

Christopher Rock
cmrock2@illinois.edu
CS410 Fall 2020
15-Nov-2020

Original Proposed Topic

Survey of current medical literature which references Information Retrieval, Text Mining, Natural Language Processing, or a selection of other relevant terms. Review will attempt to summarize the quantity, focus, and areas of interest that are currently represented in this body of published work.

Introduction

MEDLINE (Medical Literature Analysis and Retrieval System Online) is a database of published works in the life sciences[\[1\]](#) ([#fn1](#)) . This database is accessible through the free PubMed website[\[2\]](#) ([#fn2](#)) , maintained by the US National Library of Medicine, part of the US National Institute of Health[\[3\]](#) ([#fn3](#)) . The focus of this technology review is to survey the available published papers found through PubMed which contain terms related to text information retrieval and analysis.

The following topics were chosen, picked from the weekly course overview slides of CS410 on Coursera, to use as search topics: “text mining”, “information retrieval”, “natural language processing”, “text clustering”, “rochio feedback”, “text categorization”, “sentiment analysis”, “BM25”, “opinion mining”, “tf-idf”, “zipf’s law”, and “topic modelling.” Results were loosely reviewed in an ad-hoc manner to allow general exploration of each terms results. Further information will be presented below about the quantity and dates of publishing for each search.

Ultimately, the results compiled from “topic modelling” were chosen for a more in-depth human-powered review. These results were chosen over some others for several practical reasons. First, the published papers were all written in the last 7 years, making this the most temporally novel set of documents. Additionally many listed papers were focussed on topics with clinical relevance which is of particular interest to this reviewer. Finally, the practical matter of the size of the corpus (77 documents) made it possible to do a 100% review of the abstracts, and a more thorough review of the full text, given the time period available for completion of this project.

Discussion

Overview of PubMed Results

Sorted by the number of articles matching the query.

Search Term	# Papers Found	Earliest Published Year	Notes
"natural language processing"	6413	1978	
"text mining"	2785	1999	
"information retrieval"	2443	1960	
"sentiment analysis"	392	2009	112 papers for "sentiment analysis" also included "tweet"/"twitter"
"zipf's law"	150	1941	In 1941 <i>Zipf's Law of Urban Concentration</i> was published in Science magazine. The next published article was in 1978 (one), and then regularly published papers started 1994 (3 that year).
"tf-idf"	98	2002	
"text categorization"	77	1994	
"topic modelling"	55	2013	
"BM25"	50	2001	
"opinion mining"	45	2010	
"text clustering"	28	2003	
"rocchio feedback"	5	2005	2010, 2011, 2017, 2019

Note: all terms were searched using "exact phrase matching", ie with double quotes around the term, due to the large number of non-relevant results returned when using the default "free text searching" without double quotes.

“topic modelling” Results Overview

I chose to focus on the results retrieved by the query “topic modelling”. This term did not appear before 2013, making it the most recently appearing term of those retrieved.

Search query: “topic modelling” Year	Number of Papers Retrieved
2013	1
2015	5
2016	7
2017	5
2018	10
2019	19
2020	16

Using the abstracts and the full text (when available) of these papers, each was reviewed to determine why “topic modelling” was referenced in the paper, whether the paper was healthcare related, and what data source was used for topic modelling.

Of the 55 papers, 33 had a focus on healthcare whereas 22 were primarily focussed on non-healthcare fields. Of the 55 papers, six were about Topic Modelling itself, with the primary focus on the mathematical or computational process and ways to improve or optimize that process. Not surprisingly 49 papers, the bulk of the results, used Topic Modelling as a tool for some other purpose.

Of 33 papers which were focussed on healthcare, 32 used topic modelling to review some data source. The most popular single use was to assist in the review of other published works, usually by using topic modelling to summarize or narrow a large group of papers into a more manageable list which was then reviewed by a human.

The next most popular use was to cull through tweets, sometimes correlating tweet topics with sentiment analysis and geographic data. Two papers were essentially the identical, correlating tweets about human papillomavirus (HPV) vaccination to population data about cervical cancer, which is caused by HPV.

Using topic modelling to distill survey results in a clinical setting was used three times, with the focus on healthcare being somewhat incidental rather than of primary concern. For instance, one paper surveyed a large number of veterinarians about how they felt on a large array of subjects, and used topic modelling to distill the results.

Only three papers which used electronic health record (EHR) text data were produced from this query. The raw data available in EHRs data has been rapidly growing, and this low use of EHR data is somewhat surprising. There are several barriers to EHR data use. Privacy concerns regarding identifiable EHR data results in EHR data silos which are not accessible to researchers. The EHR written records in the US has many audiences, with the primary being billing, legal defense in the case of possible lawsuit, and documentation demonstrating the meeting of institutional metrics being the main focus of medical documentation. Communication of the medical plan is unfortunately often delegated to an ancillary status. This may explain the surprising paucity of literature that relied on EHR data. Thus analysis of EHR text data is likely less useful as a tool to assist medical providers and researchers, and more a tool for administrators, billing staff, and regulators.

Topic Modelling Data Sources Used In Healthcare Related Papers

Data Source	Number of Papers
Papers/Articles	13
Twitter	6
Survey Responses	3
EHR Text Data	3
Conversations (transcribed)	2
Online Forums	2
Policy/Regulatory Documents as Source	1
Labeled Images	1

Conclusion

Text information retrieval and analysis has matured over the past 30 years from a primarily theoretical and work-intensive application to a regularly applied set of tools and principles applied to digital references. Within the field of medicine Topic Modelling shows slow adoption, but review of the published works offers some insight about areas which already demonstrate the promise of being able to take advantage of these techniques. Specifically, the distillation of a large search domain of articles seems to be the most obviously relevant use of topic modelling, as it saves time and energy on the part of a human reviewer from determining the topic areas discussed in a large corpus of published works.

Interestingly, had the technique of topic modelling been utilized for this technology review, it would have likely been a way of broadening the scope of this paper and discussing a larger number of search queries. As it stands, this paper is essentially just a manually performed “topic modelling” of the

papers discovered by the query “topic modelling” in PubMed.

Further analysis should include whether or not topic modelling is in fact used in other published works which are not discovered by a simple search for that term. Papers retrieved with a search for “Natural Language Processing”, “Information Retrieval”, and other queries might actually include topic modelling without explicit reference, and this could potentially be discovered with appropriate use of topic modelling of the result from those queries.

References

- 1: Howes C, Purver M, McCabe R. Using conversation topics for predicting therapy outcomes in schizophrenia. *Biomed Inform Insights*. 2013 Jul 15;6(Suppl 1):39–50. doi: 10.4137/BII.S11661. PMID: 23943658; PMCID: PMC3740209.
- 2: Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, Zou W. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*. 2015;16 Suppl 13(Suppl 13):S8. doi: 10.1186/1471-2105-16-S13-S8. Epub 2015 Sep 25. PMID: 26424364; PMCID: PMC4597325.
- 3: Archambeau C, Lakshminarayanan B, Bouchard G. Latent IBP Compound Dirichlet Allocation. *IEEE Trans Pattern Anal Mach Intell*. 2015 Feb;37(2):321–33. doi: 10.1109/TPAMI.2014.2313122. PMID: 26353244.
- 4: Charidimou A, Fox Z, Werring DJ, Song M. Mapping the landscape of cerebral amyloid angiopathy research: an informetric analysis perspective. *J Neurol Neurosurg Psychiatry*. 2016 Mar;87(3):252–9. doi: 10.1136/jnnp-2015-310690. Epub 2015 Jun 12. PMID: 26071214.
- 5: Shi Z, Hospedales TM, Xiang T. Bayesian Joint Modelling for Object Localisation in Weakly Labelled Images. *IEEE Trans Pattern Anal Mach Intell*. 2015 Oct;37(10):1959–72. doi: 10.1109/TPAMI.2015.2392769. PMID: 26340253.
- 6: Mo Y, Kontonatsios G, Ananiadou S. Supporting systematic reviews using LDA-based document representations. *Syst Rev*. 2015 Nov 26;4:172. doi: 10.1186/s13643-015-0117-0. PMID: 26612232; PMCID: PMC4662004.
- 7: Jenkins A, Croitoru A, Crooks AT, Stefanidis A. Crowdsourcing a Collective Sense of Place. *PLoS One*. 2016 Apr 6;11(4):e0152932. doi: 10.1371/journal.pone.0152932. PMID: 27050432; PMCID: PMC4822840.
- 8: Hashimoto K, Kontonatsios G, Miwa M, Ananiadou S. Topic detection using paragraph vectors to support active learning in systematic reviews. *J Biomed Inform*. 2016 Aug;62:59–65. doi: 10.1016/j.jbi.2016.06.001. Epub 2016 Jun 10. PMID: 27293211; PMCID: PMC4981645.

- 9: Munksgaard R, Demant J. Mixing politics and crime - The prevalence and decline of political discourse on the cryptomarket. *Int J Drug Policy*. 2016 Sep;35:77–83. doi: 10.1016/j.drugpo.2016.04.021. Epub 2016 Jun 20. PMID:
- 10: Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *J Med Internet Res*. 2016 Aug 29;18(8):e232. doi: 10.2196/jmir.6045. PMID: 27573910; PMCID: PMC5020315.
- 11: Andrei V, Arandjelović O. Complex temporal topic evolution modelling using the Kullback-Leibler divergence and the Bhattacharyya distance. *EURASIP J Bioinform Syst Biol*. 2016 Sep 29;2016(1):16. doi: 10.1186/s13637-016-0050-0. PMID: 27746813; PMCID: PMC5042987.
- 12: Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Griffiths KM. What's all the talk about? Topic modelling in a mental health Internet support group. *BMC Psychiatry*. 2016 Oct 28;16(1):367. doi: 10.1186/s12888-016-1073-5. PMID: 27793131; PMCID: PMC5084325.
- 13: Wong ZS, Chan WM, Wong EL, Chau PY, Tsui KL, Fung H. Uncovering Hidden Topics in Hong Kong Clinical Research Through Hospital Authority Convention Publications. *Stud Health Technol Inform*. 2017;245:624–628. PMID: 29295171.
- 14: Székely N, Vom Brocke J. What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. *PLoS One*. 2017 Apr 12;12(4):e0174807. doi: 10.1371/journal.pone.0174807. PMID: 28403158; PMCID: PMC5389611.
- 15: Hackl WO, Ganslandt T. Clinical Information Systems as the Backbone of a Complex Information Logistics Process: Findings from the Clinical Information Systems Perspective for 2016. *Yearb Med Inform*. 2017 Aug;26(1):103–109. doi: 10.15265/IY-2017-023. Epub 2017 Sep 11. PMID: 29063547; PMCID: PMC6239233.
- 16: Huntley SJ, Mahlberg M, Wiegand V, van Gennip Y, Yang H, Dean RS, Brennan ML. Analysing the opinions of UK veterinarians on practice-based research using corpus linguistic and mathematical methods. *Prev Vet Med*. 2018 Feb 1;150:60–69. doi: 10.1016/j.prevetmed.2017.11.020. Epub 2017 Dec 5. PMID: 29406085.
- 17: McInerney F, Doherty K, Bindoff A, Robinson A, Vickers J. How is palliative care understood in the context of dementia? Results from a massive open online course. *Palliat Med*. 2018 Mar;32(3):594–602. doi: 10.1177/0269216317743433. Epub 2017 Dec 13. PMID: 29235386; PMCID: PMC5851129.
- 18: Skeppstedt M, Kerren A, Stede M. Vaccine Hesitancy in Discussion Forums: Computer-Assisted Argument Mining with Topic Models. *Stud Health Technol Inform*. 2018;247:366–370. PMID: 29677984.

19: Ceré R, Egloff M. An illustrated approach to Soft Textual Cartography. *Appl Netw Sci.* 2018;3(1):27. doi: 10.1007/s41109-018-0087-y. Epub 2018 Aug 13. PMID: 30839805; PMCID: PMC6214316.

20: Wandy J, Zhu Y, van der Hooft JJJ, Daly R, Barrett MP, Rogers S. Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics.* 2018 Jan 15;34(2):317–318. doi: 10.1093/bioinformatics/btx582. PMID: 28968802; PMCID: PMC5860206.

21: Serôdio PM, McKee M, Stuckler D. Coca-Cola - a model of transparency in research partnerships? A network analysis of Coca-Cola's research funding (2008–2016). *Public Health Nutr.* 2018 Jun;21(9):1594–1607. doi: 10.1017/S136898001700307X. Epub 2018 Mar 21. PMID: 29560842; PMCID: PMC5962884.

22: van der Veer RA, Haverland M. Bread and butter or bread and circuses? Politicisation and the European Commission in the European Semester. *Eur Union Polit.* 2018 Sep;19(3):524–545. doi: 10.1177/1465116518769753. Epub 2018 Apr 19. PMID: 30220878; PMCID: PMC6108040.

23: Onan A. Biomedical Text Categorization Based on Ensemble Pruning and Optimized Topic Modelling. *Comput Math Methods Med.* 2018 Jul 22;2018:2497471. doi: 10.1155/2018/2497471. PMID: 30140300; PMCID: PMC6081524.

24: Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam MA, McNaught J, von Elm E, Nolan K, Ananiadou S. Prioritising references for systematic reviews with RobotAnalyst: A user study. *Res Synth Methods.* 2018 Sep;9(3):470–488. doi: 10.1002/jrsm.1311. Epub 2018 Jul 30. PMID: 29956486; PMCID: PMC6175382.

25: Chen X, Liu Z, Wei L, Yan J, Hao T, Ding R. A comparative quantitative study of utilizing artificial intelligence on electronic health records in the USA and China during 2008–2017. *BMC Med Inform Decis Mak.* 2018 Dec 7;18(Suppl 5):117. doi: 10.1186/s12911-018-0692-9. PMID: 30526643; PMCID: PMC6284279.

26: Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, Leask J, Dunn AG. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. *Hum Vaccin Immunother.* 2019;15(7–8):1488–1495. doi: 10.1080/21645515.2019.1596712. Epub 2019 Apr 12. PMID: 30978147; PMCID: PMC6746515.

27: Hossain M, Abdel-Aty M, Quddus MA, Muromachi Y, Sadeek SN. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accid Anal Prev.* 2019 Mar;124:66–84. doi: 10.1016/j.aap.2018.12.022. Epub 2019 Jan 8. PMID: 30634160.

28: Shamna P, Govindan VK, Abdul Nazeer KA. Content based medical image retrieval using topic and location model. *J Biomed Inform.* 2019 Mar;91:103112. doi: 10.1016/j.jbi.2019.103112. Epub 2019 Feb 6. PMID: 30738189.

- 29: Saha SK, Prakash A, Majumder M. "Similar query was answered earlier": processing of patient authored text for retrieving relevant contents from health discussion forum. *Health Inf Sci Syst*. 2019 Feb 18;7(1):4. doi: 10.1007/s13755-019-0067-3. PMID: 30863540; PMCID: PMC6384532.
- 30: K N, R J. A Robust User Sentiment Biterm Topic Mixture Model Based on User Aggregation Strategy to Avoid Data Sparsity for Short Text. *J Med Syst*. 2019 Mar 5;43(4):93. doi: 10.1007/s10916-019-1225-5. PMID: 30834466.
- 31: Chen X, Lun Y, Yan J, Hao T, Weng H. Discovering thematic change and evolution of utilizing social media for healthcare research. *BMC Med Inform Decis Mak*. 2019 Apr 9;19(Suppl 2):50. doi: 10.1186/s12911-019-0757-4. PMID: 30961624; PMCID: PMC6454597.
- 32: Drosatos G, Kaldoudi E. A probabilistic semantic analysis of eHealth scientific literature. *J Telemed Telecare*. 2020 Aug-Sep;26(7-8):414-432. doi: 10.1177/1357633X19846252. Epub 2019 May 12. Erratum in: *J Telemed Telecare*. 2020 Apr;26(3):NP1. PMID: 31081450.
- 33: Baudoin L, Sapinho D, Maddi A, Miotti L. Scientometric analysis of the term 'microbiota' in research publications (1999-2017): a second youth of a century- old concept. *FEMS Microbiol Lett*. 2019 Jun 1;366(12):fnz138. doi: 10.1093/femsle/fnz138. PMID: 31240301.
- 34: Valenti AP MS, Chita-Tegmark M PhD, Tickle-Degnen L PhD, OTR/L, FAOTA, Bock AW BS, Scheutz MJ PhD. Using topic modeling to infer the emotional state of people living with Parkinson's disease. *Assist Technol*. 2019 Jun 13:1-10. doi: 10.1080/10400435.2019.1623342. Epub ahead of print. PMID: 31194649; PMCID: PMC6908779.
- 35: Aufegger L, Bicknell C, Soane E, Ashrafian H, Darzi A. Understanding health management and safety decisions using signal processing and machine learning. *BMC Med Res Methodol*. 2019 Jun 13;19(1):121. doi: 10.1186/s12874-019-0756-2. PMID: 31196000; PMCID: PMC6567495.
- 36: Roque C, Lourenço Cardoso J, Connell T, Schermers G, Weber R. Topic analysis of Road safety inspections using latent dirichlet allocation: A case study of roadside safety in Irish main roads. *Accid Anal Prev*. 2019 Oct;131:336-349. doi: 10.1016/j.aap.2019.07.021. Epub 2019 Aug 1. PMID: 31377497.
- 37: Lau AYS, Staccini P; Section Editors for the IMIA Yearbook Section on Education and Consumer Health Informatics. Artificial Intelligence in Health: New Opportunities, Challenges, and Practical Implications. *Yearb Med Inform*. 2019 Aug;28(1):174-178. doi: 10.1055/s-0039-1677935. Epub 2019 Aug 16. PMID: 31419829; PMCID: PMC6697520.
- 38: Nguyen PH, Henkin R, Chen S, Andrienko N, Andrienko G, Thonnard O, Turkay C. VASABI: Hierarchical User Profiles for Interactive Visual User Behaviour Analytics. *IEEE Trans Vis Comput Graph*. 2020 Jan;26(1):77-86. doi: 10.1109/TVCG.2019.2934609. Epub 2019 Aug 20. PMID: 31442992.

- 39: Garcia-Rudolph A, Laxe S, Saurí J, Bernabeu Guitart M. Stroke Survivors on Twitter: Sentiment and Topic Analysis From a Gender Perspective. *J Med Internet Res*. 2019 Aug 26;21(8):e14077. doi: 10.2196/14077. PMID: 31452514; PMCID: PMC6732975.
- 40: Green MA, Widener M, Pollock FD, Pearce J. The evolution of Health & Place: Text mining papers published between 1995 and 2018. *Health Place*. 2020 Jan;61:102207. doi: 10.1016/j.healthplace.2019.102207. Epub 2019 Oct 3. PMID:
- 41: Neresini F, Crabu S, Di Buccio E. Tracking biomedicalization in the media: Public discourses on health and medicine in the UK and Italy, 1984–2017. *Soc Sci Med*. 2019 Dec;243:112621. doi: 10.1016/j.socscimed.2019.112621. Epub 2019 Oct 22. PMID: 31677575.
- 42: García D, Massucci FA, Mosca A, Ràfols I, Rodríguez A, Vassena R. Mapping research in assisted reproduction worldwide. *Reprod Biomed Online*. 2020 Jan;40(1):71–81. doi: 10.1016/j.rbmo.2019.10.013. Epub 2019 Oct 30. PMID:
- 43: Sommeria-Klein G, Zinger L, Coissac E, Iribar A, Schimann H, Taberlet P, Chave J. Latent Dirichlet Allocation reveals spatial and taxonomic structure in a DNA-based census of soil biodiversity from a tropical forest. *Mol Ecol Resour*. 2020 Mar;20(2):371–386. doi: 10.1111/1755–0998.13109. Epub 2019 Dec 2. PMID: 31650682.
- 44: Haynes E, Green J, Garside R, Kelly MP, Guell C. Gender and active travel: a qualitative data synthesis informed by machine learning. *Int J Behav Nutr Phys Act*. 2019 Dec 21;16(1):135. doi: 10.1186/s12966–019–0904–4. PMID: 31864372; PMCID: PMC6925863.
- 45: Chiudinelli L, Dagliati A, Tibollo V, Albasini S, Geifman N, Peek N, Holmes JH, Corsi F, Bellazzi R, Sacchi L. Mining post-surgical care processes in breast cancer patients. *Artif Intell Med*. 2020 May;105:101855. doi: 10.1016/j.artmed.2020.101855. Epub 2020 Apr 15. PMID: 32505422.
- 46: Cammel SA, De Vos MS, van Soest D, Hettne KM, Boer F, Steyerberg EW, Boosman H. How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (NLP) approach. *BMC Med Inform Decis Mak*. 2020 May 27;20(1):97. doi: 10.1186/s12911–020–1104–5. PMID: 32460734; PMCID: PMC7251822.
- 47: Crooks CJ. Supervised Bayesian Statistical Learning to Identify Prognostic Risk Factor Patterns from Population Data. *Stud Health Technol Inform*. 2020 Jun 16;270:422–426. doi: 10.3233/SHTI200195. PMID: 32570419.
- 48: Giunti G, Claes M, Dorronzoro Zubiete E, Rivera-Romero O, Gabarron E. Analysing Sentiment and Topics Related to Multiple Sclerosis on Twitter. *Stud Health Technol Inform*. 2020 Jun 16;270:911–915. doi: 10.3233/SHTI200294. PMID: 32570514.

- 49: Kar AK. What Affects Usage Satisfaction in Mobile Payments? Modelling User Generated Content to Develop the “Digital Service Usage Satisfaction Model”. *Inf Syst Front*. 2020 Jul 18;1–21. doi: 10.1007/s10796–020–10045–0. Epub ahead of print. PMID: 32837261; PMCID: PMC7368597.
- 50: Thiabaud A, Triulzi I, Orel E, Tal K, Keiser O. Social, Behavioral, and Cultural factors of HIV in Malawi: Semi-Automated Systematic Review. *J Med Internet Res*. 2020 Aug 14;22(8):e18747. doi: 10.2196/18747. PMID: 32795992; PMCID: PMC7455873.
- 51: Debnath R, Bardhan R. India nudges to contain COVID–19 pandemic: A reactive public policy analysis using machine-learning based topic modelling. *PLoS One*. 2020 Sep 11;15(9):e0238972. doi: 10.1371/journal.pone.0238972. PMID: 32915899; PMCID: PMC7485898.
- 52: Mora L, Wu X, Panori A. Mind the gap: Developments in autonomous driving research and the sustainability challenge. *J Clean Prod*. 2020 Dec 1;275:124087. doi: 10.1016/j.jclepro.2020.124087. Epub 2020 Sep 11. PMID: 32934442; PMCID: PMC7484706.
- 53: Wicke P, Bolognesi MM. Framing COVID–19: How we conceptualize and discuss the pandemic on Twitter. *PLoS One*. 2020 Sep 30;15(9):e0240010. doi: 10.1371/journal.pone.0240010. PMID: 32997720; PMCID: PMC7526906.
- 54: Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, Perra N. Collective Response to Media Coverage of the COVID–19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. *J Med Internet Res*. 2020 Oct 12;22(10):e21597. doi: 10.2196/21597. PMID: 32960775; PMCID: PMC7553788.
- 55: Debnath R, Darby S, Bardhan R, Mohaddes K, Sunikka-Blank M. Grounded reality meets machine learning: A deep-narrative analysis framework for energy policy research. *Energy Res Soc Sci*. 2020 Nov;69:101704. doi:10.1016/j.erss.2020.101704. PMID: 33145178; PMCID: PMC7563684.

Footnotes

1. <https://en.wikipedia.org/wiki/MEDLINE> (<https://en.wikipedia.org/wiki/MEDLINE>) ↩ (#fnref1)
2. <https://pubmed.ncbi.nlm.nih.gov/> (<https://pubmed.ncbi.nlm.nih.gov/>) ↩ (#fnref2)
3. <https://en.wikipedia.org/wiki/PubMed> (<https://en.wikipedia.org/wiki/PubMed>) ↩ (#fnref3)
4. <https://pubmed.ncbi.nlm.nih.gov/?term=“topic+modelling”> (<https://pubmed.ncbi.nlm.nih.gov/?term=%22topic+modelling%22>) ↩ (#fnref4)