Meta
Journal des traducteurs
Translators' Journal

ΜΕΤΑ

The Social Model of Translation and Its Application to Internet Search Engines Specialized in Health: The ASEM Search Engine for Neuromuscular Diseases

Joan Miquel-Vergés and Elena Sánchez-Trigo

Cite this article

Article abstract

The use of the Internet as a source of health information is greatly increasing. However, identifying relevant and valid information can be problematic. This paper firstly analyses the efficiency of Internet search engines specialized in health in order to then determine the quality of the online information related to a specific medical subdomain like that of neuromuscular diseases. Our aim is to present a model for the development and use of a bilingual electronic corpus (*MYOCOR*), related to the said neuromuscular diseases in order to: a) on one hand, provide a quality health information tool for health professionals, patients and relatives, as well as for translators and writers of specialized texts, and software developers, and b) on the other hand, use the same as a base for the implementation of a search engine (using keywords and semantics), like the ASEM (*Federación Española Contra las Enfermedades Neuromusculares*) search engine for neuromuscular diseases.

érudit

The Social Model of Translation and Its Application to Internet Search Engines Specialized in Health: The ASEM Search Engine for Neuromuscular Diseases

JOAN MIQUEL-VERGÉS
Universidade de Vigo, Vigo, Spain
jmv@uvigo.es

ELENA SÁNCHEZ-TRIGO
Universidade de Vigo, Vigo, Spain
etrigo@uvigo.es

RÉSUMÉ

L'utilisation d'Internet comme source d'information médicale est de plus en plus importante; toutefois, repérer les informations significatives et valables peut être problématique. Le présent article analyse, d'abord, l'efficacité des moteurs de recherche en ligne spécialisés dans le domaine médical pour déterminer, par la suite, la qualité de l'information en ligne relative à un sous-domaine médical spécifique, comme celui des maladies neuromusculaires. Nous présentons un modèle de conception et d'exploitation d'un corpus électronique de textes bilingues (*MYOCOR*), spécialisé dans le domaine des maladies neuromusculaires pour: a) d'une part, fournir un outil d'information sanitaire de qualité pour les professionnels de la santé, les patients et leur entourage, ainsi que pour les traducteurs, les rédacteurs de textes spécialisés et les développeurs de logiciels; b) d'autre part, servir de base à la mise en œuvre d'un moteur de recherche (mots-clés et recherche sémantique) semblable au moteur de recherche portant sur les maladies neuromusculaires de l'ASEM (*Fédération espagnole contre les maladies neuromusculaires*).

ABSTRACT

The use of the Internet as a source of health information is greatly increasing. However, identifying relevant and valid information can be problematic. This paper firstly analyses the efficiency of Internet search engines specialized in health in order to then determine the quality of the online information related to a specific medical subdomain like that of neuromuscular diseases. Our aim is to present a model for the development and use of a bilingual electronic corpus (*MYOCOR*), related to the said neuromuscular diseases in order to: a) on one hand, provide a quality health information tool for health professionals, patients and relatives, as well as for translators and writers of specialized texts, and software developers, and b) on the other hand, use the same as a base for the implementation of a search engine (using keywords and semantics), like the ASEM (*Federación Española Contra las Enfermedades Neuromusculares*) search engine for neuromuscular diseases.

MOTS-CLÉS/KEYWORDS

corpus électronique, traduction de textes médicaux, maladies neuromusculaires, français, espagnol
electronic corpus, translation of medical texts, neuromuscular diseases, French, Spanish

1. Introduction

The use of the Internet as a source of health information by specialists as well as users or patients is greatly increasing. Millions of people around the world search for health information on the Net regarding diseases and their treatments, drugs and different prevention and diagnostic measures or even medical advice through virtual consultation, chat or e-mail. The quality of this information is extraordinarily variable; from scientific information based on evidence to home-made remedies, sometimes of doubtful origin, which can even be harmful or dangerous. The concerns of government, professional and scientific societies and institutions, as well as of the users themselves are growing and require some kind of intervention.

Thus, for example, according to a recent study by Bennett, Casebeer *et al.* (2005), in which they compare the use that the primary and specialized health care doctors make of the Internet in the United States, both consider that the Net is a good source of knowledge and they feel capable of accessing the information through this technology. The main difference between both groups lies in the fact that family doctors access Internet mainly to increase their knowledge on diseases they deal with in their everyday practice, whereas their hospital counterparts are more prone to read scientific journals online.

In this way, just as websites with biomedical information have vastly proliferated and are one of the most visited by Internet users, the initiatives to guarantee and improve the quality of the information that is offered by these sites have followed a parallel path, which led to an important number of instruments for control, surveillance, and assessment of the biomedical contents available online.

The main objective of the majority of these instruments that try to certify biomedical information has been to guarantee the Internet user that a given website complies with a series of quality requirements, mainly related to offering transparent information about the authors, sponsors, and the existence of conflicts of interest. The proposed initiatives are fundamentally oriented towards the accreditation, certification, self-regulation, qualification systems, and the concession of quality seals, which indicate the fulfilment of specific criteria (Ávila, Portillo *et al.* 2001).

There are important studies on the quality of the biomedical information that have been published and are available on the Net in English and in Spanish, but it is difficult to find works regarding the Hispanic initiatives that watch over the quality of this information through accreditation-certification methods. Amongst them, the study by García, Montesinos *et al.* (2004), is worth mentioning. It shows the experience of each of the projects that were found, comparing the methods that were applied and thus assessing the coexistence on the Net of the quality seals granted by each initiative.

But the users, mainly the patients, hardly know the accreditation systems of the websites and it is more difficult for them to understand the value they may have. Therefore, we have to reach the conclusion that the assessment of the quality of the biomedical information available on the Net should not be restricted to finding out whether websites comply with a series of static criteria. We should know to what extent the quality of this information has an influence on users, both professionals and the general public.

The proposal to create a platform, gathering the experience of all of the initiatives working on the quality of biomedical information on the Net and involving informa-

tion producers and their final users, may improve the information given to citizens, both patients and professionals and thus avoid their confusion, while also improving the counselling regarding the biomedical information available on the Internet, preventing low-quality information from reaching the users. The European Union is currently involved in this task, studying the development of a joint action among the member states in order to establish a recognisable quality-seal system for websites featuring health information. It would be desirable to achieve a multilingual platform that unifies experiences. This is the joint work of the *MedCircle*[1] (2002) and *HON*[2] (2007) projects.

What is beyond any doubt or interpretation is the fact that, in roughly a decade since its appearance, Internet has become an instrument used by our society on a daily basis, comparable to other important media such as radio, television or the telephone, and actually surpassing them in many aspects. In parallel fashion to the spectacular growth of the Internet, the accompanying technologies have experienced a fast evolution, allowing a better, wider, more powerful, more flexible and more readily understandable Web. These changes at once influence and are influenced by the transformation of the so-called World Wide Web. The dynamic creation of sites, the combination of databases, the greater interactivity with the user, the conception of the Net as a universal platform for displaying applications and the adaptation to the user are just some evolutionary trends in the last few years.

In the conclusions and recommendations of Berland, Elliott, *et al.*'s (2001) report regarding the evaluation of English and Spanish health-related information on the Internet, they affirmed that the two most common ways that users find health information on the Internet are through the use of search engines that lead to health-related websites and direct visits to health-related websites. They examined what users are likely to find when using these two methods to search for health-related information. Specifically they asked: 1) How easy is it to find relevant information on important health conditions using search engines? 2) What type of information do search engines turn up, and to what extent does this vary by search engine? 3) With regard to health-related websites, how comprehensive, accurate, and readable is the information they provide? 4) Is health-related information as readily available in Spanish as it is in English?

The results of the analysis are presented, conclusions drawn from them, and pertinent recommendations given with regard to the different aspects of health-related information on the Internet. These are issues that should be considered by the consumers and their advocates, health care providers, government policymakers and regulators, and health information providers, such as search engines and websites. Without overly extrapolating, the main conclusions of the study show that the tremendous amount of information available to consumers is clearly one of the major attractions of the Internet for obtaining health information, but it also means that consumers must sift through a large amount of material during their searches.

This study found that search engines are only moderately efficient in locating information on a particular health topic and the efficiency of finding relevant information varies significantly across search engines and conditions. In the case of the simple search strategies used in this study, the search engines were only moderately efficient in finding relevant information related to specific health topics. Only one in five links identified by 10 English-language search engines and one in eight links

from 4 Spanish-language search engines led to webpages with pertinent content. More than half of consumers who used the Internet reported using search engines to find health information, and they spent about half an hour on such searches, so efficiency and relevance of information retrieved are important aspects of performance.

The company *JupiterResearch*[3] (2006), a leading authority on the impact of the Internet and emerging consumer technologies on business, claims that 71 percent of online consumers use search engines to find health-related information. However, only 16 percent find the information they are looking for, thus highlighting the need for new strategies for general and specialized search engines to attract and retain the loyal user base. According to Levy, senior analyst of the mentioned company and author of the report we are referring to:

> Despite strong demand for health information, most online consumers' search experiences are negative. […] The combination of high demand and poor experience means there is a significant opportunity for better engines and products in the market. (Levy 2007)

The *JupiterResearch* report shows that sixty percent of online consumers use general search engines with *Google* leading the pack by a significant margin. Additionally, 42 percent of online consumers already use various health search engines such as *WebMD*, *AOL Health* and *MSN Health & Fitness*. As far as Schatsky, president of *JupiterKagan*,[3] is concerned, he considered that:

> Search engines must work toward striking the right balance between search efficiency, quality of results and proprietary feature sets. […] Online consumers are interested in features that improve and facilitate their searches as long as they don't add an unnecessary layer of complexity. (*JupiterResearch* 2006)

The study from Berland, Elliott *et al.* (2001), which we referred to, shows that although advertising and other non-explicit promotional material were common, it was beyond the scope of the study to evaluate whether or not consumers had difficulties recognizing commercial or promotional health information. Furthermore, when the top ten websites listed on each of the search engines were reviewed, the degree of overlap was quite low (generally 11 percent only). This level of variability across search engines and conditions suggests that the likelihood of finding the information one needs varies considerably depending on which search engine is used. No search engine is clearly better than another, but where users start matters.

The study has some important limitations that are worth noting to be able to interpret the final results, for example in Chapter 2, *The Study of Search Engine Performance*:

1) First, the Internet changes constantly, and we are only able to study it at one point in time. However, without concerted attention to the issue, it seems unlikely that the variability in performance is likely to change dramatically;
2) Second, the study looked at a small sample of search engines and conditions, and hence we cannot draw more general conclusions about the performance of all search engines and information on all conditions. However, because the study included the most popular search engines, the results should reflect what most people experience;
3) Third, the study shows the performance of search engines using very simple search terms describing the medical condition. The findings regarding the efficiency of

> search engines in yielding relevant content might have been quite different if more sophisticated search strategies were used;
>
> 4) Fourth, the research conducted in the study was not a naturalistic experiment (e.g., using actual consumers to search for information and testing their knowledge after such a search) so we cannot draw conclusions about what consumers actually encounter when they search for information, or about how well they are able to judge the quality of the information they find. However, the systematic nature of their methods provides a backdrop for future studies of actual consumer behaviour (we can compare what consumers are able to find with what is actually out there to find). (Berland, Elliott, *et al.* 2001)

In order to implement our documentary search engine on neuromuscular diseases we will consider all the aspects related with the performance of the search engines. This way, for example, we will limit the search threshold from our search engine to a limited, previously selected, structured group and of contrasted quality of documents of the environment in question. This will assure ahead of time the efficiency and the relevance of the recovered information. We will also consider the features that improve and facilitate the users' searches, as long as they don't add an unnecessary layer of complexity.

Amongst the latest trends that could have an impact on the medium-term future of the Net we have the *semantic web*, whose aim is to make machines understand and therefore use what the Net contains. This new Net would be inhabited by software agents or representatives able to surf and perform operations for us in order to save us work and optimize the results. In order to achieve this goal, the semantic web suggests describing the Net resources as representations that may be processed (that is, understood) not only by people, but also by software applications that may assist, represent or replace people in routine tasks or simply not viable for a human being. The technologies of the semantic web seek to develop a more cohesive Net, where it would be easier to locate, share and integrate information and services in order to profit even more from the resources available on the Net.

This conception of semantic web corresponds to a new vision of the concept of health introduced by the World Health Organization (WHO)[4] in 2001, based on the integration of two antagonist models: *the medical model* and the *social model* (WHO 2001). The former corresponds to a more traditional vision and conceives health as an individual problem, directly related to a disease requiring specific medical treatment. However, the social model is inscribed in a wider context, which necessarily refers to situations that are, at least partially, created by the social environment and therefore affect the whole society. This new approach emphasizes the importance of considering the implication of health in all social debate forums, which has led to an important change in the projection traditionally attributed to health.

2. The medical subdomain of neuromuscular diseases

Neuromuscular diseases are genetic conditions, which are normally inheritable and affect muscles and the nervous system (they are also known as myopathies). There are hundreds of these diseases duly classified according to their degrees of severity. Their onset may occur at birth or during other stages in life. Their most important characteristic is the progressive loss of muscular strength and the degeneration of the muscles and of the nerves controlling them.

While the total number of affected people is quite large, the great diversity of genetic expression and location of these conditions means that the many well-individualized diseases that have been defined are included in the sphere of so-called *rare diseases*. This situation has had very negative consequences for patients (diagnostic delays, adaptation problems, etc.), which can now be overcome through intervention from different spheres.

The *Spanish Federation of Neuromuscular Diseases* (ASEM-Spain; *Federación Española Contra las Enfermedades Neuromusculares*)[5] is a Spanish federation of associations that provide support and guidance to the associations of people suffering from neuromuscular diseases. This same support is also offered to the people affected and their families. The strategic lines of the ASEM Federation may be classified as:

a) bringing together all the associations of neuromuscular diseases existing in Spain in order to combine forces that will allow them to achieve objectives that could scarcely be achieved on an individual basis;
b) coordinating the activities of the ASEM Federation with those of all the services and institutions, both public and private, which are directly or indirectly involved in the vital process of assisting handicapped people, and finally;
c) disseminating information on neuromuscular diseases, which includes publications (of books, brochures, triptychs, journals…) on clinical, psychological, therapeutic social or general information topics addressed to those affected as well as the professionals working with them.

In accordance with the social health model promoted by the WHO and in line with the strategies of the ASEM Federation in Spain, the *Galician Association against Neuromuscular Diseases* (ASEM-Galicia),[6] which has been working in the field of neuromuscular diseases for more than 10 years, considered that it is essential to create a series of documentary resources, which could be useful for all the people involved in this issue (patients, family, doctors, health professionals), as well as the society in general.

2.1. The project to create a multilingual textual corpus about neuromuscular diseases and the elaboration of a documentary search engine

In 2004, ASEM-Galicia requested the external collaboration of a team of researchers from the University of Vigo and Meixoeiro Hospital, respectively made up of experts in translation and information science and in the treatment and research of neuromuscular diseases. The aim was to bring together the knowledge of the different teams involved in order to create a knowledge database, which would become a reference in the mentioned field and could be of use and interest for all those involved directly (patients and close environment) or indirectly (health professionals, professionals and researchers from other areas, like laboratories, translation, etc., or even the public administration).

The mentioned project's aim was to give a response to all these needs by means of creating and using a series of documentary resources on neuromuscular diseases, one of the priority fields in the health sphere nowadays. This project would allow the consolidation of a necessary collaboration with researchers from the University of Vigo and medical specialists from Meixoeiro Hospital, which despite having occurred occasionally and uninterestedly for years, needed to be intensified and consolidated

at this moment. Thus a link would be established between the research activities and the social reality in which research is imbedded. That would definitely foster the convergence that has been claimed so repeatedly by different social agents due to their interest in improving the quality of life of society as a whole.

The final aim of the project was the creation of a bilingual corpus (French and Spanish) with texts on neuromuscular diseases and a search engine, which could directly browse the corpus both in Spanish and French. First of all, the system would have to be tested in a restricted experimentation environment and require the opinion of the professionals; and later on, if the results were satisfactory, the tool would be made available online for the whole community.

2.1.1. The MYOCOR bilingual textual corpus on neuromuscular diseases

The creation of a bilingual corpus (French and Spanish) with texts on neuromuscular diseases (*MYOCOR*) is clearly connected to its obvious antecedent, which is the work done by the *Association Française contre les Myopathies* (AFM).[7] This association was created in France in 1958 and enjoys wide social recognition (it was declared of public interest in 1976), comparable to the *ONCE* (an organization mainly working for the visually handicapped amongst other disabled individuals) in Spain, and has a large budget, allowing it to totally or partially finance a series of highly prestigious research centres in the scientific world, such as: *Généthon, Institut de Myologie, Génopole, Plateforme Maladies Rares* or *European Neuromuscular Centre* (ENMC). Both AFM and the aforementioned centres produce an important number of publications on neuromuscular diseases every year, which are of high interest, apart from carrying out other actions related to research and the ongoing training of experts.

These publications are characterized by being texts produced by different types of individuals and institutions, involved in the field of neuromuscular diseases. They include texts written by doctors or health professionals with different degrees of specialization (biologists, nurses, physiotherapists, etc.), by researchers and by different institutions. On the other hand, the addressees of these texts are also widely diverse, since they not only include the aforementioned professionals and patients but also society as a whole.

ASEM-Galicia has signed an agreement thanks to which all these texts have been made available for the association so that they can be used in benefit of the patients and with non-profit making aims.

By taking texts produced by the AFM (in French) and then translated into Spanish, texts originally written in Spanish or translations into Spanish of English texts as a starting point, the research project seeks to create a documentary resources database in Spanish which will be uploaded in a specialized website on neuromuscular diseases available for the entire Galician society and state, as well as those interested in other countries (such as the Latin American patient associations, which have already shown interest in the project). The theoretical foundations for this project are those proposed by the *Corpus Translation Studies* (CTS – Sánchez-Trigo [2003, 2005, 2006]). It is a recent piece of research, which was started in the 90's and is currently contributing interesting results, as shown by the projects conducted by different researchers, as well as the recent corpora and projects (both national and international), specialized in different medical fields, such as *LIQUID* (2001)[8] or *ONCOTERM* (2002),[9] amongst many others.

2.1.2. Creation of a documentary reference tool for MYOCOR

2.1.2.1. Theoretical models for information search

Many models of text search engines available in websites depend on matching a string of exact words chosen by the user with those in the websites. Two antithetical situations are found: on one hand, the search engine finds homonymous terms and, on the other hand, synonyms are ignored. In the first case, we find sites whose topics are not of our interest, whereas in the second one, some sites of our interest are ignored. These two distorting factors should be combined with an additional one: multilingualism. This happens when we search for something using the search engine and we type terms in a given language and find sites related to our topic of interest, but written in a different language to that used for the terms.

In all these cases, the problem lies in the fact that our query does not precisely identify what we are interested in, but only identifies a term, which, in a given language, has, among its meanings, what we are looking for. This search procedure in which the results obtained are a series of documents containing the terms typed in the query is known as *syntactic search*. In order to go beyond this syntactic (or linear) search, many authors have expressed their points of view, coming up with multiple solutions (some of which are more theoretical than others) from different research fields.

Some authors (Gracia 2002) suggest an approach to Internet search engines by means of linear algebra. Since computers do not know anything about languages, the solution to the problem lies in the possibility of providing them with a mathematical model, which could *learn* the meaning of words in the semantic universe made up by all the existing websites at a given moment worldwide. It is possible to achieve this learning by means of the *latent semantic index* (LSI), which would unveil the hidden meaning of the words in websites. This latent semantic index would make a software application learn what is a given concept or entity (for example, a table) from website texts saying something about them. This analysis could be extended to the recognition and matching of images, scenes, sounds (very important aspects in the world of medicine); in other words, photography, video and audio. They formulate a series of *mathematical ideas* that could be taken to these fields. Nevertheless, it also warns us about the fact that, in such cases, difficulties would be even higher since, for example, when two pictures are almost the same or when two sound clips are virtually identical, they would have to be defined algorithmically. These applications may require the use of spectral signal processing, sampling, etc.

Other authors (Sánchez Fernández and Fernández García 2005), however, have come up with a different type of solution. If websites had associated formal annotations unmistakably identifying the concepts and main entities included in the websites, a search engine (semantic) could avoid making the three aforementioned types of mistakes and just find the documents that are exactly related to what we were actually looking for. Nevertheless, as they acknowledge, having semantic annotations may not be enough for the applications that use the semantic web. Most of these applications require a domain model to operate in including the vocabulary of the relevant concepts for this domain and probably the properties that relate the different concepts as well as the rules governing that domain. Based on that model, the system would be capable of obtaining conclusions and/or make decisions by processing the

annotations extracted from the websites. Such models are defined by means of *ontologies*.

Thus, a way of understanding a semantic search engine would be as a tool processing queries based on ontologies, which are conducted using a knowledge base and then the system would return suitable results. These techniques typically use boolean search models and are based on an *ideal* vision of the information space, consisting of formal pieces of ontological knowledge without ambiguity or redundancy. However, in practice it is well known that there are limits to the extent to which knowledge can be formalized using ontologies (Vallet Weadon, Fernández Sánchez, *et al.* 2005):

1) First of all, due to the huge volume of unstructured information available nowadays as texts and multimedia contents, turning all this amount of information into ontological knowledge with a feasible cost is a generally unsolved problem.
2) Secondly, documents have a value *per se* and they are not equivalent to the sum of their parts. Although it is useful to divide documents into smaller information units, which may be reused and assembled for different purposes, it is frequently appropriate to maintain the original documents in the system.
3) Thirdly and lastly, if no clear grading system is included, the search system could become useless if the search space is too big.

2.1.2.2. The search for information on neuromuscular diseases

In order to implement our documentary search engine on neuromuscular diseases, the extent to which knowledge can be formalized using ontologies is less problematic. On one hand, working with an initially limited volume of texts makes it much more feasible to convert the said information into ontological knowledge, structuring text documents and multimedia contents (mainly images). However, as new texts are included, the problem will become more and more evident. On the other hand, in order to respect the principle whereby documents have value *per se* and are not equivalent to the sum of their parts, we will always work with two copies of the same original document. One of the copies will remain *unalterable* (in PDF format), whereas the other copy will be conveniently structured (in XML format). Thirdly, and in order to restrict the size and volume of queries, we will take *the name of neuromuscular diseases* as a starting point. The number of neuromuscular diseases, although high (we are talking about roughly 150-200 known diseases), is limited *per se*.

Thus, from a practical point of view, it seems feasible to create a semantic search engine on neuromuscular diseases. Nevertheless, as Mayfield and Finin (2003) indicated: "it is better to use semantic search as a complement for keyword search, as long as not enough ontologies and metadata are available." Therefore, and due to the time and financial constraints of our research project and ASEM-Galicia's need for imminent results, we decided to start with the implementation of a search engine on neuromuscular diseases using keywords, so that on the second stage of the project, we could proceed in creating the ontologies and metadata and implement them in a semantic search engine on neuromuscular diseases.

Another determining factor in our research work has to do with the *social model* of health that we mentioned before and with the ultimate aim of our project. This *socialization* of the environment should also be reflected in the computer tools used. This is the reason why we decided to work with the open code software. Many

different factors encouraged us to do so, but they could be *outlined in the conclusions* of the Wheeler Report.

The open code has a significant market share, it may be the most reliable and in many cases, the best performing. The open code escalates both in the size of the problem and the size of the project. The open code generally has better security, particularly when compared to Windows. The total ownership cost for the open code is normally less than for the closed code, especially as the number of platforms grows. These claims are not merely opinions. These assertions can be proven quantitatively, using a wide range of means. This does not even consider other aspects that are difficult to measure, such as freedom vs. control by a single source, freedom *versus* licence management with the corresponding lawsuits and the growing flexibility. We believe that open options should be carefully considered whenever computer hardware or software is needed (Wheeler 2005).

Finally, we always kept in mind that neuromuscular diseases are associated with individuals and social environment. Guaranteeing and improving the quality of the information offered on our website was always our main and ultimate aim. But we also wanted to know to what extent the quality of this information could influence the users, both professionals and the general public. Therefore the information available to users at the end of their search is shown in its *complete version* (original) to provide a response to professionals and researchers, whereas for those with more specific needs (for example, pinpointing the causes, symptoms, treatment, etc of a given disease) there is *selection of the information shown*. We achieved this by working with two versions of the original text; whereas one is unaltered and retains all its graphic elements, the other is structured (following a common *pattern* for all texts) and conveniently processed (using metadata) for the search engine to easily access the specific sub-information. Before putting the tool online, accessible for everyone, it would also be necessary to establish a *test period* to evaluate its efficiency (in human and computational terms) and the quality of the information shown.

3. Conclusions

The quality of medical information available on the Net is extremely variable, ranging from scientific and evidence-based information to homemade remedies or remedies of dubious origin, which may even be harmful or dangerous. The concerns of governments, professional and scientific institutions and associations, as well as users, are growing and require some kind of intervention. The initiatives proposed are mainly focused on accreditation, certification, self-regulation, qualification systems and quality seals indicating that specific criteria are met. But users, mainly patients, hardly know all the accreditation systems of the websites and it is even more difficult for them to understand their validity.

This global assessment also applies to the medical subdomain of neuromuscular diseases, but we should not forget that these conditions are chronic, generally progressive and of variable severity and are included in the field of so-called *rare diseases*. The varied consequences of these conditions suggest a convergence of many medical specialities in this field (such as neurology, myology, cardiology, etc). Furthermore, it is a field that relates to many other areas of health in the broadest sense (such as physiotherapy, orthopaedics, psychology, nutrition, aspects related to the quality of life of patients, etc.).

Thus, in this first stage of our project, our proposal focuses on creating a platform, gathering the experience from all the initiatives working (directly or indirectly) in the said medical environment and involving information producers and final users. We believe that by doing so we can improve the information available to citizens, patients and professionals, reduce confusion, while also improving their counselling regarding medical contents and preventing low-quality information from reaching the users. Apart from the creation of the ASEM website (ASEM-Galicia), specialized in processing all types of information related to neuromuscular diseases, our actual theoretical model is based, first of all, on the creation of a bilingual electronic text corpus (Spanish-French), from a reduced volume of high quality texts related to neuromuscular diseases, created by experts in French and in Spanish (the French texts were translated into Spanish by professional translators specialized in medical translation).

Secondly, we proceeded to create a documentary search engine so that information could be extracted from the said corpus. From a practical point of view we realize that it seems feasible to create a semantic search engine on neuromuscular diseases. However, due to the time constraints of our research project and ASEM-Galicia's need for imminent results, we decided to begin with the implementation of a search engine using keywords. The ASEM search engine for neuromuscular diseases (*MYOCOR*), in its syntactic or keyword search version, has already been implemented. It is currently being reviewed and the contents are being updated (texts). It can be directly accessed through a web link, or indirectly, through the buscador myocor web link of the ASEM-Galicia website.[10]

In a second stage of the project, which is still being studied and pending on financing, our intention is to create the necessary ontologies and metadata in order to implement them in a semantic search engine on neuromuscular diseases. In order to carry out this process, once the keyword search engine has been implemented, the next step would be to assess the information and services that the current ASEM-Galicia web server offers to health professionals, patients and relatives, as well as translators and writers of specialized texts and software developers. Last but not the least, we should see how the semantic web fits in the current web, in other words, how the user will access the semantic web and, above all, how to undertake the transition from the current web (syntactic or linear) to the semantic web. Possibly, we should consider combining a small part of the manual work involved with the automation of the rest of the process. It would be desirable to at least automate the terminological component by applying an existing terminological database management system (knowledge-based).

The following have participated as external collaborators: the University of Vigo (research team led by Elena Sánchez Trigo, Professor of Translation and Interpreting, including Joan Miquel Vergés, together with members from the Computer department) and the Meixoeiro Hospital (research team led by Carmen Navarro Fernández-Balbuena, Chief of the Pathological Anatomy and Neuropathology Department). Alexandre Dapena Mora, Computer engineer, took charge of developing the computer application.

NOTES

1. MedCircle (17 December 2002): *Medical Collaboration for Internet Rating, Certification, Labeling and Evaluation of Health Information*. Visited 23 February 2007, <http://www.medcircle.org>.
2. HON (24 January 2007): *Health On the Net Foundation*. Visited 23 February 2007, <http://www.hon.ch>.
3. In 2006, *JupiterResearch* and *Kagan Research LLC* merged into JupiterKagan Inc.; and, finally, in 2008, JupiterResearch was acquired by *Forrester Research*. Visited 23 February 2007, <http://www.jupiterresearch.com>.
4. WHO (2001): *World Health Organization*. Visited 23 February 2007, <http://www3.who.int/icf/intros/ICF-Eng-Intro.pdf>.
5. ASEM-España, *Federación Española de Enfermedades Neuromusculares* (2007): Visited 23 February 2007, <http://www.asem-esp.org>.
6. ASEM-Galicia, *Asociación Gallega contra las Enfermedades Neuromusculares* (2007): Visited 23 February 2007, <http://www.asemgalicia.com>.
7. AFM, *Association Française contre les Myopathies* (2007): Visited 23 February 2007, <http://www.afm-france.org>.
8. LIQUID (2001): *Language Independent Querying for Information Discovery*. Visited 23 February 2007, <http://www.cervantes.es/seg_nivel/lect_ens/oesi/liquid02/ingles/index.htm>.
9. ONCOTERM (9 March 2002): *Sistema Bilingüe de Información y Recursos Oncológicos*. Visited 23 February 2007, <http://www.ugr.es/~oncoterm/>.
10. *The ASEM search engine for neuromuscular diseases* (2007): Visited 23 February 2007, <http://213.60.169.224/ASEM/> (direct access) and <http://www.asemgalicia.com/myocor/myocor.html> (indirect access).

REFERENCES

Ávila, José F., Portillo, Blanca E., and Pajares, Juan. M. (2001): Calidad en la información biomédica existente en Internet. *Aten Primaria*. 28(10):674-679.

Bennett, Nancy L., Casebeer, Linda L., Kristofco, Robert, *et al.* (2005): Family physicians' information seeking behaviors: A survey comparison with other specialties. *BMC Medical Informatics and Decision Making*. 5(9). Visited 23 February 2007, <http://www.biomed central.com/content/pdf/1472-6947-5-9.pdf>.

Berland, Gretchen K., Elliott, Marc N., Morales, Leo S., *et al.* (2001): Health Information on the Internet: Accessibility, Quality, and Readability in English and Spanish. *Journal of the American Medical Association*. 285(20):2612-2621. Visited 23 February 2007, <http://jama.ama-assn.org/cgi/reprint/285/20/2612.pdf>.

García-Vicente, Sergio, Montesinos, Erik, and Boyer, Célia (2004): Iniciativas en Español para la Evaluación de Contenidos Biomédicos en Internet. *In*: *Informédica 2004. Information & Communication Technologies in Healthcare Development. 3rd Virtual Congress in Internet* (online congress, 1-30 March 2004).

Gracia, Juan-Miguel (2002): *Álgebra lineal tras los buscadores de Internet*. Visited 23 February 2007, <http://www.vc.ehu.es/campus/centros/farmacia/deptos-f/depme/profesor/gracia/buscap.pdf>.

Levy, Monique (2007): Health Search: Assessing Consumers Demand for Health Vertical Search Engines. *Jupiter Research Industry Focus Health*.

Mayfield, James and Finin, Tim (2003): Information retrieval on the Semantic Web: Integrating inference and retrieval. *In: Workshop on the Semantic Web at the 26th International ACM*

SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003, Toronto, 1 August 2003). Visited 23 February 2007, <http://www.cs.umbc.edu/~finin//papers/sigir03.pdf>.

Sánchez Fernández, Luis and Fernández García, Norberto (2005): La Web Semántica: fundamentos y breve "estado del arte." *Novática, 178* (November-December), 6-11. Visited 23 February 2007, <http://www.ati.es/novatica/2005/178/178-6.pdf>.

Sánchez-Trigo, Elena (2003): Aproximación traductológica la análisis de corpus para el estudio de las convenciones textuales. *In:* Emilio Ortega Arjonilla, ed. *Panorama actual de la investigación sobre traducción e interpretación*. Granada: Atrio, 113-127.

Sánchez-Trigo, Elena (2005): Creación y explotación de recursos documentales sobre enfermedades neuromusculares." *In: Actas de ponencias del XXII Congreso nacional de ASEM sobre enfermedades neuromusculares* (Vigo, 7-8 October 2005). Vigo: ASEM, 85-94. Visited February 23rd 2007, <http://www.asemgalicia.com/biblioteca/congresos_ponencias/XXII_Congreso_ASEM_2005_Libro_de_PONENCIAS-1.pdf>.

Sánchez-Trigo, Elena (2006): MYOCOR: creación y explotación de un corpus bilingüe (fr-es) sobre enfermedades neuromusculares. *Confluências*. 4(May 2006):67-83. Visited 15 October 2009, <http://confluencias.net/n4/sanchez-trigo.pdf >.

Vallet Weadon, David, Fernández Sánchez, Miriam, and Castells Azpilicueta, Pablo (2005): Recuperación de información en la Web Semántica. *Novática*. 178, (November-December):12-15. Visited 23 February 2007, <http://www.ati.es/novatica/2005/178/178-12.pdf>.

Wheeler, David A. (2005): *Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers!* Visited 23 February 2007, <http://www.dwheeler.com/oss_fs_why.html>. Spanish version: *¿Por qué usar programas abiertos? ¡Atención a las cifras!* Visited 23 February 2007, <http://www.hispalinux.es/informes/wheeler/index.html>.