 Figure 8. Corpus data processing flow