Extracting recipe ingredients from cookbooks

Der Beginn einer Master-Arbeit

von
Torsten Knauf

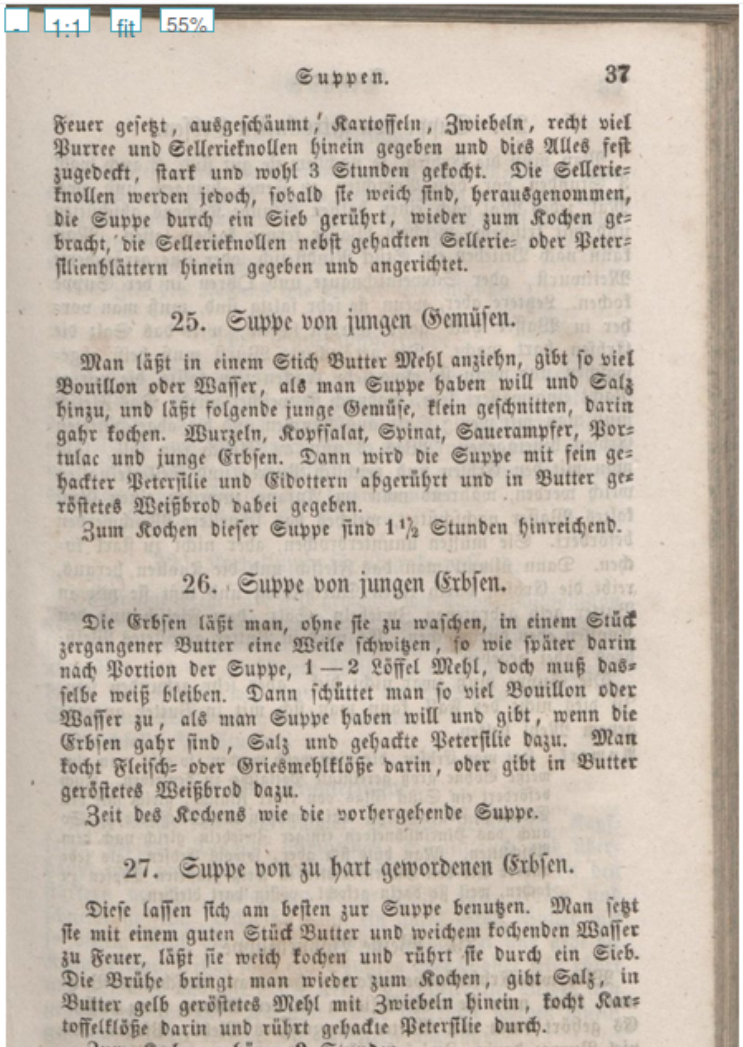
Was und wieso

- Was?
  - Zutaten der Rezepte automatisch extrahieren
- Wieso?
  - Ernährungswissenschaften
  - Soziologische Forschung wie z.B. Globalisierung, Wohlstand, ...

Digitalisierung

(Deutsches Textarchiv 2016)

Bild: 0075 : 37 << vorherige Seite nächste Seite >>



Suppen. 37

Feuer gesetzt, ausgeschäumt, Kartoffeln, Zwiebeln, recht viel Purree und Sellerieknollen hinein gegeben und dies Alles fest zugedeckt, stark und wohl 3 Stunden gekocht. Die Sellerieknollen werden jedoch, sobald sie weich sind, herausgenommen, die Suppe durch ein Sieb gerührt, wieder zum Kochen gebracht, die Sellerieknollen nebst gehackten Sellerie- oder Petersilienblättern hinein gegeben und angerichtet.

25. Suppe von jungen Gemüsen.

Man läßt in einem Stich Butter Mehl anziehn, gibt so viel Bouillon oder Wasser, als man Suppe haben will und Salz hinzu, und läßt folgende junge Gemüse, klein geschnitten, darin gahr kochen. Wurzeln, Kopfsalat, Spinat, Sauerampfer, Portulac und junge Erbsen. Dann wird die Suppe mit fein gehackter Petersilie und Eidottern abgerührt und in Butter geröstetes Weißbrod dabei gegeben.

Zum Kochen dieser Suppe sind 1½ Stunden hinreichend.

26. Suppe von jungen Erbsen.

Die Erbsen läßt man, ohne sie zu waschen, in einem Stück zergangener Butter eine Weile schwitzen, so wie später darin nach Portion der Suppe, 1—2 Löffel Mehl, doch muß dasselbe weiß bleiben. Dann schüttet man so viel Bouillon oder Wasser zu, als man Suppe haben will und gibt, wenn die Erbsen gahr sind, Salz und gehackte Petersilie dazu. Man kocht Fleisch- oder Griesmehlklöße darin, oder gibt in Butter geröstetes Weißbrod dazu.

Zeit des Kochens wie die vorhergehende Suppe.

27. Suppe von zu hart gewordenen Erbsen.

Diese lassen sich am besten zur Suppe benutzen. Man setzt sie mit einem guten Stück Butter und weichem kochenden Wasser zu Feuer, läßt sie weich kochen und rührt sie durch ein Sieb. Die Brühe bringt man wieder zum Kochen, gibt Salz, in Butter gelb geröstetes Mehl mit Zwiebeln hinein, kocht Kartoffelklöße darin und rührt gehackte Petersilie durch.

Suppen.

Feuer gesetzt, ausgeschäumt, Kartoffeln, Zwiebeln, recht viel Purree und Sellerieknollen hinein gegeben und dies Alles fest zugedeckt, stark und wohl 3 Stunden gekocht. Die Sellerieknollen werden jedoch, sobald sie weich sind, herausgenommen, die Suppe durch ein Sieb gerührt, wieder zum Kochen gebracht, die Sellerieknollen nebst gehackten Sellerie- oder Petersilienblättern hinein gegeben und angerichtet.

25. Suppe von jungen Gemüsen.

Man läßt in einem Stich Butter Mehl anziehn, gibt so viel Bouillon oder Wasser, als man Suppe haben will und Salz hinzu, und läßt folgende junge Gemüse, klein geschnitten, darin gahr kochen. Wurzeln, Kopfsalat, Spinat, Sauerampfer, Portulac und junge Erbsen. Dann wird die Suppe mit fein gehackter Petersilie und Eidottern abgerührt und in Butter geröstetes Weißbrod dabei gegeben.

Zum Kochen dieser Suppe sind 1½ Stunden hinreichend.

26. Suppe von jungen Erbsen.

Die Erbsen läßt man, ohne sie zu waschen, in einem Stück zergangener Butter eine Weile schwitzen, so wie später darin nach Portion der Suppe, 1—2 Löffel Mehl, doch muß dasselbe weiß bleiben. Dann schüttet man so viel Bouillon oder Wasser zu, als man Suppe haben will und gibt, wenn die Erbsen gahr sind, Salz und gehackte Petersilie dazu. Man kocht Fleisch- oder Griesmehlklöße darin, oder gibt in Butter geröstetes Weißbrod dazu.

Zeit des Kochens wie die vorhergehende Suppe

```
<div n="3">
    <head>25. Suppe von jungen Gemü&#x017F;en.</head>
    <lb/>
    <p>Man läßt in einem Stich Butter Mehl anziehn, gibt &#x017F;o viel<lb/>
        Bouillon oder Wa&#x017F;&#x017F;er, als man Suppe haben will und Salz<lb/>
        hinzu, und läßt folgende junge Gemü&#x017F;e, klein ge&#x017F;chnitten, darin<lb/>
        gahr kochen. Wurzeln, Kopf&#x017F;alat, Spinat, Sauerampfer, Por-<lb/>
        tulac und junge Erb&#x017F;en. Dann wird die Suppe mit fein ge-<lb/>
        hackter Peter&#x017F;ilie und Eidottern abgerührt und in Butter ge-<lb/>
        rö&#x017F;tetes Weißbrod dabei gegeben.</p> <lb/>
    <p>Zum Kochen die&#x017F;er Suppe &#x017F;ind 1½ Stunden hinreichend.</p>
</div>
```

```
<div n="3">
    <head>25. Suppe von jungen Gemü&#x017F;en.</head>
    <lb/>
    <p>Man läßt in einem Stich Butter Mehl anziehn, gibt &#x017F;o viel<lb/>
        Bouillon oder Wa&#x017F;&#x017F;er, als man Suppe haben will und Salz<lb/>
        hinzu, und läßt folgende junge Gemü&#x017F;e, klein ge&#x017F;chnitten, darin<lb/>
        gahr kochen. Wurzeln, Kopf&#x017F;alat, Spinat, Sauerampfer, Por-<lb/>
        tulac und junge Erb&#x017F;en. Dann wird die Suppe mit fein ge-<lb/>
        hackter Peter&#x017F;ilie und Eidottern abgerührt und in Butter ge-<lb/>
        rö&#x017F;tetes Weißbrod dabei gegeben.</p> <lb/>
    <p>Zum Kochen die&#x017F;er Suppe &#x017F;ind 1½ Stunden hinreichend.</p>
</div>
```

```
<cue:recipe type="Suppen." rcp-id="B-25">
    <head>Suppe von jungen Gemüsen.</head>

    <p>Man läßt in einem Stich Butter Mehl anziehn, gibt so viel Bouillon oder Wasser, als
       man Suppe haben will und Salz hinzu, und läßt folgende junge Gemüse, klein
       geschnitten, darin gahr kochen. Wurzeln, Kopfsalat, Spinat, Sauerampfer, Portulac und
       junge Erbsen. Dann wird die Suppe mit fein gehackter Petersilie und Eidottern
       abgerührt und in Butter geröstetes Weißbrod dabei gegeben. </p>
    <p>Zum Kochen dieser Suppe sind 1½ Stunden hinreichend.</p>

</cue:recipe>
```

Das Tagging

- **cueML** (culinary editions markup language)
  - Erweiterung von **TEI** (Text Encoding Initiative)
  - Erweitert mit **https://schema.org/Recipe**

...nehme man 1 Eßlöffel Mehl ...

...nehme man 1 Eßlöffel
<cue:recipeIngredient quantity="0.01" unit="kg">Mehl</cue:recipeIngredient>
...

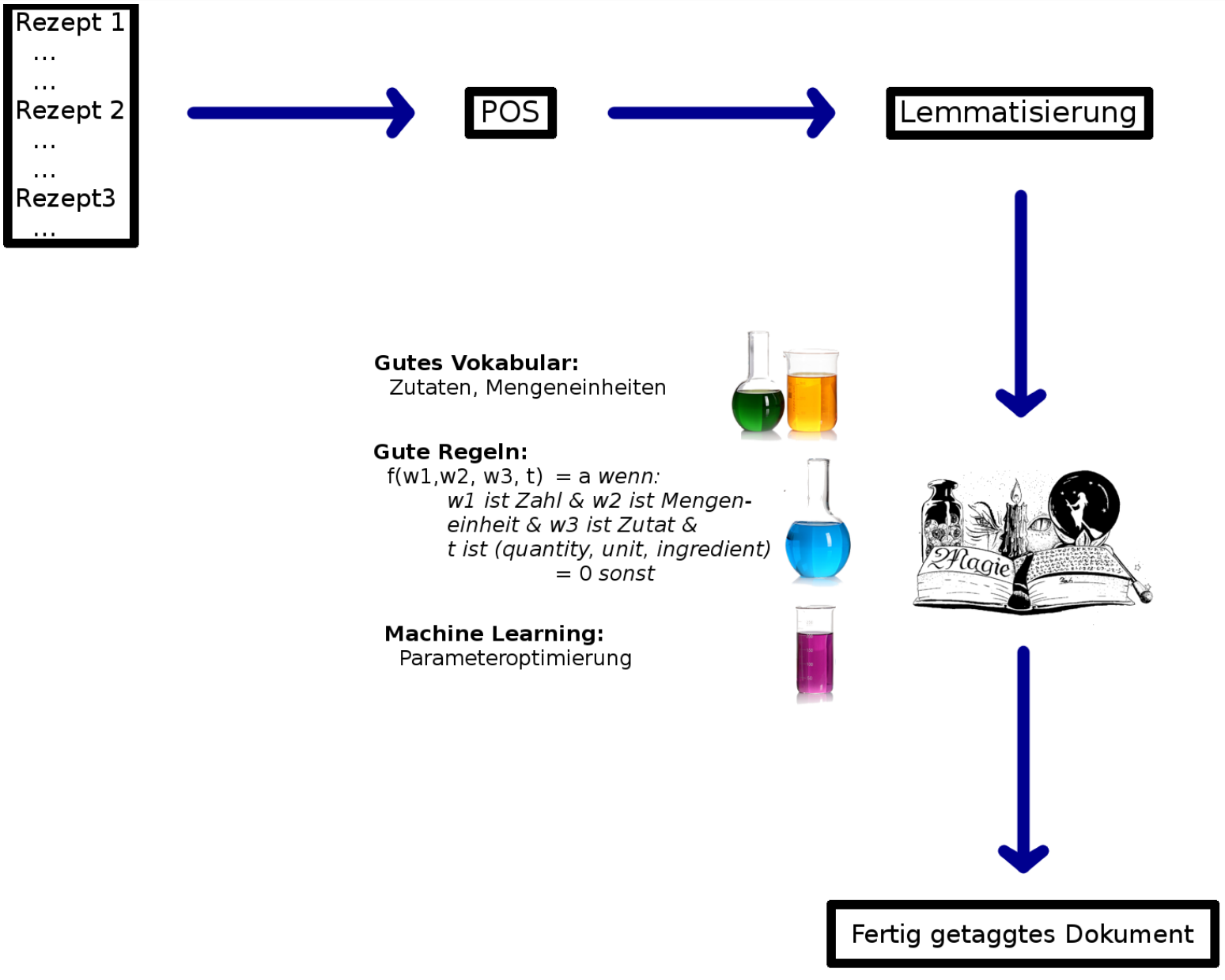
Das Tagging - Achtung

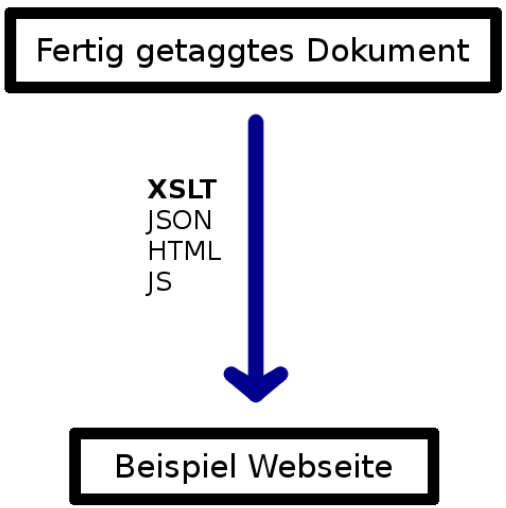
- Scorzoner-Wurzeln, Petersilienwurzel
- 8 Pfennig Weißbrot, Loth, Quart
- *"[Man] legt ihn in einen eisernen Topf, darauf 1/4 Pfund in Scheiben geschnittenen rohen Schinken oder Sommerwurst"* → <recipeIngredientAlt>
- *"Wie die vorhergehende, nur mit weiß gebranntem Mehl und Bouillonzubereitet, die Nelken bleiben weg."*
- *"Fehlt ihm die gewünschte Süße, so wird zeitig ein Stück Zucker dazu gethan so wie beim Anrichten die Brühe mit etwas Kartoffelmehl gebunden gemacht."*
  → <recipeOptionalIngredient>

Ansätze für automatisches Tagging

- Was muss die Lösung alles können?
  - **Aus reinem Text alle Zutaten eindeutig erkennen**
  - Zutaten erkennen **(Lemmatisierung)** und Zuordnen der Quantitäten und Einheiten zu den entsprechenden Zutaten **(Structure prediction for NLP)**
  - Unterscheidung ob eine Zutat verwendet werden soll, nicht verwendet werden soll, optional ist, eine Alternative ist **(Klassifizierungs-Probleme)**
  - Verweise erkennen und verlinken **(Klassifizierungs-Probleme, Structure prediction for NLP)**

- Bestehende Ansätze:
  - Reguläre Ausdrücke (http://skipthepizza.com/ 2016)
  - Kontextfreie Grammatik: (Ahmed 2009)
    $NP \rightarrow DT?JJ^{*}ING$
  - Conditional Random Field (Hamon and Grabar 2013), (Greene 2015)


Rezept 1
...
...
Rezept 2
...
...
Rezept3
...
POS
Lemmatisierung
Gutes Vokabular:
Zutaten, Mengeneinheiten
Gute Regeln:
f(w1,w2, w3, t) = a wenn:
w1 ist Zahl & w2 ist Mengen-
einheit & w3 ist Zutat &
t ist (quantity, unit, ingredient)
= 0 sonst
Machine Learning:
Parameteroptimierung
Magie
Fertig getaggtes Dokument


Fertig getaggtes Dokument
XSLT
JSON
HTML
JS
Beispiel Webseite

Literatursuche ist toll

Aus reinem Text alle Zutaten eindeutig erkennen

- Problem verstehen → Teilgebiet vom Text Mining
- Zerlegung in Teilprobleme → Structure prediction for NLP, Lemmatisierung, Klassifizierungs-Probleme, Conditional Random Field
- "Das Rad [nicht] neu erfinden"
  - Bestehende Lösungen nutzen ist viel produktiver
  - Mehrere unterschiedliche Lösungen / Ideen / Meinungen ergänzen sich oft

Literaturverzeichnis I


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.


Deutsches Textarchiv, Berlin-Brandenburgische Akademie der Wissenschaften (2016). *Davidis, Henriette: Praktisches Kochbuch für die gewöhnliche und feinere Küche. 4. Aufl. Bielefeld, 1849.* URL: http://www.deutschestextarchiv.de/book/show/davidis_kochbuch_1849.


Greene, Erica (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.

Literaturverzeichnis II


Hamon, Thierry and Natalia Grabar (2013). "Extraction of Ingredient Names from Recipes by Combining Linguistic Annotations and CRF Selection". In: *Proceedings of the 5th International Workshop on Multimedia for Cooking & Eating Activities*. CEA '13. Barcelona, Spain: ACM, pp. 63–68. ISBN: 978-1-4503-2392-5. DOI: 10.1145/2506023.2506035. URL: http://doi.acm.org/10.1145/2506023.2506035.


http://skipthepizza.com/ (2016). *Skip The Pizza*. URL: http://skipthepizza.com/blog/analyzing-the-ingredients-of-29200-recipes.