

An Automatic System for Helping Health Consumers To Understand Medical Texts

Marco Alfano^{1,2}, Biagio Lenzitti¹, Giosuè Lo Bosco^{1,3} and Valerio Perticone¹

¹*Department of Mathematics and Computer Science, University of Palermo, via Archirafi n.34, Palermo, Italy*

²*Anghelos Centre on Communication Studies, via Pirandello n.40, Palermo, Italy*

³*Euro-Mediterranean Institute of Science and Technology, via E. Amari n.123, Palermo, Italy*
marco.alfano@anghelos.org, {biagio.lenzitti, giosue.lobosco}@unipa.it, valerio@pertico.net

Keywords: E-Health, Public Health, Healthcare Management Systems, Patient Empowerment, Plain Language, Consumer Health Vocabulary, Infobutton, Electronic Health Record, Personal Health Record.

Abstract: Medical texts (reports, articles, etc.) are usually written by professionals (physicians, medical researchers, etc.) who use their own language and communication style. On the other hand, these texts are often read by health consumers (as in the case of medical reports) who do not have the same skills and vocabularies of the experts and can have difficulties in text comprehension. To help a health consumer in understanding a medical text, it would be desirable to have an automatic system that, given a text written with medical (technical) terms, translates them in simple or plain language and provides additional information with the same kind of language. We have designed such a system. It processes online medical documents and provides health consumers with the needed information for their understanding. To this end, we use a medical vocabulary for finding the technical terms in the medical texts, a consumer health vocabulary (CHV) for translating the technical terms into their consumer equivalents and a health-consumer dictionary for finding supplementary information on the terms. We have built a prototype that processes Italian medical reports and uses infobuttons next to the technical terms for allowing easy retrieval of the desired information.

1 INTRODUCTION

Medical texts (reports, articles, etc.) are usually written by professionals (physicians, medical researchers, etc.) who use their own language and communication style. On the other hand, these texts are often read by health consumers (as in the case of medical reports) who do not have the same skills and vocabularies of the experts and can have difficulties in text comprehension (Keselman and Slaughter, 2007; Seedorff and Peterson, 2013).

A language that can be understood by almost anyone, made of common every-day terms, should be possibly used in writing medical content that also goes to non-experts. As an example, *plain* language is used by the US Government for improving its communication with the public (Plain Language.com, 2014). It is defined as a kind of language that audience can understand the first time they read it. Simple English Wikipedia is another example of content developed using 2,000 common

English words (Simple English Wikipedia, 2014). Of course, it is not always easy for the medical experts to write in plain language because of the risk of losing precision and the time often required to simplify concepts.

To help a health consumer in understanding a medical text, it would then be desirable to have an automatic system that, given a text written with medical (technical) terms, translates them in plain language and provides additional information with the same kind of language. In this way, the user is not required to increase his/her knowledge to a level where he/she is able to understand the text being examined but knowledge is somehow brought back to the user level so greatly facilitating his/her comprehension. Moreover, it can help the communication between experts and non-experts (e.g., physicians and patients or scientists and laymen) by providing a sort of two-way translation of terms.

In this work, we describe a system that, given an online medical text, finds the technical terms, translates them in a plain language and provides additional information for them (in plain language) greatly facilitating text understanding. To this end, we use a medical vocabulary for finding the technical terms in the medical text, a consumer health vocabulary (CHV) for translating the technical terms into their equivalents for consumers and a consumer dictionary for finding supplementary information on the terms. We have developed a prototype that processes Italian medical reports and uses infobuttons next to the technical terms for allowing easy retrieval of the desired information.

Our system, to the best of our knowledge, constitutes one of the first solutions in this field. In fact, infobuttons have been used for years mainly to support clinicians' decisions and only recently they have been used to bring information to patients (Baorto and Cimino, 2000; Kemper, 2010). Moreover, a couple of systems that present some similarities with our system are described in (Zeng-Treitler and Goryachev, 2007) and (Kandula, 2010). However, they replace the medical terms with the consumer ones and add further information within the text, altering the original text. Our system provides the consumer information 'aside' the original text leaving it intact. Moreover, our system provides term definitions in consumer language whereas the other two systems do not.

The paper is organized as follows. The second section describes the resources that help health consumers to understand a medical text. The third section describes the infobuttons that can be used for automatic retrieval of health information. The fourth section describes the architecture and implementation details of the system we have developed together with some practical use. The final section presents some conclusions and future work.

2 MEDICAL VOCABULARIES, THESAURI AND DICTIONARIES

As said in the Introduction, when a health consumer reads a medical text written by a professional he/she has often difficulties in understanding it because of the used terms and language. He/she may then need some external help to understand the technical terms, find familiar synonyms and get additional

information. This external help comes in the form of different resources (online or not) such as vocabularies, dictionaries and thesauri:

- a 'medical vocabulary' is a selective list of words and phrases used in the medical field; it can be used to find the technical terms in a medical text;
- a 'medical-consumer thesaurus' contains synonyms and antonyms of medical terms; it can be used to find plain synonyms of the technical terms;
- a 'health-consumer dictionary' gives information about the meaning of the words; it can be used to find additional information on the technical terms.

In some cases, a single resource can have multiple functionalities, e.g., it can contain both definitions and synonyms. Of course, there are numerous resources of each type. In the next subsections we briefly describe the resources we have used for building the system presented in Section 4 that deals with Italian reports even though the proposed methodology can be applied to any language.

2.1 Medical vocabularies

Medical vocabularies, as said above, are selective lists of words and phrases used in the medical field. Usually they are created for the professionals and, for this reason, they contain all the technical terms used in the medical field. Usually, health consumers use these resources when they do not find the medical terms in the consumer dictionaries or want more technical information on the searched term.

The 'Unified Medical Language System (UMLS)' is a large collection of multilingual vocabularies that contains information about biomedical and health related concepts created and maintained by the *US National Library of Medicine* (Unified Medical Language System, 2014). It mainly uses a 'Concept Unique Identifier – CUI' (a unique identifier for each concept) to create a mapping among these vocabularies and thus allows translation among the various terminology systems. It may also be viewed as a comprehensive thesaurus and ontology of health and biomedical concepts.

The 'Dizionario di Medicina e Biologia' by *Zanichelli* is a bilingual (italian and english) vocabulary and dictionary written by medical experts and used not only by technical translators, but also by professionals, physicians and health executives (Dizionario di Medicina e Biologia, 2014). Each entry contains an encyclopaedic section

with English translation and Italian explanation and it is frequently updated.

2.2 Medical-consumer thesauri

Consumer terms are not usually well covered by the existing medical vocabularies, which mostly represent the language of health professionals (Zeng and Tse, 2006). Indeed, expressions used by consumers to describe health-related concepts and relationships among such concepts frequently differ on multiple levels (i.e., syntactic, conceptual and explanatory) from those of professionals. As a consequence, consumer health vocabularies (CHVs) have been created for translating medical terms and concepts in their equivalent for consumers and they can be very useful for translating medical (technical) terms in consumer ones (Zielstorff, 2003).

One of the best known examples of CHV is the ‘Open Access Collaboratory Consumer Health Vocabulary (OAC-CHV)’ created and maintained by the *Consumer Health Vocabulary Initiative* (Consumer Health Vocabulary Initiative, 2014). It is a relationship file that links commonly used terms to associated medical terminology represented by the UMLS. The OAC-CHV focuses on expressions and concepts that are employed in health-related communications from or to consumers.

The OAC-CHV contains around one hundred and sixty thousand rows (one for each term) and different fields among which:

- ‘Term’: The term as found in the text;
- ‘Concept Unique Identifier’ (CUI): The unique identifier of a concept as found in the UMLS;
- ‘CHV Preferred Name’: The preferred consumer term as defined in the Consumer Health Vocabulary;

In Section 4, we will show how we have used these fields for translating technical terms in consumer ones.

2.3 Health-consumer dictionaries

There are many sites created for health consumers that contain health and medical information suitable to general users. These sites usually contain health and medical dictionaries specifically created for health consumers and then use a language that can easily be understood by them. Examples of online dictionaries in English are WebMD (WebMD, 2014) and MedlinePlus (MedlinePlus, 2014)] and online dictionaries in Italian are Ok Salute (OK Salute,

2014) and Dizionario della Salute (Dizionario della Salute, 2014).

3 INFOBUTTONS FOR AUTOMATIC RETRIEVAL OF INFORMATION

Infobuttons are context-sensitive links usually inserted in online medical texts, as the ones found in a Clinical Information System (CIS), Electronic Health Record (EHR) or Personal Health Record (PHR), and allow easy retrieval of relevant information (Cimino, 2007). The online medical text shows a button next to some of its parts such as diagnoses or prescriptions. When clicked, the infobutton creates a query based on the context of the interaction and sends it to electronic health information resources retrieving information that helps understanding or completing the text being read (Infobuttons, 2014).

The Health Level Seven (HL7) Organization has created a standard for infobuttons, the “Context Aware Knowledge Retrieval Application (Infobutton), Knowledge Request”, that provides a standard mechanism for clinical information systems to submit knowledge requests to knowledge resources. The specification also defines a shared context information model to be implemented by EHR/PHR systems and knowledge resources (Infobuttons, 2014; HL7 Product Infobuttons, 2014). Based on the context, which includes characteristics of the patient, provider, care setting, and clinical task, infobuttons anticipate clinicians’ and patients’ questions and provide automated links to resources that may answer those questions. For example, an infobutton displayed in the context of a patient’s problem list may allow a clinician to retrieve treatment guidelines on a specific condition as well as relevant patient education material to share with the patient (HL7 Version 3 Implementation Guide, 2014).

As said above, infobuttons have been used for years mainly to support clinicians’ decisions regarding diagnosis and orders for tests and medications. More recently, infobuttons have also been used to trigger relevant and helpful information to patients (Baorto and Cimino, 2000; Kemper, 2010). We use infobuttons in a similar way, as we show in Section 4, i.e., to bring useful information to health consumers by automatically providing consumer synonyms and term definitions.

4 AN AUTOMATIC SYSTEM FOR HELPING COMPREHENSION OF MEDICAL TEXTS

As seen in Section 2, a health consumer reading a medical text will often use some additional resources to find the technical terms, their synonyms and related information so, ultimately, to understand the whole text. However, this approach assumes that the additional resources are readily available and this is not always the case. Moreover, the resources are not accessed in an established order leading to a disorganized process that can be time consuming and bring an information overload. It is then important to develop a process that uses the additional resources in a coherent and efficient way, even better if this process is automatized. The final objective of this process is the translation of the technical terms of a medical text to plain language and the provision of additional information for improving its understanding.

The basic steps of this process can be summarized as follows:

1. Take a medical text (possibly online) and find and highlight the technical terms (words or combinations of words) by using the medical vocabulary.
2. Translate the highlighted technical terms to non-technical, or plain, terms with the medical-consumer thesaurus.
3. Finally, provide additional plain information with the health-consumer dictionary.

By doing so, the user will have everything at hand by dealing with a single document that contains all the useful information for its understanding.

4.1 System architecture and implementation

In the frame of a collaboration with some Italian hospitals for providing advanced tools to health consumers, we have developed a system that automatically finds the technical terms in an online medical document, translates them in plain or consumer terms and provides additional information in plain language. The architecture of the system is shown in Figure 1.

The HIGHLIGHT module takes as input an arbitrary text and, using a medical vocabulary, highlights all the technical terms. The MAP module connects the technical terms previously found to their equivalent consumer terms by using a thesaurus

and puts an infobutton with a question-mark icon (?) next to the item for which a consumer translation exists. When clicked, the infobutton will show the consumer translation of the term in a tooltip near the item. The DEFINE module provides a description of the term retrieved by a consumer dictionary. It will put an infobutton with an information icon (i) next to each item for which a consumer definition exists and, when clicked, it will show the definition in a separate frame under the main text. This definition will also be processed by the whole system and transformed in an annotated hypertext that highlights the technical terms and add the related infobuttons so to allow the user a deeper analysis and navigation.

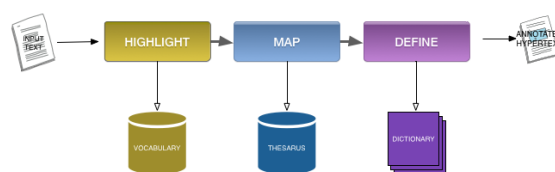


Figure 1. System architecture.

Since our system has to work with medical texts written in Italian, the HIGHLIGHT module uses, as medical vocabularies, the Italian vocabularies present in the UMLS, namely the Italian versions of the ‘Medical Subject Headings’ (MeSH), the ‘International Classification of Primary Care’ (ICPC), the ‘Medical Dictionary for Regulatory Activities Terminology’ (MedDRA) and the ‘Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy’ (MTHSMS) for a total of around 150,000 entries. Moreover, it also uses the ‘Dizionario di Medicina e Biologia’ by Zanichelli which has around 60,000 entries.

The MAP module uses, as a thesaurus, the Open Access Collaboratory Consumer Health Vocabulary (OAC-CHV) seen in Section 2.2. The mapping from technical to consumer terms is accomplished by means of the Concept Unique Identifier (CUI) when available (Keselman, 2008). This is the case of the technical terms that are found in the UMLS.

Notice that, since there is no Italian CHV available, beside the ‘Italian Consumer-oriented Medical Vocabulary’ (ICMV) that only contains a few items (Italian Consumer-oriented Medical Vocabulary, 2014), we proceeded in translating the OAC-CHV (with its 160,000 entries) from English to Italian. In particular we have translated the ‘Terms’ and the ‘CHV Preferred Names’ elements described in Section 2.2. This has been done by using the UMLS multilingual vocabularies, the

English-Italian translation of the ‘Dizionario di Medicina e Biologia’ and Google Translator. To improve the effectiveness of the Google translation, we have replaced a term translated by Google with a UMLS term with the same CUI when their ‘distance’, computed with the Levenshtein distance algorithm (Levenshtein, 2014), was short.

The DEFINE module uses the Italian consumer health dictionaries ‘Ok Salute’ (around 8,000 entries) and ‘Dizionario della Salute’ (around 6,000 entries) to find the definitions of the technical terms. When a term is not found in these dictionaries, the definition is searched in the ‘Dizionario di Medicina e Biologia’ (around 60,000 entries) although its definitions have a more technical nature. If a definition is not found in any of the dictionaries, we use the ‘Google define’ keyword to provide a definition in any case.

4.2 Practical use

We have implemented the system and created a prototype that has been tested with some Italian medical reports. It can be found at the address <http://math.unipa.it/simplehealth/simple>.

Figure 2 shows a snapshot of the prototype that presents an input form where it is possible to insert any (Italian) medical text and process it or choose among some Italian medical reports. In the case of Figure 2, we have loaded a magnetic resonance (MRI) report.

Inserisci testo

RM encefalo (senza e con contrasto) del 06.06.2014:

Indicazione all'esame: evidenza TC di sfumata opacità tondeggianti, associata a ipodensità perilesionale in sede cortico-sottocorticale occipitale sx in paziente con progresso K uterino ed ipertensione polmonare.

Dati tecnici: indagine eseguita in condizioni di base con sequenze DP e T2 FSE, FLAIR, T1 FSE, T1 FSPGR, DWI e dopo somministrazione di 10 ml di MdC (Multihance; T1 FSPGR).

Stima del GFR pari a 71.

Sia prima che dopo la somministrazione del MdC, non si apprezzano aree di alterato segnale a carico del parenchima nervoso cerebrale sovra e sottotentoriale.

Non patologiche aree di restrizione del segnale nella sequenza DWI.

Spazi liquorali peri-encefalici della volta e della base di ampiezza conservata.

Sistema ventricolare sovra e sottotentoriale di regolare morfologia e dimensioni.

Strutture della linea mediana in asse.

Input:

submit

Seleziona uno dei referti disponibili

Referto:

Figure 2. Input form for medical text processing.

Figure 3 shows the same medical report where the technical terms are highlighted after being processed through our system. Next to each technical term, we find one or two info buttons, i.e., a question mark icon for the consumer translation of

the term (in a tooltip next to the term) and an information icon for its definition (in a frame below). In particular, The word *encefalo* (*encephalon* in English) is selected and its consumer translation, *cervello* (*brain* in English), is shown together with its explanation in the below frame.

1 Referto

encefalo

RM (encefalo) (senza e con contrasto) del 06.06.2014:

Indicazione all'esame: evidenza TC di sfumata opacità tondeggianti, associata a ipodensità perilesionale in sede cortico-sottocorticale occipitale sx in paziente

Dati tecnici: indagine eseguita in condizioni di base con sequenze DP e T2 FSE, FLAIR, T1 FSE, T1 FSPGR, DWI e dopo somministrazione di 10 ml di MdC

Stima del GFR pari a 71.

Sia prima che dopo la somministrazione del MdC, non si apprezzano aree di alterato segnale a carico del parenchima nervoso cerebrale sovra e sottotentoriale.

Non patologiche aree di restrizione del segnale nella sequenza DWI.

Spazi liquorali peri-encefalici della volta e della base di ampiezza conservata.

Sistema ventricolare sovra e sottotentoriale di regolare morfologia e dimensioni.

Strutture della linea mediana in asse.

Conclusioni: non evidenza di reperti patologici.

Encefalo

Parte del sistema nervoso centrale contenuta nella cavità cranica che in unione con la parte caudale midollo spinale costituisce l'asse cerebro-spinale. Il cervello è il tronco encefalico costituito da mesencefalo, ponte di Vairoli vedi Vairolo ponte e midollo allungato o bulbo spinale.

Figure 3. Medical report with highlighted technical terms.

From a first analysis, our system is capable of finding (and highlighting) most of the technical terms present in the medical reports together with the corresponding consumer terms (even though they need further verification) and provide the related definitions. Of course, more experiments are needed and we are in the process of executing them with a group of physicians and patients.

Notice that, as said above, our system does not create any change in the original text by replacing the word or inserting an explanation in the text because, in our opinion, this could disorient the user. It only provides a translation (as a tooltip) and additional info (on a separated frame) on request, leaving the user fully in charge of his/her navigation path through the text as it was originally created.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a system that, given a medical text, automatically finds the technical terms, translates them in plain language and provides additional information for them with the same kind of language. We have implemented the system and built a first prototype that is working and provides satisfying results but needs some improvements.

As a first priority, the medical vocabulary needs to be expanded to be able to find more technical terms present in the text. Very important is the completion of the Italian CHV by adding more

technical terms and their consumer equivalents. Finally, other health-consumer dictionaries have to be found for increasing the number of definitions that come from medical sources.

A potential extension of the system comes from providing the user with direct access to an external web search engine, either a generic one (i.e. Google or Bing) or a specific one (such Quertle or Pubmed), for finding further information beside the ones already provided. Also, consumer-oriented sites could be directly accessed from the system so to allow the user an easy navigation inside familiar information.

We plan to complete the prototype and integrate it within an Electronic Health Record (EHR) or Personal Health Record (PHR) by using the HL7 “Infobutton” standard. Moreover, we want to extend it so that it is able to take as an input the URL of any health web page and automatically provide, as output, the same page with the highlighted technical terms together with their consumer translations and definitions so to help a generic user in understanding any health-related web page.

ACKNOWLEDGEMENTS

This work was partially funded by the PON Smart Cities PON04a2_C “SMART HEALTH – CLUSTER OSDH – SMART FSE-STAYWELL” project.

REFERENCES

- Baorto, D.M. & Cimino, J.J., 2000. An “infobutton” for enabling patients to interpret on-line Pap smear reports. *AMIA Annual Symposium Procs.*, pp.47–50.
- Cimino, J., 2007. An integrated approach to computer-based decision support at the point of care. *Trans. of the American Clinical and Climatological Ass.*, 118, pp.273–288.
- Consumer Health Vocabulary Initiative. <Available from: <http://consumerhealthvocab.org/>>. [6 October 2014].
- Dizionario della salute. <Available from: <http://www.corriere.it/salute/dizionario/>>. [6 October 2014].
- Dizionario di Medicina e Biologia. <Available from: <http://medicina.zanichellipro.it/>>. [6 October 2014].
- HL7 Version 3 Implementation Guide: Context-Aware Knowledge Retrieval Application (Infobutton), Release 4. <Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=22>. [6 October 2014].
- HL7 Product Infobutton. <Available from: http://wiki.hl7.org/index.php?title=Product_Infobutton>. [6 October 2014].
- Infobuttons. <Available from: <http://clinfowiki.org/wiki/index.php/Infobuttons>>. [6 October 2014].
- Italian Consumer-oriented Medical Vocabulary. <Available from: <https://ehealth.fbk.eu/resources/italian-consumer-oriented-medical-vocabulary-icmv>>. [6 October 2014].
- Kandula, S., Curtis, D. & Zeng-Treitler, Q., 2010. A semantic and syntactic text simplification tool for health content. *AMIA Annual Symposium Procs.*, pp.366–370.
- Kemper, D. et al., 2010. Getting patients to meaningful use: using the HL7 infobutton standard for information prescriptions. *Healthwise White Paper Series*.
- Keselman, A. & Slaughter, L., 2007. Towards consumer-friendly PHRs: patients’ experience with reviewing their health records. *AMIA Annual Symposium Procs.*, pp.399–403.
- Keselman, A., Smith, C. A., 2008 et al.. “Consumer health concepts that do not map to the UMLS: where do they fit?”, *Journ. Am. Med. Inform. Ass.*, vol. 15, no. 4, pp. 496-505.
- Levenshtein Distance Algorithm. <Available from: http://en.wikipedia.org/wiki/Levenshtein_distance>. [6 October 2014].
- MedlinePlus. <Available from: <http://www.nlm.nih.gov/medlineplus/healthtopics.html>>. [6 October 2014].
- Ok Salute. <Available from: <http://www.ok-salute.it/dizionario-medico>>. [6 October 2014].
- Plain Language.gov. <Available from: <http://www.plainlanguage.gov/>>. [6 October 2014].
- Seedorff, M. & Peterson, K., 2013. Incorporating Expert Terminology and Disease Risk Factors into Consumer Health Vocabularies. *Pacific Symposium in Biocomputing*, pp.421–432.
- Simple English Wikipedia. <Available from: https://en.wikipedia.org/wiki/Simple_English_Wikipedia>. [6 October 2014].
- Unified Medical Language System. <Available from: <http://www.nlm.nih.gov/research/umls/>>. [6 October 2014].
- WebMD. <Available from: <http://www.webmd.com/a-to-z-guides/common-topics/default.htm>>. [6 October 2014].
- Zeng, Q. & Tse, T., 2006. Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics*, pp.24–29.
- Zeng-Treitler, Q. & Goryachev, S., 2007. Making Texts in Electronic Health Records Comprehensible to Consumers: A Prototype Translator. *AMIA Annual Symposium*, pp.846–850.
- Zielstorff, R.D., 2003. Controlled vocabularies for consumer health. *Journal of Biomedical Informatics*, 36(4-5), pp.326–333.