Rationale

- Young people spend a lot of time online
- Online environments are not without risks
- Unfeasible for stakeholders to keep track of potentially harmful situations
- Protection: detect and curate threats

Manual monitoring infeasible because of information overload

Urgent demand for automatic monitoring

AMiCA Goals

- Detection and filtering of unwanted and illegal online content
- Cross-media analysis (text, image, video)
- Context and profile analysis
- Aggregated data => quantitative information on risk incidence
- Embedded monitoring and privacy by design

Text analytics

- Translate noisy language into its canonical form
- Approaches: spelling correction, machine translation, G2P2G, classification, ...

Profiling

- Automatic extraction of information about the author of a text: identity, gender, age, educational level, personality, etc.
- Challenges: single-out feature types and discriminative methods that are able to efficiently deal with large author set sizes, small data sizes, and a variety of topics and genres

Frame-based detection

- Script: temporal sequence of event frames with different roles (participants, action, location, time, ...)
- Script detection through an ensemble of classifiers trained on the detection of participant features and their interactions

Deep text analytics

- Text analysis pipeline that automatically analyzes text up to the level of discourse
- Modules that deal with non-propositional aspects of meaning (e.g. modality, negation), necessary for filtering and mining social media

Transgressive sexual behavior: script with series of event frames in which participants (minor, adult) experience a number of “grooming” steps

Cyberbullying: script with series of event frames in which participants (bully, bystander, victim) experience a number of interactions

Project overview

Development

AMiCA kernel

Dataflow management

Context mining & analysis

Cross-media analysis

Text Analytics

Image Processing & Audio Mining

Grounding

Issues and risks of social media use

Validation: 3 use cases

- Automutilation & suicidal behavior
- Transgressive sexual behavior
- Cyberbullying

Normalisation

<table>
<thead>
<tr>
<th>Original</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>hey sarahe tis al lang gelde dak hier ng op ben geweest ma hey bffl eh ;)</td>
<td>hey sarahe het is al lang geleden dat ik hier nog op ben geweest maar hey best friends for life he ;)</td>
</tr>
</tbody>
</table>

Text analysis pipeline that automatically analyzes text up to the level of discourse

Modules that deal with non-propositional aspects of meaning (e.g. modality, negation), necessary for filtering and mining social media

Transgressive sexual behavior: script with series of event frames in which participants (minor, adult) experience a number of “grooming” steps

Cyberbullying: script with series of event frames in which participants (bully, bystander, victim) experience a number of interactions

Frame-based detection

- Script: temporal sequence of event frames with different roles (participants, action, location, time, ...)
- Script detection through an ensemble of classifiers trained on the detection of participant features and their interactions