

Abstrak

Seiring cepatnya penambahan data pada internet, internet kini dimanfaatkan menjadi sumber data bagi berbagai keperluan. *Automatic Cataloging* (ACat) adalah sistem IE yang digunakan untuk otomasi proses pengkatalogan buku dengan input data dari internet yang berupa halaman *offline* html.

Dengan menggunakan *rule* yang dibentuk dari *learning corpus* menggunakan *natural language tools*, informasi tentang buku dapat diambil dari suatu halaman html. Nilai *precision* dan *recall* dari penggunaan *rule* hasil learning dipengaruhi oleh nilai maksimum dan minimum *slot filler length* serta penghilangan *uncoupled* tag.

Metode yang digunakan adalah *Supervised Learning of Extraction Patterns and Rules* di mana *learning corpus* perlu dibuat sesuai dengan *domain* yang diharapkan, dalam kasus ini merupakan domain buku. Sistem IE dibuat bekerja sebagai *tagger* yang berfungsi memberikan tag pada informasi relevant yang akan diekstrak.

Kata kunci: *tagger, information extraction, supervised learning, rule, natural language, POS tagger.*