Meine erste Woche ist quasi rum und hier nun meine gewonnenen Erkenntnisse: Unten stehend ist beispielsweise das erste Rezept (eingefügt übrigens über notepad++ -> Erweiterungen NppExport -> copy RTF to clipboard und vorher mittels „str j“ die Zeilen gejoint sowie 4 Leerzeilen durch eins ersetzt und dann eine einzellige/-spaltige Tabelle drum)

|  |
| --- |
| <cue:recipe type=**"Suppen."** rcp-id=**"B-1"**>  <head>**Klare weiße Rindfleischsuppe.**</head>  <p>**Man rechnet, wenn die Gesellschaft klein ist, auf jede Person ¾ Pfund, bei einer großen Gesellschaft ½ Pfund Fleisch. Dasselbe wird gewaschen und mit doppelt so vielem Wasser, als man Suppe haben will, zu Feuer gesetzt. Auf jede Person rechnet man bei einem Gesellschaftsessen ⅜ Maß Suppe, also auch ¾ Maß Wasser, wozu man gleich das nöthige Salz gibt, weil das Fleisch nur in der ersten Stunde des Kochens das Salz annimmt und das spätere Hinzugeben des Salzes nur die Suppe würzt. Ist das Fleisch sorgfältig ausgeschäumt, gibt man eine fein geschnittene Zwiebel hinzu, wodurch die Brühe an Kraft gewinnt, und läßt sie, fest zugedeckt, fortwährend, aber nicht zu stark kochen. Eine Stunde später gießt man die Brühe durch ein Haarsieb, weil sie nie ganz klar ist, spült das Stück Fleisch eben ab und setzt es mit der Brühe, die man vom Bodensatz langsam abschüttet, in dem ebenfalls umgespülten Topfe wieder zu Feuer nebst einigen Scorzoner-, einer Sellerie- und Petersilienwurzel. Hiermit läßt man die Suppe ferner kochen, während man zuweilen das Fett abnimmt, das der Bouillon den feinen Geschmack benimmt. Man kann ein halbes Stündchen vor dem Anrichten Spargel oder Blumenkohl, was die Jahreszeit gerade bietet, hinein geben, nur muß beides eben abgekocht sein, und darf ja nicht in der Bouillon zerkochen. Ist das Fleisch gahr, aber nicht zu weich gekocht, läßt man beliebige Klöße eben darin gahr kochen und richtet die Suppe an. Zugleich hat man Reis, für jede Person einen Eßlöffel gerechnet, abgebrüht und in Bouillon, wozu man so viel mehr Fleisch genommen, gahr und dick gekocht; doch müssen die Körner ganz bleiben. Derselbe wird bei der Suppe (nach Belieben mit geriebenem Parmesankäse) herumgereicht.** </p>  <p>**Rindfleischsuppe muß vom Kochen an, 3—3½ Stunden kochen.**</p>  </cue:recipe> |

Angetreten bin ich mit der (naiven) Hoffnung, einen super schlauen Parser schreiben zu können, der automatisch um die Zutaten ein <cue:recipeIngredient>–Tag schließt. Doch schon das erste Rezept belehrte mich eines Besseren:

Fleisch muss später als Rindfleisch erkannt werden, da Rindfleischsuppe vom Schwein wahrscheinlich nicht gemeint ist. Für die Mengenangabe habe ich den Mittelwert von **¾** und **½ Pfund** genommen. **Scorzoner- […]** **Petersilienwurzel** muss als Scorzonerwurzeln und Petersilienwurzel erkannt werden. Ersteres ist nicht mit *Scorzonera* (teuren Sommer-Trüffeln) zu verwechseln, sondern wahrscheinlich *Scorzonera hispanica* (Schwarzwurzeln). Zweiteres ist nicht etwa Petersilie, sondern ich kenne es unter Wurzelpetersilie.

Eine weitere Schwierigkeit ergibt sich bei Sachen wie. **[Man]legt ihn in einen eisernen Topf, darauf ¼ Pfund in Scheiben geschnittenen rohen Schinken oder Sommerwurst**. In der Zutatenliste sollte nicht getrennt voneinander roher Schinken und Sommerwurst auftreten, da nur eins von beiden benötigt wird. Daher habe ich ein <cue:recipeIngredientAlt>-Tag hinzugefügt.

|  |
| --- |
| <cue:recipeIngredientAlt><cue:recipeIngredient quantity=**"0.1"** unit=**"kg"**>**rohen Schinken**</cue:recipeIngredient>**, oder** <cue:recipeIngredient>**Sommerwurst**</cue:recipeIngredient></cue:recipeIngredientAlt> |

Damit das XML-Dokument wieder valide ist, muss nun die *relax ng*-Grammatik angepasst werden. Herr Luttenberger hat mich glücklicherweise rechtzeitig darauf hingewiesen, dass die *cueML.rng*-Grammatik automatisch erstellt wurde/wird und dass es daher sehr ungünstig ist, die Datei manuell zu ändern, falls sie (automatisch) geändert wird. Daher habe ich die Grammatik in *cueML\_with\_extension.rng* erweitert.

Zurück zu dem <cue:recipeIngredientAlt>-Tag. Was macht man nun bei sowas wie **[…] und schüttet mit Zitronenschale, Charlotten oder Zwiebeln […]**. Sind eins von den 3 oder Ersteres und Charlotten oder Zwiebeln gemeint? Schwierig – da ich mir Zitrone mit Zwiebeln nicht so gut vorstelle, habe ich mich für die erste Variante entschieden, aber sicher bin ich mir nicht.

Ein weiteres Problem wurde mir bei **mit geschmolzener Butter, Senf, fein gehackter Petersilie (jedoch jedes besonders) oder mit einer sauern Eier-Sauce und Kartoffeln zur Tafel gegeben** klar. Es ist entweder Butter, Senf, Petersiele oder Eier-Sauce gemeint und NICHT Butter oder Senf oder Petersiele oder Eier-Sauce. Als Workaround finde ich <ab>-Blöcke (anonym block) ausreichend.

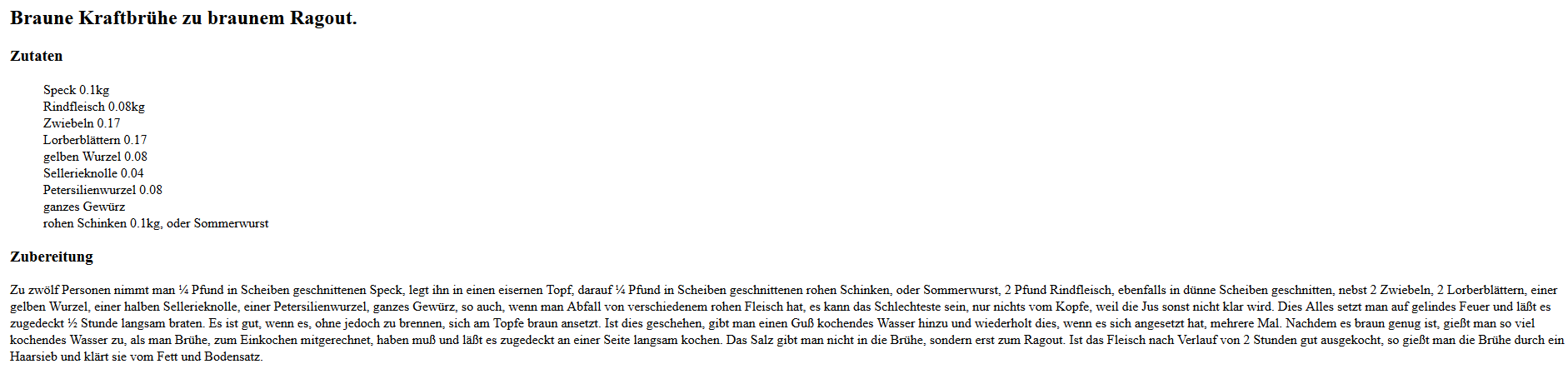
|  |
| --- |
| **mit geschmolzener** <cue:recipeIngredientAlt><ab><cue:recipeIngredient >**Butter**</cue:recipeIngredient>**,** <cue:recipeIngredient>**Senf** </cue:recipeIngredient>**, fein gehackter** <cue:recipeIngredient>**Petersilie** </cue:recipeIngredient> **(jedoch jedes besonders)**</ab> **oder mit einer** <ab><cue:recipeIngredient>**sauern Eier-Sauce**</cue:recipeIngredient><link target=**"#P-19 #P-20"** /></ab></cue:recipeIngredientAlt> **und** <cue:recipeIngredient>**Kartoffeln**</cue:recipeIngredient> **zur Tafel gegeben.** |

Gerade fällt mir dabei ein: Die cueML-Grammatik selber um ein Tag (<cue:recipeIngredient­Alt>) zu erweitern ist zwar relativ leicht und geht relativ schnell. Allerdings ist es dann nicht mehr Standardkonform. Wahrscheinlich sollte man den Machern von cueML das Problem schildern. Sie haben vielleicht bereits ein Workaround für das Problem oder könnten den Standard für alle erweitern.

**Also** habe ich die ersten drei Tage Parser „gespielt“ und nichts Anderes gemacht, außer Zutaten getagt – yay, was nen Spaß. Dieser wird auch nicht dadurch erhöht, dass die Rechtschreibung sowie Grammatik aus heutiger Sicht gewöhnungsbedürftig sind. Ich hoffe sehr, dass ich mir sie nicht ausversehen zu eigen mache.

**Gott sei Dank** kam Herr Luttenberger gestern zu mir und hat mir die Pläne für das Projekt/meine Arbeit genauer vorgestellt:

Als Ergebnis davon habe ich eine erste Bastelei / Vertraut machen mit XSLT / Plausibilitätsüberprüfung in *recipes extracted cueML to html.xsl* gebastelt – eine willkommene Abwechslung, auf die ich mich mit Freude gestürzt hatte. Das Ergebnis sieht wie folgt aus:



Eine Spielerei, die ich schon viel früher hätte machen sollen. Ganz im Sinne der iterativen Entwicklung 1-3 Rezepte tagen und dann überprüfen, wie das Ergebnis aussieht. Ich bilde mir ein, dass das Tagen mit dem Wissen, dass man die Daten wie folgt sinnvoll extrahieren kann, viel befriedigender ist – diese These werde ich demnächst validieren können, da es noch mehr als genug Rezepte zu tagen gibt. Des Weiteren ist mir so aufgefallen, dass das <cue:recipeIngredient­Alt>–Tag benötigt wird. Vorher hatte ich nur das Gefühl, dass es nicht ganz stimmt – jetzt hingegen bin ich, wenn wahrscheinlich auch nur vorerst, wieder zufrieden mit den Tags. Das *vorerst* ist aber nicht schlimm, denn wenn ich (oder wer anders) nicht mehr zufrieden ist, ist das ein Zeichen davon, dass man etwas gelernt hat. Aus meiner Sicht der Traum oder zumindest das Ziel von jedem Studenten! **\*Notiz an mich selbst: Das nächste Mal einfach selber auf die Idee kommen einmal so eine Plausibilitäts-Spielerei durchzuführen\***

Was ich noch vergessen hatte, aber sicherlich dokumentationswürdig ist, ist die Datei *Umrechnung quantity.txt*. Dort habe ich Umrechnung wie *1 Stich Butter = 10g* oder *1 EL Mehl = 10 g* vorgenommen, damit ich diese nicht immer neu suchen muss und sie auch einheitlich sind.

Was mich demnächst beschäftigen wird:

* Natürlich mehr Zutaten tagen – jetzt wo ich jedoch gesehen habe / selbst nachvollzogen habe, dass das jetzige Schema Sinn und Verstand hat und ich Abwechslung zwischen durch habe, eine akzeptable Tätigkeit (man möge mir mein Misstrauen verzeihen, dass ich Sachen erst glaube, wenn ich sie selbst nachvollzogen habe – wir sind hier ja in einer Universität und nicht in der Kirche :-P ).
* Mich mit der geplanten Webseite genauer beschäftigen. Das Ziel ist eine formvollendete wie [Vulkaneifelhaus.com](http://www.vulkaneifelhaus.com/#!haus/c1070), welche optional mit [Wix](https://www.wix.com/) gebaut werden kann.
* Beim Tagging Ungereimtheiten zu entdecken/verbessern. Z. B. sind mir folgende Sachen noch aufgefallen:
* Das Rezeot P-3 sieht im wesentliche, wie folgt aus: **Wie die vorhergehende, nur mit weiß gebranntem Mehl und Bouillonzubereitet, die Nelken bleiben weg.** Mit einem link zum vorherigen könnte man die Zutaten von dort übernehmen. Allerdings würden dann auch Nelken auftauchen, was ja explizit nicht geplant ist. Dementsprechend bräuchte man eine Möglichkeit, gelinkte-Zutaten zu entfernen. Dies wäre zum Beispiel über ein *exclude*-Attribute im *link*-Tag möglich.
* Ich habe mich bei dem Beispielfall aus Rezept P-33 **junge Sahne, in Ermangelung derselben frische Milch** dazu entschieden, Milch nicht als Zutat zu tagen, da es nur eine Ersatz-Zutat ist. Allerdings geht dadurch die Information verloren, dass Sahne augenscheinlich eine nicht selbstverständlich verfügbare Zutat darstellt und man als (billigere?) Alternative mit Milch gekocht hat. Gerade fällt mir auf, dass das bereits eingeführte <cue:recipeIngredient­Alt>–Tag dafür eine gar nicht so schlechte Lösung wäre.
* Das XSLT zieht zurzeit die Zutaten nur aus den <p>-Tags. Manchmal steht eine Zutat allerdings nur im <head>-Tag und ist deswegen dort getaggt. Das XSLT muss dahingehend noch angepasst werden.
* P-6 oder P-61 sind jeweils 2 unterschiedliche Rezepte innerhalb einem *recipe*-Tag. Eine Zutatenliste, welche beide Rezepte umschließend, finde ich irgendwie nicht akkurat.
* Mengenangabe 8 Pfennig Weißbrot verstehen

Also insgesamt sehr viel tun und neben dem (stupidem) Tagging auch viel spannende Abwechslung. Insbesondere freue ich mich auf das Basteln der Webseite und das zusammentragen der Rezepte / „Ernten des Taggings“ – Mal schauen wie weit ich komme. In diesem Bereich könnte ich mir auch eine spannende Master-Arbeit vorbei ☺