28. Lawrence D. W. and Laflamme L. Using online databases to find journal articles on injury prevention and safety promotion research: key journals and the databases that index them, Inj. Prev. 2008;14;91-95

29. Chang, C. , Beghtol, C. , Mackenzie, S. , Maurice, P. , Peck, S. , Rogmans, W. , et al. ,. Thesaurus of Injury Prevention Terminology. SMARTRISK, Toronto 2003.

30. Gallagher, L. A. , Thesaurus of Psychological Index Terms, vol. iv, 10th ed. American Psychological Association, Washington, DC 2005..

31. Gore, G. , Searching the medical literature. Injury Prevention 9, 103–104, 2003.

32. McCray, A. T. , The nature of lexical knowledge. Methods of Information in Medicine 37, 353–360. 1998

33. Lawrence D. W. , Using online databases to find peer-reviewed journal articles on injury prevention and safety promotion research: a study of textword queries by SafetyLit users, Injury Prevention 2007;13: 232–236.

34. Murphy L. S. , Reinsch S. et all, Searching biomedical databases on complementary medicine: the use of controlled vocabulary among authors, indexers and investigators, BMC Complementary and Alternative Medicine 2003, 3: 3

35. IJzereef1 L. , Kamps J. and de Rijke M. , Biomedical Retrieval: How Can a Thesaurus Help? LNCS 3761,pp. 1432-1448, 2005.

36. A. C. Yu, "Methods in biomedical ontology," J. Biomed. Ontol. , vol. 30, no. 3, pp. 252–266, 2006.

37. Taboada M. , Lal´ın R. , and Mart´ınez D. , An Automated Approach to Mapping External Terminologies to the UMLS, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 6, , page 1598-1605, JUNE 2009.

38. S. Riedel, E. Klein Genic interaction extraction with semantic and syntactic chains, In Proceedings of ICML05 Workshop on Learning Language in Logic (LLL05) , 2005.

39. Kim M. Y. , Detection of Protein Subcellular Localization based on a Full Syntactic Parser and Semantic Information, Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD. 2008) , 2008.

20. NAKAYAMA K. , HARA T. and NISHIO Sh. A Thesaurus Construction Method from Large ScaleWeb Dictionaries, Proceeding of the 21st International Conference on Advanced Networking and Applications (AINA'07) , 2007.

21. Kongthon A. , Haruechaiyasak Ch. And Thaiprayoon S. Constructing Term Thesaurus using Text Association Rule Mining, Proceedings of ECTI-CON 2008

22. Popek G. and Katarzyniak R. P. Agent-based Generation of Personal Thesaurus, Proceedings of the First Asian Conference on Intelligent Information and Database Systems (ACIIDS. 2009) , 2009.

23. Milne D. , Medelyan O. and Witten I. H. Mining Domain-Specific Thesauri from Wikipedia: A case study, Proceedings of the 2006 IEEE/ WIC/ACM International Conference on Web Intelligence (WI 2006) Main Conference Proceedings (WI'06) , 2006.

24. Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R and Musen MA, Ontology-driven indexing of public datasets for translational bioinformatics , Centre for Biomedical Informatics, Feb. 2009.

25. Shah NH, Rubin DL, Supekar KS, Musen MA. , Ontology-based annotation and query of tissue microarray data, AMIA Annu Symp Proc. 2006: 709-13.

26. Lowe, HJ; Barnett, GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA. 1994 Apr 13;271 (14) : 1103–1108.

27. Lawrence DW. Searching the scholarly literature for injury prevention-related articles. Proceedings of the National Injury & Violence Prevention Research Conference, Oct 10 2007, Columbus, OH, USA: Nationwide Children's Hospital and the Society for Advancement of Violence and Injury Research (SAVIR) 2007.

11. Sintichakis M. and Constantopoulos P. , A Method for Monolingual Thesauri Merging. In Proc. 20th International Conference on Research and Development in Information Retrieval, ACM SIGIR, Philadeplphia PA, USA, July 1997.

12. Voorhees E. M, "Using WordNet to Disambiguate Word Sense for Text Retrieval" , Proc ACM SIGIR'93, Pittsburgh, 1993, 171-180.

13. Miller G. A. , WordNet: a lexical database for English" , Communications of the ACM, 38 (11) , 39–41, 1995.

14. Liddy, E. D. and Paik W. , Statistically-guided word sense disambiguation. In Proceedings of AAAl Fall '92 Symposium on Probabilistic Approaches to Natural Language (Boston, Mass.). AAAI, Menlo Park, Calif. (1992).

15. Chodorow M. S. , Byrd R. J. and Heidorn G. , Extracting semantic hierarchies from a large on-line dictionary, In 23ed annual meeting of the association for computational Lingustics, pp 299-304, Chicago, Illinois, July, 1985.

16. Markowitz J. , Ahlswede T. and Evens M, Semantically significant patterns in dictionary definition. In Proc. of ACL-24, pp 112-119, 1986.

17. Tingting He and Jing, Li. Building relationships for hypernym and hyponym among Cilin synsets under the guidance of WordNet, SWCL, 2004.

18. Tingting HE, LI J. and Donghong, JI. Discovering ISO Relations among Synsets in a Flat Thesaurus Automatically, Proceeding ofNLP-KE'05, 2005.

19. Shepherd M. , Watters C. and Young J. , Context Thesaurus for the Extraction of Metadata from Medical Research Papers , Proceedings of the 37th Hawaii International Conference on System Sciences – 2004.

References:

1. Roe S. K. and Thomas A. R. , "The Thesaurus: Review, Renaissance, and Revision" , Routledge, 2004

2. Salton G. , McGill M. J. , Introduction to modern information retrieval. McGraw Hill, New York, 1983.

3. Aitchison J. , Gilchrist A. , and Bawden D. , Thesaurus Construction. 3ed. ASLIB, 1997.

4. Abuzir Y. , Vervenne D. , Kaczmarski P. and Vandamme F. , "Extracting Semantic Relationships between Terms using IKEM Tool" , KIM/KIT NEWS, Vol. 15, nr. 3, Nov. 2000.

5. Golub K. , The Role of Different Thesauri Terms and Captions in Automated Subject Classification, Proceedings of the 2006 IEEE/WIC/ ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06) , 2006.

6. Crouch C. J. , An approach to the automatic construction of global thesauri, Information Processing & Management, 26 (5) : 629-40, 1990.

7. Qui Y. and Frei H. P. , Concept Based Query Expansion. Proc. of the 16th Int. ACM SIGIR Conf. on R&D in Information Retrieval, Pittsburgh, SIGIR Forum, ACM Press, June 1993.

8. G. Grefenstette, Use of syntatic context to produce term association lists for text retrieval. In SIGIR'92, pp. 89--97, 1992.

9. Ruge G. , Experiments on linguistically based term associations. In RIAO'91, pp. 528-545, 1991.

10. Mili H. , Rada R. "Merging Thesauri: Principles and Evaluation" . IEEE Transactions On Pattern Analysis and Machine Intelligence, 10 (2) : 204-220, 1988.

terms and the relationships between them. The information is often presented in an implicit way, and we aim at rendering it explicitly so that (ATCT) tool can make good use of it. The choice of the prearranged list as a source of information to construct thesaurus provides us with accurate results.

The main purpose of the Medical thesaurus will involve its use to index online documents related to Medicine. Once this tool is incorporated, users will be able to search for abstracts of articles from professional journals, reports from academic institutions and organizations, and proceedings of relevant conferences in a more comprehensive and systematic fashion. It is also envisioned that this tool will have a much broader application. Specifically, the Medical thesaurus offers medical resource centers, a taxonomic structure that facilitates consistent indexing of information resources, ultimately enabling professionals to conduct more effective and efficient information search.

in the thesaurus. Each scope note will be automatically mapped to the best matching concept (s) or term (s) in the thesaurus.

Finally the results of the thesaurus merging process can be viewed through our thesaurus manager tool.

Since the field of information retrieval is constantly evolving, new online search systems is required to know all of the different words that may apply to a concept in order to conduct a thorough search. These new online search engines have the same limitations as other text keyword based retrieval systems. This endeavor will create a multilingual thesaurus that maps concepts across many disciplines at an international level.

5. Conclusion:

We described our tool, which is an automatic tool for thesaurus construction from prearranged lists (ATCT) . We examined our tool with different types of prearranged lists. These prearranged lists may be a table of content, index of books or any other arranged lists (Dictionaries) . (ATCT) can be used to extract the terms and the relationships from these lists using heuristic rules and other syntactic information in any specific domain.

(ATCT) is an ideal tool for developing fast and accurate thesaurus form prearranged lists of terms based on patterns extraction function used in our tool. Examining the other approaches, we can see that their approaches are limited to extract association or hierarchical relationships between terms while our tool (ATCT) can extract all types of relationships of thesaurus. (ATCT) tool can extract the hierarchical, equivalence , association and scope notes relationships between terms.

Our experiment shows that a prearranged list is a good place to find terms and their semantic relations to build a thesaurus automatically. The structure of these electronic lists can contain significant information about

In many cases, a scope note provides a definition of the term and suggests another term to compare. In many thesauri, scope notes are provided only for cases in which searchers and indexers have to be instructed on the differences of terms, and for a term whose scope is unclear. In our medical thesaurus, scope notes are added to the terms in most cases to make their scope, meaning, usage and history clearer. Our main goal is to provide scope notes for all terms of the thesaurus. In our work, all of the terms will have definitions taken from machine-readable medical dictionaries to extract their definitions in order to populate the Scope Note field of our medical thesaurus. We believe that the inclusion of definitions for the majority of medical terms in the thesaurus will be a substantial aid to end-users as well as search intermediaries

In our approach, we have three phases for thesaurus updating and merging. The input of this process will be machine readable medical dictionaries and the result will be the new update multilingual thesaurus, This process will pass through three phases as shown in figure 3.

The first phase is Extraction; the input of this phase is readable medical dictionaries. The terms, definitions and translation will be extracted from these dictionaries and represented into an internal way that is recognized by the second phase.

Merging is the second phase which is the core of the matching process. A mapping processes will be applied to the output of the first phase. These processes will automatically generate various index terms (keywords, concepts, and relations) and the translations of these terms into French and English languages. The structure of the thesaurus will be used to create associations between new terms and the scope notes annotations and concepts

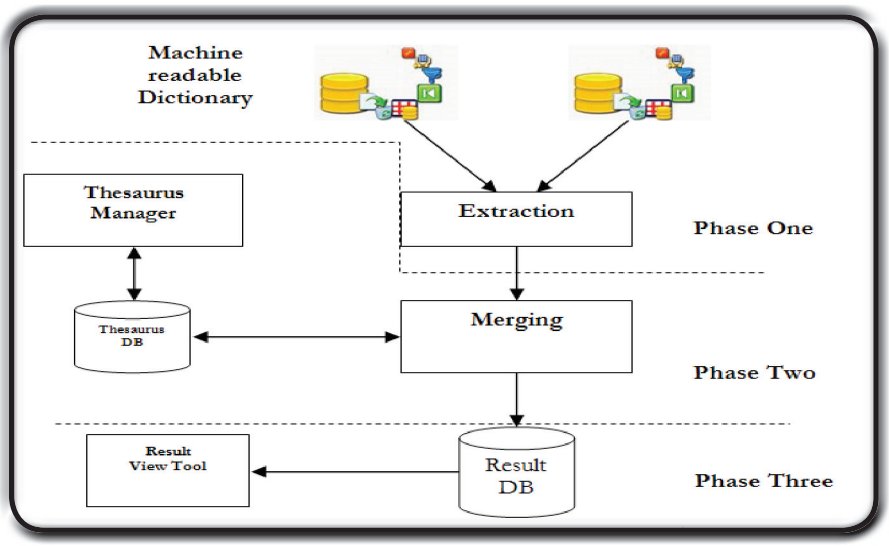
4.6 The Medical Thesaurus:

We are developing a thesaurus that clarifies important medical concepts and standardizes their corresponding terms. The primary purpose of the medical thesaurus is to index documents so that users can conduct online searches from a variety of starting points to retrieve related information with satisfactory completeness and with a minimum amount of irrelevant material. Other purposes could include providing definitions of terms, describing the scope of the concepts covered by each term, and demonstrating the interrelationships of concepts.

Although there are many reasons for creating this thesaurus, it is not intended to impose an ‘‘official language’’ upon practitioners in any field. If the standard vocabulary is to be widely accepted and successfully used, it will need to include all the various terms for each concept, their spelling variants, and words in use by non-experts that refer to the key concepts. Such a thesaurus would not only be an indispensable tool for facilitating access to documents but would also provide a multidisciplinary glossary of the medical terms.

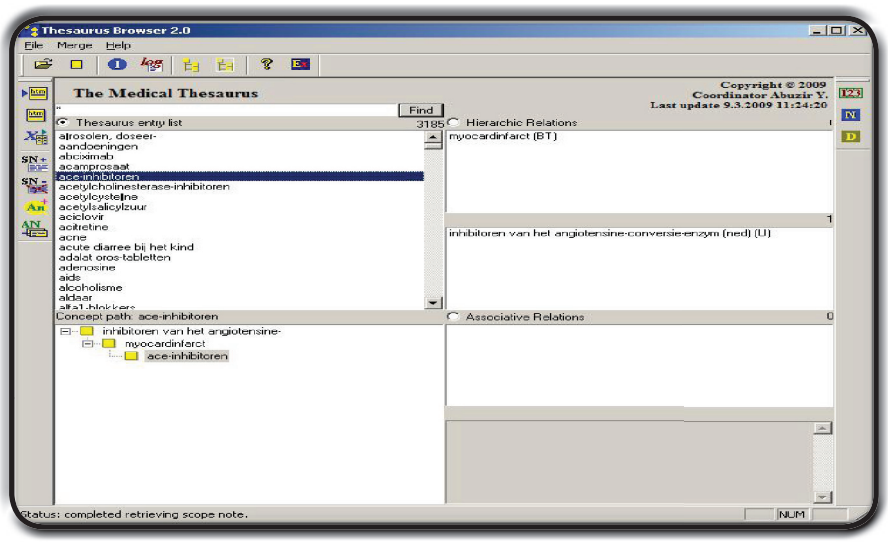
Figure (3)

Updating and Merging the Thesaurus



This regards to the different characters coding systems (see table 2) . Other functions were implemented to find these codes and replaced them with the correct characters. For the task at hand the results of the parsing are highly satisfactory. The noisy list of terms and the relationships (5%) were analyzed automatically and manually and added to the thesaurus.

Figure (2)
Medical thesaurus



The Medical Thesaurus contains different relationships, hierarchical relationships BT/ NT, equivalence relationships U/UF (Use/ Used for), associative relationship RT (Related Terms) and the SN (scope note) . Figure -2 shows the thesaurus. The first draft version of the thesaurus contains 3185 terms. The second phase of this thesaurus is concerned with building a multilingual thesaurus that support English and French languages. A tool (Figure 3) to update and merge thesauri was tested and will be used for the process of building the multilingual Thesaurus from Medical Databases[1], and machine readable Dictionaries

[1] CORK Database.

Hepatitis -B -C	24/24 - 25/77 - 26/62 23/94	The minus sign indicates a compound term. The system will form a new term by replacing by adding Hepatitis. The new term Hepatitis-A

The last example is the patterns Zie ook (see also) will give us the association relation between AIDS and HIV-infective, and the system will transform it to the thesaurus.

4.4 Summary:

A number of text preprocessing tasks were done to the list in order to prepare the terms for use in our automatic conversion tool. The preprocessing includes text operations like extracting and finding the boundary of the terms, removing special characters and character conversion. We also remove list of page numbers following the terms.

During text extraction we apply the following heuristic rules while constructing the thesaurus.

- All terms that start with uppercase characters is considered as BT term in the thesaurus structure.
- All terms that start with lowercase characters are considered as NT term in the thesaurus structure. We can see from this that all terms starting with lowercase characters considered as NT for the upper term that start with uppercase characters.
- Parenthesis was converted into UF Relationships.
- See used to represent the equivalence relationships.
- See also used to represent the association relationships.

4.5 Evaluation:

Finally, testing our approach on different real-world terminologies within the medicine or research realm helps evaluate the scope of the method's application. Our approach was tested to create a thesaurus using the (ATCT) and using the thesaurus in indexing medical databases. A sample of prearranged lists were used to construct the Medical Thesaurus and we compared our manual thesaurus with the automatic thesaurus construction automatically. At present, (95%) of the entries have been completely analyzed without error by our tool based on the set of rules used in the text extractions for single terms or to compose a compound term and relationships extraction. The other (5%) of the entries are basically spelling mistalres and some other minor errors.

- The state of the first character in the terms. There are two states, terms with capital letters and terms with small letters.
- Terms or letters have special characters like + or – at the first position or at the end of the terms

We have gathered more information about how to discover the relations; Table 2 shows sample rules used to discover the relations.

We used a set of rules to find the relations. The tool first tries parsing the entries in the list with the primary set of patterns. If no patterns are found, the tool adds terms to the thesaurus list. For example the term Bromocriptine will be added to the thesaurus as a term.

Let us take another sample, Table 2 shows an example with the terms Vitaminen that has 6 NT terms. The hierarchy is discovered by the heuristic rules. We illustrate how we extract the hierarchical relation using the above rules. This process is performed on all terms in the list and then all the hierarchical relations are found by this rule.

Table (2)
Sample from the prearrange list

Example		Description
reumato < de	23/76	ANSI to ASCII code conversion
Bromocriptine	27/2	Thesaurus term
Antipsychotic, see neuroleptica		Occurrence of Zie equivalence relation
AIDS see also HIV- infectie	24/55 - 24/56 - 24/89	Occurrence of Zie ook Association relation
Vitaminen congenitale afwijkingen supplementen vitamine A vitamine B6 en B12 vitamine E vitamine K	 24/13 26/49 24/13 27/53 27/30 24/14 - 27/81	Hierarchical relations. Terms that start with uppercase characters are considered as BT while the terms which follow that term and start with lowercase characters are considered as NT terms.

Broader Term BT. Other terms that do not start with capital letters are Narrow Terms NT. Such rules are used to identify the hierarchical relationships in the list.

4.2 Information Type:

Let us now concentrate on the nature of the information that should be used by the system. We have identified three types of information:

- Relationships in which the terms are extracted:
- Hierarchical Relationship (BT/ NT) ,
- Association relationship
- Equivalence relationship
- Determining the position and replacement of the special characters to form a complete term.
- ANSI to ASCII character table lookup

These might be termed as the assumptions of our tool.

According to the information mentioned before, the first class of information identifies and defines the terms and the relationships between the terms. The second one introduces the rules to form complete terms. This is of crucial importance in term extraction because it avoids making incorrect interpretations of the terms. The latter is called the terms correcting.

4.3 Structure of the lists:

In general, the structure of the list is a sequence of lines. Each of these lines consists of set of words represent a term or more and the relationships between terms in addition to the set of page number (s) related to that term. The main structure used in the list is as follows (Table 2) :

- Characters conversion from ANSI to ASCII code.
- Single term.
- Terms have an association or equivalence relations. These are multiple terms.

and uses special syntactic structure to represent the relationships. The advantage of the prearranged list is that some of its structures are used very often and it gives hints for the ways to create a thesaurus or extract information.

- The simplicity of the sentences (limit length, limit number of patterns) leads us to more limits of set rules to extract the relationships.
- The prearranged lists are a rich resource of terms.

4. Medical Thesaurus Construction:

4.1 General:

In this section, we present the structure of the prearranged lists and the relationships we are seeking. The methods used to extract the relationships and the results obtained are shown. This section presents the steps, ideas, process, and heuristics leading to automatic construction of thesaurus from prearranged list.

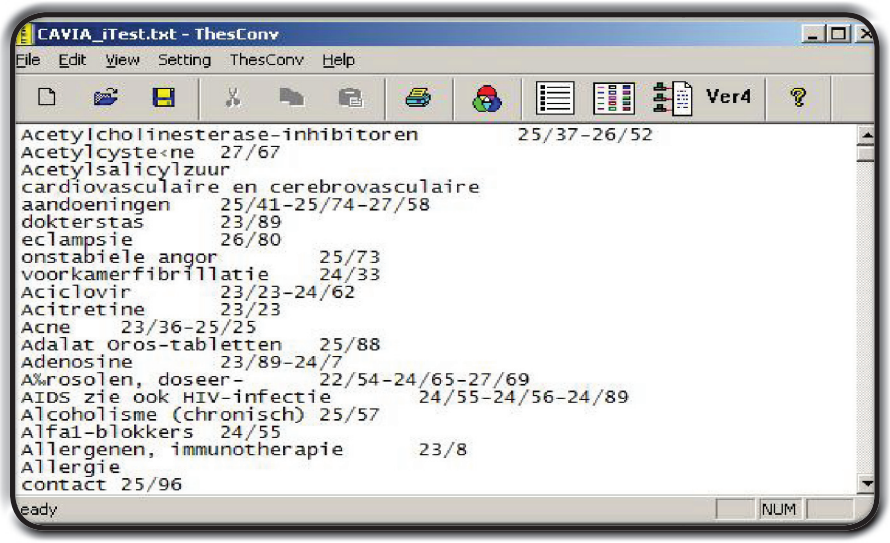
Our approach to construct a medical thesaurus is based on sets of heuristics rules. The first step involves parsing the list using our tool: Automatic Thesaurus Converter Tool (ATCT) . The resulting parse structure is then subjected to a set of heuristics rules whose goal is to identify the syntactic and lexical patterns which are consistently associated with some specific semantic relations.

The list was analyzed automatically as a whole and a number of rules were discovered. By using these rules, (ATCT) can organize the list into a structured thesaurus. A procedure for extracting and using these rules was developed and the Results were positive.

Different heuristics rules were used to discover and characterize the relationships between terms. The goal at this stage is to be existted between the terms in the list. Analyzing the list shows that uppercase or lowercase states of the letter are considered to be an indication of hierarchical relationships between terms. Terms beginning with capital letters were considered as

Figure (1)

Automatic Thesaurus Construction Tool ((ATCT))



3.2 Characteristics of the Prearranged Lists:

Deciding to work with prearranged lists reflects our belief that the prearranged lists are a good source for thesaurus construction. These lists are rich in domain information and terms.

The scope of the vocabulary used in the prearranged lists is not the same as in the unrestricted text. Moreover, the language used in the prearranged lists is appropriately called structured. The regularity of the language used within the prearranged lists definition lies in the frequent occurrence of lexical and syntactic patterns to express particular semantic relationships.

Now, let us summarize our justification for choosing the prearranged lists as our source for thesaurus construction:

- The prearranged lists give strong enough information for thesaurus construction. They are a very good starting point.
- The prearranged lists contain the same vocabulary as a plain text,

productive to use a thesaurus that was constructed to facilitate finding material[35].

Research oriented to promote using controlled vocabularies is an extensively recognized topic in a biomedical community [36]. The proliferation of biomedical terminologies and the need to use them in many health care activities, as well as in information retrieval have increased their value as knowledge resources. Providing interoperability between different knowledge sources is also a critical issue for efficient information, sharing in other communities [37].

To improve performance in detecting protein subcellular localization information, the author attempted to use semantic information from the Word Net thesaurus. Furthermore, they demonstrated that syntactic and semantic information is important for the performance of this method [38,39].

3. An Overview:

We propose an efficient thesaurus construction method based on prearranged lists. In the following sections, we describe our approach in detail after describing the characteristics of the prearranged lists.

3.1 Basic Strategy:

A thesaurus is a data structure that defines semantic relations between words. Relations have been extracted from prearranged list in a two step procedure:

- First: parsing the list,
- Second: using patterns to find the relationships.

The first step involves parsing the list using our tool (Automatic Thesaurus Construction Tool ((ATCT)) figure-1) . The resulting parsing structures are then subjected to a set of heuristic rules whose goal is to identify the occurrence of these patterns which are an indication of some specific relations. These patterns are used to decide what relationships should be attached to the term in the list.

Popek G. and Katarzyniak R. P described an approach to extract relations. Extracting relations presented in their work based on finding relations between concepts in general needs all allowed combinations of concepts and hedges to be checked results in great complexity if done without planning. There are two main ideas for this task. The first one is to remember the sets for each property X once they are extracted. Most likely the sets will be used multiple times and it is a waste of resources to calculate them every time. The second idea is to find dependencies between co-occurrence of relations. In using these dependencies, some cases can be excluded from the search. Because of that, dependencies between co-occurrence of relations are listed and used in an algorithm for a determination and update of thesaurus to potentially reduce its complexity [22]. Another important step is to take into consideration knowledge stored in the thesauri. It should be used at least in such areas as planning.

Milne D. , Medelyan O. and Witten I. H. ,Mining , have shown how to construct domain- and corpus specific thesauri from Wikipedia. Comparing terms and semantic relations to those in a manually created thesaurus demonstrates excellent coverage of domain terminology, and of synonymy relations between terms [23].

There are many studies [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34] that are a series of investigations that may help to shed light on the knowledge, skills, and practices that researchers and practitioners use when seeking medical literature. A thorough search of the literature is important. Incomplete literature search may result in a distorted interpretation of the body of research on a topic. Decisions that are based on incomplete information are poorly informed and may waste time, work effort, and money especially if that information is gathered from a few familiar sources using only search terms that are familiar.

Although searching by keywords is usually highly focused, there are cases where a keywords search may produce excessive irrelevant information, particularly for words with multiple meanings. Thus, it is probably more

The second approach is merging existing thesauri [10], [11]. This approach is appropriate when two or more thesauri for a given subject exist that need to be merged into a single unit. If a new thesaurus can indeed be served by merging two or more existing thesauri, then a merger perhaps is likely to be more efficient than producing the thesaurus from scratch.

The simplest approach to construct a thesaurus automatically is to reuse existing online lexicographic databases, such as Word Net [12], [13] Longman's subject codes [14]

One important relation between terms is the hierarchy relation. It has given a lot of importance in recent work on knowledge extraction and thesaurus construction from dictionaries [15], [16].

In recent years, novel methods for thesaurus construction based on structured documents are getting much attention. Tingting et al proposed an unsupervised method for relation discovery in a flat thesaurus. The key idea includes mapping synsets in a flat thesaurus to Mapped Areas in Word Net, and getting the relations among synsets according to the average distances between the Mapped Areas in Word Net's semantic structure. They have made experiments to discover the relations of Immediate Super-Ordinate among nouns in Cilin and they did not discover other relations in Cilin automatically [17,18].

Shepherd et al. , described an approach to the automatic extraction of metadata from medical research papers. Medical research papers tend to have stereotypic prescribed sections, such as introduction, methods, and conclusions. The approach described uses context thesauri and the semantic structure of the documents to extract metadata based on these stereotypic sections and if they appear in the appropriate sections and in the size of the context windows from that document [19].

Kongthon A. etal. , proposed a method to analyze the link structure of Web-based dictionaries to construct an association thesaurus. In their project, they extracted term associations only and not other term relations such as "is-a" or "part-of." (Hierarchical relations) [20, 21].

The work presented focuses on the extraction of the different relationships (hierarchical, association and equivalence relationships) that best characterize the term in these lists. Part of this research outlines how Medical Thesauri may help resolve a number of common difficulties encountered when searching online recourses.

An important part of this work will be the idea of using prearranged lists to construct a thesaurus automatically. The main basic elements that will be studied in our work to construct a thesaurus from prearranged lists are:

- The choice of the list.
- The type of the relationships.
- Discovering of the patterns that represent the relations between terms.
- The use of the patterns to extract the relationships.

To discover the different relationships of the thesaurus, we need to examine the format or structure of these lists. Each list has a different format to represent relationships and different patterns to discover these relationships. We developed a tool to extract these patterns automatically from the list and the terms related to these patterns to create the thesaurus.

In section 2 we discuss related work. Section 3 looks at all the steps to transform the terms found in the list into thesaurus. We also show the extraction of the relation's type as we build the thesaurus. Section 4 discusses the construction of the medical thesaurus from a prearranged list like book index. Finally, we give our conclusion in section 5.

2. Related Work:

Various techniques and approaches were used to construct thesaurus. Building manual thesauri requires a lot of human labor from linguists or domain experts and they are expensive to build. Since it is difficult and expensive to build thesauri manually, many researchers attempted to construct thesauri automatically [3]. There are different approaches to construct a thesaurus. The first approach, on designing a thesaurus from document collection, is a standard one [2], [4], [6], [7], [8], [9]. Here the idea is to use a collection of documents as the source for thesaurus construction. This assumes that a representative body of text is available. By applying statistical or linguistic procedures, we can identify important terms as well as their significant relationships. These can be viewed as the raw information from which prearranged list or dictionaries are made.

to assign appropriate terms to individual documents. It is also an important tool for formulating good search strategies. Thesauri can be used to suppress retrieval of non-relevant documents (i.e. increasing precision) by using narrow terms. It can also be used to retrieve more relevant documents (i.e. increasing recall) by expanding the query with related terms [5].

Undoubtedly, thesaurus technology indicates an innovative approach in advanced document management, especially in the field of document indexing and retrieval [1], [2], [3]. Every day millions of people are searching for information stored in documents somewhere locally, in an intranet or on the Internet, in order to find the specific information they want to find at a specific time. Another characteristic feature of a thesaurus is using concepts for indexing which allows a document to be ranked as relevant to a query even if the query term itself does not occur in the text, but only a related term which denotes the same concept applied to document retrieval. This means that thesaurus based indexing allows documents to be retrieved even if one or more of some given words in a search string do not match any word or combination of words in the documents.

The number of researches relevant to medicine is increasing rapidly in many disciplines. A Medical thesaurus identifies and clarifies important concepts from these disciplines and standardizes their corresponding terms. The Medical thesaurus may be used as an indexing and search tool so that users can conduct online searches with satisfactory completeness but with a minimum amount of irrelevant material. Without the Medical thesaurus, researchers may miss important or relevant information related to their discipline studies or researches.

This work shows that standard prearranged lists (e.g. index of books and other list) define highly connected terms linked through basic semantic relations. It deals with the exploitation of The prearranged list of terms for the automatic construction of thesaurus. The prearranged list provides us with more information that is not easily obtained from raw text to construct a thesaurus.

1. Introduction:

This work focuses on automatic construction of medical thesauri from prearranged lists and the use of these thesauri. The thesaurus is designed in order to help researchers and searchers in the domain of Medicine to gain an understanding of the semantic structure of the concepts and the keywords of their queries and to receive a quick answer concerning their queries with specific keywords. A thesaurus has long been a concern in lexicography, and recently, it has found many applications in machine translation, information retrieving, and computational lexical semantics, etc. [1], [2], [3], [4].

A thesaurus is a structured system of terms encoding explicitly semantic relations like synonyms, hyponyms, hyperonyms, part-whole relations, associations, etc. Table 1 shows the relationships between terms in thesauri:

- The equivalence relationship
- The hierarchical relationship
- The associative relationship

Table (1)

Rrelationships between terms

Relationship	Indicator	Abbreviation
Equivalence	Use	none or U
	Used for	UF
Hierarchy	Broader term	BT
	Narrower term	NT
Association	Related term	RT
Scope Notes		SN

A thesaurus provides a precise and controlled dictionary that can be used to coordinate document indexing and retrieval. It can also be used to help searchers write some precise queries. A well-designed thesaurus can be very useful in subject searching in an online document collection. It is an irreplaceable tool as an aid in indexing. An indexer can consult the thesaurus

Abstract:

There is an interest in extracting knowledge and retrieving information automatically from the current availability of a large collection of electronic resources and from the academic literature available on the Web. Much of Thus, researchers may miss relevant, even critical, information in parallel or disparate fields, which in turn may lead them to recreate information already developed or to miss important information that could advance each discipline. A Medical thesaurus identifies and clarifies important concepts from these disciplines and standardizes their corresponding terms. This paper reviews the structure, construction and use of Medical thesauri, and outlines how Medical Thesauri may help resolve a number of common difficulties encountered when searching these recourses. The increasing importance of the medical online resources as information resource makes the thesaurus an aid tool for both professionals and non-experts.

Keywords:

Medical thesaurus, Thesauri use, thesaurus construction, indexing, information retrieval.

ملخص البحث:

تتوافر في العصر الحالي لتكنولوجيا المعلومات والاتصالات العديد من المصادر الالكترونية والمؤلفات الأكاديمية و الأبحاث العلمية التي يمكن الوصول إليها عن طريقة شبكة الانترنت. أدى توافر هذه المصادر بنمط الكتروني إلى اهتمام كبير في استخلاص المعرفة، واسترجاع المعلومات بطريقة أتوماتية بهدف الحصول على المعلومات الصحيحة التي لها علاقة باهتمام الباحث. إن عدم توافر طرق استخلاص المعرفة واسترجاع المعلومات بصورة أتوماتية للباحثين قد يكون أحد الأسباب التي تجعل الباحث لا يجد المعلومات المهمة وذات الصلة بالمواضيع المتباينة، او ذات العلاقة ببحثه العلمي ، والتي قد تؤدي بهم إلى تكرار البحث العلمي والمعلومات التي توصلوا إليها سابقا في مراكز الأبحاث العلمية، أو عدم الحصول على المعلومات المهمة التي يمكن أن تقدم لكل تخصص.

يمكن استخدام هيكلية المفاهيم الطبية (المكنز الطبي) في مجال البحث العلمي في الطب. هذه الهياكل تحدد المفاهيم المهمة في هذه التخصصات وتوضحها، وتوحد آلية الرجوع إليها. يقدم هذا البحث تركيب هيكلية المفاهيم الطبية وبناءها واستخدامها ، ويبين ان هيكلية المفاهيم الطبية قد تساعد في حل العديد من الصعوبات المشتركة التي يواجهها الباحث عند البحث في هذه المصادر الالكترونية. إن الأهمية المتزايدة للموارد الطبية على الانترنت كمصدر معلومات يجعل المكنز الطبي أداة مساعدة في البحث عن المعلومات الطبية لكل من المهنيين والخبراء وغير الخبراء في هذا المجال.

إن القوائم المرتبة مسبقاً مثل محتويات الكتاب والفهارس وغيرها تعدُّ مصدراً جيداً للعثور على الكلمات المفتاحية (المفاهيم) والعلاقات بينها لبناء معجم وهيكلية المفاهيم آليا باستخدام الحزمة البرمجية ATCT التي صُمِّمت في هذا البحث. سيكون لهذه الأداة تطبيق على نطاق واسع، وبخاصة أنها ستكون المعجم الطبي المركزي للمصادر الطبية المتعددة اللغات وهيكلية تصنيف لغوية تسهل الفهرسة بشكل موحد من مصادر المعلومات، مما يمكن في نهاية المطاف الخبراء والباحثين من إجراء البحث عن المعلومات بطريقة أكثر فعالية وكفاءة.

Constructing the Medical The saurus as a Tool for Indexing

Yousef Abuzir *

* Computer Information Systems Department, AlQuds Open University, Palestine.