
ソフトウェア開発における Web 検索行動の分析

Analysis of web search action in software development

中才 恵太郎* 角田 雅照†

あらまし インターネット上には、プログラミングを行う際に有用な資料が多く存在しており、検索エンジンを活用して資料を参照することにより、作業効率を高めることができる。ただし、検索行動には個人差があり、効率のよい行動を取らないと目的の知識を得ることができない。本研究では、プログラミング時における検索行動を分析することにより、検索方法の指針を得ることを試みた。

1 はじめに

インターネット上には、プログラミングを行う際に有用な資料が多く存在しており、プログラミング時に検索エンジンを活用して資料を参照することにより、作業効率を高めることができる。プログラミング言語の公式リファレンスの多くはインターネット上に公開されている。その他に、プログラミングに有用なサイトとして、プログラミングに関する Q&A サイト、プログラミング言語入門のサイトがあげられる。Google に代表される Web 検索エンジンで検索を行うことにより、これらの Web サイトの URL (Uniform Resource Locator) を知らなくとも、これらの Web サイトを閲覧することができる。

ただし、検索行動には個人差があり、適切な検索行動を行わないと必要な情報を得ることができないため、作業が滞ってしまう。そこで本研究ではプログラミング時の検索行動の指針を得ることを目的とし、プログラミング時の検索行動を分析した。特に、検索対象の知識（専門知識）を持たない場合に、探している知識を発見できる場合とそうでない場合の検索行動の差異について着目する。

専門知識も持つ人とそうでない人とは検索行動が違い専門知識を持つ人のほうがより適切な行動を行えることが示されている[2]。しかし、ソフトウェア開発のように複雑で高度な知的活動である場合でも検索行動が違うのか、より適切な戦略を示した研究は存在しない。また、ソフトウェア開発の活動において個人の能力に差があることがいくつかの研究で示されているが[2][3][4][5][6][7]、我々の知る限り、ソフトウェア開発における Web 検索の行動に着目し、行動の差異を明らかにした研究は存在しない。

2 実験方法

本研究では、プログラミング時の検索行動を分析するために、プログラミングに関する問題を作成した。また、問題解答プログラム（被験者の解答を確認し、正解の場合に次の問題を表示する）を作成して実験環境を構築した。その後、問題を被験者に解答させ、アンケートを実施した。実験後に、プログラミング時の検索行動を、アンケート、問題の正誤、解答時間に基づいて分析した。

情報系学科に所属している学部生 9 人と修士学生 1 人の計 10 人を被験者とした。分析では、効率の高い検索行動の差異を定量的に分析するために被験者をプログラミングが

* Keitaro Nakasai, 近畿大学

† Masateru Tsunoda, 近畿大学

得意であるか不得意であるか、問題の正答率、回答時間についてグルーピングを行い 2.2 節において定義した 5 つの項目について、差異を確かめた。得意か不得意かどうかは被験者に対し、実験実施後にアンケートを行い、プログラミングが得意であるか不得意であるかを解答してもらった。

実験では Windows7 のノート型パソコンを使用し、使用するブラウザ、検索エンジン、開発環境を統一した。検索時の行動を計測するために、キー入力やアクティブにしたウィンドウタイトルなどを記録するソフトウェアを起動した。以降では、問題、分析項目のそれぞれについて詳細を述べる。

2.1 問題

被験者全員が理解しているプログラミング言語である Java を題材とした問題を作成した。また、開発環境も被験者全員が使用したことがある Eclipse を指定した。問題は全部で 4 問とした。紙面の都合上、問題 1 と問題 2 についてのみ解説する。

- 問題 1 : Java では、小数点の計算において誤差が生じる場合がある。そこで、Java で正確な計算を行うためのライブラリを使い、正しい計算を行う問題を作成した[1]。適切なライブラリを探す必要がある問題である。
- 問題 2 : ArrayList を使ったデータ構造で実装されたプログラムから、連想配列を使ったデータ構造に書き換える問題である、ただし、連想配列を直接使うとは問題文には記載しておらず、どのクラスを使うのかを推測する必要がある問題である。

いずれの問題も、検索を適切に利用することができれば容易に正解に到達できるように問題を作成した。被験者の実験に対する慣れや実験時間の制限を考慮し、前半の問題は難易度を低くし、後半の問題は難易度が高くなるように順序を設定した。検索対象の知識を持たない場合に、探している知識を発見できる場合とそうでない場合の検索行動の差異について分析するため、検索を行う際のタイミングや、検索を行う内容については特に指定せず、自由に行ってもらった。

解答時間は指定せず無制限としたが、各問題で解答時間が 20 分以上経過している場合、問題をスキップすることができるようにした。スキップした場合は不正解とした。問題を正解した場合、スキップした場合のどちらも、現在解答している問題以外の閲覧、解答はできないようにプログラムで制御した。これは、現在取り組んでいる問題における経過時間や検索回数などをできるだけ正確に計測するためである。

2.2 分析項目

以下の 5 項目を定義し、被験者に対する実験を行った後に、それぞれについて集計を行った。

2.2.1 検索結果ページ数÷表示した Web ページ数

検索結果ページ数とは Google での検索結果リストを表示した数を表す (Google の検索の設定でのページあたりの表示件数を 10 とした)。表示した Web ページ数とは Google での検索結果リストとその検索結果リストからアクセスした Web ページ、アクセスした Web ページから別のページを参照した数を合計したものである。

この項目を定義した目的は、検索結果ページから、結果に示されている各ページにアクセスせず、タイトルやスニペット (Web ページの説明文やタイトル) だけで検索キーワードを変えるかどうかなど判断をしたほうがよいのか、逆に、検索結果ページから実際に各ページへアクセスして内容を確認する、もしくは各ページの内容を深く見る (アクセスした Web サイトさらにリンクをたどる) ほうがよいのかを分析するためである。

この項目の値が高い場合、検索結果から別のページを参照した回数 (検索ページリストを多く見ている) が少ないことを示し、逆に値が低い場合、検索結果から別のページ

を参照した回数が多いことを示す。

2.2.2 キーフレーズの種類÷検索結果ページ数

キーフレーズの種類とは、検索時に入力された全ての検索キーフレーズから、重複を取り除いた数を指す。この項目の値が高い場合、被験者は検索キーフレーズを頻繁に変更していることを示す。逆に、値が低い場合、被験者は検索結果ページリストを掘り下げていることになる。なお、同一のキーフレーズで検索を何度か行ったり、ブラウザの戻るボタンを多用した場合も、この値が低くなる。

2.2.3 キーワードの種類÷キーフレーズの種類

キーワードの種類とは、検索におけるキーフレーズをキーワードに分けて、重複しているものを除いた数である。Google が公開している検索のヒントでは、検索のキーワードは少ないほうがよいと指摘されている。この方針がプログラミングにおいても適切であるかどうかを確かめるため、この項目を定義した。この項目の値が高い場合、毎回複数個のキーワードを並べて検索していることを示し、逆に、値が低い場合、少ないキーワードで検索していることを示す。

2.2.4 検索結果ページ数÷問題解答時間

この項目の値が高い場合、問題に解答している時間や検索結果以外のページを閲覧するよりも、Web 検索を多く行っていることを示す。逆に、値が低い場合、検索結果（検索結果リストからヒットしたサイトや検索結果リストのタイトルやスニペットの内容について）を詳細に閲覧していることがわかる。

2.2.5 キーワードの種類÷キーワードの合計

キーワードの合計とは、キーワードの重複を許して合計したものである。この項目の値が高い場合、同じキーワードを何度も別のキーフレーズ中で用いず、キーワードを新たに考案していることが多いことを示す。逆に値が低い場合、同じキーワードを各キーフレーズ中で多用していることを示す。この項目を定義した理由は、検索効率を高めるためには、類似したキーワードで検索したほうがよいのか、もしくは毎回キーワードを大きく変更したほうがよいのかを分析するためである。

3 実験結果

問題 1 については正解者が 10 人中 7 人、問題 2 は正解者が 8 人であるのに対し、問題 3, 4 は正解者が 6 人であった。よって、問題 1, 2 は比較的簡単な問題であり、問題 3, 4 は難しい問題であるといえる。このことから、出題順序は比較的適切であった（前半の問題は難易度が低く、後半の問題は難易度が高い）といえる。なお、3 人の被験者が全問正解し、1 人の被験者は全問不正解であった。実験後にアンケートを取ると、検索すべき技術をあらかじめ理解していた被験者は少なかった。すなわち、被験者の知識に大きな違いはないといえる。

以降、2.2 節で定義した項目を、プログラミングを得意と回答したかどうか(2 章参照)、解答時間の長短、問題の正誤によりグルーピングをした結果を示す。表 1 は各項目の平均値を示し、表 2 は各項目の中央値を示す。キーワードの種類÷キーフレーズの種類を除き、それぞれの値は 1 を下回っているため、グループ間に 0.1 程度の差がある場合、相対的には 10%以上の差があるといえる。

3.1 検索結果ページ数÷表示した Web ページ数

検索結果ページ数÷表示した Web ページ数の平均値を見ると、得意、解答時間短、正答グループのいずれの場合も、不得意、解答時間長、誤答グループよりも値が低かった。中央値においても同様の結果となった。正答と誤答のグループでは、中央値の差は 0.01

表 1 各問題における各項目の平均値

	全体	得意	不得意	解答時間短	解答時間長	正答	誤答
検索結果ページ数÷表示した Web ページ数	0.53	0.44	0.57	0.48	0.58	0.51	0.57
キーワードの種類÷検索結果ページ数	0.48	0.58	0.44	0.51	0.45	0.51	0.41
キーワードの種類÷キーワードの種類	2.44	1.80	2.71	2.24	2.64	2.40	2.52
検索結果ページ数÷問題解答時間	0.73	0.48	0.84	0.55	0.91	0.69	0.82
キーワードの種類÷キーワードの合計	0.59	0.78	0.52	0.72	0.47	0.63	0.50

表 2 各問題における各項目の中央値

	全体	得意	不得意	解答時間短	解答時間長	正答	誤答
検索結果ページ数÷表示した Web ページ数	0.56	0.47	0.59	0.50	0.58	0.55	0.56
キーワードの種類÷検索結果ページ数	0.45	0.52	0.40	0.50	0.39	0.50	0.37
キーワードの種類÷キーワードの種類	2.38	1.83	2.77	2.31	2.44	2.33	2.38
検索結果ページ数÷問題解答時間	0.61	0.46	0.69	0.48	0.73	0.59	0.69
キーワードの種類÷キーワードの合計	0.53	0.75	0.46	0.67	0.43	0.58	0.44

しかなかったが、得意と不得意、解答時間短と解答時間長のグループ間では、平均値で 0.1 以上、中央値で 0.08 以上の差があった。このことから、効率の高い検索行動をおこなった場合、この値が低くなる可能性があると考えられる。

この項目の値は、検索結果のページからそれ以外のページを多く参照した時や、検索結果以外のページから他のページにアクセスした場合にも低くなる。すなわち、スニペットだけ、もしくは Web ページのタイトルを参照しただけで検索結果が違くと判断し、すぐに別の検索キーワードを試すことは適切ではない可能性がある。一見調べようとしている内容と無関係に思える検索結果の場合でも、念のためそれらの Web ページにアクセスし、内容を確認したほうが、プログラミングの作業効率が高まる可能性がある。

3.2 キーフレーズの種類÷検索結果ページ数

キーワードの種類÷検索結果ページ数の平均値は、得意、解答時間短、正答のいずれのグループも、不得意、解答時間長、誤答のグループよりも高かった。中央値でも同様の結果となった。得意と不得意のグループの平均値の差と、正解者と不正解のグループの差は 0.1 以上あるのに対し、解答時間短と解答時間長のグループの平均の差は 0.06 とあまり大きくなかった。中央値に着目すると、得意と不得意グループ、解答時間短と解答時間長グループ、正答と誤答グループの差はそれぞれ、0.12、0.11、0.13 となった。よって、検索行動の高い行動を行った場合、この項目の値が比較的高くなる可能性がある。

この値が高い場合、検索する際に検索キーワードを頻繁に変更している、あまり検索リストを掘り下げてみていない。つまり、このことから、検索結果の 1 ページ目を精読し、その後別のキーワードを考えて検索するほうが、検索結果ページを掘り下げるより

プログラミングの作業効率が高い可能性がある。

3.3 キーワードの種類÷キーフレーズの種類

キーワードの種類÷キーフレーズの種類の平均値は、得意、解答時間短、正答のいずれのグループも、不得意、解答時間長、誤答のグループよりも値が低かった。中央値でも同様の結果となった。しかし、得意と不得意のグループにおいて、平均の差は 1.09 であったのに対し、解答時間短と解答時間長のグループの差は 0.4 であり、あまり大きくなかった。さらに、正答と誤答の差も 0.12 であり、あまり大きくなかった。中央値に関しても、得意と不得意のグループ間の差が大きく、その他のグループ間では差が大きくなかった。正答している場合でも熟練者であるとは限らず、その他のグループ間ではこの項目の値の差が小さかったことから、今回の実験結果からは、この項目とプログラミングの作業効率とに関連があるとはいえなかった。

この項目の値が低い場合、少ないキーワードで検索していることを示すが、Google の検索のヒントにおいても、検索キーワード数は少ないほうがよいと指摘されている。プログラミングにおける検索の場合も、キーワード数を少なくすることにより効率が高まる可能性があるが、今回の実験結果からは検索キーワード数は少ないほうがよいと結論付けることはできなかった。

3.4 検索結果ページ数÷問題解答時間

検索結果ページ数÷問題解答時間の平均値は、得意、解答時間短、正答グループのいずれも、不得意、解答時間長、誤答グループより値が小さかった。中央値も同様の結果となった。正答と誤答グループの平均値と中央値の差異は 0.1 ほどで他のグループより差が少ない。ただし、得意と不得意グループの差と、解答時間短と解答時間長グループの平均値の差は 0.36 であることから、効率の高い検索行動を行った場合はこの項目の値が低い傾向にあると考えられる。

この項目の値が小さい場合、問題を解くために検索を多数繰り返すことをしていないことを示す。すなわち、検索結果リストからヒットしたサイトや検索結果リストのタイトルやスニペットの内容について詳細に閲覧していることを示しており、そうすることによりプログラミングの作業効率が高まる可能性がある。

3.5 キーワードの種類÷キーワードの合計

キーワードの種類÷キーワードの合計の平均値は、得意、解答時間短、正答グループのいずれも、不得意、解答時間長、誤答グループよりも値が大きかった。中央値も同様の結果となった。なお、正答と誤答グループ間の差は、他のグループよりも小さかったが、得意と解答時間短のグループは不得意、解答時間長のグループよりもこの値が大きいことから、効率の高い検索行動を行った場合はこの値が高い傾向にある可能性がある。

この項目の値が高い場合、同じ検索のキーワードを多用していないことを示している。すなわち、一度使用した検索のキーワードを含むキーフレーズを再び用いることはプログラミング時の検索行動としては効率が高くない可能性がある。プログラミング時には、毎回新たなキーワードを用いることを考慮して検索を行うことにより、効率が高い検索ができる可能性がある。

4 おわりに

本研究では、プログラミングにおける効率の高い検索行動を明らかにするために、プログラミング時における検索行動を定量的に計測する実験を行った。その際、実験結果やアンケート結果に基づき、被験者をグルーピングし、5 つの項目に関してグループ間の差異を分析した。その結果、以下の検索行動を行うと効率が高い可能性があることが

わかった。

- スニペットだけ、もしくは Web ページのタイトルを参照しただけで検索結果が違くと判断し、すぐに別の検索キーワードを試すことは適切ではない可能性がある。
- 最初の検索結果リストや検索結果のリンクをたどった先のページを熟読することによりプログラミングの作業効率が高まる可能性がある。
- 一度使用した検索のキーワードを含むキーフレーズを再び用いることはプログラミング時の検索行動としては効率が低い可能性がある。

今後の課題は、被験者に出題する内容をさらに検討することと、分析に用いる項目を新たに定義し、被験者により、それらの項目の値が異なるかどうかを分析し、最適な検索行動についてさらに検討することである。

謝辞 本研究の一部は、文部科学省科学研究補助費（挑戦的萌芽：課題番号 26540029、基盤 C：課題番号 25330090）による助成を受けた。

5 参考文献

- [1] ジョシュア・ブロック, ニール・ガフター: 罫, 落とし穴, コーナーケース Java PUZZLERS, ピアソン・エデュケーション (2005).
- [2] Ko, A., and Uttl, B.: Individual differences in program comprehension strategies in unfamiliar programming systems, Proc. International Workshop on Program Comprehension, pp.175-184 (2003).
- [3] Rifkin, S., and Deimel, L.: Program comprehension techniques improve software inspections: a case study, Proc. International Workshop on Program Comprehension, pp.131-138 (2000).
- [4] Sackman, H., Erikson, W., and Grant, E.: Exploratory experimental studies comparing online and offline programming performance, Communications of ACM, vol.11, no.1, pp3-11 (1968).
- [5] 高田義広, 鳥居宏次: プログラマのデバッグ能力をキーストロークから測定する方法, 電子情報通信学会論文誌 D-I, vol.J77-D-I, no.9, pp.646-655 (1994).
- [6] Thelin, T., Andersson, C., Runeson, P., and Dzamashvili-Fogelstrom, N.: A replicated experiment of usage-based and checklist-based reading, Proc. International Symposium on Software Metrics, pp.246-256 (2004).
- [7] Uwano, H., Nakamura, M., Monden, A., and Matsumoto, K: Exploiting Eye Movements for Evaluating Reviewer's Performance in Software Review, IEICE Transactions on Fundamentals, vol.E90-A, no.10, pp.317-328 (2007).
- [8] White, R., Dumais, S., and Teevan J.: Characterizing the influence of domain expertise on web search behavior, WSDM '09 Proceedings of the Second ACM International Conference on Web Search and Data Mining Pages 132-141 (2009).