Abstrak

Ekstraksi informasi adalah suatu proses untuk mencari data yang spesifik dan penting dari sebuah dokumen yang tidak terstruktur (*natural language document*) menjadi dokumen yang terstruktur. Ekstraksi informasi ini merupakan solusi yang dapat mengubah *job posting* dari dokumen yang tidak terstruktur ataupun semi-terstruktur menjadi dokumen yang terstruktur. Konsepnya adalah dengan cara meng-ekstrak informasi *job posting* berdasarkan beberapa label *field*, seperti company, title atau position, city, salary, dll. Metode yang digunakan adalah metode Boosted Wrapper Induction yang dapat menangani *free text* dengan menghasilkan *rule-rule* yang dapat mengenali keberadaan *field* yang ingin diekstrak. Evaluasi performansi sistem menggunakan *precision, recall* dan *F-Measure.* Parameter yang mempengaruhi performansi sistem adalah jumlah iterasi *boosting* yang akan mempengaruhi jumlah rule detector yang dihasilkan, nilai *lookahead* yang menyatakan jumlah token yang akan diperhitungkan sebagai kandidat prefix dan suffix, serta pemakaian *wildcards.* Dari hasil yang diperoleh dapat disimpulkan keberadaan *wildcard* sangat berpengaruh untuk meningkatkan performansi sistem. Dan iterasi *boosting* juga cenderung meningkatkan performansi akan tetapi sangat bergantung pada jumlah variasi rule yang dihasilkan. Dan untuk parameter lookahead, performansi sistem bergantung pada jumlah prefix atau suffix dari detector yang selalu berpasangan.

Kata kunci: *Information Extraction, Wrapper, Wrapper Induction, AdaBoost, Boosted Wrapper Induction*