A Focused Crawler for Mining Hate and Extremism Promoting Videos on YouTube
Swati Agarwal, Ashish Sureka
Indraprastha Institute of Information Technology, Delhi (IIIT-D) New Delhi, India
swatia@iiitd.ac.in, ashish@iiitd.ac.in

Research Motivation and Aim

- More than 1 billion unique users visit each month
- Over 6 billion hours of video are watched each month
- 100 hours of video are uploaded to YouTube every minute
- High reachability, low publication barriers and anonymity
- Platform for many hate and extremist groups to share information and promote their ideologies [online radicalization]

To investigate the application of a best-first search focused crawler based approach for retrieving hate and extremism promoting videos

Solution Approach

1. Seed User
2. Width of Graph
3. Size of Graph
4. Threshold
5. N-Gram
6. Positive Class Channels

Related Work

A. Sureka et. al.; 2010 - Locating hate promoting videos, users and their groups sharing a common agenda
M. GoodWin; 2013 - Analysis of various counter-jihad, Islam and Muslim communities on web 2.0.
E. Erez; 2011 - Quantitative and qualitative assessments of the content of communications on forums.

Experimental Results

F1-SCORE : 0.81  PREDICTED
ACCURACY : 0.69

<table>
<thead>
<tr>
<th>ACTUAL</th>
<th>RELEVANT</th>
<th>IRRELEVANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREDICTED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RELEVANT</td>
<td>991</td>
<td>295</td>
</tr>
<tr>
<td>IRRELEVANT</td>
<td>55</td>
<td>29</td>
</tr>
</tbody>
</table>