

# 以中文缺字技術為基礎之多媒體設計平台

卓楷斌<sup>12</sup> 王祥安<sup>1</sup> 林德潤<sup>1</sup>

<sup>1</sup>中央研究院資訊科技創新研究中心

<sup>2</sup>元智大學資訊管理學系

{comein5559, sawang, soar}@iis.sinica.edu.tw

## 摘要

隨著網際網路的蓬勃發展，使用者日漸習慣透過網路產製專屬的數位內容，然而中文缺字問題，造成部份中文字無法被完整的呈現與使用，因而限制了文字在數位內容產業中的應用。

本研究整合中文缺字技術和多媒體技術，發展一套可同時編輯中文字與影像的線上多媒體編輯平台，使用者能夠透過此平台製作出一般電腦未編碼的漢字、古漢字以及合文等文字內容，並可以將它與多媒體內容整合，進行數位圖章、電子賀卡等加值應用。此平台可提供數位內容製作者作為中文字的應用中心，解決中文缺字製作與應用的問題。

**關鍵詞：**中文缺字、多媒體編輯、動態組字、數位典藏、數位設計

## 1. 前言

在電腦系統中，每個文字皆存在對應的編碼，早期電腦源自於英語系國家，其編碼系統只需包含英文、數字和符號等封閉式的集合，所以不存在缺字問題；然而在亞洲語系國家中的表意文字，由於電腦編碼空間有限而無法被完整對應，使得文字資訊無法適當的傳達、溝通和共享，因而產生缺字問題。

目前中文 BIG5 編碼只收納了一萬三千多個字，在文字的運用上仍存在著許多缺字，面對此問題，一般單位常以造字法作為解決方案，但造字法不僅費時、費力，在不同造字檔間也因文字編碼並不完全相同，而在整合上可能發生同字異碼或異字同碼的狀況，因此造字法不僅沒有系統不易管理外，也無法達到文字的流通和共享。

後續發展的 Unicode 計畫，增加所收錄的漢字數量，降低了缺字發生的機率，但因中文字的數量十分龐大而無法完整收錄所有的漢字，況且漢字會隨著時空環境改變，衍生出字音、字義、字形之變化，隨著日後新字的產生，

缺字問題仍會再度浮現，面對如此複雜的文字結構，最好的解決方式就是打破原先設計編碼，改以適合漢字結構的方式來處理缺字問題。

目前政府單位設立了全字庫[13]作為政府機關缺字問題的解決方案，而中央研究院文獻處理實驗室[1]亦發展出漢字構形資料庫[12]來處理缺字問題，全字庫收錄的文字共有 10 萬餘字，而漢字構形資料庫收錄了將近 12 萬的古今漢字，相較之下漢字構形資料庫在漢字收錄上較為完整，而且它不僅擁有漢字構形的知識，也收錄了古今不同字集及異體字，因此能提供給有缺字需求的使用者較好的利用。

就缺字的使用需求上，我們可以由「政府機關中文資訊應用概況調查分析結果」[4]得知，在缺字問題發生的狀況下，以特殊人名、地名、專有名詞和廠商名稱命名的造字使用頻率最高。在日常生活中此類的造字需求屢見不鮮，尤其以數位產業最為明顯，像是姓名、祖譜上的文字、特殊地名之紀念圖章，或是罕見文字廠牌的製作，而目前進行數位設計的廠商所提供的文字數量多為常用字，因此對於罕見文字的需求便顯得力有不足。

缺字問題不但限制了數位產業未來發展的延展性，同時一些較具紀念性、價值性內容的創造也會因此而面臨瓶頸；然而隨著網際網路的蓬勃發展，人們逐漸習慣透過網路進行溝通和傳達訊息，文字資訊在網路上的流通、應用變得更加普遍，目前漢字是世界上最多人使用的文字，近年來各國使用漢字的人口也持續地增加，因此解決缺字問題更顯得有其必要性。

因應此問題，本研究整合了缺字和多媒體技術，運用漢字構形資料庫以及動態組字的缺字解決技術，提出了一套中文字整合設計平台，其目的在於提供解決中文缺字問題以及缺字無法在數位產業或一般大眾之間普及共享、流通、應用的問題，並以生動有趣的介面吸引更多使用者前來使用，滿足生活中的實際需求，使用者能夠取得所需的文字資訊，不僅在數位內容產製的過程能被完整地呈現，在一般單位也能更加普及地應用。

## 2. 相關理論與技術

### 2.1 漢字構形資料庫

漢字構形資料庫是為了解決文字編碼衍生缺字問題所提出的解決方案。就中文漢字的字形結構而言，每個中文字可視為由各個不同的部件所組合而成，舉例來說，「萌」字可拆解為「艸、日、月」三個部件，以此概念對每一個中文字做延伸，便能夠解決漢字編碼的問題，因為每個文字皆能以有限的部件來表達不同字形結構的文字。

漢字構形資料庫記錄漢字的方式是以上述概念為基礎，依據不同的字形結構作構形分析，並發展出一套漢字的表達方式，稱之為構字式，構字式是以漢字拆解的部件和特定構字符號所形成的漢字表示方式。

構字符號總共有十三種，分別為： $\Delta$ 、 $\triangle$ 、 $\Delta$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ 、 $\circ$ ，以下是構字式的規則說明，如表 1 所示。

表 1 構字符號說明表

類別	符號	說明
連接符號	$\Delta$	部件的連接由左至右
	$\triangle$	部件的連接由上至下
	$\Delta$	部件的連接由外至內
部件序	$\square$	按部件書寫順序輸入，前後以起始符號( $\square$ )和終止符號( $\square$ )包夾
	$\square$	
方便符號	$\circ$	兩個相同部件直聯
	$\circ$	三個相同部件直聯
	$\circ$	兩個相同部件橫聯
	$\circ$	三個相同部件橫聯
	$\circ$	三個相同部件呈三角狀排列
$\circ$	四個相同部件橫聯	

構字式有許多優點，在文字數位化的過程中，能夠以有系統的方式將文字分門別類，並提供使用者一套規則作為日後查詢的依據。構字式切合了漢字以義構形的特點，以漢字形體為基礎來解決缺字問題，如此一來，即使未來所產生的新字也能夠以正規化的表達方式記錄在漢字構形資料庫中。

早期漢字構形資料庫的建置是以楷書字形為主，再陸續延伸至古今漢字的整理、收納各

種字集的漢字，並建立不同歷史時期的異體字表，以表達不同漢字各時期文字間的關係。

目前資料庫共收錄古今漢字 11 萬 9 千餘字，其中楷書字形約 6 萬 4 千個，小篆約 1 萬 1 千字，金文約 2 萬 1 千字，楚系簡帛文字約 1 萬 9 千字，甲骨文有 2 千餘字。而異體字表的部分共收錄《漢語大字典》1 萬 2 千餘組。

### 2.2 動態組字

漢字構形資料庫雖能解決缺字問題；然而若目前需要用的字仍未被收錄在資料庫中，即使使用者拼湊出構字式也有可能檢索不到，此時必須透過一道標準程序申請新字，但缺點是資料庫的更新不具即時性，必須等到資料庫中紀錄新的文字後才能夠使用，而動態組字便能解決這樣的不便。

目前剎那搜尋工坊[5]所開放的可攜式造字引擎可以用來實作動態組字，其組字方法是利用 Unicode 和表意文字序列(Ideograph Description Sequence)的組合來表示。表 2 分別比較漢字構形資料庫和可攜式造字引擎之間的差異。

表 2 漢字構形資料庫與動態組字比較表

漢字構形資料庫 構字表示式	可攜式造字引擎 動態組字方法
中置式表示法	前置式表示法
部件序、構字符號	Unicode、表意文字序列

由於可攜式造字引擎和構字式的表示法不盡相同，因此必須透過中置式轉前置式的步驟，才能將動態組字和漢字構形資料庫做整合，如此一來系統便能透過構字式轉換、擷取字形圖片，最後將文字即時輸出給使用者。

動態組字的需求較常發生在合文的呈現上，所謂的合文是由多個漢字組合而成，在字典中往往找尋不到，像是「孔孟好學」、「招財進寶」，這些字皆可以透過動態組字的圖片擷取模組將文字即時地呈現出來，如圖 1 所示。



圖 1 動態組字示例圖

### 3. 系統架構

本研究以中文缺字處理技術為基礎建構了一套系統，以解決數位內容產製過程所發生的缺字問題，使用者能夠利用此系統所提供的設計平台，以圖文互動的編輯方式產製出個人專屬的數位內容，而數位內容的引用和產製是透過系統中各個模組間互動達成的。

系統中涵蓋幾個模組，分別是缺字檢索模組、動態組字模組、圖像處理模組，以及多媒體編輯模組，系統架構圖如圖 2 所示，而以下部分將介紹這些模組在系統中所扮演的角色。

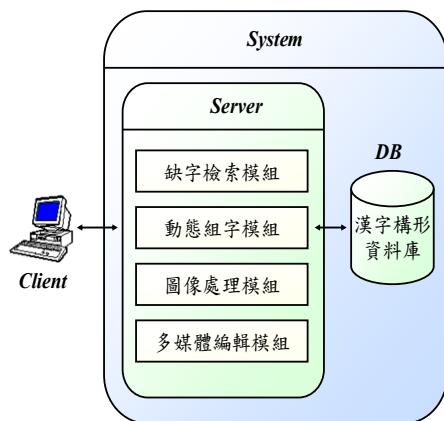


圖 2 系統架構圖

#### 3.1 缺字檢索模組

文字資訊的部份，使用者能夠透過缺字檢索模組至漢字構形資料庫中取得所需的文字，而檢索模組提供多種查詢方式讓使用者取得文字資訊，以下將以幾種不同的檢索功能說明之。

##### 3.1.1 字形檢索

字形檢索會將使用者輸入的文字送至漢字構形資料庫中作字形比對，並將符合的文字資訊回傳給使用者。除了直接輸入文字外，使用者還能透過部件檢索、部首查詢、漢語拼音檢索取得所要的文字資訊。

以部件檢索為例，其使用時機通常是使用者不知道該字的讀音，或是輸入法無法支援使用者輸入該文字，在此狀況時使用者只需輸入拆解後的部件作部件檢索，系統便能將這些部件拆解成字根並對資料庫作模糊查詢，在檢索後將相符的文字資訊回傳給使用者。

##### 3.1.2 異體字檢索

異體字檢索的部分，能夠讓使用者認識在中國文字演變過程中，各時期文字書寫上字形結構的變化，以及了解不同文字間的字義的關

係，檢索到的異體字能提供使用者作後續利用。

#### 3.2 動態組字模組

當缺字檢索模組找不到使用者所需的文字資訊時，動態組字模組能提供即時自行組字的服務。此模組是以構字式做為組字依據，但因為組字構字式對一般使用者而言進入門檻高且不易使用，因此介面設計上以圖像化為原則，其操作介面將於下一章系統實作再多加介紹。

#### 3.3 圖像處理模組

此模組負責圖像資訊的處理和應用，在進行多媒體編輯的功能前，使用者能夠提供適合的圖像資訊作為設計的元素，而此模組提供使用者上傳圖片、透過網址引用外部圖片兩種方式，當圖片來源確定後，為了符合設計規格，本模組亦提供使用者對圖像作等比裁切的功能，而處理後的圖片將呈現於平台等待進行後續的編輯動作。

#### 3.4 多媒體編輯模組

多媒體編輯模組以 Flash 開發，使用的程式語言為 Action Script 3.0，以下為此模組所開發的編輯功能，其功能包括：

- 畫線：使用者能藉由調整線段的粗細、顏色和透明度等屬性，畫出其想要的線段。
- 拖曳：拖曳已產生的線段、文字及圖片。
- 刪除：刪除點擊到的線段、文字及圖片。
- 縮放：將文字、圖片作放大、縮小的設定。
- 旋轉：以自行設定的角度作物件的旋轉。
- 變形：以自行設定的角度作物件的變形。

使用者能夠利用多媒體編輯模組所提供的功能，將文字、圖像資訊做進一步的設計，直到使用者確認其內容已設計完成後，系統會將這些產製出的數位內容作品儲存在伺服器端，以進行產製內容的應用，而這些作品也能在不同的使用者間分享、傳播、引用和再設計。

### 4. 系統流程

系統的流程簡化如圖 3 所示，而此系統的操作說明如下列分項詳加解釋。

- 選擇文字或圖片  
使用者登入系統後，在功能選單的地方可以選擇使用文字和圖片的輸出功能，以進行後續多媒體編輯引用圖文的動作。
- 選擇文字檢索  
文字資訊的部分，首先給使用者選擇文字檢索方式，例如：字形檢索、部件檢索、部首查詢、漢語拼音查詢及異體字檢索。而後系統會利用使用者輸入的查詢條件到



## ● 異體字檢索操作說明

異體字檢索的實作畫面如圖 6 所示，輸入文字為「漁」，使用者能透過不同字集的選擇檢索到不同時期使用文字的狀況，如甲骨文、金文、楚系文字等，圖 6 是選擇金文字集的結果，最後在使用者選擇所欲呈現的異體字後，系統會將使用者選擇的異體字作編輯和輸出。

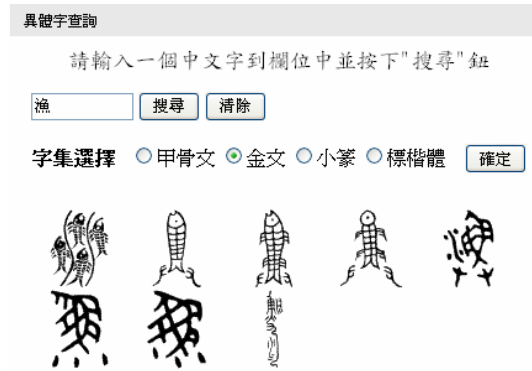


圖 6 異體字檢索實作範例圖

## ● 動態組字操作流程

在動態組字的過程中，使用者只需輸入欲組出文字的拆解部件，並將不同部件做拖拉組合，系統模組便能自動啟動構字式轉換的功能並達到動態組字效果。

圖 7 以合文「唯吾知足」為例，說明上下和左右形式的部件組合，其操作說明如後。

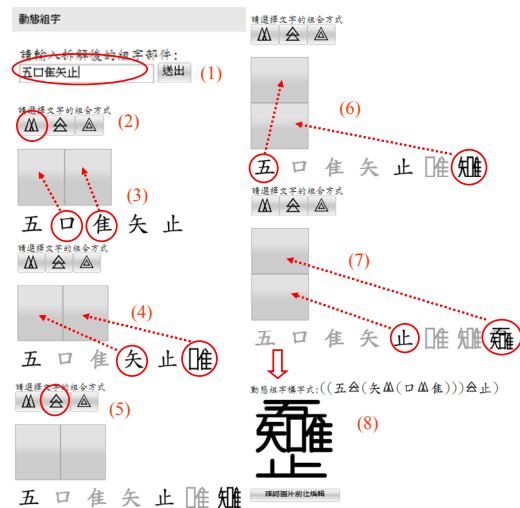


圖 7 唯吾知足之動態組字操作說明圖

(1) 進入動態組字介面後，使用者首先輸入所欲組字的拆解部件，並按下確認按鈕，系統會將其輸入的部件設成可拖拉元件，並顯示於下端。  
(2) 使用者能依個人喜好選擇組字的順序，在此例中，首先第一個欲組出的字為「唯」，其部件

的连接顺序是由左而右，因此選擇構字符號△。

(3) 將「口、隹」兩個部件拖拉到組字框內，此時系統會自動產生組出的「唯」字於下端，並將原先「口、隹」兩個部件的拖拉功能取消。

(4) 接著要組出「知」字，其連接順序一樣是由左而右，因此不需改變其構字符號型態，而唯吾知足的「知」和「唯」兩個字共用一個「口」部件，因此必須將「矢、唯」兩字拖拉至組字框中，並完成「知」字的組合。

(5) 接著要組出「吾」字，其部件的連接順序是由上而下，因此選擇構字符號△，而系統會將組字框設定為上下的拖拉型態。

(6) 同理，「吾」字和剛才組出的字擁有共用部件，所以此步驟將「五」字和「知」字作組合。

(7) 最後將「止、彙」兩字組合成「彙」。

(8) 顯示出唯吾知足的「彙」字及其構字式，當使用者確認完成後，便能將此文字做新的用途。

圖 8 是「也」字的動態組字的操作說明圖，說明內外部件的組合方式，其操作說明如後。

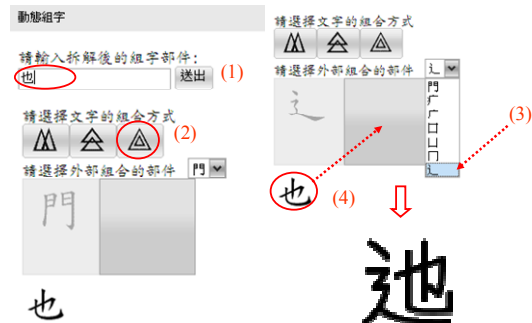


圖 8 也字之動態組字操作說明圖

(1) 使用者首先輸入部件並按下確認鈕，而系統將此部件設成可拖拉元件顯示於下端。

(2) 選擇組字符號△能完成內外部件的組合。

(3) 內外部件組合中的外部部件有「門、广、广、口、口、刀、之」幾種選擇，使用者在此步驟必須選擇外部組合的部件而不需手動輸入，此例中以「之」做為外部部件。

(4) 將「也」拖拉進組字框後能夠組出「也」字。

經過以上的說明，可以得知圖像化的方式能夠幫助使用者順利組出構字式以完成動態組字的動作，在操作上不僅變得簡易使用，使用者還能自己發揮創意、親自參與設計過程，創造出具有意義且實用的文字資訊。

## ● 數位電子圖章操作流程

數位電子圖章是提供製作個人專屬圖章的一項應用服務，使用者能利用多媒體介面對文字作拖拉、組合，以便完成固定形式的圖章，完成後的作品能被設計平台直接引用。圖 9 表示數位電子圖章的操作說明圖，其說明如後。



圖 9 數位電子圖章操作說明圖

- (1)提供幾種固定版形以選擇圖章文字的陳列方式,使用者能夠依照所欲輸入的字數做選擇。
- (2)輸入文字並且按下送出文字按鈕。
- (3)系統會依照使用者輸入的文字送至資料庫作字集的比對,並且列出文字存在的字集,而此例中使用者選擇了小篆字集。
- (4)選定圖章的顏色,並按下製作圖章按鈕。
- (5)系統會自動將製作圖章的效果呈現於右上方供使用者預覽,確認完後按下確認按鈕並能夠將該圖章存檔以作後續的應用。

透過如上所述的操作流程,使用者能夠設計出如圖 10 之數位電子圖章的成品範例。

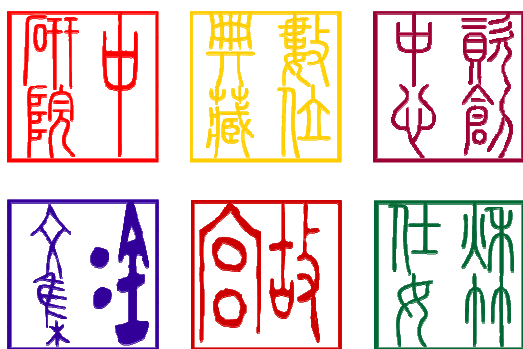


圖 10 數位電子圖章設計成品範例圖

#### ● 多媒體圖文編輯

透過中文字整合設計平台,使用者能在多媒體開發的介面下進行先前所輸出圖片、文字的編輯動作,如圖 11 所示,使用者能夠對輸入的圖片做編排、用畫筆畫線作註解,以及對圖文做拖曳、縮放、旋轉及刪除等動作,如此一來多媒體編輯功能使得文字資訊在數位產業的使用上變得更加豐富、便利。

圖 12 是以多媒體編輯介面實作出數位典藏應用的範例,在設計過程中亦能引用先前所

設計的數位電子圖章至設計畫面中,當使用者編輯完畢後,系統將其作品進行存檔和應用。

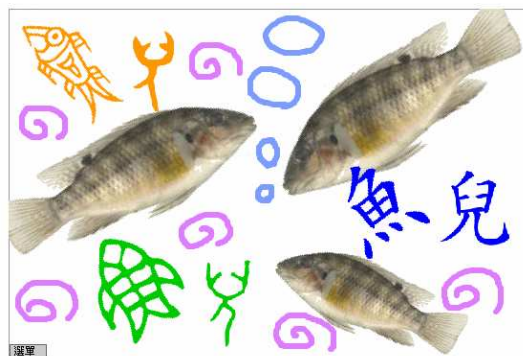


圖 11 多媒體線上編輯範例圖

此圖引用自數位典藏聯合目錄網站[2]。



圖 12 整合設計平台在數位典藏的應用

此圖引用自數位典藏聯合目錄網站[6,8,9]。

由於多媒體的介面較能體現圖文並茂的效果,對於使用者而言較為直覺,也能實現所見即所得的精神,因此應用上能夠作為數位典藏展示、古今漢字的數位學習,進而吸引更多新的使用者前來產製創新的數位內容。

#### 6. 結論與未來展望

本研究整合了以缺字技術為基礎,利用漢字構形資料庫及動態組字技術,整合了多媒體技術建構出一個中文字製作的設計平台,以解決數位內容產製因缺字問題帶來的不便。而本研究能推廣給一般有特殊文字需求的使用者,甚至讓世界各地對於漢字文化有興趣的人更方便地取得文字資訊並產製其個人數位內容。

本系統目前的應用有線上電子賀卡,能運用漢字構形資料庫中文字提供使用者在設計平台上製作卡片,寄送個人作品給認識的親友,再者,使用者能夠自行發揮創意,將不同歷史時期的文字加以運用,實現個人專屬圖章的製作,達到中文字在數位產業的實際應用。

本系統未來可以有廣泛的發展方向，諸如：姓名祖譜、特殊地名 Logo 製作、藏書章、藏書票，或是罕見商標廠牌的製作，亦或是利用目前系統所開發之編輯功能，配合 CAD(Computer-Aided Design)製作成各項中文字之需求產品，像是紀念 T-Shirt 或是紀念杯等實體產品，滿足使用者創造數位內容的需求。

本系統亦能夠作為數位學習的一套工具，因為漢字構形資料庫所收錄的文字來源大多是古籍中的文字和各時期的古今漢字，因此未來可以發展成古今漢字的教材，使用者可以對古漢字能有更多的接觸並有機會欣賞漢字的結構和其演變流程，另外，也可以作為數位典藏的傳播和展示，讓使用者將具有典藏意義的數位內容加以推廣、傳播、應用與呈現。

### 誌謝

此研究計畫由台灣行政院國家科學委員會的數位典藏與數位學習國家型科技計畫(TELDAP)資助，計畫編號：NSC98-2631-001-011、NSC98-2631-001-012。

### 參考文獻

1. 中央研究院文獻處理實驗室，  
<http://www.sinica.edu.tw/~cdp/index.html>
2. 吉利慈綱，《數位典藏聯合目錄》，  
<http://140.109.18.74/ImageCache/ImageCache/00/00/1a/19.jpg>
3. 李宏益、林德潤、王祥安，『漢字構形資料庫在網路上的應用』，TANET 2008 台灣網際網路研討會（2008）。
4. 政府機關中文資訊應用概況調查分析結果  
<http://www.cns11643.gov.tw/eng/reports.jsp>
5. 剎那搜尋工坊，<http://www.ksana.tw/>
6. 秋竹仕女，《數位典藏聯合目錄》，  
<http://140.109.18.74/ImageCache/ImageCache/00/05/b0/ec.jpg>
7. 莊德明、謝清俊，『漢字構形資料庫的建置與應用』，漢字與全球化國際學術研討會（2005）。
8. 梅竹仕女，《數位典藏聯合目錄》，  
<http://140.109.18.74/ImageCache/ImageCache/00/05/a6/ae.jpg>
9. 梅花仕女，《數位典藏聯合目錄》，  
<http://140.109.18.74/ImageCache/ImageCache/00/05/a7/74.jpg>
10. 黃俊瑋、林金龍、黃國倫，『缺字系統整合動態組字之應用』，TANET 2007 台灣網際網路研討會（2007）。
11. 數位典藏國家型計畫/數位典藏技術發展組  
<http://daal.iis.sinica.edu.tw/Chinese/System/index.htm>
12. 漢字構形資料庫，  
<http://www.sinica.edu.tw/~cdp/cdphanzi/>
13. CNS11643 中文標準交換碼全字庫  
<http://www.cns11643.gov.tw/AIDB/welcome.do>
14. T. J. Lin, J. W. Huang, Christine Lin, H. Y. Li, H. A. Wang, C. Y. Chiu, “A Mechanism for Solving the Unencoded Chinese Character Problem on the Web”, ECDL, Aarhus, Denmark, 2008.