

浮世絵閲覧システムのメタデータに対する 多言語アクセス手法の提案

木村 文則¹ Biligsaikhan Batjargal² 前田 亮¹

¹ 立命館大学 情報理工学部 〒525-8577 滋賀県草津市野路東 1-1-1
E-Mail: {fkimura, amaeda}@is.ritsumeai.ac.jp

² 立命館大学理工学研究科 〒525-8577 滋賀県草津市野路東 1-1-1
E-Mail: biligsaikhan@gmail.com

概要

本論文では、浮世絵データベースの言語横断アクセス手法の提案を行う。本システムでは、オープンソースのデジタルライブラリソフトウェアである Greenstone Digital Library Software 上でシステムの構築を行う。また、浮世絵データベースのメタデータ属性を Dublin Core のメタデータ要素にマッピングを行うことにより、他のデータベースとの横断検索を行うことを容易にしている。さらに、浮世絵データベースに対して英語による問合せを日本語に翻訳する手法を提案し実装を行った。それにより日本語が理解できない利用者でも容易に浮世絵データベースにアクセスをすることを可能とした。

キーワード

浮世絵, 画像データベース, デジタル図書館, 言語横断情報検索

Proposal of Multilingual Access Method for Metadata of Ukiyo-e Image Database

Fuminori Kimura¹, Biligsaikhan Batjargal² and Akira Maeda¹

¹ College of Information Science and Engineering, Ritsumeikan University
1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577, JAPAN
E-Mail: {fkimura, amaeda}@is.ritsumeai.ac.jp

² Graduate School of Science and Engineering, Ritsumeikan University
1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577, JAPAN
E-Mail: biligsaikhan@gmail.com

Abstract

In this paper we introduce our system that retrieves Ukiyo-e databases using an English query by customizing and utilizing freely available open source software. In our system, the Ukiyo-e metadata elements were mapped to Dublin Core. We adopted a dictionary-based query translation approach and utilized the Greenstone Digital Library Software to make available our Ukiyo-e digital collections online. The preliminary result is an easy-to-use and useful system for users who do not understand Japanese, that allows to search and view Japanese Ukiyo-e databases in English.

Keywords

Ukiyo-e, Image database, Digital library, Cross-language information retrieval

1. はじめに

浮世絵は江戸時代に成立し、演劇や古典文学から人々の日常生活や風物に至るまで、当時の様々な風俗を題材として描かれている。浮世絵を媒介とすることにより、後世の人々にも江戸時代の庶民の生活や風俗を垣間見ることも可能である。それゆえ、江戸時代の文化を理解するための資料として重要視されている。

一方、美術品としての価値は、もともと大衆文化の一部として描かれたいきさつから、それほど高かったわけではなかった。しかし、現代では浮世絵の芸術性の高さが見直され、美術品としての価値が高まっている。また海外においても、浮世絵は西洋の近代画にも影響を与えた日本を代表とする美術品として注目されている。

従来、浮世絵は美術館や大学の研究所などに所蔵されていることが多く、観賞するためには所蔵されている場所まで行く必要があった。しかし、近年では美術品や芸術作品などをデジタル化してデジタルアーカイブとして保存することが行われ始めている。浮世絵も同様であり、美術館や大学の研究所などにおいてデジタル化されている。また、これらのデジタル化されたアーカイブはインターネット上に公開することが可能である。早稲田大学坪内博士記念演劇博物館「演劇博物館所蔵浮世絵閲覧システム」[1]や立命館大学アート・リサーチセンターの「ARC 所蔵浮世絵検索閲覧システム」[2]などの Web サイトにおいて浮世絵のデジタルアーカイブが公開されている。

このように、非常に多くのデジタルアーカイブが公開されるようになってきたことで利便性は高まっているが、その反面、大量のデジタルアーカイブから目的の浮世絵を効率よく探すことが重要となっている。そのための検索手法、例えばデータベースを横断した検索などが求められている。また、浮世絵は日本発祥であるが、海外でも関心が高いことから、海外において所蔵されている浮世絵も多い。それらのうち、Minneapolis Institute

of Arts[3]などのように Web サイトでも公開されているものもある。

このように、世界中に所蔵されている浮世絵の所在を把握することは困難である。また、浮世絵に関するデータベースも世界の各所において公開されており、一元管理は困難な状態である。このような状況において、出来る限り全ての浮世絵を対象に検索を行いたいという要求にこたえるは、データベースを横断した検索の技術が必要となる。

そこで本論文では、複数の浮世絵データベースの検索を容易にするため、浮世絵閲覧システムのメタデータに対する言語横断アクセス手法の提案を行う。

2. 関連研究

2.1 古典史料に対する言語横断検索

これまでに言語横断情報検索の分野において多くの研究がおこなわれてきた。言語横断情報検索とは、検索対象文書で使用されている言語とは異なる言語を用いた問合せ語による検索を行うことである。例えば、日本語の問合せ語を用いて、英語で書かれた文書の検索を行うことである。しかし、古典史料に対しての言語横断情報検索を実現した研究はほとんどない。古典史料に対して言語横断検索の場合、利用者は現代語による問合せを行い、システムはそれを古語に翻訳して検索対象文書を検索することになる。この場合、現代語と古語の間で時代を横断していることから、「時代横断検索」に相当する。近年、いくつかの研究では古典史料に対する検索が行われている。我々は以前の研究において、伝統的モンゴル語と現代モンゴル語の時代横断検索を行う手法を提案している[4]。また、日本語における古語と現代語の時代横断検索についても行っている[5]。Koolenら[6]は、オランダ語の古語と現代語における綴りと発音の違いに焦点を当て、その類似度をもとに翻訳を行っている。また Pilzら[7]は、歴史的文書での英語とドイツ語における綴りの変化に焦点を当て、その類似度をもとに翻訳を行っている。このように、ヨーロッパにおいては綴りの相違を考慮することが多い。

それに対し、日本語においては古語と現代語でほぼ同じ意味の単語であっても表記が明らかに異なることが多い。また、日本語では単語の区切りが明記されていないことも、英語など欧米の言語と比べて処理が困難である要因となっている。さらに、日本語では漢字が使用されるが、同じ漢字でも複数の読みがあることなども、処理の複雑さを生じる要因となっている。

2.2 立命館大学 ARC 所蔵 浮世絵検索閲覧システム[2]

立命館大学アート・リサーチセンターでは 5,011 件の浮世絵がデジタル化され、「立命館大学 ARC 所蔵 浮世絵検索閲覧システム」においてインターネット上で公開されている。

ARC浮世絵閲覧システム

検索語は基本的に先頭一致で検索します。検索語を含むものを探す場合は検索語の末尾に「*」を付けてください。
項目名をクリックすると検索項目の説明を表示します。この説明からも検索を実行できます。
検索語を入れてください。

Caution! if you find the message of "proxy error", change your browser's properties following this instruction.
「プロキシエラー」が表示されたら、ここをクリックしてブラウザのプロパティを変更してください。

資料番号: 所蔵機関: 立命館ARCのみ

絵師: 落款:
 彫師等: 版元: 印1 元1 文字名
 改印: 判型種別:
 画題等: 分類:

上演: 西暦 和暦 年(閏) 月 日 劇場 地域
 演目: 演目よみ:
 幕・場名: 配役:

結果の表示形式

一度に表示する件数

同板画像を除く

分割撮影画像を除く

整列順: (複数指定可)

<input type="text" value="-なし-"/>	<input type="text" value="-なし-"/>	<input type="text" value="資料番号順"/>
<input type="text" value="西暦順"/>	<input type="text" value="西暦順"/>	<input type="text" value="-なし-"/>
<input type="text" value="月日順"/>	<input type="text" value="月日順"/>	<input type="text" value="Group番号順"/>
<input type="text" value="地域順"/>	<input type="text" value="地域順"/>	<input type="text" value="重複番号順"/>

検索条件を入れたら下のボタンをクリックしてください。検索語なしで、全件表示します。

[浮世絵検索システムのエンタランスへ戻る](#)

図 1. 立命館大学 ARC 所蔵 浮世絵検索閲覧システム 検索画面



詳細情報表示 45件の内11件目を表示

←画像をクリックすると、オリジナルサイズの画像を見ることができます。

AcNo. UP1173 CoGNo. UP1173 AIGNo.

絵師: 広重(1) 落款印章: 広重画 彫師: 二

判型: 大判/横繪 統方向横 作品位置: 01,01/001.001

異版: Col: 重複: 1 All: 重複: 1 (頁: 1)

版元文字: 長、芝神有田屋 版元備考:

版印1: 208 版元1: 29 版元名1: 有田屋 清右衛門

版印2: 版元2: 版元名2:

改印: 村 蔵印: 二

出版年月日: 〇・〇出版地: 江戸

画題等: 「忠臣蔵 大序」

上演年月日: 弘化02(1845)・(頃)

上演場所: 江戸・

興行名: 場立:

外題: 場名:

所作題: 音曲:

細目題:

配役:

組解説:

個別解説:

画中文字:

系統分類: 忠臣蔵、大序 所蔵: 立命館ARC

続き物・組み物 (同一物)を表示	same collection	1024×	800×
	all prints	1024×	800×

図 2. 立命館大学 ARC 所蔵 浮世絵検索閲覧システム 検索結果

図 1 は、立命館大学 ARC 所蔵 浮世絵検索閲覧システムの検索画面である。「絵師」、「彫師等」、「画題等」など、様々な項目をキーワードとして浮世絵の検索を行うことができる。利用者がキーワードを入力して検索を行うと、検索結果として該当する浮世絵の候補が提

示される。その中から閲覧したい浮世絵を利用者が選択すると、その浮世絵の画像および、その浮世絵に関するメタデータが提示され（図2）、閲覧することができる。

3. 浮世絵検索閲覧システムの言語横断アクセス

本論文では、立命館大学 ARC 所蔵 浮世絵検索閲覧システムに格納されているデータの言語横断アクセスの手法を提案し、その実装を行った。浮世絵データベースのメタデータ要素（以下「メタデータ属性」と呼ぶ）を標準的なメタデータである Dublin Core のメタデータ要素へのマッピングを行い、他のデータベースのメタデータ属性との互換性を図った。また、浮世絵データベースの個々のメタデータのローマ字化を行い、浮世絵検索閲覧システムに対する英語（ローマ字）による検索を行うことのできるシステムを構築した。それに加え、提示する浮世絵に関する情報もローマ字化して提示できるシステムとなっている。本論文ではオープンソースの電子図書館ソフトウェアである Greenstone Digital Library Software を用いてシステムを構築している。

システムの構成および英語の問合せによる検索の流れを図3に示す。

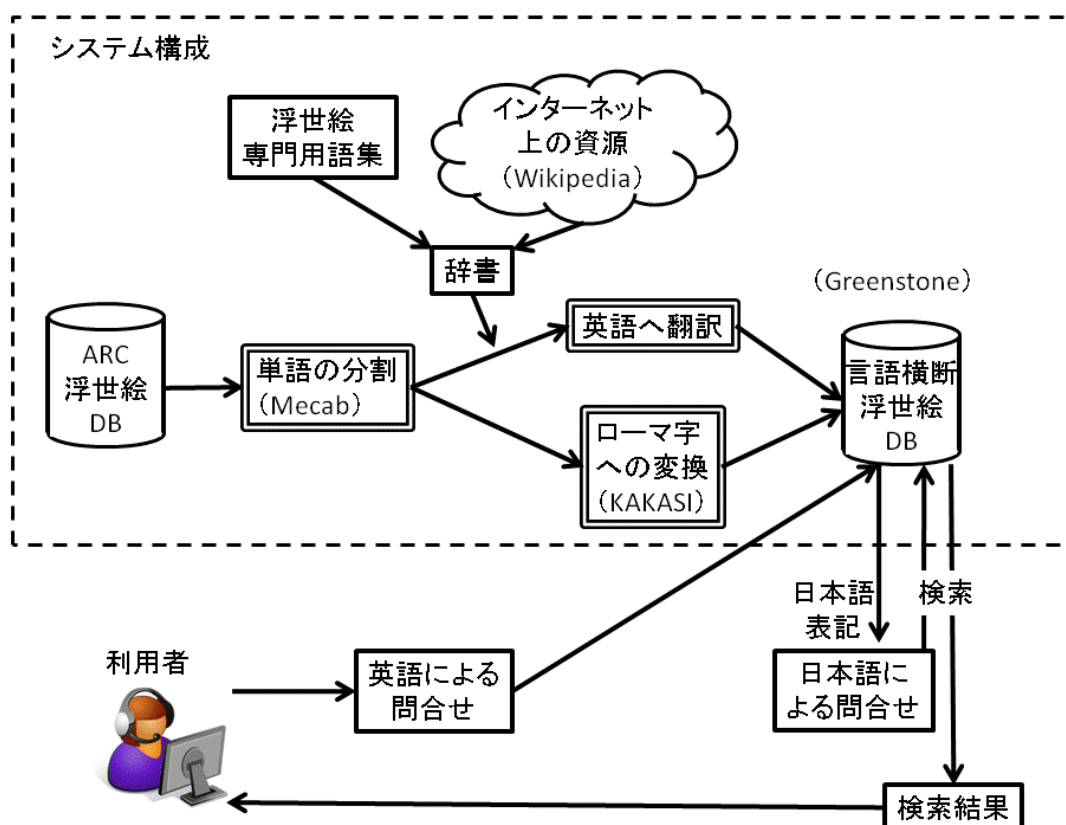


図3. システムの構成および英語の問合せによる検索の流れ

3.1 メタデータ属性のメタデータへのマッピング

立命館大学アート・リサーチセンターの浮世絵検索閲覧システムで用いられているデータ属性のうち、67属性は Dublin Core のメタデータ要素集合 11 メタデータにマッピングを行った。このメタデータのマッピングは、人手により行った。また、Dublin Core の既存のメタデータにマッピングできなかったメタデータ属性は、拡張メタデータとして割り当てた。メタデータ属性を標準的なメタデータである Dublin Core などのメタデータにマッピングを行うことにより、一度に複数のデータベースに検索を行う（横断検索）際に、データ属性名の差異を吸収することが容易となる。他の浮世絵データベースにおいても、自身以外のデータベースのメタデータ属性がどのように構成されているかについて考慮する必要が無く、自身と Dublin Core とのマッピングだけについて考慮すればよいからである。

また、浮世絵検索閲覧システム固有のメタデータ属性に関しては、新たなメタデータ要素を作成した。新たに作成したメタデータ項目は、"readingRomaji (ローマ字読み)", "readingHiragana (平仮名読み)", "JapaneseDate (和暦)" などである。"readingRomaji", "readingHiragana" は、3.3 節で述べる、英語（ローマ字）での問合せによる検索の際に用いる。また、"JapaneseDate" は日本独自の年代表記法であるため、新たなメタデータ要素として作成している。

3.2 メタデータの翻訳・翻字

本システムでは、全てのメタデータに対し、辞書により翻訳できる単語は辞書による翻訳を行い、それが出来ない単語については、ローマ字化することにより翻字を行う。これにより言語横断アクセスを実現する。

まず、各メタデータを辞書により翻訳を行うが、そのためにはメタデータを単語に分割する必要がある。一般に、日本語の単語分割を行うには形態素解析を行う。本論文では、Mecab を形態素解析器として用いた。分割されたメタデータの単語を辞書とのマッチングを行い、その単語の訳語を獲得する。もしこの段階で全ての単語が翻訳された場合、そのメタデータの翻訳は完了する。本システムでは、メタデータの翻訳の際に用いる辞書に、浮世絵に関する用語を追加している。本論文において対象とする浮世絵に関するデータは、江戸時代の人物名や浮世絵固有の専門用語などが頻出する。これらの単語は一般の辞書に存在しないことも多いためである。追加した単語は、立命館大学文学部において編纂された用語集および、Wikipedia などの Web 上の情報を用いて収集した。

辞書により全ての翻訳で全ての単語が翻訳できなかったメタデータは、残りの単語のローマ字化を行う。元のデータで、平仮名および片仮名表記されている箇所は、そのままローマ字表記を行うことが可能である。しかし元のデータで漢字表記されている箇所については、まずその漢字の読みを特定する必要がある。そこで、漢字→かな(ローマ字)変換プログラムである KAKASI(kanji kana simple inverter)を用いて漢字の読みを特定し、ローマ字表記に変換する。この時得られる元データに対するローマ字表記は、"readingRomaji" 属

性のデータとしてデータベースに追加する。こうして日本語表記とローマ字表記の対応表を作成する。この情報は、浮世絵の情報をローマ字表記で提示する際だけでなく、英語（ローマ字）の問合せで検索を行う際にも用いる。

立命館大学アート・リサーチセンターの浮世絵検索閲覧システムでは、「絵師」や「上演場所」などのように固有名詞であることがほとんどである。そのため、その固有名詞の読みをそのままローマ字表記にすることで多くの場合は英訳できていると言える。それゆえ、本システムでは、「絵師」や「上演場所」などの各データを、ローマ字で表記することにより、海外の浮世絵データベースとの横断検索を行うことを可能とする。

3.3 英語（ローマ字）での問合せによる検索

本システムでは、3.2 節において浮世絵データベースの横断検索を行うためにメタデータの翻訳・翻字を行ったが、これにより浮世絵の検索の際に日本語または英語（ローマ字）を問合せとして用いることも可能となる。本節においては、英語を問合せ語に用いる場合について述べる。

まず、利用者は検索インターフェースの”Search”ボタンを押し（図 4-①）、問合せ入力画面に切り替える。次に、利用者は問合せを英語（ローマ字）で入力し（図 4-②）、”Begin Search”ボタンを押し、検索を行う（図 4-③）。

システムは、利用者から入力された英語による問合せを、日本語に翻訳を行う。その際、3.2 節で述べた”readingRomaji”属性のデータと英語の問合せとのマッチングを行い、事前に作成していた日本語表記とローマ字表記の対応表からそのローマ字表記に対応する日本語表記を獲得する。例えば、利用者から入力された英語の問合せが”Kuniyoshi”であった場合、”Kuniyoshi”という問合せ語に対して”readingRomaji”属性のデータとのマッチングを行い、「国芳」という日本語表記を獲得する。これを浮世絵データベースに対して検索を行う際の問合せ語とする。その後、この問合せにより浮世絵データベースの検索を行い、利用者に提示する（図 4-④）。

提示された検索結果から、利用者は閲覧したい浮世絵を選択することにより、その浮世絵に関する情報を閲覧することができる（図 4-⑤）。利用者が選択した浮世絵の情報を提示する際、各データは元々の日本語表記で提示するだけでなく、事前に生成されたそれぞれの日本語表記に対応するローマ字表記も合わせて提示する。こうすることにより、提示された情報も英語化されていることになり、日本語が理解できない利用者にも提示された情報を読むことが可能となっている。

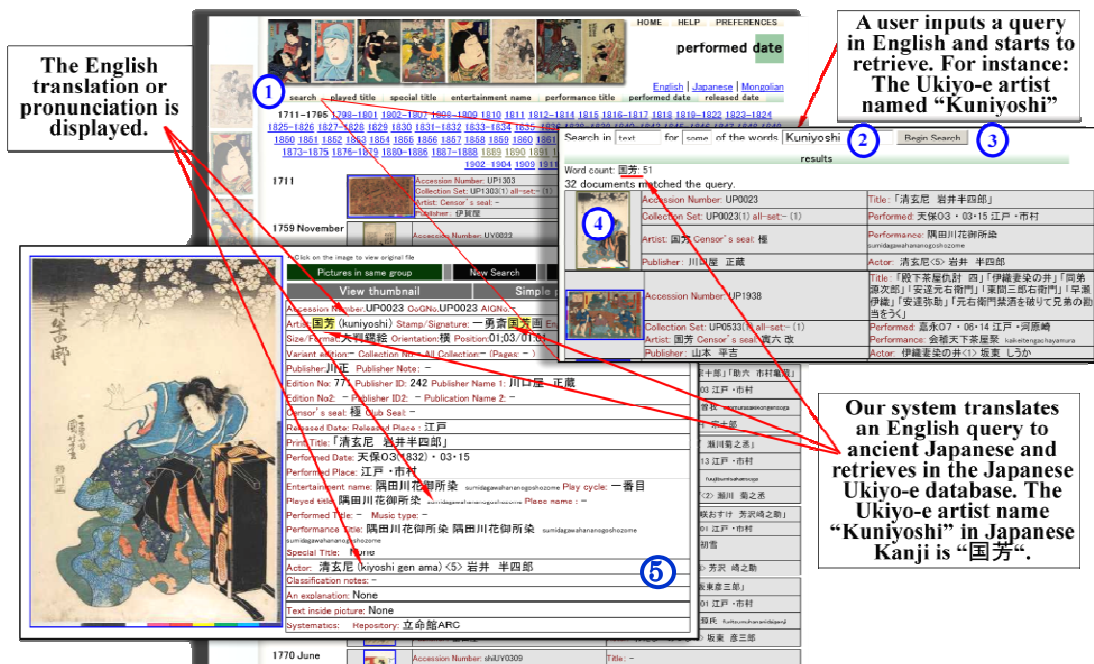


図 4. 英語での問合せによる検索

4. 考察

本論文では浮世絵データベースに格納されているメタデータを翻訳およびローマ字へ翻訳を行うことにより浮世絵データベースの横断検索の手法の提案を行った。また、それにより、英語による問合せを用いた浮世絵の検索を可能とし、提示される浮世絵の情報もローマ字化することを実現した。浮世絵データベースでは、絵師の名前や上演場所など、固有名詞の情報が多いため、ローマ字表記を利用した本手法が有効である。しかし、本手法では人物名などの読みを得る際に必ずしも正しい読みが得られるとは限らないため、適切なローマ字化が行えないこともある。

浮世絵の専門用語や固有名詞以外の情報に関しては、ローマ字表記を用いるより、適切な訳語に翻訳するほうがよい場合もある。現在のところ、辞書を用いることで翻訳を行っているが、適切な訳語が得られないこともある。この原因として、訳語の曖昧性解消が行えていないことが挙げられる。浮世絵の専門用語や絵師の名前などの固有名詞ではなく、一般語を問合せ語として用いることも考えられる。そのため、より柔軟な多言語検索を実現するには、そのような問合せにも対応することが必要である。

5. おわりに

本論文では、浮世絵検索閲覧システムの多言語アクセス手法の提案を行い、その実装を行った。また、複数データベースの横断検索のためのメタデータのマッピングを行った。さらに、浮世絵データベースのメタデータを英語に翻訳およびローマ字化を行った。これ

により、日本語が理解できない利用者でも、日本の浮世絵検索閲覧システムを検索し、閲覧することが可能となった。

今後の課題として、より柔軟な検索を行うために、一般語にも対応できる検索システムに改良することが挙げられる。また、提案システムの有効性の評価実験を行うことも課題としてあげられる。さらに、日本および海外で公開されている複数の浮世絵データベースの横断検索を実現することも課題として挙げられる。

謝辞

本研究の一部は文部科学省グローバル COE プログラム「日本文化デジタル・ヒューマニティーズ拠点」、文部科学省私立大学戦略的研究基盤形成支援事業「芸術・文化分野の資料デジタル化と活用を軸とした研究資源共有化研究」、文部科学省科学研究費補助金若手研究(B)「言語・時代・文化横断型の情報アクセスに関する研究」(研究代表者:前田亮, 課題番号:21700271) の支援を受けている。

参考文献

- [1] 早稲田大学坪内博士記念演劇博物館 演劇博物館所蔵浮世絵閲覧システム
(URL) <http://www.enpaku.waseda.ac.jp/db/enpakunishik/>
- [2] 立命館大学アート・リサーチセンター, ARC 所蔵浮世絵検索閲覧システム
(URL) <http://www.dh-jac.net/db/arcnishikie/default.htm>
- [3] Minneapolis Institute of Arts, The ART of Asia - EXPLORE THE COLLECTION
(URL) <http://www.artsmia.org/art-of-asia/explore/>
- [4] ハルタルフー・ガルマーバザル, 前田 亮. 伝統的モンゴル文字文書のデジタル図書に関する研究. 紀要『アート・リサーチ』, Vol. 8, pp. 33-43, (2008)
- [5] 木村 文則, 小傘礼 雅之, 前田 亮, 佐古 愛己, 杉橋 隆夫. 古典史料データベース検索システムの提案. 情報処理学会研究報告, 2008-CH-78, pp. 45-52, (2008)
- [6] Koolen, M., Adriaans, F., Kamps, J., Rijke, M.: A Cross-Language Approach to Historic Document Retrieval. In: 28th European Conference on IR Research: Advances in Information Retrieval, LNCS, vol. 3936, pp. 407-419. Springer (2006)
- [7] Pilz, T., Ernst-Gerlach, A., Kempken, S., Rayson, P., Archer, D.: The Identification of Spelling Variants in English and German Historical Texts: Manual or Automatic. Literary and Linguistic Computing, Vol. 23, No. 1, pp. 65-72, ALLC, ACH (2008)
- [8] University of Waikato, Greenstone Digital Library Software
(URL) <http://www.greenstone.org/>