



Extracting recipe
ingredients
from cookbooks

Ritter und Fabelwesen
von
Torsten Knauf

Der Beginn einer
Master-Arbeit

Irrlichter

Heraus aus
dem Sumpf

- ➊ Introduction
- ➋ Making a cookbook machine readable
- ➌ Related Work
- ➍ CRF-based extraction
- ➎ Dictionary- and Rule-based extraction
- ➏ Discussion
- ➐ Summary

2. Making a cookbook machine readable

- ① Digitalisation
- ② CueML ontology
- ③ Need for automation

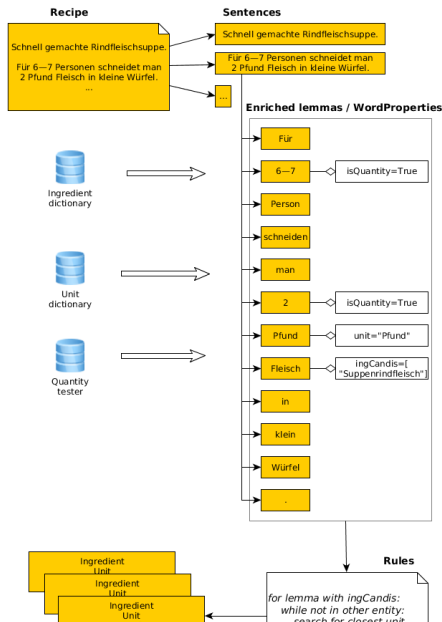
3. Related Work

- ① Skip The Pizza
- ② Extracting Structured Data From Recipes Using Conditional Random Fields
 - ① CRF
 - ② Implementation of NYT
- ③ Domain Specific Information Extraction for Semantic Annotation
- ④ Data-driven Knowledge Extraction for the Food Domain
- ⑤ Lessons for this work

4. CRF-based extraction

- ① CRF prototype
- ② Evaluation

5. Dictionary- and Rule-based extraction



5. Dictionary- and Rule-based extraction

Sentences

Für 6—7 Personen schneidet man
2 Pfund Fleisch in kleine Würfel.

Fleisch
2
Pfund

$$Recall = \frac{\#(retrieved \cap relevant)}{\#relevant}$$

Für 6—7 Personen schneidet man 2 Pfund
<recipeIngredient ref="#Suppenrindfleisch"
quantity="2" unit="Pfund"> **Fleisch**
</recipeIngredient> in kleine Würfel.

Fleisch
2
Pfund

8. Wort

5. Dictionary- and Rule-based extraction

Sentences

Für 6—7 Personen schneidet man 2 Pfund Fleisch in kleine Würfel.

Fleisch
2
Pfund

$$Precision = \frac{\#(retrieved \cap relevant)}{\#retrieved}$$

Recipe

Für 6—7 Personen schneidet man 2 Pfund
<recipeIngredient ref="#Suppenrindfleisch"
quantity="2" unit="Pfund"> **Fleisch**
</recipeIngredient> in kleine Würfel.
[...]
[...] läßt dann das Fleisch [...] rösten [...]

Fleisch
2
Pfund

8th word

possible
ref/target
values

Fleisch

nth word

possible
ref/target
values

5. Dictionary- and Rule-based extraction

Evaluation with recipes B-1 to B-50:

(Only considering ingredients)

- Recall: 0.807 (394/488)
- Precision: 0.833 (434/521)
- Time: 144.7 seconds
- Flaws:
 - Lemmatization (*Saucissen* \nrightarrow *Saucisse*)
 - Is *Brühe* an ingredient?
 - Ingredients within title not tagged

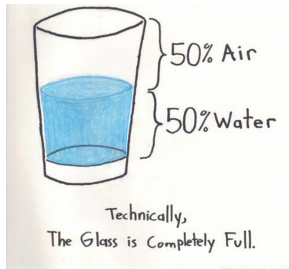
5. Dictionary- and Rule-based extraction

- ① Dictionary- and Rule-based prototype
 - ① Conceptual idea
 - ② Evaluation
- ② Refinement of prototype
 - ① Illustrative enhancements
 - ② Evaluation

6. Discussion

- ① Usefulness of automatic extraction of ingredients
- ② Contribution to field of research
- ③ Quality of cueML and the obtained data
- ④ The development process
- ⑤ Knowledge is power

(8.) Ich bin Realist



Ich glaube, ich kann:

- Eine makellose Zutatenliste für jedes Rezept automatisch extrahieren
- Alle Informationen aus dem Buch extrahieren
- Eine wunderschöne Webseite zum Kochbuch erstellen und mit Inhalt füllen
- Einen Nutellabaum pflanzen

