



# 網路內容未來發展趨勢

谷圳

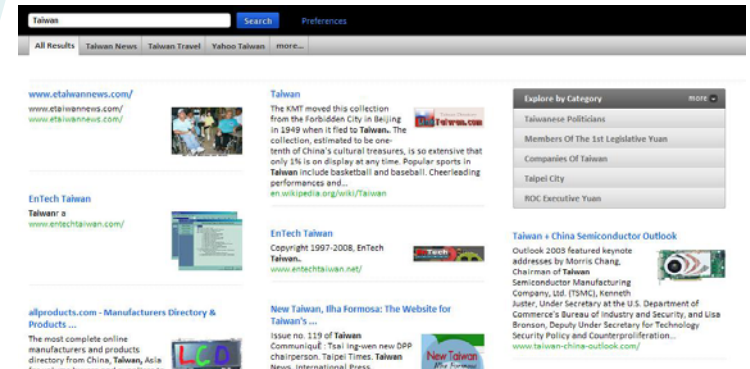
資策會 創新應用服務研究所

97.09.24



# 發展趨勢: Concept Search Engine

cuil



## Keyword Search Engine

- Keyword, Tag, Metadata
- User Generate Content
- Multi-Language

## Portal

- Taxonomic
- Published Content
- Single Language

## Concept Search Engine

- Concept Space/Map, Personalization
- Content Auto-Generation
- Cross-Language

資料整理: 創研所



# Concept Space/Map- 協同式語意空間建構技術

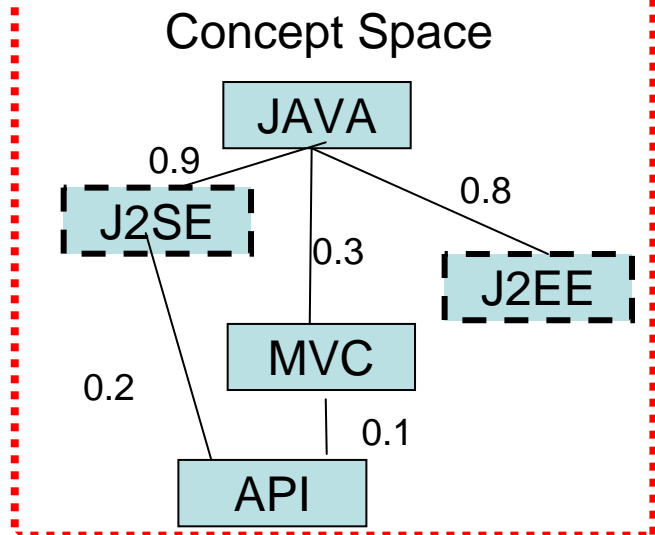
### Tag Search

資料維護 > 著錄教材管理 > 個人著錄 > 檢索教材 > 瀏覽教材 > 多維度瀏覽 > 瀏覽多媒體庫 > 討論

查詢學習元件

搜尋條件: java | 全站資源 | 送出

找"JAVA"



關聯式搜尋, 大幅增加搜尋範圍

Result:

JAVA

J2SE

J2EE



# Personalization- 概念詞彙輔助搜尋技術

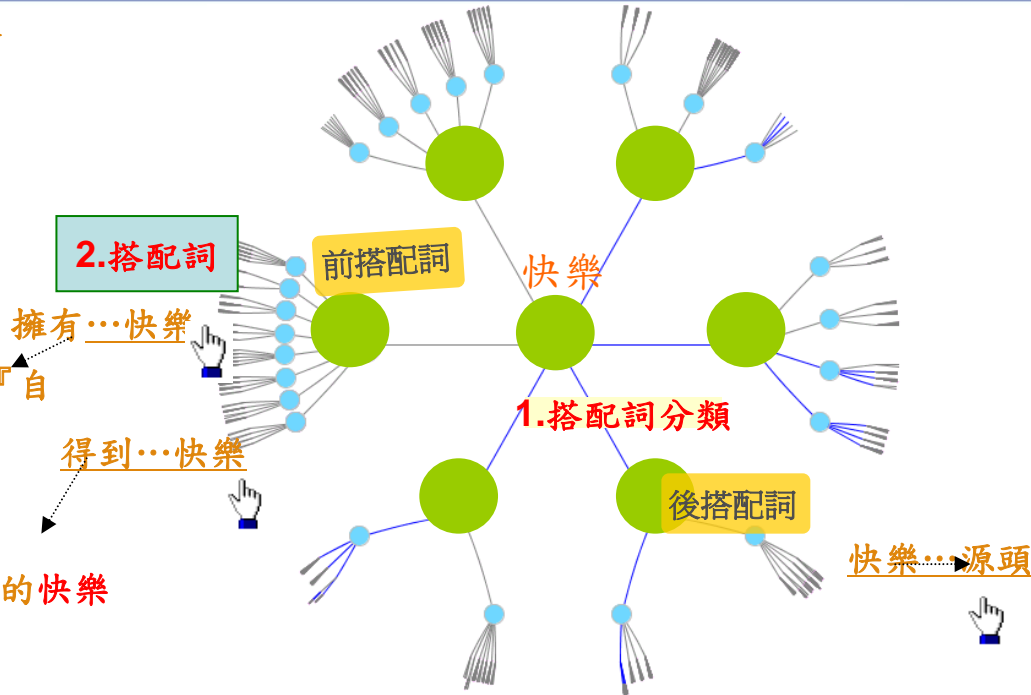
輸入想要查詢的概念

快樂

### 3. 搭配詞例句

例句：我依然能夠體驗『自己擁有的快樂』

例句：我得到從未有過的快樂

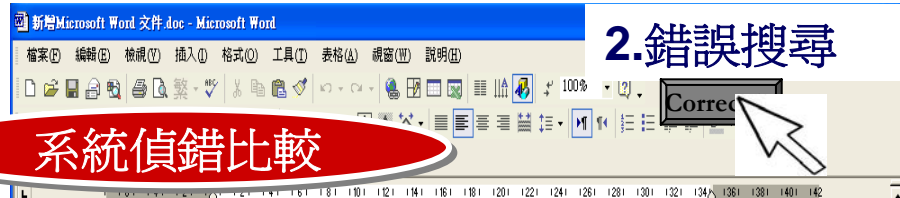


例句：一旦你了解自己快樂的源頭，就可以學習創造快樂的經驗

4. 同義詞	高興\歡喜\歡愉\開心\暢快\樂
5. 成語辭典	樂不可支\樂不可言\樂不思蜀 大快人心\大喜過望\心花怒放\助人為快樂之本
6. Wikipedia	快樂是一種感受良好時的情緒反應，而且常見的成因包括感到健康、安全、愛情和性快感等等。快樂最常見的表達方式就是笑。
7. 網路資源	快樂在哪裡, Cheers雜誌快樂工作人



# Personalization- 華文錯別字搜尋技術



更正建議：**和藹可親**  
說明：  
「和藹」是形容人和氣、

Type of False Alarm	原句	以往系統 (自然輸入法 correction rate 64%) (微軟word correction rate 35%)	本計畫
專有名詞辨識 (譯名、人名、地名...)	吳城億獲選為優秀青年	無法辨別專有名詞,直接校正成同音的短字詞 <b>無誠意</b> 獲選為優秀青年	可辨別專有名詞 <b>吳城億</b> 獲選為優秀青年
同音字	也因此嚐片各種酸甜苦辣	只找到“苦臘”短字詞的錯誤 <b>也因此嚐扁各種酸甜苦辣</b> 只考慮於兩個字的短字同音詞	將長字詞,前後文也納入考慮 <b>也因此[嚐遍]各種[酸甜甜辣]</b> 將近形長字詞納入考慮,更正正確
近形字	令他忍不住耀耀欲試	令他忍不住 <b>耀耀浴室</b> False Alert過多(容易變成另一詞、甚至另一句話)	令他忍不住 <b>[躍躍欲試]</b> 考慮前後文,避免False Alert
Word Sense (考量前後文)	想用這句話來代表我的中心思想	<b>享用</b> 這句話來代表我的中心思想	想用這句話來代表我的中心思想

近形字	1	和藹>和藹
Word Sense	0	N/A



# Personalization-

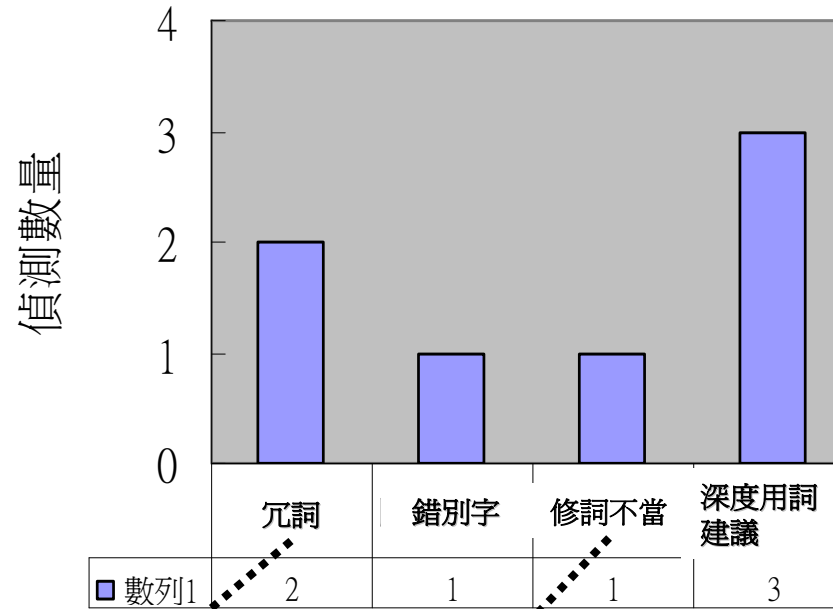
## 華文作文語法分析與診斷技術



華文寫作  
輔助工具



作文語法分析與診斷



冗詞

•我原先本來想要告訴她事實，可是我沒有勇氣。  
 →我本來想要告訴她事實，可是我沒有勇氣。

修辭不當

我爸爸很美麗  
 →我爸爸很英俊



# Personalization-

## 個人領域能力模塑與評量技術

來自國內各大部落格，累計約1059萬篇文章

Del.icio.us  
190 萬筆書籤

HEMIDEMI  
17 萬筆書籤



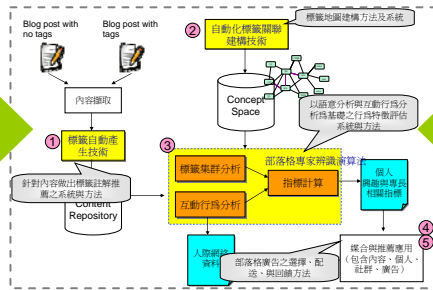
Keyword Extraction - Result

Keyword Extractor Page

Keyword Extraction

標簽自動產生技術

### 部落格專家辨識技術



Tagsphere

Automated concept weight

自動化標簽關聯建構技術

合作與應用



## 領域內容自動建置技術

### 傳統辭典建置

#### 維護更新階段-新詞

維護更新內容困難度高,需定期召集專家修訂不同版本,以符合時代

#### 多語系轉換階段

多語系發展困難,需要不同國家的專家協助,溝通協調複雜

#### 編寫階段解釋與例句

人力與時間成本高,需找多位專家花費時間編撰

#### 核心字彙建置

參考資料來源有限,參考資料限制在一開始的編撰階段

人力投入:勞力密集

大幅降低人力投入

多語百科辭典自動建置方法

#### 字典編審與修訂

- 專家審核與修訂

#### 新增辭典功能

- 搭配詞、多語系、新詞偵測、例句分類、影片推薦

#### 電腦輔助編輯

#### 維護更新階段-新詞

#### 多語系轉換階段

#### 編寫階段解釋與例句

#### 核心字彙建置

- 透過偵測網路資訊,持續更新詞彙
- 網路資源如Wikipedia,新聞...等高可信度網站自動擷取解釋,與例句
- 利用wordnet與網路資訊建立跨語系核心字彙