Healthcare information system engineering: AI technologies and open source approaches

Shen, Z.

Summary

Healthcare Information Systems (HIS) play a pivotal role in modern today's healthcare, managing patient data, supporting clinical decisions, and enabling large-scale clincal trails. However, developing effective HIS faces numerous challenges, including structured and unstructed data in various format, massive data volume, regulatory compliance, and the increasing need of using AI for better decision making. Simultaneously, a large number of open source software or projects are created to address healthcare problems, with some well-known and successful ones like openEHR (open Electronic Health Records) and cTAKES (clinical Text Analysis and Knowledge Extraction System). These openly accessible OSS (Open Source Software) can benefit the HIS development, especially for circumstances where IT resources are restrained.

This dissertation investigates how artificial intelligence technologies and open source principles can accelerate HIS development to solve real-world clinical problems. The main research question is:

> **How can we employ artificial intelligence technologies – such as machine learning algorithms, knowledge systems and natural language processing techniques – based on open source principles to accelerate healthcare information system engineering in solving real-world clinical problems?**

We approached this research question through a practical, two-step methodology. The first part of the dissertation (chapter 2 and chapter 3) built a few healthcare information systems and prototypes to address specific clinical challenges, showing the real-world user cases of AI technologies in healthcare settings. Then, building on these practical experiences, we investigated how open source software and methodologies could accelerate HIS development. This progression from concrete system development to broader engineering methodology reflects our belief that effective HIS

engineering must be grounded in practical implementation while embracing open collaboration.

The research begins with a concrete clinical need – supporting polypharmacy reviews through rule-based clinical decision support system called the STRIP Assistant (STRIPA). To supoort its use in a multinational clinical trial, we design GDPR-compliant systems while maintaining data security and accessibility in chapter 2. STRIPA has a manual data entry process which is time consuming and error-prone. To address such issue, chapter 3 proposes a lightweight, API-based architecture for clinical NLP systems, which orchestrates various external NLP services to create flexible, cost-effective solutions that overcome the limitations of monolithic clinical NLP systems. We develop a prototype which demonstrates the potential of reducing manual data entry for HIS like STRIPA. The dissertation then continues on designing an automated NLP pipeline for extracting structured adverse drug reactions from European product labels in chapter 4. It shows how specialized NLP techniques can unlock structured knowledge from regulatory documents with high precision. chapter 5 presents a cloud-based big data framework for large-scale biomedical literature mining, providing a scalable solution for processing millions of research articles while maintaining cost-effectiveness. The need for such framework becomes apparent as the size of biomedical data grows rapidly.

The second part of the research focuses on exploring how open source software can enhance HIS engineering. A reproducible methodology is developed for systematic studies of open source clinical software in chapter 6, which provides insights into the landscape of available tools and their characteristics. Furthermore, we develop a web platform, named LOCATE, that bridges the gap between open source clinical software and scientific literature and enables developers to make more informed choices about software reuse. Except this first clinical decision support system, all the source codes of this research are publically accessible on GitHub with the aim of supporting open source software.