ARRT

Aiding Research on Rheumatism using Technology

Uppsala University

&

Rose-Hulman Institute of Technology

Blankenbaker, Zachariah
Hajimohammadi, Hamid
Johansson, Håkan
Jonasson, Tobias
Knutsson, Tobias
Malmros, Robin
Marcum, Keith
Matsui, Michael
Moritz, Niklas
Nordström, Mikael
Overall, Christopher
Parks, Justin
Reust, Michael
Stuart, Jeffrey
Wallin, Christian
Ward, Scott
Åkerstrand, Patrik

Abstract

A new requirement from the EU is that health research should be centred on patients' needs and that patients should have an active role in pushing the agenda of research. At the same time the internet and social networking has emerged as a medium where patients discuss and elaborate different aspects of their disease. This creates new possibilities for researchers to understand patients' needs. Hence, this report gives ideas of how to aid research on rheumatism using web-based mediums.

Two different technical solutions are evaluated:

- Mining existing services for information
  The mediums that are used today (Facebook, blogs, etc.) contains lots of information and discussions between people with different diagnosis and this information is interesting for researchers. Developing a tool that gathers this information and provides it to the researchers is one of the solutions that are reviewed in this report.

- Creation of a new community
  The other solution that is reviewed is creating a new community that has a tool for processing the information built in to the system. This gives the creators total control of the system and the information in the system (discussions, forum posts, etc.).

The different solutions are described and elaborated from a technical aspect, along with discussions on anonymity of users, legal and ethical issues. Finally recommendations are given based on conclusions of the different solutions.

1 Introduction

1.1 Presentation

Uppsala University is the oldest university in Scandinavia and with forty thousand students it is also one of the largest universities in Sweden. Every year the course "IT in society" is taught at the department of Information Technology. The aim of the course is to "[...] provide an understanding of the interactions between technology, users and designers". This is done by letting the students partake in an actual project with a real client, and produce their result in the form of a report. Uppsala University is also collaborating with Rose-Hulman Institute of Technology on this project. Rose-Hulman is located in Terre-Haute, Indiana, in the United States. The client of this project is Ulla Lindqvist and the Uppsala Academic Hospital. Ulla, who is partly employed at the rheumatics ward, is a prefect of the Department of Medical Sciences and part of the rheumatic's research group at that department. Uppsala Academic Hospital is a research and teaching hospital. It is one of the largest hospitals in Sweden and the oldest university hospital in the world. It maintains close cooperation with the medical faculty at Uppsala University.

1.2 Statement of Purpose/Client Requirements

A new requirement from the EU is that research should be centred on patients needs and that patients should have an active role in pushing the agenda of research. The Rheumatics Society is currently educating members and encouraging them to participate and give their view of what they feel is important in future research. But Ulla Lindqvist has a concern regarding the newest generation:

"As a young person who just got a diagnosis the rheumatics society might not be that appealing because it is personified by the older rheumatic patients and the webpage addresses its users mostly as older. The younger might have different questions like for example; can I go downhill skiing with rheumatism? Is there any special equipment I could use? Can I go scuba diving? These are not questions that a 62 year old lady with rheumatism would pose. What Reumatikerförbundet (the Swedish rheumatic society) stands for is of the old times meaning that the young men who gets the disease in their late teens (18-20 years old) is not interested in the rheumatics society but perhaps in sports and will seek to join such groups but will still have questions about their affliction and how it relates to their activities."

This was the major problem stated by the client. The younger generation needs to be considered. By looking into where they create their social networks and what they discuss there, their needs can be found. The Rheumatics Society appealed mostly to the older members and was not enough to create a complete view of the patient group. The fear was that the newest generations' voices would be lost and research would not be in sync with a new society and new needs that constantly arise in an ever changing environment. Besides this basic demand, no direct requirements were stated as to functionality, compatibility with existing practises or technology. The team has accordingly worked with a high degree of freedom and looked into several possibilities and have done a lot of "outside the box thinking".

1.3 Our Vision

The overall vision of this report is to improve the quality of communications between patients, doctors and researchers. A proper solution would improve the quality of information that can be gathered by researchers, provide better quality feedback on the needs of patients, and aid doctors in understanding patients' needs. Researchers should be able to obtain more detailed information concerning the areas of rheumatism that patients are most concerned about. They could obtain graph-wise representations of common discussion topics, for example a pie graph showing a distribution of common activities that patients are having problems performing. Also, researchers would benefit greatly from a tool that could specifically track the amount of discussion that gets generated when a new medication is introduced to patients. With this information in the hands of researchers, doctors would be better able to meet their patients' needs. They can directly address problems that are commonly being discussed among patients, answering common questions that patients may not think of asking when meeting with their doctor. In addition, doctors should be able to participate in the patient community discussions and provide immediate feedback. Finally, patients would benefit the most from a proper solution. A communal discussion would provide patients with an excellent source of information about rheumatism and an excellent place of support for their individual problems. They could get specific feedback about their disease from knowledgeable doctors or from patients who have gone through similar problems.

1.4 Summary of this Document

This document is divided into three sections. Each section builds upon the information found in the previous section and works toward a solution that fulfils the client's needs. First, we analyse modern forms of communication, ranging from telephone conversations to various web-based tools. We consider how much they are used and the strengths and weaknesses associated with gathering information from them. Most current communities and support groups maintain online support for rheumatism patients. Also, purely based on the form of data, online resources prove to be the best medium for monitoring and communicating information. Thus, we next describe two potential solutions to the client's requirements: building an online community with built-in data gathering utilities or constructing a tool that will analyse the discussions of existing online communities. Each has its own strengths and weaknesses, but both have potential legal and technical issues. Weighing the qualities of each solution, our conclusion is that building an analysis tool that interfaces with existing communities would be the best solution. We give a detailed design description of such a tool in the final section of the report.

2 Background

2.1 Available Tools for Communication

Before a new tool to gather and analyse information can be planned and developed you need to do some research. This includes looking into what tools are currently available for people to communicate without seeing each other face to face. The communication which we are interested in is both between patient and doctor and between patient and patient.

The flow of information that we want to analyse is generated by numerous tools that people use in order to express their opinion, share their feelings or ask questions. To successfully gather information from this flow we need to look into the tools that are being used. We also need to find out which ones are most suitable to extract information from. We also need to take into consideration that our target group is people with rheumatism, which narrows the field.

2.1.1 Technology Platforms

Patients and researchers are always in need of ways to communicate. This report summarises the different methods that are used by patients and researchers to perform this information sharing and gathering process. They need to be examined in order to find out which ones are most suitable to gather valuable information from.

Telephone

The telephone is a great tool if you know who you want to talk to. However, if you just want to express your feelings and/or opinions to the world and get feedback, it's not so great. This is partly because the telephone is not anonymous. If you call someone they will know who you are, or find out who you are before answering your questions. There is also a problem with gathering information that is coming from a telephone call. Your options are either to write down the conversation to text, as you speak, or to record the phone call. However, storing lots of recorded telephone calls if you want to analyse what was said, is not very resource efficient. It is easier to make a computer analyse text than to make it analyse recorded conversations. Maybe in the future the telephone will be a better tool for gathering information, when we have computers with a strong enough capacity for artificial intelligence that they can understand what a person is saying.

E-Mail

One reason that makes e-mail suitable for a system where one wants to gather information is that the information is already in a digital form. It is also easy to identify the sender and receiver, which means that you can analyse how information flows between patients and medical staff. It is thereby possible to see whether the conversation is one-way or not. People also tend to speak in a more formal manner via e-mail, which should make language analysis of the text easier; on the other hand this might also mean that patients will hold back information that might seem out of place in formal communication[1].

Since a lot of the conversations that will take place between patients and medical staff will have to be confidential, it is important to consider the use of e-mail from a security point-of-view. When the

technology behind e-mail was developed back in 1965, it was intended for use in a friendly environment and therefore it was never designed with security in mind: "Mail is inherently insecure".[2] This means that there is no built-in confidentiality or integrity features into the e-mail protocol. The result is that all e-mail is sent in clear text, readable for anyone with basic computer knowledge who can place themselves somewhere on the route that the e-mail is being sent.

This does not mean that the use of e-mail has to be abandoned; it just means that you have to use additional protection when sending messages. The most common way of achieving secure e-mail[3] is to use so called S/MIME-Certificates (Secure/Multipurpose Internet Mail Extensions) which guarantees the origin, integrity and confidentiality of an e-mail from the sender to the receiver.

Mail

This is probably the best media to use if you want to send out surveys so that all the patients will get it. Everyone knows how to use this old tool so there will not be any problems with a user group that doesn't have enough knowledge of the media. The biggest problem with this is that it is hard to analyse since it is not in digital form. To perform advanced studies on the information within the mail the information would first have to be converted to digital form which would result in extra workload compared to other digital mediums.

Internet Forums

An Internet forum is a medium for holding discussions. Users post topics anonymously or with a nickname and other users post responses. Forums tend to be very public, although some instances restrict access to certain users.

Internet forums can contain massive amounts of information. This can be very useful if the information is well organised and well written, but there is no guarantee that this will be the case. Because there is often little guidance, forums can become messy and disorganised. Also, many forum topics and replies tend to be informal and thus automated information extraction could be difficult. While anonymity encourages some users to be honest, other users may abuse it to post off-topic discussions, advertisements, or other distractions.

Instant Messaging

Instant messaging is a kind of communication between two or more people in real-time. Most of instant messaging tools are based on typed text, but in a couple of instant messaging services you can talk to each other just like how a telephone is used.

One of the most used text based instant messaging services used is Windows Live Messenger with around 27.2 million active users all over the world.[4] Another big instant messenger, and one of the services that gave IP Telephony a new dimension, is Skype with over 10 million users online during peak periods (2007).[5]

2.1.2 Online Communities and Resources

Over 76% of the population in Sweden uses the Internet today in 2007.[6] This means that a lot of people turn to the Internet to find information about their ailments or use it to communicate with other people as well as with doctors and researchers. This makes online communities, and other web resources, interesting to us.

Web Resources in Sweden

Netdoktor

Netdoktor is one commonly used web community were people go to find out more about their condition. In other words, Netdoktor is a good tool that a lot people with rheumatism use to express their thoughts and gather information. Therefore, it's an interesting tool to analyse, as we want to gathering information.

They have a number of sections for different kinds of diseases, including one section for rheumatism. Here people can get information and share their thoughts with other people in the same situation. They can also ask questions directly to a board of medical staff that will try to answer them. Each member also has the opportunity to write their own diary on the webpage. On their page they also have a forum that gives people an opportunity to ask questions to other people. The exchanged information here might in all probability be a good source to extract information about what people with rheumatism have on their mind. The forum has a small number of sub categories. As mentioned above Netdoktor also has a diary function. It's mainly for the users and each user has the ability to create their own. It is also possible to search for and read other member's diaries. These diaries contain a lot of information and one way to gather users' opinions could be to analyse the data within the users' diaries. This may or may not be possible due to legal issues.

The Swedish Rheumatic Society Web Site

The organisation with the most influence on patients with rheumatism in Sweden is without a doubt the Swedish Rheumatic Society, Reumatikerförbundet. In 2007, they awarded 62 different research projects over eight million SEK through their funds from the public inheritance foundation (Allmänna arvsfonden), donations and inherited funds[7], making them one of the biggest non-governmental contributors to Swedish rheumatic research. They also arrange local meetings where people can come together and talk about their problems; during the summer there is also a camp for children with rheumatism.

The Swedish Rheumatic Society strives to be in the public's eye as much as possible to raise the concerns of their members and also to attract new members. To accomplish this they try to appear in radio and television shows such as debate and health programs. To communicate with their members about current topics, such as upcoming events and research progress, they publish their own magazine, Reumatikertidningen, six times a year.

Another channel to communicate with their members, and for their members to communicate between each other, is their forums, placed on their homepage. The average number of unique visitors in one month is currently around 10 000 members with a daily average of around 800 non-unique members

each day.[8] Inside this forum the members can discuss anything they want even if what they say cannot be proven to be correct information such that some food will make you feel better. In addition to this there is also a possibility for their members to call a nurse and ask questions or send a mail to the society. On their homepage there is also a very big information database about a large amount of different versions of rheumatic diseases.

After describing our project and asking about their interest in this they are very enthusiastic. They think that a way for the scientists to get feedback and information from the patients in an easy way would be a great thing. Their only current concern for this is the legal and ethical issues. They do not want to risk losing any members because of a computer program running analyses and providing research material from the information the users put down on the webpage.

Therefore, it was suggested that this form of data gathering should not be continuous but instead perhaps limited to a month at a time. This solution would mean that the people who do not want to participate would only be out of touch with their society for short periods. Another way to go would be to give each member a question whether they want to participate in the research project or not. This could be done on an opt-in basis to make sure that everybody who is being studied actually is aware of the projects ongoing.

Both of these solutions will have an impact on the resulting information that the users will create on the forum. This is derived from the fact that it is not possible to study human behaviour, with the subject's knowledge, without affecting the results. So from a scientific point of view, the way to achieve the optimal research results in this case would be to carry out the study without the user's knowledge. However this would not be ethically correct, in some cases it might even be against the law of integrity. Thus this should be avoided.

What then remains is a choice on how to present and perform the study on the users. The option of a limited study period could lead to people acting and writing differently during that time and then go back to their normal habits after the research is over. With the other approach this is less likely as people will probably not change their behaviour forever. Instead, many users may choose not to participate and the people who will choose to do so will likely share some properties such as sex, age, etc.

Web Resources in U.S.

MDJunction - http://www.mdjunction.com

MDJunction.com is a site based on people helping people. The core of the site consists of community forums, split up by disease, where users can interact with each other. They are encouraged to share anything and everything with each other. This often includes information about treatments, reviews of doctors they have seen, or just talking about what is going on in their life. The idea is that all of the discussion can bring about new thoughts or ideas that can greatly help a person. A nice thing about the site is that it focuses on having a very welcoming aesthetic. One example of this is that users are able to send each other hugs for support. Within the forums themselves, users are able to remain anonymous if they like. They have the option of including their name and picture or making up an alias. The site also contains links to relatively current articles about medical issues people using the forums might find interesting.

WebMD - http://www.webmd.com

WebMD is primarily a site focused on giving users the ability to find out about their diseases or play around with diagnosing themselves. In short, it brings together a set of information that users are able to parse through. One important aspect for us is that there is no apparent interaction between users of the site. It is simply a site to go and get information from.

Patientslikeme - http://www.patientslikeme.com

Patientslikeme.com is a fairly new web 2.0 site/community. It is broken into four areas: patients, treatments, symptoms, and community. Each area is interconnected and works together in a way to help the patient better understand their ailment. You can start using the site by creating your own profile; including your affliction (current options are Parkinson's, MS, and ALS). Once you have an account, you can do the following:

- Browse all the symptoms other people with the same disease are facing, and discuss them.
- Browse common treatments and discuss them
- Browse the other users' profiles and contact them
- Browse or post in the forums (the community section)

Curezone - http://www.curezone.com

Curezone is a general forum with a wide variety of topics. The Rheumatism topic includes general information, facilities to ask questions of "experts", a Frequently Asked Questions section and forums for individuals and "experts" to discuss issues. The content focuses on "cleansing" and other alternative medicine methods. The site offers simple search tools that may be useful for finding topics of interest but the link to the Rheumatism forum is apparently broken, currently (December 2007) directing to an acne forum. The site also lacks any sort of organisation for the general information it provides, relying on in-text links to navigate from topic to topic.

DailyStrength - http://dailystrength.org

DailyStrength is a strong community building site that is divided up by affliction. DailyStrength has an average of 14,000 visitors daily, each spending 82 minutes on the site and each viewing approximately 145 pages.[9] Users have profiles, friends, photos, videos, journals, activity logs and can send and track virtual hugs. Ages seem to range down to at least the lower 20's and maybe into the teens. There is a catalogue of treatments (from crying to drugs to music) that are rated and reviewed by members on their effectiveness. There is a limited amount of raw information on afflictions, an advice forum, a recommendations forum and standard discussions forum.

Social Networking Websites

A social networking website is a site where people can communicate with each other. Normally a social network website, such as MySpace and Facebook, allows the user to create a user profile. The users can write about themselves, add interests, add friends (other users), upload pictures and often much more than that. These communities often provide you with functionality as talking to your friends, chat and discuss all different things in forums.

The current (2007) most popular social network website in Sweden is Facebook; it is also the third most popular website of all type in Sweden, just behind Google and YouTube[10], and has over 57 million active users all over the world.[11]

On Facebook you can join groups about almost everything that you can have a group for. As example there is a group for young rheumatics (Unga Reumatiker). Here they can for example discuss things in their own forum.

Peter Dahlgren, who is working with usability, information architecture and marketing, writes on his blog, Backend Media, about why Facebook is so popular. In this entry he outlines five major pointers:

- Favourable for journalists
  One of Facebook's target groups are the journalists. For a lot of journalists Facebook has been a good starting point for their journalistic work.

- Focus on existing relations
  The idea with Facebook is not to make people meet new friends; rather it brings existing relations into the virtual world.

- Names, not usernames
  On Facebook you do not need to know or remember your friend's username, number etc. you just use your real name. That makes it more intuitive, and it makes it easier to find your friends.

- Marketing through users
  The large expansion of the network is made by the users themselves. The users want to collect all their friends on Facebook, and to do so they promote it to their friends and get them to use it too.

- Open Application Programming Interface (API)
  A smart thing that Facebook has done to become so popular is to let third part developers use an open application programming interface so companies can make their own applications and introduce them on Facebook. This has improved the variety of the content available to users and strengthened Facebook's standing as an online community.

Second Life

Second Life is a virtual world (e.g., The Sims) that is open and free. Users create an avatar, which is essentially a virtual character. They can then explore and venture into the virtual world like in a video game. A particular interesting concept is that users interact in a way that mimics traditional social interaction. They just walk up to each other and start chatting. As far as medicine, there are actually several medical locations inside the virtual world of Second Life. There is a large variance in the target market for each location. However, most are for people suffering from medical conditions. There is even a locale called “Wheelies @ Second Ability” where you can learn about disabilities and try out a wheelchair. Another locale called “Medical Library at Health Info Island” allows users to search for health info and chat with the information desk.

2.2 Research Value of Online Communication as Data

The Internet allows for people of all educational backgrounds and personal opinions to speak their minds, something that most times is valuable but can cause problems when specific information is desired. A public site on which anyone can contribute information will contain false information presented by people who have incorrect sources or who intentionally want to contribute incorrect information. However, as more individuals interact in the discussion, the information more closely conforms to the facts. More knowledgeable people are able to correct false information or blatantly point out something that is wrong. This is beneficial as it allows for misinformation to be corrected. Such behaviour can be seen in forums and sites such as Wikipedia. Wikipedia is an encyclopaedia which is built entirely from user contributed information. In the years since its inception, it has been subject to much criticism and praise, but recent studies show that it is as accurate as well established academic encyclopaedias such as Britannica.[12] The strength of online discussion sources comes from the cumulative knowledge of all the users.

Another problem with discussion tools is a problem with relevance. Any online discussion tool will contain off topic information. For example, the DailyStrength Rheumatism community not only has discussions about their affliction, but also about what are good gift ideas for loved ones. Since human behaviour is a factor, it is hard to determine how relevant different data are.

Like any conversation, one held in a virtual environment has its own set of limitations and characteristics and simply being aware of these gives you a possibility of analysing and retrieving information. As you would have to be aware of the circumstances in an interview you need to be aware of the different aspects of having an online conversation compared to a real life one. One aspect, which is fundamental, is the possibility of anonymity in an online conversation whereas anonymity in a real life conversation is usually not possible. A virtue of being anonymous is the ability to speak more freely since what you say will not be tracked back to you. This is one aspect of online communication which researchers need to be aware of when considering data generated by our tool.

2.2.1 Anonymity

By having the possibility of being anonymous you have a greater degree of freedom to reveal personal information since it cannot be connected to you as a physical person, but you also have a possibility of lying or distorting the truth. According to research done by Jeffrey T. Hancock, Jennifer Thom-Santelli, and Thompson Ritchie at Cornell University [13], lying is most common in face to face conversations or over the phone while instant messaging or e-mail contain significantly fewer lies or misleading statements.[14] "[...] the degree to which a medium 1) allows for synchronous interaction, 2) is recorded less, and 3) is distributed, the greater the frequency of lying that should occur in that medium" this means in regard to forums and similar tools that they should contain fewer lies than would occur in a face to face interaction. HuaQian at University of Texas and CraigR.Scott at Rutgers University state in their article Anonymity and Self-Disclosure on Weblogs[15]: "online communication lends itself to self-disclosure" so we can accordingly assume that users will be truthful and reveal personal information.

2.2.2 Other Factors to Consider

Other factors that come into play when communicating online are several; age, context, culture all play their role in affecting us and our communication. Any researcher should be aware that the average age of online communicators is lower than that of most Rheumatoid Arthritis (RA) patients and that older patients group will not be as big as the younger group.[16]

2.3 User Perspective: What is a Good Tool?

2.3.1 Patient Needs

The majority of all Rheumatism patients are familiar with the Internet and uses it in some way in their daily life, either at work or at home. But some in the older generation do not use the Internet at all, or may use it very restrictively. Almost none of the patients we interviewed use the Internet for searching for information about their disease, most of them think that they know as much as they need but some say that sometimes they want to talk about their disease with like-minded.

About their ability to use computers and belonging accessories, most of the patients can make it without any tool for their help. But some patients, depending on how their bodies are affected, may have some problems to use a computer on a daily basis.

Most of the patients currently do not use any web-based service related to their disease. But they think they will use a web service if it presents the latest research result about their disease, or if there are researchers or doctors behind it, that really use the service in their work. Patients that have recently been diagnosed are more positive about using a web service to learn more about their disease and talk to like-minded to see their problems, and discuss them. They say that a forum where you can discuss is something that they may use for that purpose.

The patients often do not have so deep a knowledge about their disease so that they think that they can discuss with researchers about what they should do to get a better life. But if they just can discuss with researchers and other patients with the same disease about how they feel, it would be very good if the researchers could take that in mind when they do research. A patient said that "I'm just not capable of discussing with a researcher. I don't have the expertise needed.".

2.3.2 Researcher Needs

According to the researchers we interviewed it is much more important that the patients can speak freely. Ulla Lindqvist said that "Free thought is much more important because if you have multiple choices you have already decided the types of questions and what answers that can be given". A future tool can be used by researchers in many different ways and one way could be that the researcher can ask a question like "What are patients talking mostly about on Facebook" and using statistics, the researcher can get the answer to that question. Some researchers' wants diagrams and statistics about everything and some want free text from the patients. If you want free text it can be good to get some ground for further analyse, like interview or questionnaire information. Maybe a good solution is to have some kind of choice about what you will get from the tool.

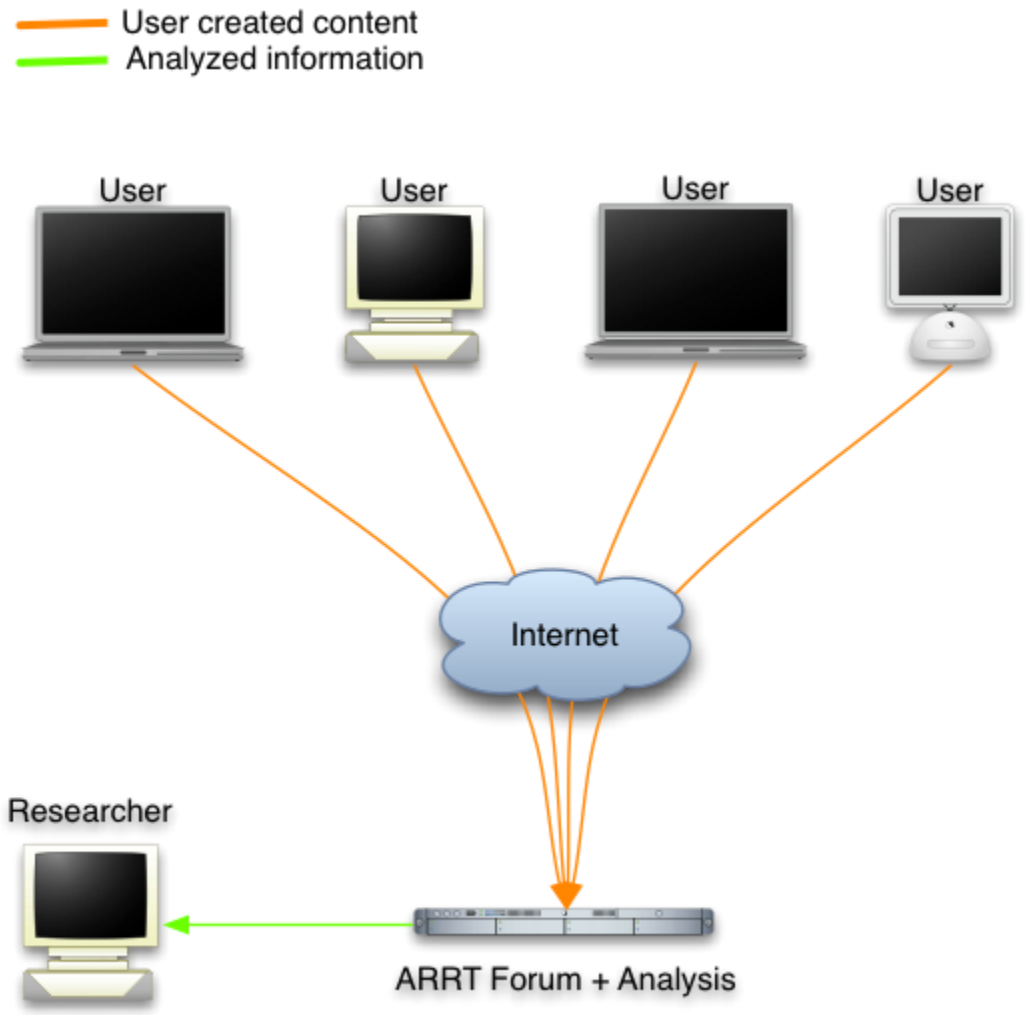
2.4 Summary of the Current Situation

There are lots of different communication tools available for communication for different user groups, but they might be separated or spread over different formats and areas. This might result in lack of communication between different user groups. Most patients use some of these communication tools in their daily life, but most often not for their disease. To overcome the shortcomings of the investigated communication tool, we can create a new tool that fulfils all of our expectations, or we can create a solution to gather information from all, or some of, the different existing communication tools.

3 Potential solutions

3.1 Building a New Web Community

Our first solution is to develop a new web community from the ground up. The web community would be a place for patients to interact with other patients through forums, blogs, and to learn more about their symptoms. This web community would be integrated into it a system of analysing tools that would inspect what the users of the system write and attempt to discover patterns that would give doctors and researchers a better idea of what their patients are worrying about. These analysis tools would look through many different pieces of data, including the text that is written on the forums, what people are saying in their blogs, and what pages in the education section of the website are being accessed the most. The first two of these would give a good idea of what patients talk about the most, and the third would help to give an idea about what patients most commonly look to learn about.



3.1.1 Legal Issues

The legal issues of this system are that everything the users say would be visible to everyone else on the site, but also to the analysis tools. This could be a potential violation of rights unless the users are notified in advance that studies are going to be conducted on everything they submit to the web community.

3.1.2 Deployment and Marketing Costs

The deployment of new software will consist largely of pushing the developed tools onto a web hosting service. There are several different hosting options depending on what technologies are used to implement the new tools (e.g. ASP.NET, PHP, Java, MySQL, etc...). As such, there are a variety of providers for these services. One popular hosting service, Go Daddy, currently charges $15.00/month for their premium plan.[17]

After the new software is hosted and available for access, the next step is directing traffic to it. Unfortunately, online marketing can be a bit of a black art. One of the primary venues with which to drive traffic to your site is via search engine results. To make things complex, each of the main search engines used throughout Europe, and the rest of the world, have unique ways of sorting search results. For example, Google uses a page ranking system that relies heavily on the number of different web sites that link to yours. So, if Site A is linked to in 50 different places throughout the Internet, and Site B is linked to in 100 different places throughout the Internet, then Site B will show up earlier in a Google search for any keyword(s) both sites share.[18] Of course, Google uses a much more complex algorithm to sort search results than simply the number of links to a site, so optimising your results in a Google search can end up being a lot of guess work and luck.

In order to receive more favourable placement in search results one might consider placing links to their own site in as many random places on the Internet as possible. However, there are some more appropriate and practical methods. In fact, Google has a detailed tutorial for optimising your web application for Google.[19]

In addition to standard search results, one could also purchase advertisement space that would appear in ad space that surrounds search result listings. Google's version of this is called Google AdWords. There are a few major advantages and disadvantages to using a paid service like Google AdWords.

Major advantages:

- Increased visibility
- Only charged when users click the advertisement
- More reliable than relying on increasing your site’s page ranking
- Allows you to set a budget, so you know the max you'll be charged each month

Major disadvantages:

- When users click, it does cost money
- Relies on users to notice the ads
- Users must be using the keywords you purchase in their search queries

- Click fraud is a risk

An effective marketing method for the tool could be word of mouth. This can be effectively done by doctors letting patients know of the tool out there where it not only provides support for the patient, but also provides valuable feedback to researchers. The effectiveness of the method lies heavily in the value of human interaction over computer interaction; users will be much more likely to trust a website recommended to them by a medical professional versus a site they find from Google. If this is emphasised enough by doctors to their patients (aided perhaps by a handout as a reminder), it will go a long way to bringing users to the site.

3.1.3 Technology Issues

The biggest issue of this solution is that the development of a whole new website with forums, blogs, information, and analysis tools such as this one would have, is quite expensive. The forums and blog could probably be developed from freely available open source projects, but the analysis tool would have to be developed from the ground up, and putting it all onto a system that could handle it all together would be very costly. It is hard to determine how costly since the preferred level of developing advanced features might differ between creators of the community. Even developing this platform on a basis level might get expensive comparing to using the already established sources that exist.

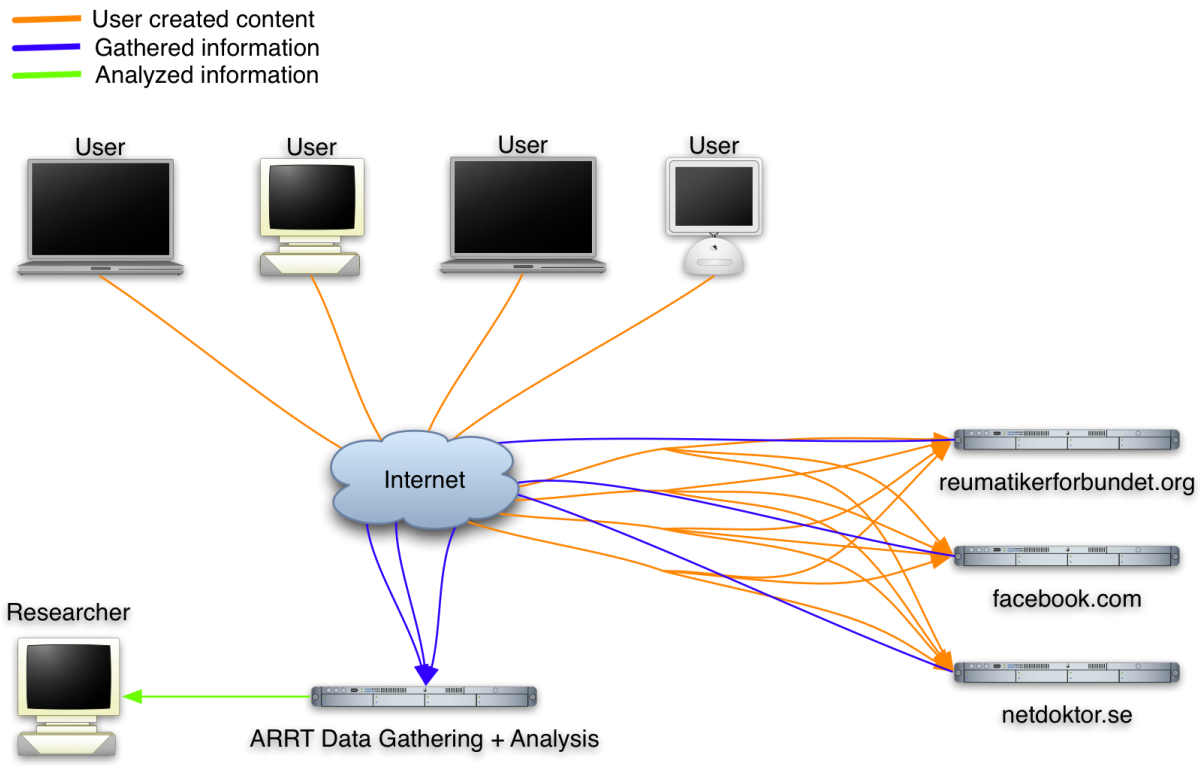
3.2 Analysis Using Existing Services

This alternative would gather data from existing social networking services available on the Internet, e.g. Forums, Mail, Blogs, and Diaries, and analyse the existing content. For example, one could gather all the messages being written on the Swedish Rheumatic Society's (Reumatikerförbundet) forum and have a computer analyse it and present it in a condensed format. The same approach can then be reused by having a system where different input modules can be attached, thus enabling analysis of many different sources by converting the gathered data into a common format. This also means that the analysis tool can be extended to analyse new services as they appear, without having to reinvent the wheel. The result of the analysis can then be displayed to a researcher in the form of diagrams, common word lists etc.

The most obvious advantage with this approach is that all the users of all online services can be used as an information source. This saves a lot of time and work since there is no need to develop any new services for the users. They can use the services they are used to and still participate in the research. This is a big advantage since people in general are reluctant to embrace new systems and changing their habits. Meanwhile the service providers will handle the administration of users, content, security and all other responsibilities that comes with running an online social network. It also means that the service providers will pay for all the computer power needed to run the service in form of processing power, storage capabilities, bandwidth and so forth. Since there is no need to attract users to a new service, a lot of resources can be saved when it comes to marketing, thanks to the fact that the system will be completely transparent and not have to depend on active participation from its users.

Another very big advantage of using the existing services is that there is already a huge set of information available to analyse right from the start, which of course means that there is no need to wait for the users to produce new information to study.

Overall, a very important advantage that comes with gathering data, from already existing services, is that a huge amount of money can be saved, which instead can be invested towards performing the actual research. The reason for this is that no web community has to be developed but also, there is no need of marketing the new tool since the users will already exist. The money can therefore be put into developing the analysing tool which had to be developed anyway. Therefore, lots of money can be used for other purposes in this solution.



3.2.1 Legal Issues

About the ethical and legal aspects of gathering data from existing resources, there is a lot to have in mind. Due to the fact that this system will run transparently to the user, in theory this could be done without the user's knowledge. Therefore, to perform data gathering from existing resources from other service providers there needs to be an explicit consent from both the service providers and the users of the service for it to be both legal and ethically correct. Also, the system must not be able to trace single users as anonymity is of utmost importance. The data could, after all, contain very sensitive information and needs to be kept private.

When it comes to data mining, users have no legal recourse if they have no "expectation of privacy." "Expectation of privacy" here means that a "reasonable person" would expect that the item in question is secret. What a "reasonable person" is when it comes to Cyberlaw is a topic for debate, but it is widely accepted that forums that do not require a login to read (reading is the only access required for data mining) provide no expectation of privacy.

The opportunity to gather data from private e-mails might be tempting and could certainly be a valuable resource. However, since it would be hard to get a proper consent to analyse the information when taking the laws governing the privacy of e-mail into consideration. Also, it is important to keep some channels free from analysis for those who do not wish to participate in the research.

3.2.2 Deployment and Marketing Costs

Deployment of this solution is quite different from the first solution. Instead of needing to host a new application on an Internet facing server (using a hosting service such as Go Daddy), this solution could simply run on a computer(s) with Internet access. In order to provide information and reports to users, a simple web server could be installed on the same machine as the main application and allow users to access it via the internal network.

Marketing is also completely different for this solution. Making the assumption that the new tool would be developed to interact with already popular tools; the marketing of those tools is not an issue. It will be handled by the companies behind those tools themselves. However, it would be advisable to use word of mouth for these tools. This can be effectively done by doctors letting patients know of the tool out there where it not only provides support for the patient, but also provides valuable feedback to researchers. The effectiveness of the method lies heavily in the value of human interaction over versus computer interaction; users will be much more likely to trust a website recommended to them by a medical professional versus a site they find from Google. If this is emphasised enough by doctors to their patients (aided perhaps by a handout as a reminder), it will go a long way to bringing users to the site.

3.2.3 Technology Issues

Without a doubt, the biggest problem we will face is the non-uniformity of the data source. There will be a lot of different formats to handle in the appropriate manner. One of the service providers might use a forum, for their users to communicate; another service provider might use a diary structure, where other users can comment on the diary posts. To be able to handle all these kinds of media it is required to convert all the data to a somewhat standardised form. A different conversion method would be needed for each different data sources, potentially creating a large amount of development time. In addition, service providers that update or change their services may force an update or rewrite of the conversion tool.

Once a uniform data format is established, the analysis tool would not be required to change. However, an initial development time would remain for the construction of the analysis tool.

3.3 Summary

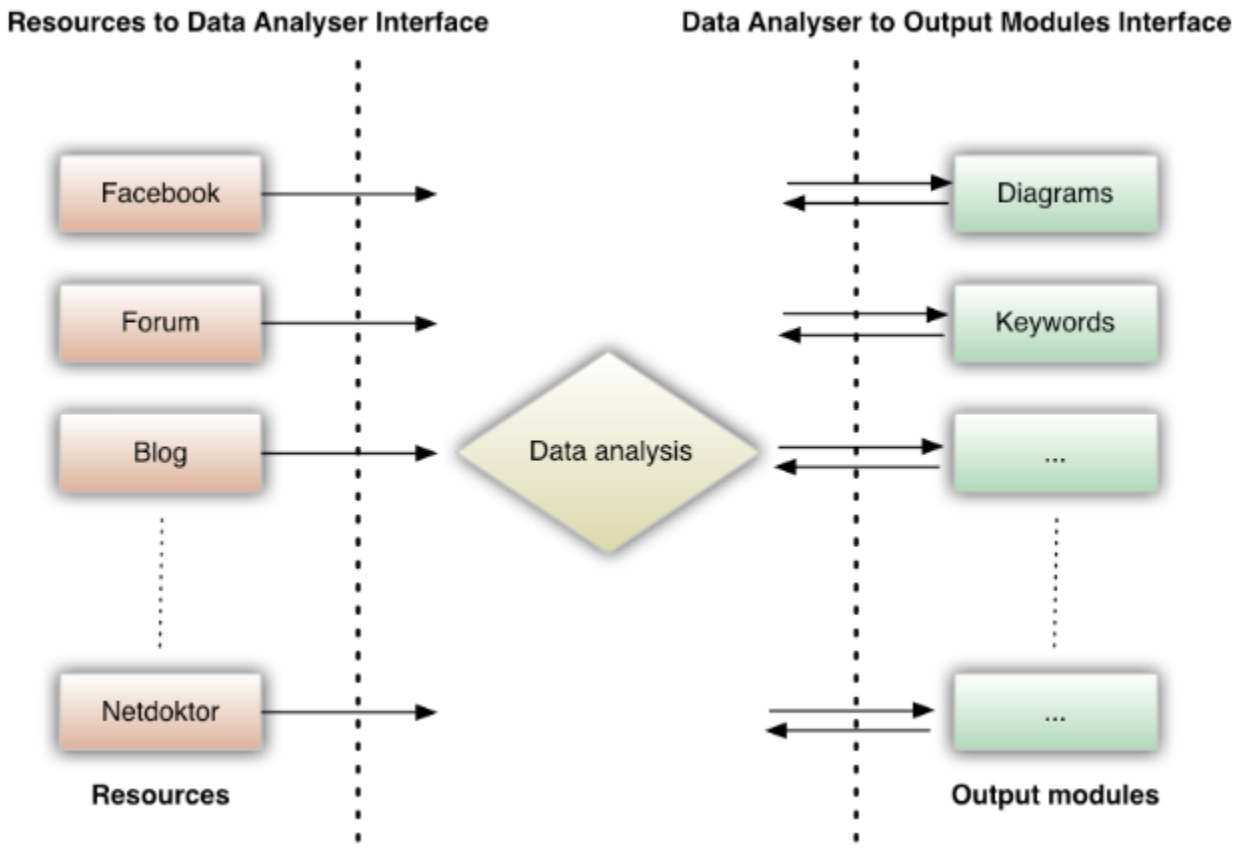
There are a lot of advantages with the solution that gathers information from already existing sources which makes it the best alternative of the two. Using it will save time, money, resources and it will also have the possibility to gather information from the past. There are advantages and disadvantages of both solutions. The major advantage of the solution to build a new web community from scratch are that the web community would be fully under the control of the development team. This would mean that it would be much easier to get access to the data that users submit to the site, and the site could be designed such that it looks and feels how the client wants it to look and feel.

An issue with this solution is negotiating with the people that are running the current sources about using the analysing tool on their web community. They may feel that they will lose users or that the privacy of their users will be violated by having this third party application using their information.

However, the advantages of gathering from existing sources are superior to the advantages of the new web community.

4 Design of a Possible Future Tool (Technical)

4.1 General Design



The general idea behind the design is to gather information from various sources, analyse it, and give some valuable results. To make the system flexible and not limited to a fixed number of sources, we are using a module based approach both for input and output. Every input module gathers data from a particular source, and then it is transformed into a standardised format before it is forwarded to the data analyser unit. This processes it and sends it to one or more output modules, depending on the request from the user. The output modules later present the information to the user.

4.2 Modules

It is seldom a good idea to create systems where all the parts/units are integrated into each other. It will result in a system that is bound to its original structure and properties. If the developers sometime want to customise and/or expand the system, changing the already existing parts becomes inevitable. Instead the optimal solution would be to implement a modular design. The main idea behind modularity is the possibility to add and remove units without having to modify the rest the system. For example let us look at the car industry. Buying a car nowadays means a lot of choices and decisions. You are not just buying the main product/system (the car itself) you are also making decisions whether to have the extra sound system, size of tires and rims, GPS and DVD-kit. You can either choose to have all or some of those things or you can choose to go with the basic product (the car). All these examples of choices you have are possible through modularity. In a nutshell modularity is about flexibility.

Considering our current situation and the needs of our system, a modular design suits us very well. The ability to add modules to the system is crucial as the Internet and everything surrounding the information society we live in changes constantly. Today the main means of communication for people with rheumatism is forums and communities. Tomorrow it might be something completely different. Therefore we have to apply a strategy that allows change and modular design.

How do we make use of modules and how do we implement a modular system? Our system contains three main parts: input, data analysis and output. It is in the input part of the system where modularity is most crucial. The set of different input models need to somehow communicate with the rest of the system. The different input models achieve this by passing a standardised text format of information and by doing this the rest of the system does not need to know where the information came from and neither adapt to the original format of the information. Having this kind of approach opens up for the possibility to add infinite kinds of different input modules without causing any effect on the rest of the system. The same ideas go for the output modules. The data analysis unit presents a standardised text format that contains the data. This data can then be processed by several different output modules that present the data in different ways ranging from diagrams and figures to plain text.

4.3 Gathering of Information

Information is gathered through the use of different input modules. An input module gathers information from a specific source such as a forum. It then transforms the information to a special format that is suitable for the analysing unit. Several modules can be active at the same time to allow extensive gathering of information.

How data is gathered from the various sources differs a lot. The only thing we can know for sure is that we have some form of information source that we want to use. We then gather that information and translate it into a form that is understandable for the analysing unit. The most common format of information that is going to be gathered is plain text. So what exactly is going to be different when reading information from various places? Some information might be stored in large SQL-databases, some might only exist on a webpage and some information sources could be ”live only” and not stored anywhere, like a chat conversation. What kind of sources to gather from is really something dynamic and could be anything. This was just some examples.

Some modules might require external software (that one could say is a part of the module). Let’s say for example we want to develop a module to gather information from a forum. First we have the main module that is executed when the program is looping through all loaded modules. But how will our forum module actually get hold of the forum-data? It could communicate in some form with an external piece of software running on the forum server. That software, in this example, is reading information from a database and returning it to the main-module, which converts it and sends it to the analysing unit.

The only thing that the main program needs to know is that you can add modules which in some way are giving information in a standardised form. If this format is well specified, a developer of a module does not need any knowledge about how the main program works to implement a new module. So how a module gets information or from where does not matter. As long as the last thing the module does is to convert the information into the standard format it will be compatible with the main program. This is an important part of developing this kind of tool since new sources will be created. These new sources may also use a new way of communicating between users (new forums, blogs) and therefore, modularity is important.

4.4 Analysing Data

4.4.1 Input Interface

The input interface abstracts the underlying technical solution from the way to access it. It should support any interaction that is required to reliably and securely collect data from the various platforms used by the system. To support various means of collecting data it might be necessary to observe the system as a whole as a distributed system. There are a number of ways to implement this, but the final decision on a solution will probably rely on both technical considerations and response requirements. A cross-platform, and widely supported, solution could utilise web services, at the cost of a bit of overhead and speed. At the opposite end of the scale all communication between the subsystems could be implemented with the native methods of the programming language of choice, which would put restrictions on the system but would probably be considerably faster and more efficient.

One of the most crucial functions of the input interface is to authenticate the incoming connection requests. This authentication mechanism should be two-way, which means that once authentication is completed both the server as well as the client can be certain that they communicate with a real entity in the system, and not an attacker. The motivation for the client to require this insurance is primarily to avoid disclosing possibly sensitive data to a third party. The reason for the server to want to have this insurance is to accept only accurate data from registered entities to avoid corruption and other attacks on the data or system as a whole.

A mutual challenge-response system in which the session key is derived from both the server's and the client's challenges, is a suitable authentication mechanism. The challenges can be exchanged either with public key encryption techniques, in which case each registered data gatherer has a public key that is stored at the analyser and the public key of the analyser is stored in each data gatherer, or by the use of certificates.

Verifier

The provided data will consist of metadata to be able to separate different inputs and also the actual message, i.e. a forum post, blog message, etc. Different types of metadata that could be used are:

- Source name - To separate the sources that the data is provided from, i.e. forums, blogs, etc.
- Entry Info - Info about the actual data
  - EntryID - To identify the data
  - ReplyTo - Refers to another EntryID, or NULL
  - Entry length - size of the data
  - Entry title
  - Entry thread - If the source is a forum, the forum post belongs to a certain thread
  - Date created
  - Date updated
  - Date extracted
- User info - Contains interesting information but still maintains user anonymity
  - Age
  - Sex

To be able to separate different inputs, some of these data fields will be mandatory. If a mandatory field within an input is empty, the data will be discarded. If an input cannot be identified, it cannot be included in any statistical outputs and therefore the data is useless and can be discarded.

The different fields that will be mandatory are:

- Source name
- EntryID
- Date extracted

Without knowing the source of the data, a statistical analysis will be misleading since the language that is used in forums differs from blogs etc. Forum posts are discussions while blogs might be written with other purposes. Therefore, the source of the data needs to be provided for more correct statistical analysis to be possible.

Since there might be replies in a forum thread after the data is gathered from the forum, an entry identification number needs to identify the different data that gets collected by the data gatherer. This unique number will be referred to as an EntryID. This ID can be used to make sure that the same forumpost is not gathered twice from the same source, this has to be avoided since it will make the statistics useless if there is no guarantee that the information is unique. The ID will also be used to know if a post is a reply to another post by being added as a reference in the database to the message that it is an answer to.

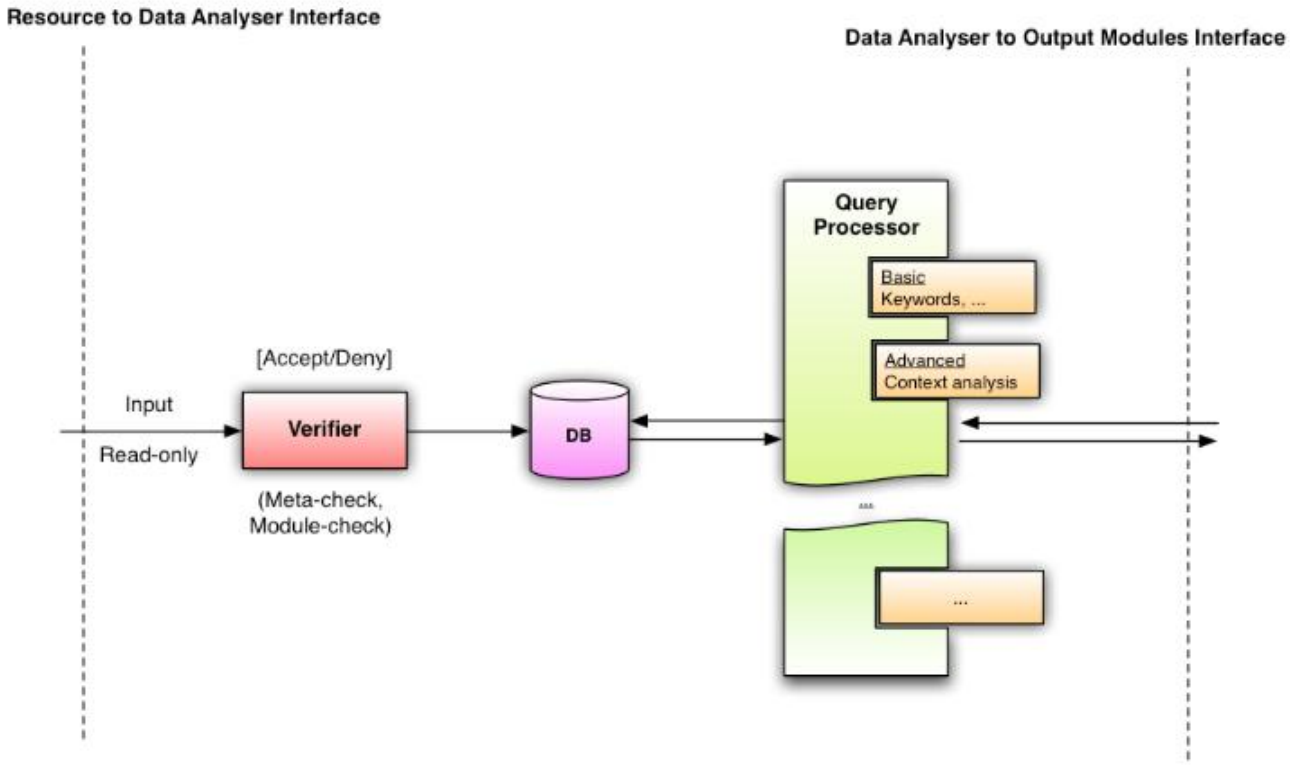
Different sources may provide creation date and updating date in different forms and some forums may only provide the date of the latest update and not both. Therefore, using date created and date updated as mandatory fields will not be possible. The date when the data was extracted will be used as a timestamp for that input. By using a timestamp and the module name telling the source of the data it will be possible to make statistical analysis of data from source “Facebook” between the dates x and y.

Depending on the source of the data, a variety of information is provided. This information might be different from, e.g. forums and blogs. A forum uses threads where the user posts their comments and replies to other comments. Every comment can have its own topic and comments can be edited by the writer at any point. All of these actions provide information about the input and the data gatherer should retrieve these different information fields if they are available. These metadata types belong to EntryInfo and it will be used to get more accurate statistical analysis when needed.

For statistical analysis, information like gender, age etc. is interesting. In this case, the users should be kept anonymous so there is only a limited amount of data about the users that can be used. Gender and age is information that keeps anonymity but still provides a variety of statistical analysis opportunities.

When the input has passed the initial checking, it is stored before any analysis is performed. This ensures that all inputs are stored and can be manually analysed even if automatic analysis fails. It also provides the data with an input id that can be used as a reference for future access. This unique input id can be the same as the EntryID.

4.4.2 Data Analysis



Data Search

Data is useless without interpretation. Just as the information available on the Internet is useless to most without a guide such as a search engine, so are the communications between rheumatic patients too vast for interpretation without some sort of aggregation. The techniques for aggregating this information are many and varied, but here is a recommendation of plausible techniques for this particular data set.

The driving force behind all of this information extraction, the data, requires some specification. For the sake of information extraction, we will assume that the communications we are analysing are constructed in sentences with correct spelling and punctuation. It is important to understand that this assumption is inaccurate. However, it is necessary for some of the data mining techniques discussed below. Pragmatic implementation of a solution must include considerations for corrections or exclusions that need to be applied to the data to allow for meaningful interpretations of the data on the part of the program.

- Keyword Search
- Keyword Search with Synonyms
- Topic Search via Automated Tagging
- Topic Search via Manual Tagging

Keyword Search

Keyword search is the simplest data gathering technique and is already in place for many systems. This technique involves little to no processing, only a simple query of the data set. It does, however, provide a starting point to move on to the option to search for keywords using synonyms.

Keyword Search with Synonyms

Keyword search with synonyms is a slightly modified keyword search. In this technique, words are grouped into collections of synonyms. Words not in the source dictionary are simply treated as having no synonyms. This provides the user with larger quantities of data per search, but requires more implementation effort. It's important to note that although this is called “Keyword Search with Synonyms” that it traditionally includes things that aren't strictly synonyms. Specifically, it includes alternate forms of the word in question, such as plural and singular forms, various conjugations, and alternate spellings.

Topic Search via Automated Tagging

The next logical step in data processing is to group words once again. From the synonym keyword search above, we add categories to groups of synonyms. For example, “ball” and “sphere” could be grouped into the synonym group “sphere” before being grouped again by the “shape” tag, which would be applied to other synonym groups such as “prism.” Implementation must be careful to consider the source data for this technique because it is very easy to be too broad in creation of tag dictionaries. For example, creating a tag “body parts” containing all of the relevant synonym groups would make the tag too broad for a data set that deals largely with medical conditions as this project's data does. It is easy to imagine, however, that there could be situations in which a general search would be useful. To remedy this, it is important to allow for tags as groups for other tags. For example, the “body parts” tag could contain “hand bones” and “leg bones,” each of which would contain their own relevant synonym groups. This would allow the user to search topics with varying levels of specificity without compromising the underlying keyword search functionality.

This data gathering technique can become very expensive to implement and maintain as tags are added for all of the relevant words. However, after the initial investment of implementing a framework, developers can add as many or as few tags as the client requires. One possibility would be to only add tags for relevant medical terms. While this would still require developers to construct a very large tag dictionary, it is a much more manageable problem than creating a tag library for the entire language.

Topic Search via Manual Tagging

A compromise between the precision of automatic tagging and affordability is manual tagging. Manual tagging has been the solution for many services with very large amounts of data, such as YouTube and Last.fm. The basic principle is to allow the users of the service to tag the data themselves as they view it. This works very well in the event that a new site is created as a result of this project, in which case tagging could be a part of data submission. However, it works more poorly on data mining for existing data sources (such as NetDoktor) because the only users available to do tagging would be the researchers who should be busy doing research and utilising the system, not maintaining it. That said, there is little reason that manual and automated tagging could not coexist. If properly implemented, the

combination of the two could provide very accurate feedback while allowing for manual corrections by users.

Unique Information

Unique cases are some of the most interesting pieces of information that can be gathered from the tool being designed. All of the aforementioned data gathering techniques can be coupled with a program that monitors new data for unusual qualities. In the case of tagging, the coupling of a “pain” tag with a “body part” which had not previously been associated would be of particular interest to the researchers. This would not be difficult to monitor. Similarly, keywords could be monitored for unusual combinations. However, unique information is much easier to discover with the tagging techniques because they provide a more descriptive representation of the data.

Implementation Suggestion

The data analysis can be performed by sub modules, each one performing increasingly complex and involved tasks. While some statistics can be precomputed at the time of data extraction, some calculations will have to be performed at a later stage depending on the query. This effectively decomposes the analysis phase into two separate phases.

The preparation phase of the analysis could involve recalculating keywords in the input text. The different keywords to use for this preparation could either be chosen from the search history or predefined in a configuration file. Other preparatory methods could be to run the input through a language interpreter to improve the structure of the text alternatively to find verbs to use as keywords. The precomputed data will be stored in the database together with its corresponding input entry.

The data analysis might also be used to calculate new information depending on the search query given by a user. This requires that the analysis modules are available to the Statistical Analyser at all times.

4.4.3 Statistical Analyser

The Statistical Analyser (SA) retrieves data from the database based on the results of a query given by the user. Before the data is presented to the user, it is converted into a standard output format. By using a standard output format, development of future modules will be easier. The standard output should have a detailed specification provided. That way, newly developed modules can retrieve the result of the query and later use it for analysis that was not in the developers thoughts on the day of creating this system. The standard output keeps the system module based.

The query specifies the information the user is interested in. The SA fetches the largest dataset satisfying the query constraints and runs any additional analysis required by the query on that dataset. The user might ask for statistics using a keyword that was not considered in the preparation phase and therefore, the SA has to consider the query and perform a new analysis of the data.

4.4.4 Data Storage

All the extracted data and subsequent analysis results must be able to be stored for later extraction and study. This requires some kind of persistent storage technique. There are a few different approaches to this, ranging from implementing persistence in the analysis tool to using some kind of database management system (dbms). We strongly encourage the use of a dbms since they are mostly highly optimised and well tested systems aimed specifically at organising and retrieving large sets of data. Another highly desirable property of a dbms is that they, mostly, support well-recognised query interfaces, greatly simplifying the tasks of storing and retrieving data. There are two major models to choose from when it comes to database management systems, either the relational model or the object model. Each model has its own advantages and disadvantages, and we will briefly discuss them.

The Relational Model [20]

The relational model is the older of the two models and built around a sound mathematical foundation of predicate logic and set theory. The stored data is arranged in tables that are linked to each other with unique identifiers called keys. The relational model has the advantage of being built around formal methods, and the correctness of the database can be guaranteed and proven. It also has efficient backup and recovery mechanisms, almost certainly guaranteeing that the database can be restored, without data loss, in the case of a system failure.

But by far the greatest advantage of the relational model is that it supports SQL (Standard Query Language). This is a very flexible language where the user states the result she wants, and the system internally translates this into the required operations to retrieve the data. The expressive power of SQL is very great and allows for almost all conceivable queries on the data, which enables for expansion of the output once the system is implemented and researchers begin to get new ideas of how they would like to correlate data.

The disadvantage of the relational model is that it has no direct mapping to the object oriented world often used in programming. This means that the system most likely will have to convert between the relational and object model, which could prove challenging.

The Object Model [21]

In the object model data is stored in the same format as it is represented in other parts of the system, as objects. Object databases have the advantage that they can be very efficient at specific tasks, at the cost of not being as general as the relational model. The object model is not based on a formal mathematical foundation and lacks a standard specification. This may lead to weaknesses in query support and might potentially hinder currently unknown queries that researchers want to ask in the future.

Object database management systems often lack interoperability due to the lack of standards, and this also affects functionality that is often taken for granted in relational databases, such as recovery, backup and connectivity. This means that once an object database management system has been chosen, it might prove exceedingly hard to change to another vendor if the need would arise.

Recommendation

Although the object model have some advantages such as efficiency under certain circumstances and no need to convert between objects and data representation it also have some disadvantages that have been discussed above. We feel that in a project that is still rather unspecified it is very important that new demands and requirements can be met. Since the relational model is based on a formal mathematical foundation, all major implementations of it supports SQL which allows very complex requests to be made when the need is discovered, and also good standards for backup, recovery and other functionality we feel that the safest and best decision is to use the relational model for storing analysis data.

4.4.5 Output Interface

The output interface interacts with the Statistical Analyser and the output modules to provide the modules with the requested information.

Authentication

Users, researchers etc. are given access to a number of output modules that provides the analysis the user is asking for. Before any information is shown to the user, an authentication action needs to be taken. This authentication can be a username and password that the user is provided to be able to login to the user interface to interact with the data and the output modules. By using a login session, the integrity of the data is protected and only authenticated users can use the system and retrieve the data.

4.5 Presenting the Information

4.5.1 User Interface

The user interface allows the user to interact with the data stored in the database by sending queries to the Statistical Analyser. The user interface will consist of two parts, a search field and a list of all modules that are added to the system.

The search area provides the users with options to put together a query to send to the database. The user can choose which input source, or sources, to use. Furthermore, the user can restrict the query within a time interval if that is preferred. The search options a user will have depends on the implemented metadata described above. All used metadata in the database should give opportunities to specify a search query. Choosing what input source, or sources, that are interesting is a choice the user can always choose since that information is provided by all input. Also the date when the data was extracted is always provided so choosing an interesting interval is always possible. Depending on what the source module is, different analysis can be made since all metadata are not provided for all module sources.

If the user sends a search query to the database and gets results, the user can now choose the type of module that would present the data in the correct manner according to the user.

In order to meet different users' different demands on presenting the data this part of the system is also module based. Every module takes data from the data analysis unit and presents it in different ways. It can reach from presenting diagrams to information in plain text.

When the data is prepared to the standard output format, the user can choose what to do with the data. The choices users have depend on the modules that are added to the system. The purpose of the modules is to take data that are converted to the standard output format and create statistics of the data.

5 Conclusion and Discussion

The client stated that there was a problem with finding out young rheumatics' wants and needs. This report has presented some possible solutions to this problem and especially focused on one solution where the information is extracted from already existing sources. One of the main concerns about this approach is the negotiation that has to take place between the users of the system and the information sources that will be used to extract information from. Some sources may not be very fond of letting a third party gain access to information written by their users.

Another concern is that the users must be informed that they are being watched by the researchers. The developers and users of this new tool must respect people's need of privacy. When people join a community of some sort they will have the feeling of that information that is said within the community stays within the community. This might be one of the reasons that people feel comfortable to express their feelings and opinions. If they know that a third party is watching and analysing what they write, will that affect how they express them self? This is something that needs to be considered if the system should be implemented.

Another thing that must be considered is how we weight information. Information that comes out of a patient's letter to a doctor probably contains more valid, and meaningful, information then a simple chat entry. As the system, through its modularity, will support the use of all kinds of different sources we must have in mind where the information was taken from. The language used in forums might differ a lot from the language used in other sources, blogs etc. Therefore it might be harder to extract high quality information from sources where the users are using a less formal language. To make researchers aware of this factor, the presented statistics should include the distribution of data over the different sources. This way, the researchers can judge the relevance of the statistics.

6 Acknowledgements

We wish to acknowledge and express our appreciation to the many people who contributed to the outcome of this report. We would especially like to thank the patients, researchers and the spokesperson for Reumatikerförbundet (The Swedish Rheumatic Society) for their cooperation and input.

We would also like to thank the faculty members Åsa Cajander, Mats Daniels, Cary Laxer and Michael Wollowski for their contributions to the project.

7 References

[1] Lovejoy, T., & Grudin, J. (2003). Messaging And Formality: Will IM Follow in the Footsteps of E-mail? Fairfax, IOS Press, Inc.

[2] IETF RFC 2821. (2007-10-05). Available: <http://www.ietf.org/rfc/rfc2821.txt?number=2821>. [2007-10-07]

[3] Wikipedia. (2007-10-21). Available: <http://en.wikipedia.org/wiki/S/MIME>. [2007-10-25]

[4] Wikipedia. (2007-11-16). Available: <http://en.wikipedia.org/wiki/Instant_messenging>. [2007-11-23]

[5] Wikipedia. (2007-11-16). Available: <http://en.wikipedia.org/wiki/Instant_messenging>. [2007-11-23]

[6] Wikipedia. (2007-11-05). Available: <http://www.internetworldstats.com/eu/se.htm>. [2007-11-08].

[7] Reumatikerförbundet. (2007-11-04). Available: <http://www.reumatikerforbundet.org/start.asp?sida=3583>. [2007-11-06].

[8] Spokesperson (2007). The Swedish Rheumatic Society. 2007-11-08, Uppsala.

[9] Comscore. (2007-11-19). Available: <http://www.comscore.com/>. [2007-11-23].

[10] Alexa. (2007-11-22). Available: <http://www.alexa.com/site/ds/top_sites?cc=SE&ts_mode=country&lang=none>. [2007-11-23].

[11] Facebook. (2007-12-05). Available: <http://www.facebook.com/press/info.php?statistics> [2007-12-05].

[12] BBC. (2007-11-05). Available: <http://news.bbc.co.uk/2/hi/technology/4530930.stm>. [2007-11-05].

[13] Cornell University, (2007-12-10). Avaliable: <http://delivery.acm.org/10.1145/990000/985709/p129-hancock.pdf?key1=985709&key2=3471489911&coll=GUIDE&dl=GUIDE&CFID=11517622&CFTOKEN=98215376>. [2007-12-10].

[14] CHI 2004 Volume 6 Number 1, April 24–29, 2004, Vienna, Austria.

[15] JCMJ (Journal of Computer-Mediated Communication), (2007-07-27). Available: <http://jcmc.indiana.edu/vol12/issue4/qian.html>. [2007-12-08]

[16] Läkartidningen. Nr 23, 2001, Volume 98. Stockholm, Sweden.

[17] GoDaddy. (2007-11-13). Available: <https://www.godaddy.com/gdshop/hosting/shared.asp?ci=9009>. [2007-11-14].

[18] Google. (2007-11-22). Available: <http://www.google.com/technology/>. [2007-11-22].

[19] Google. (2007-11-24). Available: <http://www.google.com/support/webmasters/bin/answer.py?answer=35769>. [2007-11-24].

[20] Wikipedia, (2007-12-13). Available: <http://en.wikipedia.org/wiki/Relational_model>. [2007-12-03].

[21] Wikipedia, (2007-12-13). Available: <http://en.wikipedia.org/wiki/Object_model>. [2007-11-27].