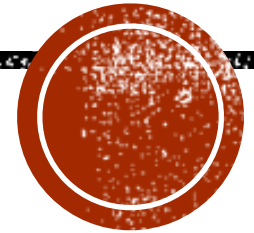
WHO NEEDS CONTROLLED VOCABULARIES WHEN WE HAVE KEYWORDS & FREE TEXT SEARCHING?

Drahomira Cupar, PhD, University of Zadar, Department of Information Sciences, Croatia

Ljiljana Poljak, Split University Library, Croatia



Department of information sciences

CONTENTS

- Introduction
- Previous work
- Research questions
- Methodology & Procedures
- Results & Discussion
- Conclusion
- References

INTRODUCTION

- Pilot study
- Small sample
- Portal Hrčak – Biomedicine and healthcare journals
- Author's Guidelines
- MeSH vs. Keywords
- Library catalog – subject headings

PREVIOUS WORK

- Gross, Taylor & Joudrey (2014) investigated importance of controlled vocabularies in keyword searching – one-third of results would be lost from hits in keyword search if there were no subject headings derived from a controlled vocabulary (library catalog; LCSH)
- Studies investigating overlap between author-assigned keywords, and controlled vocabularies to provide further insights on indexing and searching the literature
- The match between keywords and MeSH terms is mostly less than 50% (Ghazi-Mirsaeid, S. J., and F. Masoudi (2014), Roh (2012); Kim et al. (2013), Névéol et al. (2010)); and complete match in keywords vs MeSH terms around 15%.
- Beside similar result in overlap between keywords and MeSH terms, Kim et al. (2013) have noted increased number of papers where keywords and MeSH terms do not match.
- In larger scale, topic searching emphasize the importance of enhancing the MeSH thesaurus to support systematic resource discovery (Douyère et al. (2004), Kim, Yeganova & Wilbur (2016).

RESEARCH QUESTIONS

- ✓ to find out what type of instructions are given to authors regarding the creation of the keywords (sample of Author's Guidelines from 54 active journals from Biomedicine and Healthcare in Hrčak)
- ✓ to test how effective is keyword searching in Hrčak
- ✓ to compare MeSH terms and keywords in chosen articles
- ✓ to compare SH in catalogue and keywords for the sample articles

METHODOLOGY & PROCEDURES I

- Methods: content analysis and comparison.
- The research was done in two phases.
- In the **first phase** Guidelines for authors were analysed for 54 active journals in the field of Biomedicine and Healthcare included in Hrčak
- All instructions given to authors within Author's guidelines regarding the creation of the keywords for the selected journal were extracted and analysed in details

METHODOLOGY & PROCEDURES II

- In the **second phase,** research was done following 4 steps.

  Step 1. Identification of the topic/MeSH descriptor. Extraction of MeSH terms (synonyms and connected terms) and all keywords within chosen articles found by searching the topic in Hrčak.

  Step 2. Choosing the sample of articles from the journals with keywords made by using MeSH thesaurus. All data was collected into a table with journal title, article title, abstract and keywords.

  Step 3. Test searches using all variations of terms used by authors (e.g. synonyms and close synonyms) and comparison of the results in order to see how results are changed with different keywords.

  Step 4. Extraction of an exhaustive list of Main Heading (Descriptor) Terms and Entry terms from the MeSH thesaurus in order to compare:

  a) authors' keywords extracted from chosen articles and

  b) subject headings from library catalog assigned to the same articles.

1. PHASE: AUTHOR'S GUIDELINES ANALYSIS

- 54 active journals (in July)– field of Biomedicine and Healthcare (today 56 journals!)

Criteria	Results (N=54)	
Existing Author's guidelines/instructions	52 (in Word, PDF or html);	
	2 without (bulletins with news articles)	
Language of Author's guidelines	English = 26	Eng&Cro = 5
	Croatian = 21	

Regardles instructions, or language of the journal – all journals have abstracts and keywords in English and Croatian

1. PHASE: AUTHOR'S GUIDELINES ANALYSIS

<table>
<tr><th>Analysis of instructions regarding keywords</th><th colspan="3">Results (N = 52)</th></tr>
<tr><td>Who creates / assigns KW?</td><td>Authors = 47</td><td>with MeSH = 20
no system = 27</td><td>No instructions = 5 (in 4 - articles have KW)</td></tr>
<tr><td colspan="4">Detailed instructions consists of following:</td></tr>
<tr><td>With MeSH = 20</td><td colspan="3">No system required for KW creation =27</td></tr>
<tr><td>- authors (strictly) refer to; 'KW should be classified to' the medical Subject Headings (MeSH) = 9 (2 combinations with indexing)
- MeSH provided with link or detailed instructions = 8</td><td colspan="3">authors are instructed to put number of KW (3 – 5, 3 – 6) =14</td></tr>
<tr><td>KW 'assist (indexers) in (cross)indexing the article'; 'for indexing purposes'; 'for creating descriptors' = 9</td><td colspan="3">Instruction with combination of number of KW =13
- 4 'easy identification, and classification of content'
- 3 'main topic' + 1 'most important terms'
- 3 instruction says: 'must write keywords'
- 1 techical instruction ('avoid and; plural; general</td></tr>
</table>

EXAMPLE OF DETAILED INSTRUCTIONS

- "Below the abstract provide a list of 5 key terms that will be **useful for indexing or searching**. They should not be taken from the title of the manuscript but rather **reflect the content of the entire article and the field of study**. Use terms from the Medical Subject Headings (**MeSH**) list of the Index Medicus (**www.nlm.nih.gov/mesh**/), whenever possible. Key words should be listed in alphabetical order and separated by semicolons."

(Archives of Industrial Hygiene and Toxicology)

WHY CLEAR INSTRUCTIONS MATTER?

- Authors think of themselves as potential users of their own work, and provide access to their work
- Authors know the purpose of keywords in the system (and not only in individual journals)
- Authors can connect their work with others in the same area of expertise by using same terms/keywords found in the system
- Database (e.g. Hrčak) does not need 'external' indexing system/built-in thesaurus if, for example, MeSH thesaurus is used properly
- Quality instructions embeded in Author's guidelines could reduce number of misused keywords which are only 'pretending' to be from MeSH

2. PHASE. 4 STEPS

Step 1. Identification of the topic/MeSH descriptor. Extraction of MeSH terms (synonyms and connected terms) and all keywords within chosen articles found by searching the topic in Hrčak.

Chosen topic is: abortion, miscarriage.

1. MeSH does not have main term: abortion. It always is a combination of words (52 entry terms and descriptors).
2. Articles prescribing usage of MeSH in Author's guidelines often use keyword 'abortion'.
3. MeSH distinguishes humans from other living beeing. E.g. Abortion, veterinary.
4. Searching Hrčak for the topic of abortion in animal world, it is not possible to distinguish the difference between human and animals (without choosing particular journals)

Step 2. Choosing the sample of articles from the journals with keywords made by using of MeSH thesaurus. All data was collected into a table with journal title, article title, abstract and keywords.

		Keywords	MeSH
Acta clinica Croatica, Vol. 55. No. 2., 2016.	Safety of radiographic imaging in pregnancy	Pregnancy – radiography	
		Pregnancy – radiation effects	Radiation Effects
		Fetus – radiation effects	Radiation Effects
		Radiation injuries	Radiation Injuries
		Abortion, induced	Abortion, Induced
Veterinarski arhiv, Vol. 87 No. 2, 2017.	Identification of chlamydial strains causing abortions and pneumonia in sheep and goat flocks during trans Himalayan seasonal migration in the northern region of India	abortions	
		pneumonitis	Pneumonia
		chlamydiae	Chlamydia
		migratory sheep and goats	
		PCR	Polymerase Chain Reaction
Acta medico-historica Adriatica : AMHA, Vol. 11 No. 2, 2013.	An investigation into the ancient abortion laws: comparing ancient Persia with ancient Greece and Rome	Abortion	
		ancient Persia	
		human rights	Human Rights
Veterinarska stanica, Vol. 50 No. 4, 2019.	Identification and MLVA genotyping of Chlamydia abortus from abortion cases in small ruminants in Croatia	small ruminants	Ruminants
		Chlamydia abortus	
		chlamydial infection	
		abortion	
		Croatia	Croatia

Step 3. test searches: synonyms and comparison of the : are chang

Search results

Number of results: 47

[eng] **Comparison of Glycosylatic Missed Abortion**
Šerman, Ljiljana; Šerman, Alan; Lauc
Original scientific paper
Collegium antropologicum, Vol. 28, N
Full tekst: english / PDF (107 KB)
Hrčak ID: 4909

[eng] **Anxiety and Depression in**
Medjedović Marčinko, Vesna; Marčin
Short communication, Note
Collegium antropologicum, Vol. 35 su
Full tekst: english / PDF (56 KB)
Hrčak ID: 64081

[hrv] **RANI GUBITAK TRUDNOĆE NA**
[eng] **SPONTANEOUS ABORTION AF**
Finderle, Aleks; Petrović, Oleg

Collegium antropologicum, Vol. 28 No. 1, 2004

Original scientific paper

Comparison of Glycosylation Patterns of Placental Proteins Between Normal Pregnancy and Missed Abortion

Ljiljana Šerman

Abortion, Missed MeSH Descriptor Data 2020

Details | Qualifiers | MeSH Tree Structures | Concepts

MeSH Heading	Abortion, Missed
Tree Number(s)	C13.703.039.173
Unique ID	D000030
Annotation	check the tags FEMALE & PREGNANCY
Scope Note	The retention in the UTERUS of a dead FETUS two months or more after its DEATH.
NLM Classification #	WQ 225
Date Established	1966/01/01
Date of Entry	1999/01/01
Revision Date	2005/06/30

glycoconjugates might have an important role in placental development.

Keywords
glycoproteins; lectins; placenta; missed abortion

eyword: abortion
dvanced search;
l: keywords
ılts: 47
nals: 20
mple 1.
nal Collegium
opologicum is not
he list of results

Step 4. Extracted exhaustive list of Main Heading (Descriptor) Terms and Entry terms from the MeSH thesaurus compared to:

~~a) authors' keywords extracted from chosen articles and~~

~~b) subject headings from library catalog assigned to the same articles~~

Skipping the Step 4.

RESULTS & DISCUSSION

- Author's guidelines are short, not consistent and with ambigous instructions regarding keywords creation
- Even with recommendations for using MeSH, authors have insufficient information and/or training how to index with MeSH
- All journals (and articles) have keywords in Croatian and English
- Simple and Advance search options do not give satisfying results while searching and/or browsing in order to find all relevant resources
- Synonyms are not connected and results often come dispersed
- Lack of controlled vocabulary in Hrčak
- User is never sure system gave him/her all relevant results

CONCLUSIONS

- Databases need controlled vocabularies in order to provide users access to all relevant content/resources; regardless the terminology they use to access
- Hrčak might start by implementing existing vocabularies into their system which can be helpful
- In order to ensure usage of MeSH (or any other system), editors should check what keywords are supplied by authors in which case they will serve as connection between author who index / create keywords and users who search using MeSH
- Author's guidelines might be enriched with clear and detailed instructions on how to properly create keywords (in Medicine and Healthcare – with the help of MeSH)
- Small chunks of the job can be done with students of information sciences who are thaught to use tools for indexing (SH systems, thesaurus, classifictions)
- With the help of librarians, students of medicine are learning how to use MeSH during their education – so things could be much better soon ☺

REFERENCES

- Cimino J J. (1998) “Desiderata for controlled medical vocabularies in the twenty-first century”. Methods Inf Med;37(04/05): 394-403. DOI: 10.1055/s-0038-1634558
- Douyère, Magaly, et al. (2004) "Enhancing the MeSH thesaurus to retrieve French online health resources in a quality–controlled gateway." Health Information & Libraries Journal 21.4 :253-261. DOI: 10.1111/j.1471-1842.2004.00526.x
- Ghazi-Mirsaeid, S. J., and F. Masoudi. (2017) “A Comparative Review of the Compliance Rate of Abstracts Keywords of Iranian Dental Latin Journals Articles and their American Peers Indexed in PubMed with MeSH 2014." Journal of Research in Dental Sciences 14.1 DOI: 10.9790/3013-0704014347
- Gil–Leiva I, Alonso–Arroyo. (2007) “A. Keywords given by authors of scientific articles in database descriptors”. JASIST. Jun; 58(8): 1175-87. DOI: https://doi.org/10.1002/asi.20595
- Gross, Tina & Taylor, Arlene & Joudrey, Daniel. (2014) !Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching”. Cataloging & Classification Quarterly. DOI: 10.1080/01639374.2014.917447.
- Hartley J, Kostoff, R. (2003). How Useful are `Key words' in scientific journals? Journal of Information Science;29: 433-8. DOI: 10.1177/01655515030295008
- Kim, Sun, Lana Yeganova, and W. John Wilbur. (2016) "Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms." Bioinformatics 32.19: 3044-3046. DOI: 10.1093/bioinformatics/btw331
- Kim, Yun-Young & Park, Hye-Joo & Lee, Si-Woo & Yoo, Jong-Hyang. (2013). Comparison of Keywords of the Journal of Sasang Constitutional Medicine with MeSH Terms. Journal of Sasang Constitutional Medicine. 25. 34-42. DOI: 10.7730/JSCM.2013.25.1.34.
- Névéol, Aurélie et al. (2010)“Author keywords in biomedical journal articles.” AMIA ... Annual Symposium proceedings. AMIA Symposium vol. 2010 537-41.
- Roh, Jung-Suk. (2012). The Comparison of Keyword of Articles in Journal of the Korean Society of Physical Medicine with MeSH. Journal of the Korean Society of Physical Medicine. 7. 367-377. DOI: 10.13066/kspm.2012.7.3.367.

Thank you for your attention!

And once again
WELCOME TO #PubMet2019 in Zadar!

Contact:

Drahomira Cupar, dcupar@unizd.hr
Ljiljana Poljak, ljpoljak@svkst.hr