

cueML - eine kulinarische Textedition

Torsten Knauf, Norbert Luttenberger, Jesper Zedlitz

Arbeitsgruppe Kommunikationssysteme (AG ComSys)

Institut für Informatik

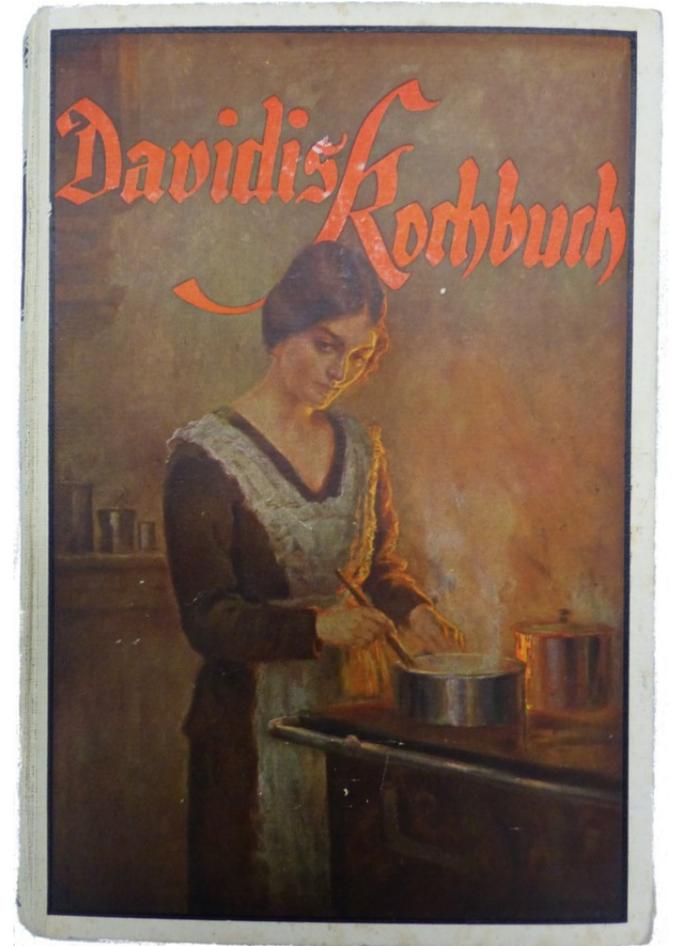
Christian-Albrechts-Universität Kiel

cueML – ein Beispiel
für eine
domänenspezifische
TEI-konforme
Auszeichnungssprache

- Motivation
- Transkription
- TEI
- Auszeichnungssprache cueML
- Anbindung an den BLS
- Automatische Auszeichnung
- Ausblick

Motivation

- Kulinarik als Kulturmerkmal
- Quellen
 - Schulkochbücher
 - verbreitete Kochbücher
- Analyse der Zutaten/
Zubereitungsmethoden
- Artikel

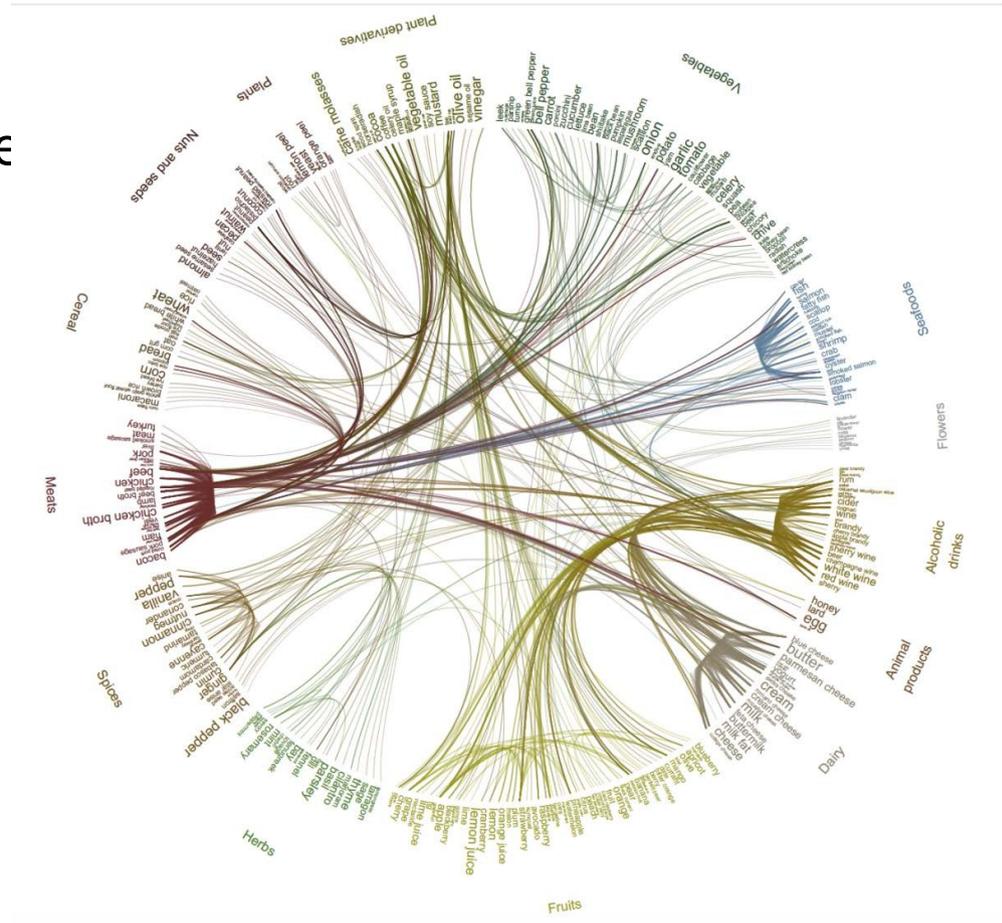


Motivation

- Inspiriert von
Ahn, Ahnert, Bagrow, Barabási:
“Flavor network and the principle
of food pairing”, in: Scientific
Reports 196 (2011)

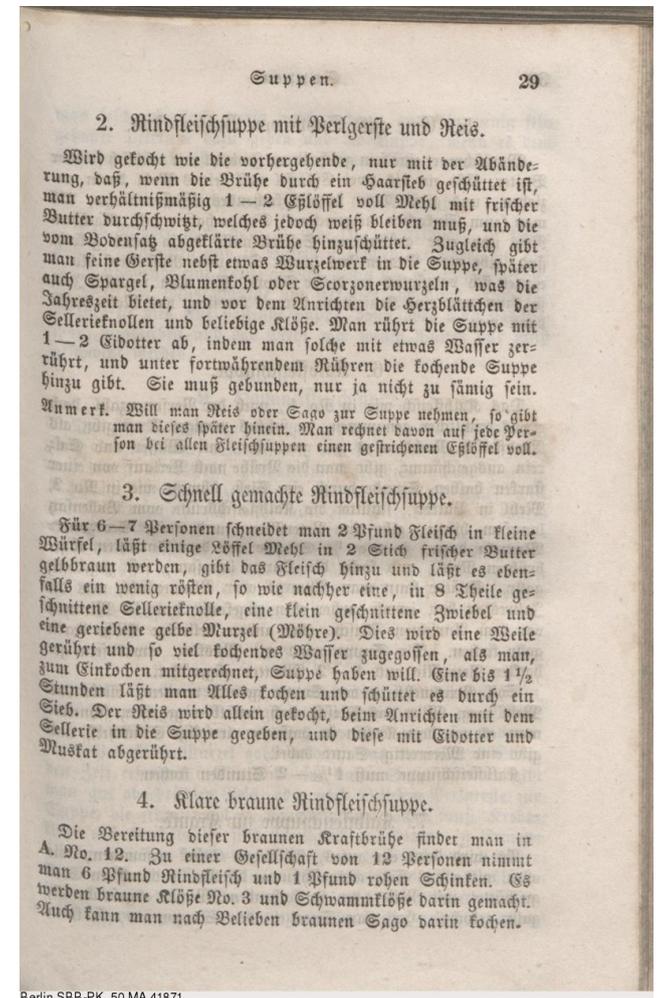
doi:10.1038/srep00196

- Forschungsfrage dort:
Ist das food pairing in
der östlichen anders als
in der westlichen Küche?



Transkription

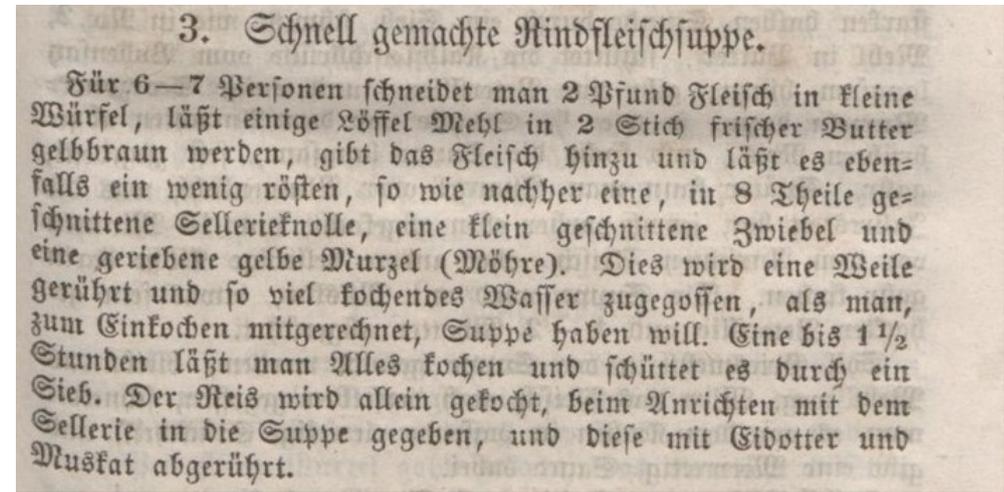
- Erster Schritt: Scannen
- Staatsbibliothek zu Berlin/
Deutsches Textarchiv



Berlin SBB-PK, 50 MA 41871

Transkription

- Scan → TEI
- Deutsches Textarchiv
- Abtippen in China

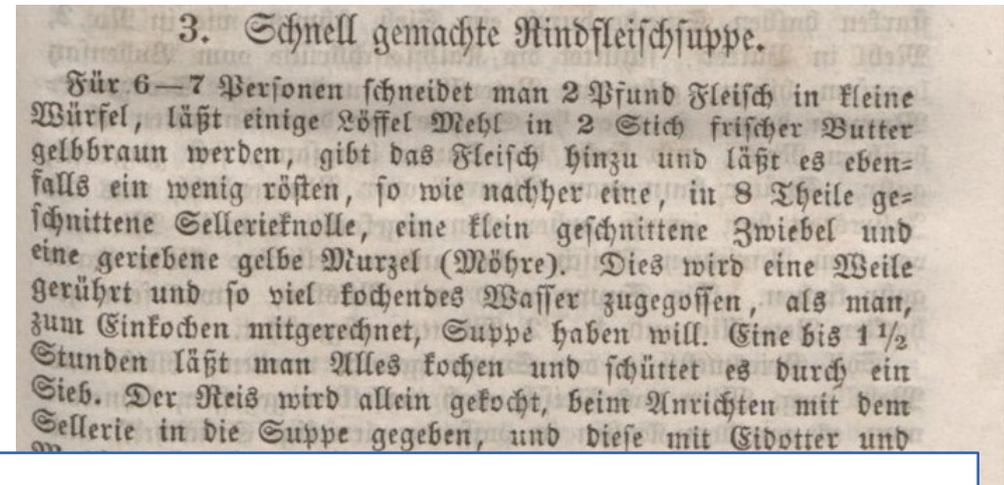


```
<div n="3">
  <head>3. Schnell gemachte
    Rindflei&#x017F;ch&#x017F;uppe.</head>
  <lb/>
  <p>Für 6&#x2014;7 Per&#x017F;onen &#x017F;chneidet
    man 2 Pfund Flei&#x017F;ch in kleine<lb/>
    W&#x017F;fel, lä&#x017F;t einige L&#x017F;ffel Mehl in 2 Stuch
    fri&#x017F;cher Butter<lb/> gelbbraun werden,
    gibt das Flei&#x017F;ch hinzu und lä&#x017F;t es
    eben-<lb/> falls ein wenig rö&#x017F;ten,
    &#x017F;o wie nachher eine, in 8 Theile
    ge-<lb/> &#x017F;chnittene Sellerieknolle, eine
    klein ge&#x017F;chnittene Zwiebel und<lb/> eine
    geriebene gelbe Murzel (Möhre). Dies wird eine
    Weile<lb/> gerührt und &#x017F;o viel kochendes
    Wa&#x017F;er zugego&#x017F;en,
    als man,<lb/> zum Einkochen mitgerechnet, Suppe
    haben will. Eine bis 1½<lb/> Stunden lä&#x017F;t man
    Alles kochen und &#x017F;chüttet es durch
    ein<lb/> Sieb. Der Reis wird allein gekocht,
    beim Anrichten mit dem<lb/> Sellerie in die
    Suppe gegeben, und die&#x017F;e mit Eidotter
    und<lb/> Muskat abgerührt.</p>
</div>
```



Transkription

- Scan → TEI
- Deutsches Textarchiv
- Abtippen in China



```
<?xml version="1.0" encoding="utf-8"?>
<?oxygen RNGSchema="http://www.deutschestextarchiv.de/basisformat.rng" type="xml"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    [...]
    <profileDesc>
      <langUsage>
        <language ident="deu">German</language>
      </langUsage>
      <textClass>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#dtamain">Fachtext</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#dtasub">Gesellschaftswissenschaften</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#dwd1main">Gebrauchsliteratur</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#dwd1sub">Kochbuch</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#DTACorpus">core</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#DTACorpus">ready</classCode>
        <classCode scheme="http://www.deutschestextarchiv.de/doku/klassifikation#DTACorpus">china</classCode>
      </textClass>
    </profileDesc>
  </teiHeader>
</TEI>
```

Suppe gegeben, und dieſe mit Eidotter
und<lb/> Muskat abgerührt.</p>

</div>

Exkurs: TEI

- TEI = Text Encoding Initiative
- Standard zur Kodierung von Texten
- Institution zur Pflege des Standards
- Community rund um den Standard
- seit 2003 auf XML basierend
- RelaxNG-Schema
- modular aufgebaut



TEI P5: Guidelines for Electronic Text Encoding and Interchange

by the TEI Consortium

Originally edited by C.M. Sperberg-McQueen and Lou Burnard for the ACH-ALLC-ACL Text Encoding Initiative
Now entirely revised and expanded under the supervision of the Technical Council of the TEI Consortium

The TEI Consortium
Version 3.0.0. Last updated on 29th March 2016, revision
89ba24e

Text Encoding Initiative Consortium
2016

The TEI Consortium

TEI P5:

Guidelines for Electronic Text Encoding and Interchange

Text Encoding Initiative Consortium 2016

W3C Recommendation



Extensible Markup Language (XML) 1.1 (Second Edition)

W3C Recommendation 16 August 2006, edited in place 29 September 2006

This version:

<http://www.w3.org/TR/2006/REC-xml11-20060816>

Latest version:

<http://www.w3.org/TR/xml11>

Previous version:

<http://www.w3.org/TR/2006/PER-xml11-20060614>

Editors:

Tim Bray, Textuality and Netscape <tbray@textuality.com>

Jean Paoli, Microsoft <jeanpa@microsoft.com>

C. M. Sperberg-McQueen, W3C <cmsmcq@w3.org>

Eve Maler, Sun Microsystems, Inc. <eve.maler@east.sun.com>

François Yergeau

John Cowan <cowan@ccil.org>

Please refer to the [errata](#) for this document, which may include some normative corrections.

- TEI-Dokument bereinigen

```
<div n="3">
  <head>3. Schnell gemachte
    Rindflei&#x017F;ch&#x017F;uppe.</head>
  <lb/>
  <p>Für 6&#x2014;7 Per&#x017F;onen &#x017F;chneidet
    man 2 Pfund Flei&#x017F;ch in kleine<lb/>
    Würfel, läßt einige Löffel Mehl in 2 Stich
    fri&#x017F;cher Butter<lb/> gelbbraun werden,
    gibt das Flei&#x017F;ch hinzu und läßt es
    eben-<lb/> falls ein wenig rö&#x017F;ten,
    &#x017F;o wie nachher eine, in 8 Theile
    ge-<lb/> &#x017F;chnittene Sellerieknolle, eine
    klein ge&#x017F;chnittene Zwiebel und<lb/> eine
    geriebene gelbe Murzel (Möhre). Dies wird eine
    Weile<lb/> gerührt und &#x017F;o viel kochendes
    Wa&#x017F;&#x017F;er zugego&#x017F;&#x017F;en,
    als man,<lb/> zum Einkochen mitgerechnet, Suppe
    haben will. Eine bis 1½<lb/> Stunden läßt man
    Alles kochen und &#x017F;chüttet es durch
    ein<lb/> Sieb. Der Reis wird allein gekocht,
    beim Anrichten mit dem<lb/> Sellerie in die
    Suppe gegeben, und die&#x017F;e mit Eidotter
    und<lb/> Muskat abgerührt.</p>
</div>
```

```
<cue:recipe type="Suppen." rcp-id="B-3">
  <head>Schnell gemachte Rindfleischsuppe.</head>
  <p>Für 6–7 Personen schneidet man 2 Pfund
    Fleisch in kleine Würfel, läßt einige Löffel
    Mehl in 2 Stich frischer Butter gelbbraun
    werden, gibt das Fleisch hinzu und läßt es
    ebenfalls ein wenig rösten, so wie nachher
    eine, in 8 Theile geschnittene
    Sellerieknolle, eine klein geschnittene
    Zwiebel und eine geriebene gelbe Murzel
    (Möhre). Dies wird eine Weile gerührt und so
    viel kochendes Wasser zugegossen, als man,
    zum Einkochen mitgerechnet, Suppe haben will.
    Eine bis 1½ Stunden läßt man Alles kochen und
    schüttet es durch ein Sieb. Der Reis wird
    allein gekocht, beim Anrichten mit dem
    Sellerie in die Suppe gegeben, und diese mit
    Eidotter und Muskat abgerührt. </p>
</cue:recipe>
```

Auszeichnungssprache für Rezepte

AG ComSys

- Anforderungen
 - Überschrift
 - Untertitel
 - Bild
 - Zubereitungszeit
 - Anzahl der Portionen
 - Zutatenliste
 - Zubereitungsanweisung



Pfannekuchen

Einfach und neutral

Bewertung

★★★★★ (93) 04,27

[Rezept bewerten](#)

Verfasser

06onkel

Zutaten

250 g Mehl

4 Ei(er)

Salz

300 ml Milch

300 ml Mineralwasser, (mit Kohlensäure)

Öl

4

Portionen

[Umrechnen](#)

Zubereitung

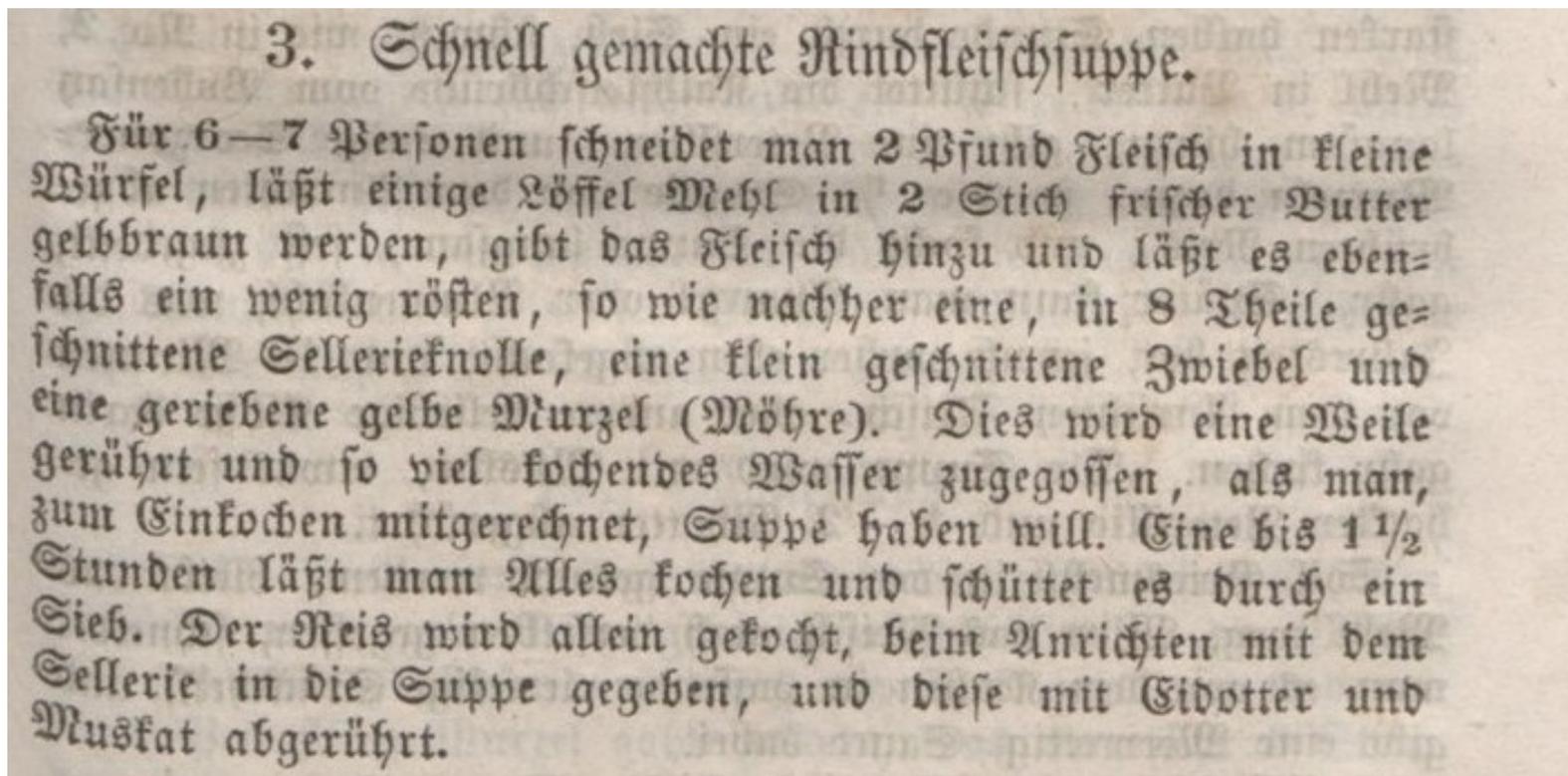
Arbeitszeit: ca. 30 Min. / **Schwierigkeitsgrad:** simpel / **Kalorien**
p. P.: keine Angabe

Den ganzen Teig mit dem Mixer zubereiten (Achtung es spritzt), so dass es ein glatter Teig ohne Klümpchen ist. Eine Pfanne mit etwas Öl heiß werden lassen und dann eine Suppenkelle Teig in die Pfanne geben, dabei die Pfanne ein wenig hin und her schwenken, dass sich der Teig gut verteilen kann, nach einer kurzen Zeit den Pfannekuchen wenden. Fertig!

Dazu passen süße sowie auch herzhafte Sachen, z. B. Nutella, Leberwurst, Konfitüren, Lachs usw.

Auszeichnungssprache für Rezepte

Historische Rezepte sehen anders aus:



Auszeichnungssprache für Rezepte

- Mengeangaben und -einheiten

„schneidet man 2 Pfund Fleisch“

„für 8 Pfennig [...] Weißbrod“

→ Verweis auf standardisierte Verzeichnisse

Auszeichnungssprache für Rezepte

- optionale/alternative Zutaten

„Der [Englische] Soja macht die Suppe gewürzreicher, kann jedoch gut wegbleiben, und statt Madeira kann man weißen Franzwein und etwas Rum nehmen.“

- Verweise

„Man kocht nach No. 1 eine gute Bouillon [...]“

Auszeichnungssprache cueML

- **culinary edition Markup Language**
- eine TEI customization

```
<define name="tei_recipe">
  <element name="recipe" ns="http://cueML/ns">
    <a:documentation>A recipe.</a:documentation>
    <ref name="tei_att.typed.attributes"/>
    <optional>
      <attribute name="rcp-id">
        <a:documentation>Identifies a recipe by giving its
          entry in the original TOC.</a:documentation>
        <text/>
      </attribute>
    </optional>
    <ref name="tei_model.headLike"/>
    <optional><ref name="tei_div"/></optional>
    <optional><ref name="tei_recipe"/></optional>
    <optional><ref name="tei_recipeInstructions"/></optional>
    <zeroOrMore><ref name="tei_p"/></zeroOrMore>
    <zeroOrMore><ref name="tei_ab"/></zeroOrMore>
    <zeroOrMore><ref name="tei_note"/></zeroOrMore>
  </element>
</define>
```

Auszeichnung von Hand

```
<cue:recipe type="Suppen." rcp-id="B-3">
  <head>Schnell gemachte Rindfleischsuppe.</head>

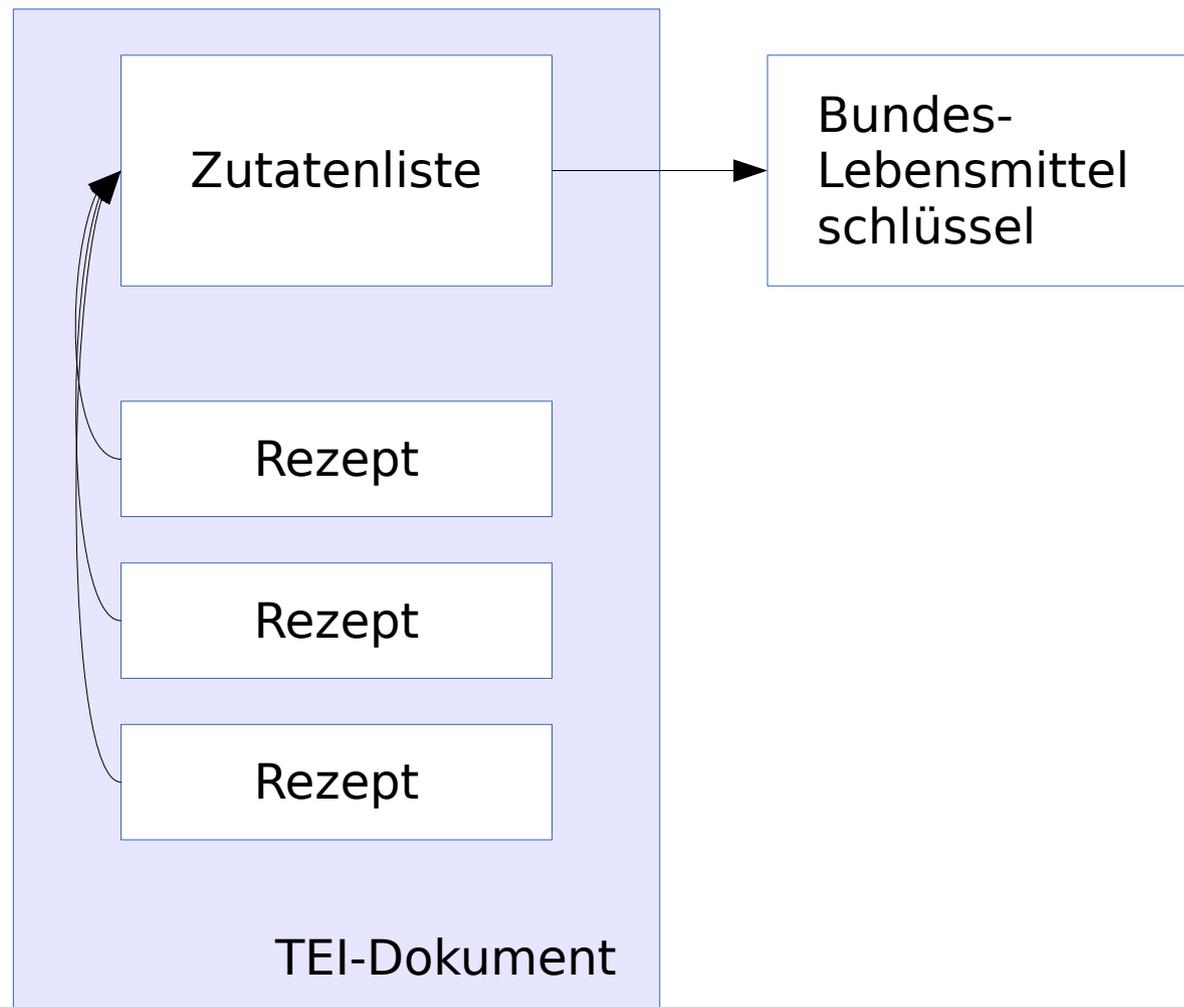
  <p>Für <cue:recipeYield atLeast="6" atMost="7" unit="people">6-7
    Personen</cue:recipeYield> schneidet man 2 Pfund
    <cue:recipeIngredient ref="#Rindkochfleisch" quantity="2"
      unit="Pfund">Fleisch</cue:recipeIngredient> in kleine Würfel,
    läßt einige Löffel <cue:recipeIngredient ref="#Mehl"
      quantity="einige" unit="EL">Mehl</cue:recipeIngredient> in 2
    Stich frischer <cue:recipeIngredient ref="#Butter" quantity="2"
      unit="Stich">Butter</cue:recipeIngredient> gelbbraun werden,
    gibt das Fleisch hinzu und läßt es ebenfalls ein wenig rösten, so
    wie nachher eine, in 8 Theile geschnittene <cue:recipeIngredient
      ref="#Knollensellerie" quantity="1"
    >Sellerieknolle</cue:recipeIngredient>, eine klein
    geschnittene <cue:recipeIngredient ref="#Zwiebel" quantity="1"
    >Zwiebel</cue:recipeIngredient> und eine geriebene gelbe
    Murzel (<cue:recipeIngredient ref="#Karotte" quantity="1"
    >Möhre</cue:recipeIngredient>). Dies wird eine Weile gerührt
    und so viel kochendes Wasser zugegossen, als man, zum Einkochen
    mitgerechnet, Suppe haben will. <cue:cookTime dur-iso="PT1H15M"
    >Eine bis 1½ Stunden</cue:cookTime> läßt man Alles kochen und
    schüttet es durch ein Sieb. Der <cue:recipeIngredient ref="#Reis"
    >Reis</cue:recipeIngredient> wird allein gekocht, beim
    Anrichten mit dem Sellerie in die Suppe gegeben, und diese mit
    <cue:recipeIngredient ref="#Hühnereigelb"
    >Eidotter</cue:recipeIngredient> und <cue:recipeIngredient
      ref="#Muskatnuß">Muskat</cue:recipeIngredient> abgerührt. </p>

</cue:recipe>
```

Anbindung an den BLS

- Bundeslebensmittelschlüssel
- ca. 15.000 Zutaten und Gerichte
- Code für jede Zutat/Gericht
- jeweils über 100 Nährwertangaben
- nur für Forschung frei erhältlich

Anbindung an den BLS



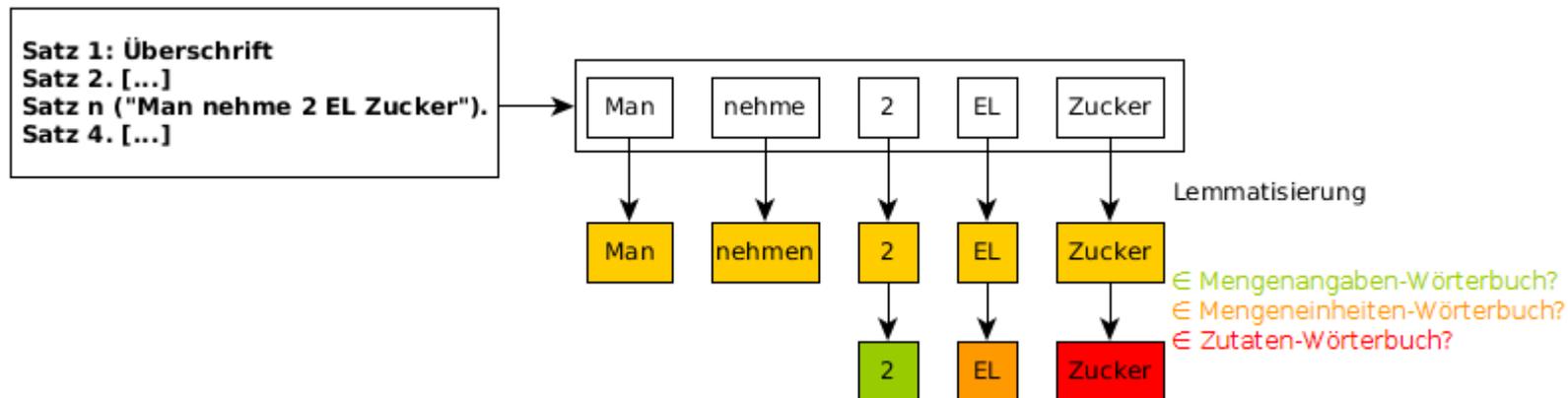
- Zutatenliste

```
<cue:ingredient xml:id="Cayennepfeffer" BLSref="R252000">
  <cue:prefBasicForm>Cayennepfeffer</cue:prefBasicForm>
  <cue:altBasicForm>Cayenne-Pfeffer</cue:altBasicForm>
</cue:ingredient>
<cue:ingredient xml:id="Midder" BLSref="V582100">
  <cue:prefBasicForm>Midder</cue:prefBasicForm>
  <cue:altBasicForm>Kalbsmidder</cue:altBasicForm>
  <cue:altBasicForm>Bries</cue:altBasicForm>
  <cue:altBasicForm>Kalbsmilch</cue:altBasicForm>
  <cue:note>"Kalbsmidder ist auch unter dem Synonym: Bries, Kalbsmilch bekannt. Kalbsmilch ist
  die Thymusdrüse des Kalbes. 100 g frische Kalbsmilch enthalten 99,8 kcal, 3,4 g Fett,
  17,2 g Eiweiß, 77,8 g Wasser, 1,91 mg Eisen, 268 mg Cholesterin und 0,42 g Purine.
  Dieses Organ ist bei Jungtieren voll entwickelt. Bei erwachsenen Tieren bildet sich das
  Organ zurück. Kalbsmilch wird zur Herstellung von Spezialitäten, wie Suppen, Klöße und
  Ragout, verwendet." (http://www.cosmiq.de/qa/show/70827/was-ist-kalbsmidder/)</cue:note>
</cue:ingredient>
<cue:ingredient xml:id="Saucisse" BLSref="W000000">
  <cue:prefBasicForm>Saucisse</cue:prefBasicForm>
  <cue:note>Franz. Name für bestimmte Würste (s.
  https://fr.wikipedia.org/wiki/Saucisse)</cue:note>
</cue:ingredient>
<cue:ingredient xml:id="Butter" BLSref="Q610000">
  <cue:prefBasicForm>Butter</cue:prefBasicForm>
</cue:ingredient>
```

- Ansätze in der Literatur
 - Reguläre Ausdrücke [Skip The Pizza, 2012]
 - Conditional Random Fields [Greene, 2015]
 - Wörterbuch- und regelbasiert
[Zeeshan, 2009] und [Wiegand et al., 2012]
- hier: Wörterbuch- und regelbasiert

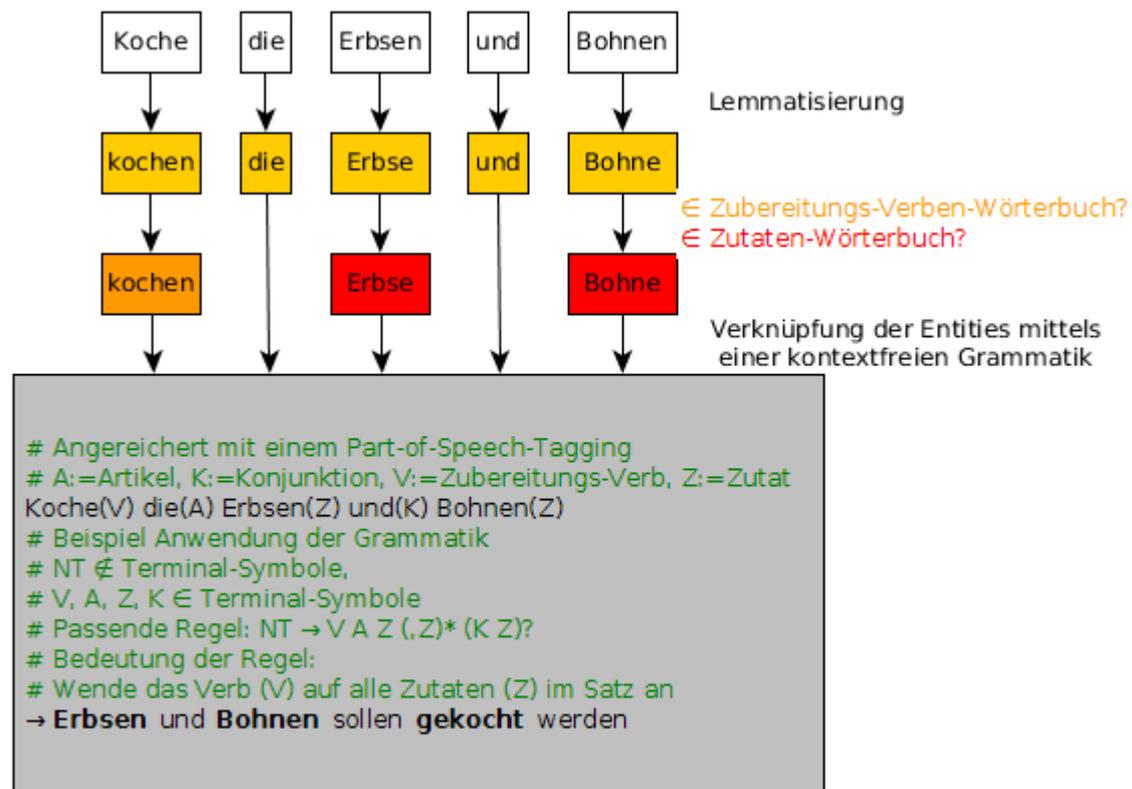
Automatische Auszeichnung

- Wörterbuch-basiert



Automatische Auszeichnung

- Regelbasiert



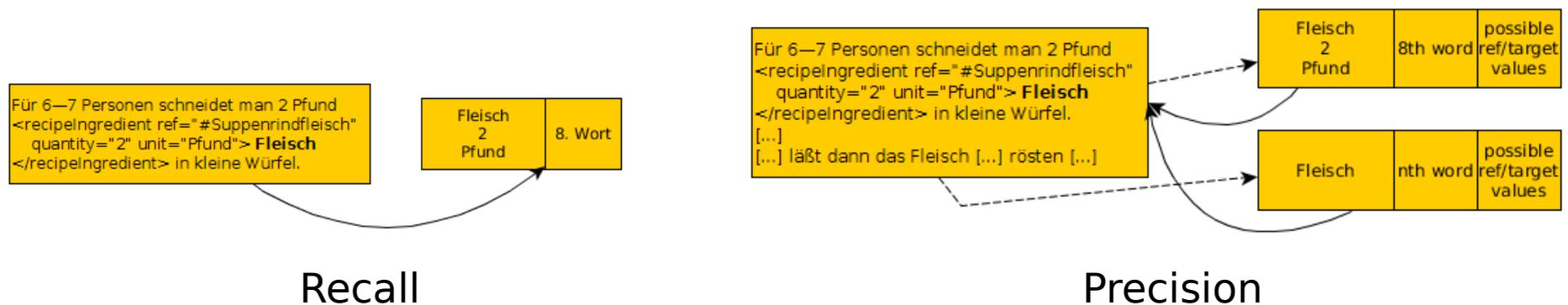
Automatische Auszeichnung

- Ergebnis der automatischen Auszeichnung

```
<cue:recipe rcp-id="B-3" type="Suppen" xmlns:cue="http://cueML/ns">
  <head> Schnell gemachte Rindfleischsuppe.</head>
  <p> Für 6–7 Personen schneidet man 2 Pfund <cue:recipeIngredient
    quantity="2" ref="#Rindkochfleisch" unit="Pfund"
    >Fleisch</cue:recipeIngredient> in kleine Würfel, läßt
    einige Löffel <cue:recipeIngredient quantity="einige"
    ref="#Mehl" unit="EL">Mehl</cue:recipeIngredient> in 2
    Stich frischer <cue:recipeIngredient quantity="2"
    ref="#Butter" unit="Stich">Butter</cue:recipeIngredient>
    gelbbraun werden, gibt das <cue:recipeIngredient
    ref="#Rindkochfleisch">Fleisch</cue:recipeIngredient>
    hinzu und läßt es ebenfalls ein wenig rösten, so wie nachher
    eine, in 8 Theile geschnittene <cue:recipeIngredient
    quantity="8" ref="#Knollensellerie"
    >Sellerieknolle</cue:recipeIngredient> , eine klein
    geschnittene <cue:recipeIngredient quantity="1"
    ref="#Zwiebel">Zwiebel</cue:recipeIngredient> und eine
    geriebene gelbe Murzel( <cue:recipeIngredient quantity="1"
    ref="#Karotte">Möhre</cue:recipeIngredient> ). Dies wird
    eine Weile gerührt und so viel kochendes Wasser zugegossen,
    als man, zum Einkochen mitgerechnet, Suppe haben will. Eine
    bis 1½ Stunden läßt man Alles kochen und schüttet es durch
    ein Sieb. Der <cue:recipeIngredient ref="#Reis"
    >Reis</cue:recipeIngredient> wird allein gekocht, beim
    Anrichten mit dem <cue:recipeIngredient
    ref="#Knollensellerie #Sellerieblätter"
    >Sellerie</cue:recipeIngredient> in die Suppe gegeben,
    und diese mit <cue:recipeIngredient ref="#HühnereiEigelb"
    >Eidotter</cue:recipeIngredient> und
    <cue:recipeIngredient ref="#Muskatnuß"
    >Muskat</cue:recipeIngredient> abgerührt. </p>
</cue:recipe>
```

Automatische Auszeichnung

- Evaluation ist schwierig
- Ground-truth von Hand erstellen
- Was sind recall und precision?



Automatische Auszeichnung

- Evaluation

	Recall	Precision
nur Zutaten	0.986	0.982
+ Mengeangaben und Einheiten	0.848	0.837
+ optionale/alternative Zutaten	0.713	0.728

- TEI → HTML
 - z.B. mit XSLT

Rezepte

Saucen	▼
Suppen	▼
Klare weiße Rindfleischsuppe.	
Rindfleischsuppe mit Perlgerste und Reis.	
Schnell gemachte Rindfleischsuppe.	
Klare braune Rindfleischsuppe.	
Eine gute Suppe von Braten-Jus.	
Kalbfleischsuppe.	
Kalbfleischsuppe für Kranke.	

Schnell gemachte Rindfleischsuppe.

Zutatenliste

- 2 Pfund Fleisch
- einige EL Mehl
- 2 Stuch Butter
- 1 Sellerieknolle
- 1 Zwiebel
- 1 Möhre
- Reis
- Eidotter
- Muskat

Zubereitung

Für 6—7 Personen schneidet man 2 Pfund Fleisch in kleine Würfel, läßt einige Löffel Mehl in 2 Stuch frischer Butter gelbbraun werden, gibt das Fleisch hinzu und läßt es ebenfalls ein wenig rösten, so wie nachher eine, in 8 Theile geschnittene Sellerieknolle , eine klein geschnittene Zwiebel und eine geriebene gelbe Murzel (Möhre). Dies wird eine Weile gerührt und so viel kochendes Wasser zugegossen, als man, zum Einkochen mitgerechnet, Suppe haben will. Eine bis 1½ Stunden läßt man Alles kochen und schüttet es durch ein Sieb. Der Reis wird allein gekocht, beim Anrichten mit dem Sellerie in die Suppe gegeben, und diese mit Eidotter und Muskat abgerührt.

- verbesserte automatische Extraktion
 - Zeitangaben
 - Zubereitungsmethoden
 - Verweise auf Teil-Rezepte
 - Beilagen
- kulinarische Analyse

Vielen Dank für
die Aufmerksamkeit!