

Concord: プロトタイプ方式の オブジェクト指向データベースの 試み

守岡 知彦

2006年5月31日

はじめに

Concord とは

- プロトタイプ方式のオブジェクト指向データベース

プロトタイプ方式の OODB とは

- オブジェクト指向
 - オブジェクト：素性（属性）の集合
- プロトタイプ方式
 - クラスがない

動機 (1)

とあるレガシーなデータベースを何とかしたい

- 時代毎に微妙に形式が違うデータが存在する
- データ形式の仕様が完全に把握できていない
- リファクタリングしたい

動機 (2)

最初、関係データベース (PostgreSQL (の改造版))
を使ったが…

- スキーマを変えるのは大変
- チューニングが大変
- もう SQL 嫌

動機 (3)

でも構造は扱いたい

- 全文データベースはちょっと嫌
- XML データベースもちょっと大変

従来の普通の OODB は

- クラス定義がある
- オブジェクトの永続化のため?

なんかちょうど良いのがないっぽい

動機 (4)

いろいろな文字が使いたい

- レガシーな DB の JEF + 謎の外字
- 常に現れるマイナーな漢字や甲骨文字、金文、その他、謎の文字

- 少なくとも UCS の拡張漢字 B は使いたい
 - Unicode 使えると称するものの多くは BMP だけしかサポートしてくれないことが多い
- できれば CHISE 技術が使いたい

動機 (5)

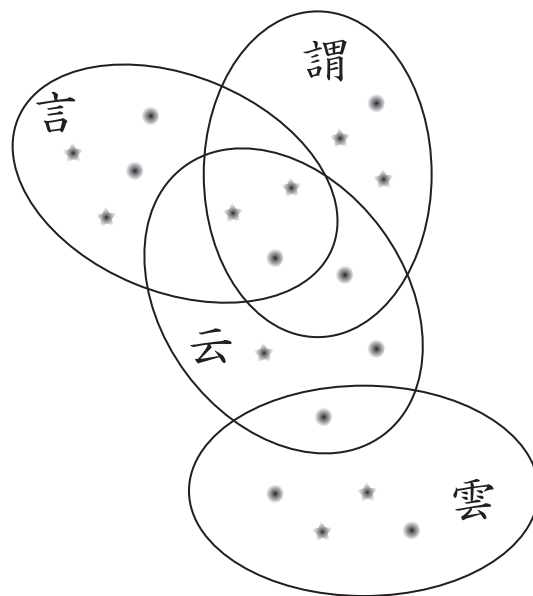
オントロジーとか知識データベースの類を意識したデータベース・エンジンが欲しい

どうするか

CHISE の手法（Chaon モデル）と実装を流用しよう

- 文字オブジェクトから一般のオブジェクトへ

Chaon モデル



文字オブジェクト：素性の集合

データモデル

- オブジェクトをノードとした有向グラフ
- ノードはオブジェクト
- オブジェクトは素性の集合
- リンクも素性的一种
- リンクはオブジェクト間の関係を表す

素性の種類

- 基礎素性：数値やシンボルなどのアトム（のリスト・配列）
- 関係素性：オブジェクト間の関係
- 構造素性：オブジェクトのリスト・配列
- ID 素性・写像素性
- メタデータ素性

実装

- libconcord
- XEmacs CHISE

libconcord

- C の API を提供
- 文字列処理レベル
 - オブジェクトはオブジェクト ID (文字列) で管理
 - 素性名、素性値は文字列

→ しょぼい

libconcord の中身

- Berkeley DB を利用
- *location/genre/feature/feature* というファイルに素性を格納
- このファイルのキーはオブジェクト ID, 値は素性値

XEmacs CHISE

- libconcord の Emacs Lisp バイndenディング
- concord-object 型
- 素性名はシンボル
- 構造を持った素性値を S 式として扱える

→ リッチ

アプリケーション

- CHISE (^_^);
- 某レガシーデータベース (現在開発中)

→ まだまだこれから

課題

- libconcord の機能が低い
 - 構造を持った素性値が扱いづらい
- XEmacs CHISE 以外のバインディングを増やす
- 全文データベースを使ったインデックス機能を付ける

- XML サポート
- など