Przetwarzanie języka naturalnego
Ćwiczenia 5
Zajęcia 24 i 28 stycznia

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

Wskazówka (trochę silniejsza niż ostatnio): spróbuj zakodować problem SAT. Symbole nieterminalne gramatyki moga przechowywać informacje o wartości podformuły oraz o wartosciowaniu wszystkich zmiennych (to drugie powinno być przekazywane bez zmian). Możesz pytać o należenie słowa pustego do generowanego języka.

Zadanie 2. Powiedz, jak zaadaptować Chart parser do gramatyk PCFG.

Zadanie 3. 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 4. Zaprojektuj struktury danych dające możliwości generowania wierszy przypominających Pana Tadeusza, złożonych w pełni ze zdań z korpusu, ale (tak jak w oryginale) bez zachowania odpowiedniości dwuwers=zdanie (czyli zdania mogą się zaczynać lub kończyć w dowolnym miejscu wersu, w którym może się zaczynać lub kończyć słowo). Zaproponuj jakiś sposób uczynienia takich tekstów ciekawszych treściowo przy użyciu zanurzeń słów (lub form bazowych).

Zadanie 5. Rozważmy następujące zadanie: na wejściu bierzemy tekst w języku polskim (każde zdanie w osobnym wierszu) i dla wszystkich przyimków w nim występujących mówimy z jakim czasownikiem, imiesłowem, rzeczownikiem lub przymiotnikiem się łączą. Przykładowe dane:

```
Stefan podczas pobytu na wywiadówce był dumny ze swojego syna.
Nad morzem można wypocząć lepiej niż w mieście.
Księżniczka o blond włosach czekała na przyjście rycerza na białym koniu.
Wiszący nad przepaścią turysta zastanawiał się, o czym powiedzieć ratownikom w pierwszej kolejności.
Stefan mieszkał w domu nad rzeką.
```

Wynik dla tych danych powinien być następujący:

```
był-podczas dumny-ze pobyt-na
wypocząć-nad wypocząć-w
księżniczka-o czekała-na#1 rycerza-na#2
wiszący-nad powiedzieć-o powiedzieć-w
mieszkał-w domu-nad
```

Zaproponuj jakąś metodę rozwiązywania tego zadania. Wiedzę o języku powinieneś czerpać z dużego nieoznakowanego korpusu oraz małego treebanku.

Zadanie 6. Model N-gramowy może generować kolejne słowa od lewej do prawej, można też użyć do generowania słów w odwrotną stronę. Na wykładzie podawaliśmy prawdopodobieństwo zdania w modelu na przykład bigramowym liczone od lewej do prawej. Czy prawdopodobieństwo liczone „od tyłu" się różni? Czy odpowiedź na to pytanie zależy od sposobu wygładzania prawdopodobieństw?