

Persbericht

13 mei 2013

Computer raadt leeftijd en geslacht twitteraar TweetGenie minstens zo nauwkeurig als de mens

Onderzoekers van het Meertens Instituut en de Universiteit Twente lanceren vandaag TweetGenie (www.tweetgenie.nl), een computerprogramma dat aan de hand van taalgebruik de leeftijd en het geslacht van Nederlandse twitteraars kan raden. In 85 procent van de gevallen raadt het programma het geslacht van twitteraars correct. Bij de inschatting van de leeftijd zit de computer er gemiddeld minder dan vier jaar naast. Hiermee is de inschatting van de computer al iets nauwkeuriger dan die van mensen.

Iedereen twittert op zijn eigen manier. Toch zijn er voor groepen allerlei wetmatigheden te vinden. Uit onderzoek van het Meertens Instituut en de Universiteit Twente komt bijvoorbeeld naar voren dat jongeren op Twitter vaker over zichzelf en over onderwerpen als school praten. Ook gebruiken ze meer smileys dan oudere twitteraars. Deze gebruiken op hun beurt bijvoorbeeld weer langere woorden, langere zinnen, links en hashtags in hun tweets.

De onderzoekers hebben nu een eerste versie van een computerprogramma ontwikkeld, TweetGenie, dat op basis van iemands tweets redelijk nauwkeurig de leeftijd en het geslacht van die persoon in kan schatten. Het programma kijkt hierbij enkel naar het taalgebruik van de twitteraar, en niet naar de naam, de foto of het profiel.

Beter dan de computer?

Je kunt elk twitterprofiel in het programma invullen (van mensen die voornamelijk in het Nederlands twitteren). Ook kun je met het programma, op basis van geanonimiseerde twitterprofielen, kijken of jij beter bent in het inschatten van leeftijd en geslacht van een willekeurige twitteraar dan de computer. Vandaag wordt de eerste versie van het programma gelanceerd. De onderzoekers hopen, op basis van de feedback die mensen geven, het programma verder te verbeteren. Zo is het op het moment bijvoorbeeld nog erg moeilijk om de leeftijd van twitteraars ouder dan 35 goed in te schatten.

Verhalen en geruchten

Het programma is onderdeel van het promotieonderzoek van Dong Nguyen. Zij wil onder meer onderzoeken hoe verhalen en geruchten zich via sociale media verspreiden. Hierbij is het van groot belang om verschillende typen gebruikers te kunnen onderscheiden. Verder is de ontwikkelde techniek mogelijk ook interessant voor allerlei marketingdoeleinden.

Onderzoek

Het programma is ontwikkeld door onderzoekers van het Meertens Instituut en de Universiteit Twente. Het Meertens Instituut – dat zich bezighoudt met de bestudering van de Nederlandse taal en cultuur, waarbij verschijnselen die het alledaagse leven in onze samenleving vormgeven centraal staan – leverde de benodigde taalkundige expertise. Onderzoekers van de vakgroep Human Media Interaction van het onderzoeksinstituut CTIT van de UT waren verantwoordelijk voor de meer technische kant van het onderzoek. Het onderzoek is financieel mede mogelijk gemaakt door NWO en KNAW.

Noot voor de pers, niet voor publicatie

Het programma vindt u op www.tweetgenie.nl. Voor meer informatie, of een digitale versie van het wetenschappelijke artikel “How Old Do You Think I Am?": A Study of Language and Age in Twitter, kunt u contact opnemen met UT-wetenschapsvoorlichter Joost Bruysters (06 1048 8228).