Harthulp

a smart question-and-answer platform for cardiac patients

Master Thesis

Ward Hendrix

Project chair
Dr. ir. Maaike Kleinsmann

Project mentor
Jiwon Jung, MSc

Final version

Delft, April 2019

Cardio
Lab

“Harthulp”: a smart question-and-answer platform for cardiac patients

Ward Hendrix

Master Thesis Delft University of Technology - Faculty of Industrial Design
Master in Integrated Product Design
Delft Design Labs - Cardiolab

Project chair
Dr. ir. Maaike Kleinsmann

Project mentor
Jiwon Jung, MSc

Delft, April 2019

This Master Thesis is confidential. No part of this document may be shared or reproduced in any form without written permission of Cardiolab.

Contents

Abstract

In common Dutch cardiac care, patients only have few follow-up meetings with their cardiologists after they have been treated at the hospital. Sometimes, they have to wait several months for their next visit. Therefore, patients often turn to online platforms where they can ask questions to other patients or healthcare professionals, namely *health-based social networks*. These platforms are managed by a small group of volunteers, cardiologists, who seek to inform patients in order to make efficient use of the limited time at the hospital. However, it is expected that the number of cardiac patients on these networks will increase due to overall growth of the number of Dutch cardiac patients and the recent trend of searching health information on the internet. Hence, it remains the question whether health-based social networks can scale up with this development.

To investigate the sustainability of health-based social networks, user posts from the Dutch social platform Hart Volgers and American social platform DailyStrength were analysed. The results show that the number of patients on these platforms grows rapidly, and they reveal that users prefer to contact cardiologists instead of fellow patients. These findings suggest that there is an urgent need for new solutions that keep these platforms future-proof. To address this problem, Harthulp is proposed, a smart question-and-answer platform for cardiac patients to provide better and more efficient aftercare. Harthulp introduces a question wizard that enables patients to quickly find information on the platform, so that it is not always necessary to ask a question to a healthcare professional and to wait for a reply.

As a core component of the question wizard, a novel search engine has been developed which employs a deep learning model that captures the semantics of words on health-based social networks. In this way, patients can search with short questions and retrieve relevant posts while they may not contain the exact same words. It has been demonstrated that the proposed search engine significantly outperforms traditional search engines when retrieving relevant question-and-answer posts from Hart Volgers. A web interface has been designed to show how all components can be embedded in a single user-centered design. This design has been evaluated together with an experienced cardiologist.

Acknowledgements

First of all, I would like to thank my project chair Maaike Kleinsmann for the opportunity to graduate at CardioLab. Since the very beginning of the graduation project, she completely put her trust into my capabilities as an Industrial Designer. Her constructive feedback during the evaluation sessions greatly contributed to the academic value of my master thesis.

I owe many thanks to my project mentor Jiwon Jung for her excellent support and feedback during the graduation project. Every week, she took the time to elaborately discuss my project, regardless of her own tight time schedule as a PhD-student. She was always available to listen to me and to help me out.

I want to express my gratitude to Stijn de Ridder, who shared his insights as a cardiologist and gave me great feedback on my design. The sessions with him gave me a lot of inspiration during the project.

I also want to thank Alexander Hilt and Bas Biersteker at LUMC, David Smeekes at the Dutch Heart Foundation, and cardiologist Natasja de Groot. Thanks to them, I better understand the needs of cardiac patients and healthcare professionals.

I would like to thank my friends at the faculty of Industrial Design for their support and inspiring ideas during the brainstorm workshop. I want to give my thanks to the students who have helped me with labelling the dataset for the experiments.

I wish to thank my family for their love and encouragement.

In particular, I would like to express a special feeling of gratitude to my twin brother Nils. A year ago, it was unimaginable that I would make it this far into the interesting, but difficult domains of Data Science and Deep Learning. I could not have accomplished this without his support.

Chapter 1
General introduction

The research described in this thesis was conducted at Cardiolab, which is a collaboration between Philips Design, Dutch Heart Foundation and TU Delft. The aim of Cardiolab is to reduce the burden of cardiovascular diseases using smart technologies: from both an individual's perspective (patients and medical doctors) and societal perspective (the Dutch healthcare system). Hence, a smart product-*service system will be proposed that fits the vision of Cardiolab. In this chapter, the central problem of this thesis is be first introduced (Section 1.1). Next, it is be discussed how the current healthcare system undergoes a paradigm shift from patient treatment to self-management (Section 1.2), and what part health-based social networks and smart technology have in this process (Section 1.3). Finally, the main research question and the sub-research questions addressed in this thesis are be presented (Section 1.4).*

1.1 Problem statement

1.1.1 Growth of cardiac patients in the Netherlands

Cardiovascular diseases (CVD) is one of the most common causes of death in the Netherlands (de Boer et al., 2018). The cardiovascular registries of the Dutch Heart Foundation show that 38.119 people died from CVD in 2017. Although the mortality from CVD has decreased over the last decades thanks to preventive measures against risk factors (such as smoking) and new medicine, the disease burden remains high: 730 Dutch people are hospitalised everyday due to CVD (Hartstichting, 2019a). It is expected that this burden will only increase in the near future, because the current population of 1,4 million Dutch cardiac patients is rapidly growing.

Two major factors can be identified that contribute to this growth. The first factor is that the average life expectancy of people is increasing[1] and thus there are more people who reach the age where they have an increased risk of developing CVD (de Boer et al., 2018). The second factor is the ageing population in the Netherlands, which is the result of the high childbirth in the sixties but relatively low childbirth today (PBL, 2013; van

[1] According Statistics Netherlands (CBS), the life expectancy at birth of both Dutch men and women has increased by approximately 10 years over the last 50 years (Volksgezondheidenzorg.info, 2017). CBS forecasts a life expectancy of 86,5 years for men and 89,9 years for women in 2060: an increase of approximately 7 years compared to 2017.

Oostrom et al., 2017). These developments emphasis the continuous need for research on prevention, diagnosis and treatment of CVD to unburden the Dutch healthcare system.

1.1.2 Limited aftercare and reach of cardiac rehabilitation

Standard cardiac care in the Netherlands consists of four phases (Achttien et al., 2011): preoperative phase, hospitalisation (phase I), outpatient cardiac rehabilitation (phase II) and post-cardiac rehabilitation (phase III). Cardiac rehabilitation is a multidisciplinary treatment process that is focused on the physical, mental and social recovery of cardiac patients. The most important aspect of cardiac rehabilitation is *secondary prevention*, which involves the reduction of high-risk behaviour and risk factors (e.g. smoking). This way, the development of new complications slows down and recurring heart failures are prevented. After the cardiac rehabilitation, the focus shifts towards *tertiary prevention* that involves long-term management of CVD to optimize the patient's quality of life (Institute for Work and Health, 2015). An overview of standard cardiac care is depicted in Figure 1.1.

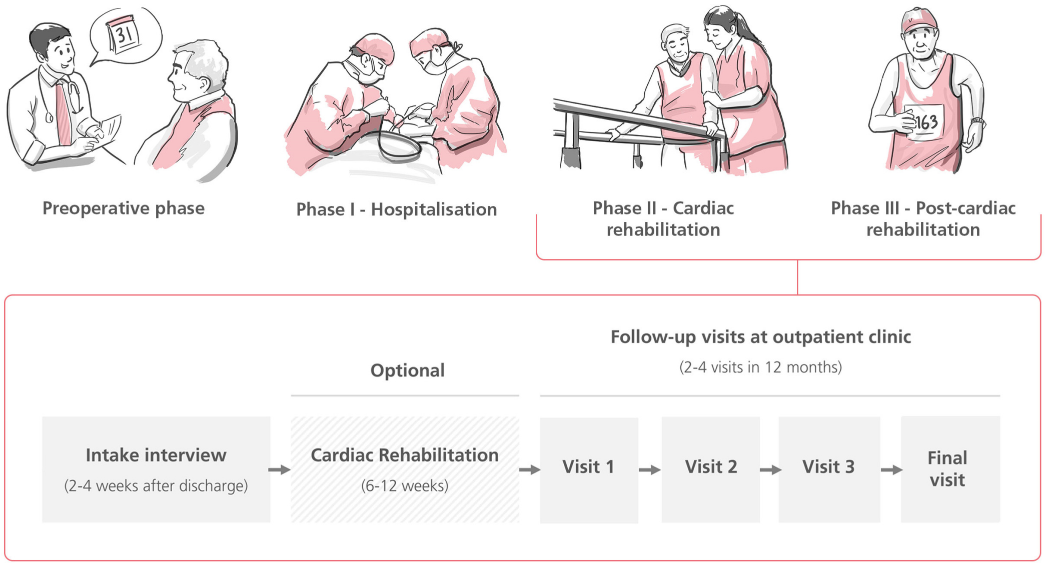

31
163
Preoperative phase
Phase I - Hospitalisation
Phase II - Cardiac rehabilitation
Phase III - Post-cardiac rehabilitation
Follow-up visits at outpatient clinic
(2-4 visits in 12 months)
Optional
Intake interview
(2-4 weeks after discharge)
Cardiac Rehabilitation
(6-12 weeks)
Visit 1
Visit 2
Visit 3
Final visit

Figure 1.1: Overview of standard cardiac care in the Netherlands.

When taking a closer look at Figure 1, one may notice that the current aftercare for cardiac patients is fairly limited. Cardiac patients may undergo an extensive cardiac rehabilitation program of six or even twelve weeks (based on the outcome of an intake interview), but afterwards there are only two or four follow-up visits at the outpatient clinic in the hospital with three or six months in between (Snaterse, 2018). It seems that hospitals can only offer little support for those who face problems or have questions in the meantime, and cardiac patients are therefore appointed to their general practitioner for CVD-related questions. Even when patients are able to speak a medical doctor, proper communication is hindered by time pressure and the prevalence of protocols (Voormolen, 2013).

Another limitation of the current aftercare is related to cardiac rehabilitation itself: cardiac rehabilitation programs in the Netherlands do not reach all cardiac patients. In Figure 1.2, it can be seen that less than half of the cardiac patients are offered cardiac rehabilitation and one-third of these patients reject the offer (e.g. due to the distance to the hospital[2]). Moreover, one-third of the patients who accepted the offer do not finish the cardiac rehabilitation program (Jonkers, 2018). Altogether, it could be assumed that there is a significantly large group of patients who still have questions and uncertainties about managing their condition after they left the hospital.

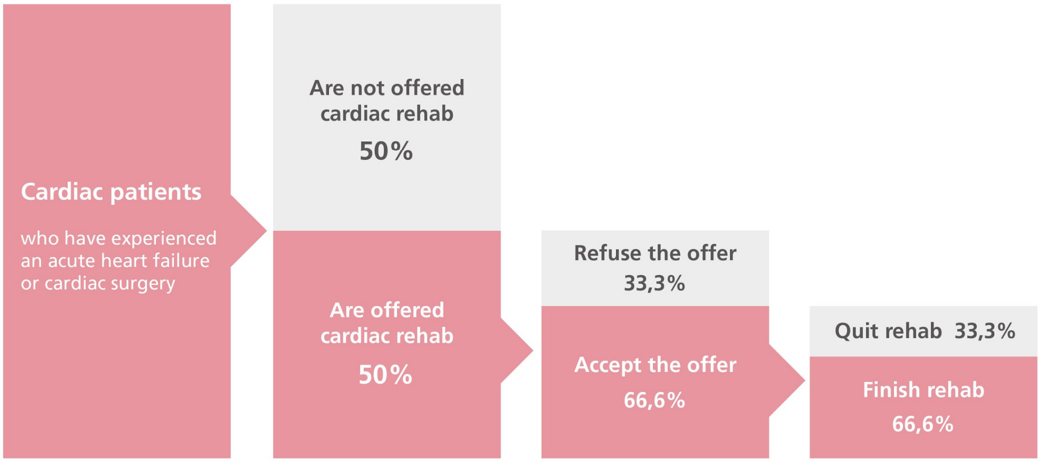


Figure 1.2: Dropout rates of cardiac patients who experienced an acute heart failure or cardiac surgery.

1.2 Self-management paradigm shift in healthcare

The ever-increasing population of Dutch cardiac patients and the limited time at the hospital necessitates that patients better inform themselves before a doctor's appointment and assume greater responsibility for their own treatment. These are important aspects of *self-management*: a fundamental shift of responsibility whereby the relationship between healthcare professionals and patients becomes more symmetric. Dubberly, Mehta, Evenson, and Pangaro (2010) observes a paradigm shift in healthcare whereby traditional healthcare is extended by self-management strategies, as summarized

[2] The findings of a European international study explain why cardiac patients reject or quit cardiac rehabilitation (De Vos, 2013): the most common reasons to refuse cardiac rehabilitation are that patients do not have time, think they can solve their issues themselves or indicate that the hospital is too far away. The most common reasons to stop with cardiac rehabilitation are that patients suffer from physical problems, think they do not need help anymore or believe that the costs are too high. However, the study of De Vos and the news article of Jonkers (2018) provide no explanation why a large group of patients are not offered cardiac rehabilitation.

in Figure 1.3. A chronic illness is no longer treated by prescribed medicine alone, but also by lifestyle interventions to maintain well-being and improve quality of life.

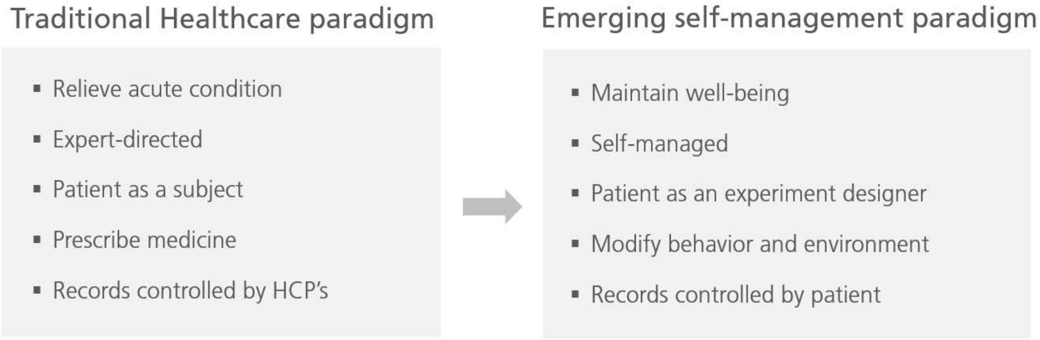

Traditional Healthcare paradigm
Relieve acute condition
Expert-directed
Patient as a subject
Prescribe medicine
Records controlled by HCP's
Emerging self-management paradigm
Maintain well-being
Self-managed
Patient as an experiment designer
Modify behavior and environment
Records controlled by patient

Figure 1.3: Paradigm shift in healthcare. *HCP* stands for healthcare professionals. Adapted from "Reframing health to embrace design of our own well-being," by H. Dubberly, R. Mehta, S. Evenson, and P. Pangaro, 2010, *Interactions*, *17*, p. 3.

Another factor that contributes to the emerging self-management paradigm is the arrival of the internet. Before that time, it was easier for a doctor to have authority over a patient due to the large knowledge gap between them and the low accessibility of information (compared to the status quo). Nowadays, patients can better inform themselves by reading medical articles on the internet before visiting the doctor. A recent study of the Leiden University Medical Center (LUMC) demonstrates that this behaviour is quite common: 4.500 Dutch participants were asked about their health-information seeking behaviour and 80% of the participants stated that they primarily use the internet (Bos, 2018). Among those who use the internet, 60% also looks for a solution that they can directly apply by themselves. This form of online self-management will probably become ever more omnipresent in the near future, as the younger generation are significantly more engaged in this behaviour than the older generation; according to an interview with dr. Lukas Dekker, cardiologist and cofounder of the Dutch social platform Hart Volgers, 55% of persons between 45 and 65 years old look up medical information on the internet and 90% of people between 18 and 24 years old (Van Bergen, 2014).

The changing, self-managing attitude of patients have led to an enormous collection of e-*health* applications on the internet. E-health is a term that describes all information and communication technologies that support healthcare (Loohuis & Chavannes, 2017). In 2016, approximately 259.000 health apps were available on the three largest app platforms (IOS, Android and Windows), which constitutes an increase of 100.000 apps compared to the previous year (Research2guidance, 2016). A downside to this trend is that most of these health applications have never been clinically validated (Bos, 2018). Hence, it can be observed that healthcare professionals at hospitals and large health foundations have started to develop their own e-health solutions that meet the medical standards (Konings et al., 2018).

An example of professional e-health applications is the telemonitoring application, which enables cardiac patients to do their own health measurements at home (e.g. measuring blood pressure) and send their data to the hospital for medical screening.

State-of-the-art applications are cVitals at UMC Utrecht and The Box at Hart Long centre Leiden. Another example of professional e-health applications is an (online) helpline, which is usually offered by large, independent health foundations (e.g. the Dutch Heart Foundation), hospitals or general practitioners. Helplines enable patients (or their family or friends) to call, chat or email a healthcare professional to get personal advices about cardiovascular diseases.

1.3 Health-based social networks

Besides telemonitoring applications and helplines, a third example of professional e-health applications is the *health-based social network*. Health-based social networks are online platforms where patients can post questions, share experiences and receive community replies. Patients can also directly ask questions to healthcare professionals, who secure the integrity of information on the platform. A well-known Dutch platform is Hart Volgers that is supported by cardiologists from Catherina Hospital Eindhoven. Another Dutch platform is the Atrial Fibrillation Innovation Platform (AFIB online), which is developed by researchers from medical centres VUmc (Amsterdam) and EMC (Rotterdam).

A major advantage of health-based social networks in comparison to other e-health applications is that they preserve useful, disease-specific information that has been posted throughout the years. All of this information is evaluated by medical experts and is accessible for all cardiac patients on the internet. By providing large volumes of high quality and widely accessible medical information, these social networks have a high potential to promote self-management of Dutch cardiac patients as a means of unburdening the Dutch healthcare system. However, the operation of a health-based social network requires time and resources in order to secure the integrity of information and maintain patient engagement. Optimizing the platform management and accessibility of information will become increasingly important, because it is expected that the number of cardiac patients on these networks will increase due to the overall growth of the number of Dutch cardiac patients and their evolving online health-information seeking behaviour. Hence, it is remains the question whether e-health applications can scale up with the ever increasing burden on the Dutch healthcare system. Therefore, the central problem in this thesis is how to keep health-based social networks future-proof.

1.4 Research questions

In the previous section, it has been argued that health-based social networks are a promising solution that helps doctors to cope with the rapidly growing population of Dutch cardiac patients. Moreover, it fits the contemporary health-information seeking behaviour of patients and helps them to better prepare for the infrequent follow-up meetings at the hospital. And yet, it is suggested that there is an urgent need for new solutions that keep these platforms future-proof. Therefore, the main research question reads as follows:

MQ: How can smart technology be used to cope with the increasingly large group of cardiac patients on health-based social networks with only a small group of healthcare professionals available?

To address the main research question, the following sub-research questions will be investigated:

RQ1: What impact has the growth of the number of cardiac patients and their evolving online health-information seeking behaviour on the sustainability of current health-based social networks?

RQ2: What are the user needs of cardiac patients on health-based social networks?

RQ3: What smart technologies can be used to optimize information retrieval in large-scale medical text data?

RQ4: What is the performance of state-of-the-art information retrieval systems on question-and-answer data from health-based social networks?

RQ5: How can health-based social networks be best designed to enable the integration of smart technologies?

A conceptual framework of a smart product-service system that fits the ambition of Cardiolab will be designed. In order to validate the identified problem of possible rapid user growth on health-based social networks, a data analysis will be conducted of two popular health-based social networks, namely the Dutch platform Hart Volgers and the American platform DailyStrength. This way, also the online behaviour of cardiac patients and healthcare professionals can be determined in order to better understand their needs. Then, a study will be conducted to determine how smart technologies can be implemented in the current design context of cardiac healthcare to improve information retrieval in large-scale medical data. The outcomes of both studies will be synthesized into a user-centered design, which will be evaluated together with an experienced cardiologist. The original graduation project brief can be found in Appendix A.

The rest of the thesis is organized as follows. Chapter 2 explains in more detail what health-based social networks are and how they can be analysed. The method and results of the data analysis are then presented and discussed. Chapter 3 explains what smart technologies are available for addressing the central problem in the thesis, and how they can be best implemented. Based on the results of the data analysis in Chapter 2, the scope of smart technologies is reduced to search and recommendation systems only. Chapter 4 presents the design proposal that integrates the research findings of the studies in Chapter 2 and 3. Finally, Chapter 5 provides a general discussion of the proposed design and concludes the thesis.

Chapter 2
Data analysis of health-based social networks

Before this research, the design project started with an open-ended assignment to design a new self-management application to improve cardiac aftercare. A meeting with experts at the Leiden University Medical Center (LUMC) steered the project towards health-based social networks: a decision that has been substantiated in the general introduction of this thesis. This chapter describes the findings of a literature research into health-based social networks and a data analysis of these platforms. The introduction of this chapter provides general information about health-based social networks and motivates the need for a data analysis (Section 2.1). Then it is explained how online posts from the platforms Hart Volgers and DailyStrength were obtained and how these were analysed (Section 2.2). The remaining chapters cover the results of the data analysis (Section 2.3) and the interpretation of the results (Section 2.4).

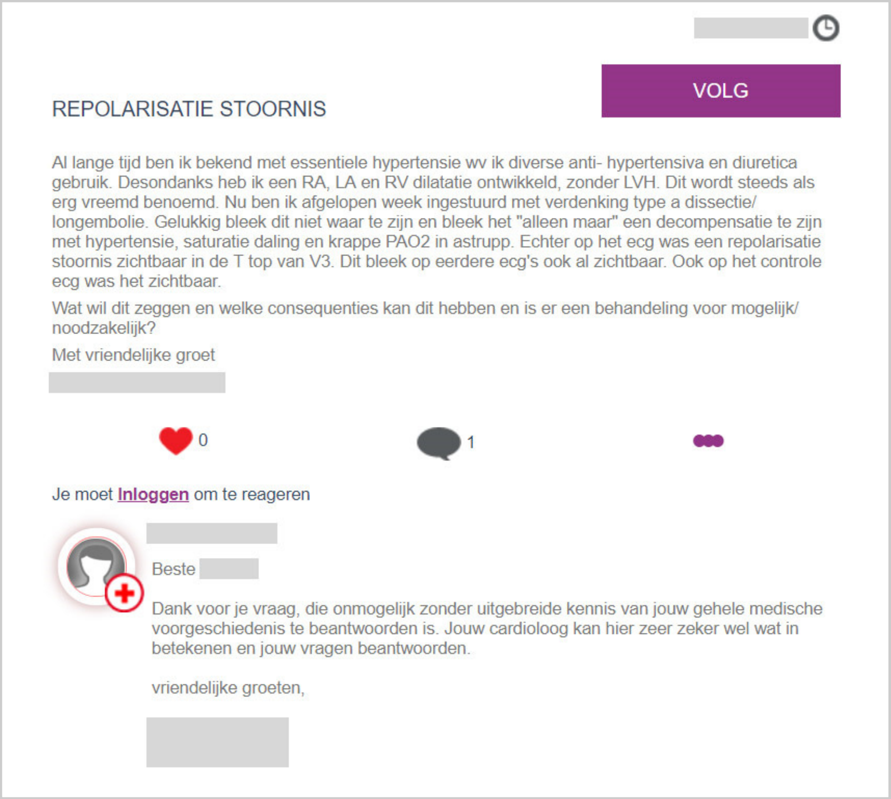
2.1 Introduction

2.1.1 Features of health-based social networks

During the design project, six health-based social networks (both Dutch and foreign ones) were evaluated on the basis of their features, namely (1) DailyStrength, (2) Hart Volgers, (3) WebMD, (4) AFIB Online, (5) NewLifeOutlook, and (6) PatientsLikeMe. An overview of the functionalities and URLs of each platform can be found appendix B. A thorough understanding of the functionalities of these platforms is necessary to develop new ideas during the ideation phase (Chapter 4) and to select platforms for the data analysis (section 2.2). In the next paragraphs, the organisation and common functionalities of health-based social networks will be briefly discussed as a way of introducing the subject matter.

Most social platforms have three basic components: forums, medical articles and search systems. Forums consist of discussion boards where patients can post questions or share experiences with other patients to receive advice or emotional support. An example of a discussion board is shown in Figure 2.1. Discussion boards are often categorized by cardiovascular diseases or stages of these diseases (i.e. diagnosis, treatment or disease management). Medical articles are always posted by healthcare professionals and contain medical information about CVD and their impact on everyday life. Most platforms offer visitors the possibility to search for specific posts on discussion boards by using keywords. The topic of search and recommendations systems will be discussed more elaborately in Chapter 3.

It is important to note that healthcare professionals do not necessarily participate in online conversations. Moreover, they might only check the medical validity of patient posts and make corrections when necessary. For example, healthcare professionals have this role on the American platform DailyStrength. Health-based social networks are therefore different from mainstream social media (such as Facebook), where information is not certified by a medical staff. Dedicated health platforms are especially convenient for healthcare professionals when they provide online consultation, because it is difficult for them to separate personal and professional usage on mainstream social media (KNMG, 2018).



[redacted]

REPOLARISATIE STOORNIS

Al lange tijd ben ik bekend met essentiele hypertensie wv ik diverse anti- hypertensiva en diuretica gebruik. Desondanks heb ik een RA, LA en RV dilatatie ontwikkeld, zonder LVH. Dit wordt steeds als erg vreemd benoemd. Nu ben ik afgelopen week ingestuurd met verdenking type a dissectie/ longembolie. Gelukkig bleek dit niet waar te zijn en bleek het "alleen maar" een decompensatie te zijn met hypertensie, saturatie daling en krappe PAO2 in astrupp. Echter op het ecg was een repolarisatie stoornis zichtbaar in de T top van V3. Dit bleek op eerdere ecg's ook al zichtbaar. Ook op het controle ecg was het zichtbaar.

Wat wil dit zeggen en welke consequenties kan dit hebben en is er een behandeling voor mogelijk/ noodzakelijk?

Met vriendelijke groet

[redacted]

Je moet Inloggen om te reageren

[redacted]

Beste [redacted]

Dank voor je vraag, die onmogelijk zonder uitgebreide kennis van jouw gehele medische voorgeschiedenis te beantwoorden is. Jouw cardioloog kan hier zeer zeker wel wat in betekenen en jouw vragen beantwoorden.

vriendelijke groeten,

[redacted]

Figure 2.1: A screenshot of a discussion board on the platform Hart Volgers. Reprinted from *Hart Volgers* website, 2019, retrieved from https://hart.volgers.org.

2.1.2 Research aim

In the previous chapter, it is suggested that health-based social networks need to be made future-proof. However, little is known about the demographic characteristics of the population of cardiac patients on health-based social networks (e.g. population size or prevalent illnesses), online advice seeking behaviour of cardiac patients, and how these characteristics correlate with the central problem of this thesis. For example, it is essential to determine the online behaviour of cardiac patients: do they mainly seek personal advice from medical doctors, or do they rather seek emotional and social support from other patients? Other demographic characteristics, such as the types of CVD among online users, will also determine the design requirements (Chapter 4).

A method to determine these demographic features is to obtain and store content of health-based social networks into datasets for a *data analysis*. A data analysis may include simple queries and reporting functions (e.g. to get the total number of online users), a statistical analysis or a more complex analysis that require algorithms to process and cluster textual data. In the context of the design project, this means that all online posts of patients and medical doctors should be downloaded from one or more health-based social networks and converted into a workable format. This practice could be considered as *data mining*: the extraction of knowledge from large collections of data (Provost & Fawcett, 2013). The next section covers the data mining method and the materials that are required for such analysis.

2.2 Materials and method

2.2.1 Materials

The programming language Python supports several modules to scrape (i.e. automatically retrieve) content from the internet. For this research, custom Python codes were written with the modules BeautifulSoup and Selenium to scrape posts from the Dutch platform Hart Volgers (7.454 posts) and American platform DailyStrength (65.385 posts). The obtained data from Hart Volgers and DailyStrength was separately stored into two datasets, where each row represents a single post. The features of both datasets are described in Table 2.1.

Table 2.1: Features of the datasets and data types.

Features	**Data type**
Forum category *	String
URL of discussion board	String
Title of discussion board	String
Number of likes of the discussion board	Integer
Name of the poster	String
Whether the poster is a medical doctor or not *(Hart Volgers only)*	Boolean
Timestamp of the post	String
Content of the post	String
Post order	Integer

* In the case of DailyStrength, a maximum of 2000 posts could be obtained per category due to technical issues with the website. Therefore, additional statistics from the website itself were used to make corrections during the data analysis (i.e. the number of users and posts per category).

2.2.2 Method

Scraping

The six health-based social networks as discussed in the introduction of this chapter were considered for the data analysis. The following (technical) criteria were used to select the most suited platforms for the data analysis:

- *Accessibility of webpages for visitors (without an account)*
  Platforms where posts are intended to be only shared within a specific community (i.e. require a user account for online access) were not taken into consideration.

- *Consistency of HTML Markup*
  Web page elements (such as the user posts) should always have the same tag in the source code of the website in order to obtain a clean dataset.

- *Static or dynamic web pages*
  Static means that all information is loaded at once on the webpage. Dynamic means that new information can be requested from the server after a webpage is loaded. It requires more effort to scrape from dynamic web pages than static web pages. The Python module BeautifulSoup can only handle static web pages, in contrast to the module Selenium.

- *Permission to scrape from the website*
  Popular websites often have policy statements about scraping data. This is important, because websites may crash if people request large collections of web pages in a short amount of time. Only if the policy statement explicitly prohibited to scrape from the website, then the website was not taken into consideration.

- *Quality of website content*
  Not every platform offers the same service quality to patients. A manual inspection of the website gives an indication of the quality of the answers given by experts. Websites that only offer standard answers (such as “we cannot help you here, please visit your doctor”) were not taken into consideration.

- *Amount of content on the website*
  Platforms with more user posts than others were prioritized. For this research, it was important that the target audience of the platform sufficiently covered the population of cardiac patients (e.g. websites should not only target cardiac patients with a specific type of condition).

Based on these criteria, the platforms Hart Volgers and DailyStrength were selected for the data analysis. Hart Volgers is one of the largest health-based social networks for cardiac patients in the Netherlands. The American platform DailyStrength was selected alongside Hart Volgers, because the posts of DailyStrength are labelled by type of CVD. The labels would make it easier to identify the different types of cardiac patients on health-based social networks.

Data analysis

The data analysis consisted of two parts: determining the demographic characteristics of the community on health-based social networks, and the online advice seeking behaviour of cardiac patients. The demographic characteristics are defined as follows:

- The number of patient and medical doctor posts over time (Hart Volgers only).
- The number of patients and medical doctors over time.
- The number of members per forum category (DailyStrength only).

As discussed in the introduction, there are no medical doctors on DailyStrength that participate in online conversions and thus only the number of patients and their posts could be obtained. For both DailyStrength and Hart Volgers, all users that are not labelled as medical doctors were considered as patients, although family of cardiac patients or interested persons might be among those users. Furthermore, only users were taken into account who have posted at least a single message. The online advice seeking behaviour of cardiac patients is described by the following features:

- The number of likes per discussion board (Hart Volgers only).
- The number of discussion boards created per patient.
- The number of replies by patients on discussion boards of others.
- The number of discussion boards per forum category.

In the results section, it will be explained that the number of likes per discussion board insufficiently captured the popularity of topics. Therefore, it was decided to conduct an additional cluster analysis to better understand the topics that are discussed among the community members of Hart Volgers. These topics may indicate what type of CVD is most prevalent on Dutch health-based social networks. This information influences how health-based social networks should be managed in the near future (this will be explained in more detail in Chapter 4). The topic clustering analysis consisted of the following steps:

1. Pre-process the posts and create a single text per discussion board by putting posts together that belong to the same discussion board.
2. Calculate TF-IDF scores for each term per discussion board.
3. Group similar discussion boards together through hierarchical agglomerative clustering (later on more about this).
4. Recalculate TF-IDF scores for each term per cluster.

5. Rank terms for each cluster based on TF-IDF score.

The pre-processing steps were (1) punctuation removal, (2) removing letter accents, (3) lower-casing, (4) stop-word removal, (5) lemmatization, and finally (6) tokenization. Stop-removal removes common words such as "the" or "a", and lemmatization is the process of bringing all inflected forms of words back to their basic lemma. Tokenization extracts words as separate features (i.e. tokens) from texts. Lemmatization decreases the dictionary size, and stop-word removal gives more weight to unique words. In the literature, it has been demonstrated that both approaches increase the performance of bag-of-words models (Maalej, et al. 2016). A bag-of-words (BOW) representation of a text means that only word counts are considered as features. In a BOW representation, the text is turned into a fixed-length vector which contains the word counts of the text, and its length is equal to the vocabulary of all unique words in the corpus. Figure 2.2 illustrates how such representation looks like for two example texts.

Text 1: "I like your advice"
Text 2: "I appreciate your advice"

	Advice	**Appreciate**	**I**	**Like**	**Your**
Text 1	1	0	1	1	1
Text 2	1	1	1	0	1

Bag-of-Words

Figure 2.2: In a bag-of-words representation, texts are represented by words counts of the entire vocabulary.

An alternative to the BOW representation is to use TF-IDF instead of raw word counts. TF-IDF stands for "term frequency (TF) - inverse document frequency (IDF)", which is a weighting scheme that makes rare words more prominent and effectively ignores common words (Casari & Zheng, 2015). TF-IDF returns normalized word counts (a value between 0 and 1) by dividing each word count by the number of documents in which this word occurs. Thus, the TF-IDF score for a particular word is close to 1 when it occurs very frequently in a certain text, but rarely in all other texts in the corpus. The corpus (i.e. all discussion boards) is represented as a sparse matrix where each cell contains a TF-IDF score for a word in a particular discussion board (or zero when the word does not appear in the discussion board). Such a matrix is very similar to the one that is shown in Figure 2.2.

When all discussion boards are represented with BOW or TF-IDF, they can be clustered so that discussion boards with similar content are placed together. In this research, TF-IDF is used instead of word counts, because TF-IDF is a representation that highlights meaningful words and therefore usually yields better clustering results. After the discussion boards are clustered and labelled, TF-IDF scores will be recalculated for each cluster (i.e. all texts in a cluster are considered as a single document) in order to get the most meaningful words per cluster. Consequently, these words can be used to determine the overall theme of a cluster.

Regarding the clustering process itself, it should be noted that the number of clusters or themes is not known beforehand. Therefore, it is a good strategy to use a flexible clustering method where the number of clusters can be defined afterwards. A clustering method that fits this criterium is *agglomerative clustering*: a procedure that iteratively merges pairs of objects (i.e. discussion boards) that are very close to each other in terms of similarity. In this research, a bottom-up approach was used where the procedure starts with as many clusters as there are discussion boards, but ends with a single, large cluster. After the clustering process, a visual (a *dendrogram,* an example can be found in Appendix C) can be produced that shows the hierarchy of clusters and their distances at each iteration. Based on the dendrogram, one can chose a sensible number of clusters. For more information about this procedure and its distance metrics, please refer to Pathak (2018).

2.3 Results

2.3.1 Demographic characteristics

Figure 2.3 shows a comparison between the number of cardiac patients on Hart Volgers and DailyStrength. Figure 2.4 shows a comparison between the number of cardiac patients and medical doctors (i.e. cardiologists and cardiac surgeons) on Hart Volgers. The number of posts of patients and medical doctors on Hart Volgers over time is shown in Figure 2.5.

2.3.2 Online advice seeking behaviour

Figure 2.6 shows the number of discussion boards per forum category on Hart Volgers. Figure 2.7 shows the number of community boards and community members per forum category on DailyStrength. On the basis of the number of likes per discussion board on Hart Volgers, it was found that 91% of the discussion boards are not liked by the community. Hence, it was decided not to use this feature as a measure for topic popularity: it is probable that users have not liked the vast majority of discussion boards, because it can be inappropriate to ‘like’ the content of those discussion boards (which might deal with severe issues). Regarding the number of discussion boards per patient, it was found that patients normally create no more than two discussion boards on DailyStrength or three discussion boards on Hart Volgers. Histograms and Tukey’s boxplots of the distribution of the number of discussion boards can be found in Appendix C.

Furthermore, it was found that 61% of the patients on Hart Volgers never reply to discussion boards that are started by other patients. This percentage increases to 66%

when only patients are considered who have never joined a discussion board outside the "Ask Us" category (i.e. the category where patients can ask a question to a medical doctor), but the percentage drops to 35% when only patients are considered who have never asked question to a medical doctor. On DailyStrength, also a percentage of 35% was found. Figures of the distribution of the number of replies that patients give on discussion boards can be found in Appendix C.

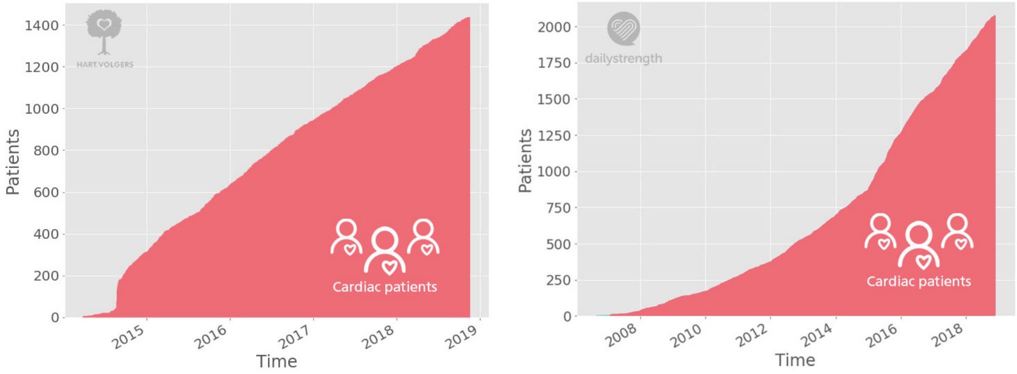


Figure 2.3: The number of cardiac patients on the platform Hart Volgers (left) and DailyStrength (right) over time.

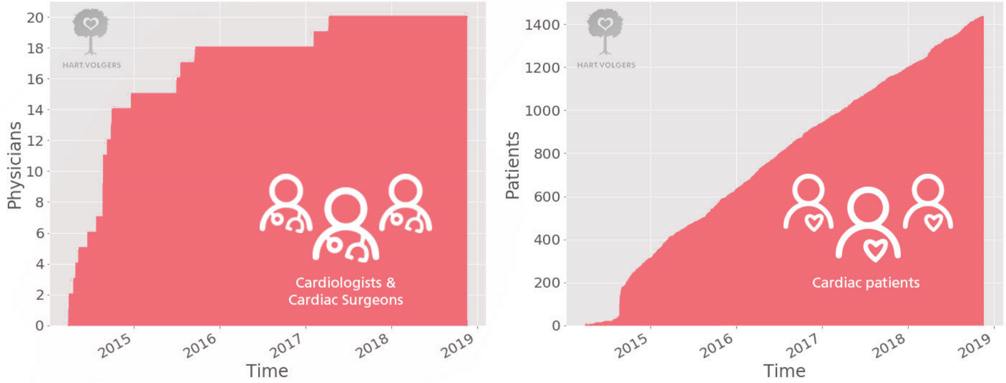
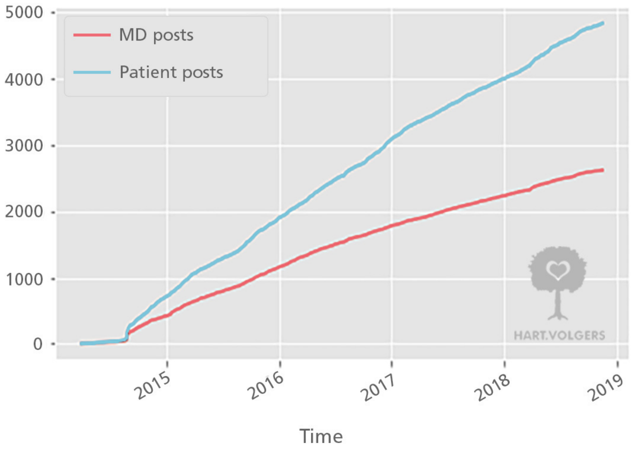


Figure 2.4: The number of medical doctors (left) and cardiac patients (right) on Hart Volgers over time.



Figures 2.5: The number of posts of patients and medical doctors on Hart Volgers over time.

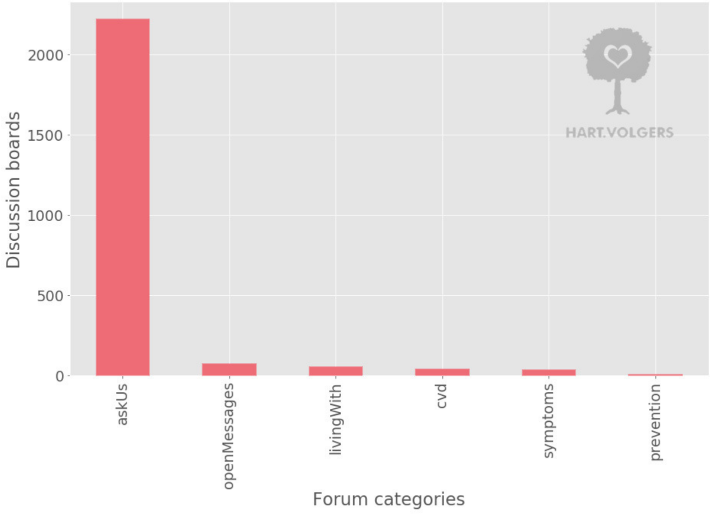


Figure 2.6: The number of discussion boards per forum category on Hart Volgers. *AskUs*: Posts with questions for medical doctors. *OpenMessages*: Posts about people's experiences. No questions are asked to the community. *LivingWith*: Posts about managing cardiovascular diseases. *Symptoms*: Posts about symptoms of cardiovascular diseases. *Cvd*: Posts about the medical nature of cardiovascular diseases. *Prevention*: Posts about prevention of cardiovascular diseases.

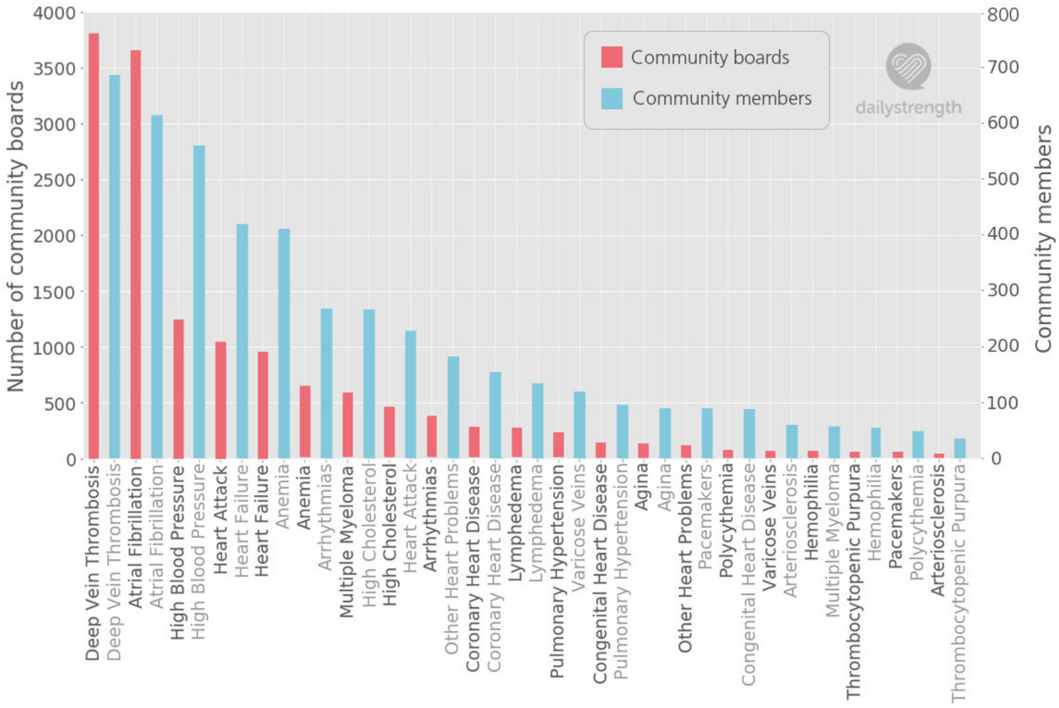


Figure 2.7: The number of community boards and community members per forum category on DailyStrength.

2.3.3 Clustering results

Cluster sizes and the highest ranked words per cluster (based on TF-IDF scores) are shown in Table 2.2. It should be noted that these words are translated from Dutch without the context of the original sentences, so the meaning of the English translations might deviate from the original meaning of the words. Figure 2.8 gives an impression of the clustering quality by means of a t-SNE diagram, which is a dimensionality reduction technique to display a projection of a vectorized corpus into two dimensions (Maaten & Hinton, 2008).

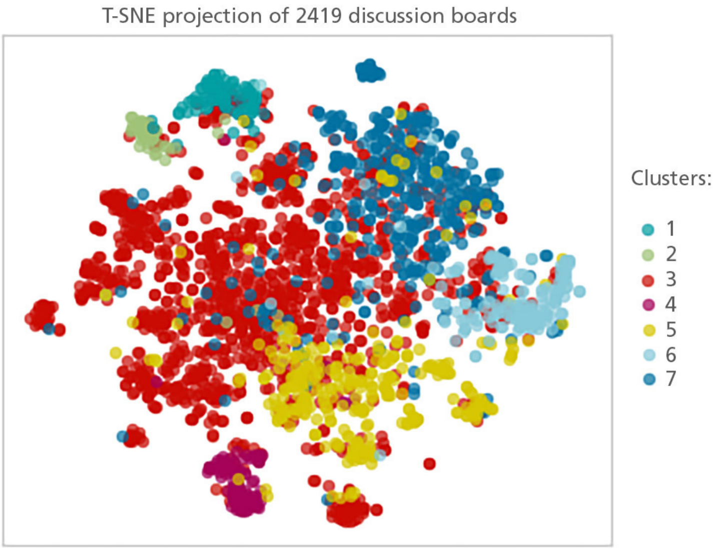


Figure 2.8: A t-SNE diagram of the clusters. The cluster numbers in this plot correspond to those in Table 2.2.

Cluster size	Cluster-id	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
1318	3	*Heart*	*Cardiologist*	*Year*	*Good/ great/ best*	*Question*	*Signs/ Symptoms*	*To get*	*Very*	*Day*	*Pain*
383	7	*Cardiac Ablation*	*Atrial fibrillation*	*Good/ great/ best*	*Question*	*Year*	*Cardiologist*	*Heart*	*Very*	*Heartbeat*	*Once/times*
312	5	*Medicine*	*Cardiologist*	*mg*	*Usage*	*Question*	*Best*	*Side-effects*	*Day*	*Year*	*To get*
161	6	*Heart palpitation*	*Heart*	*Extrasystoles*	*Day*	*Very*	*Wires*	*Cardiologist*	*Burden*	*Good/ great/ best*	*Question*
103	1	*Pacemaker*	*Cardiac ablation*	*Year*	*Good/ great/ best*	*Question*	*Heart*	*Wires*	*Cardiologist*	*Heartbeat*	*To get*
93	4	*Cholesterol*	*Statin*	*LDL*	*Statins*	*Body fat*	*Usage*	*Side-effects*	*Question*	*mg*	*heart*
49	2	*ICD*	*Question*	*Good/ great/ best*	*Wires*	*Year*	*Shock*	*CRT*	*Heart*	*Very*	*To get*

Table 2.2: The top 10 words per cluster of discussion boards on Hart Volgers.

2.4 Discussion

In general, the results indicate that there is a growing pressure on healthcare professionals on health-based social networks. The results reveal that the number of cardiac patients is rapidly increasing on health-based social networks (Figure 2.3), whereas the number of healthcare professionals is not proportionally increasing on Hart Volgers (Figure 2.4). The increase of community members naturally results in an increase of online posts (Figure 2.5). Surprisingly, half of the posts on Hart Volgers are written by medical doctors, which suggests that many discussion boards require the attention of a medical doctor. A further inspection confirms this finding: the results in Figure 2.6 show that 90% of all discussion boards on Hart Volgers are only addressed to medical doctors (i.e. discussion boards in the “Ask Us” category). Therefore, it is questionable if a platform such as Hart Volgers remains manageable in the near future.

Based on the results, it could be argued that cardiac patients prefer contact with a medical doctor rather than a fellow patient on health-based social networks. Especially on Hart Volgers, a large group of patients (61% of all patients) never replies to discussion boards that have been started by other patients. This behaviour could be explained in multiple ways. Firstly, the process of advice exchange within these networks raises issues that are associated with trust, expertise and disclosure. Especially in the specific and complex domain of cardiovascular diseases, patients have to deal with the difficult issue of ‘being an expert’ and it can be hard to indicate their credibility as a layman. Secondly, patients bear great responsibility over their health advices in this context, considering the serious nature of cardiovascular diseases.

Furthermore, the cluster analysis provides more insight into the topics that patients discuss on health-based social networks. Based on Table 2.2, the following themes could be identified based on the obtained clusters:

- Questions related to overall signs and symptoms of CVD (*cluster-id 3, 1318 items*)
- Questions related to atrial fibrillation and treatment (*cluster-id 7, 383 items*)
- Questions related to medicine: usage and side-effects (*cluster-id 5, 312 items*)
- Questions related to extrasystoles (*cluster-id 6, 161 items*)
- Questions related to pacemakers and related treatments (*cluster-id 1, 103 items*)
- Questions related to the cholesterol management (*cluster-id 4, 93 items*)
- Questions related to implantable cardioverter-defibrillators (ICD) and related treatments (*cluster-id 2, 49 items*)

The clusters suggest that most discussions boards are related to medicine, extrasystoles and *atrial fibrillation* (more about this later). The t-SNE plot in Figure 2.8 indicates that the overall clustering quality is reasonably good, but not all clusters are well defined (especially the third cluster is quite scattered). This could explain why the highest ranked words per cluster overlap with other clusters (e.g. “question”, “heart”, etc.). It can be also observed that most clusters in Table 2.2 contain non-informative words (e.g. “good/great/best”, “very” , etc.) despite the use of the TF-IDF weighting scheme.

Nonetheless, these results support the finding that patients prefer asking questions to a cardiologist, and they suggest that health-based social networks are mainly used by people with heart rhythm disorders; words such as “atrial fibrillation”, “extrasystoles”, “pacemakers” and “ICDs” are highly ranked.

The same types of cardiac patients can be found on the platform DailyStrength. Figure 2.7 shows the number of community boards and community members per type of CVD on DailyStrength. It can be observed that *Deep Vein Thrombosis* and *Atrial Fibrillation* are the most prevalent types of CVD on the platform. Deep Vein Thrombosis (DVT) is a condition where a blood clot has formed in a deep vein of the body, usually in the legs (Mayo Clinic, 2019). Atrial Fibrillation (AF) is a condition where the heart is beating irregularly and increases the risks of heart-related complications (American Heart Association, 2016). The clustering of discussion boards of Hart Volgers also indicated that AF is fairly common on health-based social networks. An explanation of these results will be provided in the next paragraph.

During the design project, the cardiologists Prof. Dr. Natasja de Groot and Stijn de Ridder were interviewed to discuss and validate the findings of this research. According to Natasja de Groot, founder of the platform AFIB Online, it is not surprising AF is very common on health-based social networks. It is one of the most common types of CVD in general and AF is a condition that is difficult to cure. Even after a surgical intervention (e.g. cardiac ablation), it is uncertain if AF will ever return again. The symptoms of AF often lead to anxiety among cardiac patients as well. With respect to DVT, it is important to point out that this condition is not treated by the cardiologist in the Netherlands (in contrast with the United States) and is treated by an internist instead. This explains why DVT has not been observed among cardiac patients on Hart Volgers. In the end, both cardiologists supported the finding that platforms such as Hart Volgers may become unmanageable in the near future.

The findings suggest that there is an urgent need for new solutions that keep health-based social networks future-proof. This form of aftercare is especially important for patients who are treated for AF, a condition that is difficult to cure and has long-term complications. From a design perspective, this problem could be addressed in two different ways:

- By reducing the number of questions for healthcare professionals. Potential solutions should provide new ways to quickly find relevant answers and experiences.
- By increasing the capacity of healthcare professionals. Potential solutions should provide new ways to quickly assess and reply to incoming questions.

In the introduction of this chapter, it has been explained that existing health-based social networks have search systems to find questions, answers and experiences that have been posted in the past. However, the next two chapters will highlight the limitations of these search systems (in terms of both technology and usability) and demonstrate how the advancements in *natural language processing* (NLP) can overcome these limitations.

Chapter 3
Search and recommendation systems

In the previous chapter, it has been observed that an increasingly large group of cardiac patients ask their questions to healthcare professionals on the internet, whereas only a small group of healthcare professionals is available. Hence, promising solutions should provide new ways to patients to quickly find relevant information, so that it is not always necessary to ask a question to a healthcare professional and to wait for a reply. Alternatively, they should provide new ways for cardiologists to quickly assess and reply to incoming questions. In this chapter, state-of-the-art technology will be explored that can assist users with finding relevant information: search and recommendation systems. First, it will be explained what search and recommendations systems are, what types of search and recommendation systems can be distinguished, and how these systems can be improved and implemented in the current context of healthcare (Section 3.1). Then, the method (Section 3.2) and experimental set-up (Section 3.3) will be outlined that is used to determine the best information retrieval system for health-based social networks. Finally, the results of the evaluation will be presented (Section 3.4) and discussed (Section 3.5).

3.1 Introduction

3.1.1 Definition of search and recommendation systems

Since the early days of the internet, researchers from information retrieval and related fields have been working on search and recommendation systems. Prior to the internet, it was common practice to manually extract features from the data or to build specific rule-based systems in small domains. For instance, a librarian had to extract the author, title and subject of each book in the library in order to make a catalogue. Nowadays, there is so much online content available that is has become impractical to *manually index* this content in similar way as librarians. Therefore, *search engines* and *recommendation systems* have been built that use advanced techniques to automatically assign identifiers to online content (*automatic indexing*) and to return items that fit the user's information need.

Search engines are information retrieval systems that focus on user-specified requirements, such as a search query (Manning, Raghavan & Shütze, 2010). They satisfy short-term information needs by immediately showing a ranking of documents that meet those requirements. Recommender systems are information retrieval systems that infer the user's interests by learning from past interactions between the user and documents

(Kembellec, Chartron & Saleh, 2014). They satisfy long-term information needs by recommending relevant, unseen documents to the user. Both types of information retrieval systems have unique advantages and disadvantages in the context of health-based social networks, but their mechanisms will be first discussed in further detail.

3.1.2 Types of recommendation systems

Kembellec, Chartron and Saleh (2014) identify four types of recommender systems: (1) content-based recommenders, (2) collaborative recommenders, (3) demographic-based recommenders, and (4) knowledge-based recommenders. An overview of these recommender systems is shown in Figure 3.1. In general, all these recommender systems work with two types of data, namely the user-item interactions (e.g. ratings) and attribute information about users and items (e.g. keywords). *Collaborative recommenders* mainly use the first type of data and suggest items based on users with similar ratings (e.g. patients who favourite the same medical articles), whereas *content-based recommenders* mainly rely on attribute information of items that a user has rated before. For both models, users have to explicitly indicate what items fit their interests (e.g. by giving a like or rating). However, this is not required for the other two models, namely *demographic-based* and *knowledge-based recommenders*.

A demographic-based recommender assumes that user profiles have been made in advance (e.g. stereotypes). It looks at demographic attributes of users (such as age, gender or location) to categorize them and to recommend relevant items accordingly. A knowledge-based recommender works similarly and uses domain knowledge to define how items in the database could meet the user's needs. Thus, they need explicitly specified user requirements and look for similarities between item attributes and these requirements for generating recommendations. In fact, one could argue that search systems are a very specific case of knowledge-based recommenders, in the sense that they look for similarities between a user's query (i.e. explicit requirements) and document keywords (i.e. item attributes). However, recommender systems are usually defined as systems that *implicitly* link users to a community with related interests by looking at their ratings of items *over time*. Thus, search engines are still considered as a separate category of information retrieval systems.

3.1.3 Types of search engines

Search engines have two major functions: one function to *index* documents (i.e. building an information structure to enable searching) and another function to process queries (Croft et al., 2015). For both processes, search engine have to decide what pieces of text are relevant to the information need of the user and therefore they need a certain representation of texts. Croft et al. (2015) define two types of search engines that use different text representations: *Boolean* and *vector space models*. Boolean models (or exact-retrieval models) only retrieve documents that exactly match the query specification (e.g. returns all documents that contain certain keyword). Vector space models represent

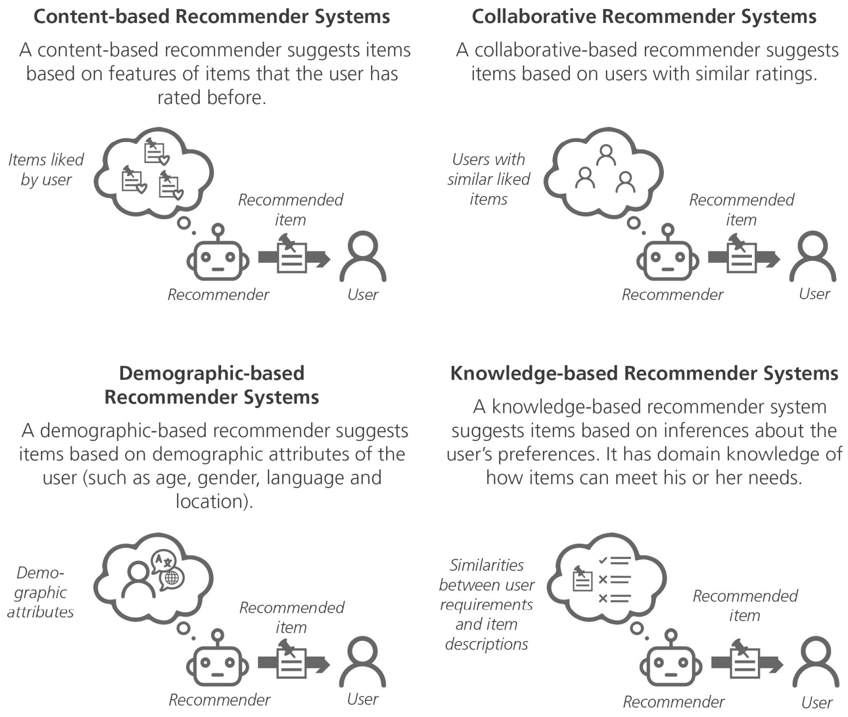

Content-based Recommender Systems
A content-based recommender suggests items based on features of items that the user has rated before.
Items liked by user
Recommended item
Recommender
User
Collaborative Recommender Systems
A collaborative-based recommender suggests items based on users with similar ratings.
Users with similar liked items
Recommended item
Recommender
User
Demographic-based Recommender Systems
A demographic-based recommender suggests items based on demographic attributes of the user (such as age, gender, language and location).
Demo-graphic attributes
Recommended item
Recommender
User
Knowledge-based Recommender Systems
A knowledge-based recommender system suggests items based on inferences about the user's preferences. It has domain knowledge of how items can meet his or her needs.
Similarities between user requirements and item descriptions
Recommended item
Recommender
User

Figure 3.1: Overview of recommender systems.

documents and queries as multi-dimensional vectors, and retrieve documents whose vectors are close to one of the query. These models can be bag-of-words or TF-IDF models, which are discussed in Chapter 2.

3.1.4 Information retrieval systems in health-based social networks

The previous chapter ended with the conclusion that smart technology should be used to provide new ways for cardiac patients to quickly find relevant information, or for healthcare professionals to quickly assess and reply to incoming questions. In general, it could be argued that recommender systems are best suited for healthcare professionals, because these systems scale well with the number of visits (i.e. interaction history of the user) and can automatically send new, relevant documents to them without an explicit request (also called *push* or *server push*). For instance, a content-based recommender could filter incoming questions based on the expertise of the healthcare professional. To train such system, the questions should be labelled by patients or by the experts themselves to guaranty high quality labels. In the latter case, it is unsure if the effort of training such a system outweighs the potential time it could save, but this aspect could be investigated in a future design project (more about this in Chapter 5).

Regarding search engines, it could be argued that they are most suited for patients instead. It has been demonstrated in the data analysis section (Chapter 2) that patients are mainly interested in asking questions to healthcare professionals, and it is therefore very likely that they are looking for specific health information on health-based social networks. A study of Medlock et al. (2015) supports this finding by demonstrating that Dutch elderly patients predominantly use the internet when searching for information on symptoms, prognosis and treatment options for their condition. This means that an information retrieval system should return documents that exactly meet their information needs, which is a task that is typically done by search engines.

However, it would be even better to design a system that combines the advantages of traditional search engines (i.e. focus on *topical relevance*) with those of recommender systems (i.e. focus on *user relevance*). For instance, when a patient would search for information about medicines, then the results could be re-ranked based on their age, gender or other user attributes. In this case, patients should create their own user profiles (similar to a demographic-based recommender), or they should use *community tags* instead, which are labels that users can give to documents that both describe their own interests and document attributes. For example, people on Twitter use hashtags to label their posts on its content, theme or mood, so that other users can easily find their posts. A search engine that incorporates user attributes would probably be the best information retrieval system in health-based social networks. Unfortunately, such a system would also inherit the disadvantages of recommender systems: their sensitivity to *cold-start* and *data sparsity*.

Cold start refers to the initial shortage of data in new, small communities (Kembellec, Chartron and Saleh, 2014). Especially collaborative recommender systems suffer from this problem, which need many community tags (or user ratings or likes) in order to function. A similar problem is data sparsity, which arises when users typically label only a few items (Guo, 2012). The data analysis in the previous chapter indicates that Dutch health-based social networks have to deal with both problems: patients only post two or three messages at most and the Dutch communities are relatively small.

Considering data sparsity and cold-start in Dutch health-based social networks, it is a good strategy to design a search engine that can initially work without labelled data and can later be personalised with user profiles. The rest of this chapter will be focused on search engines instead of recommender systems, because they are considered as the most promising solution to unburden healthcare professionals. An effective search system can take questions away among patients before they are submitted to healthcare professionals, and it is known that a large group of patients are acquainted with searching health information on the internet (see Chapter 1). In the next sections, it will be discussed what an optimal search system – without community labels - should look like for the specific case of question-and-answer communities on health-based social networks.

3.1.5 Limitations of traditional search engines

As stated before, one of the most important aspects in designing a search engine is the representation of texts. Traditional information retrieval systems normally use a simple bag-of-words (BOW) representation of texts. Although BOW has been fairly successful in the past, these representations do not assume relationships between words and so semantic structures are lost (Liu, Zhao & Volkovs, 2017). This means that BOW models consider terms such as “New York” as two independent words, namely “New” and “York”. This problem has been partly solved by using *n-grams,* a technique where texts are grouped in sequences of *n* words. Thus, the text “I am in New York” is processed as “I am”, “am in”, “in New”, “New York” with *n* equal to 2 (also called *bigrams*). However, this process increases computational costs (due to the increased vocabulary size) and does not solve the *vocabulary mismatch problem* (Croft et al., 2015). In the context of health-based social networks, this problems means that a traditional information retrieval model probably misses many relevant questions, because there are many ways to ask the same question.

Another aspect of text representation is how important words are selected or weighted in documents. Modern search engines use *vectorized* documents where words are weighted by TF-IDF scores. As explained in Chapter 2, TF-IDF is a measure that gives large weights to words that occur frequently in a particular document, but rarely occur in the entire corpus. It is a fairly simple and efficient method to ignore stop words and to prioritize meaningful words, but it might not be sophisticated enough in the context of health-based social networks. It can be observed in the experimental dataset (see Chapter 2) that most questions are preceded by the medical history of patients, which are often not relevant for other patients. The medical history can quite differ from patient to patient, so most words in these texts get a high TF-IDF score regardless. This problem will be revisited in the section 3.2, where a solution will be proposed that can mitigate this problem.

Finally, an essential aspect of search engines is how they deal with different query formulations. For example, Boolean search suggests documents based on exact keyword matches. In the hands of an experienced user, these engines can be quite powerful and return accurate results when multiple operators (AND, OR, etc.) are combined. However, it is reasonable to assume that most users do not exploit these qualities and these systems still force users to translate their problem into a few appropriate keywords to get good results. Especially in the context of question-answering communities, this is undesirable because the results become increasingly worse when longer questions are being asked. The best solution would be to design a search engine that can better handle our natural language, or at least has knowledge of semantic relationships between words. Thus, one needs a ‘smarter’ model that can better deal with the aforementioned problems.

3.1.6 Word and document embeddings

Advances in the field of Artificial Intelligence (AI) have led to systems that have the ability to acquire their own knowledge by extracting patterns from raw data. This capability is known as *machine learning* (Goodfellow, Bengio & Courville, 2016). While traditional machine learning models can only learn the mapping from representation (i.e. predefined features) to output, models with *representation learning* can also discover the representation itself (i.e. extract features from raw data). The most successful and popular approach in representation learning is *deep learning*, a technique that introduces representations that are expressed in terms of other, simpler representations. In other words, the computer learns to solve intuitive problems by building complex concepts out of simpler concepts. In this thesis, only specific aspects of deep learning will be covered that help to understand models that create *word embeddings* (a term that will be explained in the next paragraph). For more information about deep learning, please refer to Goodfellow, Bengio and Courville (2016).

Deep learning currently dominates the landscape of *natural language processing* (NLP)[3] with impressive results on tasks such as text classification and they generally outperform traditional BOW approaches (Conneau & Kiela, 2018). The key to its success is that it uses dense *word embeddings*. In word embeddings, words are represented by *n*-dimensional vectors. These word embeddings can be jointly learned along with a classification task, so that these numerical representations will start to show contextual similarities when training a model. Depending on the task at hand, syntactic (e.g. “walked”, “walking”) or semantic related words (e.g. “man”, “woman”) are placed together in vector space, as shown in Figure 3.2. This unique attribute helps computer models to better determine the textual similarity between documents and to address the vocabulary mismatch problem. Documents can be represented by a sequence of word vectors or by a single vector, but this depends on what model is used to obtain the embeddings. Popular models for obtaining word vectors are the Continuous Bag of Words (CBOW) and (Continuous) Skip-Gram of Mikolov et al. (2013), which will discussed in detail in the Materials and Method section.

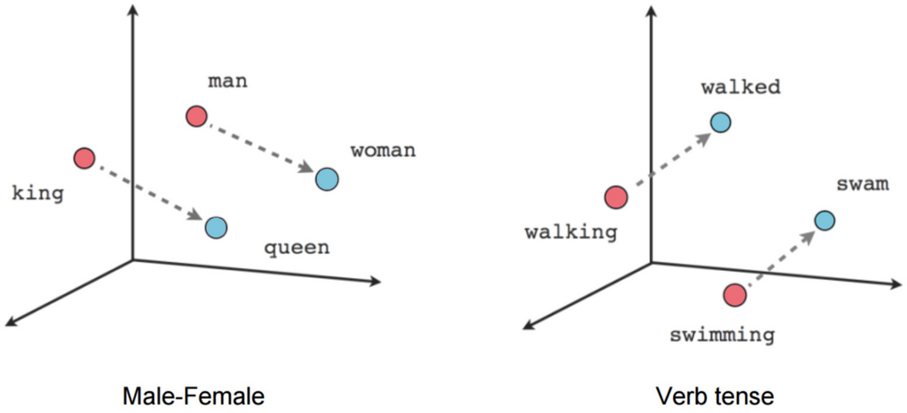


Figure 3.2: Trained word embeddings show meaningful semantic (left) or syntactic regularities (right). Reprinted from Towards Data Science website, by R. Ruizendaal, 2017, retrieved from https://towardsdatascience.com/deep-learning-4-embedding-layers-f9a02d55ac12.

[3] Natural language processing (NLP) could be defined as “a subfield of Artificial Intelligence that is focused on enabling computers to understand and process human languages” (Seif, 2018, section *Human vs Computer understanding of language*).

3.1.7 Word embeddings in information retrieval systems

Over the last few years, word and document embeddings have become very popular in tasks related to *Semantic Textual Similarity (STS)*. During STS tasks, the goal is to calculate the semantic similarity between two texts. As a result of the increased interest in STS, evaluation toolkits and benchmarks have been designed to compare the quality of *universal* word and document representations, such as SentEval (Conneau & Kiela, 2018) and SemEval (SemEval, 2019). According to these benchmarks, state-of-the-art techniques are to represent documents by averaging word vectors with a TF-IDF weighting scheme or to modify averaged word vectors with the *Smooth Inverse Frequency (SIF)*[4] (Arora et al., 2017; Ethayarajh et al. 2018). More complex techniques are (1) to train *contextual* word embeddings (i.e. multiple representations for a single word; obtained with *ELMo* from Peters et al., 2018, or *BERT* from Devlin et al, 2018), (2) to train document vectors along with word vectors (such as *Doc2Vec*; Le & Mikolov, 2014) or (3) to use other deep learning networks that learn from paired sentences or documents (such as *InferSent*; Conneau et al., 2017).

The majority of these techniques can also be used for information retrieval systems. Specifically, Brokos et al. (2016) have demonstrated that averaged word embeddings with TF-IDF weighting can be effectively implemented in a search engine in the context of question-and-answer communities. As explained before, word embeddings can be used to solve the vocabulary mismatch problem and the TF-IDF weighting makes it possible to process short questions as search queries. In addition, the same study has shown that the *Word Mover's Distance* (WMD) of Kusner et al. (2015) can be used to successfully re-rank retrieved documents. WMD is a function for calculating the semantic distance between two documents (e.g. the search query and obtained document) and will be discussed in the Materials and Method section. It should be noted that word embeddings can also be trained with help of search engines. A recent study of Zamani and Croft (2017) has shown that word embeddings can be trained on the top retrieved documents for millions of training queries (from existing search engines such as Bing), by predicting the words in the top retrieved documents for each query. These word embeddings may better capture *relevance* than word embeddings from CBOW or Skip-Gram, which mainly capture *term proximity*. However, this method also requires a vast amount of training data that is not available in the context of Dutch health-based social networks.

[4] SIF is a reweighting procedure that is very similar to TF-IDF in the sense that it reduces the influence of stop words. The difference is that SIF modifies the weighted average of word vectors with *Singular Value Decomposition (SVD),* a widely used matrix decomposition method. Please refer to the work of Arora et al. (2017) for more information about SIF and SVD.

3.1.8 Research aim

The research aim is to determine the best implementation of an information retrieval system – without community labels - in the context of question-and-answer communities for cardiac patients. The proposed model in this thesis is similar to the one of Brokos et al. (2016), but it uses an additional, novel weighting scheme for averaging word vectors that considers the overlap of semantically-related words in the patient's post and doctor's answer. Moreover, the concept of *cross validation* has been implemented in the procedure of training word vectors to obtain higher quality vectors with only a small training dataset; the experimental dataset is much smaller than in related studies. The original CBOW model of Mikolov et al. (2013) has been slightly modified as well with state-of-the-art techniques such as *Adam* (Kingma & Ba, 2014). During the evaluation, the performance of the proposed model and two variations are compared with a TF-IDF-based baseline model.

3.2 Materials and Method

3.2.1 Model overview

In this thesis, a search engine will be proposed that uses word embeddings. An overview of the proposed system is shown in Figure 3.3. In this overview, it can be seen that it consists of three major components: (1) an algorithm that pre-processes the corpus and user queries, (2) a machine learning model that learns word embeddings, and (3) a search engine that creates documents embeddings and returns a ranked list of documents. Each of these components will be discussed in detail in the following sections.

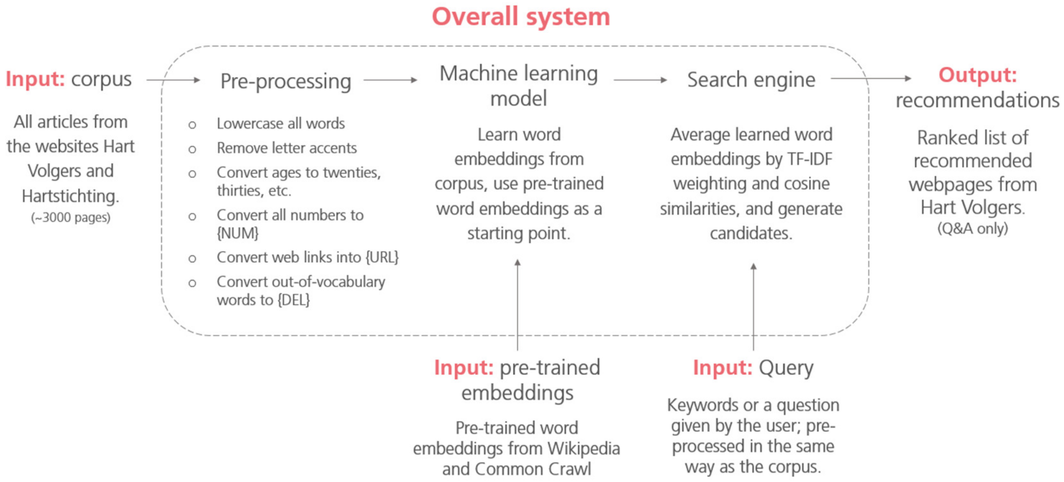

Overall system
Input: corpus
All articles from the websites Hart Volgers and Hartstichting.
(~3000 pages)
Pre-processing
Lowercase all words
Remove letter accents
Convert ages to twenties, thirties, etc.
Convert all numbers to {NUM}
Convert web links into {URL}
Convert out-of-vocabulary words to {DEL}
Machine learning model
Learn word embeddings from corpus, use pre-trained word embeddings as a starting point.
Search engine
Average learned word embeddings by TF-IDF weighting and cosine similarities, and generate candidates.
Output: recommendations
Ranked list of recommended webpages from Hart Volgers.
(Q&A only)
Input: pre-trained embeddings
Pre-trained word embeddings from Wikipedia and Common Crawl
Input: Query
Keywords or a question given by the user; pre-processed in the same way as the corpus.

Figure 3.3: Overview of the proposed information retrieval system.

3.2.2 Machine learning model

The Continuous Bag of Words (CBOW) model of Mikolov et al. (2013) was used for training the word embeddings. The original paper introduces two types of models: CBOW and (Continuous) Skip-Gram. The difference between the two models is visualised in Figure 3.4. CBOW learns word embeddings by predicting the target words based on their surrounding words, whereas Skip-Gram learns this the other way around. CBOW needs more data than Skip-Gram, but it learns faster and can obtain more accurate vectors for frequent words. Hence, CBOW may be a better choice for obtaining word embeddings for search engines, where the accuracy of the captured semantics of words are very important. In the next paragraph, the architecture of CBOW will be discussed in further detail.

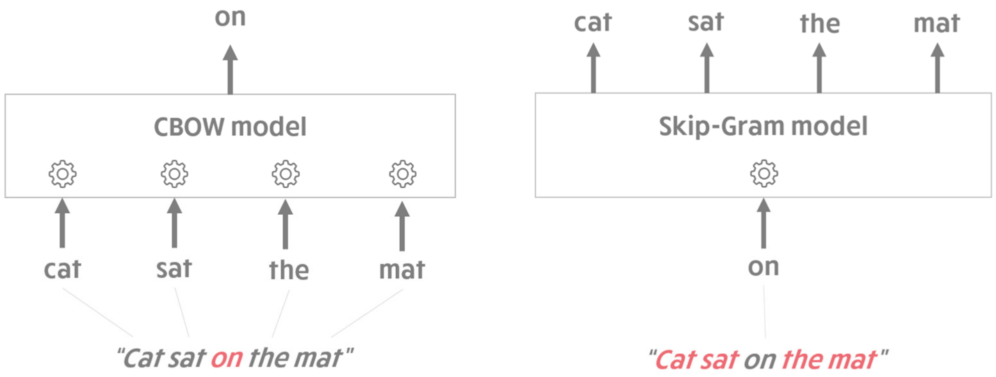


Figure 3.4: Comparison between CBOW (left) and Skip-Gram (right).

In general, CBOW is a shallow deep learning network with a single *hidden layer*, a collection of linear functions with learnable weights (also called *neurons*). The model architecture is shown in Figure 3.5. In Figure 3.5, it can be seen that the input is a set of context words and a target word. The context words are sampled within a certain distance from the target words, namely a number of words before and after the target word. In the literature, the number of context words is a hyperparameter that is called *window size*. During training, words are first represented by a unique index and then passed to an *embedding layer* (with the exception of the target word). An embedding layer is a lookup table that returns the corresponding vector of each input word. Consequently, the vectors of the context words are concatenated before they are passed on to the hidden layer. These weights are required later on for updating the word vectors that are stored in the embedding layer.

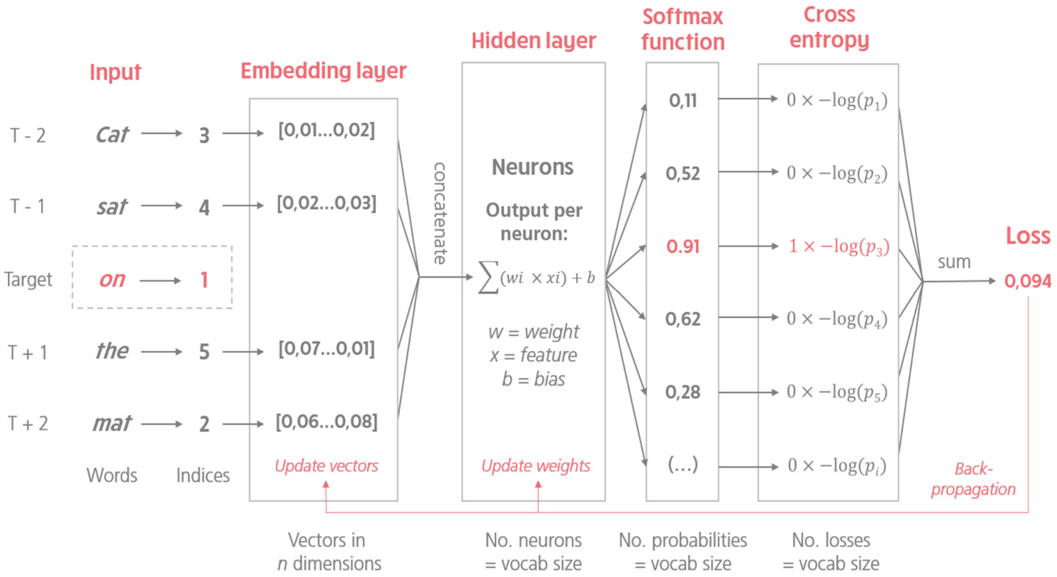


Figure 3.5: Model architecture of CBOW.

The neural network ends with a *softmax* and *cross entropy* function (which is the *cost function*) to predict the target and calculate a *loss* (i.e. the output of the cost function). The softmax function receives the vectors from the hidden layer and normalizes them into a probability distribution consisting of *n* probabilities (*n* = vocabulary size). The cross entropy function calculates the error (i.e. in this case the same as the loss) between the predicted target word probability and true target probability (which equals 1.0). The aim of the neural network is to minimize the cost function, which can be done by calculating the gradient or partial derivative of the cost function (this process also called *backpropagation*). This gradient is then used by an optimization algorithm (i.e. *gradient descent*) to adjust the weights of the neurons in the hidden layer, which in turn are used to update the stored word vectors in the embedding layer. An example of how gradient descent works is given in Figure 3.6. At the end of the learning process, a dictionary can be created that contains all the unique words in the corpus and their vectors.

For the experiments, a few modifications were made to the original CBOW algorithm with respect to the initialisation of the neural network and optimization algorithm. First of all, *pre-trained word embeddings*[5] were used to pre-initialise the neural network. Liu, Zhao and Volkovs (2017) argue that this method results in faster learning and gives a better performance compared to random initialization. The weights of the hidden layer

[5] Pre-trained word embeddings are obtained from deep learning models that have used very large corpora (such as Wikipedia) during the training process, and so they already contain useful semantic relationships.

can be further updated during training so that the model eventually finds new word embeddings that better fit the domain. Secondly, the optimisation algorithm was changed from *stochastic gradient descent* to *mini-batch gradient descent*. The first algorithm updates the model weights after processing each example and so the model may converge faster, but the training error also becomes noisier. The latter algorithm updates the model weights using a batch of examples and therefore has a larger training stability. Moreover, it can be run in parallel on the GPU (i.e. each thread simultaneously handles one sample in the batch). The experimental training set is relatively small and has a large variety in classes, so in this case batch gradient descent will be more accurate. The *Adam* optimisation algorithm, a modern extension to gradient descent, was chosen to update the weights in the network.

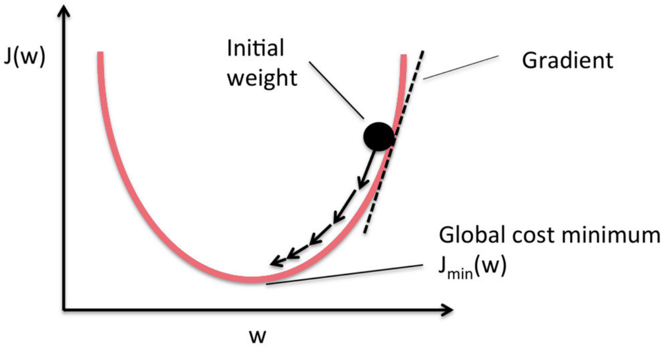
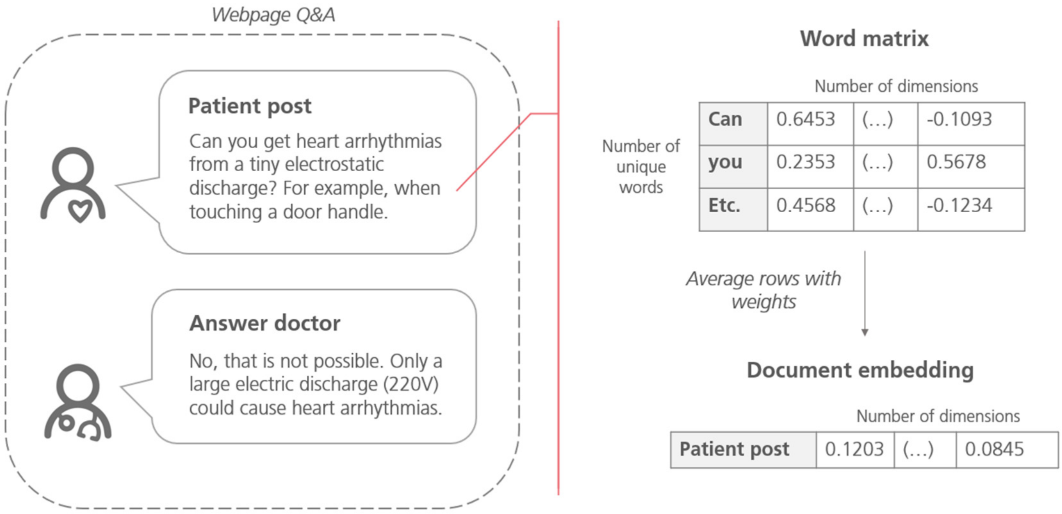


Figure 3.6: An example of how gradient descent works. The goal is to minimize the cost function *J(w)*, where *w* is a weight in the neural network. When slightly adjusting *w*, one can determine how much it affects the cost function by calculating the gradient of this function. The smaller the gradient becomes, the closer the model converges to the local or global cost minimum. Reprinted from Hackernoon website, by S. Suryansh, 2018, retrieved from https://hackernoon.com/gradient-descent-aynk-7cbe95a778da.

3.2.3 Search engine

After word embeddings have been trained on the corpus, these embeddings can be used to measure the semantic similarity between texts, and to generate recommendations. Semantic similarity between texts can be measured by calculating the angle between two document vectors (also called the *cosine distance* or *cosine similarity*). For the baseline model (TF-IDF), the documents are already vectorized and the cosine similarity can be immediately calculated. For the proposed model, the word embeddings have to be averaged first in order to obtain a document embedding. In Figure 3.7, an example is shown of how a document embedding is created. First, a text is represented as a collection of unique words where each word is represented by a vector (i.e. word embedding). Then, the word vectors are averaged in order to create a document vector (i.e. document embedding) with the same number of dimensions as the original word vectors.

However, it should be noted that all words are equally weighted when they are averaged, thus “arrhythmias” is considered equally important as “the”. A better method would be to give meaningful words (such as “arrhythmias”) a larger weight than words such as “the”. In addition, it can be observed in the dataset that most questions are preceded by the medical history of patients. The search results could be improved if the model would know which parts of the post are strongly related to the actual question. Both aspects can be addressed by incorporating two different metrics: the TF-IDF scores of words in a particular post, and the pairwise cosine distances between words in the post and those in the doctor’s answer. It is naïve to assume that a doctor always summarizes the patient’s question, but it will be demonstrated in this thesis that this assumption improves the results regardless.



Can you get heart arrhythmias from a tiny electrostatic discharge? For example, when touching a door handle.

No, that is not possible. Only a large electric discharge (220V) could cause heart arrhythmias.

Can	0.6453	(…)	-0.1093
you	0.2353	(…)	0.5678
Etc.	0.4568	(…)	-0.1234

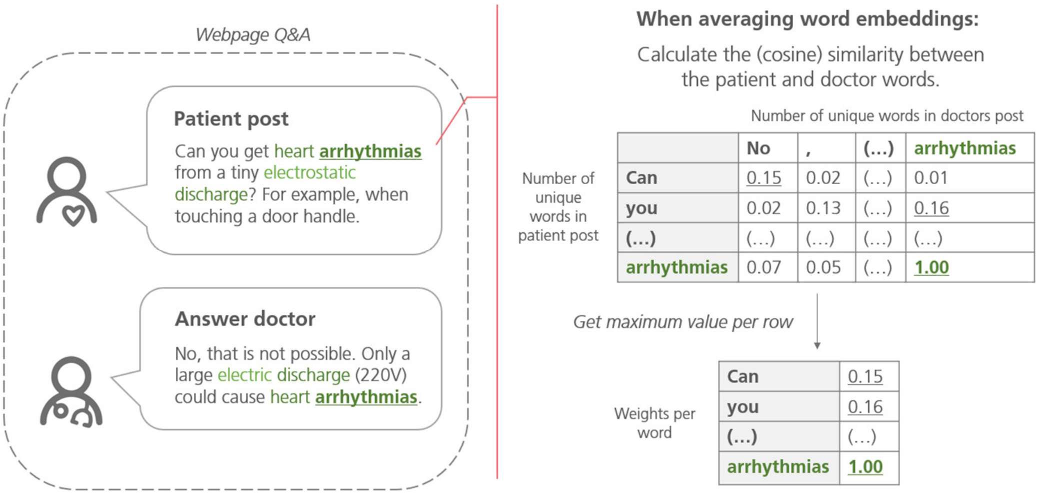
Patient post	0.1203	(…)	0.0845

Figure 3.7: An example of how a document embedding is created from word embeddings that belong to a patient post.

In Figure 3.8, an example is given of how the pairwise cosine distances are used to calculate weights per word. As explained before, each word in the patient’s post is compared with each word in the doctor’s answer. This procedure results in a matrix where each cell contains a cosine similarity score (a value between 0 and 1). If two words are semantically very similar, then this score is close to 1. If two words are identical, then this score is exactly 1. Next, one can take the maximum score per row (each row represents a word in a patient’s post) to determine the weights of the words in the patient’s post. The search system only looks at the questions of patients[6], thus the weights of the words in the doctor’s answer are not calculated. As shown in Figure 3.9, the final weights per word can be calculated by multiplying each weight with the corresponding TF-IDF score, so that stop words (especially those that occur in both question and answer) become less influential. After all weights are calculated, then the word embeddings can be averaged to obtain document embeddings.

[6] Except for the re-ranking process, where the model re-ranks the top *n* recommendations by considering all words in the question and answer.

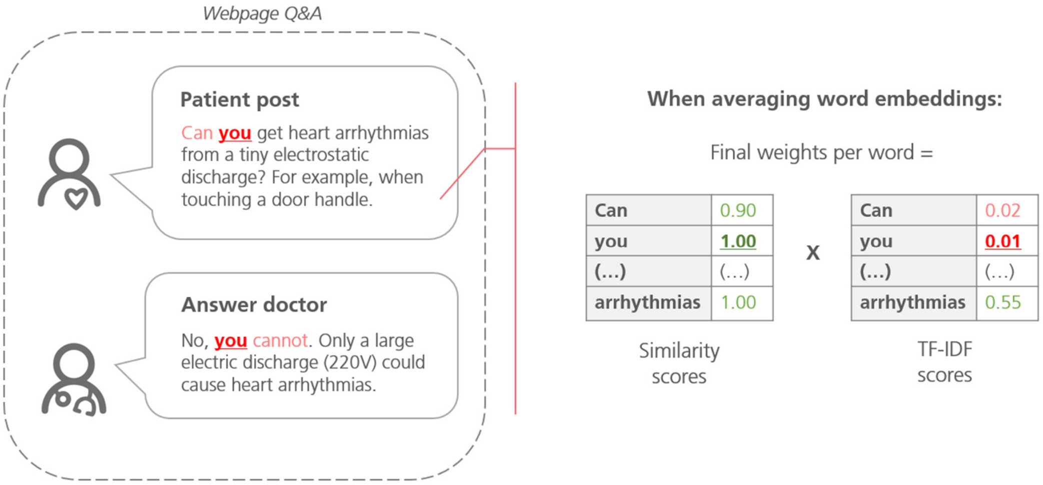
In order to generate recommendations, the recommender system calculates the pairwise cosine distances between the user's query and all document embeddings in the corpus. The user's query is pre-processed in the same way as the texts in the corpus (see section Dataset and pre-processing), but only TF-IDF scores are used as weights when averaging the words in the query. Then, the top *n* recommendations are re-ranked so that the recommendations with most similar words in comparison to the user's query are shown first (Figure 3.10). This is done by recalculating the pairwise cosine distances between the words in the user's query and those in the top *n* recommendations (with words of a single patient question and doctor's answer), but stop words and punctuation in the user's query are ignored during this process.


Webpage Q&A
Patient post
Can you get heart arrhythmias from a tiny electrostatic discharge? For example, when touching a door handle.
Answer doctor
No, that is not possible. Only a large electric discharge (220V) could cause heart arrhythmias.
When averaging word embeddings:
Calculate the (cosine) similarity between the patient and doctor words.
Number of unique words in doctors post
Number of unique words in patient post
Get maximum value per row
Weights per word

	No	,	(…)	arrhythmias
Can	0.15	0.02	(…)	0.01
you	0.02	0.13	(…)	0.16
(…)	(…)	(…)	(…)	(…)
arrhythmias	0.07	0.05	(…)	1.00

Can	0.15
you	0.16
(…)	(…)
arrhythmias	1.00

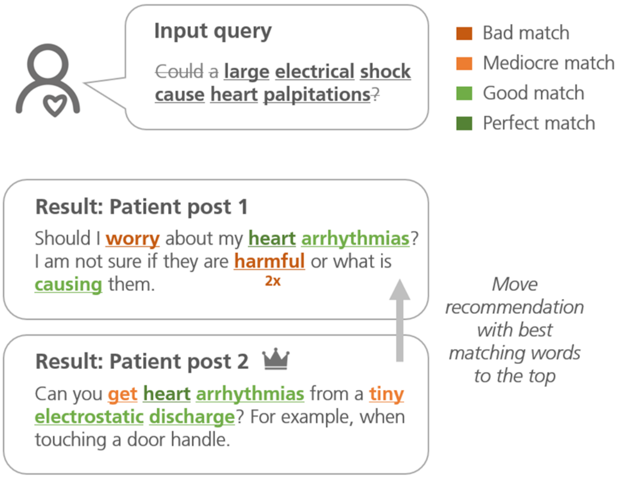
Figure 3.8: An example of how a document embedding is created from word embeddings that belong to a patient post.


Webpage Q&A
Patient post
Can you get heart arrhythmias from a tiny electrostatic discharge? For example, when touching a door handle.
Answer doctor
No, you cannot. Only a large electric discharge (220V) could cause heart arrhythmias.
When averaging word embeddings:
Final weights per word =
X
Similarity scores
TF-IDF scores

Can	0.90
you	1.00
(…)	(…)
arrhythmias	1.00

Can	0.02
you	0.01
(…)	(…)
arrhythmias	0.55

Figure 3.9: The final weights per word in the patient's post are calculated by multiplying cosine similarity scores with TF-IDF scores.



Input query

~~Could a~~ large electrical shock cause heart palpitations~~?~~

Result: Patient post 1

Should I worry about my heart arrhythmias? I am not sure if they are harmful or what is causing them.

Result: Patient post 2

Can you get heart arrhythmias from a tiny electrostatic discharge? For example, when touching a door handle.

Figure 3.10: Top *n* recommendations are re-ranked so that the recommendations with best matching words are shown first. Stop words and punctuation in the query are ignored. The re-ranking procedure also considers the words in the answer of the doctor, which are not shown in this figure.

The final similarity score per recommendation is calculated by summing the cosine similarity scores of the best matching words in a certain recommendation, thus the number of cosine similarity scores equals the number of words in the query. This method is also called the Word Mover's Distance (WMD; Kusner et al., 2015). One should choose a small number for *n,* because the computational cost of this operation rapidly increases when *n* increases and the model will increasingly override the unique document vector properties that prioritise overlapping meaningful words between the patient's question and doctor's answer.

An additional advantage of the re-ranking procedure is that it creates opportunities to make the search engine more transparent to the user. Whereas traditional exact-match retrieval models are easy to understand for users, models that use deep learning are much harder to understand due to their complex internal behaviour. However, the internal behaviour of the model does not have to be fully explained as long the user can clearly see what results are relevant and how they are related to their query. During the re-ranking procedure, each word in the user's query is compared to the words in the top *n* recommendations. Thus, one could immediately highlight the words in those recommendations which are most similar to the words in the user's query. In Chapter 4, it will be shown how this aspect looks like in the design proposal.

3.3 Experimental set-up

3.3.1 Corpus and pre-processing

A domain corpus has been built by combining articles from Hart Volgers and the website of the Dutch Heart Foundation. The articles from the Dutch Heart Foundation are publicly accessible and are obtained with the same method as described in Chapter 2.

Only educative articles on CVD and patient stories were selected and scraped from the website. The corpus contains 1.5 million tokens (i.e. words and interpunction) and counts 25 thousand unique words (i.e. vocabulary size). It should be noted that this corpus is relatively small compared to those in other studies, where word embeddings are trained on corpora that are thousand times larger (Pennington, Socher & Manning, 2014; Grave et al., 2018). Still, there are no heuristics to determine how large a corpus should be for training word embeddings, because this strongly depends on the *co-occurrences* of words in the corpus (e.g. the more words occur closely together, the better the model can capture their semantics). Considering the relatively small size of the training dataset, pre-trained Dutch word embeddings[7] with 300 dimensions were used to initialize the machine learning model and a *cross-validation* procedure was used, which will be explained in the Experiments section.

The training dataset has been pre-processed by (1) lower-casing all words, (2) removing letter accents, (3) converting ages into predefined categories (e.g. twenties, thirties, etc.), (4) converting all other numbers into a single term as “{NUM}”, (5) replacing out-of-vocabulary words by “{DEL}”, (6) replacing all web links by “{URL}”, and (7) tokenization. Punctuation was preserved so that the original distances between words are left intact. It was decided to leave abbreviations intact during the tokenization process, because they are frequently used by the community of Hart Volgers. Ages could be filtered from the text by looking at the surrounding words of the numbers in the text (e.g. the combination of “I am” and “years old” suggest that the number in question is an age). Ages are important features to keep, because it helps the system to find relevant documents that fit the profile of a user. Words that only occur once in the dataset were deleted and replaced by “{DEL}”, because CBOW cannot learn the semantics of infrequent words. All web links were filtered from the text by looking at substrings such as “http:”, “www.” or “.com” and then replaced by “{URL}”.

The queries and corpus for the experiments have been pre-processed in the same way as the training dataset (expect for out-of-vocabulary words, which are simply ignored). However, the queries and corpus for testing the baseline model have been lemmatized[8] as well, because this mitigates the vocabulary mismatch problem for a TF-IDF-based model. The corpus for the experiments only contains Q&A web pages from the website Hart Volgers. Unanswered or duplicated questions have been deleted from the corpus, and only the first answer by a medical doctor has been kept. The main reason for deleting the rest of the replies is that patients may not reply to the original question of the poster and introduce their own problem or question, which will degrade the quality of the search

[7] The word embeddings are created by Grave et al. (2018) and can be obtained from the following web page: https://fasttext.cc/docs/en/crawl-vectors.html.

[8] The lemmatization process required an open source dataset from the Dutch Language Institute (2014), which contains Dutch verbs, nouns, adjectives and their basic lemmas. The dataset can be obtained from the following web page: https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-nederlands.

results. Also, the design proposal in this thesis is intended to work with the format where questions are only answered by healthcare professionals (see Chapter 4).

3.3.2 Experiments

The CBOW model was programmed with the Python Deep Learning library Pytorch in Python 3.5. TF-IDF functions were implemented via the Scikit-learn Python library. All experiments were performed on a Windows 10 computer with an Intel Core i7-8750H CPU (2.20 GHz), 16 GB RAM and a Nvidia Quadro P1000 GPU (4 GB). GPU acceleration was enabled using the CUDA 9.0 software toolkit.

Four models have been used during the experiments, which are shown in Table 3.1. In order to validate the performance of the proposed model, two variants of this model were introduced: one that does not have cosine weighting, and another that does not have both cosine and TF-IDF weighting. This set-up makes it possible to study the effects of each weighting scheme. Besides these two models, a baseline model was used that only uses document vectors with TF-IDF weighting and no word embeddings. The goal of these models was to retrieve the most relevant Q&A pages from the website Hart Volgers based on a test set of search queries.

Table 3.1: Four models for the experiments.

Models	**Candidate selection**
Averaged word vectors	Cosine similarity + WMD re-ranking
Averaged word vectors with TF-IDF weighting	Cosine similarity + WMD re-ranking
Averaged word vectors with TF-IDF and cosine weighting*	Cosine similarity + WMD re-ranking
TF-IDF (baseline)	Cosine similarity
* The proposed model of this thesis. Cosine weighting is the pairwise cosine similarity between words in the patient’s and doctor’s post.	

The word embedding were trained by a CBOW model with the use of *early stopping*, that is, the training stops when the *validation loss* starts to increase. An example of early stopping is given in Figure 3.11. When training a machine learning model, it is recommended to split the available data into two separate datasets: one for training the model (training dataset) and another for testing the model (validation or test dataset). This procedure is also called *cross-validation*. It is essential to know how a model will generalize to an independent dataset, in order to flag problems such as *overfitting*. Overfitting means that the model has learned specific patterns in the training dataset that poorly generalize to an independent dataset. For example, when the name “John” co-occurs with the word “patient” in the training dataset, then the model could regard “John” as semantically related to “patient” if that would minimize the cross entropy loss. Studies that involve word embeddings usually do not use cross-validation, because the training datasets are so large that overfitting rarely occurs. However, the training dataset in this study is relatively small and so it was shuffled and split into 1.275.000 training samples (90 percent of all data) and 140.000 validation samples (10 percent of all data) in a *stratified* fashion. Stratified splitting means that the training and validation datasets have approximately the same label distribution. After deleting the infrequent words (see section "Corpus and pre-processing"), the datasets contained 17 thousand classes.

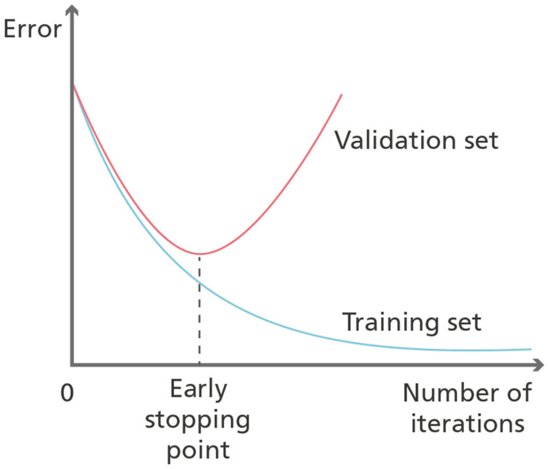


Figure 3.11: An example of early stopping. The training procedure stops when the model starts to overfit on the training dataset, that is, when the error on the validation set increases.

3.3.3 Hyperparameter settings

For the CBOW model, the window size was set to 5 (i.e. maximum distance from target word), as recommended by Mikolov et al. (2013). The embedding dimensionality was set to 300, so that the pre-trained vectors with the same dimensionality could be used to initialise the model. The learning rate of the Adam optimizer was set to 0.001 and the training batch size to 1024. For the re-ranking procedure, the top 20 (= *n*) recommendations were taken into consideration.

3.3.4 Evaluation

The evaluation of search systems is a wicked problem due two major issues. First of all, "relevance" is a complex concept and people can disagree whether a document is more relevant than the other for a given query. Therefore, it is common practice to ask a panel of judges (domain experts or users) to evaluate the relevance of the results of their information retrieval models (Croft et al., 2015). In cases where project time is limited or when these resources are unavailable, then another method is to use a benchmark dataset where documents are already judged on their relevance to the input queries. Secondly, queries can be formulated in many ways and they are often adjusted to the search results of a search engine. Thus, even a 'bad' search engine may work if users learn to formulate 'good' queries for the system. This problem can be partly addressed by preparing a test collection of search queries, which are (1) representative for the intended users, (2) sufficiently cover the themes in the dataset, and (3) are obtained from a search engine with similar properties as the one that is being evaluated.

For this thesis, the aforementioned solutions are not directly applicable. Benchmark datasets or test collections of search queries are not available in the context of question-and-answer communities of cardiac patients. Moreover, labelling a large dataset with relevance judgements is a time-consuming task[9], which would require a large number of cardiac patients or cardiologists in order finish the evaluation within the limited timespan of the design project. These specific types of users are hard to recruit as well. Therefore, it has been decided to obtain search queries by extracting questions from patient posts on Hart Volgers (in the "Ask Us" category) and to label 'duplicate' questions, that is, questions which demand the same answer or often co-occur in patient posts. Each question-and-answer post has one or more labels that correspond to the questions that it contains.

When extraction of questions, question-and-answer posts were removed from the dataset if they did not meet the following criteria:

- The post does not contain a clear question, as the user intended to share his or her experience with the community (i.e. it belongs to the wrong forum category), or requests a general advice for his or her personal situation.
- The post cannot be answered by cardiologists, because information from the patient's medical record would be needed. It is also possible that a patient requests a diagnosis, which cardiologists are not allowed to give on health-based social networks.
- The post lacks essential information, because the poster refers to another post.
- The post contains too many questions. Only posts with three or less questions were taken into consideration to facilitate the annotation process.
- The post contains too many words (i.e. more than 350 words). Skipping long patient posts facilitates the annotation process.

The extraction of the questions was done by four students from the faculty of Industrial Design of TU Delft, including the author of this thesis. The questions were summarized and split into multiple questions if they address multiple issues. The complete protocol can be found in Appendix D. Next, the questions were labelled and checked by a domain expert (project mentor). For the experiments, a dataset was used with 402 labelled question-and-answer posts and 570 unique questions (i.e. search queries), which were

[9] To give an impression of the workload for such labelling task: if one would choose 50 test queries and only consider the top 10 results of the search engine, then this means that a person has to judge the relevance of 500 results *per model*. Considering that one needs at least a single baseline model, then a person has to evaluate 1000 results.

divided into 403 categories[10]. A dataset of 402 posts was deemed representative for small, Dutch health-based social networks, since this already exceeds the number of posts currently on AFIB online.

After the dataset is labelled, then one can use multiple evaluation metrics to determine the performance of the model. Common metrics are *precision* and *recall*: precision is the fraction of search results that are relevant, and recall is the fraction of relevant results that are returned. In question-answering communities, there is often a single relevant document and so the focus should be on measuring the model performance of retrieving relevant documents at very high ranks (i.e. precision). In the literature, one usually reports *precision at rank p* (also called *precision@k*), *Mean Reciprocal Rank* (MRR) and *Mean Average Precision* (MAP) for these cases (Manning, Raghavan & Shütze, 2010; Croft et al., 2015). Precision at rank *p* corresponds to the percentage of relevant documents in the top *p* ranked results (e.g. the top 5 or 10), thus the exact rank position of the relevant documents does not matter. Therefore, this metric is combined with MRR, which considers the reciprocal of the rank at which the *first* relevant document is retrieved (averaged over a set of search queries). MAP is intuitively harder to understand and can best be explained with help from Figure 3.12. As shown in Figure 3.12, the average precision is calculated per query by averaging the precision values from the rank positions where a relevant document was retrieved. Then, the MAP score is obtained by taking the mean of all average precision values. All three measures summarize the effectiveness of a ranking algorithm over many queries. During the experiments, all three metrics were taken into account to measure the effectiveness of the four information retrieval models on the basis of the top 5 ranked documents.

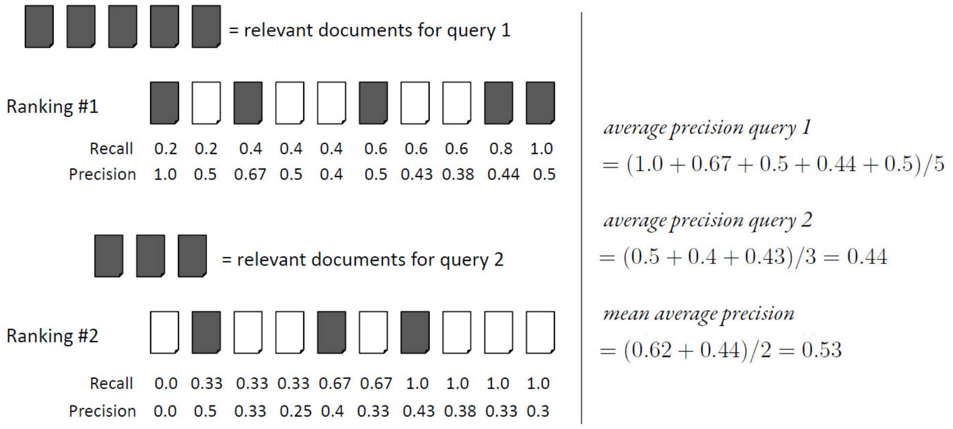

= relevant documents for query 1
Ranking #1
Recall 0.2 0.2 0.4 0.4 0.4 0.6 0.6 0.6 0.8 1.0
Precision 1.0 0.5 0.67 0.5 0.4 0.5 0.43 0.38 0.44 0.5
= relevant documents for query 2
Ranking #2
Recall 0.0 0.33 0.33 0.33 0.67 0.67 1.0 1.0 1.0 1.0
Precision 0.0 0.5 0.33 0.25 0.4 0.33 0.43 0.38 0.33 0.3
average precision query 1
$= (1.0 + 0.67 + 0.5 + 0.44 + 0.5)/5$
average precision query 2
$= (0.5 + 0.4 + 0.43)/3 = 0.44$
mean average precision
$= (0.62 + 0.44)/2 = 0.53$

Figure 3.12: Recall and precision values for rankings from two different queries (left) and calculation of the mean average precision (right). Adapted from *Search Engines Information Retrieval in Practice* (p.314), by W. B. Croft, D. Metzler, and T. Strohman, 2015, Pearson Education, Inc. Copyright 2015 by Pearson Education, Inc.

[10] Search queries should reflect at least 50 unique information needs or more according to Croft et al. (2015) and Lewandowski (2015).

3.4 Results

Table 3.2 shows an overview of the means and standard deviations of the performance of the models (mean precision@k, MRR, and MAP) on 570 search queries. All scores range from 0.0 to 1.0.

Table 3.2: Means and standard deviations of the performance of the models.

Models	Metric	*N*	*M*	*SD*
TFIDF (baseline)	Precision@k	570	.44	.45
	Reciprocal rank	570	.55	.49
	Average precision	570	.42	.44
Averaged word vectors	Precision@k	570	.55	.41
	Reciprocal rank	570	.71	.44
	Average precision	570	.53	.41
Averaged word vectors with TF-IDF weighting	Precision@k	570	.67	.37
	Reciprocal rank	570	.82	.36
	Average precision	570	.62	.37
Averaged word vectors with TF-IDF and cosine weighting	Precision@k	570	.69	.36
	Reciprocal rank	570	.84	.34
	Average precision	570	.64	.37

A Friedman test revealed that there were statistically significant differences in model performances regarding precision@k, $\chi^2(2) = 338.76$, $p < .001$, reciprocal rank, $\chi^2(2) = 245.82$, p < .001, and average precision, $\chi^2(2) = 261.94$, $p < .001$. Pair-wise comparisons based on the Wilcoxon signed-rank test were conducted to determine which models had the highest performance for each metric. As discussed in the Materials and method section, it was expected that (1) the model with averaged word vectors would generally perform better as the model with TF-IDF (baseline), (2) the model with averaged word vectors with TF-IDF weighting scheme would outperform the model with only averaged word vectors, (3) the model with averaged word vectors with TF-IDF and cosine weighting scheme would outperform the model with averaged word vectors with only the TF-IDF scheme.

The model with averaged word vectors (*M* precision@k = .55, *MRR* = .71, *MAP* = .53) performed significantly better than the model with TF-IDF on precision@k, (*M* = .44), *Z* = -5.285, *p* < .001, reciprocal rank, (*M* = .55), *Z* = -6.485, *p* < .001, and average precision, (*M* = .42), *Z* = -5.630, *p* < .001. For 21% of all queries, the former model retrieved more relevant documents than the latter model (i.e. precision@k). For 18% of all queries, the former model ranked the first relevant document higher than the latter model (i.e. reciprocal rank). For 22% of all queries, the former model obtained a higher overall precision over multiple recall levels than the latter model (i.e. average precision). An overview of the test results are shown in Table 3.3.

Table 3.3: Wilcoxon Signed Ranks Test results: comparison between the model with averaged word vectors and the model with TF-IDF.

Model comparison	Metric	Ranks	*N*	Mean Rank	Sum of Ranks	*Z*	*p*
Averaged word vectors vs. TF-IDF	Precision@k	Negative Ranks	62[a]	146.75	9098.50	-5.285[j]	<0.001
		Positive Ranks	181[b]	113.52	20547.50		
		Ties	327[c]				
		Total	570				
	Reciprocal rank	Negative Ranks	60[d]	110.72	6643.00	-6.485[j]	<0.001
		Positive Ranks	163[e]	112.47	18333.00		
		Ties	347[f]				
		Total	570				
	Average precision	Negative Ranks	70[g]	150.71	10549.50	-5.630[j]	<0.001
		Positive Ranks	194[h]	125.93	24430.50		
		Ties	306[i]				
		Total	570				

a. Averaged word vectors < TF-IDF (Precision at p)

b. Averaged word vectors > TF-IDF (Precision at p)

c. Averaged word vectors = TF-IDF (Precision at p)

d. Averaged word vectors < TF-IDF (Reciprocal rank)

e. Averaged word vectors > TF-IDF (Reciprocal rank)

f. Averaged word vectors = TF-IDF (Reciprocal rank)

g. Averaged word vectors < TF-IDF (Average precision)

h. Averaged word vectors > TF-IDF (Average precision)

i. Averaged word vectors = TF-IDF (Average precision)

j. Based on negative ranks

The model with averaged word vectors with TF-IDF weighting (*M* precision@k = .67, *MRR* = .82, *MAP* = .62) performed significantly better than the model with only averaged word vectors on precision@k, (*M* = .55), *Z* = -8.049, *p* < .001, reciprocal rank, (*M* = .71), *Z* = -6.499, *p* < .001, and average precision, (*M* = .53), *Z* = -6.444, *p* < .001. For 16% of all queries, the former model retrieved more relevant documents than the latter model (i.e. precision@k). For 11% of all queries, the former model ranked the first relevant document higher than the latter model (i.e. reciprocal rank), and also obtained a higher overall precision over multiple recall levels (i.e. average precision). An overview of the test results are shown in Table 3.4.

Table 3.4: Wilcoxon Signed Ranks Test results: comparison between the model with averaged word vectors with TF-IDF weighting and the model with averaged word vectors.

Model comparison	**Metric**	**Ranks**	***N***	**Mean Rank**	**Sum of Ranks**	***Z***	***p***
Averaged word vectors with TF-IDF weighting vs. Averaged word vectors	Precision@k	Negative Ranks	30[a]	50.68	1520.50	-8.049[j]	<0.001
		Positive Ranks	123[b]	83.42	10260.50		
		Ties	417[c]				
		Total	570				
	Reciprocal rank	Negative Ranks	32[d]	43.88	1404.00	-6.499[j]	<0.001
		Positive Ranks	94[e]	70.18	6597.00		
		Ties	444[f]				
		Total	570				
	Average precision	Negative Ranks	64[g]	68.24	4367.50	-6.444[j]	<0.001
		Positive Ranks	129[h]	111.27	14353.50		
		Ties	377[i]				
		Total	570				

a. Averaged word vectors with TF-IDF < Averaged word vectors (Precision at p)

b. Averaged word vectors with TF-IDF > Averaged word vectors (Precision at p)

c. Averaged word vectors with TF-IDF = Averaged word vectors (Precision at p)

d. Averaged word vectors with TF-IDF < Averaged word vectors (Reciprocal rank)

e. Averaged word vectors with TF-IDF > Averaged word vectors (Reciprocal rank)

f. Averaged word vectors with TF-IDF = Averaged word vectors (Reciprocal rank)

g. Averaged word vectors with TF-IDF < Averaged word vectors (Average precision)

h. Averaged word vectors with TF-IDF > Averaged word vectors (Average precision)

i. Averaged word vectors with TF-IDF = Averaged word vectors (Average precision)

j. Based on negative ranks

The model with averaged word vectors with TF-IDF and cosine weighting (*M* precision@k = .69, *MRR* = .84, *MAP* = .64) performed significantly better than the model with averaged word vectors with only TF-IDF weighting on precision@k, (*M* = .67), *Z* = -1.984, *p* = .047, reciprocal rank, (*M* = .82), *Z* = -2.032, *p* < .042, and average precision, (*M* = .62), *Z* = -3.135, *p* = .002. For 4% of all queries, the former model retrieved more relevant documents than the latter model (i.e. precision@k), and also ranked the first relevant document higher (i.e. reciprocal rank). For 8% of all queries, the former model obtained a higher overall precision over multiple recall levels than the latter model (i.e. average precision). An overview of the test results are shown in Table 3.5.

Table 3.5: Wilcoxon Signed Ranks Test results: comparison between the model with averaged word vectors with TF-IDF and cosine weighting and the model with averaged word vectors with TF-IDF weighting.

Model comparison	**Metric**	**Ranks**	***N***	**Mean Rank**	**Sum of Ranks**	***Z***	***p***
Averaged word vectors with TF-IDF and cosine weighting vs. Averaged word vectors with TF-IDF weighting	Precision@k	Negative Ranks	32[a]	43.08	1378.50	-1.984[j]	0.047
		Positive Ranks	53[b]	42.95	2276.50		
		Ties	485[c]				
		Total	570				
	Reciprocal rank	Negative Ranks	29[d]	42.52	1233.00	-2.032[j]	0.042
		Positive Ranks	52[e]	40.15	2088.00		
		Ties	489[f]				
		Total	570				
	Average precision	Negative Ranks	47[g]	70.68	3322.00	-3.135[j]	0.002
		Positive Ranks	91[h]	68.89	6269.00		
		Ties	432[i]				
		Total	570				

a. Averaged word vectors with TF-IDF and cosine < Averaged word vectors with TF-IDF (Precision at p)

b. Averaged word vectors with TF-IDF and cosine > Averaged word vectors with TF-IDF (Precision at p)

c. Averaged word vectors with TF-IDF and cosine = Averaged word vectors with TF-IDF (Precision at p)

d. Averaged word vectors with TF-IDF and cosine < Averaged word vectors with TF-IDF (Reciprocal rank)

e. Averaged word vectors with TF-IDF and cosine > Averaged word vectors with TF-IDF (Reciprocal rank)

f. Averaged word vectors with TF-IDF and cosine = Averaged word vectors with TF-IDF (Reciprocal rank)

g. Averaged word vectors with TF-IDF and cosine < Averaged word vectors with TF-IDF (Average precision)

h. Averaged word vectors with TF-IDF and cosine > Averaged word vectors with TF-IDF (Average precision)

i. Averaged word vectors with TF-IDF and cosine = Averaged word vectors with TF-IDF (Average precision)

j. Based on negative ranks

3.5 Discussion

In general, the results show that the proposed information retrieval model outperforms the baseline TF-IDF model. The TF-IDF model is one of the most common search engine on the internet, it has been implemented with an optimal pre-processing procedure (i.e. lemmatisation and stop word removal). In line with the research of Brokos et al. (2016), it has been demonstrated that (neural) word embeddings can be successfully be implemented in a search engine. The results also indicate that the quality of the obtained word embeddings is high, which means that the proposed training procedure (CBOW with cross validation) can be used on small datasets obtained from health-based social networks.

In Tables 3.2 - 3.5, it can be observed that the largest increase in performance can be observed when averaged word embeddings are used instead of TF-IDF vectors (i.e. an increase of approximately 13% averaged over all metrics). The increase in performance becomes slightly smaller when those word embeddings are averaged with TF-IDF weighting (i.e. an additional increase of approximately 11%). The cosine weighting or comparison between words in patient and doctor's post had the least impact on the performance (i.e. an additional increase of only 2%). Although this difference in performance is statistically significant, it remains uncertain if this difference is, in fact, noticeable to the users. The cosine weighting may have been more effective if long patient posts (with short answers) were kept in the dataset (see protocol in section 3.3.4). It is interesting to note that the (mean) reciprocal rank (MRR) of all models is generally higher than the precision@k and MAP. Models with a high MRR are especially useful in question-and-answer platforms, where there is often a single relevant document.

While the effectivity of the proposed model has been evaluated, its efficiency has not been evaluated yet. When the model were to be implemented in an online platform, then efficiency metrics such as *query latency* (i.e. the time between executing a query and receiving results) or *indexing time* (i.e. the time that is required for organising information before a search) would become very important. These aspects can be properly evaluated when more information is available about the daily traffic or number of search requests on the platform where the model is implemented.

A major limitation of the proposed model is that is does not 'understand' the syntax of the user's queries. Word vectors obtained by CBOW capture term proximity, which means words that co-occur often are represented by similar vectors (such as "New" and "York"). Thus, the proposed model would return very similar results for "I have AF" or "I have not AF". As discussed in the introduction of this chapter, advanced deep learning models have recently been published that can better capture syntactic relationships in texts by using contextual word embeddings (ELMo from Peters et al., 2018; and BERT from Devlin et al., 2018). However, the disadvantages of these models are that they are very slow to train compared to CBOW and they need vast amounts of training data in order to tweak their large number of parameters (e.g. a vanilla BERT model can have 350 million parameters). It is recommended for future research to investigate whether these models can be implemented by using pre-trained versions and finetuning them on the task at hand.

Another recommendation for future research is to investigate other tasks than information retrieval in the context of health-based social networks. For instance, one could develop a model for *query suggestion* or *expansion*, which means that a query is replaced by a new one without spelling errors (query suggestion), or that new words are suggested to expand the query to obtain better search results (query expansion). Furthermore, one could improve the model by looking at logs of the user's queries and the user's interactions with the platform (e.g. clickthrough data, dwelling time, etc.). In this way, one could personalise the search results without explicitly asking for the user's preferences.

Chapter 4
Design proposal

In this chapter, the design proposal will be discussed. It starts with an overview of the design requirements (Section 4.1), which are based on the findings from the literature research into the Dutch cardiac care (Chapter 1), data analysis of health-based social networks (Chapter 2), and research into search and recommendation systems (Chapter 3). Furthermore, the overall design approach will be explained (Section 4.2) as well as the features of the final design (Section 4.3). At the end of this chapter, the evaluation of the design proposal will be discussed (Section 4.4).

4.1 Design requirements

4.1.1 Needs of healthcare professionals

Professional health-based social networks in the domain of CVD are managed under the supervision of cardiologists and cardiac surgeons. These experts usually work as volunteers on these platforms, which are operated by independent health foundations, hospitals or research institutes. For instance, the platform Hart Volgers is supported by experts from the Catherina Hospital Eindhoven, and the platform AFIB online is operated by researchers from medical centres VUmc and EMC. A major component of these platforms is the question-and-answer service, which fills an important need of healthcare professionals to inform cardiac patients.

As explained in the general introduction (Chapter 1), the aftercare for cardiac patients is limited and patients visit the outpatient clinic only a few times after their treatment. In addition, proper communication between patients and medical doctors is increasingly hindered by time pressure and the prevalence of protocols (Voormolen, 2013; Van den Elsen, 2016). Therefore, it is essential that cardiac patients are well informed in advance, so that patients and cardiologists can communicate more easily and more time remains for discussing specific questions. More and more patients look for health information on the internet, thus health-based social networks are a good medium for cardiologists to reach cardiac patients and to answer their questions. These certified platforms are also essential in times where many patients can be exposed to untrustworthy information on the internet (Bos, 2018).

Requirement 1: The platform should enable healthcare professionals to efficiently inform cardiac patients to improve cardiac aftercare.

Despite the good intentions of healthcare professionals on health-based social networks, the management of these platforms could become unfeasible in the near future. An increasingly large group of cardiac patients ask questions to healthcare professionals on these platforms, while there is only a small group of healthcare professionals available (Chapter 2). Therefore, it becomes increasingly difficult to provide the same quality of care to everyone. Hence, there is an urgent need for efficient solutions to make these platforms future-proof, preferably those that incorporate smart technology (as stated in the project assignment).

Requirement 2: The platform should incorporate smart technology to unburden healthcare professionals.

Another need of healthcare professionals is related to scientific research. There is a growing interest in the applications of big data and artificial intelligence for cardiac care, especially for the early detection and prevention of cardiovascular diseases (Commit2Data, 2018). In practice, this means that vast amounts of (unstructured) data could potentially be used to enable scientific breakthroughs for improving lifestyle, CVD prevention, diagnosis or personalised treatment. For example, a recent study has concluded that algorithms could be used to help general practitioners with prescribing the correct medicine to (elderly) patients (Opondo, 2018). It is argued that the quality of prescriptions in general practice could be increased by training algorithms on a large collection of electronic medical records. The type of data on health-based social networks could be used to collect a vast amount of data about lifestyle, health and cardiovascular diseases to guide medical research.

Requirement 3: The platform should lay the foundation for collecting data about lifestyle, health and cardiovascular diseases. This data can potentially be used to guide medical research into lifestyle, CVD prevention, diagnosis or personalised treatment.

The goal of collecting data entails that health-based social networks should be publicly accessible. One could argue that these platforms should be kept within a hospital instead, so that the number of users is kept small and the workload of healthcare professionals remains acceptable. Furthermore, healthcare professionals would then have direct access to the patient's medical records, which enables them to offer better care. In fact, most Dutch hospitals already have patient portals where patients can log-in and contact healthcare professionals for help. And yet, this strategy is not suitable for collecting vast amounts of standardized data that are needed for creating intelligent systems (as discussed in Chapter 3) and possibly enabling breakthroughs in medical research. High-quality answers from healthcare professionals could also be beneficial for cardiac patients outside the hospital.

Requirement 4: The platform should be publicly accessible to facilitate the process of collecting standardized data and to help as many cardiac patients as possible.

4.1.2 Needs of cardiac patients

Cardiac patients need a place where they can ask their questions about their condition. As explained before, there are few follow-up visits at the hospital after treatment with long periods of time between them. Health-based social networks help patients to bridge those gaps (Chapter 1). These platforms are especially helpful for cardiac patients who have not participated in a cardiac rehabilitation program before, and therefore lack the knowledge to optimally self-manage their condition. Cardiac patients may also want social and emotional support from other patients who face similar issues, which is a common need on social media.

The data analysis of the platform Hart Volgers (Chapter 2) suggests that the majority of cardiac patients seek help from healthcare professionals alone and look for specific health information on health-based social networks. A study of Medlock et al. (2015) shows that elderly Dutch patients predominantly use the internet when searching for information on symptoms, prognosis and treatment options for their condition. This means that the platform should offer an information retrieval system that enables patients to find documents that exactly meet their information needs.

Requirement 5: The platform should enable patients to quickly find trustworthy and up-to-date information which is specific for their condition.

Requirement 6: The platform should enable patients to ask questions to healthcare professionals to obtain information that is hard to find in other public sources of information (e.g. brochures, websites of health foundations and hospitals, etc.).

The majority of the cardiac patients on health-based social networks suffer from heart rhythm disorders such as atrial fibrillation (Chapter 2). Atrial fibrillation is one of the most common types of CVD and it is difficult to cure as well. Although atrial fibrillation is not life-threatening, its symptoms (e.g. dizziness, pain on the chest, irregular heartbeat) can be very uncomfortable and cause anxiety among cardiac patients (Hartstichting, 2019b). Even in absence of these symptoms, patients may still have to be treated to prevent long-term complications (e.g. by taking blood thinners to prevent the development of blood clots). It is important to note that three-quarter of the patients with atrial fibrillation are older than 65 years (Hartstichting, 2019b). Altogether, atrial fibrillation is a chronic disease with frequent, sometimes unpredictable symptoms which demands additional aftercare.

Requirement 7: The platform should promote self-management among (elderly) cardiac patients with long-term conditions.

4.2 Design approach

An overview of the design process shown in Figure 4.1. The design process involved two phases: (1) analysing the design context, and (2) designing and building Harthulp. The goal of the first phase was to narrow down the scope of the project and to update the project assignment accordingly. This phase consisted of a field and desk research into Dutch cardiac care (see Chapter 1), a data analysis of health-based social networks (see Chapter 2) and other activities to synthesize the findings (i.e. a synthesis workshop and mid-term evaluation with project supervisors). The goal of the second phase was to develop a conceptual framework of Harthulp and its technology. This phase consisted of a brainstorm workshop, a technology research, prototyping activities (i.e. building and testing computer models), design activities (i.e. create overall design concept), and other activities to evaluate project outcomes. Descriptions and outcomes of all workshops throughout the design process are included in Appendix E.

4.3 Design overview

4.3.1 Introduction

The proposed design is Harthulp, a smart question-and-answering platform for cardiac patients. The most important part of the design is an online question wizard, a procedure for patients when they want to ask a question to a healthcare professional. This procedure is designed to help patients with finding relevant information, and to send their question in an efficient format for healthcare professionals. An overview of the concept is shown in Figure 4.2. Detailed screenshots of the design proposal are included in Appendix F.

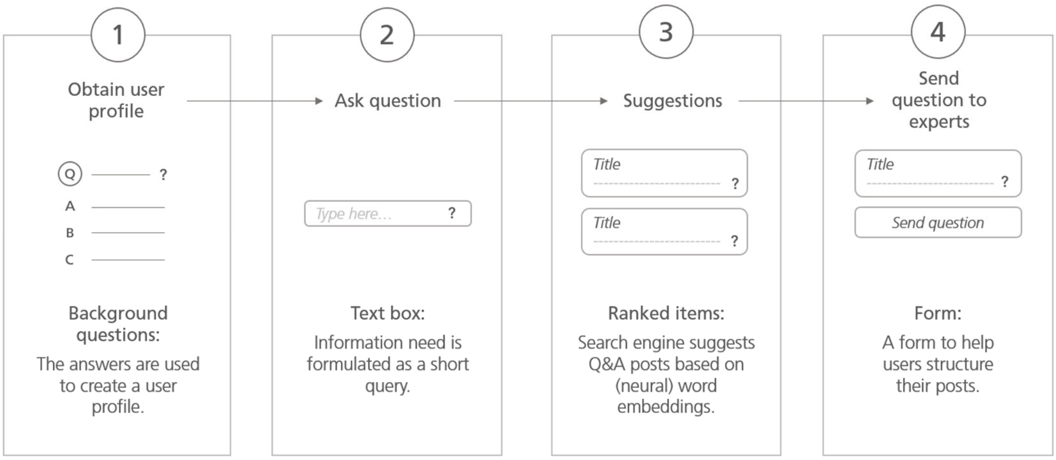

1
Obtain user profile
Q
?
A
B
C
Background questions:
The answers are used to create a user profile.
2
Ask question
Type here...
?
Text box:
Information need is formulated as a short query.
3
Suggestions
Title
?
Title
?
Ranked items:
Search engine suggests Q&A posts based on (neural) word embeddings.
4
Send question to experts
Title
?
Send question
Form:
A form to help users structure their posts.

Figure 4.2: Overview of the overall design concept: (1) Obtain the user profile, (2) Ask a question, (3) Return a list of suggestions, (4) Give the option to send a question to team of experts.

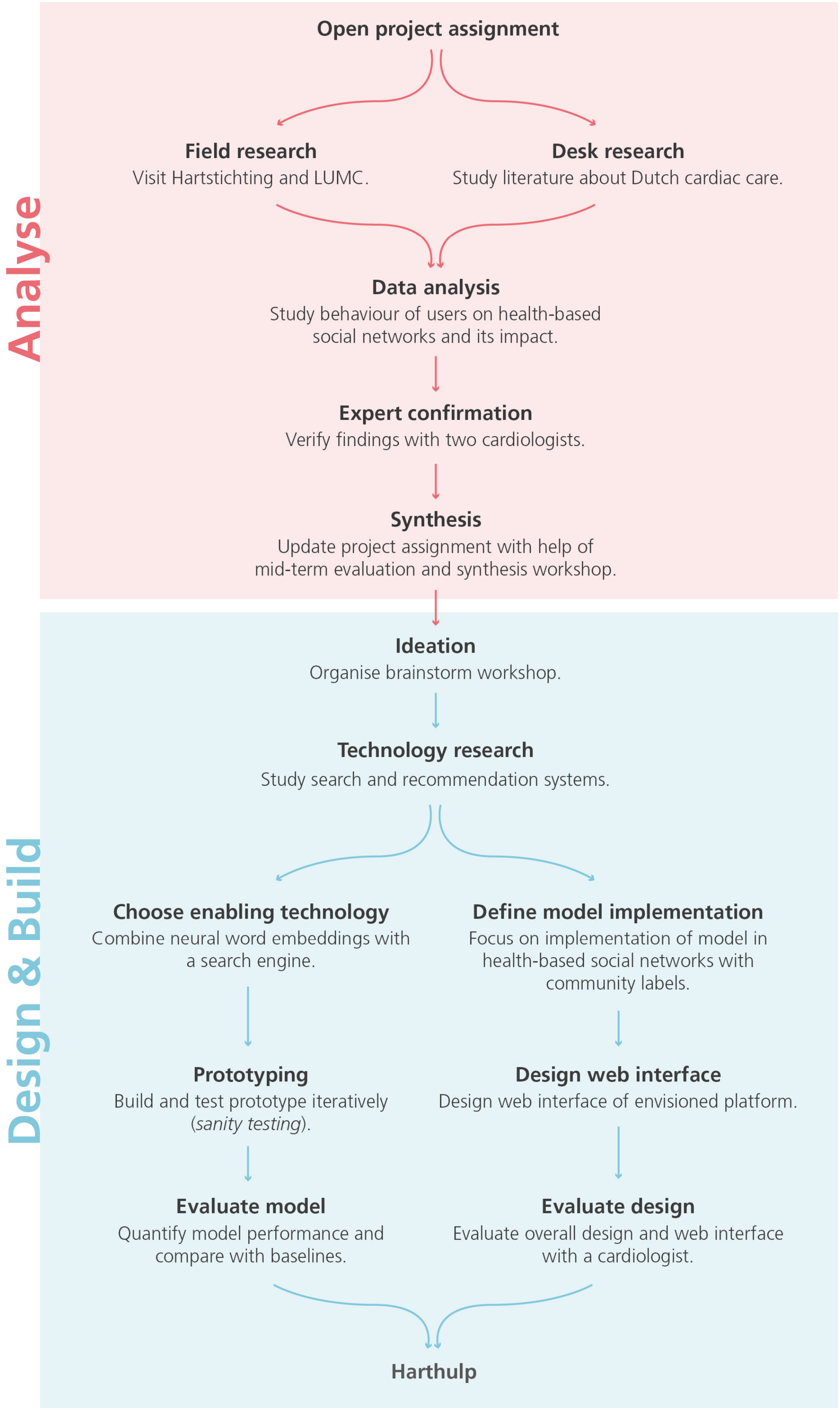

Analyse
Open project assignment
Field research
Visit Hartstichting and LUMC.
Desk research
Study literature about Dutch cardiac care.
Data analysis
Study behaviour of users on health-based social networks and its impact.
Expert confirmation
Verify findings with two cardiologists.
Synthesis
Update project assignment with help of mid-term evaluation and synthesis workshop.
Design & Build
Ideation
Organise brainstorm workshop.
Technology research
Study search and recommendation systems.
Choose enabling technology
Combine neural word embeddings with a search engine.
Define model implementation
Focus on implementation of model in health-based social networks with community labels.
Prototyping
Build and test prototype iteratively (*sanity testing*).
Design web interface
Design web interface of envisioned platform.
Evaluate model
Quantify model performance and compare with baselines.
Evaluate design
Evaluate overall design and web interface with a cardiologist.
Harthulp

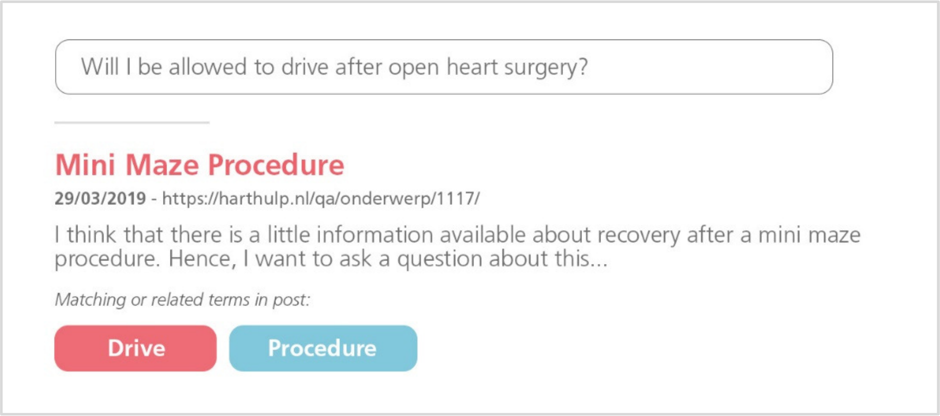
Figure 4.1: Overview of the design process.

4.3.2 Question Wizard

The procedure starts with background questions to check the identity of the user and to create an user profile. First of all, users are given a multiple-choice question where they are asked who they are: (1) a cardiac patient, (2) a family member or friend of a cardiac patient, (3) or a person who is interested in CVD. Based on the answer of the patient, the follow-up questions will be rephrased accordingly and search results can be optimized. Secondly, the user receive background questions about their gender, age and medication usage. Age and gender are important factors for cardiovascular risk management (Nederlands Huisartsen Genootschap, 2012), and healthcare professionals may change their answer based on this information. The topic clustering analysis in Chapter 2 suggests that there are many questions about medicine among cardiac patients, thus information about one's medication usage can be useful for an healthcare professional to answer these questions. When evaluating the design (Section 4.4), it will be discussed whether more background questions are needed.

The background questions can be used to improve the search results later on in the question wizard (the third step in Figure 4.2) by re-ranking the documents based on the user's profile. The current search engine only returns documents based on *topical relevance* and not necessarily on *user relevance*. Given enough data, the user's profile could be used to prioritize documents from similar patients, but the desired influence of the user's profile on the search results should be investigated in future research (see Chapter 5). The background questions can be skipped by users who are not cardiac patients, but they are still given the option to fill in these questions when their question is related to a cardiac patient. Users can create an online account where their answers are saved, so that they do not have to fill in these questions for their next visit on Harthulp.

When the background questions are filled in, then users are given the option to search through Harthulp. They can formulate their information need as a short query, which can be a set of keywords or a short question. Next, the search engine returns a ranked list of documents (i.e. question-and-answer posts) based on their relevance to the search query. The presentation of the search results is shown in Figure 4.3. Each result contains the title, data, URL and a text snippet of the question-and-answer post, and the matching or related terms to the search query. As explained in Chapter 3, the proposed search engine uses (neural) word embeddings and therefore it can retrieve documents with terms that are semantically related to the search query, but do not occur in the query itself.



Will I be allowed to drive after open heart surgery?

Mini Maze Procedure

29/03/2019 - https://harthulp.nl/qa/onderwerp/1117/

I think that there is a little information available about recovery after a mini maze procedure. Hence, I want to ask a question about this...

Matching or related terms in post:

Figure 4.3: An example search query and result. The red tag shows a direct match ("drive") and the blue tag shows a term related to the search query ("procedure" instead of "surgery"). It should be noted that this example has been translated from Dutch, where "open heart surgery" can be written as a single word and the best synonym would then be "surgery" instead of "procedure".

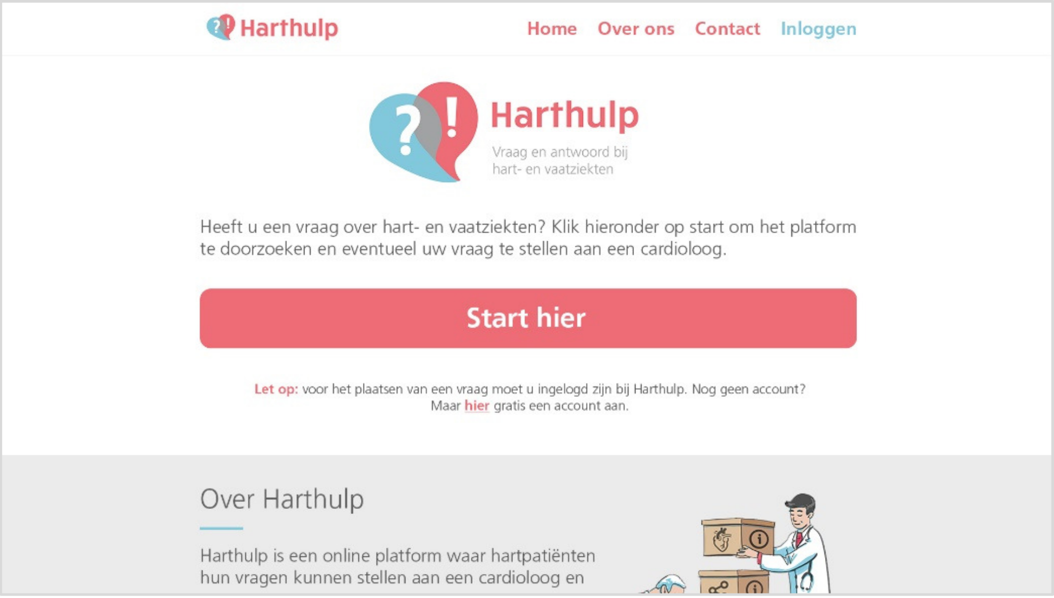
Normally, it is impossible to highlight direct keyword matches with this kind of system, because queries and documents are represented as single vectors. However, the re-ranking algorithm (i.e. word mover's distance or WMD) compares each word in the top ranked results with the those in search query and returns words which are deemed semantically related to the query (i.e. words whose cosine similarity exceeds a certain threshold). As a result, the mechanism of the search system becomes more transparent to the user and the highlighted words below the search results summarize their contents. This is an essential feature because it can be observed in the Hart Volgers dataset that patient posts rarely start with their actual question and that most posts are quite long (i.e. approximately 108 words on average, excluding outliers, see Tukey boxplot analysis in Appendix C).

After the search procedure, patients may still want to ask their question to a cardiologist when they cannot find the information that they are looking for. If they have an account on Harthulp, then the question wizard continues and they can write down their question in a form. The form consist of two text fields: one for the title (or question) and another for the explanation. The text field for the explanation has a word limit in order to keep posts concise for the cardiologist and the search algorithm (i.e. the longer the posts, the worse the document vectors becomes, because document vectors are obtained by averaging word vectors). When filling in the form, users get a reminder that cardiologists do not have access to their medical records and are not allowed to give a diagnosis on Harthulp. It can be observed in the Hart Volgers dataset that patients sometimes have wrong expectations of the platform and post questions that cardiologists cannot answer (examples are given in the labelling protocol in Appendix D).

At the end of the question wizard, users get a confirmation whether their question has been successfully sent to the Harthulp team. Users also receive additional information about the follow-up procedure: they will be notified when their question has been answered; and until that time their questions remain invisible for others, in order to prevent disinformation on the platform. A post can still be adjusted by the user as long it has not been answered yet. After the question has been answered by a cardiologist, the topic is closed and other patients cannot reply. The reason for this is that other patients sometimes reply with their own question to the cardiologist that is related to the original question (this can also be observed in the Hart Volgers dataset), which makes it difficult for cardiologists to keep an overview of all incoming questions.

4.3.3 Web interface

An impression of the web interface of Harthulp is given in Figure 4.4. The current web interface is solely designed for question-and-answering and therefore the focus is on the question wizard (Figure 4.5). The interface of the question wizard is inspired by the order procedure of large web shops, where requirements such as ease-of-use and accessibility are key factors in the user experience. Another source of inspiration is the decision tree of the well-known Dutch website "Moet ik naar de dokter?", where users can check if they have to visit their general practitioner when experiencing certain signs or symptoms. To the best of the author's knowledge, there are no health platforms yet that combine both a question wizard and a search engine (with deep learning) to retrieve community-generated medical content. To improve visual appeal, it was decided to use soft colours and a sans serif font to enlighten the mood on the platform, which contains many serious themes and discussions.



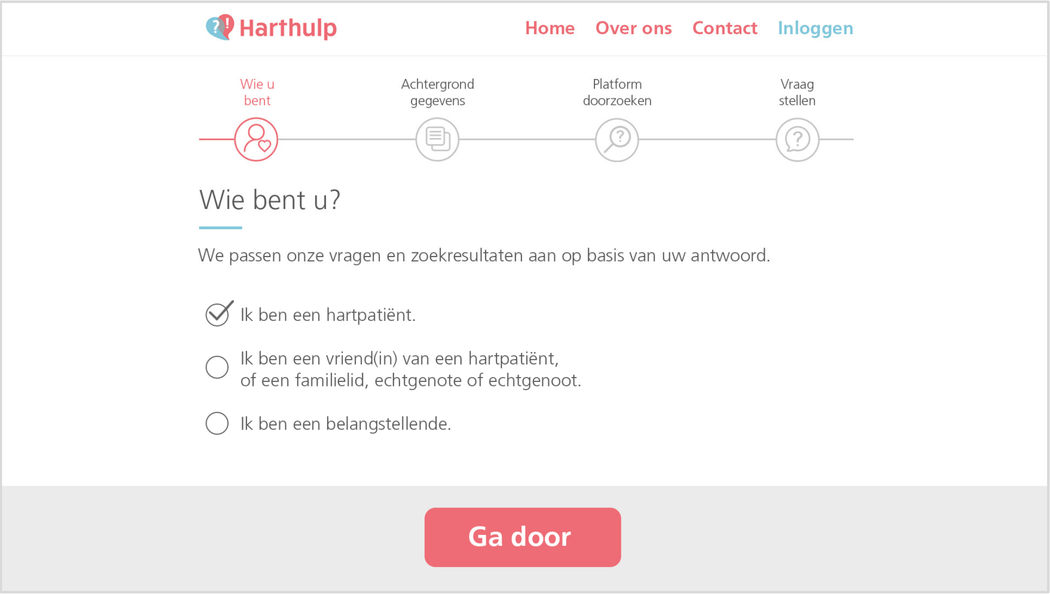
Heeft u een vraag over hart- en vaatziekten? Klik hieronder op start om het platform te doorzoeken en eventueel uw vraag te stellen aan een cardioloog.

Let op: voor het plaatsen van een vraag moet u ingelogd zijn bij Harthulp. Nog geen account? Maar hier gratis een account aan.

Over Harthulp

Harthulp is een online platform waar hartpatiënten hun vragen kunnen stellen aan een cardioloog en

Figure 4.4: Impression of the web interface of Harthulp.



Wie bent u?

We passen onze vragen en zoekresultaten aan op basis van uw antwoord.

☑ Ik ben een hartpatiënt.

○ Ik ben een vriend(in) van een hartpatiënt,
of een familielid, echtgenote of echtgenoot.

○ Ik ben een belangstellende.

Figure 4.5: The web interface of the online question wizard on Harthulp.

4.3.4 Long-term usage

The main purpose of the design is to quicker and better inform cardiac patients in order to reduce the incoming stream of online questions for healthcare professionals. Paradoxically, the more content that users generate on the platform, the more valuable the platform becomes. Frequent updates with user-generated content motivate users to use the platform over longer periods of time. For instance, information needs of patients with atrial fibrillation (which is a chronic condition) change over time, because the condition of AF can worsen when they get older and patients have to undergo new treatments (Hartstichting, 2019c). For these patients, it is helpful to periodically revisit Harthulp for new information about their condition. However, long-term engagement involves a degree of self-disclosure on the part of the user, who have to be comfortable with sharing information about themselves on the internet. The growing collection of online content also enables machine learning algorithms to learn and improve over time. Long-term applications of machine intelligence will be further discussed in Chapter 5.

4.4 Design evaluation

The design was evaluated by means of an semi-structured interview with cardiologist Stijn de Ridder. The interview questions are included in Appendix G. In the following paragraphs, the main insights from the interview will be discussed.

4.4.1 Question wizard

In general, it is expected that the current question wizard could reduce the workload of cardiologists on health-based social networks:

> "I think that the overall concept of the questions wizard is good. If you do not let patients search first, then you could get a large number of questions which already has been answered before. In daily practice, I receive the same questions from many patients, for example: "*When am I allowed to return to work after balloon angioplasty* (a procedure used to widen blocked coronary arteries)? ". The question wizard could prevent duplicate questions on the platform."

From the cardiologist's perspective, the background questions about age, gender and medication usage could be helpful to give better answers:

> "Especially the background questions about age and medication usage should be kept. As a cardiologist, almost immediately you know in what phase patients are when you have this information. For example, if a patient has severe heart failure, then this is reflected by his or her medication list. In such cases, you should be more careful with your answer, and background information can be helpful to give a more specific, nuanced answer."

From the patients' perspective, the background questions can be used to get more relevant results:

> "It is more interesting for patients to read posts from similar patients. For example, a patient of 70 years old with a heart rhythm disorder cannot use the information from a post from a patient of 20 years old, because patients in this age category often have different types of heart rhythm disorders. It can be even inappropriate for patients to read posts from patients with different profiles, because they might interpret the answers incorrectly and share inaccurate information among each other. It is essential to make use of the patient's medical background."

Moreover, the documents could be re-ranked by looking at the user's interaction with the platform or by letting patients rate the answers of cardiologists. The most voted answers could then be placed higher in the rankings. Furthermore, the number of background questions should be kept as small as possible, since visitors may quit the question wizard if there are too many of them. The question about medicine usage should be kept simple in particular (the name of the medicine alone is sufficient), because patients may need time to look up this information. Additional questions could be saved in a user profile. These profiles may contain whether patients are currently under treatment or have been treated before (including the type of treatment and treatment period).

When posting a question, the word limit on the question form could motivate patients to be more concise in their questions. Nevertheless, this procedure should include more disclaimers to manage the patient's expectations and to protect cardiologists:

> "The word limit could indeed be helpful to prevent long stories from patients. However, what I am missing here is a disclaimer so that cardiologists can freely give their opinions. Sometimes, we are entering dangerous waters on these platforms: patients could ask "Should I stop with blood thinners?". Nobody can give a good answer on such a question and patients should contact their own cardiologist instead. Of course, you still want to help them on the platform, but then you should add a disclaimer. For example, patients could be notified that answers might not be specifically applicable to them due to the absence of medical records. Cardiologists are then directly protected; they can freely give their opinion, because no rights are conferred by their answers."

4.4.2 Harthulp for medical research

The data of patients should be used to improve the services on Harthulp, such as improving the search engine and underlying algorithms. Statistics from the platform may inspire researchers or increase their awareness about the situation of cardiac patients as well, but the data itself cannot be used for medical research:

> "Using patient-reported data for medical research is notoriously difficult and I personally think that doctors do not want to use this data, because it will be very biased. If you would ask patients about their diagnosis, then half of the time they wrongly recall their own diagnosis. For example, patients say that they have had a cardiac arrest, but that is not true. Patients mention a specific type of heart rhythm disorder, but they unintentionally mixup different disorders. It is perhaps too difficult to improve such a database for medical purposes, but it can still be used to make doctors more aware about the circumstances of Dutch cardiac patients or to improve the search results."

The details of the data collection should be given in the privacy policy of the platform. Moreover, one should not ask for explicit permission to collect data for medical purposes in the question wizard, because this suggests that the data will be used in medical research itself instead of guiding medical research.

4.4.3 Long-term vision of Harthulp

A long-term vision of Harthulp could be to expand to other platforms, so that different patient groups could also benefit from Harthulp:

> "A long-term vision could be to let other online health platforms use the same question wizard and smart search engine. You could start with platforms for cardiovascular diseases, and then move to other platforms, such as the ones for diabetes. In fact, Harthulp could become a Google for health platforms. Instead of merely searching on "bladder infection", patients could combine their search with background information, such as "I am a woman of 70 years old and use this medication", to directly get the information they need. An even more ambitious goal could be to expand to English health platforms and to translate the content, so that patient with other nationalities could benefit from Harthulp as well."

Yet, it is important that Harthulp remains neutral in the future:

> "I strongly believe that Harthulp should always remain a non-profit platform, without commercial partners such as pharmaceutical companies. The lack of sponsors drastically reduces budget, but the platform would then remain neutral and does not have to conform to commercial interests. You could partner up with large health foundations instead, who would be definitely interested in the services of Harthulp."

4.4.4 Recommendations

It is recommended to develop a specific interface for cardiologists to further reduce their workload. In Chapter 3, it has been discussed that a recommender system could be designed for cardiologists, an idea that has been proposed during the interview:

> "It could be useful to narrow down the incoming streams of questions by categorizing them. For instance, you could let an algorithm read the post of a patient to determine the subject or theme of the post. If the post is about cardiac ablation, then you could send the post to an expert of this procedure and not to a general cardiologist. This would require a platform where cardiologists can sign up – with a BIG registration - and register their field of expertise."

A future design concept should also try to protect patients from tragedies which are not applicable to their situation:

> "If a patient – who has had surgical ablation - would like to check if returning arrythmias are common, then they could be confronted with tragedies from other patients who have had the procedure six or seven times or severe rebleeding afterwards. This is a very challenging problem to solve, because an algorithm would then need to know the common

> side-effects of surgical ablation. Ideally, an algorithm should be able to automatically filter all the side-effects from posts which only occur in less than 5% of all cases, which are the side-effects we also do not mention to patients. You could get a sort of placebo effect when patients read side-effects from others, that is, they start to experience side-effects that they would not have had if they did not know about them. Patients should be protected from tragedies of other patients which are not applicable to them."

Finally, the cardiologist stresses the importance of presenting Harthulp as a smart question-and-answer platform rather than a forum:

> "You should not present your design as a platform for patients only. To be honest, we usually dislike online forums – because they are unstructured and biased – and we are still trying to accept co-creation platforms. Therefore, Harthulp should not be presented as another platform where patients just post stories: patients are directly guided to the correct answers that belong to the correct questions. Content is filtered on whether they are relevant or not. This will be very difficult, but Harthulp would then be quite appealing to cardiologists."

Chapter 5

General discussion

The first sub-research question is what impact the growth of the number of cardiac patients and their health-information seeking behaviour have on the sustainability of health-based social networks. Based on the data analysis in Chapter 2, it has been concluded that the growth of the number of cardiac patients on health-based platforms is large, while the number of healthcare professionals remains small. Hence, there is an urgent need to make health-based social networks future proof.

The second sub-research question is what needs cardiac patients have on health-based social networks. Based on the same data analysis, it has been concluded that the majority of Dutch cardiac patients prefer to contact cardiologists instead of fellow patients. This is reflected by the large number of discussion boards in the “Ask Us” category on Hart Volgers and the low number of responses per patient to posts of others on the platforms Hart Volgers and DailyStrength.

The third sub-research question is what smart technologies can be used to optimize information retrieval in large scale medical text data. In Chapter 3, it has been argued that recommender systems are most suited for healthcare professionals, because these systems scale well with the many returning visits of cardiologists and they could automatically filter incoming questions based on their expertise. Search engines are probably most suited for patients, because the findings in Chapter 2 suggest that patients have very specific information needs. Search engines can help patients to find the online information that they need, so that they do not have to post a question and to wait for a reply.

The fourth sub-research question is what the performance is of state-of-the-art information retrieval systems on question-and-answer data from health-based social networks. It has been demonstrated in Chapter 3 that search engines with neural word embeddings outperform a traditional search engine when retrieving relevant question-and-answer posts from the Dutch platform Hart Volgers. The performance of search engines with neural word embeddings greatly increases when word vectors are averaged with TF-IDF weighting. Using an additional weighting scheme, where words in patient posts are compared with words in the doctor’s post, leads to a small increase in performance.

The fifth sub-research question is how health-based social networks can be best designed to enable the integration of smart technologies. In Chapter 4, it is shown how this engine can be implemented in a question wizard, which guides patients during the search process and collects labels for future improvements. A web interface has been designed that shows how users should interact with the platform, and it has been evaluated by an experienced cardiologist.

A limitation of the project is that it does not provide details of how labels or background information from patients could be exactly implemented to improve search results. In Chapter 3, it has been explained that platforms which rely on these labels suffer from a *cold start*, and therefore a platform has been designed that could already work from the beginning. Moreover, the platform could improve over time by learning from incoming labels. For future research, it is recommended to investigate how labels (e.g. age, gender and medication usage) could best be combined with search queries. Another suggestion is to explore other ways of collecting background information from patients. For instance, it has been mentioned in the interview in Chapter 4 that patients may want to rate the answers of cardiologists, so that the best answers can be ranked higher. Additionally, the ratings could be used to discover the preferences of users and to personalise search results.

Another recommendation is to conduct tests with cardiac patients to evaluate the user experience on the envisioned platform and to refine its interface accordingly. It should be investigated whether patients can easily use the question wizard and to what extent it can satisfy their information needs. A large group of patients could also be monitored over longer periods of time to check the impact of the design on the number of incoming questions. However, this would require a working web application with all the algorithms embedded, which could be developed with a web framework such as Django (www.djangoproject.com).

For now, an unique interface for healthcare professionals is missing in the design. The idea has been proposed to design a recommender system for cardiologists, which automatically assigns incoming questions to each cardiologist based on his or her expertise. It is therefore recommended to investigate how such a system should be designed, from both a technological and user perspective. For instance, it is uncertain if training such a system outweighs the potential time it could save, and how it should be implemented in the workflow of healthcare professionals. Another service for healthcare professionals could be to enable them to easily ask questions among each other. For example, the Dutch company Siilo offers an encrypted chat service (i.e. Siilo Connect) for healthcare professionals where they can discuss cases among each other (Siilo, 2019). This could be a welcome feature for health-based social networks.

Finally, the design proposal does not address the issue that cardiac patients could be confronted with negative posts from other patients. In the interview in Chapter 4, it has been suggested to create a system that automatically filters rare side-effects, complications or other redundant background information. This can be very challenging, because such a system needs to have domain knowledge. Nonetheless, deep learning could be a promising technique for text summarisation (Patel et al, 2018). One could even take it one step further and work on a chatbot that can automatically answer complex questions about CVD-related subjects. Chatbots are increasingly common on the internet and are already provided by companies such as IBM (Watson), Apple (Siri), Google (Assistant) or Amazon (Alexa). However, it is unclear how they should be implemented in health-based social networks, because they still cannot completely understand our natural language, and people also value social contact (Emerce, 2019). The design of text summary systems or chatbots for cardiac patients could an interesting research direction for the future.

References

Achttien, R. J., Staal, J. B., Merry, A. H. H., & Voort, S. S. E. M. (2011). KNGF-richtlijn Hartrevalidatie. *Supplement bij het Nederlands Tijdschrift voor Fysiotherapie*, *121*(4).

American Heart Association. (2016). What is Atrial Fibrillation (AFib or AF)? Retrieved February 25, 2019, from https://www.heart.org/en/health-topics/atrial-fibrillation/what-is-atrial-fibrillation-afib-or-af.

Arora, S., Liang, Y., & Ma, T. (2016). A simple but tough-to-beat baseline for sentence embeddings. Paper presented at the *ICLR 2017 conference.* Retrieved from https://openreview.net/pdf?id=SyK00v5xx.

Bos, K. (2019, January 26). De patiënt weet het beter dankzij dokter Google. Retrieved February 19, 2019, from https://www.nrc.nl/nieuws/2018/01/26/de-patient-weet-het-beter-dankzij-dokter-google-a1589646.

Brokos, G. I., Malakasiotis, P., & Androutsopoulos, I. (2016). Using centroids of word embeddings and word mover's distance for biomedical document retrieval in question answering. *arXiv preprint arXiv:1608.03905.*

Casari, A., & Zheng, A. (2015). The Effects of Feature Scaling: From Bag-of-Words to Tf-Idf - Feature Engineering for Machine Learning. Retrieved February 11, 2019, from https://www.oreilly.com/library/view/feature-engineering-for/9781491953235/ch04.html.

Commit2Data. (2018). Big Data & Health: early detection and prevention of cardiovascular diseases. Retrieved from https://commit2data.nl/commit2data-programma/gezondheid/big-data-health-early-detection-and-prevention-of-cardiovasculair-diseases.

Conneau, A., & Kiela, D. (2018). Senteval: An evaluation toolkit for universal sentence representations. *arXiv preprint arXiv:1803.05449.*

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364.*

Croft, W. B., Metzler, D., & Strohman, T. (2015). Search Engines and Information Retrieval. *Search Engine: Information Retrieval in Practice*, 1–9.

de Boer, A. R., Bots, M. L., van Dis, I., Vaartjes, I., & Visseren, F. L. J. (2018). *Hart- en vaatziekten in Nederland 2018.* Den Haag. Retrieved from https://www.hartstichting.nl/getmedia/a6e15c10-2710-41b9-bcf8-8185feaf54b2/cijferboek-hartstichting-hart-vaatziekten-nederland-2018.pdf

De Vos, C., Li, X., Van Vlaenderen, I., Saka, O., Dendale, P., Eyssen, M., & Paulus, D. (2013). Participating or not in a cardiac rehabilitation programme: factors influencing a patient's decision. *European journal of preventive cardiology, 20*(2), 341-348.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Dubberly, H., Mehta, R., Evenson, S., & Pangaro, P. (2010). Reframing health to embrace design of our own well-being. *Interactions, 17*(3), 56-63.

Dutch Language Institute. (2014). Referentiebestand Nederlands (RBN). Retrieved from https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-nederlands.

Emerce. (2019, January 14). Nog weinig vertrouwen in chatbot. Retrieved January 16, 2019, from https://www.emerce.nl/nieuws/nog-weinig-vertrouwen-chatbot.

Ethayarajh, K. (2018). Unsupervised random walk sentence embeddings: A strong but simple baseline. In *Proceedings of The Third Workshop on Representation Learning for NLP* (pp. 91-100).

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893.*

Guo, G. (2012, July). Resolving data sparsity and cold start in recommender systems. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 361-364). Springer, Berlin, Heidelberg.

Hartstichting. (2019a). Cijfers over hart- en vaatziekten. Retrieved February 19, 2019, from https://www.hartstichting.nl/hart-en-vaatziekten/feiten-en-cijfers-hart-en-vaatziekten.

Hartstichting. (2019b). Boezemfibrilleren [brochure]. Retrieved from https://www.hartstichting.nl/getmedia/c504aba2-4a3c-4287-9e48-98db0ec5f2d8/brochure-hartstichting-boezemfibrilleren.pdf.

Hartstichting. (2019c). Boezemfibrilleren: van kwaad tot erger voorkomen. Retrieved from https://www.hartstichting.nl/wetenschappelijk-onderzoek/hartritmestoornissen/boezemfibrilleren-van-kwaad-tot-erger-voorkomen.

Institute for Work and Health (2015, April). Primary, secondary and tertiary prevention. Retrieved from https://www.iwh.on.ca/what-researchers-mean-by/primary-secondary-and-tertiary-prevention.

Jonkers, A. (2018, November 30). Gezond verder na een hartinfarct gaat niet vanzelf. Deze drie ingrepen zijn hard nodig. *de Volkskrant*. Retrieved from https://www.volkskrant.nl/wetenschap/gezond-verder-na-een-hartinfarct-gaat-niet-vanzelf-deze-drie-ingrepen-zijn-hard-nodig-~b5281ef8/.

Kembellec, G., Chartron, G., & Saleh, I. (2014). Recommender systems. *Recommender Systems*, 1–232. https://doi.org/10.1002/9781119054252.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

KNMG (2018). Artsen en Social Media - Handreiking voor Artsen. Retrieved from https://www.knmg.nl/advies-richtlijnen/dossiers/sociale-media.htm.

Konings, H., Chavannes, N., & Kreindler, J. (2018). *Verslag eHealth congres*. Retrieved from the KNMG website: https://www.knmg.nl/web/file?uuid=949d549a-b448-435f-9c25-f5613edf5e96&owner=5c945405-d6ca-4deb-aa16-7af2088aa173&contentid=72314.

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015, June). From word embeddings to document distances. In *International Conference on Machine Learning* (pp. 957-966).

Le, Q., & Mikolov, T. (2014, January). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196).

Lewandowski, D. (2015). Evaluating the retrieval effectiveness of web search engines using a representative query sample. *Journal of the Association for Information Science and Technology, 66*(9), 1763-1775.

Liu, C., Zhao, S., & Volkovs, M. (2017). Unsupervised Document Embedding With CNNs. *arXiv preprint arXiv:1711.04168.*

Loohuis, A., & Chavannes, N. (2017). Medische apps: zorg voor de toekomst?. *Huisarts en wetenschap*, *60*(9), 440-443.

Manning, C., Raghavan, P., & Schütze, H. (2010). Introduction to information retrieval. *Natural Language Engineering, 16*(1), 100-103.

Maalej, W., Kurtanović, Z., Nabil, H., & Stanik, C. (2016). On the automatic classification of app reviews. *Requirements Engineering, 21*(3), 311-331.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research, 9*(Nov), 2579-2605.

Mayo Clinic. (2019). Deep vein thrombosis. Retrieved February 25, 2019, from https://www.mayoclinic.org/diseases-conditions/deep-vein-thrombosis/symptoms-causes/syc-20352557.

Medlock, S., Eslami, S., Askari, M., Arts, D. L., Sent, D., de Rooij, S. E., & Abu-Hanna, A. (2015). Health information–seeking behavior of seniors who use the internet: a survey. *Journal of medical Internet research*, *17*(1).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

Nederlands Huisartsen Genootschap. (2012). Cardiovasculair risicomanagement. Retrieved from https://www.nhg.org/?tmp-no-mobile=1&q=node/1803.

Opondo, D. O. (2018). *Electronic medical records and quality of prescriptions in general practice* (Doctoral dissertation). Retrieved from https://dare.uva.nl/search?identifier=6398b0e5-e131-4675-aa37-170631cc3f72.

Patel, M., Chokshi, A., Vyas, S., & Maurya, K. (2018). Machine Learning Approach for Automatic Text Summarization Using Neural Networks. *International Journal of Advanced Research in Computer and Communication Engineering, 7*(1).

Pathak, M. (2018). Hierarchical Clustering in R. Retrieved February 23, 2019, from https://www.datacamp.com/community/tutorials/hierarchical-clustering-R.

PBL Planbureau voor de Leefomgeving. (2013). Wat is de oorzaak van vergrijzing? Retrieved February 19, 2019, from https://www.pbl.nl/vraag-en-antwoord/wat-is-de-oorzaak-van-vergrijzing-0.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508.

Research2guidance. (2016). mHealth App Developer Economics 2016: the current status and trends of the mHealth app market.

Seif, G. (2018, October 2). An easy introduction to Natural Language Processing. Retrieved February 27, 2019, from https://towardsdatascience.com/an-easy-introduction-to-natural-language-processing-b1e2801291c1.

SemEval. (2019). SemEval-201: International Workshop on Semantic Evaluation. Retrieved from http://alt.qcri.org/semeval2019/index.php?id=tasks.

Sillo. (2019). Sillo Connect. Retrieved from https://www.siilo.com/nl/connect.

Snaterse, M. (2018). Rethinking management of risk factors in secondary prevention of cardiovascular disease. Retrieved from https://pure.uva.nl/ws/files/27041473/Thesis_complete_.pdf.

Van Bergen, J. (2016, Augustus 7). Persbericht: lancering van hart.volgers is nu officieel. Retrieved from https://hart.volgers.org/forum/onderwerp/223/.

Van den Elsen, W. (2016, March 29). VvAA: ‘Artsen overbehandelen vaker onder druk’. *Zorgvisie*. Retrieved from https://www.zorgvisie.nl/vvaa-artsen-overbehandelen-vaker-onder-druk/.

Van Oostrom, S. H., Gijsen, R., Stirbu, I., Korevaar, J. C., Schellevis, F. G., Picavet, H. S. J., & Hoeymans, N. (2017). *Toename in chronische ziekten en multimorbiditeit: veroudering van de bevolking verklaart maar een deel van de toename*. Retrieved from the website of NARCIS: https://www.narcis.nl/publication/RecordID/publicat%3A6654.

Volksgezondheidenzorg.info. (2017). Levensverwachting. Retrieved February 21, 2019, from https://www.volksgezondheidenzorg.info/onderwerp/levensverwachting/cijfers-context/trends#node-trend-levensverwachting-bij-geboorte.

Volksgezondheidenzorg.info. (2017). Overgewicht Internationaal. Retrieved from https://www.volksgezondheidenzorg.info/onderwerp/overgewicht/regionaal-internationaal/internationaal#node-internationale-vergelijking-overgewicht-volwassenen.

Voormolen, S. (2013, February 20). In het ziekenhuis verleert de dokter het luisteren. Retrieved February 19, 2019, from https://www.nrc.nl/nieuws/2013/07/20/in-het-ziekenhuis-verleert-de-dokter-het-luisteren-1272278-a658292.

Zamani, H., & Croft, W. B. (2017, August). Relevance-based word embedding. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 505-514). ACM.

Appendices

Appendix B: Functionalities of health-based social networks

Hart Volgers

A Dutch platform that has been developed by cardiologists of the Catherina Hospital in Eindhoven since 2014. Link: https://hart.volgers.org.

Category	**Functionalities**	**Explanation**
Moods	Navigate through recent moods.	Users can post how they feel (e.g. depressed or happy). Others can like and reply on moods, or share them on other social media.
Forum/blog posts	Navigate through recent questions.	Users can post, like and share messages. It is also possible to add discussion boards to favourites in order to receive updates when there are new online posts.
	Navigate through popular discussion boards.	
	Navigate through most watched discussion boards.	
	Navigate through recent posts on the forum.	
	Navigate through blog posts of healthcare professionals.	
Events	Navigate through events.	Users can post, like and share events. The events are sorted by date on the home page.
Social media	Navigate through recent posts on other social media (Twitter, Facebook, Instagram, Youtube).	Click on the posts to read them on the original website.
Medical information	Find information on cardiovascular diseases.	Articles are categorized by cardiovascular disease. Each article contains sections about the disease, common symptoms, diagnosis, treatment, and how the disease impacts daily life.
Polls	Navigate through recent polls.	Users can post, like and share polls. A poll contains a

		question with pre-defined answers. The community can vote for particular answer.
Search	Search through a list of community members or posts on the forum.	Find articles or posts based on keywords. Users can follow a person, then they will receive notifications of his or her online activity.

AFIB Online

A Dutch website that has been developed by Prof. Dr. Bianca Brundel (VUmc, Amsterdam) and Prof. Dr. Natasja de Groot (EMC, Rotterdam) since 2016. The number of active community members is unknown, but this number is estimated to be lower than the platform Hart Volgers based on the number of forum posts. Link: https://afiponline.org.

Category	**Functionalities**	**Explanation**
Participate in research	Navigate through medical research and sign up for an experiment.	Each research description shows what the research is about, who can participate, and how many participants have signed up.
Forum	Navigate through posts on the forum.	The forum is divided into four categories: 1) Diagnosis of AF, 2) Treatments, 3) Life with AF, 4) Females with AF. Users can start their own discussion boards, post messages, and reply to those of others.
	Navigate through new discussion boards.	
Donate	Navigate through medical research and make a donation.	Each research description shows what the research is about, how many backers support the research, how much is funded, and how much money is needed.
Medical information	Navigate through news articles about atrial fibrillation (AF).	News articles are ordered by date. Users can reply on these articles.
Search	Search through forum posts.	Find forum posts by searching on keywords of interest.

WebMD

One of the largest health-based platforms and most popular source of health information in the United States, which was founded in 1996. Link: https://www.webmd.com.

Category	Functionalities	Explanation
Symptom Checker	Gives an automatic diagnosis based on your symptoms.	The tool asks for your medical background and symptoms and generates a list of correlated diseases (rated by their correlation-strength). Each disease has a description and a link to a discussion board where people talk about this disease.
Forum	Navigate through posts on the forum.	The forum is divided into three categories: 1) family and pregnancy, 2) living healthy, 3) health conditions. People can start their own discussion boards and give them relevant tags. They can also reply to messages of others.
Medical information	Navigate through news articles about health.	Users can find informative articles about health conditions, family and pregnancy, living healthy and new scientific research. Users can also give these articles a rating.
Search	Search through forum posts.	Find forum posts by searching on keywords of interest.

PatientsLikeMe

An American platform where patients with a chronic condition can share their medical data and experiences. The platform has more than 600.000 community members and has been made public since 2011. Link: https://www.patientslikeme.com.

Category	Functionalities	Explanation
Forum	Navigate through the forum.	The forum is not publicly accessible, but it is known that they are categorized by each condition. Community members can also follow other members and specific discussion boards. They can rate each other’s comments.

Medical information	Navigate through medical research that is published by the organisation and general articles about medical conditions.	PatientsLikeMe collects self-reported data of patients and uses this data for a variety of research purposes. All members of the community can access these research files. For each condition, the website provides statistics about common symptoms reported by people with the condition and how severe these symptoms are. There are statistics about treatments for each disease. Each treatment is evaluated by the users on perceived effectiveness and side effects. One can also find statistics about age, gender and diagnosis status of community members. Each condition, symptom and treatment has its own web page with medical information and a link to a discussion board.
Search	Navigate through patients profiles with filters.	Patients can maintain detailed profiles about themselves, which they can share with the community or visitors of the website. Profiles contain a bio, age, gender, interests, conditions, treatments, symptoms, country and jobs. Each profile contains detailed charts about outcomes, symptoms and treatments plotted over time.

NewLifeOutlook

A Canadian platform with discussion boards and social media channels for people with a chronic condition. The platform has more than a million community members and was founded in 2014. Link: https://newlifeoutlook.com.

Category	Functionalities	Explanation
Patient stories	Read and publish articles about patient stories	The website has a standard format for stories, in which patients comprehensively answer specific questions about their AFib disease and post relevant photos. This format makes it unique from common forum posts. Users can reply to these stories.
Forum/social media	Navigate through posts on the forum	The forum is divided into five categories: 1) general, 2) news, 3) awareness, 4) lifestyle, 5) stories. People can start their own discussion boards and reply to messages of others. NewLifeOutlook is also connected to discussion groups on Facebook.
Medical information	Navigate through news articles about AFib	Users can find informative articles about symptoms, risks and causes of AFib. There is also information about living a healthy lifestyle, coping with AFib and treatments. Users can give articles a rating.
Search	Search through forum posts and articles	Find forum posts by searching keywords of interest.

DailyStrength

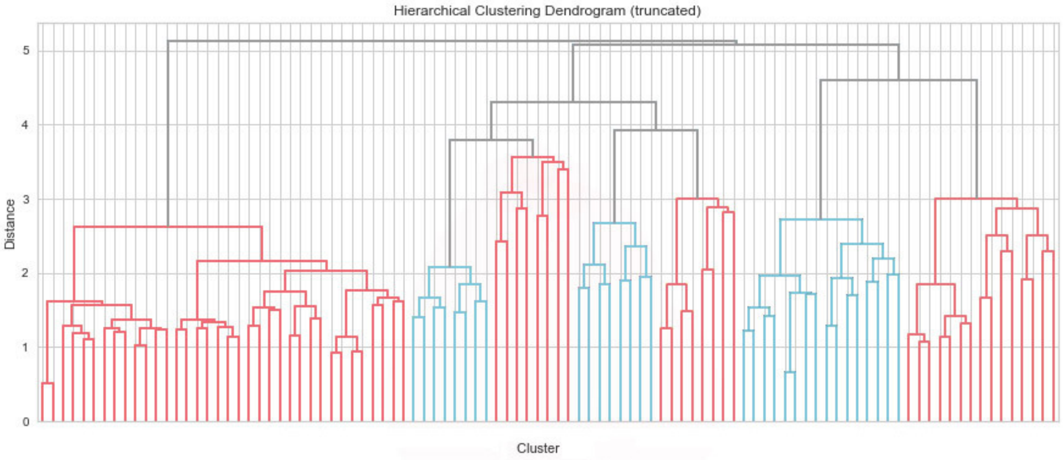
An American platform that is focused on support groups where patients can discuss their struggles and successes with each other. The platform has more than 14,000 daily visitors and was founded in 2006. Link: https://www.dailystrength.org.

Category	Functionalities	Explanation
Forum	Navigate through forum posts by community members.	The forum is divided into two groups: groups created by community members

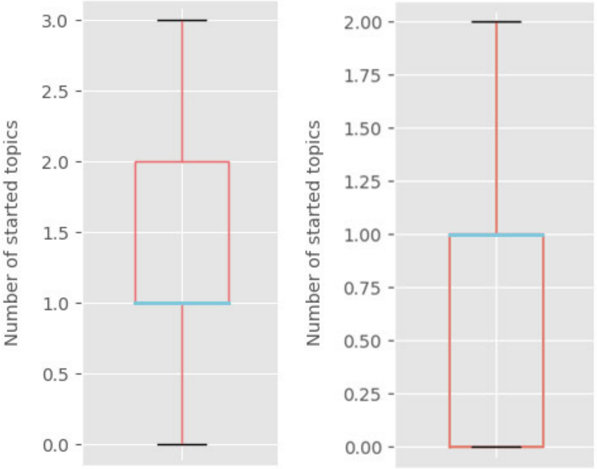
	Assist others as a community leader.	(community groups) and groups created by moderators (support groups). The community can join these groups, post messages, like posts of others. Each group contains various discussion boards which are categorized by health conditions. It is noticeable that DailyStrength also recruits community members as community leaders, when they have exhibited continually sound judgment and a high level of positive support for other members. Community leaders help the administrators with maintenance and managing the platform.
Search	Search through forum posts.	Find forum posts by searching on keywords of interest.

Appendices

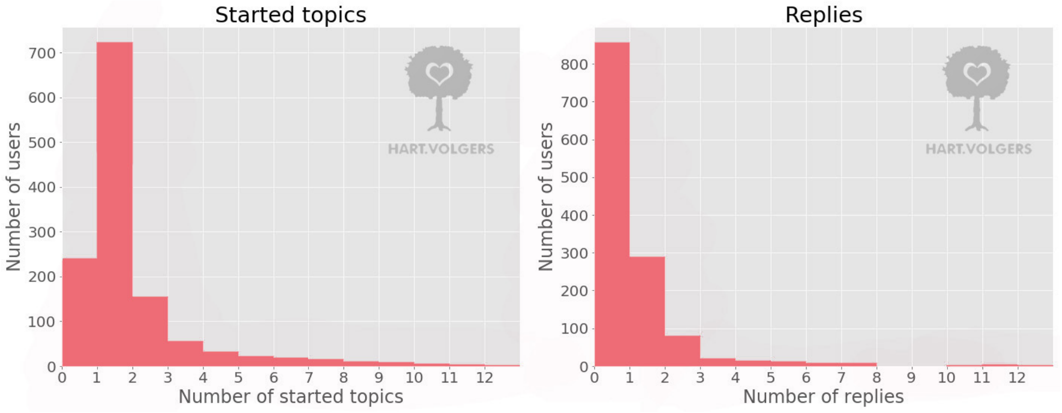
Appendix C: Additional figures of data analysis



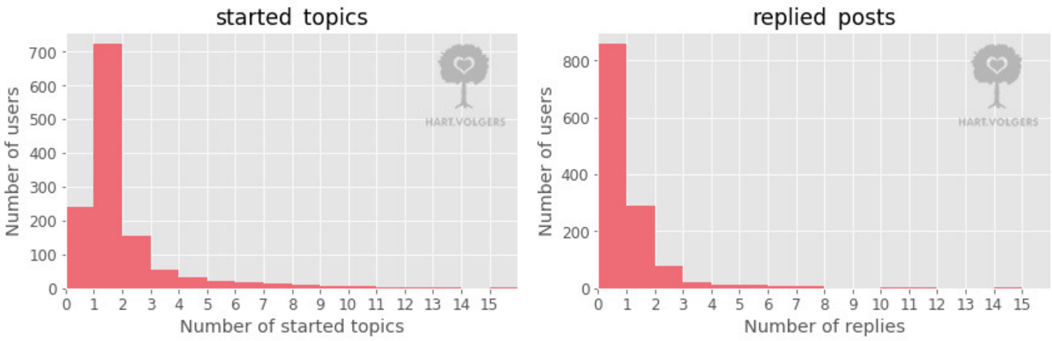
A (truncated) dendrogram of the clustering of discussion boards on Hart Volgers. Each split of a 'tree branch' is a moment when two cluster are merged together (when reading the dendrogram from the bottom to the top). The x-axis shows how many discussion boards each branch contains. The y-axis shows the distance or dissimilarity between clusters: the greater the dissimilarity between clusters, the longer the vertical branch. If a horizontal line would be drawn through the dendrogram (which truncates the dendrogram), then the number of clusters can be determined (7 clusters in this case, separated by colour).



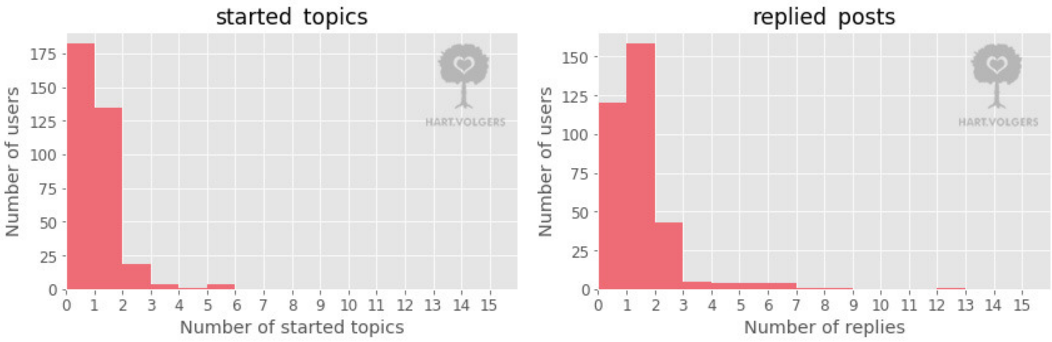
Tukey boxplot analysis of Hart Volgers (Left) and DailyStrength (Right), which shows the distribution of the number of started topics (i.e. discussion boards) per patient. Outliers are not visualized in these boxplots.



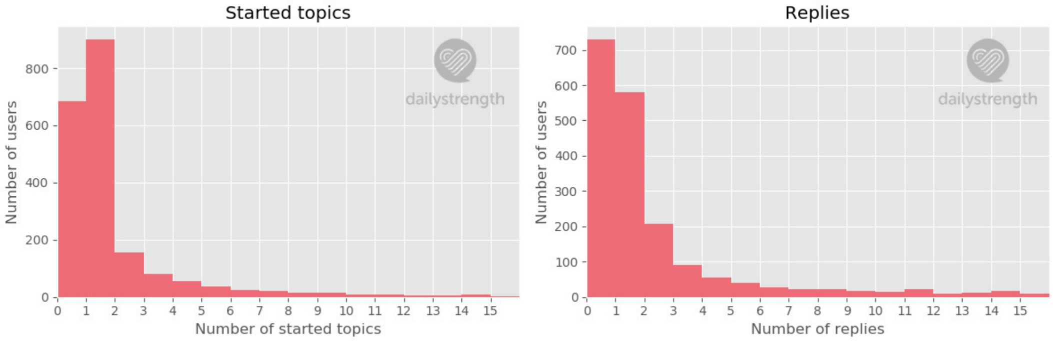
The distribution of users over the number topics that patients start on the platform Hart Volgers (left) and the distribution of the number of replies that patients give on topics which are not started by themselves (right).



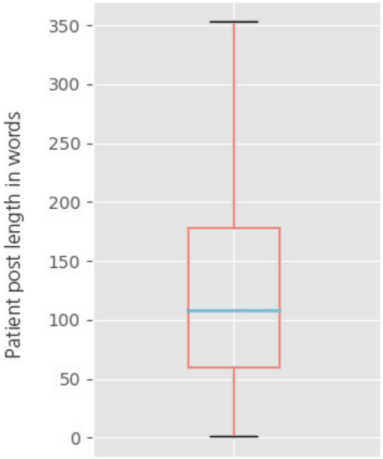
The distribution of users over the number topics that patients start *inside the "AskUs"* category on the platform Hart Volgers (left) and the distribution of the number of replies that patients give on topics which are not started by themselves (right).



The distribution of users over the number topics that patients start *outside the "AskUs"* category on the platform Hart Volgers (left) and the distribution of the number of replies that patients give on topics which are not started by themselves (right).



The distribution of users over the number topics that patients start on the platform DailyStrength (left) and the distribution of the number of replies that patients give on topics which are not started by themselves (right).



Tukey boxplot analysis of Hart Volgers, which shows the distribution of the number of words per patient post in the “Ask Us” category. Only the first patient post per discussion board was taken into account. The outliers are not visualized in this boxplot.

Appendices

Appendix D: Labelling protocol

Introduction

I am Ward Hendrix and I am a graduate Industrial Design student at CardioLab at TU Delft. My final master project is about health-based social networks, which are publicly accessible websites where cardiac patients can ask questions to healthcare professionals (i.e. cardiologists and cardiac surgeons). The problem is that an increasingly large group of cardiac patients ask questions to healthcare professionals on these platforms, while there is only a small group of healthcare professionals available. Therefore, it becomes increasingly difficult to provide the same quality of care to everyone. It should be noted that healthcare professionals do this work voluntarily and aim to inform cardiac patients, because time at the hospital is very limited. The better patients are informed before their visit at the hospital, the easier patients and cardiologists can communicate and more time remains for discussing specific questions.

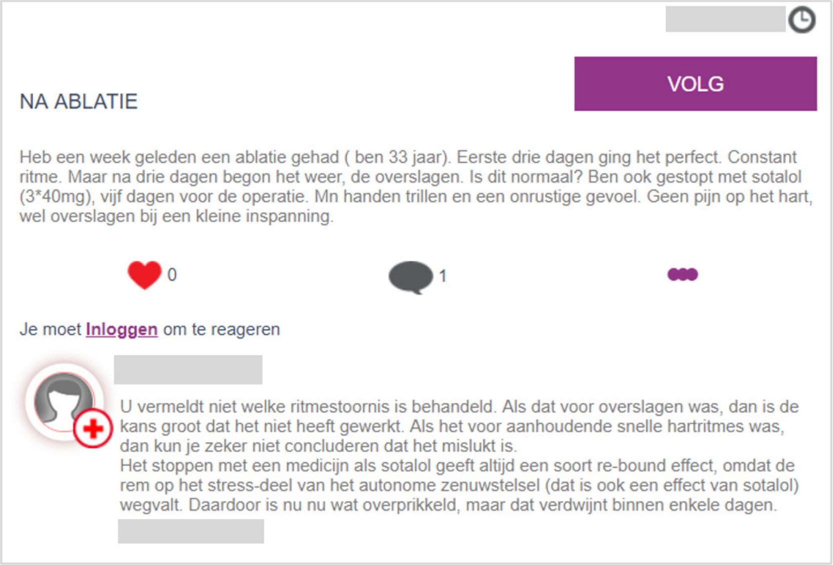
As a way to unburden healthcare professionals, I have designed a new platform that enables patients to quickly find relevant information. In this way, it is not always necessary to ask a question and to wait for a reply. A very important part of my design is a novel search engine, which incorporates a machine learning algorithm that captures the semantics of words (e.g. when searching for “cardiologist”, one also receives results with “specialist” or “doctor”). It can deal with short questions and it is no longer required to enter keywords only. I want to compare the performance of my model with traditional information retrieval systems, but a labelled dataset is required for this evaluation. In the next sections, I will explain the labelling procedure.

Labelling procedure

I have scraped (i.e. automatically received) all patient posts from the Dutch platform Hart Volgers (https://hart.volgers.org/). All the posts are publicly accessible and no account is required to read them. The overall procedure is to visit a web page on Hart Volgers and to extract the question from the patient post. When a large collection of web pages is labelled, then I can use the questions as search queries for the system. Consequently, very similar questions can be grouped together as a single label (a task that is excluded from this protocol). In the end, every web page is labelled with a question, and every question has a label. Now it is possible to enter a search query in the system and measure how well it returns relevant web pages in terms of ranking and precision (i.e. the fraction of obtained search results that are relevant).

Steps for extracting questions:

1. You should have received a CSV file called “dataset_1_part_#.csv”. Open the CSV file with Microsoft Excel (or download Notepad++ if you do not have Excel: https://notepad-plus-plus.org/download/v7.6.4.html).
2. Copy-and-paste an URL from the CSV file in your web browser. For example, https://hart.volgers.org/forum/onderwerp/3816/
3. Read the original post and the first answer of the cardiologist. In this example, the web page looks like this:



NA ABLATIE

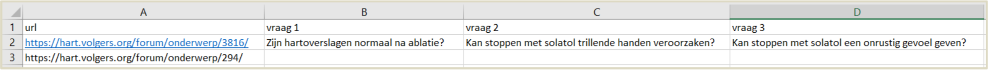
Heb een week geleden een ablatie gehad (ben 33 jaar). Eerste drie dagen ging het perfect. Constant ritme. Maar na drie dagen begon het weer, de overslagen. Is dit normaal? Ben ook gestopt met sotalol (3*40mg), vijf dagen voor de operatie. Mn handen trillen en een onrustige gevoel. Geen pijn op het hart, wel overslagen bij een kleine inspanning.

Je moet Inloggen om te reageren

U vermeldt niet welke ritmestoornis is behandeld. Als dat voor overslagen was, dan is de kans groot dat het niet heeft gewerkt. Als het voor aanhoudende snelle hartritmes was, dan kun je zeker niet concluderen dat het mislukt is.
Het stoppen met een medicijn als sotalol geeft altijd een soort re-bound effect, omdat de rem op het stress-deel van het autonome zenuwstelsel (dat is ook een effect van sotalol) wegvalt. Daardoor is nu nu wat overprikkeld, maar dat verdwijnt binnen enkele dagen.

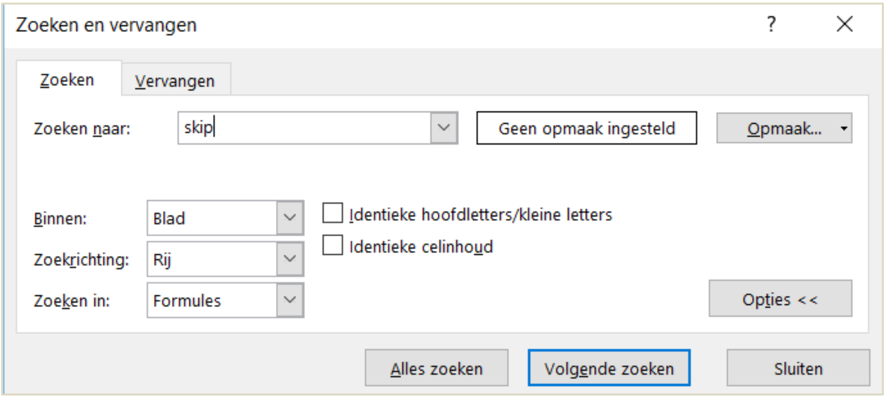
4. Check if the post is valid with help of the guidelines in the following section. If a post is invalid, write down “skip”.
5. Extract all questions from the patient post. For this example, one could extract the following question (in Dutch): “Zijn hartoverslagen normaal na ablatie?”. Cardiac ablation is a surgical procedure to treat atrial fibrillation (i.e. irregular heartbeats), so the patient wonders if it is common that the irregular heartbeats return or that the procedure has failed. However, the post also contains an indirect question, which can be reformulated in this way: “Kan stoppen met solatol trillende handen veroorzaken en een onrustig gevoel geven?”. Even better would be to split this question, because it contains two ideas: “Kan stoppen met solatol trillende handen veroorzaken?” and “Kan stoppen met solatol een onrustig gevoel geven? “. It appears that this patient post contains three questions after all. It can be useful to look at the doctor’s answer when indirect questions are involved, because he or she will respond to these questions if they truly need attention. Posts usually contain a lot of jargon words, use Google when in doubt.
6. Write down the questions in Dutch in the CSV file. Each question should be placed in a separate column. Regarding the notation of the questions, please try not to use special characters (e.g. @#%&) and especially avoid the semicolon “;” (which

separates the values in a csv file). Furthermore, please keep the questions concise (i.e. they should reflect true search queries). The CSV file should look like this:



	A	B	C	D
1	url	vraag 1	vraag 2	vraag 3
2	https://hart.volgers.org/forum/onderwerp/3816/	Zijn hartoverslagen normaal na ablatie?	Kan stoppen met solatol trillende handen veroorzaken?	Kan stoppen met solatol een onrustig gevoel geven?
3	https://hart.volgers.org/forum/onderwerp/294/			

7. Repeat the previous five steps until 100 URLs are done.
8. Press CTRL+F to search and count all "skip" instances. When pressing CTRL+F, one should the following window (click on "opties" to show everything):



Zoeken en vervangen

Zoeken naar: skip

Binnen: Blad

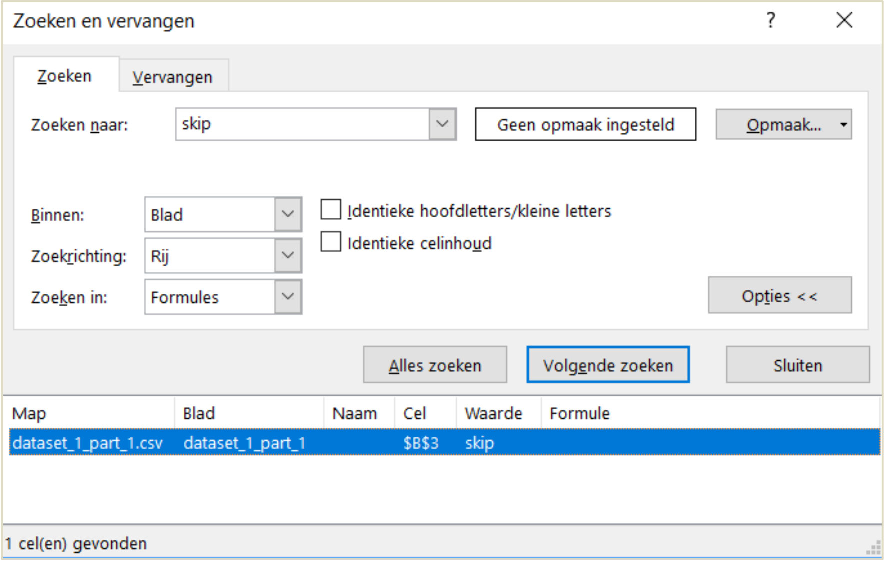
Zoekrichting: Rij

Zoeken in: Formules

☐ Identieke hoofdletters/kleine letters

☐ Identieke celinhoud

9. Type "Skip" and click on "Alles zoeken" to count all invalid instances. The following window should appear:



Zoeken en vervangen

Zoeken naar: skip

Binnen: Blad

Zoekrichting: Rij

Zoeken in: Formules

☐ Identieke hoofdletters/kleine letters

☐ Identieke celinhoud

Map	Blad	Naam	Cel	Waarde	Formule
dataset_1_part_1.csv	dataset_1_part_1		B3	skip	

1 cel(en) gevonden

10. At the bottom, the total number of found instances is shown: "# cel(en) gevonden". Write down this number on a piece of paper. If you work with Notepad++, press CTRL+F and click on "count" straightaway.

11. Continue until 100 web pages have been labelled with questions, excluding “skip” rows.

Skipping posts

Extracting questions from a patient post can be difficult task. For example, questions are often preceded by a long, medical history of the patient. To simplify this task, it is recommended to completely skip a post in the following cases:

- The patient asks for advice for his or her personal situation. It is not possible to extract question, because the question cannot be interpreted without the given context. These posts usually end with “What should I do?”.
- The post cannot be answered by a cardiologist, because:
    - Information from the patient’s medical record would be needed to answer the question.
    - The patient requests a diagnosis, which cardiologists are not allowed to give on health-based social networks.
    - The user only intended to share his or her experience with other patients, so it belongs to the wrong forum category. This post should not have been addressed to a cardiologist, and the cardiologist will let the patient know if this is the case (e.g. “Please put your post on the forum.”).
- The post is missing essential information, because the patient refers to another post.
- The post contains too many questions. Only posts with a maximum of three questions should be considered. If the original questions can be split into more than three questions (because they contain multiple ideas), please skip the entire patient post as well.
- The post contains clear questions, but some of them cannot be answered by the cardiologist or are related to the platform itself
- The post is too long, as it contains more than 350 words.

Deliverable

The deliverable is one semicolon-separated CSV file that contains questions per URL. The dataset contains four columns: the first column contains the URLs, and other three contain “Questions 1”, “Questions 2” and “Questions 3”. Each row represents an unique web page.

Appendices

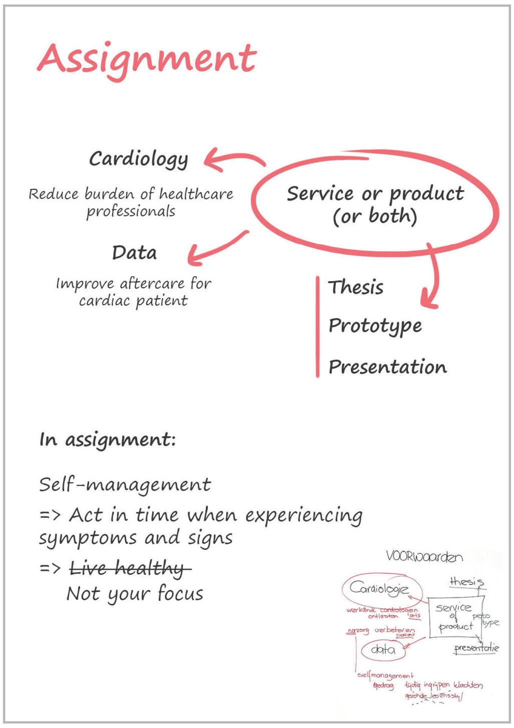
Appendix E: Workshops

Synthesis workshop

Description

The goal of this workshop was to combine the most important findings from the literature research and data analysis (see Chapter 2) and use them to update the project assignment (see Chapter 1). Three Dutch participants joined the workshop, who all had a Master's degree, but were unfamiliar with design practices. The project findings were explained to the participants and then the following topics were discussed: project assignment, user profiles, envisioned solution, and benefits. The results of the workshop are summarized in the figures below. These results were helpful to define the design requirements and questions for the brainstorm workshop. The overall outcome of this workshop was the decision to only focus on the question-and-answer service of health-based social networks in order to reduce the scope of the design project.

Results



Assignment

Cardiology

Reduce burden of healthcare professionals

Service or product (or both)

Data

Improve aftercare for cardiac patient

Thesis

Prototype

Presentation

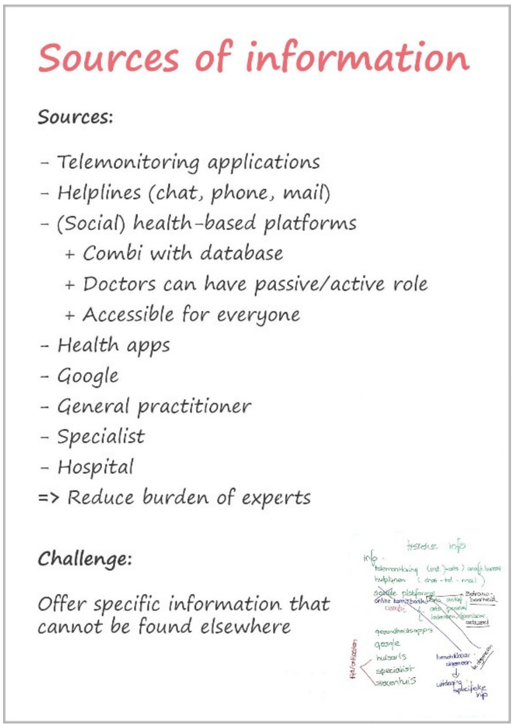
In assignment:

Self-management

=> Act in time when experiencing symptoms and signs

=> ~~Live healthy~~

Not your focus



Sources of information

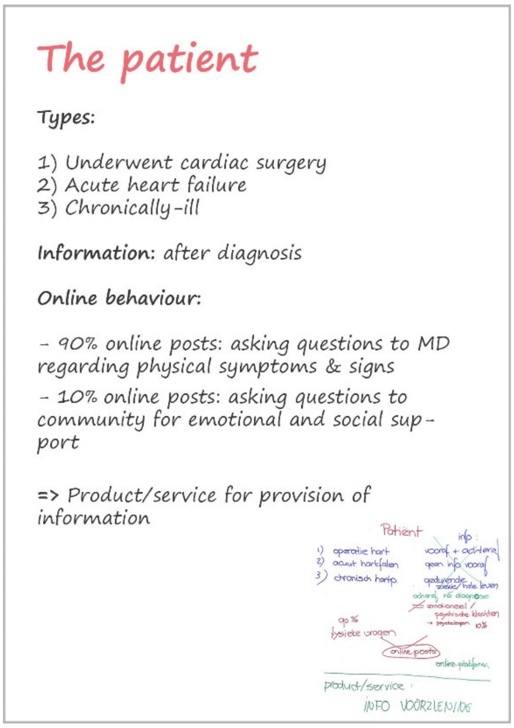
Sources:

- Telemonitoring applications
- Helplines (chat, phone, mail)
- (Social) health-based platforms
  - + Combi with database
  - + Doctors can have passive/active role
  - + Accessible for everyone
- Health apps
- Google
- General practitioner
- Specialist
- Hospital

=> Reduce burden of experts

Challenge:

Offer specific information that cannot be found elsewhere



The patient

Types:

1) Underwent cardiac surgery
2) Acute heart failure
3) Chronically-ill

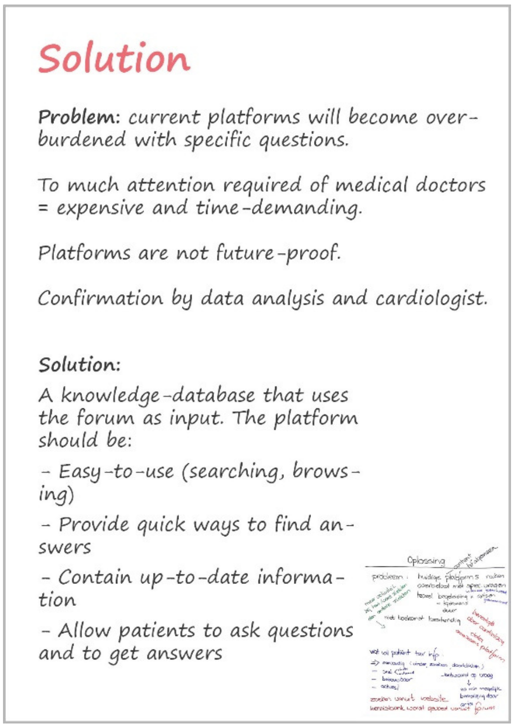
Information: after diagnosis

Online behaviour:

- 90% online posts: asking questions to MD regarding physical symptoms & signs
- 10% online posts: asking questions to community for emotional and social support

=> Product/service for provision of information

Patiënt
1) operatie hart
2) acuut hartfalen
3) chronisch hartp.
info:
vooraf + achteraf
geen info vooraf
gedurende ziekte/hele leven
achteraf nà diagnose
≠ emotioneel / psychische klachten
→ psychologen 10%
90%
fysieke vragen
online posts
online platform
product/service:
INFO VOORZIENING



Solution

Problem: current platforms will become overburdened with specific questions.

To much attention required of medical doctors = expensive and time-demanding.

Platforms are not future-proof.

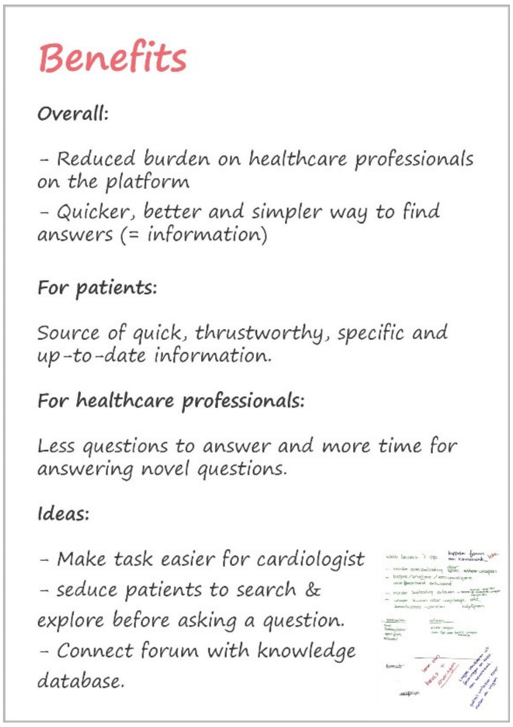
Confirmation by data analysis and cardiologist.

Solution:

A knowledge-database that uses the forum as input. The platform should be:

- Easy-to-use (searching, browsing)
- Provide quick ways to find answers
- Contain up-to-date information
- Allow patients to ask questions and to get answers

Oplossing
probleem:
huidige platforms raken overbelast met spec. vragen
teveel begeleiding + artsen
= kostbaar
duur
niet toekomst bestendig
bevestigd door cardioloog
data
analyse platform
wat wil patiënt tav info
=> eenvoudig (vinden, zoeken, doorklikken)
– snel
– betrouwbaar
– actueel
– antwoord op vraag
zo min mogelijk belasting door arts
zoeken vanuit website
kennisbank wordt gevoed vanuit forum



Benefits

Overall:

- Reduced burden on healthcare professionals on the platform
- Quicker, better and simpler way to find answers (= information)

For patients:

Source of quick, thrustworthy, specific and up-to-date information.

For healthcare professionals:

Less questions to answer and more time for answering novel questions.

Ideas:

- Make task easier for cardiologist
- seduce patients to search & explore before asking a question.
- Connect forum with knowledge database.

Brainstorm workshop

Description

The goal of the brainstorm workshop was to gather ideas for concept development. Five Dutch Industrial Design master students from TU Delft joined the workshop, whose expertise ranged from technical product design to interaction design. The participants were introduced to the design project by means of a short presentation and they were asked to generate ideas that addressed the following questions:

1. How can we seduce patients to go through medical articles and forum posts before asking a question to a cardiologist?
2. How can we make the task of answering patient questions as easy as possible for the cardiologist?
3. How can we increase patient involvement on the forum, despite the issues associated with trust, expertise and responsibility?

Next, the ideas for each of these questions were clustered based on two criteria: originality and feasibility. After the clustering procedure, the participants had to vote for the best ideas (which score high on both originality and feasibility) and refine a selection of those ideas in pairs of two. At the end of the workshop, the participants had to present their ideas for a group discussion.

The overall outcome of the workshop was that the participants thought that the organisation of information was the most important aspect in solving the central problem of the thesis. This was the overarching theme of the ideas for all three brainstorm questions and this was also reflected by their presentations. A well organised platform invites patients to explore its content (and so they become more involved and informed) and helps cardiologists to assess incoming patient posts. Smart technology, such as machine learning, can be used to ensure easy information retrieval by users. The technical details of the implementation of smart technology is explained in Chapter 3.

Schedule

Time	**Activity**
19:00 – 19:30	*Introduction (30 min)* • A short presentation to show the planning and goal of the workshop and to introduce the problem statement of the project (15 min). • Discuss (sub-) design questions with the group, which are formulated in advance. Final moment to reformulate these questions (10 min). • Q&A (5 min).
19:30 – 20:30	*Idea generation (60 min)* • Brainstorming individually, one design question per participant (with sticky notes). Circulate sheets with design question and sticky notes among the participants. • Discuss the ideas on each sheet with the group and generate new ideas together.
20:30 - 20:45	Break (15 min)
20:45 – 21:15	*Idea selection (30 min)* • Cluster ideas based on originality and feasibility. • Participants can vote for the best three ideas. Ideas can be combined during this phase. • Three ideas with the most votes are selected for refinement.
21:15 – 22:00	*Idea refinement (45 min)* Participants are grouped together in pairs of two. They work on a single design and work out the details.
22:00 – 22:20	*Presentations (20 min)* • Each group presents their designs. • A short discussion after each presentation.
22:20 – 22:30	*Reflection and conclusions (10 min)* • Reflect on the workshop: how did it go? • Draw conclusions from the reflection and design presentations.

Photos

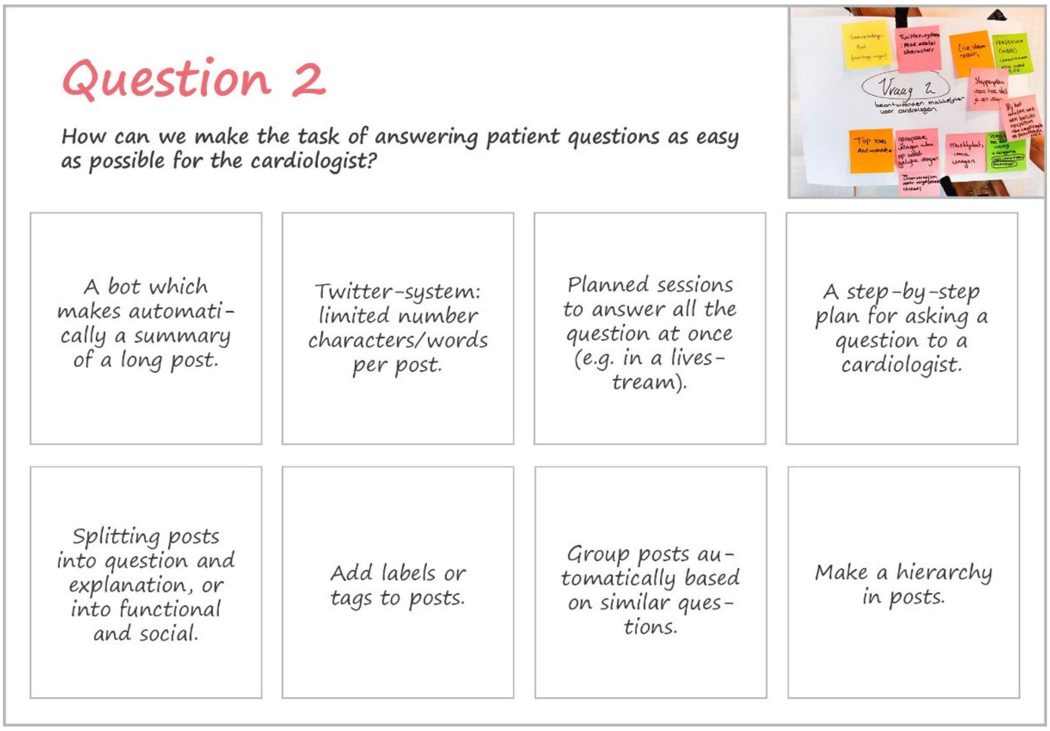
Ideas per brainstorm question

Question 1

How can we seduce patients to go through medical articles and forum posts first before asking a question to a cardiologist?

Vraag 1
Verleiden tot lezen
Persoonlijke Chatbot?
Kortere vragen
tekst opmaak hulp
Alleen de basics, geen onnodige crap
laten zien of een vraag is beantwoord (met vinkje)
videos van info / voorlichtingen
Vraag wizard
Vragen posten per onderwerp
Forums koppelen aan artikelen
Algemene zoek functie

Refer to specific parts of the forum in medical articles and the other way around.	A human-like assistent: a chatbot.	Question wizard or decision tree.	Show related questions on the forum when typing your own question.
Simplify the homepage: only show what patients want to see.	Show posts per day, week, month or all time.	Present forum questions or medical information in a video.	Swipe posts to the left or right in order to personalize your account (similar to Tinder).
Motivate patients to formulate shorter questions.	Let patients label their posts with a category.	Show if a question is already answered or not.	Help patients with the layout of their posts.



Question 2

How can we make the task of answering patient questions as easy as possible for the cardiologist?

A bot which makes automatically a summary of a long post.	Twitter-system: limited number characters/words per post.	Planned sessions to answer all the question at once (e.g. in a livestream).	A step-by-step plan for asking a question to a cardiologist.
Splitting posts into question and explanation, or into functional and social.	Add labels or tags to posts.	Group posts automatically based on similar questions.	Make a hierarchy in posts.

Question 3

How can we increase patient involvement on the forum, despite the issues associated with trust, expertise and responsibility?

'Humanize' the navigation on the website: search on experiences instead of technical terms.

Make post wall per forum category (just like Facebook).

Reputation system: patients with helpful posts obtain higher status.

Helpful replies on the forum can be up voted.

Personalize content on the website for the patient.

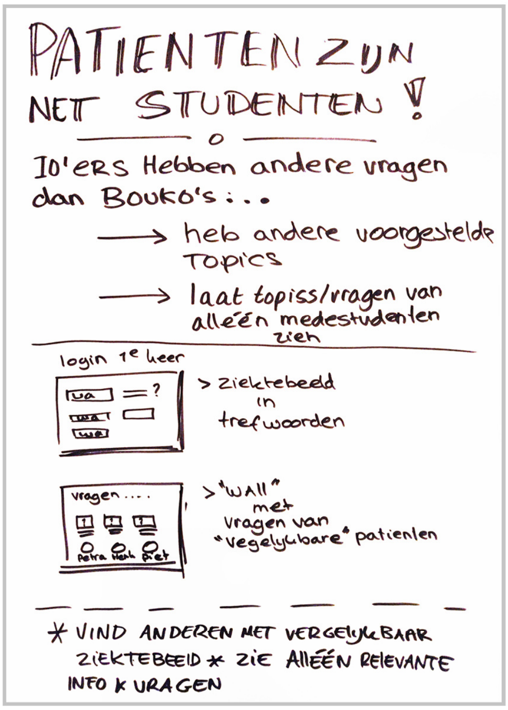
Buddy program: group patients with same interests or live in same neighbourhood.

Virtual Reality chat.

Force patients to respond to other messages before asking their question to an doctor.

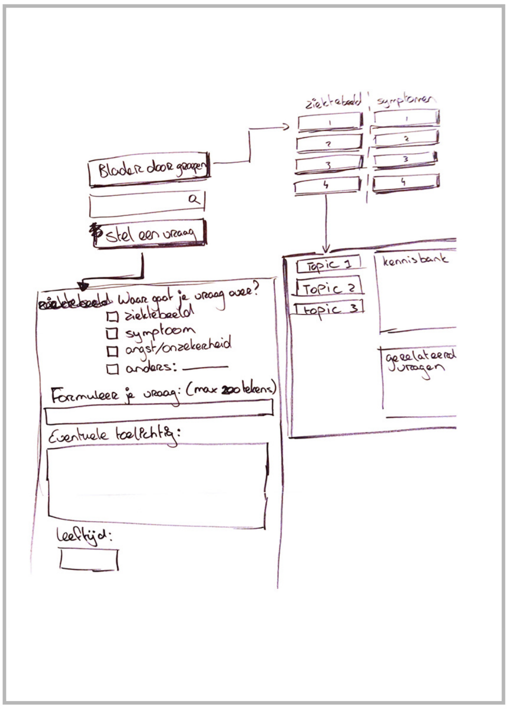
Add labels or tags to posts.

Move answered questions to knowledge-database.


PATIENTEN ZIJN NIET STUDENTEN !
IO'ERS Hebben andere vragen dan BOUKO's: . .
→ heb andere voorgestelde Topics
→ laat topiss/vragen van alleéén medestudenten zien
login 1e keer
> ziektebeeld in trefwoorden
vragen....
Petra Henk Piet
> "WALL" met vragen van "vegelijkbare" patienlen
* VIND ANDEREN MET VERGELIJKBAAR ZIEKTEBEELD * ZIE ALLEÉN RELEVANTE INFO & VRAGEN

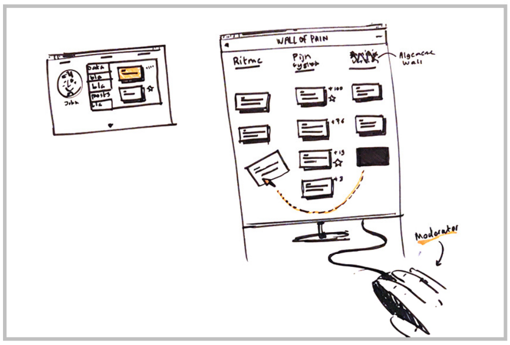
Idea 1: Personalised patient environment

Only show personalized content to the user, that is, posts from patients who are similar to the user. Initial set-up: describe the condition in keywords to personalize content.



Idea 2: Question wizard

Three options: search, navigate through groups, or ask a question. When navigating through groups: show cardiovascular diseases and their symptoms side-by-side. Per group: medical information and related posts on the forum.



Idea 3: Wall of labeled posts

Patients label their posts: (1) All posts per category are shown on a single wall, (2) the moderator can make adjustments to the categorization of these posts, (3) patients are notified when their posts are re-labeled.

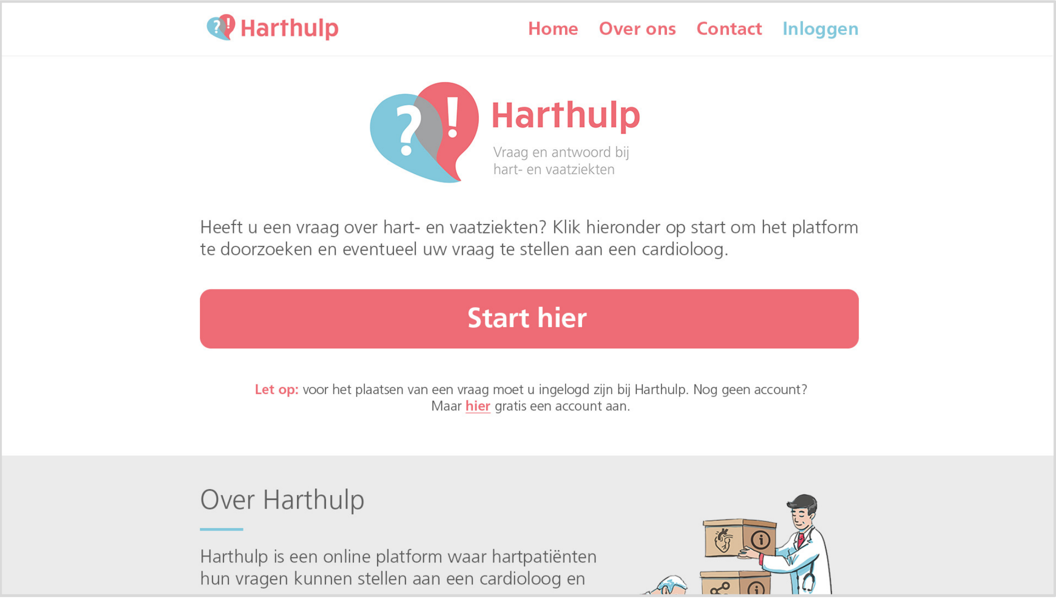
LIVE
LIVE RESPONSES:
Brood + chips
19226
ik278
Arts
A
B
C

Idea 4: Live streaming sessions

A single livestream session where patients can ask questions to experts. There will be weekly livestream sessions on the website with different themes.

Appendices

Appendix F: Screenshots of web interface



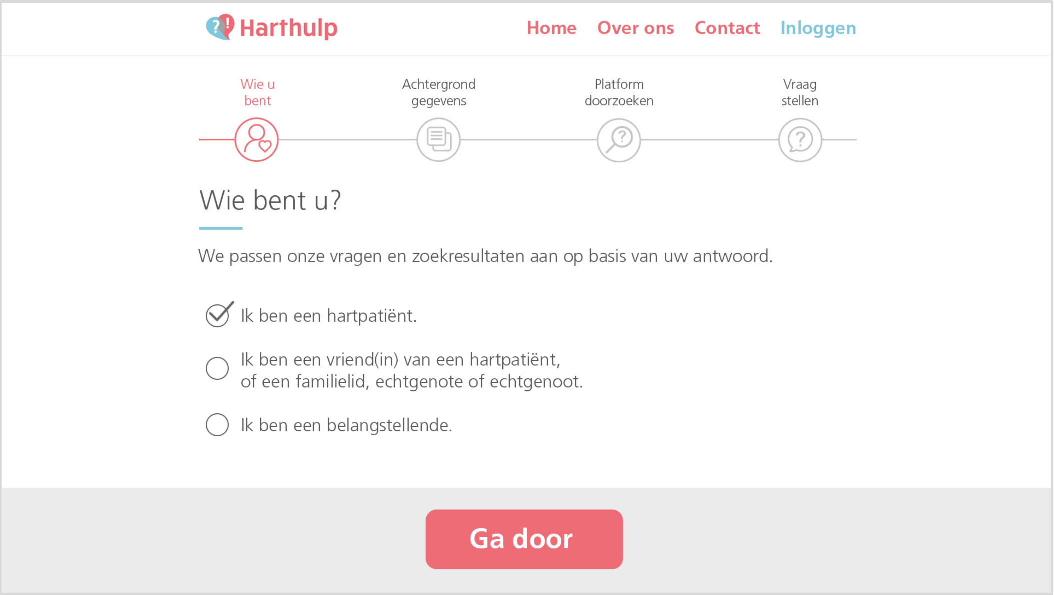
Heeft u een vraag over hart- en vaatziekten? Klik hieronder op start om het platform te doorzoeken en eventueel uw vraag te stellen aan een cardioloog.

Let op: voor het plaatsen van een vraag moet u ingelogd zijn bij Harthulp. Nog geen account? Maar hier gratis een account aan.

Over Harthulp

Harthulp is een online platform waar hartpatiënten hun vragen kunnen stellen aan een cardioloog en

Home screen: Visitors have to click on *start* to open the question wizard.



Wie bent u?

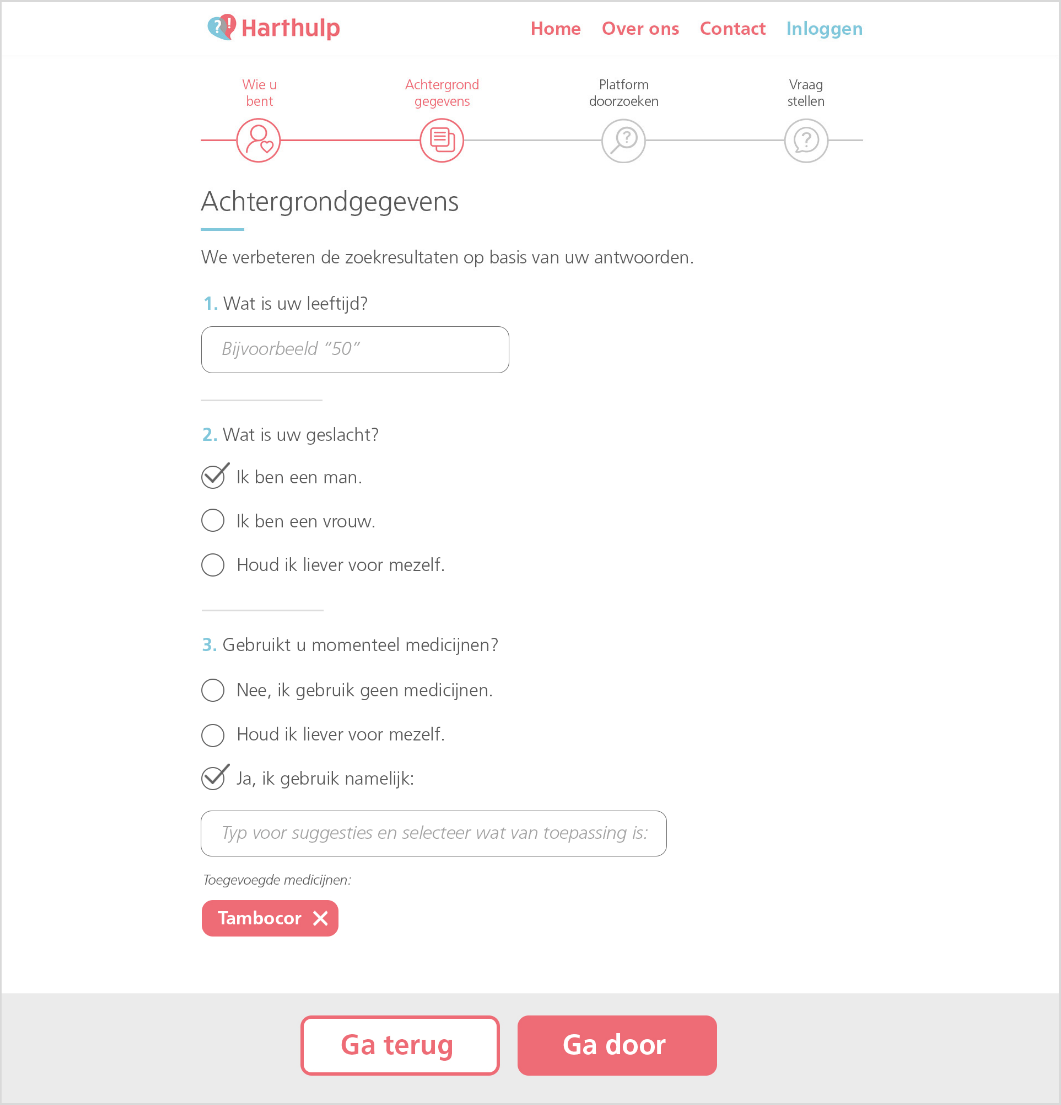
We passen onze vragen en zoekresultaten aan op basis van uw antwoord.

☑ Ik ben een hartpatiënt.

○ Ik ben een vriend(in) van een hartpatiënt,
of een familielid, echtgenote of echtgenoot.

○ Ik ben een belangstellende.

Question wizard step 1: check the identity of the user. Follow-up questions will be rephrased and search results will be improved based on the user's answer.



Achtergrondgegevens

We verbeteren de zoekresultaten op basis van uw antwoorden.

1. Wat is uw leeftijd?

Bijvoorbeeld "50"

2. Wat is uw geslacht?

☑ Ik ben een man.

○ Ik ben een vrouw.

○ Houd ik liever voor mezelf.

3. Gebruikt u momenteel medicijnen?

○ Nee, ik gebruik geen medicijnen.

○ Houd ik liever voor mezelf.

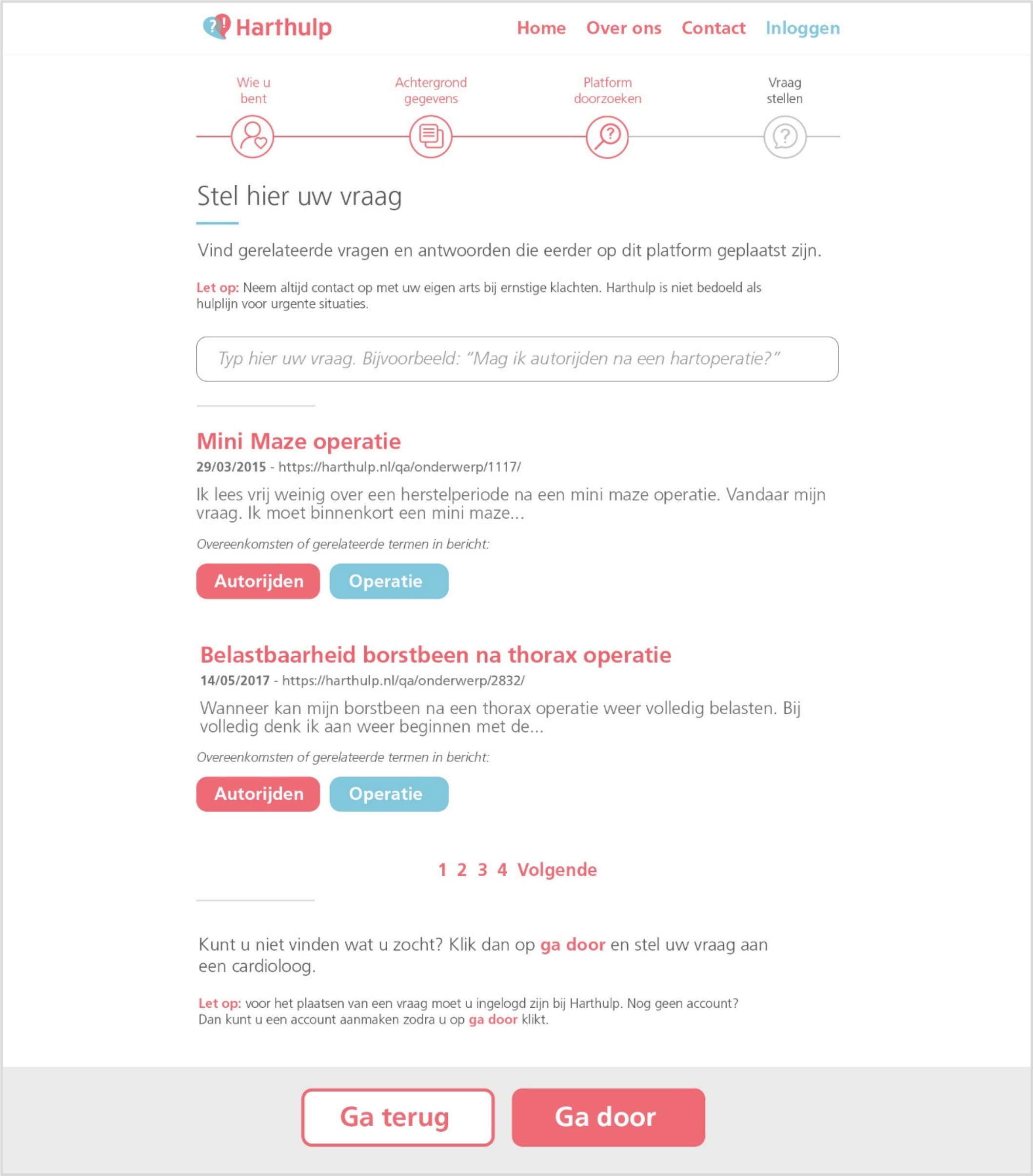
☑ Ja, ik gebruik namelijk:

Typ voor suggesties en selecteer wat van toepassing is:

Toegevoegde medicijnen:

Tambocor ✕

Question wizard step 2: ask background questions. Background questions can be used to improve the search results and to better inform cardiologists. When users are not cardiac patients or when they have an account on Harthulp, then they can skip this step (background information is already saved in their user profiles).



Stel hier uw vraag

Vind gerelateerde vragen en antwoorden die eerder op dit platform geplaatst zijn.

Let op: Neem altijd contact op met uw eigen arts bij ernstige klachten. Harthulp is niet bedoeld als hulplijn voor urgente situaties.

Typ hier uw vraag. Bijvoorbeeld: "Mag ik autorijden na een hartoperatie?"

Mini Maze operatie

29/03/2015 - https://harthulp.nl/qa/onderwerp/1117/

Ik lees vrij weinig over een herstelperiode na een mini maze operatie. Vandaar mijn vraag. Ik moet binnenkort een mini maze...

Overeenkomsten of gerelateerde termen in bericht:

Belastbaarheid borstbeen na thorax operatie

14/05/2017 - https://harthulp.nl/qa/onderwerp/2832/

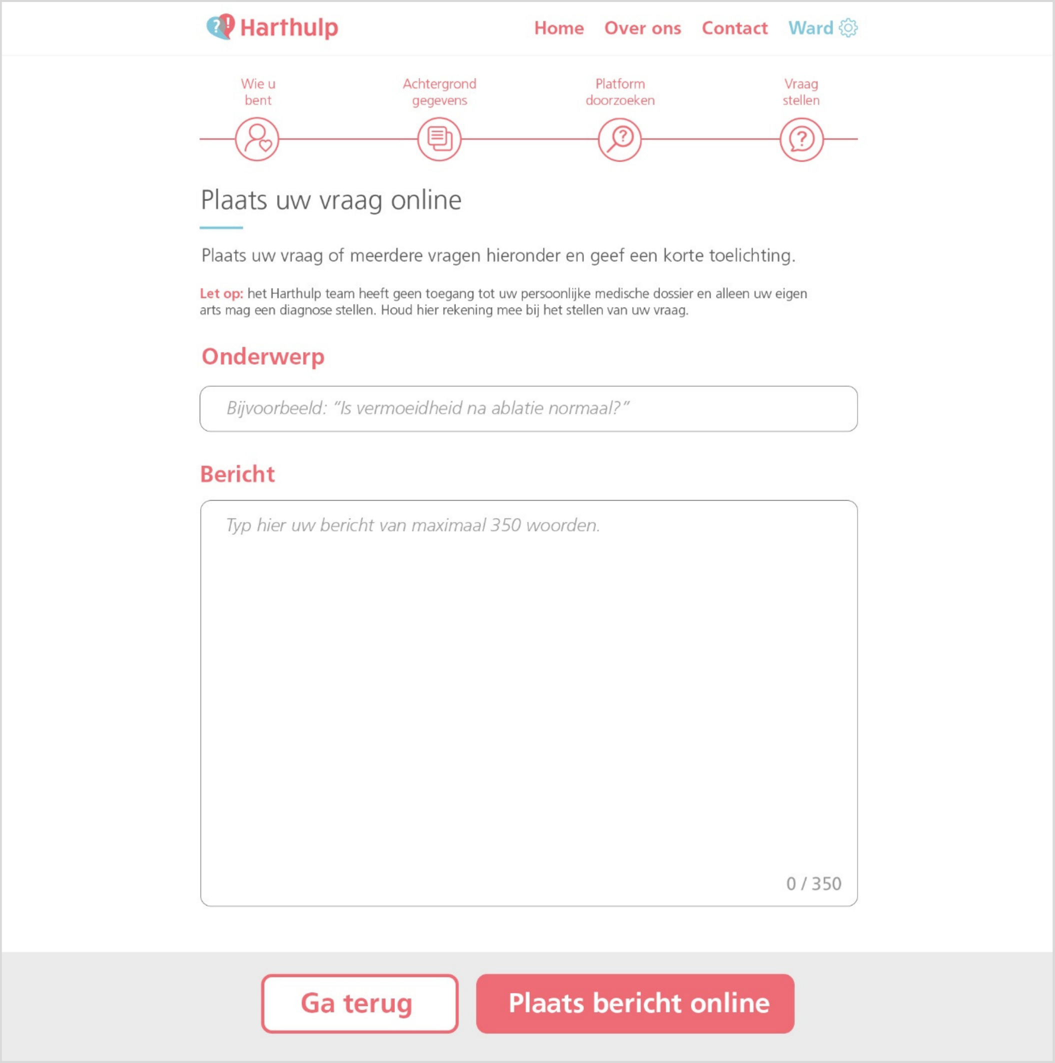
Wanneer kan mijn borstbeen na een thorax operatie weer volledig belasten. Bij volledig denk ik aan weer beginnen met de...

Overeenkomsten of gerelateerde termen in bericht:

Kunt u niet vinden wat u zocht? Klik dan op **ga door** en stel uw vraag aan een cardioloog.

Let op: voor het plaatsen van een vraag moet u ingelogd zijn bij Harthulp. Nog geen account? Dan kunt u een account aanmaken zodra u op **ga door** klikt.

Question wizard step 3: allow users to search. Harthulp incorporates a novel search algorithm that uses deep learning to capture the semantics of words, and filters the most meaningful words from (long) patients posts. When the search results do not contain the information that the user is looking for, then they can ask a question to a cardiologist. A visitor needs an account for posting questions on the platform.



Plaats uw vraag online

Plaats uw vraag of meerdere vragen hieronder en geef een korte toelichting.

Let op: het Harthulp team heeft geen toegang tot uw persoonlijke medische dossier en alleen uw eigen arts mag een diagnose stellen. Houd hier rekening mee bij het stellen van uw vraag.

Onderwerp

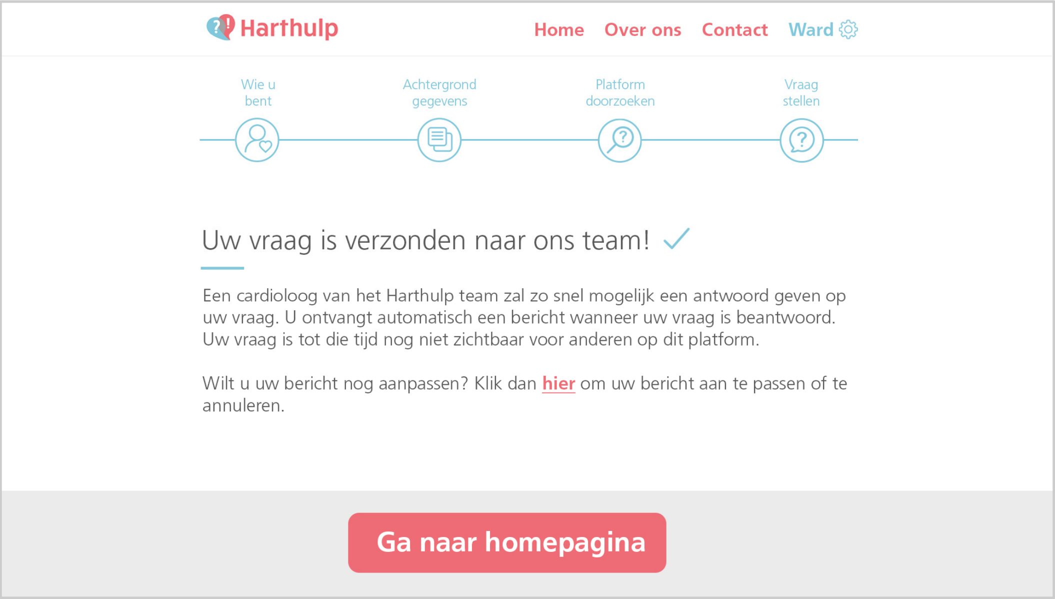
Bijvoorbeeld: "Is vermoeidheid na ablatie normaal?"

Bericht

Typ hier uw bericht van maximaal 350 woorden.

0 / 350

Question wizard step 4: optionally, post a question on Harthulp. The length of a post cannot exceed a certain word limit in order to keep the posts concise for the cardiologist (for instance, a maximum of 350 words).

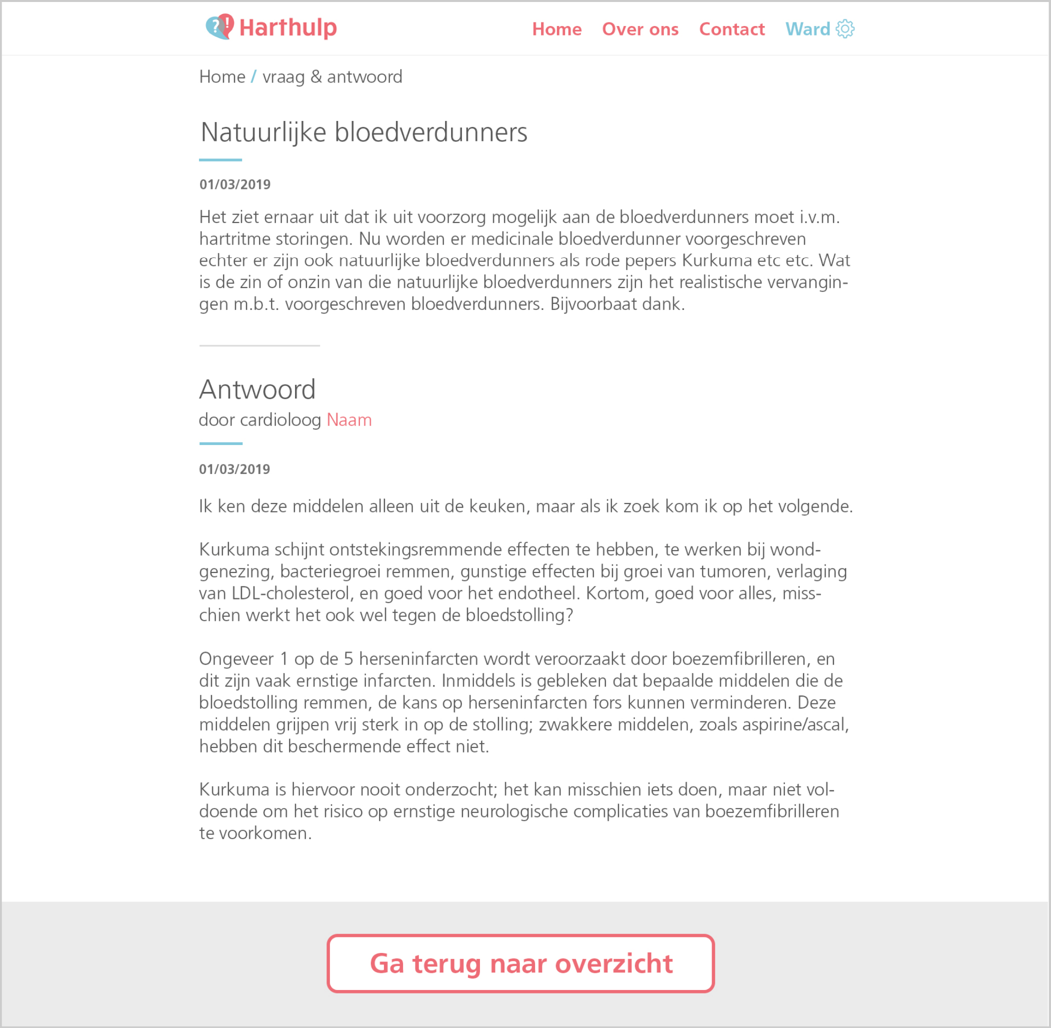


Uw vraag is verzonden naar ons team! ✓

Een cardioloog van het Harthulp team zal zo snel mogelijk een antwoord geven op uw vraag. U ontvangt automatisch een bericht wanneer uw vraag is beantwoord. Uw vraag is tot die tijd nog niet zichtbaar voor anderen op dit platform.

Wilt u uw bericht nog aanpassen? Klik dan hier om uw bericht aan te passen of te annuleren.

Question wizard step 5: confirm whether the question has been successfully sent to the Harthulp team. It also contains further instructions for the user.



Natuurlijke bloedverdunners

01/03/2019

Het ziet ernaar uit dat ik uit voorzorg mogelijk aan de bloedverdunners moet i.v.m. hartritme storingen. Nu worden er medicinale bloedverdunner voorgeschreven echter er zijn ook natuurlijke bloedverdunners als rode pepers Kurkuma etc etc. Wat is de zin of onzin van die natuurlijke bloedverdunners zijn het realistische vervangingen m.b.t. voorgeschreven bloedverdunners. Bijvoorbaat dank.

Antwoord

door cardioloog Naam

01/03/2019

Ik ken deze middelen alleen uit de keuken, maar als ik zoek kom ik op het volgende.

Kurkuma schijnt ontstekingsremmende effecten te hebben, te werken bij wondgenezing, bacteriegroei remmen, gunstige effecten bij groei van tumoren, verlaging van LDL-cholesterol, en goed voor het endotheel. Kortom, goed voor alles, misschien werkt het ook wel tegen de bloedstolling?

Ongeveer 1 op de 5 herseninfarcten wordt veroorzaakt door boezemfibrilleren, en dit zijn vaak ernstige infarcten. Inmiddels is gebleken dat bepaalde middelen die de bloedstolling remmen, de kans op herseninfarcten fors kunnen verminderen. Deze middelen grijpen vrij sterk in op de stolling; zwakkere middelen, zoals aspirine/ascal, hebben dit beschermende effect niet.

Kurkuma is hiervoor nooit onderzocht; het kan misschien iets doen, maar niet voldoende om het risico op ernstige neurologische complicaties van boezemfibrilleren te voorkomen.

An example page of the Q&A section of Harthulp.

Appendix G: Interview questions

During the semi-structured interview, cardiologist Stijn de Ridder was given screenshots of the design (see Appendix F). The cardiologist was already informed about the design project, thus the interview could immediately start with questions about the details of the design.

Questions about the design

1. The question wizard starts with background questions that visitors have to answer before they can search through the platform. In this way, search results can be re-ranked based on the patient's profile, that is, posts from patients with a similar profile are ranked higher in the results. Information from the background questions can also help the cardiologists with answering questions and it can be used to obtain medical statistics. Given this information, should visitors be obliged to answer these background questions, or is it a better strategy to let visitors decide for themselves what they prefer to report to a cardiologist?
2. If the answer is yes, what are the most important background questions?
3. How could medical data from health-based social networks, such as Hart Volgers, be useful for guiding scientific research in cardiovascular diseases?
4. Does the current form for posting questions enable cardiologists to quickly answer questions?
5. The current design is mainly designed for helping patients. During the ideation phase of the project, an idea came up to automatically categorize incoming questions based on the expertise of the cardiologists. This means that every cardiologist in the team receives the most suited questions. Do you think that this functionality could help cardiologists?
6. Are there any essential functionalities for cardiologists that are missing in the design?

Questions about the future of health-based social networks

7. A concern on health-based social networks is that patients can be confronted with posts from similar patients who are in dire straits. Should this aspect be addressed in a future design?
8. How should health platforms such as Harthulp look like in the near future?
9. Do you have any comments that you would like to add to this interview?