**поиск характерных чтений**

За основу были взяты кластеры, выделенные вручную по матрице сходства, в состав этих кластеров вошли все рукописи кроме F.п.I.6 (т.е. кластеры, представленные в документе ).

Результаты поиска характерных чтений представлены в трех таблицах.

В таблице **all\_features\_no\_F.п.I.6.xlsx** два листа, оба из которых представляют собой таблицу разночтений. При этом на первом листе выделены синим цветом характерные чтения для больших основных кластеров, а на втором листе выделены характерные чтения для подкластеров.

В таблице **mc\_features\_no\_F.п.I.6.xlsx** каждый лист отведен под один из основных кластеров и характерные для этого лишь кластера чтения. То же сделано и для подкластеров, а результаты представлены в таблице **sc\_features\_no\_F.п.I.6.xlsx.**

**Алгоритм поиска характерных чтений**

По каждому узлу разночтений для каждого кластера мы вычислили процент сходства. Например, для больших кластеров, коих было выделено 4, узлы разночтений получили следующее представление:

0: [0, 0, 0, 5], 1: [2, 20, 12, 41], 2: [97, 80, 87, 50], 3: [0, 0, 0, 2]

Здесь мы видим номера всех вариантов, которые образуют данный узел разночтений. Для каждого варианта приводится массив, который содержит процент сходства по данному варианту в соответствующем кластере. Например, в этом узле по варианту 2 процент сходства в первом кластере равен 97.

Следующий шаг – непосредственно отбор характерных чтений. Из всех чтений выбирались лишь те, что удовлетворяют двум условиям:

* процент сходства по данному чтению в некотором кластере должен быть больше либо равен 75
* процент сходства по данному чтению в остальных кластерах должен быть меньше 75.

**Результаты и корректировка выделения кластеров**

|  |  |
| --- | --- |
| номер кластера | номера узлов разночтений |
| 1 | [233, 304] |
| 2 | [] |
| 3 | [93] |
| 4 | [91, 173, 329] |
| 5 | [3, 119, 221, 247] |
| 6 | [66, 97, 98, 109, 110, 114, 115, 118, 128, 160, 320, 323] |
| 7 | [4, 10, 15, 16, 29, 80, 90, 154, 155, 156, 157, 167, 169, 186, 203, 213, 215, 228, 252, 293, 294, 295, 298, 306, 313, 327] |
| 8 | [131] |
| 9 | [40, 69, 72, 74, 75, 79, 87, 89, 101, 123, 146, 190, 245, 246, 247, 248, 268, 270, 305] |
| 10 | [5, 6, 210, 212, 216] |
| 11 | [48, 49, 50, 51, 53, 54, 55, 56, 58, 59, 60, 61, 62, 63, 117, 136] |

При первой итерации не для всех подкластеров удалось найти характерные чтения. Одним из таких кластеров оказался второй подкластер:

F.п.I.57.s ф.304.III.2.s ф.256.116.s F.п.I.20.s F.п.I.11.s F.п.I.17.s ф.37.17.s ф.256.112.s Син.67.s Син.68.s ф.354.1.s F.п.I.61.s ф.98.49.s.

При более внимательном изучении оказалось, что периферия кластера (а именно рукописи F.п.I.11.s F.п.I.17.s ф.37.17.s ф.256.112.s Син.67.s Син.68.s ф.354.1.s F.п.I.61.s ф.98.49.s) ближе по средним процентам сходства к первому кластеру. Если эти рукописи мы отнесем к первому подкластеру, то второй подкластер (F.п.I.57.s ф.304.III.2.s ф.256.116.s F.п.I.20.s) приобретает гораздо больше характерных чтений (11 чтений по сравнению с 1), но при этом у первого кластера оказывается всего 1 характерное чтение.

|  |  |
| --- | --- |
| номер кластера | номера узлов разночтений |
| 1 | [304] |
| 2 | [82, 94, 95, 120, 137, 189, 190, 217, 233, 270, 323] |
| 3 | [93] |
| 4 | [91, 173] |
| 5 | [3, 221, 247] |
| 6 | [66, 109, 110, 114, 115, 118, 320, 323] |
| 7 | [4, 10, 15, 16, 29, 80, 90, 154, 155, 156, 157, 167, 169, 186, 213, 215, 228, 252, 293, 294, 295, 298, 306, 313, 327, 337] |
| 8 | [131] |
| 9 | [69, 72, 74, 75, 79, 87, 89, 101, 123, 146, 190, 245, 246, 247, 248, 268, 270, 305] |
| 10 | [5, 6, 210, 212, 216] |
| 11 | [48, 49, 50, 51, 53, 54, 55, 56, 58, 59, 60, 61, 62, 63, 117, 136] |

1 характерное чтение выделяется и для третьего кластера: F.п.I.82.s Гильф.1.s ф.17.1189.s Син.65.s ф.256.109.s ф.256.110.s. Но данные рукописи будет правильнее назвать периферийными для первого кластера, ведь процент сходства в данном третьем кластере равен всего лишь 78.6, тогда как средний процент сходства с первым кластером равен 89. Строго говоря, рукописи третьего кластера как таковой кластер не образуют, а лишь присоединяются к первому кластеру.

При попытке объединить первый отредактированный подкластер с третьим получаем следующие результаты. Третий подкластер объединился с первым, при этом количество характерных чтений для первого подкластера не изменилось. Общее же число кластеров логично уменьшилось на один.

|  |  |
| --- | --- |
| номер кластера | номера узлов разночтений |
| 1 | [304] |
| 2 | [82, 94, 95, 120, 137, 189, 190, 217, 233, 270, 323] |
| 3 | [91, 173] |
| 4 | [3, 221, 247] |
| 5 | [66, 109, 110, 114, 115, 118, 320, 323] |
| 6 | [4, 10, 15, 16, 29, 80, 90, 154, 155, 156, 157, 167, 169, 186, 213, 215, 228, 252, 293, 294, 295, 298, 306, 313, 327, 337] |
| 7 | [131] |
| 8 | [69, 72, 74, 75, 79, 87, 89, 101, 123, 146, 190, 245, 246, 247, 248, 268, 270, 305] |
| 9 | [5, 6, 210, 212, 216] |
| 10 | [48, 49, 50, 51, 53, 54, 55, 56, 58, 59, 60, 61, 62, 63, 117, 136] |

**Поиск характерных чтений для двух групп: группа отрывков десятой недели и группа отрывков Великого понедельника**

Стало интересно посмотреть, какие чтения разделяют две этих группы. Тот же алгоритм поиска характерных чтений мы запустили на материале двух больших групп. Результаты представлены в двух следующих таблицах: **all\_features\_passages.xlsx** (характерные чтения выделены цветом), **passages\_features.xlsx** (для каждой группы отдельно приведены характерные для нее чтения).

56 характерных чтения выделилось для отрывков 10 недели. 37 характерных чтений выделилось для отрывков Великого понедельника.