京都大学情報学研究科﹕NTT通信科學實驗室聯合研究部

本文摘自[京都大学情報学研究科網頁](https://nlp.ist.i.kyoto-u.ac.jp/kuntt/index.php)，由彭靖雯統整後撰文。

這個網站是由京都大學的信息學研究科和日本最大的通信公司NTT裡的通信科學研究所的聯合研究所。在剛合作的五年間主要在做自然語言處理的研究，現在則是以面向國際的語言資源為研究目標。順帶一提的是，這個網站剛開始只有京都大學的信息學研究科和NTT裡的通信科學研究合作的研究中心，現在則是以開放聯合的方式來進行合作研究，除了有各個大學的教授外，也會跟其他地區的研究人員合作。而這個網站特別的還有一點就是它是由PukiWiki，也就是日語的維基網站廣泛使用的軟體所製作，而隨著 PukiWiki的更新，這個網站也會不斷地更改版本並且詳細的標示出出處讓使用者參考。

這個聯合研究所在2004年創辦，之後的五年間也陸陸續續發表了基礎語言、輸入法等相關論文。也積極地參加了不少國內外的學術研討會、自然語言處理協會發表了很多與日文語言處理研究相關的探討內容。而聯合研究所的資金來源大部分是由日本科學促進會青年科學家補助金，以及文部科學省科學研究特定領域研究信息IT基礎設施所提供。

近年則以能夠支持世界各國語言的語言系統做研究，接下來就讓我繼續介紹這個聯合研究所正在研究的主題吧。第一個要介紹的是MeCab，是目前廣泛使用在日本的詞法分析系統，是一款不需依靠具體的語言或語言庫仍可以進行參數估計的分詞系統。裡面也有可以讓使用者透過MeCab將文檔翻譯成中文，不過要注意的是它翻譯的效果目前還沒有很完善，有時候可能會翻譯出不理想的內容喔。

接下來是建立語言資源數據庫跟撰寫基礎單詞的部分，為了建立國際語言系統這個目標，研究所大量的分析了語言字詞的意思，像是單字的意義和表面意思的不同，或是針對政治、經濟、運動類型的專有名詞製作語言辭典等。還有研究透過部落客的文章來進行語料庫的構造分析，主要是針對京都觀光、手機、運動、美食這4個人們較常關注的話題。

最後要提到的主題是研究所透過RTE來進行的聲譽分析，那麼RTE是什麼呢？RTE是一個即時管理系統，在應用程式跟電腦系統之間做連結，可以計算時間、將工作內容分好前後順序、還可以快速地收集資料，因此可以透過這個即時管理系統來分析各式各樣的教材或文本內容。