# Dictionary compression format AG85DCF1

by Arthur Golubev 19850316

2025-08-15

AG85DCF1 defines the following entities:

1. Dictionaries – in custom way provide dictionary words so that there exists sequence numbering of the dictionary words through the dictionary;

2. Dictionary extracts – of the following format:

2.1. Dictionary extract part 1 – in either system or program/application/task defined format identification of used for defining the dictionary extract dictionaries so that there exists sequence order of the dictionaries;

2.2. Dictionary extract part 2 – considering Sequence 1 is sequence of numbers of all dictionary words of all designated by dictionary extract part 1 dictionaries with sequencing dictionary words through all dictionaries in orders of as the dictionaries are provided in dictionary extract part 1 and of as dictionary words are sequenced in the dictionaries a data structure ag85dst2 with presences representing disjoint ranges in the Sequence 1 by alternating begins and ends of the ranges and considering Sequence 2 is sequence of values in the ranges in order of Sequence 1 a data structure ag85dsp1 with present states representing numbers of being including in the dictionary extract dictionary words in Sequence 2 and not present states representing numbers of not being included in the dictionary extract dictionary words in Sequence 2 so that the ranges are such that both there are no numbers of being included in the dictionary extract dictionary words out of the ranges and the both data structures are as little information as possible to archive changing the ranges.

3. Archives – of the following format:

3.1. Archive part 1 – in either system or program/application/task defined format identification what data is compressed into the archive;

3.2. Archive part 2 – in either system or program/application/task defined format identification of used to make the archive dictionary extracts so that there exists sequence order of the dictionary extracts;

3.3. is empty if no grammar compression, Archive part 3 – in either system or program/application/task format identifiers of grammar programs and devices required for performing grammar decompression in order as processing by grammar programs and grammar devices is required;

3.4. exists if archive part 3 is not empty, Archive part 4 – pieces of information for grammar decompression in order as processing by grammar programs and grammar devices is required.

3.5. Archive part 5 – considering Sequence 3 is sequence of numbers of all dictionary words provided marked as used by all designated by archive part 2 dictionary extracts with sequencing dictionary words through all dictionary extracts in orders of as the dictionary extracts are provided in archive part 2 and of as dictionary words are sequenced in the dictionary extracts a

data structure ag85dst2 with presences representing disjoint ranges in the Sequence 3 by alternating begins and ends of the ranges and considering Sequence 4 is sequence of values in the ranges in order of Sequence 3 a data structure ag85dsp1 with present states representing numbers of not used by the archive dictionary words in Sequence 4 and not present states representing numbers of used by the archive dictionary words in Sequence 4 so that the ranges are such that both there are no numbers of not used by the archive dictionary words out of the ranges and the both data structures are as little information as possible to archive changing the ranges.

3.6. Archive part 6 – for every used dictionary word a data structure ag85dst2 data structure with presences representing one of the following:

3.6.1. if archive part 7 is not empty, position numbers of starting bytes of occurrences of the dictionary word in sequence of bytes of data compressed into the archive;

3.6.2. if archive part 7 is empty, position numbers of occurrences of the dictionary word in sequence of dictionary words corresponding to pieces of data compressed into the archive without restrictions of how to convert pieces of data into words in variate cases.

3.7. is empty if there were no such data pieces in compressed into the archive data, Archive part 7 – in the same order as in the data compressed into the archive pieces of data which are out pieces of data corresponding to dictionary words values in the data compressed into the archive.