



Universiteit  
Leiden  
The Netherlands

## **Knowledge discovery from patient forums: gaining novel medical insights from patient experiences**

Dirkson, A.R.

### **Citation**

Dirkson, A. R. (2022, December 6). *Knowledge discovery from patient forums: gaining novel medical insights from patient experiences*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/3492655>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3492655>

**Note:** To cite this publication please use the final published version (if applicable).

# SAMENVATTING

Patiënten delen waardevolle adviezen en ervaringen met hun lotgenoten op online patiëntenfora. Deze ongecensureerde ervaringen kunnen een aanvullend perspectief bieden op dat van de zorgprofessional en nieuwe hypothesen opleveren die vervolgens gevalideerd kunnen worden met medisch onderzoek. Deze dissertatie richt zich op het ontwikkelen van automatische methoden om patiëntenervaringen uit online patiëntenfora te extraheren met behulp van text mining technieken. We onderzoeken de complementaire waarde van deze kennis voor het genereren van wetenschappelijke hypothesen ten opzichte van traditionele bronnen van medische kennis. In het bijzonder richten wij ons op de extractie van bijwerkingen en copingstrategieën die patiënten gebruiken om om te gaan met hun bijwerkingen.

In het eerste deel van deze dissertatie onderzochten we hoe spelfouten in berichten op het patiëntenforum verminderd kunnen worden. We onderzochten welke unsupervised spellingcorrectiemethode het meest geschikt is voor het corrigeren van spelfouten, zonder waardevolle informatie te verliezen door valse positieven (domein-specifieke termen of termen van leken die onterecht gecorrigeerd worden omdat ze niet in het woordenboek staan). We onderzochten ook hoe we berichten met patiëntenervaringen het best kunnen onderscheiden van berichten zonder om zo irrelevante berichten uit de data te kunnen filteren. Bovendien konden we hiermee inzicht verkrijgen in de verschillende soorten ervaringen die patiënten op het forum delen. In een andere studie lieten we zien dat het meenemen van de structuur van discussiedraad tekstclassificatiemodellen niet helpt om relevante berichten te identificeren, ondanks dat relevante berichten vaak gegroepeerd zijn.

In het tweede deel gingen we in op de uitdagingen die ontstaan bij het extraheren van tekstfragmenten waar bijwerkingen in staan. We toetsten hoe goed Transformer-modellen, waaronder het populaire BERT model, werken voor deze taak. Daarnaast toetsten we de kwetsbaarheid van deze modellen voor variatie in de data: hoe makkelijk kunnen ze voor de gek gehouden worden? We presenteerden daarnaast een eenvoudige representatie voor discontinue of onderbroken concepten (bv. “pijn in handen en voeten” wat bestaat uit zowel “pijn in handen” als “pijn in voeten”) die de extractie van bijwerkingen ten goede komt.

In het derde deel introduceerden we een nieuwe taak: de extractie van coping mechanismen voor bijwerkingen van geneesmiddelen. Hoe gaan patiënten om met hun bijwerkingen en welk advies geven ze aan elkaar? We presenteerden de eerste ontologie voor coping mechanismen, vergeleken verschillende mogelijke conceptualisaties van deze taak en lieten zien dat automatisch geëxtraheerde coping strategieën uit berichten van een patiëntenforum gebruikt kunnen worden om nieuwe hypothesen te genereren.

In het vierde en laatste deel beschreven we een casus: een specifiek patiëntenforum voor Gastro-Intestinal Stromal Tumor (GIST) patiënten. We onderzochten de waarde van automatisch geëxtraheerde bijwerkingen uit dit patiëntenforum voor het monitoren van

geneesmiddelen die al op de markt zijn (ook wel post-market surveillance genoemd). We lieten zien dat onze methode niet eerder gevonden bijwerkingen en langetermijnbijwerkingen kan ontdekken. Ook kan onze methode een indicatie geven van welke bijwerkingen het meest belangrijk zijn voor de patiënt zelf. Een vergelijking van deze resultaten met de bijwerkingen die door GIST-patiënten in een enquête worden gemeld liet zien dat automatisch geëxtraheerde bijwerkingen uit patiëntenfora gebruikt kunnen worden om de meest geschikte vragenlijst te selecteren en om vragenlijsten up-to-date te houden. Tot slot hebben we onderzocht hoe representatief de online patiëntenpopulatie van GIST-patiënten is om de beperkingen van kennis uit patiëntenfora beter te overzien. Onze conclusie was dat patiënten die relatief gezonder zijn minder goed vertegenwoordigd zijn op het patiëntenforum.

Ons werk biedt aanknopingspunten voor het ontsluiten van de waardevolle kennis die gedeeld wordt op online patiëntenfora en het gebruik ervan als aanvullende informatiebron voor het genereren van hypothesen. Toekomstig werk zal moeten uitwijzen in hoeverre de complementaire waarde van patiëntenkennis verschilt voor verschillende aandoeningen, waaronder tussen zeldzame en veelvoorkomende aandoeningen en tussen acute en chronische aandoeningen.