

木簡研究支援システムにおける 視点に依存した属性付け機構に関する検討

上島紳一 (関西大学総合情報学部), 森下淳也, 大月一弘 (神戸大学国際文化学部),
杉山武司 (姫路獨協大学情報科学センター), 田中克己 (神戸大学大学院自然科学研究科)

1 はじめに

近年, 文献情報データや環境情報データなどの科学技術データベースが注目されてきている [1, 2]. これらはデータが収集された時点で, データに対して大雑把な構造しか与えられていない場合が多い. 言わば半構造化の状態のまま, 構築されているという特徴がある. また, 確定したスキーマを持つ場合でも, そのデータを利用する段階で, 視点の違いや仮説などを盛り込んだ思考実験のような自由度を与えたいと考えられる.

木簡研究支援システムは, 中国敦煌遺跡から出土した木簡一千本, 中国居延遺跡の木簡一万本を対象にしたシステムである [3]. これらの木簡 (図 1) は, 発掘された時点で基本的なデータを収集し, その後, 木簡の釋読が進むにつれて文字データが収集される. さらに, 釋読文から個々の意味が汲み取られるという過程を経て, データベースに格納される. このようなデータは,

- 文字も分からない状態のまま, 長い期間保存される.
- データの解釈が後から発見される.
- データの意味が一般に確定するには, 長い期間の研究が必要である.

など, 特異な振舞いが見られる.

我々はこのような大雑把なスキーマしか持たない半構造化状態のデータに対して, 元のデータベースのデータを保持しつつ, 利用者の視点に応じて, 属性付けを行うことで様々な構造化を同時に表現するデータモデル, 階層構造グラフを考案した [4]. ここでは, そのデータモデルを用いて木簡研究支援システムに必要な属性付け機構を検討する.

次節で階層構造グラフを説明し, 3節では属性付け機構による属性の伝播ビューを論ずる.

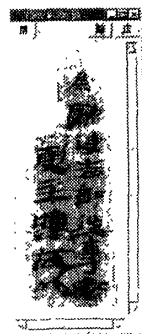


図 1. 木簡

Restructuring Data with Viewpoint for Wooden Slips Researches :
Shinichi Ueshima(Kansai University), Jun-ya Morishita, Kazuhiro Ohtsuki(Kobe University), Takeshi Sugiyama(Himeji Dokkyo University),
Katsumi Tanaka(Graduate School of Science and Technology, Kobe University).

2 階層構造グラフ

階層構造グラフは, ノードと枝からなるサイクルのない有向グラフである [4]. 各ノードと枝には属性集合を持つオブジェクトへのリンクが張られている. リンクされるオブジェクトはオブジェクト・アイデンティティと属性集合からなる. 次のような形式のものである.

$$O = \langle \text{Oid}, \{a_1 : v_1, \dots, a_n : v_n\} \rangle$$

ここで, Oid はオブジェクト・アイデンティティであり, a_i は属性, v_i は値である. この属性集合の a_i と v_i はそれぞれ次の様に $\langle \text{attribute} \rangle$ と $\langle \text{values} \rangle$ で定義されるデータである.

$$\begin{aligned} \langle \text{attribute} \rangle &::= \text{symbol}, \\ \langle \text{values} \rangle &::= \langle \text{value} \rangle \mid \langle \text{value} \rangle, \langle \text{values} \rangle, \\ \langle \text{value} \rangle &::= \text{int} \mid \text{string} \mid \dots \mid \text{Oid}. \end{aligned}$$

オブジェクトは 3 種類あり, それぞれ基本オブジェクト, カテゴリ, 関係オブジェクトと呼ばれる. 基本オブジェクトは元々の半構造化状態のデータを格納するもので, グラフの最下位のリーフに置かれる. カテゴリはノードに置かれ, 視点を表す. 関係オブジェクトは枝に置かれ, 視点とデータの間に固有の情報を書き込まれる. 図 2 に階層構造グラフの例を示す. ノードと枝は, リンクされたオブジェクトの属性集合 $\alpha_i, \beta_j, \rho_{kl}$ を各々持つ (四角で表される).

このグラフからカテゴリを定めて, 下位のノードまでのサブグラフを取り出すと, その下位のノードのデータだけでなく, 間に付与されたすべての属性を取り出すことができる. また, 同じ下位ノードを対象としてもカテゴリが異なれば, 全体として異なる属性集合を取り出せる. このサブグラフを仮想的なオブジェクトとして捉えることで, 視点に基づくいくつものビューをデータに与えることができる.

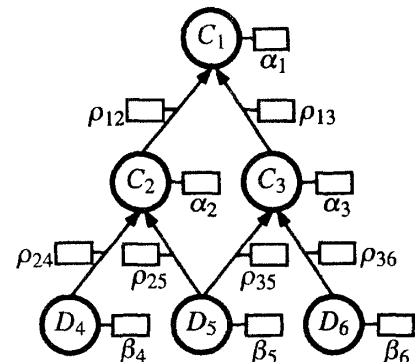


図 2. 階層構造グラフ

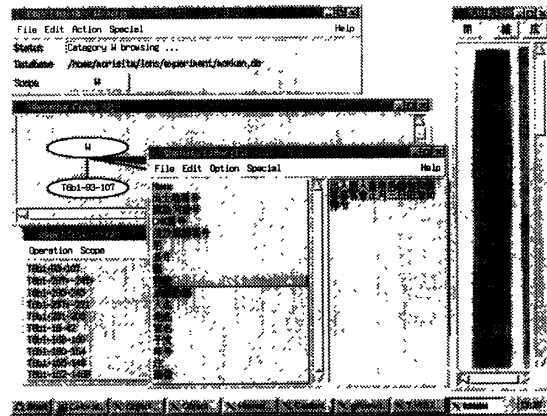


図3. 木簡研究支援システム

3 属性の伝播ビュー

階層構造グラフのサブグラフを指定して、その全体の属性集合を見るには、グラフの上位から下位へと属性を伝播させて見ることである [5]。これを属性の伝播ビューと呼ぶ。伝播ビューによって得られた全体の属性集合は、基本的に和集合となる。同じ属性に複数の値が併記される。

これに対して、木簡研究支援システムの場合、値が後ほど認識されて特定の属性がつけられるようなことが起こる。データ（あるいはデータの一部）が後から、別の属性に格納されるような場合を表現するためには、伝播ビューの際の置換ルールを定義することで、視点によって元のデータに含まれる属性をマスクしたり、置き換える機構が必要となってくる。例えば、ある木簡では、

釋読：七月己卯日罷軍病已 不迹八日丙戌

干支：己卯, 丙戌

となっているが、「己卯」が後ほど、属性「日付」として認識される。この時には、次のような伝播ビューを得られるようにしたい。

釋読：七月己卯日罷軍病已 不迹八日丙戌

干支：丙戌

日付：己卯

そのため、次のように特殊なシンボル、 f を用いて置換ルールを導入する。

$$f : "a_i : v_i \rightarrow a_j : v_j"$$

これは、伝播ビューの際に、下位のノードの伝播ビューに $a_i : v_i$ を持つものがあれば、 $a_j : v_j$ に置換して伝播せよと解釈される。これによって、先ほどの例は、

$$f : "干支:己卯 \rightarrow 日付:己卯"$$

と記述される。このルールは上位のノード、枝にあるルールが優先する。下位のデータノードから順に評価することで、明確にビューが確定する。この機構によって、不完全なスキーマを整えてみせることや、仮説をあたかもデータに直接格納されたかのようなグラフからの伝播ビューを生成することができる。

4 終わりに

階層構造グラフを用いた木簡研究支援システムにおける視点に基づく属性付け機構について考察した。このシステムは現在、Lisp によるプロトタイプシステムが稼働している (図3)。木簡システムのようなデータの特徴 (大まかに格納されたデータが認識された属性として取り出されるような場合は、木簡独自のものではなく、科学技術データを利用する立場で考えた場合、より広い範囲に適応されるであろうと思われる。グラフ構造とデータをリンクすることで、グラフからの伝播ビューによって自在なデータの描像が得られることから、データの再構築のインターフェースとしてこのデータモデルが活用できるものと期待される。

参考文献

- [1] IEEE Computer Society, "Special Issue on Scientific Databases," Bulletin of the Technical Committee on Data Engineering, Vol.16, No.1, Mar. 1993.
- [2] Zdonik, S., "Incremental Database Systems: Databases from the Ground Up," Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, Washington DC, USA, pp.408-412, May 1993.
- [3] Ueshima, S., Ohtsuki, K., Morishita, J., Qian, Q., Oiso, H. and Tanaka, K., "Incremental Data Organization for Ancient Document Databases," Proc. of the Fourth International Conference on Database Systems for Advanced Applications (DASFAA'95), pp.457-466, Singapore, Apr. 1995.
- [4] 森下淳也, 上島紳一, 大月一弘, 杉山武司, "階層構造グラフを用いた半構造化データの段階的構造化手法に関する検討," 情処研報, Vol.97, No.7, DBS96-111, pp.9-16, Jan.1997.
- [5] 森下淳也, 上島紳一, 大月一弘, 杉山武司, "階層構造グラフにおける属性の取り扱い方に関する検討," 信学技報, Vol.96, No.469, DE96-79, pp.31-36, Jan.1997.