

### Appendix 3 Knowledge Matrix morphology

Amongst Chinese textbooks and supplementary exercises published over the years, there has been a lack of systematic and standardized ways to gauge the levels of difficulty of input text in Chinese textbooks and a lack of criteria and indicators to compare and review Chinese curriculums among Hong Kong and other areas. We aim at developing a system that is capable of conducting intelligent linguistic analysis and difficulty level diagnosis so Chinese text content supporting smart learning can be analyzed. Advanced technologies from Natural Language Processing, and now with ChatGPT, will be hosted at a platform, and is able to identify level of difficulty, identify linguistic and grammatical items, identify theme or genre for Primary 1 to Secondary 6. It can also analyze exercises and conduct matching between any Chinese text and exercises according to difficulty levels. The target users of the whole system are students, tutors, teachers, parents, authors, and publishers. In addition, the development of a knowledge base is also critical to student profiling.

This project pertains to natural language processing (NLP) which is an area under artificial intelligence (AI). It will integrate theories and methodologies from linguistics with modern computational models from Natural Language Processing (NLP) and computational linguistics. Effective computational models integrated with grammar-rules for Chinese language in text segmentation, part of speech (POS) tagging, syntactic parsing, and named entity recognition (NER) will be developed to analyze and understand Chinese texts commonly found in Chinese language textbooks. Also, text theme identification models will be designed to identify Chinese text themes or genres and level of difficulty.