

1 Introdução

O desafio de organizar áreas do conhecimento remonta à criação de taxonomias. Estudiosos construíram em resposta a este desafio, sistemas únicos de classificação que abrangiam todas as áreas do conhecimento, na tentativa de possibilitar uma organização geral para os dados. Com o surgimento da Web o volume de dados disponíveis aumentou várias ordens de magnitude, tornando praticamente impossível a organização de todos estes dados manualmente. Uma forma automática de organizar e classificar esta informação tornou-se imperativa.

Em resposta à necessidade de classificação de grandes quantidades de informação apresenta-se neste trabalho uma nova abordagem na organização de dados semi-estruturados, representados por frames, sem uma estrutura de classes prévia.

Esta dissertação está organizada da seguinte forma. O capítulo 2 discute alguns dos métodos mais utilizados para classificação, introduzindo pontos positivos e negativos de cada abordagem. O capítulo 3 apresenta os principais algoritmos e técnicas utilizadas no processo de agrupamento. O capítulo 4 descreve um algoritmo, baseado no K-Medóide, capaz de organizar um conjunto de frames em classes, estruturadas sob forma de uma hierarquia estrita. Este capítulo discute também o uso de um critério de proximidade que leva em conta os atributos e valores que cada frame possui na organização dos dados e um critério de proximidade que leva em conta o nível hierárquico para a inserção de novos elementos. Por fim, o capítulo 5 contém as conclusões e as direções para trabalhos futuros.