



Cognitive Systems

Address: 141014, Moscow Oblast, Mytishchi,
12 Veri Voloshinoy St., office 714;
office@cogsys.company

Brain2Morph

Brain2Morph is a machine-learning module, which is aimed at establishing the part of speech and morphological tagging of words in natural texts. Morphological markup is an important part of machine processing of natural texts, since it helps to understand how the words in the sentence are related.

Brain2Morph module is based on a neural network, trained on a corpus of approximately 60 thousand words. Currently the module recognizes the following word properties:

1) Part of speech - noun, adjective, verb, pronoun, proper noun, adverb, determinant, preposition, particle, conjunction. In this case, participles and gerunds are considered as a form of the verb.

2) Morphological tags - gender, person, number, case, animacy, tense, voice, verb form, degree.

Also, the system is able to automatically recognize the language (Russian and English) and with the use of context to eliminate contradictions with homonyms.



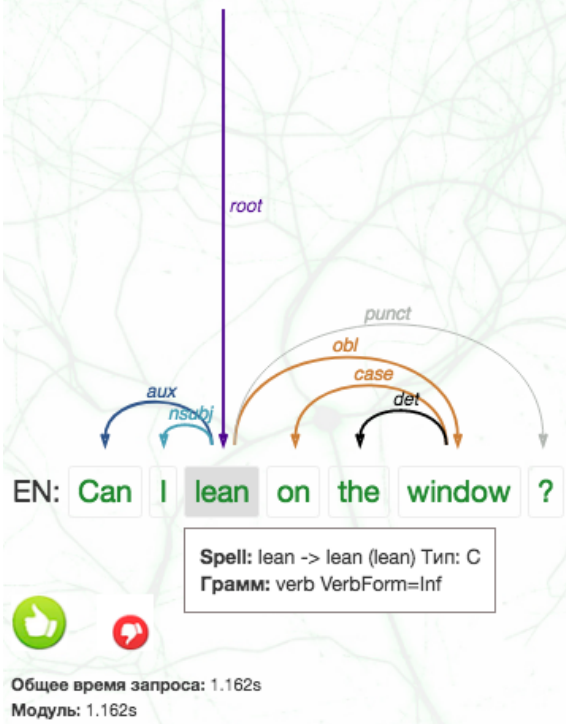
Cognitive Systems

Address: 141014, Moscow Oblast, Mytishchi,
12 Veri Voloshinoy St., office 714;
office@cogsys.company

Brain2Spell alfa RU/EN

Can I lean on the window?

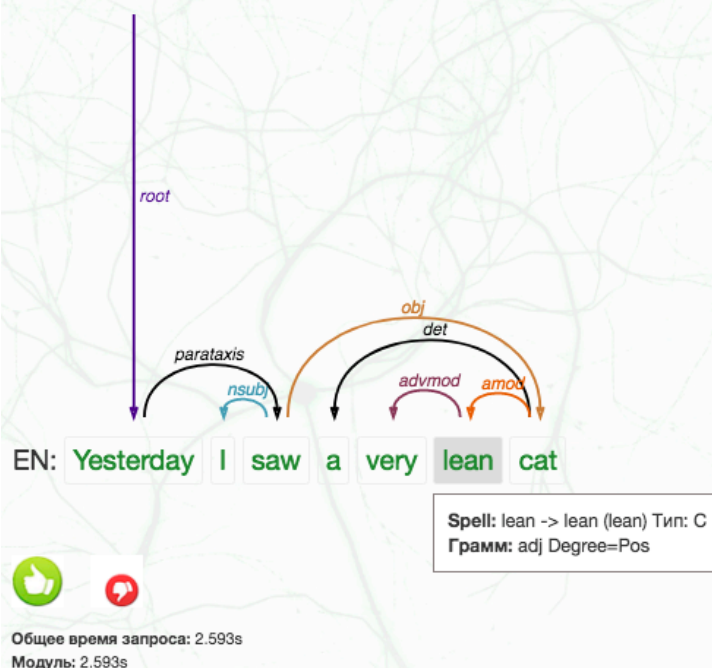
Коррекция ошибок



Brain2Spell alfa RU/EN

Yesterday I saw a very lean cat

Коррекция ошибок





Cognitive Systems

Address: 141014, Moscow Oblast, Mytishchi,
12 Veri Voloshinoy St., office 714;
office@cogsys.company

Conducted tests showed good accuracy results:

Neuromodel	Russian language	English language
Part of Speech (POS), mean value	93.5*	84.3*
Morphological Tags, mean value	84.9	79.2
Defined parts of speech	Total - 15 adj, adp, adv, aux, cconj, det, intj, noun, num, part, pron, propn, sconj, sym, verb, x	Total - 16 adj, adp, adv, aux, cconj, det, intj, noun, num, part, pron, propn, sconj, sym, verb, x
Defined morphological tags	Total - 26 'Case': {'Acc', 'Dat', 'Gen', 'Ins', 'Loc', 'Nom', 'Par', 'Voc'}, 'Gender': {'Fem', 'Masc', 'Neut'}, 'Number': {'Plur', 'Sing'}, 'Person': {'1', '2', '3'}, 'Tense': {'Fut', 'Past', 'Pres'}, 'VerbForm': {'Conv', 'Fin', 'Inf', 'Part'}, 'Voice': {'Act', 'Mid', 'Pass'}}	Total - 28 {'Case': {'Acc', 'Nom'}, 'Degree': {'Cmp', 'Pos', 'Sup'}, 'Gender': {'Fem', 'Masc', 'Neut'}, 'Mood': {'Imp', 'Ind'}, 'Number': {'Plur', 'Sing'}, 'Person': {'1', '2', '3'}, 'PronType': {'Art', 'Dem', 'Int', 'Prs', 'Rel'}, 'Tense': {'Past', 'Pres'}, 'VerbForm': {'Fin', 'Ger', 'Inf', 'Part'}, 'Voice': {'Pass'; Act}}

* The accuracy results were calculated based on the model with 100% of words having 2 or more parts of speech. In the case of mixed text analysis with only 50% of words with ambiguous parts of speech, the accuracy for Russian and English will be approximately 96% for Russian and 92.15 for English.



Cognitive Systems

Address: 141014, Moscow Oblast, Mytishchi,
12 Veri Voloshinoy St., office 714;
office@cogsys.company

We have compared our models' accuracy results on Russian language with the results obtained on a similar task by teams during the MorphoRuEval-2017 morphological markup competition. Brain2Morph algorithm demonstrated best quality results – 94.23. Results of the teams in the competition varied from 75.88 to 93.71. It is worth noting that the competition conditions implied evaluating not a full set of tags - for example, the tense of the verbs was simplified to *present* and *not present*. The comparison results can be seen in the table below.

Algorithm name or author	Annotated words in training corpus	Words* in test corpus (mixed)	Accuracy - POS + Features	Year
MSU-1	3 500 000	13 150	93,39	2017
IQMEN	3 500 000	13 150	93,08	2017
Sagteam	3 500 000	13 150	92,64	2017
Aspect	3 500 000	13 150	92,57	2017
Morphobabushka	3 500 000	13 150	90,07	2017
Pullenti POS Tagger	3 500 000	13 150	89,96	2017
Shacker	3 500 000	13 150	89,91	2017
N	3 500 000	13 150	89,86	2017
Xmorphy	3 500 000	13 150	89,46	2017
Koziev	3 500 000	13 150	88,14	2017
I	3 500 000	13 150	86,05	2017
L	3 500 000	13 150	71,48	2017
Brain2Tag	1 423 631	20 000	93,5	2017

The following link may be used to test the alpha version of the module combined with the algorithm for correcting errors in words - <http://cogsys.company/ru/brain2spell>.