

# Using Knowledge from Wikipedia to Improve Document Classification

Martin Boroš

## *(Abstract)*

The intention of the research is to understand how document classification can be improved using Wikipedia. A variety of literatures have been reviewed to get a better understanding of the topic, and to study details of the functions of Wikipedia and the effect it can have on document classification. Spoken and written communication further facilitated our technological advancement. Starting with the early Greeks, the encyclopedia emerged as a modern form of culture and knowledge transmission and has served as an important tool for the collecting and archiving of knowledge allowing future generations the opportunity to build on prior developments rather than continual rediscovery. The popularity of Wikipedia has also grown and currently (as of October, 2011) ranks fifth in overall global web traffic (“Alexa Top 500 Global Sites,” n.d.). Web users looking for information on any topic will likely come across a Wikipedia article fairly quickly. Keeping this in mind we fail to reject H1 = Document classification can be improved using Wikipedia, H3 = Algorithm models are the factors that contribute to document classification, and H6 = the current position document classification is good.