

**МОДЕЛЬ ДАННЫХ ДЛЯ ПРЕДСТАВЛЕНИЯ LLVM IR
НА ОСНОВЕ БИТОВОГО ПОТОКА**

Пименов Д.Д. (Университет ИТМО),

Кореньков Ю.Д. (Университет ИТМО),

Научный руководитель – кандидат технических наук Кореньков Ю.Д.

(Университет ИТМО)

Множество полезных библиотек и инструментов реализовано в неуправляемом коде, вследствие чего им свойственны его главные недостатки – платформенно-ориентированность и относительная небезопасность. Эти недостатки приводят к необходимости портирования подобного программного обеспечения на нужные целевые платформы и разработки дополнительных слоев безопасности, что затрудняет, а иногда и вовсе делает невозможными поддержку и использование такого ПО на некоторых платформах. Одним из вариантов решения проблемы является перенос такого ПО в управляемый код.

Для избежания необходимости повторно реализовывать ПО в управляемом коде необходимо преобразовывать LLVM IR модули, сгенерированные из исходного кода ПО, во внутреннее представление среды исполнения управляемого кода. За счёт данного преобразования обеспечивается возможность совместного использования среды исполнения управляемого кода и инструментов, реализованных в рамках LLVM. Для реализации подобного преобразования необходимо сформировать модель данных, описывающую битовый поток LLVM IR модулей.

Целью работы является разработка модели данных для представления LLVM IR на основе битового потока.

В основе решения лежит 3 слоя представления. Первый слой представляет из себя декларативную схему данных. Этот слой описывает формат всех блоков и записей, которые могут содержаться в LLVM IR модулях [1][2]. Второй слой содержит в себе набор метаданных, создаваемый на основании схемы данных. Данный слой существенно упрощает задачу сериализации LLVM IR бит-кода, так как содержит в себе достаточный набор информации, необходимых для корректной интерпретации обрабатываемых данных. Третий слой предоставляет набор структур данных, сгенерированных на основе схемы данных и снабженных необходимыми метаданными. В результате сериализации LLVM IR получается иерархически организованный набор экземпляров сгенерированных структур данных, соответствующий исходному битовому потоку LLVM IR.

Для построения структур данных из декларативной схемы, в рамках данной работы был реализован генератор. Помимо этого, был реализован сериализатор, преобразующий битовый поток LLVM IR в набор сгенерированных структур данных.

В результате работы была разработана модель данных, описывающая битовый поток LLVM IR модулей и генератор, строящий структуры данных на основе декларативной схемы данных потока. На основе созданной модели данных был разработан сериализатор, преобразующий битовый поток LLVM IR в набор сгенерированных структур данных.

Список использованных источников:

1. Документация LLVM Bitcode – Электронный ресурс URL: <https://llvm.org/docs/BitCodeFormat.html#define-abbrev>

2. Исходный код LLVM – Электронный ресурс URL: <https://github.com/llvm/llvm-project>