

**Я**ндекс

## **Коррекция грамотности**

Конференция «Диалог 2009»

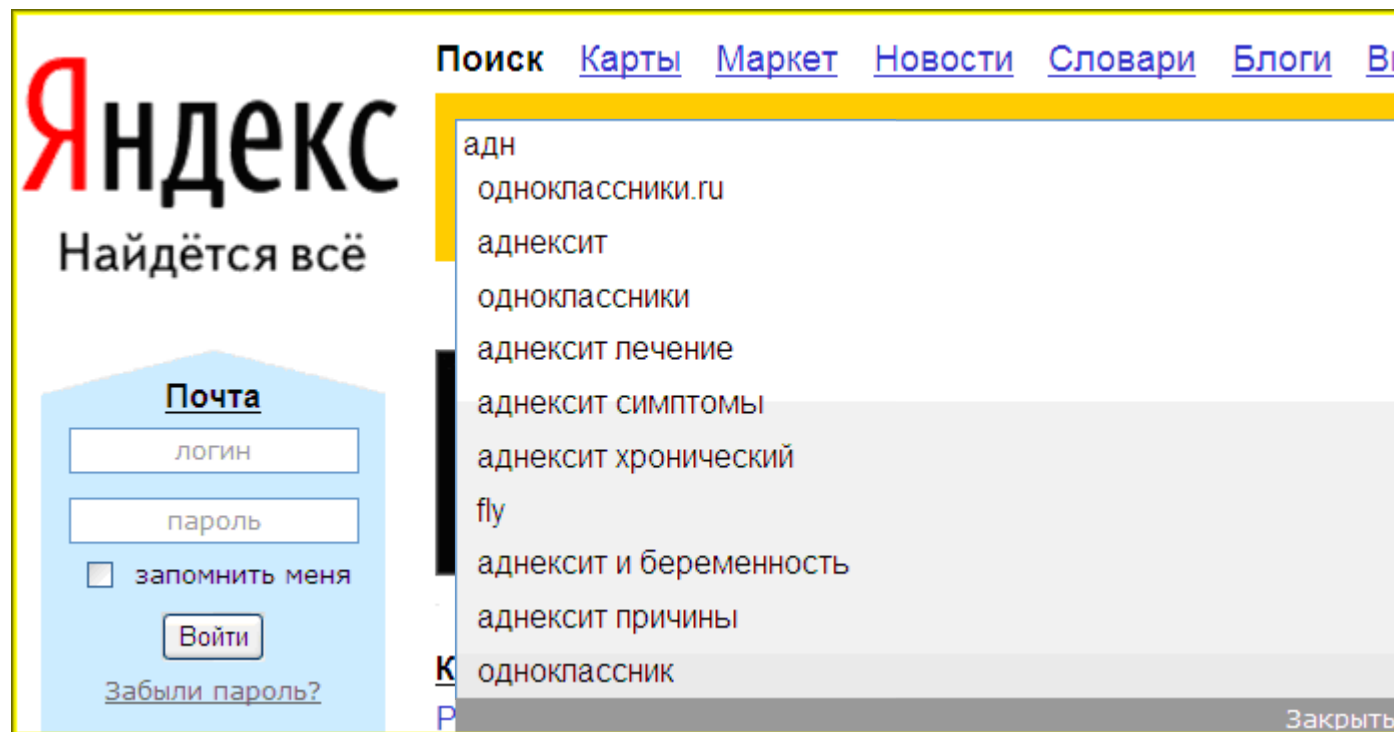
# Сервисы для проверки правописания

1. [Исправление опечаток](#) в поисковых запросах (**Я.Поиск**)
2. [Проверка правописания](#)
  - в почтовых сообщениях (**Я.Почта**)
  - в веб-формах (**Я.Бар**)
3. [Словари](#) русского языка  
**Я.Словари**

# Часть 1. Исправление запросов

Я

# Подсказка во время набора



Я

# Автоматическое исправление запроса

**Яндекс**  
Нашлось  
122 тыс. страниц

Упражнения для культуристов  
 в найденном  в Москве и области [расшир](#)

Результат поиска: страниц — **121 858**, сайтов — не менее **7 579**  
Статистика слов: упражнения — 42 198 426, для — 10 212 225 675, культуристов — 1 004 818.

В [исходном запросе](#) «Упаражнения для культуристов», возможно, есть опечатка.  
Результаты поиска для «Упражнения для культуристов».

**Яндекс**  
Нашлось  
13 тыс. страниц

чемпионат россии по хоккею 2009г итоговая таблица  
 в найденном  в Москве и области [расшир](#)

Результат поиска: страниц — **13 399**, сайтов — не менее **734**  
Статистика слов: чемпионат — 232 840 075, россии — 2 474 711 987, по — 9 835 516 836, хоккею — 86 569 959, 2009г — 43 102 261, итоговая — 27 408 255, таблица — 117 063 515.

По [исходному запросу](#) «чемпионат россии по хокею 2009г итоговая таблица» ничего не нашлось.  
Результаты поиска для «чемпионат россии по хоккею 2009г итоговая таблица».

## Подсказка после поиска

Поиск [Почта](#) [Карты](#) [Маркет](#) [Новости](#) [Словари](#) [Блоги](#)

**Яндекс**  
Нашлось  
108 тыс.  
страниц

агентство недвижимости

в найденном  в Москве и области

Результат поиска: страниц — 107 632, сайтов — не менее 3 725  
Статистика слов: агентство — 22 664 666, недвижимости — 538 250.

Быть может, вы искали: [«агентство недвижимости»](#)

Я

## Ошибки в запросах

- **70%**

- Орфографические (*прокуратруа, теливизор*)

- **25 %**

- Слитно-раздельное написание (*белыйгородок*)

- Контекстные ошибки (*меховой слон*)

- **5%**

- Английская раскладка клавиатуры (*crfxfnm*)

- Латинский алфавит (*lekarstva*)

# Ошибки – случайные и когнитивные

## Случайные vs. Когнитивные: 50 x 50

Отношение частоты ошибочного написания к частоте слова :

- $f(\text{скчать}) / f(\text{скачать}) = 1/900$
- $f(\text{однокласники}) / f(\text{одноклассники}) = 1/50$
- $f(\text{агенство}) / f(\text{агентство}) = 1/3$



## Примеры запросов с ошибками

- *юнность* кончилась внезапно мрЗ
- найти в Одессе *военое* училе*ще*
- в контакте *соце*альная сеть
- как из кукол *з*делать девочек группы ранетки
- *Денежний* знак или *обле*гация
- Фильм *Термен*атор 4 онлайн
- *Учим*ся футбольному *фрей*сталу
- взаимоотношения человека и природы в творчестве *постер*нака

# Слова чемпионы

Слово	Кол-во показов в день	Кол-во опечаток в день	Кол-во показов / Кол-во опечаток	Уникальных опечаток
<i>одноклассники</i>	213 641	<b>22 407</b>	10	<b>479</b>
<i>скачать</i>	<b>1 359 783</b>	15 986	<b>85</b>	296
<i>бесплатно</i>	782 733	11 775	66	346
<i>видео</i>	198 977	9 211	22	178
<i>онлайн</i>	431 616	8 260	52	248
<i>агентство</i>	17 240	6 861	<b>3</b>	54

Я

# Исправление запросов. Постулаты

- 1. Чаще пишут правильно**  
Частота слова в среднем на 2 порядка превосходит частоту опечатки
- 2. Все зависит от контекста**  
Корректность слова в тексте можно однозначно определить только по словарному окружению
- 3. Ошибки повторяются**  
Повторяемость ошибок очень высокая – ошибаются люди одинаково

## Исправление запросов. Источники

*Запросы содержат ВСЮ информацию, необходимую для исправления запросов*

- **частоты слов**
  - заменяем слово на более частотное
- **сочетаемость слов**
  - проверяем сочетаемость слова с соседними словами
- **замены слов**
  - смотрим как изменяют слово сами пользователи

# Исправление запросов. Вероятностная языковая модель

- Словарь двусловий

<i>мехавой</i>	17 000
<i>меховой</i>	90 000
<i>меховой салон</i>	48 000
<i>меховой слон</i>	45

- Функция близости (функция ошибки)

*крайсер* – это *крейсер* или *крайслер*?

- Словарь замен

*стреляца* → *стрелятся*

*мущина* → *мужчина*

- Обычные словари (для «надежности»)

Я

## Исправление запросов. Трудные случаи

### *Слова «похожие» на опечатку...*

- **Арфография** (диск группы Meldis)
- **Архидея** (фестиваль архитекторов)
- **Гринландия** (фестиваль авторской песни)
- **Одессей** (журнал)
- **Плотформа** (клуб в Сочи)

### *А как правильно?*

- **Предынсультный** или **прединсультный**
- **Приоритезация** или **приоритизация**
- **Контратенор** или **контртенор**

## Часть 2. **П**роверка орфографии

Я

# Я.Почта

The screenshot displays the Yandex Mail interface. On the left, a sidebar lists folders: **Отправленные** (75), **Удаленные** (—), **Спам** (—), **Черновики** (69), **Метки** (with settings and dropdown icons), **Непрочитанные** (0), **Важные** (0), **С вложениями** (with paperclip icon), and **Только мне**. The main area shows a header with **Кому** and **Тема** fields, and a toolbar with icons for bold, italic, underline, text color, undo, redo, bulleted list, numbered list, and link. The text area contains a paragraph with several words underlined in red, indicating a spell check. A popup menu is open over the word **иероглифов**, offering options: **иероглифов**, **Пропустить слово**, and **Пропустить все**. The text in the main area reads: "прочитать несколько общеобразовательных лекций для жен дипломатов, направляющихся в китай с миссией. Жены попались благодарные, интересующиеся...одна из них так восхищалась красотой иероглифов и их изяществом, что решила пр... искусству... Ей особо пригл... иероглифов, висящих на сте... занятия. Дама она была авт... попросила официанта снять картинку (что то типа рекламного листа), что и было сделано... Однако никто не предполагал, как далеко пойдет ее любовь к этим иероглифам! К банкету по случаю



# Я.Бар

Записать мысль - я.ру - Windows Internet Explorer

http://user-1594.ya.ru/posts\_add\_text.xml?ncmd=3943

Яндекс Я Найти

Записать мысль - я.ру

Заголовок

Кто увидит запись:

все

только друзья

только я

Комментировать разрешено

26 мая 2009 12:44

Отель от орода в получасе езды. А если хотите понырять с маской то пожайлуста 15 минут хотьбы в лево по бетонной дорожке к строящемуся отелю, там есть залив и вот там и глубина и кораллы и рыбы всяческие, даже на экскурсии на лодке со стеклянным дном таких не видели, просто супер!!!

Я

# Сервис **Я**.Спеллер

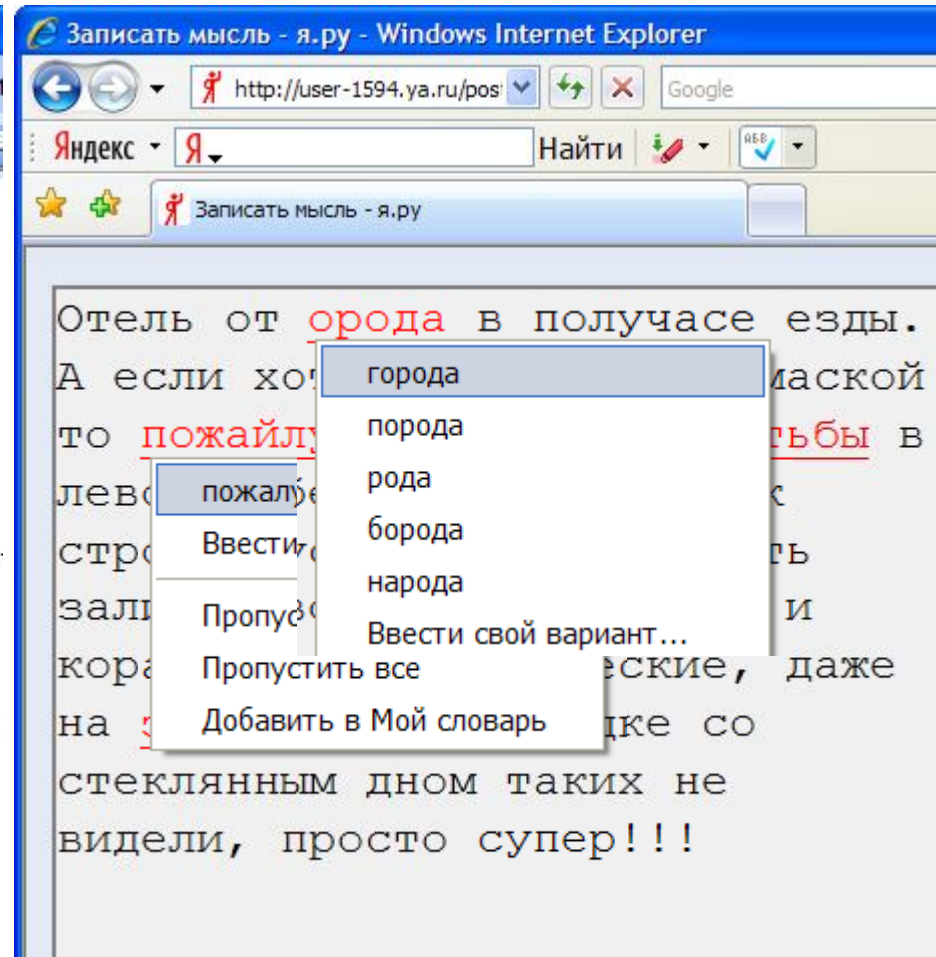
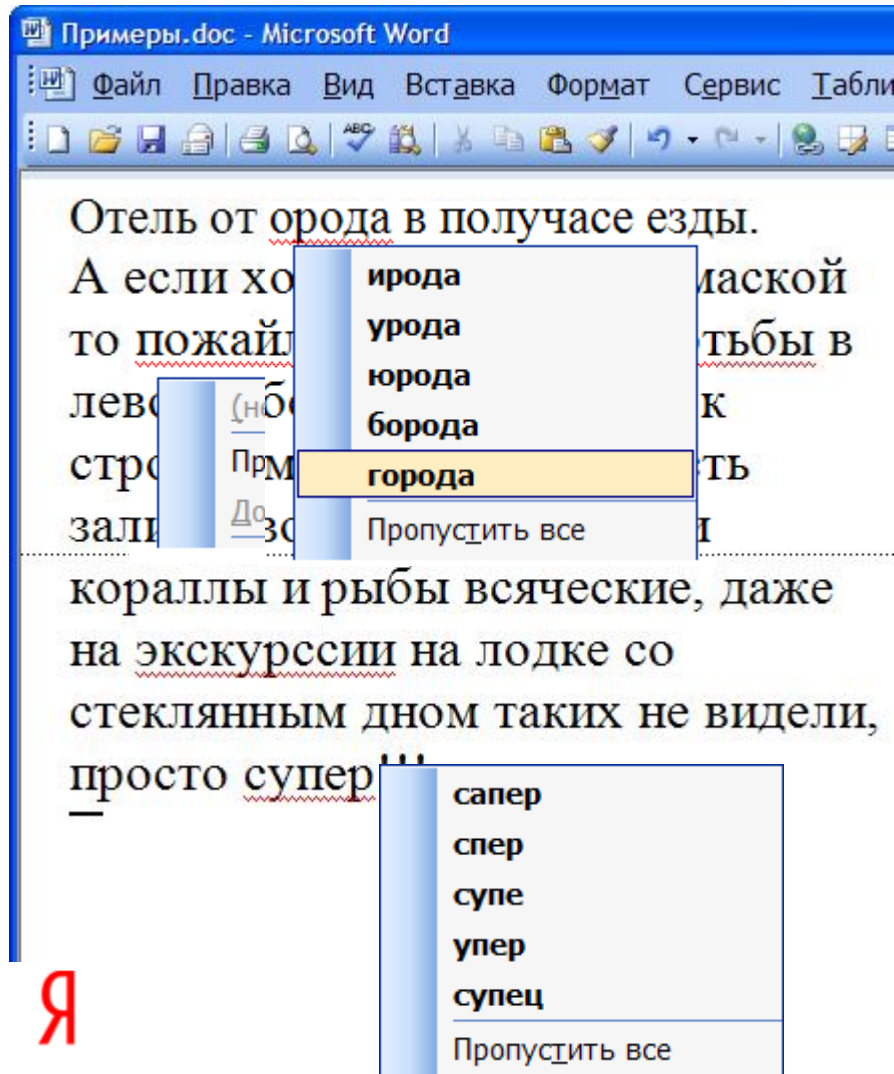
## Основы сервиса:

- **орфографический** словарь русского языка
- **вероятностная** языковая модель

## Особенности сервиса:

- **Большой** словарь русского языка  
(3.6М словоформ)
- Исправление **многобуквенных** ошибок
- **Точная** подсказка

# MS Word или Я.Спеллер?



# Часть 3. **С**ловари русского языка

Я

# Сервис Я.Словари

правписание

энциклопедии  русский язык  перевод [расширенный поиск](#)

Запрос исправлен на «**правписание**», так как для «праваписание» ничего не найдено.

## «правписание» в словарях русского языка

Морфемно-орфографический словарь

прав/о/пис/а́/ни/е [й/э]

Школьный этимологический словарь русского языка

**ПРАВОПИСАНИЕ**. Др.-рус. словообразоват. калька греч. *orthographia* (из *orthos* "правильный" и *graphō* "пишу"). См. [орфография](#).

Словарь синонимов русского языка

**ПРАВОПИСАНИЕ** (*только ед.*), **орфография** (*только ед.*).

Общепринятая система правил написания слов какого-л. языка.

## **С**ловари русского языка

1. **Морфемно-орфографический словарь / Тихонов А. Н.**
2. **Школьный этимологический словарь русского языка. Происхождение слов / Шанский Н. М., Боброва Т. А.**
3. **Словарь синонимов русского языка / Зимин В. И. и др.**
4. **Словарь русских синонимов и сходных по смыслу выражений / Абрамов Н.**
5. **Современное дополнение к «Словарю русских синонимов» Н. Абрамова**
6. **Русское словесное ударение / Зарва М. В.**
7. **Словарь русских фамилий / Никонов В. А.**
8. **Русские фамилии: популярный этимологический словарь / Федосюк Ю. А.**

## Планы

- **Служба Я.Правописание**
  - Проверка орфографии
  - Грамматическая справка
- **Новые словари русского языка**
  - Орфографический словарь
  - Грамматический справочник
  - Фразеологический словарь
  - Словарь антонимов