

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي .08 ماي 1945, بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



MEMOIRE

En Vue de l'obtention du diplôme d'ingénieur d'état en informatique
Filière : Informatique
Spécialité : Système d'Information et Web (SIW)

Thème

**Model-Driven Prompt Engineering : Vers une Nouvelle
Approche pour la Génération Automatique de Code**

Présenté par :
Mr. ALILI Ilyes

Soumis en septembre 2024

Pr. Malki Mimoun
Dr. Awad Samir
Pr. BENSLIMANE Sidi Mohamed
Dr. NIANG Boubou Thiam
Dr. VERHAEGHE Benoit

Président
Examinateur
Encadreur
Encadreur
Encadreur

Année Universitaire : 2023/2024

Résumé

Dans un contexte où les systèmes logiciels deviennent de plus en plus complexes et interconnectés, la génération automatique de code rencontre des défis majeurs. Bien que des avancées significatives aient été réalisées, les approches actuelles peinent à capturer la logique métier spécifique et à générer du code robuste, adapté aux architectures complexes. Les solutions existantes se concentrent principalement sur la génération de code syntaxiquement correct, mais manquent souvent de la compréhension contextuelle nécessaire pour produire un code conforme aux exigences fonctionnelles et métier.

Face à ces limites, cette recherche propose une approche innovante pour la génération automatique de code, en exploitant des modèles de langage avancés et une ingénierie des prompts basée sur des modèles logiciels abstraits. L'objectif est de structurer les prompts de manière à ce qu'ils intègrent des informations contextuelles riches et détaillées, permettant ainsi aux modèles de langage de mieux appréhender le cadre technique et fonctionnel des systèmes. Cette méthode permet d'améliorer la qualité du code généré, tant en termes de précision que de pertinence par rapport aux spécifications métier.

Notre approche repose sur l'utilisation de modélisation logicielle via des méta-modèles et l'ingénierie inverse, offrant une représentation abstraite des logiciels qui guide la génération de prompts parfaits. Les expérimentations réalisées sur divers systèmes ont montré des améliorations significatives, validées par des métriques telles que CodeBLEU, CodeBERTScore, et l'évaluation fonctionnelle. Cependant, des défis subsistent dans les environnements complexes et privés, où l'intégration d'informations supplémentaires s'avère nécessaire pour capter pleinement la logique métier.

En conclusion, cette recherche propose une méthode innovante qui combine la modélisation logicielle et les modèles de langage, ouvrant de nouvelles perspectives pour améliorer l'efficacité de la génération de code dans des environnements variés. Elle pose les bases d'une optimisation future pour adapter cette approche aux besoins spécifiques des systèmes industriels complexes, tout en explorant de nouvelles formes d'enrichissement des prompts.

Summary

In a context where software systems are becoming increasingly complex and interconnected, automatic code generation faces major challenges. While significant advances have been made, current approaches struggle to capture specific business logic and generate robust code suited to complex architectures. Existing solutions focus mainly on generating syntactically correct code but often lack the contextual understanding necessary to produce code that meets functional and business requirements.

In response to these limitations, this research proposes an innovative approach to automatic code generation by leveraging advanced language models and prompt engineering based on abstract software models. The goal is to structure prompts in such a way that they incorporate rich, detailed contextual information, allowing language models to better grasp the technical and functional framework of systems. This method improves the quality of the generated code, both in terms of accuracy and relevance to business specifications.

Our approach relies on software modeling through meta-models and reverse engineering, providing an abstract representation of software that guides the generation of optimal prompts. Experiments conducted on various systems have shown significant improvements, validated by metrics such as CodeBLEU, CodeBERTScore, and functional evaluation. However, challenges remain in complex and private environments, where integrating additional information is necessary to fully capture business logic.

In conclusion, this research proposes an innovative method that combines software modeling and language models, opening up new perspectives for improving the efficiency of code generation in various environments. It lays the foundation for future optimization to adapt this approach to the specific needs of complex industrial systems, while exploring new ways of enriching prompts.

ملخص

في سياق تصبح فيه الأنظمة البرمجية أكثر تعقيداً وترابطاً، تواجه عملية توليد الكود التلقائي تحديات كبيرة. على الرغم من تحقيق تقدم كبير، إلا أن الأساليب الحالية تواجه صعوبة في استيعاب المنطق التجاري المحدد وتوليد كود قوي يتناسب مع البنى المعقدة. تركز الحلول الحالية في الغالب على توليد كود صحيح من الناحية التركيبية، لكنها غالباً ما تفتقر إلى الفهم السياقي الضروري لإنتاج كود يتوافق مع المتطلبات الوظيفية والتجارية.

استجابة لهذه التحديات، تقترح هذه الدراسة نهجاً مبتكراً لتوليد الكود التلقائي، بالاعتماد على نماذج لغوية متقدمة وهندسة التلميحات المستندة إلى نماذج برمجية مجردة. الهدف هو هيكلة التلميحات بطريقة تدمج معلومات سياقية غنية ومفصلة، مما يسمح للنماذج اللغوية بفهم الإطار الفني والوظيفي للأنظمة بشكل أفضل. يؤدي هذا الأسلوب إلى تحسين جودة الكود المنتج، سواء من حيث الدقة أو الملاءمة للمواصفات التجارية.

يعتمد نهجنا على النمذجة البرمجية من خلال النماذج الفوقية والهندسة العكسية، مما يوفر تمثيلاً مجرداً للبرمجيات يوجه عملية توليد التلميحات المثالية. أظهرت التجارب التي أُجريت على أنظمة متنوعة تحسينات كبيرة، تم التحقق منها باستخدام مقاييس مثل CodeBLEU و CodeBERTScore والتقييم الوظيفي. ومع ذلك، لا تزال التحديات قائمة في البيئات المعقدة والخاصة، حيث يصبح دمج معلومات إضافية ضرورياً لفهم المنطق التجاري بالكامل.

في الختام، تقترح هذه الدراسة طريقة مبتكرة تجمع بين نمذجة البرمجيات والنماذج اللغوية، مما يفتح آفاقاً جديدة لتحسين كفاءة توليد الكود في بيئات متنوعة. كما تمهد هذه الطريقة الطريق للتحسينات المستقبلية لتكثيف هذا النهج مع احتياجات الأنظمة الصناعية المعقدة، مع استكشاف أشكال جديدة من إثراء التلميحات.