

PESQUISA NA INTERNET

Prof. Fernando Hattori



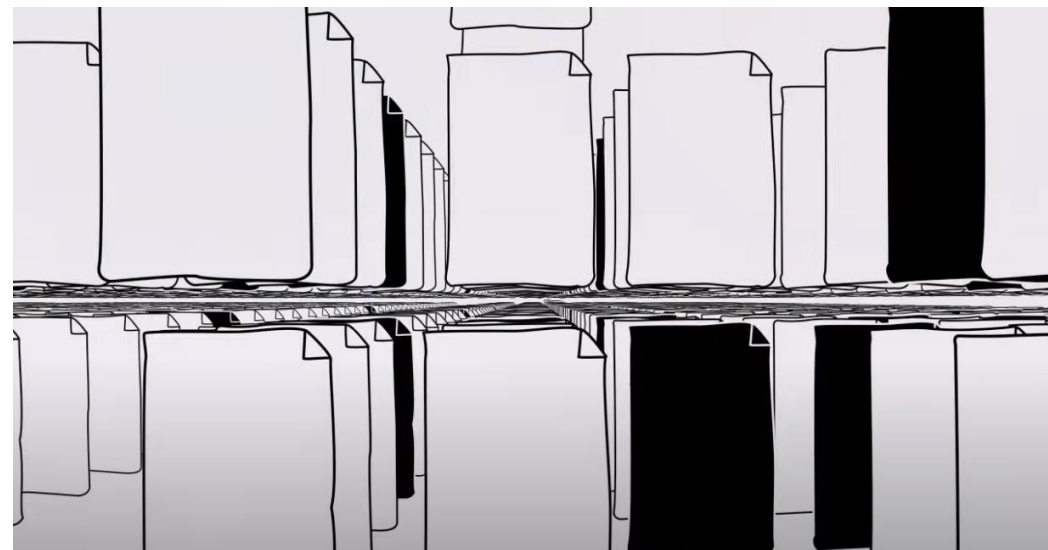



COMO FUNCIONA O GOOGLE?

- Robôs (Crawlers) varrem a internet procurando novas páginas;
- Quando encontram essas novas páginas, eles colocam essas páginas dentro do index do google.

O QUE É INDEX?

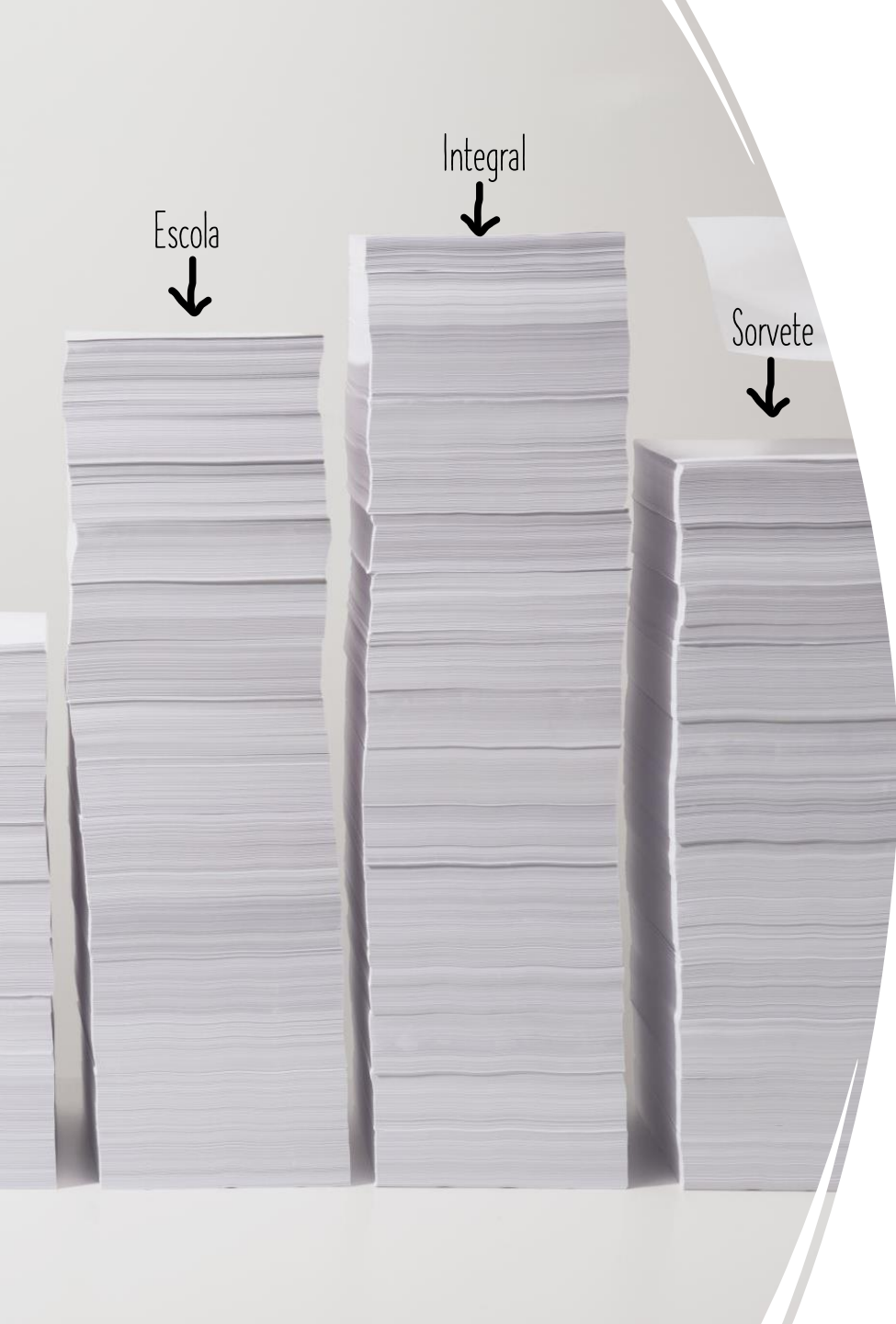
- Index é uma maneira de organizar todas as páginas que os robôs do google encontram;
- Dentro do índice, as páginas são organizadas pelas palavras que elas têm;
- Por exemplo, dentro do índice cada palavra do nosso dicionário tem uma lista e nessa lista estão todos as páginas que possuem aquela palavra.





O QUE ACONTECE QUANDO FAZEMOS UMA BUSCA?

- Quando fazemos uma busca no google, ele procura no índice as palavras que procuramos e devolve a lista de documentos que estão nessa lista;



O QUE ACONTECE QUANDO FAZEMOS UMA BUSCA?

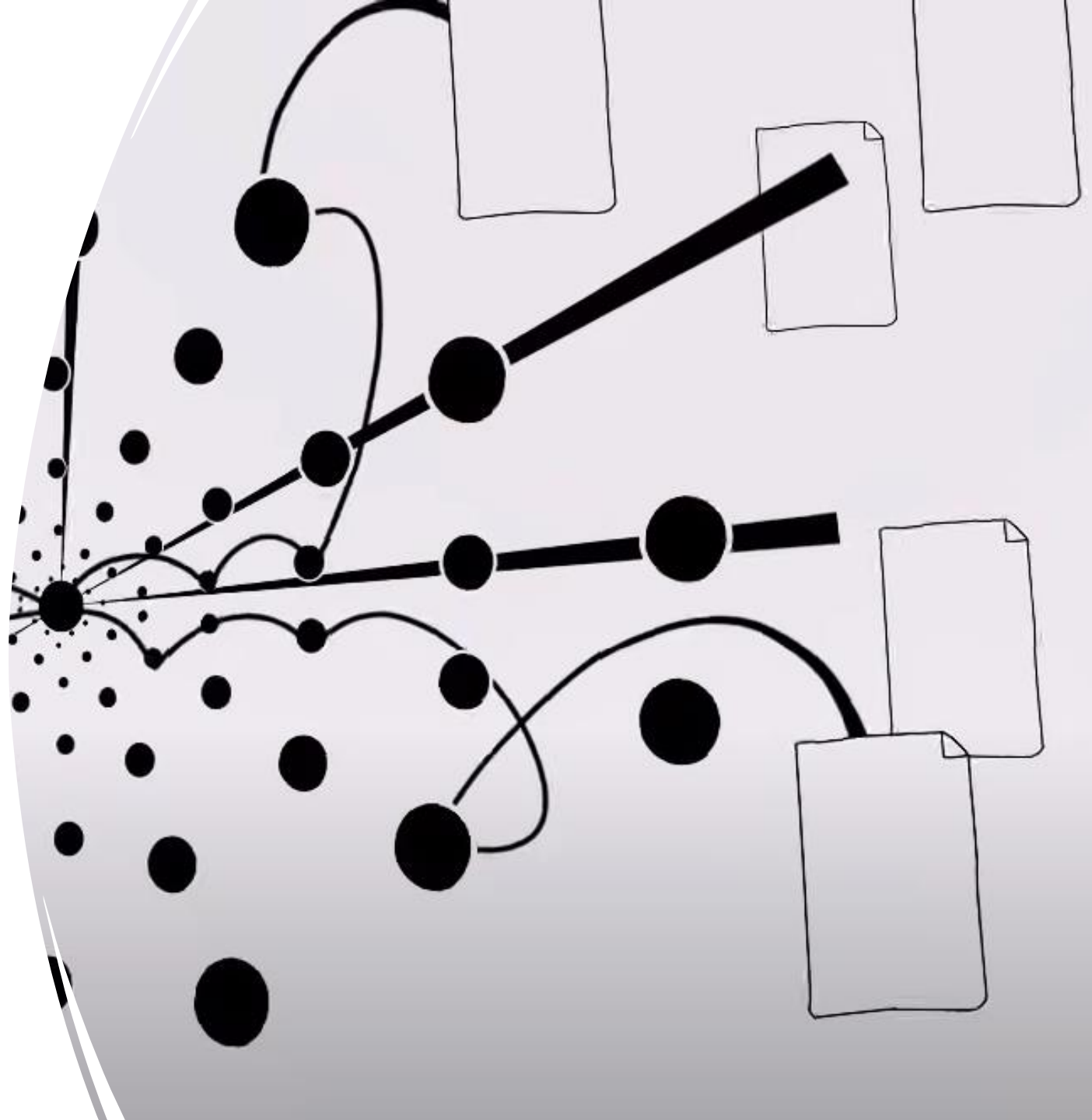
- Buscando por exemplo "escola" e "integral". O google vai buscar a lista de páginas "escola" e a lista "integral". E vai devolver as páginas que estejam nessas duas listas ao mesmo tempo;

COMO ELE ORGANIZA AS PÁGINAS QUE ELE ENCONTRA?

- Ele vai ordenar as páginas de acordo com:
 - Quantas vezes a palavra buscada é repetida nessa página?
 - Onde a palavra buscada ocorre dentro da página? No título? No texto?
 - Existem sinônimos para essa palavra?
 - Se são mais de 1 palavra buscada, elas aparecem uma do lado da outra?
 - Essa página é relevante? Page Rank?

PAGE RANK

- Esta é a grande inovação do google!



O QUE IMPORTA QUANDO FAZEMOS UMA BUSCA?

- Toda palavra importa;
 - Exemplo: "who", "the who", "a who";
- Ordem importa;
 - Exemplo: "blue sky", "sky blue";
- Letras maiúsculas não importam;
- Símbolos não importam (tem algumas exceções, ex.: "C#" ou "Google+").

QUAL COMIDA A ELEVEN MAIS GOSTA?

- Estamos procurando "a comida que a Eleven mais gosta". Então, a palavra "qual" e o sinal "?" não adiciona nada de útil na informação que estamos procurando, podemos removê-la;
- "comida a Eleven mais gosta"

"COMIDA A ELEVEN MAIS GOSTA"

- "Comida que mais gosta" poderia ser substituída por "comida favorita";
- "comida favorita Eleven"

"COMIDA FAVORITA ELEVEN"

- Estamos falando da personagem de Stranger Things. Então, seria bom adicionar algum contexto à nossa busca;

"COMIDA FAVORITA ELEVEN STRANGER THINGS"