

Serialisierung von Datenobjekten in JSON zur Übertragung von Objekten aus Energieanwendungen

PROJEKTARBEIT

für die Prüfung zum

Bachelor of Science

des Studienganges Angewandte Informatik

an der

Dualen Hochschule Baden-Württemberg Karlsruhe

von

Sebastian Rieger

Abgabedatum 15. September 2014

Bearbeitungszeitraum	13 Wochen
Matrikelnummer	7406886
Kurs	TINF12B1
Ausbildungsfirma	Karlsruher Institut für Technologie (KIT) Karlsruhe
Betreuer der Ausbildungsfirma	Dr.-Ing. Karl-Uwe Stucky

Erklärung

Gemäß §16 (3) der „Studien- und Prüfungsordnung für den Studienbereich Technik“ vom 1.11.2007.

Ich habe die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Ort Datum

Unterschrift

Inhaltsverzeichnis

1	Einleitung	4
1.1	Generisches Managementsystem für Energiedaten	4
1.2	Objektorientiertes Datenmodell	4
1.3	Strukturelle Metadaten	4
2	Aufgabenstellung	5
3	JavaScript Object Notation	6
3.1	Der Aufbau von JSON	6
3.2	JSON in Verbindung mit Programmiersprachen	6
3.2.1	JSON und JavaScript	6
4	Objektserialisierung nach JSON	8
4.1	Was ist Serialisierung	8
4.2	Möglichkeiten der JSON-Serialisierung in Java	8
4.3	Eigener Ansatz	8
4.4	Flexjson	8
4.5	Jackson	8
4.6	Auswertung der Möglichkeiten	9
4.7	Fragestellungen nach der Besprechung	9
5	Jackson und das Jackson Projekt	10
5.1	Jackson-Module	10
5.2	Serialisierung mit Jackson	10
5.3	Deserialisierung mit Jackson	11
5.4	Klassendiagramm der Serialisierung	11
5.5	Auffälligkeiten beim Testen	12
6	Der SMD-Assistent	14
7	Abkürzungsverzeichnis	15

1 Einleitung

Die folgende Arbeit befasst sich mit der Serialisierung von Datenobjekten in JSON zur Übertragung von Objektdaten.

1.1 Generisches Managementsystem für Energiedaten

Das *generisches Managementsystem für Energiedaten* (GDS) verwendet Objekte um Anwendungsdaten zu verwalten, die einem *objektorientierten Programmiermodell* (OPM) genügen und durch *strukturelle Metadaten* (SMD) beschrieben werden können. Das GDS wird zur Zeit im *Institut für Angewandte Informatik* (IAI) des *Karlsruher Instituts für Technologie* (KIT) entwickelt und ist ein System generischer Datenservices welches bei Release vollautomatisch Energiedaten managen können soll.

Durch den generischen Charakter kann es aber auch Daten aus anderen Bereichen verwalten, wenn diese dem OPM-Modell genügen.

1.2 Objektorientiertes Datenmodell

Im OPM werden Richtlinien für die Entwicklung von allen objektorientierten Softwarebausteinen festgelegt. Das gesamte Projekt ist bisher OPM-konform gehalten und somit soll auch die Schnittstelle der Serialisierung OPM-konform gestaltet werden.

Im Folgenden sind die wichtigsten OPM-Regeln dargestellt:

- Basisklasse `OPMObject` von der alle Klassen Erben
- Es gibt keine Konstanten
- Attribute sind Grundsätzlich `private` und werden durch Getter- und Setter-Methoden aufgerufen
- Programme bestehen nur aus Objekten, der Aufruf erfolgt ausschließlich Über Methodenaufrufe
- Bezeichner werden im „Camel Case“ formuliert
- In der Dokumentation eines Attributs wird immer der erlaubte Wertebereich spezifiziert
- Die Dokumentation muss den Spezifikationen der verwendete Sprache entsprechen
- Jede Klasse enthält einen der Stati `Valid`, `Experimental` oder `Depricated` in ihrer Dokumentation

Alle Klassen die vom GDS verwaltet werden sollen, müssen diesen Grundsätzen genügen um verarbeitet werden zu können.

1.3 Strukturelle Metadaten

Strukturelle Metadaten sind laut der OPM-Definition spezielle Metadaten, die den Aufbau einer Programmklasse enthalten. Die Informationen der SMD sind programmiersprachenunabhängig und stehen dem Anwendungsprogrammen zur Verfügung.

Die Metadaten werden in einer SQL-Datenbank gespeichert, wo sie wenn benötigt geladen werden können. [Zil14]

2 Aufgabenstellung

In einem GDS-System werden Anwendungsdaten mit Hilfe von Objekten verwaltet, welche den Regeln von OPM folgen und mittels der SMDs beschrieben werden können.

Für den Transfer von solchen Anwendungsdaten, sind Schnittstellen für die automatische Serialisierung der Daten vorgesehen, welche im Rahmen dieser und einer weiteren Arbeit entwickelt werden sollen. [Wal14]

Ziel der Arbeit ist es, Spannungszeitreihen und OPM-Objekte mittels der Metadaten in ein JSON-Format zu Serialisieren und diese zu Übertragen. Die Empfängerseite muss letztendlich in der Lage sein aus den übertragenen Daten wieder Objekte zu erstellen.

Abschließend soll ein Vergleich zeigen welche der Serialisierungsarten (JSON oder XML) besser geeignet ist. Hierfür wird die Arbeit „Serialisierung von Datenobjekten in XML zur Übertragung von Objekten aus Energieanwendungen“ von Herrn Achim Walz herangezogen. [Wal14]

3 JavaScript Object Notation

JavaScript Object Notation (JSON) ist ein Textbasiertes Format zum Datenaustausch, wobei jedes gültige JSON-Dokument auch ein gültiges JavaScript ist. Es wurde als Ersatz für XML geschaffen und wird hauptsächlich in Bereichen eingesetzt wo Ressourcen wie Speicherplatz, Prozessorleistung und Netzwerkverbindung stark limitiert sind. Im Aufbau erinnert JSON an die Struktur eines Arrays. Ein Beispiel für ein JSON-Objekt ist im Kapitel 3.1 zu finden. [Wik14a]

3.1 Der Aufbau von JSON

Ein Beispiel für ein gültiges JSON-Dokument ist im Beispiel unten zu finden. In der ersten und letzten Zeile sind geschweifte Klammern zu finden, da jedes JSON-Dokument ein Objekt ist und Objekte in JSON von geschweiften Klammern umschlossen werden müssen. [Sri13]

JSON ist nach dem Schlüssel/Wert Prinzip aufgebaut was bedeutet, dass jedem Schlüssel genau ein Wert zugeordnet werden kann. Im Beispiel sind alle in JSON möglichen Formattypen aufgezeigt.

Zeile zwei enthält einen String, eine Zeichenkette in der jedes Zeichen erlaubt ist. Der boolsche Wert wird genau wie ein Nullwert ohne Anführungszeichen geschrieben wie in den Zeilen drei und vier gezeigt.

Zahlen können Ganzzahlig, Fließkommazahlen oder Exponentialzahlen sein wie sie auch im Beispiel zu finden sind.

Ein Array kann mehrere Werte enthalten und wird deshalb von eckigen Klammern umschlossen. Die eigentlichen Werte im Array müssen jedoch vom selben Typ sein.

Objekte können wiederum Objekte enthalten, wie es in den Zeilen acht bis zehn dargestellt ist. Das innere Objekt wird wieder von geschweiften Klammern umschlossen.

Somit können sechs Datentypen in JSON Unterschieden werden Strings, Zahlen, Booleans, Arrays, Objekte und Nullwerte. Zu beachten ist das Booleans, Nullwerte und Zahlen ohne Anführungszeichen beschrieben werden.

```
1 {  
2     "String": "Variable123",  
3     "Boolean": true;  
4     "Nullwert": null,  
5     "Zahl": 1234567  
6     "Fließkomma": 1.23e+6,  
7     "Array": ["Beispiel", "fuer", "Array"],  
8     "Objekt": {  
9         "Zahl_2": "1234",  
10    }  
11 }
```

3.2 JSON in Verbindung mit Programmiersprachen

Viele Programmiersprachen wie PHP, Python, C#, C++ und Java unterstützen JSON sehr gut und sogar Nativ.

3.2.1 JSON und JavaScript

JSON wird unter JavaScript als ganz normale Variable geführt und kann auch als solche ausgelesen werden. Dies geschieht beispielhaft über das Kommando `alert(JSONVariablenName.Zahl);`

liefert den Wert 1234567 aus dem Beispiel, unter der Bedingung das, dass JSON-Objekt als Variable `JSONVariablenName` deklariert wurde, zurück.

4 Objektserialisierung nach JSON

Wie in der Aufgabenstellung im Kapitel 2 vorgegeben, sollen hier nun die Möglichkeiten einer Serialisierung von Java-Objekten in JSON untersucht werden.

4.1 Was ist Serialisierung

Serialisierung ist die Abbildung von Daten auf eine geeignete Darstellungsform und wird oft bei verteilten Softwarelösungen wie im Falle von GDS verwendet. Der erzeugte Datenstrom kann dann entweder über ein Netzwerk übertragen oder lokal gespeichert werden. Somit liegt das Objekt doppelt vor, zum einen als reales Objekt eines Programms und als serialisiertes Objekt. Eine Änderung des Objekts im Programm hat somit keine Auswirkung auf das serialisierte Objekt. [Wik14b]

Im Rahmen dieser Arbeit heißt das OPM konforme strukturierte Java-Objekte in einen JSON-Datenstrom zu wandeln.

4.2 Möglichkeiten der JSON-Serialisierung in Java

Grundsätzlich gibt es verschiedene Möglichkeiten eine JSON-Serialisierung in Java durchzuführen. Im folgenden werden die im Projektteam diskutierten Möglichkeiten genauer vorgestellt.

4.3 Eigener Ansatz

Eine Möglichkeit einen Funktionsfähigen Serialisierer zu erhalten, ist diesen selber zu schreiben. Hierfür müsse eine Lesefunktion für JSON-Objekte implementiert werden, was auch als Scanner bezeichnet wird.

Dieser Scanner muss in der Lage sein einen JSON-Datenstrom zu lesen und ihn in die einzelnen Bestandteile aufspalten.

Eine weitere Funktion die erfüllt werden muss, ist die eines Parsers. Dieser muss die einzelnen vom Scanner erkannten Bestandteile in Javaobjekte umwandeln.

Bei der Implementierung muss des weiteren zum Beispiel auf Rekursion und nicht valide JSON-Objekte geachtet werden.

4.4 Flexjson

Flexjson ist eine einfache Bibliothek für das Serialisieren und Deserialisieren von JSON-Objekten in Javaobjekte.

Wenn Attributnamen in JSON von dem Deklarationsnamen im Javaobjekt abweichen sollen, müssen Annotationen verwendet werden.

Nachteilig ist das beim Serialisieren immer explizit angegeben werden muss wenn geschachtelte Objekte mit serialisiert werden sollen.

4.5 Jackson

Jackson ist ähnlich wie Flexjson eine Bibliothek für die Serialisierung von Javaobjekten zu JSON-Objekten. Vorteilhaft an Jackson ist das die Bibliothek modular aufgebaut ist. So benötigt man für einzelne Aufgaben nicht die gesamte Bibliothek.

Ein weiterer Vorteil ist, dass es über ein Jackson-Modul möglich ist Annotationen von der *Java Architecture for XML Binding* (JAXB) zu verwenden. Was es ermöglicht für die JAXB und Jackson Serialisierung die selben Annotationen zu nutzen. Dies würde die Bearbeitung und die Übersichtlichkeit sehr vereinfachen, da nicht zwei unterschiedlichen Annotationen benutzt werden müssen.

4.6 Auswertung der Möglichkeiten

In einer Projektgruppenberatung wurden alle drei Vorschläge ausführlich erläutert und diskutiert.

Vorteil des eigenen Ansatzes ist es zum einen, dass man hier völlig frei von fremden Bibliotheken ist und man keine zusätzlichen „technischen Schulden“ aufnehmen muss. Zum anderen ist es sehr Aufwändig eigen Klassen für die Serialisierung und Deserialisierung zu schreiben und dies ist im Rahmen dieser Arbeit leider nicht möglich.

Aus diesem Grund schied eigene Ansatz recht früh aus.

Der zweite Ansatz über die Flexjson Bibliothek ist interessant, und wurde in der Gruppe lange diskutiert. Denn hiermit ist es möglich expliziert anzugeben welche Attribute serialisiert werden sollen, und welche nicht, wie es später einmal vorgesehen werden soll.

Im letzten Ansatz mit Jackson ist es wie bei JAXB möglich den Serialisierer über Annotationen zu steuern. Vorteilhaft ist es vor allem, dass der Jackson und JAXB Serialisierer die selben Annotationen managen kann.

Dies war am Schluss auch das entscheidende Element warum sich für eine Umsetzung mit Jackson und JAXB entschieden wurde.

Diese Arbeit behandelt jedoch nur die Jackson beziehungsweise JSON Verarbeitung. In einer anderen Arbeit die Zeitgleich entstand, ist die Verarbeitung mit JAXB und XML zu finden. [Wal14]

4.7 Fragestellungen nach der Besprechung

Wie schon erwähnt sollen nicht immer alle Attribute serialisiert werden. Ob, und wies mit Jackson möglich ist, wird im weiteren Verlauf der Arbeit geklärt.

Des weiteren kam die Frage auf, ob es möglich ist Klassenattribute gesondert zu Serialisieren und gegebenen Falls zu untersuchen wie sich dies auf die Serialisierungs- beziehungsweise Deserialisierungsgeschwindigkeit auswirkt.

5 Jackson und das Jackson Projekt

Das Jackson Projekt entwickelt eine freie und modulare Bibliothek für die Serialisierung und Deserialisierung von Java-Instanzen in JSON-Dokumente. Jackson wird unter der Contributor License Agreement (CLA) vermarktet. Die zur Zeit aktuelle Version ist 2.4.1, welche auch bei der Bearbeitung des Projektes eingesetzt wird.

5.1 Jackson-Module

Die Jackson-Bibliothek besteht aus drei Hauptmodulen, welche wie folgt bezeichnet sind:

- „jackson-core“ welches die JSON spezifische Implementierung sowie eine low-level streaming API enthält
- „jackson-annotations“ welches die Jackson spezifischen Annotationen enthält.
- „jackson-databind“ welches für das *databind* verantwortlich ist.

Unter Databind wird eine Methode verstanden, welche über ein User-Interface gesteuert werden kann. Diese Methode ist in der Lage Daten aus einem Datenstrom wie zum Beispiel einem JSON-File zu lesen oder zu schreiben.

Mit diesen drei Modulen ist Jackson voll einsetzbar und kann Java-Instanzen zu einem JSON-Datenstrom umwandeln. Der JSON-Datenstrom wiederum kann gespeichert oder an andere Programme gesendet werden.

Um jedoch einheitliche Annotationen für Jackson und JAXB zu haben, wird ein weiteres Jackson-Modul benötigt, welches in der Lage ist die JAXB-Annotationen zu verarbeiten.[Jac14]

5.2 Serialisierung mit Jackson

Um eine Serialisierung mit Jackson umzusetzen wird zuerst eine Instanz der Klasse `ObjectMapper` benötigt, welche den Databinder darstellt. Der `mapper` ist somit für die Convertierung von Java-Instanzen zu JSON-Dokumenten verantwortlich.

Jedoch wird nicht nur der „Converter“ benötigt, sondern auch ein `AnnotationInspector`. Der `inspector` wird als Instanz von `JaxbAnnotationInspector` erstellt, welchem eine `TypeFactory` mit „Default-Einstellungen“ übergeben wird. Dies bedeutet es wird auf die Original JAXB-Annotationen geparkt, ohne auf Sonderfälle zu achten. Andere Annotation werden nicht berücksichtigt. Der `inspector` wird nun dem „Converter“ übergeben, damit dieser auf die entsprechenden Annotationen reagieren kann.

Um eine *Minimale Exception Safety* zu garantieren wird nun eine `Null`-Abfrage des zu serialisierenden Elements gemacht. Mit dieser Stufe der Sicherheit soll nicht verhindert werden das eine Exception passiert. Es wird lediglich garantiert das die Methode ohne Abzustürzen durchlaufen werden kann. [Gri02]

Ist die zu serialisierende Instanz `Null` so wird eine `IllegalArgumentException` generiert und die Methode so ordnungsgemäß beendet. Ist eine Instanz vorhanden, wird diese dem `mapper` übergeben. Das Ergebnis des Aufrufs von `writeValueAsString` ist entweder bei Erfolg ein valider JSON-String oder beim scheitern eine `JsonProcessingException`.

Dieser Zusammenhang ist noch einmal im folgenden Quellcode-Beispiel beschrieben.

```

1 ObjectMapper mapper = new ObjectMapper();
2 AnnotationIntrospector inspector = new JaxbAnnotationIntrospector(TypeFactory.
    defaultInstance());
3 mapper.setAnnotationIntrospector(inspector);
4
5 if (opmObject == null) {
6     throw new IllegalArgumentException("OPMObject can not be null!");
7 }
8 try {
9     return mapper.writeValueAsString(opmObject);
10 } catch (JsonProcessingException e) {
11     e.printStackTrace();
12     return null;
13 }

```

5.3 Deserialisierung mit Jackson

Für die Deserialisierung mit Hilfe von Jackson wird wie bei der Serialisierung ebenfalls ein Databinder und AnnotationInspector benötigt, welche wie im Kapitel 5.2 erstellt werden.

Bevor dies jedoch passiert, wird geprüft ob der eingegebene String weder Null noch Empty ist. Sollte das der Fall sein, wird die Methode `readValue` mit dem übergebenden String und der Information um welche Klassen-Instanz es sich beim String handelt übergeben.

Die Schwierigkeit beim Deserialisieren besteht also nun darin, das bevor der String überhaupt deserialisiert werden kann erst festgestellt werden muss um welche Klasse es sich eigentlich handelt.

Im Codebeispiel unten wird momentan noch davon ausgegangen, das es sich immer um eine Instanz der Klasse „TestData“ handelt. Wie diese Einschränkung aufgehoben werden kann wird im folgenden Kapitel beschrieben.

```

1 public <T extends OPMObject> T deserialize(String string) {
2     if (string == null) {
3         throw new IllegalArgumentException("String can not be null!");
4     }
5     if (string.isEmpty()) {
6         throw new IllegalArgumentException("String can not be empty!");
7     }
8
9     ObjectMapper mapper = new ObjectMapper();
10    AnnotationIntrospector inspector = new JaxbAnnotationIntrospector(
        TypeFactory.defaultInstance());
11    mapper.setAnnotationIntrospector(inspector);
12    // mapper.enableDefaultTyping();
13    try {
14        return (T) (mapper.readValue(string, Class.forName("opm_serializer.
            TestData")));
15    } catch (IOException | ClassNotFoundException | ClassCastException e) {
16        e.printStackTrace();
17    }
18    return null;
19 }

```

5.4 Klassendiagramm der Serialisierung

Wie von OPM verlangt erben hier alle Klassen von OPMObject. Um diese Arbeit mit der von Herrn Achim Walz vergleichen zu können wurde sich auf eine gemeinsame abstrakte Klasse `Serializer` geeinigt.

Die Klasse `JSONSerializer` erbt um einen Direkten vergleichen durchführen zu können genau wie `XMLSerializer` von `Serializer`. `Testdata` und `TestData2OPM` sind erste Test-Klassen von denen Instanzen serialisiert und deserialisiert werden.

Main-Klasse in diesem Projekt ist `OPM_Serializer`. Die `main`-Methode setzt die Serialisierung im Test in Gang.

Damit wie gewünscht jede Klasse serialisiert werden kann, wurde die Methode `serializeMe` zu `OPMObject` hinzugefügt. Diese Methode nutzt nun bei Aufruf die `serialize`-Methode des jeweiligen Serialisierers.

Ein vollständiger Überblick ist im Klassendiagramm unten zu finden gegeben.

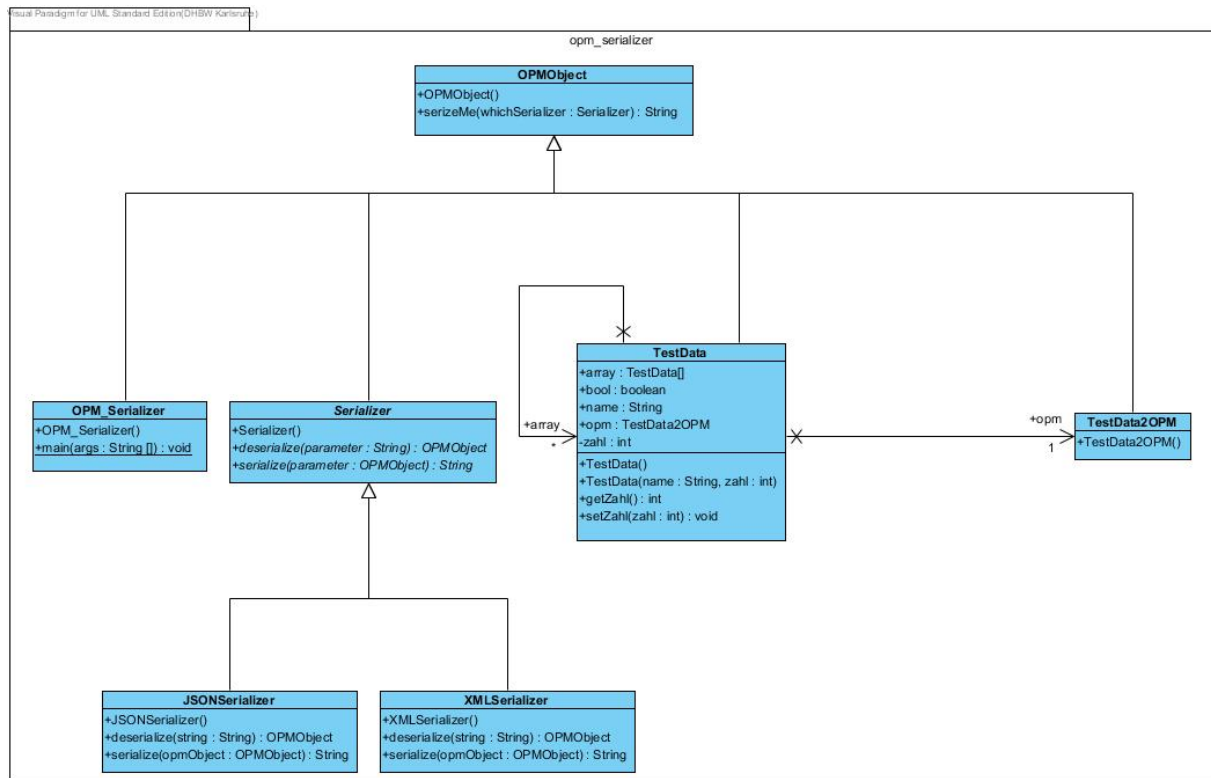


Abbildung 1: Klassendiagramm der Serialisierung

5.5 Auffälligkeiten beim Testen

Beim Serialisieren und Deserialisieren sollen Attribute in JSON-Dokumenten gespeichert werden. Hiefür benötigen alle Klassen einen Standard-Konstruktor damit der Serializer diesen Aufrufen kann.

Des weiteren benötigt der Serializer für alle nicht `public` Attribute Getter- und Settermethoden um Zugriff auf diese Attribute zu erhalten. Denn der Serializer darf durch Java-Richtlinien nur auf `public` Attribute ohne Getter- und Settermethoden zugreifen.

Um eine Serialisierung von allen Klassen zu gewährleisten muss eventuell das OPM-Modell angepasst werden. Da alle Klassen sich an die OPM-Regeln halten, ist somit gewährleistet das alle ankommenden Instanzen serialisiert oder deserialisiert werden können.

In ersten Tests mit dem JSONSerializer wurde die Funktionsfähigkeit bewiesen. Jedoch ist der derzeitige Stand des Serialisierers noch nicht in der Lage verschiedene Klassen zu deserialisieren. Bisher ist der Klassenname fest vorgegeben.

Im Projekt aber sollen unterschiedlichste Klassen deserialisiert werden können. Das wirft die Frage auf wie festgestellt wird, um welche Klasse es sich bei der Instanz handelt. Dieser Punkt wird nun in den folgenden Kapiteln ausführlich erläutert.

6 Der SMD-Assistent

sdhkfysdhfhhkgs

7 Abkürzungsverzeichnis

KIT *Karlsruher Instituts für Technologie*

GDS *generisches Managementsystem für Energiedaten*

LSDF *Large Scale Data Facility*

OPM *objektorientierten Programmiermodell*

SMD *strukturelle Metadaten*

JSON *JavaScript Object Notation*

HALO *High Altitude and Long Range Research Aircraft*

IAI *Institut für Angewandte Informatik*

JAXB *Java Architecture for XML Binding*

Literatur

- [Gri02] GRIFFITHS, Alan: *More Exceptional Java*. <http://accu.org/index.php/journals/399>. Version: Juni 2002
- [Jac14] *Jackson Project Home*. <https://github.com/FasterXML/jackson>. Version: August 2014
- [Sri13] SRIPAPASA, Sai S.: *JavaScript and JSON Essentials*. 2013
- [Wal14] WALZ, Achim: Serialisierung von Datenobjekten in XML zur Übertragung von Objektdaten aus Energieanwendungen / KIT. 2014. – Forschungsbericht
- [Wik14a] *JavaScript Object Notation*. <http://de.wikipedia.org/wiki/JSON>. Version: Juni 2014
- [Wik14b] *Serialisierung*. <http://de.wikipedia.org/wiki/Serialisieren>. Version: Mai 2014
- [Zil14] ZILIAN, Lars: Strukturelle Metadaten- Objektorientierte Beschreibung von Klassen für das generische Management von Energiedaten und -Modellen / KIT. 2014. – Forschungsbericht