

# Projectplan

---

Online sociale netwerken en discussiefora zijn het afgelopen decennium enorm in populariteit gegroeid. Dergelijke platformen bieden de mogelijkheid om gemakkelijk en snel informatie en interesses te delen met een potentieel groot publiek. Ook voor patiënten met een zeldzame of minder zeldzame ziekte zijn sociale media een effectieve manier om informatie en steun uit te wisselen met lotgenoten. In eerdere analyses hebben we gezien dat patiënten, immers ervaringsdeskundigen, veel kennis, ervaringen en tips uitwisselen die zeer relevant kunnen zijn voor zowel medisch specialisten als voor andere patiënten. Echter, de grote hoeveelheid berichten op de meeste discussiefora, het gebruik van spellingvarianten en synoniemen, en de vaak zeer beperkte zoekfunctionaliteit die bestaande platformen bieden maakt het bijzonder lastig om waardevolle informatie terug te vinden en al helemaal om nieuwe inzichten en kennis te ontdekken.

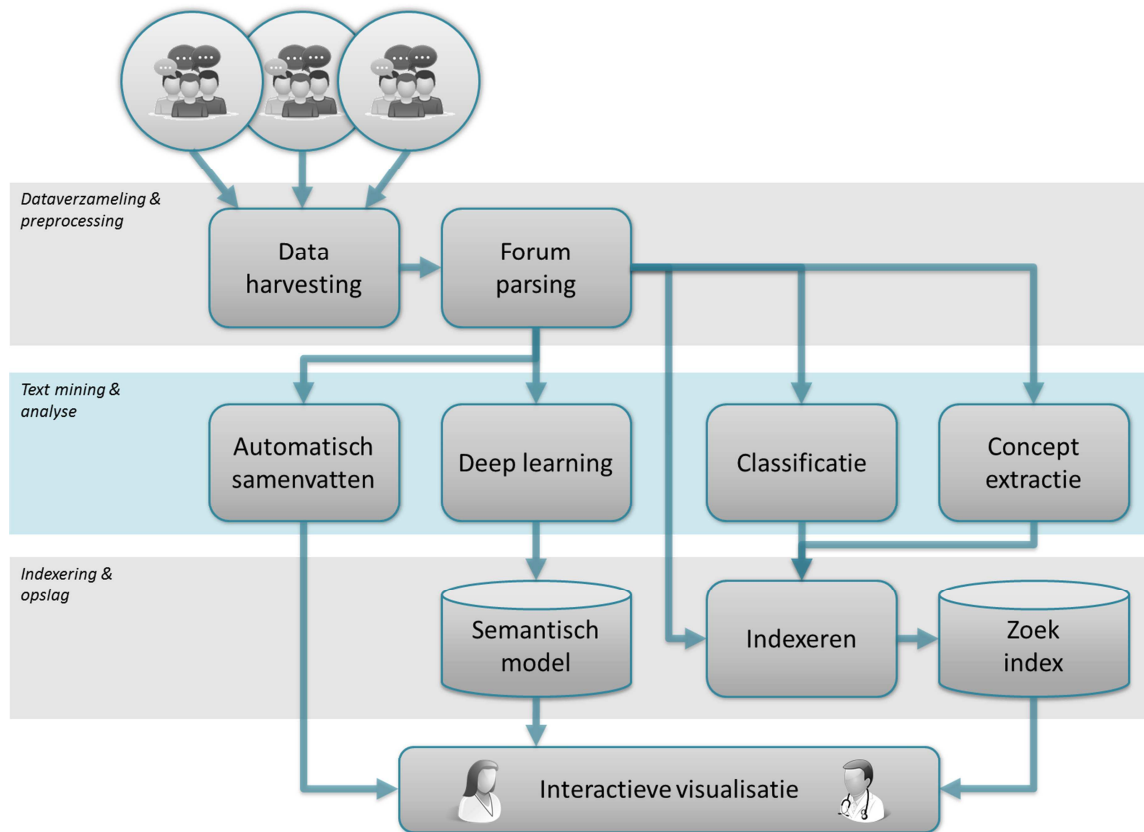
Het doel van dit project is om een prototype te ontwikkelen dat zowel de patiënt als de medisch specialist ondersteunt bij het vinden van relevante kennis en het genereren van nieuwe inzichten uit online patiëntenfora.

De primaire databron in het project is Kanker.nl. Kanker.nl is een online platform voor patiënten en hun naasten. Het is een initiatief van KWF Kankerbestrijding, patiëntenbeweging Levenmetkanker en IKNL (kenniscentrum voor zorgverleners in de oncologie). Zij hebben de handen ineen geslagen om betrouwbare informatie, ervaringskennis en het ondersteuningsaanbod rond kanker op één plek te bundelen en aan te bieden. De ambitie is te bouwen aan een betrouwbare, veilige en rijke internetomgeving waar de (ex-)kankerpatiënt en hun naasten terecht kunnen voor: informatie over kanker, lotgenotencontact en het uitwisselen van ervaringen. Begin augustus 2015 hebben 15.000 mensen een profiel aangemaakt. Daarnaast zijn er ruim 750 bloggers. Er komen dagelijks zo'n 10.000 bezoekers naar Kanker.nl.

Het prototype zal worden bediend middels een intuïtieve gebruikersinterface die een persoonlijke en interactieve blik biedt op de informatie die in de discussies verscholen ligt. Op de achtergrond zorgt een 'analyse-sstraat' van geavanceerde text mining modules voor de benodigde analyses en verrijking van de berichten. Figuur 1 bevat een schematische weergave van de architectuur van het prototype. De modules van het forumanalyse prototype kunnen worden ingedeeld in drie hoofdcategorieën: *Dataverzameling & preprocessing*, *Text mining & analyse*, en *Indexering & opslag*. Hieronder worden de modules per categorie in meer detail toegelicht.

## T2.1 – T2.2 Dataverzameling & preprocessing

**T2.1 Data harvesting:** De eerste stap is het (automatisch) binnenhalen van de data uit online discussiefora, zodat ze offline kunnen worden opgeslagen en geanalyseerd. Omdat de data bestaat uit privacy gevoelige informatie zullen we daar zorgvuldig mee omgaan. Experts op het gebied van privacy zullen gevraagd worden om advies. Mocht het toch niet lukken om de data van kanker.nl te gebruiken, dan zullen we data verzamelen van andere publieke webfora en open groepen op Facebook.



**Figuur 1: Modulaire architectuur van de beoogde aanpak.**

**T2.2 Forum parsing:** de gedownloade data wordt vervolgens geparsed. Dit houdt in dat de structuur van het forum wordt herkend en de relevante content (titels en teksten van berichten, tijdsaanduidingen, forumcategorieën, etc.) uit de HTML pagina's wordt geëxtraheerd.

### T2.3 - T2.6 Text mining & analyse

**T2.3 Automatisch samenvatten:** Individuele discussiethreads worden met geavanceerde tekstanalysetechnologie samengevat, zodat de essentie, en daarmee de relevantie voor een gebruiker van de discussie in een oogopslag kan worden beoordeeld. Deze activiteit is nauw gerelateerd aan de activiteiten in het DISCOSUMO project. In dit project van de Radboud Universiteit Nijmegen wordt een toolkit ontwikkeld om automatisch discussie forum threads samen te vatten die taal en domein onafhankelijk is. Er wordt gebruik gemaakt van deep learning technieken, die ook voor andere taken van belang zijn. De Nederlandse forum teksten die in dit project geanalyseerd worden, zijn een goede testcase voor de technologie.

**T2.4 Deep learning:** Een nieuwe stroming binnen het veld van de kunstmatige intelligentie, deep learning genaamd, maakt het mogelijk om automatisch inzichten en relevante kenmerken te ontdekken in grote hoeveelheden ongestructureerde data. Toegepast op tekst betekent dit dat bijvoorbeeld op basis van berichtcontexten semantisch gelijksoortige termen en berichten kunnen worden gevonden, zonder dat ze daarvoor dezelfde woorden hoeven te bevatten. Op deze manier kunnen automatisch

spellingvarianten en synoniemen worden ontdekt, maar ook andere instanties van hetzelfde type (bijvoorbeeld bijwerkingen, lichaamsdelen, genen, medicijnen, etc.).

T2.5 Classificatie: Om een onderscheid te maken tussen verschillende type berichten worden ze ook automatisch geclassificeerd. Zo betreft een aanzienlijk deel van de reacties op patiëntenfora een steunbetuiging of gelukwens. Deze berichtcategorie bevat over het algemeen nauwelijks medisch of praktisch relevante informatie en kan op deze manier een lager gewicht worden toegekend.

T2.6 Conceptextractie: Deze kennisgedreven module extraheert de biomedische entiteiten die in de berichten worden genoemd, op basis van publiek beschikbare medische thesauri. De module extraheert o.a. fysische processen, ziektes, lichaamsdelen, gennamen, medicijnen, bijwerkingen, en voedingsmiddelen.

## **T2.7 Indexering & opslag**

Semantisch model: Het semantisch model wordt gegenereerd door de deep learning module en wordt gebruikt om woorden en berichten met een gelijke betekenis/inhoud terug te vinden.

Zoekindex: De originele berichten en hun meta-informatie, evenals de verrijking die door de text mining modules is toegevoegd (classificaties, concepten) worden geïndexeerd en opgeslagen in een schaalbare zoekindex, zodat de gebruiker is verzekerd van een razendsnelle zoekrespons.

## **T2.8 Gebruikersinterface**

De beoogde gebruikersinterface is zeer intuïtief en makkelijk in het gebruik. Het idee is dat een persoonlijk filter wordt gegenereerd op de data, op basis van de zoekvragen en interactie en het profiel van de gebruiker (patiënt of medisch expert). In eerste instantie zoekt de gebruiker met een of meerdere zoektermen naar berichten. Op de achtergrond wordt deze zoekvraag verrijkt via het semantische model van de data. Hierdoor worden ook berichten gevonden die niet de ingevoerde term bevatten, maar wel relevant kunnen zijn voor de zoekvraag. Het resultaat van de zoekopdracht wordt real-time gevisualiseerd. De visualisatie toont de belangrijkste biomedische entiteiten die in de berichten zijn gevonden door de conceptextractiemodule, en hun onderlinge relaties. Via de visualisatie kan de ordening en omvang van het zoekresultaat direct worden beïnvloed; als het ware wordt zo ingezoomd op specifieke informatie waarin de gebruiker is geïnteresseerd. Door middel van feedback kunnen de gebruikers domeinkennis toevoegen aan het model, bijvoorbeeld nieuwe medicijnnamen die nog niet herkend worden door de conceptextractie.

## **Communicatieplan**

Door de deelname van de patiëntenorganisaties Leven met Kanker en de GIST contactgroep zijn we verzekerd van de goede lijnen om in contact te komen met patiënten en medisch professionals om deze bij het project te kunnen betrekken. Daarnaast hebben we met de partner Kanker.nl toegang tot een veelbezochte site voor kankerpatiënten. Via deze kanalen kunnen we een groot gedeelte van de beoogde eindgebruikers bereiken. Via een project website en Twitter zullen we regelmatig nieuws en resultaten publiekelijk delen.

De resultaten zullen gepresenteerd worden in een publiek eindrapport en door middel van een workshop open voor alle geïnteresseerden. Daarnaast willen we op enkele (inter)nationale conferenties de resultaten presenteren zowel medische conferenties als IT-gerelateerde conferenties.

In aanvulling op de beschrijving van het plan voor de realisatie in de projectaanvraag is een schematisch overzicht van de projectplanning te vinden in Figuur 2.

Activiteit	Deelnemers	Maand									
		1	2	3	4	5	6	7	8	9	10
<b>WP 1: Project Management</b>											
T1.1 Kick-off	Allen	■									
T1.2 Rapportage	Allen									■	■
<b>WP 2: Ontwikkeling</b>											
T2.1 Data Harvesting	TNO, RUN, Kanker.nl		■	■	■						
T2.2 Forum Parsing	TNO, RUN		■	■	■	■					
T2.3 Samenvattingen	RUN			■	■	■	■	■			
T2.4 Deep Learning & Classificatie	TNO, RUN			■	■	■	■	■			
T2.5 Classificatie	TNO			■	■	■	■	■			
T2.6 Concept extractie	TNO, RUN			■	■	■	■	■			
T2.7 Indexering en opslag	TNO				■	■	■	■	■		
T2.8 Gebruikersinterface	TNO, LMK, Kanker.nl					■	■	■	■	■	■
<b>WP 3: Gebruikers Input</b>											
T3.1 Requirements analyse	Allen	■									
T3.2 Inbrengen domeinkennis	LMK, GIST, NKI		■	■	■						
T3.3 Tussentijdse evaluatie	LMK, GIST, NKI						■		■	■	
T3.4 Evaluatie eindproduct	LMK, GIST, NKI						■		■	■	

Figuur 2: Planning

## CV's

Het project wordt uitgevoerd door een aantal partners die in een pilot project gericht op analyse berichten van het GIST forum al hebben samengewerkt. (GIST is een zeldzame kankersoort).

### TNO

dr. **Rianne Kaptein** (projectmanagement, specialist text mining). Rianne Kaptein is in 2011 gepromoveerd op het gebied van Information Retrieval aan de Universiteit van Amsterdam. In haar promotieonderzoek heeft ze gewerkt aan entiteiten ranking, zoeken in Wikipedia, en het genereren van woordenwolken. Daarna heeft zij gewerkt bij Social Media marketing bedrijf Oxyme, onder andere aan het verbeteren van de relevantie van de (sociale media) zoekresultaten en de ontwikkeling van text mining tools om content van sociale media te analyseren. Momenteel werkt ze bij TNO in de Media Mining groep aan projecten op het gebied van text mining, information retrieval en social media analyse in de rol van onderzoeker en/of projectleider.

drs. **Martijn Spitters** (text mining & machine learning specialist). Na zijn studie Taal & Kunstmatige Intelligentie aan de Universiteit van Tilburg trad Martijn in 1999 in dienst van TNO als onderzoeker op het gebied van information retrieval en machine learning. In 2003 stapte Martijn over naar Textkernel BV, waar hij bijna tien jaar werkte als onderzoeker, software ontwikkelaar en consultant op het gebied

van machine learning-gedreven oplossingen voor de human resource markt, en daarmee een substantiële bijdrage leverde aan de groei van Textkernel tot marktleider in haar segment. In 2013 keerde Martijn terug naar TNO, waar hij zich sindsdien bezighoudt met de ontwikkeling en toepassing van innovatieve technieken voor het ontdekken en extraheren van relevante kennis uit (grote hoeveelheden) tekst. Zijn meest recente focus ligt op het toepassen van dergelijke technieken voor de ondersteuning van internet forensics processen.

Prof.dr.ir. **Wessel Kraaij** (Principal Scientist TNO.) Wessel Kraaij behaalde zijn ingenieursdiploma elektrotechniek aan de TU Eindhoven in 1988. Hij werkte achtereenvolgens aan de Universiteit Tilburg (1988-1994) en de Universiteit Utrecht (1994-1995) als onderzoeker taaltechnologie en zoektechnologie (information retrieval). Sinds oktober 1995 is hij in dienst van TNO. In 2004 promoveerde hij aan de faculteit Informatica van de Universiteit Twente. Bij TNO heeft hij een groot aantal onderzoeksprojecten geïnitieerd en geleid op het gebied van de ontsluiting van multimedia en text mining technieken. Sinds 2009 leidt hij een groot nationaal R&D project op het gebied van big data en gezondheid. Daarnaast leidt hij het TNO Early Research Programma 'Making Sense of Big Data'. Wessel Kraaij is sinds 2008 bijzonder hoogleraar 'Information Filtering and Aggregation' aan de Radboud Universiteit Nijmegen.

#### *Radboud Universiteit Nijmegen:*

dr. **Suzan Verberne** (onderzoeker tekst analyse). Suzan Verberne is onderzoeker op het gebied van Text Mining. Het onderwerp van haar promotie, die ze in 2010 afrondde, was de ontwikkeling van zoeksystemen die waarom-vragen kunnen beantwoorden. Sinds 2010 heeft ze gewerkt aan projecten op het gebied van tekstclassificatie, informatie-extractie, zoeksystemen voor specifieke doelgroepen en de interactie tussen mensen en zoeksystemen. In haar huidige werk als onderzoeker aan de Radboud Universiteit houdt ze zich bezig met het extraheren van informatie uit webforums en het automatisch samenvatten van forumdiscussies. Daarnaast geeft ze onderwijs op het gebied van Text Mining en Information Retrieval.

#### *Levenmetkanker*

**Laurence Maes** (teamleider belangenbehartiging). Laurence Maes is Manager Afdeling Belangenbehartiging Patiëntenbeweging Levenmetkanker (voorheen NFK). Als hoofd van de afdeling is ze verantwoordelijk voor een team van projectleiders, beleidsmedewerkers en programmamanagers die samen invulling geven aan steeds betere kwaliteit van zorg en kwaliteit van leven bij kanker. In 2010 was zij betrokken bij de ontwikkel- en opstartfase van KankerNL. Daarnaast is Laurence Maes lid van diverse adviesraden waaronder de adviesgroep ministerie SZW Actieplan Kanker en Werk en de Raad van Advies Academisch Centrum voor Private Verzekeringsgeneeskunde (ACPV) en is zij gastdocent bij de NSPOH opleiding verzekeringsgeneeskunde en arbeidsdeskundigen

Drs. **Jeroen Schreurs** (projectmanagement, consultancy ICT). Jeroen Schreurs is afgestudeerd als bedrijfseconoom aan de Erasmus Universiteit in Rotterdam. Na een korte carrière in de accountancy maakte hij de overstap naar de automatisering. In de afgelopen 15 jaar deed hij brede werkervaring op in verschillende posities en projecten, waarbij de focus steeds lag op ontwikkeling en implementatie van ERP- en webbased applicaties. Zijn sterke eigenschap is dat hij de brug kan slaan tussen de

gebruikerswensen en de inrichting en ontwikkeling van de programmatuur. De laatste 5 jaar is hij werkzaam bij Levenmetkanker en betrokken bij de uitrol van diverse CMS systemen, CRM/ERP applicaties, en het platform [www.kanker.nl](http://www.kanker.nl).

### **Contactgroep GIST**

prof. dr. **Gerard van Oortmerssen** (expert en ervaringsdeskundige). Gerard van Oortmerssen is gepromoveerd in 1976 aan de TU Delft en heeft daarna veel ervaring opgedaan in research met als rode draad scientific computing en ICT. Van 1970 tot 1991 werkte hij bij MARIN, het Maritiem Research Instituut in Wageningen. Hij was directeur van het CWI in Amsterdam (1991 - 2003), TNO ICT (2003 - 2008) en het ICT Regie Orgaan (2008 - 2011). Daarnaast was/is hij deeltijd hoogleraar aan de Universiteit Maastricht (2006 - 2009) en Tilburg University (2009 - heden). In 2009 kreeg hij de diagnose GIST (Gastro Intestinale Stromale Tumor), een zeldzame vorm van kanker, waarna hij actief werd in de nationale en internationale kankerpatiënten beweging. Hij is voorzitter van de Contactgroep GIST, de organisatie voor GIST patiënten in Nederland en België.

drs. **Paul van Kampen** (ervaringsdeskundige). Paul van Kampen is drs. in de sociaal-culturele wetenschappen. Hij is lange tijd in het kader van de ontwikkelingssamenwerking werkzaam geweest als onderwijsspecialist in Latijns-Amerika. Tegenwoordig heeft hij een bedrijf voor de productie van puzzels. In 2007 werd bij hem GIST vastgesteld en sindsdien is hij daarvoor onder medische behandeling. Binnen de Contactgroep GIST is hij actief als vrijwilliger informatievoorziening medische zaken.

### **KankerNL**

Drs. **Mies van Eenbergen** (Hoofdredacteur Kanker.nl). Mies van Eenbergen is vanaf ideevorming betrokken bij Kanker.nl en op dit moment in deeltijd hoofdredacteur van Kanker.nl. Daarnaast werkt ze voor 16 uur als onderzoeker bij IKNL. Zij is afgestudeerd bij Letteren (digitale media). Haar afstudeeronderzoek ging over het gebruik van internet door mensen met kanker. Haar onderzoeksonderwerp is Kanker en Internet, en dan meer specifiek in hoeverre maken patiënten gebruik van internet en hoe kunnen we patiënten beter gaan bedienen via internet.

**Peter Heine** (Community Manager bij Kanker.nl). Vanaf 2012 communitymanager bij Kanker.nl, platform met informatie over kanker en ontmoetingsplek voor iedereen die geraakt is door kanker. Daarvoor was hij landelijk communitymanager van het sociale netwerk Deeljezorg.nl; vanaf 1 januari 2011 tevens landelijk coördinator van de gezamenlijke e-Panels van 11 Zorgbelangorganisaties. Ruim 10 jaar werkt Peter Heine al met en voor patiënten. Voor Kanker.nl was dat bij het RPCP (Regionaal Patiënten/Consumenten Platform) Midden-Brabant.

### **NKI/ALV (Nederlands Kanker Instituut / Antoni van Leeuwenhoek ziekenhuis)**

drs. **Sheima Farag** (Medisch expert). Sheima Farag is promovendus Medische Oncologie bij het Nederlands Kanker Instituut (NKI)/ Antoni van Leeuwenhoek ziekenhuis. Het doel van haar promotietraject is de verbetering van diagnostiek en behandeling van GIST en het opzetten van Nederlandse GIST registratie. Ze heeft een studie Geneeskunde afgerond aan de Universiteit van Amsterdam. Voor de Contactgroep GIST is zij medisch adviseur.

## **Bijlagen**

In de overige bijlagen hebben we de volgende documenten toegevoegd:

De AWTI (Adviesraad voor Wetenschap, Technologie en Innovatie) heeft recentelijk een advies over ICT uitgebracht aan de ministers Bussemaker en Kamp. Hierin staat ook een hoofdstuk over Big Data en het pilot project dat we momenteel uitvoeren rond GIST wordt uitvoerig als voorbeeld genoemd. Zie bladzijden 49 en 50 in het advies [Klaar-voor-de-Toekomst.pdf](#)