

# TagAnt

Big Data and Language

Seonmin Park

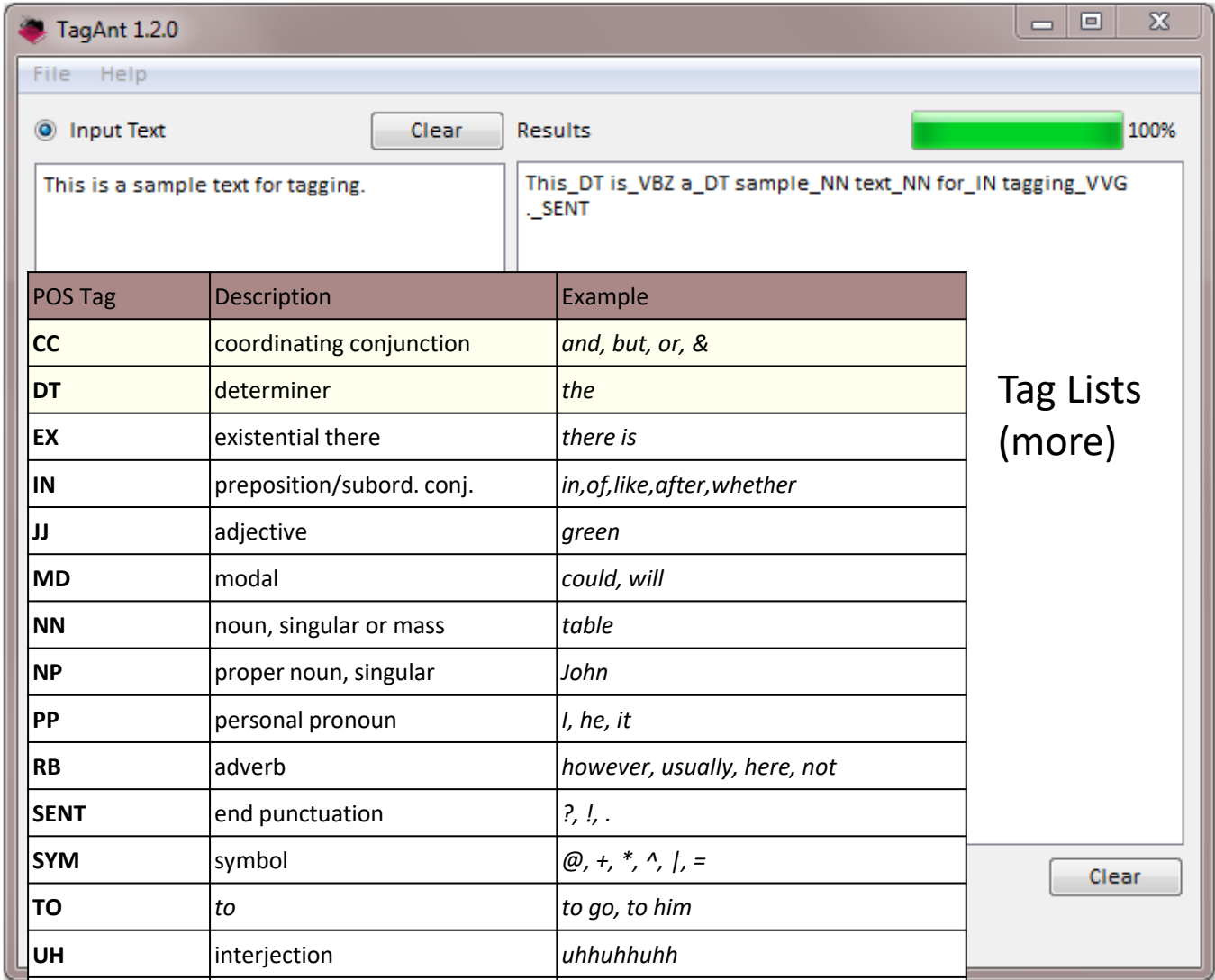
[seonminpark@kaist.ac.kr](mailto:seonminpark@kaist.ac.kr)

# TagAnt



- TagAnt
  - Laurence Anthony, Ph.D
- A freeware (non-commercial) Part-Of-Speech (POS) tagger built on TreeTagger (developed by Helmut Schmid).
- Tagger Set:  
<https://courses.washington.edu/hypertext/csar-v02/penntable.html>

# TagAnt



TagAnt 1.2.0

File Help

☒ Input Text  Results 

100%

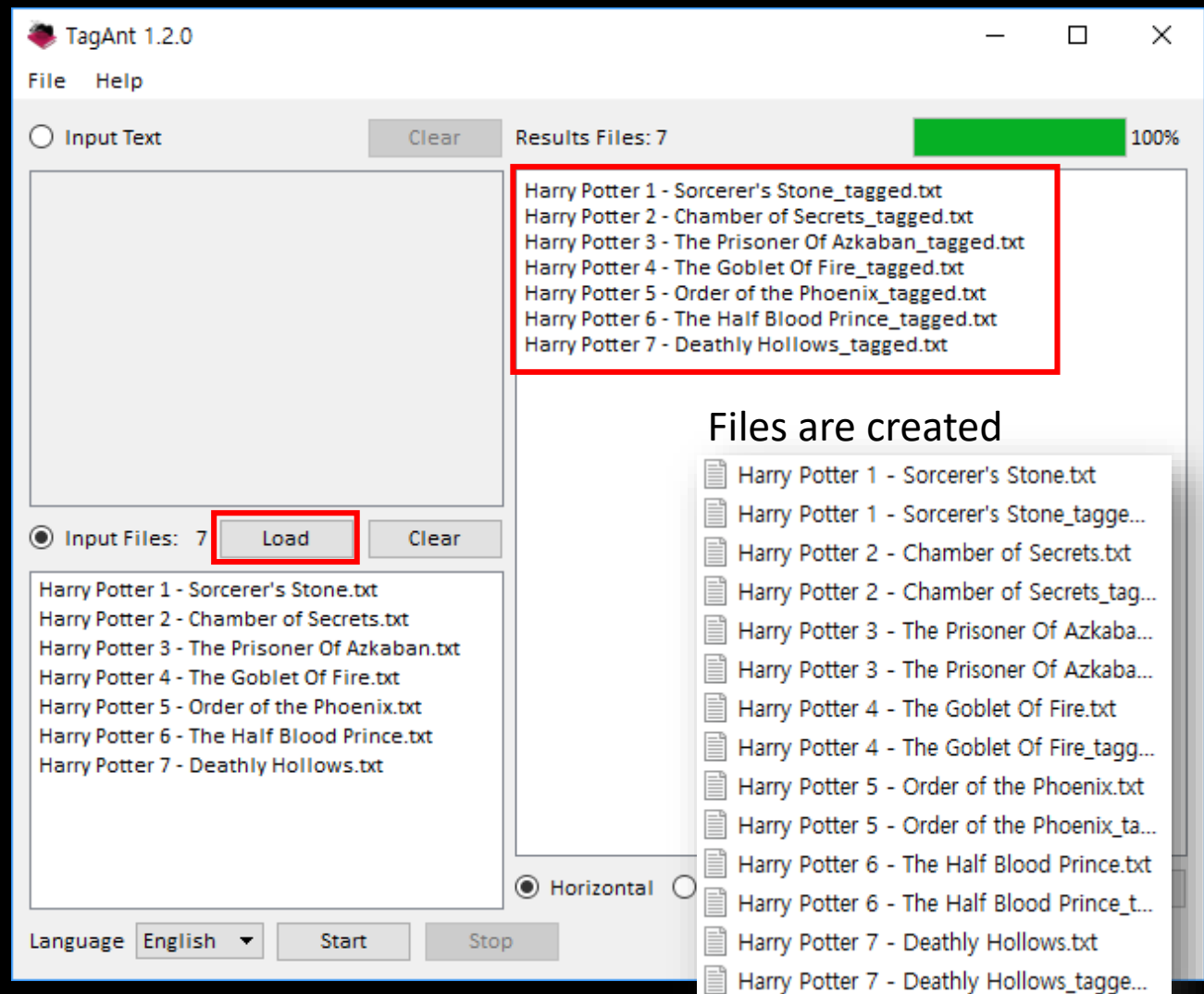
This is a sample text for tagging.

This\_DT is\_VBZ a\_DT sample\_NN text\_NN for\_IN tagging\_VVG  
\_SENT

POS Tag	Description	Example
CC	coordinating conjunction	<i>and, but, or, &amp;</i>
DT	determiner	<i>the</i>
EX	existential there	<i>there is</i>
IN	preposition/subord. conj.	<i>in, of, like, after, whether</i>
JJ	adjective	<i>green</i>
MD	modal	<i>could, will</i>
NN	noun, singular or mass	<i>table</i>
NP	proper noun, singular	<i>John</i>
PP	personal pronoun	<i>I, he, it</i>
RB	adverb	<i>however, usually, here, not</i>
SENT	end punctuation	<i>?, !, .</i>
SYM	symbol	<i>@, +, *, ^,  , =</i>
TO	<i>to</i>	<i>to go, to him</i>
UH	interjection	<i>uhhuhhuhh</i>
VB	verb <i>be</i> , base form	<i>be</i>
WP	wh-pronoun	<i>who, what</i>

Tag Lists  
(more)

# TagAnt



# TagAnt

## Tagged Text

Harry Potter and the Sorcerer's  
Stone

CHAPTER ONE

THE BOY WHO LIVED

Mr. and Mrs. Dursley, of number  
four, Privet Drive, were proud to  
say that they were perfectly normal,  
thank you very much. They were the  
last people you'd expect to be  
involved in anything strange or  
mysterious, because they just didn't  
hold with such nonsense.

Mr. Dursley was the director of a  
firm called Grunnings, which made  
drills. He was a big, beefy man with  
hardly any neck, although he did  
have a very large mustache. Mrs.

Harry\_NP Potter\_NP and\_CC the\_DT  
Sorcerer\_NP 's\_POS Stone\_NP

CHAPTER\_NN ONE\_CD

THE\_DT BOY\_NP WHO\_WP LIVED\_VVD

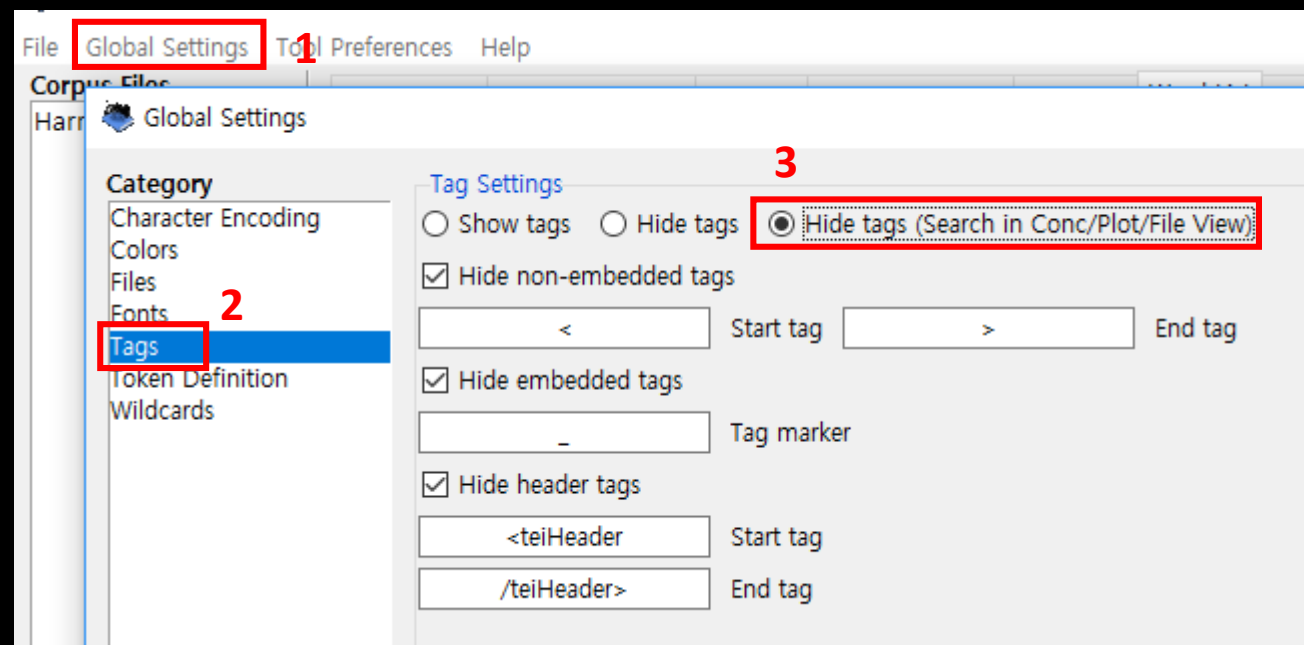
Mr.\_NP and\_CC Mrs.\_NP Dursley\_NP ,\_,  
of\_IN number\_NN four\_CD ,\_,  
Privet\_NP Drive\_NP ,\_, were\_VBD  
proud\_JJ to\_TO say\_VV that\_IN/that  
they\_PP were\_VBD perfectly\_RB  
normal\_JJ ,\_, thank\_VV you\_PP  
very\_RB much\_RB .\_SENT They\_PP  
were\_VBD the\_DT last\_JJ people\_NNS  
you\_PP 'd\_MD expect\_VV to\_TO be\_VB  
involved\_VVN in\_IN anything\_NN  
strange\_JJ or\_CC mysterious\_JJ ,\_,  
because\_IN they\_PP just\_RB did\_VVD  
n't\_RB hold\_VV with\_IN such\_JJ  
nonsense\_NN .\_SENT

# TagAnt

- Open with AntConc again
- Word List may show like this picture at the first time.

Corpus Files		Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Harry Potter 1 - Source		Word Types: 5802		Word Tokens: 168625		Search Hits: 0		
Rank	Freq	Word	Lemma Word Form(s)					
1	9496	nn						
2	9044	pp						
3	8562	in						
4	6824	np						
5	6713	dt						
6	6438	sent						
7	5907	rb						
8	5020	vvd						
9	4021	jj						
10	3717	to						
11	3630	the						
12	2854	vv						
13	2751	nns						

# TagAnt



# TagAnt

- Now it shows ignoring tag

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Word Types: 5760		Word Tokens: 81453		Search Hits: 0		
Rank	Freq	Word	Lemma Word Form(s)			
1	3630	the				
2	1924	and				
3	1861	to				
4	1758	he				
5	1691	a				
6	1327	harry				
7	1267	of				
8	1261	was				
9	1185	it				
10	1035	you				
11	1010	s				
12	967	in				
13	937	his				



# TagAnt

## Concordance search in “Show Tag”

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 17

Hit	KWIC	File
1	ID know_VV loads_NNS of_IN magic_NN already_RB .SENT	Harry Potte
2	_DT lot_NN more_RBR to_TO magic_NN ,_ as_IN Harry_NP	Harry Potte
3	ow_VV anything_NN about_IN magic_NN at_IN all_DT .SENT	Harry Potte
4	_IN my_PP\$ family_NN 's_POS magic_NN at_IN all_DT ,_ it_P	Harry Potte
5	\$ allowed_VVN to_TO use_VV magic_NN at_IN home_NN .S	Harry Potte
6	\$ eyes_NNS .SENT " _" A_DT magic_NN beyond_IN all_DT v	Harry Potte
7	" _" What_WP sort_NN of_IN magic_NN do_VVP you_PP tea	Harry Potte
8	_NNS not_RB ter_NN use_NN magic_NN in_IN the_DT Mugg	Harry Potte
9	except_IN powerful_JJ Dark_JJ magic_NN --_: no_DT kid_NN	Harry Potte
10	\$ s'pposed_JJ ter_NN use_NN magic_NN now_RB I_PP 've_V	Harry Potte
11	N and_CC force_VV some_DT magic_NN out_RP of_IN me_P	Harry Potte
12	V with_IN some_DT secret_JJ magic_NN .SENT " _" Good	Harry Potte
13	, are_VBP you_PP doing_VVG magic_NN ?_SENT Let_VV 's_V	Harry Potte
14	mprecise_JJ branch_NN of_IN magic_NN .SENT " _" The_D	Harry Potte

< > < >

Search Term ☒ Words ☐ Case ☐ Regex

magic\_NN

Advanced

Search Window Size 50

# TagAnt

## Concordance search in “Hide Tag (search in Conc/Plot/File view)”

The screenshot displays the TagAnt software interface with the 'Concordance' tab selected. The search term 'magic\_NN' is entered in the 'Search Term' field at the bottom, which is highlighted with a red box. The search results are shown in a table with columns 'Hit', 'KWIC', and 'File'. The 'Hit' column lists line numbers 1 through 14. The 'KWIC' column shows the context of the word 'magic' in various sentences, with the word itself highlighted in blue. The 'File' column indicates that all hits are from 'Harry Potte'.

Hit	KWIC	File
1	you must know loads of magic already ._SENT " The	Harry Potte
2	was a lot more to magic , as Harry quickly found	Harry Potte
3	Nobody in my family 's magic at all , it was ever	Harry Potte
4	n't know anything about magic at all ._SENT How can	Harry Potte
5	're not allowed to use magic at home ._SENT I 'm	Harry Potte
6	, wiping his eyes ._SENT " A magic beyond all we do here	Harry Potte
7	you ._SENT " " What sort of magic do you teach , Profess	Harry Potte
8	, but yer not ter use magic in the Muggle world e:	Harry Potte
9	omstick except powerful Dark magic -- no kid could do tha	Harry Potte
10	Not s'pposed ter use magic now I 've got yeh ._	Harry Potte
11	my guard and force some magic out of me -- he push	Harry Potte
12	to tingle with some secret magic ._SENT " Good aftern	Harry Potte
13	._SENT " Oh , are you doing magic ?_SENT Let 's see it ,	Harry Potte
14	a very imprecise branch of magic ._SENT " The sky had	Harry Potte

Search Term ☒ Words ☐ Case ☐ Regex  
magic\_NN Advanced Search Window Size 50

# TagAnt

## ■ Wildcard Search

Wildcard Settings

*	zero or more characters
+	zero or one character
?	any one character
@	zero or one word
#	any one word
	search term 'OR' search term
&	non word

Concordance Hits 2854

Hit	KWIC
1	d_CC he_PP could_MD n't_RB see_VW a_DT single_JJ collecti Harry Potte
2	!_SENT But_CC I_PP can_MD promise_VW a_DT wet_JJ night Harry Potte
3	W ..._: he_PP could_MD n't_RB kill_VW a_DT little_JJ boy_NN ? Harry Potte
4	thead_NN they_PP could_MD see_VW a_DT curiously_RB sha Harry Potte
5	_DT very_RB sad_JJ ,_, but_CC get_VW a_DT grip_NN on_IN y Harry Potte
6	ry_NP ,_, who_WP could_MD see_VW a_DT huge_JJ Dudley_I Harry Potte
7	_PP was_VBD careful_JJ to_TO walk_VW a_DT little_JJ way_NN Harry Potte
8	_JJ Harry_NP tried_VVD to_TO get_VW a_DT closer_JJR look_I Harry Potte
9	where_WRB he_PP could_MD see_VW a_DT tiny_JJ ray_NN o Harry Potte
10	f_IN he_PP dared_VVD to_TO ask_VW a_DT question_NN ._SI Harry Potte
11	_WDT everyone_NN got_VVD hit_VW a_DT lot_NN by_IN the Harry Potte
12	him_PP to_TO go_VV and_CC make_VW a_DT cup_NN of_IN Harry Potte
13	le_NP Vernon_NP would_MD take_VW a_DT sharp_JJ turn_NI Harry Potte
14	_SENT He_PP tried_VVD to_TO start_VW a_DT fire_NN but_CC Harry Potte

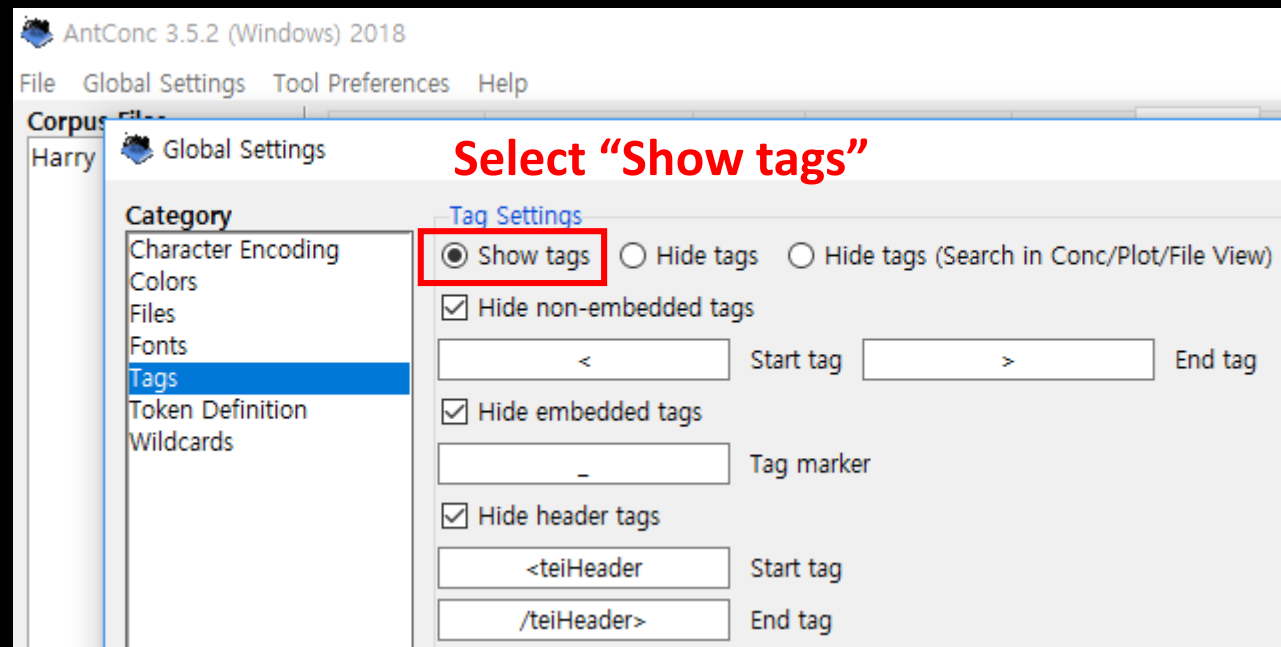
Search Term ☒ Words ☐ Case ☐ Regex

Advanced Search Window Size 50

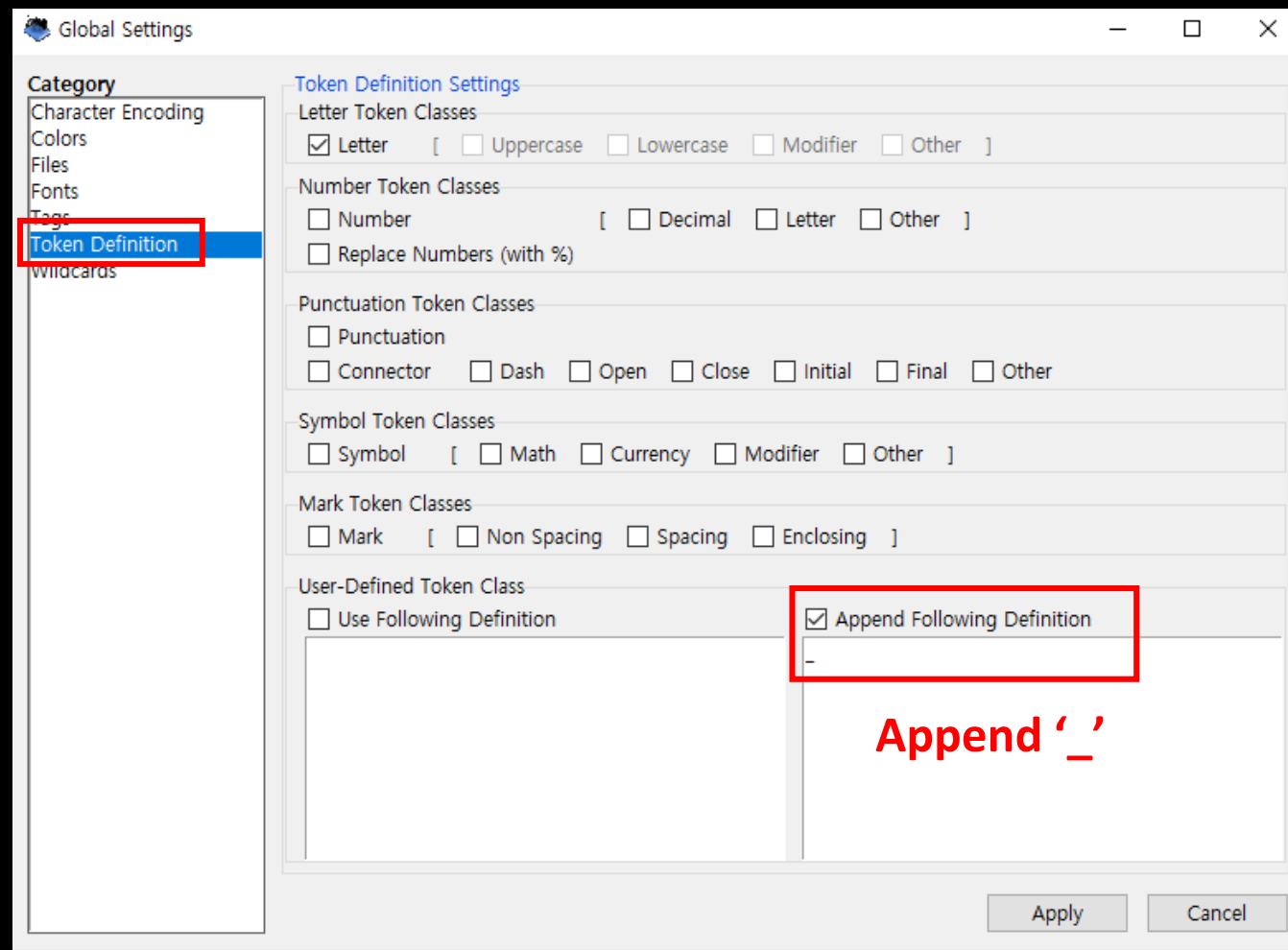
Searching all the verbs

# TagAnt

- Q: What is the frequent nouns?



# TagAnt



# TagAnt

- Now, it shows the words with tags.

Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

Word List

Keyword List

Word Types: 7375

Word Tokens: 100519

Search Hits: 465

Rank

Freq

Word

Lemma

Word Form(s)

1

11988

\_

2

6421

\_sent

3

3622

the\_dt

4

1915

and\_cc

5

1856

to\_to

6

1756

he\_pp

7

1686

a\_dt

8

1325

harry\_np

9

1261

was\_vbd

10

1258

of\_in

11

1184

it\_pp

12

1005

you\_pp

13

950

in\_in

<

>

<

>

<

>

<

>

Search Term

☒ Words

☐ Case

☐ Regex

Hit Location

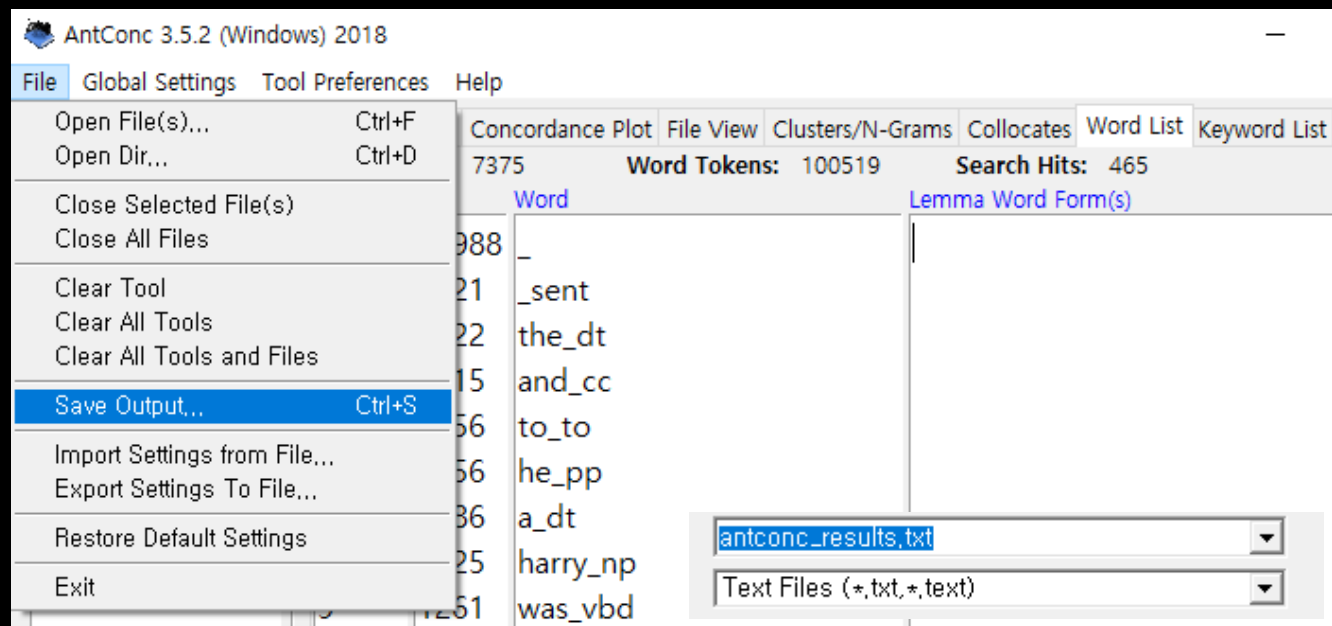
Advanced

Search Only

1

# TagAnt

- Filter out noun?
- Use Excel



# TagAnt

- Copy & Paste the result to the Excel

	A	B	C	D
1	#Word Types: 7375			
2	#Word Tokens: 100519			
3	#Search Hits: 465			
4	1	11988	_	
5	2	6421	_sent	
6	3	3622	the_dt	
7	4	1915	and_cc	
8	5	1856	to_to	
9	6	1756	he_pp	
10	7	1686	a_dt	
11	8	1325	harry_np	
12	9	1261	was_vbd	
13	10	1258	of_in	
14	11	1184	it_pp	
15	12	1005	you_pp	
16	13	950	in_in	
17	14	937	his_pp	



# TagAnt

- Modify the first line into indexes.

	A	B	C	D
1	No.	Freq.	Tag	
2	1	11988	_	
3	2	6421	_sent	
4	3	3622	the_dt	
5	4	1915	and_cc	
6	5	1856	to_to	
7	6	1756	he_pp	
8	7	1686	a_dt	
9	8	1325	harry_np	
10	9	1261	was_vbd	
11	10	1258	of_in	
12	11	1184	it_pp	
13	12	1005	you_pp	
14	13	950	in_in	
15	14	937	his_pp	
16	15	909	i_pp	
17	16	832	n	

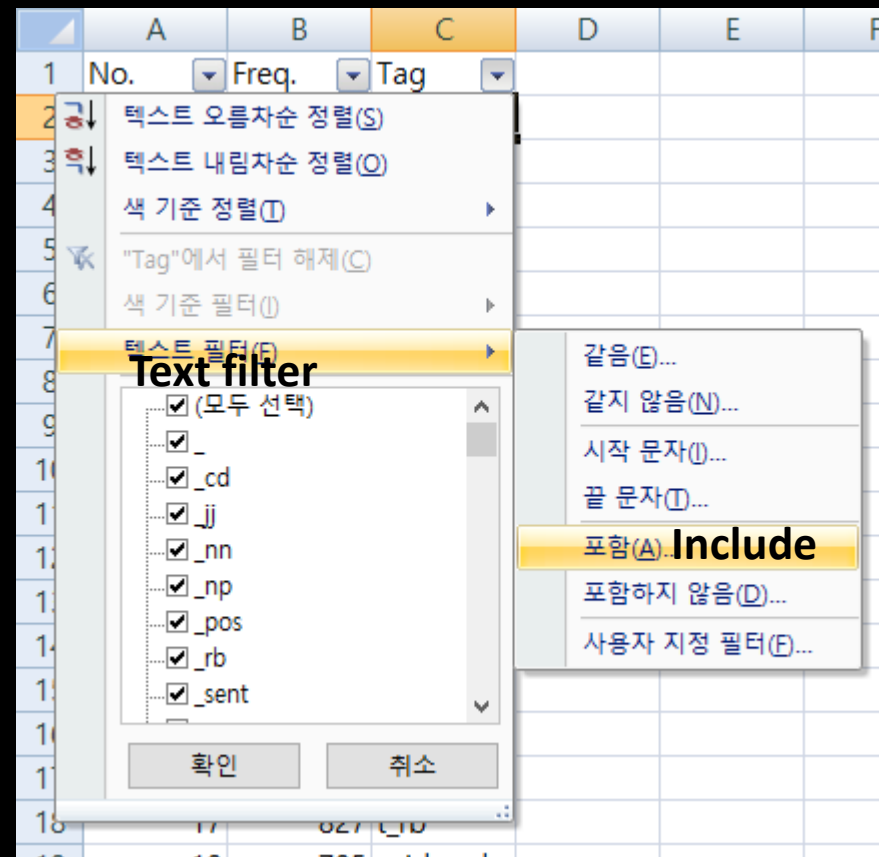
# TagAnt

- Click on Sort & Filter

	A	B	C	D
1	No.	Freq.	Tag	
2	1	11988	_	
3	2	6421	_sent	
4	3	3622	the_dt	
5	4	1915	and_cc	
6	5	1856	to_to	
7	6	1756	he_pp	
8	7	1686	a_dt	
9	8	1325	harry_np	
10	9	1261	was_vbd	
11	10	1258	of_in	
12	11	1184	it_pp	
13	12	1005	you_pp	
14	13	950	in_in	



# TagAnt



# TagAnt

	A	B	C	D	E	F	G	H
1	No.	Freq.	Tag					
96	95	129	something_nn					
104	103	119	time_nn					
114	113	105	door_nn					
131	130	94	head_nn					
132	131	94	way_nn					
137	136	90	room_nn					
146	145	82	face_nn					
157	156	78	ter_nn					
161	160	73	anything_nn					
163	162	71	boy_nn					
165	164	71	house_nn					
177	176	67	hand_nn					

사용자 지정 자동 필터

찾을 조건:  
Tag **End letter** **\_NN**

끝 문자

☒ 그리고(A) ☐ 또는(O)

? 기호를 사용하여 한 문자를 나타낼 수 있습니다.  
\* 기호를 사용하여 여러 문자를 나타낼 수 있습니다.

확인 취소

- We can conclude that the frequent nouns are *something, time, door, head, way, ...* in Harry Potter.