Exe files Compression

Or Laharty & Netanel Shimoni
Prof. Dana Shapira

1.Contribution/project goal

Design a compressor specially suited for executable files.

2. Introduction

The algorithms we have used for comparison are:

LZ77 - The algorithm is based on the division of the string encoded into sub-strings called paragraphs in a process known as punctuation. Each paragraph is fitted to a string above a final AB and a dictionary is constructed in a dynamic process. The algorithm is universal, the compression is optimal asymptotic and no prior knowledge of the compressed content is require.

PPMD- PPMD base on PPM - is an adaptive statistical data compression technique based on context modeling and prediction. PPM models use a set of previous symbols in the uncompressed symbol stream to predict the next symbol in the stream. PPM algorithms can also be used to cluster data into predicted groupings in cluster analysis.

BWT- The Burrs-Wheeler transformation is done by lexicographically sorting all the rotations of the input string. When these appear in the table row by row, sort the rows, and return the last column in the table, and the row number corresponding to the original string

3. Methods/algorithms/Alternatives or Design Considerations

The main algorithm on our project is B2:

The idea is based on adaptive arithmetic coding. The weights of the characters are initialized to 1, each character being encoded increases its weight by u.
The difference is that one to k characters divide the weights of all the characters by 2 (top value).

A. The challenge is to find the best k.
how ? Examining different values and measuring them in order to know how much entropy (equivalent to the size of the compressed file in arithmetic coding), until we achieved an optimal result.

B. Depending on k you should choose u as large as possible - to keep the archaic so that we lose less as soon as it is divided and rounded up.

4. Selected Approach

Our approach is basically a journey of testing and research for which algorithm will best compress our EXE files. Meanwhile B2Compress has given the best results. We are trying to develop another algorithm that uses B2 to further optimize the results.

5. Solution Description (Algorithms, Modulation, Patterns, Infrastructure, UI, Functionality)

result of compression all algorithms.

יחס - דחוס/מקורי (B2Compress))	יחס - דחוס/מקורי (PPMd))	יחס - דחוס/מקורי (C- LZ77)	יחס - דחוס/מקורי (Java - LZ77)	גודל הקובץ הדחוס- B2Compress (KB)	גודל הקובץ הדחוס- PPMd (KB)	גודל הקובץ הדחוס- LZZ7 (KB)-C code	גודל הקובץ הדחוס- LZZ7 (KB)-Java Code	גודל קובץ מקורי (KB))	שם הקובץ
0.570469799	0.268456376	0.563758389	0.926174497	85	40	84	138	149	alice29.txt
0.601626016	0.292682927	0.593495935	0.93495935	74	36	73	115	123	Asyoulik.txt
0.412524851	0.115308151	0.252485089	0.321073559	415	116	254	323	1006	Kennedy.xls
0.57793765	0.237410072	0.551558753	0.925659472	241	99	230	386	417	lcet10.txt
0.566878981	0.280254777	0.615711253	0.955414013	267	132	290	450	471	plrabn12.txt
0.137450199	0.097609562	0.284860558	0.298804781	69	49	143	150	502	ptt5