An Investigation of the Effectiveness of Concept-based Approach in Medical Information Retrieval GRIUM @ CLEF2014eHealthTask 3

Wei Shen, Jian-Yun Nie, Xiaohua Liu, and Xiaojie Liu

C.P. 6128, succursale Centre-ville Montreal,
Quebec CANADA H3C 3J7
{shenwei,nie,xiaohul,xiaojie}@iro.umontreal.ca

Abstract. In our participation in the CLEF 2014 eHealth task 3a, we investigate the effectiveness of concept-based retrieval techniques on medical IR. Concepts are determined using the existing resources and tools: UMLS Metathesaurus and MetaMap. We tested several methods based on concepts. Although some of these methods lead to slight improvements in retrieval effectiveness over a traditional bag-of-words method, the impact of the rich domain ressource is lower than we expected. So the whole question on whether and how such a resource can help improve medical IR effectiveness remains open. In this report, we describe the methods tested as well as their results.

Keywords: concept-based retrieval, query expansion, language model, UMLS, MetaMap, Indri

1 INTRODUCTION

Our experiments on CLEF 2014 eHealth Task 3 [1, 2] aim to investigate the effectiveness of concept-based approaches in Medical IR. Medicine is possibly the area in which there are the best manually constructed resources for identifying concepts. Metathesaurus [24] is a large thesaurus in medicine, gathering resources such as MeSH [25], Snomed [26], etc. Tools for identifying and disambiguating concepts in texts, such as MetaMap [27] have also be developed. In Metathesaurus a term is linked to a large number of other terms, denoting his synonyms, lexical variants, abbreviations and hypernyms, hyponyms, etc. Intuitively, the availability of those resources and tools should result in better IR effectiveness than the traditional bag-of-words approaches. However, the previous experimental results have been disappointing. For example, [3] did not observe any improvement using concepts recognized from texts. [4] exploited a statistical thesaurus and obtained 2.2% improvement. [5] used MetaMap to recognize concepts from texts, and used the concepts in query expansion. This led to an improvement of 4.4% over the bag-of-words approach. A number of other studies [6–21] have also used different resources and tools. However, the global conclusions are similar: In some cases, slight improvements are obtained, in other cases, no improvements or even degradations are observed. Overall, the experimental results using medical resources and tools for IR have been lower than one expects. The whole question remains: can we really benefit from the rich resources and tools in the medical area to improve IR effectiveness? Are they related to the way that the resources and tools are used?

In our experiments in CLEF 2014, we would like to examine a few more possible approaches to take advantage of medical concepts. In our experiments, we use MetaMap to recognize medical concepts from documents and queries. MetaMap identifies concepts from a text (document or query). From the concept IDs (CUI - Concept Unified Identifier) identified, we can further identify the concept word sequence (SUI - String Unified Identifier). Our experiments will test several ways to exploit either CUI or SUI. In prticular, we will focus on query expansion using concepts, as query expansion has been shown to be relatively effective in the previous experiments on medical IR.

2 METHODS

Let us first describe the bag-of-words baseline method to which our methods will be compared. Then we will describe how concepts are determined and used in our approaches.

2.1 Baseline

As baseline, we use a traditional approach based on language modeling, with Dirichlet smoothing[23]. We use Indri as the basic experimental platform for all the methods. For the baseline method, the score of a document D for a query Q is determined as follows:

$$S(Q, D) = \frac{1}{n} \sum_{i=1}^{n} \log P(q_i|D) \tag{1}$$

where n is the length of query and $P(q_i|D)$ is adjusted by Dirichlet smoothing,

$$P(q_i|D) = \frac{tf_{q_i,D} + \mu \frac{tf_{q_i,C}}{|C|}}{|D| + \mu} \tag{2}$$

Here C represent the whole collection and $|C|$ is its size. All the terms are stemmed using Porter stemmer, and stop words from PubMed are removed.

2.2 Concept-based IR

Concept identification We use UMLS Metathesaurus Release2012AB as our resource. A concept is defined as a "meaning"[1]. Each meaning is given a CUI

[1] http://www.ncbi.nlm.nih.gov/books/NBK9684/

(Concept Unified Identifier). The different synonyms and abbreviations of this concept is called a Term which is identified by LUI (Lexical Unified Identifer). Each of their lexical variant will be further subdivided into different String. SUI (String Unified Identifier) is their ID. For example, concept `C0004238` corresponds to the meaning `atrial fibrillation`. While `atrial fibrillation` and `auricular fibrillation` are two synonyms, they are identified by two different LUIs `L0004238` and `L0004237`. These two terms have both their singular and plural forms, with and without s. So in UMLS concept `C0004238` corresponds to 4 different SUIs representing its 4 different expression strings, called `SUIname` in Metathesaurus.

Concept (CUI)	Terms (LUIs)	Strings (SUIs)	Atoms (AUIs) * RRF Only
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	**L0004238** Atrial Fibrillation (preferred) Atrial Fibrillations	**S0016668** Atrial Fibrillation (preferred)	**A0027665** Atrial Fibrillation (from MSH) **A0027667** Atrial Fibrillation (from PSY)
		S0016669 (plural variant) Atrial Fibrillations	**A0027668** Atrial Fibrillations (from MSH)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	**S0016899** Auricular Fibrillation (preferred)	**A0027930** Auricular Fibrillation (from PSY)
		S0016900 (plural variant) Auricular Fibrillations	**A0027932** Auricular Fibrillations (from MSH)

Fig. 1. Concept, Term, String and Atom Identifiers [24]

MetaMap is a tool that identifies concepts from a text. Among other functionalities, MetaMap can identify the CUI corresponding to the concept string. It can also find all different string expressions (i.e. SUI names) for this concept. CUI and SUI names are the two different concept expressions that we used in our experiments. An example is shown in the figure below.

Original Expression	CUI	SUIname

`atrial fibrillation`	`C0004238`	`atrial fibrillation`
		`atrial fibrillations`
		`auricular fibrillation`
		`auricular fibrillations`

Fig. 2. Mapping original expression to CUI and SUIname

Retrieval on concept ID space We can view the whole set of concepts IDs as defining a concept space. Both document and query can then be represented as a set of CUI that MetaMap has recognized. The ranking score of a document can be determined by the matching score based on the concept IDs using the language model.

$$S(Q, D) = S(Q_{CUI}, D_{CUI}) = \frac{1}{n}\sum_{i=1}^{n} \log P(q_{CUI_i}|D_{CUI}) \tag{3}$$

It is possible that some of the concepts in documents and queries cannot be correctly identified by MetaMap. In this case, a more reasonable approach is to combine the concept-based retrieval with the traditional word-based retrieval. We implement it as follows:

$$S(Q|D) = \lambda S(Q_{orig}, D_{orig}) + (1 - \lambda)S(Q_{CUI}, D_{CUI}) \tag{4}$$

Reformulation with concept SUI name CUI is a very strict expression of concept. Another alternative expression of a concept is to enumerate all his `SUI`name in Metathesaurus. These SUInames are put into the `#syn()` operator in Indri[29], who treat all of the expressions listed as synonyms. We further test different operators `#1(), #uwN(), #uwN+1()` and `#combine()` with different flaxibility for each concept name, where `#1()` matches the term in parentheses as an exact phrase. `#uwN()` and `#uwN+1()` allows terms to appear in unordered window of size N and $N + 1$. `#combine()` just eliminate all dependence and group terms as "bag of words". This method is denoted by:

$$S(Q|D) = S(Q_{suiname}, D_{orig}) \tag{5}$$

Again, the above method can be combined with the word-based approach as follows:

$$S(Q|D) = \lambda S(Q_{orig}, D_{orig}) + (1 - \lambda)S(Q_{suiname}, D_{orig}) \tag{6}$$

Query expansion with mutual information Term co-occurrence analysis has been quite successful in traditional IR to determine related terms. Here, we try to determine related concepts using concept co-occurrences. Two concepts are considered to be related if they co-occur frequently. The relevance between two concepts x and y is measured by Point-wise Mutual Information (PMI):

$$pmi = \log \frac{p(x, y)}{p(x)p(y)} \tag{7}$$

We found that many of the determined concepts are indeed strongly related. For example, the related concepts to *Sepsis* are listed in Figure 3. We can see that they are usually related to the related drugs, diseases and treatments.

blood poison	bougardirey hemoglobin mali substance
injectable product solesta	abrasive point
brem hemoglobin mer sur	factor gamma interferon necrosis tumor
cilastatin dose mass	hazebrouck hemoglobin
glomerulosclerosis intercapillary	blanche grange hemoglobin
concord enterica entericon salmonella ser subsp	immunosuppressant macrolide
adrenergic nerve	hemoglobin henri mondor substance
aeromonadaceae family organism	dibromopropamidine product
injection mitomycin	hemoglobin maputo substance
murexide	abnormal blood find urea
blood entity fluidity	hemoglobin ibadan k
gene kdm4b	hemoglobin vaasa
cystic disease medullary uremic	gard hemoglobin ty
entire pelvis renal	phosphomannan
crotalarias	

Fig. 3. Top 30 Mutual Information concepts of *Sepsis*

In our experiment, the original query is expanded by the top mutual information concepts. In addition, the query is further expanded by the suiNames of the concepts.

$$S(Q,D) = \lambda_1 S(Q_{orig}, D_{orig}) + \lambda_2 S(Q_{suiname}, D_{orig}) + \lambda_3 S(Q_{mi}, D_{orig}) \quad (8)$$

with

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (9)$$

Markov Random Field Model In addition to taking into account synonyms, we also consider dependencies between words within a concept. Markov Random Field (MRF) model [22] can be used to account for dependencies between words. By default, one can assume that there is a dependency between two adjacent query words. Many experimental results showed that this model works better than the traditional bag-of-words method. When concepts are identified, it is possible that we only assume dependencies within a concept, and we believe that this could be a better approach than the default model. The MRF model contains three components. The first component is the traditional uni-gram language model. The second component is an ordered model, in which a concept is required to appear together and in order. This can be implemented in Indri as follows:

$$P(q_{orderedConcept}|D) = \frac{tf_{\#1(q_1,q_2,\ldots q_k),D} + \mu \frac{tf_{\#1(q_1,q_2,\ldots q_k),C}}{|C|}}{|D| + \mu} \quad (10)$$

where $tf_{\#1(q_1,q_2,\ldots q_k),D}$ is the frequency of an ordered concept in document, and k is the length of this concept.

The third component is an unordered model, in which the words within a concept can appear in any order within a text window.

$$P(q_{unorderedConcept}|D) = \frac{tf_{\#uwk+1(q_1,q_2,\ldots q_k),D} + \mu\frac{tf_{\#uwk+1(q_1,q_2,\ldots q_k),C}}{|C|}}{|D|+\mu} \quad (11)$$

where $tf_{\#uwk+1(q_1,q_2,\ldots q_k),D}$ is the frequency of the words in a window of size $k+1$[2].

Based on the above probabilities, we can define $S(q_{orderedConcept}, D)$ and $S(q_{unorderedConcept}, D)$. The final score is a combination of these three models,

$$S(Q,D) = \lambda_1 S(Q_{word}, D) + \lambda_2 S(q_{orderedConcept}, D) + \lambda_3 S(q_{unorderedConcept}, D) \quad (12)$$

where

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (13)$$

The model defined above is compared to the default MRF model, in which any two adjacent query words are assumed to be dependent (sequential dependence model).

3 EXPERIMENT

The data set for task 3 consists of a set of documents in the mdeical domain, provided by the Khresmoi project. Each document contains `#Uid,#date,#url` and `#content` fields. We convert the collection into TREC style. In the content part, we eliminate all commend, css and JavaScript part and all HTML tags. Only the remaining textual contents are indexed. Each query contains `<title>`, `<desc>`, `<discharge_summary>`. We use the short title queries.
The following 12 methods (runs) are tested:

1. baseline (Submitted as `GRIUM_EN_Run1`)
2. SUIname query, groupped by `#1()` oprator.
3. SUIname query expansion, groupped by `#1()` oprator.
4. SUIname query expansion, groupped by `#uwN()` oprator.
5. SUIname query expansion, groupped by `#uwN+1()` oprator.(Submitted as `GRIUM_EN_Run5`)
6. SUIname query expansion, groupped by `#combine()` oprator.
7. manual SUIname query expansion, groupped by `#combine()` oprator. Concepts are identified manually.
8. Pure CUI query retrieved in CUI document
9. CUI query expansion, document also contain `<original>` and `<cui>` two fields.(Submitted as `GRIUM_EN_Run7`)
10. Top mutual information and SUI name query expansion. (Submitted as `GRIUM_EN_Run6`)

[2] We only use k+1 as the window size in our experiments, although other sizes could also be used

11. Markov Random Field baseline with bigram and biterm.
12. Markov Random Field with concept dependence.

Only 4 of them (those with the run IDs) have been submitted.

4 RESULT

The experimental results are summarized in Fig. 4.[3]

			Result		
Submit	Run ID	Method	MAP	P@10	R-prec
Run1	Run 1	Baseline	0.3945	0.7180	0.4201
	Run a	#1(SUIname) query	0.2717	0.5680	0.3042
	Run b	#1(SUIname) query expansion	0.3916	0.6900	0.4217
	Run c	#uwN(SUIname) query expansion	0.4055	0.7500	0.4279
Run5	Run 5	#uwN+1(SUIname) query expansion	0.4069	0.7420	0.4283
	Run e	#combine(SUIname) query expansion	0.4112	0.7140	0.4286
	Run f	#combine(manual SUIname) query expansion	0.4185	0.7540	0.4306
	Run g	CUI query	0.2276	0.4920	0.2692
Run7	Run 7	CUI expansion	0.3495	0.6540	0.3862
Run6	Run 6	#uwN+1(SUIname) expansion + MutualInfo expansion	0.4007	0.7120	0.4156
	Run h	Markov random field baseline	0.3999	0.7320	0.4175
	Run i	Markov random field with concept dependence	0.3965	0.7260	0.4195

Fig. 4. Result of 12 runs evaluated by clef2014t3.qrels.test.binary.

First of all, we observe that the method using only strict concept space is less effective than the traditional word-based method. *Run g*, which use CUI query leads to a degradation of 42.3% compared to the baseline. If we simply compare the "bag-of-words" and "bag-of-concepts" methods, bag-of-words approach is certainly more flexible as a retrieve framework.
The result is far from what was expected. That means concept mapping procedure is still the bottleneck of the concept-based approach. Unfortunately, the mapping process is much more complicated than it seems. The definision of concept itself is not clear. An important hypothesis of "concept" is that "a meaning " should correspond only to one concept. But in fact, in UMLS a meaning can

[3]In order to keep the result comparable with other runs, we change the lambda of GRIUM_EN_Run5 from 5/6 to 1/10. The submitted result was 0.4016 for MAP, 0.7540 for P10.

be represented by a single accurate concept or be broken down into smaller concepts. For example, in query 36, for meaning *open pelvic fracture*, we can have 4 choices:

1. {Open fracture of pelvis}
2. {Fractures, Open} and {Pelvis}
3. {Open} and {Fracture of pelvis}
4. {Open} and {Fracture} and {Pelvis}

This is not simply an ambiguity, but also a granularity problem. None of them should be judged as definitly wrong, but their retrieval performance is different. In Fig.5, we show the concepts identified using different strategies:

Mapping strategy	Mapped concept expression	MAP(in Run e)
Original query	Convalescence after an open pelvic fracture and a right superior rami fracture	
MetaMap	[Convalescence] [Fractures, Open] [Pelvis] [Open] [Fracture of pelvis] [Right superior] [Branch of plant] [Fracture]	0.4958
Broad manual	[Convalescence] [Fractures, Open] [Pelvis] [Right superior] [Fracture of public rami]	0.3820
Middle manual	[Convalescence] [Open fracture of pelvis] [Right superior] [Fracture of public rami]	0.3445
Narrow manual	[Convalescence] [Open fracture of pelvis, multiple public rami - unstable]	0.3078

Fig. 5. Performance of different mapping strategies. Implemented by method e: #combine(SUIname) expansion.

the concepts identified by MetaMap, the broad concepts, narrow concepts and those in the middle level identified manually from Metathesaurus, as well as the corresponding MAP score. As we can see, the strategy that group many words into a very specific concept (Narrow manual) does not produce the best result. On the contrary, the other strategies that break long concepts into parts work significantly better. Still, the concepts that we recognize from a text have a large impact on the final retrieval result. This brings some new challenges for mapping task. [28] reported that MetaMap reached 84% in precision and 70% in recall. However, this evaluation is not done for the purpose of IR. For the 50 test queries, MetaMap identified 88 concepts. A rough evaluation indicates that only 66% of them, i.e.58 concepts seem reasonable for IR. We believe that even these concepts may not form the best way to do retrieval.

Knowing that mapping is not always acurate, some compromise solutions have to be used. Our tests show that at least two such strategies can help to reduce the impact of wrong mapping.
First, the most simple way is to also consider the original query. The concept

Run name	Method	MAP	Compared with baseline	Compared with Run g: CUI query
Run 1	Baseline	0.3945		
Run g	CUI query	0.2276	-42.3%	
Run a	#1(SUIname) query	0.2717	-31.1%	+19.4%
Run b	#1(SUIname) expansion	0.3916	-0.7%	+72.1%
Run e	#combine(SUIname) expansion	0.4112	+4.2%	+80.7%
Run f	#combine(manual SUIname) expansion	0.4185	+6.1%	+83.9%

Fig. 6. The benefits of query expansion strategie and retrieval flexibility

based synonyms are only treated as a complement to the original query. In our test, *Run b*, #1(SUIname) expansion brought an improvement of 57.2% over a pure #1(SUIname) query. At *Run c, 5, e, f*, the combination query brought an improvement.

Second, instead of strict CUI Id, we use SUIname as the expression of concept. As we can see in the result, *Run a* produced 19.4% less mistake than *Run g*. In addition, taking into account the fact that concepts IDs can share many words.Using SUIname can further help us retrieving documents on related concepts. That is why, with *#combine()* operator, *Run e* achieved the best performance over all 11 automatic runs. Our two MRF runs (*Run h* and *Run i*) showed in another way that naive concept-based dependence does not bring any improvement.

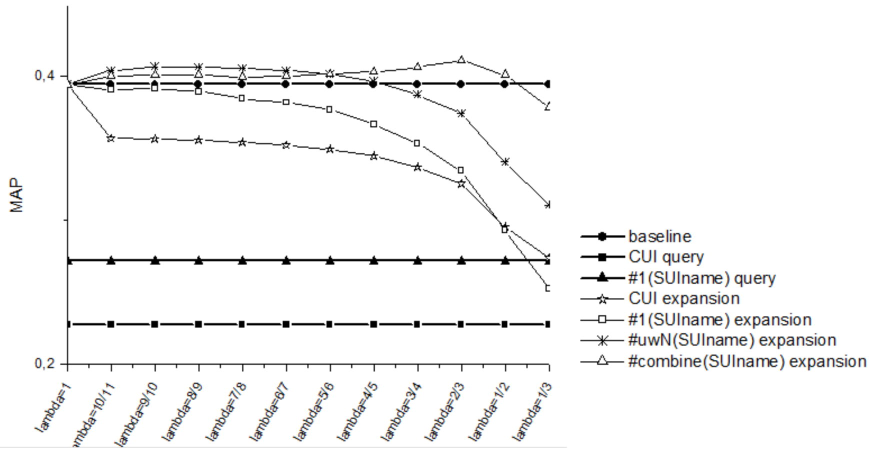

Fig. 7. Query expansion Vs. Pure CUI, SUIname query and baseline at MAP. The three straight lines represent respectively pure CUI, SUIname query and baseline. The four curves show their different combination result.

Fig.7 shows the impact of using different values for lambda.
At last, our naive mutual information expansion did not bring any additional information as expected.

5 CONCLUSION

This year in task 3, we tested several different ways of integrating concept knowledge. Our results showed that the "bag-of-concepts" is less effective than "bag-of-words" approach. We further discuss about two effecive ways of reducing the impact of incorrect concept mapping. Original query is indispensable, and SUIname is a more flexiable way of using a concept. The mapping performance is still the bottleneck of the concept-based approach. This is a question that we will examine in our future research.

References

1. Liadh Kelly, Lorraine Goeuriot, Hanna Suominen, Tobias Schrek, Gondy Leroy, Danielle L. Mowery, Sumithra Velupillai, Wendy W. Chapman, David Martinez, Guido Zuccon and Joao Palottim. Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. Proceedings of CLEF 2014. Lecture Notes in Computer Science (LNCS). Springer. (2014)
2. Lorraine Goeuriot, Liadh Kelly, Wei Li, Joao Palotti, Pavel Pecina, Guido Zuccon, Allan Hanbury, Gareth Jones and Henning Mueller. ShARe/CLEF eHealth Evaluation Lab 2014, Task 3: User-centred health information retrieval. Proceedings of CLEF 2014. (2014)
3. Hersh, William R.; David D. Hickam; and T. J. Leone. Words, concepts, or both: Optimal indexing units for automated information retrieval. Mark E. Frisse (ed.) Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care, 644-648 (1992)
4. Srinivasan P. Query expansion and MEDLINE. Information Processing and Management, 32(4): 431-443 (1996)
5. Aronson, A. R., & Rindflesch, T. C. Query expansion using the UMLS Metathesaurus. In Proceedings of the AMIA Annual Fall Symposium. American Medical Informatics Association, p. 485 (1997)
6. BOUDIN, Florian, NIE, Jian-Yun, et DAWES, Martin. Clinical information retrieval using document and PICO structure. In : Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, p. 822-830 (2010)
7. ZHOU, Wei, YU, Clement, SMALHEISER, Neil, et al. Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature. In : Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM. p. 655-662. (2007)
8. Zhu, Dongqing, et al. "Using discharge summaries to improve information retrieval in clinical domain." Proceedings of the ShARe/-CLEF eHealth Evaluation Lab (2013)

9. ZUCCON, G., KOOPMAN, B., et NGUYEN, A. Retrieval of health advice on the web: AEHRC at ShARe/CLEF eHealth evaluation lab task 3. In : Proceedings of CLEF Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis. (2013)
10. Choi, Sungbin, and Jinwook Choi. "SNUMedinfo at CLEFeHealth2013 task 3." Proceedings of the ShARe/CLEF eHealth Evaluation Lab (2013)
11. Bedrick, Steven, and G. Sheikhshabbafghi. "Lucene, metamap, and language modeling: OHSU at CLEF eHealth 2013 ." Proceedings of the ShARe/CLEF eHealth Evaluation Lab (2013)
12. CALLEJAS, P., MIGUEL, A., WANG, Yue, et al. Exploiting Domain Thesaurus for Medical Record Retrieval. DELAWARE UNIV NEWARK, (2012)
13. OZTURKMENOGLU, Okan et ALPKOCAK, Adil. DEMIR at TREC Medical: Power of Term Phrases in Medical Text Retrieval. In : TREC. (2011)
14. QI, Yanjun et LAQUERRE, Pierre-Franois. Retrieving Medical Records with sennamed: NEC Labs America at TREC 2012 Medical Records Track.(2012)
15. KOOPMAN, Bevan, BRUZA, Peter, SITBON, Laurianne, et al. AEHRC & QUT at TREC 2011 Medical Track: a concept-based information retrieval approach. In : Proceedings of 20th Text REtrieval Conference (TREC 2011). National Institute of Standards and Technology (NIST), p. 1-7 (2011)
16. KOOPMAN, Bevan, ZUCCON, Guido, NGUYEN, Anthony, et al. Exploiting SNOMED CT concepts and relationships for clinical information retrieval: Australian e-Health Research Centre and Queensland University of Technology at the TREC 2012 Medical Track. (2012)
17. KING, Benjamin, WANG, Lijun, PROVALOV, Ivan, et al. Cengage Learning at TREC 2011 Medical Track. In : TREC. (2011)
18. FUJITA, Sumio. Revisiting Again Document Length Hypotheses TREC 2004 Genomics Track Experiments at Patolis. In : TREC. (2004)
19. DARWISH, Kareem et MADKOUR, Amgad. The GUC Goes to TREC 2004: Using Whole or Partial Documents for Retrieval and Classification in the Genomics Track. In : TREC. (2004)
20. KRAAIJ, Wessel, RAAIJMAKERS, Stephan, WEEBER, Marc, et al. MeSH Based Feedback, Concept Recognition and Stacked Classification for Curation Tasks. In : TREC. (2004)
21. LI, Jiao, ZHANG, Xian, ZHANG, Min, et al. THUIR at TREC 2004: Genomics Track. In : TREC. (2004)
22. ARONSON, Alan R. et RINDFLESCH, Thomas C. Query expansion using the UMLS Metathesaurus. In : Proceedings of the AMIA Annual Fall Symposium. American Medical Informatics Association, p. 485. (1997)
23. ChengXiang Zhai: Statistical Language Models for Information Retrieval. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2008)
24. BODENREIDER, Olivier. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, vol. 32, no suppl 1, p. D267-D270. (2004)
25. LIPSCOMB, Carolyn E. Medical subject headings (MeSH). Bulletin of the Medical Library Association, vol. 88, no 3, p. 265 (2000)
26. SPACKMAN, Kent A., CAMPBELL, Keith E., C, R. A., et al. SNOMED RT: a reference terminology for health care. In : Proceedings of the AMIA annual fall symposium. American Medical Informatics Association. p. 640. (1997)
27. ARONSON, Alan R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In : Proceedings of the AMIA Symposium. American Medical Informatics Association, p. 17. (2001)

28. PRATT, Wanda et YETISGEN-YILDIZ, Meliha. A study of biomedical concept identification: MetaMap vs. people. In : AMIA Annual Symposium Proceedings. American Medical Informatics Association. p. 529. (2003)
29. STROHMAN, Trevor, METZLER, Donald, TURTLE, Howard, et al. Indri: A language model-based search engine for complex queries. In : Proceedings of the International Conference on Intelligent Analysis. p. 2.6. (2005)
30. MILLER, George A. WordNet: a lexical database for English. Communications of the ACM, vol. 38, no 11, p. 39-41. (1995)