

Abstract. The problem of characterizing and detecting recurrent sequence patterns such as substrings or motifs and related associations or rules is pursued ubiquitously in order to compress data, unveil structure, infer succinct descriptions, extract and classify features, etc. The discovery, particularly on a massive scale, of significant patterns and correlations thereof poses interesting methodological and algorithmic problems, in so far as it often exposes scenarios in which tables and descriptors grow faster and bigger than the phenomena they are meant to encapsulate. While part of this problem is endemic, part of it can be attributed to the traditional approaches to the definition of a pattern, that hinge alternatively on syntactic or statistical properties alone. It has been seen recently that at least this part of the problem may be mitigated by more prudent paradigms, in which the syntactic description of a pattern and the list of all of its occurrences are tightly intertwined. This approach leads to identify regions of monotonicity for some scores of surprise in use, within which it is enough to consider and weigh only extremal terms and values. This talk reviews concepts, constructs, and application results obtained along this line of research.