APPENDIX

A. GRID SEARCH HYPERPARAMETERS FOR AUTOMATED TAGGING

Table 4: Hyperparameters and Value Ranges for Logistic Regression Classifiers

Parameter	Parameter Description	Range	Best
Penalty	Normalization to use in penalization	[l1, l2]	l2
C	Inverse regularization strength	0.1–30	21
Intercept Scaling	When including intercept, the scaling of the intercept term	0.1–1.0	0.5
Class Weight	*Uniform* counts all classes of labels equivalently; *balanced* adjusts classes based on their frequencies	[uniform, balanced]	balanced
Prob. Threshold	Probability threshold to determine that a tag is assigned to a resource	0.1–0.9	0.4
Title Weight	Include title *title weight* times in resource document	1	1
Subtitle Weight	Include subtitle *subtitle weight* times in resource document	[0, 1]	0
Description Weight	Include description *description weight* times in resource document	1	1
Syllabus Weight	Include syllabus *syllabus weight* in resource document	[0, 1]	1
Stop Words	Stop word collection removal	[none, English]	English
Max. Document Frequency Threshold	Remove terms that occur in more than this proportion of resource documents	0.4–1.0	0.7
Min. Document Frequency Threshold	Remove terms that occur in less than this number of resource documents	3–10	8
N-gram Range	Include n-grams in vectorizations	$(1,1)$ $(1,2)$ $(1,3)$	$(1,2)$
NMF	Reduce vectorization with non-negative matrix factorization	[true, false]	false

Table 5: Hyperparameters and Value Ranges for TF–IDF Vector Comparison

Parameter	Parameter Description	Range	Best
Title Weight	Include title *title weight* times in resource document	2–4	4
Subtitle Weight	Include subtitle *subtitle weight* times in resource document	[1, 2]	1
Description Weight	Include description *description weight* times in resource document	[1, 2]	1
Syllabus Weight	Include syllabus *syllabus weight* in resource document	1	1
Stop Words	Stop word collection removal	[English]	English
Max. Document Frequency Threshold	Remove terms that occur in more than this proportion of resource documents	0.6–0.9	0.6
Min. Document Frequency Threshold	Remove terms that occur in less than this number of resource documents	2–5	2
N-gram Range	Include n-grams in vectorizations	$(1,2)$ $(1,3)$ $(1,4)$	$(1,2)$
Sublinear Term Frequency	Make term frequency equal to $1 + \log(\text{tf})$	[true, false]	true
Normalization	l1 is the Manhattan Distance, l2 is the Euclidean norm	[l1, l2]	l2
Inverse Document Frequency	Multiply term frequency by inverse document frequency	[true, false]	false
Similarity Aggregation	Use this aggregation function to assign resource–tag pairs	[mean, max, median, min]	max
Rank	Select up to *rank* value to assign tags to a resource	5–10	10