



Cognitive Systems

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

## **Brain2Spell**

Brain2Spell is a word-recognition service that is able to correct spelling mistakes and typos in words. Being one of the modules of the Big Linguistic Model, Brain2Spell algorithm is aimed at recognizing basic building blocks of the language – words.

Brain2Spell service functions on the basis of a neural network, trained on the corpus of 120 thousand words. Conducted tests showed high accuracy results - 90% on the test set of 25 thousand words with errors. We have compared the quality of the algorithm with the results obtained on a similar task by teams during the SpellRuEval-2016 contest. Brain2Spell algorithm demonstrated the best quality indicators on words with errors. It should be noted that at the moment the service recognizes such types of errors as orthographic mistakes (except for conjoint writing of words and particles), missing letters, extra letters and letters written in the wrong order.



## Cognitive Systems

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

Group	Algorithm name or author	Words in corpus	Incorrect Words in sample	Accuracy/Recall		Year	Source
				Words with errors	All types		
English words	Dronen	120 000	4 349		<b>0,92</b>	2016	1
	Hodge, Austin	81 717	600		<b>0,93</b>	2003	3
	Wikipedia ≥ 10			<b>0,78</b>		2013	9
	Wikipedia ≥ 8		1100	<b>0,75</b>		2013	9
	Wikipedia ≥ 6	42 622		<b>0,71</b>		2013	9
	Wikipedia ≥ 4			<b>0,68</b>		2013	9
	Ahmad, Kondrak	580 000	508		<b>0,79</b>	2005	6
Russian words	<b>Brain2Word</b>	<b>118 637</b>	1) 25 000 2) 660	<b>0,90</b>	<b>0,71*</b>	<b>2017</b>	-
	MS Word 2007			<b>0,88</b>		2015	8
	Yspell		56 000	<b>0,88</b>		2015	8
	Yandex		Not publicly available	<b>0,84</b>		2015	8
	Google			<b>0,80</b>		2015	8
	CIGR Lemming			<b>0,80</b>		2015	8
	Aspell			<b>0,70</b>		2015	8
	MSU*				<b>0,70</b>	2016	4
	Orfogrammatika*				<b>0,61</b>	2016	4
	MIPT*	10 000	1100		<b>0,58</b>	2016	4
	ISP RAS*	for tune up + dictionaries	publicly available		<b>0,56</b>	2016	4
	HSE*				<b>0,50</b>	2016	4
	NLP@Cloud*				<b>0,34</b>	2016	5
	InfoQubes*				<b>0,25</b>	2016	4
Dereza et al.	900 000	-		<b>0,40</b>	2016	8	

The algorithm was tested on two data sets:

1) 25 thousand words were selected from a set of 60 thousand words based on most common and relevant writing errors of people - from the research of D. Kulagin Word Map: Rethinking the Approach to Compiling Online Dictionaries in the Postmobile Era. Computer linguistics and intellectual technologies: proceedings of the international conference "Dialogue 2017", May 31 - June 3, 2017.

[http://cogsys.company/data/test\\_25k\\_errs.csv](http://cogsys.company/data/test_25k_errs.csv)



Cognitive Systems

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

2) The initial set of 1100 words with errors contained various kinds of generated mistakes – two words without spaces, scrambled letters etc. We used the dataset of 1100 words with errors only with orthographical mistakes and typos - proceedings of the Spellrueval Competition: The First Competition on Automatic Spelling Correction for Russian. Sorokin A. A., Baytin A.V., Galinskaya I.E., Shavrina T.O. "Dialogue-2016". June 1-4, 2016  
[http://cogsys.company/data/test\\_1200w\\_60-40.csv](http://cogsys.company/data/test_1200w_60-40.csv)

In the course of algorithm development, various options of combining the features and classes of the model were put to test to achieve the best result.









## Cognitive Systems

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

### Brain2Spell alfa RU/EN

supmre cort reaalvd new casse  [Check!](#)

supreme court revealed new case

reaalvd   

[Все](#) [Карты](#) [Картинки](#) [Видео](#) [Новости](#) [Ещё](#) [Настройки](#) [Инструменты](#)

Результатов: примерно 213 000 (0,49 сек.)

Показаны результаты по запросу **realvd**

Искать вместо этого [reaalvd](#)

reaalvd    [Найти](#)

[ПОИСК](#) [КАРТИНКИ](#) [ВИДЕО](#) [КАРТЫ](#) [МАРКЕТ](#) [НОВОСТИ](#) [ПЕРЕВОДЧИК](#) [ЕЩЁ](#)

[Check the fines for vehicle license plate](#)

[youfines.com](#) > [SD/REAALD](#)

Check the fines for vehicle license plate...

Нашёлся 1 результат

The same can be said about Microsoft Word error correction system, the examples of which in comparison with Brain2Spell are presented in the table below.

CORRECT	MISTAKE	MS_WORD	BRAIN2SPELL
Revealed	Reaavld	Repaved	Revealed
Premature	Prmeatruue	Prmeatruue	Premature
Panic	Painc	Painc	Panic
Changing	Changnig	Changnig	Changing
Issuing	Issiung	Issiung	Issuing
Cleaning	Claennig	Claennig	Cleaning
Thrilled	Thrliled	Thrliled	Thrilled



## Cognitive Systems

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

The following link may be used to test the alpha version of the Brain2Spell service - <http://cogsys.company/ru/brain2spell>.

The next steps in the creation of the Big Linguistic Model will be the development of an algorithm for determining morphological features of a word, and an algorithm for sentence parsing.