



DRUID
society

Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

Patent-to-Patent Similarity: A Vector Space Model

Kenneth Younge

EPFL

CDM

kenneth.younge@epfl.ch

Jeffrey Kuhn

University of California

Haas School of Business

jeffrey_kuhn@haas.berkeley.edu

Abstract

Current measures of patent similarity rely on the manual classification of patents into taxonomies. In this project, we leverage information retrieval theory and Big Data methods to develop a machine-automated measure of patent-to-patent similarity. We validate the measure and demonstrate that it significantly improves upon existing patent classification systems. Moreover, we illustrate how a pairwise similarity comparison of any and every two patents in the USPTO patent space can open new avenues of research in economics, management, and public policy. We make the data available for future scholarship through the Patent Research Foundation.