

TITLE: Mathematics in the era of Big Data

SPEAKER: Marek Rychlik

ABSTRACT: Mathematics is gravitating towards more complex and more "applied" problems, often coupled with computation. In this talk I will describe my own research and projects extending beyond mathematics. All are related to the overarching theme of "Big Data".

One such project involves coding theory, and error correcting codes (ECC). It combines finite field (or Galois field) theory of algebra to build cloud storage. I will describe an invention which resulted from an RTG project 3 years ago that has led to the formation of Xoralgo, a company which will build software and hardware to do just that. I will tell the story of how an RTG project became a patent and a start-up company in the hot tech industry.

I will also describe the current project in Optical Character Recognition (OCR), for converting scanned images of documents in Traditional Chinese and Pashto (language of Afghanistan). Chinese documents span a thousand years since the printing press was invented in China, and use over 60,000 characters. Pashto is a language based on the Arabic writing system, which uses cursive even for print, and it has a hundred characters. But since the characters are connected, this creates about 5,000 ligatures. OCR is a challenging problem in machine learning and neural networks. Recently the National Endowment for Humanities funded this project. A variety of graduate research opportunities in machine learning, statistics and pure math are present. Another opportunity involves natural language processing (NLP), which is typically used to increase accuracy of OCR and extract meaning from documents.