

Molecule: Sistema de organização e visualização de *tags*

José Nuno Almeida Neves Vigário Martins

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas Gráficos e Multimédia**

Orientadora: Isabel Azevedo

Júri:

Presidente:

Fátima Rodrigues, ISEP

Vogais:

Paulo Maio, ISEP

Isabel Azevedo, ISEP

Porto, Outubro 2013

Resumo

No dia-a-dia existe regularmente a necessidade de rotular um item com informação adicional de forma a poder ser mais facilmente recuperado ou identificado posteriormente. Diversas plataformas permitem que os utilizadores rotulem recursos com *tags* que habitualmente são partilhadas com outros utilizadores. Assim, ao longo do tempo foram propostas várias formas de visualização das *tags* associados aos recursos, com o intuito de não só facilitar aos utilizadores a pesquisa dos mesmos, mas também permitir a visualização do *tag space*. A nuvem de *tags* destaca-se como a forma mais comum de visualização.

Este documento apresenta um estudo efetuado sobre formas de visualização de *tags*, as suas vantagens e limitações, e propõe uma forma de visualização alternativa. Sugere-se também uma nova interpretação sobre como pesquisar e visualizar recursos com *tags* associadas: o sistema Molecule, uma solução viável e inovadora, para vários dos problemas associados à tradicional nuvem de *tags* que, para além de permitir aos seus utilizadores associem *tags* aos recursos, proporciona uma abordagem multivista para os mesmos navegarem no *tag space* e pesquisarem informação.

Palavras-chave: *Tag*, Nuvem de *tags*, Similaridade, Sinónimo, Tradução, Visualização de *tags*

Abstract

In the day-to-day life, there is regularly the need to label an item with additional information so that it can be more easily recovered or identified later. Several platforms allow users to tag resources and share information. Over time various forms of tags visualization have been proposed in order to facilitate the visualization of the tag space and the search and retrieval of resources. The tag cloud stands out as the most common form of tags visualization.

This dissertation presents a study carried out on tags visualization, their characteristics and proposes an alternative. It also suggests a new interpretation on how to search and view resources based on their tags. As a result of this interpretation, a viable and innovative system was achieved, Molecule, that overcome some of the problems associated with the traditional tag cloud, that in addition to enable its users to associate tags to their resources, provides a multiview approach to navigate in the tag space and to shearch for information.

Keywords: Tag, Tag cloud, Similarity, Synonymous, Translation, Tags visualization

Agradecimentos

Concluir esta etapa é o culminar de um trajeto que encerra uma fase importante e decisiva na minha vida e, como não poderia deixar de ser, não posso passar sem apresentar alguns breves, mas sentidos, agradecimentos.

Começaria por agradecer mais uma vez aos meus pais pois sem eles esta caminhada, que agora encerro, não teria sido possível.

Agradeço a todos os meus amigos, com quem contei nos bons e maus momentos, e sem os quais seguramente esta viagem não teria sido a mesma. Eles sabem quem são, e mais que amigos, são como irmãos para mim.

À professora Isabel Azevedo, minha orientadora, por toda a ajuda dada durante a realização deste trabalho, a qual se mostrou sempre disponível para me guiar, tirar dúvidas e tecer críticas construtivas de modo a aumentar a qualidade deste trabalho e o poder levar a bom porto.

Ao departamento de Engenharia Informática, na pessoa da sua diretora, a Dra. Ana Maria Almeida, pela disponibilização de um computador que serviu para alojar o sistema Molecule, acessível assim de qualquer ponto do mundo. Também aos técnicos de informática do departamento e ao engenheiro Luís Pereira, agradeço o apoio dado neste processo.

Agradeço também ao engenheiro Michel Almeida, por toda a disponibilidade e ajuda dada na fase final deste processo.

À Ana, minha namorada, por todo o apoio, ajuda e compreensão na elaboração deste trabalho. Sem a sua ajuda, não teria sido possível finalizar com sucesso este desafio.

Por fim, mas não por último, a todos aqueles que de alguma forma me ajudaram a chegar a esta fase, desde os colegas de curso no instituto, aos colegas das várias empresas por onde passei. Com todos aprendi e cresci muito.

A todos, o meu muito obrigado!

Índice

1	Introdução	1
1.1	Contexto	1
1.2	Objetivos.....	3
1.3	Estrutura da dissertação	3
2	Tags	5
2.1	Tipos de tags	5
2.1.1	Tags objetivas	6
2.1.2	Tags subjetivas	6
2.2	Sistemas que utilizam tags	8
2.2.1	Delicious	9
2.2.2	Yahoo! Bookmark.....	10
2.2.3	CiteULike	11
2.2.4	Flickr	12
2.2.5	Last.fm	13
2.2.6	Technorati.....	14
2.2.7	Upcoming	14
2.2.8	Amazon.....	15
2.2.9	Comparação entre plataformas.....	15
2.3	Formas de visualização de tags	18
2.3.1	Tag Clouds.....	18
2.3.2	Outras formas de visualização	20
2.3.3	Comparação entre formas de visualização	30
2.4	Limitações conhecidas às formas de visualização de tags	32
2.4.1	Limitação da amostragem do <i>tag space</i>	32
2.4.2	Sistemas que colmatam algumas das limitações	34
3	Molecule - Sistema de organização e visualização de tags	37
3.1	Introdução	37
3.2	Formas de organização da informação	38
3.2.1	Tradução de termos	39
3.2.2	Sinonímia de termos	39
3.2.3	Similaridade de termos.....	39
3.2.4	Termos mais utilizados	40
3.3	Vantagens da forma de organização do sistema Molecule	41
3.4	Modelação gráfica do sistema Molecule	42
3.4.1	Átomos	42
3.4.2	Elétrões	43
3.4.3	Diferentes moléculas geradas pelo sistema	44
3.4.4	Vantagens da abordagem de átomos comparativamente às nuvens de <i>tags</i> convencionais	45

3.5	Especificação técnica da aplicação	47
3.5.1	Descrição das tecnologias mais relevantes	47
3.5.2	Vistas	50
3.5.3	Arquitetura da aplicação.....	58
3.5.4	Esquema Relacional.....	60
3.5.5	Interfaces (WCF).....	61
3.5.6	Outros trabalhos realizados	64
4	Conclusões	69
4.1	Objectivos atingidos	69
4.1.1	Inovações introduzidas pelo sistema Molecule	70
4.2	Limitações e trabalho futuro	70
4.3	Apreciação Final	71
	Anexo A - Artigo aceite na 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI)	77
	Anexo B - Artigo publicado na revista RISTI número 11 de Junho de 2013.....	85

Lista de Figuras

Figura 1 - Modelo de um sistema de <i>tags</i>	2
Figura 2 - Comparação de frequência de algumas <i>tags</i> utilizadas nos sistemas Citeulike, Connotea e Delicious.	6
Figura 3 - Comparação de frequência de algumas <i>tags</i> sentimentais utilizados nos sistemas Citeulike, Connotea e Delicious.	7
Figura 4 - <i>Tags</i> não relacionadas em utilização com outras <i>tags</i> nos sistemas Citeulike, Connotea e Delicious.	8
Figura 5 - Interface do Delicious	10
Figura 6 – Interface original do Yahoo! MyWeb	11
Figura 7 – Interface de pesquisa do CiteULike.....	12
Figura 8 – Interface de pesquisa do Flickr.....	13
Figura 9 - Tag Cloud.....	19
Figura 10 – <i>Tag Cloud</i> circular	19
Figura 11 - Visualizações de um <i>Tag Index</i>	20
Figura 12 - <i>Tag Graph</i>	22
Figura 13 - <i>Tag Soup</i> (Delicious).....	23
Figura 14 - Topigraphy <i>Cloud</i> sobre o termo “Cooking”.	24
Figura 15 - <i>Taglines</i> (Yahoo!).	24
Figura 16 - <i>Elastic Tag Map</i>	25
Figura 17 - Sixpli.	26
Figura 18 – SpaceNav.	27
Figura 19 – Grouper	28
Figura 20 - Flickr <i>Clusters</i>	29
Figura 21 - Sistema Molecule.....	38
Figura 22 - Átomo de similaridade	43
Figura 23 - Átomo de sinonímia	43
Figura 24 - Átomo de tradução	43
Figura 25 - Átomo de termos mais utilizados	43
Figura 26 - átomo inicial.....	44
Figura 27 - Pesquisa pelo termo “ <i>references</i> ” com retorno de dois termos similares	45
Figura 28 - Átomo de termos mais utilizados para a pesquisa pela <i>tag</i> estratégia	46
Figura 29 - Página inicial do sistema Molecule	50
Figura 30 - Opções de pesquisa de resultados no Molecule.....	51
Figura 31 - Página inicial do sistema Molecule após autenticação.....	52
Figura 32 - Página de listagem dos documentos	53
Figura 33 - Página criação de novo documento.....	54
Figura 34 - Página de detalhes de um documento	56
Figura 35 - Página de eliminação de um documento.....	56
Figura 36 – <i>Deployment Diagram</i>	59
Figura 37 – Modelo de dados.....	60

Figura 38 - Primeira versão do sistema Molecule	66
---	----

Lista de Tabelas

Tabela 1 - Comparação entre plataformas	17
Tabela 2 - Comparação entre formas de visualização	31
Tabela 3 - Número de resultados retornados na plataforma Ebay	34
Tabela 4 - Operações disponíveis no serviço	62

Acrónimos e Símbolos

Lista de Acrónimos

API	Associação Ibérica de Sistemas e Tecnologias de Informação
AISTI API	Associação Ibérica de Sistemas e Tecnologias de Informação <i>Application Programing Interface</i>
CD	<i>Compact Disc</i>
CISTI	Conferência Ibérica de Sistemas e Tecnologias de Informação
DVD	<i>Digital Versatile Disc</i>
EF	<i>Entity Framework</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IIS	<i>Internet Information Services</i>
IPS	<i>Inter-process Communication</i>
JSON	<i>Javascript One Notation</i>
Linq	<i>Language-Integrated Query</i>
MP3	<i>MPEG Layer 3</i>
MPEG	<i>Moving Picture Experts Group</i>
MVC	<i>Model View Controller</i>
ORM	<i>Object-relational mapping</i>
RISTI	Revista Ibérica de Sistemas e Tecnologias de Informação
SOA	<i>Service Oriented Architecture</i>
TCP	<i>Transmission Control Protocol</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Universal Resource Locator</i>
WCF	<i>Windows Communication Foundation</i>

XALM *eXtensible Application Markup Language*

XML *eXtensible Markup Language*

1 Introdução

Neste capítulo introdutório apresenta-se o tema da tese de mestrado (secção 1.1), descrevem-se os seus objetivos (secção 1.2) e a estrutura deste documento (secção 1.3).

1.1 Contexto

No nosso dia-a-dia somos confrontados regularmente com a necessidade de rotular algo com informação adicional de forma a podermos mais facilmente recuperar ou identificar posteriormente o item em questão. Quando se marca uma página de um livro ou, por exemplo, se adiciona aos favoritos do navegador de *internet* um sítio com uma descrição para mais tarde se rever, esta-se a fazer exatamente isso. Esta técnica já antiga pode ser vista, por exemplo, nas primeiras catalogações feitas pelas bibliotecas, que utilizavam palavras-chave como títulos ou tópicos para ajudar os seus utilizadores a encontrar o que necessitam.

Assim sendo, desde há muito tempo que o ser humano usa este tipo de técnica adaptada às diversas realidades como forma de se organizar e poder recuperar mais facilmente informação.

Com o advento da *internet*, como não poderia deixar de ser, não tardou a que sítios na *web* adotassem estes mecanismos, e comesçassem a permitir aos seus utilizadores usar estas palavras-chave, designadas por *tags*, para poderem encontrar mais facilmente conteúdo. Por definição, *tags* são palavras atribuídas por utilizadores para anotar e caracterizar vários tipos de recursos na *web*. Trata-se, na prática, de rotular informação de forma simples e intuitiva, recorrendo a uma ou mais palavras-chave. Assim, pode-se dizer que são uma forma de melhorar e facilitar o acesso a recursos, pois estes já foram previamente caracterizados e “indexados” pelo utilizador, com palavras que ele próprio associa e relaciona com os itens em questão, facilitando assim futuras pesquisas e a recuperação de informação [Elhussein, 2012].

Dadas as vantagens notórias de se colocar *tags* associadas a recursos (também designado por *tagging*), este tipo de sistemas generalizaram-se e foram adotados e implementados com sucesso em diversas plataformas *web*, que são apresentadas na secção 2.2. Estes sistemas permitem organizar e partilhar diversos recursos *on-line*, como por exemplo: livros, fotografias, vídeos, publicações de blogues, músicas, entre outros [Choi, 2010]. Desta forma, os utilizadores têm acesso a um repositório de recursos rotulados, também designado de *tag space*, que pode ser pesquisado e explorado de diversas formas pela comunidade de utilizadores e resulta do conjunto de todas as *tags* de todos os utilizadores associados aos recursos do sistema em questão (Figura 1).

1. Introdução

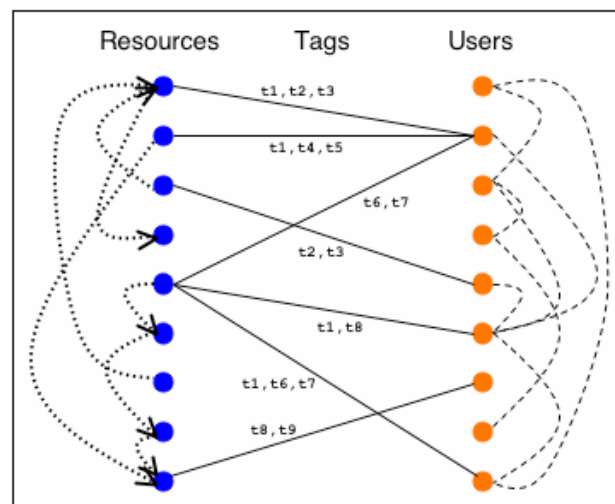


Figura 1 - Modelo de um sistema de tags

[Marlow, Naaman, Boyd, Davis, 2006]

O *tag space* pode ser também designado por folksonomia (*folksonomy*, na terminologia anglo-saxónica). Curiosamente, a criação da palavra é atribuída a Thomas Vander Wal [Pink, 2005] [Wal, 2007] sendo que para o mesmo, *folksonomy*, que deriva da junção das palavras *folk* e *taxonomy*, é apenas o ato de um utilizador adicionar uma *tag* pessoal a um determinado recurso *web*, que apenas serve para o mesmo o poder recuperar, através de proporcionar um meio para o utilizador se relacionar com os itens de uma forma que ele compreenda, não definindo assim nenhum ato colaborativo, mas sim individual [Wal, 2005].

No entanto, o conceito de taxonomia está pouco presente no de folksonomia, pois taxonomia refere-se a uma forma ontológica e hierárquica de organização de categorias, enquanto tal não existe na folksonomia, pois teoricamente todas as categorias são “iguais” [Nakamura, Endo, Morita, Morinishi, 2008] [Wal, 2007]. Neste contexto, a junção das palavras *folk* e *taxonomy* é usada para referir um conjunto orgânico, informal de terminologias relacionadas, criadas por pessoas (*folk*), que no contexto para o qual foram rotuladas adquirem um significado adicional.

Thomas Vander Wal define ainda que *social tagging* trata-se de um esforço para interligar socialmente diferentes recursos em que os utilizadores criam as *tags* com o propósito destes estarem disponíveis para outros [Wal, 2005].

Pode-se então deduzir que da junção das definições de folksonomia e de *social tagging* originais de Thomas Vander Wal nasceu o conceito mais generalizado e atual de folksonomia.

Independentemente da definição (atual ou a original) verifica-se que, quando um utilizador observa como uma comunidade rotulou um recurso, a probabilidade desse utilizador adotar *tags* com vocabulário semelhante para recursos idênticos é elevada. Assim, neste contexto, folksonomia define-se como um vocabulário emergente entre a comunidade de utilizadores que partilham interesses semelhantes [Cattuto, Loreto, Pietronero, 2006].

Assim, surgiram várias formas de visualização desta informação, destacando-se, de entre todas, a nuvem de *tags* como a forma mais comumente utilizada. Esta caracteriza-se por numa nuvem constituída por várias palavras e que normalmente variam em cor e tamanho, de acordo com a sua utilização.

Não obstante a proliferação deste tipo de sistemas e desta forma de visualização, existem ainda várias limitações existentes à sua utilização e torna-se, assim, imperativo recorrer a alternativas que permitam ultrapassar estes problemas para potenciar ainda mais a sua utilização.

1.2 Objetivos

Este trabalho teve como objetivos principais:

- Estudar as formas de visualização de *tags* existentes e em utilização em vários sistemas diferentes, compreendê-las e contextualizá-las, entendendo as suas vantagens e caracterizando as suas desvantagens;
- Propor uma nova forma de visualização de *tags* que introduza melhoramentos às já existentes e desenvolver um protótipo.

Assim, pretendeu-se compreender e conhecer melhor esta temática, obtendo-se uma visão clara sobre quais as limitações decorrentes destes tipos de visualização e de sistemas. Num segundo momento, pretendeu-se o desenvolvimento de abordagens e soluções às limitações identificadas anteriormente, conjugadas de forma a criar uma nova forma de visualização.

1.3 Estrutura da dissertação

Esta dissertação encontra-se organizada em quatro capítulos. Neste primeiro capítulo foi apresentada uma pequena abordagem à temática das *tags* e da sua utilização e foram ainda referidos os objetivos e desafios deste trabalho.

No segundo capítulo é apresentado um estudo mais aprofundado sobre *tags*. É feita uma introdução inicial ao tema, abordando-se vários conceitos sobre o mesmo. Descreve-se seguidamente diversos sistemas que implementam este tipo de solução e ainda várias formas de visualização de *tags*. Por fim, são ainda apresentadas várias limitações conhecidas a estes sistemas, centrando-se o estudo nas nuvens de *tags*.

No terceiro capítulo é apresentado e descrito o sistema Molecule que pretende colmatar as várias restrições apresentadas anteriormente e implementar melhorias às atuais nuvens de *tags* e, assim, atingir o objetivo principal deste trabalho.

1. Introdução

No quarto capítulo é feita uma conclusão sobre o trabalho realizado, de acordo com os objetivos inicialmente propostos. São também referidos vários problemas identificados, assim como trabalho futuro a ser realizado.

Por fim estão incluídos dois anexos, os quais contêm trabalhos científicos elaborados no decorrer do desenvolvimento do sistema Molecule. No primeiro anexo podemos encontrar um paper submetido para a conferência CISTI enquanto que no segundo anexo podemos ver um paper submetido e publicado na revista RISTI.

2 Tags

Neste capítulo são descritos os diversos tipos diferentes de *tags*, seguindo-se a apresentação de vários sistemas que utilizam o conceito de *tag*, assim como a caracterização de diversas formas de visualização.

Por fim, são enumeradas várias limitações levantadas, decorrentes do estudo efetuado sobre os sistemas e sobre as formas de visualização de *tags*.

2.1 Tipos de tags

Tags são apenas palavras atribuídas a recursos, mas podem ser categorizadas e tipificadas, de acordo com a forma como o utilizador se relaciona com o recurso em questão.

Estudos anteriores de plataformas como Delicious e Citeulike apuraram que, embora muitas das rotulações feitas pelos utilizadores estejam intrinsecamente ligadas aos itens a que são referentes, e formem de facto uma descrição precisa o suficiente a nível de indexação, muitas outras não estão diretamente relacionadas com o item, mas são do tipo afetivas e resultantes da resposta emocional do utilizador ao item, ou então são do tipo temporal ou tarefa [Kipp, 2007].

As *tags* existentes nas plataformas, podem ser divididas em dois grupos: *tags* relacionados com características objectivas dos itens, como assunto, autor, data de criação, e as *tags* relacionadas com outros aspetos. Na Figura 2 são visíveis várias *tags* utilizadas nas plataformas Delicious, Citeulike e Connotea (descritas brevemente na secção 2.2), ficando patente a discrepância de utilização de *tags* entre plataformas, sendo facilmente visível a utilização de *tags* que descrevem os itens de forma mais subjetiva além de outros tipos de *tags*.

2. Tags

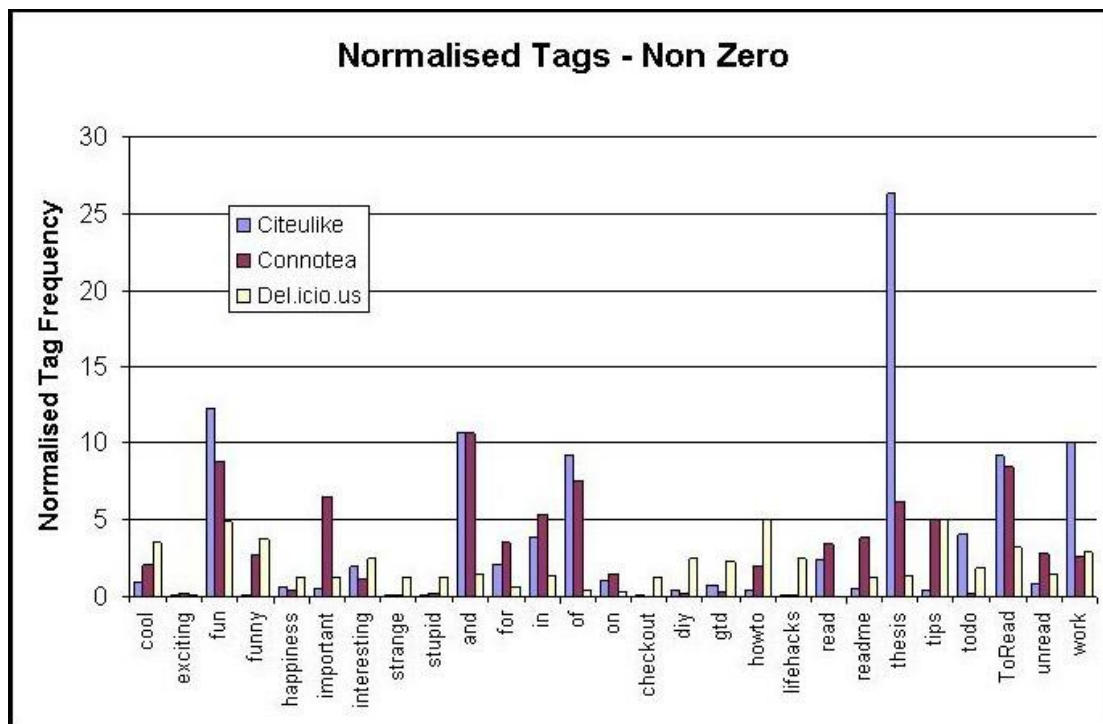


Figura 2 - Comparação de frequência de algumas *tags* utilizadas nos sistemas Citeulike, Connotea e Delicious.

[Kipp, 2007]

2.1.1 Tags objetivas

Tags objetivas descrevem, geralmente, o assunto do item em questão mas também podem estar relacionadas com outros aspetos, como o seu autor, um local ou data de criação. Para uma determinada publicação educacional, *tags* como *physics* e *mathematics*, podem ser utilizadas para descrever o conteúdo ou características da publicação, segundo a opinião dos seus utilizadores.

É a elevada utilização de termos dos títulos e assuntos como *tag*. Verifica-se muitas vezes também o uso de acrónimos e siglas relacionadas com termos técnicos ou académicos.

2.1.2 Tags subjetivas

Tags subjetivas descrevem outras características do item, sendo resultantes de uma qualquer reação que o item causou no utilizador e dessa sua experiência muito particular.

2.1.2.1 Tags relacionadas com tempo e tarefa

Tags do tipo tarefa podem ser do tipo *"toread"*, *"todo"*, *"tobuy"*, entre outras, e nascem da necessidade e desejo de combinar informação sobre tarefas e atividades na classificação dos

termos. A utilização deste tipo de tags varia muito de plataforma para plataforma, estando relacionada com a natureza da mesma. Por exemplo, a tag “toread” (bastante utilizados no Delicious) é pouco usado na plataforma Citeulike (cerca de 1000 artigos em 10 de Outubro de 2013¹). Este tipo de tags pode ser útil num sistema de recomendações como, por exemplo, o da Amazon (“toread” - 503 produtos em 10 de Outubro de 2013), uma vez que demonstram o interesse de outros utilizadores naquele item em questão [Kipp, 2007].

Existem ainda tags relacionadas com tempo, sendo estas usadas para caracterizar o item de forma temporal, mas sem caracterizar nenhum pormenor do item em si. Assim, por exemplo, tags como Agosto, aniversário ou Natal, são consideradas tags relacionadas com tempo.

2.1.2.2 Tags Sentimentais

As tags sentimentais traduzem uma opinião sobre o recurso. Na Figura 3 pode-se observar a utilização de algumas dessas tags e a comparação da sua frequência nas plataformas Citeulike, Connotea e Delicious.

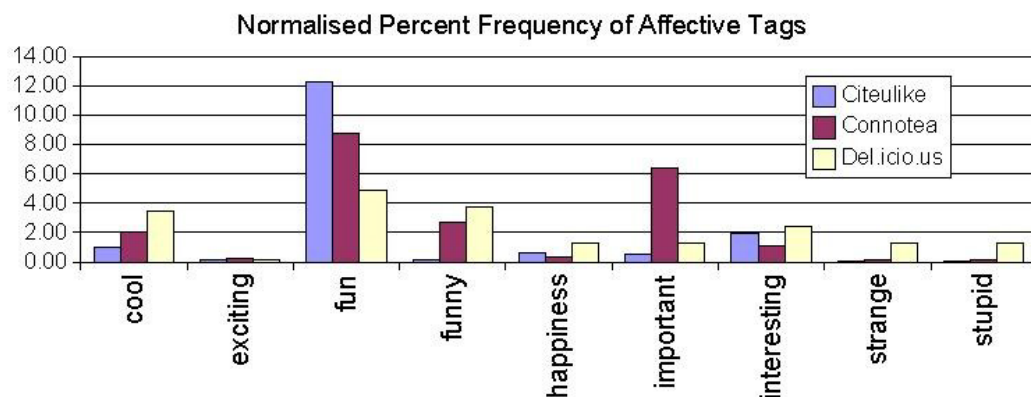


Figura 3 - Comparação de frequência de algumas tags sentimentais utilizados nos sistemas Citeulike, Connotea e Delicious.

[Kipp, 2007]

Tags como “cool” ou “fun”, por exemplo, aparentemente não são úteis para a classificação do recurso. No entanto, foi sugerido que estas tags indicavam a reação emocional do utilizador ao recurso [Kipp, 2007]. Estes termos são obviamente subjetivos e, desde sempre, foram excluídos de qualquer tipo de sistema de classificação, por essa mesma razão. No entanto, o uso continuado destes termos em sistemas de “bookmarking” social, indica que eles possuem significado para os utilizadores.

¹ <http://www.citeulike.org/tag/toread/page/20>

2. Tags

É também usada a combinação de *tags* relacionadas com o assunto dos recursos com *tags* subjetivas (Figura 4), possivelmente para se obter uma maior expressividade na caracterização dos itens.

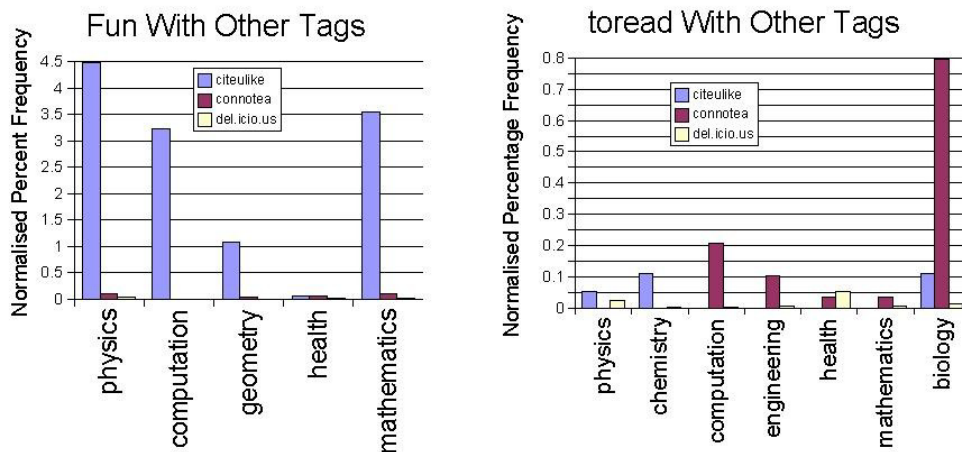


Figura 4 - Tags não relacionadas em utilização com outras tags nos sistemas Citeulike, Connotea e Delicious.

[Kipp, 2007]

2.1.2.3 Outros tipos de tags

Existem ainda uma infinidade de outros tipos de *tags* não relacionadas como, por exemplo, *tags* sobre a performance, auto preservação, ativismo, entre outras [Zollers, 2007]. Por exemplo, a tag “defectivebydesign” é muito popular no sistema Amazon.com², sendo utilizada como uma forma de ativismo.

2.2 Sistemas que utilizam tags

Existem inúmeras plataformas que usam o recurso a *tags* como forma de organização e pesquisa. A natureza dessas plataformas pode ser distinta e terem temas diversos associados às mesmas, mas uma coisa é comum entre todas: a necessidade de catalogar e rotular informação. Nesta secção serão descritas algumas das plataformas que recorrem a *tags*, tendo para isso sido selecionadas plataformas que possuem formas de navegação do seu *tag space* e demonstrada a diversidade de tipos de plataformas que as utilizam.

² Ver http://www.amazon.com/tag/defectivebydesign/products/ref=tag_dh_istp.

2.2.1 Delicious

Esta plataforma é um serviço *web* de *bookmarking* (<http://delicious.com>) gratuito para armazenar, partilhar e descobrir *bookmarks* na *web*. O sítio *web* foi fundado por Joshua Schachter em 2003 e adquirido pela Yahoo! em 2005. No final de 2008, estimava-se que a plataforma teria mais de 5.3 milhões de utilizadores e 180 milhões de *URL's* diferentes marcados. A plataforma foi adquirida pela AVOS Systems em Abril de 2011 e é atualmente detida pela mesma.

O Delicious usa uma classificação não hierárquica em que os utilizadores podem rotular todos os seus *bookmarks*. Está disponível uma vista combinada dos *bookmarks* de todos os utilizadores para um determinado recurso. Assim, por exemplo, o *URL* <http://delicious.com/tag/wiki> mostra todos os *links* rotulados recentemente com a palavra “wiki”. A sua natureza colaborativa faz com que seja possível ver *bookmarks* de outros utilizadores.

O Delicious também permite aos utilizadores agrupar *links* com tópicos semelhantes em pilhas, e incluir um título e descrição a essa pilha. As pilhas podem ser trabalhadas de forma colaborativa com outros utilizadores e podem ser seguidas e partilhadas para outros utilizadores. Este permite também importar *bookmarks* dos *browsers* para o seu sítio, de forma a facilitar a integração de novos utilizadores.

Todos os *bookmarks* postado no Delicious são públicos por defeito, mas *bookmarks* específicos podem ser marcados como privados pelos utilizadores e todas as importações são privadas por defeito [delicious, 2003].

Para se marcar algo basta estar registado, sendo que registos de plataformas externas como o Facebook podem ser usados para criar uma conta. No entanto, se a intenção for apenas pesquisar não há qualquer necessidade de registo.

A interface do Delicious é simples e, após pesquisa, mostra os recursos relacionados com a mesma, assim como todas as *tags* relacionadas, e as *tags* relacionadas com cada recurso (Figura 5).

Muitas formas de visualização de *tags* podem ser usadas adicionalmente, dado que nos é permitido, como já referido, fazer *download* de informação. Exemplo dessas formas de visualização são por exemplo os sistemas SpaceNav (2.3.2.9) e Grouper (2.3.2.10).

2. Tags

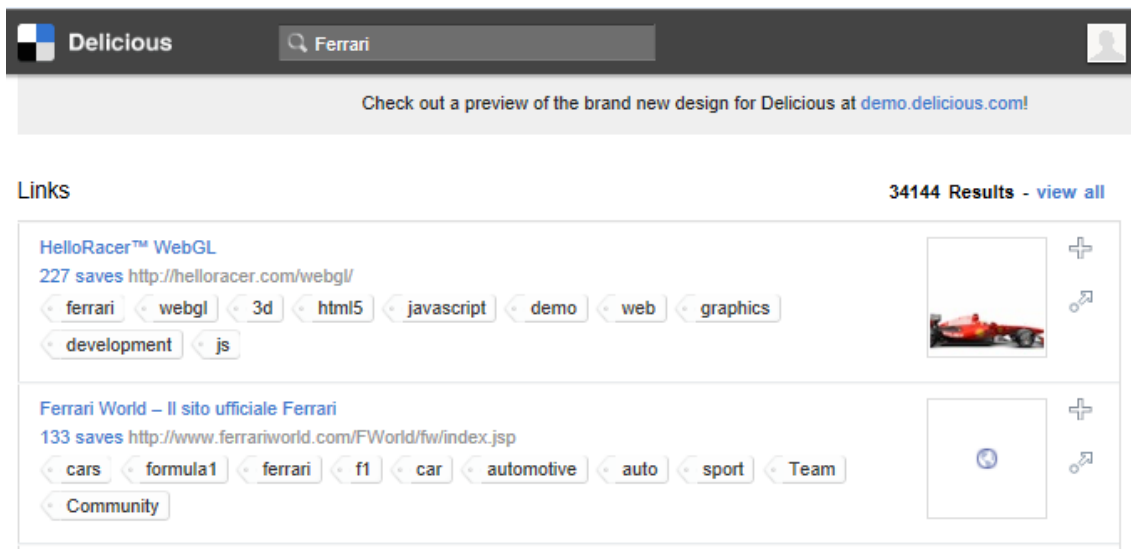


Figura 5 - Interface do Delicious

2.2.2 Yahoo! Bookmark

O serviço MyWeb (<http://myweb.yahoo.com>), nasceu em Junho de 2005 e foi criado pela Yahoo! como um serviço de *bookmarking* (Figura 6).

Através deste, era possível aos utilizadores rotularem, desde páginas web, ao próprio utilizador, assim como para contatos selecionados ou todos os utilizadores.

O sistema permitia também a pesquisa por *tags* com que outros utilizadores tinham marcado os seus sítios *web* (*bookmarked*). Era também permitido adicionar contatos à sua conta MyWeb, que permitia que se pudesse visualizar as *tags* dos outros utilizadores.

Em Fevereiro de 2009, foram anunciados planos para encerrar o MyWeb em 18 de Março de 2009. Os *bookmarks* MyWeb ficaram automaticamente disponíveis no Yahoo! Bookmarks [Cashmore, 2009].

A plataforma que nasceu como um sistema de pesquisas revolucionário e que iria permitir aos utilizadores participarem ativamente na forma como o motor de busca encontrava a informação foi, assim, encerrada, ficando apenas esta anotação do mesmo pela inovação que esta ideia poderia ter trazido [Battelle, 2005].

O sistema Yahoo! Bookmarks é um sistema gratuito e simples de *bookmarking* que permite após registo ou login através da plataforma Facebook ou Google fazer *bookmarks* privados. Permite também, tal como o Delicious, importar ou exportar as *tags*, sendo possível depois utilizar inúmeras ferramentas para se obter visualizações diferentes dos mesmos dados.

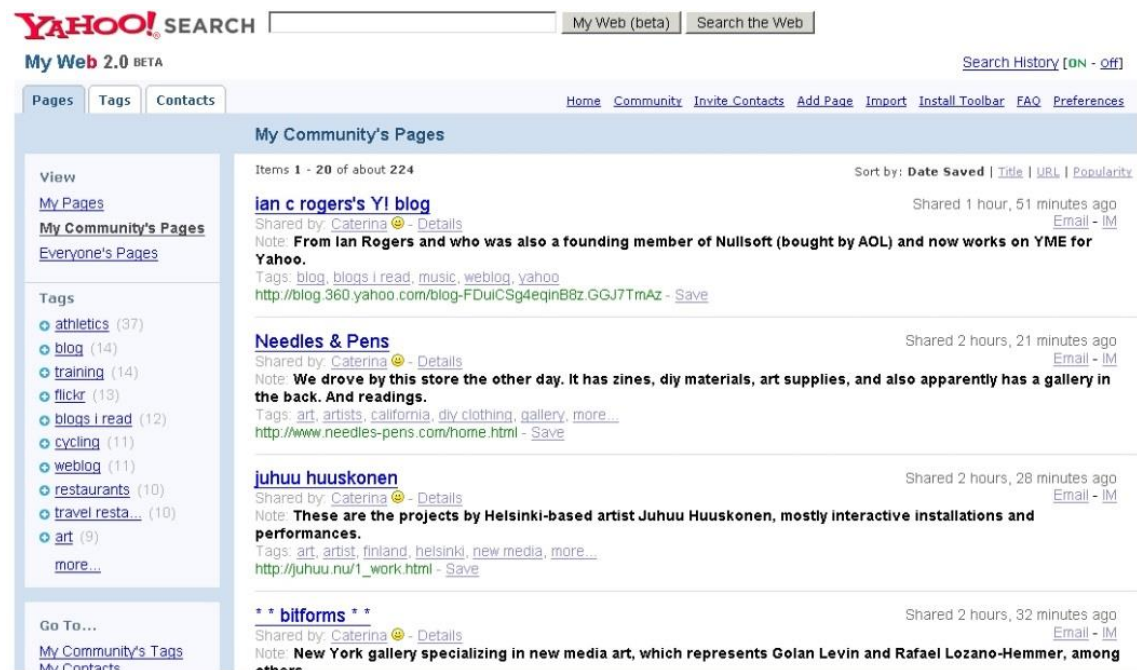


Figura 6 – Interface original do Yahoo! MyWeb

2.2.3 CiteULike

CiteULike (<http://www.citeulike.org/>) é uma plataforma de *bookmarking* social e tem como objetivo promover e desenvolver a partilha científica de referências para investigadores. Da mesma forma que é possível rotular sítios *web* (com o Delicious, por exemplo), ou fotografias (com o Flickr, por exemplo), os cientistas e investigadores podem partilhar informação sobre documentos científicos e académicos com esta ferramenta. O CiteULike foi criado por Richard Cameron em Novembro de 2004 e, em 2006, foi estabelecido que a Oversity Ltd. iria continuar o desenvolvimento e suporte [Citeulike, 2013].

Esta plataforma é completamente gratuita, fazendo já parte dela 260.866 utilizadores, sendo apenas necessário um registo ou login através da plataforma Facebook, para fazer *bookmarks* privados e públicos [Citeulike, 2009].

Os mesmos podem ser pesquisados tanto pelas *tags*, como por outras informações (título, autor, etc), sejam esses do próprio utilizador, ou públicos de outros utilizadores da comunidade. Não é necessário, no entanto, ter-se uma conta na plataforma para se poder pesquisar, uma vez que as pesquisas são públicas.

Na Figura 7 é visível uma pesquisa pelo termo “*folksonomy*” onde é possível ver parte dos resultados da mesma. Nesta plataforma, são apresentadas também em forma de *tag cloud* as próprias *tags* do utilizador como forma de pesquisa alternativa, nos seus *bookmarks*.

2. Tags



Figura 7 – Interface de pesquisa do CiteULike

2.2.4 Flickr

Flickr (<http://www.flickr.com>) é um sítio *web* que permite o armazenamento de imagens e vídeo. Este foi criado por Ludicorp em 2004 e adquirido pela Yahoo! em 2005. Para além de ser um dos sítios *web* mais conhecidos para a partilha e armazenamento de fotografia a nível pessoal, é também usado como ferramenta de armazenamento, por blogues, e não só. Em 2011, a Yahoo! anunciou que a comunidade total de utilizadores tinha atingido um total de 51 milhões. Em Agosto do mesmo ano, foi anunciado que o sítio *web* possuía mais de 6 biliões de fotografias [Parfeni, 2011].

Fotografias e vídeos podem ser acedidos no Flickr sem necessidade de registo, mas para se poder fazer *upload* e rotular conteúdos já é obrigatório. O registo pode ser feito também através do login na plataforma Facebook e permite também ao utilizador criar um perfil, e criar uma rede de perfis de utilizadores e contactos. O Flickr pode ser gratuito ou pago, dependendo da conta de utilizador que se escolher, sendo que pago permite mais regalias, entre elas *downloads* ilimitados e ausência de publicidade.

A atribuição de *tags*, dado estar associada à fotografia, pode ser feita tanto em privado como público, ou apenas para a rede do utilizador [Flickr, 2005].

A apresentação das *tags* (Figura 8) associadas a um recurso é feita em forma de lista e é possível pesquisar por estes, assim como nome do recurso. Não obstante, existem ferramentas que

permitem visualizações alternativas como, por exemplo: Flickr Related Tag Browser (2.3.2.14), Tag's Life (2.3.2.13), entre outras.

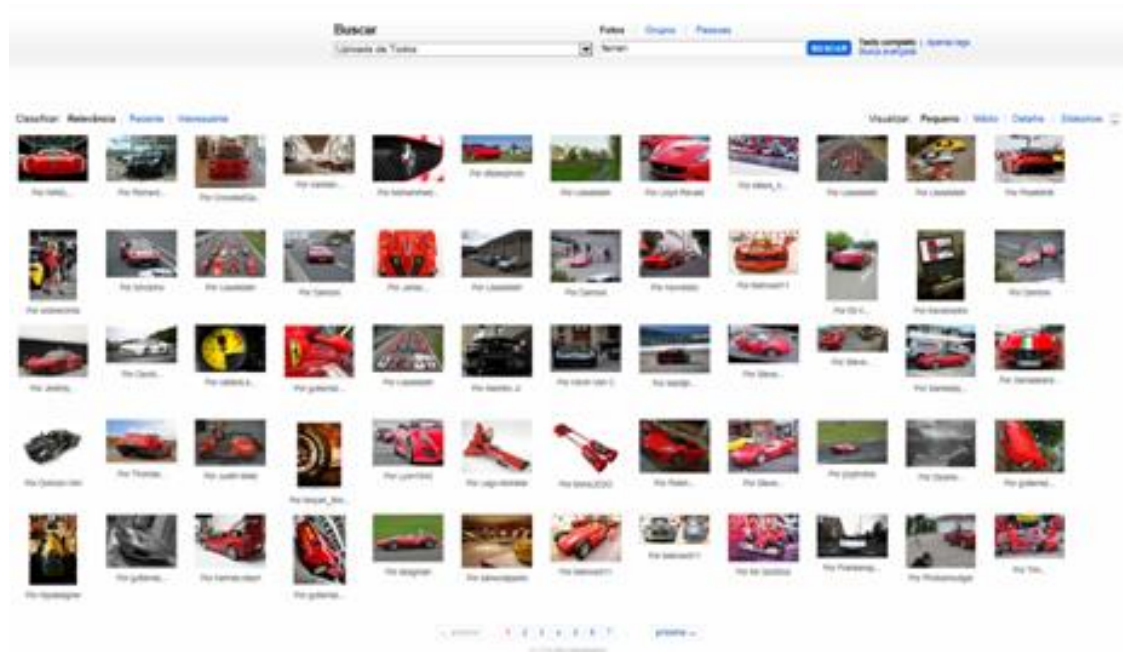


Figura 8 – Interface de pesquisa do Flickr

2.2.5 Last.fm

Last.fm (<http://www.last.fm>) é um sítio *web* sobre música, criado em 2002 no Reino Unido por Felix Miller, Martin Stiksel e Richard Jones. Em Março de 2009 já tinha mais de 30 milhões de utilizadores ativos [Jones, 2009]. Em 30 de Maio de 2007, a CBS Interactive comprou a plataforma por 140 milhões de libras.

Usando um sistema de recomendação de música chamado “Audoscrobber”, a Last.fm constrói um perfil detalhado sobre o gosto musical de cada utilizador ao guardar detalhes sobre as músicas que o utilizador ouve, seja através de rádios na Internet, ou músicas do computador ou dispositivos portáteis de música. Esta informação é transferida para a base de dados da plataforma, tanto diretamente pelo leitor de música (Rdio, Spotify, Clementine, Amarok), como por via de um *plugin* instalado no mesmo. A informação é depois mostrada ao utilizador no seu perfil e compilada para criar páginas referenciadas para cada artista. Em Abril de 2011 foram reportados mais de 50 bilhões de entradas na base de dados [thenextweb, 2011].

Este serviço é gratuito para utilizadores no Reino Unido, Estados Unidos e Alemanha. Noutros países requer uma subscrição para usar o serviço de rádio (3 euros por mês após 50 faixas gratuitas). O sítio oferece ainda vários outros serviços (venda de álbuns, informações de concertos, entre outras) e faz reprodução de artistas semelhantes aos favoritos do utilizador.

2. Tags

É possível efetuar pesquisas sem uma conta e inclusive ouvir-se algumas músicas resultantes dessa pesquisa, mas outras ficarão bloqueadas para reprodução.

A Last.fm suporta também *tags* (públicas) atribuídas pelos utilizadores (mas apenas para utilizadores registados) para os artistas, álbuns e músicas. Os utilizadores podem fazer pesquisas pelas *tags*, mas o benefício mais importante é permitir aos utilizadores reproduzir músicas que estão rotuladas de certa forma. A *tag* pode ser o tipo de música, sentimento, característica artística, ou outra qualquer que defina uma forma de classificação do utilizador. No entanto, como a rotulação não é controlada, esse facto propicia a manipulação dos utilizadores, não havendo consenso quanto à atribuição da *tag*.

As *tags* são apresentadas com os recursos aquando da pesquisa, na forma de uma lista simples.

2.2.6 Technorati

Technorati (<http://www.technorati.com/>) é um motor de busca para blogues. Em Junho de 2008, a plataforma já tinha indexado mais de 1.128 milhões de blogues e mais de 250 milhões de artigos da comunicação social [Technorati, 2008]. O nome Technorati advém da mistura das palavras *technology* e *literati*, que têm a denotação de inteligência tecnológica ou intelectual.

A plataforma foi fundada por Dave Sifry e tem a sua sede nos Estados Unidos, mais precisamente na Califórnia, em São Francisco.

O Technorati aproveita as *tags* que os autores colocaram nas suas próprias publicações e sítios *web*. As *tags* mais recentes podem ainda ser listados sobe a forma de *tag index* na plataforma, para pesquisa pelas mesmas [Technorati, 2002], ou em forma de tag cloud na página inicial.

A plataforma atribui uma nota a cada blogue, através do número de blogues que possuem ligações para o blogue.

A plataforma disponibiliza também aos seus utilizadores a possibilidade de verem num *tag index* as tags utilizadas no último mês.

2.2.7 Upcoming

Upcoming (<http://upcoming.org/>) é um calendário de eventos sociais criado em 2003 por Andy Baio. A 4 de Outubro de 2005 este foi comprado pela Yahoo! [Kawamoto, 2005].

A plataforma combina funcionalidades de um calendário de eventos e uma rede social. O sítio *web* é um motor de busca de eventos que irão acontecer, como concertos de música, exposições de arte, conferências empresariais, entre outros. A informação dos eventos é criada pela comunidade de utilizadores e os utilizadores podem indicar os seus planos, referindo se vão ou se têm interesse no evento, assim como estabelecer amizades com outros utilizadores

e receber notificações (via email ou mensagem) sobre os planos dos amigos e eventos prestes a acontecer [info.yahoo, 1995].

Upcoming usa uma conta Yahoo! (gratuita) para ser acedido pelos seus utilizadores e assim poderem adicionar os seus eventos. Os eventos podem conter *tags* (públicas) para melhor pesquisa e organização.

Não é necessário, no entanto, ser-se detentor de uma conta para se poder pesquisar sobre os eventos da plataforma, bastando para isso fazer a pesquisa através do campo de texto disponível em todas as páginas sem necessidade de autenticação.

2.2.8 Amazon

Amazon.com (<http://www.amazon.com/>) é uma multinacional de comércio eletrónico com sede em Seattle nos Estados Unidos da América. Trata-se do maior revendedor *online* no mundo, mas também produz equipamentos eletrónicos próprios, como o “Amazon Kindle” e “Kindle Fire *tablet computer*”, e é também um dos maiores fornecedores de serviços de “*cloud computing*”.

A empresa foi criada por Jeff Bezos, em Julho de 1994, com o nome de Cadabra. Em 1995, a plataforma foi *online* sob o nome de amazon.com. A plataforma que começou como uma vendedora de livros *online* rapidamente se diversificou e hoje vende desde livros a DVDs, CDs, MP3 *downloads*, *softwares*, jogos de vídeo, equipamentos eletrónicos, móveis, comida, brinquedos, joias, entre outros.

Em 2010, a Amazon já contava com mais de 121 milhões de utilizadores [businessweek, 2010].

O registo no sítio *web* é gratuito e necessário para a aquisição e venda de bens. Embora atualmente não exista a possibilidade de adição de *tags* aos produtos, anteriormente era possível. Por defeito, essa atribuição era tida como pública, mas era possível fazê-la como privada [Amazon, 1996].

Aos utilizadores sem registo apenas é possível pesquisar produtos e ver seus comentários, sendo necessário sempre uma autenticação para outras operações como comentar, comprar, vender ou rotular, por exemplo.

É possível navegar através de uma *tag cloud* nativa da plataforma, como forma alternativa à pesquisa direta por *tags*.

2.2.9 Comparação entre plataformas

Na Tabela 1 pode-se observar as várias plataformas anteriormente descritas e comparar alguns aspetos das mesmas, resumindo assim algumas das suas características para uma melhor visão geral sobre as mesmas.

2. Tags

De notar a coluna “*Ranking* Alexa”, que contém a posição no *ranking* Alexa das várias plataformas (<http://www.alexa.com/>). O *ranking* qualifica os sítios *web* mediante o número de utilizadores e visitas médias por página nos últimos 3 meses. Assim, quanto mais perto do número 1, maior é o número de visitantes do sítio, teoricamente. Os dados são recolhidos através da colaboração dos utilizadores, pelo que os resultados do *ranking* não são exatos, mas apenas uma ideia aproximada da realidade. Desta forma, incluiu-se a coluna para dar apenas uma expressividade ao movimento que as plataformas poderão ter a nível de visitas.

Tabela 1 - Comparação entre plataformas

Plataforma	Endereço Web	Tipo de Sistema	Data de criação	Grátis	Tags públicas	Tags privados	Ranking Alexa (Nov 2012)
Delicious	http://delicious.com	Bookmarking	2003	Sim	Sim	Sim	528
Yahoo! Bookmark	bookmarks.yahoo.com	Bookmarking	2005	Sim	Não	Sim	-
CiteULike	http://www.citeulike.org/	Bookmarking	2004	Sim	Sim	Sim	12.080
Flickr	http://www.flickr.com	Partilha de fotografia	2004	Opcional	Sim	Sim	57
Last.fm	http://www.last.fm	Partilha e venda de produtos e serviços relacionados com música	2002	Opcional	Sim	Não	935
Technorati	http://www.technorati.com/	Motor de busca de blogues	2007	Sim	Sim	Não	1.634
Upcoming	http://upcoming.org/	Calendário de eventos sociais	2003	Sim	Sim	Não	-
Amazon	http://www.amazon.com/	Venda e revenda de artigos e serviços.	1994	Sim	Sim	Sim	10

2.3 Formas de visualização de tags

Para possibilitar a navegação e pesquisa do *tag space*, as plataformas que proporcionam este tipo de serviço necessitam de modelar uma interface para a sua visualização pela sua comunidade de utilizadores. A visualização de informação envolve a utilização de representações visuais abstratas para melhorar a cognição [Ware, 2004]. Interfaces visuais eficazes permitem a interação com grandes volumes de dados de forma rápida e eficaz, potenciando a descoberta de características que podem não ser tão evidentes numa apresentação tabular, por exemplo.

Assim, começaram a surgir formas de visualização de *tags* que permitem aos utilizadores rapidamente verificar como as *tags* são utilizadas numa determinada plataforma. Em algumas o número de *tags* é elevado, o que inviabiliza a apresentação de todas as *tags* que foram atribuídas por todos os utilizadores, mas apenas são visualizadas as mais relevantes segundo determinados critérios.

Nas restantes subsecções deste capítulo, discutem-se diferentes formas de visualização, desde a mais conhecida *Tag cloud* até outras menos habituais que, dependendo do contexto ou utilizador, podem ser mais apropriadas.

Independentemente da forma de visualização, um aspeto comum a todas as que são discutidas nesta parte da dissertação é que, ao contrário da pesquisa por *query* (com recurso a uma caixa de texto), que requer que o utilizador formule as suas necessidades de informação, o tipo de pesquisa por *tag*, implícito nas formas de visualização de *tags*, permite que o utilizador reconheça a informação, não tendo assim que formular a pesquisa para poder ver os primeiros resultados [Choi, 2010].

2.3.1 Tag Clouds

Tag cloud não é mais que uma lista das *tags* mais populares, normalmente organizadas por ordem alfabética (Figura 9) e com diversos tamanhos de letra e cores, de forma a fazer sobressair os termos mais frequentemente utilizados. Esta forma de visualização é bastante simples e largamente usada em inúmeras plataformas *on-line*.

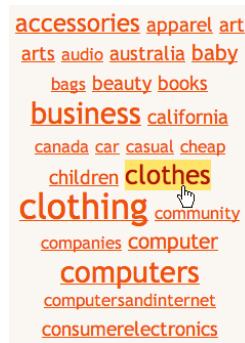


Figura 9 - Tag Cloud

[Friedman, 2007]

Também é habitual a utilização de *tags* que não são ordenadas de forma alfabética e todos os termos têm o mesmo tamanho e “peso”. Os termos mais importantes são sobressaídos com uma mudança da cor da letra ou da cor de fundo do termo, estando as *tags* ordenadas por importância ou frequência.

Outra variação comum é o uso de uma mistura de tamanho de letra com “peso” e cor nos termos, não havendo qualquer ordenação nas *tags*.

Por vezes, numa nuvem de *tags*, os termos semelhantes aparecem na vizinhança de outros, sendo aplicadas várias variações da formatação visual.

As *tags* também podem ser dispostas de forma circular com a posição e o tamanho da letra a variar do centro para a “órbita”, por relevância (Figura 10).



Figura 10 – *Taq Cloud* circular

[Werts, 2009]

2. Tags

Por fim, note-se que existem ainda *tag clouds* que combinam várias das características anteriormente descritas.

2.3.2 Outras formas de visualização

Existem, também, formas de visualização alternativas às *tag clouds*, sendo várias apresentadas nesta secção.

2.3.2.1 Tag Index

Num *Tag Index* todas as *tags* aparecem indexadas alfabeticamente, como se pode ver na Figura 11.

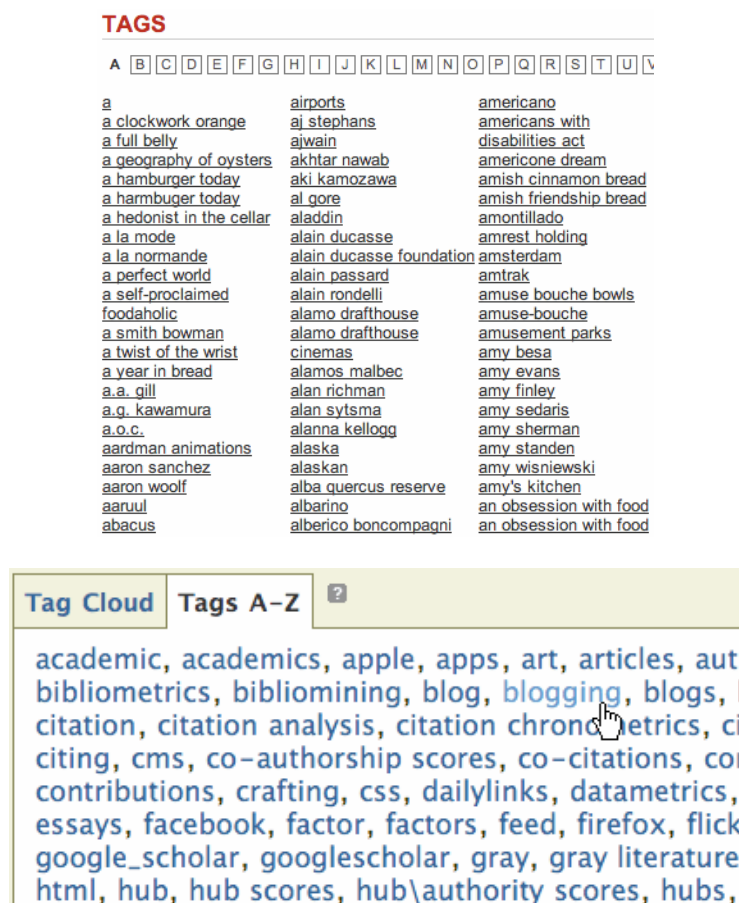


Figura 11 - Visualizações de um *Tag Index*

[Friedman, 2007]

2.3.2.2 *Tag Graph*

Tag Graph mostra os termos e as suas ligações, assumidamente em forma de grafo, mostrando as “relações” de coocorrência entre os mesmos. Na Figura 12 pode-se ver o exemplo do *Tag Graph*, estabelecendo uma relação entre vários termos, tanto com a *tag* “*Folksonomy*” como com a *tag* “*Ontology*”, e relacionando ambos através da *tag* “*metadata*”, estabelecendo assim uma ligação entre as *tags* “*Ontology*” e “*Folsonomy*”.

2. Tags

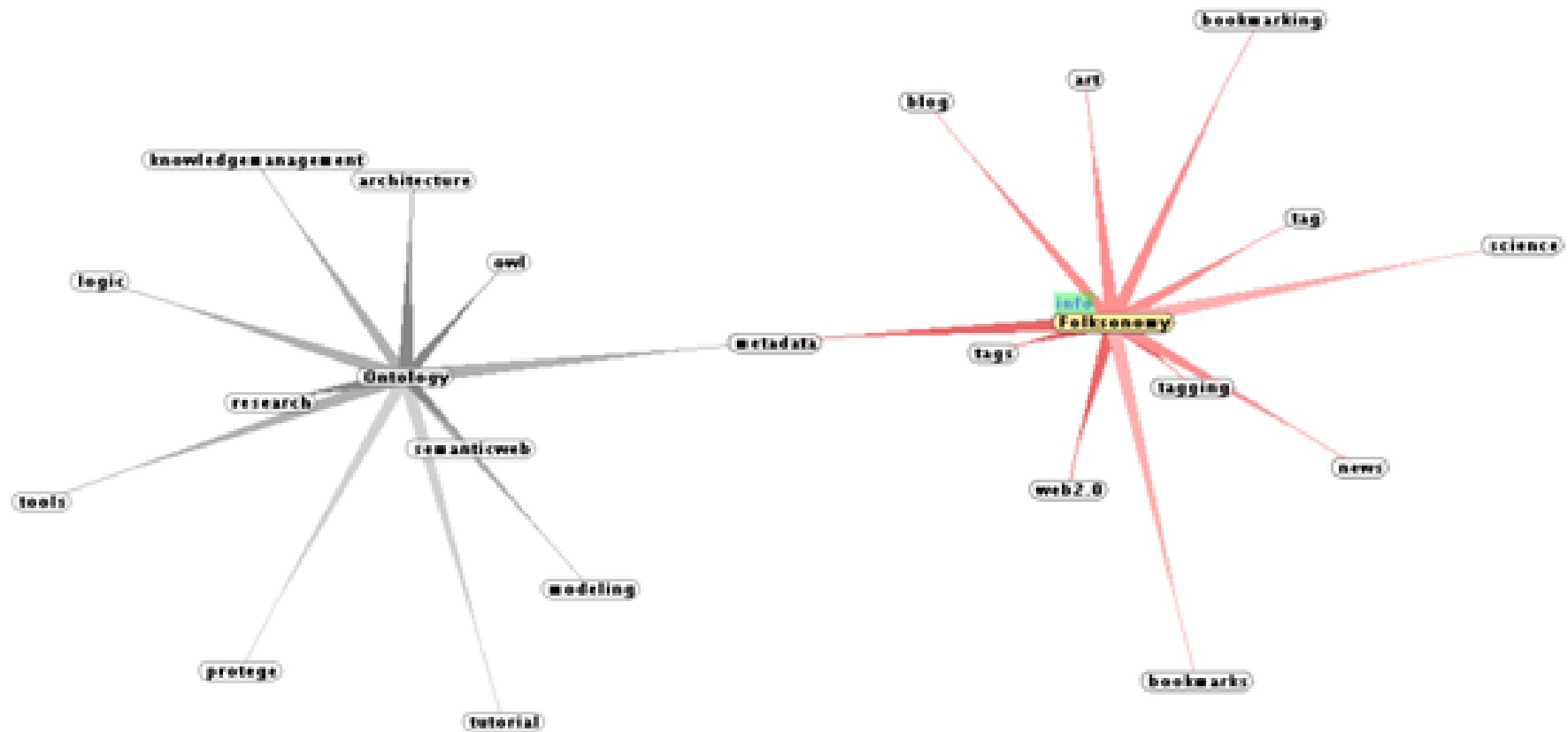


Figura 12 - Tag Graph

[Albrecht, 2006]

2.3.2.3 Tag Soup

Tag Soup é outra forma distinta de visualização de *tags* em que, de uma forma diferente do *Tag Graph*, é também mostrada a relação de coocorrência entre os termos, mas em forma circular e ao estilo de uma *tag cloud* circular (Figura 13).

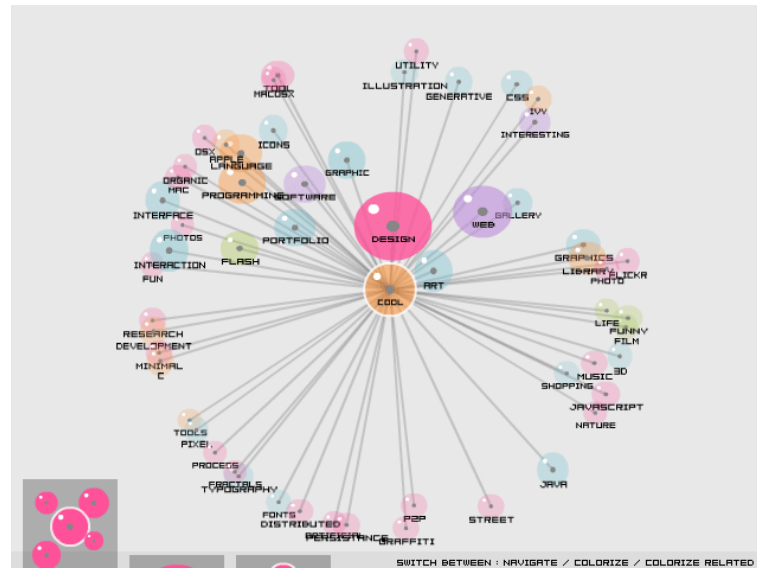


Figura 13 - *Tag Soup* (Delicious)

[Zitvogel, 2005]

2.3.2.4 Topigraphy Cloud

Topigraphy (*topic* + *topography*) nasce da necessidade da identificação de uma *tag* específica quando, no meio de uma amostragem muito grande de *tags*, especialmente se utilizada a ordenação alfabética típica, as relações entre *tags* não são visíveis. Assim, é usada uma imagem topográfica como fundo, sobre a qual os termos são mostrados. A "altura" representa a centralidade do conceito de termos relacionadas, enquanto o *layout* a duas dimensões aborda semelhança dos termos.

Na Figura 14 é visível uma *Topigraphy Cloud* sobre o termo "Cooking".

2. Tags

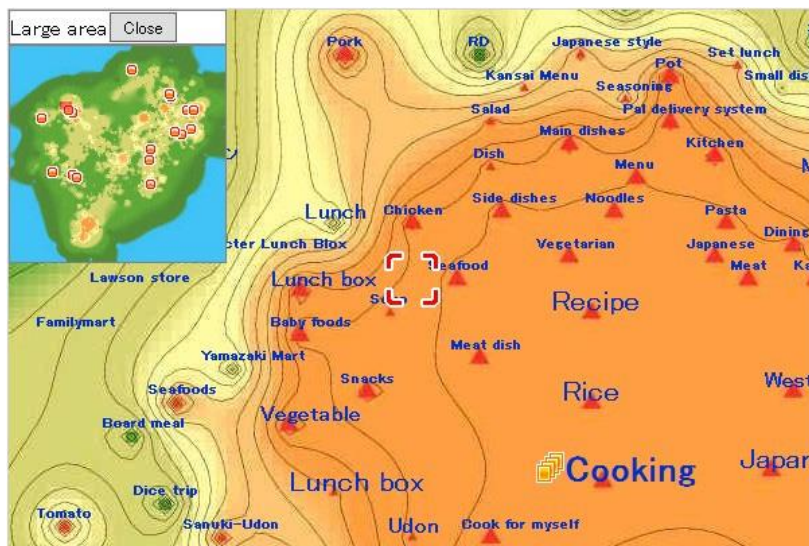


Figura 14 - Topigraphy *Cloud* sobre o termo “Cooking”.

[Fujimura, Fujimura, Matsubayashi, Yamada, Okuda, 2008]

2.3.2.5 Taglines

Tagline é um método de visualização de *tags* desenvolvido pela Yahoo! para representar *tags* com miniaturas a “flutuar”, ao longo de um eixo primário temporal, para a plataforma Flickr.

Esta forma de modelação gráfica de *tags*, visível na Figura 15, mostra as 8 *tags* mais populares em cada dia na plataforma Flickr sendo que, entre outros aspetos, permite clicar e segurar nas *tags* ou miniaturas e, claro, mudar a velocidade com que estas se movem pelo “tempo”, assim como listá-las.



Figura 15 - *Taglines* (Yahoo!).

[Don Turnbull, 2006]

2.3.2.6 Elastic Tag Map

Nesta modelação gráfica são visíveis as relações emergentes entre *tags*. A rotina que produz o *layout*, insere as *tags* que são frequentemente usadas juntas, mais perto num plano 2D. Quando o utilizador passa o rato por cima de uma *tag*, as *tags* que costumam também ocorrer com a *tag* em questão são trazidas para a frente. Clicar permite fixar essa *tag* e explorar melhor as relações seleccionadas (Figura 16). A este tipo de técnica dá-se o nome de “*Pivot browsing*”.

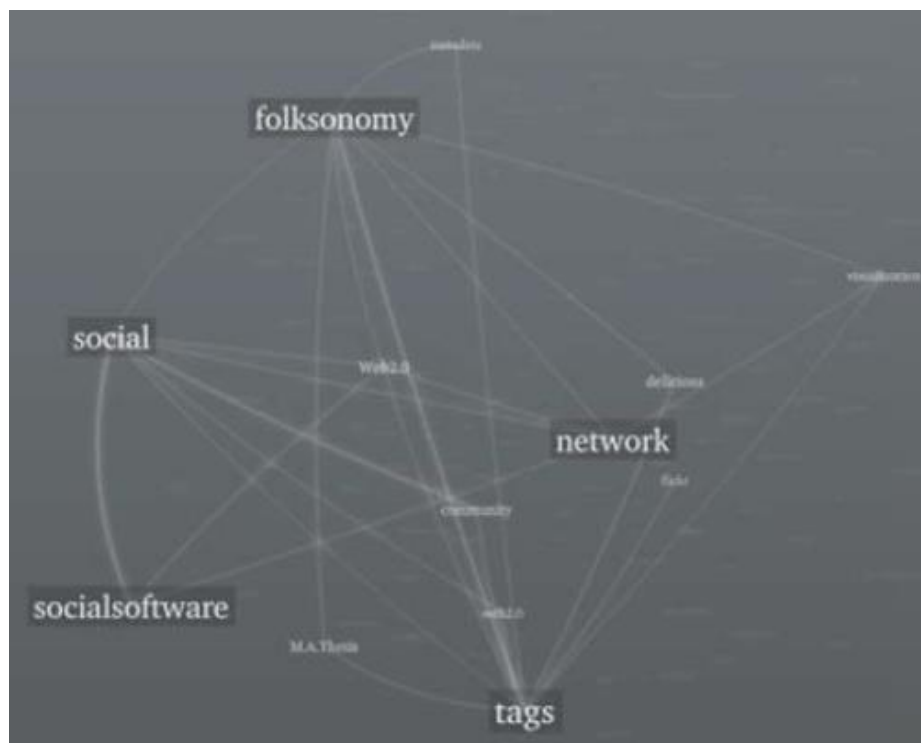


Figura 16 - Elastic Tag Map.

[Zamora, Nistal, 2009]

2.3.2.7 Sixpli

Sixpli, ou 6pli, é uma ferramenta de visualização desenvolvida para a plataforma Delicious. Os utilizadores podem navegar pela sua rede de *tags* de uma forma interativa, que inclui visualização do tipo “*elastic tag maps*” em três dimensões e círculos em duas dimensões.

Os recursos são listados do lado direito da interface, como está patente na Figura 17.

2. Tags

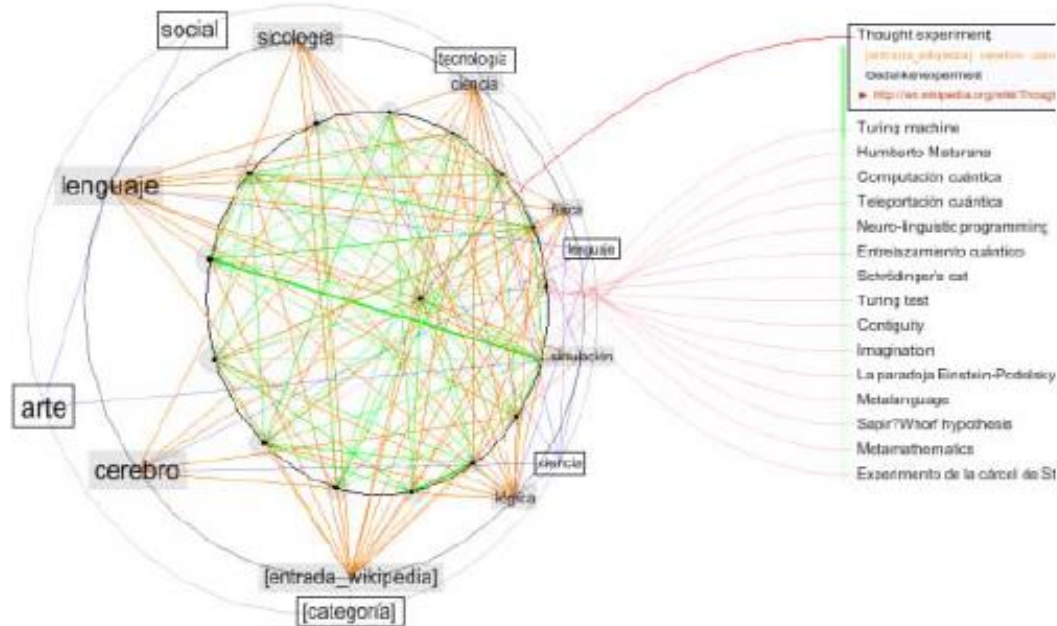


Figura 17 - Sixpli.

[Zamora, Nistal, 2009]

2.3.2.8 Cloudalicious

Cloudalicious é uma ferramenta de visualização *on-line* para o Delicious, que foi desenvolvida para podermos ver a evolução de *tag clouds* e *tag spaces* através do tempo.

Esta ferramenta recebe um *URL*, importando toda a informação sobre as *tags* do Delicious para o mesmo. Por fim, mostra todas as informações dos utilizadores através de um gráfico temporal sobre esse *URL*.

Assim, pode-se interpretar os padrões resultantes, como a opinião coletiva dos utilizadores sobre como rotular um determinado *URL*, dado as linhas do gráfico irem estabilizando através do tempo [Russell, 2008].

2.3.2.9 SpaceNav

A ferramenta SpaceNav representa as ligações entre *tags* de forma empírica. Como visível na Figura 18, a *tag "information"* foi aplicada a 9% dos recursos no Delicious e foi relacionada em 16 % com outras *tags* no *tag space*.



Figura 18 – SpaceNav.

[Revealicious, 2005]

2.3.2.10 Grouper

O Grouper foi desenvolvido para uma melhor organização dos *bookmarks* e para providenciar uma melhor vista geral sobre a personomia. O *layout* (Figura 19) é separado em três áreas de popularidade: 1) *most used tags*; 2) *frequently used tags*; 3) *less frequently used tags*. Acrescenta a isto que o Grouper oferece a possibilidade de visualização de *tags* relacionadas. Na Figura 19, pode-se ver que a *tag* “Mac” (em evidência) é a mais relacionada com as *tags* a *bold* (como por exemplo: “Cool” e “Tool”).

2. Tags



Figura 19 – Grouper

[Revealicious, 2005]

2.3.2.11 Flickr Clusters

Flickr Clusters é uma ferramenta que, dada uma *tag*, mostra *tags* relacionadas agrupadas em “clusters” ou grupos, considerando o *tag space* do Flickr. Por exemplo, ao procurar pela palavra “jaguar”, são visíveis os grupos formados que derivam de várias categorias semânticas da palavra: animal, carro e música (Figura 20).

Assim, torna-se mais fácil explorar a informação e forma-se para o utilizador um caminho definido para a informação (ex: “Explorar/Tags/jaguar/clusters/car,classic,auto” - se o primeiro *cluster* fosse seleccionado).

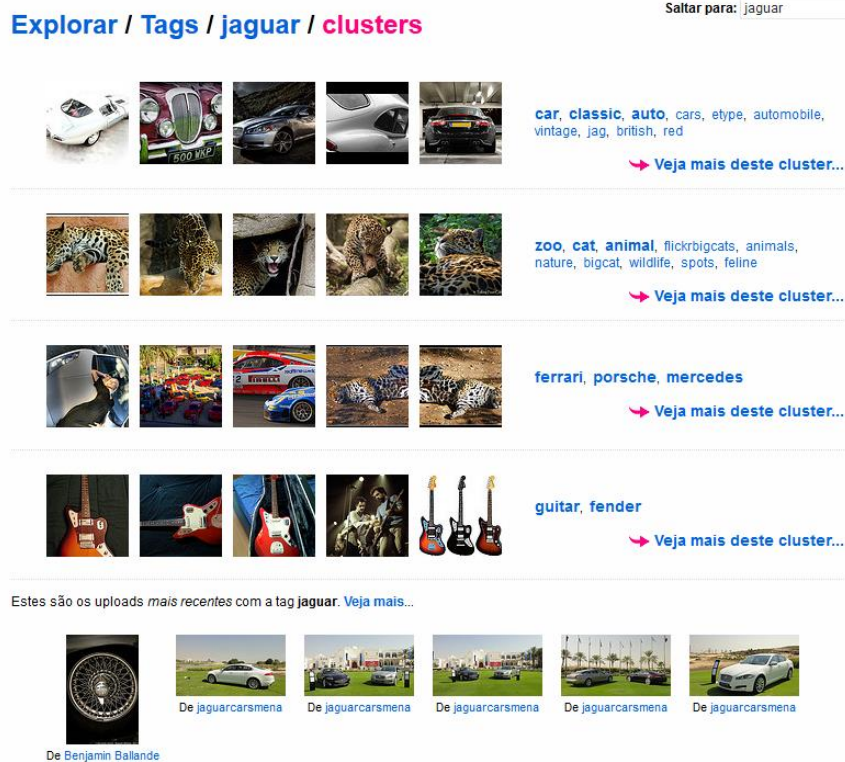


Figura 20 - Flickr Clusters.

[Flickr, 2002]

2.3.2.12 Flickr Tag Visualization

Flickr Tag Visualization, visível em <http://smilehyun.com/flickr-tag-visualization/>, é uma forma de visualização de quantidades massivas de informação utilizando *Processing*, que é uma linguagem baseada em Java. A informação do Flickr é importada para um ficheiro XML. Já num vetor, todos os termos são comparados de forma a ver quais os utilizados mais de 10 vezes. Atribui uma cor aleatória, e o utilizador pode então navegar através do espaço 3D com o teclado.

É possível, ainda, visualizar a coocorrência de tags, e assim ver as relações entre as mesmas [smilehyun, 2009].

2.3.2.13 Tag's Life

Tag's Life (visível em <http://festivalenter.wordpress.com/2009/04/09/tag-ties-and-affective-spies/>) é uma aplicação desenvolvida para o Flickr e analisa o tempo de vida das tags. Ela responde à questão de como as tendências na internet evoluem em termos de tempo e progresso, e demonstra o aparecimento e desaparecimento de tags.

2. Tags

2.3.2.14 Flickr Related Tag Browser

Existem muitos sítios *web* que colecionam as *tags* dos seus utilizadores. Assim, e dado o interesse de se poder explorar estes *tag spaces* com uma modelação visual gráfica, nasceu este sistema que permite percorrer as imagens do Flickr. Numa pesquisa por palavra são apresentados todos os termos associados, assim como uma compilação na parte central de todas as fotos associadas à pesquisa. O mesmo pode ser visível em http://www.airtightinteractive.com/projects/related_tag_browser/app/ [airtightinteractive, 2005].

2.3.3 Comparação entre formas de visualização

Na Tabela 2 pode-se observar as várias formas de visualização anteriormente descritas e comparar alguns aspetos relevantes das mesmas ao nível da informação apresentada, podendo assim identificar e resumir as várias diferenças.

De notar que nenhuma forma de visualização implementa uma distinção entre os vários tipos de *tags* que os utilizadores usam como descrito anteriormente na secção 2.1 havendo, no entanto, em algumas, uma organização intrínseca e própria.

Tabela 2 - Comparação entre formas de visualização

Formas de visualização	Multiplataforma	Visualização de relações	Conteúdos em múltiplos formatos	Detalhes disponíveis	Formas de organização
Tag Clouds	Sim	Não	Não	Não	Sim
Tag Index	Sim	Não	Não	Não	Indexação alfabética
Tag Graph	Sim	Através de linhas	Texto e linhas	Não	Não
Tag Soup	Sim	Através de linhas	Texto, linhas e esferas	Não	Não
Topigraphy Cloud	Sim	Através de proximidade no mapa topográfico	Texto e imagem topográfica	Não	Não
Taglines	Flickr	Não	Texto e imagens	Informação temporal	Eixo Temporal
Elastic Tag Map	Sim	Através de linhas	Texto e linhas	Não	Não
Sixpli	Sim	Através de linhas	Texto e linhas	Não	Não
Cloudalicious	Delicious	Não	Texto e gráfico	Informação temporal e número de utilizadores	Não
SpaceNav	Delicious	Através de percentagens	Não	Informação percentual de utilização	Não
Grouper	Delicious	Através da utilização de <i>bold</i>	Não	Número de itens relacionados	Popularidade

2. Tags

Flickr Clusters	Flickr	Através da listagem de <i>tags</i> associadas	Texto e imagens	<i>Tags</i> associados	Categoria
Flickr Tag Visualization	Flickr	Através de linhas	Texto e linhas	Não	Não
Tag's Life	Flickr	Não	Texto, imagens e gráficos	Informações numéricas de utilização	Não
Flickr Related Tag Browser	Flickr	Através da amostragem de <i>tags</i> associadas	Texto e imagens	Não	Não

2.4 Limitações conhecidas às formas de visualização de tags

Apesar das inúmeras vantagens já referidas ao longo dos subcapítulos anteriores, existem também alguns problemas associados à utilização de sistemas de *tags*. Assim, esta secção centra-se essencialmente na identificação de vários problemas existentes na utilização das diversas formas de visualização, nos diversos sistemas estudados, sendo que todas estas limitações se encontram presentes na forma de visualização mais utilizada, a nuvem de *tags*.

2.4.1 Limitação da amostragem do *tag space*

Em sistemas que utilizam nuvens de *tags* (e não só), verifica-se que o utilizador é sugestionado a pesquisar sempre os mesmos termos, pois estes aparecem destacados na forma de visualização, o que origina uma inclinação para a sua escolha. Como normalmente estes sistemas baseiam a escolha da amostragem de resultados no número de vezes que os termos foram utilizados, a forma de visualização irá tendencialmente mostrar sempre os mesmos resultados, dado serem estes teoricamente os mais populares.

Assim, há uma limitação da amostragem do *tag space*, uma vez que apenas são mostradas as *tags* mais utilizadas que poderão, ou não, ser as mais indicadas para a pesquisa que um utilizador pretende.

2.4.1.1 *Limitação do espaço de amostragem*

Adicionalmente, existe uma restrição clara no espaço visual de amostragem e, também devido a esse facto, a característica descrita no ponto anterior tem outra relevância, pois existe uma redução ainda maior no número de elementos no espaço de amostragem que, por si só, já é bastante limitado e pouco ou nada variado.

2.4.1.2 *Diversificação e aumento da granularidade do tag space*

Estando estes sistemas normalmente baseados na colaboração da comunidade que integra os mesmos, existem também diversas desvantagens inerentes a este facto.

2.4.1.2.1 Termos sinónimos

Uma das mais notórias limitações pode ser entendida se se pensar que cada utilizador é livre de rotular os recursos como bem entender, levando a que dois utilizadores, embora possam potencialmente estar a atribuir o mesmo significado, utilizem dois termos sinónimos mas, como não existe qualquer ligação direta entre os mesmos, irá consequentemente ocorrer uma diversificação e aumento da granularidade do *tag space*. Por conseguinte, as pesquisas tornam-se menos eficazes, pois existe uma dispersão dos recursos por termos sinónimos, mas sem ligação aparente. Assim, por exemplo, se um utilizador pesquisar pela palavra carro, terá resultados diferentes se pesquisar pela palavra automóvel. Embora ambas as *tags* tenham o mesmo significado, como estão associados a recursos diferentes, a pesquisa não irá retornar a união dos resultados, mas apenas o conjunto derivado de um deles.

2.4.1.2.2 Tradução de termos

O mesmo se passa se for feita uma reflexão ao nível de tradução de termos. Um determinado utilizador pode escolher rotular um recurso em inglês, enquanto um outro pode querer rotular um outro recurso com o mesmo significado, mas em português. Assim, se por exemplo pesquisarmos por carro, ou por *car*, embora ambos os termos tenham sido utilizados para rotular o mesmo tipo de recursos, não será retornada na pesquisa a união dos resultados, mas apenas um deles. Note-se que a utilização de termos em várias línguas é comum em várias plataformas mundialmente conhecidas. Assim, muitas vezes os utilizadores atribuem *tags*, não apenas escritas na língua inglesa, mas também na sua língua nativa. Mais uma vez assiste-se a um aumento da granularidade do *tag space*, sem que sejam apresentadas vantagens resultantes desse fenómeno.

2.4.1.2.3 Termos semelhantes

Outro problema também comum e presente neste tipo de sistemas e que surge na linha dos dois anteriormente apresentados, é a “repetição” de termos semelhantes, ou seja, existirem dois termos que diferem apenas, por exemplo, na conjugação entre o singular e o plural. O termo automóvel e o termo automóveis são de facto termos diferentes, embora queiram potencialmente significar o mesmo mas, como não existe qualquer espécie de mapeamento entre eles, é impossível, numa pesquisa, retornar o universo dos recursos de ambos os termos, mas apenas do pesquisado. Esta característica não se estende só a plural/singular, mas a uma infinidade de casos como a utilização de diferentes separadores entre palavras como “*race_car*” ou “*race car*”. Neste último exemplo, embora ambos os termos signifiquem exatamente o

2. Tags

mesmo, por terem sido escritos de maneiras diferentes, nunca será mostrado o resultado de ambos mas, mais uma vez, apenas o conjunto de resultados para o termo pesquisado.

2.4.1.2.4 Tag gardening

A expressão *tag gardening* refere-se ao reajuste das *tags*, como alusão ao reajuste de plantas na jardinagem. Uma das atividades do processo é justamente lidar com as variantes ortográficas das *tags* [Weller, Peters, 2008].

Um sistema de tagging pode ser melhorado através das considerações das variações de escrita [Hess, Maass, Dierick, 2008]. Note-se que a propensão para existência de erros em *tags* é mais elevada que nos sistemas de indexação tradicionais [Hammond, Hannay, Lund, Scott, 2005].

Assim, neste tipo de sistemas é natural assistir-se a uma dispersão dos recursos, pois cada utilizador pode rotular os mesmos como entender, ficando comprometida, desde logo, a eficácia de retornar posteriormente os recursos. Acresce a estes factos, ainda, o problema de os recursos serem erradamente rotulados, tanto de forma propositada, como devido, por exemplo, a diferenças no significado entre o utilizador que atribui o termo e o utilizador que o pesquisa.

A Tabela 3 ilustra os problemas apresentados anteriormente na plataforma Ebay, mostrando a diferença de número de resultados retornados entre os diversos termos.

Tabela 3 - Número de resultados retornados na plataforma Ebay

Termos pesquisados na plataforma Ebay	Número de resultados
Carro	652
Automóvel	6
Automóveis	2
<i>Car</i>	3.504.866
<i>race_car</i>	40.085
<i>race car</i>	177.376

2.4.2 Sistemas que colmatam algumas das limitações

Diversas propostas de visualização têm tentado dar resposta a algumas das limitações das nuvens de *tags* mais comuns, nomeadamente pela visualização para cada *tag* de outras que comumente são associadas aos mesmos recursos. Isto mesmo pode ser visto em algumas das

formas de visualização descritas anteriormente, tomando *layouts* diversos, mas sempre com o intuito de estabelecer relações entre os termos, baseando-se para isso, normalmente, na análise de frequência de *tags* e de coocorrência de *tags*.

Assim, o mesmo pode ser visto, por exemplo, no capítulo 2.3.2.3, onde é descrita na forma de visualização *Tag Soup* da plataforma Delicious.

No entanto, outras abordagens surgiram, nomeadamente por Sanchez-Zamora e seus colegas, que vão além da análise de frequência de *tags* e de coocorrência de *tags*. Estes propõem a visualização de *tags* numa rede semântica que considera também relações entre *tags* como, por exemplo, a sinonímia [Zamora, Nistal, 2009].

Não obstante, das diversas propostas descritas anteriormente, nenhuma singrou como a abordagem a seguir, uma vez que não colmatam todos os problemas comumente associados às nuvens de *tags*, mas apenas respondem a alguns deles. Assim, a escolha varia de sistema para sistema e é adotada a forma que melhores soluções oferece ao sistema em questão.

2. *Tags*

3 Molecule – Sistema de organização e visualização de tags

Consideradas as limitações apresentadas no capítulo anterior, enumeradas após o estudo de diversas formas de visualização e diversos sistemas que utilizam *tags*, com o objetivo de corrigir as mesmas, mas ao mesmo tempo aproveitar as potencialidades da visualização de informação com nuvens de *tags*, nasceu o sistema Molecule (<http://molecule.dei.isep.ipp.pt/>), que é apresentado e descrito neste capítulo.

3.1 Introdução

Molecule (molécula) foi a designação escolhida para o sistema, pois tal como uma molécula é composta por átomos, e os átomos compostos por eletrões, o sistema Molecule é, na verdade, a combinação de vários termos de áreas diversas, o que permite a visualização de dados relacionados, formando assim uma molécula de informação conectada.

Com este sistema, é possível aos utilizadores partilhar com a comunidade informação diversa e rotulá-la com termos que permitam a sua posterior utilização.

No entanto, o sistema Molecule vai além da normal forma de funcionamento destes sistemas, e permite criar vários “átomos” de informação relacionada com a pesquisa, permitindo assim ao utilizador ter uma visão mais alargada de todos os recursos relacionados com os termos da pesquisa.

O sistema Molecule procura, de forma automática, mais informação relacionada com a consideração de traduções, sinónimos e similares aumentando, assim, claramente o seu universo de informação relacionada, obtendo maior interligação entre os recursos e determinadas *tags*.

Dado o *tag space* do sistema ter um crescimento exponencial, o sistema Molecule apresenta ainda outra mais-valia que o demarca de todos os outros sistemas: a sua modelação gráfica original que dá o nome ao mesmo. Como as *tags* associadas aos dados provêm de diversos campos, o Molecule apresenta dados de todos estes campos em átomos distintos, o que permite apresentar a informação de forma mais completa (Figura 21).

- Tradução de termos;
- Similaridade de termos;
- Sinonímia de termos;
- Termos mais utilizados.

Esta última é a forma normalmente vista neste tipo de sistemas que reúne os termos mais utilizados, com a particularidade de estes estarem relacionados com o termo pesquisado.

3.2.1 Tradução de termos

Com o intuito de solucionar os problemas relacionados com a tradução de termos apresentados na secção 2.4.1.2.2, o sistema Molecule faz pedidos ao serviço Google Translate, para obter as traduções para as línguas portuguesa, inglesa, alemã, espanhola e italiana.

De notar que não é usada a API da Google, que recentemente deixou de ser gratuita, mas sim um pedido http directo, o qual é tratado na resposta, sendo inseridos apenas termos novos à plataforma. O número de línguas, embora limitado a estas cinco, poderá ser facilmente aumentado. O sistema Molecule utiliza apenas estas cinco línguas porque, para já, o único intuito é a demonstração do conceito.

3.2.2 Sinonímia de termos

Para colmatar as desvantagens descritas na secção 2.4.1.2.1, o sistema Molecule, para o termo original, assim como para cada tradução, faz um pedido ao serviço words.bighugelabs.com, para que este retorne sinónimos para o termo em questão. Este pedido é feito à API do sistema, a qual retorna uma lista de termos no formato JSON, que é tratada e introduzida na plataforma, criando-se assim todas as ligações entre os termos sinónimos.

De notar que, inicialmente, o sistema Molecule tinha sido integrado com a API do sistema WordReference.com, mas dado este ter deixado de estar disponível, deu-se a integração posterior com o serviço words.bighugelabs.com. Não obstante o transtorno e trabalho perdido, esta situação veio confirmar a facilidade de integração do Molecule com outros serviços.

3.2.3 Similaridade de termos

Para solucionar as questões levantadas nas secções 2.4.1.2.3 e 2.4.1.2.4, o Molecule está preparado para efetuar comparações entre termos semelhantes.

No momento do registo de uma nova *tag*, esta, as suas traduções e os seus sinónimos são comparados com os termos já existentes na base de dados da aplicação. Se um determinado

3. Molecule – Sistema de organização e visualização de *tags*

coeficiente de similaridade for obtido, será criada uma relação de similaridade entre a *tag* e o termo existente na base de dados.

Acresce que o sistema Molecule, numa pesquisa em que não seja retornada nenhuma correspondência com o *tag space* ou recursos, efetua uma pesquisa em tempo real, procurando encontrar similaridades entre o termo e as *tags* já armazenadas, retornando os resultados similares num único átomo.

Estes processos são feitos através do recurso ao algoritmo de Levenshtein, desenvolvido em 1965 por Vladimir Levenshtein. Trata-se de um algoritmo de distâncias que calcula o número de modificações necessárias para transformar um termo num outro [Levenshtein, 1965]. Através da utilização deste algoritmo, o sistema Molecule mede a similaridade entre dois termos.

Existem outros algoritmos de similaridade, mas que tendencialmente medem a similaridade através de outros pressupostos, como por exemplo o número de letras diferentes entre uma palavra e outra (algoritmo de Hamming).

No caso do sistema Molecule, e de forma a tornar a operação de detecção de similaridade o mais exate possível, à métrica resultante da execução do algoritmo, é tido ainda em conta o número de letras que constituem a palavra uma vez que quando se trata de uma palavra de pequena dimensão, corre-se o risco de o número de trocas ser pequeno e a palavra adquirir um significado muito diferente.

Assim, em casos como os termos “race car” e “race_car”, será possível encontrar os recursos referentes a ambos, dado que é estabelecida uma relação de similaridade entre os vários termos da base de conhecimento, tanto no momento da sua introdução, como no caso de uma pesquisa em que nenhum termo é retornado.

3.2.4 Termos mais utilizados

Existe ainda uma quarta esfera que reúne os termos mais utilizados relacionados com o termo introduzido na pesquisa.

Assim, neste átomo são mostrados os termos mais utilizados do conjunto de termos constituídos por todos os termos similares, mais todas as traduções, mais todos os sinónimos.

De notar que, no limite, caso os termos mais utilizados provenham todos da mesma área, então a esfera será constituída exclusivamente por esses termos, uma vez que não existe limite quanto ao número máximo de termos provenientes de uma área. Isto permite dar ao utilizador um indicador e uma melhor ideia, sobre a utilização de *tags* da comunidade, relacionadas com a pesquisa.

3.3 Vantagens da forma de organização do sistema Molecule

Dada a maior interligação entre recursos, através de *tags* que podem diferir mas que estão de alguma forma relacionadas, o sistema Molecule apresenta várias mais-valias se comparado com os sistemas tradicionais de nuvens de *tags*.

Assim, dado a geração automática de um enorme volume de *tags* a partir de uma única, associadas desde logo entre elas, são resolvidas várias questões intrinsecamente ligadas à inserção manual de *tags* pelo utilizador. Com este *tag space* e, não sendo permitida a duplicação de *tags* no mesmo, a probabilidade do utilizador usar uma *tag* previamente introduzida é muito elevada. Este comportamento leva a que todas as desvantagens provenientes da duplicação de termos como, por exemplo, por erros ortográficos, singular/plural, entre outras, fiquem virtualmente eliminados, uma vez que o utilizador irá tendencialmente recorrer a uma *tag* que foi previamente inserida de forma automática e devidamente mapeada, para rotular um determinado recurso.

Tal não acontece, por exemplo, na plataforma Amazon.com, que é uma loja *online* utilizada por pessoas de diferentes países e nem todas fluentes na língua inglesa, assim como em várias outras plataformas. Em Amazon.com existem 278 produtos aos quais foi atribuída a *tag* “excelent” (http://www.amazon.com/tag/excelent/products/ref=tag_dh_istp – *tag* com um erro ortográfico), 148 produtos com a *tag* “excelente” (http://www.amazon.com/tag/excelente/products/ref=tag_dh_istp – *tag* possivelmente escrita noutra língua como o português) e 5.382 produtos com a *tag* “excellent” (http://www.amazon.com/tag/excellent/products/ref=tag_dh_istp – *tag* escrita na língua inglesa). Mas as *tags* “excelent” e “excelente” não aparecem como *tags* usualmente utilizadas como a *tag* “excellent”, pelo que os sistemas discutidos na secção 2.2 não conseguiriam lidar bem com situações deste tipo, para o qual o sistema Molecule foi também pensado.

Assim, no sistema Molecule, a dispersão de termos no *tag space* e, consequentemente, a dispersão de recursos por falta de relações pré estabelecidas entre *tags*, é também atenuada porque, por exemplo, embora um utilizador possa estar a usar um sinónimo ou tradução de um outro termo já referenciado, ambos serão retornados no universo da pesquisa, uma vez que o sistema compreende que se tratam potencialmente do mesmo, ou de assuntos relacionados.

Mesmo assim, e apenas aplicando as mais-valias já descritas, continuar-se-ia a obter resultados distintos em termos semelhantes como “race car” e “race_car”. Isto acontece porque, por exemplo, um utilizador teria introduzido manualmente a *tag* “race_car” não utilizando o previamente introduzido “race car”, nem fazendo o mapeamento manual. Para colmatar este problema, o Molecule efetua comparações entre termos semelhantes como referido na secção anterior (3.2.3).

Desta forma, os resultados obtidos na Tabela 3 seriam bastante diferentes e, virtualmente, verificar-se-ia o mesmo número de ocorrências nos termos “carro”, “automóvel”, “automóveis”, “car”, “race car” e “race_car”.

Por fim, o sistema Molecule tem uma característica, habitual nos sistemas que utilizam *tags*, que é a descontextualização dos termos. Como não existe um contexto associado aos termos, isto permite ao utilizador em última análise, decidir o que é relevante de acordo com os seus interesses. Existe no entanto uma desvantagem que pode ser identificada quando um termo pode ter significados diferentes mediante a situação em que é utilizado e consequentemente induzir o utilizador em erro. A palavra “banco” exemplifica o problema pois num contexto de mobiliário adquire um significado diferente de se utilizada num contexto de instituição financeira.

3.4 Modelação gráfica do sistema Molecule

O sistema Molecule é caracterizado pelo agrupamento de *tags* distintas mas de alguma forma relacionadas e a sua apresentação ao utilizador, para que este tenha uma melhor visão do *tag space* relevante.

A utilização de múltiplas vistas, é particularmente indicada quando uma única seria demasiado complexa e cognitivamente exigente para os utilizadores. Baldonado et al. propuseram, entre outras, a regra da diversidade para o uso de várias vistas [Baldonado, Woodruff, Kuchinsky, 2000]. Segundo esta regra deve-se utilizar mais do que uma vista quando existem múltiplos atributos, perfis, modelos ou níveis de abstração. Nestas situações, uma única visão dos dados pode requerer demasiado esforço de assimilação para a compreensão da multiplicidade dos dados. Para tal ser possível, foi necessário o desenvolvimento de uma forma original de modelação da informação, que é apresentada de seguida.

3.4.1 Átomos

Dado o sistema ser constituído por várias áreas, começou a surgir o conceito de esfera de conhecimento, que mais tarde acabou por estar na origem do nome do sistema, uma vez que esta foi associada a um átomo que por sua vez forma uma molécula.

Para cada uma das suas áreas de conhecimento o sistema Molecule forma um átomo.

Os átomos da molécula do sistema, giram em volta do seu centro permitindo, por um lado, uma maior organização do espaço visual e, por outro, permitindo ao utilizador poder ter em destaque o átomo que mais lhe interessa para a pesquisa em questão, tendo assim mais uma forma de “filtrar” a informação mantendo, na mesma, toda a restante disponível.

Para ajudar o utilizador na identificação do significado da esfera na molécula e assim compreender e chegar mais rapidamente à informação que pretende, todos os átomos têm um ícone característico relacionado com os seus eletrões e, por conseguinte, com o conhecimento que representam.

Isso mesmo é visível nas próximas figuras, nas quais se pode identificar o átomo de similaridade (Figura 22), o átomo de sinonímia (Figura 23), o átomo de tradução (Figura 24) e por fim o átomo que contém os termos mais utilizados (Figura 25).



Figura 22 - Átomo de similaridade



Figura 23 - Átomo de sinonímia



Figura 24 - Átomo de tradução



Figura 25 - Átomo de termos mais utilizados

3.4.2 Eletrões

Como descrito anteriormente, o Molecule forma para cada uma das suas áreas organizacionais, um átomo de informação. Assim, cada átomo é constituído por eletrões, caso no sistema existam termos relacionados com a área organizacional em questão.

3. Molecule – Sistema de organização e visualização de *tags*

Cada eletrão representa um termo no qual o utilizador pode clicar de forma a refinar a sua pesquisa e navegar pelo sistema.

De notar que cada eletrão tem um tamanho diferente relacionado com a sua utilização por parte da comunidade do sistema. Assim, o utilizador pode identificar de forma mais fácil quais os termos relacionados com a sua pesquisa, que são os mais utilizados pelos restantes utilizadores.

De forma a cada átomo poder comportar mais termos, e o utilizador poder ter em evidência aqueles que lhe convêm, os eletrões rodam em volta do seu centro, tal como se de um átomo químico se tratasse.

Assim, tal como no átomo em sí, o utilizador tem por um lado uma melhor organização do espaço visual e, por outro, dispõe de mais uma forma de poder ter em evidência a informação que mais lhe convém para a pesquisa em questão, nunca perdendo de vista toda a restante.

3.4.3 Diferentes moléculas geradas pelo sistema

O sistema Molecule modela três moléculas diferentes mediante o estado da pesquisa atual. Assim, se for pretendido carregar o átomo inicial, uma molécula de átomo único e sem núcleo é gerada (Figura 26).

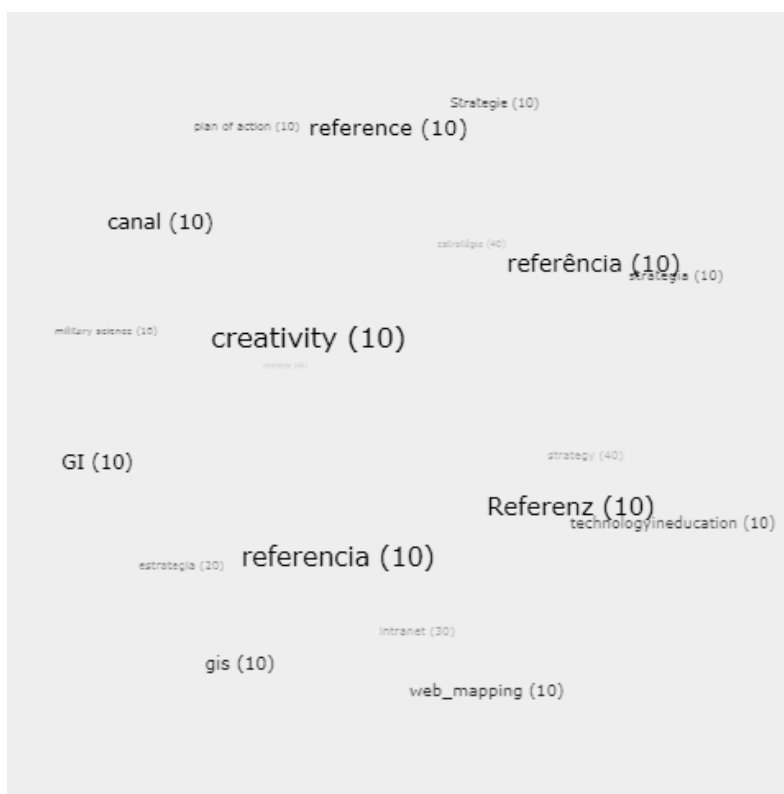


Figura 26 - átomo inicial

Se for pretendido o carregamento de toda a informação proveniente das diferentes áreas, uma molécula completa é gerada (Figura 21). Por fim, caso nenhum termo seja encontrado, é feita uma procura de similaridade em tempo real por todos os termos da base de conhecimento.

Nesse caso, é gerada uma molécula de átomo único, sem núcleo, em que estarão como eletrões os termos encontrados resultantes dessa pesquisa (Figura 27).



Figura 27 - Pesquisa pelo termo “*references*” com retorno de dois termos similares

3.4.4 Vantagens da abordagem de átomos comparativamente às nuvens de *tags* convencionais

Com este conceito de visualização, o sistema Molecule consegue lidar com várias desvantagens recorrentes da utilização de nuvens de *tags*, descritas anteriormente na secção 2.4.1.1.

Assim, para além das vantagens apresentadas no capítulo 3.3, sobre a forma de organização da informação, o Molecule apresenta ainda outras vantagens derivadas da sua modelação gráfica.

A limitação ao nível da variedade do universo do espaço de amostragem, tão característica noutros sistemas como as nuvens de *tags*, é ultrapassada, dado não mostrar apenas os recursos rotulados exatamente com um termo específico, mas também todos os recursos e *tags* que, de alguma forma, se encontram ligados/mapeados ao mesmo (Figura 21).

A molécula apresentada é, na verdade, o resultado das várias interpretações do termo, oferecendo assim ao utilizador uma maior escolha. Este facto traduz-se, também, num aumento exponencial e claro da probabilidade de sucesso, em caso de diferenças ao nível semântico, da interpretação do termo usado pelo utilizador que pesquisa e do utilizador que rotulou a informação. Desta forma, o utilizador que pesquisa pode rapidamente encontrar o que realmente procurava, pois tem acesso visual direto a várias outras interpretações/opções para o termo e, portanto, pode muito rapidamente identificar uma outra alternativa e, dessa forma, explorar e/ou refinar a sua pesquisa, aumentando a taxa de sucesso da mesma, não estando restringido só a resultados diretos.

3. Molecule – Sistema de organização e visualização de *tags*

Outra das características intrínsecas a esta forma de visualização de *tags* é o grande aumento do número de conceitos retornados, uma vez que cada “átomo” (excetuando o central) é uma “esfera” de termos, como visível na Figura 28.



Figura 28 - Átomo de termos mais utilizados para a pesquisa pela *tag* estratégia

O conceito de esfera de termos permite que os termos que estão intrinsecamente ligados estejam dispersos por um determinado espaço 3D e, portanto, ao ser acrescentada uma dimensão, torna-se claro que irá forçosamente existir um aumento da área disponível para alojar termos, em comparação à área disponível numa “normal” nuvem de *tags* (2D) e, por conseguinte, como pode ser visto na Figura 21, existe potencialmente um espaço muito maior de amostragem, se comparada com a Figura 9.

Assim, o utilizador pode ver os vários termos dispersos pela esfera, podendo rodá-la para navegar na mesma e ver em maior plano um determinado termo.

Acresce a este facto que cada molécula é constituída por quatro átomos, sendo que todos estes rodam, por sua vez, à volta do átomo principal e estático que é o termo pesquisado, ficando ainda disponível um maior espaço para a amostragem de informação.

Desta forma, a interpretação que o sistema Molecule faz da modelação gráfica do *tag space* permite tratar as restrições do espaço de visualização, pois permite a disponibilização de muito mais espaço de informação, quando comparado aos sistemas tradicionais de nuvens de *tags* retirando ao mesmo tempo, todas as vantagens já enumeradas anteriormente da utilização de múltiplas vistas.

Para além de todas as suas mais-valias funcionais, o sistema Molecule é caracterizado, também, por uma grande atratividade para os seus utilizadores resultante, não só do seu aspecto visual inovador derivado do seu conceito de modelação gráfica dos termos, mas também por convidar o utilizador a interagir com ele, de uma forma muito intuitiva e natural.

3.5 Especificação técnica da aplicação

Neste subcapítulo é apresentada a aplicação numa perspetiva técnica. Assim, em primeiro lugar serão enumeradas e descritas as tecnologias mais relevantes empregues na conceção da aplicação (secção 3.5.1).

De seguida, serão descritos os vários *layouts* existentes na aplicação, sob um ponto de vista funcional (secção 3.5.2).

Após esta subsecção, será então apresentada a arquitetura da aplicação, começando pelas diferentes camadas que a mesma possui (secção 3.5.3), passando pelo esquema relacional (secção 3.5.4), e finalizando com a apresentação das interfaces desenvolvidas (secção 3.5.5).

3.5.1 Descrição das tecnologias mais relevantes

No desenvolvimento desta aplicação, que atua em ambiente web, foram empregues diversas tecnologias, sendo as mais relevantes abaixo descritas de forma breve.

3.5.1.1 C#

C# (“C Sharp”) é uma linguagem de programação criada para o desenvolvimento de uma variedade de aplicações que executam sobre a plataforma .NET. C# é uma linguagem simples, poderosa, com tipos seguros e orientada a objetos. O C# tem como grande foco e objetivo permitir o desenvolvimento rápido de aplicações, mantendo a expressividade e a elegância das linguagens *C-style* [oficinadanet, 2007].

3.5.1.2 ASP.NET MVC

MVC (*Model View Controller*) fornece uma maneira de dividir “fisicamente” as camadas de modelo, apresentação e de controlo.

O modelo representa os dados da aplicação e as regras do negócio. O modelo mantém o estado persistente do negócio e fornece ao controlador a capacidade de aceder às funcionalidades da aplicação encapsuladas pelo próprio modelo.

Um componente de visualização (*view*) faz o *render* do conteúdo de uma parte particular do modelo e encaminha para o controlador as ações do utilizador. Acede também aos dados do modelo via controlador e define como esses dados devem ser apresentados.

Um controlador define o comportamento da aplicação. É este que interpreta as ações do utilizador e as mapeia para chamadas do modelo [macoratti, 2008].

3. Molecule – Sistema de organização e visualização de *tags*

3.5.1.3 WCF

O WCF (*Windows Communication Foundation*) é uma parte da plataforma.NET, e fornece um modelo unificado de programação para construir de forma rápida aplicações distribuídas orientadas a serviço (SOA).

Um serviço WCF é constituído pelos seguintes componentes:

- *Service Class* - Uma classe de serviço WCF que implementa um serviço com um conjunto de métodos;
- *Host Environment* - Este é o grande diferencial do WCF, pois pode-se usar qualquer tipo de aplicação como *host*. Um *Host Environment* pode ser uma aplicação *Console*, um *Windows Service*, um *web service*, uma aplicação *Windows Forms* ou o *Internet Information Service*, no caso de um *web service* normal;
- *EndPoints* - Todas as comunicações com um serviço WCF irão acontecer via *endpoints*. O endpoint é composto de 3 partes (*ABC's endpoints*):
 - *Address* - Consiste em especificar um endereço que define onde o *endpoint* está hospedado. Cada *endpoint* possui um endereço especificado que é usado para identificar e localizar o *endpoint* (geralmente é definido pela instância da classe *Uri*).
 - *Binding* - Especifica como o cliente irá comunicar com serviço (qual o tipo de transporte que será usado: HTTP, TCP, IPS, etc.) e o endereço onde o *endpoint* está hospedado. A definição do *Binding* pode ser feita no arquivo de configuração ou via programação.
 - *Contract* - Especifica um contrato que define que métodos da classe de Serviço serão acessíveis via *endpoint*, onde cada *endpoint* poderá expor um conjunto diferente de métodos (o contrato é representado por uma *Interface* que deverá ser decorada com o atributo *ServiceContract*) [macoratti, 2009a].

3.5.1.4 Silverlight

Microsoft Silverlight é uma tecnologia para navegadores e *plugins* da nova geração.

As principais características do Silverlight são:

- Ser formado essencialmente por quatro arquivos - dois arquivos em XAML (que é a linguagem utilizada, tanto pelo Silverlight, quanto pelo Windows Presentation Foundation) e mais dois arquivos *de code-behind*, ou seja, escritos em alguma linguagem de programação para manipular o XAML.

- Oferecer uma experiência ampla na incorporação de vídeos, animações e interfaces interativas.
- Ser pequeno e de fácil instalação, possuindo menos de 6 MB.
- Poder criar e trabalhar com gráficos vetorizados, assim como textos, animações e sobreposições que interagem com gráficos e efeitos de alta qualidade.
- Criarem-se aplicativos programando em diversas linguagens, como JavaScript, utilizando AJAX, C#, Python, Ruby, Visual Basic entre outras.
- Ter sido projetado para trabalhar em diversas arquiteturas, sistemas operacionais e navegadores [msdn, 2013].

3.5.1.5 Entity Framework

Entity Framework - EF é um ORM (*Object-relational mapping*) que usa um conjunto de tecnologias do ADO.NET para suportar o desenvolvimento de aplicações orientadas a dados.

A *Entity Framework*, permite trabalhar com dados na forma de propriedades e objetos específicos do domínio, entre outros, sem ter que relacioná-los com as tabelas da base de dados e as colunas onde os dados estão armazenados. Isto é possível pela elevação do nível de abstração no qual os programadores podem trabalhar quando estão a tratar com os dados e pela redução do código que é necessário para manter as aplicações orientadas a dados.

A *Entity Framework* renova o modelo conceitual, permitindo consultar entidades e relacionamentos, enquanto traduz estas operações para comandos específicos da base de dados, libertando as aplicações das dependências de código específicas e criando, assim, uma abordagem orientada a objetos para interagir com a base de dados [macoratti, 2009b].

3.5.1.6 Linq

Language-Integrated Query (LINQ) é uma inovação introduzida no Visual Studio 2008 e na plataforma .NET 3.5, e permite aproximar o universo dos objetos ao dos dados.

Tradicionalmente, pesquisas de dados são expressas em texto simples sem tipo, verificadas na compilação ou através de suporte por *IntelliSense*. Acresce que é necessário saber diferentes linguagens de pesquisa para cada tipo de sistema de armazenamento de informação: base de dados SQL, documentos XML, serviços web variados, etc. O LINQ torna uma pesquisa numa linguagem de construtores em C#. O LINQ permite escrever pesquisas sobre coleções de objetos fortemente tipados, usando uma linguagem de palavras-chave e operadores relacionais.

3. Molecule – Sistema de organização e visualização de *tags*

No Visual Studio podem ser escritas pesquisas de LINQ tanto em Visual Basic como C# para bases de dados SQL, documentos XML, ADO.NET *Datasets* e coleções de objetos que suportem *IEnumerable* ou *IEnumerable<T>* [msdn, 2012].

3.5.2 Vistas

Neste capítulo serão descritas as várias vistas desenvolvidas para a aplicação, assim como explicados para cada uma delas qual a sua funcionalidade e dados alguns extratos de código para suportar a explicação de algumas particularidades das mesmas.

Dada a aplicação ter sido desenvolvida no modelo MVC, esta secção está organizada por controlador, descrevendo todas as vistas por ele utilizadas.

3.5.2.1 Home

O controlador *Home* é responsável pela apresentação da molécula e seus resultados através de uma única *action* denominada de *Index*.

3.5.2.1.1 Index

Quando o utilizador entra no sistema Molecule é apresentada a página inicial. No canto superior direito é requerida a autenticação para se poder aceder às funcionalidades de administração, como visível na Figura 29.



Figura 29 - Página inicial do sistema Molecule

Não obstante, é permitido ao utilizador pesquisar recursos no sistema tanto através da molécula inicial formada pelas *tags* mais utilizadas, ou através da caixa de pesquisa.

Por forma a evitar que o número de resultados retornado seja muito grande, dado o mesmo ser resultante da recolha de várias fontes de conhecimento, é possível ainda ao utilizador filtrar que tipo de resultados quer ver apresentados, aquando da pesquisa (Figura 30):

- Se todos (ALL), ou seja, se o conjunto de resultados do termo, dos seus similares, das suas traduções e dos seus sinónimos;
- Se apenas os resultados diretamente relacionados com o termo (*Search only*);
- Se apenas os resultados relacionados com as traduções do termo (*Translations*);
- Se apenas os resultados relacionados com as similaridades do termo (*Similar*s);
- Se apenas os resultados relacionados com os sinónimos do termo (*Sinonyms*);

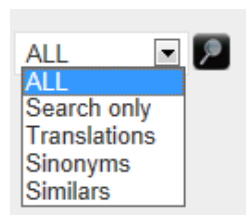


Figura 30 - Opções de pesquisa de resultados no Molecule

O mecanismo de autenticação usa a *ExtendedMembershipProvider* e reescreve os seus métodos de forma a utilizar as suas funcionalidades base mas, implementando uma lógica diferente e adaptada ao modelo de dados existente, como visível no extrato do método abaixo apresentado.

```
public class MyProvider : ExtendedMembershipProvider
{
    ...
    public override bool ValidateUser(string username, string password)
    {
        var u = te.User.Where(s => s.Username == username && s.Password
        == password);
        if (u != null && u.Count() >0)
            return true;
        else
            return false;
    }
    ...
}
```

Código 1 - Método ValidateUser

Após autenticação, o utilizador passa a ter acesso ao menu que permite administrar todos os conteúdos do sistema (Figura 31).

3. Molecule – Sistema de organização e visualização de *tags*

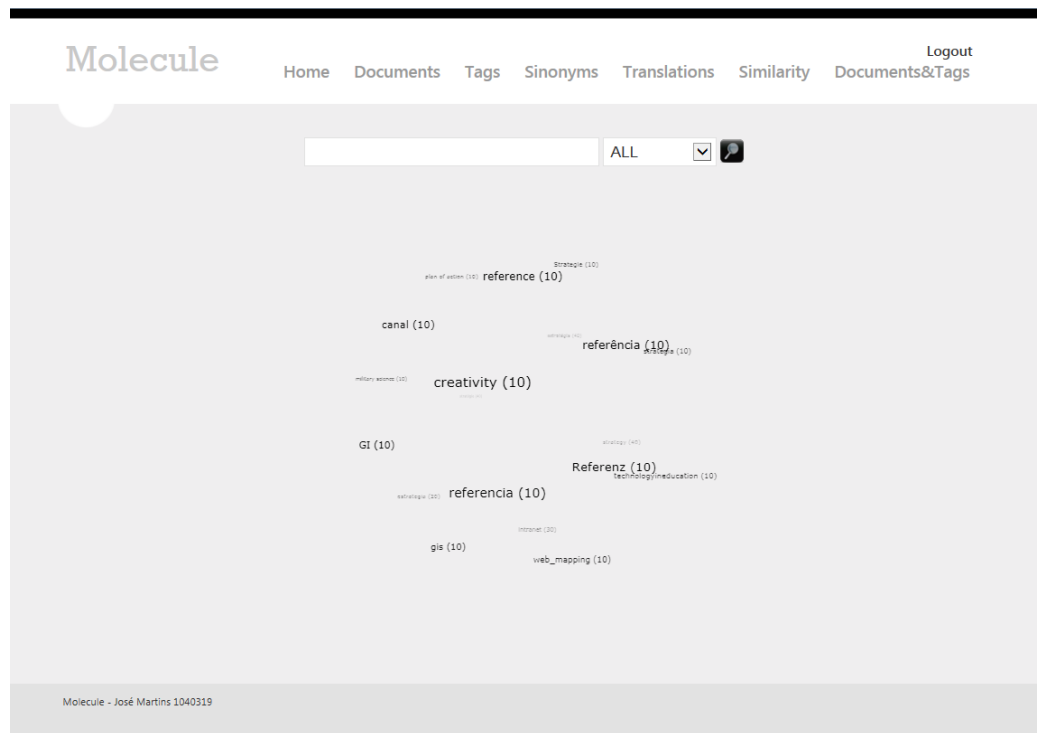


Figura 31 - Página inicial do sistema Molecule após autenticação

As várias funcionalidades estão organizadas segundo diferentes menus:

- No menu *Home* é mostrada a página inicial após autenticação;
- No menu *Documents* é dado acesso à página de administração dos documentos disponíveis para pesquisa no sistema;
- No menu *Tags* é dado acesso à página de administração das *tags* que existem no sistema para classificar documentos;
- No menu *Sinonyms* é dado acesso à página de administração das relações de sinónimos entre *tags*;
- No menu *Translations* é dado acesso à página de administração das relações de tradução entre *tags*;
- No menu *Similarity* é dado acesso à página de administração das relações de similaridade entre *tags*;
- No menu *Documents&Tags* é dado acesso à página de administração das relações entre *tags* e documentos;
- No menu *Logout* o utilizador perde a sessão e, por conseguinte, o acesso aos menus de administração, voltando à página e estado inicial.

3.5.2.2 Documents

O controlador *Documents* é responsável pela criação, eliminação, detalhe e listagem de documentos.

Para isso este usa quatro *actions* diferentes, sendo estas a *Index* (para listar documentos), *Create* (para criar documentos), *Details* (para visualizar os detalhes de um documento) e *Delete* (para eliminar um documento).

3.5.2.2.1 Index

Quando o utilizador acede ao menu *Documents*, é-lhe apresentada uma lista de todos os documentos, paginada, mostrando os dez primeiros documentos, como visível na Figura 32.

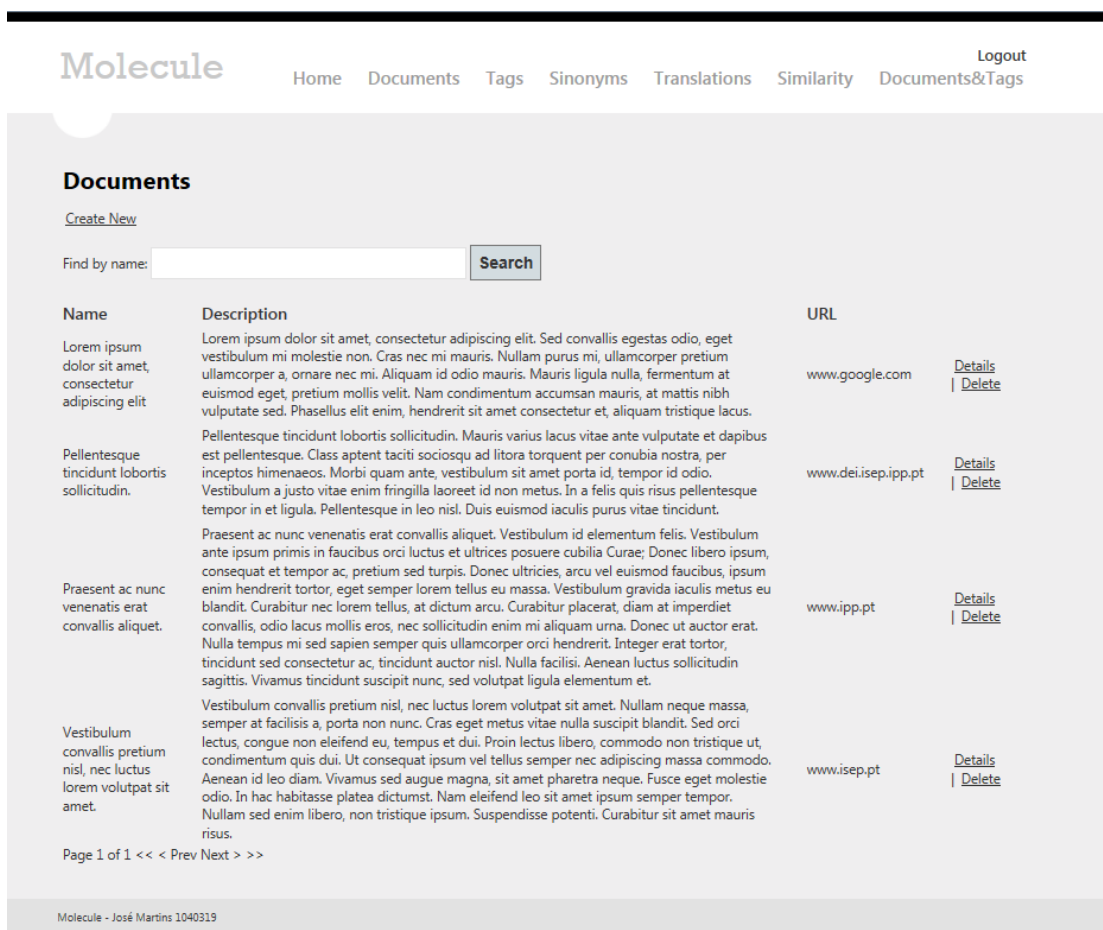


Figura 32 - Página de listagem dos documentos

O utilizador pode navegar na lista através do menu situado no inferior da página.

Nesta página pode ainda ordenar os documentos pelas colunas *Name*, *Description* e *URL*.

3. Molecule – Sistema de organização e visualização de *tags*

Para além destas ações é possível, ainda, filtrar por essas mesmas colunas através da caixa de texto.

Por fim o utilizador pode criar um novo documento ou então ver os detalhes, ou apagar algum já existente.

De notar o pormenor de ser identificado no início da página, a página em que o utilizador se encontra para o mesmo estar sempre situado no sistema.

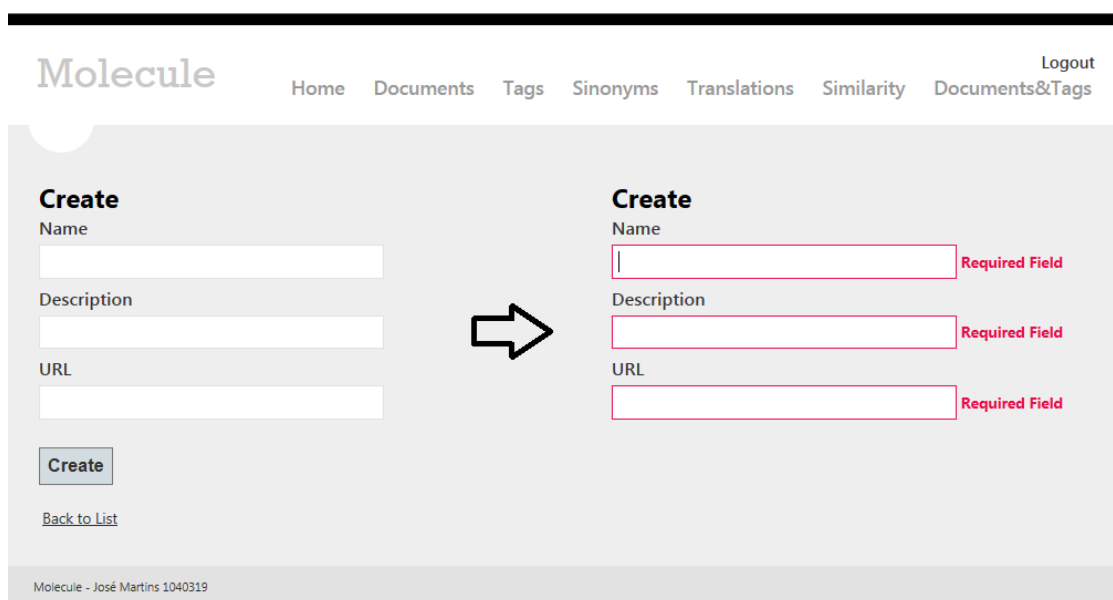
As funcionalidades de menu descritas na secção 3.5.2.1 mantêm-se, uma vez que estas estão presentes em todas as páginas após autenticação.

3.5.2.2.2 *Create*

Através do menu *Documents* é possível aceder ao menu *Create*.

Neste, o utilizador pode criar novos documentos ou voltar para a lista de documentos na qual se encontrava.

De notar que o nome do documento é único e existem mensagens de validação que forçam a integridade lógica e física dos dados, tanto do lado do cliente como do servidor, como visível na Figura 33.



The screenshot displays the 'Molecule' web application interface. At the top, a navigation bar includes links for Home, Documents, Tags, Synonyms, Translations, Similarity, and Documents&Tags, along with a Logout button. The main content area is titled 'Create' and contains three input fields: Name, Description, and URL. Each field is accompanied by a red 'Required Field' validation message. A large black arrow points from the left side of the form to the right side, indicating a transition or validation state. Below the input fields, there is a 'Create' button and a 'Back to List' link. The footer of the page reads 'Molecule - José Martins 1040319'.

Figura 33 - Página criação de novo documento

Estas e todas outras validações de integridade lógica e de negócio, são conseguidas através da adição de uma classe ("*buddy class*") que anota metadados às classes que posteriormente são instanciadas e enviadas para a *view*, como mostrado no extrato de código abaixo para a classe "Doc", que depois de instanciada representa um documento.

```

[MetadataType(typeof(Documents_validator))]
public partial class Doc
{
}

public class Documents_validator
{
    [Required(ErrorMessageResourceType = typeof(Global),
ErrorMessageResourceName = "Required")]
    [LocalizedDisplayName("Id", NameResourceType = typeof(Global))]
    public int Id { set; get; }

    [Remote("IsUniqueDoc", "Validation")]
    [Required(ErrorMessageResourceType = typeof(Global),
ErrorMessageResourceName = "Required")]
    [StringLength(250, ErrorMessage = "Must be under 250 characters in
length")]
    [LocalizedDisplayName("Nome", NameResourceType = typeof(Global))]
    public string Nome { set; get; }

    [Required(ErrorMessageResourceType = typeof(Global),
ErrorMessageResourceName = "Required")]
    [StringLength(1000, ErrorMessage = "Must be under 1000 characters
in length")]
    [LocalizedDisplayName("Desc", NameResourceType = typeof(Global))]
    public string Desc { set; get; }

    [Required(ErrorMessageResourceType = typeof(Global),
ErrorMessageResourceName = "Required")]
    [StringLength(5250, ErrorMessage = "Must be under 250 characters in
length")]
    [LocalizedDisplayName("Url", NameResourceType = typeof(Global))]
    public string Url { set; get; }
}

```

Código 2 – Classe Documents_Validator

3.5.2.2.3 Details

Através do menu *Documents* é possível aceder ao menu *Details*.

Neste, o utilizador pode visualizar os detalhes de um documento ou voltar para a lista de documentos (Figura 34).

3. Molecule – Sistema de organização e visualização de *tags*

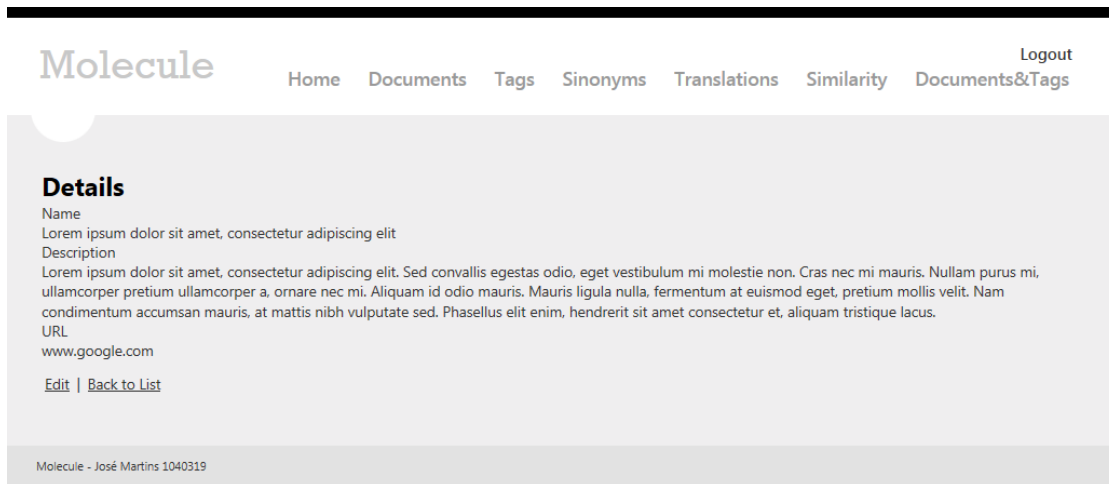


Figura 34 - Página de detalhes de um documento

3.5.2.2.4 Delete

Através do menu *Documents* é possível aceder ao menu *Delete*.

Neste, o utilizador visualiza os detalhes de um documento e decide se o apaga ou volta para a lista de documentos (Figura 35).

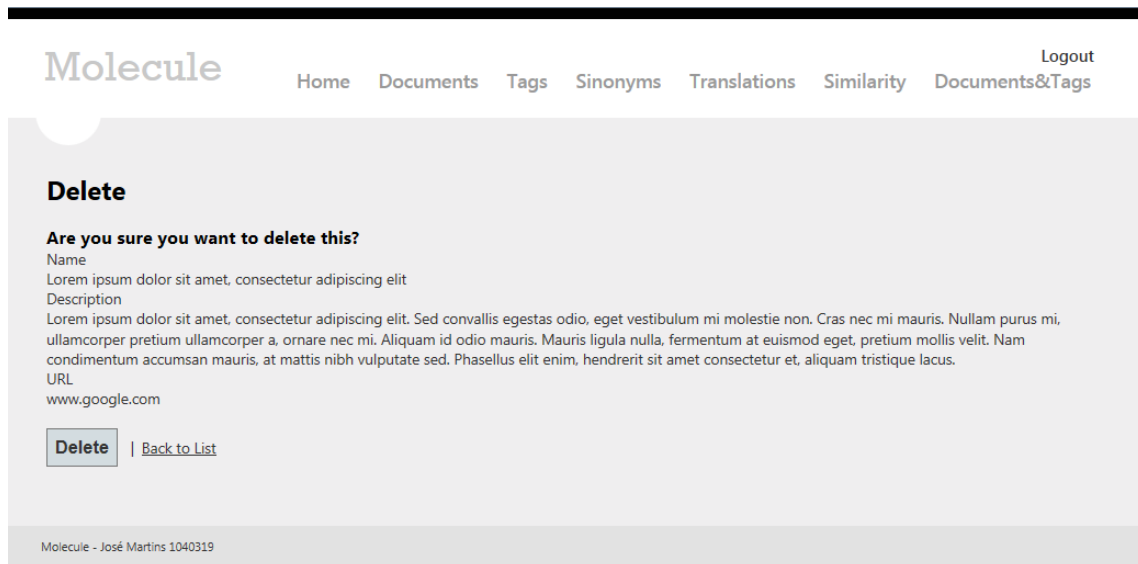


Figura 35 - Página de eliminação de um documento

3.5.2.3 Tags

O controlador *Tags* é responsável pela criação, eliminação, detalhe e listagem de *tags*.

Para isso, este usa quatro *actions* diferentes, sendo estas a *Index* (para listar *tags*), *Create* (para criar *tags*), *Details* (para visualizar os detalhes de um *tags*) e *Delete* (para eliminar uma *tag*).

O comportamento destas é idêntico aos referidos para o controlador *Documents* pelo que não serão descritas em pormenor.

3.5.2.4 *Sinonyms*

O controlador *Sinonyms* é responsável pela criação, eliminação, detalhe e listagem de relações entre *tags* sinónimas.

Para isso, este usa quatro *actions* diferentes, sendo estas a *Index* (para listar relações entre *tags* sinónimas), *Create* (para criar uma relação entre *tags* sinónimas), *Details* (para visualizar os detalhes de uma relação de *tags* sinónimas) e *Delete* (para eliminar uma relação de *tags* sinónimas).

O comportamento destas é idêntico aos referidos para o controlador *Documents* pelo que não serão descritas em pormenor.

3.5.2.5 *Translations*

O controlador *Translations* é responsável pela criação, eliminação, detalhe e listagem de relações entre *tags* traduzidas.

Para isso, este usa quatro *actions* diferentes, sendo estas a *Index* (para listar relações entre *tags* traduzidas), *Create* (para criar uma relação entre *tags* traduzidas), *Details* (para visualizar os detalhes de uma relação de *tags* traduzidas) e *Delete* (para eliminar uma relação de *tags* traduzidas).

O comportamento destas é idêntico aos referidos para o controlador *Documents* pelo que não serão descritas em pormenor.

3.5.2.6 *Similarity*

O controlador *Similarity* é responsável pela criação, eliminação, detalhe e listagem de relações entre *tags* similares.

Para isso, este usa quatro *actions* diferentes, sendo estas a *Index* (para listar relações entre *tags* similares), *Create* (para criar uma relação entre *tags* similares), *Details* (para visualizar os detalhes de uma relação de *tags* similares) e *Delete* (para eliminar uma relação de *tags* similares).

O comportamento destas é idêntico aos referidos para o controlador *Documents* pelo que não serão descritas em pormenor.

3. Molecule – Sistema de organização e visualização de *tags*

3.5.2.7 *Documents&Tags*

O controlador *Document&Tags* é responsável pela criação, eliminação, detalhe e listagem de relações entre *tags* e documentos.

Para isso, este usa quatro *actions* diferentes, sendo estas a *Index* (para listar relações entre *tags* e documentos), *Create* (para criar uma relação entre uma *tag* e um documento), *Details* (para visualizar os detalhes de uma relação de uma *tag* e um documento) e *Delete* (para eliminar uma relação de uma *tag* e um documento).

O comportamento destas é idêntico aos referidos para o controlador *Documents* pelo que não serão descritas em pormenor.

3.5.3 Arquitetura da aplicação

A aplicação foi desenvolvida com uma arquitetura multicamada, de forma a promover a modularização do código, separação de conceitos e facilitar a manutenção e evolução futura. As seguintes camadas foram consideradas:

- *Silverlight* - Camada que é responsável pela geração das três animações de moléculas descritas anteriormente. Esta camada recebe informações da camada MVC (qual o termo pesquisado) e requer informações à camada WCF para poder gerar a molécula de termos.
- MVC - Camada onde a aplicação web se encontra. Fornece informação à camada de *Silverlight* para geração das moléculas e interage com a camada de modelo, sempre que necessita de operar sobre a base de dados (pesquisar, criar, apagar dados).
- WCF - Camada que serve como porta de acesso à camada *Silverlight* para poder recolher informações da base de dados. Para interagir com a base de dados, o WCF interage com a camada modelo (pesquisar dados).
- Modelo - Camada que permite usar uma abordagem orientada a objetos na interação com a base de dados. Assegura também validações de negócio junto das entidades que expõe. Esta camada serve as camadas MVC e WCF, sempre que é necessário uma interação com a base de dados.
- Base De Dados - Camada que é usada para o armazenamento da informação do sistema. Esta apenas interage com a camada modelo.

Na Figura 36 é mostrado uma esquematização da aplicação.

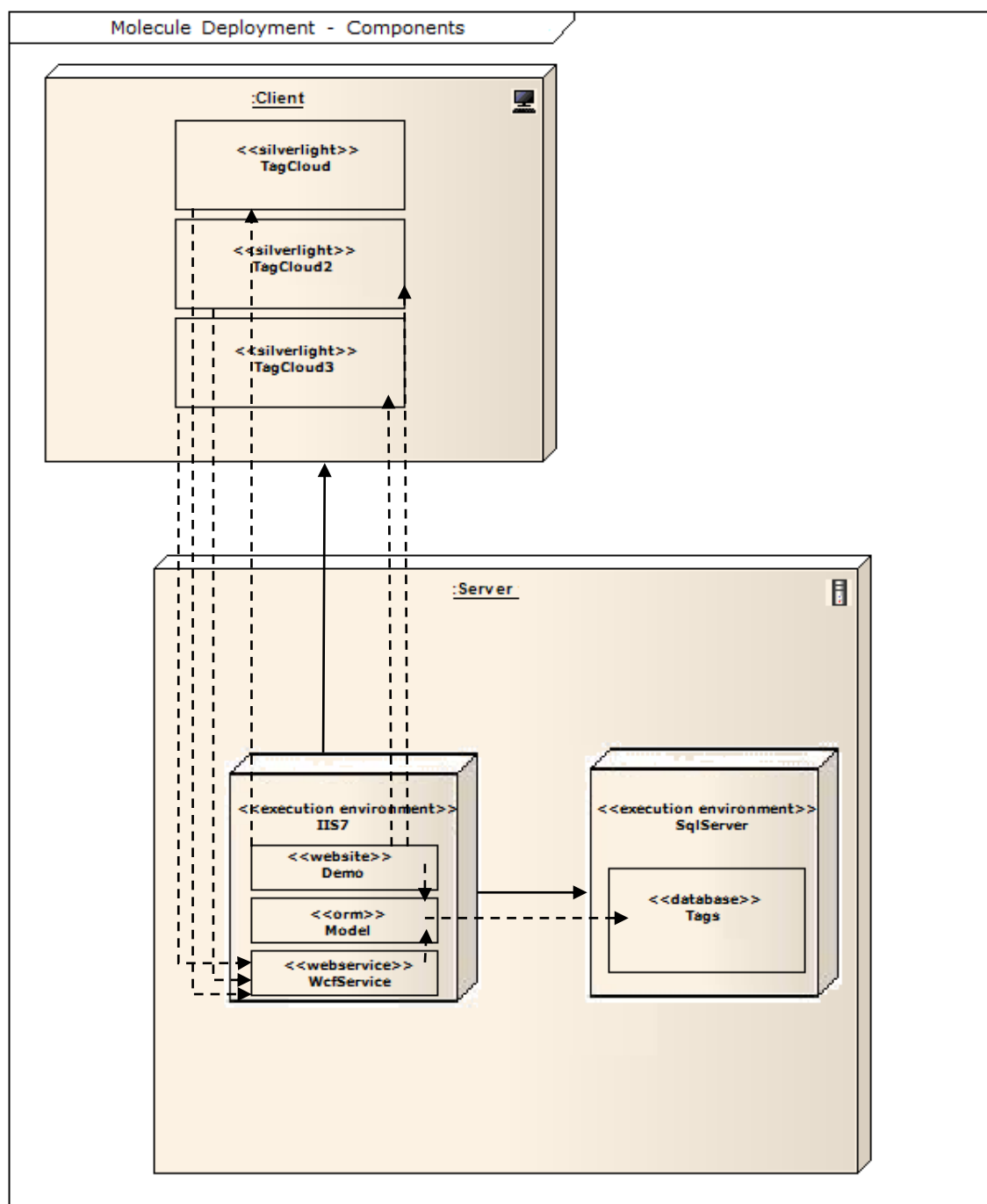


Figura 36 – Diagrama de implantação

3. Molecule – Sistema de organização e visualização de *tags*

3.5.4 Esquema Relacional

Na Figura 37 é mostrado o modelo de dados do Molecule, podendo ser visto como as diversas tabelas que integram o modelo se relacionam e a cardinalidade dessas relações.

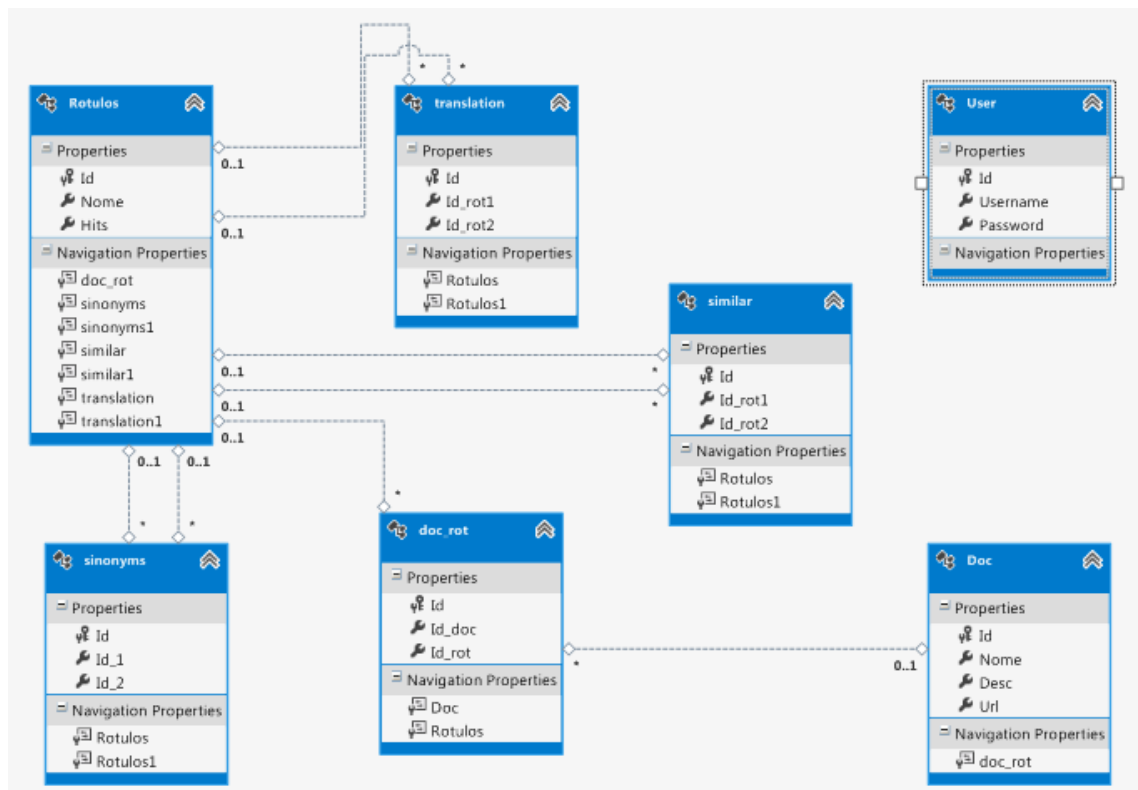


Figura 37 – Modelo de dados

As principais tabelas são as seguintes:

- **Doc** - Na tabela “Doc” são armazenados todos os documentos do sistema.
- **Doc_rot** - Na tabela “doc_rot” são armazenadas todas as relações entre *tags* e documentos, ou seja, quais as *tags* com que um determinado documento foi referenciado.
- **User** - Com a tabela “User” é armazenada a informação dos utilizadores da comunidade do Molecule. É através desta tabela que é dado o acesso à parte de administração do sistema.
- **Rotulos** - A tabela “Rotulos” armazena todas as *tags* existentes no sistema.
- **Similar** - Na tabela “similar” são armazenadas todas as relações de similaridade entre as diversas *tags* do sistema.

- Synonyms - Através da tabela “synonyms” são guardadas todas as relações de sinónimos entre *tags*.
- Translation - Na tabela “translation” ficam armazenadas as relações de tradução existentes entre as diversas *tags*.

3.5.5 Interfaces (WCF)

As diferentes modelações gráficas do sistema Molecule servem-se de um serviço *web* para carregarem os seus átomos de termos, sendo este serviço público, e portanto, podendo ser integrado em diversos sistemas e, assim, reutilizado e reaproveitado em diversas plataformas, mesmo se estas não implementam o mesmo número de “átomos” ou a mesma forma de visualização que o sistema nativo.

Esta solução de carregamento por serviço *web* permite, também, que seja virtualmente infinito o número de termos que podem ser carregados para cada átomo ou serviço, uma vez que recebe por parâmetro, em cada um dos seus métodos, o número de termos a devolver (ordenando pelos mais utilizados), ficando apenas limitado pelo bom senso do seu utilizador, dado que não existem limitações ao nível da quantidade de informação que poderá ser passada, como poderiam existir, por exemplo, se utilizada uma solução de passagem de dados por parâmetros para as animações.

3.5.5.1 Operações

Nesta secção serão descritas as várias operações disponíveis no serviço na Tabela 4, assim como o modelo do contrato de dados. Alguns extratos de código serão incluídos de forma a completar a descrição das operações.

3. Molecule – Sistema de organização e visualização de *tags*

Tabela 4 - Operações disponíveis no serviço

Operações	Descrição
GetTranslation	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno e por que valor deve procurar relações de tradução. É utilizado sempre que necessário carregar o átomo de traduções. Abaixo pode ser vista a implementação do método (Código 6).
Get Synonym	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno, e por que valor deve procurar relações de sinónimos. É utilizado sempre que necessário carregar o átomo de sinónimos.
GetSimilar	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno, e por que valor deve procurar relações de similaridade. É utilizado sempre que necessário carregar o átomo de similaridade.
GetHits	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno, e por que valor deve procurar os termos relacionados. É utilizado sempre que necessário carregar o átomo de mais utilizados. Abaixo é visível a implementação do método, que reutiliza os métodos já implementados anteriormente (Código 7).
GetHits2	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno. É utilizado sempre que necessário carregar o átomo inicial.
GetHits3	Esta operação retorna um inteiro. Recebe como parâmetro o valor pelo qual deve retornar o número de utilizações. É utilizado sempre que necessário carregar o núcleo de um átomo.
GetHits_Similar	Esta operação retorna uma lista de dados do tipo " <i>Tag</i> ". Recebe como parâmetro o número de elementos máximo da lista de retorno, e por que valor deve procurar os termos similares. É utilizado sempre que numa pesquisa não é encontrada qualquer correspondência com os dados do sistema, e é gerado em tempo real uma pesquisa de similaridade com toda a base de dados.

```
public List<Tag> GetTranslation(string value, int num)
{
    var lista = db.translation.Where(s => (s.Rotulos.Nome == value
    && s.Rotulos1.Hits>0) || (s.Rotulos1.Nome == value && s.Rotulos.Hits>0));
```

```

List<Tag> trans = new List<Tag>();

for (int i = 0; i < lista.Count(); i++)
{
    Tag t = new Tag();

    translation aux = lista.ToList()[i];

    if (aux.Rotulos.Nome == value)
    {
        t.Name = aux.Rotulos1.Nome.ToString();
        t.Count = aux.Rotulos1.Hits.Value;
        trans.Add(t);
    }
    else
    {
        t.Name = aux.Rotulos.Nome.ToString();
        t.Count = aux.Rotulos.Hits.Value;
        trans.Add(t);
    }
}

return trans.OrderByDescending(s =>
s.Count).Take(num).ToList();
}

```

Código 3 - Método GetTranslation

```

public List<Tag> GetHits(string value, int num)
{
    List<Tag> hit = GetTranslation(value, num);
    List<Tag> hit2 = GetSinonym(value, num);
    List<Tag> hit3 = GetSimilar(value, num);

    var result = hit.Concat(hit2).Concat(hit3).ToList();

    return result.OrderByDescending(s =>
s.Count).Take(num).ToList();
}

```

Código 4 - Método GetHits

Contrato de dados - A estrutura de dados “Tag” é utilizada como tipo de retorno em diversas operações do serviço. Esta é constituída pelos campos “name” e “count”, tendo como valores por defeito vazio no campo “name” e 0 no campo “count”. A estrutura é descrita no código abaixo:

```

[DataContract]
public class Tag
{
    string name = "";
    int count = 0;
}

```

3. Molecule – Sistema de organização e visualização de *tags*

```
[DataMember]
public string Name
{
    get { return name; }
    set { name = value; }
}

[DataMember]
public int Count
{
    get { return count; }
    set { count = value; }
}
}
```

Código 5 – Estrutura de dados “Tag”

3.5.6 Outros trabalhos realizados

No decorrer da elaboração e desenvolvimento deste sistema, foram desenvolvidos vários trabalhos paralelos relacionados que merecem a sua alusão. Assim, nesta secção serão descritos detalhadamente alguns destes trabalhos e referida qual a sua relevância para o resultado final do sistema.

3.5.6.1 *DataSet do Delicious*

Durante o desenvolvimento do sistema Molecule, grande parte das suas *tags* foram obtidas de um grande conjunto de dados utilizados num sistema de *bookmarking* social, (<http://delicious.com>) de Setembro de 2007 a Janeiro de 2008, e disponibilizado no âmbito de um trabalho que o analisou [Wetzker, Zimmermann, Bauckhage, 2008].

Estas foram importadas através de uma rotina desenvolvida que ia buscar, através de um ficheiro de texto, todas as *tags* e, para cada uma delas, pesquisava suas traduções, sinónimos e similares, criando todas as relações necessárias.

Atualmente, derivado de problemas técnicos, não é possível ver esses dados importados. No entanto, está previsto a importação destes de novo, embora limitado pelo número de utilizações (mensais) dos serviços que o Molecule usa para gerar informação, e que levará a que seja possível quantificar novamente, no futuro, a diferença de retorno nos resultados de pesquisa entre o Molecule e o sistema original.

3.5.6.2 *Publicação do sistema Molecule on-line*

Resultante da submissão do sistema Molecule para a conferência CISTI, deu-se a necessidade de o publicar pela primeira vez *on-line*. Para isso, foi requerido um servidor ao Departamento

de Engenharia Informática do ISEP, o qual forneceu uma máquina virtual que responde ao endereço <http://molecule.dei.isep.ipp.pt/>.

Foi necessário configurar toda a máquina, que trazia apenas o Windows Server 2008 R2, procedendo-se à instalação de diversos softwares, como o SQL Server e IIS, necessários para o bom funcionamento do sistema Molecule.

Durante o seu desenvolvimento, foram efetuadas várias publicações a cada *milestone* atingida, sendo que o sistema sofreu evoluções tanto a nível visual como de funcionalidade e melhoramentos de performance.

3.5.6.3 Artigo para a conferência CISTI

A CISTI é um evento técnico-científico anual, que visa a apresentação e a discussão de conhecimentos, novas perspetivas, experiências e inovações no domínio dos sistemas e tecnologias de informação, organizada pela AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação) e pela Universidade Nova de Lisboa [Cisti, 2013].

Assim, o sistema Molecule, não obstante de ainda estar em construção e evolução, foi submetido à conferência sob o tema “*Organizational Models and Information Systems*” sob a forma de um “*Full Paper*”. O artigo foi aceite como um “*Poster Paper*” devido à fase muito preliminar do seu desenvolvimento, mas reconhecendo o seu interesse e utilidade.

Na Figura 38 pode-se encontrar uma imagem com o *layout* da primeira versão do sistema Molecule o qual foi apresentado na conferência.

3. Molecule – Sistema de organização e visualização de *tags*

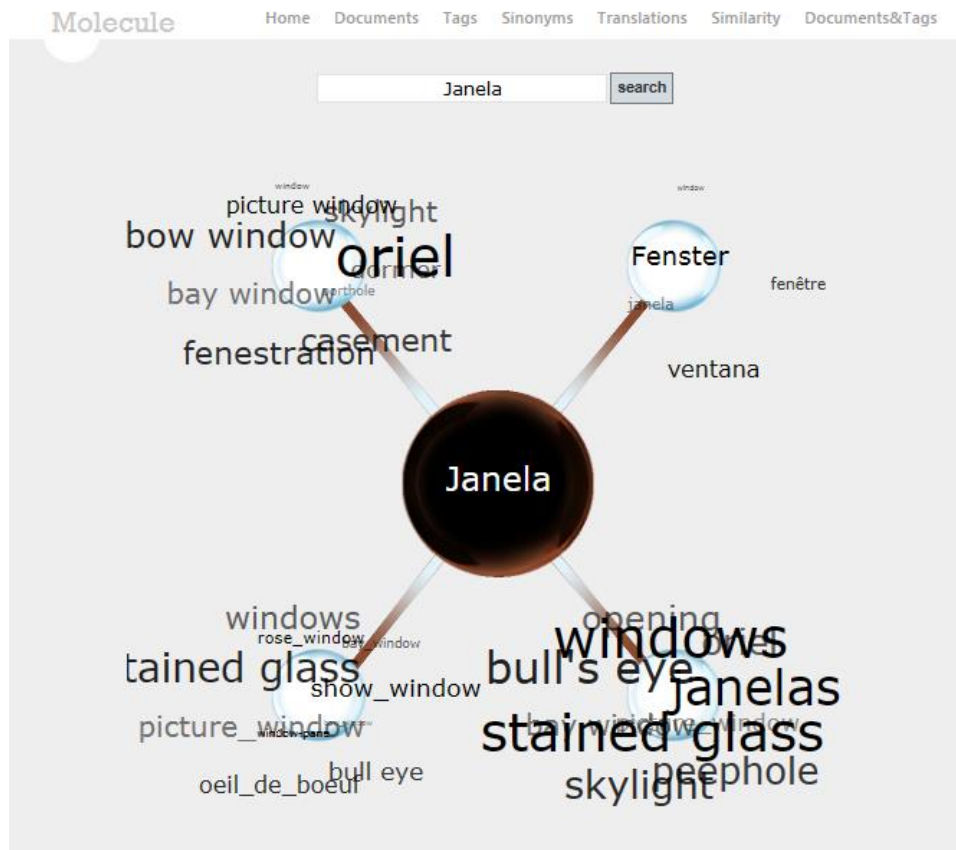


Figura 38 - Primeira versão do sistema Molecule

Dado não ter sido cumprido o objetivo, o sistema ser aceite como *Full Paper*, muito embora este o tenha sido como *Poster Paper*, o autor tomou a decisão de não manter a sua presença na conferência uma vez que esta implicaria custos associados tanto ao nível de deslocações como de inscrição.

3.5.6.4 Artigo para a revista RISTI

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é uma revista científica, publicada pela AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação), que se centra na investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação. Assim, esta revista publica semestralmente artigos originais e inovadores que são avaliados por pelo menos três membros do seu conselho científico.

Cada número da RISTI aborda uma temática específica, sendo aceites somente 5 a 8 artigos para publicação. A taxa média de aceitação é bastante apertada, situando-se abaixo dos 15%.

A RISTI é uma revista indexada em Academic Journals Database, Compendex, CrossRef, Dialnet, DOAJ, DOI, EBSCO, GALE, IndexCopernicus, Index of Information Systems Journals, Latindex, ProQuest, Scielo, SCOPUS e Ulrich's.

Assim, com o intuito de melhorar o resultado obtido junto do júri da CISTI (secção 3.5.6.3), foi submetida a evolução e nova versão do sistema Molecule sob o mesmo tema que na conferência CISTI, *“Organizational Models and Information Systems”*, para publicação na revista RISTI, número 11, de Junho de 2013. Esta versão foi aceite para publicação e a mesma saiu nessa versão da revista. Esta versão continha os mesmos *layouts* que o sistema possui atualmente, assim como autenticação e melhoramentos ao nível da eficácia de pesquisa do mesmo.

3. Molecule – Sistema de organização e visualização de *tags*

4 Conclusões

Neste capítulo serão apresentadas diversas conclusões resultantes do estudo sobre visualização de *tags*, do desenvolvimento do sistema Molecule e da elaboração deste documento.

4.1 Objectivos atingidos

No início deste trabalho, o autor propôs-se a dois grandes objetivos:

- Estudar as formas de visualização de *tags* existentes e em utilização em vários sistemas diferentes, compreendê-las e contextualizá-las, entendendo as suas vantagens e caracterizando as suas desvantagens;
- Propor uma nova forma de visualização de *tags* que introduza melhoramentos às já existentes e desenvolver um protótipo.

Desta forma, para poder cumprir o primeiro objetivo, o autor teve que efetuar uma extensa pesquisa, que o levou a estudar largas dezenas de artigos científicos e efetuar pesquisas em várias plataformas *on-line*. Esse trabalho resultou na documentação de mais uma dezena de sistemas que utilizam *tags*, indo além da sua pesquisa bibliográfica, inscrevendo-se e utilizando estes sistemas, de forma a poder melhor compreendê-los. Alguns desses sistemas acabaram por não ser incluídos neste trabalho como é o caso do Youtube, LiveJournal, eBay, Connotea, entre outros. A razão pela sua não inclusão neste documento, deve-se ao facto de que, embora estes utilizem *tags*, a sua presença e relevância não é considerada importante, levando a que não exista, por exemplo, nenhuma forma de visualização nos mesmos para a navegação do seu *tag space*. Do mesmo modo, muitas formas de visualização de *tags* não foram consideradas, embora estudadas, uma vez que podiam por um lado ser apenas uma variação de alguma já descrita, como o caso do Extispicious (*tag cloud* circular para *tags* do Delicious) ou, por outro lado, por não se inserirem dentro do contexto como é o caso do Emosi Sosial, que embora formasse uma *tag cloud*, esta era referente aos termos que os utilizadores usavam na plataforma eBay nos seus comentários e, portanto, não eram *tags* como as caracterizadas neste estudo.

Como resultado deste estudo, formou-se uma visão clara de algumas das limitações dos sistemas e formas de visualização atuais. Isto levou a que o autor pudesse estudar como colmatar alguns desses problemas e criar uma forma de visualização diferente, que acabou por se materializar no sistema Molecule, levando-o a atingir o segundo objetivo a que se tinha proposto.

4. Conclusões

4.1.1 Inovações introduzidas pelo sistema Molecule

O sistema Molecule, disponível em <http://molecule.dei.isep.ipp.pt/>, vem introduzir diversas inovações à comum nuvem de *tags*, começando pela sua forma de visualização da informação. Assim, o sistema considera quatro perspetivas diferentes para apresentar a informação pretendida, ao invés de uma única, introduzindo múltiplas vistas e ficando cada uma visível num átomo diferente:

- Tradução – As traduções do termo em diferentes línguas;
- Sinónimos – Os diferentes sinónimos do termo na língua inglesa;
- Similares – Os diferentes termos similares ao termo em questão;
- Mais utilizados – Esta última é a forma normalmente vista neste tipo de sistemas que reúne os termos mais utilizados, com a particularidade de estes estarem relacionados com o termo pesquisado.

Desta forma, é dada ao utilizador uma visão mais completa sobre todos os recursos relacionados com o objeto da sua pesquisa, uma vez que facilita a navegação entre conceitos interligados. Por outro lado, permite também ao utilizador aumentar o sucesso da sua pesquisa, uma vez que o resultado da mesma é mais abrangente e completo e combate a dispersão de resultados pelo *tag space*.

Outra novidade em relação aos tradicionais sistemas de *tags* é a sua capacidade de conexão com outras plataformas, o que vem permitir que os seus serviços sejam consumidos por terceiros caracterizando-se, assim, como um sistema útil e de partilha de informação.

De destacar, ainda, a geração automática de informação relacionada que o distingue de todos os outros sistemas e que permite, por um lado, reduzir drasticamente o fator erro humano no momento da adição de novos termos, assim como criar conexões entre a informação, de forma a unificá-la e tornar as pesquisas no sistema mais eficazes, permitindo ao utilizador ter uma visão mais completa dos conteúdos do sistema.

4.2 Limitações e trabalho futuro

Apesar dos objetivos deste trabalho terem sido atingidos com sucesso, existem ainda algumas limitações e lugar para futuras melhorias. Estão previstos diversos melhoramentos ao sistema, por forma a tornar o mesmo ainda mais eficaz, mais partilhável, mais completo e mais seguro, eliminando assim algumas limitações existentes.

Assim, ao nível da performance, poder-se-á melhorar a rapidez do sistema na inserção e pesquisa, através da introdução de vários *workers* em *background*. Desta forma, irá obter-se uma diminuição no tempo de resposta do sistema na execução dessas tarefas, nomeadamente na inserção de novas *tags*, não ficando estas pendentes das respostas dos serviços de sinónimos e tradução. O próprio mecanismo de similaridade também iria beneficiar destas alterações.

Ao nível da segurança, pode ser adicionado ao sistema uma infraestrutura de partilha de chaves pública, de forma a que a troca de informação entre o sistema Molecule e sistemas que usem os seus serviços *web* seja confidencial, autenticada e não repudiável, ao contrário do que acontece atualmente.

Poderá ser também revisto, ao nível da usabilidade, a interação do utilizador com a molécula em termos do movimento da mesma, de forma a aperfeiçoar ainda mais este mecanismo.

Por fim referir uma característica que em determinado contexto pode ser vista como uma limitação já referida na secção 3.3. Dado o sistema molecule não permitir a contextualização dos termos, é necessário dependendo da situação, ter que se utilizar outras *tags* para caracterizar um item. Por exemplo, se as outras *tags* fossem "dinheiro" e "crédito", o significado de banco seria o de instituição financeira.

4.3 Apreciação Final

Através das *tags* que atribuem a diferentes recursos, os utilizadores apresentam a sua própria visão, com a utilização de uma linguagem que não é considerada uniforme, pois a própria cultura e o nível de conhecimento variam de utilizador para utilizador. Esta forma de indexação mais alargada reflete os termos geralmente aplicados pelos utilizadores dos recursos e não das pessoas que os criaram ou disponibilizaram. Isto pode ser visto como uma desvantagem mas, simultaneamente, permite que as pessoas possam beneficiar dessa variação semântica e linguística. No entanto, e para que isso seja possível, é necessário que os sistemas de visualização de informação adotados permitam capturar essa diversidade, sem sobrecarga cognitiva dos seus utilizadores.

Esta tese de Mestrado vem colmatar diversos dos problemas característicos dos sistemas de *tags*, nomeadamente da nuvem de *tags*, a forma de visualização normalmente mais utilizada, sugerindo uma solução inovadora para a pesquisa e visualização de *tags*.

Este trabalho foi também um desafio a nível pessoal. Foram adquiridos conhecimentos em áreas que se apresentavam desconhecidas para o autor (tanto técnicas como conceptuais) até então. Adicionalmente a esses, por si só, já grandes motivos, a conciliação da realização deste trabalho académico com a vida profissional acrescentou ainda mais dificuldades. Juntando a tudo isto, a submissão do trabalho tanto em conferência como em revista veio trazer um esforço extra significativo.

4. Conclusões

Não obstante todas as dificuldades, foi sentida uma enorme gratificação pela possibilidade de evoluir nas mais diferentes áreas e o vencer os desafios constantes impostos pela natureza inovadora deste trabalho. Para além disso, a aceitação do sistema tanto em conferência como em revista foi a confirmação, por um lado do potencial do sistema, mas também a satisfação pelo reconhecimento de todo o trabalho desenvolvido.

Referências

- [airtightinteractive, 2005] Airtightinteractive, http://www.airtightinteractive.com/projects/related_tag_browser/app/ [último acesso: Out 2012].
- [Albrecht, 2006] Albrecht C., Folksonomy, Master's thesis, Vienna University of Technology, Vienna, Austria, 2006.
- [Amazon, 1996] Amazon, About Tags, <http://www.amazon.com/gp/help/customer/display.html?nodeId=16238571> [último acesso: Nov 2012]
- [Andrews, 2009] Andrews R., paidContent.org - LiveJournal Lays Off San Francisco Staff, Will Operate From Moscow, <http://www.washingtonpost.com/wp-dyn/content/article/2009/01/06/AR2009010601216.html> [último acesso: Set 2013].
- [Baldonado, Woodruff, Kuchinsky, 2000] Baldonado M., Woodruff A., Kuchinsky A., Guidelines for using multiple views in information visualization (pp. 110–119), Proceedings of the working conference on Advanced visual interfaces, 2000.
- [Battelle, 2005] Battelle J., More on Yahoo's MyWeb2.0, 2005, http://battellemedia.com/archives/2005/06/more_on_yahoos_myweb20.php [último acesso: Out 2012].
- [Cashmore, 2009] Cashmore P., Yahoo MyWeb Shuts Down, Becomes Delicious, <http://mashable.com/2009/02/14/yahoo-myweb-shuts-down/> [último acesso: Set 2013].
- [Cattuto, Loreto, Pietronero, 2006] Cattuto C., Loreto V. e Pietronero L., Collaborative tagging and semiotic dynamics, <http://arxiv.org/pdf/cs/0605015.pdf>, 2006.[último acesso: Out 2013]
- [Choi, 2010] Choi J., TagNavi – Interactive Tool for Tag Space Navigation, 2010.
- [Cisti, 2013] Cisti, <http://www.aisti.eu/cisti2013/>, [último acesso: Set 2013].
- [Citeulike, 2009] Citeulike, Group: CiteULike-discussion - Forum Thread, 2009, <http://www.citeulike.org/groupforum/1784> [último acesso: Out 2012].
- [CiteULike, 2013] CiteULike, Frequently Asked Questions, <http://www.citeulike.org/faq/faq.adp> [último acesso: Set 2013].
- [delicious, 2003] Delicious, Help, <http://delicious.com/help> [último acesso: Nov 2012].
- [Don Turnbull, 2006] Turnbull D., Taglines – tag visualization from Yahoo! research , <http://tagsonomy.com/index.php/taglines-tag-visualization-from-yahoo-research/> [último acesso: Out 2012].
- [Elhussein, 2012] Elhussein M., A Descriptive Model for Arabic-English Cross-Lingual Tagging, University Reading, 2012.
- [Flickr, 2002] Flickr, Explorar / Tags / jaguar / clusters, <http://www.flickr.com/photos/tags/jaguar/clusters/> [último acesso: Out 2012].
- [Flickr, 2005] Flickr, Ajuda / Perguntas frequentes / Públicos/Privados, <http://www.flickr.com/help/privacy/> [último acesso: Set 2013].
- [Friedman, 2007] Friedman V., Tag Clouds Gallery: Examples And Good Practices, <http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices/> [último acesso: Out 2012].

- [Fujimura, Fujimura, Matsubayashi, Yamada, Okuda, 2008] Fujimura K., Fujimura S., Matsubayashi T., Yamada T. e Okuda H., Topigraphy: Visualization for Large-scale Tag Clouds, Proceedings of the 17th international conference on World Wide Web, 2008.
- [Greenall, 2012] Greenall R., LiveJournal: Russia's unlikely internet giant, <http://www.bbc.co.uk/news/magazine-17177053> [último acesso: Set 2013].
- [Hammond, Hannay, Lund, Scott, 2005] Hammond T., Hannay T., Lund B. e Scott J., Social Bookmarking Tools (I): A General Review., D-lib Magazine 2.4, 2005.
- [Hess, Maass, Dierick, 2008] Hess A., Maass C., Dierick, F., Web 2.0 to Semantic Web: A Semi-Automated Approach, CISWeb 2008 conference, 2008.
- [Kawamoto, 2005] Kawamoto D., Yahoo acquires Upcoming.org, http://news.cnet.com/Yahoo-acquires-Upcoming.org/2100-1026_3-5889189.html [último acesso: Set 2013].
- [Kipp, 2007] Kipp M., @toread and Cool: Tagging for Time, Task and Emotion, 8th Information Architecture Summit, 2007.
- [Levenshtein, 1965] Levenshtein V., "Binary Codes for Correcting Deletions, Insertions, and Reversals," Doklady Akademii Nauk SSSR, vol. 163, no. 4, pp. 845–848, 1965.
- [macoratti, 2008] macoratti, http://www.macoratti.net/08/06/asp_mvc1.htm, [último acesso: Jul 2013].
- [macoratti, 2009a] macoratti, http://www.macoratti.net/09/09/net_wcf1.htm, [último acesso: Jul 2013].
- [macoratti, 2009b] macoratti, http://www.macoratti.net/09/06/vb_iaef.htm, [último acesso: Jul 2013].
- [Marlow, Naaman, Boyd, Davis, 2006] Marlow C., Naaman M., Boyd D. e Davis M., HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read. Proceedings of the seventeenth conference on Hypertext and hypermedia, 2006.
- [msdn, 2012] msdn, <http://msdn.microsoft.com/en-us/library/vstudio/bb397897.aspx>, [último acesso: Jul 2013].
- [msdn, 2013] msdn, [http://msdn.microsoft.com/en-us/library/cc838158\(v=vs.95\).aspx](http://msdn.microsoft.com/en-us/library/cc838158(v=vs.95).aspx), [último acesso: Jul 2013].
- [Nakamura, Endo, Morita, Morinishi, 2008] Nakamura C., Endo E., Morita L., Morinishi V., 4.5 Taxonomia x Folksonomia, <http://webfx2.blogspot.pt/2008/11/web-10-x-web-20-taxonomia-x-folksonomia.html>, [último acesso: 2013].
- [NBC, 2006] NBC, Google buys YouTube for \$1.65 billion, http://www.nbcnews.com/id/15196982/ns/business-us_business/t/google-buys-youtube-billion/#.UbECrBG9KSM [último acesso: Set 2013].
- [ncbi, 2008] ncbi, Connotea, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2268226/> [último acesso: Out 2012].
- [oficinadanet, 2007] oficinadanet, C# (CSharp) o que é está linguagem?, http://www.oficinadanet.com.br/artigo/526/c_sharp_csharp_o_que_e_esta_linguagem, [último acesso: Jul 2013].
- [Pink, 2005] Pink D., Folksonomy, http://www.nytimes.com/2005/12/11/magazine/11ideas1-21.html?_r=0, [último acesso: Jul 2013].
- [Revealicious, 2005] Revealicious, <http://ivy.fr/revealicious/demo/spacenav.html> [último acesso: Out 2012].
- [Risti, 2013] Risti, http://www.aisti.eu/index.php?option=com_content&task=view&id=20, [último acesso: Set 2013].
- [Russell, 2008] Russell T., Cloudalicious – Watching Tag Clouds Over Time, <http://cloudalicio.us/> [último acesso: Out 2012].

- [Seifert, Kump, Kienreich, Granitzer, 2008] Seifert C., Kump B., Kienreich W., Granitzer G. e Granitzer M., On the beauty and usability of tag clouds, Information Visualisation 12th International Conference, 2008.
- [seroundtable, 2012] Seroundtable, YouTube Hides Tags Due To Abuse , <http://www.seroundtable.com/youtube-hide-tags-15588.html> [último acesso: Out 2012].
- [smilehyun, 2009] smilehyun.com, FLICKR TAG VISUALIZATION, <http://smilehyun.com/flickr-tag-visualization/> [último acesso: Out 2012].
- [Technorati, 2008] Technorati, About Us, <http://web.archive.org/web/20080504092447/http://www.technorati.com/about/> [último acesso: Set 2013].
- [Wal, 2005] Wal T., Folksonomy Definition and Wikipedia, <http://www.vanderwal.net/random/entrysel.php?blog=1750> [último acesso: Out 2012].
- [Wal, 2007] Wal T., Folksonomy Definition and Wikipedia, <http://vanderwal.net/folksonomy.html> [último acesso: Nov 2012].
- [Ware, 2004] Ware C., Information Visualization: Perception for Design. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 2004.
- [Weller, Peters, 2008] Weller K. e Peters I., Seeding, Weeding, Fertilizing. Different Tag Gardening Activities for Folksonomy Maintenance and Enrichment, Triple-I Conference, Proceedings of I-Semantics '08, 2008.
- [Werts, 2009] Werts M., Word cloud from Brewer's State of the City speech, <http://blogs.kansas.com/gov/files/2009/01/brewers-word-cloud1.jpg> [último acesso: Out 2012].
- [Zamora, Nistal, 2009] Zamora F. e Nistal M., Visualizing Tags as a Network of Relatedness, Frontiers in Education Conference, FIE '09. 39th IEEE, 2009.
- [Zitvogel, 2005] Zitvogel O., Delicious Soup, a tool that reveals your del.icio.us activities, <http://www.zitvogel.com/deliciousoup/screenshots.html> [último acesso: Out 2012].
- [Zollers, 2007] Zollers A., Emerging Motivations for Tagging: Expression, Performance, and Activism, WWW Workshop on Tagging and Metadata for Social Information Organization, 2007.

***Anexo A - Artigo aceite na 8ª Conferência
Ibérica de Sistemas e Tecnologias de
Informação (CISTI)***

Molecule

Sistema de organização e visualização de tags

Resumo— Diversas plataformas informáticas permitem que os utilizadores rotulem recursos e partilhem informação com outros utilizadores. Assim, foram desenvolvidas várias formas de visualização dos rótulos associados aos recursos, com o intuito de facilitar a pesquisa dos mesmos, e pelos mesmos, aos utilizadores. De entre os vários conceitos desenvolvidos, a nuvem de *tags* destaca-se como a forma mais comum de visualização destes rótulos ou *tags*. Este documento apresenta um estudo efetuado sobre as suas limitações e propõe uma forma de visualização alternativa às já comuns nuvens de *tags*. Propõe também uma nova interpretação sobre como pesquisar e visualizar *tags*, diferindo assim do método de pesquisa direta do termo na base de dados que atualmente é maioritariamente utilizado. Com esta proposta, pretende-se implementar uma solução viável e inovadora para vários dos problemas com que a tradicional nuvem de *tags* se debate.

Palavras Chave – *tag*; nuvem de *tags*; rótulos; *tag space*; *tagging*;

(<http://www.ebay.com>). Desta forma, consegue-se organizar e partilhar diversos recursos *on-line*, como livros, fotografias, vídeos, publicações de blogues, músicas, entre outros [2].

O resultado destas e outras implementações, foi a possibilidade dos utilizadores terem acesso a um repositório de recursos rotulados, também designado de *tag space*, que pode ser pesquisado e explorado de diversas formas pela comunidade de utilizadores e que resulta do conjunto de todos os rótulos de todos os utilizadores, associados aos recursos do sistema em questão.

Para possibilitar a navegação e pesquisa do *tag space*, as plataformas que proporcionam este tipo de serviço necessitam de modelar uma interface gráfica para a sua comunidade de utilizadores. Assim, essa modelação da interface normalmente toma a forma de uma nuvem de *tags* (Figura 1). Esta não é mais do que uma lista dos rótulos mais populares, normalmente organizados por ordem alfabética e com diversos tamanhos de letra e cores.

1. Introdução

Por definição, *tags* são palavras atribuídas por utilizadores, para anotar e caracterizar vários tipos de recursos na Web. Trata-se, na prática, de rotular informação, de forma simples e intuitiva, recorrendo a uma ou mais palavras-chave.

Assim, podemos dizer que permitem melhorar e facilitar o acesso a recursos, pois estes já foram previamente caracterizados e “indexados” pelo utilizador, com palavras que ele próprio associa e relaciona com os itens em questão, facilitando assim futuras pesquisas e a recuperação de informação [1].

Dadas as vantagens notórias de se colocar *tags* associadas a informação (também designado por *tagging*) para a pesquisa e recuperação da mesma, este tipo de sistemas generalizou-se e foi adotado e implementado com sucesso em diversas plataformas Web como o Delicious (<https://delicious.com>) e o Ebay



Figura 1- Nuvem de tags [3]

Embora seja a forma mais conhecida de visualização, existem muitas outras alternativas e variações para a modelação do *tag space* numa interface. Essas alternativas são também relevantes e, dependendo do contexto ou utilizador, podem revelar-se mais úteis e apropriadas.

Não obstante, uma característica comum a todas é o facto de que, ao contrário da pesquisa por palavra-chave numa caixa de texto, que requer que o utilizador formule as suas necessidades de informação, a pesquisa por *tag* permite que o utilizador reconheça a informação enquanto visualiza a mesma, dado esta ser-lhe logo apresentada sobre a forma de rótulos, não tendo assim necessidade de formular a pesquisa para poder ver os primeiros resultados.

O resto do artigo está organizado da forma que se descreve a seguir. Esta primeira secção faz uma introdução ao tema do artigo, fazendo a sua contextualização. Na próxima secção são apresentados alguns problemas associados a utilização de nuvens de *tags*. Na terceira secção algumas nuvens de *tags* que permitem a visualização das *tags* de forma mais rica. Na quarta secção descreve-se o sistema Molecule e as suas características. Por fim, na última secção apresentam-se algumas conclusões e trabalho futuro a realizar nos próximos meses.

2. Limitações conhecidas aos sistemas que utilizam nuvens de tags

Apesar das inúmeras vantagens referidas no capítulo anterior, existem também alguns problemas associados à utilização de nuvens de *tags*.

Assim, especialmente em sistemas que utilizam nuvens de *tags*, verifica-se que o utilizador é sugestionado a recorrer sempre aos mesmos termos, pois estes aparecem destacados na forma de visualização, o que origina uma inclinação para a escolha dos mesmos. Como normalmente estes sistemas baseiam a escolha da amostragem de resultados no número de vezes que os termos foram utilizados, a nuvem de *tags* irá tendencialmente mostrar sempre os mesmos rótulos, dado serem estes teoricamente os mais populares.

Assim, há uma limitação da amostragem da base de conhecimento, uma vez que a tendência será o crescente recurso às mesmas *tags* para rotular novos recursos.

Adicionalmente, existe uma restrição clara no espaço visual de amostragem e, também devido a esse facto, aumenta ainda mais o sintoma descrito no parágrafo anterior, pois existe uma redução ainda maior no número de elementos no espaço de amostragem que, por si só, já é bastante limitado e pouco ou nada variado.

Estando estes sistemas normalmente baseados na colaboração da comunidade que integra os mesmos, existem também diversas desvantagens inerentes a este facto.

Uma das desvantagens notórias pode ser vista se se pensar que cada utilizador é livre de rotular os recursos como bem entender, levando a que dois utilizadores, embora possam potencialmente estar a atribuir o mesmo significado, utilizem dois termos sinónimos, mas como não existe qualquer ligação direta entre os mesmos, irá consequentemente ocorrer uma diversificação e aumento da granularidade do *tag space*. Por conseguinte, as pesquisas tornam-se menos eficazes, pois existe uma dispersão dos recursos por termos sinónimos, mas sem ligação aparente. Assim, por exemplo, se um utilizador pesquisar pela palavra carro, terá resultados diferentes se pesquisar pela palavra automóvel. Embora ambos os rótulos tenham o mesmo significado, como estão associados a recursos diferentes, a pesquisa não irá retornar a união dos resultados, mas apenas o conjunto derivado de um deles.

O mesmo se passa se for feita uma reflexão ao nível de tradução de termos. Um determinado utilizador pode escolher rotular um recurso em inglês, enquanto um outro pode querer rotular um outro recurso com o mesmo significado, mas em português. Assim, se por exemplo pesquisarmos por carro, ou por *car*, embora ambos os termos tenham sido utilizados para rotular o mesmo tipo de recursos, não será retornada na pesquisa a união dos resultados, mas apenas um deles. Note-se que a utilização de termos em várias línguas é comum em várias plataformas mundialmente conhecidas. Assim, muitas vezes os utilizadores atribuem *tags* não apenas escritas na língua inglesa mas também na sua língua nativa. Mais uma vez assiste-se a um aumento da granularidade do *tag space*, sem que sejam apresentadas vantagens resultantes desse fenómeno.

Outro problema também comum e presente neste tipo de sistemas, que surge na linha dos dois anteriormente apresentados, é a “repetição” de termos semelhantes, ou seja, existirem dois termos que diferem apenas, por exemplo, na

conjugação entre o singular e o plural. O termo automóvel e o termo automóveis são de facto termos diferentes, embora queiram potencialmente significar o mesmo, mas como não existe qualquer espécie de mapeamento entre eles, é impossível, numa pesquisa, retornar o universo dos recursos de ambos os termos, mas apenas do pesquisado. Esta característica não se estende só a plural/singular, mas a uma infinidade de casos como erros ortográficos, ou utilização de diferentes separadores entre palavras como “*race_car*” ou “*race car*”. Nesta última, embora ambos os termos signifiquem exatamente o mesmo, por terem sido escritos de maneiras diferentes, nunca será mostrado o resultado de ambos mas, mais uma vez, apenas o conjunto de resultados para o termo pesquisado.

Assim, neste tipo de sistemas é natural assistir-se a uma dispersão dos recursos, pois cada utilizador pode rotular os termos como entender, ficando comprometida, desde logo, a eficácia de retornar posteriormente os recursos.

Acresce a estes factos ainda o problema de os recursos serem erradamente rotulados, tanto de forma propositada, como devido, por exemplo, a diferenças no significado entre o utilizador que atribui o termo, e o utilizador que o pesquisa.

A tabela que se segue, ilustra os diversos sintomas apresentados anteriormente na plataforma Ebay, mostrando a diferença de número de resultados retornados, entre os diversos termos.

TABELA I. NÚMERO DE RESULTADOS RETORNADOS NA PLATAFORMA EBAY

Termos pesquisados na plataforma ebay	Número de resultados
carro	652
automóvel	6
automóveis	2
car	3.504.866
race_car	40.085
race car	177.376

3. Trabalho relacionado

Diversas propostas de visualização têm tentado dar resposta a algumas das limitações das nuvens de tags mais comuns, nomeadamente pela visualização para cada tag de outras que comummente são associadas aos mesmos recursos (Figura 2).

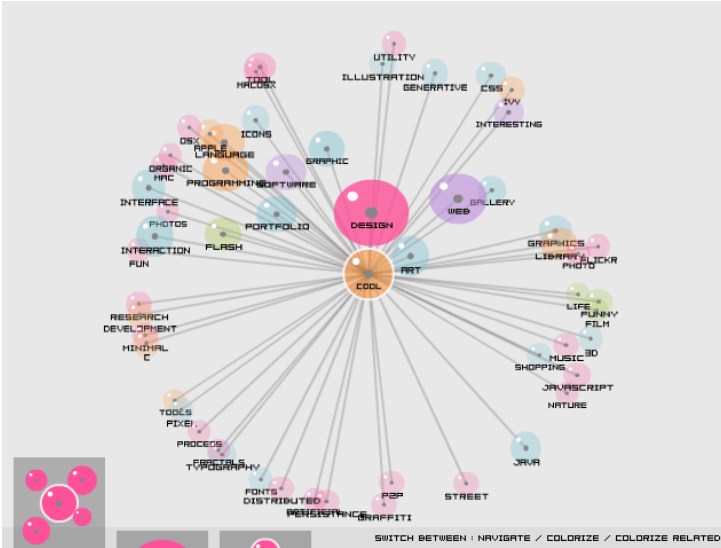


Figura 2 - Del.icio.us soup (<http://www.zitvogel.com/deliciousoup/screenshots.html>)

Outras propostas surgiram com propósito similar mas com diferente layout [4][5] como se pode observar na Figura 3 e na Figura 4, respectivamente. Note-se que na Figura 3 e na Figura 4 as tags estão ligadas por linhas que indicam tags relacionadas.

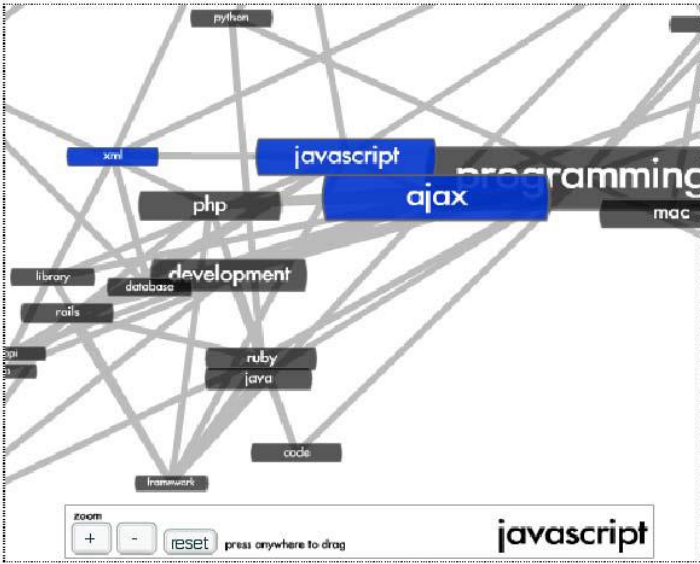


Figura 3- Tag map

Zamora e seus colegas vão além da análise de frequência de tags e de coocorrência de tags. Propõem a visualização de tags numa rede semântica que considera também relações entre tags como, por exemplo, de sinonímia [6].

O trabalho descrito neste artigo tem uma abrangência maior e pretende contornar também outras limitações das nuvens de tags e dos sistemas que utilizam tags, conforme se descreve na secção seguinte.

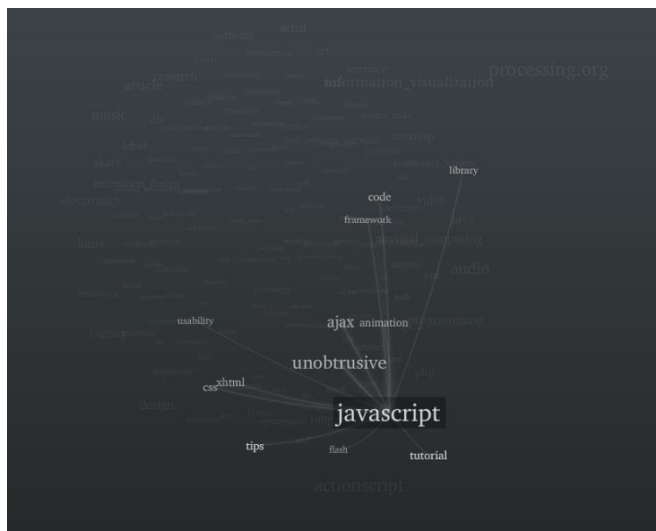


Figura 4 - Elastic tag map (obtida de http://well-formed-data.net/experiments/tag_maps_v5/)

4. Molecule – Sistema de organização e visualização

Considerados os problemas apresentados anteriormente, e com o objetivo de diminuir a dimensão dos mesmos, mas ao mesmo tempo aproveitar as potencialidades da visualização com nuvens de *tags*, nasceu o sistema Molecule, que é apresentado nesta secção.

Tal como uma molécula é composta por átomos, o sistema Molecule é, na verdade, a combinação de vários termos, permitindo a visualização de dados relacionados, formando assim uma molécula de informação conectada.

Esta aplicação, tal como a maioria deste tipo de sistemas, utiliza vários formulários para a inserção dos conteúdos como os recursos, rótulos e associação de rótulos a recursos. No entanto, contrariamente às restantes plataformas, aquando do registo de um novo rótulo, o sistema vai procurar automaticamente tanto sinónimos, como traduções, como termos similares, inserindo-os automaticamente e de forma interligada.

Não obstante, é dada a opção ao utilizador de gerir estas ligações, tanto de sinónimos, como de traduções, ou mesmo de termos similares, através de interface gráfica, caso os termos ou ligações geradas automaticamente não se revelem de alguma forma corretas.

A plataforma usa o recurso a três sistemas distintos para criar os termos e as ligações de forma automática.

Para traduções, o sistema faz pedidos ao Google Translate, para as línguas portuguesa, inglesa, alemã, espanhola e italiana. De notar que não é usada a API da Google, que recentemente deixou de ser gratuita, mas sim um pedido *http* directo, o qual é tratado na resposta, sendo inseridos apenas termos novos à plataforma. O número de línguas, embora limitado a estas cinco, poderá ser facilmente aumentado. O sistema Molecule utiliza apenas estas cinco línguas porque, para já, o único intuito é a demonstração de conceito.

Desta forma, para o termo original, assim como para cada tradução, o sistema vai fazer um pedido ao serviço da Worldreference.com, para que este retorne sinónimos para o termo em questão. Este pedido é feito à JSON API do sistema, a qual retorna uma lista de termos, que é tratada e introduzida na plataforma, criando-se assim todas as ligações entre os termos sinónimos.

Atualmente, o sistema Molecule estabelece sinónimos apenas para as línguas espanhola e inglesa. Mais uma vez, esta limitação não é impeditiva de, futuramente, a inserção de sinónimos ser alargada a outras línguas, bastando para tal, por exemplo, mudar o fornecedor do serviço.

Assim, dado a geração automática de um enorme volume de rótulos a partir de um único, associados desde logo entre eles, são resolvidas outras questões intrinsecamente ligadas à inserção manual de rótulos pelo utilizador. Com esta base de conhecimento, e não sendo permitida a duplicação de rótulos na mesma, a probabilidade do utilizador usar um rótulo previamente introduzido é muito elevada. Este comportamento leva a que todas as desvantagens provenientes da duplicação de termos como, por exemplo, por erros ortográficos, singular/plural, entre outras, fiquem virtualmente eliminadas, uma vez que o utilizador irá tendencialmente recorrer a um termo que foi previamente inserido de forma automática e devidamente mapeado.

Tal não acontece, por exemplo, na plataforma Amazon.com, que é uma loja *online* utilizada por pessoas de diferentes países e nem todas fluentes na língua inglesa, assim como em várias outras plataformas. Em Amazon.com existem 278 produtos aos quais foi atribuída a *tag* excelente (http://www.amazon.com/tag/excelent/products/ref=tag_dh_istp-tag com um erro ortográfico), 148 produtos com a *tag* excelente (http://www.amazon.com/tag/excelente/products/ref=tag_dh_istp-tag possivelmente escrita noutra língua como o português) e 5.382 produtos com a *tag* excellent (http://www.amazon.com/tag/excellent/products/ref=tag_dh_istp-tag escrita na língua portuguesa). Mas as *tags* excellent e excelente não aparecem como *tags* usualmente utilizadas como a *tag* excellent, pelo que os sistemas discutidos na secção III não conseguiriam lidar bem com situações deste tipo, para o qual o sistema Molecule foi também pensado.

Assim, no sistema Molecule, a dispersão de termos no *tag space*, e consequentemente a dispersão de recursos, por falta de relações pré estabelecidas entre *tags*, será também atenuada porque, por exemplo, embora um utilizador possa estar a usar um sinónimo ou tradução de um outro termo já referenciado, ambos serão retornados no universo da pesquisa, uma vez que o

sistema compreende que se tratam da mesma coisa, ou de assuntos relacionados.

Desta forma, os resultados obtidos na Tabela I seriam bastante diferentes e virtualmente verificar-se-ia o mesmo número de ocorrências nos termos *carro*, *automóvel*, *automóveis* e *car*.

Mesmo assim, e apenas aplicando estas medidas, continuar-se-ia a obter resultados distintos em termos semelhantes como *race car* e *race_car*. Isto aconteceria porque, por exemplo, um utilizador teria introduzido manualmente o rótulo *race_car* não utilizando o previamente introduzido *race car*, nem fazendo o mapeamento manual. Para colmatar este problema, o Molecule está preparado para efetuar comparações entre termos semelhantes.

No momento do registo de uma nova *tag*, esta, suas traduções, e sinónimos, são comparados com todas as já existentes, e se um pré determinado coeficiente de similaridade for obtido, será criada uma relação de similaridade entre a *tag*, e a *tag* comparada.

O sistema Molecule, numa pesquisa em que não seja retornada nenhuma correspondência com a base de conhecimento, permite ainda a opção de procurar em tempo real similaridades entre o termo e as *tags* já armazenadas.

Estes processos são feitos através do recurso ao algoritmo de Levenshtein. Em 1965 Vladimir Levenshtein criou um algoritmo de distâncias que calcula o número de modificações necessárias para transformar um termo num outro [7]. Através da utilização deste algoritmo, o sistema Molecule mede a similaridade entre dois termos.

Assim, em casos como os termos *race car* e *race_car*, será possível encontrar os recursos referentes a ambos, dado que é estabelecida uma relação de similaridade entre os vários termos da base de conhecimento, tanto no momento da sua introdução,

como no caso de uma pesquisa em que nenhum termo é retornado.

Desta forma, começa a tornar-se clara a escolha do termo Molecule para o nome da plataforma, uma vez que esta usa diversos sistemas distintos (átomos) que, por sua vez, se interligam e formam um único sistema (molécula).

Não obstante, existe outro motivo para a escolha do nome. Este prende-se com a interpretação visual que o sistema Molecule gera, resultante do seu *tag space*.

Ao entrar no Molecule o utilizador poderá efetuar a pesquisa dos recursos por campo de texto, ou por um “átomo” inicial carregado com os termos mais utilizados, similar à nuvem de *tags* tradicional.

Quando o utilizador realizar uma pesquisa pelo campo de texto, ou pelo átomo inicial, observará a formação de uma “molécula” caso seja encontrada correspondência entre o termo e o *tag space* do sistema, aparecendo todos os recursos associados, listados posteriormente.

Assim, a “molécula” formada tem em cada um dos seus “átomos” uma representação de cada uma das diferentes particularidades de pesquisa existentes no sistema Molecule.

O utilizador verá os termos mais utilizados ao nível de traduções num dos “átomos” (no canto superior direito), enquanto noutro “átomo” estarão os termos mais utilizados ao nível de sinónimos (no canto superior esquerdo), noutro os termos mais utilizados ao nível de similaridade (no canto inferior esquerdo), e por fim, os termos mais utilizados absolutos relacionados com o termo (no canto inferior direito), formando assim a “molécula” referente ao termo inicialmente pesquisado.

Na Figura 5, pode se ver isso mesmo, referente a uma pesquisa no sistema pelo termo “janela”, tendo sido gerada a molécula correspondente.

Para além de todas as suas mais-valias funcionais, o sistema Molecule é caracterizado, também, por uma grande atratividade para os seus utilizadores, resultante não só do seu visual inovador, derivado do seu conceito de modelação gráfica dos termos, mas também por convidar o utilizador a interagir com ele, de uma forma muito intuitiva e natural, que o cativa a usá-lo de um modo recorrente.

O sistema Molecule serve-se de um serviço web para carregar os seus átomos de termos, podendo ser integrado em diversos sistemas e, assim, reutilizado e reaproveitado em diversas plataformas, mesmo se estas não implementam o mesmo número de “átomos” ou a mesma forma de visualização que o sistema nativo.

Esta solução de carregamento por serviço web permite, também, que seja virtualmente infinito o número de termos que podem ser carregados para cada átomo, ficando apenas limitado pelo bom senso do seu utilizador, dado que não existem limitações ao nível da quantidade de informação que poderá ser passada, como poderiam existir, por exemplo, se utilizada uma solução de passagem de dados por parâmetros.

5. Conclusão e Trabalho Futuro

O trabalho apresentado neste artigo vem colmatar diversos dos problemas característicos dos sistemas de *tags*, nomeadamente da forma normalmente mais utilizada, a nuvem de *tags*, abrindo novos caminhos, e apresentando soluções inovadoras para os mesmos.

Dada a sua capacidade de conexão com outras plataformas, o sistema Molecule, vem ainda permitir que os seus serviços sejam consumidos por terceiros, caracterizando-se assim como um sistema útil e de partilha de informação.

Estão previstos diversos melhoramentos ao sistema, por forma a tornar o mesmo ainda mais eficaz, mais partilhável, mais completo e mais seguro.

Assim, ao nível da performance, será revista a rapidez de inserção e pesquisa do sistema, através da introdução de vários *workers* em *background* obtendo-se, dessa forma, um aumento na rapidez de execução dessas tarefas. Sofrerá também uma revisão a vários métodos e classes, de forma a ser feita uma optimização do código ao nível do peso que o mesmo tem na memória em execução.

Ao nível da segurança será adicionado ao sistema uma infraestrutura de partilha de chaves pública, de forma a que a troca de informação entre o sistema Molecule e sistemas que usem os seus serviços web seja confidencial, autenticada e não repudiável.

Com o intuito de melhorar a partilha de informação automática, a interface gráfica do Molecule será enriquecida com dados RDFa, de forma a permitir a outros serviços consultar informação diretamente através das suas páginas.

Por fim, tendo em mente sempre a melhoria do serviço, será planeada a interligação entre o sistema Molecule e diversos serviços que proporcionem conhecimento ontológico, levando assim à geração de um quinto átomo na molécula visual, permitindo aumentar ainda mais a eficácia da pesquisa do sistema.

4.3.1.1.1 Referências

- [1] M. Elhussein, “A Descriptive Model for Arabic-English Cross-Lingual Tagging,” presented at the The International Conference on Communications and Information Technology (ICCIT-2012), Hammamet, Tunisia, 2012.
- [2] I. Peters., *Folksonomies indexing and retrieval in Web 2.0*. De Gruyter, 2009.
- [3] V. Friedman, “Tag Clouds Gallery: Examples And Good Practices”, <http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices/>.
- [4] B. Shaw, “Utilizing Folksonomy: Similarity Metadata from the Del.icio.us System”, <http://www.metablake.com/webfolk/web-project.pdf>.
- [5] M. Stefaner, “Visual tools for the socio-semantic web,” Master’s thesis, University of Applied Sciences Potsdam, 2007.
- [6] F. Sanchez-Zamora and M. Llamas-Nistal, “Visualising tags as a network of relatedness,” presented at the The 39th ASEE/IEEE Frontiers in Education Conference, 2009.
- [7] V. Levenshtein, “Binary Codes for Correcting Deletions, Insertions, and Reversals,” *Doklady Akademii Nauk SSSR*, vol. 163, no. 4, pp. 845–848.

***Anexo B – Artigo publicado na
revista RISTI número 11 de Junho de
2013***

Molecule: Sistema de organização e visualização de Tags

José Martins ¹, Isabel Azevedo ²

Josenuno7@hotmail.com, ifp@isep.ipp.pt

¹ Instituto Superior de Engenharia do Porto, Rua Dr. António Bernardino de Almeida 431, 4200-072, Porto, Portugal

² Instituto Superior de Engenharia do Porto, Rua Dr. António Bernardino de Almeida 431, 4200-072, Porto, Portugal

DOI: 10.4304/risti.11.61-75

Resumo: Diversas plataformas permitem que os utilizadores rotulem recursos com *tags* e partilhem informação com outros utilizadores. Assim, foram desenvolvidas várias formas de visualização das *tags* associados aos recursos, com o intuito de facilitar aos utilizadores a pesquisa dos mesmos, assim como a visualização do *tag space*. De entre os vários conceitos desenvolvidos, a nuvem de *tags* destaca-se como a forma mais comum de visualização. Este documento apresenta um estudo efetuado sobre as suas limitações e propõe uma forma de visualização alternativa. Sugere-se também uma nova interpretação sobre como pesquisar e visualizar informação associada a *tags*, diferindo assim do método de pesquisa direta do termo na base de dados que atualmente é maioritariamente utilizado. Como resultado desta implementação, obteve-se uma solução viável e inovadora, o sistema Molecule, para vários dos problemas associados à tradicional nuvem de *tags*.

Palavras-chave: *tag*; nuvem de *tags*; *tag space*; similaridade; sinónimo; tradução; visualização de *tags*.

Abstract: Several platforms allow users to tag resources and share information. Over time various forms of tags visualization have been developed in order to facilitate the visualization of the tag space and the search and retrieval of resources. The tag cloud stands out as the most common form of tags visualization. This paper presents a study carried out on their limitations and proposes an alternative. It also suggests a new interpretation on how to search and view information associated with tags, thus differing from the method of direct search term in the database that is currently used mostly. As a result of this interpretation, a viable and innovative system was achieved, Molecule, that overcome some of the problems associated with the traditional tag cloud.

Key-words: tag; tag cloud; tag space, similarity; synonymous; translation; tags visualization.

1. Introdução

Por definição, *tags* são palavras atribuídas por utilizadores para anotar e caracterizar vários tipos de recursos na *Web*. Trata-se, na prática, de rotular informação de forma simples e intuitiva, recorrendo a uma ou mais palavras-chave. Assim, podemos dizer que permitem melhorar e facilitar o acesso a recursos, pois estes já foram previamente caracterizados e “indexados” pelo utilizador, com palavras que ele próprio associa e relaciona com os itens em questão, facilitando assim futuras pesquisas e a recuperação de informação (Ellhussein, 2012). O termo folksonomia refere-se a forma de indexação resultante da atribuição de *tags* pelos utilizadores aos recursos.

Dadas as vantagens notórias de se colocar *tags* associadas a informação (também designado por *tagging*) para a pesquisa e recuperação da mesma, este tipo de sistemas generalizou-se e foi adotado e implementado com sucesso em diversas plataformas *Web* como o Delicious (<https://delicious.com>) e o Ebay (<http://www.ebay.com>). Desta forma, consegue-se mais facilmente organizar e partilhar diversos recursos *on-line* como livros, fotografias, vídeos, publicações de blogs, músicas, entre outros (Peters, 2009).

O resultado destas e outras implementações foi a possibilidade dos utilizadores terem acesso a um repositório de recursos rotulados, também designado de *tag space*, que pode ser pesquisado e explorado de diversas formas pela comunidade de utilizadores e que resulta do conjunto de todas as *tags* de todos os utilizadores, associados aos recursos do sistema em questão.

Para possibilitar a navegação e pesquisa no *tag space*, as plataformas que proporcionam este tipo de serviço necessitam de modelar uma interface gráfica para a sua comunidade de utilizadores. A visualização de informação envolve a utilização de representações visuais abstratas para melhorar a cognição (Ware, 2004). Interfaces visuais eficazes permitem a interação com grandes volumes de dados de forma rápida e eficaz, potenciando a descoberta de características que podem não ser tão evidentes numa apresentação tabular, por exemplo.

Assim, começaram a surgir formas de visualização de *tags* que permitem aos utilizadores rapidamente verificarem como as *tags* são utilizadas numa determinada plataforma. Em algumas, o número de *tags* é elevado, o que inviabiliza a apresentação de todas as *tags* que foram atribuídas por todos os utilizadores.

De entre várias formas de visualização de *tags* surgidas, destaca-se a nuvem de *tags* como a mais comum. Esta não é mais do que uma lista dos rótulos mais populares, normalmente organizados por ordem alfabética e com diversos tamanhos de letra e cores para evidenciar quais as *tags* mais populares. Embora seja a forma mais conhecida de visualização, existem muitas outras alternativas e variações para a modelação do *tag space* numa interface. Essas alternativas são também relevantes e, dependendo do contexto ou utilizador, podem revelar-se mais úteis e apropriadas.

Não obstante, uma característica comum a todas é o facto de que, ao contrário da pesquisa por palavra-chave numa caixa de texto, que requer que o utilizador formule as suas necessidades de informação, a pesquisa por *tag* numa nuvem de *tags* permite que o utilizador reconheça a informação enquanto visualiza a mesma, dado esta ser-lhe logo apresentada sobre a forma de rótulos, não tendo assim necessidade de formular a pesquisa para poder ver os primeiros resultados.

Assim, propõe-se neste artigo uma forma alternativa de visualização de *tags* integrada no sistema Molecule. São utilizadas múltiplas vistas, o que é particularmente indicado quando uma única seria demasiado complexa e cognitivamente exigente para os utilizadores. Baldonado et al. propuseram, entre outras, a regra da diversidade para o uso de várias vistas (Baldonado, Woodruff, & Kuchinsky, 2000). Segundo esta regra deve-se utilizar mais do que uma vista quando existem múltiplos atributos, perfis, modelos ou níveis de abstração. Nestas situações uma única visão dos dados pode requerer demasiado esforço de assimilação para a compreensão da multiplicidade dos dados.

Após esta primeira secção onde se faz uma introdução e contextualização ao tema, segue uma secção onde são apresentados alguns problemas associados a utilização de nuvens de *tags*. Na terceira secção apresentam-se algumas nuvens de *tags* que permitem a visualização das *tags* de forma mais rica. Na quarta secção descreve-se o sistema Molecule e as suas características, finalizando-se com a apresentação de algumas

conclusões e trabalho futuro a realizar nos próximos meses.

2. Limitações Conhecidas aos Sistemas que Utilizam Nuvens de Tags

Apesar das inúmeras vantagens referidas no capítulo anterior, existem também alguns problemas associados à utilização de nuvens de *tags*. Em sistemas que utilizam nuvens de *tags*, verifica-se que o utilizador é sugestionado a pesquisar sempre os mesmos termos, pois estes aparecem destacados na forma de visualização, o que origina uma inclinação para a sua escolha. Como normalmente estes sistemas baseiam a escolha da amostragem de resultados no número de vezes que os termos foram utilizados, a nuvem de *tags* irá tendencialmente mostrar sempre os mesmos resultados, dado serem estes teoricamente os mais populares.

Assim, há uma limitação da amostragem do *tag space*, uma vez que apenas são mostradas as *tags* mais utilizadas, que poderão ou não, ser as mais indicadas para a pesquisa que um utilizador pretende.

Adicionalmente, existe uma restrição clara no espaço visual de amostragem e, também devido a esse facto, a característica descrita no parágrafo anterior tem outra relevância, pois existe uma redução ainda maior no número de elementos no espaço de amostragem que, por si só, já é bastante limitado e pouco ou nada variado. Estando estes sistemas normalmente baseados na colaboração da comunidade que integra os mesmos, existem também diversas desvantagens inerentes a este facto.

Uma das mais notórias pode ser entendida se se pensar que cada utilizador é livre de rotular os recursos como bem entender, levando a que dois utilizadores, embora possam potencialmente estar a atribuir o mesmo significado, utilizem dois termos sinónimos, mas como não existe qualquer ligação direta entre os mesmos, irá consequentemente ocorrer uma diversificação e aumento da granularidade do *tag space*. Por conseguinte, as pesquisas tornam-se menos eficazes, pois existe uma dispersão dos recursos por termos sinónimos, mas sem ligação aparente. Assim, por exemplo, se um utilizador pesquisar pela palavra carro, terá resultados diferentes se pesquisar pela palavra automóvel. Embora ambos os rótulos tenham o mesmo significado, como estão associados a recursos diferentes, a pesquisa não irá retornar a união dos resultados, mas apenas o conjunto derivado de um deles.

O mesmo se passa se for feita uma reflexão ao nível de tradução de termos. Um determinado utilizador pode escolher rotular um recurso em inglês, enquanto um outro pode querer rotular um outro recurso com o mesmo significado, mas em português. Assim, se por exemplo pesquisarmos por carro, ou por car, embora ambos os termos tenham sido utilizados para rotular o mesmo tipo de recursos, não será retornada na pesquisa a união dos resultados, mas apenas um deles. Note-se que a utilização de termos em várias línguas é comum em várias plataformas mundialmente conhecidas. Assim, muitas vezes os utilizadores atribuem *tags* não apenas escritas na língua inglesa mas também na sua língua nativa. Mais uma vez assiste-se a um aumento da granularidade do *tag space*, sem que sejam apresentadas vantagens resultantes desse fenómeno.

Outro problema também comum e presente neste tipo de sistemas, que surge na linha dos dois anteriormente apresentados, é a “repetição” de termos semelhantes, ou seja, existirem dois termos que diferem apenas, por exemplo, na conjugação entre o singular e o plural. O termo automóvel e o termo automóveis são de facto termos diferentes, embora queiram potencialmente significar o mesmo, mas como não existe qualquer espécie de mapeamento entre eles, é impossível, numa pesquisa, retornar o universo dos recursos de ambos os termos, mas apenas do pesquisado. Esta característica não se estende só a plural/singular, mas a uma infinidade de casos como erros ortográficos. Note-se que a propensão para existência de erros em *tags* é mais elevada que nos sistemas de indexação tradicionais (Hammond, Hannay, Lund, & Scott, 2005). Também é comum a utilização de diferentes separadores entre palavras como “race_car” ou “race car”. Neste último exemplo, embora ambos os termos signifiquem exatamente o mesmo, por terem sido escritos de maneiras diferentes, nunca será mostrado o resultado de ambos mas, mais uma vez, apenas o conjunto de resultados para o termo pesquisado.

A expressão *tag gardening* refere-se ao reajuste das *tags*, tal como a jardinagem permite reajustar plantas. Uma das atividades do processo é justamente lidar com as variantes

ortográficas das *tags* (Weller & Peters, 2008). Um sistema de *tagging* pode ser melhorado através das considerações das variações de escrita (Hess, Maass, & Dierick, 2008).

Assim, neste tipo de sistemas é natural assistir-se a uma dispersão dos recursos, pois cada utilizador pode rotular os mesmos como entender, ficando comprometida, desde logo, a eficácia de retornar posteriormente os recursos. Acresce a estes factos ainda o problema de os recursos serem erradamente rotulados, tanto de forma propositada, como devido, por exemplo, a diferenças no significado entre o utilizador que atribui o termo e o utilizador que o pesquisa.

A Tabela 1 ilustra os problemas apresentados anteriormente na plataforma Ebay, mostrando a diferença de número de resultados retornados, entre os diversos termos.

Tabela 1 – Número de resultados retornados na plataforma Ebay

Termos pesquisados na plataforma Ebay	Número de resultados
carro	652
automóvel	6
automóveis	2
car	3.504.866
race_car	40.085
race car	177.376

3. Trabalho Relacionado

Diversas propostas de visualização têm tentado dar resposta a algumas das limitações das nuvens de *tags* mais comuns, nomeadamente pela visualização para cada *tag* de outras que comumente são associadas aos mesmos recursos (Figura 1).

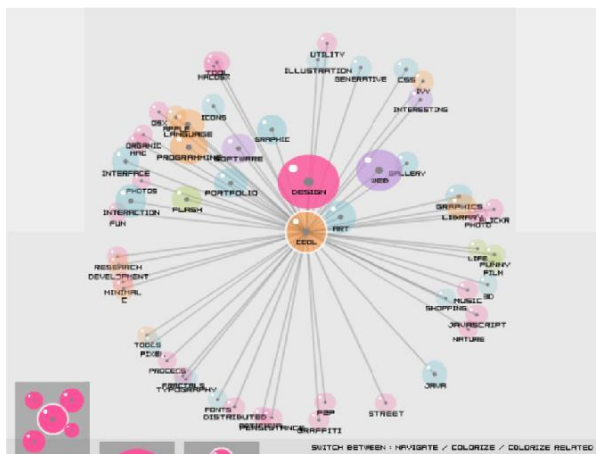


Figura 1 – Delicio.us soup (<http://www.zitvogel.com/delicioussoup/screenshots.html>)

Outras propostas surgiram com propósito similar mas com diferente *layout* (Shaw, n.d.) (Stephaner, 2007) como se pode observar na Figura 2, respectivamente. Note-se que nestas figuras as *tags* estão ligadas por linhas que indicam *tags* relacionadas.

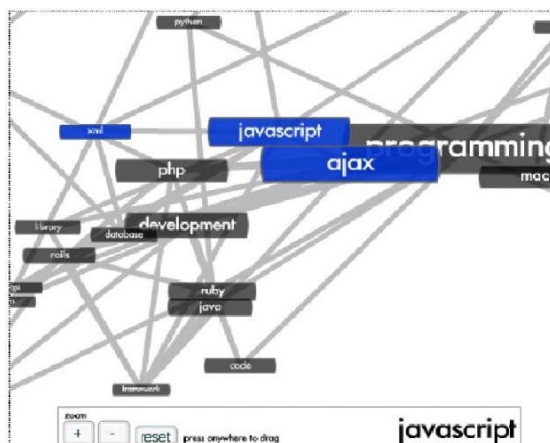


Figura 2 – Tag map

Sanchez-Zamora e seus colegas vão além da análise de frequência de *tags* e de coocorrência de *tags*. Propõem a visualização de *tags* numa rede semântica que considera também relações entre *tags* como, por exemplo, a sinonímia (Sanchez-Zamora & Llamas-Nistal, 2009).

O trabalho descrito neste artigo tem uma abrangência maior e pretende contornar também outras limitações das nuvens de *tags* e dos sistemas que utilizam *tags*, conforme se descreve na secção seguinte.

4. Molecule – Sistema de Organização e Visualização

Considerados os problemas apresentados anteriormente, e com o objetivo de diminuir a dimensão dos mesmos, mas ao mesmo tempo aproveitar as potencialidades da visualização com nuvens de *tags*, nasceu o sistema Molecule, que é apresentado nesta secção.

Tal como uma molécula é composta por átomos, e os átomos compostos por electrões, o sistema Molecule é, na verdade, a combinação de vários termos, permitindo a visualização de dados relacionados, formando assim uma molécula de informação conectada.

O sistema baseia-se em tecnologia *Web* de forma a potenciar a facilidade de acesso, e está acessível no endereço <http://molecule.dei.isep.ipp.pt/>.

Esta aplicação, tal como a maioria deste tipo de sistemas, utiliza vários formulários para a inserção dos conteúdos como os recursos, rótulos e associação de rótulos a recursos. No entanto, contrariamente às restantes plataformas, aquando do registo de um novo rótulo, o sistema vai procurar automaticamente tanto sinónimos, como traduções, como termos similares, inserindo-os automaticamente e de forma interligada.

Não obstante, é dada a opção ao utilizador de gerir estas ligações, através de interface gráfica, caso os termos ou ligações geradas automaticamente não se revelem de alguma forma corretas, ou queira criar novas associações, como pode ser visto na Figura 3.

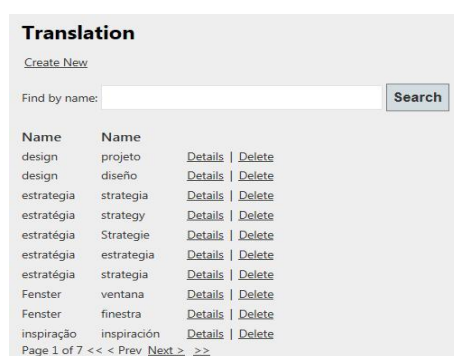


Figura 3 – Menu de gestão de traduções

A plataforma usa o recurso a três sistemas distintos para criar os termos e as ligações de forma automática. Para traduções o sistema faz pedidos ao serviço Google Translate, para obter traduções de um dado termo nas línguas portuguesa, inglesa, alemã, espanhola e italiana.

De notar que não é usada a API do Google Translate, que recentemente deixou de ser gratuita, mas sim um pedido http directo, o qual é tratado na resposta, sendo inseridos na plataforma Molecule apenas termos e ligações entre termos novos. O número de

línguas, embora limitado a estas cinco, poderá ser facilmente aumentado. O sistema Molecule utiliza apenas estas cinco línguas porque, para já, o único intuito é a demonstração de conceito.

Desta forma, para o termo original, assim como para cada tradução, o sistema vai fazer um pedido ao serviço Big Huge Thesaurus (disponível em <http://words.bighugelabs.com>), para que este retorne sinónimos para o termo em questão. Note-se que este serviço baseia-se no serviço WordNet (Miller, 1995), mas considera também outras fontes de informação. O pedido é submetido através da API JSON do sistema, que retorna uma lista de termos, que é tratada e introduzida na plataforma, criando-se assim todas as ligações entre os termos sinónimos. Atualmente o sistema Molecule estabelece sinónimos apenas para a língua inglesa. Mais uma vez, esta limitação não é impeditiva de, futuramente, a inserção de sinónimos ser alargada a outras línguas, bastando para tal, conectá-lo a outros serviços que permitam a aquisição de informação para outras línguas.

Assim, dada a geração automática de um enorme volume de *tags* a partir de uma única, associadas desde logo entre elas, são resolvidas outras questões intrinsecamente ligadas à inserção manual de *tags* pelo utilizador. Com este *tag space*, e não sendo permitida a duplicação de *tags* no mesmo, a probabilidade do utilizador usar uma *tag* previamente introduzida é muito elevada. Este comportamento leva a que todas as desvantagens provenientes da duplicação de termos como, por exemplo, por erros ortográficos, singular/plural, entre outras, fiquem virtualmente eliminados, uma vez que o utilizador irá tendencialmente recorrer a uma *tag* que foi previamente inserida de forma automática e devidamente mapeada, para rotular um determinado recurso.

Tal não acontece, por exemplo, na plataforma Amazon.com, que é uma loja *online* utilizada por pessoas de diferentes países e nem todas fluentes na língua inglesa, assim como em várias outras plataformas. Em Amazon.com existem 278 produtos aos quais foi atribuída a *tag* “excelent”¹ (*tag* com um erro ortográfico), 148 produtos com a *tag* “excelente”² (*tag* possivelmente escrita noutra língua como o português) e 5.382 produtos com a *tag* “excellent”³ (*tag* escrita na língua inglesa). Mas as *tags* “excelent” e “excelente” não aparecem como *tags* usualmente utilizadas conjuntamente com a *tag* “excellent”, pelo que os sistemas discutidos na secção III não conseguiriam lidar bem com situações deste tipo, para o qual o sistema Molecule foi também pensado.

Assim, no sistema Molecule, a dispersão de termos no *tag space*, e consequentemente a dispersão de recursos, por falta de relações pré estabelecidas entre *tags*, é também atenuada porque, por exemplo, embora um utilizador possa estar a usar um sinónimo ou tradução de um outro termo já referenciado, ambos serão retornados no universo da pesquisa, uma vez que o sistema compreende que se tratam da mesma coisa, ou de assuntos relacionados. Desta forma, os resultados obtidos na Tabela 1 seriam bastante diferentes e virtualmente verificar-se-ia o mesmo número de ocorrências nos termos “carro”, “automóvel”, “automóveis” e “car”.

Mesmo assim, e apenas aplicando estas medidas, continuar-se-ia a obter resultados distintos em termos semelhantes como “race car” e “race_car”. Isto acontece porque, por exemplo, um utilizador teria introduzido manualmente o rótulo “race_car” não utilizando o previamente introduzido “race car”, nem fazendo o mapeamento manual. Para colmatar este problema, o Molecule está preparado para efetuar comparações entre termos semelhantes.

No momento do registo de uma nova *tag*, esta, suas traduções, e sinónimos, são comparados com todas as já existentes, e se um pré determinado coeficiente de similaridade for obtido, será criada uma relação de similaridade entre a *tag*, e a *tag* comparada.

Acresce que o sistema Molecule, numa pesquisa em que não seja retornada nenhuma correspondência com o *tag space* ou recursos, efectua uma pesquisa em tempo real, procurando encontrar similaridades entre o termo e as *tags* já armazenadas, retornando os resultados similares num único átomo.

Estes processos são feitos através do recurso ao algoritmo de Levenshtein. Em 1965 Vladimir Levenshtein criou um algoritmo de cálculo de distâncias entre dois termos que

¹ http://www.amazon.com/tag/excelent/products/ref=tag_dh_istp

² http://www.amazon.com/tag/excelente/products/ref=tag_dh_istp

Figura 5 – Resultado gerado pelo Molecule para o termo “strategy”

Por forma a evitar que o número de resultados retornado seja muito grande, dado o mesmo ser resultante da recolha de várias fontes de conhecimento, é possível ao utilizador ainda filtrar que tipo de resultados quer ver apresentados, aquando da pesquisa (Figura 6). Se todos (*ALL*), ou seja, se o conjunto de resultados do termo, dos seus similares, das suas traduções e dos seus sinónimos. Se apenas os resultados diretamente relacionados com o termo (*Search only*), ou com as traduções do termo (*Translations*), ou com as similaridades do termo (*Similarars*), ou por fim com os sinónimos do termo (*Sinonyms*).

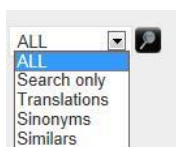


Figura 6 – Opções de pesquisa de resultados no Molecule

Após pesquisa com sucesso, a molécula formada tem em cada um dos seus átomos uma representação de cada uma das diferentes particularidades de pesquisa existentes no sistema Molecule, estando representadas essas particularidades através de diferentes ícones, para melhor identificação do átomo em questão. Ficam visíveis quais os termos mais utilizados ao nível de traduções num dos átomos, enquanto noutro átomo estarão os termos mais utilizados ao nível de sinónimos, noutro os termos mais utilizados ao nível de similaridade, e por fim, os termos mais utilizados absolutos relacionados com a procura, formando assim a molécula referente ao termo inicialmente pesquisado que está contido num átomo central.

Caso não seja possível fazer qualquer mapeamento com o *tag space*, então o Molecule tentará encontrar relações de similaridade geradas em tempo real entre o termo pesquisado e todos os previamente introduzidos como referido anteriormente.

Da resultante desta pesquisa entre termos será gerado um átomo de similaridade do termo procurado (Figura 7), tendo como eletrões todos os termos em que a similaridade foi estabelecida.

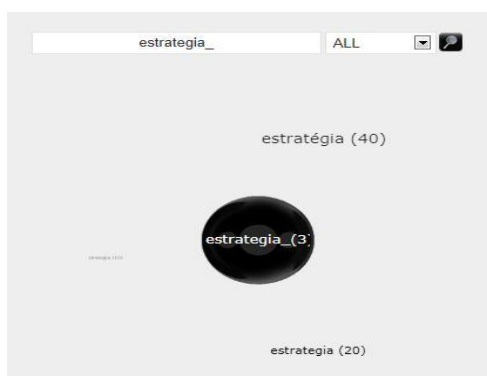


Figura 7 – Átomo gerada para o termo estratégia_

Com este conceito de visualização, o sistema Molecule consegue evitar a limitação ao nível da variedade do universo do espaço de amostragem, tão característica noutros sistemas como as nuvens de *tags*, dado não mostrar apenas os recursos rotulados exatamente com um termo específico, mas também todos os recursos e *tags* que de alguma forma se encontram ligados/mapeados ao mesmo.

A molécula apresentada é, na verdade, o resultado das várias interpretações do termo, oferecendo assim ao utilizador, uma maior escolha. Este facto traduz-se, também, num aumento exponencial e claro da probabilidade de sucesso, em caso de diferenças ao nível semântico, da interpretação do termo usado pelo utilizador que pesquisa e do utilizador que rotulou a informação. Desta forma, o utilizador que pesquisa pode rapidamente encontrar o que realmente procurava, pois tem acesso visual directo a várias outras interpretações/opções para o termo e, portanto, pode muito rapidamente identificar uma outra alternativa e, dessa forma, explorar e/ou refinar a sua pesquisa, aumentando assim a taxa de sucesso da mesma, não estando restringido só a resultados directos.

Outra das características intrínsecas a esta forma de visualização de rótulos é o grande aumento do número de conceitos retornados, uma vez que cada átomo (excetuando o central) é uma “esfera” de termos.

O conceito de esfera de termos, permite que os termos que estão intrinsecamente ligados, estejam dispersos por um determinado espaço 3D, e portanto, ao ser acrescentada uma dimensão, torna-se claro que irá forçosamente existir um aumento da área disponível para alojar termos, em comparação à área disponível numa “normal” nuvem de *tags* (2D), e portanto como podemos ver na Figura 5, existe potencialmente um espaço muito maior de amostragem, em comparação com a nuvem de *tags* habitual.

Assim, o utilizador pode ver os vários termos dispersos pela esfera, podendo rodá-la para navegar na mesma e ver em maior plano um determinado termo.

Acresce a este facto que cada molécula é constituída por quatro átomos, sendo que todos estes rodam por sua vez á volta do átomo principal e estático que é o termo pesquisado, ficando ainda disponível um maior espaço para a amostragem de informação.

Desta forma, fica demonstrado que a interpretação que o sistema Molecule faz da modelação gráfica do *tag space* é um grande passo em frente no que diz respeito às restrições do espaço de visualização, pois permite a disponibilização de muito mais espaço de informação, quando comparado aos sistemas tradicionais de nuvens de *tags*.

Para além de todas as suas mais-valias funcionais, o sistema Molecule é caracterizado, também, por uma grande atratividade para os seus utilizadores, resultante não só do seu visual inovador, derivado do seu conceito de modelação gráfica dos termos, mas também por convidar o utilizador a interagir com ele, de uma forma muito intuitiva e natural, dado todo o movimento de termos e de átomos possível, que o cativa a usá-lo de um modo recorrente.

Esta modelação gráfica recorre à tecnologia Silverlight, através de 3 modelações diferentes mediante se é pretendido carregar o átomo inicial (Figura 4), a molécula (Figura 5) ou ainda um átomo com termos relacionados (que ocorre quando nenhum termo é encontrado e é feita uma procura de similaridade por todos os termos como descrito anteriormente e visível na Figura 7).

O sistema Molecule, mais precisamente as suas diferentes modelações gráficas, servem-se de um serviço *Web* para carregarem os seus átomos de termos, sendo este serviço público, e portanto, podendo ser integrado em diversos sistemas e, assim, reutilizado e reaproveitado em diversas plataformas, mesmo se estas não implementam o mesmo número de “átomos” ou a mesma forma de visualização que o sistema nativo.

Esta solução de carregamento por serviço *web* permite, também, que seja virtualmente infinito o número de termos que podem ser carregados para cada átomo ou serviço, uma vez que recebe por parâmetro em cada um dos seus métodos o número de termos a devolver (ordenando pelos mais utilizados), ficando apenas limitado pelo bom senso do seu utilizador, dado que não existem limitações ao nível da quantidade de informação que poderá ser passada, como poderiam existir, por exemplo, se utilizada uma solução de passagem de dados por parâmetros.

5. Conclusão e Trabalho Futuro

Através das *tags* que atribuem a diferentes recursos, os utilizadores apresentam a sua própria visão, com a utilização de uma linguagem que não é considerada uniforme, pois a própria cultura e o nível de conhecimento variam de utilizador para utilizador. Esta forma de indexação mais alargada reflete os termos geralmente aplicados pelos utilizadores dos recursos ou não das pessoas que os criaram ou disponibilizaram. Isto pode ser visto como uma desvantagem mas simultaneamente permite que as pessoas possam beneficiar dessa variação semântica e linguística. No entanto, e para que isso seja possível, é necessário que os sistemas de visualização de informação adotados permitam capturar essa diversidade, sem sobrecarga cognitiva dos seus utilizadores.

O trabalho apresentado neste artigo vem colmatar diversos dos problemas característicos dos sistemas de *tags*, nomeadamente da forma de visualização normalmente mais utilizada, a nuvem de *tags*, sugerindo uma solução inovadora para a visualização de *tags*, que considera 4 perspetivas, cada uma visível num átomo diferente:

- Tradução – As traduções do termo em diferentes línguas;
- Sinónimos – Os diferentes sinónimos do termo na língua inglesa;
- Similares – Os diferentes termos similares ao termo em questão;
- Mais utilizados – As *tags* utilizadas mais vezes em conjunto com o termo inicialmente considerado (a abordagem considerada habitualmente nas nuvens de *tags*);

Dada a sua capacidade de conexão com outras plataformas, o sistema Molecule, vem ainda permitir que os seus serviços sejam consumidos por terceiros, caracterizando-se assim como um sistema útil e de partilha de informação.

A grande maioria das *tags* utilizadas atualmente no sistema Molecule foi obtida de um grande conjunto de dados utilizados num sistema de bookmarking social (<http://delicious.com>) de Setembro de 2007 a Janeiro de 2008, e disponibilizado no âmbito de um trabalho (Wetzker, Zimmermann, & Bauckhage, 2008) que o analisou, estando prevista a importação de muitas mais, limitada pelo número de utilizações (mensais) dos serviços que o Molecule usa para gerar informação, e que levará a que seja possível quantificar, no futuro, a diferença de retorno nos resultados de pesquisa entre o Molecule e o sistema original.

Estão previstos também diversos melhoramentos ao sistema, por forma a tornar o mesmo ainda mais eficaz, mais partilhável, mais completo e mais seguro. Assim, ao nível da performance, será revista a rapidez de inserção e pesquisa do sistema, através da introdução de vários *workers* em *background* obtendo-se, dessa forma, um aumento na rapidez de execução dessas tarefas, nomeadamente na inserção de novas *tags*, não ficando estas pendentes das respostas dos serviços de sinónimos e tradução. Ao nível da segurança será adicionado ao sistema uma infraestrutura de partilha de chaves pública, de forma a que a troca de informação entre o sistema Molecule e sistemas que usem os seus serviços web seja confidencial, autenticada e não repudiável. Por fim, será planeada a interligação entre o sistema Molecule a diversos serviços que proporcionem conhecimento ontológico, levando assim à geração de um sexto átomo na molécula visual, permitindo aumentar ainda mais a eficácia da pesquisa do sistema.

Referências

- Baldonado, M. Q. W., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization (pp. 110–119). Presented at the Proceedings of the working conference on Advanced visual interfaces.
- Elhussein, M. (2012). A Descriptive Model for Arabic-English Cross-Lingual Tagging. Presented at the The International Conference on Communications and Information Technology (ICCIT-2012), Hammamet, Tunisia.
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 11(4). Retrieved from <http://www.dlib.org/dlib/aprilo5/hammond/o4hammond.html>
- Hess, A., Maass, C., & Dierick, F. (2008). From Web 2.0 to Semantic Web: A Semi-Automated Approach. Presented at the 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008), Tenerife, Spain.

- Levenshtein, V. (1965). Binary Codes for Correcting Deletions, Insertions, and Reversals. *Doklady Akademii Nauk SSSR*, 163(4), 845–848.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. New Horizons in Commercial and Industrial Artificial Intelligence. *Communications of the ACM*, 38(11), 39–41.
- Peters., I. (2009). *Folksonomies indexing and retrieval in Web 2.0*. De Gruyter.
- Sanchez-Zamora, F., & Llamas-Nistal, M. (2009). Visualising tags as a network of relatedness. Presented at the The 39th ASEE/IEEE Frontiers in Education Conference.
- Shaw, B. (n.d.). Utilizing Folksonomy: Similarity Metadata from the Del.icio.us System. Retrieved from <http://www.metablake.com/webfolk/web-project.pdf>
- Stefaner, M. (2007, June). *Visual tools for the socio-semantic web* (Master's thesis). University of Applied Sciences Potsdam.
- Ware, C. (2004). *Information Visualization: Perception for Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Weller, K., & Peters, I. (2008). Seeding, Weeding, Fertilizing. Different Tag Gardening Activities for Folksonomy Maintenance and Enrichment (pp. 110–117). Presented at the Proceedings of I-Semantics '08, International Conference on Semantic Systems, Graz, Austria.
- Wetzker, R., Zimmermann, C., & Bauckhage, C. (2008). Analyzing social bookmarking systems: A del.icio.us cookbook. Mining Social Data (MSoDa) Workshop Proceedings, pp. 26–30. ECAI 2008.