

Анотация (РЕШЕНИЕ)

Задача 1.

В компютърната лингвистика в много случаи е важно да се определи честотата на думите и тяхното групиране по различни признаци. Числата в скобите показват колко пъти се е срещнала всяка от изброените думи (независимо от значението си, ако има еднакви форми) в Българския национален корпус, който съдържа 4,5 милиона думи (search.dcl.bas.bg). След тях са дадени означения, които илюстрират определен тип групиране между думите.

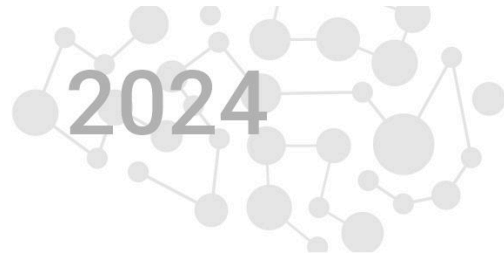
<i>вечер</i>	(52 919)	C49
<i>степен</i>	(29 576)	C49
<i>филм</i>	(14 578)	C7
<i>дъб</i>	(4547)	C1
<i>безгрижен</i>	(4393)	П79
<i>търг</i>	(2947)	C1
<i>локомотив</i>	(1968)	C7
<i>каляска</i>	(1688)	C41
<i>диаграма</i>	(1207)	C41
<i>низ</i>	(1105)	C1
<i>дрешник</i>	(778)	C14
<i>чест</i>	(553)	П76
<i>царствен</i>	(325)	П76
<i>чартърен</i>	(298)	П79
<i>топчест</i>	(93)	П76
<i>синус</i>	(70)	C7
<i>безбрежен</i>	(54)	П79
<i>осмоъгълник</i>	(22)	C14

А) Възстановете пропуснатото:

<i>зъб</i>	(37 044)	?
<i>пасквил</i>	(46)	?
<i>ерес</i>	(559)	?
<i>видим</i>	(1686)	?
<i>безредица</i>	(915)	?
<i>жанр</i>	(968)	?
<i>живителен</i>	(83)	?
<i>крепост</i>	(10 751)	?



СЪСТЕЗАНИЕ по компютърна лингвистика



<i>рояк</i>	(1780)	?
<i>рус</i>	(23 806)	?

Решение:

зъб (37 044)

С7 – съществително от м. р. с окончание **-и** в мн. ч. Думата *зъб* образува формите си по същия начин като думите *локомотив*, *синус* и *филм*: *зъб*, *зъба*, *зъбът*, *зъби*, *зъбите*, (два) *зъба*. Записът С7 означава, че думата е съществително име от флективен тип 7.

пасквил (56)

С7 – съществително от м. р. с окончание **-и** в мн. ч. Думата *пасквил* образува формите си по същия начин като думите *локомотив*, *синус* и *филм*: *пасквил*, *пасквила*, *пасквилът*, *пасквили*, *пасквилите*, (два) *пасквила*. Записът С7 означава, че думата е съществително име от флективен тип 7.

ерес (559)

С49 – съществително от ж. р., което завършва на съгласна. Думата *ерес* образува формите си по същия начин като думите *степен* и *вечер*: *ерес*, *ереста*, *ереси*, *ересите*. Записът С49 означава, че думата е съществително име от флективен тип 49.

видим (1686)

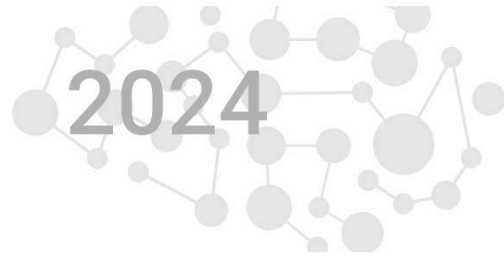
П76 – прилагателно без особености. Думата *видим* образува формите си по същия начин като думите *топчест* и *царствен*: *видим*, *видимия*, *видимият*, *видима*, *видимата*, *висимо*, *видимото*, *видими*, *видимите*. Записът П76 означава, че думата е прилагателно име от флективен тип 76.

безредица (915)

С41 – съществително от ж. р. с окончание **-а**. Думата *безредица* образува формите си по същия начин като думите *диаграма* и *каляска*: *безредица*, *берицицата*, *безредици*, *безредиците*. Записът С41 означава, че думата е съществително име от флективен тип 41.

жанр (968)

С1 – съществително от м.р. с окончание **-ове** в мн.ч. Думата *жанр* образува формите си по същия начин като думите *дъб*, *низ* и *търг*: *жанр*, *жанра*, *жанрът*, *жанрове*, *жанровете*, (два) *жанра*. Записът С1 означава, че думата е съществително име от флективен тип 1.



живителен (83)

П79 – прилагателно, в което при формоизменението изпада **-е-**. Думата *живителен* образува формите си по същия начин като думите *безбрежен*, *безгрижен* и *чартърен*: *живителен*, *живителния*, *живителният*, *живителна*, *живителната*, *живително*, *живителното*, *живителни*, *живителните*. Записът П79 означава, че думата е прилагателно име от флективен тип 83.

крепост (10 751)

С49 – съществително от ж. р., което завършва на съгласна. Думата *крепост* образува формите си по същия начин като думите *степен* и *вечер*: *крепост*, *крепостта*, *крепости*, *крепостите*. Записът С49 означава, че думата е съществително име от флективен тип 49.

рояк (1780)

С14 – съществително от м. р., при което **-к-** се променя в **-ц-** пред **-и** в мн.ч. Думата *рояк* образува формите си по същия начин като думите *дрешник* и *осмогълник*: *рояк*, *рояка*, *роякът*, *рояци*, *рояците*, (два) *рояка*. Записът С14 означава, че думата е съществително име от флективен тип 14.

рус (23 806)

П76 – прилагателно без особености при формообразуването. Думата *рус* образува формите си по същия начин като думите *смръщен*, *топчест* и *царствен*: *рус*, *русия*, *русият*, *руса*, *русата*, *русо*, *русото*, *руси*, *русите*. Записът П76 означава, че думата е прилагателно име от флективен тип 76.

Задача 2.

1. Флективният тип на думата *жанр* е С1. Кои от изброените думи принадлежат на същия флективен тип?

Изберете една или повече: *лист*, *стоп*, *зрач*, *грис*, *рибарник*, *хипертекст*.

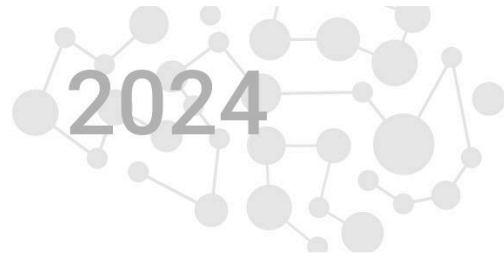
Решение:

Думите *лист*, *стоп* и *хипертекст* имат същите форми като думата *жанр*: *жанр*, *жанра*, *жанрът*, *жанрове*, *жанровете*, (два) *жанра*.

Думата *лист* има значение, при което формообразуването е същото като на думата *жест*: ‘Тънък, четвъртит къс хартия като част от книга, тетрадка, бележник и под. или взет отделно’.

Думите *зрач* и *грис* нямат множествено число и бройна форма.

Думата *рибарник* образува множествено число с промяна на **-к-** в **-ц-**.



2. Флективният тип на думата *пасквил* е С7. Кои от изброените думи принадлежат на същия флективен тип?

Изберете една или повече: *агнец*, *аметист*, *децибел*, *стрихнин*, *разпад*, *послеслов*.

Решение:

Думите *аметист*, *децибел*, *разпад*, *послеслов* имат същите форми като думата *пасквил*: *пасквил*, *пасквила*, *пасквилът*, *пасквили*, *пасквилите*, (два) *пасквила*.

Думите *агнец* и *стрихнин* нямат форми за множествено число и бройна форма.

3. Флективният тип на думата *рояк* е С14. Кои от изброените думи принадлежат на същия флективен тип?

Изберете една или повече: *гръбнак*, *мостик*, *грък*, *потомък*, *скициник*, *заек*.

Решение:

Думите *гръбнак*, *мостик*, *скициник* имат същите форми като думата *рояк*: *рояк*, *рояка*, *роякът*, *рояци*, *рояците*, (два) *рояка*.

Думата *грък* образува форми за множествено число и бройна форма с редуване на **-ръ-** в **-ър-**.

Думата *потомък* образува форми за множествено число и бройна форма с изпускане на **-ъ-**.

Думата *заек* образува форми за множествено число с вмьтане на **-й-** и изпускане на **-е-**.

4. Флективният тип на думата *безредица* е С41. Кои от изброените думи принадлежат на същия флективен тип?

Изберете едао или повече: *вендуза*, *отмора*, *изневяра*, *подагра*, *фреза*, *мряна*.

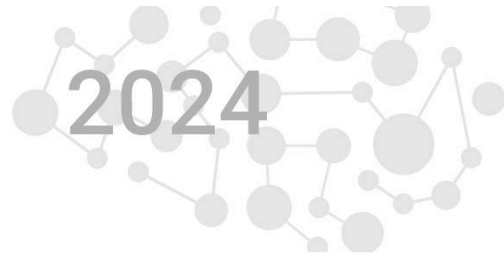
Решение:

Думите *вендуза*, *фреза* имат същите форми като думата *безредица*: *безредица*, *бедрицата*, *безредици*, *безредиците*.

Думите *отмора* и *подагра* нямат форми за множествено число.

Думите *изневяра* и *мряна* образуват множествено число с промяна на **я** (предходна мека съгласна и **а**) в **е**.

5. Флективният тип на думата *крепост* е С49. Кои от изброените думи принадлежат на същия флективен тип?



Изберете една или повече: *личност, местност, пакост, декоративност, женственост, унилоост.*

Решение:

Думите *личност, местност, пакост* имат същите форми като думата *крепост*: *крепост, крепостта, крепости, крепостите.*

Думите *декоративност, женственост* и *унилоост* нямат форми за множествено число.

6. Флективният тип на думата *рус* е П76. Кои от изброените думи принадлежат на същия флективен тип?

Изберете една или повече: *неканен, плашлив, риж, полувиси, ведър, врял.*

Решение:

Думите *неканен, плашлив, риж* имат същите форми като думата *рус*: *рус, русия, русият, руса, русата, русо, русото, руси, русите.*

Думата *полувиси* образува среден род с окончание *-е*.

Думата *ведър* образува форми за женски и среден род и множествено число с изпадане на *-ъ*.

Думата *врял* образува членувани форми за мъжки род и множествено число с промяна на **я** (мека гласна и **а**) в *е*.

7. Флективният тип на думата *живителен* е П79. Кои от изброените думи принадлежат на същия флективен тип?

Изберете една или повече: *водоравен, благотворен, витален, аварен, яносин, ухаен.*

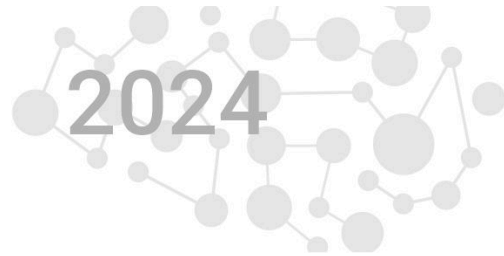
Решение:

7. Думите *водоравен, благотворен, витален* имат същите форми като думата *живителен*: *живителен, живителния, живителният, живителна, живителната, живително, живителното, живителни, живителните.*

Думите *аварен* и *ухаен* образуват формите си с изпадане на *е* и вмязане на **й**.

Думата *яносин* образува формите си без изпадане на гласна. Формите за среден род се образуват с окончание *-ьо* (мека съгласна и *о*).

Задача 3.



В компютърната лингвистика често срещан подход е да се създават модели на базата на данни, обогатени с лингвистична информация. Главните букви означават граматични класове думи, а малките букви и цифрите – категории в зависимост от образуването на формите на думите. **Определете правилните означения на подчертаните думи, като изберете един от следните формални записи.**

Гм2е Гм3е Гм3е Гм3е Гп2е Пмн Пмн Пмч Псеч СЖеч СЖеч СЖмн СЖмн СМмн СМмн СМмч ССен ССеч

Петър **покани** ____ много **гости** ____, като разпрати **бели** ____ и **червени** ____ **покани** ____ . Пригответи три **кани** ____ с лимонада и цял ден **бели** ____ картофи.

– **Покани** ____, Петре, и **сладкото** ____ момиче отсреща!

– Добре! – каза Петър, отиде до **къщата** ____ на **момичето** ____, спря се да купи **сладко** ____ от череша и изгори пържените **картофи** ____ .

Накрая **гости** ____ **гостите** ____ само с лимонада и сладко.

Решение:

Много от думите могат да имат повече от едно граматично значение: *покани* – съществително име в множествено число, *покани* – глагол в трето лице, единствено число, сегашно време; във второ или трето лице, единствено число, минало свършено време или във второ лице единствено число, повелително наклонение. В контекста думите се реализират както с едно от лексикалните си значения, ако думата е многозначна, така и с точно едно граматично значение. С главни букви са означени характеристиките на основната форма (т.е. характеристиките, които се отнасят до всички форми на дадена дума), а с малки букви и цифри – характеристиките на конкретната форма, с която е употребена думата. Означенията са позиционни: един и същи знак на различна позиция може да означава различна граматична характеристика.

В използваната конвенция означенията са:

С – съществително име

М – мъжки род на съществително име

Ж – женски род на съществително име

С – среден род на съществително име

Г – глагол

П – прилагателно име

с – среден род на прилагателно име

м – минало време на първа позиция след Г (глагол)

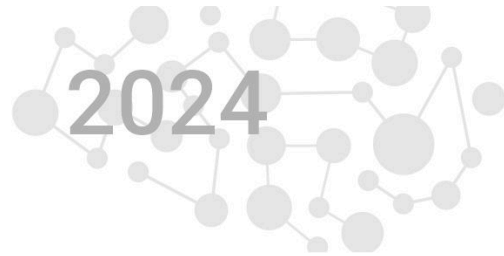
п – повелително наклонение

2 – второ лице

3 – трето лице



СЪСТЕЗАНИЕ
по компютърна
лингвистика



е – единствено число

м – множествено число на първа позиция след С (съществително) или П (прилагателно)

н – нечленувана форма

ч – членувана форма

Така формалните записи към подчертаните думи изразяват следните граматични характеристики:

покани Гм3е – глагол, минало свършено време, 3 л., ед. ч.

гости СМмн – съществително име, м. р., мн. ч., нечленувано

бели Пмн – прилагателно име, мн. ч., нечленувано

червени Пмн – прилагателно име, мн. ч., нечленувано

покани СЖмн – съществително име, ж. р., мн. ч., нечленувано

кани СЖмн – съществително име, ж. р., мн. ч., нечленувано

бели Гм3е – глагол, минало свършено време, 3 л., ед. ч.

поканѝ Гп2е – глагол, повелително наклонение, 2 л., ед. ч.

сладкото Псеч – прилагателно име в ср. р., ед. ч., членувано

къщата СЖеч – съществително име, ж. р., ед. ч., членувано

момичето ССеч – съществително име, ср. р., ед. ч., членувано

сладко ССен – съществително име, ср. р., ед. ч., нечленувано

картофи СМмн – съществително име, м. р., мн. ч., нечленувано

гости Гм3е – глагол, минало свършено време, 3 л., ед. ч.

гостите СМмч – съществително име, м. р., мн. ч., членувано

Петър **покани** Гм3е много **гости** СМмн, като разпрати **бели** Пмн и **червени** Пмн **покани** СЖмн. Пригответе три **кани** СЖмн с лимонада и цял ден **бели** Гм3е картофи.

– **Покани** Гп2е, Петре, и **сладкото** Псеч момиче отсреща!

– Добре! – каза Петър, отиде до **къщата** СЖеч на **момичето** ССеч, спря се да купи **сладко** ССен от череша и изгори пържените **картофи** СМмн.

Накрая **гости** Гм3е **гостите** СМмч само с лимонада и сладко.

Няма подчертани думи, които имат следните граматични характеристики, и следователно тези отговори са грешни:

Пмч – прилагателно име, мн. ч., членувано

Гм2е – глагол, минало свършено време, 2 л., ед. ч.

СЖен – съществително име, ж. р., ед. ч., нечленувано