System report Knowledge Technology Practical

Group 29 Genetic Mutations

Gijs Hendriks (s2410540)
Marnix Jansma (s2784866)
Hidde van den Bos (s3219496)

December 2018

1 Problem

We are making a knowledge system that can give advice to the general public about genetic mutations. If you know that you have a higher risk for some genetic disease, you can go to the doctor which can refer you to a clinical geneticist. Then you can get tested for that disease. It is important to know early if you have high risk so you can take action. This can be to start treatment early, adjust your lifestyle or to rethink whether or how to have children.

An example of early screening in the Netherlands is population screening for breast cancer with women between 50 and 75, because they have high risk and still can be treated. However this is still a huge population and screening is expensive, takes time and can be quite painful. There are much more factors other than age and gender that influence your chances of getting breast cancer. Also there are a lot of women below 50 or men who get breast cancer who won't get caught by this screening. Our system is designed to give a basic assessment whether you need to be referred to a clinical geneticist. This is done by asking basic questions that can be answered by people without the need for tests like blood tests or genetic sequencing. The system is meant for the general population, not patients. Our system is low threshold and has potential to save lives of people who would otherwise not go to the doctor.

2 Experts

We have two experts that helped us in constructing the model:

Lies Hoefsloot
Hoofd Laboratoriumdiagnostiek
`https://www.erasmusmc.nl/klinische_genetica/verwijzer/specialisten/hoefsloot/`
Relation: Mother of Gijs

Carlijn Frantzen
Klinisch geneticus
`https://www.rug.nl/staff/c.frantzen/research`
Relation: Sister-in-law of Gijs

We took one interview with Lies about the scope and what factors play a role. We had a second interview with Carlijn about the specifics of when you need to be referred to flesh out our model.

We went into this subject rather unknowingly, but we let the experts guide us in the shape and size of the model. We first wanted to do only breast cancer, however Lies pointed out that that might be a small scope so we changed it to contain the three main categories of genetic mutations: Cancer, Heart and vascular disease and metal disorders. Carlijn helped us shape the model further by specifying which factors where the most important. Also she pointed out that there are a lot of cancer types, which all need to be diagnosed differently so that it was best we focus on the two most common types which coincidental have the most easy guidelines of referral. We took this to heart and only test for breast cancer and colon cancer. She used the official guidelines for this. The ones for cancer can be found `here`. Furthermore, we also decided to only ask questions which are easy to answer for anyone using our system, like the gender and age of the family members, and to skip the questions which most users would be unable to answer, like blood values, the size or the type of the tumor.

Both experts told us that our system has a lot of potential and maybe could help general practitioners in the decision whether to refer a patient, because they often have trouble with this problem. However this problem is complicated and every case is different. A lot of problems arrive when you need to give a binary advice. This is why clinical geneticists, experts for this very problem, exist. If this system were to be worked out with more time so a bigger and better model could be constructed this could be very interesting!

3 Role of knowledge technology for the problem

Medical diagnosis is a field where Knowledge Technology can and has been applied successfully. It is possible to formulate rules for diagnosis that can give a strong indication of whether a person may or may not have some ailment. By asking questions about certain aspects of a person's health and family relations it is possible to infer whether or not they fall within a group that has a higher risk for some genetic mutation/disease.

4 The knowledge models: problem solving model, domain model, rule model

We use an inference engine called Flora to do our reasoning for us. Flora allows us to create a knowledge base in the form of a source file that contains rules and facts (that may yet need to be inferred). The syntax of Flora is quite readable, and because it is separate from the rest of the program it is easy to modify parts of it and doesn't require hard-coding of logic statements in the program.

The knowledge-base can be updated during run-time of the program and this allows us to do inference as new information is obtained from the user.

In practice some of the inference is done explicitly using the order of questions to be asked by the program. This choice had to be made due to the limitations we encountered working with flora.

5 User interface, functionality

We made a GUI where we show a question. The user can cycle trough questions and your answers are saved. The user can go back and change a question. On the right side of the interface a log can be seen of the answers given so far. The user can get extra information if you do not understand a question (figure 2). The input is checked by the program to prevent illegal input by the user (figure 3).

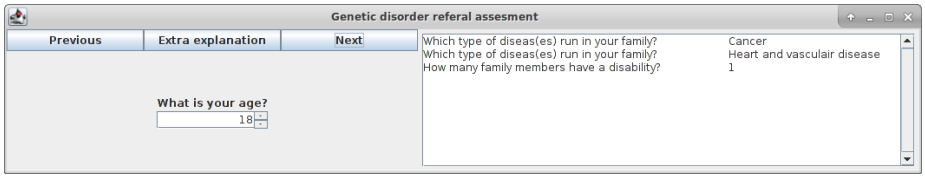

Figure 1: A screenshot of the current GUI.

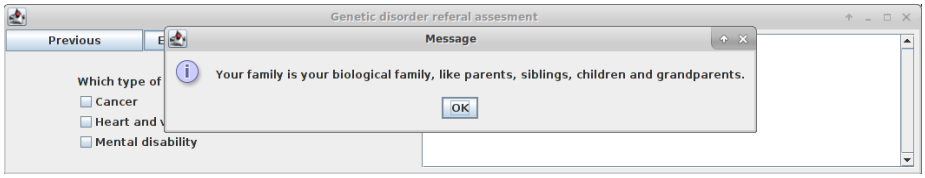

Figure 2: A screenshot of the current GUI.

The information given by the user is then passed to the flora engine to reason about that information. When enough information is given to reach a conclusion, the user is informed and can either exit the program or start over.

One of the problems with the program in its current state is that it is not very intuitive to install. The user is required to install flora (this is a prerequisite

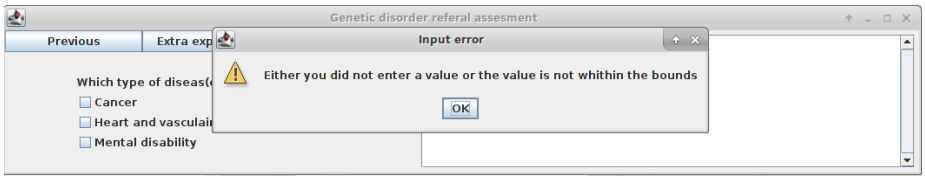

Figure 3: A screenshot of the current GUI.

that we can not get around) and some system configuration has to be done as well. The whole process is explained in a readme file that can be found on the front page of the program repository.

One unfortunate problem we encountered was that we found it difficult to package the program into an executable .jar file. This is due to inexperience in developing using java and because flora complicates things a lot. In its current state the user needs to download the source code and compile and run it using provided scripts.

6 Walkthrough of a session

The user starts the program and is shown a disclaimer. Then the user fills in questions. The user might go back and change one or more answers. The answers to them are send to the Flora engine that generates new knowledge and determines which questions need to be asked next. When all the required questions are answered the user is presented with the result: an advice on whether or not to go to the doctor. The user can choose to re-do the test or exit the program.

7 Validation of knowledge models

7.1 Structure of validation

For the validation of the project, we had an interview with both experts. We followed an structure we created beforehand with five use cases.

1. A 30 year old women has a sister that was diagnosed with breast cancer at 32 years old and is now worried what to do. (We refer her)

2. An 81 year old man is worried about cancer in his family. (We do not refer him)

3. An 18 year old man had a grandfather suddenly die when he was 47. (We do not refer him)

4. A women wants to have children and has a sister with down syndrome. (We refer her)

5. A combination of use case three and four. (We refer him)

We also inquired about the ease of use of the program, any bugs we encountered and general feed back the experts might have.

7.2 Evaluation with Lies

Lies was positive about our program. She had some comments about spelling mistakes and wording of questions and other information. We adjusted this accordingly. She had no comments on the model itself. She did comment on the fact that our program is not able to give advice like: "Your sister should go to the doctor". This is something a clinical geneticist is able to do. This is out of the scope of our project however. The sister is able to do the test herself of course.

She also had some comments about the interface, mainly that it should be bigger, so it is more clear. She also noted that there was no option to have no disease in your family, we added this afterwards.

7.3 Evaluation with Carlijn

Carlijn was impressed by our system. She said that the questions would be the same ones she would ask a potential patient. The order would also be the similar. She also noted that it was a good choice that we only determined for breast, colon and other type of cancer.

For use case 2 she noted that if a 81 year old man came to her with high risk (for example three sisters with breast cancer at a young age), she would still want to see him to see if he is a carrier of the gene. This could help for his offspring. This is the same comment Lies made about the fact that our program is not able to give advice for someone else than the user.

8 Task division among group members

We divided the tasks so we can work on the project on the same time, however we helped each other where we could of course. There was an overlap for each task. The division we settled on is: Gijs is doing the interface and the contact with the experts, Marnix the interaction between Java and Flora and Hidde the Flora model. All interviews/evaluations where conducted by Gijs and Hidde. One interview was done by Gijs alone.

9 Reflection

Especially in the beginning progress has been somewhat sluggish. The decision to use Flora has been interesting certainly but it was slow to pay off. The difficulty lied in the the lack of prior knowledge about it or similar inference engines/languages like for example Prolog. The software is also relatively new

and has seemingly few (known) usages and applications. Because of this it was hard to find solutions to problems that are likely common, since all information has to come from the documentation. This documentation is vast and at points difficult to get through, however it is thorough and seemingly complete.

Because of this large hurdle, it took a lot of work and testing to implement the knowledge of our experts. The experience has been rewarding however, as from an technical standpoint we have learned a lot about working with this software in particular and it has been a real eye-opener in what it can be like to work with cutting edge software. Both positively and negatively.

We also found it interesting to dive a little bit in a field of clinical genetics.