Word Indexing Versus Conceptual Indexing in Medical Image Retrieval
(ReDCAD participation at ImageCLEF Medical Image Retrieval 2012)

Karim Gasmi, Mouna Torjmen-Khemakhem, and Maher Ben Jemaa

Research unit on Development and Control of Distributed Applications (ReDCAD),
Department of Computer Science and Applied Mathematics,
National School of Engineers of Sfax, University of Sfax
{karimgasmi@yahoo.fr,torjmen.mouna@redcad.org,maher.benjemaa@enis.rnu.tn}
http://www.redcad.org

Abstract. This paper presents our participation in medical image retrieval task of ImageCLEF 2012. Our aim is to study the effectiveness of using conceptual indexing comparing to word indexing in medical image retrieval. For this aim, we have used in the one hand the Terrier tool for textual indexing and for textual retrieval, and on another hand, the MetaMap tool for conceptual indexing and Vector model for conceptual retrieval.
More precisely, the run of the BM25 model is considered as a baseline. For textual indexing, we tried to compare different weighting formulas. However, for conceptual indexing, we Used BM25 model results to extract concepts and rerank results using vector model.
Results show that the use of the textual indexing is more useful than the conceptual indexing. However, the conceptual indexing improves the result of some queries, which encourages us to continue the study of conceptual indexing and retrieval.

Keywords: medical image retrieval, information retrieval model, reranking, conceptual indexing, metamap

1 Introduction

Classical Information Retrieval (IR) models retrieve documents that have the same words (at least in part) that the query. But meaning can be expressed by different words, and the same word can express different meanings in different contexts. This false assumption is exactly the pitfalls of traditional approaches to IR. Overcome these limitations is the subject of several recent research projects. This is particularly true of the IR approach known as "based concepts."
The choice of information retrieval model is a crucial task, which directly affects the result of any system of information retrieval, for that, we decided to work on the evaluation of different information retrieval models using the Terrier IR platform [1] and we tried to improve these models by using a conceptual indexing.

---

[1] http://terrier.org/docs/v3.5/

Our model uses two types of indexing: words and or concepts to re-rank the result obtained by the BM25 model. The goals of this research are:

1. To study the influence of using of each retrieval model on the information retrieval system performance,
2. To study the influence of using two series retrieval models on the information retrieval system performance.
3. To study the effects of using concepts for indexing on the information retrieval system performance.

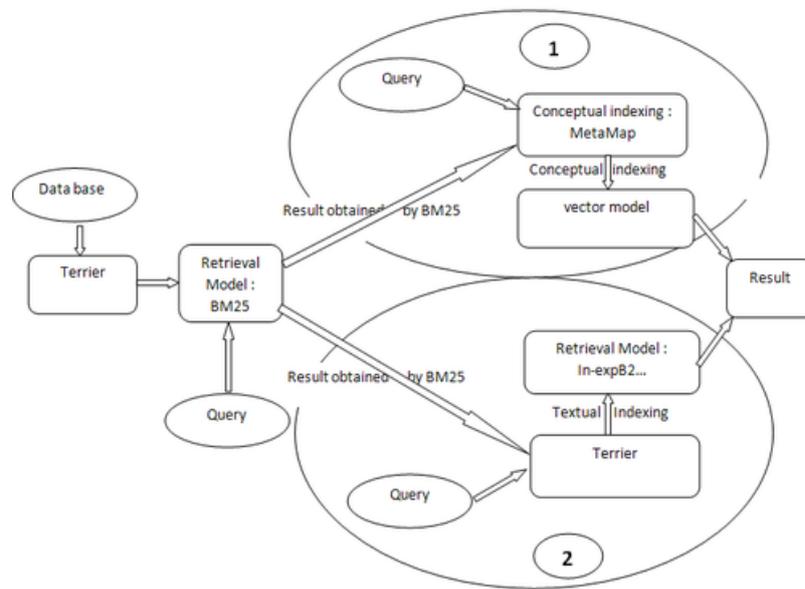We have summarized the two indexing methods in the following figure (1).

Fig. 1. An illustrationof: (1) Conceptual Indexing (2) Textual Indexing

Our paper is organized as follows: in section 2, we describe the models of image retrieval used in different runs. Then we describe in Section 3, the conceptual indexing of medical image and before the conclusion we are done with the section 4, which describes the run and the result obtained.

2   Word Indexing of medical images

The manual textual indexing of images is usually performed by a librarian named iconographer. Its role is to categorize and index images by associating them to categories and groups of words, often taken from a thesaurus, to quickly find the

images. Unfortunately, the choice of terms for indexing is a problem for picture researchers, because it is impossible that the user choose the same keywords as those chosen by the iconographer. So the indexing of an image is subjective, because several indexing are possible.

Despite its subjectivity, manual indexing is an effective method to associate a meaning to images. However, to index a large volume of images, this work quickly becomes tedious or impossible, which is not the case for automatic indexing.

The automatic textual indexing of images is to associate words in an image using a computer system without human intervention. The indexing textual images on the web can be done from the words in the page title or the most frequent or relevant words to this page.

Every system of information retrieval needs a weighting model , but the weighting term process must provide an iconic representation, compact and informative content of the documents regarding the terms of queries. It should provide an indicator of importance to discriminate the terms towards each other.

Although several approaches and techniques have been developed using this factor of importance (weight terms), yet they almost all use these two terms:

TF (term frequency): a term more frequently in a document, it is more important in the document.

IDF (Inverse Document Frequency): a term is uncommon in the collection, it is more important in the document.

2.1 Model BM25 [5]

This is a ranking function used by search engines to rank matching documents according to their relevance to a given query. It is a probabilistic model (1) :

$$BM25 = \sum_{t \in q \cap d} \left(\frac{tf}{tf + k_1.n_b}.log\left(\frac{N - df_t + 0.5}{df_t + 0.5} \right).qtf \right) \tag{1}$$

with:

- tf: frequency of term occurrences,
- N: total number of documents in the collection,
- dft: number of documents containing a term t,
- qtf: frequency of occurrences of a term t in the query,
- $k1$: parameters influencing the frequency of terms that is adjusted to 1.2 by default,
- nb: normalization factor is calculated as follows:

$$n_b = (1 - b) + b.\frac{tl}{tl_{avg}} \tag{2}$$

with:
  - tl: Number of terms in the document (document length),
  - tl_{avg} : Average number of words in a document,

2.2   Model TF_IDF (Term Frequency Inverse Document Frequency)

This model works by determining the relative frequency of words in a specific document compared to the inverse proportion of that word over the entire document corpus (3) [2].

$$TF - IDF = Roberston\_tf * idf * K_f \qquad (3)$$

with:

- $idf = log\left(\frac{n_d}{d_f+1}\right)$
- Roberston_tf=$k_1 * \frac{tf}{tf+k_1*\left(1-b+\frac{b*d}{tl_{avg}}\right)}$
- tf: The term frequency of the term in the document
- d_l: The document's length
- d_f: The document frequency of the term
- K_f: The term frequency in the query
- n_d: Nombre de documents

b= 0.75 , $k_1 = 1.2$;

2.3   Model In_expB2 [1][4]

Inverse expected document frequency model for randomness, the ratio of two Bernoulli's processes for first normalisation, and Normalisation 2 for term frequency normalisation (4).

$$\omega\left(t,d\right) = \frac{F+1}{n_t.\left(tfn+1\right)}\left(tfn.log_2\left(\frac{N+1}{n_e+0.5}\right)\right) \qquad (4)$$

2.4   Model BB2 [1][4]

Bose-Einstein model for randomness, the ratio of two Bernoulli's processes for first normalisation, and Normalisation 2 for term frequency normalization (5).

$$\omega\left(t,d\right) = \frac{F+1}{n_t.\left(tfn+1\right)}\left(-log_2\left(N-1\right) - log_2\left(e\right) + f\left(N+F-1, N+F-tnf-2\right) - f\left(F, F-tnf\right)\right) \qquad (5)$$

- $\omega\left(t,d\right)$ is the within-document term weight of the term t in the document d,
- tf is the within-document frequency of the term t in the document d,
- F is the term frequency of the term t in the whole collection,
- N is the number of documents in the collection,
- nt is the document frequency of the term t,
- λ is given by $\frac{F}{N}$,

---

[2] http://terrier.org/docs/v3.5/

- $n_e = N.\left(1 - \left(1 - \frac{nt}{N}\right)^F\right),$
- $f(n, m) = (m + 0.5).log_2\left(\frac{n}{m}\right) + (n - m)\,log_2 n,$
- $tfn = tf.log_2\left(1 + c.\frac{avg-l}{l}\right),$
- $tfn_e = tf.log_e\left(1 + c.\frac{avg-l}{l}\right)$

where c is a parameter. l and $avg - l$ are the document length of the document d and the average document length in the collection respectively.

3   Conceptual Indexing of medical images

For extracting concept from "caption + title" of each image, we choose to use MetaMap[3] which performs the following steps [3]

- *1-Parse the text into noun phrases*

- *2-Look for variants for each nominal sentence, with a variant consists of a noun phrase or words with all its variant spellings, abbreviations, acronyms, synonyms, inflectional and derivational variants, and meaningful combinations of these;*

- *3-Look for different candidates from all metathesaurus strings containing one of the variants found in step 2;*

- *4-By using an evaluation function, compute the mapping from the noun phrase and calculate the strength of the mapping, this step is performed for each candidate, finding during stage 3.;*

- *5-Combine candidates involved with disjoint parts of the noun phrase, recompute the match strength based on the combined candidates, and select those having the highest score to form a set of best Metathesaurus mappings for the original noun phrase.*

The evaluation function used to calculate the strength of the mapping, is based on four components: centrality, variation, coverage, and cohesiveness. A normalized value between 0 (the weakest match) and 1 (the strongest match) is computed for each of these components.

4   Evaluation

The collection used in the medical retrieval task is sized of 2.5 GB and it consists of 306,528 documents. The number of queries is 22[2].

---

[3] http://metamap.nlm.nih.gov/

For the word indexing runs, we used Terrier IR platform[4], the open source search engine written in Java and developed at the School of Computing, University of Glasgow.

For the conceptual indexing runs, MetaMap has been used: a document (respectively a query) is represented as a set of weighted concepts extracted using the UMLS [5].

For all runs, *title*, *caption* and *abstract* were used to represent images. Our runs are described as follows:

4.1   Evaluation of the use of textual indexing

– Run1-**Terrier_CapTitAbs_BM25b0.75 (BM25)**: this is our official run, which uses BM25 as an information retrieval model and Terrier as a tool for textual indexing.
– Run2-**Terrier_CapTitAbs_BB2 (BB2)**: this run uses model BB2,which is described in the 2.4 subsection, to compare the result obtained with that obtained by BM25.
– Run3-**Terrier_CapTitAbs_In_expB2 (In_expB2)**: using the In_expB2 model,which is described in the 2.3 subsection, to compare the result obtained with that obtained by BM25.
– Run4-**Terrier_CapTitAbs_TF_IDF (TF_IDF)**: TF_IDF used for the calculation of weight, and as an information retrieval model, to compare the result obtained with that obtained by BM25 as an information retrieval model .

Table 1. Results obtained by using different information retrieval models

	BM25	BB2	In_expB2	TF_IDF
Map	0.1678	0.1893	0.1920	0.1632
P@5	0.3273	0.3636	0.3545	0.3182
P@15	0.2455	0.2182	0.2242	0.2303
P@30	0.1712	0.1712	0.1606	0.1606

we used Map (Mean Average Precision) as evaluation measure, results of table 1 show the difference between the four information retrieval models. According to the results, we observed that In_expB2 model has higher Mean Average Precision than other models. TF_IDF model has lower Mean Average Precision

---

[4] http://terrier.org/docs/v3.5/
[5] http://www.nlm.nih.gov/research/umls/

than other models.

- Run5-**Terrier_CapTitAbs_ BM25-DFR_BM25**: For this run, we used the model DFR_BM25 for the task of re-rank the result obtained by the run-1
- Run6-**Terrier_CapTitAbs_ BM25-In_expB2**: we used the same principle as that of run-5, but we used the In_expB2 model to re-rank.
- Run7-**Terrier_CapTitAbs_ BM25-TF_IDF**: For this run, we used the model TF_IDF for the task of re-rank the result obtained by the run-1

Table 2. Evaluation the results obtained by using two retrieval models together

	BM25	BM25-DFR_BM25	BM25-In_expB2	BM25-TF_IDF
Map	0.1678	0.1591	0.1718	0.1517
P@5	0.3273	0.2909	0.3000	0.2909
P@15	0.2455	0.2000	0.2061	0.2182
P@30	0.1712	0.1409	0.1530	0.1500

To improve the results obtained by the BM25 model, we try to sort the result by another model.Table 2 show obtained result.These Map confirm that the use of another model to re-rank the baseline result can help and improve the result obtained by BM25 only.

But the BM25 model gives the best results for P @ 5. So according to the needs of information retrieval system, we can choose between different models.

4.2   Evaluation of the use of conceptual indexing

- Run8-**Terrier_CapTitAbs_ BM25-Concept-Vector**: For this run, and after running run-1, we have implemented MetaMap to extract the concept and vector model as model of information retrieval for re-rank the results obtained.
- Run9-**Terrier_CapTitAbs_BM25-Concept_seuil-vector**: same principle as the Run8, but a threshold equals to 0.7 is used for concepts that have participated in the task of information retrieval, this threshold is calculated from the evaluation function, what described in Section 3.

With the BM25 model, as shown in Table 3, we obtain a result which is really better than that achieved by the use of concepts. However, analyzing results query by query, we discovered that , using conceptual indexing can improve results for some queries.

Table 3. Results obtained by using concept indexing

	BM25	BM25-Concept-Vector	BM25-Concept_seuil-vector
MAP	0.1678	0.0778	0.0623
Query 7	0.1073	0.1207	0.0488
Query 20	0.0242	0.0400	0.0333
Query 21	0.1255	0.1840	0.1564

Also, the results obtained by the concepts can be improved if used right from the beginning, without indexing text as a first step, because it perhaps that the result obtained by BM25 affects negatively on the result achieved by the concept. Because these are two different types of indexing.

We can improve also the result obtained by the use of the concept by the implementation of another model instead of vector model.

5   Conclusion and future Work

Along this paper, we have compared the use of word indexing and conceptual indexing.

Results show that using word indexing is better than using conceptual indexing. However, we note that conceptual indexing improves significatively some queries. This finding encourages us to more work in conceptual indexing, and also in conceptual retrieval. In future work, we plan to continue studying conceptual indexing and to propose a conceptual retrieval model for medical image retrieval. We plan also to propose a mixed approach that combines the visual appearance of an image and conceptual description.

References

1. Ben He and Iadh Ounis. A query-based pre-retrieval model selection approach to information retrieval. In *RIAO*, pages 706–719, 2004.
2. Jayashree Kalpathy-Cramer Dina Demner Fushman Sameer Antani Ivan Eggel Henning Mller, Alba Garcia Seco de Herrera. Overview of the imageclef 2012 medical image retrieval and classification tasks. *CLEF 2012 working notes, Rome, Italy*, 2012.
3. Quanzhi Li and Yi fang Brook Wu. Identifying important concepts from medical documents. pages 668–679, 2006.
4. Sobhana N.V. Enhancing retrieval of geological text using named entity disambiguation. *International Journal of Emerging Technology and Advanced Engineering*, 2(1):2250–2459, 2012.
5. Stephen E. Robertson, Steve Walker, Micheline Hancock-Beaulieu, Aarron Gull, and Marianna Lau. Okapi at TREC. In *Text REtrieval Conference*, pages 21–30, 1992.