

COMMIT/



Wie, Wat, Waar, Wanneer
UNIVERSITEIT TWENTE. ARCADIS



Door: Maurice van Keulen^(*), Victor de Graaff^(*), Zhemin Zhu^(*), Rolf de By^(S), Andreas Wombacher^(*), Jan Flokstra^(*) ^(*)DB-groep¹, Faculteit EWI; ^(S)GIP-groep², Faculteit ITC

Het project en onze partners

Het TimeTrails-project³ gaat over data mining in grote hoeveelheden gegevens over gebeurtenissen in ruimte en tijd, d.w.z. met coördinaten en time-stamps. Dergelijke gegevens worden doorgaans vergaard door mensen, sensoren en wetenschappelijke observaties. Gegevensanalyse richt zich vaak op de vier W's: Wie, Wat, Waar en Wanneer. Een belangrijke kwestie is het kunnen behappen van de grote hoeveelheden gegevens, d.w.z. "big data". Vanuit de UT werken we, d.w.z. de groepen EWI/DB en ITC/GIP, aan twee applicaties:

- Het in kaart brengen van de mening van het publiek bij grote infrastructuurproject zoals de aanleg van een nieuw stuk snelweg. Dit doen we met Twitter-analyse en data-visualisatie.
- Het vinden van goede vakantiebestemmingen. Hierbij spelen Social media, web harvesting en analyse van GPS-traces een rol.

De COMMIT/-manier van onderzoek doen

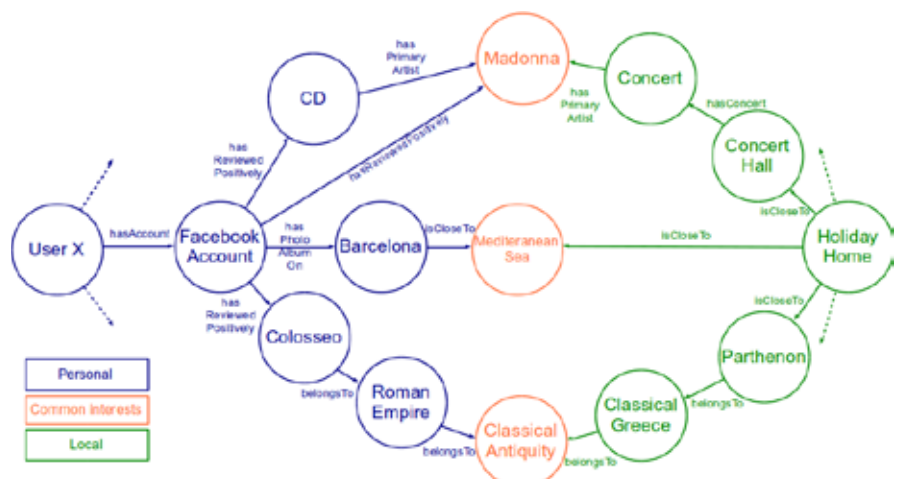
TimeTrails is één van 16 projecten in het COMMIT/-programma. Een van de principes van COMMIT/ is het werktafel-model: elke universiteit of onderzoeksinstelling werkt nauw samen met één of enkele bedrijven. Zo werken wij in de eerste applicatie samen met Arcadis⁴ en in de tweede met EuroCottage⁵. Een ander principe van COMMIT/ is

"Geografische data kan ons helpen bij het maken van beslissingen. De vraag is nu: hoe?"

use-inspired: de bedrijven bepalen niet wat er onderzocht moet worden, maar zijn inspiratie voor onderzoek dat uiteindelijk de maatschappij en de economie van Nederland echt kan verbeteren. Er wordt in COMMIT/ veel aandacht besteed aan onderzoeksresultaten uitwerken tot bruikbare software en communiceren naar een breed publiek.

Arcadis is een internationaal bedrijf dat onder meer diensten levert rondom grote infrastructuurprojecten. Arcadis wil graag meer grip krijgen op wat het publiek vindt over zo'n project, bijvoorbeeld wat omwonenden vinden van een bepaald snelwegtraject of de locatie van een afrit. Twitter zou hen die informatie kunnen geven. Hoewel tweets over van alles kunnen gaan, worden ze voornamelijk gebruikt om meningen en gevoelens te uiten over wat er om ons heen

De mening van het publiek bij grote infrastructuurprojecten



Figuur 1: Een goede match tussen gebruiker en vakantiehujsje.

gebeurt. We werken zowel aan het kunnen visualiseren van discussies op een kaart en tijdslijn als ook aan taalanalysetechnologie. Met het laatste hebben we recentelijk een Twitter-analyse challenge gewonnen [3]. Uiteindelijk hopen we de redenen voor bepaalde meningen te kunnen achterhalen.

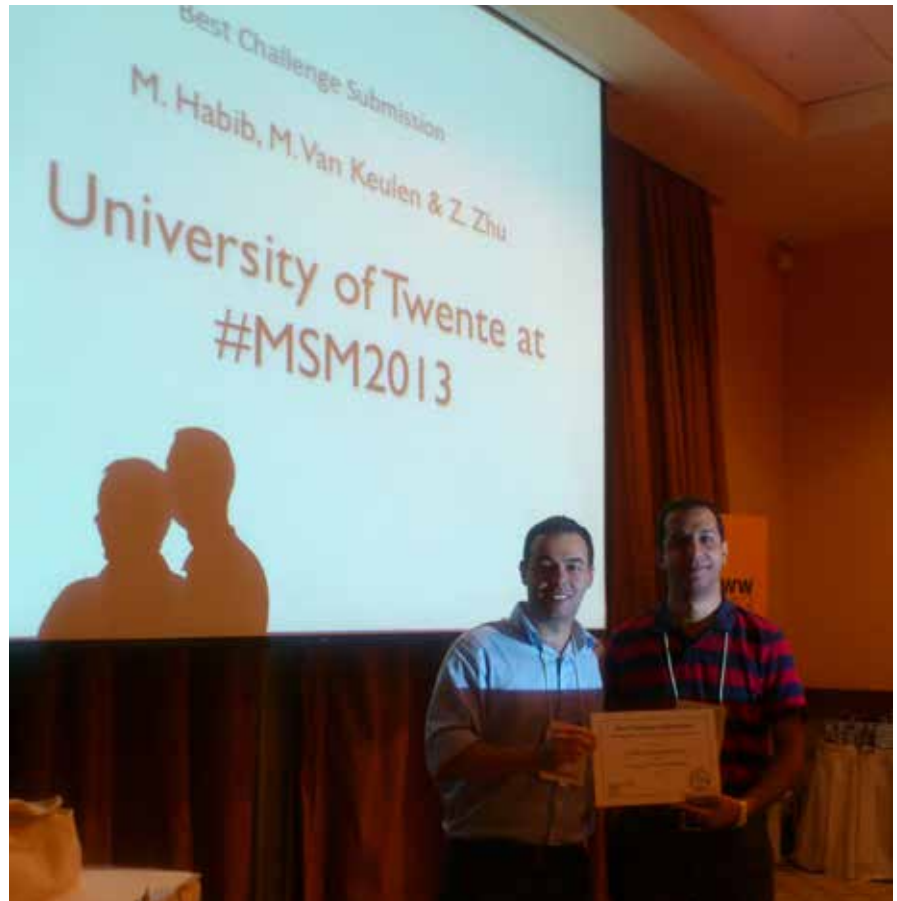
Het vinden van goede vakantiebestemmingen

EuroCottage is een bedrijf dat vakantiehuizen door heel Europa aanbiedt. Vakantiegangers hebben over het algemeen een helder beeld van hoe de vakantie eruit moet zien: een wintersport-vakantie, een strandvakantie, een fietstocht, etc. In welk land dit plaats zal gaan vinden is meestal van ondergeschikt belang. Toch is juist dit laatste de vraag die als eerst gesteld wordt op vakantie-websites (zie Figuur 2). Door middel van “geo-social recommender systems” proberen wij de match tussen vakantieganger en vakantiehuisje op een hele andere manier te benaderen.

We zoeken voor ieder vakantiehuisje naar interessante objecten in de buurt. Deze informatie stamt uit vele verschillende online bronnen, zoals Wikipedia, de Gouden Gids, restaurant-websites, etc. De vakantieganger kan dan d.m.v. andere bronnen (bijv. een Facebook-profiel, GPS traces, of d.m.v. een wizard) een gebruikersprofiel vullen. Het systeem zoekt dan hiermee naar goede matches met vakantiehuisjes. Als een



Figuur 2: op zoek naar een strandvakantie.



gebruiker bijvoorbeeld graag op vakantie gaat rondom de Middellandse zee, en een fan is van Madonna en de klassieke oudheid, dan kan deze gebruiker een aanbeveling krijgen voor een Grieks vakantiehuisje in de buurt van een Madonna-concert (zie Figuur 1).

Tot slot

Ben je geïnteresseerd in data mining, natuurlijke taal-analyse, data-visualisatie, big data, web harvesting of het omgaan met beperkte gegevenskwaliteit [2]? Of zoek je naar een afstudeeropdracht, mogelijk deels bij het ITC? Wil je op de hoogte gehouden worden? Neem gerust contact op met Maurice van Keulen (m.vankeulen@utwente.nl).

Referenties

[1] de Graaff, V. and van Keulen, M. and de By, R.A. “Towards Geosocial Recommender Systems”. In: 4th International Workshop on Web Intelligence & Communities (WI&C 2012), 16 Apr 2012, Lyon, France. pp. 8:1-8:4. <http://eprints.eemcs.utwente.nl/21619/>

[2] Keulen, M. (2010) “Onzekere databases”. DB/M: database magazine, 21 (4). pp. 22-27. ISSN 0925-6911. <http://eprints.eemcs.utwente.nl/18030/>

[3] Habib, M.B. and van Keulen, M. and Zhu, Zhemín, “Concept Extraction Challenge: University of Twente at #MSM2013”. In: Proceedings of the 3rd workshop on ‘Making Sense of Microposts’ (#MSM2013), 13 May 2013, Rio de Janeiro, Brazil. Best IE challenge submission. <http://eprints.eemcs.utwente.nl/23249/>