

Вязников Максим Николаевич

**Разработка и исследование
методов повышения стойкости к
автоматическому распознаванию
изображений для тестов Тьюринга**

научный руководитель

доцент, к.т.н.

Ржеуцкая Светлана Юрьевна

Больше 7.5 млн. сайтов используют CAPTCHA*

* по данным SimilarTech <https://www.similartech.com/categories/captcha>

	Technology	Websites
1	 General Captcha	5,637,208
2	 reCAPTCHA	2,176,562
3	 WP reCAPTCHA	33,622
4	 KeyCAPTCHA	7,765
5	 Sweet Captcha	4,877
6	 AreYouAHuman	3,196
7	 NuCaptcha	532

Примеры символно-графического CAPTCHA



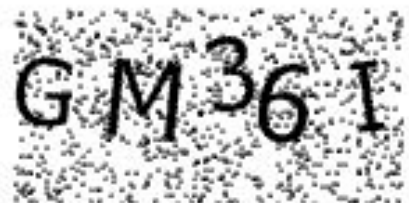
N°1



N°2



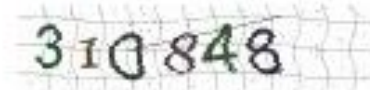
N°3



N°4



N°5



N°6



Слабые стороны программ автоматического распознавания:

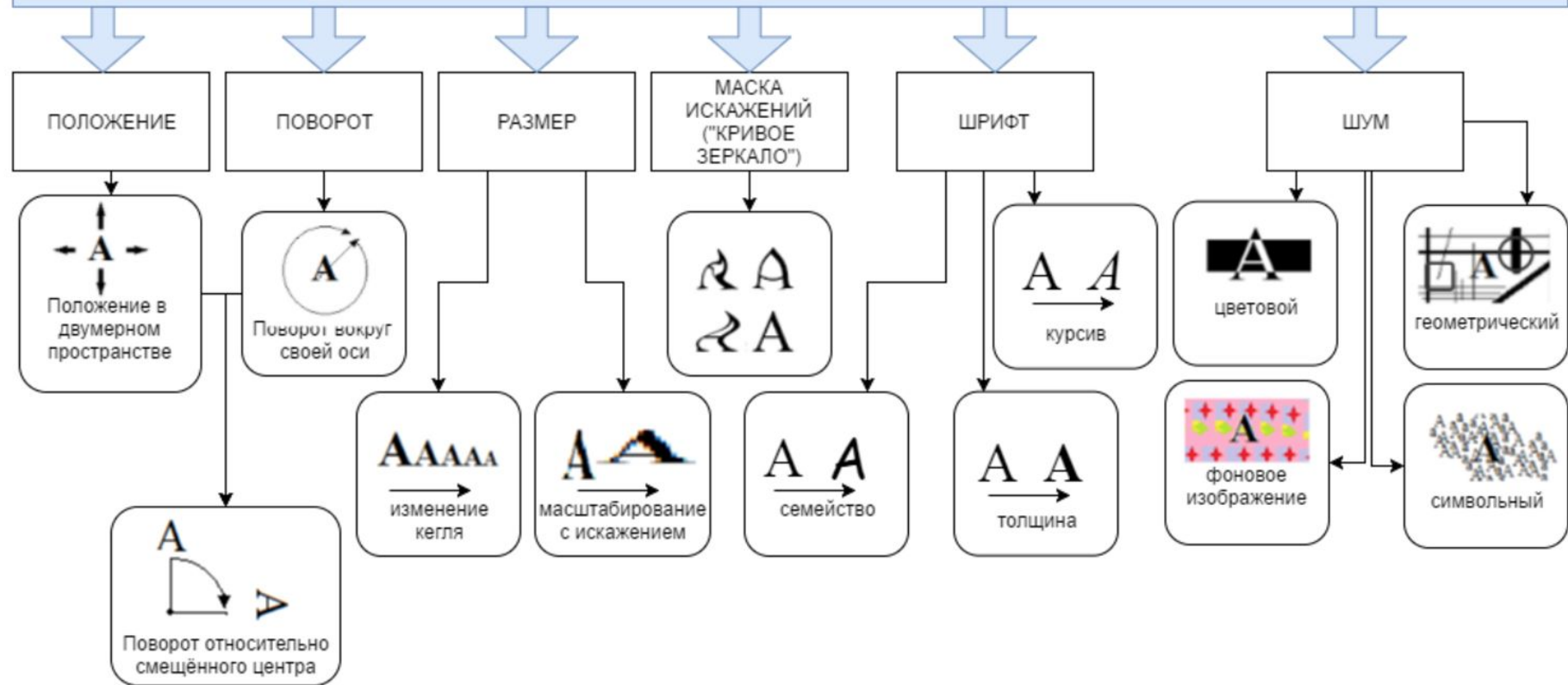
Очистка изображений от шума
(компенсация после распознавания символов и анализ)

Сегментация изображения
(поиск по словарю)

Таким образом формируется строка-ответ на тест CAPTCHA

Разбор каждого символа отдельно
(компенсация шума слипшихся после распознавания каждого)

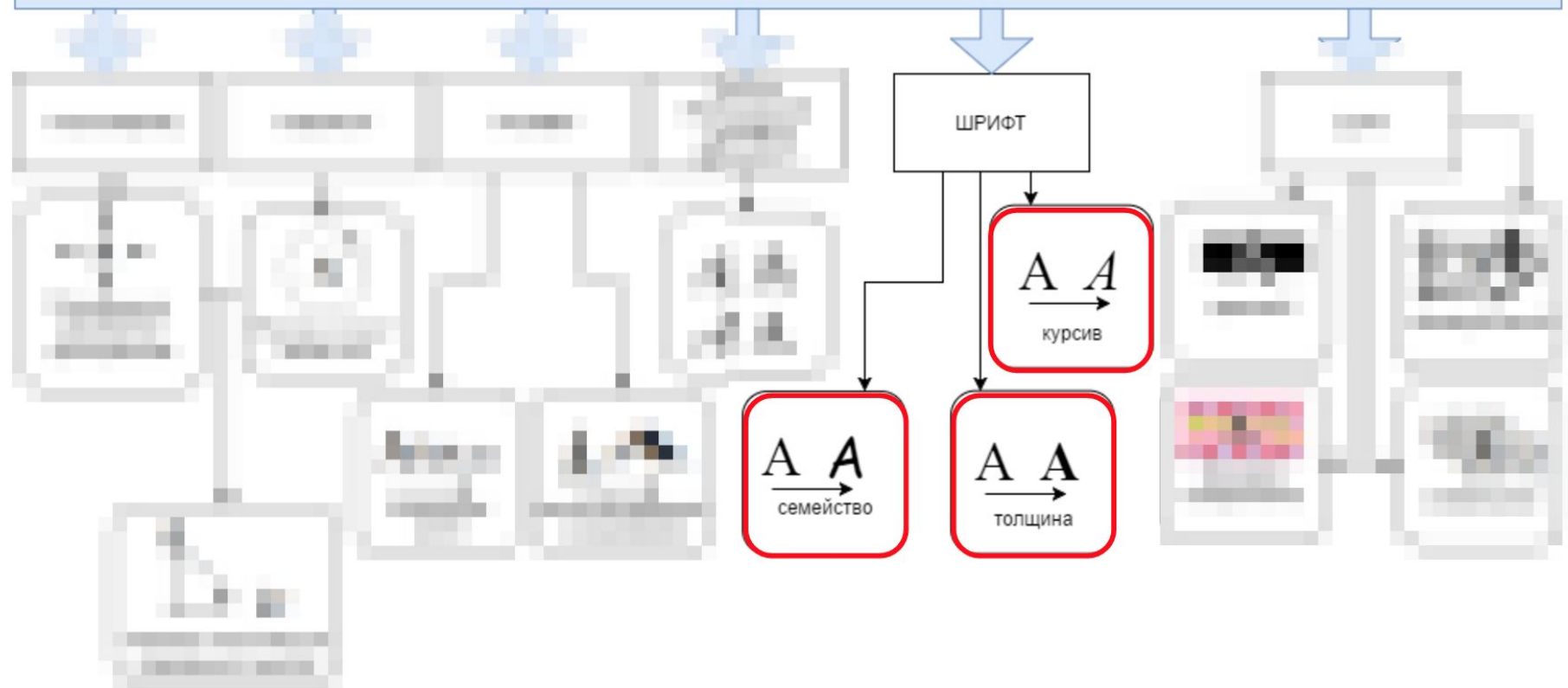
КЛАССИФИКАЦИЯ ИСКАЖЕНИЙ СИМВОЛЬНО-ГРАФИЧЕСКОГО САРТСНА



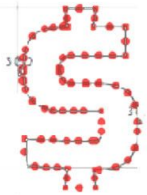




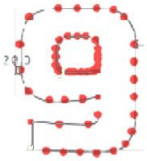
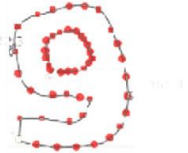
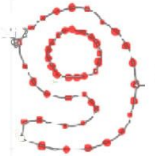







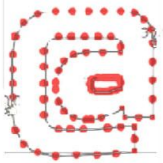


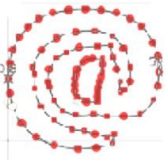
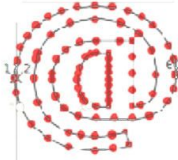
Разработана методика создания алгоритмов генерации САРТСНА

- коэффициенты искажений/шума выбираются случайным образом в заданных рамках
- случайное количество символов
ещё лучше, когда символы разной ширины
- наличие шумов, характеристики которых соотносятся с очертаниями символов
- разные шрифты?
желательно контурные

КЛАССИФИКАЦИЯ ИСКАЖЕНИЙ СИМВОЛЬНО-ГРАФИЧЕСКОГО САРТСНА



Пример морфинга

100% шрифт №1	75% шрифт №1 25% шрифт №2	50% шрифт №1 50% шрифт №2	25% шрифт №1 75% шрифт №2	100% шрифт №2
				
				
				
				

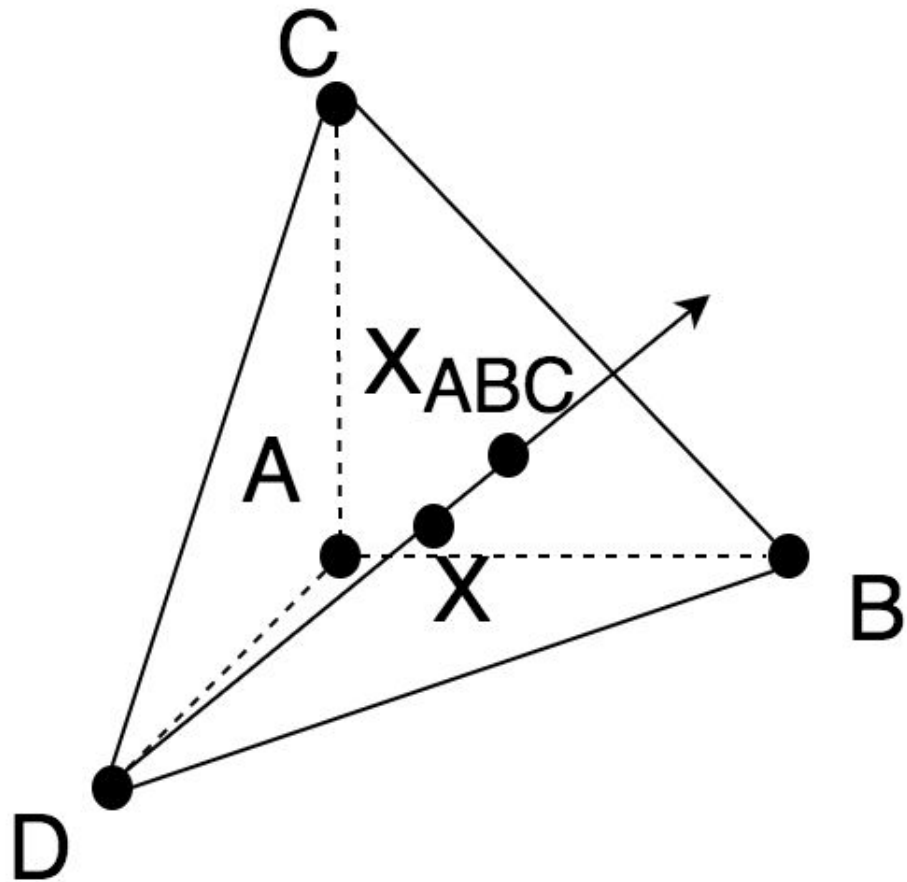
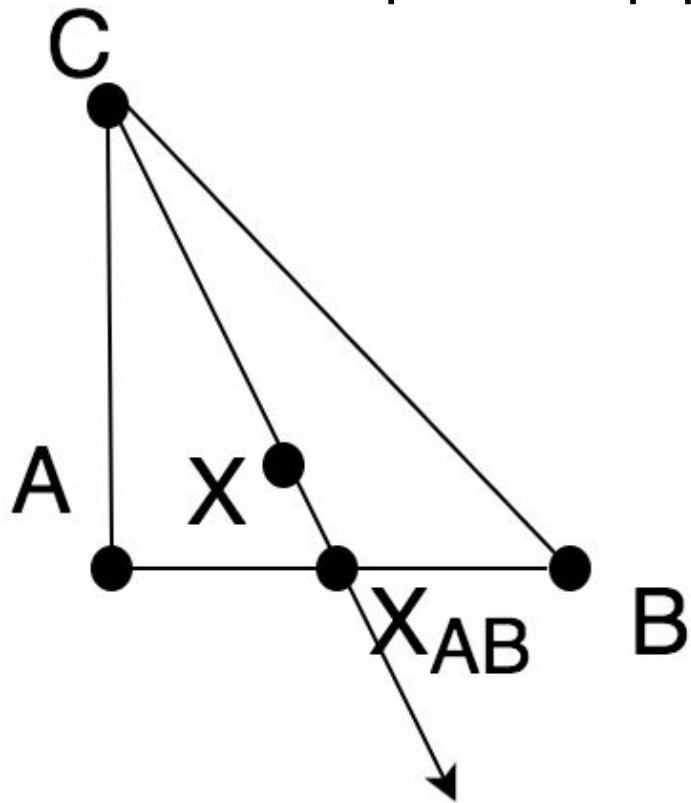
Распознавание нового шрифта нейронной сетью

- Шрифты Times New Roman и Georgia очень похожи между собой. НС при различных настройках распознавала отдельные символы. Значит, смогла бы распознать их все.
- В аналогичных опытах с новыми шрифтами полученными из шрифтов отличных от Georgia и Times New Roman, НС ничего не распознавала
- В зависимости от настроек, морфинг может оказывать влияние на распознаваемость символов нейронными сетями

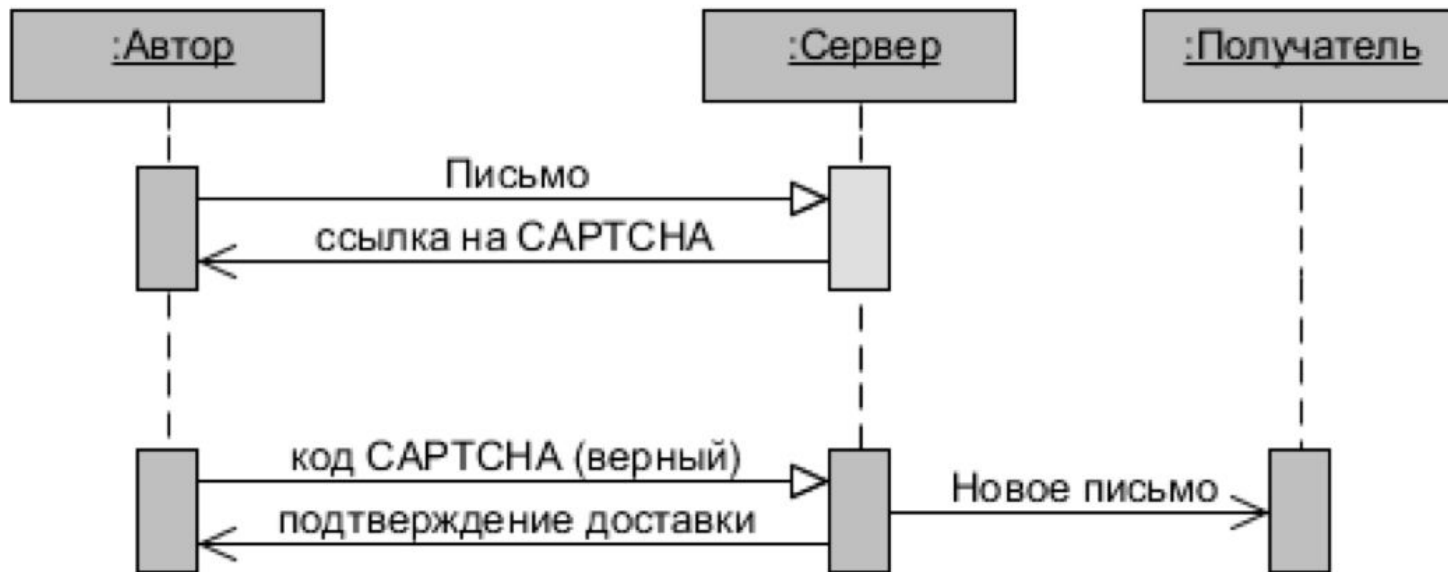


Морфинг - это ещё одно искажение

Алгоритм морфинга нескольких шрифтов



Использование CAPTCHA в качестве дополнительной меры борьбы со спамом



Итоги кратко

- классификация различных искажений для символьно-графического CAPTCHA
 - морфинг шрифтов как новый способ искажения СИМВОЛОВ (патент № 2608262)
 - использование CAPTCHA в качестве дополнительной меры борьбы со спамом
- вопросно-ответная система