

文書作成過程の自動蓄積とその文章化

奥村 明俊 佐藤 研治 池田 崇博
日本電気株式会社 C&C メディア研究所

業務文書の作成過程において、作成者は、種々の情報源にアクセスして参照・引用できるコンテンツを獲得する。本稿では、文書作成過程を文書作成ノウハウと位置付け、エージェントによって自動抽出・蓄積し、一般化して文章として提示する手法を提案する。エージェントは、作成者のマシンに常駐して、文書作成に利用されたコンテンツとその利用個所を抽出し、ノウハウとして作成文書と共に保存する。このように蓄積されたノウハウは、他の作成者に対する文書作成支援情報として提示される。また、コンテンツに関する階層的な知識ベースと鳥瞰要約の手法によって一般化し文章として生成される。このエージェントは、グループで試験運用され、文書作成作業の効率化、著作権管理、発想支援において効果を確認した。

Documentation Know-how Extraction by Automatic Process Tracking

Akitoshi Okumura Kenji Satoh Takahiro Ikeda
NEC C&C Media Research Laboratories

In producing office documentation, one obtains information by accessing several information resources. This common documentation process can become helpful to others if collected and shared as documentation know-how. This paper proposes an agent system which extracts documentation know-how by automatic process tracking, and a generalization method of the know-how based on bird's-eye summarization. The agent stores information used for making documents by watching user operation. The information can be used by others when they make similar documents. The proposed agent was tested in an office group and proved useful for documentation support, copyright management and enhancing user creativity.

1. はじめに

文書の電子化と情報検索技術の発展に伴い、組織やグループにおいて情報を共有し、類似文書や関連文書を活用して文書を作成することが、ごく一般的に行われている。従来、文書作成の共有・協調作業支援の為の技術として、グループ作業を目的としたトップダウンなアプローチが提案されている[1][2]。たとえば、電子メールをベースとした共同執筆[3]やワークフロー等のシステム[4][5]、これらを統合したもの[6]である。トップダウンなアプローチは、共有すべき情報を手順として明示するための作業が必要となる。文書作成の場合、トップダウンアプローチは、明示化容易な定型的な文書の作成支援としては有効であるが、様々なバリエーションのある文書に関して、手順を一々記述・更新することは現実的に極めて困難である。一方、情報検索は、過去の事例をベースとしたボトムアップなアプローチを可能としている。参考となる文書データを大量に蓄積すれば、様々なタイプの文書に対してロバストに類似文書・関連文書を参考情報として提供可能である。しかしながら、新規に文書を作成する者にとって、必要な情報は、検索された文書自身ではなく、その文書がどのように作成されたかという、いわゆる、文書作成に関するノウハウである。一般に、文書を作成する際、さまざまな情報源にアクセスして必要なコンテンツを獲得する。例えば、オフィスのネットワーク環境下で、文書を作成する場合、図1に示すように、他の類似文書の参照・引用、Webのような外部データベースの検索、他の人に電子メールで記入すべき内容を質問することなどが一般的に行われる。

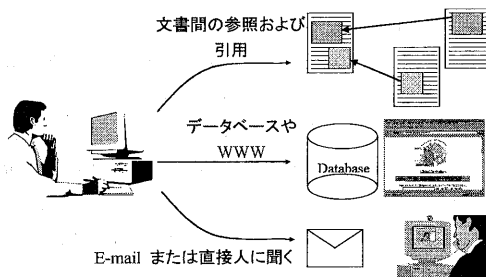


図1: 文書作成ノウハウの情報源アクセスモデル

このように、文書は、一連の情報アクセスによって得られたコンテンツに修正・追加が施された結果であることが多い。参考情報として検索された文書について、構成要素であるコンテンツ毎に、どの情報源にアクセスして作成されたかを提示できれば、新規文書作成者は、その作成過程を文書作成ノウハウとして活用することができる。

ノウハウのように明示的に保存されなかった情報を蓄積・共有することは、作業の生産性と信頼性を向上させる上で極めて重要である[2]。しかしながら、文書作成者にとって、文書作成が本来の目的であって、共有すべきノウハウ情報の記録を強いることは余分な負荷となり、実際にはノウハウが蓄積されないことになる。また、ノウハウが共有情報として蓄積されたとしても、ノウハウ情報参照・選択において、文書作成のための思考や操作を妨げてはならない。

これらの問題を解決するために、文書作成者の操作をクライアントマシンに常駐してウオッチし、文書作成過程を自動抽出・提示するノウハウ共有エージェントを提案する。さらに、エージェントによって獲得・蓄積されたノウハウを一般化して文章の形で提示する手法について述べる。

まず、ノウハウ共有エージェントの機能と動作例について述べる。次に、蓄積されたノウハウを文書に関する階層的な知識ベース(オントロジ)を用いて一般化・文章化する手法について説明する。最後に、ノウハウ共有エージェントを実装しグループで試験運用した結果について報告する。

2. ノウハウ共有エージェントの機能

ノウハウ共有エージェントは、以下の3つの機能を有する[3]。

- 1) 文書の作成作業履歴の取得 - ノウハウ抽出:
文書作成中に行われる、文書の参照および引用、インターネットなどによるデータベースへの問い合わせ、電子メールによる人への問い合わせという情報獲得操作を自動的に抽出し、獲得した情報と情報源を文書作成ノウハウとして文書ごとに管理する。

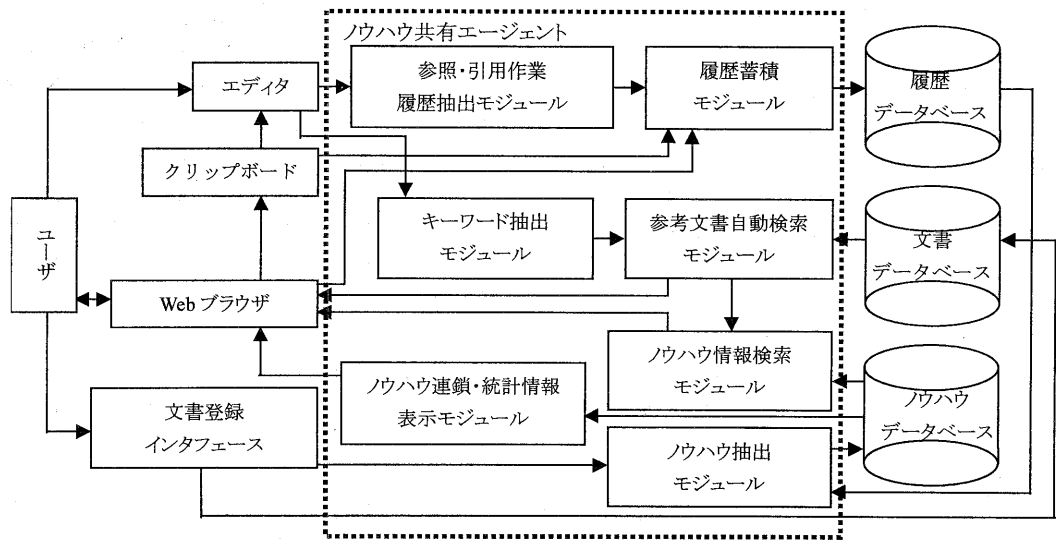


図2: ノウハウ共有エージェントのモジュール構成

- 2) 文書作成支援 - ノウハウ提示:
文書作成者に対して、作成中の文書に類似する文書を自動的に検索し、参考文書としてそれが作成された際のノウハウ情報とともに提示する。
- 3) 統計情報表示 - ノウハウ連鎖:
文書作成に最も有効なノウハウを選択するために、ノウハウ情報を、連鎖的な文書間関係を示すハイパーリンクとして表示し、また、参考文書に対する作成者、利用者、利用頻度など利用状況を提示する。

これらの3つの機能を、モジュール構成図(図2)に従って詳細に説明する。

2.1 文書の作成作業履歴の取得・ノウハウ抽出

利用者が、テキストエディタ/ワープロ等を用いて文書を作成している時、履歴抽出モジュールにおいて、他の文書の参照/引用作業をチェックし、その作業履歴を履歴データベースへ保存する。現在の実装では、Webブラウザおよびメールソフト上での参照および引用をチェックしている。保存する情報は文書の文字列・アドレス(URL, From)・作業者・時刻・端末情報等である。

利用者が文書作成を完了して文書データベースに登録する時、ノウハウ抽出モジュールにおいて、履歴データベースの作業履歴と登録文書の照合を行う。その

文書を作成していた時点での他文書の参照や引用が実際に関連のあったものかどうかを判定し、ノウハウデータベースへ文書間の相互リンクとして保存する。参照や引用の判定は、文書間における単語の一致度を用いて行う。一定割合以上の単語が一致している時、関連ありとして双方向のリンクを生成する。図に関しては、完全に一致している場合のみ関連ありと認定する。

ノウハウデータベースに保存される情報としては、履歴データベースに保存されている情報に加え、相互リンク・参照/引用個所を保存する。参照/引用個所については、認定時におのおのの文書中の前後範囲をも含めて認定し、その個所を保存する。履歴データベース内の履歴情報は、ノウハウ抽出モジュールで参照/引用関係が認定されたものは消去される。また、認定されなかった履歴については、他の文書作成についてのノウハウが抽出できる可能性がある為、現在のところは消去しないようにしている。

これらの履歴取得・ノウハウ自動抽出機能により、利用者に文書作成以外の余分な作業を強わずに、文書作成ノウハウの自動抽出が可能となっている。

2.2 ノウハウ提示・文書作成支援

利用者が、テキストエディタ/ワープロ等を利用して文書を作成している時点で、キーワード抽出モジュールにおいてその文字列入力をチェックし、自動的に入力文字列からキーワードを抽出する。その抽出したキ

ワードを用いて、参考文書自動検索モジュール、ノウハウ情報検索モジュールにおいて、文書データベース内の文書を自動検索し提示する。文書の提示は Web ブラウザ上で行われる。その文書提示の際に、その文書に関連する他文書への相互リンクを併せて提示することで、文書作成ノウハウを、利用者による検索作業を行わずに提供している。相互リンクを張られている文書の参照は、Web ブラウザ上のリンクのクリックにより行える。

キーワードによる検索は、文書に対してはフルテキストサーチを行っている。また、入力文字列からのキーワードの抽出に関しては、キーボードからの連続入力をチェックし、一番後(最近)に入力された単語の重みが最大になるように重み付けを行って検索キーとして用いる¹。このノウハウ提示・文書作成支援機能により、文書検索が行われる。利用者は、過去の類似文書を探すために検索インタフェースを起動したり、キーワードを打ちこんだりする必要はない。その結果、文書作成作業に集中することができる。利用者は、提示された文書とそのノウハウをブラウズし必要なものを利用することにより、文書作成を容易に行うことができる。

2.3 ノウハウ連鎖・統計情報表示

ノウハウが複数ある場合、どのノウハウを選択すればよいかを判断する手がかりが必要である。例えば、文書Aに引用されている文書Bが更に引用している文書Cといった連鎖的な情報や誰がどの程度引用している情報かといった統計的情報は、複数のノウハウ情報がある場合の有効な選択情報となる。これらの情報を提供するのが、ノウハウ連鎖・統計情報提供モジュールである。

本モジュール機能の起動は、自動提示された文書の上のアイコンをクリックすることで起動され、その文書を中心として相互リンクを辿って上下2リンクまでの文書が表示される。また、表示されている任意の文書を選択して中心へ移動することにより、相互リンク関係を辿りながらリンクの全体像を把握することが可能である。現プロトタイプシステムでは、Web ブラウザのプラグインソフトとして実現されている。

また、本モジュール機能では、文書データベース全体の相互リンクの統計情報も提供する。提供する統計

情報は、各文書毎の参照／引用数、被参照／被引用数、および各著者毎の登録文書数・総参照引用数・総被参照被引用数・1文書毎参照引用割合・1文書毎被参照被引用割合である。これらの連鎖・統計情報により、多くの人が参照している文書を探すことができ、参考とすべき文書の選択が可能となる。また、著者毎の統計情報は、多くの人に利用される文書の作成者を明らかにし、従来認知されにくかった業務への貢献を明確化する効果がある。

3. ノウハウの文章化

本章では、蓄積されたノウハウを一般化して文章とする手法について述べる。

ノウハウ共有エージェントは、ノウハウをハイパーリンクとして提示するが、より直感的に理解しやすい形で提示するために、鳥瞰要約の手法を用いて一般化し文章として提示する。鳥瞰要約は、複数のイベント記述文から抽出された5W1H情報(いつ、どこで、誰が、何を、どうした、..)をシソーラス上の上位概念を用いて、要約文を生成する手法である[4]。図3は、「NECがPC98NXを発売」「××通信機が次世代パソコンを開発」「××電気が携帯パソコンを発売」「××通信が仮想商店街を開発」「××電話がインターネットサービスを強化」という5文に対する鳥瞰要約の例である。各文からWho要素とWhat要素を抽出し、それぞれの要素に対してシソーラスの上位概念を照合する。上位概念を「誰が何をどうした」という文書テンプレートにあてはめると、これら5文に対して、例えば、「電機企業3社がパソコンを開発・販売した」、「通信企業2社がネットワークサービスを開設・強化した」という鳥瞰的な要約を生成することができる。

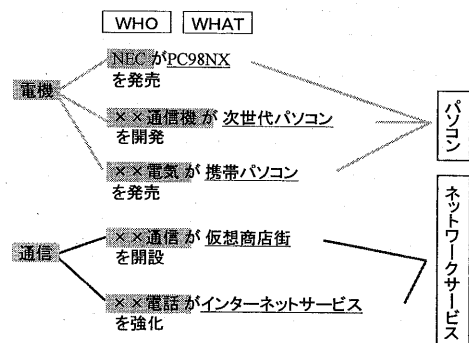


図3 鳥瞰要約例

¹ ある時点より過去に入力されたキーワードについては重みwを0とし、その時点で検索キーとして用いられなくなる。

ノウハウ共有エージェントで記録されているデータは、文書作成に関するイベント記述である。例えば、「文書Xが文書Yを引用した。」という記録の集合に関して、文書Xと文書Yの種別に関する階層的な知識ベース(オントロジ)を用いて、文書の引用に関する一般的な傾向を表現することができる。さらに、各文書の属性として記録されている、作成者、引用個所、引用情報源に関するオントロジを用いることによって、より詳細な鳥瞰要約が可能である。様々な属性の組み合わせでノウハウ情報を一般化・文章化できるが、ノウハウとして文章化する場合には、文書集合に対して何を引用したかという観点からの一般化と、個々の文書についてどのように引用されているかという被引用の観点からの一般化が効果的である。

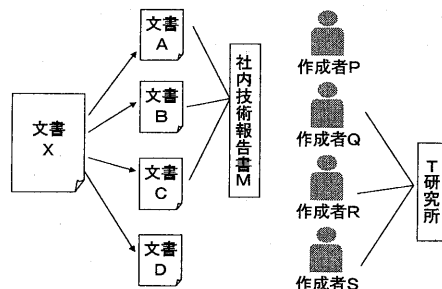


図5 被引用情報の鳥瞰要約

1) 文書集合に対する引用情報の一般化

例えば、図4のように、ノウハウ情報として文書集合、項目とその情報源が得られた時、オントロジを用いて共通するノウハウ、「購入要求書の目的の項目は、<http://www.xyz.com>を参照すること」を生成することができる。利用者が作成する文書の種類が決まっている場合、あらかじめ提示する情報として有効である。

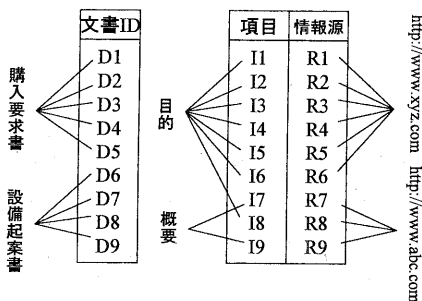


図4 ドキュメントオントロジによる鳥瞰要約

2) 個々の文書に対する被引用情報の一般化

例えば、図5のように、ある文書Xが、文書A(作成者P)、B(作成者Q)、C(作成者R)、D(作成者S)に引用されており、文書A、B、Cが社内技術報告書Mに投稿されているとき、「文書Xは、社内技術報告書Mでよく引用された。」と一般化できる。また、作成者P、Q、R、Sの所属がT研究所である場合、「文書Xは、T研究所所属者によく引用された。」という文章が生成できる。参考文書として、多くの文書が利用者に提示された場合、このように一般化された要約は、どの文書を参考にしたらよいかの選別情報として有効である。

4. ノウハウ共有エージェント動作例

2章で説明したノウハウ共有エージェントの動作をパソコンの購入要求書作成を例にとり説明する。図6. 1から図6. 8までが、その一連の画面イメージである。

まず、ユーザがエディタを開き(図6. 1左ウィンドウ)タイトルや必要項目を入力する。すると、ノウハウ共有エージェントがその入力を把握し、参考とすべき文書を文書サーバ中から自動的に検索し、その検索結果をユーザへ提示する(図6. 1右ウィンドウ)。提示される情報は、参考文書のリストおよびそのリストの中で最も優先されている文書の内容、および最優先文書の作成ノウハウ情報である。ここでは、最優先文書として「電子電話帳システム」に対する購入要求書が表示されており、その購入要求書の作成者が文書を作成する際に ValueStar について書かれたホームページ等を利用して文書を書いたこと等が表示されている。これらの情報から、ユーザは自分が書こうとしているシステムでも ValueStar をパソコンとして利用すればよいことが判る。また、その ValueStar に対する情報がどこにあるかをユーザが知らなくても、ノウハウとして表示されている ValueStar の行をクリックするだけで、そのホームページを開いて参照することが可能となる(図6. 2)。こうして入手したパソコンの価格や仕様を、自らが書いている購入要求書に書き込むため、ユーザは次にそのホームページの必要部分をコピーし(図6. 3)、購入要求書内にペーストする(図6. 4)。これらの一連のユーザ作業は、ノウハウ共有エージェントにより把握され、ユーザがパソコンのホームページを利用したという履歴が自動的に蓄積される。購入要求書を完成し、その文書を保存すると(図6. 5)、ノウハウ共有エージェントによって、保存した文書と利用した文書(この場合はパソコンのホームページ)の間に

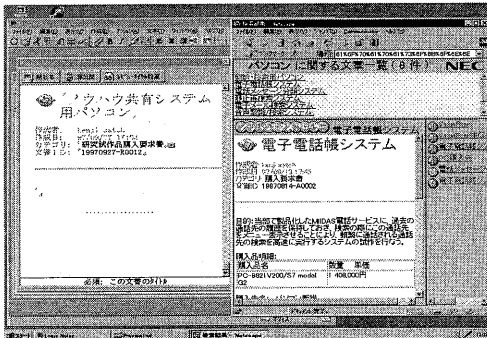


図6. 1: 文書作成支援例(1)
参考文書提示



図6. 2: 文書作成支援例(2)
ノウハウリンクの利用

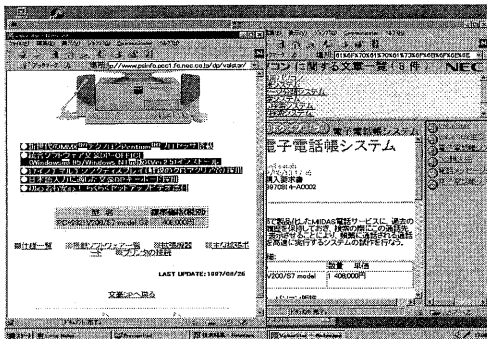


図6. 3: 文書作成支援例(3)
文書の引用(コピー)

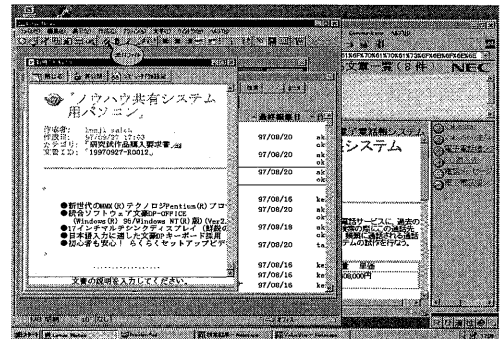


図6. 4: 文書作成支援例(4)
文書の引用(ペースト)

「引用関係」のノウハウが自動的に生成され保存される。保存された文書の作成ノウハウは、文書の保存後にすぐに有効となる。ここで、例えば、先ほどの購入要求書を作成したユーザや他のユーザが、新たな文書を書く場面で、この購入要求書が参考文書として提示された場面を想定する。その表示画面(図6. 6)を見ると、購入要求書のノウハウ情報として、ValueStar のホームページを引用したという情報が表示されており、また、そのノウハウ情報をクリックするだけで、ValueStar のホームページも、新しい購入要求書からすぐに辿ることが可能となっており、文書作成に対するノウハウが自動的に蓄積されていることが判る。

図6. 7において文書の上に表示されている「ノウハウエクスプローラー」という名前のアイコンをクリックすることで、表示されている文書を中心とした「連鎖・統計情報提示インタフェース」が表示される(図6. 7)。この連

鎖・統計情報提示インタフェースでは、Webブラウザ上に表示されている文書を中心のアイコンとして、左側に被参照被引用文書のアイコンが、右側に参照引用文書のアイコンが、それぞれ2階層まで表示されている。この表示されている任意のアイコンを右クリックすると「中心へ移動」メニューが表示され、それを選択することでこの連鎖情報を左右へスクロールし、文書間の連鎖の全体像を把握することが可能である。また、任意のアイコンをダブルクリックすると、Webブラウザへその文書および文書作成ノウハウが表示される。このインタフェースを利用することで、情報の発生源となった文書の検索や、更新されつつある情報中、最新情報の取得が可能となる。また、この連鎖・統計情報提示インタフェースのメニューより統計情報の表示を選択すると、文書データベース全体の文書作成ノウハウ情報に対する統計情報が提示される(図6. 8)。

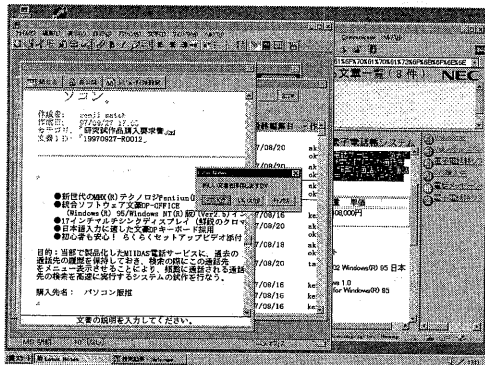


図 6. 5: 文書作成支援例(5)
文書の保存(ノウハウ抽出)

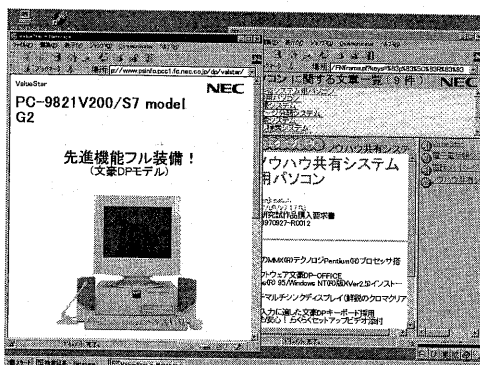


図 6. 6: 文書作成支援例(6)
新規作成ノウハウリンクの利用

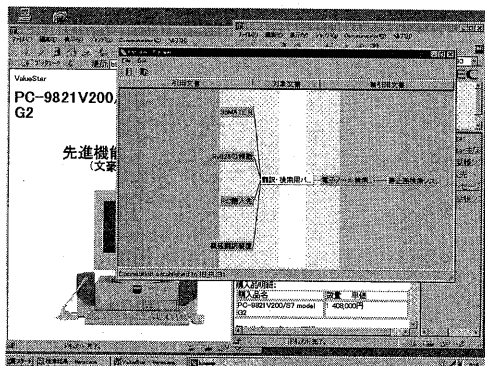


図 6. 7: 文書作成支援例(7)
ノウハウ連鎖情報の利用

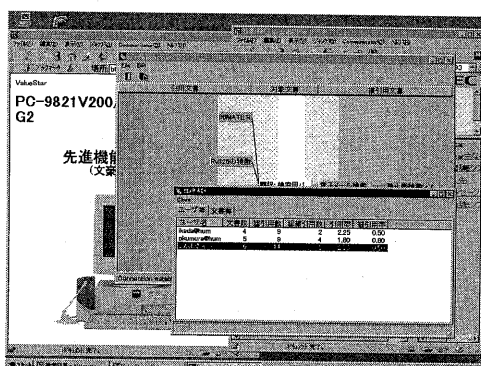


図 6. 8: 文書作成支援例(8)
ノウハウ統計情報の利用

統計情報の提示では、文書毎の統計情報と、著者毎の統計情報が表示可能である。図 6. 8では、著者毎の統計情報として、各著者毎の登録文書数・総参照引用数・総被参照被引用数・1文書毎参照引用割合・1文書毎被参照被引用割合、が表示されている。この情報を見ることによって、「作成した文書がよく人に利用される著者」という情報を得ることが可能で、ノウハウの選択をより容易に行うことが可能となっている。

5. 考察

ノウハウ共有エージェントのプロトタイプシステムを構築し、試験運用を行った。プロトタイプシステムは、文書および履歴・ノウハウデータベースとして Lotus Notes を利用し、Webサーバプログラムとしては Notes DOMINO Server を利用している。また、入力文字列からの文書自動検索プログラム、および連鎖・統計情報提示インターフェースは、クライアントとなる Windows マシン上で稼動する独立プログラムとし、これらがWebサーバ上のCGIプログラムおよび Notes クライアントプロ

グラムと通信することでデータの受け渡しを行っている。対象文書のデータタイプとしては、テキスト文書、Notes 文書、HTML文書を対象としている。

研究者5人のグループにおいて、購入要求書、特許提案書、定期活動報告書の合計500文書程度で本プロトタイプシステムを利用し、以下のような定性的評価をえた。

1) 文書作成作業の効率化

利用者に特別な作業を課することなくノウハウを抽出できた。また、文書作成時には、参考となる類似文書を自動的に検索してノウハウとともに表示することが、有効な支援となることを確認した。ノウハウ抽出と自動検索は、利用者に負荷をかけずにおこなわれるので、利用者は文書作成作業に集中できる。反面、自動検索は、常に画面にその結果が動的に表示されるので、見づらい場合もあった。今後は、利用者の指定によって表示モードを切り替えるなどの工夫が必要である。

2) 著作権情報・個人情報管理

ノウハウ共有エージェントは、文書の著作権情報を文字単位に管理しており、文書の一部を他の人が利用すると、元の文書の著作者はその事実を容易に知ることができる。著者毎の統計情報は、多くの人に利用される文書の作成者を明らかにし、従来認知されにくかった業務への貢献を明確化する効果があった。このように、著作権情報を自動管理することで、個人の文書とノウハウを安心してサーバに登録することができた。しかしながら、個人の文書作成過程をノウハウとしてすべて登録することは、利用者にとって違和感を覚える場合もある。利用者が文書を保存するときに、ノウハウも登録するかどうかの確認機構を設け、最新の注意を払っていく。

3) 発想支援

特許作成において、自分の作成した特許に関する引用状況が、発想支援に有効であることを確認した。利用者は、本システムによって、自分のアイデアが他人によって自分が気づけなかった別の分野や他の目的に利用されていることを容易に知ることができる。その結果、本人自身が、さらに新たな応用を思いつくことがあった。この効果は、Batesが提案している進化的探索モデルにおける引用元の参照(citation chasing)の効果と同様のものと考えられる[13][14]。

6. おわりに

本論文では、文書作成において利用された情報とその際の操作をノウハウ共有エージェントによって自動獲得して文書作成の支援とする手法と、ノウハウを一般化・文章化して提示する手法を提案した。ノウハウ共有エージェントは、プロトタイプを装束し、文書作成作業の効率化、著作権情報管理、発想支援などにおいて効果があることを確認した。今後、より大規模なグループで運用・評価するために、支援ツール等の環境整備を進めていく。

ノウハウの文章化に関しては、作成文書の種類、目的、利用者の選択基準などによって鳥瞰要約の対象とする属性を制御する必要がある。今後、文章化ノウハウ連鎖・統計情報表示機構の中に文章化機構を搭載して、実証・評価を行う予定である。

参考文献

- [1] Ellis, C.A., Gibbs, S.J. and Rein, G.L.: Groupware Some Issues and Experiences, *Comm. ACM*, Vol. 14, No.1, pp. 33-58 (1991).
- [2] Ackerman, M.S. and Mandel, E.: Memory in the Small: An Application to Provide Task-Based Organizational Memory for a Scientific Community, *Proc. of Hawaii International Conference of System Sciences*, Vol. IV, pp. 323-332 (1995).
- [3] Fish, R.S., Kraut, R.E. Root, R.W. and Rice, R.E.: Evaluating Video as a Technology for Informal Communication, *Proc. of CHI '92*, pp. 37-48 (1992).
- [4] Froles, F., Graves, M. et al.: Computer System and the Design of Organization Interaction, *ACM Trans. Office Information System*, Vol. 6, No. 2, pp. 153-172 (1988).
- [5] 市村 哲, 松下 温: 発言と行動の管理に基づいた協調作業支援電子メールPilot Mail, *情報処理学会論文誌*, Vol. 33, No. 7, pp. 955-963 (1992).
- [6] 垂水 浩幸, 田淵 篤, 吉府 研治: ルールベースの電子メールによるワークフローの実現, *情報処理学会論文誌*, Vol. 36, No. 6, pp. 1322-1331 (1995).
- [7] Arai, T., Aust, D. and Huson, S.E.: PaperLink: A Technique for Hyperlinking from Real Paper to Electronic Content, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'97)*, pp. 327-334 (1997)
- [8] Davis, H.C., Hall, W., Heath, I., Hill, G., Wilkins, R.: Towards an Integrated Information Environment with Open Hypermedia Systems; *Proceedings of the ACM Conference on Hypertext*, ACM Press, 181-190 (1992).
- [9] 門脇 千恵, 爰川 知宏, 他: 情報取得アウェアネスによる組織情報の共有促進支援, *人口知能学会誌*, Vol. 14, No.1, pp. 111-121 (1999).
- [10] Terveen, L., Hill, W., Amento, B, et al: PHOAKS: A System for Sharing Recommendation, *Communications of the ACM*, Vol. 40, No. 3, pp. 59-62 (1997)
- [11] Satoh, K., Okumura, A. and Muraki, K.: Documentation Know-how Sharing by Automatic Process Tracking, *Proc of International Conference on Intelligent User Interfaces*, pp.49-56, 1999
- [12] 奥村明俊, 他: 5W1H情報抽出・分類によるテキスト要約, *自然言語処理学会誌*, Vol.6, No.6, pp.27-44, 1999
- [13] Bates, M. J. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface, *Online Review*, 13 (5), 407-424.
- [14] 徳永健伸: 情報検索と言語処理, *言語と計算*5, 辻井潤一編, 東京大学出版会, pp.169-171, 1999