

# 個人の連想辞書を利用した図書検索システム

及川 拓弥<sup>\*1</sup> 角 康之<sup>\*1</sup>

## Book Search System Using Personalized Associative Dictionary

Takumi Oikawa<sup>\*1</sup>, Yasuyuki Sumi<sup>\*1</sup>

**Abstract** – We propose a book search system using the diversity of word association. Word association varies from people to people. We regarded it kind of personalities and tastes of individuals. On the system, there is one word association agent per one user, created from sentences written or read by one user. The user can acquire useful information with changing keyword by using and selecting word association agents. We also explain the effect of multiple agent associative search on people's behavior.

**Keywords** : 情報検索, 単語連想

### 1. はじめに

本研究では、書籍のキーワード検索に、人間の単語連想に基づく付加情報を与えることで、ユーザが検索によって意図していなかった意外性の高い書籍を発見することを支援するシステムを提案する。個々人が体験、見聞きしたものの違いによって、ある単語について思い浮かぶ単語が違うという点に着目する。そこで、ユーザの読み書き体験に則した単語連想を行うエージェントをユーザごとに作成し、ユーザがエージェントを相互的に利用しながら検索を行う仕組みを開発する。ユーザは、目的や気分に合わせて、検索に利用する連想エージェントや連想単語の切り替えを行い、検索キーワードを変更させていながら検索を行う。これにより、一検索単語から、広く、深く情報の検索を行っていくことを可能にする。

#### 1.1 キーワード検索によって捨てられてしまう情報たちの存在

キーワードによる情報検索は、電子化された情報の絞込みにおいて、一般的で便利な方法である。しかし、キーワード検索が行っていることは基本的に単純で、キーワードがタイトルや本文中に含まれているかどうかで取捨選択を行うということのみである。もしキーワードが含まれてはいないが、探したい情報として十分な内容で、代わりに関係の深い語句がタイトルや本文中に現れているものがあっても、条件に合致しないものとして破棄されてしまう。特に書籍や音楽などは、検索の際に参照する文字列がタイトルや作者、少量の説明ぐらいしかないので、より捨てられてしまう危険性が高くなる。

#### 1.2 零れ落ちてしまった情報を取り戻す

このような問題を解決するための試みとして、ある単語に関連の深い単語もあわせて検索の手がかりにして検索を行うというものが考えられる。電子化された類語辞典などから取得した関連語を用いて検索結果の幅を広げたり、最適な検索結果になるよう修正を与える。実際に、「WebCatPlus」、「想-IMAGIZE」などといった検索サービスの中に、キーワードから連想される単語を、大規模な書籍データベースや、wikipediaといった Web 辞書を用いてリストアップし、検索単語候補とするという仕組みがある。しかし、同じ検索キーワードであっても、ユーザはそれぞれに異なる検索意図を持っている。単一の辞書によって生まれた、どのユーザに対しても同じ関連語群による検索では、多様に変化する検索意図に対応することはできないと考える。

#### 1.3 人によって違う単語連想

ある事物から、体験・記憶をもとに別の事物を思い浮かべることを連想 (association) という。連想は個人の体験したもの、見聞きしたものによって、非字義的に、多種多様に変化する。本研究では、連想元となる個人の体験・記憶の一部を Web 閲覧履歴や、投稿した文章から形作り、これらを元に連想される単語による関連語検索を行うことで、ユーザそれぞれの検索意図に沿った検索結果を提供できると考える。

#### 1.4 単語連想の個人差を利用した検索結果の補完

また本研究では、ユーザの読み書き体験から作成された単語連想関係の集合体を、ユーザの単語連想行為を代理で行うものとして、連想エージェントと呼ぶこととする。各連想エージェントは、それぞれ違う読み書き体験から構成されているので、ある単語から連想するものにそれぞれ違いがあると考えられる。これによりユーザが、他のユーザの連想エージェントを使用

<sup>\*1</sup>: 公立はこだて未来大学

<sup>\*1</sup>: Future University Hakodate

することで、自分や自分のエージェントでは思い浮かべることのできなかった語彙を用いた関連語検索を行うことができる。本研究ではこのように、ユーザが情報のキーワード検索を行う際に、個人化された単語連想エージェントを複数使い分けて自分自身の語彙を広げていくことによってユーザの発見できる情報を拡張することを目指す研究である。

## 2. 関連研究

### 2.1 履歴による個人化

本研究のように、ユーザの過去の閲覧履歴などを利用してユーザの個性を人数分表現し、利用するものとして協調フィルタリングがある。特に、推薦の精度だけでなく、情報の新規性に注目した研究として、清水らの研究<sup>[1]</sup>がある。協調フィルタリングでは、システムが自動的に性質が同じ他ユーザの情報を利用して、探している情報のうちで最適な情報推薦を行う。本研究では、利用する他ユーザの情報の使い分けを一部ユーザ自身に預けることによって、性質が異なるユーザからも価値のある情報を取得できる可能性を探る。

### 2.2 書籍の検索

本研究のようにユーザの書籍検索を支援するサービスに「WebCatPlus」<sup>1</sup>、「想-IMAGIZE」<sup>2[2]</sup>などといったものがある。これらが提供する機能の中に、単語連想による検索結果の補完機能があるが、本研究では、これらのサービスで扱っているような単一の大規模なデータベースによる連想ではなく、それぞれに元となる情報に偏りのある複数のデータベースを基にしたエージェントによる連想によって、より有用な検索結果の取得を目指す。

### 2.3 キーワード検索の拡張

検索に用いるクエリ情報を拡張または変換をすることによって検索結果の質を向上させる研究として大石らの研究<sup>[3]</sup>や、松本らの研究<sup>[4]</sup>がある。大石らは、キーワード検索から得られた検索結果の本文から、キーワードの関連語を抽出し、関連語を検索のためのクエリに追加することで、検索精度を向上させることを狙っている。また、松本らは、拡張された複合クエリを一語に置き換えることによってクエリの多義性を無くしている。上記の2研究では、キーワード検索によって絞り込まれた情報集合の形を変えていくことによってユーザの求める情報を発見していくが、本研究では、キーワード一つ一つが形作る情報集合を渡り歩いていくことによって、求める情報を発見していく。

### 2.4 単語の連想関係を用いた情報推薦

本研究と同じく、単語の連想関係から、その人の個性なるものを見出し、それに基づいてその人に最適な情報推薦を行う研究として、福井らの研究がある<sup>[5]</sup>。本研究ではさらに、個人化された単語連想を他人に適用し合うことによって、相互作用的にユーザが価値ある情報を提供し合う仕組みを作成する。

### 2.5 セレンディピティの向上

また、意外性のある情報に価値を見出す能力を指す表現として Serendipity という言葉がある。システムの Serendipity を向上させるための研究がいくつかある。例えば、山崎ら<sup>[6]</sup>は、ユーザごとに興味領域を作り、あるユーザと似た領域をもつユーザ集合間での共有度は高いが、そのユーザは共有していない領域の情報をユーザに提示する仕組みによってシステムの Serendipity を向上させることを提案している。本研究では、ユーザにエージェントの相互的利用を行なわせ、概念の渡り歩きを行わせていくことで、システムではなくユーザ自身の Serendipity を養っていくことを目標とする。

## 3. 個人化された単語連想エージェントを用いた連想検索

本章では、提案するシステムの基本的なアイデアと、期待される効果について述べる。

### 3.1 個人の体験・見聞から作られる連想エージェントによる検索

ある単語を入力したら、それに対してその人なりに連想される単語群を出力する連想エージェントを、一人につき一体作成する。連想エージェントは、宿主の閲覧した文章や書いた文章を基に作成される。ユーザがキーワードによる検索を行う際に、連想エージェントがキーワードから別の単語を連想し、ユーザはキーワードを連想された単語に差し替えて検索を行ったり、キーワードと連想単語の AND 検索を行うことによって、単一キーワードによる検索結果とは異なる結果を得ることができる。

### 3.2 ユーザ間での連想エージェントの相互的な利用

自分が考え付いた関連語句によるキーワード検索や自分の連想エージェントによる連想単語に満足が行かなかった場合は、他のユーザの連想エージェントを利用することで、また新たな単語による検索を行うことができる。推理小説を調べたいのなら、推理小説に詳しい人のエージェントを使って、尊敬する人に本を勧めてもらいたいのなら、尊敬する人のエージェントを使って、というように、ある程度自分にとって価値のあるエージェントに目星をつけていきながら、価値の

1: <http://webcatplus.nii.ac.jp/> (2015/7/15 参照)。

2: <http://imagine.bookmap.info/index.jsp> (2015/7/15 参照)。

ある情報をさがしていく。

上記のようなシステムの使用によって期待される効果として、以下のようなものがある。

### 3.3 検索対象に関する分野の深い理解の促進

元となる情報の内容が偏った辞書データを複数使い分けるといった形を取ることで、ユーザがより高速に、調べたい情報の深い領域まで手を伸ばすことができるという期待される。「人工知能」という単語で検索を行った場合を例に説明する。大規模な単一の辞書によって単語連想を行うと、「AI」、「コンピュータ」、「ロボティクス」などといった、抽象度が検索単語と同程度か少し小さい程度の単語しか連想されない。本研究で提案するシステム上では、人工知能のある分野に造詣の深いユーザがいたとすれば、そのユーザのエージェントを利用すると、「ミンスキー」、「コネクショニズム」、「ディープ・ラーニング」など、検索単語よりもより具体的な単語を連想し、それを利用して検索を行うことができる。このように、調べたい分野の専門的な情報をよりすばやく取得することができると考えられる。

### 3.4 自分では思いつかない連想単語での検索による新たな書籍、概念の獲得

複数の連想エージェントの使い分けによって、自分では思いつくことのできない単語を利用した検索が可能であり、ユーザの発見できる書籍の幅を広げることができるという期待される。他のユーザのエージェントによる連想単語によって、自分が予期する検索結果とはまったく違う検索結果だが、そのエージェントの宿主にしてみれば意味のある結果を得ることで、自分にとって未知の書籍や概念を獲得することにつながると考えられる。

## 4. システムの実装

本章では、複数の単語連想エージェントによる書籍検索システムを作成するまでの詳細を述べる。

### 4.1 個人化された単語連想エージェントの作成

まず、読み書きした文章を元にした、単語連想エージェントを複数作成した。エージェントによって作成の方法が多少異なるが、基本的に、単一の人間による Web 閲覧履歴やブログなどの投稿文章のまとまりから、word2vec<sup>3</sup> と呼ばれるニューラルネットワーク作成モジュールを用いて語句同士の関係をベクトル空間上で表現した。word2vec とは、Mikolov ら<sup>[7][8]</sup> によって提唱された、言語モデルをニューラルネットワーク上で表現するための手法を、ツール化したもののことである。この手法は、一単語を 200 次元前後のベクトル

ルとして表し、単語間の関連性をベクトル間のコサイン距離によって求めることができるようにする手法である。word2vec は、この手法における、skip-gram と呼ばれるモデルを用いている。skip-gram モデルとは、ある語の次もしくは前に出現する単語は関連が深い、つまり、もし他の文章である単語が現れたときその前後に現れる確率が他よりも高いだろうと仮定し、そうなるように各単語ベクトルのパラメータを調整することで、言語モデルを成り立たせようという手法である。具体的に以下の (1) 式を最大化するよう調整を行う。ちなみに  $T$  は、ベクトル空間上で表現されている単語の数、 $c$  は吟味する周辺単語の広さ（例えば  $c = 1$  のときは、着目単語の前後の語しか評価されない）、 $w_t$  は今着目している単語で、 $w_{t+j}$  はそこから  $j$  語だけずれた場所にある単語のことである。

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t) \quad (1)$$

ただし、 $p$  は以下の (2) 式で表される。

$$p(w_{t+j} | w_t) = \frac{\exp(v'_{w_{t+j}} v_{w_t})}{\sum_{w=1}^W \exp(v'_w v_{w_t})} \quad (2)$$

ここでの  $v$  は、前述の単語  $w$  などを添え字にした例の 200 次元前後のベクトルのことである。 $v$  のパラメータを調整していくことによって、(1) 式が最大化されるようになっている。

ある単語に関係の深い単語群を知りたいときは、その単語ベクトルとの距離が近い順に単語ベクトルをリストアップすることで実現できる。言語モデル構築のために必要なものは、学習用テキストのみである。ただし、単語と単語がスペースで区切られていなければ、単語のかたまりを認識することができないため、日本語のテキストに関しては、MeCab<sup>4</sup> を用いて文章の分かち書きを行った。また、word2vec では単語ベクトルの次元数を自由に変更することができる。本研究で作成したエージェントは、200 次元の単語ベクトルを用いている。

入力単語から連想される 8 単語を出力する単語連想エージェントを作成した。今回試験的に作成した単語連想エージェントは多岐にわたるが、本稿では、動作例でも扱われる、4 つの代表的なエージェントを紹介する。それぞれ、学習データの収集方法とともに、「連想」、「検索」という単語での、単語連想の様子を示し

3: <https://code.google.com/p/word2vec/> (2015/7/18 参照)

4: <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> (2015/7/15 参照).

ていく。

#### 1) 閲覧履歴から作成されたエージェント

第一筆者が閲覧した Web ページの本文（約 1,2 週間分）を集めたものを学習データとした。学習用の文章は、GoogleChrome の拡張機能を用いて、ブラウザによって Web ページが開かれるたびに、そのページの url をサーバに送信し、集まった url をたどり html ファイルを取得し、そこから本文を抽出しテキストとして蓄積することで実現した。表 1 にこのエージェントによる単語連想の例を示す。

表 1 閲覧履歴から作成されたエージェント

Table 1 Agent created from web browsing history.

学習データのサイズ		1.5 MB	
単語ベクトル数		4657	
「連想」		「検索」	
連想単語	類似度	連想単語	類似度
結果なし		労働	0.74345427
		イメージ	0.71812844
		読む	0.71446949
		自称	0.71029877
		手段	0.71013134
		運命	0.70804440
		所有	0.70537406
		サウンド	0.69312512

このエージェントは、学習データが小さいためか、単語ベクトル数が 4657 とあるように、ボキャブラリに乏しい。「連想」というキーワードに関しては、対応する単語ベクトルを持ち合わせていなかったために、連想が行えなかった。「検索」というキーワードに関しては、正しい言い換え表現というよりは、「労働」、「手段」、「運命」などの、検索行為に対して感情的に湧き上がるもの、「イメージ」、「サウンド」などの、普段どのようなメディアを検索しているのかがうかがい知れるもの、といったものが連想された。このエージェントの宿主が 1,2 週間の間に閲覧したものとして、プロ野球、陰謀論、未確認生物、先史遺産に関する Web サイトが多かった。例で用いた 2 単語以外による連想を一部紹介すると、「大陸」に対して「ムー」（太古の昔にあったとされる伝説の大陸）、「春」に対して「センバツ」など、自分の興味に関連した単語に関しては一定の反応を示す。また、地名を入力するとその場所を本拠地とする球団をある程度正しく連想することができる。閲覧したサイトに関連した単語連想が行われていることがわかる。

#### 2) ユーザのブックマークから作成されたエージェント

一人の学生から協力をもらって、彼のブックマークに

登録されている Web ページを集めたものを学習データとした。1) と同じように、html ファイルを取得し本文抽出し、テキストを集めた。しかし、プログラミング言語のリファレンスなどといった、まとまった文章があまりないものは取得を避けた。表 2 にこのエージェントによる単語連想の例を示す。

表 2 ユーザのブックマークから作成されたエージェント

Table 2 Agent created from bookmarks.

学習データのサイズ		3.2 MB	
単語ベクトル数		6953	
「連想」		「検索」	
連想単語	類似度	連想単語	類似度
プル	0.73039579	指	0.78000128
ビルド	0.72696328	ビルド	0.75165444
プッシュ	0.72066909	分割	0.68612879
使用	0.70454561	リポジトリ	0.68147379
ダウンロード	0.69467461	ロック	0.68085247
アドレス	0.67949122	描画	0.67601066
保持	0.67550814	マウス	0.67385011
抽出	0.6730662	ビュー	0.65479713

このエージェントの宿主は、技術系の記事を多くブックマーク登録していたため、「連想」というキーワードに対して、一部のプログラミング言語で利用される、「連想配列」を思い浮かべ、辞書的な言い換えではなく、プログラミングに関連したワードを連想した。また、1) と同じく学習用のテキストが少なかったためか、「ビルド」などのように、別のキーワードに対して同じ単語を使い回して連想することが多かった。

#### 3) 単独ユーザによる投稿文書から作成されたエージェント

これまでの二つの連想エージェントは、ユーザの読み体験に基づいたエージェントであったが、これからの二つは、ユーザの書き体験に基づいた連想エージェントである。このエージェントは、「松岡正剛の千冊千夜」<sup>5</sup> と呼ばれる、一人の文筆家による、Web 上に公開されている文章を学習データとした。千冊とあるが、すでに 1500 を超える書評文が、公開されており、これまでの二つのエージェントのものよりはるかに規模の大きい学習データを取得できた。表 3 にこのエージェントによる単語連想の例を示す。

「知の巨人」と呼ばれるほどの人物による文章から作成された学習データを用いたこと、学習データも大きく、単語ベクトル数 269733 の豊富なボキャブラリをもったエージェントであることもあって、どちらの

5: <http://1000ya.isis.ne.jp/> (2015/7/15 参照)。

表 3 単独ユーザによる投稿文書から作成されたエージェント

Table 3 Agent created from posting text by single user.

学習データのサイズ		19.9 MB	
単語ベクトル数		269733	
「連想」		「検索」	
連想単語	類似度	連想単語	類似度
創発	0.75987660	ユーザー	0.68051528
稀少	0.74694943	グーグル	0.67923933
類似	0.73171132	索引	0.61592638
不変	0.71754038	エンジン	0.61592096
必然	0.70871704	経路	0.61445873
良心	0.70787751	入力	0.61240464
文脈	0.69999498	クローラー	0.60875606
比喻	0.69394534	サーチ	0.60670047

単語に対しても、関連する単語の確実な連想を行っている。また、このエージェントは、歴史や文化史に関するボキャブラリが豊富であり、例えば「キリシタン」と単語入力を行うと、戦国時代のキリシタン大名である細川忠興にちなんだ単語を連想したりする。

#### 4) 複数ユーザによる投稿文章から作成されたエージェント

Uncyclopedia<sup>6</sup>といわれる不特定多数により編集された Web デクシオナリの本文を学習データとした。Uncyclopedia は、wikipedia と同じく、不特定多数のユーザが自由に編集していくことで出来上がる辞書である。ただし、内容に多くの偏見、嘘偽り、皮肉を盛り込みつつ、面白おかしく記事を作成することをポリシーとしている、他の辞書にはない特徴を持った辞書である。今回は、この複数ユーザによる文章の集合体を偏りのある意見を持った一人の人間による文章と捉え、連想エージェントを作成する。表 4 にこのエージェントによる単語連想の例を示す。

表 4 複数ユーザによる投稿文章から作成されたエージェント

Table 4 Agent created from posting text by multiple users.

学習データのサイズ		167.9 MB	
単語ベクトル数		84358	
「連想」		「検索」	
連想単語	類似度	連想単語	類似度
髭髯	0.54256182	Google	0.50011342
想起	0.53615325	URL	0.39536860
彷彿	0.49063530	グーグル	0.38974666
平伏さ	0.41337269	Yahoo	0.38499709
ひらひら	0.40660411	入力	0.37724983
強要	0.39038258	ヤフー	0.36163410
助長	0.38639599	翻訳	0.35203111
痛感	0.38113451	エンジン	0.34682309

このエージェントも 3) のエージェントのように確実な連想を行っている。「検索」という単語に関しては、3) のエージェントと似た連想を行っており、一部単語も重複している。「連想」に関しては、3) のエージェントとは、別の言い換え表現を連想しており、書き体験の違いによる、連想単語の多様性が確認できたと考えられる。

この時点で、読み書き体験の違いによる単語連想の個人差が確認され、後の検索インターフェースの開発への足掛けとなった。

#### 4.2 連想単語による書籍のキーワード検索の拡張

エージェントによって連想された単語を用いてキーワード検索に付加情報を与えて書籍の検索を行うよう実装を行った。まず、キーワードと連想単語の AND 検索を試み、書籍が見つからない場合は、連想単語単独によるキーワード検索を行うようになっている。書籍の検索には、Amazon、楽天などが考えられるが、今回は試験的に検索クエリの入力や書誌情報の取得が容易であった Junkdo 書店ネットストア<sup>7</sup>による検索結果を用いた。

#### 4.3 書籍検索インターフェースの実装

エージェントによる単語連想や使用するエージェントの切り替え、連想単語による書籍検索を一括して行うための簡単な検索インターフェースを Web アプリケーションの形で実現した。図 1 は、作成した Web アプリケーションの機能概要である。

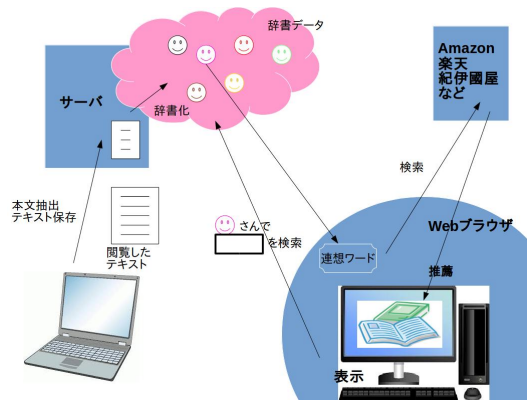


図 1 作成した Web アプリケーションの概要  
Fig. 1 Figure of the Web Application.

サーバに各ユーザの学習データとなる文章を格納し、それから作成した言語モデルを辞書データとして保存しておく。ユーザの Web ブラウザから検索単語と使用するエージェントの情報をサーバに送り、サーバがエージェントに対応する辞書データを参照し検索単語から連想される単語リストを Web ブラウザに送

6: ja.uncyclopedia.info/ (2015/7/15 参照).

7: <http://www.junkudo.co.jp/>

る。同時に、連想単語で、書籍を検索し、検索結果をWebブラウザに送る。Webブラウザでは、送られてきた情報に基づいて情報を表示する。

図2が作成した書籍検索インターフェースの画面である。画面上の入力フォームに検索キーワードを入力し、画面左で、キーワードから連想を行うエージェントを選択する。検索を行うと、画面右に単語連想がリストアップされ、上位2単語を利用した書籍の検索が行われ、結果が画面中央に表示される。途中で、使用する連想エージェント、検索する連想単語の切り替えが可能で、どちらかが行われた場合、書籍の再検索が行われる。



図2 作成した検索インターフェース  
Fig.2 Figure of the Web Application.

## 5. 動作例と課題

### 5.1 動作例

まず、単一のエージェントの使用による動作例を、第一筆者が実際にシステムを利用した際に遭遇した事例を踏まえて紹介する。第一筆者は関連研究を調べるために「セレンディピティ」をキーワードとして検索を行った。そこで、松岡正剛の文章を基にしたエージェントが「サガシティ」という単語を連想した。サガシティとは、察知力と訳され松岡正剛氏によるとセレンディピティとともに二つでひとつの働きを行う重要な概念だという<sup>8</sup>。第一筆者は、この連想が行われるまでに、この言葉を知らなかったため、新たな本人の意図しない重要な概念の獲得が行われたケースとすることができるが、「サガシティ」をキーワードに含む書籍は発見されなかった。このように、たとえ連想単語が本人意図しないものであり、そこから有用な情報を取り出すことができても、連想単語によるキーワード検索でひとつの情報も捜すことができなかったケースが多数あった。

第一筆者は哲学者の「ニーチェ」について、持っている知識がニーチェという名前程度しかなかったため、書籍を読んで勉強するために、「ニーチェ」というキーワードで検索を行った。そうしたところ、エージェントにより「ショーペンハウアー」という単語が連想された。単純に「ニーチェ」でキーワード検索を行うと「日めくりニーチェ」などといった、ニーチェに関する書籍が結果として表れるが、連想された単語による検索を行うと、「ショーペンハウアー 新装版 (Century Books)」、「ショーペンハウアー兵役拒否の哲学」などが、検索結果として現れる。ショーペンハウアーは、ニーチェと同じくドイツの哲学者で、ニーチェの思想に多大な影響を与えた哲学者として知られている。第一筆者は、書籍から、ニーチェの思想の下地となるこの哲学者を先に理解することの重要性をそこから認識した。実際にニーチェの本を先に手にとるか、ショーペンハウアーの本を先に手にとるか、どちらがユーザにとって有効なのかは定かではないが、エージェントの宿主による背景知識から連想された単語による推薦が行われたケースとして示した。

また、複数エージェントの使用によって一検索単語から複数の検索単語候補を創出できるケースを図3に示す。あるエージェントを用いて「単語」というキー

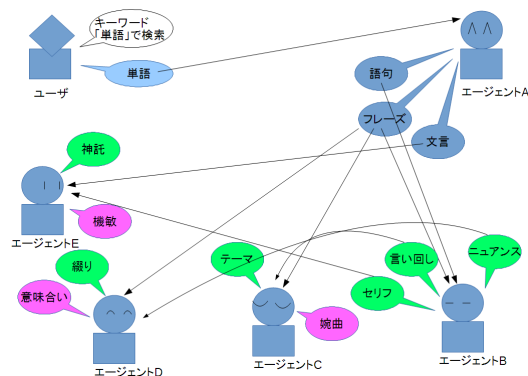


図3 一単語からの連想の広がり  
Fig.3 associative vocabulary enrichment from one word.

ワードで検索を行うと、「語句」、「フレーズ」、「文言」という言い換え表現が連想された。そこからさらに、連想された3つの単語で別のエージェントを利用して検索を行うと、「ニュアンス」、「言い回し」、「セリフ」、「テーマ」、「綴り」といった単語が新たに連想された。そこからさらに、連想された単語で別のエージェントを用いて検索を行うと、「神託」、「意味合い」、「婉曲」、「機微」などが新たに連想された。このように、連想エージェントの切り替えによって、一検索単語から、多くの関連語を獲得でき、そのつど関連する書籍を検



索することができる。

また、2体のエージェントの掛け合いによって、別分野の書籍の新たな発見ができるケースを図4に示す。

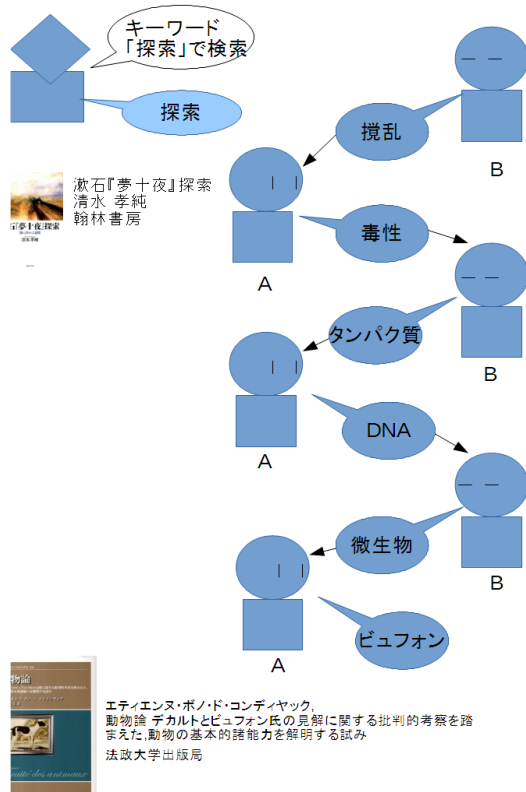


図4 一単語からの連想の移り変わり

Fig.4 associative word transition from one word by multiple agent interaction.

「探索」という単語で、エージェントAが「攪乱」という単語を連想した。それに対して、エージェントBが「毒性」と連想し、それに対してエージェントA「タンパク質」と連想した。それに対しエージェントBが「DNA」、それに対しエージェントA「微生物」と連想した。最終的にエージェントBが「ビュフォン」という16世紀のフランスの博物学者の名前を連想し、彼の著作が検索結果として現れた。このように、エージェント同士による言葉のやり取りによって、一検索単語から、ユーザがまったく予想できない書籍の発見がもたらされることが分かった。

## 5.2 課題

連想単語こそユーザの個性が現れていても、キーワード検索によってユーザの個性が度外視された書籍がリストアップされることが多かった。複数の書籍検索ソースを用意する必要性や、書籍表示順のユーザに即した並べ替えの必要性が示唆された。また、個人の読み体験から作成された連想エージェントは、ボキャブラリに乏しいものばかりだった。もう少し長いスパ

ンの体験情報からエージェントを構成すればこの問題は解消され则认为る。ユーザとエージェントのインタラクションだけでなく、エージェントとエージェントのインタラクションによっても効果的な情報推薦が行える可能性が動作例で示唆された。今後は、エージェント同士の掛け合いをユーザのエージェント利用履歴などに基づいて、ある程度自律的に行うことができるよう実装を行いたい。

## 6. 今後の展望

### 6.1 エージェントの相互的な利用による、社会的コミュニケーション

本研究では、ユーザが別のユーザのエージェントを利用して検索を行うという行為は、お勧めの本を紹介してくださいと尋ねるようなもので、一種のコミュニケーションであると捕らえている。ユーザーは自分のエージェントが他のユーザーが使用するとき、有用で、自分らしく、検索結果を与えられるようにしたいと望む。そのために閲覧・投稿する文章の質を高めたり、普段触れない分野の記事や書籍を読むなどすることが予想される。人と会う前に身だしなみを整えるように、今述べたような自分のエージェントの最適化のための努力が見られ、結果的にシステムによって情報リテラシの向上が図れるのではないかと期待している。また、友人や気になる人の理解のためにその人の連想エージェントで検索をしたり、尊敬する人のエージェントによる検索を行うなど、特定のエージェントによる連想単語に特別な価値を感じて、検索結果によって得られた書籍の読書意欲の向上などが見られるのではないと思われる。

### 6.2 エージェントによる、ユーザへの新たな単語連想の刷り込み

また、本人に身に覚えのない連想であっても、エージェントが連想を行ったというエピソードがきっかけとなって、新たな単語間の関連が、ユーザの頭の中に暗に形成されてしまう可能性がある。連想エージェントの使用履歴や、検索履歴を蓄積し、その情報をもとに、単語の連想関係を更新していく必要性が考えられる。

### 6.3 単語連想エージェントの時間軸による切り分け

ある人の、興味関心は時の流れとともに変化していき、それに伴いある単語について思い浮かべる単語も変化すると考えられる。2014年の自分を表現したエージェントなどのように、時間による切り分けによって、一人の人間から新たに別の特徴をもったエージェントを作成することができると考える。

## 7. おわりに

本研究では、書籍のキーワード検索に、人間の単語連想に基づく付加情報を与えることで、ユーザが検索によって意図していなかった意外性の高い書籍を発見することを支援するシステムを提案した。個人によって異なる単語の連想関係に着目し、ユーザの読み書き体験に則した単語連想を行うエージェントをユーザごとに作成した。これらを利用して、ユーザがエージェントを相互的に利用しながら検索を行う仕組みを開発した。ユーザは、目的や気分に合わせて、検索に利用する連想エージェントや連想単語の切り替えを行い、検索キーワードを変更させていながら検索を行うことにより、一検索単語から、広く、深く情報の検索を行っていくことを可能にすることを目指した。連想エージェントを word2vec により作成し、ユーザの体験の違いによる、連想単語の違いを確認し、検索インターフェースを作成した。本研究の意図するような動作例が確認されたが、実際に推薦された書籍のユーザにおける有効性を示すまでには至らなかった。ただ、複数エージェントの協調による検索単語の拡張の可能性が確認された。今後は、実際に単独ユーザ、複数ユーザによるシステムの利用実験を行い、システムによって推薦された書籍の有効性を確認できるケースを探す。また、一般的な書籍検索サイトや Google などの検索エンジンを使って自分で思いつくままに書籍を検索する場合と、本研究で開発したシステムを用いて検索する場合では、出会う本の内容に違いが出るのか、または、読書体験に違いが出るのかを、検証していく。同時に、検索インターフェースの改良やユーザ間での相互利用環境の構築、複数エージェントの協調による、検索単語の拡張の仕組みの開発を行い、システムによって実現できる書籍探索の可能性を広げていく。

## 参考文献

- [1] 清水, 土方, 西田 : 発見性を考慮した協調フィルタリングアルゴリズム, 電子情報通信学会論文誌 . D, 情報・システム J91-D(3), pp538-550, 2007.
- [2] 丸川, 阿辺川 : 横断的連想検索サービス「想-I M A G I N E」 データベース連携が拓く新たな可能性, 情報管理 53(4), pp198-204, 2010.
- [3] 大石, 峯, 長谷川, 藤田, 越村 : 関連単語抽出アルゴリズムを用いたクエリ拡張, 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM2009), C4-3, 2009.
- [4] 松本, 北山 : 上位下位関係に基づく複合クエリの集約手法, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM2015), C5-3, 2015.
- [5] 福井, 佐藤, 小山, 栗原 : Web 閲覧履歴に表れる単語の連想関係を考慮した情報推薦のためのユーザモデルの生成, 全国大会講演論文集 2013(1), pp665-667, 2013.
- [6] 山崎, 中島 : アイテムに対する Community 内認知度とユーザ嗜好度を考慮した Serendipity 指向情報推薦

方式, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM2015), B2-3, 2015.

- [7] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient estimation of word representations in vector space, arXiv:1301.3781, 2013.
- [8] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean: Distributed Representations of Words and Phrases and their Compositionality, arXiv:1310.4546, 2013.