How similar are sources of content for courses?

Joseph A. Price III, Ph.D., Dept. of Pathology, OSU-COM.

ABSTRACT

For course content selection, to establish a measure of consensus on topics of relevance, and to compare sources of content, a novel quantitative method was developed and applied to a medical school course. Sources included texts, board review books, and a listing of topics currently taught in the course. Data mining of topics from sources developed data as binary encoded lists of what was present (among 350 topics) before two classical similarity measures were used to compute relatedness in pairwise comparisons of 13 sources. Relatedness was not always as expected. Total topics included ranged from highest in the course handouts, to lower in all other sources. This quantitative indication is opposed to reliance on subjective impressions, and can help faculty make better choices on content topics to include in a course and to compare texts.

INTRODUCTION

When choosing content for courses, the responsible person(s) could choose to teach from a standard text in the field. However, many courses address specialized audiences and particular curricular needs, so the task becomes how to choose material from various technical sources, and to what extent is there consensus, which in turn means asking which sources would be most useful, and how different are they from each other in content? For large content domains, a subjective impression may not be enough.

METHODS

A binary matrix was created by data mining leading sources of course content for the Musculoskeletal & Integument course at OSU-CHS for occurrence of topics. Sources included texts, board review books, and a listing of topics currently taught in the course. Rows were topics, a lexicon of topics found, columns were sources, and the entries a 1 or 0, indicating presence or absence. The relatedness of the descriptive binary lists in pairwise comparisons was calculated using two classical similarity measures reflecting two possible philosophies of content choice: Simple matching similarity coefficient (Sokal and Michener 1958) and the Legendre index (Gower 1985, Gower and Legendre 1986, Ellis, Furner-Hines et al. 1993). These are closely related. The Legendre index is a quotient of total exact agreement on presence of a topic (1,1) out of the entire list, while the Simple index is a quotient of the sum of exact agreement to include or exclude (1,1 or 0,0) out of the entire list. A lexicon list of 350 topics was generated, used for 78 pairs of lists to compare by the two measures of similarity, done with a spreadsheet and MSvba.

RESULTS

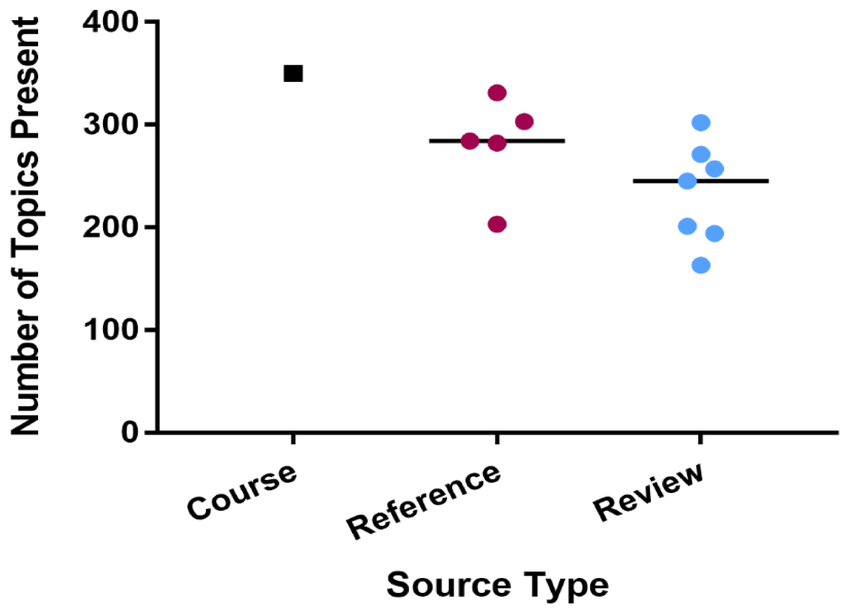

Figure 1. Number of Topics found in the source by type of text compared to the course. Course = 350, Reference Texts (Median = 284), Review Texts (Median = 245).

	2016 Course Topics	Cecil Medicine	Robbins	First Aid	Rubin's Clin. Path.	Robbins Basic	First Aid 2016	Rapid Review	Crush	Essential Rubin	USLME Secrets	Step-up	BRS Pathology
2016 Course		0.39	0.31	0.35	0.29	0.30	0.30	0.29	0.26	0.22	0.25	0.23	0.18
Cecil Medicine	0.58		0.36	0.39	0.34	0.32	0.36	0.33	0.34	0.26	0.29	0.27	0.22
Robbins	0.47	0.63		0.31	0.38	0.35	0.30	0.28	0.28	0.29	0.22	0.21	0.24
First Aid	0.55	0.68	0.56		0.30	0.30	0.42	0.29	0.34	0.25	0.31	0.28	0.20
Rubin's Clin. Path.	0.46	0.61	0.73	0.59		0.34	0.28	0.28	0.27	0.31	0.22	0.21	0.22
Robbins Basic	0.48	0.57	0.69	0.58	0.69		0.27	0.27	0.25	0.26	0.21	0.19	0.22
First Aid 2016	0.50	0.68	0.60	0.84	0.60	0.57		0.28	0.32	0.24	0.27	0.26	0.21
Rapid Review	0.51	0.64	0.60	0.60	0.62	0.60	0.64		0.27	0.22	0.20	0.19	0.21
Crush	0.47	0.67	0.60	0.72	0.62	0.58	0.75	0.67		0.22	0.23	0.24	0.19
Essential Rubin	0.46	0.60	0.71	0.62	0.77	0.68	0.65	0.65	0.66		0.18	0.19	0.19
USLME Secrets	0.53	0.65	0.57	0.74	0.59	0.59	0.72	0.61	0.69	0.66		0.20	0.16
Step-up	0.49	0.63	0.56	0.71	0.59	0.56	0.70	0.60	0.72	0.69	0.72		0.14
BRS Pathology	0.46	0.58	0.66	0.60	0.67	0.67	0.67	0.69	0.67	0.75	0.69	0.66	

Table 1. Similarity indices for pairwise comparisons of source material. The lexicon of topics was compared pairwise and the similarity measures calculated by the methods: Simple matching similarity coefficient, and the Legendre index.

DISCUSSION

A novel quantitative approach was developed to quantitatively compare relatedness of sources of course content using classical similarity measures. The course as was taught had the most topics followed by reference texts. Board review texts results were variable. There are several possible interpretations of this pattern, including anecdotal observations that lecturers may be over including topics into their presentations, or current texts are out of date.

The relative consistence (relatedness) among sources, medical texts with related goals, was expected in a science area, and supports the use of this methodology. It might be less in a more subjective knowledge domain. One might infer which texts are more inclusive references. Because the data mining and analysis is for topics, course detail at finer granularity is not addressed, although the method is applicable to a wide variety of applications. Consensus among expert sources is only one tool in deciding what should be taught, or which materials are best for study preparing for national level 1 board exams. This quantitative approach is opposed to reliance on subjective impressions, and can help faculty make better choices on content topics to include in a course, and to compare texts.

LITERATURE CITED

Ellis, D., Furner-Hines, J., Willett, P. (1993) Measuring the degree of similarity between objects in text retrieval systems. Perspectives in Information Management 3(2): Pg. 128-149

Gower, J. C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. Journal of Classification 3: Pg. 5-48.

Sokal, R. R. & Michener, C. D. (1958) A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: Pg. 1409-1438.

AUTHOR

Joseph A. Price III, Ph.D.
Professor of Pathology, OSU-COM
918-561-1441
joseph.price@okstate.edu