

Zagreb, 11. ožujka 2016.

Predmet: **Strojno učenje**

DIPLOMSKI ZADATAK br. 1323

Pristupnik: **Vjeran Crnjak (0036467899)**

Studij: **Računarstvo**

Profil: **Računarska znanost**

Zadatak: **Učenje pretraživanja za rješavanje zadataka obrade prirodnoga jezika**

Opis zadatka:

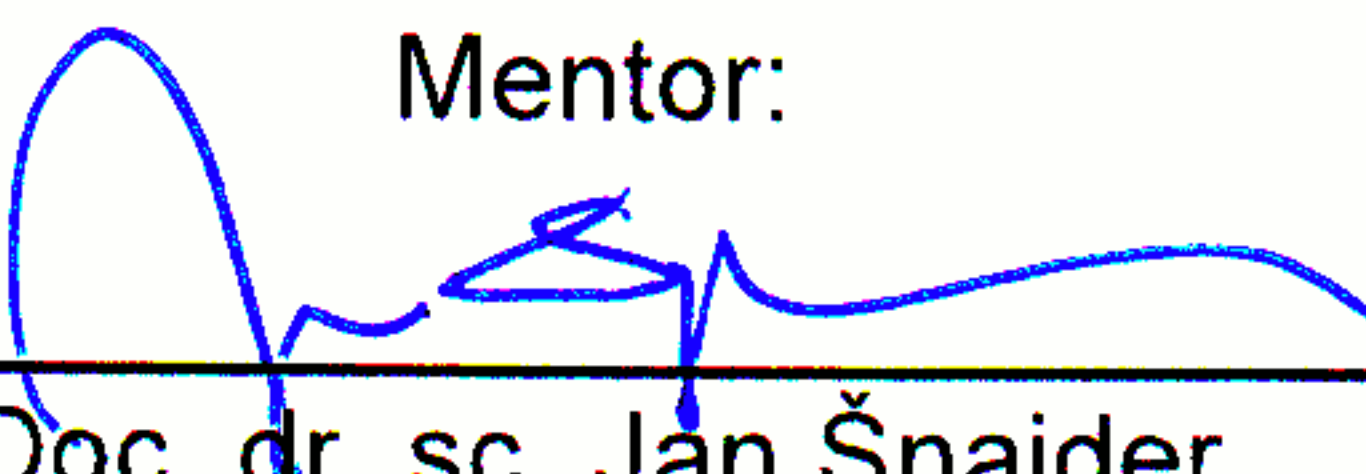
Mnogi problemi u obradi prirodnoga jezika uključuju predviđanje strukture (npr. sintaktička analiza, ekstrakcija relacija i sl.) ili slijednu obradu kroz više razina (npr. označavanje imenovanih entiteta u kombinaciji s razrješavanjem koreferencije, semantičko parsanje, i sl.). Jednostavni pristupi temeljeni na strojnom učenju strukturu izgrađuju na temelju lokalnih klasifikacijskih odluka, a višerazinsku obradu ostvaruju slijedom nezavisnih klasifikatora. Takvi pristupi međutim ne iskorištavaju ovisnosti između zadataka i skloni su propagiranju pogrešaka kroz razine. Alternativu predstavljaju pristupi temeljeni na združenom učenju.

U okviru diplomskoga zadataka potrebno je proučiti pristupe paradigme temeljene na združenome strojnom učenju. Posebnu pažnju posvetiti paradigmi "učenja pretraživanja" (engl. learning to search), kao što su SEARN (Daumé III i dr., 2006) i LOLS (Kai-Wei Chang i dr., 2015), te proučiti radni okvir Vowpal Wabbit. Proučiti tri zadatka obrade prirodnog jezika koji uključuju predviđanje strukture, višerazinsku obradu ili oboje. Razraditi prilagodbu tih zadataka paradigmi učenja pretraživanja. Izgraditi združeni model za svaki od zadataka te provesti iscrpno vrednovanje modela, uključivo analizu pogrešaka te usporedbu s referentnim modelima. Radu priložiti izvorni i izvršni kod razvijenog sustava, skupove podataka i programsku dokumentaciju te citirati korištenu literaturu.

Zadatak uručen pristupniku: 18. ožujka 2016.

Rok za predaju rada: 1. srpnja 2016.

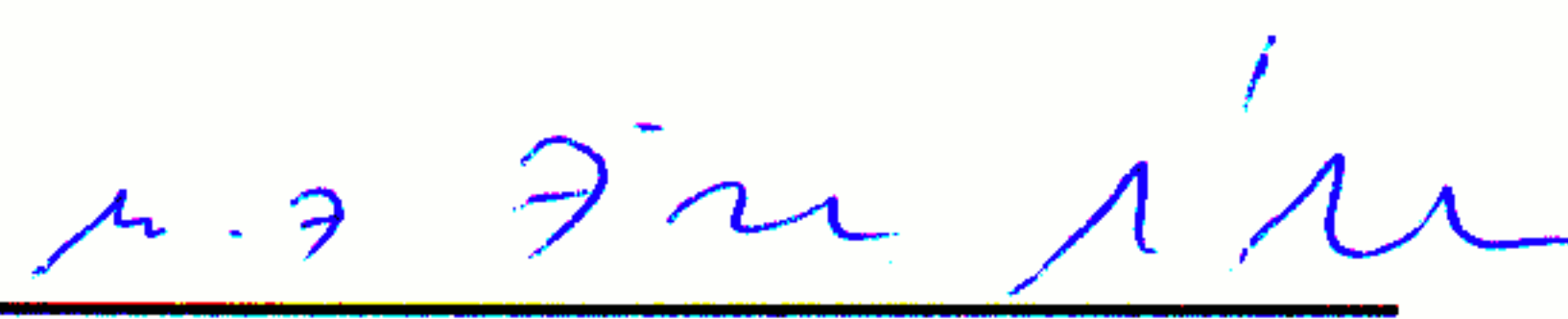
Mentor:


Doc. dr. sc. Jan Šnajder

Djelovoda:


Doc. dr. sc. Tomislav Hrkać

Predsjednik odbora za
diplomski rad profila:


Prof. dr. sc. Siniša Srbljić