MTMDD

Multi-Terminal MDD (MTMDD) is an extension of MDD (Multi-valued Decision Diagrams), which has more than two terminal nodes, i.e., it may have a set of multi-terminal nodes. In this work we use an MTMDD structure to store a set of words, i.e., a dictionary, where the terminal nodes represent the word classes associated to different words. Such dictionary representation using a MTMDD becomes the fast, efficient and scalable dictionary retriever.

WAGGER

The Word clas taGGER using MTMDD-based dictionaries is a command line application that can be integrated to any third party software. This software was implemented using C++ language and versions compiled to Linux and Mac OS X are available at http://www.inf.pucrs.br/afonso.sales/wagger where the reader may find examples and files for WAGGER.

Primary Results

This table shows the memory used to store the corresponding MTMDD structure, the time to create it from a set of words (each dictionary or set of dictionaries), and to load it in memory considering the WAGGER software running over a machine with Intel Core i7 quad-core, which runs at 2.2 GHz under Mac OS X 10.6.8, with 4 GB of main memory. For all dictionaries the time to seek the word class of a word is negligible (it takes less than 3 seconds to classify 2 million words.

Dictionary Definition

The dictionary format is composed by four sections:

- The maximum length of words, an integer (the maximum depth of the MTMDD structure);
- The alphabet, i.e., the valid letters to the dictionary (the maximum output cardinality of MTMDD structure nodes);
- The word classes for this dictionary (the number of terminal nodes in the MTMDD structure);
- The list of words of the dictionary with its classes.