

Diss. ETH No. 22203

# Interacting with the Web of Things

A dissertation submitted to  
ETH ZURICH

for the degree of  
DOCTOR OF SCIENCES

Presented by

**Simon Mayer**

MSc in Computer Science, ETH Zurich

born July 16, 1987

citizen of Austria

accepted on the recommendation of

Prof. Dr. Friedemann Mattern, examiner, ETH Zurich

Prof. Dr. Sanjay Sarma, co-examiner, Massachusetts Institute of Technology

Prof. Dr. Gustavo Alonso, co-examiner, ETH Zurich

2014

A fundamental paradigm shift is currently taking place in the field of computing: due to the miniaturization of computing devices and the proliferation of embedded systems, tiny, networked computers can now be easily integrated into everyday objects, turning them into *smart things*. In the resulting *Internet of Things*, physical items are no longer disconnected from the virtual world but rather become accessible through computers and other networked devices, and can even make use of protocols that are widely deployed in the World Wide Web, in a paradigm that we call the *Web of Things*. Eventually, smart things will be able to communicate, analyze, decide, and act – and thereby provide an invisible background assistance that should make life more enjoyable, entertaining, and also safer. However, in an environment that is populated by hundreds of Web-enabled smart things, it will become increasingly difficult for humans to interact with devices that are relevant to their current needs, and to find, select, and control them.

The objective of this thesis is to investigate how human users could be enabled to conveniently interact with individual smart objects in their surroundings and to interconnect devices and configure the resulting *physical mashups* to perform higher-level tasks on their behalf. To achieve basic interoperability between devices, we rely on the World Wide Web with its proven protocols and architectural patterns which emphasize scalability, generic interfaces, and loose coupling between components.

As a first step to facilitate the interaction with smart things on top of the basic Web principles, we propose the embedding of metadata for automatically generating user interfaces for smart devices. Our specific approach enables not only the generation of more intuitive graphical widgets but also the mapping of interactive components to gesture-based, speech-based, and physical interfaces by describing the *high-level interaction semantics* of smart devices instead of specifying purely interface-specific information. The provisioning of an interaction mechanism with a smart object is thus reduced to the embedding of simple interaction information into the representation of the smart thing. Before users can start *interacting* with a smart device, it must, however, first be *selected*. To permit users to choose which of the many smart objects in their surroundings should be involved in an interaction, we propose to use technologies for *optical image recognition*.

The visual selection of smart things and automatically generated user interfaces enable end users to conveniently interact with individual services in their surroundings that are embodied as specific physical objects. To complement the direct interaction with smart devices, the second part of this thesis focuses on more complex use cases where

*multiple smart objects* must collaborate to achieve the user's goal. Such situations arise, for instance, in home or office automation scenarios, or in smart factories, where machines or assembly lines could adjust to better support the operator.

To put users more in control of entire environments of smart devices, we present a system that records *interactions between smart things and with remote services* and displays this data to users in real time. To do this, we use an augmented reality overlay on the camera feed of handheld or wearable devices such as smartphones and smartglasses. Next, we propose a management infrastructure for smart things that makes the services they offer discoverable and composable, and fully integrates them with more traditional Web-based information providers. This system enables humans to *find and use* data and functionality provided by physical devices and allows machines to support users in finding services within densely populated smart environments and even to discover and use required services themselves, on behalf of the user. The basis for these applications is a generic mechanism that allows smart devices to provide semantic descriptions of the services they offer. Specifically, our infrastructure supports the embedding of *functional semantic metadata* into smart things that describes which functionality a concrete object provides and how to invoke it. Based on this metadata, a semantic reasoning component can find out which composite tasks can be achieved by a user's smart environment and can provide instructions about how to reach concrete goals, thus enabling the *configuration of entire smart environments* for end users.

As a concrete use case, we present a platform that applies our proposed interaction modes with smart things to automobiles: a mobile application recognizes cars, downloads information about them from a back-end server, and displays this information – as well as interaction capabilities with the car and its services – on the user's interface device. The back-end server furthermore exposes functional metadata about the capabilities of individual cars to make their services automatically usable within physical mashups. Finally, it records client interactions to enable car owners to monitor in real time who accesses which kind of data and services on their vehicles.

The overarching objective of this thesis is to show how current technologies could support the interaction of end users with Web-enabled smart devices. To achieve this, we make use of a number of technologies from different areas of the computer science discipline: A management infrastructure makes smart things discoverable for human users and machines and builds upon current research in the distributed systems domain. State-of-the-art computer vision technologies allow users to select devices in their environment using handheld or wearable computers such as smartphones or smartglasses. Novel methods from the field of computer-human-interaction enable the embedding of metadata that allows for automatically generating user interfaces. Finally, semantic technologies enable flexible compositions of smart things that collaborate to achieve the user's goal.

Anhaltende Fortschritte bei der Miniaturisierung von Mikroelektronik und Sensorik sowie bei Kommunikationstechnologien ermöglichen die Einbettung von vernetzten Kleinstcomputern in Alltagsgegenstände. Solche sogenannten „Smart Things“ – schlaue, wenn auch nicht im eigentlichen Sinne intelligente Dinge – sind Geräte, die mit einer virtuellen Präsenz im *Internet der Dinge* gepaart sind. Sie können miteinander und mit Menschen kommunizieren, ihre Umwelt durch Sensoren wahrnehmen, autonom Entscheidungen treffen und auf die Welt mittels ihrer Aktoren einwirken. Wenn dies im Sinne der menschlichen Benutzer geschieht, agieren vernetzte, schlaue Dinge wie eine unsichtbare Hintergrundassistentin, die unser Leben angenehmer, unterhaltsamer und auch sicherer machen kann. Sollten diese Geräte zudem die Fähigkeit mitbringen, Kommunikationsprotokolle, welche im World Wide Web eingesetzt werden, zu verwenden, so sprechen wir vom „Web of Things“, dem Web der Dinge.

Das Ziel dieser Arbeit ist, zu untersuchen, wie menschliche Benutzer dabei unterstützt werden können, sich in Umgebungen zurechtzufinden, die hunderte schlauer Dinge enthalten. In solchen *smarten Umgebungen* ist es für Benutzer insbesondere schwierig, jene Geräte, die sie gerade benötigen, effizient aufzufinden und intuitiv auszuwählen, und mit ihnen in geeigneter Weise zu interagieren. Zudem soll es Benutzern ermöglicht werden, Geräte in derartigen Umgebungen so zu konfigurieren, dass sie in kooperativer Weise Aufgaben erledigen können, welche für ein einzelnes schlaues Ding zu komplex sind. Die Basis für diese Arbeit bildet dabei das World Wide Web, das durch seinen Aufbau und seine offenen Protokolle eine grundlegende Interoperabilität zwischen schlauen Dingen ermöglicht.

Zunächst wird in der vorliegenden Arbeit die direkte, unmittelbare Interaktion von Benutzern mit schlauen Dingen behandelt: Hierfür stellen letztere Metadaten zur Verfügung, welche die *automatische Erzeugung von Nutzungsschnittstellen* auf tragbaren Geräten wie Smartphones, Tablets und Smartglasses ermöglichen. Indem schlaue Dinge ihre Interaktionssemantik auf hoher Ebene beschreiben, anstatt nur schnittstellenspezifische Informationen bereitzustellen, ermöglicht unser Konzept nicht nur die automatische Erzeugung von grafischen Widgets, sondern gleichzeitig auch gestenbasierte Interaktion sowie Sprachsteuerung. Darüber hinaus vereinfacht unser Ansatz das Beschreiben der Interaktionssemantik selbst, sodass dies sogar Laien ermöglicht wird. Bevor allerdings eine Nutzungsschnittstelle geladen werden kann, um mit einem schlauen Gegenstand zu interagieren, muss dieser vom Benutzer ausgewählt werden. Hierfür verwenden wir aktuelle

Technologien aus der optischen Bilderkennung: Um die *Interaktion mit einem Gerät zu initiieren*, müssen Benutzer in unserem Ansatz lediglich mit der Kamera ihres Smartphones oder Tablets auf das Gerät zielen – falls sie über Smartglasses verfügen, ist es ausreichend, das Gerät einfach nur anzusehen.

Im zweiten Teil der Arbeit wird die Interaktion mit smarten Umgebungen als Ganzes untersucht, wobei mehrere schlaue Dinge selbstständig zusammenarbeiten sollen, um Benutzer bei komplexeren Aufgaben zu unterstützen. Mithilfe der in dieser Arbeit entwickelten Technologien können beispielsweise Automatisierungsszenarien zu Hause und in Fabrikumgebungen umgesetzt werden – dort sollen sich in Zukunft einzelne Geräte oder ganze Fertigungsanlagen automatisch abstimmen und schnell anpassen, um den Produktionsprozess effizienter zu gestalten, insbesondere bei kleinen Losgrößen. Zunächst beschreiben wir eine Managementinfrastruktur, die das Auffinden und Zusammensetzen von Diensten, die von schlaue Dingen bereitgestellt werden, vereinfacht. Dieses System unterstützt insbesondere die *Suche nach Geräten und Diensten* in dichten smarten Umgebungen und ermöglicht durch *eingebettete semantische Beschreibungen* anderen Geräten, diese im Sinne des Benutzers anzusteuern. Auf Basis dieser semantischen Metadaten, welche die Funktionalität von einzelnen schlaue Geräten charakterisieren, stellen wir sodann ein System vor, welches ermitteln kann, welche Aufgaben Geräte in schlaue Umgebungen *gemeinsam* erledigen können. Hierfür gibt der Benutzer lediglich den erwünschten Zielzustand seiner Umgebung an, und unser System findet mithilfe eines semantischen Reasoners selbst heraus, ob und wie dieser Zustand erreicht werden kann. Um die Kontrolle über solche dynamischen Abläufe in smarten Umgebungen zu behalten, stellen wir ein System vor, welches Interaktionen zwischen schlaue Dingen protokolliert und sie in Echtzeit auf Geräten wie Smartphones oder Smartglasses in Form einer Augmented-Reality-Einblendung visualisiert.

Abschließend wird ein konkreter Anwendungsfall der in der vorliegenden Arbeit vorgestellten Konzepte und Technologien behandelt: „Connected Cars“. Eine von uns entwickelte Anwendung erkennt Fahrzeuge optisch und erzeugt automatisch Schnittstellen, welche benutzerfreundliche Interaktion mit Fahrzeugsensoren und -aktoren ermöglichen – beispielsweise das Auslesen des aktuellen Tankfüllstands und des Treibstoffverbrauchs oder die Bedienung der Fahrzeugverriegelung. Zusätzlich visualisiert die Anwendung Interaktionen zwischen Fahrzeugen und anderen schlaue Dingen, wie auch mit Online-Diensten, um insbesondere unberechtigte Zugriffe auf die durch das Fahrzeug bereitgestellten Daten und Dienste aufzuzeigen. Des Weiteren ermöglichen eingebettete semantische Metadaten die Einbindung von Fahrzeugdaten in zusammengesetzten Anwendungen, welche komplexe Aufgaben übernehmen können.

Das übergeordnete Ziel dieser Arbeit besteht darin, zu zeigen, wie aktuelle Technologien aus verschiedenen Bereichen der Informatik die Interaktion zwischen Benutzern und schlaue Dingen substantiell vereinfachen können. Die Arbeit vereint dazu Methoden aus der grafischen Datenverarbeitung, künstlichen Intelligenz, Mensch-Maschine-Interaktion und verteilten Systemen, um Nutzern das Auffinden und Auswählen von schlaue Dingen zu ermöglichen und mit ihnen, sowie mit smarten Umgebungen insgesamt, zweckmäßig und effizient zu interagieren.