Soc. Netw. Anal. Min. (2014) 4:225
DOI 10.1007/s13278-014-0225-1

ORIGINAL ARTICLE

A medical social network for physicians' annotations posting and summarization

Jalel Akaichi

Received: 5 January 2014 / Revised: 20 July 2014 / Accepted: 29 July 2014 / Published online: 19 August 2014

Abstract Patients are often anxious to quickly discover reliable analysis and concise explanation of their medical images while waiting for the physician decision. The fact of making important choices individually in his own corner may lead the physician to commit errors leading to malpractices and consequently to unforeseeable damages. In order to minimize medical errors by fostering collaboration between physicians and/or patients, we propose in this paper, as a first contribution, a medical social network destined to gather patients' medical images and physicians' annotations expressing their medical reviews and advices. The need, to automatically extract information and analyze opinions, becomes obviously a requirement due to the huge number of comments expressing specialists' recommendations and/or remarks. For this purpose, we propose a second contribution consisting of providing a kind of comments' summary which extracts the major current terms and relevant words existing on physicians' reports. Furthermore, this extracted information will present a new and robust input for image indexation enhanced methods. In fact, significant extracted terms will be used later to index images in order to facilitate their search through the underlying social network. To overcome the above challenges, we propose an approach which focuses on algorithms mainly based on statistical methods and external semantic resources destined to filter selected extracted information.

Keywords Medical social network · Images indexation · Mixed approach · Relevant words extraction · MeSh thesaurus

1 Introduction

Currently, the unprecedented growth of social network media in various fields together with the amazing propagation on the global scale of social applications, have provided strong innovation in the information technology field and an interesting part of mainstream culture. Personal health information is recognized as one of the most sensitive types of personal information (Cavoukian 2004) presented on social networks; however, the use of these applications in the health care domain is still in its infancy (Williams and Weber-Jahnke 2010). In fact, physicians were very late to embrace new technology based social networks; however; there are already Web applications providing e-health services which seem to be lacking of interaction and/or analysis tools. Many personal health record systems allow individuals to follow their health data historic, such as immunizations, past procedures, allergies, etc., plans, which can be all shared with their knowledge. Social networking applications pledge to do more than just providing access to personal health information. One of its chief tasks is to promote collaboration between patients, caregivers and health care providers. Franklin and Greene (2007) consider that participation in the health care management can render patients longer health conscious. The main objective behind these medical networks is to foster collaboration between medical actors and to place the patient at the heart of the health system (Grenier 2003). As Marion and Omotayo (2011) quoted the idea behind the research is that projects already done by someone else do

J. Akaichi (✉)
Department of Computer Sciences, Institut Supérieur de Gestion, Tunis, Tunisia
e-mail: jalel.akaichi@isg.rnu.tn

not have to be developed again; instead it should be improved upon. The paramount goal of this work is to design and implement a social network dedicated to physicians and patients in order to cooperate through sharing knowledge, about patients' medical images, expressed through a collection of comments showing heterogeneous reports established by several physicians belonging to various specialties. In other words, this paper intends to provide a networking platform based on collaboration paradigms and knowledge diffusion. This will, obviously, help to save time and better serve the patients through medical errors minimization.

The need, to automatically extract information and analyze opinions, becomes obviously a requirement due to the huge number of comments expressing specialists' recommendations and/or remarks. For this purpose, we propose an additional contribution consisting of providing a kind of comments' summary which extracts the major current terms and relevant words existing on physicians' reports. Furthermore, this extracted information will present a new and robust input for image indexation enhanced methods. In fact, significant extracted terms will be used later to index images in order to facilitate their search through the underlying social network. To overcome the above challenges, we propose a Terminology Extraction of Annotation (TEA) approach which focuses on algorithms mainly based on statistical methods and external semantic resources destined to filter selected extracted information.

The remainder of the paper is organized as follows. In Sect. 2, we present the radiologist social network main requirements and data structure. In Sect. 3, we describe our TEA approach destined to filter selected extracted information from the intended social network. In Sect. 4, we detail our implementation and the experimentation results. Finally, in Sect. 5, we conclude and present the targeted future work.

2 Radiologist social network

Nowadays, the progress observed in medical imaging, either in the hardware or in the analysis techniques represent a very fruitful area in terms of research and innovation. Considering the time devoted by the physician in order to interpret and to annotate the underlying medical images; this effort may be considered to be complicated and time-consuming. Moreover, being the decision maker through making choices on his own, the physician may commit malpractices and generate unpredictable and paramount damages to patients' health states. The social internet has recently exploded due the huge number of different cases requesting the opinion of professionals, accompanied by an enormous number of comments and reviews describing various points of views. Yet, in spite of the minority of social networks in the medical field, an emerging trend is observed in different fields (Bruyere 2004) such as PatientsLikeMe (http://www.patientslikeme.com/), Radiolopolis (http://radiolopolis.com/), Docadoc (http://docadoc.com/), Carenity (http://www.carenity.com/), Sermo (http://www.sermo.com/), etc., which offer to their members an opportunity to be connected with others and share experiences and knowledge.

Medical images and associated comments are among the major fields inescapable for interaction between patients and patients and/or physicians such as radiologists. The collection of annotations and reports, from different medical experts' information sources, is very useful, when analyzed, to better serve patients and to overcome the problem of availability of physicians on site.

Our first goal is to design a social network dedicated to physicians and patients. The basic model of the targeted social media should take into account the management of a:

- Set of patients which provide personal information in their health care profile.
- Set of physicians providing information enabling their identifications.
- Set of mechanisms permitting to patients to upload the medical images related to their diseases.
- Set of mechanisms permitting to physicians to comment the uploaded medical images.
- Set of search functions by which patients and physicians can locate easy and efficient information about medical images.
- Site operator who controls the site and triggers a set of mechanisms permitting to collect comments about medical images in order to process them for various purposes such as concepts' extraction and medical images' indexation.

Our social network addressed to radiologists contains all of these features situated above.

3 Indexing medical annotations

The need to process automatically opinions and to extract information is therefore becoming a necessity due to the huge number of comments expressing opinions. Information extraction aims to automatically excerpt relevant information from texts related to a particular job or domain (Gaussier et al. 2007). In our case, we aim to extract powerful words from physicians' comments published on a dedicated social network in order to index medical images. Indexation is therefore the process of describing and characterizing documents, composed by comments, using the representation of their contents. Therefore, its purpose is to indicate in a concise form the content of the posts and to allow an efficient search of information in a collection of documents without having to analyze them in each time of need (Karbasi 2007). In short, an index is a relation that connects each collection of posts to all descriptors telling the theme it treats. Descriptors can be single words, lemmas, terms of a thesaurus, external semantic resource concepts, summaries of paragraphs or any other unit of information that describes the contents of the document.

3.1 Indexation approaches

In the literature, there exist many approaches of indexation such as linguistic, statistical, mixed, etc., methods. Linguistic approaches use syntactic partial analysis or pattern syntactic such as ontology, thesaurus, etc., to describe the content of the document. Several studies such in Maisonnasse et al. (2009), Gaussier et al. (2007) Zhou et al. (2007) have also used semantic resources in the process of indexation.

Many statistical approaches have been proposed such as probabilistic latent semantic indexing (Hofmann 1999), latent Dirichlet allocation (Blei et al. 2003), etc., in which several probabilistic or statistical combinations of the choice of weighting words were used. According to Van Rijsbergen (1979) and Salton and McGill (1986), the frequency of occurrence of words in natural language texts is indicative of the importance of these words for the sole purpose of representing the content of these texts.

The mixed approaches combine both syntactic and statistic information to improve the accuracy in the detection of index terms. Harrathi (2010) showed that the use of methods of extracting purely statistical terms, provide results of equivalent quality; therefore it is not necessary to use linguistic tools adapted to a given language. Subsequently, the author demonstrates that with an external semantic resource of sufficient quality, his prosed purely statistical approach gives greater results than those using linguistic techniques. Thus, the author found that with this statistical method, he will not have to change the linguistic parser whenever the document language evolves. In addition, the statistical approach is simple to implement as opposed to linguistic approaches.

3.2 The terminology extraction of annotation approach

In this section, we present our approach which we consider as a derivation of a mixed method involving a combination of statistical methods with an external semantic resource. The following algorithm 1 denotes the TEA approach.

Algorithm 1;

```
Input: Corpus of comments;
Output: Concepts;
Begin
  Preprocess the corpus of comments provided with medical image.
  Repeat
    Clean comments using an anti-dictionary;
    Perform lemmatization;
    Extract simple terms;
    Update the anti-dictionary;
    Extract compound terms;
  Until true;
  Extract concepts using MeSH thesaurus;
End.
```

As illustrated by the above algorithm, the TEA approach initiates by the creation of the textual corpus containing comments extracted from the medical social network and performed on a medical images. Then the following steps are accomplished:

- The preprocessing step is composed of four preliminary stages:
  - The decomposition of the corpus sentences.
  - The removal of the punctuation points.
  - The conversion of sentence to lowercase.
  - The cutting of the sentence into words.

In this step, we use pretreatment algorithm mainly based on four functions; one for each sub-step or stage.

- The cleaning step seeks to remove the stop words. To achieve this task, we use an anti-dictionary containing the blackest words that seem useless in the medical field. The anti-dictionary is a standard list containing words not to be used as index such as prepositions, articles, pronouns, some adverbs and adjectives, etc.
- Lemmatization step: for grammatical reasons, comments may include different forms of words belonging to various families. Lemmatization is used to regroup words in their appropriate ones. To find the lemmas, we implement a stemmer algorithm which seeks the root or prefix, and then assigns the suffix for a parent noun. Stemming algorithms are used in many applications related to natural language processing such as text analysis systems, information retrieval, database search systems, etc. (Frakes and Fox 2003). We use Porter algorithm (Manning et al. 2008) which is considered as the best known manner corresponding to our requirements. Porter's algorithm consists of five phases, of word reductions, applied sequentially. Within each phase there are various conventions to select appropriate rules. We implemented the entire algorithm using all grammatical rules associated to French language.
- Extraction of simple terms step: in this step, we measure the weight of each term using local and global

weighting tf.idf or term frequency–inverse document frequency. This method combines two factors, the local weighting (tf) which quantifies the local representation of a term in the corpus of comments and the overall weighting (idf) which measures the global representative of the term on the collection of corpus based on provided comments. This measure rises with the number of occurrences within a document and with the inadequacy of the term in the collection.

The formula expressing the measure is the following:

$$\text{tf.idf} = 0.4 + 0.6 \times \frac{\text{tf}_{ij}}{\text{tf}_{ij+0.5+1.5\frac{\delta i}{\delta}}} \times \frac{\log\left(\frac{N+0.5}{n_i}\right)}{\log(N+1)},$$

where N is the total number of documents in the corpus, n_i is the number of documents containing term i, tf_{ij} is the local weighting of term i in document j, δ_i is the length of document j in words, and δ is the average of the lengths of the documents of the corpus in words.

Note that, we keep only the terms that the output value exceeds a certain threshold set. In our case the threshold is fixed to 0.125.

- Extraction of compound terms step: this step seeks to distinguish single and compound words. Collocation measure attempts to check the terms which occur together more often; these terms are often reduced to two or three words present in a concept. This measure, called mutual information (MI), compares the occurrence of the probability of co-occurrence of words and the probability of these words separately. The process of extracting complex terms is iterative and incremental. We build complex terms of length $n + 1$ words from the words of length n. We start from the simple list of one word length; for each sequence of words, we compute the value of the MI. If the sequence of words exceeds the threshold set to 0.15 in our case, the sequence will be comprised on the list of compound terms. Note that the computation of MI is ensured by the following formula.

$$\text{MI}(m_1, m_2) = \log\frac{P(m_1, m_2)}{P(m_1) \times \text{P}(m_2)},$$

where $P(m_1)$ and $P(m_2)$ are an estimation of the probability of occurrence of the words m_1, m_2, and $P(m_1, m_2)$ is an estimation of the probability that the two words appear together.
- Extraction of concepts step: this is a verification step which comes to use an external semantic resource. More precisely a medical thesaurus MeSH is used to filter keywords obtained in the previous step, in order to verify that the extracted term belongs, or not, to the medical vocabulary. For that purpose, we propose the algorithm 2 to extract appropriate concepts.

Algorithm 2;
```
Input:
  D: The set of documents;
  TSimpleTerms: The list of simple terms;
  TCompoundTerms:  The list of compound terms;

Output: TConcepts: The extracted concepts;

Begin
  Foreach di in D do
    Begin
      While not the end of document di do
        Begin
          Read ti; //ti is a string of di
          If ti is a stop word then perform the cleaning process on ti;
        End;
      Lemmatize ti;
      Extract the simple term from ti and store it into TSimpleTerms;
      Extract the compound term from ({ti ∈ di},{ti+1 ∈ di}) and store it into
      TCompoundTerms;
      Extract concepts from (TSimpleTerms,TCompoundTerms) and store it into
      TConcepts;
    End;
End.
```

4 Implementation and experimental results

4.1 Social network description

A social network, described as a graph of associations and exchanges within a collection of individuals, plays an important role as a medium for the spread of information, ideas, and impact among its members.

The proposed social network, as many others described in the literature Gong and Sun (2011), Almansoori et al.

(2011), aims to allow patients and physicians to connect with each other by overwhelming geographical and time frontiers. Both users exploit it to seek advices and to share experiences related to medical images' interpretations.

The implemented social network site is a place where registered patients create profiles about themselves and upload their medical images to be commented by various registered physicians.

The targeted medical social network aims to enable patients and physicians to exchange information, share experiences and knowledge. It, also, offers to its members an opportunity to connect with others in need of support and encouragement related to encountered health care problems. Consequently, the patient is placed at the heart of the health system surrounded by experts' interpretations expressed by comments. A content management system Word press was used in the development of our social network. It ensures many features that allow creating and managing an entire social network website. The programs were written in HTML, PHP, Ajax, and JavaScript, pages were designed using Dreamweaver, and the social network underlying database was implemented thanks to MySQL.

We describe in the following some examples of the social network interfaces:

- In the main interface, the core functions dedicated to patients and physicians are displayed. Both can perform activities updates, get access to posting and commenting functions, etc.
- In the posting interface, mixed posts performed by different physicians and patients are entered and displayed. Posts will be treated by back office functions implementing our TEA approach.
- In the communication interface, synchronous communications between patients and physicians, physicians and physicians are performed. They are dedicated to urgent questions and/or advices.
- In the identification interface, the patient identification or the registration process is provided. An almost similar interface is dedicated to physicians; however, profiles have different structures.

Our social network is equipped with interfaces trying to provide users with a spontaneous mapping between their intent and targeted functions. Essentially, user interfaces describe the way patients and physicians interact with the social network site and the way they can access its functions allowing them to interact with each other's.

Provided interfaces are simple in terms of color scheme and graphics, and context sensitive, displaying many features only on demand. The color scheme, calm and supportive, usually consists of a few colors along with slight monochromatic disparities; the background is generally white, comments' updates are often highlighted with light colors associated with the kind of user (physician or patient) as well.

The social network is equipped with a functional search having multiple dimensions and advanced features. It permits connections' establishment between patients and physicians having common interests such as patients suffering from same diseases, physicians sharing same patients and/or treating same diseases. A mutual feature in social applications is the use of live exploration outcomes and filtering. When typing into the box, the outcomes are filtered out in a drop-down style. The filtering supports users to rapidly find the content they are searching for.

The proposed social network comprises many functions that need to be communicated in some manner to physicians and patients. Therefore, buttons and links are placed almost on every page of the social network. Some links are related to the navigation processes and some others permit to users to regulate specific functions. Buttons are used to associate users to actions; they are larger, brighter and more remarkable.

To make the content understandable and easy to recognize, content blocks are visually separated. Each element is well defined and presented separately. In fact, the split-up of elements in a layout is one of the easiest manners to succeed a cleaner user interface that patients and physicians can easily interact with.

While content could be a vast block of text accompanying uploaded medical images, user interface includes intelligent variations of text color, background, font sizes and link presentation to make the content more reachable. Apart from refining usability, considering annotations as user interface helps to increase readability and to associate back office tools destined to analyze interface content for various purposes such the application of our TEA approach.

Forms and inputs are used almost everywhere from sign-up, login, posting medical images, commenting them, posting a patient request, replying to a patient post, etc. A number of features will aid to keep our forms serviceable. It is ensured by keeping them as short as possible by including only essentials and leaving the extra information, for example, for the patients and physicians to edit once their accounts are generated.

Physicians and patients need to see incoming messages or posts in real time. For this reason, we equipped our social network with a real-time update feature ensuring the delivery of updates (medical images uploading, physicians' annotations, etc.) as soon as they are submitted.

Our social network, also, offer smart features and provide functions that would have benefits both online and offline. This is ensured by utilizing personalization and offering more flexible and more adaptive user interfaces. Personalization is a method that is used to deliver only relevant content to patients and physicians. Recommendation procedures are used to provide them with the information that is very likely to concern them. Continuous learning of interests, based upon users' accomplishments, will be used to enhance the model of the user and search for more personalized propositions.

Our aim is to provide physician and patients with a user-centric interface that is strongly focused on their personal interests gathered partially from details of their profiles. Many actions could be performed such as suggesting new "friends", interests, events, and groups. This is performed to extend their social circles allowing implicitly the extension of the shared knowledge about medical images and the associated diseases. Our social network tries to provide comprehensive information about updates and notifications and provides functions that make it easy to update the present status, hide selected information and it updates patients and physicians about actors they may know, groups or conversations that they may be involved in.

Note that, information and queries provided by physicians and patients will be used in the future to enhance the system though taking in account various users' contexts and preferences. Moreover, classification or clustering tasks will be studied, implemented and triggered to identify and create communities, such in (Chorbev et al., 2011) according to the identified characteristics of uploaded and commented medical images, and their relationships with the underlying considered diseases.

Our social network will remain open for enrichments inspired and enhanced from various works such as those presented in Yin et al. (2012), where authors describe a system using natural language processing and data mining techniques to extract circumstances' awareness information from Twitter messages generated during numerous disasters and crises.

We will also consider security aspects such those presented in Li (2014), where the author proposed methods to secure healthcare social networking sites providing users with tools and services to easily establish contact with each other around shared problems and utilize the wisdom of masses to outbreak disease.

Mining tools will be also on our focus in order to extract more knowledge from various posts such those presented in Xie et al. (2013), where authors consider the emergence and pervasivity of online social networks that have enhanced web data with developing interactions and communities both at large scale and in real-time.

4.2 Physicians' annotations posting and summarization

Subsequent to the implementation of our medical social network, we focused on the extraction of relevant concepts

Table 1 Cleaning and extraction of full words

No.	Annotation cleaned	Full words	
1	*présence d un gros* hématome fronto pariétale droit *de 8 cm sur 4 cm associé à une* inondation ventriculaire et une hémorragie méningée *visible* notamment *au niveau de l* hémisphère cérébral droit déviation gauche *de la ligne médiane témoignant d un début d un* engagement sous falcoriel	*hématome* *pariétale* *inondation* *hémorragie* *hémisphère* *déviation* *engagement*	*fronto* *droit* *ventriculaire* *méningée* *cérébral* *gauche* *falcoriel*
2	*nous remarquons la présence d une* hémorragie méningée hématurie *et d une* inondation ventriculaire	*hémorragie* *hématurie* *ventriculaire*	*méningée* *Inondation*
3	*l* image *est peu nette pour décider mais nous constatons la présence d une* hématome fronto pariétale piat *et d une* hémorragie méningée	*image* *fronto* *piat* *méningée*	*hématome* *pariétale* *hémorragie*
4	hémorragie méningée hématome *début d* engagement	*hémorragie* *hématome*	*méningée* *engagement*
5	*cette* radiographie *présente une* hématome fronto pariétale droit hémorragie méningée inondation ventriculaire *et un début d* engagement cérébral	*radiographie* *fronto* *droit* *méningée* *ventriculaire* *cérébral*	*hématome* *pariétale* *hémorragie* *inondation* *engagement*
6	*c est une* hématome fronto pariétale droit *une* hémorragie méningée	*hématome* *pariétale* *hémorragie*	*fronto* *droit* *méningée*
7	hémorragie méningée hématome fronto pariétal déviation *de la ligne médiane*	*hémorragie* *hématome* *pariétal*	*méningée* *fronto* *déviation*
8	*Il s agit d une* hématome fronto pariétal *encore une* inondation ventriculaire *et une* hémorragie méningée	*hématome* *pariétal* *ventriculaire* *méningée*	*fronto* *inondation* *hémorragie*

from posted comments in order to annotate and index medical images. For this purpose, we collected 30 examination cases, from Tunisian physicians involved in various specialties, related to six different medical images. Each examination consists on a comment related to a medical image. We used the TEA method to preprocess, to clean, to lemmatize comments and to extract relevant medical terms intended to be used for medical images annotation and indexation. We attempt then to prove that the theoretical results are confirmed in practice.

We assume that comments are written only in the French language. The following sample (Table 1) presents eight Tunisian radiologists annotations on a brain radiation

Table 2 tf.idf computation

	A1	A2	A3	A4	A5	A6	A7	A8	AVG
hématome	0.0667	0.0000	0.1429	0.2500	0.0909	0.1667	0.1667	0.1429	0.1283
fronto	0.0943	0.0000	0.2021	0.0000	0.1286	0.2358	0.2358	0.2021	0.1374
pariétal	0.0943	0.0000	0.2021	0.0000	0.1286	0.2358	0.2358	0.2021	0.1374
droit	0.4000	0.0000	0.0000	0.0000	0.2727	0.5000	0.0000	0.0000	0.1466
inondation	0.1333	0.4000	0.0000	0.0000	0.1818	0.0000	0.0000	0.2857	0.1251
ventriculaire	0.1333	0.4000	0.0000	0.0000	0.1818	0.0000	0.0000	0.2857	0.1251
hémorragie	0.0667	0.2000	0.1429	0.2500	0.0909	0.1667	0.1667	0.1429	0.1533
méningée	0.0667	0.2000	0.1429	0.2500	0.0909	0.1667	0.1667	0.1429	0.1533
hémisphère	0.2667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0333
cérébral	0.2667	0.0000	0.0000	0.0000	0.3636	0.0000	0.0000	0.0000	0.0788
déviation	0.2667	0.0000	0.0000	0.0000	0.0000	0.0000	0.6667	0.0000	0.1167
gauche	0.2667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0333
engagement	0.1610	0.0000	0.0000	0.6038	0.2195	0.0000	0.0000	0.0000	0.1230
falcoriel	0.2667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0333
image	0.0000	0.0000	0.5714	0.0000	0.0000	0.0000	0.0000	0.0000	0.0714
piat	0.0000	0.0000	0.5714	0.0000	0.0000	0.0000	0.0000	0.0000	0.0714
radiographie	0.0000	0.0000	0.0000	0.0000	0.3636	0.0000	0.0000	0.0000	0.0455

Table 3 MI computation

	hématome	*fronto*	*pariétal*	*droit*	*inondation*	*ventriculaire*	*hémorragie*	*méningée*
hématome	****	0.234	0	0	0	0	0	0
fronto	0	****	0.234	0	0	0	0	0
pariétal	0	0	****	0.142	0.013	0	0	0
droit	0	0	0	****	0.005	0	0.035	0
inondation	0	0	0	0	****	0.178	0	0
ventriculaire	0	0	0	0	0	****	0.035	0
hémorragie	0	0	0	0	0	0	****	0.272
méningée	0.07	0	0	0	0.035	0	0	****

Asterisks indicate undefined values

image. We extracted the full words from the corpus of comments while applying the preprocessing and cleaning steps.

In Table 2, we present the results related to the computation of the tf.idf. The threshold used for these tests was set to 0.125 and subsequently adjusted after several experimentations.

After determining the simple terms whose average exceeds 0.125, we compute the mutual information and present in Table 3. The threshold, used for MI formula, was set to 0.15 for our experiments. This value was also subsequently adjusted after several experimentations.

The extracted compound terms are “hématome fronto”, “fronto pariétale”, “hémorragie méningée”, “inondation ventriculaire”. The last step is the use of MeSh thesaurus for the extraction of concepts.

Finally, the extracted terms used to index the brain radiation image loaded by a patient and annotated by several specialists are “hématome fronto pariétale”, “hémorragie méningée”, and “inondation ventriculaire”.

5 Conclusion and future work

In order to minimize medical errors by fostering collaboration between physicians and/or patients, we proposed in this paper a medical social network destined to gather patients’ medical images and physicians’ annotations expressing their medical reviews and advices. We, also, proposed a kind of comments’ summarization which extracts the major current terms and relevant words existing on physicians’ annotations. The extracted information presents a new and robust input for image indexation enhanced methods. In fact, significant extracted terms are used to index images in order to facilitate their search and reuse through the underlying social network. To overcome the above challenges, we proposed an approach which focuses on algorithms mainly based on statistical methods and external semantic resources destined to filter selected extracted information.

Future work will focus on the establishment of relationships between uploaded and commented images. Two types of relationships will be considered: the relationship between versions, gathered in time, leading to the tracing of the trajectory of a disease enhanced by summarized annotations; and the relationship between images uploaded by different users, leading to an understanding of a same disease behavior. The study of the two relationships involves the design, the implementation, and the analysis of a specific data warehouse which will gather medical images and annotations published on the social network. This latter have to be enhanced to take into account various aspects such those related to indexation, versioning and analysis and/or mining tasks.

References

Almansoori W, Zarour O, Jarada TN, Karampales P, Rokne J, Alhajj R (2011) Applications of social network construction and analysis in the medical referral process. In: Proceedings of the 2011 IEEE ninth international conference on dependable, autonomic and secure computing (DASC '11)

Blei D, Ng A, Jordan M (2003) Latent Dirichlet allocation. J Mach Learn Res 3(2003):993–1022

Bruyere C (2004) Medico-social networks and the coordination of knowledge. DEA in Management Sciences, Logistics and Organization. Mediterranean University (Aix-Marseille II)

Cavoukian A (2004) Privacy and healthcare: you need both. https://www.backbonemag.com/files/PDF/Speakers/2011-09-privacy-and-healthcare.pdf

Chorbev I, Sotirovska M, Mihajlov D (2011) Virtual communities for diabetes chronic disease healthcare. Int J Telemed Appl 2011:721654

Frakes WB, Fox CJ (2003) Strength and similarity of affix removal stemming algorithms. In: Newsletter of ACM SIGIR Forum Homepage archive, vol 37 Issue 1. New York, pp 26–30

Franklin V, Greene S (2007) Sweet talk: a text messaging support system. J Diabetes Nurs 11(1):22–26

Gaussier E, Maisonnasse L, Chevallet JP (2007) Multiplying concept sources for graph modeling. In: Proceedings of the 8th workshop of the Cross-Language Evaluation Forum (CLEF), Budapest, Hungary, pp 585–592

Gong J, Sun S (2011) Individual doctor recommendation model on medical social network. In: Proceedings of the 7th international conference on Advanced Data Mining and Applications (ADMA'11)

Grenier C (2003) The role of intermediate subject to understand the structuring of an organizational network of actors and technology—case of a care network. In: Proceedings of the 9th conference of the association information and management, Grenoble

Harrathi F (2010) Extraction de concepts et de relations entre concepts à partir des documents multilingues: approche statistique et ontologique. PhD Thesis, INSA Lyon

Hofmann T (1999) Probabilistic latent semantic indexing. In Proceeding of SIGIR '99 of the 22nd annual international ACM SIGIR conference on research and development in information retrieval, New York, pp 50–57

Karbasi S (2007) Terms weighting in information retrieval; weighting model based on the ranking of terms in documents. PhD thesis, University Paul Sabatier

Li J (2014) Data protection in healthcare social networks. J IEEE Softw 31(1):46–53

Maisonnasse L, Gaussier E, Chevallet JP (2009) The combination of semantic analysis for searching medical information. In: Proceedings of the INFORSID Conference, Toulouse

Manning C, Raghavan P, Hinrich SH (2008) Introduction to information retrieval. Cambridge University Press Book, UK

Marion A, Omotayo O (2011) Development of a social networking site with a networked library and conference chat. J Emerg Trends Comput Inform Sci 2(8):396–401

Salton G, McGill MJ (1986) Introduction to modern information retrieval. McGraw-Hill, Inc., New York

Van Rijsbergen CJ (1979) Information retrieval. Butterworths, Information Retrieval Group, University of Glasgow, London

Williams J, Weber-Jahnke J (2010) Social networks for health care: addressing regulatory gaps with privacy-by-design. 8th Annual International Conference on Privacy Security and Trust (PST), Ottawa, pp 134–143

Xie Y, Chen Z, Cheng Y, Zhang K, Agrawal A, Liao WK, Choudhary A (2013) Detecting and tracking disease outbreaks by mining social media data. In: Proceedings of the twenty-third international joint conference on artificial intelligence (IJCAI'13)

Yin J, Lampert A, Cameron M, Robinson B, Power R (2012) Using social media to enhance emergency situation awareness. J IEEE Intell Syst 27(6):52–59

Zhou W, Yu C, Smalheiser N, Torvik V, Hong J (2007) Knowledge intensive conceptual retrieval and passage extraction of biomedical literature. In: Proceeding of SIGIR '07 of the 30th annual international ACM SIGIR conference on research and development in information retrieval, New York, pp 655–662