Jacob Murphy
COSI 175
2013-3-20
Lossless Text Compression

My text compression program (compress.c) and decompress program (decompress.c) use a dynamic dictionary approach to lossless text compression. They have two dictionary updating methods:

- “FC" (First character): Add the previous match followed by the first character of the current match.
- "CM" (Current match): Add the previous match followed by the current match.

and three dictionary deletion methods:

- "FREEZE": Stops adding new words when the dictionary is full.
- "RESTART": Restart learning when the dictionary is full. Resets the dictionary to the base alphabet
- "LRU": Delete the least recently used entry that will allow the prefix property to be preserved

Below is a comparison of the file sizes of given compressed files using a combination of the above update and deletion methods. There is currently no data using LRU deletion due to that feature currently not having a working implementation.

File Name	**Original Size**	**FC/FREEZE**	**CM/FREEZE**	**FC/RESTART**	**CM/RESTART**	**FC/LRU**	**CM/LRU**
bib.txt	111,261 bytes	*53,704 bytes*	77,392 bytes	53,704 bytes	77,392 bytes		
book1.txt	768,773 bytes	*325,148 bytes*	488,744 bytes	358,130 bytes	518,096 bytes		
book2	610,856 bytes	*255,264 bytes*	405,724 bytes	270,722 bytes	415,178 bytes		
geo	102,400 bytes	*85,678 bytes*	94,722 bytes	85,678 bytes	94,722 bytes		
news	377,109 bytes	*186,892 bytes*	270,338 bytes	200,250 bytes	278,892 bytes		
obj2	246,814 bytes	*136,494 bytes*	179,808 bytes	139,306 bytes	179,808 bytes		
paper2	82,199 bytes	*42,674 bytes*	60,082 bytes	42,674 bytes	60,082 bytes		
pic	513,216 bytes	*72,618 bytes*	143,690 bytes	72,618 bytes	143,690 bytes		
progc.txt	39,611 bytes	*23,962 bytes*	31,238 bytes	23,962 bytes	31,238 bytes		
progl.txt	71,646 bytes	*33,110 bytes*	49,906 bytes	33,110 bytes	49,906 bytes		

Taking the best results from the above table, you can compare my results to the compression of other compression utilities.

File Name	**Original Size**	**Compress (.tar.gz)**	**Gzip (.gz)**	**Bzip2 (.bz2)**	**My Best**
bib.txt	111,261 bytes	40,960 bytes	35,069 bytes	27,467 bytes	*53,704 bytes*
book1.txt	768,773 bytes	317,440 bytes	313,387 bytes	232,608 bytes	*325,148 bytes*
book2	610,856 bytes	215,040 bytes	206,694 bytes	157,434 bytes	*255,264 bytes*
geo	102,400 bytes	71,680 bytes	68,483 bytes	56,885 bytes	*85,678 bytes*
news	377,109 bytes	153,600 bytes	144,854 bytes	118,540 bytes	*186,892 bytes*
obj2	246,814 bytes	81,920 bytes	81,537 bytes	76,342 bytes	*136,494 bytes*
paper2	82,199 bytes	30,720 bytes	29,752 bytes	25,041 bytes	*42,674 bytes*
pic	513,216 bytes	61,440 bytes	56,442 bytes	49,759 bytes	*72,618 bytes*
progc.txt	39,611 bytes	20,480 bytes	13,281 bytes	12,544 bytes	*23,962 bytes*
progl.txt	71,646 bytes	20,480 bytes	16,278 bytes	15,579 bytes	*33,110 bytes*