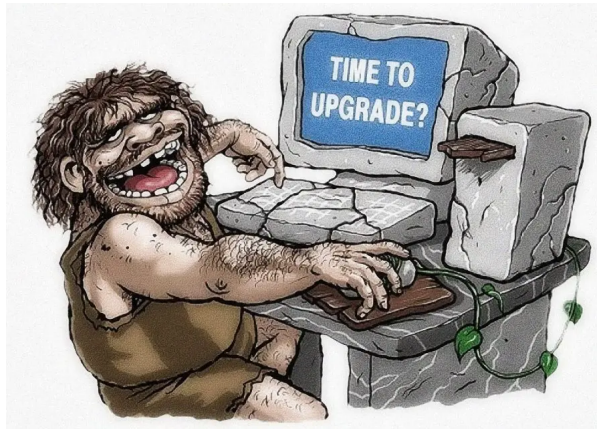


Processos Estocásticos

Berilhes

A Internet de Antigamente

A Internet de Antigamente



GOOGLE E PAGEARNK

- Infoseek

GOOGLE E PAGEARNK

- Infoseek
- Alta Vista

GOOGLE E PAGEARNK

- Infoseek
- Alta Vista
- Yahoo

GOOGLE E PAGEARNK

- Infoseek
- Alta Vista
- Yahoo
- Lycos

PROBLEMA

PROBLEMA

- O problema é que estas máquinas de pesquisa trabalhavam exclusivamente casando palavras.

PROBLEMA

- O problema é que estas máquinas de pesquisa trabalhavam exclusivamente casando palavras.
- Ou seja você podia facilmente enganar o sistema simplesmente fazendo páginas que contivessem palavras chaves repetidas centenas de vezes.

PROBLEMA

- O problema é que estas máquinas de pesquisa trabalhavam exclusivamente casando palavras.
- Ou seja você podia facilmente enganar o sistema simplesmente fazendo páginas que contivessem palavras chaves repetidas centenas de vezes.
- Como resolver este problema?

- Coletar todas as páginas que possuem um encaixe textual descente, e então ordenar estas páginas por alguma medida de autoridade ou qualidade.

- Coletar todas as páginas que possuem um encaixe textual descente, e então ordenar estas páginas por alguma medida de autoridade ou qualidade.
- Yahoo! fazia isso a mão.

- Coletar todas as páginas que possuem um encaixe textual descente, e então ordenar estas páginas por alguma medida de autoridade ou qualidade.
- Yahoo! fazia isso a mão.
- Entram em cena dois estudantes de doutorado em computação: Larry Page e Sergey Brin.

- Como nós podemos medir a qualidade ou autoridade de uma página web de forma automática?

- Como nós podemos medir a qualidade ou autoridade de uma página web de forma automática?
- A ideia é utilizar uma análise de hiperlink.

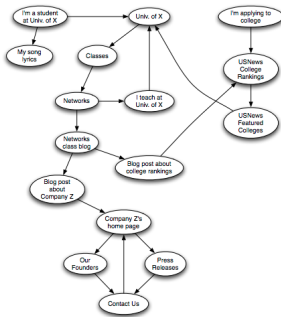
PAGERANK

- Como nós podemos medir a qualidade ou autoridade de uma página web de forma automática?
- A ideia é utilizar uma análise de hiperlink.
- Ou seja ver a web como um grafo dirigido.

PAGERANK

- Como nós podemos medir a qualidade ou autoridade de uma página web de forma automática?
- A ideia é utilizar uma análise de hiperlink.
- Ou seja ver a web como um grafo dirigido.

THE WEB AS A DIRECTED GRAPH



PAGERANK (CONT.)

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?



Ideia 1

Citações - Talvez nós pudéssemos ordenar as páginas pelo seu grau de incidência.

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?



Ideia 1

Citações - Talvez nós pudéssemos ordenar as páginas pelo seu grau de incidência.

- Problemas:

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?



Ideia 1

Citações - Talvez nós pudéssemos ordenar as páginas pelo seu grau de incidência.

- Problemas:
 - Você poderia fazer a sua página simplesmente criando 100 páginas falsas que apontam para a sua página.

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?



Ideia 1

Citações - Talvez nós pudéssemos ordenar as páginas pelo seu grau de incidência.

- Problemas:
 - Você poderia fazer a sua página simplesmente criando 100 páginas falsas que apontam para a sua página.
 - Outra problema é que um link de um especialista pode valer mais do que cinco de leigos.

PAGERANK (CONT.)

- Como estes links podem nos dizer alguma coisa a respeito da importância de uma página?



Ideia 1

Citações - Talvez nós pudéssemos ordenar as páginas pelo seu grau de incidência.

- Problemas:
 - Você poderia fazer a sua página simplesmente criando 100 páginas falsas que apontam para a sua página.
 - Outra problema é que um link de um especialista pode valer mais do que cinco de leigos.
 - Você poderia dar uma citação (voto) para cada um dos seus amigos simplesmente criando um link para a página deles.

PAGERANK - SOLUÇÕES

PAGERANK - SOLUÇÕES



Ideia 2 - Citações fracionadas

Se uma página possui d links de saída, então cada página apontada deveria obter somente $\frac{1}{d}$ de uma citação.

PAGERANK - SOLUÇÕES



Ideia 3 - Definição Recursiva

Se uma página x possui d links de saída , um dos quais dirige-se a página y , então esta deveria contribuir com $1/d$ para a importância de y .

Mas nós deveríamos levar em consideração a importância de x !
Ou seja nós deveríamos ponderar a contribuição de x pela sua importância.

PAGERANK - SOLUÇÕES

PAGERANK - SOLUÇÕES

- Suponha que π_v é a importância da página v . Para cada outra página u , esta possui uma fração dos seus links de saída que apontam para v . Vamos chamar esta fração de K_{uv} . Então

$$\pi_v = \sum_u \pi_u K_{uv}.$$

PAGERANK - SOLUÇÕES

PAGERANK - SOLUÇÕES

- Nós temos uma Cadeia de Markov de Tempo Discreto.

PAGERANK - SOLUÇÕES

- Nós temos uma Cadeia de Markov de Tempo Discreto.
- Onde os estados são as páginas.

PAGERANK - SOLUÇÕES

- Nós temos uma Cadeia de Markov de Tempo Discreto.
- Onde os estados são as páginas.
- As transições são os links.

PAGERANK - SOLUÇÕES

- Nós temos uma Cadeia de Markov de Tempo Discreto.
- Onde os estados são as páginas.
- As transições são os links.
- Em cada passo nós seguimos um link aleatório sob a página.

PAGERANK - SOLUÇÕES

- Nós temos uma Cadeia de Markov de Tempo Discreto.
- Onde os estados são as páginas.
- As transições são os links.
- Em cada passo nós seguimos um link aleatório sob a página.
- A importância, ou “PageRank” de uma página é apenas a distribuição estacionária da cadeia de Markov.