

WWWからの情報獲得・整理支援 - 思考・興味空間ブラウザ -

村上 晴美[†] 平田 高志[‡]

[†] 大阪市立大学学術情報総合センター [‡] 奈良先端科学技術大学院大学情報科学研究科

個人の外化記憶を構築し、知的生産活動を支援するシステム Memory-Organizer を試作した。本システムは、(a) ユーザのアイデアから外化記憶の作成・編集を支援する思考空間ブラウザ、(b) ユーザの Web 閲覧時の外化記憶の作成、収集、抽出を支援するオーバレイ Web ブラウザ、(c) ユーザの時系列の興味空間上に外化記憶を配置することにより外化記憶の想起を支援する興味空間ブラウザから構成される。本稿では、WWW からの情報獲得、整理に焦点をあて、本システムの機能を説明する。また、ユーザの Web ブラウジング履歴から個人の興味空間を生成して Web ページを整理する実験を行ったところ、興味空間ブラウザがユーザの興味空間を表していることや、過去の想起や Web ページの整理に役立つ可能性があることがわかった。

Information Acquisition and Organization from WWW

MURAKAMI Harumi[†] HIRATA Takashi[‡]

[†] Osaka City University [‡] Nara Institute of Science and Technology

We have developed a system called Memory-Organizer which enables users to construct externalized memory and support users' creative activities. The system consists of (a) a thinking space browser which helps users to input and edit externalized memory and support creative thinking, (b) an overlay Web browser which helps users to gather, cut, and create information while browsing the World-Wide-Web, and (c) an interest space browser which helps users to recall externalized memory by displaying it chronologically on the users' interest space. In this paper, we focus on information acquisition and organization while browsing the Web and describe the system's functions. We conducted some experiments and found that the interest space browser displays the users' interest space, and it is useful for recalling their past and organizing browsed web pages.

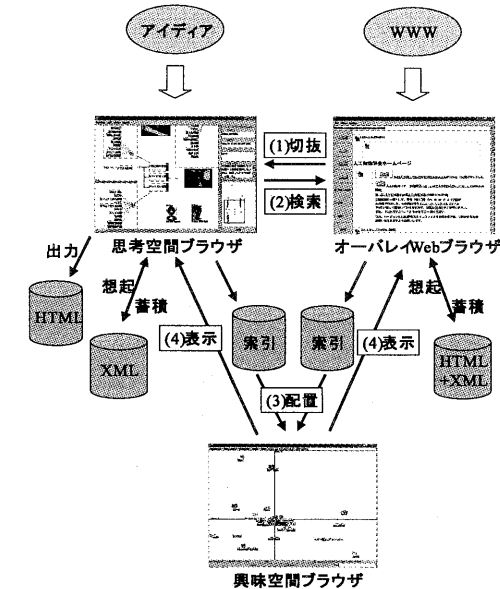
1 はじめに

インターネット、特に WWW における情報収集は我々の日常生活に浸透している。これまで、インターネット上の膨大な情報源から情報を収集、検索、組織化するために非常に多くの研究開発がなされてきた。しかし、それらの多くは、情報収集、抽出、検索、視覚化などを中心とした個々のアルゴリズムの提案や精度の向上などに焦点があてられており、日常生活において個人が収集した大量の Web ページや画像ファイルなどのリソースを、どのようにして知的生産活動に活用するかに関する総合的な環境の提案や実用システムに関する研究はさほど多くない。

我々は、個人の外化記憶を構築し、知的生産活動を

支援するシステム Memory-Organizer¹ を研究開発している。外化記憶とは、外在化された人間の記憶という意味で、比喩的な概念として使用している。我々のアプローチでは、人間の記憶を、とりあえず短期記憶（または作動記憶）と長期記憶からなる二重貯蔵システムとしてとらえ、短期記憶と長期記憶を補助するために、それらの内容をコンピュータ上に外化記憶として蓄積する。一般的なメモツールは、ユーザの短期記憶から長期記憶への意識的な記録処理を助ける役割を果たすが、本システムでは、たまたま見かけたような情報をユーザが意識することなく取り込む役割も果たす。Memory-Organizer では、外化記憶の作成、編集、蓄積、検索などを支援するとともに、外化記憶を用いて行われる、思考、意志決定、執筆などの人間の知的生

¹ 前身は、日常記憶の構築を支援するシステム CoMeMo。



- (1) 切抜:オーバーレイ Web ブラウザ上からテキストや図を思考空間ブラウザ上に切り抜く。
- (2) 検索:思考空間ブラウザ上で作成した検索式を利用してオーバーレイ Web ブラウザ上で Web 情報検索を行う。
- (3) 配置:外化記憶を興味空間ブラウザ上に配置する。
- (4) 表示:外化記憶を思考空間、オーバーレイ Web ブラウザ上に表示する。

図 1: システムの概要

産活動を支援することを目指している。

本稿では、Memory-Organizer の中で、WWW から外化記憶として情報を獲得、整理する過程の支援に焦点をあてて述べる。

2 個人の外化記憶構築システム Memory-Organizer

Memory-Organizer は、個人の外化記憶を構築し、知的生産活動を支援することを目指している。外化記憶とは、外在化された人間の記憶という意味で、比喩的な概念として使用している。

Memory-Organizer は、(a) ユーザのアイデアから外化記憶の作成、編集を支援する思考空間ブラウザ、(b) ユーザの Web 閲覧時の外化記憶の作成、収集、抽出を支援するオーバーレイ Web ブラウザ、(c) ユーザの時系列の興味空間上に外化記憶を配置することにより外化記憶の想起を支援する興味空間ブラウザから構成される。本システムを Java 言語 (JDK1.2) で試作した。外化記憶の対象となる情報源として、現在は主にアイデアと Web ページを対象としている。図 1 にシステムの

コマンド名	マウス操作の軌跡	編集後	説明
範囲指定	複数の概念ユニットを囲む [情報] [編集]	情報編集	一つの概念ユニットに統合
	ユニット群を囲む A B C D	検索式の作成 (A+B)*(C+D) 一つのユニットに統合 要約レベルを付与 Sunite	検索式作成 要約ユニット作成
消去	ユニット上で斜線を引く [] []	ユニットの削除に伴い 関係を失った連想関係 ユニットも自動で消去	ユニット削除 連想関係 ユニット削除
リンク	ユニット間に線を引く [] []		連想関係 ユニット作成
分割	概念ユニット内の文字列間に 位置に線を引く [情報] [編集]	情報 [編集]	一つの概念ユニットを 二つの概念ユニットに 分割

図 2: ペンジェスチャによる情報編集

概要を示す。

以下では、3、4、5 で各ブラウザの機能を説明し、6 でオーバーレイ Web ブラウザと興味空間ブラウザを用いた実験について報告する。

3 思考空間ブラウザ

思考空間ブラウザはユーザのアイデアから外化記憶の作成、編集を支援する。アイデアを書きとめたり、丸で囲んだり線をひいたりなどの、ノートなどの紙で行なう動作をディスプレイ上でできる環境を目指している。

3.1 外化記憶の作成・編集・想起

思考空間ブラウザで扱われる外化記憶の最小単位をユニットと呼ぶ。ユニットには、キーワードなどの概念を表す概念ユニット、外部ファイルへのリンクを示す外部データユニット、ユニット間の関係を表す関係ユニットがある。また、複数のユニットをまとめるユニットとして要約ユニットがある。

現在のところ、外部データユニットとしてテキスト、画像、Web ページを、関係ユニットとして連想関係を扱っている。連想関係は、ユニット間に有方向の関連性があることだけを示す関係で、多対多を扱うことができる。連想関係の表示方法は、Memory-Organizer の前身である CoMeMo における連想表現 ([1],[2],[3]) を使用している。連想関係の起点となるユニットを連想元ユニット、終点となるユニットを連想先ユニットと呼ぶ。図 2 中の範囲指定コマンド欄に連想関係の例を



図3: 思考空間ブラウザ

示す。A、B、C、Dが概念または外部データユニットであり、●が連想関係ユニットである。連想関係の方向は矢印で表現され、A、Bが連想元ユニット、C、Dが連想先ユニットである。

図3に、思考空間ブラウザの画面例を示す。本ブラウザは、ユニットの作成、編集を行なう思考空間表示部、選択されたユニットの作成日時、作成者、URL(Uniform Resource Locator)などを表示する詳細情報表示部、ユニットの全体配置を概観できる全思考空間表示部から構成される。

思考空間表示部上では、ペンジェスチャと呼ぶ、ペンやマウス操作を中心としたユニットの編集が可能である。思考空間ブラウザ上では、思考の流れを止めずに記憶の外在化ができるインタフェースの実現を目指している。図2に代表的なペンジェスチャの例を示す。また、思考空間上へのフリーハンド記述も可能である。フリーハンドで記述する場合には、ペンジェスチャとの混乱が予想されるためモード切り替え(ペンジェスチャ⇄フリーハンド)をして利用する。

本ブラウザは、過去に蓄積した外化記憶を再利用するために、現在の思考空間表示部に表示されている情報、現在のユーザの位置情報、日付指定を基にした想起ができる。想起/非想起はモード設定が可能であり、想起モードにしておくと、システムが思考空間表示部を常に監視し、新たな概念ユニットが作成されるたびに、蓄積されている外化記憶を検索する。検索方式は新規に作成される概念ユニットと蓄積されたユニットにつけられているキーワードとの単純なパターンマッチで行なっている。探索されたユニットは、いきなり思考空間表示部の概念ユニットに関連付けて表示されるのではなく、要約ユニットの形で表示される。ユーザは、表示された要約ユニットが望みのものであれば、それを展開してオリジナルの情報(ユニット)を展開し、

現在の思考表示部上で再利用することができる。

3.2 Web 情報検索

思考空間表示部上に表示されているユニットを使って、サーチエンジンに検索依頼してWeb情報検索を行ない、結果をオーバーレイWebブラウザに表示することができる。検索は、以下に示す2通りの方法が可能である。

1. 概念ユニットを用いた検索

選択された概念ユニットをキーワードとして、Yahoo! Japanなどのサーチエンジン²の簡易検索を実行する。検索式は、すべてのキーワードをAND演算子で結んだものである。

2. 連想関係ユニットを用いた検索

選択された連想関係ユニットによって関連付けられている概念ユニットから検索式を作成し、Yahoo! Japanのエキスパート検索³を実行する。検索式は、連想元ユニットをAND演算子で結び、連想先ユニットをOR演算子で結んで作成する。図2中に検索式作成の例を示す。

検索結果はオーバーレイWebブラウザ上に表示され、ユーザは、検索結果からWebブラウジングを行ない必要な情報のあるWebページを見つけることができる。4.2で後述するとおり、オーバーレイWebブラウザ上に表示されたWebページから必要な部分を思考空間表示部に取りこむことができ、さらに、取りこんだ情報を用いて再び新たな検索を実行できる。このように、検索と抽出をシームレスに行なうことによって、思考や情報収集を中断することなくスムーズに行なえる。特に、Web情報検索においては、現在表示されている文字の中から新たな検索キーワードを作成することがよくあるため、本手法は有効であると考えられる。

3.3 HTML ファイル出力

思考空間ブラウザ上で編集された情報をHTMLファイルとして出力することが可能である。HTMLファイルは、思考空間上に表示されている各ユニットの位置関係や、連想関係ユニットによる関連付けを基に作成される。この機能を使うことにより、思考内容やWebから収集・抽出した様々な情報を編集して新しいHTMLファイルを作ってWebにおくことができる。

² <http://www.yahoo.co.jp/>

³ <http://www.yahoo.co.jp/docs/search.html>

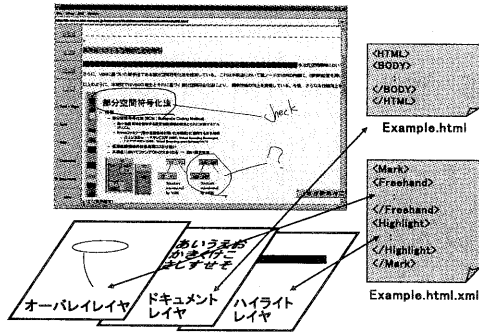


図 4: オーバレイ Web ブラウザ

4 オーバレイ Web ブラウザ

オーバレイ Web ブラウザはユーザの Web 閲覧時の外化記憶の作成、収集、抽出を支援する。Web ページ上への注釈上書きや、Web ページからの情報切抜きができることが特徴である。

4.1 Web ページへの注釈上書き

本ブラウザ上に表示されている Web ページに対して自由に注釈(ハイライト、アンダーライン、コメント)が付けられる機能を提供する。ハイライト以外の注釈は、フリーハンド記述で行う。フリーハンド記述は、3種類の太さ、3色の線を使って書込みが可能であり、線の色により、過去に注釈を付けた Web ページの検索が可能である。これによって、ペンの色を用途毎に決めておけば、ある用途についた注釈付きの Web ページのみを集めることが可能となる。

図 4 にオーバレイ Web ブラウザの画面例を示す。本ブラウザは、3層構造のレイヤから構成されており、対象となる Web ページを表示するドキュメントレイヤを中心として、フリーハンドの記入が可能なオーバーレイレイヤ、ドキュメントにハイライトをつけるハイライトレイヤがある。元々の Web ページであるドキュメントレイヤは 1つの HTML ファイルとして、注釈であるオーバーレイレイヤとハイライトレイヤはまとめて 1つの XML ファイルとして、関連付けられてローカルに保存される。この XML ファイルには注釈が付けられた箇所からキーワードが抜き出され保存される。

本機能の応用例として、電子文書の添削が考えられる。従来は、一旦、紙面に印刷しないと行なえなかった添削を、本ブラウザを被添削者と添削者の両者が利用することにより可能となる。

オーバレイWebブラウザ上のテキスト

クリントン米大統領は20日夕、ホワイトハウスで、チベット仏教の精神的指導者ダライ・ラマ14世と会談し...

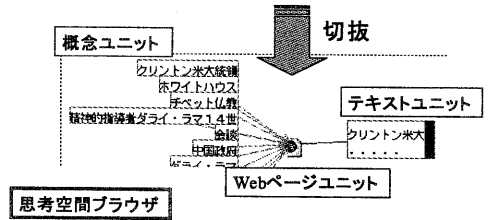


図 5: Web ページからの情報切抜き

4.2 Web ページからの情報切抜き

Web ページからユーザのペンジェスチャにより情報を思考空間ブラウザに取り込む機能を提供する。現在はテキストと画像ファイルを取り込むことができる。

テキストは、必要な箇所をペンまたはマウスでなぞることにより切り出され、概念ユニットまたはテキストユニットとして思考空間上に表示される。句読点を含まないテキストはそのまま概念ユニットとして作成される。句読点を含むテキストは文章とみなされ外部テキストユニットとして切り出され、自動的にキーワード(概念ユニット)が抽出される。

本研究ではリアルタイムな切抜きを想定しており、キーワードの抽出処理時間を可能な限り短くするため、(1) ひらがなと特殊文字以外の文字が 2 つ以上続いた文字列を抽出し、(2) 簡単なヒューリスティックや不要語リスト法による不要語処理を行うという単純な手法を用いている。簡単なヒューリスティックとは、たとえば、「具体的」というように最後に「的」のついた文字列や、「10」「200」のように数字のみの文字列をキーワードとしない、というようなものである。不要語リスト法とは、不要語リストに登録された語をキーワードとしないというもので、ユーザは思考空間ブラウザと興味空間ブラウザ上で選択した語を不要語リストに登録することができる。

テキストの切り出し時に Web ページユニットが作成され、概念ユニットやテキストユニットと関連付けられる。Web ページユニットは、Web ページアイコンと呼ぶ地球マークのアイコンで表示される。思考空間ブラウザ上で Web ページアイコンをダブルクリックすると、オーバレイ Web ブラウザで Web ページを見ることができる。図 5 に思考空間ブラウザへのテキストの切抜きの例を示す。

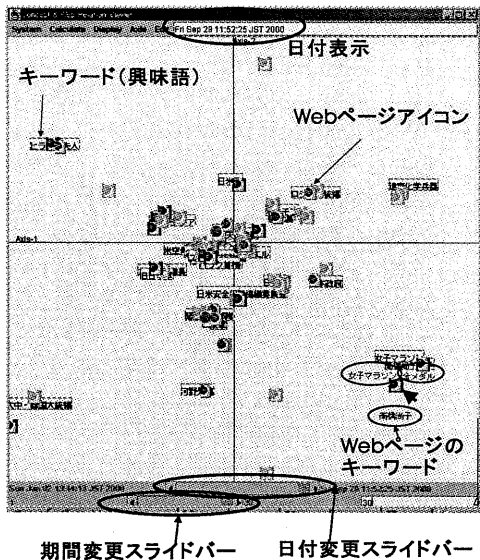


図 6: 興味空間ブラウザ

画像は、直接その画像をクリックすると思考空間ブラウザ上に画像ユニットとして取り込まれる。IMG 要素に ALT 属性が付加されている場合は、ALT 属性に指定されているテキストを連想元ユニット、画像データユニットを連想先ユニットとして連想関係ユニットを作成して表示する。

5 興味空間ブラウザ

興味空間ブラウザは、ユーザの時系列の興味空間上に外化記憶を配置することにより外化記憶の想起を支援する。本研究では、外化記憶から興味語として抽出されたキーワード群の類似度に基づいて生成される 2 次元空間を興味空間と呼ぶ。ユーザは興味空間ブラウザを通して、過去に自分が何に興味を持ったかを知ることや、興味を持った外化記憶にアクセスすることができる。

5.1 Web ブラウジング時の興味語抽出

思考空間ブラウザとオーバーレイ Web ブラウザでは、後で外化記憶の想起を容易にするために、利用時に自動的にキーワードの抽出を行っている。以下は、オーバーレイ Web ブラウザから、Web ブラウジング時に興味語としてキーワードを抽出する手法について述べる。

ユーザがオーバーレイ Web ブラウザを用いて Web ブラウジングする毎に、クリックしたテキストである起

点アンカー (と で囲まれる文字列、以下リンクテキストと呼ぶ) の中からキーワードを自動的に抽出する。現在は、リンク先の Web ページからのキーワード抽出は行っていない。これは、キーワードの抽出がユーザの興味語の抽出を目的としていることと、可能な限り抽出処理時間を短くするためである。本研究では、ユーザのリンク選択は、そのリンク内容や周辺の情報に関心があるために起こす行為であると仮定している。実際にリンク先を見てみるとあまり関係のない Web ページであることがしばしばあり、再現率よりも適合率の向上に重点をおくこととした。選択したテキストの中からのキーワードの抽出手法は、4 で述べたオーバーレイ Web ブラウザからのキーワード抽出手法と同様である。

抽出されたキーワードは、Web ページを閲覧した日付及びリンク先の URL と組で索引として保存され、興味空間の生成に利用される。現在は、数量化 3 類を用いてキーワード群の相関を計算して 2 次元空間上に表示している。数量化 3 類は、数値化されていないサンプル同士の相関を多変量解析する手法(質的データに対する主成分分析)であり、相関の強いキーワード同士が近傍に配置される。閲覧した Web ページを表す Web ページアイコンも相関を計算し同じ空間上に表示することが可能である。

5.2 興味空間の探訪

図 6 に興味空間ブラウザの例を示す。興味空間ブラウザ上では、キーワードは四角で囲まれた概念で、Web ページは Web ページアイコンで表示される。Web ページアイコンをダブルクリックすると、オーバーレイ Web ブラウザが開き、再びその Web ページを閲覧することができる。配置されている Web ページアイコンにマウスカーソルを合わせると、その Web ページから抽出されたキーワードが表示される。たとえば、図 6 の右下のマウスカーソル近傍の Web ページアイコンの周りにキーワードが 3 つ表示されている。これにより、Web ページを実際に表示する前に、この Web ページアイコンのリンク先のページ概要を把握できる。キーワードが 3 つ表示されるだけであるが、自分が過去に閲覧した Web ページなので概ねどのようなページか理解できると考える。

各 Web ページアイコンは閲覧した日付を保存しているため、興味空間の変化を時系列に見ることが可能である。日付変更スライダーで、システムの使用初め日付から現在の日付の中から表示の対象とする日付を

選ぶ。期間変更スライドバーは、表示対象の日付から過去に遡る期間（過去1日から30日の間）を指定する。Web ページアイコンは表示対象日から過去に遡る日数に応じて、カラー（3日以内）、濃い白黒（1週間以内）、薄い白黒（1週間以降）の3段階に表示が変化する。また、Web ページアイコンは閲覧回数に応じて、表示を大（3回以上）、中（3回未満）、小（1回）の3段階で表示する。表示を変化させることにより、よく閲覧するWeb ページの発見や、過去に閲覧したWeb ページの記憶を想起することが容易になる。

興味空間ブラウザは、過去に興味を持って閲覧したWeb ページの、時系列な履歴探訪や、自動的なブックマークとして利用できる。

6 実験

オーバーレイWebブラウザと興味空間ブラウザの有効性について検証するための実験を行った。

6.1 実験1

日常生活における自然な状態でのオーバーレイWebブラウザの動作確認を行い、興味空間生成のためのキーワード抽出手法の有効性について検討する。

6.1.1 方法

本システムの開発者である第2著者（30歳男性）が、オーバーレイWebブラウザを実際にWebブラウジングに用いて利用した結果を考察する。利用期間は2000年5月26日から2001年1月31日までの約36週間（約8か月）である。

6.1.2 結果と考察

オーバーレイWebブラウザの利用に関しては、実装上の問題によりフレームを使ったページの一部の表示がずれるなどの問題があったが、通常のページにおいてはブラウジングを行うことができた。

Webブラウジングの過程からキーワードとして切り出された語は2545語であり、閲覧されたWebページの総数は1052であった。内訳は、朝日新聞のwebサイト（www.asahi.com）757（72.0%）、Yahoo! Japan（www.yahoo.co.jp）74（7.0%）、goo（www.goo.ne.jp）30（2.9%）、その他191（18.2%）であった。1つのWebページに対して平均2.42個のキーワードが抽出された。不要語リ

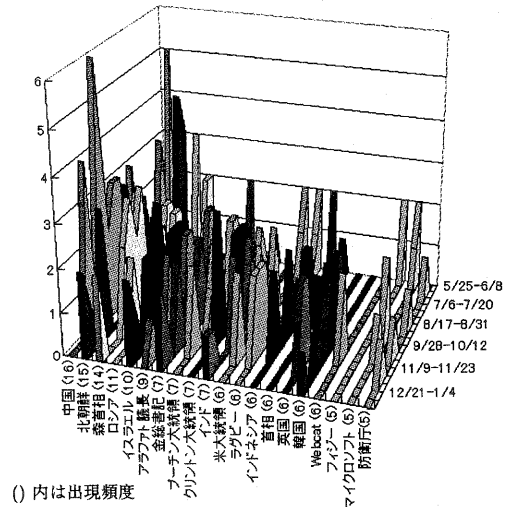


図7: 高頻度語20語の推移

ストに登録された不要語は715語である。

抽出されたキーワードを見てみると、(1) 語の切り出しとして妥当なもの2403(94.4%)、(2) 修正を加えれば妥当なもの119(4.7%)、(3) 妥当でないもの23(0.9%)であった。単純な抽出方法であるが実用性の高い手法であると考えられる。本手法の主要な欠点は、たとえば「神の国」や「つくば市」などのひらがなを含む語が切り出せないことであり、今後の課題である。

図7にシステム利用期間における高出現頻度語20語の推移を示す。最もよく出現した語は「中国」であり、以下「北朝鮮」「森首相」と続く。抽出されている上位語はユーザの興味を反映する語がほとんどであった。また、抽出回数の多い語の推移を見ると、時期によってその注目度(頻度)に差が見られる。期間全体を通して山が続く語は、将来においても興味の対象となる語であろう。反対に、大きな山がきてもその後は一切出現しない語は、その時期の一過性の興味であると考えられる。実験では「Webcat」と「フィジー」が一過性の興味を示す語にあたる。このように、ある期間単位で高い出現頻度の語の推移を見ることにより、ユーザにとって不変な興味と一過性の興味を表す語を選別できると考える。

この実験では偶然ユーザが主として朝日新聞のWebサイトを閲覧しており、どのようなWebサイトにおいても本システムが有効かどうかは不明である。実際、Yahoo! Japanのホームページから掲示板やオークションなどをブラウジングした場合、リンクテキストが存在しないためにほとんど意味のあるキーワードが抽出

されない。今後は、一般的な Web ページなどを対象としてシステムの実際的な有効性を検討する必要がある。

6.2 実験 2

新聞記事を対象とした Web サイトにおける興味語としてのキーワード抽出手法と興味空間ブラウザの有用性について検討する。

6.2.1 方法

被験者は 32 歳、37 歳、37 歳、24 歳の女性 4 名であり、順番に A、B、C、D と呼ぶ。A と C は専業主婦、B はシステムエンジニア、D は秘書であり、それぞれ Web 利用経験は 2、3 年程度である。

毎日インタラクティブの Web サイト (www.mainichi.co.jp) からダウンロードした 2000 年 6 月から 8 月の 3 か月の HTML ファイル (1 ファイル 1 記事) を対象として、同年 12 月にブラウジング実験を行った。何でもいから見たい記事を自分のペースで 1 日に 1 つ以上を見るよう指示した。不要語リストは一切使用しなかった。ブラウジング後に履歴データを回収し、アンケート調査を行った。以下では、(1) 抽出されたキーワードに関して行った興味度の評価と、(2) 興味空間ブラウザの主観的評価について述べる。評価は、5 段階 (5:非常に思う、4: やや思う、3:どちらともいえない、2:あまり思わない、1:全く思わない) で行った。(2) については、抽出されたキーワードのみを表示するテスト 1 と、キーワードと Web ページアイコンを表示するテスト 2 の 2 種類を順番に行った。

6.2.2 結果と考察

抽出されたキーワードの興味度の評価 Web ブラウジングの過程からキーワードとして切り出された語はリンクテキストを対象とする場合 (以下「リンク対象」) で平均 1299 語、全文記事テキストを対象とする場合 (以下「記事対象」) で 11450 語、閲覧された Web ページの総数は平均 372 であった。一つの Web ページに対して、リンク対象で平均 3.62 個、記事対象で 33.86 語が抽出された。

被験者 A の履歴データから抽出されたキーワード 416 個を実験 1 と同様に評価したところ、(1) 妥当 384 (92.3%)、(2) 修正すれば妥当 28 (6.7%)、(3) 妥当でない 4 (1.0%) と、不要語リストを使わなくても高い精度でキーワードが抽出できた。

表 1: 抽出された興味語に対する興味度

	A	B	C	D	平均
総閲覧頁数	109	240	902	236	372
キーワード数 (リンク)	416	867	3039	872	1299
キーワード数/頁 (リンク)	3.82	3.61	3.37	3.69	3.62
キーワード数 (記事)	4298	8788	25324	7388	11450
キーワード数/頁 (記事)	39.43	36.62	28.08	31.31	33.86
興味度 (リンク)	3.85	4.08	3.73	3.50	3.79
興味度適合率 (リンク)	61.5%	76.9%	54.5%	58.3%	62.8%
興味度 (記事)	3.50	3.33	3.36	3.00	3.30
興味度適合率 (記事)	27.3%	50.0%	20.0%	36.4%	33.4%

リンク:リンクテキストを対象とした場合
 記事:記事テキストを対象とした場合
 興味度:高頻度語 10 語における 5 段階評価の平均
 興味度適合率:高頻度語 10 語における評価が 4 または 5 であった興味語の割合

表 2: 興味空間ブラウザの主観的評価

質問	A	B	C	D	平均
テ 興味空間を表しているか	4	4	4	5	4.25
ス 今後も興味空間を見たいか	4	1	4	5	3.50
ト 自分を知るのに役立つか	3	3	5	5	4.00
1 Web ページの整理に役立つか	4	4	4	3	3.75
過去の想起に役立つか	4	5	4	4	4.25
テ 興味空間を表しているか	3	4	4	5	4.00
ス 今後も興味空間を見たいか	2	1	3	4	2.50
ト 自分を知るのに役立つか	2	2	4	5	3.25
2 Web ページの整理に役立つか	3	4	3	3	3.25
過去の想起に役立つか	4	5	4	4	4.25

テスト 1:キーワード表示のみ、テスト 2:キーワード + Web ページアイコン表示

リンク対象と記事対象の抽出語のどちらが、直観的に興味語としてふさわしいかという質問に対して、被験者全員がリンク対象と回答した。上位 10 語について 5 段階評価した結果、リンク対象で平均 3.79、記事対象で平均 3.30 であった。また、回答 4 (やや思う)、5 (非常に思う) と回答した語の割合を適合率として算出したところ、リンク対象で 62.8%、記事対象で 33.4% であった。

予想どおり、新聞記事の場合は、リンクテキストが記事の見出しであるため、記事から興味語を抽出するよりも、リンクテキストから興味語を抽出する方が精度が高いことが確認された。もちろん再現率を高めるためには記事から抽出した語を加える必要があることは自明であるが、今後の課題とする。

興味空間ブラウザの主観的評価 テスト1、2ともに、興味空間ブラウザは、個人の興味空間を表しており(テスト1:平均4.25、テスト2:平均4.00。以下同様)、過去の想起に役立つ(4.25、4.25)ことが示唆された。Webページの整理に役立つ(3.75、3.25)か、自分を知ることに関与する(4.00、3.25)かは、やや肯定的な結果である。

総じて、テスト1よりもテスト2の方が評価が低い、一つにはキーワードとWebページアイコンが重なって見にくくなることと、表示されるキーワードとWebページアイコンの類似度が必ずしも対応していないことが理由である。両者とも表示の問題であり、特に後者に関しては、数量化3類に起因する問題であり、他のアルゴリズムも含めて検討する必要がある。

全体として、興味空間に配置した時系列の提示が過去のWebページの検索、整理に関与する可能性が示唆されたが、表示や操作方法に関するインターフェースの問題が指摘されており、今後は表示アルゴリズムと、インターフェースの改善が課題であると考えている。

7 関連研究

人間の記憶を支援するための研究としては、たとえば、エピソード記憶に似た方法でデータを体制化するForget-me-not[4]がある。Forget-me-notは、人間の記憶を支援するという研究課題とモデルを比較的初期に提示したという点で意義があるが、当時はインターネットの浸透以前でもあり、応用システムに関する研究開発はあまり進展しなかった。

Webページの集合から抽出したキーワードやWebページを2次元空間上に配置する研究としては、自己組織化や統計に基づく手法などの配置アルゴリズムの研究が数多くある。また、現実の課題に応用するためにユーザの興味を2次元空間上に配置する研究もいくつかある(たとえば[5])。本研究は手法に焦点をあてておらず、応用研究であると考えているが、Webブラウジング履歴から記憶の支援やWebページの整理に応用しようという研究は我々の知る限りほとんどない。

Webブラウジングにおけるナビゲーション支援研究としては、WebWatcher[6]やLetizia[7]などがよく知られている。これらは、ユーザの興味を取得するための機械学習などの手法に焦点があてられている。本研究ではブラウジング自体を支援するというよりも、ブラウジングによって得られた履歴データを、Webページを中心とした外化記憶の想起に利用することを目的としている。

8 おわりに

個人の外化記憶を構築し、知的生産活動を支援するシステムMemory-Organizerの中で、(a)ユーザのアイデアから外化記憶の作成・編集を支援する思考空間ブラウザ、(b)ユーザのWeb閲覧時の外化記憶の作成、収集、抽出を支援するオーパレイWebブラウザ、(c)ユーザの時系列の興味空間上に外化記憶を配置することにより外化記憶の想起を支援する興味空間ブラウザを試作した。本稿では、WWWからの情報獲得、整理に焦点をあて、本システムの機能を説明した。本システムでは、(1)Webページから必要な箇所のみ切抜きキーワードをつけて整理することや、(2)Webページに注釈をつけて保存し検索、再利用することや、(3)Webページから獲得した情報とアイデアなどを統合して利用することや、(4)Webブラウジング履歴から個人の興味空間を生成して興味空間上で時系列にWebページを検索することなどができる。

ユーザのWebブラウジング履歴から個人の興味空間を生成してWebページを整理する実験を行ったところ、興味空間ブラウザがユーザの興味空間を表していることや、過去の想起やWebページの整理に関与する可能性があることがわかった。今後は、表示方法を中心としたユーザインタフェースの改善などを行う予定である。

参考文献

- [1] 西田 豊明, 前田 晴美, 平田 高志, 日常記憶の共有支援 - CoMeMo, システム/制御/情報, 1997.
- [2] 前田 晴美, 日常記憶の共有支援に関する研究, 奈良先端科学技術大学院情報科学研究科博士論文, 1998.
- [3] 平田 高志, 村上 晴美, 西田 豊明, 植村 俊亮, 連想表現を用いた創造的思考環境, 人工知能学会全国大会, pp.359-360, 2000.
- [4] Mik Lamming and Mike Flynn, "Forget-me-not" intimate Computing in Support of Human Memory, by Mik Lamming and Mike Flynn, 1993. 村越真(訳)「Forget-me-not」: 人間の記憶を支援する密着型コンピューティング, 認知科学, Vol.2, No.1, pp.16-26, 1995.
- [5] 門林 理恵子, 西本 一志, 角 康之, 間瀬 健二, 学芸員と見学者を仲介して博物館展示の意味構造を個人化する手法の提案, 情報処理学会論文誌, Vol.40, No.3, pp.980-989, 1999.
- [6] Armstrong, R., Freitag, D., Joachims, T. and Mitchell, T., WebWatcher: A Learning Apprentice for the World Wide Web, Proceedings of AAAI Symposium on Information Gathering from Distributed, Heterogeneous Environments, pp.6-12, 1995.
- [7] Lieberman, H., Letizia: An Agent That Assists Web Browsing, Proceedings of IJCAI95, pp.924-929, 1995.