BRIEF REPORT: "Where Do We Teach What?"

Finding Broad Concepts In The Medical School Curriculum

Joshua C. Denny, MD,¹ Jeffrey D. Smithers, MD,² Brian Armstrong, BS,³
Anderson Spickard III, MD, MS⁴

¹Department of Medicine, Vanderbilt School of Medicine, Nashville, TN, USA; ²Department of Medicine and Pediatrics, Good Samaritan Regional Medical Center, Phoenix, AZ, USA; ³Vanderbilt School of Medicine, Nashville, TN, USA; ⁴Department of Medicine, Department of Biomedical Informatics, Vanderbilt School of Medicine, Nashville, TN, USA.

BACKGROUND: Often, medical educators and students do not know where important concepts are taught and learned in medical school. Manual efforts to identify and track concepts covered across the curriculum are inaccurate and resource intensive.

OBJECTIVE: To test the ability of a web-based application called KnowledgeMap (KM) to automatically locate where broad biomedical concepts are covered in lecture documents in the Vanderbilt School of Medicine.

METHODS: In 2003, the authors derived a gold standard set of curriculum documents by ranking 383 lecture documents as high, medium, or low relevance in their coverage of 4 broad biomedical concepts: genetics, women's health, dermatology, and radiology. We compared the gold standard rankings to KM, an automated tool that generates a variable number of subconcepts for each broad concept to calculate a relevance score for each document. Receiver operating characteristic (ROC) curves and area-under-the-curve were derived for each ranking using varying relevance score cutoffs.

RESULTS: Receiver operating characteristic curve areas were acceptably high for each broad concept (range 0.74 to 0.98). At relevance scores that optimized sensitivity and specificity, 78% to 100% of highly relevant documents were identified. The best results were obtained with the application of 63 to 1437 subconcepts for a given broad concept. The search time was fast.

CONCLUSIONS: The KM tool capably and automatically locates the detailed coverage of broad concepts across medical school documents in real time. Use of KM or similar tools may prove useful for other medical schools to identify broad concepts in their curricula.

KEY WORDS: medical education; medical informatics; curriculum; abstracting and indexing.

DOI: 10.1111/j.1525-1497.2005.0203.x

J GEN INTERN MED 2005; 20:943-946.

"Where do we currently teach genetics?" is a typical question asked in medical schools. A dean may pose this question in preparation of the school's report to the Liaison Committee on Medical Education (LCME). A faculty member may ask this question to prepare to teach a new genetics lecture added to the medicine clerkship. A medical student clerk may ask this question to prepare for her postcall bedside presentation of her patient with complications of an inherited disease. Questions about where concepts and topics are covered in the curriculum are typically answered through lengthy collaborative meetings and numerous emails and phone calls. These labor- and communication-intensive approaches may

not be sufficiently detailed and sustainable to yield accurate, up-to-date answers.

Educators have turned to the web to display medical school information in order to improve the ability to manage, track, and coordinate curricular content. Some medical schools support a full text electronic curriculum so that teachers and learners have quick access to the current curriculum.^{1,2} Other medical schools take a step further by placing electronic curricular documents and images into a database.^{3,4} Each document is labeled with a manually derived set of keywords or titles to allow for searching and sharing of curricular content. Improvements on these efforts would obviate the need for manual identification and entry of selected concepts that describe a teaching session, and yet grant the ability to locate all concepts within each curricular document automatically. We have developed KnowledgeMap (KM) to provide these functions.

BACKGROUND

KnowledgeMap is a web-based application that uses a concept identifier to locate biomedical concepts automatically from search queries of medical education documents.⁵ Faculty members use the KM web application to upload presentations and lecture handouts in HTML, Microsoft® Word®, Adobe Acrobat® PDF, and Microsoft® PowerPoint® formats. As each document is uploaded, the KM concept identifier uses a rigorous algorithm⁶ to find all Unified Medical Language System (UMLS, a composite vocabulary containing more than one million terms)^{6,7} concepts located in the document. KM has performed well in identifying biomedical concepts in large sets of medical curriculum documents.⁶

KnowledgeMap can also search for broad concepts, termed metaconcepts, in the curriculum. Examples of metaconcepts include "genetics" or "women's health." After a user submits a metaconcept query, such as "genetics," the tool constructs a large list of related subconcepts (such as chromosome, point mutation, phenotype, etc.) using relationships defined in the UMLS. A user can decide how many subconcepts he or she wishes to search for by selecting from 6 expansion levels ranging from narrow (fewer subconcepts) to wide (more subconcepts). Once a user has selected the desired number of subconcepts and submits them for a search, KM returns the documents that contain the submitted subconcepts ranked by relevance. Analogous to a web search engine, the output lists all documents matching any subconcepts with

No outside sources of funding were used for this study.

Address correspondence and requests for reprints to Dr. Spickard, III: 7040, Medical Center East, Vanderbilt School of Medicine, Nashville, TN 37232 (e-mail: anderson.spickard-iii@vanderbilt.edu).

Received for publication April 7, 2005

and in revised form May 20, 2005

Accepted for publication May 20, 2005

FIGURE 1. A screen view of a document from a "Genetics" metaconcept search. Concept matches in the search are highlighted in gray. (Included in file *Spickard-MS#572.tif*.)

the most relevant documents listed first. The user may select a document to see the display of subconcepts located in that document as shown in Figure 1.

METHODS

We tested the ability of KM to locate metaconcepts within the Vanderbilt School of Medicine curriculum. The Institutional Review Board approved the study. First, we determined 4 metaconcepts of interest; next, we established gold standard rankings for a set of documents; and then we searched the document set for metaconcepts.

The authors identified 4 metaconcepts relevant for curriculum-type coverage queries: "genetics," "women's health," "dermatology," and "radiology." Two of these, "genetics" and "women's health," were actual concept queries posed by curriculum review committees at the Vanderbilt School of Medicine.

One author (A.S.) scored 383 documents from 19 first- and second-year medical school courses in order to establish the gold standard set of documents. The author has 12 years' experience in teaching in all 4 years of medical school and was not familiar with the concept coverage algorithm or the vocabulary sets. He scored each document as 1) having little or no information ("low" documents), 2) having a moderate amount

of information ("moderate" documents), or 3) having a large amount of information ("high" documents), which pertains to each of the 4 metaconcepts tested. References and "suggested readings" sections in documents were ignored when ranking. To validate the gold standard rankings, an expert in each field (genetics, women's health, dermatology, and radiology) scored a set of 10 documents. The percentage of exact agreement between each expert and author A.S. ranged from 90% to 100%; the overall κ between the author and 4 experts was 0.84 (calculated with PRAM⁸).

Using each expansion level of the KM metaconcept search (up to 40,000 subconcepts), we generated relevance scores for each of the 383 documents. We calculated sensitivity and specificity by iterating through all possible score "cutoffs" to identify "high," "moderate or high," and "low" documents among all documents scored. Documents with a score of zero were classified as "low." For the gold standard "high" set, we considered a true positive any document ranked "high" in the gold standard set and above a given score cutoff in KM; "moderate" and "low" documents were gold standard negatives. For the "moderate-high" set, we considered a true positive those documents above a given score cutoff that were ranked "high" or "moderate" in the gold standard set; "low" documents were the only gold standard negatives. We derived a receiver operating characteristic (ROC) curve from the calculated sensitivity

Table 1. Number of Subconcepts Generated and the Corresponding ROC Areas for each Metaconcept Search

Metaconcept	Metaconcept Alone*	UMLS Expansion						Optimal†	
		1	2	3	4	5	6	Sensitivity	Specificity
Genetics									
Number of subconcepts	1	355	421	1,437	21,182				
High GS docs (N=44)‡	0.56	0.96	0.96	0.98	0.94			0.96	0.94
Moderate-high GS docs (N=80)	0.58	0.94	0.94	0.95	0.90			0.85	0.92
Women's health									
Number of subconcepts	1	1	40	1,311	9,594	35,412			
High GS docs (N=18)	0.50	0.50	0.88	0.93	0.88	0.82		0.78	0.87
Moderate-high GS docs (N=43)	0.50	0.50	0.86	0.89	0.86	0.84		0.84	0.78
Dermatology									
Number of subconcepts	1	1	8	8	542	4,158	20,728		
High GS docs (N=4)	0.50	0.50	0.62	0.62	0.95	0.94	0.73	1.00	0.85
Moderate-high GS docs (N=32)	0.50	0.50	0.61	0.61	0.72	0.74	0.67	0.63	0.71
Radiology									
Number of subconcepts	1	63	387	999	12,636				
High GS docs (N=3)	0.65	0.97	0.95	0.62	0.69			1.00	0.96
Moderate-high GS docs (N=7)	0.55	0.75	0.78	0.71	0.72			0.57	0.96

*A search performed for a metaconcept without expansion of additional subconcepts.

†Optimal sensitivity and specificity is calculated from the best-performing expansion set as described in methods.

‡GS docs refers to the documents in the gold standard set ranked as "high" or "moderate-high." The corresponding row displays the ROC areas for each ranking. The highest ROC areas are in bold.

ROC, receiver operating characteristic; UMLS, Unified Medical Language System.

and specificity for each expansion set of each metaconcept. To compare the efficacy of each ranking, we calculated the area-under-the-curve for each ROC curve.⁹ The optimal sensitivity and specificity were determined by finding the northwest corner of each curve. For all statistical calculations, we used Stata[®] version 7 (College Station, TX) and Microsoft[®] Excel (Redmond, WA).

RESULTS

Table 1 shows the number of documents in the gold standard set found with high or moderate-high amount of information related to the 4 metaconcepts studied. In addition, the table displays the number of concepts yielded for each metaconcept with corresponding ROC areas using the UMLS vocabulary. A submission of a metaconcept alone without additional subconcepts yielded results not significantly different from random chance (expected ROC area of 0.5); the aggregate average ROC areas were 0.55 for high documents and 0.53 for moderate-high documents when using meta concept alone. The best ROC areas were found with an expansion set ranging from 63 to 1437 concepts in the high gold standard documents (average ROC area 0.93, standard error of the mean (SEM) 0.03), and 387 to 4158 concepts in the moderate-high gold standard documents (average ROC area 0.82, SEM 0.04). The average optimized sensitivity and specificity were 0.93 (SEM 0.06) and 0.90 (SEM 0.03) for high documents and 0.72 (SEM 0.08) and 0.84 (SEM 0.07) for the moderate-high set.

In general, the search time was fast. It took 5 to 20 seconds for a user to generate an expansion set of subconcepts and submit them to the corpus of documents. The search time to find all the locations of 350 subconcepts in the entire curriculum was 3 seconds to complete page delivery; an 8000-subconcept list took 17 seconds to complete page delivery.

DISCUSSION

This study presents a simple yet effective method for finding large-scale metaconcepts in medical school documents. The

application uses UMLS-defined relationships to generate a list of pertinent subconcepts to a metaconcept. Submission of these subconcepts to the curriculum database yields a substantially greater accuracy than submitting the metaconcept alone to a search across the curriculum. For example, using the optimum number of subconcepts ($n=1437$) for genetics, 42 (95%) of the high documents were found in the first 63 documents, and the top 19 documents were all "high" documents. Using just the search term "genetics" only identified 7 (16%) "high" documents.

The ROC areas found in this study are similar to the results reported for automatic classification of documents and the use of clinical tests. Wilcox and Hripesak¹⁰ reported ROC areas of 0.8 to 0.9 for inductive learning algorithm classification of discharge summaries and radiology reports. In clinical practice, ROC areas from 0.70 to 0.95 are common. The ROC area for B-type natriuretic peptide for diagnosing congestive heart failure as a cause for shortness of breath is 0.91¹¹; the ROC area for PSA for prostate cancer is 0.72 to 0.86, depending on subject age.¹²

Owing to limitations, these data should be interpreted with caution. KM yielded false-positive results caused by properties inherent in the document. Several short documents ranked highly by KM because of a high proportion of target concepts were ranked low by the gold standard because they were "only one paragraph" or "too short." Other false-positive ratings resulted from finding target concepts in "suggested reading" or misidentification of abbreviations. With regard to the "radiology" metaconcept, the most common misidentified abbreviations were "rt" (meaning "right" but interpreted as "radiotherapy"), "brs" (meaning "branches" but interpreted as "basic radiographic system"), and "CT" (meaning "connective tissue" in some documents but interpreted as "computed tomography"). Finally, the algorithm was tested on medical educational material and may not apply to the ranking of abstracts, clinical documents, or research documents.

Quick and accurate identification of concepts within the medical school curriculum has potential advantages for

students, teachers, and administrators. At our institution, students can readily find all concepts located in their class notes. Recently, using KM, teachers and administrators found a paucity of coordinated curriculum in the area of genetics, leading to the formation of a new free-standing course in genetics for second-year medical students. Finally, using KM to detail where important concepts are covered in the curriculum, curriculum task force members and administration were quickly able to produce detailed reports for a recent site visit by the Liaison Committee on Medical Education. Future studies will address the deployment of KM to other institutions to assist in identifying medical concepts in local curricula.

REFERENCES

1. **University of Florida College of Medicine.** Available at: <http://medinfo.ufl.edu/>. Accessed April 6, 2005.
2. **McGowan J, Abrams M, Frank M, Bangert M.** Creating a virtual community of learning predicated on medical student learning styles. *Proc AMIA Symp.* 2003;435-44.
3. **University of Pittsburg School of Medicine Navigator.** Available at: <http://labedutech.medschool.pitt.edu>. Accessed April 4, 2005.
4. **Lee MY, Albright SA, Alkasab T, Damassa DA, Wang PJ, Eaton EK.** Tufts health sciences database: lessons, issues, and opportunities. *Acad Med.* 2003;78:254-64.
5. **Denny JC, Irani PR, Wehbe FH, Smithers JD, Spickard A III.** The KnowledgeMap Project: development of a concept-based medical school curriculum database. *Proc AMIA Symp.* 2003;195-9.
6. **Denny JC, Smithers JD, Miller RA, Spickard A III.** "Understanding" medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc.* 2003;10:351-62.
7. **Unified Medical Language System.** Available at: <http://www.nlm.nih.gov/research/umls/>. Accessed December 10, 2004.
8. **PRAM: a Program for Reliability Assessment with Multiple Coders.** Available at: <http://www.geocities.com/skymegsoftware/pram.html>. Accessed January 20, 2005.
9. **Hanley JA, McNeil BJ.** The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29-36.
10. **Wilcox AB, Hripcsak G.** The role of domain knowledge in automating medical text report classification. *J Am Med Inform Assoc.* 2003;10:330-8.
11. **Maisel AS, Krishnaswamy P, Nowak RM, et al.** Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. *N Engl J Med.* 2002;347:161-7.
12. **Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM.** Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. *N Engl J Med.* 2003;349:335-42.