Abstract of Thesis presented to UFF as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

Exploring Direct Counting for Frequent Itemset Mining

Adriana Bechara Prado

January/2005

Advisor: Alexandre Plastino

Department: Computer Science

During the last ten years, many algorithms have been proposed to mine frequent itemsets. In order to fairly evaluate their behavior, the *IEEE/ICDM Workshop on Frequent Itemset Mining Implementations (FIMI'03)* has been recently organized. According to its analysis, kDCl++ is a state-of-the-art algorithm. However, it can be observed from the *FIMI'03* experiments that its efficient behavior does not occur for low minimum supports on sparse databases. Aiming at improving kDCl++ and making it even more competitive, we present the kDCl-3 algorithm. This proposal directly accesses candidates not only in the two initial iterations but specially in the third one, which represents, in general, the highest computational cost of kDCl++ for low minimum supports. Results have shown that kDCl-3 outperforms kDCl++ in the conducted experiments. When compared to other important algorithms, kDCl-3 enlarged the number of times kDCl++ presented the best behavior.