

Molecule: Sistema de organização e visualização de Tags

José Martins¹ Isabel Azevedo²

ata, citation and similar papers at core.ac.uk

brought to you

provided by Repositório Científico do Instituto Pol

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 blogues, 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 abstractas 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 exactamente 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.

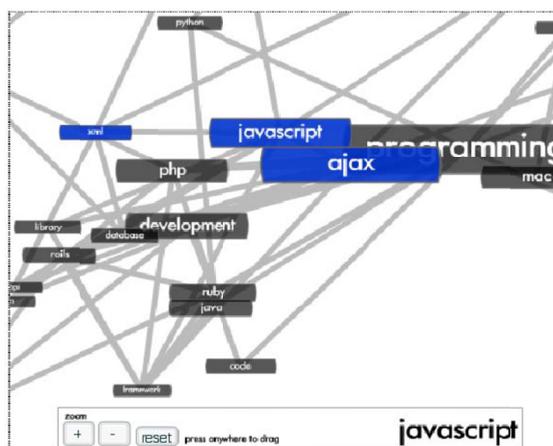


Figura 2 – Tag map

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, quando 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.

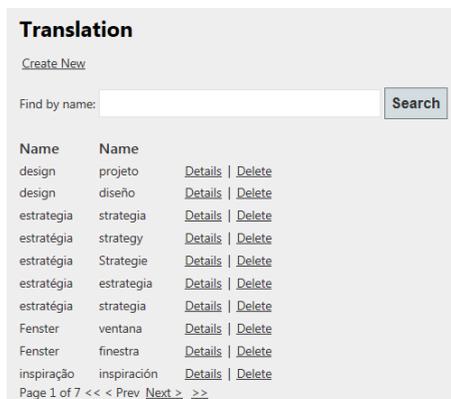


Figura 3 – Menu de gestão de traduções

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 conseqüentemente 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

³ http://www.amazon.com/tag/excellent/products/ref=tag_dh_istp

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.

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 têm diversas *tags* associadas (electrões), e tudo se interliga num único sistema de informação relacionada (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 na sua funcionalidade, materializando-se este numa esfera 3D (átomo) como visível na Figura 4.

Para além desta informação, é sempre visível o número de recursos associados a uma determinada *tag* em qualquer um dos modos de visualização do Molecule, variando o seu tamanho mediante esse número.

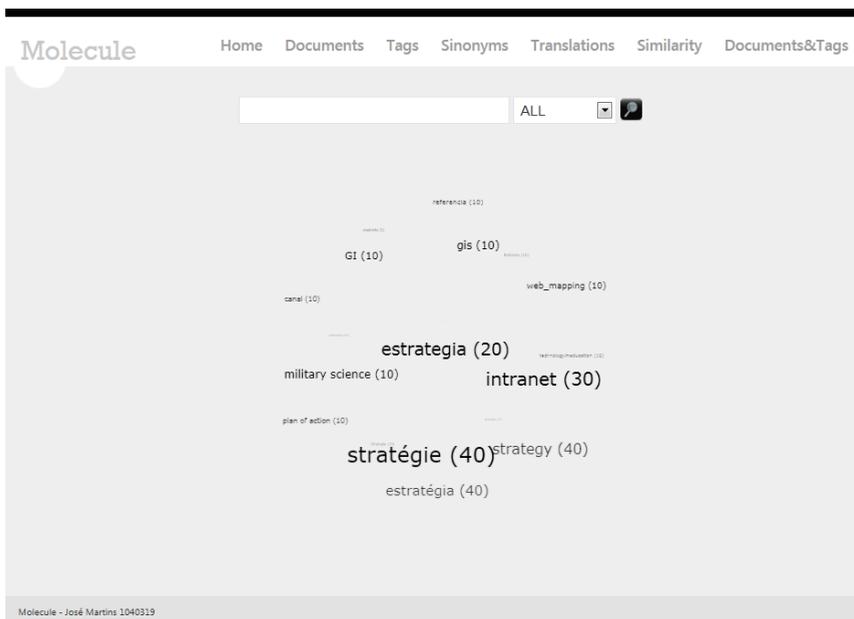


Figura 4 – Vista inicial após autenticação

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 infra-estrutura 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/april05/hammond/04hammond.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.