A Conversational Agent for Use in the Identification of Rare Diseases

Ana Olivia Caballero Lambert[1(✉)], Cesar Horacio Torres Montañez[2], Monica Bueno Martinez[2], and Marcelo Funes-Gallanzi[3]

[1] Universidad Iberoamericana, León, Guanajuato, Mexico
ana.caballero@leon.uia.mx
[2] Natura Xalli, Tepic, Nayarit, Mexico
{cesar.torres,monica.bueno}@naturaxalli.com
[3] AVNTK, S.C., Guadalajara, Jalisco, Mexico
mfg@avntk.com

Abstract. This paper presents work in progress on implementing a conversational agent, Dr. Rachael, in the form of a virtual caregiver, for use in helping to identify rare diseases. The rationale for the system is grounded in the fact that rare disorders by their very nature are difficult to diagnose unless the caregiver or doctor is familiar with a wide range of these conditions. The conversational agent uses unstructured free-flowing natural language together with a large database of rare disorders, and is easily updatable by human caregivers without any technical expertise. Matching of users' comments with database entries is performed using a general cognition engine; which is able to understand natural language regardless of specific wording or grammar. In this paper we give a comprehensive background to and an overview of the system, with a focus on aspects pertaining to natural language processing and user interaction. The system is currently only implemented for English.

Keywords: General cognition engines · Artificial cognition systems · Diagnosis · Rare diseases

1 Introduction

Rare diseases pose a large challenge to society, and gravely affect the lives of many children. Approximately 75 % of rare diseases affect children and 30 % of rare disease patients die before the age of 5. There are approximately 7,000 rare diseases identified. A rare disease is a life-threatening or chronically debilitating condition. A disease is defined as rare in Mexico when there is one sufferer per 2000 inhabitants (or nearly 10 million people currently), in Europe when fewer than 5 in 10,000 people are affected, and in the US when fewer than 200,000 people are affected. Almost 5 % of the world's population are living with a rare disease [1]. The longer it takes to diagnose a rare disease, the more physicians the patient needs to see. 40 % of rare disease patients are misdiagnosed at least once. For people with a rare disease, the mean average length of time from symptom onset to accurate diagnosis is approximately 4.8 years [2]. People with a rare disease may experience low quality

E. Sucar et al. (Eds.): AFI 2016, LNICST 179, pp. 128–139, 2017.
DOI: 10.1007/978-3-319-49622-1_15

of life and high levels of disability. This difficult journey to diagnosis and care can increase medical, economic and social burdens [3]. Rare diseases often hide behind common symptoms of other more common illnesses, making the diagnosis extremely challenging and often leading to mis-diagnosis. Despite progress made in this area, there remains a need to better understand the obstacles patients and caregivers within the rare disease community face in obtaining a correct and timely diagnosis. The path to improve diagnosis can only be helped by all stakeholders joining together to find solutions to help ensure accurate diagnosis of rare diseases at an early stage. An accurate diagnosis may be the first step to improving the care for those living with a rare disease and their families. In more than 50 % of cases it is those living with a rare disease, or their families, who ultimately manage to discover the correct diagnosis of the patient and inform the doctor. There are resources such as http://www.ncbi.nlm.nih.gov/pubmed/ & http://rarediseases.org/for-patients-and-families/information-resources/rare-disease-information/ but these are not easy to use, they list conditions alphabetically and assume the user has a relatively high level of scientific knowledge that normally lay people do not possess.

There is a need therefore for a user-friendly tool to help families or caregivers find the correct diagnosis for a patient at an early stage, without limitations of time, a specific doctors' knowledge or any preconceptions. We are committed to helping raise awareness of rare diseases, including the diagnostic challenges, and to implementing initiatives to support an improved diagnosis journey. One such initiative is to make available a conversational agent, initially provided with knowledge of about 1,000 diseases, specifically targeted at rare diseases patients and their families. Our aim is that the system could form a part of the tools available to doctors in the future. In this paper we give an overview of the "Dr. Rachael" system, its rationale, design and implementation. The next section gives a background to and a rationale for the system. We then describe the conversational agent and the user interaction in Sects. 3 and 4. Following this, in Sect. 5 we describe the characteristics of the domain and the users, which lead to the choice of a general cognition engine to do the matching of users' comments with the available knowledge, which described in Sect. 6. We then describe the structure of the database, the interface for updating this knowledge base, and some sample interactions in Sect. 7. Finally, in Sect. 8, we present some conclusions and suggestions for future work.

2 Background and Rationale

Patients with rare diseases form a rather heterogeneous group with respect to their thoughts and questions regarding their disorder; especially in the early stages. Anecdotal clinical experience indicates that patients tend to ask the same kinds of questions using similar language for a given disorder. The aim of Dr. Rachael is to be able to interact in an easy-going non-structured way for as long as is necessary in order to provide likely leads for the condition affecting a patient, at times when a human caregiver is not available. Dr. Rachael is a conversational agent powered by a general cognition engine, embodied as a virtual caregiver, accessible through a web interface. Comments and questions are posed to the system using written natural language. By providing an

embodiment of a conversational agent, the hope is to support a more natural feeling dialogue, finding likely leads and to promote users' trust in the answers. In addition, it is hoped that such a system will be user-friendly and easy-going in providing suggestions and advice, lessening the focus on the patient's problems.

The reader might raise concerns regarding the use of a conversational agent in helping to find a likely diagnosis for patients whose condition could be long-term and life-threatening. The rationale for the use of Dr. Rachael is that research by the patient or his family, who are prepared to put in the time and reading necessary, is often the way the correct diagnosis to rare diseases is reached and any means that can be devised to shorten the time to reach the correct diagnosis should be tried. The scientific basis for this rationale is discussed in detail in [4]. By focusing on providing answers to questions related to the likely disorder, based on scientific facts and expressed in a professional, yet comprehensible, encouraging, motivating and empathetic language, the hope is that the system will help in disrupting the worries and thoughts of the patients. There are of course potential problems with Internet based treatments and diagnostics tools. For example, Internet access might be a limiting factor [5]. However, for instance a survey shows that 95 % of the Swedish population in the age group of 15-24 year olds, have access to the Internet in their homes, and that 86 % use the Internet on a daily basis [6], the statistics being quite similar throughout the developed world. Furthermore, it should be noted that the aim of leveraging a virtual caregiver is not to replace human caregivers. Instead, one should view these as complementary parts of the long and often fraught diagnosis process for rare diseases. More generally, Internet-based treatments have previously been used successfully for example in the treatment of agoraphobia [7] and ADHD [8]. These naturally focused on the psychological aspects of the disorders and were based on Cognitive Behavioral Therapy and working memory training, respectively. The present work is not directly inspired by this previous work, since the nature of the disorders are very different.

3 Conversational Agent

While one should not take for granted that the use of a conversational agent will benefit all information system applications, there are some associated properties that make them especially beneficial to this specific application. As discussed earlier, patients with rare diseases find it hard to obtain the correct diagnosis in a reasonable period of time, and often need to do some research on their own while not having much technical or scientific knowledge.

Numerous studies have shown that these aspects are not currently being addressed effectively. For example, some ideas present in the system can be conveyed more directly, by use of facial expressions and body gestures, than is possible with text alone. Studies have also shown that by using friendly small talk, verbal and non-verbal expression of emotion and expressions of expertise, empathy and trust can be conveyed [9]. Furthermore, by using relaxed knowledgeable artificial characters a sense of confidence, which can have motivational effects, can be conveyed [10]. Dr. Rachael is capable of holding a normal conversation discussing a range of topics, as well as having a large

amount of knowledge about symptoms of more than 1,000 rare diseases; while body movement, speech generation and recognition are currently being implemented.

4 User Interaction

Although we make use of a conversational agent, the primary communication channel used in the system is written text. Users interact with Dr. Rachael by posing questions using natural language entered using a standard text-field, see Fig. 1. If a comment is made by the user, Dr. Rachael will make a suitable related comment, based on a association of ideas very much as a person would. If a posed question has a high similarity with a question in the database, the answer is presented directly to the user by Dr. Rachael. If no such question is found, the closest available general comment is given.

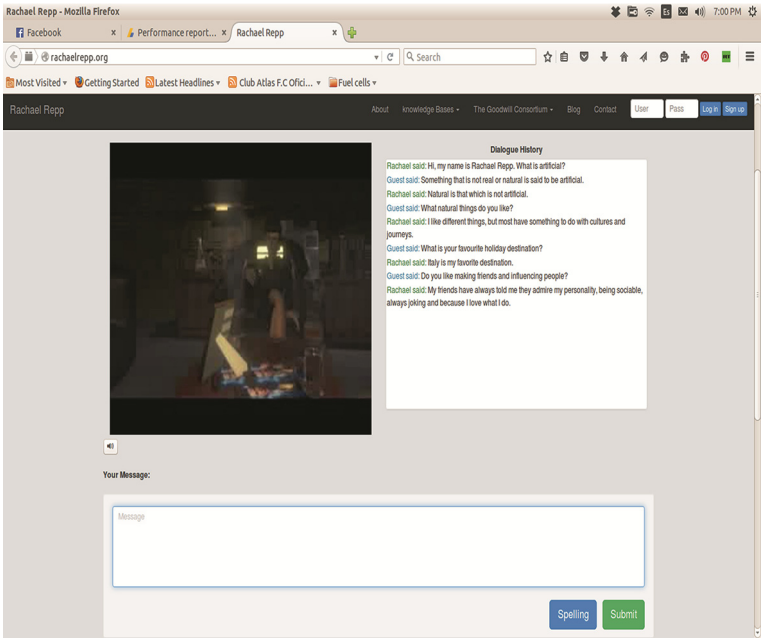

Fig. 1. Screen-shot of Dr. Rachael answering the question: "Do you like making friends and influencing people?"

Thus, a conversation can range from discussing general concepts to replying to specific questions, as shown by the following example:

Rachael said: Hi, my name is Rachael Repp. If You were a spy what would your alias be?
Guest said: I would like to be called Superman.
Rachael said: I would like to be remembered as a good person.
Guest said: Are you good or just want to convince me you are nice?

Rachael said: I am just a girl who knows what she wants.
Guest said: I often have a stomach ache.
Rachael said: You are a collective of every experience you have had in your life.
Guest said: I have lost weight recently.
Rachael said: Weight loss is common in people with Crohn's disease.
Guest said: I sometimes have constipation, pain and swelling in the stomach.
Rachael said: When Crohn's disease causes intestinal obstruction, the symptoms may include pain, constipation, swelling of the abdomen, and/or vomiting.
Guest said: What is it called when you get an infection by the bacteria that are in the intestines?
Rachael said: The name is septicemia.

After each session, the user is provided with the most likely condition that the user or family member is thought to be at risk of, and a written pre-formulated general description is given, so that suitable health care services can be contacted. Each comment most of the time can relate to more than one condition so counters are kept, and when a given statement relates to more than one condition all conditions counters are increased by one. If a given statement is related to only one condition, then only the counter associated to this condition is increased. The final rating of the patient's likely condition is calculated based on which counter is largest, and where there is not sufficient confidence of a given diagnosis, the three top conditions are displayed, yielding a probability that a like diagnosis has been found. The user can also see, at the end of the session, a short description of the condition(s) most likely to apply to the symptoms discussed, see

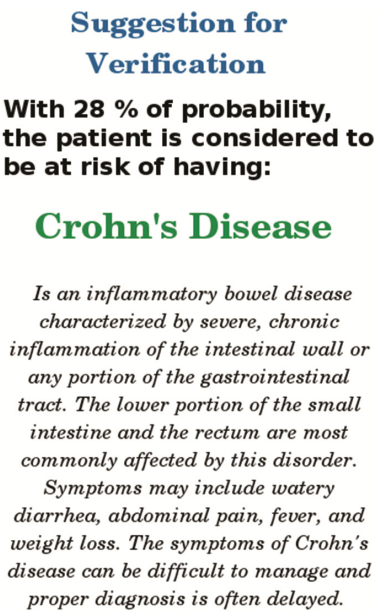

Fig. 2. Screen-shot of the "Suggestion for verification" showing a brief text giving a condition the patient is considered to be at risk of.

Fig. 2. This feature allows the patient or the family to get a firm suggestion even if the session lasts a long time or is even completed over many sittings, hopefully helping in the right direction towards a suitable diagnosis. This kind of feedback is an integral part of the method on which Dr. Rachael is based, which has been previously applied by our group to the diagnosis of eating disorders and personality tests.

5 Domain Characteristics

A number of characteristics of the domain and the patient group, in combination with the rather small scale of the project so far, has influenced the choices made in designing the system. First of all, the domain which the system needs to cover is too broad for the system to cover using limited resources. Secondly, since the system may be employed in a real-world situation either by doctors, patients or their families, users with no technical training should be able to quickly expand the knowledge base using only natural language, in case the system lacks coverage of a certain sub-domain. It follows that the knowledge base in principle should consist of a set of concise information field together with more or less frequently asked questions and answers to these questions.

Matching users questions to this database on the surface looks like a standard information retrieval (IR) task. However, the domain and task at hand differs from those considered in traditional IR in a number of ways:

- Large data set – the number of answers is quite large and apply in many cases to more than one condition (currently about 100 out of 1,200 considered in the first stage of testing).
- Short description – each condition is described by its symptoms, some questions and answers and a short description, so quite a few words (on average about 300)
- Narrow sub-domains – several descriptions relate to distinct conditions, although many details apply to more than one condition so there is a large amount of overlapping.
- Stressed, fatigued users – patients or carers often find it hard to concentrate for a long period of time, and so might need to come back many times and perhaps benefit from some light entertainment!

Traditional IR methods, using a traditional bag-of-words (BoW) representation or latent semantic indexing (LSI) [11] for instance, work well on large data sets, with longer documents, however they often place a high burden on the users to have a lot of knowledge, both of IT and the specific field, and sift through search result sets in order to find what they are looking for. In the case of patients with rare diseases, we want to keep session results down to a minimum, preferably delivering the correct diagnosis in answer to a conversation directly if a sufficiently high probability of disease identification has been achieved, or at least to keep details through a database, user code and password for the user to come again. Due to the wide variation in educational level and vocabularies, the system must handle morphological variation and possess a considerable vocabulary. The system should also be able to handle spelling errors gracefully, since the users are often to some degree fatigued spelling errors are to be expected.

Furthermore, since the users interact with the system in English, both grammar and vocabulary can vary widely so natural language processing must be handled appropriately. These issues are not as important when many long documents are drawn from a large domain, and when there is no limit on search result size, since the probability of a certain morphological form, or spelling variation, occurring at least somewhere in at least some relevant document is high.

6 General Cognition Engine

As previously discussed, standard IR methods are not suited to the present task. Instead we need a representation and similarity measure that handles morphological variation, spelling errors, and compound words.

In order to facilitate interaction, we firstly acquire information by converting all text to a simplified version using Basic English implemented in a proprietary tool called "Simplish" capable of understanding standard English of 100,000 words. We then create a multidimensional ideogram to represent each sentence, similar to the logic used in creating Chinese symbols, and store these in the memory of the general cognition engine in such a way that two sentences having the same meaning, regardless of vocabulary or grammar, result in ideograms of similar shape and position in the memory space. Details of this method are beyond the scope of this article and will be the subject of a related publication in the near future. Memory retrieval is achieved using a nearest-neighbor search of ideograms in 12 dimensions. We have therefore enabled the computer to "understand" what the user says, moving artificial intelligence from "pattern matching" strategies to "concept matching".

Possibly the most fundamental aspect of cognition is memory, which is itself broken down into three distinct tasks: acquisition, storage and retrieval; bearing in mind that it is well-established that order helps memory. Knowledge and experience are most easily transferred through language and at the root of the concept of language lies the very definition of a word. For centuries words like "virtue" have defied an exact generally-accepted definition, though this is a subtle complex term. However, as first noted by Wittgenstein (1953) even simple words like "dog" or "game" can be equally challenging. His suggestion was that there exists a "family resemblance" which allows us to identify a particular instance as a member of a group such as a "dog". Using the idea of meaning being defined in terms of a word's relationship to others is attractive but involves deriving a matrix of word relations that implies many more entries than the average number of neurons in a human brain, if a standard vocabulary is used. This fact leads us to two conclusions: first, that it is likely that basic-level categorization is in fact used by the brain, and second that it would therefore be useful to find a way to represent a full vocabulary in terms of a reduced vocabulary, that can act as a proxy for this set of basic-level categorizations. The conversion of text into a reduced-vocabulary form is achieved by *Simplish*, a proprietary tool that converts standard English into a reduced-vocabulary version of 1,000 words - 850 basic words, 50 international words and 100 specialised words – said reduced vocabulary being employed as a proxy for a set of basic-level categorizations. This representation therefore yields an effective means of

knowledge acquisition. A reduced-vocabulary also has the advantage of reducing ambiguity as a by-product of the translation process.

Using a reduced-vocabulary representation of knowledge, where words have been associated to each other, enables the mapping of language, through the use of standard multivariate techniques, to a low-dimensionality space where a multidimensional ideogram (a graphic symbol that represents an idea or concept) can be produced as an illustrative point in this subspace, as might be represented in the brain (using variables such as coordinates x, y, z, potential, neurotransmitter, frequency, phase, etc.). The low-dimensionality ideogram is the storage medium. These ideograms will be similar even if different words or grammar are used, because their form and position is given by the relationship of a semantic unit to all other members of the basic-level categorization.

The result of this strategy is to establish a means to map semantic similarity into spatial and shape proximity, i.e. the distance between two points (concepts), or between all points in more complex shapes such as those due to ideograms, is a measure of how similar their meanings are. Spatial proximity can be used to yield a means for information retrieval for a given query, via the association of ideas as a human does, implemented here through the application of nearest-neighbour search algorithms for instance, a method that is well-known in the art, either to a single point or to sets of points describing an ideogram. This strategy enables concepts to be mapped either as part of a data-driven step or concept-driven strategy for contextual information needed in problem-solving. We can plot and step along an evolving path and come across both types of information if relevant.

This approach enables us to merge equivalent concepts, do summaries, order data, find relevant contextual data easily, identify sentences having similar meaning, do true abduction (as originally proposed by C. S. Peirce) on knowledge, and implement a means to identify metaphors. It basically places knowledge in a position and shape depending on meaning and so requires no prior training to understand any given topic, hence the term "general cognition engine". Note that no pattern-matching method, ontology, topic map, or prior training of any kind is used. Current work is centered on deduction using the STUDENT world-algebra algorithm [12] combined with the E-theorem prover from the Technical University Munich [13], while decision-making is achieved through a Bayes decision tree obtained automatically directly from the text. We have implemented two prize-winning (http://www.innovation-council.org.uk/2014-awards/) applications: a simple summary generator available in www.simplish.org and a conversational agent available at www.rachaelrepp.org that is being developed for competing for the Loebner Prize and on which Dr. Rachael is based. The system uses a single-processor implementation for the ANN, which is fast enough given the database size, but current work centers on implementing a fully parallelized version using the FLANN library [14] since as the size of the database increases this will be needed.

7 Knowledge, Comments, Questions and Answers

As described above, the knowledge base of the system consists of a knowledge base containing symptoms, a description and some questions and answers for each condition,

formulated in natural language. Each comment or answer can be linked to several different conditions and counters are kept for each. Comments and answers are assigned to one or more conditions. Responses are further assigned a mood, which is currently used to control the body language and facial expressions of Dr. Rachael. In order to collect data and involve the users early on in the design of the system, a small-scale pilot study was performed, covering only 150 conditions. The experiment resulted in an initial database of symptoms, descriptions, questions and answers, which were mostly based on the NORD database and have subsequently been refined for our purposes. It also gave valuable information on the characteristics of the language use by the patients, the degree of spelling errors and other aspects of language variation.

An important point worthy of mention is that not all rare diseases are difficult to diagnose, such as Fibrodysplasia Ossificans Progressiva or Tourette's. A sub-group of particular interest is that made up of conditions which are rare *and* difficult to diagnose, mostly because they are easily confused with more common conditions. It is this latter group upon which we have concentrated our attention.

A web-based "editor interface" aid in the expansion of the database is currently in progress. Using this interface, a human caregiver can add, remove and update symptoms, descriptions, questions and answers, add alternative formulations to an answer, and browse and search comments/questions posed to the system that were left unanswered.

While the database is necessary for the conversational agent, it could also be potentially useful in providing human caregivers with a common ground, so that patients are not given conflicting answers when asking different caregivers the same question. It could also potentially be useful in highlighting mildly frequently asked questions and topics, allowing the human caregivers to better prepare for these questions.

8 Conclusions and Future Work

In this paper we presented some early results of a conversational agent being used to aid in the identification of 150 rare diseases, based on one of the few available conversational agents able to "understand" natural language effectively. An initial study and some trials shows that there is indeed a need for this kind of system among patients with rare diseases, and suggests that the current system fulfills the demands of these patients satisfactorily. Currently, data is being collected to raise the number of diseases covered up to 1,200, the results of trials will then be used to determine the final refinement steps of the implementation and possible improvements to all aspects of the system. We have high hopes on the utility of the final system, which is to be finished later this year.

Although the current system is well received by medical practitioners, patients and family members, it is important to point out that it is an "informational" tool and there is yet no evidence that the system is an effective tool to help identify rare diseases. This can only be proved by conducting a randomized controlled trial in a clinical setting prior to it being regarded as a "diagnostic tool" for any disease. The system was developed on a small budget, using mostly our own proprietary software, some open source software and standard web technology throughout, showing that it is indeed possible to build usable systems in this domain as a small scale research project with limited funding.

Currently, the system is only implemented for English, but the natural language processing modules are in principle language agnostic. Of course there are still much room for improvement, both to the dialogue and interaction aspects and to the natural language processing aspects of the system. We are happy to receive comments and suggestions for further improvements from the participants of this symposium, and will look eagerly towards any offers of collaboration in this project.

Appendix: Crohn's Disease

General Discussion

Crohn's disease is an inflammatory bowel disease characterized by severe, chronic inflammation of the intestinal wall or any portion of the gastrointestinal tract. The lower portion of the small intestine and the rectum are most commonly affected by this disorder. Symptoms may include watery diarrhea, abdominal pain, fever, and weight loss. The symptoms of Crohn's disease can be difficult to manage and proper diagnosis is often delayed. The exact cause of Crohn's disease is unknown[1].

Signs & Symptoms

Crohn's disease typically affects the lower portion of the small intestine and/or the colon, especially the right colon. Occasionally, inflammation may occur in the middle and lower portions of the small intestine. In some cases, there is inflammation of the membranes that line the mouth, the esophagus, and/or stomach. The symptoms of Crohn's disease may begin abruptly or appear slowly over a long period of time. Symptoms that may develop over time include nausea, vomiting, fever, night sweats, loss of appetite, a general feeling of weakness, waves of abdominal pain and discomfort, diarrhea and/or bleeding. Weight loss is common in Crohn's disease.

Acute attacks of Crohn's disease may cause fever, elevated white blood cell counts, and/or severe pain in the lower right abdomen. These symptoms are frequently confused with appendicitis. Crohn's disease may cause lesions in the intestinal wall and the surrounding lymph nodes. Abscesses may occur before the appearance of other symptoms. Grooves on the inner surface of the intestines may also occur. These may feel like a solid mass in the abdomen and when the mucosal lining of the intestines becomes thickened, it may feel like cobblestones. Deep open abscesses, scarring, and some degree of intestinal obstruction may occur as a result of chronic inflammation of the intestine. In some cases, fistulas and abscesses may create an opening through the intestinal wall and result in infection by the bacteria that occur naturally in the intestines. Massive, abnormal enlargement of the colon is a serious complication of Crohn's disease and may result in intestinal bleeding into the abdomen and septicemia.

When Crohn's disease causes intestinal obstruction, the symptoms may include pain, constipation, swelling of the abdomen, and/or vomiting. This may be due to the

[1] Source for the material: http://rarediseases.org/rare-diseases/crohns-disease/ .

accumulation of fluid in the intestines or thickening of the muscosal layers of the intestinal walls. Inflammation and obstruction may occur together and can impair digestion and the absorption of food and may lead to malnutrition. Crohn's disease rarely occurs in children, and is characterized by failure to thrive, fever, and/or abnormally low levels of circulating red blood cells. Children may also experience joint pain and stiffness. Growth and sexual development are often delayed. Initially, children with Crohn's disease may not experience diarrhea or abdominal pain. People with Crohn's disease may have anemia, abnormally low levels of albumin in the blood, abnormally high white blood cell counts, and/or a deficiency of vitamin B-12. Other laboratory findings may include abnormally low blood levels of sodium, potassium, calcium, and/or magnesium. Individuals with Crohn's disease may also have symptoms that are not related to intestinal dysfunction. These may include joint pain, or skin and eye problems. A fatty like substance may accumulate in various parts of the body. Blood circulation may be impaired by abnormally thick blood, dehydration, and/or lack of movement or exercise. In some cases, arthritis may occur resulting in swollen and painful joints. In rare cases of Crohn's disease, liver function may be impaired. In some cases, affected individuals may experience diminished bone mass resulting in thinning and weakness of the bones.

References

1. Aronson, J.: Rare diseases and orphan drugs. Br. J. Clin. Pharmacol. **61**(3), 243–244 (2006). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1885017/
2. Engel, P.A., et al.: Physician and patient perceptions regarding physician training in rare diseases: the need for stronger educational initiatives for physicians. J. Rare Disord. **1**(2) 2013. http://www.journalofraredisorders.com/pub/IssuePDFs/Engel.pdf
3. Rare Diseases UK. Key Statistics from the RDUK Report 'Experiences of Rare Diseases: An Insight from Patients and Families'. http://raredisease.org.uk/index.htm
4. Rare Disease Impact Report: Insights from patients and the medical community, Shire, April 2013. http://www.geneticalliance.org.uk/docs/e-update/rare-disease-impact-report.pdf
5. Carlbring, P., Andersson, G.: Internet and psychological treatment. How well can they be combined? Comput. Hum. Behav. **22**(3), 545–553 (2006)
6. Nordicom-Sveriges Internetbarometer 2007. MedieNotiser 2/2008. ISSN 1101–4539 (2008)
7. Carlbring, P.: Panic! Its Prevalence, Diagnosis and Treatment via the Internet. Doctoral thesis, Department of Psychology, Uppsala University (2004)
8. Klingberg, T., Forssberg, H., Westerberg, H.: Training of working memory in children with ADHD. J. Clin. Exp. Neuropsychol. **24**(6), 781–791 (2002)
9. de Rosis, F., Cavalluzi, A., Mazotta, I., Novielli, N.: Can embodied conversational agents induce empathy in users? In: AISB'05 Virtual Characters Symposium (2005)
10. Berry, D., Butler, L., de Rosis, F., Laaksolahti, J., Pelachaud, C., Steedman, M.: Final evaluation report, the Magister project. January (2004)
11. Dumais, S.T., Furnas, G.S., Landauer, T.K., Deerwester, S.C.: Using latent semantic analysis to improve information retrieval. In: Proceedings of CHI 1988: Conference on Human Factors in Computing, pp. 281–285 (1988)
12. Bobrow, D.G.: Natural Language Input for a computer problem solving System. Ph.D Thesis, MIT (1964)

13. Schulz, S.: System description: E 1.8. In: McMillan, K., Middeldorp, A., Voronkov, A. (eds.) LPAR 2013. LNCS, vol. 8312, pp. 735–743. Springer, Heidelberg (2013). doi: 10.1007/978-3-642-45221-5_49
14. Muja, M., Lowe, D.G.: Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans. Pattern Anal. Mach. Intell. **36**(11), 2227–2240 (2014)
15. Wittgenstein, L.: Philosophical Investigations. Blackwell, Oxford (1953). (G.E.M. Anscombe, trans.)