Exploiting Temporal Information in Documents and Query to Improve the Information Retrieval Process: Application to Medical Articles

Jihen MAJDOUBI
Department of Computer Science
College of Science and Humanities at AlGhat
Majmaah University, P.O. Box 66, Majmaah 11952
Kingdom of Saudi Arabia

Ahlam Nabli
Department of Computer Science
College of bandaq
Albaha university
Kingdom of Saudi Arabia

***Abstract*—In the medical field, scientific articles represent a very important source of knowledge for researchers of this domain. But due to the large volume of scientific articles published on the web, an efficient detection and use of this knowledge is quite a difficult task. In this paper, we propose a novel method for semantic indexing of medical articles by using the semantic resource MeSH (Medical Subject Headings) and the temporal information provided in the documents. The proposed indexing approach was evaluated by intensive experiments. These experiments were conducted on document test collections of real world clinical extracted from scientific collections, namely, CISMEF and CLEF. The results generated by these experiments demonstrate the effectiveness of our indexing approach.**

***Keywords*—Biomedical information retrieval; semantic indexing; temporal criteria; Medical Subject Headings (MeSH) thesaurus**

I. Introduction

The WWW becomes a very vast repository of data and the volume of information generated in this digital world is increasing day by day. This, however, would be wasted if necessary information could not be found, analyzed, and exploited. The goal of any Information Retrieval System (IRS) is to retrieve relevant information to a users query.

This goal is quite a difficult task with the rapid and increasing development of the Internet. Indeed, web information retrieval becomes more and more complex for the user who IRS provides a lot of information. However the user often fails, to find the best information in the context of his/her information need.

The problem in searching over documents is that documents are time-dependent and accumulated over time which results in a large number of irrelevant documents in a set of retrieved documents. Therefore, the users have to spend more time in finding the documents that are satisfying his/her information need. Traditional Information Retrieval approaches based on topic similarity alone is not sufficient for the search in growth document collections. Much research is going on the field of temporal information retrieval to improve the retrieval results. The time criterion has already been the core concept of recent IR ranking models, given that most of documents include a high level of temporal information [23]. Indeed, several works show that a large amount of web documents become time-dependent [2], [21]. The authors in [13] have argued that about 7% of queries have implicit temporal intent, while other studies show that only 1.5% of queries are explicitly provided with temporal intent.

In this paper, we are interested in the temporal information and its impact in the process of medical article indexing. Our motivation is that timeliness is one of the key aspects that determine a documents credibility besides relevance, accuracy, objectivity and coverage.

The treatment of medical information has made the interest of several research works and a lot of solutions have been proposed so far, based on context query, online ranking model, semantic model. However, to the best of our knowledge, there is no prior attempts dealing with the use of the temporal criteria in the biomedical IR field. In this paper, we propose a novel method for conceptual indexing of medical articles by using the semantic resource MeSH (Medical Subject Headings) and the temporal information provided in the documents. Specifically, both temporal relevance and topic similarity are needed for efficient retrieval. The remainder of this paper is organized as follows. In the next section, we attempt to prove the effectiveness of exploiting temporal criteria in the information retrieval process. Section 3 summarizes our context and motivations. In Section 4, we review the related work. Section 5 details our conceptual indexing approach. An experimental evaluation and comparison results are discussed in Sections 6 and 7. Finally, Section 8 presents some conclusions and future work directions.

II. Using Temporal Information to Improve the Retrieval Results

The notion of using time as an important factor becomes more important for a large number of searches. In the following, we attempt to prove the effectiveness of exploiting temporal criteria in the information retrieval process.

Consider an example of historian interested in knowing about the Tunisia revolution that occurs in past years. He searches in the news archives expecting to retrieve the details of the event- not necessarily the latest news, but a report on

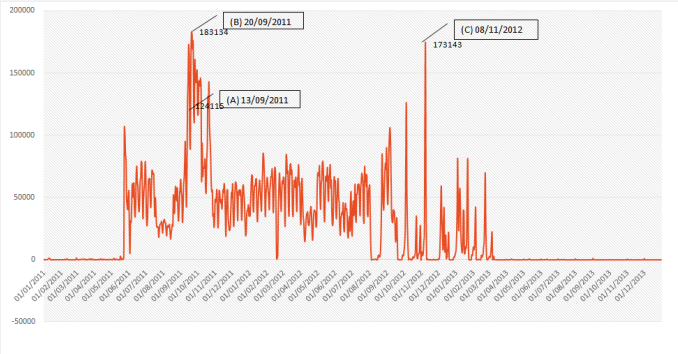

Fig. 1. Number of documents containing the term "barrack obama".

- **September 8 –** President Obama presents the American Jobs Act, his plan to create jobs and revive the American economy.
- **September 12 –** The President delivers a speech in the White House Rose Garden to promote his American Jobs Act.
- **September 16 –** President Obama signs the America Invents Act, (H.R. 1249), a major overhaul of the U.S. patent system, into law.
- **September 19 –** The President releases his debt reduction plan and the Buffett Rule.

Fig. 2. Timeline of the presidency of Barack Obama in 2011.

the latest news about that query is retrieved. The most of the relevant documents for that query is for the time period of $2010 - 2011$ or associated with the time that event happened. This example shows clearly that timeliness is one of the key aspects that determine a documents credibility besides relevance, accuracy, objectivity and coverage. Both temporal relevance and topic similarity are needed for efficient retrieval.

In order to better prove this, we try to analyze the frequency of the term "barrack obama" in the "LivingKnowledge sub collection" over the time. This collection spans from May 2011 to March 2013 and contains around 3.8M documents collected from about 1500 different blogs and news sources. The data is split into 970 files, named after the date of that day and some information about its sources (there might be more than one file per day). We plot in Fig. 1 curve representing the number of documents containing the term "barrack obama" in the "Living Knowledge news and blogs" corpus.

The *x-axis* represents time in months (From $01/01/2011$ to $01/12/2013$), and the *y-axis* indicates the number of documents containing the term "barrack obama" over the corpus.

Fig. 1 clearly shows that the number of documents containing the term "barrack obama" increases significantly during specific time periods.

For example as highlighted at Fig. 1 in $13/09/2011$, we have 124116 blogs and news sources discuss about "obama". This number has grown to reach a value of 183134 in $20/09/2011$. By referring to the timeline of the presidency of Barack Obama in 2011 (see Fig. 2), we can remark that peaks (A) and (B) presented in Fig. 1 are well-lined up with the timeline of most actions made by Barack Obama to revive the American economy.

Also, the third peak (C) off the graphic appearing at $08/11/2011$ with 173143 news corresponds to date of re-election Barack of Obama $(November 6, 2012)$. This is mainly due to the fact that people tend to talk about the Obama's news mainly during or slightly after time periods when the action was held and number of documents created beyond these time periods increases significantly.

On the basis of examples presented in this section we can confirm that time dimension must be exploited as a highly important relevance criterion to improve the retrieval effectiveness of document ranking models.

III. Context and Motivations

Each year, the rate of publication of biomedical literature grows, making it increasingly harder for researchers to keep up with novel relevant published work. In recent years several researches have been devoted to attempt to manage effectively this huge volume of information.

In [14], the authors proposes a tool called MAIF (MesH Automatic Indexer for French) which is developed within the CISMeF team. To index a medical resource, MAIF follows three steps: analysis of the resource to be indexed, translation of the emerging concepts into the appropriate controlled vocabulary (MeSH thesaurus) and revision of the resulting index.

In [15], the authors proposed the MTI (MeSH Terminology Indexer) to index English resources. MTI results from the combination of two MeSH Indexing methods: MetaMap Indexing (MMI) [16] and a statistical, knowledge-based approach called PubMed Related Citations (PRC) [9].

The conceptual indexing strategy proposed by [8] involves three steps. First they compute for each concept MeSH C its similarity with the document D. After that, the candidate concepts extracted from step 1 are re-ranked according to a correlation measure that estimates how much the word order of a MeSH entry is correlated to the order of words in the document. Finally the content based similarity and the correlation between C and the document D are combined in order to compute the overall relevance score. The N top ranked concepts having the highest scores are selected as candidate concepts of the document D.

The indexing approach presented in this paper differs from previous works. In this paper, we are interested in integrating temporal information in the process of medical article indexing. Our motivation is that Temporal information is crucial in biomedical information systems. Healthcare providers normally record the progress of a disease or a hospital course chronologically in text, and procedures and laboratory tests are stored in databases with time-stamps. Therefore, automatically reasoning about temporal information can help us understand the dynamics of medical phenomena and may potentially improve the quality of patient care.

IV. Related Research Works

Temporal Information Retrieval has started to be considered as a subdivision of the field of information retrieval. In this section, we provide a comprehensive and a comparative overview of most important work on both time and IR.

Li and Croft [11] defined two types of time-based queries in TREC collections that contain news archives. The first favors the most recent documents and the second is shown to have relevant documents within a specific period in the past. To incorporate time information into retrieval models, they proposed a time-based language model using a prior based on an exponential or a normal distribution depending on the types of recency queries.

In [19], the authors proposed an extension to the Query Likelihood Model that incorporates query-specific information to estimate rate parameters. They also introduced a temporal factor into language model smoothing and query expansion using pseudo-relevance feedback. These extensions were evaluated using a Twitter corpus and two newspaper article collections. Results showed that, compared to prior approaches, models proposed are more effective at capturing the temporal variability of relevance associated with some topics.

In [12], the authors proposed a query expansion model for microblogs, which selects terms temporally closer to the query submission time. Their model is supposed to work well for finding documents related to events currently happening but, not as well for past events.

In [10], the authors suggested a general language model that incorporates time into the ranking model in a principled manner. For a given time-sensitive query over a news archive, the approach automatically identifies significant time intervals for the query and uses them to adjust the document relevance scores by boosting the scores of documents published within the important intervals. They presented an extensive experimental evaluation, including TREC as well as an archive of news articles, and showed that proposed techniques improve the quality of search results, compared to the existing state-of-the-art algorithms.

In [26], the authors presented an adaptive temporal query modeling for blog feed retrieval, in that they analyzed the top retrieved documents in terms of temporal histogram to find the bursts. They used documents with the highest scores from the bursts for query expansion and weighted each feedback document with the distance from the peak that contains most documents.

In [13], [24], the authors proposed a temporal query expansion method for microblogs based on the temporal co-occurrence of terms in a timespan. They first performed pseudo-relevant timespan retrieval for an event query and used those timespans for query expansion. Although their goal was retrieving a ranked list of historical event summaries, the temporal query expansion method showed that selecting relevant timespan is crucial for query expansion for microblog documents.

The state-of-the-art presented in this section shows that temporal information retrieval has shown its performance in many scopes. In this paper, we try to exploit temporal information in medical documents to improve the information retrieval process. Our choice is motivated by the fact that temporal information is crucial in biomedical information systems and procedures and laboratory tests are stored in databases with time-stamps.

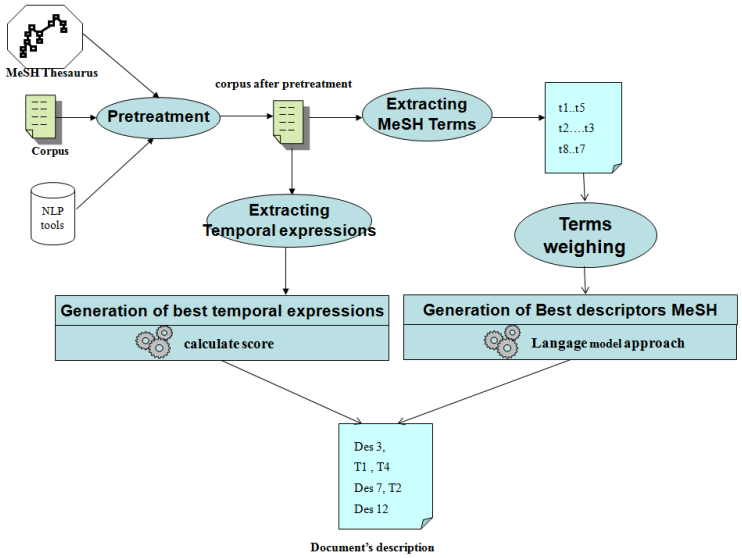

Fig. 3. Architecture of proposed indexing approach.

V. Proposed Approach

In [5], we have proposed an approach for conceptual indexing of medical articles by using the MeSH (Medical Subject Headings) thesaurus. More precisely, we have tried to determine for each document, the most representative MeSH descriptors. For this reason, we deduced a language model for each document and rank Mesh descriptor according to our probability of producing each one given that model. The proposed indexing approach was evaluated by intensive experiments in [6], [7]. These experiments were conducted on document test collections of real world clinical extracted from scientific collections, namely PUBMED and CLEF. The results generated by these experiments demonstrated the effectiveness of proposed indexing approach.

To improve these results, we integrate the time criteria in the indexing process. we made an assumption that the Time plays important roles in medical articles because healthcare providers normally record the progress of a disease or a hospital course chronologically in text.

Our indexing methodology as schematized in Fig. 3, consists of five steps: 1) Pretreatment; 2) Extracting MeSH concepts; 3) Extracting temporal expressions; 4) Generation of Best descriptors MeSH; and 5) Generation of best temporal expressions. We describe below the structure of MeSH vocabulary and then we detail the steps of proposed indexing method.

A. MeSH Thesaurus

The language of biomedical texts, like all natural language, is complex and poses problems of synonymy and polysemy.

Therefore, many terminological systems have been proposed and developed such as Galen [1], UMLS [2] and MeSH [3].

In our context, we have chosen MeSH because it meets the aims of medical librarians and it is a widely used tool for indexing literature.

The structure of MeSH is centered on descriptors, concepts, and terms.

- Each term can be either a simple or a composed term.
- A concept is viewed as a class of synonymous terms, one of then (called Preferred term) gives its name to the concept.
- A descriptor class consists of one or more concepts closely related to each other in meaning. Each descriptor has a Preferred Concept. The descriptor's name is the name of the preferred Concept. Each of the subordinate concepts is related to the preferred concept by a relationship (broader, narrower).

Cardiomegaly [Descriptor]
Cardiomegaly [Concept, Preferred]
Cardiomegaly [Term, Preferred]
Enlarged Heart [Term]
Heart Enlargement [Term]
Cardiac Hypertrophy [Concept, Narrower]
Cardiac Hypertrophy [Term, Preferred]
Heart Hypertrophy [Term]

Fig. 4. An example of MeSH.

As shown by Fig. 4, the descriptor "Cardiomegaly" consists of two concepts: "Cardiomegaly" and "Cardiac Hypertrophy". Each concept has a preferred term, which is also said to be the name of the Concept. For example, the concept "Cardiomegaly" has three terms "Cardiomegaly" (preferred term), "Enlarged Heart" and "Heart Enlargement". As in the example above, the concept "Cardiac Hypertrophy" is narrower than the preferred concept "Cardiomegaly".

B. Pretreatment

The first step is to split text into set of sentences. We use the Tokeniser module of GATE in order to split the document into tokens, such as numbers, punctuation and words. Then, the TreeTagger stems these tokens to assign a grammatical category (noun, verb...) and lemma to each token. Finally, system prunes the stop words for each medical article of the corpus.

This process of pretreatment is also carried out on the MeSH thesaurus.

Fig. 5 outlines the basic steps of the pretreatment phase.

[1]http://www.opengalen.org
[2]http://www.nlm.nih.gov/researchlumls/
[3]http://www.nlm.nih.gov/mesh/

La prise en charge de la pathologie dépressive repose sur les médicaments antidépresseurs. Ces médicaments doivent entre toutefois prescrits avec prudence chez les sujets âgés.
(a) Document

la prise en charge de la pathologie dépressive repose sur les médicaments antidépresseurs.
ces médicaments doivent entre toutefois prescrits avec prudence chez les sujets âgés.
(b) Spliting into senteces

la prise en charge de la pathologie dépressif repose sur le médicament antidépresseur.
Ce médicament doit entre toutefois prescrire avec prudence chez le sujet âgé.
(c) Lemmatisation

prise pathologie dépressif médicament antidépresseur.
médicament prescrire sujet âgé.
(d) Pruning stop words

Fig. 5. Pretreatment step.

C. Extracting MeSH Terms

As mentioned above, a term can be either simple or composed. To extract the simple term, we project the Mesh thesaurus on the document by applying a simple matching. More precisely, each lemmatized term in the document is matched with the canonical form or lemma of MeSH terms. To recognize the composed terms, we have chosen to use YateA [27]. YateA (Yet Another Term ExtrAtor) [29] is an hybrid term extractor developed in the project ALVIS. After text processing, YateA generates a file composed of two columns: the inflected form of the term and its frequency. For instance, as shown in Fig. 6 which describes the result of the term extraction process by using YateA, the term "exercice physique" occurs 6 times.

```
# Inflected form      Frequency
activité physique     16
activité sportive     9
exercice musculaire   8
exercice physique     6
effets bénéfiques     6
g de glucides    5
contrôle glycémique   5
insuffisance coronaire      5
index glycémique 4
risque cardiovasculaire     4
adaptation des doses  4
glycémie capillaire   4
sensibilité à l' insuline   4
fréquence cardiaque   3
hydrates de carbone   3
acides gras libres    3
autosurveillance glycémique 3
patient dnid     3
acides gras      3
dernier repas    3
profil lipidique 3
activité physique régulière 3
insuline rapide  3
```

Fig. 6. An excerpt of the result of YaTeA.

D. Term Weighing

Given a set of extracted terms issued from the step of "Extracting MeSH Terms", we calculate the terms weight by using two measures: the Content Structure Weight (CSW) and the Semantic Weight (SW) [4].

1) Content Structure Weight: We can notice that the frequency is not a main criterion to calculate the CSW of the term. Indeed, the CSW takes into account the term frequency in each part of the document rather than the whole document. For example, a term of the Title receives a higher importance ($*10$) than to a term that appears in the Paragraphs ($*2$). Table 1 shows the various coefficients used to weight the term locations. These coefficients were determined in an experimental way in [3].

TABLE I. WEIGHING COEFFICIENTS

term location	Weight of the location
Title (T)	10
Keywords (K)	9
Abstract (A)	8
Paragraphs (P)	2

The CSW of the term t_i in a document d is given as follows:

$$CSW(t_i, d) = \frac{\sum_{A \in T,K,A,P} f(t_i, d, A) \times W_A}{\sum_{A \in T,K,A,P} f(t_i, d, A)} \quad (1)$$

Where,

- W_A is the weight of the location A (see Table I),
- $f(t_i, d, A)$ is the occurrence frequency of the term t_i in the document d at location A.

For example, the term *cancer* exists in the document d_{1683}: 1 time in the title, 2 times in the abstract and 9 times in the Paragraphs,

$$CSW(cancer, d_{1683}) = \frac{1*10 + 2*8 + 9*2}{1+2+9}$$

.

2) Semantic Weight (SW): The Semantic Weight of term t_i in the document d depends on its synonyms existing in the set of Candidate Terms ($CT(d)$) generated by the term extraction step. To do so, we use the $Synof$ function that associates for a given term t_i, its synonyms among the CT(d).
Formally the measure SW is defined as follows:

$$SW(t_i, d) = \frac{\sum_{g \in Synof(t_i, CT(d))} f(g, d)}{|Synof(t_i, CT(d))|} \quad (2)$$

For a given term t_i, we have on the one hand its Content Structure Weight ($CSW(t_i, d)$) and on the other its Semantic Weight ($SW(t_i, d)$), its Local Weight (($LW(t_i, d)$)) is determined as follows:

$$LW(t_i, d) = \frac{CSW(t_i, d) + SW(t_i, d)}{2} \quad (3)$$

By examining the equation 3, we can notice that the terms (simple or composed) are weighted by the same way. Despite the several works dealing with the weighing of composed terms, there is so far no weighing technique shared by the community [17]. In our approach, we applied the weighing method proposed by [17]. For a term t composed of n words, its frequency in a document depends on the frequency of the term itself, and the frequency of each sub-term. For this purpose, it proposes the measure cf is defined as follows:

$$cf(t, d) = f(t, d) + \sum_{st \in subterms(t)} \frac{length(st)}{length(t)}.f(st, d) \quad (4)$$

where,

- $f(t, d)$: the occurrences number of t in the document d.
- $Length(t)$ represents the number of words in the term t.
- $subterms(t)$ is the set of all possible terms MeSH which can be derived from t.

For example, if we consider a term "cancer of blood", knowing that "cancer" is itself also a MeSH term, its frequency is computed as:

$$cf(cancer\ of\ blood) = f(cancer\ of\ blood) + \frac{1}{2}.f(cancer)$$

Consequently, in an attempt to take into account the case of composed terms, we calculate the csw measure as follows:

$$CSW(t_i, d) = \frac{\sum_{A \in T,K,A,P} f(t_i, d, A) \times W_A}{\sum_{A \in T,K,A,P} f(t_i, d, A)} + \sum_{st \in subterms(t_i)} \frac{length(st)}{length(t_i)}.f(st, d) \quad (5)$$

where, f(st,d) is the occurrences number of st in the document d.

It's important to note that in the case of simple terms, $subterms(t_i) = \emptyset$. Consequently the formulas presented by (5) and (1) are equivalent.

Finally, the weight of a term t_i in a document d_j ($Weight(t_i, d_j)$) is calculated as follows:

$$Weight(t_i, d_j) = LW(t_i, d_j).ln(N/df) \quad (6)$$

where,

- N: the total number of documents,
- df (document frequency): the number of documents which term t_i occurs in.

E. Generation of Best Descriptors MeSH

A term MeSH may be located in different hierarchies at various levels of specificity, which reflects its ambiguity. As an illustration, Fig. 7 depicts the term "Pain", which belongs to four branches of three different hierarchies (descriptors) whose the most generic descriptors are: Nervous System Disease (C10); Pathological Conditions, Signs and Symptoms (C23); Psychological Phenomena and Processes (F02); Musculoskeletal and Neural Physiological Phenomena (G11).

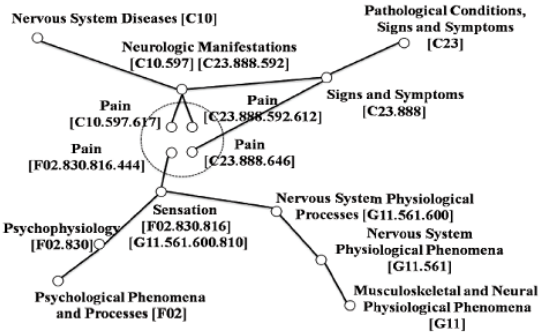

Fig. 7. Term Pain in MeSH.

In the last years, due to the amount of ambiguous terms and their various senses used in biomedical texts, term ambiguity resolution becomes a challenge for several researchers [1] [18]. For an ambiguous term, the task of WSD consists in answering the question: among its several senses, which is the best descriptor that can represent this term. The task of the WSD system is then to estimate, for each candidate descriptor MeSH, which is most likely to be the ideal concept. Differently from the proposed works in the literature, our method assign the appropriate descriptor related to a given term by using the language model approach.

In proposed approach, to determine for an ambiguous term, its best descriptor, we have adapted the language model of [20] by substituting the query by the Mesh descriptor. Thus, we infer a language model for each document and rank Mesh descriptors according to their probability of producing each one given this model. We would like to estimate $P(des|d)$, the probability of generation a Mesh descriptor des given the language model of document d.

For a collection (D), a document (d) and a MeSH descriptor (des) composed of n concepts, the probability $P(des|d)$ is done by:

$$P(des_k|d) = \prod_{c_j \in relatedtoDes(des_k,d)} ((1-\lambda).P(c_j|d) + \lambda.P(c_j|D)) \quad (7)$$

$RelatedtoDes$ (respectively RelatedtoCon) is the function that associates for a given descriptor des (respectively concept con) and a document d, the concepts (respectively terms) MeSH which are related to des (respectively con) in d.

In this equation, we need to estimate two probabilities:

- $P(c_j|D)$ the probability of observing the concept c_j in the collection D.

$$P(c_j|D) = \frac{\sum_{t_i \in relatedtoCon(c_j,d)} df(t_i, D)}{\sum_{c' \in D} f(c', D)}$$

$df(t, D)$: df (document frequency) is the number of documents which term t occurs in D.

- $P(c_i|d)$ the probability of observing the concept c_i in document d.

$$P(c|d) = \frac{f(c,d)}{|concepts(d)|}$$

$$f(c_j, d) = \sum_{t_l \in relatedtoCon(c_j,d)} LW(t_l, d)$$

$LW(t, d)$ is determined by using (3).

Based on this approach, to assign the appropriate sense (Best Descriptor (BD)) related to an ambiguous term (t_i) in the context of document (d_j), we must go through these steps:

1) *Compute the descriptor relevance score*
Let

$$senses^i_{d_j} = \left\{ des^{i1}_{d_j}, des^{i2}_{d_j} \dots des^{in}_{d_j} \right\} :$$

the set of descriptors MeSH that can represent the term t_i in the document d_j.
For each descriptor des_k existing in this set, we need to measure its ability to represent the term (t_i) in the document (d_j). To do so, we calculate $P(des_k|d_j)$.

2) *Selection of the best descriptor* The best descriptor (BD) to retain is the one which maximizes $P(des_k|d_j)$:

$$BD(t_i, d_j) = \max_{des_k \in senses^i_{d_j}} P(des_k|d_j)$$

Finally, in document's description of document d, we retain its Semantic Index (SI).

$$SI(d_j) = \bigcup_{t_i \in CT(d_j)} BD(t_i, d_j)$$

F. Extracting Temporal Expressions

This step extracts all temporal expressions in document, including the explicit temporal expressions and the implicit temporal expressions.

- Explicit temporal expressions:
These temporal expressions directly describe entries in some timeline, such as an exact date or year. For example, the token sequences "December 2017" or "September 12, 2011" in a document are explicit temporal expressions and can be mapped directly to chronons in a timeline.

- Implicit temporal expressions:
These temporal expressions represent temporal entities that can only be anchored in a timeline in reference to another explicit or implicit, already anchored temporal expression. For example, the expression "last Friday" or "next week" alone cannot be anchored in any timeline.

To extract explicit temporal expressions, we employ the GUTime tool [22]. The implicit temporal expressions are extracted by using the method presented in [27].

TABLE II. PRECISION (P) AND MAP(MEAN AVERAGE PRECISION) GENERATED BY THE INDEXING SYSTEMS USING TITLE AS INPUT

System	(P@10)	MAP
MTI	0.18	0.16
MetaMap	0.17	0.14
EAGL	0.18	0.17
KNN	0.43	0.47
TemPIndex	0.40	0.45

TABLE III. PRECISION AND MAP (MEAN AVERAGE PRECISION) GENERATED BY THE INDEXING SYSTEMS USING TITLE AND ABSTRACT AS INPUT

System	(P@10)	MAP
MTI	0.32	0.25
MetaMap	0.19	0.16
EAGL	0.21	0.19
KNN	0.45	0.50
TemPIndex	0.33	0.28

G. Generation of Best Temporal Expressions

The score of a temporal expression as a combination of an explicit score and an implicit score. The explicit score is related to the term frequency of a temporal expression, and accordingly the implicit score is related to the contribution made by all its children expressions [27]. The score of T_i, denoted as Score(T_i), is the sum of its explicit score, denoted as ES(T_i), and its implicit score, denoted as IS(T_i).

$$ES(T_i) = TF_{ETE}(T_i) + d * TF_{ITE}(T_i)$$

$TF_{ETE}(T_i)$ refers to the term frequency of the explicit temporal expressions which are recognized as T_i.

$TF_{ITE}(T_i)$ refers to the term frequency of the implicit temporal expressions which are calculated as T_i. d is the weighting factor, if d is set to 1, it means that the explicit and implicit temporal expression have the same credible level; if d is set to 0, it means that we take no account of implicit temporal expressions. Finally, the N[4] top-ranked temporal expressions with the highest score are selected in document's description.

VI. COMPARISON OF PROPOSED SYSTEM WITH OTHERS INDEXING SYSTEMS

To prove the effectiveness of our indexing method, we compared system (TemPIndex) to other medical indexing systems. We evaluate the performance of five indexing systems (MetaMap, EAGL, KNN, MTI and TemIndex) in terms of generating the manual MeSH annotations. For this evaluation, we used the same corpus[5] used by [28] composed of 1000 random MEDLINE citations.

Table II shows the results generated by indexing systems using the title of a 1000 random MEDLINE citations.

Table III shows the results generated by indexing systems with the title and abstract of a 1000 random MEDLINE citations.

Fig. 8 illustrates the obtained results by the five indexing systems on the 1000 random MEDLINE citations.

[4]The N value is calculated experimentally

[5]The corpus can be downloaded in ($http://www.ebi.ac.uk/triesch/meshup/testset_v1.xml$)

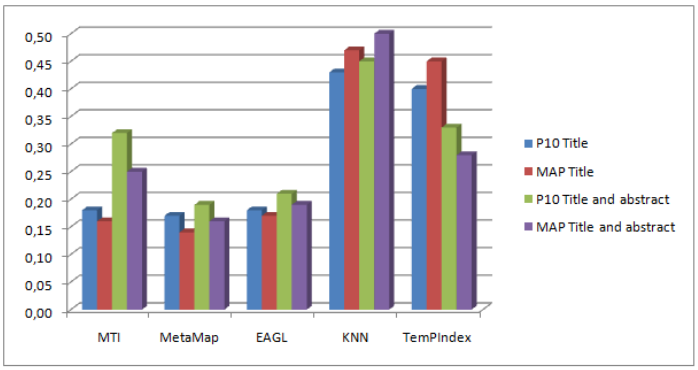

Fig. 8. Experimental results generated by the five indexing systems.

The system TemPIndex serves as the baseline against which the other systems are compared. Both indexing systems MetaMap and EAGL perform worse than BIOINSY in all metrics. Indeed, MetaMap performs similarly to or slightly worse than EAGL when presented with the title only or both title and abstract of the citation to index. MTI performs worse than TemPIndex when the title was available for indexing. For example, when title used as input, the value of P10 generated by MTI is equal to $0,18$. Concerning TemPIndex, it generates $0,40$ as value of P10. When using title and abstract, MTI performs similarly to or slightly better than TemPIndex in terms of MAP and P10. By using title as input, KNN and TemPIndex echoed very similar performance. Given the title and abstract of a citation, KNN shows the best results in all metrics. The obtained results confirm the well interest to use the temporal criteria in the indexing process.

VII. RESULTS AND DISCUSSION

In this section, we try to answer the following question: Can proposed temporal indexing approach (described and evaluated above) improve the information retrieval process. The overview of this section is as follows. In subsection 7.1 we will present the test collection. In subsection 7.2 we will describe the experimental setup. In subsection 7.3, the experimental results will be analyzed and discussed.

A. Test Collection

To evaluate the retrieval effectiveness based on our conceptual indexing method, we use the ImageCLEF med 2007 collection[6]. Started from 2004, the ImageCLEFmed (medical retrieval task) aims at evaluating the performance of medical information systems, which retrieve medical information from a mono or multilingual image collection. This corpus [25] is based on a dataset containing images from the Casimage, MIR, PEIR, PathoPIC, CORI, myPACS and Endoscopic collections. For each image of this corpus, a textual description called diagnosis is attributed. This corpus comprises 47680 cases, 66662 images and 55485 Annotations.

B. Experimental Setup

The ImageCLEF data contains the qrels file (TREC format) which specifies the set of relevant images to a given query. In

[6]CLEF (Cross Language Evaluation Forum)

TABLE IV. THE COMPARISON OF OUR SYSTEM WITH OFFICIAL RUNS PARTICIPATED IN IMAGECLEF MED 2007

Run	(P@5)	MAP
LIG-MRIM-LIGMU	0.44	0.32
OHSU	0.42	0.27
IPAL4	0.39	0.27
miracleTxtFRT	0.43	0.17
IRIT RunMed1	0.05	0.04
system TemPIndex in experiment 1	0.38	0.24
system TemPIndex in experiment 2	0.43	0.33

our indexing method we are interested by the textual document. Hence, to evaluate the proposed approach we assume that "If a query is relevant to an image then it is also relevant to its textual description (diagnosis)".

This evaluation process is structured around the following steps:

- *Indexing of diagnosis and queries* The indexing process is carried out on the diagnosis and queries. Thus, documents and eventually queries are expanded with MeSH descriptors and temporal expressions identified by our indexing method.
- *Calculation of Retrieval Status Value (RSV (q, d))* The relevance score of the document d_j with respect to the query q is given by

$$RSV(q, d_j) = \sum_{des,temp \in q} TF_j(des, temp) * IDF(des, temp)$$

Where,
- TF_j: the normalized term frequency of the current descriptor MeSH (des) or expression temporal (temp) in document d_j.
- IDF: the normalized inverse document frequency of the current descriptor MeSH (des) or expression temporal (temp) in the collection.

C. Results and Discussion

To evaluate the effectiveness of integrating temporal criteria in the indexing process, we carried out two sets of experiments:

Experiment 1: Indexing without temporal criteria: indexing process consists of four main steps: (a) Pretreatment (b) term extraction (c) term weighing and (d) generation of best descriptors MeSH.

Experiment 2: Indexing with temporal criteria: indexing process consists of five main steps: (a) Pretreatment (b) Extracting MeSH concepts c) Extracting temporal expressions (d) Generation of Best descriptors MeSH and (e) Generation of best temporal expressions.

We had compared the results of the indexing system TemPIndex to official runs in medical retrieval task 2007 discussed as follows:

Table IV summarizes the results obtained by the participants in medical retrieval task 2007 and system TemPIndex in experiment 1 and experiment 2.

In order to make clear these experimental results, we propose Fig. 9 which presents the precision and MAP value generated by each system. By examining this figure, we can note that the results generated by our system (even without using temporal criteria) close to those of the best run (LIG-MRIM-LIGMU).

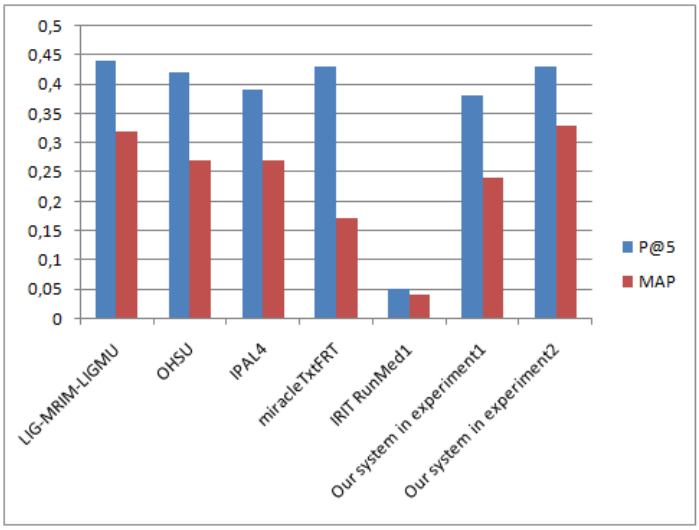

Fig. 9. Experimental results generated by the participants in medical retrieval task 2007 and system TemPIndex.

As shown in Fig. 9, our temporal indexing approach (experiment 2) is really significant compared to the classical indexing approach (experiment 1). The obtained results confirm the well interest to integrate the temporal criteria in the indexing process. For instance, our system displayed 0.38 as precision in the case of experiment 1 and 0.43 in the experiment 2. Thus, we conclude that our temporal indexing approach proposed in this paper would significantly improve the IR performance.

VIII. CONCLUSION

The work developed in this paper outlined a temporal indexing approach using the Mesh thesaurus for representing the semantic content of medical articles. Our proposed approach consists of three main steps. At the first step (Term extraction), being given an article, MeSH thesaurus and the NLP tools, the system TemPIndex extracts the article's lemma. After that, these sets are used in order to extract the Mesh terms existing in the document. At step 3, these extracted terms are weighed by using the measures CSW and SW that intuitively interprets MeSH conceptual information to calculate the term importance. The step 4 aims to recognize the MeSH descriptors that represent the document by using the language model. At step 5, the system TemPIndex extracts the list of temporal expressions. This list is used in step 6 to determine the best temporal expressions.

In order to assess its feasibility, our indexing approach was experimented on through training data sets containing 1000 random MEDLINE citations. An experimental evaluation and comparison of our system with others indexing tools confirms the well interest to use the temporal criteria in the indexing process.

Our future work aims at incorporating a kind of temporal smoothing into the language modeling approach.

ACKNOWLEDGMENT

We thank the deanship of Scientific Research of Majmaah University to support the research project number 27/40.

REFERENCES

[1] B. Andreopoulos D. Alexopoulou and M. Schroeder. *Word Sense Disambiguation in biomedical ontologies with term co-occurrence analysis and document clustering*, IJDMB, 2008.

[2] R. Campos G. Dias AM. Jorge and A.Jatowt. *Survey of temporal information retrieval and related applications*, ACM Computing Surveys 2014; 47(2): 15:115:41.

[3] J. Gamet. *Indexation de pages web*, Report of DEA, universit de Nantes, 1998.

[4] J. Majdoubi H.Loukil M.Tmar and F.Gargouri. *Using the Mesh Thesaurus to Index a Medical Article:Combination of Content, Structure and Semantics*, In KES Journal 16-4 pp. 278285, 2009.

[5] J. Majdoubi H.Loukil M.Tmar and F.Gargouri. *An approach based on langage modeling for improving biomedical information retrieval*, In KES Journal,16-4 pp. 235-246, 2012.

[6] J. Majdoubi H.Loukil M.Tmar and F.Gargouri. *Medical Case-based Retrieval by Using a Language Model: MIRACL at ImageCLEF 2012*, Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012.

[7] J. Majdoubi M.Tmar and F.Gargouri. *Thesaurus based Semantic Representation in Language Modeling for Medical Article Indexing*, In 2th International Conference on Enterprise Information Systems, Volume 2, AIDSS, Funchal, Madeira, Portugal, pp. 65-74, June 8 - 12, 2010

[8] W. Kim A. Aronson and W. Wilbur. *Automatic MeSH term assignment and quality assessment*, in: AMIA, 2001.

[9] D. Dinh and L. Tamine. *Biomedical concept extraction based on combining the content-based and word order similarities,* in: SAC, 2011, pp. 11591163.

[10] W. Dakka L. Gravano and P. Ipeirotis. *Answering general time-sensitive queries*. In CIKM08, 2008.

[11] X. Li and W. B. Croft. *Time-based language models*. In CIKM03, 2003.

[12] K. Massoudi E. Tsagkias M. de Rijke and W. Weerkamp. *Incorporating query expansion and quality indicators in searching microblog posts*. In ECIR11, 2011.

[13] D. Metzler R. Jones F. Peng and R. Zhang. *Improving search relevance for implicitly temporal queries*, 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR 09. New York, NY, USA: ACM, 2009, pp. 700701.

[14] A. neveol. *Automatisation des taches documentaires dans un catalogue de santé en ligne*, Institut National des Sciences Appliques de Rouen, 2005.

[15] A.Aronson J. Mork C.Gay S.Humphrey and W.Rogers. *The NLM Indexing Initiatives Medical Text Indexer*, in: Medinfo, 2004.

[16] A. Aronson. *Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program*, in: AMIA, 2001, pp. 17-21.

[17] M. Baziz. *Indexation conceptuelle guide par ontologie pour la recherche dinformation*. PhD thesis, Univ. of Paul sabatier (2006).

[18] B. Dinh and L. Tamine. *Sense-based biomedical indexing and retrieval*. In: NLDB. (2011) pp- 2435.

[19] M. Efron and G. Golovchinsky. *Estimation Methods for Ranking Recent Information*. In SIGIR11, 2011.

[20] Hiemstra, D. *Using Language Models for Information Retrieval*. PhD thesis, University of Twente (2001).

[21] Yu. PS X. Li and B.Liu. *On the temporal dimension of search*. In: Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters. WWW Alt. 04; New York, NY: ACM, 2004, pp.448449.

[22] I. Mani and G.Wilson. *Robust Temporal Processing of News*, In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, 2000, pp. 69–76

[23] K. Mathews and S. Deepa Kanmani. *A Survey on Temporal Information Retrieval Systems*, International Journal of Computer Applications, November 2012, pp 24-28.

[24] D. Metzler C. Cai and E. Hovy. *Structured Event Retrieval over Microblog Archives*, NAACL-HLT 12, 2012.

[25] H. Mller T. Deselaers T. Deserno and P.Clough, E.Kim and W. Hersh. *Overview of the imageclefmed 2006 medical retrieval and annotation tasks.* In: In: CLEF 2006 Proceedings. Lecture Notes in Computer Science (2006) pp- 595608.

[26] M-H. Peetz E. Meij M. de Rijke and W. Weerkamp. *Adaptive Temporal Query Modeling*, In ECIR12, 2012.

[27] Sheng Lin Peiquan Jin Xujian Zhao and Lihua Yue. *Exploiting temporal information in Web search, Expert Systems with Applications*, 2014, pp. 331 - 341.

[28] D.Trieschnigg P.Pezik V.Lee W.Kraaij F. Jong and D.Rebholz-Schuhmann. *MeSH Up: Effective MeSH Text Classification and Improved Document Retrieval.* Bioinformatics 25(11), (2009), pp- 14121418.

[29] S. Aubin and T. Hamon. *Improving Term Extraction with Terminological Resources*, Advances in Natural Language Processing, 2006.