Sheet 5 Aufgabe 3

Permutation: Wie si

nd Wörter angeordnet?

Sinn: Wie verhält sich Ergebnis mit Minhashing vs Ergebnis nur Spaltverglecich

A)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| D | D1 | D2 | D3 | 1 | 2 | 3 | 4 |
| Your | 1 | 0 | 0 | 1 | 3 | 1 | 5 |
| New | 1 | 1 | 0 | 7 | 10 | 9 | 6 |
| Red | 1 | 1 | 1 | 9 | 1 | 6 | 4 |
| Bag | 1 | 0 | 1 | 2 | 6 | 3 | 7 |
| Lies | 1 | 1 | 1 | 3 | 9 | 5 | 10 |
| On | 1 | 1 | 1 | 0 | 5 | 8 | 8 |
| My | 1 | 1 | 1 | 8 | 2 | 2 | 3 |
| Table | 1 | 1 | 1 | 5 | 4 | 4 | 9 |
| Cat | 0 | 1 | 0 | 6 | 8 | 7 | 2 |
| old | 0 | 0 | 1 | 4 | 7 | 10 | 1 |
|  |  |  |  |  |  |  |  |

b) Signatur Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | D1 | D2 | D3 |
| 1 | 1 | 3 | 2 |
| 2 | 1 | 1 | 1 |
| 3 | 1 | 2 | 2 |
| 4 | 5 | 2 | 1 |

c)

|  |  |  |  |
| --- | --- | --- | --- |
|  | D1d2 | D2d3 | D1d3 |
| Col | 6/9 | 5/9 | 6/9 |
| sig | 1/4 | 2/4 | ¼ |

* Schlussfolgerung: Bei einzelnen Wörtern sind D1D2/D1D3 bzw D2D3 am ähnlichsten

d.a)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| D | D1 | D2 | D3 | 1 | 2 | 3 | 4 |
| Your new red | 1 | 0 | 0 | 7 | 12 | 3 | 10 |
| New red bag | 1 | 0 | 0 | 9 | 9 | 1 | 12 |
| Red bag lies | 1 | 0 | 1 | 1 | 5 | 9 | 2 |
| Bag lies on | 1 | 0 | 1 | 6 | 6 | 5 | 9 |
| Lies on my | 1 | 1 | 1 | 10 | 2 | 10 | 4 |
| On my table | 1 | 1 | 1 | 5 | 1 | 4 | 6 |
| My new red | 0 | 1 | 0 | 3 | 7 | 6 | 8 |
| New red cat | 0 | 1 | 0 | 12 | 10 | 7 | 1 |
| Red cat lies | 0 | 1 | 0 | 11 | 3 | 8 | 3 |
| Cat lies on | 0 | 1 | 0 | 8 | 8 | 11 | 5 |
| My old red | 0 | 0 | 1 | 4 | 11 | 12 | 7 |
| Old red bag | 0 | 0 | 1 | 2 | 4 | 2 | 11 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | D1 | D2 | D3 |
| 1 | 1 | 3 | 1 |
| 2 | 1 | 1 | 1 |
| 3 | 1 | 4 | 2 |
| 4 | 2 | 1 | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | D1d2 | D2d3 | D1d3 |
| Col | 2/10 | 2/10 | 4/8 |
| sig | 1/4 | 1/4 | 3/4 |

* Nimmt man Shingles wird eher der Sinn der Texte verglichen, andere Texte sind ähnlicher