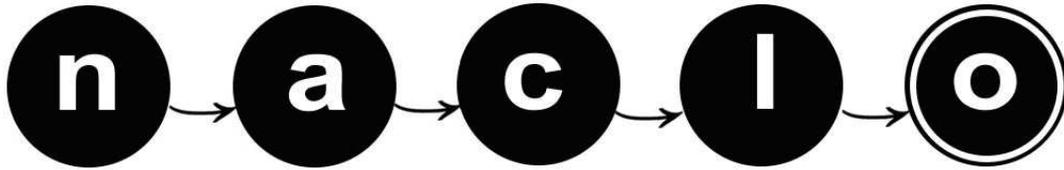


The Association for Computational Linguistics
North American Chapter



***The Third
Annual***
**North American
Computational
Linguistics
Olympiad**
2009
www.naclo.cs.cmu.edu

Open Round
February 4, 2009



The North American Computational Linguistics Olympiad
www.naclo.cs.cmu.edu

Contest Booklet

Your Name: _____

Registration Number: _____

Your School: _____

City, State, Zip: _____

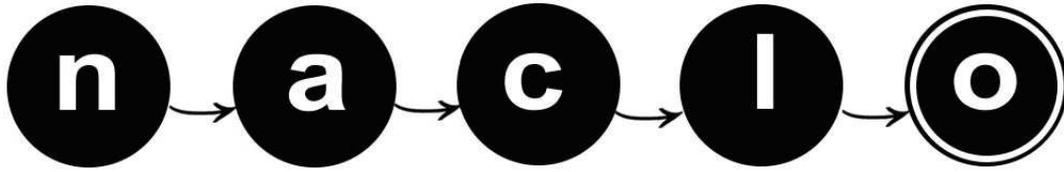
Your Grade: _____

Start Time: _____

End Time: _____

Your Teacher's Name: _____

Please also make sure to write your registration number and your name on each page that you turn in. Each problem will be graded by a different judge and pages with no registration numbers will not be graded.



Welcome to the third annual North American Computational Linguistics Olympiad! You are among the few, the brave, the brilliant, to participate in this unique event! In order to be completely fair to all participants across North America, we need you to read, understand and follow these rules completely.

Rules

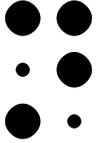
1. The contest is three hours long and includes six problems, labeled A to F.
2. Follow the facilitators' instructions carefully.
3. If you want clarification on any of the problems, talk to a facilitator. The facilitator will consult with the jury before answering.
4. You may not discuss the problems with anyone except as described in item 3.
5. Each problem is worth a specified number of points, with a total of 100 points. On all problems, points are given for "practice," that is, for getting the right answers. Some problems also assign points for "theory," that is for written descriptions of how you solved the problem. You should therefore show all your work.
6. We will grade only work in this booklet. All your answers should be in the spaces provided in this booklet. **DO NOT WRITE ON THE BACK OF THE PAGES.**
7. Write your name and registration number on each page:
Here is an example: Jessica Sawyer #850
8. The top 100 participants (approximately) across the continent in the open round will be invited to the second (invitational) round on March 11, 2009.
9. Each problem has been thoroughly checked by linguists and computer scientists as well as students like you for clarity, accuracy, and solvability. Some problems are more difficult than others, but all can be solved using ordinary reasoning and analytic skills. You don't need to know anything about linguistics or about these languages in order to solve them.
10. If we have done our job well, almost no one will solve all these problems completely in the time allotted. So don't be discouraged if you don't finish everything.
11. If you have any comments, suggestions or complaints about the competition, we ask you to remember these for the web based evaluation. We will send you an e-mail shortly after the competition is finished with instructions on how to fill it out.
12. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE!**

Oh, and have fun!

(20 points)

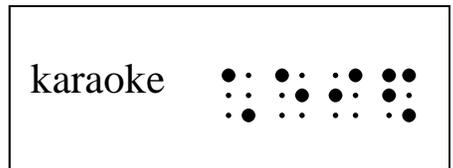
(A) Tenji karaoke (I/I)

Braille is a tactile writing system, based on a series of raised dots, that is widely used by the blind. It was invented in 1821 by Louis Braille to write French, but has since been adapted to many other languages.



English, which uses the Roman alphabet just as French does, required very little adaptation, but languages that do not use the Roman alphabet, such as Japanese, Korean, or Chinese, are often organized in a very different manner!

To the right is a Japanese word written in the *tenji* ("dot characters") writing system. The large dots represent the raised bumps; the tiny dots represent empty positions.



A1. The following *tenji* words represent *atari*, *haiku*, *katana*, *kimono*, *koi*, and *sake*. Which is which? You don't need to know either Japanese or Braille to figure it out; you'll find that the system is highly logical.

a. _____	b. _____
c. _____	d. _____
e. _____	f. _____

A2. What are the following words?

g. _____	h. _____
----------	----------

A3. Write the following words in *tenji* characters. Please write very legibly!!

i. samurai	j. miso
------------	---------



(5 points)

(B) Nok-nok! (1/2)

Christopher Robin is the most educated friend of Winnie-the-Pooh, and he can spell a lot of simple words, and even some long and delicate words, but unfortunately he often spells them incorrectly. For example, he has helped Owl to write the following two notices on the door of Owl's residence at The Chestnuts:

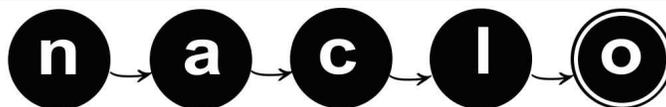
Ples ring if an rnsr is reqird.

Plez cnoke if an rnsr is not reqid.

Although Owl is very proud of these notices, he is also concerned that some of his friends may not understand them; in fact, even Owl himself finds the notices a bit confusing.

To help Christopher Robin with his spelling, Winnie-the-Pooh and Owl have bought him an electronic spelling tutor, which pronounces various words and asks the user to spell them. If the user makes a mistake, the tutor shows the correct spelling, along with a comment on the accuracy of the user's spelling; it uses four comments: *almost right*, *quite close*, *a bit confusing*, and *very confusing*. For instance, Christopher Robin has received the following feedback during his initial experiments with the tutor:

Spelling by Christopher Robin	Correct Spelling	Comment
flocinaucinihilipilification	floccinaucinihilipilification	almost right
owll	owl	almost right
pseudopseudohipoparathyroidism	pseudopseudohypoparathyroidism	almost right
ples	please	quite close
reqird	required	quite close
rnsr	answer	quite close
antidisestablishmentarianism	antidisestablishmentarianism	quite close
wol	owl	quite close
humuhumunukunukuapua'a	humuhumunukunukuapua'a	quite close
plez	please	a bit confusing
cnoke	knock	a bit confusing
rnsr	answer	a bit confusing
reqid	required	a bit confusing
pneumonoultramicroscopic-cilиковolkanokoniosis	pneumonoultramicroscopic-silicovolcanokoniosis	a bit confusing
mispeln	misspelling	very confusing
mestipenk	mistyping	very confusing



(B) Nok-nok! (2/2)

Your task is to determine how the tutor chooses its comments and give the appropriate comment for each of the following six misspellings of the word “*typo*”. You do *not* need to explain your answers; just indicate the right comments.

For each misspelling, put the “X” sign in the column with the appropriate comment. Please note that each row in the table must have *exactly one* “X” sign. If you mark multiple comments for the same misspelling, it will be graded as a wrong answer. Also note that there is no penalty for a wrong answer; thus, if you are unsure of the right

Misspelling of “ <i>typo</i> ”	Comment by the Tutor			
	Almost right	Quite close	A bit confusing	Very confusing
oooo				
opyt				
pyto				
typ				
typa				
typotypo				

As a side note, the dictionary definitions of the long and delicate words misspelled by Christopher Robin are as follows; these definitions are unrelated to the problem.

floccinaucinihilipilification: act or habit of estimating or describing something as worthless, or making something to be worthless by deprecation.

pseudopseudohypoparathyroidism: inherited disorder that closely simulates the symptoms, but not the consequences, of pseudohypoparathyroidism; thus, it has mild or no manifestations of hypoparathyroidism or tetanic convulsions.

antidisestablishmentarianism: nineteenth century movement in England opposed to the separation of church and state.

humuhumunukunukuapua’a: one of several species of triggerfish.

pneumonoultramicroscopicsilicovolcanokoniosis: lung disease caused by the inhalation of very fine silica dust, mostly found in volcanoes



(20 points)

(C) Letters for Cuzco (1/4)

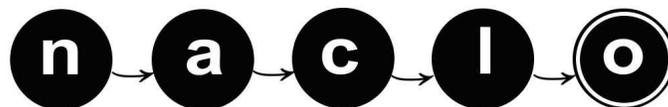
Orthography design is the process of developing an alphabet and spelling rules for a language. A good orthography has several features:

- * Given a spoken word, there's no question of how to spell it.
- * Given a written word, there's no question of how to pronounce it.
- * In the modern world, it's increasingly important that it be reasonably easy to type!

Quechua is spoken today by millions of people in Peru, Ecuador, and Bolivia, the descendants of the citizens of the Incan Empire. Quechua speakers are rapidly joining the Information Age, and both Google and Microsoft Windows now come in Quechua!

Like in English, there are more sounds in Quechua than there are letters on a keyboard, but there are ways around that. For example, we can assign one letter to multiple sounds so long as a reader can always predict, from its position in the word or from other letters in the word, which sound is meant. So if the sound [b] only ever occurs right after [m], and [p] never occurs right after [m], we can just write "p" for both, since you'll be able to predict from the previous letter whether "p" means [b] or [p].

This "phonemic principle" is the central principle of most orthographies, not just because it reduces letters but also because our minds categorize sounds in the same way.

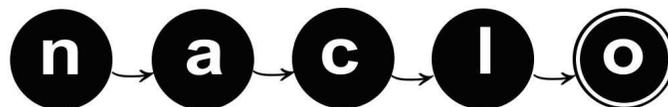


(C) Letters for Cuzco (2/4)

Here are 33 words in Cuzco Quechua, as they are pronounced but not necessarily as they are written. [q] and [χ] represent special sounds that don't occur in English.

awtu	car	qasi	free	seqay	to climb
kanka	roasted	qatoχ	merchant	sikasika	caterpillar
karu	far	qatuy	to barter	sipeχ	murderer
kiru	teeth	qatisaχ	I will follow	sipiy	to kill
kisa	nettle	qelqax	writer	soχtaral	sixty cents
kisu	cheese	qelqay	to write	sunka	beard
kunka	neck	qolqe	silver	toχra	ball of ash
kusa	great	qosa	husband	uyariy	to listen
layqa	witch	qosqo	Cuzco	uywax	caretaker
oqe	spotted	saqey	to abandon	waleχ	a lot
qasa	frost	saxsa	striped	weqaw	waist

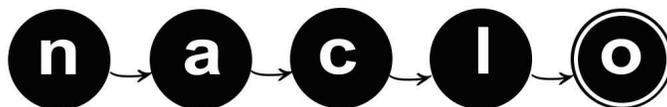
C1. Show that we don't need separate letters for [q] and [χ].



(C) Letters for Cuzco (3/4)

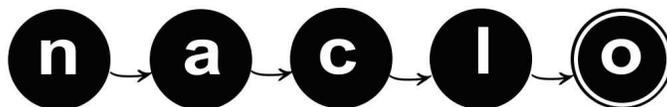
C2. Show that we can't represent [a] and [i] by the same letter.

C3. Show that we can't represent [a] and [e] by the same letter.



(C) Letters for Cuzco (4/4)

C4. Most modern Quechua orthographies get by with only *three* of the five vowels [a], [e], [i], [o], and [u]. Show how this is possible.



(25 points)

(D) You will be laughing (I/I)

The following Guaraní verb forms are listed along with their English translations.

N.	Guaraní	English translation
1	japyhyta	We will be catching
2	nohyvykõiri	He is not enjoying
3	ombokapu	He is shooting
4	pemomaitei	You are greeting
5	ndokarumo'ãi	He will not be eating
6	ndapevo'oima	You were not taking
7	napekororõmo'ãi	You will not be crying
8	noñe'ëi	He is not talking
9	okororõ	He is crying
10	ndajajupirima	We were not waking up
11	ahyvykõima	I was enjoying
12	añe'ëta	I will be talking
13	namomaiteiri	I am not greeting
14	japurahei	We are singing

D1. Translate into English.

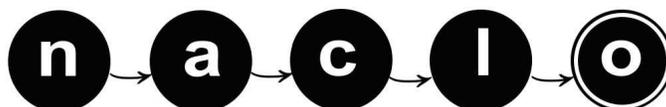
- (a) *akaruma* (b) *ojupita* (c) *ndavo'omo'ãi* (d) *napekororõ* (e) *ndapyhyima*
 (a) (b) (c) (d) (e)

D2. Translate into Guaraní.

- (f) *you are not shooting* (g) *he is not singing* (h) *we will be eating* (i) *I will not be singing*
 (f) (g) (h) (i)

Notes: “you” is always plural in the sentences above. A squiggle over a vowel indicates that it is nasal (pronounced partly through the nose). The letter ñ is pronounced like the sound in the middle of “piñata” or “onion”. The letter y is pronounced like the “u” in “cut”. The letter j and the apostrophe (') are specific consonants.

Guaraní is one of the official languages (along with Spanish) of Paraguay, where it is spoken by 94% of the population.



(25 points)

(E) Summer Eyes (1/2)

Below are two news stories, each of which has had three sentences automatically selected as a summary by a computer, based on a number of criteria. The criteria are the same for each story. Within each story, some scores may depend on other sentences. Below each story, a change has been proposed to one of the sentences.

E1. Rescore the sentences after the change, using the same criteria: in each score box, either write the appropriate new score if it's different from the old one, or LEAVE IT BLANK IF IT WOULD BE UNCHANGED. For instance, a correctly formatted (but wrong) possible answer for the first sentence is:

1	X	3.4	0	1	4	0	-1	7.4	Taiwan authorities say a powerful earthquake has struck the southeastern part of the island.
		1.7	3				-1	2	

E2. For each story, put the symbol X in exactly three of the boxes in the first column, the ones corresponding to the sentences in the new summary.

E3: Give the added sentence in the second story scores according to the same criteria.

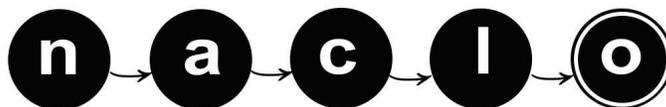
Story 1 input:

Powerful earthquake strikes Taiwan

Sentence Number	Criteria						Total score	Sentence	
	1st	2nd	3rd	4th	5th	6th			
1	X	3.4	0	1	4	0	-1	7.4	Taiwan authorities say a powerful earthquake has struck the southeastern part of the island.
2		2.3	0	0	0	0	-1	1.3	There were no immediate reports of damages or injuries from the Tuesday morning quake, authorities said.
3	X	1.2	1	2	2	2	-1	7.2	The Taiwan Central Weather Bureau says the magnitude 6.0 quake struck just offshore, near a sparsely populated area about 20 miles (30 kilometers) north of the city of Taidung.
4		0.1	2	1	0	1	-1	3.1	However, the U.S. Geological Survey says the quake had a magnitude of 5.2.
5	X	0.0	3	1	0	1	-1	4.0	Buildings shook in Taipei about 90 miles (150 kilometers) to the northwest of the epicenter.

Story 1 output: An “extractive summary” consisting of sentences 1, 3, and 5.

Story 1 proposed change: Add “in Taipei” between “damages or injuries” and “from”. in sentence 2.



(E) Summer Eyes (2/2)

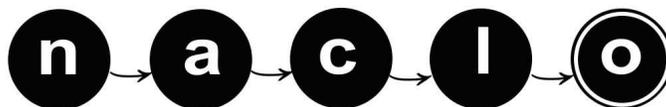
Story 2 input:

Interior minister of Mexico dies in plane crash that killed eight								
Sentence Number	Criteria						Total score	Sentence
	1st	2nd	3rd	4th	5th	6th		
1 X	3.8	0	1	7	0	-3	8.8	Mexico's interior minister has died in a plane crash, after the small aircraft he was travelling in plummeted into rush hour traffic in the nation's capital.
2	2.7	0	1	1	1	-1	4.7	He was one of the architects of conservative Felipe Calderón's 2006 election victory.
3	1.6	0	2	1	1	-2	3.6	Calderón described Mouriño as "a compatriot who worked for the service of his county."
4	0.5	0	1	3	1	-1	4.5	Jose Luis Santiago Vasconcelos, the deputy attorney-general until recently and a key player for years in the war on drug cartels, also died in the crash.
5 X	0.4	0	3	2	3	-3	5.4	Communications minister Luis Tellez said everything pointed to the crash being an accident but Jorge Lara, vice-president of a national civil aviation rescue committee, told Reuters "It could have been anything, from mechanical failure to sabotage."
6	0.3	0	1	3	0	-4	0.3	A colleague of Lara at the scene said it appeared something happened to the plane in the air and it nosedived into the ground.
7	0.2	1	0	0	0	-1	0.2	Surrounding buildings were not damaged.
8	0.1	2	2	2	2	-3	5.1	Newspaper El Universal reported an air traffic controller as saying the aircraft had been coming in to land at Mexico City's airport when it hit the ground between tall office buildings.
9 X	0.0	3	2	4	1	-3	7.0	Tellez said there was no contact with the plane, which had been returning from a trip to the central city of San Luis Potosí, in the moments before the crash.

Story 2 output: An "extractive summary" consisting of sentences 1, 5, and 9.

Story 2 proposed change: Add the following sentence between the current second and third sentences:

n/a	Seven other people were killed and 40 injured when the government jet smashed into the streets of Mexico City yesterday, setting several cars ablaze, Reuters reported.							
3								



(5 points)

(F) Help my camera! (I/I)

In the following dialogues, the computer is misunderstanding the human because it is misinterpreting certain words or phrases:

Example 1

Human: There's this restaurant on Bancroft that's supposed to be really good that I heard about from my mother. Can you help me find it?

Computer: Where did you last see your mother?

Example 2

Human: My friend is looking for information about on-line movie databases. What should I tell him?

Computer: You can tell movie databases about movies that you have seen.

Example 3

Human: My parents are having a hard time choosing a new digital camera. How can I help them?

Computer: I'm afraid I don't understand your question. People don't help cameras.

F1. What are the words (or phrases) that the computer is misinterpreting?

Ex1: _____

Ex2: _____

Ex3: _____

F2. What rule is the computer using to interpret those words (or phrases)?

F3. Give a better rule that would make the computer interpret the words (or phrases) correctly in these examples.



NACLO 2009 organizers

General chair:

Lori Levin, Carnegie Mellon University

Program committee chair:

Dragomir Radev, University of Michigan

Program committee:

Emily Bender, University of Washington
Aleka Blackwell, Middle Tennessee State University
Bozhidar Bozhanov
Eric Breck, Rhodes College
Ivan Derzhanski, Bulgarian Academy of Sciences
Jason Eisner, Johns Hopkins University
Eugene Fink, Carnegie Mellon University
Adam Hesterberg, Princeton University
Anatole Gershman, Carnegie Mellon University
Boris Iomdin, Russian Academy of Sciences
Joshua Katz, Princeton University
Lori Levin, Carnegie Mellon University
Patrick Littell, University of British Columbia
Scott Mackie, University of British Columbia
Ruslan Mitkov, University of Wolverhampton
David Mortensen, University of Pittsburgh
Barbara Partee, University of Massachusetts
Thomas Payne, University of Oregon
James Pustejovsky, Brandeis University
Harold Somers, Dublin City University
Ekaterina Spriggs, Carnegie Mellon University
Richard Sproat, Oregon Health and Science University
Todor Tchernakov, Université Lumière Lyon 2
Xiaojin "Jerry" Zhu, University of Wisconsin
Richard Wicentowski, Swarthmore College

School liaison:

Amy Troyani, Taylor Allderdice High School

Administrative assistant:

Mary Jo Bensasi, Carnegie Mellon University



NACLO 2009 organizers (cont'd)

Problem credits:

Problem A: Patrick Littell
Problem B: Eugene Fink
Problem C: Patrick Littell
Problem D: Bozhidar Bozhanov
Problem E: Dragomir Radev and Adam Hesterberg
Problem F: Emily Bender

Other members of the organizing committee:

Eugene Fink, Carnegie Mellon University
Ketty Gann, Boeing
Patrick Littell, University of British Columbia
Thomas Payne, University of Oregon
Susanne Vejdemo, the Linguist List
Richard Wicentowski, Swarthmore College

Web site and registration:

Ida Mayer, Carnegie Mellon University
Eugene Fink, Carnegie Mellon University

Jury members

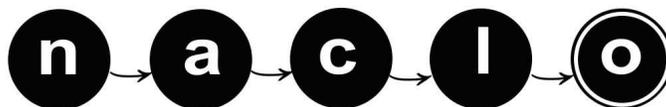
The program committee plus:
Jae-Kyu Lee, Princeton University
Helen Mukomel, DeVry University

US Team coaches:

Dragomir Radev, University of Michigan, head coach
Lori Levin, Carnegie Mellon University, coach
Adam Hesterberg, Princeton University, assistant coach

Student assistants:

Adam Emerson, University of Michigan
Blumie Gurarie, University of Michigan
Mercedes Harvey, University of Michigan
Nate LaFave, University of Michigan
Victor Pudeyev, University of Michigan
Vahed Qazvinian, University of Michigan
Meredith Rogan, University of Michigan
David Ross, University of Michigan
Laine Stranahan, University of Michigan



NACLO 2009 organizers (cont'd)

Contest site coordinators:

Brandeis University: James Pustejovsky
Carnegie Mellon University and University of Pittsburgh: Lori Levin and David Mortensen
Cornell University: Claire Cardie, Randy Hess, Lillian Lee, and Jennifer Wofford
Dalhousie University: Connie Adsett
Georgetown University: Graham Katz
Middle Tennessee State University, Aleka Blackwell
Minnesota State University, Mankato: Rebecca Bates
New York University: Adam Meyers and Satoshi Sekine
Ohio State University: Chris Brew
Princeton University: Christiane Fellbaum, Adam Hesterberg, and Jae-Kyu Lee
San Diego State University: Mark Gawron
San José State University: Roula Svorou
Simon Fraser University: John Alderete, Cliff Burgess, and Maite Taboada
Stanford University: Josh Falk, Jenny Finkel, Dan Jurafsky, and Kyle Noe
University of Great Falls: Porter Coggins
University of Illinois: Roxana Girju and Julia Hockenmaier
University of Illinois, Chicago: Robert Sloan
University of Lethbridge: Kevin Grant
University of Maryland, Baltimore County: Marjorie McShane and Sergei Nirenburg
University of Memphis: Vasile Rus
University of Michigan: Steve Abney and San Duanmu
University of North Texas: Rada Mihalcea
University of Oregon: Thomas Payne
University of Ottawa: Diana Inkpen and Stan Szpakowicz
University of Pennsylvania: Mitch Marcus and Cheryl Hickey
University of Southern California: David Chiang
University of Texas at Dallas: Vincent Ng
University of Texas at El Paso: Gabby Gandara
University of Utah: Naomi Fox
University of Washington: Jim Hoard
University of Wisconsin: Nathanael Fillmore and Xiaojin Zhu
High school sites: Dragomir Radev



NACLO 2009 sponsors

Sponsorship chair:

James Pustejovsky, Brandeis University

Corporate, academic, and government sponsors

National Science Foundation

Google

Microsoft

Everyzing

M*Modal

JUST. Systems

The North American Chapter of the Association for Computational Linguistics (NAACL)

Cambridge University Press

Oxford University Press

Carnegie Mellon University's Language Technologies Institute

University of Michigan

Brandeis University

University of Pittsburgh Linguistics Department

Individual contributions

Helga and Charles Marqusee

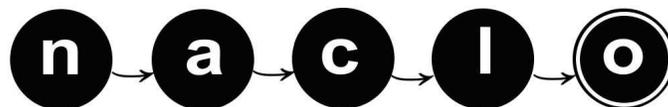
Linnae and David Evans

Special thanks to:

Tanya Korelsky, NSF

More than 80 high school teachers from 27 states and provinces

And many other individuals and organizations



NACLO 2009 sites



Cornell University



THE UNIVERSITY OF UTAH

Carnegie Mellon



San José State
UNIVERSITY



UMBC
AN HONORS UNIVERSITY IN MARYLAND



THE UNIVERSITY
of
WISCONSIN
MADISON



USC
UNIVERSITY
OF SOUTHERN
CALIFORNIA

University of
Lethbridge




uOttawa



NEW YORK UNIVERSITY

SFU

SIMON FRASER UNIVERSITY
THINKING OF THE WORLD



University of Great Falls



SAN DIEGO STATE
UNIVERSITY

THE UNIVERSITY OF
MEMPHIS



Penn
UNIVERSITY OF PENNSYLVANIA

STANFORD
UNIVERSITY



ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

UNIVERSITY OF
NORTH TEXAS



Brandeis University



UNIVERSITY OF MICHIGAN


MINNESOTA STATE
UNIVERSITY
MARSHFIELD



PRINCETON
UNIVERSITY

MIDDLE
TENNESSEE
STATE UNIVERSITY



UNIVERSITY OF OREGON



DALHOUSIE
UNIVERSITY
Inspiring Minds

UIC

Georgetown
UNIVERSITY

as well as more than 80 high schools throughout the USA and Canada