

Enhancing English Teaching Research with Natural Language Processing Technologies

LingConc: A Tool for Language Learning with Top-down Access

Jason s. Chang

Professor of Computational Linguistics
Department of Computer Science
National Tsing Hua University

Abstract

Corpus linguistic techniques and encourage a more bottom-up rather than top-down processing of text. More specifically most concordance tools support only keyword search (not general linguistic query) and ordered search results not taking linguistic properties (phraseology and collocations) into consideration. In this talk, I will first introduce *LingConc*, a corpus analysis system for user-provided text dataset, supporting linguistic queries with part-of-speech wildcard, displaying search results top-down according to key-phrase frequency. Additionally, *LingConc* also extracts keywords, lexical bundles and linking Wikipedia information. The system accepts user submitted pdf files or URLs, and generates various text analyses including the word list, keyword list, collocation list and lexical bundles. Users can then view this top-level summary of the dataset, and click on a word or phrase to explore relevant documents on the sentence and document levels. By clicking a document name, users can read the content in an interactive environment and retrieve on demand relevant information (e.g., word definitions, collocations, grammar patterns and Wikipedia information). Additionally, the system also includes the functions of *Linggle* (linggle.com) to retrieves common phrases, collocation, or lexical bundles using intuitive and expressive linguistic queries with wildcards and wild parts of speech (PoS). The search results are displayed in a top-down fashion (phrases first and then sentences) according to phrase frequency.

Keywords

corpus linguistics, data-driven learning, concordancer, top-down processing, phraseology analysis