

利用者の問合せ理解度を考慮した Web ページ検索手法

水野広大[†] 鈴木優[†] 川越恭二[†]

[†]立命館大学 情報理工学部

1 はじめに

利用者がプログラミングを学習する際に、Web 上から学習内容の問合せに対する情報を利用者が取得することが行われている。例えば、アプリケーションを作成するために、利用者が Web ページのプログラムソースを参考にするといった事例が挙げられる。この場合利用者は、問合せであるキーワードに対する Web ページの適合度合いによる検索結果を検索エンジンから得ることになる。しかし、プログラミングの学習において、検索された Web ページ内のプログラムソースが利用者にとって知らない事項が多く、理解することができない内容である場合がある。それは利用者にとって有効な検索結果であるとはいえない。そのため、同じ学習内容に関する問合せであっても知らない事項が多い利用者と少ない利用者との間で、有効であると考えられる情報に差異が生じることが考えられる。つまり、利用者が学習を行う際に、利用者の問合せだけから Web ページを提示することは必ずしも適切ではないと考えられる。

そこで本研究では、利用者の問合せ結果である Web ページ群に対する利用者ごとの理解度を算出するための手法の提案を行う。提案手法では、まずプログラムソースの処理内容を利用者にとって知っている事項と知らない事項に分別する。次に、利用者ごとに各 Web ページ内のプログラムソースに対する理解度を算出する。算出した理解度から、利用者が Web ページ内のプログラムソースを理解できる検索結果を出力する。プログラミングの学習ではプログラムの内容を理解することが必要であるため、この検索結果はプログラミング学習の支援に有効であると考えられる。

2 理解度を考慮した Web ページ検索手法

2.1 利用者の理解度

本研究における利用者の理解度とは、利用者がプログラムソースをどの程度理解できるかを表す指標である。ここでいう理解度とは、プログラムフォルダ内で使用した事のあるメソッドの個数から算出される数値のことである。また、本研究におけるメソッドとは、利用者がプログラム作成の際に JAVA ライブラリから参照したメソッドを指す。提案手法を利用する際、利用者のプログラムフォルダには、利用者が作成した全てのプログラムソースが保存されているとする。これらのプログラムソースに含まれるメソッドは、少なくとも一度は利用者が理解した経験を持っていると考えられるため、利用者が知っている事項であると本研究では評価する。また、プログラムフォルダに存在しないメソッドについては学習経験のない内容であると考えられるため、利用者の知らない事項であると本研究では評価する。

2.2 Web ページに対する理解度の算出

Web ページに対する理解度とは、利用者が Web ページ内に存在するプログラムソースのメソッドを理解している割合を表した値である。本章において Web ページに対する理解度の算出手法を述べる。

提案手法では Web ページに対する理解度を、Web ページのプログラムソース内に含まれる JAVA ライブラリのメソッドのうち利用者のプログラムソースに含まれていないメソッドが占める割合から算出する。

ここで Web ページ i のプログラムソース内のメソッドの個数を m_i 、そのうち利用者のプログラムソースに含まれていないメソッドの個数を m_u とすると、Web ページ i に対する理解度 U_i は、式 (1) により算出される。

$$U_i = 1 - \frac{m_u}{m_i} \quad (1)$$

$0 \leq U_i \leq 1$ であり、 U_i が 0 に近い Web ページのプログラムソースについては、知らないメソッドの割合が大きい点から、利用者は理解が困難であるといえる。

A Web Page Retrieval Method Based on Users' Programming Knowledge

Koudai MIZUNO[†], Yu SUZUKI[†] and Kyoji KAWAGOE[†]

[†]College of Information Science and Engineering, Ritsumeikan University.

mizuno@coms.ics.ritusmei.ac.jp

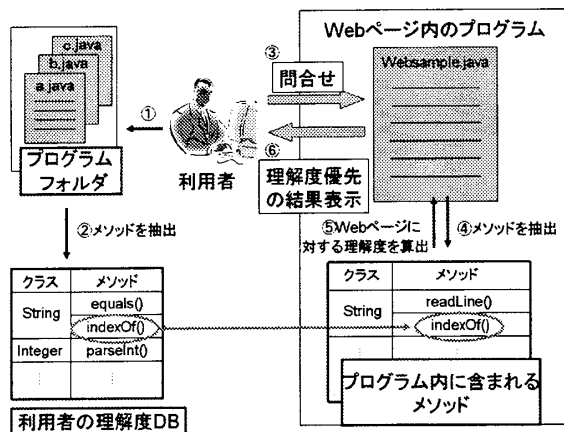


図 1: 提案手法の概要図

図1は提案手法の概要図である。まず2.1節より、利用者が問合せを行う前に、プログラムフォルダの利用によって、あらかじめ理解度DBに利用者が知っているメソッドを登録しておく。利用者が問合せを行ったときに理解度DBを利用し、問合せに適合したWebページ内に含まれるプログラムソースに知らないメソッドが少ないWebページから優先して表示する。

3 実験

3.1 実験条件

提案手法の有効性を示すために評価実験を行った。本実験では、従来手法と従来手法の後処理として提案手法を加えた手法との比較により、出力されるWebページの違いを示した。従来手法として、Webページを検索する際に用いられるGoogleを用いた。それぞれの検索結果における上位30件のWebページに対し、利用者にはプログラミング学習をする上で役に立ったかどうかという観点で、それぞれ10件の正解集合をあらかじめ決定させた。そして手法ごとに、問合せであるキーワードに対する検索結果を出力した。

3.2 実験結果及び考察

JAVAのプログラミング演習の授業を修了した学生を対象に、「JAVA GUI サンプル」という問合せに対する従来手法と提案手法の有効性を、再現率適合率曲線で示した結果を図2に示す。

図2から、従来手法に比べ提案手法を加えた手法が正解を上位に出力させている点から、提案手法に有効性があることが確認できる。提案手法で正解が上位に集まった理由として、知らない事項が多いプログラムソースに比べ、知らない事項が少ないプログラムソ

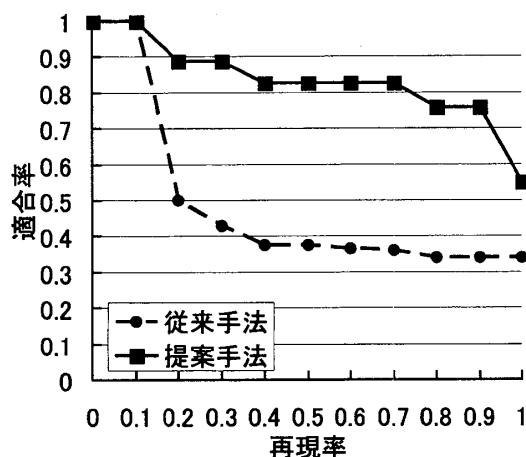


図 2: 実験結果

スは理解が容易であったことが考えられる。

一方で、提案手法は、利用者の知っている事項と知らない事項を分別する際にプログラムソースを利用しているが、正解集合の中にはプログラムソースを含んでいないWebページも確認された。よって、知っている事項と知らない事項の分別に、プログラムソースの利用だけでは、不十分であると考えられる。

4 おわりに

本稿では、利用者の作成したプログラムソースを利用することにより、利用者の問合せに対して知っている事項と知らない事項を分別することによって、利用者の理解度を考慮したWebページの検索手法を提案した。

今後は、利用者の理解度DBを構築する際に、利用者がプログラムソースを作成する際の編集履歴やコンパイラが出力したエラー履歴を利用することによって、より正確な理解度を算出できるよう改善を図る予定である。また、学習内容の問合せに対するWebページごとの信頼性を、blogを解析することで算出し[1]、プログラミング学習を行う上でより有効なWebページを検索できるよう改善を図る予定である。

参考文献

- [1] 中島伸介, 竹原幹人, 舘村純一, 日野洋一郎, 原良憲, 田中克己: “blog 解析に基づく web 情報検索の信頼性向上技術”, 人工知能学会第6回セマンティックウェブとオントロジー研究会 (2004).