



「専門用語コレ・モカモ」をネタにした 図書館ツール開発のお話

平成23年6月22日

図書系職員のためのアプリケーション開発講習会

専門用語コレ・モカモ

このツールについて

東京大学学内の各学部・研究所の図書館室ではその専門分野の学術図書を選定し蓄積することで、長年かけてコレクションを作ってきました。そのコレクションの情報を元に、個々の学術分野に特化した日本語の関連語提示を行います。

① フレーズ入力

○日本語の文章かフレーズを入力
遺伝

② 学術文分野 (図書館) セレクト

○東京大学の図書館室を指定 農学生命図書館

連想検索 リセット

③ クリック

遺伝

KeyWord	Score	WebCat
遺伝	5.002700e+00	BOOK
学	1.022370e+00	BOOK
育種	4.351482e-01	BOOK
実験	3.983614e-01	BOOK
資源	3.920508e-01	BOOK

Webcat Plus Minus にリンク

クリックした語でさらに連想

<https://mbc.dl.itc.u-tokyo.ac.jp/Colle-mokamo/> にて実験サービス中

専門用語コレ・モカモの3つのねらい

サービス

- ・ 研究・学習者向けの検索補助

研究

- ・ 関連語提示手法のひとつの提案

図書館業界

- ・ 図書館の蔵書構築をフォーカス





制約を逆にヒントにする

自作の関連語提示フレームワークを、OPAC
検索ログから大学の学部図書館室の蔵書コレ
クションに換えただけ

外的要因が
理由として
大きい...

検索ログ利用の利用許諾が簡単ではない

東京大学OPACにはすでに東京大学蔵書
を元にしたキーワード補完機能がある

図書館システムリプレイスにより、図書館職
員が図書館所蔵データを入手可に

本質的に重要なのは、ドメインに特化したコーパスを使うこと

検索ログとコレクション情報の違いを考える

検索ログ	コレクション情報
利用の実態に即している	図書館の図書選定力が問われる
利用者のニーズを拾いやすい	
重みづけが働きやすい	各図書館はすべて等価



基本はWEBCAT PLUSの連想ワード

The screenshot shows the Webcat Plus website interface. The browser window title is "Webcat Plus - Dell により提供された Internet Explorer". The address bar shows "http://webcatplus.nii.ac.jp/#/130ad2d0c03". The search bar contains "webcat". The main content area shows "文章から連想" (Associations from articles) with a search box containing "医学" (Medicine) and a button "この文章で連想する" (Associate with this article). On the right side, there is a section titled "連想ワードで絞込み" (Filter by association words) containing a list of related terms:

- 歯学 (Dentistry)
- 医術 (Medicine)
- 歯科 (Dentistry)
- 歯科学 (Dentistry)
- 医学 (Medicine)
- 美容 (Cosmetics)
- 鍼灸 (Acupuncture)
- 鍼灸 (Acupuncture)

Below this list, there is a section titled "人物" (People) containing:

- 浅野伍朗 (Asano Gorō)
- 京大 医研 (Kyoto University Medical Research)
- 北京中医学院 (Beijing University of Traditional Chinese Medicine)
- 中医研究院 鍼灸研究所 (Institute of Traditional Chinese Medicine, Acupuncture Research Institute)

An orange callout bubble points to the list with the text: 書誌情報から得た「連想ワード」 (Association words obtained from bibliographic information).

と

言選WEBによる解消

GETAの関連語は形態素
(語の最小単位)で示される

しかし!

○ 専門用語は複合語となることが多い

- 言選Web(専門用語自動抽出システム)をカスタマイズし、複合語と形態素の双方を出力する
- 複合語の重みづけを強くする

インターフェイスをみる(1)

～機能を絞る～

このツールについて

東京大学学内の各学部・研究所の図書館室ではその専門分野の学術図書を選定し蓄積することで、長年かけてコレクションを作ってきました。そのコレクションの情報を元に、個々の学術分野に特化した日本語の関連語提示を行います。

○日本語の文章かフレーズを入力

○東京大学の図書館室を指定 経済学部図書館

(ここに連想した専門用語を表示します)

連想検索のパラメータ指定も可能にできるが、あえてオミット

シンプルな操作

このツールについて

東京大学学内の各学部・研究所の図書館室ではその専門分野の学術図書を選定し蓄積することで、長年かけてコレクションを作ってきました。そのコレクションの情報を元に、個々の学術分野に特化した日本語の関連語提示を行います。

日本語の文章かフレーズを入力
遺伝

東京大学の図書館室を指定 農学生命図書館

連想検索 リセット

KeyWord	Score	WebCat
遺伝	5.002700e+00	BOOK
実験	1.022370e+00	BOOK
資源	4.351482e-01	BOOK
分子遺伝学入門	3.983614e-01	BOOK
	3.920508e-01	BOOK
	2.970712e-01	BOOK

語をクリックすると、その後でさらに連想

Webcat Plus Minus にリンク (ゼロ件ヒットなし 東大限定のツールにしない)



インターフェイスをみる(2)

～項目の並び順～

このツールについて

東京大学学内の各学部・研究所の図書館室ではその専門分野の学術図書を選定し蓄積することで、長年かけてコレクションを作ってきました。そのコレクションの情報を元に、個々の学術分野に特化した日本語の関連語提示を行います。

○日本語の文章かフレーズを入力

○東京大学の図書館室を指定

連想検索 リセット

(ここに連想した専門用語を表示)

当ツールで使用した和書タイトル数

経済学部図書館	192,9
社会科学研究所図書室	78,7
教育学部図書室	50,3
文学部図書室	5,4

経済学部図書館
経済学部図書館
社会科学研究所図書室
教育学部図書室
文学部図書室
史料編纂所図書室
理学部物理図書室
理学部化学図書室
理学部生物学図書室
理学部地球惑星図書室
工学部1号館B建築
工学部5号館化学系
工学部14号館航空
物性研究所図書室
数理科学図書室
農学生命図書館
薬学図書館
医学図書館

主観だが、
近い学術分野を
近くに配置

次回予告

- 「言選Web」を使った選書補助用キーワードリストの作成

