

1.

,

:

		()	
	1. , : , , ; ;	-2 ,	
	2.	-2	
	1. , ; / ;	-2 ,	
	2. ;	-2	
(/)	1. -	-2 ,	

	2.	-2	

2.

« »

2 4
« »

3. (,) ,
- 1.

4.

4.1.

(-2)-

()	()				
		« »	« »	« »	« »
()	:				

	,	,	,	,	,
	:				
	,				
	,				
	,				
	;				
()	:				
	,	,	,	,	,
	,	,	,	,	,
	;	;	;	;	;
	;	;	;	;	;
	/	/	/	/	/
	;	;	;	;	;

			/ ;	/ ;	
()	:	-	-	-	-
		-	-	-	
		.			

-2-

()	(«	«	«	«
)	»	»	»	»

1-	1.	-2	,
	2.	-2	,
2-	1.	-2	,
	2.	-2	,
3-	1.	-2	,
	2.	-2	,

4.3

2.

1. PyCharm Edu Editor. Guided Tour
2. Managing Editor Tabs
3. Opening and Reopening Files in the Editor
4. Closing Files in the Editor
5. Saving and Reverting Changes
6. 'Lens' Mode

7. Multicursor
8. Scratches
9. Zooming in the Editor
10. TODO Comments
11. Basic Editing Procedures
12. Advanced Editing Procedures
13. PyCharm Edu Usage Guidelines
14. Creating and Managing Projects
15. Creating Projects
16. Opening, Reopening, and Closing Projects
17. Opening Multiple Projects
18. Cleaning System Cache
19. Configuring Project and IDE Settings
20. Accessing Settings
21. Configuring Keyboard Shortcuts
22. Configuring Menus and Toolbars
23. Configuring Quick Lists
24. Configuring Scopes and File Colors
25. Configuring Colors and Fonts
26. Configuring Code Style
27. Copying Code Style Settings
28. Configuring Individual File Encoding
29. Configuring Line Separators
30. Creating and Registering File Types
31. Switching Between Schemes
32. Configuring Project Structure
33. Accessing Project Structure
34. Configuring Content Roots
35. Configuring Folders Within a Content Root

:

« - »

« »

1

...

« _____ » _____ 2018 .

1. :

2. *P(not|do)*

<s> I am Sam </s>

<s> Sam I am </s>

<s> I do not like green eggs and ham </s>

... : ... , 1998 (1987). « »,
 « » . . . « »,
 « ».

(): 5
 - 4-5 , ,
 , ;
 - 1-3 , , ;
 -0- , , ;
 , , ;
 () :
 (- , , , ...
) , ,
 () , () .

() , .
 (British National Corpus (BNC)

[<http://www.natcorp.ox.ac.uk/>])

[<http://www.ruscorpora.ru/>]

() , ()
 / 1 2 , ()
 () : 1 (Mod1); 2
 (, gr401); 3 : Mod1_gr401_Petrov.

(): 5
 - 4-5 , ,
 , ;
 - 1-3 , , ;
 -0- , , ;
 , , ;

1. . Suppose "the" is the most frequent word in our corpus and occurs 200,000 times. The second most frequent word is "of", which occurs 190,000 times. Given just what we have explained so far, what would be the Good-Turing estimated count of "the"?

2. Suppose we train a language model with add-one smoothing on a given corpus C . How does the perplexity of the smoothed language model on that corpus, $PP_{smoothed}(C)$, compare with the perplexity of the maximum-likelihood language model, $PPML(C)$?

1. $PP_{smoothed}(C) = PPML(C)$
2. $PP_{smoothed}(C) < PPML(C)$
3. $PPML(C) < PP_{smoothed}(C)$
4. The answer depends on the corpus.

3. A traffic signal has three colors: green, yellow, and red, which appear the following probabilities. Using a unigram model, what is the perplexity of the sequence (green, yellow, red)?

- $P(\text{green}) = 2/5$
- $P(\text{yellow}) = 1/5$
- $P(\text{red}) = 2/5$
- $(2/5 \ 2/5 \ 1/5)^{1/3}$
- $(2/5 \ 2/5 \ 1/5) - 1/3$
- $(5/2 \ 5/2 \ 5/1) - 1/3$
- 3

4. If we estimate a bigram language model from the following corpus, what is $P(\text{not}|\text{do})$?

- $\langle s \rangle$ I am Sam $\langle /s \rangle$
- $\langle s \rangle$ Sam I am $\langle /s \rangle$
- $\langle s \rangle$ I do not like green eggs and ham $\langle /s \rangle$
- $2/3$
- $1/3$
- 1
- $1/2$

5. Which is assigned higher probability by a unigram language model for English?

- $P(\text{I like ice cream})$
- $P(\text{the thethe the})$
- $P(\text{Go to class daily})$
- $P(\text{class daily go to})$

6. Compute min edit distance.

7. What is the value of the cell labeled X in the following partially-completed Levenshtein distance table (substitutions cost 2):

- 5
- 4
- 3
- 2

8. A period (".") occurs at the end of four words that each have one of the following four wordshapes. Which of the four periods (all else being equal) is more likely than the other to be a sentence boundary?

- Upper
- Lower
- Cap
- Number

9. The Porter stemmer would stem the word 'aching' as

- aching
- ach
- ache
- aches

10. How many wordform types and tokens are in this sentence:

I want to cook a dish that I never cooked before

- 10 types 11 tokens
- 11 types 10 tokens
- 11 types 11 tokens
- 10 types 10 tokens

11. We saw that the words 'd' and 's' appear frequently in the word tokenization results. Which of following reasons might explain this?

- The words "s" and "d" are archaic words that were more popular during the time that Shakespeare was writing.
- The words "s" and "d" were not very common during the time that Shakespeare was writing, but they are among the many words which Shakespeare appears to have made up himself.
- The tokenizer splits on all non-alphabetical characters, which includes the apostrophe. This means that words like "Ophelia's" and "dimmed" will be tokenized to "Ophelia s" and "dim d".
- The Tokenizer splits words into smaller subparts named morphemes. "s" and "d" are two very common such morphemes because they tend to occur at the end of words.

(): 15
-11- 15 , , ,
- 6-10 , ;
;

5. -

5.1.

1. / . . . , . . . ;
 , 2017. - 109 . : . , . - . : . 104. - ISBN 978-5-7410-1736-4 ; []. -
URL: <http://biblioclub.ru/index.php?page=book&id=481823> (19.11.2018).
2. / - : , 2014. - 448 . - . - ISBN 978-5-392-12318-6 ; []. -
URL: <http://biblioclub.ru/index.php?page=book&id=251602> (19.11.2018).
3. , - : , 2011. - . 2. - 75 . - ISBN 978-5-7782-1873-4 - « » -
URL: <http://biblioclub.ru/index.php?page=book&id=228749>
4. , . . . / - : , 2008. - 106 . - ISBN 978-5-85117-320-2 - « » -
URL: <http://biblioclub.ru/index.php?page=book&id=88672>
5. , . . . / - : , 2012. - 111 . - ISBN 978-5-7638-2234-2 - « » -
URL: <http://biblioclub.ru/index.php?page=book&id=229302> .
6. , . . . = Computer for Work and Leisure:
 / . . . , - : « » -
URL: <http://biblioclub.ru/index.php?page=book&id=277672>

5.2.

7. - « : »:
http://uisrussia.msu.ru/linguist/B7_komp_tehn_v_prepodavanii_jazykov.jsp
8. - - .
http://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/GERMANSKIE_YAZIKI.html
9. . : <http://ru.wikipedia.org/wiki/>
10. - : <http://www.i-u.ru/biblio>
11. : - : <http://www.lingvoinfo.com>
12. - : <http://www.krugosvet.ru>
13. Philology.ru.: <http://philology.ru/linguistics1.html>
14. : <http://www.mccme.ru/ling/index.htm>.

- 15. http://www.nyu.edu/projects/politicsdatalab/localdata/workshops/NLTK_Presentation.pdf
- 16. Windows 8 Russian. Windows Professional 8 Russian Upgrade.
- 17. Microsoft Office Standard 2013 Russian.

6. - ,

(
); (

1	2	3
<p>1. : 10 (, 15 (, 04 ((19, . , 1)</p>		<p>15 - Canon – 1 ; HP Pavilion Slimline S3500F AMD Athlon 64X2 5400+/2.8 GHz, 4Gb, LG 19 L1942S SF 1280*1024, 5ms, 8000:1, black, , ; , , 4 10</p>
<p>2. : 15 (, 10 (, 04 (((. , .</p>		<p>120 Televic Confidea; Televic ID 2500D – 5 ; Televic L-DD – 1 ; 8 Televic AOP2500;. 8- Televic R8; 8- Televic 8; Cooler Master, Intel Core i5/2*2 / GigaByte/ Seagate/ Zotac / Optiarc AD, Genius, Genius 130 , Viewsonic VG2239M 2. () : NEC M402W (M402WG). Full3D.DLP.4000 ANSI</p>

19, . , 1)		Lm.WXGA; Lumien Master Control 129x200 (88"); SmartBoard 680, 77"/105,6
3. : () (. 13 , . 19, , 1)		<p>(10 .)</p> <p>() .</p> <p>04</p> <p>, , - 12 .</p> <p>, -</p> <p>()</p> <p>13</p> <p>, - ,</p> <p>, - -1 , .</p> <p>-2 .</p> <p>-</p> <p>() ,</p> <p>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. 104 17.06.2013 .</p> <p>2. Microsoft Office Standard 2013 Russian. 114 12.11.2014 .</p> <p>3. Adobe Audition CS6 5 AcademicEdition License International English MultiplePlatforms 228 06.11.2012</p>

« - »

« 4 »

(/)	3/108
:	33,2
/	32
(,) ()	1,2
()	11,8
/ / ()	63

: :4

/		: , , , ()				(,)		(, , .)
1	2	3	4 /	5	6	7	8	9
1.	1: , . 1: , . 2: , , . , , . 3 . .		16		10	: 1,2 : 1,3	,) 1,2,3) 1,2,6	, ,
2.	2: 4: . 5 . .		16		1,8	: 1,2 : 1, 3	,) 1,2,3) 1,2,5,6	, ,

	6							
	7							
	8.							
	:		32		11,8			

-

1				
1. ()	5	2	0	10
2.	5	2	0	10
1.	15		0	15
1			0	35
2				
1. ()	5	2	0	10
2.	5	2	0	10
1.	15		0	15
2			0	35
, , , ...			0	10
()				
				-6
				-10
				30
				110