

SastraKB : Knowledge Base Sastra Indonesia = SastraKB : Knowledge Base of Indonesian Literature

Faraya Agatha Putri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20511037&lokasi=lokal>

Abstrak

Karya sastra merupakan hal yang perlu dilestarikan, karena melestarikan karya sastra juga berarti melestarikan bahasa. Upaya pelestarian dapat dilakukan dengan berbagai cara, salah satunya dengan memanfaatkan teknologi. Implementasi upaya yang dapat dilakukan dengan memanfaatkan teknologi adalah dengan melakukan ekstraksi entitas karya sastra secara otomatis. Dari data ekstraksi tersebut dapat dibangun knowledge base agar informasi menjadi lebih terstruktur dan dapat diatur dengan mudah. Penelitian ini menggunakan sumber data dari 435 halaman sastrawan Indonesia pada Wikipedia berbahasa Indonesia. Terdapat dua proses ekstraksi pada penelitian ini, yaitu ekstraksi daftar dan ekstraksi tabel. Pada akhir penelitian ini, diperoleh 4953 entitas karya sastra yang terpetakan ke dalam 14 kategori karya sastra. Kualitas hasil ekstraksi pada penelitian ini diukur dengan nilai precision dan recall. Nilai precision dan recall didapatkan dari hasil perbandingan data hasil ekstraksi dengan data golden result yang merupakan data yang disusun secara manual dari halaman-halaman sastrawan Indonesia. Nilai precision dan recall pada penelitian ini adalah 0.608 untuk precision dan 0.571 untuk recall.

Literature work needs to be preserved because it also means preserving a language. There are many preserving methods, one of them is using technology. The implementation of using technology as a preserving method is by automatically extracting the literature work entities. From that data extraction, a knowledge base can be built to make the information more structured and easy to manage. This research used 435 Wikipedia pages about Indonesian litterateur as a source of data extraction. Two extraction processes have been implemented, which are list extraction and table extraction. At the end of this research, 4953 literature work entities that mapped into 14 literature work categories were obtained. The quality of the data extraction results in this research was measured by precision and recall value. The precision and recall value was obtained from comparing the data extraction result with the golden result which is data that was organized manually from Wikipedia pages about Indonesian litterateur. The precision and recall value of this research are 0.608 for precision value and 0.571 for recall value.