Proceedings Paper:

Sen, B.A. (2011) The viability of automatic indexing for biomedical literature. In: International Journal of Health Information Management Research. 15th International Symposium on Health Information Management Research, 8 - 10 September 2011, Zurich, Switzerland. ISHIMR .

The viability of automatic indexing of biomedical literature.

Jane Sharp[1], Barbara Sen[2]

[1]British Library, Boston Spa, Wetherby, UK. jane.sharp@bl.uk

[2]Information School, University of Sheffield, Regent Court, Sheffield, UK. b.a.sen@sheffield.ac.uk

Automatic indexing is evaluated as an aid/replacement to manual indexing for biomedical literature. Manual indexing is costly and labour intensive. Technological innovations have the potential to increase efficiency and reduce costs. British Library produces a bibliographic database of allied and complementary medicine (AMED). This study compares articles which have been indexed manually for AMED with the same documents submitted to an automated indexing tool. The indexing tool selected was Helping Interdisciplinary Vocabulary Engineering, (HIVE) which is a jointly funded project by the University of North Carolina and the National Evolutionary Synthesis Center, North Carolina. A random selection of 100 records from a total of 1059 articles was selected. Each manually indexed document was compared with results returned by HIVE. Data analysis was made using SPSS. Results showed that HIVE does not provide a suitable replacement for the skills of a human indexer. Continued development of automatic indexing tools is recommended.

Keywords
AMED, automatic indexing tools, HIVE, knowledge organisation systems, medical indexing.

1. Introduction

This study evaluates the properties of an automatic indexing tool to establish if it can be a viable replacement or aid to manual indexing methods for biomedical literature. The context is within the British Library Health Care Information Service that produces a bibliographic database of allied and complementary medicine. Manual indexing is a costly labour intensive process and technological innovations that increase effectiveness and improve efficiency could help reduce costs and bring improvements.

Indexing involves assigning terms or concepts to a document usually from a controlled vocabulary or thesaurus to describe it. A controlled vocabulary is an established list of standardised terminology for use in indexing and retrieval of information. Adding terms from a controlled vocabulary that both the indexer and searcher understand aids the recall of a search. Manual indexing, which has previously been the norm for assigning concepts or descriptors to a document, is being replaced by automatic methods particularly for science, technology, engineering and mathematics (STEM) subjects.

The US National Library of Medicine (NLM) use a controlled vocabulary, the Medical Subject Headings (MeSH) for indexing items on to its MEDLINE bibliographic database via the Data Creation and Maintenance System (DCMS) interface. MeSH provides indexers with descriptors or subject headings to enable them to describe the content of a document. Subheadings or qualifiers which make the subject heading more specific are assigned as well as other descriptors such as check tags and publication types (a check tag is a MeSH heading that is looked for routinely by indexers in every journal article e.g. HUMANS). Increasingly, due to the volume of biomedical literature and sheer size in MEDLINE, automatic indexing tools are needed to accommodate these documents [1]. Manual indexing is labour intensive [1], and costly [2], which has prompted the development of automatic tools in an effort to reduce time and costs. NLM indexers are trained over several weeks and it can take up to two years to become a competent indexer [3]. The possibility that an automatic indexing tool can compete with the knowledge and skills of trained indexers is considered in this study.

2. Context for the study

The British Library employs medical indexers to index material for a healthcare database, called the Allied and Complementary Medicine Database (AMED). AMED is a unique bibliographic database produced by the Health Care Information Service of the British Library which covers a selection of journals in complementary medicine, palliative care, and several professions allied to medicine. The database covers the years from 1985 to present and is updated monthly. At the time of its conception allied and complementary medicine topics were not fully represented in the medical literature, or fully available on bibliographic databases such as Medline. AMED's purpose was to make this material more available via commercial data providers such as OVID and Dialog.

2.1 Automatic indexing

In order to establish if automatic indexing is comparable to manual methods this study compares articles which have been indexed manually by human action with the same documents that are submitted to an automated indexing tool. The online indexing tool selected for this study was Helping Interdisciplinary Vocabulary Engineering, or HIVE [4] which is a jointly funded project by the Institute of Museum and Library Services (IMLS) involving the Metadata Research Center (MRC) at the School of Information and Library Science, University of North Carolina at Chapel Hill and the National Evolutionary Synthesis Center (NESCent) in Durham, North Carolina. HIVE is described as having an 'automatic generation approach' that 'integrates discipline-specific controlled vocabularies encoded with the Simple Knowledge Organisation System (SKOS)'. SKOS is a World Wide Web Consortium (W3C) standard [5]. The advantages of HIVE, as described by its researchers, are in its cost, interoperability and usability [4]. HIVE is intended to assist with subject cataloguing. Hive is experimental but functional, and is freely available online. HIVE allows users to select a vocabulary source such as MeSH, upload a document or enter a URL. MeSH is mapped to SKOS in this process. **Figure 1** shows the HIVE screen for selecting the vocabulary source and uploading the document.

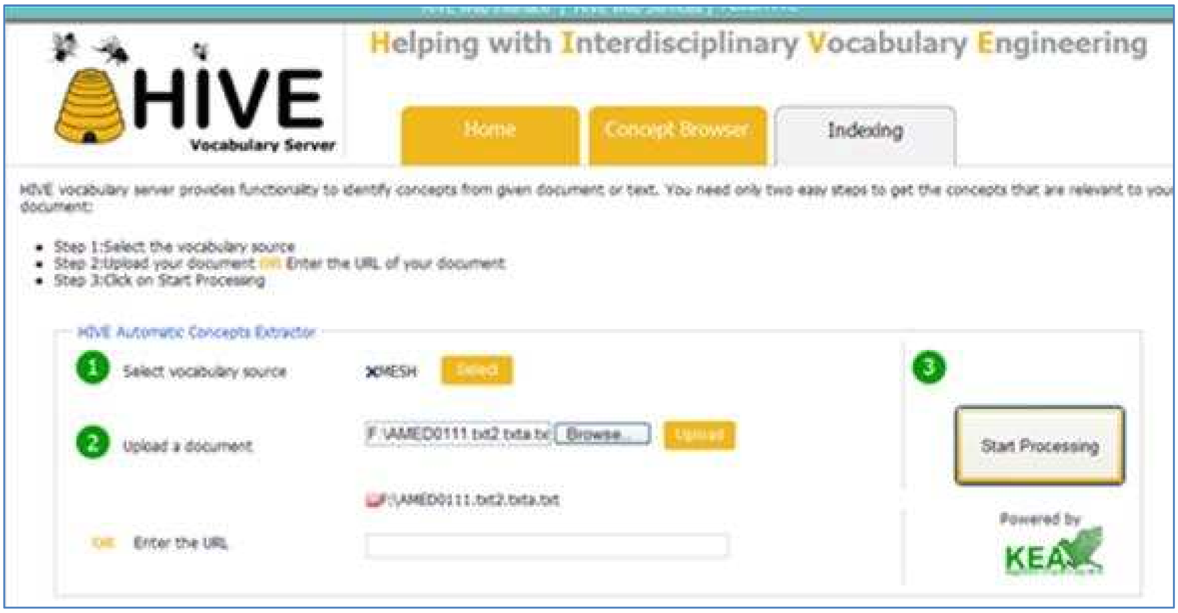
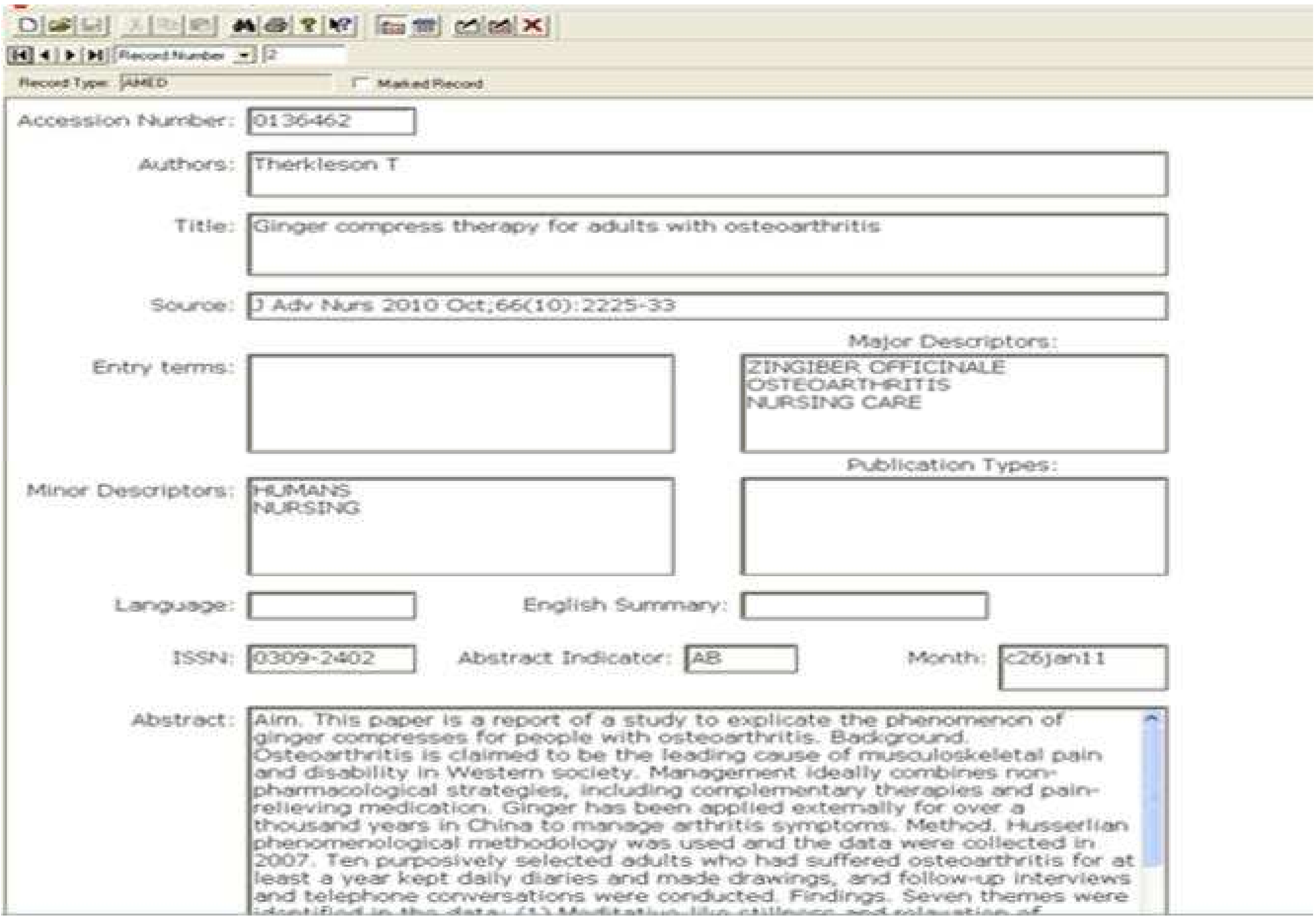


Figure 1. HIVE Screen shot for selecting vocabulary source and uploading documents.

2.2 Manual indexing via AMED

The process of manual indexing involves selecting articles from journals which are considered relevant to the subject area for example *Planta Medica*. Articles are then scanned to *Library Master* which is a system that manages bibliographic information. The article can consist of just the title but an abstract can be available. A photocopy of the front page of the article is usually provided for the indexer. Once on the system the human indexer will assign concepts (subject headings e.g. Cerebral Palsy) to the article to describe it. This is done using the knowledge and expertise of the indexer, and the AMED bespoke thesaurus which is based on MeSH. Indexers attempt to cover the concepts of the title in particular but also the rest of the article/document. The indexer will also assign subheadings which are qualifiers that make the concepts more specific i.e. PAIN with PREVENTION. Minor concepts such as RISK FACTORS or COSTS AND COST ANALYSIS or check tags such as HUMANS, RATS etc. and Publication types such as REVIEW or RANDOMIZED CONTROLLED TRIAL (RCT) are also added. The latest AMED thesaurus in use was updated in 2006 so it lacks any new concepts that might appear on MeSH since that date. It has approximately four thousand descriptors as opposed to MeSH which has in excess of twenty-six thousand as well as entry terms which direct indexers to appropriate headings. The thesaurus retains some local terms and does not have the scope or detail of MeSH. A typical finished AMED record for one article that appears on *Library Master* after indexing can be seen in **Figure 2**.



Record Number: 2
Record Type: AMED
Marked Record
Accession Number: 0136462
Authors: Therkleson T
Title: Ginger compress therapy for adults with osteoarthritis
Source: J Adv Nurs 2010 Oct;66(10):2225-33
Entry terms:
Major Descriptors: ZINGIBER OFFICINALE; OSTEOARTHRITIS; NURSING CARE
Minor Descriptors: HUMANS; NURSING
Publication Types:
Language:
English Summary:
ISSN: 0309-2402
Abstract Indicator: AB
Month: c26jan11
Abstract: Aim. This paper is a report of a study to explicate the phenomenon of ginger compresses for people with osteoarthritis. Background. Osteoarthritis is claimed to be the leading cause of musculoskeletal pain and disability in Western society. Management ideally combines non-pharmacological strategies, including complementary therapies and pain-relieving medication. Ginger has been applied externally for over a thousand years in China to manage arthritis symptoms. Method. Husserlian phenomenological methodology was used and the data were collected in 2007. Ten purposively selected adults who had suffered osteoarthritis for at least a year kept daily diaries and made drawings, and follow-up interviews and telephone conversations were conducted. Findings. Seven themes were

Figure 2. A typical completed record for one article that appears on *Library Master* after manual indexing.

The indexing carried out on the *Library Master* system is split between four different fields:

- **Semi-controlled field** i.e. survivors, supination etc. (concept not included in AMED thesaurus which has to be placed in this field as *Library Master* will not accept it as a major heading, but is a searchable field).
- **Major concept** (subject heading) i.e. CEREBRAL PALSY, DIABETES MELLITUS etc.
- **Minor concept** (including check tags, subheadings etc.) i.e. HUMANS, CHILD, ANIMAL, RATS, RISK FACTORS, PREVENTION etc.
- **Publication type** i.e. Randomized Controlled Trial, Review etc.

In any one month approximately 1,000 articles are indexed and appear on *Library Master* as completed records. Up to three indexers can be involved in the process. After proof checks are made the file containing the records is exported to a WORD document. Final bibliographic checks are then made before the file is sent to the commercial information hosts (OVID, Dialog etc.).

2. Related Literature

The development of automatic approaches to assist in indexing has been the subject of considerable research [2]. For MEDLINE, indexers have several aids to assist them in assigning MeSH terms or concepts to a document such as NLM's Medical Text Indexer (MTI). MTI recommends MeSH terms to indexers based on the title and abstract of an article [2]. According to Vasuki and Cohen it does this by "using a combination of distributional and vocabulary-based methods" [2]. The MTI system relies on MetaMap indexing where mapping occurs using the Unified Medical Language System (UMLS) Metathesaurus. UMLS represents information about the biomedical language and domain knowledge of biomedical text in machine readable form [6]. Concepts from UMLS are then ranked and restricted to MeSH descriptors. The MTI system also involves PubMed Related Citations. Vasuki and Cohen [2] stated that related citations are obtained by "computing the similarity between citations".

Vasuki and Cohen looked at combining the nearest neighbour search employed by NLM, which looks at similar indexed articles, with Reflective Random Indexing (RRI) [2]. Vasuki and Cohen noted that RRI "draws on meaningful associations between terms that do not occur in the same document" [2]. Their research concluded that the results obtained by this method outperformed those of MTI.

In similar research Neveol et al. [1], supported by NLM, looked at the automatic assignment of subheadings to recommend to indexers using methods such as Natural Language Processing and machine learning methods. Subheadings are an important part of the indexing process as they make the subject heading or concept more specific (i.e. CEREBRAL PALSY with PHYSIOPATHOLOGY. They also aid retrieval as the narrow the focus of the search. They can also be used for statistical purposes. Pair recommendations of subject heading and subheadings obtained with the methods described by Neveol et al. are an important automatic indexing advancement for MTI [1].
Neveol et al. had previously looked at pair recommendations as well as single terms to assist the Catalogue et Index des Sites Medicaux de langue Francaise (CISMeF) indexers for the cataloguing and indexing of medical on-line resources [7]. The system was designed to be used as a tool for preliminary indexing or as an aid to indexers when reviewing. The CISMeF indexing system interacts with the INTEX platform for MeSH term extraction. Neveol et al. stated that INTEX is a "powerful corpus analysis tool, which may also be used as a linguistic toolbox" [7]. Their results concluded that the system was helpful to indexers and that the precision was comparable to similar systems. Check tag retrieval was proved to have an increased recall and precision.

The literature reveals that a number of technological aids to manual indexing are being developed to aid medical indexers. The British Library is currently exploring the use of an automatic indexing tool (HIVE) as an aid or replacement manual indexing methods. This study presents a comparison of data generated by both manual and automatic indexing of a random sample of 100 records from a total of 1059 articles available for indexing during a one month period.

3. Methodology

A random selection of 100 article records from a total of 1059 articles that had been indexed on to *Library Master* from January 2011 were selected. Of the 100 bibliographic records selected randomly (using a random sample table):

- 54 records had the title and abstract from which to generate indexed concepts
- 46 records had the title only
- 96 records were in English
- 4 records were not English language

All three indexers employed by British Library had contributed to the indexing. Each record was exported from *Library Master* to Notepad and then this was copied to an Excel spread sheet with all the indexing in place. A copy was made of each document and then all the indexing was removed to leave the basic original details (**Table 1**). Each document without indexing was then individually uploaded to HIVE via the online link. MeSH was selected for each document as this is comparable to the AMED thesaurus. HIVE returned its results in the 'extracted concepts cloud' (**Figure 3**).

Table 1. Example - AMED document with indexing removed prior to uploading to HIVE and with indexing

Bibliographic record with indexing removed
Therkleson T
Ginger compress therapy for adults with osteoarthritis
J Adv Nurs 2010 Oct;66(10):2225-33
0309-2402
Aim. This paper is a report of a study to explicate the phenomenon of ginger compresses for people with osteoarthritis. Background. Osteoarthritis is claimed to be the leading cause of musculoskeletal pain and disability in Western society. Management ideally combines non-pharmacological strategies, including complementary therapies and pain-relieving medication. Ginger has been applied externally for over a thousand years in China to manage arthritis symptoms. Method. Husserlian phenomenological methodology was used and the data were collected in 2007.
Ten purposively selected adults who had suffered osteoarthritis for at least a year kept daily diaries and made drawings, and follow-up interviews and telephone conversations were conducted. Findings. Seven themes were identified in the data: (1) Meditative-like stillness and relaxation of thoughts; (2) Constant penetrating warmth throughout the body; (3) Positive change in outlook; (4) Increased energy and interest in the world; (5) Deeply relaxed state that progressed to a gradual shift in pain and increased interest in others; (6) Increased suppleness within the body and (7) More comfortable, flexible joint mobility. The essential experience of ginger compresses exposed the unique qualities of heat, stimulation, anti-inflammation and analgesia. Conclusion. Nurses could consider this therapy as part of a holistic treatment for people with osteoarthritis symptoms. Controlled research is needed with larger numbers of older people to explore further the effects of the ginger compress therapy.

Bibliographic indexed record
AMED}
[Accession Number] 0136462
[Authors] Therkleson T
[Title] Ginger compress therapy for adults with osteoarthritis
[Source] J Adv Nurs 2010 Oct;66(10):2225-33
[Major Descriptors] ZINGIBER OFFICINALE
[Major Descriptors] OSTEOARTHRITIS
[Major Descriptors] NURSING CARE
[Minor Descriptors] HUMANS
[Minor Descriptors] NURSING
[ISSN] 0309-2402
[Abstract Indicator] AB
[Month] c26jan11
[Abstract] Aim. This paper is a report of a study to explicate the phenomenon of ginger compresses for people with osteoarthritis. Background. Osteoarthritis is claimed to be the leading cause of musculoskeletal pain and disability in Western society. Management ideally combines non-pharmacological strategies, including complementary therapies and pain-relieving medication. Ginger has been applied externally for over a thousand years in China to manage arthritis symptoms. Method. Husserlian phenomenological methodology was used and the data were collected in 2007.
Ten purposively selected adults who had suffered osteoarthritis for at least a year kept daily diaries and made drawings, and follow-up interviews and telephone conversations were conducted. Findings. Seven themes were identified in the data: (1) Meditative-like stillness and relaxation of thoughts; (2) Constant penetrating warmth throughout the body; (3) Positive change in outlook; (4) Increased energy and interest in the world; (5) Deeply relaxed state that progressed to a gradual shift in pain and increased interest in others; (6) Increased suppleness within the body and (7) More comfortable, flexible joint mobility. The essential experience of ginger compresses exposed the unique qualities of heat, stimulation, anti-inflammation and analgesia. Conclusion. Nurses could consider this therapy as part of a holistic treatment for people with osteoarthritis symptoms. Controlled research is needed with larger numbers of older people to explore further the effects of the ginger compress therapy.

Figure 3. HIVE screen shot showing 'Extracted Concepts Cloud'

A screen capture was taken of the results returned by HIVE for addition to the spread sheet. It was vital that HIVE received the document with all the indexing removed as this would have influenced the results. A printout was produced for all articles with indexing and also of the HIVE result. Each manually indexed AMED record was then compared with the results returned by HIVE. Data was collated for both AMED and HIVE based on the four types of headings (major, minor, semi-controlled, publication type). As well as the total numbers for heading types, an analysis was made of the results using Excel and SPSS, working on the assumption that the AMED indexing should be correct, due to the expert knowledge and experience of the indexer. An analysis was made of each concept and type to ascertain if there was:

- a match between the human indexers and the automatic indexing tool
- errors identified in the output from the human indexers and/or the automatic indexing tool (HIVE)

A small selection of 5 records was timed from upload to processing. The average time taken from record to processing was 2 minutes 34 seconds. Human indexers take on average 10 minutes to manually index.

Data for al the outputs were input into SPSS. Descriptives, frequencies, and T-tests were carried out in SPSS in order to explore the differences between the human indexers and automatic indexing tool.

4. Findings

The data was assessed for normality using the Kolmogorov-Smirnov test [8]. A non-significant result with values of more than 0.5 was reported across all data indicating a normal distribution [8].

Table 2. A comparison of the total number of terms generated by both the human indexer and HIVE with figures for both major, minor and publication type terms.

	Human Indexer	**Automatic Tool (HIVE)**
Major terms generated (inc 13 terms in semi-controlled field which were all major terms)	370	350
Minor terms generated	344	22
Publication type terms generated	17	0
Totals number of terms generated	**731**	**372**

When considering the overall number of terms generated, the human and automatic indexing tool are comparable across major terms (H=370; A=350, **Table 2**). Differences occur with the generation of minor terms, and publication types. Though it is to be noted here that HIVE the automatic indexing tool cannot distinguish between major and minor terms and this categorisation had to be carried out by an indexer. The automatic indexing tool was ineffective in identifying any publication type terms.

The human indexer generated between 3 and 20 terms per record with a median average of 7 terms per record, mean of 7.31, and a mode of 5. In comparison, the automatic indexing tool generated between 0 and 10 terms per record with a median average of 2 terms, mean of 3.72 and a mode of 0 (**Table 3** and **Table 6**).

Table 3 Frequencies of terms generated per record on major terms, minor terms, and publication types.

Number of terms generated per record	**Frequency Human Indexer**				**Frequency Automatic Indexing Tool (HIVE)**			
	Total	**Major**	**Minor**	**Pub Types**	**Total**	**Major**	**Minor**	**Pub. types**
0	0	0	0	83	41	42	80	0
1	0	9	8	17	6	7	19	0
2	0	23	34	0	6	5	1	0
3	8	21	21	0	3	5	0	0
4	14	14	10	0	6	5	0	0
5	18	17	12	0	4	3	0	0
6	9	9	8	0	2	3	0	0
7	12	3	3	0	5	5	0	0
8	7	1	1	0	6	5	0	0
9	8	2	2	0	4	10	0	0
10	5	1	1	0	17	10	0	0
11	8	0	0	0	0	0	0	0
12	4	0	0	0	0	0	0	0
13	2	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0
15	2	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0
20	1	0	0	0	0	0	0	0
Totals	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The most number of matches between the human and automatic indexing tool were achieved across major terms. Few matches were achieved on minor terms (9 terms across 9 records). No matches were achieved on publication type terms, as the automatic indexing tool failed to generate publication type terms (**Table 4**).

Table 4. The number of terms generated that were matched by both the human indexer and HIVE.

Terms types	**Matches achieved between human indexer and automatic indexing tool out of a total of 100 records**
Major terms (inc semi-controlled field)	66 terms matched across 42 records 22 records matched on 1 term 16 record matched on 2 terms 4 records matched on 3 terms
Minor terms	9 terms matched across 9 records
Publication type terms	0 terms matched
Total number of matched terms across the sample	75 matches in total across 44 records

The data out-put was checked for errors by an medical indexer. The number of invalid terms generated across minor terms was comparable for both the human and automatic indexer. With regard to major terms, the human indexer was more effective generating fewer errors than the HIVE automatic indexing tool (H=18, A= 155 – **Table 5**)

Table 5. The number of invalid or erroneous terms generated by both the human indexer and the automatic indexing tool.

Invalid terms generated (errors)	Human Indexer	Automatic Tool
Invalid major terms generated	18 across 12 records	155 across 48 records
Invalid minor terms generated	6 across 6 records	3 across 3 records
Total invalid terms (errors) generated	24 across 18 records	158 across 51 records

An independent-samples t-test was conducted to compare the scores for human indexers and the automatic indexing tool across the total number of terms generated, the major terms, minor terms, and publication type terms, and the number of errors made across both major and minor terms generated (**Table 6**).

Table 6. Independent samples t-test

Sig. (2 tailed)	Category	N	Mean	Std. Deviation	Std Error Mean
Major terms Sig. 2 tailed .644	Human Automatic	100 100	3.70 3.50	1.925 3.873	.193 .387
Minor terms Sig. 2 tailed .000	Human Automatic	100 100	3.44 .22	1.935 .484	.193 .048
Publication Type Terms Sig. 2 tailed .000	Human Automatic	100 100	.17 .00	.378 .000	.038 .000
Total Terms Sig 2 tailed .000	Human Automatic	100 100	7.31 3.72	3.463 4.013	.346 .401
Errors Major Terms Sig. 2 tailed .000	Human Automatic	100 100	.18 1.55	.626 2.129	.063 .213
Errors Minor Terms Sig. 2 tailed .309	Human Automatic	100 100	.06 .03	.239 .171	.024 .017

There was no significant difference between the human and automatic indexers on the major terms generated. Human (M=3.70, SD= 1.925), and Automatic (M= 3.50, SD= 3.873 (**Table 6**).

There was a significant difference in the scores for the *total terms* generated, the *minor terms* generated, and the terms generated on *publication types* (Table 6):

Minor Terms*:* Human(M=3.44, SD = 1.935) and Automatic (M=.22, SD=.484)
Publication Types: Human (M= .17, SD=.378) and Automatic (M=.00, SD=.00)
Total Terms Generated: Human (M=7.31, SD=3.463) and Automatic (M=3.72, SD=4.013).

Where there was a significant difference, the magnitude of this difference was calculated [8]:

Totals: The magnitude of the difference in the means (mean difference=3.59, 95% CI: 4.65 to 4.65) was large (eta squared =0.188).
Minors*:* The magnitude of the difference in the means (mean difference = 3.22, 95% CI: 3.613 to 3.615) was moderate (eta squared = 0.568).
Publication types: The magnitude of the difference in the means (mean difference =.170, 95% CI: .244 to .245) was quite large (eta squared 0.093).

There was also a significant difference (p=.05 or less) between the human and automatic indexing tool on the errors made across major and minor terms (Table 6):
Major errors: Human - M=.18, SD=.626; Automatic - M= 1.55, SD=2.129
Minor errors: Human - M=.06, SD=.239; Automatic - M=.03, SD=.171 **Wrong. Not sig/**

The magnitude of the difference in the means on the major terms (mean difference = -1.370, 95% CI: -.932 to -.930) was large (eta squared =-0.238).
The magnitude of the difference in the means on the minor terms (mean difference = 0.30, 95% CI: .088 to .088) was small (eta squared = 0.005).

On reviewing the data output and matches across the data, errors in term generation were also identified. Errors were made by both human indexers and the automatic indexing tool. HIVE itself does not differentiate between major or minor terms, this categorisation had to be carried out by an indexer. Post categorisation, HIVE, the automatic indexing tool appears more effective with those categorised as minor terms, generating fewer invalid terms (i.e. errors), rather than those categorised as major terms.

The statistical analysis demonstrates that HIVE the automatic indexing tool is closest to the human indexer when generating major terms, though the automatic indexing tool generates invalid terms more frequently than the human indexer on major terms. These tests consider numerical output only, and the quality of output is important in indexing as poor quality indexing is likely to affect the user experience from an information retrieval perspective.

5. Discussion

All comparisons of the figures are made on the premise that the human indexer is the expert. Findings suggest that HIVE has the capacity to produce concepts when it has a title and abstract in English. Very few concepts are obtained when only the title is available. HIVE extracts approximately half as many concepts overall compared to that of human indexers. As results are presented in a concepts cloud it was the task of the researcher to interpret these. HIVE does not distinguish between major and minor concepts, check tags and publication types so the researcher carried out this categorisation for evaluation purposes. Concepts were considered to be either a match, unmatched, or errors. HIVE is more effective at generating major headings and the figures were more comparable with that of the human indexer (see **Table 3**). Out of 100 records 46 had a title only and the other 54 included an abstract. Without an abstract HIVE produced an almost zero extraction rate. Where an abstract was present HIVE was more effective and the results are more comparable to those of the human indexer. Of the 4 articles not in English, which were also title only, HIVE failed to extract any concepts. Matches between HIVE and the human indexer were low as only 66 major automatic concepts were evident in total (see **Table 4**). When compared to the major human concepts this resulted in a match of 17.84%. An example of where HIVE failed to produce any major concept matches can be seen in **Table 7**. Matches can be seem in **Table 8**.

Table 7. Record showing negative major concept matches.

Record	Human indexer	HIVE	Matches major concepts
ARTICLE 279	ORTHOMOLECULAR THERAPY PROFESSIONAL PRACTICE (major) HUMANS (minor)	Vitamin B Complex Plasma Health Care Health Urination Normality Self Care	No concept matches

Table 8. An example of where HIVE produced positive major concept matches.

Record	Human indexer	HIVE	Matches major concepts
ARTICLE 381 (in English, includes abstract)	WOUND HEALING WOUNDS HYPERBARIC OXYGENATION STEM CELL TRANSPLANTATION ORGONOMY	Wounds Wound Healing Hyperbaric Oxygenation	3 major concept matches

(major)
HUMANS
TRANSPLANTATION
THERAPY
(minor)
Review
(pub type)

18 errors in the human major indexing were exposed by HIVE making this an error rate of 4.86%. On review, these were missed concepts rather than incorrect ones as all other major terms were considered correct. **Table 9** shows a record with five human indexing errors.

Table 9. Human major concept errors.

Record	Human indexer	HIVE	Human errors Major concepts
ARTICLE 237	DISABLED DISASTERS THERAPY (major) HUMANS (minor)	Disasters Mental Health Health Theory of Mind Cities Bombs Hurricanes Survivors Set (Psychology) Population	5 human errors MENTAL HEALTH CITIES BOMBS HURRICANES SURVIVORS

HIVE generated 129 valid major concepts. Many of these had very little value in indexing terms. Minor concepts were mainly check tags and qualifiers and of these the match was 2.62%.The human indexer generated 6 errors. Only 3 HIVE minor errors were evident but although HIVE extracted check tags such as RATS it failed to produce the hierarchical main heading term ANIMALS which is needed for correct indexing as seen in **Table 10.**

Table 10. ANIMAL not extracted by HIVE as main heading.

Record	Human indexer	HIVE	Check tag with ng main heading
ARTICLE 34	lactic acid COLON HORDEUM (major) ANIMALS RATS (minor)	Rats Lactic Acid Fermentation Colon Diet Germination Barley C Fibers Bacteria Constipation	Concept with RATS should have ANIMAL as main heading

Article records that are not in English seem to be a problem for both human indexers and HIVE. Human indexers are able to translate these articles to some extent before indexing unlike HIVE. No publication types were retrieved by HIVE indicating that this seems to be a problem area. To test its ability to extract a publication type an additional record with (RCT) in the title was submitted to HIVE. It extracted Random Allocation for this. Failure to return any RCTs is an important finding as they are often used in data retrieval.

In addition to these findings HIVE also extracted the following concepts for analysis:

- More specific and less specific concepts in the same cloud i.e. plant genus and species
- Cancer was extracted but not converted to neoplasms which is the correct concept

- Adults age 50+ was extracted as adult which should be middle aged

HIVE seems to be unable to distinguish which is the correct more specific concept required in terms of indexing rules.

The overall findings are that HIVE is more comparable to the human indexer when extracting major concepts with regard to actual terms generated. The matches between these however were low in percentage terms. Several errors in the human major concepts were exposed by the automatic indexing. HIVE does not compare well to the human indexer when generating minor terms, or publication types. The minor concepts figures for HIVE had fewer errors and this was comparable to the human indexing ones.

Timings revealed that it takes on average 2 minutes 34 seconds to select a document without indexing and for HIVE to process it. Human indexers take on average 10 minutes to index a record. The time taken to analyse the extracted concepts was not calculated. The usage costs of the HIVE system are not known.

6. Conclusions

This study has evaluated if an automatic indexing tool such as HIVE can be a viable aid or replacement to human indexing for biomedical literature. Indications suggest that while HIVE can extract just over half the number of concepts there is considerable variation in match quality. HIVE's inability to extract a high enough percentage number of correct concepts raises doubts to its use as a replacement for human indexing. HIVE's ability to extract the correct check tag hierarchical combination cannot be relied on. The failure of HIVE to extract any publication types is an important result as these are essential in biomedical indexing. It would seem that the lack of an abstract for many articles reduces the amount of extracted concepts to a minimum by HIVE. This may be different for say, NLM indexing, but articles indexed for AMED very often have a title only. AMED could be enhanced by the inclusion of articles that have an abstract and title. The potential for HIVE may be more fully achieved in articles or documents with more text. While human indexers are reliant on the AMED thesaurus with its limited number of descriptors HIVE has the advantage of linking to MeSH with its extensive vocabulary to contribute to retrieval success. Human indexers do have access to other tools such as translation tools to assist them which HIVE does not. The process time for a document upload to HIVE is faster but this is not a saving if only a small percentage of correct concepts are extracted. Again the same principles apply to costs which would only be saved if the time taken by the human indexer were reduced. The human indexer would still have to spend time on scanning the document for concepts to assign adding to time and costs. They would have to analyse the extracted concepts to decide which ones were useful and which were incorrect. HIVE's performance level is closer to the human indexer when extracting major concepts but it does not work well elsewhere. It needs to extract a higher percentage of correct concepts to make any impact on quality, timings and costs. HIVE's ability to bring to the attention some missed concepts by the human indexer could be considered an aid to indexing but as the error rate is quite low its usefulness is questionable.

It would seem that the human indexer is more efficient at selecting the correct concepts overall. HIVE's ability to assist in subject cataloguing has succeeded to some extent. HIVE needs to be developed and increase its success rate in retrieving the correct concepts to achieve its potential. It can only be considered as a possible aid to AMED indexing and not a replacement at this present time. However, HIVE's developers are continuing to add new vocabularies to the system, and continuing their research with practitioner partners to more closely align HIVE vocabularies with partner requirements, improving metadata generation [9]. As technological advancements are made such tools will become increasingly important in offering practical solutions for metadata creators in real world situations.

7. Limitations

The basis of the results relied on the competence of the human expert in indexing articles to the correct standard for comparison with the automatic indexing. Analysis of major and minor terms was subjective and open to interpretation as HIVE does not make a distinction between the two.

8. Recommendations for further study

An independent analysis of a small number of records is recommended by a group of AMED users. Their opinions as to whether the human indexer or HIVE covers the concepts of the article most precisely would aid this study.

9. Acknowledgements

Gratitude is extended to HIVE for the use of the system, for permission to include screen shots within this study, and for their continued valuable and interesting research Grateful thanks are also extended to The British Library for the AMED data used within this research.

10. References

[1] Neveol, A., Rogozan, A. and Darmoni, S. *A recent advance in the automatic indexing of the biomedical literature.* Journal of Biomedical Informatics, 2009: 42: (5) 814-823.
[2] Vasuki, V. & Cohen, T. *Reflective random indexing for semi-automatic indexing of the biomedical literature.* Journal of Biomedical Informatics, 2010 43: (5), 694-700.
[3] Roberts, D. and Souter, C. *The automation of controlled vocabulary subject indexing of medical journal articles.* Aslib Proceedings (2000) 52:(10), 384-401.
[4] HIVE. *HIVE Web Interface*. Available from: http://hive.nescent.org:9090/home.html 2011.
[5] WC3 *World Wide Web Consortium* WC3. Available from: http://www.w3.org/ 2011.
[6] Demner-Fushman, D., Mork, J.G. Shooshan, A. and Aronson, A. R.. *UMLS content views appropriate for NLP processing of the biomedical literature vs. clinical text*. Journal of Biomedical Informatics, 2010: 43: (4) 587-594.
[7] Neveol, A., Shooshan, A., Humphrey, S.M., Mork, J.G. and Aronson, A. R. (2006). *Automatic indexing of online health resources for a French quality controlled gateway.* Information and Processing Management, 2006: 42: (3) 695-709.
[8] Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*. Maidenhead: Open University Press. 2010.
[9] Greenberg, J., Losee, R., Pérez Agüera, J.R Scherle, R. White, H, and Willis, C., *HIVE: Helping Interdisciplinary Vocabulary Engineering.* Bulletin of the American Society of Information Science and Technology, 2011: 37: (4) 23-26. Available from: http://www.asis.org/Bulletin/Apr-11/AprMay11_Greenberg_etAl.html.