

Kommentare

Praktischer Nutzen der Arbeit	Hier werden ergänzend zur bisherigen Ausarbeitung Gedankengänge präsentiert, die wir während dieser Arbeit hatten.
Schwierigkeiten für die kulinarische Analyse mit cueML und Frau Davidis' Kochbuch	<h2>Praktischer Nutzen der Arbeit</h2> <p>Ich selber koche gerne. Beim Beschäftigen mit Auszeichnungssprachen wie auch Information Extraction habe ich darüber hinaus viel gelernt. Beides finde ich klasse, ist jedoch keine echte Rechtfertigung für ein Master-Arbeits-Thema. Folgend werden daher ein paar Beispiele vorgestellt, die durch (mit cueML) ausgezeichnete Rezepte ermöglicht werden.</p> <p>Wie bereits hier erwähnt, speichern die kommerziellen Seiten Chefkoch.de und Cooking.nytimes.com intern Meta-Daten zu ihren Rezepten. Dies zeugt von dem Bedarf für Auszeichnungssprachen in der Koch-Domäne. Nahe liegende, dadurch ermöglichte Services, sind Empfehlungssysteme für Rezepte, welche auf den Zutaten eines Rezeptes aufbauen. Das in (Ueda et al., 2011) vorgestellte System empfiehlt Rezepte, aufbauend auf den Zutaten der Rezepte, welche ein Nutzer in seiner Vergangenheit besucht hat. (Freyne, Berkovsky, 2010) und (Geleijnse et al., 2010) stellen Empfehlungssysteme vor, welche gesunde Rezepte bevorzugen. Wie gesund ein Rezept ist, wird aus den aufsummierten Nährwertangaben der verwendeten Zutaten bestimmt. (Teng et al., 2012) bietet darüber hinaus die Möglichkeiten nach Alternativen für eine Zutat eines Rezeptes zu suchen. So können beispielsweise einzelne Zutaten durch fettärmere ersetzt werden.</p>
Wissenschaftliche Arbeiten im 21. Jahrhundert	Aus Rezepten lassen sich auch gesellschaftswissenschaftliche Informationen ableiten. (Henning, 2008) beschäftigt sich zum Beispiel mit charakteristischen Merkmalen von Kochbüchern im Verlauf der deutschen Geschichte. So gibt es in Kochbüchern vor dem 18. Jahrhundert oft Hinweise, wie man den Eigengeruch von verdorbenen Zutaten überdecken kann, was ein Zeichen von Nahrungsmangel und Armut ist. Der prozentuale Anteil von Gerichten explizit für die christliche Fastenzeit zeugt von dem Einfluss der Religion. Die ersten Vorkommen von nicht lokalen Zutaten sind Indikatoren für internationalen Austausch. Ein anderes Beispiel ist (Ahn et al., 2011). Dort wird anhand von über 56.000 Rezepten die westliche mit der asiatischen Küche verglichen.
Wissen ist Macht	
Final Retrospective	

Schwierigkeiten für die kulinarische Analyse mit cueML und Frau Davidis' Kochbuch

Das Ziel dieser Arbeit ist es, eine kulinarische Analyse von Frau Davidis' Kochbuch vorzubereiten. Folgend gehen wir noch auf unbehandelte Aspekte ein. Dies beinhaltet zum einen Ungenauigkeiten im Kochbuch und zum anderen noch offene Punkte in cueML.

Nährwertangaben sind nur sinnvoll, wenn sie auf ein **Vergleichbares Maß** wie *pro Person* umgerechnet werden können. Ein Nudelsalat für 12 Personen wird insgesamt mehr Kalorien haben, als ein fetthaltiges 200g Steak. Bei Frau Davidis' Rezepten fehlen jedoch oft jegliche Angaben zu der angestrebten Portionen-Menge. Dementsprechend müsste diese geraten werden, was die Nährwertanalyse ungenau macht.

Gleiches gilt für die einzelnen Mengenangaben der Zutaten. Zum einen müssen vage Mengenangabe wie *ein Stich Butter* interpretiert werden. Meine Oma versteht z. B. sicherlich mehr als die fünffache Menge Butter darunter als meine Freundin. Dies führt zu weiteren Ungenauigkeiten bei der Nährwertanalyse. Zum anderen müssen historische Mengeneinheiten wie *ein Maß* oder *für 8 Pfennig Weißbrot* in heute übliche Einheiten umgerechnet werden, damit sie vergleichbar werden. Dafür ist historische Recherche nötig. Auch für manche Zutaten wie *ganzes Gewürz* sind weitere Recherchen nötig, um herauszufinden, was damit gemeint ist.

Weitere Schwierigkeiten bereitet die **schlechte Struktur von Frau Davidis' Kochbuch**. Manche Informationen sind in anderen Rezepten versteckt. Im Rezept *Rindfleischsuppe mit Perlgerste und Reis* steht z. B. „Anmerk. Will man Reis oder Sago zur Suppe nehmen, so gibt man dieses später hinein. Man rechnet davon auf jede Person bei allen Fleischsuppen einen gestrichenen Eßlöffel voll“ (Davidis, 1849, S. 29). Zu allen Fleischsuppen gehört also als optionale Zutat pro Person ein Eßlöffel Reis oder Sago, was man nur wissen kann, wenn man dieses Rezept gelesen hat.

Überhaupt das Vorkommen von optionalen Zutaten erschwert die kulinarische Analyse. Je nachdem, ob eine optionale Zutat verwendet wird, hat ein Rezept eine andere Geschmacksrichtung und natürlich auch andere Nährwertangaben. Den Mittelwert der Nährwertangaben vom Verwenden und nicht Verwenden einer optionalen Zutat zu bilden, macht kulinarisch keinen Sinn. Gleiches gilt für alternative Zutaten. Das Rezept für *Barsch auf deutsche Art* (Davidis, 1849, S. 62) gibt entweder als Soße geschmolzene Butter mit Senf oder eine Eier-Sauce hinzu. Je nachdem welche Sauce verwendet wird, hat das Rezept ein anderes Flair. Streng genommen ist jede alternative Zutat wie auch jede optionale Zutat ein eigenes Rezept. Ist in einem Rezept eine Zutat optional und sind drei weitere alternativ zueinander, enthält das Rezept eigentlich sechs Rezepte. Zwei je nachdem, ob die optionale Zutat verwendet wird und jeweils drei dazu, je nachdem welche der alternativen Zutaten verwendet wird. Eine kulinarische Analyse muss daher aus den Rezepten die *echte* Anzahl der Rezepte extrahieren.

Die Anbindung von cueML an den **Bundeslebensmittelschlüssel** (BLS) erfüllt die Anforderung der frei verfügbaren Ressource nicht. Ich sehe ihn allgemein aus folgenden Gründen kritisch:

- Er ist nicht öffentlich zugänglich (wir dürfen ihn unter einer Forschungslizenz verwenden, jedoch in keiner Weise veröffentlichen). Damit sind durch ihn berechnete Nährwertangaben nicht nachvollziehbar. Somit geht die Transparenz verloren, die ein wichtiger Bestandteil des semantischen Webs ist (Berners-Lee et al., 2001).
- Wie die Nährwertangaben einzelner Zutaten berechnet werden, ist bereits im BLS nicht transparent. Nach dem BLS haben z. B. *Nichtalkoholische Getränke* einen Energiegehalt von 47kcal/100g, 198kJ/100g, ... Da frage ich mich, wieso manche Leute aus Diät-Gründen auf Cola verzichten, wenn Wasser wie Cola als nicht alkoholisches Getränk doch die gleichen Nährwertangaben hat. Als anderes Beispiel steht in einer Liste über derzeit bekannte Diskrepanzen in BLS - Version 3.02: „Der Fettgehalt von U661100 „Schwein Schulter (Bug) (ma) roh“ hat sich von Version II.3

zu 3.02 fast verdoppelt. Ursache: Bereits im Zuge des Updates für Version 3.0 wurde gegenüber Version 2.3 ein aktuellerer Analysenwert übernommen. WICHTIG: Dabei handelt es sich nicht um einen Fehler.“ Die Korrektur einer Angabe um fast 100% ist nicht sehr vertrauenerweckend, wenn als Begründung lediglich pauschal ein aktuellerer Analysewert angegeben wird.

- Ein Mapping Zwischen einer Zutat und dem BLS kann sich als überraschend schwierig erweisen. Zu *Hauskaninchen* hat der BLS z. B. 17 Einträge (Hauskaninchen gegart, Hauskaninchen roh, Hauskaninchen gekocht, ...). Dem liegt das allgemeine Problem zugrunde, dass es nicht die perfekte Muster-Zutat gibt. Streng genommen hat auch jedes individuelles Hauskaninchen (roh, gekocht, ...) individuelle Nährwertangaben. Eine kulinarische Analyse basierend auf Nährwertangaben kann im Allgemeinen nicht exakt sein, da selbst zwei benachbart gewachsene Möhren aus dem selben Feld unterschiedliche Nährwertangaben haben. Eine Verallgemeinerung erhöht die Übersichtlichkeit und somit auch die Nutzbarkeit und Vergleichbarkeit.
- Die Struktur des BLS-Schlüssels lässt zwar eine Taxonomy erahnen, allerdings ist diese nicht konsequent eingehalten und auch nicht dokumentiert.

Da wir cueML entwickelt haben, sind wir noch die einzigen, die dieses Vokabular verwenden. Idealerweise würden nach einer Veröffentlichung **sämtliche Kochseiten cueML adoptieren**. Das würde die in Praktischer Nutzen der Arbeit erwähnten Anwendungen ermöglichen.

Diese Arbeit hat sich für den Prototypen zur automatischen Auszeichnung erst einmal nur mit den ersten 50 Rezepten beschäftigt. Das Wörterbuch aller Zutaten konnte aus den manuell ausgezeichneten Rezepten extrahiert werden. Um neue Rezepte auszuzeichnen wird ein **allgemeines Wörterbuch von Zutaten** benötigt. Wie zuvor erwähnt, ist der BLS dafür nicht ideal. Als Alternative haben wir auch GermaNet in Betracht gezogen. In diesem Thesaurus gibt es *Nahrung* als Unterkategorie von Nomen. Allerdings ist GermaNet auch nicht frei verfügbar. Des Weiteren sind zu den einzelnen Begriffen keine Nährwertangaben angegeben. Einträge wie *Fressen*, *Mahlerzeugnis*, *Gang*, *Biokleidung*, *Dübel*, *Haschzigarette*, usw. lassen außerdem auch Zweifel an der Qualität dieser Unterkategorie aufkommen. Ich habe den allgemeinen Verdacht, dass frei zugängliche Datenquellen nicht nur nachvollziehbar sind, sondern durch die Transparenz und der Teilung mit vielen Interessierten auch qualitativ besser sind.

Wissenschaftliches Arbeiten im 21. Jahrhundert

Als ich 2014 meine Bachelor-Arbeit in *Word* geschrieben habe, wurde ich von allen gefragt, ob ich noch bei klarem Verstand bin - eine wissenschaftliche Arbeit ist doch selbstverständlich in *LaTeX* zu schreiben. Dem widerspreche ich vehement. „Vor meiner Zeit“ mag *LaTeX* das mit Abstand beste und stabilste Textsatzsystem gewesen sein, so dass größere Arbeiten nur mit diesem sinnvoll zu schreiben waren. Inzwischen wurde *Word* hingegen seit mehr als 34 Jahre weiter entwickelt und Funktionen wie Bibliographien, Inhaltsverzeichnisse, Formeln, usw. funktionieren in *Word* problemlos.

Wie anhand dieser Webseite zu sehen ist, will ich jedoch auch nicht *Word* als das Mittel der Wahl verkaufen. Wir denken, dass Papier im Allgemeinen für Abschluss-Arbeiten nicht mehr das ideale Medium ist. Gebundene exemplare tendieren dazu im Regal einzustauben und nie mehr gefunden zu werden. Sie können auch schlecht weiterentwickelt werden.

Bereits 1989 hat Tim Berners-Lee in seinem Proposal am CERN, welches als Grundstein für das World Wide Web gilt, geschrieben, dass Dokumentation durch ein Buch zu einem System, welches ständig von unterschiedlichen Leuten verwendet und weiterentwickelt wird, unbrauchbar ist (Berners-Lee, 1989). Eine Master-Arbeit verstehen wir als so ein, zeitlich begrenztes Forschungsprojekt, welches idealerweise in weiteren Forschungsarbeiten von anderen Leuten als dem nun fertig absolvierten Studenten weiterentwickelt wird.

Wie diese Arbeit zeigt, kann eine Abschluss-Arbeit auch als Webseite dargeboten werden. Abbildungen profitieren vom dem potenziell größeren Bildschirm als das feste A-4 Format. Auch sind interaktive Grafiken möglich. Des Weiteren können Webseiten zu Hause am PC, auf der Couch am Tablet oder unterwegs am Handy gelesen werden. Links auf einer Webseite zu folgen ist angenehmer, als sie über ein gedrucktes Literaturverzeichnis manuell aufzufinden. Über die URL ist die Arbeit leicht verteilbar und durch das World Wide Web frei zugänglich und auffindbar. Durch eine schicke Webseite kann ein Thema auch (teilweise) dem Laien zugänglich gemacht werden.

Ein weiterer Vorteil bietet die Integration eines Blogs in die Webseite. Seine Gedanken niederzuschreiben ist immer sinnvoll. Dies hilft die Gedanken zu ordnen und früh zu merken, wo sie noch nicht stimmen. Der Blog bietet auch eine gute Plattform zum Austausch von Informationen. Der betreuende Professor ist nach dem Lesen stets im Bilde, was der Student erarbeitet hat. Unter die Blog-Einträge kann er dann ggf. sofort Feedback geben. So ist ein aneinander vorbeireden unwahrscheinlich und die stattgefundene Kommunikation ist später noch nachvollziehbar. Über den Blog können auch andere Wissenschaftler Feedback geben und ihre Erkenntnisse teilen.

Meiner Meinung nach hätten mich also bereits 2014 alle fragen sollen **wieso schreibst du keine Webseite für deine Abschluss-Arbeit?** Unten stehend noch eine schöne Karikatur von einer schicken, aufgeräumten und vernetzten Webseite vs. einer verstaubten, gebundenen Abschluss-Arbeit.

*Gemalt von Fenja Severing nach Idee von Torsten Knauf



Wissen ist Macht

Mit Forschung ist auch immer die Frage der Verantwortung verbunden. Ein einprägsames Beispiel dafür ist die Frage, inwieweit Physiker, die an der Atom-Bombe geforscht haben, Verantwortung für den Einsatz einer Atom-Bombe tragen. Eine sehr empfehlenswerte Lektüre zu diesem Thema ist „Die Physiker“ von Friedrich Dürrenmatt. Unsere Arbeit beschäftigt sich damit, wie Wissen automatisiert extrahiert und ausgezeichnet werden kann. Daher möchte ich folgend erörtern, dass nicht nur technische Forschungen wie an der Atom-Bombe oder die Erfindung des Internets eine Gesellschaft verändern können sondern auch Wissen.

Bereits im 7.Jh. wird dem Schwiegersohn des Propheten Mohammeds das Zitat „*Knowledge is power...*“ zugeordnet. Aber „*With great power comes great responsibility*“. Dies möchte ich folgend an ein paar Beispielen verdeutlichen.

Nahj Al-Balagha, Saying 146 (Imam Ali, ca. 599-661 n.Chr.)

„Knowledge is power and it can command obedience. A man of knowledge during his lifetime can make people obey and follow him and he is praised and venerated after his death. Remember that knowledge is a ruler and wealth is its subject“

Bekanntes Sprichwort mit unbekannten Ursprung (s. hier für genauere Informationen)

„With Great Power Comes Great Responsibility“

Das für mich persönlich krasseste Beispiel, wie Wissen instrumentalisiert und missbraucht werden kann, ist die Nazi-Zeit unter Adolf Hitler. Ein wesentlicher Bestandteil der Unrechts-Diktatur war sicherlich falsche Anschuldigungen gegenüber Juden, um ein gemeinsames Feindbild zu haben. Ein anderer die Hitlerjugend welche früh der Jugend eine Ideologie und Weltanschauung eingetrichtert hat. Beides sind Formen von Wissen.

Ein sehr aktuelles Beispiel ist die Diskussion, in wie weit *Fake News* öffentliche Meinungen und Wahlen beeinflussen (s. z. B. hier). Ich persönlich denke, dass die Methodik der bewusst falschen Informationen zum eigenen Vorteil kein neues Phänomen ist. Ich behaupte, dass es bereits seit dem Bestehen von Zeitungen Zeitungsenten gibt, um die Verkaufszahlen der Zeitung in die Höhe zu treiben. In den beiden Weltkriegen hat man kritische wie auch ausländische Presse oft als *Lügenpresse* angeprangert (ein Begriff der traurigerweise aktuell wieder im Umlauf ist). Die Amerikaner sind 2003 in den Irak einmarschiert mit der Begründung, dass der damalige Diktator Saddam Hussein Massenvernichtungswaffen besitzt, was sich im Nachhinein als eine Lüge herausgestellt hat. Erst kürzlich wurden die im Baltikum stationierten Bundeswehrtruppen fälschlicherweise einer Vergewaltigung beschuldigt, um die Toleranz ihnen gegenüber zu vermindern (s. z. B. hier). All dies verdeutlicht die Macht von Informationen / Nachrichten / (falschem) Wissen. Was meiner Meinung nach neu ist, ist dass durch das Internet quasi jeder, anonym und ohne Zeitverzögerung ein großes Publikum erreichen (und auch belügen) kann.

Ein weiteres Beispiel ist die bloße Existenz von Geheimdiensten, zu deren Hauptaufgabe es gehört, Informationen zu beschaffen. So war z. B. in den Medien, dass der russische Geheimdienst Material gesammelt hat, um den aktuellen amerikanischen Präsidenten Donald Trump zu erpressen (s. z. B. hier). „*Knowledge is power and it can command obedience*“ lässt grüßen. Seit 2002 bis vielleicht sogar heute, wird auch von den amerikanischen Geheimdiensten das Handy des deutschen Bundeskanzlers abgehört (s.

z. B. hier). Nur ein Schelm (wie ich) würde behaupten, dass das etwas mit *mangelnden command obedience* von dem damaligen Bundeskanzler Gerhard Schröder, den Amerikanern in den Irak-Krieg zu folgen, zu tun haben könnte.

In der Koch-Domäne die Macht des Wissens zu diskutieren, mag vielleicht übertrieben sein. Ich bin jedoch der Meinung, dass es nie falsch ist, das Thema zu diskutieren. Daher möchte ich diese kleine Abschweifung mit einem kurzen Gedankenspiel beenden. Dieses soll zeigen, wie selbst Wissen, welches durch unser cueML-Vokabular zugreifbar geworden ist, Macht ausüben könnte. Stellen Sie sich vor, es sind Wahlen und Sie wollen natürlich zukünftig einen gesunden BundeskanzlerIn haben. Nun lesen Sie in der Zeitung, dass jemand den Internet-Verkehr ihres Kandidaten mitgeschnitten hat. Seit zwei Monaten öffnet er verstärkt Rezepte, die diese X Nährstoffe meiden. Wie Ihnen jeder Arzt bestätigen kann, macht es Sinn diese Nährstoffe zu meiden, wenn man Bauchspeicheldrüsenkrebs hat. Der Artikel kommt also zu dem Ergebnis, dass der Kandidat schwer krank ist. Der Kandidat selber weigert sich jedoch aus Prinzip, seine Kranken-Akte wegen einer so lächerlichen Anschuldigung öffentlich zu machen, da damit viele private Informationen verbunden sind. Denken Sie, dass dieses Szenario ihre Wahlentscheidung oder die von anderen beeinflussen würde? Der Gesundheitszustand von Hillary Clinton war auf jeden Fall ein Thema im amerikanischen Wahlkampf 2016, aufgebauscht durch einen kleinen Schwächeanfall von ihr und ihrer Weigerung ihre Kranken-Akte zu veröffentlichen (s. z. B. hier).

Final Retrospective

Abschließend möchte ich Revue passieren lassen, was ich in dieser Arbeit gelernt habe. Ich denke bei jedem Software Projekt sollte man etwas lernen. Dies gilt insbesondere für eine Abschluss-Arbeit. Dieses Revue orientiert sich dabei an folgenden vier Fragen aus (Kerth, 2001): 1. *What did we do well, that if we don't discuss we might forget?* 2. *What did we learn?* 3. *What should we do differently next time?* 4. *What still puzzles us?*

1. Zuerst *quick and dirty* Prototypen als Machbarkeitsanalyse zu erstellen war eine gute Idee. So wurden früh Probleme erkannt und der CRF-based Ansatz nicht weiter verfolgt. Den dictionary- and rule-based Prototypen vor der Weiterentwicklung neu zu designen war ebenfalls sehr wichtig. Die anschließende Integration der rule-based Weiterverarbeitung, sowie das zukünftige Hinzufügen von weiteren Regeln, ist auf Grund des guten Designs trivial. Die Unit-Tests zeitgleich mit dem Source Code zu schreiben hat auch gut funktioniert. Dies war noch eine wichtige Lektion aus meiner Bachelor-Arbeit.
2. Natürlich habe ich viele Kleinigkeiten davon gelernt, interessante Artikel zu lesen und ein größeres Projekt vom Anfang bis zum Ende durchzuführen. Mein einschlagendes *Aha-Erlebnis* hatte ich jedoch sicher bezüglich Domänen-spezifischen Vokabulare. Ich habe stets gedacht, dass sich jemand ein Vokabular trivial aus den Fingern saugt und es dann von vielen Leuten verwendet wird. Aber Vokabulare aufzubauen ist zeitaufwendig, benötigt Fachwissen und ist daher auch teuer. Auch ist es nicht einfach ein passendes Vokabular zu finden. Ein universelles Vokabular für alle Zwecke ist illusorisch. Diese (für mich neue Erkenntnisse) werden u.A. in (Zeeshan, 2009) und (Hage et al., 2008) bestätigt.
3. Folgende zwei Sachen würde ich verbessern, wenn ich die Möglichkeit hätte, noch einmal von vorne anzufangen:
 - Im Rahmen meiner SHK-Tätigkeit habe ich viele Rezepte aus Davidis' Kochbuch manuell ausgezeichnet. Dies war gut, da ich dadurch ein Gefühl

dafür bekommen habe, wie schwierig das Auszeichnen der Rezepte ist. Allerdings habe ich sehr viele Rezepte ausgezeichnet, bevor ich *Stopp* gesagt habe - *Wir müssen unser cueML Vokabular ändern*. Dann habe ich wieder sehr viele Rezepte mit dem geänderten Vokabular ausgezeichnet, bevor ich erneut *Stopp* gesagt habe - *Das stimmt noch immer nicht*. Stattdessen hätte ich mich an das Prinzip aus dem Agilen Manifest *Working software is the primary measure of progress* halten sollen. Das cueML Vokabular alleine ist keine funktionierende Anwendung. Der Sinn des Vokabulares ist es, Informationen so auszuzeichnen, dass sie von einer Anwendung verwendet werden können. Dementsprechend hätte ich nach einer Änderung im Vokabular den Workflow bis zur Webseite einmal exerzieren sollen; ein paar Rezepte auszeichnen, diese Rezepte in eine Webseite umwandeln und mir dann als Nutzer anschauen, ob ich die Rezepte passend dargestellt finde. Daraufhin kann das Vokabular erneut angepasst werden, usw. So hätte verhindert werden können, dass ein Großteil meines Auszeichnens als SHK letztendlich nicht verwendet wird, da es mit einer alten Version des cueML-Vokabulares ausgezeichnet ist.

Intuitiv stellt sich die Frage, ob dies auch durch eine *Requirement Analyse* hätte verhindert werden können. Alle Anforderungen wie optionale Zutaten, oder Verweise zwischen (Teil)-Rezepten *upfront* zu erkennen, halte ich jedoch für illusorisch. Dementsprechend ist der oben beschriebene *iterative* Ansatz besser.

- Des Weiteren würde ich, wie in Wissenschaftliches Arbeiten im 21. Jahrhundert erläutert, von Anfang an einen echten Blog führen.
- 4. Beim nächsten größeren Projekt werde ich versuchen meine Anfangserwartungen herunterzuschrauben. Am Anfang habe ich gedacht, dass ich selbstverständlich alle Informationen aus Davidis' Kochbuch extrahieren kann. Nach und nach musste ich jedoch erkennen, dass das gelinde gesagt übermotiviert war. Unser finaler vorgestellte Prototyp ist davon noch weit entfernt, was ihn in keiner Weise schlecht reden soll.