

Prüfungsprotokoll CL-Diplom

Konrad Anton

22. September 2004

1 Umfang

Die Prüfung bei Prof. Udo Hahn dauerte 30 min. und erstreckte sich über die Themen

- Computerlinguistik I, II, die Grundvorlesungen
- Grundlagen der Automatischen Sprachverarbeitung, eine Mischung aus CLI und Natürlichsprachliche Systeme.
- Seminar Medieninformatik, Thema Semantic Web

2 Fragen

Im Gegensatz zu den beiden Prüfungen vor mir ging Hahn zunächst auf allgemeine CL-Themen ein und stürzte sich nicht sofort aufs Vertiefungsgebiet. Die hier wiedergegebenen Fragen sind natürlich nur aus dem Gedächtnis rekonstruiert und daher notwendigerweise ungenau.

2.1 Morphologie

Was ist denn das dominierende Verfahren in der Morphologie? Zwei-Ebenen-Morphologie.

Und worauf baut es auf? Transduktoren.

Wie definiert man einen endlichen Automaten? Tupel-hinschreib. So ist er deterministisch und so nichtdeterministisch.

Wie macht man nun daraus einen Transduktor? Zwei Zeichen statt einem pro Übergang, etwa "lies e und schreib i".

Und wie kann man sich die Arbeitsweise vorstellen? Zwei Bänder, lies eines und schreib das andere.

Wie sieht das dann in der Anwendung aus? Ein Band mit Morphemfolge, ein Band mit Oberflächenform. Dazwischen ein Transduktor, der übersetzt, etwa mit y-ie-Übergang im Englischen. Der y-ie-Transduktor lässt sonst alles durch.

Zeichnen Sie mal einen Transduktor für die y-ie-Umwandlung hin. Bastel bastel.

Jetzt verarbeiten Sie mal die Morpheme “try+s” und das Wort “python” . Es kommt “tries” und “python” raus.

2.2 Syntaxanalyse

Kann man diese endliche Automaten auch für Syntaxanalyse einsetzen? Nicht wenn man die Phrasenstruktur vollständig extrahieren will.

Was nimmt man dann? Kontextfreie Grammatiken mit den dazu passenden Algorithmen wie Earley.

Schreiben Sie mal eine kontextfreie Grammatik für “der große, schöne,... Tisch”. (Ich schreibe eine mit einer Epsilon-Produktion, um ADJP nötigenfalls leer ableiten zu können).

Und wenn Sie keine Epsilon-Produktionen zur Verfügung haben? (Ich schreibe sie um.)

Was wäre das Schöne an endlichen Automaten? Die Laufzeit.

Nämlich? Deterministische linear, weil man jedes Zeichen nur einmal anschaut. Nicht-deterministische kann man mit Backtracking parsen, dann länger, oder mit irgendeiner Darstellung der Zustandsmenge, dann linear. Man kann auch nichtdet in det umwandeln.

Was ist aber dabei das Problem? Schlimmstenfalls hat der DFA exponentiell mehr Zustände als der NFA. Etwa die Sprache der Wörter über $\{0,1\}$, bei denen das n -letzte Zeichen 1 ist. Der NFA kann so aussehen (hinmal), aber es gibt exponentiell viele Äquivalenzklassen unter der Myhill-Nerode-Relation. . . *an dieser Stelle hat er genug Info-3-Tutoren-Monolog gehört.*

Wie ist die Laufzeit von Earley? Zwischen quadratisch und kubisch.

Na gut, reiner Earley ist kubisch, mit Optimierungen wird es etwas besser, ja. Welche Schritte gibt es bei Earley? Den Scanner, den Completer und den Predictor. (was sie tun, was die Items mit dem Punkt drin bedeuten, was für eine Tabelle man füllt).

Parsen Sie mal “der große schöne Tisch” mit Ihrer Grammatik und Earley. Pars pars.

2.3 Medieninformatik: Semantic Web

Was versteht man unter dem Semantic Web bzw. was ist der Unterschied zum heutigen Web? Computerverarbeitbare Annotationen auf Webseiten.

Was ist RDF? Eine Sprache, in der man mittels Tripeln aus Subjekt, Prädikat und Objekt Aussagen notiert. Die Subjekte, Prädikate und Objekte sind jeweils URIs.

Wie würden Sie mit RDF ausdrücken, dass das Bild Soundso den Herrn Hahn zeigt? Ein Tripel mit Subjekt Bild-URL, Objekt Hahn-URI, Prädikat wähle ich “stelltDar” in meiner Ontologie.

Wie verhalten sich RDF und XML und HTML zueinander? RDF ist eine XML-Anwendung, HTML und XML stammen beide von SGML ab, wobei man mit XML beliebige Datenstrukturen aufschreiben kann und HTML auf Webseiten beschränkt ist. *(auf irgendetwas will er hinaus)*

Könnte man in HTML ausdrücken, dass mein Name Udo Hahn ist? Nein.

Und in XML? Man kann ein Name-Element innerhalb eines Person-Elements einsetzen.

Woher weiß man dann, dass das bedeutet, dass das der Name ist? Weiß man nicht. Das Wissen steckt in den Programmen, die diese Elemente interpretieren, und in dem Kopf des Autors.

Könnte man daher einfach XML zur Annotation einsetzen? *oder so ähnlich. Was meint er? Nach etwas Herumgedruckse kommen wir drauf, dass er hören will, dass man diese Begriffe in einer Ontologie definieren muss.*

Was ist eine Ontologie? Zitat der Gruberschen Definition.

Womit würden Sie eine Ontologie schreiben? OWL, weil W3C-Standard und verbreitet.

Welche computerlinguistischen Schwierigkeiten sehen Sie bei einem Semantic-Web-Ansatz, der auf Annotationen basiert? *oder so... Ich erzähle etwas davon, dass man die Annotationen ja gerne aus natürlicher Sprache extrahieren würde, dass das aber KI-vollständig sei. Das ist aber nicht, was er hören will. Nach der Prüfung erfahre ich, dass er darauf hinaus wollte, dass man mit solchen Annotationen nicht beliebige Fragen sondern nur innerhalb feststehender Kategorien beantworten kann.*

2.4 Ergebnis

Es gab 1,3. Ich bin's zufrieden.

3 Literaturliste

Andere am selben Prüfungstag hatten längere Prüfungslisten zu mehr Vertiefungsgebieten.

Allen, James; Natural Language Understanding; 1995; 2nd Ed;
Redwood City/CA: Benjamin/Cummings

Manning, Christopher D + Schuetze, Hinrich
Foundations of Statistical Natural Language Processing
Cambridge/MA, London/England: MIT Press, 1999, 680pp.

Jurafsky, Daniel; Martin, James A. [2000]
Speech and Language Processing:
An Introduction to Natural Language Processing, Computational
Linguistics and Speech Recognition
Upper Saddle River/NJ: Prentice Hall, 2000, 934pp

S.Staab + R.Studer (eds) 2004
Handbook on Ontologies.
Springer 2004

Fensel, Dieter; Hendler, James; Lieberman, Henry; Wahlster, Wolfgang; (Eds)
Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential.
Cambridge, MA. MIT Press 2002, 392 pp

Fensel, Dieter:
Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce.
Berlin etc Springer 2001, 138pp

Davies, John; Fensel, Dieter; Harmelen, Frank van (Eds)
Towards the Semantic Web: Ontology-Driven Knowledge Management
Wiley, 2003