

Е.Е. МАКАРОВА (КАНУНОВА),
Л.В. АНТОНОВ

**Информационная система
графологического анализа текста**

УДК 004.932.4:004.912

Муромский институт
(филиал) ФГБОУ ВПО
«Владимирский
государственный
университет имени
Александра
Григорьевича и Николая
Григорьевича
Столетовых», г. Муром

В статье рассматриваются вопросы автоматизации процесса графологического анализа текста и формирования психологического портрета личности. Описана последовательность этапов обработки и анализа текстовых рукописей. Приведены результаты работы информационной системы.

In article questions of automation of process of the graphological analysis of the text and formation of a psychological portrait of the person are considered. The sequence of stages of processing and the analysis of text manuscripts is described. Results of work of information system are resulted.

Целью работы является разработка комплексной системы графологического анализа текста, позволяющей пользователю получить достоверную информацию о характере человека. В ходе разработки необходимо было выполнить ряд задач: проанализировать предметную область; выявить ряд правил и закономерностей, с помощью которых разрабатываемая система будет предоставлять достоверный результат; разработать алгоритм анализа рукописных текстов.

Графология — область знания о почерке и методах его исследования с точки зрения отражающихся в нем психических состояний и особенностей личности пишущего. Различные характеристики почерка дают в совокупности ценную информацию о темпераменте человека, его характере, состоянии в момент написания, отношении

к предмету и содержанию написанного. Психологический анализ почерка предполагает учет всех известных сведений о данном индивиду и обстоятельствах написания анализируемого текста. Фактически, графологический метод позволяет определить основные качества человека по определённым критериям.

Сферы применения графологии весьма обширны. Графология в сфере занятости и бизнесе представляется анализом информации, собираемой при приёме кандидатов на работу. Клиническая графология занимается выявлением различных психоэмоциональных и психических расстройств, таких как личностные, пограничные, неврозы и психозы - графологическим синдромам. Медицинская графология занимается исследованием связей между различными заболеваниями и признаками почерка. Криминологическая графология исследует почерк преступников и осуждённых, обвиняемых по различным преступлениям, начиная с административных и коррупционных, и заканчивая различными видами агрессии и насилия. Судебная графология предлагает приемы для идентификации или сравнения документов (почерков, автографов), для судебного рассмотрения разногласий по поводу подписи, либо завещания, почерка и цифр, фигурируемых на чеках и т.д [1].

Входными данными системы является изображение рукописного документа (РД) в формате *bmp или *jpeg. Изображение РД имеет повреждения и дефекты, такие как перегибы вертикальные и горизонтальные, слабый контраст, серый фон. Немаловажным является отсутствие чёткой связной линии между буквами, так как это может привести к неверной обработке текста.

Поэтому первым этапом обработки изображения является автоматизированное устранение перечисленных выше дефектов. Для устранения дефекта в виде вертикальных и горизонтальных перегибов применяется метод устранения линейных дефектов в виде перегибов (складок) на основе операций математической морфологии[2]. Слабый контраст восстанавливается с использованием метода восстановления контраста изображения текстового документа (ТД) с использованием опорных точек [3]. Дефект серого фона устраняется на основе метода восстановления неравномерной фоновой составляющей изображения ТД [4].

Следующим этапом обработки является разбивка документа на отдельные слова. С этого момента документ рассматривается с точки зрения геометрии написания. Определяются следующие параметры (Таблица 1):

- Процент пространства отступа сверху от общей площади рукописи;
- Процент пространства отступа слева от общей площади рукописи (вверху и внизу);
- Процент пространства отступа справа от общей площади рукописи (вверху и внизу);
- Градусная величина наклона строк (минимальный, максимальный и средний показатель)
- Величина высоты строк в собственной градации (минимальный, максимальный и средний показатель)
- Высота символов в процентах от высоты строк (минимальный, максимальный и средний показатель)
- Расстояние между словами в процентах от размера символов (минимальный, максимальный и средний показатель)
- Общий наклон почерка (минимальный, максимальный и средний показатель).

Пользователь, опираясь на информацию о геометрии страницы и особенностей написания, заполняет значения правил системы.

Заключительным этапом обработки является составление психологического портрета личности на основе перечисленных выше параметров и особенностей написания текста и отдельных букв. На основе базы эталонных букв и набора соответствующих характеристик происходит процесс распознавания и отождествления каждой буквы РД. Затем формируется результат.

Предположим, что загруженный нами текст, имеет прямой наклон, узкие со всех сторон поля; почерк – средний и легко наклонён вправо; расстояние между словами пропорционально; нажим на ручку испытуемого – средний (рисунок 1).

3.4) Эксперименты показывают, что более понятными изображениями являются изображения с непрерывными границами между объектом и фоном. Такие изображения субъективно лучше и применяются в фотографии и телевизионных передачах.

Для того чтобы уменьшить яркость границ, оставив яркость картинки в целом, ее нужно свести.

Одним из способов порегулировки границ является статистическое дифференцирование. Это один из способов минимизации яркости изображения. Для этого исходное изображение $f(x_i, y_i)$ для построения минимизации изображения делаем из $\sigma(x_i, y_i)$:

$$g(x_i, y_i) = f(x_i, y_i) / \sigma(x_i, y_i)$$

$$\text{где } \sigma(x_i, y_i) = \sqrt{\sum_{i,j \in N(i,j')} (f(x_i, y_i) - f(x_j, y_j'))^2}, \quad i, j \in N(i, j')$$

где $f(x_i, y_i)$ - исходное изображение представляется в виде двумерной матрицы. Матрица $\sigma(x_i, y_i)$ является матрицей исходного изображения с усредненными значениями яркости.

Рис. 1. Пример рукописного документа

Для данной рукописи система определяет следующую личностную характеристику: “Вы тёплая, дружеская натура и открыто выражаете свои чувства. Однако Вы не очень импульсивная натура и можете держать свои эмоции под контролем. Вы можете ладить с другими людьми, потому что в Вашем характере существует искренняя тяга к общению, и Вы не будете счастливы, если будете долго находиться в одиночестве. Вы не любите скрывать свои чувства и хотите, чтобы предмет Ваших чувств отвечал Вам взаимностью. Вас нелегко расстроить мелкими событиями или заявлениями других людей. Вы тщательно обдумываете свои мысли и поступки и удерживаете свои настроения под контролем. Вы являетесь весьма бережливой натурой. У Вас есть хороший вкус и инстинктивная тяга к изысканным предметам, а стало быть, и желание потратить деньги для приобретения таковых. Ваши временные всплески расточительности быстро подавляются внутренним призывом к бережливости. Порой у Вас возникают приливы щедрости. Однако Вы не способны сохранять верность таким обязательствам и Вас одолевает страсть к бережливости. Вы практичный, спокойный, рассудительный, но с богатым творческим потенциалом человек.”

Система была протестирована на ряде испытуемых, выбранных из разных общественных слоёв, а так же максимально охватывающих возрастной диапазон. Объем тестирующей выборки составил 55 рукописей. В результате обработки 55 рукописных текстов были выявлены 10 распространенных типов личности.

Таблица 1

Параметры рукописного документа для выборки из 10 испытуемых

№ Испытуемого Признак	1	2	3	4	5	6	7	8	9	10
Верхнее поле (%)	4,33	3,86	3,76	1,91	5,73	6,68	12,68	17,66	6,52	4,84
Левое поле (вверху) (%)	10,57	6,34	7,51	4,94	11,74	6,44	9,14	5,46	9,19	8,83
Левое поле (внизу) (%)	5,63	4,62	6,82	19,56	12,59	4,72	0,69	6,7	11,47	8,62
Правое поле (вверху) (%)	5,95	6,2	4,41	5,34	6,98	3,63	4,77	8,57	6,37	11,54
Правое поле (внизу) (%)	0,77	1,74	4,74	7,78	9,78	6,6	5,01	2,28	0,35	-0,89
Наклон строк (минимальное) (°)	-5,15	-0,87	-2,25	0	-0,11	-0,34	-0,86	-1,26	-1,66	0
Наклон строк (максимальное) (°)	1,65	0,72	2,38	2,18	1,09	0,57	0,39	0	0	1,64
Наклон строк (среднее) (°)	0,12	0,2	-0,8	1,23	0,67	0,11	-0,23	-0,37	0,59	0,65
Высота строк (минимальное)	62	95	48	80	82	81	124	48	82	83
Высота строк (максимальное)	336	143	258	242	116	167	431	218	239	262
Высота строк (среднее)	136	125,2 5	123,3 6	157,4	96,44	102,8 6	277,5	122,3 3	141,6 7	159,6
Высота символов (минимальное) (%)	1,79	4,69	4,49	3,19	5,41	3,59	1,39	3,21	2,51	2,29
Высота символов (максимальное) (%)	212,9	118,9 5	314,5 8	73,94	65,85	118,8 9	110,4 8	162,5	132,9 3	156,6 3
Высота символов (среднее) (%)	39,63	40,56	46,59	28,26	25,27	39,97	13,25	29,58	40,39	29,9
Расстояние между словами (минимальное) (%)	14,84	7,87	13,92	8,99	8,21	17,03	5,44	5,53	10,49	16,77
Расстояние между словами (максимальное) (%)	807,1 6	490,0 9	506,2 8	242,7 8	574,4	389,1 5	160,4 6	433,9 2	246,4	901,2 1
Расстояние между словами (среднее) (%)	123,6 4	137,3 8	95,71	88,35	137,8 6	99,34	41,73	103,1 4	71,62	95,61
Наклон почерка (минимальное) (°)	48,51	53,14	57,18	59,07	45,99	45,33	49,66	68,31	50,69	47,6
Наклон почерка (максимальное) (°)	114,8 6	124,4 9	96,94	127,1 7	115,7 4	121,2 3	128,0 2	117,1 8	109,7 9	118,2
Наклон почерка (среднее) (°)	79,32	94,32	72,07	85,85	68,67	83,4	83,61	89,24	78,84	73,81

Для оценки корректности работы системы, все испытуемые прошли также ряд традиционных психологических тестов, включающих тест Холланда и методику Резапкиной, а также собеседование с психологом. Результаты, обработанные и предоставленные психологом, практически полностью совпадают с выходными данными системы анализа почерка. Вероятность ошибки составляет 0,04% и при большем объеме выборки будет стремиться к нулю.

Литература

1. Инесса Гольдберг Почерк и типы личности. Учебник – М. У-Фактория. 2011, 234 с.
2. Канунова Е.Е. Алгоритм поиска и устранения дефектов в идее перегибов на изображениях текстовых документов [Текст]/ Канунова Е.Е., Полякова Е.В. //Алгоритмы, методы и системы обработки данных, 2008. № 13. С. 58-62
3. Садыков, С.С. Автоматизированная реставрация изображений архивных текстовых и фотографических документов [Текст]/ С.С. Садыков, Е.Е. Канунова, А.Д. Варламов //Автоматизация и современные технологии. 2007. №8. С. 10-12.
4. Садыков С.С. Алгоритмы устранения дефектов на архивных рукописных материалах [Текст]/ Садыков С.С., Канунова Е.Е.// Труды международного симпозиума «Надежность и качество», 2005 Т.1. С. 46-48

E-MAIL: ASP_KAT@RAMBLER.RU