

A Bayesian Classification System for Facilitating an Institutional Risk Profile Definition

Authors : Roman Graf, Sergiu Gordea, Heather M. Ryan

Abstract : This paper presents an approach for easy creation and classification of institutional risk profiles supporting endangerment analysis of file formats. The main contribution of this work is the employment of data mining techniques to support set up of the most important risk factors. Subsequently, risk profiles employ risk factors classifier and associated configurations to support digital preservation experts with a semi-automatic estimation of endangerment group for file format risk profiles. Our goal is to make use of an expert knowledge base, accquired through a digital preservation survey in order to detect preservation risks for a particular institution. Another contribution is support for visualisation of risk factors for a required dimension for analysis. Using the naive Bayes method, the decision support system recommends to an expert the matching risk profile group for the previously selected institutional risk profile. The proposed methods improve the visibility of risk factor values and the quality of a digital preservation process. The presented approach is designed to facilitate decision making for the preservation of digital content in libraries and archives using domain expert knowledge and values of file format risk profiles. To facilitate decision-making, the aggregated information about the risk factors is presented as a multidimensional vector. The goal is to visualise particular dimensions of this vector for analysis by an expert and to define its profile group. The sample risk profile calculation and the visualisation of some risk factor dimensions is presented in the evaluation section.

Keywords : linked open data, information integration, digital libraries, data mining

Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020