



Jaime Carbonell

Carnegie Mellon University

### Massively Multilingual Language Technologies

The so-called "digital divide" addressed the challenge of poorer people not having access to the internet and all the myriad resources that come from being digitally connected. With the advent and spread of affordable smart phones, the present challenge is the "linguistic divide" where speakers of rare languages, often in poorer nations or regions, lack access to the vast content of information contained only in major languages. We -- CMU and InterAct -- have long sought solutions, via machine translation and via methods that extend to rare languages with minimal on-line data, well beyond the major languages with economic clout that the major search engines address. The presentation outlines some of the challenges and methods to bridge the linguistic divide.

Dr. Jaime Carbonell is the *Director of the Language Technologies Institute* and Allen Newell Professor of *Computer Science at Carnegie Mellon University*. He received SB degrees in Physics and Mathematics from *MIT*, and MS and PhD degrees in *Computer Science from Yale University*. His current research includes machine learning, data/text mining, machine translation, rare-language analysis and computational proteomics. He invented *Proactive Machine Learning*, including its underlying decision-theoretic framework. He is also known for the *Maximal Marginal Relevance* principle in information retrieval, for derivational analogy in problem solving and for example-based machine translation and for machine learning in structural biology, and in protein interaction networks. Overall, he has published some 350 papers and books and supervised some 60 PhD dissertations.