The North American Computational Linguistics Olympiad

 What? A competition for high school students interested in linguistics, languages, and computation Paper and pencil contest No prerequisites needed Easy problems in the Open Round Harder problems in the Invitational Round - national team selection 	 When? Open Round – January 30, 2014 Invitational Round – March 13, 2014 Where? 200 sites in the USA and Canada Some at universities Others are local schools Register at the NACLO site shown below Contact: naclo14org@umich.edu
 Did you know? More than 7,000 languages are spoken in the world Human language is central to human communication and social interaction Human languages exhibit interesting patterns and structure You can practice scientific reasoning (forming hypotheses and using data to support them) 	 Linguistics The study of human language Phonetics: how spoken sounds are produced and heard Syntax: how sentences are structured Semantics: what do words and sentences mean Sociolinguistics: how language use varies socially
 Computational Linguistics Computational Linguistics is about teaching computers to understand human language It is the basis of search engines such as Google, Yahoo!, and Bing Apple's Siri and IBM's Watson are also built using computational linguistics technology Automatic translation programs such as Google Translate use it as well 	 The International Linguistics Olympiad (ILO) <u>http://www.ioling.org</u> More than 30 countries participate The US teams have won the most first places, including at ILO 2013 in Manchester, England ILO 2014 will be held in Beijing, China
National Science Foundation WHERE DISCOVERIES BEGIN	North American Chapter





Aymara Fish (by Pat Littell)

Aymara is a South American language spoken by more then 2 million people in the area around Lake Titicaca, which, at 12,507 feet above sea level, is the highest navigable lake in the world. Among the speakers of Aymara are the Uros, a fishing people who live on artificial islands, woven from reeds, that float on the surface of Lake Titicaca.

Below, seven fishermen describe their catch. Who caught what?

Watch out! One of the is lying.



- ____ 1. "Mä hach'a challwawa challwataxa."
- ____ 2. "Kimsa hach'a challwawa challwataxa."
- ____ 3. "Mä challwa mä hach'a challwampiwa challwataxa."
- ____ 4. "Mä hach'a challwa kimsa challwallampiwa challwataxa."
- ____ 5. "Paya challwallawa challwataxa."
- ____ 6. "Mä challwalla paya challwampiwa challwataxa."
- ____ 7. "Kimsa challwa paya challwallampiwa challwataxa."

Your daily catch is pictured to the right. Describe it in Aymara, and don't lie!

(Answers at the NACLO URL below)

http://www.naclo.cs.cmu.edu



Lost in Yerevan (by Dragomir Radev)

On her visit to the country of Armenia, Millie has gotten lost in Yerevan, the nation's capital. She is now at the Metropoliten (subway) station named Shengavit, but her friends are waiting for her at the station named Barekamutyun. Can you help Millie meet up with her friends?



1. Assuming Millie takes a train in the right direction, which will be the first stop after Shengavit? Note that all names of stations listed below appear on the map.

- a. Gortsaranayin
- b. Zoravar Andranik
- c. Charbakh
- d. Garegin Njdehi Hraparak
- e. none of the above

2. After boarding at Shengavit, how many stops will it take Millie to get to Barekamutyun (don't include Shengavit itself in the number of stops)?

3. What is the name (transcribed into English) of the end station on the short, five-station line that is currently in construction, shown in a different shade on the map?

http://www.naclo.cs.cmu.edu



Automata (by Pat Littell)

Finite-state automata (FSA) are a type of abstract "machine" with many possible uses. One possible use is to guess what language a document (such as a webpage) is in. If we make an automaton that can distinguish between possible English words and impossible ones, and then give it a webpage with a bunch of words that are impossible in English (like "*aioaepa*" or "*ragaiiare*"), we can be pretty sure that the webpage isn't written in English. (Or, at least, isn't *entirely* written in English.)

Here is a finite state automaton that can distinguish between possible and impossible words in Rotokas, a language spoken on the island of Bougainville, off the coast of New Guinea. Rotokas has a very simple system of sounds and allows us to create a very small FSA.



An FSA works like a board game. Choose a word, and place your pencil on the space marked "Start". Going through the letters of the word one at a time, move your pencil along the path marked with that letter. If the word ends and you're at a space marked with a thicker circle, the word succeeds: it's a possible Rotokas word! If the word ends and you're not at a thicker circle, or you're midway through the word and there's no path corresponding to the next letter, the word fails: it's *not* a possible Rotokas word!

Try it out with these possible and impossible words; the automaton should accept all the possible words and reject the impossible ones.

Possible Rotokas words		Impossible Ro	Impossible Rotokas words	
tauo	kareveiepa	grio	ouag	
puraveva	ovokirovuia	ovgi	vonoka	
avaopa	ouragaveva	gataap	оарра	

Now, using the automaton above, put a check mark next to each possible Rotokas word:

iu	uente	VOAV
idau	urioo	uaia
oire	raorao	oratreopaveiepa

http://www.naclo.cs.cmu.edu