

メタデータのブラウジングシステムの構築

○阪口哲男*, 嶋田恭子†, 沼尻花名†, 田畑孝一*

Building a Browsing System of Metadata

○ Tetsuo Sakaguchi*, Kyoko Shimada†, Kana Numajiri†, Koichi Tabata*

Abstract

The Library of the University of Library and Information Science launched its digital library service (ULIS-DL) in February 1999. The major functions of ULIS-DL are to build and provide a collection of metadata of network resources for libraries and library and information science. This paper describes building a browsing system of metadata. The system provides a list of words included in the metadata. Users choose one or two words from the list and get a subset of the metadata which include the words. It also provides a list of the creator element of the metadata and a list of Japanese noun phrase which includes chosen word. Users easily get metadata for accessing network resources with this system. The system is available on the Internet.

1 はじめに

インターネットと World Wide Web (WWW) の普及に伴い、ネットワークを通じてアクセス可能な情報資源が急増している。その膨大な情報資源の中から利用者が目的に合うものを捜し出すことについて、様々な研究・開発が行われている。情報資源をアクセスする手がかりとして提案されたものにメタデータがある。本研究では、情報資源のアクセスをより容易にするために、収集・蓄積されたメタデータをブラウジングするシステムを構築する。

2 メタデータを用いた情報資源のアクセス

メタデータとは情報資源の二次情報で、その情報資源がどこから入手可能でどんなものであるのかという情報を記述したものである。メタデータを収集、蓄積、提供することによって、利用者が情報資源をアクセスする手がかりとすることができる。

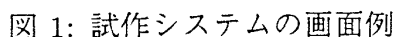
2.1 図書館情報大学デジタル図書館におけるメタデータ

図書館情報大学では、1999年2月よりデジタル図書館システム (ULIS-DL) のサービスを行っている (URL: <http://lib.ulis.ac.jp/>)[1]。ULIS-DL の中心的な機能は、図書館情報学に関する情報資源のメタデータを作成・蓄積し、提供することである。

ULIS-DL のメタデータは、Dublin Core[2] の基本 15 エレメントに出版国と文字コード、そして管理用の ID を追加した計 18 エレメントから成っている。各要素には「読み」を表すサブエレメントも加えられている。インターネット上の情報資源については、自動収集ソフトウェア (ロボット) が収集し、タイトルなどの情報を自動抽出したものを参考にしてメタデータを作成している。本稿執筆時点で約 2 万件のメタデータが蓄積されている。

2.2 メタデータの検索とブラウジング

ULIS-DL にはメタデータの検索機能が備わっている。これは利用者が入力した語を含むメタデータを全文検索し、その結果を呈示するものである。この場合、利用者が選んだ語によっては結果が膨大な数となることや、結果が全く得られないようなこともあり得る。その場合、利用者はもう一度語を選定することになる。場合によっては適切な語を選ぶまでに何度も試行錯誤することもあり得る。



— 66 —

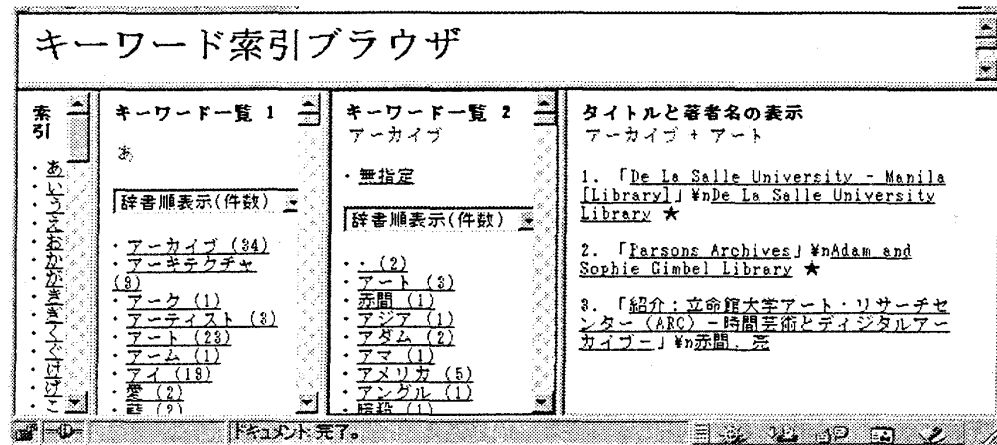


図 2: キーワード一覧からのブラウジングの画面例

データにのみ含まれる語である。これにより 2 語の組合せ指定を可能にする。なお、1 語のみ指定する際は「無指定」を選ぶ。

- 各キーワード一覧において、それぞれの語を含むメタデータの件数を添えて表示する。
- 各キーワード一覧において、語の表示順序を五十音順と、メタデータ件数順に切替えて表示することができる。

3.2 著者名一覧からのブラウジング

人名や組織名と一般的な分野や主題を表す語は性質が異なるので、別に扱う方が良いと考えられる。本システムではメタデータの著者名エレメントを著者名一覧として表示し、選択された著者名を持つメタデータより抽出した語の一覧を選ぶという機能を備える。

著者名は語に分割せずエレメントに含まれる値そのものを一覧する。形態素解析により抽出した語を選びたい場合は、キーワード一覧からのブラウジング機能を用いることができる。

3.3 名詞句一覧からのブラウジング

前述までの機能では語と語の相互関係やそれを含む文脈は問われない。そこで日本語について名詞と隣接する語から一種の句が構成されると考え、文脈に関する情報を提供するためにこれを利用者に呈示する。この句を名詞句と定義する。形態素解析結果に品詞情報が含まれるので、その以下のパターンをもって名詞句とした。なお、これらの品詞分類名称は形態素解析システム「茶筌」[3] に準拠している。

- 名詞＋名詞（未定義語でないもの）
- 名詞＋助詞「の、と、や」＋名詞
- 名詞＋助詞「の」＋名詞「ため」＋助詞「の」＋名詞
- 名詞＋助詞「に」＋動詞基本形「関する、よる、おける」＋名詞
- 名詞＋動詞連用形
- 名詞＋助詞「の」＋形容詞語幹＋助詞（接尾辞）「さ」
- 形容詞連用形＋名詞
- 形容動詞連体形＋名詞

これらのパターンに一致する名詞句を図 3 のように選択された語を中心として KWIC 索引のように表示し、利用者が選択した句を含むメタデータの一覧を呈示する。

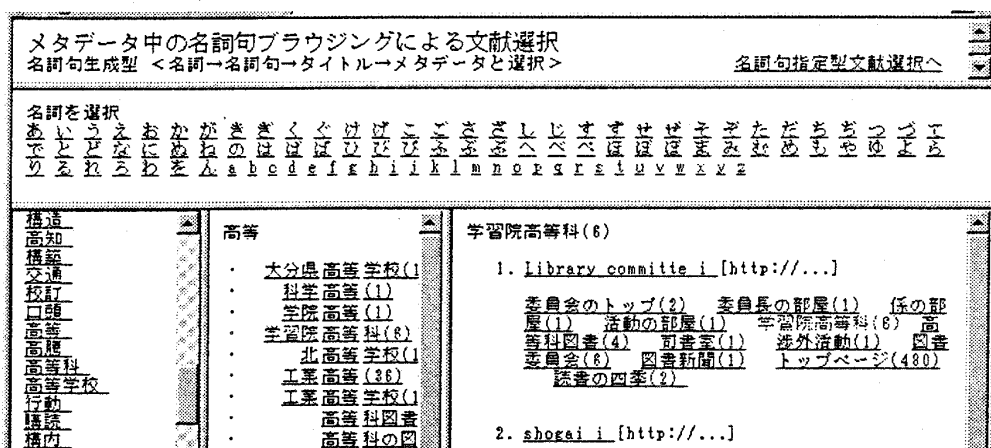


図 3: 名詞句一覧からのブラウジングの画面例

4 実現手法

ULIS-DL ではインターネットを通じて利用することを想定しており、WWW の仕組みをその基盤として使用している。そのため、本研究においても同様に WWW の仕組みを用いている。

本システムは Perl Ver.5 を用い、件数表示を行うため、検索性インデックスは Perl に組み込みのハッシュ機能により独自に作成した。開発はワークステーション HP9000 (HP-UX) 上で行い、日本語形態素解析システムに茶筌 Ver.1.51 を、WWW サーバに Netscape Commerce Server を用いている。ULIS-DL のメタデータには「読み」を表すサブエレメントがあるが、語との対応が明示されていないため、語を一覧する際には茶筌により付与された読み情報を用いた。一方、著者名一覧に関してはサブエレメントの読みを利用している。

5 おわりに

ULIS-DL におけるメタデータのブラウジングシステムとその実現について述べた。本システムでは検索語の選定に関して、ブラウジングすることで利用者が熟練していなくても容易に情報資源をアクセスすることができる。今後は蓄積されているメタデータの件数の増加に伴って、さらに機能向上の検討を進める必要があると思われる。なお、本システムのキーワード一覧と著者名一覧の機能については現在 ULIS-DL においてインターネット上で利用可能である。

参考文献

- [1] 平岡博, 真中孝行, 横山敏秋, 阪口哲男, 杉本重雄, 田畑孝一. 図書館情報大学デジタル図書館システム. 情報管理, Vol.42, No.6. p.471-479. 1999.
- [2] Dublin Core 参照記述. (URL: <http://www.DL.ulis.ac.jp/DC/>)
- [3] 茶筌ホームページ『茶室』. (URL: <http://cl.aist-nara.ac.jp/lab/nlt/chasen/>)

* 図書館情報大学 University of Library and Information Science

† 同上 (現在、会社員)