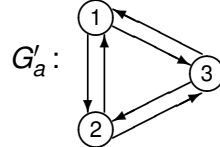


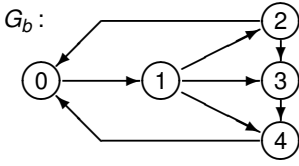
Update:	$i=0$	$i=1$	$i=2$	$i=3$
$\bar{h}_i^{k+1}$	$a_1^k + a_2^k$	$a_2^k + a_3^k$	$a_1^k + a_3^k$	0
$\bar{a}_i^{k+1}$	0	$h_0^k + h_2^k$	$h_0^k + h_1^k$	$h_1^k + h_2^k$



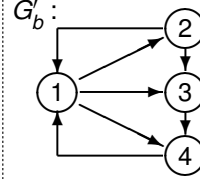
Vergleich Page Rank

$$p_{G_a} \approx \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0.09, & 0.27, & 0.27, & 0.36 \end{pmatrix}^T$$

$$p_{G'_a} \approx \begin{pmatrix} 1 & 2 & 3 \\ 0.33, & 0.33, & 0.33 \end{pmatrix}^T$$



Update:	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$
$\bar{h}_i^{k+1}$	$a_1^k$	$a_2^k + a_3^k + a_4^k$	$a_0^k + a_3^k$	$a_4^k$	$a_0^k$
$\bar{a}_i^{k+1}$	$h_2^k + h_4^k$	$h_0^k$	$h_1^k$	$h_1^k + h_2^k$	$h_1^k + h_3^k$



Vergleich Page Rank

$$p_{G_b} \approx \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.27, & 0.27, & 0.09, & 0.14, & 0.23 \end{pmatrix}^T$$

$$p_{G'_b} \approx \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.38, & 0.13, & 0.19, & 0.31 \end{pmatrix}^T$$

Hub-Vektor

	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{h}^0$	1	1	1	1	1	1	1	1	1	1	1	1
$\bar{h}^1$	3	3	3	0	2	2	0	0	2	2	0	0
$\bar{h}^2$	9	9	9	0	4	4	0	0	4	4	0	0
$\bar{h}^3$	27	27	27	0	8	8	0	0	8	8	0	0

Authority-Vektor

	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{a}^0$	1	1	1	1	1	1	1	1	1	1	1	1
$\bar{a}^1$	0	0	0	3	3	3	2	2	0	0	2	2
$\bar{a}^2$	0	0	0	9	9	9	4	4	0	0	4	4
$\bar{a}^3$	0	0	0	27	27	27	8	8	0	0	8	8

Sei  $N$  = Anzahl der Dokumente in der Dokumentensammlung  
 $DF(T_j)$  = Anzahl der Dokumente, in denen der Term  $T_j$  vorkommt.  
 $DL(D_i)$  = Länge des Dokuments  $D_i$  gemessen in Termvorkommen =  $\sum_k RTF(D_i, T_k)$ .  
 $RTF(D_i, T_j)$  = Anzahl der Vorkommen von Term  $T_j$  in Dokument  $D_i$ .

Dann ist

$$TF\_IDF(D_i, T_j) = \frac{RTF(D_i, T_j)}{DL(D_i)} \cdot \log \frac{N}{DF(T_j)} = \frac{RTF(D_i, T_j)}{\sum_k RTF(D_i, T_k)} \cdot \log \frac{N}{DF(T_j)}$$

input:  $[(id_1, d_1), \dots, (id_N, d_N)]$

1. Termfrequenzen  $rtf_{ij} = RTF(D_i, T_j)$

map:  $(id, [Term]) \rightarrow [(id, Term), Int]$   
 $(id_i, d_i) \rightarrow [ \dots, ((id_i, t_j), 1), \dots ]$   
 für alle Term-Vorkommen  $t_j \in d_i$

reduce:  $((id, Term), [Int]) \rightarrow [Int]$   
 $((id_i, t_j), [ \dots, 1, \dots ]) \rightarrow [rtf_{ij}]$  für  $rtf_{ij} = length(\ell)$

output:  $[ \dots, ((id_i, t_j), rtf_{ij}), \dots ]$

2. Dokumentlängen  $dl_i = DL(D_i) = \sum_k RTF(D_i, T_k)$

map:  $((id, Term), Int) \rightarrow [(id, (Term, Int))]$   
 $((id_i, t_j), rtf_{ij}) \rightarrow [(id_i, (t_j, rtf_{ij}))]$

reduce:  $(id, [(Term, Int)]) \rightarrow [(Term, Int, Int)]$   
 $(id_i, [ \dots, (t_k, rtf_{ik}), \dots ]) \rightarrow [ \dots, (t_k, rtf_{ik}, dl_i), \dots ]$

für  $dl_i = \sum_{(t_k, rtf_{ik}) \in \ell} rtf_{ik}$

output:  $[ \dots, (id_i, (t_j, rtf_{ij}, dl_i)), \dots ]$

3. Dokumentfrequenzen  $df_j = DF(T_j)$

map:  $(id, (Term, Int, Int)) \rightarrow [(Term, (id, Int, Int))]$   
 $(id_i, (t_j, rtf_{ij}, dl_i)) \rightarrow [(t_j, (id_i, rtf_{ij}, dl_i))]$

reduce:  $(Term, [(id, Int, Int)]) \rightarrow [(id, Int, Int, Int)]$   
 $(t_j, [ \dots, (id_k, rtf_{kj}, dl_k), \dots ]) \rightarrow [ \dots, (id_k, rtf_{kj}, dl_k, df_j), \dots ]$

für  $df_j = length(\ell)$

output:  $[ \dots, (t_j, (id_i, rtf_{ij}, dl_i, df_j)), \dots ]$

4. Werte  $tfidf_{ij} = TF\_IDF(D_i, T_j)$

map:  $(Term, (id, Int, Int, Int)) \rightarrow [(id, Term), (Int, Int, Int)]$   
 $(t_j, (id_i, rtf_{ij}, dl_i, df_j)) \rightarrow [(id_i, t_j), (rtf_{ij}, dl_i, df_j)]$

reduce:  $((id, Term), [(Int, Int, Int)]) \rightarrow [Float]$

$((id_i, t_j), [(rtf_{ij}, dl_i, df_j)]) \rightarrow \begin{cases} [tfidf_{ij}] & \text{für } tfidf_{ij} = \frac{rtf_{ij}}{dl_i} \cdot \log \frac{N}{df_j} \\ & \text{falls } dl_i \neq 0 \text{ und } df_j \neq 0 \text{ und } N \neq 0 \\ [0.0] & \text{andernfalls} \end{cases}$

output:  $[ \dots, ((id_i, t_j), tfidf_{ij}), \dots ]$

**MapReduce:**

<i>input:</i>	[	]
	Ergebnis des <i>Map</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Grouping</i> -Schritts:	
<i>intermediate:</i>	[	]
	Ergebnis des <i>Reduce</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Ungrouping</i> -Schritts:	
<i>output:</i>	[	]

**MapReduce:**

<i>input:</i>	[	]
	Ergebnis des <i>Map</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Grouping</i> -Schritts:	
<i>intermediate:</i>	[	]
	Ergebnis des <i>Reduce</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Ungrouping</i> -Schritts:	
<i>output:</i>	[	]

**MapReduce:**

<i>input:</i>	[	]
	Ergebnis des <i>Map</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Grouping</i> -Schritts:	
<i>intermediate:</i>	[	]
	Ergebnis des <i>Reduce</i> -Schritts:	
	[	]
<hr/>		
	Ergebnis des <i>Ungrouping</i> -Schritts:	
<i>output:</i>	[	]