**Хорошо различимые данные**

**Base Uncased**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.61 0.61 0.62  V measure:  0.05 0.05 0.16  Mutual info score:  0.05 0.05 0.16  Fowlkes Mallows:  0.31 0.32 0.38 | Adjuster Rand:  0.53 0.60 0.61  V measure:  0.03 0.05 0.05  Mutual info score:  0.03 0.04 0.05  Fowlkes Mallows:  0.31 0.32 0.36 |
| Существительные и прилагательные | Adjuster Rand:  0.63 0.63 0.63  V measure:  0.19 0.19 0.19  Mutual info score:  0.19 0.19 0.19  Fowlkes Mallows:  0.40 0.40 .40 | Adjuster Rand:  0.63 0.62 0.63  V measure:  0.03 0.19 0.19  Mutual info score:  0.03 0.18 0.19  Fowlkes Mallows:  0.36 0.40 0.40 |
| Существительные | Adjuster Rand:  0,62 0,62 0,62  V measure:  0,18 0,18 0,19  Mutual info score:  0,18 0,18 0,19  Fowlkes Mallows:  0,39 0,39 0,39 | Adjuster Rand:  0,62 0,62 0,62  V measure:  0,19 0,19 0,19  Mutual info score:  0,18 0,19 0,19  Fowlkes Mallows:  0,39 0,39 0,39 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.58 c:0.44 a:0.28 s:0.28  V measure:  w:0.09 c:0.01 a:0 s:0  Mutual info score:  w:0.12 c:0.03 a:0 s:0  Fowlkes Mallows:  w:0.16 c:0.06 a:0.02 s:0.02 | Adjuster Rand:  c:0.56 a:0.28 s:0.28  V measure:  c:0.02 a:0 s:0  Mutual info score:  c:0.03 a:0 s:0  Fowlkes Mallows:  c:0.04 a:0.02 s:0.02 |
| Существительные и прилагательные | Adjuster Rand:  w:0.64 c:0.40 a:0.28 s:0.28  V measure:  w:0.18 c:0.02 a:0. s:0.  Mutual info score:  w:0.21 c:0.05 a:0. s:0.  Fowlkes Mallows:  w:0.26 c:0.13 a:0.02 s:0.02 | Adjuster Rand:  c:0.48 a:0.28 s:0.28  V measure:  c:0.03 a:0. s:0.  Mutual info score:  c:0.05 a:0. s:0.  Fowlkes Mallows:  c:.0.09 a:0.02 s:0.02 |
| Существительные | Adjuster Rand:  w:0.64 c:0.52 a:0.28 s:0.28  V measure:  w:0.18 c:0.03 a:0. s:0.  Mutual info score:  w:0.23 c:0.03 a:0. s:0.  Fowlkes Mallows:  w:0.28 c:0.05 a:0.02 s:0.02 | Adjuster Rand:  c:0.60 a:0.29 s:0.28  V measure:  c:0.04 a:0. s:0.  Mutual info score:  c:0.04 a:0. s:0.  Fowlkes Mallows:  c:.0.05 a:0.02 s:0.02 |

**Base Cased**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.60 0.60 0.64  V measure:  0.02 0.03 0.05  Mutual info score:  0.02 0.03 0.05  Fowlkes Mallows:  0.27 0.29 0.30 | Adjuster Rand:  0.50 0.59 0.60  V measure:  0.02 0.03 0.03  Mutual info score:  0.02 0.02 0.03  Fowlkes Mallows:  0.29 0.30 0.38 |
| Существительные и прилагательные | Adjuster Rand:  0,58 0,60 0,64  V measure:  0,03 0,04 0,07  Mutual info score:  0,03 0,04 0,06  Fowlkes Mallows:  0,28 0,31 0,31 | Adjuster Rand:  0,59 0,61 0,64  V measure:  0,03 0,04 0,07  Mutual info score:  0,03 0,04 0,07  Fowlkes Mallows:  0,28 0,31 0,31 |
| Существительные | Adjuster Rand:  0.58 0.59 0.64  V measure:  0.03 0.03 0.06  Mutual info score:  0.03 0.03 0.06  Fowlkes Mallows:  0.28 0.31 0.31 | Adjuster Rand:  0.58 0.60 0.64  V measure:  0.03 0.04 0.08  Mutual info score:  0.03 0.04 0.08  Fowlkes Mallows:  0.28 0.31 0.31 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.59 c:0.50 a:0.25 s:0.28  V measure:  w:0.02 c:0.01 a:0 s:0  Mutual info score:  w:0.03 c:0.01 a:0. s:0.  Fowlkes Mallows:  w:0.03 c:0.03 a:0.16 s:0.02 | Adjuster Rand:  c:0.45 a:0.28 s:0.28  V measure:  c:0.01 a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.03 a:0.03 s:0.03 |
| Существительные и прилагательные | Adjuster Rand:  w:0.58 c:0.39 a:0.25 s:0.25  V measure:  w:0.03 c:0.01 a:0. s:0.  Mutual info score:  w:0.03 c:0.02 a:0. s:0.  Fowlkes Mallows:  w:0.03 c:0.06 a:0.08 s:0.15 | Adjuster Rand:  c:0.50 a:0.25 s:0.28  V measure:  c:0.02 a:0. s:0.  Mutual info score:  c:0.03 a:0. s:0.  Fowlkes Mallows:  c:.0.05 a:0.12 s:0.02 |
| Существительные | Adjuster Rand:  w:0.55 c:0.57 a:0.25 s:0.25  V measure:  w:0.01 c:0.02 a:0. s:0.  Mutual info score:  w:0.02 c:0.03 a:0. s:0.  Fowlkes Mallows:  w:0.03 c:0.04 a:0.08 s:0.15 | Adjuster Rand:  c:0.48 a:0.25 s:0.28  V measure:  c:0. a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.03 a:0.15 s:0.02 |

**Distile Bert**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.61 0.62 0.62  V measure:  0.06 0.06 0.09  Mutual info score:  0.06 0.06 0.09  Fowlkes Mallows:  0.32 0.32 .07 | Adjuster Rand:  0.63 0.64 0.67  V measure:  0.07 0.07 0.21  Mutual info score:  0.06 0.07 0.20  Fowlkes Mallows:  0.31 0.32 0.39 |
| Существительные и прилагательные | Adjuster Rand:  0,61 0,62 0,66  V measure:  0,06 0,12 0,22  Mutual info score:  0,06 0,12 0,22  Fowlkes Mallows:  0,32 0,35 0,41 | Adjuster Rand:  0,63 0,65 0,67  V measure:  0,08 0,12 0,20  Mutual info score:  0,08 0,12 0,20  Fowlkes Mallows:  0,33 0,34 0,39 |
| Существительные | Adjuster Rand:  0.61 0.62 0.67  V measure:  0.06 0.11 0.20  Mutual info score:  0.06 0.11 0.19  Fowlkes Mallows:  0.32 0.35 0.39 | Adjuster Rand:  0.62 0.66 0.67  V measure:  0.06 0.16 0.20  Mutual info score:  0.06 0.16 0.20  Fowlkes Mallows:  0.31 0.36 0.39 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.60 c:0.57 a:0.39 s:0.25  V measure:  w:0.06 c:0.01 a:0. s:0.  Mutual info score:  w:0.06 c:0.02 a:0.1 s:0.1  Fowlkes Mallows:  w:0.07 c:0.03 a:0.02 s:0.15 | Adjuster Rand:  c:0.59 a:0.43 s:0.25  V measure:  c:0.03 a:0.01 s:0.  Mutual info score:  c:0.06 a:0.02 s:0.  Fowlkes Mallows:  c:.0.07 a:0.05 s:0.15 |
| Существительные и прилагательные | Adjuster Rand:  w:0.64 c:0.60 a:0.43 s:0.25  V measure:  w:0.14 c:0.05 a:0.01 s:0.  Mutual info score:  w:0.13 c:0.05 a:0.01 s:0.  Fowlkes Mallows:  w:0.15 c:0.06 a:0.02 s:0.15 | Adjuster Rand:  c:0.49 a:0.47 s:0.25  V measure:  c:0.01 a:0.01 s:0.  Mutual info score:  c:0.01 a:0.01 s:0.  Fowlkes Mallows:  c:.0.02 a:0.03 s:0.15 |
| Существительные | Adjuster Rand:  w:0.62 c:0.57 a:0.47 s:0.25  V measure:  w:0.11 c:0.07 a:0.01 s:0.  Mutual info score:  w:0.15 c:0.06 a:0.01 s:0.  Fowlkes Mallows:  w:0.18 c:0.09 a:0.03 s:0.15 | Adjuster Rand:  c:0.60 a:0.25 s:0.25  V measure:  c:0.08 a:0. s:0.  Mutual info score:  c:0.08 a:0. s:0.  Fowlkes Mallows:  c:.0.10 a:0.17 s:0.15 |

**Плохо различимые данные**

**Base Uncased**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.57 0.57 0.58  V measure:  0.02 0.02 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.34 0.34 0.34 | Adjuster Rand:  0.57 0.57 0.58  V measure:  0.02 0.02 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.34 0.34 0.34 |
| Существительные и прилагательные | Adjuster Rand:  0.56 0.56 0.56  V measure:  0.03 0.03 0.03  Mutual info score:  0.03 0.03 0.03  Fowlkes Mallows:  0.36 0.36 0.36 | Adjuster Rand:  0.52 0.56 0.56  V measure:  0.02 0.02 0.03  Mutual info score:  0.02 0.02 0.03  Fowlkes Mallows:  0.35 0.35 0.42 |
| Существительные | Adjuster Rand:  0.55 0.56 0.56  V measure:  0.02 0.03 0.03  Mutual info score:  0.02 0.02 0.03  Fowlkes Mallows:  0.36 0.36 0.36 | Adjuster Rand:  0.55 0.55 0.56  V measure:  0.02 0.02 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.36 0.36 0.36 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.55 c:0.50 a:0.35 s:0.35  V measure:  w:0.03 c:0.02 a:0. s:0.  Mutual info score:  w:0.04 c:0.02 a:0. s:0.  Fowlkes Mallows:  w:0.04 c:0.04 a:0. s:0. | Adjuster Rand:  c:0.45 a:0.35 s:0.35  V measure:  c:0. a:0. s:0.  Mutual info score:  c:0.02 a:0. s:0.  Fowlkes Mallows:  c:.0.03 a:0.01 s:0.01 |
| Существительные и прилагательные | Adjuster Rand:  w:0.56 c:0.53 a:0.35 s:0.35  V measure:  w:0.04 c:0.01 a:0. s:0.  Mutual info score:  w:0.03 c:0.02 a:0. s:0.  Fowlkes Mallows:  w:0.03 c:0.02 a:0.01 s:0.01 | Adjuster Rand:  c:0.52 a:0.35 s:0.35  V measure:  c:0.02 a:0. s:0.  Mutual info score:  c:0.02 a:0. s:0.  Fowlkes Mallows:  c:.0.03 a:0.01 s:0.01 |
| Существительные | Adjuster Rand:  w:0.53 c:0.52 a:0.35 s:0.35  V measure:  w:0.01 c:0.03 a:0. s:0.  Mutual info score:  w:0.02 c:0.04 a:0. s:0.  Fowlkes Mallows:  w:0.02 c:0.05 a:0. s:0. | Adjuster Rand:  c:0.52 a:0.35 s:0.35  V measure:  c:0.03 a:0. s:0.  Mutual info score:  c:0.02 a:0. s:0.  Fowlkes Mallows:  c:.0.03 a:0. s:0. |

**Base Cased**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.53 0.57 0.58  V measure:  0.01 0.01 0.02  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.31 0.33 0.38 | Adjuster Rand:  0.53 0.57 0.58  V measure:  0.01 0.01 0.02  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.31 0.34 0.38 |
| Существительные и прилагательные | Adjuster Rand:  0.57 0.57 0.57  V measure:  0.01 0.02 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.33 0.34 0.34 | Adjuster Rand:  0.50 0.57 0.59  V measure:  0.01 0.02 0.02  Mutual info score:  0.01 0.02 0.02  Fowlkes Mallows:  0.30 0.34 0.43 |
| Существительные | Adjuster Rand:  0.49 0.56 0.59  V measure:  0.01 0.01 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.31 0.34 0.44 | Adjuster Rand:  0.49 0.56 0.59  V measure:  0.02 0.02 0.02  Mutual info score:  0.02 0.02 0.02  Fowlkes Mallows:  0.30 0.34 0.44 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.57 c:0.49 a:0.34 s:0.35  V measure:  w:0.01 c:0.01 a:0. s:0.  Mutual info score:  w:0.01 c:0.01 a:0. s:0.  Fowlkes Mallows:  w:0.01 c:0.01 a:0.05 s:0. | Adjuster Rand:  c:0.52 a:0.34 s:0.35  V measure:  c:0.01 a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.02 a:0.10 s:0. |
| Существительные и прилагательные | Adjuster Rand:  w:0.55 c:0.46 a:0.34 s:0.35  V measure:  w:0.02 c:0.01 a:0. s:0.  Mutual info score:  w:0.02 c:0.01 a:0. s:0.  Fowlkes Mallows:  w:0.02 c:0.03 a:0.13 s:0. | Adjuster Rand:  c:0.45 a:0.34 s:0.34  V measure:  c:0.01 a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.02 a:0.07 s:0.13 |
| Существительные | Adjuster Rand:  w:0.54 c:0.48 a:0.34 s:0.34  V measure:  w:0.02 c:0.01 a:0. s:0.  Mutual info score:  w:0.02 c:0.01 a:0. s:0.  Fowlkes Mallows:  w:0.02 c:0.02 a:0.14 s:0.14 | Adjuster Rand:  c:0.45 a:0.34 s:0.34  V measure:  c:0.01 a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.02 a:0.13 s:0.13 |

**Distile Bert**

K-means algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  0.58 0.58 0.58  V measure:  0.01 0.01 0.01  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.31 0.31 0.32 | Adjuster Rand:  0.58 0.58 0.58  V measure:  0.01 0.01 0.01  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.31 0.31 0.31 |
| Существительные и прилагательные | Adjuster Rand:  0.56 0.56 0.57  V measure:  0.01 0.01 0.01  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.33 0.33 0.34 | Adjuster Rand:  0.57 0.57 0.57  V measure:  0.01 0.01 0.02  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.33 0.33 0.33 |
| Существительные | Adjuster Rand:  0.56 0.56 0.57  V measure:  0.01 0.01 0.01  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.33 0.34 0.34 | Adjuster Rand:  0.57 0.57 0.57  V measure:  0.01 0.01 0.01  Mutual info score:  0.01 0.01 0.01  Fowlkes Mallows:  0.33 0.33 0.33 |

Hierarchical algorithm

|  |  |  |
| --- | --- | --- |
| Предобработка\метрика | Евклидова | Косинусная |
| Все части речи | Adjuster Rand:  w:0.57 c:0.54 a:0.40 s:0.34  V measure:  w:0.01 c:0.01 a:0. s:0.  Mutual info score:  w:0.02 c:0.01 a:0. s:0.  Fowlkes Mallows:  w:0.01 c:0.01 a:0. s:0.09 | Adjuster Rand:  c:0.47 a:0.40 s:0.34  V measure:  c:0. a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.01 a:0. s:0.13 |
| Существительные и прилагательные | Adjuster Rand:  w:0.54 c:0.54 a:0.42 s:0.34  V measure:  w:0.01 c:0. a:0. s:0.  Mutual info score:  w:0.02 c:0. a:0. s:0.  Fowlkes Mallows:  w:0.02 c:0. a:0. s:0.12 | Adjuster Rand:  c:0.49 a:0.43 s:0.34  V measure:  c:0. a:0. s:0.  Mutual info score:  c:0. a:0. s:0.  Fowlkes Mallows:  c:.0.01 a:0.01 s:0.12 |
| Существительные | Adjuster Rand:  w:0.53 c:0.47 a:0.47 s:0.34  V measure:  w:0. c:0. a:0. s:0.  Mutual info score:  w:0.01 c:0. a:0. s:0.  Fowlkes Mallows:  w:0.01 c:0.01 a:0.01 s:0.13 | Adjuster Rand:  c:0.50 a:0.44 s:0.34  V measure:  c:0.01 a:0. s:0.  Mutual info score:  c:0.01 a:0. s:0.  Fowlkes Mallows:  c:.0.01 a:0. s:0.13 |