

# Ajax を用いた諺データベース検索支援システムの開発

米澤 俊昭\*、倉田 孝之\*\*、前田 桂子\*\*\*

## Development of Retrieval Support System for Proverb Databases using Ajax

Toshiaki YONEZAWA \*, Takayuki KURATA \*\*, Keiko MAEDA \*\*\*

**Abstract:** In this research, a prototype of proverb database retrieval support system has been built. Book-type proverb dictionary has difficulty in management and inconvenience when carrying around. In addition, there is almost no Web application which deals with proverbs edited in a large scale. In these circumstances, it is considered that portable proverb database edited into universal contents is useful. However, one of the problems is that users tend to search proverbs without accurate memory and accurate spell typing because of their complexity. So in this research, using PHP and MySQL of the server side technology and Ajax of client-side technology, this system has been developed. Consequently, the developed retrieval support system enabled the users to make input easier than before, and was able to improve the convenience of the proverb database.

**Key words:** Ajax, relational database, MySQL, PHP, proverb

### 1. 序論

現在、日本語を取り扱う大規模な検索システムとして国立国語研究所の日本語書き言葉コーパス<sup>[1]</sup>などがあげられる。しかしながら、諺(ことわざ)を大規模に編纂して取り扱っている電子辞典はほとんど見当たらない。また、書籍の諺辞典においては、印刷物であるが故に、管理が困難で、持ち運びが不便などといった欠点がある。さらに、諺は様々な側面から意味を捉えることができるため、編纂した者によって記述している内容に違いが生じることがある。以上のことから、持ち歩きも容易で、普遍的な内容に編纂された大規模な諺データベースが必要である。

しかし、諺には難解で複雑なものが多く、検索の際にアプリケーションの利用者が、検索対象となる語句が曖昧なまま検索を行う可能性もある。そのような場合、Web ブラウザのページを全て書き換えて検索結果を表示するサーバ側の技術のみで構築された諺データベースでは、キーワードを間違えたまま検索すると、利用者がある間違いに気づかず、検索後に再入力をする手間が発生する。

(2010年11月30日受理)

\*宇部工業高等専門学校制御情報工学科

\*\*九州工業大学大学院生命体工学研究科

(平成22年3月 専攻科生産システム工学専攻修了)

\*\*\*宇部フロンティア大学人間社会学部児童発達学科

このような諺データの特性上、諺データベースの検索システムには、ユーザにとって利便性の高い検索支援システムが必要である。そこで、本研究では、検索したい単語を全て入力した上で検索するのではなく、一語ずつ入力するたびに逐次候補を表示させる検索手法である、インクリメント検索に着目した。この手法を用いることで、曖昧な諺の検索を支援し、検索語句入力中に検索対象となる諺の候補を表示することで、キーワードの打ち間違いなどを減少できると考えた。

### 2. 概要

インクリメント検索を実現するために、Ajax と呼ばれる技術を採用した。本研究では、諺データベースにインクリメント検索で検索の支援を行うシステムを構築する。図1にシステムの概要を示す。

図に示すように、サーバ側の技術として動的にWeb ページを作製することが可能なPHP を使い、諺データを保管・運用

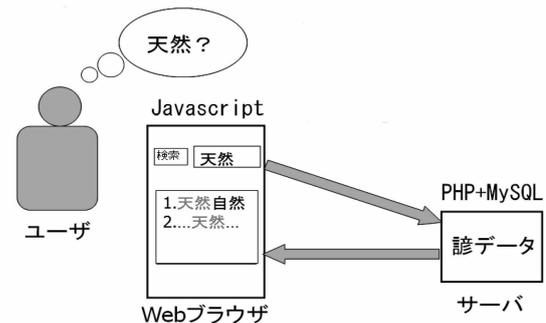


図1 検索支援システムの概要

するためのデータベース管理システムには MySQL を用いる。また、クライアント側には、Web ブラウザに搭載されている Javascript による Ajax を使用している。また、動作としては、クライアント側でユーザが文字入力するたび、Javascript 関数が立ち上がり、サーバ側の PHP と MySQL が検索処理を行うように、Javascript で指示を出す。その後、再び Javascript によって結果を取得し Web ブラウザに表示する。以上が本検索支援システムを開発した。なお、MySQL、PHP、Ajax についての詳細は 3. 基本事項で説明する。

### 3. 基本事項

基本事項では、検索支援システムに用いた MySQL、PHP のアプリケーション、Ajax の詳細について説明する。

#### 3. 1 MySQL

MySQL とは、オープンソースのリレーショナルデータベース管理システム (RDBMS : Relational DataBase Management System) のことで、リレーショナルデータベース (RDB : Relational DataBase) の管理・運用を目的として開発された実装の一つである。さらに、RDB とは、1 件のデータを複数の項目 (カラム、フィールド) の集合として表現し、データの集合をテーブルと呼ばれる表で表す方式に基づいて設計されたデータベースのことである。例として表 1 にスーパーの商品の価格と在庫を RDB のテーブルで表す。表 1 のように、データベース内はテーブルと呼ばれる表構造になっており、テーブルの各列をフィールドという。したがって、表 1 では商品名、価格、在庫が各フィールドの示すもので、テーブルの各行は、1 件 1 件のデータそのものを表すレコードという<sup>[2]</sup>。

データベースには RDB のほかに階層型データベースや網型データベースなどが存在するが、RDB は他のデータベースと比較して以下の 5 つの利点を持つ。

- ・ 保存可能なデータ容量が大きい。
- ・ データ管理が厳密に行なわれる。
- ・ 検索や並び替え、抽出、集計などの機能が高度かつ高速である。
- ・ 複数のデータを同時に取り扱える。
- ・ SQL 文を使ってデータを操作することが可能である。

使用する諺のデータ量は何万件にも上る大量のデータの集合であり、保存可能なデータ容量の大きい方がデータベースとして適している。また、後述するように一件の諺データは

表 1 RDB のテーブル例

ID	商品名	価格	在庫
1	キャベツ	¥100	20
2	パブリカ	¥100	30
3	ローライト	¥290	5

14 項目の記述項目から構成され、階層型データベースや網型データベースでは 1 件のデータが長くなり、管理・運用が難しい。さらに、諺データベースは一度設計すれば更新を殆ど行う必要が無い。これらの要素は、諺データを管理するために有用であり、RDB の利点と合致している。

#### 3. 2 PHP

PHP (Hypertext Preprocessor) とは、オープンソースのスク립ト言語である。諺データベースを作成するには、出力結果がアプリケーションの利用者の処理要求に対して、動的に生成されなければならない。したがって、動的な Web アプリケーションの作成に適している言語が諺データベースと親和性の高い言語となる。使用するプログラミング言語としては、Perl もしくは PHP などを使用するのが一般的であるが、Perl 自体が古くかつ仕様が複雑化しているため、ここでは PHP を用いる。なお、PHP は Apache HTTP Server のモジュールとして動作させた。これは、CGI を用いた検索プログラムを動作させるよりも、リクエストのたびに CPU に掛かる負荷が減少するという利点があるためである。

#### 3. 3 Ajax

Ajax とは Asynchronous Javascript + XML (非同期的な Javascript および XML) の頭文字を取って名付けられた Javascript をベースとした非同期通信というアプローチを指すもので、Javascript、ダイナミック HTML、XML などを中心とした技術である<sup>[3]</sup>。本研究では XML は使用せず、主に Javascript とダイナミック HTML を使用した。

Ajax の大きな特徴の一つである非同期通信は、一般的な Web アプリケーションに使用されている同期通信と、サーバとクライアントの間で通信する方法が異なる。同期通信を用いた Web アプリケーションは、検索処理の実行時にサーバと通信が行われ、HTTP リクエストが Web ページを全て書き換えて結果を表示する。

一方、Ajax を用いた Web アプリケーションは、マウスをクリックするなどのユーザのアクションに応じて、XMLHttpRequest を作製し、サーバと非同期に通信を行ってデータの検索などを行うことができる。したがって、通信はデータ単位で行うので処理の負荷も少なく、ブラウザのページを更新することなく検索結果を表示することも可能である。このような同期通信のモデルを図 2 に、非同期通信のモデルを図 3 に示す。

図 2 では、ユーザが何らかの処理をサーバに要求した場合、Web サーバはブラウザから要求を受け取り、入力データの処理を行った後に、その結果を Web ブラウザのページにそのまま反映する。このように、通常の同期通信を用いる Web アプリケーションは、アクションを一つ実行するごとに全てのデータがサーバで再処理され、ページ全体がリロードされる。したがって、サーバとクライアント間で通信を行うたび、ユ

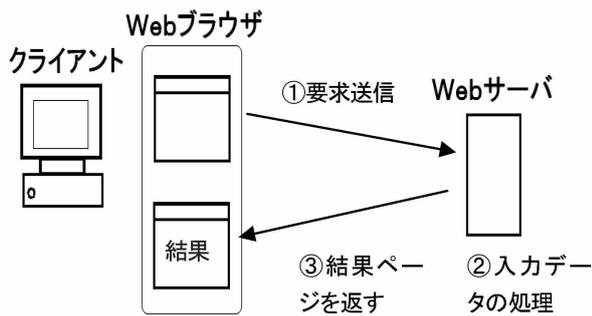


図2 同期通信の動作モデル

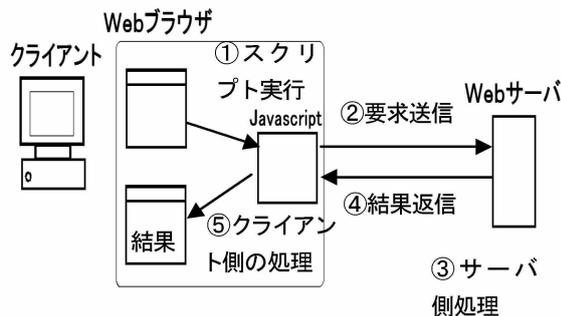


図3 非同期通信の動作モデル

ブラウザは結果の出力まで待つ必要がある。一方、図3では、ユーザがブラウザで何らかの処理を行うと、まず Javascript が実行される。Javascript はユーザがブラウザを操作した結果、発生した要求を Web サーバに送信する。受け取った要求に対する結果を再び Javascript が Web サーバから受け取り、そこからクライアント側の処理を行った後、結果をブラウザに反映する。このように、サーバへのリクエストは中間層となる Javascript を通してバックグラウンドで行われるため、従来の Web アプリケーションとは異なり、通信中であっても他の処理を行うことが出来る。

### 3.3 XMLHttpRequest

XMLHttpRequest とは、サーバに対して非同期で HTTP リクエストを発行や、テキストや XML データの取得などを行う Javascript のオブジェクトであり、Ajax による非同期通信の根幹技術である。本研究では、ユーザの入力を支援する検索支援システムを構築するにあたり、検索結果をサーバから非同期で取得する必要不可欠な要素となっている。XMLHttpRequest による非同期通信のデータ取得から表示までの流れは図4のようになる。

Ajax では Web ブラウザが行っていた HTTP 通信クライアントの役割を Javascript によって生成された XMLHttpRequest が肩代わりするため、その生成は `createXMLHttpRequest()` で行うが、使用する Web ブラウザの違いを考慮してプログラミングする必要がある。図5のように、XMLHttpRequest オブ

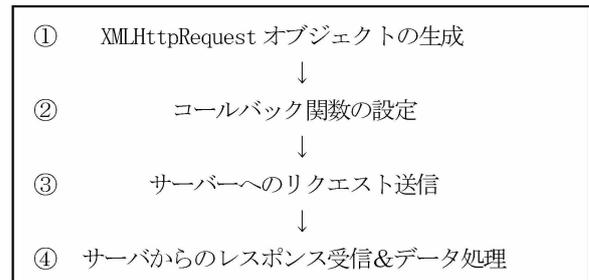


図4 XMLHttpRequest による処理の流れ

ジェクトを生成できるかどうかは if 文で判定される。XMLHttpRequest オブジェクトを利用できる Safari などの Web ブラウザであれば、XMLHttpRequest オブジェクトを生成して利用できる。XMLHttpRequest オブジェクトに未対応の IE6 などの Web ブラウザでは HTTP 通信オブジェクト生成の際に ActiveX を利用しなければならないため、ActiveXObject クラスが利用される。

このようにして生成された XMLHttpRequest オブジェクトで、さまざまなプロパティやメソッドを設定できる。

## 4. 諺のデータベース化

本研究で用いる諺データは、西九州諺研究会によって編纂された意味分類諺辞典<sup>[4]</sup>を用いる。あらかじめテキスト形式で編集された諺データを、リレーショナルデータベースのレコード形式に変換し、MySQL データベースに格納して利用する。

### 4.1 諺データの概要

諺のテキストデータの収録内容は、漢籍からの出展が 6165 句、西洋の諺が 9754 句、四字成句 11161 句、日本の諺 36148 句、その他東洋の諺 244 句、箴言を言った個人の名が特定できる箴言若干、総計は 65536 句になる。これには、諺だけに限定せず、諺に関連する章句も採録されており、同じような意味の諺は隣接して記述し、関連のある諺や反義の諺は分類番号で示している。一つの諺データは、表2に示すように全 14 項目に区切られている。また、実際にテキストデータに記述されている内容を「天然自然」、「自然天然」などを例に表2に示す。

次に表3の一行目を例にとり、表2に示す項目を説明する。まず、漢字で諺が書かれている第一項目が章句に該当し、「/」で区切られた次の項目が「てんねんしぜん」という読み方に該当し、「/」で区切られた次の項目が分類記号に該当する。このように、「/」で区切られた各項目が表5の記述項目の一つずつ該当していく。

### 4.2 諺データのテーブル格納方法

諺データをデータベースで利用するためには、テキストデータに書かれた諺を MySQL で作製したテーブルに格納する必

要がある。そのために、テキストデータの内容を insert 文に整形する。insert 文は、MySQL のテーブルにデータを格納するための SQL 文である。

原本となるテキストデータには、表 2 のとおり項目ごとに、「/」で区切られた内容が記述されている。これを変換するために、下記の手順を行う。

- ① テキストデータをエクセルで「/」ごとに分解して読み込み、諺データの内容が表 3 の一項目ごとに Sheet1 のセルに格納する。
- ② Sheet1 の内容を、Sheet2 で MySQL にデータを挿入する insert 文の形に整形し、Sheet2 の内容を MySQL 上に準備したデータテーブルへ格納する。Sheet2 には Sheet1 で分解された諺データを insert 文に自動で整形する関数を準備しておく。

### 5. 検索支援システムの詳細

クライアント側に Ajax を取り入れ、サーバ側に MySQL と PHP で連携して諺データベースに検索支援システムを実装する方法を採用した。データベース全体の構成とデータの動きを図 5 に示す。

検索指示画面はキーワードを入力して検索を指示するための PHP ファイルで構成されており、イベント処理は Javascript で記述してある。また、検索処理はサーバ側で検索するための PHP ファイルである。

次に処理の流れを図中の番号で説明する。

- ① 検索指示画面が表示される。検索指示画面にキーワードが入力されると、Javascript 関数によって XMLHttpRequest が生成される。

表 2 記述項目

1. 章句	2. 読み	3. 分類記号
4. 語種	5. 典拠	6. 句積
7. 表現	8. 関連句	9 反義語
10. 収録書	11. 出現頻度	12. 代表句
13. 略解句	14. 備考	

表 3 記述内容

天然自然/てんねんしぜん/00000/4//自然にあるが まの姿。/////1///
自然天然/しぜんてんねん/00000/4//自然ありのま まの姿/////1///
自然は最良の教師/しぜんはさいりょうのきょうし /00026/8/ドイツ///// f /1///

- ② XMLHttpRequest が生成され、半角スペース・全角スペース・カンマ (,) を区切り文字にして、検索キーワードを取得・加工した後、検索 SQL を組み立てる。
- ③ 検索 SQL に従い検索処理が実行され、検索キーワードを受け取った検索処理部が、MySQL への接続、SQL 文の発行、MySQL の切断までを行い、最後に検索結果を HTML 形式で返す。再び XMLHttpRequest を利用して検索結果を取得する関数を作成し、Web ブラウザの検索指示画面の一部分を書き換えて結果を表示する。結果が読み込まれた後、再びユーザーイベントの監視を開始する。

### 6. 実行結果

Ajax により構成した諺データベースの検索支援システムの実行結果を図 6、7 に示す。

図 6 は、キーワードを入力するためのテキストボックスに「天」と入力すると、天という文字を含むデータをサーバから取得するように検索処理を行う。検索 SQL を組み立て、天を含む文字列をサーバから探し、結果を検索画面を構成する HTML の一部分として返す。結果を受け取った検索指示画面では、検索候補として天が含まれた諺を、入力された文字をハイライトして表示する。図 7 では「天」に続いて「然」と入力した場合の検索結果を示している。ユーザーがキーワードの入力を継続して行う、もしくは文字を削除するといった操作を監視しており、継続して文字が入力された場合、再度検索 SQL を組み立てなおして表示する。

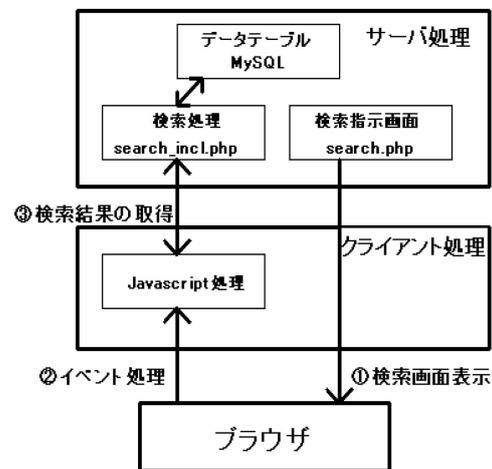


図 5 諺データベース全体の構成と処理の流れ

keyword: 因

---

**Search Result (82)**

1. 自然天然 - 自然ありのままの姿
2. 自然天然 - 自然ありのままの姿
3. 天然自然 - 自然にあるがままの姿。
4. 天然自然 - 自然にあるがままの姿。
5. 天地自然 - 大自然。「(韓詩外伝1)伝曰、喜名者必多怨、好善者必多辱、能随天地自然為能勝理」(俚諺集覽)
6. 天地自然 - 大自然。「(韓詩外伝1)伝曰、喜名者必多怨、好善者必多辱、能随天地自然為能勝理」(俚諺集覽)
7. 天人の影向 - 天人が具体的な姿をとって現れること
8. 天人の影向 - 天人が具体的な姿をとって現れること
9. 荒れ野にこそ、天地自然が保存されている - アメリカの思想家、作家。
10. 荒れ野にこそ、天地自然が保存されている -
11. 天地は万物の父母 - 天地は万物を生成するものである。「惟天地万物父母、惟人万物之靈」(書経・泰誓・上)
12. 天地は万物の父母 - 天地は万物を生成するものである。「惟天地万物父母、惟人万物之靈」(書経・泰誓・上)

図6 一字入力したときの実行結果

keyword: 天然

---

**Search Result (4)**

1. 自然天然 - 自然ありのままの姿
2. 自然天然 - 自然ありのままの姿
3. 天然自然 - 自然にあるがままの姿。
4. 天然自然 - 自然にあるがままの姿。

図7 二字入力したときの実行結果

## 7. 考察

一般的に公開されているWebアプリケーション型の諺辞典では、諺のデータ量はおよそ1000件前後である。用意した諺データの件数は65536件のうち、試験的に諺データ2000件を登録して運用した。その場合、検索支援システムは、ユーザにストレスを感じない速度でインクリメント検索を行った。

しかし、65536件の諺データベースでは、若干のタイムラグが発生した。これは、現状のプログラムではデータ数と検索文字入力数が乗算された回数だけサーバとクライアントの間で通信を行うためである。入力に応じて出力結果を表示するため、通信量が減っても通信回数が増大するため、検索時間が長くなるためであると予測される。したがって、この問題を解消するためには、検索プログラムの改善が必要である。その方法として考えられるのは、一文字目を入力した時に検索結果を表示する諺に付属するidを別に用意した配列に控えておき、二文字目を入力した時にはこの配列から検索を行うようにするなどが考えられる。

## 8. 結論

本研究では、Ajaxを用いた諺データベースの検索支援システムを開発した。一般的なWebアプリケーション型の諺辞典の倍以上のデータ数で運用される本システムは十分有用であると考えられる。また、複雑な諺の検索を容易にするために、入力中に検索候補を表示するインクリメント検索を導入したことで、部分的にしか記憶していない諺の検索が可能となった。さらに検索途中の誤字・脱字の確認も容易になった。

今後の課題として、6万件以上の諺データを全て使用した諺データベースの構築が挙げられる。その際、検索システムを運用したときに生じる結果出力までのタイムラグが問題となる。上述の改善方法だけでなく、Ajaxを用いて、検索中に負荷の少ない待機メッセージを表示するなどして、視覚的にユーザのストレスを緩和する方策も考えられる。また、現在は検索処理で参照するデータを章句と句釈の二点に絞って運用しているが、分類記号や意味などからも検索できるようにしたい。いずれにせよ、Ajaxの導入によってかなり柔軟性の増すデータベース構築が可能となる。

## 参考文献

- [1] 山崎誠・他, 代表性を有する現代日本語書き言葉コーパスの設計, 言語処理学会第12回発表論文集, pp.440-443, 2006
- [2] 星野努: PHP+MySQL で作る最速 Web システム, 技術評論社, pp.200-204, 2005
- [3] 深津貴之・増井雄一郎・川崎有亮・台場圭一・佐藤真介, Ajax 実装のための基礎テクニック, 技術評論社, pp.13-20, 2006
- [4] 前田勝己, 意味分類諺辞典分類目録, 西九州諺研究会, pp.2, 2001
- [5] 高井正三・喜田啓太・米田恭章, Ajax による古文書データベース検索語類推支援ツールの一提案, 電気関係学会北陸支部連合大会, Vol.2006, 2006