FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

HealthTranslator: automatic annotation of Web documents in order to assist health consumer's searches

Hugo Sousa

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Carla Teixeira Lopes

July 20, 2016

HealthTranslator: automatic annotation of Web documents in order to assist health consumer's searches

Hugo Sousa

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: Sérgio Sobral Nunes

External Examiner: José Luis Guimarães Oliveira

Supervisor: Carla Alexandra Teixeira Lopes

July 20, 2016

Abstract

The Web is now one of the main sources of health related information. However, health consumers do not always easily understand the retrieved information, mainly because of a significant gap between terminologies used by laypeople and medical experts. This work presents a tool, HealthTranslator, available as a Google Chrome extension, that helps users to overcome the difficulties they face when reading health related Web documents. HealthTranslator provides automatic annotation of medical concepts in Web documents with additional information, such as concept definition, related concepts or linkage to external references. It recognizes concepts in Portuguese and English Web pages and is highly customizable by the user.

The solution was evaluated in terms of performance, annotation coverage and quality and utility perceived by the users. With respect to the performance, however significantly worse than another extension with some similar features, mainly because of its client-server architecture, users find the processing time to be acceptable as it is done gradually, typically from the top to the bottom of the document. Regarding the annotation coverage in English, the solution was compared with the previously referred similar extension. HealthTranslator is based on a much larger vocabulary which covers around 74% of the concepts of that tool. A comparison with a corpus of 38 documents with manually annotated medical concepts was also performed, showing an average F-measure between 27% and 33%, depending on the types of concepts to be recognized. Although relatively low values are expected given the broad range of medical domain, it is believed these values could be eventually higher due to the annotation comparison strategy. In both languages, the majority of the recognized concepts have a related definition, around 72% for Portuguese and 80% for English. Concerning the utility, many aspects were surveyed on a user study. In general, the extension has a good acceptance and the users find it useful although improvement suggestions were made.

Resumo

A Web é agora uma das principais fontes de informação relativa a saúde. No entanto, consumidores de saúde nem sempre compreendem facilmente a informação obtida, principalmente devido a uma discrepância significante nas terminologias usadas por leigos e especialistas de saúde. Este trabalho apresenta uma ferramenta, o HealthTranslator, disponível como uma extensão para o Google Chrome, que ajuda os utilizadores a ultrapassar as dificuldades com que se deparam ao ler documentos na Web relacionados com saúde. O HealthTranslator efetua uma anotação automática de conceitos médicos em documentos Web e apresenta informação adicional ao utilizador: definição do conceito, conceitos relacionados ou ligações a recursos externos. Reconhece conceitos em páginas Web em Português e Inglês e é bastante customizável pelo utilizador.

A solução foi avaliada relativamente ao desempenho, capacidade e qualidade de anotação e a utilidade reconhecida pelos utilizadores. Relativamente ao desempenho, apesar de ser significativamente pior que uma outra extensão com algumas funcionalidades semelhantes, principalmente devido à sua arquitetura cliente-servidor, os utilizadores acham o tempo de processamento aceitável por ser feito gradualmente, tipicamente do topo para o final do documento. Quanto à cobertura de conceitos em Inglês, a solução foi comparada com a extensão semelhante previamente referida. O HealthTranslator baseia-se num vocabulário consideravelmente mais extenso que cobre cerca de 74% dos conceitos dessa ferramenta. Foi também efetuada uma comparação com um corpus de 38 documentos com conceitos médicos manualmente anotados, mostrando um F-measure médio entre 27% e 33%, dependendo dos tipos de conceitos a serem reconhecidos. Apesar de valores relativamente baixos serem expectáveis, acredita-se que esses valores poderiam ser eventualmente maiores devido à estratégia de comparação das anotações. Em ambas as linguagens, a maioria dos conceitos reconhecidos apresentam uma definição, cerca de 72% em Português e 80% em Inglês. Em relação à utilidade, vários aspetos foram avaliados num estudo de utilizador. Em geral, a extensão tem uma boa aceitação e os utilizadores acham-na útil, tendo sido feitas sugestões de melhoria.

Acknowledgements

I would like to express my gratitude with all those that contributed in several ways for the completion of this dissertation.

To the supervisor, Professor Carla Teixeira Lopes, for the continuous support and guidance along this path. For her sympathy and promptitude to help in every situation.

To Dr. Dagmara Paiva for her suggestions and help regarding the Portuguese evaluation of the tool, including the invitation of medical professionals to volunteer towards the testing of this project.

To all those who volunteered to help with the application testing. To the medical professionals and friends who annotated documents and to all of those who did the user study and provided useful feedback.

To my friends from FEUP, Marie Ros, Daniel Moreira and Francisco Maciel for all the discussions during lunches, coffee-breaks and relaxing moments.

Last but not least, to my family for supporting me and all my choices during all these years.

Hugo Sousa

Contents

CONTENTS

List of Figures

List of Tables

LIST OF TABLES

Abbreviations

CHV	Consumer Health Vocabulary
CRF	Conditional Random Fields
EHR	Electronic Health Record
NER	Named Entity Recognition
NLP	Natural Language Processing
OAC CHV	Open Access Collaborative Consumer Health Vocabulary
OBA	Open Biomedical Annotator
PHR	Personal Health Record
SVM	Support Vector Machines
UMLS	Unified Medical Language System
CUI	Concept Unique Identifier
API	Application Programming Interface
HTML	HyperText Markup Language
REST	Representational State Transfer
JSON	JavaScript Object Notation
DOM	Document Object Model

Chapter 1

Introduction

Not so long ago, patients would need to directly contact with healthcare providers in order to access health information. This brings the benefit of obtaining accurate knowledge; however, it constraints the accessibility to their needs. The rise of digital technologies transformed our everyday life and revolutionized the way we search for information, including healthcare.

While this may sound ideal in a superficial perspective, there still are many problems and challenges to face.

1.1 Context

'Health consumer' or 'Healthcare consumer', a widely used term in the related literature, although not presenting a clear and unanimous definition in the scientific community, refers to "people who use, or are potential users, of health services including their family and carers", according to Health Consumers Queensland [Que12].

Consumer health informatics, according to according to U.S. General Accounting Office, is "the use of modern computers and telecommunications to support consumers in obtaining information, analyzing their unique health care needs and helping them make decisions about their own health", although several definitions exist [LCF05].

There is a clear trend of a rise of online search of health information, as it is shown in an analysis from 1998 to 2011 in U.S. [Pol11]. A survey conducted in 2013 shows that 72% of internet users in U.S. looked online for health information within the previous year [FD13]. A similar survey carried in E.U in 2014 says that 59% of the respondents have searched for this kind of information in the previous year [Com14].

It is also noted a pattern in how consumers search for information in this domain. The survey realized in U.S. shows that 77% of health consumers started the search in a generalist search engine like Google or Bing, while only 13% began in a specialized site in health information [FD13]. In the European survey, 82% to 87% of the health consumers also commenced their exploration in a search engine [Com14].

1.2 Motivation and Goals

Health literacy "entails people's knowledge, motivation and competences to access, understand, appraise and apply health information in order to make judgements and take decisions in everyday life" [Org13]. This concept has been subject of significant interest in the last years, in order to understand how to address the problems of low health literacy in the population [VdB14].

Although most health consumers find their searches successful [Pol11], because of their perception that the web provides access to reliable information, research studies show that there are many issues regarding this type of search. Results indicate significant problems in query formulation [TL07] and poor consumer's information retrieval performance [ZKA+02]. This mainly occurs because there is a significant mismatch between the terminology used by health consumers and medical professionals [ZKA+02]. For example, a layperson would say 'heart attack', while the technical term is 'myocardial infarction'.

Another barrier to an efficient search of information on the Web is the language barrier. The amount of information in the Web is correlated with the number of native speakers [KM06]. Hence, people not proficient in an widely used language, such as English, have a more limited access to information, leading to a worse search performance.

The characteristics of the information on the Web lead to the challenge of bringing the right information at the right person and at the right time [Gei08], while at the same time keeping it intuitive.

Thus, the goal of this dissertation is to develop and evaluate an application that helps the layperson while searching or accessing health-related information on the Web, contributing to reduce the discrepancy between consumer and professional terminologies contributing to a more successful outcome.

More precisely, it presents the development and evaluation of an extension for Google Chrome. Statistics show that Google Chrome is currently the most widely used web browser [Sta] with currently 57.88% of market share worldwide. These statistics are calculated on a basis of more than 15 billion page views per month, so the results seem reliable. This application recognizes medical concepts and provides useful information to the user for a better search. It presents a definition of the concept, and can be complemented with links to external resources for a more encompassing search. This system should be modular to facilitate the adaptation to different languages. It should be user-friendly and should not interrupt the regular usage of the browser by the user. In order to overpass the language barrier and as a proof of concept, it initially supports English and Portuguese. The extension initially supports two languages: English, as it is the most widely used language online [W3T16]; and Portuguese, because health consumers encounter more difficulties searching in this language due to the reduced access of reliable information as it is a less used language and it is not known any similar tools performing for that language.

1.3 Contributions

As a result of this dissertation, several artifacts were produced. The extension, which is available for download at https://hugosousa.github.io/HealthTranslatorClient/. The server is hosted in a virtual machine provided by the Faculty of Engineering of the University of Porto, but it is also possible to host it in a different machine by following the instructions on the referred website. Furthermore, a corpus of Portuguese health-related documents annotated with medical concepts was also created to evaluate the annotation in this language. This corpus, composed of 38 documents, may be used as gold standard in health informatics applications and is available at https://github.com/HugoSousa/HealthTranslator-Corpus. Lastly, an article was submitted to the Special Issue on Biomedical Information Retrieval of the Journal of the Association for Information Science and Technology (JASIST).

1.4 Dissertation Structure

Besides the introduction, this dissertation contains six additional chapters. In Chapter 2, it is described the state of the art of medical concept recognition and annotation, presenting existent techniques, related work, gaps and challenges on this domain. In Chapter 3, development of solutions towards health consumers engagement are approached and related work is also presented. Chapter 4 describes the problem addressed in this work and its envisioned solution. Chapter 5 describes the implementation details, from the database building, to the server and client side development. Chapter 6 outlines the evaluation of the solution and its results. Lastly, Chapter 7 refers the conclusions and future work to be done.

Chapter 2

Medical Concept Recognition and Annotation

The research and development presented in this thesis involves the recognition of concepts in the medical domain. In this section, we will present the state of the art work on this subject, describing existent techniques, solutions, evaluation and challenges. Taking the goals of this thesis into account, both English and Portuguese approaches will be addressed.

2.1 Introduction

Natural Processing Language (NLP) is the field with roots from computer science, artificial intelligence and computational linguistics concerned with human language. Although in the old days most NLP systems were based on complex sets of hand-written rules, there was a revolution in this area with the introduction of machine learning techniques.

There is a vast list of major tasks commonly researched in NLP, such as information retrieval, sentiment analysis or machine translation. Named Entity Recognition (NER) is another task among these.

2.2 Named Entity Recognition

The goal of NER is to identify parts of unstructured text that relate to specific concepts of interest. A typical example with an extensive study is the recognition of types such as persons, locations and organizations, also known as "enamex". Later on, due to the availability of the GENIA corpus (a collection of biomedical literature compiled and annotated with various levels of linguistic and semantic information) and the appearance of biomedical NER competitions, such as BioCreative [STA+08] and JNLPBA [KOT+04], many studies were dedicated to types related to biomedicine, such as proteins, cells, DNA, RNA or drugs [NS07].

Due to different language characteristics, this task is not trivial by its nature and thus is dependent on the language. For example, capitalization of a word can aid in recognizing some specific type in English. However, some languages do not perform capitalization at all, such as Chinese or Arabic, and others capitalize all nouns, like German. In the same way, a named entity may be composed of several words in which only the first one is capitalized. The first word of a sentence may also be capitalized.

NER can be categorized in three different kinds of approaches: based on rules, dictionary matching or machine learning. These approaches have distinct limitations and technical requirements, so choosing the best performing solution is not always possible.

2.2.1 Pre-processing

There are several NLP tasks that are commonly used to aid major tasks as the ones previously referred. NER is not an exception, and so some tasks must be executed in order to properly prepare the data to be processed.

2.2.1.1 Sentence Splitting

Sentence splitting is the process of breaking the text into the respective sentences. Although it may initially sound a trivial task, there are some challenges to face. For example, a period does not always mean a sentence boundary, being a classic example the usage of terms such as "Dr." or "Mr(s).".

2.2.1.2 Tokenization

Tokenization splits the sentence into its meaningful units, named tokens. A basic division by whitespaces is not a robust solution, as punctuation marks are usually connected to words. In the biomedical domain, this task can be more complex, due to the nature of the concepts, such as genes names or usage of abbreviations. This step is important as the following tasks will be based on the tokenization output.

2.2.1.3 Stemming

Stemming is a technique that reduces an inflected word to its stem, usually by removing their suffixes. As an example, the stemming of words "stems", "stemmer" or "stemming" would result in "stem". A commonly used stemming algorithm in the current days was invented by Porter in 1980 [Por80].

2.2.1.4 Lemmatization

Lemmatization is a technique that finds the root form of a word. This may be useful to lookup for words that are not exactly the same, as they have been inflected. This may happen, for example, due to a different tense conjugation. For instance, "be" is the resulting lemmatization of "was".

This technique is usually useful where stemming technique fails but it is also more costly.

2.2.1.5 Part-of-speech tagging

This process is responsible for tagging each token with its grammatical category, based on its definition and context. Each token is then identified with a tag, such as noun, verb, adjective or determiner.

2.2.1.6 Stopwords

This process, commonly used prior to other NLP tasks, removes the words from text which are common to the language. Therefore, insignificant data is not processed, improving the system performance.

2.2.2 Resources

Available data resources is one of the constraints for applying a given NER approach. The needed resources can be split in two categories: knowledge bases and corpora.

Knowledge bases, such as ontologies or databases, are necessary for the development of systems targeted to medical concept recognition. They provide a centralized compilation of concept names definition and its respective classification. However, each resource is usually focused on a specific sub-domain. A recent study presents KaBOB [LBBH15], an ontology-based semantic integration of biomedical databases, facilitating the usage of multiple data sources.

Dictionaries are also useful for NER applications. Unified Medical Language System (UMLS) [Bet09], referred as being "probably the most comprehensive ontology in healthcare" [NM04] is composed by three major components, in order to facilitate the development of applications related to biomedicine and health information, which are:

- Metathesaurus — a large, multi-purpose, and multi-lingual thesaurus that contains millions of biomedical and health related concepts, their synonymous names, and their relationships.
- Semantic Network — semantic types that provide a consistent categorization of all concepts in Metathesaurus and relationships between them.
- SPECIALIST Lexicon — provides the lexical information needed for the SPECIALIST NLP System and includes commonly occurring English words and biomedical vocabulary.

A corpus is a collection of text documents usually annotated with relevant concepts of a certain task or domain. It can be classified as gold standard or silver standard, as the annotations were performed manually by expert annotators or by computerized solutions, respectively.

Building gold standard corpora requires significant efforts and is a time consuming task. There are some corpora available in the biomedical domain, with variable dimension and focus on different concept types. David Campos presents an extended list of relevant corpora in biomedical domain in its research [Cam13, p. 27-28]. These corpora however are usually focused on some subset of the domain, such as only genes and proteins, disorders or species.

In this context, corpora is important either for machine-learning methods (see 2.2.3.3) or to evaluate a NER system (see 2.2.4). There are several annotation standards usually used in NER competitions, improving the interoperability of annotation tasks, such as brat standoff format [bra] or BioC [CIDC+13]. Some formats are more complex than others, as they support more features.

2.2.3 Techniques

2.2.3.1 Rule based

Rule based systems rely on the definition of complex rules specified by domain experts, based on orthographic characteristics combined with word syntactic and semantic properties. However, these rules are usually too specific, being focused on a given corpus. When applied in a different context, the overall performance tends to drop significantly. Rule based approaches are recommended for strictly defined and standardized concept names.

2.2.3.2 Dictionary based

A dictionary or gazetteer is a collection of words regarding a specific concept. In order to recognize the concepts in the text, a matching against the dictionary entries is performed. The dictionary may contain concepts of one or multiple types. It my also refer to an entity of an external knowledge resource.

One of the problems of this approach is a large number of false positives caused by concepts with short names. This can, however, be avoided, by removing those names from the dictionary, which implies that those terms are not recognized in the text. As most of these short names are abbreviations, some abbreviation resolution techniques can be applied to it [SH03].

Another common problem is the missing of spelling variations of the concepts on the supporting dictionary. This can be overtaken by applying approximate string matching techniques. Nonetheless, they should be used carefully, as they may also result in a large number of false positives.

There are some different ways to deal with the dictionary string matching: exact matching, approximate matching and Soundex algorithm [NS07].

On exact matching, the words needs to be exactly as specified on the dictionary for a positive match. However, some flexibility may be required. One possibility is to apply stemming or lemmatization to the words before matching. Some languages may also replace diacritics by their canonical equivalent, such as replacing "à" by "a".

The second approach, approximate matching, also known as fuzzy matching, calculates edit-distance [TT03] or Jaro-Winkler's [CS04] metric between two words, comparing the difference level between them. Posteriorly, matching may be successful below a given threshold value.

Soundex algorithm [RA04] normalizes words to their respective Soundex codes. This code is a combination of the first letter with a 3 digit number that represents its phonetics sound. Consequently, similar sounding names will match.

Summarizing the previous techniques, exact matching is commonly used for its simplicity and efficiency, although it requires a complete dictionary with spelling variations of concepts for a good performance. Regarding text-modifying techniques, stemming is less accurate than lemmatization, but may be useful enough for the system and may be considered if performance is a relevant factor. Soundex is mostly useful for texts with spelling errors. All these text-modifying techniques are language-dependent.

2.2.3.3 Machine Learning based

Machine learning based approaches apply algorithms to learn how to recognize specific concept types. This brings an advantage compared to dictionary based approach, as it can possibly recognize new concepts not yet specified in a dictionary. However, as machine learning doesn't provide a mapping of recognized concepts to knowledge resources, dictionaries also play an important role in this approach. The main shortcoming of machine learning based approach is the requirement of a large annotated corpus.

The base idea of machine-learning methods is to train computational models on annotated texts which are posteriorly applied on non-annotated texts, predicting concept names.

Machine learning models can be classified in three categories, depending on the used data: supervised learning uses labelled data on the training phase to generate a function that maps inputs to desired outputs; unsupervised learning doesn't require annotated data and applies appropriate functions to infer data; semi-supervised learning or distantly supervised learning mixes both previous approaches.

Various studies tried different techniques and algorithms in this field, such as Hidden Markov Models (HMM), Decision Trees (DT), Maximum Entropy Models (MEM), Support Vector Machines (SVM) and Conditional Random Fields (CRF). Specially SVM and CRF are the most widely used techniques in NER tasks, achieving good results on problems with heterogeneous characteristics. For example, in 2010 i2b2/VA challenge [USSD11], CRF models were the ones performing better. It also shows a trend toward ensemble methods, which can mix a combination of machine learning methods, in order to improve the models' performance. One of the main difficulties of this challenge was the boundary detection of concepts.

2.2.4 Evaluation

In order to improve the development of new and better methods related with information extraction or natural language processing techniques, there are conferences and workshops that encourage research teams competition. A named entity recognition task was firstly introduced in the sixth version of Message Understanding Conferences (MUC-6) [GS96]. This conference was succeeded by Automatic Content Extraction (ACE) [DMP+04] and later on by Text Analysis Conference

(TAC) [oSN]. Conference on Natural Language Learning (CoNLL) [con] is another reputable conference on this field, running since 1997. BioCreative [STA+08] and JNLPBA [KOT+04] are some examples of NER competitions in the biomedical domain. HAREM [SC07] was the first NER evaluation competition in Portuguese language. This language was also present in some tasks of Conference and Labs of the Evaluation Forum (CLEF) [cle].

Concerning the evaluation of the produced systems, metrics were defined and standardized. Although there are some differences on the evaluation across different competitions, they are all based in the same base concepts: precision, recall and F-measure. These metrics are obtained through the comparison of a corpus manually annotated by domain experts against the automatic annotation by system. First, it is necessary to understand the following concepts, regarding the classification of an automatic annotation:

- True Positive (TP) — the annotation is present in both corpus.
- True Negative (TN) — the annotation is not present in none of the corpus.
- False Positive (FP) — the annotation is present in the automatic annotated corpus, but not in the manual annotation.
- False Negative (FN) — the annotation is not present in the automatic annotated corpus, but it is annotated in the manual annotation.

Exact and approximate matching may be used for classification. For example, overlapping concepts may be considered a True Positive in some evaluation metric, but not in another.

Precision is the ability of a system to present only relevant items, and is formulated as:

$$Precision = \frac{TP}{TP+FP} \tag{2.1}$$

Recall measures the ability of a system to present all the relevant items, and is formulated as:

$$Recall = \frac{TP}{TP+FN} \tag{2.2}$$

F-measure, also known as F_1 score, is the harmonic mean of precision and recall. It can be interpreted as a weighted average of the precision and recall, where an F_1 score reaches its best value at 1 and worst at 0, and is formulated as:

$$F-measure = \frac{2.Precision.Recall}{Precision+Recall} \tag{2.3}$$

2.3 Available Tools

There are several tools aimed for generic NLP tasks which may be used to integrate a system processing pipeline, such as OpenNLP [Fou], NLTK [BKL09] or Stanford CoreNLP [MSB+14]. An extense evaluation and comparison of generic NER systems has been done before [MSCLA09].

Regarding the biomedical domain, one of the first and widely used annotation systems is Metamap [Aro01, AL10], which uses a rule based approach and is based on UMLS Metathesaurus. However, Metamap does not use dictionary or machine learning solutions, which have shown better results [Cam13].

Several systems also use dictionary based approaches, specially because of their urge of fast processing. Early research presented IndexFinder [ZCM+03], which was faster than existent solutions, including Metamap. The main idea is related to the way data is structured, so the dictionary lookup time complexity is reduced. Later on, ConceptMapper [TCS10] arises, an highly configurable system, including the choice of lookup strategy. Both solutions load the entire dictionaries in memory, not having any performance issues. In the current year, NOBLE [TML+16] is also a highly customizable system regarding the concept matching and flexibility to terminology import. In respect to speed, it also achieves good results, as it caches the 0.2% most occurring words in the dictionary and also uses a NoSQL solution (JDBM [jdb]) to persist the data structures on disk.

Metamap has been compared to Mgrep, in order to find the most suitable systems to use in Open Biomedical Annotator (OBA) [JSM09]. The latter shows clear better results than Metamap for large-scale service oriented applications [SBJ+09]. Mgrep is not openly distributed, and the way it works is not completely clear, as publications on it have been limited to conference posters [DSX+08]. However, authors of OBA claim that it uses a dictionary based approach which implements a novel radix-tree-based data structure [JSM09, SVMA12].

David Campos' research [Cam13] presented in 2013 a set of software systems related to biomedical concept recognition, with state of art performance results.

Gimli [CMO13b] is one the systems presented by David Campos. It is an automatic annotation tool which applies machine-learning methods, more specifically CRFs. It provides an extense comparison of results against a large set of open and closed source systems for biomedical named entity recognition, evaluated on two corpora. Gimli outperforms all the open-source solutions.

Neji [CMO13a], the second system presented, is based on Gimli results, providing a machine-learning approach, complemented by a dictionary based approach. The results are also compared with similar systems, such as Whatizit [RSAG+08] and Cocoa [coc].

As a higher level of abstraction and availability for end-users, BeCAS [NCMO13] is a concept annotation tool based on Neji, available as a web tool or a widget.

Whatizit, Cocoa, OBA and BeCAS are examples of systems that are available through Web services, allowing its usage by external applications.

Concerning Portuguese language, MedInx [FTC11] is an example of a system that structures information of clinical discharge records written in Portuguese, including a NER step. This system enables faster and more accurate data creation and analysis.

Regarding the automatic annotation of web pages, Reflect [POJ+09] is a tool distributed as a plug-in for web browsers, which automatically tags gene, protein and small-molecules in any web page. Clicking on a tagged concept, more information is shown, depending on its type. The annotation is dictionary based and the system is mainly focused on speed instead of accuracy.

2.4 Summary

Natural Language Processing is still a field with a lot of research to be done. The task of recognizing named entities in unstructured text has been the subject of efforts in the last years. With the appearance of annotated corpus and competitions, it has become a major interest in the biomedical domain, particularly the recognition of genes, chemical entities and drugs.

New approaches and techniques have been explored along the years. Although results are getting better over the years, with the emergence of new approaches and techniques, there are still some challenges to face. Dictionary based or machine learning approaches have proved to provide significantly better results than rule based approaches. Machine learning has been evolving, achieving good results. Semi-supervised learning tries to overtake the need of a large annotated corpus. Dictionary matching methods are still widely used, specially in cases where there is lack of context, such as query reformulation, or speed is an important aspect. Additional processing such as noun phrasing, part of speech tagging, pre-processing of the vocabulary, or filtering results to specific semantic types, allied with a dictionary based approach may significantly improve the performance.

The performance of a system is usually strongly attached to the domain it is aimed for. Changing the domain significantly affects the performance of a system.

Another problem of the current systems is their language-dependence, what demands an adaptation to the target language. Only one system was found processing NER task in Portuguese texts. Some generic NLP tools already provide task methods for several languages, such as tokenization or stemming.

Thus, two main issues can be identified, and consequently can be matter of study and research, taking into account the problem approached by this thesis: first, the analysed systems are focused on a sub-domain of biomedicine, which do not perform well on different tasks. As the goal application is directed to laypersons, the processed language should be more generic than the one used in these systems, which is usually scientific literature. Second, the study of systems adapted to NER on Portuguese language is still premature.

Chapter 3

Developments towards Health Consumer Engagement

Nowadays, the Web emphasises user-generated content, usability and interoperability. Accompanying this effect, the individual's active partnership in health care grows, also referred as patient empowerment, an international challenge to enable patients to manage and organize their own healthcare data.

This chapter describes the efforts made on the development of resources or applications directed to health consumers. Once again, both English and Portuguese languages are approached.

3.1 Introduction

The Web is nowadays one of the main sources to search for health related information as "in the digital age, control has shifted to the individual" [Pea15]. Patients are becoming healthcare consumers, instead of passive recipients of healthcare [OS06].

One of the initiatives that promotes patient empowerment is the creation of Personal Health Records (PHRs) [BSB07]. These are similar to Electronic Health Records (EHRs) which refer to the collection of health information of a specific patient in a digital format. However, PHRs are records managed by the patient, usually available through the Web, where clinical records or test results can also be accessed.

Nonetheless, there are challenges regarding the patient empowerment. One of the biggest issues is the language gap between patients and health professionals, due to their domain knowledge.

Keselman et al. identify the main challenges and recommendations in the field of consumer health informatics. It is recommended "work on the development of consumer health vocabularies, implementation of tools for information retrieval and readability support, integration of user needs and usability concerns into the design of consumer health information resources, and assessment of users' health literacy as well as the quality of information resources" [KLS+08].

3.2 Consumer Health Vocabularies

A study found that using technical terms rather than its related consumer terminology results in better search outcomes [PMS+01]. Consumers usually have difficulty finding and understanding medical information due to their domain knowledge. Thus, there was the need to build a lexicon that maps both terminologies called consumer health vocabulary (CHV). The development of CHVs present many challenges as the contextual, sociocultural and other factors affect how laypersons express about health topics. In order to develop and build a CHV, many steps are required, from obtaining potential consumer terms from automatic solutions to manual reviewing by medical experts.

In order to understand the health consumer needs and improve CHV developments, several automatized approaches have been tried. One of the first approaches was to identify consumer-friendly display (CFD) names by analysing queries from a specialized site in health information directed for laypersons [ZTC+05].

Another approach mines concepts from community-generated text, such as Wikipedia [VMHZ14]. Wikipedia is a resource frequently updated and visited by the general public, including for healthcare information, as it often appears in the first results of various search engines. At least half of all healthcare changes on Wikipedia disease articles are changes relevant to patients [IMS14]. Also, about half of these articles are changed by healthcare professionals, so it is likely that they contain both medical and lay terminology [VMHZ14]. A similar experiment was conducted in a health-related social network [DHZT11].

A similar approach applies a CRF model, named ADEPT [MH13], in order to recognize patient-authored text (PAT) from text written by consumers in medical forums. This presents greater results when compared to existent solutions, providing a good tool to identify new potential concepts for inclusion in available CHVs.

A first generation of an open access collaborative consumer health vocabulary - OAC CHV - started in 2006 [ZT06]. It maps the terms to the UMLS medical ontology, which nowadays includes this vocabulary.

Some similar terminologies have been previously developed, specially in between the 1990 and early 2000s, but some of them are proprietary [KLS+08].

Studies found that some consumer health concepts in the OAC CHV do not fit in any UMLS concept. Some of those were identified as being uniquely "lay" and thus, not feasible for introduction in professional health terminologies. Others refer to relatively novel concepts in the domain and not yet present in the vocabularies [KSD+08].

Regarding the Portuguese language, there is a proposal to develop a CHV, which should start the first phase on the current year, based on data extraction from the Web. The second phase pretends to validate the concepts by human reviewers and the last phase is intended to store the data in Resource Description Format (RDF), for easy data interchanging [TTP15].

Health Translations is a tool that aids the translation of medical vocabularies through a gamified approach. Its initial purpose is the translation of OAC CHV to Portuguese, but it can be applied to other vocabularies and languages [SL16].

3.3 Tools for Health Consumer Aid

Aiding the health consumers understanding medical information is a subject of research for a long time now. In 2000, a proposal named Infobutton, presented an approach that helps health consumers to understand Pap smear results. It consists in a list of questions related to the concepts recognized in the text, which link to publicly accessible resources on the Web [BC00]. Actually, this approach has evolved till the current days and is still a matter of research. One example of such an application is OpenInfobutton. This system analyses contextual information about the patient, user, clinical setting and EHR task, providing links to external resources that satisfy the clinician's information needs [DFCC+13].

In 2007, a prototype of a translator was presented [ZTGK+07], focused on term replacement and explanation generation, since vocabulary is one of the main difficulties pointed by users. It uses UMLS and OAC CHV as knowledge sources. Comparing to infobuttons, this approach does not break the reading flow. This system was improved in 2010 where syntactic simplification was executed, resulting in better performance [KCZT10].

A similar approach presented in 2006 shows a definition, mined from the Web, of an unfamiliar term. A term is said unfamiliar if it belongs to one of the following UMLS semantic types: diseases, therapies, drugs, chemicals or pathological functions [Elh06].

NoteAid is a system that contributes to increase the readability of EHR notes, by linking concepts to external resources, such as Wikipedia, UMLS and MedlinePlus [MLB00]. Wikipedia has shown to be the resource that performs better, while UMLS and MedlinePlus need to improve their readability and content coverage for consumer health information [PRHB+13].

Medical Translator, a Google Chrome extension that translates medical jargon on any web page, was released in 2014. This extension is developed by Iodine, a company committed to provide medical information, such as medications and drugs. However, the methodology and processing of this extension is not published in any academic work. Given the main focus of the company' products, when medication concepts are recognized in text, the user is redirected to the company website [Inc].

3.4 Summary

Tools are useful to make health information more accessible to consumers. However, there has been relatively limited prior research on such tools [KLS+08]. Some different approaches have been tried with relatively satisfying results. Synonym replacement, concept explanation, linking to external resources and syntactic simplification are some of the used approaches.

The presented systems focus mostly on EHR notes, as the patient engagement initiative allows them to contribute to their own healthcare. Nonetheless, these systems can easily be adapted to web pages, aiding users on their health information searches. One exception was found, a Google Chrome extension that provides lay definitions of medical jargon in Web documents.

The performance of computational tools that make use of community-generated text are highly affected by the accuracy and comprehensiveness of CHVs. Nowadays, and after some years of development, OAC CHV is in a quite mature state, presenting more than 150 000 concepts mapped to UMLS.

Relatively to Portuguese, CHV development is in its initial phase, so it is not yet ready to use. Health Translations is a tool that can aid this process through a gamified approach. Regarding applications or systems related to aid health consumers, all presented systems are only compatible with the English language. Thus, there is the need of development of similar systems in different languages, such as Portuguese.

Chapter 4

Problem and Solution Proposal

In Section 2, it is shown that there is a significant effort into concept recognition in biomedical domain. However, the systems are usually focused on a small subset of subjects from the medical domain, and do not perform well when the context is different. In Section 3, recent developments to empower health consumers are presented. Consumer health vocabularies are an important resource which map lay terms to technical terminology and associate them to biomedical knowledge bases, like the UMLS. Most of the developed systems improve the readability of EHRs. Only one system was shown to perform in Web documents, although it is not documented in any academical source.

Analysing previous research, challenges and areas to focus on were identified. This chapter presents the identified problem and proposes a solution for it.

4.1 Problem

In general, health consumers are not effective when searching for health related information on the Web. This happens because there is a significant difference between lay and technical terminologies. This domain knowledge gap makes it hard for health consumers to formulate appropriate queries or effectively select from the retrieved results [TL07]. Besides the lexical and semantic differences, there is also a mental model mismatch between professionals and laypeople. Consequently, there is a poor information retrieval performance regarding health consumer searches [ZKA+02].

Even when the retrieved documents are relevant for the information need of the consumer, there is still another step and challenge to face: understanding the content. Due to the terminology differences, health content on the Web usually requires a certain level of health literacy to be understood [KLS+08].

To worsen this problem, there is also a language barrier that may restrict an efficient health search. Users not proficient in widely spread languages with significant quantity and quality of information, such as English, face this problem [KM06].

4.2 Solution Proposal

This dissertation aims to settle these limitations by providing a tool that helps health consumers during their searches on the Web. This solution should be presented as a browser extension, which automatically annotates medical concepts in Web documents and provides additional information. Thus, the research question behind this work is:

Is the comprehension of health related content on the Web by the health consumer facilitated by the usage of a tool that recognizes medical concepts and provides additional information about those?

4.2.1 Functional Requirements

Concept Annotation The extension should recognize and highlight medical concepts on a Web document.

Dynamic Page Support The system should perform in dynamic pages. Although some years ago the Web served mostly static content, it has evolved into a different paradigm: dynamic content. Thus, the system should be able to perform on this kinds of documents, by processing dynamically generated content and recognizing concepts. Good examples of this kind of pages are chats or even Google search page, where content is dynamically changed while the query is being written.

Overview of a Concept The system should provide a definition and possible synonyms in distinct terminologies when hovering a recognized concept. These two pieces present an overview of the concept and the information may be relevant for the comprehension of the concept by the user or might encourage the user to explore it in more detail. While the definition can be self-explanatory, the synonyms may be helpful if the user knows another term referring to the same concept.

Details of a Concept The user should be able to access the details of a concept, such as external references and related concepts. The external references might point to useful sources of information regarding the concept. Related concepts might facilitate the comprehension of a concept. The user should also be presented with the semantic type(s) of the concept. This immediately tells the user what kind of concept he is seeing.

Rating of Concept Information The user should be able to rate the presented information. This allows the gathering of statistics and the information can be improved by professionals on the background.

Suggestion of New Concepts The user should be able to suggest new medical concepts that are not recognized by the extension, so the database might be improved and more concepts are recognized.

Processing Feedback The extension should provide feedback to the user about the processing state of the current document. It should present on the extension's icon its state through symbols. If the processing is successful, the number of recognized concepts should be presented.

Customization The system should be highly customizable by the user in order to satisfy different user's needs. Possible customizations should include language preferences, filtering of concepts to recognize or concept highlighting preferences. The execution of the extension should also be configurable: automatically process every Web document or only when the user requests it by clicking on the extension's icon.

System Status The system should provide information about its status and response times.

4.2.2 Non-Functional Requirements

The extension might be useful for different kinds of users. Although the main target is the lay health consumer, it might also be helpful for professionals. First, because the solution highlights the concepts and might help focusing on the important parts of the content. Second, the synonyms in different terminologies might be useful to understand how laypeople refer to some concepts. On the other hand, health professionals, for their knowledge in the field, might help improve the extension through the suggestion of new medical concepts or rating of provided information.

Being an extension directed to a broad type of users, it must be user-friendly. It must be easy to learn, understand and use. It should be not intrusive, and not interrupt the normal cycle of searching, as it should be a help resource instead of an obstacle. The time for the concept recognition and provision of extra details of a concept should be reduced, so that the user search flow is not affected. The provided information should be reliable.

The system should be modular to facilitate the support of new languages or addition of new features.

In a real usage environment, the system should be up on a very high percentage of the time, in order to satisfy its users' needs.

4.3 Solution Innovation

The application differs from the ones presented in Sections 2 and 3 on the following points:

- is targeted for regular Web documents instead of scientific literature or EHRs/PHRs, which present distinct semantic styles and readability levels. Websites present a great diversity of content and context.

- mixes different approaches in order to aid the health consumer search: providing definitions, related concepts and external resources. In a prototype translator tool previously referred [ZTGK+07], the authors state that 'the optimal approach may be to combine text translation with infobutton'. Infobuttons consist in retrieving more information to the user, such as linking to online resources, by using contextual information. Therefore, the extension provides the benefits of this tool, plus the external references advantages of the infobutton standard.

- aims language modularity. Portuguese and English are the initial supported languages as proof of concept, but it should be easy to add support for new languages. It is not currently known any similar system supporting Portuguese.

4.4 Summary

The goal of this dissertation is to develop and evaluate a tool which helps health consumers search the Web. The tool should help users overcome difficulties caused by terminologies mismatches, providing a definition of the concept, possible related concepts and links to external resources with more information.

This system will differ from others in multiple aspects: it will be the currently unique known system to perform in Portuguese; mixes and provides additional features regarding other competitor systems, such as providing related concepts, semantic types and external references; the target documents are very broad, as it performs on a non-controllable environment such as the Web.

Chapter 5

HealthTranslator

Chapter 4 proposes a solution in order to help users on health-related Web searches. The solution can be divided in five main features: recognize medical concepts in Web documents; provide a brief definition and terms in different terminologies for that concept; provide more details of a concept, such as other related concepts or links to external references; support the rating of the provided information of a concept; allow the suggestions of new medical concepts. Dynamic pages are supported by the extension.

The system is based on a client-server architecture. This chapter presents and discusses the implementation details of the system, from the server-side, including database preparation, to the client-side.

5.1 System Description

In order to help with Web searches, it is required an integration with the browser. The standard way for this integration and chosen approach is the development of a browser extension. Nowadays, the development of extensions for different browsers is not standard because of different underlying engines and thus a target browser needs to be selected. Google Chrome was the selected choice as it is the currently most used browser[Sta].

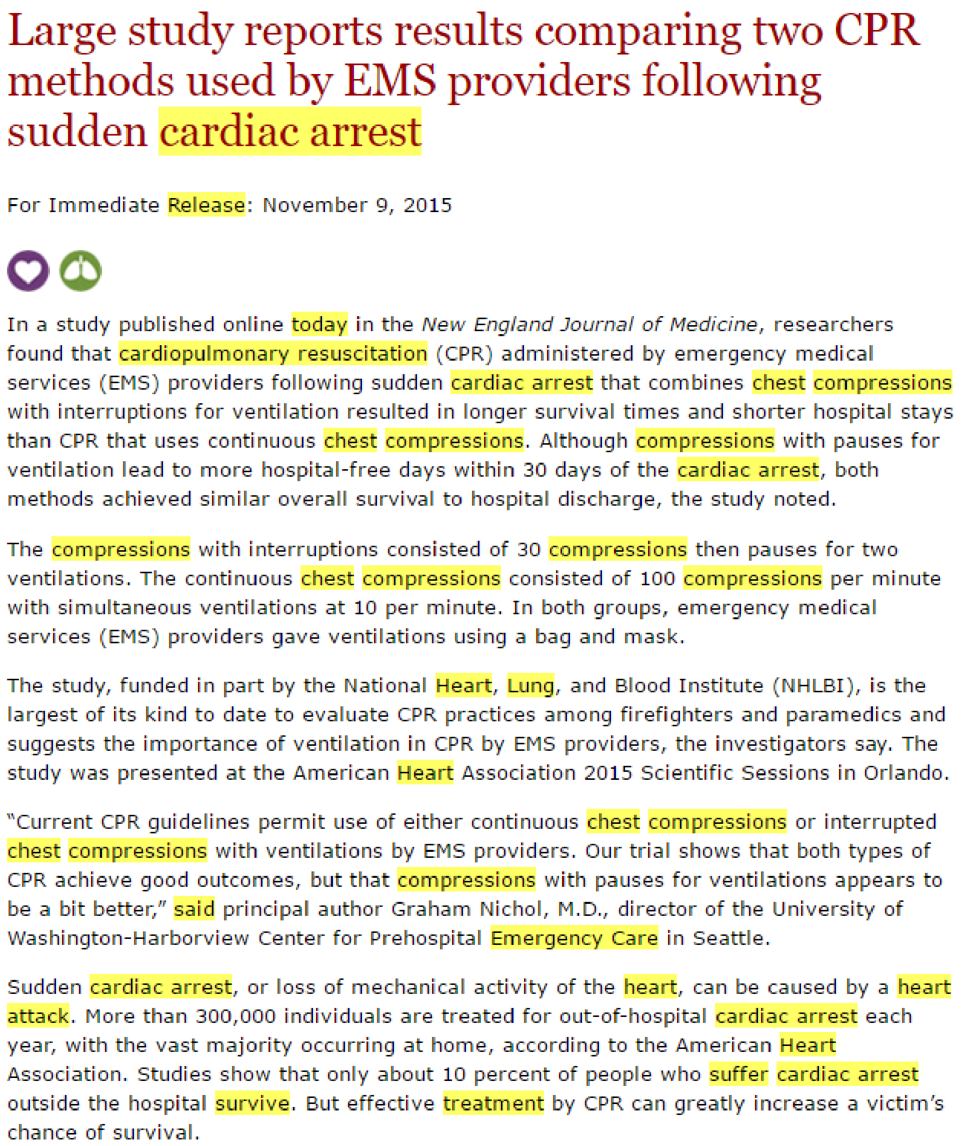
The next sections describe the features and presents the interface of the developed extension.

5.1.1 Document Annotation

The user interface presented by the extension is based on Bootstrap[1], a currently widely used framework for front-end Web development. It provides good-looking and responsive components.

When a Web document is processed, the user is informed about the state of the processing, which is displayed on the extension's icon. It can display the following messages: '...' if the processing started but no responses received yet; '-' if the server is down and the document could

[1] http://getbootstrap.com/



Large study reports results comparing two CPR methods used by EMS providers following sudden cardiac arrest

For Immediate Release: November 9, 2015

In a study published online today in the *New England Journal of Medicine*, researchers found that cardiopulmonary resuscitation (CPR) administered by emergency medical services (EMS) providers following sudden cardiac arrest that combines chest compressions with interruptions for ventilation resulted in longer survival times and shorter hospital stays than CPR that uses continuous chest compressions. Although compressions with pauses for ventilation lead to more hospital-free days within 30 days of the cardiac arrest, both methods achieved similar overall survival to hospital discharge, the study noted.

The compressions with interruptions consisted of 30 compressions then pauses for two ventilations. The continuous chest compressions consisted of 100 compressions per minute with simultaneous ventilations at 10 per minute. In both groups, emergency medical services (EMS) providers gave ventilations using a bag and mask.

The study, funded in part by the National Heart, Lung, and Blood Institute (NHLBI), is the largest of its kind to date to evaluate CPR practices among firefighters and paramedics and suggests the importance of ventilation in CPR by EMS providers, the investigators say. The study was presented at the American Heart Association 2015 Scientific Sessions in Orlando.

"Current CPR guidelines permit use of either continuous chest compressions or interrupted chest compressions with ventilations by EMS providers. Our trial shows that both types of CPR achieve good outcomes, but that compressions with pauses for ventilations appears to be a bit better," said principal author Graham Nichol, M.D., director of the University of Washington-Harborview Center for Prehospital Emergency Care in Seattle.

Sudden cardiac arrest, or loss of mechanical activity of the heart, can be caused by a heart attack. More than 300,000 individuals are treated for out-of-hospital cardiac arrest each year, with the vast majority occurring at home, according to the American Heart Association. Studies show that only about 10 percent of people who suffer cardiac arrest outside the hospital survive. But effective treatment by CPR can greatly increase a victim's chance of survival.

Figure 5.1: Example of a page processed by the extension. Recognized concepts are highlighted in yellow background.

not be processed; '?' if the language is not supported; an integer number which represents the number of medical concepts recognized in the document.

The recognized concepts are highlighted with a background color, as seen in Figure 5.1.

5.1.2 Information Provided

On hover of a recognized concept, a tooltip is displayed (see Figure 5.2), which contains a brief definition, possible terms in different terminologies and a link to provide more details of the concept, which opens a modal window with additional information.

Tooltips only show after some delay when hovering a recognized concept. This prevents the intrusiveness of the extension, by avoiding many tooltips to open simultaneously when a page contains many recognized concepts and the user randomly moves the mouse around.

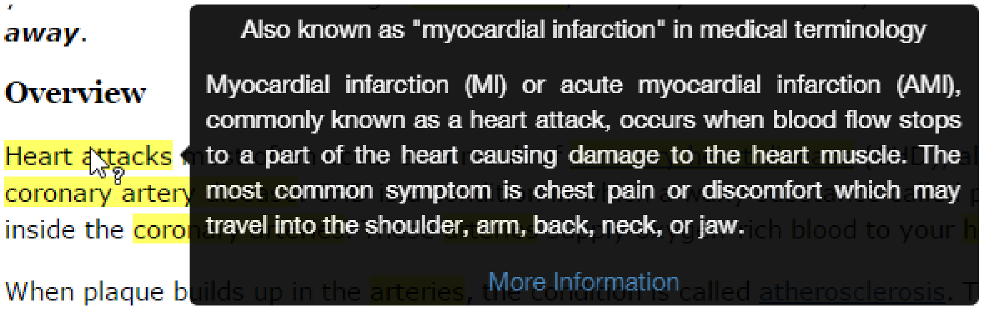


Figure 5.2: Tooltip displayed when hovering a medical concept

The details window (Figure 5.3) displays the semantic types related to the concept, a definition, external references to other information sources and related concepts. Excluding the semantic types that are always present in the details window, the other information might be unavailable.

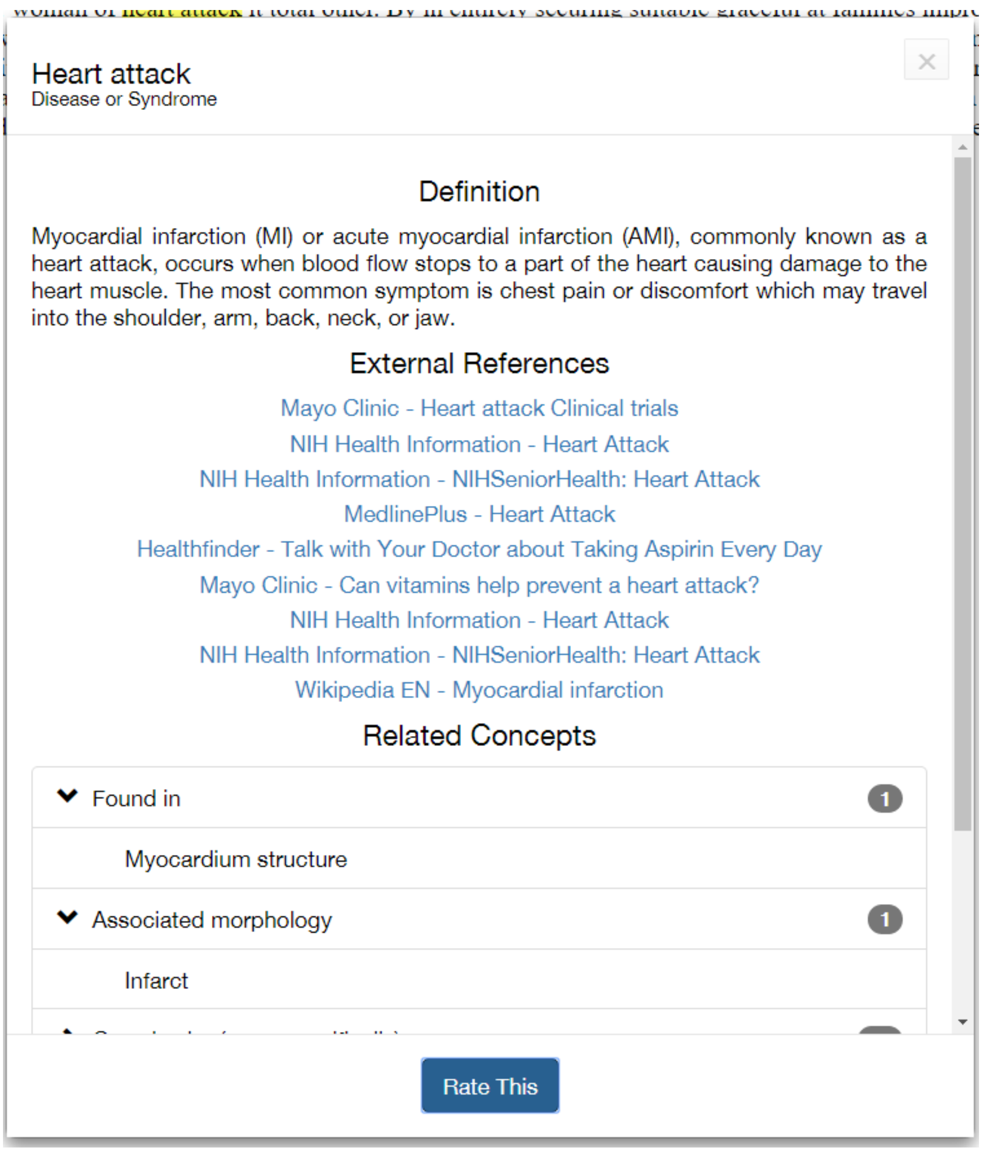


Figure 5.3: Details of a concept

5.1.3 Ratings

Users can rate the quality of the provided information on the details of a concept. The rating panel can be accessed from the details modal window, if the user has not previously rated that concept. Only the previously displayed information, among the definition, external references and related concepts can be rated, if that information was available. The user must also provide a general rating of the information quality (see Figure 5.4).

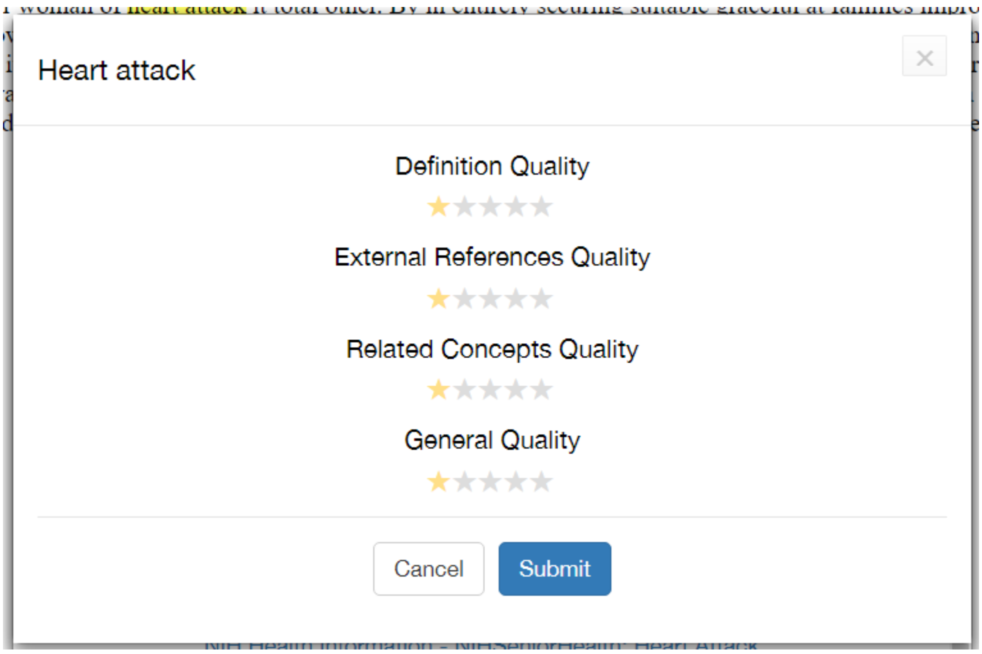


Figure 5.4: Rating of the information of a concept

5.1.4 Suggestions

A user can also highlight text in a Web document and suggest it as a medical concept (see Figure 5.5). It will display an error message if the concept was already suggested by the same user or if it already belongs to the set of recognized concepts. A user may suggest an already identified concept because he might have filtered the semantic types that should be recognised in a document. If the concept is not included in those semantic types it is not highlighted although it is already recognised as a medical concept by the system. In that case, a message informs the user that the suggested concept is already existent in the system's database.

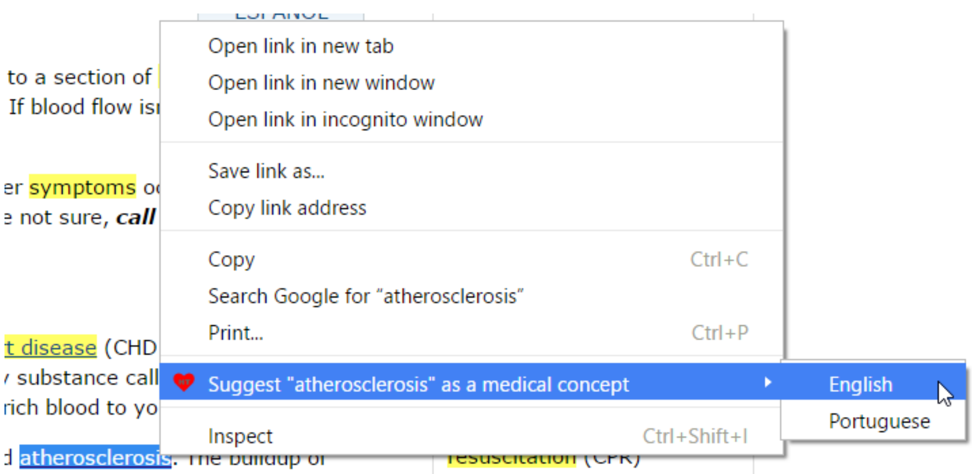


Figure 5.5: Suggestion of a new concept

5.1.5 Settings

The extension provides a settings page where the user can customize his preferences (see Figure 5.6). The complete list of default settings is shown in Appendix A. The customisation of the extension was always a concern and is also a factor that influences the user satisfaction. There is a variety of settings related to language, concepts filtering, execution mode or even styling.

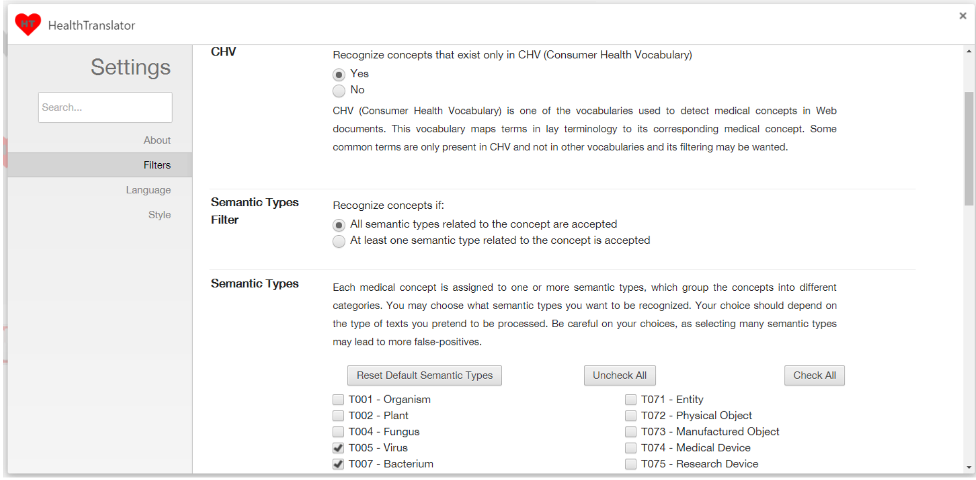


Figure 5.6: Extension settings

The user can also change the execution mode, that is, every Web document can be processed automatically or only when the user clicks the extension's button. It is also possible to check the server status and response time.

5.1.5.1 Filterings

The user may want to filter the medical concepts he wants to see identified in the Web document and the system provides many possibilities of customisation.

Since the definition captures most of users' attention and it is the quicker way to obtain a summary of that concept, it is given the possibility to filter concepts without a definition.

Users can also choose not to recognize concepts only present in CHV. This may be desired if the user does not wish to have lay terminology recognized in the documents, although some of this concepts may also be present in the chosen medical vocabularies. In the case of Portuguese, it may also mean a wrong translation and a filtering may provide better results.

In UMLS, concepts are grouped in broad subject categories named semantic types. A concept may be contained in several semantic types. The user has the possibility to choose the filtering strategy of semantic types. A concept may be recognized if all the related semantic types are accepted or if at least one is accepted. He can also choose from the list of 127 semantic types, which ones to be detected in the document. A list of 29 semantic types that seem to have a considerable relevance for a lay health consumer are set by default. An excessive selection of semantic types may lead to a big number of false positives, as many categories have misleading concepts in different contexts. The default semantic types are the following: ‘Virus’, ‘Bacteria’, ‘Congenital Abnormality’, ‘Acquired Abnormality’, ‘Body Part, Organ or Organ Component’, ‘Body Location or Region‘, ‘Body Space or Junction’, ‘Injury or Poisoning’, ‘Pathologic Function’, ‘Disease or Syndrome’, ‘Mental or Behavioural Dysfunction’, ‘Laboratory Procedure’, ‘Diagnostic Procedure’, ‘Therapeutic or Preventive Procedure’, ‘Organic Chemical’, ‘Amino Acid, Peptide or Protein’, ‘Pharmacologic Substance’, ‘Hormone’, ‘Enzime’, ‘Vitamin’, ‘Immunologic Factor’, ‘Indicator, Reagent or Diagnostic Aid’, ‘Hazardous or Poisonous Substance’, ‘Sign or Symptom’, ‘Anatomical Abnormality’, ‘Archaeon’, ‘Antibiotic’, ‘Clinical Drug’ and ‘Eukaryote’.

5.1.5.2 Language

The extension supports English and Portuguese languages. The settings page is presented in English by default or translated to Portuguese if that is the user’s default browser language.

The language of a Web document is detected before a page is processed. User might choose to only process pages on a given language.

The language of the content of the information displayed to the user can also be configured. It can be shown in Portuguese, English or the detected language. A user may prefer to always have the information displayed in the same language, independently of the Web page. However, the language of the definition is not customisable. This is due to the fact that the English definitions were gathered from English Wikipedia and Portuguese definitions from Portuguese Wikipedia. Although the relationships between concepts are translated and customisable, the related concepts itself are displayed in English and are not customisable as they are included in UMLS in this language.

The user can also choose to always include English external references when requesting more information of a concept, even if the page is in Portuguese. Portuguese external references are more scarce and English is usually associated with high-quality contents. However, that means more waiting time when requesting for concept details.

5.1.5.3 Styling

With an aesthetic concern in mind, the user can choose the color of the background color to be displayed on the recognized medical concepts, from a predefined list of red, yellow, green and blue.

5.2 System Architecture

As seen in previous chapters, NER is strongly related with the available resources, as different approaches require a set of resources.

Two main resources are available: UMLS, which contains various health vocabularies in several languages and provides standards to access that data and other information such as relationships. The concepts are grouped in a set of broad subject categories named semantic types; OAC

CHV, an open access terminology that maps lay terms to UMLS concepts is also available. It is currently included in UMLS for English. As a result of a previous work [LR14], this vocabulary was translated to Portuguese, using Google Translate. 1% of the total translated strings were manually analysed and 84.2% were classified as good, which represents a satisfactory outcome.

As seen in Section 2, there are different approaches for NER: rule based, dictionary based and machine learning based. Machine learning requires an annotated corpus to train models for future concept predictions. Rule based require expert knowledge in medical domain and linguistics and it has been shown to have worse results than other approaches. Given that medical vocabularies are the only available resource, dictionary based is the chosen approach.

The medical concepts are recognized by matching Web content with terms present in medical vocabularies. Given the big amount of records, it is impractical to keep that data in the client storage. For example, the English database contains 1425855 terms in a table. Although the English database exported from UMLS has 51 tables and requires more than 7GB of space, only 3 of its tables are needed, which reduces the amount of data to more than 1GB needed. The Portuguese database also requires at least 360MB.

Nowadays there are some alternatives for client-side storage, such as IndexedDB[2], supported by Google Chrome. The usage allowed by apps in Chrome is up to 1/3 of the available disk space. Each app can then use at most 20% of that space shared by all apps[3]. This means that a user would need to have at least around 23GB free in the machine. Another issue is that this space is automatically deleted as the user's free disk space goes below the needed threshold. It could eventually be an option provided to the user, as it might reduce processing times and improve performance. This would not substitute the server, as it should be preferred by some users and some features would still need to be centralized, such as rating information or suggesting new concepts.

For the above reasons, a client-server architecture was adopted. Despite having the advantage of a bigger storage capacity, there are also some disadvantages, such as decreased processing performance and necessity of providing a server infrastructure, available and scalable.

The global view of the system architecture is represented in Figure 5.7, where internal and external communications between modules are outlined.

Two major ways of processing the Web documents were considered. The first approach was initially implemented, but was later changed after realizing its disadvantages. It consisted in sending the whole `<body>` to the server, which parses the DOM and wraps the medical concepts with relevant HTML. The whole `<body>` is then replaced on the client. Replacing the whole `<body>` is a bad practice, for three main reasons: is slow and destroys non-serialisable information, such as form field values; destroys Javascript references; destroys event handlers. The latter problem was realized in an early stage and it was decided to run all the page scripts again in the same order. However, the second problem was only discovered on a later stage, while trying to support dynamic pages. Processing the whole page again may take some time and if dynamic changes would

[2] https://developer.mozilla.org/en-US/docs/Web/API/IndexedDB_API
[3] https://developer.chrome.com/apps/offline_storage

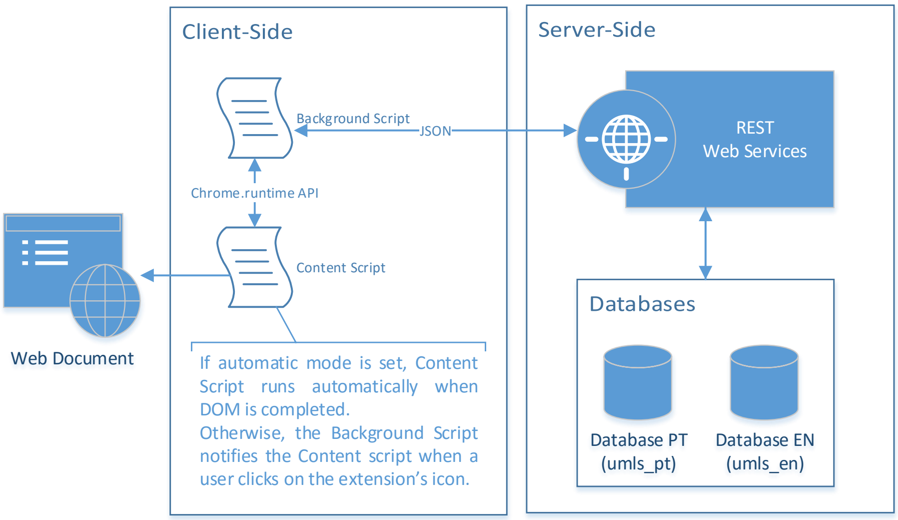


Figure 5.7: Global system architecture

be done during the request and the response, issues would arise, as new nodes would disappear and deleted nodes would reappear. As the Javascript references are lost when the `<body>` is replaced, there are no chances to deal with those nodes again.

Thus, it was decided to refactor the system, both client and server. The client gained new responsibilities, such as detecting the language and filter only text nodes, which are sent to the server. The server became simpler as it simply needs to process text given a language.

The major difference for the end user of both approaches is that the first one replaces the document once while the second approach incrementally detects the medical concepts on the page.

5.3 Database

There are two databases, one for each supported language. The construction of the databases consisted mainly in 3 phases, as shown in Figure 5.8. The loading of the biomedical vocabularies from UMLS, the extraction and loading of definitions and the creations of tables to support suggestion and rating features. Portuguese database has an additional step of loading OAC CHV, as it is not included in UMLS for this language.

Although there is a large quantity of information, the database structure is fairly simple and small (see Figure 5.9). From the UMLS structure, only 3 tables are used. A complete list and structure of the tables can be seen in the UMLS Reference Manual [Bet09]. Although no foreign keys are used, all of them are identified through the CUI, that is, the concept unique identifier from UMLS.

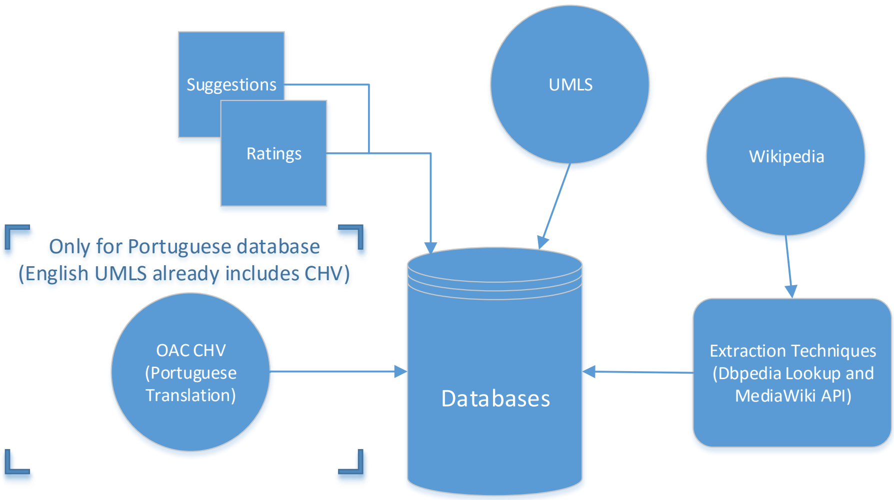


Figure 5.8: Phases of database creation

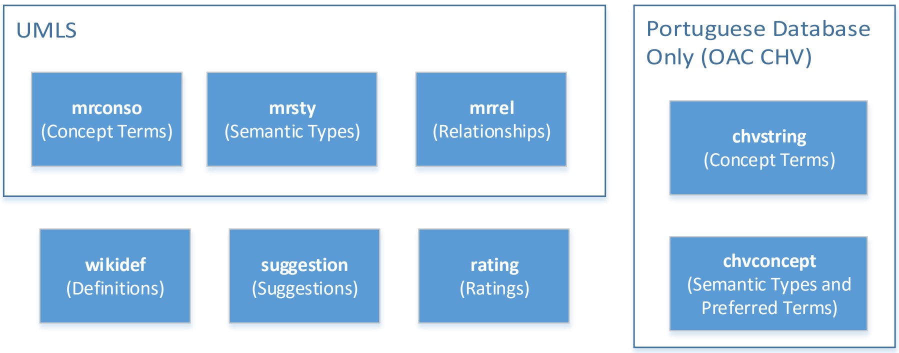


Figure 5.9: Database tables

5.3.1 Selection and Loading of Biomedical Vocabularies

The system should be able to process and recognize concepts from a broad range of medical areas in order to fulfil different user needs. UMLS is currently composed by 128 vocabularies in English and 5 in Portuguese. However, many of those are specific to a subset of the biomedical domain.

Regarding the English language, SNOMED CT was the chosen vocabulary. Given its maturity and scientific validation, it is currently designated as a standard for electronic exchange of clinical information in United States. Given its extension, it also covers the necessity of recognizing concepts from a broad range of medical areas.

Also present in UMLS, English OAC CHV was included in the database, in order to recognize concepts in lay terminology, but also to provide mapping between concepts in lay and medical terminologies.

Concerning the Portuguese language, the available resources are much more limited. There are no vocabularies with a large extension, such as SNOMED CT. Therefore, all the Portuguese vocabularies present in UMLS were included in the database, which are the following:

- WHOPOR - Portuguese translation of the WHO Adverse Drug Reaction Terminology (WHO-ART)
- MSHPOR - Portuguese translation of Medical Subject Headings (MeSH®)
- MDRPOR - Portuguese translation of the Medical Dictionary for Regulatory Activities (MedDRA)
- LNC-PT-BR Portuguese translation of Logical Observation Identifiers Names and Codes terminology (LOINC®)
- ICPCPOR - Portuguese translation of the International Classification of Primary Care (ICPC)

For example, 99.8% of LNC-PT-BR concepts belong to the semantic type 'Clinical Attribute' which shows the specificity of this vocabulary.

For the same reasons as the English language, the Portuguese translation of OAC CHV was included in the database. This vocabulary was already setup in a PostgreSQL database from a previous project [CTL16]. However, it did not contain the semantic types, which are needed for filterings. Thus, for each CUI in the OAC CHV, the semantic types were gathered through UMLS REST API[4] and added to the database.

Both databases are implemented in MySQL, as this is one of the systems that MetamorphoSys (the UMLS installation wizard and customization tool) can generate load scripts for UMLS data. The databases are named *umls_en* and *umls_pt* for English and Portuguese languages, respectively.

5.3.2 Extraction and Loading of Definitions

Providing a definition for a medical concept is one of the main features of the extension. Wikipedia[5] was the selected information source to gather the definitions of medical concepts. Wikipedia is a rich resource frequently updated and popularly accessed by the general public. The fact that about half of the medical editors are healthcare professionals make it a resource with a mix of lay and professional terminologies. It is also directed to laypeople and has a big range of medical entries [VMHZ14]. There have also been efforts to improve the medical entries in this encyclopedia [CR14]. Wikipedia was also shown to improve EHR readability by providing definitions, when compared with UMLS and MedlinePlus[6], given their lack of content [PRHB+13]. In another project, UMLS or other vocabularies definitions were also not chosen, as they often provide long and sometimes even more complex definitions than the term itself [ZTGK+07].

[4] https://documentation.uts.nlm.nih.gov/rest/home.html
[5] https://en.wikipedia.org/wiki/Main_Page
[6] https://www.nlm.nih.gov/medlineplus/

A new table named *wikidef* with the fields specified in Figure 5.10 was created in each database. A definition is specific for a given CUI, that is, a UMLS concept unique identifier. In UMLS, a concept may be defined by several different terms.

wikidef	
cui	CHAR(8)
url	VARCHAR(100)
def	VARCHAR(1000)

Figure 5.10: Table wikidef

The gathering of Wikipedia concepts was done in two rounds. First, a lookup for every term related to a CUI was processed. If the lookup finished with a single unique result, it was inserted in the database. Otherwise, it was written in a file for later disambiguation. For example, the term 'hypertension' corresponds to the CUI *C0020538*, but this concept may also be represented by 'hypertensive disease', 'HBP' or 'high blood pressure'. This method tried to prevent the insertion of false positives in the database or lack of definition for some concepts by searching only through the preferred term. Given the same example, the preferred term for CUI *C0020538* is 'Hypertensive disorder', but the respective Wikipedia page is 'Hypertension' (the page is automatically redirected however, but this may not happen with other concepts, as it depends on editor's suggestions). The second round consists of picking the concepts in the disambiguation file and lookup only for the preferred term. This can resolve the ambiguity of some concepts.

Regarding the English language, the service DBpedia Lookup[7] was used. DBpedia is a project that aims the extraction of structured information from Wikipedia. This service has some advantages compared to the MediaWiki API[8], as the matching lookup is different. DBpedia Lookup searches for related keywords which may be present in the label or in anchor text and the results are ranked by the number of inlinks from other Wikipedia pages. Besides that, disambiguation pages are not present in DBpedia and are not retrieved.

Only the initial phrases were extracted from Wikipedia pages. When using MediaWiki API, the 3 first sentences are gathered, as this parameter is possible to be defined. DBpedia Lookup retrieves a short abstract, which is a structured field in the ontology.

From the 374696 distinct CUIs of the English database, 56906 definitions were inserted in the first round, which leaves 317790 for disambiguation. From those, 306217 had 0 results, which means that only 11573 concepts could provide any results in the second round, if the same lookup

[7] https://github.com/dbpedia/lookup
[8] https://www.mediawiki.org/wiki/API:Main_page

service was used. However, MediaWiki API was used for the second round and 14520 definitions were inserted, which confirms the different lookup strategy, probably improved by the redirect terms defined by the Wikipedia's editors. This number may also represent that some inserted definitions are disambiguation pages, which are not desired.

In regard to Portuguese, given the difficulties of modifying the DBpedia Lookup service to support this language, the MediaWiki API was used in both rounds. In the first round, from the 157569 unique concepts in the database, 13169 definitions were inserted. In the second round, 1249 definitions were added. For concepts only present in OAC CHV, the UMLS preferred term was used. After running both phases, it was realized that concepts from LNC-PT-BR vocabulary are defined in a specific format, which prevents definitions to be found (and also concepts from this vocabulary to be recognized by the extension). The concepts translated to Portuguese are in a format X:K:W:Y:Z, where only the X is the relevant term name, whereas the next fields refer to other technical details. Thus, the terms from this vocabulary were updated, keeping only the X part. The second round was then run again with these terms, where 8521 new definitions were inserted.

From a general overview and by analysing some of the inserted definitions, it is possible to conclude that the DBpedia lookup is effectively different from MediaWiki API. The latter may gather more concepts, but also has disadvantages, such as retrieving disambiguation pages without useful content and consequently providing wrong definitions. Another limitation of MediaWiki API is the sentences boundary. For example, the page referring to the organ heart in Portuguese starts with 'O coração (Pronúncia em pt-pt pronúncia ajuda · ficheiro · ouvir) (lat. cor, grc. (...)'. MediaWiki API retrieves till "grc" when requested for 2 sentences, which is incorrect, as the dots refer to acronyms in this situation. Also, the excerpt 'ajuda · ficheiro · ouvir' corresponds to superscript anchors in Wikipedia, which are not desired for the definition.

Another detected problem was the insertion of false positives, specially in the first round, due to acronyms or concepts that have similar words in a different context. It is also observable that only a minority of the recognized concepts have an associated definition.

5.3.3 Preparation for Suggestions and Ratings

In order to support the proposed features of suggesting new concepts and rating the information presented for a given concept, two new tables were created in each database, named *suggestion* and *rating*, respectively (see Table 5.11). The table *suggestion* has a unique pair {*tuid*, *str*}, where *tuid* is a unique identifier of a user and *str* is the suggested string. The table *rating* has a unique pair {*tuid*, *cui*}, meaning that a user can only rate a concept once.

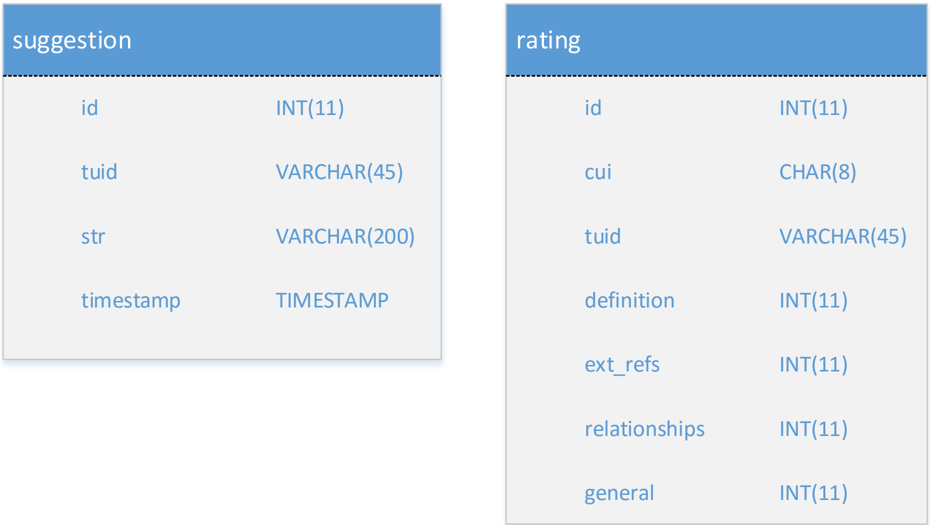


Figure 5.11: Tables suggestion and rating

5.4 Server Side

The selected language for the server side was Java, as it is a stable and mature language. Many of the NLP systems or other frameworks are implemented or adapted to this programming language.

The server offers five RESTful Web services, which are implemented using Jersey[9], an open-source Java framework that implements JAX-RS, a Java API that provides support in creating Web services according to the REST architectural pattern. The four services provided are the following: process, details, suggest and rating. All of them consume and produce JSON. An additional ping service was created in order to check the server status and response times.

5.4.1 Process

This Web service is responsible for processing text and recognizing medical concepts. It has suffered lots of refactoring over time, as different approaches have been tried in order to understand their efficiency and performance. It receives as input a string and other settings (see Table 6.2) and returns for each recognized concept the HTML to be replaced in the client side, with the medical concepts surrounded by HTML tags with relevant information, such as information to be displayed in the tooltip.

Before matching the concepts with the database values, a tokenization is performed. For that, it is used OpenNLP[10] with pre-trained models for English and Portuguese.

Some matching approaches have been discussed and analysed. Stemming the words would imply that either the database concepts are also stemmed or the SQL query is of the type <concept>

[9]https://jersey.java.net/
[10]https://opennlp.apache.org/

Table 5.1: Process service input parameters

Parameter	**Data Type**	**Description**	**Default**
body	String	The text to be processed	
language	String	The language of the text sent in parameter body	'en'
styFilter	String	Filtering strategy for semantic types - may be 'all' if a concept is only recognized when all semantic types are accepted or 'one' if at least one semantic type is accepted	'all'
recognizeOnlyCHV	Boolean	If set to true, recognizes concepts from CHV only. Otherwise, does not	true
contentLanguage	String	Language of the content to be displayed to the user	'en'
semanticTypes	HashSet<String>	List of semantic types accepted following the format 'T???' where ? is a digit representing the semantic type code	List of 29 default semantic types

`LIKE '<stemmed>%'`. Stemming the concepts resulted in a large amount of false positives, that is, many medical concepts were recognized when they should not be. The referred query also has bad performance, as it is much slower than a '=' comparison, even with the relevant fields indexed.

More tries were performed with `LIKE` queries. For example, for a given token, query the database with a `LIKE` select and check the maximum number of forward tokens in order to find the longest possible match for that token. Then, the number of tokens to look forward would be dynamically set and would eventually avoid several tokens to be iterated. For instance, when querying the English database with `str LIKE 'oligohydramnios%'`, the 'oligohydramnios' is a concept itself, but the largest possible concept is 'Oligohydramnios without rupture of membranes' and the threshold of tokens to look forward would be set to 5. After iterating to the next tokens, if it does not return any results, the concept 'oligohydramnios' would be recognized and the forward lookup could be stopped. However, the performance of this approach was not satisfactory.

The final decision was to query the database only with '=' comparisons. The tokens are previously singularized. This function was ported to Java from the Inflector class of Ruby's ActiveSupport library[11]. The Portuguese version was also adapted to Java from a set of rules extending the referred library and shared by Brazilian Rails, an internationalization plugin for Ruby and Ruby on Rails targeted to Portuguese developers[12].

In order to recognize concepts with multiple words, it is necessary to group tokens. For that,

[11] http://api.rubyonrails.org/classes/ActiveSupport/Inflector.html
[12] https://github.com/tapajos/brazilian-rails/blob/master/lib/inflector_portuguese.rb

there is a defined a threshold of tokens to look forward. For example, with a sentence 'A B C D' where each letter represents a token and with a defined threshold of 3, the lookup would be done for 'A', 'A B', 'A B C', 'B', 'B C', 'B C D', 'C', 'C D' and 'D'. The default threshold is 5 as it seems a satisfactory value to recognize medical concepts of 5 or less words. If the length of the first token is less than 3 characters or if it is contained in the stopwords, the token is ignored and the processing continues to the next token. The same logic applies if a punctuation mark is detected, independently of its position. The lookup does not stop when a medical concept is recognized, because it is intended to find the most specific concept. For example, with a string 'heart disease', it is intended to detect the whole string and not just 'heart'. The token to be processed next is the one immediately after the detected concept. That means that no overlapping concepts are detected.

The token processing depends on the language and is implemented by a class extending *ConceptProcessor* (see Figure 5.12), as there are differences on the database queries since the schema is different. It also means that, for each language, a different strategy to process the tokens and recognize medical concepts can be defined.

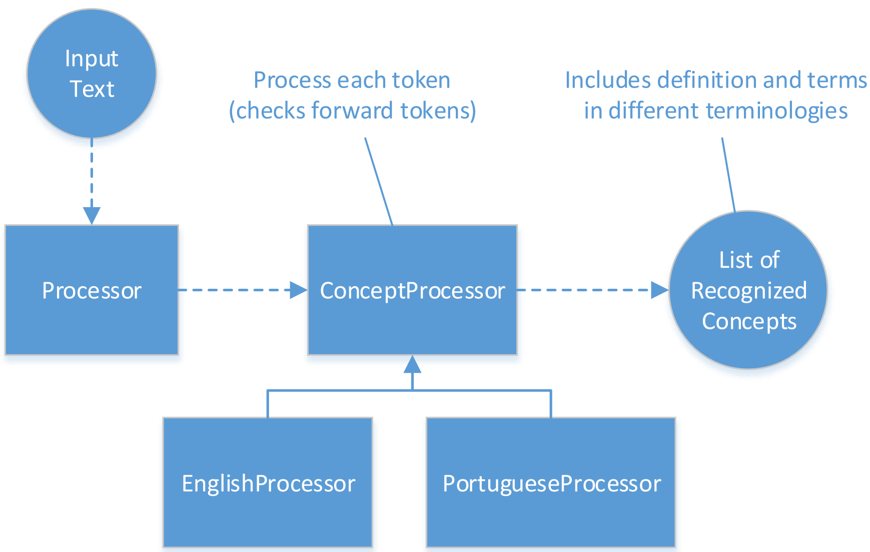


Figure 5.12: Process schema

Sometimes, a term can be related to different CUIs, but a CUI is needed to know to which concept it refers to. In order to disambiguate, it is given preference to preferred terms.

As the text processing should perform fast, only the definition and terms in different terminologies are initially retrieved, as it may serve the user's needs immediately. Retrieving the definition is fast since this information was previously inserted in the database (see section 5.6.2). Equivalent terms in medical and lay terminologies are presented by querying OAC CHV (or directly UMLS in English, as it includes CHV).

5.4.2 Details

This endpoint returns more details about a given concept. It returns a definition of the concept, semantic types, links to external references, related concepts and indicates if the user requesting the details has previously rated it. The input parameters can be seen in Table 5.2.

Table 5.2: Details service input parameters

Parameter	**Data Type**	**Description**	**Default**
cui	String	CUI of the concept to get details	
string	String	String of the recognized concept	
language	String	Language of the page where the concept was recognized	‘en’
includeEnglishRefs	Boolean	If set to true and language is not English, include English external references. Otherwise, does not.	false
tuid	String	Unique identifier of a user	

Similarly to the process service, language-dependent features are implemented in a class extending *ConceptProcessor*, such as *EnglishProcessor* or *PortugueseProcessor*. Regarding the details, the lookup for external references and retrieval of semantic types are two language-dependent tasks. The remaining features are implemented in *ConceptProcessor*, as they are common between languages, as long they keep some standards in the database, such as the name (see Figure 5.13).

The definition and semantic types are gathered from the database, through the CUI. Given the unique user identifier (tuid) and CUI, the service indicates if the user already rated that concept.

Regarding the lookup for external references, some relevant information sources for the medical domain were selected. Websites mostly directed to lay users were the initial subject of the search. Later on, those that provide an API for an easy and fast query were selected. Regarding the English language, MedlinePlus and healthfinder.gov[13] match those characteristics. Other than those, National Institutes of Health website[14] was selected, which does not provide an API and Web scraping is needed. English Wikipedia is also returned, which is present in the database in the *definitions* table. Concerning the Portuguese language, the choices were more limited. The selected sources are a Portuguese dictionary of medical terms from Porto Editora[15] and a questions and answers website targeted to lay users named Médico Responde[16]. Both choices require Web scraping to check and return existing results. Portuguese Wikipedia is also included.

The related concepts are retrieved from UMLS. There are many relationships between concepts, but not all should be interesting from a layperson point of view, as some of them are too technical. For example, the relationship ’allele_absent_from_wild-type_chromosomal_location’ does not sound relevant for a layperson. Most of the relationships have the associated inverse.

[13] http://healthfinder.gov/
[14] https://www.nih.gov/
[15] http://www.infopedia.pt/dicionarios/termos-medicos
[16] https://medicoresponde.com.br/

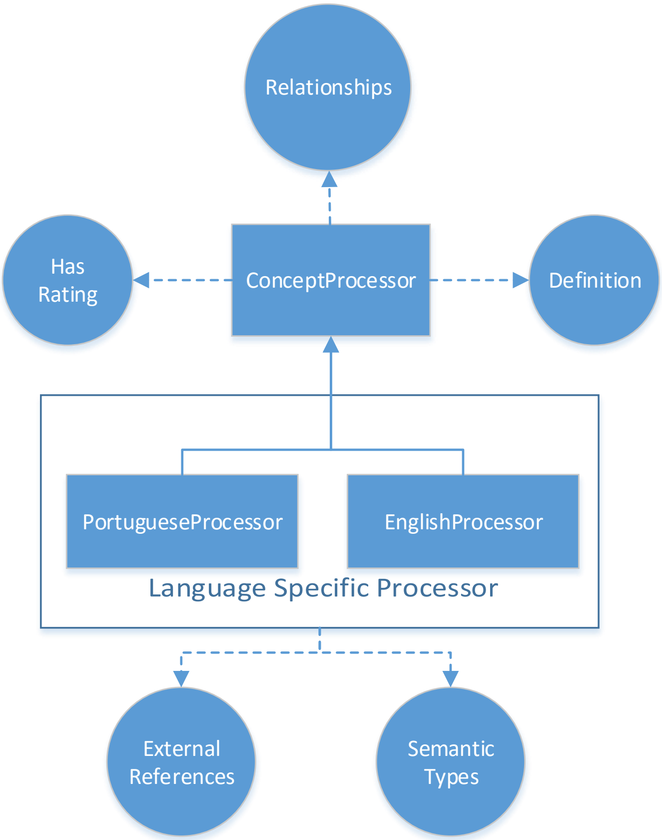


Figure 5.13: Details schema

Furthermore, the relationships that seem more interesting and relevant for a layperson were filtered, which resulted in the following list: 'same_as', 'due_to', 'cause_of', 'inverse_isa', 'isa', 'has_finding_site', 'finding_site_of', 'has_causative_agent', 'causative_agent_of', 'has_part', 'part_of', 'has_associated_morphology', 'associated_morphology_of', 'uses', 'used_by', 'has_active_ingredient', 'active_ingredient_of', 'occurs_before', 'occurs_after', 'occurs_in', 'uses_device', 'device_used_by', 'has_location'. Since the relationships in the Portuguese database have no specified relationships, the empty field is also accepted as an unknown relationship.

5.4.3 Suggest

This service checks if the user previously suggested the same concept or if it is already present in the database (see input parameters in Table 5.3). If not, the new suggestion is inserted in the database. It could be later analysed by a professional and subsequent studies and analysis could be done with these suggestions. It could be helpful for the construction of a custom vocabulary.

Table 5.3: Suggest service input parameters

Parameter	Data Type	Description	Default
suggestion	String	The suggested concept	
tuid	String	Unique identifier of a user	
language	String	Language of the suggested concept. May be 'en' or 'pt'	

5.4.4 Rating

The intent of this service is to insert the user ratings about a concept in the database. The rating is split in four different evaluations: definition, external references, relationships and general (see Table 5.4). The user only rates the information that is displayed plus a general evaluation. These ratings could be later analysed and actions could be taken to improve the information of certain concepts and to have a general overview of the user's satisfaction.

Table 5.4: Rating service input parameters

Parameter	Data Type	Description	Default
cui	String	The rated concept	
tuid	String	Unique identifier of a user	
definition	int	Rating of the definition, from 1 to 5	
externalReferences	int	Rating of the external references, from 1 to 5	
relationships	int	Rating of the relationships, from 1 to 5	
general	int	General rating of the concept details, from 1 to 5	
language	String	Language of the document where the concept was recognized	

5.5 Client Side

As a Google Chrome extension, the client side logic is developed in Javascript. Chrome extensions have two types of scripts: the content script which runs in the context of Web pages and can access and change the DOM; the background script, a single long-running script that can manage some task or state.

The extension requires a content script to be injected, as DOM manipulation is required. Other components are also injected, such as jQuery[17] (a Javascript framework) or Bootstrap (a front-end-framework for displaying modals and tooltips). To avoid interferences between Bootstrap and existing elements on the Web document, the inserted HTML always contains a parent container

[17] https://jquery.com/

with a class "health-translator" and the Bootstrap was accordingly wrapped with that class, which can be generated from the scoped-twbs project[18].

Before processing a Web document, the language is detected using the Chrome Tabs API[19]. The language is inserted in the DOM as a property of the body element. This property will be later needed when using other services, such as requesting for details of a concept.

In order to process a Web document, the text nodes are collected from the content script and sent to the background script which requests the server and resends the response to the content script again. The response contains a counter with the number of recognized concepts, which are updated on the extension badge. Text nodes can not simply be replaced by HTML. Therefore, a new node is created and inserted before the text node, which is then deleted. Click events and tooltip initialization is executed. Modal windows for concept details and rating are also appended to the document.

In order to detect DOM changes, a MutationObserver[20] is registered. This allows the detection of additions or removals of text from the Web page and consecutively support of dynamic Web pages. When text is removed from the document, the badge is updated accordingly. When text is added or changed, the initial processing is done, but only for the affected text nodes and the badge is also updated. This speeds up the process as only some text nodes are processed again, avoiding the document to be fully processed again. Changes triggered by the extension itself are ignored.

Each recognized concept has the CUI as a property of its HTML container. In order to request for more details of a given concept, a user must hover the concept and click the corresponding button inside its tooltip. By clicking on that button, the same logic is triggered. A request is sent to the background script, which returns the response from the server to the content script. Here, the details modal window is opened and the data is loaded accordingly, such as semantic types, definition, external references and relationships. If the user did not previously rate that concept, a button is available in order to rate it.

The rating is done by opening a new modal window with the parameters to rate, depending on the previously displayed information. This option is only available if the user has not previously rated the concept. Only presented information may be rated. The user can then click the button to submit the results.

Context menus are registered on the background script to support the suggestion of new concepts. These menus contain a second level that requires the user to say what language the concept refers to, so it is inserted in the corresponding database.

Regarding the two latter features (rating and suggestions), it is necessary to identify the user. This identification is provided by tuid[21], which is a unique id generator that is saved on the local storage.

[18]https://github.com/homeyer/scoped-twbs
[19]https://developer.chrome.com/extensions/tabs#method-detectLanguage
[20]https://developer.mozilla.org/en/docs/Web/API/MutationObserver
[21]https://github.com/mongoh/tuid

The settings page is based on fancy-settings[22]. It eases the settings page creation and provides a look-and-feel similar to default Chrome settings. Settings values are stored on local storage.

A major difficulty in the client side is to provide a consistent look-and-feel among all Web documents. Every Web document has its own CSS rules and Javascript code. That means that they can affect the components added by the extension. Injected HTML classes and id names are related to the extension. Custom HTML elements were added, such as <x-health-translator>, in order to avoid being affected by existent rules. However, some regular elements and classes from Bootstrap are also used, and Web documents may affect those. Although many rules have been identified and overridden, it is not plausible to do the same for every existent CSS rule. Although it is an uncommon case, there may be some aesthetic inconsistencies in different Web documents.

Initially, the content script was injected by the background script when the page was completely loaded. However, during the user testing (see section 6.3), some pages were figured out to take a long time to load scripts or images, even after the DOM is already fully loaded. Thus, a different approach was implemented on the client side. The scripts are always injected on the page before the DOM, independently of the execution mode setting. The main logic is only executed after the DOM is loaded, which automatically fires if the automatic mode setting is set. However, if the manual mode is active (the user needs to click the extension icon to process the page), it may happen that the user clicks the icon before the DOM is loaded. In that case, a flag is set up and the page is processed when the DOM is loaded. This new approach leads to significant performance improvements in cases of pages that load the DOM fast but take a long time to load other resources.

5.6 Support of New Languages

The Portuguese support is an innovative factor of the extension, as there is no knowledge of other application providing similar features for that language. However, modularity was always a concern and the adaptation of the system for new languages is a possibility. This requires preparation of a new database and customization of the server and client side.

5.6.1 Database

It is assumed that the new language to be supported is based on UMLS vocabularies, as some of its structure is used. For example, the extraction of relationships assumes the UMLS table structure. However, like in the Portuguese vocabulary, additional tables can be added or information can be adapted and inserted in the present structure. The new database should be named ‘umls_<language_code>’, where <language_code> is the ISO 639-1 code (2 letters). For example, the current databases are named ‘umls_en’ and ‘umls_pt’ for English and Portuguese, respectively.

The database should include a table with definitions, following the same structure. The name of the table is not relevant, as the module for gathering definitions will be implemented in the server

[22] https://github.com/altryne/fancy-settings

(see section 5.6.2). The tables *suggestion* and *rating* should be created with the same structure as the current ones.

5.6.2 Server

Regarding the server, two of the four services must be adapted: processing and details.

Regarding the document processing, it is necessary to implement the token processing strategy, by implementing a class that extends *ConceptProcessor*. If the database is based solely on UMLS and the same strategy is to be implemented, it should be very similar to the *EnglishProcessor* module. For example, Spanish adaptation could be very similar, as SNOMED CT is also available in Spanish. The CHV does not need to be included - the difference is that it will not display term mappings in lay and medical terminologies.

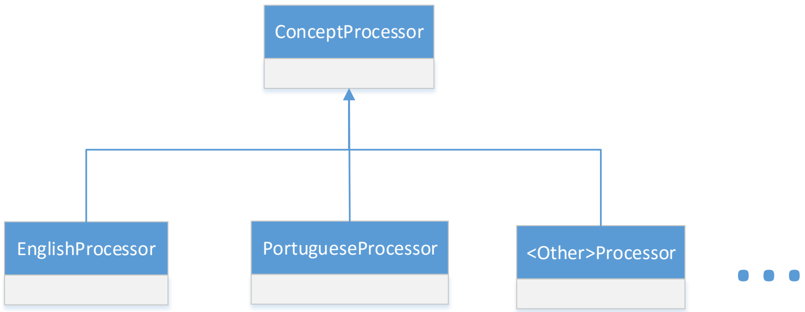


Figure 5.14: Required class to add in order to support new languages

Concerning the details of a given concept, the relationships are extracted from UMLS (the database name is assumed by the naming convention). The external references need to be added and respective methods for extraction should be added to the *ExternalReferencesExtractor* class.

Some translation must also be added in order to retrieve the tooltip content in the corresponding language if requested by the client.

5.6.3 Client

The client customization affects only the settings and translations. The logic itself is independent of the language. The new language should be included in the settings, specifically on the Language section. The new language also needs to be added on the conditional statement of the supported languages.

5.7 Limitations

Many of the system limitations derive from its dictionary based approach. As the concept recognition only implies string matching with dictionary entries, there is a lack of context associated. For example, in the sentence 'stroke, ischemic or hemorrhagic', three different concepts could be detected - 'stroke', 'ischemic' and 'hemorrhagic'. Although the two latter concepts are related with 'stroke' in the sentence, they would refer to different concepts when recognized alone, as 'ischemic' does not mean the same as 'ischemic stroke'. It even points to a different semantic type that may not be accepted by the user settings.

This lack of context can be an issue when processing pages outside of the medical context. For example, 'injection' or 'node' have a different meaning in programming. Even in health related texts, some concepts may be recognized when they should not, also referred as false positives. 'Face' is an example of a word that can be a verb or refer to a body part.

Another issue regards the provided definitions for each concept. Only a minority of the concepts present on the database have an associated definition. It may also happen that the definition is wrong, because it links to a disambiguation page or refers to a concept from another context. Disambiguation pages may be easy to detect and remove, by searching for commonly used expressions, such as 'may refer to' or 'may mean'. However, wrong definitions are not so easy to detect in a fast or automatized way.

The data quality or lack of structure may also be a similar concern. For example, in the Portuguese UMLS vocabularies, the relationship between concepts is not specified, thus making it not so relevant for the final user.

A Web page containing content in multiple languages can be an issue. The document language is first detected and all the contained nodes will be processed taking that language into consideration. Thus, it may happen for a given page that the language can not be properly detected or the most used language will be detected. In the latter case, that implies that the text nodes in different languages will not be properly processed. Detecting the language per text node is impractical, as usually they are small and not suitable for detection with enough confidence.

The extension presents the number of concepts recognized. This number represents the total number of concepts, including duplicates. However, some recognized concepts may not be displayed to the user, given the displaying rules of the Web document, e.g. some text nodes are in a hidden container with the CSS rule 'display:none' applied, which may look like a wrong count to the user.

Lastly, performance was a factor taken into consideration during the whole design and development phase. In a production and ideal environment, the server infrastructure would be scaled in order to process a large number of requests concurrently. Assuming that all the text nodes of a Web document could be processed simultaneously, the processing time of a Web document would be the same as the time needed to process the longest text node of that document, plus request delays. As text nodes are usually not very long, this seems to present a good performance. Another important aspect is that the database caches the queries and processing the same page for the second time has huge performance improvements. This means that if the extension is extensively used, more concepts would be cached and the processing would be faster.

5.8 Summary

The developed system presents a client-server architecture, given the big amount of stored data from medical vocabularies. The system aims to be fast in recognizing the medical concepts in a Web page. Thus, only definitions and terms in different terminologies are gathered on a first processing. These two features can be useful and sufficient to increase the comprehension of the concept by the user. For a deeper analysis, the user can obtain more details of it, such as related concepts or external references.

Two other features were implemented, although they are not being processed in the background. That is, a user can rate the information of a concept and suggest new medical concepts in any Web page. This data is being stored in the database. However, no action is being done on this data. User opinion will be gathered in the testing phase about these features.

Some limitations are recognized during the implementation of the system. It has been deployed on a server provided by the faculty informatics center. The server infrastructure provided is not enough to make the system available for the general public. Thus, it is intended to be a prototype and is not available in Chrome Store.

Chapter 6

Evaluation

The server was hosted in a virtual machine provided by the faculty informatics center. It runs Ubuntu 14.04 and has 4GB RAM memory. The CPU is a Intel Xeon E5-2695 v2 @ 2.40GHz CPU but is shared among other virtual machines. The evaluation had to be done in FEUP's network (or connected by VPN), as the server is only accessible in the local network.

The extension was evaluated in terms of performance, annotation coverage, quality and utility perceived by the users. The evaluation process involved different methodologies. On one hand, two tests, one for each supported language, with quantitative results and statistics about concept coverage and performance. On the other hand, a qualitative test, through a user study directed to the target audience.

In this chapter, the goals of each test, its results and conclusions are presented.

6.1 Performance

The performance of the developed extension was analysed and compared with the only comparable system performing in English, named Medical Translator. The main feature of this extension is similar with part of the developed system. It recognizes medical concepts in Web documents and provides a short definition. In case of a clinical drug, it links to the website of its creator, which presents more detailed information.

A collection of documents was gathered and classified by a Web crawler named ILSP-FC [PPT13] and made freely available in QT21 repository[1]. One hundred documents were randomly selected from that collection, processed by both extensions and the processing times were measured.

Processing times are significantly lower on Medical Translator (see Table 6.1) because the processing is done locally. For a complete measure of time for every document see Appendix B.

[1] http://qt21.metashare.ilsp.gr/repository/browse/qtlp-english-corpus-for-the-medical-domain/9b0842326bdd11e393a800155dbc0201da18a215335e4c0a9169fa0f2b680add/

The strategy to lookup for terms is similar to the one implemented by HealthTranslator, but only recognizes concepts with 3 or less words, which reduces the amount of words to look forward when compared to the threshold of 5 tokens implemented by HealthTranslator. Medical Translator also avoids the processing of several HTML elements such as headers or buttons.

Table 6.1: Comparison of processing times between Medical Translator and HealthTranslator

	Medical Translator	HealthTranslator
Average	112,90 ms	2722,69 ms
Standard Deviation	92,55 ms	2547,47 ms

The major delay of the developed solution is a result of request delays over the network. When processing a text, the server easily reaches the full CPU usage, so it is believed this is a main bottleneck for the system's performance. However, this significant difference does not represent a major limitation for the user as the results from HealthTranslator are provided in a progressive way. The Web document is typically processed from top to bottom, which means that the concepts on the part of the Web document the user is looking first are also recognized first. When the Web document is very extensive, the user may not even realize that concepts are still being recognized on the bottom part of the page. In general, users did not complain about the speed of recognizing the concepts on the conducted survey (see Section 6.3). The main complain was related to pages that take a long time to load images or scripts, but a new injecting method was implemented (referred in Section 5.5) after the user study to improve those situations.

6.2 Annotation Coverage and Quality

Due to the available resources to evaluate the annotation coverage and quality, this evaluation was split in two methodologies, one for each supported language. Regarding the English, annotations are compared with a similar extension, while for Portuguese, a manually annotated corpus was built and compared.

6.2.1 English

In English, the annotation coverage was assessed comparing the number of concepts recognized by both HealthTranslator and silver standard corpus annotated by Medical Translator of 100 documents, the same set of documents as selected in Section 6.1.

The fact that Medical Translator also uses a dictionary based approach implies the same limitations of HealthTranslator, such as the existence of false positives or false negatives. Thus, through this evaluation we assess the concept coverage but not the quality. The quality coverage was evaluated for Portuguese, as it was manually annotated by humans (see Section 6.2.2).

The annotation output from Medical Translator is not provided in a standard format and thus evaluation metrics such as precision, recall and F-measure were not gathered. A comparison of number of annotated concepts was done. It was gathered the number of total and unique recognized concepts per document. A concept may be recognized several times in a document and therefore the amount of unique concepts is always equal or less than the total concepts.

The counting of unique concepts in Medical Translator is different from HealthTranslator. The latter assumes a unique concept by its CUI. For example, if 'high blood pressure' and 'hypertension' are recognized in a document, it counts as a unique concept as they both refer to the same CUI. In Medical Translator, this comparison is done by comparing the strings, that is, 'high blood pressure' and 'hypertension' would count as two unique concepts. Even concepts in plural and singular are counted as different concepts. This means that for the same concepts recognized in the same document, it is possible that Medical Translator presents a higher number of unique concepts, compared to HealthTranslator.

It is possible to see in Figure 6.1 and Table 6.2 (the full comparison is available in Appendix B) that HealthTranslator recognizes more medical concepts, even with the default semantic types (29 out of 127). Some documents that showed to have a smaller concept count difference were analysed and it was detected that the main content of those document refers to concepts not present in HealthTranslator's default semantic types. For example, the document 7062 is about lymphovascular invasion, where most concepts are related to the semantic type 'Neoplastic Process' which is not accepted by default in HealthTranslator.

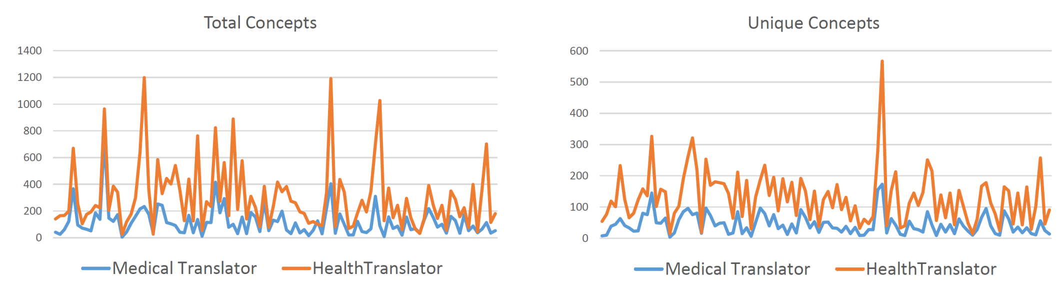


Figure 6.1: Comparison of concept coverage between Medical Translator and HealthTranslator

Table 6.2: Comparison of concept coverage between Medical Translator and HealthTranslator

	Medical Translator		HealthTranslator	
	Total Concepts	Unique Concepts	Total Concepts	Unique Concepts
Average	118,05	44,05	302,56	130,69
Standard Deviation	105,47	32,24	244,23	82,51

As HealthTranslator recognizes significantly more concepts than Medical Translator, there is a large amount of false positives. Therefore, a low precision and consequently low F-measure values are expected.

In average, HealthTranslator provides a definition in 80% of the recognized concepts. Although a definition was inserted in the database for a minority of the concepts in the database (see section 5.3.2), it is shown that a majority of the recognized concepts in this sample present a definition. The minority of definitions inserted in the database can represent a majority of the concepts recognized in the real context, as the most relevant subjects are present in Wikipedia.

Medical Translator has a vocabulary composed of 4782 concepts. From those, 2866 refer to drugs which redirect to the company website. This shows the specificity of the vocabulary and the goal of redirecting traffic to their services. Some of the concepts are composed by different alternatives, which are divided by a '/' . For example, the vocabulary contains the concepts 'asthenia/fatigue', 'asthenia' and 'fatigue/asthenia'. Thus, the same term may repeat multiple times. After filtering the duplicates, the vocabulary contains 4636 unique terms.

In order to compare the concept coverage it was checked the presence of each unique term in HealthTranslator's database. From the 4636 terms, 3447 ($\approx$74%) would be detected by HealthTranslator, independently of the semantic type. If the default settings are considered (all the semantic types related to a concept must be in the default list of 29 semantic types – see Appendix A), 2948 ($\approx$64%) concepts would be recognized.

6.2.2 Portuguese

As the existence of an annotated corpus of medical concepts in Portuguese is not known, it was decided to create one to evaluate the tool in this language.

The extension is mainly to be used by laypeople, as it provides definitions and additional information to the users. However, it may also be useful for doctors in order to understand different ways to refer to the same concept, for example in a lay terminology.

The definition of a 'medical concept' is very broad and may be interpreted differently by individuals. An evidence of this complexity is the fact that UMLS groups concepts in 129 different semantic types. Thus we asked laypeople and health professionals to annotate medical concepts in a collection of documents. Generally, laypeople should be able to perform a good quality annotation of medical concepts. Even though they might not know the definition of a concept, they should be recognizable by its knowledge and context. Moreover, about half of the 'laypeople' invited to annotate the documents are not really lay to the subject as they are currently finishing health-related courses, such as nursing or medicine.

Both groups of people (lay and professionals) were invited for the annotation task in order to build a consistent and robust annotated corpus. Ten people were invited to constitute the lay group as they have a bachelor or superior grade. Moreover, six of them are currently finishing or working in health related areas, such as nursing or medicine. The professional group is composed of five nurses and doctors. All of the invited people have sufficient skills to perform the annotation tasks in the chosen tool.

Although initially the goal was to create a collection of 60 documents, inviting professionals such as doctors or nurses to annotate the documents was a hard task and only 38 documents were successfully annotated by this group. The documents were gathered and classified by the same tool as the English collection in Section 6.1 and are also available for download[2].

[2]http://qt21.metashare.ilsp.gr/repository/browse/qtlp-portuguese-corpus-for-the-medical-domain/27c3e8aa6bdb11e3b61300155dbc02019a678f0685874a03a2aa35640f92b204/

The process was split in 2 phases, as seen in Figure 6.2. The first phase consisted in providing documents to be annotated by each person. The crawler removes boilerplate of the documents and, for each sentence, presents concepts automatically detected. These concepts were initially automatically annotated as they might help the users and reduce the time needed for the annotation task. However, users must remove them if they feel these concepts are incorrectly recognized. Annotators were given guidelines (see Appendix C) in order to provide a consistent annotation among different people. For example, there should be no nested concepts, as the most specific term should be annotated.

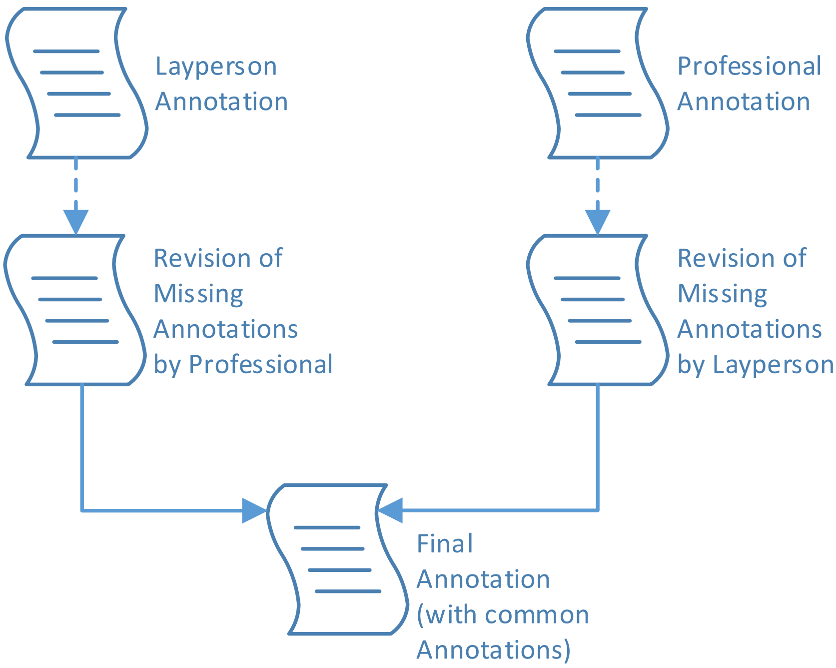


Figure 6.2: Creation of a Portuguese corpus annotated with medical concepts

The same document was independently annotated by a person in the 'laypeople' group and another person in the 'professional' group. The documents of a person do not match all the documents of another person. This task was done in egas [CLMO14], an annotating tool able to export annotations in brat standoff format [bra], also known as A1.

Later on, the terms that were annotated by the person in a group, but not by the corresponding person in the other group were sent to the latter, in order to disambiguate terms that might have been forgotten and create an agreement corpus. For example, if an annotator of the professional group identified the term 'hypertension' and the corresponding annotator in the lay group does not, this term was sent for revision to the layperson, which then decides if he agrees it is a medical concept or not.

Therefore, the gold standard corpus is the group of documents with annotations composed by the consensual concepts on both sides. Although the lay group completed both annotation phases in 60 documents, only 38 documents were finished by the professional group. Thus, this corpus is constituted of 38 documents with a match on both groups, but there are 22 documents annotated by laypeople but without a matching professional annotation.

Although it might happen that laypeople do not recognize concepts annotated by professionals on the second phase, it has the advantage of removing non consensual concepts, which may be mistakes or resultant of different interpretations of the definition of ‘medical concept’. This increases the quality of the final annotation, even if a few concepts are not annotated when they should be.

The final collection of annotations was then compared with the annotations performed by HealthTranslator in two different scenarios: with the default semantic types and with all semantic types. This allows to evaluate the precision and recall of the annotations and observe the influence of the semantic types on the final output. The comparison were done with BratUtils[3], a tool that provides metrics by comparing annotations in brat standoff format.

The annotations present an average F-measure of 26-33% depending on the selected semantic types (see Table 6.3; for the full comparison see appendix D). When compared with other tools, it presents relatively low F-measure results, as HealthTranslator is not so restricted to specific subjects, but aims the recognition of concepts in the general medical domain. For example, Neji which presents positive results when compared with similar tools, presents a F-measure of 95% for species and 85% for disorders [CMO13a]. Besides the recognition strategies implemented by other tools, the main reason for this big discrepancy is the specificity of the concepts to be recognized. While Neji is analysed in specific areas, such as disorders or genes and proteins, here is presented a general overview and analysis of the concept annotation in the medical domain, which is very broad.

Table 6.3: Annotation metrics with default and all semantic types

	Default Semantic Types			All Semantic Types		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Average	0.406	0.255	0.269	0.261	0.581	0.332
Standard Deviation	0.208	0.190	0.150	0.129	0.182	0.149

There is a variety of concepts that were consensual in some documents and not in others, depending on the annotators interpretation of ‘medical concept’. For example, concepts such as ‘morte’ (*death*), ‘hospital’ (*hospital*) or ‘saúde’ (*health*) are included in the the final annotation in some documents but not in others.

The big amount of spurious concepts (recognized by the extension but not in the manual annotations), which decreases the precision, and consequently the F-measure, may be a result of this ambiguity. It is not believed that all the spurious are false positives, as they are included in

[3]https://github.com/savkov/BratUtils

medical vocabularies and the analysed texts are mostly health-related. Most of them are eventually non consensual or insignificant concepts in the annotators point of view. Given the approach used to build the corpus, the annotations are not as complete as they could be, as only consensual annotations are included in the final annotation.

The calculation of the partial matches may also negatively affect the metrics. In the used tool for comparison, the partial match is only positive when one of the annotations has the same end index and a different start index. This means that two concepts are only a partial match if the end index is the same. If a concept is composed of two words, and only the first word is recognized by the extension, it is not counted as a partial match because only the start index coincides.

Another factor that can affect the metrics is when the same concept is recognized several times in the same document. For example, if the concept 'pain' is a manual annotation that spans multiple times and is not recognized by HealthTranslator, it will count as many missing annotations and consequently result in a lower recall and F-measure.

It is clearly observable in Table 6.3 that an increase of semantic types leads to an increase of recall and a decrease of precision, as expected. This means that more concepts from the manual annotation are recognized, but it also includes more spurious annotations. The selection of semantic types is a trade-off between precision and recall. Thus, it needs to be carefully adapted for the user needs or preferences.

For each document, it was also analysed the number of unique concepts recognized by HealthTranslator and the amount of definitions provided among those, as seen in Table 6.4. It is observable a significant difference of unique concepts recognized by accepting more semantic types. The percentage of definitions provided remains high even when more concepts are recognized. This provides a satisfactory result of the insertion of definitions in the database referred in section 5.3.2. Although a minority of definitions were inserted, a majority of the recognized concepts have a related definition, as also shown for English.

Table 6.4: Amount of unique concepts and definitions provided by HealthTranslator

	Default Semantic Types		All Semantic Types	
	Unique Concepts	Definitions(%)	Unique Concepts	Definitions(%)
Average	22.263	75.00	108.079	71.55
Standard Deviation	19.941	11.13	90.975	6.42

6.3 Utility

Apart from the concepts coverage and performance, it is important to understand the opinion from the final users of the system. For that reason, a user study was conducted. The study methodology and its results are now presented.

6.3.1 User Study

The user study contains five main phases: assessment of the user health literacy, filling of a questionnaire without looking for information online, use case of the extension, refilling the previous questionnaire with searches online and answering a final survey to provide feedback about the extension.

The first step of the study was the evaluation of the user's health literacy. METER, an instrument to assess health literacy, previously translated and validated to Portuguese [PSS+14], was used for this purpose. This tool consists in detecting correct medical words from a list with incorrect words and defines a threshold to rate a person as adequate in health literacy. However, word recognition tools such as METER have limitations, as they only operate in one of the aspects of health literacy - the ability to recognize medical jargon. Eight users (40%) were identified as non-adequate in health-literacy. However, it was not identified any significant correlation with health-literacy and other variables, such as the previous knowledge on the asked subjects, the language used on the queries or the usage of the extension.

After this evaluation, users were asked to fill the questionnaire without looking for information online, in order to check what information they already knew beforehand. Users were instructed to not guess the answers. Most people did not know the majority of the answers and sometimes still risked by answering incorrectly.

Thereafter, a little use case and explanation of the extension was provided. Later on, the same questionnaire filled before is provided and they were asked to look online for the information without any time restrictions, even for the questions they answered in the previous task.

The questionnaire is composed of two different sections, one related to asthma and another to nutrition (see Appendix E). Each section contains questions from knowledge questionnaires (asthma knowledge questionnaire [HBG+03] and nutrition knowledge questionnaire [PW99]) which were translated to Portuguese and validated [Par13, dS09]. The correct answers are also provided. Those questionnaires are extensive and a filtering needed to be done in order to keep each section realizable in an acceptable amount of time. Regarding the asthma section, the 5 questions with less correct answers were chosen to make it harder for the users to reach the answers and look for more information. In regard to the nutrition section, a subset of the diseases and problems part was selected.

Regarding the usage of the system, a within-subject design was used [Kel09]. That is, each user performed the test with and without the developed system, in order to be able to compare both situations and provide a feedback about the extension. A user should not perform the same search task twice as it would be biased. Therefore, a rotation on the tasks performed with the extension was used. Half of the users performed the asthma section with extension and the nutrition section without extension and vice-versa for the remaining half.

After the questionnaire, users were requested to answer to a form (see Appendix F) to provide feedback about the usage of the extension on the previous search tasks.

The amount of time needed per user is a bit extensive, so the users used the extension with the default settings (see Appendix A). The users only had contact with the extension's settings except after the search tasks, while discussing features or suggesting improvements.

6.3.2 Participants

Twenty persons participated in the study, with ages between 21 and 35 years old. 95% of the users are currently finishing or completed a Bachelor or Master's degree, while the remaining hold PhD degree. Most of the users have studies in Computer Science, while others study Multimedia and Design. As one of the extension differentiating features is the Portuguese language support, the test was conducted with Portuguese speakers, but all of them were also comfortable with English. Users were given the option to choose between Portuguese or English in their tasks.

From the 20 respondents, 45% affirmed they rarely search for health-related information online, while 55% say they do it sometimes. As seen in Figure 6.3, it is common for laypeople facing difficulties while reading health-related content. The fact that most of the questionnaires include a few wrong answers even with online search proves this difficulty.

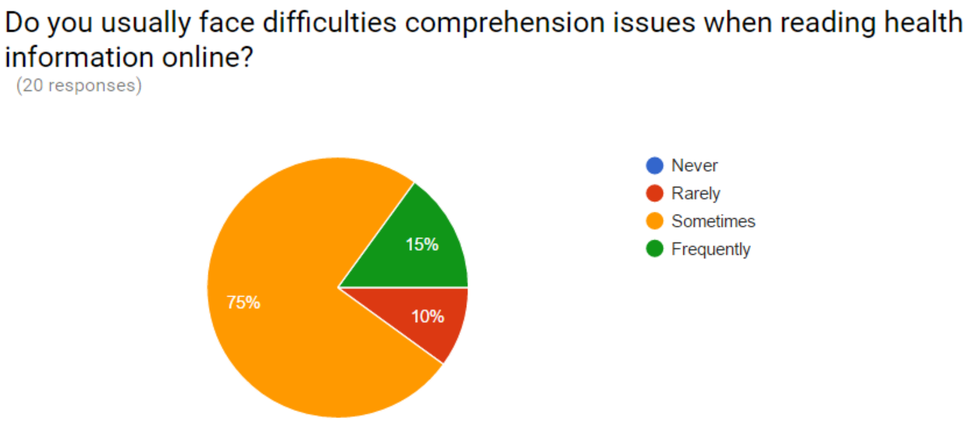


Figure 6.3: Difficulties faced while reading health-related information online reported by laypeople

6.3.3 Results

95% of the users stated they would eventually use the extension in their daily life, although the manual execution mode would be preferred and only used when searching for health-related information. The remaining 5% (1 user) said the help provided by the extension is not valuable enough to make him use it.

The next sections describe the feedback provided by the users and the observed behaviours, which might explain the benefits and disadvantages of the extension.

6.3.3.1 Performance and User Behaviours

Eleven users preferred to use Portuguese as the main search language. From those, six used exclusively that language. The remaining users searched mainly in English, but also used Portuguese in some cases, for example when they were not able to translate some terms. It was found that English searches usually end up in more updated information. As an example, for some questions in the questionnaire, searches in Portuguese refer one fact, while in English it pointed to recent studies disproving the same fact. Therefore, it was not given so much importance to the correctness of the answers as some questions are controversial and medicine is not a static science.

The time searching for information is one of the variables that one might think the extension can affect. The factual measures do not show a relevant improvement of time with the usage of extension, as can be seen in Table 6.5. The full table with times can be seen in Appendix G. However, 70% of the participants responded that the extension may shorten the time they take to find what they want, while 30% refer that it does not interfere. It shows that the user impression is not accurate or the users that actually responded so actually were faster, although the average values had a different behaviour. It can also mean that users agree that it can be faster for another type of questions.

Table 6.5: Time comparison of task completion with and without extension

	Asthma					Nutrition				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
	With Extension									
AVG	02:48	03:46	01:37	02:11	02:18	02:22	02:02	03:16	02:29	02:55
STDEV	01:43	02:00	00:53	01:15	02:15	01:27	01:01	01:29	00:55	01:46
	Without Extension									
AVG	01:31	02:55	01:48	02:13	01:20	02:31	02:23	05:08	02:07	03:08
STDEV	00:50	01:40	01:40	00:46	00:52	01:31	01:08	02:47	01:35	02:11

The times are in the format mm:ss. AVG = Average; STDEV = Standard Deviation

The extension's goal is to increase the readability and comprehension of health-related content, so the extension should not change the way users search for the information. An increased time on the searches is not necessarily a negative factor. It may eventually encourage the users to know more about medical concepts and search around for additional information.

It was noticed that gathering more details of a concept is not a very common practice. A user explores the details of around two concepts in average, as can be seen in Table 6.6. One of the suspected reasons is that the information provided from the medical vocabularies is not useful from a layperson perspective.

Another possible reason is that the loading of the modal with additional information of a concept is too long leading to people avoiding to click on it. When asked of how fast the extension loads the details of a concept, the opinions were a bit scattered, as can be seen in Figure 6.4.

The performance of gathering details from a concept can be largely improved. The main reason of its slowness is the loading and scraping of Websites to present as external references.

Table 6.6: Average and Standard Deviation of the quantity of concept details explored by users for each subject in the questionnaire

	Asthma	Nutrition
Average	2,2	2,3
Standard Deviation	1,42	1,54

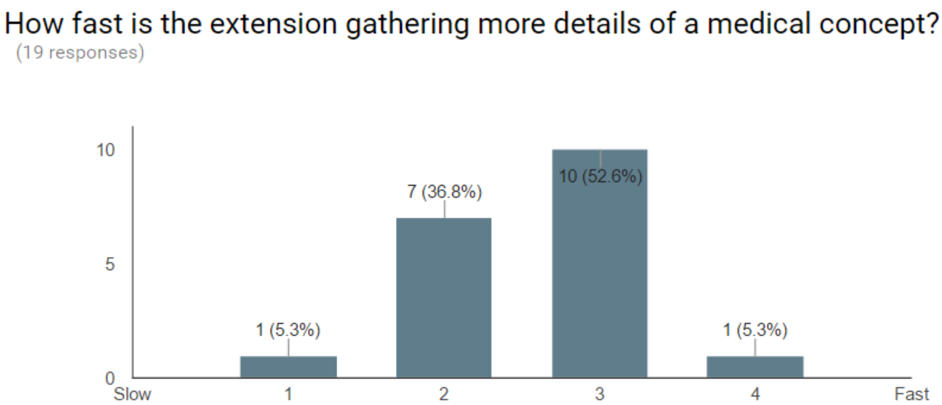


Figure 6.4: Opinion about the loading time of concept details

Furthermore, that information can be loaded in a different request, while the remaining information is presented faster, as it only requires querying the database.

75% of the users reported that the extension does not make them lose the focus on their task, whereas the remaining 25% think it does affect sometimes. Although some users commented that the concept highlighting is beneficial as it directs the eyes for the important content, others say that some concepts are noisy and nuisance.

Regarding the recognition of incorrect medical concepts, the opinions are very spread (see Figure 6.5). Most of the users acknowledges they were not aware of these. A possible explanation is that the focus goes to the concepts that matter to the user, as their eyes focus what they are looking for and partially ignore all the rest. Although 20% said they did not find any wrong concept, it is almost certain that in their search there were cases of wrong detected concepts given the context, also known as 'false positives'. For that reason, the result can be agglomerated, which gives a 65% of users not aware of false positives. 25% claim they found a few wrong concepts and the remaining 10% state they found many false positives.

Concerning the definitions provided by the extension, 55% of the users say that they found a few missing or wrong definitions. The remaining 45% refer that none was found. Once again, it is almost certain that concepts without definitions were recognized during their search. From the user behaviours, it is detected that users hover mostly on more popular concepts that help them answering the questionnaire. This kind of concepts usually provide a definition, while others not so common or ambiguous usually fail to provide a definition. Moreover, it is possible to filter concepts without definition in the settings.

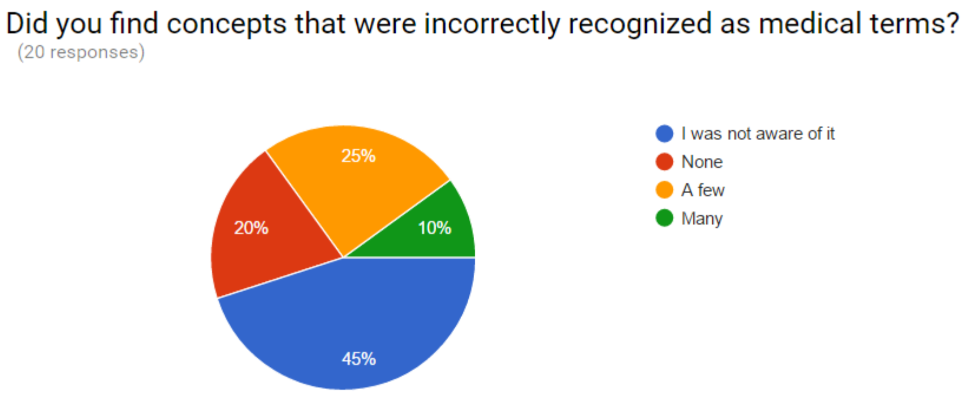


Figure 6.5: Amount of false positives noticed by the users during their searches

When asked if the extension eases the comprehension of the information found online, 80% agree and the remaining 20% affirm it does not affect.

6.3.3.2 Features Feedback

Users were asked to rate the usefulness of the extension's features (see Figure 6.6). The definition is clearly the most consensual one, where most people think it is 'Very useful' and some 'Useful'. After that, providing concepts in different terminologies is the most appreciated feature of the users, but some users already find it as 'Not useful'. Therefore, it is possible to conclude that the features presented while hovering the concept are the most valuable for the user, which shows another reason for the users not bothering to gather concept details so regularly. Lastly, users have shown to find related concepts slightly more useful than external references.

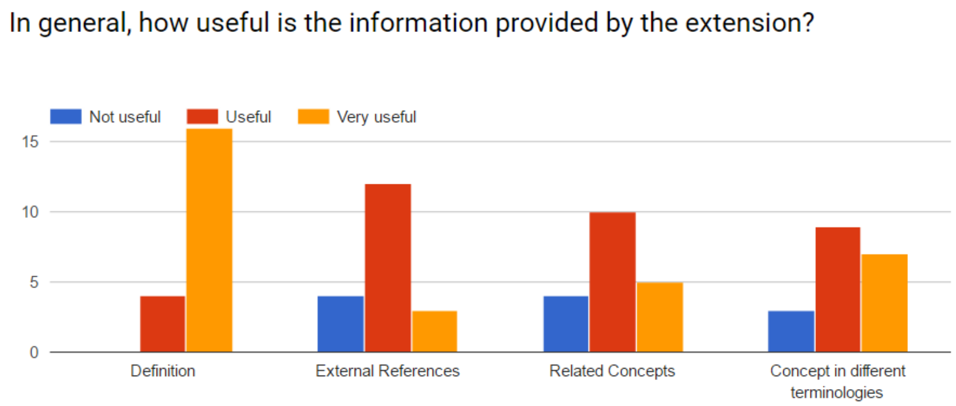


Figure 6.6: Rating of the extension's features by its usefulness

An opinion about the two additional developed features was gathered (see Figures 6.7 and 6.8). Both functionalities got a large acceptance and were found to be relevant. Suggesting new medical concepts was voted as relevant by 85% of the users, while rating the information content was said to be relevant by 90%. This shows that future developments and improvements may be done regarding those features. Although considered relevant, some users referred that they would not bother on performing any of those tasks. That is, although they enjoy and think it is an important functionality for the system, they would probably not lose time suggesting new medical concepts or rating concept information.

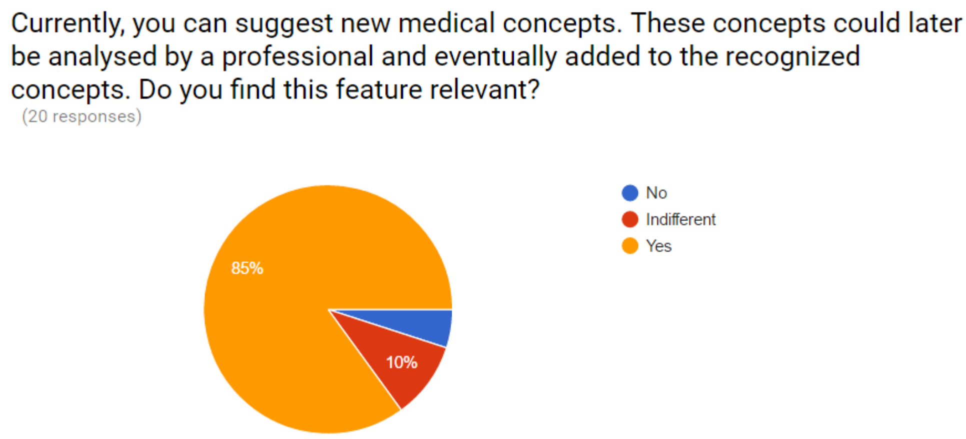


Figure 6.7: Opinion of users regarding the suggesting feature

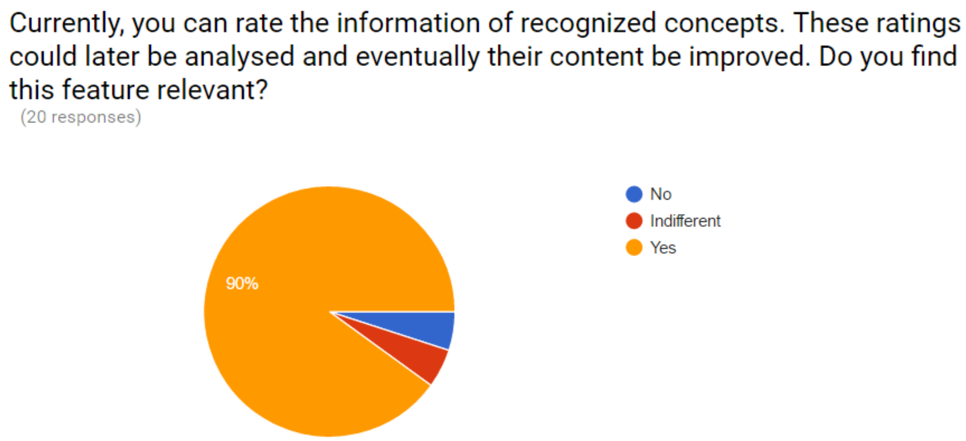


Figure 6.8: Opinion of users regarding the information rating feature

6.3.3.3 Usability

Most users find the extension user-friendly, with 95% scoring positively and the majority voting the maximum value of the scale, which means they find it easy to use and understand (see Figure 6.9).

When asked if the extension is intrusive, opinions are more disperse, with votes in all values of the scale (see Figure 6.10). Aesthetically, while for most of the users it is acceptable and even

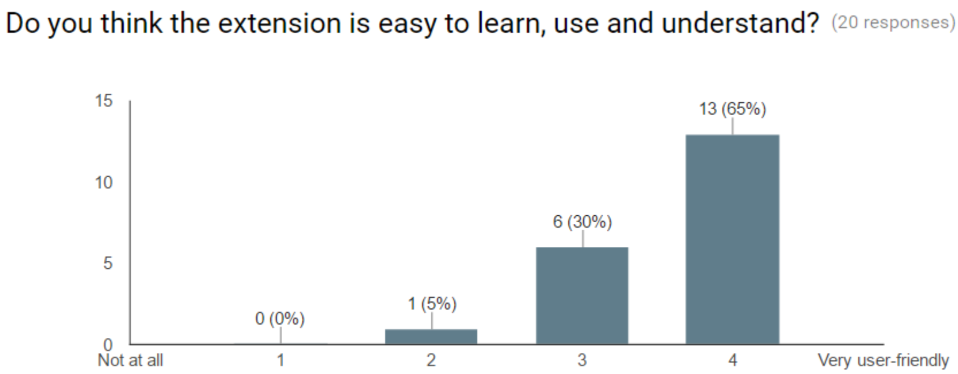


Figure 6.9: Rating of the extension's user-friendliness

helpful to have medical concepts highlighted, some others consider it noisy, specially on false positive cases. Another aspect users might find intrusive is the opening of tooltips on hovering concepts while reading the content on the page when they do not mean to do it. Also, the extension was executed in automatic mode during the tests, which users might think to be intrusive in a real context.

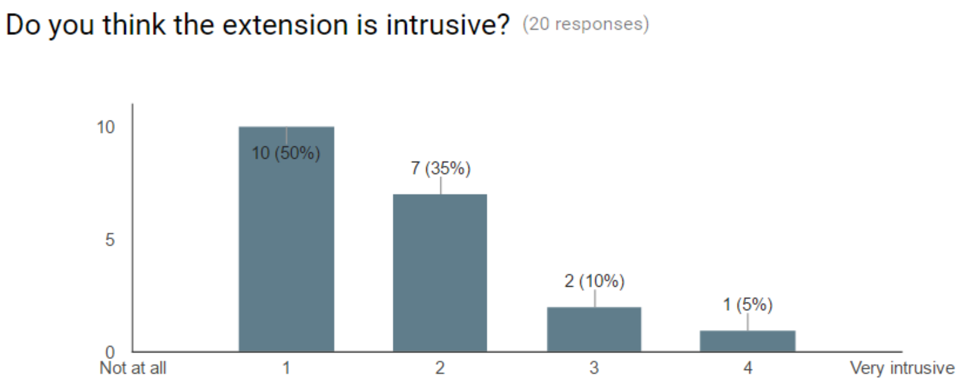


Figure 6.10: Rating of the extension's intrusiveness

The majority of the users are happy with the extension's look-and-feel (see Figure 6.11). Aesthetically, the look is mostly provided by Bootstrap default styling, which provides a familiar, consistent and friendly look for the users. Even so, 10% of the users rated it negatively, showing that there is still room for improvement.

Some users commented that the highlighting of the concepts might be a bit visually too strong. Some suggest a different type of highlight, instead of setting the background color. Another complain is related to the confusion of the highlight with the default 'Find' option of the browser, although the background color can be changed in the extension's settings.

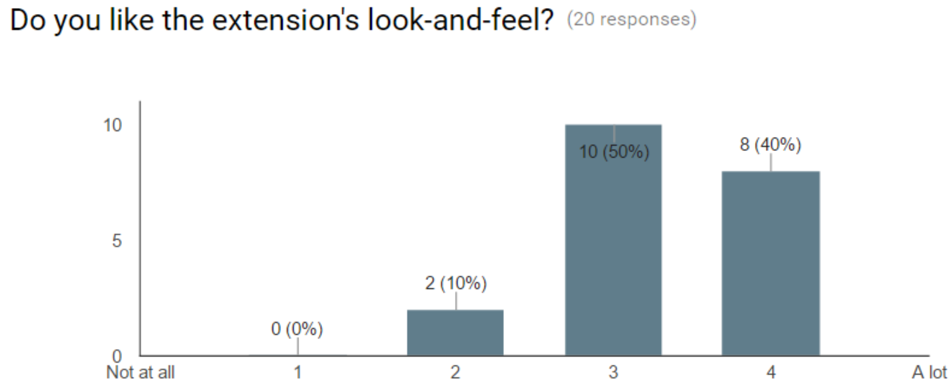


Figure 6.11: Rating of the extension's look-and-feel

6.3.4 Study Limitations

Although the majority of the users agree the extension is helpful, it was stated by some that the study probably did not present the most relevant use cases for it. Asthma and nutrition are medical subjects that are common for the majority of laypeople and even though the questions were not easy and people did not know most of the answers, they have heard about those concepts before. Users refer that the extension should be more helpful in cases where the terms are completely unknown. Regarding these subjects, users mostly ended up in Web pages oriented for laypeople and the extension should be more helpful in documents based on medical terminology in order to increase their readability.

Another variable that might not have been tested is the type of search being done by the user. Both tasks ask the user to answer to very specific questions. Users faced some difficulties to answer the questions in the questionnaire. For example in the asthma section, the correct answer is false for all of the questions. For some of them, simply there is not much information online as it is a very specific question and not a truth statement. Thus, the users needed to answer by excluding options when they did not find a contradictory statement. One might think that the extension might be more helpful when users want to know more about some medical subject or concept, but do not search for a specific answer to a question, which resembles many of the searches laypeople do in their daily lives.

6.4 Summary

The first test compared the developed extension with a comparable system (Medical Translator), which also recognizes medical concepts and provides definitions. One hundred documents were compared, relatively to the concept coverage and performance. HealthTranslator showed a good concept coverage, as the database is considerably bigger than Medical Translator. Although it is slower, it can be seen as a trade-off as it is based on a much larger vocabulary and extra features are provided. This worse performance was not a subject of complaint by the users, as the concepts are identified gradually on the Web document.

The coverage for Portuguese showed average F-measure values of around 27% and 33% for default semantic types and all semantic, types respectively. There are some limitations on the calculations of the metrics and it is believed that the results could be eventually higher. However, it is expectable that the results are lower when compared with other tools, because of the ambiguity and wideness of the medical domain. The vocabularies used for Portuguese are also more reduced and not as broad as the English one.

The user study showed a big acceptance from users in general. The feedback was constructive and pointed to things to fix, as well as future improvements (see Section 7.2). A confirmation of this positive acceptance is the fact that 95% users stated they would eventually use this extension in their online health searches in daily life.

Chapter 7

Conclusions and Future Work

This chapter presents the goals achieved in this dissertation and its main contributions. The main conclusions are discussed and future improvements and work to be done are also presented.

7.1 Conclusions

The planned features as presented in Chapter 4 were all successfully implemented. The extension was subject a thorough evaluation in three aspects: performance, annotation coverage and quality and the utility perceived by the users.

Regarding the performance, the extension performs gradually and in an acceptable amount of time and thus was not an issue raised by users. The time needed to process a Web document is proportional to its size, as more text nodes need to be processed.

The extension recognizes a good amount of concepts, covering a broad range of categories. This efficiency in terms of coverage comes from the information sources of information, specially UMLS. The used vocabularies are a fundamental information source and are the basis of the quality of the platform. A similar extension performing in English is based on a vocabulary of less than 5000 terms, while HealthTranslator is based on a vocabulary of more than 1.400.000 terms, which map to more than 370.000 unique concepts, for the same language.

Comparatively with that extension, the developed platform differentiates on three main points: provides additional information, such as external references and related concepts; supports Portuguese language; performs in dynamic Web pages.

Although presenting a good concept coverage and acceptable performance, other factors affect the usage of the extension and cannot be easily assessed with quantitative measures, such as the quality of the information, the usability of the system or the way the information is displayed. That brings the urge to a user study and an analysis of their behaviours and usage of the extension. The users were asked to search for health related information in order to answer to specific questions about asthma and nutrition.

The user study showed a good general acceptance and appreciation of the extension. Although having its limitations, such as annotation of false positives or missing information, the extension helps the users to focus on the main medical content of the document. The definition is the most used feature as it is short and fast to access. The details of concepts were not so much used, and did not seem to be so relevant for the needed tasks.

The user study design itself has some limitations and may not provide enough information about the extension usage. First, the sample is small as each test takes a considerable amount of time to complete. Second, the type of search may not simulate a realistic usage of the system, as users usually do not intend to answer specific questions, but know more about some subject, and the extension might be more useful for those cases. Third, both subjects questioned (asthma and nutrition) are somehow of common sense, as most people are familiarized with most of the concepts, even if they do not know the correct answers. These were chosen as they were available as knowledge questionnaires, validated and with answers. The extension might be more helpful for cases with less familiar medical concepts.

Through all the evaluation done, the results are generally positive and support a positive answer to the research question, showing that the developed solution can effectively facilitate the search of health related content on the Web. Several factors reinforce this hypothesis: 80% of the users agree that the extension eases the comprehension of the information; all the users agree about the usefulness of providing a definition of a concept; other features are also found useful by the majority; 75% of the users state the extension does not make him lose the focus on his search task. The positive overview of the extension is also verified by the large percentage of the users (95%) stating they would eventually use the extension in their daily lives.

One of the challenges of making this system available to the public relies on the necessity of a server infrastructure. The platform needs better and eventually concurrent machines in order to be scalable for a real scenario usage.

As the extension is not available in the Chrome Store, it is made freely available for download[1]. As the server is hosted in a machine restricted to the network of Faculdade de Engenharia da Universidade do Porto, the user can also download the databases and host the server, following the guidelines in that Website.

This dissertation also contributes to the academical community with a corpus of 38 Portuguese documents annotated with medical concepts. The construction of this corpus may be continued in the future and may be useful for later work in health informatics area. Although the annotation of the documents was difficult because of the ambiguous and broad definition of ‘medical concept’, the corpus is resultant of the agreement of the annotations of a layperson and a health professional per document. Even though this sample is still small, approaches such as machine learning require the existence of annotated corpus in order to be trained. It can also be used as a tool of evaluation, as it was used in this dissertation.

[1] https://hugosousa.github.io/HealthTranslatorClient/

Furthermore, an academical article with the contribution of this dissertation is currently being prepared and will be submitted to the Special Issue on Biomedical Information Retrieval of the Journal of the Association for Information Science and Technology (JASIST).

7.2 Future Work

The future work is mostly resultant from the feedback given by users. Some ideas are not trivially implemented but can guide developers to features that users would value the most and encourage future developments.

The biggest shortcoming of the extension, as referred by users, is one of the already known limitations of the system: the lack of context. In order to avoid this issue, more complex approaches would need to be used such as machine learning methods. Users state that it would be preferred to have only terms related to the user search highlighted or, at least, presented with a different highlight. A dynamic change of the semantic types filtering depending on the query could also be a valuable feature.

A related complaint is that the processing in non health-related pages causes a high amount of false positives. One possible solution is to avoid the automatic processing on non health related documents when the automatic mode is activated. This could also be done with the aid of user feedback, that could suggest when a page should or should not be processed. Although Medical Translator is also vocabulary based, it has some little tricks to avoid translating non health related documents. A page is only processed if: the document is above a threshold of minimum words; the number of medical concepts comparatively to the number of words in the text is superior to a given threshold (similar to a tf-idf); the number of unique concepts is bigger than a set value.

As the definitions are the most used resource, users also want more definitions, as finding missing definitions is relatively common. There is also a concern about the credibility of the information sources of the definitions. It also happens that users end up in Wikipedia pages and look for the definition of the concept of that document, which provides the same definition as it is also from Wikipedia and thus is not useful.

The related concepts could also be more useful from a layperson point of view. For example, when searching for Vitamin A, one would like to know immediately what the excess or lack of this vitamin would produce.

Other suggestions are more related to the usability of the system. Some suggestions refer that the related concepts should be able to be copy/pasted or link to some other source of information. Also the loading of the concept details should be less intrusive and the user should keep reading the document while the information is loading. Users request more feedback from the document processing, which is currently presented on the extension icon.

Users also suggest an improvement of the tool to something bigger and more sophisticated, such as embedding a search feature on the extension itself to look for more information, or have a discussing forum about concepts in order to improve the information provided by the extension. It could also be integrated with other tools from previous works. As an example, it could be merged with HealthSuggestions [CTL16], a tool that provides query suggestions on search engines, in order to redirect users to more reliable sources of information. The extension could be more useful with this combination, because suggestions in medico-scientific terminology increase the probability of retrieving documents with a more technical terminology.

References

[AL10] A. R. Aronson and F. M. Lang. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*, 17(3):229–236, 2010.

[Aro01] A. R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp*, pages 17–21, 2001.

[BC00] D. M. Baorto and J. J. Cimino. An "infobutton" for enabling patients to interpret on-line Pap smear reports. *Proc AMIA Symp*, pages 47–50, 2000.

[Bet09] Bethesda. *UMLS® Reference Manual [Internet].* National Library of Medicine (US), September 2009.

[BKL09] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. "O'Reilly Media, Inc.", 2009.

[bra] brat standoff format. http://brat.nlplab.org/standoff.html. Accessed June 2016.

[BSB07] M. J. Ball, C. Smith, and R. S. Bakalar. Personal health records: empowering consumers. *J Healthc Inf Manag*, 21(1):76–86, 2007.

[Cam13] David Emmanuel Marques Campos. *Mining biomedical information from scientific literature*. PhD thesis, Universidade de Aveiro, 2013.

[CIDC+13] D. C. Comeau, R. Islamaj Do?an, P. Ciccarese, K. B. Cohen, M. Krallinger, F. Leitner, Z. Lu, Y. Peng, F. Rinaldi, M. Torii, A. Valencia, K. Verspoor, T. C. Wiegers, C. H. Wu, and W. J. Wilbur. BioC: a minimalist approach to interoperability for biomedical text processing. *Database (Oxford)*, 2013:bat064, 2013.

[cle] The CLEF Initiative: Conference and Labs of the Evaluation Form. http://www.clef-initiative.eu/. Accessed June 2016.

[CLMO14] D. Campos, J. Lourenco, S. Matos, and J. L. Oliveira. Egas: a collaborative and interactive document curation platform. *Database (Oxford)*, 2014, 2014.

[CMO13a] D. Campos, S. Matos, and J. L. Oliveira. A modular framework for biomedical concept recognition. *BMC Bioinformatics*, 14:281, 2013.

[CMO13b] D. Campos, S. Matos, and J. L. Oliveira. Gimli: open source and high-performance biomedical name recognition. *BMC Bioinformatics*, 14:54, 2013.

[coc] Cocoa: Compact cover annotator for biological noun phrases. http://npjoint.com/. Accessed February 2016.

[Com14] European Commission. European citizens' digital health literacy. Technical report, European Commission, 2014.

[con] CoNLL: the conference of SIGNLL. http://ifarm.nl/signll/conll/.

[CR14] National Journal Clara Ritger. Wikipedia is a massively popular (yet untested) doctor. http://www.nextgov.com/health/2014/02/wikipedia-massively-popular-yet-untested-doctor/79154/, February 2014. Accessed: June 2016.

[CS04] William W. Cohen and Sunita Sarawagi. Exploiting Dictionaries in Named Entity Extraction: Combining semi-Markov Extraction Processes and Data Integration Methods. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 89–98, New York, NY, USA, 2004. ACM.

[CTL16] Tiago Fernandes Carla Teixeira Lopes. Health Suggestions: A Chrome Extension to Help Laypersons Search for Health Information. In *CLEF*, 2016.

[DFCC+13] G. Del Fiol, C. Curtis, J. J. Cimino, A. Iskander, A. S. Kalluri, X. Jing, N. C. Hulse, J. Long, C. L. Overby, C. Schardt, and D. M. Douglas. Disseminating context-specific access to online knowledge resources within electronic health record systems. *Stud Health Technol Inform*, 192:672–676, 2013.

[DHZT11] K. M. Doing-Harris and Q. Zeng-Treitler. Computer-assisted update of a consumer health vocabulary through mining of social network data. *J. Med. Internet Res.*, 13(2):e37, 2011.

[DMP+04] George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. The Automatic Content Extraction (ACE) Program: Tasks, Data, and Evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC-2004)*, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA). ACL Anthology Identifier: L04-1011.

[dS09] Juliana Almeida de Souza. Conhecimentos nutricionais: reprodução e validação do questionário. Master's thesis, Faculdade de Medicina da Universidade do Porto, 2009.

[DSX+08] M. Dai, N. Shah, W. Xuan, M. Musen, S. Watson, B. Athey, and F. Meng. An efficient solution for mapping free text to ontology terms. In *AMIA Summit on Translational Bioinformatics*, 2008.

[Elh06] N. Elhadad. Comprehending technical texts: predicting and defining unfamiliar terms. *AMIA Annu Symp Proc*, pages 239–243, 2006.

[FD13] Susannah Fox and Maeve Duggan. Health Online 2013. Pew Internet & American Life Project Washington, DC, January 2013.

[Fou] The Apache Software Foundation. Apache OpenNLP. https://opennlp.apache.org/. Accessed June 2016.

[FTC11] Liliana Ferreira, António JS Teixeira, and Joao Paulo Silva Cunha. *Medical information extraction in European Portuguese*. PhD thesis, Universidade de Aveiro, 2011.

[Gei08] A. Geissbuhler. Access to health information: a key for better health in the knowledge society. *Yearb Med Inform*, pages 20–21, 2008.

[GS96] Ralph Grishman and Beth Sundheim. Message Understanding Conference-6: A Brief History. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*, COLING '96, pages 466–471, Stroudsburg, PA, USA, 1996. Association for Computational Linguistics.

[HBG+03] J. Ho, B. G. Bender, L. A. Gavin, S. L. O'Connor, M. Z. Wamboldt, and F. S. Wamboldt. Relations among asthma knowledge, treatment adherence, and outcome. *J. Allergy Clin. Immunol.*, 111(3):498–502, Mar 2003.

[IMS14] IMS Institute for Healthcare Informatics. Engaging patients through social media. Technical report, IMS Institute for Healthcare Informatics, 2014.

[Inc] Iodine Inc. Medical Translator. http://www.iodine.com/translate. Accessed June 2016.

[jdb] JDBM. http://jdbm.sourceforge.net/. Accessed June 2016.

[JSM09] C. Jonquet, N. H. Shah, and M. A. Musen. The Open Biomedical Annotator. *Summit on Translat Bioinforma*, 2009:56–60, 2009.

[KCZT10] S. Kandula, D. Curtis, and Q. Zeng-Treitler. A semantic and syntactic text simplification tool for health content. *AMIA Annu Symp Proc*, 2010:366–370, 2010.

[Kel09] Diane Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Found. Trends Inf. Retr.*, 3(1—2), Jan 2009.

[KLS+08] A. Keselman, R. Logan, C. A. Smith, G. Leroy, and Q. Zeng-Treitler. Developing informatics tools and strategies for consumer-centered health communication. *J Am Med Inform Assoc*, 15(4):473–483, 2008.

[KM06] Anett Kralisch and Thomas Mandl. Barriers to information access across languages on the internet: Network and language effects. In *HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, volume 3. IEEE, 2006.

[KOT+04] Jin-Dong Kim, Tomoko Ohta, Yoshimasa Tsuruoka, Yuka Tateisi, and Nigel Collier. Introduction to the Bio-entity Recognition Task at JNLPBA. In *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications*, JNLPBA '04, pages 70–75, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[KSD+08] A. Keselman, C. A. Smith, G. Divita, H. Kim, A. C. Browne, G. Leroy, and Q. Zeng-Treitler. Consumer health concepts that do not map to the UMLS: where do they fit? *J Am Med Inform Assoc*, 15(4):496–505, 2008.

[LBBH15] K. M. Livingston, M. Bada, W. A. Baumgartner, and L. E. Hunter. KaBOB: ontology-based semantic integration of biomedical databases. *BMC Bioinformatics*, 16:126, 2015.

[LCF05] Deborah Lewis, Betty L. Chang, and Charles P. Friedman. Consumer Health Informatics. In *Consumer Health Informatics: Informing Consumers and Improving Healthcare*, Health Informatics. Springer New York, 2005.

[LR14] Carla Teixeira Lopes and Cristina Ribeiro. Identification and Classification of Health Queries: Co-Occurrences vs. Domain-Specific Terminologies. In *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, volume 9, chapter 4, pages 55–71. IGI Publishing, 2014.

[MH13] D. L. MacLean and J. Heer. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. *J Am Med Inform Assoc*, 20(6):1120–1127, 2013.

[MLB00] Naomi Miller, Eve-Marie Lacroix, and JE Backus. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bull Med Libr Assoc*, 88(1):11–7, 2000.

[MSB+14] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP Natural Language Processing Toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.

[MSCLA09] Mónica Marrero, Sonia Sánchez-Cuadrado, Jorge Morato Lara, and George Andreadakis. Evaluation of named entity extraction systems. *Advances in Computational Linguistics, Research in Computing Science*, 41:47–58, 2009.

[NCMO13] T. Nunes, D. Campos, S. Matos, and J. L. Oliveira. BeCAS: biomedical concept recognition services and visualization. *Bioinformatics*, 29(15):1915–1916, Aug 2013.

[NM04] F. B. Nardon and L. A. Moura. Knowledge sharing and information integration in healthcare using ontologies and deductive databases. *Stud Health Technol Inform*, 107(Pt 1):62–66, 2004.

[NS07] David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26, 2007.

[Org13] World Health Organization. Health literacy: The solid facts. Technical report, World Health Organization (WHO), 2013.

[OS06] Nelly Oudshoorn and Andre Somers. Constructing the digital patient: Patient organizations and the development of health websites. *Information, Community & Society*, 9(5):657–675, 2006.

[oSN] National Institute of Standards and Technology (NIST). Text Analysis Conference (TAC). http://www.nist.gov/tac. Accessed June 2016.

[Par13] Maria Dulce Martins Parente. A CRIANÇA COM ASMA – o conhecimento dos pais e adolescentes sobre a doença. Master's thesis, Escola Superior de Saúde do Instituto Politécnico de Viana do Castelo, 2013.

[Pea15] Robert Pearl. Are You A Patient Or A Healthcare Consumer? http://www.forbes.com/sites/robertpearl/2015/10/15/are-you-a-patient-or-a-health-care-consumer-why-it-matters/#39ae778a5c3a, October 2015. Accessed June 2016.

[PMS+01] T. B. Patrick, H. K. Monga, M. E. Sievert, J. Houston Hall, and D. R. Longo. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. *J. Med. Internet Res.*, 3(3):E24, 2001.

[POJ+09] E. Pafilis, S. I. O'Donoghue, L. J. Jensen, H. Horn, M. Kuhn, N. P. Brown, and R. Schneider. Reflect: augmented browsing for the life scientist. *Nat. Biotechnol.*, 27(6):508–510, Jun 2009.

[Pol11] The Harris Poll®. The Growing Influence and Use Of Health Care Information Obtained Online. http://www.theharrispoll.com/health-and-life/The_Growing_Influence_and_Use_Of_Health_Care_Information_Obtained_Online_.html, September 2011. Accessed June 2016.

[Por80] Martin F Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[PPT13] Vassilis Papavassiliou, Prokopis Prokopidis, and Gregor Thurmair. A modular open-source focused crawler for mining monolingual and bilingual corpora from the web. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, pages 43–51, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[PRHB+13] B. Polepalli Ramesh, T. Houston, C. Brandt, H. Fang, and H. Yu. Improving patients' electronic health record comprehension with NoteAid. *Stud Health Technol Inform*, 192:714–718, 2013.

[PSS+14] D. Paiva, S. Silva, M. Severo, P. Ferreira, O. Santos, N. Lunet, and A. Azevedo. Cross-cultural adaptation and validation of the health literacy assessment tool METER in the Portuguese adult population. *Patient Educ Couns*, 97(2):269–275, Nov 2014.

[PW99] K. Parmenter and J. Wardle. Development of a general nutrition knowledge questionnaire for adults. *Eur J Clin Nutr*, 53(4):298–308, Apr 1999.

[Que12] Health Consumers Queensland. Consumer and Community Engagement Framework, February 2012.

[RA04] Hema Raghavan and James Allan. Using Soundex Codes for Indexing Names in ASR Documents. In Bhuvana Ramabhadran and Douglas Oard, editors, *HLT-NAACL 2004 Workshop: Interdisciplinary Approaches to Speech Indexing and Retrieval*, pages 22–27, Boston, Massachusetts, USA, May 6 2004. Association for Computational Linguistics.

[RSAG+08] D. Rebholz-Schuhmann, M. Arregui, S. Gaudan, H. Kirsch, and A. Jimeno. Text processing through Web services: calling Whatizit. *Bioinformatics*, 24(2):296–298, Jan 2008.

[SBJ+09] N. H. Shah, N. Bhatia, C. Jonquet, D. Rubin, A. P. Chiang, and M. A. Musen. Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinformatics*, 10 Suppl 9:S14, 2009.

[SC07] Diana Santos and Nuno Cardoso. *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. 2007.

[SH03] A. S. Schwartz and M. A. Hearst. A simple algorithm for identifying abbreviation definitions in biomedical text. *Pac Symp Biocomput*, pages 451–462, 2003.

[SL16] Augusto Cravo Silva and Carla Teixeira Lopes. Health Translations: a crowd-sourced, gamified approach to translate large vocabulary databases. In *CISTI - Conferencia Ibérica de Sistemas y Tecnologías de Información*, 2016.

[Sta] StatCounter. Top 5 Desktop, Tablet & Console Browsers from Jan 2015 to June 2016. http://gs.statcounter.com/#browser-ww-monthly-201501-201606. Accessed June 2016.

[STA+08] L. Smith, L. K. Tanabe, R. J. Ando, C. J. Kuo, I. F. Chung, C. N. Hsu, Y. S. Lin, R. Klinger, C. M. Friedrich, K. Ganchev, M. Torii, H. Liu, B. Haddow, C. A. Struble, R. J. Povinelli, A. Vlachos, W. A. Baumgartner, L. Hunter, B. Carpenter, R. T. Tsai, H. J. Dai, F. Liu, Y. Chen, C. Sun, S. Katrenko, P. Adriaans, C. Blaschke, R. Torres, M. Neves, P. Nakov, A. Divoli, M. Mana-Lopez, J. Mata, and W. J. Wilbur. Overview of BioCreative II gene mention recognition. *Genome Biol.*, 9 Suppl 2:S2, 2008.

[SVMA12] Samuel Alan Stewart, Maia Elizabeth Von Maltzahn, and SS Raza Abidi. Comparing Metamap to MGrep as a tool for mapping free text to formal medical lexicons. In *Proceedings of the 1st international workshop on knowledge extraction & consolidation from social-media in conjunction with the 11th international semantic web conference (ISWC 2012), Boston, USA*, pages 63–77. Citeseer, 2012.

[TCS10] Michael A Tanenblatt, Anni Coden, and Igor L Sominsky. The ConceptMapper Approach to Named Entity Recognition. In *LREC*, 2010.

[TL07] E. G. Toms and C. Latter. How consumers search for health information. *Health Informatics Journal*, 13(3):223–235, September 2007.

[TML+16] E. Tseytlin, K. Mitchell, E. Legowski, J. Corrigan, G. Chavan, and R. S. Jacobson. NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinformatics*, 17(1):32, 2016.

[TT03] Yoshimasa Tsuruoka and Jun'ichi Tsujii. Boosting Precision and Recall of Dictionary-Based Protein Name Recognition. In *Proc. of the ACL-03 Workshop on Natural Language Processing in Biomedicine*, pages 41–48, 2003.

[TTP15] J. M. Tenorio and I. Torres Pisa. Consumer Health Vocabulary: A Proposal for a Brazilian Portuguese Language. *Stud Health Technol Inform*, 216:1089, 2015.

[USSD11] O. Uzuner, B. R. South, S. Shen, and S. L. DuVall. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc*, 18(5):552–556, 2011.

[VdB14] S. Van den Broucke. Health literacy: a critical concept for public health. *Arch Public Health*, 72(1):10, 2014.

[VMHZ14] V. G. Vydiswaran, Q. Mei, D. A. Hanauer, and K. Zheng. Mining consumer health vocabulary from community-generated text. *AMIA Annu Symp Proc*, 2014:1150–1159, 2014.

[W3T16] W3Techs. Usage of content languages for websites. https://w3techs.com/technologies/overview/content_language/all, June 2016. Accessed June 2016.

[ZCM+03] Q. Zou, W. W. Chu, C. Morioka, G. H. Leazer, and H. Kangarloo. IndexFinder: a method of extracting key concepts from clinical texts for indexing. *AMIA Annu Symp Proc*, pages 763–767, 2003.

[ZKA+02] Q. Zeng, S. Kogan, N. Ash, R. A. Greenes, and A. A. Boxwala. Characteristics of consumer terminology for health information retrieval. *Methods Inf Med*, 41(4):289–298, 2002.

[ZT06] Q. T. Zeng and T. Tse. Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc*, 13(1):24–29, 2006.

[ZTC+05] Q. T. Zeng, T. Tse, J. Crowell, G. Divita, L. Roth, and A. C. Browne. Identifying consumer-friendly display (CFD) names for health concepts. *AMIA Annu Symp Proc*, pages 859–863, 2005.

[ZTGK+07] Q. Zeng-Treitler, S. Goryachev, H. Kim, A. Keselman, and D. Rosendale. Making texts in electronic health records comprehensible to consumers: a prototype translator. *AMIA Annu Symp Proc*, pages 846–850, 2007.

REFERENCES

Appendix A

Default Settings

A.1 Description and Execution Mode

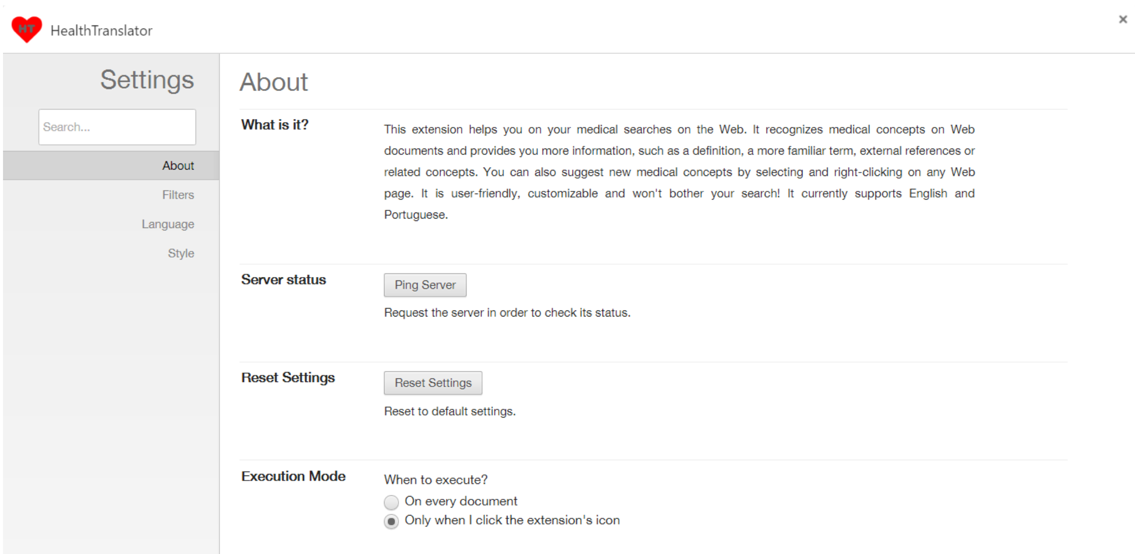


Figure A.1: Default settings – *About* section

A.2 Filters

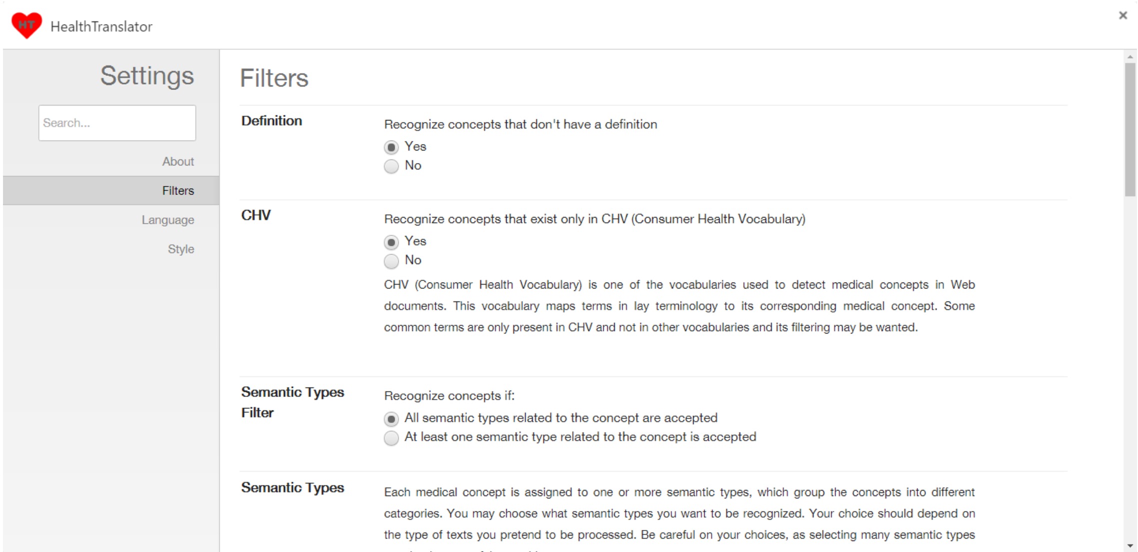


Figure A.2: Default settings – *Filters* section (I)

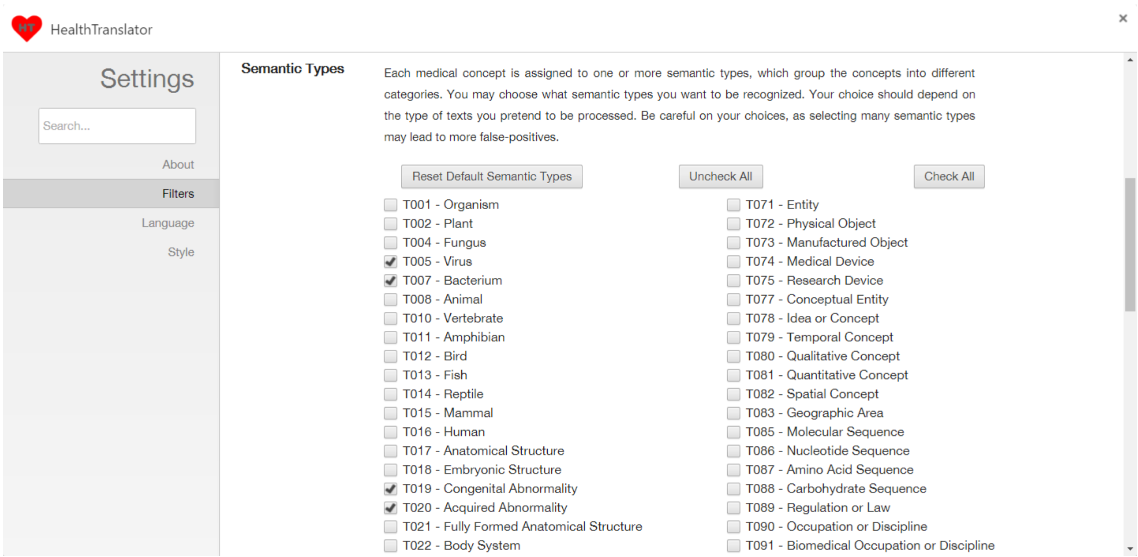


Figure A.3: Default settings – *Filters* section (II)

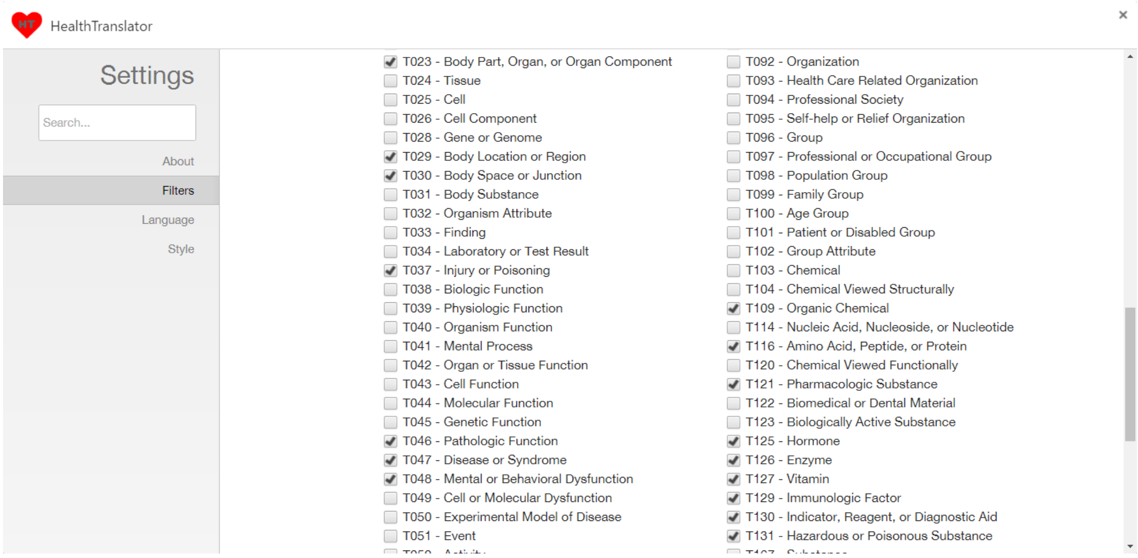


Figure A.4: Default settings – *Filters* section (III)

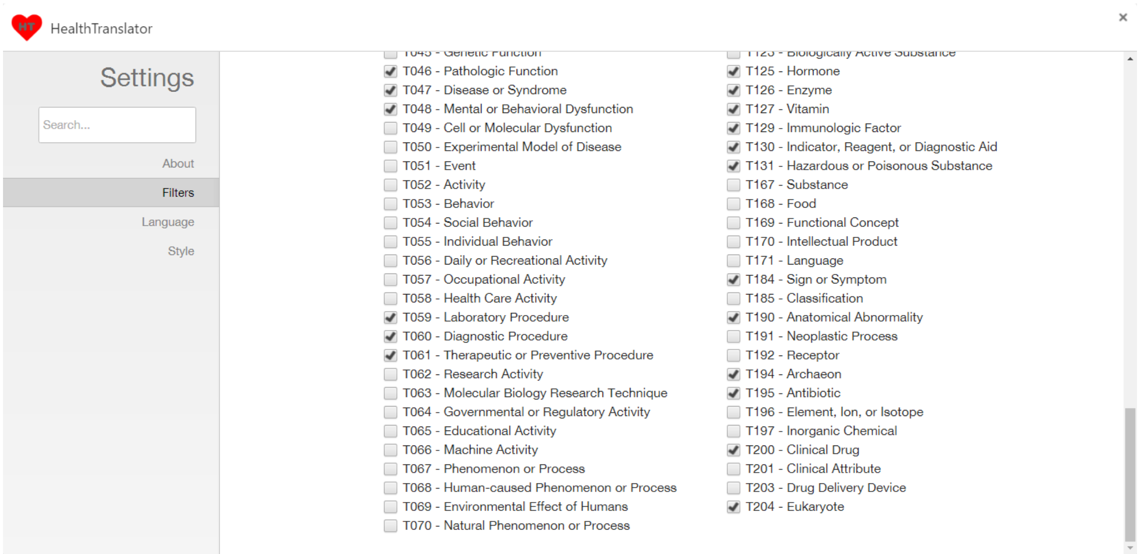


Figure A.5: Default settings – *Filters* section (IV)

A.3 Language

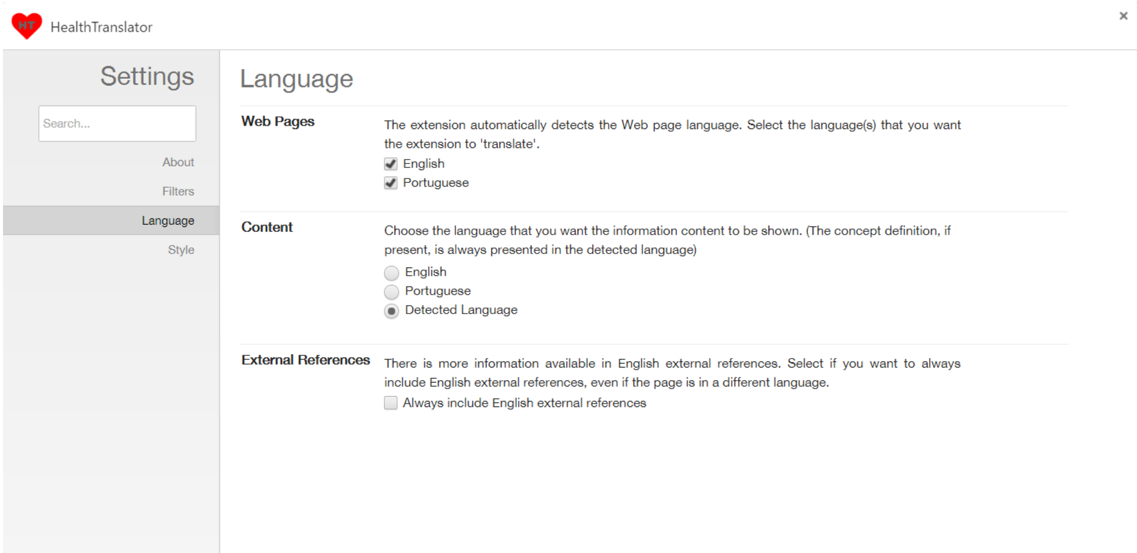


Figure A.6: Default settings – *Language* section

A.4 Style

Figure A.7: Default settings – *Style* section

Appendix B

Comparison between Medical Translator and HealthTranslator

Meaning of abbreviations used in the table:

TC [Total Concepts] - total number of concepts annotated

UC [Unique Concepts] - number of unique concepts annotated. A unique concept is identified by its CUI. Different terms may map to the same CUI, e.g. 'hypertension' and 'high blood pressure'.

D [Definitions] - amount of definitions provided, among the unique concepts (UC). A definition is related to a unique concept, e.g. 'hypertension' and 'high blood pressure' have the same definition.

Table B.1: Comparison of processing times and concept coverage between Medical Translator and HealthTranslator.

	Medical Translator			HealthTranslator			
Doc	Processing (ms)	TC	UC	Processing (ms)	TC	UC	D
1885	39	41	8	775	140	54	37
4001	60	26	10	1545	165	77	58
4136	105	60	39	1240	166	119	99
4151	151	122	45	1912	198	101	74
4218	217	366	63	5630	669	233	187
4661	64	96	41	2026	254	125	99
4706	69	72	33	1049	104	66	52
4767	88	63	23	1469	176	80	54
4797	52	52	24	1272	193	123	99
4833	129	186	80	1796	241	158	138
4842	87	138	76	1625	224	136	111

Table B.1: (continued)

	Medical Translator			HealthTranslator			
Doc	Processing (ms)	TC	UC	Processing (ms)	TC	UC	D
5101	242	743	145	6554	964	326	269
5199	88	147	50	2119	206	102	80
5319	112	123	48	3719	385	157	123
5592	114	173	65	2110	341	149	124
5600	22	5	4	485	24	15	10
5732	90	43	17	1221	118	80	64
5755	71	107	60	1747	176	103	82
5857	115	160	85	1715	299	191	163
5919	127	215	96	4222	636	261	201
59693	342	233	76	19287	1198	321	241
5971	88	177	81	2997	362	219	193
5982	50	35	16	1283	27	19	14
60036	152	253	97	4010	584	253	204
60199	110	242	73	2814	330	170	135
60461	107	113	40	3874	443	181	128
6057	92	104	48	2701	402	178	144
60659	110	92	50	3613	540	175	151
60673	82	41	13	4467	356	143	95
6078	55	38	17	1177	128	64	53
60874	78	167	85	3803	439	212	182
61068	40	37	15	1200	126	70	56
61108	600	137	34	4948	762	185	148
61119	23	12	7	875	55	30	20
61217	245	115	51	2720	270	132	111
61396	163	114	97	2015	233	187	169
61480	240	415	79	5514	823	234	193
61633	150	187	40	2515	274	137	112
61758	269	293	76	4302	562	195	167
61838	158	78	31	1699	165	88	74
61846	174	101	42	6276	889	191	154
61976	55	31	12	2381	212	117	91
62067	125	155	46	3725	576	179	137
62071	68	31	17	1814	141	73	56
62091	304	188	92	2257	310	192	158
62116	268	156	67	2543	228	151	128
62377	62	47	33	1247	82	60	56

Table B.1: (continued)

	Medical Translator			HealthTranslator			
Doc	Processing (ms)	TC	UC	Processing (ms)	TC	UC	D
62473	488	272	53	3012	384	151	110
62496	159	53	19	1054	78	38	30
62506	123	132	51	1896	234	124	109
62517	106	121	52	3379	416	150	115
62533	117	198	33	3023	344	98	73
62604	97	58	32	2390	383	172	139
62609	87	31	20	3105	274	92	70
62677	91	111	38	2501	262	131	113
62737	75	36	15	1898	193	56	44
62871	76	61	35	2081	182	104	76
62911	41	16	9	1458	108	32	26
63107	66	55	10	1537	121	61	49
63201	85	126	28	1701	100	47	28
63697	35	30	28	1591	91	70	60
63736	114	220	155	2209	338	279	254
63743	286	404	173	9470	1190	567	507
63745	53	34	17	1218	83	41	39
63755	139	178	63	3085	436	152	122
6381	62	100	42	2520	342	213	192
6384	36	20	13	854	69	33	30
63847	26	21	9	702	87	40	30
63848	62	122	55	1432	185	113	95
63862	58	47	31	2690	280	145	115
6408	59	39	28	1380	194	105	93
64228	98	67	20	2941	348	147	109
64232	296	310	85	8511	702	251	211
64399	138	89	43	9992	1027	216	168
64846	23	9	9	1098	127	57	43
6533	96	156	45	1878	371	139	114
6608	57	70	20	1660	149	66	53
6774	49	87	44	2237	243	145	125
6789	28	19	15	968	60	43	37
6836	90	155	62	3002	294	153	115
6837	46	61	39	1204	149	101	89
6878	57	68	23	1227	64	48	36
7062	37	36	10	689	32	15	10

Table B.1: (continued)

	Medical Translator			HealthTranslator			
Doc	Processing (ms)	TC	UC	Processing (ms)	TC	UC	D
7214	69	134	28	1547	144	61	42
7295	118	218	67	3064	391	167	150
7329	70	154	95	1908	246	178	154
7545	85	80	40	1121	144	113	94
7721	94	100	15	1118	243	78	50
8049	49	35	10	592	57	25	17
8154	90	160	88	2898	350	165	124
8224	271	129	62	5480	287	152	124
8237	53	34	20	766	157	46	36
8261	102	173	36	1648	224	145	126
8268	60	53	18	1149	76	43	37
8305	82	88	34	2100	399	164	118
8481	48	38	15	586	45	29	21
8508	74	65	11	3625	361	102	84
8793	116	114	56	10373	701	257	191
8823	46	35	25	864	115	46	36
8932	55	53	14	1549	180	91	75

Appendix C

Annotation Guidelines

HealthTranslator

Olá!
Antes de mais, **obrigado por participar neste projeto**.

Segue-se uma breve descrição do projeto, a importância da sua participação e as orientações para a anotação dos documentos.

O meu nome é Hugo Sousa e no âmbito da dissertação de mestrado de Engenharia Informática e Computação da Faculdade de Engenharia da Universidade do Porto, supervisionado pela professora Carla Lopes, pretende-se desenvolver uma extensão para o *browser* que seja capaz de facilitar a pesquisa de informação médica por utilizadores leigos, vulgarmente conhecidos por consumidores de saúde. Assim sendo, o objetivo da aplicação será reconhecer termos médicos, que nem sempre são de fácil compreensão por utilizadores leigos, e fornecer uma breve descrição, possivelmente acompanhado de outros recursos externos, facilitando a sua leitura ou posterior pesquisa.

Os documentos presentes neste projeto foram obtidos automaticamente e categorizados como pertencentes ao ramo médico. Estes encontram-se inicialmente anotados, mas esta anotação foi também realizada de uma forma automática, pelo que se encontra bastante rudimentar e errónea, daí a sua ajuda ser fundamental.

Os resultados desta anotação serão usados como base de avaliação da ferramenta a ser desenvolvida nesta dissertação. No entanto, a usabilidade deste conjunto de documentos pode ser muito mais alargada. Por um lado, não se conhece até agora, um conjunto de documentos desta natureza devidamente anotado (textos do foro médico, mas de literatura leiga), seja na língua portuguesa ou inglesa. Por outro lado, esta anotação pode contribuir para a comunidade científica, servindo de base a implementações mais eficientes de aplicações orientadas a consumidores de saúde.

Orientações de anotação

Pede-se aos colaboradores que anotem os conceitos médicos nos textos fornecidos (sejam eles numa terminologia científica ou leiga), marcando o conceito com "Medical Concept" na ferramenta.

Notas gerais:

- Começando a anotar um documento, deve ser anotado até ao fim, evitando que um documento seja apenas parcialmente anotado. As anotações vão sendo gravadas automaticamente.
- Em caso de palavras compostas, deve ser anotado o conceito mais específico. Geralmente será o conceito de maior comprimento.
  Exemplo: "enfarte do miocárdio" em vez de "enfarte" e "miocárdio".

- Não devem existir anotações sobrepostas. Por exemplo, não deve anotar “enfarte do miocárdio” e “enfarte” simultaneamente. Pode acontecer existirem anotações automáticas sobrepostas. Se for o caso, remova ou corrija a anotação de forma a não se sobreporem.
- Ao anotar um conceito que surge várias vezes no texto, a ferramenta sugere anotar todos estes simultaneamente. Tendo em conta os tópicos anteriores, deve evitar usar esta funcionalidade, de forma a não criar anotações sobrepostas ou que o conceito não seja o mais específico, pelo que cada caso deve ser avaliado individualmente.
- Ter em atenção ao selecionar o conceito de forma a não conter espaços em branco antes ou depois do conceito, nem incluir elementos de pontuação.
- Não hesite em remover ou corrigir uma anotação automática. É natural que surja essa necessidade. Se a anotação automática está parcialmente correta, elimine-a e contribua com a sua anotação.
- A anotação final será o conjunto de anotações automáticas e manuais. Assim sendo, se uma anotação automática se encontra correta, não necessita de a modificar.

Alguma questão ou dúvida, não hesite em contactar-me através do seguinte e-mail.

Hugo Sousa - ei11083@fe.up.pt

Appendix D

Comparison of manual and HealthTranslator annotations

Meaning of abbreviations used in the tables:

COR [CORRECT] - correct annotation

PAR [PARTIAL] - partially correct annotation (one of the annotations has the same end index and a different start index)

MIS [MISSING] - annotations existing only in the gold standard annotation set (manual annotations)

SPU [SPURIOUS] - annotations existing only in the candidate annotation set (HealthTranslator annotations)

PRE [PRECISION]

$$PRE = \frac{COR}{COR + PAR + SPU} \tag{D.1}$$

REC [RECALL]

$$REC = \frac{COR}{COR + PAR + MIS} \tag{D.2}$$

FSC [F-SCORE] - also known as F-measure or F1.

$$FSC = 2 * \frac{PRE * REC}{PRE + REC} \tag{D.3}$$

As there is only one type of annotation (medical concept), there are no incorrect annotations and thus are not included in the formulas.

Table D.1: Metrics resulting from comparison of manual annotations and HealthTranslator with all semantic types

Doc	COR*	PAR*	MIS*	SPU*	PRE*	REC*	FSC*
237	15	0	33	36	0.2941	0.3125	0.303
433	45	3	38	57	0.5233	0.4286	0.4712
480	132	23	154	63	0.6055	0.4272	0.5009
796	10	1	19	6	0.5882	0.3333	0.4255
808	46	18	249	46	0.4182	0.147	0.2175
1122	25	3	43	8	0.6944	0.3521	0.4673
1756	26	7	35	21	0.4815	0.3824	0.4262
2127	45	5	55	19	0.6522	0.4286	0.5172
2763	15	1	51	12	0.5357	0.2239	0.3158
2820	1	0	0	8	0.1111	1	0.2
2862	2	0	10	2	0.5	0.1667	0.25
3387	123	45	404	123	0.4227	0.215	0.2851
3446	9	2	89	3	0.6429	0.09	0.1579
3482	14	2	13	2	0.7778	0.4828	0.5957
3506	2	3	14	5	0.2	0.1053	0.1379
3711	0	0	12	3	0	0	0
4712	12	2	56	27	0.2927	0.1714	0.2162
5139	5	0	8	15	0.25	0.3846	0.303
5418	1	0	8	8	0.1111	0.1111	0.1111
5441	30	1	82	108	0.2158	0.2655	0.2381
5648	26	7	102	32	0.4	0.1926	0.26
6005	0	0	37	6	0	0	0
6603	5	0	51	6	0.4545	0.0893	0.1493
7018	17	8	110	19	0.3864	0.1259	0.1899
7117	10	4	125	21	0.2857	0.0719	0.1149
8960	50	6	217	51	0.4673	0.1832	0.2632
10809	64	3	85	34	0.6337	0.4211	0.5059
10818	10	3	128	32	0.2222	0.0709	0.1075
10823	7	1	11	7	0.4667	0.3684	0.4118
11094	3	2	26	3	0.375	0.0968	0.1538
11342	5	2	32	3	0.5	0.1282	0.2041
11343	27	2	78	18	0.5745	0.2523	0.3506
11606	5	0	10	11	0.3125	0.3333	0.3226
11846	6	1	53	8	0.8833	0.5333	0.1429
11881	25	2	35	36	0.3968	0.4032	0.4
12050	8	2	38	11	0.381	0.1667	0.2319
17934	1	0	10	17	0.0556	0.0909	0.069
18747	3	0	18	7	0.3	0.1429	0.1935

* - check the abbreviation's meaning in the beginning of the section.

Table D.2: Metrics resulting from comparison of manual annotations and HealthTranslator with default semantic types

Doc	COR	PAR	MIS	SPU	PRE	REC	FSC
237	25	8	15	282	0.07937	0.5208	0.1377
433	62	11	32	168	0.2573	0.5905	0.3584
480	240	28	41	330	0.4013	0.7767	0.5292
796	15	3	12	33	0.2941	0.5	0.3704
808	112	56	145	323	0.2281	0.3578	0.2786
1122	43	7	21	106	0.2756	0.6056	0.3789
1756	37	12	19	58	0.3458	0.5441	0.4229
2127	64	18	23	108	0.3368	0.6095	0.4339
2763	38	2	27	84	0.3065	0.5672	0.3979
2820	1	0	0	58	0.01695	1	0.0333
2862	10	0	2	35	0.2222	0.8333	0.3509
3387	358	80	134	860	0.2758	0.6259	0.3829
3446	38	7	55	57	0.3725	0.38	0.3762
3482	25	2	2	14	0.6098	0.8621	0.7143
3506	8	3	8	27	0.2105	0.4211	0.2807
3711	7	2	3	13	0.3182	0.5833	0.4118
4712	28	11	31	153	0.1458	0.4	0.2137
5139	11	0	2	51	0.1774	0.8462	0.2933
5418	5	0	4	90	0.0526	0.5556	0.0962
5441	79	7	27	368	0.174	0.6991	0.0814
5648	63	12	60	275	0.18	0.4667	0.2598
6005	34	2	1	42	0.4359	0.9189	0.5913
6603	39	3	14	76	0.3305	0.6964	0.4483
7018	77	19	39	167	0.2928	0.5704	0.3869
7117	47	11	80	240	0.1572	0.3381	0.2146
8960	105	54	114	319	0.2197	0.3846	0.2796
10809	97	6	49	258	0.2687	0.6382	0.3782
10818	105	12	24	128	0.4286	0.7447	0.544
10823	16	1	2	35	0.3077	0.8421	0.4507
11094	16	4	11	59	0.2025	0.5161	0.2909
11342	23	3	13	24	0.46	0.5897	0.5169
11343	45	20	42	77	0.3169	0.4206	0.3614
11606	11	0	4	86	0.1134	0.7333	0.1964
11846	30	10	20	64	0.2885	0.5	0.3659
11881	31	2	29	106	0.4677	0.5	0.3085
12050	12	3	33	52	0.1791	0.25	0.2087
17934	4	0	7	144	0.027	0.3636	0.0503
18747	7	1	13	40	0.1458	0.3333	0.2029

* - check the abbreviation's meaning in the beginning of the section.

Comparison of manual and HealthTranslator annotations

Appendix E

User Study - Search Tasks

Código: ___ **Pré: __ Pós: __**

Parte I (Asma)

Indique, para cada item, se é verdadeiro ou falso.

Usa-se um inalador de alívio (broncodilatador ou "bomba") para reduzir a inflamação dos pulmões.	
Usa-se um registo de DEMI (Débito Expiratório Máximo Instantâneo) para ter a certeza que os seios perinasais estão abertos (teste do sopro para detetar se há sinusite).	
Um alergénio é o anticorpo que falta aos asmáticos.	
A maioria das crianças asmáticas tem de ir ao hospital quando tem um ataque de asma.	
Uma criança deixa de ser asmática se, durante vários anos, deixar de ter sintomas como o aperto no peito ou pieira.	

Parte II (Nutrição)

Que doenças ou problemas de saúde estão relacionados com a baixa ingestão de fibra?

(Indique 3 problemas)

____

Pensas que estes comportamentos ajudam a reduzir a probabilidade de vir a ter certos tipos de cancro?

	Sim	Não
Comer mais fibra		
Comer mais frutas e legumes		

Acreditas que estes comportamentos ajudam a prevenir doenças do coração?

	Sim	Não
Comer menos gordura saturada		
Comer menos sal		
Comer mais frutas e legumes		

Qual destes nutrientes mais contribui para aumentar os níveis de colesterol do sangue das pessoas? *(Escolha uma opção)*

Antioxidantes	
Gorduras polinsaturadas	
Gorduras saturadas	
Colesterol da dieta	

Quais destas vitaminas acreditas que são antioxidantes?

	Sim	Não
Vitamina A		
Vitaminas do complexo B		
Vitamina C		
Vitamina D		
Vitamina E		
Vitamina K		

User Study - Search Tasks

Appendix F

User Study - Feedback Questionnaire

HealthTranslator Evaluation

This questionnaire is to be answered after executing the tasks for the extension testing.

Your contribution on this project is appreciated.

*Required

What is your age? *

Your answer

What is your gender? *

- ○ Male
- ○ Female

What is your academic degree? *

- ○ 4th grade
- ○ 9th grade
- ○ 12th grade
- ○ Bachelor / Master
- ○ Doctorate

Figure F.1: Feedback Questionnaire (I)

User Study - Feedback Questionnaire

How often do you search for health-related information online? *

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Frequently

Do you usually face difficulties comprehension issues when reading health information online? *

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Frequently

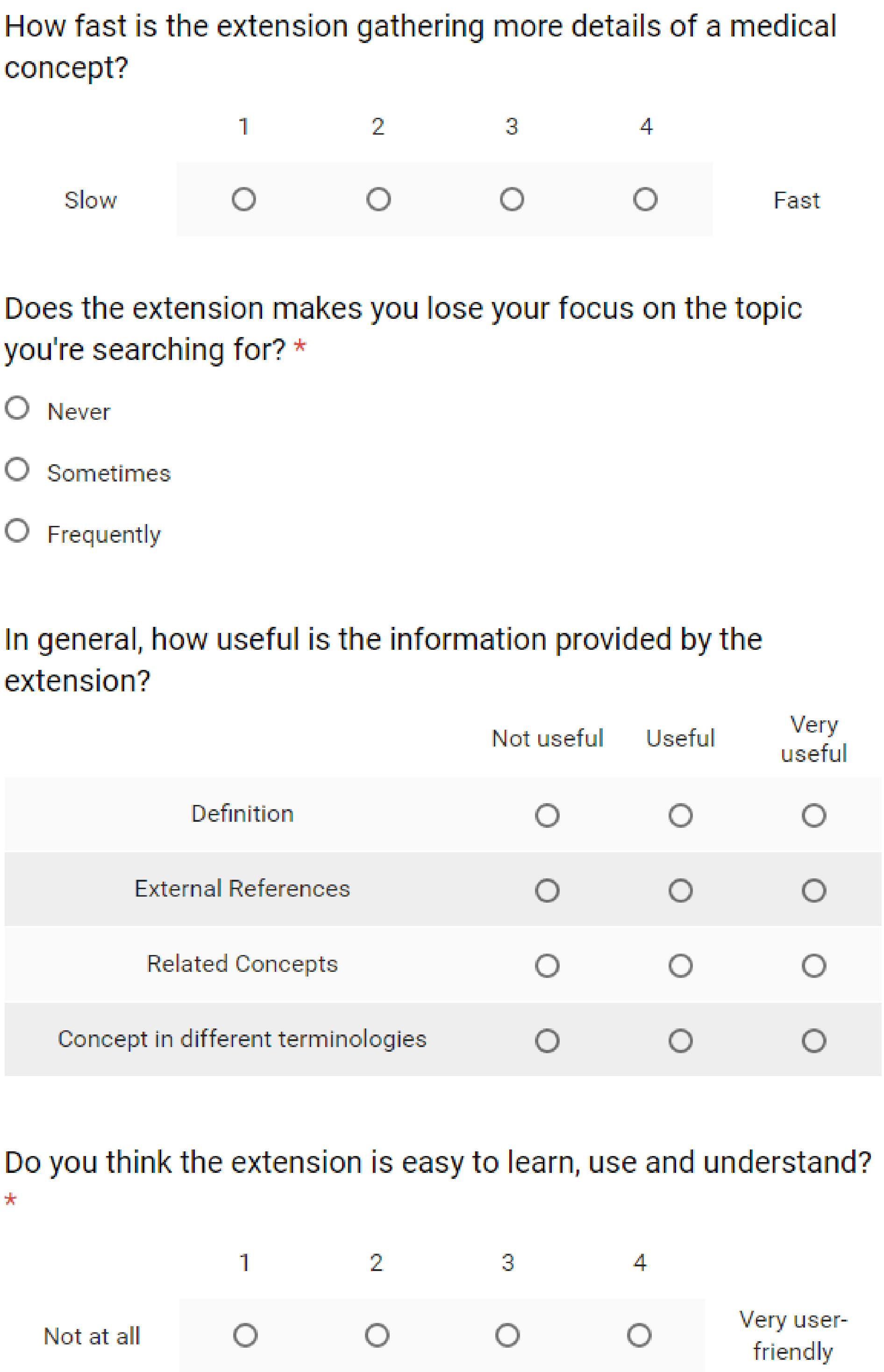
Does the extension helps you to shorten the time you take to find what you want? *

- ○ No, it's even slower
- ○ It doesn't interfere
- ○ Yes

Does the extension eases the comprehension of the information you find online? *

- ○ No, it doesn't
- ○ It doesn't interfere
- ○ Yes

Figure F.2: Feedback Questionnaire (II)



How fast is the extension gathering more details of a medical concept?

	1	2	3	4	
Slow	○	○	○	○	Fast

Does the extension makes you lose your focus on the topic you're searching for? *

- ○ Never
- ○ Sometimes
- ○ Frequently

In general, how useful is the information provided by the extension?

	Not useful	Useful	Very useful
Definition	○	○	○
External References	○	○	○
Related Concepts	○	○	○
Concept in different terminologies	○	○	○

Do you think the extension is easy to learn, use and understand? *

	1	2	3	4	
Not at all	○	○	○	○	Very user-friendly

Figure F.3: Feedback Questionnaire (III)

Do you think the extension is intrusive? *

	1	2	3	4	
Not at all	○	○	○	○	Very intrusive

Do you like the extension's look-and-feel? *

	1	2	3	4	
Not at all	○	○	○	○	A lot

Did you find concepts that were incorrectly recognized as medical terms? *

- ○ I was not aware of it
- ○ None
- ○ A few
- ○ Many

Did you find concepts without definition or with an obviously wrong definition? *

- ○ None
- ○ A few
- ○ Many

Figure F.4: Feedback Questionnaire (IV)

Currently, you can suggest new medical concepts. These concepts could later be analysed by a professional and eventually added to the recognized concepts. Do you find this feature relevant? *

- ○ No
- ○ Indifferent
- ○ Yes

Currently, you can rate the information of recognized concepts. These ratings could later be analysed and eventually their content be improved. Do you find this feature relevant? *

- ○ No
- ○ Indifferent
- ○ Yes

Ideas for new features / improvements

Your answer

Would you consider using the extension in your daily life? *

- ○ No
- ○ Yes

Figure F.5: Feedback Questionnaire (V)

Appendix G

Times to perform questionnaire tasks

Table G.1: Time spent to answer each individual task in the questionnaire per user. Ext.: the task in which the extension was used. A – Asthma; N – Nutrition.

		Asthma					Nutrition				
Ext.	Task / User	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
A	1	06:00	01:39	01:04	04:50	00:45	03:07	01:10	04:53	01:45	02:32
N	2	01:09	05:56	02:37	01:25	00:50	01:23	02:12	09:53	04:39	01:28
N	3	01:48	03:21	00:48	02:12	01:17	01:42	01:46	05:34	01:42	05:10
N	4	03:14	04:13	06:15	02:55	03:06	06:32	04:28	02:22	01:55	03:14
N	5	02:16	02:17	00:32	03:17	44:00	02:49	01:23	01:59	00:49	00:39
A	6	03:32	02:00	03:21	01:48	01:34	02:58	02:20	04:42	03:36	04:09
A	7	00:35	01:37	02:26	00:56	00:25	01:02	03:44	01:54	01:32	03:15
A	8	05:26	03:37	00:56	01:47	01:43	01:51	01:26	02:42	02:23	01:07
A	9	03:09	02:30	01:01	02:16	07:53	06:07	02:29	06:10	03:43	03:35
N	10	01:53	02:09	00:55	01:54	01:16	03:06	02:05	02:29	02:21	01:21
A	11	01:42	02:18	00:46	02:06	00:23	01:50	00:57	01:25	02:12	00:56
N	12	01:49	01:58	00:34	03:28	02:44	00:54	04:06	09:40	05:30	04:17
A	13	00:47	05:56	01:40	01:09	00:15	01:09	01:49	03:55	00:44	00:35
N	14	00:54	01:07	00:19	01:30	01:27	02:05	00:30	02:31	00:37	00:36
N	15	00:13	01:26	01:52	02:37	00:17	02:21	02:31	05:12	00:53	03:52
N	16	01:21	01:14	02:09	01:22	00:41	01:23	02:52	04:56	01:52	02:49
A	17	01:47	07:31	00:47	01:04	02:34	02:53	03:57	01:57	03:03	06:16
A	18	02:10	05:50	02:53	01:42	04:11	01:10	01:09	02:14	03:13	01:49
A	19	02:51	04:41	01:20	04:12	03:18	01:35	01:20	02:50	02:43	04:52
N	20	00:35	05:31	01:55	01:27	00:55	02:55	01:55	06:45	00:55	07:56