



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 16830

The contribution was presented at AIRS 2014 :
<https://www.insight-centre.org/content/10th-asia-information-retrieval-society-conference>

To cite this version : Gasmi, Karim and Torjmen-Khemakhem, Mouna and Tamine, Lynda and Ben Jemaa, Maher *A New Combination Method Based on Adaptive Genetic Algorithm for Medical Image Retrieval*. (2014) In: 10th Asia Information Retrieval Society Conference (AIRS 2014), 3 December 2014 - 5 December 2014 (Sarawak, Malaysia).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

A New Combination Method Based on Adaptive Genetic Algorithm for Medical Image Retrieval

Karim Gasmi¹, Mouna Torjmen-Khemakhem¹, Lynda Tamine²,
and Maher Ben Jemaa¹

¹ ReDCAD Laboratory, ENIS Soukra km 3,5, University of Sfax, TUNISIA

² IRIT Laboratory, University of Paul Sabatier, Toulouse
gasmikarim@yahoo.fr

Abstract. Medical image retrieval could be based on the text describing the image as the caption or the title. The use of text terms to retrieve images have several disadvantages such as term-disambiguation. Recent studies prove that representing text into semantic units (concepts) can improve the semantic representation of textual information. However, the use of conceptual representation has other problems as the miss or erroneous semantic relation between two concepts. Other studies show that combining textual and conceptual text representations leads to better accuracy. Popularly, a score for textual representation and a score for conceptual representation are computed and then a combination function is used to have one score. Although the existing of many combination methods of two scores, we propose in this paper a new combination method based on adaptive version of the genetic algorithm. Experiments are carried out on Medical Information Retrieval Task of the ImageCLEF 2009 and 2010. The results confirm that the combination of both textual and conceptual scores allows best accuracy. In addition, our approach outperforms the other combination methods.

Keywords: medical image retrieval, score combination, genetic algorithm, concept mapping.

1 Introduction

Traditional of information retrieval models uses the exact correspondence between the document terms and the query terms to select a document to return to the user. The problem of these models is that the meaning of a word can be expressed in different words, and one word can express different meanings in different contexts. This is due to the richness of the mechanisms of reflection and linguistic expression.

Some studies as [11], [26] have highlighted the inadequacy of document representation based on simple words [11], [26]. The authors in [8] showed that only 20% of Internet users use an application 100% accurately to their needs. Indeed, this wealth can be a source of ambiguity in natural language. To overcome the problem of term ambiguity, some research works proposed the use of

relevance feedback. This technique allow to partially circumvent the problem of term synonymy, but it is not a satisfactory solution.

Recently, many approaches based on semantic indexing have been proposed [2], [6], [19]. The idea of the semantic indexing is to identify all the document terms, project them on an external resource [17](as ontologies) to extract concepts. The authors in [16] have shown that this technique does not cover all the meanings of words. Therefore, this approach needs a semantic resource with a rich terminology covers all terms addressed in the documents of the collection area. All areas require an aspect of optimization, including optimization algorithm we quote the genetic algorithm that is inspired by the genetic operations. Generally, GA is used either for optimization or for selection of parameters [18], [21]. In this paper, we focus on medical image retrieval using surrounding text as the annotation. We propose to compute two scores: the first based on the textual representation of the image annotation, and the second based on conceptual representation of the image annotation. Then, we adapt the genetic algorithm to combine these two scores. Wherein the two weights are two independent numbers. The rest of the paper is organized as follows. Some related work about the combination of two scores is described in Section 2. Following this, in section 3 we give the details of our approach about computing textual and conceptual scores for medical images. In section 4, we describe the genetic algorithm and how we adapt it to combine textual and conceptual scores to one medical image final score. Section 5 presents the corpus used in our experiments, the evaluation metrics and experimental results. Finally, we conclude in Section 6 with possible future work.

2 Related Work

Recently, many studies have highlighted the inadequate representation of documents (D) and requests (R) based on simple words [3], [26] and proposed to explore the semantics textual representation of D / R [19]. The idea of these approaches is to represent the D / R in the form of concepts extracted by projecting the text of the D / R to an external resource such as a semantic ontology. Thus, for a given text, only the words or recognized by the semantic resource sentences will be translated into concepts. We are talking about a conceptual representation or a conceptual indexing [5], [17], [19]. The disadvantage of this design is that indexing is based on the assumption that all the terms of the D / R exist in the semantic resource and therefore the passage of a textual representation to a conceptual representation will be well done. This hypothesis requires the use of a semantic resource with a rich terminology that covers the entire area covered in the documents of the collection. In its approach, Baziz [5] indicated that an ontology (e.g. Wordnet) does not cover all the vocabulary used in the collection. Therefore, Baziz [5] and Hinrich [14] proposed to combine two types of indexing: one using keywords and one using concepts. This idea allows to have significant results. One of the factors that influence performance is probably the quality of the ontology and especially its coverage of the vocabulary in the corpus.

Several methods are typically based on well known voting-based data fusion techniques [9] (e.g., CombMAX, CombLin, CombRank, etc.) that have been used to combine data from different information sources [4], [20], [24]. The existing methods of combination are manual methods, and do not cover the entire search space. In addition, these methods uses dependent weights, which minimizes the number of possible combinations. Table 1 depicts some voting techniques. They are grouped into two categories according to the source of evidence used.

Table 1. Voting techniques used for a multi-annotation

Category	Technique	Score
Rank-based	CombRank	$\alpha R_{text_i} + (1 - \alpha) R_{con_i}$
Score-based	CombMAX	$Max(S_{text_i}, S_{con_i})$
	CombMin	$Min(S_{text_i}, S_{con_i})$
	CombMed	$Median(S_{text_i}, S_{con_i})$
	CombLin	$score_i = \alpha S_{text_i} + (1 - \alpha) S_{con_i}$

3 Computing Textual and Conceptual Scores for Medical Image Retrieval

Four phases are the components of our attempt to develop a retrieval model for medical information. This model combines two different annotation methods: conceptual and textual. The first phase is the pre-processing. The second step is the step of indexing. For the textual indexing, we used Terrier IR platform, the open source search engine written in Java and developed at the School of Computing, University of Glasgow, and the conceptual indexing step is discripted in the next section. The third phase is the calculation of the score for each image by both textual and conceptual model, using the vector space model. The last step is consists of searching the optimal weights of two scores from both textual and conceptual model, this step are based on the genetic algorithm. An overview of our model is done in Fig. 1.

3.1 Mapping Text to Concepts

The aim of this step is to map text into concepts. For this purpose, we start by preprocessing the collection and remove the stop word in order to keep only significant words. After the preprocessing step, we extract the concepts from the text, the switch of text to concepts is realized by an innate procedure system named MetaMap [3], followed by the U.S. National Library of Medicine. MetaMap scrutinized biomedical free-text and identified concepts that are derived from Unified Medical Language System (UMLS). MetaMap is broadly implemented in clinical NLP and IR. By means of MetaMap, the both queries and papers are transformed, [3] endow us with supplementary fine points of this course of action. The extraction of concepts step is a process which allows to highlight the most significant topics of the document by extracting the most relevant concepts. For example, the paradigm of "brain cancer trial" MetaMap

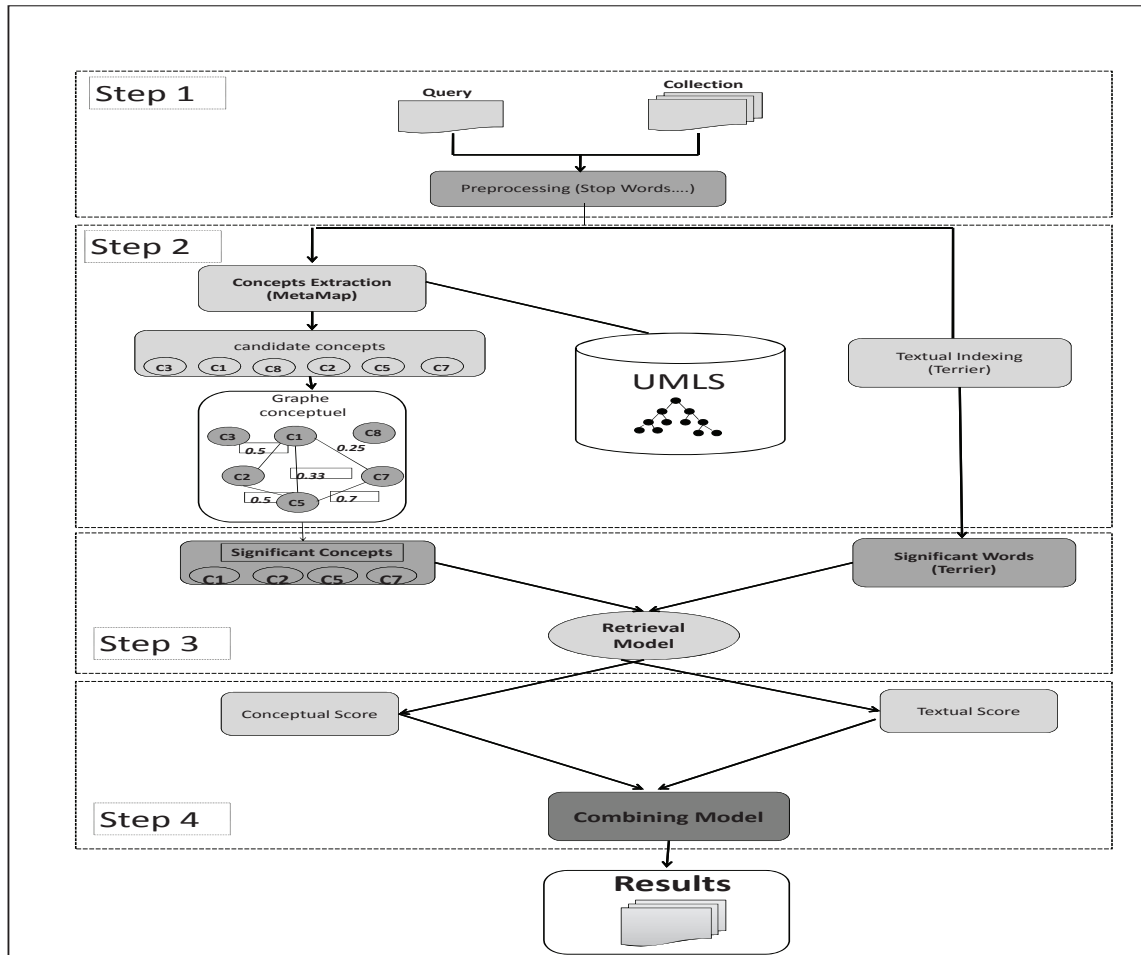


Fig. 1. Overview of our model of medical image retrieval

instrument engenders the concepts ID C0153633 and C0008976, which stand for correspondingly "brain cancer" from the semantic brand "Neoplastic Process" and "Trial" from the semantic brand "Research Activity", as the paramount consequences amid the seven meta-nominees. After a thorough study on the UMLS ontology, we found that it suffers from several problems in relations between concepts such as erroneous relationship.

Authors in [7] reveal that a total of 17 022 (24.3%) of associations (parent-child) between UMLS notions can not be justified according to the semantic categories of concepts. Among several cases that can produce artificial relations, we cite:

- Cases where the semantic category of the child is very broad whereas the parent's semantic type is too specific;
- Situations where the parent-child relationship is erroneous;
- Cases where a parent-child relationship is lacking and have to be added to the UMLS semantic network;
- Conditions where the parent or the child is missing a semantic category;

In fact, these problems can lead to false conceptual annotations. It should, however, be noted those mistaken concepts may be chosen and hence the implicating

of the article will be influenced. To surmount this problem, we use the method of enhancement concepts based on graph theory proposed by Gasmi [12], they put forward the computation of associations between concepts, and no more than associating concepts will be chosen.

Therefore, this method aims to select relevant concepts and to eliminate misconceptions. This method involves calculating distances between two types of concepts generated. Both distances are: one based on the arcs and another based on the information content of a concept. So, they have found as shown by a graph of concepts linked by arcs formed by the hybridization of two types of distances. Thereafter, the concepts which are not connected with other concepts will be deleted [12].

According to Fig. 1, D1 is a document presented by C1, C2, C3, C5, C7 and C8. After the weighting step, D1 is presented only by C1, C2, C5 and C7. C3 and C8 are removed because they do not have any relationship with other concepts.

3.2 Conceptual and Textual Score Fusion

To automate the combination of conceptual and textual results, we chose to implement the genetic algorithm to find the most optimal weight for each method, we followed the following equation:

$$score_i = \alpha score_{text_i} + \beta score_{con_i} \quad (1)$$

$score_{text_i}$: This is the score of the document i, obtained by the textual method
 $score_{con_i}$: This is the score of the document i, obtained by the conceptual method
 α, β : are the weights of the two scores. Wherein the two weights are two independent numbers.

4 Score Combination Based on Adaptive Genetic Algorithm

Several researchers have used a genetic algorithm (GA) to find the most optimal solution in the retrieval information field. Including the use of this algorithm, the authors in [18], [21] use this algorithm to find the optimal combination of query. In its thesis, Yahya [1] proposed a method of combination of multiple similarity measures in the field of chemical information retrieval. This combination based on genetic algorithm can produce better results.

In genetic algorithms [13], the basic idea is to simulate the population evolution process. We start from a population of N solutions to the problem represented by individuals. The randomly selected population is called a relative population. The individual adaptation degree to the environment is expressed by the value of the cost function $f(x)$, where x is the solution that the individual represents. It is declared that more an individual is better adapted to its environment, more the cost of the solution is lower.

For each problem to solve, a fitness function f should be provided. Its choice is crucial for the proper functioning of the algorithm. Given a chromosome, the fitness function must return a numeric value that represents its utility. This score will be used in the selection process of the parents, so that the fittest individuals will have a greater chance of being selected.

Within this population, a random selection of one or both parents is done, producing a new solution through genetic operators such as crossover and mutation. The new population, obtained by the choice of N individuals among parent and child populations is called next generation. By repeating this process, it produces a richer individuals that are better adapted to the population. Fig. 2 shows the procedure of the proposed GA weighting combining-annotation method as follows:

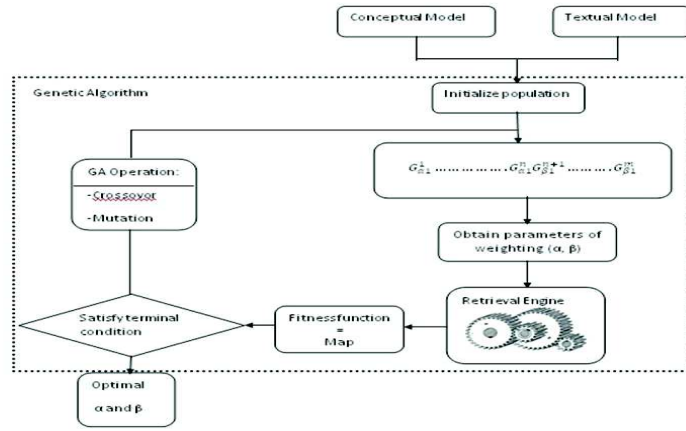


Fig. 2. Genetic algorithm process

Most operations require the Genetic Algorithm fitness function to calculate the adaptation of the individual. In the proposed approach, our fitness function depends on two factors: (1) it depends on results returned by the retrieval model according to mean average precision (MAP), MAP is taken as a user who selected in each iteration the relevant documents found for calculated the Fitness for each chromosome; (2) the absolute value of subtraction between the two weights is added. The aim of the second part is to minimize the cost of any influence of annotations compared to the other on the Fitness function 2.

$$Fitness = MAP - \frac{1}{\|\alpha - \beta\|} \quad (2)$$

With:

$MAP = \sum_{j=1}^M wq_i * AP_{qj} / M$
 where APq is the average precision of a query q , and M is the number of queries.

$$Precision = |relevanddocuments \cap retrievaldocuments| / |retrievaldocuments|$$

In our proposed method, we use the MAP measure, that is connected directly by the number of relevant documents found as he shows his formula, MAP is taken as a user who selected in each iteration the relevant documents found for calculated the Fitness for each chromosome.

- Step 1 : Encodes the chromosomes and the parameters representing the weighting indexing method as a binary string. Fig. 3 shows an example of a chromosome creation from two weights.

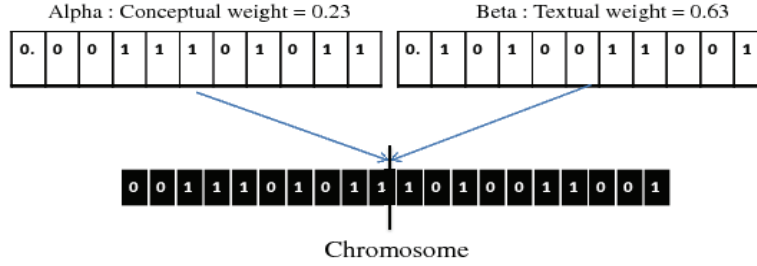


Fig. 3. Coding of a Chromosome

- Step2 : Initializes the population and produces the initial population of chromosomes arbitrarily.
- Step3 : The fitness for each chromosome must be computed, this is related to the calculated results obtained by the fitness equation.
- Step 5 : The main feature is that the fitness value decreasing during the last M generation or N is reached as the maximum generation number.
- Step 6 : The iteration process stops only when the two criteria are achieved. Otherwise you have to move to step 5.
- Step 7 : To generate a offspring generation, genetic operations should be performed. The offspring generation has as components : crossover, mutation and tournoi reproduction.

4.1 Crossover

The main operator acting on the population of parents is the crossover, which is applied with a certain probability, called crossover rate Pc (typically close to unity). The crossing is to choose two individuals represented by their chains of genes randomly selected from the general population and define random or more crossing points. The new children are, then, created in inter changing different parts of each string.

Let G_α and G_β be two selected parent chromosomes, which are represented respectively as follows:

$$G_\psi = G_\psi^1, G_\psi^2, \dots, G_\psi^n$$

$$G_\omega = G_\omega^1, G_\omega^2, \dots, G_\omega^n$$

Then:

$$G_\xi = \begin{cases} G_\psi^i & \text{if } 1 < i < k \\ G_\omega^i & \text{else} \end{cases}$$

With :

$$k = \text{random}(1, n - 1), 1 < k < n - 1;$$

4.2 Mutation

This operation protects genetic algorithms premature loss of relevant information. It allows introducing some information in the population, which could be lost during the crossing operation. Thus, it helps to maintain diversity, useful for a good exploration of the research area. The mutation operator is applied with a certain probability, called mutation rate Pm, typically between 0.05 and 0.10. In binary code, the mutation involves changing a 1 bit at bit 0, and vice versa, for each bit of the string, with the probability Pm. Let G_α be the parent chromosome,

$$G_\alpha = G_\alpha^1, G_\alpha^2, \dots, G_\alpha^n$$

$$\text{Then : } G_\alpha^k = 1 - G_\alpha^1$$

$$\text{With : } k = \text{random}(1, n - 1), 1 < k < n - 1;$$

5 Evaluation

5.1 Data Sets and Evaluation Metrics

To evaluate our approach, we use the 2009 and 2010 ImageCLEF collection composed respectively of 74,902 and 77000 medical images and annotations associated with them. This collection contains images and captions from Radiology and Radiographic, two Radiological Society of North America (RSNA) journals. The number of queries is 25 from 2009 collection and 16 from 2010 collection. They are queries selected by experts in information retrieval company to evaluate results by ImageCLEF collection. Table 2 contains the parameters used for the genetic algorithm.

Table 2. Parameters of GA

GA property	Value/Method
Size of generation	100
Initial population size	30
Selection method	Tournoi
Number of crossover points	1
Crossover Probability	0.9
Mutation method	Uniform mutation
Mutation Probability	0.05

To evaluate our approach, we have used P@5, P@10 and Mean Average Precision MAP. To statistically validate our results, we used the signed-rank test of the Wilcoxon test [27] which is the non-parametric equivalent of the paired sample test. This test consists in evaluating a value of significance $p \in [0, 1]$ which

estimates the probability that the difference between the two methods is due to chance. We can thereby conclude that two methods are statistically different when $p \leq \alpha$, where $\alpha \leq 0.05$ is commonly used [15]. More precisely, the more $p \rightarrow 0$, the more two methods are supposed to be different.

In our experiments, we consider that the difference between two methods is significant when $p \leq 0.1$, and it is very significant when $p \leq 0.05$.

5.2 Retrieval Model

For both textual and conceptual representations, we have used the vector space model [22,23] to compute the similarity between documents and queries. Each dimension in this model represents a term or an ontology concept. Documents and queries are represented by a vector with n dimensions where n is the number of the terms or ontological concepts [25].

The value of system relevance is calculated using the similarity function $RSV(Q, d)$ (Retrieval Status Value) where Q is a query and D_j is a document. The RSV is calculated as follows:

$$RSV(Q, D_j) = \frac{\sum_{i=1}^N w_{qi} \cdot w_{ij}}{\left(\sum_{i=1}^N w_{qi}^2\right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^N w_{ij}^2\right)^{\frac{1}{2}}} \quad (3)$$

– The term/concept frequency is:

$$w_{ij} = cf_{ij} * idf_{ij} \quad \text{and} \quad w_{qi} = cf_{qi} * idf_{qi} \quad (4)$$

where : cf_{ij} (respectively cf_{qi}) is the number of occurrences of the concept/term C_i in the document D_j (respectively the query q); And idf_{ij} that stands for the inverse document frequency, is equal to divide the number of documents containing the concept/term i , by the number of all documents in the collection.

5.3 Experimental Results

In this section, we present the different results obtained by different combination methods. Results according to P@5, P@10 and MAP measures for 2009 and 2010 data sets are presented in Table 3.

We note that our model results are better than conceptual and textual model. This observation affirms that the use of a combination solution improves the retrieval accuracy. Consequently, we can conclude that the use of both model conceptual and textual is a good solution, on the one hand to improve the outcome, and on the other to improve the semantic representation of the document.

Best results are obtained by our combination method for both data sets. Gains in MAP measure are presented in Table 4.

Symbol * after the gain indicates statistical significance using the Wilcoxon test at $p \leq 0.1$. Symbol ** after the gain indicates statistical significance using the Wilcoxon test at $p \leq 0.05$. Table 4 shows how our indexing approach is statistically significant and the improvements of our method compared to other

Table 3. Comparison between conceptual, textual, MaxComb, LinComb, RankComb and GentComb Methods

Category	P@5	P@10	MAP	Category	P@5	P@10	MAP
ImageClef 2010 Collection				ImageClef 2009 Collection			
Conceptual	0.362	0.293	0.208	Conceptual	0.480	0.450	0.264
Textual	0.399	0.393	0.298	Textual	0.620	0.600	0.386
MaxComb	0.333	0.3733	0.269	MaxComb	0.520	0.516	0.3451
LinComb	0.400	0.440	0.324	LinComb	0.640	0.600	0.390
RankComb	0.426	0.433	0.319	RankComb	0.640	0.584	0.381
GentComb	0.413	0.440	0.329	GentComb	0.660	0.630	0.398

Table 4. Gain between GentComb Method and the other Methods for 2009 and 2010 Collection

Category	P@5	P@10	MAP
ImageClef 2010 Collection			
GentComb /Conceptual	14% (*)	50% (**)	58.1% (**)
GentComb /Textual	3.5%	11.9% (*)	10% (*)
GentComb /CombMax	24.02% (**)	17.96% (*)	22.30% (**)
GentComb/CombLin	3.25%	0%	1.54%
GentComb/CombRank	-3.1%	1.61%	3.16%
ImageClef 2009 Collection			
GentComb/Conceptual	+37.5% (*)	+40% (**)	+50.75% (**)
GentComb/Textual	+6.45%	+5%	+3.1%
GentComb /CombMax	26.92% (**)	22.09% (**)	9.21% (**)
GentComb /CombLin	3.1%	5%	2%
GentComb Method/CombRank	3.1%	7.87%	4.46%

methods rates. We computed the Wilcoxon test between means of each ranking obtained by each indexing method. The Wilcoxon test validated our method for the 2010 query set ($p \leq 0.1$) compared to the textual baseline. Additionally, Table 4 shows that our proposed method (GENTComb) is statistically significant compared to the conceptual baseline ($p \leq 0.05$) for the 2010 and 2009 query set. Concerning the Map measure, our method outperforms every combination method rated at a result and on the other hand the proposed method uses two annotation-weight which are two automatic and separate numbers, unlike other method used an manually and dependent weight.

5.4 Comparison of Our Method with Official Submissions of Medical IMAGE CLEF 2009 and 2010

In this section, we compare our method with runs submitted for medical ImageCLEF 2009 and 2010. Only runs based on textual approaches are taken into account because our method is based only on textual annotation of images. For

Table 5. Comparison of our method with official runs of IMAGE CLEF 2009

	Rang	MAP	P5	P10
LIRIS maxMPTT extMPTT	1	0.43	0.70	0.66
Our Method (GentComb)	2	0.398	0.660	0.630
sinai CTM t	3	0.38	0.65	0.62
york.In expB2c1.0	4	0.37	0.61	0.60
ISSR text 1	5	0.35	0.58	0.56
ceb-essie2-automatic	6	0.35	0.65	0.62
deu run1 pivoted	7	0.34	0.58	0.52
clef2009	8	0.34	0.67	0.60
BiTeM EN	9	0.32	0.52	0.50
UNTtextrf1	10	0.26	0.53	0.44
OHSU SR1	11	0.18	0.59	0.54
MirEN	12	0.17	0.62	0.55
uwmTextOnly	13	0.13	0.44	0.40
Alicante-Run3	14	0.13	0.34	0.36

Table 6. Comparison of our method with official runs of IMAGE CLEF 2010

	Run Type	MAP	P5	P10
Our Method (GentComb)	Automatic	0.329	0.440	0.440
Information Processing Laboratory	Automatic	0.3235	0.3109	0.4687
OHSU	Automatic	0.3029	0.344	0.4313
UESTC	Automatic	0.2789	0.297	0.3125
ISSR	Automatic	0.2583	0.2667	0.3187
HES-SO VS	Automatic	0.2568	0.278	0.35
ITI	Automatic	0.188	0.2158	0.375
Bioingenium Research Group	Automatic	0.1005	0.1289	0.1875
XRCE	Feedback	0.2925	0.3027	0.4125
SINAI	Feedback	0.2672	0.2683	0.4125
XRCE	Not applicable	0.338	0.3828	0.5062

Medical IMAGE CLEF 2009, Table 5 compares our method (Gent-Comb) with the other official submissions.

For the best result of the IRIS team in 2009, the authors use a combination of two analysis tools, one with the MiniPar parser and the second with the TreeTagger tool. Then, they compare the combination of conceptual representation with the Kullback-Leiber divergence instead of the combination with likelihood function. Two RSV are used : one based on log-probability and the other based on divergence function. The best result is obtained with the log-probability method. This method is improved by using a relevance feedback extension of queries with the n first returned documents. Best results are obtained with $n = 100$. Using our proposition, we are ranked second. We recall that we do not use the relevance feedback.

For Medical IMAGE CLEF 2010, Table 6 compares our method (Gent-Comb) with the other official submissions.

Organisers consider the XRCE laboratory run is wrong and not applicable. Then, the best official result in 2010 is the one of Information Processing Laboratory team, the authors indexed the text usind Lucene framework. For the retrieval, first, the documents are expanded with the Mesh-terms, after that a combination of textual score and the score obtained by Mesh terms is released. To compute the fibal score of each medical image, a function proposed by Fang and Zhai [10] is used. Thanks to our proposed method, we are classified on first rank.

6 Conclusion and Future Works

The purpose of this paper is to better study the importance of combining two retrieval methods; textual and conceptual method. To do this, we used the UMLS as a specific ontology for the medical domain, which allowed us to extract concepts representative of each document. Our approach begins with extracting concepts by MetaMap tool, after words, it uses the graph theory to retain only the relevant concepts, and later to refine the results obtained we used the genetic algorithm to combine the two types of indexing: textual and conceptual. Our results showed that the combination method can on the one hand to improve the efficiency of our model and secondly to improve the semantic representation of the document.

References

1. Abderahman, Y.A.: Fusion of similarity measures using genetic algorithm for searching chemical database (2008)
2. Abdulahhad, K., Chevallet, J., Berrut, C.: Solving concept mismatch through bayesian framework by extending umls meta-thesaurus. In: Conference en Recherche d'Informations et Applications - CORIA 2011, pp. 311–326 (2011)
3. Aronson, A.R.: Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In: Annual Symposium AMIA, pp. 17–21 (2001)
4. Aslam, J.A., Montague, M.: Models for metasearch. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 276–284 (2001)
5. Baziz, M.: Indexation conceptuelle guidée par ontologie pour la recherche d'information. In: Thesis, University Paul Sabatier, Toulouse, France (2005)
6. Chevallet, J., Lim, J., Le, D.T.H.: Domain knowledge conceptual inter-media indexing: application to multilingual multimedia medical reports. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, CIKM 2007, pp. 495–504 (2007)
7. Cimino, J.J., Min, H., Perl, Y.: Consistency across the hierarchies of the umls semantic network and metathesaurus. *Journal of Biomedical Informatics* 36(6), 450–461 (2003)
8. Crestani, F.: Exploiting the similarity of non-matching terms at retrieval time. *Journal of Information Retrieval* 2, 25–45 (1999)
9. Dinh, D., Tamine, L.: Combining global and local semantic contexts for improving biomedical information retrieval. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) ECIR 2011. LNCS, vol. 6611, pp. 375–386. Springer, Heidelberg (2011)
10. Fang, H., Zhai, C.: An exploration of axiomatic approaches to information retrieval. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2005, pp. 480–487 (2005)
11. Fieschi, M., Aronson, A.R., Morfy, J.G., Gay, C.W., Humphrey, S.M., Rogers, W.J.: The.nlm indexing initiative's medical text indexer. In: Proceedings of the 11th World Congress on Medical Informatics Demner-Fushman and Lin Answering Clinical Questions, pp. 268–272 (2004)

12. Gasmi, K., Khemakhem, M., Jemaa, M.B.: A conceptual model for word sense disambiguation in medical image retrieval. In: AIRS, pp. 296–307 (2013)
13. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc. (1989)
14. Hinrich, J.O.P.: Information retrieval based on word senses (1995)
15. Hull, D.: Using statistical testing in the evaluation of retrieval experiments. In: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1993, pp. 329–338 (1993)
16. Jimeno-Yepes, A.J., McInnes, B.T., Aronson, A.R.: Exploiting mesh indexing in medline to generate a data set for word sense disambiguation. BMC Bioinformatics 12, 223 (2011)
17. Khan, L., McLeod, D., Hovy, E.: Retrieval effectiveness of an ontology-based model for information selection. VLDB Journal, 71–85 (2004)
18. Lopez-Herrera, A., Herrera-Viedma, E., Herrera, F.: Applying multi-objective evolutionary algorithms to the automatic learning of extended boolean queries in fuzzy ordinal linguistic information retrieval systems. vol. 160, p. 2192–2205 (2009)
19. Mihalcea, R., Tarau, P., Figa, E.: Pagerank on semantic networks with application to word sense disambiguation. In: International Conference on Computational Linguistics (COLING), pp. 1126–1132 (2004)
20. Ogilvie, P., Callan, J.: Combining document representations for known-item search. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, SIGIR 2003, pp. 143–150 (2003)
21. Malo, P., Pyry Siitari, A.S.: Automated query learning with wikipedia and genetic programming 194, 86–110 (2013)
22. Salton, G.: The smart retrieval system: Experiments in automatic document processing (1970)
23. Salton, G., McGill, M.: Introduction to modern information retrieval (1983)
24. Shaw, J.A., Fox, E.A., Shaw, J.A., Fox, E.A.: Combination of multiple searches. In: The Second Text REtrieval Conference (TREC-2), pp. 243–252 (1994)
25. Ventresque, A., Cazalens, S., Lamarre, P., Valduriez, P.: Improving interoperability using query interpretation in semantic vector spaces. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 539–553. Springer, Heidelberg (2008)
26. Weeber, M., Mork, J., Aronson, A.: Developping a test collection for biomedical word senes disambiguation. In: Annual Symposium. AMIA Symposium, pp. 746–750 (2001)
27. Wilcoxon, F.: Individual Comparisons by Ranking Methods. Biometrics Bulletin pp. 80–83 (1945)