Transitioning Existing Content: inferring organisation-specific documents

Arijit Sengupta, Sandeep Purao

Abstract

A definition for a document type within an organization represents an organizational norm about the way the organizational actors represent products and supporting evidence of organizational processes. Generating a good organization-specific document structure is understanding among the organizational actors about how certain business processes should be performed. Current tools that generate document type definitions focus on the underlying technology, emphasizing tags created in a single instance document. The tools, between organizational actors about how a given document type should be represented. We propose a method for inferring organization-specific document structures using multiple instance documents as inputs. The method consists of heuristics that combine individ compiled using standard algorithms. We propose a number of heuristics utilizing artificial intelligence and natural language processing techniques. As the research progresses, the heuristics will be tested on a suite of test cases representing multiple instance documen methodology will be implemented as a research prototype

Full Text: PDF