

博物館資料情報の検索のための 発見的検索手法

Heuristic Search for Information on Museum's Holdings

山田 篤・安達文夫・小町祐史

YAMADA Atsushi, ADACHI Fumio and KOMACHI Yushi

①はじめに

②歴博の資料管理の特徴と館蔵資料データベース

③資料集合における関連度の計算

④関連語を用いた発見的検索

⑤資料集合の構成法

⑥異なる分野の混在の効果

⑦まとめ

【論文要旨】

本稿では、博物館資料情報の高度な検索の一手法として、既存の資料名称から語を取り出し、それらの間の関連度を用いて、一般的な語彙から専門的な語彙にたどり着く発見的な検索手法を提案し有用性を検証する。

関連度の計算においては、一般的な文書に含まれる語の間の関連度を測る方法を適用することができるが、資料名称に含まれるすべての語を対象とするのではなく、その中から主要な語を選定した方がより適切な関連語が得られる。

このようにして得られた関連語を関連度の高いものから順に提示し、関連語を順にたどっていくことで一般的な語彙から専門的な語彙にたどりつくことが可能になる。

関連語の計算においては、資料集合の設定が大きすぎると、何でも関連づけられてしまい、結果として望ましくない関連が発生し、資料集合を細かく分割し過ぎると、関連語としてふさわしい語までが排除されてしまうため、適切な資料集合の設定が課題となる。これに対して、コレクションの木構造を利用した資料集合の構成法を示し、その効果を検証している。

最後に、分野別のコレクションを併せて関連語を探す際には、全般的な傾向としては、語彙数の増加により、到達容易性は低くなるが、異なる分野において共通に出現する語を介して、分野間を横断した検索が可能になることもあることを示している。

【キーワード】 資料名称、関連度、到達容易性、木構造