PilotCity Math Clinic Team

Final Presentation

Team Members:
Madison Hobbs (PM)
Aanya Alwani
Selasi Adedze
Dominique Macias
Evan Liang

Faculty Advisor: Dr. Talithia Williams

Liaison: Derick Lee
CEO & Founder, PilotCity

Overview

Two Stakeholders

- PilotCity
- Institute of Educational Sciences (IES)

Project Goals

- Automate & Digitize PilotCity Programming
- Extract Educational Information from Uncurated Data Sources

Stakeholders

PilotCity

- Small startup in San Leandro, CA
- Connects high school classrooms with local employers.

ies INSTITUTE OF EDUCATION SCIENCES

- U.S. Department of Education
- Focuses on extracting educational information from uncurated data sources.

PilotCity Problem Statement

The primary goal of the clinic is to create a **proof of concept** solution to **scale** the current program.

This clinic project will extract **educational priorities** from partnering employer and classroom data, develop an **engine** to **automate** program delivery, and engineer a **web-based application** for students, teachers, and employers.

IES Problem Statement

Education evaluations depend on several sources of educational data from websites from public, private, and charter schools, school handbooks among others.

This clinic is tasked to explore, develop, and refine **algorithmic approaches** and software to **extract insight** from these publically-available, unstructured educational data sources.

Objectives

- Build a web-based application that students, teachers, and employers interface with
- Design and build a recommender system for participants of PilotCity's programming.
- Use topic modeling to extract keywords from teachers' course syllabi to assess curriculum priorities.

Objectives

- **Build a web-based application that students, teachers, and employers interface with**
- Design and build a recommender system for participants of PilotCity's programming.
- Use topic modeling to extract keywords from teachers' course syllabi to assess curriculum priorities.

User Interviews

OAKLAND UNIFIED SCHOOL DISTRICT

Community Schools, Thriving Students

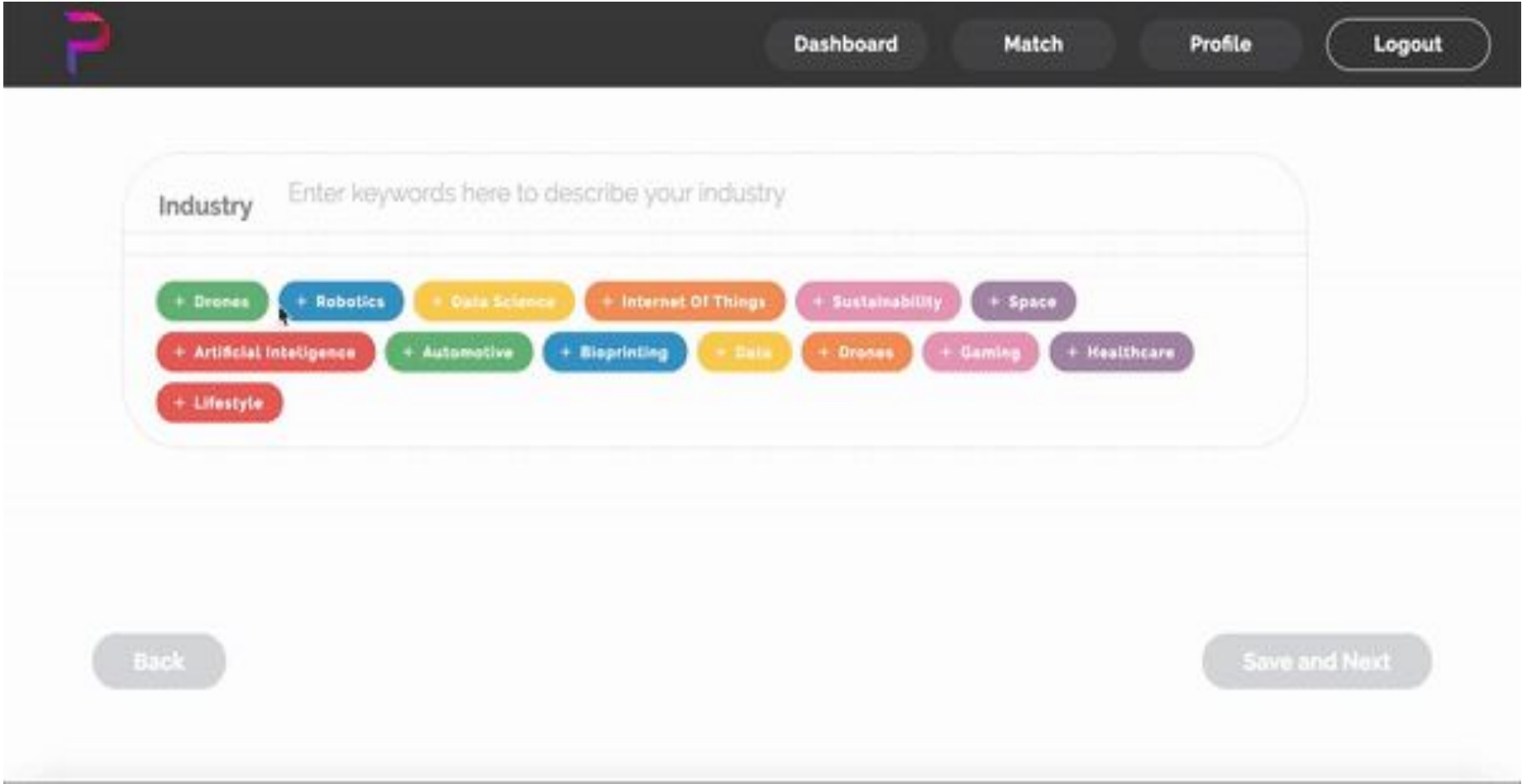
seeed studio

The IoT Hardware Enabler

ALAMEDA COUNTY OFFICE OF EDUCATION

SONY

WebDev

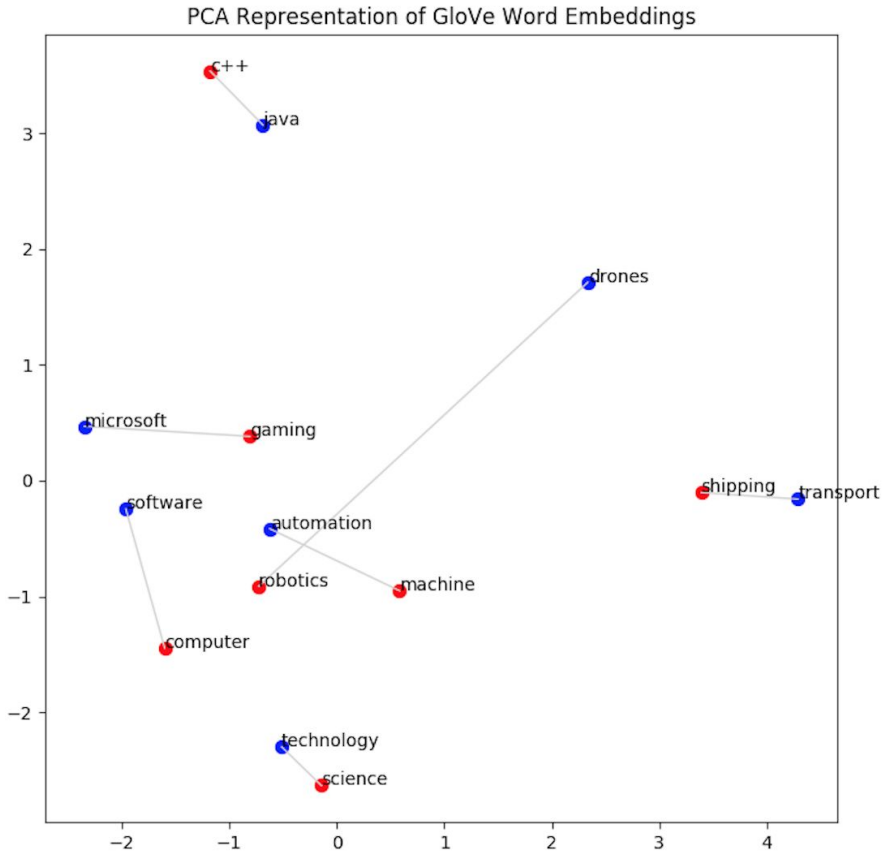

Dashboard
Match
Profile
Logout
Industry
Enter keywords here to describe your industry
+ Drones
+ Robotics
+ Data Science
+ Internet Of Things
+ Sustainability
+ Space
+ Artificial Inteligence
+ Automotive
+ Bioprinting
+ Data
+ Drones
+ Gaming
+ Healthcare
+ Lifestyle
Back
Save and Next

Objectives

- Build a web-based application that students, teachers, and employers interface with
- **Design and build a recommender system for participants of PilotCity's programming.**
- Use topic modeling to extract keywords from teachers' course syllabi to assess curriculum priorities.

GloVe Model

- Unsupervised learning algorithm producing vector representations of words.
- Allows for measurement of semantic similarity between words.
- Trained on 2014 Wikipedia data

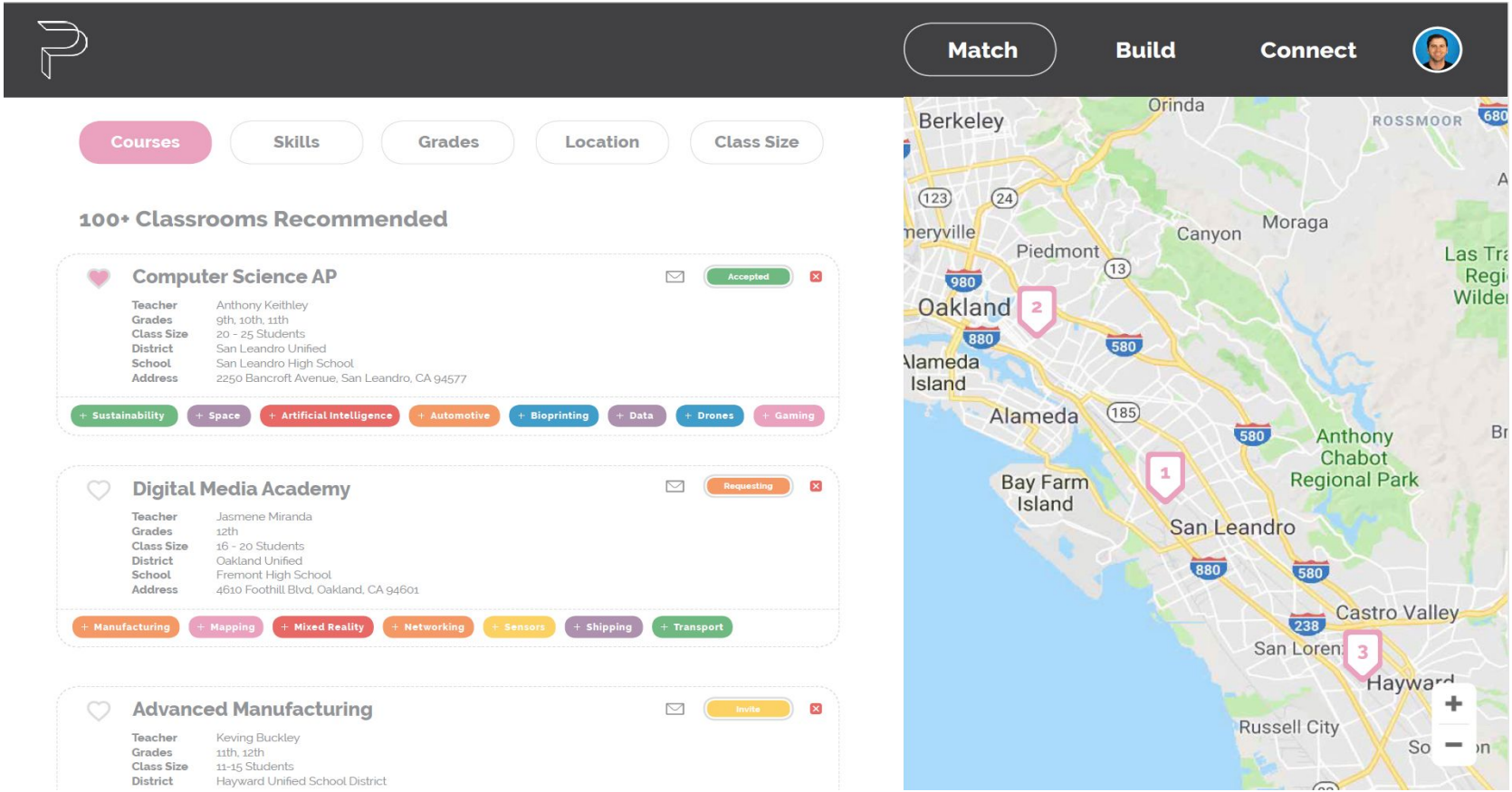


Inputs to recommender system

Classroom
Course name
Industry preference of teacher
Tools, technologies, and skills taught
City

Employer
Industry
Service
Product
Vision for project
City

Recommender System



Objectives

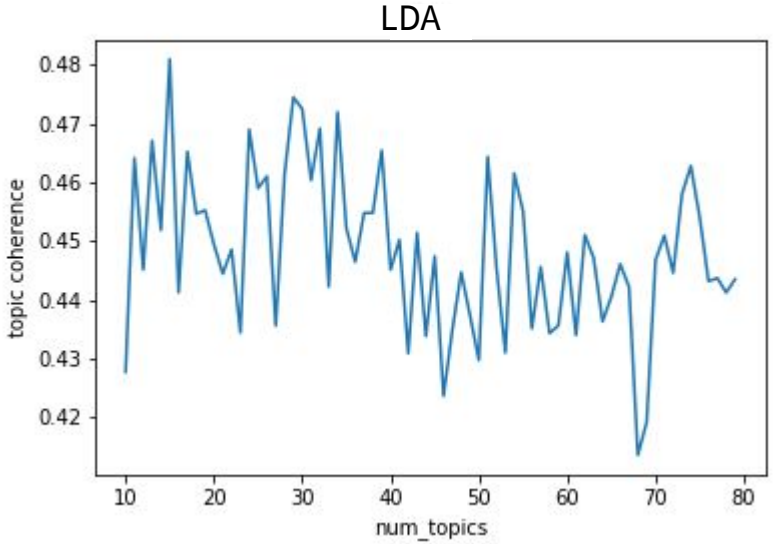
- Build a web-based application that students, teachers, and employers interface with
- Design and build a recommender system for participants of PilotCity's programming.
- **Use topic modeling to extract keywords from teachers' course syllabi to assess curriculum priorities.**

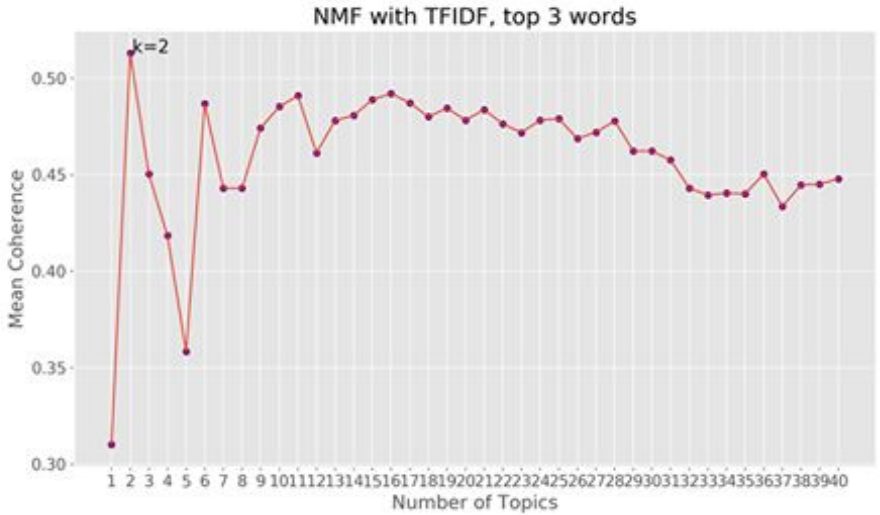
Topic Modeling - What is Topic Modeling?

- A **topic model** is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents.
- Popular techniques include Latent Dirichlet Allocation (**LDA**) and Non-negative Matrix Factorization (**NMF**).

Choosing the number of topics

- Topic coherence - how similar are the top words representing each topic?
- Use GloVe model to compute similarity
- Plot topic coherence vs. number of topics.

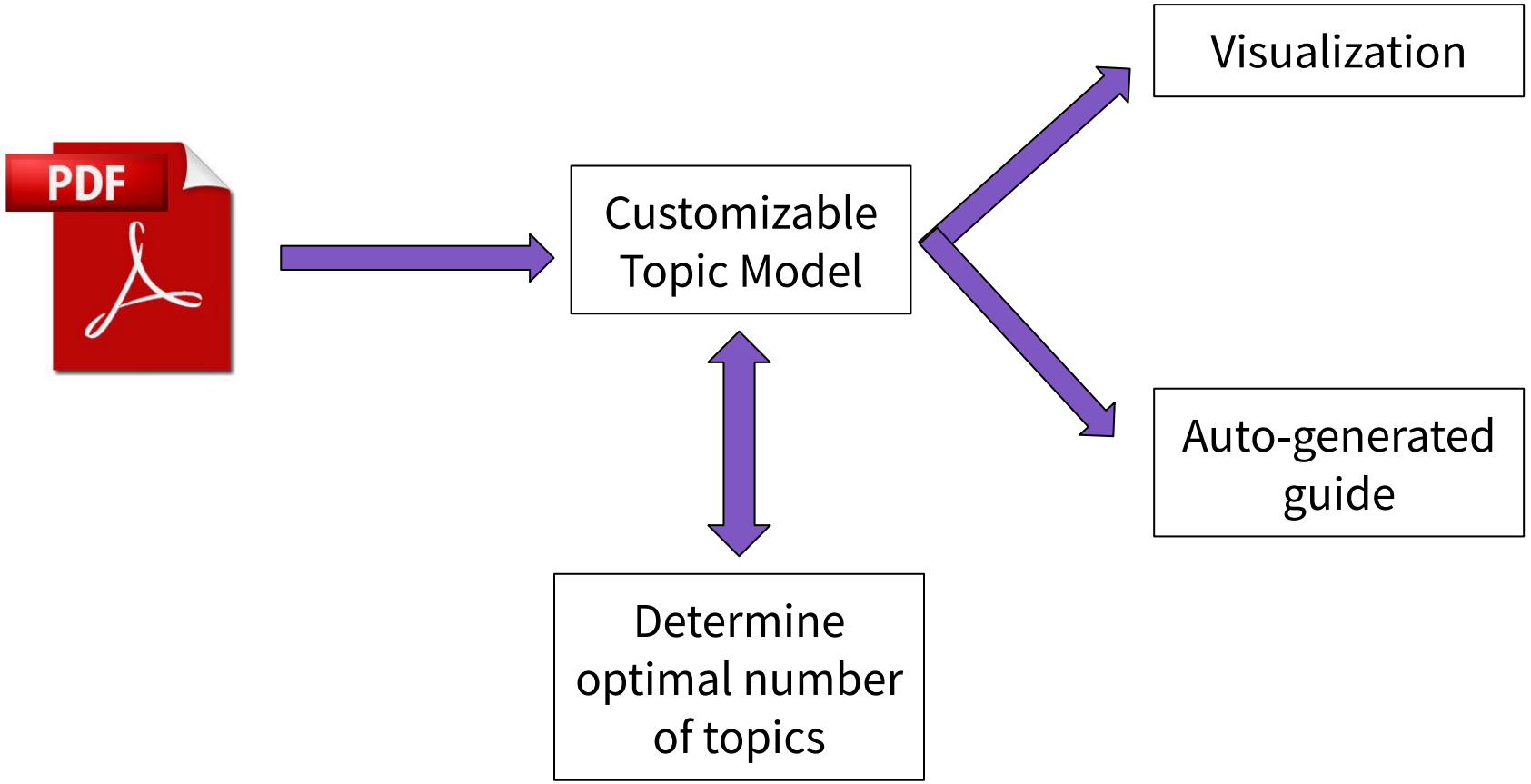




Example

Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12	Topic # 13	Topic # 14	Topic # 15	Topic # 16
independent	fitness	math	theater	basketball	web	network	wine	swimming	painting	yoga
project	exercise	linear	musical	game	design	configure	winery	polo	color	relaxation
study	training	algebra	music	intercollegiate	site	cisco	grape	swim	drawing	breathing
end	strength	solve	production	team	create	routing	tasting	water	design	strength
noted	endurance	exponential	performance	competition	data	security	vineyard	backstroke	studio	flexibility
semester	aerobic	quadratic	ensemble	shooting	use	configuration	sensory	training	art	yo
develop	walking	rational	acting	participation	office	operating	production	stroke	critique	balance
form	kin	logarithmic	vocal	flag	lab	server	viticulture	butterfly	value	kin
instructor	muscular	intermediate	jazz	passing	page	lan	fermentation	kin	lighting	mat
lab	heart	learning	stage	football	user	wireless	world	competitive	composition	increase

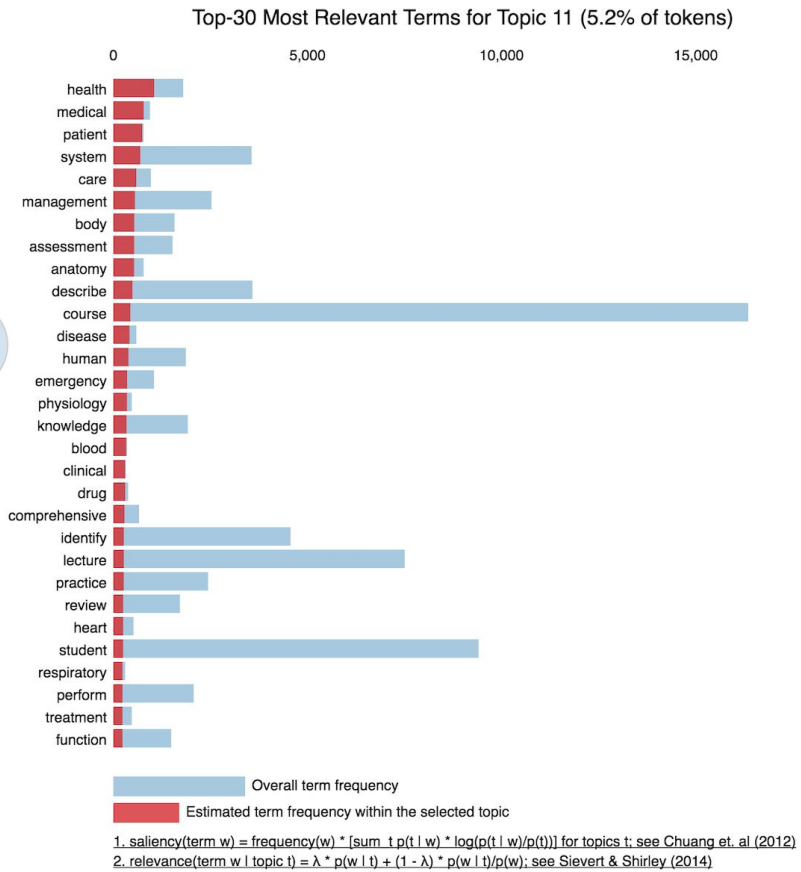
What we did


PDF
Customizable Topic Model
Visualization
Auto-generated guide
Determine optimal number of topics

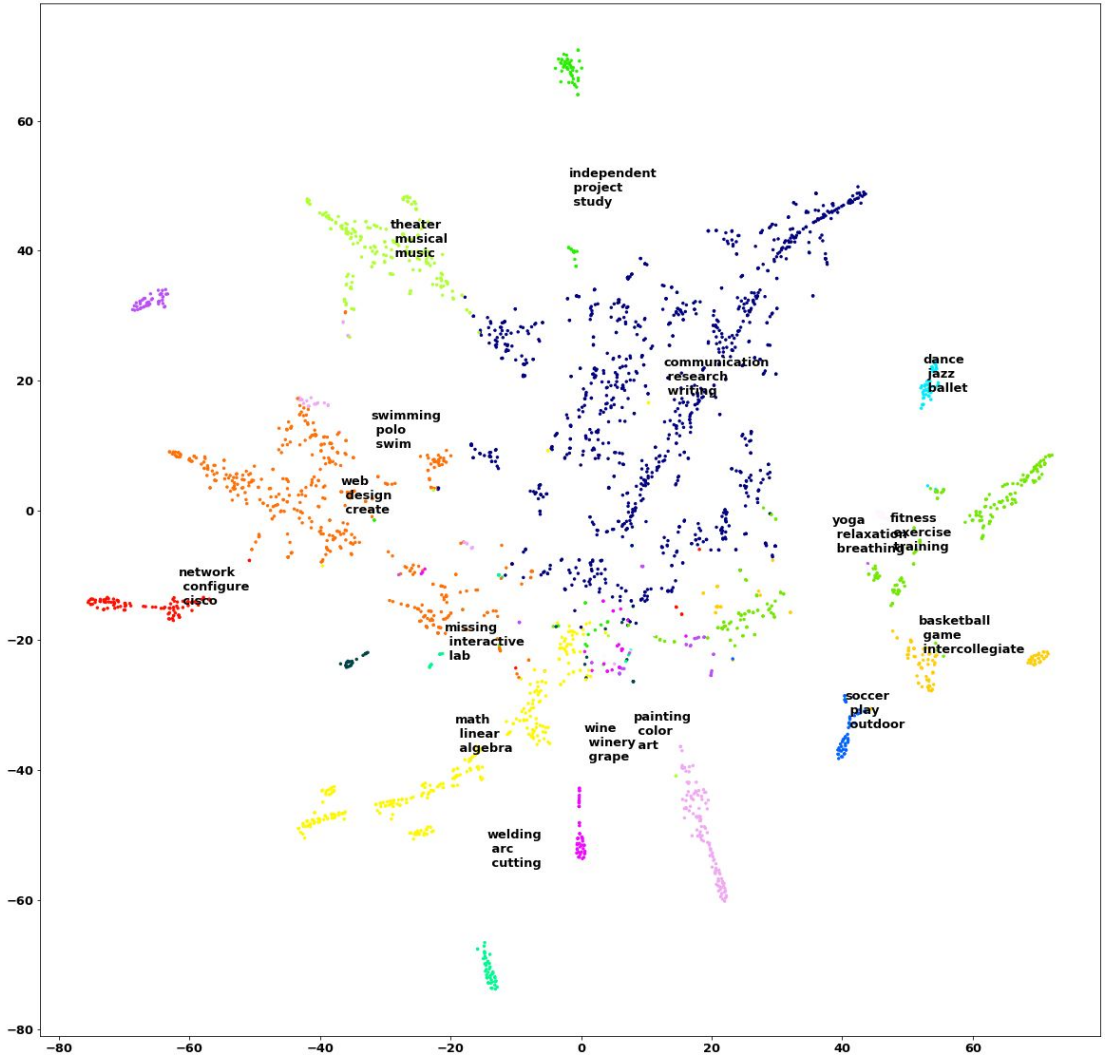
Choice of Topic Modeling Data

LDA


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
9
6
2
3
8
4
14
11
5
1
13
12
10
15
7
Marginal topic distribtion
2%
5%
10%


Top-30 Most Relevant Terms for Topic 11 (5.2% of tokens)
0
5,000
10,000
15,000
health
medical
patient
system
care
management
body
assessment
anatomy
describe
course
disease
human
emergency
physiology
knowledge
blood
clinical
drug
comprehensive
identify
lecture
practice
review
heart
student
respiratory
perform
treatment
function
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

NMF Topics


independent
project
study
theater
musical
music
communication
research
writing
dance
jazz
ballet
swimming
polo
swim
web
design
create
yoga
relaxation
breathing
fitness
exercise
training
network
configure
cisco
missing
interactive
lab
basketball
game
intercollegiate
soccer
play
outdoor
math
linear
algebra
wine
winery
grape
painting
color
art
welding
arc
cutting
60
40
20
0
−20
−40
−60
−80
−80
−60
−40
−20
0
20
40
60

Assessing Changes in Curriculum Over Time

Data:

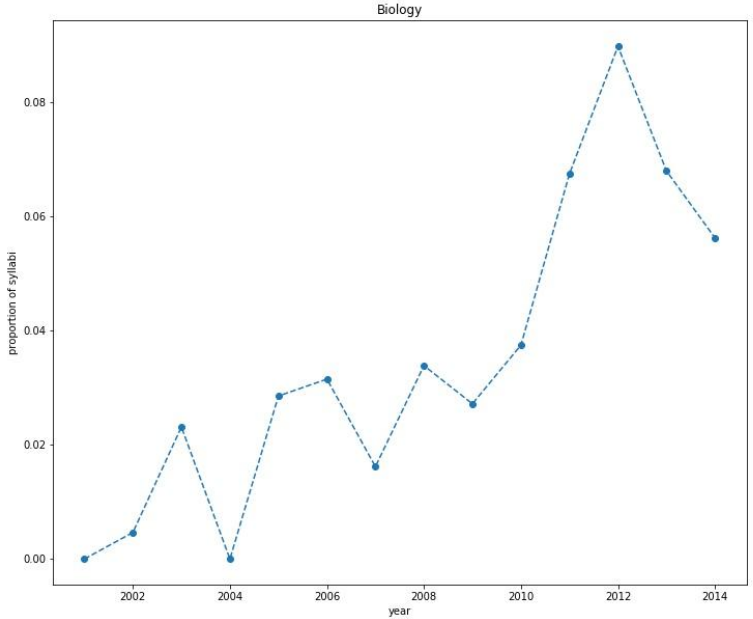
Oxford College at Emory University
3,778 syllabi from 2001-2014

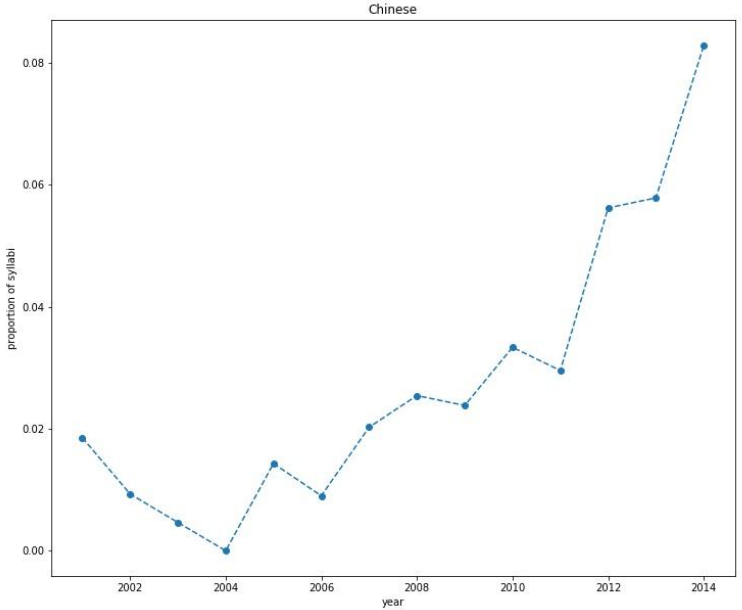
Methodology:

1. Best model: NMF with 100 topics
2. Hand-selected 43 topics
3. Grouping syllabi by year, fed each group into the model and got back the proportion of syllabi most representing each topic

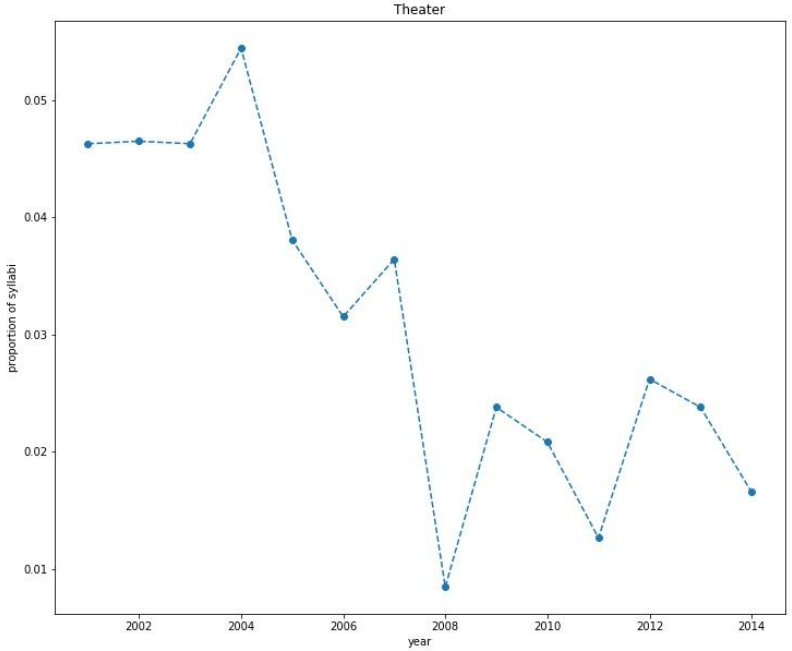
Topic Description	Word 1	Word 2	Word 3	Word 4	Word 5
Anatomy	dissection	lab	physiology	anatomy	laboratory
Anthropology	anthropology	park	culture	cultural	archaeology
Art	studio	drawing	color	charcoal	art
Astronomy	laboratory	astronomy	observation	universe	heavens
Biology	biology	lab	genetics	laboratory	scientific
Botany	field	plant	woody	trip	identification
Child Development	development	child	discussion	childhood	group
Classical Studies	metamorphoses	homer	myth	mythology	tragedy
Dance	dance	ballroom	folk	cha	cultural
Economics	economic	march	policy	demand	market
Environmental Science	environmental	ozone	lab	stream	science
Ethics	ethics	ethical	morals	utilitarianism	philosophy
Finance	accounting	financial	business	time	assets
French	sur	pour	dissertation	reprise	lire
Geology	geology	earth	lab	laboratory	geologic
German	mitt	german	thema	die	sie
Gerontology	aging	aged	dying	death	surrounding
Golf	golf	game	score	chipping	swing
Health	fitness	activity	physical	training	running
Linear Algebra	linear	algebra	differential	matrices	problem
Literature	fiction	poetry	portfolio	short	march
Logic	logic	reasoning	categorical	syllogism	ordinary
Mandarin Chinese	dialogue	workbook	character	mandarin	cheng
Martial Arts	tai	skill	chi	practice	form
Mathematics	gateway	calculus	trigonometric	derivative	logarithmic
Media Studies	screening	film	cinema	reserve	sound
Meteorology	lab	weather	climate	meteorology	atmospheric
Music Education	music	musical	western	concert	classical
Musical Performance	dress	rehearsal	black	concert	music
Philosophy	philosophy	philosophical	philosopher	reverse	ken
Physical Education	cycling	fitness	indoor	workout	physical
Poetry	poetry	workshop	mid	poem	story
Political Philosophy	republic	utopia	book	political	politics
Political Science	politics	political	science	syllabus	international
Probability and Statistics	statistics	statistical	probability	data	hypothesis
Proofs	mathematics	theory	mathematical	landau	analysis
Racket Sports	singles	play	smash	badminton	net
Spanish	leer	lunes	antes	para	las
Swimming	pool	swim	swimming	water	underwater
Theater	theater	play	theatrical	performance	production
US Government and Politics	political	federalist	march	federalism	bureaucracy
Weight Training	lift	weight	training	fitness	muscle
Western History	art	ancient	architecture	paleolithic	aesthetic

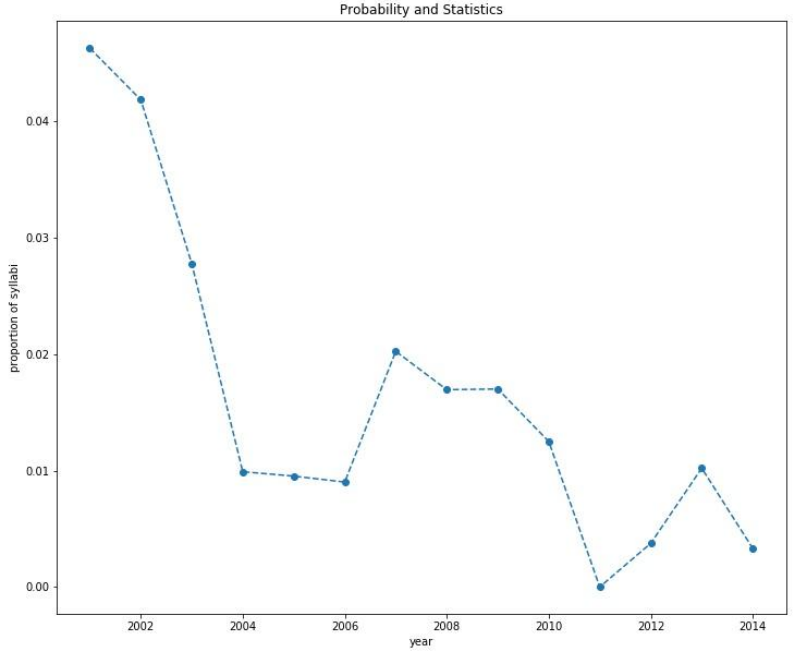
Topics Increasing in Prevalence





Topics Decreasing in Prevalence





Auto-Generated Guide

communication: Communication (from Latin communicare, meaning "to share") is the act of conveying meanings from one entity or group to another through the use of mutually understood signs, symbols, and semiotic rules.

writing: Writing is a medium of human communication that represents language and emotion with signs and symbols.

research: Research comprises "creative and systematic work undertaken to increase the stock of knowledge, including knowledge of humans, culture and society, and the use of this stock of knowledge to devise new applications." It is used to establish or confirm facts, reaffirm the results of previous work, solve new or existing problems, support theorems, or develop new theories.

soccer: Association football, more commonly known as football or soccer, is a team sport played with a spherical ball between two teams of eleven players.

outdoor: Wilderness or wildland is a natural environment on Earth that has not been significantly modified by human activity.

game: A game is a structured form of play, usually undertaken for enjoyment and sometimes used as an educational tool.

interactive: Across the many fields concerned with interactivity, including information science, computer science, human-computer interaction, communication, and industrial design, there is little agreement over the meaning of the term "interactivity", although all are related to interaction with computers and other machines with a user interface.

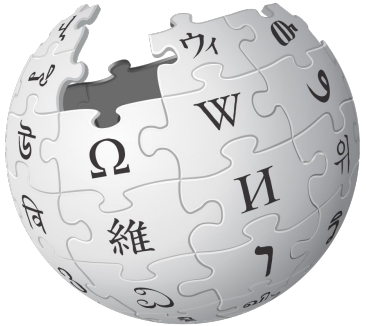
transistor: A transistor is a semiconductor device used to amplify or switch electronic signals and electrical power.

computer: A computer is a device that can be instructed to carry out sequences of arithmetic or logical operations automatically via computer programming.

welding: Welding is a fabrication or sculptural process that joins materials, usually metals or thermoplastics, by using high heat to melt the parts together and allowing them to cool causing fusion.

cutting: Cutting is the separation or opening of a physical object, into two or more portions, through the application of an acutely directed force.

metal: A metal (from Greek μ■ταλλον métallon, "mine, quarry, metal") is a material that, when freshly prepared, polished, or fractured, shows a lustrous appearance, and conducts electricity and heat relatively well.



WIKIPEDIA
The Free Encyclopedia

Future Directions

Topic Modeling

- Incorporate Wikipedia clickstream data
- Try larger corpuses from other universities, high schools, or community colleges
- Look at curriculum changes across time within a particular field (ie: statistics)

Improve Recommender System

- Employ machine learning techniques to automate the input weights
- Employ topic modeling module on uploaded syllabus to eliminate onboarding flow

Questions?

How was the recommender's performance quantified?

For a Single Employer:

classroom	A	B	C	D	E	F	G
our rank	13	16	14	15	30	91	29
ideal	0	1	2	3	4	5	6

Median - Ideal = Score

16 - 3 = 13

Taking the Median Over All Employers = 34