In modern software development, software modeling is considered to be an essential part of the software architecture and design activities. The Unified Modeling Language (UML) has become the de facto standard for software modeling in industry. Surprisingly, there is only little empirical evidence on the practices and impacts of UML modeling in software development.

This PhD thesis contributes to this matter by describing a method to build and curate a big corpus of open-source-software (OSS) projects that contain UML models. Subsequently, this thesis offers observations on the practices and impacts of using UML in OSS projects.

We combine techniques from repository mining and image classification in order to successfully identify more than 24.000 open source projects on GitHub that together contain more than 93.000 UML models. Various empirical studies have been carried out across this set of projects.

The results show that UML is generally perceived to be helpful to new contributors. The most important motivation for using UML seems to be to facilitate collaboration within teams. Our study also shows that the use of UML correlates with lower defect proneness. Further, we find out that vizualisation of design concepts, such as class role-stereotypes, helps developers perform better in software comprehension tasks.

We hope researchers in the field will find our corpus and the findings from this thesis a valuable source for further empirical studies.

Truong Ho Quang
Department of Computer Science and Engineering
Division of Software Engineering

ISBN 978-91-7833-608-1

DOCTORAL THESIS

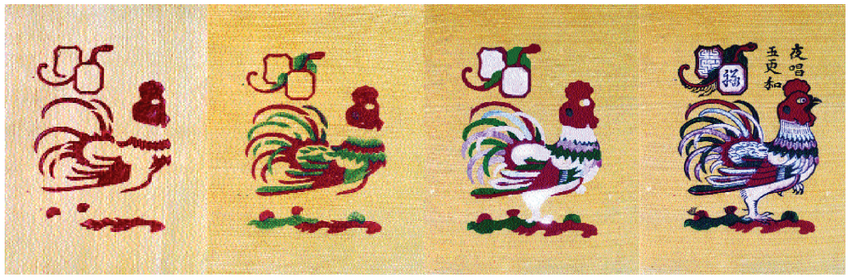
Truong Ho Quang Empowering Empirical Research in Software Design: Construction and Studies on a Large-Scale Corpus of UML Models 2019

DOCTORAL THESIS
IT FACULTY

Empowering Empirical Research in Software Design

Construction and Studies on a Large-Scale Corpus of UML Models

Truong Ho Quang



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIVERSITY OF GOTHENBURG