

ユーザの視点に基づく情報獲得のための知識整理インタフェース

A Knowledge Arrangement Interface for the Information Acquisition based on a User's Viewpoint

田中 大智
Taichi Tanaka

砂山 渡
Wataru Sunayama

広島市立大学 情報科学部
Faculty of Information Science, Hiroshima City University

Since there are a lot of digital resources on WWW, we can take those information into our creative activity. Useful topics and relationships among topics are buried in the information flood. In this paper, we propose a knowledge arrangement interface to support information acquisition to complement and to copyread a knowledge from WWW. We had some experiments to verify the effects of the interface.

1. はじめに

近年、コンピュータの高速化や、WWW(World Wide Web)の普及に伴い、我々がインターネットから取得できるデータ量は年々増加している。Web データベースには膨大な量の情報が存在しており、有用な情報が多くある。そのため、Web 上の情報を視覚化し情報獲得支援を行う研究が行われている。本研究では、Web から取得した複数のキーワード間の関係をユーザの興味に応じて整理し、ユーザの知識整理・補完を行うインタフェースを提案する。

2. 本研究の位置づけ

従来の情報獲得支援の主な目的は、目的情報の取得の支援、有用な情報の探索などである。

前者は、主にユーザに目的とする Web ページのイメージがある場合に用いられ、目的情報に素早く直観的に到達することを目的とする。主な方法は、検索結果の分類や、Web ページ間の関係の強さをを用いて情報の視覚化を行うことで実現される。Web ページ間の関係を維持したまま対話的に Web ページに辿り着けるようにしたものに、納豆ビュー[塩澤秀和 1997]がある。納豆ビューは、抽出したキーワードと関係が強いキーワードを繋ぐノードがリンクの強さに応じた納豆の糸となっており、目的に関連するキーワードを引っ張ることで、関連キーワードが接続されたままくっついてくる。そのため、目的に応じたキーワードを引っ張ると目的カテゴリに容易に辿り着ける。

後者は、目的情報の取得に多量の Web ページ閲覧を必要とする場合や、検索キーワードと関係の強い情報の取得を目的とする場合に用いられ、ユーザにとって未知の情報の取得支援を目的とする。主な方法は、相関ルールを用いて、関係の強い Web ページやキーワードを取得し、視覚化することで実現される。主に、Web マイニングによって行われ、Web マイニングツールに「Visual Mining Studio」[(株)数理システム]がある。

従来研究の多くは目的キーワードと関係の強い情報を表示することで情報獲得を狙うものであり、ユーザの知識の整理・補完を行うものではない。そこで本研究では、ユーザの知識を整理・補完するインタフェースを提案する。

3. 提案するインタフェース

本研究で提案するインタフェースは、ユーザの知識の整理・補完を行うことを目的とする。例えば、ユーザがあるテーマに対

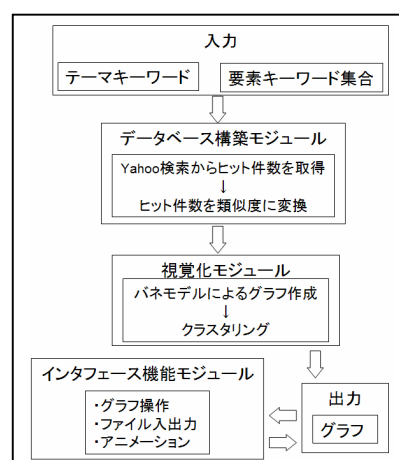


図1 インタフェースの概要

して関係が強い要素を調べたい場合や、ユーザが予想する考えに欠落はないかを確認したい場合に使用するインタフェースである。

3.1 提案するインタフェースの概要

提案するインタフェースでは、ユーザが知識を整理したいキーワード(テーマキーワード)と、テーマに関して関係を調べたい要素集合(要素キーワード集合)を入力する。例えば、ユーザが、サッカーに関して都道府県の間を知りたい場合は、テーマキーワードはサッカーとなり、要素キーワード集合は 47 都道府県名となる。

インタフェースは入力されたキーワード間の類似度に応じてキーワード間の距離を決定し、距離の近いノード集合でクラスタ分類を行い、無向グラフを出力する。対話的機能として、ユーザはノードの移動や類似度の変更や所属クラスタの変更を行うことができ、ユーザは自身の視点に基づいたグラフを作成できる。また、作成したグラフの保存と読み込みが可能のため、自身や他者が作成したグラフとの比較が行うことができ、複数のグラフの切り替え表示やアニメーション機能によって知識の整理・補完が行える。

3.2 システム構成

本システムは、「データベース構築」「視覚化」「インタフェース機能」の3つのモジュールから構成される(図1)。インタフェースの概観を図2に示す。以下に、各機能の詳細について述べる。

3.3 データベース構築モジュール

データベース構築モジュールは、入力されたテーマキーワードと要素キーワード集合に対する Web ページヒット件数を取得し類似度に変換する。N 個の要素キーワード集合 K の要素 k_i, k_j と、テーマキーワード T の AND 検索を行い、Web ページヒット件数をデータベース Hit に格納する(式 1)。次に、データベース Hit を i 行 j 列の行列に見立てたときの対角成分でそれぞれの列の除算を行い、類似度データベース DB を作成する(式 2)。また、本研究では、広く一般に用いられているという理由から検索エンジンとして Yahoo[Yahoo]を用いる。

$$Hit_{ij} = hit(K_i \text{ AND } K_j \text{ AND } R) \quad (1)$$

$$DB_{ij} = Hit_{ij} / Hit_{ii} \quad (2)$$

3.4 視覚化モジュール

視覚化モジュールは、データベース DB を元にグラフの作成とクラスタリングを行い表示する。グラフの描画にはパネモデルのアルゴリズムを用いており、類似度によってノード間の距離が変化するように KK 法[Kamada, Kawai 1989]を使用する。また、まとまったノード集合を分類し直観的に理解できるように K-mean 法によるクラスタリングを行う。

3.5 インタフェース機能モジュール

インタフェース機能モジュールは、グラフ操作、ファイル入出力、グラフ切替えなど、ユーザに対話的機能を提供する。グラフ操作は、ノード位置、所属クラスタ、リンクの強さの変更を行う。ファイル入出力はグラフ情報の保存や複数のグラフ読み込みを行う。グラフ切替え機能は、複数のグラフの切り替え表示やアニメーション機能を行う。

4 提案システムの評価実験

提案したシステムによって、ユーザの知識の整理・補完が行えるかを判定する評価実験を行う。

4.1 実験設定

今回の実験では、テーマキーワードをサッカー、野球、スノーボードの 3 つとし、要素キーワード集合を 47 都道府県名とする。

4.2 実験手順

まず、被験者にテーマキーワードに基づいて関係があると思われる都道府県の組合せと理由を挙げてもらう。その後、提案インタフェースを用いた上で、同じ質問に回答してもらった。

4.3 実験結果

テーマキーワードがサッカーの場合は、被験者は J リーグチームを持つ都道府県が一つの集合となると考えた。しかし、被験者は、全てのチームを思い出せなかったが、インタフェースによって関連が表示されると被験者は J リーグチームの情報を獲得・整理できた。しかし、サッカーに詳しくない被験者はインタフェースによって関係が表示されても情報の獲得が行えなかった。テーマキーワードが野球の場合は、被験者はプロ野球チーム、高校野球の強いチームを持つ都道府県の関係が強いと予想した。サッカーと同様に、インタフェースの使用によってチームをもつ都道府県の情報も取得できたが、野球に興味がなく予備知識を持たない被験者は情報の獲得できなかった。

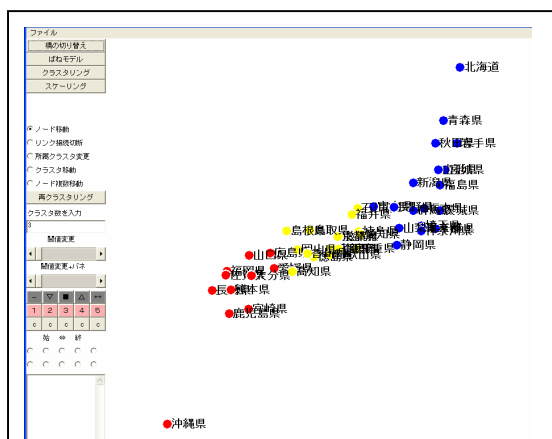


図 2 インタフェースの概観

テーマキーワードがスノーボードの場合は、被験者は雪が多く、スキー場がある都道府県の関係が強いと予想した。被験者はスキー場がある都道府県を確認できた。

4.4 考察

被験者は、インタフェースを使用し知識の補完を行えた。これはインタフェースが要素キーワードの関係を資格表示したことが被験者の欠落した知識を補完したと考えられる。被験者にある程度の予備知識があればインタフェースの使用によって知識の整理が行え、あまり予備知識がなければ知識は整理されない。また、関係が強い理由に気づく被験者は少なかったが、これは被験者に十分な予備知識がなかったことが原因と考えられる。

5 まとめ

本研究では、ユーザの知識を整理・補完するインタフェースを提案した。提案したインタフェースは、ユーザに予備知識がある場合とそうでない場合に結果に差が生じた。ユーザが十分な予備知識を持っていた場合には、ユーザの知識が補完されたが、そうでない場合にはユーザの知識は補完されなかった。一方、ユーザ自身が全く知らないテーマでは知識の整理は難しいことがわかった。

今後は、ユーザに予備知識がない場合でも対話的に情報の取得を行いながら知識の整理・補完を行うことが課題である。

参考文献

- [塩澤秀和 1997] 西山晴彦 松下温: 納豆ビューの対話的な情報視覚化における位置づけ, 情報処理学会論文誌, Vol.38 No.11 pp.2331-2342, 1997 年 11 月
- [Kamada, Kawai 1989] Tomihisa Kamada, Satoru Kawai, An algorithm for drawing general undirected graphs, Information Processing Letters, 1989
- [数理システム] <http://www.msi.co.jp/vmstudio/>
- [yahoo] <http://www.yahoo.co.jp>