
Towards Reactive Information Systems and their Services

MASTER THESIS

Author:
Dominic BOSCH

Supervisors:
Prof. Dr. Helmar BURKHART
Dr. Martin GUGGISBERG

June 3, 2014



Abstract

This is very abstract

Acknowledgment

I would like to express my deep gratitude to my master thesis advisor, Prof. Dr. Helmar Burkhart, for his patience and constant encouragement to keep going further. I also want to thank Dr. Martin Guggisberg for his ever helpful advices, hints to existing technologies and guidance through the whole process of this thesis. Besides my advisors, I would like to thank the members of the research group: Danilo Guerrero, Antonio Maffia, Alexander Gröflin and Robert Frank for their help, many interesting discussions and inputs. My thanks also go to Manus Bonner for reading through this and coming up with well formed sentences. Finally I would like to thank my family; my parents Silvia and Peter Bosch, my Sister Svenja, her husband Andreas Buser, their seven months old son Nico and my girlfriend Kathrina for their inspiring ideas, the joyfulness, irresistible smiles and financial support they gave me on my journey.

Contents

1	Introduction	1
2	Related Work	3
2.1	Services in the Web	3
2.2	Rule Engines & Rule Languages	6
3	Conceptual Model for Reactive Information Systems and their Services	8
3.1	From Real Events to Events in the Web	8
3.2	The Web's Event & Action Information Space	9
3.3	Impose Reactivity to the Web Information Space	10
4	Applicability	12
4.1	Augment existing Web Applications	12
4.2	Service Functionality and Availabilty Checking	14
4.3	Exploiting the Web of Things	14
5	Prototype System	16
5.1	Event Trigger	17
5.2	Action Dispatcher	17
5.3	ECA Rules in the Engine	17
5.4	Web Programming	17
6	Discussion & Results	20
	Bibliography	20
	Index	23
	Appendices	25

Chapter 1

Introduction

The Web is an ever growing institution(*entity*), in all aspects that it covers. The number of information and functionality providers for the Web is growing in the whole spectrum from bigger computing centers down to smaller devices. Computing centers, growing in size and quantity, allow for massive amounts of data to be stored and accessed, but they also enable the construction and accessibility of more complex functionalities. Ever smaller devices provide informations and functionalities to the Web, through their quickly increasing number. Many of them are even providing access to the Web itself and thus leverage the effect of the growing Web. For example mobile phones can act as a hotspot to grant Web access to others through WiFi. All the smart Things with access to the Web start to form the Web of Things [13]. This can be everyThing from a temperature sensor to all electronical devices within a house. These Things do not only provide sensor data but they can also be easily controlled through the Web.

TODO use the word task or workflow?

Confronted with this rapid growth of the Web, an increasing number of human beings is exposed to it in their daily life and they get literally flooded with data and possibilities to govern them. Even though they have access to all these services in the Web, they often lack the knowledge, necessary time or right approach to weild them. Users need ways to get appropriate informations, in the right moment, automatically and in a condensed matter that suits them best. They have to be able to automate tedious tasks, e.g. detecting relevant changes in information resources and react on behalf of such changes. This requires the identification of and filtering for user-specific information, appropriate timing, assembly and preprocessing of different information resources and finally the forwarding of outcomes to services in the web.

Users don't want to be bound to specific services or applications, they want to use their preferred one, which they are used to and which helps them best to fulfill their needs. Some of these services in the web offer ways to spread their data to other pre-defined services, but those are often mere copy/paste task. Therefore users end up mashing informations or functionalities from different locations within the Web by hand. This often means to execute similar tasks repeatedly by hand. Moreover the completion of such a task suffers from latency due to the deferred detection of the initialization or because the user is just not able to execute the work at that time. And also if the bits and pieces that form such a task could be separated, it is likely that some parts could be reused for other tasks or even by other people to get similar work done.

Since the data and the functionality already exists in the web, the users could automate their work to some extent by orchestrating services and data. Even though the access to services and data gets simpler, the average user is still not able to program the Web. Another challenge is that often a lot of effort needs to be invested, in understanding how the service works, before it

can be fully exploited. It is desirable to retain manageable mashing of services and data, while still exploiting their full potential. There are a few possibilities in the Web that go towards easing the programmability of the Web, but they are either complicated to weild themselves or mere data copy or mashup approaches.

A big part of the informations, that become available to the users, are short-lived informations corresponding to certain state changes, and therefore can be modeled as events. In this thesis we to introduce an event-driven conceptual model that enables programmability of the Web and through this imposes reactivity on the Web. We claim the whole Web as our information space, in which we listen for triggered events and on which we execute actions as a result. Such an user-specific reactivity allows a personalization of the whole Web and a tool to govern its information flood. This would allow users to orchestrate the Web in an intuitive way. Since current approaches concentrate on data flow rather than on event flow, which is mere data copy and spreading than reactivity, we believe that a conceptual model can overcome certain shortcomings of the existing solutions.

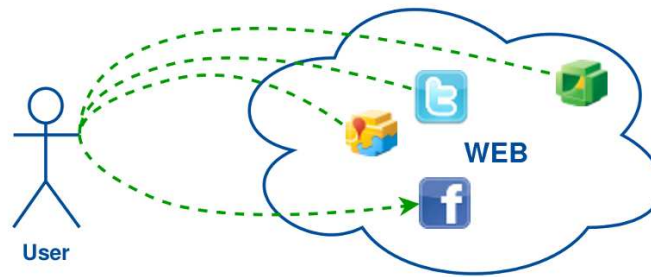


Figure 1.1: Users mashup the Web's Information and Data

Chapter 2

Related Work

Research on behalf of the Web attracts a great deal of attention. This is not surprising since modern business and life is already impossible without the internet. Great opportunities arise with it and we cover a part of the Web service orchestration in it. There Service-oriented Architecture (SOA) [21] With the wide adoption of SOA, we see an increasing number of Web accessible services and their compositions. [14] By the end of 2013 this number grew to 10'000 <http://www.programmableweb.com/news/programmablewebs-directory-hits-10000-apis.-and-counting./analysis/2013/09/23> At the

2.1 Services in the Web

What are services Other recent research [14] has shown that the number of accessible Web APIs follows power law distribution and thus provides an ever growing source of data and functionality. [20] An important development within the Web are the increasing number of available services. The term service in the Web is not very precise and the Web's short history has already seen a lot of different kinds.

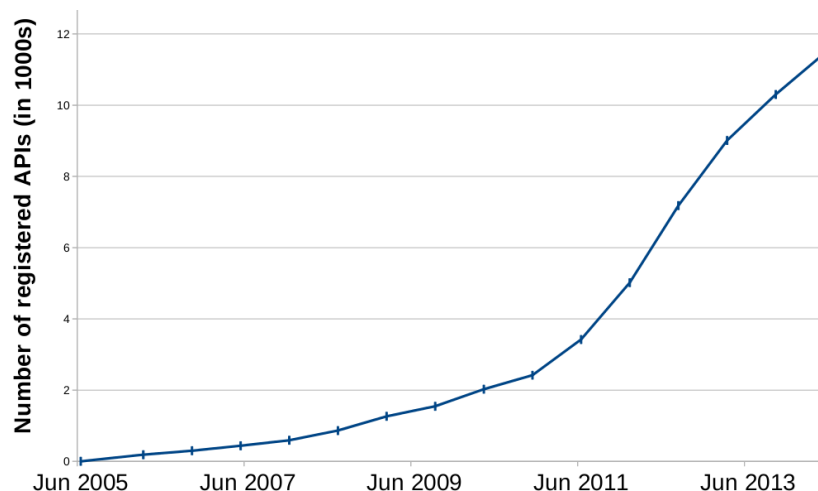


Figure 2.1: Number of registered APIs in the ProgrammableWeb directory by date

A fairly early adoption of the service concept to computers were the Remote Procedure Calls (RPC) [4]. Through RPC a piece of code can be executed on a different machine, than the one which is calling the procedure. It is basically achieved via inter-process communication and

doesn't necessarily require the Web. Even more since when RPC was invented, the World-Wide Web [3] didn't even exist. RPC also found its use in grid computing [24] and through this, opened doors into the field of distributed computation. The RPC paradigm isn't bound to certain technologies and thus, has been implemented in a lot of different programming languages. These implementations were tightly bound to the respective language that was used, which resulted largely in incompatibility among them. It became necessary to enhance RPC's in order to get cross platform compatibility. By abstracting RPC functionality with the Extensible Markup Language (XML) [7], compatibility between services that used different technologies was easier to achieve.

Since XML-RPC was held relatively simple but received a lot of attention, it was further enhanced. Together with additional functionality, XML-RPC turned into Simple Object Access Protocol (SOAP) [6]. SOAP is accompanied by the Web Service Description Language (WSDL) [9] which is used to describe the interfaces to SOAP services. Through SOAP and WSDL a client for the service can issue a request for the WSDL information of the service and retrieves all interface specifications he requires in order to issue a call to the actual service. The service specifications are then incorporated into the existing application as if it is a local function call. The SOAP layer takes care of marshalling the request and unmarshalling the response.

Another initiative that aimed for eased communication between different platform is the Common Object Request Broker Architecture (CORBA) [12]. As the name already suggest it is an object-oriented approach and it allows the exchange of whole objects. CORBA relies on its communication layer, the Object Request Broker (ORB), which forms the basis of its architecture. The platform-specific ORB's provide the communication abstraction, which free the application from platform dependencies. Similar to SOAP's WSDL, CORBA has its Interface Definition Language (IDL) to provide information about the objects to be offered and accessed. An object is instantiated by an application and the interface to this instance is offered through the ORB. Another application attached to the ORB can then access all public variables, data structures and functions of this object. This means not only remote access to variables and data structures, but also remote function invocation. CORBA requires the implementation of object-oriented mechanisms in programming languages which aren't object-oriented. This can be technically difficult and become an eventually tedious task. CORBA allows communication between applications written in different programming languages and which are running on the same physical computer, as well as the communication between different computers in the same network. With the Internet Inter-ORB Protocol (IIOP) it is also possible to connect ORB's over the Web. Through this, the offered objects can become services in the Web, though they are shielded by the ORB.

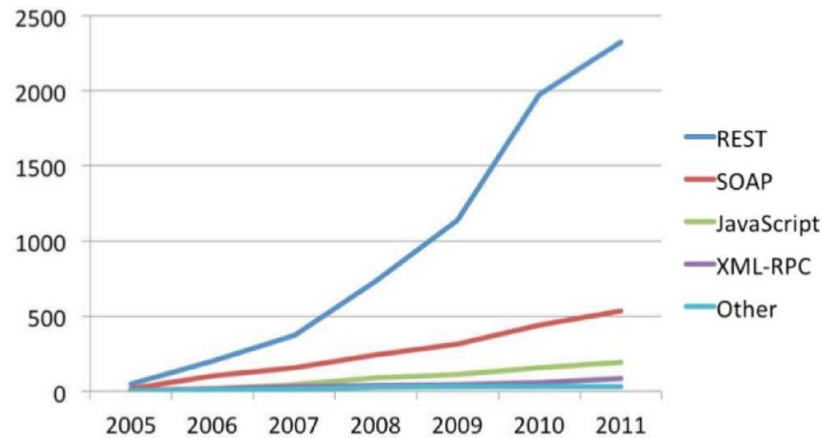
Naturally, as the Web grows quickly in its offered services and also efforts to access them, trends lead toward simple ways to access services in the Web. During our research it turned out that SOAP's overhead compared to REST diminishes its popularity on behalf of the latter one. But still since SOAP services are request responders they have their place in our model, both as event trigger and action invoker. Web of Things [13] has REST interfaces

Several different resources [Figure 2.2 (John Musser, ProgrammableWeb, 2011)] draw the same picture, an increasing popularity of REST over SOAP by an order of magnitude and a exponential growth of the number of accessible Web API's.

An interesting trend that could be observed was the growing popularity of REST over SOAP. [22]

Web APIs are Web accessible endpoints for users to invoke.

Twelve Theses on Reactive Rules for the Web [8]: This article investigates issues of relevance in designing high-level programming languages dedicated to reactivity on the Web. It presents



Distribution of API protocols and styles

Based on directory of 3,200 web APIs listed at ProgrammableWeb, May 2011

Figure 2.2: REST vs. SOAP: Simplicity wins again
(John Musser, ProgrammableWeb, 2011)

twelve theses on features desirable for a language of reactive rules tuned to programming Web and Semantic Web applications.

Data and functionalities in the Web were always accessible via Web services, whatever this means. reengineering of services in the beginning with standardised REST API's we got to Web API, easy access and understandable

The fast evolving Web has brought up a trend towards easy to master interfaces to services, the so called Web APIs. They do not only provide access to simple functionality but whole applications that allow access over Web APIs. These trending Web APIs benefit from a RESTful architecture which predominantly uses HTTP and thus relies on the most basic and powerful operations and the basis of the Web itself, the HTTP protocol.

Mashups combine information and functionality of more than one Web service in a single place. The mashing up of such Web services allows data to be enhanced with new informations, processing/refinement of the information, or even ways to interact with them, e.g. through Google Maps. Simple functions from different sources can be combined into more powerful ones, which influence data and services in a way their founders eventually didn't even think of. Web service mashups have been developped ever since services in the Web started to exist and were accessible in a more or less convenient way. We introduce Paul Rademacher as an example for how recent the invention of Web service mashups are. He's one of the first inventors of such a Web service mashup. In the same year after Google Maps came up in 2005, he invented a site that displayed Craigslist houses on a Google Map. With no Google Maps API at that time, he needed time and skills to reverse engineer Google Map's functionalities.

A large number of such "static" mashups were and are still developped. They are static in the way that they aggregate a fixed (and mostly low) number, of either data or functionality resources, to provide an enhanced resource in a specialized domain. Of course Mashups can be mashed up again, to provide even more sophisticated functionality and data. Some latest example Mashups, taken from the ProgrammableWeb collection, are:

- Wifi and Plugs: MapBox, Google Docs and Import.io API's used to display where Wi-Fi and plugs are available in London.

- MapLight: GovTrack.us and OpenSecrets API's used to combine political results with financial contributions to show how capital contributions affect voting.
- Shared Count: Facebook, LinkedIn, Pinterest and Twitter API's used to display informations about how well spread a URL is on social media sites.

In the past few years, research and development for platforms to allow users to flexibly mashup Web APIs got attention. With IFTTT and Zapier, two platforms have evolved out of this process. Users that register on those platforms are provided with a multitude of Web API functions that act as event triggers and such that are used to execute actions. The user is then free to combine these event triggers and actions in the way it suits best, creating helpful Web API mashups on their own.

2.2 Rule Engines & Rule Languages

It turns out that Web App mashing up is not able to bring reactivity to the Web. They are merely aggregations of services, which only provide data or functions but no write possibilities such as Web applications provide.

Thus

Several different rule languages have been developped for different purposes over the last years and they vary grately in their purpose. A compilation of research on different emerging rule-based languages and technologies [17] gives an oveview over such efforts. We examined different existing rule languages with respect to a certain use case to identify its applicability for reactivity in the Web. The use case is defined such that the rule needs to suffice the ECA paradigm: In Appendix A.1 In Appendix ??

- Event: Receipt of an Email
- Condition: Check for a certain sender
- Action: Store it remotely via a Web API

Definition of events[1] CEP or Composite event detection by means of using event templates[15]

We defined an email event which the rule languages need to be able to process. The JSON representation of the given email event as depicted in the appendix.

An early ECA Rule Language for XML repositories [16] was postulated in 2003 and was picked up by many researches afterwards. It was designed to react on insert and delete events within XML repositories and as an action change XML documents.

Now apart from implementing a rules engine, we would also need to add an XML document event manager which interpretes and pushes events into the XML file *inbound_queue.xml*. Then again this instance would interpret the ouptuts of the ECA engine, which would theoretically manifest in other XML documents, and produce meaningful actions on remote hosts. This wouldn't be an architecture which has its focus on the solution of our use case and, as a result, add complexity and create an unnecessary overhead.

To make the lengthy RDF definitions smaller and more readable, Notation 3 [2] was designed and announced in 2005. Through the implies operator(\Rightarrow) an "event" can be connected to an "action", both expressed in RDF's subject, predicate, object notation, which makes the expression of ECA rules a complicated and not very intuitive task. A solution to our use case would look as follows:

This language is used to express relations between entities and thus not really suitable for our use case, since we would require another interpreter to infer the actions. But concepts and ideas of the work that was done in these consortias could eventually still find influence into our solution.

The rule language XChange [19] was the outcome of the REVERSE ([23], Reasoning on the Web with Rules and Semantics) project, which was funded by the EU and Switzerland. Their work influenced a number of future research. The language was designed to add reactive behaviour to a "static" Web which is represented through XML resources. Thus we have action logics to alter such resources through insertions and deletions. Since we aim to utilize Web API's for our rule language we need a more generic approach which adds flexibility in term of the API provided. But the thorough research done with the language XChange holds valuable concepts, especially in terms of temporal event composition. This could be a rule according to our use case:

But XChange is designed to access other resources in an action and thus provides powerful tools:

In 2008 *JSON Rules* [11] was introduced as a language to easily react on specific DOM tree compositions. The usage of JavaScript allowed them to provide simple functions which could be called directly by the actions, thus abstracting functionality from the language. This key concept found influence into our language as it allows different layers of abstractions. Through this it is possible to provide generic functions for expert user as well as very limited functions with only few possibilities for parameterization to be used by unexperienced persons. A drawback of this language is its binding to DOM tree events, where we would want to react on any events happening in the world. Also the temporal composition to complex events is not a subject of their work and needs further attention.

A recent open-source development is the Kinetic Rules Engine together with the Kinetics Rule Language [25]. It is built for the purpose of adding reactivity to the cloud. The language is based on declarative syntax, enriched with imperative elements. But it is a tedious task to get into a whole new language and their caveats. *authorization?*

The basis of *RuleML* [5] is datalog, a language in the intersection of SQL and Prolog. In 2012 the *Reaction RuleML* [18] language incorporated several different types of rules into the RuleML syntax, to establish a uniform syntax and interchangeability of rules. *Reaction RuleML* is a valuable resource in terms of manifold research that has been done in the domain of rule languages, but the syntax is not user-friendly.

R2ML allows usage for RuleML together with many other dialects. Really!?

Most of the examined rule languages are designed for the interchangeability of rules between different service providers. We do not attempt to jump into this domain but we rather pick up important concepts to manifest Web API's as first class citizens of our rule language. This allows the ad-hoc design and implementation of reactive rules between existing Web API's without the need for their cooperation in setting up their endpoint in a special way.

Chapter 3

Conceptual Model for Reactive Information Systems and their Services

In the previous chapter we have shown that services in the Web and reactivity through programmability have received a great deal of attention. Our goal is to combine both research fields in order to achieve reactivity within the Web by orchestrating its information space.

3.1 From Real Events to Events in the Web

Real events are always bound to a spatial location and a point on the time axis. An earthquake for example always has an epicenter and an occurrence time. Different points on earth's surface would feel the earthquake, which originates from the same epicentre, at a different point in time with a different intensity. A Web event model of an earthquake would consist of a large number of identical **ground-shake** events that occur at different points in time and places. Therefore they would hold different spatial location informations and intensities. These events can be thought of as emitted into the Web by a seismometer sitting at the corresponding location.

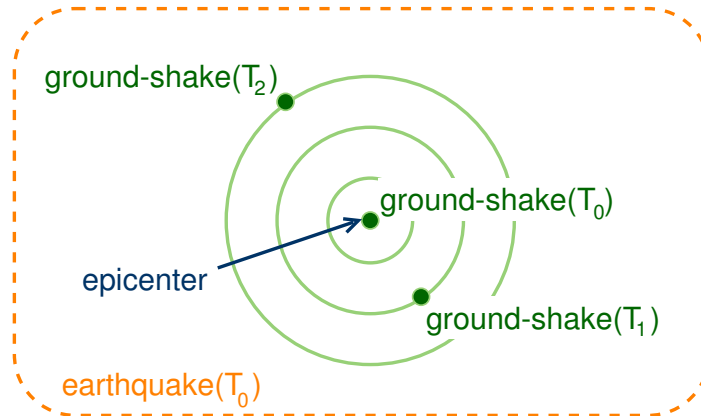


Figure 3.1: Web Event Model of an Earthquake

Within the Web, events lose their tight coupling to locations and retain only a time component. The event instances keep this information as descriptive metadata. A reactive system such as we

envision it, could detect these **ground-shake** events and react on behalf of each one of them. Because of the Web's latency these events do most certainly not arrive at systems within the Web in the original order, in which they were triggered. They also do most likely not arrive in the exact same order for all systems. This leaves us with time as the only important factor left, to distinguish events from each other in the first place. To get an earthquake event out of all these ground shaking informations floating through the Web, somebody would need a reactive system that detects these events and assembles them into one earthquake event, together with a computed epicenter and magnitude. Such a system (we call it **earthquake-tracker**) would own an earthquake model that allows it to decide whether a **ground-shake** event belongs to one physical earthquake or to another one, depending on its spatial location information and the intensity at that point in time. It could then emit a more complex **earthquake** event (with epicenter and magnitude) that allows other systems to interpret this physical event and react on behalf of it.

Let's take another system that reacts on a physical earthquake. It is now left with a multitude of different options on how to react. It could only react on the **earthquake** event which is coming from the system above (**earthquake-tracker**) that applies its earthquake model to the incoming **ground-shake** events. But how long will it take for this system to deploy its **earthquake** event? Eventually it waits for one round-trip of a seismic wave around the world, which takes approximately half an hour. What if it waits two or three round-trip times in order to collect more accurate data? And what if our new system wants to react as fast as possible in order to warn people all around the world. It would then need to react on a small subset of the **ground-shake** events in order to quickly identify a real earthquake and take measurements, e.g. immediately send out text messages to people, or to deploy yet another (this time **earthquake-alert**) event into the Web's information space. This relatively simple example discloses the complex nature of event-driven systems, but also their high flexibility and fine grained tuning possibilities.

3.2 The Web's Event & Action Information Space

For a conceptual model, the information space in which the events are triggered and the actions are invoked, needs to be identified. During our research we encountered many different event or action providing subsets of the Web that can be incorporated into our model:

- World-Wide Web
- Web of Things
- Services in the Web

We have already shown in chapter "Related Work", that there are basically two different ways how the Web's information space is accessible, i.e. either functionality and data have to be requested, or data is pushed through Webhooks. All of the above listed information space subsets require at least one of these two access methodologies. And through these access methodologies are we able to turn the information space of the Web into events and actions. The World-Wide Web [3], as envisioned by Tim Berners-Lee, is an information universe of interlinked documents, that a user can browse through. In our model, we can pull events directly from the World-Wide Web. For example, most documents in the World-Wide Web are subject to changes and such changes can be translated into events.

We gave an introduction into Services in the Web in chapter "Related Work"

Either there is an event producer which proactively pushes events into the Web, or a service is offered whose responses can be turned into events. These events and actions can have a solely virtual nature or, in the case of data from the "Web of Things", also a physical nature. A virtual nature can be anything from a static webpage to offered services, such as a detected change on a webpage can become an event as well as a service answer is interpreted as such for example a new mail arriving. A virtual action could be a babelibu. A physical nature for events could for example be measurements from a rain detector, an action could be a window shutting automatically.

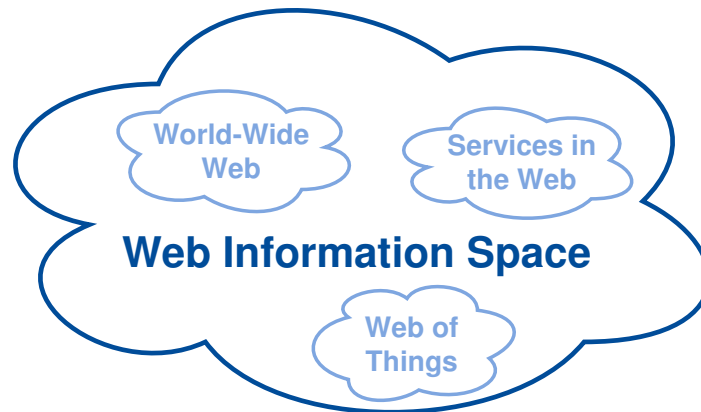


Figure 3.2: The Web's Information Space

There are different categories of events and also different ways how they can get into the Web:

We have seen in the related work chapter that there is a growing number of Web APIs which become accessible.

There are different categories in which we could

Actions

- Event Redirection
- Event Enrichment
- WebApp Actions

3.3 Impose Reactivity to the Web Information Space

In the last section we showed how mashups create additional value for the Web by combining several WebAPI's. But it turned out, that such mashups are closed systems, which often only allow little degree of parametrization. To get past such limitations and define a conceptual model for reactive Web systems, it is necessary to define a

existing rule languages, rule engines,

Existing ECA systems all act on local data. Looking at (Wikipedia...) their definition is actions on local data. This does only add reactivity to these systems and not to the Web per se.

Such systems are merely event sinks which add fairly any value to the Web, except for the individual users and the system itself.

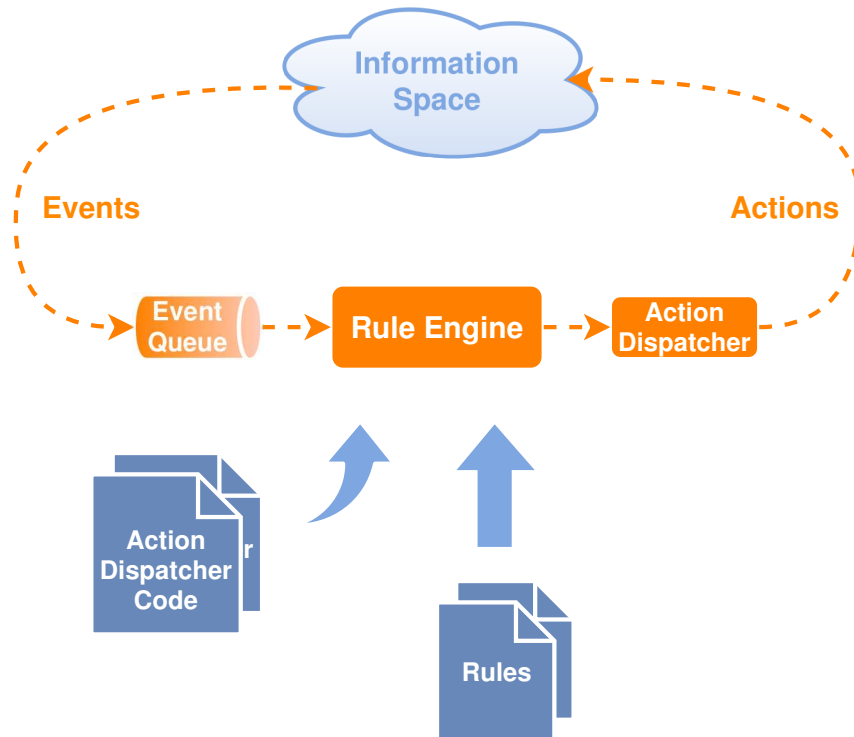


Figure 3.3: Conceptual Model for Reactive Information Systems

3.3.1 Conceptual Rule Language

Describe conceptual rule language ON (existing categories) IF (condition boundaries) DO (call to existing action modules with parameters)

a lot possible, but dangerous.

```

on mail
if sender="sender@mail.com"
do webapi->newcontent(subject)
  
```

Chapter 4

Applicability

We have introduced a conceptual model for reactive information systems and its services. In this chapter we will show how this model applies to certain use cases.

4.1 Augment existing Web Applications

Web applications, such as webmails, social networks or content management systems (CMS) are widely spread and used by a large number of internet users. But often users or developer miss some features or interoperability with other web applications, which would result in augmented functionality and also in less work. Features of that kind could also include data and functionality from other sites on the web. This would require web applications to mashup together and to grab data or invoke functionality on each other. Many kinds of augmentations will not be implemented by the web applications themselves because they are very specific to a small number of their users. With a reactive information system, users and developers could realize such features on their own.

4.1.1 Enrich CMS Posts with Additional Information

Every new post to a content management system can be modeled as an event. In the case that a user would like to enrich such a system with founded knowledge from a remote resource, reactivity in the Web can be used to enrich it. Augmenting an existing CMS can be realized by a rule that evaluates new posts and checks whether knowledge tags are included in the post. Whenever there are tags included within the post, the reactive system will enrich the post with additional knowledge to these tags from a remote service of the user's choice.

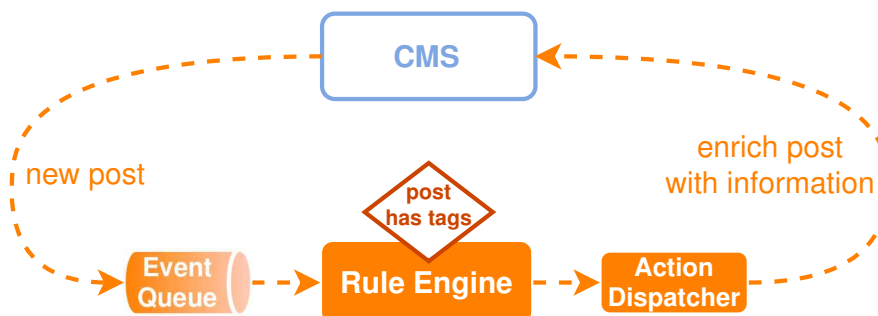


Figure 4.1: Enrich CMS Post with Remote Knowledge Data

4.1.2 Workflow Automation

Within such web applications, users often have very specific workflows. And because workflows always start with an event, they are predestinated to be automated by a reactive information system. As an example for workflow automation, course and student exercise submission administration can be taken care of by a reactive information system. Whenever a new semester starts, the reactive system will detect this through one of its rules and command an action dispatcher to set up infrastructures for courses. This can also include grabbing course data from an official web page and including it into the infrastructure, thus eliminating the need for manual data copy tasks.

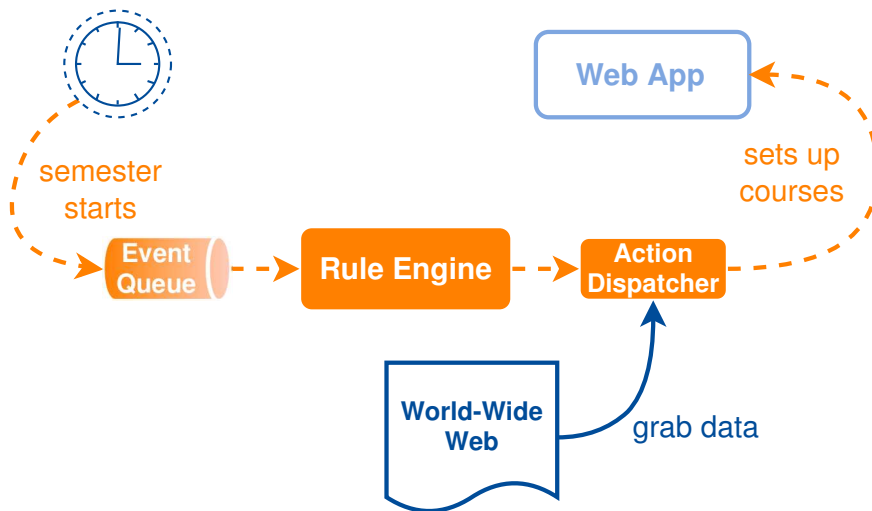


Figure 4.2: Create course resources at semester start

After setting up the semester courses, the reactive system is ready to process new student registrations for these courses. It automatically associates students into the afore mentioned infrastructure and sets up additional infrastructure such as an exercise submission container. Whenever a course tutor submits a new exercise to the course resource, the system will detect this and spread this information to the students, together with a deadline. The students are expected to submit their exercise solutions before the deadline to the exercise submission container that was created reactively for them.

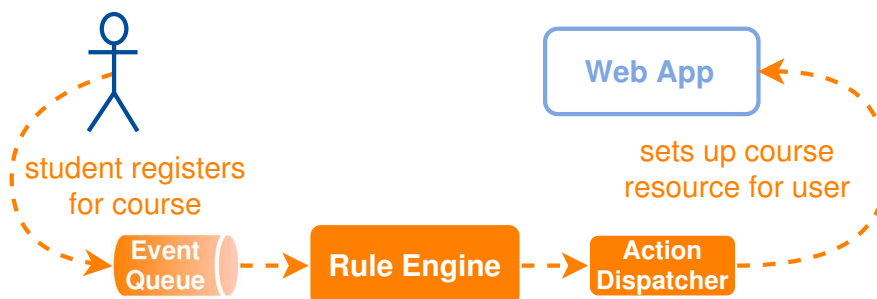


Figure 4.3: Create course resource for registered student

A certain amount of time before the deadline, e.g. one day, the reactive system will detect the deadline and process events that depict the current exercise submission status per student. If the system detects a student who hasn't uploaded her exercises yet, it will notify her about

the deadline. This is an additional service that gives students the chance to react on a missed exercise submission deadline. As soon as the deadline passed, the system will revoke write-rights to the exercise submission container and therefore disallow submissions which are too late.

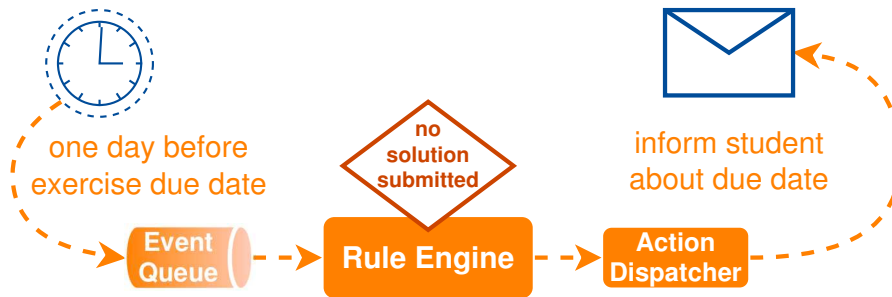


Figure 4.4: Notify student before exercise due date

4.2 Service Functionality and Availability Checking

Services offered through the Web aren't monitored or tested by users or developers from other sites. If they rely on correct functionality or availability they need a way to assert this. It is also possible that an owner of such a service doesn't have the tools to monitor his own services. Whenever such a service isn't working correctly anymore or stops responding, these users or developers need to be able to react on this before it is too late. With a reactive rule in place that evaluates Service Testing results, measurements can be taken early. One action to such a failing service test could be an automatic switching of the utilized service within an action dispatcher, so that from then on it uses one which still works correctly.

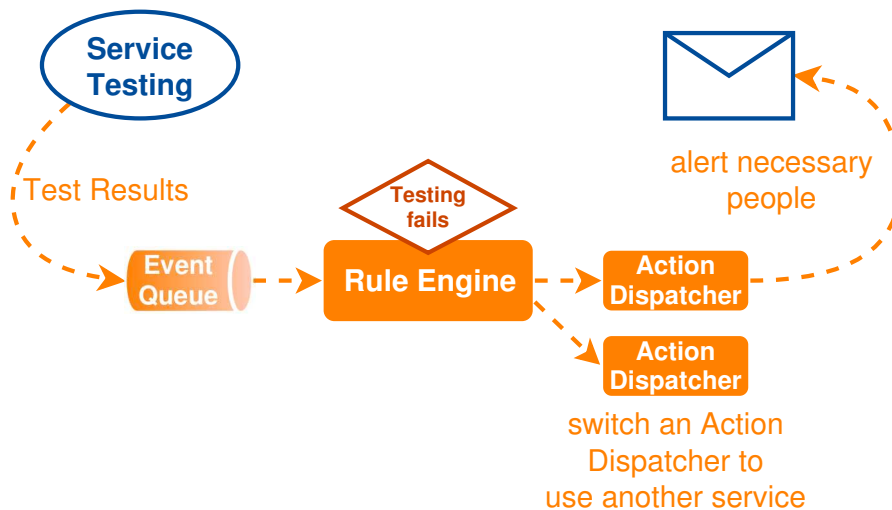


Figure 4.5: Test proper Service Functionality and Availability

4.3 Exploiting the Web of Things

A model for reactive information systems becomes also very interesting in the context of the Web of Things. Through the small connected devices, a lot of sensor data become accessible

via the Web and can be used as events to trigger actions. These actions could also be part of the Web of things, if there are Things that offer services. One example of a reactive rule, that has parts in the Web of Things, is that of a server room which has a defective cooling. The increasing temperature eventually causes the servers to shutdown or even fail. Servers in this room could push current state information into a reactive system. The reactive system could take measurements if it detects a certain pattern that will lead to an overheating of all systems. It could inform certain (not so important) servers to gracefully shutdown and additionally inform administrators, who otherwise might miss the shutdown. Even better would be if it would have the power to kick in an additional emergency cooling system to prevent the shutdown of any of the servers.

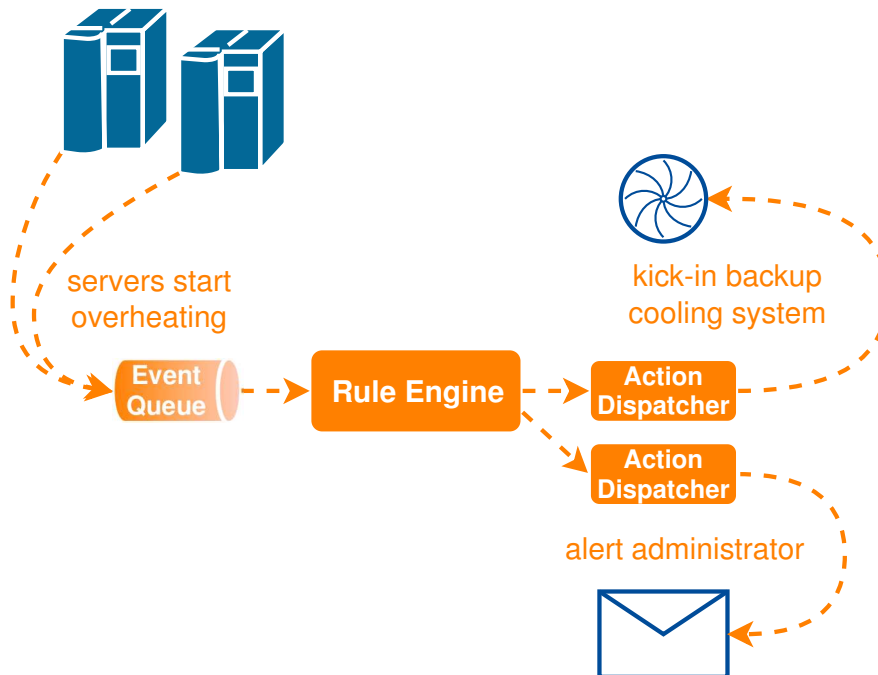


Figure 4.6: Measurements on Server Failure

Another scenario gets more realistic with the increasing number of homes that are connected to the Web. A home or apartment owner has his light controls attached to the Web. The first thing a reactive system could do, is that it detects holidays in the owner's agenda and automatically sets the light control to somewhat reasonable random during his absence. This would make suspicious characters, which are eventually interested in his wealth, think that he's still at home. In combination with another Thing that is connected to the Web and always accompanies people, the cell phone, an even more interesting application scenario can be thought of. The phone would push location information about the owner into the system. Whenever the owner gets close to his house, the reactive system could turn on the light in the entry area, pull up some nice music. On a wednesday evening it could also inform the delivery service that they can deliver the owner's preferred dish now, because the owner is doing this always on a wednesday evening.

Chapter 5

Prototype System

The prototype system is the realisation of a reactive Web system. It was developed during the research for this thesis and acts as a platform for feasibility studies of certain use cases.

Prototype consists of a queue in which all incoming events are pushed, and an engine that picks the events from the end of the queue whenever it is idle. Since Prototype's core functionality is the communication with resources in the Web, the architecture bases on HTTP protocol in several parts. For example the events are meant to be retrieved completely via HTTP, the user interface is a Webpage which posts requests to the system and most actions are also meant to be HTTP requests, or at least using them to gather information.

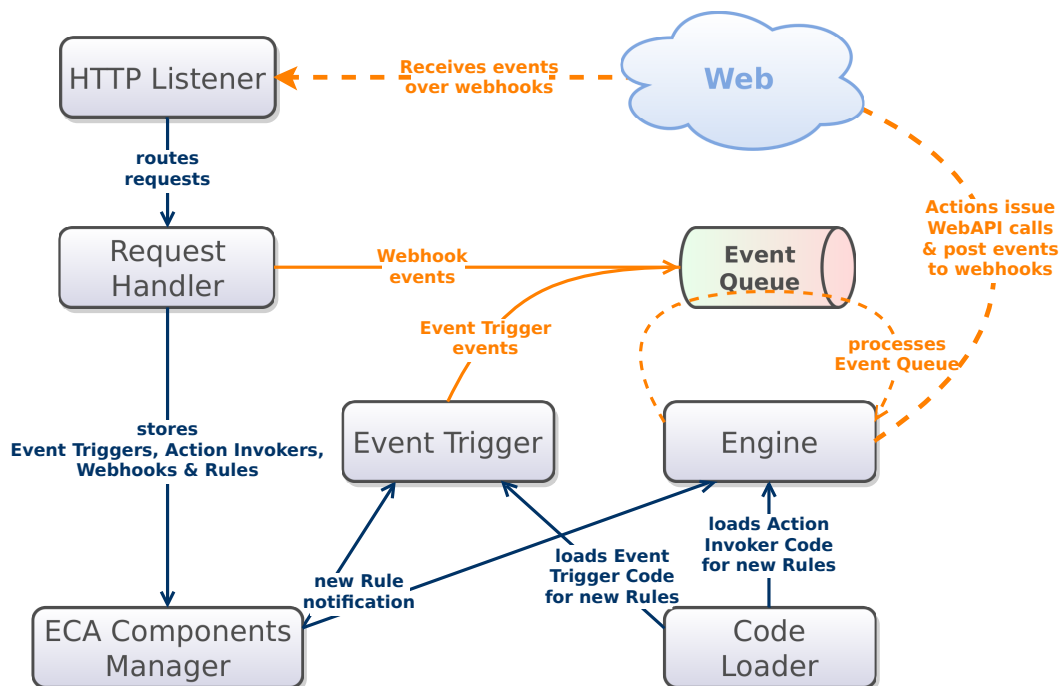


Figure 5.1: Prototype Process diagram

5.1 Event Trigger

Event Gathering is the E in ECA and without one of these letters such a system would not run. It is of utmost importance to find as much as possible ways to get data into a system.

5.1.1 Polling

5.1.2 Webhooks

5.2 Action Dispatcher

5.3 ECA Rules in the Engine

During our research we found a troublesome server room that shows how the Web of Things can be exposed through our model. This server room suffered from a defective cooling system which lead to a drastical increase of temperature in certain circumstances. As a consequence certain server automatically shut themselves down as safety measurements. Eventually, these shutdowns weren't detected immediately by the people that administered these servers, therefore unnecessary downtimes were the result. As a very quick fix to inform certain administrators about the shutdown of their server, we started pinging these servers and pushed the results int

5.4 Web Programming

5.4.1 Node.js

5.4.2 Callback Functions

5.4.3 Asynchronous Closures

Often, optimization approaches and programming language concepts require special attention to avoid common pitfalls. When closures are used as asynchronous functions, developers need to be very careful not to end up with race conditions.

Looking at an example of sequential code execution in Figure 5.2, we see that function execution of `fA` is halted until function `fB` is finished. If `fB` happens to be a latency-driven I/O operation the completion of `fA` could be deferred for a relatively long time. While the application waits for the completion of the I/O operation, some remaining operations in `fA` could eventually already be executed without causing any race conditions.

Asynchronous code execution, as shown in Figure 5.3, allows non-blocking and thus scalable applications. Non-blocking operations are a remedy for optimized resource allocation and open up ways to overcome previously described unnecessary resource bindings. Processing any kind of latency-driven I/O operation asynchronously (e.g. filesystem access and socket communication) exploits resources that would otherwise be bound while waiting for completion. Such operations are processed and completed whenever required resources are available.

Often other operations depend on the completion of asynchronous operations, hence their execution needs to be deferred. This necessary code execution deferral is achieved through the

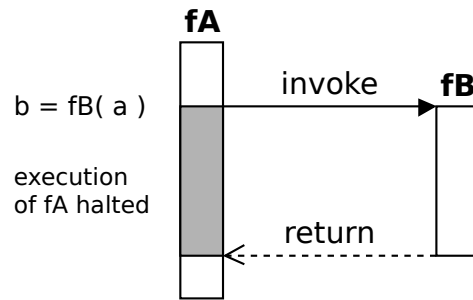


Figure 5.2: Synchronous Function Call

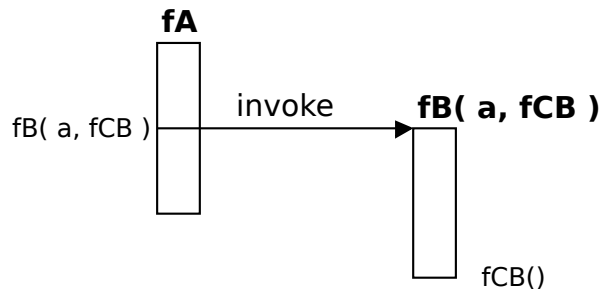


Figure 5.3: Asynchronous Function Call

use of callback functions, denoted fCB in Figure 5.3. Any code placed in a callback function, which is assigned to an asynchronous operation, is only executed after the respective asynchronous operation completed. This allows stacking of functions and operations upon each other which automatically results in a flexible and event-driven application.

Now we take closures into this asynchronous context, as defined in ECMAScript[10], which is the base for widely-spread script languages like JavaScript, JScript and ActionScript. Closures in ECMAScript[10] are defined such as they have access to the context of the function they were created in. This is shown in Figure 5.4 where c from fA 's context is accessible from within fB , assuming that fB was created in fA and not only invoked from there. Using asynchronous closures it becomes evident, that the context in the invoking function can change while the closure is still computing and eventually referencing the outer context, thus causing race conditions. This will be most obvious in a loop that immediately invokes fB several times, as shown in Figure 5.5. In such a setup c will have different values in the same part of different invocations of fB .

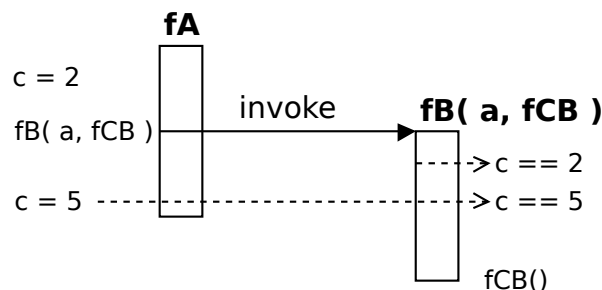


Figure 5.4: Closure Scope and referenced context

Those event-driven context overwrites can be taken care of by shielding the closure from context changes, as shown in Figure 5.6. To shield the closure from context changes, closure fB needs to create another closure fC and return it to fA . The argument passed to fB is the

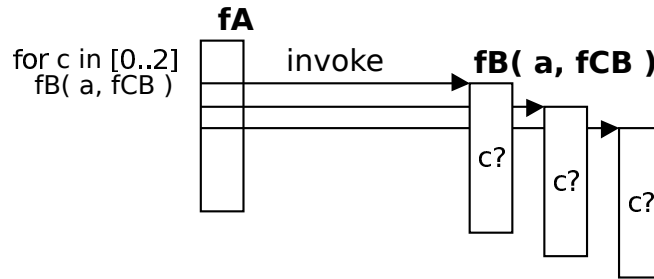


Figure 5.5: Closure context changes in a loop

context (c in Figure 5.6) that might change but requires to be persistent for one invocation. fC has now c as a fixed context, which can't be overwritten anymore. Now the only thing left is fC needs to be invoked and it will retain the original context. This implementation is necessary when the closure acts as a callback function for asynchronous operations, to preserve the original context in case it is required within the callback function.

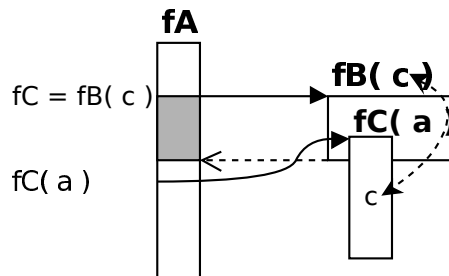


Figure 5.6: Closure context shielding

Benchmarking JavaScript vs. Java

refer to listings

Chapter 6

Discussion & Results

We have seen that the ECA approach is already a powerful one to make the Web reactive. A future improvement of this could be to adopt Complex Event Processing (CEP). This would mean that several events could be stored in a rule and be evaluated in terms of time constraints. Through this more complex events can be created as a result of several atomic events which would lead into semantically more complex events. A change in paradigm will result in an approach where events are not just processed when they are entering the system and evaluated against rules, but these events would need to be stored for quite a long time. Also the rules will not all be checked for each event but they are subject to a scheduler. It can be decided when and how often a rule is evaluated and all events will be checked at these point in times, whether they are candidates for firing the rule. A relational database will be needed in order to search through the timestamps

Bibliography

- [1] Raman Adaikkalavan and Sharma Chakravarthy. Event Specification and Processing for Advanced Applications: Generalization and Formalization. In Roland Wagner, Norman Revell, and Günther Pernul, editors, *Database and Expert Systems Applications*, volume 4653 of *Lecture Notes in Computer Science*, pages 369–379. Springer Berlin Heidelberg, 2007.
- [2] Tim Berners-Lee. Notation 3 Logic. <http://www.w3.org/DesignIssues/Notation3.html>, 2005. Accessed: 2013-10-21.
- [3] Tim Berners-Lee, Robert Cailliau, Jean-François Groff, and Bernd Pollermann, World-Wide Web: The Information Universe. *Electronic Networking: Research, Applications and Policy*, 1(2):74–82, 1992.
- [4] Andrew D. Birrell and Bruce Jay Nelson, Implementing Remote Procedure Calls. *ACM Trans. Comput. Syst.*, 2(1):39–59, February 1984.
- [5] Harold Boley. The RuleML family of web rule languages. In *Principles and Practice of Semantic Web Reasoning*, pages 1–17. Springer, 2006.
- [6] Don Box, David Ehnebuske, Gopal Kakivaya, Andrew Layman, Noah Mendelsohn, Henrik Frystyk Nielsen, Satish Thatte, and Dave Winer. Simple object access protocol (SOAP) 1.1, 2000.
- [7] Tim Bray, Jean Paoli, C Michael Sperberg-McQueen, Eve Maler, and François Yergeau, Extensible markup language (XML). *World Wide Web Consortium Recommendation REC-xml-19980210*. <http://www.w3.org/TR/1998/REC-xml-19980210>, 1998.
- [8] François Bry and Michael Eckert. Twelve Theses on Reactive Rules for the Web. In Torsten Grust, Hagen Höpfner, Arantza Illarramendi, Stefan Jablonski, Marco Mesiti, Sascha Müller, Paula-Lavinia Patranjan, Kai-Uwe Sattler, Myra Spiliopoulou, and Jef Wijsen, editors, *Current Trends in Database Technology – EDBT 2006*, volume 4254 of *Lecture Notes in Computer Science*, pages 842–854. Springer Berlin Heidelberg, 2006.
- [9] Erik Christensen, Francisco Curbera, Greg Meredith, Sanjiva Weerawarana, et al. Web services description language (WSDL) 1.1, 2001.
- [10] ECMA International. *Standard ECMA-262 - ECMAScript Language Specification*. 5.1 edition, June 2011.
- [11] Adrian Giurca and Emilian Pascalau, Json rules. *Proceedings of the Proceedings of 4th Knowledge Engineering and Software Engineering, KESE*, 425:7–18, 2008.
- [12] Object Management Group. *The Common Object Request Broker (CORBA): Architecture and Specification*. Object Management Group, 1995.

-
- [13] Dominique Guinard, Vlad Trifa, Friedemann Mattern, and Erik Wilde. From the Internet of Things to the Web of Things: Resource-oriented Architecture and Best Practices. In Dieter Uckelmann, Mark Harrison, and Florian Michahelles, editors, *Architecting the Internet of Things*, pages 97–129. Springer Berlin Heidelberg, 2011.
 - [14] Keman Huang, Yushun Fan, and Wei Tan. An Empirical Study of Programmable Web: A Network Analysis on a Service-Mashup System. In Carole A. Goble, Peter P. Chen, and Jia Zhang, editors, *ICWS*, pages 552–559. IEEE, 2012.
 - [15] Aparna Nagargadde, Sridhar Varadarajan, and Krithi Ramamritham. Semantic Characterization of Real World Events. In Lizhu Zhou, BengChin Ooi, and Xiaofeng Meng, editors, *Database Systems for Advanced Applications*, volume 3453 of *Lecture Notes in Computer Science*, pages 675–687. Springer Berlin Heidelberg, 2005.
 - [16] George Papamarkos, Alexandra Poullovassilis, Ra Poullovassilis, and Peter T. Wood. Event-Condition-Action Rule Languages for the Semantic Web. In *In: Workshop on Semantic Web and Databases*, pages 309–327, 2003.
 - [17] Adrian Paschke and Harold Boley. Rules Capturing Events and Reactivity. In Adrian Giurca, Dragan Gasevic, and Kuldar Taveter, editors, *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches*, pages 215–252. IGI Publishing, May 2009.
 - [18] Adrian Paschke, Harold Boley, Zhili Zhao, Kia Teymourian, and Tara Athan. Reaction RuleML 1.0: Standardized Semantic Reaction Rules. In Antonis Bikakis and Adrian Giurca, editors, *Rules on the Web: Research and Applications*, volume 7438 of *Lecture Notes in Computer Science*, pages 100–119. Springer Berlin Heidelberg, 2012.
 - [19] Paula-lavinia Patranjan. *The Language XChange*. PhD thesis, Ludwig-Maximilians-Universität München, 2005.
 - [20] C. Peltz, Web services orchestration and choreography. *Computer*, 36(10):46–52, Oct 2003.
 - [21] Randall Perrey and Mark Lycett. Service-oriented architecture. In *Applications and the Internet Workshops, 2003. Proceedings. 2003 Symposium on*, pages 116–119. IEEE, 2003.
 - [22] Adina Ploscar, XML-RPC vs. SOAP vs. REST web services in Java—uniform using WSWrapper. *Int. J. Comput*, 4:215–223, 2012.
 - [23] REVERSE - Reasoning on the Web with Rules and Semantics. <http://reverse.net>. Accessed: 2014-05-08.
 - [24] Keith Seymour, Hidemoto Nakada, Satoshi Matsuoka, Jack Dongarra, Craig Lee, and Henri Casanova. Overview of GridRPC: A remote procedure call API for Grid computing. In *Grid Computing—GRID 2002*, pages 274–278. Springer, 2002.
 - [25] P. Windley. *The Live Web: Building Event-Based Connections in the Cloud*. Cengage Learning PTR, 2011.

List of Figures

1.1	Users mashup the Web's Information and Data	2
2.1	Number of registered APIs in the ProgrammableWeb directory by date	3
2.2	REST vs. SOAP: Simplicity wins again (John Musser, ProgrammableWeb, 2011)	5
3.1	Web Event Model of an Earthquake	8
3.2	The Web's Information Space	10
3.3	Conceptual Model for Reactive Information Systems	11
4.1	Enrich CMS Post with Remote Knowledge Data	12
4.2	Create course resources at semester start	13
4.3	Create course resource for registered student	13
4.4	Notify student before exercise due date	14
4.5	Test proper Service Functionality and Availability	14
4.6	Measurements on Server Failure	15
5.1	Prototype Process diagram	16
5.2	Synchronous Function Call	18
5.3	Asynchronous Function Call	18
5.4	Closure Scope and referenced context	18
5.5	Closure context changes in a loop	19
5.6	Closure context shielding	19

Index

C

CMS, 12
CORBA, 4

E

Engine, 6

I

IDL, 4
Information Space, 9

J

JSON Rules, 7

K

KRL, 7

M

Mashups, 5

N

Notation3, 6

O

ORB, 4

R

RDF, 6
REST, 4
RPC, 3
Rule, 6
Rule Language, 6
RuleML, 7

S

Services, 3, 9
SOA, 3
SOAP, 4

W

Web API, 3
Web of Things, 1, 9, 14
Web Service, 3
World-Wide Web, 9
WSDL, 4

X

Xcerpt, 7
XChange, 7
XML, 4, 6
XML-RPC, 4

Appendices

Appendix A

Rule Languages

A.1 Example JSON Event for Rule Languages

```

1 {
2   "eventname": "email",
3   "body": {
4     "sender": "sender@mail.com",
5     "subject": "Important subject!",
6     "textbody": "Hi User,\n\nThis is a lengthy mail body"
7   }
8 }

```

A.2 E-Mail Example Rule expressed in RDF

```

1 ON INSERT document("inbound_queue.xml")/mails/mail
2 IF $delta/sender[.="sender@mail.com"]
3 DO DELETE document("inbound_queue.xml")/mails/mail;
4 LET $api = resource("www.webapi.com") IN
5 INSERT ($api, newcontent,
6   <content>New mail: {$delta/subject}</content>)

```

A.3 E-Mail Example Rule expressed in Notation 3

```

1 { ?x :event "email". ?x :sender "sender@mail.com" }
2 => { :webapi :newcontent ?x }

```

```

1 TRANSACTION
2   in {
3     resource { "http://www.webapi.com"},
4     newcontents {{
5       insert newcontent { var Mail }
6     }}
7   }
8 ON
9   xchange:event {{
10    xchange:sender { "http://mailserver.com" },
11    var Mail -> email {{
12      sender { "sender@mail.com" }
13    }}
14  }}
15 END

```

```

1 TRANSACTION
2   [...]
3 ON
4   [...]
5 FROM
6   in {
7     resource { "http://www.weather.com"},
8     temperatures {{
9       var T -> temperature {{
10        datetime { "2013-10-20-08:00:00 AM" }
11      }}
12    }}
13  }
14 END

```

A.4 E-Mail Example Rule expressed in XChange/Xcerpt

A.5 XChange/Xcerpt Remote Resource Access

A.6 E-Mail Example Rule expressed in JSON Rules

```

1 {
2   "id": 0,
3   "conditions": [
4     {
5       "type": "email",
6       "constraints": [
7         {
8           "propertyName": "sender",
9           "operator": "EQ",
10          "restriction": {
11            "type": "String",
12            "value": "sender@mail.com"
13          }
14        },
15        {
16          "bind": "$S",
17          "propertyName": "subject"
18        }
19      ]
20    }
21  ]
22 }

```

```

18         }
19     ]
20 }
21 ],
22 "actions": [
23     "webapi('addcontent', $S)"
24 ]
25 }

```

A.7 E-Mail Example Rule expressed in Kinetics Rule Language (KRL)

```

1 rule store_mail {
2     select when mail newmail
3     sender re#sender@mail.com#
4     subject re### setting(subj)
5     http:post("http://www.webapi.com/newcontent")
6     with params = {
7         "text": subj
8     }
9 }

```

Listing A.1: E-Mail Example rule in KRL

A.8 E-Mail Example Rule expressed in (Reaction) RuleML

```

1 <Rule style="active">
2   <on>
3     <Event>
4       <Atom>
5         <Rel per="value">mail</Rel>
6         <Var>sender</Var>
7         <Var>subject</Var>
8       </Atom>
9     </Event>
10  </on>
11  <if>
12    <Atom>
13      <op><Rel>equals</Rel></op>
14      <Var>sender</Var>
15      <Ind>sender@mail.com</Ind>
16    </Atom>
17  </if>
18  <do>
19    <Atom>
20      <oid><Ind uri="http://webapi.com"/></oid>
21      <Rel>newcontent</Rel>
22      <Var>subject</Var>
23    </Atom>
24  </do>
25 </Rule>

```


A.9 Prototype Rule transformed into JSON

```
1 {  
2   "event": "mail",  
3   "conditions": [  
4     { "sender": "sender@mail.com" },  
5   ],  
6   "actions": [  
7     {  
8       "api": "webapi",  
9       "method": "newcontent",  
10      "arguments": {  
11        "text": "$X.subject"  
12      }  
13    }  
14  ]  
15 }
```

Appendix B

Rules

B.1 Binder Annotations

B.1.1 Binder Annotations

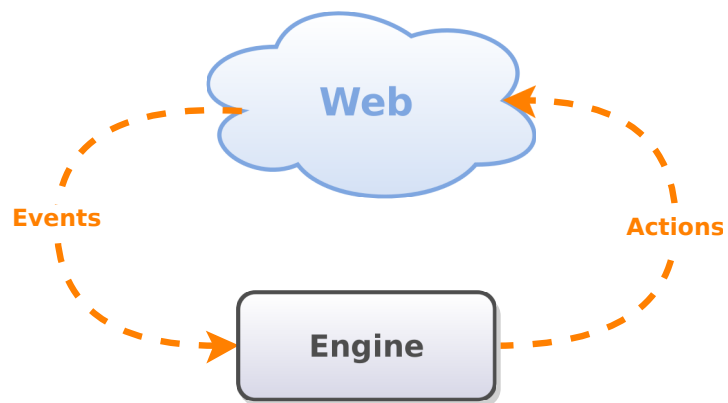


Figure B.1: SHOULD NOT SHOW UP

Appendix C

Benchmarking

C.1 Java

```

1  /*
2   * BenchmarkingDeferred.java
3   */
4  import java.util.concurrent.ScheduledExecutorService;
5  import java.util.concurrent.Executors;
6  import java.util.concurrent.TimeUnit;
7  import java.util.HashMap;
8
9  public class BenchmarkingDeferred {
10
11     private static Runtime runtime = Runtime.getRuntime();
12     private static final ScheduledExecutorService worker =
13         Executors.newSingleThreadScheduledExecutor();
14
15     private static void deferFunctionCall( int numScopeVars, int delay,
16         String scopeId ) {
17         HashMap<String, String> mapVars = new HashMap<String, String>();
18         for( int i = 0; i < numScopeVars; i++ ) {
19             mapVars.put( "id" + i, "12345678" ); // 8 bytes per stored scope
20             variable
21         }
22         Object context = new TimeoutContext( "TimeoutFunction" );
23         Runnable task = new RunnableCallbackFunction( mapVars, context );
24         worker.schedule( task, delay, TimeUnit.SECONDS );
25     }
26
27     public static void main( String[] args ) {
28         long startTime, stopTime;
29         int numVars = 10, firstArg = 0;
30         firstArg = Integer.parseInt( args[0] );
31         numVars = Integer.parseInt( args[1] );
32         int j = 0, numFuncs = 1 << firstArg;
33
34         startTime = System.nanoTime();
35         while( j++ < numFuncs ) {
36             deferFunctionCall( numVars, numFuncs * 10, numFuncs + "(" + j + ")"
37                 );
38         }
39         stopTime = System.nanoTime();
40     }
41 }

```

```

38     // [...] benchmark system out
39
40     worker.shutdownNow();
41 }
42 }
43
44 /*
45  * RunnableCallbackFunction.java
46  */
47 import java.util.HashMap;
48
49 /*
50  * The Callback function instance.
51  */
52 public class RunnableCallbackFunction implements Runnable {
53
54     // The hashhmap is used to store variables and their value as the scope
55     private HashMap<String, String> mapScope;
56     private Object context;
57
58     public RunnableCallbackFunction( HashMap<String, String> scope, Object
59         context ) {
60         this.mapScope = scope;
61         this.context = context;
62     }
63
64     // If this is executing, we didn't wait long enough and the
65     // benchmark time is compromised
66     public void run() {
67         System.out.println( mapScope.toString() );
68     }
69 }
70
71 /*
72  * TimeoutContext.java
73  */
74 public class TimeoutContext {
75     private long idleTimeout = 1;
76     private long idlePrev;
77     private long idleNext;
78     private long idleStart = 140000505;
79     private String onTimeout = null;
80     private boolean repeat = false;
81
82     public TimeoutContext( String cb ) {
83         this.onTimeout = cb;
84     }
85 }

```

Listing C.1: Closure Benchmarking: Java Code

C.2 JavaScript

```
1  /*
2  The function deferral measurements in node.js
3  */
4
5  var deferredFunction = function ( numScopeVars, delay, scopeId ) {
6      var scope = {};
7      for ( var i = 0; i < numScopeVars; i++ ) {
8          scope[ "id" + i ] = "12345678"; // 8 bytes per stored scope variable
9      }
10     setTimeout( function () {
11         // If this is executed we didn't wait long enough
12         console.log( JSON.stringify( scope, null, ' ' ) );
13     }, delay );
14 }
15
16 var numOfFunctions,
17     numOfScopeVars = process.argv[ 3 ];
18
19 numOfFunctions = Math.pow( 2, process.argv[ 2 ] );
20
21 var time = process.hrtime();
22 for (var i = 0; i < numOfFunctions; i++) {
23     deferredFunction( numOfScopeVars, 1000 * numOfFunctions, numOfFunctions
24         + "(" + i + ")" );
25 };
26 var diff = process.hrtime( time );
27
28 var mem = process.memoryUsage();
29 // [...] benchmark system out
process.exit( 0 );
```

Listing C.2: Closure Benchmarking: JavaScript Code

UNIVERSITÄT BASEL

PHILOSOPHISCH-NATURWISSENSCHAFTLICHE FAKULTÄT

Declaration on Scientific Integrity

(including a Declaration on Plagiarism and Fraud)

Bachelor's / Master's Thesis (*Please cross out what does not apply*)

Title of Thesis (*Please print in capital letters*):

First Name, Surname (*Please print in capital letters*): _____

Matriculation No.: _____

I hereby declare that this submission is my own work and that I have fully acknowledged the assistance received in completing this work and that it contains no material that has not been formally acknowledged.

I have mentioned all source materials used and have cited these in accordance with recognised scientific rules.

In addition to this declaration, I am submitting a separate agreement regarding the publication of or public access to this work.

☐ Yes ☐ No

Place, Date: _____

Signature: _____

Please enclose a completed and signed copy of this declaration in your Bachelor's or Master's thesis .

