

WWW を用いた事例ベース構築に関する研究

924125 尾林一世 924143 白神嘉人 (藤原研究室 W-Group)

1.はじめに

事例ベース推論(CBR)は,与えられた問題に類似する過去の事例(成功または失敗の経験)を直接利用して解を導く問題解決方式をいう.

一方,インターネット上で分散しているさまざまな情報を,多くの人々が共有し,簡単に見ること,知ることができることを目標にして,情報やサービスを公開するために作られたシステムとして WWW(World Wide Web)がある.

図やテキストが扱えてネットワーク・ワイドで使える事例ベースを実現するためにはオブジェクト指向データベースやマルチメディア・データベースが必要であり,これらは一般に高価である.本研究では,フリーの WWW を利用して事例ベースを構築することを目的とする.

具体的には HTML 言語による事例の表現法の検討,CGI による事例検索システムの作成,HTML 言語による入出力インターフェースの作成を行ないシステムの有効性を料理レシピの検索を例題として評価した.

2.システムの構成

本システムは,検索システムと入出力インターフェースから構成される.検索システムは C 言語と Perl 言語で作成し,入出力インターフェースは,HTML 言語によって作成した.図1にシステムの流れを示す.

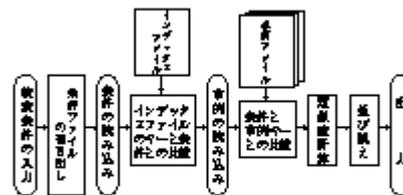


図1: システム処理の流れ

3.事例の表現

(1)事例ファイル

図2のaに示すように,事例を HTML 言語で記述する.事例ファイルでは図形ファイルの指定,検索用キー(keys)項目の指定,レシピ等のテキストの指定を HTML 形式に従って表記する.

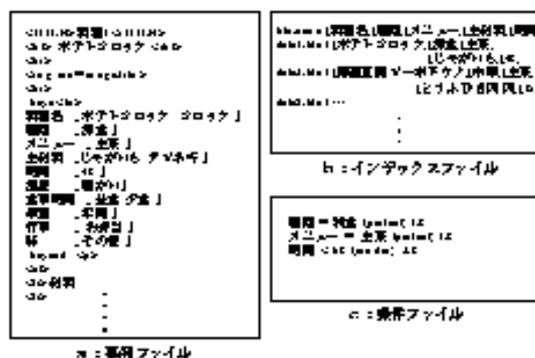


図2: ファイル例

検索キー項目は,事例の 10 個の属性(料理名,主材料,種類,メニュー,時間,温度,食事時間,季節,行事,味)であり,その事例の特徴を表す.以下にその形式を示す.

属性(条件)名 [属性値] …… (1)

属性値が複数の場合,複数表記も可能である.その場合,[属性値 1 属性値 2 …]の様に記述する.検索キーはaに示すように,-keys- ~ -keyend-内に(1)の書式に従って記述し,他の部分は HTML 形式に従い自由に書くことができる.

(2)インデックスファイル

検索の際,事例を絞り込むために,インデックスファイルを用いる.検索用インデックスファイルは,事例ファイル名と,条件として多用されると考える特定した 5 つの検索(条件)キー,すなわち料理名,種類,メニュー,主材料,時間の対で表現する.図のbに例を示す.

(3)条件ファイル

条件ファイルは,ユーザーが入力した条件の属性名とその属性値,範囲を示すオペレーター,完全照合か類似でよいかを表す照合条件,重要度を表す重みからなる.図のcに例を示す.

4.検索機能

検索システムは、まず、入力された条件をインデックスファイルに記述されているデータと比較し、検索される事例を絞り込む。絞り込まれた事例ファイルのそれぞれに記述されている検索用のキー項目と入力条件を比較し、該当する事例の類似度を計算し、類似度順に事例をソートして出力する。

事例ベース推論では、条件に完全に一致している事例だけでなく、条件と似ている事例も検索の対象とされなければならない。そこで本システムにおいて、条件の類似度を算出し、事例を検索する。類似度の計算方法は条件の種類により3つに分け、条件と事例の同名の属性値を比較する。

1. 条件値が文字列なら、その文字と事例の値を比較し、一致すれば0、そうでなければ1とする。
2. 条件値がリストなら、条件と事例のリストに含まれる値のうち的一致した数を e 、リストの値の数を x とすると、次式を用い計算する。

$$\text{類似度} = (x - e) / x \quad (x \neq 0)$$

3. 条件値が数値なら、条件値の数を x 、検索キーの値の数を k とし、次式を用い計算する。

$$\text{類似度} = |k - x| / 2x \quad (x \neq 0)$$

条件が複数存在する場合、各条件の類似度の値を合計した値を事例の類似度とする。

5.システムの評価

アンケート形式で評価を行なった。評価内容として、事例総数 100 のシステムにおいて、料理名、材料を条件とした検索結果の評価、時間を条件とした検索結果の評価、条件数の違いによる検索、類似の評価、また操作性について回答を得た。結果を表1に示す。

表1：評価結果

●：1人

悪い	評価内容	良い
●●●	料理名	●●●●●●
●●●●●●●●	主材料	●
●●	時間	●●●●●●●
●●●●●●●●	条件(単一)	●
●●	条件(複数)	●●●●●●●
●●●●●	操作性	●●●

表 1 より,主材料,条件の単一入力,操作性に問題点がうかがえる.主材料については,事例ベース内の事例数が少ないためと考えられる.条件の入力では,条件が一つの場合は,広範囲での事例検索となり評価が落ちたものと思われる.また操作性では,WWWブラウザの性能により評価がわかれたのであろう.その他の項目については概ね良好な結果を得た.

6.おわりに

WWW を用いた事例ベース・システムの構築法について述べた.今後の課題としては,事例の追加,文字属性の類似表現の検討,事例ベースの分散化,ネットワーク・セキュリティなどがあげられる.

SHIRAGA Yoshito & OBAYASHI Kazuyo