

Verslag RD1 2008/2009 - Groep 4

ONDERZOEK NAAR DE BEHOEFTE AAN  
INTELLIGENTE E-MATCHING BIJ  
PROFIELPAGINA'S D.M.V. RECOMMENDER  
SYSTEMS EN COLLABORATIVE FILTERING

**Michiel van Lierop**

# 1 Abstract

*In dit onderzoek wordt gekeken naar enkele mogelijkheden die er zijn om een datingsite uit te breiden met enkele zoekalgoritmes geïnspireerd op het principe van Collaborative Filtering. Uiteindelijk wordt een methode gevonden die Item-to-Item heet, die op een realistische en bruikbare wijze geïmplementeerd kan worden. Uit deze methode zijn twee functies voortgevloeid, die beiden een significante bijdrage kunnen leveren aan de effectiviteit van datingsites. De eerste functie maakt het mogelijk om gemakkelijk nieuwe, relevante en interessante gebruikers te vinden, uitgaande van huidige interesses van gebruikers. De tweede functie draait de principes van Item-to-Item filtering om, teneinde de kansen die een gebruiker maakt bij een profiel waarin hij/zij geïnteresseerd is te bepalen. Een vragenlijst hierover heeft aangetoond dat er in ieder geval één marktsegment is, jong volwassenen tussen de 18 en 25 jaar, dat meerwaarde ziet in deze uitbreidingen.*

## 2 Inleiding

We leven in een tijd waar het internet een steeds grotere rol gaat spelen in ons dagelijks leven. Contact houden met mensen geschied niet langer voornamelijk via ontmoetingen, of zelfs via de telefoon. Social Network sites als Facebook en Hyves zijn een geaccepteerde en populaire manier geworden om contact te houden met mensen, en zelfs om nieuwe mensen mee te ontmoeten. De moderne single kan zijn dates vinden met behulp van datingsites zoals Relatieplaneet.nl en Parship.nl. Blogs worden gebruikt om meningen zichtbaar te maken voor de hele digitale wereld. Informatie over vrijwel alles is vrij beschikbaar dankzij initiatieven zoals Wikipedia. Complete instructiefilms zijn online gezet op YouTube, en maken het mogelijk vrijwel iedere vaardigheid te leren van het internet.

In deze digitale tijd is de voornaamste moeilijkheid het vinden van de correcte informatie, de juiste mensen. Slimme zoekmachines als Google voorzien hier gedeeltelijk in, maar zelfs de intelligente algoritmes van Google schieten wel eens tekort. Het is dus tijd te zoeken naar nieuwe mogelijkheden, die ons helpen bij onze speurtochten over het WWW.

Dit onderzoek probeert niet te voorzien in de nieuwste zoektechnologie, maar probeert een voorzichtige eerste stap te zetten in die richting. Het streven is om de collaborative filtering techniek, vooral bekend van sites als Amazon.com, Bol.com en Last.fm (“Mensen die dit product kochten, kochten ook...”), te combineren met profielpagina’s. Als voorbeeld worden datingsites genomen. Gebruikers maken een profiel aan en zoeken vervolgens

naar geschikte partners door de profielen van andere mensen te bekijken en beoordelen. De datingsite helpt hierbij door te voorzien in een zoekmachine, die profielen met elkaar vergelijkt en vervolgens rangschikt op overlap. Het doel van dit onderzoek is een lijst van voorwaarden te formuleren, d.m.v. een model, die suggereren hoe deze zoektechniek uit te breiden is met een collaborative filtering techniek, die we baseren op het systeem van Bol.com.

## 3 Theorie

De achtergrondinformatie van dit onderzoek kan in ruwweg twee delen gesplitst worden: theorie over hoe men op zoek gaat naar geschikte partners op een datingsite, en informatie over collaborative filtering. In de komende secties zal deze theorie besproken worden.

### 3.1 Methoden in het online zoeken naar geschikte partners

Online een geschikte partner vinden is een fenomeen dat nog niet veel onderzocht is. Hoe mensen naar een profiel kijken, en hoe ze daaruit de persoonlijkheid van de persoon achter het profiel afleiden, alvorens te besluiten of deze persoon al dan niet een geschikte partner is, valt nog niet concreet uit te drukken.

Wel is inmiddels bekend dat het voornamelijk de foto en de vrije tekst op datingsites zijn die het verschil uitmaken (3). De “standaard” invulopties die gebruikers hebben, waar ze een keuze moeten maken uit een lijst van antwoorden die vaststaan, blijken niet tot nauwelijks effect te hebben op de aantrekkelijkheid van een profiel. Mensen waarderen vooral de eigen inbreng van mensen, omdat dit meer zegt over iemand persoonlijkheid. Niet alleen de antwoorden van een persoon doen ertoe, ook uit schrijfstijl en correctheid worden conclusies getrokken.

De foto is significant belangrijker dan alle andere elementen van een profiel. Mensen met een aantrekkelijke foto krijgen meer blijk van interesse van andere leden dan mensen met een kwalitatief mindere foto. Mensen met foto worden 10 tot 20 keer vaker benaderd dan mensen zonder foto. (3)

### 3.2 Collaborative Filtering

Klanten bij e-commerce winkels zoals Bol.com en Amazon.com hebben toegang tot een enorme catalogus van duizenden en soms zelfs miljoenen producten. Iets vinden in deze enorme berg van aanbod is zoals te verwachten

valt niet een gemakkelijke taak. Sterker nog, het zou bijna onmogelijk zijn iets te vinden dat precies aan de verwachtingen van een klant voldoet, als de klant geen precieze gegevens heeft over het gezochte product.

Als antwoord op dit probleem is een zoekmethode ontwikkeld die collaborative filtering heet. Collaborative Filtering (afgekort CF) is het systeem achter de welbekende zin "Mensen die dit product kochten, kochten ook" die klanten op de site zien als zij een bepaald product bekijken. Op deze manier is het mogelijk om vanuit een product, dat wellicht niet helemaal is wat je zoekt, bij soortgelijke, relevante artikelen te komen, die je zelf niet zo snel zou hebben gevonden. De manier waarop dit gebeurd is, simpel gezegd, door alle andere gebruikers mee te laten zoeken. Als mensen een product kopen of waarderen, wordt dit opgeslagen. Wanneer andere mensen vervolgens een product bekijken dat door anderen is gekocht, zien zij ook welke andere producten die personen gekocht hebben. Omdat er waarschijnlijk veel personen zijn die een bepaald product hebben gekocht, worden alleen de producten getoond van de klanten die het meest op de gebruiker lijken. Dit wordt uitgedrukt aan de hand van Similarity. Een mogelijke formule is (1):

$$\text{Similarity}(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} \bullet \vec{B}}{\|\vec{A}\| * \|\vec{B}\|} \quad (1)$$

Deze formule berekent in feite hoezeer twee klanten, A en B, overeenstemmen door ze uit te drukken als een vector waar in uitgedrukt wordt welke producten ze wel en niet hebben gekocht, en vervolgens uit te rekenen hoeveel dezelfde producten ze hebben gekocht. Om de aangeboden artikelen vervolgens nog wat relevanter te maken kan ook nog berekend worden welke artikelen het vaakst samen met het bekeken artikel werden gekocht.

Hoewel dit algoritme een complexiteit heeft van  $O(KP)$  in het slechtste geval, waarbij  $K$  het aantal klanten is en  $P$  het aantal producten, en waarbij alle klanten met alle producten moeten worden vergeleken, is het realistischer om te stellen dat  $O(K + P)$  dichter bij de eigenlijke complexiteit is. Niet alle klanten bestellen immers alle artikelen, maar alle klanten hebben wel iets besteld, en alle klanten doorlichten kost  $O(K)$ , en alle producten zijn wel een keer gekocht, waarbij alle producten doorlichten nog eens  $O(P)$  kost, opgeteld dus  $O(K + P)$ .

Echter is dit bij grote klantenbestanden en catalogi (denk aan enkele honderdduizenden of miljoenen klanten, en een catalogus van ongeveer 1 miljoen artikelen) nog steeds een erg groot getal, vooral als dit in real-time berekend moet worden en een antwoord dus binnen een halve seconde gewenst is, om klanten niet te lang op de website te laten wachten. Dit is bij allerlei methoden van CF een probleem (5).

Daarom heeft Amazon.com een nieuw algoritme ontwikkeld, dat *Item-to-Item* heet. Dit algoritme vergelijkt niet de klanten met elkaar terwijl ze aan het winkelen zijn, maar de artikelen die zij bekijken. Het grote voordeel is dat deze berekening offline uitgevoerd kan worden, waardoor deze niet bijdraagt aan de wachttijden van klanten. Het algoritme loopt als volgt, zie (1):

Voor elk product  $P_1$  in de catalogus

Voor elke klant  $K$ , die product  $P_1$  heeft aangeschaft

Voor elk product  $P_2$ , dat aangeschaft is door klant  $K$

Noteer dat een klant  $P_1$  en  $P_2$  heeft aangeschaft

Voor elk product  $P_2$

Bereken de overeenkomst tussen  $P_1$  en  $P_2$

De complexiteit van dit algoritme ligt, vergeleken met normale collaborative filtering technieken, hoger. Deze is namelijk, in het ergst geval,  $O(P^2K)$ . In realiteit ligt dit deze waarde echter meer in de buurt van  $O(KP)$ , omdat de meeste klanten niet zoveel artikelen besteld hebben. Echter, de volledige berekening kan offline worden uitgevoerd. Klanten hoeven dus niet te wachten op de site, terwijl het algoritme berekent wat de meest overeenkomstige producten zijn van een willeurig ander product. Het achterhalen van de producten die aanbevolen worden is vervolgens alleen nog maar afhankelijk van het aantal producten  $p$  dat de klant al eerder heeft gekocht.

Het berekenen van de overeenkomsten tussen de artikelen kan op verschillende manieren gaan. De formule die hierboven genoemd werd voor het berekenen van de overeenkomst tussen een klant  $A$  en een klant  $B$ , kan ook gebruikt worden om de overeenkomsten tussen artikel  $P_1$  en  $P_2$ . Een andere optie is het *Slope One* algoritme, dat niet binaire werkt (wel gekocht of niet gekocht) maar dat kan omgaan met een waarderingssysteem. Zo kunnen klanten een getal toewijzen aan een artikel, uit een range van bijvoorbeeld 1 tot 5, en dat wordt opgeslagen. Als klanten vervolgens andere artikelen bekijken, kan een schatting worden gemaakt van de waardering die zij aan het artikel zouden geven aan de hand van waarderingen die zij eerder hebben gegeven.

De werking van het *Slope One* algoritme wordt uitvoerig beschreven in (2)

Het *Item-to-Item* algoritme is momenteel het meest gebruikte algoritme voor grote e-commerce sites, zoals Amazon.com en Bol.com, en bovendien bewezen efficiënt. Om deze redenen zullen we gedurende dit onderzoek dit algoritme aanhouden als hetgeen waarmee ook wij willen gaan werken, en dat we willen implementeren in een datingsite.

## 4 Onderzoek

### 4.1 Onderzoeksvraag

*Wat is er nodig om een persoonlijk profiel geschikt te maken voor intelligente e-matching in de vorm van recommender systems?*

Teneinde deze vraag te beantwoorden, zal gepoogd worden een theoretisch model te maken voor een datingsite waarin een recommender system is geïmplementeerd. Daarbij is het essentiële punt hetgeen waarop het recommender system vergelijkingen maakt tussen gebruikers en/of potentiële dates.

### 4.2 Hypothese

Naar verwachting kan een recommender systeem op meerdere manieren in een datingsite verwerkt worden. Sommige opties zullen, realistisch bekeken, onwenselijk zijn, maar er kan in ieder geval gekeken worden of de implementatie van een recommender systeem volgens een bepaalde methode theoretisch plausibel is, zelfs al is deze methode onwenselijk voor daadwerkelijk gebruik. Immers, wellicht zijn er andere applicaties denkbaar.

Een wat minder optimistisch vooruitzicht wordt gesteld door het feit dat de meest bepalende elementen van een datingsite de foto en de vrije tekst zijn<sup>(3)</sup>. Hoewel dit in principe geen verschil maakt voor het recommender systeem, omdat dit functioneert door het gedrag van gebruikers, en niet door informatie op hun profiel, sluit het wel de mogelijkheid van partities uit, tenzij deze zeer ruim worden opgevat. Partities zouden eventueel gebruikt kunnen worden om het aantal producten en/of klanten waar een algoritme zich doorheen werkt te verkleinen, door te specificeren op een bepaalde doelgroep.

### 4.3 Onderzoeksplan

Ons plan van aanpak bestaat uit drie stappen:

1. Uitzoeken hoe mensen gebruik maken van de zoekopties van profiel-sites; waar selecteren mensen op?  
Hoewel het idee oorspronkelijk was om dit met een enquête na te gaan bij gebruikers van datingsites, is dit plan door tijdgebrek gewijzigd. Uiteindelijk is een bron gevonden die exact datgene had onderzocht wat wij nodig hadden. De informatie uit deze bron is te lezen in bovenstaande sectie over methoden in het online zoeken naar een geschikte partner.

2. Een site bestuderen die op effectieve wijze gebruik maakt van collaborative filtering; hoe gaat dit in zijn werk?

De sites die overduidelijk succesvol gebruik maken van recommender systems, en *Item-to-Item* in het bijzonder, zijn **Bol.com** en **Amazon.com**. Er is gezocht naar datingsites die al succesvol gebruik maken van een recommender system, maar deze zijn niet gevonden. De website **Last.fm** geeft echter wel een voorbeeld van een site die gebruik maakt van zowel profielen als recommender systems, zoals uitgevonden in Fase 1 van ons onderzoek.

3. Het samenvoegen van alle informatie om tot een theoretisch model te komen waarin de *Item-to-Item* benut wordt.

Er zal per opgeleverd model worden gekeken naar de mogelijkheden en wenselijkheid van een dergelijk systeem, d.m.v. een enquête.

De uitkomst van dit onderzoek is hopelijk een of meerdere modellen voor een functie die bijdraagt aan het gebruiksgemak van een datingsite, en die collaborative filtering gebruikt om gebruikers te helpen vinden wat zij zoeken.

## 5 Resultaten

De resultaten van het onderzoek zijn ingedeeld naar de drie stappen van het onderzoeksplan.

### 5.1 Zoekmethoden

Een recommender system is gebaseerd op het gegeven dat het de overeenkomst tussen twee vectoren kan berekenen. In dit geval zijn de vectoren gebruikers. Wat we willen vergelijken is hetgeen waar de gebruikers naar zoeken. Uiteindelijk moet een gebruiker dus samengevat kunnen worden in een reeks die het recommender system verteld waar de gebruiker in geïnteresseerd is.

Gedurende fase 1 van het onderzoek is de datingsite **Parship.nl** bestudeerd. Deze site maakt matches op basis van een psychologisch profiel, dat aangeemaakt wordt als gebruikers zich aanmelden op de site, door middel van een vragenlijst. Met deze vragenlijst, waar slechts een beperkt aantal antwoorden mogelijk zijn, kan de overeenkomst tussen gebruikers ook berekend worden, simpelweg door te berekenen in hoeverre hun antwoorden overeenkomen.

Uit onderzoek blijkt echter dat de gemiddelde gebruiker meer aandacht heeft voor de foto's op een profiel en de vrije tekst die op het profiel ingevoerd is (3). Tot in hoeverre het dus nuttig is om via een dergelijke vragenlijst te bepalen of twee gebruikers bij elkaar passen is onbekend. **Parship.nl** geeft

hier ook geen onderbouwing voor, in de vorm van wetenschappelijke artikelen of onderzoek, hoewel ze wel beweren wetenschappelijk gestaafd te zijn.

Huidige zoekmethoden op een datingsite zijn dus voornamelijk gebaseerd op zoekmachines die van een niveau zijn dat niet hoger is dan dat van de Google zoekengine. Dit zal dus weinig ‘intelligente’ antwoorden opleveren. Sites die gericht voorstellen voor matches doen, doen dit op basis van gemeenschappelijke tags. Meer overeenkomende tags wordt over het algemeen gekoppeld een grote slagingskans.

## 5.2 Evaluatie model Bol.com

Hier zijn niet veel resultaten aan te verbinden. Gebleken is dat **Bol.com** met een zeer hoge waarschijnlijkheidsgraad gebruik maakt van het *Item-to-Item* algoritme van **Amazon.com**. De klanttevredenheid hierover is hoog. Bol.com is meerdere malen verkozen tot beste webwinkel van Nederland(4), onder andere vanwege het brede assortiment en het gebruiksgemak van de site. De recommend functie draagt hier vrijwel zeker aan bij, omdat op klanten op deze manier sneller gemakkelijker artikelen vinden die bij hen in de smaak zullen vallen. Zowel voor de klant als de webwinkel levert dit dus voordeel op.

## 5.3 Samenvoegen modellen

Er zijn in totaal 2 algoritmes gevonden die van nut zouden kunnen zijn voor een datingsite, geïnspireerd op het *Item-to-Item* algoritme, zoals beschreven bij de Theorie.

In onderstaande algoritmes is er sprake van dat gebruikers interesse in elkaar tonen. Hoe “interesse tonen” precies gedefinieerd wordt is uiteraard aan de beheerders van de webpagina, maar dit onderzoek gaat ervan uit dat als een gebruiker interesse toont in een andere gebruiker, dit op actieve wijze dient te gebeuren, net als het bestellen van een product. Het bezoeken van een pagina zegt te weinig over het daadwerkelijke niveau van interesse om nuttige informatie op te leveren. Als dit toch gebruikt zou moeten worden, is er een systeem nodig wat de tijd bijhoudt die iemand actief besteed op een profiel, of het aantal keren dat een gebruiker een specifiek profiel bezoekt, of een andere dergelijke indirecte indicatie van langdurige interesse.

### 5.3.1 Functie 1

Dit algoritme is een bijna letterlijke vertaling van het *Item-to-Item* algoritme zoals uitgewerkt door **Amazon.com**(1). Het voornaamste verschil is dat er



geen onderscheid meer wordt gemaakt tussen Klant en Product. In feite is het zo dat een datingsite gezien kan worden als een e-commerce site waar de Producten hun eigen klanten mogen uitzoeken. Het tweede algoritme gaat dieper in op dit gegeven.

Voor elke gebruiker  $G_1$

Voor elke gebruiker  $G_2$ , die genteresseerd is in  $G_1$

Voor elke gebruiker  $G_3$ , die genteresseerd is in  $G_2$

Noteer dat een gebruiker genteresseerd is in  $G_1$  en  $G_3$

Voor elke gebruiker  $G_3$

Bereken de overeenkomst tussen  $G_1$  en  $G_3$

Het vermoeden is dat de overeenkomst het best te berekenen valt met het eenvoudige algoritme dat ook beschreven wordt in de Theorie en in (1). Het *Slope-One* algoritme voldoet waarschijnlijk ook prima, maar in eerste instantie is wel of geen interesse tonen in een gebruiker voldoende informatie.

### 5.3.2 Functie 2

Bij de eerste versie viel vooral op dat er bij datingsites geen onderscheid wordt gemaakt tussen klanten en producten. Alles is een gebruiker. Het bijzondere daaraan is dat het matchen niet meer vanuit 1 partij gebeurt. Niet alleen de klanten zoeken naar producten, de producten zoeken zogezegd ook actief naar klanten. In andere woorden: een product kan zijn eigen doelgroep uitkiezen.

Vanuit het oogpunt van datingsites is dit ook niet raar. Iedere gebruiker op een dergelijke site is op zoek naar een partner, en heeft bepaalde voorkeuren. Een eventuele interessante applicatie is nu dat gebruikers zouden kunnen controleren of zij voldoen aan de zoekcriteria van anderen. Het algoritme hiervoor gaat als volgt:

Voor elke gebruiker  $G_1$

Voor elke gebruiker  $G_2$ , waar  $G_1$  interesse in heeft getoond

Voor elke gebruiker  $G_3$ , waar  $G_1$  interesse in heeft getoond

Bereken de overeenkomst tussen  $G_2$  en  $G_3$

Tel de overeenkomst van  $G_2$  en  $G_3$  op

Met dit algoritme wordt in principe de overeenkomst tussen alle gebruikers waar  $G_1$  ooit in geïnteresseerd was berekend. De gebruiker die de grootste

opgetelde overeenkomst heeft met alle andere gebruikers die interesse opwekken in  $G_1$  wordt bepaald als de meest representatieve gebruiker  $G_r$  waarin  $G_1$  genteresseerd is.

Dit cree(e)rt de mogelijkheid een functie te maken, die een gebruiker kan activeren, als hij wil weten tot in hoeverre hij overkomt met de interesses van  $G_1$ . Het enige wat nu immers hoeft te gebeuren is de overeenkomst tussen  $G_r$  en de functie-activerende gebruiker  $G_a$  bepalen, daar een slagingspercentage aan te verbinden (bijvoorbeeld door te delen door de gemiddelde overeenkomst van  $G_r$  met andere gebruikers  $G_2$ ).

Een interessante situatie die dit kan opleveren is wanneer  $G_1$  interesse toont in  $G_a$  en  $G_a$  vervolgens gaat controleren tot in hoeverre hij/zij representatief is voor de gebruikers waarin  $G_1$  geïnteresseerd is. Dit kan namelijk tot gevolg hebben dat  $G_a$  uit de berekening komt als  $G_r$ , wat waarschijnlijk erg goed is voor het zelfvertrouwen van die gebruiker, en de slagingskans van de match.

## 5.4 Wenselijkheid

Beide algoritmen leveren een mogelijke functie op die als uitbreiding op een datingsite gebruikt kan worden. Om objectief te beoordelen of deze functies wenselijk zijn, is een vragenlijst uitgedeeld onder een groep mensen met de gemiddelde leeftijd van 21,3 jaar. Alle ondervraagden hadden een leeftijd tussen 18 en 25. De vragenlijst was kort, en bestond uit een tweetal simpele vragen:

1. Stel, je zit op een datingsite, en je vindt een persoon waarin je genteresseerd bent. Zou je het fijn vinden, als er een functie was waarmee je kon zien of je voldoet aan wat hij/zij zoekt?
2. Stel, je zit op een datingsite, en je vindt een persoon waarin je genteresseerd bent. Zou je het fijn vinden, als er een functie was waarmee je gemakkelijk soortgelijke personen kon vinden?

Beide vragen waren beantwoordbaar met een simpel “Ja” of “Nee”.

Onder de ondervraagden was 35,7% vrouw (10 in totaal), en 64,3% man (18 in totaal). De resultaten waren als volgt:

	Ja		Nee	
	Mannen	Vrouwen	Mannen	Vrouwen
Vraag 1	14	7	4	3
Vraag 2	14	8	4	2

Uit de vragen blijkt dat 77,8% van de mannen beide functies zeer zou waarderen. Van de vrouwen zou 70% het op prijs stellen van tevoren te weten wat hun kansen zijn. Een hoger percentage, 80%, zou het fijn vinden als ze gemakkelijk relevante nieuwe gebruikers konden vinden m.b.v. hun eerdere interesses. Bij beide geslachten is er dus een hoge mate van draagkracht voor de functies.

Bijna alle ondervraagden die “Ja” antwoorden op de eerste vraag gaven aan dat zij het zeer prettig zouden vinden om een indicatie te krijgen van hun kansen om in de smaak te vallen bij de ander.

Van degenen die “Nee” antwoorden op vraag 1 was er één persoon (mannelijk) die aangaf niet vooraf te willen weten wat zijn kansen waren. Dit zou de spanning uit het daten halen.

Ongeveer de helft van de ondervraagden die “Ja” antwoorden op de tweede vraag gaf aan dat ze het prettig zouden vinden om gemakkelijk hun eigen doelgroep uit te breiden.

Bij degenen die “Nee” antwoorden op de tweede vraag was er één persoon (vrouwelijk) die aangaf dat deze functie haar teveel deed denken aan **Bol.com**, en dat zij andere mensen niet als artikelen wenste te beschouwen.

## 6 Discussie

De resultaten waar de focus op ligt zijn de twee algoritmes die zijn voorgesteld in sectie 5.3 en de daaropvolgende inventarisatie van de wenselijkheid van dergelijke functies zoals te zien in sectie 5.4.

Het vermoeden bestaat dat beide functies daadwerkelijk kunnen bijdragen aan een datingsite. Er zijn echter wel wat nadelen denkbaar. Zo heeft het omschrijven van een algoritme voor een e-commerce site eigenlijk tot gevolg gehad dat gebruikers van een datingsite werden teruggebracht tot klanten en producten. Ethisch lijkt deze kwestie zeer kwalijk. Men zou deze stap kunnen interpreteren als het vergroten van de mogelijkheden om vooral met zoveel mogelijk mensen te daten. Dit zou kunnen bijdragen aan het versterken van het gevoel van losbandigheid, of wellicht het reduceren van de menselijke maat in het gebruik van de site. Omdat het zo makkelijk is om andere mensen te vinden, zullen sommige mensen zich wellicht geïntimideerd voelen. Dit is niet iets wat een dergelijke site wil uitstralen.

Een ander mogelijk gevolg zou kunnen zijn dat mensen teveel waarde hechten aan de functies. Dit geldt voor beide functies, maar voor functie 1 in het bijzonder. De functie geeft slechts een grove indicatie van de smaak van een ander, gebaseerd op eerdere interesses. Zoals met alles in de liefde zijn er echter geen concrete voorspellingen mogelijk. Het is zeer goed mogelijk

dat een persoon die volgens de eerste functie niet in de doelgroep valt van de ander toch een zeer goede indruk weet te maken. Hoogstwaarschijnlijk weten mensen hier overigens goed mee om te gaan. Het is immers ook niet zo dat gebruikers, als zij zoekmachines gebruiken, alleen vertrouwen hebben in de eerste tien resultaten.

Het tegendeel blijkt echter uit de enquête. Het merendeel van de ondervraagden zou beide functies zeer op prijs stellen, en verbond weinig tot geen negatieve implicaties aan de functies. Slechts een enkeling voelde zich oncomfortabel bij de wijze waarop de informatie werd aangeboden, of wat deze impliceerde. Helaas is verdere nuancering niet mogelijk, gezien het binaire karakter van de vragenlijst. Een vragenlijst met drie of vijf opties was, achteraf, waarschijnlijk beter geweest dan een hard ja/nee.

Een ander ethisch gevolg wat werd aangekaart is het probleem van deelname. Deelname aan het *Item-to-Item* systeem is passief. De gebruiker merkt er dus niets van. Maar ondertussen wordt wel zijn smaak vastgelegd, en andere mensen kunnen die zien en vergelijken. Dit zou gezien kunnen worden als een inbreuk op privacy. Gebruikers die niet wensen mee te werken aan een grootschalig koppelingsalgoritme zouden de mogelijkheid moeten hebben zich hieraan te onttrekken. Ondanks het gegeven dat de gebruikers van datingsites juist graag gezien willen worden, hebben zij toch recht op de vrijheid hier niet aan mee te doen. Daarom stelt dit onderzoek voor om, mocht implementatie volgen, gebruikers de optie te geven wel of niet meegerekend te worden in de *Item-to-Item* berekening. Een simpele boolean, die gebruikers via hun profiel aan of uit kunnen schakelen, die bepaald of de gegevens van de gebruiker meegenomen mogen worden, voldoet als oplossing voor dit probleem.

## 7 Conclusie

De meest voor de hand liggende conclusie van dit onderzoek is dat er zeker een marktsegment is, namelijk jonge mensen tussen de 18 en 25 jaar, dat een uitbreiding van de zoekfuncties van een datingsite, zoals voorgesteld door dit onderzoek, op prijs zou stellen.

Een logische en praktische toevoeging zou zijn om de twee functies toe te voegen op zo'n wijze dat gebruikers zelf bepalen of zij er gebruik van maken of niet. Een inschatting van de kansen van een gebruiker zou pas getoond worden op het moment dat de gebruiker heeft aangegeven dit daadwerkelijk te willen weten. Ook moeten gebruikers de keuzemogelijkheid krijgen deelname aan de collaborative filtering te weigeren. Dit zou eventuele complicaties met morele grenzen van gebruikers moeten vermijden.

Het antwoord op de onderzoeksvraag is dan ook duidelijk: datingsites zijn bij uitstek geschikt om collaborative filtering op toe te passen. De structuur die inherent is aan alle communities die met gebruikersprofielen werken, dus ook datingsites als **Parship.nl**, is vergelijkbaar met die van e-commerce sites, alleen wordt er geen onderscheid gemaakt tussen klanten en producten. De input van e-commerce sites (het kopen van producten) kan uitstekend worden vervangen (door het actief aangeven van interesse) waardoor algoritmes voor Collaborative Filtering prima functioneren.

## 8 Bronnen

1. **Greg Linden, Brent Smith, Jeremy York**, “Amazon.com Recommendations: Item-to-Item Collaborative Filtering”, IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan./Feb. 2003, doi:10.1109/MIC.2003.1167344
2. **Daniel Lemire, Anna Maclachlan**, “Slope One Predictors for Online Rating-Based Collaborative Filtering”, In SIAM Data Mining (SDM’05), Newport Beach, California, April 21-23, 2005
3. Conference on Human Factors in Computing Systems - Proceedings, 2008, Pages 797-806: **Fiore, A.T., Taylor, L.S., Mendelsohn, G.A., Hearst, M.**, “Assessing attractiveness in online dating profiles”
4. <http://www.bol.com/nl/m/persberichten/html-pers-consument-verkiest-bol-beste-webwinkel/index.html>
5. [http://www.nljug.org/pages/events/content/jfall\\_2007/sessions/00007/](http://www.nljug.org/pages/events/content/jfall_2007/sessions/00007/)

## 9 Logboek

10 juni:

- 10:30 - 12:30 Michiel bouwt eerste opzet enquête
- 12:30 - 13:00 Michiel en Gaby overleg
- 13:00 - 13:30 Gaby overleg met Erik en Sjaak, plan van aanpak wordt omgegooid

12 juni:

- 10:00 - 12:00 Michiel doet poging het Slope One algoritme te begrijpen. Bestuderen collaborative filtering. Proberen Bol.com te modelleren. Verscheidene pogingen mislukten, strandden op niet relevante informatie, teveel bijzaken, of voldeden niet aan het doel (namelijk aantonen hoe het item-to-item algoritme gecomplementeerd is in de site)

17 juni:

- 12:30 - 13:30 Michiel en Nurer overleg
- 13:30 - 14:00 Gaby werkt aan tweede versie enquête
- 13:30 - 18:00 Michiel bestudeert Collaborative Filtering, Item-to-Item Collaborative Filtering, en probeert dit te modelleren naar dating/profiel site. Enige vooruitgang wordt geboekt, onder andere door het vinden van literatuur over het onderwerp. Vermoed wordt dat het incorporeren van een dergelijk systeem lastig wordt in een datingsite, ook ivm morele consequenties. Desalniettemin gaat het onderzoek door.

18 juni:

- 14:00 - 16:00 Michiel bouwt verder aan het voorlopig verslag.
- Gaby stapt uit onderzoeksgroep

20 juni

- 14:00 - 18:00 Michiel typt verslag
- 16:00 Vragen wenselijkheid online

21 juni

- 12:00 - 19:30 Michiel typt verslag, interpretatie resultaten

22 juni

- 10:00 - 12:00 Michiel doet afronding verslag, klaar voor Peer Review

1 juli

- 14:00 - 15:00 Laatste aanpassingen aan verslag. Eindresultaat inleveren