



INTISARI

QUR'AN LINK : EKSTRAKSI DAN VISUALISASI SITASI AL-QUR'AN PADA ARTIKEL ISLAM DI WEB

Oleh

Muhammad Zha'farudin Pudya Wardana

14/365977/PA/16182

Internet telah dijadikan media dakwah untuk pencarian informasi tentang Islam oleh ummat muslim. Mesin pencari yang ada saat ini belum mengakomodasi sumber ayat Al-Qur'an pada perangkian pencarian. Ekstraksi sitasi Al-Qur'an pada artikel di web mungkin dapat mengakomodasi sumber Al-Qur'an pada sistem pemeringkatan mesin pencari. Penelitian ini mengusulkan sebuah kerangka kerja untuk ekstraksi sekaligus visualisasi sitasi Al-Qur'an dari artikel yang dipublikasikan di web.

Sistem ekstraksi yang diusulkan pada penelitian ini menggunakan aturan ekspresi reguler yang telah ditentukan untuk mengekstraksi sitasi. Selanjutnya algoritme koreksi ejaan diterapkan untuk pengenalan entitas nama surat. Pada penelitian ini juga dirancang mesin pencari untuk membuat artikel islam dapat dicari. Penelitian ini menyediakan metode untuk mereduksi kompleksitas visualisasi data besar hasil ekstraksi dengan memvisualisasikan setiap artikel dari hasil pencarian bersama dengan sitasi Al-Qur'an hasil dari ekstraksi setiap dokumen.

Hasil ekstraksi sitasi berupa graf dapat digunakan untuk mempengaruhi nilai relevansi dalam proses pemeringkatan hasil pencarian. Sistem mampu mengekstraksi 31.062 sitasi dari 9.022 artikel yang dikumpulkan dari empat website islam yang berbeda. Sistem ekstraksi menunjukkan hasil kinerja dengan Presisi 0,996, *Recall* 0,991, *F-measure* 0,993, dan *Fallout* 0,075. Visualisasi interaktif pada penelitian ini juga dapat digunakan untuk analisis dan eksplorasi pengetahuan Al-Qur'an pada penelitian selanjutnya.

Kata-kata kunci : ekstraksi informasi, ekstraksi sitasi, visualisasi big data, Al-Qur'an, ekspresi reguler, mesin pencari



ABSTRACT

QUR'AN LINK : EXTRACTION AND VISUALIZATION OF QUR'AN CITATIONS IN ISLAMIC ARTICLES ON THE WEB

By

Muhammad Zha'farudin Pudya Wardana

14/365977/PA/16182

Internet has been used for searching of Islamic knowledge by moslem community. Currently, most search engine has not accommodated source of Al-Qur'an for scoring in their ranking system yet. Extraction of Qur'an citations from articles on the web may be able to accommodate source of Al-Qur'an to relevance score in search engine's ranking system. This research proposes a framework to extract and visualize the citation of Al-Qur'an from articles published on the web.

The extraction system proposed in this research uses the given rules of regular expression to extract the citation. After extraction, we use spelling correction algorithm in order to recognize named entity of surat's name of Al-Qur'an. A search engine has been designed in order to make the article searchable. We provide a method to reduce the complexity of big data visualization in extraction's results by visualizing each article from search engine's results together with citations of Al-Qur'an which extracted from individual document.

Result of extraction represented as graph of article citing verses from Al-Qur'an can be used for formulating the relevance score in ranking process of search engine too. System is able to extract 31,062 citations from 9,022 articles collected using our crawler from four different islamic website. Performance of extraction in this research shows Precision 0.996, Recall 0.991, F-measure 0.993, and Fallout 0.075. Interactive visualization proposed in this research can also be used for analysis and exploration of Quranic knowledge in future study.

Keywords : information extraction, big data visualization, regular expression, named entity recognition, search engine, Al-Qur'an