

Tekniker för storskalig parsning

Inlämningsuppgifter

1 Introduktion och grundbegrepp (F1–F3)

- Obligatoriska uppgifter:
 - O1 Förklara vad problem med *robusthet*, *disambiguering* och *effektivitet* innebär i samband med storskalig parsning.
 - O2 Ge exempel på vad en *morfologisk* respektive *syntaktisk* analys av följande text kan innebära:

Första matchen gav frågetecken. Mötet mot Taylor Dent svar på tal. Seger i tre raka set, 6–2, 6–2, 6–4, gör att Robin Söderling är vidare till tredje omgången i US open och det på ett övertygande sätt.
 - O3 Ge exempel på *modeller* och *algoritmer* som kan användas för att utföra den morfologiska respektive syntaktiska analysen.
 - O4 Förklara på vilket sätt statistik och statistiska modeller kan användas för att minska problemen med *robusthet*, *disambiguering* och *effektivitet* i storskalig parsning.
- Extrauppgifter:
 - E1 Ge exempel på hur en *utvärdering* av den morfologiska respektive syntaktiska analysen i O2 skulle kunna gå till.
 - E2 Vad är en approximeringsalgoritm? Förklara skillnaden mellan *modellfel* och *sökfel* vid användningen av approximeringsalgoritmer.

2 Dependensparsning (F4–F5)

- Obligatoriska uppgifter:
 - O5 Förklara hur en transitionsbaserad dependensparser är uppbyggd och hur inlärning och inferens (parsning) går till.

- O6 Vilka egenskaper har en transitionsbaserad parser med avseende på *robusthet, disambiguering* och *effektivitet*?
- O7 Förklara hur en grafbaserad dependensparser är uppbyggd och hur inlärning och inferens (parsning) går till.
- O8 Vilka egenskaper har en grafbaserad parser med avseende på *robusthet, disambiguering* och *effektivitet*?

- Extrauppgifter:

- E3 Redogör för begreppet projektivitet och förklara vad som menas med pseudo-projektiv parsning.
- E4 Beskriv översiktligt någon grammatikbaserad metod för dependensparsning.

3 Stokastiska grammatiker och modeller (F7, F9, F11)

- Obligatoriska uppgifter:

- O9 Konstruera en mini-trädbank innehållande (minst) tre meningar och härled motsvarande trädbanks-PCFG. Trädbanken ska vara så konstruerad att flera icke-terminaler har mer än en produktion. Konstruera en ny mening (som ingår i strängspråket för din trädbanksgrammatik) och härled den mest sannolika analysen enligt trädbanksgrammatiken.
- O10 Vilka egenskaper har en trädbanks-PCFG med avseende på *robusthet, disambiguering* och *effektivitet*?
- O11 Förklara vad som menas med en *generativ* modell i statistisk parsning. Ge exempel på två olika generativa modeller.
- O12 Transformera din mini-trädbank från uppgift 3 så att alla noder är lexikaliserade och har ”parent annotation”. Extrahera en ny trädbanksgrammatik ur den transformerade trädbanken.
- O13 Förklara vad som menas med en *diskriminativ* modell i statistisk parsning. Ge exempel på två olika diskriminativa modeller.
- O14 Förklara skillnaden mellan *lokala* och *globala* diskriminativa modeller. Vilka fördelar och nackdelar har de två typerna av modell?

- Extrauppgifter:

- E5 Visa i detalj hur den probabilistiska varianten av CKY härleder den mest sannolika analysen för den nya meningen i uppgift 3. Transformera din trädbanksgrammatik om det behövs.
- E6 Vad innebär *data-oriented parsing*? Förklara översiktligt hur metoden fungerar och hur den skiljer sig från PCFG-parsning.
- E7 Många diskriminativa modeller bygger på en linjär modell som har formen $f(x, y) \cdot w$. Förklara vad de olika faktorerna i denna ekvation står för och vad inlärningsproblemet för en sådan modell består i.

4 Finita automater (F13)

- Obligatoriska uppgifter:

O15 Förklara hur en *viktad finit transduktor* (weighted finite state transducer, WFST) är uppbyggd.

O16 Ge minst tre exempel på parsningsproblem som kan lösas med en (viktad) finit transduktor.

- Extrauppgifter:

E8 Diskutera för- och nackdelar med att använda finita transduktorer för syntaktisk parsning.