

AVALIAÇÃO DOS MOTORES DE PESQUISA COMO FERRAMENTA DE CONHECIMENTO

(Estudo de Caso Google, SlideShare, YouTube)

Fernando Bandeira

Professor Auxiliar
CECLICO, Faculdade de Ciências Humanas e Sociais
Universidade Fernando Pessoa, Porto, Portugal
fband@ufp.edu.pt

Tânia Marisa Fernandes Dias

Aluna do Mestrado em Ciências da Comunicação
Faculdade de Ciências Humanas e Sociais
Universidade Fernando Pessoa, Porto, Portugal
20633@ufp.edu.pt

RESUMO

O presente artigo visa avaliar o desempenho dos motores de pesquisa na Internet. Metodologicamente recorre a duas linhas de investigação que derivam, uma dos sistemas de recuperação de documentos e a outra do modelo de análise de documentos da Internet. Para responder as estas questões avalia: a adequação, a precisão, a densidade média de termos, o peso percentual dos descritores, o tipo e a fonte dos documentos. As conclusões evidenciam que o Google ocupa a primeira posição, o SlideShare a segunda e o YouTube a terceira.

PALAVRAS-CHAVE

comunicação de crise, densidade termos, Google, Internet, motores de pesquisa, precisão, SlideShare, YouTube

ABSTRACT

This article aims to evaluate the performance of Internet search engines. In methodological grounds it applies two different research traditions, one used to evaluate text retrieval systems the other is the model used to assess Internet documents. To answer these questions it uses the following indicators: precision, recall, average terms density, concept percent wage, document type and source. Conclusions are that Google is leading search engine followed by SlideShare and YouTube.

KEYWORDS

Concept density, crisis communication, Google, Internet, precision, recall, search engines, SlideShare, YouTube.

1. INTRODUÇÃO

A Internet teve a sua origem na necessidade de troca expedita e rápida de informação técnica e científica, porém o seu crescimento exponencial conduziu à disponibilização de uma grande diversidade informação *on-line*, contudo esta evolução quantitativa não correspondeu, em muitos casos, a informação criteriosa, relevante e fidedigna (Mintz).

Simultaneamente, foram desenvolvidas aplicações, vulgo motores de pesquisa, destinadas a auxiliar os utilizadores a localizar a informação. De uma forma muito simplificada, essas aplicações localizam ficheiros das mais variadas naturezas (sites, imagens, filmes, etc.), classificam-nos e disponibilizam-nos de uma forma hierarquizada (ranking) (Miller).

Entre esses motores o de maior nomeada é, sem dúvida, o Google, cujo domínio hegemónico, pode ser ilustrado com o neologismo “googlar,” sinónimo de pesquisar informação na Internet. No entanto, existe uma variada plêiade de motores disponíveis (genéricos, meta pesquisa, directórios, especializados em determinados formatos de documento, etc.), por isso ocorre perguntar se existem diferenças de desempenho entre eles.

O artigo que ora se apresenta, visa justamente avaliar o comportamento de três motores de pesquisa diferentes (Google, SlideShare e YouTube), à luz de metodologias e indicadores científicos.

2. ENQUADRAMENTO TEÓRICO

Ao questionar a eficácia dos resultados obtidos na pesquisa de informação, o presente artigo reclama-se de uma sólida linha de pesquisa, que trata a pertinência dos sistemas de recuperação de documentos (*text retrieval systems*), nas suas diversas facetas: qualidade da informação, funcionalidades do *software* de pesquisa, indexação/thesaurus, estratégia de pesquisa, etc.

Com efeito e restringindo o horizonte aos formatos digitais, ela remonta a 1964, ano da criação do primeiro sistema comercial de recuperação de documentos Dialog (Summit), embora seja só nos anos 70, com a maturidade da designada “*online industry*,” que este tipo de investigação começou a ganhar sólida reputação em revistas científicas e livros.

De facto, a proliferação de serviços comerciais que ofereciam sistemas de pesquisa diversificados, o crescimento exponencial dos textos nas bases de dados e os custos de acesso (Bjørner), a breve trecho conduziram investigadores e *search brokers* a questões como: Quais são as bases de dados que devolvem melhores resultados? As que melhor permitem expressar o objecto de pesquisa? As que apresentam melhor qualidade de indexação?

Mais tarde, com o crescimento exponencial da Internet e dos motores de pesquisa, a avaliação dos recursos *online*, ganhou uma nova perspectiva, que era algo secundária na abordagem anterior — a análise da qualidade e fidedignidade dos próprios recursos de informação, o célebre modelo ACOAC (*Authority, Coverage, Objectivity, Accuracy, Currency*), cuja explicação está sobejamente documentada na Internet (existe uma óptima descrição de Elizabeth Kirk para a Johns Hopkins University (Kirk).

Por outro lado, reduziu-se a especificidade da análise, já que os motores de pesquisa não oferecem as mesmas funcionalidades que os sistemas de recuperação de documentos "... search engines don't have nearly the sophistication that even our early systems did..." (Roger Summit *apud* Bjørner 7), sendo por isso impossível aplicar-lhes os sofisticados sistemas de análise até então usados.

No presente trabalho, retoma-se relativamente aos sistemas de recuperação de documentos a questão de saber: "Quais as bases de dados que devolvem melhores resultados?" e relativamente às metodologias de avaliação de conteúdos Internet, o item *Authority* (fonte e tipo de documento).

Tendo em consideração estas linhas orientadoras, procedeu-se à escolha dos motores de pesquisa, recaindo esta sobre: o Google, o YouTube e o SlideShare. Como se mencionará na metodologia, cada um destes motores, propõe-se oferecer informação em formatos diferenciados: diapositivos (SlideShare), vídeos (YouTube), documentos indiferenciados (Google).

Esta escolha tem por base duas ordens de razão: a da oportunidade e a científica. Relativamente às razões de oportunidade, são recursos de pesquisa profusamente usados no meio académico, que os autores consultam recorrentemente, tendo por isso um conhecimento profundo da sua utilização, da interpretação dos resultados de pesquisa neles obtidos e da análise da informação recolhida.

No conjunto das escolhas o Google representa os motores de pesquisa clássicos e indiferenciados, por oposição ao SlideShare e YouTube, motores especializados e conotados com a Web 2.0, mais especificamente com o que Tim O'Reilly designa por potenciar da inteligência colectiva "Harnessing Collective Intelligence" (O'Reilly).

Complementarmente, trata-se em qualquer das escolhas, de líderes dos respectivos segmentos de oferta de informação, o que não sendo propriamente um considerando científico é, sem dúvida, um argumento superlativo para atestar da importância da escolha feita, no contexto dos recursos disponíveis na Internet.

Como se mencionou, na escolha também pesaram argumentos científicos. Nesta linha de argumentação convocou-se McLuhan, designadamente o conceito o "meio é a mensagem". (McLuhan e Fiore).

Embora se esteja ciente da impossibilidade de fazer uma transposição mecânica da proposta conceptual dos autores: pois o contexto em que foi feita é distinto, afigura-se pertinente explorá-la, tendo em consideração que o tipo de documentos disponibilizados por cada um dos motores de pesquisa (texto, vídeo, apresentações) deve obedecer a narrativas específicas, envolver o destinatário de feição distinta, condicionar a forma de apreensão da mensagem, estabelecer contornos singulares para cada conteúdo a transmitir³.

3 A título de exemplo: As apresentações pressupõem a presença do emissor como parte intrínseca do próprio dispositivo de comunicação. Com efeito, elas são construídas tendo como base a dinâmica que o narrador estabelece quer com a informação quer com a audiência. Por outro lado, os diapositivos segmentam a mensagem não só do ponto de vista lógico mas também físico, cada diapositivo forma uma unidade sintagmática com um peso

Consequentemente, afigura-se pertinente assumir o seguinte axioma: cada um destes recursos usando meios diferentes, veicula mensagens diferentes, isto é pontos de vista distintos sobre o conceito em análise.

3. METODOLOGIA DO ESTUDO

Apresentado o enquadramento teórico, mandam os cânones do trabalho científico que se explicita a metodologia usada na pesquisa, em obediência passar-se-á a referir o respectivo referente:

3.1. QUESTÕES DE PARTIDA

Sumariamente, o presente artigo visa aquilatar da informação relacionada com a comunicação de crise, disponível na Internet e acessível a partir de motores de pesquisa previamente escolhidos. Mais concretamente, propõe-se responder às seguintes questões:

Qual é o melhor motor de pesquisa? Qual é o que apresenta mais resultados e mais pertinentes?

Qual é o que apresenta maior densidade de conceitos?

Como é que o conceito em análise é tratado pelos motores de pesquisa?

Qual o motor de pesquisa que apresenta maior diversidade de conteúdos?

Qual o tipo de documentos mais referido?

Quais são as fontes mencionadas nos documentos?

3.2. OBJECTO DE ANÁLISE

Como já foi referido, o objecto de análise será o conceito “comunicação de crise”. A sua explicitação em inglês (*crisis communication*), fica a dever-se ao facto da pesquisa com termos em português devolver um conjunto exíguo de documentos, considerando-se aprioristicamente que tal acervo poderia não ser significativo da informação disponível *on-line* sobre o tema.

Porquê “comunicação de crise”? Não existem razões de monta, trata-se de uma área de investigação próxima dos autores, na qual se sentem particularmente à-vontade para avaliar os documentos e respectivos conteúdos.

específico próprio. Mais existem boas práticas incontestadas que determinam, o número de parágrafos e o número de palavras por diapositivo e o próprio número de diapositivos por apresentação. (Duarte)

Outrossim, veja-se a o caso dos vídeos, como mensagem só fazem sentido enquanto unidade narrativa, ao contrário de um texto ou de um diapositivo, num vídeo é muito difícil identificar expeditamente segmentos específicos da mensagem, a informação retida depende em larga medida da memória auditiva/visual, enfim existem diversos níveis de significação, a linguagem corporal, a entoação verbal, os adereços a iluminação, etc. (Moran)

3.3. UNIVERSO DE ANÁLISE

Tomou-se como universo de análise os documentos disponibilizados na Internet.

3.4. OBJECTO DE ESTUDO

Actualmente é disponibilizada aos internautas uma grande variedade de motores de pesquisa, no caso foram seleccionados três: Google, YouTube e SlideShare. Os fundamentos dessa escolha foram já avançados no enquadramento teórico, são razões de oportunidade (curva de aprendizagem, liderança no segmento de mercado e especialização) e a expectativa teórica de que tratando-se de diferentes média, estes devem ser portadores de diferentes mensagens. Mais, em conjunto os três motores cobrem uma parte substantiva dos formatos informação disponível para pesquisa na Internet.

Com efeito, sendo o Google um motor de busca tradicional e genérico, permite a busca de documentos: institucionais, científicos, relatórios oficiais, informação comercial e marketing, etc. Ao mesmo tempo é também amplo o espectro de informação que disponibiliza, quer em termos qualitativos quer em termos quantitativos, pois pode ir da simples brochura comercial ao site de uma empresa até a um livro completo.

Já o YouTube é uma plataforma especializada, que funciona como um serviço de partilha de vídeo online. Nele predominam conteúdos relacionados com o entretenimento, mas também com formação e divulgação comercial. Trata-se de recursos com uma duração predominantemente curta (menos de cinco minutos), em que receptor da mensagem assume uma atitude mais passiva comparativamente com a informação providenciada pelos outros recursos. Em termos de conteúdos científicos, enquadra-se na linha da divulgação científica não apresentado, via de regra, profundidade conceptual ou recorte teórico assinalável.

Quanto ao SlideShare, é um motor de pesquisa especializado em apresentações (*slideshows*). Trata-se de documentos que raramente proporcionam a complexidade teórica de um livro, relatório ou artigo científico, todavia revelam-se excelentes fontes quando objectivo é ter um conhecimento integrado e diacrónico sobre um assunto, normalmente a narrativa numa apresentação é do estilo telegráfico, onde são apresentados de uma forma sintética e segmentada, conceitos, ideias, teorias, exemplos... mais importantes e basilares das matérias que se propõe tratar.

Pesquisando o SlideShare, tanto se encontram documentos com idoneidade científica e académica, por exemplo aulas, apresentações de estudantes, resumos de dissertações, resenhas de obras, como apresentações comerciais, estudos de consultoras, etc. Complementarmente, dos três motores é provavelmente aquele que mais encarna a filosofia Web 2.0 e que pode ser utilizado em diferentes contextos e lugares.

Pelas razões avançadas, os três motores identificados, afiguram-se ser aqueles que melhor servem os propósitos enunciados nas questões de partida, por serem líderes dos respectivos segmentos de mercado e porque, dada sua especialização, acredita-se espelham de modo superlativo a variedade de informação disponível na Internet, ao mesmo tempo que revelam de forma mais abrangente as diferentes abordagens ao conceito objecto de análise.

3.5. A AMOSTRA

No que se refere à amostra, colocaram-se problemas de grande monta à sua determinação científica. A apresentação dos documentos disponibilizada por cada um dos motores é probabilística, pois eles passam por uma ordenação (*ranking*) determinada por um algoritmo, todavia esse algoritmo não é do conhecimento público, pelo contrário é dos segredos mais bem guardados (Michael). Assim a investigação encontrou-se face a um paradoxo, que é o de encontrar a dimensão de uma amostra probabilística, mas ignorando a probabilidade de um documento ocorrer numa determinada posição na série de resultados obtidos.

Assim, optou-se por analisar as 30 primeiras entradas de cada motor de pesquisa. Número que se estimou ser pertinente para o tipo de análise pretendida e para o tempo que é razoável despendido na investigação de um artigo científico. Obviamente está-se ciente que uma amostra desta dimensão, está aquém dos valores que uma investigação de maior envergadura obrigaria.

No total foram analisados 90 documentos.

3.6. DATA DAS PESQUISAS

As pesquisas foram efectuadas na semana que decorreu entre 19 e 25 de Junho de 2009.

3.7. TERMOS DE PESQUISA

Os termos de pesquisa usados foram: “comunicação de crise”.

Optou-se pela pesquisa em inglês “*crisis communication*” devido ao exíguo número dos documentos encontrados usando os vocábulos em português, nomeadamente no SlideShare e no YouTube.

Optou-se ainda pela utilização do termo de pesquisa entre aspas, isto é, restringiu-se a pesquisa aos documentos em que os vocábulos aparecessem em posições adjacentes, uma vez que se pretendia que os resultados correspondessem o mais rigorosamente possível ao conceito em pesquisa, condição que não seria praticável no caso de os vocábulos ocorrerem aleatoriamente nos documentos.

4. CARACTERIZAÇÃO DO ESTUDO DE CASO

4.1. RECOLHA DA INFORMAÇÃO

Acedeu-se à página principal de cada motor de pesquisa. Nela inseriram-se os termos de pesquisa antes mencionados. De seguida, procedeu-se à extracção dos documentos identificados: no Google a gravação da página, no YouTube o *download* do vídeo e no SlideShare descarregaram-se as apresentações para disco.

Houve alguns problemas relacionados com os resultados de pesquisa que obrigaram à substituição de documentos, as listas de reprodução foram ignoradas (YouTube), os documentos repetidos ou que apresentavam problemas técnicos (por exemplo, que não se conseguiram visualizar) também foram ignorados, em todos os casos essas referências foram substituídas pelo documento imediato.

4.2. TRATAMENTO DA INFORMAÇÃO

Após a recolha dos dados procedeu-se à análise e sistematização dos resultados obtidos, os documentos foram analisados individualmente, tendo sido consideradas as seguintes variáveis:

Conceitos ou termos mencionados no documento e relacionados com o tema de pesquisa;
 Tipo de documento;
 Tipo de fonte.

Durante a análise dos dados foram excluídos documentos que se mostraram irrelevantes para o estudo (por exemplo a crise financeira ou a crise da água).

No final da análise de conteúdo dos documentos pesquisados foram encontrados no total 224 termos que posteriormente foram reagrupados em dez grandes grupos. Deste processo de codificação, resultaram os termos apresentados na seguinte tabela.

Tabela 1. Termos em análise

Termos	Motores de pesquisa			
	Google	Slideshare	Youtube	Total
Plano de comunicação	21	9	1	31
Canais de comunicação	9	13	2	24
Construção da notícia	14	8	6	28
Aspectos legais e éticos	2	0	3	5
Formação/equipa de crise	19	5	11	35
Relações com os media	7	2	2	11
Identificar públicos	4	3	6	13
Hardware	5	0	1	6
Análise da informação	7	0	3	10
Boas práticas/auditoria/guidelines	4	1	1	6
Total	92	41	36	169

5. ANÁLISE DE DADOS

Sem prejuízo de medidas específicas que se possam estabelecer em função dos propósitos de cada investigação, existem dois indicadores obrigatórios para avaliar uma pesquisa de documentos, seja ela electrónica ou não: a precisão e a adequação.

5.1. ADEQUAÇÃO

A adequação (*recall*) obtém-se com o seguinte cociente: nº de registos identificados/total de registos na base de dados. Este indicador visa medir a pertinência da base de dados para o assunto em pesquisa (Meadow).

Tabela 2. Resultados da Pesquisa

Google	Slideshare	Youtube	Tot.Docs
62 100 000	77 000	668	62 108 368

Todavia e apesar de a adequação permitir em larga medida responder às interrogações colocadas, tal medida é impossível de calcular no presente estudo, pois o denominador, isto é o total de documentos indexados, em qualquer dos motores de pesquisa é desconhecido.

Não obstante, é inegável que os números são elucidativos, as diferenças de cifras manifestam-se em ordem de grandeza, o que significa que se está a falar de realidades distintas, com vantagem para o Google, seguido pelo SlideShare e, por fim, em posição modesta o YouTube.

5.2. PRECISÃO

A precisão é obtida pelo cociente nº de registos relevantes/total de registos identificados e visa identificar a eficácia da estratégia de pesquisa (Meadow).

Como se tratou de uma pesquisa simples, isto é, sem recurso ao “modo avançado” e os termos de pesquisa são os mesmos (*crisis communication*), só é possível retirar ilações relativamente ao que se poderia designar por “sensibilidade” de cada motor no que respeita ao conceito em avaliação.

Tabela 3. Precisão da Pesquisa

	Pertinentes	Tot. docs	Precisão (%)
Google	18	30	60
Slideshare	9	30	30
Youtube	15	30	50
Tot. docs	42	90	47

Nota. os valores reportam à amostra na impossibilidade de analisar a totalidade dos documentos resultantes da pesquisa

Os números relevam uma grande variação nos resultados, sendo que o SlideShare assume valores extremamente modestos, em torno de metade dos outros motores, também na relação YouTube/Google a vantagem de 10% para o último, não se pode considerar despcienda.

Todavia, parece ser a leitura do total dos documentos pertinentes (47%) que aponta uma evidência a ter em devida conta: pelos dados mais de metade do esforço dispendido nas pesquisas Internet é não produtivo.

5.3. DENSIDADE MÉDIA DE TERMOS POR MOTOR DE PESQUISA

Este indicador, permite centrar a análise no conjunto dos documentos e obtém-se com o cociente Σ termos/ Σ de documentos pertinentes.

Tabela 4. Densidade média de termos por motor de pesquisa

Google	Slideshare	Youtube	Tot.Docs
5,1	4,6	2,4	4,0
42%	38%	20%	33%

Como se pode constatar, a densidade média de termos por documento revela uma grande heterogeneidade entre os três motores e, conseqüentemente, uma grande dispersão em torno da média. Sendo que o Google e o SlideShare apresentam uma maior riqueza de conteúdos face ao YouTube com resultados manifestamente modestos

Será porventura interessante comparar este quadro com o anterior (precisão), pois se uma medida fornece informação sobre a riqueza de conteúdos a outra revela a sua importância, funcionando por isso em complementaridade.

Antes de proceder a essa comparação, é necessário avisar que se estão a comparar valores estruturalmente diferentes, quer nos conceitos que consubstanciam, quer na metodologia de cálculo (no primeiro caso pesos percentuais e no segundo médias), em termos de análise esta circunstância só permite, em rigor, comparar tendências ou semelhanças, não valores directamente.

Constata-se que a posição relativa entre os motores de pesquisa se altera, o Google lidera nas duas tabelas, em termos de precisão o YouTube ocupa o segundo lugar e o SlideShare o terceiro, já na densidade média de termos o SlideShare ocupa o segundo lugar e o YouTube o terceiro.

Dito por outras palavras, os resultados de pesquisa no YouTube devolvem mais documentos com informação sobre o tema, todavia embora em menor número os documentos SlideShare apresentam uma maior riqueza de conceitos.

Em todo o caso é interessante reter que em média, cada documento versa quatro conceitos, valor algo modesto se comparado com as fontes de documentação clássica.

5.4. PESO PERCENTUAL DOS DESCRITORES POR MOTOR DE PESQUISA

O peso percentual dos descritores por motor de pesquisa, consiste na desagregação ao nível de descritor, do indicador densidade média de termos por motor de pesquisa e mede a importância de cada um destes no total de informação recolhida.

O seu cálculo processa-se em duas etapas, a primeira visa neutralizar o efeito induzido pelas diferenças de precisão dos motores de pesquisa e obtém-se pelo quociente: fi descritor/Σdocumentos pertinentes no motor de pesquisa, na segunda etapa, com os resultados obtidos, constrói-se uma tabela de contingência, onde os valores são apresentados no seu peso percentual relativamente ao total.

A leitura global dos números evidência comportamentos algo difusos entre os motores de pesquisa. Partindo dos valores em totais, pode afirmar-se categoricamente que a primeira posição vai para o Google segunda para o SlideShare e a terceira para o YouTube.

Todavia a análise do desempenho descritor a descritor mostra que o Google contribui com 6 valores máximos e 1 valor mínimo, o SlideShare 2 valores máximos e 4 mínimos e por fim o YouTube 2 valores máximos e 5 mínimos, esta perspectiva de coloca os dois últimos motores praticamente em igualdade de circunstâncias.

Por outro lado os coeficiente de variação permitem constatar que o Google e o YouTube apresentam uma maior consistência entre valores que o SlideShare. Este último revela um comportamento tanto mais errante quanto apresenta o maior valor absoluto (10,5%), juntamente com três ausências de valores.

Tabea 5. Peso percentual dos descritores por motor de pesquisa

Descritores	Google	Slideshare	Youtube	Total	Acumulado
Formação/equipa de crise	↑8,7%	↓4,6%	6,1%	19,0%	19,0%
Plano de comunicação	↑9,7%	8,3%	↓0,6%	19,0%	38,0%
Canais de comunicação	4,1%	↑12%	↓1,1%	17,0%	55,0%
Construção da notícia	6,4%	↑7,4%	↓3,3%	17,0%	72,0%
Identificar públicos	↓1,8%	2,8%	↑3,3%	8,0%	80,0%
Relações com os media	↑3,2%	1,8%	↓1,1%	6,0%	86,0%
Análise da informação	↑3,2%	↓0%	1,7%	5,0%	91,0%
Boas práticas/auditoria/guidelines	↑1,8%	0,9%	↓0,6%	3,0%	95,0%
Hardware	↑2,3%	↓0%	0,6%	3,0%	97,0%
Aspectos legais e éticos	0,9%	↓0%	↓1,7%	3,0%	100,0%
Total	42%	38%	20%	100%	

legenda. (↑) valor máximo, (↓) valor mínimo

Tabela 6. Coeficiente de variação (%)

Google	Slideshare	Youtube	Total docs
6,8	10.5	8,3	6,8

Não obstante, é preciso não descurar a concentração de valores, pois quatro dos termos em análise, em conjunto computam 72% dos valores (ver coluna acumulado). Restringindo a análise a estes valores, a perspectiva altera-se radicalmente, o SlideShare ocupa uma posição de relevo (acumulando 32% em 72%), seguido pelo Google (29% em 72%), o YouTube revela a posição muito mais modesta (11% em 32%). Como se constata, este cenário apresenta-se singular e contrário às tendências predominantes anteriormente detectadas.

Com efeito os valores obtidos para o peso percentual dos descritores por motor de pesquisa, conduzem a diferentes vectores de análise, em grande parte contraditórios, sendo que as principais conclusões postulam:

Globalmente, se se pretende uma maior densidade de termos deve preferir-se o Google seguindo-se-lhe do SlideShare e por fim a uma distância considerável o YouTube.

Por outro lado, se a intenção de pesquisa é especialização, isto é uma menor dispersão de conteúdos o YouTube apresenta-se como a melhor escolha, seguido pelo Google já o SlideShare com um maior coeficiente o que revela uma grande dispersão de conteúdos.

Para concluir deve relembrar-se a grande concentração de valores, centrando a análise nesse núcleo central de descritores o SlideShare ganha vantagem sobre os restantes motores, o Google assume uma posição relevante enquanto o YouTube apresenta um desempenho objectivamente muito pobre.

5.5. FONTE E TIPO DE DOCUMENTO

Relativamente às fontes e tipo de documento, trata-se de indicadores da maior importância para avaliar da qualidade da pesquisa e do motor de pesquisa, pois as primeiras garantem credibilidade e o tipo de documento, e o segundo o rigor do tratamento e riqueza de conteúdos.

Antes de avançar com a leitura dos dados, convém mencionar que um documento pode assumir duas ou mais categorias⁴. Posta esta ressalva metodológica, o Google e YouTube surgem como os motores cujos tipos de documento são mais específicos, pois são na sua maioria única e exclusivamente caracterizados por um único termo, contudo os resultados do Google são quase exclusivamente teóricos, ao passo que os do YouTube se encontram distribuídos por todas as categorias.

⁴ Por exemplo, um estudo de caso com um enquadramento teórico, seria classificado nas duas categorias, esta circunstância justifica que os totais apresentem frequências superiores a número de documentos e impede que se achem valores percentuais, como foi regra nos outros casos deste artigo.

Por seu turno mais de 50% dos documentos SlideShare tem dupla caracterização o que, em princípio, os torna mais generalistas. Quanto à distribuição por categorias, de acordo com a experiência dos autores, elas revelam uma distribuição típica, com um maior peso para os textos teóricos, seguidos de estudo de caso e por fim, informação empresarial.

Tabela 7. Tipo de Documento

	Google	Slideshare	Youtube	Tot. docs
Estudo de caso	1	5	6	12
Seminário/formação/serviços	2	–	4	6
Teórico	18	9	15	42

As fontes dos documentos são sempre um bom indicador, que o pesquisador experimentado avalia para decidir sobre a natureza e a credibilidade da informação, sob este ponto de vista os resultados merecem algumas notas. A primeira é a fraca contribuição das universidades e organismos estatais para o cômputo dos documentos, este facto pode, eventualmente, ser obnubilado pelo peso da variável biblioteca/publicação digital, já que houve publicações provavelmente oriundas da academia e de organismos do estado que foram aí classificados em obediência à lógica de classificação estabelecida.

Tabela 8. Tipo de Fonte

	Google	Slideshare	Youtube	Tot. docs
Biblioteca/publicação digital	3	5	7	15
Empresa	7	–	6	13
Universidade/estado	8	4	2	14

Por outro lado, é de sublinhar o relevo que documentação oriunda do meio empresarial assume, será provavelmente resultado de actividades e serviços em expansão ou mesmo consequência da importância crescente que as empresas dão à visibilidade proporcionada pela Internet, contudo não se deve perder de vista que no quadro anterior a categoria correlativa (seminário/formação/serviços) só agrega seis em 42 resultados.

Centrando a análise nos motores de pesquisa, os valores não deixam de ser menos atípicos. O Google e o YouTube apresentam distribuições inversas, no primeiro caso crescentes e no segundo decrescentes, já o SlideShare apresenta valores próximos, mas distribuídos somente por duas variáveis, por isso uma leitura conjunta dos resultados obtidos pelos motores de pesquisa afigura-se pouco razoável.

Com a informação coligida não parece pertinente qualquer leitura ulterior, justificar-se-ia testar outras hipóteses para confirmar ou infirmar estas tendências, porém quer no âmbito quer na economia própria deste tipo de trabalho, tal não tem cabimento aqui.

6. CONCLUSÕES

Finda a análise de dados, é chegada a altura de contrastar as questões de partida com os resultados no sentido de as esclarecer.

A questão central é: Qual o melhor motor de pesquisa? O que apresenta mais resultados e mais pertinentes? O que apresenta maior densidade de conceitos?

Pode-se afirmar em termos quantitativos os motores apresentam resultados totalmente díspares, com efeito, se a bitola se colocar no número de documentos, o Google ganha indiscutível vantagem com milhões de documentos, seguido do SlideShare com milhares e por fim o YouTube com centenas (ver tabela 2). Todavia este indicador é muito rudimentar, sobretudo quando se fala de documentos *on-line*, onde os resultados das pesquisas frequentemente estão eivados de problemas como, documentos repetidos, *links* quebrados, sítios inactivos, etc., convém por isso analisar a qualidade dos documentos. Nesta perspectiva e no que respeita à precisão (tabela 3), embora o Google mantenha a liderança, o YouTube assume a segunda posição relegando o SlideShare para o último lugar. No que respeita à densidade de termos por documento (tabela 4), verifica-se novamente a vantagem do Google, seguido do SlideShare e do YouTube, deve no entanto notar-se que em ambos os casos existe uma polarização nítida de resultados.

A segunda questão versava a diversidade de conteúdos. O Google consegue obter frequências em todos os termos, quanto ao SlideShare, não apresenta valores em quase um terço dos termos. Todavia a esta análise tem um valor relativo, que só ganha relevância com a análise conjunta dos termos, contemplada na terceira questão colocada pelo artigo.

Como foi referido existe uma grande concentração (40% dos termos congregam quase dois terços dos resultados), restringindo a análise a este grupo, a tendência inverte-se, com o SlideShare a ocupar o papel de liderança seguido do Google e do YouTube. Já na análise dos termos menos versados, aflora o padrão dominante: Google SlideShare, YouTube.

Acresce que existem termos fundamentais na teoria da comunicação de crise (públicos, relações com os media, ...), mas que nos dados analisados revelam uma posição modesta.

Outra questão importante em qualquer pesquisa é "Qual o tipo de documentos são mais referidos?" Talvez não seja surpresa encontrar em posição de liderança os documentos de enfoque teórico, todavia a sua preponderância 70% contrasta significativamente com as restantes estudo de caso (20%), seminário/formação/serviços (10%).

Sinopse: O Google é, de todos os motores o que apresenta o melhor e mais regular desempenho. Ao SlideShare, embora com algumas reservas, merece ser creditada a segunda posição, pois apresenta na maioria dos indicadores essa posição. Por fim e com surpresa para os autores, surge o YouTube, sobretudo dado o seu fraco desempenho na maioria dos parâmetros, em contraste com a sua reputação no meio académico.

Adicionalmente emergiram três ideias, que embora não tivessem directamente contempladas pelas questões iniciais, parecem merecer alguma reflexão.

A primeira, é a fraca precisão dos instrumentos de pesquisa em análise (47%), isto é, mais de metade do esforço dispendido nas pesquisas é inútil.

A segunda, é a grande concentração de valores, o que pode indiciar um possível enviesamento da informação resultante da pesquisa.

A terceira, consiste na menor valorização de alguns dos tópicos em análise se comparados com as fontes de informação impressas sobre o tema, o que deverá conduzir à necessidade de questionar a consistência teórica dos documentos avaliados.

7. LIMITAÇÕES E DESENVOLVIMENTO POSSÍVEL DO ARTIGO

Como limitações podem apontar-se duas de maior importância: a data dos documentos e a dimensão da amostra. Relativamente à primeira, via de regra, foi muito difícil averiguar a data de publicação dos documentos pesquisados, embora no YouTube e o SlideShare fosse mencionado quando eles foram disponibilizados online (o que não corresponde necessariamente à data da sua publicação).

Para um pesquisador traquejado a data é um elemento de avaliação de suma importância, por um lado ela pode determinar a sua pertinência (se tratar de leis ou documentos com múltiplas versões, como é o caso de normas e comunicados), por outro lado, a data pode plasmar um variado conjunto de indicações complementares em função da natureza e objectivos da pesquisa (tecnologias em uso, teorias dominantes, contextos socioeconómicos, ...). Trata-se de uma limitação exógena ao artigo, que embora possa colmatada com o uso de formatos normalizados de apresentação dos documentos como Dublin Core Metadata Standard, é muito pouco provável que venha a ser ultrapassado se se mantiver o paradigma que enforma a Internet actualmente.

A segunda limitação diz respeito à dimensão da amostra, provavelmente deveria incidir sobre um maior número de documentos, circunstância que poderia propiciar maior fidelidade e consistência dos dados recolhidos e mitigar o reduzido número de documentos em avaliação. No entanto, tendo em conta o tipo de análise pretendida e o tempo que é razoável dedicar a escrita de um artigo, considerou-se pertinente o critério que determinou o número de documentos analisados.

Para a realização de estudos futuros, recomenda-se alargar a pesquisa a um segundo conceito da mesma área ou de área diferente, para comparativamente se aferir e validar os resultados obtidos.

Complementarmente em vez de se usar um sistema de codificação que resultou da própria pesquisa, seria talvez mais criterioso usar um thesaurus da respectiva área de conhecimento ou definir uma lista de termos a partir do *corpus* teórico, por exemplo do índice remissivo de obras científicas relacionadas.

Por fim seria de considerar como bitola de comparação dos resultados obtidos *text retrieval system* como por exemplo a B-On.

BIBLIOGRAFIA

- Bjørner, Susanne. "Online Before the Internet, Early Pioneers Tell Their Stories". *Searcher* 11.7 (2003): 52-61.
- Duarte, Nancy. *Slide:ology, the Art and Science of Creating Great Presentations*. Sebastopol: O'Reilly, 2008.
- Fleischner, Michael H. *SEO Made Simple: Strategies for Dominating the World's Largest Search Engine*. Totowa: Lightning Press, 2009.
- Kirk, Elizabeth E. "Evaluating Information Found on the Internet". *Sheridan Libraries*. Internet. 10 Jun. 2009. <<http://www.library.jhu.edu/researchhelp/general/evaluating>>
- Marshall, McLuhan, e Quentin Fiore. *The Medium is the Message*. New York: Penguin, 2008.
- Maxwell, James C. *A Treatise on Electricity and Magnetism*. Oxford: Clarendon, 1982.
- Meadow, Charles T. *Text Information Retrieval Systems*. Cambridge: Cambridge UP, 2008.
- Miller, Michael. *Complete Idiot's Guide to Search Engine Optimization*. New York: Alpha Books, 2009.
- Mintz, Anne P. *Web of Deception: Misinformation on the Internet*. New Jersey: CyberAge Books, 2002.
- Moran, José M. "O vídeo na sala de aula". *Comunicação & educação* .2 (1995): 27-35.
- O'Reilly, Tim. "What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software". *O'Reilly*. Internet. 10 Jun. 2009. <<http://oreilly.com/web2/archive/what-is-web-20.html>>
- Summit, Roger. "Reflections on the Beginnings of Dialog, The Birth of Online Information Access". *Dialog*. 10 Jun. 2009. <support.dialog.com/publications/chronolog/200206/1020628.shtml>.