Conversation Level Constraints on Pedophile Detection in Chat Rooms  
Notebook for PAN at CLEF 2012

Claudia Peersman, Frederik Vaassen, Vincent Van Asch, Walter Daelemans

Computational Linguistics and Psycholinguistics Research Center, Antwerp University

{claudia.peersman, frederik.vaassen, vincent.vanasch, walter.daelemans}@ua.ac.be

In this paper we present a new approach for detecting online pedophiles in chat rooms that combines the results of predictions on the level of the individual post, the level of the user and the level of the entire conversation, and describe the results of this three-stage system in the PAN 2012 competition. Also, we describe a resampling and a filtering strategy to circumvent issues regarding the unbalanced dataset. Finally, we describe the creation of a dictionary of words and expressions relating to predators’ grooming stages, which we used to identify which posts in the predators’ conversations were most distinctive for their grooming behavior.