

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

本橋 洋介[†] 坂上 秀和[†] 内藤 圭三[†] 一色 友宏[†]

[†]NEC サービスプラットフォーム研究所

ナレッジワーカーは情報検索に多大な時間を割いており、特に、既知情報の検索に大きな時間をかけている。既知情報検索が困難になるのは、人間のエピソード記憶と情報管理のモデルの違いが探索効率を悪化させる要因となるからであると考えられる。我々は、ナレッジワーカーのエピソード記憶に相当する情報を関係性（＝情報シナプス）として管理し、企業内情報の連想的な探索を可能にする情報探索アプリケーション「シナプスビューア」を提案する。試作を用いた情報探索実験により、シナプスビューアを用いることで関係する関係情報の探索時間が短縮化されることを確認した。本稿では、シナプスビューアの構想・システム概要・評価実験について述べる。

Implementation and Evaluation of Associative Retrieval System for Enterprise Information.

YOSUKE MOTOHASHI[†] HIDEKAZU SAKAGAMI[†] KEIZO NAITO[†] TOMOHIRO ISSHIKI[†]

[†]NEC SERVICE PLATFORMS RESEARCH LABORATORIES

Knowledge-workers spend much time trying to search information, especially to search known-information. That is caused by difference between model of episodic memory and model of information-managemant. We propose an information retrieval system, called "Synapse Viewer", which stores relationships correspond to episodic memory, and enables to search enterprise information associatively. We have experimentally examined the effectiveness that the synapse viewer reduces the retrieval time. This paper describes concept, implementation and evaluation of the Synapse Viewer.

1. はじめに

ナレッジワーカーの業務活動では、業務中の情報検索に多大な時間をかけていると言われている。例えば、米国の調査会社が行った調査^[1]では、ナレッジワーカーは 15～35%の時間を情報検索に費やしており、探しているものが見つかる確率は 50%以下であるという調査結果が得られている。

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。新たな業務を開始する際に自分の知らない知識を得るために社外の web サイトを検索する行為等が未知情報の検索であり、過去の業務の成果物や作業過程で作られた文書・メール等を探す行為が既知情報の検索である。

筆者らは、ナレッジワーカーが、未知情報・既知情報のどちらの検索に時間を費やしているのかを調査するために、簡易なサンプルとして、筆者らのグループの 1 名の 5 日間の業務を観察した^[2]。図 1 は、調査対象者が業務中に情報探索を行った発生回数とその所要時間を調査した結果である。図 1 のように、既知情報の検索は未知情報の検索に比べ発生回数・時間ともに多くなっている。筆者の過去の調査実験^[2]では、未知

情報の検索は新規に業務を立ち上げる際に多く発生し、業務が継続するに従って検索回数が減少しているのに対し、既知情報の検索は業務継続中に万遍無く発生している。このように、ナレッジワーカーの情報検索の負担は、未知情報検索よりも既知情報検索の方が大きいと考えられる。

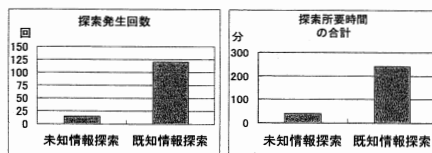


図 1 業務中の情報探索時間

2. 情報システムでの既知情報検索の課題

企業内の情報システムにおいて、既知情報の検索が困難になる状況を考える。既知情報の検索が困難になるのは、過去の状況を漠然と覚えているが正確な保存場所やファイル名を忘れている場合である。このような時に、ナレッジワーカーはフォルダの探索やキーワード検索を利用して既知情報の検索を行うが、時間がかかることや検索に失敗することがしばしばある。

これは、人間の記憶の想起方法と、全文検索エンジンやフォルダ探索による情報の取り出し方法が異なることに起因していると考えられる。

人間の記憶のうち、長期記憶の一部であるエピソード記憶^[3]は過去の行為や体験の記憶であるといわれている。エピソード記憶は人間が体験したエピソードに基づいて符号化された情報同士が連結されて脳内に格納されている。人間がその記憶を取り出すときには、連結された情報を連想的にたどっていきと言われて^[4]。つまり、人間の記憶の想起方法は、図2のようにネットワーク型の記憶を辿る行為と言える。

一方、従来の情報システムでの電子ファイルやメールの探索では、人間はネットワーク型の探索ではなく階層型やジャンプ型の探索を実行している(図2)。

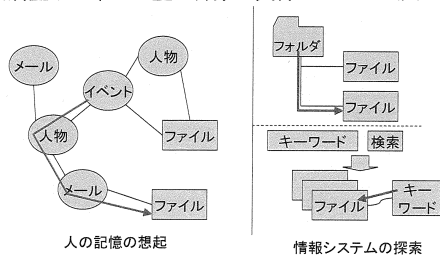


図2 人の記憶の想起と情報システムでの探索の違い

以上の理由から、人間の記憶の想起方法と、情報システムでの情報探索方法は異なっている。この違いによって、断片的な情報のみ覚えている状況において、既知情報の検索が困難になるという問題が発生していると考えられる。

3. 情報探索アプリケーション「シナプスビューア」

本稿では、ナレッジワーカーの既知情報検索に関する課題を解決するために、エピソード記憶に相当する情報を関係性(=情報シナプス^[5])として管理し、関係性を用いた情報探索を可能にするアプリケーション「シナプスビューア」を提案する。シナプスビューアと、それを含むシステムの主な特徴を以下に記す。

- ・メールや文書管理等のアプリケーションのログから自動的に関係性を収集し、再利用可能にすること
- ・関係性に基づく連想的な情報探索を可能にすることで、既知情報の探索効率を向上すること
- ・発見した情報や、作業の途中状態の文書等を自分の作業領域に配置することができ、配置した状態を記録し再現可能にすることで、業務再開時におけるユーザの情報探索コストを低減すること

本章では、シナプスビューアのシステム構成・アプリケーション概要と、関係情報の登録方法・シナプスビューアを利用した情報探索方法について述べる。

3.1 システム構成

図3に提案システムの構成を示す。システムは、メールや文書管理等のアプリケーションを含むグループウェアと、「シナプスエンジン」「ログマネージャ」「シナプスビューア」で構成される。

ログマネージャは、メールや文書管理等のアプリケーションから送られるログを元に、情報間の関係性を判定してシナプスエンジンへ登録する機能を持つ。関係付けの方法は3.3節で述べる。

シナプスエンジンは、情報の関係性の管理と検索を行う基盤である。シナプスエンジン内のDBは、メール・人物・ファイルなどの情報をエンティティ情報として保存した上で、エンティティ情報同士の関係性(=情報シナプス)を関係の重み値と共に保存する。シナプスエンジンは関係性を登録・削除・検索するAPIを備え、ログマネージャやシナプスビューアはこのAPIを用いて情報の保存・取り出しを行う。

シナプスビューアは、シナプスエンジンに保存された関係性に基づいて、情報を表示するwebアプリケーションである。表示された情報には、実体へのリンクが貼られ、ユーザはクリックすることで実体情報を開くことができる。また、シナプスビューア上でユーザは情報同士を関係付けすることが可能であり、関係付けられた情報はシナプスエンジンに登録される。シナプスビューアの詳細な機能は3.2節で述べる。

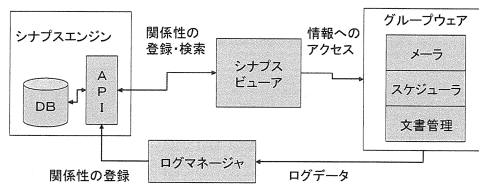


図3 システム構成

3.2 シナプスビューアの機能

図4はシナプスビューアの画面例である。ユーザは、シナプスビューアで、自分の仕事に関係する情報を情報の種類によらず管理できると共に、蓄積された関係性に基づいて過去の経緯を素早く探索して利用することができる。シナプスビューアは以下の機能を持つ。

・カードメタファによる情報表示

シナプスビューアには、ユーザに関係する文書・メール・人物等の情報がカードメタファによってデスクトップ風の画面上に表示される。図4のように、デスクトップ画面にはカードが表示されており、個々のカードがメールや人物などを表す。また、各タスクは個別のデスクトップ画面を持ち、ユーザはタスクを選択することで、タスクに関する情報をデスクトップ画面上に表示することができる。

・カードからのアプリケーション呼び出し

ユーザは、カード内のアイコンをクリックすることで、実体情報を持つアプリケーションを呼び出し、メールやファイル等の実体情報を表示することができる。

・関係情報の探索

シナプスビューアは、指定した情報に関する他の情報を検索して表示する機能を持つ。これにより、ユーザは関係する情報を探索し、見つけた情報からさらに関係する情報を探索することができる。関係情報の探索方法は3.4節で述べる。

・カードの自由な配置と、配置の保存

ユーザは、デスクトップ上に自由にカードを配置することができる。ユーザが配置したカードの位置は保存され、業務を再開した際に以前の配置状態が再現される。これにより、業務の割り込みなどで中断した作業や、日を跨った業務の再開時に、ユーザは以前の思考過程や作業経緯を迅速に思い出すことができる。

・情報の関係付け

ユーザは、デスクトップ上のカードを重ね合わせることで情報同士を関係付けることができる。関係付け方法の詳細は3.3節で述べる。

・実体情報からのシナプスビューア呼び出し

シナプスビューアは、各アプリケーションのメールや文書を指定して立ち上げると、指定した情報が含まれるデスクトップを表示する機能を持つ。

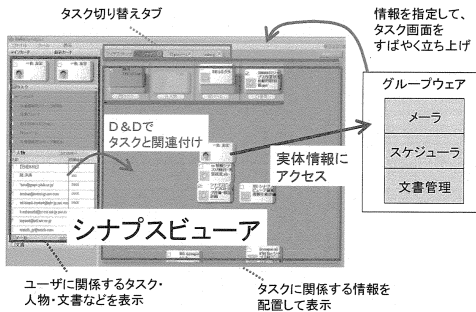


図4 シナプスビューアの画面例

3.3 情報間の関係付け方法

情報の関係付け方法は、ログマネージャによる自動関係付けとシナプスビューアでユーザが行う手動関係付けの2種類がある。

はじめに、ログマネージャによる自動関係付け方法を説明する。表1は、試作したログマネージャが自動的に関係付ける関係性の一覧である。表1のように、ログマネージャは、メールと関係人物や、メールを送りあった人物同士等の明示的な関係性を、アプリケーションのログを元に抽出し登録する。

また、ログマネージャは、ユーザのタスクを自動的に推定し、文書やメールを、推定したタスクと関係付ける。タスクの推定は以下の手順で行う。

まず、メールのサブジェクトや文書のタイトルから重複する単語を抽出し、タスク候補とする。次に、タスク候補とした単語の一般性(品詞等)や、タスク候補とした単語が含まれるメールに關係する人物の偏りからタスクとして適切か判定し、適切と判定した単語をタスクとする。

推定したタスクはシナプスエンジンにエンティティ情報として保存する。その上で、タスクとした単語(=タスクのキーワード)を含むメールやファイルをタスクと関係付ける。また、関係付けたメールのヘッダ情報から人物を抽出し、一定数以上同じ人物が抽出された場合に、タスクとその人物を関係付ける。

表1では、ファイルや文書等、限定した情報の種類に限られているが、対象をこれらに限定することを意図していない。ログマネージャは、関係付けルールを設定ファイルとして持っており、システム管理者が関係付けルールを編集することで関係付ける情報の種別を拡張することができる。

表1 試作したログマネージャで自動的に関係付けられる関係性

関係情報	関係情報が登録されるきっかけ (ログから抽出する)
メール-メール	返信・転送される
メール-ファイル	添付される
メール-人物	メールのfrom・toに含まれる
ファイル-人物	登録、編集、閲覧、ダウンロードする
人物-人物	メールを送りあう
タスク-メール	メール件名にタスクのキーワードを含む
タスク-ファイル	ファイル名にタスクのキーワードを含む
タスク-人物	タスク内のメールのfrom・toに一定数以上含まれる

次に、シナプスビューアによる手動関係付けについて説明する。図5はシナプスビューアでの関係付け方法を示した図である。図5のように、ユーザは情報をあらかずカード同士を重ね合わせることにより情報同士を関係付けることができる。

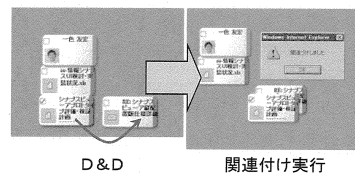


図5 情報の手動関係付け方法

3.4 関係情報の探索方法

図6に、シナプスビューアでの関係情報探索方法を示す。図6のように、ユーザが情報（カード）を1つ選択してメニューから関係カード集積機能を選択すると、選択したカードに関係するカードが集められ、山になって表示される。表示された山の中からさらに別のカードを選択して関係する情報を探索することで、芋づる式に情報を探索することができる。集められたカードはデスクトップ上にストックしておくことができ、後の業務で利用することができる。

4. 関係情報の探索効率に関する検証実験

我々は、シナプスビューアを用いることで関係情報の探索時間が短縮され、また、関係情報を正確に探索できるようになるという仮説を持っている。そこで、上記仮説を検証するための実験を実施した。以下に実験の概要と結果を述べる。

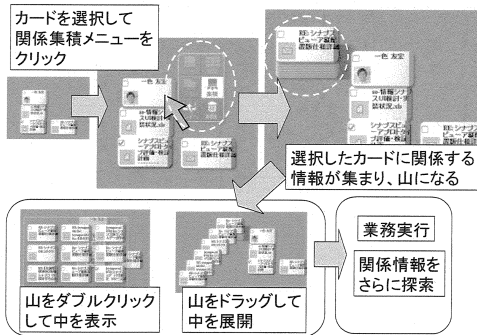


図6 関係情報の探索方法

4.1 実験の概要

被験者に簡単な作業を実行させ、その後、作業の経緯を元にした情報の探索課題を実行させた。表2は被験者に実行させた作業と探索課題である。

まず、被験者に実行させた作業を説明する。被験者に、他の資料を引用して新しい資料を作る作業を実行させた。予め用意された画像データ30個・空のpowerpointファイル30個を与え、課題指示用のwebページを用いて被験者に作業を実行させた。指示した作業は2種類ある。1種類目は、powerpointファイルに画像を貼り付ける作業である。例えば、「岩手県の名産はりんごとウニです。りんごとウニの画像を岩手県.pptに貼り付けて下さい」という指示を提示し、被験者に貼り付け作業を実行させた。2種類目は、powerpointファイル内の全ての画像を別のpowerpointファイルに引用する作業である。例えば、「東北地方の名産の一覧を作ります。東北地方.pptに岩手県.pptと青森県.pptと…の全ての画像を貼り付けて下さい」

という作業指示を提示し、被験者に資料を作成させた。これら2種類の作業を各10回ずつ実行させた。

次に、探索課題について説明する。2種類の探索課題を被験者に実行させた。1種類目は、資料から引用元画像を探索させる課題である。前述の作業の成果物の1つを指定し、「東北地方.pptの中に使われている画像を全て探索して下さい」という指示を提示し、被験者に画像を探索させた。2種類目は、元画像からそれが使われている資料を探索させる課題である。画像ファイルを1つ指定し、「りんご.jpgが使われている資料を全て探して下さい」という指示を提示し、被験者にpowerpointファイルを探索させた。以上の探索課題を被験者に各5回ずつ実行させた。

表2 被験者が実行した作業・探索課題

分類	番号	概要
1. 作業	1-A	指定した画像を新しい資料に貼り付ける
	1-B	指定した資料にある画像をすべて別の資料に貼り付ける
2. 探索課題	2-A	成果物(資料)を1つ指定し、中に使用されているすべての画像の元ファイルを探す
	2-B	画像を1つ指定し、使用された資料をすべて探す

表3に実験条件を示す。被験者は成人男性6名である。被験者に、Windowsのエクスペローラとシナプスビューアで同じ作業・探索課題を実施させた。

エクスペローラには、「画像」フォルダと「powerpoint」フォルダがあり、それぞれの中に全ての画像とpowerpointファイルが予め入っている。エクスペローラを用いた実験では、被験者には、これらのフォルダ内のファイルリストを表示させ、作業や情報探索を実行させた。

シナプスビューア内のデスクトップには、全ての画像・powerpointファイルのカードが予め置いてある。シナプスビューアを用いた実験では、被験者には、貼り付け作業を実施した後、貼り付け元の情報と貼り付け先の情報を関係付けさせた。また、探索課題の際に、シナプスビューアでの関係情報探索機能を用いて情報を探索させた。

探索課題実施中のアプリケーションの操作ログ・探索課題の成否情報・課題を完了するまでの所要時間を収集した。

表3 実験条件

被験者	成人男性6名
使用アプリケーション	1.エクスペローラ 2.シナプスビューア
実験環境	各被験者の日常業務用PC
収集データ	探索中のアプリケーション操作ログ 探索課題の成否情報 探索課題完了までの所要時間

4.2 実験結果

実験結果について述べる。まず、図7はエクスペローラとシナプスビューアの探索成功率の比較を示した図である。図7のように、2つの探索課題とも探索成功率に大きな差がなかった。しかし、シナプスビューアで探索が失敗した原因を被験者の操作ログより分析したところ、貼り付け作業の際に、手動での関係付けを一部失敗しており、正確な検索ができなかったことに起因していることが分かった。関係付けが正しく行われている被験者のシナプスビューアでの探索成功率は両探索課題とも90%以上であり、正しい関係性がある場合の探索成功率は高くなることがわかった。

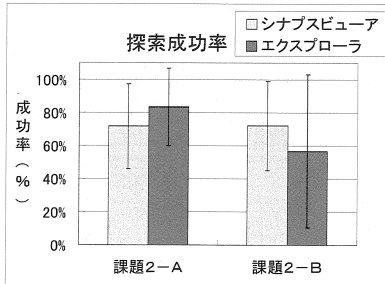


図7 探索成功率の比較

次に、エクスペローラとシナプスビューアで探索時間を比較した。被験者間のPC操作の習熟度が探索時間に与える影響を排除した上で、シナプスビューアとエクスペローラの相対的な結果の差を明らかにするために、以下の手順で探索時間の正規化を行った。

- ① 被験者個人のシナプスビューア・エクスペローラ双方の課題を合わせた、探索時間の平均値を求める
- ② 被験者個人の個々の課題における探索時間を、①の値で割る

図8は、上記②で求めた値の平均値とばらつきを、シナプスビューアとエクスペローラの課題で比較した図である。図8のように、画像からpowerpointファイルを探る課題(課題2-B)において、シナプスビューアの探索時間はエクスペローラに比べて有意に短くなった(有意確率 $p<0.01$)。これは、エクスペローラを用いた場合、課題2-Aではpowerpointファイル中の画像の数や内容を確認できるのに対し、課題2-Bでは画像の引用経緯を記憶から思い出すことを強いられることが原因であると考えられる。

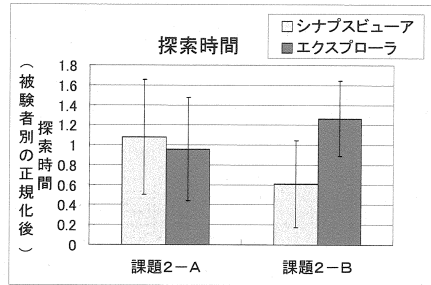


図8 探索時間の比較

4.3 考察

実験の結果では、シナプスビューアはエクスペローラと比較すると探索時間が短くなることが確認された。また、関係付け操作の失敗という別の要因でシナプスビューアでの探索成功率が下がったものの、関係付けを正しく行えた場合の探索成功率は高いことが確認された。これらにより、シナプスビューアを用いることで関係情報の探索成功率が上がり探索時間が減少するという仮説が支持されたと考える。

特に、課題2-Bのように画像ファイルから成果資料を探す課題において探索時間に顕著な差があった。これは、従来の階層型情報管理では再現できない関係情報の探索をシナプスビューアが実現することによって探索効率が向上しているからであると考えられる。このような探索はシナプスビューアの探索方式が最も効果を発揮する状況の一つと考えられる。

本実験を行うことで、関係付け操作が分かりづらく失敗しやすいことがわかった。今回は手動での関係付けを強いたことで失敗が発生したが、ユーザの操作ログ(コピー&ペーストや同時参照等)を元に、正しい関係性を自動的に収集するようにログマネージャを改良することで、探索の成功率が向上すると考えられる。このような、操作履歴に基づく自動関係付け方法については今後の研究課題とすると共に、エラーの少ない手動関係付けUIも今後の課題としたい。

また、被験者にインタビューを行ったところ、エクスペローラの場合は背景知識(青森県はりんごの名産である等)を利用して思考しながら探索を実行したのに対し、シナプスビューアの場合は背景知識を利用せず探索を行っていたことが分かった。これはエクスペローラでの探索時に、エピソード記憶が十分でなく、意味記憶によるエピソード記憶の補完⁴⁾が行われたと考えられ、意味記憶がない場合はエクスペローラでの探索の効率は落ちることが推測される。つまり、シナプスビューアの方が、他人が生成した関係性の探索など背景知識がない情報の探索時の効率悪化が少ないと予想できる。この点は今後の実験で検証していきたい。

本実験はまだシナプスビューアの効果の一面について簡単な評価を行ったのみである。本実験以降に行われるべき他の検証実験について考察する。

まず、関係情報の探索コストに関して、関係情報量の増加・複雑化によって、どのように探索効率が落ちるのかについての検証が必要である。これについては、エンティティ数や関係情報数が多い実験データを用意し、探索経路長が大きくなる探索課題を実行させ、検証を行う予定である。

また、探索経路にスケジュール情報やメール情報を挟むような複雑な探索課題を実行させる実験を行い、探索課題の複雑さが探索効率に及ぼす影響の検証も行ってみたいと考えている。

また、シナプスビューアの想定効果の一つである、業務再開時の立ち上げコストを低減する効果についても検証をしたいと考える。これは強制的に業務課題を切り替える状態を作り、シナプスビューア上で業務課題を頻繁に切り替えながら作業を行わせることで、効果の検証が行えると考えている。

また、最終的に、実業務における有効性検証を行う必要がある。実業務上でのデータは今回行った実験のような単純な関係構造には必ずしもならないと考えられる。例えば、人物に関する情報は長期間の業務後は大量になるはずであり、探索を困難にする要因になると考えられる。

5. 関連研究

本研究は memex^[6]から続くハイパーテキストの研究に大きな影響を受けており、情報間の関係リンクを管理するという狙いはハイパーテキストと一致している。

情報の関係性を管理し探索を支援するアプリケーションの先行研究として、Chau ら^[7]の研究ある。Chau らの研究は、自分が過去に関係した情報についてエピソードを元に絞り込み、情報を探ることができる。しかし、探索方式が絞り込みに限定されており、発見した情報からさらに関係する情報を検索することができない。それに対し、本稿で提案するシステムは、関係する情報を渡り歩きながら情報を探索する点に違いがある。関係する情報を渡り歩くことにより、試行錯誤しながら情報を探するという、より人間が記憶を想起する方法に即した探索が可能になると考えている。

人間のエピソード記憶の想起を支援するアプリケーション及びシステムには、Dumais ら^[8]の過去に見た情報を提示する想起支援システムや、Rekimoto^[9]、Plaisant ら^[10]のタイムライン上に情報を表示するアプリケーションがある。これらの研究で用いられる手法は、想起を支援する情報として主に時間情報を用いており、他の手がかりのみを記憶している時に情報を探

しにくくなることが考えられる。それに対し、本稿で提案するシステムは時間的な関係性だけでなく、論理的关系性や空間的关系性など様々な関係性を手掛かりに情報を探索することができる点に特徴がある。

6. おわりに

本稿では、業務中に発生するエピソード情報を関係性として管理し、後に探索可能にする情報探索アプリケーション「シナプスビューア」と、それを含む情報システムを提案した。また、試作したシナプスビューアを用いて関係情報を探索させる実験を実施し、関係情報の探索時間を低減させる効果があることを確認した。今後は、実業務を対象にした実証実験等を実施し、実用化に向けた課題の洗い出しとさらなる価値の探究を行う予定である。

謝辞 本研究の遂行にあたり、NEC 第 2 システムソフトウェア事業部の三浦真氏よりアプリケーションの機能検討に関して有益な助言を多数頂きました。ここに記して感謝いたします。

参考文献

- 1) S.Feldman; Enterprise Search Technology : Information Disasters and the High Cost of Not Finding Information. Portals Magazine, No.IDC Special Report , December, 2003.
- 2) 本橋,坂上,坂口,飛田; オフィスワーカーの業務種別と情報探索行動の関係に関する調査実験,情報処理学会全国大会,2008
- 3) E.Tulving; Episodic memory and common sense: how far apart?: Philosophical transactions of the Royal Society of London. Series B, V.356, .1413,pp.1505-1515,2001
- 4) 太田他編; 記憶研究の最前線:北大路書房,2000
- 5) 坂上,本橋,坂口,内藤; 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: 情報処理学会研究報告, No.2008-GN-070,2009
- 6) V.Bush; 'As we may think':The Atlantic Monthly,1945
- 7) D.H.Chau 他; What to do when search fails: finding information by association: Proceedings of CHI'2008,pp.999-1008,2008
- 8) S.Dumais 他; Stuff I've seen: a system for personal information retrieval and re-use : Proceedings of SIGIR 26th,pp.72-79,2003
- 9) J.Rekimoto; Time-machine computing: a time-centric approach for the information environment : Proceedings of UIST'99,pp.45-54,1999
- 10) C.Plaisant 他 ; LifeLines:visualizing personal histories:Proceedings of CHI'96,pp.221-ff,1996