



Extracting recipe ingredients from cookbooks

Irrlichter
von
Torsten Knauf

Der Beginn einer
Master-Arbeit

Irrlichter

Beispiel cueML

```
<cue:recipe type="Suppen." rcp-id="B-16">
<head>Mock Turtle Suppe.</head>

<p>Es wird hierzu für <cue:yield atLeast="24" atMost="30">24 30 Personen</
  cue:yield> eine kräftige <ref target="#Bouillon">Bouillon</ref> von 8
  10 Pfund <cue:recipeIngredient ref="#Rindkochfleisch" atLeast="8"
  atMost="10" unit="Pfund">Rindfleisch</cue:recipeIngredient> mit <cue:
  recipeIngredient ref="#Wurzelwerk">Wurzelwerk</cue:recipeIngredient>
  gekocht. [...]</p>

<note>Anmerk. Der <cue:recipeIngredient ref="#Englische_Soja" optional="True"
>Soja</cue:recipeIngredient> macht die Suppe gewürzreicher, kann jedoch gut
wegbleiben, und statt <cue:recipeIngredient ref="#Madeira" altGrp="1"
>Madeira</cue:recipeIngredient> kann man <cue:recipeIngredient
ref="weißen_Franzwein" altGrp="2">weißen Franzwein</cue:recipeIngredient>
und etwas <cue:recipeIngredient ref="#Rum" altGrp="2" quantity="etwas"
>Rum</cue:recipeIngredient> nehmen<cue:alt target="1_2"/>. Sowohl die Bouillon
als Kalbskopf können schon am vorhergehenden Tage, ohne Nachtheil der
Suppe, gekocht werden.</note>
</cue:recipe>
```

Wie man es hätte auch machen können I

```
und statt <cue:recipeIngredient ref="#Madeira" altGrp="1">Madeira</cue:
recipeIngredient> kann man <cue:recipeIngredient ref="weißen_Franzwein"
altGrp="2">weißen Franzwein</cue:recipeIngredient> und etwas <cue:
recipeIngredient ref="#Rum" altGrp="2" quantity="etwas">Rum</cue:
recipeIngredient> nehmen <cue:alt target="1_2"/>.
```

VS

```
und statt <recipeIngredient ref="#Madeira" xml:id="B-16-Madeira">Madeira</
recipeIngredient> kann man <recipeIngredient ref="weißen_Franzwein" xml:
id="B-16-Franzwein">weißen Franzwein</recipeIngredient> und etwas <
recipeIngredient ref="#Rum" xml:id="B-16-Rum">Rum</recipeIngredient>
nehmen. Sowohl die Bouillon als Kalbskopf können schon am vorhergehenden
Tage, ohne Nachtheil der Suppe, gekocht werden.
```

```
<recipeIngredientGrp xml:id="B-16-alt1" target="#B-16-Madeira"/>
<recipeIngredient xml:id="B-16-alt2" target="#B-16-Franzwein_#B-16-Rum"/>
<alt target="#B-16-alt1_#B-16-alt2" mode="excl"/>
```

Wie man es hätte auch machen können II

```
8 10 Pfund <cue:recipeIngredient ref="#Rindkochfleisch" atLeast="8" atMost="10" unit="Pfund">Rindfleisch</cue:recipeIngredient>
```

VS

```
<atLeast target="B-16-Rindfleisch">8</atLeast> <atMost target="1">10</atMost>  
> <unit target="B-16-Rindfleisch">Pfund</unit> <ingredient xml:id="B-16-Rindfleisch" ref="#Suppenrindfleisch">Rindfleisch</ingredient>
```

Trainingsdaten

für	...	O
24	...	B-Yield
	...	I-Yield
30	...	I-Yield
Personen	...	O
kräftige	...	O
Bouillon	...	O
von	...	O
8	...	B-Quantity
	...	I-Quantity
10	...	I-Quantity
Pfund	...	B-Unit
Rindfleisch	...	B-Ingredient
mit	...	O
Wurzelwerk	...	B-Ingredient
gekocht	...	O
.	...	O

Ergebnis vom Prototyp

wort | wahrheit | getaggt

Nächste Schritte

- Mehr Rezepte labeln
- Label-Engineering
- Feature-Engineering
- Skript um Zutatenliste zu extrahieren



Ahmed, Zeeshan (2009). “Domain Specific Information Extraction for Semantic Annotation”. Diploma thesis. Charles University in Prague, Czech Republic and University of Nancy 2 in Nancy, France.



Schema.org (2016). *Schema.org*. URL: <https://schema.org/> (visited on 10/30/2016).



Skip The Pizza (2012). URL: <http://skipthepizza.com/blog/analyzing-the-ingredients-of-29200-recipes> (visited on 10/28/2016).



Sutton, Charles and Andrew McCallum (2012). “An Introduction to Conditional Random Fields”. In: *Found. Trends Mach. Learn.* 4.4, pp. 267–373. ISSN: 1935-8237. DOI: 10.1561/22000000013. URL: <http://dx.doi.org/10.1561/22000000013>.



The New York Times (2015). *Extracting Structured Data From Recipes Using Conditional Random Fields*. URL:

http://open.blogs.nytimes.com/2015/04/09/extracting-structured-data-from-recipes-using-conditional-random-fields/?_r=1 (visited on 10/30/2016).