

INF2820 Datalingvistik 2014

Detaljert pensumoversikt

Pensum består av ulike elementer

1. Presentasjoner fra forelesningene som er lagt ut på web
2. Alle oppgavesett, både obligatoriske innleveringsoppgaver og ukeoppgaver
3. Noen små Python-program
4. Utvalg fra S. Bird, E. Klein and E. Loper: Natural Language Processing with Python (kalt NLTK under)
5. Utvalg fra Jurafsky, D., & Martin, J. M. : Speech and Language Processing, 2008 (kalt J&M under)
6. En artikkel fra Wikipedia

Noen seksjoner/paragrafer fra bøker vil være ND (=not the details). Det betyr at du er forventet å ha lest seksjonene og være kjent med innholdet, men at du ikke vil bli spurt om tekniske detaljer på eksamen.

Oversikten er organisert under temaer.

0. Introduksjon

1. Presentasjon
 - a. 16 Jan: Forel.1.1
5. J&M, Kap. 1

1. Endelige tilstandsteknikker

1. Presentasjoner:
 - a. 16. Jan: Forel.1.2,
 - b. 22 .Jan: Forel.2,
 - c. 30. Jan: Forel.3
 - d. 6. Feb: Forel.4
2. Oppgaver
 - a. 21. Jan: Oppgavesett1 + Oppgavesett1B
 - b. 28. Jan: Oppgavesett2
 - c. 4. Feb: Oppgavesett3
 - d. 14 Feb: Obligatorisk oppgave 1
3. Python-program:
 - a. dfa_recog.py
 - b. dfa_recrec.py
 - c. nfa_naive_recog.py
 - d. nfa_smart_recog.py
4. NLTK: ingenting
5. J&M: Seksj. 2.2-2.3
 - a. Underseksjonene 2.2.5 og 2.2.6 er ND. For detaljer, bruk i stedet forelesnings-presentasjonene, Python-programmene og oppgaver inkludert JFLAP-oppgavene.

2. Kontekstfrie grammatikker

1. Presentasjoner
 - a. 6. Feb: Forel.4
 - b. 13. Feb: Forel.5
2. Oppgaver
 - a. 11. Feb: Oppgavesett4
 - b. 18. Feb: Oppgavesett5
 - c. 25. Feb: Oppgavesett6
 - d. 4. Mar: Oppgavesett7
 - e. 10. Mar: Obligatorisk oppgave 2, oppgave 1, 2
 - f. 9. Apr: Obligatorisk oppgave 3, oppgave 2,3
3. Python-program: ingen
4. NLTK: Seksj. 8.1-8.3
5. J&M: Seksj. 12.0-12.6, men
 - a. Seksj. 12.3 er ND i den forstand at du ikke trenger å huske de enkelte grammatikkregler. Forskjellige grammatikker vil bruke tildels forskjellige regler.
 - b. Underseksj. 12.4.3 og 12.4.4 er ND
6. Wikipedia: [X-bar theory](#)

3. Parsing, Recursive-Descent og Shift-Reduce parsing

1. Presentasjoner:
 - a. 13. Feb: Forel.5
 - b. 20. Feb: Forel.6
2. Oppgaver
 - a. 25. Feb: Oppgavesett6
 - b. 4. Mar: Oppgavesett7
 - c. 11. Mar: Oppgavesett8
 - d. 10. Mar: Obligatorisk oppgave 2, oppgave 3-6
 - e. 9. Apr: Obligatorisk oppgave 3, oppgave 1
3. Python-program:
 - a. rd_recognizer.py
 - b. rd_parser.py
 - c. sr_recognizer.py
 - d. sr_parser.py
4. NLTK: Seksj. 8.4
5. J&M: Seksj. 13.0-13.3

4. Tabellparsing: CKY- og chart-parsing

1. Presentasjoner:
 - a. 27. Feb: Forel.7
 - b. 6. Mar: Forel.8
 - c. 13. Mar: Forel.9
2. Oppgaver
 - a. 18. Mar: Oppgavesett9
 - b. 25. Mar : Oppgavesett10
 - c. 1. Apr: Oppgavesett11
 - d. 9. Apr: Obligatorisk oppgave 3, oppgave 4
3. Python-program:
 - a. cky.py
 - b. chart.py
4. NLTK:
 - a. Seksj. 8.4
 - b. Web-extras: Sec 8.1 Chart-parsing to A*-parser
5. J&M: Seksj. 13.4:
 - a. Underseksjon 13.4.2 er ND. Skift ut detaljene med forelesningspresentasjonene og Python-programmet chart.py.

5. Trekk og unifikasjon

1. Presentasjoner:
 - a. 20. Mar: Forel.10
 - b. 27. Mar: Forel.11
 - c. 3. Apr: Forel.12
2. Oppgaver
 - a. 8. Apr: Oppgavesett12
 - b. 29. Apr: Oppgavesett 14
 - c. 7. Mai: Obligatorisk oppgave 4, oppgave 1-3
3. Python-program:
 - a. grammar_2_1.fcfg
4. NLTK: Kap. 9, alt unntatt
 - a. Avsnittet om "Unbounded Dependency Constructions" I seksjon 9.3.
5. J&M: Seksj. 15.1-15.3
 - a. Underseksjon 15.3.3 er ND. "Subcat" skal ikke være en "HEAD"-feature.

6. Semantikk

1. Presentasjoner:
 - a. 10. Apr: Forel.13
 - b. 24. Apr: Forel.14
2. Oppgaver
 - a. 22. Apr: Oppgavesett13
 - b. 29. Apr: Oppgavesett 14
 - c. 7. Mai: Obligatorisk oppgave 4, oppgave 4
 - d. 13. Mai: Oppgavesett15
3. Python-program:
 - a. simple-sem.fcfg
4. NLTK: Seksj. 10.0-10.4, but
 - a. Unntatt delen "Model building" i seksj. 10.3.
 - b. Unntatt delen "Quantifier Ambiguity revisited" i seksjon 10.4.
5. J&M: Seksj. 17.0-17.3

7. Formell språkteori

1. Presentasjoner:
 - a. 6. Mai: Forel.15
5. J&M: Kap. 16