```
< dict size : uint32>
< word : char[] > \0 < doc frequency : uint32 >
                                                                    Corpus
                                                                  Level Data
           (Repeated dict size times)
                                                                                 Legend
                                                                                 uint32 : 32-bit unsigned int
                                                                                 char[] : character array
< num docs : uint32>
                                                                                          null byte
                                                                                 float32: 32-bit floating point
< doc id : char[] > \setminus 0 < num features : <math>uint32 >
      < feature id : uint32 > < feature weight : float32 >
                                                                   Document
                                                                   Level Data
                   (Repeated num features times)
                  (Repeated num docs times)
```