

# 北京大學計算語言研究所

## 詩詞語言語意分類體系之建置與應用

報告人：台灣元智大學·羅鳳珠  
湖北大學客座教授

E-mail : [gefjulo@saturn.yzu.edu.tw](mailto:gefjulo@saturn.yzu.edu.tw)

網路展書讀網站：<http://cls.hs.yzu.edu.tw>

2007年9月4日

# 大綱

人與電腦

詩詞與電腦

電腦的侷限

韻文的語言特色

資訊檢索 (Information Retrieval) 的發展歷程

詩詞文體語言的特性所衍生的

資訊檢索需求

傳統文學網站資訊檢索功能及缺點

語意標註及語意概念分類建置方法：詞彙切分

語意標註及語意概念分類建置方法：分類之思考過程

語意標註及語意概念分類建置方法：語意標記與分類

詩詞語言語意標記對全文檢索的重要性

語意概念分類體系在全文檢索的應用

詩詞語言語意概念分類體系在教學的應用

詩詞語言語意概念分類體系在研究上的應用

詩詞語意標記及分類的困難與未來的發展

# 人與電腦

人：

- 能分辨字形的符碼
- 能解讀字詞含義的意碼

電腦：

- 只能分辨字形的符碼
- 電腦理性客觀，處理有規則而合邏輯的資料。

所以引用電腦做為偏重主觀抽象感知的詩詞文學研究輔助工具，比客觀知識的自然科學困難得多。

# 詩詞與電腦

詩詞：

感性主觀，講究：

只可意會而不可言傳

不著一字而盡得風流

言有盡而意無窮

電腦：

電腦理性客觀

處理有規則而合邏輯的資料。

# 電腦的侷限

- 電腦處理有規則及合邏輯的資料，其效率高於人腦何止千萬倍，但是電腦只能分辨字形的符碼，無法解讀字詞含義的意碼，所以引用電腦做為偏重主觀抽象感知的詩詞文學研究輔助工具，比客觀知識的自然科學困難得多。
- 電腦縱然有強大的記憶與搜尋能力，因為只能以字形分辨，無法辨別語意內涵，不具有思考能力，不具有人的認知、感知能力，不具備人的知識體系，無法滿足使用者的需求。

# 韻文的語言特色

- 袁行霈教授在《中國詩歌藝術研究·原序》論及：「如果從語言學的角度給詩歌下一個定義，不妨說詩歌是語言的變形，他離開了口語和一般的文書語言，成為一種特異的語言形式。」
- 無論從語意或語言形式的角度，詩歌非但與一般口語、文書語言不同，與其他文體也存在著差異，再加上詩歌的形式受到詩歌格律之平仄聲調及字數、句法上的限制，語言更精鍊，語意更豐富，修辭更講究，語境更複雜，需要更多聯想，解讀更困難。
- 相對於「自然語言」而言，詩詞語言是文學語言，是「人工語言」（葉聖陶：詩詞是特殊的語言形式）

# 資訊檢索 (Information Retrieval) 的發展歷程

科學使得人際間的交流非常的快速，科學也提供了思想的記錄，使人類能夠運用這些思想，知識就不再限於個人，而是屬於人類的所有生命。（Vannevar Bush）

- 記錄思想的文獻儲存在圖書館，圖書館的資料按照字母或數字排列的索引，只能一層一層往下尋找，不符合人類大腦以思維聯繫尋找資料的方式，因此，Vannevar Bush便構思著建立一個以聯繫（association）代替索引（indexing）以選擇資料的機械化個人文件夾和圖書館（mechanized private file and library）（Vannevar Bush）
- 資訊無論以任何物質／非物質載體、任何媒體形式呈現，從資訊產生者到資訊接收者之間，都會有認知上的落差，提取方法的不同，而產生給予、接受與取得的障礙。

# 詩詞文體語言的特性所衍生的 資訊檢索需求1/2

- 詩詞是特殊的語言形式，以不同的修辭技巧達到精緻凝練、語意無窮，情境高遠、**言意與形神兼具**的境界。
- 透過修辭的技巧，將作者主觀內化的情思意念，投注在客觀外顯的事物上，作為作者與讀者之間引發共鳴的溝通媒介。**詩詞所使用的詞彙，有一些是客觀知識的描述闡釋**（蘇軾〈少年遊〉：「去年相送，餘杭門外，飛雪似楊花。今年春盡，楊花似雪，猶不見還家。」），**有一些是主觀情意的引導感發**（蘇軾〈水龍吟〉：「春色三分，二分塵土，一分流水。細看來、不是楊花，點點是離人淚。」）。

# 詩詞文體語言的特性所衍生的 資訊檢索需求2/2

- 主觀情意的感發，往往因為讀者的學養、經驗、年齡、心理氣質、思想境界等各種背景差異而使感受各有不同，因此相同的詞彙，相同的作品，感受上存在著個別的差異，即便是同一個人，經過不同時空，重讀相同的作品，感受也會不同，這是客觀認知與主觀感知的問題。
- 文學表達的情感通常是主觀而抽象的，即便是藉助客觀的物象作為作者與讀者之間溝通的媒介，仍然存在著感知層次的不同，因此，資訊檢索的技術，應用在文學上，其困難度比自然科學領域高得多，應用在特殊語言形式的詩詞，困難度又高於小說、戲劇、散文等文體。

# 傳統文學網站資訊檢索功能及缺點1/2

- 傳統的文學網站檢索都只能比對字形，不能分辨詞形、詞義。文學講究修辭變化，講究文采意境，一詞多義，一義多詞的現象比比皆是。只能比對字形的檢索，檢索結果的完整性及正確都不足，不能滿足一般的使用需求，對於文學領域以研究及教學為目的的使用者而言，尤其不足。

以「月」為例：

## 一詞多義：

「月」常用的基本含義有四種，其一是月亮，其二是曆法單位，其三是季節代稱，其四是圓形如月亮的物品，這是一詞多義。

- 一詞多義會使檢索時出現字形、詞形相符但語意不符的資料，影響正確性

## 一義多詞

「月」若作「月亮」解時，同義詞有「月亮＝玉兔＝銀蟾＝玉杯＝玉鉤＝玉盤＝玉輪＝玉蟾＝生魄＝冰輪＝兔降＝孤光＝明霽＝金波＝金輪＝桂華＝桂魄＝素娥＝掩夜＝清光＝寒暉＝嬋娟＝瓊瑤＝蟾影……」，這是一義多詞。

- 一義多詞會遺漏很多語意相同但字形、詞形不同的資料，影響完整性。



# 語意標註及語意概念分類建置方法： 詞彙切分1/3

- 一、建立人名、地名……等各種專有名詞詞彙資料庫（婆餅焦時須早去，婆餅焦是飛禽）
- 二、建立詩詞典故詞彙資料庫
- 三、建立詩譜、詞譜資料庫，分析其句法

# 語意標註及語意概念分類建置方法： 詞彙切分2/3

## 四、詞彙切分方法及步驟：

1. 第一優先比對專有名詞資料庫，以專有名詞做為詞彙切分之最優先順序。
2. 第二優先之切分順序是領字：依據所建立的宋詞詞譜優先將領字切分，其餘詞句依「宋詞詞譜」之步驟切分，例如：「**漸**／霜風／淒緊／，關河／冷落／，殘照／當樓／。」（只有詞需要此步驟）
3. 第三優先比對典故詞彙資料庫

詩詞之句首，除了使用專有名詞、典故或領字之外，絕大多數都是雙字詞。

# 語意標註及語意概念分類建置方法：

## 詞彙切分3/3

### 4. 其餘句子之切分步驟

- 句子字數為偶數句之句子，兩兩切分為雙字詞，例如：「離恨／恰如／春草／，更行／更遠／還生／」。
- 句子字數為奇數句之句子，末三字之外的句子，兩兩切分為雙字詞，例如：「菡萏／香消／翠葉殘／，西風／愁起／綠波間／，還與／韶光／共憔悴／，不堪看／。」。
- 奇數句句末三字，其句法或作上二下一，或作上一下二，並無準則，切分的方法需仰賴詞庫以提高切分之正確率。切分時以前述步驟所獲得之雙字詞比對句末之三字詞，若這三個字之前二字或後二字是前述步驟所獲得之雙字詞，則優先取為雙字詞，予以切分。

不過也有末三字之前二字或後二字都是詞彙的句子，此時便需以人工校正，例如：「歸心正似**三春草**。」（蘇軾〈虞美人〉）之「三春」、「春草」都是宋詞常用的詞彙，遇到這種情況，便需以人工校正。

# 語意標註及語意概念分類建置方法： 分類之思考過程

- 一、嘗試引用《同義詞詞林》之分類，但是格格不入。
- 二、以蘇軾詩詞為範圍，以人工進行詞彙切分，建立詞彙切分之規則（如前所述）
- 三、以蘇軾詩詞所切分之詞彙，以人工進行語意標註，並作分類。
- 四、分類的考慮以符合詩詞語言特性、保留詩詞語意的內涵、滿足文學研究需要為主，其次考慮分類學的意義。
- 五、盡可能兼顧詞類的類別。
- 六、詞類的類別參考《現代漢語語法信息詞典》
- 七、語意類別參考：唐·徐堅《初學記》、明·張之象《古詩類苑》、當代·馬東田主編《唐詩分類大辭典》。

# 語意標註及語意概念分類建置方法：

## 語意標記與分類

目前總共50大類，279小類，元素詞共10,807種（不含重複）。

- 全台詩總共23,066首詩，詩作共2,617,823字，切分為700,848個詞彙，去除重複後為152,670個詞彙，目前已標上95,361個詞彙（不含重複）。
- 文學作品不外乎描寫：「人」、「事」、「時」、「地」、「物」，故以此五項作為大類，另加「其他」，總計六大類。

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（人1-6/14）

1. **人名**：本名、別名字號、他稱、法號、合稱、外國人名、虛構人名、待考
2. **朝代君王**：君王帝號、君王陵墓、君王年號
3. **神鬼仙人**：神鬼專稱、神鬼泛稱、菩薩尊號
4. **稱謂**：自稱謙稱、尊稱美稱、貶稱惡稱、親人眷屬、指代稱謂、一般稱謂、職業身份、職官爵位、皇室稱謂、宗教稱謂
5. **身體器官**：五官顏面、鬚鬢髮膚、身體四肢、五臟器官、分泌物、合稱
6. **感官詞**：嗅覺、味覺、觸覺、聽覺、視覺

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（人7-9/14）

### 7. 情感詞（綜合情感）：

- (1) 正面情感（喜悅欣樂、自我肯定、景仰羨慕、悠然自適、憐惜愛寵）
- (2) 負面情感（自我否定、惱怒怨妒、悲哀傷痛、感懷惆悵、憂愁掛慮、厭煩憎惡、懊惱憾恨、害怕恐懼、辛酸艱苦、慚愧自責、驕矜自恃、思鄉念舊、尷尬為難）

### 8. 情感詞（兩性情感）：正面情感、負面情感

9. 人際互動：宗族、鄰里、姻戚、交游、僚友、耆老、懷思、尋訪、參謁、招致、會遇、話舊、送別、留別、贈別、話別、酬別、寄別、宴別

# 語意標註及語意概念分類建置方法： 語意標記與分類（人11-14/14）

10. 品德操行：正面（聖、賢、忠、孝、友悌、恭敬、聰敏）、負面
11. 抽象意識：心神
12. 人為：帝王出巡、治國之道
13. 動作：行為動作、言語動作
14. 狀態：生命狀態、生活狀態、思維活動、合稱

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（事3）

1. **禮俗**：郊丘、社稷、宗廟、祭祀、釋奠、封禪、明堂、籍田、巡幸、扈從、朝會、禮見、召對、省直、誕辰、拜表、生日、滿月、晬日、宴饗、酺宴、婚姻、拜掃、送葬、輓歌
2. **典故**：佛典、神話典故、其他
3. **法令條文**：刑法、條文

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（時2）

1. **四時節氣**：節氣、節慶、四季
2. **時間**：明確時間、範圍時間（今昔、日、月、世代、生、年、清晨、白晝、黃昏、夜晚、時刻、其他）、年齡

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（地1-7/7）

1. **地名**：行政地名、山嶺地名、河湖地名、島嶼地名、關津地名、虛構地名、外國地名、地名他稱（行政地名、山嶺地名、河湖地名、島嶼地名、關津地名）、地名合稱
2. **地理**：區域範圍、邦國都城、郊原村野、街道巷弄、特殊場域
3. **自然景觀**：山峰崖嶺、水澤湖泊、沙洲島嶼、洞穴坑谷、林原莽漠、山川泛稱、關津渡口
4. **閬苑仙境**：閬苑、仙境
5. **幽冥之境**：幽冥
6. **人世之境**：人間
7. **空間**：方向、位置、距離

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（物1-6/14）

1. **生物**：神物
2. **生物**：動物專名（走獸、飛禽、昆蟲爬蟲、水生兩棲、合稱）、動物通稱（單稱、合稱）
3. **生物**：植物專名（木本、草本、禾本、苔菌、瓜果、合稱）、植物通稱（花、木、葉、枝、果、草、莖、根、苗、合稱）
4. **生物**：生物合稱
5. **生物狀態**：動物生命狀態、植物生命狀態
6. **生物動作**：動物動作

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（物7-8/14）

### 7. 器物：

- 生活用品（飲食器具、烹飪器具、裝置容器、起居用品、其他用品、梳妝用品）
- 交通工具（水路、陸路、空路、合稱）
- 工具用品（兵器單稱、兵器合稱、戰爭物品、農牧漁獵工、生活工具單稱、生活工具合稱、刑具）
- 絲帛服飾（衣冠腰帶、布料、鞋襪、配飾、合稱）
- 宗教用品、喪葬祭祀、娛樂器具、金帛錢幣、公堂器物、旗幟、帳幕、神器、樂器、器物合稱

### 8. 建築物：建築專名、宮室屋廬、亭臺樓閣、園林院落、寺廟道觀、碑塔陵墓、公署建築、建築部件、軍事攻防

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（物9-14/14）

9. **飲食**：飲品、食品、食材、糕點、調味品、菸草、宴席、合稱
10. **自然資源**：玉石、火、金屬、礦物、燃料、合稱
11. **文藝**：書籍泛稱、書籍專名、文具、書篇名、文體、引文、樂曲泛稱、樂曲專名、
12. **技藝**：書畫通稱、書畫專名、書法、字帖專名、棋、舞蹈、戲劇
13. **天候氣象**：風霜雪露、雲霧煙霞、雷電電霓、冷寒、熱旱、晴天、陰雨、合稱
14. **天文**：日、月、星、天空、銀河、合稱

# 語意標註及語意概念分類建置方法：

## 語意標記與分類（其他1-8）

1. **行政體系**：部門名稱、朝代名號、朝代名號合稱、民族邦國名、民族邦國名合稱、國名
2. **顏色**：本色、借色、間色、合稱
3. **災厄**：旱災、水災、地震、兵災、其他災害
4. **疾病**：疾病療法、疾病類別
5. **狀聲詞**：自然界聲音、動物聲音、人類聲音、器物聲音
6. **單位詞**：文藝、長度、重量、面積、容量、量詞
7. **數量詞**：數字
8. **專業術語**：佛家語、作戰語、法術語、道家語

# 詩詞語言語意標記對全文檢索的重要性1/10

- 詞彙的語意標記，可以幫助讀者解讀詞彙、體會詩作，與詩人的情思意念共鳴；語意標記資料庫的建立，也可以使電腦從只能分辨字形提升到可以理解字詞義，進而達到提升全文檢索質量的目標。
- 詩是特殊的語言，特別講究聲律及形式之美，受到字數及平仄押韻等格律的限制，詩的用語與語體文不同。這種特殊文體的特殊用語，若以傳統比對字形的方式開發檢索系統，檢索的質量勢必無法滿足使用者的需求。
- 詩句的語意不一定在單一詞彙顯示，或說從單一詞彙解讀，不見得能夠完整理解整首詩的含義。整首詩的含義，有時需要從共現的詞彙、從作者及寫作的背景綜合判斷，非常複雜，也非常困難。

# 詩詞語言語意標記對全文檢索的重要性2/10

以「酒」為例

寫酒的詩，不一定使用「酒」字，舉凡：瓊漿、玉液、濁醪、花釀、新醅……等詞彙指的都是酒。

李白「舉杯邀明月，對影成三人」、「飛蓬各自遠，且盡手中杯」，雖然沒有寫到酒字，杯中盛的都是酒。

白居易〈效陶潛體詩〉詩：「清光入杯杓，白露生衣巾。乃知陰與晴，**安可無此君**。」的「此君」是酒。然而宋之問〈綠竹引〉詩：「含情傲睨慰心目，**何可一日無此君**。」的「此君」是竹子。

王翰〈涼州詞，二首之一〉詩：「**醉臥沙場君莫笑**，古來征戰幾人回。」  
使之醉臥沙場的是葡萄美酒。

李白〈和盧侍御通塘曲〉詩：「**偶逢佳境心已醉**，忽有一鳥從天來。」  
使他醉的，是佳境不是酒。

## 詩詞語言語意標記對全文檢索的重要性3/10

詩人寫詩，會創造新的詞彙，也會在已有的詞彙上賦予新的詞義，二者皆有可能為後代詩人沿用。

詩詞用語可以區分為：

- 一、一般用語
- 二、特殊用語：典故、意象、比喻……
- 三、情境用語

# 詩詞語言語意標記對全文檢索的重要性4/10

特殊用語：典故、意象、比喻

因某位詩人所使用的意象，廣為後世詩人沿用而成為典故的詞彙，都有固定的詞義，在語意標記時，可以標記其典故詞義。

例如「南浦」，最早出現於《楚辭》卷第二〈九歌·河伯〉：「與女遊兮河之渚，流澌紛兮將來下。子交手兮東行，送美人兮南浦。」其後江淹〈別賦〉：「春草碧色，春水淥波。送君南浦，傷如之何！」這二處的「南浦」都附著了離別感傷色彩。

唐代王維寫下〈齊州送祖二〉：「送君南浦淚如絲，君向東州使我悲。為報故人憔悴盡，如今不似洛陽時。」之後，詩人寫水邊送行的感傷，都使用「南浦」，從此之後，「南浦」已經牢牢的與離別感傷結合在一起。

# 詩詞語言語意標記對全文檢索的重要性5/10

特殊用語：深層義、雙關義

深層義：杜甫〈江南逢李龜年〉：「岐王宅裡尋常見，崔九堂前幾度聞。正是江南好風景，落花時節又逢君。」

雙關義：韓翃〈章台柳〉

章台柳，章台柳，昔日青青今在否？縱使長條似舊垂，也應攀折他人手。

柳姬〈楊柳枝〉

楊柳枝，芳菲節，苦恨年年贈離別。一葉隨風忽報秋，縱使君來豈堪折。

電腦無法理解**語義**，更無法理解**雙關義、情韻義**

# 詩詞語言語意標記對全文檢索的重要性6/10

## 情境用語

- 詩人們以某些詞彙共現在同一首詩中，以達到傳遞心境，烘托情境的效果，所使用的詞彙，或所選擇共現的詞彙，與所要烘托的情境，必然具有物象本身的特徵及某些人際間共有的認知脈絡可以依循。這個共同的認知脈絡，引發讀者的認知、感知與聯想，成為作者與讀者之間共鳴的媒介，詩人因心境的感知，以聯想創造文學藝術，讀者以聯想、感知體會詩人的創意與心境、情境。

# 詩詞語言語意標記對全文檢索的重要性7/10

情境用語：以「苔蘚」為例

- 苔蘚植物是由水生生活趨向陸生生活的植物族群，生長於極端潮濕的向陽地區或是陰暗的環境中，以配合其親水性結構，不然便很快枯死。這些地方往往也是人跡罕至之處，這是我們對於「苔蘚」的認知基礎。
- 詩詞中寫到苔蘚時，往往是「江邊道路多苔蘚，塵土無由得上衣。」的失意文人，或是「玉階陰深殿，苔蘚蒼蒼闕虛院。」的失寵佳人，或是「天香寂寞，深羅綺點成苔蘚斑。」的塵外僧人，或是「肌膚銷盡雪霜色，羅綺點成苔蘚斑。」的空閨女子，或是「花木閒門苔蘚生，澆川特去得吟情。」的山林雅士。
- 「苔蘚」若是長於門前或台階，詩人將之與「門」、「階」並列呈現時，更核心的集中在描寫「因失寵而幽居空閨女子」的淒清冷落。鄭谷〈長門怨，二首之二〉：「流水君恩共不回，杏花爭忍掃成堆。殘春未必多煙雨，淚滴閒階長綠苔。」以不再使用而長滿綠苔的蒼涼景象。

## 詩詞語言的語意分類對全文檢索的重要性8/10

- 使用者還具備由各種單一詞彙的知識所建構的知識體系，這是只能分辨字形的電腦所無法具備的。詞彙經過語意標記，雖然已經能夠使電腦初步具有分辨語意的能力，但也只是零星片段沒有組織的語意，不符合人的知識體系。無法符合人類以「思維聯繫」搜尋資料的行為模式。
- 詞彙經過語意標記之後，需要建立符合知識結構及人的知識體系的知識架構，建立知識連結的軌跡，才能夠使全文檢索從「索引式」提升為「思維聯繫式」的檢索。

## 詩詞語言的語意分類對全文檢索的重要性9/10

- 梅家駒主編，上海辭書出版社出版的《**同義詞詞林**》最具代表性。本書第一版發表於1983年，第二版發表於1996年，收錄詞彙54,235筆，將詞義分為12大類，95中類，1,428小類，3,925詞群。每一次類所包含的詞彙以及該詞彙的同近義詞依序列出，成為研究詞彙語意學及語意概念分類的重要工具書。

# 詩詞語言的語意分類對全文檢索的重要性10/10

《同義詞詞林》從語言學的學理做專家分類，詩詞韻文語意分類，最優先的考慮是文學文本內容的修辭特性與文體語言的特性。

文學的內容以人為主軸，人、事、時、地、物五個要素是小說、戲曲內容必備的元素，詩詞因為受到字數的限制，在有限的文字裡時常省略一些用語，所以詩詞內容的文字，雖然不一定直接寫及這五項元素，但整體而言，所描述的內容，仍然包含了這些元素，例如：柳宗元的〈江雪〉：「千山鳥飛絕，萬徑人蹤滅。孤舟蓑笠翁，獨釣寒江雪。」詩中有人（孤舟蓑笠翁）、有事（獨釣）、有時（雪，冬天）、有地（寒江）、有物（孤舟）。

# 語意概念分類體系在全文檢索的應用

## 全臺詩全文檢索設計

The screenshot shows a web browser window with the following elements:

- Browser Address Bar:** http://localhost/TWP/b/b02.htm
- Navigation Bar:** Includes links for '全臺詩全文索引區', '全臺詩檢索區', '臺灣詩社資料庫', and '時空資訊系統'.
- Search Interface:**
  - Section Header:** 全文檢索
  - Form Fields:** 作者或別名, 詩題, 詩序, 詩句.
  - Dropdowns:** 大類: 天文, 小類: 星, 元素: 請選擇.
  - Search Criteria:** 標誌項目, 作品出處, 寫作時間, 相關地名.
- Navigation and Links:** 作者索引, 作者全文檢索, 出處索引, 時間資訊系統, 空間資訊系統.
- Decorative Elements:** A vertical banner on the left reads '全臺詩 智慧型全臺詩知識庫'. A red vertical box contains '全臺詩檢索區' and '全臺詩全文檢索區'.

# 語意概念分類體系在全文檢索的應用 優點

- 一、使電腦從只能分辨字形提升到理解字詞義
- 二、提升全文檢索的完整性（使用者的知識程度不同）
- 三、情感詞的分析，使電腦具備辨識及理解情感的能力，往人工智慧邁前一大步。
- 四、建構完整的知識結構體系，符合人類以聯繫（association）代替索引（indexing）的資料搜尋模式。

# 詩詞語言語意概念分類體系在教學的應用1/4

## 自動作詩填詞系統

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

上一頁 搜尋 我的最愛

網址 http://cls.hs.yzu.edu.tw/CSP/W\_stave/index.htm

Google 开始 已拦截 15 个 拼写检查 翻译 发送至 设置

自作新詞韻最嬌 淺斟低唱

曲譜區

回首頁

倚聲填詞 曲中依約 斷人腸

此曲有意 無人傳 曲譜檔

倚聲填詞

倚聲填詞系統 · 關於本系統

- 詞牌格律
- 詞韻韻表
- 文字韻目
- 訪客作品
- 相關論文

【倚聲填詞】格律自動檢測索引教學系統

國科會研究計畫

計畫主持人：羅鳳珠·朱四明

系統設計人：羅鳳珠 曹偉政·李元萍

程式設計人：曹偉政·李元萍

網際網路

# 詩詞語言語意概念分類體系在教學的應用2/4

## 提供的功能

- 詞牌格律檢索顯示
- 以韻目查詢韻字，以韻字查詢所屬韻目
- 格律自動檢測並提示錯誤
- 跨資料庫查詢全宋詞

舉例：采桑子

- 最難消受簷前雨朝也難收暮也難休雁唳蛩鳴添客  
愁當年一別音書絕人也難留夢也難求獨擁孤衾憶  
舊遊。

# 詩詞語言語意概念分類體系在教學的應用3/4

能（做得比人好的部分）：

記誦所有詞牌的格律、韻目、韻字並快速檢索、自動批改不合格律之處。

還不能作的：

不能判斷多音字、不能更改修辭、意境。

所以：

電腦能填一闕格律完全正確的詞，但不能填一闕好詞。

電腦所填出的「格律完全正確的詞」，有可能是「詞句完全不通的詞」。

電腦能將用錯的字改成對的，但不能將用得不好的字改成好的。

# 詩詞語言語意概念分類體系在教學上的應用4/4

- 一、語意概念之分類體系，幫助使用者豐富詞彙，提升應用比喻、意象、典故等修辭的能力。
- 二、系統具有修改修辭的功能
- 三、共現詞彙的比對功能，幫助使用者創造詩詞的情境、意境。
- 四、能分辨多音字，並且正確使用：詞彙讀音資料庫提高平仄格律檢測的正確率
- 五、能分辨、理解詞意

# 詩詞語言語意概念分類體系在研究上的應用

## 以情感計算為例1/6

### 詩詞中表達情感的方式

- 一、直接以情感詞彙描寫，如高興、斷腸、消魂
- 二、間接烘托：李白〈長干行〉：「門前遲行跡，一一生綠苔。苔深不能掃，落葉秋風早。」周邦彥〈玉樓春〉：「當時相候赤欄橋，今日獨尋黃葉路。」
- 三、寄託外物：蘇軾〈卜算子〉：「缺月挂疏桐，漏斷人初靜。誰見幽人獨往來，縹緲孤鴻影。驚起卻回頭，有恨無人省，揀盡寒枝不肯棲，寂寞沙洲冷。」李商隱〈暮秋獨遊曲江〉：「荷葉生時春恨生，荷葉枯時秋恨成。深知身在情長在，悵望江頭江水聲。」
- 四、以感官感覺（情以物遷，辭以情發）：柳永〈八聲甘州〉：「對瀟瀟、暮雨灑江天，一番洗清秋。漸霜風淒緊，關河冷落，殘照當樓。是處紅衰翠減，苒苒物華休。」

# 詩詞語言語意概念分類體系在研究上的應用 以情感計算為例2/6

- 李清照留下來的詞作53闕，抽取的詞彙數量共計1,754條（扣除重複，有1,371條），情感詞彙61條，負面情感58條，正面情感1條。
- 正面情感1條是：
- 「更喜」（〈鷓鴣天〉：酒闌更喜團茶苦），這裡喜的是團茶，不屬於心性情感；
- 這項統計，更具體明確的顯示李清照詞的情感表現與情感風格。

# 詩詞語言語意概念分類體系在研究上的應用

## 以情感計算為例3/6

- 以溫庭筠詞分析，溫庭筠留下來的詞作74闕，抽取的詞彙數量共計1,656條（扣除重複，有1,182條），情感詞彙54條，全部都是負面情感，沒有正面情感。
- 這項統計資料，與溫庭筠詞全都是描寫歌樓酒肆之中的歡場女子苦悶的心緒相符合。
- 這些負面情感的詞彙，依使用頻率多寡，依序是：斷腸（10）、愁（9）、惆悵（7）、恨（7）、傷（3）、寂寞（3）、惜（3）、苦（2）、驚（2）、羞（2）、嗚咽（2）、淒（1）、憐（1）、怨（1）。
- 溫庭筠擅長於寫歡場女子的心事，所以非常受到這些女子的喜愛，他筆下這些女子，處於歡場之中，送往迎來，但她們畢竟也是人，也有動真情的時候，只是她們動情的對象幾乎都是逢場作戲的尋芳客，情感的落空是多數人的宿命，因此心境上的黯然、悲傷、惆悵都可以想見，溫庭筠深刻細膩的寫出這些女子最幽微隱密，最不足為外人道的心事，從這些分析統計數據上可以看出詞作的風格與情感的取向。

# 詩詞語言語意概念分類體系在研究上的應用

## 以情感計算為例4/6

- 除此之外，將情感詞彙與其他類別詞彙比較，還可以進一步做「共現詞彙」的研究，可以分析「共現詞彙」所傳遞的語義以及文學情感的表達受到那些外在因素影響的研究。
- 依據《文心雕龍·物色》所說的：「春秋代序，陰陽慘舒，物色之動，心亦搖焉。」、「情以物遷，辭以情發」的論點，「春秋代序」指的是季節，「陰陽慘舒」指的是天色明暗，「物色之動」、「情以物遷」的「物」是客觀的景與物，以此論點檢視李清照與溫庭筠詞的相關類別的詞彙，表列如下：

# 詩詞語言語意概念分類體系在研究上的應用

## 以情感計算為例5/6

|     | 正面<br>情感 | 負面<br>情感 | 季節  | 時間   | 天候                            | 冷寒暖熱                          |
|-----|----------|----------|---|--|-------------------------------|-------------------------------|
| 李清照 | 1 / 61   | 58 / 61  | 春24 / 30<br>秋6 / 30<br>夏0 / 30<br>冬0 / 30 | 夜晚18 / 40<br>黃昏12 / 40<br>清晨7 / 40<br>白晝3 / 40 | 陰雨<br>16 / 19<br>晴天<br>3 / 19 | 冷寒<br>15 / 16<br>暖熱<br>1 / 16 |
| 溫庭筠 | 0 / 54   | 54 / 54  | 春14 / 15<br>秋1 / 15<br>夏0 / 15<br>冬0 / 15 | 夜晚11 / 20<br>黃昏5 / 20<br>清晨4 / 20<br>白晝0 / 20  | 陰雨<br>20 / 21<br>晴天<br>1 / 21 | 冷寒6 / 8<br>暖熱2 / 8            |

# 詩詞語言語意概念分類體系在研究上的應用 以情感計算為例6/6

- 詩詞裡傷春悲秋的传统：春天傷失偶，秋天悲失路
- 由統計表分析，可知溫、李詞的負面情感都屬於失偶之悲，所以多數寫於春天，與詩詞傷春悲秋的传统相符合。這種感傷的情懷，在陰冷的天候，在黯淡的黃昏與漫長的夜晚最容易引發。（辛棄疾：春天147次，秋天49次；正面情感200次、負面情感622次）
- 語意標記與分類，使電腦具備理解及表達情感的能力（辨識那些詞彙是情感詞，那些詞彙是感官詞，不同詞彙共現時烘托出那一種情感）

# 詩詞語意標記及分類的困難與未來的發展1/10

一、詩詞語言的特殊性，包含構詞的特殊、詞義的特殊，篇章結構的特殊，這些特殊性增加了詩詞詞彙切分及語意標記、語意分類的困難，也使得詩詞語意概念的分類，除了必須符合知識認知層次的語言知識分類，還需顧及情感認知層次的意象取向分類。

二、詩詞語意標記的困難，還包括詩詞喜用多義字來創造聯想的空間，以引出無窮的趣味及意在言外的情境。



# 詩詞語意標記及分類的困難與未來的發展3/10

- 從以上所述可知，可以透過語意標記及語意概念分類，讓電腦從分辨字形提升到理解詞義，使電腦的認知能力更接近人的知識體系，以提升全文檢索的品質。
- 人的知識體系分為「對知識的認知」、「對情意的感知」、「對哲理的辨知」，認知是表述的層次，感知需要聯想輔助，哲理詩的解讀體會，卻不是單純的邏輯思維能解決，也就更難以用語意標記處理。
- 全唐詩網站所包含的文本數量龐大，人工校對所需要花費的人力很多，加上虛詞的語意標記及分類的高成本研究，撰寫本文時，還來不及完成人工校對，語意的分類也不完整、不精確，尚待後續的研究修正及補足。

# 詩詞語意標記及分類的困難與未來的發展4/10

## 語意分類的模糊界線

- 人：朝代君王：君王陵墓：孝陵、長陵、安陵、陽陵、茂陵、平陵、黃陵、端陵
- 應該分在「人：朝代君王」或分在「物：建築物：碑塔陵墓」
- 分在「人」的原因是詩詞寫及「君王陵墓」語意常常是人，但有時又指「陵墓建築」

# 詩詞語意標記及分類的困難與未來的發展5/10

## 語意分類的模糊界線

一、「相思」是正面情感或負面情感？

二、「束脩、宵夜」放在「飲食」、「行李」  
放在「器物」是否合適？

# 詩詞語意標記及分類的困難與未來的發展6/10

## 並列詞彙是否要切分

合稱（1. 保留詩詞雙音節詞的特性；2. 合稱產生新的語意；3. 相反詞或相似詞之分合具有某些意義）

- 二喬 { 大喬（三國吳）、小喬（三國吳） } 項劉 { 項籍（秦）【項羽】、劉邦（漢） } 三楚 { 西楚、東楚、南楚 }、伊洛 { 伊水、洛水 } 魚稻、蝦筍
- 肝腸、肝腦、肝膽、眉目、眉睫、唇舌、唇齒
- 悲喜、憂樂、苦樂、喜怒、





# 詩詞語意標記及分類的困難與未來的發展8/10

## 語意標註的側重面

- 白鶴：白鶴在何處
- 白鶴：試看中間白鶴
- 白鶴：白鶴飛來似可
- 有鶴：有鶴止庭隅
- 老鶴：老鶴高飛
- 夜鶴：惱殺晨猿夜鶴
- 夜鶴：曉猿夜鶴驚猜
- 孤鶴：我夢橫江孤鶴去
- 招鶴：且自多招鶴
- 非鶴：非月非雲非鶴
- 秋鶴：春與猿吟秋鶴飛
- 野鶴：因風野鶴飢猶舞
- 野鶴：野鶴溪邊留杖屨
- 野鶴：□欠歸來野鶴
- 黃鶴：聽取雙黃鶴
- 椿鶴：莫祝靈龜椿鶴
- 雙鶴：雙鶴一千年
- 騎鶴：試教騎鶴
- 騎鶴：騎鶴揚州

# 詩詞語意標記及分類的困難與未來的發展9/10

## 語意標註的側重面

聽雨（聽風聽雨小窗眠）：天候氣象→陰雨→  
雨

秋露（秋露下）：天候氣象→風霜雪露→露

秋霄（目斷秋霄落雁）：四時節氣→四季→秋

雁霜（雁霜寒透幕）：生物→動物專名（飛禽）  
→雁

# 詩詞語意標記及分類的困難與未來的發展10/10

## 詩詞詞類標註的必要性

王安石：春風又綠江南岸

柳永〈雪梅香〉：「漁市孤煙裊寒碧，水村殘葉舞愁紅。」

王安中〈卜算子〉：「客舍兩三花，並臉開清曉。一朵涓涓韻已高，一朵纖纖裊。」

王之道〈望海潮〉：「楓葉露痕，荻花風色，人言今日重陽。芳菊裊秋香。」

已有數字化資源及建置網站1994~迄今

網路展書讀：<http://cls.hs.yzu.edu.tw/>

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

網址(D) <http://cls.hs.yzu.edu.tw/> 移至 連結 >>

# 網路展書讀

最新消息

相關評介及報導

- 發展理念
- 成果榮譽
- 成員介紹
- 線上申請
- 意見回饋
- 網路討論
- 版權宣告

—快速連結—

## 網路私塾

[網路教學](#)

[古典文學館](#) (持續建置中)

[卓越計畫網路教育園區](#)

[蘇軾詩標誌系統研究](#)

[閩南語計劃典藏—歷史語言與分布變遷資料庫](#)

[蒸鏡記](#) (陳三五娘)

[唐宋流行歌](#)

[唐詩三百首](#)

[宋詞三百首](#)

[搜文解字](#)

[語文知識網路](#)

[文國尋寶記](#)

[全球華語文數位教與學資源中心](#)

[數位典藏國家型科技計劃](#)

[時空之旅—蘇軾](#)  
(教育部六大學管網計劃)

[淺斟低唱—宋詞古唱虛擬實境](#)

## 古典小說

[紅樓夢 \(English Version\)](#)

[三國演義](#)

[水滸傳](#)

[金瓶梅](#) (金瓶梅全文)

## 數位文物

[賴和數位博物館](#)

[鍾理和數位博物館](#)

[台灣客家文學館](#)

網際網路

# 現有數位資料清單與統計數據1/8

## 注音符號、漢語拼音與國語正音資料庫

- 一、注音符號讀音資料庫：17,261筆，約178,313字
- 二、漢語拼音讀音資料庫：1,346筆，約12,518字
- 三、閩南語讀音資料庫：17,269筆，約82,199
- 四、客語讀音資料庫：17,261筆，約74,740字
- 五、其他地方讀音資料庫：91,600筆
- 六、詩詞語料讀音資料庫：89,400筆

# 現有數位資料清單與統計數據2/8

## 簡、繁體常用字資料庫

- 一、繁簡體常用字對照資料表：3,145筆
- 二、常用字詞讀音資料庫：4973筆，約71,856字

# 現有數位資料清單與統計數據3/8

## 詩詞典故資料庫

一、詩詞典故資料庫：18,585筆，約4,142,621字

二、詩詞典故出處資料庫：9,124筆，約1,163,977字

三、詩詞典故例句資料庫：21,687筆，約593,655字

# 現有數位資料清單與統計數據4/8

## 詩詞故事資料庫

- 一、唐詩三百首資料庫：321首詩，約29,367字
- 二、宋詞三百首資料庫：301闕詞，約411,316字
- 三、全唐詩資料庫：51,170首詩，約4,767,979字
- 四、全宋詩資料庫：35,696首詩，約2,886,831字
- 五、唐宋詞資料庫：24,226闕詞，約3,980,049字
- 六、全台詩資料庫：約152,672字

# 現有數位資料清單與統計數據5/8

## 文學文本等文獻資料庫

- 古典詩詞文、章回小說、當代小說散文、戲曲、史料等（包含文字、聲音、影像、圖片、教學軟體等媒材）：約6,000萬字

# 現有數位資料清單與統計數據6/8

## 工具書資料庫

- 一、詩詞典故資料庫（ 18,585筆）
- 二、中國古典文學人名資料庫（唐宋代、明代、清代、民國）：
  - （一）宋人傳記：24,889筆，約3,890,443字
  - （二）全台詩作者：754筆，約75,288字
  - （三）人名（本名、別名字號）：28,376筆，約2570,548字
- 三、《辭源》通俗地名：610筆，約8,307字
- 四、廣群芳譜：864筆，約13,435字
- 五、歷代帝王譜系：6,527筆，約68,664字

# 現有數位資料清單與統計數據7/8

## 民間歌謠資料庫

台灣民間文學：

- (一) 客家歌謠、故事集：約371,778字
- (二) 閩南語歌謠、故事集：約797,619字

# 現有數位資料清單與統計數據8/8

## 地名資料庫

- 一、歷代地名語料：38,661筆，約1,457,746字
- 二、古典詩詞地名：22,161筆，約269,807字
- 三、《辭源》通俗地名：610筆，約8,307字

感謝聆聽

敬請指教