FACULTY OF ARTS
AND PHILOSOPHY

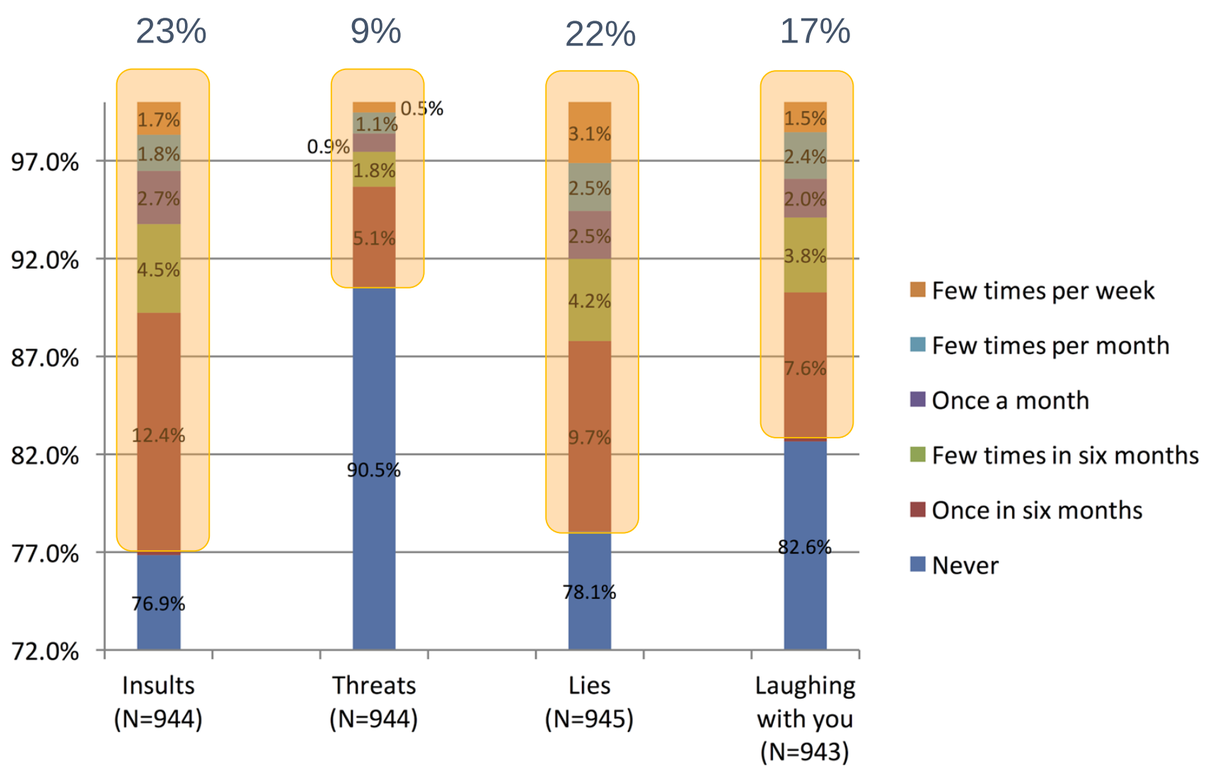
DEPARTMENT OF TRANSLATION, INTERPRETING AND COMMUNICATION
LT3 LANGUAGE AND TRANSLATION TECHNOLOGY TEAM

AUTOMATISCHE TEKSTANALYSE VOOR CYBERPESTDETECTIE

Veronique Hoste, 3 mei 2017

GHENT
UNIVERSITY

language and
translation
technology
team



Grafiek 1: Frequentie van verschillende soorten cyberpesten op sociale media (Van Royen et al., 2014).

- Cyberpesten = belangrijk probleem

- **Gevolgen** van cyberpesten:

  - Woede, verdriet, angst, laag zelfbeeld, schoolproblemen,...

    (Price & Dalgleish, 2010)

  - Automutilatie en zelfdoding

    (Hinduja & Patchin, 2010)

GHENT UNIVERSITY

- Initiatieven voor online veiligheid van jongeren bestaan.

MAAR:

- Ongewenste berichten vaak (lange tijd) online.
- Te veel data voor moderatoren van sociale netwerken.
- Nood aan intelligente systemen die cyberpestberichten **automatisch** opsporen.

GHENT UNIVERSITY

SYSTEMEN VOOR AUTOMATISCHE DETECTIE NODIG

GHENT UNIVERSITY

NOOD AAN DATA, VEEL DATA

- Om lerende systemen te trainen zijn voldoende DATA nodig
- Dataverzameling voor Nederlands en Engels
  - Data relevante sociale media

    MAAR: weinig / private data

    ⬇

  - Mediacampagne voor donatie voorbeelden van cyberpesten

    MAAR: gevoelige data!

    ⬇

  - Cyberpestsimulaties

GHENT UNIVERSITY

NOOD AAN DATA, VEEL DATA

- Om lerende systemen te trainen zijn voldoende DATA nodig
- Dataverzameling voor Nederlands en Engels
  - Data relevante sociale media

    MAAR: weinig / private data

    ⬇

  - **Mediacampagne voor donatie voorbeelden van cyberpesten**

    MAAR: gevoelige data!

    ⬇

  - **Cyberpestsimulaties**

GHENT UNIVERSITY

MEDIACAMPAGNE

HET NIEUWSBLAD WOENSDAG 27 NOVEMBER 2013 11

VLAAMSE ONDERZOEKERS ONTWIKKELEN SYSTEEM TEGEN CYBERPESTEN

'Stuur ons uw haatmail'

Waarom draagt gij altij make up aan, Verberg u lelijkheid nie helpt tog nie

gij zijt egt ne mega nerd hoe gij durft buitenkomen??? ik snap het ni

dikzak ik ga u morge kapot slage en u zus gooi ik derbove ik maak u kappot

Enkele haatcommentaren op Ask.fm. De onderzoekers zullen het taalgebruik in pestberichten analyseren. Foto: rr

Onderzoekers van de Vlaamse universiteiten roepen slachtoffers van cyberpesten op om de pestberichten die ze ontvangen *en masse* aan hen te bezorgen. Met die gegevens willen ze een programma ontwikkelen dat pestgedrag op sociale media detecteert en hulporganisaties alarmeert.

JESSE VAN REGENMORTEL

Maar liefst één Vlaamse jongere op de tien kreeg recentelijk pestberichten via internet of gsm, zeggen onderzoekers van de universiteiten van Antwerpen, Leuven en Gent. Cyberpesten is een groot en groeiend probleem. Maar de aanpak van dat probleem staat nog niet op punt. Dus slaan ze de handen in elkaar voor een groot wetenschappelijk project.

Pestfilter

Het doel van AMiCA (Automatic Monitoring for Cyberspace Applications) is een systeem te ontwikkelen dat automatisch cyberpestgedrag herkent op sociaalnetwerksites. 'Omdat te kunnen doen, hebben we enorm veel gegevens nodig', zegt Véronique Hoste van UGent. 'E-mails, sms'en, Facebook- en chatberichten ...'

Vandaar de opvallende oproep die zij en haar collega's gisteren lanceerden. 'Of je nu het slachtoffer bent van cyberpesten of een getuige, leerkracht of ouder, bezorg ons de berichten die je beschouwt als cyberpesten.' De onderzoekers benadrukken dat die volledig anoniem behandeld worden.

Het systeem dat ze met die gegevens zullen ontwikkelen, moet weldra elk bericht scannen dat binnenkomt op Twitter of eender welke andere sociaalnetwerksite. Daarbij moet het programma uitmaken of het al dan niet om een pestbericht gaat. De moderatoren van die sites zullen de problematische berichten vervolgens signaleren aan organisaties die pesten bestrijden.

'Vergelijk het met een meldknop waarmee je nu al her en der kwetsende berichten kunt melden', zegt Hoste. 'Wij willen dat automatiseren. Zulke filters bestaan nu al om bijvoorbeeld racisme op te sporen.'

Uniek project

Door zoveel mogelijk pestberichten te verzamelen, willen de onderzoekers zicht krijgen op welke woorden en zinsneden zoal gebruikt worden. 'Want met woorden alleen komen we er niet. Het woord *zelfmoord* kan gebruikt worden in een suïcidaal bericht, maar ook in een persbericht. Je moet dus kijken naar het bericht in zijn totaliteit en naar de berichten die voorheen en nadien geschreven zijn.'

VÉRONIQUE HOSTE Universiteit Gent

Er bestaan nu al knoppen waarmee je kwetsende berichten kunt melden. Wij willen dat systeem automatiseren

Het project van de Vlaamse universiteiten is door zijn omvang nu al uniek. Maar voorlopig focussen ze zich enkel op het Nederlands. 'Al kan het systeem in een later stadium ook voor het Engels gebruikt worden. Maar dan zullen we opnieuw massaal gegevens moeten verzamelen.'

INFO Bezorg je cyberpestberichten via data@amicaproject.be

AMiCA

Wat is AMiCA?

Met het **AMiCA**-project (Automatic Monitoring for Cyberspace Applications) willen we mogelijk bedreigende situaties op sociale netwerken automatisch herkennen door middel van tekst- en beeldanalyse om zo de online veiligheid van kinderen te waarborgen.

In dit project wordt technologie ontwikkeld om deze situaties te herkennen en wordt nagegaan hoe de ontwikkelde technologie het best kan worden ingezet om slachtoffers van cyberpesten te helpen.

Meer informatie over dit project vindt u op de website:
http://www.amicaproject.be/

Gezocht: cyberpestberichten

Beste,

In het kader van ons onderzoek naar cyberpesten aan de universiteiten van Antwerpen, Leuven en Gent hebben we een oproep gelanceerd om zoveel mogelijk cyberpestberichten te verzamelen. Die berichten willen we gebruiken als voorbeeldmateriaal om een systeem te ontwikkelen dat gevaarlijke situaties op sociale netwerken automatisch kan herkennen.

We willen onze oproep in het bijzonder richten aan jongeren die met cyberpesten geconfronteerd worden. Graag vragen we aan u als hulpverlener/leerlingbegeleider om onze vraag door te geven aan de jongeren waarmee u in contact komt. Concreet willen we u vragen dat u kort ons onderzoek schetst en vraagt of ze bereid zijn te helpen. Daarna kunt u de jongeren een brief meegeven waarin onze doelstellingen staan en waarin ze praktische richtlijnen vinden om de gezochte berichten op te sturen.

Onze dataverzameling loopt vanaf heden tot en met juni 2014, gedurende dit half jaar zijn alle initiatieven om berichten te verzamelen welkom. Elke bijdrage aan ons onderzoek is immers waardevol en kan ons helpen om een beter systeem te ontwikkelen waar we jongeren in de toekomst mogelijk mee kunnen helpen.

Wij bedanken u alvast voor uw hulp!
Het AMiCA-onderzoeksteam

Voor meer informatie over deze oproep of ons project kan u steeds terecht bij:

Cynthia Van Hee	Universiteit Gent	cynthia.vanhee@ugent.be
Ben Verhoeven	Universiteit Antwerpen	ben.verhoeven@uantwerpen.be

UNIVERSITEIT GENT · Universiteit Antwerpen · iMinds

---

radio 2 altijd dicht bij jou

De Universiteit van Gent zoekt cyberpestberichten

Tags: *cyberpesten*, *gent*, *universiteit*

Een aantal onderzoeksgroepen van de **universiteiten van Antwerpen, Gent en Leuven** zijn een systeem aan het ontwikkelen dat automatisch cyberpestgedrag, en andere gevaarlijke situaties op het internet, kan herkennen op sociale netwerksites. Op die manier proberen zij een veilige internetomgeving te creëeren voor jongeren.

De wetenschappers die hier mee bezig zijn, lanceren nu een oproep aan slachtoffers, en getuigen van cyberpesten, ouders en leerkrachten, om zoveel mogelijk berichten door te geven die zij als cyberpesten beschouwen. Dat kunnen **e-mails zijn, sms'en, chatgesprekken, of berichten van Facebook, netlog, Twitter en Ask.fm**. Alles wordt anoniem behandeld, en wordt alleen gebruikt voor intern onderzoek. Dus maak je geen zorgen, er wordt niks doorgegeven aan derden !

Dus als je dit wil, en kan doen, surf dan snel naar deze website: ***www.amicaproject.be*** of stuur meteen een e-mail: **data@amicaproject.be**. Uw e-mail wordt vertrouwelijk behandeld door het AMiCA-onderzoeksteam.

26 nov 2013 - 16:18

Nieuw systeem spoort online pestgedrag automatisch op

SAM FEYS
BRUSSEL

Belgische onderzoekers willen cyberpesten tegengaan door pestgedrag op Facebook en Twitter automatisch op te sporen. Het systeem moet op termijn zelfs ingrijpen bij zelfmoordgedachten.

Cyberpesten is vandaag nog erg moeilijk op te sporen, zeker omdat slachtoffers er zelden voor uitkomen. Verschillende onderzoekers van de universiteiten van Gent, Leuven en Antwerpen slaan daarom de handen in elkaar voor AMiCA, een systeem dat automatisch pestgedrag op het internet moet herkennen. Zo kunnen zowel sitebeheerders als politie ingrijpen voor het te laat is.

"We willen zowel tekst als beeld op sociale media analyseren om zo volautomatisch alarmsignalen te laten afgaan bij verontrustende situaties", legt Véronique Hoste van UGent het project uit. "Zo kunnen we zowel gevallen van online pestgedrag en pedofilie opsporen, maar ook ingrijpen als jongeren met suïcidale gedachten kampen."

Het onderzoeksteam is daarom op zoek naar slachtoffers van cyberpesters die ooit beledigende sms'jes, e-mails of berichten kregen, ook via socialenetwerksites als Twitter, Facebook en Ask.fm. "Die zullen we uitgebreid analyseren. Op basis daarvan kunnen we accurate computermodellen bouwen die kunnen afleiden wanneer het om cyberpesten gaat. Hoe meer bruikbare voorbeelden we binnenkrijgen, hoe beter we dat model kunnen voeden."

De bedoeling is om uiteindelijk elk nieuw bericht op Facebook of Twitter te toetsen aan dat model. "Elk bericht zal een label meekrijgen", verduidelijkt Hoste. "Als bij meerdere berichten van een gebruiker rode lampjes gaan branden, zal een moderator of sitebeheerder automatisch op de hoogte worden gebracht. Die kan dan de ernst van de situatie inschatten en eventueel de politie inschakelen."

'Elk bericht op Facebook zal een label meekrijgen'
VÉRONIQUE HOSTE
ONDERZOEKER UGENT

De onderzoekers werken samen met onder meer Netlog, Child Focus en met de VRT, dat dagelijks tienduizenden kinderen naar ketnet.be ziet komen. "Het is voor ons heel belangrijk dat een site als Ketnet een veilige omgeving vormt voor kinderen", beseft VRT-woordvoerster Anneke Ernon. "We zetten sterk in op sensibilisering en de webredactie leest ook nog eens alle berichten na die de kinderen online plaatsen."

"Op sites als Facebook is volledige manuele controle echter onmogelijk", weet Hoste. "Er is gewoon te veel input."

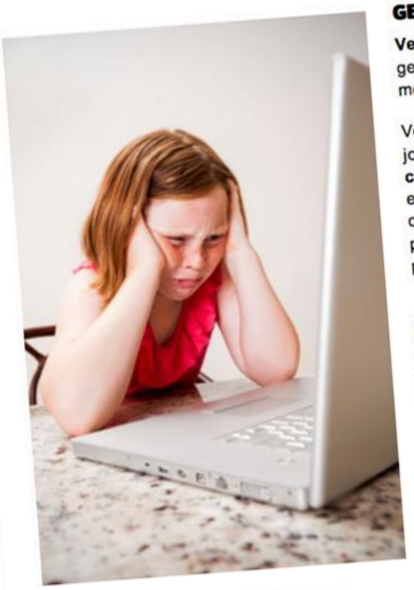
Het Belgische onderzoeksteam speelt internationaal een voortrekkersrol in het automatisch opsporen van pestgedrag. Of het project uiteindelijk zal slagen, zal grotendeels afhangen van de medewerking van Twitter en vooral Facebook, waar veel berichten niet openbaar zijn.

Cyberpestberichten kunnen via data@amicaproject.be worden doorgestuurd aan het onderzoeksteam.

K voor OUDERS

ACTUEEL · OVER KETNET.BE · VEILIG INTERNET · OVER KETNET · KAATJE VOOR OUDERS

GEZOCHT: CYBERPESTBERICHTEN



GEZOCHT: CYBERPESTBERICHTEN

Veilig internetten is belangrijk, maar niet altijd even gemakkelijk. Heb jij al een keer **rare** of **onveilige** dingen meegemaakt op het internet?

Verscheidene studies geven aan dat **één** Vlaamse jongere **op tien** recent het slachtoffer was van **cyberpesten**. De impact van dergelijke situaties is vaak erg groot doordat kwetsende berichten soms langere tijd online blijven staan en slachtoffers alleen met dit probleem zitten omdat ze er niet over willen of durven te praten.

SAMEN STERK

In het wetenschappelijk project AMiCA slaan een aantal onderzoeksgroepen van de universiteiten van Antwerpen, Leuven en Gent de handen in elkaar om een systeem te ontwikkelen dat automatisch kritieke situaties zoals cyberpestgedrag en seksueel grensoverschrijdend gedrag herkent op sociale netwerksites om zo een **veilige internetomgeving** te kunnen garanderen voor jongeren.

MEDIACAMPAGNE

HET NIEUWSBLAD WOENSDAG 27 NOVEMBER 2013 11

VLAAMSE ONDERZOEKERS ONTWIKKELEN SYSTEEM TEGEN CYBERPESTEN

'Stuur ons uw haatmail'

Waarom draagt gij altij make up aan, Verberg u lelijkheid nie helpt tog nie

gij zijt egt ne mega nerd hoe gij durft buitenkomen??? ik snap het ...

Gezocht: cyberpestberichten

Beste,

In het kader van ons onderzoek naar cyberpesten aan de universiteiten van Antwerpen, Leuven en Gent hebben we een oproep gelanceerd om zoveel mogelijk

radio 2 altijd dicht bij jou

De Universiteit van Gent zoekt cyberpestberichten

Tags: cyberpesten, gent, universiteit

Een aantal onderzoeksgroepen van de universiteiten van Antwerpen, Gent en Leuven zijn een systeem aan het

RESULTAAT

- 30-tal reacties
- ± 368 berichten (Facebookberichten, haatpagina's, Netlog, mail, chat, enz.)

Het doel van AMiCA (Automatic Monitoring for Cyberspace Applications) is een systeem te ontwikkelen dat automatisch cyberpestgedrag herkent op sociaalnetwerksites. Omdat te kunnen doen, hebben we enorm veel gegevens nodig', zegt Véronique Hoste van UGent. 'E-mails, sms'en, Facebook- en chatberichten ...

Vandaar de opvallende oproep die zij en haar collega's gisteren lanceerden. 'Of je nu het slachtoffer bent van cyberpesten of een getuige, leerlijk ook of ouder, bezorg ons de berichten die je beschouwt als cyberpesten.' De onderzoekers benadrukken dat die volledig anoniem behandeld worden.

Het systeem dat ze met die gegevens zullen ontwikkelen, moet weldra elk bericht scannen dat binnenkomt op Twitter of ander

VÉRONIQUE HOSTE Universiteit Gent

Er bestaan nu al knoppen waarmee je kwetsende berichten kunt melden. Wij willen dat systeem automatiseren

Het project van de Vlaamse universiteiten is door zijn omvang nu al uniek. Maar voorlopig focussen ze zich enkel op het Nederlands. Al kan het systeem in een later stadium ook voor het Engels gebruikt worden. Maar dan zullen we opnieuw manmaal gegevens moeten verzamelen.'

INFO Bezorg je cyberpestberichten via data@amicaproject.be

vindt u op de website:
http://www.amicaproject.be/

Voor meer informatie over deze oproep of ons project kan u steeds terecht bij:

Cynthia Van Hee	Universiteit Gent	cynthia.vanhee@ugent.be
Ben Verhoeven	Universiteit Antwerpen	ben.verhoeven@uantwerpen.be

Nieuw systeem spoort online

SAM FEYS
BRUSSEL

Belgische onderzoekers willen cyberpesten tegengaan door pestgedrag op Facebook en Twitter automatisch op te sporen. Het systeem moet op termijn zelfs ingrijpen bij zelfmoordgedachten.

Cyberpesten is vandaag nog erg moeilijk op te sporen, zeker omdat slachtoffers er zelden voor uitkomen. Verschillende onderzoekers van de universiteiten van Gent, Leuven en Antwerpen slaan daarom de handen in elkaar voor AMiCA, een systeem dat automatisch pestgedrag op het internet moet herkennen. Zo kunnen zowel sitebeheerders als politie ingrijpen voor het te laat is.

"We willen zowel tekst als beeld op sociale media analyseren om zo volautomatisch alarmsignalen te listen afgaan bij verontrustende situaties", legt Véronique Hoste van UGent het project uit. "Zo kunnen we zowel gevallen van online pestgedrag en pedofilie opsporen, maar ook ingrijpen als jongeren met suicidale gedachten kampen."

Het onderzoeksteam is daarom op zoek naar slachtoffers van cyberpesters die ooit beledigende sms'jes, e-mails of berichten kregen, ook via sociale netwerksites als Twitter, Facebook en Ask.fm. "De zullen we uitgebreid analyseren. Op basis daarvan kunnen we accurate computermodellen bouwen die kunnen afleiden wanneer het om cyberpesten gaat. Hoe meer bruikbare voorbeelden we binnenkrijgen, hoe beter we dat model kunnen voeden."

De bedoeling is om uiteindelijk

pestgedrag automatisch op

'Elk bericht op Facebook zal een label meekrijgen'

VÉRONIQUE HOSTE
ONDERZOEKER UGENT

elk nieuw bericht op Facebook of Twitter te toetsen aan dat model. "Elk bericht zal een label meekrijgen", verduidelijkt Hoste. "Als bij meerdere berichten van een gebruiker rode lampjes gaan branden, zal een moderator of sitebeheerder automatisch op de hoogte worden gebracht. Die kan dan de ernst van de situatie inschatten en eventueel de politie inschakelen."

De onderzoekers werken samen met onder meer Netlog, Child Focus en met de VRT, dat dagelijks tienduizenden kinderen naar ketnet.be ziet komen. "Het is voor ons heel belangrijk dat een site als Ketnet een veilige omgeving voor het voor kinderen", beseft VRT-woordvoerster Anneke Ernon. "We zetten sterk in op sensibilisering en de webredactie leest ook nog eens alle berichten na die de kinderen online plaatsen."

"Op sites als Facebook is volledige manuele controle echter onmogelijk", weet Hoste. "Er is gewoon te veel input."

Het Belgische onderzoeksteam speelt internationaal een voortrekkersrol in het automatisch opsporen van pestgedrag. Of het project uiteindelijk zal slagen, zal grotendeels afhangen van de medewerking van Twitter en vooral Facebook, waar veel berichten niet openbaar zijn.

Cyberpestberichten kunnen via data@amicaproject.be worden doorgestuurd aan het onderzoeksteam.

GEZOCHT: CYBERPESTBERICHTEN

GEZOCHT: CYBERPESTBERICHTEN

Veilig internetten is belangrijk, maar niet altijd even gemakkelijk. Heb jij al een keer rare of onveilige dingen meegemaakt op het internet?

Verscheidene studies geven aan dat één Vlaamse jongere op tien recent het slachtoffer was van cyberpesten. De impact van dergelijke situaties is vaak erg groot doordat kwetsende berichten soms langere tijd online blijven staan en slachtoffers alleen met dit probleem zitten omdat ze er niet over willen of durven te praten.

SAMEN STERK

In het wetenschappelijk project AMiCA slaan een aantal onderzoeksgroepen van de universiteiten van Antwerpen, Leuven en Gent de handen in elkaar om een systeem te ontwikkelen dat automatisch kritieke situaties zoals cyberpestgedrag en seksueel grensoverschrijdend gedrag herkent op sociale netwerksites om zo een veilige internetomgeving te kunnen garanderen voor jongeren.

GHENT UNIVERSITY

DATAVERZAMELING: SIMULATIES

Resultaat:

3500 berichten

GHENT UNIVERSITY

DATA-ANNOTATIE

- Welke berichten bevatten cyberpesten?
- Welke vormen van cyberpesten komen voor?

CATEGORIEËN

- Bedreiging
- Belediging
- Verwensing/uitsluiting
- Reputatie schaden
- Seksueel getinte berichten
- Verdedigende berichten (slachtoffer, bijstaander)
- Aanmoedigingen aan de pester

GHENT UNIVERSITY

DATA-ANNOTATIE

Hi bitchess, zondag afspreken? x

Kmoet naar de Chiro.

Ok, kom je chat dan?

Pester

Waarom doet ge alsof ge een supermodel zijt? Ge zijt dik

Slachtoffer

u mening en door wie heb ik hier minder zelfvertrouwen gekregen door fucking ask.fm :(

Bijstaander

Luister daar niet naar als je daar iets wil aan doen (niet grof) probeer beetje te vermageren niemand is perfect blijf wie je bent en blijf sterk..

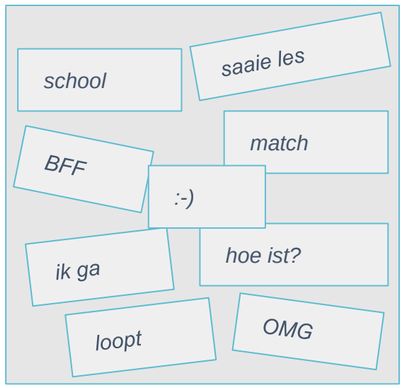
GHENT UNIVERSITY

LEREND SYSTEEM VOOR CYBERPESTDETECTIE

- Geannoteerde data = **voorbeelden** om lerend systeem te trainen
- Binaire taak:

  positieve klasse → cyberpesten

  negatieve klasse → geen cyberpesten
- Tekstuele kenmerken om **cyberpestberichten** te detecteren:
  - Woordsequenties
  - Polariteitsinformatie (positief vs. negatief)
  - Kenmerkende themas's (vb. uiterlijk, intelligentie, …)
  - …



GHENT UNIVERSITY

RESULTATEN

Hoe goed kunnen we pestberichten detecteren?

	Precisie	Bereik *(recall)*	F1-score
NL	76%	56%	65%
EN	74%	55%	63%

MAAR:

- Impliciet taalgebruik → *Mijn vuisten jeuken al..*
- Wereldkennis nodig → *Een kleedje zou u niet misstaan ;)*
- Beperkte context → *Mah gij seut!*

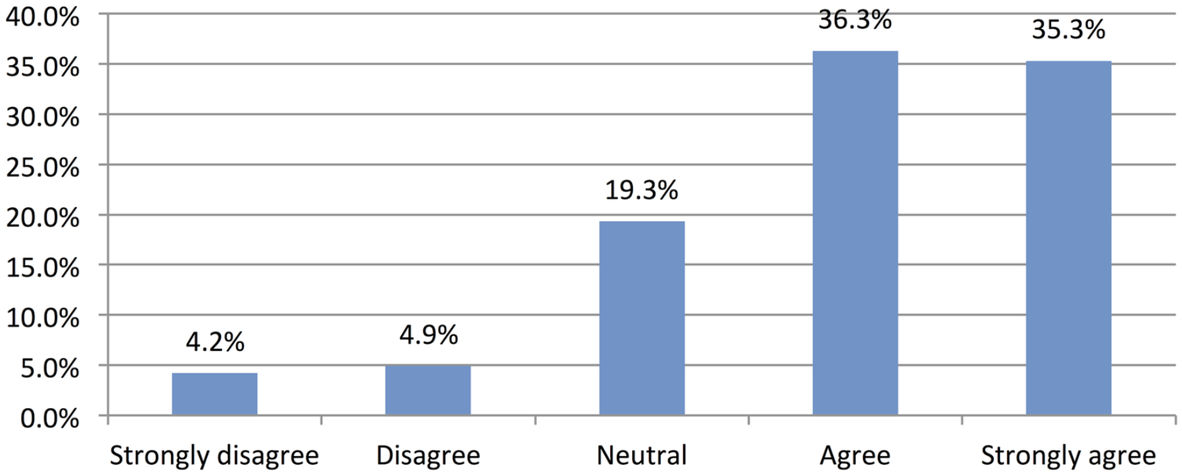
GHENT UNIVERSITY

Levert gesimuleerde data werkbaar materiaal op voor de detectie van echte cyberpestberichten?

JA,

- ASKfm data: F1 = 69%
- Gesimuleerde data: F1 = 66%
- ASKfm + gesimuleerd: F1 = 73%

WENSELIJK?



Grafiek 2: Wenselijkheid van cyberpestdetectie op sociale media (***N*** = 770) volgens jongeren (Van Royen et al., 2014).

Experten*:

- Follow-up nodig
- Privacy jongeren moet gerespecteerd worden
- Technisch haalbaar?

*(Van Royen et al., 2014).

GHENT UNIVERSITY

VERVOLG

- Simuleren van cyberpestsituaties m.b.v. roldetectie:

  Voordelen:

  + Accurater detecteren van cyberpesten
  + Betere inschatting nodige tussenkomst / follow-up

- Monitoren van cyberpesten over de tijd heen

HAHA dwazen! Het scheelt dat je Charlotte kunt krijgen, puistenkop!

Laat Charlotte erbuiten...

‘T is niet voor niks da je altijd alleen staat op de speelplaats...

Heet hij u iets misdaan? Ik dacht het niet toch? Laat hem met rust!

Trek het u niet aan.. Hij is zelf zo dwaas als iets :)

GHENT UNIVERSITY

www.amicaproject.be