

# 多様な情報源の統合と知識空間の作成による記憶想起支援

## Supporting Human Recollection by Integrating Diverse Information and Creating Knowledge Space

仙波 圭大<sup>\*1</sup>  
Keita Senba

三橋 謙太<sup>\*2</sup>  
Kenta Mitsuhashi

村上 晴美<sup>\*1</sup>  
Harumi Murakami

<sup>\*1</sup> 大阪市立大学大学院創造都市研究科 <sup>\*2</sup> 株式会社オーグス総研  
Graduate School for Creative Cities, Osaka City University OGIS-RI Co.,Ltd.

We presented a data integration method using a simple information structure called history structure, which is constructed from time, keywords, and URI sets. Our Proposed algorithms generate history structure from such information usages as web browsings, twitters, emails, calendars, and book purchases to create a user knowledge space. We developed a system based on our approach to support human recollection. The system's performance was evaluated in three experiments, which revealed the usefulness of our approach and the implemented system.

### 1. はじめに

人の記憶は曖昧なものである。思い出したいことがあっても思い出せないことが日常茶飯事である。また、社会の情報化が進むにつれて個人が扱う情報量が多くなっている。記憶媒体が日進月歩で発達し、写真や音声、映像など多くの情報を保存することができるようになった。人の記憶の限界に比べて情報量は増大する一方で管理することが難しくなっている。

このような背景に関して、住所録やスケジュールなどといった個人情報管理する PIM [Jones 07] や、個人の生活を記録するライフログの研究が盛んに行われている [Gemmell 02]。また、古くから人間の記憶の想起支援を目的とした研究がある [Lamming 94]。本研究室でも長年にわたり個人の記憶を支援する研究を行い、個人の外化記憶構築支援システム Memory-Organizer を開発してきており [村上 01, Murakami04]、本研究はその一環である。多様な情報源から情報を収集して、「日時、キーワード群、URI」から構成される履歴構造と呼ぶシンプルな情報表現を用いて蓄積し、ユーザの知識空間を作成することにより、記憶の想起支援を目指す。先行研究 [村上 04] においては情報源として Web 閲覧履歴のみを扱ってきたが、本研究では、多様な情報源として、Twitter、メール、カレンダー、書籍購入などの情報も扱う [三橋 10]。さらに、知識空間作成アルゴリズムと知識空間ブラウザの再設計、実装を行い、評価実験を行った。

図 1 に本研究の概要を示す。個人の記憶の想起を支援するために、(1) 多様な情報源 (Web 閲覧, Twitter, メール, カレンダー, 書籍購入の履歴) から情報を収集して履歴構造を作成し、(2) 履歴構造からユーザの知識空間を作成・表示する。

以下、2, 3 節では提案するアプローチを示す。4 節では提案手法とシステムに対する評価実験を述べる。

### 2. 履歴構造の作成

#### 2.1 多様な情報源からのデータの収集

多様な情報源から、ユーザの記憶の想起を支援するために有用な、日時を伴う、思考や行動を表すテキストの収集を目指

す。テキストデータは一部の例外 (Web 閲覧の検索クエリやメールの To 行) を除いて、次節のキーワード抽出アルゴリズムによりキーワードに分割してキーワード群とする。

予備調査の結果、Web 閲覧、メール、カレンダーの履歴は Google サービスを利用し、書籍購入履歴はプレーンテキストをユーザに入力してもらうことにした。

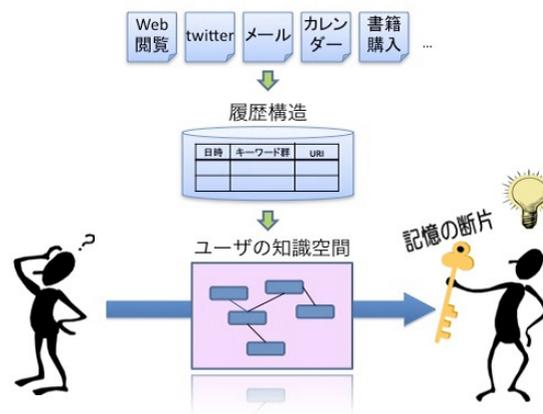


図 1: 研究の概要

#### (1) Web 閲覧

ユーザの興味を表す情報として検索エンジン利用履歴 (Google の検索クエリと、閲覧した検索結果) を利用した。検索クエリの場合は「検索した日時、検索クエリ、検索結果の URI」、閲覧ページの場合は「閲覧した日時、閲覧したページのタイトル、閲覧したページの URI」を取得する。検索クエリの場合はユーザが明示的に語を分割すると考えられるために空白等でキーワードに分割する。閲覧ページの場合は次節のアルゴリズムを用いてタイトルからキーワードを抽出する。

#### (2) Twitter

@返信は儀礼的なやり取りが多いと考えて取り除き、@ではじまらないツイートを対象とし、「ツイートした日時、@ではじまらないツイートの文字列、ツイートの URI」を取得する。@ではじまらないツイートの文字列から次節のアルゴリズムを用いてキーワードを抽出する。

### (3) メール

広告やメーリングリスト宛のメールなどの本人の思考や行動とは直接関係のないものを除くために、送信メールだけを利用することにした。「メール送信日時, To 行と Subject 行の文字列, メール URI」を取得する。送信メールの To 行の中, 氏名が含まれている場合は氏名, 含まれていない場合はアドレスをキーワードとする。さらに, Subject 行をテキストデータとして次節のアルゴリズムでキーワードを抽出する。

### (4) カレンダー

「イベントの日時, イベントのタイトル, カレンダーの URI」を取得する。日時に期間がある場合には開始時間とする。イベントのタイトルは文字列そのものがユーザにとって意味があると考えてそのままキーワードとし, さらに, 次節のアルゴリズムにかけてキーワードを抽出する。

### (5) 書籍購入

書籍や読書管理ツールなどはさほど普及していないため, ユーザが購入あるいは読んだ書籍について日付やタイトルや ISBN をプレーンテキストで入力して, 履歴構造に変換する方式とした。時刻までは管理していないことが多いため省略可能とした。「書籍の購入(あるいは利用)日時, 書籍のタイトル, ISBN」を取得し, 書籍のタイトルから次節のアルゴリズムでキーワードを抽出する。

## 2.2 キーワード抽出アルゴリズム

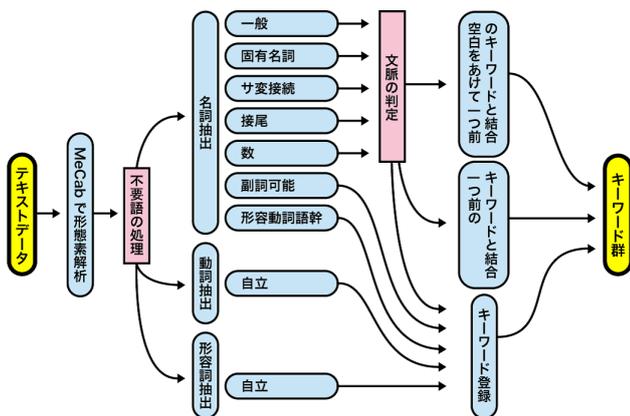


図 2: キーワード抽出アルゴリズムの概要

取得されたデータ中のテキストデータ(閲覧したページのタイトル, ツイート, メール Subject 行, イベントのタイトル, 書籍のタイトル)から, キーワード抽出アルゴリズムを用いてキーワードを抽出してキーワード群を作成する。図 2 にアルゴリズムの概要を示す。

まず, 収集したテキストデータから, MeCab により形態素解析を行い, 不要語(8 語)を除去後, 名詞, 動詞, 形容詞を抽出する。

名詞の場合は, 情報源と品詞細分類 1 に応じて, そのままキーワードとするか, 一つ前のキーワードと結合してキーワードとするか, 一つ前のキーワードと半角空白で結合してキーワードとする(英語等と判定)。

以下に名詞の場合の処理の一部を示す。

品詞細分類 1 が「一般」の場合で, もし一つ前の単語の品詞細分類 1 が一般, 固有名詞または数の場合で, 単語と一つ前

の単語の読みがない場合(英語や記号を意味する), 一つ前のキーワードとの間に半角空白をはさんで結合する(以下省略)。

品詞細分類 1 が「サ変接続」の場合は情報源によって異なる処理を行う。

動詞と形容詞の場合, 品詞細分類 1 が「自立」の場合は基本形をキーワードとする。ただし, 動詞の場合は基本形が一致した場合のみとする。

## 3. 知識空間の作成

### 3.1 知識空間作成アルゴリズム

知識空間作成の基本的な考え方は, 履歴構造において共起するキーワードをつなぐことである。リンクにより自然に発生するクラスタ(まとまり)がユーザの記憶の想起を支援する。

基本的なアルゴリズムは以下のとおりである。

- 履歴構造において共起するキーワードをリンクする。
- 出現頻度が高いキーワードは重要であると考えて大きく表示する。
- 複数の情報源を利用しているため, 情報源がわかるように色分けを行っている。調査の結果, Web 閲覧を薄青, twitter を薄緑, メールを薄赤, カレンダーを薄黄, 書籍購入を薄緑としている。複数の情報源で出現しているキーワードは重要であると考えて赤色で強調表示する。

履歴構造に存在するすべてのキーワードを表示するアルゴリズム 1(視覚化アルゴリズム 1)と, 2 回以上出現及びリンクのあるキーワードのみ表示するアルゴリズム 2(視覚化アルゴリズム 2)を考案した。図 3 に視覚化アルゴリズムの概要を示す。

日時	キーワード群	URI	備考
Mon Sep 13 00:00:00 JST 2010	大学行く, 大学, 行く	http://www.google.com/calendar/feeds/default...	カレンダー
Mon Sep 13 00:28:00 JST 2010	新潟, 佐渡	ISBN-10:4398153187	書籍購入
Mon Sep 13 15:19:14 JST 2010	佐渡島, レンタカー	http://www.google.com/search?q=%E4%BD%90...	Web閲覧 (検索クエリ)
Mon Sep 13 16:06:18 JST 2010	アイランドレンタカー, 佐渡島, 一番, 安い, レンタカー屋	http://www2.tabiplaza.net/district?ps_cd=1&pref=15...	Web閲覧 (閲覧ページ)
Mon Sep 13 17:14:33 JST 2010	明後日, 学会, 佐渡島	http://twitter.com/mitsuhashikenta/status...	Twitter
Mon Sep 13 19:07:55 JST 2010	アイランドレンタカー, 佐渡島, 一番, 安い, レンタカー屋	http://park22.wakwak.com/~islandrentacar	Web閲覧 (閲覧ページ)
Mon Sep 13 19:09:35 JST 2010	アイランドレンタカー, アイランドレンタカー予約	https://mail.google.com/mail/	メール

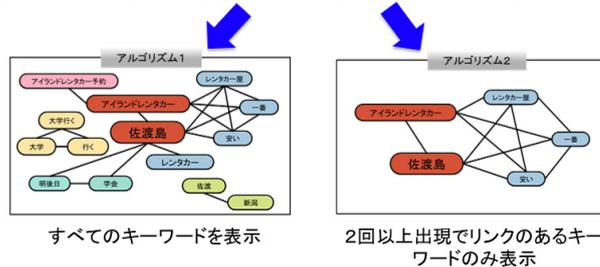


図 3: 視覚化アルゴリズムの概要

### 3.2 知識空間ブラウザ

ユーザの知識空間の作成, 表示, 操作を可能とする知識空間ブラウザを実装した。

知識空間ブラウザは, (a) 選択した期間の知識空間を表示する知識空間表示部, (b) 選択した期間の履歴構造を一覧形式で表示する履歴構造表示部, (c) 知識空間の操作を行う操作・説明表示部, (d) 知識空間上で選択したキーワードを含む履歴構造を表示する選択キーワード表示部から構成される。

図 4 はユーザが佐渡島の学会に出かける前日である 2010 年 9 月 13 日の画面例を示している。知識空間表示部は、視覚化アルゴリズム 1 を用いて表示後にズームした画面である。複数の情報源で出現するキーワードとして「佐渡島」「学会」「アイランドレンタカー」などが赤色で表示されている。履歴構造表示部には時系列の履歴構造、選択キーワード群表示部には「佐渡島」を含む履歴構造が表示されている。

佐渡島内の交通手段として安いレンタカーを探して予約したことや、発表準備のためにファイル読み込みの Java プログラムを書いていたことや、Photoshop の体験版をインストールしようとしていたことなどを思い出せる。

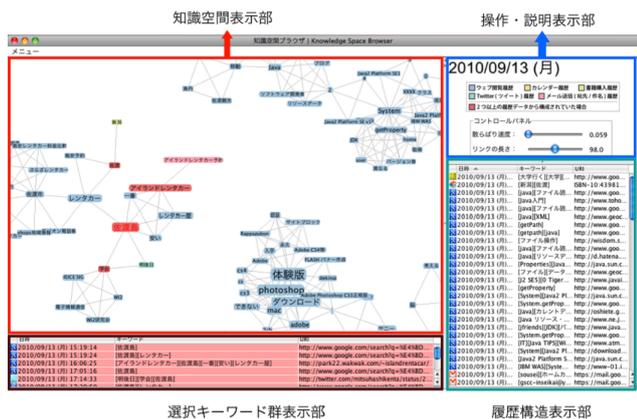


図 4:知識空間ブラウザ

## 4. 実験

### 4.1 実験 1

5 つの情報源からキーワード群を作成するアルゴリズムの有効性を評価する。

#### (1) 方法

被験者は 23-24 歳の情報系分野の男子学生 10 名である。5 つの情報源から 4 つのアルゴリズム (提案手法と 3 つの比較手法) を用いてキーワード群を作成した。

比較手法 1 は名詞を抽出するだけ、比較手法 2 は名詞を抽出して TermExtract にかけるもの、比較手法 3 は提案手法の中名詞だけ扱うものである。表 1 を参照されたい。たとえば、「佐渡島の学会発表では、友達も沢山で楽しかった。」というテキストがあった場合、提案手法では「佐渡島, 学会発表, 友達, 楽しい」と出力され、比較手法 1 では「佐渡, 島, 学会, 発表, 友達」、比較手法 2 では「学会発表, 友達, 佐渡島」、比較手法 3 では「佐渡島, 学会発表, 友達」となる。

作成された履歴構造のキーワード群を確認させ「記憶を表すキーワードとして分割が適切か」を 5 段階評価させた。被験者毎に 10 個×4 手法ずつ行った。

表 1: 比較手法と提案手法

	比較 1	比較 2	比較 3	提案手法
取得する品詞	名詞	名詞	名詞	名詞, 動詞, 形容詞
手法	なし	TermExtract	提案手法	提案手法

## (2) 結果と考察

情報源毎の平均値を表 2 に示す。

表 2: 実験 1 の結果

	比較 1	比較 2	比較 3	提案手法
Web 閲覧	2.71	2.81	4.02	4.05
Twitter	2.56	2.43	3.56	3.91
メール	2.85	3.26	4.10	4.15
カレンダー	3.50	3.54	3.89	4.00
書籍購入	2.78	2.69	3.88	4.18

全体的に提案手法が最も良く、提案手法の有効性を確認した。比較手法 1 と 3 の比較より、名詞は単名詞よりも提案手法を用いて連結させたほうがよいことがわかった。比較手法 2 と 3 の比較から、名詞の連結・提示手法として提案手法が TermExtract より優れていることがわかった。比較手法 3 と提案手法の比較より、名詞だけでなく動詞や形容詞を利用したほうがよいことがわかった。

## 4.2 実験 2

記憶の想起に有用なインタフェースを調べた。視覚化アルゴリズム 1 と 2 を比較した。

### (1) 方法

被験者は 23-24 歳の情報系分野の男子学生 6 名である。

概ね 1 週間前の一日分の履歴構造の中、データ量の多いものの 3 日分を選択して無作為に 3 つの表示方法 (履歴構造, 視覚化アルゴリズム 1 と 2 で表示した知識空間) に割り当てた。

まずシステムを利用して履歴構造を操作させて一日の記憶を想起・説明させ、インタビュー (この表示方法の良かった点, 悪かった点など) を行った。次に同様に、視覚化アルゴリズム 1, 視覚化アルゴリズム 2 の作成した知識空間の順番で、実験を行った。視覚化アルゴリズムのインタビューにおいては、クラスタ数についての質問を行った。最後に、3 つの表示方法について、「記憶の想起に有用か」「一日の要約に有用か」を順位付けさせた。

### (2) 結果と考察

「記憶の想起に有用か」については、6 人全員が履歴構造が最も有用であると答えた。2 位以下は視覚化アルゴリズム 1 と 2 が同等であった。「一日の要約に有用か」については、視覚化アルゴリズム 2 が 3 名、視覚化アルゴリズム 1 が 2 名、履歴構造が 1 名であった。

記憶の想起に有用なのは履歴構造の一覧形式で、一日の要約に有用なのは知識空間であると言える。また、6 人全員から「履歴構造と知識空間の両方が一度に見られるとすれば便利」との回答があった。

表示方法の良かった点と悪かった点について、複数回答が得られたものを表 3 に整理した。履歴構造は時系列, 知識空間は一覧性が良かった点であると言える。

キーワードが 3 つ以上つながったまとまりをクラスタと呼ぶ。クラスタ数が視覚化アルゴリズム 1 では平均 10.5 個, アルゴリズム 2 では平均 4.2 個であった。クラスタ数が適切かどうかはアルゴリズム 1 では 4 人, アルゴリズム 2 では 3 人が適切と答えた。

表 3: 表示方法の良かった点と悪かった点

	良い点	悪い点
履歴構造	時系列でキーワードが並ぶ [4]	一覧性がない [2]
視覚化アルゴリズム 1	クラスタから一日の大枠がつかめる / 一覧性がある [6]	似たクラスタが多いのでまとめるべき [3]
	大体一つのクラスタが一つのテーマである [3]	
	直感的にわかる [3]	
視覚化アルゴリズム 2	情報が整理されている / 思い出しやすいキーワードが多い [4]	情報量が少なすぎる [2]
	アルゴリズム 1 よりもまとまっている [2]	

表 4: 想起に関する平均時間

	履歴構造	視覚化 1	視覚化 2
履歴構造の平均行数	62.7	42.5	95.7
想起までの平均時間	2分 06 秒	2分 48 秒	1分 35 秒
想起を語る平均時間	1分 59 秒	4分 33 秒	1分 48 秒

表 4 に、履歴構造、視覚化アルゴリズム 1, 2 毎に履歴構造の行数と想起までに要した時間と想起を語る時間をまとめた。

視覚化アルゴリズム 2 はデータ量が多いにも関わらず想起までの時間も想起を語る時間も短い。視覚化アルゴリズム 1 はデータ量が少ないにも関わらず想起までの時間も想起を語る時間も長い。

要約については想起までの時間が短いほどよいと考え、視覚化 2 が視覚化 1 や履歴構造より優れていると言えるだろう。想起については、想起する量が多いほどよいと考え、時間が長いほうがよいと考えられ、視覚化 1 が視覚化 2 や履歴構造より優れていると言えるだろう。

### 4.3 実験 3

知識空間ブラウザが記憶の想起に役立つかどうかを検証する。被験者は実験 2 と同じで実験 2 の終了後に実施した。

#### (1) 方法

被験者に知識空間ブラウザを使用して、ある 1 日の記憶を想起・説明させた。システムは自由に操作させた。想起後にはいくつかの質問を 5 段階で評価させた。

#### (2) 結果と考察

システムを使用してその効果について質問したところ表 5 の結果となった。被験者の多くが知識空間ブラウザに有用性を感じていることがわかった。

また、被験者全員がよく似た使い方をしていることがわかった。知識空間ブラウザを使用した場合、次のような順序で記憶の想起を行っていた。

- ① 知識空間を見て一日の行動の概要を想起する。
- ② 見直しや詳細把握のために履歴構造を閲覧する。履歴構造の中に、知識空間では見つけられなかったキーワードを発見する。

- ③ 再度知識空間を見て一度目には目にとまらなかったキーワードを発見する。
- ④ 一日の記憶の想起を行う。

表 5: 実験 3 の評価結果

	質問	評価
Q1	システムを利用することで、より早く記憶の想起ができたか	4.8
Q2	システムを利用することで、記憶の想起が容易になったか	5.0
Q3	システムを利用することで、自分では思い出せないであろう記憶が思い出せたか	4.7
Q4	システムが役に立つあるいは役に立ちそうだと感じるか	4.2
Q5	システムをまた使いたいか	4.5
Q6	システムを利用すると記憶の想起が容易にできそうか	4.7

## 5. おわりに

Web 閲覧や Twitter など日常的な情報源から履歴構造を用いて情報を収集してユーザの知識空間を作成して個人の記憶の想起を支援するシステムを開発した。実験により提案したキーワード抽出アルゴリズムの有効性やシステムの有用性を確認した。

今後は、キーワード抽出アルゴリズムの改良、視覚化アルゴリズム 1 と 2 の相違を精査して最適なアルゴリズムを検討するとともに、一日単位以上(一週間や一年)などの期間における有用性を検討したい。

## 参考文献

- [Jones 07] William Jones: Personal Information Management, Annual Review of information Science and Technology, Vol. 41, pp.453-504, 2007.
- [Gemmell 02] Jim Gemmel, Gordon Bell, Roger Lueder, Steven Drucker, Curtis Wong: MyLifeBits: Fulfilling the Memory vision, Proceedings of the tenth ACM international conference on Multimedia, pp.235-238, 2002.
- [Lamming 94] Mik Lamming, Mike Flynn: Forget-me-not :Inimate Computing in Support of Human Memory, International Symposium on Next Generation Human Interfaces, Proc. FRIEND21, 2004
- [村上 01] 村上 晴美, 平田 高志: Memory-Organizer: 個人の外化記憶構築システム, 2001 年度人工知能学会全国大会(第 15 回)論文集, 2001.
- [Murakami 04] Harumi Murakami and Takashi Hirata: A System for Generating User's Chronological Interest Space from Web Browsing History, International Engineering Systems, Vol.8, pp.149-166, 2004.
- [村上 04] 村上 晴美, 平田 高志: 記憶を中心とする人生の記録 - ユーザの知識空間の作成による Web ブラウジング履歴の想起支援 -, 情報処理学会研究報告 Vol.2004, No.7, pp.19-24, 2004.
- [三橋 10] 三橋 謙太, 村上 晴美: 一日の要約: 一日分の履歴からの知識空間の作成, 電子情報通信学会第二種研究会資料(第 18 回 Web インテリジェンスとインタラクション研究会), pp.47-48, 2010.