Författare till examensarbetet

Olof Falk
S171279@student hb.se
198611147839
Publicera uppsatsen i DiVA?: JA

Examensarbetet

Akademi *: Akademin för bibliotek, information, pedagogik och IT
Huvudområde *: Biblioteks- och informationsvetenskap
Examen *: Master
Examensarbetsnummer *: X
Handledare: Mikael Gunnarsson
Examinator: Johan Eklund
Språk: Engelska

Engelsk titel (kopiera från uppsats) *: Automated fiction classification - an explorative study of fiction classification using machine-learning techniques
Engelska nyckelord (kopiera från uppsats) *: Skönlitteratur, klassifikation, genre, särdrag, ämne, stil, maskininlärning.
Engelsk sammanfattning (kopiera från uppsats) *: This thesis aims to explore the possibilities and components of employing automated text classification techniques to classify collections of narrative fiction by genre, and also, what linguistic features are prominent in distinguishing genres of fiction. The historical traditions and current practices and theories in the field of fiction classification are outlined, along with central concepts of classification and genre theory. Linguistic features are also introduced, and hypothesized to carry capabilities of distinguishing genres of fiction. The thesis also reviews the foundations and current state of automated text classification, and reasons on what constitutes topical and stylistic features in relation to fiction. Knowledge gaps are identified between automated text classification and traditional fiction classification, and also, concerning the potentially genre-distinguishing qualities of topical and stylistic features. The main experiment, around which the thesis is centered, is divided into two parts. The first part employs and evaluates kNN and SVM classifiers on a collection of fiction documents across four genres of fiction. In the second part, some feature selection methods are employed for inspection of distinguishing features across the collection. Findings suggest a potential of using automated techniques to classify fiction, and also illustrates feature patterns that are argued to distinguish each of the four different genres of fiction. Some suggestions for further research are also proposed.