**Кластер 0 (3%)**

Медиана количества слов 11, среднее количество слов 19.

В 15% есть логи/архивы. В 7% xml. В 15% есть Stack Trace, в 21% Java Code. 3% STR, 5% EOB.

10% лексики совпадает с важнейшей лексикой старого кластера[[1]](#footnote-1) (0 STR + stack trace) во всех дескрипшнах этого кластера.

Л10% (6 STR) в 54% дескрипшнов этого кластера[[2]](#footnote-2),

Л10% (10 EOB) в 52% дескрипшнах этого кластера (ДЭК).

Л10% (12 log/zip+xml) в 91% ДЭК.

Аномалий не обнаружено.

**Кластер 1 (6%)**

Медиана количества слов 14, среднее количество слов 20.

В 13% xml. В 20% есть Stack Trace, в 16% Java Code.

Л10% (3 xml + stack trace) во всех ДЭК.

Аномалий не обнаружено.

**Кластер 2 (9%)**

Медиана количества слов 16, среднее количество слов 22.

3% картинки, 7% xml, 3% STR, 1% EOB. 3% Java Code.

Аномалия - Л10% (11 STR + xml) во всех ДЭК.

**Кластер 3 (6%)**

Медиана количества слов 13, среднее количество слов 16.

В 18% есть логи/архивы. В 13% xml. В 8% есть Stack Trace, в 10% Java Code. 5% STR, 1% EOB.

Л10% (12 log/zip+xml) в 98% ДЭК.

Аномалия - Л10% (6 STR) во всех ДЭК.

**Кластер 4 (2%)**

Медиана количества слов 37, среднее количество слов 35.

Логи/архивы 5%, картинки 2%, xml 7%, STR во всех ДЭК, EOB в 47%, Java Code 3%.

Л10% (9 STR + EOB) в 40% ДЭК, Л10% (10 EOB) в 37% ДЭК. Л10% (11 STR + xml) в 26% ДЭК.

['reproduce','step'] в 71% ДЭК. ['reproduce','click'] в 13% ДЭК.

Аномалий не обнаружено.

**Кластер 5 (8%)**

Медиана количества слов 17, среднее количество слов 21.

Картинки 4%, xml 5%, Java Code 9%.

Л10% (12 log/zip+xml) во всех ДЭК. Л10% (11 STR + xml) в 31% ДЭК.

Аномалий не обнаружено.

**Кластер 6 (22%)**

Медиана количества слов 17, среднее количество слов 25.

Ничего нет.

Аномалий не обнаружено.

**Кластер 7 (6%)**

Медиана количества слов 16, среднее количество слов 24.

Картинки 2%, xml 6%, STR 2%, EOB 9%, Java Code 4%.

Л10% (9 STR + EOB) 24%.

Аномалия - Л10% (10 EOB) во всех ДЭК.

**Кластер 8 (5%)**

Медиана количества слов 20, среднее количество слов 26.

Логи/архивы во всех ДЭК, Stack Trace во всех ДЭК. Картинки 2%, xml 25%, STR 5%, EOB 1%, Java Code 10%.

Л10% (8 log/zip + xml+ stack trace) - 20%.

Л10% (10 EOB) - 19%.

Л10% (12 log/zip+xml) - 22%.

Л10% (5 log/zip + stack trace) - 17%.

Аномалий не обнаружено.

**Кластер 9 (6%)**

Медиана количества слов 21, среднее количество слов 30.

Логи/архивы во всех ДЭК. Картинки 4%, xml 16%, STR 7%, EOB 2%, Java Code 9%.

Л10% (11 STR + xml) - 40%.

Л10% (12 log/zip+xml) - 34%.

Л10% (10 EOB) - 26%.

Аномалий не обнаружено.

**Кластер 10 (2%)**

Медиана количества слов 26, среднее количество слов 30.

Логи/архивы 20%, xml 6%, STR во всех ДЭК, EOB 92%, Java Code 22%, Stack Trace 42%.

Л10% (12 log/zip+xml) - 72%.

Л10% (8 log/zip + xml+ stack trace) - 60%.

Л10% (0 STR + stack trace) - 38%.

Л10% (9 STR + EOB) - 34%.

Л10% (10 EOB) – 26%.

Л10% (11 STR + xml) – 16%.

['fails','test'] – 10%.

['reproduce','step','result'] – 86%.

['expected','actual'] – 86%.

['start','problem'] – 10%.

Аномалий не обнаружено.

**Кластер 11 (6%)**

Медиана количества слов 15, среднее количество слов 19.

Xml 8%, STR 2%, EOB 1%, Java Code 9%, Stack Trace во всех ДЭК.

Л10% (10 EOB) - 20%.

Л10% (5 log/zip + stack trace) - 8%.

Л10% (8 log/zip + xml+ stack trace) - 7%.

Аномалия – почти нет лексики, сопряжённой со Stack Trace в старой кластеризации.

**Кластер 12 (3%)**

Медиана количества слов 15, среднее количество слов 18.

Логи/архивы 9%. Картинки 3%. Xml 10%, STR 3%, EOB 3%, Java Code 38%, Stack Trace во всех ДЭК.

Л10% (12 log/zip+xml) – во всех ДЭК.

Л10% (10 EOB) - 29%.

Л10% (6 STR) - 24%.

Л10% (8 log/zip + xml+ stack trace) - 20%.

Л10% (3 xml + stack trace) - 18%.

Л10% (5 log/zip + stack trace) – 16%.

Л10% (0 STR + stack trace) – 15%

Л10% (9 STR + EOB) – 13%

Аномалий не обнаружено.

**Кластер 13 (5%)**

Медиана количества слов 26, среднее количество слов 34.

Логи/архивы 25%. Картинки 3%. Xml 17%, STR 4%, EOB 1%, Java Code 10%, Stack Trace во всех ДЭК.

Л10% (11 STR + xml) – во всех ДЭК.

Л10% (12 log/zip+xml) - 35%.

Л10% (10 EOB) - 26%.

Л10% (3 xml + stack trace) - 18%.

Аномалий не обнаружено.

**Кластер 14 (6%)**

Медиана количества слов 11, среднее количество слов 20.

Логи/архивы 3%. Картинки 9%. Xml 3%, STR 10%, EOB 1%, Java Code 1%.

Л10% (13 images) – во всех ДЭК.

Л10% (6 STR) - 49%.

Л10% (10 EOB) - 20%.

Л10% (8 log/zip + xml+ stack trace) - 15%.

Аномалия – слова характерные для картинок в старой кластеризации во всех ДЭК.

**Кластер 15 (1%)**

Медиана количества слов 69, среднее количество слов 77.

Логи/архивы 6%. Xml 3%, STR 56%, EOB 26%, Stack Trace 13%.

Л10% (13 images) – 60%.

['logged'] – 10%

['reproduce','step'] – во всех ДЭК.

['reproduce','step','result'] – 90%.

['reproduce','login'] – 70%.

['reproduce','click'] – 43%

['expected','actual'] – 17%

Аномалии не обнаружены.

**Кластер 16 (2%)**

Медиана количества слов 54, среднее количество слов 71.

Логи/архивы 51%. Картинки 14%. Xml 69%, STR 77%, EOB 3%, Stack Trace 49%. Java Code – 3%.

Л10% (11 STR + xml) – 71%.

Л10% (10 EOB) – 60%.

Л10% (12 log/zip+xml )– 20%

Л10% (8 log/zip + xml+ stack trace) – 17%.

['attached','start'] – 68%

['file','broker'] – 40%

['attached','configuration'] – 25%

['following','producer'] – 20%

['following','configuration'] – 17%

['following','directory'] – 11%

--

['start','problem'] – 51%

['start','note'] – 31%

['reproduce','directory'] – 23%

['start','stop'] – 23%

['reproduce','step'] – 17%

['reproduce','restart'] – 14%.

Аномалии не обнаружены.

**Кластер 17 (3%)**

Медиана количества слов 14, среднее количество слов 22.

Логи/архивы 5%. Картинки 2%. Xml 5%, STR 5%, Java Code 5%, Stack Trace 20%.

Л10% (0 STR + stack trace) – во всех ДЭК.

Л10% (6 STR) - 23%.

Л10% (7 xml) - 20%.

Л10% (10 EOB) - 12%.

Л10% (9 STR + EOB) - 11%.

Аномалий не обнаружено.

**Кластер 18 (3%)**

Медиана количества слов 11, среднее количество слов 19.

Логи/архивы 16%. Xml 7%, STR 4%, EOB 5%, Java Code 21%, Stack Trace 16%.

Л10% (0 STR + stack trace) – во всех ДЭК.

Л10% (12 log/zip+xml) - 92%.

Л10% (6 STR) - 54%.

Л10% (10 EOB) - 53%.

Аномалия – высокая доля лексики из старых кластеров STR/EOB.

**Кластер 19 (3%)**

Медиана количества слов 14, среднее количество слов 21.

Логи/архивы 1%. Картинки 2%. Xml 13%, STR 4%, EOB 5%, Java Code 16%, Stack Trace 20%.

Л10% (3 xml + stack trace) – во всех ДЭК.

Л10% (5 log/zip + stack trace) -21%.

Л10% (12 log/zip+xml) - 19%.

Л10% (7 xml) - 19%.

Л10% (10 EOB) – 19%.

Л10% (13 images) – 13%.

Аномалий не обнаружено.

1. Обозначим эту фразу Л10% [↑](#footnote-ref-1)
2. Обозначим эту фразу ДЭК [↑](#footnote-ref-2)