ООО «Когнитивные системы»
ОГРН: 1165029057490; ИНН: 5029214276
Адрес: 141014, Московская обл, г.Мытищи,
ул. Веры Волошиной дом 12, офис 714;
office@cogsys.company

Brain2Morph

Brain2Morph представляет собой модуль Большой Лингвистической Модели, ответственный за определение части речи и морфологических признаков слова. Морфологическая разметка является важной составляющей машинной обработки естественных текстов, так как помогает понимать, как связаны слова в предложении.

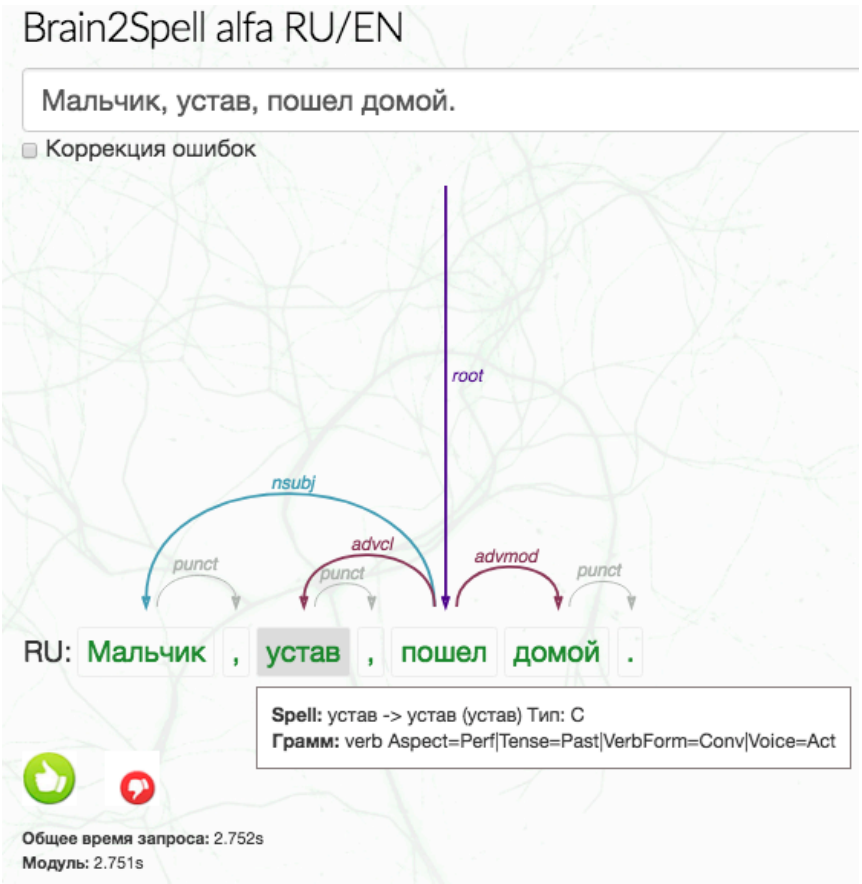
Модуль Brain2Morph создан на основе нейронной сети, обученной на корпусе в 60 тысяч слов. На текущий момент модуль распознает следующие свойства слова:

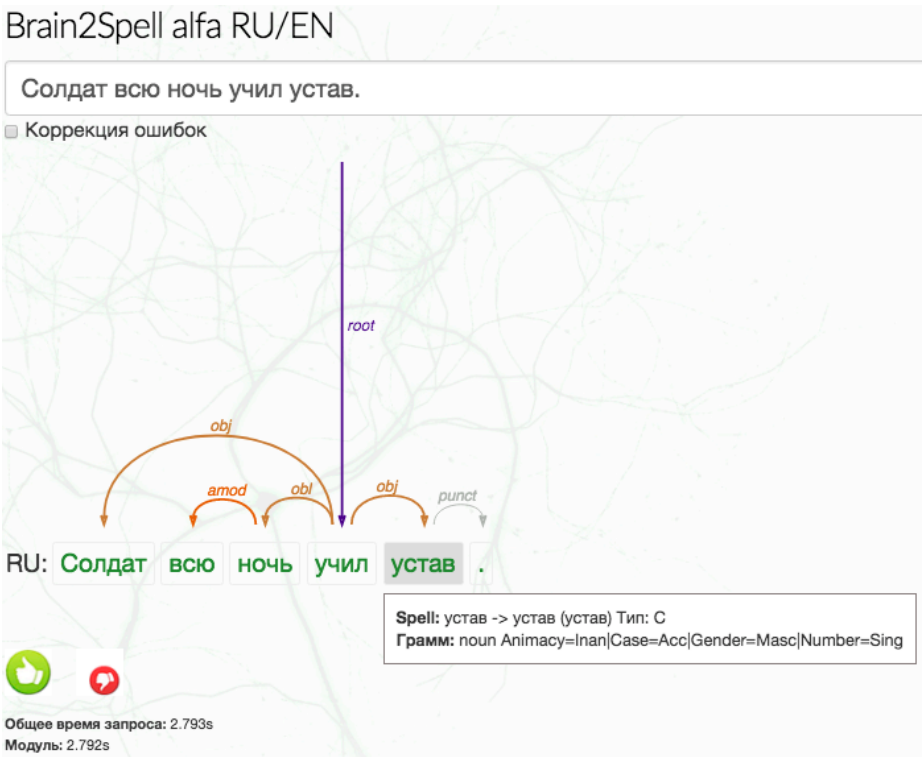
1) Часть речи - существительное, прилагательное, глагол, местоимение, имя собственное, наречие, определяющее слово, предлог, частица, союз. При этом причастие и деепричастие рассматриваются, как форма глагола.

2) Морфологические признаки - род, лицо, число, падеж, одушевленность, время, залог, вид глагола, степень сравнения.

При этом система способна автоматически распознавать язык (русский и английский), а также засчет контекста система производит устранение противоречий с омонимами.

ООО «Когнитивные системы»
ОГРН: 1165029057490; ИНН: 5029214276
Адрес: 141014, Московская обл, г.Мытищи,
ул. Веры Волошиной дом 12, офис 714;
office@cogsys.company





ООО «Когнитивные системы»
ОГРН: 1165029057490; ИНН: 5029214276
Адрес: 141014, Московская обл, г.Мытищи,
ул. Веры Волошиной дом 12, офис 714;
office@cogsys.company

Проведенные тесты показали хорошие результаты точности:

Нейромодель	**Русский язык**	**Английский язык**
Части речи (POS), ср.значение	93.5*	84.3*
Грамматические теги, ср.значение	84.9	79.2
Определяемые части речи	Всего - 15 adj, adp, adv, aux, cconj, det, intj, noun, num, part, pron, propn, sconj, sym, verb, x	Всего - 16 adj, adp, adv, aux, cconj, det, intj, noun, num, part, pron, propn, sconj, sym, verb, x
Определяемые теги у частей речи	Всего - 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'}}	Всего - 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}}

* Приведенные результаты точности были рассчитаны по модели исходя из 100% слов, имеющих 2 или более частей речи. В случае анализа смешанного текста, в котором присутствует только 50% слов с неоднозначными частями речи, точность для русского и английского языков составит около 96% для русского языка и 92,15 для английского языка.

ООО «Когнитивные системы»
ОГРН: 1165029057490; ИНН: 5029214276
Адрес: 141014, Московская обл, г.Мытищи,
ул. Веры Волошиной дом 12, офис 714;
office@cogsys.company

Было проведено сравнение точности работы моделей с результатам, полученными на аналогичной задаче командами в ходе соревнования по морфологической разметке естественных текстов MorphoRuEval-2017. Алгоритм Brain2Morph продемонстрировал лучшие результаты качества – 94.23, при разбросе результатов команд на смешанном тексте с 75.88 до 93.71 для смешанной оценки по частям речи и грамматическим тэгам. Стоит отметить, что в ходе данного соревнования оценивался не полный набор тэгов – так, время глаголов было упрощено до настоящего и «не настоящего», т.е. остальных.

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	94,23	2017

Протестировать альфа-версию модуля вместе с алгоритмом исправления ошибок в словах можно по следующей ссылке: http://cogsys.company/ru/brain2spell.