

個人連想辞書群を用いたセレンディピティ指向 書籍探索システム

及川 拓弥^{1,a)} 角 康之¹

概要: 本研究は、他人の視点を取り入れることによって、ユーザに新たな価値や視点をもたらすことのできる書籍探索手法を提案する。様々な他者の思考空間を代替する個人連想辞書を複数用意する。その連想辞書群を使った仮想的な対話を通して、新たな観点を見出し、書籍探索を豊かにすることを目指す。人それぞれの読み書きデータを用いて、同じキーワードに対してそれぞれに多様な単語を連想する辞書群を作成し、それらを利用して情報を探索するための Web アプリケーションである、AssocSearcher を開発した。この AssocSearcher 上で、単一の大規模な関連語辞書を使用して探索を行う条件と、個人化された連想辞書を 10 体使用して探索を行う条件で比較実験を行った。これによって、ユーザが意外性の高い書籍との出会いを行うことができたかという観点で評価を行ったとともに、探索方法、ユーザの見つけられた書籍の違いについて分析を行った。

Serendipitous Book Explorer Using Personal Associative Dictionaries

TAKUMI OIKAWA^{1,a)} YASUYUKI SUMI¹

1. はじめに

本研究は、他人の視点を取り入れることによって、ユーザに新たな価値や視点をもたらすことのできる書籍探索手法を提案する。ユーザがあらかじめ具体的な目標を持っていない、検索意図が明確でない探索を楽しむことを支援したい。

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。ある単語に対してどんな連想を行うかは、人それぞれ、経験や興味の違いによって異なり、そういった他者と違う連想単語は、その人の独自性や個性のある思考が反映されているものだと考える。そこで本研究では、「あの人がいたらどういった本を薦めるだろう」という、他者による擬似的な推薦を実現するために、様々な他者の思考空間を代替する単語連想辞書 (以下、個人連想辞書と記す) を複数用意する。

それらの個人連想辞書を使用した仮想的な対話を通し

て、事前には想定していなかった新たな視点に気づかされたり、知らなかった観点をを用いて取得した情報に新しい価値を見出すことで、書籍の探索を豊かにすることを目指す。

本研究では、ユーザの体験や興味の一部として Web 閲覧履歴や、投稿した文章などを利用し、それぞれに異なる連想を行う辞書を人の数だけ作成する。それらを利用して、書籍探索を行うためのアプリケーションである AssocSearcher を開発する。

また、単一の大規模な関連語辞書を使用して探索を行う条件と、個人化された連想辞書を 10 体使用して探索を行う条件で比較実験を行う。これによって、ユーザが意外性の高い書籍との出会いを行うことができたかという観点で評価を行うとともに、探索方法、ユーザの見つけられた書籍の違いについて分析を行う。

2. 関連研究

2.1 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

¹ 公立はこだて未来大学
Future University Hakodate

^{a)} t-oikawa@sumilab.org

WebCatPlus^{*1}や想-IMAGINE^{*2}[1] 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辞書による連想によって、より臨機応変に変化する検索結果の取得を目指す。

2.2 キーワード検索の拡張

本研究は、検索に用いるクエリ情報を拡張または変換することによって検索結果の質を向上させる研究の一種である。大石ら [2] 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、その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 [3] 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。キーワードを拡張する関連語を WordNet などと言った字義的な関連語コーパスをもとにしたものにし、検索結果の質を上げる研究 [4] がある。また、拡張するクエリをユーザの特性ごとに個人化し支援するアプローチがあり、Web 閲覧履歴、OS レベルの利用情報などから個人化された関連語を提示する研究 [5], [6] などがある。本研究では、様々な個人連想辞書を複数用意することで、多目的に対応出来るようにするとともに、他人の個人化された関連語を利用することで、ユーザが思いつかない関連語句を取得できる。

2.3 Serendipitous Search

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 (セレンディピティ) という言葉がある。本研究も情報探索において、この能力を高めるための研究 (Serendipitous Search, Serendipitous Exploration) であると考えている。ユーザの興味の範囲内の中で、まだ見つけられていないもの、もしくは、ユーザの興味の範囲外を認識し、その領域の情報からの新たな発見を促す研究がある。山崎ら [7] 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。また、部分的に一致しているユーザ集合を用いる方法 [8] もある。また、概念の渡り歩きによって新たな情報との出会いを促進させる研究などがあり、例えば、エンティティ同士の関係ノードを自動的にランダムで辿っていった結果となる情報を提示するなど、検索結果に調整されたランダム性を利用するもの [9] だったり、エンティティ同士の関連性を視覚的に分かりやすいグラフとし

て表現し、そこからユーザに選択させる試み [10], [11] などがある。本研究では、他者の視点をを用いた、人間による擬似的な推薦によって、Serendipity の向上だけでなく、実際に新しく見つかった情報を手にとって理解するというユーザ自身のアクションまでを促進することを目指す。

2.4 検索意図を考慮した検索結果

ある単語に対してどんな検索結果がふさわしいかは、ユーザの好みや意図の違いによって変化する。本研究では、そういったユーザごとに変化する検索の意図に柔軟な情報探索も可能だと考える。これまでは、その柔軟性を高めるために、検索時点でのユーザの意図はどのようなものなのかを、認識する方法がとられた。例えば、検索クエリそのものやこれまでの履歴などの周辺情報から検索意図を抽出し、それに応じて検索結果のリランクを行うなどのアプローチ [12], [13] などがある。本研究では、意図は、同じユーザ単位でも際限なく変化すると考えており、また意図がはつきりしているならば、ユーザ自身がよくわかっているだろうと仮定する。よって意図の抽出を行う代わりに、ユーザが自分で認識している意図の変化に応じて使用する辞書を変えることを可能にすることによって柔軟性を高めることを目指す。また、あらかじめ検索意図が定まっていなかったユーザが辞書との対話の中で、自分では認識していなかった新たな意図を知ることを支援したい。

3. 提案手法

本章では、個人化された連想辞書の作成方法と、それらを利用した Web アプリケーションである AssocSearcher について説明する。

3.1 個人化された単語連想辞書の作成方法

本人の体験・好みに基づいた、固有の単語連想を行う辞書を作成するために、個人ごとの読み書きデータをひとまとまりとした学習データから言語モデルを個人ごとに構築した。

単一の人間による閲覧履歴や投稿文章のまとまりから、ニューラルネットワーク作成モジュールである word2vec^{*3}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 [14], [15] 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

*1 <http://webcatplus.nii.ac.jp/> (2017/6/24 参照)

*2 <http://imagine.bookmap.info/index.jsp> (2017/6/24 参照)

*3 <https://code.google.com/p/word2vec/> (2017/6/7 参照)

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。ここで 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とする。

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

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。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。本研究で作成した辞書は、窓を 5、次元数を 500 として学習を行った。学習データが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、MeCab*4 を用いて文章の分かち書きを行った。

また、glove[16] と呼ばれる手法も用いた。類似度計算の際に前述の手法では確率ベースで行っていたが、こちらはまず、全体的な共起行列を作成しそこから単語ベクトルに変換する手法である。こちらの手法では、次元数を 100、窓を 10、学習率を 0.05、エポック数を 10 とした。

後述するシステムで各連想辞書が連想する単語に両手法によって生成された連想単語を含め運用し、どちらがより検索のための連想単語として使用されるかを比較したが、差が認められなかった。そこで、本稿では、word2vec と glove での結果を両方使用し、入力単語との類似度のそれぞれ上位 7 単語を重複なしで結合した、10~14 単語に着目しそれらを辞書が単語に対して連想する単語群とした。

3.2 作成した辞書と連想の例

読み書きした文章を元にした、単語連想辞書を複数作成した。ユーザの閲覧履歴、登録したブックマーク、ユーザが執筆したブログ記事、twitter や Facebook で投稿した文章といった読み書きデータを元に、ユーザごとにそれぞれ作成した。また、特定の分野を扱った wiki や Web サイトからも、ある偏った知識を持ち合わせた偏りのある意見を

*4 <http://taku910.github.io/mecab/> (2017/6/7 参照)

持った一人の人間による文章とみなして辞書作成を行った。以下に、実際に作成した単語連想辞書の例を示す。また、各連想辞書の「言語」というキーワードに対しての連想単語を例として示す。

表 1 学生 A の連想の例 (キーワード「言語」)

手法	word2vec	手法	glove
学習データのサイズ	11.4 MB	学習データのサイズ	11.4 MB
単語ベクトル数	12201	単語ベクトル数	58120

連想単語	類似度	連想単語	適合率
処理	0.8229	自然	0.9622
自然	0.7438	型	0.8600
応用	0.7268	副作用	0.8458
音声	0.6903	処理	0.8350
ファクトイド	0.6876	関数	0.8336
シンポジウム	0.6564	プログラミング	0.8330
従来	0.6536	音声	0.8074
ことば	0.6535	イーディッシュ	0.7704

表 2 松岡正剛辞書の連想の例 (キーワード「言語」)

手法	word2vec	手法	glove
学習データのサイズ	19.9MB	学習データのサイズ	19.9MB
単語ベクトル数	26973	単語ベクトル数	77644

連想単語	類似度	連想単語	適合率
数理	0.6981	宗教	0.9896
コミュニケーション	0.6801	生命	0.9772
認知	0.6503	論理	0.9762
優生	0.6439	数学	0.9722
生態	0.6409	環境	0.9712
地理	0.6353	精神	0.9702
生理	0.6322	行動	0.9701
記号	0.6312	システム	0.9687

表 3 教授 A の連想の例 (キーワード「言語」)

手法	word2vec	手法	glove
学習データのサイズ	717 KB	学習データのサイズ	717 KB
単語ベクトル数	8470	単語ベクトル数	14121

連想単語	類似度	連想単語	適合率
出現	0.6541	社会	0.9887
表出	0.6452	的	0.9884
パターン	0.6173	表現	0.9878
着目	0.5873	実現	0.9875
共起	0.5729	意味	0.9874
合算	0.5673	体験	0.9867
プロトコル	0.5535	情報	0.9850
文法	0.5527	術	0.9846

表 1 に示す辞書は、第一筆者が閲覧した Web ページの本文 (約 2,3 ヶ月分) を集めたものを学習データとした「学生 A 辞書」である。「言語」に対して「処理」、「音声」、「プログラミング」など、情報技術よりの連想を行うことがわかる。この辞書の宿主が収集期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、プログラミングに関する Web サイトが多かった。例で用いた単語以外に

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が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。

表2に示す辞書は、「松岡正剛の千冊千夜」*5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辞書である。「言語」に関する連想では、「認知」、「コミュニケーション」など、学生Aよりも、より一般的な、人間同士の意思疎通としての言語に関する連想を行っている。また、この辞書は、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連想単語が得られる。

表3に示す辞書は、第二筆者が投稿したtwitter、執筆した論文を学習データとした辞書である。「言語」に対して、「プロトコル(言語プロトコル分析)」、「共起(言語の共起反応)」、「パターン(パターンランゲージ)」、「表出(表出言語)」など自身の研究的な関心に基づいた、より専門的な連想単語が得られる。

このように、それぞれ違うテキストから学習するとそれぞれの属人的な背景に基づいた個性的な連想を行うという状況を作れる。このことをを利用して、探索システムを開発した。

3.3 複数の辞書を利用した探索アプリケーション

個人化された辞書群を用いて書籍を探索するための AssocSearcher を開発した。個人連想辞書のマネジメント、辞書を利用した書籍・Web検索、あとで読むやブックマークといった書籍情報の管理ができる。マイページ画面(図1)で、登録した書籍の管理や、検索で使用する個人連想辞書リストの確認・管理が行える。サーチ画面(図2)では、左に書籍・ウェブ検索結果が表示され、右には自分が辞書リストに登録している辞書が並んでいる。辞書のアイコンの左下のボタンを押下すると、検索キーワードに対しての連想単語の一覧が表示される。その一覧の中から特定の単語をクリックすると、左の検索結果が、検索キーワードと連想単語の AND 検索を行った際の結果に更新される。各書籍画像の下に位置する詳細ボタンを押下すると、その書籍の詳細ビュー(図3)が表示され、書名、著者名、詳細な説明文、連想単語と連想辞書のアイコンが表示される。右下にある閉じるボタンの左隣に「あとで読む」ボタンがあり、押下すると、マイページで管理する「あとで読む」書籍リストとして登録がされる。

なお、書籍の検索結果や詳細情報は honto ストア*6の検索

*5 <http://1000ya.isis.ne.jp/> (2017/6/26 参照)

*6 <https://honto.jp/netstore.html>(2017/6/24 参照)

結果を用いた。理解補助のための Web 検索結果は bing*7から取得した。ユーザのサイト上での利用データを収集している。対象とするイベントは、検索単語の入力、書誌情報の閲覧、あとで読むへの登録、ウェブサイトの閲覧、辞書の連想、連想単語を用いた検索である。それぞれのイベントに対して、日時、検索キーワード、書誌情報、連想単語が関係している場合は、どの辞書の何番目の連想か、といった情報が収集される。



図1 マイページ画面



図2 サーチ画面

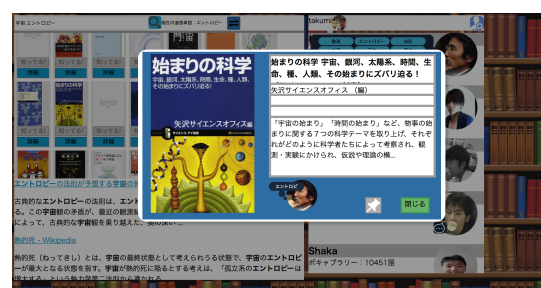


図3 書籍の詳細ビュー

4. 評価実験

個人連想辞書群の使い分けによる様々な他人の視点を用いた探索によって、自分にとって新たな発見となる書籍をより多く得られる(「あとで読む」に登録する書籍が増える)、また、見つけられた書籍の幅が広がる(書籍の作り出す名刺空間の広がりが大きくなる)と仮定した。その検証

*7 <https://www.bing.com/>(2017/6/24 参照)

を行うために、複数個人連想辞書を使用した時と、大規模な単一関連辞書を使用した時で、比較実験を行った。

4.1 実験概要

学生 6 人に実験に参加してもらった (平均年齢 22 歳)。表 4 に実験参加者の基本情報を示す。いずれの参加者も研究活動を行っており、普段本に目を通す頻度 (パラパラめくるだけなど、読み通していないものもカウントする) は月に 1~2 冊が 2 人、3~10 冊が 4 人であった。

表 4 各参加者の詳細

参加者	性別	本に目を通す頻度	キーワード	タスクの順番
A	女	3~10 冊/月	生物	単一 → 複数
B	男	1~2 冊/月	視線	複数 → 単一
C	男	3~10 冊/月	感覚	単一 → 複数
D	男	3~10 冊/月	ロボット	複数 → 単一
E	男	3~10 冊/月	言語	複数 → 単一
F	男	1~2 冊/月	学習	単一 → 複数

参加者にあらかじめ用意された 10 個のキーワードの中から一つ、検索を行う際の最初のキーワード選択してもらう。その単語を始まりとしてシステムを利用し、気になる本に関しては「あとで読む」に登録するというタスクを行ってもらった。キーワードの選定は、事前に聞き込みを行い、参加者の調べたい分野で、辞書が反応できる抽象度の単語を収録した。辞書リストに登録されている辞書が、「はてなキーワード連想語 API^{*8}」のみの状態で行うタスク (単一辞書タスク) と個人連想辞書が 10 体 (表 6) いる状態で行うタスク (複数辞書タスク) の二つを用意した。参加者に同じキーワードでそれぞれのタスクを 15 分ずつ行ってもらった。タスクの順番は半数の実験者で入れ替えた。タスク中は発話思考をしてもらい、使用の仕方や辞書の特性に関して実験者に質問することは許可した。

両タスク終了後、二種類のアンケートに答えてもらった。一つは、AND 検索に利用した各連想単語に対して、その単語を選択した理由について聞いたものである。各連想単語に対して、「この連想単語を選択した理由として当てはまるものにチェックをつけてください」という設問文で以下のような回答で質問を行った。

- (1) 関連性があったから
- (2) 意外だったから
- (3) この単語を連想した辞書に信頼、親しみ、期待などがあったから

なお、その連想単語によって見つかった書籍があればその表紙の画像を、また、その連想単語を連想した辞書の名前とアイコンを併せて表示している。もう一つは、ResQue[17]を参考に作成した、タスクの全体的な印象評価に関するア

^{*8} <http://developer.hatena.ne.jp/ja/documents/keyword/apis/association> (2017/7/3 参照)

ンケートである。

4.2 全体的な傾向

AssocSearcher の利用データから得られた、各参加者各タスクにおける利用データを示す (表 5)。「連想による検索回数」とは連想単語を選択し、キーワードとの AND 検索を行った回数の中で、「書籍のバリエーション」では「あとで読む」に登録した書籍の説明文中から名詞を抽出し、重複を除いた単語数を重複を含む単語数で割った値を算出した。また、各イベントのタイムスタンプから、同じキーワードで、探索を行っていた時間である「トピック停留時間」と、詳細ビューを開いてから「あとで読む」に登録するまでの時間である「登録詳細ビュー停留時間」を算出した。各項目において、2 タスク間の t 検定を行ったところ、全ての項目に関して、有意差が認められなかった ($p > .05$)。連想による検索回数 ($t(6) = -2.05$)、トピック停留時間 ($t(6) = -2.05$)、関連性があると評価された連想単語の割合 ($t(6) = -2.53$)、の 3 項目で、2 タスク間での差に有意傾向が見られた ($p < .10$)。

4.3 観察された事例

ここでは、発話思考や、参加者へのヒアリングから観測された、成功事例、探索戦略の違いや辞書との関わり方の特徴について、表 5 での参加者別の結果、参加者が利用した辞書 (表 6)、AND 検索に利用された連想単語を連想した辞書の分布表 (表 7) 中の値に言及しながら述べる。

4.3.1 新たな観点からの書籍探索が行えた事例

参加者 A は、単数辞書タスクでは、キーワードである生物から始まり、辞書が連想した単語の中から選びトピックを変え、分類学、フクロウ、人間など、生物からそれほど離れずにトピックが移り変わっていき、参加者自身の印象でも「自分の好きな生物系の本が見つかった印象」と話しており、連想単語に対しての評価では、「関連性」の割合が 0.52、「意外性」が 0.16 と全体的に関連度の高い連想単語を用いて探索を行っていた。複数辞書タスクでは、「あとで読む」に登録する書籍での詳細ビューで「おー」や「へー」、 「んー」などと発話することが半数以上の登録書籍で確認され、アンケートでも「生物にどこか関連しつつ意外性があった面白そうな本が見つかった」と答えていた。また、連想単語に対しての評価においても、「関連性」0.11、「意外性」0.44 という割合になっており、単一辞書タスクと反対の傾向が見られた。

4.3.2 意図に応じた連想単語の選択

参加者 E は、複数辞書タスクでは、連想による AND 検索が 3 回と少ないが、全体的に、全ての連想辞書による連想を眺めており、「もうちょっと簡単な本」、「もうちょっと、聴き方に関することとか」といった、出てきてほしい本についての発話が見られ、そういった本が出てきそうな

表 5 参加者全体の利用データの平均

項目	A		B		C		D		E		F		平均	
	複数	単一	複数	単一	複数	単一	複数	単一	複数	単一	複数	単一	複数	単一
連想による検索回数	18	19	8	10	17	19	11	28	3	10	12	14	11.5	16.6
詳細ビュー表示回数	12	14	33	33	14	18	26	21	33	26	9	14	21.2	21.0
「あとで読む」登録回数	8	8	8	6	6	5	7	6	12	13	3	8	7.33	7.66
書籍のバリエーション	0.82	0.83	0.80	0.80	0.93	0.81	0.59	0.83	0.65	0.76	0.78	0.83	0.61	0.67
トピック停留時間(秒/連想)	105.0	109.5	327.5	195.8	141.7	116.4	339.5	139.8	442.4	154.1	147.0	165.4	250.5	143.8
「関連性がある」連想単語の割合	0.11	0.52	0.25	0.3	0.41	0.47	0.64	0.64	0.60	0.90	0.25	0.43	0.39	0.55
「意外性がある」連想単語の割合	0.44	0.16	0.38	0.40	0.59	0.52	0.18	0.29	0.33	0.30	0.33	0.29	0.38	0.33
登録時詳細ビュー停留時間	5.89	6.24	9.25	13.06	8.20	12.0	8.61	7.02	12.76	12.16	4.08	14.72	6.81	9.19

表 6 複数辞書タスクで参加者が利用する辞書

辞書	データの種類の	詳細情報
D1	書評	学者, 文化, 科学に詳しい
D2	小説, 評論	文豪
D3	ブログ記事	AI 系の記事の集成
D4	twitter, 研究論文	参加者達の指導教員
D5	ウェブ閲覧履歴	同じ研究室の学生
D6	twitter	同じ研究室の学生
D7	ブックマーク, ブログ, メモ	同じ研究室の学生
D8	twitter	同じ研究室の学生
D9	小説, 評論	文豪
D10	文書	六法全書本文

表 7 利用辞書の分布

参加者	辞書の分布 (%)
A	D8(33.3),D1(27.8),D5(16.7),D3(11.1),D4(5.6),D6(5.6)
B	D4(37.5),D1(25.0),D6(25.0),D3(12.5)
C	D6(23.5),D4(23.5),D8(17.6),D3(17.6),D5(11.8),D7(5.9)
D	D4(45.5),D7(27.3),D5(18.2),D3(9.1)
E	D1(66.7),D6(33.3)
B	D5(33.3),D2(33.3),D1(25.0),D3(8.3)

単語は何かを吟味していた。辞書 D1 の納得感の高い連想単語で「認知」「コミュニケーション」と言った、「言語」に関する書籍を登録していき、「せっかくだから突拍子もない単語で」と発話しながら、「英語」というキーワードに対しての D6 による「バカ」というワードで AND 検索を行い、書籍を 2 冊登録した。その際の本に対しての評価でも、「こういう本がむしろいい」と発話していた。

4.3.3 辞書の属人性に基づいた辞書の選択

参加者 C は、序盤から「文豪とかは硬い連想とかしてくれない気がする」、「身内の方が面白いと思う」と発話し、馴染みのない、D1, D2, D9, D10 にはほとんど連想させなかった。参加者 D は、「ロボット」に関係する単語を連想してくれそうな辞書に辺りをつけており、使用する辞書も、自分の指導教員と研究室のメンバーの辞書 (D4,D5,D7), AI 系のブログ記事を元にした D3 に使用が集中した。また、自分の予想では反応してくれるだろうと予想していた同僚の辞書である D6 が、「ロボット」という単語に答えられなかったことを残念がるという場面も見られた。参加者 B でも、検索に使用した辞書が 4 体と少なく、自分の研究分野に近い単語の時は、自身の指導教員や先輩の辞書を、話題を広げる目的では、文豪の辞書を使用するなど、連想辞書の使用を絞り込むという行動が観測された。また、どの参加者も自身にとって馴染みがない、D9, D10 を全く使

用しないといった状況が観測された。

4.3.4 探索過程で掴んだ辞書の特性に基づく使用辞書の選択

参加者 F は、前半での全辞書の連想を眺め特性を掴んだのち、「これは D2 に聞いてもわからないだろう」、「このキーワードなら D3 は無限に答えられる」「D2 横文字に弱い」など、あらかじめ辞書の特性に基づいて満足いく連想ができるかできないかを吟味する発話が多く見られた。参加者 A では、複数辞書タスクでは、基本的に、全辞書に連想させ、出現した連想単語を網羅的に眺めていたが、「D8 さんの連想面白いですね」と発話するなど、その後の AND 検索での使用回数は、言及した辞書に集中し、当該辞書の使用頻度が最も高い 33.3%であった。

4.3.5 未知の概念に対する反応

連想単語が自分にとって未知の単語の際に、その意味を把握しようとする行動が観測された。参加者 A は、D1 辞書が連想した単語が理解できない時に、その単語をシングルクエリとして検索を行って下のウェブ検索結果から意味を把握しようとし、また、その検索で出現した書籍を「あとで読む」に登録した。その連想単語に関するアンケートでは、「この単語を連想した辞書に信頼、親しみ、期待などがあつたから」という項目にチェックがされており、辞書の擬人性の影響があることも示唆された。参加者 B は、自身の指導教員の辞書である D4 の連想単語が未知の単語だったが、その単語で検索を行い、出てきた書籍の詳細情報を次々に開いていってその連想単語の意味をつかもうとする場面が観測された。

逆に、参加者 D は、「連想単語には意外性があつたが、その単語で検索しても全然関連のない本が出てきてしまったため、本の探索自体は楽しめなかった」とアンケートで答えており、意外性の高い連想単語の割合も 0.18 と低く、そういった単語を用いて登録された書籍もなく、個人連想辞書による未知の連想単語が効果的でない場面もあつた。

4.3.6 トピックの変更

トピック停留時間が単一辞書タスクにおいて短い参加者が多く、また、参加者 B では、最初のキーワードは「視線」であったが、中盤で自身が今気になっている「夢」にトピックを大幅に変更するなど、自分で考えた単語を新たにキーワードとする場面も見られた。ヒアリングでも「人に頼れ

ない、明確な目的もない探索なので、何をクエリにすればわからず、出てきた本を吟味していくしかなかった」「漠然と本を探すことは難しく感じた」などという声があった。

4.4 アンケートに基づくユーザ反応のまとめ

全体的なアンケートの結果を表8に示す。Q1, Q2, Q3に関してはそう思う～そう思わないの五段階評価、Q4は「足りない」、「ちょうど良いタイミングで終わった」、「時間が余った」の三段階評価である。t検定を行ったところ、Q1に関して両タスク間での有意差が認められた ($p < .05$)。

表8 アンケート結果の平均

アンケート項目	複数辞書	単一辞書
Q1 出てきた書籍は、意外性がある面白かった。	4.0	3.0
Q2 このタスクは新しい書籍を発見する助けになった。	4.3	3.7
Q3 見つけることのできた書籍の種類は、多様だった。	4.3	3.7
時間が足りないと感じた人数	5	1

5. 考察と今後の課題

5.1 考察

5.1.1 探索速度の低下

両タスク間で「あとで読む」に登録した書籍の数に変わりがなく、書籍のバリエーションにも差が見られなかった。全体として、時間単位でより多くの書籍の発見をもたらすことはできなかった、また、見つけれられた書籍の幅は広がらなかったと考える。辞書に連想させ眺める、連想単語を選ぶことに時間がかかってしまったため、検索数、連想数がより少ない傾向にあり、複数辞書タスクにおいて、全体的な探索速度を低下させることにつながってしまったと考える。

5.1.2 関連性を認識しない単語を用いた書籍の探索

複数辞書タスクにおいて、全体的により少ない「関連性」の連想単語を用いて探索していた傾向が見られ、また、アンケートでも、「出てきた書籍は、意外性がある面白かった」という項目において、より有意な差があったことから、より、ユーザが関連性を意識しない単語を用いた探索ができたと考えられ、また、全体として、両タスク間で「あとで読む」に登録した書籍の数に変化がないことから、その探索活動が、読みたいと思う書籍を見つけることを阻害せずに行われたということが示唆された。

5.1.3 探索活動の両条件での特性と持続性

単一辞書では、元の複数辞書に比べ選択肢が少ないこともあり、トピックを変更することが多く、キーワード空間を渡り歩いていくといった探索傾向が見られ、トピック停留時間もより短い傾向が見られた。逆に、複数辞書の場合は、いろいろな辞書の連想単語を使用することで、同じキーワードとのAND検索が続くことが多かった。比較的キーワード空間の周りを異なる切り口で照らしていくといった

探索傾向が見られ、漠然とテーマがあり、そこから関連書籍を探したり、関連がないとは言えないが、独特な切り口になりうる書籍を集めるといった目的においてより効果を発揮するのではないかと示唆された。

単数辞書タスクでは、「ネタ切れ」が起きる場面が何度か観測されたが、多くのAND検索候補を取りうる、複数辞書タスクではそのようなことは起こらなかった。タスク時間に関するアンケートに置いて、6人中5人が、「足りない」と答えており、トピック停留時間にも差がある傾向が見られたことから、一つのトピックから、その周辺トピックに関する書籍を探す活動を持続させるための支援が可能であると示唆された。

5.1.4 辞書の意図的な選択行為

参加者の発話思考から、タスクの中で、特定の辞書やその連想単語への関心から、よりその辞書に連想をさせたり、その連想単語を使用して検索をすることがあった。辞書を意図的に選択することが、書籍の発見に影響を及ぼしているということが示唆された。

使用する辞書を身内に絞る参加者がいたり、どの参加者も自身にとって馴染みがない、D9, D10を全く使用しないといった状況が観測された。辞書が多数あっても、特定辞書の使用が集中し、長期的な観点で見ると次第に自分の興味範囲外での新しい書籍の発見を促進できなくなってしまう可能性も考えられた。

5.1.5 ユーザによる連想辞書の個人差の認識

本研究では、辞書の連想単語そのものに関しての、定量的な評価は行わなかったが、参加者の発話の中でも、辞書に対しての、評価や連想単語の傾向予測、好みによる使用する辞書の選定に関する発話が多く見られた。また、アンケートでも、「辞書ごとに特性が見られたのでどんな単語が連想できるのか見てただけでも楽しかった」などといった評価が得られた。これらから、元となるデータの違いにより、それぞれその人らしきのある連想辞書が存在するという状況を作り出したのではないかと示唆された。

5.2 今後の課題

あまりにも関係のない単語群を眺める、試しにやってみても全く納得感のない結果が現れるなど、連想辞書が多様過ぎたり、その連想単語が一般性を欠きすぎて、逆効果が働く場面があった。また、今回の実験ではどの参加者も同じ10体の辞書を使用しており、辞書も身内の辞書が多く辞書の利用の変遷などが観測できた場面が少なかった。実験内容に関して、今回のタスクは、無目的に探索するタスクだったため、戸惑う参加者がいたり、詳細ビュー停留時間など、参加者単位で比較すると大きく変化が見られたが、全体の傾向としては認められなかった測定値があった。今後は、長期的な期間での実験を行い、詳細な参加者の思考傾向の分類を行った上での探索空間の違いや、使用する辞

書のマネジメントの変遷などに着目した検証を行いたい。

6. 今後の展望

実験でタスク中に、「他の人がこれについてどう思っているのかが気になる」と発話し、自分が気になっているワードで他の辞書に連想させる場面が複数見られた。他の人がある単語に対してどのようなことを考えているのかという情報は、ユーザが関心をもつことである。本研究で開発した AssocSearcher を用いて、それらが共有されているという状況下での人間の行動変容や、社会的インタラクションについて検証ができると考えている。自身の辞書が、的外れだったり、恥ずかしい連想をしないように、または、より多くの分野をカバーし、様々な人に使用してもらえるように、普段見ない情報のインプットを行うなど、行動変容があるのではないかと考えている。また、辞書の元となるデータを、時間や公私などによって切り分けることによって、一人の人間から新たに別の特徴をもった辞書を作成することができると思う。

7. まとめ

本研究では、他人の視点を取り入れることによって、ユーザに新たな価値や視点をもたらすことのできる書籍探索手法を提案した。そのために、それぞれのユーザの読み書き体験から個人連想辞書群を作成し、それらを利用し情報を探索するための Web アプリケーションである、AssocSearcher を開発した。

この AssocSearcher 上で単一大規模関連語辞書のみを利用し情報探索を行う条件と、複数の個人化辞書を利用し情報探索する条件で、見つけられる情報や、ユーザの行動に関する定量的、定性的な比較を行った。結果として、両条件で「あとで読む」に登録した書籍の数に変わりがなく、時間単位でより多くの新たな発見をもたらすことはできなかった。しかし、一つのトピックに関して、その周辺トピックに関しての書籍を探す活動を持続させるための支援が可能であると示唆された。また、元となるデータの違いによる各辞書の特性の違いを、参加者が認識し、参加者の好みや意図に応じた辞書の選択を行ったり、自分では思いつかなかった観念の単語を用いた、新たな書籍の発見をもたらすことができると示唆された。

参考文献

- [1] 丸川, 阿辺川: 横断的連想検索サービス「想 - IMAGINE」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.
- [2] 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.
- [3] 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.

- [4] J. Zhang, B. Deng, and X. Li, "Concept Based Query Expansion Using WordNet, pp. 52-55, 2009.
- [5] P.-A. Chirita, C. S. Firan, and W. Nejdl, Personalized query expansion for the web, In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 7-14, 2007.
- [6] T.-P. Liang, Y.-F. Yang, D.-N. Chen, and Y.-C. Ku, A semantic-expansion approach to personalized knowledge recommendation, Decision Support Systems, vol. 45, no. 3, 401-412, 2008.
- [7] 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.
- [8] J. Kamahara, T. Asakawa, S. Shimojo, and H. Miyahara A Community-Based Recommendation System to Reveal Unexpected Interests, In 11th International Multimedia Modelling Conference, 433-438, 2015.
- [9] O. Van Laere, I. Bordino, Y. Mejova, and M. Lalmas: DEESSE: entity-Driven Exploratory and sErendipitous Search SystEm, 2072-2074, 2014.
- [10] A. Thudt, U. Hinrichs, and S. Carpendale: The bohemian bookshelf: supporting serendipitous book discoveries through information visualization, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1461-1470, 2012.
- [11] T. Sakai and K. Nogami: Serendipitous search via wikipedia: a query log analysis, In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 780-781, 2009.
- [12] Z. Cheng, B. Gao, and T.-Y. Liu: Actively predicting diverse search intent from user browsing behaviors, In Proceedings of the 19th international conference on World wide web, 221-230, 2010.
- [13] O. Chapelle, S. Ji, C. Liao, E. Velipasaoglu, L. Lai, and S.-L. Wu: Intent-based diversification of web search results: metrics and algorithms, Information Retrieval, vol. 14, no. 6, 572-592, Dec. 2011.
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient estimation of word representations in vector space, arXiv:1301.3781, 2013.
- [15] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean: Distributed Representations of Words and Phrases and their Compositionality, arXiv:1310.4546, 2013.
- [16] Jeffrey Pennington, Richard Socher, and Christopher D Manning: Glove: Global vectors for word representation, Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014), 2014.
- [17] B. P. Knijnenburg, D. Bollen, and L. Schmidt-thieme: A User-Centric Evaluation Framework of Recommender Systems, In Proceedings of the ACM Conference on Recommender Systems (RecSys), Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI), 14-21, 2010.