

نام خدا

آزمونک شماره پنج – درس پردازش زبان طبیعی – نام و نام خانوادگی:

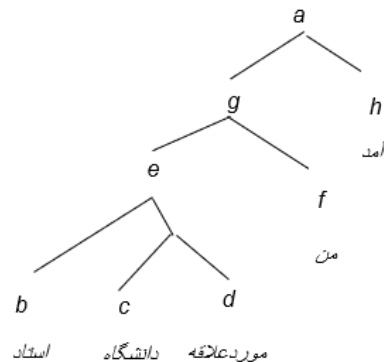
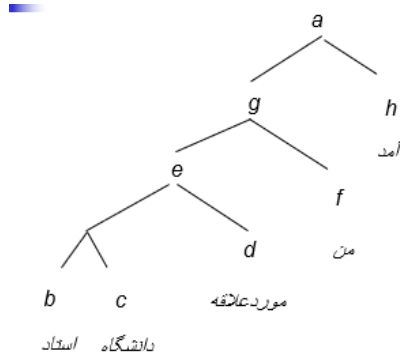
سوال ۱) با استفاده از الگوریتم Transition-based dependency parsing جمله‌ی زیر را تجزیه کنید. مراحل را به ترتیب در یک جدول شامل ۴ ستون stack، word list، action و relation added بنویسید. (۳۰ نمره)
 “استاد دانشگاه موردعلاقه من آمد”

stack	Word list	action	
ریشه	استاد-دانشگاه-موردعلاقه-من-آمد	shift	
استاد-ریشه	دانشگاه-موردعلاقه-من-آمد	shift	
دانشگاه-استاد-ریشه	موردعلاقه-من-آمد	rightarch	دانشگاه → استاد
استاد-ریشه	موردعلاقه-من-آمد	shift	
موردعلاقه-استاد-ریشه	من-آمد	rightarch	موردعلاقه → استاد
استاد-ریشه	من-آمد	Shift	
من-استاد-ریشه	آمد	rightarch	من → استاد
استاد-ریشه	آمد	shift	
آمد-استاد-ریشه		leftarch	آمد ← استاد
آمد-ریشه		rightarch	آمد → ریشه
ریشه		Done	

سوال ۲) درخت یا درخت‌های سازه‌ی ممکن (حداقل دو درخت نحوی سازه‌ی ای) برای جمله‌ی مطرح شده در سوال ۱ را بدون لیبل بکشید (۲۰ نمره)
 دو حالت دارد.

[[استاد دانشگاه موردعلاقه من] آمد]

[[استاد [دانشگاه موردعلاقه]] من] آمد]



سوال ۲) در جای خالی کلمه یا عبارتی مناسب قرار دهید. (۲۰ نمره)

الف. در یک عبارت اسمی، به سایر کلمات به غیر از هسته (Head) می‌گویند.

پاسخ: توضیح‌دهنده، اصلاح‌کننده یا - modifier یا وابسته (dependant): هر دو درست است.

ب. به عبارتی که هسته آن اسم است و به عبارتی که هسته آن فعل است گویند.

پاسخ: NP – VP

ج. در جمله‌ی «She cats pizza with a fork» ابهام از نوع وجود دارد.

پاسخ: PP attachment

د. درخت حاصل از الگوریتم CYK یک درخت است.

پاسخ: دودویی

سوال (۳) می‌خواهیم یک مترجم ماشینی فارسی به انگلیسی داریم ولی متأسفانه دادگان موازی با حجم خیلی بزرگ نداریم. خوشبختانه مترجم گوگل در دسترس است. چه راهکاری پیشنهاد می‌دهید؟ (۳۰ نمره)

پاسخ: backtranslation

backtranslation is a way of using monolingual data for improving machine translation. In this technique monolingual data is translated using a machine translation system to create synthetic parallel data. Finally a new model trained on existing parallel data and synthetic parallel data in reverse direction.