

探索型検索を用いた学習ナビゲーションツールの検討 A study of learning navigation tool based on exploratory search.

村上 貴彦[†] 横山 誠[†] 中村太 戯留[†] 上林 憲行[†]
Takahiko Murakami Makoto Yokoyama Tagiru Nakamura Noriyuki Kamibayashi

1. 目的

我が国は近年、自律的・能動的に学習を行い、変化し続ける社会において自らの力で問題の解決を図れる人材を求めており、受動的な教育の場を見直し自律的・能動的な学習態度の育成が必要であると指摘している[1]。理由として、すべての学習者に対して最適な教材をあらかじめ用意することは物理的に不可能であるという問題があるからと考えられる。最先端の技術や知識が大学などのカリキュラムとして取り入れられ、教科書となるにはどうしてもタイムラグが生じる。また、一般的ではない専門的・文脈依存的な技術や知識には、明確な教材が用意されないこともある。これら教材が存在しない分野に取り組むには、学習者自らが情報を集め、教材をまとめ、計画し、学習・実践・評価をする必要がある。本研究では、このような学習者が教材を自ら見つけ、学習していく自ら学ぶ学習方略に向けた支援ツールの提案と実証的検討を行った。

2. 研究背景

最先端の分野、すなわち教師や教科書がない分野や文脈依存性の高い分野の学習では、最新かつ大量の情報へ容易にアクセスできる Web 検索が重要な役割を果たす。しかし、現状では Web 検索は主に答えの発見に活用し、コンテンツの内容をそのままレポートなどにコピーするという行動様式が目立つ。すなわち、学生は Web による調査レポートなど学習の機会が与えられていても、レポート提出などの目的を果たすだけの機械的なコピー作業に終始することが多く、Web コンテンツの内容を学習するまでには至っていない。そのため、Web 検索を答え発見の道具ではなく学習の教材として認識し、知識の獲得と実践へつなげる Web 学習の仕組みが必要である。

Marchionini[2]は、検索者にとって未知の課題に対する Web 検索では、Web 検索を行いながら探索の目標を明確化していく Exploratory Search (以下、探索的検索)を行っており、この探索的検索を支援する研究の必要性を主張している。三輪ら[3]は、学生のレポート学習および旅行のプランという課題において探索的検索の行動を観察した結果、探索的検索のスキルは「検索技能」「学習テーマごとのウェブ情報源の知識」、「探索的検索の経験的知識」、「学習テーマの事前知識」などの要因から、学習テーマごとに個別に定義される、すなわち学習テーマごとにスキルレベルが異なる一般化することのできないスキルであることを示唆している。また、中島ら[4]は、高度な検索サービスの登場により、学習者の検索技能よりも目的知識に到達できるキーワードが創出できる領域知

[†] 東京工科大学, Tokyo University of Technology

識の獲得が重要になったと指摘している。したがって、探索的検索には、領域知識の獲得が必要不可欠であるが、特定の分野の探索的検索の経験が、別の分野の探索的検索に寄与しづらく、学習者の必要とする分野ごとに領域知識の獲得のための支援が必要であることが考えられる。

三輪ら[3]は、探索的検索が充足された場合、通常の参照型検索 (Lookup) に移行すること、参照型検索時に気になるワードを発見した場合に探索的検索に移行することを指摘している。White[5]は、探索的検索は探索的閲覧 (Exploratory Browsing) と絞り込み検索 (Focused Searching) の二つの検索プロセスによる、検索空間の拡大と絞り込みの繰り返しによって行われるモデルを提唱している。したがって、探索的閲覧と絞り込み検索双方とその繰り返しに対して、それぞれ別のアプローチによる支援が必要であると考えられる。

Web による自律的・能動的学習を進めていくには、膨大な Web コンテンツの中から、教材を学習者自らが選択し学習する必要があるが、学習者は教材の発見や教材の妥当性の判断のために領域知識を必要とする。つまり、目的とする知識に到達するための領域知識を得るために繰り返し領域知識獲得のための教材の発見と学習を繰り返す探索的検索による学習を行う必要がある。その際に、学習者の知識レベルと合わないコンテンツしか見つかからない場合、学習が困難となってしまう。解決案としては、学習者に合わせてテキストの難易度を推定し推薦する仕組みが考えられる。佐藤[6]は、日本語書き言葉均衡コーパスを用い、対象テキストを難易度順に並べたテキストと照合することで、相対的難易度を 9 段階に測定する仕組みを示した。

3. 研究領域

3.1 自律的・能動的学習様式の検討

すべての学習者に求められている教材をあらかじめすべて移用することができないため、生涯にわたって自律的・能動的に学び続けられる人材が求められており、教育課程の質的転換を必要としている。そこで、現在の学習様式と新しく求められている学習様式について定義した。図 1 に現在のカリキュラムによる学習様式、図 2 に本研究の目指す自律的・能動的学習様式の概略図を示す。



図 1 既存の学習様式



図 2 自律的・能動的学習様式

図 1 のように、既存の学習方式は学習者の学習態度は受動的なものになってしまうことがわかる。原因として、A)計画 B)教材発掘 C)学習の評価 D)他者との協調などの要素が学習者以外の要因によって決定されており、学習者自らが判断・行動・決定する要素少なくなってしまうためである。したがって、学習者が自律的・能動的に学習する経験を得るためには、学習様式そのものが自律的・能動的である必要があると考えられる。そこで本研究では、自律的・能動的な学習様式を定義した(図 2)。自律的・能動的学習様式では、学習全般にわたって学習者が主体となって A)計画、B)教材発掘、C)評価、D)協調の各要素を実施していく。学習者以外人間は、あくまで協調学習者やアドバイザーとして関わり、学習のサイクルそのものには関与しない。つまり、教師という立場の存在しない学習様式が必要である。

本研究では、自律的能動的学習様式の A)計画、B)教材発掘、C)評価、D)協調の 4 つの要素のうち、既存の学習様式では学習者が関わることが少ないが学習の準備として必要と考えられる B)教材発掘に着目し、教材の発見の支援を目的としたナビゲーションツールを検討した。

3.2 探索的検索による教材発掘

自律的・能動的学習では学習者は自ら教材を発見する必要があることを述べたが、本研究では、Web 検索による教材発掘の枠組みとして探索的検索モデルを活用する。これは、学習者が目的となる教材を発見するために、領域知識の獲得を目的とした教材の発見と学習の繰り返す学習モデルである。White[5]による探索的検索モデルを参考にし、探索的検索による学習のタスクを定義する。①～⑤が探索的閲覧フェーズ、⑥～⑧が絞り込み検索フェーズである。また、図 3 にその概略図を示す。

問題解決のタスク

- ①課題の発見
- ②キーワードの発見
- ③クエリ創出・検索
- ④Web 教材の選択・学習
- ⑤学習が充足するまで②へ、充足したら⑥へ
- ⑥課題解決方法のためのクエリ創出
- ⑦解決方法の絞り込み
- ⑧解決法決定まで⑥へ、決定したら⑨へ、方針転換や気になるキーワードがある場合は②へ
- ⑨課題解決の実践。

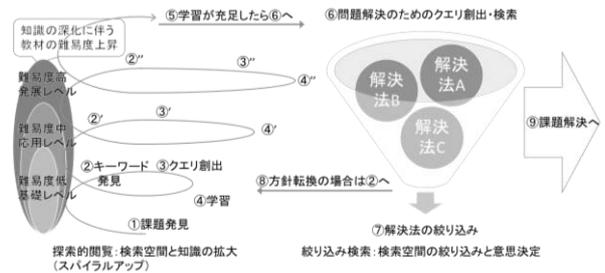


図 3 探索的検索モデル

本研究では、中島ら[4]により、探索的検索における領域知識の獲得の重要性から、②～④の探索的閲覧フェーズにおける、領域知識の獲得のために、検索結果から得られた Web コンテンツによる学習・理解の支援に着目した。

3.3 学習・理解の支援

White[5]は、学習・理解の支援のために「概要を含んだページを返す」「わかりやすいページを返す」「他のユーザの検索履歴を利用する」などを指摘している。このうち「わかりやすいページを返す」という点に着目し、そのための「学習・理解の支援」を学習者が学習目的の領域知識を深めていくために学習者のレベルに合わせた教材を提示することと定義し(図 3 左)、方法としてテキスト難易度の測定に着目し実証的実験を行った。

4. 実験

4.1 実験 1

佐藤[6]のテキスト難易度測定の手法から、プログラム obi2[7]を用い、実際に Web ページの難易度を測定した。難易度測定を行った Web ページのテーマは下記 A～D の 4 テーマである。

- A) 東京都 23 区の自治体サイトの出産・育児のテーマの母親向け Web ページの計 4214 ページ
- B) 東京都 23 区の自治体サイトの保険・年金のテーマの Web ページを計 1767 ページ
- C) Wikipedia の「インフォームド・コンセント」の記事における、記事のあるキーワード 29 ページ
- D) 「インフォームド・コンセント」の記事から手繰れるページを計 7312 ページ

情報提供を目的とした一般的な Web ページのサンプルとして、東京都 23 区の自治体サイトの A)出産・育児のテーマの母親向け Web ページ(計 4214 ページ)と B)保険・年金テーマの Web ページ(計 1767 ページ)の難易度を測定した。

専門知識の提供を目的とした Web ページのサンプルとして Wikipedia の「インフォームド・コンセント」の記事における、記事のあるキーワード 29 ページについて同様に分析し(以下、Wikipedia 分析結果 A)、比較した。URL の抽出方法は、実験者が各サイトに実際にアクセスし、内容を確認したのちにブラウザ(chrome)の履歴からコピーリスト化を行った。

また、Wikipedia の記事内リンクを自動で手繰るツールを開発し、「インフォームド・コンセント」の記事から手繰れるページを計 7312 ページリスト化し、難易度測定を行った(以下、Wikipedia 分析結果 B)。

難易度スケールは、obi2 の 13 段階学年区分を利用した。これは、小学 1 年レベルが難易度 1、高校 3 年レベルが難易度 12、それ以降を大学・一般レベル難易度 13 としている。

4.2 実験 2

人間の判断と obi2 の分析結果を比較し、難易度の判断に差が見られるかを検証した。Wikipedia 分析結果 B で使用した Wikipedia の記事を難易度 9 と難易度 13 の分析結果に分け、それぞれランダム抽出し、ペアとした。このペアをランダムに配置したリストを作成し、ペアごとに言葉づかいのより難しいページを選択してもらい、Wikipedia 分析結果 B の結果と比較し、人間の判断と obi2 による難易度判定の一致度を測った。ランダム抽出にあたって、記事ではないページと内容に問題があるページは排除した。記事のペア 32 組に対して、大学生の 4 年 9 名に比較をしてもらった。(図 4)

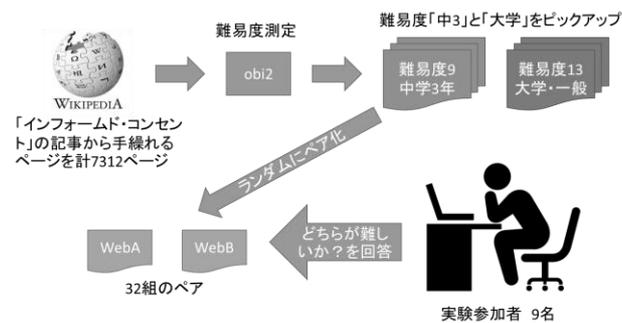


図 4

5. 結果

5.1 実験 1

東京都自治体の出産子育てページでは、東京都 23 区中 21 区について測定することができ、計 4214 ページのテキスト難易度を測定した。測定不能ページが 13 ページあった。表 1 に難易度別のページ数を示す。平均難易度は約 8.5 (中学 2 年~3 年) で全体の 91% のページが難易度 8 ~ 9 であった。

表 1 東京都 23 区出産・子育て難易度別ページ数

難易度	1	2	3	4	5	6	7	8	9	10	11	12	13
ページ数	0	1	1	6	14	67	118	1557	2292	120	18	4	3

東京都自治体の保険・年金のページでは、東京都 23 区中 21 区について測定することができ、計 1767 ページのテキスト難易度を測定した。測定不能ページが 156 ページあった。表 2 に難易度別のページ数を示す。平均難易度は約 9.15 (中学 3 年) で約 73.6% のページが難易度 8 ~ 9 で、約 26.3% が難易度 10 ~ 11 (高校 1 年~2 年) であった。

表 2 東京都 23 区保険・年金難易度別ページ数

難易度	6	7	8	9	10	11
ページ数	1	0	171	860	350	18

Wikipedia 分析結果 A, 29 件の内訳は表 3 に示す。全体の平均は約 11.8 (高校 3 年レベル) であった。

表 3 Wikipedia 分析結果 A

難易度	9	10	11	12	13
ページ数	1	5	3	9	11

Wikipedia 分析結果 B, 計 7312 ページの結果を表 4 に示す。全体の平均は約 10.6 (高校 1 年~2 年レベル) であった。なお分析不能ページが 60 ページあった。

表 4 Wikipedia 分析結果 B

難易度	3	4	5	6	7	8	9	10	11	12	13
ページ数	1	2	1	13	102	77	1749	2121	796	1219	1171

5.2 実験 2

図 5 に一致数別の記事のペア分布図、図 6 と図 7 に記事のペアそれぞれの有効 bigram の差 2000 (絶対値) を基準としてグループ化した一致数の分布を示す。Bigram とは 2 文字だけ続いた文字列のことである。実験協力者と obi2 のテキスト難易度判断の一致数の平均は全 32 ペア中の約 22.4 ペアで割合としては 70.1% だった。有効 bigram の差が >2000 でグループ分けした場合、一致数において、コンテンツ量の多いグループと少ないグループに有意差が見られた(有効 bigram の差 2000 基準による差: $t=1.754$, $df=30$, $p<.1$)。また、一致数と有効 bigram の差に負の相関がみられた($r=-.514$, $p<.01$)。

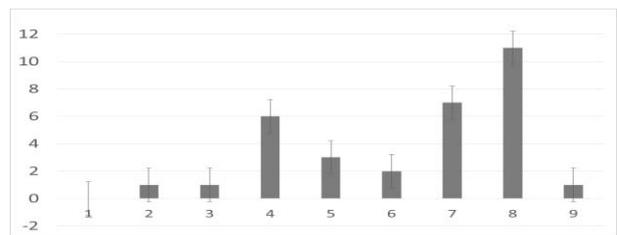


図 5 一致数ごとの分布

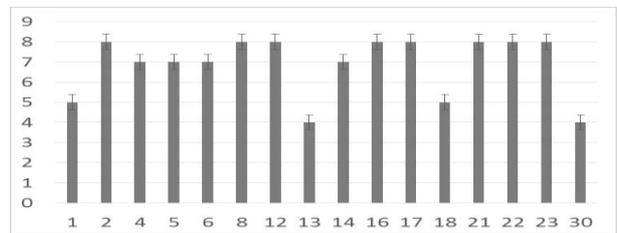


図 6 有効 bigram の差 <= 2000 グループにおける一致数

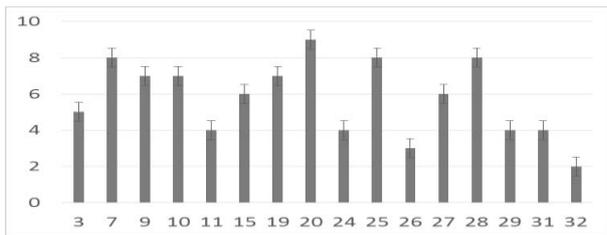


図7 有効 bigram の差 > 2000 グループにおける一致数

6. 考察

6.1 実験1

6.1.1 情報サイトの難易度とテーマによる難易度の差

東京都の21区行政ページにおける、出産・育児のテーマの母親向けページでは、中学2年生～3年生のテキスト難易度が一般的である可能性が示唆された。これらのページは、母親の学歴や初めて出産・育児をすることを考慮して作成されていると考えられた。

また、保険・年金テーマのページでは高校生2年までの難易度に分布していた。子育てテーマより平均で約0.64難易度の差があり、また、すべての自治体において、保険年金のテーマの方が高難易度であるという分析結果になった。(図8)このため、テーマによって作成される情報ページの難易度に差がでる可能性があることが示唆された。

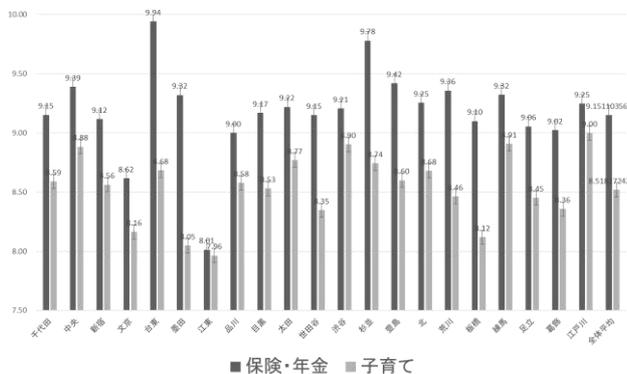


図8 出産・子育てと保険・年金テーマの難易度比較

6.1.2 Wikipedia における専門知識ページの難易度分析

Wikipedia 分析結果 A では約 69% が、Wikipedia 分析結果 B (計 7312 ページ) の結果では約 33% が高校3年～大学・一般レベルであった。Wikipedia では専門性の高い記事が多いため、一般的な情報サイトと比較すると難しくなると考えられた。したがって、Web ページの目的や対象とするユーザの違いによって、Web ページのテキスト難易度に差がある可能性が示唆された。そのため、段階的に理解を深めていく探索的検索による学習において、学習者のレベルに合わせたページの推薦が「わかり易いページを返す」探索的検索支援の一つとして必要であると考えられた。

6.2 実験2

6.2.1 人間の判断とテキスト難易度測定ツールの比較

判断の一致率は約 70.1% であった。そのため 3 割の不一致の原因を突き止める必要がある。今回実験 2 ではテキストの量 (bigram) の差が大きいページのペアと小さいページのペアで分け、一致率を比較した。

テキスト難易度判断の一致数において、テキスト量の差の大きいグループと小さいグループ間に有意傾向があったことと、有効 bigram の差とテキスト難易度判断の一致数に負の相関がみられたことから、実験協力者のテキスト難易度判断において、テキスト量の差が多いページはテキスト難易度判定ツールとの判断と異なる確率が増える可能性が示唆された。したがって、探索的検索における、「わかり易いページを返す」支援に、テキスト難易度以外の要素も考慮する必要がある可能性が示唆された。

6.3 総合考察

本研究では、自律的能動・能動的に学習し能動的に問題解決を図れる人材の学習支援を目的とし、自律的・能動的学習様式の定義を行った。また、自律的・能動的学習様式では教材発掘を学習者自らが行う必要があることに着目し、Web を活用した探索的検索による教材発見の支援を目的としたナビゲーションツールを検討した。そのために、「わかり易いページを返す」という探索的検索の支援の一つに着目し、実証的検討を行った。Web ページにはサイトごとや学習テーマごとにテキスト難易度の偏りがあることがわかった。そのため、学習進度別の教材を推薦する支援の必要性ある可能性が示唆された。また、「わかり易いページを返す」支援のために、テキスト難易度について分析を行ったが、テキスト難易度だけでなくテキストサイズなど、テキスト難易度以外にも考慮する必要がある可能性が示唆された。

謝辞

本研究を進めるにあたって、Web 公開されているテキスト難易度判定ツール obi2 を活用した。提供者の名古屋大学佐藤理史教授に心より感謝する。

参考文献

- [1] 中央教育審議会 答申 新たな未来を築くための大学教育の質的転換に向けて～生涯学び続け、主体的に考える力を育成する大学へ～, 9p, (2014).
- [2] Marchionini, G.: EXPLORATORY SEARCH: FROM FINDING TO UNDERSTANDING, COMMUNICATIONS OF THE ACM, Vol. 49, No. 4, pp.41-46 (2006).
- [3] 三輪 眞木子: Web 上の exploratory search の特徴: 発話プロトコルと事後インタビュー分析結果より, 情報処理学会研究報告. 情報学基礎研究会報告 2009-FI-96(2), pp.1-8, (2009).
- [4] 中島悠: 検索経験と領域知識の WWW 情報検索行動に与える影響, ヒューマンインターフェース学会, pp.1-10, (2006).
- [5] White, R. and Roth, R.: Exploratory Search: Beyond the Query-Response Paradigm, Morgan & Claypool Publishers (2009)
- [6] 佐藤理史: 均衡コーパスを規範とするテキスト難易度測定情報処理学会論文誌 Vol. 52 No. 4 pp.1777-1789, (2011).
- [7] Satoshi Sato 2014.10.10 帯 2: 日本語テキストの難易度推定 <http://kotoba.nuce.nagoya-u.ac.jp/sc/obi2/obi.html> (参照日 2015.01.24)