

O-011

即時的なドキュメント共有を行う情報システムの構成法

A Construction Method of Information System Immediately Share Documents

飯田雅広†

Masahiro Iida

千葉貴史‡

Takashi Chiba

津田和彦†

Kazuhiko Tsuda

1. はじめに

企業内で作成するドキュメントを電子化することは、すでに一般的になっている。同時に文書管理システムなどが提案、販売されている。このようなシステムで必要とされるメタデータの付加工程などは、ドキュメントが多数に上る場合、作業量が膨大になるという欠点がある。また共有の是非を作成者が判断することは、無駄な情報流通を抑えるという利点があるが、情報が故意に隠蔽される可能性がある。さらに組織内の知識の創造を活性化するという点では、その個人にとって必要ないと思われる情報でも共有されていることの重要性(情報冗長性)が指摘されている。

Sanders⁵⁾は情報の評価基準として、情報が必要なとき存在している価値(適時性)を指摘するが、共有までのタイムラグがある場合、適時性は低下する。同時に企業組織内の非定型なコミュニケーションにおいては、遠いオーソリティより身近な人間関係が優先され、結果として貴重な知識が個人の中にとどまっているという指摘がある¹⁾。

このような中、本研究ではこれまで共有されなかった情報として個人の端末内に格納されているドキュメントに着目し、1)共有の効率化による情報適時性の底上げ、2)網羅的な共有による情報冗長性の拡大、3)端末間のドキュメント関連情報による気付き(awareness)とコミュニケーションの促進、を目指している。またこのドキュメント共有システムを実現する手段として、分散環境下における統合的な形態素解析の実現方法を提案する。

2. 情報共有

90年代初めに提唱されたBPR²⁾(Business Process Reengineering)では、現場で発生する情報を共有することで担当者が業務プロセス全体を理解し、管理者に代わって高度な意思決定が可能になるとされた。しかしこうした合理化は人間を機械とみなす傾向があり、田坂⁶⁾は日本企業内部のアンバランスな世代構成の是正とあわせて、中間管理職のレイオフが進んだと指摘する。同時に、BPRが進んだ組織での管理職の機能として、i)膨大なデータから要点をつかむ、ii)知識を言葉として伝える、iii)ノウハウの共

有を進める、などとした。

こうした組織内部の知的生産性の向上を強調する議論はKM(Knowledge Management)として90年代から議論されている。寺本⁷⁾は、KMにおける知の創造プロセス(SECIモデル)において特に言語化可能な形式知の内面化プロセスを強調する。また、情報冗長性を高めるために、言語で記述可能な知識が多く共有されている必要性を指摘し、これが個人の内省や経験を経て新たなイメージの観点になっているとしている。

KMを情報科学の手法を用いて実現しようとする研究、あるいはCSCW(Computer Support Cooperative Work)領域の研究は多数存在する。中山ら⁴⁾は、システムへの情報登録を暗黙知の形式知化とし、蓄積するコンテンツの量と質に焦点を当てた。しかし情報提供者が自身の持つ情報の価値を判断できるときのみ共有の場に提供するという、提供者と参照者のマッチングに関して課題を残した。

Dignum¹⁾は、知識共有において組織の目標と個人の要求が必ずしも一致しない点を指摘し、知識リポジトリへの情報登録と参照をエージェントによるモデルで仲介することで組織内のマッチメイキングを行う手法を提案した。エージェントが組織の目標や規定、従業員間の交流における規範などを考慮することで円滑な知識交換を目指す、形式知の共有に偏りすぎている。

3. システム構成

本研究では、言語で記述された形式知としてのドキュメントに着目し、見えない形式知である個人の端末内に保存

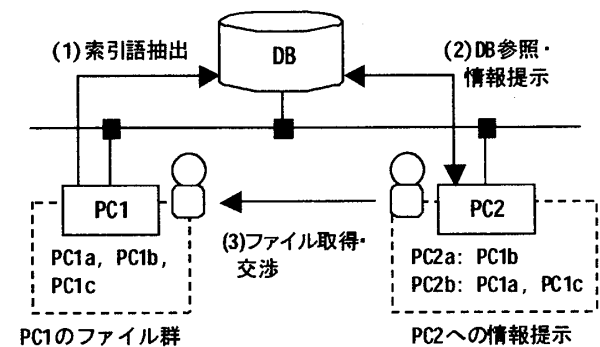


図1. システム構成

† 筑波大学大学院ビジネス科学研究科, Graduate School of Systems Management, University of Tsukuba

‡ (株) 建築ピボット, archipivot inc.

されているドキュメントを扱う。具体的には LAN 内の各端末から DB サーバーへ、ドキュメント情報を自動的に登録する仕組みを提案する。また自らの端末にあるドキュメントと類似した他端末のドキュメントを提示することで、これまで知ることが難しかった組織内での知識の所在を提示する(図1)。

これによって、1章で述べた課題に対する解決策として、1) 必要なときに必要な情報がある適時性、2) 形式知の内面化プロセスや情報交換に十分な情報冗長性、3) 所在の特定が困難な知識への気付き、を実現する。

3.1. 索引語の登録部

端末内に、ドキュメントの変更を監視する常駐プログラムを配置し、このプログラムにより即時的に形態素解析を行って索引語を抽出し、LAN 内のデータベースサーバーに登録する。これにより従来の文書共有システムの欠点であった登録作業の手間を解消し、かつ網羅的な共有を実現する。

また索引語は固有名詞だけでなく、文書の性質をより説明すると思われる複合語を検出する。具体的には、名詞/動詞/形容詞/未知語など一定品詞の連続を複合語として、索引語とする。例えば「ビジネス|科学|研究|科」では接続する形態素から索引語を10語抽出できる(図2)が、サブセットになり得る高頻度の索引語(例:「科学研究」)も抽出することにより、過剰抽出への対策を行っている。

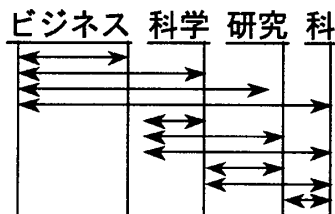


図2. 複合語の検出例

3.2. 関連情報の提示部

取り出した索引語から文書間の類似度をTF/IDF指標による余弦³⁾で計算し、各端末に対して他端末にある類似文書を提示する。提示した類似文書の取得は、文書提供者のポリシーによって可否を決めるが、共有拒否の場合でもその後の依頼次第で共有の道が開けると考えられる点から、参照者と提供者の通信手段を提供する。

3.3. 形態素辞書の問題と解決手法

3.1 節で提示した複合語検出はコストの高い手法であり、可能であれば形態素解析の処理段階で形態素として抽出したい。この場合、複合語を形態素辞書に登録する必要があるが、各端末で分散して形態素解析を行う本システムでは、辞書の共有が大きな問題となる。

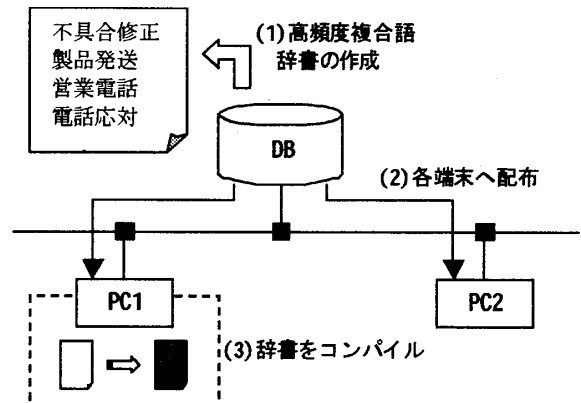


図3. 辞書共有の概略

ここでは複合語検出処理で取り出した複合語のうち、頻度の高いものを形態素辞書に追加し、この辞書を各端末へ差し戻す手法を提案する。これによって各端末での索引語抽出を均質化し、かつ文書集合全体の索引語のトレンド変化にも対応できる。

4. おわりに

見えない形式知として個人の端末に保存されたままの文書を即時的に登録・共有し、他端末の文書と比較して類似文書を提示するシステムの提案を行った。また分散環境下の形態素解析で問題となる辞書の共有について解決手法を提案した。今後は、提案する構成法を企業に適用し、システム開発と評価を行う予定である。

参考文献

- 1) Dignum V.: Personalized Support for Knowledge Sharing, Dutch Directions in HCI, SIGCHI.NL, (2004)
- 2) Hammer M. and Champy J.: Re-engineering the Coporation: A Manifesto for Business Revolution, Harper Business, New York, (1993)
- 3) 北研二, 津田和彦, 獅子堀正幹: 「情報検索アルゴリズム」, 共立出版, (2002)
- 4) 中山康子ほか: 知識情報共有システム(Advice/Help on Demand)の開発と実践: 知識ベースとノウハウベースの構築, 情報処理学会論文誌, Vol.39, No.5, pp.1186-1194, (1998)
- 5) Sanders, Donald H: "Computers in Business: An Introduction" 2nd Ed., MacGraw-Hill, New York, (1972)
- 6) 田坂広志: 「なぜ日本企業では情報共有が進まないのかーナレッジ・マネジャー7つの心得ー」, 東洋経済新報社, (1999)
- 7) 寺本義也: 「学習する組織」, 同文館出版, (1993)