Novices face much difficulty in comprehending even small computer programs. A framework is needed to help novices’ attain sufficient program comprehension skills. To develop such framework, this research re-examines the three main factors that influence the novices’ program comprehension; the programmer, the program code, and the task. Most studies so far focused on the first two especially programmer, with tasks traditionally applied only as a measure for program comprehension. Very few studies have inverted this supplementary role of task and instead examined the task factor itself. There is a research gap on the effects of tasks on novices’ program comprehension. Moreover, current program comprehension mental models (PCMMs) have been mostly constructed for expert programmers, which do not match the novices’ mental model. To solve these two problems, the research proposes a task-based program comprehension framework for novices. Employing empirical software engineering research design approach, 1) fourteen effective programming tasks for novices have been identified, 2) a novices’ PCMM has been constructed, and 3) the effects of the tasks on the PCMM have been investigated. Consequently, the tasks were classified according to the cognitive domain of Revised Bloom’s Taxonomy. The tasks in each cognitive category were then ranked based on their effectiveness on novices’ program comprehension using a survey of instructors. Subsequently, novices’ PCMM with four core abstraction levels, i.e. Statement, Block, Module, and Program, and an extended level, i.e. Domain, was developed, and validated by experts. The effects of eight tasks on the novices’ PCMM were investigated through an experiment involving 69 novices, and six tasks were validated through another experiment involving 178 novices in three universities. Both experiments also validated the ranking of these tasks on the novices’ PCMM. The first experiment demonstrated that all the tested tasks were able to significantly improve novices’ PCMM. The first and second ranked tasks were studied more closely, where the first ranked tasks consistently showed higher improvements than the second ranked tasks in each of the four tested Bloom’s categories of Remember, Understand, Analyze, and Create. The second experiment demonstrated that different tasks improved the abstraction levels and the information categories differently. It also indicated that higher cognitive category tasks improve PCMM at higher abstraction levels. The general implication is that the framework can be an effective tool for computing educators to incorporate program comprehension in programming, and thus shift from merely teaching tracing and debugging tasks only. These tasks need to be introduced in stages in the teaching of programming, starting initially from the lower cognitive categories’ tasks such as Recall and culminating at the higher cognitive categories’ tasks such as Modification in possibly team project assessments. However, these tasks should be applied with taking the consideration of the novices’ programming levels and the information categories need to be improved. The key contribution of this thesis is a new developed framework, which includes novices’ PCMM, and a set of classified and ranked effective tasks that can improve novices’ PCMM.

The Structured Document Classification System (SDCS) is an industrial-driven technology that has the ability to classify piles of structured documents collected everyday efficiently in different places. Although the SDCS technology has advanced tremendously, one of the most challenging tasks is to propose a classifier that supports various layouts for different categories and different script languages in a high accuracy and efficient time. To solve the issue of supporting various layouts for different categories and different script languages, a Robust Generic Structured Document Classifier has been proposed (RGSDC). RGSDS starts with finding the best objects that can be used to fit the target and solve the issue. Detailed study for all the previous thresholding techniques is conducted to introduce a new categorization method based on the transformation value of input images. This study is a good base for finding reliable thresholding algorithm. A new thresholding technique based on ordinal structure fuzzy logic (OSFM) is proposed to provide a robust generic image thresholding technique (RGT) that is able to extract clear mixed predefined objects for different languages and multi layouts problems. Two different set of features that distinguish different languages and multi layouts structured documents are proposed. Lines-based features are totally relying on dimensions, locations, and slanting properties. On the other hand, blob-based features represent the shape, appearance, and distribution of the nominated objects. All the collected features are used to build a statistical feature vector for the classification stage. Based upon the need of a fast and accurate process to propose a practical structured document classification, a creative and fast skewing process based on nominated referencing lines out of the group of lines in the object selection process is created as the first stage. An algorithm focused on evaluating performance of different pattern classification techniques namely; neural network, support vector machine, Pearson correlation, and dynamic time wrapping (DTW) is used. At the end, computational calculations to prove the accuracy of the proposed algorithms are performed in four different stages. The results of experiments demonstrate that the proposed RGSDC is capable of performing classification of 3.5 forms per second with a 4.9% misclassification error rate (ME). Reliability of the research is verified by benchmarking the results of RGSDS with another well-known SDCS developed by Austrian banking solution company (xyzmo).