Medical document indexing and retrieval: AMTEx vs. NLM MMTx

Medical Document Indexing and Retrieval: AMTEx vs. NLM MMTx

Angelos Hliaoutakis, Kalliopi Zervanou, Euripides G.M. Petrakis

Dept. of Electronic and Comp. Engineering, Technical University of Crete (TUC), Chania, Greece,
angelos@softnet.tuc.gr, kelly@intelligence.tuc.gr, petrakis@intelligence.tuc.gr

AMTEx **is a medical document indexing method, specifically designed for the automatic indexing of documents in large medical collections, such as MEDLINE, the premier bibliographic database of the U.S. National Library of Medicine (NLM).** ***AMTEx*** **combines MeSH, the terminological thesaurus resource of NLM, with a well-established method for term extraction, the C/NC-value method. The performance evaluation of two** ***AMTEx*** **configurations is measured against the current state-of-the-art, the MMTx method in indexing and retrieval tasks in three experiments. In the first, a subset of MEDLINE (PMC) full document corpus was used for the indexing task. In the second and third, a subset of MEDLINE (OHSUMED) abstracts was used for indexing and retrieval respectively. The experimental results demonstrate that AMTEx achieves better precision in all tasks, in 50-20% of the processing time compared to MMTx.**

Keywords

document indexing, medical document retrieval, term extraction

1. Introduction

The availability of large medical online collections, such as MEDLINE [1], poses new challenges to information and knowledge management. MEDLINE documents are currently indexed by human experts, based on the MeSH thesaurus [2]. The automatic mapping of biomedical documents to UMLS [3] term concepts has been undertaken by MMTx [4]. MMTx was originally developed to improve retrieval of bibliographic material [5]. The limitations of MMTx in term extraction and in the UMLS Metathesaurus mapping have been analysed in detail in [6] and [7]. Our experiments in a pilot study of MMTx and AMTEx on a small MEDLINE corpus showed that MMTx performance was low in precision and that its output greatly suffers by over-generating terms, which diffuse the document concept leading to inaccurate indexing of MEDLINE documents [8]. This reflects a design choice in MMTx, which attempts to favour recall by not focusing on MeSH, whereupon MEDLINE indexing has been based, and by incorporating a variant generation process which leads to term over-generation.

In this paper, we briefly review the MMTx approach and we present our alternative method, the Automatic MeSH Term Extraction method (*AMTEx*). *AMTEx* aims at improving the efficiency of automatic term extraction, using a hybrid linguistic/statistical term extraction method, the C/NC value method [9]. Additionally, *AMTEx* aims at improving efficiency and accuracy in indexing of MEDLINE documents, based on the extraction and mapping of document terms to the MeSH Thesaurus, rather than the full UMLS Metathesaurus mapping of MMTx.

2. Term Extraction

Term Extraction aims at the identification of linguistic expressions denoting specialised concepts, namely domain or scientific terms. The automatic identification of terms is of particular importance because these linguistic expressions convey the informational content of a document. Term extraction approaches largely rely on the identification of term formation patterns (e.g. [10], [11], [12]). Statistical techniques may also be applied to measure the degree of unithood or termhood of the candidate multi-word terms (e.g. [13]). Later and current approaches tend to follow a hybrid approach combining both statistical and linguistic techniques (e.g. [9], [14], [15]).

The extraction of terms for the medical, biological and biomedical domain has greatly motivated research for both indexing, as well as knowledge extraction purposes [12], [16], [17], [18]. In the specific context of term extraction for indexing purposes, the main objective of the term extraction process is the identification of discrete content indicators, namely *index terms*. A traditional technique for automatic indexing has been the *tf · idf* method [19]. In traditional indexing techniques, query and document representations ignore multi-word and compound terms, which may perform quite efficiently, split into isolated single-word index terms. However, multi-word terms are very common in the biomedical domain [14] and are often used in indexing medical documents. Multi-word terms carry important classificatory content information, since they comprise of modifiers denoting a specialisation of the more general single-word, head term [11]. Currently machine learning techniques are also applied for indexing, such as the Naïve Bayes learning model implemented in the KEA (Automatic Keyphrase Extraction, [20]).

3. Background

3.1 The MMTx Approach and Resources

The MMTx approach uses the UMLS Metathesaurus® and SPECIALIST™ lexicon as its lexicographic resources. The *Unified Medical Language System (UMLS)* is a source of medical knowledge developed and maintained by the U.S. NLM. UMLS consists of the Metathesaurus, the Semantic Network and the SPECIALIST lexicon. The *Metathesaurus* is a large, multi-purpose and multi-lingual vocabulary database. The Metathesaurus on its own does not have a hierarchical structure, neither fulfils ontological requirements. The *Semantic Network* provides a categorisation of all concepts represented in the Metathesaurus and a set of useful relationships among these concepts. Finally, the *SPECIALIST lexicon* is intended to be a general English lexicon which includes many medical and biomedical terms. MMTx uses the Metathesaurus and SPECIALIST lexicon resources during the term extraction process. This process maps arbitrary text to Metathesaurus terms and works in the following steps [5]:

Parsing: The document text is parsed and a simple linguistic filter isolates noun phrases [22].
Variant Generation: Variant generation is performed in iterative manner [23]. First, the multi-word term phrase is split into *generators*. A *variant generator* is any meaningful subsequence of words in the phrase. In the second phase, for each of the generators, all possible semantic and derivational variants are identified using the SPECIALIST lexicon and a supplementary database of synonyms. At this stage please note that, although we have started the process of variant generation of a noun phrase, we may have derivational and semantic variants belonging to other parts-of-speech, such as verbs. All these variants are in turn used as

generators and their respective variants are recomputed. Finally, inflectional and spelling variants are generated based on all word-forms found in the previous processes.

Candidate Retrieval: The main criterion for candidate retrieval is that the mapped term string contains at least one of the variants found during the variant generation process [24]. The mapping is not always an exact match [5].

Candidate Evaluation: The candidate terms are evaluated by computing their *mapping strength* based on linguistic criteria [25].

3.2 The AMTEx Method Resources

The C/NC-Value Method for Term Extraction

The C/NC value method [9] is a domain-independent method for term extraction. It combines statistical and linguistic filtering information to extract multi-word and nested terms.

The statistical part defining the termhood of the candidate phrases aims at improved term detection, compared to the mere frequency of occurrence method, being especially designed for the detection of terms appearing as nested within longer terms, such as the term *enzyme inhibitors* nested in *Angiotensin-converting enzyme inhibitors*. The measurement used for this estimation is C-value [9]. The C-value algorithm produces a list of terms ranked by decreasing term likelihood value. The NC-value measure takes into account the context of each term for its final weighting. It assigns weights to specific grammatical categories that tend to appear in term context.

C/NC-value has been successfully tested in various domains, such as molecular biology [26], eye pathology medical records [9] and biomedical business newswire texts [18]. For the purposes of AMTEx, any efficient term extraction method would do. However, comparative experiments of *tf · idf*, KEA and the C/NC value term extraction methods by Zhang et al. [21] show that C/NC value significantly outperforms both *tf · idf* and KEA in a narrative text classification task using the extracted terms.

The MeSH Thesaurus

The MeSH Thesaurus (Medical Subject Headings) is a taxonomy of medical and biological terms and concepts suggested by the US NLM. The MeSH terms are organized in fifteen IS-A taxonomies. Each MeSH term is described by several properties, the most important being:

- *MeSH Heading (MH):* the term name or identifier;
- *Scope Note:* a text description of the term;
- *Entry Terms:* mostly synonym terms to the MH.

A fragment of the MeSH IS-A hierarchy is illustrated in Fig.1.

4. The AMTEx method

Based on the study of the MMTx algorithm and resources, we observe the following:

- During the variant generation stage, the iterative expansion of the initial text phrase to all possible variants is quite exhaustive. MMTx extracts term variants, not only based on the terms found in the original text phrase, but also from their variant terms. However, this process also results in term over-generation and increased term ambiguity, which diffuse the original term concept, leading to inaccurate indexing.
- MMTx extracts general Metathesaurus terms, not MeSH terms.

- Term selection is based on a scoring function, for evaluating the importance of all candidate terms, using the SPECIALIST lexicon as an external lexical resource. This function, though partly based on valid linguistic principles, it is arbitrarily and empirically defined, making it possible for unrelated terms to be included in the list of extracted terms. The C/NC-value scoring functions are especially tuned to multi-word terms, taking into consideration nested terms and term context words. Additionally, C/NC-value has been proven to extract up to 98% of correct terms [26], [9], [18].

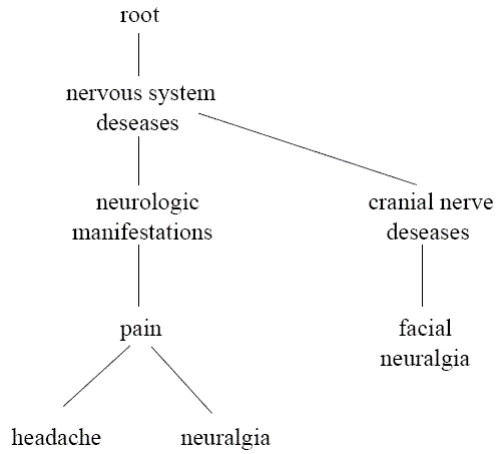

Figure 1 A fragment of the MeSH IS-A hierarchy.

Based on the above observations we propose two basic changes towards the development of an improved term extraction method that could substitute MMTx:

1. Term extraction based on a well-established method, the C/NC-value method;
2. Use of MeSH Thesaurus as lexical resource, both for (limited) term variant retrieval, and candidate term mapping.

Table 1 AMTEx Algorithm.

Input: Document *d*, MeSH taxonomy.
Output: MeSH terms *t*.
1. *Multi-word Term Extraction:* C/NC-value method
2. *Term Ranking:* NC-value ranking
3. *Term Mapping:* Only MeSH terms are retained.
4. *Single-word Term Extraction:* Single-word MeSH terms are added.
5. *Term Variants:* Stemmed terms are added.
6. *Term expansion:* Semantically similar terms from MeSH

An outline of the *AMTEx* procedure is illustrated in Table 1. In particular, the *AMTEx* method has the following processing stages:

1. *Multi-word Term Extraction:* The C/NC-value method is used for term extraction. During term extraction in *AMTEx* the document text is parsed, using the C/NC-value part-of-speech tagger and linguistic filters.
2. *Term Ranking:* Extracted candidate terms are evaluated, first by C-value and subsequently by NC-value score. The final candidate term list is ranked by decreasing

term likelihood. Top ranked terms are more important than terms ranked lower in the list and are more likely to be included in the final list of extracted terms.

3. *Term Mapping:* Candidate terms are mapped to terms of the MeSH Thesaurus, by complete, exact string matching. The list of terms now contains only MeSH terms.
4. *Single-word Term Extraction:* For the multi-word terms which do not fully match MeSH, their single word constituents are used for matching. If mapped to a single word MeSH term, the mapped term is added to the term list.
5. *Term Variants:* Term variants are included in the candidate term list. The C/NC-value implementation in *AMTEx* includes inflectional variants of the extracted terms. Also, MeSH itself is used for locating variant terms, based on the MeSH term, Entry Terms property. However, only the stemmed term-forms are used in *AMTEx,* since the full list of Entry Terms may contain terms, which often are not synonymous.
6. *Term Expansion:* The list of terms is augmented with semantically similar terms from MeSH. Fig. 2 illustrates this process: a term is represented by its MeSH tree hierarchy (hypernyms/hyponyms). The neighbourhood of the term is examined and all terms with similarity greater than threshold $T_{Expansion}$ are also included in the query vector. This expansion may include terms more than one level higher or lower than the original term, depending on the value of $T_{Expansion}$.

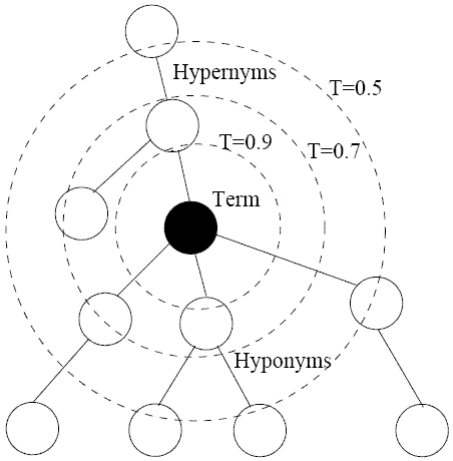

Figure 2 Term expansion thresholds using MeSH.

Our approach to *Term Variant* generation, it is more limited than MMTx. This constrains our term recall to terms that are closer to the original term in text. As we observe in the results of our experiments in section 5, we manage to achieve better precision in a fraction of the processing time taken for MMTx. We believe that this is partly due to the fact that our term extraction method outperforms MMTx in suggesting candidate terms. It is also due to the fact the *AMTEx* approach to variant generation is limited to MeSH and does not operate iteratively, generating variants out of already found variants, thus avoiding the diffusion of the original concept to unrelated concepts.

Term Expansion, the method used in *AMTEx* for discovering semantically similar terms, is based on the semantic similarity method by Li et al. [27]. The evaluation of the semantic similarity methods indicated that this method is particularly effective, achieving up to 73% correlation with results obtained by humans [28]. Because no synonymy relation is defined in MeSH, we did not apply expansion to the Entry Terms of terms.

4.2 Refining the AMTEx Method

In order to determine the optimal set of indexing terms, namely one increasing recall and precision, there exist three thresholds in the *AMTEx* process that could be refined:

- *C-Value threshold* (T_{Cvalue}) for the term extraction, which in our initial experiments presented in [8] was set to its recommended value ($T_{Cvalue} = 1.5$) to limit output to the most valid terms;
- *Term expansion threshold* ($T_{Expansion}$), whereupon we have experimented in our pilot small scale experiments with *AMTEx* [8] ;
- *Final list threshold* ($T_{FinalList}$), which determines the minimum value a term mapped to MeSH must have to be included in the final index term list. In our experiments presented in [8], all candidate terms were retained.

The optimal value for each of these thresholds is not easy to determine, as each of these affects term recall at different stages of the *AMTEx* process. In our pilot experiments in [8], an increase of the $T_{Expansion}$ improved precision but largely affected recall at that stage of *AMTEx* processing. Thus, independent alterations of one threshold are bound to affect the other two.

A simple approach to this optimisation problem would be to consider only the threshold applied at the end of the process, the $T_{FinalList}$. Moreover, precision or recall alone should not determine an optimal threshold, since an increase in precision for example, simultaneously affects recall. A balanced measure, such as an F-measure, where recall and precision are equally weighted (shown on Equ. 1 below), would provide us a better indicator for our final threshold.

$$F = \frac{2 * precision * recall}{precision + recall} \quad (1)$$

Thus, in our *AMTEx v2*, we have chosen to be exhaustive with both T_{Cvalue} (i.e. $T_{Cvalue}=0$) and $T_{Expansion}$ (i.e. $T_{Expansion}=0.5$) thresholds and use the maximum F-measure to determine the $T_{FinalList}$. Moreover, in the *Term Expansion* step, the semantically similar terms ($T_{Expansion}=0.5$) added to the candidate list are assigned a weight, as shown on Equ. 2 below:

$$weight(w) = sim * weight(s) \quad (2)$$

where a term w, semantically similar to term s, has ranking weight, *weight(w)*, combining its semantically similar term weight, *weight(s)*, and the similarity value, *sim*, by which w is similar to s. In this way, in *AMTEx v2* the final candidate list ranks accordingly terms which are added to it by the *Term Expansion* process. In *AMTEx v1*, these terms were merely assigned the *weight(s)*.

In our pilot experiments with *AMTEx v1* [8], in the *Single-word Term Extraction* step, we were attempting to find partial matches in MeSH, for all word constituents of an unmatched multi-word term. We have observed that single term insertion in our candidate list through that process produced worse results. In our *AMTEx v2*, we have chosen to conceptually limit our search for single-word mappings using only the *head* word of the multi-word term. The experiments presented in section 5.1 of this paper show that this type of *Single-word Term Extraction* improves both recall and precision. Regarding ranking weight for these terms, we consider it equal to its source, i.e. the original multi-word term weight.

5. Experiments and evaluation

5.1 Developing AMTEx v2

Defining $T_{FinalList}$ threshold for AMTEX v2

In order to determine the $T_{FinalList}$, we have experimented with a corpus of 5,819 full PMC documents selected out of 60 Journals. The documents were selected on the basis of having an UID number, which we used to retrieve their respective MEDLINE index sets. This index set for each document is manually assigned by MEDLINE experts and is used in our experiment as our ground truth. Thus, in our evaluation, *precision* is the total number of correctly extracted terms, compared to the MeSH terms appearing in the respective document index. Similarly, *recall*, in our evaluation, is the total number of correctly retrieved terms, compared to the total number of terms in the MeSH index gold standard. In this experiment F-measure of equally weighted precision and recall is used.

The results of this experiment are shown in Fig. 3. The peak of the curve in Fig. 3 indicates the optimal F-measure performance for our corpus, showing the 22nd term to reach the maximum F-Measure value. Thus, in *AMTEx v2*, the $T_{FinalList}$ is set to the 20 top terms in the list.

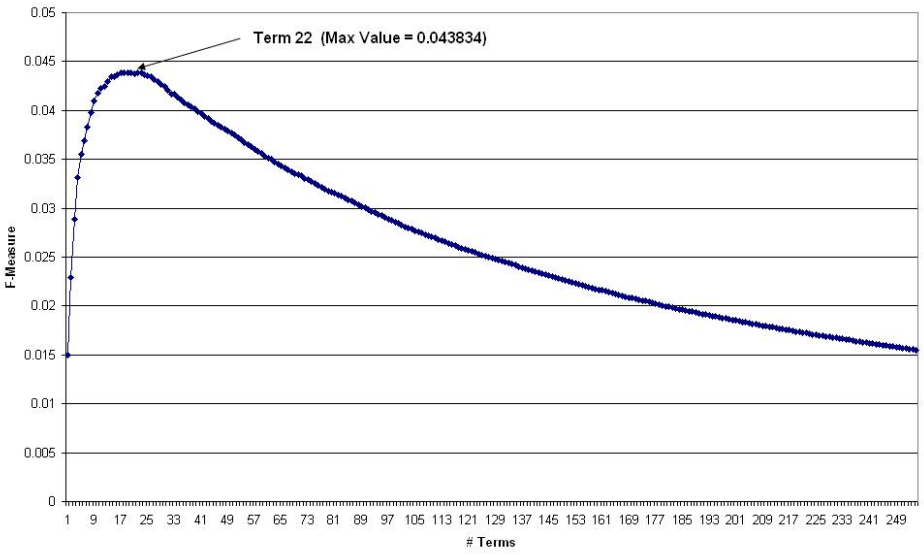

Figure 3 Single-word step performance and $T_{FinalList}$ threshold in PMC dataset.

Refining Single-word Term Extraction

In AMTEx v2, as discussed in section 4.2, we attempted to modify the Single-word Term Extraction process, using only the *head* term constituent for MeSH mapping. To determine the results we conducted two experiments, using the same 5,819 full PMC documents and the same evaluation process as in the $T_{FinalList}$ experiment.

In the first experiment, we indexed our corpus including the modified version of the Single-word Term Extraction process. The results, as shown in Fig. 3 were satisfactory. Nevertheless, we needed to ascertain that the single-word term extraction step significantly contributes to *AMTEx* performance, rather than unnecessarily complicating the *AMTEx* algorithm. Thus, we conducted a second experiment on the same dataset, where the *single-*

word term extraction step was not included in the process. The comparative results in Fig. 4 show clearly that *single-word term extraction* improves *AMTEx* performance.

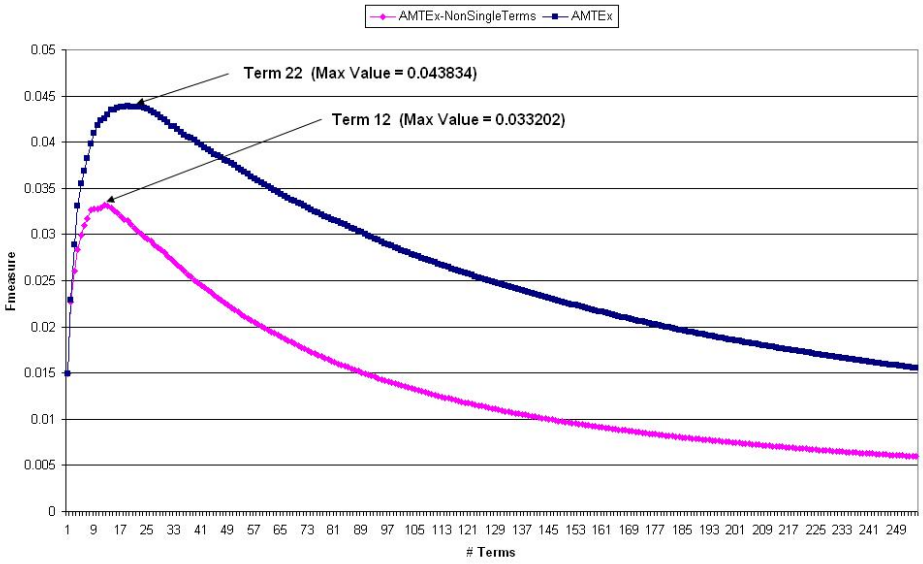

Figure 4 *AMTEx* with/without single-word extraction and $T_{FinalList}$ threshold in PMC dataset.

5.2 MMTx vs AMTEx Method

In our pilot experiments presented in [8], we have compared our first *AMTEx* version performance to MMTx, which is considered the benchmark method, using a small set of 61 full documents. In this paper, we present a series of comparative experiments we conducted to test our approach in:

- a significantly larger corpus of full documents,
- a corpus of document abstracts,
- using both versions of AMTEx, v1 and v2,
- for indexing and retrieval tasks,
- against MMTx, v24B.

For this reason, we conducted three experiments, comparing AMTEx v1 and v2 to MMTx v2.4B: In the first, the *Full Doc Indexing experiment*, we compare AMTEx vs MMTx for the indexing task in a large data set of full documents. In the second, the *Abstract Indexing experiment*, we compare performances on indexing document abstracts, rather than full documents. Finally, in the *Retrieval experiment*, we compare performances on the retrieval task.

We should note that in MMTx term ranking is less rigorous than AMTEx. In MMTx valid term output has mostly a weight value of 1000, whereas in AMTEx each term is ranked based on its individual weight. Thus, the evaluation score value of the 10th or 100th best answer of MMTx is not particularly adequate, since all its results may be equally weighted.

Please also note that for our indexing experiments, we have thought it fair for MMTx to restrict its term mapping process to MeSH, rather the full UMLS, similarly to our AMTEx, since our ground truth consists for the MEDLINE provided index sets, which are based on MeSH.

Full Doc Indexing experiment

For this experiment, we have again used the 5,819 PMC full document corpus. Our two versions of AMTEx and the MMTx v2.4B were used for document indexing. The results were evaluated for precision and recall, against our ground truth, i.e. the MEDLINE document index set.

The results in Table 3 show average term output, precision and recall for each document, for all three systems. We observe that *AMTEx v1*, shows a precision result that is higher than MMTx, whereas the average extracted terms are much less. *AMTEx v2* demonstrates the best recall of the two AMTEx systems, for a fraction of the average MMTx term output.

Table 3 *AMTEx* vs. MMTx performance on the PMC data set.

PMC Dataset	AMTEx v1.0	AMTEx v2.0	MMTx 2.4B
Average Terms	16	25	72
Precision	0.052	0.034	0.033
Recall	0.054	0.062	0.162

Abstract Indexing experiment

The second experiment was conducted to test the performance of the three systems in a document abstracts corpus. The problems related to processing document abstracts were first identified in our pilot experiments with *AMTEx* [8]. These relate to the abstract size, which is quite limited to be used as input to a method using statistics, such as *AMTEx*. Moreover, the content of the abstract has not been found to contain all necessary textual information for accurately indexing the full document. We have concluded at the time that we needed to consolidate our *AMTEx* approach before embarking into such an experiment.

For the *Abstract Indexing experiment* presented in this paper, we selected a corpus subset of the OHSUMED standard TREC collection corpus [29]. OHSUMED is a collection of MEDLINE document abstracts used for benchmarking information retrieval systems evaluation. Our selected subset consisted of 10% of OHSUMED, i.e. 30,000 document abstracts. These were again evaluated in terms of precision and recall against the MEDLINE provided MeSH index term sets.

For processing of document abstracts, *AMTEx* algorithm was slightly modified to respond to the problems of document limited size and content that we have identified. Thus, both *AMTEx* versions first treat the totality of the corpus as a single document input during the term extraction step. Subsequently the extracted terms are associated to their respective source document by string matching. This modification of the *AMTEx* process has been thought necessary, since *AMTEx* term extraction is not only linguistic but also statistically based.

Table 4 demonstrates the comparative performance of AMTEx v1 and v2 against MMTx v2.4B in terms of average document precision and recall. We observe again the *AMTEx* improved precision compared to MMTx, and a reasonable recall by merely a fifth of the average term output compared to MMTx.

Table 4 *AMTEx* vs. MMTx performance on the OHSUMED data set.

OHSUMED Dataset	AMTEx v1.0	AMTEx v2.0	MMTx 2.4B
Average Terms	8	8	40
Precision	0.124	0.125	0.089
Recall	0.101	0.101	0.336

Finally, table 5 illustrates the comparative results of all systems, in both full PMC and OHSUMED abstract indexing experiments in terms of time efficiency. We observe that the time taken for OHSUMED processing was longer in all systems. Nevertheless, both AMTEx systems are shown to perform much faster than MMTx. We believe that this is due to the algorithmic simplicity of *AMTEx* compared to MMTx especially with regards to variant generation and term expansion processes (even though MMTx was tested using MeSH rather than the full UMLS).

Table 5 Time intervals of *AMTEx* and MMTx for PMC & OHSUMED data set.

Time Intervals	AMTEx v1.0	AMTEx v2.0	MMTx 2.4B
PMC Dataset	1721.4	4994.6	9819.5
OHSUMED Dataset	9161.9	26582.5	52261.8

Retrieval experiment

In our last experiment we attempted to test *AMTEx* performance in the medical document retrieval task. In this experiment our AMTEx versions were again compared to MMTx, which was considered the benchmarking method for this task. We have used the OHSUMED standanrd TREC collection corpus [29] subset used in our indexing experiment. However, for this task the results were evaluated against the TREC provided queries and answers. These constituted our ground truth for all systems performance.

Fig. 5 Illustrates the performance of AMTEx v1 and v2 compared to MMTx and the MEDLINE provided index term sets, i.e. the terms used as ground truth in our indexing experiments. In Fig. 5, each method is represented by a precision/recall curve. For each query, the best 100 answers were retrieved, i.e. the precision/recall plot of each method contains exactly 100 points. Precision and recall values are computed from each answer set and therefore, each plot contains exactly 100 points. The top-left point of the precision/recall curve corresponds to the precision/recall values for the best answer or best match (which has rank 1), while the bottom right point corresponds to the precision/recall values for the entire answer set. Document matching is performed by Vector Space Model (VSM, [30]).

In this task, we observe that the increased term recall of MMTx results in significantly better retrieval performance than *AMTEx*, nearing the performance of the manually assigned MeSH index terms for large answer sets. For small answer sets, MMTx is shown to perform better

than the MeSH representation. The reason for the poor performance of *AMTEx* lies mostly in the way VSM works: For a query and a document to be similar, the terms of the query vector may be a subset of the terms of the document vector. Thus, VSM clearly favours representation with many terms (such as MMTx representations). Based on all three experiments we conclude that MMTx increased term recall is well suited for retrieval, whereas *AMTEx* less noisy output serves indexing best.

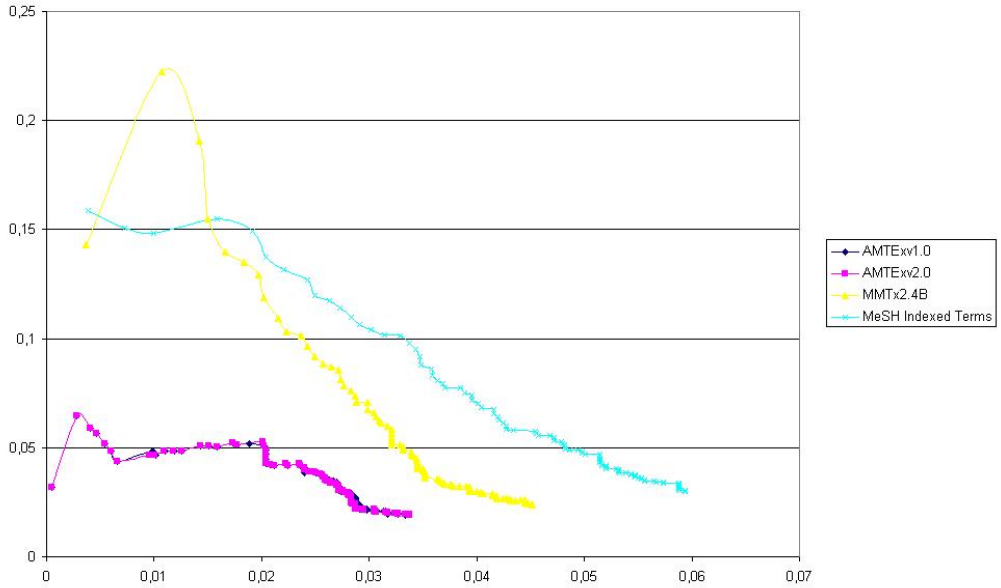

Figure 5 Interpolated Precision/Recall of AMTEx vs. MMTx on OHSUMED dataset retrieval task.

6. Conclusions

This paper discusses about the document mapping process to the correct MeSH index terms automatically. We presented the term extraction problem for the automatic indexing of documents in large medical collections, such as the MEDLINE collection. We have briefly presented related approaches to this problem, focusing on the MMTx method, which attempts to map terms in medical documents to UMLS Metathesaurus concepts.

We have developed an alternative method, the *AMTEX* method, which is specifically designed for the indexing of MEDLINE documents, using the MeSH Thesaurus resource and a well-established method for extraction of domain terms, the C/NC-value method. We presented the experiments we conducted for the refinement of the *AMTEx* method. *AMTEx* has been also compared to MMTx in the indexing and the retrieval tasks.

References

1. MEDLINE: Medical Literature Analysis and Retrieval System Online. [cited 2007 March]; http://www.nlm.nih.gov/databases/databases_medline.html.
2. MeSH: The Medical Subject Headings (MeSH) thesaurus. [cited 2007 March]; http://www.nlm.nih.gov/mesh.
3. UMLS Metathesaurus: The Unified Medical Language System. [cited 2007 March]; http://www.nlm.nih.gov/research/umls.
4. MMTx: MetaMap Transfer tool. [cited 2007 March]; http://mmtx.nlm.nih.gov.

5. Aronson A R. Effective Mapping of Biomedical Text to the UMLS® Metathesaurus®: The MetaMap Program. In: Proc American Medical Informatics Association Symposium; 2001. p. 17-21.
6. Pratt W, Yetisgen-Yildiz M. A Study of Biomedical Concept Identification: MetaMap vs. People. In: Proc American Medical Informatics Association Symposium; 2003 Nov; Washington DC, USA. p. 529-33.
7. Divita G, Tse T, Roth L. Failure Analysis of MetaMap Transfer (MMTx). In: Fieschi M, Coiera E, Li Y C J, editors. MEDINFO 04; 2004 Aug; Amsterdam: IOS Press; 2004. p. 763-7.
8. Hliaoutakis A, Zervanou K, Petrakis E G M, Milios E. Automatic Document Indexing in Large Medical Collections. In Proc ACM International Workshop on Health Information and Knowledge Management; 2006 Nov 11; Arlington, VA, USA. p.1-8.
9. Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms: The C-Value/NC-value Method. Int J Digital Libraries 2000; 3(2):117-32.
10. Ananiadou S. A Methodology for Automatic Term Recognition. In: COLING-94; 1994 Aug 5-9; Kyoto, Japan. p. 1034-8.
11. Bourigault D, Gonzalez-Mullier I, Gros C. LEXTER, a Natural Language Tool for Terminology Extraction. In: Gellerstam M, Järborg J, Malmgren S G, Norén K, Rogström L, Papmehl C R, editors. Seventh EURALEX International Congress on Lexicography; 1996 Aug 13-18; Göteborg, Sweden. p. 771-9.
12. Gaizauskas R, Demetriou G, Humphreys K. Term Recognition in Biological Science Journal Articles. In: Ananiadou S, Maynard D, editors. Proc NLP 2000 Workshop on Computational Terminology for Medical and Biological Applications; June 2000; Patras, Greece. p. 37-44.
13. Daille B, Gaussier E, Lange J. Towards Automatic Extraction of Monolingual and Bilingual Terminology. In: COLING-94; 1994 Aug 5-9; Kyoto, Japan. p. 515-21.
14. Maynard D, Ananiadou S. TRUCKS: A Model for Automatic Multi-Word Term Recognition. J Natural Language Processing 2000; 8(1):101-5.
15. Jacquemin C. Spotting and Discovering Terms through Natural Language Processing. Cambridge: MIT Press; 2001.
16. Yu H, Hatzivassiloglou V, Rzhetsky A, Wilbur W J. Automatically Identifying Gene/Protein Terms in MEDLINE Abstracts. J Biomed Inform 2002; 35: 322-30.
17. Yakushiji A, Tateisi Y, Miyao Y, Tsujii J. Event Extraction from Biomedical Papers using a Full Parser. In: PSB 2001 Proc Sixth Pacific Symposium on Biocomputing; 2001; Hawaii, USA. p. 408-19.
18. Zervanou K, McNaught J. A Domain-Independent Approach to IE Rule Development. In: LREC 2004 Proc 4th International Conference on Language Resources and Evaluation; 2004 May 26-28; Lisbon, Portugal. p. 745-8.
19. Manning C, Schüzte H. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press; 1999.
20. Witten I, Paynter G, Frank E, Gutwin C, Nevill-Manning C. KEA: Practical Automatic Keyphrase Extraction. In: Proc 4th ACM Conference on Digital Libraries; Aug 1999; Berkeley, USA. p. 254-5.
21. Zhang Y, Milios E, Zincir-Heywood N. Narrative Text Classification and Automatic Key Phrase Extraction in Web Document Corpora. In: Proc 7th ACM International Workshop on Web Information and Data Management; 2005 Nov 5, Bremen, Germany. p.51-8.
22. Aronson A R. MetaMap: Mapping Text to the UMLS® Metathesaurus®. 1996 March [cited 2007 March]; http://skr.nlm.nih.gov/papers.
23. Aronson A R. MetaMap Variant Generation. 2001 May [cited 2007 March]; http://skr.nlm.nih.gov/papers.
24. Aronson A R. MetaMap Candidate Retrieval. 2001 July [cited 2007 March]; http://skr.nlm.nih.gov/papers.
25. Aronson A R. MetaMap Evaluation. 2001 May [cited 2007 March]; http://skr.nlm.nih.gov/papers.
26. Ananiadou S, Albert S, Schuhmann D. Evaluation of Automatic Term Recognition of Nuclear Receptors from Medline. Genome Inform Ser Workshop Genome Inform 2000 Dec 11;11:450-1.
27. Li Y,Bandar Z A, McLean D. An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. IEEE Trans Knowl Data Eng 2003 Jul/Aug; 15(4):871-82.

28. Petrakis E G, Varelas G, Hliaoutakis A, Raftopoulou P. Design and Evaluation of Semantic Similarity Measures for Concepts Stemming from the Same or Different Ontologies. In Proc 4th Workshop on Multimedia Semantics; 1998; Chania, Greece. p. 44-52.
29. TREC:Text REtrieval Conference TREC-9 Filtering Track Collections: OHSUMED [cited 2007 March]; http://trec.nist.gov/data/t9_filtering.html.
30. Salton G. Automatic text processing: the transformation analysis and retrieval of information by computer. Reading, MA: Addison-Wesley; 1989.