Citation parsing made easy

code4lib 2008 / lightning talks

Erik Hetzner
California Digital Library
erik.hetzner@ucop.edu
title: Anton Pannekeoeok and the Socialism of Worker’s Self Emancipation, 1873-1960

author: Gerber, John P.

copyright: Springer

date: 1990
Harder than it seems
Can’t use standard parsing techniques
Rules-based systems require tuning for different styles
Statistical parsing methods can help
Hidden Markov models
Train on your data
> 75% field level $F_1$
Submitted to JCDL 2008
Erik Hetzner
California Digital Library
erik.hetzner@ucop.edu