



## Googles PageRank

1. Die Formel
2. Das Eingabeformular
3. Die Anzeige
4. Die Berechnung
5. Fragen zur Realität



# 1. Die Formel

PageRank meiner Seite

$$\text{PR} = (1-d) + d * ($$
$$\quad [ \text{PR} (L1) / L1 (\text{out}) ] +$$
$$\quad [ \text{PR} (L2) / L2 (\text{out}) ] +$$
$$\quad \dots )$$

d bezeichnet einen Dämpfungsfaktor, er wird mit 0,85 beziffert.



# 1. Die Formel

PageRank meiner Seite

$$\text{PR} = (1-d) + d * ($$
$$\quad [ \text{PR}(\text{L1}) / \text{L1}(\text{out}) ] +$$
$$\quad [ \text{PR}(\text{L2}) / \text{L2}(\text{out}) ] +$$
$$\quad \dots )$$

PR(L1) ist der PR einer Seite mit Link auf mich! L1(out) ist die Anzahl der Links, die von L1 wegzeigen!



# 1. Die Formel

PageRank meiner Seite

$$PR = (1-d) + d * ( [PR(L1) / L1(out)] + [PR(L2) / L2(out)] + \dots )$$

Da sich  $PR(L1)$  nicht von selbst ergibt, errechnet man ihn iterativ, dh. man beginnt mit einem (beliebigem) Startwert, als PR z.B. 1 für jede Seite.



# 1. Die Formel

PageRank meiner Seite

$$PR = (1-d) + d * ($$

$$[PR(L1) / L1(out)] +$$

$$[PR(L2) / L2(out)] +$$

$$\dots ] )$$

Jetzt muss dieses Verfahren nur n mal durchgeführt werden, um ein hinreichend genaues Ergebnis zu erhalten, z.B. 40 Mal.



## 2. Das Eingabeformular

Wir wollen in PHP den  
PageRank einer kleinen Welt  
berechnen, in der es nur 10  
Seiten gibt!



## 2. Das Eingabeformular

Die Links lassen sich am einfachsten in einer sog. Adjazenzmatrix darstellen und im Programm bearbeiten.



## 2. Das Eingabeformular

Simulation PageRank - Mozilla Firefox

Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

http://192.168.1.107/pr/ Go

Kostenlose Hotmail Links anpassen RealPlayer Windows Media Windows

Diese Matrix stellt die Links der Seiten zueinander dar, 1 bedeutet Link, 0 kein Link.

|   | 0                              | 1                              | 2                              | 3                              | 4                              | 5                              | 6                              | 7                              | 8                              | 9                              |
|---|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 0 | <input type="text" value="0"/> |
| 1 | <input type="text" value="0"/> |
| 2 | <input type="text" value="0"/> |
| 3 | <input type="text" value="0"/> |
| 4 | <input type="text" value="0"/> |
| 5 | <input type="text" value="0"/> |
| 6 | <input type="text" value="0"/> |
| 7 | <input type="text" value="0"/> |
| 8 | <input type="text" value="0"/> |
| 9 | <input type="text" value="0"/> |

PageRank

Fertig



## 2. Das Eingabeformular

```
<form action=... method="post">
```

```
<input type="text" name="p1">
```

```
<input type="text" name="p2">
```

```
...
```

```
<input type="submit" name="Schalter"  
value="PageRank">
```

```
</form>
```

```
<! Beschriftung nicht vergessen !>
```



## 2. Das Eingabeformular

idealerweise in Form einer Tabelle:

```
<table><tr><td>  
<input type="text" name="p1"></td> <td>  
<input type="text" name="p2"></td>  
...</tr><tr>  
<input type="submit" name="Schalter"  
  value="PageRank"><tr></table>  
  
</form>  
  
<! Beschriftung nicht vergessen !>
```



# 3. Die Anzeige

Ein Array ist eine Liste mit Variablen!

```
$Zahlen[0]=0; $Zahlen[1]=1; $Zahlen[2]=1;  
$Zahlen[3]=0; $Zahlen[4]=1; $Zahlen[5]=0;
```

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 | 0 |
|---|---|---|---|---|---|



# 3. Die Anzeige

Ein zweidimensionales Array ist wie ein Koordinatensystem

```
$Zahlen[0][0]=0; $Zahlen[1][0]=1;  
$Zahlen[2][0]=1; $Zahlen[3][0]=0;  
$Zahlen[4][0]=1; $Zahlen[5][0]=0;  
$Zahlen[0][1]=0; $Zahlen[1][1]=1;...
```

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 0 |



## 3. Die Anzeige

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 1 | 0 |

Variablen werden mit

```
$Seiten[0][0]=$_GET[p1] eingelesen.
```

Wir packen 100 Zahlen von p0 bis p99 einfach in ein zweidimensionales Array.

Dafür zählen wir zwei Schleifen verschachtelt durch jeweils von 0 bis 9 und können so 100 Werte einlesen!

```
$Seiten[$i1][$i2]=$_GET["p".(($i1*10)+$i2)]
```



## 3. Die Anzeige

Wenn wir die Zahlen alle in einem Feld haben, sollte die Anzeige einer Tabelle mit den Werten ein Kinderspiel sein!



## 4. Die Berechnung

Bisher haben wir ein zweidimensionales Array `$Seiten`, in dem wir alle Links angegeben haben!

Wir haben 10 Seiten.

Wir benötigen einen Array `$PageRank`, in der wir den PageRank jeder Seite eintragen. Wir starten mit einem Wert 1 für jede Seite!

`$PageRank`

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|



## 4. Die Berechnung

Top-Down-Ansatz:

```
// $Seiten ist vorhanden
```

```
// Setze für jede Seite den PR auf 1
```

```
$PageRank=PageRankStart();
```

```
$iter=40;
```

```
for ($i=0;$i<$iter;$i++){
```

```
$PageRank=GetIterPageRank($Seiten,$PageRank);
```

```
}
```

```
PageRankAusgeben();
```



## 4. Die Berechnung

```
$PageRank=GetIterPageRank  
($Seiten,$PageRank);
```

Berechnung PR für die `$Seite=0`:

`$SeitenLinks[]` sei ein Array, in die ich alle Seiten schreibe, die auf `$Seite` zeigen

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 4 | 5 | 7 | 8 | 9 |
|---|---|---|---|---|---|



## 4. Die Berechnung

Mit der Funktion

```
GetLinksOut($Seiten,$Seite) {
```

```
...
```

```
return $AnzahlLink;}
```

erhalte ich die Anzahl der Links, die von  
der Seite 2 weggehen

**\$AnzahlLink**

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 4 | 5 | 7 | 8 | 9 |
|---|---|---|---|---|---|



## 4. Die Berechnung

Für jede Seite die auf mich zeigt ermittle ich also den Wert der Links, die von der Seite wegführen!

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 4 | 5 | 7 | 8 | 9 |
|---|---|---|---|---|---|

\$SeitenLinks  
(zeigt auf mich)

|   |   |   |   |   |    |
|---|---|---|---|---|----|
| 3 | 5 | 4 | 3 | 1 | 12 |
|---|---|---|---|---|----|

AnzahlLink

```
$Ergebnis=0;  
for ($i=0;$i<count($SeitenLinks);$i++)  
$Ergebnis=$Ergebnis+$PageRank[$SeitenLinks  
[$i]]/GetLinksOut($Seiten,$SeitenLinks[$i]);
```

```
$DiesesPageRank[$Seite]=0.15+(0.85*$Ergebnis);
```



## 5. Fragen zur Realität

Wie groß ist die Adjazenzmatrix in Wirklichkeit. Wie stark steigt sie an wenn die Anzahl der Seiten von 10 auf 20 steigt?

Wie lange dauert es, bis Google das WWW erneut gescannt hat?

Was ist mit den Seiten, auf die kein Link zeigt, weil z.B. eine Eingabe erforderlich ist, bei Blogs & Co.?

