

文献情報共有支援システム Papits における Know-Who 検索機能の実現について

後藤 将志, 大園 忠親, 新谷 虎松
名古屋工業大学 知能情報システム学科

An Implementation of Know-Who Search for Documents Sharing Support System Papits

Shoji GOTO, Tadachika OZONO, and Toramatsu SHINTANI,
Dept. of Intelligence and Computer Science, Nagoya Institute of Technology

Abstract: In this paper, we present a method for searching users who know answers of users' questions. This searching method is called a "Know-Who" search. The method is used in Papits, that is a documents sharing support system developed in our laboratory. In order to realize the Know-Who search, we construct users' knowledge models from various files that users have made out.

1 はじめに

本研究では、グループでの研究活動を支援するため、マルチエージェントに基づくナレッジマネジメントツールの開発に取り組んでいる。今回発表する Papits は、そのためのエージェントシステムの一つとして試作された文献情報共有支援システムである。研究活動において、他の関連研究のサーベイは欠かすことができない重要な作業である。本システムでは、研究活動中の文献サーベイを重点的にサポートするためにさまざまな機能を提供している^{1, 2)}。文献サーベイにおいて、自分の知らないことが文献の中で挙げられていた場合、自分で調べるより、知識を持つ人に直接聞いた方が効率的である。このように、ユーザからの質問に対し、適切な答えを持っていると考えられる人を検索する機能を Know-Who 検索と呼ぶ。Know-Who 検索を実現するためには、各ユーザの知識をモデル化する必要がある。本研究では、Papits の運用において、ユーザモデル構築のためのユーザの負担が問題として挙げられている²⁾。本研究では、極力ユーザが機能実現のために手間をかけることなく Know-Who 検索を実現するために、各ユーザが研究の過程で作成したファイルからユーザの知識をモデル化する。本稿では、Papits における、Know-Who 検索のためのユーザモデルの構築手法とユーザモデルに基づく Know-Who 検索手法について述べる。

2 文献情報共有支援システム Papits

文献情報共有支援システム Papits は、研究活動における、文献サーベイを支援するために開発されているシステムである。Papits では、新たなサービスの追加等を容易にするため、マルチエージェントシステムとして実現されている。Fig.1 に Papits の実行画面を示す。

文献サーベイを主に文献を読むことによる情報収集活動と捉え、文献を探すこと、読むこと、文献に対する情報を管理することが必要となる。これらの作業に費やされる時間的コストはたいへん大きい。Papits では、論文を探すことや、管理することに関する問題に対し、グループ内での情報の共有機能や、ユーザに適した情報を推薦するといったような機能を実現してきた^{1, 2)}。

Papits では、論文を読むことをサポートするための機能として、Know-Who 検索機能を提供している。Know-Who 検索とは、ユーザの質問に対する答えや知識を持っている人を検索する機能である。文献を読んでいて、自

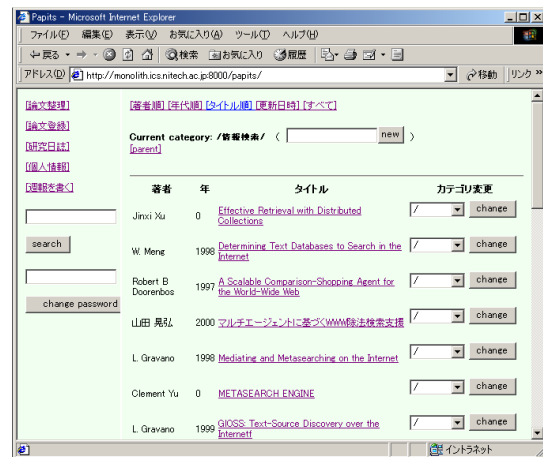


Fig. 1 Interface of Papits

分の知らないことが現れたときは、その内容に関する知識を持つユーザに尋ねるのが最も効率的である。以下では、Know-Who 検索において必要とされるユーザの知識のモデル化と、モデルに基づいた検索について述べる。

3 Know-Who 検索

Know-Who 検索を提供するエージェントの構成を Fig. 2 に示す。本システムは、文献管理エージェント、ファイル管理エージェント、そして Know-Who 検索エージェントの 3 つのエージェントで構成される。更に、Know-Who 検索エージェントは二つのモジュールで構成されている。一つは User Modeling Module、もう一つが Matching Module である。User Modeling Module は、ユーザの知識をモデル化するモジュール、Matching Module は、ユーザモデルとユーザの検索要求のマッチングを行い、Know-Who 検索を実現するモジュールである。以下でそれぞれの詳細を述べる。

3.1 ユーザモデルの構築

本システムでは、Know-Who 検索実現の際の、ユーザへの負担を極力軽減することを目指し、ユーザモデルの構築に際し、ユーザの直接的な協力を必要としない、2 つ

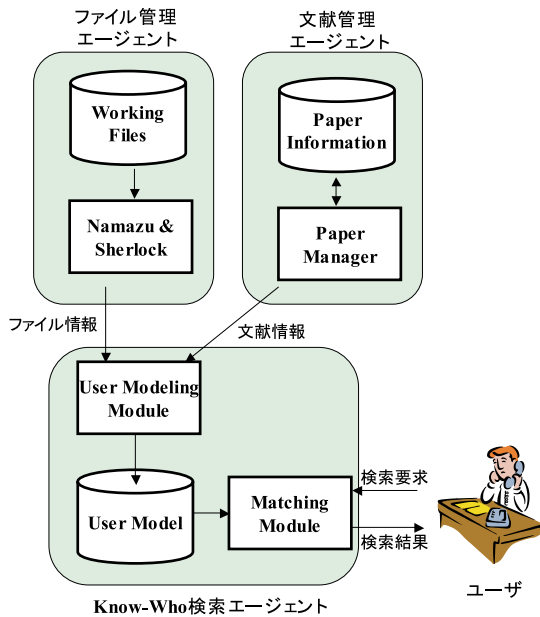


Fig. 2 System Architecture

の情報を利用する。1 つは各ユーザが作成した作業ファイル、もう 1 つが論文管理の状態から得られる情報である。ユーザは研究の過程でメモや論文など、さまざまなファイルを作成する。そのような作業ファイルにはその人が持つ知識や興味に関する情報が含まれている。本システムでは、ユーザモデル構築に作業ファイルを利用することにより、ユーザに特別な作業を強いることなくサービスを提供することができるようになる。本システムでは、UNIX 上の NFS で共有されたホームディレクトリの情報と Macintosh の AppleShare で共有されたファイルを対象にしている。ホームディレクトリ上のファイル解析のため、全文検索エンジン Namazu¹ を利用し、ホームディレクトリ上の各ユーザのファイルのインデックスを作成する。また、AppleShare で共有されたファイルサーバ上のファイルは、Macintosh 標準のファイル検索エンジンである Sherlock を AppleScript で制御することにより、各ユーザのファイル情報を取得している。もう一つの情報として、ユーザが Papits に登録している論文情報を利用する。登録された論文に含まれる情報はユーザにとって興味の対象となっている。これらの情報はユーザの研究過程の中で必然的に作成されるものであり、ユーザは Know-Who 検索実現のために特別な作業をする必要はない。

作業ファイルには雑多な情報が含まれている。文書から構築されたユーザモデルはキーワードベクトルで表現されることが多い。しかし、作業ファイルのような雑多な情報を対象とした場合、キーワードに基づいた手法では、雑多な情報の中の部分的な情報しか取得することができない。本システムでは、得られる情報を元にシソーラスを構築し、ユーザの知識のモデルとして利用する。シソーラスとは語と語の関係を表現する辞書であり、語を頂点に、関係を辺にとるとグラフ構造で表現できる。ある語が異なった意味で出現する情報からは異なったシソーラスが構築される。なぜなら、語は意味によって関連する語が異なるためである。本手法では、シソーラスを用いて語の意味を同定することにより検索要求とユーザの知識の関連性を計算し、適切なユーザを選択する。

シソーラスの構築において、人手での作成は多大なコストを必要とするため、共起関係に基づく自動構築を行

う。共起関係に基づくシソーラスでは、ある範囲に同時に現れる語は関係があるという仮定に基づいて、シソーラスを自動構築する。本研究では、シソーラスにおける語の類似尺度として、二つの語の出現についての条件付き確率の平均値を利用する。この類似尺度は式 (1) で定義できる。

$$\text{similarity}(x, y) = \frac{P(x|y) + P(y|x)}{2} \quad (1)$$

$P(x|y)$ は、語 y の出現ときに、語 x が同時に出現する条件付き確率である。

3.2 検索要求とのマッチング

本手法では、ユーザから与えられた検索要求を元に、各ユーザについて検索要求との関連度を計算し、評価値の高い n 個のユーザを選択する。以下で、検索要求とユーザの関連度を計算するためのアルゴリズムについて述べる。

本手法では、検索要求には複数の語が含まれることを前提とし、検索要求中の語は関連が強いという仮定に基づく。関連の強い語は、グラフ構造をしたシソーラス中では隣接しているか、近い距離でリンクがはられている。本アルゴリズムでは、検索要求中の二つの語のシソーラスにおける距離を利用し、検索質問中の語同士の関係の強さの平均を計算する。平均値が高いほど検索要求とそのユーザが適合しているといえる。

検索要求を $q = \{t_1, t_2, \dots, t_n\}$ とする。 t_i は検索要求中の語とする。 Q を q の直積 $Q = q \times q$ とする。このとき、 $\forall (t_i, t_j) \in Q$ について $t_i \neq t_j$ であれば、式 (2) で語同士の関係を計算する。

$$r_{t_i t_j} = \frac{tf_{t_i} + tf_{t_j}}{\text{distance}(t_i, t_j)} * \text{similarity}(t_i, t_j) \quad (2)$$

tf_{t_i} はユーザが持つファイルに語 t_i が出現した回数、 $\text{distance}(t_i, t_j)$ は自動構築されたシソーラスにおける語 t_i と t_j の距離、 $\text{similarity}(t_i, t_j)$ は、自動構築されたシソーラスにおける語 t_i と t_j の関連度である。

最終的に、 Q のすべての要素に関して計算された $r_{t_i t_j}$ の平均値が、検索要求に対する人間の評価値とされ、その値が上位 n 人を Know-Who 検索の結果として提示する。

4 まとめ

本論文では、グループにおける研究活動を支援するためのシステムである Papits の概略を説明し、その中で提供されているサービスである Know-Who 検索機能の実現について説明した。Papits では、ユーザの作業ファイルと管理されている文献の情報からユーザモデルを構築し、検索要求とユーザモデルとのマッチングを実現することにより Know-Who 検索を実現した。本システムでは、存在の情報を活用することにより実現されているため、ユーザに特別な作業をしてもらう必要がないよう設計されている。

参考文献

- [1] 林真暢, 後藤将志, 大園忠親, 新谷虎松: “サーベイ支援システム Papits における論文情報推薦エージェントの試作”, 第 62 回情報処理学会全国大会論文集 (3), 情報処理学会, 2001.
- [2] 藤巻伸洋, 後藤将志, 大園忠親, 新谷虎松: “知識共有支援システム Papits における推薦機構について”, 平成 13 年度電気関係学会東海支部連合大会論文集, 2001.(掲載予定)

¹<http://www.namazu.org/>