Sehr geehrter Herr Luttenberger,

gerne würde ich noch einmal ihre Meinung zum Tag-Schema hören. Anbei sind 3 Versionen mit jeweils den gleichen zwei getagten Rezepten, die ich folgend kurz erläutern möchte:

„alter Stiel.xml“:  
Im wesentlichem das ursprüngliche Tag-Schema, mit der Abänderung, dass ich als *unit*-Zusatz *pP* (per person) hinzugefügt habe und statt dem ISO 8691 Zeitformat *quantity* und *unit* verwende, da man so eine variable Zeitspanne von 3-3.5 Stunden darstellen kann.

Daran stören mich zwei Dinge, wobei ich Ihnen glauben würde, dass das unwichtig ist, da Sie deutlich mehr Erfahrung als ich haben:  
1. Für mich suggeriert das einschließende *ingredient*-Tag, dass alle relevanten Informationen eingeschlossen sind, was nicht der Fall ist. Z.B. in *<ingredient name="Selleriewurzel" quanitity="1">Sellerie-</ingredient>* aus Rezept B-1, geht aus „Sellerie“ nicht hervor, dass „wurzel“ abgekürzt ist, und die *unit* und *quantity* auch nicht.  
2. Auch wenn ich weiß, dass es das Modell sehr verkompliziert, finde ich ein Weglassen der Unterscheidungen, ob eine Zutat eine Alternative zu einer anderen ist, oder optional ist, bei Frau Davidis Kochbuch eine zu starke Vereinfachung. Nachträgliche Ergänzungen wie „statt Madeira kann man weißen Franzwein und etwas Rum nehmen“ in B-16 können so nicht wirklich erfasst werden.

„mit ingredientPhrases.xml“:  
Bezüglich Erstens ist nun um den gesamten erläuterten Teil einer Zutat ein *ingredientPhrase*-Tag. Um die meiner Meinung nach falsche Suggestion ganz zu entkräften, sollte das *ingredient*-Tag wahrscheinlich auch leer sein und an der früh möglichsten Stelle stehen, wo alle Informationen im Text bereits standen, anstatt um die Zutat herum. Dies würde auch die menschliche Lesbarkeit etwas erhöhen.  
Bezüglich Zweitens stellt das keine Verbesserung dar.

„mit metaElement.xml“:  
Auch wenn Sie bei meinem Vortrag am Mittwoch gar nicht von dem *meta*-Element begeistert waren, habe ich mir noch einmal darüber Gedanken gemacht. Aus meiner Sicht hat das folgende Vorteile:  
- Es ist leicht zu lesen. Als „Entwickler“ finde ich das sehr angenehm, aber für ein Schema ist das vielleicht kein valides Argument, da es ja von einem Computer gelesen werden soll.  
- Der zweite Punkt der nachträglichen Kommentare  stellt kein Problem mehr dar.  
- Ich denke, dass ich meine Extraktions-Algorithmen wie folgt evaluieren möchte: Ich vergleiche die extrahierte Zutatenliste, gegen ein Modell der echten Zutatenliste (ein Beispiel Modell wäre, dass das *isOptional*-Attribute ignoriert wird) und gegen die echte Zutatenliste durch eine Metrik. Ein Beispiel für so eine Metrik, die mir im Kopf schwebt, ist:

* 1 Strafpunkt für einen Fehler gegen das Modell (wie eine Zutat extrahiert, die keine ist)
* 1/4 Strafpunkt für einen systematischen Fehler des Modells (wie Zutat extrahiert, obwohl es eine optionale Zutat ist, wenn das *optional*-Attribut im Modell ignoriert wird)

Die Evaluation gegen eine bereits vorhandene Zutatenliste ist trivial. Um aus den Tags im Text eine Zutatenliste zu extrahieren, braucht man hingegen einen Algorithmus, der trivial ist, wenn man nur Zutaten betrachtet, aber nicht trivial ist, wenn man optionale und alternative Zutaten berücksichtigen möchte.

Wenn Sie mir jetzt allgemein sagen, dass optionale und alternative Zutaten zu berücksichtigen zu ehrgeizig ist, fände ich es zwar schade, da ich denke, dass das den Wert der Auswertung von Frau Davidis Buch einschränkt, aber glaube Ihnen das natürlich.

Ich freue mich auf ihre Meinung und hoffe, dass ich das Tag-Schema anschließend für mich endlich als vorläufig abgearbeitet abhacken kann;  
mit freundlichen Grüßen,  
Torsten knauf