

源氏物語湖月抄のXML化の試み

2000年07月09日 DB-West

早稲田大学MNC非常勤講師
大内英範

2000/07/09

湖月抄とは(1)

- 源氏物語の注釈書
- 北村季吟著
- 延宝元年(1763)成立
- 注釈54巻、他6巻、計60巻
- 江戸時代を通じて最も広く流布
- 契沖・真淵・宣長らにも影響大

2000/07/09

湖月抄とは(2)

- 物語の全本文を掲げ、上部に諸注を集成し、自説を加え、傍注にも諸注・自説あり
- ↓
- 「旧注データベース」としての性格

2000/07/09

旧注データベースとしての湖月抄

- データフォーマット
 - 本文
 - 読点あり
 - 頭注
 - 同一見出し語に複数の注
 - 傍注
 - 同一見出し語に複数の注

2000/07/09

XML版湖月抄

- インターネットへの対応
 - UTF-8等による国際化
 - さまざまな情報へのリンク
- 独自のツール・システムが不要
 - データ整備に専念
 - コスト削減
- IDによる検索
 - 『日本古典文学総合事典』との連携

2000/07/09

なぜXMLか

- eXtensible Markup Language
 - 拡張性
 - HTMLではできないデータ管理
 - 独自タグの一人歩き
 - テキスト形式
 - わかりやすい
 - IE5.xで表示
 - Perlで検索

2000/07/09

作成上の問題点(1)

- ID化にまつわる問題
 - 一定の基準のもとにIDを付与
↓
 - 単純検索では区別のできない同音異語を区別して検索可能
- 「一定の基準」とは？
→たとえば『日本古典文学総合事典』

2000/07/09

作成上の問題点(2)

- データレコードの「単位」
 - 単語？ 短すぎる？
 - 文節？ 短すぎる？
 - 文？
 - 段落？ 長すぎる？
- 本文データベースでは単語
- 本文異同データベースでは文節
- 注釈データベースでは(とりあえず)「文」

2000/07/09

将来性

- さまざまな情報との(相互)リンク
 - オンライン化された論文と
 - 音声データベースと
 - 画像データベースと
 - 本文データベースと

2000/07/09

参考

- 仏教学データベースにおけるXMLの活用—INBUDSにおけるID検索の実現にむけて(師茂樹氏 漢字文献情報処理研究会第2回大会 99/12/18)
- マルチメディア(パソコン)国語国文学 DB-West編著 95/01/18)
- 古典文学総合事典データベース構築の試み(伊井春樹氏 人文科学データベース研究創刊号)

2000/07/09