

Przetwarzanie języka naturalnego
Ćwiczenia 5
Zajęcia w przedostatnim tygodniu semestru

Zadanie 1. (2p) ★ Pokaż, że problem należenia słowa do języka generowanego przez gramatykę atrybutową (z atrybutami pochodzącymi ze skończonego zbioru) jest NP-trudny.

Wskazówka: (rot13.com): fcebohwx mnxqbqbjnp ceboyrrz FNG. Flzobyr avrgrezvanyar tenznqlxvz bbtm cemrpubljlnp vasbeznpwr b jnegbfpv cbqsbezhyl benm b jnegbfpvbjnavh jfmflfgxvpu mzvraalpu. Jlcebijnqmna! wrmlx zbmr olp qbfj geljvnyal.

Zadanie 2. (2p) Ograniczoną przez $p > 0$ gramatyką PCFG nazwiemy parę składającą z liczby p oraz gramatyki PCFG. Gramatyka (p, G) generuje te słowa, które są generowane przez G , dla których najbardziej prawdopodobne drzewo ma prawdopodobieństwo większe niż p .

Scharakteryzuj zbiór języków generowanych przez k -ograniczone PCFG. Jak widać (?) nie jest to szczególnie ciekawe pojęcie. Zaproponuj jakiś jego wariant, który wyda Ci się mniej trywialny. Uwaga: o gramatykach PCFG jest na wykładzie 12 (już opublikowanym)

Zadanie 3. Połączenie reguł Bottom-Up Predict oraz Fundamental Rule (zob. wykład 11) nie daje kompletnego parsera. Jakiej reguły (reguł) brakuje?

Zadanie 4. Połączenie reguł Top-Down Predict oraz Fundamental Rule (zob. wykład 11) nie daje kompletnego parsera. Jakiej reguły (reguł) brakuje?

Zadanie 5. Powiedz, jak zaadaptować Chart parser do gramatyk PCFG.

Zadanie 6. Przesłuchaj piosenkę „My jesteśmy tanie dranie” z Kabaretu Starszych Panów (bez problemu znajdziesz na YT). Dwie jej zwrotki mają bardzo specyficzną konstrukcję rytmiczno/gramatyczną. Powiedz, jak można wykorzystać tę wiedzę do automatycznego generowania podobnych tekstów (ale o potencjalnie innej tematyce). Postaraj się, by Twoja odpowiedź dawała możliwie pełne wskazówki: „jak to zaimplementować”. Jak chcesz, to możesz zamienić na pracowni Pana Tadeusza na Tanich Draniów.

Zadanie 7. Poniżej 3 zdania z korpusu polskiego zależnościowego

Ale w wielu przypadkach znacznie groźniej wyglądają koszty dodatkowe.
Chodzili między regałami i włożyli do koszyka aparat fotograficzny.
Ewa ma bliźniowca, operacja plastyczna mogłaby to jeszcze pogorszyć.

Zaproponuj rozbiór zależnościowy tych zdań (bez etykiet na krawędziach). Sprawdź, czy udało Ci się zaproponować ten rozbiór, co twórcom korpusu, skomentuj ewentualne różnice. (Plik: `p1_lfg-ud-test.conllu`)

Zadanie 8. Wybierz zdanie z powyższej trójki i przedstaw etykiety znajdujące się na strzałkach (zgadując ich semantykę, lub odczytując z dokumentacji korpusu).

Zadanie 9. W korpusie LFG (tym z zadań powyżej) są również 'dodatkowe zależności'. Znajdź zdanie, w którym występuje *Wójeł Józiek* i wyjaśnij na przykładzie, do czego one służą. Znajdź również i zaprezentuj inny przykład, w którym zależność „dodatkowa” (siostrzana) wskazuje na inny węzeł, niż zależność główna.