

## Пофразов превод (РЕШЕНИЕ)

При обучението на система за пофразов машинен превод (*phrase-based machine translation*) на системата се дават изречения на два езика с установени съответствия между отделни думи и от тях тя учи двойки преводни еквиваленти на непрекъснати последователности от съответстващи си множества от думи. (Името на метода идва оттам, че на английски тези последователности се наричат *phrases 'фрази'*, макар и често да не образуват едно цяло и в друг контекст да не се наричат така.)

Например от английското изречение *She<sub>1</sub> gave<sub>2</sub> him<sub>3</sub> water<sub>4</sub>* и съответното българско *Тя<sub>1</sub> му<sub>3</sub> даде<sub>2</sub> вода<sub>4</sub>* (индексите показват коя дума на коя отговаря) могат да се научат трите двойки последователности

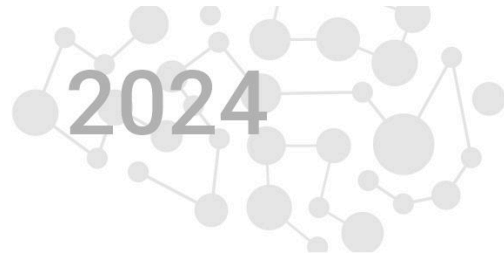
- *She gave him* – *Тя му даде*,
- *gave him* – *му даде*,
- *gave him water* – *му даде вода*,

както и двойките *She* – *Тя*, *gave* – *даде*, *him* – *му*, *water* – *вода* и *She gave him water* – *Тя му даде вода*, които обаче са известни, тъй като са между думите и целите изречения.

### Задача 1.

Маркирайте всички последователности, които могат да бъдат извлечени от изречението на български *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>*.

*три бели*  
*три котки*  
*бели котки*  
*бели пият*  
*три пият*  
*котки пият*  
*три мляко*  
*бели мляко*  
*котки мляко*  
*пият мляко*  
*три бели котки*  
*три котки пият*  
*бели котки пият*  
*бели котки мляко*  
*бели пият мляко*  
*котки пият мляко*  
*три бели котки пият*  
*три котки пият мляко*  
*бели котки пият мляко*  
*мляко пият котки бели*



### Решение:

Трябва да се избераат само непрекъснати последователности от 2, 3 или 4 думи, които присъстват в българското изречение, затова правилните отговори са отбелязаните по-долу:

- ✓ три бели  
три котки
- ✓ бели котки  
бели пият  
три пият
- ✓ котки пият  
три мляко  
бели мляко  
котки мляко
- ✓ пият мляко
- ✓ три бели котки  
три котки пият
- ✓ бели котки пият  
бели котки мляко  
бели пият мляко
- ✓ котки пият мляко
- ✓ три бели котки пият  
три котки пият мляко
- ✓ бели котки пият мляко  
мляко пият котки бели

### Задача 2.

По-долу ще видите преводите на българското изречение *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>* на няколко други езика, като еднаквите индекси на думите в изречението на български и на чуждия език показват преводните съответствия между думите.

За всеки чужд език определете преводните еквиваленти на дадените последователности, които могат да се научат от двойката изречения (на български и на чуждия език), по следния начин:

- За всяка последователност на български запишете съответстващия ѝ преводен еквивалент на чуждия език като поредица от цифри (без разделители), където всяка цифра съответства на индекса на дума в изречението на чуждия език, а редът на цифрите отразява последователността на думите в изречението на чуждия език.
- Ако някоя последователност на български не може да бъде съотнесена с валидно преводно съответствие на другия език, означете това с 0.

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>*.

ирландски: *Ólann<sub>4</sub> trí<sub>1</sub> chat<sub>3</sub> bhána<sub>2</sub> bainne<sub>5</sub>*.

кюрдски: *Sê<sub>1</sub> pisîkên<sub>3</sub> spî<sub>2</sub> şîr<sub>5</sub> dixwin<sub>4</sub>*.

малгашки: *Misotro<sub>4</sub> ronono<sub>5</sub> saka<sub>3</sub> fotsy<sub>2</sub> telo<sub>1</sub>*.

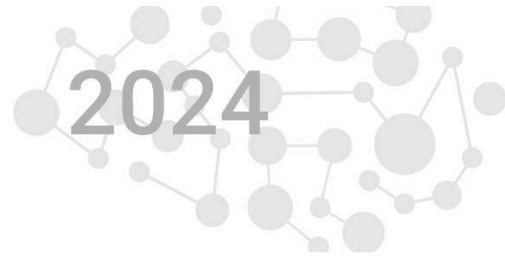
португалски: *Três<sub>1</sub> gatos<sub>3</sub> brancos<sub>2</sub> bebem<sub>4</sub> leite<sub>5</sub>*.

турски: *Üç<sub>1</sub> beyaz<sub>2</sub> kedi<sub>3</sub> süt<sub>5</sub> içer<sub>4</sub>*.

чеченски: *Klâйн<sub>2</sub> кхо<sub>1</sub> цициг<sub>3</sub> шура<sub>5</sub> молу<sub>4</sub>*.



## СЪСТЕЗАНИЕ по компютърна лингвистика



хавайски: *Inu<sub>4</sub> waii<sub>5</sub> 'eko<sub>1</sub>u<sub>1</sub> pōroki<sub>3</sub> ke 'oke 'o<sub>2</sub>.*

### Решение:

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*

ирландски: *Ólann<sub>4</sub> trí<sub>1</sub> chat<sub>3</sub> bhána<sub>2</sub> bainne<sub>5</sub>.*

<i>Три бели</i>	0
<i>бели котки</i>	32
<i>котки пият</i>	0
<i>пият мляко</i>	0
<i>Три бели котки</i>	132
<i>бели котки пият</i>	0
<i>котки пият мляко</i>	0
<i>Три бели котки пият</i>	4132
<i>бели котки пият мляко</i>	0

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*

кюрдски: *Sê<sub>1</sub> pisikên<sub>3</sub> spî<sub>2</sub> şîr<sub>5</sub> dixwin<sub>4</sub>.*

<i>Три бели</i>	0
<i>бели котки</i>	32
<i>котки пият</i>	0
<i>пият мляко</i>	54
<i>Три бели котки</i>	132
<i>бели котки пият</i>	0
<i>котки пият мляко</i>	0
<i>Три бели котки пият</i>	0
<i>бели котки пият мляко</i>	3254

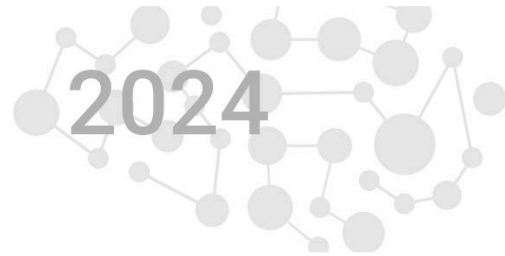
български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*

малгашки: *Misotro<sub>4</sub> ronono<sub>5</sub> saka<sub>3</sub> fotsy<sub>2</sub> telo<sub>1</sub>.*

<i>Три бели</i>	21
<i>бели котки</i>	32
<i>котки пият</i>	0
<i>пият мляко</i>	45
<i>Три бели котки</i>	321



## СЪСТЕЗАНИЕ по компютърна лингвистика



бели котки пият	0
котки пият мляко	453
Три бели котки пият	0
бели котки пият мляко	4532

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*  
португалски: *Três<sub>1</sub> gatos<sub>3</sub> brancos<sub>2</sub> bebem<sub>4</sub> leite<sub>5</sub>.*

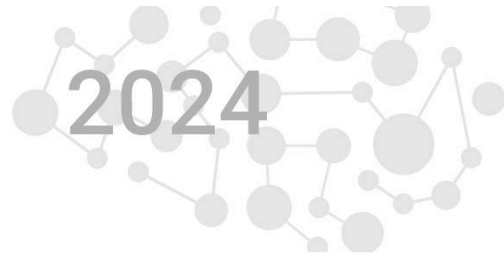
Три бели	0
бели котки	32
котки пият	0
пият мляко	45
Три бели котки	132
бели котки пият	324
котки пият мляко	0
Три бели котки пият	1324
бели котки пият мляко	3245

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*  
турски: *Üç<sub>1</sub> beyaz<sub>2</sub> kedi<sub>3</sub> süt<sub>5</sub> içer<sub>4</sub>.*

Три бели	12
бели котки	23
котки пият	0
пият мляко	54
Три бели котки	123
бели котки пият	0
котки пият мляко	354
Три бели котки пият	0
бели котки пият мляко	2354

български: *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>.*  
чеченски: *Клайн<sub>2</sub> кхо<sub>1</sub> цициг<sub>3</sub> шура<sub>5</sub> молу<sub>4</sub>.*

Три бели	21
бели котки	0



котки пият	0
пият мляко	54
Три бели котки	213
бели котки пият	0
котки пият мляко	354
Три бели котки пият	0
бели котки пият мляко	0

български:  $Три_1$   $бели_2$   $котки_3$   $пият_4$   $мляко_5$ .  
хавайски:  $Ини_4$   $waii_5$  'eko $lu_1$  pōroki $_3$  ke 'oke 'o $_2$ .

Три бели	0
бели котки	32
котки пият	0
пият мляко	45
Три бели котки	132
бели котки пият	0
котки пият мляко	0
Три бели котки пият	0
бели котки пият мляко	0

### Задача 3.

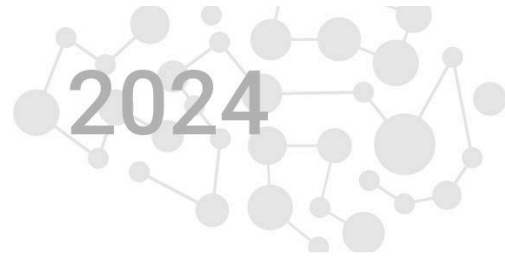
Изречението на български  $Три_1$   $бели_2$   $котки_3$   $пият_4$   $мляко_5$  се предава на суахили с изречение, което се състои от пет думи, дадени тук в азбучен ред: *hunuwa* 'пият', *maziwa* 'мляко', *nyeupe* 'бели', *paka* 'котки', *tatu* 'три'. Както на български, на суахили съответствието на *пият* е по-близо до началото на изречението, отколкото съответствието на *мляко*, а съответствията на *три*, *бели* и *котки* са три съседни думи.

Подредете думите, за да възстановите изречението на суахили, като знаете, че от него и българското му съответствие биха могли да се извлекат точно 5 двойки преводни еквиваленти на последователности от думи (освен състоящите се от по една дума или от цялото изречение).

**Решение:**  $Paka_3$   $nyeupe_2$   $tatu_1$   $hunuwa_4$   $maziwa_5$ .

С израза в скоби **(1, 2, 3)** означаваме последователността, която съответства на *три бели котки* на суахили без значение от реда на думите, т.е. може да означава всяка от последователностите **123, 132, 213, 231, 312** или **321**.

От това, че съответствието на *пият* е по-близо до началото на изречението, отколкото съответствието на *мляко*, получаваме възможностите **4(1, 2, 3)5** или **(1, 2, 3)45**.



Знаем също, че от изречението на суахили и българското му съответствие биха могли да се извлекат точно 5 двойки преводни еквиваленти.

В първия вариант  $4(1, 2, 3)5$  имаме преводни съответствия на: 1) *три бели котки* –  $(1, 2, 3)$ ; 2) на *три бели котки пият* –  $4(1, 2, 3)$ ; и 3) вътре в рамките на  $(1, 2, 3)$  имаме максимум още две последователности с преводно съответствие. От това следва, че в този случай не можем да получим пет преводни съответствия.

Следователно решението е измежду възможните подреждания на  $(1, 2, 3)45$ . В този случай имаме преводни съответствия на: 1) *три бели котки* –  $(1, 2, 3)$ , независимо от подредбата; 2) на *пият мляко* –  $45$ ; 3) на *три бели котки пият* –  $(1, 2, 3)4$ ; и 4) вътре в  $(1, 2, 3)$  имаме поне едно преводно съответствие от две думи с поредни индекси. За да получим точно пет двойки преводни еквиваленти, в  $(1, 2, 3)$  трябва да имаме точно 2 преводни съответствия, а това е възможно само при подредба  $321$  или  $123$ . Ако  $3$  е последна дума вътре в  $(1, 2, 3)$ , то ще имаме още 2 съответствия –  $34$  и  $345$ , с което надхвърляме разрешените пет. Остава подредбата  $321$ .

Решението е: **32145**.

#### Задача 4.

На даден чужд език  $X$  българското изречение *Три<sub>1</sub> бели<sub>2</sub> котки<sub>3</sub> пият<sub>4</sub> мляко<sub>5</sub>* се превежда с пет думи, а съответствията на *три*, *бели* и *котки* са три съседни думи (в един или в друг ред).

Колко най-малко двойки преводни съответствия ще могат да се извлекат от изречението на български и изречението на езика  $X$  (освен състоящите се от по една дума или от цялото изречение)?

Оценете валидността на всяко от по-долните твърдения въз основа на дадената информация относно преводните съответствия в изреченията на езика  $X$  и на български. Изберете „Няма достатъчно информация“, ако наличната информация за преводните съответствия в изреченията на езика  $X$  и на български не е достатъчна, за да се оцени валидността на твърдението.

Възможности за избор: „Вярно“, „Невярно“, „Няма достатъчно информация“.

Последователността от български думи *три бели котки* има съответстваща преводна последователност от думи в изречението на езика  $X$ .

Последователността от думи *три бели котки* в българското изречение може да няма съответстваща преводна последователност от думи в изречението на езика  $X$ .

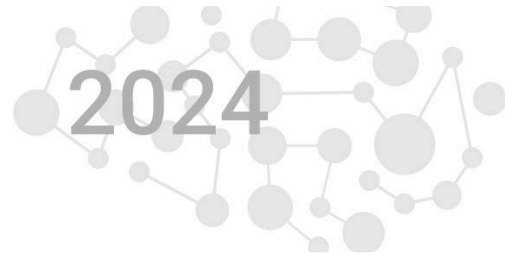
В последователността от думи в изречението на езика  $X$ , която отговаря на *три бели котки*, има две съседни думи с преводно съответствие в изречението на български.

Последователността *бели котки* от българското изречение може да се съотнесе с преводна последователност в изречението на езика  $X$ .

Последователността *пият мляко* от българското изречение може да се съотнесе с преводна последователност в изречението на езика  $X$ .

Ако съответствието на *пият* в изречението на езика  $X$  не е съседно на съответствието на *три бели котки*, то ще е съседно на съответствието на *мляко*.

Съответствието на *пият* в изречението на езика  $X$  е съседно както на съответствието на *три бели котки*, така и на съответствието на *мляко*.



### Решение:

От изречението на български и изречението на езика X могат да се извлекат най-малко три преводни съответствия (освен състоящите се от по една дума или от цялото изречение).

За двойката изречения на български и на езика X знаем, че:

1. В изречението на езика X имаме преводно съответствие на *три бели котки*.
2. В изречението на езика X имаме преводно съответствие или на *три бели*, или на *бели котки* (или и на двете).
3. Съответствието на *пият* в изречението на езика X се намира или до съответствието на *три бели котки* и имаме преводен еквивалент на *три бели котки пият*, или до съответствието на *мляко* и имаме преводен еквивалент на *пият мляко* (или и двете).

От това следва, че могат да се извлекат поне три преводни съответствия (освен състоящите се от по една дума или от цялото изречение).

- Последователността от български думи *три бели котки* има съответстваща преводна последователност от думи в изречението на езика X.

**Вярно.** От дадената информация можем да заключим, че твърдението е вярно, защото е дадено, че съответствията на *три*, *бели* и *котки* са три съседни думи в изречението на езика X, независимо от реда, което е преводно съответствие на българското словосъчетание *три бели котки*.

- Последователността от думи *три бели котки* в българското изречение може да няма съответстваща преводна последователност от думи в изречението на езика X.

**Невярно.** От дадената информация можем да заключим, че твърдението е невярно, тъй като знаем, че *три*, *бели* и *котки* са три съседни думи в изречението на езика X и образуват последователност, която е преводен еквивалент на *три бели котки*.

- В последователността от думи в изречението на езика X, която отговаря на *три бели котки*, има две съседни думи с преводно съответствие в изречението на български.

**Вярно.** От дадената информация можем да заключим, че твърдението е вярно, защото възможните последователности в изречението на езика X са: **123** (две двойки последователности от съседни думи **12** и **23**), **132** (последователност от съседни думи е **32**), **213** (последователност от съседни думи е **21**), **231** (последователност от съседни думи е **23**), **312** (последователност от съседни думи е **12**), **321** (две двойки последователности от съседни думи **32** и **21**).

- Последователността *бели котки* от българското изречение може да се съотнесе с преводна последователност в изречението на езика X.

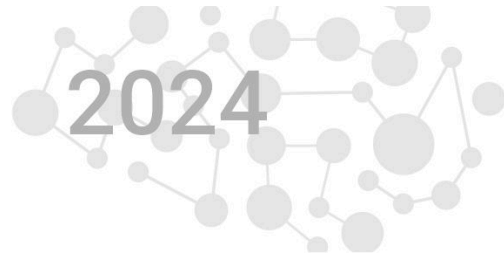
**Няма достатъчно информация.** От дадената информация не можем да заключим, че твърдението е вярно или невярно, тъй като на езика X може да има подредба, в която съответствията на *бели* и *котки* не са съседни и не образуват последователност, например **213** или **312**.

- Последователността *пият мляко* от българското изречение може да се съотнесе с преводна последователност в изречението на езика X.

**Няма достатъчно информация.** От дадената информация не можем да заключим, че твърдението е вярно или невярно, тъй като на езика X може да има



СЪСТЕЗАНИЕ  
по компютърна  
лингвистика



подредба, в която съответствията на *пият* и *мляко* не са съседни и не образуват последователност – едното да бъде преди съответствието на *три бели котки*, а другото да бъде след, например **4(1, 2, 3)5** или **5(1, 2, 3)4**.

- Ако съответствието на *пият* в изречението на езика *X* не е съседно на съответствието на *три бели котки*, то то ще е съседно на съответствието на *мляко*.

**Вярно.** От дадената информация можем да заключим, че твърдението е вярно, защото възможните последователности в изречението на езика *X*, при които глаголят **4** не е съседен на **(1, 2, 3)**, са: **(1, 2, 3)54** (съседен е на **5** – ‘мляко’) и **45(1, 2, 3)** (отново е съседен на **5** – ‘мляко’).

- Съответствието на *пият* в изречението на езика *X* е съседно както на съответствието на *три бели котки*, така и на съответствието на *мляко*.

**Няма достатъчно информация.** От дадената информация не можем да заключим, че твърдението е вярно или невярно, тъй като на езика *X* съответствието на *пият* може да е съседно и на двете **(1, 2, 3)45**, но може и да не е – например в конфигурацията **4(1, 2, 3)5**.