

CodeConcierge:クラウド知識に基づいたプログラミング支援

高橋 裕太^{†1} 鵜林 尚靖^{†1} 細合 晋太郎^{†1}
亀井 靖高^{†1} 渡邊 卓也^{†2}

ソフトウェア開発において、コーディングに関するクラウド知識を利用する機会は往々にしてある。しかしながら、年々増加するクラウドデータの中から開発者が必要とする情報を得るのは難しい。本稿では開発者のクラウド知識の利用を支援するために我々が提案するツール CodeConcierge の概要を示す。このツールはプログラマが欲する情報をその利用コンテキストに応じて、複数のクラウド知識群を検索し、動的にパッケージ化して提示する。

1. はじめに

ソフトウェア開発において、コーディングの際に必要な情報をクラウド上の知識から取得する機会は往々にしてある。クラウド知識の量は年々増加しており、またその中には類似した知識も多く存在し、開発者が必要な情報を見つけることは難しい。Q&A サイトの StackOverflow(以下 SO)^{*1} を例に挙げると、2008 年 8 月から 2014 年 5 月までの総投稿数 21,736,594 件のうち内容が似通った投稿が 1,632,590 件あるという調査結果があった¹⁾。このことは多くの開発者が同じような問題に直面しているということと同時に、開発者が関心のある知識がクラウド上に飽和していて信頼のある情報を見つけることが困難であることを示している。このような問題の解決の為、クラウド知識の利用に関して多くの研究がされている。

本研究では、開発者のコンテキストを推測して、必要な情報をパッケージ化&ランキングして推薦するツール CodeConcierge を作成し、クラウド知識の利用による開発の支援を行いたいと考える。このツールは、キーワードやソースコードといった入力を元に、利用者のコンテキストを推測する。コンテキストとは、開発者が悩んでいるポイント(APIの使い方がわからない、効率の良いアルゴリズムが知りたい、バグの原因が知りたい)、開発者の状況(コーディング中、デバッグ中)、開発中の環境(使用中のプログラミング言語、OS、フレームワーク)などである。推測したコンテキストを考慮して複数のクラウドサービスを横断して検索を行い、取得した複数の種類の情報(APIドキュメント、優れたコード例、バグ修正例)をまとめ、必要と思われる順にランク付けをした上で推薦を行う。

2. クラウド知識とソフトウェア開発支援

クラウド知識を用いた開発支援の既存研究として、SOなどの単一のクラウドサービスの検索を行うものや、それをデバッグに役立てる研究¹⁾がある。本研究で提案するプログラミング支援ツールは「ユーザコンテキストを考慮した検索」「複数のクラウドデータベース(Q&Aサイト、バグ管理システム、コードリポジトリ、オンラインAPIドキュメント)を横断し検索」「結果はコンテキストに応じてパッケージ化&ランキングしたものを提示」という3つの観点を統合して開発支援を行う。この3つの観点は相互に関係するものであるため、これらを統合した環境による開発支援は有用であると考えられる。

3. CodeConcierge の機能

3.1 ツール概要

本研究では CodeConcierge というプログラミング支援ツールを提案する。このツールは利用者がいくつか入力を与えることで、そのコンテキストを推測し、必要な情報をパッケージ化&ランキングして提供する。

CodeConcierge の入力として、「検索キーワード」や「ソースコード」を与えることができ、これらを元にコンテキストの推測を行う。「検索キーワード」からは利用者の悩んでいるポイントや状況、「ソースコード」からは利用者の開発環境といったコンテキストを推測する。例えば、「検索キーワード」に API メソッド名が含まれる場合は利用者が API の使い方が分からず困っていると推測したり、「ソースコード」の特徴語を元に使用中の言語やフレームワークを推測したりする。また、入力には「利用者の状況」「プログラミング言語」「OS」「利用者のスキル」といった利用者のコンテキストの一部を明示的に与えることもできる。

3.2 利用シナリオ

実際にツールが利用されるシナリオとして、Android アプリケーション開発を行う例を考える。開発者が

^{†1} 九州大学

^{†2} エザリウム株式会社

*1 <http://stackoverflow.com/>

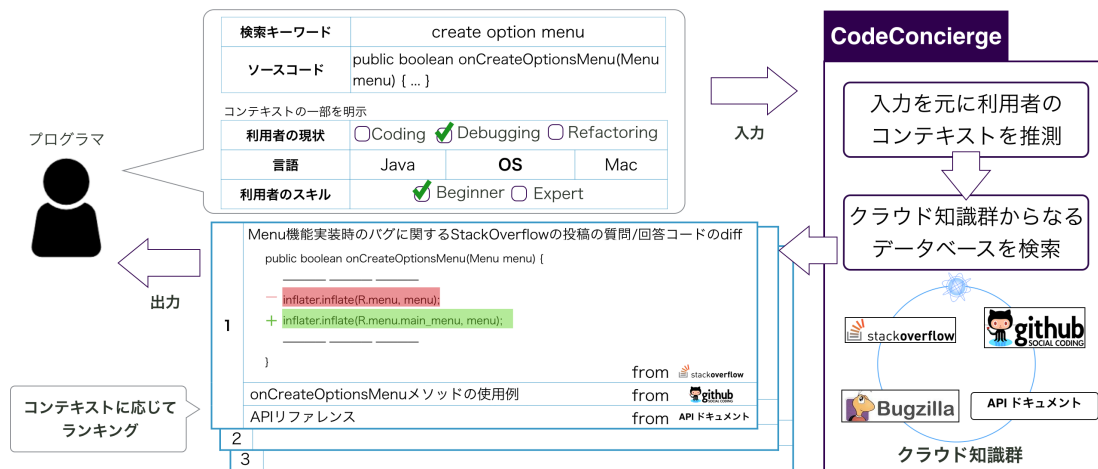


図 1 CodeConcierge 利用シナリオ

Menu 機能を実装するためコードを記述したが、意図した通りの Menu が実装されなかった。このようなとき、CodeConcierge に図 1 に示すような入力を与えた。出力として今回の例では、開発者が直面しているバグと似たようなバグに関する SO の投稿の質問コードと回答コード、使用したメソッドの OSS での使用例、API リファレンスといった情報をまとめたものが一番役に立つ情報のパッケージとして推薦される。

3.3 情報の検索, パッケージ化, ランキング

情報の検索, パッケージ化, ランキングは利用者のコンテキストを考慮して行う。考慮すべきコンテキストの例として「利用者の状況」が挙げられ、このコンテキストによって優先して検索すべきクラウドサービスやパッケージ化して同時に提供する情報が変わる。例えば、利用者がコーディング中と推測した場合は SO の投稿や github のソースコードの実例が参考になれる情報と思われるため、それらのクラウドサービスを検索する。デバッグ中と推測した場合は、デバッグに関する SO の投稿だけでなく、Bugzilla を検索し、得られたバグ修正事例を付随して提供する。これは SO の投稿のうち Debugging カテゴリの投稿は全体の 5% にも満たないという調査結果があり³⁾、デバッグ情報は SO の投稿だけでは不十分と思われる、追加の情報が必要と考えられるためである。

またランキングに影響するコンテキストとして、「利用者のレベル」が挙げられる。利用者が初心者だった場合、SO の投稿の中でも質問者がプログラミング初心者の投稿や、コードの可読性が高い投稿のランクを上げる、といったランキングの変更が考えられる。

4. 今後の課題

今後の課題として、コンテキストの推測アルゴリズムの向上や、本稿で提案した機能の実装が挙げられる。現状コンテキストは入力されたキーワードやソースコードの単語を元に推測するものであるが、検索の履歴から利用者が何をしようとしているかというコンテキストを推測するといった機能も実現したいと考えている。

謝辞 本研究は、文部科学省科学研究補助費 基盤研究 (A) (課題番号 26240007) による助成を受けた。

参考文献

- 1) Fuxiang Chen, Sunghun Kim, Crowd Debugging, In *Proceedings of the 10th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE '15)*, pp.320-332, 2015.
- 2) Fei Lv, Hongyu Zhang, Jian-guang Lou, Shaowei Wang, Dongmei Zhang, and Jianjun Zhao, CodeHow: Effective Code Search based on API Understanding and Extended Boolean Model, In *Proceedings of the 30th IEEE/ACM International Conference on Automated Software Engineering (ASE '15)*, pp.260-270, 2015.
- 3) Lucas B. L. de Souza, Eduardo C. Campos, and Marcelo de A. Maia, Ranking Crowd Knowledge to Assist Software Development, In *Proceedings of the 22nd International Conference on Program Comprehension (ICPC '14)*, pp.72-82, 2014.