Praxis der Forschung – Wintersemester 2015/16

Teilnehmende Lehrstühle im WiSe 2015/16

- ITI Prof. Beckert, Arbeitsgruppe für Anwendungsorientierte formale Verifikation
- TM Prof. Beigl, Lehrstuhl für Pervasive Computing Systems / TECO
- IPD Prof. Reussner / Jun.-Prof. Koziolek, Software Design and Quality (SDQ)
- IAR Prof. Asfour, Arbeitsgruppe für Hochperfomante Humanoide Technologien (H^2T)
- ITI Prof. Müller-Quade, Arbeitsgruppe Kryptoraphie und Sicherheit

Kontakt bei allgemeinen Fragen zu „Praxis der Forschung“:

- Sarah Grebing, ITI Prof. Beckert, sarah.grebing@kit.edu, +49 721 608-45253
- Matthias Budde, TM Prof. Beigl, matthias.budde@kit.edu, +49 721 608-41705

Ausgeschriebene Themen im WiSe 2015/16

Automatisierte Unterstützung von dynamischen Updates in Java

Durch Software können Systeme leicht an neue Anforderungen angepasst werden, beispielsweise um Funktionalität zu erweitern oder Sicherheitslücken zu schließen. Dynamische Software Updates (DSU) bieten eine möglichst unterbrechungsfreie Aktualisierung von Software im laufenden Betrieb. Aktuelle Systeme in der Forschung sind beispielsweise *Kitsune* für C und *Rubah* für Java bzw. das an unserem Lehrstuhl entwickelte System. Diese DSU-Systeme automatisieren bereits viele Schritte, die zur Installation eines dynamischen Updates nötig sind, um den gesamten Hauptspeicherzustand eines Programmes zu transformieren. Einige Transformationen, wie komplexere Umstrukturierungen der Datenstrukturen im Hauptspeicher können jedoch nicht automatisiert umgesetzt werden. Auch einfachere Fälle, wie die Umbenennung von Variablen, können aufgrund der fehlenden Eindeutigkeit nicht automatisiert sicher entschieden werden. Es verbleibt die Aufgabe des Programmierers hier geeignete Transformationsfunktionen zu entwickeln, die ein dynamisches Update vervollständigen.

Problematisch ist, dass prinzipiell diese Transformer update-spezifisch sind, und somit ein individueller Transformer für jedes Update von Version A auf Version B nötig ist, sei eine Veränderung auch noch so gering, z.B. vX.Y.1 auf vX.Y.2. In dieser Gruppenarbeit (2-3 Personen) soll ein Annotationsrahmenwerk für Java auf Basis unseres DSU-Systems entworfen und umgesetzt werden mittels welchem Programmierer die Bedeutung von Veränderungen in ihrem Code klar spezifizieren können. Die Annotationen setzen dadurch aktuellen Code in eindeutige Verbindung mit dessen direkter Vorgängerversion. Diese Annotationen werden vom DSU-System genutzt, um den dynamischen Transformationsprozess soweit wie möglich zu automatisieren.

Der Entwicklungsaufwand bisheriger Transformationsfunktionen soll abschließend im Vergleich zum entworfenen Annotationsrahmenwerk empirisch bewertet werden.

Kontakt / Betreuer: Martin Alexander Neumann (TM Beigl)
martin.neumann@kit.edu

Exploring Context-Recognition with Dust Sensors

Kontext- und Aktivitätserkennung mit günstigen Sensoren, vor allem mit 3D-Beschleunigungs-sensorik, ist gut erforscht. Der Fokus liegt dabei oft auf physischen Aktivitäten (Gehen, Rennen, etc.). „Exotische“ Sensoren werden jedoch noch weniger betrachtet. Vergangene Arbeiten haben beispielsweise gezeigt, dass günstige Staubsensoren zur Feinstaubmessung eingesetzt werden können. In dieser Arbeit soll ihre Eignung, bzw. die von Feinstaub als Messgröße, zur Kontext- und Aktivitätserkennung untersucht werden. Forschungsfragen:

- Inwiefern kann Feinstaubsensorik für Kontext-Erkennung eingesetzt werden?
- Welche Features sind dazu geeignet?
- Welche Kontexte / Umgebungen / Aktivitäten können erkannt bzw. unterschieden werden?
- Kann Kontext-Information zur Erhöhung der Messgenauigkeit bei Feinstaubmessungen beitragen?

Grundkenntnisse / Erfahrung mit Machine Learning sollten mitgebracht werden. Eigene Ideen sind willkommen und können gerne berücksichtigt werden.

Kontakt / Betreuer: Matthias Budde (TM Beigl)
matthias.budde@kit.edu

Developing a Framework for Programmer-Centered API Design

Nowadays, software development focusses on the needs of the end-user: A new software should be usable in an easy and intuitive way without making errors. However, less effort has been invested to the people who actually create and design this software, namely the programmer. Programmers are humans as well, with different needs and preferences that make their work with different APIs hard or easy. In the last years some theoretical approaches have been developed to assess e.g. the usability of APIs but it is unclear how to map programmer preferences onto different aspects of the usability of an API. The goal of this “Praxis der Forschung” project is to develop a framework that allows a programmer-centered API design. The work in this project may include the development of a standardized questionnaire to measure programmer preferences in a reliable and valid way, to derive guidelines for a programmer-centered design, to develop a tool that allows the mapping of preferences onto different design aspects of the API and to validate the results in a summative study.

Kontakt / Betreuer: Dr. Till Riedel (TM Beigl) till.riedel @kit.edu

PD Dr. Andrea Schankin (TM Beigl)
andrea.schankin@kit.edu

Cross-Platform Mobile Sensing and Activity Recognition ("Bringing Activity Recognition Classifiers into Web Browsers via JavaScript")

There are many smartphone apps on the market that track a user’s activity, eating behavior, fitness, sleep and so on. Though, most of these apps work for one operating system and require the developer to have programming knowledge. Moreover, they do not offer their recognition services and results to other apps.

What we want to achieve is a cross-platform mobile activity recognition platform which enables a crowdsourcing of training data and a delivery of activity classifiers to a wide range of web and cross-platform applications. We will build up upon existing TECO systems such *ActiServ* and *jActivity* and leverage their crowdsourcing and sensing capabilities.

You will implement a way to deliver activity recognition classifiers in a unique format to several applications. Hence, you will review available classifier formats and define a suitable, common output format such as JavaScript or JSON.

The converter will be tested with at least three different input classifier types. Moreover, the generated classifier will be exemplary implemented into an

example web application. This might be a website that adjusts its font size depending on the user activity. In a next step, user preferences such as “preferred walking font size” might be collected and stored locally on the device.

Among others, tasks within this topic are reviewing common classifier systems and formats, implementation of a classifier converter and recording real world training and test data.

Kontakt / Betreuerin: Anja Bachmann (TM Beigl)
anja.bachmann@kit.edu

RF signal based context recognition

We are surrounded by Radio Frequency (RF) signals every day: WiFi signal, cellular signal, radio signal, GPS signal, etc. However, the capability of RF signals has not been fully explored. RF signal based context recognition has a series of advantages.

1) In field infrastructure deployed. Today our daily life has been full of various kinds of wireless devices, the most popular ones are WiFi routers and mobile phones. There are huge opportunities for us to leverage these wireless devices with considerable storage and computation capability to realize sophisticated context recognition.

2) Less privacy concern. Compared with video camera based methods, RF signal contains far more less sensitive information. People are more likely to accept RF signal based context recognition technologies.

3) Capability to “see through walls”. Since electromagnetic wave can penetrate through obstacles, it provide us possibility to fulfill context recognition beyond obstacles.

RF signal based context recognition has been a hot topic recently. Activity recognition, indoor localization, counting, state identification are all possible topics for context recognition. For RF signal, there are also various dimensions we can explore: RSSI (Received Signal Strength Indication), CSI (Channel State Information), Doppler effect and so on.

Kontakt / Betreuer: Long Wang (TM Beigl)
wanglong@teco.edu

NOTE: As the advisor is not fluent in German, all communication and all exams, both written and oral, will be in English.

Ein prädiktives Regelungskonzept für HVAC Systeme zur Erhöhung des Nutzerkomforts

Ein modernes HVAC System ist ein cyber-physisches System in der Büroumgebung. Bestehende Steuerungsansätze für HVAC Systeme sollen hinsichtlich Nutzerkomforts optimiert werden. Einfach konstruierte Automatisierung der HVAC-Steuerung nach Zeit und/oder Temperatur befindet sich bereits als Stand der Technik im Einsatz. Das Ziel des Projektes ist es ein AI-basiert lernendes Regelungssystem zu entwickeln, damit sich das HVAC cyber-physische System prädiktiv orientiert nach Nutzerkomfort regeln lässt.

In diesem Vorhaben beabsichtigen wir neben der Systemidentifikation der Temperaturentwicklung in einer Büroumgebung ein prädiktives Regelungskonzept zu entwickeln, um Raumtemperatur vorausschauend nach Energieeffizienz und Nutzerkomfort automatisiert zu steuern. Zwei wichtige Forschungsfragen dieses Vorhabens sind:

1) Inwieweit kann der Nutzerkomfort in Abhängigkeit zur Temperatur modelliert und beeinflusst werden?

2) Welches Regelungskonzept mit Prädiktionskomponente ist zur Erhöhung des Nutzerkomforts geeignet?

Kontakt / Betreuer: Yong Ding (TM Beigl)
yong.ding@kit.edu

Ein Optimierungskonzept zur Realisierung einer Markt-Netz-Kopplung

Die Echtzeitüberwachung des Energieflusses, die die physikalische Realität des Stromnetzes widerspiegelt, spielt eine entscheidende Rolle im Energiemarkt, da das Marktverhalten oft von der langfristigen Marktprognose abweicht. Die Echtzeit-Marktergebnisse haben wiederum großen Einfluss auf die Stabilisierung des Energiesystems im Sinne von Energieerzeugung, -verteilung und -verbrauch im physikalischen Netz. Aus diesem Grund erforscht dieses Vorhaben ein Kopplungsmodell, das eine interoperable Regelung zwischen dem Markt und dem Netz ermöglicht. Eine gute Möglichkeit der beabsichtigten Kopplung kann durch gekoppelte Optimierungsfragestellungen zwischen dem Markt und dem Netz formuliert werden. Das Hauptziel des Kopplungsmodells ist damit ein optimaler Betriebszustand des Kraftwerkeinsatzes in Abhängigkeit von Marktpreismechanismen zu erzielen, und gleichzeitig die Marktpreise durch eine dynamische Anpassung des Kraftwerkeinsatzes zu stabilisieren.

Kontakt / Betreuer: Yong Ding (TM Beigl)
yong.ding@kit.edu

Stochastische Verknüpfung von menschlicher Ganzkörperbewegung und natürlicher Sprache

Menschliche Bewegung bildet die Grundlage für eine Vielzahl von Anwendungsfeldern der Robotik. Marker-basiertes Motion Capture erlaubt die präzise Erfassung menschlicher Bewegung in großem Umfang und das Erstellen entsprechender Bewegungsdatenbanken. Die Repräsentation menschlicher Bewegung kann dann durch unterschiedliche Techniken erfolgen, z.B. Hidden Markov Modelle (HMM) zur Modellierung von Bewegungs-(Proto-) Symbolen. Die natürliche Sprache bietet eine mächtige und vielseitige Möglichkeit zur Beschreibung von Ganzkörperbewegung (z.B.: „Thomas wirft einen schweren Ball.“), die gleichzeitig ein intuitives Verständnis für Menschen ermöglicht und daher insbesondere für die Mensch-Roboter-Interaktion interessant ist. Auch zur Modellierung natürlicher Sprache existieren verschiedene Ansätze, z.B. statistische Modellierung in Form eines n-Gramm-Sprachmodells.

Im Rahmen des Projekts sollen Methoden zur stochastischen Verknüpfung von Bewegungs- und Sprachmodell untersucht werden. Hierzu werden beide Modelle zunächst unabhängig voneinander auf einem Trainingsdatensatz gelernt. Im Anschluss wird der stochastische Zusammenhang zwischen beiden Domänen gelernt. Beispielsweise können beide Modelle verknüpft werden, indem gelernt wird, mit welcher Wahrscheinlichkeit bestimmte Zustände einer Bewegung mit einem jeweiligen Wort der natürlichen Sprache assoziiert sind. Das so trainierte System erlaubt anschließend sowohl die Abbildung einer beobachteten

Bewegung auf eine natürlichsprachliche Beschreibung als auch die Synthese einer Bewegung ausgehend von einer natürlichsprachlichen Beschreibung.

Das hier vorgestellte Thema ist Bestandteil unserer Forschung in den EU-Forschungsprojekten KoroiBot und SecondHands.

Kontakt / Betreuer: Christian Mandery (IAR Asfour)
christian.mandery@kit.edu

Prof. Tamim Asfour (IAR Asfour) tamim.asfour@kit.edu

A Mobile Context Recognition System for Detecting Anomalies in Everyday Life Behavior Patterns

Especially in psychology and user experience research the monitoring of subjects in everyday life is a matter of concern. There is an increasing need for unobtrusive and ubiquitous sensing systems. The smartphone, our prime personal wearable, is a suitable sensing platform for this task. It is fairly unobtrusive and able to continuously log sensor data from which context information can be derived.

Of high interest is the recognition of activities of everyday living related to the current user context such as the current location. An example for this might be the recognition of a physical activity (derived from accelerometer data) at the current location (derived from GPS data).

It is your task to log sensor information from the smartphone to identify behavior patterns, train a behavior model and detect anomalies in everyday behavior. You should use active learning to adapt existing models to behavior characteristics and changes of the subject, e.g. due to a switch from lecture period to examination period.

Among others, your task within this topic are builing up on existing logging tools, deriving criteria and features for a (daily) behavior model, defining contexts of interest for specific personae and use case scenarios, applying pattern recognition and machine learning and gathering training and test data.

Kontakt / Betreuerin: Anja Bachmann (TM Beigl)
anja.bachmann@kit.edu

Graph-based Scene Representation for Whole-Body Humanoid Actions

The reliable detection of possible ways of interaction with environmental structures is a key capability for autonomous robots, especially when dealing with unknown environments. To tackle this problem we employ a perceptual pipeline that segments the environment into geometric primitives and assigns affordance labels, i.e. action possibilies, to these.

This perceptual pipeline works well for affordances that refer to individual primitives, e.g. a large, vertical plane suggests the whole-body action of leaning. However, there are more abstract affordances that refer to more complex environmental structures, like: “*A staircase affords climbing”, “A chair affords sitting”, “A door affords opening or closing”, “A ladder affords climbing”, “A car affords driving”.* These affordances refer to environmental structures, formed by specific combinations of environmental primitives. The goal of this work is to enhance the perceptual pipeline in order to identify known structures within the set of environmental primitives.

The scene will be represented in a topological graph, each node referring to an environmental primitive. This graph allows for searching geometric relationships that indicate known environmental structures, such as staircases or doors. These structures will be identified and abstract affordances will be assigned to them.

The goal of the work is the design implementation of the graph-based scene representation. This includes querying mechanisms as well as pruning strategies. The graph and the query algorithms will be evaluated using real-world examples captured by the humanoid robot ARMAR-III.

Kontakt / Betreuer: Markus Grotz (IAR Asfour) markus.grotz@kit.edu

Peter Kaiser (IAR Asfour) peter.kaiser@kit.edu

Testen von Informationsflusseigenschaften

Es ist wünschenswert, dass Programme sensible Informationen wie Kontodaten oder Passwörter nicht veröffentlichen. Um dies sicherzustellen, kann man mit Hilfe von Verifikationswerkzeugen die Datenflüsse eines Programms analysieren und zeigen, dass keine sicherheitsrelevante Informationen geleckt werden. Die Eingaben und Ausgaben des Programms werden in zwei Kategorien eingeteilt – vertraulich (high) und öffentlich (low). Man muss dann zeigen, dass die öffentlichen Ausgaben unabhängig von den vertraulichen Eingaben sind, also bei gleichen öffentlichen Eingaben müssen gleiche öffentliche Ausgaben herauskommen, unabhängig von den vertraulichen Eingaben. Diese Eigenschaft (non-interference) spricht also über zwei beliebige Programmabläufe und ist deswegen schwieriger zu beweisen als funktionale Eigenschaften, die nur einen Programmablauf betrachten. Da das Beweisen dieser Eigenschaften oft zu kompliziert ist, möchten wir sie testen. Dafür müssen ausführbare Testfälle aus einem Teilbeweis generiert werden. Das Testen erfolgt automatisch und ist somit einfacher als das Beweisen, wo interaktive Schritte nötig sein können. Allerdings kann ein Test die gewünschte Eigenschaft nur für die beim Testen verwendete Eingaben zeigen. Daher müssen die Testeingaben so ausgewählt werden, dass sie das Programmverhalten so weit wie möglich abdecken.

Mögliche Aufgaben im Rahmen dieses Themas sind:

1)Definieren des Konzepts eines Informationsfluss-Testfalls

2) Definieren von Abdeckungskriterien für Informationsfluss-Testfälle wie aussagekräftig ist eine Testsuite?

3) Erweiterung des Testfallgenerierung-Frameworks in KeY, so dass neben funktionalen jetzt auch Informationsflusseigenschaften getestet werden können

4) Exploit-Generation - statt eine Testsuite zu generieren, die möglichst viel abdeckt, gezielt nach Inputs suchen, die die Eigenschaft verletzen und Testfälle nur für diese Inputs erzeugen. Die generierte Testsuite soll einen Angreifer simulieren, der die öffentliche Eingabe beliebig setzen kann, um Informationen über die vertrauliche Eingabe zu gewinnen.

Kontakt / Betreuer: Mihai Herda (ITI Beckert) herda@kit.edu

-

-

Präziseres Programm-Slicing durch semantische Analyse

-

Programm-Slicing ist ein Verfahren, um Teile von Programmen zu entfernen, die für einen bestimmen Aspekt bzw. eine Eigenschaft des Programms irrelevant sind (z.B. Anweisungen, die keinen Einfluss auf den
Wert einer Variable an einem Punkt im Programm haben).
Slicing findet unter anderem im Programmverstehen und der Code-Optimierung Verwendung.
Viele existierende Ansätze arbeiten allerdings syntaktisch; sie abstrahieren von der Bedeutung von Anweisungen und erkennen nicht, dass
z.B. die beiden Anweisungen x++; x--; auf das Endergebnis von x keinen
Einfluss haben. Dadurch werden Slices oft überapproximiert und damit zu
groß und wenig nützlich.

Im Rahmen dieses Projekts soll eine neue Methode zum semantischen Slicen
entwickelt werden, die auch die Semantik von Programmen berücksichtigt.
Die Methode soll über den Stand der Technik hinausgehen, indem sie ausdrucksvollere Kriterien erlaubt und präzisere – also kleinere – Slices berechnet. Es ist ein Ziel, die Methode prototypisch zu implementieren, sowie sie zu evaluieren.

Kontakt / Betreuer: Mihai Herda (ITI Beckert) herda@kit.edu
Thorsten Bormer (ITI Beckert) bormer@kit.edu
Mattias Ulbrich ulbrich@kit.edu

Inkrementelle Modellsichten

Moderne Software-Systeme weisen eine hohe Komplexität und eine umfangreiche Größe auf. Daher werden in der Entwicklung solcher Systeme mehrere Sprachen und Modelle verwendet, um unterschiedliche Gesichtspunkte und Abstraktionsebenen der Systeme zu beschreiben. Beispielsweise können Softwaresysteme unter anderem mit Komponentenmodellen der Architektur, Performance-Modellen zur Analyse der nichtfunktionalen Eigenschaften oder mit der Implementierung im Code dargestellt werden. Obwohl all diesen Artefakten unterschiedliche Konzepte und Formalismen zugrunde liegen, beschreiben sie dasselbe System aus unterschiedlichen Blickwinkeln. Zwischen diesen Modellen kommt es häufig zu semantischen Überlappungen, beispielsweise sind Komponenten aus der Architektur auch in der Implementierung in Form von Klassen wieder zu finden. Um ein vollständiges Bild einer Komponente zu

erfassen, müssen Informationen daher aus verschiedenen Modellen aggregiert werden in sog. Sichten. Die Entwicklung dieser Sichten wird dabei von Sprachen wie beispielsweise dem am Lehrstuhl Reussner entwickelten ModelJoin unterstützt. Ein grundsätzliches Problem hierbei ist es jedoch, dass sich während des Softwareentwicklungsprozesses die zugrundeliegenden Modelle der Architektur, der Performance-Modelle und der Implementierung ändern. Da diese Änderungen potentiell Auswirkungen auf die Sichten haben, müssen diese neu generiert werden. Für große Softwaresysteme ist die ständige Neugenerierung dieser Sichten aber zeit- und rechenintensiv, gerade weil mit jeder Änderung nur kleine Teile des Systems geändert werden. Es ist daher lohnenswert, möglichst große Teile der bestehenden Sicht zu belassen und nur die Teile zu aktualisieren, die tatsächlich von einer Änderung betroffen sind. Man spricht hierbei von einer inkrementellen Ausführung der Sicht. Im Rahmen der Projektgruppe soll untersucht werden, inwiefern neue Forschungsansätze zu impliziter Inkrementalität dazu geeignet sind, eine solche inkrementelle Ausführung von Sichten automatisch aus der Spezifikation eines Sichttyps abzuleiten. Dazu soll ein System entwickelt werden, dass mit der Sprache ModelJoin spezifizierte flexible Sichten automatisch inkrementelle ausführen kann. Hierbei kann auf das ebenfalls am Lehrstuhl Reussner entwickelte inkrementelle Ausführungssystem NMF Expressions zurückgegriffen werden. Die Korrektheit des Systems soll durch Fallstudien belegt und deren Nützlichkeit durch Messung des Laufzeitverhaltens und des Speicherverbrauchs belegt werden. Studierende werden dabei durch Mitarbeiter des Lehrstuhls Reussner betreut.

Kontakt / Betreuer: Erik Burger (SDQ Reussner) burger@kit.edu
Georg Hinkel (SDQ Reussner) georg.hinkel@kit.edu

Anonymisierung von Datenbanken

- Zu statistischen Zwecken werden regelmäßig anonymisierte Datenbanken veröffentlicht, z.B. anonymisierte Patientendatenbanken von Krankenhäusern. Forscher können diese dann auf statistische Korrelationen hin untersuchen, die z.B. Hinweise auf Nebenwirkungen von Medikamenten geben können. Aufgrund der sehr sensiblen personenbezogenen Daten in solchen Datenbanken ist eine gute Anonymisierung unbedingt notwendig.
- Es gibt jedoch zahlreiche Fälle von nicht hinreichend anonymisierten Datenbanken. Mit verschiedenen Methoden konnten einzelne Personen oder ganze Personengruppen in den anonymisierten Datenbanken re-identifiziert („deanonymisiert“) werden. In der Folge entstanden wissenschaftliche Arbeiten, die Anonymisierungsverfahren untersuchten. Inzwischen gibt es eine Vielzahl von Anonymitätsbegriffen und Anonymisierungsverfahren, die gewisse Garantien bieten.
- Im Rahmen der Seminararbeit zum Stand der Forschung sollen ein Vergleich der Anonymisierungsleistung von verschiedenen Verfahren durchgeführt werden. Hierbei sollen sowohl Verfahren aus der Forschung als auch real eingesetzte Verfahren berücksichtigt werden.

- Im weiteren Verlauf von „Praxis der Forschung“ können dann z.B. gängige Verfahren genauer untersucht werden oder bestehende Ansätze weiterentwickelt werden.
- Ansprechpartner: Gunnar Hartung (ITI Crypto)
  gunnar.hartung@kit.edu