



Using Human Language Technology to Improve MOOC Learning

Victor Zue

with Daniel Li and Hung-yi Lee

Research sponsored by Quanta Computer, Inc.

HLT Meets MOOC



- **Human language integral part of education**
 - Lectures, books, tutorials, discussions, Q&A, etc.
- **Some challenges for MOOC**
 - **Heterogeneity:** Student's background (preparedness, language competence, learning style, etc.)
 - **Scale:** One-size-fits-all solutions would not suffice; need mass customization
- **Human language technologies could help**
 - Process and manage contents
 - Develop speech-based interfaces for easy access
 - But need general and scalable solutions whenever possible

Some Example Uses



- **Transcription of course materials**
 - Currently done by humans
- **Translation into multiple languages**
 - Long term research
- **User identification and authentication**
 - Security and privacy concerns
- **Information management**
 - **Categorization:** e.g., managing Q&A
 - **Linking:** e.g., different forms, different sources
 - **Summarization:** e.g., précis of a discussion
 - **Search:** e.g., dialogue-based search engine
- ...

Some Example Uses

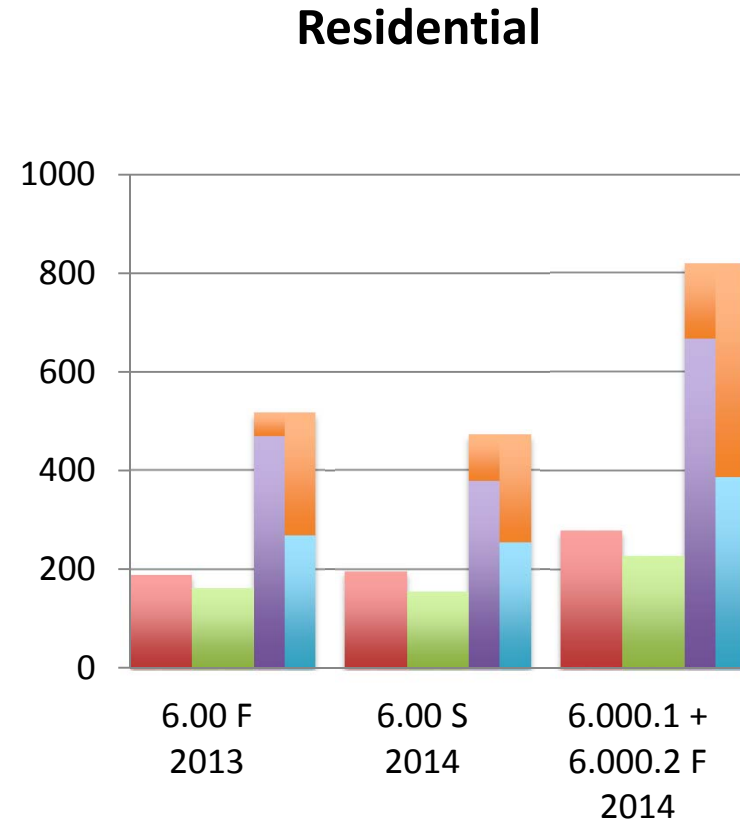
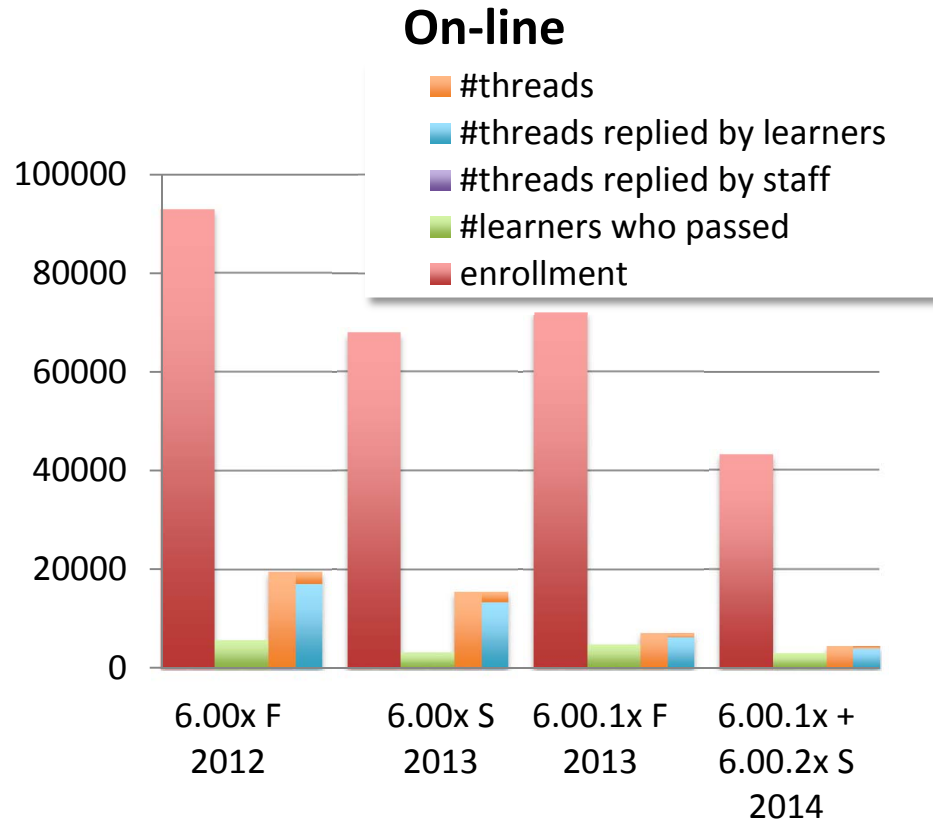


- **Transcription of course materials**
 - Currently done by humans
- **Translation into multiple languages**
 - Long term research
- **User identification and authentication**
 - Security and privacy concerns
- **Information management**
 - **Categorization:** e.g., managing Q&A
 - **Linking:** e.g., different forms, different sources
 - **Summarization:** e.g., précis of a discussion
 - **Search:** e.g., dialogue-based search engine
- ...



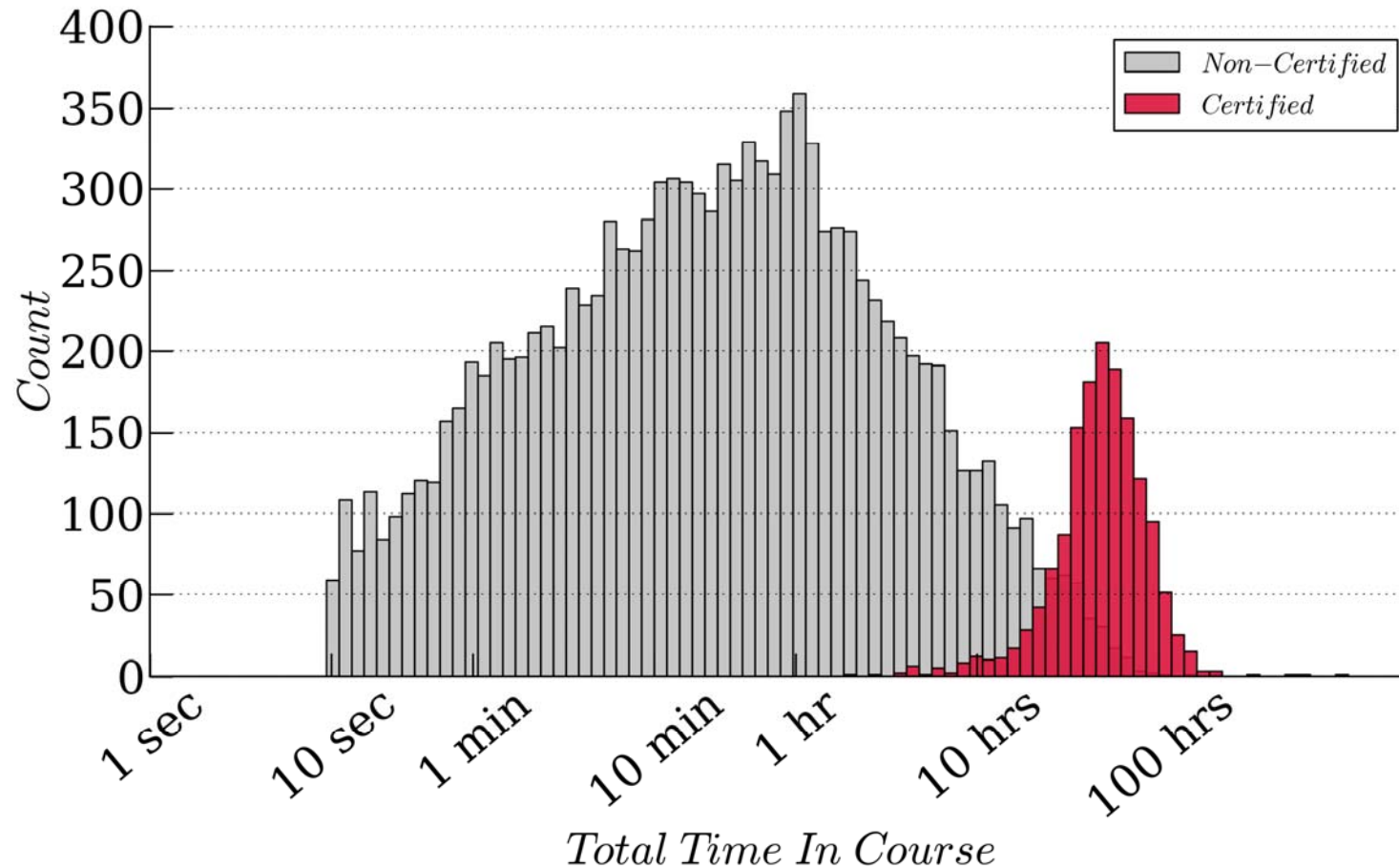
**On-line and residential
courses are different in
many respects**

Some Comparisons ...



- Multiple offerings of the same course at MIT and on edX
- Class size, drop out rate, manners of interaction, etc.

Why so many dropouts in MOOC?



- **Browsing, heterogeneous background, commitment, insufficient help, ...**

Data courtesy of Prof. John Guttag



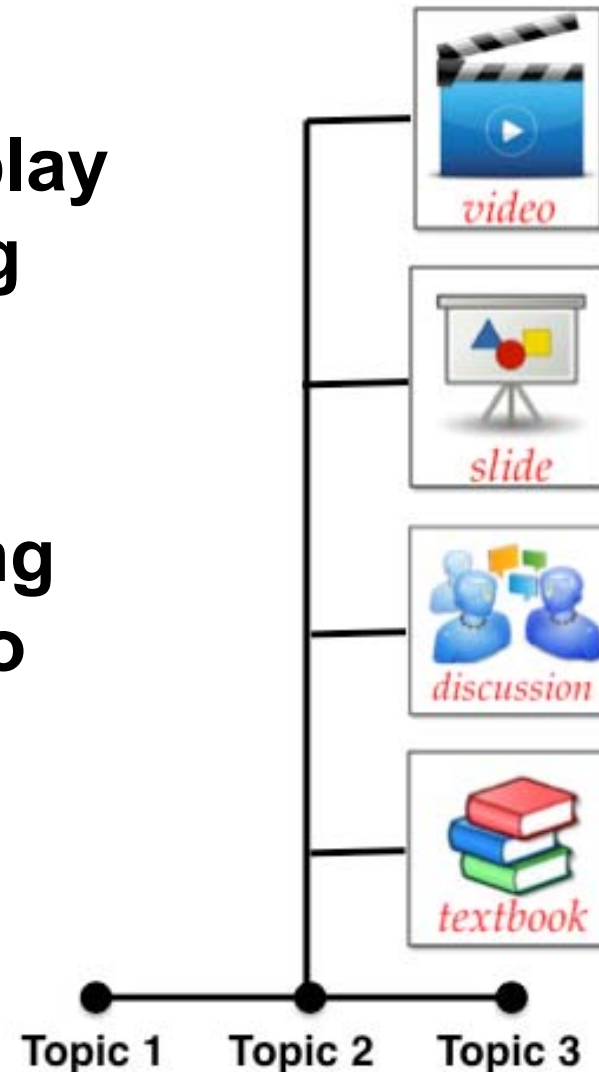
Multimedia Content Linking for MOOC

with Daniel Li

Linked Knowledge



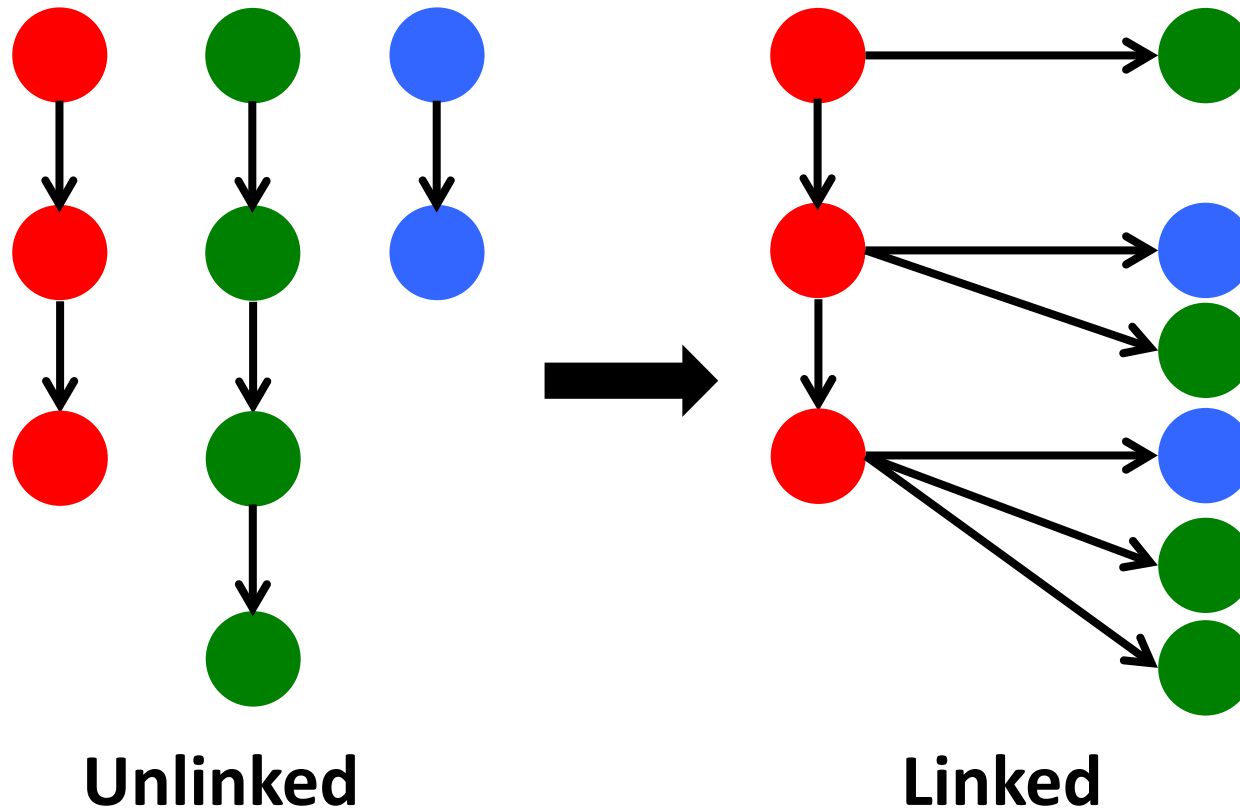
- Lectures, slides, textbook, forum, etc., of a topic can play reinforcing roles in learning
- Link various contents of a given topic
- Create an adaptable learning environment for students to navigate freely



A Hypothesis



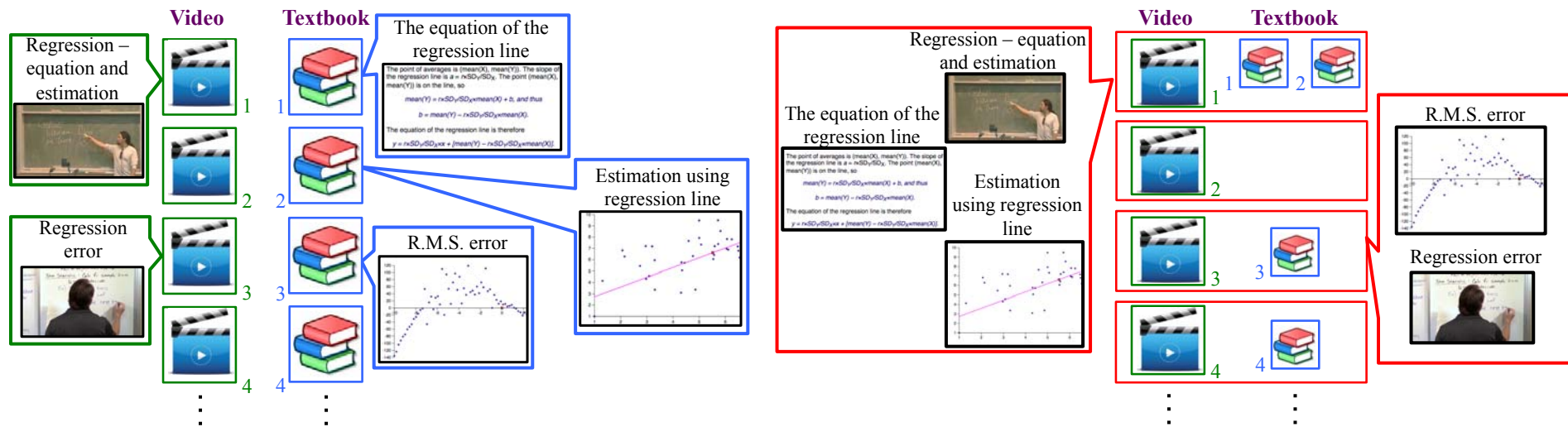
- Organizing and linking related contents will improve learning



A Hypothesis



- Organizing and linking related contents will improve learning



Unlinked

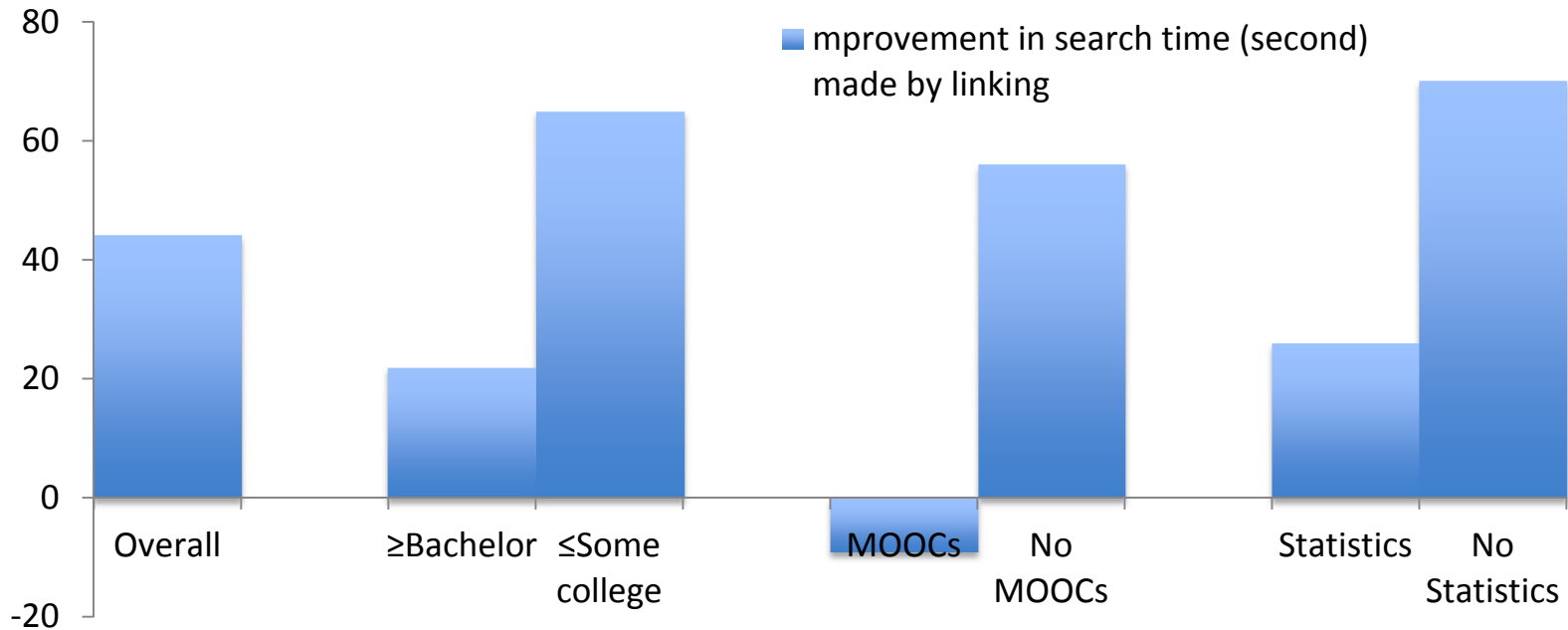
Linked

Could linking really help?



- **Conducted crowd sourcing study online**
 - **Course:** Stat2.1X from Berkeley
 - Videos, slides, textbook
 - **Subjects:** Amazon Mechanical Turk workers with varying background
 - Education? Experience with MOOC? Statistics?
 - **Ground Truth:**
 - Established by human experts
 - **Measurements:**
 - **Information Search:** How fast and accurately can subjects retrieve information
 - **Concept Retention:** How well can they retain information

Information Search (Speed)



Significance
($p=0.01$) ✓

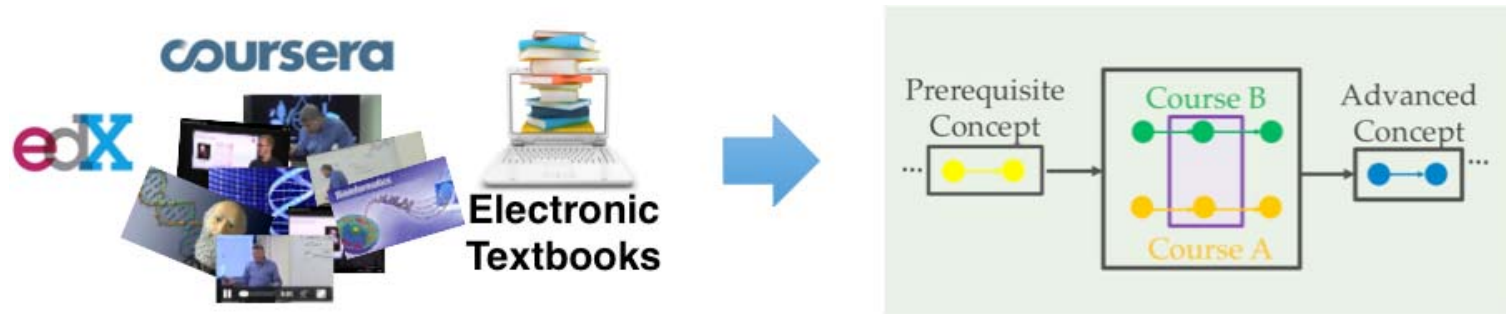
✓

✓

- **Linking improves speed in all cases involving novice learners**
- **Improvement made without sacrificing accuracy**

Visualization and Inference of Learning Paths from Online Courses

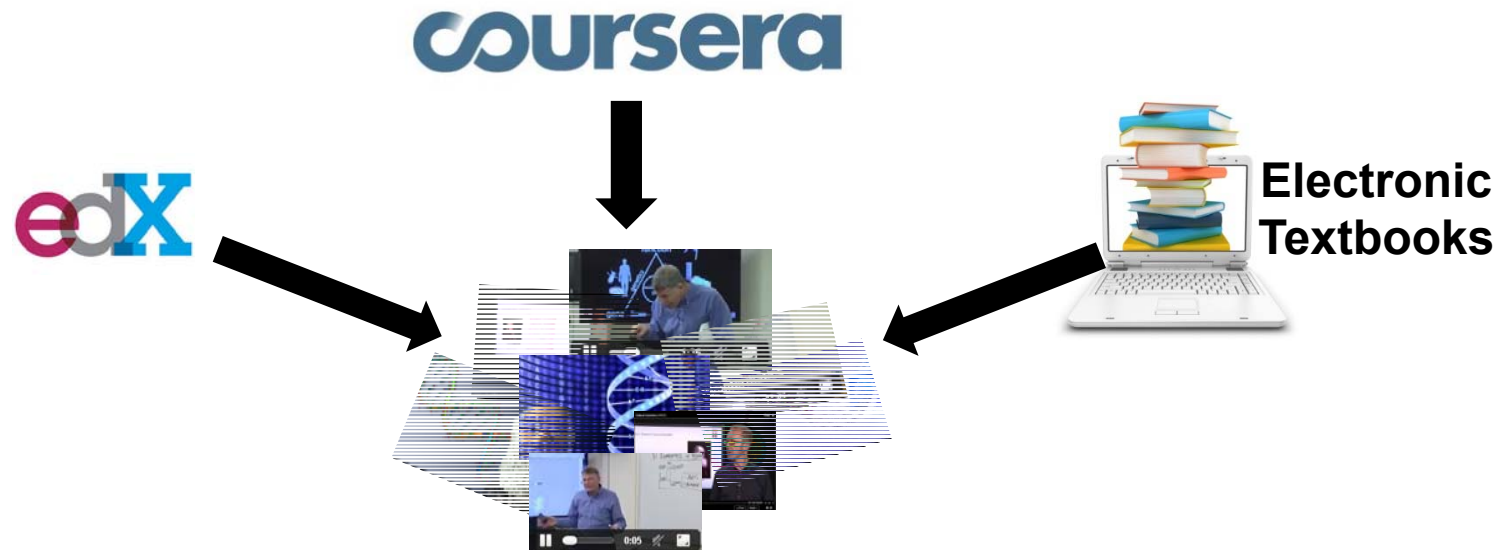
with Hung-yi Lee



1. Data Gathering

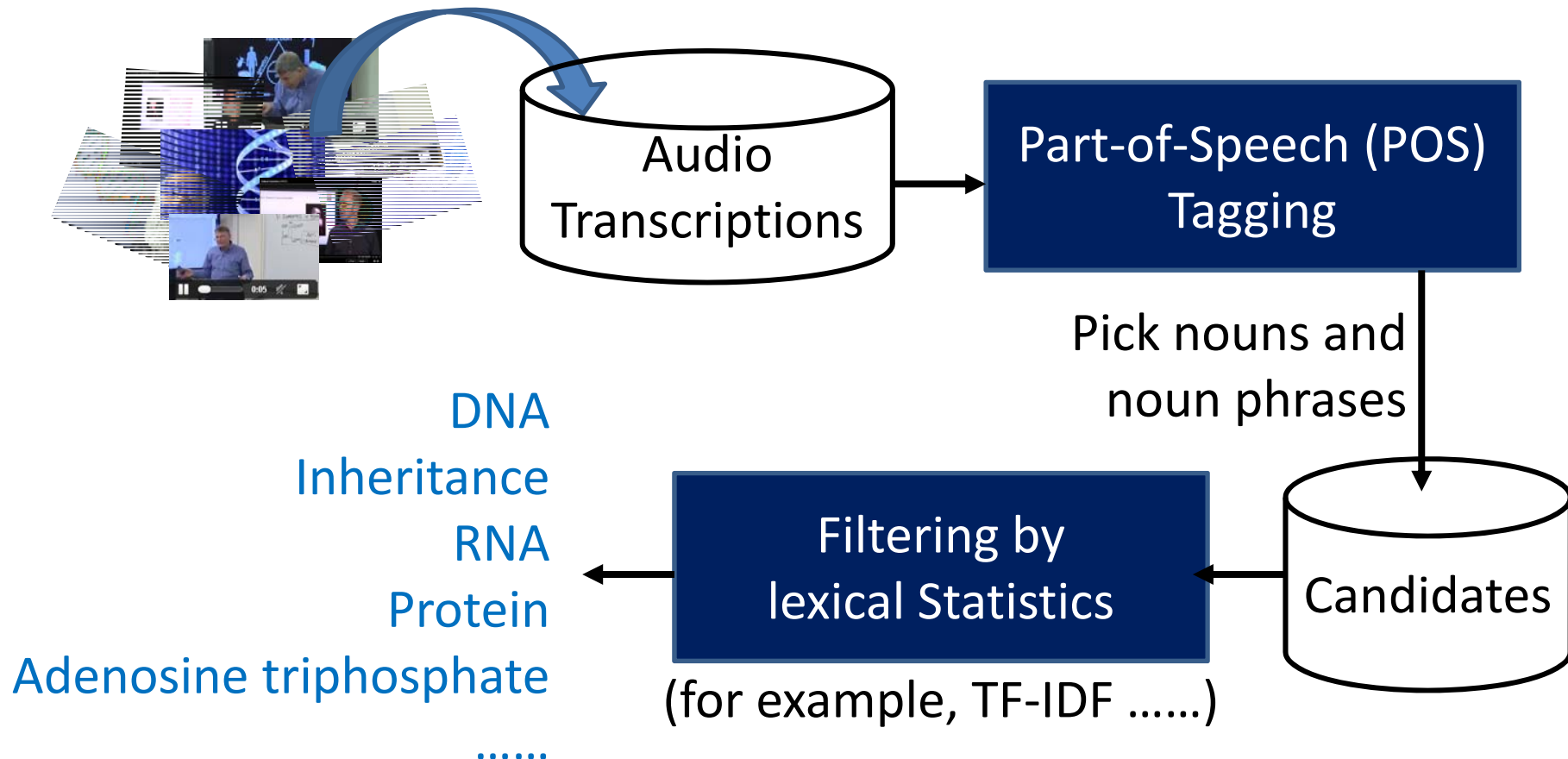


- **Collect video clips from MOOC platforms and electronic textbooks from the Internet**



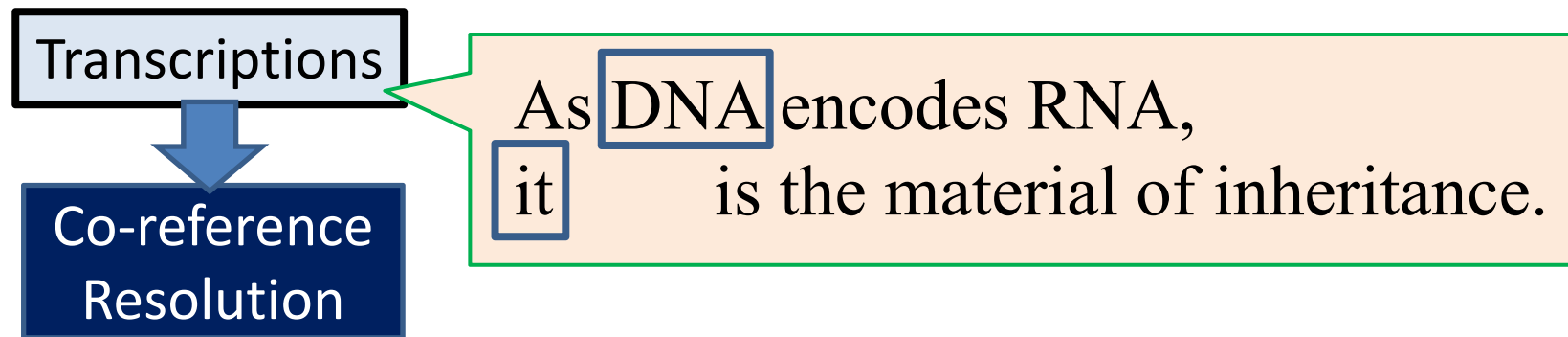
2. Key Concept Extraction

- **Extract terms representing key concepts from the audio transcriptions of video clips**



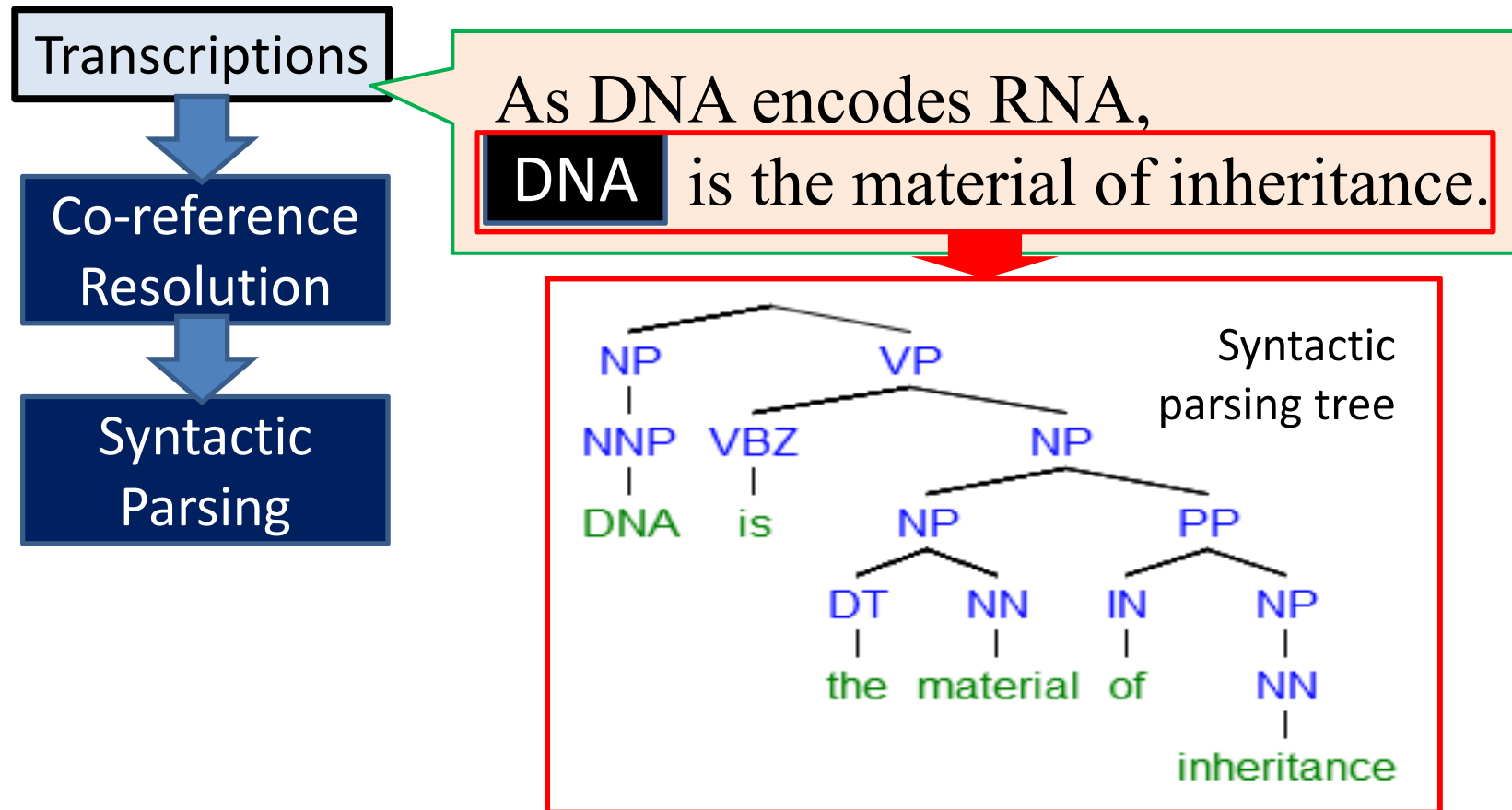
3. Relation Extraction

- “Read” the text/transcriptions to find the relation between the key concepts



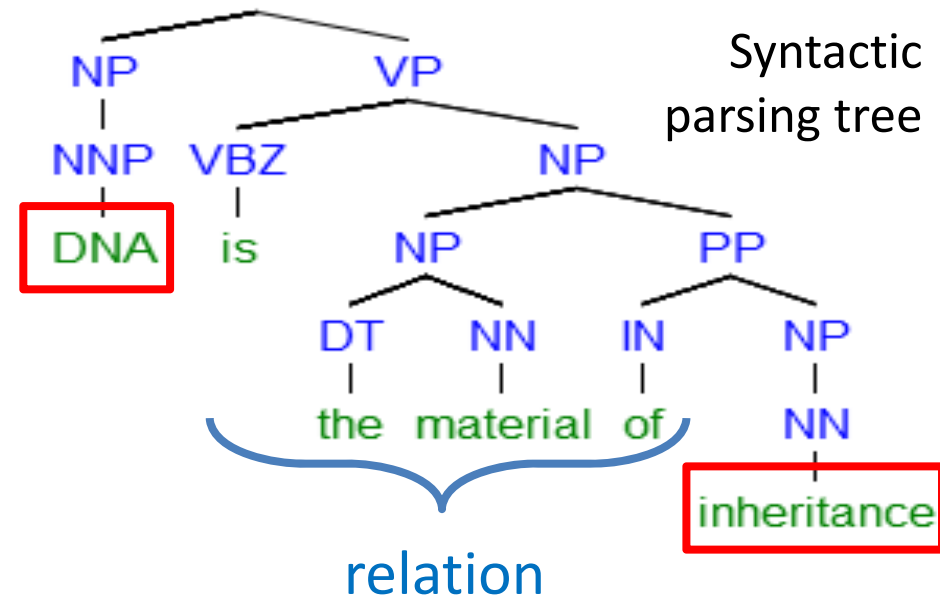
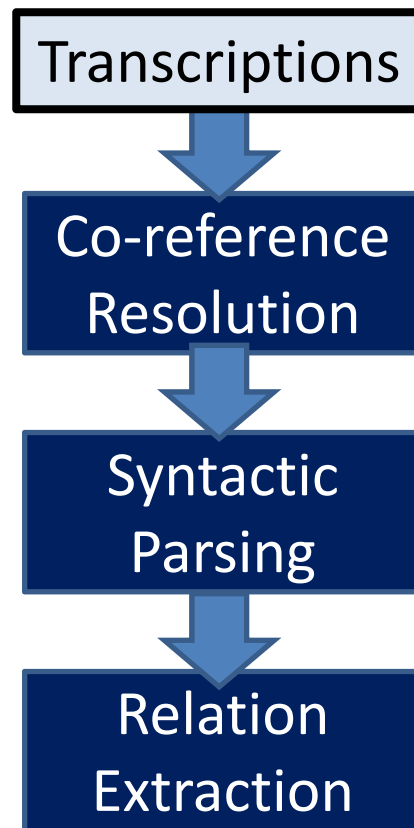
3. Relation Extraction

- “Read” the text/transcriptions to find the relation between the key concepts



3. Relation Extraction

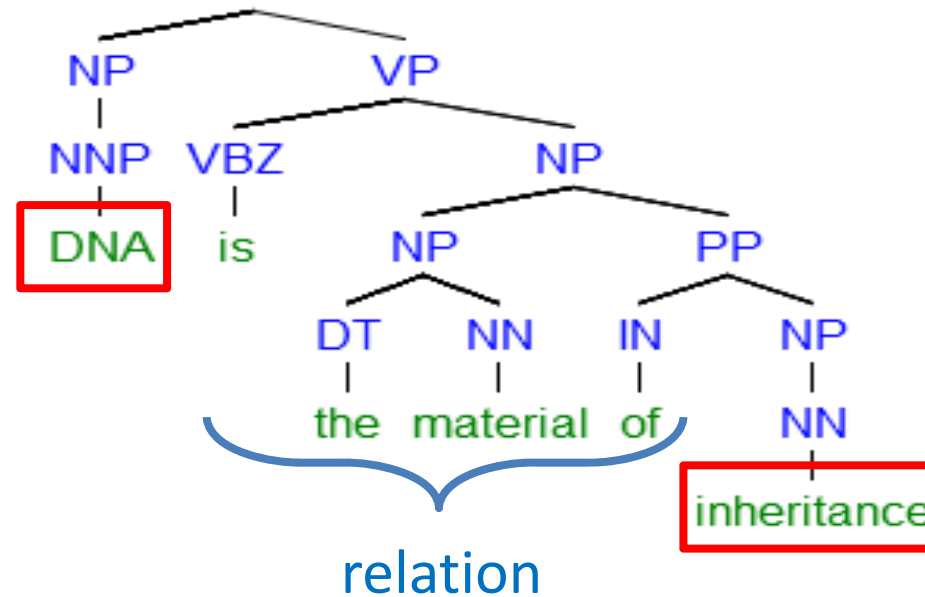
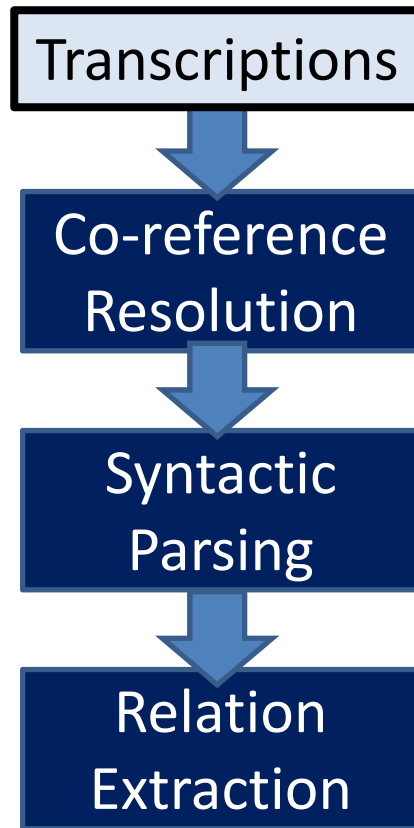
- “Read” the text/transcriptions to find the relation between the key concepts



A statistical model finds the relation between concepts on the syntactic tree [Mausam, EMNLP'12].

3. Relation Extraction

- “Read” the text/transcriptions to find the relation between the key concepts

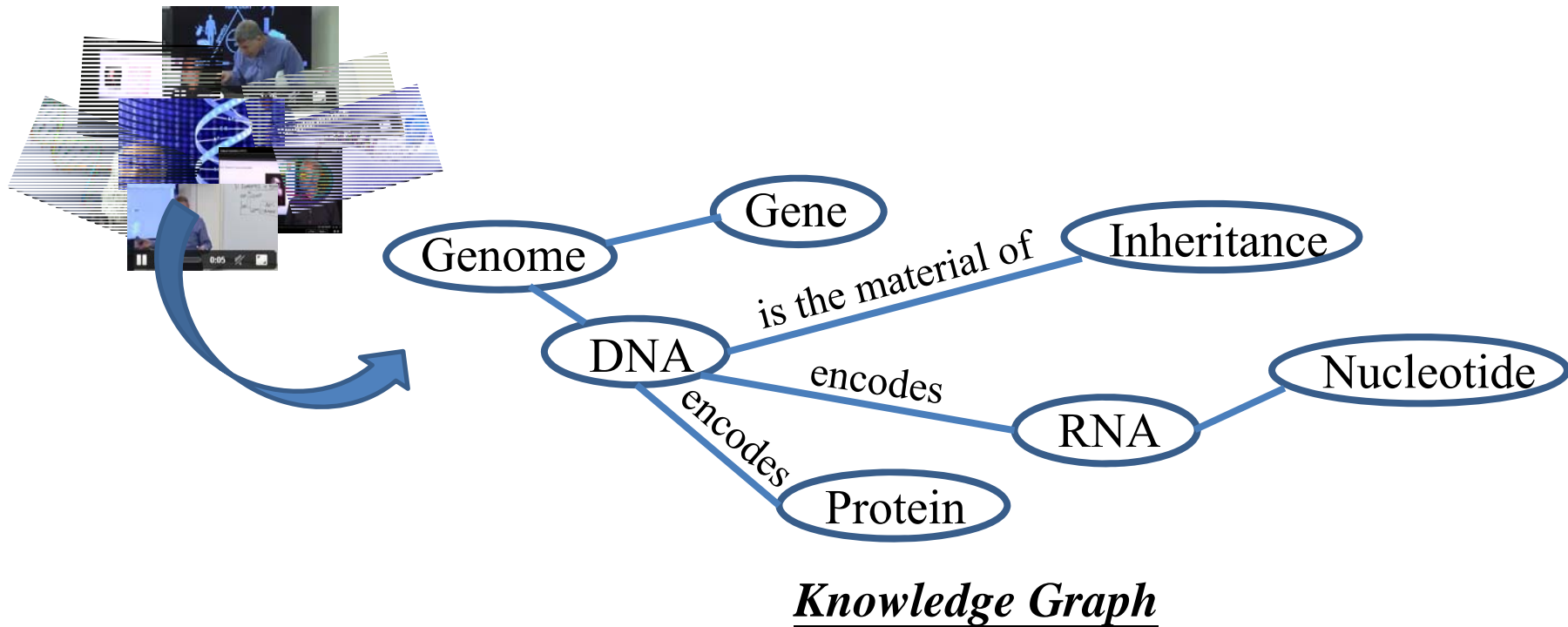


The relation of “DNA” and “inheritance” is “is the material of”.

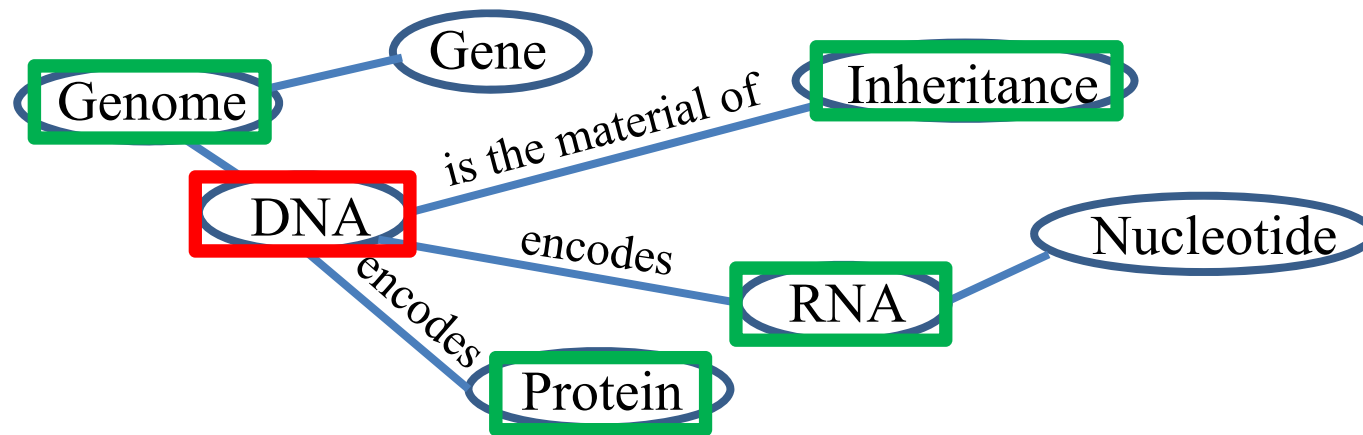
4. Knowledge Graph Construction



- **Nodes:** key concepts
- **Edges:** relation between the key concepts

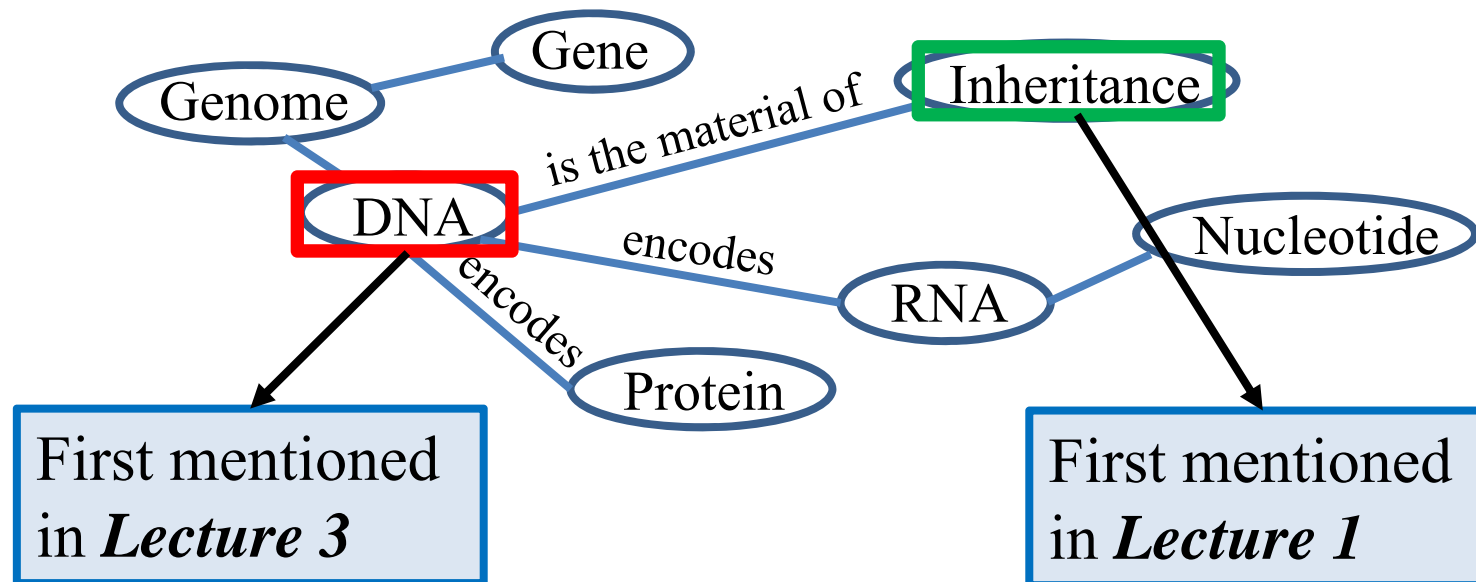


5. Prerequisite Concept Inference



- “DNA” is related to “Inheritance”, “RNA”, etc.
- Which one is the prerequisite concept of “DNA”?

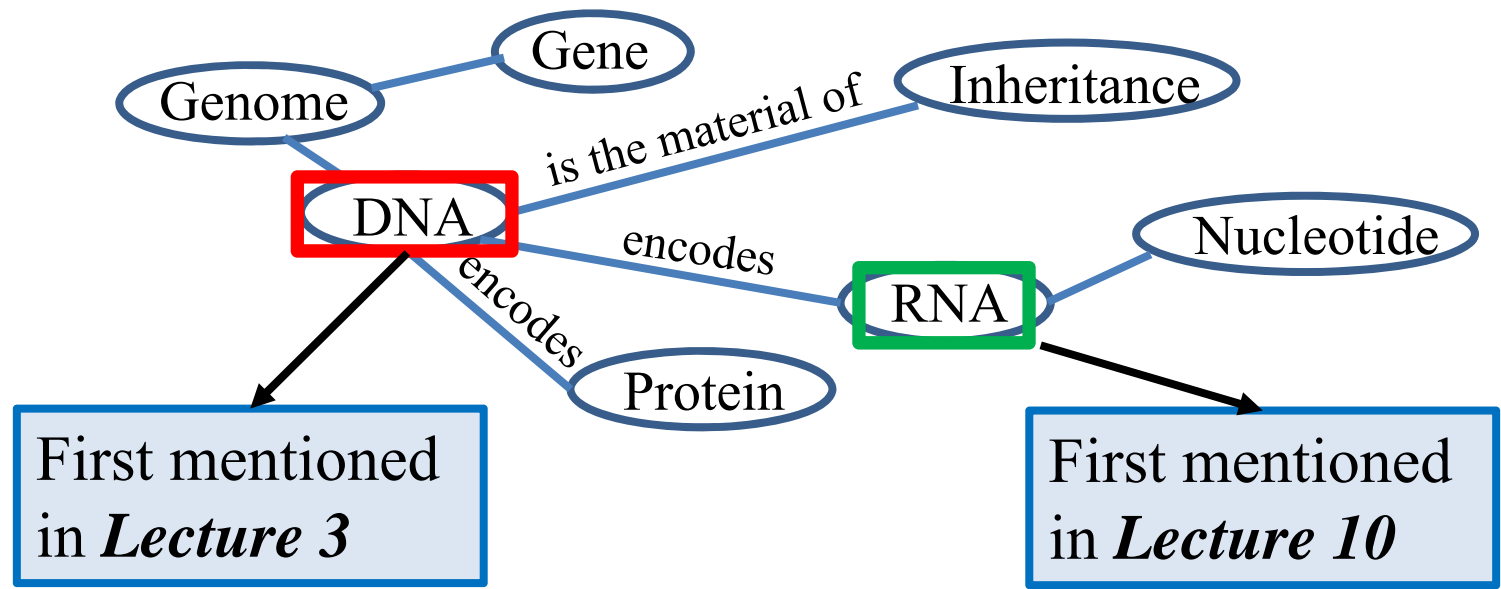
5. Prerequisite Concept Inference



- Analyze the positions where the concepts are mentioned the first time in a course

“Inheritance” is the prerequisite concept of “DNA”

5. Prerequisite Concept Inference



➤ Analyze the positions where the concepts are mentioned the first time in a course

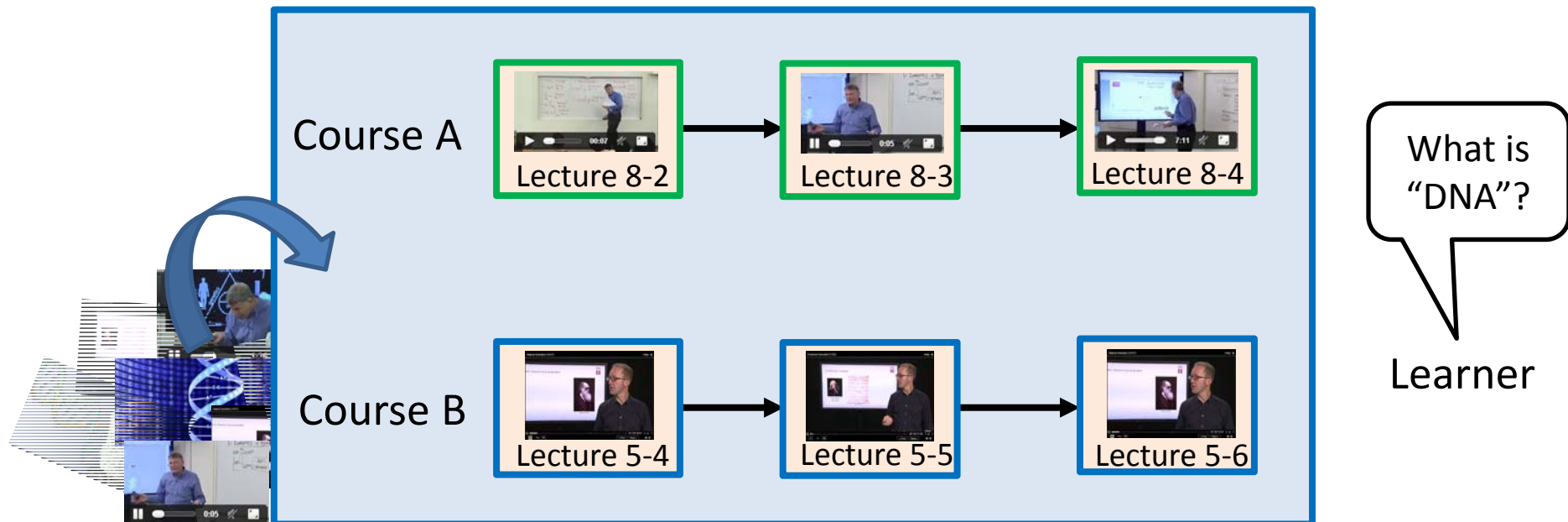
“Inheritance” is the prerequisite concept of “DNA”

“RNA” is the advanced concept of “DNA”

6. Merge of Courses



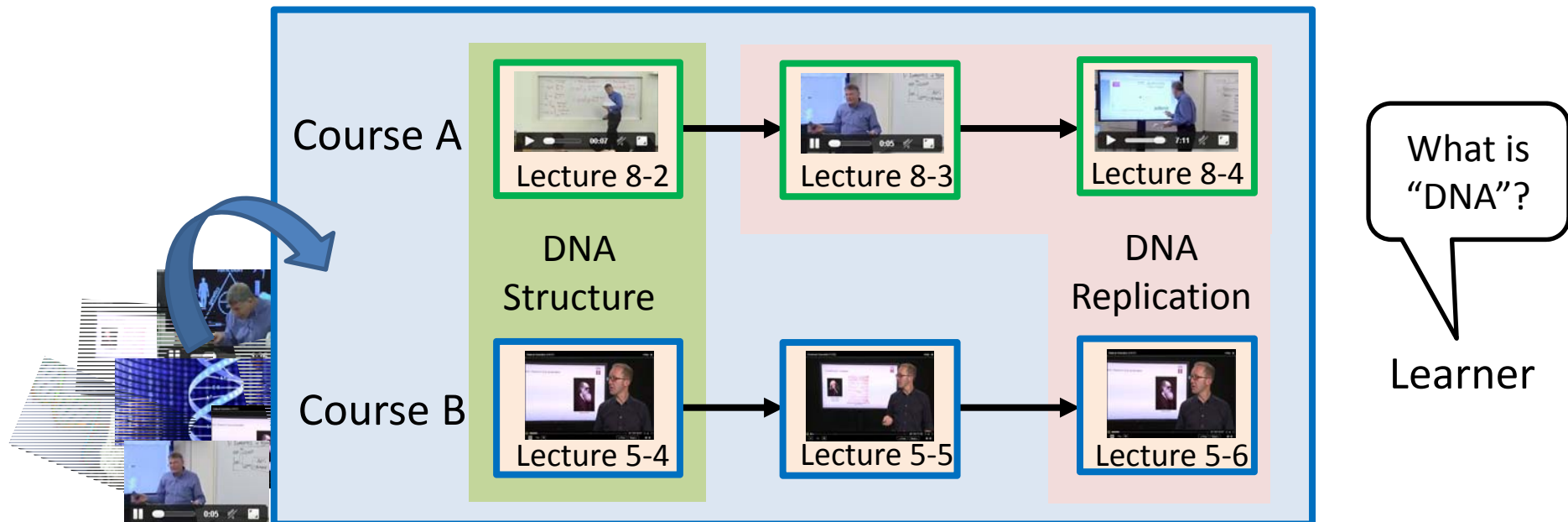
- Find lectures related to “DNA”



6. Merge of Courses



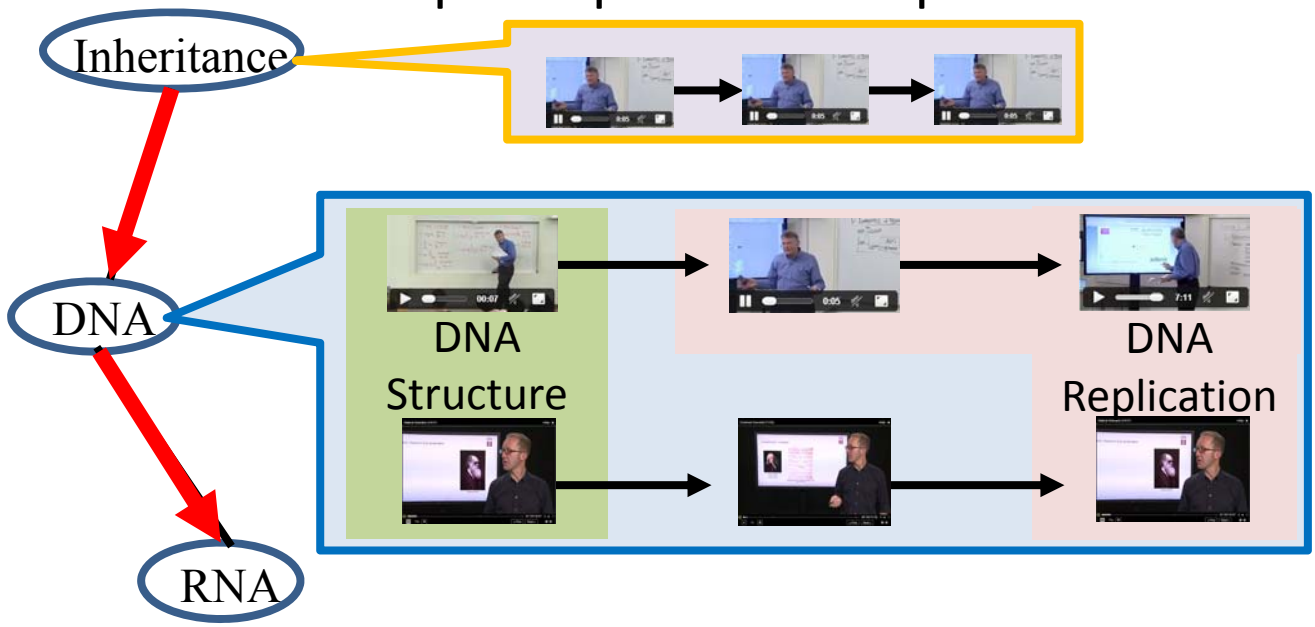
- Use vector space models to represent the audio transcriptions of each lecture
 - Compute **cosine similarity** between the models
 - Merge the courses with high cosine similarity



7. Cannot Understand?



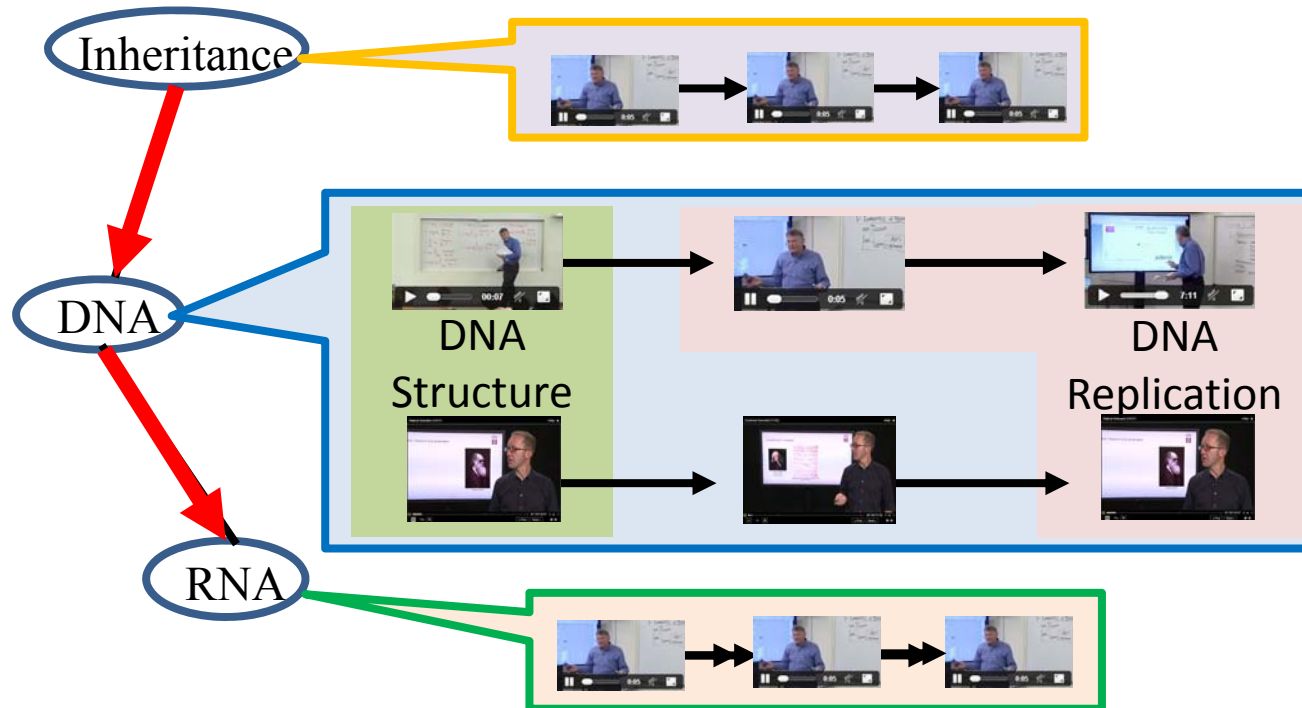
Show the lectures of the prerequisite concept of "DNA"



I cannot understand.

Learner

8. Learn More



Learn More!
Learner

Show the lectures of the advanced concept of "DNA"

Video Demonstration



Ongoing Research



- **Automatic content linking**
- **Topic modeling**
- **Video processing**
- **Sentiment analysis**

Summary



- **We have achieved some positive results using HLT for MOOC**
- **Improvements, extensions, and evaluation are ongoing**
- **We hope to collaborate with colleagues outside of MIT to achieve greater impact**



Thank You!





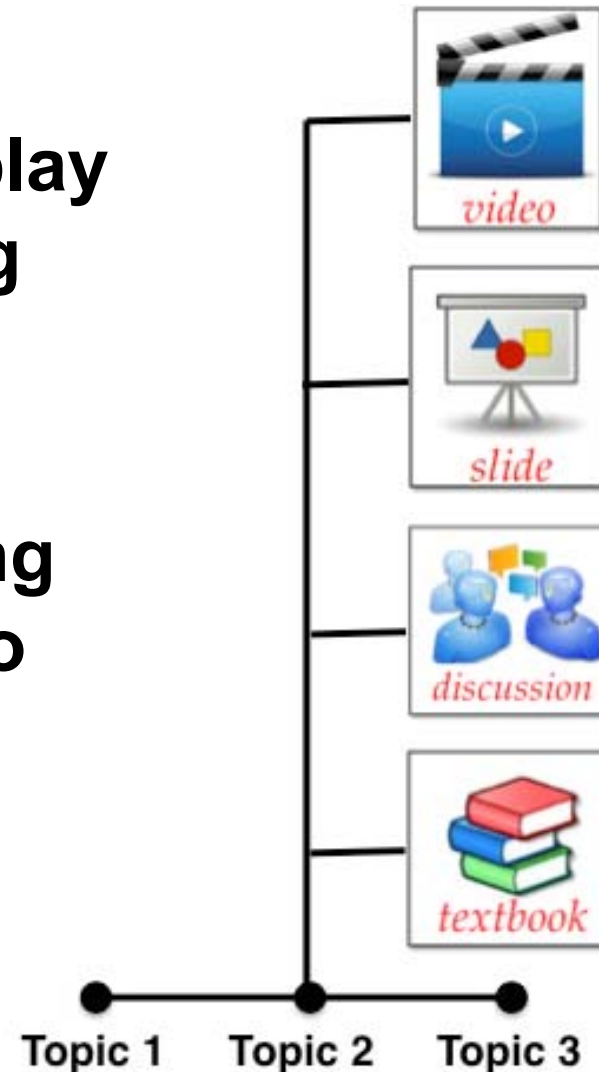
Multimedia Content Linking for MOOC

Daniel Li

Linked Knowledge

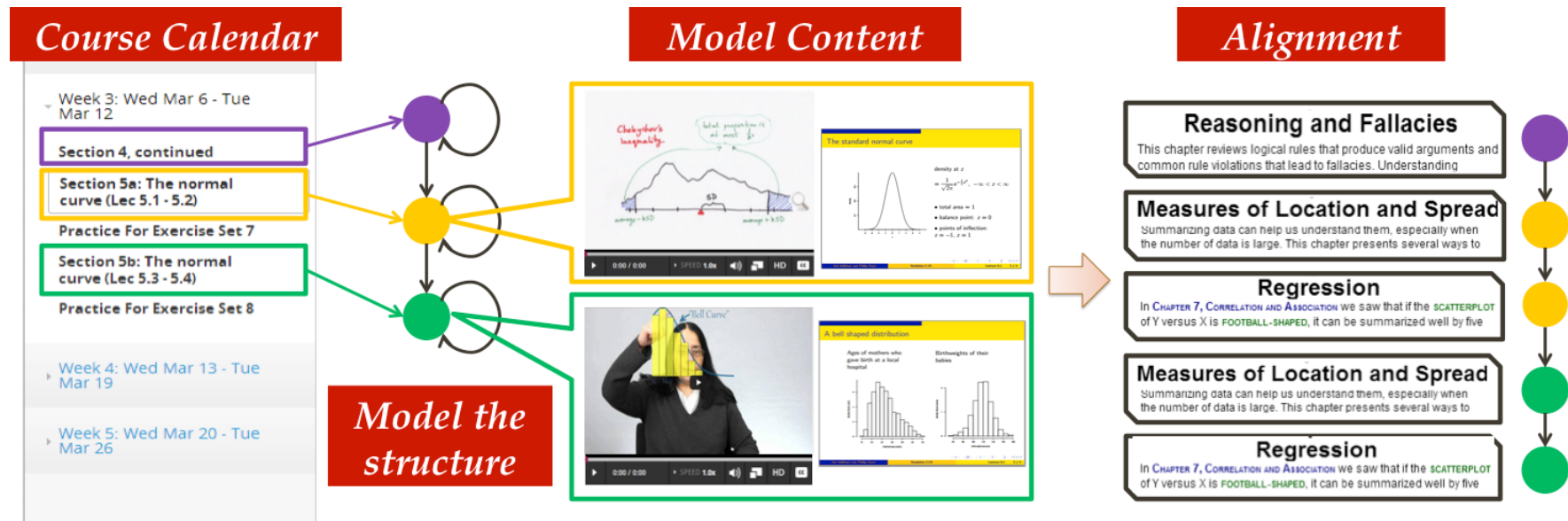


- Lectures, slides, textbook, forum, etc., of a topic can play reinforcing roles in learning
- Link various contents of a given topic
- Create an adaptable learning environment for students to navigate freely



Multimedia Content Linking

- Discover the content linking using Hidden Markov Models (HMM)
- Align textbook sections into HMM states
- Link contents assigned to the same state



Experimental Setup



- **Dataset from Berkeley's Intro to Statistics**
 - **Video Transcription:** 7 hours/56K words
 - **Slides:** 184 pages/9K words
 - **(Recommended) Textbook:** 162 sections/42K words
- **Evaluate against hand-labeled ground truth**
 - **Three Conditions:** transcription only; transcription + # of pages of slides; transcription + slides
 - **Three Models:** baseline (measuring cosine similarity using simple word statistics); HMM; and HMM weighing keywords higher than other words

Preliminary Results



Model \ Condition	Baseline	HMM only	HMM+ Feature Sel.
Transcription only	36.9	48.1	41.2
Transcription + #pages	48.5	53.5	62.4
Transcription + slides	49.6	55.8	65.6

- **HMM results better than baseline**
- **Even better results with Feature Selection**
- **Video transcription + slides yield the best performance**

User Interface



LECTURE 2.2: HOW TO DRAW A HISTOGRAM Segment 5



Video

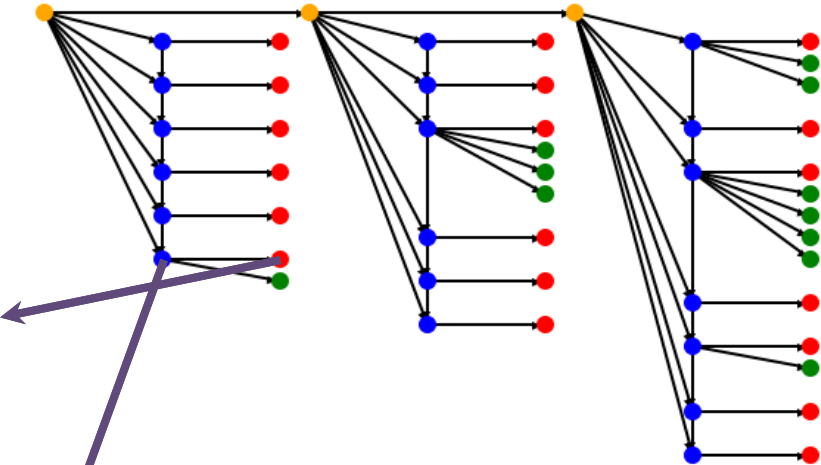
BERSTAT2SP13-G020200_100

Something like a bar graph?

- Want to draw a diagram analogous to a bar graph, to show how the incomes are distributed.
- Want income on the horizontal axis, and vertical rectangular bars.

income (thousands of dollars)	% of households
0 - 10	20
10 - 25	20
25 - 50	18
50 - 100	18
100 - 150	7

- Course start
- Selected lectures
- Slides
- Course end
- Video clips
- Textbook sections





Opinion Summarization



Course: Ancient Greek Hero

Show All Discussions ▾	
SORT BY: DATE VOTES COMMENTS	
Welcome message from Professor Gregory Nagy	+341 577
[FROM STAFF] Hour 3: Professor Nagy's Discussion Question-- Performance of Song in the Iliad	+31 230
What is everybody's background?	+30 226
[From Prof Nagy] Discussion Question for Hour 4	+30 212
[FROM CLAUDIA] Gratitude and Question: What is Your Favorite Passage from the Iliad?	+35 189
PROF. NAGY SHARES "Strangeness and Lento," a thread begun by tristandestry and members of Briseis	+158 186
[FROM PROF. NAGY] Important message about assessments and certificates	+124 158
Any people over 50?	+8 130
Note from Professor Nagy on the use of the Discussion Forums	+89 128
Wich is your favorite greek hero?	+9 123
[From Staff] Discussion of Hour 3 Close Reading Exercise	+8 114
[From Staff] Discussion of Hour 4 Close Reading Exercise	+9 97
Hour 5 Release Date	+14 86
[From Prof. Mueller] Discussing Assessments	+21 48

Topic Modeling

desertbob
24 days ago

Referring to the Professor's comments on the fact that laments tend to be more of a female medium, I can't help to notice that **the laments of both Andromache and Kleopatra seem to be more attuned to their own fate**, and the fates of those with whom they are acquainted, than their mates or lovers.

That evaluation would seem to be in tune with the general honor of teller of the tale, even considering the active participation of Penthesileia in combat, that female roles are here generally confined to worry, concern, lament, over the destiny, should the fates have woven skein against the victory of the heroes, the victors, and their supporting casts.

Achilles, on the other hand, is a remarkably different theme than those of the women, in that he is prominently present.

Yes, I agree. He keeps having to pacify his lesser colleagues who are not up to the level. Even human beings are prone to revolt. His power is founded on some basic "rules of law" and constant negotiations in which everyone needs to get some due and to save face/honour.

-posted 22 days ago by dcaillat

It will be interesting to **compare Hesiod with Homer** on these issues. I have read a little and Hesiod seems to have an almost reverence for how Zeus manages the cosmos (but then perhaps he is not intending to describe the Age of Heroes, but his own Age of Iron).

-posted 21 days ago by hriggs

My view is that **the performance of the song in The Iliad confers the text a multi-dimensional, almost 3D, perspective of the story.** And now, that I think about it, I find it quite amazing for the genius behind it. You have epic and lyric. Two mediums which will help the people engage at different levels. One conveys a message while the other privileges emotions. It's like being "sucked" into this "time-space" worm whole to find yourself, the reader/listener as silent but yet active participant of the plot.

-posted 19 days ago by Luis_Castro

"The laments of both Andromache and Kleopatra seem to be more attuned to their own fate."

"The result is more sacrifice, rather than redemption."

"Zeus is by no means an omnipotent ruler."

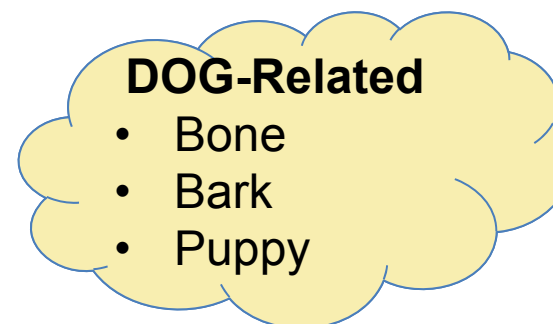
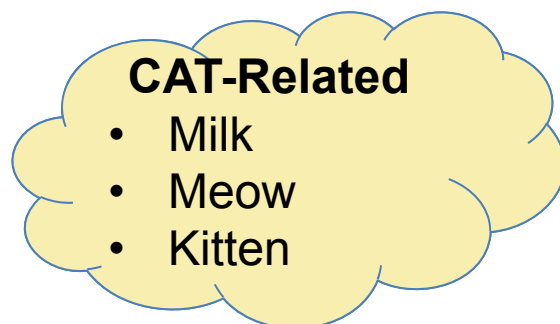
"compare Hesiod with Homer"

"The performance of the song in the Iliad confers the text a multi-dimensional, almost 3D, perspective of the story."

Topic Modeling



- **Goal:**
 - To apply statistical models for discovering the abstract "topics" that occur in a collection of documents
- **Use Latent Dirichlet Allocation (LDA)**
 - A generative graphical model



- Each document (e.g., a forum post) is a mixture of a small number of topics; and each topic has probabilities of generating various words

Topic Modeling

- Topic clustering on the forum of “Biology” class by LDA

Diseases body
health cancer
understanding
disease research
lives medicine
medical

Ancestors neanderthals
humans modern africa
neanderthal mated
recent genome
interbred interbreeding
migrated

Experiment flies
wings food
olfaction source
wing smell ability
navigate winged
obstacle reach

Organism life level
diversity complex
biological building
appearance cellular
functions process
mechanisms forms
living

Research science
scientific area project
professional field
development scientist
discoveries technology

Mutation gene genes
related function
mutations connection
relation separate
involved linked caused
mutant mutants

Molecule group
structure strong
functional close
making large
amount create
separated
chemical

Atom bonds polar
amino hydrogen acid
covalent atoms form
acids oxygen ionic
chain carbon bonding
electrons nitrogen

DNA trait dominant
affected recessive
linked phenotype
sex breeding
individual relevant
inheritance color

Pset problem set
week exam info
practice grade
due deadline
complete date
posted certificate

Topic Modeling

- Topic clustering on the forum of “Ancient Greek Hero”

Song andromache
singing hector songs
narrative laments sung
achilles klea andron
narrator patroklos
kleopatra lyre

War agamemnon
achilles anger
apollo kleos zeus
angry achaeans
athena god iliad
goddess king

Timeline week
hours starting
june beginning
weeks end
schedule date
april

Spanish hola
mexico saludos
madrid desde
spain costa
barcelona
argentina

Feeling thought
feel agree means
human thinking
comment feelings
mind

Reading read texts
iliad homer odyssey
poetry works book
nagy literature past
fast poems homeric

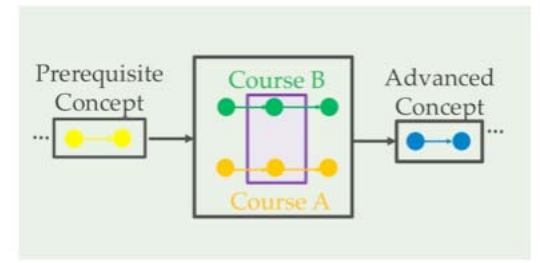
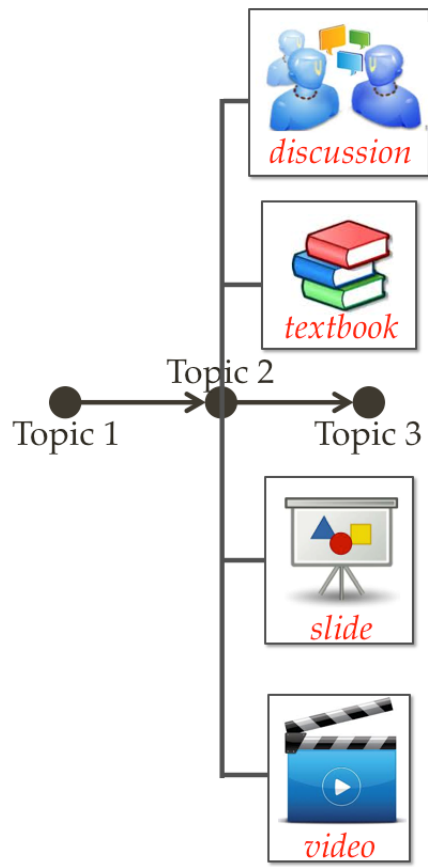
Heroes hero gods
human heroic man
humanity power fight
born strong actions
humans values notion
modern god

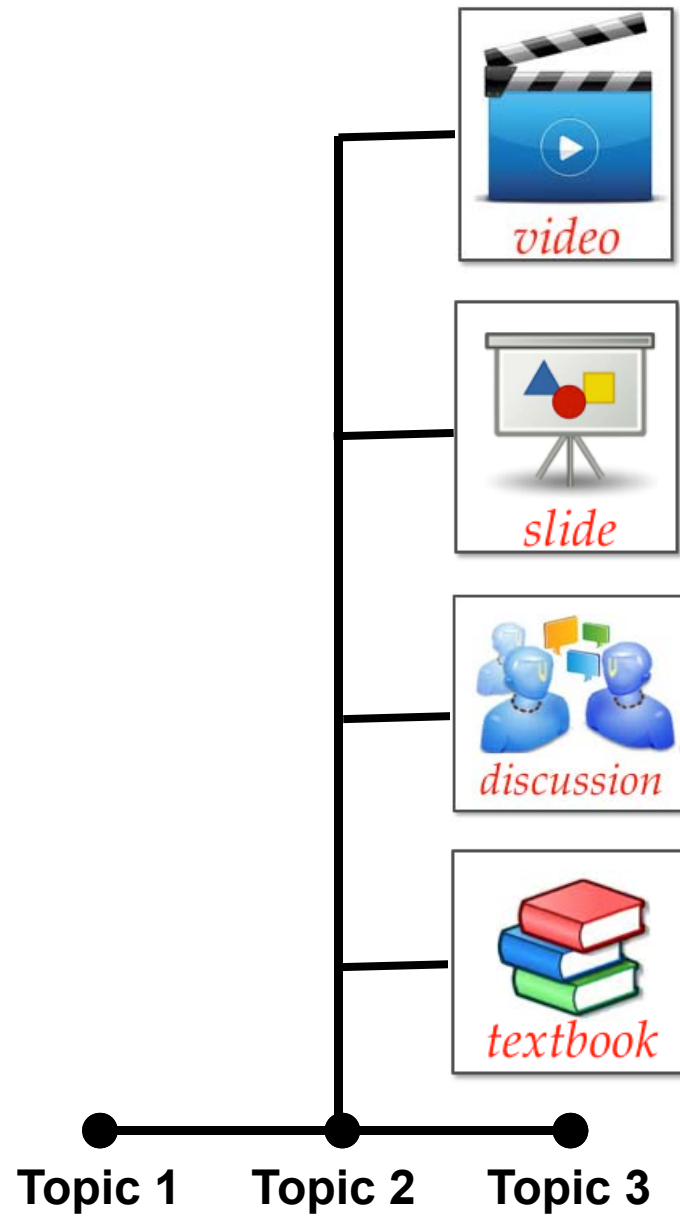
Edx online facebook
courses free mooc
enrolled joined
university global
education
educational friend
request fb

Athens greeks
greece thessaloniki
maria eleni naxos
hellas crete eirini
tirnavos evangelia
hallo island

Death lament thetis
muses funeral
nereids laments
dead die goddess
grief lamentation
lamenting

Pain achilles son
mother iliad life
sorrow hero kleos
epic patroklos
homer memory
sisters pindar



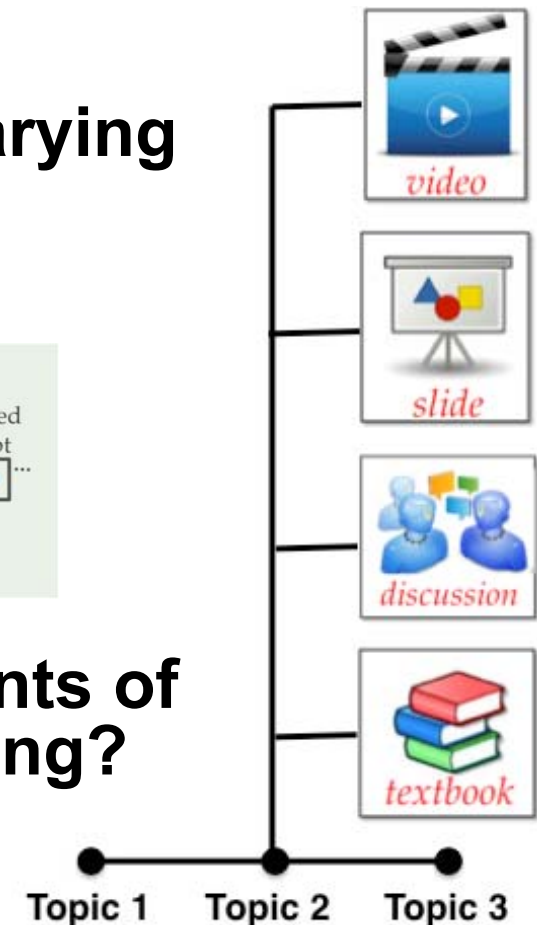


Three Related Efforts

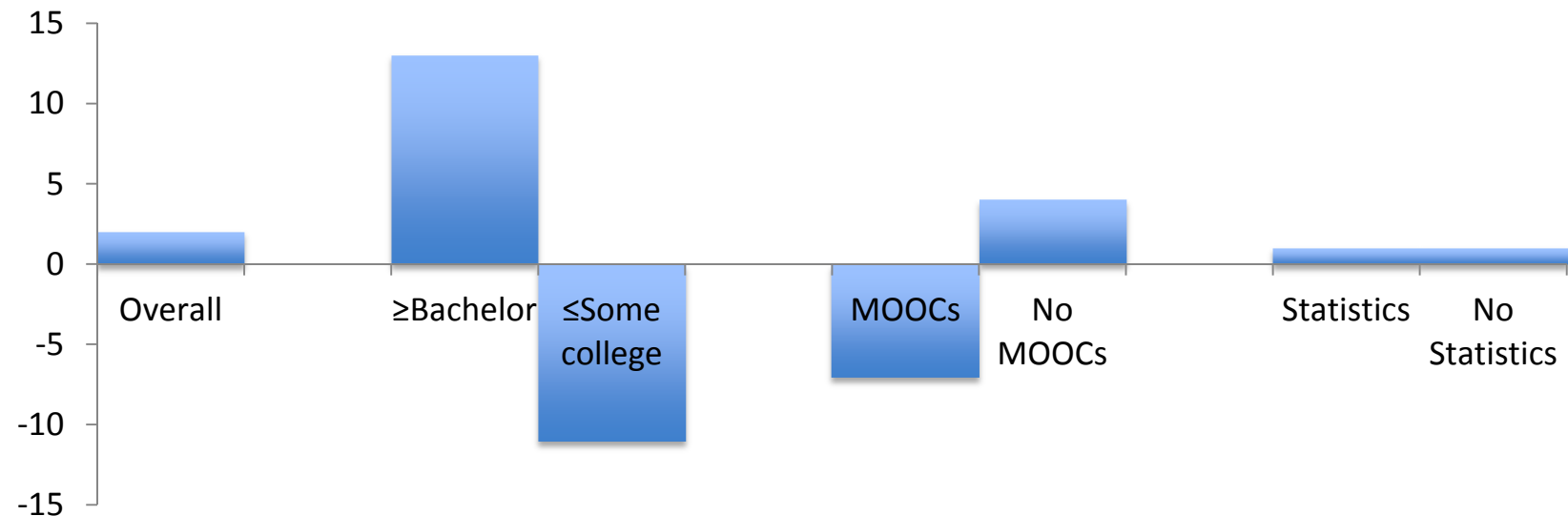
1. **How can different threads of on-line discussion be categorized?**
 - Topic Modeling (Jingjing Liu; May, 2013)
2. **How can we provide a choice of learning paths for students with varying backgrounds?**
 - Learning Path Discovery (Hung-yi Lee)



1. **How can we link the various contents of an online course to enhance learning?**
 - Multimedia Content Linking (Daniel Li)



Information Search (Accuracy)

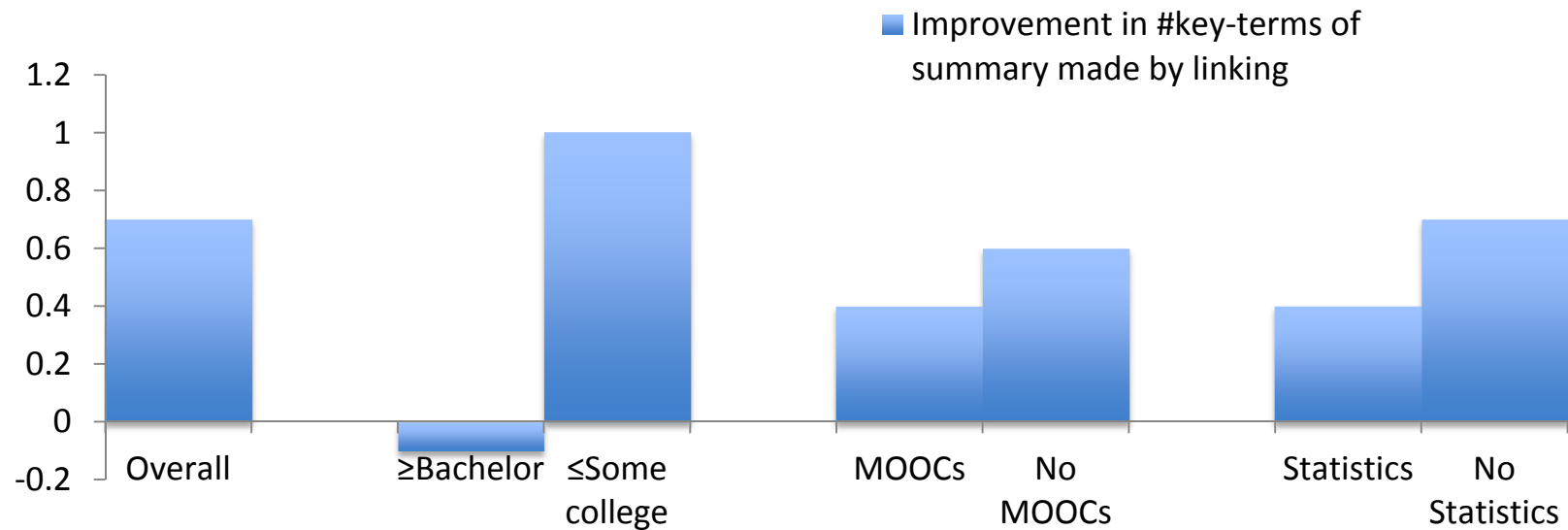


Significance
($p=0.05$)



- **Linking doesn't improve accuracy (except in one case)**

Concept Retention



Significance
($p=0.05$)



- **Linking doesn't improve retention (except in one case for novice)**