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).

Samenvatting

Healthcare Information Systems (HIS) spelen een cruciale rol in de moderne gezondheidszorg. Deze systemen beheren patiëntgegevens, ondersteunen klinische besluitvorming en faciliteren grootschalige klinische trials. De ontwikkeling van effectieve HIS wordt echter geconfronteerd met talrijke uitdagingen, waaronder de verwerking van gestructureerde en ongestructureerde data in verschillende formaten, enorme datavolumes, naleving van wet- en regelgeving, en de toenemende behoefte aan het gebruik van kunstmatige intelligentie voor betere besluitvorming. Tegelijkertijd wordt er een groot aantal open source projecten gecreëerd om gezondheidszorgproblemen aan te pakken. Prominente en succesvolle voorbeelden zijn openEHR (open Electronic Health Records) en cTAKES (clinical Text Analysis and Knowledge Extraction System). Deze publiek toegankelijke open source software (OSS) kan de ontwikkeling van HIS bevorderen, met name in omstandigheden waar IT-middelen beperkt zijn.

Dit proefschrift onderzoekt hoe kunstmatige intelligentie technologieën en open source principes de ontwikkeling van HIS kunnen versnellen om klinische problemen uit de praktijk op te lossen. De hoofdonderzoeksvraag luidt:

> **Hoe kunnen we kunstmatige intelligentie technologieën – zoals machine learning algoritmen, kennissystemen en natuurlijke taalverwerking technieken – op basis van open source principes inzetten om de ontwikkeling van healthcare information systems te versnellen bij het oplossen van klinische problemen in de dagelijkse praktijk?**

Deze onderzoeksvraag hebben we benaderd door middel van een praktische, tweestaps-methodologie. Het eerste deel van dit proefschrift (Hoofdstuk 2 en Hoofdstuk 3) ontwikkelde verschillende HIS en prototypes om specifieke klinische uitdagingen aan te pakken, waarmee praktische toepassingen van AI-technologieën in de

gezondheidszorg worden gedemonstreerd. Voortbouwend op deze praktische ervaringen hebben we onderzocht hoe open source software en methodologieën de ontwikkeling van HIS kunnen versnellen. Deze progressie van concrete systeemontwikkeling naar een bredere ontwikkelingsmethodologie weerspiegelt onze overtuiging dat effectieve HIS-ontwikkeling op praktische implementatie moet zijn gebaseerd en tegelijkertijd open samenwerking moet omarmen.

Ons onderzoek begint met een concrete klinische behoefte – het ondersteunen van polyfarmacie-reviews door middel van een regel-gebaseerd klinisch beslissingsondersteunend systeem genaamd de STRIP Assistant (STRIPA). Om het gebruik ervan in een multinationale klinische trial te ondersteunen, ontwerpen we in Hoofdstuk 2 GDPR-conforme systemen die tegelijkertijd databeveiliging en toegankelijkheid waarborgen. In eerste instantie had STRIPA een handmatig data-invoerproces dat tijdrovend en foutgevoelig is. Om dit probleem aan te pakken, introduceert Hoofdstuk 3 een lichtgewicht, API-gebaseerde architectuur voor klinische NLP-systemen. Dit is een architectuur die verschillende externe NLP-diensten orkestreert om flexibele, kosteneffectieve oplossingen te creëren die de beperkingen van monolithische klinische NLP-systemen overwinnen. We ontwikkelen een prototype dat het potentieel aantoont om handmatige data-invoer voor HIS zoals STRIPA te verminderen. Het proefschrift vervolgt met het ontwerpen van een geautomatiseerde NLP-pipeline voor het extraheren van gestructureerde bijwerkingen van geneesmiddelen uit Europese productlabels in Hoofdstuk 4. Dit laat zien hoe gespecialiseerde NLP-technieken gestructureerde kennis kunnen ontsluiten uit regelgevingsdocumenten met hoge precisie. Hoofdstuk 5 presenteert een cloud-gebaseerd big data framework voor grootschalige biomedische literatuur mining, dat een schaalbare oplossing biedt voor het verwerken van miljoenen onderzoeksartikelen met behoud van kosteneffectiviteit. De behoefte aan een dergelijk framework wordt duidelijk aangezien de omvang van biomedische publicaties nog altijd gestaag sneller groeit.

Het tweede deel van het onderzoek richt zich op het verkennen van hoe open source software de ontwikkeling van HIS kan verbeteren. In Hoofdstuk 6 wordt een reproduceerbare methodologie ontwikkeld voor systematische studies van open source klinische software, die inzicht biedt in het landschap van beschikbare tools en hun kenmerken. Verder ontwikkelen we een webplatform, genaamd LOCATE, dat de kloof overbrugt tussen open source klinische software en wetenschappelijke literatuur en ontwikkelaars in staat stelt beter geïnformeerde keuzes te maken over het hergebruik van software. Met uitzondering van het eerste klinische beslissingsondersteunende systeem, zijn alle broncodes van dit onderzoek publiekelijk toegankelijk

op GitHub met als doel aan open source software bij te dragen.