Quick Guide

Loading and preprocessing of data. 1

Preprocessing results. 7

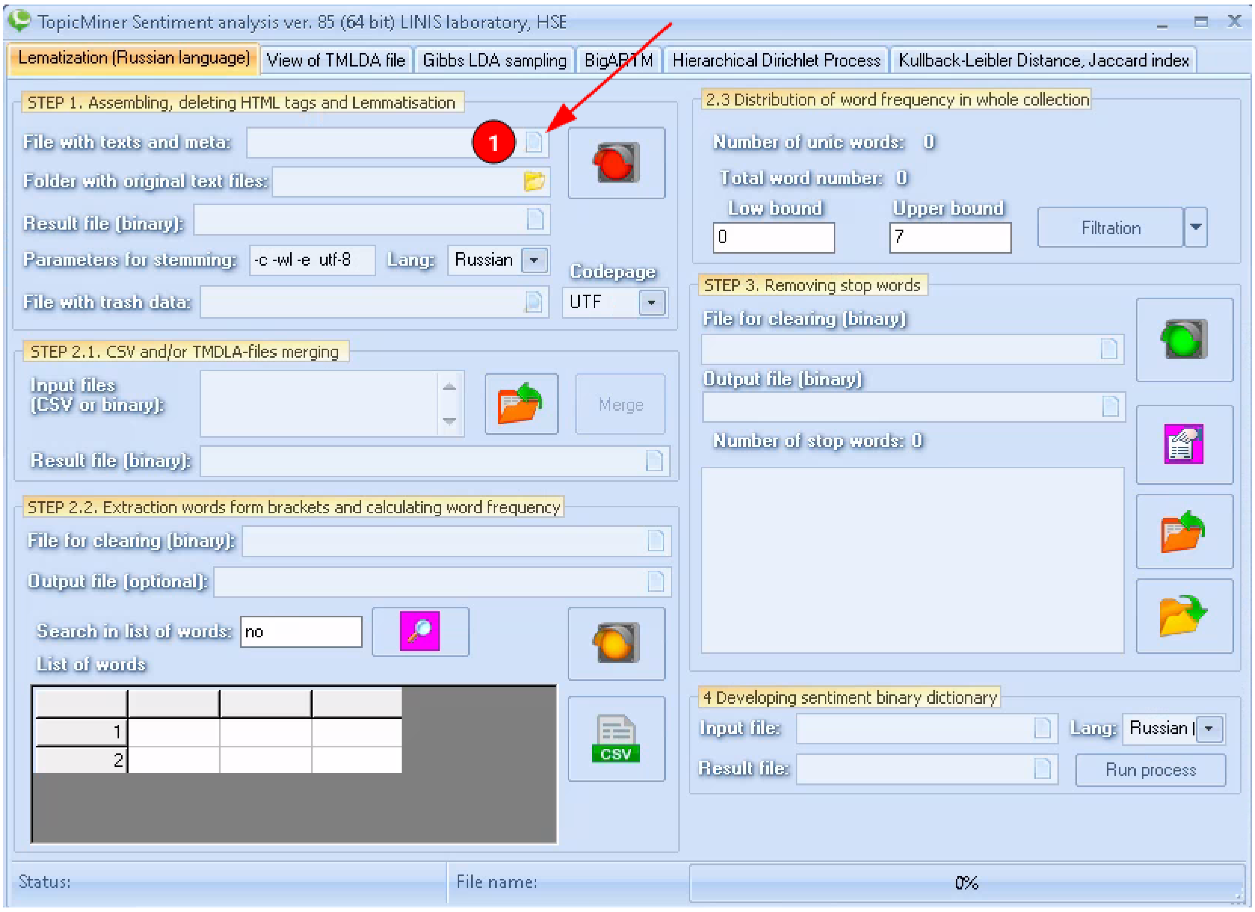
Topic modeling. 10

Viewing the results obtained and uploading to an external file. 12

Saving calculation results as a project: 13

1. Loading and preprocessing of data.

Run **TopicMiner.exe** file (in TopicMiner directory). Press the button with a file icon to load textual data.



TopicMiner Sentiment analysis ver. 85 (64 bit) LINIS laboratory, HSE

Parameters for stemming: -c -wl -e utf-8

Lang: Russian

Codepage: UTF

Number of unic words: 0

Total word number: 0

Low bound: 0

Upper bound: 7

Number of stop words: 0

Search in list of words: no

Lang: Russian

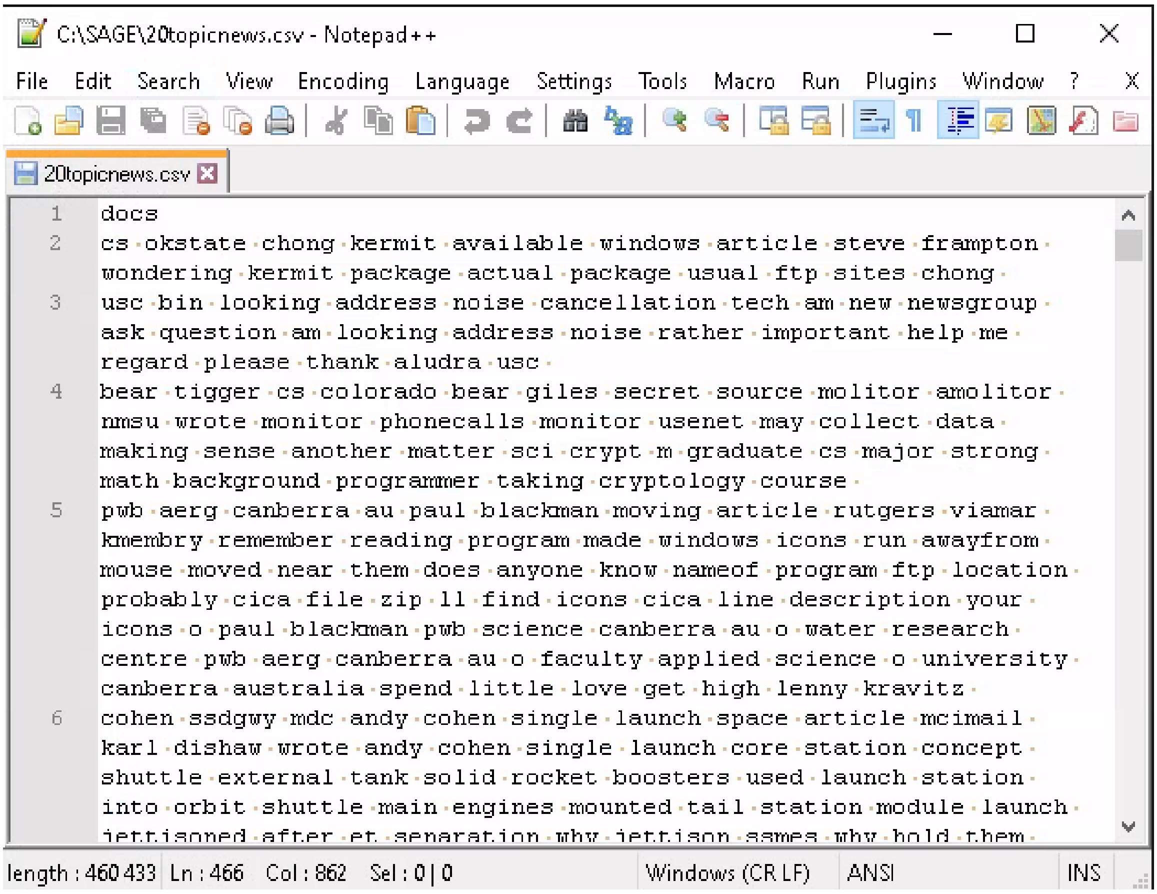
Status:

File name:

0%

Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
STEP 1. Assembling, deleting HTML tags and Lemmatisation
File with texts and meta:
1
Folder with original text files:
Result file (binary):
File with trash data:
STEP 2.1. CSV and/or TMDLA-files merging
Input files (CSV or binary):
Merge
Result file (binary):
STEP 2.2. Extraction words form brackets and calculating word frequency
File for clearing (binary):
Output file (optional):
List of words
1
2
CSV
2.3 Distribution of word frequency in whole collection
Filtration
STEP 3. Removing stop words
File for clearing (binary)
Output file (binary)
4 Developing sentiment binary dictionary
Input file:
Result file:
Run process

Next, you will need to select a file with textual data, in which one line is one document.

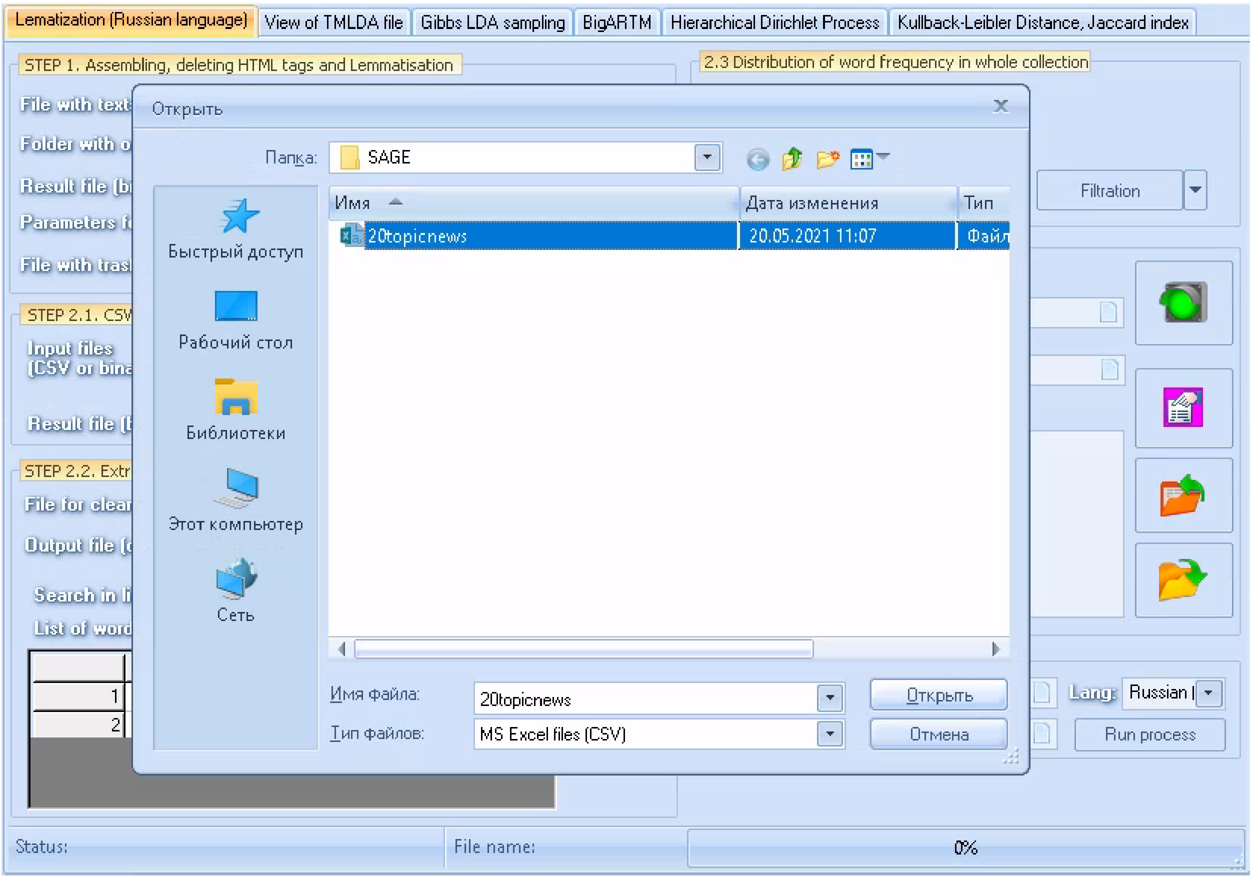


C:\SAGE\20topicnews.csv - Notepad++

```
docs
cs okstate chong kermit available windows article steve frampton
wondering kermit package actual package usual ftp sites chong
usc bin looking address noise cancellation tech am new newsgroup
ask question am looking address noise rather important help me
regard please thank aludra usc
bear tigger cs colorado bear giles secret source molitor amolitor
nmsu wrote monitor phonecalls monitor usenet may collect data
making sense another matter sci crypt m graduate cs major strong
math background programmer taking cryptology course
pwb aerg canberra au paul blackman moving article rutgers viamar
kmembry remember reading program made windows icons run awayfrom
mouse moved near them does anyone know nameof program ftp location
probably cica file zip ll find icons cica line description your
icons o paul blackman pwb science canberra au o water research
centre pwb aerg canberra au o faculty applied science o university
canberra australia spend little love get high lenny kravitz
cohen ssdgwy mdc andy cohen single launch space article mcimail
karl dishaw wrote andy cohen single launch core station concept
shuttle external tank solid rocket boosters used launch station
into orbit shuttle main engines mounted tail station module launch
jettisoned after et separation why jettison ssmes why hold them
```

length : 460 433 Ln : 466 Col : 862 Sel : 0 | 0 Windows (CR LF) ANSI INS

For example, open a CSV file (see picture).


Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
STEP 1. Assembling, deleting HTML tags and Lemmatisation
2.3 Distribution of word frequency in whole collection
Filtration
Открыть
Быстрый доступ
Рабочий стол
Библиотеки
Этот компьютер
Сеть
Открыть
Отмена
Run process
Status:
File name:
0%

Папка: SAGE

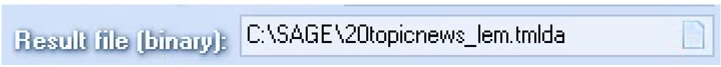
Имя	Дата изменения	Тип
20topicnews	20.05.2021 11:07	Файл

Имя файла: 20topicnews

Тип файлов: MS Excel files (CSV)

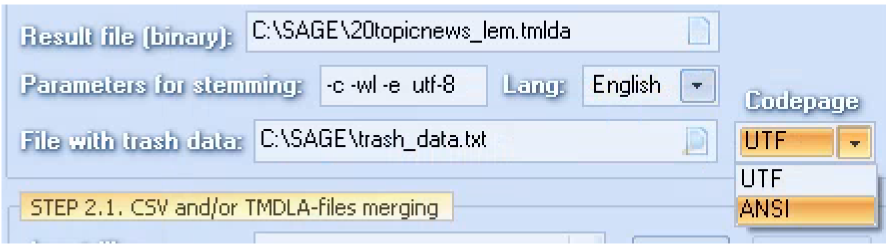
Lang: Russian

Next, you need to specify the file to store the lemmatization results.



Result file (binary): C:\SAGE\20topicnews_lem.tmlda

For example, **C:\SAGE\20topicnews_lem.tmlda**. Next, you need to specify text encoding (in this version, two options are implemented). Select **ANSI**.



Result file (binary): C:\SAGE\20topicnews_lem.tmlda

Parameters for stemming: -c -wl -e utf-8

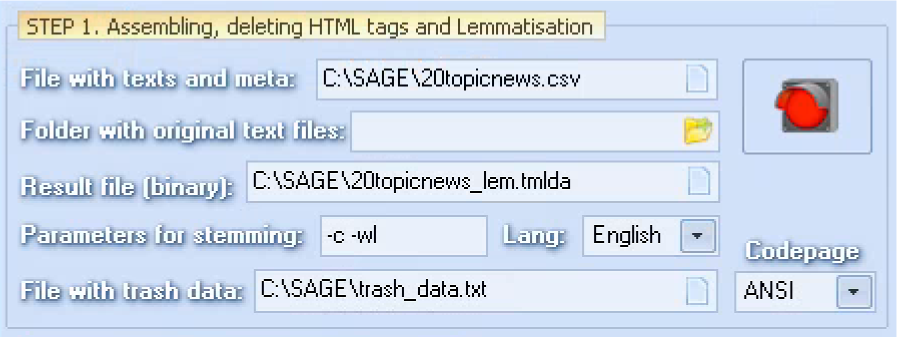
Lang: English

File with trash data: C:\SAGE\trash_data.txt

Codepage: UTF

You also need to specify the language (Russian and English are implemented in this version). Select **English**.

Lastly, you need to specify a file that contains some trash words/entities, which can also be deleted. See **trash_data.txt** file for an example of such a file.



STEP 1. Assembling, deleting HTML tags and Lemmatisation

File with texts and meta: C:\SAGE\20topicnews.csv

Folder with original text files:

Result file (binary): C:\SAGE\20topicnews_lem.tmlda

Parameters for stemming: -c -wl

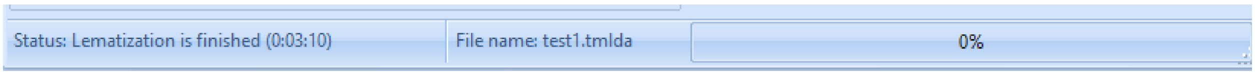
Lang: English

Codepage: ANSI

File with trash data: C:\SAGE\trash_data.txt

As a result, when everything you need is filled in, click on the red button.

You will start the process of lemmatization (see the percentage of completion at the bottom of the screen). As a result, you will have a file containing the original and lemmatized texts. When the first stage of preprocessing (lemmatization procedure) is over, you will see the following message in the status bar.

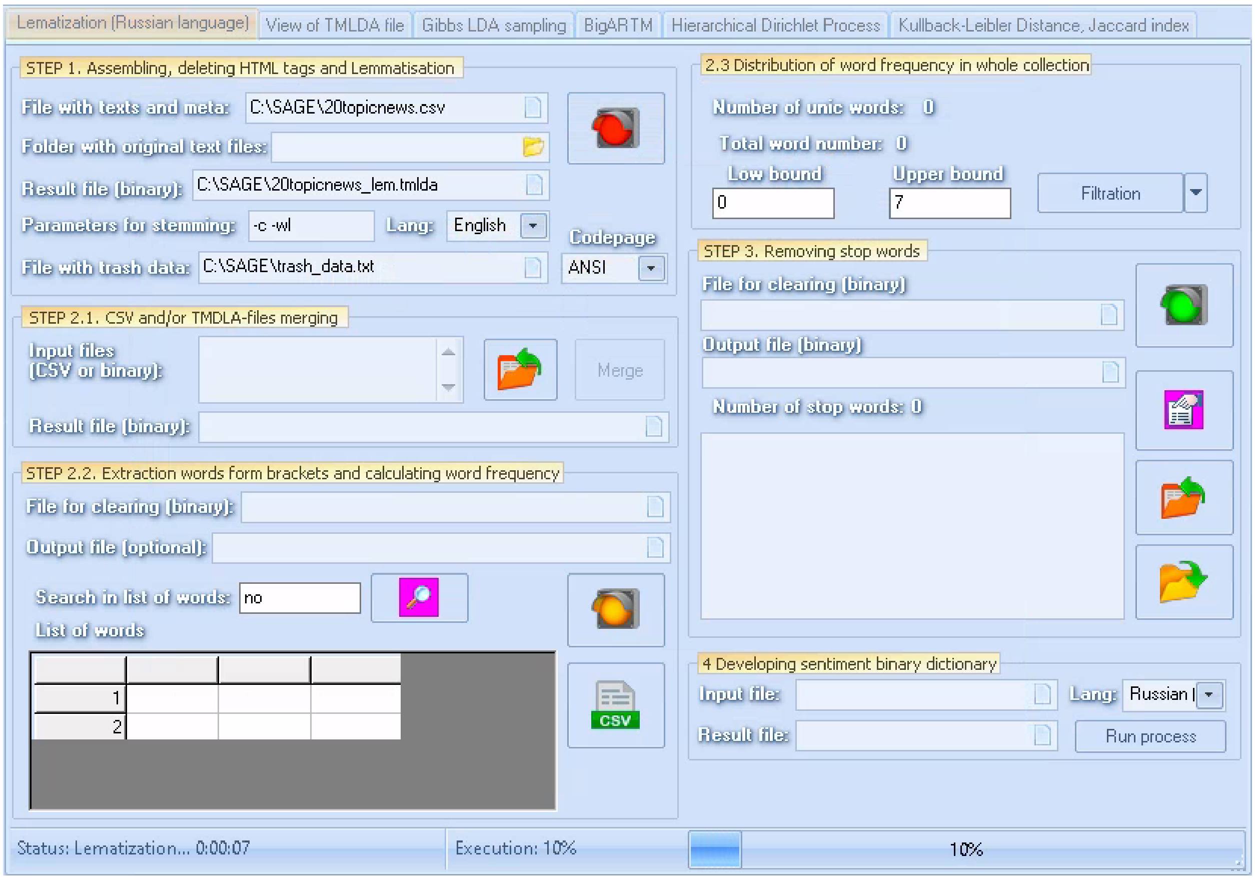


Status: Lematization is finished (0:03:10)

File name: test1.tmlda

0%

The time of lemmatization is also indicated there.


Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
Merge
Filtration
Run process
CSV

STEP 1. Assembling, deleting HTML tags and Lemmatisation

File with texts and meta: C:\SAGE\20topicnews.csv

Folder with original text files:

Result file (binary): C:\SAGE\20topicnews_lem.tmlda

Parameters for stemming: -c -wl

Lang: English

Codepage: ANSI

File with trash data: C:\SAGE\trash_data.txt

STEP 2.1. CSV and/or TMDLA-files merging

Input files (CSV or binary):

Result file (binary):

STEP 2.2. Extraction words form brackets and calculating word frequency

File for clearing (binary):

Output file (optional):

Search in list of words: no

List of words

1			
2			

2.3 Distribution of word frequency in whole collection

Number of unic words: 0

Total word number: 0

Low bound: 0

Upper bound: 7

STEP 3. Removing stop words

File for clearing (binary)

Output file (binary)

Number of stop words: 0

4 Developing sentiment binary dictionary

Input file:

Lang: Russian l

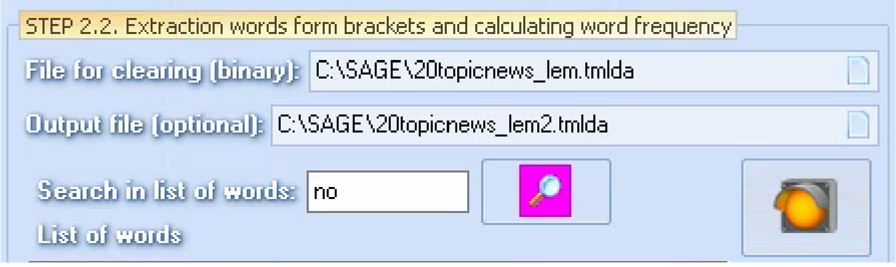
Result file:

Status: Lematization... 0:00:07

Execution: 10%

10%

At the second stage, you need to specify **C:\SAGE\20topicnews_lem.tmlda** as the input file, and **C:\SAGE\20topicnews_lem2.tmlda** as the output (it is optional). Then press the yellow button.



STEP 2.2. Extraction words form brackets and calculating word frequency

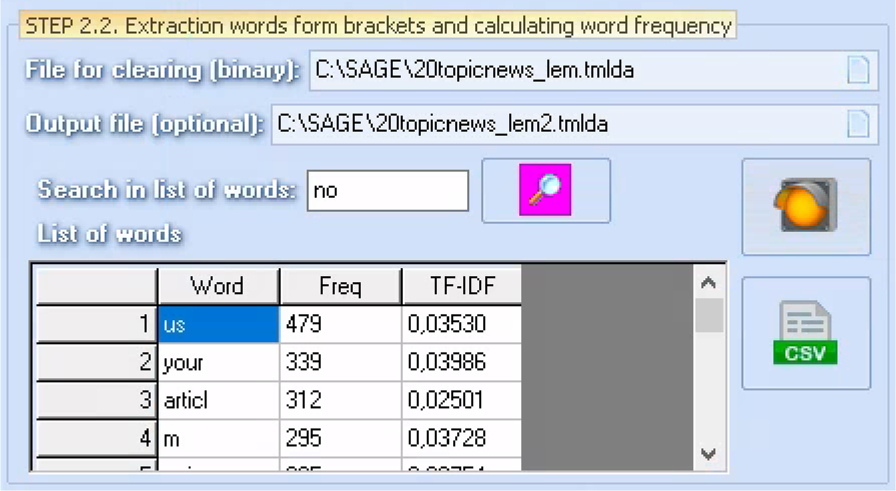
File for clearing (binary): C:\SAGE\20topicnews_lem.tmlda

Output file (optional): C:\SAGE\20topicnews_lem2.tmlda

Search in list of words: no

List of words

The process of the second stage (in percentage) is also visualized. A word list with frequencies will also appear. This list can be saved by pressing “CSV” button. This option is useful to form a list of stop words. For example, you can select the most frequent words as a list of stop words and put a filter that will save these words to a test file.



STEP 2.2. Extraction words form brackets and calculating word frequency

File for clearing (binary): C:\SAGE\20topicnews_lem.tmlda

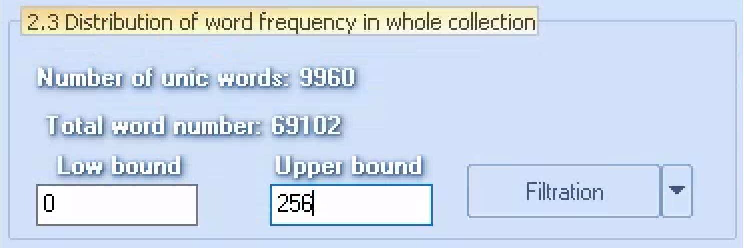
Output file (optional): C:\SAGE\20topicnews_lem2.tmlda

Search in list of words: no

List of words

	Word	Freq	TF-IDF
1	us	479	0,03530
2	your	339	0,03986
3	articl	312	0,02501
4	m	295	0,03728

For example, in this case, let's take **256** as the upper bound.



2.3 Distribution of word frequency in whole collection

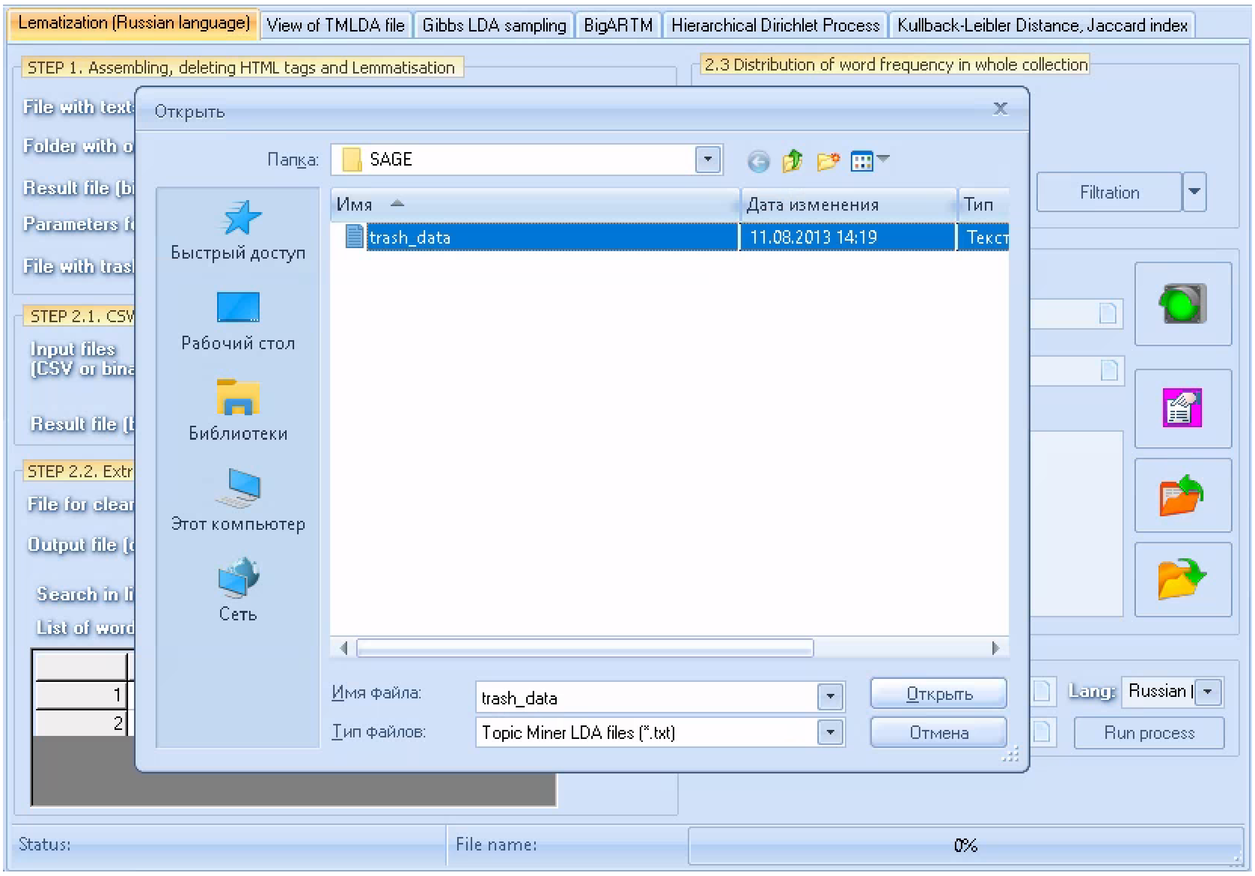
Number of unic words: 9960

Total word number: 69102

Low bound: 0

Upper bound: 256

Next, click on the “filtration” button and specify the file in which the words that lie outside the specified range will be saved (that is, everything that is above the frequency of **1258**). If you open such a file, you will see a list of stop words (most frequent).


Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
STEP 1. Assembling, deleting HTML tags and Lemmatisation
2.3 Distribution of word frequency in whole collection
Filtration
STEP 2.1. CSV
STEP 2.2. Extr
Открыть
Быстрый доступ
Рабочий стол
Библиотеки
Этот компьютер
Сеть
Открыть
Отмена
Run process

Папка: SAGE

Имя	Дата изменения	Тип
trash_data	11.08.2013 14:19	Текст

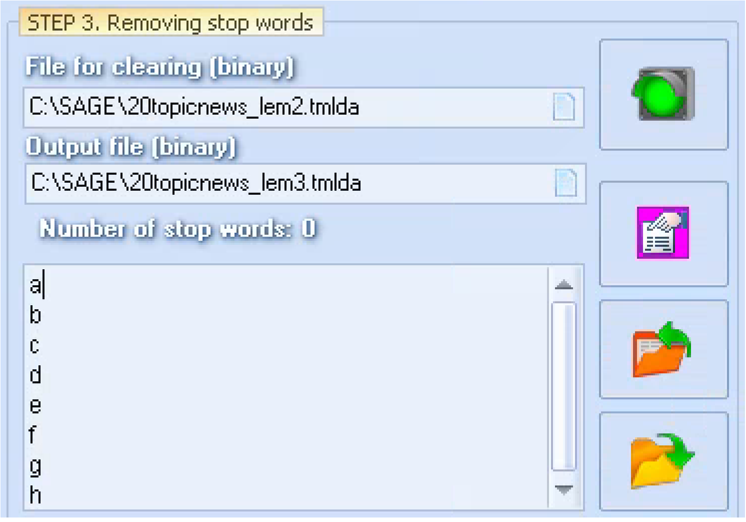
Имя файла: trash_data

Тип файлов: Topic Miner LDA files (*.txt)

Lang: Russian

Status: File name: 0%

At the last stage of preprocessing. You need to point to the input file **C:\SAGE\20topicnews_lem2.tmlda**, and the output to **C:\SAGE\20topicnews_lem3.tmlda**. And also download a file with a list of stop words.


STEP 3. Removing stop words

File for clearing (binary): C:\SAGE\20topicnews_lem2.tmlda

Output file (binary): C:\SAGE\20topicnews_lem3.tmlda

Number of stop words: 0

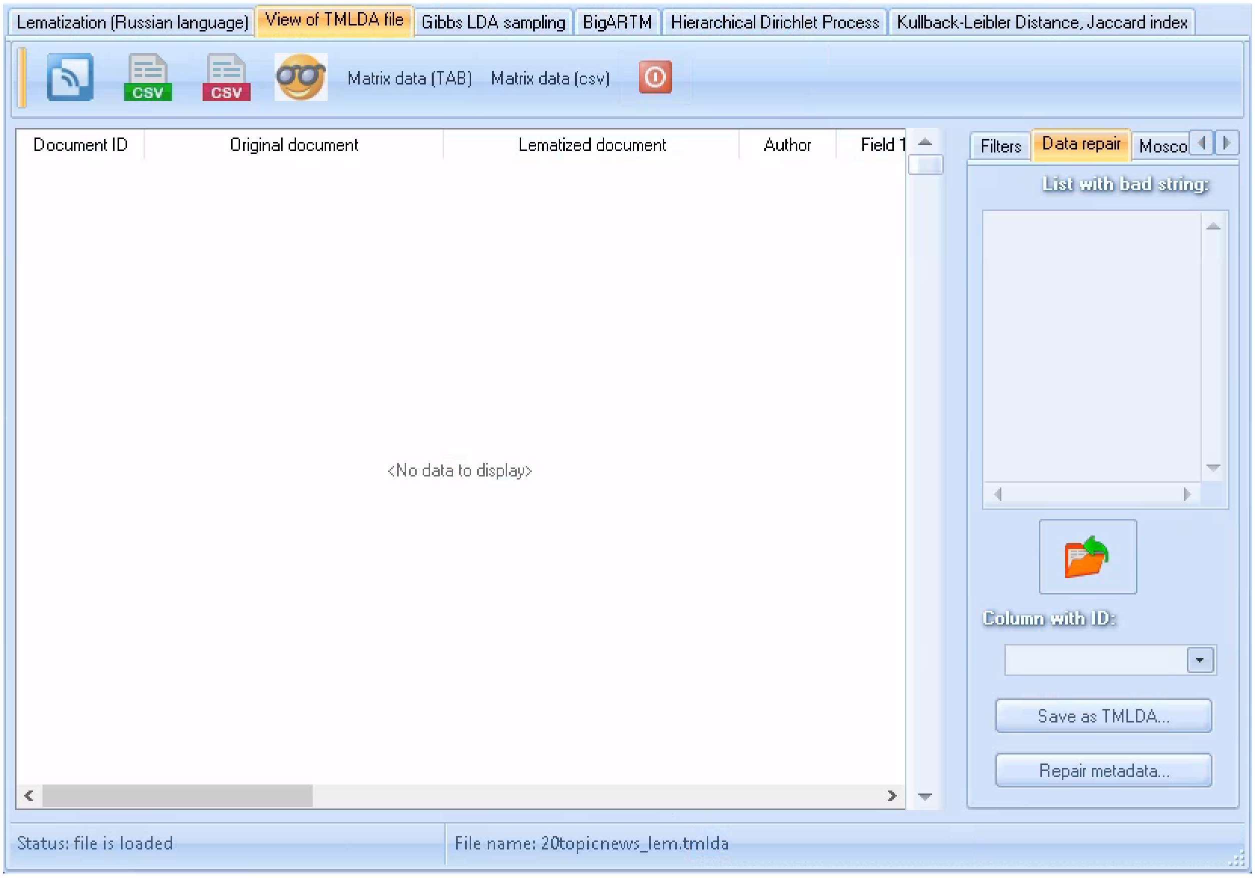
```
a
b
c
d
e
f
g
h
```

Now press the green button.

2. Preprocessing results.

If you need to see what you have learned (and also to start filtering texts), then you need to go to the next tab. To observe the results, use the following button:

Specify **C:\SAGE\20topicnews_lem3.tmlda** file.

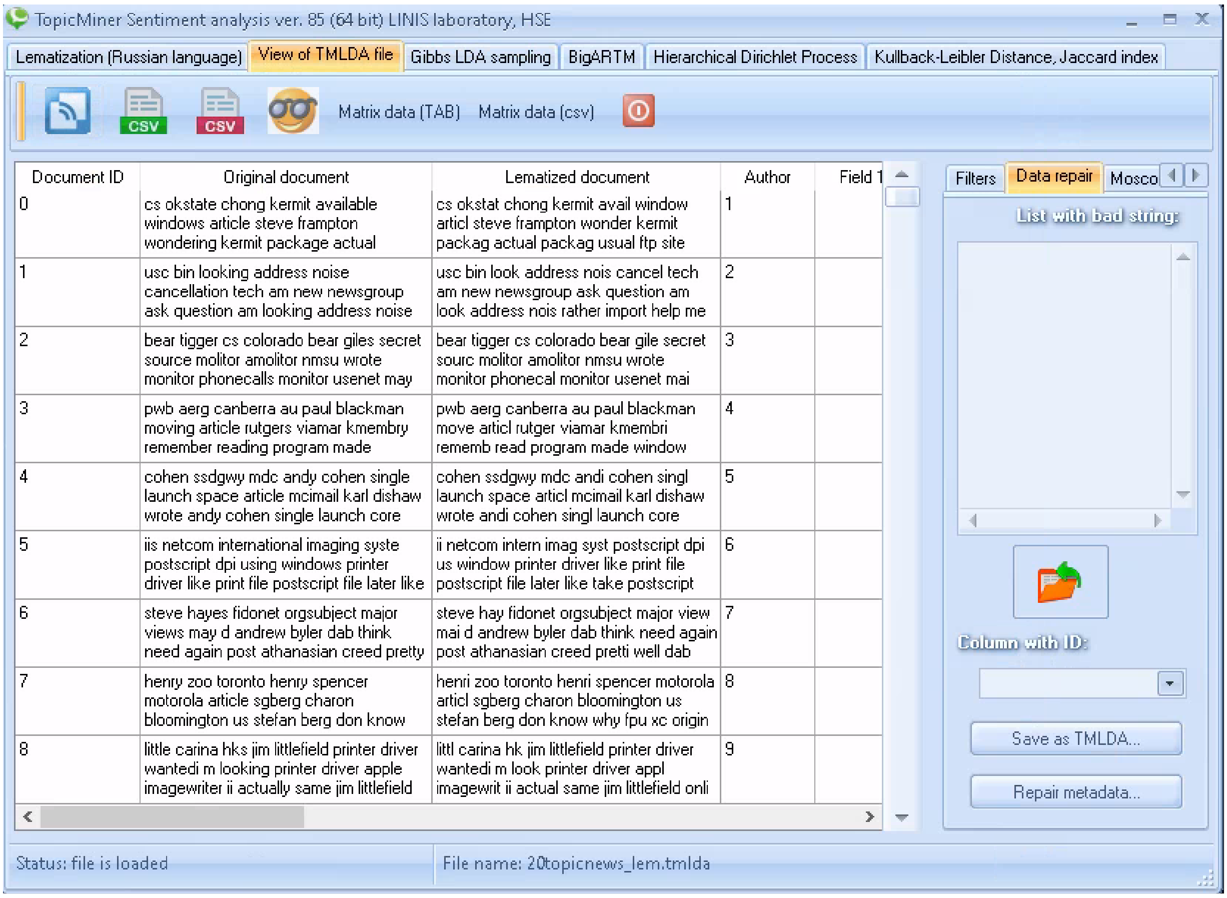


<No data to display>

Status: file is loaded

File name: 20topicnews_lem.tmlda

Table will be displayed:



Document ID	Original document	Lematized document	Author	Field 1
0	cs okstate chong kermit available windows article steve frampton wondering kermit package actual	cs okstat chong kermit avail window articl steve frampton wonder kermit packag actual packag usual ftp site	1	
1	usc bin looking address noise cancellation tech am new newsgroup ask question am looking address noise	usc bin look address nois cancel tech am new newsgroup ask question am look address nois rather import help me	2	
2	bear tigger cs colorado bear giles secret source molitor amolitor nmsu wrote monitor phonecalls monitor usenet may	bear tigger cs colorado bear gile secret sourc molitor amolitor nmsu wrote monitor phonecal monitor usenet mai	3	
3	pwb aerg canberra au paul blackman moving article rutgers viamar kmembry remember reading program made	pwb aerg canberra au paul blackman move articl rutger viamar kmembri rememb read program made window	4	
4	cohen ssdgwy mdc andy cohen single launch space article mcimail karl dishaw wrote andy cohen single launch core	cohen ssdgwy mdc andi cohen singl launch space articl mcimail karl dishaw wrote andi cohen singl launch core	5	
5	iis netcom international imaging syste postscript dpi using windows printer driver like print file postscript file later like	ii netcom intern imag syst postscript dpi us window printer driver like print file postscript file later like take postscript	6	
6	steve hayes fidonet orgsubject major views may d andrew byler dab think need again post athanasian creed pretty	steve hay fidonet orgsubject major view mai d andrew byler dab think need again post athanasian creed pretti well dab	7	
7	henry zoo toronto henry spencer motorola article sgberg charon bloomington us stefan berg don know	henri zoo toronto henri spencer motorola articl sgberg charon bloomington us stefan berg don know why fpu xc origin	8	
8	little carina hks jim littlefield printer driver wantedi m looking printer driver apple imagewriter ii actually same jim littlefield	littl carina hk jim littlefield printer driver wantedi m look printer driver appl imagewrit ii actual same jim littlefield onli	9	

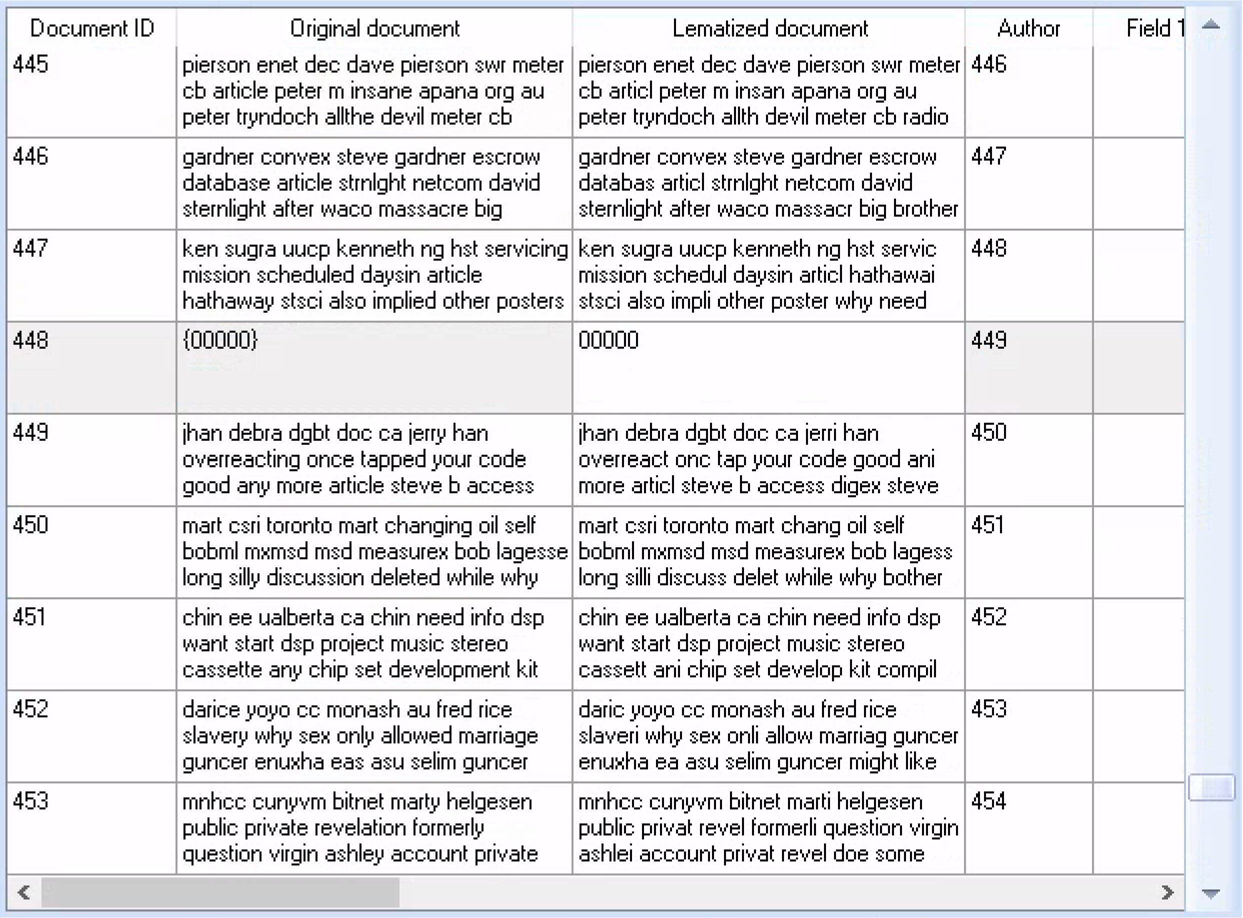
Status: file is loaded

File name: 20topicnews_lem.tmlda

The first column is the original data, the second column is the lemmatized data. The third and subsequent columns are metadata.

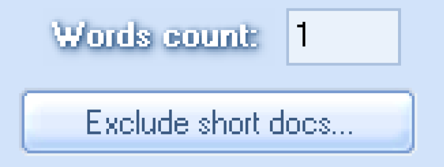
As a result of preprocessing, empty documents can be obtained (for example, due to unspecified line ends or words in the texts were deleted using stop words). Of course, such documents must be removed from the collection. This can be done using filtering.

For example, let's delete short documents that contain only one word (**00000**):



Document ID	Original document	Lematized document	Author	Field 1
445	pierson enet dec dave pierson swr meter cb article peter m insane apana org au peter tryndoch allthe devil meter cb	pierson enet dec dave pierson swr meter cb articl peter m insan apana org au peter tryndoch allth devil meter cb radio	446	
446	gardner convex steve gardner escrow database article strnlght netcom david sternlight after waco massacre big	gardner convex steve gardner escrow databas articl strnlght netcom david sternlight after waco massacr big brother	447	
447	ken sugra uucp kenneth ng hst servicing mission scheduled daysin article hathaway stsci also implied other posters	ken sugra uucp kenneth ng hst servic mission schedul daysin articl hathawai stsci also impli other poster why need	448	
448	{00000}	00000	449	
449	jhan debra dgbt doc ca jerry han overreacting once tapped your code good any more article steve b access	jhan debra dgbt doc ca jerri han overreact onc tap your code good ani more articl steve b access digex steve	450	
450	mart csri toronto mart changing oil self bobml mxmsd msd measurex bob lagesse long silly discussion deleted while why	mart csri toronto mart chang oil self bobml mxmsd msd measurex bob lagess long silli discuss delet while why bother	451	
451	chin ee ualberta ca chin need info dsp want start dsp project music stereo cassette any chip set development kit	chin ee ualberta ca chin need info dsp want start dsp project music stereo cassett ani chip set develop kit compil	452	
452	darice yoyo cc monash au fred rice slavery why sex only allowed marriage guncer enuxha eas asu selim guncer	daric yoyo cc monash au fred rice slaveri why sex onli allow marriag guncer enuxha ea asu selim guncer might like	453	
453	mnhcc cunyvm bitnet marty helgesen public private revelation formerly question virgin ashley account private	mnhcc cunyvm bitnet marti helgesen public privat revel formerli question virgin ashlei account privat revel doe some	454	

To do this, we will indicate the number of words:

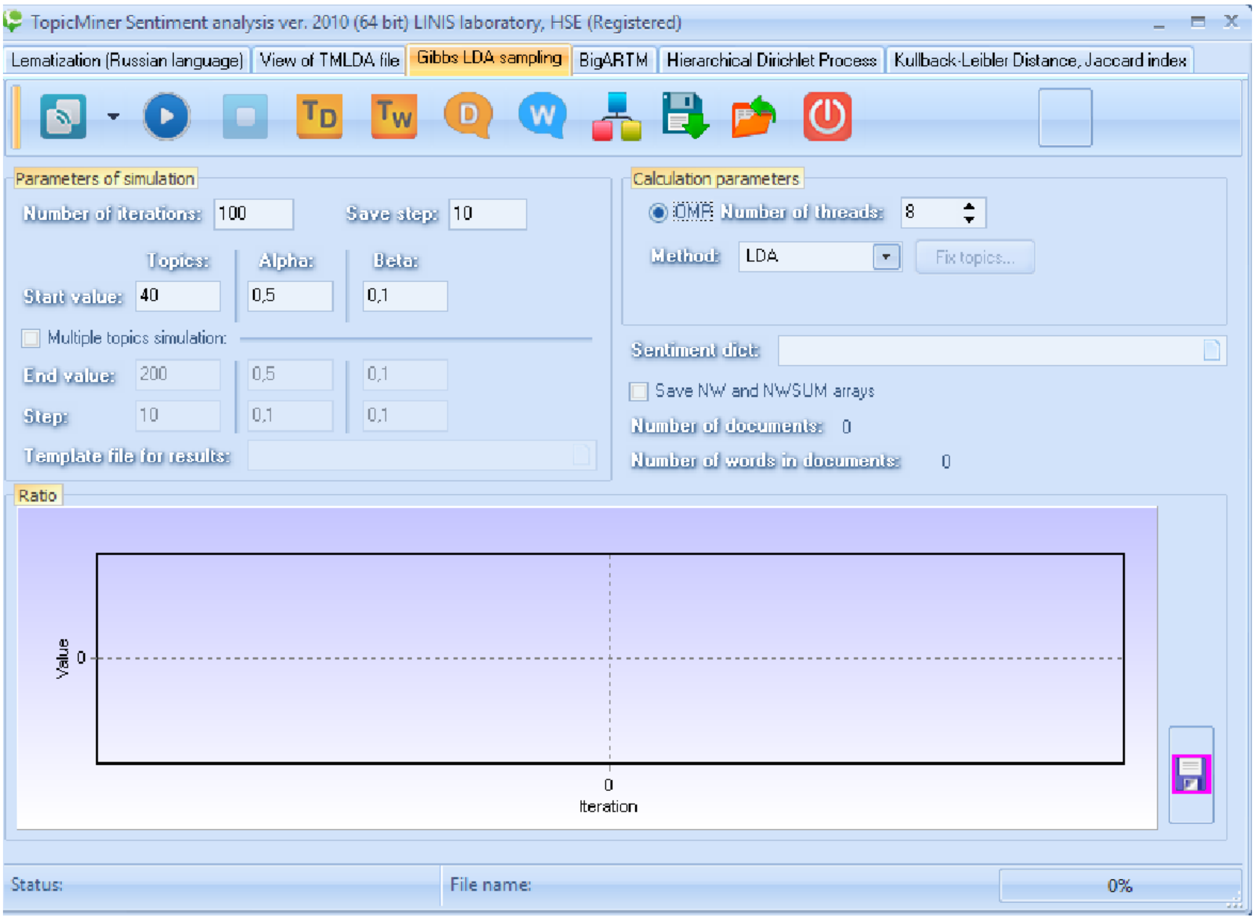


Words count: 1

And click on **Exclude short docs** button. As a result, **20topicnews_lem3_we.tmlda** file will be created, which will no longer contain these documents.

3. Topic modeling.

In order to start thematic modeling, for example, based on Gibbs sampling (Monte Carlo method), you need to go to another tab:


TopicMiner Sentiment analysis ver. 2010 (64 bit) LINIS laboratory, HSE (Registered)
Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
Ratio
Value
0
0
Iteration

Parameters of simulation
Number of iterations: 100 Save step: 10

	Topics:	Alpha:	Beta:
Start value:	40	0,5	0,1

□ Multiple topics simulation:

End value:	200	0,5	0,1
Step:	10	0,1	0,1

Template file for results:

Calculation parameters
● OMP Number of threads: 8
Method: LDA
Fix topics...

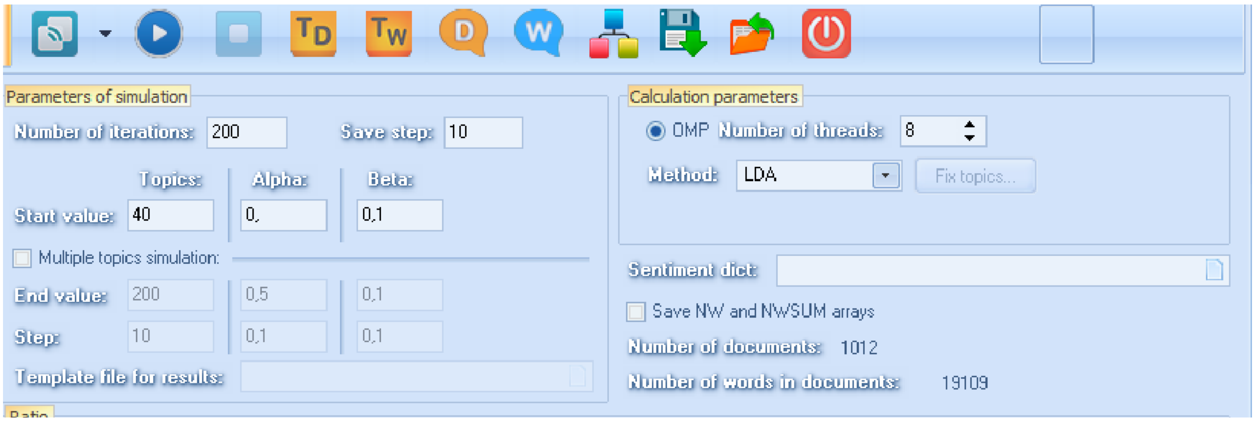
Sentiment dict:
□ Save NW and NWSUM arrays
Number of documents: 0
Number of words in documents: 0

Status: File name: 0%

You need to load the file after preprocessing and removing empty documents.

For example, load the file **C:\SAGE\20topicnews_lem3_we.tmlda**. Ignore the download percentage (in the current version, this is just a measure of the length of the collection).

Next, specify the parameters of the model (number of topics, number of iterations, the rendering step, and *alpha*, *beta* parameters). This can be done as shown in the picture.



Parameters of simulation
Number of iterations: 200 Save step: 10

	Topics:	Alpha:	Beta:
Start value:	40	0,	0,1

□ Multiple topics simulation:

End value:	200	0,5	0,1
Step:	10	0,1	0,1

Template file for results:

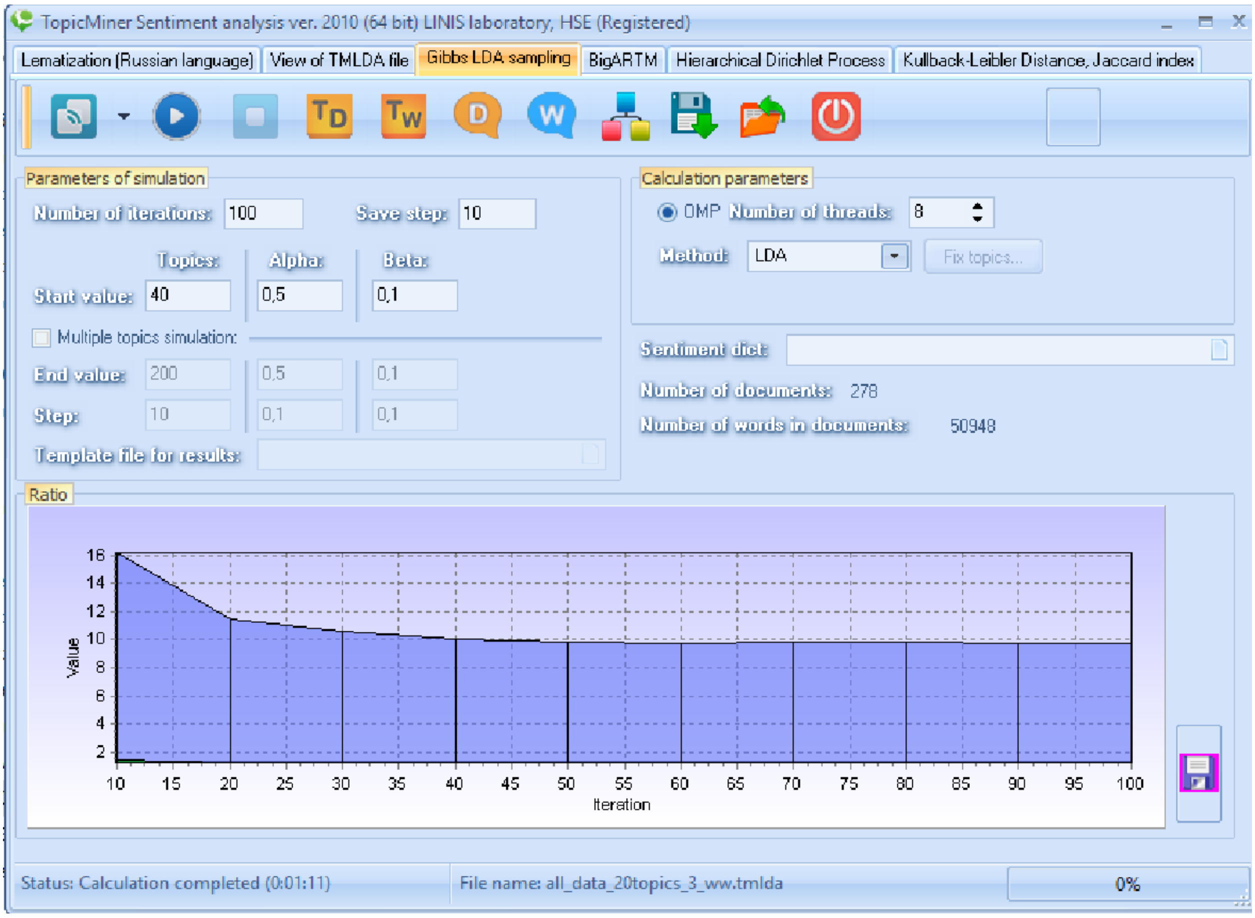
Calculation parameters
● OMP Number of threads: 8
Method: LDA
Fix topics...

Sentiment dict:
□ Save NW and NWSUM arrays
Number of documents: 1012
Number of words in documents: 19109

And you need to choose a model, for example, **LDA**.

Then click on the button:

The modeling progress will be displayed on the graph:



TopicMiner Sentiment analysis ver. 2010 (64 bit) LINIS laboratory, HSE (Registered)
Lematization (Russian language)
View of TMLDA file
Gibbs LDA sampling
BigARTM
Hierarchical Dirichlet Process
Kullback-Leibler Distance, Jaccard index
TD
TW
D
W
Ratio
16
14
12
10
8
6
4
2
Value
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Iteration

Parameters of simulation

Number of iterations: 100

Save step: 10

	Topics:	Alpha:	Beta:
Start value:	40	0,5	0,1
End value:	200	0,5	0,1
Step:	10	0,1	0,1

□ Multiple topics simulation:

Template file for results:

Calculation parameters

● OMP Number of threads: 8

Method: LDA

Fix topics...

Sentiment dict:

Number of documents: 278

Number of words in documents: 50948

Status: Calculation completed (0:01:11)

File name: all_data_20topics_3_ww.tmlda

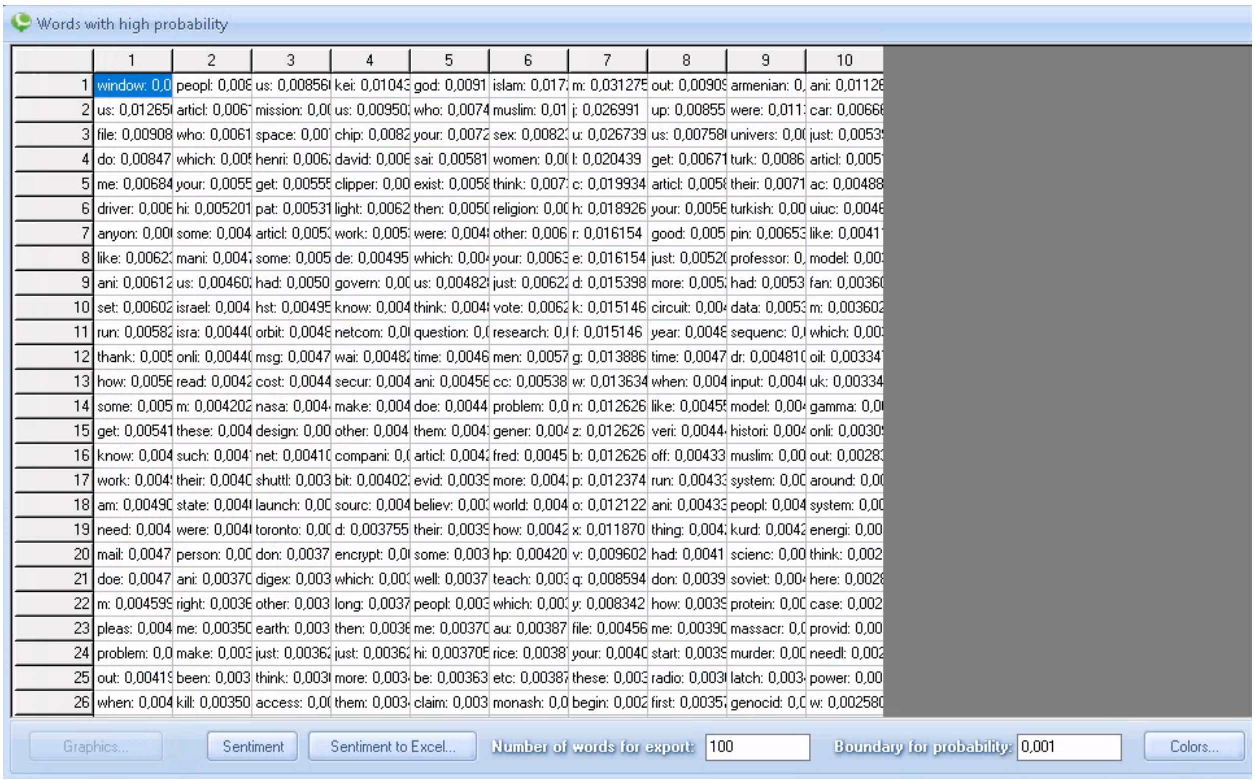
0%

The blue line is the percentage of words with high probability. It can be seen from the graph that after 100 iterations, the percentage of words stops changing, which means a further increase in iterations does not need to be done. The green line is the percentage of high probability documents.

4. Viewing the results obtained and uploading to an external file.

To see the already sorted calculation results, you need to click on the button:

As a result, you will see the following picture:



Words with high probability

	1	2	3	4	5	6	7	8	9	10
1	window: 0,0	peopl: 0,008	us: 0,00856	kei: 0,01043	god: 0,0091	islam: 0,017	m: 0,031275	out: 0,00909	armenian: 0,	ani: 0,01126
2	us: 0,01265	articl: 0,0061	mission: 0,0	us: 0,00950	who: 0,0074	muslim: 0,01	j: 0,026991	up: 0,00855	were: 0,011	car: 0,00666
3	file: 0,00908	who: 0,0061	space: 0,00	chip: 0,0082	your: 0,0072	sex: 0,0082	u: 0,026739	us: 0,00758	univers: 0,0	just: 0,0053
4	do: 0,00847	which: 0,005	henri: 0,006	david: 0,006	sai: 0,00581	women: 0,0	l: 0,020439	get: 0,00671	turk: 0,0086	articl: 0,005
5	me: 0,00684	your: 0,0055	get: 0,00555	clipper: 0,00	exist: 0,0058	think: 0,007	c: 0,019934	articl: 0,005	their: 0,0071	ac: 0,00488
6	driver: 0,006	hi: 0,005201	pat: 0,00531	light: 0,0062	then: 0,005	religion: 0,0	h: 0,018926	your: 0,0056	turkish: 0,00	uiuc: 0,0046
7	anyon: 0,00	some: 0,004	articl: 0,005	work: 0,005	were: 0,004	other: 0,006	r: 0,016154	good: 0,005	pin: 0,0065	like: 0,0041
8	like: 0,0062	mani: 0,004	some: 0,005	de: 0,00495	which: 0,00	your: 0,006	e: 0,016154	just: 0,0052	professor: 0,	model: 0,00
9	ani: 0,00612	us: 0,00460	had: 0,0050	govern: 0,0	us: 0,00482	just: 0,0062	d: 0,015398	more: 0,005	had: 0,0053	fan: 0,0036
10	set: 0,00602	israel: 0,004	hst: 0,00495	know: 0,004	think: 0,004	vote: 0,0062	k: 0,015146	circuit: 0,00	data: 0,005	m: 0,003602
11	run: 0,00582	isra: 0,0044	orbit: 0,004	netcom: 0,0	question: 0,	research: 0,	f: 0,015146	year: 0,004	sequenc: 0,	which: 0,00
12	thank: 0,005	onli: 0,0044	msg: 0,0047	wai: 0,0048	time: 0,0046	men: 0,0057	g: 0,013886	time: 0,0047	dr: 0,00481	oil: 0,00334
13	how: 0,005	read: 0,004	cost: 0,0044	secur: 0,004	ani: 0,0045	cc: 0,00538	w: 0,013634	when: 0,004	input: 0,004	uk: 0,00334
14	some: 0,005	m: 0,004202	nasa: 0,004	make: 0,004	doe: 0,0044	problem: 0,0	n: 0,012626	like: 0,0045	model: 0,00	gamma: 0,0
15	get: 0,00541	these: 0,004	design: 0,00	other: 0,004	them: 0,004	gener: 0,004	z: 0,012626	veri: 0,0044	histori: 0,004	onli: 0,0030
16	know: 0,004	such: 0,004	net: 0,0041	compani: 0,	articl: 0,004	fred: 0,0045	b: 0,012626	off: 0,00433	muslim: 0,00	out: 0,0028
17	work: 0,004	their: 0,004	shuttl: 0,003	bit: 0,00402	evid: 0,003	more: 0,004	p: 0,012374	run: 0,0043	system: 0,0	around: 0,0
18	am: 0,0049	state: 0,004	launch: 0,0	sourc: 0,004	believ: 0,00	world: 0,004	o: 0,012122	ani: 0,0043	peopl: 0,004	system: 0,0
19	need: 0,004	were: 0,004	toronto: 0,0	d: 0,003755	their: 0,003	how: 0,0042	x: 0,011870	thing: 0,004	kurd: 0,0042	energi: 0,00
20	mail: 0,0047	person: 0,0	don: 0,0037	encrypt: 0,0	some: 0,003	hp: 0,00420	v: 0,009602	had: 0,0041	scienc: 0,00	think: 0,002
21	doe: 0,0047	ani: 0,0037	digex: 0,003	which: 0,00	well: 0,0037	teach: 0,00	q: 0,008594	don: 0,0039	soviet: 0,00	here: 0,002
22	m: 0,00459	right: 0,003	other: 0,003	long: 0,0037	peopl: 0,00	which: 0,00	y: 0,008342	how: 0,003	protein: 0,0	case: 0,002
23	pleas: 0,004	me: 0,0035	earth: 0,003	then: 0,003	me: 0,0037	au: 0,00387	file: 0,00456	me: 0,0039	massacr: 0,	provid: 0,00
24	problem: 0,0	make: 0,003	just: 0,0036	just: 0,0036	hi: 0,003705	rice: 0,0038	your: 0,004	start: 0,003	murder: 0,0	needl: 0,002
25	out: 0,0041	been: 0,003	think: 0,003	more: 0,003	be: 0,00363	etc: 0,0038	these: 0,003	radio: 0,003	latch: 0,003	power: 0,00
26	when: 0,004	kill: 0,00350	access: 0,0	them: 0,003	claim: 0,003	monash: 0,0	begin: 0,002	first: 0,0035	genocid: 0,	w: 0,002580

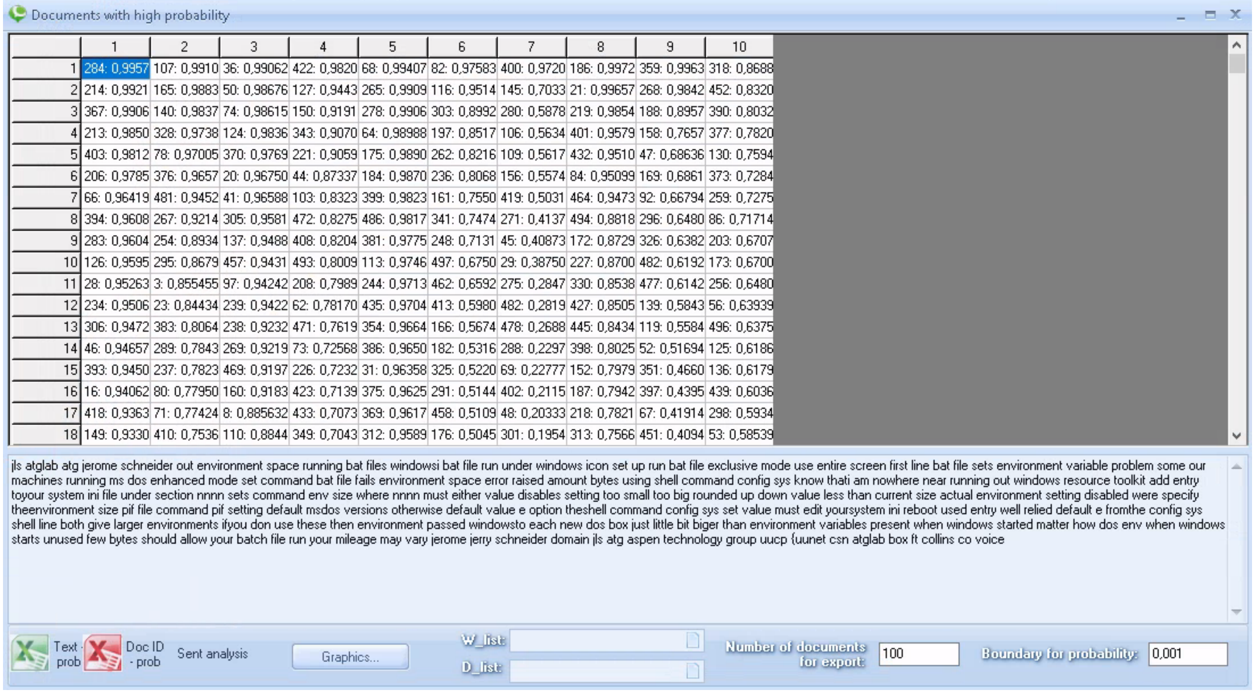
Number of words for export: 100

Boundary for probability: 0,001

If you want to read the most probable documents in any topic, then you can click on the button:

D

You will see the following:



Documents with high probability

	1	2	3	4	5	6	7	8	9	10
1	284: 0,9957	107: 0,9910	36: 0,99062	422: 0,9820	68: 0,99407	82: 0,97583	400: 0,9720	186: 0,9972	359: 0,9963	318: 0,8688
2	214: 0,9921	165: 0,9883	50: 0,98676	127: 0,9443	265: 0,9909	116: 0,9514	145: 0,7033	21: 0,99657	268: 0,9842	452: 0,8320
3	367: 0,9906	140: 0,9837	74: 0,98615	150: 0,9191	278: 0,9906	303: 0,8992	280: 0,5878	219: 0,9854	188: 0,8957	390: 0,8032
4	213: 0,9850	328: 0,9738	124: 0,9836	343: 0,9070	64: 0,98988	197: 0,8517	106: 0,5634	401: 0,9579	158: 0,7657	377: 0,7820
5	403: 0,9812	78: 0,97005	370: 0,9769	221: 0,9059	175: 0,9890	262: 0,8216	109: 0,5617	432: 0,9510	47: 0,68636	130: 0,7594
6	206: 0,9785	376: 0,9657	20: 0,96750	44: 0,87337	184: 0,9870	236: 0,8068	156: 0,5574	84: 0,95099	169: 0,6861	373: 0,7284
7	66: 0,96419	481: 0,9452	41: 0,96588	103: 0,8323	399: 0,9823	161: 0,7550	419: 0,5031	464: 0,9473	92: 0,66794	259: 0,7275
8	394: 0,9608	267: 0,9214	305: 0,9581	472: 0,8275	486: 0,9817	341: 0,7474	271: 0,4137	494: 0,8818	296: 0,6480	86: 0,71714
9	283: 0,9604	254: 0,8934	137: 0,9488	408: 0,8204	381: 0,9775	248: 0,7131	45: 0,40873	172: 0,8729	326: 0,6382	203: 0,6707
10	126: 0,9595	295: 0,8679	457: 0,9431	493: 0,8009	113: 0,9746	497: 0,6750	29: 0,38750	227: 0,8700	482: 0,6192	173: 0,6700
11	28: 0,95263	3: 0,855455	97: 0,94242	208: 0,7989	244: 0,9713	462: 0,6592	275: 0,2847	330: 0,8538	477: 0,6142	256: 0,6480
12	234: 0,9506	23: 0,84434	239: 0,9422	62: 0,78170	435: 0,9704	413: 0,5980	482: 0,2819	427: 0,8505	139: 0,5843	56: 0,63939
13	306: 0,9472	383: 0,8064	238: 0,9232	471: 0,7619	354: 0,9664	166: 0,5674	478: 0,2688	445: 0,8434	119: 0,5584	496: 0,6375
14	46: 0,94657	289: 0,7843	269: 0,9219	73: 0,72568	386: 0,9650	182: 0,5316	288: 0,2297	398: 0,8025	52: 0,51694	125: 0,6186
15	393: 0,9450	237: 0,7823	469: 0,9197	226: 0,7232	31: 0,96358	325: 0,5220	69: 0,22777	152: 0,7979	351: 0,4660	136: 0,6179
16	16: 0,94062	80: 0,77950	160: 0,9183	423: 0,7139	375: 0,9625	291: 0,5144	402: 0,2115	187: 0,7942	397: 0,4395	439: 0,6036
17	418: 0,9363	71: 0,77424	8: 0,885632	433: 0,7073	369: 0,9617	458: 0,5109	48: 0,20333	218: 0,7821	67: 0,41914	298: 0,5934
18	149: 0,9330	410: 0,7536	110: 0,8844	349: 0,7043	312: 0,9589	176: 0,5045	301: 0,1954	313: 0,7566	451: 0,4094	53: 0,58539

jls atglab atg jerome schneider out environment space running bat files windowsi bat file run under windows icon set up run bat file exclusive mode use entire screen first line bat file sets environment variable problem some our machines running ms dos enhanced mode set command bat file fails environment space error raised amount bytes using shell command config sys know thati am nowhere near running out windows resource toolkit add entry toyour system ini file under section nnnn sets command env size where nnnn must either value disables setting too small too big rounded up down value less than current size actual environment setting disabled were specify theenvironment size pif file command pif setting default msdos versions otherwise default value e option theshell command config sys set value must edit yoursystem ini reboot used entry well relied default e fromthe config sys shell line both give larger environments ifyou don use these then environment passed windowsto each new dos box just little bit biger than environment variables present when windows started matter how dos env when windows starts unused few bytes should allow your batch file run your mileage may vary jerome jerry schneider domain jls atg aspen technology group uucp {uunet csn atglab box ft collins co voice

W_list:

D_list:

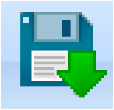
Number of documents for export: 100

Boundary for probability: 0,001

Each cell contains the document ID and its probability in the topic (that is, in the column). To view the contents of a document (original non-lematized text), just place the mouse cursor onto the cell.

5. Saving calculation results as a project:

The calculation results can be saved as a project. The next time, this project can be opened in TopicMiner and you can continue working with it. To save the project, you need to use the following button:



And to open the project use the following button: