

Auszeichnen von Rezepten

Informationen zugänglich machen
TEI: Text Encoding Initiative
Schema.org/Recipe
In kommerziellen Kochseiten

Nachdem wir im vorherigen Abschnitt (Transkription.html) eine Transkription zu Davids' Kochbuch erarbeitet haben, wollen wir uns nun damit beschäftigen, wie wir die Informationen aus dem Buch automatisch auswerten können. Dafür gehen wir zuerst darauf ein, wie Informationen zugänglich gemacht werden können. Anschließend betrachten wir existierende Auszeichnungssprachen, die Informationen aus unserer Domäne zugänglich machen.

Informationen zugänglich machen

„The key to utilizing the knowledge of an application domain is identifying the basic vocabulary consisting of terms or concepts of interest to a typical user in the application domain and the interrelationships among the concepts in the ontology.“ (Kashyap et al., 2008, S. 30 f.)
(Quellen.html#SemanticWeb2)

Eine *Ontology* wiederum ist ein *„set of interest in a particular information domain and the relationships among them“* (Kashyap et al., 2008, S. 79) (Quellen.html#SemanticWeb2) Ein wohldefiniertes *Vokabular* kann demnach eine *Ontology* sein. Dieses kann beispielsweise durch ein *Extensible Markup Language Schema (XML Schema)* definiert werden (vgl. Kashyap et al., 2008, S. 84 ff.) (Quellen.html#SemanticWeb2) Alle Informationen eines XML Dokumentes nach einem Schema können dann automatisch extrahiert werden.

Das Ziel ist somit, die Rezepte aus Davidis' Kochbuch in einer Auszeichnungssprache zu erfassen, dessen Vokabular und Beziehungen untereinander wohldefiniert sind und die den Bereich einer kulinarischen Analyse abdeckt. Das Erfassen von Informationen durch eine Auszeichnungssprache nennen wir *auszeichnen*. Folgend stellen wir einige Auszeichnungssprachen für unsere Koch-Domäne vor.

TEI: Text Encoding Initiative

TEI (<http://www.tei-c.org/index.xml>) ist ein Standard, um Texte digital darzustellen, mit dem Schwerpunkt die Texte für Geistes- und Sprachwissenschaften maschinenlesbar zu machen. (TEI-Konsortium (A), 2017) ([Quellen.html#TEI](#))

Wie in Abb. 1 zu sehen ist, wird die Struktur eines Kochbuches eins zu eins abgebildet. Die strukturelle Verschachtelungstiefe wird durch die nummerierten *div*-Elemente abgebildet, die Überschriften durch *head*-Elemente, die Paragraphen in *p*-Elementen und die Listen durch *list*- und *item*-Elemente.

Abb. 1: Rezept mit TEI (Nach: TEI-Konsortium (B), 2017, S. 1036 f.) (Quellen.html#TEIP5)

Teil 1

```
<?xml version="1.0" encoding="UTF-8"?>
<div2 type="chapter">
  <head>Recipes</head>
  <head>Chapter VI.</head>
  <div3>
    <head>Fruit and vegetable soups</head>
    <!-- ... -->
    <div4 type="recipe">
      <head>Apple soup</head>
      <div5>
        <head>Ingredients</head>
        <list>
          <item>2 lbs. of good boiling apples,</item>
          <item>3/4 teaspoonful of white pepper,</item>
          <item>6 cloves,</item>
          <item>cayenne or ginger to taste,</item>
          <item>3 quarts of medium stock</item>
        </list>
      </div5>
      <div5>
        <head>Mode</head>
        <p>Peel and quarter the apples taking out their cores; put
          them into the stock, stew them gently till tender, Rub
          the whole through a strainer, add the seasoning. give
          it one boil up, and serve.</p>
      </div5>
    </div4>
  </div3>
</div2>
```

Teil 2

```

      <div5>
        <head>Time</head>
        <p>1 hour.</p>
      </div5>
      <div5>
        <head>Sufficient</head>
        <p> for 10 persons</p>
      </div5>
      <!-- ... -->
      <div5>
        <head>The apple</head>
        <p>This useful fruit is mentioned in Holy Writ; and Homer
          describes it as valuable in his time... As a food, the
          apple cannot be considered to rank high, as more than
          the half of it consists of water, and the rest of its
          properties are not the most nourishing. It is however a
          useful adjunct to other kinds of food, and, when
          cooked, is esteemed as slightly laxative.</p>
      </div5>
    </div4>
    <!-- other recipes here -->
  </div3>
</div2>
```

Schema.org/Recipe

Schema.org/Recipe (<https://schema.org/Recipe>) ist ein Vokabular, um Treffer von Suchmaschinen anzureichern. Die einzelnen Bereiche eines Rezeptes werden durch Microdata, RDFa oder JSON-LD kenntlich gemacht und können so von einer Suchmaschine ausgelesen werden. Abb. 2 zeigt einen Suchtreffer von Google zur Anfrage „Pfannekuchen“.

Abb. 2: Beispielhafter Google Treffer zu „Pfannekuchen“



In Abb. 3 sind die hinterlegten Meta-Informationen auf der Webseite des Treffers in JSON-LD-Format dargestellt. Es wird unter anderem ersichtlich, dass das Vorschau-Bild in Google eine Vorschau des *image*-Attributes ist. Die Bewertung ist der Wert des *aggregateRating.ratingValue*-Attributes und der Vorschau-Text ist der Beginn des *description*-Attributes.

Die wesentlichen Bestandteile der Webseite mit dem eigentlichen Rezept sind **hier** zu sehen.

Abb. 3: JSON-LD Anwendung von Schema.org/Recipe auf das Pfannekuchen-Rezept (Nach: 06onkel von Chefkoch.de, 2017) ([Quellen.html#ChefkochPfannekuchen](#))

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Recipe",
  "datePublished": "2005-07-20",
  "description": "Pfannekuchen, ein gutes Rezept [...]",
  "image": "http://[...].jpg",
  "recipeIngredient": [
    "250 g Mehl", "4 Ei(er)", "Salz", "300 ml Milch",
    "300 ml Mineralwasser, (mit Kohlensäure)", "Öl"
  ],
  "name": "Pfannekuchen",
  "author": {
    "@type": "Person",
    "name": "06onkel"
  },
  "prepTime": "PT30M",
  "recipeInstructions": "Den ganzen Teig mit [...]"
  "recipeYield": "4"
  ,
  "aggregateRating": {
    "@type": "AggregateRating",
    "ratingValue": "4.27",
    "reviewCount": "93",
    "worstRating": 0,
    "bestRating": 5
  },
  "recipeCategory": ["Braten", [...]]
}
</script>
```

In kommerziellen Kochseiten

Wie im vorherigen Abschnitt gesehen, verwendet Chefkoch.de (<http://www.chefkoch.de/>) Schema.org/Recipe. Zusätzlich speichert Chefkoch.de jedes Rezept in einem JSON-Objekt, was am Beispiel des Pfannekuchen-Rezeptes in Abb. 4 zu sehen ist. Neben einer zusätzlichen ID (id) für das Rezept, werden die Einträge der Zutatenliste, welche bereits mit Schema.org/Recipe erfasst wurden, genauer aufgespalten. Jede Zutat hat eine eigene ID (id), Mengenangabe (base_amount) mit Mengeneinheit (amount_unit), sowie optional einen Kommentar (suffix).

Abb. 4: Extra JSON-Attribute für Zutaten (Nach: 06onkel von Chefkoch.de, 2017)
(Quellen.html#ChefkochPfannekuchen)

```
<script type="text/javascript">
var _simplora_params = { recipes: [] };
_simplora_params.recipes.push({
  id: '363861121870267',
  name: 'Pfannekuchen',
  servings: 4,
  ingredients: [
    { id: '71', base_amount: '62.5', amount_unit: 'g', name: 'Mehl' },
    { id: '36', base_amount: '1', amount_unit: '', name: 'Ei(er)' },
    { id: '6', base_amount: '0', amount_unit: '', name: 'Salz' },
    { id: '73', base_amount: '75', amount_unit: 'ml', name: 'Milch' },
    { id: '1012', base_amount: '75', amount_unit: 'ml', name: 'Mineralwasser', suffix: ', (mit Kohlensäure)' },
    { id: '184', base_amount: '0', amount_unit: '', name: 'Öl' },
  ]
});
</script>
```

Cooking.nytimes.com (<https://cooking.nytimes.com/>) verwendet ebenfalls Schema.org/Recipe. Aus (Greene, 2015) (Quellen.html#CRFNYT) geht hervor, dass Cooking.nytimes.com zusätzlich ähnlich wie Chefkoch.de die Zutaten der Rezepte in einer Datenbank abspeichert, aufgespalten nach Name, Mengenangabe, Mengeneinheit, Kommentar und Anderes.