



Formale Kategorisierung

openTA-Newsdienst

openTA-Dokument Nr. 1

Stand: 12.2.2014, Version 1.0

Autor: Christian Schmitt (KIT-IAI)

Zusammenfassung und Zielgruppe: Dieses Dokument beschreibt, wie in News-Feeds der NTA-Institutionen einzelne News-Einträge kategorisiert werden können. Es richtet sich in erster Linie an die Internetbeauftragten der institutionellen Mitglieder des NTA. Der openTA-Newsdienst wird durch zwei weitere Dokumente ergänzend beschrieben:

- openTA-Newsdienst – Ziele, Verfahren, Nutzung (openTA-Dokument Nr. 2)
- openTA-Newsdienst – Technisches Handbuch (openTA-Dokument Nr. 3)

Kontakt: info@openta.net

Die openTA-Dokumente sind abrufbar unter der Adresse
<http://www.openta.net/veroeffentlichungen-und-openta-dokumente>

Versionsgeschichte:

März 2013 erster Entwurf

Januar 2014 Version 1.0 (nur redaktionelle Änderungen)

Rechtehinweis: CC BY-SA 4.0

Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International

Im Rahmen des Forschungsvorhabens „Kooperativer Aufbau eines Fachportals Technikfolgenabschätzung“, gefördert durch die DFG, durchgeführt von den KIT-Instituten ITAS, IAI und der Bibliothek.

Inhaltsverzeichnis

1	Vorwort.....	5
2	Formale Kategorien von openTA.....	5
3	Verwendung in Web-Feeds	5
3.1	Verwendung in Atom-Feeds	6
3.2	Verwendung in RSS-Feeds.....	8
4	Anhang A: APP-Kategorie-Dokument für die formalen Kategorien von openTA	10

Tabellenverzeichnis

Tab. 1:	Formale openTA-Nachrichtenkategorien (de/en).....	5
Tab. 2:	Formale openTA-Kategorien – hierarchische und flache Struktur (de/en).....	6
Tab. 3:	Formale openTA-Kategorien (de/en) mit Term-Attribut	7

1 Vorwort

Im Rahmen des DFG-Projektes openTA wird ein Dienst entwickelt, welcher Neuigkeiten aus der TA-Community aufammelt, aufbereitet und integriert bereitstellt. Im Folgenden wird beschrieben, wie in News-Feeds der NTA-Institutionen einzelne News-Einträge kategorisiert werden können, damit diese sich im openTA-Newsdienst themenspezifisch und institutionsübergreifend zusammenfassen lassen.

2 Formale Kategorien von openTA

Im Rahmen des openTA-Projekts wurde folgende Hierarchie von Kategorien für die formale Klassifizierung von Nachrichten (und anderen Inhalten) definiert:

Tab. 1: Formale openTA-Nachrichtenkategorien (de/en)

Deutsch	Englisch
1. Angebot	1. offer
1. Stellenangebot	1. job offer
2. Kooperationsangebot	2. cooperation offer
3. Sonstiges	3. other
2. Institutionelles	2. institutional matters
1. Personalia	1. personel matters
2. Sonstiges	2. other
3. Projekt	3. project
4. Publikation	4. publication
5. Veranstaltung	5. event
1. Call for Papers	1. call for papers
2. Ankündigung	2. announcement
3. Bericht	3. report
6. Sonstige	6. other

3 Verwendung in Web-Feeds

Zur Verbreitung von Nachrichten in Form von Web-Feeds gibt es zwei relevante Standards, die RSS-Format-Familie und das Atom-Format. Beide Standards erlauben die Angabe von Kategorien um eine semantische Einordnung der jeweiligen Einträge zu ermöglichen.

Eine hierarchische Gliederung der einzelnen Kategorien ist jedoch nicht vorgesehen, so dass die Hierarchie in einer flachen Struktur abgebildet werden muss. Für openTA wurde dies durch Voranstellen der übergeordneten Kategorie realisiert, wobei die Hierarchiestufen durch einen Schrägstrich (/) getrennt werden. Daraus ergibt sich folgendes Schema:

Tab. 2: Formale openTA-Kategorien – hierarchische und flache Struktur (de/en)Hierarchische Struktur

	Flache Struktur/de	Flache Struktur/en
1. Angebot	Angebot	offer
1. Stellenangebot	Angebot/Stellenangebot	offer/job offer
2. Kooperationsangebot	Angebot/Kooperationsangebot	offer/cooperation offer
3. Sonstiges	Angebot/Sonstiges	offer/other
2. Institutionelles	Institutionelles	institutional matters
1. Personalia	Institutionelles/Personalia	institutional matters/personnel matters
2. Sonstiges	Institutionelles/Sonstiges	institutional matters/other
3. Projekt	Projekt	project
4. Publikation	Publikation	publication
5. Veranstaltung	Veranstaltung	event
1. Call for Papers	Veranstaltung/Call for Papers	event/call for papers
2. Ankündigung	Veranstaltung/Ankündigung	event/announcement
3. Bericht	Veranstaltung/Bericht	event/report
6. Sonstige	Sonstige	other

Sofern ein Content-Management-System (CMS) für die Erzeugung der Feeds eingesetzt wird und dieses das Anlegen von Kategorien ermöglicht, sollten dort die Einträge wie in den beiden rechten Spalten der obigen Tabelle (flache Struktur, de oder en) angelegt werden. Wenn diese Möglichkeit nicht besteht, kann der zuständige System-Administrator eventuell weiterhelfen.

Sollte im CMS die Angabe eine Referenz auf ein Klassifikationsschema für die jeweiligen Kategorien möglich sein, ist die korrekte Referenz:

`http://service.openta.net/feeds/atompub/scheme/formal-category`

Wenn es nicht möglich ist, eine solche Schema-Referenz zu setzen und die Kategorien von openTA dennoch klar von anderen Kategorien getrennt werden sollen, kann jeder Kategorie die Zeichenfolge „openTA/“ vorangestellt werden.

3.1 Verwendung in Atom-Feeds

Im Zusammenhang mit Atom-Feeds sind zwei Standards relevant. Zum Beschreiben von Feed-Inhalten gibt es das *Atom Syndication Format (ASF)*¹, zum anderen für die standardisierte Verbreitung und Bearbeitung von ASF-basierten Inhalten das *Atom Publishing Protocol (AtomPub / APP)*².

Für jeden Eintrag in einem ASF-Feed kann eine beliebige Anzahl von Kategorien in folgendem Format angegeben werden:

```
<category term="foo" scheme="http://example.org/bar" / label="Foo">
```

Die Angabe für das *term*-Attribut definiert hierbei den maschinenlesbaren Schlüssel der Kategorie, die optionale Angabe des *scheme*-Attributs definiert die Referenz auf ein Kategorie-Dokument, welches den kompletten Katalog der möglichen Kategorien beschreibt. Das *label*-Attribut beschreibt für Menschen lesbar den Namen der jeweiligen Kategorie. Die Werte für das *Term*-Attribut sind wie folgt:

¹ The Atom Syndication Format: <http://tools.ietf.org/html/rfc4287>

² The Atom Publishing Protocol: <http://tools.ietf.org/html/rfc5023>

Tab. 3: Formale openTA-Kategorien (de/en) mit Term-Attribut

Deutsch	Englisch	Term
1. Angebot	1. offer	angebot
1. Stellenangebot	1. job offer	angebot-stelle
2. Kooperationsangebot	2. cooperation offer	angebot-koop
3. Sonstiges	3. other	angebot-misc
2. Institutionelles	2. institutional matters	inst
1. Personalia	1. personel matters	inst-pers
2. Sonstiges	2. other	inst-misc
3. Projekt	3. project	projekt
4. Publikation	4. publication	pub
5. Veranstaltung	5. event	veranst
1. Call for Papers	1. call for papers	veranst-call
2. Ankündigung	2. announcement	veranst-ankuend
3. Bericht	3. report	veranst-bericht
6. Sonstige	6. other	misc

Das Format des Dokuments ist im Rahmen des APP standardisiert (vgl. Anhang A). Der Wert des *scheme*-Attributs sollte die URL sein, unter welcher das Dokument abrufbar ist. Für die formalen Kategorien von openTA ist dies:

de: <http://service.openta.net/feeds/atompub/scheme/formal-category>
en: <http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en>

Beispiele

Die Beispiele enthalten jeweils zwei Kategorie-Einträge, zum einen zur openTA-Kategorie „Stellenangebot“ unterhalb von „Angebot“ (1.1) und zum anderen zur Kategorie „Projekt“ (3):

Beispiel 1: Empfohlene Notation

```
<entry xmlns="http://www.w3.org/2005/Atom">
  <title>Beispiel</title>
  <id>Beispiel-ID</id>
  <updated>2013-02-25T12:30:02Z</updated>
  <category
    scheme="http://service.openta.net/feeds/atompub/formal-category"
    term="angebot-stelle" />
    label="Angebot/Stellenangebot" />
  <category
    scheme="http://service.openta.net/feeds/atompub/formal-category"
    term="projekt"
    label="Projekt" />
</entry>
```

Beispiel 2: Notation, falls Angabe der Schema-Referenz nicht möglich

Wenn die Angabe des *scheme*-Attributs nicht möglich ist, kann dies auch weggelassen werden. Der Eintrag wird trotzdem von openTA korrekt erkannt. Die Angabe verliert jedoch an Kontext, da von neutralen Readern nicht mehr ausgewertet werden kann, welche Kategorien aus demselben Schema stammen. Zum besseren Verständnis kann dann auch das deutsche oder englische Label verwendet werden:

```
<entry xmlns="http://www.w3.org/2005/Atom">
  <title>Beispiel</title>
  <id>Beispiel-ID</id>
  <updated>2013-02-25T12:30:02Z</updated>
  <category term="Angebot/Stellenangebot"/>
  <category term="Projekt"/>
</entry>
```

Beispiel 3: Notation, falls Angabe der Schema-Referenz nicht möglich und klare Abtrennung der openTA-Kategorien von anderen Kategorien gewünscht

Wenn aufgrund der nicht vorhandenen *scheme*-Angabe die formalen Kategorien von openTA trotzdem als solche gekennzeichnet werden sollen, um sie evtl. von anderen verwendeten Schemata zu unterscheiden, kann dies durch Voranstellen des Präfix „openTA/“ bewerkstelligt werden.

```
<entry xmlns="http://www.w3.org/2005/Atom">
  <title>Beispiel</title>
  <id>Beispiel-ID</id>
  <updated>2013-02-25T12:30:02Z</updated>
  <category term="openTA/Angebot/Stellenangebot"/>
  <category term="openTA/Projekt"/>
</entry>
```

3.2 Verwendung in RSS-Feeds

RSS ist eine Familie von Dateiformaten, welche alle XML basiert sind. Je nach Version des Formats werden unterschiedliche Funktionalitäten unterstützt. Die Angabe der Kategorien ist jedoch bei sämtlichen Versionen identisch.

Für jeden Eintrag in einem RSS-Feed kann eine beliebige Anzahl von Kategorien in folgendem Format angegeben werden:

```
<category domain="http://example.org/bar">foo</category>
```

Der Wert des *category*-Tags definiert hierbei den Namen der Kategorie, die optionale Angabe des *domain*-Attributs definiert die Taxonomie. Für die Angabe dieser Taxonomie gibt es keinen Standard, es wird jedoch eine URL empfohlen. Da in RSS auch kein Standard für die über die URL referenzierte Beschreibung der Kategorien existiert, soll das Kategorie-Dokument des APP referenziert werden:

```
de: http://service.openta.net/feeds/atompub/scheme/formal-category
en: http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en
```

Beispiele

Die Beispiele enthalten jeweils zwei Kategorie-Einträge, zum einen zur openTA-Kategorie „Stellenangebot“ unterhalb von „Angebot“ (1.1) und zum anderen zur Kategorie „Projekt“ (3):

Beispiel 1: Empfohlene Notation

```
<item>
  <title>Beispiel</title>
  <guid>Beispiel-ID</guid>
  <pubDate>2013-02-25T12:30:02Z</pubDate>
  <category
    domain="http://service.openta.net/feeds/atopub/formal-category">
    Angebot/Stellenangebot</category>
  <category
    domain="http://service.openta.net/feeds/atopub/formal-category">
    Projekt</category>
</item>
```

Beispiel 2: Notation, falls Angabe der Domain-Referenz nicht möglich

Wenn die Angabe des *domain*-Attributs nicht möglich ist, kann dies auch weggelassen werden. Der Eintrag wird trotzdem von openTA korrekt erkannt. Die Angabe verliert jedoch an Kontext, da von neutralen Readern nicht mehr ausgewertet werden kann, welche Kategorien aus demselben Schema stammen.

```
<item>
  <title>Beispiel</title>
  <guid>Beispiel-ID</guid>
  <pubDate>2013-02-25T12:30:02Z</pubDate>
  <category>Angebot/Stellenangebot</category>
  <category>Projekt</category>
</item>
```

Beispiel 3: Notation, falls Angabe der Domain-Referenz nicht möglich und klare Abtrennung der openTA-Kategorien von anderen Kategorien gewünscht

Wenn aufgrund der nicht vorhandenen *domain*-Angabe die formalen Kategorien von openTA trotzdem als solche gekennzeichnet werden sollen, um sie evtl. von anderen verwendeten Schemata zu unterscheiden, kann dies durch Voranstellen des Präfix „openTA/“ bewerkstelligt werden.

```
<item>
  <title>Beispiel</title>
  <guid>Beispiel-ID</guid>
  <pubDate>2013-02-25T12:30:02Z</pubDate>
  <category>openTA/Angebot/Stellenangebot</category>
  <category>openTA/Projekt</category>
</item>
```

4 Anhang A: APP-Kategorie-Dokument für die formalen Kategorien von openTA

Deutsch

URL:

<http://service.openta.net/feeds/atompub/scheme/formal-category>

Inhalt:

```
<categories xmlns="http://www.w3.org/2007/app"
xmlns:atom="http://www.w3.org/2005/Atom">
  <atom:category term="angebot" label="Angebot" />
  <atom:category term="angebot-stelle" label="Angebot/Stellenangebot" />
  <atom:category term="angebot-koop" label="Angebot/Kooperationsangebot" />
  <atom:category term="angebot-misc" label="Angebot/Sonstiges" />
  <atom:category term="inst" label="Institutionelles" />
  <atom:category term="inst-pers" label="Institutionelles/Personalia" />
  <atom:category term="inst-misc" label="Institutionelles/Sonstiges" />
  <atom:category term="projekt" label="Projekt" />
  <atom:category term="pub" label="Publikation" />
  <atom:category term="veranst" label="Veranstaltung" />
  <atom:category term="veranst-call" label="Veranstaltung/Call for Papers" />
  <atom:category term="veranst-ankuend" label="Veranstaltung/Ankündigung" />
  <atom:category term="veranst-bericht" label="Veranstaltung/Bericht" />
  <atom:category term="misc" label="Sonstige" />
</categories>
```

Englisch

URL:

<http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en>

Inhalt:

```
<categories xmlns="http://www.w3.org/2007/app"
xmlns:atom="http://www.w3.org/2005/Atom">
  <atom:category term="angebot" label="offer" />
  <atom:category term="angebot-stelle" label="offer/job offer" />
  <atom:category term="angebot-koop" label="offer/cooperation offer" />
  <atom:category term="angebot-misc" label="offer/other" />
  <atom:category term="inst" label="institutional matters" />
  <atom:category term="inst-pers" label="institutional matters/personnel matters" />
  <atom:category term="inst-misc" label="institutional matters/other" />
  <atom:category term="projekt" label="project" />
  <atom:category term="pub" label="publication" />
  <atom:category term="veranst" label="event" />
  <atom:category term="veranst-call" label="event/call for papers" />
  <atom:category term="veranst-ankuend" label="event/announcement" />
  <atom:category term="veranst-bericht" label="event/report" />
  <atom:category term="misc" label="other" />
</categories>
```