

Sven Wenzel:

Unique Identification of Elements in Evolving Models: Towards Fine-Grained Traceability in Model-Driven Engineering

Promotion: Universität Siegen, Naturwissenschaftlich-Technische Fakultät

Erstgutachter: Prof. Dr. U. Kelter, Universität Siegen

Zweitgutachter: Prof. T. Systä, PhD, Techn. Universität Tampere, Finnland

Drittgutachter: Prof. Dr. J. Ebert, Universität Koblenz

Datum der Prüfung: 4. März 2011

Veröffentlichung: <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2011/524/>

Kurzfassung:

Model-driven engineering (MDE) is a widely accepted methodology in software engineering. At the same time, the ability to retrace the engineering process is an important success factor for software projects. In MDE, however, such traceability is often impeded by the inadequate management of model evolution. Although models have a very fine-grained structure, their different revisions and variants are prevalently managed as monoliths in a file-based software configuration management (SCM). This causes the identification problem: if the fine-grained elements are not assigned with globally unique identifiers, we cannot identify them over time. If such identifiers would be given, they can be misleading. As a consequence, we cannot comprehend the evolution of elements and traceability relationships among the elements cannot be managed sufficiently.

This thesis presents a novel solution to the identification problem. It establishes a representation to describe the history of a model and its fine-grained elements inside. The key feature of the representation is a new kind of traceability relationship, called *identification links*. They allow us to identify elements of a given revision in other revisions or variants of the model. The identification is even applicable to anonymous elements and model fragments. It provides us with a broad spectrum of opportunities: e.g. management of fine-grained traceability links, evolution analysis, merging of development branches. Due to the expression of model evolution in the history representation, we are further able to capture the changes that have been applied to the traced elements. This thesis further presents an algorithm to infer the identification links automatically. The approach does not rely on persistent identifiers, but it utilizes a similarity-based model comparison technique to locate the model elements in other revisions.

The algorithm and the history representation have been implemented in a prototype. It is metamodel and tool independent and can work with an arbitrary SCM. Existing modeling environments do not have to

be modified. Traceability information and evolution information is accessible through a service interface and can thus be integrated in arbitrary tools. The evaluation of our approach by means of controlled experiments with data from real models attested excellent precision and recall values for the identification of model elements over time. Furthermore, different evolution analysis tools have already been built on our approach, which documents the practical applicability of our solution.

deutsch:

Die modellgetriebene Entwicklung ist eine weit verbreitete Methode zur Softwareentwicklung. Die ungeeignete Versionierung von Modellen verhindert jedoch oftmals die Nachverfolgbarkeit des Entwicklungsprozesses. Trotz ihrer feinkörnigen Struktur werden Modelle oft monolithisch auf Basis dateibasierter Konfigurationsmanagementsysteme verwaltet. Modellelemente können in diesem Fall nicht über die Zeit hinweg identifiziert werden, weil globale Identifizierer entweder fehlen oder irreführend sein können. Aufgrund dieses Identifikationsproblems können Änderungen an Modellelementen nur sehr schwer nachvollzogen werden. Außerdem können Referenzen zur Nachverfolgbarkeit zwischen Elementen verschiedener Revisionen nicht sinnvoll verwaltet werden.

Diese Dissertation löst das Identifikationsproblem. Sie führt eine Repräsentation ein, mit der die Historie von Modellen und deren feinkörnigen Elementen abgebildet werden kann. Ein zentraler Bestandteil dieser Repräsentation sind *Identifizierungslinks*, die es ermöglichen ein gegebenes Element in anderen Revisionen oder Varianten des Modells wiederzufinden. Der Ansatz unterstützt auch anonyme Elemente und komplette Modellfragmente. Diese neuartige Identifizierung ermöglicht z.B. die Verwaltung von feinkörniger Nachverfolgbarkeitsinformation, die Analyse von Modellevolution oder das Mischen von Entwicklungszweigen. Da die Repräsentation auch die Evolution eines Modells abbilden kann, können die Veränderungen identifizierter Elemente besser erfasst werden. Zudem wird in dieser Dissertation ein Algorithmus entwickelt, mit dem Identifizierungslinks zwischen Modellelementen verschiedener Revisionen inferiert werden. Dieser stützt sich nicht auf persistente Identifizierer, sondern nutzt einen ähnlichkeitsbasierten Differenzalgorithmus, um Elemente in anderen Revisionen wiederzufinden.

Der Algorithmus und die Historienrepräsentation wurden in einem modelltyp- und werkzeugunabhängigen Prototyp implementiert, der mit beliebigen Konfigurationsmanagementsystemen zusammenarbeitet, ohne dass diese angepasst werden müssen. Die Informationen zur Nachverfolgbarkeit und Evolution von Modellelementen sind über eine Programmierschnittstelle abfragbar, die sich in beliebigen

Werkzeugen nutzen lässt. Der beschriebene Ansatz wurde mit kontrollierten Experimenten auf Basis realer Modellhistorien erfolgreich evaluiert. Darüber hinaus wurde seine praktische Anwendbarkeit durch verschiedene darauf aufbauende Werkzeuge zur Evolutionsanalyse belegt.