

Abstract

As fast as the adding of information in internet, nowadays internet is used for data resource for many purposes. Automatic Cataloging (ACat) is IE system for automaton process in book cataloging using html offline page as the input data.

Using the rule that made from learning corpus using natural language tools, book information can be found from a html page. Precision and recall values in the tagging process using the rule is depend on the value of minimum and maximum slot filler length and the uncoupled tag removal.

Supervised Learning of Extraction Patterns and Rules is the method in used, where learning corpus is needed to be made based on the domain, in this case in book domain. IE system is made to be a tagger that for tagging the relevant information that will be extracted.

Keywords: tagger, information extraction, supervised learning, rule, natural language, POS tagger.