

Stageopdracht analyse van Nederlandstalige tweets

Bedrijfsinformatie

RTreporter is een start-up die zich bezig houdt met het detecteren van nieuws op Twitter voordat dit bij de massamedia bekend is. Op dit moment draait het systeem van RTreporter bij NU.nl en ANP en worden er enkele nieuwsberichten per dag op die manier gevonden die door NU.nl uitgewerkt worden en geplaatst worden. RTreporter is ontstaan uit een samenwerkingsverband van de Vrije Universiteit Amsterdam, CCinq en NU.nl. RTreporter valt tegenwoordig onder ORTEC Living Data en is gevestigd in Amsterdam. Bij ORTEC houden 650+ medewerkers zich bezig met optimalisatie.

Opdracht

We zijn op zoek naar enthousiaste studenten die ons willen helpen met de ontwikkeling van RTreporter op het gebied van geautomatiseerde analyse op Twitter. Binnen RTreporter kun je aan een breed scala aan problemen werken, maar we staan ook altijd open voor nieuwe suggesties.

De lijst van mogelijkheden bestaat uit (maar is dus niet gelimiteerd tot):

- Het ontwikkelen van een software-component die tientallen miljoenen (Nederlandstalige) tweets per dag kan analyseren op aanwezige entiteiten
- Het formuleren van een samenvatting van gerelateerde tweets naar aanleiding van de beschikbare informatie van deze tweets.
- Het analyseren van de ontwikkeling van de tweets met betrekking tot een specifieke gebeurtenis.
- Een analyse van de betrouwbaarheid van de gebruikte tweets.
- Het opzetten van taal- en regiodetectie voor tweets.
- Big Data oplossingen zoals Twitters storm-project.net of andere real-time varianten van Hadoop.

Studierichting

Opdrachten zitten in de hoek van Computer Science/Artificial Intelligence/Data Mining/Computational linguistics/Communication Science, maar een andere studierichting hoeft geen enkel probleem te vormen.

Technologie

Onze huidige software is JVM-gebaseerd. We staan natuurlijk open voor andere technieken.

Interesse?

Peter Kampstra, peter@rtreporter.com, skype: pkampst

Website: www.rtreporter.com