

技術ドキュメントを対象としたインデックスによる検索手法の研究

1Q-5

井上健ⁱ 植田正彦ⁱⁱ 田中孝一ⁱⁱ 高岡良行ⁱⁱ
池田満ⁱⁱⁱ 溝口理一郎ⁱⁱⁱ 東都雄ⁱ

ⁱ関西電力株式会社 ⁱⁱ東光精機株式会社 ⁱⁱⁱ大阪大学産業科学研究所

1. はじめに

近年情報技術の発展に伴い企業ではコンピュータ・ネットワークにおける情報の伝達・共有と言った利用環境が整備され、各種文書の電子化が進んでいる。しかし、この蓄積された文書の活用と言った利用面では、組織、業務形態、目的に沿って効率よく活用されていると言いがたい状況である。

本稿では、的確な答えが要求されるサービス提供段階における、技術ドキュメントを対象としたインデックスによる検索手法について取り組んだので報告する。

2. 対象業務と検索手法について

情報を業務形態から眺めた場合、オフラインで新たな発見、知識獲得を期待して欲しい情報を広くノイズも含め集める調査段階と、オンラインとして次の行動が前面に控えており、時間をかけることが許されない的確な答えを要求するサービス提供段階に分けられる。また、目的から眺めた場合、一過性で確認したら終了するケースと整理して再度活用するケースが考えられる。

ここでは、企業において様々な部門・業務段階で直面するサービス提供段階において繰り返し情報が活用される業務を前提としている。

次に、情報を探す方法としては大きく分けて全文検索とインデックス検索に分けられる。これらの手法は1長1短があり上記に述べた業務形態や利用目的により使い分けられている。全文検索では、「具体的なキーワードが思い付かない」、「知りたいのは内容であり、キーワードでない」と言った問題があり、また、インデックス検索では、「分類目を作成する考え方が明確でない」と言った問題がある。

我々はオントロジ技術を適用して従来の問題点を解決し、的確性と柔軟な質の良い技術ドキュメントを対象としたインデックスによる検索手法を提案する。

本検索手法での目的は下記の3点である。

- ・ 「具体的なキーワードが思い付かない」問題を解決する。
- ・ 内容による検索を実現する。
- ・ 明確な分類方法を提案する。

この3点の目的を解決するため、我々はオントロジに基づいたインデックスによる検索手法を提案しプロトタイプシステムを作成した。

3. プロトタイプシステムの機能

今回提案する手法により我々はプロトタイプを作成した(図1参照)。プロトタイプでは先に上げた3つの目的を解決するために、次の機能を実現している。

- ・ 柔軟な検索
- ・ 内容による検索
- ・ 関心度に従った優先表示順位

以下に3つの機能について説明を行う。

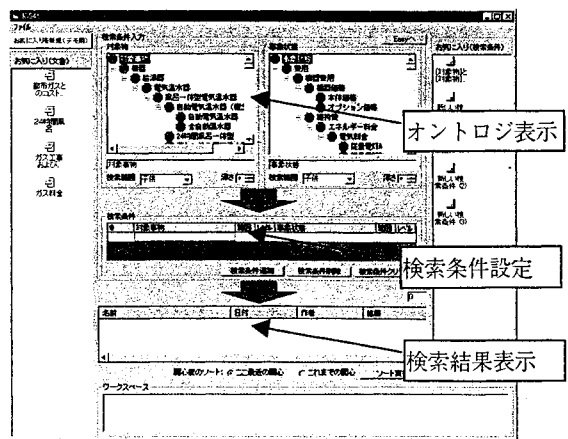


図1. プロトタイプシステム画面図

1) 柔軟な検索

我々の提案する検索手法はインデックスを用いた検索である。したがって、前処理として対象の文書にインデックスを貼り付ける必要がある。前処理のコストがかかっても使いやすい検索が行えると我々は考える。

インデックスはオントロジ技術を用いて表現する。オントロジから選択した語彙よりインデックスをタグ付けするため、ユーザごとのインデックス決定能力に依存せずにタグ付けすることができる。ユーザ間の語彙の揺れを解消することができ、一定の水準でインデックスをタグ付けすることが可能となる。

オントロジを用いることで検索条件が抽象的であっても、オントロジの概念間の関係により具体的な内容を検索することが可能となる。具体的なキーワードが思い付かなくても検索が可能となる。本システムでは図1の検索条件設定で、キーワードと検索範囲を広げる階層数や広げ方を設定する。

2) 内容による検索

1)で説明したように対象となる文書にはインデックスを貼り付けている。このインデックスは内容を表現したものをを用いる。今回対象とする業務形態や目的では2軸で十分であると考え、技術ドキュメントの内容は物事のある側面を指すとし、「対象事物」と「事象状態」のセットで表現する。本システムではこの2軸の明確な分類方法により分類する。

この「対象事物」と「事象状態」のセットにより知りたい内容を表現し検索条件とすることで、内容による検索を実現する。図1のオントロジ表示により2軸となるオントロジを表示し選択することで、検索条件の内容を表現する。

3) 関心度に従った優先表示順位

ユーザのプロファイルにより関心のある内容をシステムが推測し、関心度の高いものから優先的に表示する。オントロジの体系からインデックス体系の深さは詳細度を、広がりには範囲度を表していると考えられる。

範囲度は各ユーザの関心のある内容がなにであるかを決定する。企業内においては関心のある内容は業務

によって多くは決定される。例えば、補修業務の者は「しくみ」について関心が高く、管理業務の者は「費用」について関心が高いはずである。また、詳細度は各ユーザによって知りたい内容の詳細度に傾向があると思われる。例えば、初心者は大まかな内容を知りたい傾向にあり、上級者はより具体的な内容を知りたい傾向があると考えられる。

この詳細度と範囲度を基に検索を行うユーザの関心度を計算する。個人やグループの最近(過去)の関心傾向をスコアリングすることで選択文書を優先的に表示することが出来る。本システムでは図1の検索結果表示に表示された時点で関心度によるソートを行う。

4. プロトタイプシステムの作成検証

技術マニュアルを対象としてプロトタイプシステムを構築し機能を検証した。入力ドキュメントは約100枚とし、1ドキュメントに対し内容が複数あることを許しインデックスは複数付与できることとした。

探したい情報の内容を表わす2軸のインデックスを選択するだけで、抽象的な内容から具体的な内容に絞り込みながら検索することができることを確認した。

この手法におけるインデックス体系の作成では、業務形態や目的に応じた詳細度のオントロジの作成が要求される。また、インデックス体系を意識させないインターフェースが必要となることが分かった。

5. おわりに

本稿では、オントロジ技術を活用したインデックス検索手法について述べ、従来の検索手法での問題点を改善できることを示した。また、企業における迅速的確な答えを要求されるサービス提供段階等、業務・利用形態によっては、本手法がより高いコストパフォーマンスを与えることができる可能性があることから、今後更に機能の充実を図っていきたい。

参考文献

[溝口 97] 溝口理一郎, 池田満: オントロジ工学序説-内容研究の基礎技術と理論の確立を目指して-, 人工知能学会誌, Vol.12, No.4, pp559-569, (1997)