Overview of the   
International Sexual Predator Identification   
Competition at PAN-2012  
Extended Abstract

Giacomo Inches and Fabio Crestani

IR Group, Faculty of Informatics  
University of Lugano, Switzerland

{giacomo.inches, fabio.crestani}@usi.ch

“Chat messages” or “online/IRC conversations” are part of almost everybody's everyday life with services like Skype, Yahoo Messenger, MSN Messenger, ICQ but also IRC networks like Freenode or Quakenet. Although these services facilitate the establishment of new connections between persons or reinforce existing ones, they also allow for misbehaviors or cybercriminal acts. The Sexual Predator Identification competition ran for the first time in 2012 within PAN Lab and aimed at providing to researches a common framework to test methods for identifying such misbehaviors or cybercriminal activities.

We released a challenging collection where different types of conversations were mixed in a realistic scenario: suspicious conversations were combined with false positives and generic ones and we asked the participant to the laboratory to i) identify the “sexual predators” (within all the users) in the different conversations and ii) identify the part (the lines) of the predator conversations which are the most distinctive of the predator bad behavior.

The 16 teams participating in the subtask came from different part of the world and consisted of experts working for years in the field as well ad newcomers. The proposed methods made possible the identification of common pattern for predator identification (e.g. no preprocessing of the conversations, turn-taking/behavioral analysis, blacklisting of predator terms) as well as possible extension to existing systems (e.g. victim-predator distinction, pre-filtering of not relevant conversation).