

УДК 004.056

Версионирование спецификаций из внешнего источника в системах управления конфигурацией

Шпиталев Даниил Олегович

Кузьмина Инна Анатольевна (*научный руководитель*)

МГТУ им. Н.Э. Баумана, Москва, Россия

Рассмотрена разработка модуля-микросервиса для системы управления конфигурацией для восстановления утраченных связей и метаданных при импорте спецификаций из внешних источников (сторонних систем управления конфигурацией) на основе анализа текстового содержимого базовой и импортируемой версий спецификаций. Рассмотрены такие алгоритмы сравнения текстов, как кодировщики предложений, расстояние Левенштейна, мешок слов и др.), что позволяет автоматически переносить связи и атрибуты между версиями, сокращая ручную работу. Решение устраняет проблему потери данных при экспорте спецификаций в системах управления конфигурацией, так как экспорт обычно происходит только в человекочитаемый формат.

Ключевые слова: *система управления конфигурацией, импорт спецификаций, автоматизация переноса связей, алгоритмы нечеткого сравнения, большие языковые модели*

Введение. Разработка современных сложных систем требует управления множеством разнородных компонентов: программных, кибернетических, механических и электронных. Их сложные взаимосвязи друг с другом и взаимодействия с внешней средой создают значительные риски: несогласованность, непредвиденные ошибки интеграции и резкий рост затрат на их исправление. Например, согласно исследованию Standish Group, в среднем, около 16,2 % проектов в сфере информационных технологий являются успешными и не сталкиваются с перерасходом бюджета или нарушением сроков, при этом основными причинами являются недооценка сложности, плохое планирование, выраженное в плохо составленных спецификациях требований [1].

Системный подход становится ключевым инструментом минимизации этих рисков. Он обеспечивает целостность системы через декомпозицию и анализ взаимодействий подсистем, согласованность компонентов за счет формального описания интерфейсов и требований, предсказуемость разработки благодаря моделированию поведения системы на ранних этапах [2].

Системы управления конфигурацией (Configuration Management Systems, СМ-системы) созданы для ведения разработки сложных систем в рамках системного подхода. Конфигурация системы — совокупность компонентов, организованных для выполнения заданных функций. Управление конфигурацией — дисциплина идентификации конфигурации системы в определенные (заданные) моменты времени, с целью систематического контроля изменений

конфигурации, а также поддержки и сопровождения целостной и отслеживаемой (трассируемой) конфигурации на протяжении всего жизненного цикла системы [3].

Трассируемость — это возможность однозначно отслеживать взаимосвязи между объектами конфигурации. В различных стандартах, регламентирующих разработку ответственных систем в авиационной отрасли требуется наличие трассируемости [4–6]. Она является фундаментальной основой для проведения анализа влияния при изменении требований, помогая предсказывать эффект от внесения таких изменений [7].

Контроль изменений предполагает их версионирование. Существует понятие базовой версии (baseline) — формально зафиксированная версия конфигурации системы или ее компонентов, утвержденная для дальнейшего использования и служащая точкой отсчета для изменений. Обновленная версия (revision) — версия, предназначенная для замены старой. Вариант (variant) — версия, добавляемая в конфигурацию без замены старой. Также версия может быть рассмотрена в качестве определенного состояния конфигурации развивающейся системы.

Импорт спецификаций из внешнего источника. В некоторых случаях может возникнуть необходимость импорта данных из сторонних систем управления конфигурацией:

- в случае миграции из одной системы управления конфигурацией в другую, по причине того, что исходная система перестала поддерживаться или появилась новая система с более обширным функционалом и удобством его использования;

- в случае ведения совместной разработки несколькими компаниями, привыкшими использовать разные системы управления конфигурацией.

Существует несколько проблем в задаче реализации механизма импорта из внешнего источника. Во-первых, в исходной системе скорее всего уже хранится большое количество наработок по всей проектной деятельности, и необходимо перенести всю конфигурацию проекта из одной системы в другую. Конфигурация может состоять из разнородных артефактов разработки: спецификаций с требованиями, программного кода, файлов из САД-систем, например, различных 3D-моделей, файлов из САМ систем, специфичных схем, диаграмм. Проектов, требуемых переноса, может быть несколько, и задача импорта всей конфигурации за один шаг является трудноосуществимой. Во-вторых, экспорт из систем управления конфигурацией обычно происходит в специфичном формате. В случае экспорта спецификаций — это человекочитаемый формат, в котором теряются идентификаторы, метаданные и связи между артефактами (требованиями, различными определениями, блоками кода, описаниями тестов и т. п.).

В данном исследовании предполагается решить задачу импорта спецификаций из сторонних СМ-систем и их последующего версионирования на основе анализа текстового содержимого базовой и импортируемой версий спецификаций.

Существующие решения. Подобные решения для импорта спецификаций из внешнего источника не слишком распространены, ввиду того что разработчики систем управления конфигурацией, как и любые другие разработчики, не заинтересованы в предоставлении инструментов для возможности миграции с их системы на другую. Однако существует стандарт обмена требованиями ReqIF (Requirements Interchange Format).

ReqIF — формат файлов, предназначенный для обмена требованиями и связанными метаданными между системами управления конфигурацией, основанный на xml. ReqIF — открытый стандарт, поддерживаемый Object Management Group (OMG). ReqIF имеет древовидную структуру. Корневой элемент содержит метаданные о требованиях ReqIFHeader (время создания, версия ReqIF, идентификаторы: самого экспортируемого документа, репозитория и инструмента управления требованиями, из которого экспортируются требования), основное содержание, то есть сами требования, ReqIFContent, и опционально специфичные для конкретного инструмента управления требованиями расширения ReqIFToolExtension, нужные для обмена специфичными артефактами, которые не поддерживаются ReqIF по умолчанию. Основное содержание (ReqIFContent) состоит из SpecObjects, которые представляют требования на основе их атрибутов; SpecificationS — иерархической структуры, ссылающейся на SpecObjects; SpecRelations, содержащих связи между требованиями; SpecType — коллекции AttributeDefinitions, которая определяет какие значения атрибутов может иметь объект. SpecObjects, specifications, и SpecRelations также могут иметь SpecTypes, и соответственно атрибуты [8].

ReqIF поддерживают такие системы, как DOORS Next, Visure Solutions ALM, Siemens Polarion, но далеко не все системы управления конфигурацией. Недостатком ReqIF является то, что, хотя стандарт изначально разрабатывался как универсальный, обмен данными между системами разных вендоров может быть затруднен из-за разной реализации стандарта в некоторых системах [9]. Кроме того, обмен артефактами со сложными метаданными не работает автоматически.

Предложенный механизм импорта. Исходя из перечисленных проблем реализации и ограничений, предлагается следующий механизм, общими этапами которого являются вне зависимости от формата, в который экспортируются спецификации (ReqIF, проприетарный или человекочитаемый формат):

- предобработка (парсинг) исходного экспортированного из другой СМ-системы файла, включающая декомпозицию спецификации на структурные фрагменты, приведение к формату, в котором спецификации хранятся в целевой для импорта системе;
- сопоставление фрагментов базовой версии спецификации и импортируемой предобработанной спецификации, полученной на предыдущем этапе, на основе анализа схожести текстового содержимого фрагментов спецификаций;
- на основании сопоставления фрагментов, перенос идентификаторов, связей и метаданных из базовой версии для неизменных (почти неиз-

менных) фрагментов, удаление и добавление новых фрагментов, приведение редакции к импортируемой версии.

Преимуществами предложенного механизма импорта являются: автоматический перенос связей для всех версий, кроме базовой, определение конкретных изменений между версиями и возможность их журналирования. Недостатком является необходимость ручного выставления связей в базовой версии.

Рассматриваемые алгоритмы сравнения текстовых данных. Для определения схожести текстовых данных рассматривались алгоритмы, основанные на больших языковых моделях (кодировщики предложений на архитектуре BERT) и алгоритмы нечеткого сравнения, такие как расстояние Левенштейна, нахождение наибольшей общей подпоследовательности, сходство Жаккара, мешок слов.

В результате прототипирования был сделан выбор в пользу алгоритмов нечеткого сравнения по следующим причинам:

- большие языковые модели работают на порядок дольше более традиционных алгоритмов (при вычислениях на центральном процессоре);
- большие языковые модели требуют дообучения для реального использования, так как в реальных спецификациях много специализированных терминов и аббревиатур;
- большие языковые модели, хотя и способны сопоставить более сильно видоизмененные фрагменты спецификаций разных версий, но для таких сильно видоизмененных фрагментов не всегда корректно автоматически переносить связи и всю метайнформацию.

После прототипирования был реализован модуль для импорта из внешнего источника и версионирования. Он был интегрирован в микросервисную архитектуру системы управления конфигурацией АИС-Т. Алгоритм сравнения текстовых данных выбирается динамически. Основным алгоритм — мешок слов. Для фрагментов спецификаций, для которых некорректно применять мешок слов, используются другие алгоритмы. Например, в спецификациях могут быть формулы и листинги с кодом, сопоставление таких фрагментов выполняется с помощью расстояния Левенштейна с символьной токенизацией. Данные о том, чем являются сопоставляемые фрагменты, получаются на этапе парсинга экспортированного из сторонней системы файла.

Заключение. Таким образом, был реализован модуль импорта спецификаций из внешнего источника и версионирования в системах управления конфигурацией. Он был протестирован и успешно переносит связи из базовой версии в импортируемую, основываясь на схожести некоторых фрагментов двух версий спецификаций. Были рассмотрены и протестированы в рамках прототипирования некоторые методы сравнения текстовых данных. Был выбран подход, основанный на алгоритмах нечеткого сравнения. Основным алгоритмом стал мешок слов. Преимуществами представленного решения являются автоматический перенос связей для всех версий, кроме базовой, и возможность журналирования конкретных изменений между версиями. Не-

достатком является необходимость ручного выставления связей в базовой версии. Данный недостаток продиктован сложностью реализации импорта всей конфигурации за один шаг.

Литература

- [1] Chaos Report. URL: <https://artemushanov.ru/?go=all/otchet-chaos-report-2020/> (accessed 14.11.2024).
- [2] Guide to the Systems Engineering Body of Knowledge (SEBoK). Version 2.11. URL: https://sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_%28SEBoK%29 (accessed 21.05.2025).
- [3] IEEE 828–2012. IEEE Standard for Configuration Management in Systems and Software Engineering. IEEE, 2012, 86 p.
- [4] *Квалификационные требования КТ-178С. Требования к программному обеспечению бортовой аппаратуры и систем при сертификации авиационной техники.* Москва, АР МАК, 2016, 131 с.
- [5] *Квалификационные требования КТ-254. Руководство по гарантии конструирования бортовой электронной аппаратуры.* Москва, АР МАК, 2011, 89 с.
- [6] *Руководство Р-4754А по разработке воздушных судов гражданской авиации и систем.* Москва, АР МАК, 2016, 131 с.
- [7] Guide to the Software Engineering Body of Knowledge (SWEBoK). Version 4.0. URL: <https://www.computer.org/education/bodies-of-knowledge/software-engineering> (accessed 21.05.2025).
- [8] Requirements Interchange Format (ReqIF) Specification version 1.2. URL: <https://www.omg.org/spec/ReqIF/About-ReqIF/> (accessed 21.05.2021).
- [9] Горелиц Н.К., Кильдишев Д.С., Хорошилов А.В. Управление требованиями к ответственным системам. Обзор решений. *Труды ИСП РАН*, 2019, т. 31, вып. 1, с. 25–48. [https://doi.org/10.15514/ISPRAS-2019-31\(1\)-2](https://doi.org/10.15514/ISPRAS-2019-31(1)-2)

Versioning of requirements specifications from third-party configuration management systems

Shpitalev Daniil Olegovich

Kuzmina Inna Anatolievna (*scientific advisor*)

BMSTU, Moscow, Russia

The article discusses the development of a microservice module for a configuration management system to restore lost links and metadata when importing specifications from external sources (third-party configuration management systems) based on an analysis of the text content of the baselines and imported versions of specifications. Some text comparison algorithms are being considered: sentence encoders, Levenshtein distance, bag of words, etc.), which allows you to automatically transfer links and attributes between versions, reducing manual work. The solution eliminates the problem of data loss when exporting specifications in CM systems, as exports usually occur only in human-readable format.

Keywords: configuration management system, import of specifications, automation of link transfer, fuzzy comparison algorithms, large language models