Universidade de Aveiro

DETI- Departamento de Electrónica,
Telecomunicações e Informática

Report Work 4 DiseaseCard

Alexandre Ribeiro- 108122
Guilherme Amorim- 107162
Matilde Teixeira- 108193
Miguel Pinto- 108481
Rafael Ferreira- 107340
Rodrigo Graça- 107634

Aspetos Profissionais e Sociais da Engenharia Informática
June 7, 2024

Contents

1 Introduction

Our project focuses on enhancing the DiseaseCard platform, a centralized repository for information on rare diseases. The previous version of DiseaseCard had significant usability issues, dispersed data, and relied on outdated technologies.

We aimed to entirely remake the platform to provide a more accessible, user-friendly, and modern experience.

Key improvements include:

- Aggregating diseases into families;
- Eliminating the use of I-Frames to display articles;
- Implementing new technologies, moving away from RDF, SPARQL, and COEUS.

Additionally, we aimed to develop a mobile app that acknowledges the increasing use of mobile devices for information access. The new platform uses web scrapers to obtain content directly from sources, resulting in a cleaner and more up-to-date interface.

The primary objectives of this project were:

- Concise and Accurate Information Presentation: Ensuring that information is presented clearly and accurately for all users.
- Rebuilding DiseaseCard: Developing an entirely new platform using modern technologies to replace the outdated system.
- Accessible Administration Interface: Creating an easy-to-use administration interface for efficient system monitoring and management.
- Mobile App Development: Building a mobile application to complement the web platform, providing a seamless experience across devices.

1.1 State of Art

The development of technology and data processing has led to the creation of many databases and tools dedicated to providing comprehensive information on rare genetic diseases.

These resources are essential for facilitating the search, discovery, and understanding of these diseases by offering access to a wide range of data, including genetic sequences, literature, and specific disease information.

Critical databases such as OMIM, PubMed, GenBank, and tools like Orphanet and Malacards support research and collaboration among healthcare

professionals and scientists.

However, these platforms need help with their usability and effectiveness. Common issues include poor user interfaces, outdated information, and a need for advanced search capabilities.

Our evaluation focused on two primary factors: functional and technical aspects.

Functional Aspects

- Disease Summary: Availability of a summary for each disease.
- Search: Effectiveness and efficiency of the search mechanism.
- Advanced Search: Ability to perform searches using advanced parameters and filters.
- Disease Categories: Availability of disease classification and categorization.
- Related Diseases: Information on diseases related to the researched one.
- News: Updates and news related to different diseases.
- Drugs/Treatments: Information on treatments and drugs available for diseases.
- Symptoms: Details on the symptoms associated with diseases.

Technical Aspects:

- Number of Sources: The diversity and number of data sources integrated into the database or tool.
- Mobile Interface: Availability and quality of mobile platform versions.
- User Interface: Overall usability and user experience of the platform.

Table 1: Comparison Matrix

Aspects	Orphanet	Malacards	Disease Ontology	GARD	NORD	NCBI	Our DiseaseCard
Functional							
Disease Summary	✓	✓	x	✓	✓	✓	✓
Search	✓	✓	x	✓	✓	✓	✓
Advanced Search	✓		✓	x	x	x	✓
Auto-complete			x	✓	x	✓	✓
Disease Categories	✓	✓	✓	✓	x	x	✓
Related Diseases		✓	✓		✓		✓
News	✓	✓	x	✓	✓		✓
Drugs/Treatments	✓	✓	x	✓	✓		✓
Symptoms	✓	✓	x	✓	✓		✓
Technical							
Number of Sources	Unknown	Unknown	>190	341	Unknown	Unknown	20
User Interface	Good	Terrible	Bad	Good	Good	Ok	Excellent
Mobile	✓	x	x	x	x	x	

2 Results

Throughout this project, we aimed to fulfil the goals we established at the beginning, and we can now proudly say we have achieved them all. From aggregating diseases into families to developing a mobile app for broader accessibility of our services to users, we have succeeded on all fronts.

We have created a clean and intuitive user interface that is visually appealing, making users feel comfortable while using it. With a cache implementation, we save users from constantly loading disease data, minimizing wait times between searches. Users only need to experience a loading screen when starting the website.

Within the disease section, users have three different ways to visualize the disease, allowing them to choose the best option that fits their needs and preferences.

Our web scrapers ensure that our disease information is always clean, accurate, and up-to-date. This information is correlated with a reporting system that allows us to improve our features continuously based on user feedback.

For admins, our service includes extensive data management capabilities. They can keep track of the number of families, diseases, sources, and articles, monitor active endpoints and accept or decline user reports. This makes the admin work much easier and more reliable.

Admins are divided into two types: technical admins, who handle endpoints and database information, and medical admins, specialists in the health area who can verify the scientific accuracy of user reports, ensuring they are not just from individuals with malicious intent.

In our mobile app, we maintained a clean and intuitive layout, with fewer options to view diseases, making it more straightforward for users to navigate.

Overall, we accomplished the comprehensive and user-friendly services we had planned, ensuring both user satisfaction and efficient administrative management.

3 Technologies

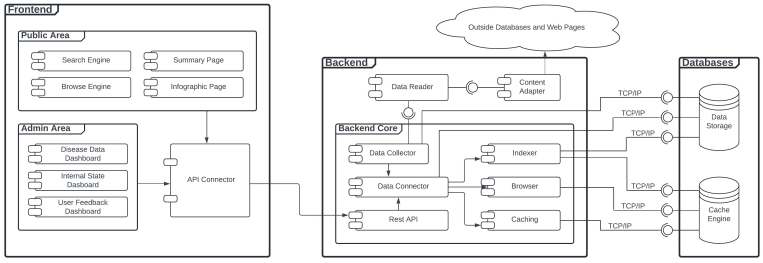
3.1 Architecture

Our architecture can be subdivided into three sectors: backend, frontend and databases. Regarding the backend, we opted to use SpringBoot, given the team's familiarity with the framework to build our main API. This API will connect to a Postgres database that will keep all articles, sources, and disease information.

We opted to use Postgres as our central database, a database known for its relationality due to the interconnected tables and data presented.

To get the different information from the different sources, we used scrappers to recover information from the various web pages and databases. For that, we utilized a plug-in architecture to simplify the addition of a new source and recollect as much information as possible from different origins.

We used a Mongo database to keep the article information persistent; given that Postgres could not support so many characters, the maximum is 255. This document is connected to others, sources, and diseases through the corresponding IDs and thus can store all information. For the front, we utilized NextJs, given the novelty of the framework.



The system architecture was designed with modularity in mind. Many system components, such as Public View, Backend Core, Scraping Mechanism, or Database System, might be run independently.

4 Commercialization

Two primary ideas for commercializing the platform have emerged to ensure that the Diseasecard can generate revenue.

4.0.1 Publicity

One way to generate revenue is through the integration of advertising. We can display ads specifically tailored to the interests and needs of individual users. Consequently, this will improve the relevance of the ads to the user, increase the possibility of engagement, and thereby boost advertising revenue.

4.0.2 Paid Priority for Articles

Another potential revenue stream is offering a paid priority service to the many article sources. Companies could pay a fee to have their articles placed in a priority position on the platform, which can appeal to entities that wish to increase the visibility of their research on diseases.

4.1 Product as a Service

Last but not least, our system design's modularity allows us to separate it into distinct actions. These actions may be used in a product-as-a-service system, allowing for extra revenue.

5 Patents Required

The project uses an MIT license, which means that the code will be open-source and, as such, there is no interest in trying to patent the important algorithms we have.

However, Diseasecard has a logo and a wordmark, so we must trademark these assets to legitimize our brand further. For that purpose, we searched TMview to check if our assets did not conflict with existing brands, and we discovered that both the marks and the name were available. To continue this process, we would start by applying for EU-wide protection by contacting EU-IPO and then take steps to trademark the brand internationally.

6 Open Software and Impacts

DiseaseCard is open-source software that uses an MIT License.

6.1 MIT License

The MIT License is a permissive free software license. Its simplicity and broad usage motivated users to use it. It allows users to use, copy, modify, merge, publish, distribute, sub-license, and/or sell copies of the software and to permit others to do so under the following conditions: the license and copyright notice must be included in all copies or substantial portions of the software.

6.1.1 Pros

- Flexibility: Its use imposes minimal restrictions on reuse.
- Business Friendly: Easy adoption, allowing use, modification and distribution with little concern for legal issues.
- Encourages usage: The simplicity and permissiveness broaden the use and adoption.
- Limited Contributions: The license does not require modifications to be open-sourced, limiting contributions from outside users.

6.1.2 Cons

- Limited Contributions: The license does not require modifications to be open-sourced, limiting contributions from outside users. Concerning end-users, this is a drawback.
- Potential for Proprietary Forks: Companies can take over open-source code, modify it and keep those modifications. We intend to help people access information, and due to distributing this service freely, other companies that use this can not have a significant advantage over us and take over the user base.

6.2 Impact on product

6.2.1 Community Engagement

- Benefit: The permissive nature can lead to widespread use and adoption, leading to a larger use base and more contributors. With more contributors adding new features, our system can scale and be accessible to new users whilst maintaining its traditional use base.
- Drawback: Due to the license not requiring derivative works to be open-sourced, it may result in fewer contributions back to the original project compared to copyleft licenses.

6.2.2 Legal Considerations

- Benefit: The MIT License simplifies legal compliance by imposing very few restrictions.
- Drawback: The simplicity and permissiveness might not offer as much legal protection for keeping improvements public compared to more restrictive licenses. Due to DiseaseCard being open to the public and free to all users, this is not a relevant drawback.

By leveraging the MIT License, Diseasecard can benefit from wide adoption, ease of integration, and the potential for commercial use while fostering an open and collaborative community.

7 Cybersecurity Issues

The central system is unlikely to be attacked because it does not contain sensitive data. However, if we considered the cookies implementation because of the advertisement, this could increase the interest of someone breaching the system.

First, it's imperative to ensure that the implementation of cookies for advertising purposes complies with GDPR regulations. This means providing users

with information about the use of cookies, obtaining explicit consent, and allowing users to manage their cookie preferences easily.

We must regularly review and update the cookies policy to reflect any changes in regulations. There are already implemented processes to prevent the risks of cyberattacks.

7.0.1 Encryption

We guarantee that the administrators' passwords are encrypted when stored in the database using Spring Security Password Encoder. This makes it more difficult for someone to enter the system using an admin credential.

7.0.2 Access Control

Many access control mechanisms, namely Role-Based Access Control, have been implemented to ensure that only authorized personnel can perform specific actions: Two roles—ADMIN and MEDICAL ADMIN—have permission to change data. Furthermore, the ADMIN can create or delete any user with the role MEDICAL ADMIN. Besides that, it also uses a multi-factor authentication using a TOTP code when an administrator logs into the system.

7.0.3 API Security

As mentioned in the last item, some specific endpoints are restricted to users with administrative functions and a specific role inside the administrator role. Besides that, the API only accepts requests from specific IP addresses, and thanks to Spring Boot Security, input sanitization protects against attacks such as SQL Injections.

8 Recover from Attacks

First, it would be crucial to understand how the attack happened and implement more restricted measures. The next step would be to analyze the data loss. As the Diseasecard contains many disease-related data, recovering from this data loss is crucial for maintaining the system's kilter. For that reason, there are regular data backups.

In addition to reporting to the appropriate authorities in case of a cookie data breach, it is essential to notice the users and indicate specifically which data was exposed.

9 Privacy Protection

GDPR compliance must be adhered to from the beginning to avoid legal issues related to such incidents. A data breach that becomes public can also affect the

project because people trust the system. That is why Diseasecard would use the minimum data necessary and implement many security protocols, as referenced in the previous sections.

10 Ethical Aspects

Ensuring ethical standards are followed is crucial to preserving the integrity of the DiseaseCard, so there are two aspects to be considered.

10.0.1 Sponsored Content

First, it is necessary to control the advertisements shown on the website to avoid misleading users.

Furthermore, regarding the paid priority system for articles from specific sources, it is essential to ensure that users know that these articles are being shown with priority over others because of the paid partnership and not due to importance or content.

10.0.2 Free Access to Information

DiseaseCard is committed to providing information on rare diseases free of charge. The team believes financial barriers should allow users to access medical information, so our service does not offer premium plans or build paywalls.

11 AI - Use of Artificial intelligence

The use of AI in the context of our project, despite not being initially implemented, could have significantly impacted the value of itself. There could be various goals we would aim to achieve as below we explain them:

11.1 Chatbot

One potential application of AI in our disease aggregation website could be developing a new feature, such as a chatbot. This chatbot could serve multiple purposes:

- Symptom Checker: The chatbot could interact directly with users by asking them about their symptoms and providing a list of possible diseases that match those symptoms. Although it would not be able to diagnose with certainty, lowering the number of diseases to search for and giving the user some information would be precious.
- Disease Information Provider: Users could ask the chatbot about a specific disease, and it would respond with a resume based on the articles in our database. At the end of the resume, the chatbot could also provide a link to redirect the user to the graph of the disease on our website.

11.2 User Preferences

In addition to the chatbot, AI could personalize user experiences. AI could identify patterns and preferences by analyzing user interactions on the site, such as clicks on diseases, articles, and sources. This would allow us to tailor the information presented to each user, ensuring they find the most relevant content quickly and efficiently.

11.3 Predictive Analysis

Another innovative idea for using AI in our project is predictive analytics. We could use machine learning algorithms to predict disease trends based on historical data and patterns. AI could also analyze data from various sources to predict which diseases are likely to increase incidence, levelling up the enhancement and search effectiveness on our website.

12 Relation to Hyperscalers

If our project took off and we managed to attract many customers, we would likely need to migrate our services to a more capable infrastructure. In this case, we would have to interact with the current hyperscalers, particularly AWS or Azure, for their cloud computing and data processing services.

Regarding publicity, we would use Google's AdSense to monetize the website and generate revenue through user traffic while also streamlining the implementation of the personalized ads that were previously mentioned, seeing as Google Adsense not only enables ads based on the user's Google Account but also allows ad customization, in case we want to restrict any ad categories. As such, our services would become dependent on these hyperscalers.

13 Network effects

Social networking has brought a new era to the everyday panorama. As information is highly and vastly available, more users are gaining knowledge in several areas. Concerning disease information, the panorama is on the same page.

However, like the rest of the information on the web, there is a large amount of wrongful information, and even with suitable sources, the information is spread and challenging to find. DiseaseCard addresses these issues by aggregating and curating medical articles, mainly on rare genetic diseases.

13.1 Direct and Indirect Network Effects

Network effects, as learned in this class, occur when the product or service value increases as more people use the said product or service. The more people that use the service, the more its value increases.

This can be seen in a website and online app, DiseaseCard's platforms. As said before, DiseaseCard provides a system that allows for a better reading and searching approach to articles and medical information regarding rare genetic diseases. Its different sources consist of many trusted sources in the scientific panorama.

13.1.1 Direct Networks

Concerning direct network effects, users can only add value to the website through the feedback option. This option allows the user to inform the platform if an article lacks information or is associated with the wrong disease.

Regarding DiseaseCard, the user can also present their feedback when a disease is in the wrong family or if the name is incorrect. By doing this, the user ensures a more informed platform and that the information is more scientifically accurate, ensuring a more accurate use for another user.

It is important to note that this feedback has to be authorized by the admin or medical admin to be changed in the global platform.

13.1.2 Indirect Networks

Regarding indirect network effects, they are also known as cross-site network effects. These are characterized by when the use of it influences the use of the product or service. DiseaseCard depends on multiple trusted sources. The more articles they publish, the more information about DiseaseCard remains updated and accurate.

The longer DiseaseCard stays this way, the more users will be attracted to the service, creating a positive feedback loop. A more comprehensive range of users will be reached by word of mouth, having promoters, or even suggesting partnerships with healthcare providers and organizations, universities, and marketing campaigns.

An interesting approach would be to share stories of families that gained much more information about a family member's condition and well-being and to understand the possible causes of it. Ensuring more users access the website and platform will create a cycle of continuous improvement and expansion.

13.2 Asymptotic Effects and Negative Effects

In the context of DiseaseCard, asymptomatic effects refer to the typical phenomena where the platform value continues to increase with additional users but at a decreasing rate.

Even though Diseasecard's information is constantly updated, these adverse effects can still occur. An option is to add new features, such as a chatbot interface and the option to have user preferences and ensure that DiseaseCard is the most comprehensible tool in the market, ensuring that our users will only access disease Cards in disease information searching.

It is important to note that Diseasecards experiences minimal adverse effects associated with network growth, such as congestion or reduced quality of service. The platform is designed to be scalable and maintain high standards of information accuracy as the user base expands. Regarding feedback, because it has to be authorized by the admin, our service's information remains accurate.

13.3 Graph Connections and Metrics

Examining and analyzing the graph connections is essential to studying the connections and metrics and guaranteeing a vast user base. The platform can be represented as a network graph where nodes represent users and articles, and edges represent the multiple interactions that they interact in.

It is essential to study degree centrality, betweenness and clustering coefficients to know where traffic comes from, see the cycle of information flow as well as identify critical users, and examine how users interact with the platform and how its use could be potentialized to ensure a broader use base, contributing to the platform's value.

Examining and analyzing the graph connections is essential to studying the connections and metrics and guaranteeing a vast user base. The platform can be represented as a network graph where nodes represent users and articles, and edges represent the multiple interactions that they interact in.

It is crucial to study degree centrality, betweenness, and clustering coefficients to understand where traffic comes from, see the cycle of information flow, identify critical users, examine how users interact with the platform, and determine how its use could be potentialized to ensure a broader user base, contributing to the platform's value.

13.3.1 Metcalfe's Law

According to Metcalfe's Law, a network's value is proportional to the square of its number of users.

This law represents DiseaseCard's structure, where the platform's value increases as users join and interact with the system.

For example, if the user base doubles, the potential interactions and feedback contributions increase fourfold, enhancing the platform's value.

13.4 Scale-Free Models

These models are another aspect worth considering. The platform benefits from direct and indirect network effects and cross-site interactions and maintains scalability with minimum asymptomatic effects.

DiseaseCard can be seen as a scale-free network where artists or users serve as hubs that attract more connections and interactions. This structure ensures robustness and reliability, as the node removal does not affect the overall network use.

In conclusion, network effects are essential to studying user interactions with the platform and guaranteeing the system's use. The platform benefits from direct and indirect network effects, especially interactions, maintaining use and scalability with minimal aspect effects. Understanding these connections and metrics is essential to optimizing their impact and building a broader user base. Metcalfe is a law that represents the interactions. Additionally, scale-free models ensure long-term sustainability, use, and growth.

14 Conclusion

In conclusion, DiseaseCard is a comprehensive platform that enhances global knowledge of rare genetic diseases by aggregating and visualizing information from reputable sources. Available for free as an app and website, DiseaseCard maintains profitability through advertising and prioritizing high-quality sources. To legitimize our brand, we will trademark our logo and wordmark.

We prioritize user security and privacy by employing robust cybersecurity measures and adhering to GDPR regulations. Our commitment to ethical content ensures that users receive accurate and responsible information. To grow our user base and offer personalized experiences, we plan to introduce AI-driven features, such as chatbots and user preference settings.

As DiseaseCard expands, we will migrate to a more capable infrastructure and utilize Google AdSense to support our operations. Our ultimate goal is to provide free, reliable information while fostering a secure and engaged community.

15 References

- APSEI Slides
- Google Adsense - Personalized ads
- How Google infers interest
- MIT License - DevX
- The importance of open source licensing models in software development - Diqq
- GDPR compliance with Spring Boot applications Part I: External databases - Medium
- The importance of open source licensing models in software development - Diqq (duplicate)
- Network effects hurt economies - MIT News