

# A HYBRID METHODOLOGY FOR CONSUMER-ORIENTED HEALTHCARE KNOWLEDGE ACQUISITION

Elena Cardillo, Andrei Tamin and Luciano Serafini  
FBK-IRST, via Sommarive 18, 38100 Povo (TN) , Italy

**Keywords:** Medical Terminology, Medical Vocabulary Acquisition, Knowledge Acquisition, Consumer Healthcare.

**Abstract:** In spite of the improvements in Healthcare Informatics in answering consumer needs, it is still difficult for laypersons who do not have a good level of healthcare literacy, to find, understand, and act on health information. This is due to the communication gap which still persists between specialized medical terminology used by healthcare professionals and “lay” medical terminology used by healthcare consumers. So there is a need to create consumer-friendly terminologies reflecting the different ways consumers and patients express and think about health topics. An additional need is to map these terminologies with existing clinically-oriented terminologies. Following this direction, this work suggests a hybrid methodology to acquire consumer health terminology for creating a Consumer-oriented Medical Vocabulary for Italian that mitigates this gap. This resource could be used in Personal Health Records to provide translation, search, and classification services, helping users to improve access to their healthcare data. In order to evaluate this methodology we mapped “lay” terms with standard specialized terminologies to find overlaps. Results showed that our methodology of acquisition provided many “lay” terms that can be considered good synonyms for medical concepts.

## 1 INTRODUCTION

With the advent of the Social Web and Healthcare Informatics technologies, we can recognize that a linguistic and semantic discrepancy still exists between specialized medical terminology used by healthcare providers or professionals, and the so called “lay” medical terminology used by healthcare consumers. The medical communication gap became more evident when consumers started to play an active role in healthcare information access. In fact they have become more responsible for their personal healthcare, exploring information sources on their own, consulting decision-support healthcare sites on the web, and using patient-oriented healthcare systems, which allow them to directly read and interpret clinical notes or test results and to fill in their Personal Health Record. During this disintermediated interaction consumers can use only their own knowledge, experience and preferences, and this can often generate a wrong inference of the meaning of a term, or the mis-association of a term with its context (Zeng and Tse, 2006).

To help consumers fill this gap, the challenge is to sort out the different ways consumers communicate within distinct discourse groups and map the common, shared expressions and contexts to the more constrained, specialized language of healthcare professionals. Though much effort has been spent on the creation of medical resources such as terminologies or classification systems, used above all to help physicians in filling in Electronic Health Records, there is little work based on the use of consumer-oriented medical terminology, and in addition most existing studies have been done only for English.

A consumer-oriented medical terminology can be defined as a “*collection of forms used in health-oriented communication for a particular task or need by a substantial percentage of consumers from a specific discourse group and the relationship of the forms to professional concepts*” (Zeng and Tse, 2006). Such terminology can be mainly used for three bridging roles between consumers and health applications: a) Information Retrieval, to facilitate automated mapping of consumer-entered queries to technical terms, producing better search results; b)

Translation of Medical Records, supplementing medical jargon terms with consumers-understandable names to help patients interpretation; c) Health Care Applications, to help integrating different medical terminologies providing automated mapping of consumer expressions to technical concepts (e.g. querying for “short of breath” and receiving information also for the concepts “dyspnea”).

Given this scenario, the present work proposes a hybrid methodology for the acquisition of consumer-oriented medical knowledge and “lay” terminology expressing particular medical concepts, such as symptoms and diseases, for consequent creation of a Consumer-oriented Medical Vocabulary for Italian. We are particularly interested in performing analysis of the clinical mapping between this consumer-oriented terminology and the more technical one used in standards Medical Classification Systems or Nomenclatures. This resource could be integrated with existing lexical and semantic medical resources, and used in healthcare systems, like Personal Health Records, to help consumers during the process of querying and accessing healthcare information, so as to bridge the communication gap.

## 2 BACKGROUND

### 2.1 Consumer-oriented Medical Terminologies

Over the last two decades research on Medical Terminologies has become a popular topic and the standardization efforts have established a number of terminologies and classification systems such as UMLS Metathesaurus<sup>1</sup>, SNOMED International<sup>2</sup>, ICD-10 (International Classification of Diseases)<sup>3</sup> and ICPC-2 (International Classification of Primary Care)<sup>4</sup>, as well as conversion mappings between them to help medical professionals in managing and codifying their patients health care data. They concern, in fact, “the meaning, expression, and use of concepts in statements in the medical records or other clinical information systems” (Rector, 1999). Despite the wide use of these terminologies, as we have already mentioned, the vocabulary problem continues to plague not only health professionals and

their information systems, but also consumers and in particular laypersons, who are the most damaged by the increased communication gap.

To respond consumer needs to support personal healthcare decision-making, during the last few years, many researchers have labored over the creation of lexical resources that reflect the way consumers/patients express and think about health topics. One of the largest initiatives in this direction is the Consumer Health Vocabulary Initiative<sup>5</sup>, by Q. Zeng and colleagues at Harvard Medical School, resulted in the creation of the Open Access Collaborative Consumer Health Vocabulary (OAC CHV) for English. It includes lay medical terms and synonyms connected to their corresponding technical concepts in the UMLS Metathesaurus. They combined corpus-based text analysis with a human review approach, including the identification of consumer forms for “standard” health-related concepts. Also (Soergel *et al.*, 2004) tried to create such a vocabulary to identify consumer medical terms and expressions used by lay people and health mediators. In fact they associated a Mediator Medical Vocabulary with the consumer-oriented one, and mapped them to a Professional Medical Vocabulary. Even in this case the standard terminology used for mapping was UMLS.

These and other similar studies examined large numbers of consumer utterances (i.e., hundreds of thousands of tokens) and consistently found that between 20% and 50% of consumer health expressions were not represented by professional health vocabularies. Furthermore, a subset of these unrepresented expressions underwent human review. In most of these cases they performed automatic term extraction from written texts, such as healthcare consumer queries on medical web sites, postings and medical publications. An overview of all these studies can be found in (Keselman *et al.*, 2008).

It is important to stress that there are only few examples of the applications as far as these initiative are concerned. For example, in (Kim *et al.*, 2007) we find an attempt to face syntactic and semantic issues in the effort to improve PHR readability, using the CHV to map content in Electronic Health Records (EHR) and PHR; (Zeng *et al.*, 2007) designed and implemented a prototype text translator to make EHR and PHR more comprehensible to consumers and patients. On the

<sup>1</sup> <http://www.nlm.nih.gov/research/umls/>

<sup>2</sup> <http://www.ihtsdo.org/snomed-ct/>

<sup>3</sup> <http://www.who.int/classifications/icd/en/>

<sup>4</sup> <http://www.fmrc.org.au/icpc2/>

<sup>5</sup> <http://www.consumerhealthvocab.org>

other hand (Rosebloom *et al.*, 2007) developed a clinical interface terminology, a systematic collection of healthcare-related phrases to support clinicians' entries of patient-related information into computer programs such as clinical "note capture" and decision support tools, facilitating display of computer-stored patient information to clinician-users as simple human-readable texts.

Concerning multilingual consumer-oriented health vocabularies, we can mention the initiative of the European Commission Multilingual Glossary of Popular and Technical Medical Terms<sup>6</sup>, in nine European languages, but it is a limited medical vocabulary for medicinal product package inserts accessible to consumers. In fact, it consists of a list of 1,400 technical terms frequently encountered in inserts, with corresponding consumer terms in all the languages of the European Community. Greater overlap between technical and lay terms was observed for Romance languages and Greek than for Germanic languages (except English), and some technical terms had no lay equivalent.

## 2.2 Knowledge Acquisition in Healthcare

Knowledge Acquisition process aims at identifying and capturing knowledge assets and terminology to populate a knowledge repository for a specific domain. Central areas of this task are: terminology work, relevant for the special subject field, including terminography; content analysis of documents; extraction of knowledge from various sources. A major part of Knowledge Acquisition is capturing knowledge from experts, a task that is made cost-effective and efficient by using knowledge models and special elicitation techniques. These techniques should be used in different phases of the process, since each of them permits the capture of a specific typology of knowledge and the achievement of specific aims. The most common techniques are *Interviews*, direct observation of expert performances to extract procedural knowledge, mostly connected to manual skills, such as *Think Aloud Problem Solving*, *Self-report*, and *Shadowing*. Other techniques, such as *Card Sorting*, *Repertory Grid*, and *Twenty Questions*, are useful for understanding how experts conceptualise knowledge related to their own domain of reference (Milton, 2007).

<sup>6</sup> <http://users.ugent.be/~rvdstich/eugloss/welcome.html>

In a task of Knowledge Acquisition, it is important to identify two main components: knowledge types, referred to knowledge orientation and domain; and modalities, referred to the representation medium in which knowledge exists. In Knowledge Acquisition for the Healthcare domain, according to (Abidi, 2008), many different types of knowledge, which directly contribute to clinical decision-making and care planning, can be identified: Patient, Practitioner, Medical, Resource, Process, Organizational, Relationship and, finally, Measurement Knowledge. In the present work we will only deal with Medical Knowledge and Patient Knowledge. These knowledge types are represented by different knowledge modalities. The most common ones are: Tacit Knowledge of practitioner, Explicit Knowledge, Clinical experiences, Collaborative Problem-solving discussions, Social Knowledge, etc. In our work we focus on Explicit Knowledge, Clinical Experience and Social Knowledge. In particular the last modality can be viewed in terms of a community of practice and their communication patterns, interest and expertise of individual members.

## 3 METHODOLOGY

In this study we focused on the acquisition of consumer-oriented knowledge about a specific subset of healthcare domain that includes Anatomy, Symptomatology and Pathology. For that task we have chosen to use a hybrid methodology for the identification of "lay" terms, words, and expressions used by Italian speakers to indicate Symptoms/Signs, Diseases and Anatomical Concepts. Three different target groups were considered for the application of our approach: First Aid patients subjected to a Triage Process; a community of Researchers and PhD students with a good level of healthcare literacy, and finally a group of elderly people with a low level of healthcare literacy. The proposed methodology consists of the following steps:

1. Familiarization with the domain and exploitation of existing common lexical resources (Glossaries, Thesauri, Medical Encyclopedias, etc);
2. Choice and application of three different Elicitation Techniques to the mentioned groups of people:
  - a. Collaborative Wiki-based Acquisition;
  - b. Nurse-assisted Acquisition;

- c. Interactive Acquisition combining traditional elicitation techniques (Focus Groups, Concepts Sorting and Games);
3. Automatic Term Extraction and analysis of acquired knowledge by means of a Text Processing tool;
4. Clinical review of extracted terms and manual mapping to a standard medical terminology, performed by physicians;
5. Evaluation of results in order to find candidate terms to be included in the Consumer-oriented Medical Vocabulary.

### 3.1 Wiki-based Acquisition

The first method for acquiring consumer-oriented medical knowledge is based on the use of a Semantic Media Wiki system<sup>7</sup>, an easy to use collaborative tool, allowing users to create and link, in a structured and collaborative manner, wiki pages on a certain domain of knowledge. Using our online eHealthWiki<sup>8</sup> system, users created wiki pages for describing symptoms and diseases, using “lay” terminology, specifying in particular the corresponding anatomical categorization, the definition and possible synonyms.

The system has been evaluated over a sample of 32 people: researchers, PhD students and administrative staff of our research institute (18 females, 14 males, between 25 and 56 years old). In one month, we collected 225 wiki pages, 106 for symptoms and 119 for diseases, and a total of 139 synonyms for the inserted terms. It was very interesting to test here also the understanding of the collaborative nature of the Wiki for the specific task, which gave users the possibility not only to insert medical terms by creating wiki pages, but also to update, cancel or correct the inserted information, and above all to modify wiki pages added by other users, in order to reach a convergence on the common sense of medical terminology. In our case, users were reluctant to modify concepts added by others, even in the case of evident mistakes in definitions or categorization (only 7 people out of 32 provided changes to wiki pages). Some examples of categorization mistakes that had not been modified are “Singhiozzo” (Hiccup), and “Mal di Testa” (Headache), both categorized as Diseases instead of Symptoms. In some cases, when users were in doubt about the right categorization of a concept, they inserted it in both the categories, e.g., “Ustione”

(Burning). This test highlighted the fact that users had problems in categorizing medical terms - mainly due to their clinic ambiguity - and also the erroneous use of these terms by them daily.

### 3.2 Nurses-assisted Acquisition

The second acquisition technique involved the nurses of a First Aid Unit (in a Hospital of the Province of Trento)<sup>9</sup>, as figures of mediation for the acquisition of terminology about patient symptoms and complaints. Nurses here help patients to express their problems using the classical subjective examination performed during the Triage Process. The Triage activity has the aim of prioritizing patients according to the severity of their condition, on the basis of examination costing few minutes. This acquisition method involved 10 nurses, around 60 patients per day and a total of 2.000 Triage Records registered in one month. During this period nurses acquired the principal problems (symptoms and complaint) expressed by their patients using “lay” terminology and inserted them in the Triage Record together with the corresponding medical concepts usually used for codifying patient data. For example, the lay expression “Ho i crampi alla pancia” (I have a stomach ache) was inserted in the Triage Record together with the corresponding medical concept “Addominalgia” (Abdominal pain).

### 3.3 Focus-group Acquisition

The last method used in our study consisted in merging three different traditional elicitation techniques: Focus Group, Concepts Sorting, and the more experimental Board Games, in order to allow interaction and sharing situations to improve the process of acquisition. We have applied this techniques to a community of 32 elderly persons in a Seniors Club in the Province of Trento, between 65 and 83 years old.

During the process of acquisition we divided participants into 4 groups, of 8 people each, assigning to each one a specific body part category (for instance, head and neck, abdomen and back, arms, hands and chest, pelvic area, legs and feet). Each group was asked to write on little cards all known symptoms and diseases related to the assigned area, starting from personal experience, and comparing their idea with other members of the

<sup>7</sup> [http://semantic-mediawiki.org/wiki/Semantic\\_MediaWiki](http://semantic-mediawiki.org/wiki/Semantic_MediaWiki)

<sup>8</sup> <http://ehealthwiki.fbk.eu>

<sup>9</sup> Medicina d’Urgenza e Pronto Soccorso del Presidio Ospedaliero di Cles (Trento): <http://www.apss.tn.it/Public/ddw.aspx?n=26808>

group to find a common definition for the written terms. Time allowed for card writing was about 20 minutes. About 160 medical terms were collected. Then, all these terms were analyzed together, creating discussions, exchanging opinions on term definitions, synonyms, and recording preferences and common sense. At the end of the discussion about each medical concept, all participants gave preferences for choosing the right body system categorization of that concept. For this categorization we provided a panel with 14 different problem areas and body systems (digestive, neurological, musculoskeletal, lymphatic, endocrine, etc.). This allowed us not only to collect lay terminology used by elderly people, but also to understand how they define and categorize medical concepts, in order to compare these results with these obtained with the other two techniques mentioned above.

#### 4 TERM EXTRACTION

Three sets of collected data, including the transcription of the Focus Group activity with elderly persons, were further processed and analyzed, to detect candidate consumer-oriented terms, with the tool Text-2-Knowledge (T2K), developed at the Institute of Computational Linguistics of Pisa<sup>10</sup>. This tool allowed us to automatically extract terminology from the data sets and to perform typical text processing techniques (normalization, pos tagging, chunking, etc.), calculating, in addition, statistics such as term frequency on the extracted data. The computational system adopted by the tool includes a specific plugin for the analysis of Italian. It provides, as final output, a term-based vocabulary whose added value is represented by the terms' semantic and conceptual information regarding the vocabulary itself. These terms, which can be either single or multi-word terms, are organized in a hierarchical hyponym/hyperonym relation depending on the internal linguistic structure of the terms (Bartolini et al., 2005); that is, by sharing the same lexical head.

In spite of the advantages of the automatic extraction process, allowing for extraction of many compound terms, a good amount of terms, certainly representative of consumer medical terminology, were not automatically extracted, since, due to the quantitative limits of the corpus dimensions, their

occurrence was inferior with respect to the predefined threshold value. Consequently, we performed an additional manual extraction to take into account such rare terms, usually mentioned by a single participant. Statistical results about the three different data sets are further discussed in Section 6.

#### 5 CLINICAL REVIEW

Term extracted by T2K were reviewed by two physicians to find errors and incongruities in categorization and synonymy. Many mistakes were found in the first set (Wiki-based), where a wrong categorization was assigned to 25 terms, and were wrong synonyms were expressed for 8 terms. Many mistakes were also found in the third set (Elderly people), where wrong categorization were assigned to 40 terms, e.g., "Giramento di Testa" or "Vertigini" (Vertigo or Dizziness), in the *Cardiovascular System* instead of the right *Neurological* one. Concerning the second data set, clinical review was performed during the process of Triage by a nurse and a physician.

During the second part of our clinical review, physicians have been asked to map a term/medical concept pair by using a professional medical terminology - in this study the International Classification for Primary Care 2<sup>nd</sup> Edition (ICPC2). ICPC2 addresses fundamental parts of healthcare process: it is used in particular by general practitioners for encoding symptoms and diagnosis. It has a biaxial structure that consider medical concepts related to Symptoms, Diseases and Diagnoses, and Medical Procedure, according to 17 Problem Areas/Body Systems. In previous work we encode ICPC-2-E using a recently developed Web Ontology Language (OWL)<sup>11</sup> (both for English and Italian), that also provides the formalization of the existing clinical mapping with the ICD10 classification system, as shown in (Cardillo *et. al.*, 2008). By means of this mapping we want to reconstruct the meaning (concept) inherent in the lay usage of a term, and then to agree that consonance between lay and professional terms exists on the basis of this deeper meaning, rather than the lexical form. Five different types of relations are possible between consumer terms and ICPC2 medical concepts:

- Exact Mapping between the pairs; this occurs when the term used by a lay person

<sup>10</sup> <http://www.ilc.cnr.it>

<sup>11</sup> <http://www.w3.org/TR/owl-features/>

can be found in ICPC2 terminology and both terms correspond to the same concept. E.g., the lay term “Febbre” (Fever) would map to the ICPC2 term “Febbre”, and both will be rooted to the same concept.

- **Related Mapping;** it involves lay synonyms and occurs when the lay term does not exist in the professional vocabulary, but corresponds to a professional term that denotes the same (or closely related) concept. E.g., lay term “Sangue dal Naso” (Nosebleed) corresponds to “Epistassi” (Epistaxis).
- **Hyponymy Relation;** this occurs when a lay term can be considered as term of inclusion of a ICPC2 concept. E.g., lay term “Abbassamento della Voce” (Absence of Voice) is included in the more general ICPC2 concept “Sintomo o disturbo della voce” (Voice Symptom/Complaint).
- **Hyperonymy Relation;** in this case the lay term is more general than one or more ICPC2 concepts, so it can be considered as its/their hyperonym. E.g., the term “Bronchite” (Bronchitis) is broader than “Bronchite Acuta/ Bronchiolite” (Acute Bronchitis/ Bronchiolitis) e “Bronchite Cronica” (Chronic Bronchitis) ICPC2 concepts.
- **Not Mapped;** it comprises those lay terms that cannot be mapped to the professional vocabulary. These can be legitimate health terms whose omission reflects real gaps in existing professional vocabularies; or they can represent unique concepts reflecting lay models of health and disease. E.g., the lay term “Mal di mare” (Seasickness).

## 6 RESULTS EVALUATION

As previously mentioned, our methodology of acquisition allowed us to acquire varied consumer-terminology and to perform an interesting terminological and conceptual analysis. Tables 1-4 provide term extraction and mapping evaluation in terms of a statistical analysis. By means of the term extraction process, we were able to extract a total of 692 medical terms from 225 Wiki pages, 375 of which were not considered pertinent to our aim. We performed mapping analysis on 587 terms as summarized in Table 1.

Table 1: Wiki term collection.

	Tot. Terms	Exact Map.	Related Map.	Iponym. Relation	Iperonym. Relation
Symptoms	306	26	50	40	9
Diseases	140	42	19	38	38
Anatomy	141	105	11	16	4
Other	375	/	/	/	/
Tot.	962				
Not Mapped.	186	Not Considered:375			

We can observe that most of the exact mappings with ICPC2 are related to anatomical concepts, and which many synonyms in lay terminologies and inclusion terms were found for symptoms. Table 2 shows the results related to the Triage acquisition data.

Table 2: Nurses-assisted term collection.

	Tot. Terms	Exact Map.	Related Map.	Not Mapped
Symptoms	508	134	197	177
Diseases	325	86	94	145
Anatomy	275	120	95	60
Other	1281	/	/	/
Tot.	2389	Not Considered: 1281		

We extracted a total of 2389 terms from 2.000 Triage records, but about half of these terms were considered irrelevant for our evaluation, so mapping was provided only for 1108 terms. Contrary to the previous results, here is interesting to highlight the high presence of lay terms used for expressing symptoms with exact mappings to ICPC2, but also many synonyms in lay terminology for ICPC2 symptoms and diseases. This is particularly related to the context chosen for the acquisition, where patients just ask for help about suspected symptoms and complaints. Table 3 shows the results related to the data acquisition from Elderly people.

Table 3: Focus Group /Game with Elderly Persons.

	Tot. Terms	Exact Map.	Related Map.	Not Mapped
Symptoms	79	35	44	/
Diseases	87	29	54	4
Anatomy	77	51	18	8
Other	68	/	/	/
Tot.	321	Not Considered: 78		

Concerning the last data set, 321 medical terms were extracted by the transcription of the Focus Group/Game activity. Here is interesting to note that all the symptoms extracted had corresponding medical concept in ICPC2 terminology.

Table 4 compares the three data set together and shows that the most profitable methods for acquiring consumer-oriented medical terminology was the one assisted by nurses. But the limit of this method is that it is time-consuming for nurses who have to report all the patient “lay” health expressions. While Wiki-based method, even if not exploited for the collaborative feature, has demonstrated good qualitative and quantitative results. Furthermore, are interesting the results concerning mapping to ICPC2, because 2/3 of the lay terms collected are covered by ICPC2 terminology.

Table 4: Results Overview.

Sources	Tot. Terms	Tot. Mapped	Not Mapped
eHealthWiki	962	398	186
Nurse-assisted	2389	726	382
Focus Groups	321	231	12
Tot.	3662	1355	580

To conclude our evaluation we have to highlight that comparing the three sets of collected terms, the overlap is only of 60 relevant consumer medical terms. The total overlap with ICPC2 is about 508 medical concepts on a total of 706 ICPC2 concepts. This means that all the other mapped terms can be considered synonyms or quasi synonyms of the ICPC2 concepts. The large number of not mapped terms and the low overlap between the three sets of extracted terms demonstrate that we collected a very variegated range of medical terms, many compound terms and expressions. These terms can be representative of the corresponding technical terms present in standard medical terminologies, and can be used as candidate terms for the construction of our Consumer-oriented Medical Vocabulary for Italian.

## 7 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a hybrid methodology for acquiring consumer-oriented

medical Knowledge and Terminology for Italian, consisting of lay expressions and terms used to indicate Symptoms, Diseases and Anatomical Concepts. We applied three explorative elicitation techniques to three different groups of people, and we compared results on the basis of a term extraction process, for statistical analyses, and on a clinical mapping procedure, for finding overlaps between extracted lay terms and specialized medical concepts in the ICPC2 medical terminology. Our methodology showed encouraging results, because it allowed us to acquire many consumer-oriented terms, to find a low overlap with medical concepts and a high number of related mappings (most of the time synonyms) to the ICPC2 terminology.

Taking each of these acquisition techniques alone, they do not allow a good coverage of the whole domain of pathology and symptomatology. In the first case, in fact, most of the terms provided by Researchers and PhD students are related to *Digestive* and *Musculoskeletal Systems*, and *Skin*, while in the case of the Triage activity patients expressed most of the time symptoms related to *Musculoskeletal System* (due to the geographic context and period of the acquisition, i.e. mountain-skiing area, end of winter), *Respiratory System*, and *Cardiovascular System*. Furthermore, also in the third acquisition, terms are mostly related to the these Body Systems. Using a hybrid approach in merging these techniques and involving a more varied sample of people would improve the results, both from the qualitative and the quantitative point of view. Another limit could be seen in the process of manual mapping performed by physicians. After this pilot study we plan to implement a semi-automatic procedure for mapping lay and specialized terminology, which will be associated to the process of automatic term extraction, and validated by the review of physicians.

In the approaches described in Section 2, consumer-oriented health vocabularies were developed by working only on big written corpora (forum postings and queries to medical websites), using machine learning algorithm and statistical methods (naive Bayesian classifiers, C-value, Logistic Regression etc.) to extract consumer-oriented terminology. In comparison to these approaches we gave more importance to qualitative data, focusing on different methods for acquiring not merely lay terminology but also knowledge directly from consumers in different scenarios related to General Practice. This allowed also to try to

understand how consumers make use of medical terminology, how common expressions daily used in health communication really match onto medical concepts used by professionals.

To improve the results of the Knowledge Acquisition process and to extract more variegated consumer terminology, not related to the regional context, one of the future tasks is to perform a Knowledge Acquisition Process involving people in a Social Network. This would allow to extend our sample, including younger people. This task would be very interesting for comparing results with what resulted from the previous methodologies. Another important improvement would be the analysis of written texts such as forum postings of an Italian medical website for asking questions to on-line doctors<sup>12</sup>. Data extracted in this way could also be used to validate the acquired verbal terminology, by providing preferences between terms according to frequency and familiarity score.

## ACKNOWLEDGEMENTS

We would like to thank Antonio Maini and Maria Taverniti, who provided us with useful support respectively in the process of Knowledge Acquisition and Term Extraction.

## REFERENCES

- Abidi, S.S.R., 2007. Healthcare Knowledge Management: The Art of the Possible. In Proceedings of the Knowledge Management for Health Care Procedures' Conference, K4CARE 2007, Springer Berlin, pages 1-20.
- Bartolini, R., Lenci, A., Marchi, S., Montemagni, S., and Pirrelli, V., 2005. *Text-2-knowledge: Acquisizione semi-automatica di ontologie per l'indicizzazione semantica di documenti*, Technical Report for the PEKITA Project, ILC. Pisa p. 23;
- Cardillo, E., Eccher, C., Taminin, A., and Serafini, L., 2008. Logical Analysis of Mappings between Medical Classification Systems. In Proceedings of the 13<sup>th</sup> International Conference on Artificial Intelligence: Methodology, Systems, and Applications, AIMS2008, Springer Berlin, pages 311-321
- Keselman, A., Logan, R., Smith, C. A., Leroy, G., and Zeng, Q., 2008. Developing Informatics Tools and Strategies for Consumer-centered Health Communication. *Journal of the American Medical Informatics Association*, 14(4):473-483.
- Kim, H., Zeng, Q., Goryachev, S., Keselman, A., Slaughter, L., and Smith, C.A., 2007. Text Characteristics of Clinical Reports and Their Implications for the Readability of Personal Health Records. In Proceedings of the 12<sup>th</sup> World Congress on Health (Medical) Informatics, MEDINFO2007, IOS Press, pages 1117-1121
- Milton, N. R., 2007. *Knowledge Acquisition in Practice: A Step-by-step Guide*. London, Springer;
- Rector, A. (1999). Clinical Terminology: Why is it so hard?. *Methods of Information in Medicine*, 38(4):239-252.
- Rosebloom, T.S., Miller, R.A., Johnson, K.B., Elkin, P.L., and Brown, H. S., 2006. Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems. *Journal of American Medical Informatics Association*, 13(3):277-287.
- Smith, B. and Rosse, C., 2004. The Role of Foundational Relations in the Alignment of Biomedical Ontologies. In Proceedings of the 28<sup>th</sup> American Medical Informatics Association's Annual Symposium, AMIA2004.
- Soergel, D., Tse, T., and Slaughter, L., 2004. Helping Healthcare Consumers Understand: An "Interpretative Layer" for Finding and Making Sense of Medical Information. In Proceedings of the International Medical Informatics Association's Conference, IMIA2004, pages 931-935.
- Zeng, Q., Goryachev, S., Keselman, A., and Rosendale, D., 2007. Making Text in Electronic Health Records Comprehensible to Consumers: A Prototype Translator. In Proceedings of the 31<sup>st</sup> American Medical Informatics Association's Annual Symposium, AMIA2007, pages 846-850.
- Zeng Q. and Tse, T., 2006. Exploring and Developing Consumer Health Vocabularies. *Journal of the American Medical Informatics Association*, 13:24-29

<sup>12</sup> <http://medicitalia.it>