Creating and Augmenting Domain Ontologies with Machine Learning: An Industry Perspective

Rayid Ghani

Accenture Technology Labs, 161 N Clark St, Chicago, IL 60601, USA
`rayid.ghani@accenture.com`

Abstract. Most companies have internal databases that consist of both structured and unstructured content. This environment is similar to the Web setting where the "semantic" Web contains structured, machine-readable content, and the rest of the Web contains unstructured content. In this talk, I will give examples where the structured content can be augmented (both in terms of structure and content) by applying machine learning techniques to the unstructured content for a variety of business problems. I will focus on describing techniques using supervised and semi-supervised learning algorithms to efficiently create and populate ontologies for a variety of product categories.