



UNIVERSIDAD CATÓLICA
de Colombia

PROYECTO FINAL ESTOCASTICOS Y SIMULACIÓN

Docente:

Roger Enrique Guzman

Presentado por:

David andres rubiano

Ever Moreno

Santiago Arias

Gabriel Jimenez

Jean Carlo Arévalo

Leonardo Oliveros Coral

Camila Castillo

Facultad de Ingeniería

25/11/2017



UNIVERSIDAD CATÓLICA
de Colombia

PAGERANK

Para el cálculo del PageRank es necesario numeración de cada una de las páginas web 1, ..., n, como se observa en la siguiente imagen, Pues para cada número de página n, hay un PageRank asociado qj que mide su importancia, El valor del PageRank es evaluado entre 0 y 1, entre más alto es su valor, más importante es la página.

```
0 http://unal.edu.co/
1 http://administrativos.unal.edu.co/
2 https://dninfoa.unal.edu.co/
3 http://redessociales.unal.edu.co/
4 http://unal.edu.co/buzon-de-notificaciones.html
5 http://www.dre.unal.edu.co/movilidad.html
6 http://www.unal.edu.co/control_interno/index.html
7 http://unal.edu.co/registro-y-matricula.html
8 http://unal.edu.co/la-universidad/naturaleza.html
9 http://quejasyreclamos.unal.edu.co/
10 http://www.palmira.unal.edu.co/
11 http://unal.edu.co/la-universidad/historia.html
12 http://www.virtual.unal.edu.co/
13 http://unal.edu.co/la-universidad/recorrido-virtual.html
14 http://unal.edu.co/internacionalizacion/la-universidad-en-el-mundo.html
15 http://unal.edu.co/acerca-de-este-sitio-web.html
```

$$q = (q_1, \dots, q_n)$$

El vector "q" contiene los valores calculados para cada pagina, este vector es una distribución de probabilidad, pues la suma de cada elemento del vector suma 1.

El cálculo del pageRank surge a raíz de la ejecución del web crawler el cual genera un archivo de texto llamado "Matriz.txt", este contendrá el total de páginas recorridas por el crawler, para este caso un total, de 200 de páginas y 3 diccionarios que se dividen de la siguiente manera: el primero contendrá los links de llegada para cada url, el segundo el número de links de salida que tiene una url y por último las páginas las cuales no contienen ninguna salida.

```
-->NUMERO DE PAGINAS 14
-->LINKS DE LLEGADAS {0: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 1: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 2: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 3: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 4: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 5: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 6: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 7: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 8: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 9: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 10: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 11: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 12: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 13: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], 14: [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]}
-->LINKS DE SALIDA{0: 14, 1: 11, 2: 14, 3: 11, 4: 14, 5: 11, 6: 14, 7: 14, 8: 14, 9: 14, 10: 14, 11: 11, 12: 14, 13: 11, 14: 14}
-->LINKS VACIOS{}
```

A continuación se podran observar los primeros 20, valores calculados del pageRank, se observara la llave asignada para cada url y el valor del pageRank.

[[1.95000000e+02 2.85667593e-02]]

[[1.94000000e+02 2.85667593e-02]]

[[1.92000000e+02 2.85667593e-02]]

[[1.99000000e+02 2.84596983e-02]]

[[1.98000000e+02 2.84596983e-02]]

[[1.97000000e+02 2.84596983e-02]]

[[1.96000000e+02 2.84596983e-02]]

[[1.93000000e+02 2.84596983e-02]]

[[1.91000000e+02 2.84596983e-02]]

[[7. 0.0284597]]

[[6. 0.0284597]]

[[5. 0.0284597]]

[[4. 0.0284597]]

[[3. 0.0284597]]

[[1.56000000e+02 1.20178574e-02]]

[[1.26000000e+02 9.50802691e-03]]

[[1.37000000e+02 9.49095568e-03]]

[[1.29000000e+02 9.37264381e-03]]

[[1.23000000e+02 9.36504410e-03]]

[[1.01000000e+02 9.22220591e-03]]

El Tiempo de ejecución del PageRank para 200 pagina(s) fue de: 16.0 seg.

REFERENCIAS

- <http://michaelnielsen.org/blog/using-your-laptop-to-compute-pagerank-for-millions-of-webpages/>
- <http://www.reportlab.com/>
- <https://www.humanlevel.com/diccionario-marketing-online/pagerank-google>