|  | Skipgram | fastText, h=10, bigram | MDLText |  |  |
|---|---|---|---|---|---|
| Number | 2,00 | 6,00 | 34,00 |  |  |
| average | 57,50% | 79,87% | 86,62% |  |  |
| min | 43,00% | 60,20% | 67,20% |  |  |
| max | 72,00% | 95,70% | 98,50% |  |  |
| difference | 29,00% | 35,50% | 31,30% |  |  |

| type | dataset | parent dataset | Skipgram | fastText, h=10, bigram | MDLText |
|---|---|---|---|---|---|
| news articles | 20Newsgroups | - |  |  | 90,20% |
| news articles | Irish economic | - |  |  | 67,20% |
| news articles | La1S | TREC-5 |  |  | 87,70% |
| news articles | La2S | TREC-5 |  |  | 88,30% |
| news articles | Latimes | LA Times |  |  | 84,70% |
| news articles | RCV1 Top-4 | - |  |  | 92,00% |
| news articles | Re0 | Reuters-21578 |  |  | 81,50% |
| news articles | Re1 | Reuters-21578 |  |  | 81,20% |
| news articles | Re8 | Reuters-21578 |  |  | 91,40% |
| news articles | Reuters | - |  |  | 91,20% |
| news articles | Reviews | San Jose Mercury newspaper |  |  | 90,30% |
| news articles | New3S | - |  |  | 85,80% |
| news articles | AG News | - |  | 92,50% |  |
| web pages | 7Sectors | - |  |  | 91,30% |
| web pages | Techtc300-1092-135724 | - |  |  | 97,80% |
| web pages | Techtc300-1092-789236 | - |  |  | 98,50% |
| scientific documents | ACM | - |  |  | 82,20% |
| scientific documents | CSTR | Computer Science Technical Reports |  |  | 89,20% |
| scientific documents | Oh0 | HSUMED |  |  | 92,50% |
| scientific documents | Oh10 | HSUMED |  |  | 81,90% |
| scientific documents | Oh15 | HSUMED |  |  | 85,90% |
| scientific documents | Oh5 | HSUMED |  |  | 90,70% |
| scientific documents | Ohscal | HSUMED |  |  | 78,00% |
| scientific documents | Pubmed-Cancer | - |  |  | 92,00% |
| scientific documents | Pubmed-Cancer-2000 | - |  |  | 84,80% |
| Mail | Enron | - |  |  | 71,60% |
| Mail | Trec7-3000 | - |  |  | 98,10% |
| Q&A | Opinosis | - |  |  | 67,40% |
| Q&A | Yelp P. | - |  | 95,70% |  |
| Q&A | Yelp F | - |  | 63,90% |  |
| Q&A | Yah A | - |  | 72,30% |  |
| Q&A | Amz F | - |  | 60,20% |  |
| Q&A | Amz P | - |  | 94,60% |  |
| Text Retrival Conference | Tr11 | TRE |  |  | 82,70% |
| Text Retrival Conference | Tr12 | TRE |  |  | 87,90% |
| Text Retrival Conference | Tr21 | TRE |  |  | 83,60% |
| Text Retrival Conference | Tr23 | TRE |  |  | 92,10% |

| Text Retrival Conference | Tr31 | TRE | | | 82,60% |
|---|---|---|---|---|---|
| Text Retrival Conference | Tr41 | TRE | | | 92,40% |
| Text Retrival Conference | Tr45 | TRE | | | 90,30% |
| word pairs | RW | - | 43,00% | | |
| word pairs | WS353 | - | 72,00% | | |

19.02.2022