

Super Sense Tagging: UKWaC super sensed

Diana McCarthy and Siva Reddy

Lexical Computing Ltd

29th Feb 2012

Semantic Class Tagging

Super Sense Tagger (SST) [Ciaramita and Altun, 2006]
(<http://sourceforge.net/projects/supersensetag/>)

- ▶ semantic tags are WordNet lexicographer classes
- ▶ supervised (requires training data) HMM system (tag sequences)
- ▶ SemCor used as training data
- ▶ Named Entity Recognition
- ▶ Multiword tagging using WordNet

Noun Classes

act	acts or actions
object	natural objects (not man-made)
animal	animals
quantity	quantities and units of measure
artifact	man-made objects
phenomenon	natural phenomena
attribute	attributes of people and objects plant plants
food	food and drinks
...	...

Verb Classes

body	grooming, dressing and bodily care
emotion	feeling
change	size, temperature change, intensifying
motion	walking, flying, swimming
cognition	thinking, judging, analyzing, doubting
perception	seeing, hearing, feeling
communication	telling, asking, ordering, singing
possession	buying, selling, owning
creation	sewing, baking, painting, performing
...	...

Semantic Tags in the Concordance

Corpus: UKWaC super sensed
Hits: 37657 (101.8 per million)

Page of 1883 [Next](#) | [Last](#)

#10858	<mwe> room temperature </mwe> is a new <mwe>	ball /NN/act.n	game </mwe> . One approach is to anchor a
#18643	possible improvements . '' You shall go to the	Ball /NNP/other.n	" We all wanted to be a princess when
#18818	buffet table . Brides attire A huge <mwe>	ball /NN/artifact.n	gown </mwe> is must and tiara . Clear shoes
#41021	playing football , <mwe> picked up </mwe> the	ball /NN/artifact.n	and then threw himself onto the ground
#41062	doing exactly the same thing : chucking the	ball /NN/artifact.n	down a little line of blokes . Neither
#41103	waitagoddamnminute ! They 're going to chuck the	ball /NN/artifact.n	down the little line of blokes ! I 'll
#41244	get to get 2 more points by kicking the	ball /NN/artifact.n	between two very wide and very tall posts
#41283	difficult stuff . They have to be able to kick a	ball /NN/artifact.n	in a certain direction ! It 's their full
#41314	only element of chance when kicking the	ball /NN/artifact.n	, unobstructed , through two very wide
#41359	. That 's it . If they manage to get the	ball /NN/artifact.n	between the two very wide and very tall
#116062	players and they keep stopping and passing the	ball /NN/artifact.n	back instead of <mwe> getting on with </mwe>
#140081	you have a planet like Jupiter , a great	ball /NN/artifact.n	of gas and liquid hydrogen , if you go <mwe>
#140440	want your rocket to escape the Earth 's	ball /NN/artifact.n	and not crash back into the sea but to
#147305	what it will do , just like a <mwe> billiard	ball /NN/artifact.n	</mwe> , it will hit a nucleus in the silicon
#293085	helped us <mwe> look into </mwe> the <mwe> crystal	ball /NN/artifact.n	</mwe> to create four 2012 worlds . Stepping
#391787	catch it . Oh , then the huge disco <mwe>	ball /NN/artifact.n	thing </mwe> appears , with the white lights
#439151	</mwe> analogous to pushing one <mwe> billiard	ball /NN/artifact.n	</mwe> with another . <mwe> In addition </mwe>
#472963	circular <mwe> billiard table </mwe> must a <mwe> cue	ball /NN/artifact.n	</mwe> be aimed so as to hit a given target
#472973	</mwe> be aimed so as to hit a given target	ball /NN/artifact.n	? <mwe><ne> Omar Khayyam </ne></mwe> (1048
#482745	visit <mwe><ne> www.scottishkpls.co.uk EAA	Ball /NNP/group.n	& Awards </mwe> Presentation 17 March 2005

Page of 1883 [Next](#) | [Last](#)

Semantic Tags in the Concordance

Corpus: UKWaC super sensed
Hits: 15772 (42.6 per million)

[First](#) | [Previous](#) Page of 789 [Go](#) [Next](#) | [Last](#)

#1493049	credibility that surrounds the <mwe><ne> Mickey	Mouse /NNP/other.n	</ne></mwe><mwe> teaching certificate </mwe>
#1496976	hovering , <mwe> seeking out </mwe> a vole or	mouse /NN/animal.n	. INFORMATION FOR VOLUNTEERS If you would
#1545720	audio visuals and the <mwe><ne> Soviet Spy	Mouse /NNP/other.n	Trail </ne></mwe> . Open daily all year round
#1561637	even after disruption . Control through the	mouse /NN/animal.n	Interactive stories make no use of the
#1561653	keyboard . They are controlled entirely by the	mouse /NN/animal.n	which moves the on-screen pointer and allows
#1561673	of effects . In a very real sense , the	mouse /NN/animal.n	represents control of the activity . Other
#1561693	importance of access to the controlling device (mouse /NN/animal.n	or keyboard , <mwe> depending on </mwe> the
#1561728	</mwe> , 1993) . Seen in this light , the	mouse /NN/animal.n	might be considered a symbol of power ,
#1561813	different methods were used when passing the	mouse /NN/animal.n	to the <ne> next 'turn' </ne> . Sometimes ,
#1561826	Sometimes , <mwe> for example </mwe> , when the	mouse /NN/animal.n	had been left in a 'neutral ' position
#1561862	remaining member of the group moved the	mouse /NN/animal.n	towards the next turn . This was thought
#1561928	within the group . If simply leaving the	mouse /NN/animal.n	is seen as unhelpful to the next turn ,
#1562003	members who were not in possession of the	mouse /NN/animal.n	issued a significant number of commands
#1562048	in <mwe> actual possession </mwe> . While the	mouse /NN/animal.n	gave the undisputed right to control the
#1562081	directed the same commands to the holder of the	mouse /NN/animal.n	, adding extra psychological pressure to
#1608486	and dragging the control points using the	mouse /NN/animal.n	(left button) or inputting the exact
#1608563	dialog or dragging the points using left <mwe>	mouse /NN/artifact.n	button </mwe> . A view of all the surfaces
#1664217	with a cenral finger-ball (e.g. marble	mouse /NN/animal.n) . Just more physical stability , thats
#1768071	students attending workshops (eg. <mwe>	mouse /NN/artifact.n	mats </mwe> , pens etc) 4.2 How the tutorial
#1831217	</ne></mwe> include the <mwe><ne> Yellow Necked	Mouse /NNP/other.n	</ne></mwe> , Weasels , <mwe><ne> Fallow Deer

[First](#) | [Previous](#) Page of 789 [Go](#) [Next](#) | [Last](#)

Semantic Tags in the Word Sketch (selected)

eat *(verb)* UKWaC super sensed freq = 26329 (71.2 per million)

<u>transframe</u>	<u>1241</u>	<u>6.6</u>	<u>Intransframe</u>	<u>1021</u>	<u>2.7</u>
person.n_*consumption.v_food.n	<u>178</u>	11.42	animal.n_*consumption.v	<u>85</u>	10.84
group.n_*consumption.v_food.n	<u>57</u>	10.25	person.n_*consumption.v	<u>382</u>	10.25
person.n_*consumption.v_plant.n	<u>37</u>	9.77	group.n_*consumption.v	<u>145</u>	9.82
person.n_*consumption.v_animal.n	<u>35</u>	9.62	0_*consumption.v	<u>61</u>	9.73
animal.n_*consumption.v_animal.n	<u>30</u>	9.56	state.n_*consumption.v	<u>20</u>	8.89
animal.n_*consumption.v_plant.n	<u>25</u>	9.32	time.n_*consumption.v	<u>19</u>	8.74
animal.n_*consumption.v_food.n	<u>24</u>	9.21	communication.n_*consumption.v	<u>30</u>	8.57
person.n_*consumption.v_person.n	<u>52</u>	8.97	artifact.n_*consumption.v	<u>30</u>	8.51
0_*consumption.v_food.n	<u>20</u>	8.92	food.n_*consumption.v	<u>19</u>	8.4
animal.n_*consumption.v_artifact.n	<u>19</u>	8.79	other.n_*consumption.v	<u>15</u>	8.1

Semantic Tags in the Word Sketch (selected)

laugh (*verb*) UKWaC super sensed freq = 6489 (17.5 per million)

<u>V_PP</u>	<u>148</u>	<u>9.1</u>	<u>Intransframe</u>	<u>1101</u>	<u>8.9</u>
*body.v_at_cognition.n	<u>14</u>	11.02	person.n_*body.v	<u>556</u>	10.49
*body.v_at_communication.n	<u>12</u>	10.94	group.n_*body.v	<u>143</u>	10.02
*body.v_at_person.n	<u>7</u>	9.77	0_*body.v	<u>102</u>	10.33
*body.v_as_person.n	<u>6</u>	9.78	artifact.n_*body.v	<u>49</u>	9.19
*communication.v_at_location.n	<u>6</u>	7.8	time.n_*body.v	<u>23</u>	8.49
*body.v_in_cognition.n	<u>5</u>	9.58	location.n_*body.v	<u>16</u>	7.88
*communication.v_at_communication.n	<u>5</u>	8.39	event.n_*body.v	<u>9</u>	7.56
*communication.v_at_person.n	<u>5</u>	8.14	other.n_*body.v	<u>8</u>	7.38
			cognition.n_*body.v	<u>7</u>	7.2
			communication.n_*body.v	<u>7</u>	6.89

Semantic Word Sketch, using several collocations together

=intransframe

COLLOC "%(2.sense)-%(1.sense)-x"

2:any_noun rel_start? adv_aux_string_incl_be 1:verb_not_pp
not_np_start

2:any_noun rel_start? adv_aux_string_incl_be aux_have
adv_string 1:past_part not_np_start

MWEs: detected by SST

bread *(noun)* UKWaC super sensed freq = 8355 (22.6 per million)

<u>mwe</u>	<u>1887</u>	<u>0.8</u>
bread_and_butter_possession.n	<u>203</u>	11.37
loaf_of_bread_food.n	<u>180</u>	11.25
white_bread_food.n	<u>171</u>	11.3
garlic_bread_food.n	<u>74</u>	10.22
brown_bread_food.n	<u>71</u>	10.16
french_bread_food.n	<u>53</u>	9.77
unleavened_bread_food.n	<u>42</u>	9.45
bread_street_location.n	<u>34</u>	9.15
bread_maker_person.n	<u>31</u>	9.03
rye_bread_food.n	<u>30</u>	8.98
bread_knife_artifact.n	<u>21</u>	8.48

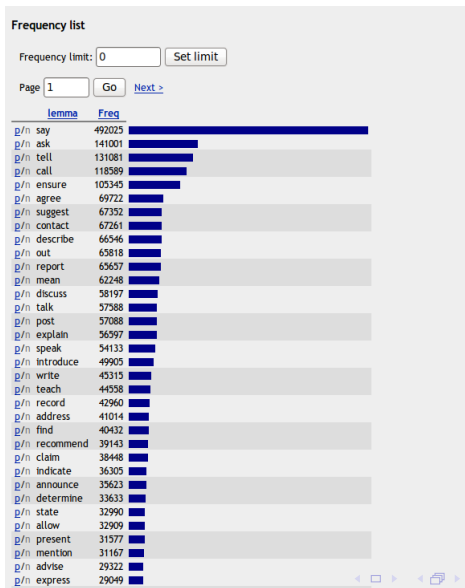
MWEs: Sketch Diff chip (green) vs chips (red)

mwe	325	314	-2.0	-2.0
fish_and_chip_food.n	0	<u>70</u>	0.0	10.4
tortilla_chip_food.n	0	<u>14</u>	0.0	8.3
potato_chip_food.n	<u>10</u>	<u>40</u>	7.7	9.7
memory_chip_artifact.n	<u>25</u>	<u>32</u>	9.0	9.4
gene_chip_artifact.n	<u>11</u>	<u>10</u>	7.9	7.8
poker_chip_artifact.n	<u>14</u>	<u>12</u>	8.3	8.0
silicon_chip_artifact.n	<u>40</u>	<u>27</u>	9.7	9.1
bargaining_chip_attribute.n	<u>29</u>	<u>7</u>	9.3	7.2
chocolate_chip_food.n	<u>14</u>	0	8.3	0.0
chip_shot_act.n	<u>15</u>	0	8.4	0.0

Portion of Sketch Diff laugh (green) vs cry (red)

V_PP	148	117	9.1	7.0
*communication.v_in_location.n	0	<u>6</u>	0.0	7.3
*stative.v_for_act.n	0	<u>11</u>	0.0	6.6
*communication.v_at_location.n	<u>6</u>	0	7.8	0.0
*communication.v_at_person.n	<u>5</u>	0	8.1	0.0
*communication.v_at_communication.n	<u>5</u>	0	8.4	0.0
*body.v_in_cognition.n	<u>5</u>	0	9.6	0.0
*body.v_at_person.n	<u>7</u>	0	9.8	0.0
*body.v_as_person.n	<u>6</u>	0	9.8	0.0
*body.v_at_communication.n	<u>12</u>	0	10.9	0.0
*body.v_at_cognition.n	<u>14</u>	0	11.0	0.0

Semantic Word Lists: CQL + word Frequency (Communication Verbs)



Semantic Word Lists: FindX (communication verbs)

0.1	122.9	say-v	271
0.1	119.4	tell-v	62
0.1	112.0	ask-v	75
0.2	101.9	out-v	53
0.2	100.0	humour-v	53
0.2	100.0	critique-v	142
0.2	100.0	underline-v	2166
0.2	100.0	stammer-v	50
0.2	100.0	reintroduce-v	501
0.2	100.0	re-introduce-v	109
0.2	100.0	shriek-v	116
0.2	100.0	exhort-v	88
0.2	100.0	publicise-v	1244
0.2	100.0	chide-v	116
0.3	100.0	interrogate-v	730
0.3	100.0	fate-v	67
0.3	100.0	bemoan-v	277
0.3	100.0	absolve-v	136
0.3	100.0	signpost-v	160
0.3	100.0	unrated-v	321
0.3	100.0	chronicle-v	487
0.3	100.0	telegraph-v	75
0.3	100.0	spam-v	218
0.3	100.0	misquote-v	119
0.4	100.0	extol-v	87
0.4	100.0	eschew-v	322
0.4	100.0	nominates-v	71
0.4	100.0	evince-v	156
0.4	100.0	spook-v	66
0.4	100.0	rejuvenate-v	205
0.4	100.0	symbolise-v	292
0.4	100.0	pardon-v	445

Reference



Ciaramita, M. and Altun, Y. (2006).

Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger.

In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 594–602, Sydney, Australia. Association for Computational Linguistics.