

Development and Validation of a Web-Crawler-Based Medical Records Information Aggregation Tool

Hongjia Liu¹, Yuliang Huang¹, Yichen Pu¹, Hao Wu¹, Yibao Zhang^{*1}

¹Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital & Institute, Beijing, China

*Corresponding author: Yibao Zhang; mail: zhangyibao@pku.edu.cn

INTRODUCTION

The medical applications of artificial intelligence have been strongly dependent on the data aggregation and organization. Conventional manual collection is labor-intensive and time-consuming. However, the lack of uniform standards for the medical record information stored in the Hospital Information System has brought challenges to the automatic mining and analysis of big data applications.

AIM

This work aims to develop and validate a Web-Crawler-based medical records information aggregation tool for effective data mining from existing electronic information systems.

METHOD

Based on Selenium framework and Python programming language, a Web-Crawler-based medical records information aggregation tool was designed, which was validated under two illustrative scenarios:

- To identify radiation pneumonitis (RP) cases from Hospital Information System (HIS), as an application example of quick data search
- To summarize an organized table combining desired data from various examination reports, to test the application of facilitating clinical workflow. Automated and manual methods were compared in terms of efficiency and accuracy.

RESULTS

- For the first scenario, automated method identified 110 RP cases out of 3541 patients in about 54 seconds per patient based on a Raspberry Pi 4B, without any human interference. Manual methods identified the same group of RP cases but took about 90 seconds per patient.
- For the other scenario, automated and manual methods needed about 10 or 75 seconds respectively for each patient.
- The automated tool showed superior efficiency and accuracy than manual method. Automated method also avoided typos that were frequently observed in manual report filling.
- It took longer to confirm a non-RP case because more data need to be excluded to avoid false-negative, suggesting even greater advantage of automated method in searching small-probability events, especially from huge patient volume.

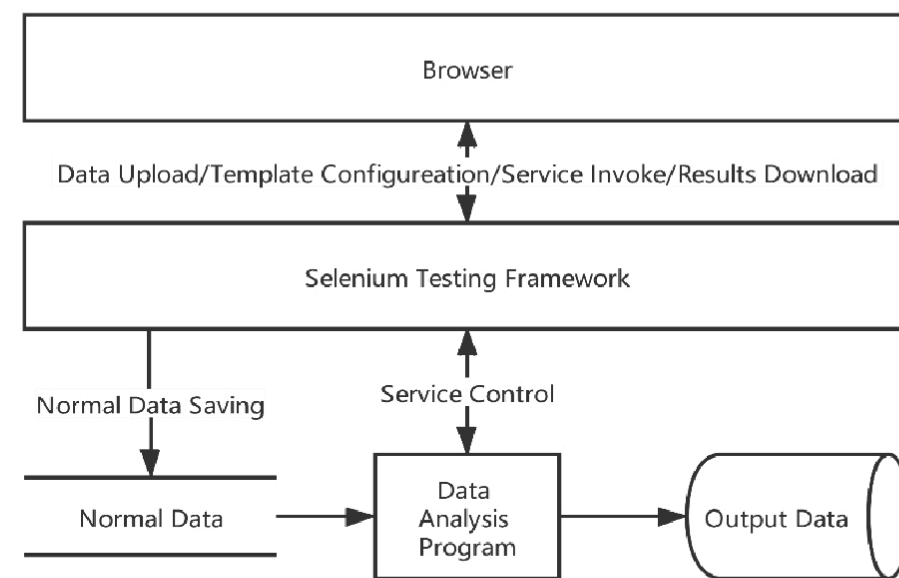


Figure 1: Program architecture of aggregation tool

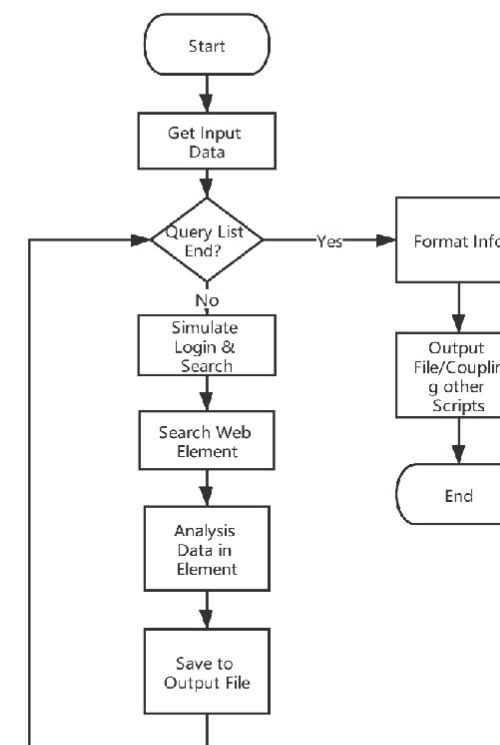


Figure 2: Workflow chart of aggregation tool

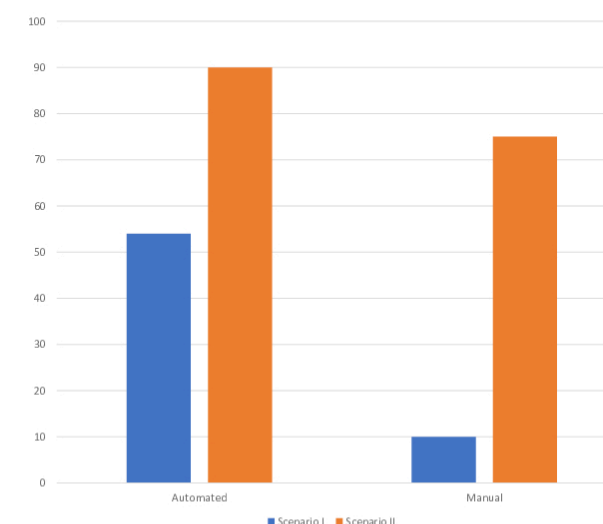


Figure 3: Test results for specific work

CONCLUSIONS

A Web-Crawler-based medical records information aggregation tool has been successfully developed. The superior efficiency and accuracy of auto-aggregation has been validated based on specific clinical scenarios. With the advantage of cross-platform and easy-to-extend, this application could improve radiologists' and physicists' productivity in their clinical and research practice.

ACKNOWLEDGEMENTS

The authors thank all physicists at Beijing Cancer Hospital for their assistance.

This work was jointly supported by Capital's Funds for Health Improvement and Research [2018-4-1027]; National Natural Science Foundation of China [11505012, 11905150, 81672969]; Fundamental Research Funds for the Central Universities/Peking University Clinical Medicine Plus X - Young Scholars Project [PKU2020LCXQ019]; National Key R&D Program of China [2019YFF01014405]; Ministry of Education Science and Technology Development Center [2018A01019]; Sichuan Science and Technology Program [2018HH0099]; Beijing Municipal Administration of Hospitals Incubating Program [PX2019042].

REFERENCES

- Huang, Y., Yue, H., Wang, M., Li, S., Zhang, J., Liu, Z., & Zhang, Y. (2018). Fully automated searching for the optimal VMAT jaw settings based on Eclipse Scripting Application Programming Interface (ESAPI) and RapidPlan knowledge-based planning. *Journal of applied clinical medical physics*, 19(3), 177-182.
- Xie, C., Yang, P., & Yang, Y. (2018). Open knowledge accessing method in IoT-based hospital information system for medical record enrichment. *IEEE Access*, 6, 15202-15211.

CONTACT INFORMATION

Mail: zhangyibao@pku.edu.cn

Address: Peking University Cancer Hospital, Fucheng Road 52, Haidian District, Beijing, China