# Der erste Monat ist um und von den 150 Stunden habe ich bereits 75 geleistet (Viele Stunden gingen auch in die Webseite „Todesopfer durch Luftangriffe“ – Eine schicke Webseite, an der ich viel JavaScript lernen durfte ☺ ). Zeit einmal zusammenzutragen, auf was für Schwierigkeiten ich beim Kochbuch gestoßen bin und darüber zu sinnieren, wie man sinnvoll weitermacht.

Zunächst einmal Punkte, bei denen ich mir unsicher bin, ob ich sie durch mein Tagen korrekt erfasst habe:

* Ich bin mir unsicher, ob Beilagen für Zutaten wichtig sind. Ich tendiere dazu, dass sie frei wählbar sind und daher nicht zur Zutatenliste beitragen. Also habe ich sie bis jetzt nicht getagt. Falls wir sie doch tagen wollen, sind „*Beilage“* und „*dabei“* gute Stichwörter um im Text nach ihnen zu suchen.
  + Vielleicht ist ein eigenes **Tag für Beilagen** sinnvoll. Ich bin überrascht, dass es das nicht im *schema.org/Recipe* zu geben scheint.
* Bis jetzt habe ich beispielsweise *drei bis vier Eier* wie folgt getagt: *drei bis vier <cue:recipeIngredient unit="3.5">Eier</cue:recipeIngredient>*. Damit ist die Information, dass es 3-4 Eier sind, nicht mit eingetagt. Es wie folgt zu machen, finde ich im Nachhinein besser: *<cue:recipeIngredient unit="3.5">drei bis vier Eiern</cue:recipeIngredient>*. Auch könnte man sich überlegen *unit="3-4"* zu verwenden.
* Ich habe Zweifel, ob *cue:cookTime* später richtig ausgewertet werden kann. Bei Rezept D-58 steht beispielsweise *wenigstens 2 Stunden*, was ich schwer zu quantifizieren finde…

Oder bei Rezept P-64 habe ich ein „einsames“ *<cue:cookTime dur-iso="PT15M">¼ Stunde</cue:cookTime>* getagt, welches später nochmal durchgekocht wird. Die gesamte Kochzeit ist also länger als die ¼ Stunde. Mir ist auch nicht klar, wie die Zugehörigkeit der ¼ Stunde maschinell zuordenbar sein soll.

* Beispielsweise bei Rezept D-58 steht *Fehlt ihm die gewünschte Süße, so wird zeitig ein Stück <cue:recipeIngredient>Zucker</cue:recipeIngredient> dazu gethan, so wie beim Anrichten die Brühe mit etwas <cue:recipeIngredient>Kartoffelmehl </cue:recipeIngredient> gebunden gemacht.* Zum einen ist unklar, ob das Binden mit Kartoffelmehl nur erfolgen soll, wenn es nicht süß genug ist. Zum anderen bin ich mir nicht sicher, ob der Zucker als Zutat anzugeben ist. So etwas wie *recipeOptionalIngredient* wäre zutreffender als *recipeIngredient*.
* Bei Rezept D-74 heißt *Kartoffeln mit verschiedenen Saucen*. Ist das überhaupt ein richtiges Rezept? Ist es sinnvoll *verschiedene Saucen* mit dem gesamten Saucen Kapitel zu verlinken?
* *Kochen* habe ich bis jetzt nicht mit *<cue:cookingMethod>* getagt. Allerdings finde ich es mittlerweile inkonsequent z. B. *braten* zu tagen und *kochen* nicht.
* Beispielsweise bei Rezept B-89 steht *Diese wird eben so gemacht<link target="#B-88"/>, nur lasse man statt Hollunderblüthen […]*. Wenn die Zutaten aus B-88 hinzugenommen werden, werden Hollunderblüthen übernommen, was ja explizit nicht gewünscht ist. Dementsprechend sind Metadaten notwendig, um bei *link* zu entscheiden, welche Zutaten zu übernehmen sind.
* Aus *Regeln beim Zubereiten der Klöße* ergibt sich für alle Klöße (L-Rezepte), bei denen nichts Anderes angeben ist, eine Kochzeit von 10-15min. Eigentlich sollte man diese Information dann auch überall entsprechend eintagen, was ich noch nicht getan habe.

Ein paar triviale Sachen die mir aufgefallen sind:

* Ab und an begegne ich noch einem *f*, welches lieber ein *s* wäre. Manchmal allerdings auch *ss*, welche lieber *ff* wären wie bei *Eßlössel*. Bei diesen Fällen versuche ich stets die entsprechenden Buchstaben zu ihrem Glück zu verhelfen.
* Die G-Rezepte haben als Typ *Eier-Milch-undMehl-Speisen* Ich nehme an, dass da ein Leerzeichen fehlt.

Insgesamt habe ich das Gefühl, dass es schwer ist, Davidis’s sehr unstrukturierte Rezepte zu tagen. Daher habe ich den Verdacht, dass sich noch herausstellen wird, dass ich manche Sachen lieber anders getagt hätte.

Es wäre sicherlich nicht optimal, zunächst alle Rezepte zu tagen und beim weiterem Vorgehen festzustellen, dass man manche Sachen lieber anders getagt hätte. Dann müsste man nämlich alle Rezepte korrigieren. Ich finde es schlauer, zunächst das weitere Vorgehen an den bisherigen Rezepten durchzuexerzieren (das spezifizieren der Zutaten durch den Bundeslebensmittelschlüssel, sowie die Webseite). Diese sind immerhin bereits einige hundert. Ergeben sich dabei bezüglich der Tags Aha-Effekte, können diese gleich bei den restlichen Rezepten berücksichtigt werden.

Randnotiz: Beim Schreiben ist mir nebenbei aufgefallen, dass ich oft Wörter wie tagen, getagt, usw. verwende. Leute die mich gut kennen, würden dies sicherlich mit „typisch Torstisch“ kategorisieren. Ich gelobe über die richtige Verwendung/Übersetzung zu *ich tage nachzudenken*.

Das Tagen ist übrigens eine Arbeit, die bei mir nicht die größten Freudegefühle auslöst. Dazu ist sie gar nicht so stupide, wie man vielleicht denken mag. Wie bereits erwähnt ist die Struktur von der lieben Frau Davidis nicht gut, was definitiv eine Untertreibung ist. Das Tagen ist auch deutlich anstrengender als gedacht und nimmt mehr Zeit in Anspruch. Ein Lichtblick ist [dieser Artikel über Extracting Structured Data From Recipes Using Conditional Random Fields](http://open.blogs.nytimes.com/2015/04/09/extracting-structured-data-from-recipes-using-conditional-random-fields/?_r=2). [Hier](https://deref-gmx.net/mail/client/vrjYRY7TUOc/dereferrer/?redirectUrl=https%3A%2F%2Fgithub.com%2FNYTimes%2Fingredient-phrase-tagger) ist deren GitHub Repository. Daraus ist die Idee geboren eine Texterkennung für unser Kochbuch zu schreiben. Gelingt dies, können viele ähnliche Kochbücher mit deutlich weniger Aufwand durch Metadaten angereichert werden. Auch wenn ich noch keine Ahnung davon habe, freue ich mich darauf, mich in diesen Themenbereich einzuarbeiten und viele neue, spannende Sachen und Algorithmen zu lernen.

Wie im Artikel erwähnt, hat der Algorithmus auch Fehler in den von Menschen getagten Rezepten gefunden. Wenn die Texterkennung gut wird, wird diese sicherlich auch den ein oder anderen Fehler bei mir finden (wer will schon ernsthaft behaupten, dass er fehlerfrei ist?). Wenn die Texterkennung unterschiedliche Zutaten wie ich tagt, dann kann dies eventuell sogar einen Hinweis liefern, wo das Tag-Schema noch nicht voll ausgereift ist.

**Also nächste To-dos ergeben sich für mich damit:**

* Besprechen wie die Zutaten spezifiziert werden (Anbindung an den Bundeslebensmittel­schlüssel)
* Überlegen wie die Webseite aufgebaut wird und wie welche Daten dargeboten werden
  + Datenbank für finale Rezepte?
  + Nur client-seitige Technologien oder auch Server?
* **Literatursuche**/einarbeiten in Masterarbeit/automatische Texterkennung (von Kochbüchern)

(**Literatursuche** -aufgrund von freundlichen Winks mit dem Eisernen Zaunpfahlfett geschrieben ☺ )