



Speicherung der Quellen mittels SemanticForms

Seminararbeit

von

Cedric Kulbach, Steffen Strobl

Am Institut für Angewandte Informatik und Formale
Beschreibungsverfahren(AIFB)

Datum: 3. Februar 2015

Inhaltsverzeichnis

1	Einleitung	3
1.1	Hintergrund	3
1.2	Herangehensweise	4
1.3	Funktionale und nicht funktionale Anforderungen	5
2	Konzept	6
2.1	User Story	6
2.2	Lösungsansatz	6
2.3	Einfügen der Quellen in Semantic Forms	7
2.3.1	Problematik	7
3	Umsetzung	8
3.1	Einlesen der Quellen in SemanticForms	8
3.2	Quellenausgabe	10
3.3	Anzeige der Vertrauenswürdigkeit	10
3.4	Installation	11
3.4.1	Benötigte Extensions	11
3.4.2	Durchführung	11
4	Fazit und Ausblick	12
4.1	Erfüllung der Anforderungen	12
4.1.1	Bestehende Probleme	12
4.1.2	Verbesserungsmöglichkeiten	13
	Literaturverzeichnis	14
	Abbildungsverzeichnis	15

1 Einleitung

1.1 Hintergrund

In MediaWiki oder Semantic MediaWiki ist es oft wichtig und interessant zu wissen woher das Wissen einer Seite eigentlich stammt. Für den Leser eines Wikis ist es meist nicht unerheblich wie alt eine Quelle ist oder wie vertrauenswürdig sie ist. Derzeit besteht bei MediaWiki und Semantic MediaWiki nicht die Möglichkeit Quellen optisch herauszuheben oder semantisch zu hinterlegen. Dies führt in der Regel dazu, dass der Nutzer die Quellen schlichtweg ignoriert. Zudem gibt es keine Möglichkeit die Güte einer Quelle zu bestimmen oder die Häufigkeit von Verweisen statistisch zu erfassen, da sie nicht semantisch erfasst sind.

Die Ziele des Seminars sollen sich auf diese zwei Bereiche konzentrieren. Erstens sollen die Quellen optisch für den Nutzer aufbereitet werden, so dass dieser einen schnellen Zugriff auf die Quellen hat und das er bei einer Textstelle, bei der eine Quelle hinterlegt ist, direkt eine Anzeige mit allen zur Quelle gehörenden Metadaten erhält. Das zweite Ziel ist die Erfassung der Quellen selbst. Bisher werden diese in MediaWiki nicht semantisch erfasst, was die Arbeit mit ihnen, wie zum Beispiel die Erstellung von Statistiken, erschwert bzw. nicht ermöglicht. Deshalb soll im Laufe des Seminars eine semantische Erfassung für Quellen ermöglicht werden um eine Struktur für diese hinterlegen zu können.

Zudem soll dem Nutzer die Option zur Verfügung gestellt werden, Quellen zu bewerten. Daraus wird dann die Vertrauenswürdigkeit berechnet, die dem Nutzer eine Einschätzung geben soll, in wie weit er sich auf die Richtigkeit des gelesenen Textes verlassen kann. Quellen, die eigene Wiki-Seiten haben, bekommen die Vertrauenswürdigkeit auf diesen angezeigt. Der Ablauf der zu implementierenden Extension lässt sich somit in folgendem Sequenzdiagramm darstellen.

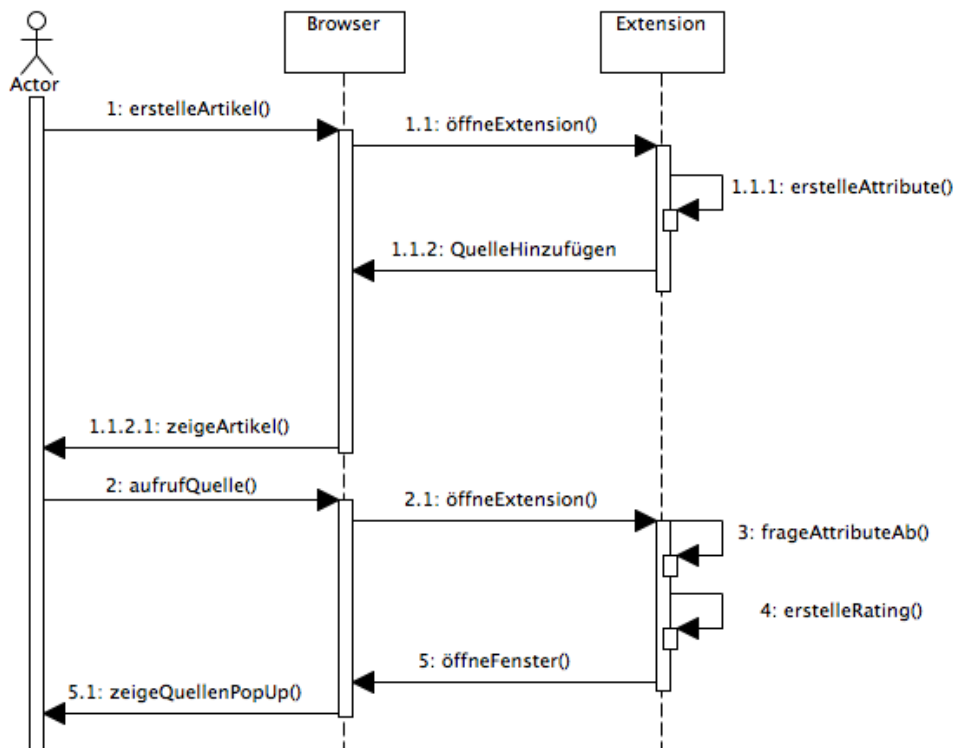


Abbildung 1.1: Sequenzdiagramm

Hierbei erstellt der Nutzer zunächst in seinem Browser auf der Oberfläche von Semantic MediaWiki einen Artikel. Dabei kann er verschiedene Attribute erstellen und diese mit Inhalt füllen. Anschließend kann er sich den Artikel anzeigen lassen und Quellen zu einem Text hinzufügen. Dazu ist in der Oberfläche von Semantic MediaWiki ein Button integriert, welcher die eigentliche Extension öffnet. Klickt man auf diesen Button öffnet sich ein neues Fenster. Dies fragt die Attribute der Quelle ab und übergibt sie der Extension. Diese berechnet das Rating der Quelle und erstellt sie. Das geöffnete Fenster wird geschlossen und der Nutzer kann die Bearbeitung des Artikels abschließen.

1.2 Herangehensweise

Um das eigentliche Problem der Speicherung der Quellen zu verstehen und anschließend einen möglichen Lösungsansatz zu erarbeiten, wurde sich zu Beginn des Seminars mit dem eigentlich Aufbau von Semantic MediaWiki befasst. In wöchentlichen Treffen wurden die bisherigen Fortschritte diskutiert und weitere Aufgaben verteilt. Um eine einheitliche Extension zu entwickeln, wurde als

Versionsverwaltungssoftware Git, als Dokumentenmanagementplattform Google Drive und als Entwicklungsumgebung XAMPP genutzt. Um einen Überblick über die Mächtigkeit der Aufgabe zu erlangen, wurde das Problem zunächst in seine Anforderungen zerlegt. Hierbei wurden funktionale und nicht funktionale Anforderungen berücksichtigt.

1.3 Funktionale und nicht funktionale Anforderungen

Die funktionale Anforderung der Extension ist es vor allem Quellen zu verwalten. Dazu gehört zum einen das Abfragen einer neuen Quelle, sowie das eigentliche erstellen der Quelle. Dazu sollte die Extension in der Lage sein eine Vorlage für die Quelle anzulegen, sowie ein Rating der Quelle zu berechnen. Zusätzlich sollte die Quelle ausgelesen werden können. Zu den nicht funktionalen Anforderungen gehört vor allem eine für den Nutzer einfache Darstellung der Quelle.

Noch einmal aufgeführt lauten die funktionalen Anforderungen:

- Eingabe der Quellen über separates Formular.
- Textmarkierung im Artikel, welches Attribut eine Quelle hinterlegt hat.
- Aufruf der Quelle per 'Klick'.
- Berechnung der Vertrauenswürdigkeit einer Quelle.
- In Artikeln, die als Quelle fungieren, soll die Vertrauenswürdigkeit der Seite angezeigt werden.

Nicht funktionale Anforderungen sollen sein:

- Dem Nutzer wird kein Quell-/Programmiercode angezeigt
- Es wird keine Vorkenntnis benötigt um die Extension zu benutzen.

2 Konzept

2.1 User Story

Hiwi Karl U. soll bei Semantic-MediaWiki einen Artikel mittels Semantic Forms erstellen. Das Formular ist bereits erstellt und Karl muss es nur noch Aufrufen und kann mit der Erstellung des Artikels beginnen. Er hat dabei im Moment noch keine Möglichkeit Quellenangaben zu tätigen, was bei einem wissenschaftlichen Thema unabdingbar ist. Des Weiteren hat Karl keine Möglichkeit die Vertrauenswürdigkeit seiner Quelle zu bewerten. Deshalb entsteht am Ende des Vorgangs ein Artikel ohne Quellenangabe. Würde Karl SemanticSources verwenden, könnte er ohne Vorwissen Quellen anlegen und diese bewerten.

2.2 Lösungsansatz

Im Laufe des Seminars sind 3 Aufgaben zu bewältigen.

1. Die Implementierung der Benutzeroberfläche
2. Die Integration und der Aufruf der Quellen in und aus MediaWiki
3. Die nachvollziehbare Berechnung der Vertrauenswürdigkeit der Quellen

Die genannten Punkte sollen grundsätzlich als Extension in MediaWiki integriert werden, so dass ein Nutzer diese ohne großen Aufwand installieren kann. Die Erweiterung hat den von MediaWiki genutzten Aufbau. Die Struktur soll, wie von Haenelt [1] beschrieben, erstellt werden. Für Semantic Sources soll allerdings nur eine php-Datei verwendet werden, in der bestimmt wird, wann auf welche Scripte zugegriffen werden soll. In den .js-Scriptdateien wird per jQuery die Funktionalität der Extension implementiert, sowohl das Erstellen der Benutzeroberfläche als auch die Integration und der Aufruf der Quellen.

2.3 Einfügen der Quellen in Semantic Forms

Für die Speicherung der Quellen werden sogenannte Subobjects verwendet. Diese ermöglichen das Speichern von Attributen ohne dass diese eine eigene WikiSeite erhalten. Um eine Quelle zu speichern markiert der Nutzer die gewünschte Textstelle und betätigt den 'Add source'-Button. Es öffnet sich ein Fenster in dem er alle nötigen Metadaten eintragen kann. Nachdem die Eingabe bestätigt wurde, wird das Subobject im Freitext-Textfeld(für den Nutzer nicht sichtbar) der SemanticForms angehängt. Zudem wird eine CITE-Referenz [2] erstellt mit Hilfe der Nutzer später auf die Quellen zugreifen kann. Dem Nutzer wird dabei jeglicher Quellcode ausgeblendet, so dass er nur das Formular sieht, dass er mit SemanticForms erstellt hat.

2.3.1 Problematik

Die Tabelle, die durch SemanticForms für den Artikel erstellt wird, hat durch die verwendeten Attribute vorgefertigte Felder mit festgelegten Datentypen. Deshalb kann der Code für die CITE Referenz nicht in die Tabelle geschrieben werden, denn MediaWiki erwartet den vorgegebenen Datentypen ohne Abweichungen. Wie in Abbildung 2.1 zu sehen ist, wird keine CITE Fußnote eingefügt, sondern eine Fehlermeldung zurückgegeben.

The image shows two parts of a form. The top part is a label 'Gründungsjahr:' followed by a text input field containing the code '1878<ref> </ref>'. The bottom part is a table with two columns. The first column is labeled 'Gründungsjahr' and the second column contains the value '1878' followed by a yellow warning triangle icon.

Gründungsjahr	1878⚠
---------------	-------

Abbildung 2.1: Cite Problematik

Um dies zu umgehen muss der Code in den Freitext geschrieben werden und per HTML dem Nutzer vorgetäuscht werden, dass die Referenzen in den Tabellenfeldern stehen. Ein weiteres Problem bei MediaWiki ist, dass jeder Quellcode, der von MediaWiki geparkt werden soll, vor dem Parsing der Seite in ein Feld von MediaWiki geschrieben werden muss. D.h. man kann in der Regel nicht nachträglich per HTML oder jQuery MediaWiki Funktionen ausführen. Zudem funktioniert das Einlesen der Quellen vorerst nur in SemanticForms in den Feldern, die der Nutzer per Vorlage erstellt. Textstellen im Freitext mit Quellen zu versehen, ist bisher nicht integriert. Genauso wenig die Möglichkeit pro Feld mehr als eine Quellenangabe zu tätigen. Dies liegt an der Verknüpfung der Subobjects zu den Datenfeldern. Pro Feld ist nur eine Verknüpfung möglich.

3 Umsetzung

3.1 Einlesen der Quellen in SemanticForms

Das Einlesen der Quellen lässt sich grob in 3 Schritte einteilen. Zuerst mit der Einbindung des Buttons und des Formulars, in dem der Nutzer alle Angaben zur Quelle tätigen kann. Dann kommt der Schritt mit der Erstellung der Subobjects, dem ask-Query auf die Subobjects und der CITE Referenz im Freitext der SemanticForms. Zuletzt wird dann das Anzeigen der Referenzen auf der Artikelseite beschrieben.

Um das Script, das den Button und das Metadatenformular zu SemanticForms hinzufügt, anzuzeigen, wird der MediaWiki Hook 'BeforePageDisplay' verwendet. Da dieser bei jeder Wiki- Seite reagiert, wird auf das Vorhandensein der HTML-Element Klasse createboxInput geprüft, die nur auf der SemanticForms-Seite existiert. Ist dies der Fall wird der Button und das Formular als HTML Element erstellt, wobei das Formular unsichtbar ist, solange der Button nicht betätigt wird. Um den markierten Text einzulesen, wird das jQuery-PlugIn 'fieldSelection' [3] verwendet. Dieses ermöglicht es die Position des Cursors auszulesen, damit die CITE Referenz dem richtigen Feld zugewiesen werden kann.

Mit dem 'insert-'Button beginnt der zweite Schritt. Durch den 'Klick' läuft ein Script ab, dass zuerst ein nicht sichtbares Element('Hidden Fields') erstellt. Dieses Element hat eine ID zusammengesetzt aus dem Namen des Textfelds in dem die

Freitext:

```
{{#subobject:1422457748456 Name der Uni
|source=Der Tag
|title=Sein oder nicht sein
|link=http://www.spiegel.de
|author=Bernd Meyer
|date=20.11.2012
|access=15.01.2015
|trustworthiness={{#ask: [[source::Der Tag]] ?rating|format=average}}|rating=4
}}
```

Abbildung 3.1: Subobject

Quelle stehen soll und dem Anhang '_semsources'. Zudem hat das Element einen Wert zusammengesetzt aus dem Attribut, für das die Quelle angegeben wird und dem CITE Tags '<ref></ref>'. Außerdem wird ein Subobject mit dem Namen des Attributs, allen eingegeben Daten und der berechneten Vertrauenswürdigkeit der Quelle in den Freitext geschrieben (Abbildung 3.1). Die Vertrauenswürdigkeit wird dabei als Durchschnitt aller bisher angegebenen Ratings zu der jeweiligen Quelle berechnet.

Nach dem der Nutzer alle benötigten Quellen hinzugefügt hat, kann er den Artikel speichern. Hier läuft ein weiteres Script ab, dass die Werte aus den 'Hidden Fields' an den Anfang des Freitextes in ein Tag <semsources> schreibt (Abbildung 3.2).

Dies wird benötigt um später feststellen zu können, an welcher Stelle die CITE Referenzen platziert wurden. Befindet sich bereits ein <semsources>-Tag im Freitext wird dieses gegebenenfalls aktualisiert oder erweitert. Zudem wird ein #ask Query [4] erstellt, dass auf die Subobjects des Artikels prüft und alle Quellenattribute ausgibt.

```
<semsources>
Name der Uni<ref> </ref>
</semsources>

{{#ask: [[-Has subobject::{{FULLPAGENAME}}]]|maintitle=semsources|?source|?title|?link|
?author|?date|?place|?isbn|?publisher|?access|?trustworthiness|?rating}}
```

Abbildung 3.2: Semsources Tag und Query

Nachdem gespeichert wurde befindet man sich auf der Artikelseite. Hier läuft das letzte Script zu Einlesung der Quellen ab. Ist die Prüfung, ob es sich bei der Seite um einen Artikel handelt und dieser mindestens ein Subobject enthält positiv, wird innerhalb des <semsources>-Tag nach CITE Referenzen gesucht. Nun wird anhand des Attributnamens zugeordnet in welches Feld der Verweis eingefügt wird. Über die entstandene Fußnote kann der Aufruf der Quellen erfolgen.

Name der Uni	Uni Basel ^[1]
---------------------	--------------------------

Abbildung 3.3: CITE Referenz in Datenfeld

3.2 Quellenausgabe

Das Fenster, das dem Nutzer bei Aufruf der Quelle angezeigt wird, wird mit dem JQuery PlugIn WebUI-Popover [5] realisiert. In dieses Fenster werden per jQuery die Daten, aus der durch das #ask erstellten Tabelle, eingefügt. Hierfür wird der abgefragte Attributsname in der Tabelle gesucht. Wird die richtige Zeile gefunden, werden die nicht leeren Felder in der Zeile in das Popover Fenster geschrieben.

3.3 Anzeige der Vertrauenswürdigkeit

In Semantic MediaWiki existieren meist neben den Artikeln auch die Quellen als Seite. Wenn der Nutzer nun mehr über eine Quelle erfahren möchte, konnte er sich über deren Seite über die Quelle informieren. Jedoch existiert die Möglichkeit nicht sich über die Vertrauenswürdigkeit der Quelle zu informieren. Diese sollte die Ratings, die zu dieser Quelle gehören, zusammenfassen und benutzerfreundlich in die Seite der Quelle integrieren. Der Zugriff auf die Ratings der Quelle erfolgt mit der Hilfe von eingebetteten Abfragen. In Semantic MediaWiki werden solche Abfragen mit der Parserfunktion #ask realisiert. Dadurch, dass der Titel der Quelle meist dem Quellennamen entspricht, fragt die Parserfunktion zunächst alle Subobjects ab, die den Titel der Seite als Quelle gespeichert haben. Nun wird mithilfe der Extension 'SemanticResultFormat' das arithmetische Mittel der abgefragten Subobjects berechnet und ausgegeben. Diese Abfrage liefert somit Werte zwischen 1 und 5 und hat folgende Gestalt:

```
{{#ask: [[Quelle::{{BASEPAGENAME}}]] |? Rating | format = average}}
```

Würde diese Abfrage nun in die Seite der Quelle integriert werden, erscheint lediglich ein Wert zwischen 1 und 5. Für den Nutzer ist dies jedoch nicht von großem nutzen, da dieser kein Wissen über die Skala hat und erst in der Seite der Quelle suchen muss um die Vertrauenswürdigkeit zu finden. Um die Übersichtlichkeit und den Nutzen zu steigern, wird die Vertrauenswürdigkeit mit der Hilfe der Extension 'SemanticRating' im Titel der Quellenseite dargestellt. Die Abfrage der Ratings wird somit in folgender Form erweitert:

```
{{#ratingAfterTitle:{{#ask: [[Quelle::{{BASEPAGENAME}}]] |? Rating | format = average}}}}
```

Um die Vertrauenswürdigkeit nun auf jeder Quellenseite im Titel anzeigen zu lassen, muss die Zeile in das Bearbeitungsfenster jeder Quellenseite integriert

werden. Dazu wird die Extension 'SemanticSources' um die Funktion RatingFunction.js erweitert. Diese fügt in jedes Bearbeitungsfenster falls sie noch nicht vorhanden ist die Zeile der Abfrage hinzu. Falls kein Rating der Quelle vorhanden ist, soll das Rating im Titel der Seite nicht dargestellt werden. Dies wird mit der Parserfunktion #ifeq ermöglicht. Der Nutzer sieht nun sehr schnell im Titel der Quellenseite die Vertrauenswürdigkeit dieser Quelle:



Abbildung 3.4: Quellentest

3.4 Installation

3.4.1 Benötigte Extensions

Benötigte Extensions Um die Extension in Semantic MediaWiki lauffähig zu integrieren werden bei der Installation die folgenden Extensions vorausgesetzt:
LINKS HINZUFÜGEN

- SemanticForms [6]
- SemanticRating [7]
- ParserFunctions
- SemanticResultFormats [8]
- Cite

Diese können via Git oder Manuell in Mediawiki eingebunden werden.

3.4.2 Durchführung

Die entwickelte Quellenverwaltung 'SemanticSources' wird in das Unterverzeichnis extensions/ des MediaWikis entpackt. Im MediaWiki- Ordner befindet sich die Datei LocalSettings.php, wobei am Ende des Scripts

```
include_once '$IP/extensions/SemanticSources/SemanticSources.php';
```

eingefügt wird.

4 Fazit und Ausblick

4.1 Erfüllung der Anforderungen

Die funktionalen Anforderung der Extension bestanden vor allem in der Erstellung und Verwaltung von Quellen in semantischem Kontext. Der Nutzer kann mithilfe der Extension in Formularfeldern Quellen hinzufügen, dessen Attribute durch Subobjects weiter abgefragt werden. Die Quelle wird dabei vom Autor bewertet, wobei sich das Gesamtrating aus dem arithmetischen Mittel der Einzelratings ergibt. Somit wird die funktionale Anforderung, dass die Quelle erstellt werden kann, erfüllt. Die Quellenverwaltung findet weitestgehend über Parserfunktionen in Semantic MediaWiki statt. Hierbei wird die Quelle, wie schon beschrieben, in einem PopUp neben dem Attribut des Formulars abgefragt und dem Nutzer übersichtlich alle relevanten Informationen zur Quelle angezeigt. Zudem sind durch die Verwendung von Subobjects der Quellenabfrage und somit der Nutzung der Quellen kaum Einschränkungen gegeben. Eine weitere funktionale Anforderung war, dass in Artikeln der Quelle dem Nutzer ebenfalls die berechnete Vertrauenswürdigkeit angezeigt wird. Dies wird durch die Anzeige von Sternen im Titel des Artikels ermöglicht. Zu den nicht funktionalen Anforderungen gehörte vor allem eine einfache Bedienung der Extension. Dies wurde nur erfüllt, da die Bedienbarkeit keine Vorkenntnisse voraussetzt und dem Nutzer kein Quellcode angezeigt wird.

4.1.1 Bestehende Probleme

Ein grundlegendes zu hinterfragendes Problem der Extension, ist die Annahme, dass der Autor bzw. der Nutzer, der die Quelle anlegt eine objektive Einschätzung über die Glaubwürdigkeit der Quelle angibt.

Browserabhängig kann es vorkommen, dass für einen Artikel die Vertrauenswürdigkeit angezeigt wird, obwohl dieser nicht als Quelle angegeben wurde. Es werden dann 5 leere Sterne angezeigt.

4.1.2 Verbesserungsmöglichkeiten

Die Vertrauenswürdigkeit im PopUp Fenster könnte auch mit Sternen dargestellt werden. Die SemanticRating Extension funktioniert hier leider nicht, da die Quellentabelle, die durch das ask-Query erstellt wird, einen Link als Datentypen erwartet und keine Bilder.

Des Weiteren könnte es ermöglicht werden mehrere Quellen pro Datenfeld anzugeben. Zudem könnte die Extension erweitert werden, damit sie auch bei dem klassischen 'Artikel erstellen' funktioniert und nicht nur bei SemanticForms und den Artikeln, die daraus entstehen.

Literaturverzeichnis

- [1] Programmierung von Extensionen für das MediaWiki , K. Haenelt, 2013
- [2] <http://www.mediawiki.org/wiki/Extension:Cite>
- [3] jQuery-PlugIn fieldSelection, A. Brem, 2006, Zugriff: 13.01.2015,
<https://github.com/localhost/jquery-fieldselection/tree/master>
- [4] http://semantic-mediawiki.org/wiki/Help:Inline_queries
- [5] jQuery-PlugIn webUI, S. Duan, Zugriff:25.01.2015,
<https://github.com/sandywalker/webui-popover>
- [6] http://semantic-mediawiki.org/wiki/Semantic_Forms
- [7] http://semantic-mediawiki.org/wiki/Semantic_Rating
- [8] http://www.mediawiki.org/wiki/Extension:Semantic_Result_Formats

Abbildungsverzeichnis

1.1	Sequenzdiagramm	4
2.1	Cite Problematik	7
3.1	Subobject	8
3.2	Semsource Tag und Query	9
3.3	CITE Referenz in Datenfeld	9
3.4	Quellentest	11

Hiermit versichern wir, dass wir die vorliegende Seminararbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt haben.

Karlsruhe, den 3. Februar 2015