

# 在 Blog 中運用標籤設計文章註解與自動化回應機制之研究

## Study on Article Annotation and Automatic Reply Mechanism Using Tagging in Blog

陳永昇 葉富銓 蔡昫達  
國立台北教育大學資訊科學所  
yschen@tea.ntue.edu.tw

### 摘要

近年來隨著網際網路技術的進步，許多以數位學習系統為輔助的教學方式也在各級學校機關單位中實行。但功能強大相對也較複雜的 LMS/LCMS(學習管理系統/課程管理系統)反而會造成授課老師使用及維護上的負擔；必竟目前大部份的授課老師對於這些系統都是很陌生的，LMS/LCMS 的使用及維護可能會造成授課老師更大的困擾。本研究利用 Blog(部落格、網誌)這種頗為大眾所接受的內容管理系統平台，來建置 LMS/LCMS；為了減少授課老師維護系統及回覆學生們所回應的問題的負擔，我們還加入了 Agent Process(代理人程序)，可以幫助授課老師自動回覆，而這些回覆內容與回應中的問題，有一定程度的關聯。這種及時的自動回覆機制，可以讓學習者很快的收到回覆，相信也可以大幅提升學習者使用的興趣。

**關鍵詞：**部落格、代理人、擴充標記語言、本體論、標籤。

### Abstract

With the advance of network technologies, many teaching patterns supported by e-LMS were implemented in many schools. Although LMS/LCMS were strong and complicated, those were also confused with teachers. Due to teachers were new to the system, it could become more problems for the teacher to facilitate and maintain the LMS system. In this research, we proposed the blog system, a popular content management system, to implement the CMS. According to reduce the time of replying article and overhead on teachers, we propose an Agent Process mechanism to make automatic reply possible. There were some guaranteed related association between comments and reply content. Under the in-time automatic reply mechanism, the learners could receive the reply comments as soon as possible. We expected that it could increase the learning passion.

**Keywords:** Blog, Agent, XML, Ontology, Tag.

### 1. 前言

由於資訊科技不斷進步以及寬頻網路的發展，數位輔助教學也漸漸被廣泛應用[3, 4, 5]，但是一套數位學習系統往往需要龐大的建置成本。於是許多人討論以 Blog 來做為線上教學的一個新方向，它的好處除了降低成本之外，也可以利用其特性促進網路學習社群的形成。

許多相關研究都提出以 Blog 為基礎的數位學習平台[2, 6, 9, 15]，在這個 Blog 的架構下可以讓數位學習邁向 E-moderation 的五個階段。而這五個階段分別是：1.使用與推動、2.線上社交、3.資訊交換、4.知識建構、5.發展知識。這樣的一個多媒體(multimedia)、超文件、同步(synchronicity)、互動(interactivity)的學習平台，相信可以有效的提升數位輔助教學的效能[2]，甚至可以型成一個被大家所接受且可以廣泛應用在學習上的平台[15]。

註解(annotation)是一種額外的描述，對事物作額外的描述或語意上的加註[10]，有許多的相關研究關於在網頁文件中加入註解[7, 8, 10, 14]；但是註解的撰寫通常需要一些特殊的工具軟體，相當不便。所以本研究利用 Blog 中的 tag 機制作為文章註解，tag 在 Blog 中是非常有用的資訊，藉由 tag 我們可以對 Blog 中的文章作分類，使文章在語意上有關係連繫，進一步促成本研究所提出的自動化回應機制。

#### 1.1 研究背景

##### 1.1.1 Blog

部落格(Blog)[11]，一個以文章發佈時間作為排序基礎的內容管理系統平台，提供all-in-one套件的架設網站機制，讓每個使用者專注於成為一個部落格使用者(Blogger)，而不需多花費心思於網站維護、網域申請或是架站主機維護。換句話說，部落格提供一個易於撰寫文章的網路平台，允許自由的發表，也能夠對於其他的文章給予回應或引用文章，透過回應或是引用的機制，若有似無的把文章分類在一起，達成在分散式環境下的知識管理，也因為部落格彈性的分享機制，可將其他相關的資訊匯集在部落格裡，如書籤、相片、影片。目前主流的 LCMS/CMS ( Learning Content Management System/Content Management System ) 如 XOOBS、

Noodle，功能強大但並不一定適合每個人使用，架設此類網路平台需要一定的系統與架站經驗，有相當的門檻，但是部落格簡單編輯，方便又好用的後台介面，讓部落格在近幾年成長飛快，根據網站blogherald.com的統計，截至今年七月，全世界已有超過七千萬個部落格，族群從小孩到老年人，各種年齡層皆有，加上3G時代的來臨以及手機製造科技的進步，利用手持裝置上網以不再被頻寬所限制，可傳輸較大量的資料。

### 1.1.2 Web2.0

根據Tim O'Reilly所提出的Web2.0概念[12]，Web2.0代表分享，網路上的資訊允許可以被分享，被允許自行做社會化分類，被重複使用，並且彼此之間有著彼此操控性；簡單來說，以是一個以使用者為中心的設計理念。為了要讓使用者方便去使用，所以設定這些網路資源可以分享、可以重複使用、可以被其他網路服務所利用，或是合成一個新的網路服務；部落格也可以說是符合此概念的實體。部落格因為簡單且人性化的使用介面，不僅適合一般的用途也可以做為線上學習的彙整平台，對老師而言，可以輕鬆的從事電子教材的製作、蒐集與彙整、學習者回饋感想或是問題於部落格上，同時也可以藉由種子訂閱的訊息通報，可以即時收到一些消息的更新，減少利用電子郵件的往返所耗掉的時間成本。

本研究提出以Blog為學習平台設計一線上學習的模式與自動化回應機制，減少老師回應學生所花的重複時間成本，轉而可以更專注於教材的豐富性。老師在教學部落格中，發表相關的教學文章或是其他影音資源，或是上課的投影片等等；學生可以針對各種類型的文章，回應感想或是提出疑問。這時本研究提出的代理人(agent)將會針對回應做解析，比對是否有相關聯性的文章或是資源於原有的教學部落格中，並自動回覆於學生的個人部落格，再透過RSS feed使即時的更新通知，不僅可以增加師生間的互動性，也讓學習歷程可以被完整的記錄下來，讓老師可掌握更豐富的學習回饋資料。

### 1.1.3 中文斷詞

詞是最小且有意義，可以自由使用的語言單位。任何語言處理的系統都必須先能分辨文本中的詞才能進行進一步的處理，例如機器翻譯、語言分析、語言了解、資訊抽取。因此中文自動分詞的工作成了語言處理不可或缺的技术。基本上自動分詞多利用詞典中收錄的詞和文本做比對，找出可能包含的詞，由於存在歧義的切分結果，因此多數的中文分詞程式多半討論如何解決分詞歧義的問題，而較少討論如何處理詞典中未收錄的詞出現的問題(新詞如何辨認)。

由於中文詞集是一個開放集合，不存在任何一個詞典或方法可以盡列所有的中文詞。當處理不同

領域的文件時，領域相關的特殊詞彙或專有名詞，常常造成分詞系統因為參考詞彙的不足而產生錯誤的切分。為了解決這個問題，最有效的方法是補充領域詞典加強詞彙的搜集。因此新的詞彙或關鍵詞的自動抽取成為分詞的先期準備步驟。領域關鍵詞彙多出現在該領域的文件中而少出現在其它領域，因此抽取關鍵詞時多利用此特性。高頻的關鍵詞比較容易抽取，少數低頻的新詞不容事先搜集，必須線上辨識。構詞律、詞素、詞彙及詞彙共現訊息，為線上新詞辨識依據。本系統提供了一個解決方案，可以自動抽取新詞建立領域用詞或線上即時分詞功能。為一具有新詞辨識能力並附加詞類標記的選擇性功能之中文斷詞系統。此一系統包含一個約拾萬詞的詞彙庫及附加詞類、詞頻、詞類頻率、雙連詞類頻率等資料。分詞依據為此一詞彙庫及定量詞、重疊詞等構詞規律及線上辨識的新詞，並解決分詞歧義問題。除了基本詞彙庫外，使用者可需要附加領域專屬詞庫。詞類標記為選擇性功能，可附加文本中切分詞的詞類解決詞類歧義並猜測新詞之詞類。分詞系統採用之詞典俱可擴充性，使用者可依據不同領域文件，補充以領域詞典做為分詞之用。[1]

## 1.2 系統目標

在目前的數位學習環境中，學習資源(例如:教材、試題、解答等)相當豐富，因此取得學習資源並不困難，反倒是學習者可能迷失在這個資料量過大的學習資料庫中(例如:學習者在搜尋引擎下輸入某些關鍵字，但是通常會找出成千上萬筆資料；而這些資料之中，哪些是真正學習者所想要得到的資料?)，而找不到自己所要的學習資訊。

本研究提出一個以Blog為基礎的數位學習平台，由授課者來發佈一些學習資訊在blog平台上，學習者可以藉由回應或引用來發問或撰寫心得。而Agent Process會自動判斷學習者的回應(comment)或引用(trackback)，做出自動化回應的回覆，藉由blog中文章的分類(tag)與本研究設計的分類樹(ontology)及比對演算法，可以讓自動化的回應與學習者所回應的問題(comment or trackback)能夠有一定程度的關聯，幫助學習者在這個浩瀚的學習資料庫中找到最適合的學習資訊，也大量減少授課者維護blog平台回應的負擔。

## 2. 研究方法與系統架構

本研究在系統實作上提出以Blog為基礎架構的資訊學習平台作為系統展示方法，其架構圖如圖1所示，共計有下列幾個模組，模組功能如以下所敘述：

### (1) Blog System

本研究所建置的部落格除了新增、修改、刪除文章的的基本功能外，還具備了以下的功能及條件，可供本研究所使用：

(a) 文章分類標籤(tag)：

在針對文章專業領域的分類後，尚需要增加敘述性或是有涵義的分類標籤(tag)，此時可以透過web2.0概念下的社會化大眾標籤分類法(如圖2所示)，將一般詞彙定義成是附有語意的標籤。例如找尋一個地址時，一般人可能會說在台北101一家鐵板燒店的附近，附近明顯的地標(台北101、鐵板燒店)可稱為專業的分類，如果要在更準確的命中的話，再透過可能是外觀有彩色的霓虹燈招牌，霓虹燈招牌便可作為一個有意義的分類標籤。

而社會化大眾標籤分類法根據wikipedia[16]的解釋，有以下三大特點如下所列：

1. 大眾分類法由個人自發性定義。
2. 標籤分類是公開共享的，可以被所有人看到。
3. 此分類方式由用戶全體定義的頻率來決定。

大眾分類法擺脫了傳統分類的一致性，且跟一般人的認知程度密切結合，並讓全體用戶跟訊息之間有一個聯繫關係，雖然不準確、不嚴謹，但是這樣的分類方式，在社會化網路下，很自由的擴展網路並不受限制，因此在目前流行的社會性網路服務，都可以看到自訂標籤的大眾分類法，如del.icio.us、Flickr、43things等等的網路服務網站。換句話說，使用者可以很自由的定義標籤，想怎麼去描述，就可以怎麼定義分類標籤。

(b) 標籤分類樹(ontology)的建立：

一個要求語意的服務(Semantic Web)[7]，必須仰賴核心的語意推論分類樹，或稱Ontolog，Ontolog必須要能夠完整描述出整個系統的邏輯性以及彼此之間的關聯性，通常這都需要依靠各個領域的專家學者才有辦法詳細且深入地用XML文件表示出來，但是目前的狀態下，專家學者都專精於某些特定領域，如醫療、社會科學、古文詩詞等，然而要集合所有的專家學者來定義出完整的分類樹以供參考或實驗用，幾乎是不可能的。若是由老師自行去建立簡易的分類，不僅是老師本身方便使用，也因為是完全一手掌握整個分類的方法，整個邏輯性也很清楚明確，而可以給予分類間上下層的關聯程度。再配合部落格提供定義分類的功能，以及簡單且易使用的系統管理介面，使得分類樹能夠容易新增、修改、刪除節點。依照分類樹圖，老師針對部落格內得相關文章給予適當的分類，分的越細密越可以達到日後比對的準確度，這也是需要專家學者來建置的主因。

但是假如學生們也可以無限制的去定義自己的tag，那tag與tag之間的關係就無法被完整的建立，所以本研究提出以下的方法來建立tag的分類樹(ontology)：

1. 我們先建立出一個基本的分類樹(default ontology)，將文章先分為文學、數學、資訊科學...等的學科分類，每個學科下再建立更精確的科目名稱分類，並以樹狀結構儲存。

2. 當使用者想要新增一個tag時，必須去選擇一個default ontology中的tag做為這個tag的父親節點，這樣才能使所有tag之間的關聯性才能被建立起來。同時我們強烈建議編寫文章tag時，最好加入文章的屬性，例如：問題、解答、心得...等，這樣可以使得比對後的結果更加的匹配；因為若是使用者想要得到的是有關TCP問題的解答文章，但是在TCP這個tag之下，有可能找到的反而是別人提出問題的文章，如果加上了文章屬性的tag後，可以大幅提升回覆的精確度。

3. 最後我們定義tag與tag之間的關係以ontology樹狀結構中節點(node)與節點之間的關係為基礎，我們定義直系(也就是父子節點或直接mapping)的關係度為1，橫向(也就是兄弟節點)的關係度為0.5~0.9(依node離root的距離為考量，離root越遠的node是被分類比較精確的，所以相對的關係度也就比較高，tree的最大level限制為5)，如此一來我們就可以利用數個tag來計算出文章與文章或文章與comment(或引用通知)之間的關係度。

以圖2為例，Ajax這個tag與Javascript的關係度為0.6；Ajax與RSS的關係度就為 $0.7*0.6 = 0.42$ ；Ajax與Tag的關係度就為 $0.5*0.6*0.7 = 0.21$ 。透過tree之間degree與hop數的關係，可以建立出一個具有關聯性ontology。

(2) Agent

Agent會收到來自回應者/學生所寫下的comment及發佈文章的引用通知(或稱Track-back Notification)，這些都會觸發代理人去抓取遠方回文者的回應內容並轉發至Parser做分割字串處理。

最後將比對處理後的結果轉換成自動回應的內容，再度利用當初回應者在部落格上提問文章的引用通知，透過引用通知發佈所回傳的相關文章，回應者本身的部落格，則接收到來自老師部落格的引用通知。

(註：條件必須假設雙方的部落格允許任何人都可以發表留言，也沒有過濾垃圾留言的機制。如果agent沒有找到適合的文章來回覆，則會發e-mail通知管理者。)

(3) Comment Parser

本系統的設計是利用中研院的中文斷詞系統[1]來做斷詞的動作，而解析出來的字詞則透過中翻英翻譯程式，產生英文的字詞，然後提供給Fuzzy Mapping做比對用。

在這裡我們可以從學習者的comment或是引用文章通知取出一些字詞來與標籤分類樹做比對，例如：學生的comment是「請問TCP是什麼呢？」，parser會將將這個comment分解成「請問(VE) TCP(FW) 是(SHI) 什麼(Nep) 呢(T) ? (QUESTIONCATEGORY)」，括弧內為該字詞的詞性。字串透過斷詞後，可以藉由裡面的字詞去判斷它的語意，這將會對自動化回應的比對有相當大的幫助，使得字詞不再只是一個keyword(只能做完

全比對)，而是可以利用它的語意來做模糊化的比對。

#### (4) Fuzzy Mapping

Step 1：做完全字詞比對，找到表示尋找到所謂的專業詞組，沒找到表示沒有專業詞組。

Step 2：再對剩下的字詞，一一透過WordNet[17]所提供的同義字詞組模組，找出其與之相關的同義詞彙集合。

Step 3：藉由比對演算法，找出最有可能的部落格敘述性分類標籤。

Step 4：回傳找到的相關性文章id。

(附註：如果有找到專業詞組(Prof)，對於在自動搜尋相關類別的文章，整個搜尋範圍可以縮小，加快找到的時間及準確性。以條件機率表示P(找到相關的資料或是文章) =  $w * P(\text{Tag}|\text{Prof}) + (1-w) * P(\text{Tag})$ , w means weight)。

為了處理的速度不會被WordNet回應時間給拖垮，會把透過WordNet查詢到的同義詞組，額外儲存在table，以利之後查詢之便，系統預設表格大小為50，tag彼此之間是ontology的概念樹組成，換句話說，tag之間有上下關係，越靠近上層越是一般的分類法則，給予的權重數值應該要比較小，反之，越下層分析越細，給予的權重分配理應比較大。且必須是在類似的專業關鍵字分類下做一般字詞的tag比對。

當一個comment或引用通知被parse成幾個字詞後，我們可以藉由詞性判斷這個comment或引用通知的涵義；以剛才所舉的例子來說，「請問(VE) TCP(FW) 是(SHI) 什麼(Nep) 呢(T) ？(QUESTIONCATEGORY)」，可以從「問號」與「TCP」發現是一個有關TCP的問題，這時我們就會先以「TCP」去做比對，選出關係度最大的tag(或許是TCP或Network)，再利用「問號」(agent會判斷他是要找分類為「解答」的文章)，而去搜尋TCP或Network的分類下tag中有「解答」的文章；當比對的次數越多時，精確度也就越來越高。



圖2、標籤(tag)分類圖

### 3. 系統成果

藉由上述的研究方式，我們建置出系統雛型，而系統自動回應內容後的畫面如圖3所示，左側欄位是相關tag的連結，中間上方是文章本文，中間下方有track-back引用的通知，並提供此文章之RSS訂閱種子，右側欄位會列出與本文章相關的主題。

當學生在老師的教學blog上對某篇文章做出回應(comment)或是在自己的blog上寫一篇文章並引用(trackback)自老師的教學blog時，agent就會被觸發並去截取學生所回應或引用的內容，並交給parser去做斷詞處理，接下來透過一些語意上的分析與模糊的比對，找出適當的文章或書籍來做為自動化回應的內容，雖然agent在處理上必須花費一些時間，但是卻有效的減少了老師們在回覆問題(尤其是一些重複性或頻連性很高的問題)與維護系統時的負擔，也可以很即時的掌握學生的學習狀況(利用RSS feed的功能)。

本研究所提出的自動化回應機制的precision rate(正確率)與recall rate(回覆率)取決於使用者對於tag定義時的看法及Blog中文章數的多寡，因此在tag定義的客觀性以及文章數量上仍需更深入的討論與系統實證，這也是本研究要繼續探討的目標之一。

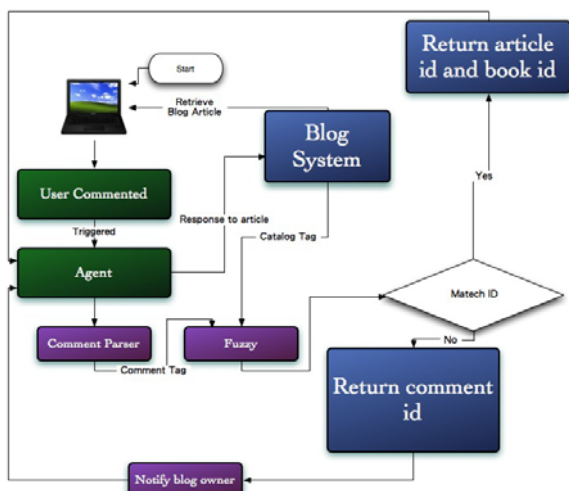


圖1、系統架構圖



圖3、系統雛型畫面

#### 4. 結論與未來發展方向

利用部落格匯集內容，使之成為個人學習的入口，也是老師學生之間的溝通平台。根據經驗法則，通常學生提出的問題或建議，大部份都是在部落格彙整文章可以找到相關資訊，除此之外，可提供相關的課外書籍以供參考；或是提供相關聯的中英文關鍵字，讓學生可以利用搜尋引擎搜尋，且透過代理人的協助，允許自動語意式回應，加快了回覆的效率與回文速度。以學生的角度來說，覺得能夠得到立即的回答與互動，是可以增加學習效率的；同時在自動回應之際，也發佈通知給老師，使老師從一個回應角色轉變成為一個審稿的稽核人員，透過自動化回應的代理人模式，老師可以不用時時刻刻去注意是否有新的回應或問題，同時讓數位學習具有立即回饋的互動性，讓學習、思考、發問與回饋可以連成一氣。透過部落格完善且彈性的設計，以及可以自由修改網站樣式的配置、排版以及跟其他網路服務的彙整服務，讓整個數位教材更加生動；加上 3G 通訊已漸為主流，行動部落格也儼然成型，對於未來的無所不在學習(ubiquitous e-learning)似乎又往前邁進一步。

線上學習絕對非城市小孩的特權，只要有了網路，縱使山區的師資比較缺乏，仍然可以達到相同的學習效果。目前世界上正在推動OLPC計畫(One Laptop Per Child)，此非營利組織的理想是，對發展中國家提供廉價PC，協助拉近城鄉數位落差，並鼓勵老師可以自願請調山區教學；不如要求老師有線上學習教材編輯能力，現在的孩子碰電腦的時間，肯定比摸書本或是讀書的時間多，本研究建議適度的使用英文撰寫文章，並非要求要看懂，而是培養一種不害怕英文的心理障礙，英文與電腦能力，絕對是未來不可或缺的個人能力。

最後本研究希望先透過引起教師/學生有能力且有意願去使用線上學習的啟發為主，期待透過簡單的使用者介面與系統管理和學生喜歡新鮮事物的好奇心態，讓師生能嘗試這種學習模式。一旦blog的機制產生作用後(發文者會因為回應者的回應而更有發文的動力)，相信老師也會有興趣，並開始專研如何增加套件的寫作，使數位學習系統能夠更加的貼切且適性化。

Blog並不是一個完整的數位學習平台，必須藉由套件的擴充與其他網路服務的合作才能發揮出接近於LMS/LCMS的功能；必竟LMS/LCMS才是一個正統且有學習方法理論為基礎所架構的線上學習平台，未來希望將此機制更深入探討且研發為更有效的LMS/LCMS套件形式。

#### 參考文獻

- [1] 中研院中文斷詞系統，<http://ckipsvr.iis.sinica.edu.tw/>
- [2] 王小惠，“部落格數位學習平台效果之初探性研

- 究”，行動與無所不在數位學習研討會論文集，Apr. 2006，pp. 118-125。
- [3] 陳永昇、廖尚儀、黃俊嘉。“行動校園整合資訊之研究與實作”，TANET2003 台灣網際網路研討會，Oct. 2003，pp. 549-554。
- [4] 陳永昇、廖尚儀、李恩慈，“建立基於訊息整合技術之互動式師生聯絡網”，TANET2004 台灣網際網路研討會，Oct. 2004，pp. 80-85。
- [5] 陳永昇、郭庭榕，“運用理人伺服器建置行動數位學習平台之設計與研究”，TANET2005 台灣網際網路研討會，Oct 26-28，2005，p. 86。
- [6] M. Divitini, O. Haugalokken and E. M. Morken, “Blog to support learning in the field: lessons learned from afiasco”, *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies*, 2005, pp. 219-221.
- [7] J. Euzenat, “Eight questions about Semantic Web annotations”, *Intelligent Systems, IEEE*, 17, 2, Mar./Apr. 2002, pp. 55-62.
- [8] M. Fernandes, M. Alho, J. A. Martins, J. S. Pinto, and P. Almeida, “Web Annotation System Based on Web Services”, *Proceedings of the International Conference on Next Generation Web Services Practices*, 2005.
- [9] L. Giarre and L. Jaccheri, “Learning Research Methods and Processes via Sharing Experience in a BLOG”, *2005 European Decision and Control Conference*, Dec. 2005, pp. 2716-2720.
- [10] J. Kahan, M. Koivunen, E. Prud'Hommeaux and R. Swick, “Annotea: An open RDF infrastructure for shared Web annotations”, *Proceedings of the WWW10 International Conference*. Hong Kong, 2001.
- [11] B. A. Nardi, D. J. Schiano, M. Gumbrecht and L. Swartz, “Why we blog”, *Communications of the ACM*, 47, 12, 2005, pp. 41-46.
- [12] Tim O'Reilly, *What Is Web 2.0*, 2005, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-2.0.html>
- [13] Ultimate Tag Warrior, <http://www.neato.co.nz/ultimate-tag-warrior/>
- [14] V. Vasudevan and M. Palmer, “On Web annotations: promises and pitfalls of current Web infrastructure”, *System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference*, 2, Jan. 1999.
- [15] Wei-Jane Lin, Hsiu-Ping Yueh, Yi-Ling Liu, M. Murakami, Koh Kakusho and M. Minoh, “Blog as a tool to develop e-learning experience in an international distance course”, *Proceedings of the Sixth International Conference on Advanced Learning Technologies*, 2006.
- [16] Wikipedia, <http://en.wikipedia.org/wiki/Folksonomie/>
- [17] WordNet, <http://wordnet.princeton.edu/>