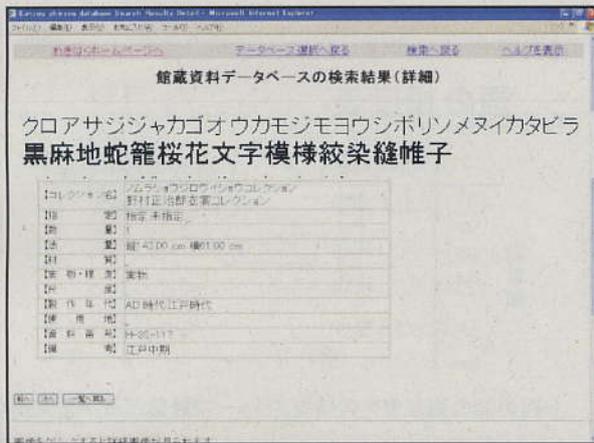
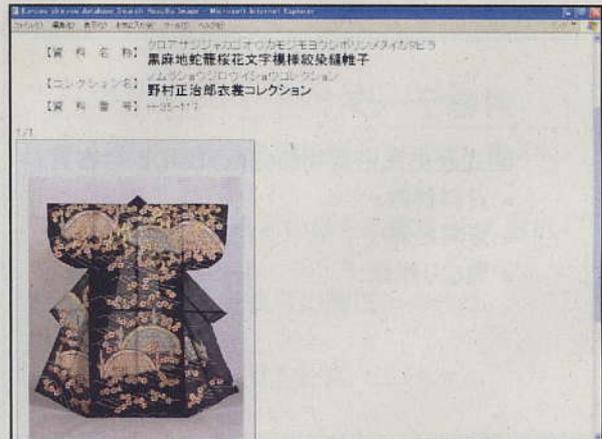
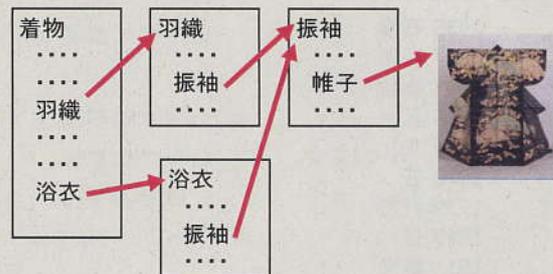


## 博物館資料群中の語の共起関係を用いた 関連後抽出における主要語選定の効果

画像電子学会 画像ミュージアム研究会  
博物館・美術館DTD-SG  
山田 篤(京都高度技術研究所)  
安達 文夫(国立歴史民俗博物館)  
小町 祐史(大阪工業大学)



## 関連語検索



## 資料群の構造

- A-539 榎の木遺跡出土品
  - A-539-1 土器類
    - A-539-1-1 注口土器
    - A-539-1-2 注口土器
    - ⋮
    - A-539-1-645 土器破片
  - A-539-2 石器類
    - A-539-2-1 石篋
    - A-539-2-2 石匙
    - ⋮
    - A-539-2-2601 軽石製品

## 資料群の構造

- A-258 亀ヶ岡遺跡出土縄文時代遺物
  - A-258-1 土器
    - A-258-1-1 壺形土器
    - A-258-1-2 壺形土器
    - ⋮
    - A-258-1-77 鉢形土器
  - A-258-2 石器・その他
    - A-258-2-1 岩版
    - A-258-2-2 土偶
    - ⋮
    - A-258-2-230 籃胎漆器

## 対象データ

国立歴史民俗博物館館蔵資料DB 考古資料

- 資料群数: 633
- 資料名称: 21,935
- 異なり総数: 4,492

## 高頻度の資料名称(上位10語)

- 1252 石鏃
- 575 打製有茎石鏃
- 537 深鉢形縄文式土器
- 451 石片類
- 339 石匙
- 303 文字平瓦
- 254 石七
- 244 磨製石斧
- 233 出土地不明打製有茎石鏃
- 225 平瓦

## 形態素解析後の高頻度語

- 2935 石器
- 2537 石鏃
- 2507 複製
- 2375 瓦
- 2015 文
- 1870 石
- 1779 打製
- 1717 形
- 1511 縄文
- 1485 式

## 語の頻度表

	語彙					
	v1	v2	v3	v4	・	vn
G1	*	*				
G2		*	*		*	
G3		*	*	*		
G4		*	*		*	
・					*	*
Gm					*	*

語の間の類似度を内積型メジャーで計算 (GETAを使用)

## 連想検索結果 Word: 土器

- |       |            |
|-------|------------|
| 1: 鉢形 | 11: 錐      |
| 2: 広口 | 12: 碟      |
| 3: 深鉢 | 13: 器      |
| 4: 土器 | 14: 篋      |
| 5: 人面 | 15: 石槍     |
| 6: 小形 | 16: 品      |
| 7: 浅  | 17: 片      |
| 8: 縄文 | 18: スクレイパー |
| 9: 小型 | 19: 軽石     |
| 10: 口 | 20: 石皿     |

## 主要語の選定

楕円/皿形/縄文/式/土器/  
 明/青花/竜/紋/輪花/皿/  
 黒色/土器/杯/  
 鉄/絵皿/(/瀬戸/美濃/窯/)/

## 連想検索結果 Word: 土器

主要語を対象	全形態素解析結果
1: 土器	1: 鉢形
2: 石鏝	2: 広口
3: 敲石	3: 深鉢
4: 石斧	4: 土器
5: 石	5: 人面
6: 破片	6: 小形
7: 土偶	: ……
8: 石器	135: 敲石
9: ヒ	171: 石斧
10: 錘	181: 石鏝

## 連想検索までのプロセス

- 資料名称の形態素解析 MeCab, UniDic
- ( ) を除く末尾の語を主要語として選択
- 資料群—主要語の頻度表の作成
- 語間の類似度を計算 GETA

## 専門語

石鏝(セキゾク)	琴柱(コトジ)
梵鐘(ボンショウ)	瓔珞(ヨウラク)
高坏(タカツキ)	銅戈(ドウカ)
鉄滓(テツサイ)	管玉(クダタマ)
刀子(トウス)	大甕(オオガメ)

## 専門語

石鏝(セキゾク)	琴柱(コトジ)
梵鐘(ボンショウ)	瓔珞(ヨウラク)
高坏(タカツキ)	銅戈(ドウカ)
鉄滓(テツサイ)	管玉(クダタマ)
刀子(トウス)	大甕(オオガメ)

## 専門語の出現順位

	土器
石鏝(セキゾク)	2
高坏(タカツキ)	13
管玉(クダタマ)	30
大甕(オオガメ)	129
鉄滓(テツサイ)	135
刀子(トウス)	176

## 連想検索結果 Word: 土器

主要語を対象	
1: 土器	11: 錘
2: 石鏝	12: 磗
3: 敲石	13: 器
4: 石斧	14: 篋
5: 石	15: 石槨
6: 破片	16: 品
7: 土偶	17: 片
8: 石器	18: スクレイパー
9: ヒ	19: 軽石
10: 錘	20: 石皿

### 専門語の出現順位

	土器	石斧	石槍	石皿
石鏃(セキゾク)	2			
高坏(タカツキ)	13			
管玉(クダタマ)	30			
大甕(オオガメ)	129	135	-	-
鉄滓(テツサイ)	135	141	-	-
刀子(トウス)	176	172	-	-

### 連想検索結果 Word: 土器

主要語を対象

1: 土器	11: 錐
2: 石鏃	12: 磔
3: 敲石	13: 器
4: 石斧	14: 筥
5: 石	15: 石槍
6: 破片	16: 品
7: 土偶	17: 片
8: 石器	18: スクレイパー
9: 七	19: 軽石
10: 錘	20: 石皿

### 主要語抽出の課題

- 形態素解析1
  - 壺/形/須恵/器 → 壺/形/須恵器
  - 土器/注/口 → 土器/注口
- 形態素解析2
  - 銅/鐸 → 銅鐸
  - 石/七 → 石七
- 主要語選択
  - 三/彩釉/陶器/複製 → 三/彩釉/陶器/複製
  - 壺/形/埴輪/片(破片) → 壺/形/埴輪/片

### まとめ

- グループ化された資料の名称から、関連語の抽出が可能
- 主要語を選定することで、連想検索の効果が向上
- より有効な主要語の選定方法について、検討が必要