

# **METASEARCH PROGRAMMIERHANDBUCH**

17. 9. 2013  
Manfred Schauer  
<http://www.metajob.at>

# 1. EINLEITUNG

METASearch ist eine Sammlung von Schnittstellen, die es ermöglichen, Suchen nicht nur über die grafische Benutzeroberfläche von METAJob durchzuführen, sondern sie auch in beliebige Client-Programme zu integrieren und dadurch die Suchergebnisse maschinell weiterverarbeiten zu können. METASearch folgt einer REST<sup>1</sup>-Architektur. Anfragen werden über HTTP-GET-Requests an den METAJob-Server gesendet, wie z.B.

```
http://www.metajob.at/metasearch?key=xyz&kw=php&location=4020&radius=20
```

Suchargumente werden dabei als URL-Parameter spezifiziert.

Der METAJob-Server antwortet darauf mit einem XML-Dokument, welches das Suchergebnis kodiert, Beispiel:

```
<?xml version="1.0" encoding="UTF-8"?>
<metasearch>
  <state>
    <code>0</code>
    <text>ok</code>
  </state>
  <jobs>
    <job>
      <title>PHP Entwickler Klagenfurt</title>
      <location>Wien, Graz</location>
      <found>2010-08-28</found>
      <snippet>... suchen wir einen PHP-Entwickler ... Kenntnissen.</snippet>
      <url>http://www.mycompany.com/offer4711</url>
      <company>XYZ GmbH</company>
    </job>
    <job>
      ...
    </job>
  </jobs>
</metasearch>
```

In den folgenden Abschnitten sind die einzelnen Abfragen beschrieben. Vorher folgt jedoch ein allgemeiner Abschnitt mit Konventionen, die für alle Abfragen gelten.

## 2. ALLGEMEINE KONVENTIONEN

### QUERY-PARAMETER

Suchargumente werden im URL als Teil des *Query-Strings* übergeben. Diese sog. *Query-Parameter* müssen kodiert werden, da hier nicht alle Zeichen verwendet werden können. Der URL

```
...metasearch?key=xyz&kw=Büro
```

ist nicht gültig. Der Wert 'Büro' muss geeignet kodiert werden, der URL sieht danach richtigerweise so aus:

```
...metasearch?key=xyz&kw=B%C3%BCro
```

Diese Form der Kodierung heißt auch *URL-Encoding* oder *Percent-Encoding* und muss mit dem Zeichensatz UTF-8 durchgeführt werden. Jede Programmiersprache bietet passende Routinen an, wie z.B: *encodeURIComponent* in JavaScript, *URLEncode.encode* in Java oder *urlencode* in PHP.

Alle Abfragen enthalten den Parameter 'key'. Das ist ein Code zur Identifikation des Clients. Er wird von METAJob bereitgestellt.

---

<sup>1</sup>[http://en.wikipedia.org/wiki/Representational\\_State\\_Transfer](http://en.wikipedia.org/wiki/Representational_State_Transfer)

## ERGEBNIS-XML

Abfragen liefern als Ergebnis ein Dokument mit Content-Type `text/xml` und encoding `UTF-8`. Das retournierte XML-Dokument besteht aus 2 Elementen (siehe dazu auch das Beispiel im einleitenden Abschnitt):

1. Das Element `<state>` zur Übermittlung von Statusinformation, siehe Tabelle 1. Es ist in der Antwort einer jeden Abfrage vorhanden.
2. Ein zweites Element, abhängig von der Abfrage selbst. Falls Jobs zurückgeliefert werden, ist es `<jobs>`. Siehe Tabelle 2.

XML-Element	Beschreibung
state	Ergebnisstatus der Abfrage
state/code	<b>0</b> ... Erfolg <b>kleiner 0</b> ... Warnung. Ergebnis in <code>&lt;jobs&gt;</code> kann verarbeitet werden, der Client wird jedoch auf eine bestimmte Konstellation der Query-Parameter hingewiesen, die verbessert werden kann. <b>größer 0</b> ... Fehler. <code>&lt;jobs&gt;</code> ist leer. Abfrage konnte nicht erfolgreich durchgeführt werden.
state/text	Text zum Statuscode.

Tabelle 1: `<state>`-XML-Element

XML-Element	Beschreibung
jobs	Enthält für jeden retournierten Job ein <code>&lt;job&gt;</code> -Child-Element.
jobs/job	Element für einen Job.
jobs/job/title	Jobtitel des Jobs. <code>&lt;title&gt;</code> kann auch fehlen, falls kein Titel ermittelt werden konnte.
jobs/job/location	Ermittelte Einsatzorte bzw. Einsatzregionen. Element kann fehlen. Enthält einen oder mehrere durch Komma getrennte Regionen.
jobs/job/found	Datum im Format <code>yyyy-mm-dd</code> . Gibt an, wann der URL des Jobs vom METAJob-Crawler zum ersten Mal gefunden wurde.
jobs/job/snippet	Durch "..." getrennte Auszüge aus dem Nutzttext des Jobinserats. Nur vorhanden, wenn der Parameter <code>ynText</code> nicht gesetzt ist, siehe auch <code>jobs/job/text</code> .
jobs/job/text	Der beschreibende Text des Jobinserats im Plain-Text-Format. Dieses Element wird anstelle von <code>snippet</code> geliefert, falls der Parameter <code>ynText=true</code> gesetzt ist. Die Größe des zurückgelieferten XMLs erhöht sich dadurch wesentlich.
jobs/job/url	URL des Jobangebots im Internet.
jobs/job/company	Firmenname des ausschreibenden Unternehmens, wie auf <code>www.metajob.at</code> angezeigt.

Tabelle 2: `<jobs>`-XML-Element

## BERECHTIGUNGEN, ZUGRIFFSVERHALTEN

Alle Abfragen benötigen einen `key`-Parameter, das ist ein Schlüssel, den METAJob an jeden Client einmalig vergibt. Er berechtigt zur Durchführung eines bestimmten Kontingents von Abfragen pro Tag, max. 100000, das entspricht etwa 1 Abfrage pro Sekunde. Sollten Abfragen vom Client-Programm iterativ abgesetzt werden, muss zwischen zwei aufeinander folgenden Abfragen mindestens 1 Sekunde Sleep-Time eingehalten werden.

## 3. ABFRAGE: SUCHE NACH KRITERIEN

Die *Suche nach Kriterien* liefert Jobs nach spezifizierten Suchkriterien wie Keywords, geografischer Region und Alter. Diese Abfrage entspricht jener Suchfunktionalität, die auch interaktiv auf der grafischen Benutzeroberfläche von METAJob in der Detailsuche zur Verfügung steht. Tabelle 3 zeigt die verfügbaren Parameter.

XML-Element	Beschreibung
key	Von METAJob zur Verfügung gestellter Schlüssel. Parameter muss angeführt werden.
kw	Keywords. 'kw' ist ein Synonym für den 'textAll'-Parameter. Dient nur der kürzeren und einfacheren Schreibweise.
textAll	Eines oder mehrere durch Blank getrennte Keywords. Qualifiziert sind nur Jobs, die alle Keywords im Titel oder Inseratentext enthalten.
textSome	Eines oder mehrere durch Blank getrennte Keywords. Qualifiziert sind Jobs, die mindestens eines der Keywords in Titel oder Inseratentext enthalten.
textNone	Eines oder mehrere durch Blank getrennte Keywords. Qualifiziert sind Jobs, welche die genannten Keywords nicht enthalten. Diese Form von negativen Keywords machen nur Sinn in Kombination mit anderen wie z.B. textAll oder textSome.
titleAll	Eines oder mehrere Keywords. Alle Wörter müssen im Jobtitel enthalten sein.
titleSome	Eines oder mehrere Keywords. Mindestens ein Wort muss im Jobitel enthalten sein.
titleNone	Eines oder mehrere Keywords. Jobs mit min. einem der genannten Keywords im Titel werden nicht geliefert.
location	Eine oder mehrere geografische Regionen, angegeben durch deren Namen oder durch die Postleitzahl. Die Verwendung von Postleitzahlen empfiehlt sich dann, wenn Ortsnamen nicht eindeutig (homonym) sind, wie z.B. 'Sankt Johann', oder das Risiko durch eine falsche Schreibweise zu hoch ist.
radius	Radius zur Umkreissuche in km. Falls dieser Parameter vorhanden ist, werden Jobs nicht nur in den durch 'location' bestimmten Gemeinden geliefert, sondern auch in deren Umkreis mit dem spezifiziertem Radius. Mögliche Werte sind 10, 20, 30, 50, 100, 150, 200.
maxAge	Maximales Alter der gelieferten Jobs in Tagen. Unter Alter eines Jobs versteht man in METAJob die Zeit seit dem erstmaligen Auffindens des URLs der Job-Seite im Internet. Zulässige Werte sind ganze Zahlen größer gleich 1.
maxResults	Maximale Anzahl der zu retournierenden Jobs. Wert muss innerhalb von [10, 1000] liegen. Falls nicht gesetzt, gilt 200 als Defaultwert.
ynText	Wenn auf true gesetzt, wird für jeden Job anstelle von <snippet> das Element <text> retourniert.

Tabelle 3: Parameter bei der Suche nach Kriterien

#### 4. ABFRAGE: JOBS ZU EINER DOMAIN

Diese Abfrage liefert alle auf METAJob verfügbaren Jobs zu einer bestimmten Internet-Domain, siehe Tabelle 4.

XML-Element	Beschreibung
key	Von METAJob zur Verfügung gestellter Schlüssel. Parameter muss angeführt werden.
domain	Ein Domain-Name, wie z.B.: <code>www.mycompany.com</code> oder <code>mycompany.at</code> . Muss zumindest aus 2 Namensteilen bestehen, die durch Punkte getrennt sind. Subdomains, wie z.B. 'www', können inkludiert sein, werden aber generell ignoriert.

XML-Element	Beschreibung
ext	<p>Mögliche Werte: 'true' und 'false'. Default ist 'true'.</p> <p>Bei ext='false' werden nur Jobs zur angegebenen Domain geliefert. Falls domain gleich 'mycompany.at' können z.B. Jobs der Domains 'www.mycompany.at', 'jobs.mycompany.at' oder 'mycompany.at' retourniert werden.</p> <p>Mit ext='true' werden Jobs einer <i>erweiterten</i> Domain-Menge retourniert. Firmen neigen dazu, Ihren Content unter mehreren Domains zu hosten. Variationen entstehen z.B. durch die Verwendung von unterschiedlichen Top-Level-Domains, wie z.B: mycompany.at versus mycompany.net, oder auch durch die Verwendung von Bindestrichen, wie z.B. my-company.at versus mycompany.at. <i>METASearch</i> liefert bei ext="true" die Vereinigungsmenge der Jobs der so verwandten Domains, jedoch nur dann, falls die Domains auf Rechnern gehostet werden, die sich im selben IP-C-Netz befinden. Beispiel: Falls domain gleich 'www.mycompany.at', können Jobs zu folgenden Domains retourniert werden:</p> <ol style="list-style-type: none"> <li>1. mycompany.at, jobs.mycompany.at, www.mycompany.at. Diese werden auch bei ext='false' retourniert.</li> <li>2. mycompany.com, www.mycompany.net, jobs.mycompany.net. Diese werden retourniert, falls die Domains auf "den selben Rechnern" wie mycompany.at gehostet werden, d.h. selbes IP-C-Netz.</li> <li>3. my-company.at, www.my-company.com. Wie voriger Punkt.</li> </ol>
ynText	<p>Wenn auf true gesetzt, wird für jeden Job anstelle von &lt;snippet&gt; das Element &lt;text&gt; retourniert.</p>

Tabelle 4: Parameter bei der Suche zu einer Domain