

弱い構造を使った情報収集と整理*

1 S-3

平田 高志† 前田 晴美† 西田 豊明†
 奈良先端科学技術大学院大学 情報科学研究科‡

1. 概要

情報ネットワーク上の不均質で多様な情報を統合的に扱うため、情報収集・整理支援システム CM-2 (Contextual Media version 2) を構築した。CM-2 の情報ベースは、不均質情報表現の弱い構造を用いたゆるやかな関連付けから構成される。本稿では、CM-2 の情報収集・整理機能と評価実験について述べる。

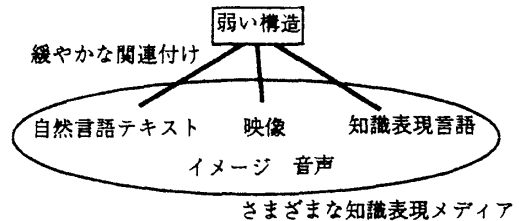


図 1: 弱い構造

2. 弱構造知識メディアと CM-2

我々はネットワーク上の大量の不均質情報を扱うためには弱い構造を用いたゆるやかな情報の関連付け (図 1) が有効であるという作業仮定に基づいてシステムの構築を進めている。弱構造とは図 2 のように何が与えられると何が想起されるかという連想関係を中心とした情報表現である。従来の知識表現を厳密で形式的に定義しようという方向に対して、ここでは表現の意味を漠然としか定義せず、表現の意味解釈を書き手と受け手に任せようというアプローチをとっている。このアプローチの是非については別の機会に論じるが、情報の構造化がすぐに実現できない知識獲得のボトルネックという技術上の限界ばかりでなく、情報表現が人間のメモ書き同様非常に手軽に生成でき、解釈の違いから新しい発想が期待できるなどの積極的理由もある。

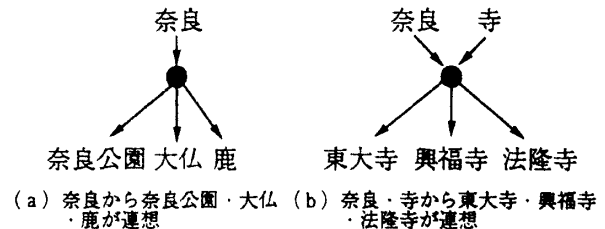


図 2: 連想構造

CM-2 の情報ベースは概念ユニットの集合であるワークスペースから成る。ワークスペースは必要に応じてユーザが作成し、アイコン化されたワークスペースユニットを他のワークスペース上に配置することができる。CM-2 システムはワークスペースを表示し、ユーザはマウスのドラッグ等により概念ユニットを再配置し構造の整形や、ユニットの付加・削除も任意に行なうことができる (図 3)。

3. CM-2 の情報収集・整理機能

3.1 情報の収集

WWW ページから形態素解析によりキーワードを切り出し、HTML 文書の構造などを手がかりに CM-2 で取り扱う弱い構造の中に取り込む [1]。

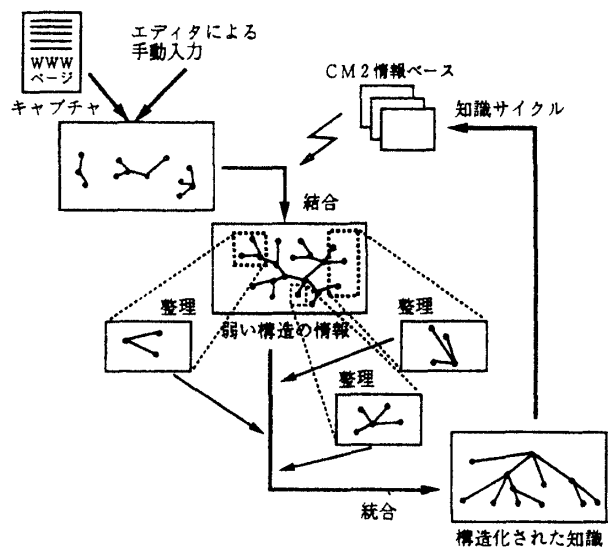


図 3: CM-2 における情報ベースの構造化と知識サイクル

*Information Gathering and Reorganizing Using Weakly-Structured Knowledge Representation

†Takashi Hirata Harumi Maeda Toyoaki Nishida

‡Graduate School of Information Science Nara Institute of Science and Technology

3.2 情報の結合

WWW ページから収集した情報ベース、CM-2 の基本編集機能を用いて手動で作成した情報ベースから、ユーザの関心のある情報を含む部分を集めて情報ベースを作る。CM-2 はこの過程を支援する情報ベース統合機能を有する。

3.3 情報の整理

情報の結合で再構成された情報ベースは普通冗長な情報が多分に含まれていたり、観点が一定していない構造が含まれている。ユーザが構造の修正・整理を行なう過程を支援するため以下の機能を提供している。

- フォーカス 指定したユニットに無関係なユニットの隠蔽
- 近傍展開 指定されたユニットに関連する連想関係表示
- パス探索 指定された複数のユニット間の連想関係表示
- 概念探索 ユニット群の中から指定されたユニット検索

ここで行なう操作により CM-2 の構造にユーザの持つ背景知識が少しずつ組み込まれていくと考えられる。

3.4 情報の分節と埋め込み

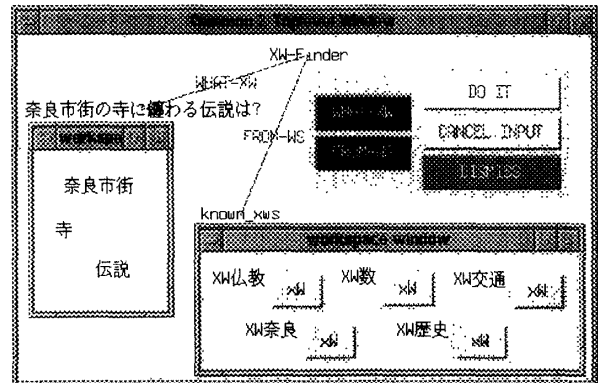
統合によって大きくなった情報ベースを再び扱いやすい大きさの断片に分ける過程を支援する。

分節 肥大し過ぎた情報ベースをある視点から切り分け細分化をはかる。

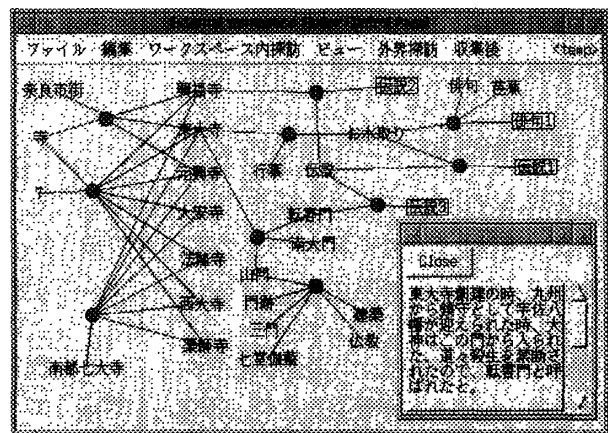
埋め込み ワークスペース上のまとまった情報を一つの情報ベースに埋め込む。

4. 実験と評価

CM-2 は、gcl(Common Lisp + Tcl/TK) を用いて実現した。図 4 にいくつかの情報ベースから奈良市街の寺に纏わる伝説に関する断片的な情報を収集し編集する例を示す。まず、キーワードとして「奈良市街」「寺」「伝説」のユニットを与え、既存の情報ベースを探索して、関連する情報の収集と結合を行なっている。(図 4(a))。ユーザは結合されて出来た情報ベースの構造を整理・統合し情報を精練する。図 4(b) は最終的に構造化された情報ベースである。この例では、「奈良市街」「寺」「伝説」に関する情報を含む情報ベース「xw 仏教」から仏教建築に関する門等の情報が、「xw 奈良」からは奈良市街の寺の行事・伝説等の情報が、「xw 数」からは 7 と寺から南都七大寺の情報が集められ整理されている。



(a) キーワードユニットと既存情報ベースの結合



(b) 構造化された情報ベース

図 4: CM-2 画面

システムの評価を行なうため、奈良の地域情報、新聞・雑誌等からの断片的情報を収集し、総計数千個の概念を含んだ情報ベースを試作した [1]。また、WWW からの情報獲得 [1]、オントロジー獲得支援システム [2]、モバイルコンピューティングを用いた会議支援 [3]、社会科教育への適用など応用システムに適用し連想構造の使われ方や有用性などについて評価を行なっている。

5. まとめ

本稿では既存の情報ベースを利用した知識の収集・結合機能と構造化のための整理・統合機能について報告した。今後は大規模実験による評価を行ないたい。

参考文献

- [1] 前田, 梶谷, 西田. 情報ベースのユーザフレンドリなインタフェースのための連想構造の提案. *Progress in Human Interface vol.5*, pp.49-56, 1996.
- [2] 梶谷, 前田, 西田. 文脈メディアによる弱構造情報の提示と構造化. 第 11 回ヒューマンインタフェースシンポジウム論文集 pp. 589-594, 1995.
- [3] 前田, 梶原, 足立, 沢田, 武田, 西田. InfoCommon : コミュニティにおける情報共有の支援 (ICMAS'96 Mobile Assistant Project 情報共有サービス). 情報処理学会第 54 回全国大会, 1997.