

5-5 Web上に存在する情報の収集支援ツールの研究開発

桃井幹夫、高瀬博人（㈱中電シーティーアイ）

1 研究の目的と概要

現在、インターネット上にはさまざまな情報が公開されており、その中には再資源化に関する有用な技術情報も多数含まれている。これらの情報は、検索サイトを利用することで見つけることは可能であるが、検索サイトで表示されるウェブページには、技術情報以外のページも多く含まれ、有用な情報にたどり着くまでに、数多くのサイトを訪れることになりがちである。また、検索結果のページを訪れても、目的としている情報がどこにあるのか探すのに手間取ることも多い。このように、インターネット上から目的とする情報を探し出すには、かなり効率が悪く、多くの労力が必要となる。

そこで本研究では、手間がかかり効率の悪いインターネット上からの技術情報検索を支援するため、再資源化技術検索に特化した収集支援ツールの開発を目指す。

2 フェーズⅠの成果

本研究はフェーズⅡから始まった課題である。

3 フェーズⅡの成果

3-1 目的及び目標

再資源化技術データベースの充実を図るため、情報検索と情報更新、データベース登録に特化した再資源化技術データベースのための情報検索支援ツールの開発を行う。具体的には、Yahoo等の検索サイトを利用して、インターネット上のホームページのキーワード検索を行い、その結果をデータベースに蓄積することにより、技術情報の検索を効率良く行えるように支援する。

3-2 方法及び結果

3-2-1 システム

本システムは、インターネットにWebとして公開されている再資源化技術情報を検索することを目的としているもので、アプリケーション内で情報を入力し、インターネットにアクセス、結果を表示し、必要な情報をマウスでコピー&ペーストし、データベースに情報を登録していく。主な機能としては、以下があり、散在する再資源化技術データを収集する際に役立つシステムとなる。

- ・ YAHOO 検索サイトを利用し、ホームページ掲載データをキーワード入力により探索
- ・ 検索結果内の重要単語マーキング表示
- ・ データベースによる検索結果（URL）の管理
- ・ 再資源化技術データベースへのオンライン登録

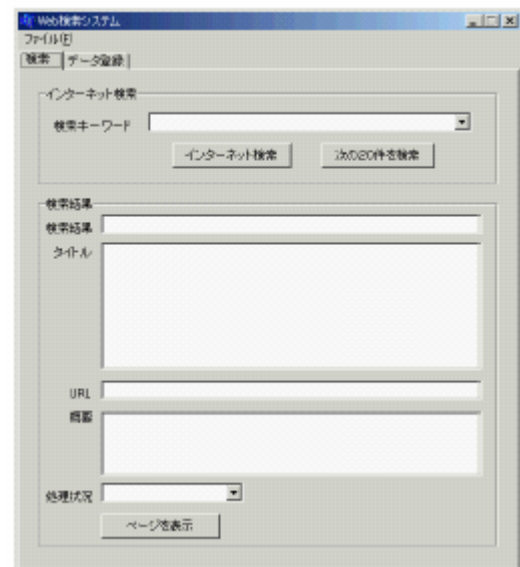


Figure 1 検索画面

- ・ プロキシサーバ対応

3-2-2 機能

■ キーワード検索

キーワード欄にキーワードを入力し検索を行う。過去に検索したキーワードは、自動的にプルダウンメニューに登録される。検索結果は一度に20件表示し、20件以上ある場合はNEXTボタンにより、次の20件を表示する。

■ マーキング表示

ページを表示する際、キーワードを色付けによってマーキングすることにより、ページ内で記述されている位置を識別しやすくする。検索結果を表示の際、「マーキング実行」ボタンをクリックすると、通常、検索キーワードが登録されていますので、現在表示しているWeb ページを取り込みマーキングを行う。キーワードを追加する場合は、キーワード欄にマーキングしたいキーワードを入力し、色を選んで「追加」ボタンを押す。すると、キーワードリストにキーワードと色が表示されるので、入力終了後に「マーキング実行」ボタンを再度押す。(Figure 2)

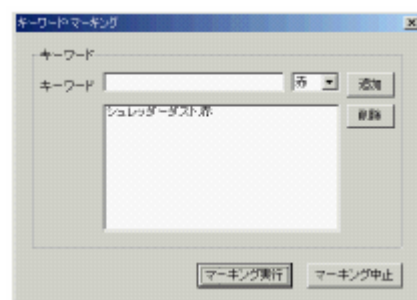


Figure 2 マーキング追加

■ URL 管理

検索サイトで検索した結果およびページの登録状況をデータベースで管理することにより、再度検索した場合のページの重複を防止、登録状況などを管理することができる。

■ オンライン登録

ページを表示しながら、オンラインでデータベースに登録することが可能。

■ プロキシサーバ対応

プロキシサーバ経由でインターネットに接続する場合、プロキシサーバの IP アドレスおよびポートの設定 (Figure 3) を行うことにより、接続可能にしている。

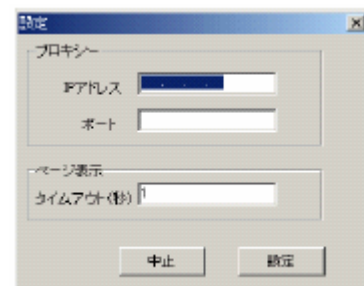


Figure 3 プロキシ設定

3-2-3 基本操作

タイトルリストからタイトルの1つをマウスで選択する (Figure 4)。すると、概要欄にそのページの概要が表示される。概要だけで、ページの内容がよく理解できない場合は、「ページを表示」ボタンを押して、該当ページを表示する (Figure 5)。この時表示されるページでは、検索に使用したキーワードが赤色でマーキングされている。ページの内容を調べ、有用な情報が含まれており技術データベースに登録すると決定した場合は、処理状況を「登録する」に変更。有用な情報が含まれておらず、今後このページは表示する必要がないと判断した場合には、処理状況を「登録しない」に変更する。タイトルリストに表示されたページについて、順番に上記の作業を繰り返す。



Figure 4 操作画面



Figure 5 該当ページ表示

3-2-4 URL テーブル

URL テーブルは検索サイトで検索した URL 情報を管理するためのテーブルで、テーブルの内容を Table 1 および Table 2 に示す。

Table 1 URL テーブル基本構成

| | |
|----------|--------------|
| テーブル名 | URLLIST |
| スキーマ | DBTEQ |
| Table 領域 | USERS |
| プライマリーキー | URL, KEYWORD |

Table 2 URL データベーステーブル

| | 項目 | データベース列名 | データ型 | サイズ |
|---|----------|----------|----------|------|
| 1 | URL | URL | VARCHAR2 | 255 |
| 2 | 検索キーワード | KEYWORD | VARCHAR2 | 255 |
| 3 | ページタイトル | TITLE | VARCHAR2 | 255 |
| 4 | ページ概要 | ABST | VARCHAR2 | 1024 |
| 5 | 処理状況 | STATE | NUMBER | 1 |
| 6 | 再資源技術 ID | SHIGENID | NUMBER | 10 |

3-2-5 データベースとの照合

検索ウィンドウのキーワード欄に、検索したい技術情報に関連するキーワードを入力し検索する。検索結果は、ウィンドウの下側に表示される。この時、以前に検索したキーワードの場合、URL データベース (Table 2) と照合し、「登録の必要なし」と判断されたページおよび「登録済み」のページについては除外される。

次に、検索結果から有用な技術情報が含まれているページを取捨選択します。リストアップされたページの中から有用な技術情報が含まれているか否かによって、ページの処理状況を変更する。(Table 3)

Table 3 ページの処理状況

| | 状況 | 説明 |
|---|-------|--------------------------------------|
| 未 | 未処理 | まだ、有用か否か判断されていない |
| 登 | 登録予定 | 有用な情報が含まれているため、データベースに登録すると判断された |
| 済 | 登録済み | 既に登録したページ |
| × | 登録しない | 有用な情報が含まれていないためデータベースに登録しないと判断されたページ |

3-2-6 オンライン登録ウィンドウの入力とデータの登録

データベースに登録する情報が見つかった場合は、「データベース登録」ウィンドウの各項目欄に入力する。コピー&ペーストする場合は、一度 Web ページを「テキストに変換」ボタンを押してテキスト形式に変換し、「Web ページ表示」ウィンドウの「テキスト形式」ページを選び (Figure 6)、選択してコピー&ペーストする。登録ウィンドウ上にある各項目に情報を入力し、入力が完了した時点で登録ボタンを押すと、技術情報がデータベースに登録される。



Figure 6 検索結果テキスト表示画面

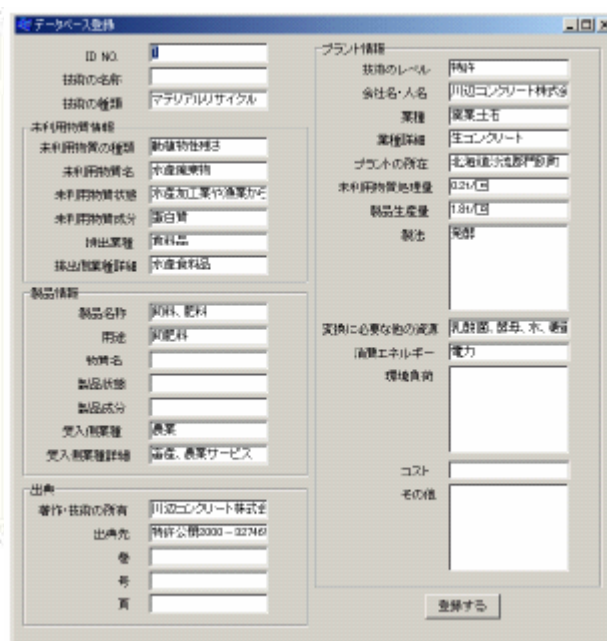


Figure 7 データベース登録画面

3-2-7 検索結果

検索結果は、タイトルリストにページのタイトルがリストとして表示され、各ページの URL や概要はタイトルを選択すると、URL 欄および概要欄に表示されます。過去に検索したキーワードの場合、同じ URL で処理状況が「登録済み」あるいは「登録しない」になっているページはタイトルリストに表示されません。これにより、同じページを何度も開かず済み効率的に技術情報を検索することが可能になる。

Table 4 データベースデータ詳細項目

| | |
|------|-----------------------|
| 検索結果 | 検索結果全件数および表示件数 |
| タイトル | 検索結果のページタイトルをリスト表示 |
| URL | タイトルリストで選択されたタイトルのURL |
| 概要 | タイトルリストで選択されたタイトルの概要 |
| 処理状況 | ページの処理状況 |

3-2-8 文字コード指定

通常ホームページには、そのページを記述する文字コード情報が含まれているが、一部のページで文字コード情報が含まれておらず、マーキング機能を利用した場合などに正しく表示できない場合がある。このような場合は、文字コードを指定してマーキングを実行することで正しく表示される。変更手順は、まず文字コードを自動から EUC あるいは Shift-JIS に変更してからマーキングを再度実行する。

3-2-9 データベース

検索結果を登録するデータベースの項目は以下のようにになっている (Figure 9)。この項目は、再資源化技術データベースのものと同じである。



Figure 8 検索結果データベース登録

Table 5 再資源化技術データベース項目

| | 項目 | 説明 |
|---------|----------|---|
| | ID | 技術毎に一意につけられる番号です。登録プログラムがつけるので、入力する必要はありません。 |
| | 技術の名称 | 技術の名称です。 |
| | リサイクルの種類 | リサイクルの種類です。「サーマルリサイクル」、「ケミカルリサイクル」、「マテリアルリサイクル」から選んでください。 |
| 未利用物質情報 | 未利用物質の種類 | リスト2 |
| | 未利用物質名 | 未利用物質の化学物質名あるいは一般的な名称 |
| | 未利用物質状態 | 未利用物質の形状、荷姿、混合状態など |
| | 未利用物質成分 | 未利用物質の雑質な成分の組成 |
| | 排出側業種 | 未利用物質を排出する業種（複数可） |
| | 排出側業種詳細 | 未利用物質を排出する業種の詳細（複数可） |
| 製品情報 | 製品名称 | 製品の名称 |
| | 用途 | 製品の用途（複数可） |
| | 物質名 | 製品の物質名 |
| | 製品状態 | 製品の状態（ペレット、液体等） |
| | 製品成分 | 製品を構成する成分 |
| | 受入側業種 | 製品を使用する業種（複数可） |
| | 受入側業種詳細 | 製品を使用する業種の詳細（複数可） |

(Table 5 再資源化技術データベース項目の続き)

| | | |
|------|------------|---|
| | 技術のレベル | 技術の実用化レベルで「製品化・プラント化」、「開発（特許）」、「研究（学術論文）」から選択 |
| | 会社名・人名 | プラントの所有会社あるいは研究者など |
| | 業種 | プラント所有会社の業種 |
| | 業種詳細 | プラント所有会社の業種詳細 |
| | プラントの所在 | 製造プラントの所在地 |
| | 未利用物質処理量 | 1日あたりの未利用物質処理量 |
| | 製品生産量 | 1日あたりの生産量 |
| | 製法 | 製造方法 |
| | 変換に必要な他の資源 | リサイクルするために必要となる他の資源 |
| | 消費エネルギー | リサイクルするために必要なエネルギー |
| | 環境負荷 | リサイクルすることによって与える環境負荷（二酸化炭素発生量など） |
| | コスト | リサイクルに係る費用 |
| | その他 | その他追記事項 |
| 出典情報 | 著者・技術の所有 | 出典先の著者あるいは技術の所有者 |
| | 出典先（特許番号） | 出典先の名称 |
| | 巻 | 出典先の巻番号 |
| | 号 | 出典先の号番号 |
| | 頁 | 出典先のページ番号 |

3-2-10 タイムアウト

ホームページを検索し取り込む際に、システム上でタイムアウトする場合があります。これは、Yahoo 等の検索サイトに接続したり、ホームページをローカルに取込む際にタイムアウトするまでの時間を秒で指定する。しかしあまり、大きくすると表示まで待たされることが多くなる。逆に、インターネットへの接続回線が細い場合にあまり短くすると頻繁にタイムアウトするケースが出てくるので、2 秒から 10 秒ぐらいの範囲で環境に合わせて設定する。

3-3 考察及びフェーズIIのまとめ

再資源化技術データベースを充実させるため、Web 上に散在する技術データを効率よく収集するためのツールを完成させた。

4 成果の達成度

当初の予定通りシステムは完成した。ただ、まだ完全に使いこなしている訳ではなく、プログラムバグも潜んでいる可能性はあるため、今後とも使い込みつつ随時修正していきたい。

5 今後の課題

今後の課題として、以下のものがあげられる。

- ・ 技術データベース項目変更に伴う、登録システム変更