

Pengenalan entitas bernama pada dokumen wikipedia dan berita bahasa Indonesia dengan pendekatan conditional random field = Named-entity recognition on indonesian wikipedia and news document using conditional random field approach.

Alif Ahsanil Satria, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20527558&lokasi=lokal>

Abstrak

Pengenalan entitas bernama (named-entity recognition atau NER) adalah salah satu topik riset di bidang pemrosesan bahasa alami (natural language processing atau NLP). Pengenalan entitas bernama merupakan langkah awal mengubah unstructured text menjadi structured text. Pengenalan entitas bernama berguna untuk mengerjakan NLP task yang lebih high-level seperti ekstraksi informasi (information extraction atau IE), Question Answering (QA), dan lain-lain. Penelitian ini memanfaatkan data berita dan wikipedia masing-masing sebanyak 200 dokumen yang digunakan untuk proses pengujian dan pelatihan. Penelitian ini mencoba mengeksplorasi entitas bernama baru yang tidak sebatas Person, Location, dan Organization. Named entity baru tersebut adalah Event, Product, Nationalities Or Religious or Political groups (NORP), Art, Time, Language, NonHuman or Fictional Character (NHFC), dan Miscellaneous. Jadi, penelitian ini menggunakan 11 entitas bernama. Dalam penelitian ini, permasalahan tersebut dipandang sebagai sequence labelling. Penelitian ini mengusulkan penggunaan model conditional random field sebagai solusi permasalahan ini. Penelitian ini mengusulkan penggunaan fitur tambahan seperti kata sebelum, kata sesudah, kondisi huruf kapital di awal kata, dan lain-lain, serta word embedding. Penelitian ini menghasilkan performa dengan nilai F-measure terbaik sebesar 67.96% untuk data berita dan 67.09% untuk data wikipedia.

.....