

情報間の関係性管理に基づく企業内情報構造化方式 情報シナプスの構想

坂上 秀和[†] 本橋 洋介[†] 坂口 基彦[†] 内藤 圭三[†]
[†]NEC サービスプラットフォーム研究所

企業のみドルマネージャたちは、1日に2時間、就業時間の1/4以上を情報検索に費やしており、しかも得られた情報で有用なものは半分に満たない、という情報収集や活用に関する調査結果がある。検索技術の急速な進歩にも関わらず、企業内で業務に必要な情報を素早く検索し、有効に再利用することの困難さに変化はない。本稿では、企業内で生成される情報・知識の再利用性を高める方式として、情報間の関係性に基づく情報構造化方式を提案する。我々は、企業内の情報活動全体を、情報の生成・活用・価値創出・消滅のライフサイクルが営まれるエコシステムと捉え、データモデル・アプリケーションインタフェースなどの情報アーキテクチャと、取り組むべき研究課題について述べる。

Information SYNAPSE: A vision of structuralizing enterprise information based on their relationship

Hidekazu Sakagami[†] Yosuke Motohashi[†] Motohiko Sakaguchi[†] Keizo Naito

[†]NEC Service Platforms Research Laboratories

According to a survey of information gathering and utilizing, middle managers spend up to two hours a day, which corresponds to more than a quarter of their time, searching for information, and more than 50 percent of the information they obtain has no value to them. Despite a rapid growth in the information retrieval technologies, it remains still difficult to find information they need for their work promptly, and utilizing them effectively. We propose a method of structuring enterprise information for increasing their reusability based on their relationships. We consider the enterprise information activities as an ecosystem in which information lifecycle of creation, use, utilization, and disposition is carried out. Based on this viewpoint, this paper discusses on the information architecture such as data model and API, and also refers to research issues.

1. はじめに

計算機やインターネットの普及は情報の爆発的な増大をもたらしたが、近年はキーワード検索などの技術の進歩により、ユーザがインターネット上や自分のPC上に保存された情報の検索などでは、以前と比較すると圧倒的に迅速に大量の情報を取得できるようになっている。

一方、ナレッジワーカーと呼ばれる企業内の知識労働者たちの生産性について目を向けると、企業内の情報検索・活用について、検索に多大なコストがかかり、十分に活用できていないという調査結果がある[1,2]。多くの企業は情報管理の基本的な仕組みとして階層型のファイル管理方式が用いられているが、その限界についてはBush[3]の時代から認識されており、情報の検索性や再利用性を高めるための試みとして、Memex[3]のように情報間の関係性に基づく管理モデルや、情報をタスクにひもづけて管理するモデルなどが提案されてきたが、十分には普及していない。これ

らの多くは情報の管理方式のみに着目しているものである。

そこに格納されるべき情報は、企業内活動の様々な場面で生成されるものであり、また、それらの情報は、様々な場面で自由に呼び出して活用されるものである。単なるデータ管理方式の提供や、検索のための特定のアプリケーションを提供するのではなく、情報を取得する仕組みや、取り出して活用する仕組みとのインタフェースを含めて企業内の情報活動全体のアーキテクチャを設計する必要があると考えている。また、蓄積される情報は、そのまま取り出すだけでなく有用性は高いが、それらの情報を集合として解析することで重要な情報を発見するなど、さらに価値を増大させることが可能である。システムが機能し続けるためには、情報を一方的に増大させるだけではなく、価値の低い情報を消去することでデータサイズを一定内に収めるなどの仕組みも必要となる。

我々は、企業内の情報活動全体を、情報生成・情報活用・情報価値創出・情報消滅のライフサイクルが営

まれるひとつのエコシステムとして捉えている。本論文は、この観点に基づき、企業内の情報を構造化するモデルと、アプリケーションインタフェースなどの情報アーキテクチャを提案する。

2. 企業内情報検索の課題

2.1. 企業内情報検索の現状

情報検索に時間がかかることや、既存情報の再利用性が低いことは、ナレッジワーカー自身が本来担うべき、意思決定や付加価値を創出するために費やす労働量の低下につながるため、組織全体の知的生産性に多大な悪影響を与える。国際的なコンサルティング企業が、米国・英国で売上高5億ドル以上の大企業に勤務する管理職1009名に対して2007年に行った調査[1]によれば、

- 彼らは情報を探すために1日約2時間を費やし、しかも、53%の人は探し出した情報のうち有用なものは半分に満たないと考えている
- 業務に必要な社内にあるはずの情報が手に入らないと感じているが59%いる
- 自社内の他部署の情報を集めるのは至難の業であると45%が考えており、競合他社の情報が得にくい(31%)と考える人より多い
- 情報を集めるために複数の情報源を当たらなければならぬことが負荷である(57%)

などの結果が出ている。

2003年に米国の市場調査会社が行った調査[2]でも、ほぼ同様の結果が報告されている。

- ナレッジワーカーは15%-35%の時間を情報検索にあてている
- 探しているものが見つかる確率は50%以下
- 40%の企業ユーザが、イントラネットで必要な情報を見つけられないと回答
- ナレッジワーカーが報告書などの成果物を新たに作成する場合、90%の時間は既に存在する情報を作り直すことに費やされている

2.2. 企業内情報検索の課題

これらの2つの調査が行われた数年間は、Google社[4]に代表されるキーワード検索をはじめとした情報検索技術が急速に発達・普及した時期と重なっており、検索作業自体は一般的に便利になったと考えられる。しかし、企業内の情報再利用に関しては状況が改善されていないのは何故であろうか。

情報検索には、未知情報の検索と既知情報の検索の2種類がある。例えば新たに新商品開発業務を開始する場合に、市場動向に関する情報をインターネット上

のリソースからキーワード検索などによって収集する作業などは、未知情報の検索である。一方、商品企画書を作成する際に、既に収集済みの資料を検索したり、過去に作成した類似の企画書を探す行為などは既知情報の検索である。

図1は簡単な思考実験によって業務のサイクルをモデル化したものである。新たな業務を開始する際には未知の情報を収集するが、その後、収集した情報を参考に自分自身で文書の作成・修正を繰り返したり、作成した文書を共同作業メンバと共有したり、他メンバが作成した資料を参考に新たな資料を作成する際には、いずれも、未知情報ではなく、既知情報の再利用が主に行われている。

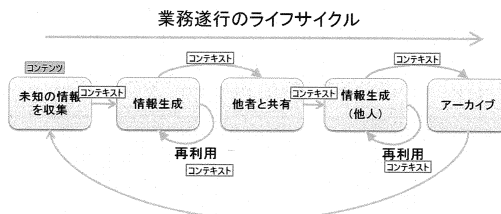


図1 業務サイクルモデルと必要な検索機能

未知情報と既知情報は、情報の探し方も異なる。

未知情報は、キーワード検索など情報の中身に基づく情報検索（コンテンツベース）が多く行われる。

既知情報は、「資料を作成しているときに参照していたファイル」「先週の会議で、Aさんが提出した資料」などのように、ある情報の断片をもとに関連する情報を取り出したい場合や、「X社向けシステム提案に関連する提案書や設計書」などのように特定の業務に関連する情報を一元的に取り出したい場合などが多い。このような場合、情報の利用や生成に関わるコンテキスト情報に基づいて情報を探すことが多く、キーワード検索などの検索手法だけでは所望の情報が見つけれなかったり、時間がかかってしまうことが多い。

また図1からも分かる通り、企業内での情報検索では、このようなコンテキストに基づいて既知情報を検索するケースが、コンテンツに基づく情報検索を行うケースよりもはるかに頻度が高く、これらの情報検索の支援は、企業やナレッジワーカーの生産性向上の上で、重要な鍵を握ると考えられる。

3. Related Work

コンテキストに基づく情報管理についての初期概念は、V.Bush[3]にさかのぼることができる。V.Bushは、人間の思考や記憶について、「脳細胞の複雑精妙な網目状経路に基づく連想作用」によって機能することを指摘し、その考え方に基いて図書館の本やフィルムを対象に、情報間のリンクを相互参照可能にする装置

Memex を考案した。この思想は後のハイパーテキストの創造に影響を与え、現代のインターネットに至っている。また、Bush は、「頻繁にたどられることのない連想経路は薄らいでいく傾向がある」など、情報間のリンクの特性やその変化に関する知見についても言及している。

近年も情報の再利用性を高める試みとして、情報間の関係性を利用した管理方式が提案されている。

Muller[5]らは、共同作業で業務を開始する際に“Activity Thread”を定義し、業務に関連する一連のオブジェクト（ファイル、メッセージ、等）を Activity に関連付けて管理することで、業務単位でのリソース共有やプロセスの把握を容易にさせる管理方式を提案している。

梅木[6]、芦川[7]らは、非定型で協調的な知的作業において、最終成果物や、成果に至る過程でやりとりされた情報を、業務の文脈の中で（返信、改訂、参考等として）関連付けて蓄積し、この情報同士のつながり（ナレッジリンクと呼ぶ）に基づいて業務ごとに経緯を可視化し、現在の業務で行うべき作業や活用すべき情報を提示する方法を提案している。

これらの手法は、いずれも情報の生成や利用におけるコンテキストに基づいて情報間の関係性を管理するデータモデルの提案であるが、管理方式に焦点が当てられており、そこで管理されるべき情報がどのように生成・蓄積されるのかという情報の蓄積部分や、様々な作業・ツールから利用されることを想定したインタフェースなどの情報の活用部分が十分には考慮されておらず、局所的な検索アプリケーションの提案などになっている。

4. 情報シナプスの提案

4.1. 情報シナプスの基本モデル

ナレッジワーカーが業務中に（既知情報を）探し出すシーンには、2.2 で述べた通り、特定の情報の断片をもとに関連する情報を取り出したい場合と、特定の業務目的に対して関連する情報を一元的に取り出したい場合がある。

前者は、ファイルやメール、会議などを情報の単位として、それらの間に張られた連想リンクのネットワ

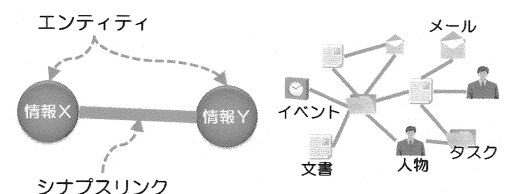


図2 情報シナプスの基本構造

ークを辿るモデルであり、後者は、業務目的などの概念をひとつの仮想的な情報単位として、その概念にひもづけられたファイルなどへの連想リンクを一括して取り出すモデルと捉えることができる。

このような連想リンク探索型の情報の呼び出し方を可能にする管理モデルとして、我々は、情報間の関係性に基づく情報構造化する方式「情報シナプス」を提案する。図2に情報シナプスの基本構造を示す。

図2に示す通り、情報シナプスは、任意の情報（エンティティと呼ぶ）同士をシナプス・リンクで結ぶ基本構造を持ち、それらがネットワーク状に組み合わさって情報空間を形成する。

エンティティには、文書・email・タスク・人物など、様々な種類の情報を定義できる汎用性を考慮している。そのため、各エンティティの実体情報（例えば文書エンティティについての文書実体ファイルなど）の管理は、そのオブジェクトを管理している他の手段（文書管理サービスなど）に委ね、情報シナプスのデータベース内では、それらへのURI情報を保持するモデルとしている。

シナプス・リンクは、エンティティ同士の関係性に関する情報を管理する。情報間の関係性を定義する点で、ハイパーリンクの一種と考えることもできるが、実体情報とは独立に管理されているため、内部にリンク情報が埋め込まれるHTML文書などは異なり、後から第三者が関係情報を付与したり変更することが可能である。また、関係性についての重み情報を定義できるほか、任意の属性（リンク更新日時や、関係性の意味づけなど）を追加拡張できる構造としている。

4.2. シナプスで管理する業務中のコンテキスト情報

情報シナプスは、業務中のコンテキストに基づいて得られる情報間の関係性を構造化して保存する役割を持ち、日常業務の様々な場面で生み出される。図3に、「電話中に資料を見ながら議論する」というシーンで生み出される情報シナプスの例を示す。

上記シーンで蓄積された情報シナプスは、例えば、次回、同じ人同士が電話をする際に、前回の会話で参照した資料を推薦する、などの活用が可能である。

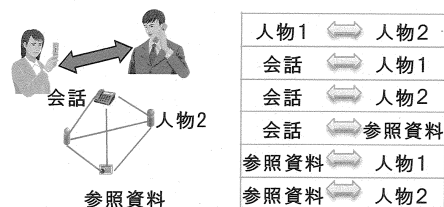


図3 業務中に生成される情報シナプスの例

このように、業務中の様々な場面やツールで生成される関係情報を蓄積するとともに、蓄積された関係情報を様々なツールから活用できるようにするためには、特定のアプリケーションではなく、任意のアプリケーションに対して情報シナプス・データベースへのインタフェースを提供する必要がある。

4.3. アーキテクチャ

図4に情報シナプスを実現するためのアーキテクチャを示す。シナプス・データベースを中心とし、任意のアプリケーションに対してデータベースにアクセスするためのインタフェースを提供するサービスAPI部と、蓄積されたシナプス・データをもとに情報価値の評価や重要情報の発見などを行うシナプス解析エンジンと、蓄積されたシナプスのデータ量を削減する機能などを持つメンテナンスエンジンから構成される。

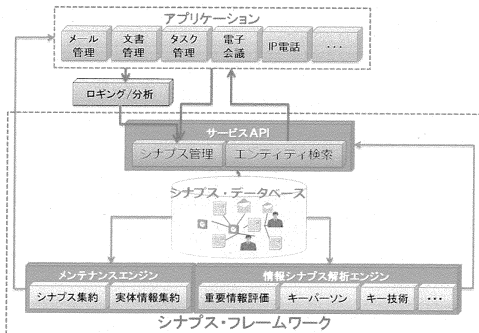


図4 情報シナプス実現アーキテクチャ

・情報シナプスの蓄積

サービスAPIでは、シナプス・データの登録・更新等を行うためのシナプス管理APIを提供する。アプリケーション内に埋め込まれた管理APIが、ユーザの操作等に基づいて情報シナプス登録や更新を行う処理を呼び出すことを想定している。例えば、文書管理システム上でユーザが文書を作成・登録すると、そのユーザと文書との間にシナプス・リンク生成処理を行ったり、ユーザが文書の閲覧(開く操作)をする都度、ユーザと文書の間のシナプス・リンクの重要度を変化させる、などのAPIの使い方を想定している。

このほか、アプリケーションから生成されるログ情報を分析する独立モジュールが、ログから情報間の関係性を抽出した結果をシナプス管理APIを通してデータベースに登録する形態も可能である。

・シナプス情報の参照

サービスAPIは、情報シナプス情報に基づくエンティティ検索APIも提供する。

最も基本的な検索APIは、任意のエンティティの1リンク先のエンティティを検索する関連エンティティ検索である。このとき、エンティティの種別での絞り込みや、シナプス・リンク更新日等による日時情報等による絞り込みも可能である。

エンティティ検索APIも、任意のアプリケーションから呼び出されることを想定している。これにより、文書管理システム上で任意の文書から関連資料を推薦する使い方や、電話帳に登録された人物からその人が作成した文書を呼び出す、などのように、情報の断片から関連する情報を素早く呼び出す使い方が可能である。また、日常業務の中で、特定の業務目的に対して関連する情報を一元的に取り出したい場合もあるが、そのようなケースでは、「業務」に相当する抽象概念をタスク管理ツールの「タスク」などのエンティティとして表現し、タスクとシナプス・リンクが張られた情報を取り出す、などの使い方ができる。

また、シナプス・リンクを順次辿っていくことも可能であるため、「文書 → 文書の作成者 → その人が参加する別プロジェクト → プロジェクトの提案書」などのように、情報の渡り歩きを行いながら、探索的に情報を発見するような使い方も可能である。

他に、指定された複数のエンティティに共通する関連エンティティを検索する共通関連エンティティ検索などを用意しているが、任意の2エンティティ間の経路検索なども将来的に必要となると考えられる。

・情報シナプス解析エンジン

シナプス情報は、関連情報の検索に利用できるだけでなく、蓄積された情報の分析等を行うことでさらに活用価値を高めることができる。情報シナプス解析エンジンには、そのような各種エンジン群が追加される。

ここでは、我々が研究中の解析エンジン例を挙げる。これらは一例であり、今後、拡充していく必要がある。

i) 重要エンティティ発見

エンティティ間のシナプス・リンクによるネットワーク構造を対象に、PageRank[8]などと同様のリンクおよびノード評価を再帰的に繰り返すことにより、エンティティの重要度を評価する機能を提供する。

ii) キーパーソン発見

キーワードマッチによって検索されたドキュメントやメールなどのエンティティ集合に対し、それらの近隣の人物エンティティの重なり度を評価することで、入力キーワードに関係の深い人物を発見する機能を提供する。重要度の高い文書との距離が近い人物を高く評価することなどで、さらに精度を上げることができる。また、人物と他のエンティティとの関係性分析に基づいた人物特性(意思決定者や、ゲートキーパーナ

ど)を加味した評価や、検索者と関係が深い人物に範囲限定した検索などの応用も考えられる。

iii) キー特性発見

キーパーソン発見とは逆に、ある人物の近傍のエンティティ集合の内容分析を行うことで、その人物がどのような技術、業務、顧客などに対して深い関係にあるのかを発見する機能を提供する。

iv) 絞り込みキー推薦

あるエンティティから検索した関連エンティティの件数が多い場合に、さらに1~2リンクの先のエンティティを調べ、重なり度が高い人物エンティティなどを、検索結果を絞り込むためのキーとしてユーザに提示する。これは、システムがユーザに対して連想検索を引き起こすように働きかける効果となる。

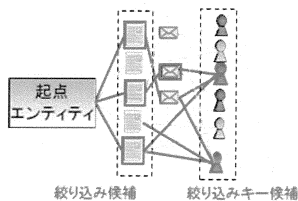


図5 絞り込みキーの推薦

・メンテナンスエンジン

i) シナプス・データ集約

シナプス・リンクやエンティティは、企業内の様々な情報活動から生成されるため、莫大なデータ量になる可能性がある。一般に、企業内情報システムは、一旦稼働した後は、システム強化を頻繁には行うことが出来ないため、あらかじめ十分な余裕をもったシステム設計を行うとともに、稼働中のデータサイズを一定量内に維持し続ける仕組みが必要となる。

シナプス・リンクについては、時間の経過や、参照頻度の低下によって重みを減らしながら最終的に消去していく。また図6のように、2つのエンティティから共通のエンティティに張られているシナプス・リンクを、1本の(重みを増加させた)シナプス・リンクに置き換えるアルゴリズムなどによって、データサイズを減少させることが考えられる。

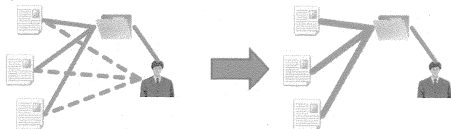


図6 シナプスの集約例

ii) 実体情報集約

シナプス・データベース内のエンティティ情報は、実体情報へのポインタ(URI)情報であるためデータ量が小さいが、それらの実体を管理している各種情報シ

ステム(文書管理システムなど)に日々蓄積されていく情報量は非常に大きく、保存価値が低い情報を判別してそれらを削減することは重要な課題である。単純に、更新やアクセス日時情報から古い情報を削除することは可能であるが、重要な情報まで消してしまう危険がある。シナプス・リンク情報の解析により、他のエンティティとの関係性の低さや、近隣エンティティも含めて長期間活用されていない情報などを発見し、そのエンティティに対応する実体情報を削除の対象とすることで、重要な情報を残したままシステム全体のデータサイズを削減することが可能になると考えられる。

5. 応用研究

上記の考えに基づき、我々は情報シナプス・エンジン、情報シナプスを利用したアプリケーション、解析エンジンの試作などを行っている。それらの一部と、今後取り組むべき研究課題について述べる。

5.1. シナプスビューア

シナプス・データベースに格納される情報を探索・閲覧するための汎用ビューア・アプリケーションを試作している[9]。シナプスビューアは、指定エンティティを起点に、関係するタスク・文書・メール・人物のエンティティ情報をカードメタファで表現し、デスクトップ風の画面上に表示する。各カードからは、実体情報を管理するアプリケーションの呼び出しや、関連エンティティ・カードを呼び出す機能を持ち、シナプス・リンクのネットワークを渡り歩きながら関連情報を探索していくような操作が可能である。

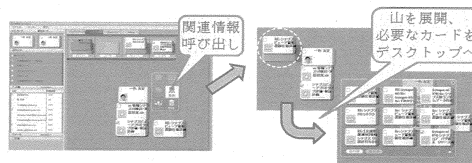


図7 シナプスビューア画面例

5.2. シナプス応用会議システム

情報シナプスを用いて、会議で生み出される情報と日常のオフィス業務で行う作業や生成物を関係づけることで、会議での決定事項や議論内容情報の再利用性を高める会議システムを試作している。ここでは、電子ホワイトボード上の書き込みの一部を選択し、担当者を割り当ててToDo化すると、そのToDoと書き込みデータの間にはシナプス・リンクを生成し、ToDo業務遂行時に簡単に議論経緯を参照できる機能を持つ。また、ホワイトボード上に会議や会議参加者に関連した資料などを呼び出すことも可能である。その際、業務中の作業履歴分析によって、元資料と改編後の資料

の関係や、編集中に参照した参考情報との関係を自動的に抽出し、可視化して提示する機能を持っている。

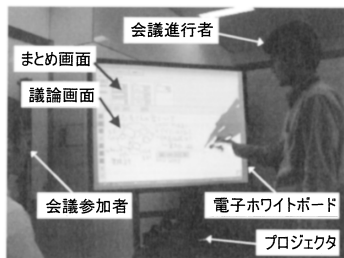


図8 会議風景

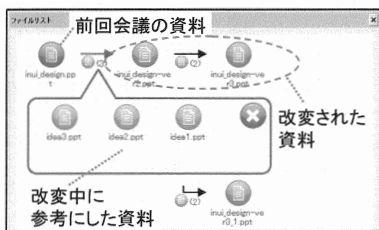


図9 資料の関係情報の可視化例

5.3. 今後取り組むべき研究課題

・シナプス情報の生成技術

情報シナプスのデータは、データベースに蓄積される情報が多いほど、その価値が高まっていく。ユーザに明示的に情報間の関係性を設定してもらうことは負荷が高いため、なるべくユーザに負担をかけずに情報シナプスを多く獲得する技術が必要である。ユーザが通常の業務遂行をしている状態から、自然な形で質の高い情報シナプスを抽出できることが望ましい。非定型業務を対象としたタスク管理ツールによって、業務・人・成果物や中間生成物などを相互に自動的に関係づけることや、IP電話やIM(インスタントメッセージャー)などのコミュニケーション履歴とコミュニケーション中に使用した情報を関係づけるなど、各種の業務ツールにおいて、情報シナプスを生成する仕組みを拡充する必要がある。

また、ファイル操作やWebサイト閲覧などの、より低レベルでのユーザの操作履歴分析に基づくシナプス抽出も試みるべきである。

・シナプス情報の利用

情報シナプスの複雑かつ広大なネットワーク空間を探索するための、情報可視化技術が必要である。特に、エンティティの種類によっては大量のシナプス・リン

クが関係付いている場合があり、それらを分かりやすくユーザに提示する技術が求められる。

また、ユーザの作業状態を把握し、その時点で必要と思われる情報を自動的に推薦するような、適用型情報推薦技術なども、シナプスの特性を活用する重要な研究課題の一つである。

6. おわりに

本稿では、コンテキストに基づく情報検索の重要性を述べ、企業内で生成される情報・知識の再利用性を高める方式として、情報間の関係性に基づく情報構造化手法である情報シナプスと、それが有効に機能するための、企業内の情報活動全体を一つのエコシステムとしてとらえた情報アーキテクチャを提案した。情報シナプスにより、ある情報の断片から関係する情報を素早く取り出すことや、作業や思考の経緯に関わる一連の情報をまとめて扱うことなどが可能となる。また、情報シナプスを共有することで、プロジェクト・会議・委員会などのあらゆるグループワークにおいて、個人が自律的に収集・作成した情報や関係情報を、必要な時にいつでも誰でも取り出すことが可能となる

現在、筆者らは、本提案に基づいて、情報シナプス・データベース、サービスAPI、情報シナプス解析エンジンなどの研究試作を進めている。今後は、業務ツールのバリエーション、メンテナンスエンジン・情報シナプス解析エンジンの拡充を図り、実業務での情報再利用性の向上と有効性の評価検証をしていきたい。

参考文献

- [1]Accenture. Managers say the majority of information obtained for their work is useless. Accenture Newsletter, US, http://newsroom.accenture.com/article_display.cfm?article_id=4484, 2007.
- [2]Feldman S. Enterprise Search Technology : Information Disasters and the High Cost of Not Finding Information. Portals Magazine, No.IDC Special Report , December, 2003.
- [3]Vanneuver Bush, 'As we may think',1945, The Atlantic Monthly.
- [4] Google. <http://www.google.com>.
- [5]Muller,et al., One Hundred Days in an Activity-Centric Collaboration Environment based on Shared Objects., CHI2004.
- [6]梅木ほか. ナレッジワーク支援システム Trino の構想 情報処理学会研究報告, No. 2004-GN-055, pp.1-6, 2005
- [7]芦川ほか, ナレッジワーク支援システム Trino の開発と評価. 情報処理学会研究報告, No.2007-GN-063, pp.13-18,2007
- [8] L. Page, S. Brin, R. Motwani and T. Winograd The PageRank Citation Ranking: Bringing Order to the Web Stanford Digital Library Technologies, January 1998.
- [9]本橋,坂上,内藤,一色 企業内情報を連想的に探索する情報探索アプリケーションの試作と評価. 情報処理学会研究報告, No.2008-GN-070, 2009.