

Visualização dos resultados do Yahoo em nuvens de texto: uma aplicação construída a partir de web services

Visualizing of Yahoo results in text clouds: a web services-based application

nuvens de texto, visualização de informações, design da informação

Este artigo apresenta os resultados de uma pesquisa que avaliou as vantagens da utilização da nuvem de texto, uma técnica de visualização de informações, para apresentar os resultados de um sistema de busca na web. A hipótese dessa investigação é que a visualização dos resultados de um sistema de busca em nuvens de textos pode auxiliar os usuários a encontrar o que procuram, facilitando a construção de consultas, em buscas exploratórias. Para testar essa hipótese, a pesquisa aqui relatada foi desenvolvida em duas fases. A primeira consistiu no desenvolvimento de uma aplicação integrada ao sistema de busca Yahoo para mostrar seus resultados em nuvens de textos, e a segunda, na avaliação dessa aplicação com a finalidade de testar seus benefícios.

Text clouds, visualization of information, information design

This article presents the results of a survey that evaluated the advantages of using text clouds, a technique for visualization of information, used to present the results of a search system in the Web. The assumption of this investigation is that by visualizing the results of a search system in text clouds may help users to find what they are looking for, making easier the development of queries, in exploratory searches. To test this assumption, the search we report here was developed in two stages. The first stage consisted of the development of an application integrated to the Yahoo search system to show the results in text clouds, and, the second, consisted of the evaluation of this application in order to test its benefits.

1. Introdução

Os sistemas de busca estão entre as páginas mais acessadas na web hoje em dia. São eles que permitem que os usuários encontrem as informações que procuram em meio a uma infinidade de documentos. No entanto, com o crescimento exponencial de informações disponíveis na rede, as consultas apresentam cada vez mais resultados. Estes aparecem listados de forma textual e seqüencialmente em inúmeras páginas, exigindo dos usuários um esforço extra no sentido de refinar e localizar o que procuram. a ser, um assunto do Design. (Bonsiepe, 2000, citado em Meirelles & Moura, 2007)

A área de buscas na web tem sido objeto de pesquisa sob vários enfoques, com contribuições de diversos campos do saber. As pesquisas têm se dado principalmente no âmbito da Interação Humano-Computador, Ciência da Computação, Recuperação de Informações e Mineração de Dados.

Ainda quanto à questão da abundância de informações na web, e a relevância dessa pesquisa para a área, Meirelles e Moura (2007) ressaltam a posição de Bonsiepe quanto à importância do design como mediador entre usuário e informação, no sentido de viabilizar a aquisição de conhecimento e o seu compartilhamento:

conhecimento como experiência acumulada precisa ser comunicado e compartilhado entre indivíduos. O processo de comunicar e compartilhar está ligado à apresentação do conhecimento e esse é, ou pode vir

¹ Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP.

² Globo. com

³ Escola Superior de Desenho Industrial - ESDI

Apesar de os sistemas de busca oferecerem recursos de pesquisa avançada que possibilitam um refinamento das consultas, pesquisas demonstram que os usuários além de não utilizarem esses recursos, usam em média apenas duas ou três palavras-chave por consulta, e na maioria das vezes não olham além da primeira página de resultados (Silverstein et al., 1999; Hoscher & Strube, 2000; Jansen et al., 2000, 2005; Wolfram et al., 2001; Spink et al., 2001, 2002).

Os principais sistemas de busca adotam algoritmos próprios para localizar e listar os resultados por ordem de relevância. Esses algoritmos vêm sendo constantemente aprimorados e atendem bem ao seu propósito quando os usuários têm objetivos bem definidos (Bates, 1989). No entanto, quando os usuários têm conhecimentos reduzidos sobre o assunto da consulta, objetivos pouco definidos ou muito complexos, eles não sabem que palavra-chave utilizar. Para contornar esse problema, geralmente adotam estratégias que envolvem utilizar palavras-chave genéricas como uma tentativa de obter nos resultados palavras-chave mais específicas para uma nova consulta (Baldonado & Winograd, 1997; Bates, 1989; Pirolli & Card, 1995).

Esse tipo de estratégia utilizada pelos usuários demanda tempo e esforço cognitivo para se chegar ao objetivo pretendido e tem recebido especial atenção por parte da comunidade científica. Foco de diversos estudos, esse comportamento é classificado como busca exploratória (Marchioni, 2006). Análises recentes sobre o objetivo das consultas sugerem que cerca de 20 a 30% de todas as consultas realizadas na *web* são exploratórias por natureza (Rose & Levinson, 2004).

Quando os usuários iniciam uma busca exploratória a falta de uma visão geral dos resultados é particularmente problemática (Baldonado & Winograd, 1997; Bates, 1989; Pirolli & Card, 1995; Dumais et al., 2001; Teevan et al., 2004). É essa visão geral que propicia uma visualização dos resultados de uma consulta em contexto.

A partir da constatação da importância do contexto na localização de resultados de busca, surgiram recentemente propostas de sistemas, ainda em suas versões beta, que utilizam técnicas de visualização de informações para mostrar graficamente os seus resultados. Técnicas de visualização de informações atendem ao objetivo de aumentar a compreensão dos resultados comunicando informações contextuais através de variações na forma como os dados são mostrados e também apresentam soluções para a disposição de grandes quantidades de resultados de busca em uma única tela. Dois exemplos dessa nova geração de sistemas de busca são o Grokker⁴ e o KartOOVISU⁵

Figura 1. Grokker



⁴ <http://www.grokker.com>

⁵ <http://www.beta.kvisu.com>

Figura 2. KartOOVIsu



A abordagem desses sistemas parte da categorização automática de documentos. Apesar de bastante promissora, a categorização automática é complexa e nem tudo está resolvido. Alguns problemas observados na visualização gráfica dos resultados desses sistemas muitas vezes vêm de uma categorização mal estruturada e não da visualização propriamente dita.

Esses sistemas se encontram ainda em fase de aprimoramento, no entanto, abrem caminho para novas experimentações como a apresentada por esta pesquisa.

A aplicação proposta por esta pesquisa apresenta os resultados de um sistema de busca em contexto através da sua visualização em nuvens de texto. Esse benefício é intrínseco às técnicas de visualização de informações em geral, e a abordagem adotada por esta pesquisa é puramente estatística.

De forma resumida, a motivação para a realização desta pesquisa surgiu da soma das seguintes constatações:

- Os sistemas de busca têm importância crescente na localização de informações na *web*;
- Os usuários de sistemas de busca utilizam apenas duas ou três palavras-chave por consulta e, na maioria das vezes, não olham além da primeira página de resultados;
- Os sistemas de busca não atendem bem nos casos de busca exploratória, e estas correspondem a até um terço de todas as consultas realizadas na *web*;
- Existe uma nova geração de sistemas de busca que utilizam técnicas de visualização para mostrar seus resultados, porém de forma ainda bastante embrionária;
- Ainda existe espaço para a experimentação de novas formas de visualização de resultados de sistema de busca.

2. Nuvens de texto x *tag clouds*

A nuvem de texto é uma forma de visualização de dados lingüísticos que mostra a frequência com que as palavras aparecem em determinado texto.

Na visualização padrão de uma nuvem as palavras são dispostas seqüencialmente em ordem alfabética, e apresentam tamanhos variados de uma mesma fonte, diretamente proporcionais ao número de vezes que aparecem no texto.

Recentemente ela se tornou bastante popular na *web* e o seu sucesso se deve, em parte, ao grande uso em redes sociais⁶. Nessas redes, as nuvens são compostas por *tags*, palavras-chave atribuídas livremente pelos usuários, para classificar determinadas páginas. Logo, uma *tag cloud*, como é conhecida, envolve taxonomias criadas e compartilhadas pelos usuários.

⁶ Redes sociais são grupos de pessoas que têm interesses em comum. Na *web* elas utilizam como ponto de encontro e canal de comunicação sites, fóruns de discussão ou listas de e-mails.

As nuvens de texto são compostas, basicamente, como seu nome indica, por textos e têm como objetivo principal proporcionar uma compreensão rápida, um resumo do conteúdo de determinado texto ou conjunto de textos, a partir de suas palavras mais freqüentes.

3. Níveis de informação de uma nuvem de texto

As nuvens atendem ao objetivo de aumentar a compreensão comunicando informações contextuais através de variações na forma como os dados são visualizados, proporcionando a visualização de dimensões adicionais da informação de forma a tornar o contexto explícito.

Os dados extraídos dos textos e comunicados por uma nuvem nada mais são do que uma listagem de palavras, e o número de vezes que elas aparecem em um texto com um diferencial qualitativo proporcionado pela forma de visualização. Uma nuvem de texto é uma lista hierarquizada visualmente.

Quando esses dados são apresentados em forma de nuvem, é possível perceber a importância de determinada palavra em comparação ao todo, no caso, o número total de palavras. Essa informação adicional comunica a importância semântica, ou o contexto das palavras mostradas. Uma nuvem de texto é uma proposta visual para comunicar relações importantes e dimensões adicionais de significados dentro das limitações de um espaço plano.

Uma nuvem é uma visualização filtrada de um universo multidimensional e pode mostrar relações mais complexas do que aquelas apresentadas por uma listagem que possui apenas uma dimensão.

Essa capacidade de mostrar mais de uma dimensão permite que uma nuvem seja o reflexo da estrutura do texto que ela representa. Dessa forma, elas não se restringem simplesmente a representar uma enumeração das palavras de um determinado conteúdo, que é a principal conquista de uma lista associada a valores.

Ainda, segundo Bertin (1986), os dados por si sós não fornecem a informação. É necessário ver as relações que o conjunto de dados estabelece. A informação útil para a decisão é dada pelas relações de conjunto.

O objetivo de uma representação gráfica é apresentar um nível superior da informação. Uma nuvem de texto permite que sejam visualizados outros conjuntos e relações entre os dados que não podem ser visualizados em uma listagem.

Bertin (1986) afirma que: “A percepção visual é espacial, e permite qualquer um utilizar um novo princípio de reclassificação: a tomada em consideração, simultânea, de muitos elementos”.

Nuvens de texto e *tag clouds* também podem ser usadas como sistemas de acesso e navegação na *web*.

A diferença entre a navegação e a leitura baseadas em nuvens e as tradicionais em listas, como os menus verticais à esquerda das páginas e os resultados de sistemas de busca, é que as nuvens não são lineares e não são desenhadas para serem consumidas de forma linear.

A forma como a nuvem é apresentada permite múltiplos pontos de entrada usando diferenciadores visuais como cor e tamanho da fonte tipográfica para atrair a atenção do usuário. Dessa maneira, as nuvens diminuem o esforço visual e cognitivo por parte do usuário já que este não tem que percorrer uma série de itens até encontrar o que deseja. Uma nuvem permite que o usuário vá direto a qualquer ponto de interesse, e destaca as palavras que são possivelmente mais relevantes em razão de sua alta freqüência.

Da mesma forma que a estrutura das listas encoraja interações lineares, interfaces baseadas no uso de nuvens encorajam diferentes tipos de comportamentos de busca e localização de informações. É importante ressaltar que nada impede que as nuvens também sejam lidas de forma linear, da esquerda para a direita, no padrão ocidental de leitura.

Dois estudos recentes de avaliação de *tag clouds* apresentados por Halvey e Keane (2007) e Rivadeneira et al. (2007) confirmaram a leitura não linear das nuvens.

Figura . Nuvens de textos geradas a partir da transcrição dos discursos dos candidatos à eleição presidencial do partido democrata americano em 27 abr. 2007.⁷

Senator Clinton



Senator Obama

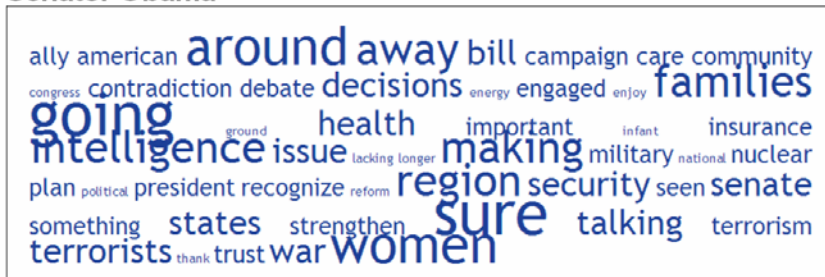
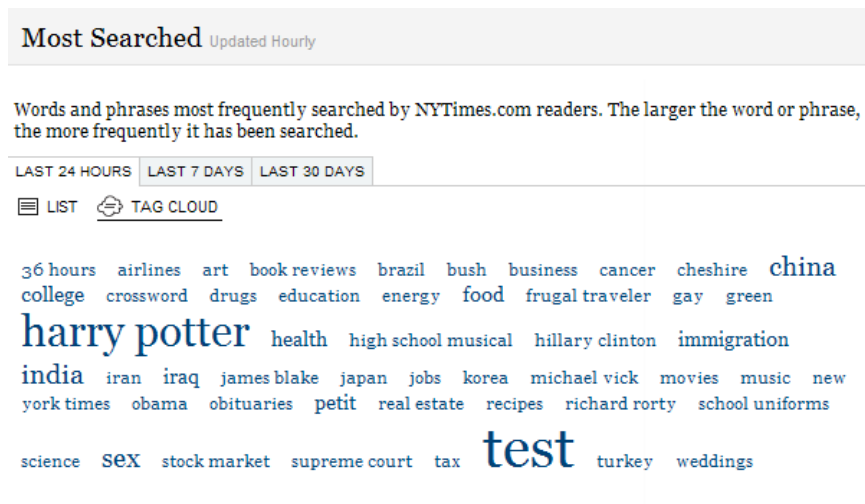


Figura 4. Nuvem das palavras mais frequentes que foram pesquisadas na versão online do jornal *The New York Times*⁸ no dia 28 jul. 2007.



4. Visualização de dados lingüísticos

A área de visualização de informações é uma área multidisciplinar que estuda formas e técnicas de representação visual para revelar tendências e relações entre dados estatísticos levando à percepção de uma nova dimensão da informação (Ware, 2000).

Como já foi dito, a nuvem de texto é uma forma de visualização de dados lingüísticos. Um texto é uma seqüência de dados codificados que chegam ao receptor de maneira mais lenta do que uma imagem. No entanto, autores (Triesman & Gormican, 1988; Duncan Humphreys, 1989; Chaud & Yeh, 1995, citado em Ware, 2000) defendem que técnicas de visualização de

⁷ [http:// www.pollster.com](http://www.pollster.com)

⁸ <http://www.nytimes.com>

informações podem ser muito efetivas, principalmente para codificar padrões e relações nos dados lingüísticos usando as propriedades gráficas pré-conscientes⁹ Assim, as conexões podem ser exploradas de forma mais rápida, sem que sejam necessários o tempo e a atenção requeridos na leitura.

A motivação para estudos na área de visualização de dados lingüísticos vem da constatação de que os leitores da internet são mais impacientes que os leitores de materiais impressos:

A impaciência do leitor digital vem da cultura, não da natureza da tela. Os usuários de sites têm expectativas diferentes dos usuários de impressos. Eles querem sentir-se “produtivos”, não contemplativos; não querem processar, querem buscar; esperam ser desapontados, distraídos e atrasados por pistas falsas. (Lupton, 2006)

Segundo Jacob Nielsen (2000) os usuários da internet não gostam de ler, q (...) Querem continuar movendo-se e clicando.

5. Fase 1 – Desenvolvimento da aplicação

A teoria

Em uma consulta na *web*, os resultados aparecem listados em diversas páginas. Através de uma nuvem de textos é possível a visualização de uma síntese, de um resumo automático, do conteúdo dos resultados listados em várias páginas sem que elas tenham de ser percorridas e os sites acessados individualmente.

Figura 5. Construção da nuvem de texto dos resultados.



⁹ Estudos das capacidades visuais indicam que existe um processamento pré-consciente de informações relacionadas à forma que é automático. Propriedades como cor, tamanho, contraste, linhas e, orientação são visualizadas por esse processo (Ware, 2000).

A nuvem, nesse contexto, funciona como uma ferramenta auxiliar para que o usuário possa gerenciar a grande carga de informação que é disponibilizada nos resultados das consultas.

É formada a partir de uma consulta inicial em que cada resultado corresponde a uma página. A partir do conteúdo de cada página, é construída uma nuvem de texto com suas palavras mais freqüentes. A nuvem dos resultados corresponde à soma das nuvens das principais páginas listadas pelo sistema de busca segundo seu algoritmo de relevância.

O desenvolvimento da aplicação

Os designers devem conhecer melhor os serviços online e pensar além da aparência do site, buscar a implementação de APIs e organização de conteúdo, que não devem ser tratados como blocos desvinculados e autônomos, mas sim pensados como um todo. (Meirelles & Moura, 2007)

A primeira fase desta pesquisa consistiu no desenvolvimento de uma aplicação que permite a visualização dos resultados do sistema de busca Yahoo ¹⁰ A aplicação foi desenvolvida na linguagem Java a partir de um serviço do sistema ¹¹, disponibilizado livremente e de forma gratuita.

A aplicação desenvolvida funciona como um *plugin* para o Yahoo, Ele interage com o navegador e permite que certas funções sejam executadas sob demanda.

Para gerar a nuvem de textos a partir de determinada consulta, a aplicação desenvolvida varre o conteúdo das quarenta primeiras páginas retornadas no *set* de resultados do Yahoo. Extrai o texto da linguagem de marcação e demais códigos de programação e cria um índice com as palavras e suas respectivas ocorrências.

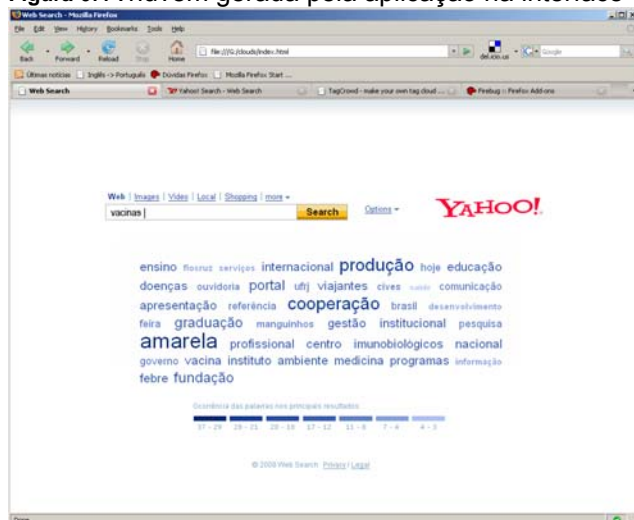
O índice da aplicação é gerado pelo Apache Lucene ¹² um software de busca de código aberto que possui uma API ¹³ de indexação de documentos.

A listagem inicial é filtrada e são eliminadas palavras freqüentes da língua portuguesa, como artigos e preposições, e finalmente ela é tabulada no formato de uma nuvem de textos. Essa nuvem de textos é construída através de Javascript e formatada por um arquivo CSS, *Cascading Style Sheet*, que é uma linguagem de estilo utilizada para definir a apresentação de documentos escritos em uma linguagem de marcação, como HTML. Seu principal benefício é prover a separação entre a forma e o conteúdo de um documento.

O funcionamento da aplicação

A aplicação é acionada quando um usuário digita uma ou mais palavras no campo de busca do Yahoo e pressiona a barra de espaço. Automaticamente, uma nuvem de texto aparece na sua página principal, antes mesmo que a consulta seja submetida em definitivo.

Figura 6. A nuvem gerada pela aplicação na interface do Yahoo.



¹⁰ <http://www.yahoo.com>

¹¹ <http://developer.yahoo.com>

¹² <http://lucene.apache.org>

¹³ Uma Application Programming Interface é composta por uma série de funções que são disponibilizadas por um programa principal para que outros programas e novas funcionalidades possam ser desenvolvidos a partir dos seus serviços.

A partir dessa nuvem, o usuário pode obter sugestões de palavras-chave para complementar a consulta inicial e então submetê-la ao sistema, ou ainda, apenas pressionar a barra de espaço para obter uma nova nuvem. Para adicionar uma palavra da nuvem à consulta, basta clicar sobre a palavra desejada para que esta automaticamente seja incluída no campo de busca ao lado da palavra inicial. Cada vez que uma palavra nova é adicionada à consulta e a barra de espaço é pressionada, a nuvem é reformulada.

A construção gráfica da nuvem na aplicação

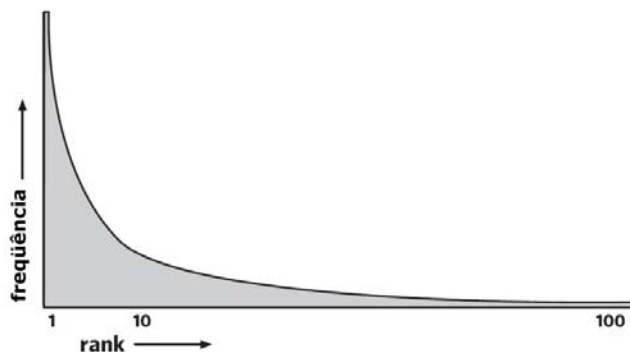
Integrar fisicamente forma e conteúdo talvez tenha sido o impulso mais persistente da arte e do design do século XX. No século XXI, forma e conteúdo voltaram a ser separados. Folhas de estilo, por exemplo, impelem os designers a pensarem global e sistematicamente ao invés de se concentrarem na construção fixa de uma superfície particular. (Lupton, 2006)

Na aplicação desenvolvida para esta pesquisa, seguimos essa tendência. Nela, forma e conteúdo são separados. O conteúdo da nuvem vem de uma base de dados e muda a cada consulta. A forma é toda predefinida em uma única folha de estilo e, tem que ser pensada para todas as possíveis variações que esse conteúdo possa apresentar.

Lei de Potência

Um padrão observado em nuvens de texto e *tag clouds* é conhecido como lei de potência. Uma lei de potência implica que pequenas ocorrências são muito comuns ao contrário de grandes instâncias que são muito raras, apesar de existirem.

Figura 7. Típica curva de uma lei de potência.



O predomínio da lei de potência influencia em muitas das escolhas relacionadas ao design da nuvem. A alta frequência de algumas poucas palavras e a baixa frequência de muitas têm de ser compensadas (Smith, 2007).

Na construção de uma nuvem de texto, a primeira decisão a ser tomada é a de quantas palavras irão fazer parte da sua composição, decisão relacionada com o espaço destinado na interface para a exibição da nuvem. Optamos por mostrar quarenta palavras com alinhamento justificado e destinamos uma área de 650 pixels de largura para a exibição da nuvem. A altura é variável, pois o número de linhas varia em função do tamanho das palavras.









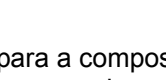
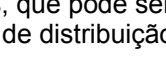
A escala da nuvem

A escolha da escala é uma questão estética e matemática, e é preciso chegar a uma fórmula que crie a melhor nuvem para atender ao seu propósito.

A escala adotada para a nuvem de texto na aplicação foi tamanho de fonte de 12 pixels para o patamar mínimo, e de 36 pixels para o patamar máximo. O equilíbrio entre tamanho máximo e mínimo de fonte foi observado para que as palavras com o tamanho máximo não se sobreponham às do tamanho mínimo. A entrelinha também é de extrema importância na composição desse equilíbrio, e para essa aplicação foi adotada entrelinha de 36 pixels.

Uma vez definidos os patamares máximo e mínimo, a escala foi dividida em sete patamares para os quais foi associada uma escala de valor de azul.

Figura 8. Escala adotada na nuvem da aplicação.

Tamanho da fonte em pixels	Tamanho da fonte Arial	Escala de valor de azul	Valores hexadecimais
12	palavra		# ACC1F3
13	palavra		# 86A0DC
14	palavra		# 607EC5
15	palavra		# 395CAE
16	palavra		# 264CA2
18	palavra		# 133B97
21	palavra		# 072771
24	palavra		
30	palavra		
36	palavra		

Distribuição das freqüências

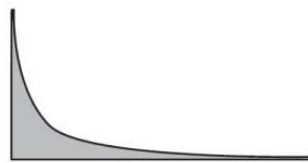
Outro aspecto de extrema importância para a composição visual de uma nuvem é a distribuição das freqüências entre esses patamares, que pode ser feita segundo vários métodos. Para a construção de uma nuvem os métodos de distribuição mais relevantes são o proporcional e o linear (Smith, 2007).

Na **distribuição proporcional**, o tamanho das palavras¹⁴ é diretamente proporcional à sua freqüência. Logo, é feita uma distribuição entre a freqüência máxima e a freqüência mínima pelo número de patamares escolhido para a nuvem. Essa distribuição gera uma nuvem que reflete a curva típica de lei de potência com poucas palavras de tamanho grande, e muitas de tamanho pequeno.

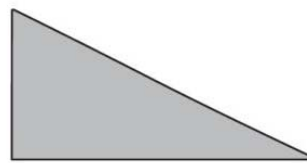
Outro método para a distribuição das freqüências é a **distribuição linear**. Neste método a curva de lei de potência é transformada em uma reta através de uma função logarítmica. Essa mudança faz que a diferença entre a maior e a menor palavra passe a ser linear em vez de exponencial.

¹⁴ Neste artigo, para fins didáticos a expressão “tamanho da palavra” é adotada de forma genérica para se referir ao tamanho do corpo da fonte tipográfica da palavra.

Figura 9. A distribuição proporcional pode resultar em poucas palavras grandes e muitas pequenas. A distribuição linear levanta o meio da distribuição, suavizando as diferenças.



DISTRIBUIÇÃO PROPORCIONAL
O tamanho das palavras é diretamente proporcional às suas freqüências.



DISTRIBUIÇÃO LINEAR
O tamanho das palavras é baseado no logaritmo das suas freqüências.

A Tabela 1 mostra uma lista de palavras com suas respectivas freqüências. O tamanho direto é o tamanho atribuído segundo uma escala proporcional. O tamanho linear é o tamanho atribuído para essas palavras segundo a função logarítmica da escala linear. Como se pode ver na última coluna, as palavras com freqüência moderada ficam maiores usando a distribuição linear.

Tabela 1. Comparação do tamanho das palavras segundo as escalas proporcional e linear

Palavra	Freqüência	Tamanho (direto)	LOG	Tamanho (linear)
design	120	48	4.78	48
web 2.0	43	24	3.76	38
internet	34	22	3.52	36
cultura	12	15	2.48	26
colaboração	4	12	1.39	15

A diferença entre uma nuvem de texto com distribuição proporcional e uma com distribuição linear pode ser visualizada nas figuras 10 e 11, ambas geradas a partir da mesma lista de palavras e freqüências (Tabela 2).

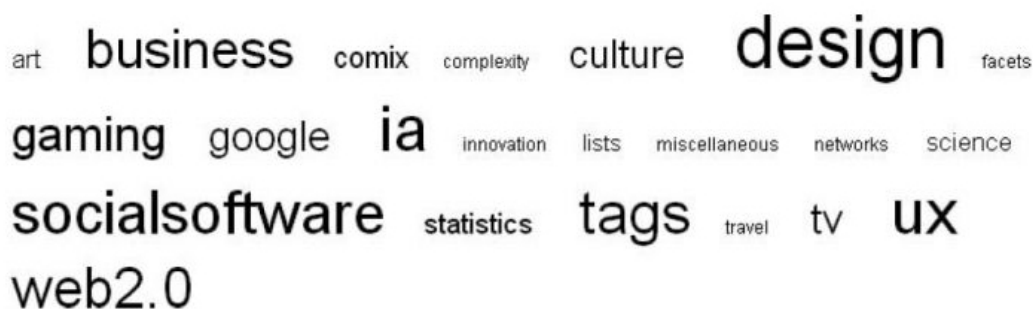
Tabela 2. Lista de palavras e respectivas freqüências que serviu de base para a geração das nuvens nas figuras 10 e 11

Palavra	Freqüência	Palavra	Freqüência	Palavra	Freqüência
design	120	gaming	23	science	5
ux	68	google	14	lists	4
ia	65	tv	14	innovation	3
socialsoftware	54	culture	12	miscellaneous	3
tags	46	comix	7	complexity	3
web2.0	43	statistics	6	facets	3
business	34	art	5	networks	3

Figura 10. Nuvem com distribuição proporcional.



Figura 11. Nuvem com distribuição linear.



De acordo com Smith (2007), a distribuição linear gera nuvens mais atrativas e mais fáceis de ler. Outro benefício aparece nos casos em que a nuvem oferece algum tipo de interação ou é usada como mecanismo de navegação. Nesse caso, as palavras ficam maiores e são clicadas mais facilmente.

Como a lista de palavras mais freqüentes de um resultado de busca segue a lei de potência, optamos por adotar a distribuição linear na aplicação desenvolvida.

Finalmente, escolhemos a fonte digital Arial por fazer parte do sistema operacional Windows, uma vez que a aplicação desenvolvida utiliza as fontes do sistema operacional do computador do usuário.

6. Fase 2 – Avaliação da aplicação

A partir da construção dessa aplicação foi possível realizar duas formas de avaliação, a partir de métodos empíricos, para as quais foram selecionados dez participantes.

Avaliação cooperativa

Primeiro foi feita uma avaliação cooperativa. Por se tratar de um conceito novo, essa avaliação inicial foi usada para registrar a compreensão do aplicativo da nuvem de texto pelos participantes. Também atendeu ao objetivo de familiarizar esse participante com o aplicativo para o experimento controlado.

De acordo com Monk et al. (1993), em uma avaliação cooperativa os participantes desempenham tarefas previamente formuladas e, juntos, com o pesquisador, avaliam a usabilidade de determinado sistema. Uma de suas principais vantagens é a possibilidade de se trabalhar com protótipos e simulações parciais, caso do aplicativo avaliado nessa pesquisa.

Foi proposta, como tarefa, a realização de duas consultas de caráter exploratório utilizando a aplicação desenvolvida.

Experimento controlado

Depois, foi conduzido um experimento controlado que comparou as mesmas consultas sendo realizadas no Yahoo padrão, que serviu de controle, e no Yahoo com o auxílio da nuvem de texto.

Para o experimento controlado, utilizaram-se os mesmos dez participantes da avaliação cooperativa distribuídos em dois grupos homogêneos, A e B.

Como tarefa, todos os participantes tiveram de responder a duas questões de caráter exploratório utilizando o sistema de busca Yahoo. Para cada questão só existia uma resposta correta, que poderia ser obtida em diferentes sites e a partir de palavras-chave variadas.

Os cinco participantes do grupo A responderam à questão 1 utilizando o sistema de busca Yahoo padrão, e a questão 2 utilizando a versão do Yahoo com o aplicativo da nuvem de texto. Para o grupo B, as questões foram invertidas. Dessa forma, cada questão foi respondida 10 vezes, 5 em cada versão do Yahoo.

Nesse experimento não houve a interferência do pesquisador. Os participantes receberam as instruções por escrito, e as questões foram entregues em cartões separados.

Ambas as avaliações foram realizadas no mesmo ambiente físico, utilizando o mesmo computador, com as mesmas configurações e a mesma velocidade de conexão com a internet.

Esse experimento controlado verificou especificamente:

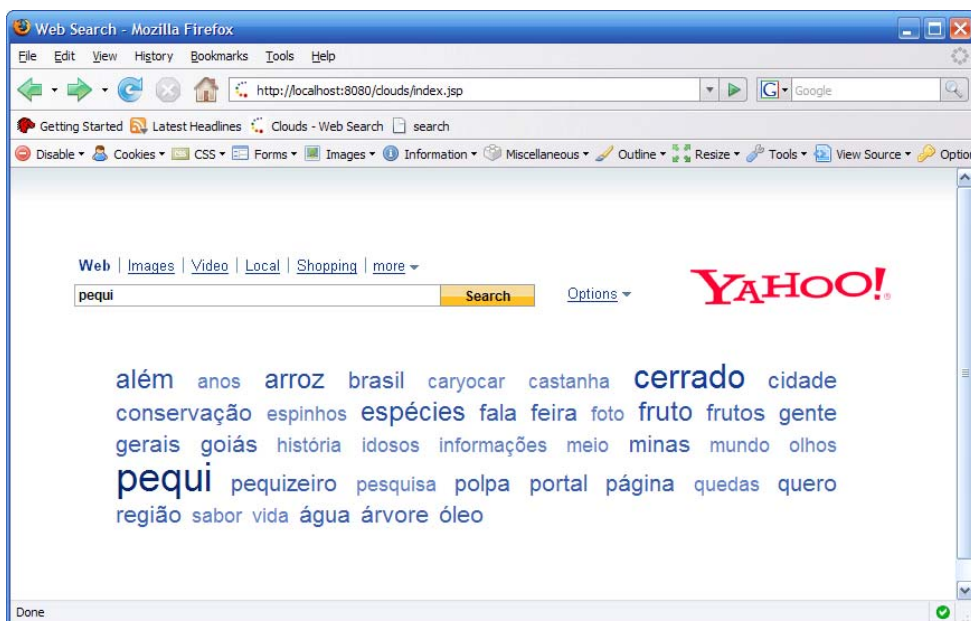
- se os usuários utilizaram palavras da nuvem de texto para refinar suas consultas;
- nos casos positivos, se as buscas foram concluídas de forma mais rápida e satisfatória com o auxílio da nuvem, comparadas às buscas realizadas no controle.

7. Resumo dos resultados

Pontos positivos

- A aceitação da aplicação foi muito boa por todos os participantes envolvidos na pesquisa, e sua utilização foi compreendida com facilidade.
- Constatou-se que a tarefa influencia diretamente na utilidade e na eficácia da nuvem.
- O aproveitamento das nuvens está ligado diretamente ao conhecimento geral do usuário e requer um conhecimento prévio mínimo em relação ao assunto da consulta. Esse aspecto é positivo em alguns casos e negativo em outros. Alguns participantes não apresentavam nenhum conhecimento prévio, o mínimo requerido, para a obtenção da resposta das questões. Nesses casos, mesmo a resposta aparecendo na nuvem, ela não foi identificada. Já nos casos em que os participantes tinham apenas uma vaga idéia da resposta, a nuvem ajudou, pois foi possível, a partir das palavras sugeridas, fazer algumas associações.

Figura 12. Nuvem de resultados para a consulta pelo termo PEQUI.



- As nuvens se mostraram eficazes nos casos em que as respostas são objetivas e compostas por apenas um termo. Caso da questão 1, por exemplo, cuja resposta era CERRADO. (Questão: O pequi é uma fruta típica de que vegetação?)
- Foi identificado como positivo, também, o fato de a nuvem apresentar os resultados de forma resumida e ainda informações adicionais sobre o contexto do assunto.
- As nuvens apresentam mais contexto nas consultas compostas por apenas um ou dois

termos, mas isso atende à maioria das consultas.

- Nas consultas cujas respostas são mais complexas as nuvens têm sua eficácia diminuída. Nesses casos, a nuvem não apresenta a resposta, apenas apresenta novas palavras em contexto, para serem adicionadas à consulta inicial.
- De maneira geral, os participantes foram unânimes em informar que as nuvens no mínimo ajudaram, sugerindo palavras-chave para refinar ou reformular suas consultas.
- A utilização da nuvem reduziu o esforço cognitivo dos usuários e o número de páginas visitadas até a obtenção dos resultados.

Pontos negativos

- As consultas compostas por palavras-chave que são associadas a diversos assuntos, ou seja, a diferentes universos semânticos, não apresentam nenhum contexto específico.
- O tempo de conclusão das tarefas não foi menor com o auxílio das nuvens, na maioria dos casos.

8. Conclusões

Podemos concluir, a partir desta pesquisa, que a visualização dos resultados de um sistema de busca em uma nuvem de texto pode auxiliar os usuários a encontrar o que procuram, facilitando a construção das consultas. Porém, a eficácia da nuvem está ligada diretamente à tarefa em questão e, também, ao conhecimento prévio do usuário em relação ao assunto da consulta.

A conclusão das tarefas nem sempre foi mais rápida com o auxílio das nuvens, contudo, a sua utilização reduziu o esforço cognitivo dos usuários e o número de páginas visitadas até a obtenção dos resultados. Observou-se, também, que o tempo percebido foi menor do que o tempo cronológico.

É importante ressaltar que as avaliações iniciais realizadas contaram como um número reduzido de participantes e, além disso, a ausência de familiaridade desses participantes com o uso da aplicação suscita questões com relação à curva de aprendizagem que devem ser consideradas.

Entretanto, foi possível constatar que o desenvolvimento de aplicações para a visualização de resultados de sistemas de busca é algo bastante desafiador, e diante do entusiasmo com que a aplicação foi recebida pode-se dizer que essa área é, de fato, um caminho que deve continuar sendo explorado.

A visualização dos resultados em nuvens deve ser vista não como uma alternativa, mas sim como mais uma possibilidade, mais uma ferramenta disponível, da qual os usuários podem tirar proveito para realizar suas consultas.

Referências

- Baldonado, M. Q. W.; Winograd, T. (1997). SenseMaker: An information-exploration interface supporting the contextual evolution of a user's interests. In: Conference on User Interface and Software Technology. *Proceedings...* S.l: s.n. p.11-18.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, v.13, n.5, p.407-424.
- Bertin, J. (1986). *Neográfica e o tratamento gráfico da informação*. Curitiba: Ed. UFPR.
- Dumais, S.; Cutrell, E.; Chen, H. (2001). Optimizing search by showing results in context. In: SIGCHI Conference on Human Factors in Computing Systems, Seattle (WA). *Proceedings...* New York: ACM Press. p.277-284.
- Halvey, M.; Keane, M. (2007). An assessment of Tag Presentation Techniques. In: WWW2007, Banff, Alberta, Canada, 2007. *Proceedings...*

- Hoscher, C.; Strube, G. (2000). Web search behavior of Internet experts and newbies. In: International World Wide Web Conference on Computer Networks. The International Journal of Computer and Telecommunications Networking. 9. *Proceedings...* North-Holland Publishing Co. p.337-346.
- Jansen, B. J.; Spink, A.; Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management*, v.36, n.2, p.207-227.
- Jansen, B. J.; Spink, A.; Pedersen, J. (2005). A temporal comparison of AltaVista Web searching: research articles. *Journal of the American Society for Information Science and Technology*, v.56, n.6, p.559-570.
- Lupton, E. (2006). *Pensar com tipos*. São Paulo: CosacNaify.
- Marchioni, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, v.49, n.4, p.41-46.
- Meirelles, J. C.; Moura, M. (2007). Web 2.0: novos paradigmas projetuais e informacionais. *Infodesign – Revista Brasileira de Design da Informação*, v.4, n.2, p.12-19.
- Monk, A. F.; Wright, P. C.; Davenport, L.; Haber, J. (1993). *Improving your human-computer interface: a practical technique*. Prentice Hall Practitioner series.
- Nielsen, J. (2000). *Projetando websites*. Rio de Janeiro: Campus.
- Pirolli, P.; Card, S. (1995). Information foraging in information access environments. In: SIGCHI conference on Human factors in computing systems. *Proceedings...* New York: ACM Press; Addison-Wesley. p.51-58.
- Rivadeneira, A.; Gruen, D.; Muller, M.; Millen, D. (2007) Getting our head in the clouds: toward evaluation studies of tagclouds. In: SIGCHI conference on Human factors in computing systems 2007. San Jose, California, USA. *Proceedings...* p.995-998.
- Rose, D. E.; Levinson, D. (2004). Understanding user goals in web search. In: International Conference on World Wide Web, 13. *Proceedings...* New York: ACM Press. p.13-19.
- Silverstein, C.; Marais, H.; Henzinger, M.; Moricz, M. (1999). Analysis of a very large Web search engine query log. *ACM SIGIR Forum*, v.33, n.1, p.6-12.
- Smith, G. (2007). *Tagging: people-powered metadata for the social web*. Pearson Education, New Riders, Peachpit Press.
- Spink, A.; Wolfram, D.; Jansen, M. B. J.; Saracevic, T. (2001). Searching the web: the public and their queries. *Journal of the American Society for Information Science and Technology*, v.52, n.3, p.226-234.
- Spink, A.; Jansen, B. J.; Wolfram, D.; Saracevic, T. (2002). From e-sex to e-commerce: web search changes. *IEEE Computer Society*, v.35, n.3, p.107-109.
- Teevan, J.; Alvarado, C.; Ackerman, M. S.; Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. In: SIGCHI conference on Human factors in computing systems, Vienna, Austria. *Proceedings...* p.415-422.
- Ware, C. (2000). *Information visualization: perception for design*. s.l.: Morgan Kaufmann.
- Wolfram, D.; Spink, A.; Jansen, B. J.; Saracevic, T. (2001). Vox Populi: the public searching of the web. *Journal of the American Society for Information Science and Technology*, v.52, n.12, p.1073-1074.

Sobre os autores

Márcia S. Lunardi

Mestre em Design pela Esdi/Uerj (2008) na linha de pesquisa Design e Tecnologia e Bacharel em Publicidade e Propaganda pela PUC-Rio. Dez anos de experiência na área de design e desenvolvimento de websites, portais corporativos e intranets. Participação nos projetos desde a

conceituação, atuando na arquitetura da informação, prototipação, design de interfaces e implementação. Atualmente trabalha na Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP.

lunardi@pobox.com

José Muanis F. de Castro

Arquiteto de Sistemas e Desenvolvedor com mais de dez anos de experiência. Há seis anos trabalhando na Globo.com, atua no momento como *Scrum Master*, coordenando projetos de sistemas de internet.

muanis@corp.globo.com

André S. Monat

Doutor em engenharia de sistemas e computação pela University of East Anglia. Mestre na mesma área pela Coppe/UFRJ. Engenheiro pelo Instituto Tecnológico de Aeronáutica (ITA). É professor da Uerj desde 1994 e desde 2000 da Escola Superior de Desenho Industrial (Esdi). Atualmente é professor do programa de Mestrado em Design da Esdi na linha de pesquisa design e tecnologia.

andresmonat@yahoo.com.br