

Instituto Superior de Engenharia do Porto

VISUALIZAÇÃO COM NUVENS DE TAGS

Joana Filipa Pinto dos Santos Silva

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática

Área de especialização em

Sistemas Gráficos e Multimédia

Orientador: Mestre Isabel de Fátima Silva Azevedo

Co-Orientador: Prof. Doutor Carlos Miguel Miranda Vaz de Carvalho

Júri:

Presidente:

Doutora Maria de Fátima Coutinho Rodrigues

Vogais:

Mestre Nuno Filipe Fonseca Vasconcelos Escudeiro

Mestre Isabel de Fátima Silva Azevedo

Doutor Carlos Miguel Miranda Vaz de Carvalho

Porto, Novembro 2010

Agradecimentos

Agradeço a minha família, em especial aos meus pais, por terem sempre conseguido reunir ao longo dos anos as condições necessárias para que eu chegasse até aqui, pelos valores e disciplina inculcados que me ajudaram neste longo percurso. Obrigada pelo apoio e confiança.

Ao meu irmão Ricardo, pelo apoio quando foi necessário mais um par de olhos para as leituras do documento.

Ao Prof. Doutor Vaz de Carvalho, um obrigado pelas sugestões, inclusive na escolha da dissertação. Ao GILT, na pessoa do seu responsável, o Prof. Doutor Vaz de Carvalho, agradeço a oportunidade de investigação, utilização de recursos vários, bem como o suporte financeiro que permitiu a apresentação de um artigo sobre o tema explorado no trabalho de mestrado descrito neste documento.

Um agradecimento especial à orientadora, Eng.^a Isabel Azevedo, não só pelo acompanhamento e pela disponibilidade, como pelas elucidações, discussões sobre a condução do trabalho a desenvolver e contributo no aperfeiçoamento deste documento.

Agradeço à Rute Matos e ao Arsénio Ferraz pela colaboração prestada no melhoramento dos textos escritos na língua inglesa.

Aos meus amigos, pela boa disposição, apoio constante durante a escrita da dissertação e pelos inúmeros momentos de descontração. Um obrigado por sempre poder contar convosco em todos estes anos.

Ao meu namorado Bruno, pelo apoio, incentivo a nunca desistir, paciência, dedicação e constante partilha de ideias, um muito obrigado.

Aos meus pais e ao Bruno.

Resumo

Com a evolução da Internet houve uma conseqüente transformação da forma como os resultados são apresentados e das técnicas de visualização e apresentação de conteúdos.

A presente dissertação é sobre visualização com nuvens de tags e foi a conjugação de diversos factores que originou a investigação sobre este tema. Uma nuvem de tags (em inglês: *tag clouds*) é uma concepção visual de um conjunto de tags isoladas com uma representação ilustrativa da sua importância, ou seja visualmente hierarquizadas e em que cada tag se hiperliga a um ou mais sítios Web ou documentos.

São extremamente úteis para a navegação ou para a descoberta de informação genérica, podendo requerer menor carga cognitiva do utilizador durante consultas e pesquisas se alguns aspectos forem considerados na sua construção. Nesta dissertação discutem-se alguns dos factores visuais que podem contribuir para que as nuvens de tags sejam ferramentas efectivas e intuitivas para os utilizadores.

O conceito de folksonomia está relacionado com nuvens de tags, permitindo categorizar as definições atribuídas a conteúdos recorrendo a palavras-chave (tags). Neste documento são ainda exploradas as suas vantagens e desvantagens.

Discutem-se ainda alguns modos de visualização e apresentação de nuvens de tags, desde o tipo de ordenação ao algoritmo de geração da nuvem de tags. Aborda-se ainda o impacto do tipo de fonte e a utilização de cores monocromáticas ou coloridas que estas possam ter na visualização da nuvem. A própria tag também é categorizada e explorada ao nível do seu posicionamento na nuvem e a percepção que causa no utilizador.

Sobre a visualização de nuvens de tags, refere-se ainda que as conclusões obtidas foram utilizadas num repositório de objectos educativos.

Palavras-chave: Folksonomia, visualização, nuvem de tags, redes sociais

Abstract

Along with Internet evolution there was a consequent transformation on how the results are presented, as well as on the visualization and presentation techniques of content generated by user research.

The present work is about tag clouds and a combination of several factors led to research on this subject. Tag clouds are visual representation of a set of tags whose font size is dependent on its importance, which is visually graded and each tag is linked to one or more websites or documents.

Tag clouds are extremely useful for navigation or to find generic information which may require less cognitive load of the user during navigations if some aspects are considered in their construction. This dissertation brings to light the visual factors that may contribute to tag clouds to be considered as an effective and intuitive tool for users.

The concept of folksonomy is related with tag clouds and allows categorizing content by using keywords. In this document their advantages and disadvantages are explored.

We also explore further visualization modes and different types of tag clouds, through the type of ordination to the algorithm used to generate the tag cloud. Furthermore, it is studied the impact of different font and color usage, in the visualization of the cloud. The tag itself is also categorized and explored at their location in the cloud and the perception that causes in the user.

About this subject we also refer the conclusions obtained were used in a repository of learning objects.

Keywords: Folksonomy, visualization, tags clouds, social networks

Índice

Agradecimentos	i
Resumo.....	v
Abstract.....	vii
Índice	ix
Índice de figuras.....	xi
Índice de tabelas	xiii
Notação e glossário.....	xv
Capítulo I. Enquadramento.....	1
1.1 Introdução.....	1
1.2 Objectivos	3
1.3 Motivações.....	4
1.3.1 Motivações pessoais	4
1.3.2 Motivações institucionais	4
1.4 Convenções de escrita.....	6
1.5 Metodologia de investigação.....	7
1.5.1 Estudos de casos	9
1.6 Organização do relatório.....	9
Capítulo II. Nuvens de tags	11
2.1 Tag	11
2.1.1 Categorias de tags	11
2.2 Nuvens de tags.....	13
2.2.1 Utilizações distintas das nuvens de tags	16
2.2.2 Ênfase emocional nas nuvens de tags	24
2.2.3 Outras formas de visualização	26
2.3 Folksonomia	29
2.3.1 Vantagens e desvantagens	31
Capítulo III. Construção de nuvens de tags.....	37

3.1 Algoritmos para o cálculo do peso de cada tag.....	37
3.1.1 Algoritmos Linear e Logarítmico	38
3.1.2 Algoritmo de selecção	39
3.2 Posicionamento, fonte e ordenação das tags.....	46
3.2.1 Posicionamento das tags.....	47
3.2.2 O tipo de fonte “Serif”	47
3.2.3 Ordenação das tags na nuvem.....	49
3.3 Análise de alguns estudos de caso.....	50
3.3.1 Caso 1 - Leitura e apresentação de nuvens de tags.....	51
3.3.2 Caso 2 - Percepção de nuvens de tags	51
3.3.3 Caso 3 - Legibilidade das fontes aplicadas às tags	54
Capítulo IV. Estudo experimental	57
4.1 Contextualização	57
4.2 Tecnologias e Ambientes de desenvolvimento usados	58
4.3 Implementação	59
4.3.1 Processo de implementação	60
4.4 Explicação da experiência	61
4.5 Avaliação do estudo.....	65
Capítulo V. Conclusões.....	71
5.1 Avaliação do trabalho realizado	71
5.2 Limitações.....	73
5.3 Trabalho futuro	74
Bibliografia	77
Anexo A – Inquérito realizado.....	83
Anexo A.1. Apresentação	83
Anexo A.2. Perguntas abertas	84
Anexo A.3. Perguntas abertas	85
Anexo A.4. Finalização	86

Índice de figuras

Figura 1 - 43things	2
Figura 2 - Flickr com tags na lateral direita das fotos.....	2
Figura 3 - Parte da página principal de acesso ao repositório	6
Figura 4 - Nuvem de tags geralmente colocada na parte lateral dos sítios Web.....	14
Figura 5 - Nuvem de tags com utilização cuidada do tamanho da fonte	14
Figura 6 - Nuvem de tag que não é apenas determinada pelo tamanho da fonte mas também pela cor utilizada.	15
Figura 7 - Nuvem de tags com aproveitamento das fontes grandes.	15
Figura 8 - Exemplo nuvem de tags.....	17
Figura 9 - Nuvem de tags 3D.....	18
Figura 10 - Exemplo de nuvem de tags Web 2.0	18
Figura 11 - Nuvem de tags (em lista) para visualizar movimento de acções.....	19
Figura 12 - Nuvem de tags utilizada no Jornal de Noticias	19
Figura 13 - Nuvem de tags de Steve Jobs na MacWorld 2007	20
Figura 14 - Nuvem de tags do discurso de Bill Gates nas CES 2007	21
Figura 15 - Número de ocorrências das palavras que foram utilizadas por cada 25 mil palavras ditas	23
Figura 16 - Nuvem de tags normal (à direita) e nuvem de tag emocional	24
Figura 17 - Visão geral dos <i>layouts</i> de nuvens de tags existentes: <i>Layout</i> pubcloud.....	27
Figura 18 - Visão geral dos <i>layouts</i> de nuvens de tags existentes: tabelas HTML.....	27
Figura 19 - Visão geral dos <i>layouts</i> de nuvens de tags existentes: estrutura gráfica	27
Figura 20 - Visão geral dos <i>layouts</i> de nuvens de tags existentes: mapa elástico de tags ...	28
Figura 21 - Visão geral dos <i>layouts</i> de nuvens de tags existentes: <i>layout</i> circular	28
Figura 22 - Sítio Web Delicious Web bookmarking.....	32
Figura 23 - Folksonomia com inflexões	33
Figura 24 - Folksonomia com sinonímia	33
Figura 25 - Tags controladas (Buzzillions.com)	35
Figura 26 - Sugestões de tags / tags pré-definidas.....	35
Figura 27 - Top 21 da selecção de palavras do passo 1.....	40
Figura 28 - Frequência das tags por ordem de ocorrência	40
Figura 29 - Top 21 da selecção de palavras determinadas pelo número de ocorrências e por ordem alfabética.....	41
Figura 30 - Gráfico com número de ocorrências das palavras seleccionadas	41
Figura 31 - Visualização logarítmica com limitação de peso (uma cor).....	44

Figura 32 - Visualização Linear com limitação de peso (uma cor)	44
Figura 33 - Visualização logarítmica com limitação de peso (colorida)	45
Figura 34 - Visualização linear com limitação de peso (colorida).....	45
Figura 35 - Visualização logarítmica Vs. Visualização Linear (sem condicionante, monocromática).....	46
Figura 36 - Visualização logarítmica Vs. Visualização Linear (sem condicionante, colorida) 46	
Figura 37 - Exemplo de fontes <i>Sans-Sarif</i> e <i>Serif</i>	48
Figura 38 – Visualização logarítmica colorida ordenada alfabeticamente	49
Figura 39 - Visualização Linear monocromática ordenada alfabeticamente	50
Figura 40 - Frequência da utilização dos diferentes tamanhos das fontes.....	50
Figura 41 - <i>Layout</i> das diversas nuvens de tags utilizadas no estudo	52
Figura 42 - Dataset Delicious disponibilizado por ftp	59
Figura 43 - Algumas das nuvens de tags utilizadas na experiência	61
Figura 44 - Nuvens apresentadas ao participante para que este fornecesse resposta as questões apresentadas no inquérito.....	63
Figura 45 - Exemplo de trabalho futuro	75

Índice de tabelas

Tabela 1 - Exemplos de categorias de tags.....	12
Tabela 2 - Percentagem de participantes que seleccionaram nuvens "emocionais".....	25
Tabela 3 - Percentagem de participantes que seleccionaram nuvens "emocionais" com classificação "Interesse alto" do conteúdo do sítio Web.	26
Tabela 4 - Quadro resumo	29
Tabela 5 - Diferenças entre algoritmos linear e logarítmico	39
Tabela 6 - Dados para realização das experiências dos estudos apresentados.....	51
Tabela 7 - Sumário dos resultados da pesquisa sobre percepção de nuvens de tags e desempenhos contrastantes com trabalhos relacionados.	53
Tabela 8 - Nuvens apresentadas na experiência.....	61
Tabela 9 - Resultados do inquérito realizado com respectivos tempos consumidos por nuvem na pesquisa pela palavra solicitada pelo sistema.....	62
Tabela 10 - Respostas dos participantes quando comparavam diferentes nuvens de tags ..	63
Tabela 11 - Tempos médios registados.....	64
Tabela 12 - Tempos médios gastos nas nuvens coloridas e monocromáticas independentemente do algoritmo	64
Tabela 13 - Votos dos participantes classificados por grau de experiência com nuvens de tags às questões da nuvem 9.....	66
Tabela 14 - Tempos gastos em cada nuvem por grau de experiência na utilização de nuvens de tags	67
Tabela 15 - Resultado da análise T-Test.....	68
Tabela 16 - Tempos gastos divididos por intervalo de idades.....	68
Tabela 17 - Anova: factor único.....	69

Notação e glossário

CAD – Computer aided design

CASPOE – Caracterização Semântica e Pragmática de Objectos Educativos

CES - Consumer Electronics Show

CSS – Cascading Style Sheets

GILT – Graphics, Interaction & Learning Technologies

MELT - Metadata Ecology for Learning and Teaching

PDA – Personal Digital Assistant

TREE - Teaching Resources for Engineering Education

URL – Uniform Resource Locator

Capítulo I. Enquadramento

Neste primeiro capítulo faz-se o enquadramento desta dissertação de mestrado. Apresentam-se as razões que estiveram na sua génese, os seus objectivos, e explica-se a organização de todo este documento, sendo apresentados, de forma breve, os capítulos e os anexos que o compõem.

1.1 Introdução

Os blogues e as redes sociais tendem em atrair grandes atenções de vários tipos de público. A natureza colaborativa de diversos serviços tem aumentado nos últimos anos e os utilizadores têm aderido a esta realidade, sendo frequente a incitação à publicação de anotações, informações e notícias.

Erik Qualman menciona que a utilização de redes sociais tornou-se a actividade mais popular na Internet (Qualman, 2009). Segundo o autor, as pessoas sentem vontade de actualizar constantemente uma espécie de diário *online* para se sentirem ligadas ao resto do mundo e deste modo serem aceites pela sociedade.

As mentalidades das pessoas e cultura estão a mudar em consequência deste fenómeno. A publicidade que passa nas televisões já não é tão poderosa como há alguns anos atrás. Nomeadamente, um simples comentário, ou actualização do estado de um utilizador numa rede social podem-se revelar mais eficientes: “*The 30-second commercial is being replaced by the 30-second review, tweet, post, status update, and so on*” (Qualman, 2009).

Actualmente as pessoas tendem a valorizar mais o comentário de um amigo de uma rede social do que um resultado obtido num motor de pesquisa em relação a qualquer produto que pretendam pesquisar, seja um destino de férias, seja um atendimento num restaurante: “*People care more about what their friend and peers think is the best Italian restaurant in Manhattan than what Google thinks*” (Qualman, 2009).

A natureza social e colaborativa de sistemas bastantes populares como o **Delicious**¹, o **Flickr**², o **43things**³, o **Furl**⁴ e o **Technorati**⁵, têm aumentado a popularidade de um determinado tipo de visualização, as nuvens de tags (*tag clouds*, na nomenclatura anglo-

¹ <http://del.icio.us>

² <http://www.flickr.com>

³ <http://www.43things.com>

⁴ <http://www.furl.net>

⁵ <http://www.technorati.com>

saxónica). Estas são representações visuais baseadas em texto de um conjunto de tags cujo tamanho da fonte está dependente da sua importância. As imagens seguintes (Figura 1 e Figura 2) exemplificam nuvens de tags utilizadas no sistema 43things e Flickr, respectivamente.

43 Things is the world's largest goal-setting community. People here want to:

Learn to cook lose 40 pounds weigh 110 pounds learn how to twirl drumsticks calm down pick strawberries run a marathon Sans domicile fixe wants to Make a soundtrack for my life get skinny get a Dachshund Play Come Sail Away on the Piano you Be a beach bum moisturize daily improve my spoken english send a postcard to Post Secret make someone smile every day Quit Smoking watch south of nowhere JulieJordanScott wants to Write a best selling book sell my idea to google learn to use chopsticks learn about wine learn guitar again restore a classic car Find an old friend become a flight attendant get a job lose 20 pounds in two weeks Create a mash up song JustCoolTom wants to organize my finances take belly dancing lessons find inner peace get better at yoga **lose weight** reunite with an old friend Be just as compassionate towards myself, as I try to be to others Have cybernetic implants atultayade wants to Pay off all of my debt find my passion get a new nickname that everyone calls me instead of my real name explore be an au pair for a year volunteer for something memorize all the U.S. presidents, in order the oc season 1 get Hannah Montanas phone number!!! duckyhead wants to Learn to be alone without feeling lonely paint the porch work because I like to, not because I have to Rent my house out learn to fight with a sword **stop procrastinating** do a split finish my novel Learn how to properly throw kunai and shuriken Finish my daughter's scrapbook go to Niagra Falls buy eyeliner watch all episodes of Captain Tsubasa J start an open source project make a video for youtube laugh more often painting_piglet wants to Sketch Daily Forgive, but not forget Write more be less materialistic Become a good photographer go spelunking cubalibre wants to learn yoga have a white christmas Visit Ireland **go on a road trip with no predetermined destination** Read more books free myself go hiking be closer to my siblings go on a cruise Dance with the Pussycat Dolls I Want A Good Job Witness a miracle decide what I want to do with my life Drastically change my appearance lhatetopaint wants to stop buying clothes have a diaper free baby make a difference practice yoga daily Raise funds for Cancer Research Volunteer at a hospital live life to the best i can Work on a cruize ship create my own website Learn to defend myself Watch all the seasons of The OC go to vegas blood sisters Learn to type without looking at the keyboard Schaap wants to Read 50 books this year go skinny dipping be a more open person live in Malibu brew my own beer win a tony Finish watching Rurouni Kenshin eman4629 wants to be a bartender watch casablanca from beginning to end switch from coffee to green tea buy a gun live with integrity Read the entire Bible

Figura 1 - 43things

(Fonte: <http://www.43things.com>)



Starburst Cluster Shows Celestial Fireworks

Esta foto foi tirada em 8 de julho de 2010.

3.351 22 113 4

Esta foto pertence a

[Galeria de NASA Goddard Photo and Video](#) (1899)

Esta foto também aparece em

- [Explored!](#) (álbum)
- [Hubble Space Telescope](#) (álbum)
- [Featured Images](#) (álbum)

Tags

Star • Nebula • Space • Hubble Space Telescope • HST • NASA • Goddard Space Flight Center

Informações adicionais

- Upload feito usando Flickr Uploadr 3.0 (Mac)

Licença

Figura 2 - Flickr com tags na lateral direita das fotos

(Fonte: <http://www.flickr.com/photos/gsfrc/4774472286/in/photostream/>)

As tags são extremamente úteis para a navegação ou para a descoberta de informação genérica, requerem menor carga cognitiva do utilizador durante consultas e pesquisas, conforme já mencionado anteriormente por alguns investigadores:

- *“These keywords, which are referred to as ‘tags’ on the site, allow users to describe and organize content with any vocabulary they choose”* (Mathes, 2004),
- *“In order to formulate a query, a participant must recall relevant words to enter into the search box - a cognitively more expensive task than recognizing relevant words in the list. A number of the comments from the exit survey indicated that the tag cloud ‘suggested’ useful search terms, saving the participant some cognitive effort (...) It appears that scanning the tag cloud requires less cognitive load than formulating specific query terms (...) When individual tags link to relevant subsets of repository content, tag clouds also become powerful instruments of topical browsing”* (Sinclair, et al., 2007),
- *“Tag clouds occupy a peculiar niche in the domain of visualization. Designers have to balance aesthetic considerations, for example with regard to font size and tag arrangement, against the capabilities of algorithms to produce optimized layouts”* (Seifert, et al., 2008).

1.2 Objectivos

Esta dissertação de mestrado é sobre visualização com *nuvens de tags* e há um conjunto de objectivos genéricos e específicos que se pretenderam atingir.

O objectivo genérico definido para este trabalho foi o seguinte:

- Estudar os conceitos de folksonomia, tags, nuvens de tags, com uma caracterização da utilização dos mesmos em vários sítios Web.

Os objectivos específicos foram:

- Determinar qual o algoritmo mais apropriado para apresentação de nuvens de tags num repositório de recursos educativos, com a análise dos seguintes aspectos:
 - Influência do tamanho da fonte para determinar a relevância da tag num contexto,
 - Tipos de letra que melhor se adequam a nuvens de tags, segundo diferentes aspectos (por exemplo, fontes de letras espaçadas ou condensadas, as que facilitam a legibilidade ao utilizador), qual a melhor disposição, alfabética ou aleatória, tamanho da fonte ou estilo ou conjugação de todos estes factores

referidos, exemplo: fontes mais largas indicam que são tags recentes, tags que aparecem em negrito diferenciam as mais populares.

- Estudar e comparar algoritmos de construção de nuvens de tags capazes de organizar de forma intuitiva os dados provenientes de uma folksonomia, potenciando assim a redução de carga cognitiva do utilizador no processo de pesquisa de informação.

1.3 Motivações

Nesta secção discutem-se as motivações pessoais (secção 1.3.1) e institucionais (secção 1.3.2) que estiveram na génese deste trabalho.

1.3.1 Motivações pessoais

Foram diversas as razões que conduziram à realização deste trabalho:

- Tendo na parte lectiva do mestrado realizado um trabalho sobre nuvens de tags, na óptica de implementação apenas, considerou-se francamente oportuna a hipótese de aprofundar o estudo do tema, considerando outros factores,
- Foi também vista como aliciante a possibilidade de realizar um trabalho numa área ainda pouco explorada em Portugal e verificar se os resultados de estudos conduzidos noutros países seriam similares aos obtidos num estudo realizado maioritariamente com pessoas portuguesas,
- Saber que o resultado do estudo serviria para definir o *layout* de uma nuvem de tags a utilizar num repositório, trouxe ainda uma motivação adicional para a realização deste trabalho.

1.3.2 Motivações institucionais

O trabalho desenvolvido nesta dissertação está relacionado com a colaboração activa dos utilizadores, mais especificamente em como torná-la visível num repositório, de forma eficaz no que diz respeito também a pesquisa e recuperação de informação. O desenvolvimento deste projecto teve lugar no centro de investigação GILT⁶ no âmbito de um projecto de investigação.

O grupo de investigação **GILT - Graphics, Interaction and Learning Technologies** é uma estrutura localizada no Instituto Superior de Engenharia do Instituto Politécnico do Porto, orientada para a análise, construção, aprofundamento e difusão do conhecimento técnico e

⁶ <http://gilt.isep.ipp.pt>

científico nos domínios da Realidade Virtual, Multimédia, Interacção, Tecnologias de Aprendizagem e áreas afins.

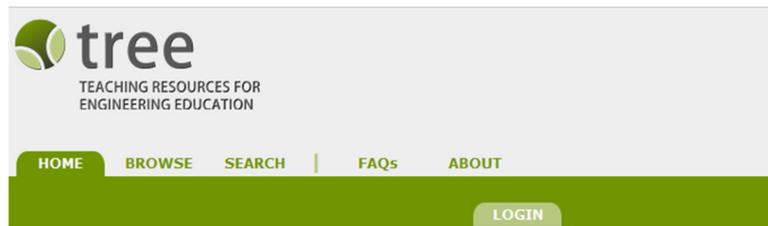
O GILT teve origem por consequência de dois factores, primeiro deveu-se ao trabalho de investigadores do Instituto Superior de Engenharia nas áreas de Tecnologias de Aprendizagem no âmbito da Unidade de Ensino a Distância (1997 - 2000) e no âmbito do Laboratório de e-Learning (2001 - 2005), o segundo factor tem em conta a experiência científica prévia no grupo de Computação Gráfica e CAD (Computer Aided Design) do INESC-Porto, o qual permitiu muitos investigadores seniores do grupo colaborarem neste projecto.

O GILT origina-se então por intermédio da conjugação dos factores descritos anteriormente e da necessidade de constituir um grupo com as competências necessárias para o desenvolvimento de trabalhos de qualidade. Para atingir esta meta, o GILT procura estabelecer as bases para um trabalho sistemático de Investigação e Desenvolvimento, articulando-o com acções de pós-graduação, nomeadamente ao nível de Mestrado e Doutoramento e com iniciativas comuns com outras Instituições, públicas ou privadas, nacionais ou estrangeiras.

No Instituto Superior de Engenharia do Porto (ISEP), no âmbito do projecto CASPOE – Caracterização Semântica e Pragmática de Objectos Educativos (PTDC/EIA/65387/2006), tem sido desenvolvido um repositório de objectos educativos (ROE) com funcionalidades avançadas, designado TREE (Teaching Resources for Engineering Education) (Seiça, et al., 2009) (Azevedo, et al., 2010). Um ROE é um repositório de objectos e seus metadados. Este permite que alunos, professores ou qualquer utilizador efectuem pesquisas, de acordo com diversos campos de metadados, e o acesso aos recursos armazenados.

O objectivo primordial do projecto CASPOE é promover a reutilização dos objectos educativos e definir um modelo para um repositório, contendo objectos educativos e padrões genéricos de aprendizagem utilizados como templates para os objectos educativos de menor granularidade.

O TREE (ver Figura 3) apesar de ser um repositório de recursos que podem ser utilizados em cursos de engenharia é um repositório bastante heterogéneo, pois abrange diferentes áreas do conhecimento, como Direito, Linguística, Matemática e Ciências da Computação, entre outras. Neste momento o repositório está numa Intranet apenas acessível a estudantes, professores e funcionários da instituição.



Welcome to Instituto Superior de Engenharia do Porto's institutional digital repository

Quick Search TREE 

Use [Advanced Search](#) to search by multiple criteria.

Did you know that you can customize your TREE RSS feeds by multiple search criteria? [What is RSS?](#)

Figura 3 - Parte da página principal de acesso ao repositório

A utilização de folksonomias foi considerada uma forma de obter informações adicionais sobre os recursos, capturando o ponto de vista das pessoas que utilizam o repositório, mas também devido à capacidade desta forma de estruturação de conhecimento fornecer funcionalidades interessantes e intuitivas aos seus utilizadores.

Um projecto inovador na área de e-learning, designado Metadata Ecology for Learning and Teaching (MELT)⁷, decorreu entre 1/10/2006 e 31/03/2009 e estudou o impacto da utilização de tags por parte dos professores, com resultados muito promissores. As tags foram posteriormente analisadas por especialistas, e verificou-se que em 25% dos casos, as tags continham valor adicional para o repositório, enquanto em outros casos, muitas das tags já estavam de alguma forma representadas em descritores de metadados. No seu relatório final é dito que a potencialidade das folksonomias está sub-explorada (Assche, et al., 2009).

1.4 Convenções de escrita

Com o propósito de permitir uma leitura agradável deste documento, foram adoptadas as seguintes convenções:

⁷ http://info.melt-project.eu/ww/en/pub/melt_project/welcome.htm

- As palavras inglesas aparecem em itálico, excepto as palavras *tag* e *tags*, pela frequência com que são utilizadas e por serem usualmente utilizadas na linguagem portuguesa.
- Sempre que se utiliza um acrónimo pela primeira vez menciona-se o seu significado, mas na secção Notação e Glossário está disponível uma lista de todos os acrónimos e abreviaturas utilizadas neste documento.
- A fonte *Courier New* é utilizada quando se apresentam extractos de código, com o objectivo de diferenciá-los convenientemente do restante texto;
- Citações aparecem entre aspas e em itálico, e na língua original, de forma a respeitar integralmente as palavras proferidas.
- É utilizada a seguinte sequência de caracteres (...) quando parte do que foi afirmado por alguém não é citado na sua totalidade.

1.5 Metodologia de investigação

Nesta secção é descrita a metodologia adoptada para o desenvolvimento das fases que contemplaram o estudo experimental assim como alguns procedimentos verificados na literatura abordada.

Durante o processo de investigação foram estudados diversos artigos sobre a construção de nuvens de tags, o que permitiu auferir as conclusões obtidas e confrontar com os resultados adquiridos no estudo desenvolvido.

O estudo experimental desenvolvido para esta dissertação está dividido em fases que são seguidamente apresentadas. No entanto, todos os procedimentos efectuados para a realização deste trabalho de investigação são justificados posteriormente no capítulo 4.

O objectivo do estudo, realizado no âmbito desta dissertação, era conhecer as opiniões dos participantes sobre o tema visualização de nuvens de tags quando confrontados com situações reais, com solicitação de escolha de determinadas tags e o registo do tempo gasto na pesquisa da mesma.

Para atingir o objectivo que foi proposto neste estudo, foi elaborado um questionário. Com vista a obtenção de dados para a concretização do objectivo deste estudo, foi seleccionada a técnica de inquérito por questionário.

Esta técnica é vulgarmente utilizada em pesquisas que pressupõem análise quantitativa de dados uma vez que a sua estrutura é padronizada, tanto no texto das questões como na

ordem apresentada (Borg, et al., 2002). Outros factores que determinaram a escolha desta técnica foram:

- O tempo reduzido na obtenção de dados,
- A não influência do investigador nas respostas dos participantes no estudo (Bravo, 1991).

No entanto, esta técnica tem uma desvantagem, o facto de o investigador não estar presente no momento da participação do indivíduo que está a responder ao inquérito, não permite ao participante solicitar esclarecimento sobre perguntas que possam levar a interpretações dúbias (Ghiglione, et al., 1997). Para evitar este tipo de problemas, e para que os dados pudessem ter credibilidade, as perguntas devem ser bem formuladas e inequívocas (Soares, 2004).

O questionário I integrou questões de resposta directa, bem como de escolha múltipla. Neste último tipo de questão foi solicitada a apresentação de uma justificação ou razão para a escolha efectuada, para dar a possibilidade ao respondente de clarificar a sua opinião/escolha sobre o assunto em causa. Desta forma foi possível obter informação mais “rica” e detalhada sobre o motivo da escolha dos inquiridos (Soares, 2004).

Uma investigação empírica pressupõe uma recolha de dados. Por vezes, a investigação centra-se em grupos muito alargados de sujeitos, mas para o estudo realizado considerada apenas uma parte constituinte da população, designada por amostra, o que é habitual neste tipo de estudos.

No estudo realizado considerou-se para a amostra maioritariamente alunos do Mestrado de Engenharia Informática do ramo Sistemas Gráficos e Multimédia do Instituto Superior de Engenharia do Porto. Os dados recolhidos permitiram tirar conclusões e extrapolá-las.

Os dados foram recolhidos entre o mês de Maio e Agosto de 2010. O inquérito foi respondido *online* pelos participantes sem qualquer intervenção do investigador.

Os dados recolhidos permitiram efectuar uma análise:

- Quantitativa: com o intuito de verificar o número de repostas obtidas em cada ecrã do inquérito e tempo gasto por ecrã/huvm apresentado.
- Qualitativa: das questões abertas, para assim complementar a informação quantitativa.

O estudo desenvolvido neste trabalho de mestrado é de carácter exploratório e descritivo. Como referido anteriormente foi realizado um inquérito com o propósito de recolher dados que permitam descrever comportamentos, atitudes, valores e situações.

Este tipo de estudo descritivo denomina-se *survey*, uma vez que os dados obtidos foram por meio de perguntas feitas aos sujeitos que participaram no estudo sobre os conceitos que se pretendiam analisar, além desta denominação, o estudo também é catalogado com tipo exploratório, dada a novidade do objecto de estudo e consequente existência de reduzida pesquisa prévia no domínio (Coutinho, 2008).

1.5.1 Estudos de casos

Alguns da literatura abordada contemplava estudos de caso sobre temas semelhantes ao abordado nesta dissertação. O termo estudo de caso foi definido como “*uma família de métodos de pesquisa cuja decisão comum é o enfoque numa instância*” (Adelman, 1976). Para outros autores, o estudo de caso consiste na “*observação detalhada de um contexto ou indivíduo, de uma única fonte de documentos ou de um acontecimento específico*” (Bogdan, et al., 1994).

O estudo de caso é um método de investigação qualitativa em que são transportados para casos gerais, factos observados num caso específico. A observação detalhada de um contexto ou indivíduo e de um acontecimento específico (o caso) juntamente com a observação de diversas variáveis, permite a comparação com outros casos.

Num estudo de caso o investigador segue um esquema que depende do enquadramento teórico, das suas finalidades, objectivos e recursos disponíveis.

Os depoimentos de um estudo de caso são utilizados como veículos para a compreensão de aspectos básicos do comportamento humano. Nas secções 3.3.1, 3.3.2 e 3.3.3 são descritos alguns casos de estudo que foram seleccionados e analisados em detalhe neste documento.

1.6 Organização do relatório

Esta dissertação é constituída por 5 capítulos e 1 anexo. O tema principal deste trabalho é a visualização com utilização de nuvens de tags e foi abordado de forma a clarificar os vários conceitos que vão sendo mencionados e com uma estrutura que a seguir se pormenoriza.

No primeiro capítulo (ou seja, neste capítulo) explica-se a importância do tema visualização com nuvens de tags e descreve-se ainda o âmbito do trabalho que foi realizado, fazendo a sua contextualização.

No segundo capítulo apresenta-se uma descrição pormenorizada de vários conceitos relevantes para o trabalho desenvolvido, como folksonomia, tags e nuvens de tags. Este capítulo centra-se no objectivo genérico descrito no 1º capítulo.

No terceiro capítulo é apresentado o processo de construção de nuvens de tags, é exposta a selecção dos algoritmos e descrevem-se alguns modelos e projectos mais marcantes na área de apresentação de nuvens de tags.

O quarto capítulo é sobre o desenvolvimento e implementação do estudo experimental. É feita a contextualização do estudo e as suas etapas principais.

No quinto capítulo são apresentadas as conclusões obtidas após a realização do estudo teórico e prático sobre o tema. Mencionam-se ainda as limitações do estudo, e as melhorias a introduzir no futuro.

Em complemento aos capítulos que constituem este documento existe 1 anexo. Nesse anexo é apresentado o inquérito apresentado aos participantes na iniciativa que serviu para realizar o estudo apresentado nesta dissertação.

Capítulo II. Nuvens de tags

Neste capítulo descreve-se detalhadamente alguns conceitos relevantes para esta dissertação de mestrado, como nuvens de tags e alguns outros relacionados, com um enquadramento histórico da utilização destes termos.

Neste capítulo são também apresentados alguns exemplos de nuvens de tags que são actualmente utilizadas em alguns sítios Web, nas suas diversas formas de apresentação de conteúdos, com variação de *layout*, cores, estilos e fontes.

2.1 Tag

Tag é a palavra inglesa para “rótulo”, “etiqueta”, “marca” e serve como palavra-chave para catalogar determinado elemento multimédia. Permite, assim, obter-se uma classificação dos recursos.

Estas tags normalmente têm origem em descrições atribuídas a conteúdos de sítios Web, fotografias, imagens, vídeos, etc. Qualquer um destes elementos multimédia pode ser categorizado com uma ou mais tags que permitam descrever e auxiliar a identificação do conteúdo.

A partir de meados de 2005 a popularidade das tags cresceu muito o que levou muitos sítios Web passarem a permitir que os utilizadores realizassem anotações ou utilizassem simples palavras, ou pequenas frases, na forma de um texto não-estruturado (Russel, 2006).

A designação atribuída à prática de atribuição de tags de forma colaborativa/livre por qualquer utilizador denomina-se *tagging*, estando o conceito de folksonomia (ver secção 2.3) relacionado com esta prática.

2.1.1 Categorias de tags

O sistema Delicious gere quantidades enormes de tags diariamente, que se tornam também objecto de vários estudos. Golder e Huberman concluíram que as tags atribuídas pelos utilizadores podem ser categorizadas em sete grupos (Golder, et al., 2006), resultantes da natureza das tags atribuídas, sendo estas:

- ‘Temas’ - agregam descrições genéricas,
- ‘Tipos’ - relacionada com o formato do conteúdo/recurso,

- ‘Adjectivos’ - agregam tags que reproduzem um parecer sobre o conteúdo/recurso marcado,
- ‘Qualificação’ - categoria que aperfeiçoa uma outra categoria. A título de exemplo, as tags deste tipo podem ser apenas números,
- ‘Propriedade’ - categoria que agrega tags que identificam o recurso marcado pelo próprio utilizador,
- ‘Auto-referência’ - categoria que os autores utilizam para identificar as tags que foram marcadas pelo utilizador, as quais iniciam por *meu* ou *my*, por exemplo *minhas coisas*, *meuscomentários*, *mystuff*, *mycomments*,
- ‘Organização das tarefas’ - categoria que agrega as tags relacionadas com as tarefas a serem executadas pelo utilizador com o conteúdo/recurso marcado. Estas tags costumam ser referenciadas por *toread* ou *jobsearch*, por exemplo.

No livro “Tagging: People Powered Metadata for the Social Web”, o autor Gene Smith dá a conhecer exemplos de situações reais para cada uma das categorias acima apresentadas e que podem ser vistos na tabela seguinte (Smith, 2008):

Tabela 1 - Exemplos de categorias de tags

Categoria	Exemplo
Temas	Css, webdesign, ajax, drama, zen, music, sushi, halo3
Tipos	Blog, book, video, photo
Adjectivos	Cool, funny, lame, crap, defective by design.
Qualificação	Squaredcircle, seenlive, aka vogan poetry
Propriedade	Nytimes, genesmith (author), newriders
Auto-referencia	Mystuff, mine, me
Organização de tarefas	Toread, todo, work

Sen e colegas reduziram a categorização das tags para apenas 3 classes gerais (Sen, et al., 2006):

- Tags Factuais, que estão relacionadas com factos sobre um conteúdo/recurso.
- Tags Subjectivas, que reflectem as opiniões dos utilizadores sobre os recursos/conteúdos marcados com tags.
- Tags Pessoais, aquelas cujo público é o próprio utilizador que atribui a tag.

2.2 Nuvens de tags

Uma nuvem de tags (em inglês: *tag clouds*) é uma concepção visual de um conjunto de tags isoladas com uma representação ilustrativa da sua importância, ou seja, visualmente hierarquizadas, e em que cada tag se hiperliga a um ou mais sítios Web ou documentos.

As nuvens de tags são formas de organizar palavras-chave e ajudar na visualização e procura, por exemplo, fazendo indexações em pesquisas. Estas estão dispostas proporcionalmente pela relevância fazendo-se notar no tamanho da fonte ou pelas cores destacadas.

As nuvens de tags podem ser representadas de várias formas, como se pode verificar em alguns exemplos ilustrados, considerando o tamanho da fonte, ou pela variação de cores ou formato de árvores:

- Na Figura 4 é mostrada uma nuvem de tags geralmente colocada nas laterais dos sítios Web,
- Na 5 apresenta-se uma nuvem de tags com utilização cuidada do tamanho de fontes para moderar o espaço utilizado num sítio Web,
- Na Figura 6 a nuvem de tags não é apenas determinada pelo tamanho da fonte mas também pela cor utilizada, sendo a cor mais destacada referente às tags com maior importância. As tags com menos relevância são determinadas com cores semelhantes à cor de fundo da página Web,
- Na Figura 7 é apresentada uma nuvem de tags em que as fontes grandes são utilizadas com moderação para aproveitamento do espaço consumido na página Web.



Figura 4 - Nuvem de tags geralmente colocada na parte lateral dos sites Web
(Fonte: http://media.smashingmagazine.com/cdn_smash/images/tagclouds/gearc.gif)



Figura 5 - Nuvem de tags com utilização cuidada do tamanho da fonte
(Fonte: http://media.smashingmagazine.com/cdn_smash/images/tagclouds/wptags.gif)



Figura 6 - Nuvem de tag que não é apenas determinada pelo tamanho da fonte mas também pela cor utilizada.

(Fonte: http://media.smashingmagazine.com/cdn_smash/images/tagclouds/cjung.gif)



Figura 7 - Nuvem de tags com aproveitamento das fontes grandes.

(Fonte: http://media.smashingmagazine.com/cdn_smash/images/tagclouds/webdesignerwall.gif)

No caso do tamanho da fonte, normalmente a relação de maior ou menor tamanho é definida pelo número de vezes que uma tag é utilizada. Desta forma, uma nuvem de tags permite, de uma forma visual e prática, a consulta das tags mais populares (Welie, 2008).

É ainda de realçar que existem vantagens e desvantagens em atribuir muitas tags ao mesmo assunto/elemento multimédia. Ao atribuir muitas tags a um conteúdo este fica mais completo e abrange mais pontos relacionados com o recurso. No entanto, também se poderá tornar uma desvantagem, uma vez que cada tag irá ter tendência para diminuir de tamanho numa nuvem de tags.

Numa análise mais profunda, admite-se que a diminuição da relação entre a tag e o conteúdo classificado originará uma perda de usabilidade na navegação, o que leva a afirmar que as nuvens de tags não substituem a navegação pela informação hierarquizada. Contudo, quando estas se encontram bem estruturadas, facilitam e são consideradas um utensílio complementar à navegabilidade, proporcionando uma visão diferente dos conteúdos apresentados (Welie, 2008): *“A tag cloud gives a visual depiction of relative frequency rather than absolute frequency. This helps people to understand the most often used ones versus the lesser used one, which often is an indication of popularity or high activity. Alternatively, users could be presented with an ordered list and frequency numbers but that does not facilitate easy comparisons very visually. Besides, a tag cloud looks cool, doesn't it? Another important aspect of tag clouds is that it says something about what people do on the site, rather than what the site designers have prepared for visitors. Some claim that this may provide a different, and possibly effective, way to navigate to information. However, there does not seem to be strong evidence that this is really the case. Hence, tag clouds are no replacement for proper information hierarchy supported by an appropriate Main Navigation. They are probably mainly useful for providing a different view on the content and an alternative way of navigating content”* (Welie, 2008).

2.2.1 Utilizações distintas das nuvens de tags

A Internet permite a divulgação de informação a custo reduzido e de forma rápida. Com a actual facilidade de acesso à Internet e com o crescente incentivo à criação e publicação de conteúdos, estes tornam-se factores importantes para o enorme crescimento e propagação de informação.

Com a evolução da Internet houve uma conseqüente transformação dos resultados apresentados e das técnicas de visualização e apresentação de conteúdos originados pelas pesquisas dos utilizadores.

As nuvens de tags constituem uma funcionalidade da denominada Web 2.0. A expressão “nuvens de tags” foi utilizada pela primeira vez em Outubro de 2004 pela empresa O’Reilly Media, como foi referido anteriormente, e foi implementada num sistema de partilha de fotografias Flickr⁸ (Fevereiro de 2002). Mais tarde foram adoptadas também nos sítios Web Delicious⁹, no final de 2003, e Technorati¹⁰, em Julho de 2006, entre outros.



Figura 8 - Exemplo nuvem de tags

(Fonte: <http://pt.tag-clouds.com/>)

A figura acima apresenta uma nuvem de tags em que o algoritmo destaca as tags de um determinado assunto, isto é, as que são frequentemente utilizadas para promover tópicos relacionados com imobiliária, com fonte maior e cor mais destacada, enquanto as restantes tags aparecem noutras cores e fontes distintas.

⁸ www.flickr.com/

⁹ <http://delicious.com>

¹⁰ <http://technorati.com>

A Figura 10 mostra uma nuvem com termos relacionados com a Web 2.0.

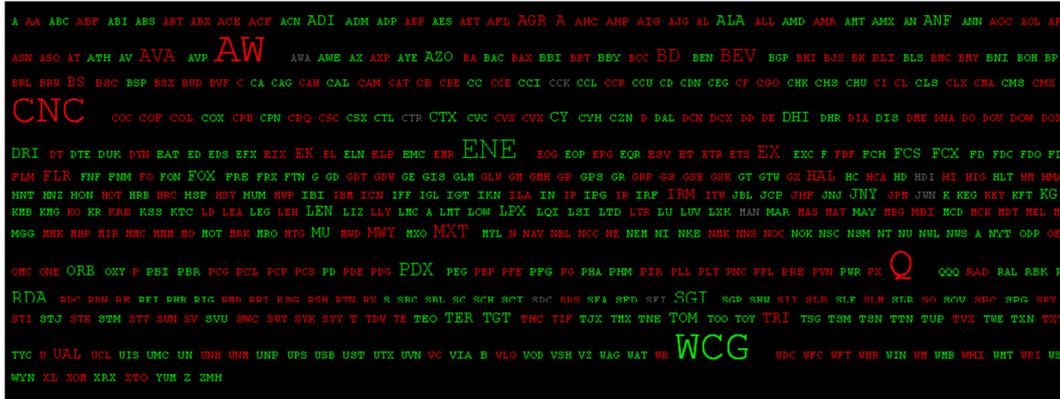


Figura 11 - Nuvem de tags (em lista) para visualizar movimento de ações

(Fonte: <http://en.wikipedia.org/wiki/>)

A imagem acima mostra o movimento do preço das ações. A cor indica a variação positiva ou negativa, o tamanho da fonte indica a mudança percentual.



Figura 12 - Nuvem de tags utilizada no Jornal de Noticias

(Fonte: <http://jn.sapo.pt/paginainicial/>)

A nuvem representada na Figura 12 é um exemplo, em que não recorre a nenhum estilo para evidenciar qual o tema mais em destaque.



Figura 14 - Nuvem de tags do discurso de Bill Gates nas CES 2007

(Fonte:<http://blog.seattlepi.com/microsoft/library/gatescloudtwo-a.jpg>)

Carolina Leslie, a autora do blogue Petitpois e arquitecta de informação, integrou a equipa do Instituto de Arquitectura de Informação dos EUA¹². Esta instituição é uma organização mundial que apoia indivíduos e organizações especializadas na concepção e construção de ambientes de informação compartilhadas. Leslie pertenceu ao quadro de conselheiros entre 2006 e 2008 estando ligada uma equipa de traduções de diversos artigos sobre Arquitecturas de Informação.

Carolina Leslie escreveu para a revista Techbits o seu comentário quando analisou o estudo efectuado sobre os discursos: “a *nuvem de palavras do Steve Jobs (acima)* é mais “*simples, concisa e eficiente*” do que a *do Bill Gates (abaixo)*, descrita como “*sisuda e difícil de entender*” (Techbits, 2007).

A conclusão obtida por Carolina Leslie veio ainda evidenciar a diferença entre os discursos, quer a nível de vocabulário, quer a nível de estruturação. O modo como as palavras foram empregues e a particular escolha das mesmas feita por cada um dos oradores influenciou a opinião do público sobre a escolha dos produtos vendidos pela empresa de cada um dos oradores.

¹² <http://iainstitute.org/>

Alexandre Fugita aficionado por tecnologias de informação refere no seu sítio Web que “*com base nessas nuvens de tags é possível visualizar quais são os assuntos mais importantes e descobrir que uma nuvem vale mais do que mil palavras*” (Techbits, 2007).

A revista internacional New York Times também escreveu um artigo semelhante ao relatado anteriormente¹³. Foram criadas 2 nuvens de tags com as diferenças de discursos entre republicanos e democratas nos Estados Unidos, conforme indicado na figura seguinte:

¹³ Disponível em
http://www.nytimes.com/interactive/2008/09/04/us/politics/20080905_WORDS_GRAPHIC.html

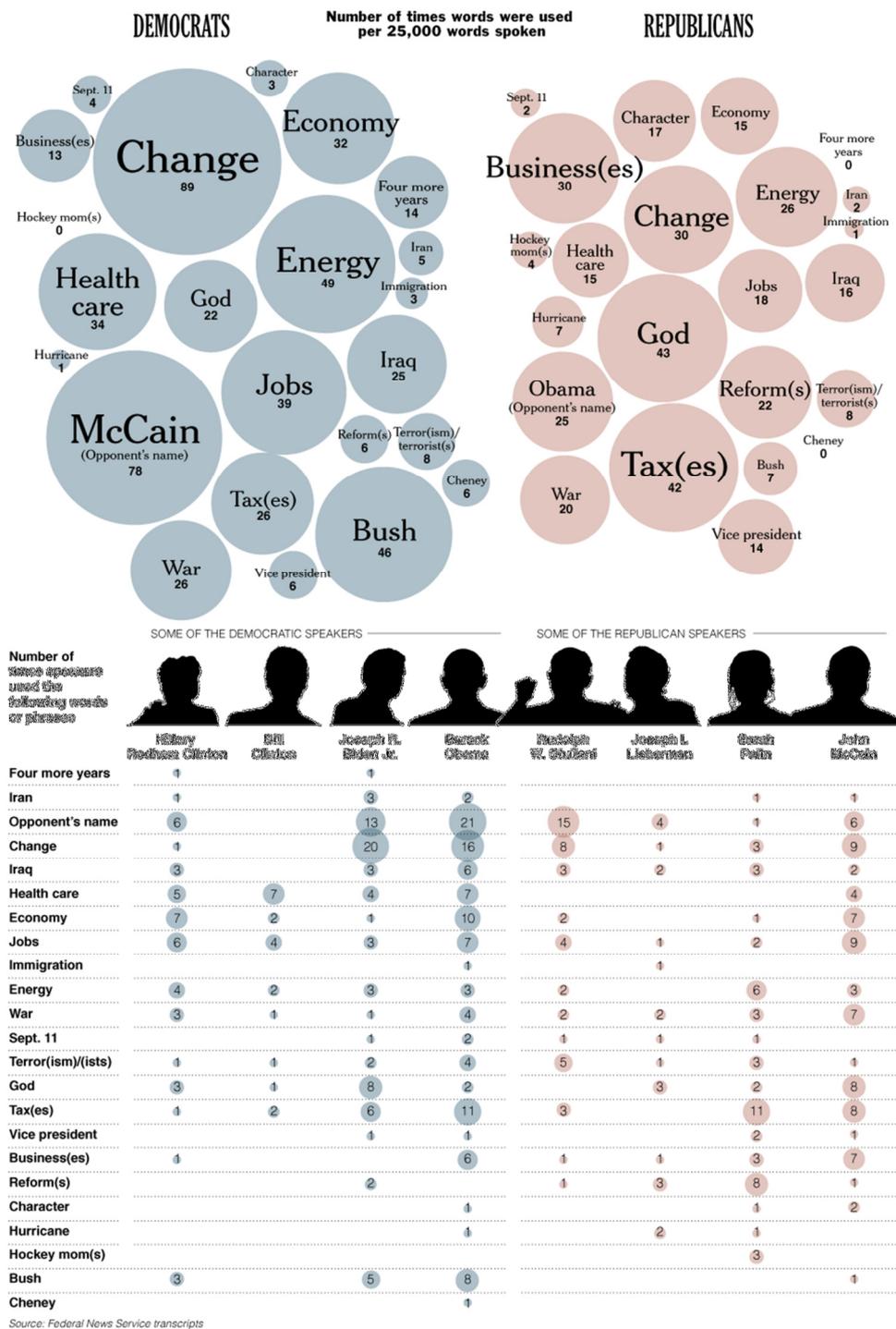


Figura 15 - Número de ocorrências das palavras que foram utilizadas por cada 25 mil palavras ditas (Fonte: http://www.nytimes.com/interactive/2008/09/04/us/politics/20080905_WORDS_GRAPHIC.html)

As palavras que os oradores utilizaram nas duas convenções políticas evidenciam os temas que ambas as partes destacaram. Os oradores republicanos falaram sobre reforma e carácter com mais frequência dos que os democratas, no entanto os republicanos também discursaram mais sobre negócios e impostos, ao passo que os democratas utilizaram um discurso mais orientado ao emprego e economia (Times, 2008).

2.2.2 Enfâse emocional nas nuvens de tags

Com intuito de criar as nuvens de tags mais interessantes e cativantes, um estudo desenvolvido por Takeharu Eda e colegas, consistiu em variar o tamanho da fonte de uma tag pelo seu conteúdo emocional e não pelo seu peso. Para tal realizam uma experiência que mostrou que este tipo de visualização suscita maior interesse no utilizador na sua primeira visualização da nuvem de tags. Tal acontece porque as tags constituintes da nuvem enfatizam o lado emocional de quem as lê. (Takeharu, et al., 2009).

Os autores partiram do pressuposto, encontrado em estudos anteriores realizados pelos próprios, que os utilizadores que já estavam familiarizados com nuvens de tags tinham maior facilidade na leitura de nuvens de tags ordenadas alfabeticamente.

O número de participantes que entrou neste estudo foi 806661, o número de tags foi 561952 e ainda foram utilizados 54336 URLs associados às tags.

Os autores, após analisarem as descrições que estavam associadas aos URLs, obtiveram duas conclusões importantes:

- A popularidade da tag é um bom critério para seleccionar palavras/tags das descrições observadas.
- Algumas das tags seleccionadas incluem-se no grupo de palavras que identificam qualidades ou características, tais como: “*cool, fun, weird*”.

Eda e colegas denominam este tipo de tags de emocionais, e revelaram que estas tendem a despertar maior atenção por parte dos utilizadores do que tags normais, ou seja, sem conotação emocional (Takeharu, et al., 2009) (Golder, et al., 2006). Diante destes fundamentos, o objectivo do estudo foi criar nuvens de tags emocionais.



Figura 16 - Nuvem de tags normal (à direita) e nuvem de tag emocional

(Fonte: (Takeharu, et al., 2009))

O estudo teve diversas etapas. Numa primeira fase foram escolhidos os 9 URLs mais populares. Para cada URL, foi mostrado ao utilizador, o título do sítio Web referente ao URL, uma amostra do sítio Web e duas nuvens de tags: uma normal, ou seja, o tamanho da fonte da palavra foi determinada pela popularidade e uma “emocional”. As duas nuvens de tags foram mostradas lado a lado, conforme se pode visualizar na imagem anterior, e a ordem foi aleatória, ora apresentada a nuvem normal à esquerda, ora à direita. De salientar que os participantes não sabiam o que as distinguiam nem qual era o objectivo do estudo.

Aos participantes foi-lhes solicitado que respondessem a duas questões em duas fases distintas:

- Primeira fase:
 - Identificar qual a nuvem mais atractiva.
 - Identificar qual a nuvem que fazia melhor descrição do conteúdo do URL apenas analisando a amostra que lhes foi fornecida.
- Segunda fase:
 - Identificar qual a nuvem mais atractiva.
 - Identificar qual a nuvem que fazia melhor descrição do conteúdo do URL depois de o conhecerem.

A última fase consistiu em classificar o conteúdo do URL quanto ao nível de interesse com uma das 3 escolhas possíveis:

- Interesse alto
- Interesse neutro
- Sem interesse

A tabela seguinte mostra os resultados obtidos pelas classificações que os participantes atribuíram ao conteúdo dos sítios Web em geral, sem distinção da classificação:

Tabela 2 - Percentagem de participantes que seleccionaram nuvens "emocionais".

	Primeira fase	Segunda fase
Avaliação da “atração”	69,4%	59,7%
Avaliação da descrição	50%	47,2%

(Fonte: (Takeharu, et al., 2009))

A tabela 3 exhibe os resultados obtidos apenas dos que seleccionaram “interesse alto” na primeira e segunda fase.

Tabela 3 - Percentagem de participantes que seleccionaram nuvens "emocionais" com classificação "Interesse alto" do conteúdo do sítio Web.

	Primeira fase	Segunda fase
Avaliação da "atração"	76,3%	60,5%
Avaliação da descrição	57,9%	47,4%

(Fonte: (Takeharu, et al., 2009))

Analisando as Tabelas 2 e 3 conclui-se que as nuvens de tags emocionais perdem interesse depois do utilizador conhecer o conteúdo do sítio Web.

Na primeira fase, ou seja, à primeira vista, as nuvens de tags emocionais são as mais interessantes, e que captam maior atenção.

Em termos de nuvem mais atractiva, a nuvem de tags emocionais conquistam mais votos dos participantes, apesar de se notar um decréscimo nas votações na segunda fase, isto é, após conhecer o conteúdo do sítio Web, verifica-se uma perda de 10% (de 69,4% para 59,7%) na Tabela 2 e de 16% (de 76,3% para 60,5%) na Tabela 3, ou seja esta diferença percentual revela que alguns participantes mudaram a sua opinião e consideraram a nuvem de tags normal mais funcional.

Os autores revelam no seu estudo que as nuvens de tags normais são mais fáceis de descrever conteúdos, depois de estes serem vistos, ou seja, facilitam a percepção.

Outras conclusões que foram obtidas indicam que:

- Os utilizadores à primeira vista apenas vêem as tags maiores, o que permite afirmar que é importante saber quais as tags que devem ser colocadas com fonte maior,
- O interesse dos utilizadores no conteúdo das tags é importante. Durante este estudo, a tag "cool" surgiu diversas vezes o que levou bastantes utilizadores a achar este facto aborrecido ou até a estarem perante o fenómeno de spam.

2.2.3 Outras formas de visualização

A visualização "standard" de nuvens de tags comum em vários sítios Web, tais como o Delicious e o Flickr, utilizam um *layout* representado por espaços rectangulares dispostos linha a linha com diferentes tamanhos da fonte para indicar a relevância das tags. As tags podem ser ordenadas pelo seu grau de importância ou alfabeticamente, por vezes recorrendo a utilização de cores.

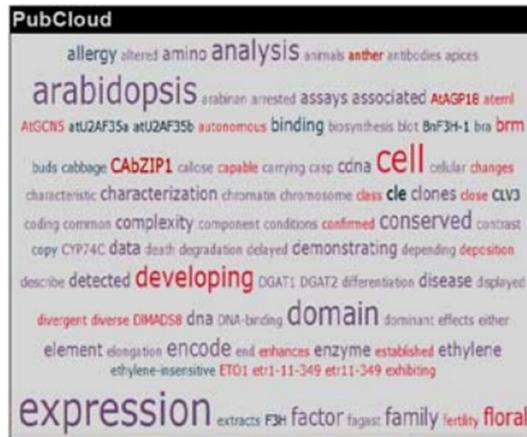


Figura 17 - Visão geral dos layouts de nuvens de tags existentes: Layout pubcloud
(Fonte: (Seifert, et al., 2008))



Figura 18 - Visão geral dos layouts de nuvens de tags existentes: tabelas HTML
(Fonte: (Seifert, et al., 2008))

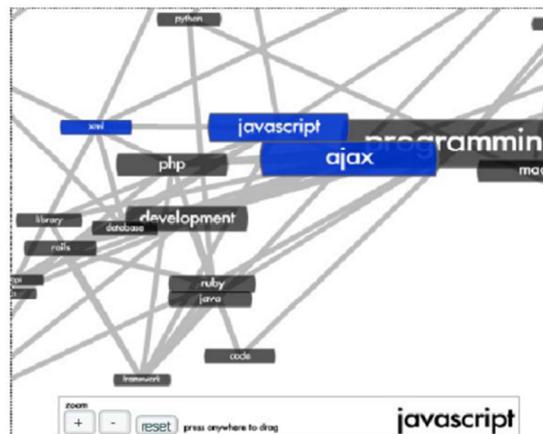


Figura 19 - Visão geral dos layouts de nuvens de tags existentes: estrutura gráfica
(Fonte: (Seifert, et al., 2008))

A primeira nuvem que rompeu com o estilo rectangular e assumiu outra forma de visualização, utilizando a relação entre tags, foi criada por Shaw (Shaw, 2005) (ver Figura 19). A nuvem de Shaw é um mapa gráfico 2D onde as tags são nós e sempre que existe relação entre tags, os nós são unidos por uma aresta. Apesar de este *layout* ser inovador, continua a manter algumas limitações, como muitos espaços brancos e sobreposição de tags, entenda-se por espaços brancos neste estudo o facto de mesmo existindo informação, a distribuição das tags não ocupar todo o espaço disponível e gerar espaçamentos sem nada.

Apesar de conhecidos os problemas da nuvem proposta por Shaw, Stefaner não os conseguiu contornar quando apresentou uma nuvem de tags elástica (Stefaner, 2007). As tags são dispostas num círculo 2D e as tags mais relevantes situam-se nos extremos dos espaços expandidos (ver Figura 20).

Bielenberg e Zacher também propuseram um *layout* circular onde o tamanho da fonte e posição do centro para as órbitas mostra a importância das tags (Bielenberg, et al., 2005) (ver Figura 21). As limitações referidas nos *layouts* anteriores continuam a ser uma constante, espaços brancos e tags sobrepostas ainda não foram corrigidas neste tipo de disposição (Seifert, et al., 2008) .

A seguinte tabela descreve resumidamente os tipos de visualização descritos anteriormente:

Tabela 4 - Quadro resumo

Figura	Observações	Estilo/Layout
Figura 17	Evidencia as tags mais actuais e revela as mais antigas recorrendo à utilização de cores.	Estilo rectangular.
Figura 18	Evidencia as tags mais actuais e revela as mais antigas, origina muitos espaços brancos.	Estilo rectangular.
Figura 19	Mapa gráfico 2D que possui relação entre tags. Origina espaços brancos e tags sobrepostas.	Estilo rectangular.
Figura 20	Nuvem elástica em que as tags mais relevantes situam-se nos extremos.	Estilo circular.
Figura 21	A importância das tags verifica-se pelo tamanho da fonte e pela posição da tag (do centro para as órbitas)	Estilo circular.

2.3 Folksonomia

O conceito “folksonomia” foi criado por Thomas Vander Wal (Wal, 2007) em 2007, e consiste numa taxonomia criada por utilizadores de um sítio Web para categorizar as definições atribuídas a conteúdos, recorrendo a palavras-chave: *“Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The*

tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information” (Wal, 2007).

Roberto Pereira afirma que “os sistemas baseados em folksonomia (folksonomy-based systems) representam uma iniciativa para auxiliar no processo de organização e de atribuição de significado aos conteúdos disponíveis na Internet. Estes sistemas permitem que seus usuários categorizem (Mathes, 2004), por meio da atribuição de um conjunto de termos, ou rótulos (tags), livremente seleccionados, qualquer objecto acessível na WEB (URLs, fotos, vídeos), e são denominados como sistemas baseados em folksonomia por fazerem do uso das tags pelos usuários sua estrutura central de funcionamento” (Pereira, 2008).

O termo “Folksonomia” nasce da junção das palavras “folk” (povo, gente) e “taxonomia” que veio integrar o conceito Web 2.0 na Internet. Os sítios Web pioneiros neste novo conceito foram o Delicious (serviço de partilha de tags), seguido do Flickr (sistema de partilha de fotografias) e o Youtube (serviço de partilha de vídeos). Através deste modo de organização, os utilizadores criam o seu próprio sistema de organização e criam relações entre conteúdos diferentes.

Joshua Schachter numa entrevista intitulada “*The time-saving trend of "tagging" is luring legions of Web surfers -- and Yahoo!*”, escrita por G. Heather e Robert D. Hof declarou que *tagging* é uma acção que origina uma melhor classificação da informação disponível na internet. Permite que várias pessoas participem numa acção conjunta de organização de informação, facilitando posteriormente a contextualização e compreensão da informação obtida: “*Tagging, however, lacks the algorithmic wizardry of search engines. But it lets people work together organically to create the context traditional search typically misses. It's like a grassroots Dewey Decimal Classification System for the Web. The essence: the combined work of people busily tagging content creates another way to make sense of the mountains of information online*” (Heather, et al., 2005).

Thomas Vander Wal explica no seu sítio Web¹⁴ a origem do termo “folksonomia” e menciona que a origem deste conceito surgiu quando Vander Wal questionou a plateia do IA Institute sobre a existência de algum conceito que permitisse classificar a informação social: “*On July 23, 2004 in the IA Institute (then called the Asylomar Institute for Information Architecture (AIFIA)) closed list serve Gene Smith asked, "Some of you might have noticed services like Furl, Flickr and Del.icio.us using user-defined labels or tags to organize and share information. Is there a name for this kind of informal social classification?" After a few other*

¹⁴ <http://www.vanderwal.net/>

people answered some other related questions Eric Scheid of Ironclad Information Architecture responded with "folk classification". On July 24, 2004 I responded just after that with, "So the user-created bottom-up categorical structure development with an emergent thesaurus would become a Folksonomy?" (Wal, 2007).

Uma folksonomia permite a classificação de conteúdos de diversas maneiras e com os termos adoptados pelos utilizadores. Assim, permite uma maior participação e uma crescente quantidade de informação num pequeno espaço de tempo gerada por todos os utilizadores/contribuintes com interesses comuns.

Graças a este modo de classificação associado (folksonomia) e utilizando este tipo de interfaces (nuvens de tags), os utilizadores pode ter uma abordagem diferente durante a navegação na Internet. Desta forma a pesquisa pode tornar-se mais rápida, personalizada, contextualizada e permite ainda identificar quais os grupos de utilizadores interessados em determinados assuntos.

O conceito de folksonomia, apesar de inovador e representar uma mais-valia para os utilizadores, apresenta vantagens e desvantagens, conforme se menciona na secção seguinte.

2.3.1 Vantagens e desvantagens

As principais vantagens que podem ser conseguidas com a utilização de folksonomias são as seguintes:

- Traduz-se numa economia em mão de obra especializada, isto é, não são necessários profissionais especializados para categorizar os conteúdos. Assim, os custos associados são diminutos, basta apenas o conteúdo multimédia estar disponível *online* para que qualquer utilizador possa efectuar *tagging*.
- Permite que o utilizador use o seu próprio vocabulário e não esteja sujeito a linguagens técnicas conforme os profissionais da área utilizam, tornando muito mais fácil e acessível a atribuição de tags. Assim, consegue-se também uma diminuição das dificuldades de cooperação.
- Relacionado com o ponto anterior, está a possibilidade de se atenuar o problema do "vocabulário" em sistemas de recuperação de informação, o utilizador tem a possibilidade de utilizar vocabulário simples e não técnico, o que possibilita uma recuperação mais rápida dos recursos marcadas nas visitas posteriores ao conteúdo marcado.
- Permite encontrar conteúdos diferentes da pesquisa inicial, graças à correlação entre tags, quando essa funcionalidade é utilizada.

No entanto, existem alguns pontos fracos neste modo de organização.

O obstáculo inicial começa por acarretar custos para a organização nomeadamente na aquisição de programadores que implementem um sistema capaz de interpretar este conceito de classificação de conteúdos. Esta atribuição livre de palavras-chave também origina outras barreiras. Por exemplo, a liberdade de atribuição de tags a conteúdos leva a inconsistências e ambiguidades, tais como: polissemia (palavras com muitos significados, ver Figura 22), sinonímia (múltiplas palavras com o mesmo significado, ver Figura 24), inflexões (variações do género, número e grau da palavra ver Figura 23), erros de ortografia. Outra desvantagem é a possibilidade de introdução de tags imprecisas ou irrelevantes (Mathes, 2004), que podem prejudicar a interpretação dos utilizadores que não estejam contextualizados com o conteúdo do recurso marcado, para o utilizador que a classificou esta questão já não é levantada.

Metadados podem ser obtidos pela introdução de informações adicionais acerca dos recursos pelos utilizadores. Por exemplo, uma descrição do recurso ou a sua categoria (Bielenberg, et al., 2005). Através da introdução de tags os utilizadores contribuem para uma caracterização dos recursos.

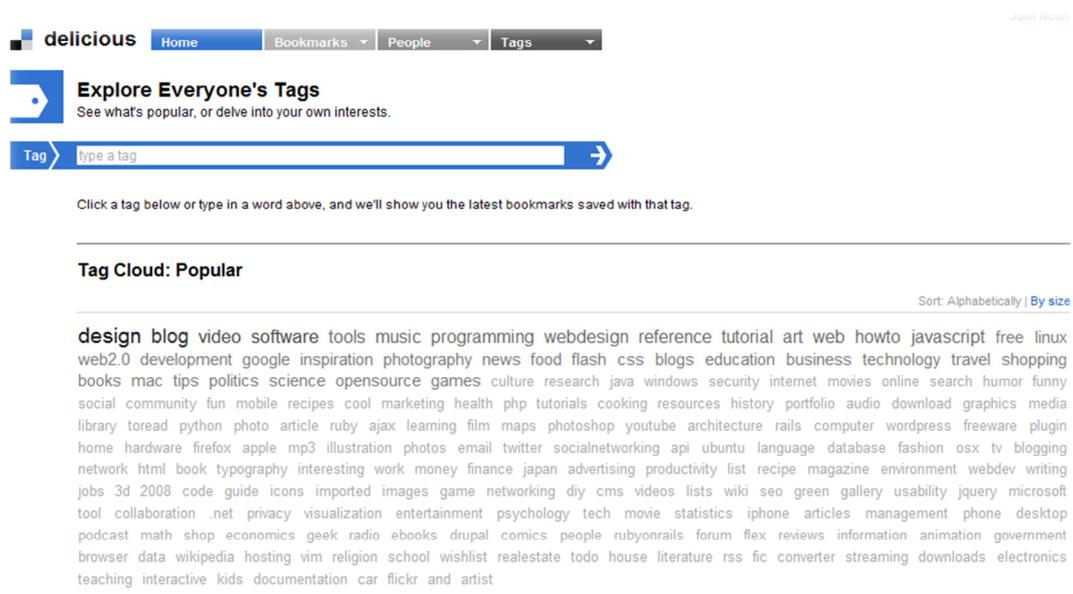


Figura 22 - Sítio Web Delicious Web bookmarking

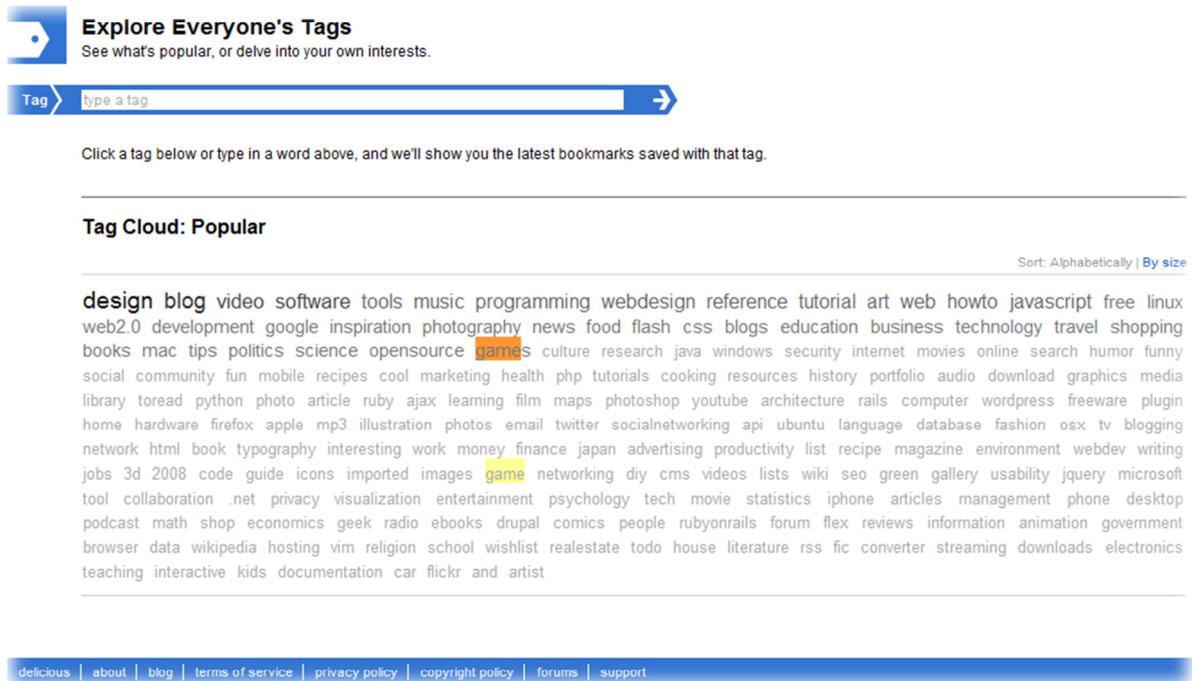


Figura 23 - Folksonomia com inflexões

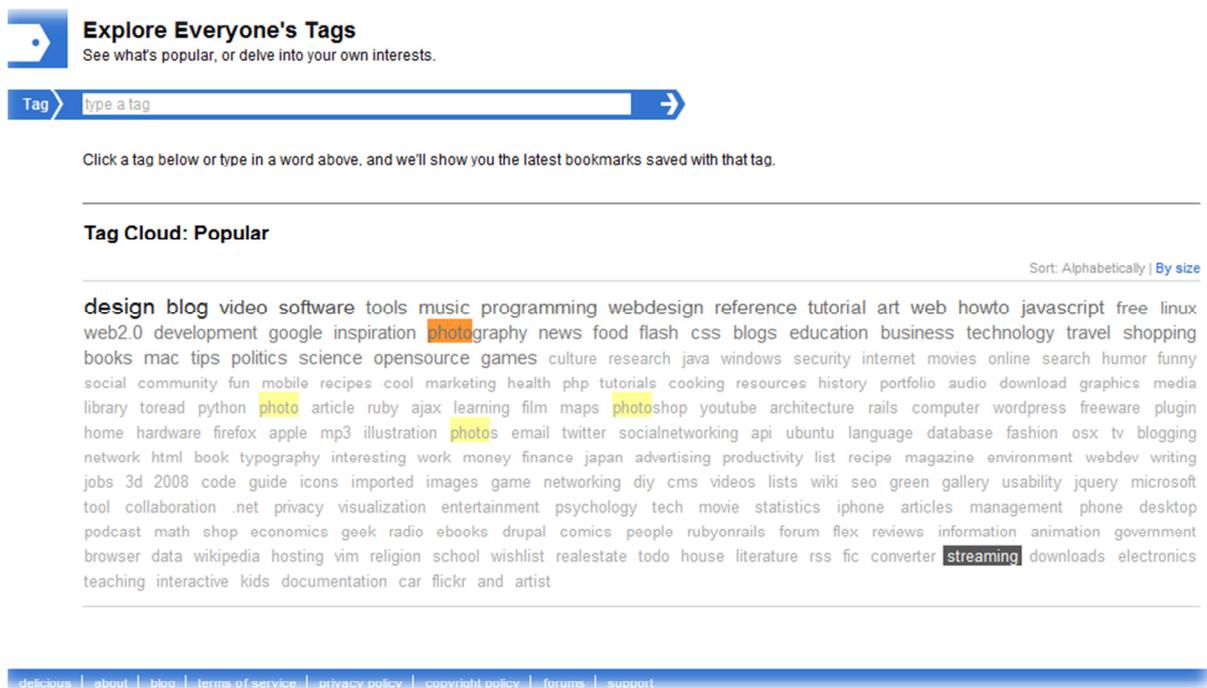


Figura 24 - Folksonomia com sinonímia

Guilherme Reis afirma que o individualismo pode ser uma desvantagem na utilização de folksonomias (Reis, 2007). Já Shilad e, Shyong Lam afirmam o contrário, categorizam este

tipo de tags como tags pessoais, um tema que será abordado ainda nesta secção. Por exemplo, uma tag muito detectada no Delicious é “toread”, no Flickr é “me”. As tags mencionadas, apesar de individualistas, são bastante perceptíveis para os utilizadores que as atribuíram e ao seu grupo de pessoas mais próximas (família, amigos, colegas), pois tiveram como propósito identificar um artigo para mais tarde ler, ou uma foto em que o utilizador aparece.

A utilização de tags mais individualistas, conforme mencionado no parágrafo anterior, dificulta o carácter colaborativo do sítio Web. Assim, e para minimizar estas situações, surgiu o conceito de “folksonomia controlada” (Reis, 2007) (Riu, 2008).

Apesar das desvantagens, Emanuele Quintarelli defende que são preferíveis as tags originadas pelo vocabulário não controlado e livre do que depender recursos, tempo e formação aos utilizadores e fazer controlo do vocabulário utilizado (Shirky, 2005): *“folksonomies are better than nothing, when traditional classification is not viable”* (Quintarelli, 2005).

A folksonomia controlada consiste no controlo da atribuição de tags dos utilizadores, isto é, a escolha das tags a utilizar fica limitada. Um sítio Web colaborativo tem várias opções de implementar este conceito. Pode, por exemplo, disponibilizar instruções sobre como definir/escolher as tags, apresentar uma listagem de tags já utilizadas por outros utilizadores, apresentar sinónimos, aplicar correctores ortográficos, disponibilizar tags pré-definidas.

Um exemplo disponível e com utilização da folksonomia controlada encontra-se no sítio Web <http://www.buzzillions.com/> e pode ser visualizada nas Figuras 25 e 26. Note-se que na Figura 26 é destacado o sistema de controlo de tags, sendo apresentadas sugestões e tags pré-definidas.

Review Snapshot™ Overview Description Specs Q&A (16)

HP 2009m 20 LCD Monitor



★★★★★ 4.6
based on 305 reviews

2,99 € - 179,00 €
from 13 stores

[Read Reviews](#) [Write a Review](#) [Add to List](#)



More Images



PROS	CONS	BEST USES
Vivid colors (248)	Poor adjustability (18)	Home computer (213)
Sharp image (242)	Little detail (6)	Multimedia (92)
Easy setup (226)	Inaccurate colors (3)	Gaming (54)
Lightweight (200)		Hooked to a
Adjustable (176)		

[See more](#)

Share this page:

Figura 25 - Tags controladas (Buzzillions.com)

Pros: Click all that apply:

- Adjustable
- Easy Setup
- Lightweight
- Sharp Image
- Vivid Colors
- Fast*

Or Add your *own* : (one at a time)

Suggestions: (click link to add to list on left)

[adjustable](#)
[affordable](#)
[accurate colors](#)

Cons: Click all that apply:

- Dead Pixels
- Inaccurate Colors
- Little Detail
- Poor Adjustability

Or Add your *own* : (one at a time)

Start entering your own tags and we will display suggestions if others have added words in similar letter combinations.

Best Uses: Click all that apply:

- Gaming
- Home Computer
- Hooked To A Laptop
- Multimedia

Or Add your *own* : (one at a time)

Start entering your own tags and we will display suggestions if others have added words in similar letter combinations.

Figura 26 - Sugestões de tags / tags pré-definidas

(Fonte: Buzzillions.com)

Capítulo III. Construção de nuvens de tags

Existem alguns procedimentos para exibir uma nuvem de tags. Irão ser descritos de forma sequencial alguns dos algoritmos que, com a mesma origem de dados, resultam em diferentes formas de visualização das tags.

A selecção do número de palavras a considerar e a respectiva contagem de ocorrências são dois elementos muito importantes para a construção de uma nuvem de tags.

Chow e seu colega Cheung conduziram uma investigação que permitiu confirmar que a utilização de 150 tags é o mais acertado para nuvens de tags que pretendem disponibilizar um grande número de informação: *“We have defined an acceptance range of total tags displayed in tag clouds. A tag cloud displaying 150-200 tags is recommended as it gives users enough information about tags but does not distort the tag cloud visualization effect”* (Chow, et al., 2008).

Neste capítulo irão ser descritos os principais algoritmos utilizados na construção de nuvens de tags.

3.1 Algoritmos para o cálculo do peso de cada tag

Esta secção aborda os 2 principais algoritmos que determinam o cálculo do peso, ou seja, o destaque que a tag deve ter dentro do espaço de uma nuvem de tags.

Existe o risco de determinada tag ser várias vezes usada, e estar constantemente na nuvem. No entanto, existe também o risco de que outras tags nunca cheguem a entrar para a nuvem, devido a serem pouco utilizadas. Um método para contornar este problema seria programar a consulta à base de dados e informar o número de vezes que determinada tag já foi exibida na nuvem e apenas obter os registos seguintes. Assim, o algoritmo utilizaria um contador do número de vezes que uma tag já foi exibida. Sempre que uma tag fosse visualizada um número de vezes igual ao limite definido, seria enviada para o fim da lista. Desta forma, as tags com menos relevância poderiam ser também mostradas na nuvem.

3.1.1 Algoritmos Linear e Logarítmico

Estes 2 algoritmos são os mais referenciados. Ambos permitem determinar o cálculo do peso, ou seja, o destaque que a tag deve ter numa nuvem de tags, mas diferem em alguns aspectos:

- O algoritmo linear é o mais simples e actualmente o mais utilizado nas implementações de nuvens de tags. Consiste no cálculo do intervalo de pesos, ou seja, o número de ocorrências da tag mais frequente (valor máximo do peso) e o número de ocorrências da tag menos utilizada (valor mínimo do peso);
- O algoritmo logarítmico permite uma melhor distribuição das tags na nuvem. Com este algoritmo é possível suavizar a diferença entre a tag actual e a seguinte no momento do cálculo. Ao invés de definir um intervalo e dividir as tags por secções, como acontece no algoritmo linear, este efectua um cálculo independente para cada tag que surge, definindo uma subdivisão diferente.

Um estudo realizado por Kevin Hoffman indicou o algoritmo logarítmico como o que mais evidenciava todas as tags e permitia uma melhor harmonização na visualização. Já a utilização do algoritmo linear, resultava numa nuvem mais rígida e evidenciava uma diferença grande entre as tags (Hoffman, 2006).

Acerca destas duas opções de cálculos para conceber nuvens de tags, conclui-se que, quando é utilizado o algoritmo logarítmico, os cálculos são baseados na frequência da ocorrência e não na sua frequência simples. Já no algoritmo linear existe uma relação directa entre o número de ocorrências e o tamanho da fonte, o que faz com que muitas tags sejam do mesmo tamanho e não haja disparidade entre elas, o que é facilmente revelado no algoritmo logarítmico. Assim, neste algoritmo (logarítmico) existe um maior desfasamento entre o tamanho das fontes das tags.

Além destes algoritmos, o *clustering* também poderá ser um método para gerar nuvens de tags. O *clustering* é uma técnica através da qual os dados são agrupados com base na semelhança entre os dados. O sistema da Amazon¹⁵ utiliza este algoritmo para prever eventuais compras dos seus utilizadores. O mecanismo de recomendação deste sistema regista a navegação e o histórico de compras do utilizador e deste modo agrupa os dados, o que permite efectuar previsões ou apresentar sugestões ao utilizador e expor informação ou conteúdos que este poderia estar interessado.

A título informativo, quando aplicado o algoritmo de *clustering*, este oferece uma distribuição mais coerente que a tradicional ordenação alfabética, permitindo a diferenciação entre os

¹⁵ <http://www.amazon.com>

principais temas, assim como melhor método para inferir conhecimento semântico sobre as tags vizinhas (Hassan-Montero, et al., 2006).

Este algoritmo, por requerer um agrupamento preliminar das tags, foi considerado estar fora do âmbito do estudo apresentado nesta dissertação.

A tabela seguinte expõe resumidamente as diferenças entre os dois algoritmos abordados em detalhe:

Tabela 5 - Diferenças entre algoritmos linear e logarítmico

Critério de comparação	Algoritmo Linear	Algoritmo Logarítmico
Complexidade implementação	na Implementação simples	Implementação mais complexa
Visualização das tags	Evidencia a diferença entre as tags	Suaviza a diferença entre tags
Tipo de apresentação	Define intervalo e divide as tags por secções	Cálculo independente por cada tag criando diferentes subdivisões
Tipo de distribuição	Distribuição rígida	Melhor distribuição visual
Tipo de cálculo	Cálculo baseado na frequência simples	Cálculo baseado na frequência da ocorrência
Escolha da fonte	Relação directa entre número de ocorrências e tamanho da fonte	Maior desfasamento entre o tamanho das fontes das tags

3.1.2 Algoritmo de selecção

Nesta secção irão ser descritos os passos utilizados para criar o algoritmo de selecção de tags utilizado no estudo experimental que é apresentado e discutido no capítulo 4.

1º) Selecção das palavras com mais ocorrências.

Segue o extracto de código SQL utilizado para a selecção de palavras que irão constituir a nuvem de tags.

```
SELECT top 60 [Tag],count([Tag])as peso FROM
[Delicious].[dbo].[Dados]
where tag is not null group by tag order by peso desc
```

Extracto de código 1. Selecção das 60 palavras a serem utilizadas

	Tag	peso
1	design	20523
2	blog	20473
3	web	19981
4	software	19803
5	music	18643
6	programming	16930
7	css	13094
8	reference	12866
9	news	12567
10	linux	12449
11	art	11594
12	tools	10953
13	java	9969
14	blogs	9515
15	tech	9258
16	mac	8838
17	imported	8336
18	politics	7773
19	photography	7737
20	games	7518
21	security	7444

Figura 27 - Top 21 da selecção de palavras do passo 1

Na Figura 27 mostra o resultado da execução da *query* do passo 1. A Figura 28 mostra graficamente os resultados apurados, a relação é feita pelo número de ocorrências de cada tag e a sua ordenação decendente.

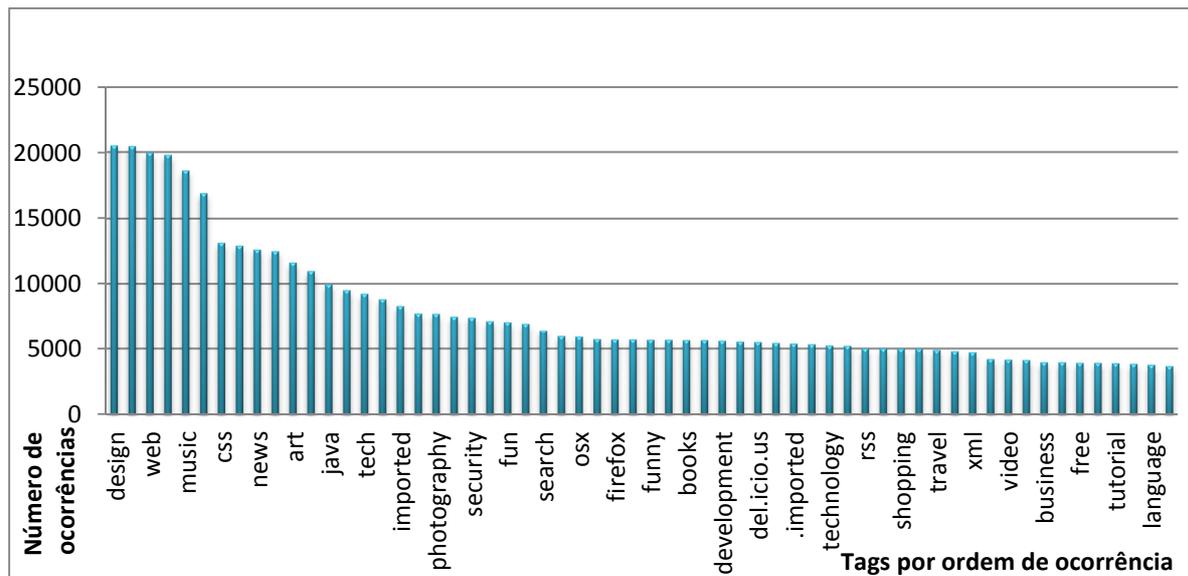


Figura 28 - Frequência das tags por ordem de ocorrência

2º) Selecção das palavras com mais peso ordenadas alfabeticamente.

```

select Tag, peso from
(
  SELECT top 60 [Tag], count([Tag]) as peso FROM
  [delicious].[dbo].[Dados]
  where tag is not null group by tag order by peso desc )
as tabela order by tag

```

Extracto de código 2. Selecção das 60 palavras determinadas pelo número de ocorrências e por ordem alfabética

A figura seguinte exhibe o resultado da *query* executada no passo 2. Na Figura 30 é possível visualizar-se a representação gráfica da mesma execução.

	Tag	peso
1	imported	5388
2	apple	3848
3	art	11594
4	blog	20473
5	blogs	9515
6	books	5666
7	business	3968
8	computer	3925
9	cool	5023
10	css	13094
11	del.icio.us	5522
12	design	20523
13	development	5601
14	firefox	5712
15	flash	5550
16	free	3930
17	fun	7094
18	funny	5692
19	games	7518
20	google	5981
21	history	3671

Figura 29 - Top 21 da selecção de palavras determinadas pelo número de ocorrências e por ordem alfabética

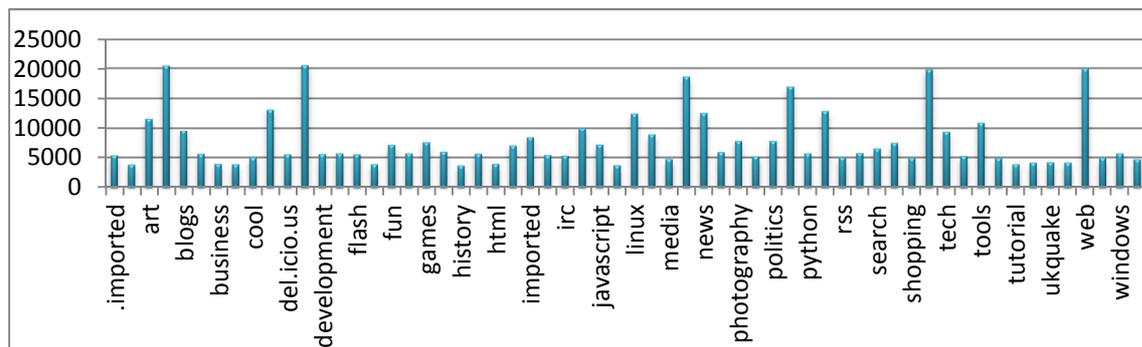


Figura 30 - Gráfico com número de ocorrências das palavras seleccionadas

3º) Codificação dos algoritmos

A codificação dos algoritmos aplicados neste estudo resultou de investigações realizadas por 2 autores distintos: Paul Hayman e o grupo criador do blogue <http://blogs.dekoh.com/dev>. Ambos relatam alguns projectos desenvolvidos para decidir qual a melhor apresentação para uma nuvem de tags.

Um dos autores, Paul Hayman, é o director de operações da empresa Ltd kwiboo. Conta com mais de uma década de experiência em consultoria TI (Tecnologias de Informação). É solicitado várias vezes para vários projectos dos seguintes sectores: *Utilities*,

Telecomunicações, Seguros, *Media*, *Investment Banking*, Lazer, Direito, CRM, Farmacêutica, jogos interactivos, Comunicações Móveis, Serviços Online. Paul Hayman é ainda o COO e co-fundador da kwiboo¹⁶, é também o criador do GeekZilla¹⁷.

Os métodos de criação e apresentação de nuvens de tags com linguagem C# e *aspx* propostos por Paul Hayman foram utilizados como base para a criação da apresentação da nuvem do estudo. O aspecto e os estilos utilizados foram adaptados para a criação da primeira nuvem de teste, seguindo-se depois a criação de novos estilos CSS (Cascading Style Sheets) para criação de outras nuvens de tags na realização do estudo.

A codificação dos algoritmos linear e logarítmico, propostos no blogue Dedok¹⁸, com mensagens de interesse técnico para programadores, também foram amplamente consultados durante a preparação do estudo realizado nesta dissertação.

```
double  occurrencesOfCurrentTag    =  double.Parse(row["peso"].ToString());
double  minOccurs = min;
double  maxOccurs = max;
double  weight = (Math.Log(occurrencesOfCurrentTag) - Math.Log(minOccurs)) /
(Math.Log(maxOccurs) - Math.Log(minOccurs));
double  minFontSize = 9;
double  maxFontSize = 32;
double  fontSizeOfCurrentTag = minFontSize + Math.Round((maxFontSize -
minFontSize) * weight);
```

Extracto de código 3. Implementação do algoritmo logarítmico (Venkat, 2007)

```
double  occurrencesOfCurrentTag = double.Parse(row["peso"].ToString());
double  minOccurs = min;
double  maxOccurs = max;
double  weight = (occurrencesOfCurrentTag-minOccurs)/(maxOccurs-minOccurs);
double  minFontSize = 9;
double  maxFontSize = 32;
double  fontSizeOfCurrentTag = minFontSize + Math.Round((maxFontSize -
minFontSize) * weight);
```

Extracto de código 4. Implementação do algoritmo linear (Venkat, 2007)

O cálculo do peso e conseqüente atribuição do tamanho da fonte à tag no algoritmo linear podem ser problemáticos. O facto do resultado da fórmula originar muitas vezes valores inferiores a 0, cria uma limitação. Esta restrição, que será mostrada no excerto de código

¹⁶ <http://www.kwiboo.com/>

¹⁷ <http://geekzilla.co.uk>

¹⁸ <http://blogs.dekoh.com/dev/2007/10/29/choosing-a-good-font-size-variation-algorithm-for-your-tag-cloud/>

seguinte, faz com que por vezes a tag seja perdida, isto é, excluída da nuvem, pois não tem “peso” suficiente para ser mantida.

No excerto de código abaixo, a zona sombreada mostra a condicionante de apenas apresentar tags que tenham a variável *weight* superior a 0. Deste modo, com este algoritmo são perdidas várias tags. Considere-se o seguinte exemplo:

```
Tag: Business
occurrencesOfCurrentTag: 3968
weight = 0.017624020887728461 (segundo a fórmula apresentada acima no
Extracto de código que implementa o algoritmo linear)
fontSizeOfCurrentTag = 9.0
weight * 100 = 1.76
```

```
if (weight * 100 >= 99)
{
    //heaviest
    weight = 1;
}
else if (weight * 100 >= 70)
{
    weight = 2;
}
else if (weight * 100 >= 40)
{
    weight = 3;
}
else if (weight * 100 >= 20)
{
    weight = 4;
}
else if (weight * 100 >= 3)
{
    //weakest
    weight = 5;
}
else
{
    // use this to filter out all low hitters
    weight = 0;
}

string aux = "<span class='fontAux'>(" + row["peso"].ToString() +
")</span>";
if (weight > 0)
    outputBuffer.Append(cloudTemplate.Replace("$weight$",
        weight.ToString()).Replace("$tag$",
        row["tag"].ToString() + aux).Replace("$urlencodetag$",
        HttpUtility.UrlEncode(row["tag"].ToString())).Replace("$title$",
        row["peso"].ToString()).Replace("$font$",
        fontSizeOfCurrentTag.ToString()));
```

Excerto do código 3. Implementação utilizada para a escolha da fonte (Hayman, 2006)

O código fonte foi modificado para que não houvesse exclusões de tags, ou seja, para aceitar todos os pesos calculados pelo número de ocorrências.

As figuras 31 e 32 mostram a nuvem originada pelo algoritmo com a condicionante do peso superior a 0.

Visualização Logarítmica



Figura 31 - Visualização logarítmica com limitação de peso (uma cor)

Visualização Linear



Figura 32 - Visualização Linear com limitação de peso (uma cor)

Um exemplo, que se verifica é a tag “fun” tem nas 2 nuvens o mesmo peso (ver Figura 33 e Figura 34), mas aparece com tamanhos diferentes.

Visualização Logarítmica



Figura 33 - Visualização logarítmica com limitação de peso (colorida)

Visualização Linear



Figura 34 - Visualização linear com limitação de peso (colorida)

A figura abaixo mostra a nuvens geradas por ambos os algoritmos, sem condicionantes (peso superior a 0) e monocromáticas.

Visualização Logarítmica



Visualização Linear



Figura 35 - Visualização logarítmica Vs. Visualização Linear (sem condicionante, monocromática)

A figura seguinte mostra a nuvens geradas por ambos os algoritmos, sem condicionantes e coloridas.

Visualização Logarítmica



Visualização Linear



Figura 36 - Visualização logarítmica Vs. Visualização Linear (sem condicionante, colorida)

A utilização do algoritmo logarítmico parece evidenciar mais todas as tags e criar uma melhor distribuição das tags na área de apresentação da nuvem. Já a nuvem gerada através do algoritmo linear parece mais rígida, como tinha sido anteriormente referido na secção 3.1.1 (Hoffman, 2006).

3.2 Posicionamento, fonte e ordenação das tags

Nesta secção discute-se a importância da ordenação das tags apresentadas na nuvem, do seu posicionamento e do tipo de fonte utilizada, nomeadamente entre os tipos “*Serif*” e “*Sans-Serif*”.

3.2.1 Posicionamento das tags

Não existindo um local específico que pode ser denominado o melhor sítio para colocar uma nuvem de tags dentro de uma página Web e sendo variável o local, o espaçamento e a apresentação final que irá ser exposta ao utilizador, existe assim “a melhor disposição de uma nuvem de tags”, isto referindo-se à disposição das tags dentro da nuvem. Esse factor deve ser considerado durante o processo de desenvolvimento desta. Os autores do artigo “On the beauty and usability of tag clouds” afirmam que o posicionamento das tags influencia o tempo gasto na pesquisa, e que as que se situam no topo da lista ou no canto superior esquerdo da nuvem são encontradas de forma mais célere: “*Tags at the beginning of the list or in the upper-left corner of the tag cloud were found much quicker*” (Seifert, et al., 2008).

Também o autor Lohman afirma que a melhor posição para as tags dentro de uma nuvem de tags é no quadrante superior esquerdo desta (Lohman, et al., 2009).

Os autores do artigo “*Seeing Things in the clouds: The effect of visual features on tag cloud selections*” também corroboram outros mencionados anteriormente: “*A position effect was found, where the upper-left hand quadrant was strongest*” (Bateman, et al., 2008) .

3.2.2 O tipo de fonte “Serif”

Muitas são as tentativas de inovação na disposição das nuvens tags. Uma das características que ainda não foi muito explorada e que pode eventualmente contribuir para uma melhor disposição das nuvens são as fontes das letras.

Segundo um conjunto de boas práticas proposto por Joe Lamanti, é aconselhável a utilização de fontes da família “*Sans-Serif*”, (Lamanti, 2007). Esta gama de fontes melhora a legibilidade global das tags apresentadas numa nuvem, segundo um estudo levado a cabo por Dr. Ralph F. Wilson que irá ser descrito ainda neste capítulo (secção 3.4.3). O autor concluiu que o ser humano lê palavras ao invés de letras individuais, assim as letras aparentam unir-se devido aos seus prolongamentos (ver Figura 37). Por outro lado, os tipos “*Sans-Serif*” costumam ser usados em títulos, pois valorizam cada palavra individualmente e tendem a ter maior peso e presença para os olhos, captando mais a atenção do utilizador, já que parecem mais limpos (Wilson, 2001).

Alex Poole, *designer* de sistemas interactivos e investigador na área de usabilidade Web, após uma indagação relativa à dicotomia “*Sans-Serif*” ou “*Serif*”, deduziu que fontes “*Sans-Serif*” são mais legíveis (Poole, 2005). Um estudo (Rubinstein, 1988) abordado por Poole identificou que quando um documento é digitalizado, as letras têm de caber num espaço de um *pixel*, o que promove uma má resolução da visualização. Pelo contrário, a utilização de

fontes “*Serif*”, quando digitalizadas, permitem obter um aspecto menos confuso, o que permite melhor legibilidade.

As fontes “Arial”, “Calibri”, “Verdana” são do tipo “*Sans-Serif*”. Já as fontes “Times New Roman” e “Liberation Serif”, por exemplo, pertencem ao grupo das fontes *Serif*.



Figura 37 - Exemplo de fontes *Sans-Serif* e *Serif*

(Fonte: http://www.alexpoole.info/academic/literaturereview.html#Tinker_1932)

Gunnar Swanson, Web *designer*, escritor, e professor de design gráfico, atualmente a leccionar na School of Art & Design em East Carolina University, e autor do livro “Graphic Design and Reading: explorations of an uneasy relationship”, refere que, apesar dos inúmeros estudos e testes realizados para perceber qual o tipo de letra, “*Sans-Serif*” ou “*Serif*”, que melhoram a percepção do utilizador na leitura, todos obtêm resultados inconclusivos: “*One extensive study suggests that “the neurological structure of the human visual system benefits from serifs in the preservation of the main features of letters”¹⁹. It seems that since words are perceived by their outline shape, single letters with serifs, which have a more distinctive outline shape over sans serif letters, may also be more easily perceptible than are sans serif letters. Another study found that serif typefaces are marginally more legible than sans serif designs. Two highly similar trial stories were presented in both serif and sans serif type*” (Swanson, 2000).

Outro aspecto importante na escolha das fontes que Swanson destaca é a aplicação de “negrito” e “itálico” na apresentação do texto (Swanson, 2000). As letras em negrito têm tendência a dificultar a leitura, por se tornarem pesadas aos olhos, por outro lado as que são extremamente claras tornam-se de difícil legibilidade.

¹⁹ D.O Robinson, M. Abbamonte, and S. H. Evans, “Why Serifs Are Important: The perception of small print” *Visible Language* 4 (1971) 353-359

As letras em itálico são geralmente utilizadas para dar ênfase a determinado texto. No entanto, Gunnar Swanson refere que vários utilizadores manifestaram desagrado na utilização deste estilo, defendendo que este estilo promove a diminuição da velocidade na leitura das palavras (Swanson, 2000).

3.2.3 Ordenação das tags na nuvem

As tags podem aparecer numa nuvem de tags ordenadas alfabeticamente ou por maior número de ocorrências, quando utilizado algoritmos lineares e logarítmicos.

Will McGugan sugere a utilização de cores nas tags relevantes e a não utilização da ordem alfabética, como é comum na maioria das nuvens de tags (McGugan, 2007). Aconselha a ordenação baseada na popularidade da tag, ou seja, no seu peso.

As figuras seguintes (ver Figura 38 e 39) exibem respectivamente o exemplo das visualizações com implementação do algoritmo logarítmico com tags coloridas e ordenadas alfabeticamente e a visualização com implementação do algoritmo linear com tags monocromáticas e ordenadas alfabeticamente.

Visualização Logaritmica

design₍₂₀₅₂₃₎ blog₍₂₀₄₇₃₎ web₍₁₉₉₈₁₎ software₍₁₉₈₀₃₎ music₍₁₈₆₄₃₎
programming₍₁₆₉₃₀₎ css₍₁₃₀₉₄₎ reference₍₁₂₈₆₆₎ news₍₁₂₅₆₇₎ linux
(12449) art₍₁₁₅₉₄₎ tools₍₁₀₉₅₃₎ java₍₉₉₆₉₎ blogs₍₉₅₁₅₎ tech₍₉₂₅₈₎ mac₍₈₈₃₈₎ imported
(8336) politics₍₇₇₇₃₎ photography₍₇₇₃₇₎ games₍₇₅₁₈₎ security₍₇₄₄₄₎ javascript₍₇₁₇₁₎ fun
(7094) humor₍₆₉₈₃₎ search₍₆₄₇₉₎ google₍₅₉₈₁₎ osx₍₅₉₃₀₎ science₍₅₇₄₅₎ firefox₍₅₇₁₂₎ windows₍₅₇₀₃₎
funny₍₅₆₉₂₎ python₍₅₆₈₇₎ books₍₅₆₆₆₎ howto₍₅₆₅₉₎ development₍₅₆₀₁₎ flash₍₅₅₅₀₎ del.icio.us₍₅₅₂₂₎
internet₍₅₄₃₂₎ .imported₍₅₃₈₈₎ irc₍₅₃₃₄₎ technology₍₅₂₅₄₎ php₍₅₂₁₀₎ rss₍₅₀₃₈₎ cool₍₅₀₂₃₎ shopping₍₅₀₁₀₎
webdesign₍₅₀₁₀₎ travel₍₄₉₁₆₎ media₍₄₇₉₇₎ xml₍₄₇₄₀₎ ukquake₍₄₂₂₅₎ video₍₄₁₇₀₎ ukq:irc₍₄₁₅₇₎ business₍₃₉₆₈₎ html
(3965) free₍₃₉₃₀₎ computer₍₃₉₂₅₎ tutorial₍₃₈₉₅₎ apple₍₃₈₄₈₎ language₍₃₇₆₉₎ history₍₃₆₇₁₎

Figura 38 – Visualização logarítmica colorida ordenada alfabeticamente

Visualização Linear



Figura 39 - Visualização Linear monocromática ordenada alfabeticamente

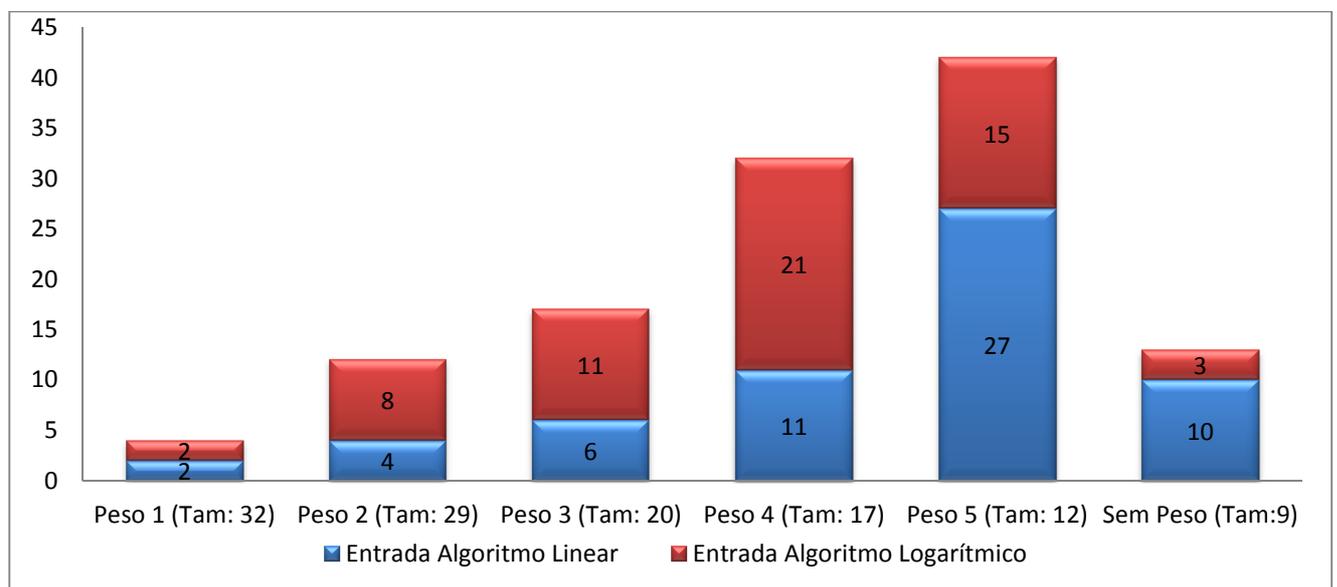


Figura 40 - Frequência da utilização dos diferentes tamanhos das fontes

A figura acima mostra a representação gráfica da frequência da utilização dos diferentes tamanhos das fontes nos 2 algoritmos mencionados anteriormente.

3.3 Análise de alguns estudos de caso

Nuvens de tags têm sido objecto de investigação em vários estudos de usabilidade. Existem alguns estudos realizados com objectivos semelhantes aos do trabalho descrito neste documento. Nesta secção são abordados três estudos pela relevância dos mesmos para o trabalho de mestrado desenvolvido.

A tabela 6 mostra a informação geral relativa ao caso 1 e caso 2, que serão seguidamente explicados.

Tabela 6 - Dados para realização das experiências dos estudos apresentados

	Local de estudo	Número de participantes	Função dominante	Idade Média (anos)	Observações
Caso 1	Dublin, Irlanda	62	Estudante	27 (Min: 23; Max: 44)	Não aplicável.
Caso 2	Duisberg, Alemanha	36	Estudante	26 (Min: 17; Max: 57)	O grau médio de familiaridade dos participantes era 4 numa escala de 1 a 10, e 5 indivíduos nunca tinham visto nuvens de tags.

3.3.1 Caso 1 - Leitura e apresentação de nuvens de tags

Halvey e Keane documentaram alguns resultados comparativos de diferentes apresentações de nuvens de tags (Halvey, et al., 2007). O estudo baseou-se em nuvens de tags sem atribuição de pesos, com apresentação das tags dispostas em listas horizontais e listas verticais que foram apresentadas aos participantes. Estes deveriam localizar e escolher determinadas tags nas nuvens com os formatos apresentados.

Os resultados obtidos indicaram que as listas de tags sem diferenças de pesos proporcionavam ao utilizador um melhor desempenho na pesquisa da tag do que a apresentação de uma nuvem de tags. Se a listagem das tags fosse ordenada alfabeticamente, ainda melhorava o processo de pesquisa das palavras, isto é, o tempo gasto a encontrar a palavra diminuía. Outra conclusão importante foi que, numa nuvem, a variação da fonte é muito determinante para se encontrar as palavras.

Também as tags que estavam localizadas no canto superior esquerdo da nuvem foram encontradas mais facilmente. Tendo em consideração os tempos gastos pelos participantes na pesquisa por determinadas palavras em regiões específicas da nuvem, os autores concluíram que as nuvens de tags são “digitalizadas” e não lidas pelos utilizadores, isto é, o utilizador faz um “*scanning*” da palavra, que não é lida na totalidade.

3.3.2 Caso 2 - Percepção de nuvens de tags

Com o intuito de identificar qual a disposição que melhor promove a apresentação de uma nuvem de tags, Lohman conduziu um estudo (Lohman, et al., 2009). Este baseou-se na apresentação de várias nuvens com as mesmas tags a um grupo de utilizadores com idade média de 26 anos (idade mínima de 17 e máxima de 57 anos). Os participantes, numa escala de 0-10, estavam quase todos no nível 4 de familiarização com nuvens de tags (0 representando o menor grau de familiaridade e 10 o maior).

Existem diversos factores que condicionam qual a melhor concepção, como os objectivos específicos do utilizador e as intenções do *designer*. No entanto, a experiência permitiu apurar os seguintes dados:

1. *Layout* que permite encontrar uma tag específica mais rapidamente: *layout* sequencial e com ordenação alfabética das tags,
2. *Layout* que permite encontrar as tags mais populares: *layout* circular, com a diminuição da importância,
3. *Layout* que permite encontrar as tags que pertencem a um determinado tema: *layout* com agrupamento temáticos.

Na tabela 7 são apresentados os resultados obtidos pelo caso 2, com uma comparação com outros trabalhos referenciados no artigo de Lohman (Lohman, et al., 2009).

Tabela 7 - Sumário dos resultados da pesquisa sobre percepção de nuvens de tags e desempenhos contrastantes com trabalhos relacionados.

Pontos observados	Resultado dos estudos e literatura analisados por Lohman para comparação	Avaliação da investigação de Lohman
Tamanho da tag	Tags maiores captam mais rapidamente a atenção do utilizador do que as mais pequenas. Também são identificadas mais rapidamente.	Os resultados obtidos suportam os resultados analisados na literatura abordada por Lohman, embora também indiquem que outras propriedades. Nomeadamente, alguns dos factores mencionados são o número de caracteres da tag, (palavras mais longas são melhor visualizadas?), a posição de uma tag na nuvem, ou tags vizinhas influenciam o efeito.
Scanning	Os utilizadores não lêem as tags, fazem um <i>scan visual</i> à nuvem.	Os dados de rastreio ocular apoiam esta conclusão. No entanto, numa passagem de olhos geral, as tags poderão não ser identificadas.
Centralização	As tags situadas no meio da nuvem captam mais atenção do que as situadas na periferia.	Segundo os resultados obtidos, este facto parece ser verdadeiro para a maioria dos tipos de nuvens de tags. Na maior parte das nuvens, as tags dispostas no centro foram encontradas mais rapidamente.
Posição	As tags posicionadas no canto superior esquerdo são mais vezes seleccionadas e encontradas com maior rapidez.	Os dados mostram que o rastreio ocular no canto superior esquerdo recebe mais fixações em todos os <i>layouts</i> .
Layout	O <i>layout</i> apresentado na nuvem de tags influencia a percepção do utilizador	Esta constatação é fortemente apoiada pelos dados de rastreio ocular obtidos.
Exploração	As nuvens de tags fornecem um bom método de pesquisa quando se pretende localizar uma determinada tag.	Em geral, esta conclusão é apoiada.

(Fonte: (Lohman, et al., 2009)).

Outro ponto interessante analisado pelos participantes foi a importância relativa da performance e da estética. O feedback atribuído pelos participantes revelou que preferiam que a nuvem tivesse um aspecto interessante e de visualização agradável, do que fosse funcional: *“The feedback and ratings of the study participants suggest that fun and aesthetic aspects largely affect the user's interaction with tag clouds. Participants partly preferred layouts that did not yield the best performance. Thus, such aspects should be included in any design decision and usability evaluation of tag clouds”* (Lohman, et al., 2009).

3.3.3 Caso 3 - Legibilidade das fontes aplicadas às tags

Com intenção de avaliar qual o impacto da fonte “*Serif*” e “*Sans-Serif*” na leitura de artigos disponíveis na Web, Dr. Ralph F. Wilson levou a cabo um estudo acerca deste tema (Wilson, 2001).

Ao longo do tempo, a ideia de que as fontes “*Serif*” facilitavam a leitura foi implementada e adoptada em quase todos os livros, jornais e revistas, a fonte utilizada costuma ser “Times New Roman” e “Bookman”.

No entanto, apesar deste pressuposto de que o tipo de fonte “*Serif*” permite maior legibilidade em qualquer meio, é necessário considerar que um monitor ou um ecrã é muito diferente do papel impresso, por exemplo. A resolução do papel (cerca de 72 pontos por polegada) é muito menor que a resolução de um ecrã de um computador (cerca de 180dpi ou 300dpi).

Foram efectuados 2 questionários com a finalidade de comparar 2 tipos de fontes, com os participantes a escolher o que lhes parecia mais legível no ecrã. O primeiro questionário confrontou a fonte “*Serif*” Times New Roman, com tamanho 12pt, a que existe por defeito em diversos *browsers* Web, com a fonte “*Sans-Serif*” Arial, com tamanho 12pt e o resultado obtido surpreendeu o autor ao mostrar que os 68% dos participantes preferiram a fonte “*Sans-Serif*”, o que pôs em causa a supremacia de legibilidade das fontes “*Serif*”.

No segundo questionário foram comparadas duas fontes “*Serif*”, com a opção do participante admitir que não conseguiu distinguir as fontes. As fontes apresentadas foram Times New Roman, tendo sido esta desenvolvida propositadamente pela Microsoft para leitura em ecrãs de computadores e a segunda fonte para comparação foi Georgia. Os resultados mostraram que a fonte Georgia não está disponível em tantos computadores como a Times New Roman, esta limitação da fonte deu uma importância significativa à opção que afirmava que o utilizador não conseguiu distinguir as fontes. Os resultados obtidos mostraram que 15% dos respondentes não conseguiram distinguir as fontes e 52% preferiu a fonte Georgia.

Com os resultados obtidos, Wilson realizou 2 questionários adicionais, comparando fontes “*Sans-Serif*”, e utilizou a fonte Verdana, também criada pela Microsoft, comparando-a com a fonte Arial. Desta vez foram utilizados os tamanhos 12pt, 10pt e 9pt. Os resultados obtidos mostraram que no tamanho 12pt a fonte Arial é preferida por 59% dos utilizadores, no tamanho 10pt e abaixo a preferência muda para a fonte Verdana.

Uma das conclusões obtidas baseando-se nas respostas dadas pelos utilizadores aos questionários, é que a fonte Verdana tem uma letra mais aberta e ocupa mais espaço do que a Arial, o que contribui para a legibilidade, conclui-se ainda que o ser humano lê palavras e não letras individualmente. Os tipos de fonte “*Sans-Serif*” costumam ser usados em títulos, pois valorizam cada palavra individualmente e tendem a ter maior presença para os olhos, captando mais a atenção do utilizador, já que parecem mais limpos (Wilson, 2001). Para diferentes tamanhos de fontes e consoante a finalidade seja papel ou um monitor a escolha do tipo de fonte é importante.

Capítulo IV. Estudo experimental

Neste capítulo irá ser abordado o estudo experimental que foi realizado para o desenvolvimento e desta dissertação. Será apresentada a contextualização da experiência, em que consistiu e os resultados obtidos.

4.1 Contextualização

Este estudo empírico centra-se na investigação das abordagens mais apropriadas para a construção de nuvens de tags. Em termos de metodologia, este estudo baseia-se na investigação qualitativa e pode ser caracterizado como um estudo experimental, utilizado para analisar os métodos mais intuitivos e eficazes de construção de nuvens de tags, considerando-as como ferramentas de auxílio na pesquisa de recursos de interesse para os utilizadores.

Foi realizada uma experiência com a colaboração de participantes que respondendo a um inquérito com nuvens de tags elaboradas a partir dos algoritmos descritos anteriormente, o qual permitiu aferir qual o método e disposição que melhor facilita a leitura das nuvens e menor carga cognitiva do utilizador.

Solicitou-se a participação de 98 pessoas, tendo-se obtido a colaboração de 28. Seguiu-se a recomendação de Kumar na escolha da audiência (Kumar, 2005), e pediu-se a participação de pessoas que estariam mais motivadas para a participação num estudo desta natureza, sem esquecer que os resultados interessavam para um repositório institucional.

Assim, os participantes no estudo eram maioritariamente alunos do Mestrado de Engenharia Informática do ISEP, com frequência da disciplina Concepção e Autoria Multimédia. No entanto, optou-se por alargar o questionário a outros utilizadores, maioritariamente com formação superior, de forma a obterem-se mais respostas. Houve ainda o cuidado de que os participantes fossem portugueses ou vivessem em Portugal, para que os resultados traduzissem a realidade neste país. O questionário foi disponibilizado no dia 1 de Maio e respondido entre dia 6 de Maio e dia 6 de Agosto de 2010.

Utilizou-se um conjunto de tags atribuídas por diversos utilizadores nos meses de Junho de 2003 e Maio de 2007 os quais ficaram registados num *dataset* do sistema Delicious, estes correspondem a mais de 1 milhão de registos. Esta informação foi utilizada para realizar a experiência, mas dada a enorme quantidade de informação, apenas as 30 mais utilizadas foram empregues para não sobrelotar a nuvem.

A primeira parte deste inquérito consistiu num levantamento da informação relativa ao participante, da qual se destaca o grau de familiaridade com nuvens de tags, localidade de residência e sexo. As restantes partes da experiência são explicadas na secção seguinte.

4.2 Tecnologias e Ambientes de desenvolvimento usados

Esta secção apresenta sucintamente as plataformas e tecnologias estudadas durante a fase inicial do projecto e posteriormente utilizadas no desenvolvimento do mesmo. Para estas tecnologias, o foco principal são os proveitos e as contrariedades que trouxeram em usá-los no âmbito do estudo realizado.

C# (C SHARP)

O C# é uma linguagem orientada a objectos, foi criada pela Microsoft para uso na plataforma .NET, neste caso foi usada para desenvolver aplicações Web em ASP.NET.

Microsoft Visual Studio

Microsoft Visual Studio foi o ambiente de desenvolvimento escolhido para construir esta aplicação Web.

O Microsoft Visual Studio 2008 é a versão mais recente do IDE (Integrated Development Environment) da Microsoft.

Framework. NET

A Framework. NET é uma plataforma cuja função é facilitar e simplificar o desenvolvimento e utilização de componentes para a criação de aplicações Web.

Microsoft Sql Server 2005

SQL Server é um SGBD (Sistema de Gestão de Base de Dados), que permite gerir bases de dados relacionais, e que garante as seguintes funcionalidades:

- Os dados são guardados de forma correcta;
- As relações entre esses dados também são correctas;
- Capacidade de recuperar todos os dados até um ponto de consistência conhecido em caso de falha do sistema por diversas razões;
- Os dados sejam consumidos por aplicações desenvolvidas utilizando a *framework*. Net e o Microsoft Visual Studio.

4.3 Implementação

Para o desenvolvimento da experiência que irá ser relatada na secção 4.4 deste capítulo foi utilizado um *dataset* do sítio Delicious, contendo todas as tags com respectivas hiperligações, desde o ano 2003 até 2007, tendo sido utilizada apenas a informação relativa a Setembro de 2003 e Fevereiro de 2005 que corresponde a 1.615.451 registos como foi mencionado anteriormente. Esta selecção de datas foi aleatória, tendo sido considerado apenas como critério o tamanho dos ficheiros e sendo escolhidos os ficheiros mais pequenos do dataset do Delicious.

Note-se que o facto do repositório TREE estar em desenvolvimento, impossibilitou que fossem utilizados dados do mesmo.

O *dataset* original foi obtido pelo contacto com o autor de um dos artigos consultados para este trabalho (Wetzker, et al., 2008), Alan Said, o qual forneceu acesso aos registos catalogados por mês e ano no formato bz2 (ver Figura 42). Os registos foram importados para uma base de dados SQL Server 2005, o que permitiu uma melhor consulta e manipulação da informação. Note-se que foram obtidos mais de 1 milhão de registos, utilizando apenas 2 meses de informação de diferentes anos.

Index of /corpus/delicious

Name	Last modified	Size	Description
 Parent Directory		-	
 200309.bz2	09-May-2009 09:53	378K	
 200310.bz2	09-May-2009 09:53	495K	
 200311.bz2	09-May-2009 09:53	578K	
 200312.bz2	09-May-2009 09:53	738K	
 200401.bz2	09-May-2009 09:53	1.3M	
 200402.bz2	09-May-2009 09:53	1.7M	
 200403.bz2	09-May-2009 09:53	1.9M	
 200404.bz2	09-May-2009 09:53	2.2M	
 200405.bz2	09-May-2009 09:53	2.4M	
 200406.bz2	09-May-2009 09:53	2.9M	
 200407.bz2	09-May-2009 09:53	4.1M	
 200408.bz2	09-May-2009 09:53	4.9M	
 200409.bz2	09-May-2009 09:53	7.2M	
 200410.bz2	09-May-2009 09:54	9.5M	
 200411.bz2	09-May-2009 09:54	12M	
 200412.bz2	09-May-2009 09:54	15M	
 200501.bz2	09-May-2009 09:54	22M	
 200502.bz2	09-May-2009 09:54	25M	

Figura 42 - Dataset Delicious disponibilizado por ftp

4.3.1 Processo de implementação

Existem bastantes combinações para exibir uma nuvem de tags. Irão ser descritos de forma sequencial algumas das etapas para a construção de nuvens de tags

- Introdução de informação na base de dados
- Selecção das tags
 - Selecção das palavras com mais ocorrências
 - Selecção das palavras com mais peso ordenadas alfabeticamente.
- Aplicação dos algoritmos
 - Logarítmico
 - Linear
 - Verificação do maior número de ocorrências e atribuição de um "peso" representativo da sua importância
- Criação da nuvem de tags
 - Apresentação das tags ordenadas alfabeticamente ou por número de ocorrências.
 - Tags distinguidas por tamanho e estilo da fonte
- Apresentação através de uma lista rectangular
 - Como o sistema Flickr já referido anteriormente
- Disponibilização das nuvens de tags geradas pelos algoritmos estudados e implementados, a um grupo de utilizadores, que respondendo a algumas questões, irão permitir aferir qual a nuvem que permite maior usabilidade, qual a razão da escolha dos participantes, perceber até que ponto a apresentação das tags num modo ou de outro influencia a escolha.

Com os dados obtidos do sistema Delicious, conforme explicado anteriormente, foi possível desenvolver 2 nuvens de tags, com algoritmos diferentes mas com os mesmos dados.

Segundo Welie, a exposição das tags deve ser feita num espaço rectangular e deve contemplar a variação do tamanho e da fonte: *“The relative popularity of each tag (i.e. the amount of items having the tag divided by the total amount of items compared to the most popular tag) is then depicted by varying the font size, and sometimes also the font weight. The tags are usually in a rectangular area, either in the main content area if it is a page dedicated to tags or in the right-hand column if it is secondary to the main content”* (Welie, 2008).

4.4 Explicação da experiência

Sucessivamente foram apresentadas nuvens para as quais seria solicitada a selecção de uma determinada palavra da nuvem. O participante tinha como função observar a nuvem e clicar na palavra requerida (ver imagem seguinte).



Figura 43 - Algumas das nuvens de tags utilizadas na experiência.

A tabela seguinte exhibe as categorias de cada uma das nuvens mostradas no inquérito.

Tabela 8 - Nuvens apresentadas na experiência

	Algoritmo Linear		Algoritmo Logarítmico	
	Ordenada por peso	Ordenada alfabeticamente	Ordenada por peso	Ordenada alfabeticamente
Monocromática	Nuvem 4	Nuvem 6	Nuvem 5	Nuvem 1
Colorida	Nuvem 8	Nuvem 2	Nuvem 7	Nuvem 3

Os resultados obtidos do total de 28 participantes, com idade média de 27 anos (mínima de 12 anos, máxima de 45 anos), dos quais 10 do sexo feminino e 18 do sexo masculino, com localização maioritária na cidade do Porto e Matosinhos, mas também com participantes portugueses residentes na Holanda, estão registados na tabela seguinte. Nesta é possível visualizar-se o tempo consumido desde a apresentação da nuvem e respectiva palavra até a selecção da tag (ver Tabela 9).

Tabela 9 - Resultados do inquérito realizado com respectivos tempos consumidos por nuvem na pesquisa pela palavra solicitada pelo sistema.

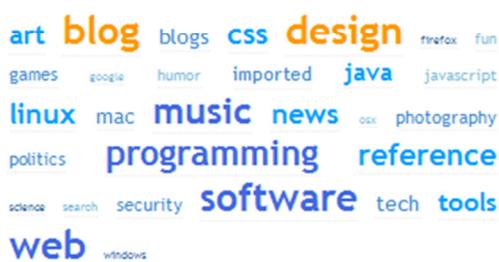
Tarefa	Descrição do algoritmo aplicado	Tempo mínimo consumido (s)	Tempo máximo consumido (s)	Tempo médio consumido (s)
1	Nuvem logarítmica monocromática ordenada alfabeticamente	6	41	20
2	Nuvem linear colorida ordenada alfabeticamente	6	28	14
3	Nuvem logarítmica colorida ordenada alfabeticamente	6	50	14
4	Nuvem linear monocromática ordenada por peso	3	43	15
5	Nuvem logarítmica monocromática ordenada por peso	7	46	17
6	Nuvem linear monocromática ordenada alfabeticamente	5	52	15
7	Nuvem logarítmica colorida ordenada por peso	3	30	13
8	Nuvem linear colorida ordenada por peso	7	35	14

Na selecção das palavras apresentadas pelo sistema, nem todos os participantes seleccionaram a tag solicitada correctamente. Para a nuvem 3, houve um participante que errou e para a nuvem 7 erraram 2 participantes.

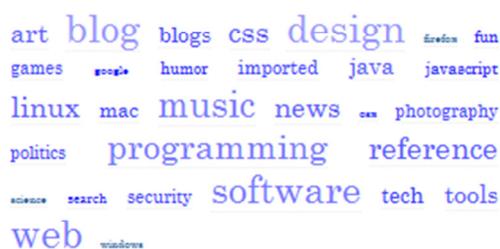
A cor avermelhada e verde, utilizadas em algumas células, permitem visualizar mais facilmente numa visão global os melhores e piores tempos gastos em cada em cada nuvem comparativamente com as restantes apresentadas no decorrer da experiência. O tipo de nuvem que melhor referência obteve foi a nuvem da tarefa 7 que obteve excelentes tempos nos critérios avaliados.

Para além do participante ter de encontrar a palavra solicitada nas diferentes nuvens apresentadas, o inquérito ainda tinha uma componente de avaliação qualitativa em que era requerida a opinião do indivíduo, consideradas 4 questões relacionadas com 5 nuvens de tags distintas, como pode ser verificado na Figura 44.

Nuvem 1



Nuvem 2



Nuvem 3



Nuvem 4



Nuvem 5



Figura 44 - Nuvens apresentadas ao participante para que este fornecesse resposta as questões apresentadas no inquérito.

As respostas revelaram que a nuvem 1 é a aparentemente mais simples de consultar e a mais indicada para colocar num blogue. As nuvens 1 e 4 destacaram-se como as esteticamente mais agradáveis, ao passo que a nuvem 4 foi maioritariamente seleccionada como a menos interessante.

Quando confrontados com as questões indicadas na Tabela 10, a nuvem 1 foi considerada a esteticamente mais agradável assim como a que seria aparentemente mais fácil de consultar. O número de escolhas para cada nuvem está registado na Tabela 10.

Tabela 10 - Respostas dos participantes quando comparavam diferentes nuvens de tags

Questão	Nuvem 1	Nuvem 2	Nuvem 3	Nuvem 4	Nuvem 5
1.Qual das seguintes nuvens lhe parece mais fácil de consultar?	17	3	3	3	2

2.Qual escolheria para colocar no seu blogue?	18	4	2	4	0
3.Qual a esteticamente mais agradável?	14	5	2	7	0
4.Qual a esteticamente menos agradável?	1	7	1	13	6

Um dado interessante que emerge neste estudo é que, apesar de os participantes terem indicado que a nuvem 1 (a que utilizou cores e algoritmo logarítmico e uma só fonte “Sans-Serif”) seria a mais fácil de consultar, confrontando esta escolha com os dados da Tabela 10, verifica-se que os tempos medidos apontam o *layout* com “algoritmo logarítmico e nuvem colorida ordenada por peso” da nuvem 7 como a mais rápida para consultar/detectar tags.

Outro ponto observado e que difere de outros estudos abordados anteriormente é que as nuvens que utilizam o algoritmo logarítmico apresentam piores tempos de resposta na obtenção da palavra pedida pelo sistema.

Nas nuvens com utilização do algoritmo logarítmico foi registado um tempo médio de resposta de 16 segundos, ao passo que com o algoritmo linear o tempo médio de resposta foi 14,5 segundos (ver Tabela 11). Assim, é possível concluir que neste estudo a disposição com o algoritmo linear oferece uma melhor harmonia na visualização, ao contrário dos resultados de outros estudos conduzidos por outros autores sobre este tema. A tabela seguinte mostra os tempos médios registados com base nos resultados na Tabela 10.

Tabela 11 - Tempos médios registados

	Tempo mínimo médio (s)	Tempo máximo médio (s)	Média do tempo gasto (s)
Algoritmo Logarítmico	5,5	41,75	16
Algoritmo Linear	5,25	39,5	14,5

Também através dos tempos médios obtidos e indicados na Tabela 12, é possível concluir que as nuvens de tags coloridas promovem a rapidez do utilizador na leitura das tags disposta na nuvem, independentemente do algoritmo, conforme os dados apresentados na tabela seguinte:

Tabela 12 - Tempos médios gastos nas nuvens coloridas e monocromáticas independentemente do algoritmo

	Nuvem de tags colorida (s)	Nuvem de tags monocromática (s)
Tempo médio	16,75	13,75

4.5 Avaliação do estudo

Após a análise do peso das tags neste estudo, verificamos que se não fossem calculadas as fontes algebricamente, e o tamanho fosse directamente relacionado com o número de ocorrências, existiriam tags muito pequenas e outras muito grandes, de tal modo que o sistema desenvolvido nem permitiria enquadrá-las numa página de Internet.

O número de tags que devem constituir uma nuvem depende do espaço que existe ao dispor. Numa página normal, em que a nuvem seja informativa na lateral da página o ideal será variar entre 30 a 50 tags, já numa página que tenha espaço e pretenda valorizar a nuvem de tags, esta deverá ter uma listagem com um grande número de tags para que sejam perceptíveis ao utilizador e permita a sua fácil navegabilidade.

No estudo efectuado verificou-se que é dada grande importância à disposição e utilização de cores nas nuvens assim como ao tipo de fontes utilizadas. Por exemplo, as nuvens 4 e 5 foram bastantes criticadas por não parecerem tão perceptíveis como as restantes, devido à utilização de mais do que um tipo de fonte.

Aparentemente também a adopção de fontes “*Sans-Serif*” influenciou a escolha de algumas nuvens por parte dos participantes.

Alguns comentários dos participantes permitem suportar as conclusões acima descritas:

- “Escolhia a Nuvem 2 por ter cores do mesmo tom e por estar ordenado alfabeticamente, o que torna mais fácil encontrar as palavras.”
- “A nuvem 1 além de ter várias cores que simplifica a consulta (embora não tenha percebido se têm algum significado), utiliza sempre o mesmo tipo de letra. A nuvem 4, além de ter apenas uma cor, tem alguns tipos de letra associados ao negrito que tornam-se bastante imperceptíveis.”
- “Nuvem 1 como esteticamente mais agradável e fácil de consultar: o tipo de letra e a variação das cores contribuem para uma leitura mais rápida dos tags. Além disso, do ponto de vista estético parece-me haver um equilíbrio mais eficaz a nível gráfico.”
- “É uma nuvem bastante colorida, tornando agradável na sua utilização. A nuvem 5 não é muito aconselhável pois tem poucas cores, sendo um bocado “seco” e a utilização das palavras em maiúsculas causa agressividade.”
- “A nuvem 1 será a que apresenta uma estética mais agradável e visualmente mais acessível para o utilizador final, porque o tipo de letra não é “carregado” e apresenta fácil leitura, tendo também a distinção das palavras mais importantes através de cores e tamanhos. A nuvem 4 é a pior, não gosto mesmo nada.”

Estes resultados permitem assim responder a um dos objectivos propostos neste documento, nomeadamente no que diz respeito ao tipo de letra, verificou-se que é grande a influência que os tipos de letra têm numa nuvem de tags. Segundo as opiniões dos participantes as fontes de letras espaçadas são as que facilitam a legibilidade ao utilizador.

Em relação às diferenças encontradas entre o estudo apresentado nesta dissertação e outros abordados na literatura, é possível que se devam a diferenças culturais resultantes de terem sido realizados em países distintos, ou diferenças nas amostras consideradas, as quais serão discutidas no capítulo da conclusão desta dissertação.

O algoritmo de visualização de nuvens de tags em utilização no repositório foi modificado, de forma a considerar os resultados do estudo de caso descrito neste documento como foi mencionado na secção de Implementação.

Observando os resultados obtidos, e tendo sido a nuvem 1 a eleita pela maioria dos participantes como a ideal a ser colocada num sítio Web, foi analisada se existe relação entre a escolha da nuvem 1 e o grau de experiência do utilizador que a seleccionou.

Dos 28 participantes de todo o estudo, apenas 17 elegeram a nuvem 1 como primeira resposta, à pergunta número 1 enunciada na Nuvem 9 do inquérito. O grau de experiência de cada utilizador foi registado previamente pelo próprio na primeira parte do inquérito e foi avaliado numa escala de 0 a 10, sendo que 0 não tinha qualquer contacto com este tipo de conceito e 10 estaria muito familiarizado, no entanto, o conceito de familiaridade com nuvens de tags pode ser interpretado de modo diferente por cada participante (ver Tabela 13).

Tabela 13 - Votos dos participantes classificados por grau de experiência com nuvens de tags às questões da nuvem 9.

	Grau experiência <5	Grau experiência >=5	Total de votos
Votou nuvem 1	6	11	17
Votou nuvem 2	1	2	3
Votou nuvem 3	0	3	3
Votou nuvem 4	2	1	3
Votou nuvem 5	2	0	2

Também foi analisado o tempo gasto a responder a cada pergunta do inquérito por cada grupo. Os resultados estão especificados na tabela seguinte:

Tabela 14 - Tempos gastos em cada nuvem por grau de experiência na utilização de nuvens de tags

	Grau experiência <5 (segundos)	Grau experiência >=5 (segundos)
Nuvem 1	17	20
Nuvem 2	11	19
Nuvem 3	17	12
Nuvem 4	14	14
Nuvem 5	15	18
Nuvem 6	15	14
Nuvem 7	13	13
Nuvem 8	14	13
Média de tempo	14,5	15,375

A primeira conclusão que se tira desta análise é que o tempo de resposta do servidor pode influenciar os dados registados na Tabela 14. O facto de o tempo de resposta do servidor demorar 2 segundos com um participante ou demorar 4 segundos com outro é o suficiente para criar resultados artificiosos. Para minimizar este factor, o estudo foi realizado na mesma semana, com igualdade de circunstâncias, ou seja, em horário laboral, para que o tráfego não fosse uma condicionante no tempo gasto em cliques nas respostas.

A conclusão seguinte que se obtém desta tabela de tempos gastos por grupo de experiência com nuvens de tags é que aparentemente os participantes com experiência igual ou superior a nível 5 gastaram mais tempo a analisar a nuvem para responder às questões.

Os grupos apresentam igual tempo de resposta quando se trata das nuvens 4 e 7, sendo que em comum estas nuvens são ordenadas por peso, ou seja, pode-se concluir que o facto de ter experiência com nuvens não influencia a escolha da nuvem e que de facto a ordenação por peso melhora a utilidade da nuvem de tags já que permite a ambos os grupos obter igual tempo de pesquisa.

Ainda se conclui que, num universo de 8 questões com as respostas temporizadas, os participantes que normalmente já contactam com nuvens de tags apenas obtêm melhores tempos em 3 perguntas (nuvens 3, 6 e 8).

Recorrendo a análise de dados utilizando ferramentas do Microsoft Excel obtém-se a seguinte informação que resulta de uma análise denominada T-Test (ver Tabela 15), este serve para comparação de resultados experimentais. Este teste foi desenvolvido por William Gosset em 1908. Gosset era químico numa cervejaria e desenvolveu o T-Test para

assegurar que cada lote de cervejas era o mais semelhante possível a todos os outros fabricados (Microbiologybites, 2009).

Tabela 15 - Resultado da análise T-Test

Teste T: duas amostras com variâncias iguais		
	<i>Grau Exp. <5</i>	<i>Grau Exp. >= 5</i>
Média	14,5	15,375
Variância	4	9,696428571
Observações	8	8
Variância agrupada	6,848214286	
Hipótese de diferença de média	0	
GI	14	
Stat t	-0,6687278	
P(T<=t) uni-caudal	0,25727261	
t crítico uni-caudal	1,761310136	
P(T<=t) bi-caudal	0,514545221	
t crítico bi-caudal	2,144786688	

A conclusão que se obtém da tabela acima é que a variância não é significativa, isto é, não existe relação entre e o grau de experiencia dos participantes tendo em conta o tempo gasto em cada tarefa do questionário. Esta informação é verificada pelo valor P(T<=t) bicaudal-que é superior a 0.05, este valor limite de comparação foi definido como sendo o valor de hipótese de diferença de média.

A fim de tentar perceber se a idade influenciou as respostas dadas a este tema, foram analisados os tempos de resposta dados pelos participantes com a diferenciação das idades como se pode ver na tabela seguinte:

Tabela 16 - Tempos gastos divididos por intervalo de idades

Tarefa	Idades 0-24 (s) *	Idades 25-29 (s) **	Idades 30-50 (s) ***
1	18	16	28
2	11	14	20
3	14	10	22
4	11	15	20
5	16	12	24
6	11	15	22
7	12	12	15
8	15	13	11

* Grupo com um total de 13 pessoas.

** Grupo com um total de 17 pessoas.

*** Grupo com um total de 10 pessoas.

Efectuada a análise de dados denominada ANOVA de factor único, recorrendo mais uma vez à ferramenta de análise de dados do Microsoft Excel, obtém-se a seguinte tabela:

Tabela 17 - Anova: factor único

SUMÁRIO						
Grupos	Contagem	Soma	Média	Variância		
0-24	8	108	13,5	7,142857		
25-29	8	107	13,375	3,982143		
30-50	8	162	20,25	27,64286		
ANOVA						
Fonte de variação	SQ	gl	MQ	F	valor P	F crítico
Entre grupos	247,5833	2	123,7917	9,579456	0,001106	3,4668
Dentro de grupos	271,375	21	12,92262			
Total	518,9583	23				

Analisando os dados da Tabela 17, verifica-se que do grupo mais novo para o grupo mais velho a diferença verificada no tempo médio consumo não é significativo. A hipótese nula traduz-se na não existência de diferença entre os 3 grupos comparados. Neste caso, com um nível de significância de 95% ($\alpha=0,05$) e uma vez que o valor verificado em F (9,5795) é maior do que F crítico (3,4668) então rejeitamos a hipótese nula de que os 3 grupos de idades desempenham um papel igual.

A idade média de um dos estudos analisados na literatura tinha uma média de idades de 26 anos (Lohman, et al., 2009). Uma vez que a análise ANOVA (Plonsky, M., 2009) mostra que existe diferença de comportamento consoante o intervalo de idade do participante, então é possível afirmar que a idade poderá ter sido um factor de influência para a diferença de conclusões obtidas.

Capítulo V. Conclusões

Neste capítulo avalia-se o trabalho desenvolvido no âmbito desta dissertação de mestrado e mencionam-se os objectivos que foram concretizados. Enumeram-se as limitações do estudo realizado, e por último, referem-se alguns aspectos que poderão estar na base de um trabalho futuro na área de nuvens de tags.

5.1 Avaliação do trabalho realizado

Nuvens de tags podem ser consideradas um instrumento de gestão de sobrecarga de informação. Actualmente existe imensa informação disponível na Internet, esta informação é crescente e irá saturar-se. Um modo simples e rápido de permitir a automatização e sumarização desta informação é recorrendo as nuvens de tags.

O objectivo primordial dos Repositórios de Objectos Educativos é promover a reutilização dos objectos educativos, o que passa também por permitir que sejam facilmente encontrados pelos seus utilizadores. Neste documento descreveu-se o estudo que foi efectuado de forma que a utilização de tags e nuvens de tags constituíssem ferramentas poderosas na pesquisa e recuperação de informação.

Foi conclusiva a influência que existe na distribuição de tags numa nuvem. Os algoritmos que foram estudados neste documento (linear e logarítmico), que permitem uma distribuição razoavelmente uniforme pelo espaço disponível, com conjugação de diferentes modos de ordenação e utilização de cores distintas ou não, proporcionaram ao utilizador diferentes perspectivas.

Também foi possível concluir que a disposição e o algoritmo utilizados em cada nuvem de tags varia de acordo com os seguintes factores: espaço disponível, contexto do sítio Web e interface que se pretende mostrar ao utilizador.

Observando os resultados obtidos pelo estudo empírico, conclui-se que a nuvem a ser implementada no repositório deve ter as características da nuvem número 7 (ver Figura 43 da secção 4.4), a que utiliza o algoritmo logarítmico e onde se apresentam tags coloridas ordenadas por peso, ou seja, pelo maior número de ocorrências. A ideia de Will McGugan (McGugan, 2007) relativamente a ordenação por peso ser mais eficiente do que a ordenação alfabética, foi confirmada no estudo efectuado.

No entanto, quando verificados os resultados da nuvem 9, averiguou-se que a maior parte dos participantes elegeram a nuvem 1 (que utiliza o algoritmo logarítmico, com tags coloridas e ordenadas alfabeticamente) como a mais provável escolha para colocar num sítio Web seu. Assim, o algoritmo logarítmico e a utilização de múltiplas cores constituem a melhor escolha para a criação de uma nuvem de tags, segundo o estudo realizado. No entanto, o utilizador final deve escolher entre a eficiência de uma ordenação por peso, e a apresentação mais agradável de uma ordenação alfabética.

É de frisar também a importância e diferença de resultados encontrados em relação aos trabalhos conduzidos por autores de outros estudos abordados nesta dissertação. Todos os estudos apontavam o algoritmo logarítmico como melhor opção de implementação e a escolha dos utilizadores para a construção de nuvens. No entanto, analisando os resultados desta experiência, e tendo em conta que a nuvem 7 foi a mais votada, o estudo revela que o algoritmo linear foi o que apresentou tempos mais curtos na obtenção da palavra solicitada pelo sistema e uma melhor harmonia na visualização.

Os dados obtidos permitem afirmar que a teoria proposta por Kevin Hoffman (Hoffman, 2006) neste estudo não foi confirmada, isto é, segundo os resultados apurados o algoritmo logarítmico mesmo evidenciando todas as tags, não promoveu a melhor harmonização na visualização.

A ordenação das tags dentro da nuvem também é um ponto que está em desconformidade com o estudo realizado por Halvey e Keane. No estudo, exposto na secção 3.4.1 Caso 1 – Leitura e apresentação de nuvens de tags, os autores referem que a ordenação alfabética favorecia o processo de pesquisa de palavras e melhorava o tempo de tempo gasto na pesquisa da palavra. No estudo realizado nesta dissertação o público inquirido obteve melhores resultados nos tempos de procura das palavras nas nuvens com ordenação por peso.

Halvey e Keane também fazem referência à variação da fonte. Neste estudo apenas foi abordada esta questão na Nuvem 9, em que os participantes justificaram algumas das suas escolhas de nuvens para utilizarem no seu blogue. Esta avaliação não foi quantitativa mas sim qualitativa, logo não foi possível recolher informação suficientemente válida para que a variação de letra fosse um critério que influenciasse a escolha do utilizador. No entanto, tanto os resultados apurados como alguns dos comentários permitem corroborar a teoria de Lohman, os participantes revelaram antepor a estética e a apresentação da nuvem à sua eficiência ou funcionalidade, isto é, dão prioridade à nuvem esteticamente mais atraente do que à nuvem comprovadamente mais funcional e rápida.

Foi ainda provado no capítulo 4 na secção 4.5 que a experiência com nuvens não influencia a escolha da nuvem, foi verificado que tanto os indivíduos com experiência prévia em nuvens de tags como os indivíduos que nunca contactaram com este tipo de visualização obtiveram igual tempo de pesquisa- Pode-se então afirmar que a simplicidade deste tipo de visualização traz vantagens para indivíduos que procurem informação, mesmo que tenham pouco ou nenhum contacto com este tipo de visualização, o que constitui uma mais-valia das nuvens de tags. Este tipo de visualização de informação torna-se assim um benefício num repositório de dados sabendo que este será acedido por alunos, professores ou funcionários e que é desconhecido o grau de experiência de quem pretenda recolher informação contendo objectos educativos e padrões genéricos de aprendizagem catalogados por tags.

Foi de enorme importância os resultados obtidos nos estudos de caso referidos na literatura e os estudos abordados no âmbito deste trabalho, pois permitiram um termo de comparação com o estudo experimental realizado.

Este trabalho também foi um meio de fomentar entre a comunidade académica este conceito, pois apesar dos participantes serem pessoas ligadas ao mundo das tecnologias e Web 2.0, uma grande parte ainda desconhecia este tipo de visualização.

Concluindo, as diferenças culturais dos participantes que intervieram nesta experiência comparativamente com os participantes de outros estudos abordados na literatura ou as diferenças nas amostras consideradas, poderão ser as explicações mais viáveis para a discordância das conclusões acima mencionadas. Note-se que estas duas hipóteses não foram exploradas como sucedeu com a análise da influência da idade dos participantes relatada no capítulo 4.5. O factor idade pode ter sido uma condicionante na diferença de resultados comparativamente com outros estudos, conforme foi referido no capítulo 4.

No âmbito desta dissertação de mestrado foi escrito um artigo científico intitulado “Visualização com nuvens de tags”, aceite no I Encontro Internacional TIC e Educação (<http://ticeduca.ie.ul.pt>), que se realiza de 19 a 20 de Novembro em Lisboa.

5.2 Limitações

Foram convidadas 98 pessoas para participar no estudo apresentado nesta dissertação, das quais apenas 28 aceitaram a proposta. Isso limitou o universo de respostas.

O trabalho realizado não contemplou de forma significativa o estudo sobre a disposição das nuvens de tags, isto é, não foi possível estudar especificamente qual a importância que o

tamanho da caixa de exposição de tags e qual a proporção das tags dentro da área a utilizar. No entanto, esta limitação foi antecipada, pelas seguintes razões:

- Já existem dados suficientes obtidos pelas várias conclusões obtidas noutros estudos semelhantes, que foram anteriormente citados (ver secção 3.2).
- Não desmotivar a participação das pessoas. Um inquérito com muitas perguntas e/ou que obriguem a um dispêndio significativo de tempo, é automaticamente descartado pela maior parte do público.

Note-se que com o objectivo de angariar o máximo de participantes possível, optou-se por um estudo que permitisse obter conclusões e estas fossem comparáveis com os estudos já realizados noutros países.

5.3 Trabalho futuro

Durante a realização desta dissertação e com a literatura lida para investigação deste tema, foi possível encontrar algumas lacunas que mostram que a investigação sobre nuvens de tags deve continuar, inclusive noutras áreas.

Como trabalho futuro, seria interessante abordar dois tópicos, um seria aliado à visualização, outro seria dar continuidade à experiência realizada nesta dissertação.

✓ Visualização com tags:

- Aplicações para PDAs; Não existe nada no mercado que permita utilizar nuvens de tags em telemóveis, pda, entre outros dispositivos. No entanto, um aficionado por nuvens de tags e programação resolveu desenvolver para o seu iPhone um aplicativo para mostrar as suas categorias de música. No entanto, encontrou imensas dificuldades, pois não existem frameworks nem classes SDK capazes de simplificar a tarefa, o programador teve que implementar esta ideia em UI e OpenGL. Esta implementação pode ser visualizada em <http://surgeworksmobile.com/iphone/how-to-build-a-3d-tag-cloud-on-iphone---challenges-and-solutions>. Futuramente a aposta neste conceito para dispositivos móveis que cada vez mais são direccionados para o uso exclusivo *touch screen* seria interessante. A criação de software aberto capaz de responder a estas necessidades, ou que permita a criação de nuvens de tags para aparelhos deste tipo deve ser considerada no futuro.

Bibliografia

Adelman C. Re-thinking case study: notas from the second Cambridge Conference [Relatório]. - [s.l.] : Cambridge Journal of Education, 1976.

Assche F.V [et al.] MELT Final Report (1/10/2006 – 31/03/2009) [Relatório]. - 2009. - Available at http://info.melt-project.eu/shared/data/melt/MELT_1_3_Final_Project_Report.pdf. - ECP 2005 EDU 038103.

Azevedo I. [et al.] Applying and Reusing Knowledge in a Repository [Conferência] // The 10th IEEE International Conference on Advanced Learning Technologies (ICALT 2010). - Sousse, Tunisia : [s.n.], 2010.

Bateman Scott, Gutwin Carl e Nacenta Miguel Seeing things in the clouds: the effect of visual features on tag cloud selections [Conferência] // Proc. of the 19th ACM conference on Hypertext and Hypermedia. - New York : ACM Press, 2008. - pp. 193–202.

Bielenberg K. e Zacher M. Groups in social software: Utilizing tagging to integrate individual contexts for social navigation. [Relatório] : Master's thesis / University of Bremen. - [s.l.] : Program of Digital Media, 2005.

Bielenberg K. e Zacker M. Groups in social software: Utilizing tagging to integrate individual contexts for social navigation [Relatório] / Master's thesis ; Universitat Bremen. - 2005.

Bogdan Robert e Biklen Sari Investigação qualitativa em educação: uma introdução à teoria e aos métodos [Livro]. - Porto : Porto Editora, 1994.

Borg W e Gall M Educational research: An introduction [Livro]. - Nova Iorque : Longman, 2002. - Vol. 7º.

Bravo R Técnicas de Investigación Social [Livro]. - Barcelona : Editorial parainfo sa, 1991. - Vol. 7º.

C.H. Brooks and N. Montanez Improved annotation of the blogosphere via autotagging and hierarchical clustering [Conferência] // WWW '06: Proceedings of the 15th international conference on World Wide Web (ACM Press). - 2006. - Edinburgh, Scotland.

Chow K.O. e Cheung K.S. Enhancing Learning Through Technology - Research on Emerging Technologies and Pedagogies [Livro] = A study on tag cloud quality in e-learning 2.0 / ed. Kwan Reggie [et al.]. - [s.l.] : World Scientific Publishing Co. Pte. Ltd., 2008. - pp. 63-79.

- Cohen L. e Manion L.** Research Methods in Education [Livro]. - London : Routledge, 1989.
- Coutinho C.** Del.icio.us:uma ferramenta da Web 2.0 ao serviço da investigação em educação. [Relatório]. - [s.l.] : Educação, Formação & Tecnologias, 2008. - pp. 104-115. - Disponível em:<http://eft.educom.pt/index.php/eft/article/viewFile/24/26>.
- Ghiglione R. e Matalon B.** O Inquérito, Teoria e Prática [Livro]. - Oeiras : Celta Editora, 1997.
- Golder S.A. e Huberman B.A.** Usage patterns of collaborative tagging systems [Jornal] // Journal of Information Science. - 2006. - Vol. 32(2). - pp. 198–208.
- Halvey Martin e Keane Mark T.** An Assessment of Tag Presentation Techniques [Conferência] // Proceedings of the 16th international conference on World Wide Web / ed. School of Computer Science and Informatics UCD Dublin. - Banff, Alberta, Canada : New York, NY, US, 2007. - Vols. ACM 978-1-59593-654-7/07/0005. - pp. 1313-1314. - <http://www2007.org/htmlposters/poster988/>.
- Hassan-Monteroa Yusef e Herrero-Solana Víctor** Improving Tag-Clouds as Visual Information Retrieval Interfaces [Conferência] // I International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006. - Mérida, Spain : [s.n.], 2006. - October 25-28.
- Hayman Paul** Creating a Tag Cloud in C# [Online] // geekzilla. - 22 de August de 2006. - 01 de 2010. - <http://www.geekzilla.co.uk/View960C74AE-D01B-428E-BCF3-E57B85D5A308.htm>.
- Heather G. e Hof Robert D** Picking Up Where Search Leaves Off [Entrevista]. - [s.l.] : BusinessWeek, 11 de Abril de 2005.
- Hoffman Kevin** Whitepaper : In Search of the Perfect Tag Cloud [Online] = inserachofperfecttagcloud.pdf // .Net Developers Journal / ed. Ferguson Derek. - 11 de Agosto de 2006. - Maio de 2010. - <http://dotnetaddict.dotnetdevelopersjournal.com/tw.htm>.
- Kaser O. e Lemire D.** Tag-cloud drawing: Algorithms for cloud visualization [Conferência] // Proceedings of the Workshop on Tagging and Metadata for Social Information Organization (WWW2007). - Banff, Alberta, Canada : [s.n.], 2007.
- Kumar Ranjit** Research Methodology: A Step-by-Step Guide for Beginners [Livro]. - [s.l.] : Sage Publications, 2005. - p. 129.

Kuo B. Y. L. [et al.] Tag clouds for summarizing web search results. [Conferência] // Proceedings of the 16th International World Wide Web Conference (WWW2007). - New York : ACM Press, 2007. - pp. 1203–1204.

Lamanti Joe 10 Best Practices For Displaying Tag Clouds [Online] // Joe Lamantia.com. - 25th de February de 2007. - Maio de 2010. - <http://www.joelamantia.com/tag-clouds/10-best-practices-for-displaying-tag-clouds>.

Lohman Steffen, Ziegler Jürgen e Lena Tetzlaff Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration [Conferência] // Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I. - Uppsala, Sweden : Springer-Verlag Berlin, Heidelberg, 2009. - pp. Section: Part Two: Long and Short Papers: HCI and Web Applications 1 (392 - 404).

Mathes A. Folksonomies - Cooperative Classification and Communication Through Shared Metadata [Relatório]. - [s.l.] : <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>, 2004.

Mathes Adam Folksonomies - Cooperative Classification and Communication Through Shared Metadata [Relatório] / Computer Mediated Communication ; University of Illinois Urbana - Champaign. - 2004. - Graduate School of Library and Information Science.

McGugan Will Tag clouds look better sorted [Online] // Will McGugan's Blog. - 31st de October de 2007. - Maio de 2010. - <http://www.willmcgugan.com/2007/10/31/tag-clouds-look-better-sorted/>.

Microbiologybites Inferential Statistics- Comparing Groups I [Online] // Microbiologybites. - February de 2009. - 2010. - <http://www.microbiologybytes.com/math/1011-20.html>.

Pereira Roberto Folkauthority: a aplicação do conceito de autoridade cognitiva por meio de folksonomia [Relatório]. - Maringá : [s.n.], 2008. - p. 161.

Plonsky, M. Analysis of Variance - One Way [Online] // Psychological Statistics. - APA, 2009. - <http://www.uwsp.edu/psych/stat/12/anova-1w.htm>.

Poole Alex Literature Review - Which Are More Legible: Serif or Sans Serif Typefaces? [Online] // Alex Poole - Interaction design and research. - 7 de April de 2005. - 07 de 2010. - <http://www.alexpoole.info/academic/literaturereview.html#Bernard>.

Qualman Erik Socialnomics: How social Media Transforms The Way We Live And Do Business [Livro]. - Hoboken, New Jersey : John Wiley & Sons Inc., 2009.

Quintarelli Emanuele Folksonomies: power to the people [Online] // ISKO Italia. Documenti. - 24 de June de 2005. - Abril de 2010. - <http://www.iskoi.org/doc/folksonomies.htm#overview>.

Reis Guilherme Arquitetura de Informação, usabilidade e webdesign [Online] // GUILHERMO.com. - 2007. - Abril de 2010. - http://www.guilhermo.com/ai_biblioteca/artigo.asp?referencia=288.

Riu Douglas Vantagens e desvantagens da folksonomia [Online] // Design, web design. - 2 de September de 2008. - Abril de 2010. - <http://douglasriu.wordpress.com/2008/09/02/vantagens-e-desvantagens-da-folksonomia/>.

Rubinstein R Digital Typography [Relatório]. - [s.l.] : Addison Wesley Longman, 1988.

Russel Terrel Cloudalicious: Folksonomy Over Time [Relatório] / School of information and Library Science ; University of North Carolina at Chapel Hill. - 2006.

Seiça R. [et al.] Tree - Um Repositório de Objetos de Aprendizagem para Engenharia baseado em Software Aberto. [Conferência] // FINTDI (Fomento e Innovación con Nuevas Tecnologías en la Docencia de la Ingeniería) . - Vigo, Spain : [s.n.], 2009.

Seifert Christin [et al.] On the beauty and usability of tag clouds [Jornal] // Proceedings of the 12th International Conference information Visualisation. - 2008. - pp. 17-25.

Sen S. [et al.] tagging, communities, vocabulary, evolution [Conferência] // Conference on Computer Supported Cooperative Work Banff. - Alberta, Canada : [s.n.], 2006. - pp. 181-190.

Shaw B. Utilizing folksonomy: Similarity metadata from the del.icio.us system [Relatório] : Project Proposal. - 2005.

Shirky Clay Folksonomies + controlled vocabularies [Online] // Corante. - 7 de January de 2005. - Abril de 2010. - http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php.

Sinclair James e Cardew-Hall Michael The Folksonomy Tag Cloud: When is it Useful? [Jornal] // Journal of Information Science. - 2007. - Vol. XX(X). - pp. 1-18.

Smith Gene Tagging: People Powered Metadata for the Social Web [Livro]. - United States of America : New Riders, 2008. - pp. 72-102.

Soares Anabela Henriques de Matos A Química e a imagem da ciência e dos cientistas na banda desenhada : uma análise de livros de B. D. e de opiniões e interpretações de

investigadores, professores de C. F. Q. e alunos do 3º ciclo [Relatório] = Dissertação de Mestrado em Química - Especialização em Ensino / Universidade do Minho. - 2004. - Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/590>.

Stefaner M. Visual tools for the socio-semantic web [Relatório] : Master's thesis / University of Applied Sciences Potsdam. - 2007.

Swanson Gunnar Graphic Design and Reading : explorations of an uneasy relationship [Livro]. - Ventura : Allworth Press, 2000. - pp. 99-106.

Takeharu Eda [et al.] Signaling Emotion in Tagclouds [Relatório]. - Madrid : [s.n.], 2009.

Techbits Techbits [Online] // Uma nuvem vale por mil palavras. - 18 de Janeiro de 2007. - Abril de 2010. - <http://techbits.com.br/2007/uma-nuvem-vale-por-mil-palavras/>.

Times Matthew Ericson / The New York The Words They Used [Online] // The New York Times (Politics). - 4 de Setembro de 2008. - Agosto de 2010. - http://www.nytimes.com/interactive/2008/09/04/us/politics/20080905_WORDS_GRAPHIC.html.

Venkat The Dekoh Blogs [Online] // Choosing a good font size variation algorithm for your tag cloud. - 29 de OCTOBER de 2007. - 01 de 2010. - <http://blogs.dekoh.com/dev/2007/10/29/choosing-a-good-font-size-variation-algorithm-for-your-tag-cloud/>.

Wal Thomas Vander Folksonomy [Online] // VanderWal.net. - 2 de Fevereiro de 2007. - 31 de 03 de 2010. - <http://www.vanderwal.net/folksonomy.html>.

Welie Martijn van Welie.com [Online] // Patterns in Interaction Design - Tag Cloud. - 2008. - <http://www.welie.com/patterns/showPattern.php?patternID=tag-cloud>.

Wetzker Robert e Zimmermann Carsten & Bauckhage, Christian Analyzing Social Bookmarking Systems: A del.icio.us Cookbook [Online]. - 2008. - disponível em: http://www.dai-labor.de/fileadmin/Files/Publikationen/Buchdatei/wetzker_delicious_ecai2008_final.pdf.

Wikipédia Serifa [Online] // Tipografia. - 29 de Abril de 2010. - Maio de 2010. - <http://pt.wikipedia.org/wiki/Sans-serif>.

Wilson Ralph F. HTML E-Mail: Text Font Readability Study [Online] // Web Marketing Today. - 1 de 3 de 2001. - 08 de 2010. - <http://www.wilsonweb.com/wmt6/html-email-fonts.htm>.

Anexo A – Inquérito realizado

Neste anexo é apresentada a sequência cronológica dos ecrãs que constituíram o inquérito apresentado aos participantes do estudo experimental realizado nesta dissertação.

Anexo A.1. Apresentação

Imagem do ecrã inicial com instruções e breve descrição do tema do trabalho.



The screenshot displays the initial screen of the 'tagclouds' application. At the top, there is a logo with two green speech bubbles and the text 'tagclouds' in blue and red, followed by 'visualização com nuvens de tags'. Below this, the interface is split into two main sections. On the left, under the heading 'nuvens de tags', there is a 'TAGCloud' logo and a login form with fields for 'Nome' and 'Senha', and a 'Validar' button. On the right, under the heading 'Visualização com nuvens de tags', the author 'Joana Silva' is listed with her affiliation 'ISEP | Engenharia Informática'. A paragraph of text explains the research project, and a link 'Solicitar credenciais...' is provided. Below this is a 'Definição' section with a descriptive paragraph. At the bottom of the page, a footer contains the text 'Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP' and a small copyright notice.

Anexo A.2. Perguntas abertas

Imagens das perguntas fechadas do inquérito.

tagclouds visualização com nuvens de tags

tarefa 1
Por favor detecte a seguinte palavra
reference

Nuvem 1

imported art blog blogs books games google howto
design development firefox flash fun funny linux mac media
humor imported internet irc java javascript linux mac media
music news oss photography php politics programming
reference science search security shopping software
tech technology tools travel travel ukraine ukraine web webdesign windows
art

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 2
Por favor detecte a seguinte palavra
tech

Nuvem 2

imported art blog blogs books games google howto
development firefox flash fun funny games google howto
irc java javascript linux mac media music news oss photography php
politics programming python reference science search security shopping
software tech technology tools travel travel ukraine ukraine web webdesign
windows art

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 3
Por favor detecte a seguinte palavra
firefox

Nuvem 3

imported art blog blogs books games google howto
design development firefox flash fun funny games google howto
humor imported internet irc java javascript linux mac media
music news oss photography php politics programming
python reference science search security shopping software
tech technology tools travel travel ukraine ukraine web webdesign windows
art

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 4
Por favor detecte a seguinte palavra
tools

Nuvem 4

design blog web software music programming
css reference new(1998) art tools java blogs tech mac imported
politics photography games security javascript fun humor search google oss science firefox
windows funny python books howto development flash delicious internet imported irc technology php
irc irc shopping webdesign travel media art ukraine ukraine ukraine ukraine ukraine ukraine ukraine
language history

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 5
Por favor detecte a seguinte palavra
books

Nuvem 5

design blog web software music programming
css reference news linux art tools java blogs tech mac
imported politics photography games security javascript fun humor
search google oss science firefox windows funny python books howto development
flash delicious internet imported irc technology php irc irc shopping webdesign
travel media art ukraine ukraine ukraine ukraine ukraine ukraine ukraine ukraine
language history

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 6
Por favor detecte a seguinte palavra
ukraine

Nuvem 6

imported art blog blogs books games google howto
development firefox flash fun funny games google howto
irc java javascript linux mac media music news oss photography php
politics programming python reference science search security shopping
software tech technology tools travel travel ukraine ukraine ukraine ukraine ukraine
windows art

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 7
Por favor detecte a seguinte palavra
ukraine

Nuvem 7

design blog web software music programming
css reference news linux art tools java blogs tech mac
imported politics photography games security javascript fun humor
search google oss science firefox windows funny python books howto development
flash delicious internet imported irc technology php irc irc shopping webdesign
travel media art ukraine ukraine ukraine ukraine ukraine ukraine ukraine ukraine
language history

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

tagclouds visualização com nuvens de tags

tarefa 8
Por favor detecte a seguinte palavra
art

Nuvem 8

design blog web software music programming
css reference news linux art tools java blogs tech mac imported
politics photography games security javascript fun humor search google oss science firefox
windows funny python books howto development flash delicious internet imported irc technology php
irc irc shopping webdesign travel media art ukraine ukraine ukraine ukraine ukraine ukraine ukraine
language history

Desenvolvido no âmbito da tese do Mestrado de Engenharia Informática por Joana Silva | ISEP

Anexo A.3. Perguntas abertas

Imagem do ecrã com perguntas abertas.

 **Nuvem 9**

Qual das seguinte nuvens lhe parece mais fácil de consultar? 1 2 3 4 5

Qual escolheria para colocar no seu blogue? 1 2 3 4 5

Qual a esteticamente mais agradável? 1 2 3 4 5

Qual a esteticamente menos agradável? 1 2 3 4 5

Porquê das suas escolhas? (não necessita comentar todas, pelo menos uma das suas escolhas)

Nuvem 1

art **blog** blogs **css** **design** firefox fun
games google humor imported **java** javascript **linux**
mac **music** **news** osx photography politics
programming **reference** sterna search
security **software** **tech** **tools** **web**
windows

Nuvem 2

art **blog** blogs **css** **design** firefox fun
games google humor imported **java** javascript
linux mac **music** **news** osx photography
politics **programming** **reference**
sterna search **software** **tech** **tools**
web windows

Nuvem 3

art **blog** blogs **css** **design** firefox fun
games google humor imported **java** javascript
linux mac **music** **news** osx
photography politics **programming**
reference sterna search security
software **tech** **tools** **web** windows

Nuvem 4

art **BLOG** BLOGS **css** **DESIGN** firefox FUN
GAMES GOOGLE HUMOR IMPORTED **java** **JAVASCRIPT** **linux**
MAC **music** **news** osx PHOTOGRAPHY POLITICS
programming **reference** sterna **SEARCH**
SECURITY **software** TECH **tools** **web** windows

Nuvem 5

art **BLOG** BLOGS **css** **DESIGN** firefox FUN
GAMES GOOGLE HUMOR IMPORTED **java** **JAVASCRIPT** **linux**
MAC **music** **news** osx PHOTOGRAPHY POLITICS
programming **reference** sterna **SEARCH**
SECURITY **software** TECH **tools** **web** windows

Anexo A.4. Finalização

Imagem que finalizava o questionário.

