

Imię i nazwisko:

Nr indeksu:

1. Rozpatrujemy parę zdań „jest to książka” = „this is a book” i dopasowanie $1 \rightarrow 2, 2 \rightarrow 1, 1 \rightarrow 3, 3 \rightarrow 4$. Która z podanych fraz NIE zostanie pozyskana:
 - „jest to” = „is a”
 - „jest” = „is a”
 - „książka” = „book”
 - „jest to książka” = „this is a book”
2. Przyjmijmy prostą estymację prawdopodobieństw unigramów wykonanych na jakimś typowym polskim korpusie. Jakiej wartości logprob (przy podstawie 10) należałoby oczekiwać dla najczęstszego polskiego wyrazu?
 - ≈ -2
 - ≈ 0.95
 - ≈ 0.01
 - ≈ -0.5
3. System przetłumaczył zdanie ”Ala ma kota” na ”Alice has cat”, tłumaczenie referencyjne brzmi ”Alice has a cat”. Precision i recall (na unigramach):
 - precision wynosi 100%, recall — 75%
 - precision wynosi 66%, recall — 50%
 - precision wynosi 75%, recall — 100%
 - precision wynosi 100%, recall — 66%
4. Rozpatrujemy korpus złożony z 2 par zdań: „foo bar” = „rach ciach” i „bar baz” = „rach ciach bach” (zakładamy, że w ogóle nie ma innych wyrazów). Jaka będzie wartość $P(\text{rach}|\text{bar})$ po inicjalizacji równym prawdopodobieństwem (krok wstępny algorytmu EM dla Modelu 1)? Uwzględniamy token NULL.
 - $1/3$
 - $1/4$
 - $2/5$
 - $1/2$
5. Tłumaczymy zdanie „Ala ma kota” na angielski. W czasie dekodowania rozpatrujemy frazy (1) „Ala” = „Alice”, (2) „ma” = „has”, (3) „ma” = „is”, (4) „kota” = „a cat”, (5) „kota” = „the cat”, (6) „ma kota” = „has the cat” (7) „ma kota” = „is crazy”, Przyjmujemy trigramowy model języka.
 - Hipotezy (1)-(3)-(5) i (1)-(6) można połączyć, ale nie (1)-(2)-(4) i (1)-(3)-(5).
 - Hipotezy (1)-(3)-(4) i (1)-(6) można połączyć, ale nie (1)-(2)-(4) i (1)-(2)-(5).
 - Hipotezy (1)-(2)-(4) i (1)-(2)-(5) można połączyć, ale nie (1)-(2)-(5) i (1)-(6).
 - Hipotezy (1)-(2)-(4) i (1)-(2)-(5) można połączyć, również (1)-(6) i (1)-(7).
6. Rozpatrzmy podzielone na frazy zdania: „Alice|has got|a white cat” i „Białego kota|ma|Ala”. Ile wynosi reordering distance dla „ma” (zakładamy naturalne, zgodne z sensem, dopasowanie fraz)?

- -5
 - -4
 - -3
 - -2
7. System tłumaczy z angielskiego na polski zdanie "Ala ma kota" i zwraca jako najlepsze tłumaczenie "Alice has a dog" z oceną 0.64, podczas gdy tłumaczenie "Alice has a cat" ma ocenę 0.72.
- Mamy do czynienia z błędem przeszukiwania.
 - Mamy do czynienia z błędem modelu.
 - Mamy do czynienia zarówno z błędem modelu, jak i przeszukiwania.
 - Zachowanie systemu jest bezbłędne.
8. Dla Modelu 1 dla korpusu złożonego z 2 par zdań: „foo bar baz”=„rach ciach trach” i „bar baz foo”=„aha ciach rach”, jaka będzie wartość $P(\text{rach}|\text{bar})$ po wielu iteracjach algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy NULL).
- 1/2
 - 1/3
 - 2/3
 - 0
9. Tłumaczymy z polskiego na angielski. Zdanie polskie ma 4 wyrazy, zaś zdanie angielskie — 3 wyrazy. Ile jest możliwych funkcji urównoleglenia (alignment functions) - niekoniecznie zgodnych z sensem wyrazów)?
- 125
 - 256
 - 25
 - 64
10. Symetryzacja dopasowań. Dla zdań „Czy Ala kota ma” i „Does Ala has a cat” mamy funkcję dopasowania: (w kierunku polsko-angielskim) $1 \rightarrow 4, 2 \rightarrow 2, 3 \rightarrow 4, 4 \rightarrow 3, 5 \rightarrow 3$; (w kierunku angielsko-polskim) $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 5, 4 \rightarrow 3$. Efekt GROW-DIAG-FINAL to:
- 1-1, 2-2, 3-4, 3-5, 4-3
 - 1-1, 2-2, 3-4, 3-5, 4-1, 4-3
 - 2-2, 3-5, 4-3
 - 2-2, 3-4, 3-5, 4-3
11. Dla Modelu 1 dla korpusu złożonego z 2 par zdań: „foo bar baz”=„rach ciach trach” i „bar baz foo”=„trach ciach rach”, jaka będzie wartość $P(\text{rach}|\text{bar})$ po wielu iteracjach algorytmu EM (inicjalizujemy równomiernym prawdopodobieństwem i uwzględniamy NULL).
- 1/3
 - 1/2
 - 2/3
 - 0

12. Tłumaczymy z polskiego na angielski. Jako tłumaczenie zdania „Ala ma psa” rozpatrujemy zdanie „Alice has a dog”. Możliwa, naturalna (zgodna z sensem wyrazów) funkcja urównoleglenia (alignment function) to:
- $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow \text{NULL}, 4 \rightarrow 3$
 - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 4$
 - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3$
 - $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 2, 4 \rightarrow \text{NULL}$
13. Mamy korpus złożony z 2 par zdań: „foo bar”=„rach ciach aha” i „foo bar”=„rach trach ciach” (zakładamy, że w ogóle nie ma innych wyrazów). Jaka będzie wartość $a(2|1,3,2)$ po inicjalizacji równym prawdopodobieństwem (krok początkowy algorytmu EM dla Modelu 2)? Uwzględniamy NULL.
- $1/3$
 - $2/3$
 - $1/4$
 - $1/2$
14. Symetryzacja dopasowań. Dla zdań „Ala ma kota” i „Alice has a cat” mamy funkcję dopasowania: (w kierunku polsko-angielskim) $1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow \text{NULL}, 4 \rightarrow 3$; (w kierunku angielsko-polskim) $1 \rightarrow \text{NULL}, 2 \rightarrow 2, 3 \rightarrow 4$. Suma to:
- 1-1, 2-2, 3-4
 - 2-2
 - 2-2, 3-4
 - 1-1, 2-2, 3-3, 3-4
15. Załóżmy, że $P(\text{rach}|\text{foo})=0.4, P(\text{ciach}|\text{foo})=0.6, P(\text{rach}|\text{bar})=0.3, P(\text{ciach}|\text{bar})=0.7, P(\text{rach}|\text{baz})=0.8, P(\text{ciach}|\text{baz})=0.2, P(\text{ciach}|\text{NULL})=P(\text{rach}|\text{NULL})=0.1$. Jakie jest najbardziej prawdopodobne dopasowanie w Modelu 1 dla zdania źródłowego „foo bar baz” i zdania docelowego „rach ciach”?
- $1 \rightarrow 3, 2 \rightarrow 2$
 - $1 \rightarrow 2, 2 \rightarrow 2, 3 \rightarrow 1$
 - $1 \rightarrow 3, 2 \rightarrow 3,$
 - $1 \rightarrow \text{NULL}, 2 \rightarrow 1$
16. Na podstawie pewnego korpusu równoległego stwierdzono, że prawdopodobieństwo przetłumaczenia wyrazy „kapelusz” na „hat” wynosi 0,65. Prawdopodobieństwo, że odpowiednikiem wyrazu „hat” jest „kapelusz” wynosi w tym wypadku:
- pytanie zawiera za mało danych, by podać to prawdopodobieństwo
 - 0,65
 - 1
 - 0,325