

探索的手法を用いた情報検索能力の分析

菅谷 克行[†]

[†]東京大学先端科学技術研究センター
sugaya@ai.rcast.u-tokyo.ac.jp

狩野 紀子[‡]

[‡]相模女子大学
n-kanou@star.sagami-wu.ac.jp

[概要] 情報検索能力は勉学を進めたり論文を書いたりする際、必要不可欠である。近年のインターネット技術の普及により、従来の図書館の文献目録に代わり、WWW 情報検索の機会が増えてきている。それにともない情報検索にはより多様な知識やスキルが求められるようになってきた。しかし最近、学生の情報検索能力の個人差が大きくなってきていると感じられる。この個人差の要因を明確にし、情報教育に生かすことが重要である。本稿では、WWW 上での情報検索能力に焦点を絞り、探索的手法を用いて情報検索能力を分析した結果を報告する。

1. はじめに

近年のコンピュータの大学教育現場への普及は目覚ましいものがあり、9 割近い学生が、ワープロ、インターネット、メールなどの操作ができると報告している(吉野, 野沢, 狩野, 2003)。一般の授業や研究にコンピュータを使用する機会は明らかに多くなっている。また、情報検索作業は勉学を進めたり論文を書いたりする際、必要不可欠であるが、近年のインターネット技術の普及により、従来の図書館の文献目録に代わり、WWW 情報検索の機会が増えてきている。それにともない情報検索にはより多様な知識やスキルが求められるようになってきた。この情報検索に必要な知識は、一般的な知識とは異なり、通常の学業とはあまり関連性がない上に、検索スキルが体系的・総合的に教授される機会もほとんどないため、ユーザーは自学自習によって自己流のスキルを身に付けてられているのが現状である(Vansickle, 2002)。そのためか、学生の情報検索能力の個人差が大きくなってきていると感じられる。

情報検索能力差を生じさせる要因として、WWW 情報検索経験、感情要因、認知要因などが研究されてきた。情報検索経験に関して、Lazander, Biemans, & Wopereis(2000)の実験では、WWW 情報検索経験が豊富な学生は経験が少ない学生より検索時間が短く、作業工程も少なく、検索結果もより適切であることを検証している。感情要因、特に Anxiety は、検索の持続性に影響を及ぼすことが報告されている(Wang, Hawk, & Tenopir, 2000)。認知要因に関する研究報告は多く、たとえば Palmquist & Kim(2000)は、認知的要因のひとつである「場依存—場独立」と情報検索能力との関係を調査し、場依存型の情報検索 Novice は場独立型の Novice よりも検索にかかる時間も長く、行う操作も多いと報告している。また、Ford ら(2001)は、認知スタイルと WWW 上の情報検索能力の関係を調べ、Imager は Verbalizer よりも検索に長けていると結論付けている。一方、メタ認知と情報検索能力に関する研究も進められており、吉岡(2002)はメタ認知(プランニングとモニタリング)過程を意識化させることが WWW 検索には重要であることを明らかにしている。

これまで概観したように、情報検索能力や方略の一側面を扱った研究は多い。しかし情報検索とは、様々な要因が複雑に関わっている複雑な過程である。そこで本研究では、探索的手法によるリサーチを行い、情報検索の Expert と Novice を比較分析することにより、情報検索能力に影響を与える要因を総合的に考察することを目的とする。さらに、情報検索に関わる認知要因を明らかにした上で、論文指導や情報処理の授業における適切な教授内容および教授方法を提案することを大きな目標としている。特に本稿では、学生の日常のコンピュータやネットの利用状況をアンケート調査した結果と情報検索能力の関係を、データマイニングの一手法である決定木学習を利用して分析した結果を報告する。

2. 調査方法

文系の学部所属する女子大学生の 57 名が、情報処理の授業の一部として調査・実験に参加した。情報検索実験は、学内のコンピュータ室(WindowsXP 環境)で WWW ブラウザを用いて検索課題を解いてもらった。被験者には、通常の情報検索の状態課題に取り組んでもらうために、利用するブラウザや検索エンジンには制限をしなかった。被験者は 15 分間で 10 問(20 点満点)の検索課題を二回 20 問(40 点満点)取り組んだ。各問は小問 2 つで構成されており、小問 1 は問題文中から適切なキーワードを入力しさえすれば答が容易に得られる問題、小問 2 は小問 1 で得られた答をもとに新たな検索をしなくては答を得られない問題である。

情報検索実験終了後に、普段のコンピュータやネットワーク利用に関するアンケートに回答してもらった。アンケート項目は、PC 利用経験(ある、なし)、自宅の PC 保有(個人、家族、なし)、自宅でのネット利用(可能、不可能)、WWW 利用歴(月数)、WWW 利用頻度(週×回)、主に利用する検索エンジン名、一回の検索に入力するキーワード数とその理由、他に知っている検索エンジン、論理式(AND, OR, NOT)を知っているか、Advanced Search を利用するか、とした。アンケート調査で得たデータと、情報検索課題の点数をもとに分類した被験者のレベル

(Expert, Novice) データを、決定木学習にかけて分析を行った。決定木学習は、機械診断やマーケティングなどの分野で多く利用されている手法で、傾向分析、診断、予測などの点で優れている (Witten & Frank, 2000)。

3. 結果と考察

検索課題の解答結果は平均 8.7 点 (標準偏差 3.13) だった。この平均点を基準とし、9 点以上の学生を Expert, 8 点以下の学生を Novice と分類した。この結果, Expert は 30 名, Novice は 27 名に分類できた。この分類結果とアンケートで得たデータを機械学習の一つである決定木学習にかけて、データマイニングの手法を用いて、分析に利用した。今回、決定木学習アルゴリズムとして一般的な C4.5 を用い、生成された決定木は危険率 0.25 で枝刈りした。生成された決定木を図 1 に示す。

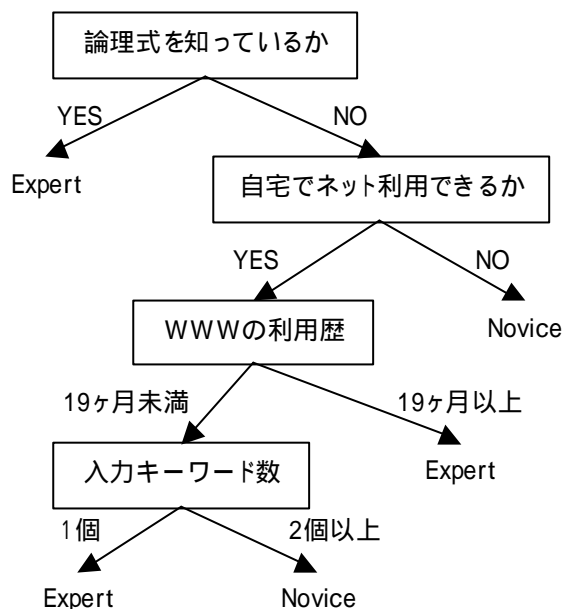


図 1: C4.5 で生成された決定木

生成された決定木から、情報検索能力に影響を与える要因として、以下のようなルールを得ることができた。それは、論理式を知っている 自宅でインターネット利用ができる WWW利用歴が一年半以上 入力キーワード数が 2 個以上 という 4 つである。は、検索範囲を絞り込んだり広げたりする際に必要な知識であるが、普段はあまり意識しない知識である。今回の結果から、情報検索の授業でも論理式をしっかりと教えることの重要性が明確になった。とは、事前に予想していた結果と言える。近年、自宅でもインターネット利用できる学生は増えているが、様々な理由から、そうでない学生もいる。デジタル・デバイドの影響が、このような形で明らかになったことは、教育機関の環境整備への一つの提言になると考えている。は、非常に興味深い結果であった。検索範囲を絞り込む際に、複数のキーワードを入れることは重要であるが、論

理式の知識がなく WWW 利用経験も少ない学生には、複数キーワードがうまく機能していないということである。この点からも、と同様、WWW 上で必要な情報を得るために、検索空間を適切に絞り込んだり広げたりすることができるのに必要な知識を教示することの重要性を再確認できた。

4. まとめと今後の課題

本稿では、日常の PC 利用状況が WWW 情報検索能力に影響を与える要因を、決定木学習を利用して探索的に分析を行った結果を報告した。分析の結果、論理式を知っている 自宅でインターネット利用ができる WWW利用歴が一年半以上 入力キーワード数が 2 個以上 という 4 つの要因を得ることができた。機械学習から得たものであるため、100%正しいということではないが、被験者の傾向を十分示している。この結果より、論理式の知識の重要性、デジタル・デバイドに対する配慮の必要性、などが明らかになった。

今後の課題としては、今回得られた決定木を用いて他の学生の能力を予測し検証すること、データ数を増やして頑健なルールを得ること、他のデータマイニング手法を利用した分析、などが挙げられる。

また、情報リテラシーや情報処理の授業において、ここで得た知見を生かして教育内容を検討・提案することも重要な課題だと考えている。

参考文献

- Ford, N., Miller, D., and Moss, N. (2001) The Role of Individual Differences in Internet Searching: An Empirical Study. *Journal of the American Society for Information Science & Technology*, 52 (12), 1049-1066.
- Lazander, A.W., Biemans, H.J.A, and Wopereis, I.G.J.H. (2000) Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science & Technology*, 51 (6), 576-581.
- Palmquist, R.A., and Kim, K.S. (2000) Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science*, 51 (6), 558-566.
- Vansickle, S. (2002) Tenth Graders' Search Knowledge and Use of the Web. *Knowledge Quest*, 30 (4), 33-37.
- Wang, P., Hawk, W.B., and Tenopir, C. (2000) User's interaction with World Wide Web resources. *Information Processing and Management*, 36, 249-270.
- Witten, I.H., and Frank, E. (2000) *Data Mining*. Morgan Kaufmann.
- 吉岡敦子 (2002) インターネット情報検索行動に及ぼすメタ認知過程の意識化の効果. 日本教育工学会論文誌, 26 (1), 1-10.
- 吉野志保, 野沢智子, 狩野紀子 (2003) 日本人大学生の英語自主学习・英語学習の動機・英語授業に対する態度: コンピュータ所有・コンピュータ技量・CALL 授業経験との関係. *Language Education & Technology* 40, 211-229.