

T-(538|725)-MALV, Málvinnsla Lokaprófið

Hrafn Loftsson¹ Hannes Högni Vilhjálmsson¹

¹Tölvunarfræðideild, Háskólinn í Reykjavík

Nóvember 2007

1 Efni til prófs

2 Tilhögun prófs

1 Efni til prófs

2 Tilhögun prófs

- Málheildir
- Endanlegar stöðuvélar
- Reglulegar segðir
- Perl
- Tilreiðing
- N-stæður
- Orðhlutafræði
- Mörkun

- Setningagreining
- Merkingarfræði
- Orðræða og tilvísanaráðning
- Uppbygging og nýnæmi upplýsinga
- Ásetningur og orðræðumerki
- Samstilling samræðna
- Líkamstjáning

Málheildir

- Hvað eru málheildir?
- Hvað geyma þær?
- Til hvers notaðar?

Endanlegar stöðuvélar

- Skilgreining
- DFA vs. NFA
- Af hverju skilvirkar (e. efficient)?
- Aðgerðir

Málheildir

- Hvað eru málheildir?
- Hvað geyma þær?
- Til hvers notaðar?

Endanlegar stöðuvélar

- Skilgreining
- DFA vs. NFA
- Af hverju skilvirkar (e. efficient)?
- Aðgerðir

Reglulegar segðir

- Skilgreining
- Aðgerðir og tákn
- Kanna að skrifa reglulegar segðir
- Tengsl við endanlegar stöðuvélar

Perl - Geta skrifað einföld forrit

- Skráarhöld
- Skilyrtar segðir og lykkjur
- Vektorar og tætitöflur
- Reglulegar segðir og Perl

Reglulegar segðir

- Skilgreining
- Aðgerðir og tákn
- Kanna að skrifa reglulegar segðir
- Tengsl við endanlegar stöðuvélar

Perl - Geta skrifað einföld forrit

- Skráarhöld
- Skilyrtar segðir og lykkjur
- Vektorar og tætitöflur
- Reglulegar segðir og Perl

Tilreiðing

- Hver eru vandamálin?
- Lesgreining - Lex/Flex/JFlex
- Geta notað JFlex til að leita eftir mynstrum í texta
- Unix tól: grep, tr, sed

N-stæður

- Einstæður, tvístæður, þrístæður, ...
- Gerð n -stæðulíkana
- Unix tól: `sort`, `uniq`, `head`, `tail`
- Tölfræðilíkon: Sennileikalíkur
- Tenging sennileikalíkinda við líkur á tilteknum n -stæðum
- Sléttun

Orðhlutafræði

- Orðflokkar og beygingarleg einkenni
- Myndön og aðskeyti
- Orðmyndun í íslensku
- Orðhlutafræðileg greining, lemmun
- Stöðuferjöld og tveggja laga orðhlutagreining

Mörkun

- Skilgreining
- Markamengi
- Hver eru vandamálin?
- Grunnmörkun
- Nákvæmni í mörkun

Mörkun

- Geta lýst mismunandi mörkunaraðferðum
- Málfræðilegir reglumarkarar vs. gagnamarkarar
- Tölfræðimarkarar og Markov líkön
- Sameining markara

Setningagreining

- Hvað er setningagreining?
- Hver eru vandamálin?
- Setningaliðir og setningafræðileg hlutverk
- Samhengisfrjáls mállýsing
- Afleiðsla og þáttunartré
- Full þáttun vs. hlutaþáttun
- Þekkja DCG í Prolog

Setningagreining

- Nákvæmni (e. precision) og griphlutfall (e. recall)
- Ofansækin vs. neðansækin þáttun
- Þekkja grundvallarhugmynd að baki:
 - Tölflupáttunar (e. chart parsing)
 - Tölfræðiþáttunar (e. probabilistic parsing)

Merkingarfræði

- Hvað felur merkingargreining í sér?
- Tengsl λ -reiknings við merkingargreiningu.
- Umsagnarrökfræði
- Geta lesið úr setningum sem settar eru fram á rökrænu formi.

Orðræða og tilvísanaráðning

- Út á hvað gengur orðræðugreining (e. discourse analysis)?
- Hver er munurinn á upplýsingamáli (e. transactional) og félagslegu máli (e. interactional)?
- Hver er munurinn á orðræðumarkmiði (e. discourse function/intent) og orðræðuáhaldi (e. discourse device)?
- Hver eru Grice heilræðin og hvernig nýtast þau við málskilning?
- Hvað er orðræðusamhengi (e. discourse context), orðræðulíkan (e. discourse model) og orðræðueining (e. discourse entity)?
- Hvernig má finna vísun (e. reference) og samvísun (e. coreference)?

Uppbygging og nýnæmi upplýsinga

- Pekkið tegundir upplýsingaástands (e. information status).
- Geta lýst uppbyggingu upplýsinga í setningu (e. information structure), þ.e. merkja þemu (e. theme) og remu (e. rheme) útfrá samhengi.
- Kannast við fókus (e. focus).
- Átta sig á tengslum upplýsingauppbyggingar og ítónunar í tali (og þar með gæðum tölvutals).

Ásetningur og orðræðumerki

- Hvað er orðræðuskipan (e. discourse structure) og hvað bendir til þess að um slíka skipan sé að ræða í orðræðu?
- Hvað getur einkennt orðræðuhluta (e. discourse segments)?
- Hvað er orðræðutilgangur (e. discourse purpose) og hlutatilgangur (e. discourse segment purpose)?
- Hvernig virkar athyglisstafla (e. attentional stack) og hvað eru orðræðustöður (e. discourse states)?
- Hvað eru orðræðumerki (e. discourse marker) og hvernig tengjast þau athyglisstafla?
- Hvað eru mælskuvensl (e. rhetoric relations) (gróflega)?

Samstilling samræðna

- Geta búið til stöðuvél einfalds samræðukerfis.
- Þekkja grannpör (e. adjacency pairs) og 3ja setninga inngríps (e. interventions) útgáfuna.
- Kunna skil á hinum þremur hlutum talgjörðar (e. speech act):
 - Locutionary
 - Illocutionary
 - Perlocutionary
- Það þarf *ekki* að leggja talgjörðarflokka á minnið!
- Athuga hvaða eiginleika texta mætti nýta til að marka talgjörð.
- Vita hvað grunnprófun er (e. grounding) og hvernig fólk getur framkvæmt slíkt.

Líkamstjáning

- Lýst dæmi um margháttu samskipti (e. multimodal communication).
- Kunnið skil á aðferðinni við að kanna hvort líkamleg hegðun þjónar samskiptatilgangi.
- Eigið að geta komið með dæmi um einhvern tilgang sem augnhreyfingar, handahreyfingar, svipbrigði og líkamsstillingar þjóna (nóg að muna eitt dæmi fyrir hverja hegðun/hátt).
- Geta lýst því hvar þekking á líkamstjáningu getur nýst í máltækni.

1 Efni til prófs

2 Tilhögun prófs

Spurningar á prófinu verða blanda af (ekki ljóst að öll atriðin verði með):

- Skilgreiningum
- Forritunarverkefnum
- Krossaspurningum
- Efnisspurningum

Vægi einstakra þátta

Vægi einstakra þátta á prófinu verður í samræmi við þann tíma sem notaður var í námskeiðinu á viðkomandi þátt.