**文學中的文本分析、數據挖掘及可視化**

本文摘自MLA Commons的〈[Text Analysis, Data Mining, and Visualization in Literary Scholarship](https://dlsanthology.mla.hcommons.org/text-analysis-data-mining-and-visualizations-in-literary-scholarship/)〉，由邱永泰統整後撰文。

在過去，人們可能普遍認為人文科學研究與數位工具沾不上邊。然而，這樣的觀點正逐漸改變中。尤其是資料探勘、視覺化、協作技術等等，正在改變人文學者進行研究的方式。

數位工具對人文科學研究主要帶來的影響是，透過工具分析，能幫我們找出文本背後隱藏的特徵或特定模式。運用工具協助分析文本的方式被稱作「遠讀」（Distant Reading）；與之相對的，傳統閱讀文本研究的的方式被稱作「細讀」（Close Reading）。所謂的「差異閱讀」（Differential Reading），便包含了遠讀法以及近讀法這兩種研究的形式，而不論是細讀或是遠讀的運用皆可以幫助我們從主觀和客觀的層面上來理解、闡釋文本的內容。有些學者嘗試對文本進行一些假設，並從文本結構的層面進行分析，例如分析某些字詞的出現頻率或位置等等，雖然說分析出來的東西不一定有意義，有時反而會得出令人誤解的結果，但失敗固然有其意義，在不斷嘗試，不斷錯誤的過程中，有時反而能有意外的發現，映證一些抽象層面上對文本的猜想。

另外，運用工具來促進差異閱讀的實踐，能幫助我們闡釋在文本中看到的特定模式，也可以促進自我反思或自我覺察的實踐。舉例來說，作者以[*Vocabulary Management Profiles*](http://web.missouri.edu/~youmansc/vmp/)(*VMP*)工具，對*《美國人的形成》(The Making of Americans, Gertrude Stein, 1925)*一書進行分析研究，並以視覺化呈現。就結果而言，該工具能基於文本的文本模式或形式的變化，標記出文中特定片段。這項成果即證明了使用工具分析的價值，甚至透過工具，還發現了以細讀法分析未曾發現的事實。我們不妨試問自己，為何以細讀的方式從文本中看到了一些模式，卻沒發現到其他的模式呢？

另一個值得提到的點是「協作技術」。所謂的「協作技術」對於資料探勘項目中的細讀與遠讀來說非常有幫助。一般而言，使用資料探勘或視覺化這類新時代的技術做研究就代表說，不是所有的學者都熟悉這些工具。因此，資料探勘項目特別需要協作技術以及眾人的努力，這也代表著，無論是人力、金錢或是機器的運算能力，這些資源的缺乏也將為研究帶來許多阻礙，尤其是對於那些獨立的研究學者而言。

雖說人文學者做不到的事，電腦也不一定能做得到。電腦能做到的事情，有些人類也做得到，但是，最大的差別在於，電腦的速度要快的多。尤其電腦可以幫我們審視大量的文本，以人類的閱讀速度來說，這點是作不到的。此外，雖然說電腦分析出來的東西不一定為真、不一定總是對的，但換個方式想，只要把它視為一個參考，提供我們一個新的視角看待文本就好。這就是，以數位方式進行人文科學研究總是能令人感到驚喜連連的原因之一。