

電子図書館によるアプリケーションデータ管理*

4Y-03

佐藤 慶三 松永 和久 高橋 洋光 伊藤 哲郎†

大分大学工学部‡

1. はじめに

イントラネット等の普及により、複数のユーザが互いのアプリケーションデータを共有しながら作業する環境が増えてきている。本稿では、アプリケーションデータの集合を電子図書館によって効率的に管理する手法について述べる。

従来の文書データを対象とした電子図書館[2][3]では、文書データはキーワードにより特徴づけられており、このキーワードによるマッチングで文書データとユーザの検索要求あるいは文書データ同士の関連性が求められていた。

文書データがキーワードで特徴づけられているのに対し、アプリケーションデータは一般に複数のフォルダに分類されていることから、この分類情報も特徴とみなせる。ここでは、アプリケーションデータのパス名も特徴として利用する。従来のキーワードに加え、パス名を利用することで、所望のアプリケーションデータを効率的に取り出すための、ブラウジング支援が期待できる。

2. 電子図書館の概要

電子図書館のユーザインタフェースおよび、ブラウジング法について述べる。

2.1 ユーザインタフェース

データの配置は、伝統的な図書館に倣い、キーワードマッチングで類似度の高いデータ同士が近くに置かれるような線形配置とし、複数列の書架表示で提示する。

書架のレイアウトは、(i)参照するデータのある棚、(ii)適合と判断されたデータがある棚そして(iii)それ以外の棚で以下のように棚幅を割り当てる。

- (i) 中央に配置し、幅も広く取る
- (ii) 部分表示 (1~2 単語分)
- (iii) 幅を縮めて表示

このレイアウトにより、重要な部分を強調しながらより多くのデータを提示できる。

2.2 ブラウジング法

ユーザの、各データに対する適合判断結果を基に、未参照データの適合可能性を判定する WiB1[2][3]が定式化されている。これは、判定対象のデータが、不適合とされた任意のデータに対するより、ある適合とされたデータに対し、キーワードマッチングでより高い類似度を示すかどうかで適合可能性を判定する。

ユーザは、適合可能性の示唆された文献から参照していくことで、効率的なブラウジングが行える。

3. アプリケーションデータの管理

アプリケーションデータは複数のフォルダに分類されていることから、アプリケーションデータの管理にはパス名やフォルダ単位での情報も利用する。

3.1 パス名のマッチング

アプリケーションデータの分類先が近い場合は、作業内容などの観点でそのアプリケーションデータ同士の関連性が高いと考えられることから、各データのパス名をデータ間の関連性を求めるのに利用する。

パス名の類似度計算には、構造マッチング[1]を用いる。これは、文同士の類似性を求める測度で、これによりパス名の一致の度合いを類似度として表せるようになる。

図 1 はパス名同士の類似度計算の例である。パス名のリストと類似度行列を示してある。

/semi_z4/Semi10/input.java	1.00	0.07	0.11	0.57	0.05
/reference_1/article002.pdf	0.07	1.00	0.16	0.08	0.04
/semi_m1/Report03.ppt	0.11	0.16	1.00	0.08	0.09
/semi_z4/Semi10/Shelf.java	0.57	0.08	0.08	1.00	0.03
/temp_3/JAVA11.HTM	0.05	0.04	0.09	0.03	1.00

図1. パス名の類似度計算の例

3.2 類似測度の動的切り替え

ユーザの検索要求がキーワードによるものか、パス名によるものかは状況によって異なり、ブラウジングの過程でも変化しうる。このため、キーワードによる類似測度とパス名による類似測度を、ユーザの適合判断結果により動的に切り替えるようにする。

*Managing the Application Data in a Digital Library

†Keizo Sato, Kazuhisa Matsunaga, Hiromitsu Takahashi and Tetsuro Ito

‡Faculty of Engineering, Oita University
700 Dannoharu, Oita, 870-1192 Japan

適合判断の結果から、2種の類似測度の何れが適合可能性示唆に適切かを判定し、類似測度を切り替え、適合可能性示唆や電子図書館上での配置もその類似測度を基に行う。

これにより、適合判断の傾向が変化しても、その傾向にあわせて類似度の高いデータ同士が常に近くに置かれるようになり、ブラウジングに適した配置を取れるようになる。また、適合可能性の判定も適合判断の傾向にあわせて行うため、より柔軟な適合可能性の示唆が行える。

3.3 フォルダのインスタンス化

各フォルダには、作業内容の共通した複数のアプリケーションデータが分類されており、作業内容によってフォルダが区別されることから、フォルダをインスタンスとして扱い、他のアプリケーションデータ同様に、電子図書館上に提示する。

フォルダ中の各アプリケーションデータがもつ特徴の集まりをフォルダの特徴とし、これを電子図書館上で一つのアプリケーションデータとして提示することで、フォルダ単位の操作が可能になる。

フォルダ単位で適合可能性を示唆したり、逆にフォルダへの適合判断結果を、他のフォルダに分類されている関連データの適合可能性を導き出したりするのに利用できる。

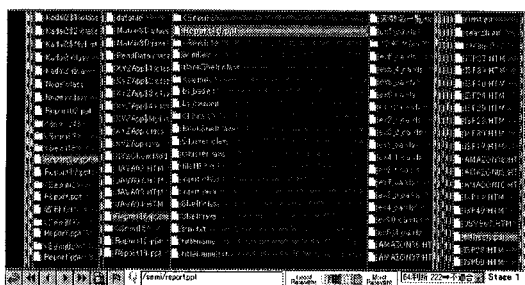


図2. 電子図書館のユーザインタフェース

4. 評価方法

類似測度の動的切り替えおよびフォルダのインスタンス化によるブラウジング支援の効果を検証していくための、評価方法について述べる。

評価に使用するアプリケーションデータは、研究室内で実際に使用されているアプリケーションデータをテスト用データとし、以下のような状況を想定した検索要求と、それに対する正解データを設定しておく。

- ・ 特定のデータを取り出す
- ・ 条件（作者等）に該当するデータを取り出す

- ・ 「～」について調べる

設定した検索要求に対するブラウジングを、計算機によるシミュレートあるいは被験者実験により実施する。

類似測度の動的切り替えによるブラウジング支援の効果については、キーワード、パス名それぞれ単独の類似測度のみでブラウジングを行う場合とのブラウジングコスト（ステップ数や移動距離など）を比較し評価する。動的に類似測度の切り替えを行った場合のブラウジングコストがより小さければ、ブラウジング支援の効果があるとみなせる。

フォルダのインスタンス化による効果については、ブラウジングの過程で適合可能性の示唆がなされたフォルダを適合とした場合の、未参照データの適合可能性を観察する。フォルダを適合とすることで、正解データのうちのどれだけのデータに適合可能性が生じるかを調べる。これにより、フォルダのインスタンス化による、関連データの適合可能性を導き出す効果を検証することができる。

5. まとめ

アプリケーションデータが複数のフォルダに分類されていることに着目し、この分類情報を利用して効率的なブラウジングを支援する、電子図書館によるアプリケーションデータの管理について提案した。

今後は、4.で述べた評価方法に従った実験を通して、類似測度の動的切り替えおよびフォルダのインスタンス化によるブラウジング支援の効果の検証を行う。

類似測度の動的切り替えにおいては適合判断の傾向をどのように判定するか、またフォルダのインスタンス化についてはフォルダ中のアプリケーションデータがもつ特徴の反映の仕方や、子フォルダの扱いなどの、具体的な計算方法について検討していく。

参考文献

- [1] 安部 隆之, 佐藤 浩史, 重松 修一, 中島 誠, 伊藤 哲郎: 構造マッチングによる文献の知的検索と結果の色空間表示, 情報処理学会研究報告, 人文科学とコンピュータ, vol.29-5, pp.25-30, Jan. 1996.
- [2] Yanhua Qu, Keizo Sato, Makoto Nakashima, and Tetsuro Ito: Browsing in a Digital Library Collecting Linearly Arranged Documents, Proc. SIGIR' 2001, New Orleans, USA, pp.426-427, Sept. 2001.
- [3] 曲 艶華, 佐藤 慶三, 中島 誠, 伊藤 哲郎: 電子図書館のための適合可能性示唆によるブラウジング支援, 電子情報通信学会論文誌 D-I, vol.J84-D-I, no.7, pp.1009-1020, July 2001.