PID.8 - Administrative Sex), or even message types (e.g. HL7 ORU or HL7 BAR).

Acknowledgment

This research was funded by the German Federal Ministry for Education and Research (BMBF, grant number 01ZZ1802T).

References

- [1] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, Knaup-Gregori P, Bavendiek U, Dieterich C, Brors B, Kraus I. HiGHmed—an open platform approach to enhance care and research across institutional boundaries. *Methods Inf Med.* 2018 Jul;57(S 01):e66-81, doi: 10.3414/ME18-02-0002.
- [2] Cheng KY, Pazmino S, Schreiweis B. ETL Processes for integrating healthcare data—tools and architecture patterns. *Stud Health Technol Inform.* 2022 Nov;299:151-6, doi: 10.3233/SHTI220974.