

Looking Across Languages: Mass Translation of the World's News

Kalev H. Leetaru

Founder, GDEL Project (<http://gdeltproject.org/>)

Senior Fellow, George Washington University, USA

ABSTRACT

Imagine a world where language was no longer a barrier to information access, where anyone can access real-time information from anywhere in the world in any language, seamlessly translated into their native tongue and where their voice is equally accessible to speakers of all the world's languages. Authors from Douglas Adams to Ethan Zuckerman have long articulated such visions of a post-lingual society in which mass translation eliminates barriers to information access and communication. Yet, even as technologies like the web have broken down geographic barriers and increasingly made it possible to access information from anywhere in the world, linguistic barriers mean most of those voices remain steadfastly inaccessible. What would it look like if one simply translated the entirety of the world's news coverage each day in real-time using massive machine translation? For the past two years the GDEL Project (<http://gdeltproject.org/>) has been monitoring global news media, identifying the people, locations, counts, themes, emotions, narratives, events and patterns driving global society. Working closely with governments, media organizations, think tanks, academics, NGO's, and ordinary citizens, GDEL has been steadily building a high resolution catalog of the world's local media, much of which is in a language other than English. Enabling GDEL to look across this material holistically required construction of a massive machine translation infrastructure capable of real-time translation of the world's daily journalistic output in 65 languages, including Afrikaans, Albanian, Arabic (MSA and many common dialects), Armenian, Azerbaijani, Bengali, Bosnian, Bulgarian, Catalan, Chinese (Simplified), Chinese (Traditional), Croatian, Czech, Danish, Dutch, Estonian, Finnish, French, Galician, Georgian, German, Greek, Gujarati, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Kannada, Kazakh, Korean, Latvian, Lithuanian, Macedonian, Malay, Malayalam, Marathi, Mongolian, Nepali, Norwegian (Bokmal), Norwegian (Nynorsk), Persian, Polish, Portuguese (Brazilian), Portuguese (European), Punjabi, Romanian, Russian, Serbian, Sinhalese, Slovak, Slovenian, Somali, Spanish, Swahili, Swedish, Tamil, Telugu, Thai, Turkish, Ukrainian, Urdu, and Vietnamese. This talk will explore what it looks like to use massive computing power to look across the world's information, cultures, and languages, to try and better understand the driving forces of global society and what it means to be human.

ABOUT THE SPEAKER



Kalev H. Leetaru - One of Foreign Policy Magazine's Top 100 Global Thinkers of 2013, Kalev is a Senior Fellow at the George Washington University Center for Cyber & Homeland Security and a member of its Counterterrorism and Intelligence Task Force. From 2013-2014 he was the Yahoo! Fellow in Residence of International Values, Communications Technology & the Global Internet at Georgetown University's Edmund A. Walsh School of Foreign Service, where he was also an Adjunct Assistant Professor. His work has been profiled in Nature, the New York Times, The Economist,

BBC, Discovery Channel and the presses of more than 100 nations, while he has been an invited speaker throughout the globe, from the United Nations to the Library of Congress, Harvard to Stanford, Sydney to Singapore. In 2011 The Economist selected his Culturomics 2.0 study as one of just five science discoveries deemed the most significant developments of 2011, while the following year HPCWire awarded him the Editor's Choice Award for Edge HPC (High Performance Computing) "representing the highest level of honor and recognition given to the thought leaders in the HPC community" and in 2013 noted "his research helped usher in the era of petascale humanities." He has recently concluded projects with NBC Universal's SyFy channel, United States Institute of Peace, World Bank and United States Army, exploring how "big data" can reshape the way we understand the world around us. Kalev founded and leads the GDELT Project, supported by Google Ideas, which monitors a considerable cross-section of the world's broadcast, print, and web news media from nearly every country each day in over 100 languages and identifies the people, organizations, locations, themes, and events driving global society. His work focuses on how innovative applications of the world's largest datasets, computing platforms, algorithms and mind-sets can reimagine the way humanity understands and interacts with the global world. More on his latest projects can be found on his website at <http://www.kalevleetaru.com/> or <http://blog.gdelproject.org>