

# Big Data Analytics for Teaching & Learning

**Jihie Kim, PhD**

Senior Vice President / Director

Future Technologies Lab

Advanced Institute of Technology | 2013. 11



# kt Advanced Institute of Technology (AIT) at-a-glance

## R&D Goals

- Developing future core technologies and services
- Securing capability to lead the mid/long-term growth for the kt group

<b>Established</b>	Jan. 1984 (as an organization for business support head office)
<b>No. of R&amp;D staffs</b>	454 (as of Aug. 2013, doctoral degree 19%, masters degree 56%)
<b>No. of Patents</b>	5,892 (as of Sept. 2013, 782 for global patents)



**kt R&D center**  
(Gangnam, Seoul)

**kt R&D complex**  
(Daejeon)



# The Evolution of Telco Services

- Fast and reliable communication services: from Bells to mobile broadband, to IT



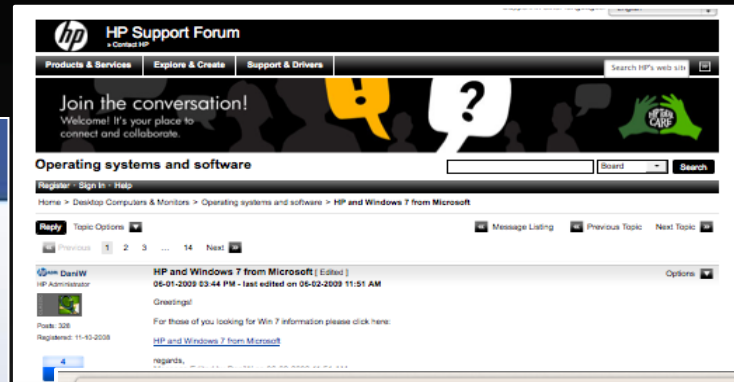
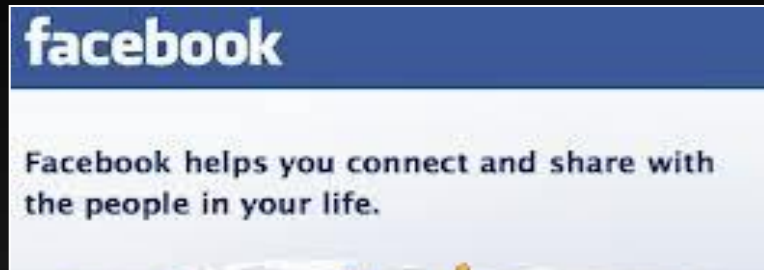
# kt AIT: Research Divisions

<b>Infra Laboratory</b>	<ul style="list-style-type: none"><li>• Network Innovation (Wireline/Wireless)</li><li>• SDN (Software Defined Network)</li><li>• Next Generation Cloud Computing &amp; Data Center</li></ul>
<b>Service Laboratory</b>	<ul style="list-style-type: none"><li>• Media Technology</li><li>• Virtual Goods</li><li>• Software Engineering</li><li>• Smart Home Service</li></ul>
<b>Convergence Laboratory</b>	<ul style="list-style-type: none"><li>• Security &amp; Surveillance</li><li>• Next Generation Monitoring &amp; Control, Smart Stadium</li><li>• Smart Grid &amp; Energy</li><li>• Healthcare</li></ul>
<b>Future Technology Laboratory</b>	<ul style="list-style-type: none"><li>• Enabling Technologies</li><li>• Intelligence and Big Data</li><li>• Next Generation UI/UX</li><li>• <b>Smart Education</b></li></ul>

# Outline

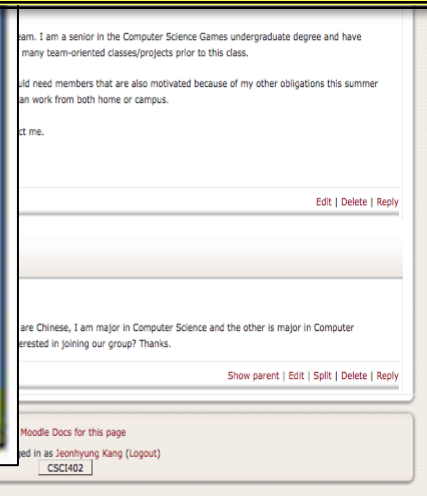
- 
- 1 Online Activities**
  - 2 Data Boom in Education**
  - 3 Example: Student Performance vs. Online Activity**
  - 4 Big Data Platform for Education**
  - 5 Summary and Conclusion**
-

# Proliferation of Online Activities



- *Business*
- *Politics*
- *Education*
- *Entertainment*
- *Household*

## Game Changing





# Face-to-Face Learning → Online/Blended



E-books



Educational Mobile Apps



Online Social Learning



Online Courses

# Intelligent Tutoring Systems

**ANDES Physics Workbench - [Exs2a-Solution.FBD]**

File Edit Diagram Variable View Help

A model airplane hangs from two strings S1 and S2 which are attached to the ceiling. String S1 is inclined at 45 degrees, and string S2 is inclined at 60 degrees, as shown in the figure below.

If the tension in string S1 is 50 N

a) find the mass of the airplane

b) find the tension in string S2

Variables

Name	Definition	X-Comp	Y-Comp
mp	mass of airplane		
F12	magnitude of the Tension For...	F12_x	F12_y
Fw	magnitude of the Weight Forc...	Fw_x	Fw_y
F11	magnitude of the Tension For...	F11_x	F11_y
a	magnitude of the average Acc...	a_x	a_y

- $Fw = mp * g$
- $g = 9.8$
- $F11_x = -F11 * \cos(45)$
- $F11_y = F11 * \sin(45)$
- $F11_y + F12_y + Fw_y = mp * a_y$
- $F12_2 = F12 * \sin(0)$
- 
- 
- 
- 
- 
- 

For Help, press F1

NUM 00:42:18



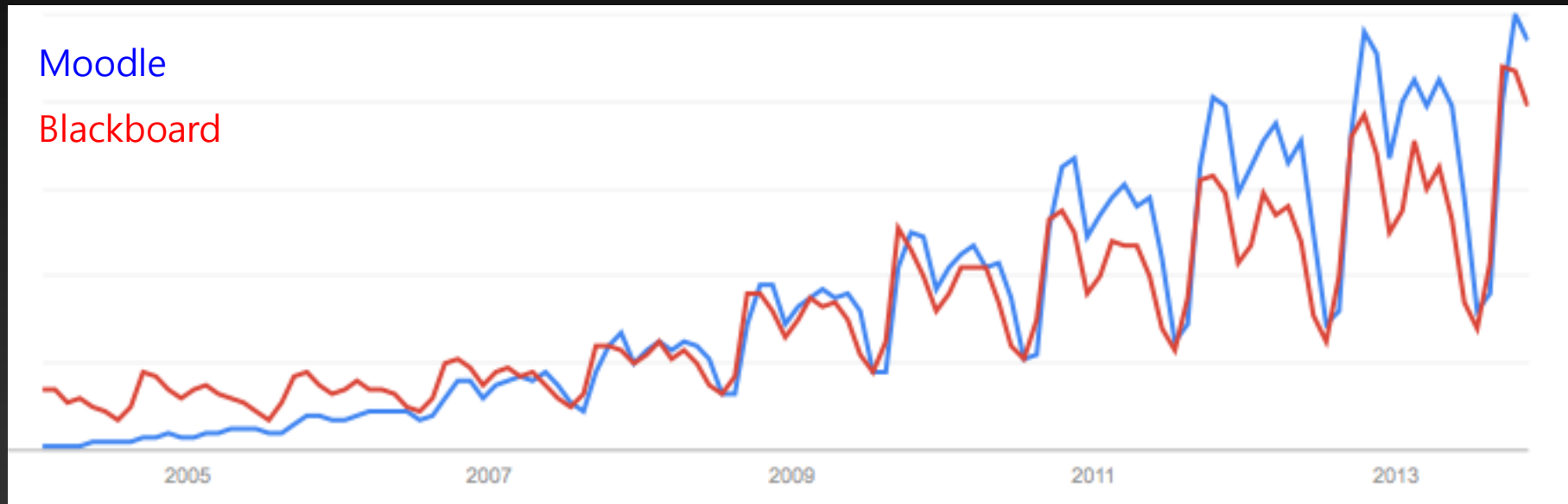
- Assess student skills using domain models
- Provide intelligent feedback



# LMS (Learning Management Systems)

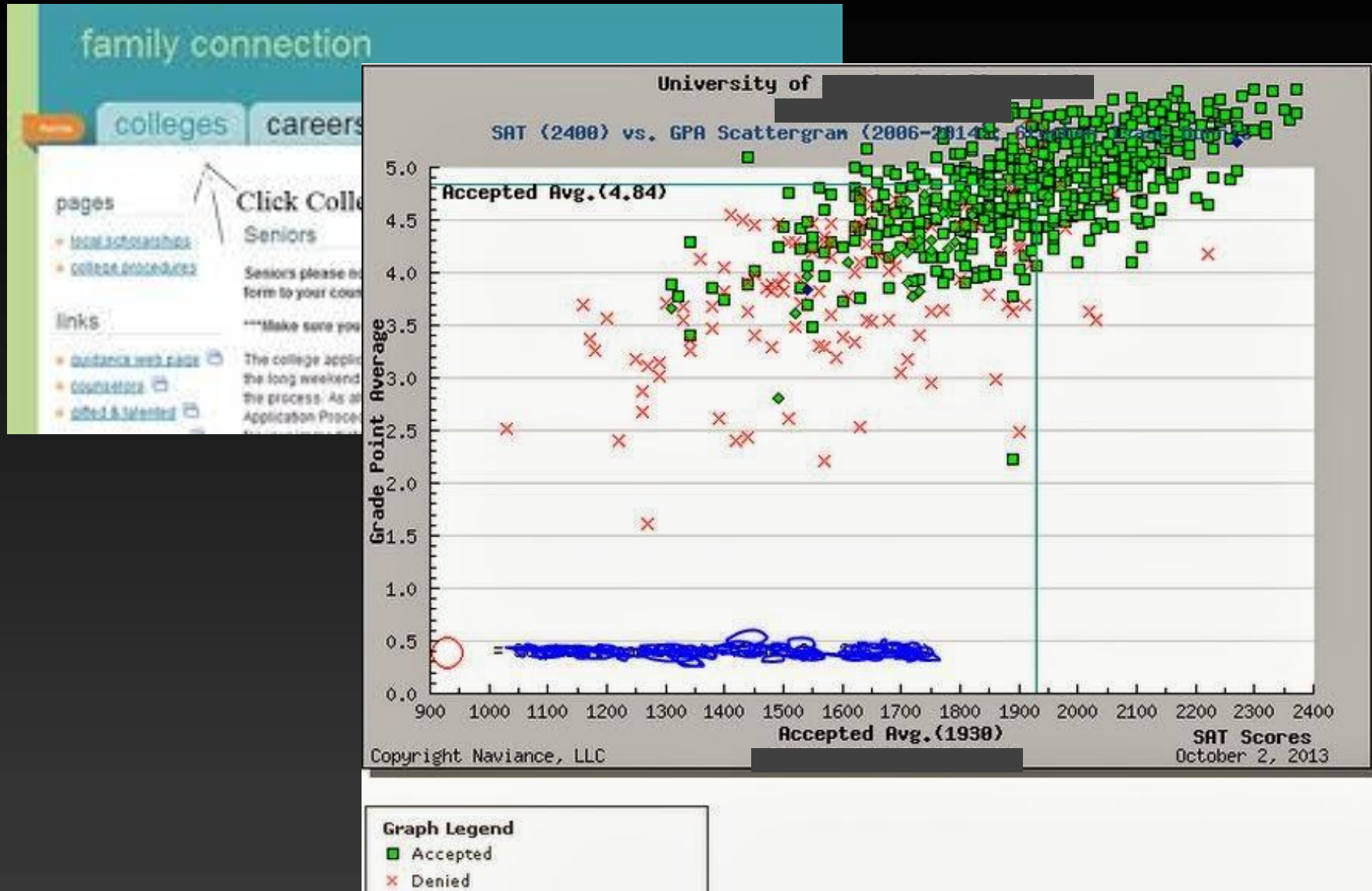


...

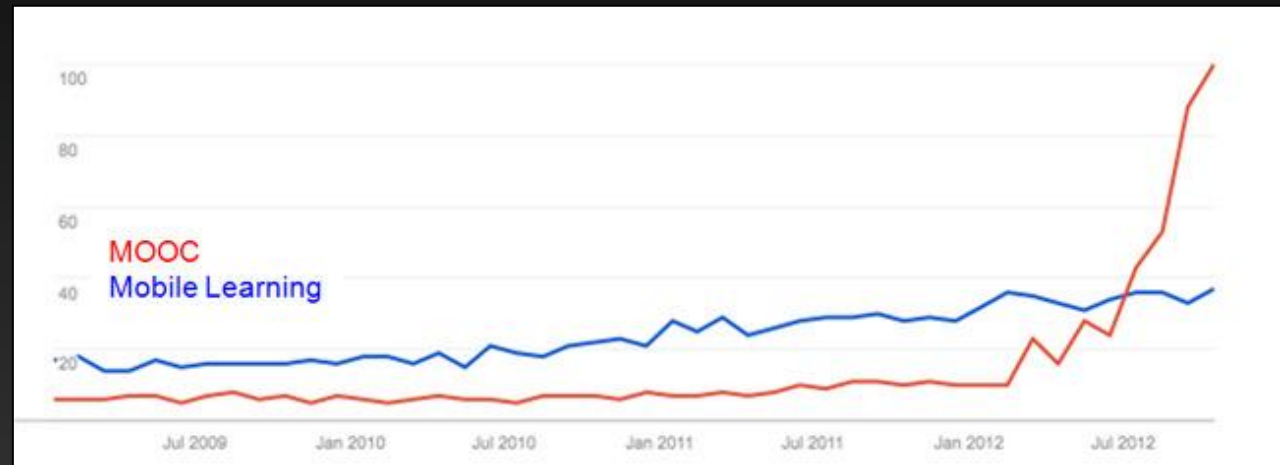


Google Trends

# Naviance in USA

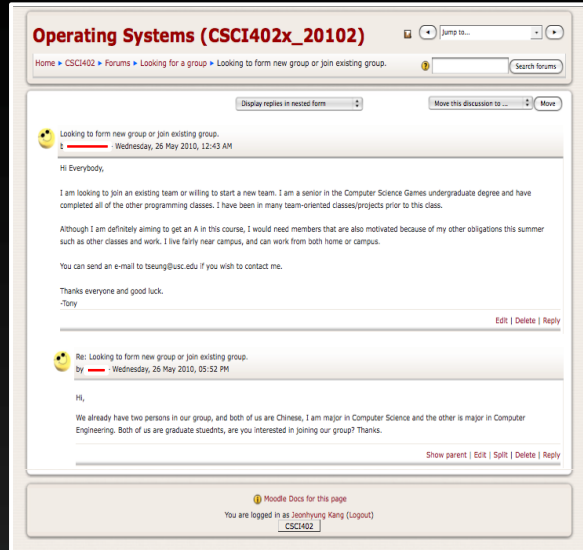
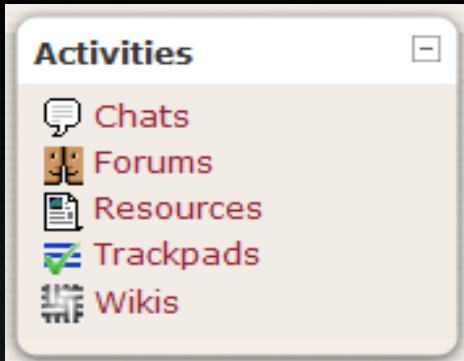


# Massive Open Online Courses (MOOC)

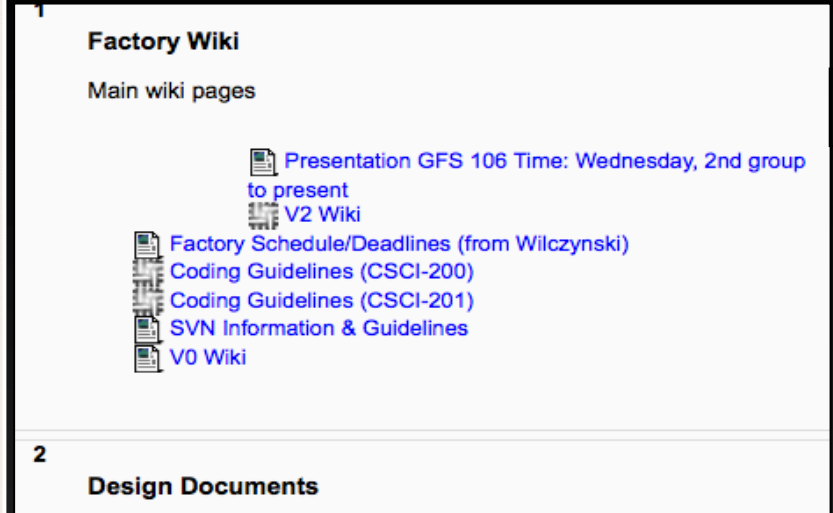


Google Trends

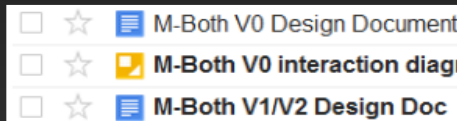
# Data Poor → Data Rich



## Discussion Board



## Wiki



## Google Docs

M-Both V0 Design Document

- Smart popup- it keeps track of what it gave the robots, and then the robot is free or that it is giving the part back.
- Special case conveyors: will know if they are front back, or neither.
- When a conveyor is trying to send a part onwards, it asks to see if the next conveyor is free, then immediately stops. It only restarts once the next thing has given the part.

## GlassReceiver interface

All agents that receive glass will implement the GlassReceiver interface



## Smart Education

# Big Data in Education: Potentials (NSF 2013)

- Online courses that get better the more students use them
- A revolution in education research -- "Internet-scale experimentation" with > 100,000 students
- Personalized and adaptive instruction -- Identify students' interests, prior knowledge, and misconceptions
- Effective innovations -- increased access and democratization
- Ability to predict which students are at greatest risk



## Publications & Products

## Surveys & Programs

## Data & Tools

## Fast Facts

## School Search

- [Achievement Effects of Four Early Elementary School Math Curricula: Findings from First Graders in 39 Schools](#)
- [Baccalaureate & Beyond \(B&B:93/94/97/03 & B&B:2000/01\)](#)
- [Beginning Postsecondary Students \(BPS:90/92/94, BPS:96/98/01 & BPS 04/06\)](#)
- [Common Core of Data \(CCD\)](#)
- [Early Childhood Longitudinal Studies](#)
- [Education Longitudinal Study of 2002 \(ELS:2002\)](#)
- [The Effectiveness of Mandatory Random Student Drug Testing](#)
- [Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort](#)
- [Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts](#)
- [Effectiveness of Selected Supplemental Reading Comprehension Interventions: Findings from Two Student Cohorts](#)
- [Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students](#)
- [The Evaluation of Enhanced Academic Instruction in After-School Programs: Final Report](#)
- [An Evaluation of Teachers Trained Through Different Routes to Certification](#)
- [High School & Beyond \(HS&B:80/82/84/86/92\)](#)
- [High School Transcript Studies](#)
- [Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study](#)

## Learners

- [Middle School Mathematics Professional Development Impact Study: findings After the First Year of Implementation](#)
- [National Assessment of Adult Literacy \(NAAL\)](#)
- [National Assessment of Educational Progress \(NAEP\)](#)
- [National Education Longitudinal Survey of 1988 \(NELS:88/92/94/00\)](#)
- [National Household Education Survey \(NHES\)](#)
- [National Longitudinal Transition Study-2 \(NLTS2\)](#)
- [National Postsecondary Student Aid Study \(NPSAS:90/93/96/00 /04/08\)](#)
- [National Study of Postsecondary Faculty \(NSOPF:88/93/99/04\)](#)
- [NCES-Barron's Admissions Competitiveness Index Data Files](#)
- [Postsecondary Education Quick Information System \(PEQIS\)](#)
- [Pre-Elementary Education Longitudinal Study \(PEELS\)](#)
- [Preschool Curriculum Evaluation Research \(PCER\)](#)
- [Program for International Student Assessment \(PISA\)](#)
- [Public Libraries Survey](#)
- [Reading First Impact Study \(RFIS\)](#)
- [Reading First Impact Study: Interim Report](#)
- [School Library Media Centers](#)
- [School Survey on Crime and Safety \(SSOCS\)](#)
- [Schools and Staffing Survey \(SASS\)](#)
- [Social and Character Development \(SACD\)](#)





## Learning Analytics and Knowledge

[Join/Renew](#)

[Home](#)   [Keynotes & Program](#)   [Video Streaming](#)   [Key Dates](#)   [Committees](#)   [Sponsors](#)   [Registration](#)

# Outline

- 
- 1 Online Activities
  - 2 Data Boom in Education
  - 3 **Example: Student Performance vs. Online Activity**
  - 4 Big Data Platform for Education
  - 5 Summary and Conclusion
-

# Can online participation data predict performance?

A predictive relation between discussion contributions and student performance in undergrad engineering courses [Yoo and Kim 2013]

- What is the relationship between the degree of participation and the project performance?
- What is the relationship between the kinds of words used by the student or coherency of sentences and the project performance?
- What is the relationship between the expressions of emotion and the project performance?
- What is the relationship between the style of participation in the discussion, as information seeking vs. giving, and the project performance?
- What is the relationship between work pacing and work performance?

# Online Discussion Boards

The screenshot shows a web browser window displaying the DEN/CARTE BBS website. The page is titled "Operating Systems (CSCI402x\_20102)" and is part of the University of Southern California Distance Education Network (DEN). The sidebar on the left contains a welcome message, login information, and a list of recent topics. The main content area shows a list of discussion topics, including "Lectures", "ISI Discussion Board Questions", "Looking for group partners", "Grader/TA Office Hour Notices", "Nachos General Questions", "402 Humor", "Project 1 Questions", "Project 2 Questions", "Project 3 Questions", and "Project 4 Questions". A detailed view of a post is shown on the right, featuring a user profile, a post title, and the text of the post. The post is titled "Looking for group partners" and is by a user named Tony. The post content reads: "I am looking to join an existing team or willing to start a new team. I am a senior in the Computer Science Games undergraduate degree and have completed all of the other programming classes. I have been in many team-oriented classes/projects prior to this class. Although I am definitely aiming to get an A in this course, I would need members that are also motivated because of my other obligations this summer such as other classes and work. I live fairly near campus, and can work from both home or campus. You can send an e-mail to tseung@usc.edu if you wish to contact me. Thanks everyone and good luck. -Tony". The post has a timestamp of Wednesday, 26 May 2010, 12:43 AM. Below the post, there is a section for "Project 1 Questions" and "Project 2 Questions". The footer of the page includes a Moodle Docs link and a login status for Jeonhyung Kang.

University of Southern California  
Distance Education Network (DEN)

Welcome lord cartedentem

Last login: 18 Sep 2005 13:29  
Current time: 18 Sep 2005 13:42  
Your total topics: 2  
Your total posts: 12  
Your courses:  
CSCI402x\_20053, CSCI555\_20053,  
CSCI577, REV101,  
Archived courses:  
CSCI402\_20043, CSCI402x\_20051,  
CSCI402x\_20052.  
XML Rss feed of this discussion board

New Topics

Reading list #54 [09/18/05 13:31]  
DNS Granularity [09/18/05 13:19]  
At laptop table in f... [09/18/05 12:44]  
who is the other gro... [09/18/05 10:44]  
Assertion failed Err... [09/18/05 10:21]

Your Recent Topics

[CSCI402] Discu... [08/04/05 10:04]  
tool [06/14/05 10:01]

Search

Discussions

Today's Dilibert

RSS Feed: <http://www.osnews.com/>

Firefox vs. IE security: Is Time Gre...  
Cairo, Libresetup, LibXML2 Ported...  
An Overview of File Paging Applicat...  
\*Palm Opens Retail Store: New Tungs...  
Random Linux Blue O...

Courses

Discussion Search

CSCI402x\_20053

Lectures  
Post questions and comments about the course lectures in t...

ISI Discussion Board Questions  
Post questions and comments about the ISI Discussion Board...

Looking for group partners

Grader/TA Office Hour Notices

Nachos General Questions

402 Humor

Project 1 Questions

Project 2 Questions

Project 3 Questions

Project 4 Questions

CSCI555\_20053

Operating Systems (CSCI402x\_20102)

Home > CSCI402 > Forums > Looking for a group > Looking to form new group or join existing group.

Jump to...

Search forums

Display replies in nested form

Move this discussion to ... Move

Looking to form new group or join existing group.  
Looking to form new group or join existing group.  
Wednesday, 26 May 2010, 12:43 AM

Hi Everybody,

I am looking to join an existing team or willing to start a new team. I am a senior in the Computer Science Games undergraduate degree and have completed all of the other programming classes. I have been in many team-oriented classes/projects prior to this class.

Although I am definitely aiming to get an A in this course, I would need members that are also motivated because of my other obligations this summer such as other classes and work. I live fairly near campus, and can work from both home or campus.

You can send an e-mail to tseung@usc.edu if you wish to contact me.

Thanks everyone and good luck.  
-Tony

Edit | Delete | Reply

Looking to form new group or join existing group.  
Looking to form new group or join existing group.  
Wednesday, 26 May 2010, 05:52 PM

Hi,

We already have two persons in our group, and both of us are Chinese, I am major in Computer Science and the other is major in Computer Engineering. Both of us are graduate students, are you interested in joining our group? Thanks.

Show parent | Edit | Split | Delete | Reply

Moodle Docs for this page

You are logged in as Jeonhyung Kang (Logout)

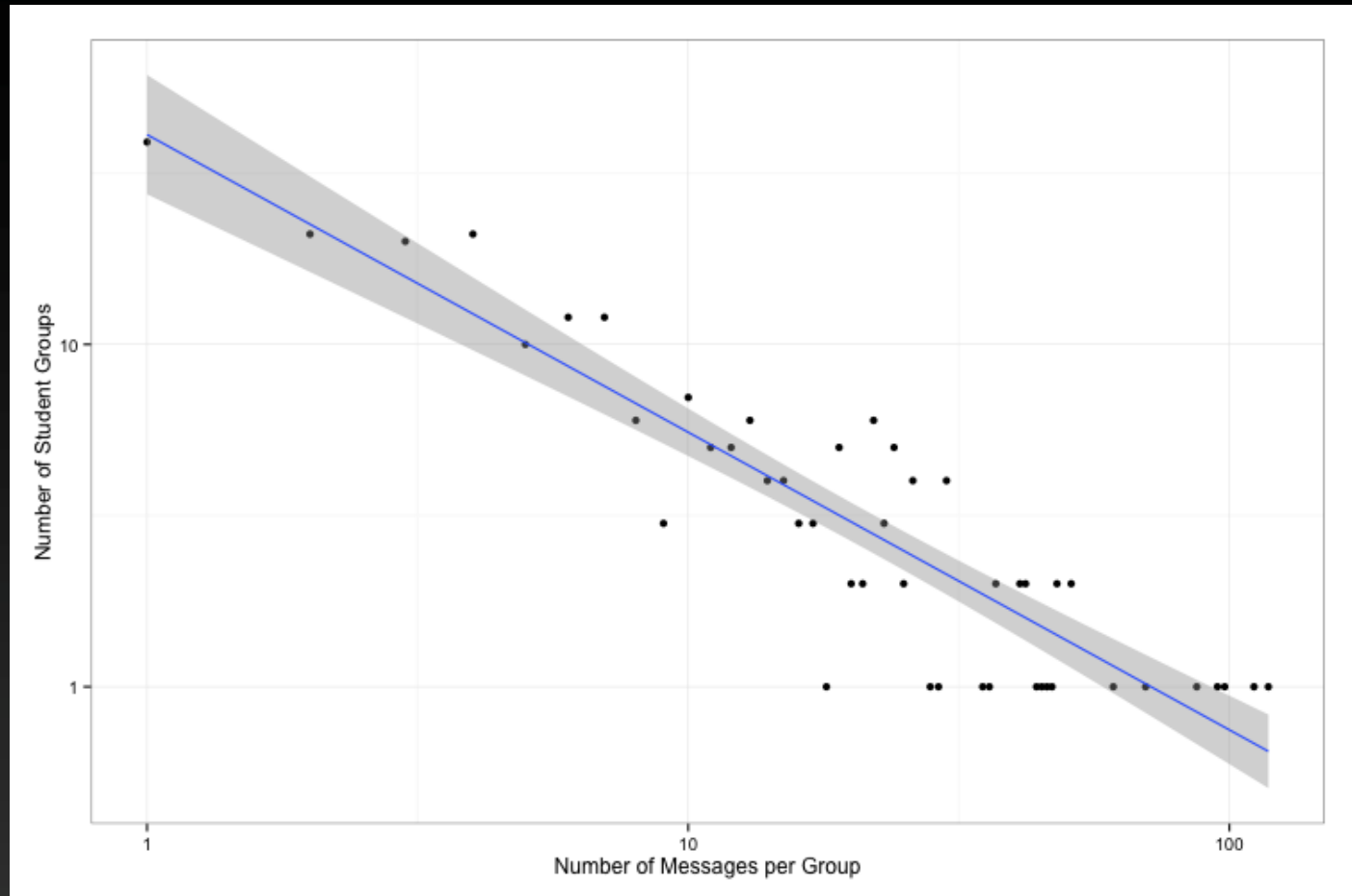
CSCI402

# Can online participation data predict performance?

## Predictive Variables

- What is the relationship between the degree of participation and project performance?
  - Number of participation/messages
- What is the relationship between the kinds of words used by the student or coherency of sentences and the project performance?
  - Linguistic measures
- What is the relationship between the expressions of emotion and the project performance?
  - Emotion words used
- What is the relationship between the style of participation in discussions, such as information seeking vs. giving, and the project performance?
  - Contribution type or roles played
- What is the relationship between work pacing and work performance?
  - Participation time and procrastination

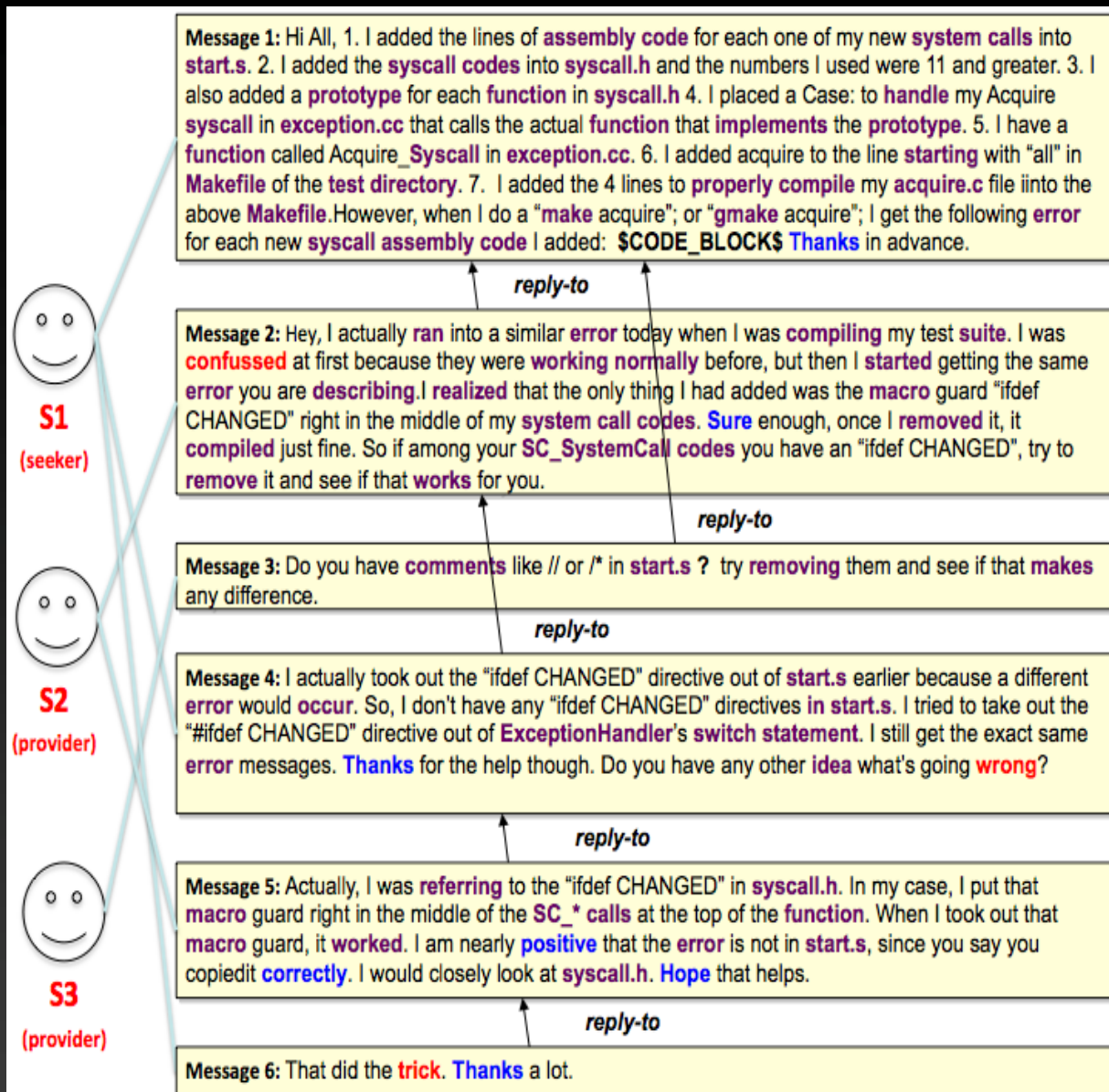
# Degree of Participation



Distribution of Number of Messages per Student Group (log-log scale)



# Message Content: Kind of Words Used



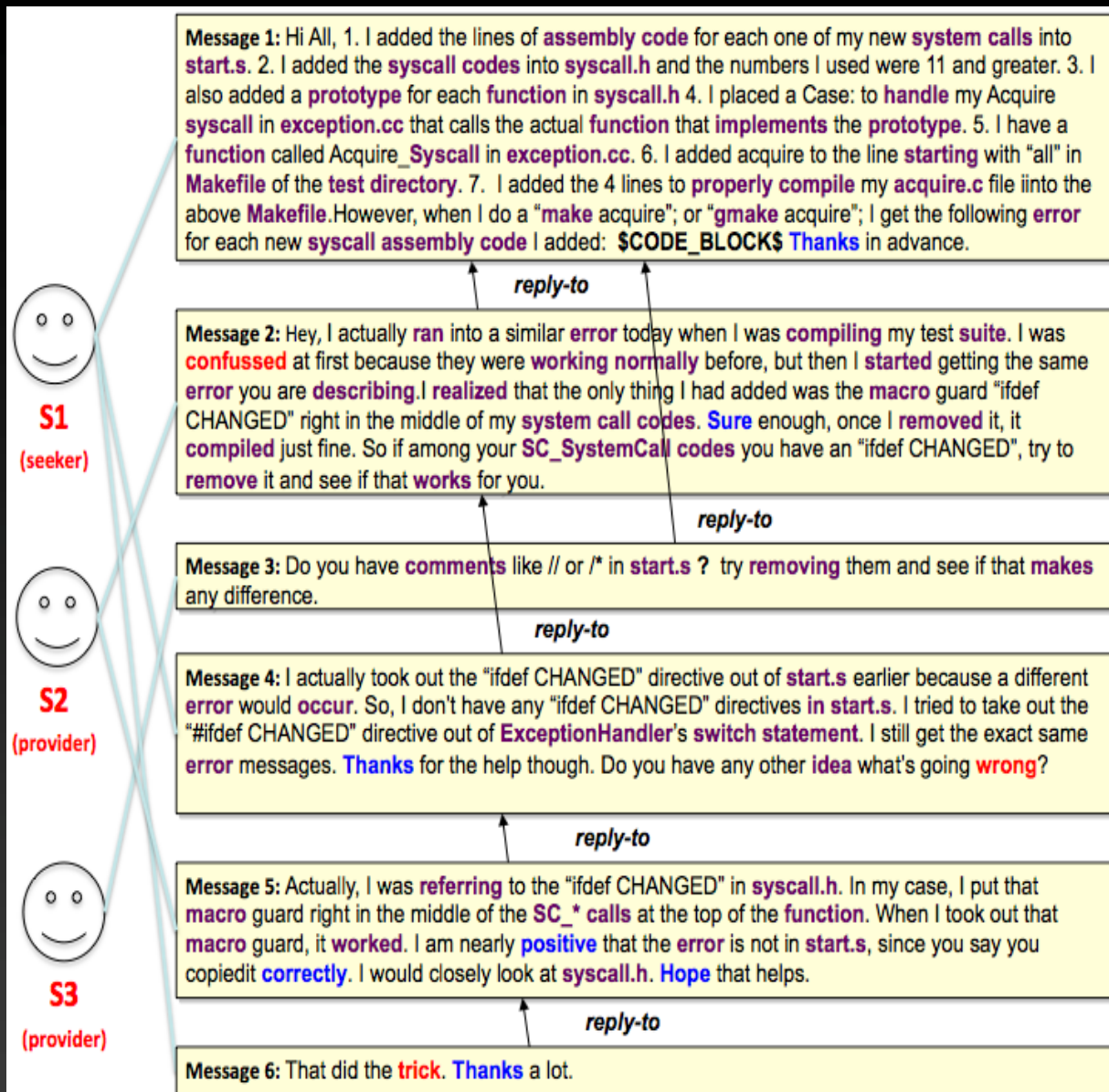
Blue : positive  
Red: negative  
Violet: technical

# Message Content: Kind of Words Used

Positive emotion		Negative emotion		Certainty		Achievement	
Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
create	794	problem	1089	all	1510	able	2316
like	754	low	781	sure	558	king	1252
ease	727	numb	677	correct	494	create	794
thank	649	fault	581	must	332	first	479
sure	558	nag	437	never	249	acquire	405
please	494	interrupt	383	every	223	try	347
value	462	wrong	267	exact	180	work	344
fine	297	fail	262	always	165	fail	262
good	281	destroy	165	true	157	gain	256
well	265	argue	146	fact	150	require	248
ok	172	evil	109	certain	99	solution	132
free	166	sorry	82	clear	98	better	126
ready	159	worry	69	total	93	creative	117

Frequent Emotional and Psychological Words

# Roles Played: Information Seeker vs. Provider



Question

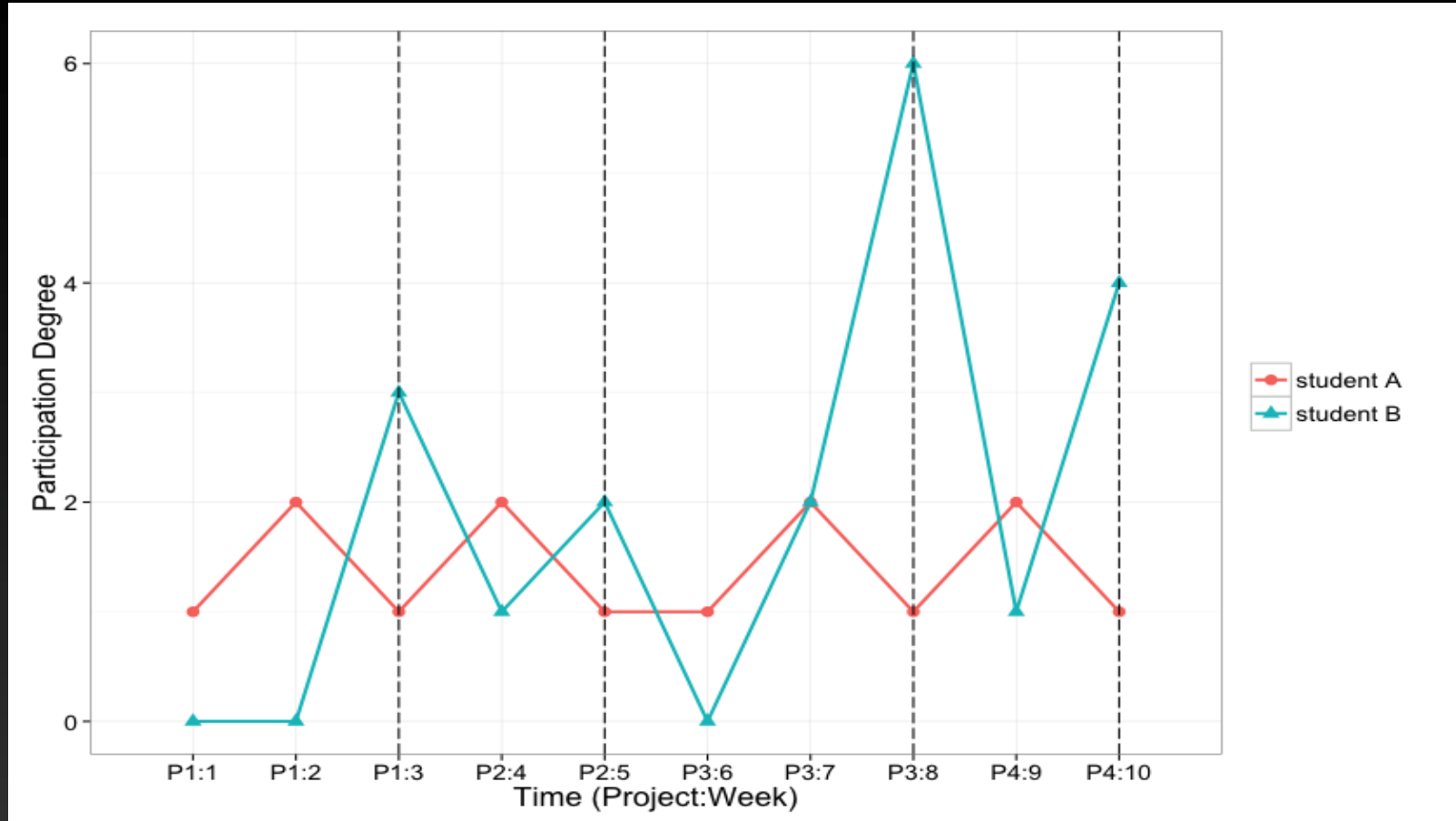
Answer

Answer

Question

Answer

# Participation Time



Degree of Participation over Time

# Procrastination

Posting Time To Deadline (PTTD):

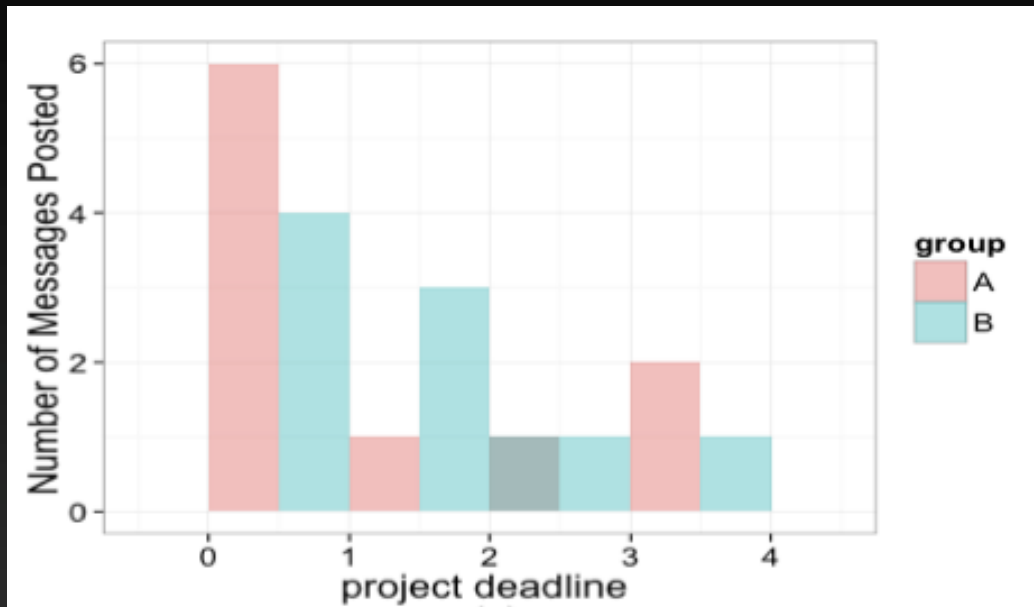
normalized distance between message posting time and project deadline

$$PTTD_S^{(i)} = \frac{end_p - msg_s^{(i)}}{end_p - start_p}$$

Smaller PTTD can indicate procrastination

# Procrastination

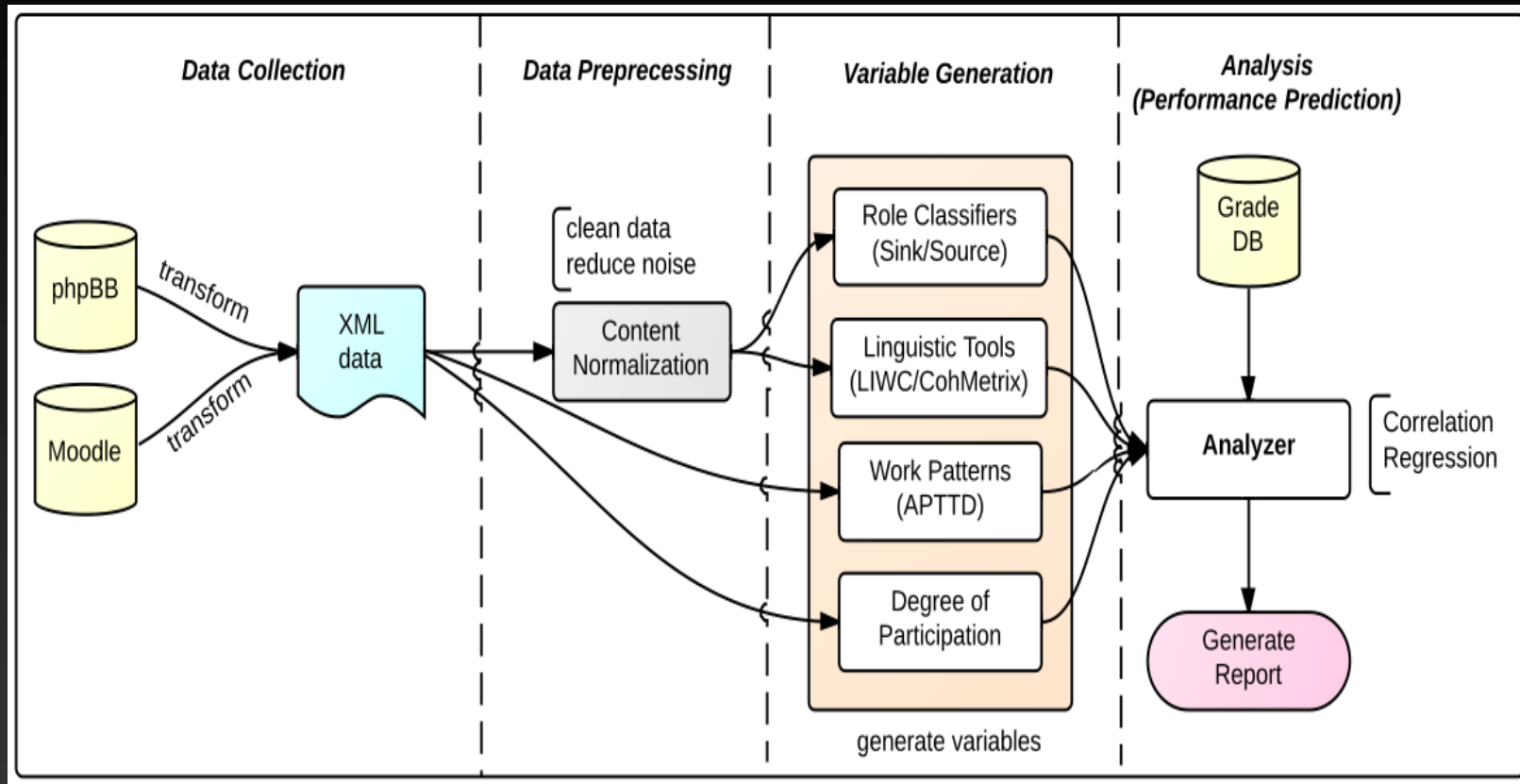
Posting Time To Deadline (PTTD): normalized distance between message posting time and deadline



Group A (high performer) posted messages earlier than Group B (low performer)



# Pipeline for Data Analysis



# Data Pre-Processing

View previous topic :: View next topic

Message	Author
<p>Posted: 13 Oct 2004 17:11</p> <p><a href="#">nlp</a> <a href="#">Reply</a> <a href="#">edit</a> <a href="#">x</a></p> <pre>/auto/home-scf-22/csci402/testgnu/decstation-ultrix/bin/ld -G 0 -c -I../userprog -I../threads start.o sleepingbarber.o -o sleepingbarber.coff /auto/home-scf-22/csci402/testgnu/decstation-ultrix/bin/ld: cannot open linker script file -I../userprog: No such file or directory gmake: *** [sleepingbarber] Error 1  i cant get my sleeping barbers test compiled???? what shud i do</pre>	<p><b>Anonymous</b></p> <p>Joined: 28 Jan 2004 Posts: 135</p> <p><a href="#">email</a></p>

Handling noisy data

# Variable Generation

Categories	Metrics	Variables	Tool
Participation quantity	Quantitative	The number of Words, Sentences, Paragraphs, Messages (Total, Initials, and Replies)	Programming
Work pattern	Quantitative	APTTD (Average Posting Time To Deadline)	Programming
Technical	Quantitative	Technical terms	Dictionary
Information Roles	Qualitative	Question (Sink), Answer (Source)	Speech Act Classifiers
Linguistic	Quantitative	Past, Present, Future tense and Negations, Swear words	LIWC
	Qualitative	Flesch Reading Ease Score, Type-token ratio, Concreteness, Hypernym, Log Frequency	Coh-Metrix
Emotional & Psychological	Qualitative	Positive emotions, Negative emotions, Insight, Causation, Discrepancy, Certainty, Tentative, Inhibition, See, Time, Achievement, Assent	LIWC
Semantic factors	Qualitative	LSA sentence adjacent, LSA sentence all	Coh-Metrix
Situation Model	Qualitative	Casual, Temporal, Spatial cohesion	Coh-Metrix

Result:  
Correlation  
Analysis

N = 173;  
\*p < .05; \*\*p < .01

Categories	Metrics	Variables		Tool	Correlation (Grade)
Simple Measures	Quantitative	Number of Words		Programming	-.09
		Number of Sentences			-.10
		Number of Paragraph			-.08
		Number of Messages	Total		.17*
			Initial		.13
			Replies		.17*
Time	Quantitative	APTTD		Programming	.21**
Technical	Quantitative	Technical terms		Dictionary	.08
Speech Act	Qualitative	Question (Sink)		Speech Act	.03
		Answer (Source)		Classifiers	.22**
Linguistic	Qualitative	Past tense		LIWC	-.09
		Present tense			.10
		Future tense			.08
		Negations			-.11
		Swear words			-.09
		Flesch Reading Ease Score		Coh-Metrix	-.08
		Type-token ratio			-.06
		Concreteness			-.08
		Hypernym			-.08
		Log frequency			-.07
Psychological	Qualitative	Positive emotions		LIWC	.16**
		Negative emotions			.10
		Insight			-.01
		Causation			.05
		Discrepancy			.16
		Certainty			.10
		Tentative			.05
		Inhibition			-.02
		See			.06
		Time			.01
		Achievement			-.02
		Assent			.02
Semantic	Qualitative	LSA sentence adjacent		Coh-Metrix	-.07

# Summary of Multiple Regression Analysis

Variables in the Equation			
Variable	<i>B</i>	Std. Error	Beta
Answer Role (Source)	.47	.06	.48***
APTTD	.20	.07	.20**
Positive Emotion	.02	.01	.13*
<b>Note:</b> $R=.57$ ; $N = 173$ ; * $p < .05$ ; ** $p < .01$ ; *** $p < .001$			

Important factors in predicting the grade

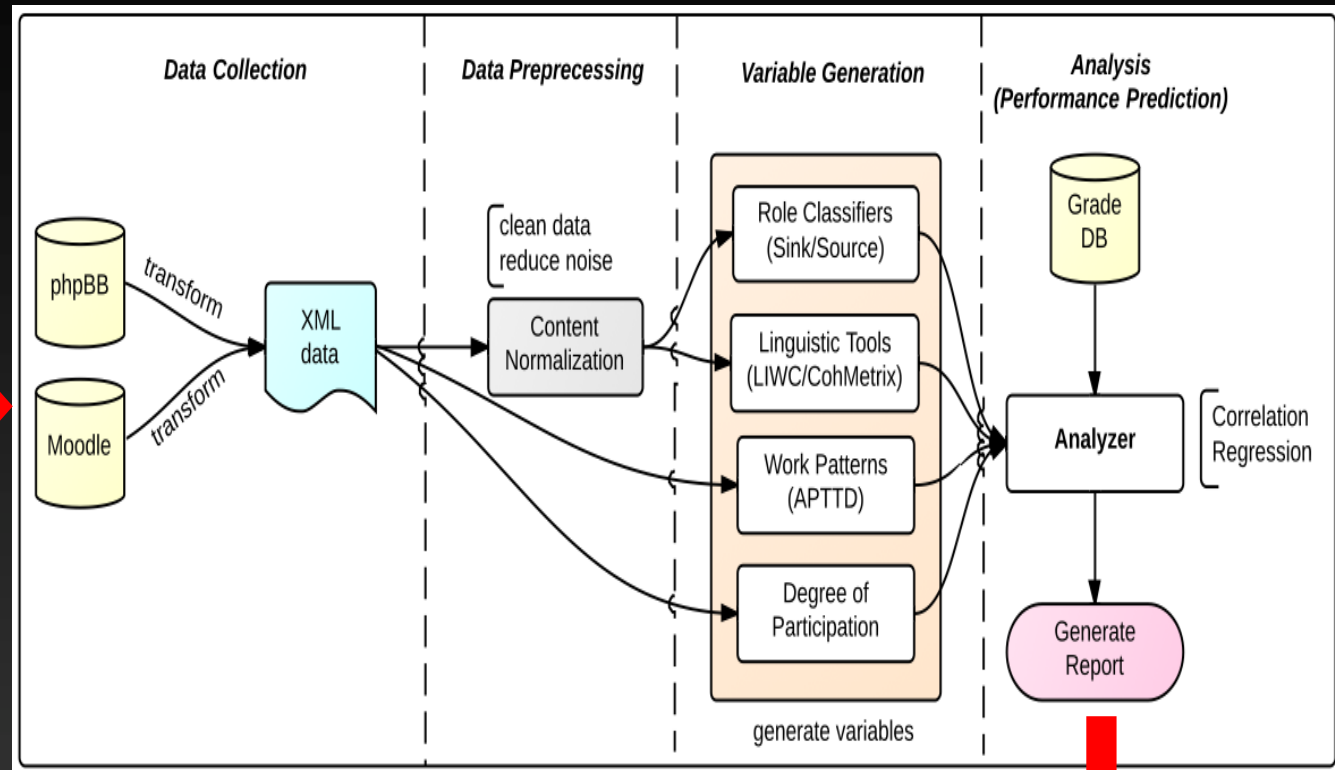
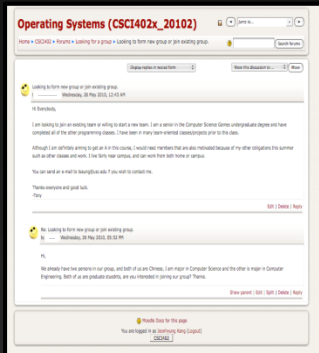
- Qualitative dialogue features (the degree of information provided to others) and
- How early students discuss their problems before the deadline

# Outline

- 
- 1 Online Activities
  - 2 Data Boom in Education
  - 3 Student Performance vs. Online Activity
  - 4 **Big Data Platform for Education**
  - 5 Summary and Conclusion
-



# Pipeline for Data Analysis



# Big Data Platform for Education

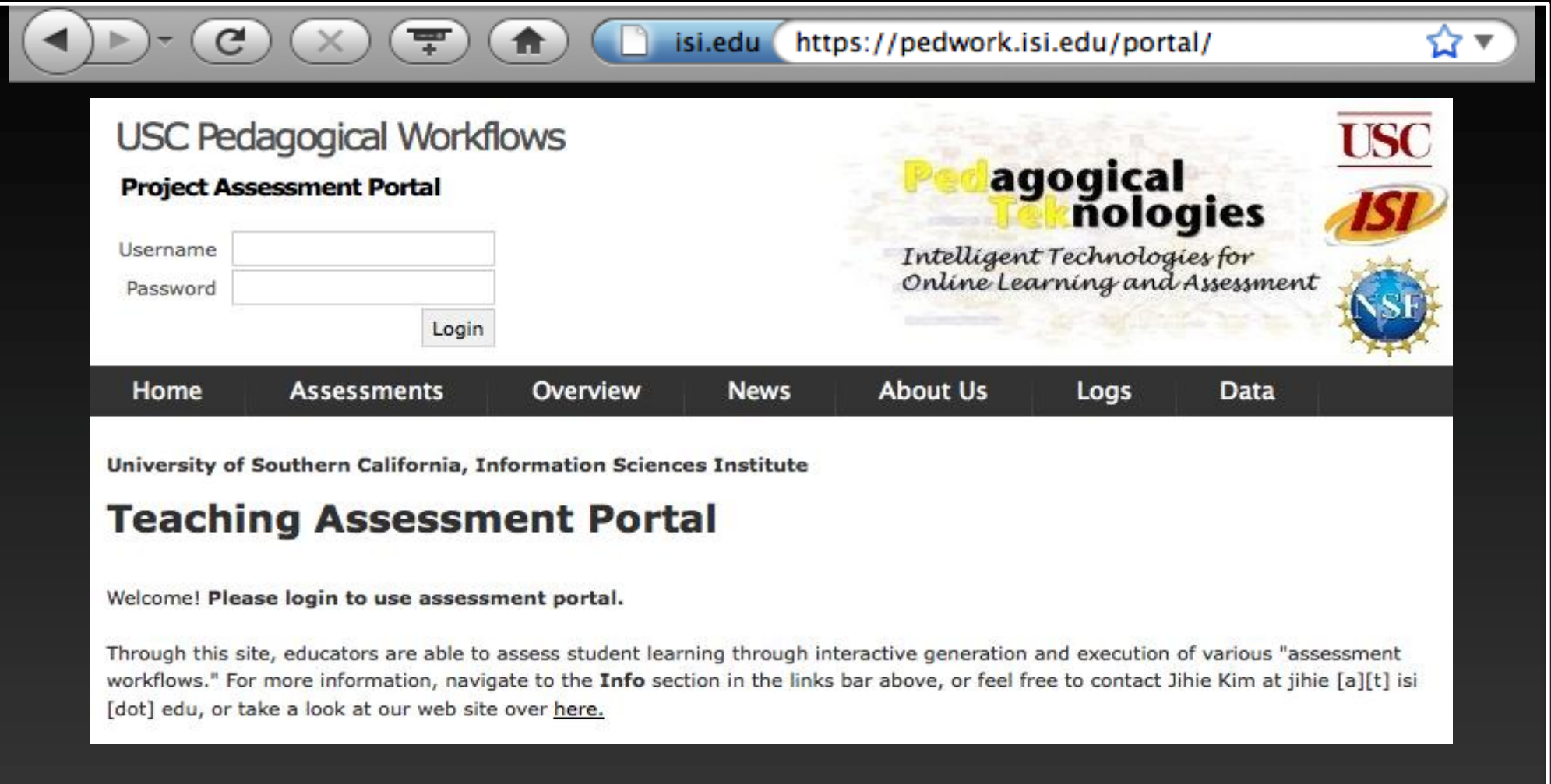
- New Assessment Challenges
  - Continuous and formative assessment
  - Scalable and accessible approaches for handling large amounts of online data
  - Integrating new practices and traditional assessment

# PAWS: Pedagogical Assessment Workflow System

E-Science Workflows → E-Learning Workflows

- Exploit workflow technologies that have been used for large-scale scientific analyses in Physics, Earthquake science, etc
  - Share experiment/analysis process
  - Robust and repeatable analysis
  - Handle a large amount of data
- Wings (Kim et al., 2008; Gil et al., 2011)
  - Knowledge-based workflow reasoning
  - Grid computing environment

# PAWS Portal



The screenshot shows a web browser window with the address bar displaying `https://pedwork.isi.edu/portal/`. The page header includes the title "USC Pedagogical Workflows Project Assessment Portal" and a login form with fields for "Username" and "Password", and a "Login" button. To the right of the login form is a logo for "Pedagogical Technologies" with the tagline "Intelligent Technologies for Online Learning and Assessment", and logos for "USC", "ISI", and "NSF". Below the header is a navigation bar with links: "Home", "Assessments", "Overview", "News", "About Us", "Logs", and "Data". The main content area features the text "University of Southern California, Information Sciences Institute" and "Teaching Assessment Portal". A welcome message states: "Welcome! Please login to use assessment portal." Below this, a paragraph explains the site's purpose: "Through this site, educators are able to assess student learning through interactive generation and execution of various 'assessment workflows.'" It provides contact information for Jihie Kim at `jihie [a] [t] isi [dot] edu` and a link to the website over [here](#).

USC Pedagogical Workflows  
Project Assessment Portal

Username   
Password

**Pedagogical Technologies**  
*Intelligent Technologies for  
Online Learning and Assessment*

USC  
ISI  
NSF

Home Assessments Overview News About Us Logs Data

University of Southern California, Information Sciences Institute

## Teaching Assessment Portal

Welcome! **Please login to use assessment portal.**

Through this site, educators are able to assess student learning through interactive generation and execution of various "assessment workflows." For more information, navigate to the **Info** section in the links bar above, or feel free to contact Jihie Kim at jihie [a] [t] isi [dot] edu, or take a look at our web site over [here](#).

# PAWS for Question-Driven Assessment

Priority Rating		Question
L	M	Which topics have been discussed in the last three semesters?
H	H	Which topics do students ask the most questions about?
H	M	Were all of the questions about topic x answered?
L	H	Which questions were unanswered?
L	M	Do students who participate more often receive better grades?
H	M	Do gender and politeness affect participation?
H		Do more motivated students perform better?
H		Do more confident students participate more?
H	M	Who are the mentors for topic x?
H	H	Which students are confused about topic x?
H	H	Is a student a mentor or help seeker?
H	L	What are his/her strengths?
H	L	Were there similar questions or answers in previous semesters?
H	M	How long did students have to wait for an answer?
H	M	How has student participation changed over time?
H	L	How do currentonline activities differ from previous semesters?

# Assessment Categories Identified with Instructors (Ma et al., ITS 2010)

Category	Workflow Description
Analysis of online activities	Composition of discussion data processing/ Classification steps
Correlation between online activities & performance	Composition of discussion data processing steps, student profiles, and correlation analysis
Correlation between online activities & self-assessment	Composition of self-assessment survey, student activity profiles, and correlation analysis
Student profiling	Composition of student information and discussion data processing/ classification steps
Discussion profiling	Composition of discussion data processing/classification steps and relation analysis
Trend analysis	Splitting of discussion data and iterative analysis
Group comparison	Composition of discussion data processing/ classification steps, student profiles and relation analysis

# PAWS: User Selects An Assessment Question

[Home](#) [Assessments](#) [Overview](#) [News](#) [About Us](#) [Logs](#) [Data](#)

## Select an assessment

Select an assessment from the pulldown menu if you know which assessment you wish to run and its input data requirements.

ForumParticipation\_AnswerWaitTimeAnaysis  
ActivityChangeOverSemester  
ForumParticipation\_FrequencyStatistics\_Project  
Forum\_Topic\_Analysis

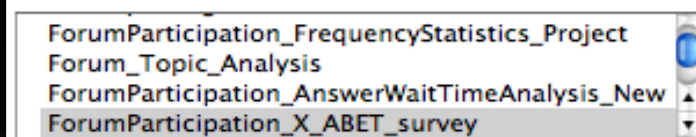
## ...Or browse assessments and select

Browse a selection of current assessments and select one to run by clicking on its link. Examples are provided for each type of input file via its link, for testing and for populating with personal data.

<b>Assessment</b>	<b><u>Forum Participation: Answer Wait Time Analysis</u></b>
<b>Example Input</b>	<u>Discussion Board Posts</u> , <u>User IDs</u>
<b>Description</b>	This assessment computes how long students waited for a reply to their initial questions within discussion forums, per forum. Current statistics include number of students, and the wait time in minutes. There are two input files for this assessment, both derived from the discussion forum data*.
<b>Assessment</b>	<b><u>Forum Topic Analysis</u></b>
<b>Example Input</b>	<u>Label Model</u> , <u>LDA Model</u> , <u>Discussion threads</u>
<b>Description</b>	This assessment analyzes terms used in discussions and computes a topic distribution per forum. In creating term-topic relation statistics, we use a Labeled LDA (Latent Dirichlet allocation) model that is created from the course assignment documents. No manual input is required: The PedWorkflows project team will work with individual course instructors to create this assessment.



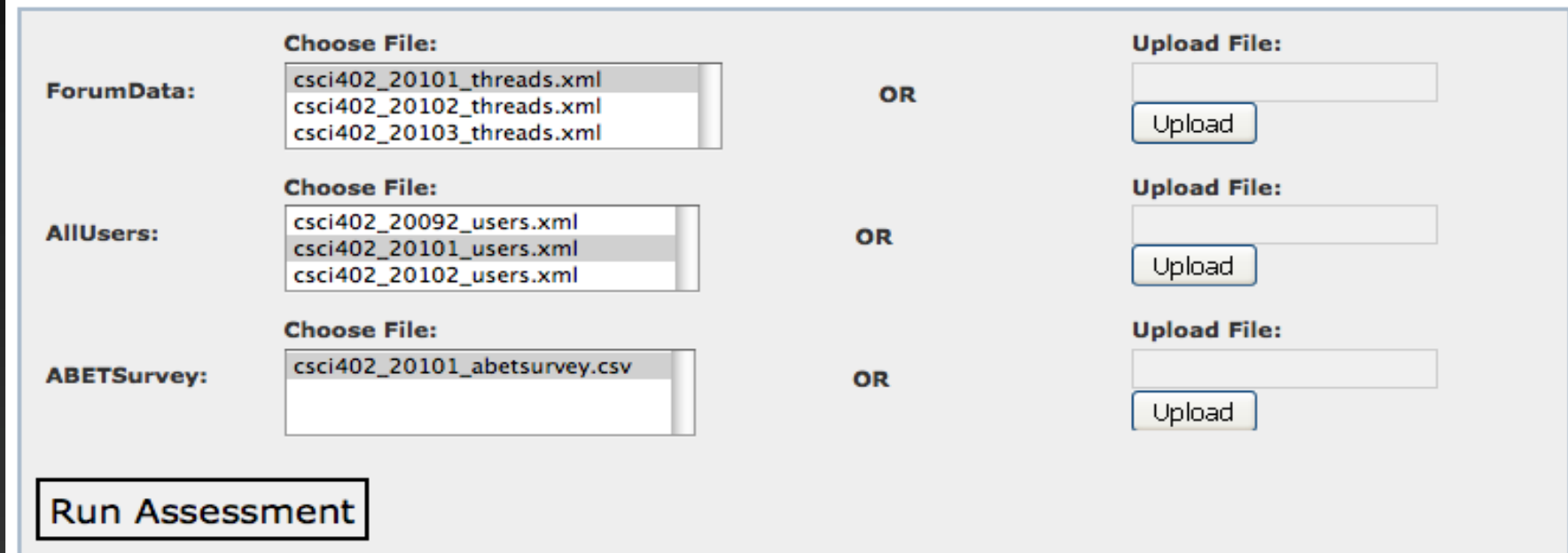
# PAWS: User selects dataset to use



A screenshot of a web application's dataset selection dropdown menu. The menu is open, showing four options: 'ForumParticipation\_FrequencyStatistics\_Project', 'Forum\_Topic\_Analysis', 'ForumParticipation\_AnswerWaitTimeAnalysis\_New', and 'ForumParticipation\_X\_ABET\_survey'. The last option is highlighted with a grey background.

## *ForumParticipation X ABET survey*

This assessment computes correlations between message incidence and the ABET measurement constructs, Multidisciplinary, Motivation, Leadership and Efficacy. ABET Survey data must be manually uploaded to the assessment. Starting in the fall of 2010, courses that use the ISI Moodle virtual learning environment will have automated access to the survey results for use in assessments.

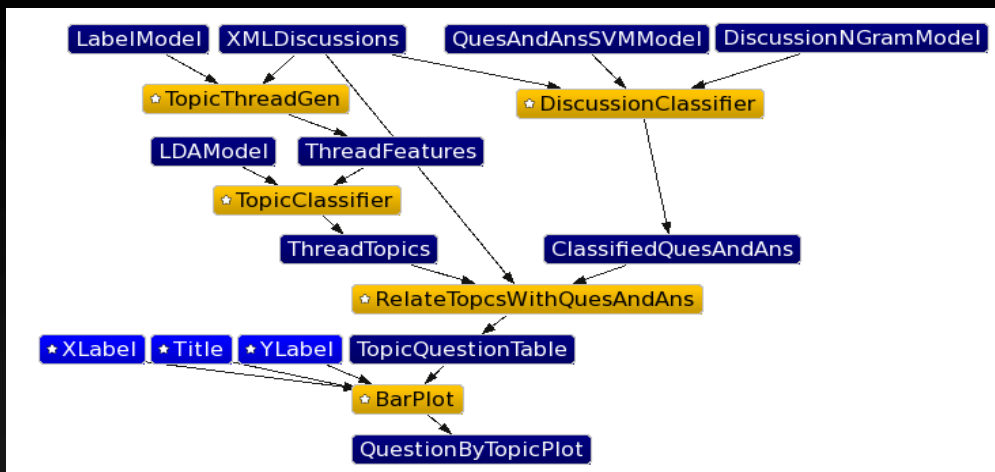


A screenshot of the PAWS assessment configuration interface. It features three rows of input fields for data selection, each with a 'Choose File' and an 'Upload File' option, separated by 'OR' labels. The first row is for 'ForumData' with files 'csci402\_20101\_threads.xml', 'csci402\_20102\_threads.xml', and 'csci402\_20103\_threads.xml'. The second row is for 'AllUsers' with files 'csci402\_20092\_users.xml', 'csci402\_20101\_users.xml', and 'csci402\_20102\_users.xml'. The third row is for 'ABETSurvey' with the file 'csci402\_20101\_abetsurvey.csv'. At the bottom left is a 'Run Assessment' button.

<b>ForumData:</b>	<b>Choose File:</b> csci402_20101_threads.xml csci402_20102_threads.xml csci402_20103_threads.xml	OR	<b>Upload File:</b> <input type="text"/> <input type="button" value="Upload"/>
<b>AllUsers:</b>	<b>Choose File:</b> csci402_20092_users.xml csci402_20101_users.xml csci402_20102_users.xml	OR	<b>Upload File:</b> <input type="text"/> <input type="button" value="Upload"/>
<b>ABETSurvey:</b>	<b>Choose File:</b> csci402_20101_abetsurvey.csv <input type="text"/>	OR	<b>Upload File:</b> <input type="text"/> <input type="button" value="Upload"/>

[Show/Hide Workflow Diagram](#)

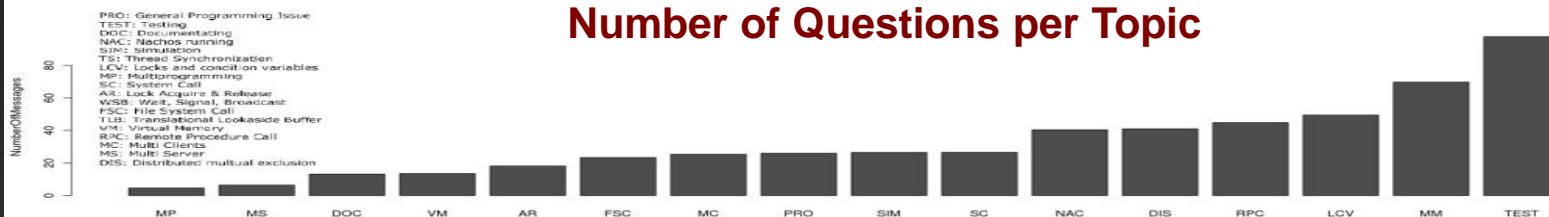
# Which topics do students ask questions about?



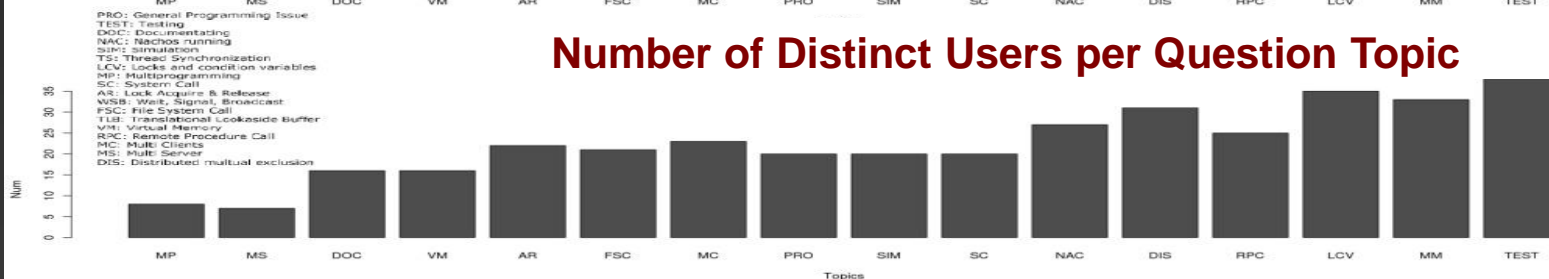
Workflow:  
pipeline for data analysis



Number of Questions per Topic

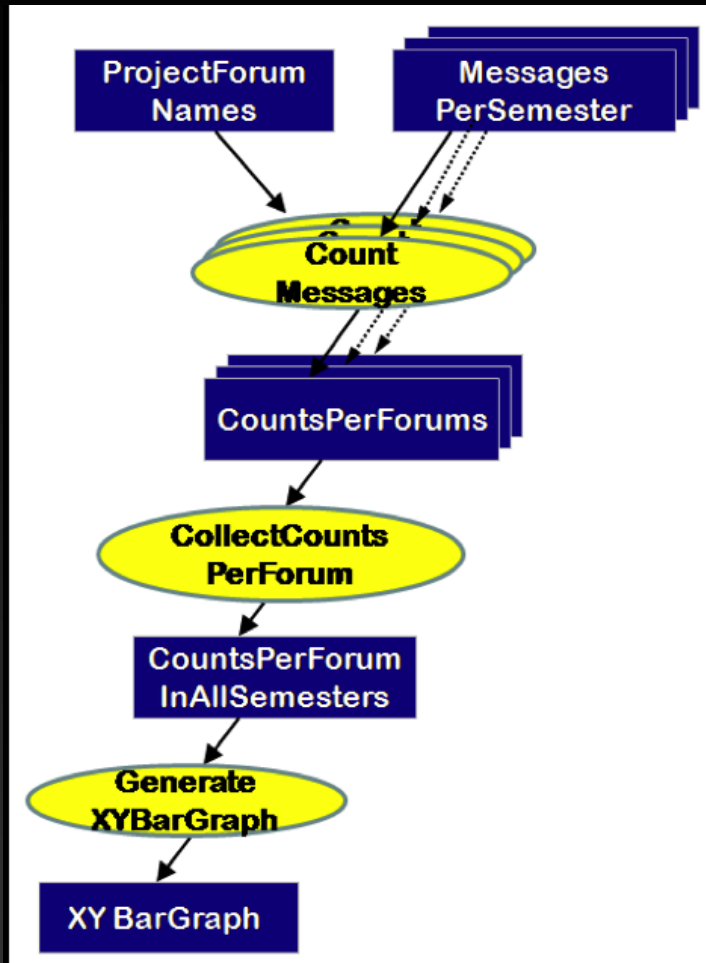


Number of Distinct Users per Question Topic

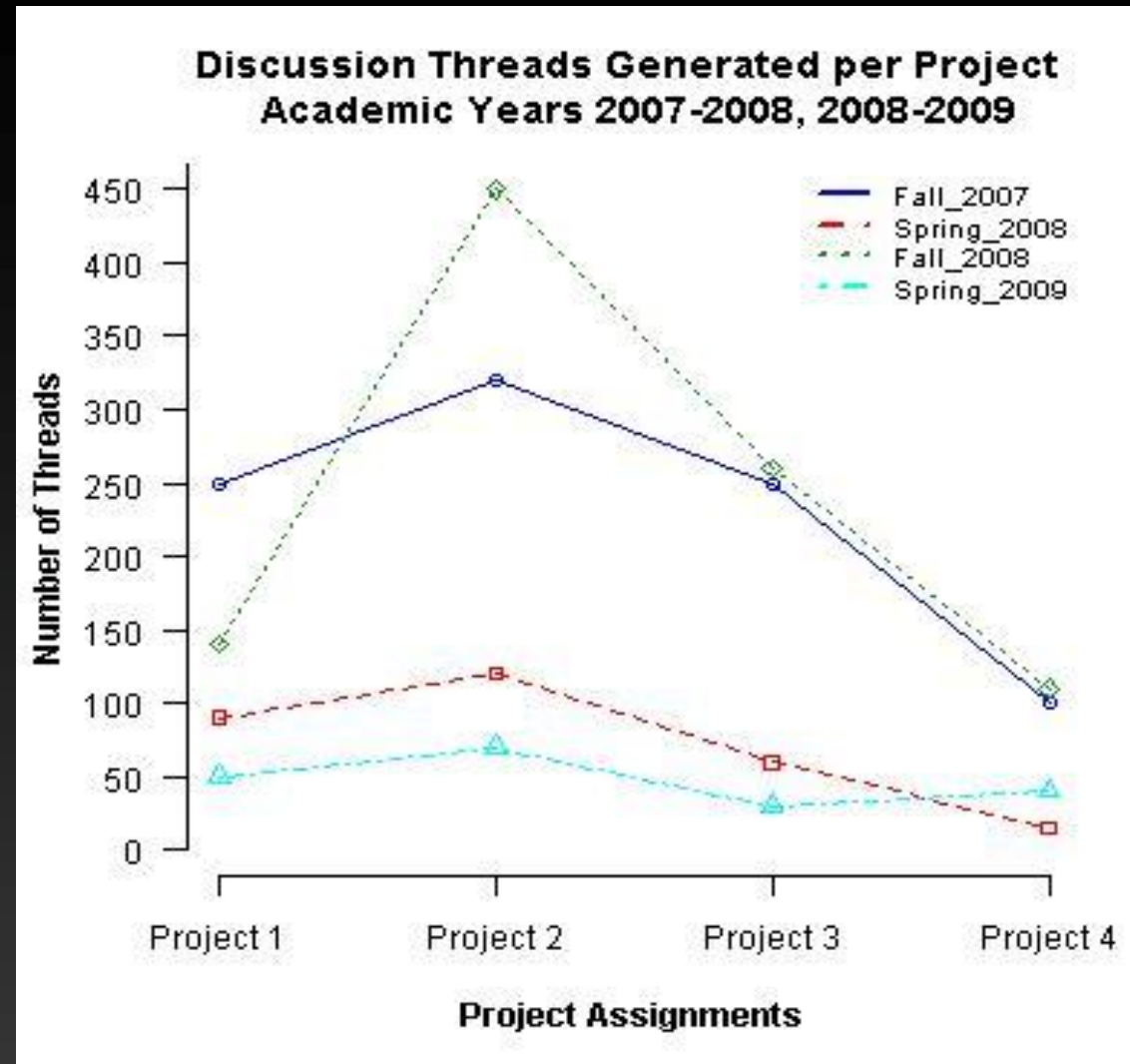


Report for teachers

# Compare Student Activities across Multiple Semesters



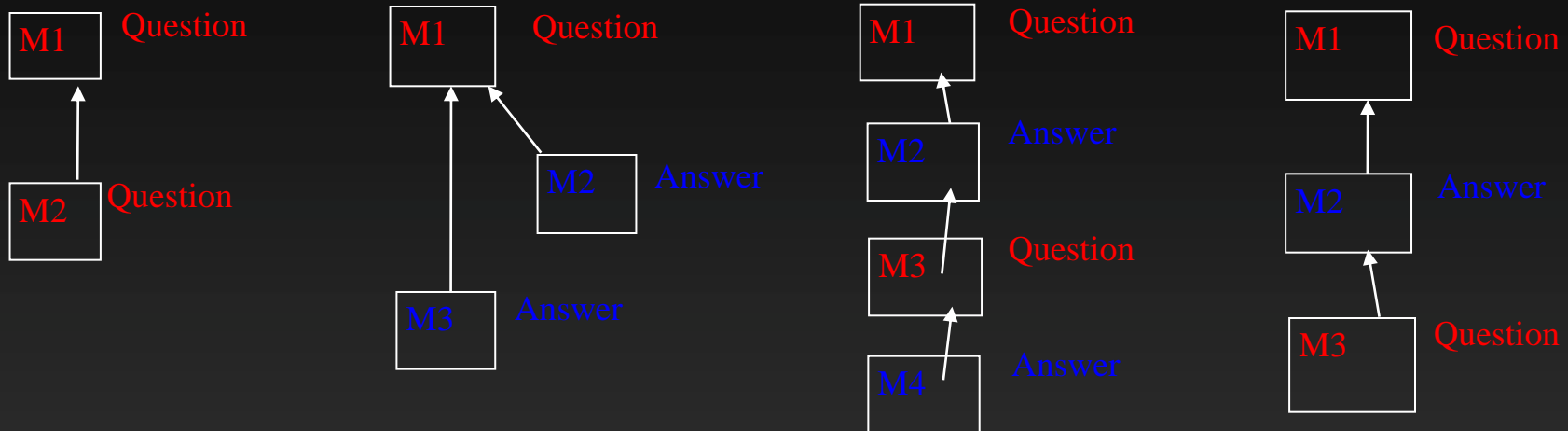
Workflow:  
pipeline for data analysis



Report for teachers

# Finding Discussion Threads that Need Attention: Unanswered Questions

- 1) whether the given thread contains questions
- 2) whether the questions were answered or not



- 70-75% of the predictions from the system were consistent with human answers (Ravi & Kim, AIED 2007)

# Conclusion

## Big Data Platform for Education:

### *Scalable and Robust Analytics*

- Powerful tool for assessing impact of online technologies
- Support continuous, robust/repeatable assessment
- Make use of state-of-the-art NLP and machine learning tools
- Combine traditional methods with computational analysis
- Platform independent
- Increase accessibility

# Outlook: Data Analytics for Education

- Formative assessment
- Support individualized and adaptive learning
- Continuous feedback to students and teachers
- Adaptive courses/ e-books
- Information sharing



Thank you

