Indexed Captioned Searchable Videos

Jaspal Subhlok

University of Houston
Department of Computer Science

Outline

- Technology and Learning (& MOOCs)
- Motivation for Employing & Enhancing videos
- How does it work: Demo
- Building It: The Technology
- Using It: Deployment & Lessons from Surveys
- Wrap Up

[The opinions in this talk are those of the speaker and do not represent the official view of Department of Computer Science or any funding agency]

Technologies Influence Education

1. (1480s) **The Printing Press -> Textbook**
   - + Fears of Professors were misplaced (settled in ~20 years)
   - + for education quality (not sure about memory capacity)
2. Computer, Viewgraphs/Powerpoint Lectures
   - + Professor & Student convenience
   - Influence on education less clear
3. Clickers, Video, Social Media, Online xxx, Tablets, Mobile Devices, Wikipedia,….
   - Not clear where this is heading
   - Model of lecturing+homeworks+exams intact so far!

What About MOOCs

Presented as an alternative model, but

- Motivated by cost reduction, but no convincing business model or integration with classrooms
- Evolution of distance learning and Evolution of textbooks
- Very valuable in many scenarios but overhyped
- Distraction from other technological developments with potential for impact

Drivers of Change ?

- Learning literature says “students not learning much” (critical thinking, reasoning, writing)
- Financial squeeze as governments reduce/stop funding education. More like a business.

Education enterprise will change with maturing of new technologies and approaches (e.g.“Inverted/Flipped” classrooms)

- Events like shutting/scaling down of (smaller) universities more likely than ever!
- Combination of technology and external pressures

ICS Videos

Research project to enhance the value of video as a learning resource employing automated (where possible) Indexing, Search and Captioning.

Motivation for ICS Videos Project

- University of Houston has been a leader in supporting video for coursework (NSM IT)
  - Tablet PCs to teach and record lectures and make them available online as study material

Surveys showed videos are a powerful, versatile learning resource

Survey Results - 1

- **N=2,349** taken from 43 sections in biology, computer science, geology, chemistry, and mathematics/physics between 2009 to 2011.
- Course sizes varied from 8 students to >300.
- Each professor posted ~25 videos per semester

These surveys are recent but the basic results have been similar even before this project started

Survey instrument details are often skipped in this talk to save time – available in papers.

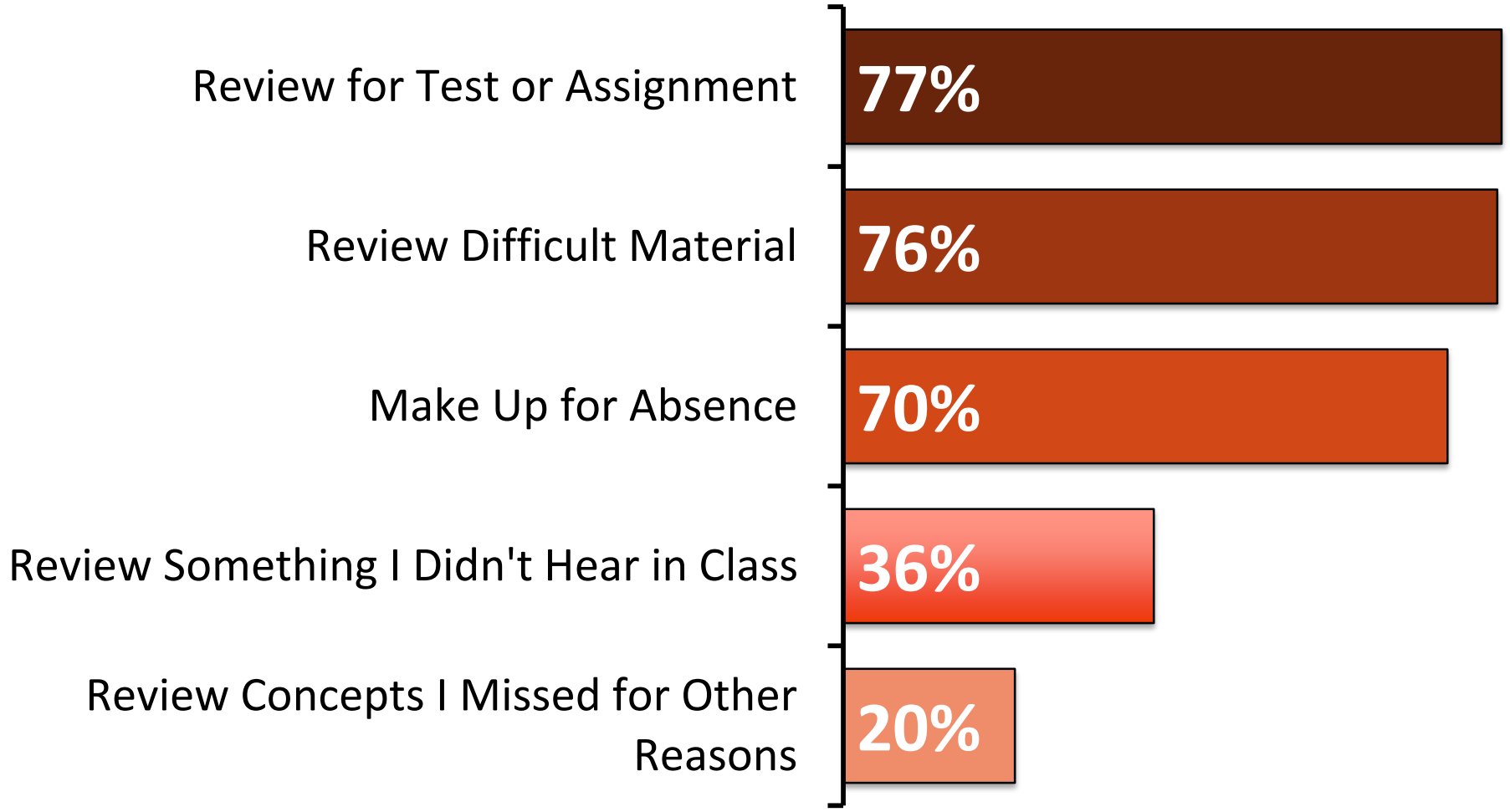
Students Watch Videos

84%
used lecture videos at least once

81%
watched the entire video, not only a part they needed

47%
watched a single video more than once

Reasons Students Used Videos



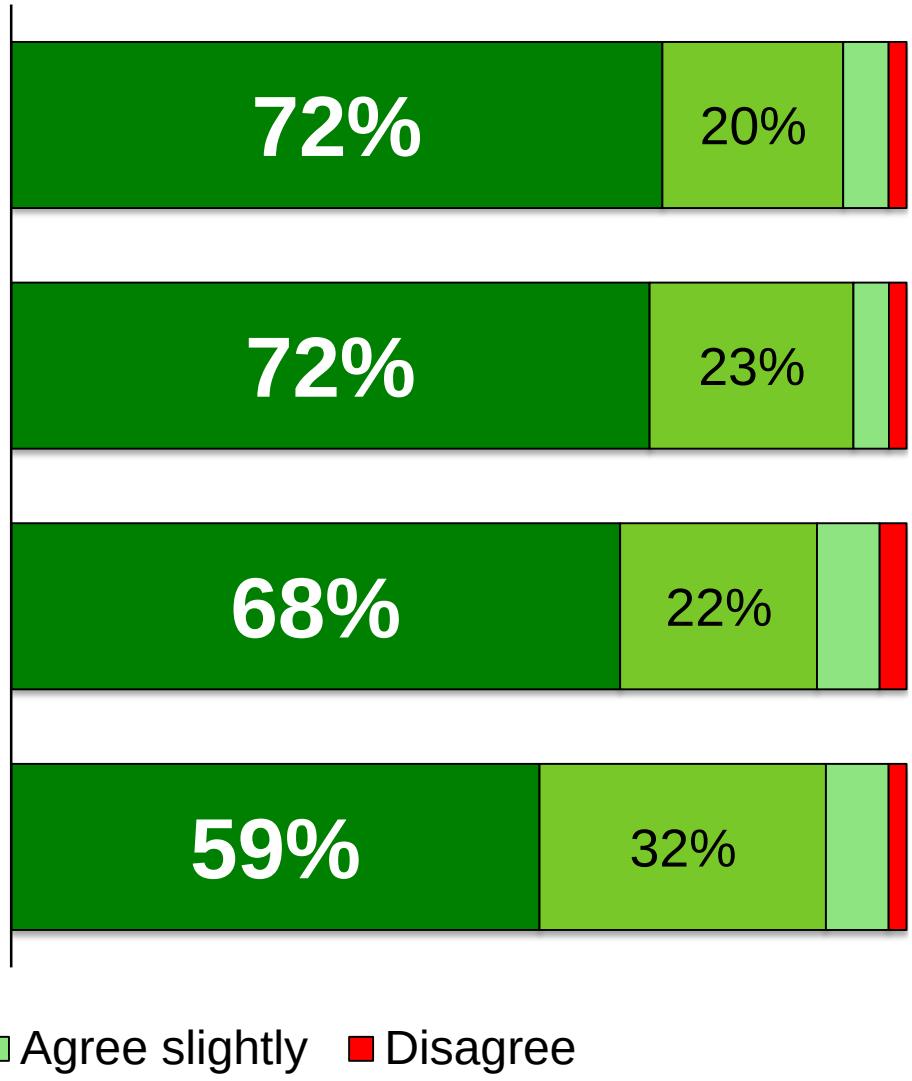
Students Strongly Value Videos

"Having access to lecture videos for this class is important to me."

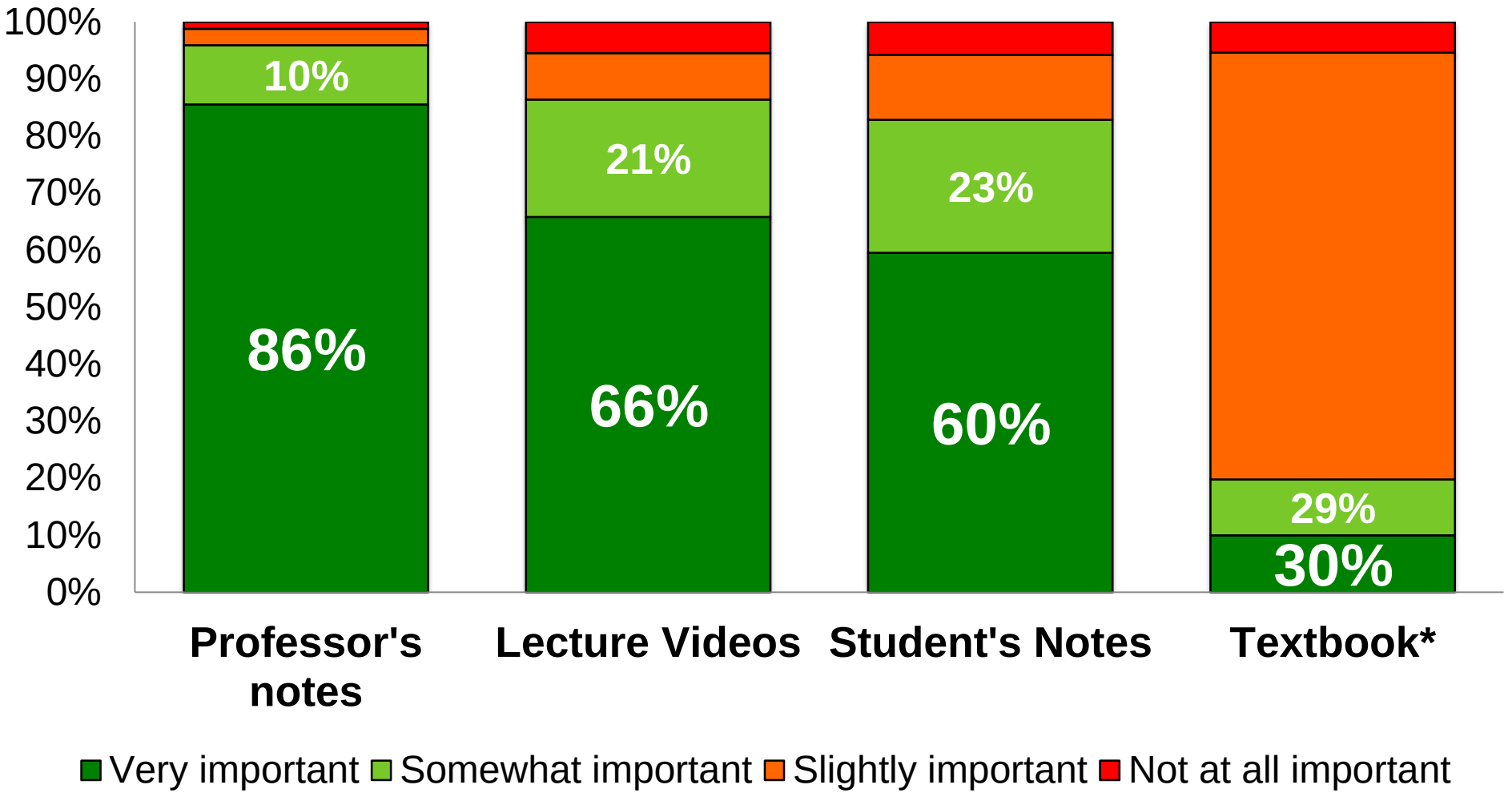
"Lecture videos are useful for reviewing."

"The lecture videos helped me to study for quizzes or tests."

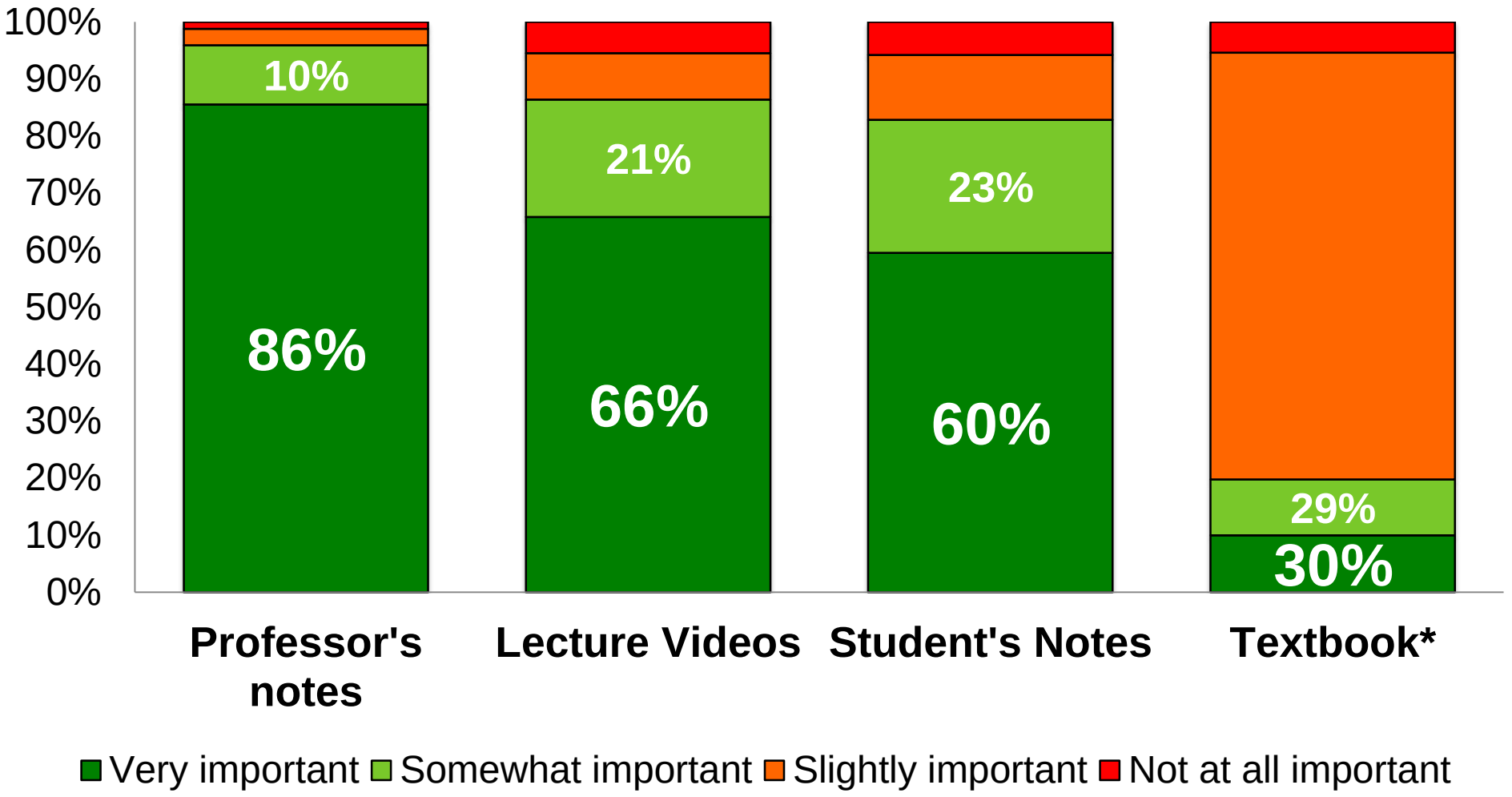
"Lecture videos help me to clarify material that was not clear in class."



Comparison with other Resources



Comparison with other Resources



Motivation (continued)

Videos are valuable but:

- Key shortcoming of video format is the inability to quickly access content of interest
  - Loud and clear in surveys and interviews (not shown)
  - Students wants answers to questions for review, not watch an hour long video!

Goal: ICS video player with advanced Indexing, Captioning, and Search

A Project to bring clear and present benefits to students with cool computing research on the way!

ICS Videos


Computer Science - Introduction to Computing - COSC 1300 Lecture 21
Help
On
Transcript
0:05 Okay so I need 3 more volunteers for today to give your powerpoint.
0:10 presentation. Let's see you are Shannon, are you Shannon?
0:38 Jasmine, okay so you're in the third. Jasmine Scott.
0:57 Whats your name now? Your last name is?
1:35 okay I got 2, I need one more? We'll go with these two okay. First one is going to be Jasmine Scott here.
2:38 So Jasminen is going to talk to us about TranSwitch: Engines for Global Connectivity. Hello everybody I'm Jasmine. The company I chose was TranSwitch.
2:56 Basically its a company, they make integrated circuits and intellectual property solutions.
00:00:00
01:02:08
Speed 1x
quiz
00:01
04:05
06:33
15:55
23:05
25:51
31:01
35:31
quiz(1)
quiz(1)
quiz(9)

Building It – Challenges

Search: Keyword search Inside video

- OCR on video frames
- Semantic search

Indexing: Divide a lecture into topic segments

- Identifying topic changes
- Images, text, or audio?

Captioning:

- Speech recognition ineffective for classroom videos

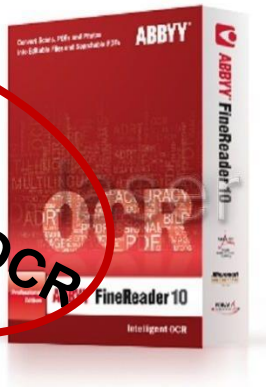
Keyword Search

- Keyword Search requires text detection in video frames
- Can be accomplished by OCR tools

Google GOCR

Adobe

TESSARACT OCR

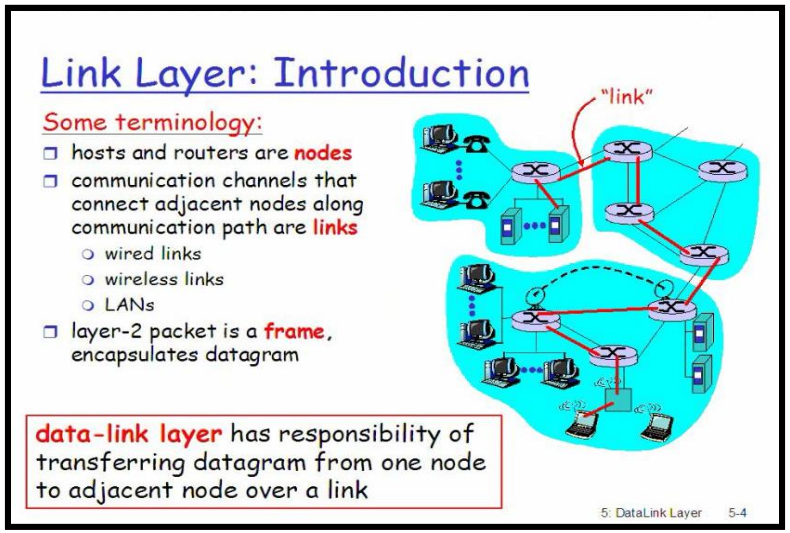


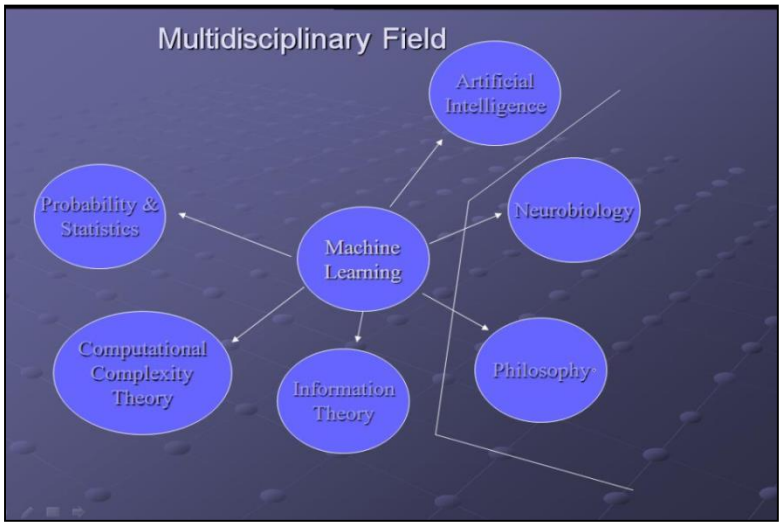
Microsoft Office

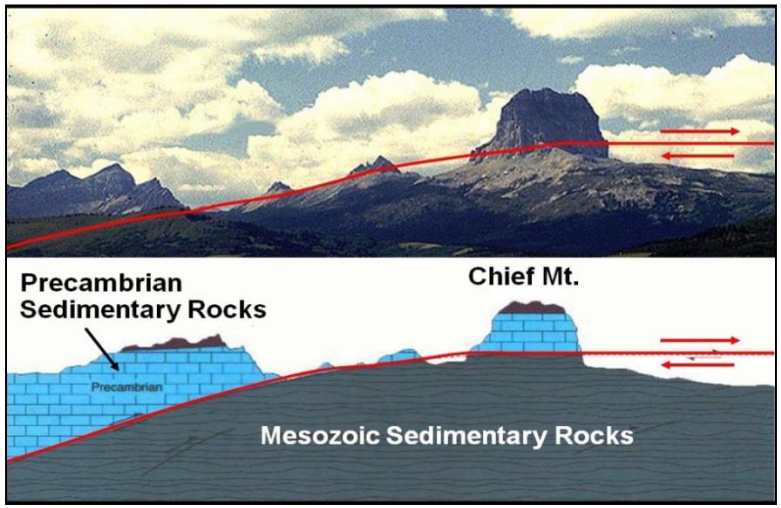
- Accuracy on lecture video images?

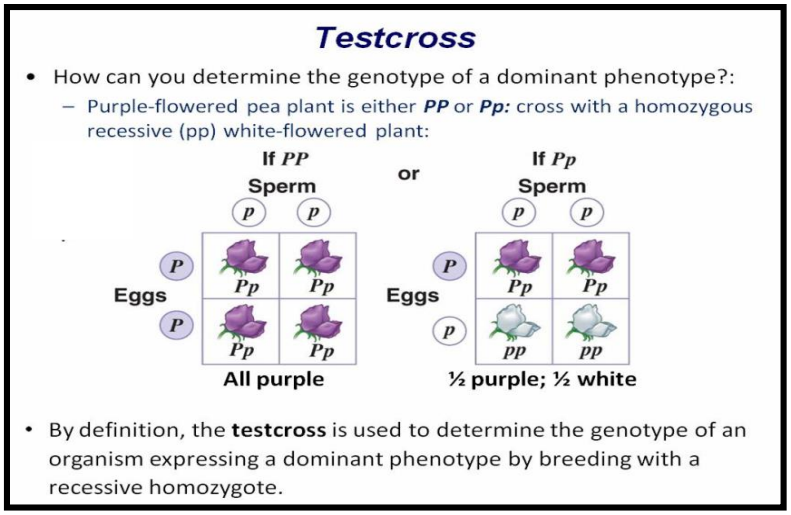
Images from Lecture Videos

- OCR: good for scanned images
- Lecture video images with colorful different layouts

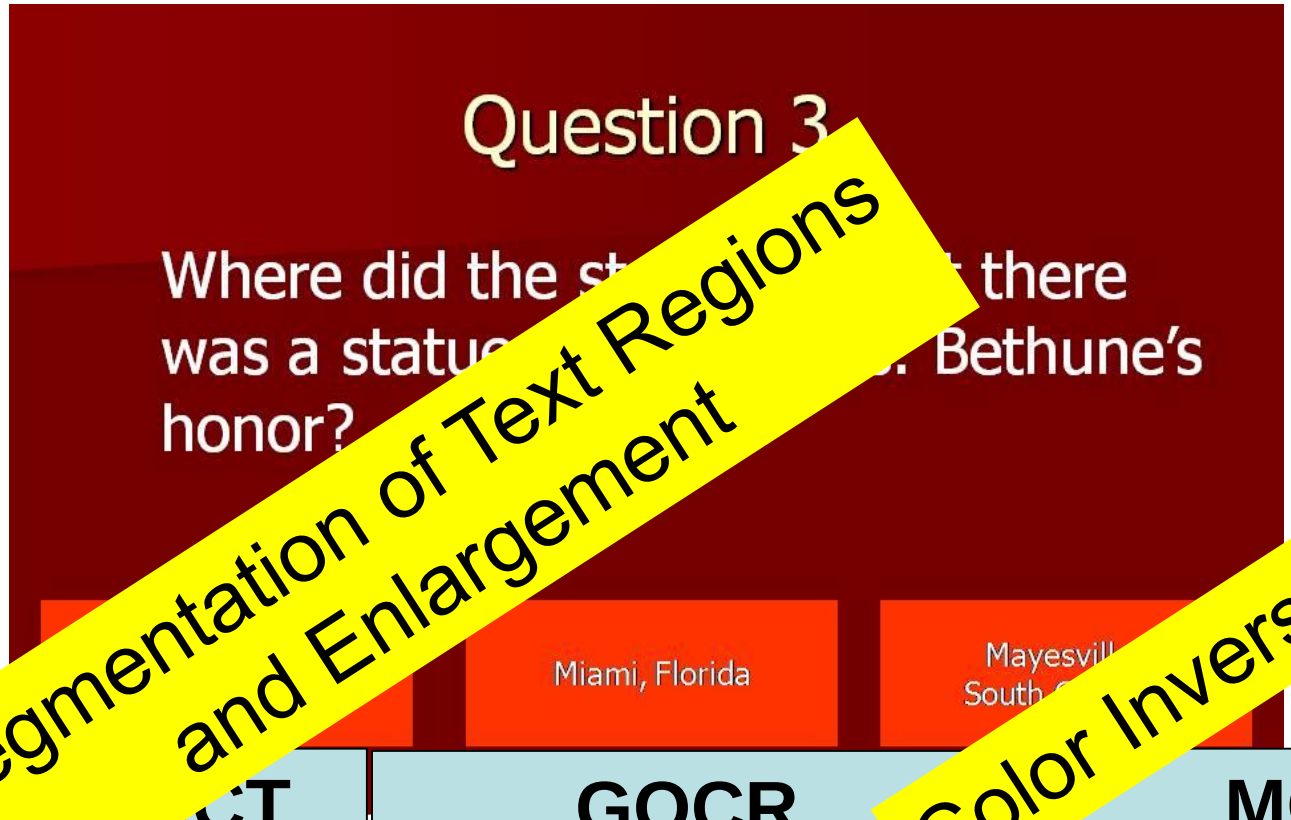








Output of OCR Tools

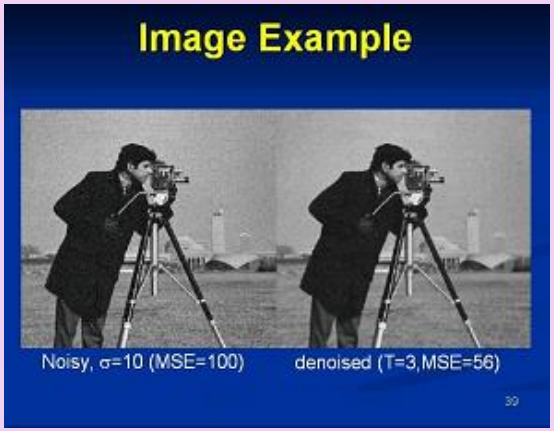


Segmentation of Text Regions and Enlargement

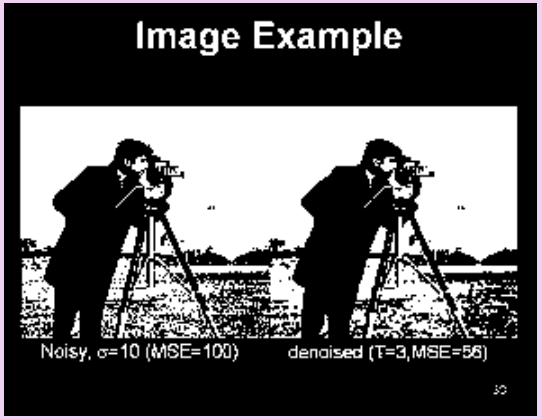
Color Inversion

T CT	GOCR	MODI
Questio Where did the story say that there was a statue raised in Mrs. bethune's honor? B-Nik is	Where did the stov say that there was a statue raised in Mrs. Bethune's honor? ;_n ton Dc , m._am._ F_or.__a Mayetvil South0rolina	Where did the story say that there was a statue raised in Mrs Bethune's hon Washinton, D.C.Miami, Florida Mayesville, South Carolina

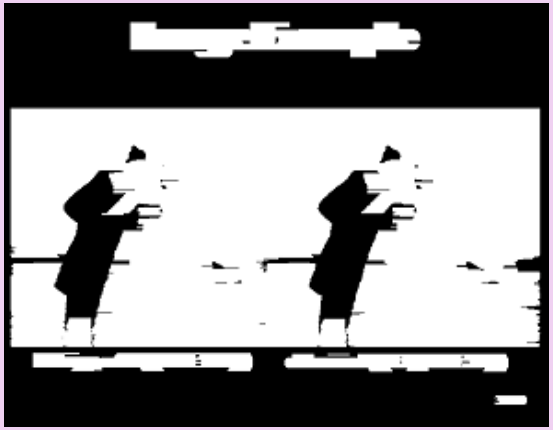
Segmentation of Text Regions



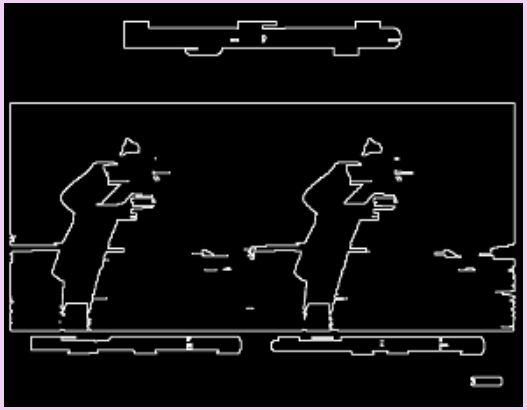
a)Original image



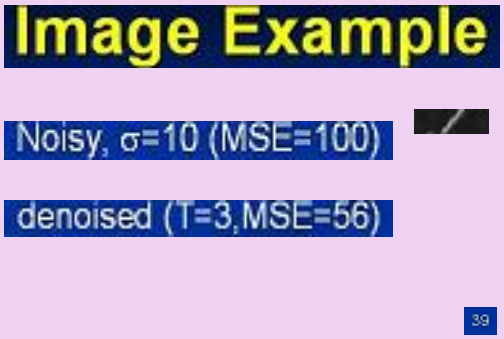
b) Binarization



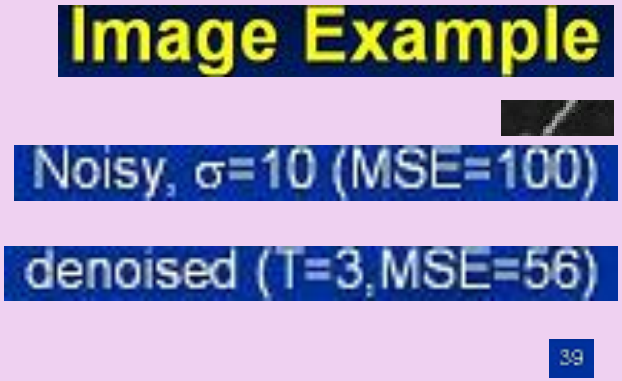
c) Dilation effect



d)Edge detection



e) Blob extraction



f) Enlargement

T. Tuna, J. Subhlok, S. Shah "*Indexing and Keyword Search to Ease Navigation in Lecture Videos*" , AIPR 2011

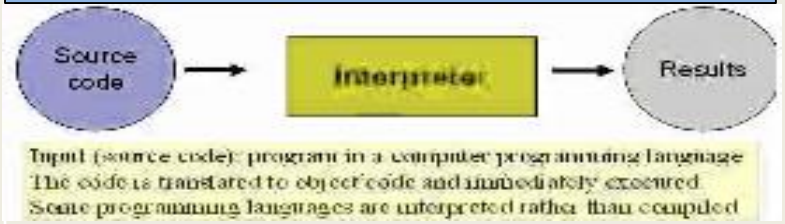
Color Inversion

Original Image R / G / B

Inversion1 255-R / G / B

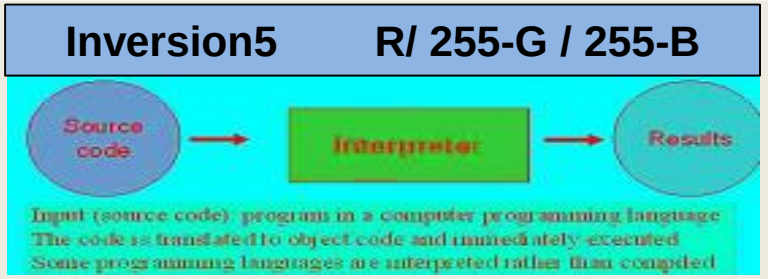
Inversion2 R/ 255-G / B

Inversion3 R/ G / 255-B



Inversion4 255-R/ 255-G / B

Inversion5 R/ 255-G / 255-B

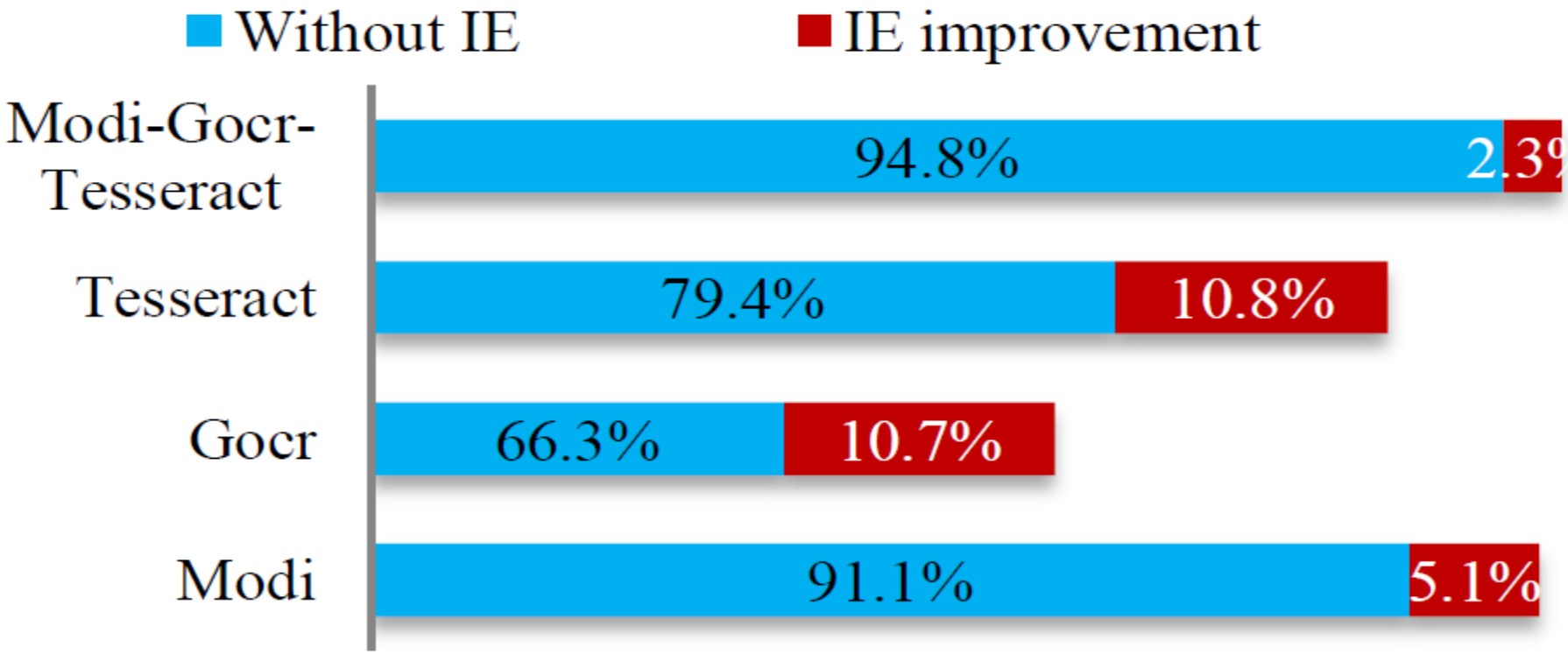


Inversion6 255-R/ G / 255-B

Inversion7 255-R / 255-G / 255-B

Impact of Image Enhancement (IE)

Accuracy = %age of words detected correctly:

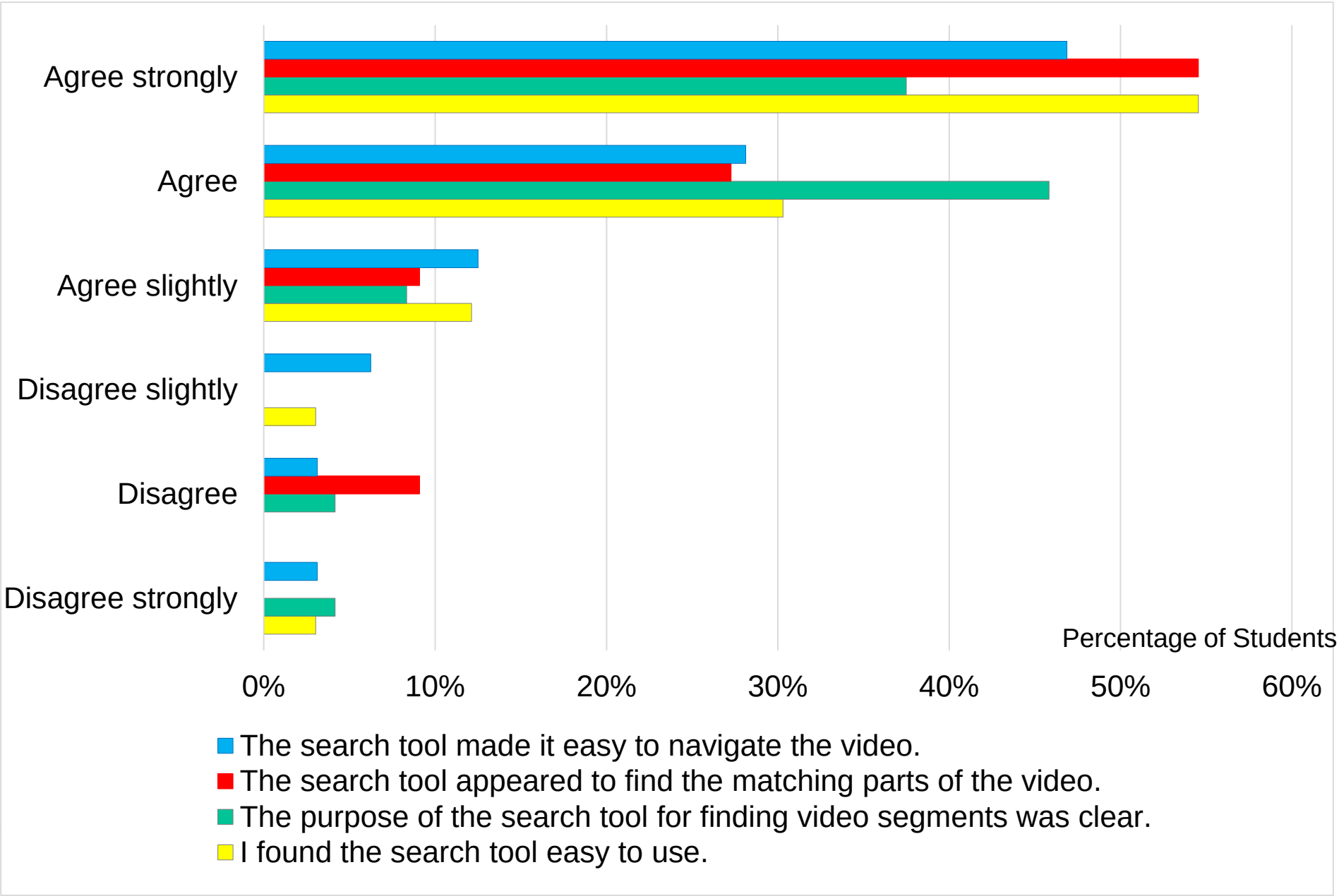


Many false positives – *not a major issue for Search*

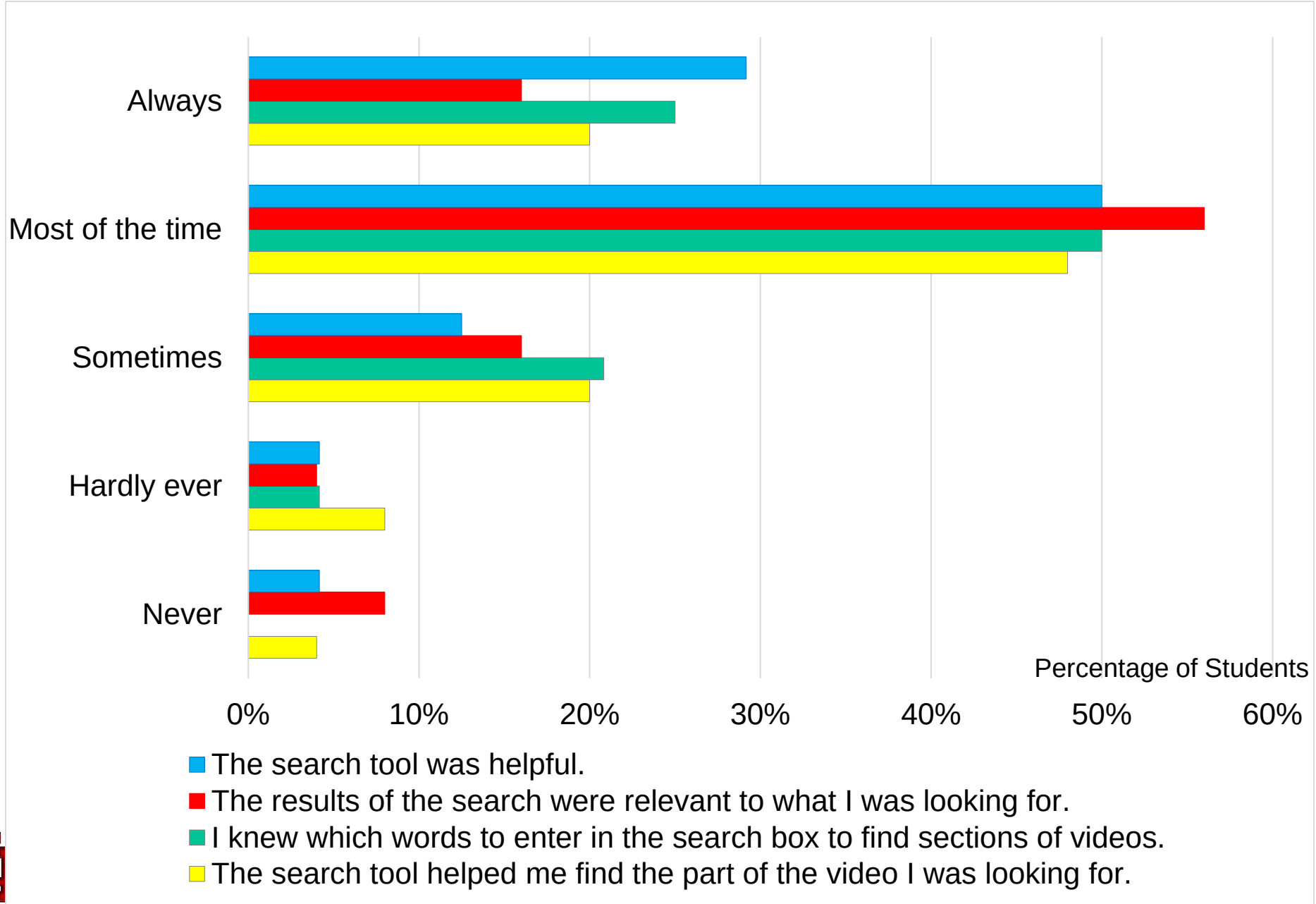
Value of Search

Results from Survey
of Student Users

Survey-Search (1)



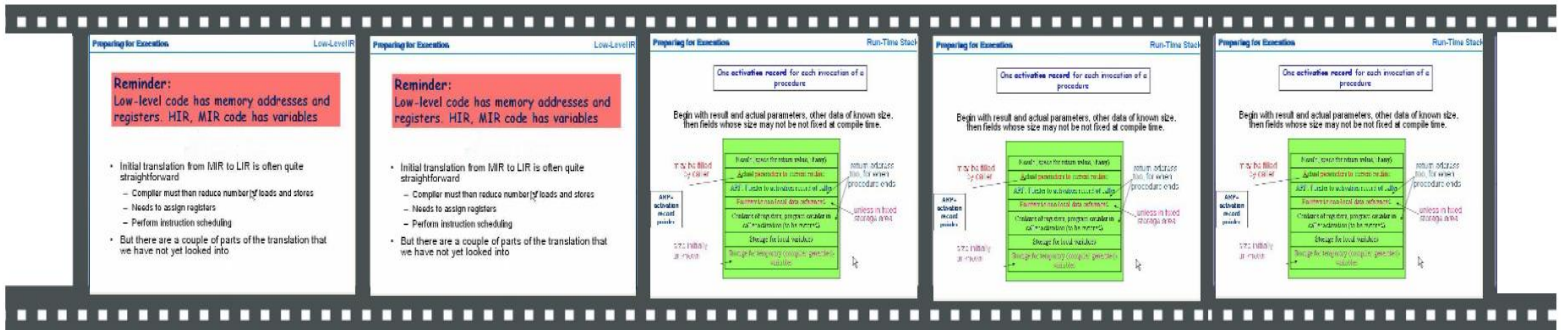
Survey-Search (2)



Video Indexing

Objective : Split video into meaningful, topical segments

1. Identify *Transition Points (TP)* where video scene changes (image- difference)



2. Identify subset of Transition Points that represent Topic changes!

Methods for Video Indexing

Split video in equal intervals of time

- Lazy reference method

Split video based on image difference

- Does not work well (J. Li MS thesis)

Split video based on text difference

Indexing by Machine Learning

Text Based Indexing Algorithm

Input:

A list of transition points

Required number of index points

Output:

List of index points

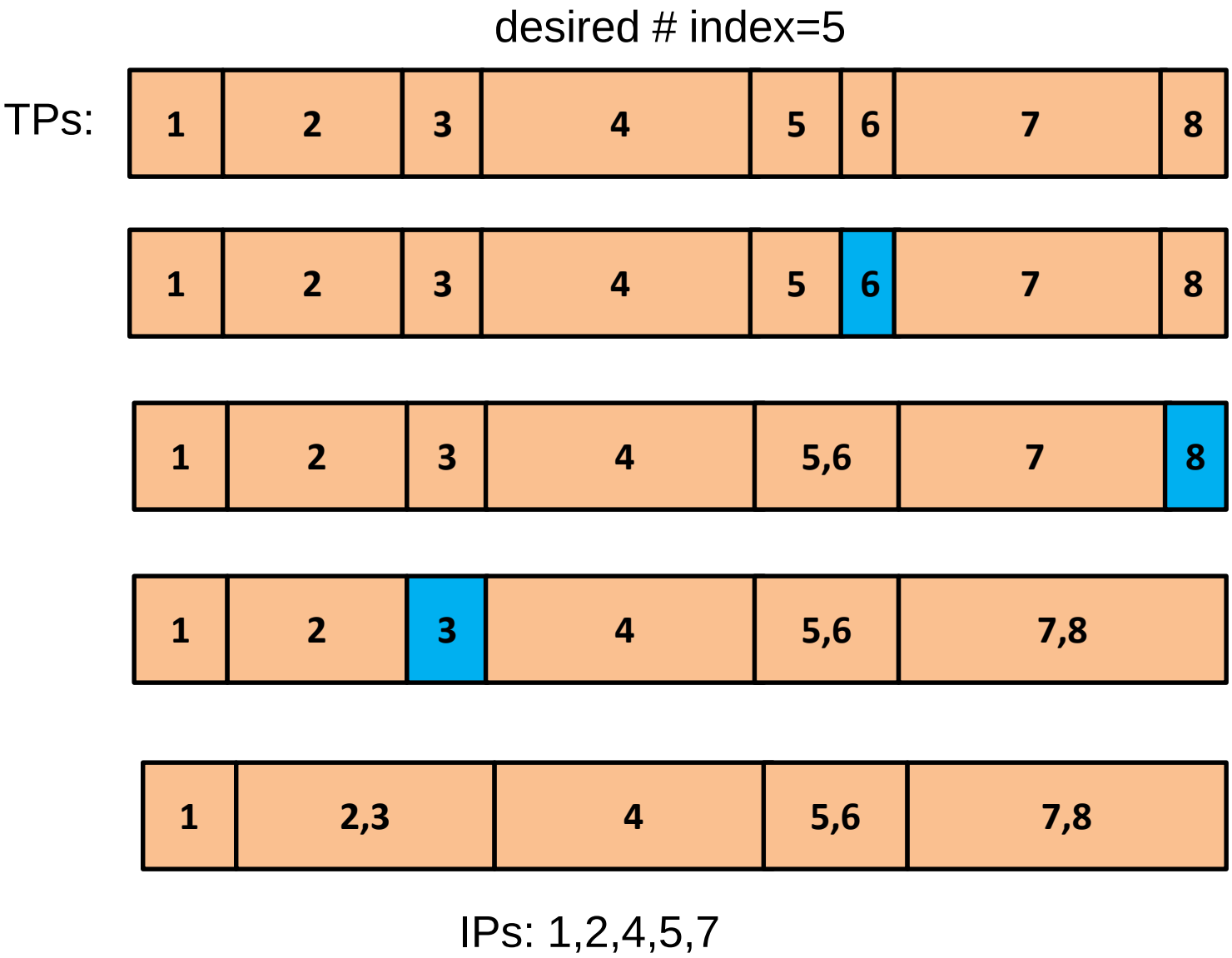
Repeat:

1. Select transition point/segment with smallest duration
2. Merge it left or right neighbor **based on text similarity**

Until

Reach required number of index points

Uniform Indexing Algorithm



Text Similarity Metric: Cosine

Frame 1

Birds fly.
Birds chase fish.

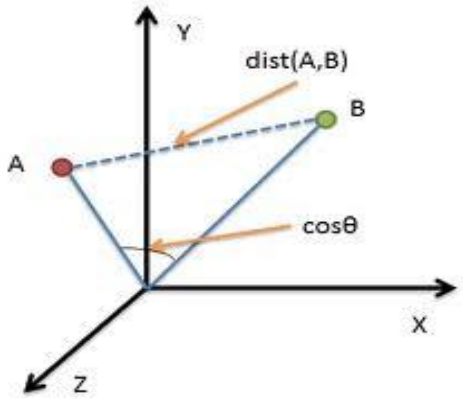
Frame 2

Cats chase birds.
Cats eat fish.

Frame 3

Dogs swim.
Dogs chase cats.

Word/Frame	Frame1	Frame2	Frame3
birds	2	1	0
cats	1	2	1
chase	1	1	1
dogs	0	0	1
fish	1	1	0
fly	1	0	0
swim	0	0	1



$$cos(\theta) = \frac{A.B}{\|A\|.\|B\|}$$

Cos(frame2,frame1)=0.80

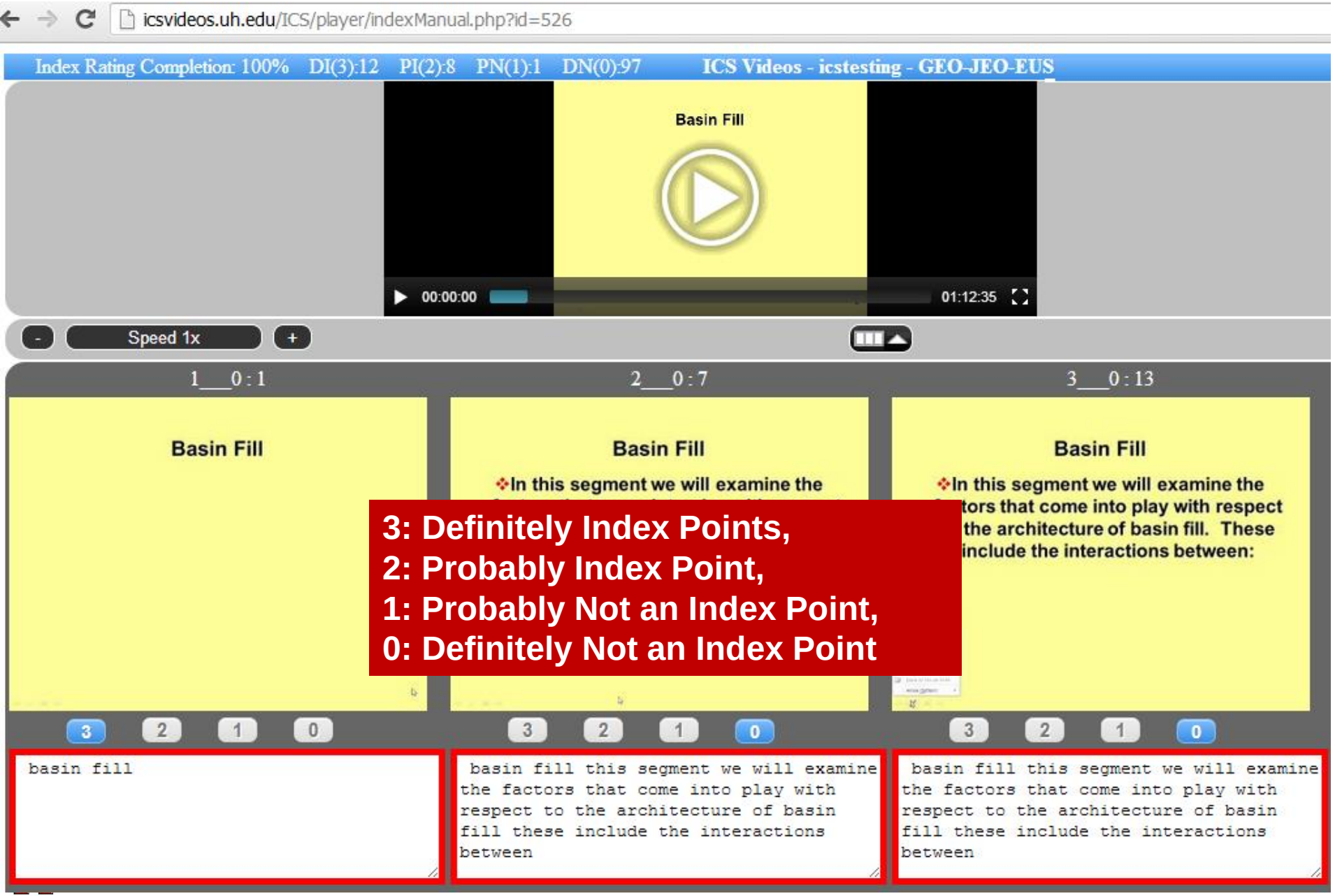
Cos(frame2,frame3)=0.57

Video Indexing Experiment

- 25 different videos
  - Computer Science 19, Biology 3, Geology 3
  - 10 different course
  - Average of 75 minutes per video,
  - Total 30+ hours of video

- 3 Manual Ground Truths
  - 1700 transition points

Interface to Identify Ground Truth



Indexing Accuracy Metric

Manual Ground Truths

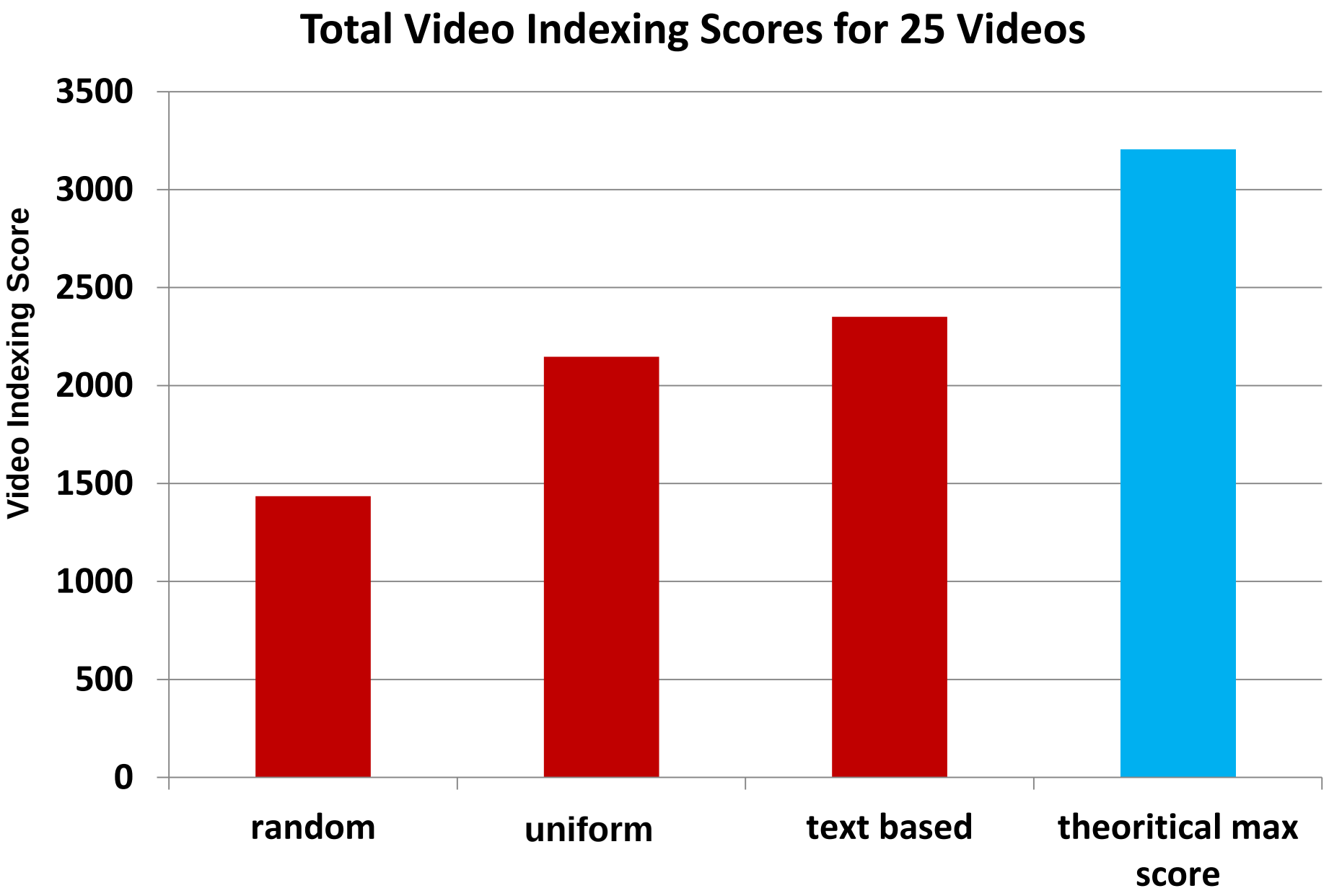
Algorithm Output	Definitely Not IP	Probably Not IP	Probably IP	Definitely IP
0 (Not IP)	(+2)	(+1)	(-1)	(-2)
1 (IP)	(-2)	(-1)	(+1)	(+2)

TIS: Transition Indexing Score

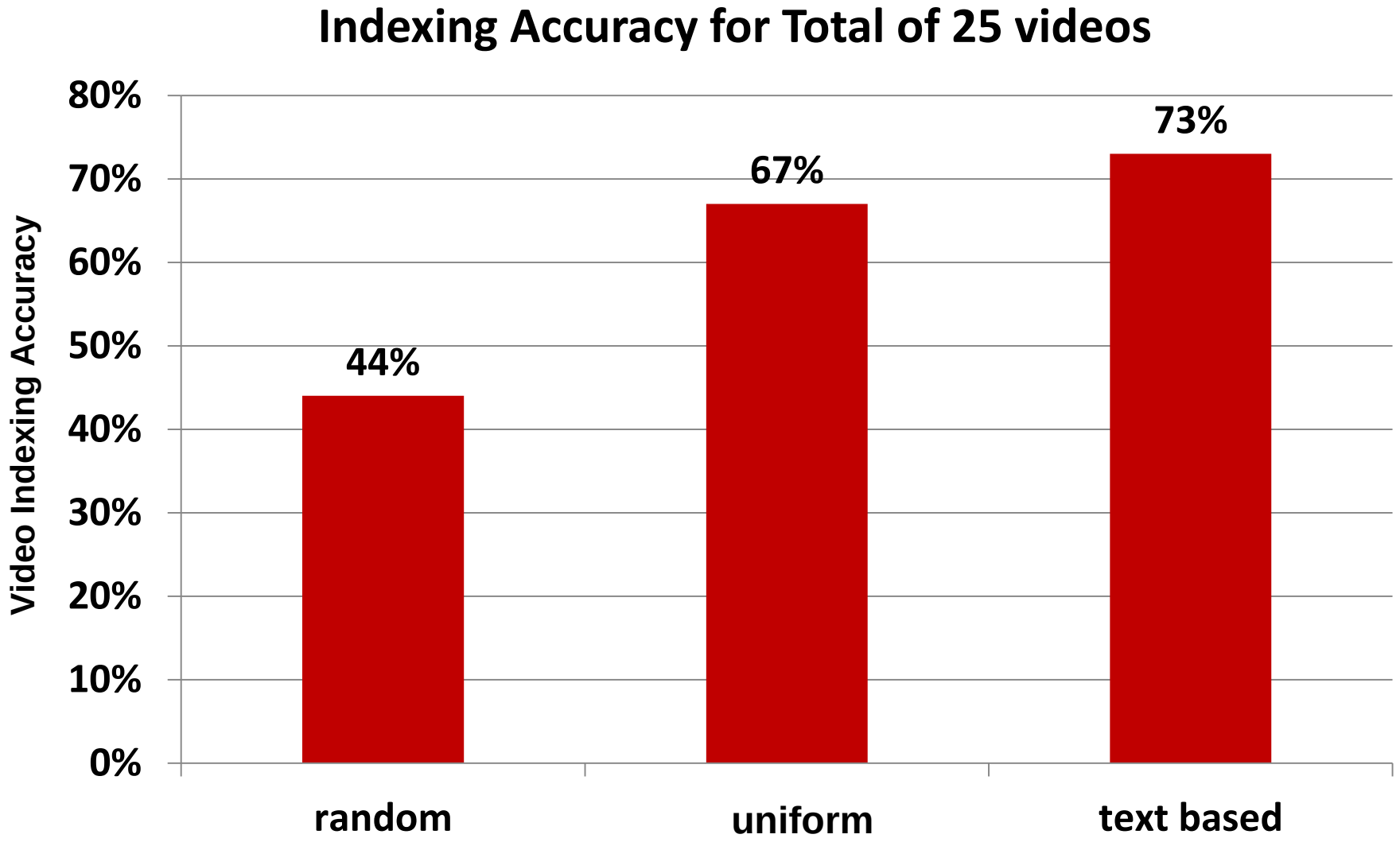
n: number of TP in a video

$$\text{Video Indexing Score} = \sum_{i=1}^{n} (TIS_i)$$

Indexing Algorithms Accuracy

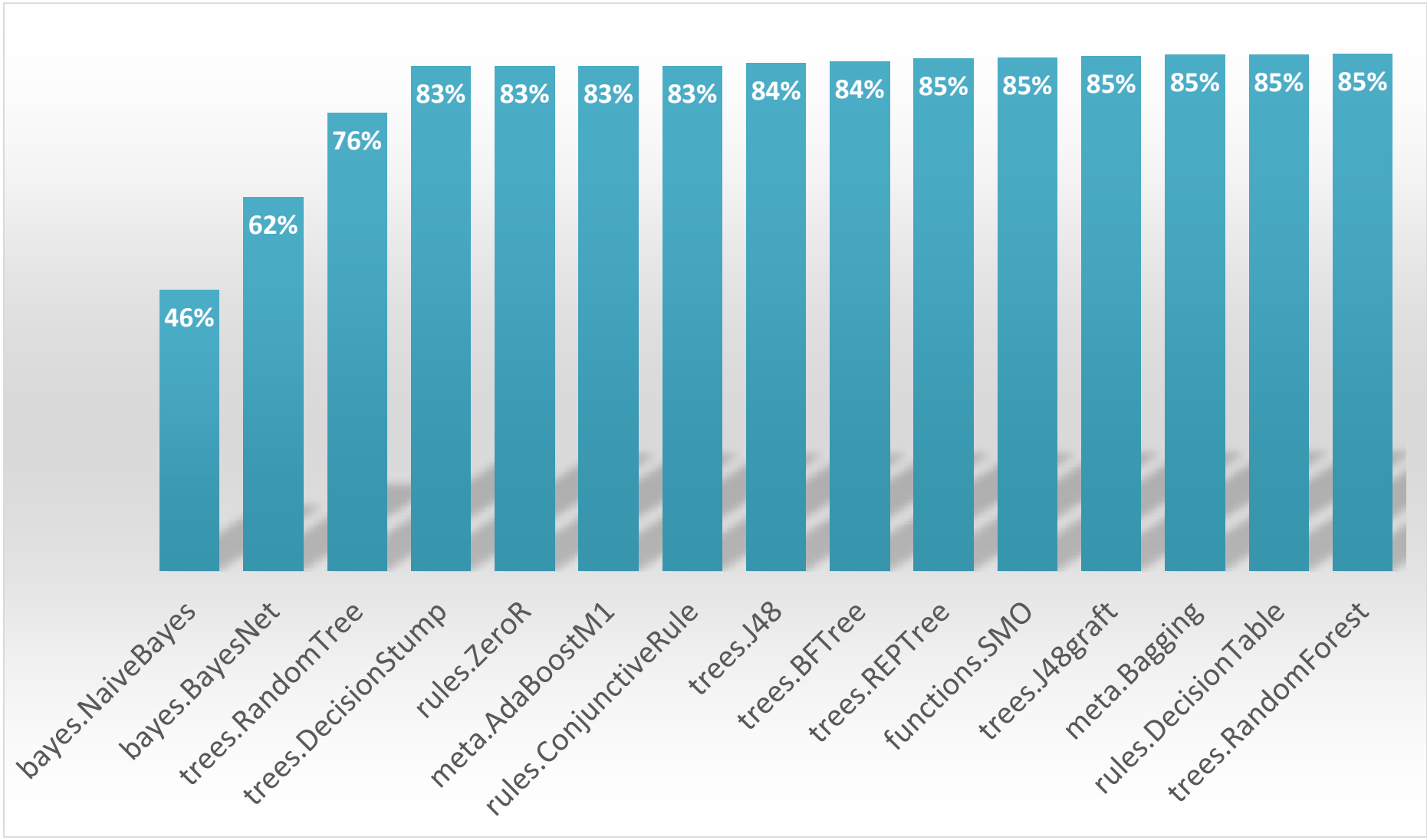


Indexing Algorithms Accuracy



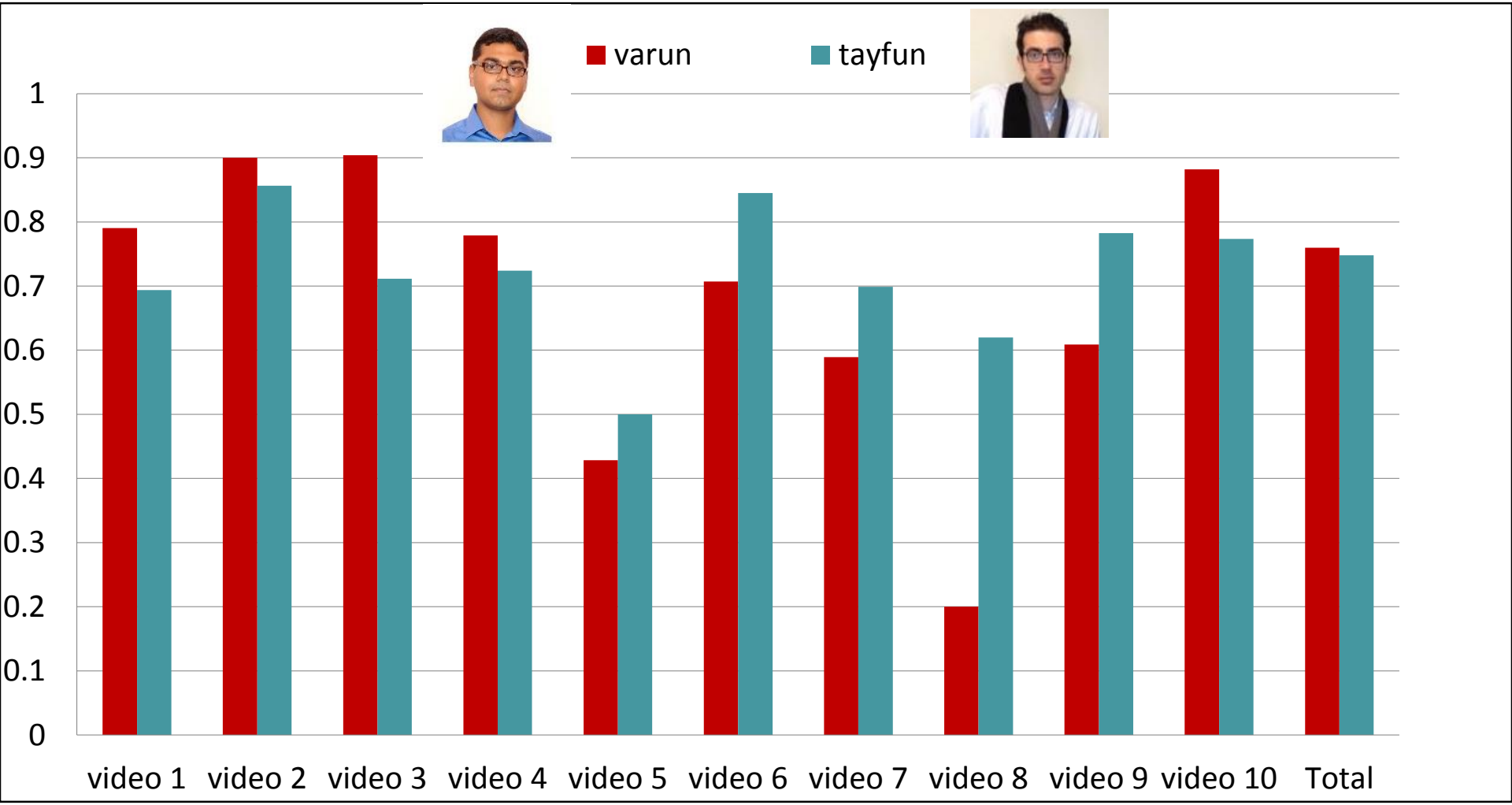
$$\text{Video Indexing Accuracy} = \frac{\text{Video Indexing Score}}{\text{Theoretical Max Score}}$$

Indexing by Machine Learning



- Promising but not fully validated! (Tuna's Ph.D. thesis)

Limits of Indexing: Humans

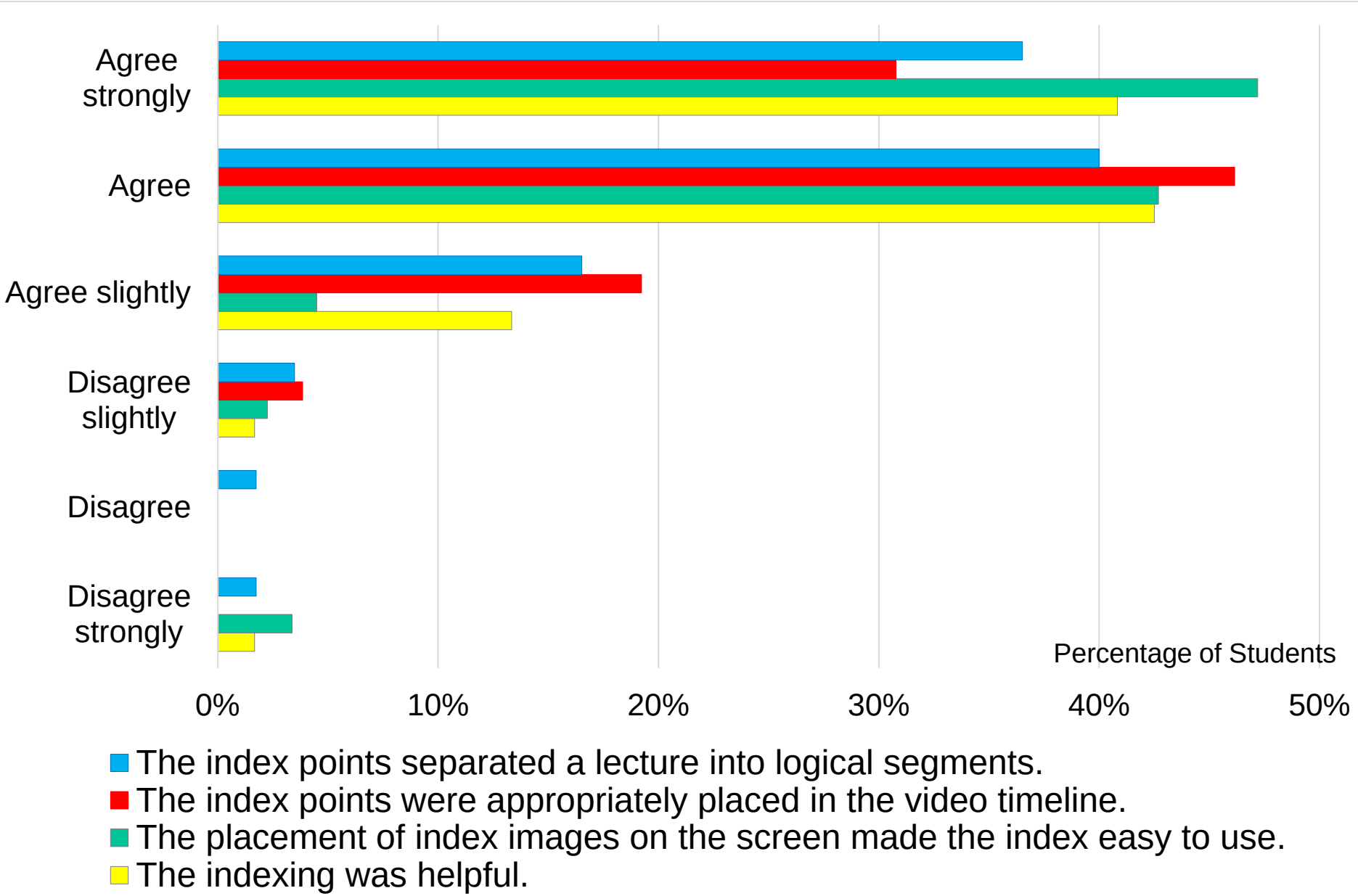


More on Indexing

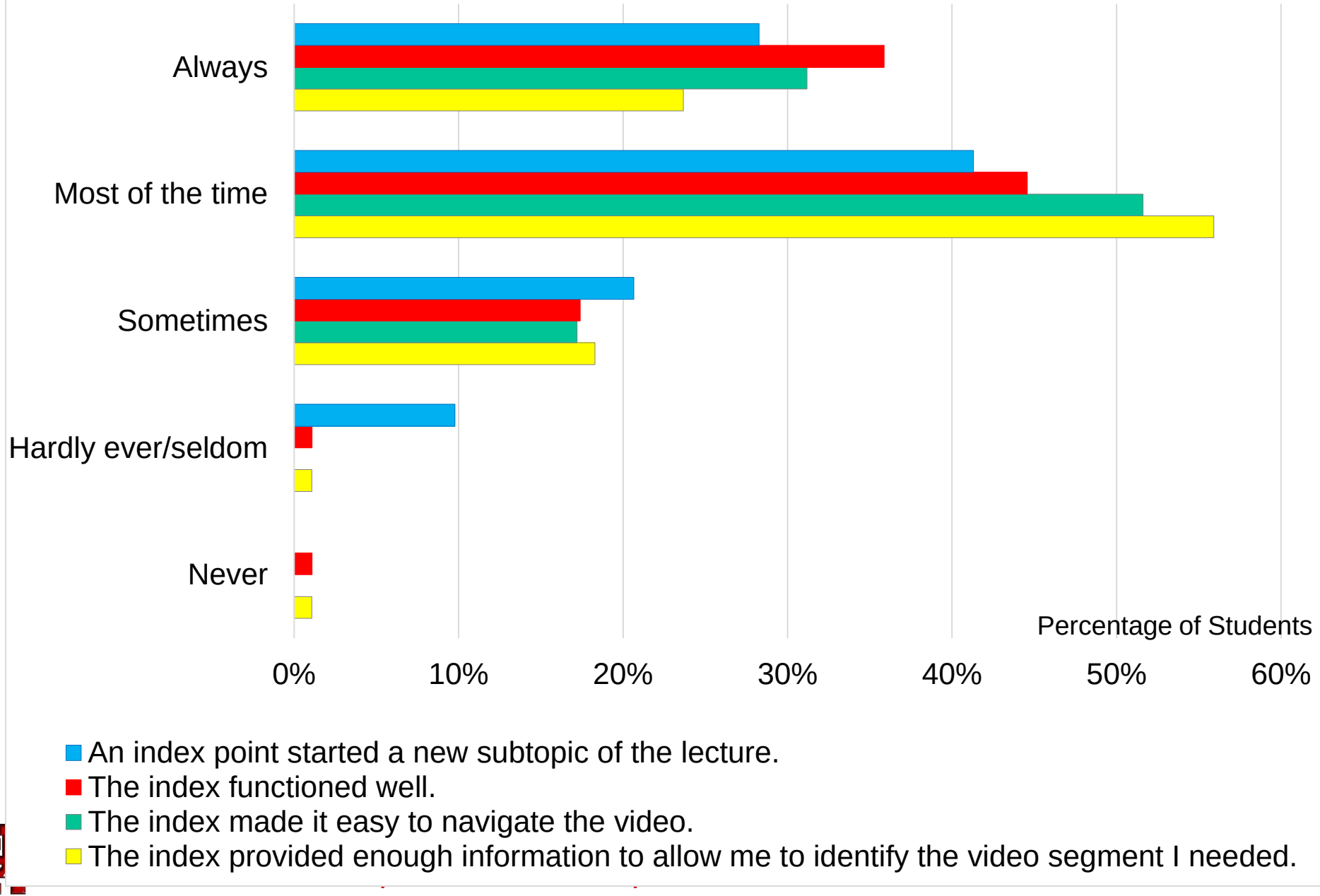
- Employment of Speech-text with slide-text being studied (M. Joshi M.S. Thesis)
- Semi-automatic user driven indexing may be a useful option

Student Surveys of Value of Indexing

Survey – Indexing (1)



Survey – Indexing (2)



ICS Videos: Captions

On screen Caption and Transcript panel

Captioning

- Captions (and transcript) are valuable for students
- Automatic speech recognition is inadequate to generate meaningful captions today
- Goal is to leverage ASR to generate captions efficiently

Automatic Speech Recognition (ASR)

Participant	Lecture Speech Recognition	Dictation	Parroting
Professor A	71.40	89.93	94.76
Professor B	62.14	83.13	96.63
Professor C	70.80	83.03	96.1
Average	68.11	85.36	95.83

Percentage Accuracy

Results shown for the best of 3 ASR tools: YouTube, Windows, Dragon

Dictation: Participants read a prescribed text that is recorded
Parroting: A trained speaker repeats the text

Normal Lectures cannot be transcribed with ASR today

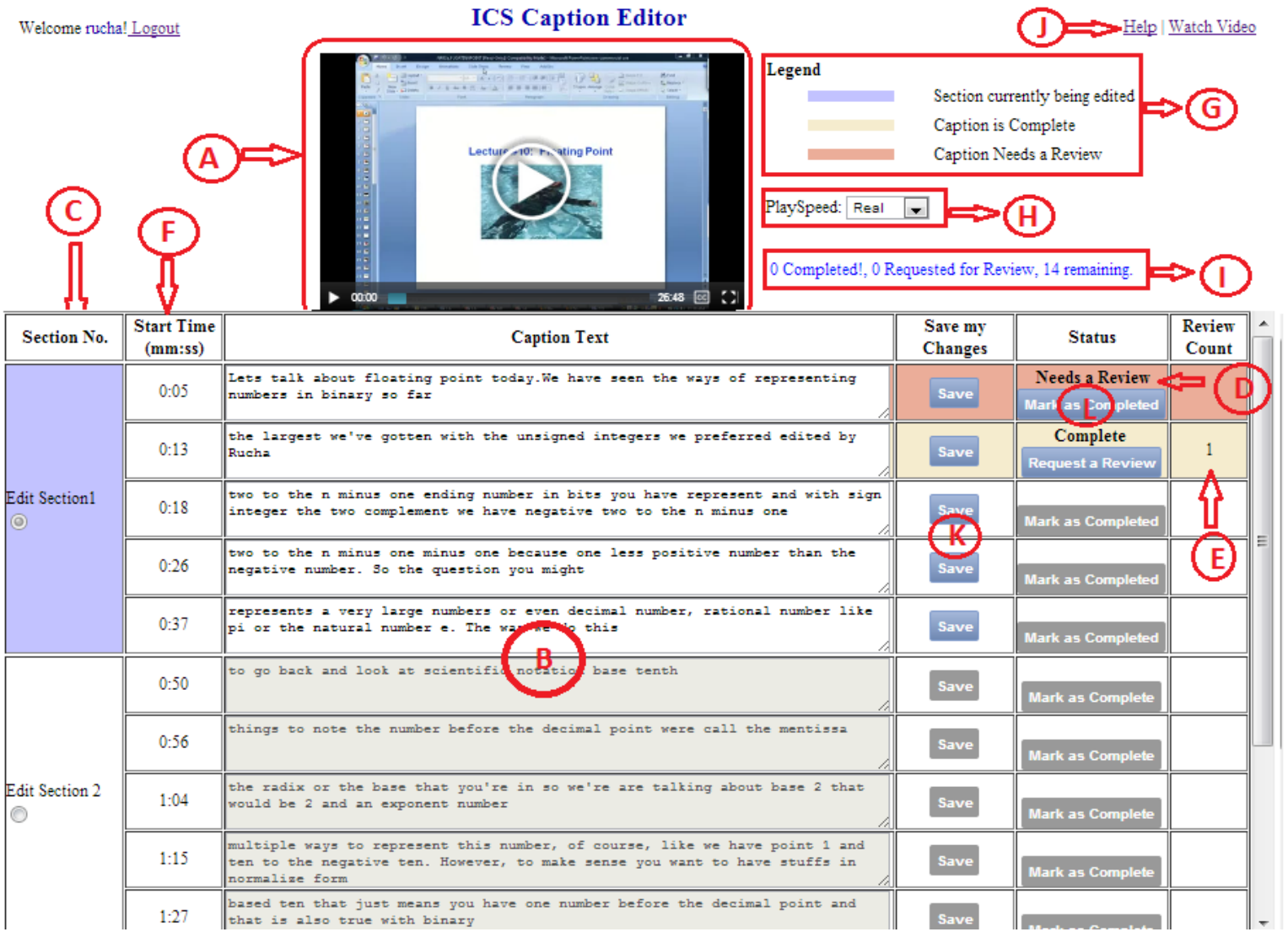
Caption Editor

Motivation: A mechanism to provide captions for STEM classroom lectures.

Design Objectives/Challenges:

- *Teamwork:* Allow co-operative/crowdsourced captioning
- *Technical content+Poor audio quality:* Captions/transcript may not be obvious
- *Easy navigation*: for non-professionals.

Caption Editor

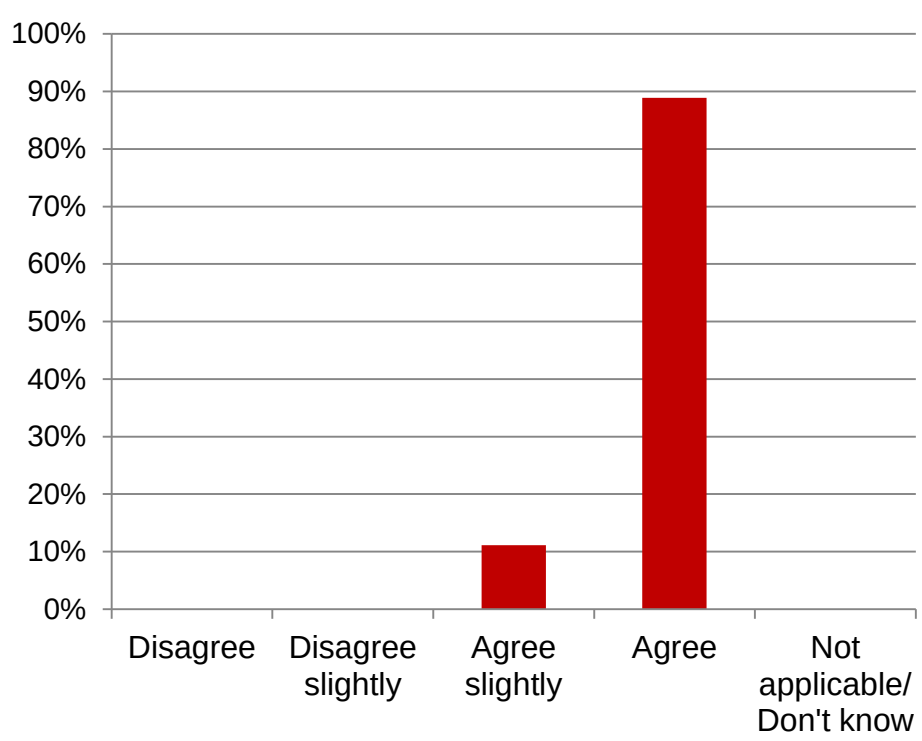


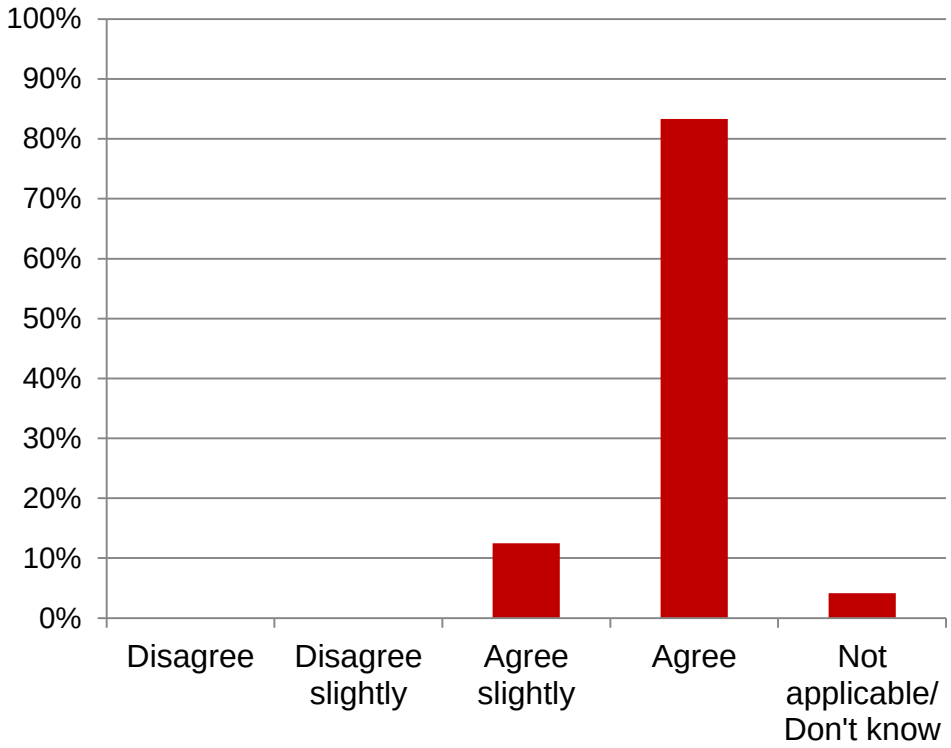
Caption Editor: Evaluation

- Deployed across 2 Computer Science courses to successfully caption selected lectures
- Work distributed across participating students. Typical example: one lecture with 11 participants
  - Each student worked on between 9 and 52 segments
  - Spent between 10 and 76 minutes over 4 days

(hour long lecture takes around 8-10 hours to caption by an untrained person)

Captions Helpful & Valuable



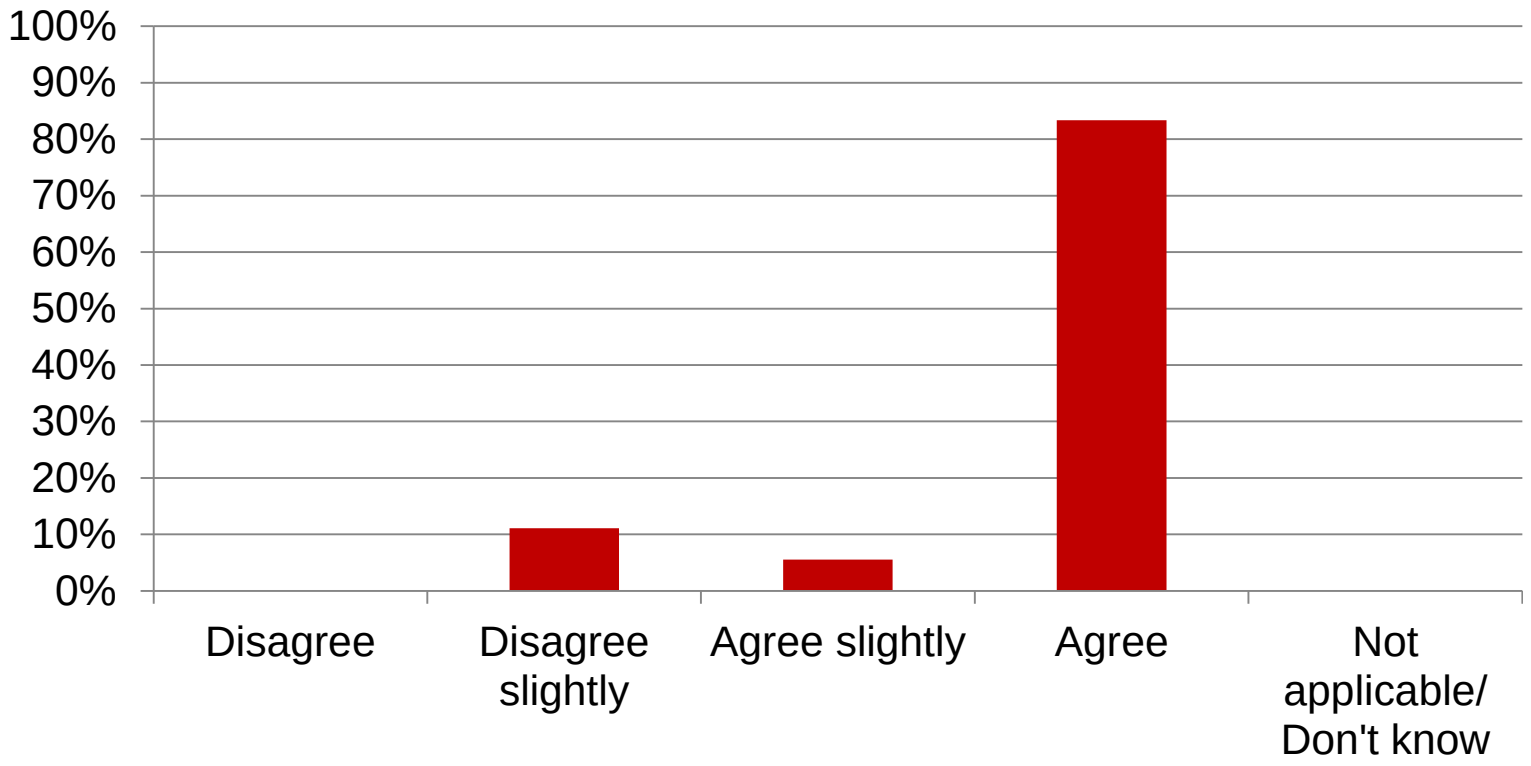


Q:The captions and transcript helped me understand what the professor was saying

Q:The videos with caption (& transcript) are preferable than videos without them

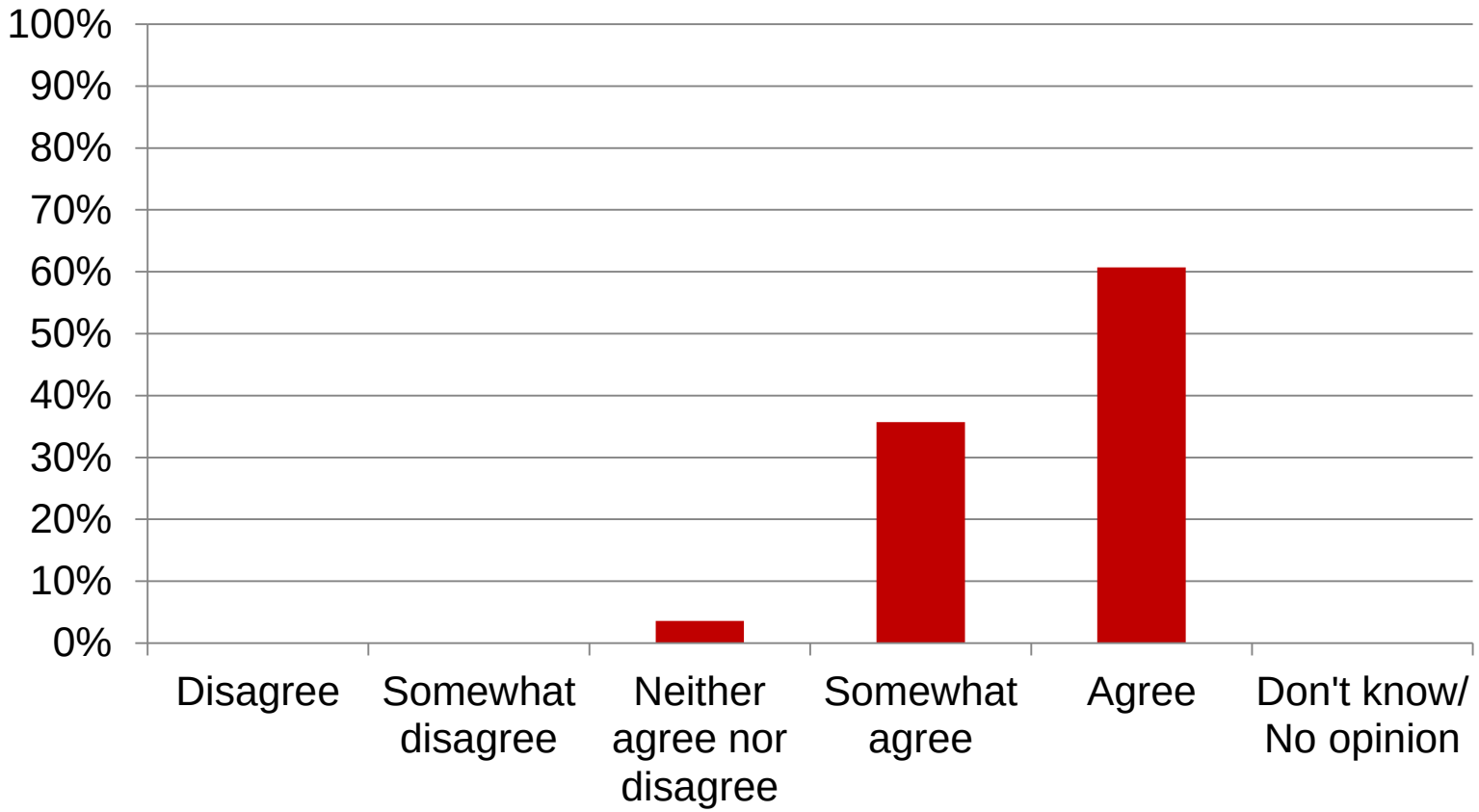
Most helpful for: Learning, Attention, Efficiency, Note-taking
Somewhat helpful for: Motivation, Quiz Performance

Caption Editor: Evaluation



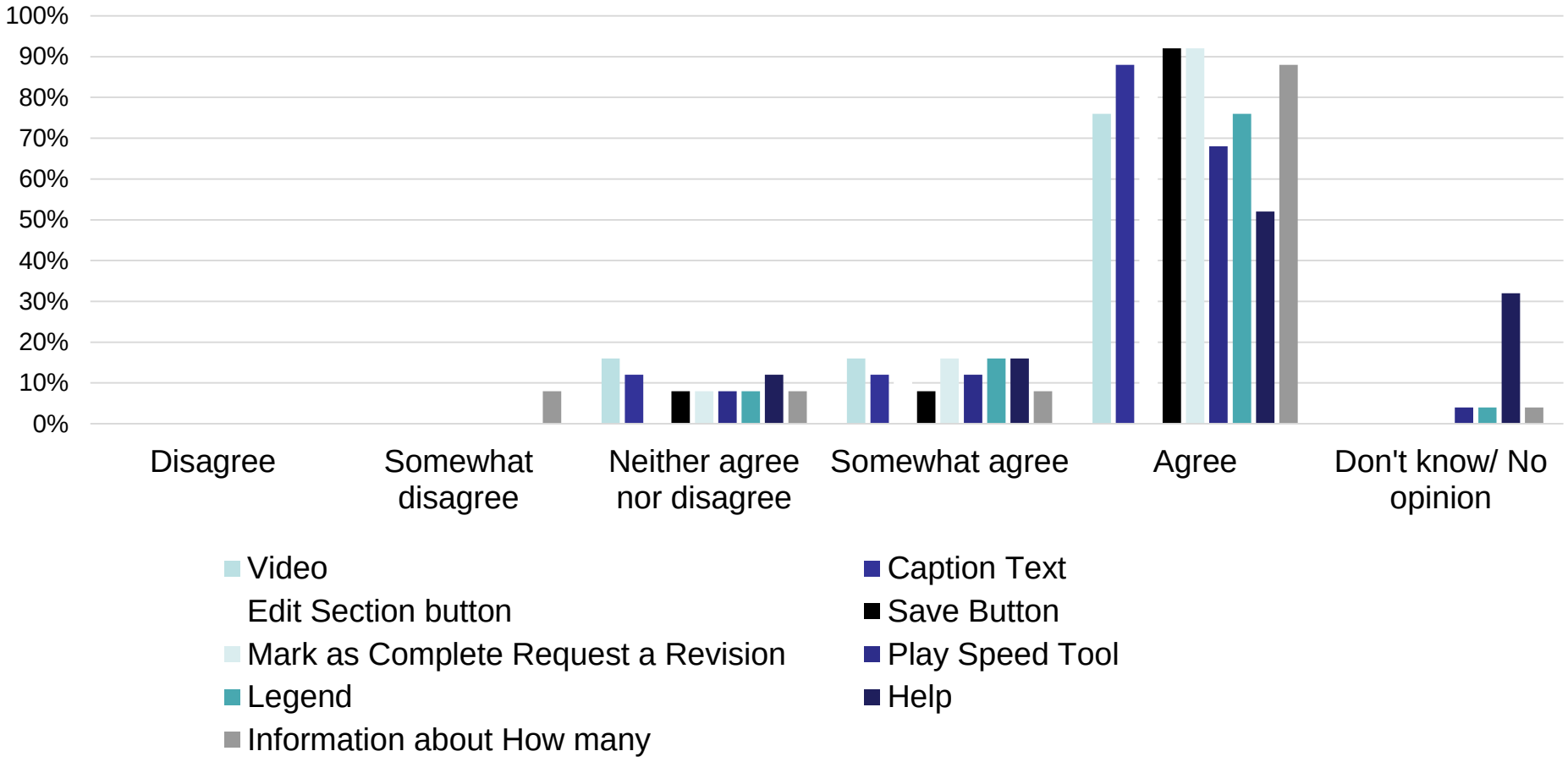
Question: The captions and the transcript represented accurately what the professor said

Caption Editor: Evaluation



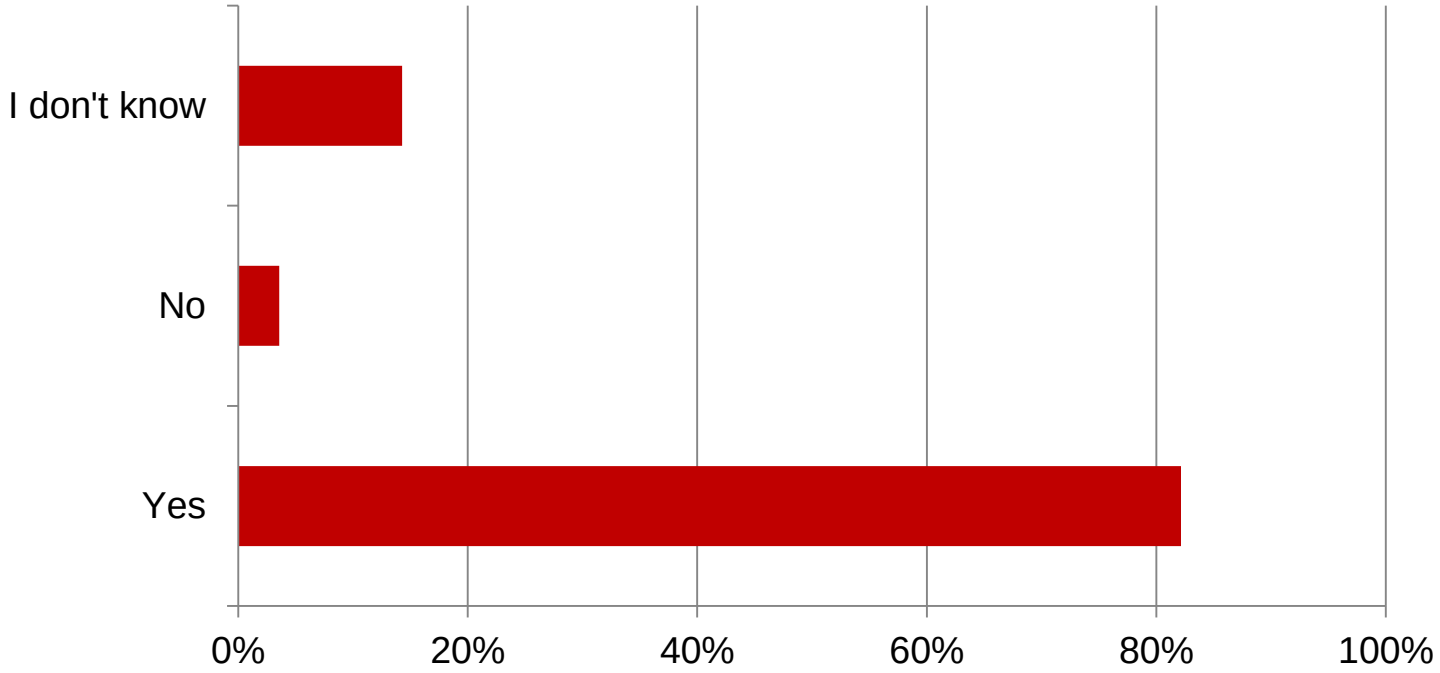
Question: The ICS Caption Editor is easy to use.

Caption Editor: Evaluation



Question: The placement (position) of the following elements and controls on the Caption Editor interface was appropriate?

Caption Editor: Evaluation



Would you be interested in working with other students to correct captions for your class lectures using this caption editor if you receive some incentive (for example academic credit)?

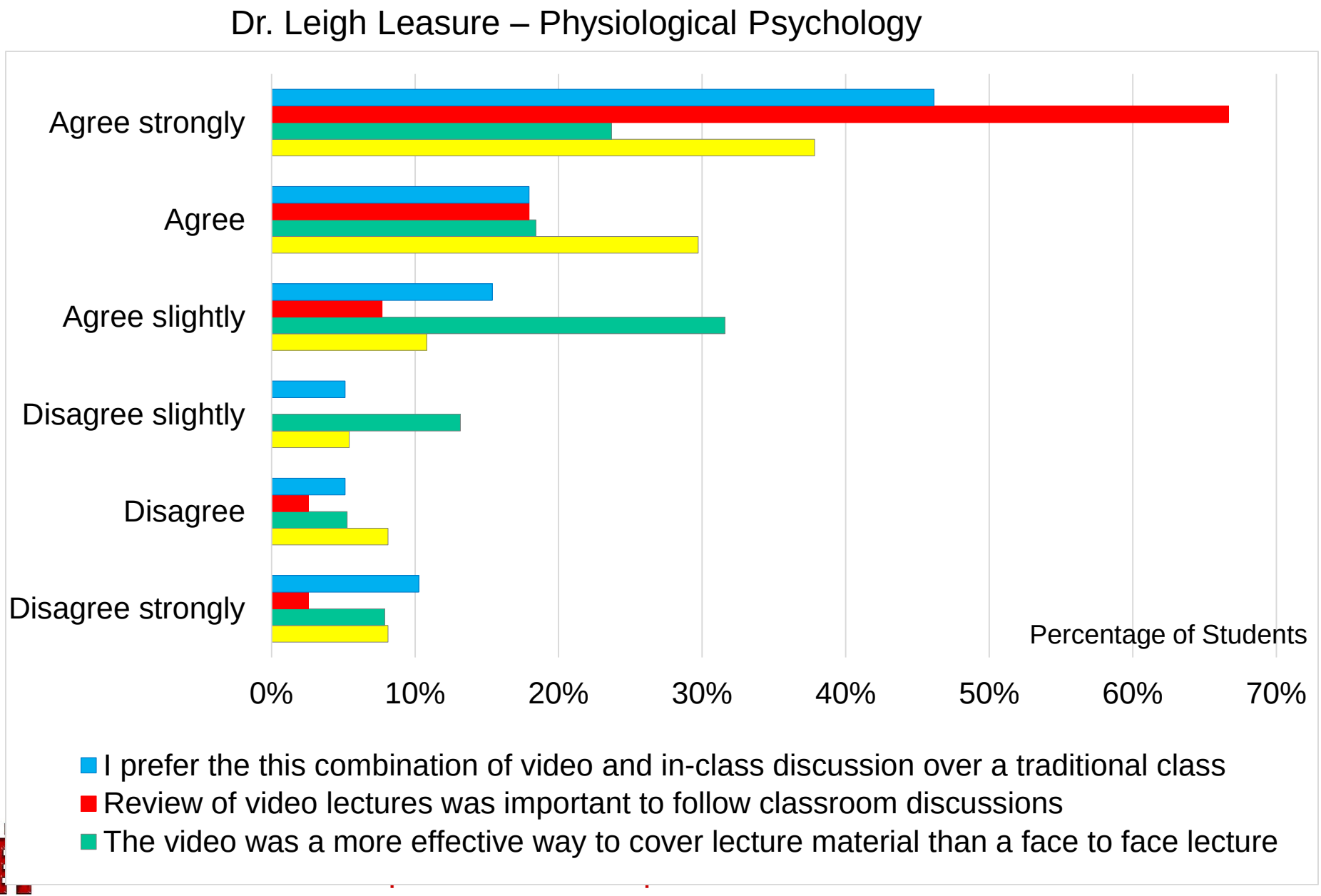
ICS Videos: Usage and Experience

Employed in over 100 class sections at UH over the past 5 with many1000s of student users total!

Freely available

Recently being employed as the technology to enable flipped classrooms (UH TIP Award 2014)

Inverted Classroom



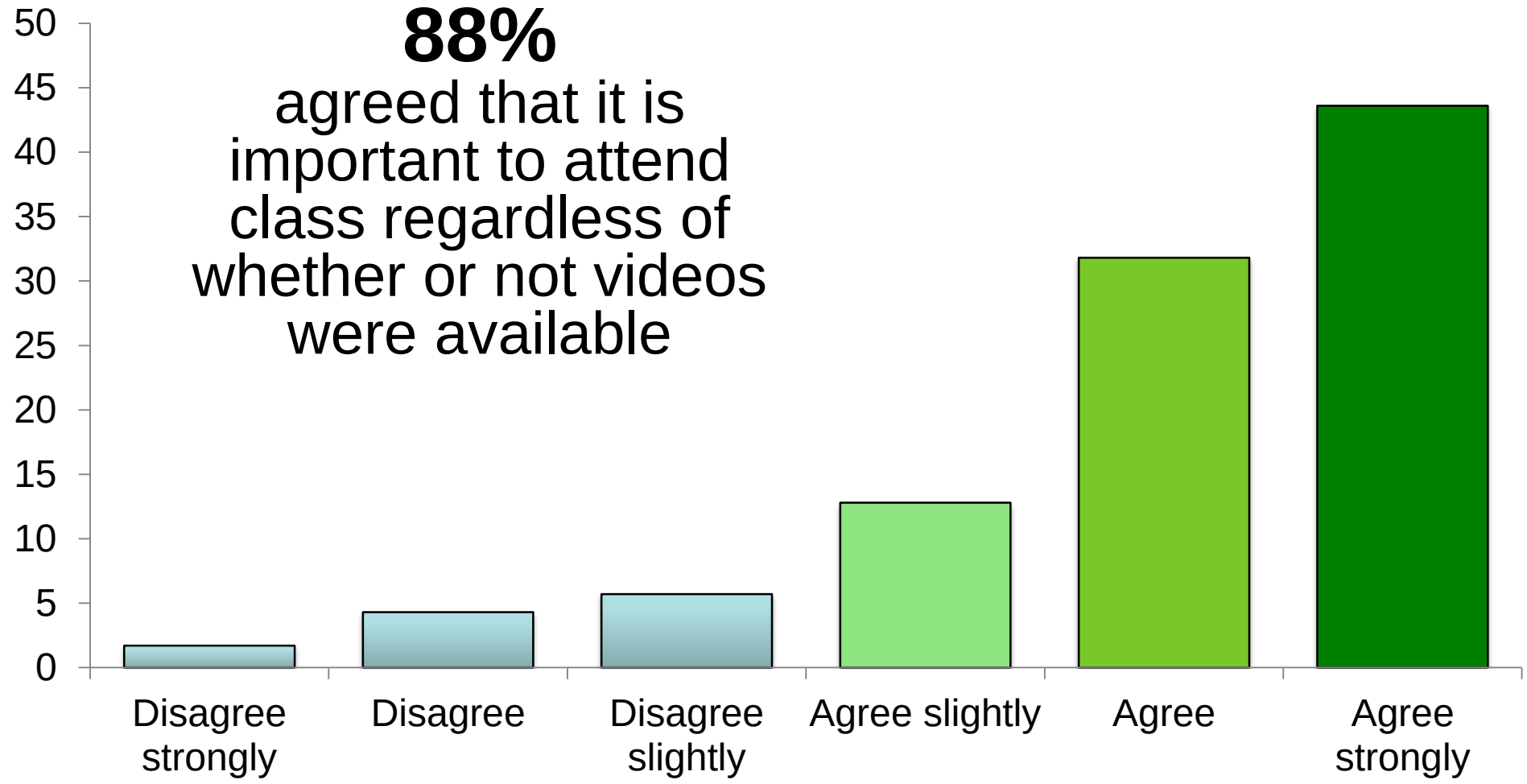
Videos & Attendance

Students who watched at least one video had higher attendance rates than those who did not watch any videos.

Video use and **attendance** were slightly positively correlated

Videos & Attendance

88%
agreed that it is important to attend class regardless of whether or not videos were available



Survey Comments

Comments in surveys and focus groups were positive beyond expectations (& very satisfying)

"*I can tell that I learned 80% of the class by videos*"

"This is a great tool for us as a student to go over everything. I would hope that this would be available in all classes"

"Indexing enabled me to jump directly to my trouble spots".

"I'm used to listening and then looking at the caption because it helped me tremendously to learn the new words and vocabulary"

"The search feature function have instant results, like a Google search box"

Conclusions (sort of)

Classrom Videos are an important learning companion (akin to a textbook)

- Automatic methods are effective at making video content more accesible
- CS research challenges remain
  - Smarter/Semantic Indexing and Search
  - Merging of audio, text and image information
  - HCI for a better learning experience
- Ongoing efforts to make the technology widely available

Co-conspirators

CS Faculty: Zhigang Deng, Olin Johnson, Shishir Shah, Rakesh Verma, Christoph Eick

Students: Tayfun Tuna, Varun Varghese, Mahima Joshi, Tuhin Dey (Education)

X-Students: J. Li, C. Yun, G. Bhatt, T. Tuna. A. Verma, R. Kushalnagar, Rucha Borgaonkar,

NSMIT Staff: S. Baez-Franseschi, Pradeep Krishnan, Andrea Arias

Assessment: Lecia Barker (UT Austin), Yumei Liu, Chris Hovey (NorthEastern)

Deployment, usage and assessment

UH Computer Science, UH Geosciences, *UH Biology and Biochemistry*
UH Downtown ***(Richard Alo),*** *Texas School for the Deaf* ***(David Coco)***

Last Slide

- ICS Videos System

  www.icsvideos.uh.edu

- Username: student
- Password : icsstudent

- Contact email:

  icsvideoscontact@gmail.com

  or jaspal@uh.edu