Graph Matching Algorithms for Business Process Model Similarity Search

As organizations advance in Business Process Management (BPM) maturity, they typically accumulate repositories containing hundreds of business process models, such as the reference model proposed by SAP and the Dutch local authority reference model, each of which includes over 600 models, as well as larger repositories such as the IT Infrastructure Library (ITIL), which each contain over a thousand models, underlining the need for a tool that can efficiently retrieve relevant models for tasks such as documentation, communication of internal procedures, and business process redesign and automation.

The aim is to find models in the repository that closely resemble a given search model or its fragment. Similarity search serves several purposes, such as detecting duplicates when adding models, or retrieving reference models that overlap with existing models. It involves evaluating similarity between dimensions such as text, structure and behavior, based on metrics that compare labels, graph topology and runtime semantics.

It aims to abstract away from specific modeling notations, such as event-driven process chains (EPC), UML activity diagrams and business process modeling notation (BPMN), in order to facilitate the measurement of similarity between business processes modeled using different notations. To this end, it adopts an abstract perspective, defining a process model as a directed attributed graph, enabling a more general approach to business process comparison.