

COLABORAÇÃO EM MASSA OU AMADORISMO EM MASSA?

Fernando S. Nifrário Rodrigues¹, Paulo Resende da Silva²

1) Investigador independente, Portugal

fernando.nifrario@gmail.pt

2) Dep. Gestão/Escola de Ciências Sociais - Universidade de Évora, Portugal

pfs@uevora.pt

Resumo

O alfa e o ómega deste estudo é a tentativa de provar a credibilidade da colaboração em massa enquanto pedra angular do modelo de negócio comumente denominado por *open source*. Para o demonstrar, realizou-se um estudo comparativo entre a qualidade dos artigos da Wikipédia e da Enciclopédia Britânica, tendo sido dado particular relevo à metodologia utilizada de modo a que os resultados obtidos fossem os mais fidedignos e abrangentes possíveis.

É utilizada uma amostra aleatória representativa do universo composto pelos artigos que integram ambas as enciclopédias, tendo estes sido avaliados por peritos proficientes na temática em que estes se inserem. A qualidade média dos artigos da Wikipédia analisados revelou-se superior à dos seus pares, sendo a diferença encontrada estatisticamente significativa.

Um inquérito previamente realizado junto dos avaliadores mostrou a reduzida confiança destes na qualidade da informação contida na Wikipédia. Esta percepção da qualidade e os resultados diametralmente opostos obtidos na avaliação dos artigos através de uma prova cega reforça a isenção do painel de avaliadores e a robustez dos resultados obtidos.

Em termos de estrutura iremos começar com um pequeno enquadramento teórico; apresentando de seguida a metodologia utilizada no estudo e as implicações e constrangimentos que a escolha desta acarreta; finalizando com a apresentação dos resultados obtidos e as conclusões que destes podemos extrair.

Palavras chave: Colaboração em Massa, Inteligência Colectiva, Crowdsourcing, Wikipédia, Enciclopédia Britânica.

1. Introdução

Para [Shirky 2008], os avanços tecnológicos tendem a ser disruptivos quando permitem a massificação de uma nova tecnologia e não apenas a substituição – evolução – de uma tecnologia por outra mais barata e eficiente. [Christensen 1997] considera que este primeiro tipo de avanço teve sempre como consequência que competências que antes estavam na mão de um grupo reduzido de peritos fossem assimiladas por uma multidão de amadores, o que, a médio prazo, torna o trabalho destes obsoleto, relegando-os para nichos de mercado e, a longo prazo, ao reajustamento das suas competências ou ao seu desaparecimento.

Em termos de acesso ao conhecimento, a invenção da impressão por Gutenberg no século XV, o aparecimento das rotativas com a revolução industrial no século XIX e a Internet no século XX foram responsáveis pelo aumento do conhecimento disponível e do número dos seus potenciais consumidores e criadores [Sunstein 2006].

O surgimento da Internet, segundo [Anderson 2006], potenciou o aumento substancial do número de criadores e consumidores de informação, teve – especialmente desde a chamada reinvenção desta, conhecida por Web 2.0 – um aumento sem precedentes no número de criadores, que se auto-organizaram utilizando ferramentas que permitem a partilha e consulta de informação entre milhões de pessoas.

O advento da Web 2.0 potenciou uma mudança radical no modo como entendemos a colaboração. Devido à Internet, uma multidão de pessoas não enquadradas nas tradicionais organizações hierárquicas das empresas e organizações estão a produzir bens, serviços e conteúdos.

A passagem de um modelo de pensamento baseado na ideia de *cogito* (eu penso), para o seu plural *cogitamus* (nós pensamos) prognosticada por [Lévy 1997], não se limita a concatenar os pensamentos individuais num único: baseia-se na crença de que a manutenção das singularidades e divergências dentro do grupo são a pedra de toque para a evolução e crescimento de uma inteligência colectiva em que o todo é maior do que a soma das partes.

O que [Lévy 1997] considerou uma “alcançável utopia” veio a tornar-se uma realidade anos mais tarde, pelo menos parcialmente, com o surgimento da Wikipédia e do conceito de *prosumers* – junção das palavras “produtor” e “consumidor” – apresentado em Wikinomics por [Tapscott e Williams 2006].

A Wikipédia é um dos exemplos apontados por [Tapscott e Williams 2006] que conseguiu utilizar o modelo de produção que os autores denominaram colaboração em massa, em conjunto com o software wiki, misturando as motivações não financeiras com as inovações tecnológicas da ferramenta para criar uma realidade que dificilmente alguém acreditaria ser possível. Neste caso, parece ter-se verificado a premissa de [Shirky 2008], e a tecnologia foi suficientemente disruptiva e conseguiu induzir um aumento exponencial no número de pessoas que coligem e consomem informação enciclopédica de uma forma colaborativa, auto-regulada e gratuita – uma

área que, historicamente, era dominada por um punhado de técnicos especializados e consumida por uma minoria disposta a pagar pelo privilégio. Estamos, assim, em presença de uma revolução de comportamentos que vai além da revolução tecnológica que potenciou o seu aparecimento.

Segundo [Tapscott e Williams 2006], podemos tentar definir colaboração em massa como um modelo de actuação colectiva que ocorre quando um grande número de pessoas trabalha voluntariamente na prossecução de um objectivo comum. De um modo geral, os projectos escolhidos são modulares e baseados na Internet tirando partido das ferramentas de colaboração online – como por exemplo o software *Wiki* – que facilitam a auto-organização dos participantes, potenciando a discussão de ideias e a partilha de conhecimento. Para [Ghazawneh 2008], o que distingue a colaboração em massa de outras formas de colaboração é o facto de ser o próprio conteúdo que está a ser criado que gere o processo colaborativo e não as interações sociais, reais ou virtuais, entre os participantes - a ênfase é posta no trabalho que está a ser feito, servindo este de mediador entre os colaboradores.

De acordo com [Tapscott e Williams 2006], a inteligência colectiva é sinónima de colaboração em massa sempre que esta garanta a partilha aberta de ideias, renunciando à propriedade intelectual, entre pares auto-organizados não limitados às barreiras geográficas – i.e., a inteligência colectiva é apenas um dos tipos de projectos de *crowdsourcing* [Howe 2009], sendo que [Benkler 2006] e [Shirky 2008] utilizam amiúde o termo *peer production*, com um sentido em tudo idêntico.

Este estudo tem assim como objectivo estratégico aferir as potencialidades da colaboração em massa através da determinação da qualidade do seu produto final, utilizando para tal o estudo de um caso concreto em que esta é aplicada; em termos tácticos, a organização escolhida é a Wikipédia, cujo modelo de produção assenta na colaboração em massa. A qualidade do seu produto final – informação – será aferida por comparação com a qualidade do produto final de uma enciclopédia generalista de referência, que concorre pelo mesmo mercado, a Enciclopédia Britânica¹ – cujo modelo de produção é suportado por uma estrutura hierárquica convencional.

O interesse desta comparação não se prende com a avaliação dos produtos em si mesmos, mas sim com as metodologias que lhe estão subjacentes. Assim, vamos assumir que as diferenças de qualidade destes produtos reflectem, maioritariamente, os processos que utilizam na criação de conteúdos que nos propomos avaliar. A escolha da Britânica como elemento de comparação deve-se à credibilidade desta junto da academia – os seus conteúdos foram criados e editados por peritos – e, também, por ter sido utilizada em estudos anteriores [Giles 2005] como termo de comparação.

2. Modelo Funcionamento da Wikipédia e da Britânica

A Britânica usa um processo tradicional em que os conteúdos são produzidos por peritos remunerados escolhidos pela própria empresa e depois revistos pelos editores antes de serem disponibilizados aos utilizadores. Os editores têm a última palavra no processo e podem mandar reformular, ou mesmo excluir, um artigo que julguem não ter qualidade para ser publicado.

Segundo [Keen 2007], é precisamente a escolha de peritos para a elaboração dos artigos e posterior revisão que torna a Britânica uma fonte de informação credível e confiável, por contraponto à Wikipédia. Por outro lado, [Tapscott e Williams 2006] alertam que o poder unilateral por parte do editor pode conduzir a artigos com uma abordagem enviesada que reflectem a visão pessoal do editor sobre o tema.

[Lih 2009] explica que a Wikipédia contrapõe à abordagem académica “editar e depois escrever” a abordagem “escrever e depois editar”: para a colaboração em massa funcionar, é necessário despertar a atenção de eventuais contribuidores, pelo que a criação de um pequeno artigo, mesmo que incompleto e mal estruturado, irá atrair outros contribuidores cuja actividade atrairá outros mais, aumentando assim as hipóteses de se chegar a um artigo de qualidade.

Na Figura 1 podemos ver uma comparação simplificada dos modelos de negócio das enciclopédias em questão.

¹ A Enciclopédia foi publicada pela primeira vez no século XVIII, tendo como editor William Smellie, sendo actualmente a mais antiga enciclopédia de língua inglesa que ainda continua a ser publicada. A sua reputação foi sendo construída ao longo dos tempos graças às contribuições de peritos de renome em diversas áreas do conhecimento que com ela colaboraram, tais como Sigmund Freud, Albert Einstein, Marie Curie, Leon Trotsky, Milton Friedman e Carl Sagan, entre outros. No início de 2012, a Britânica descontinuou a publicação da edição em papel da sua enciclopédia, passando a disponibilizar apenas a versão digital desta.

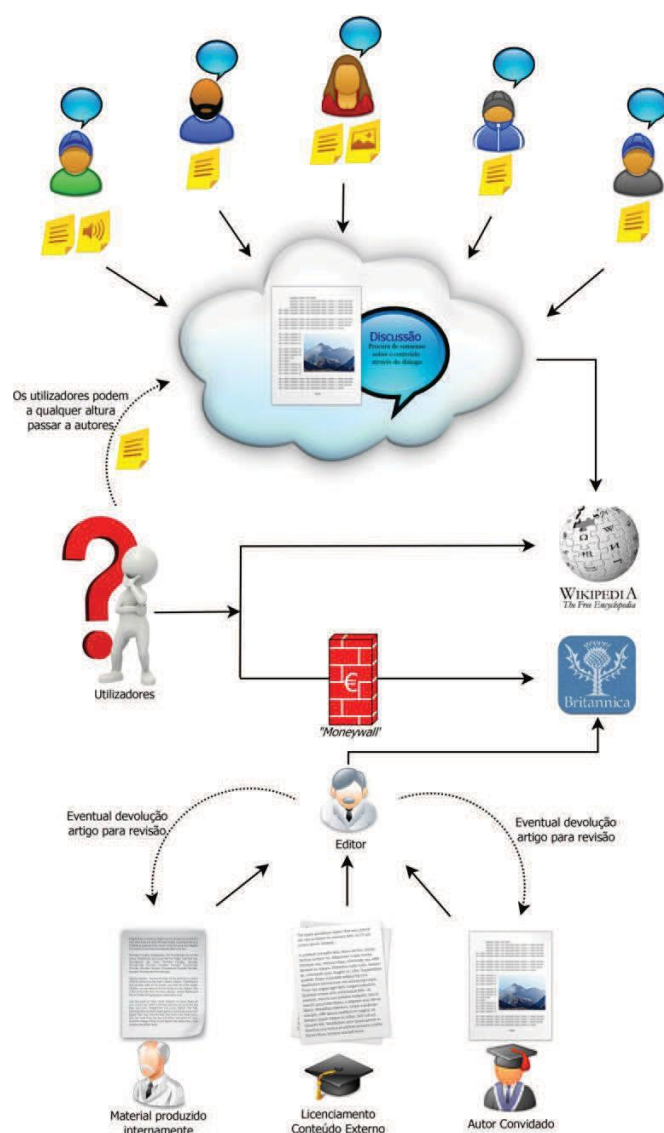


Figura 1 - Modelo Funcionamento Wikipédia e Britânica

A Wikipédia não escolhe os seus contribuidores, não os remunera, e os artigos não passam por nenhuma revisão antes de serem publicados. Segundo [Broughton 2008], o software *wiki* permite que cada artigo tenha associado uma página² de discussão em que os contribuidores deste possam conversar e tentar chegar a consenso sobre o seu conteúdo através do diálogo. Caso tal não seja possível, é comum usar a votação para decidir se determinado conteúdo deve ser ou não incluído no artigo.

Nos casos extremos em que a natureza polémica de um artigo suscite “guerras” de edição e mesmo actos de vandalismo que passem pela publicação de conteúdo ofensivo e a remoção dos pontos de vista de outros sobre o tema, a Wikipédia criou a figura do administrador, um utilizador que, pela quantidade e qualidade das suas contribuições, granjeou um estatuto reconhecido pelos seus pares. É-lhe dado o poder de bloquear temporariamente a edição de um

² Esta página não é visível para quem estiver apenas a consultar a enciclopédia. No entanto, basta escolher a opção “editar artigo” para a visualizar.

determinado tópico, banir contribuidores pela prática reiterada de conduta imprópria e apagar trechos que infrinjam a lei dos direitos de autor ou incluam conteúdos impróprios. É de salientar que estas intervenções apenas têm o intuito de impedir actos de vandalismo e violações das regras de conduta e não decidir o que irá constar no artigo em causa – um pouco à semelhança da justificação da existência da Polícia numa democracia [Lih 2009].

Para [Gobillot 2009], as potenciais falhas da política editorial da Wikipédia têm, apesar disso, dois grandes benefícios. O primeiro passa pelo facto de o grande número de contributos potenciar a recolha de conhecimento que até então não tinha sido explorado, providenciando uma visão mais abrangente e com diferentes perspectivas do tema do artigo em causa, por contraponto à visão unilateral do editor. O segundo prende-se com a inexistência de censura editorial que possibilita a abordagem de temas e assuntos de interesse comum mas sem interesse académico e que, numa política editorial clássica são, normalmente por decisão editorial, deixados de fora do conteúdo de uma enciclopédia.

[Rousseau 1762] defende no seu livro *O Contrato Social* que o poder deve ficar na mão dos cidadãos (*Multitudo*), enquanto o conceito de estado Leviatã de [Hobbes 1651] pressupõe que os cidadãos (*Populus*) devem abdicar da sua liberdade em favor do estado, de modo a conseguirem viver em sociedade.

As teorias defendidas por Rousseau e Hobbes podem, *latus sensus*, ajudar a compreender o modelo de funcionamento das enciclopédias em estudo, pondo em evidência as diferenças de pensamento que lhes estão subjacentes, em termos de confiança na natureza humana, na pluralidade, e na necessidade (ou não) de delegar a vontade individual num governante (editor).

Ao analisar o modelo de criação de conteúdos da Britânica, e se entendermos os editores desta como guardiões da qualidade do produto, podemos questionar-nos sobre os perigos que advêm de estes falharem, como [Sokal 1996] demonstrou³ em relação ao *peer review*.

O senso comum alerta-nos que o modelo da Wikipédia não deveria, de todo, funcionar, como afirma [Cohen 2005] num artigo⁴ publicado no *The New York Times*, em que resume a questão abordada neste artigo, ao transcrever um comentário de um contribuidor da Wikipédia: “*The problem with Wikipedia is that it only works in practice. In theory, it can never work.*”

3. Metodologia

O problema da qualidade da informação contida na Wikipédia já foi objecto de estudo por parte da comunidade científica. Foram deixados de fora deste levantamento, efectuado em 2010, os artigos de opinião e os estudos cujo principal enfoque não seja a problemática da qualidade da informação. Todos os estudos encontrados concluem que a Wikipédia tem uma qualidade inferior às enciclopédias com que é comparada, deixando no entanto transparecer que a qualidade científica dos artigos desta é superior ao que seria expectável.

Na Tabela 1 estão elencados os estudos encontrados e um resumo das metodologias utilizadas. A análise destas põe em evidência as diferenças entre o presente estudo e os anteriormente realizados.

³ Em 1994, Alan Sokal submeteu um ensaio à revista *Social Text*, premeditadamente escrito de modo a agradar à corrente de pensamento dos editores da revista e estava bem escrito e estruturado. No entanto, continha várias incongruências e disparates deliberados, chegando mesmo a afirmar que o significado da gravidade mais não era que uma ficção e que dependia da perspectiva cultural. Em 1996, depois de analisado pelo rigoroso crivo dos editores da revista, o artigo foi aceite para publicação e publicado num número especial da revista dedicada às guerras da ciência.

⁴http://www.nytimes.com/2007/04/23/technology/23link.html?_r=1&ex=1178510400&en=c0eb1b23e5c579f7&ei=5070

Autor	Título	Avaliação	Nº Artigos	Nº Avaliadores	Amostragem	Prova Cega	Área	Comparação
<i>The Guardian</i> (2005)	"Can you trust Wikipedia?"	N	7	7	C	Não	Global	Não comparativo
Giles (2005)	"Internet Encyclopedias Go Head to Head"	E	42	42	C	Sim	Global	<i>Britannica</i>
Rosenzweig (2006)	"Can History be Open Source? Wikipedia and the Future of the Past"	N	25	?	C	Não	História	<i>Britannica</i> <i>Encarta</i> <i>American</i> <i>National</i>
Devgan et al. (2007)	"Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference"	E	35	3	C	Não	Medicina	Obras referência
Bragues (2007)	"Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries"	N	7	1	C	Não	Filosofia	Obras referência
Pender et al. (2008)	"Putting Wikipedia to the Test: A Case Study"	N	3	?	C	Não	Medicina	Obras referência
Clauson et al. (2008)	"Scope, Completeness, and Accuracy of Drug Information in Wikipedia"	E	8	?	C	Não	Farmácia	<i>Medscape Drug Reference*</i> (MDR)
Rector (2008)	"Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth"	N	9	1	C	Não	História	<i>Britannica</i> <i>Dictionary of American History*</i> <i>American</i> <i>National</i>
Este Estudo	"Colaboração em Massa ou Amadorismo em Massa?"	N	195	66	A	Sim	Global	<i>Britannica</i>

(N) Atribuição Nota numa Escala; (E) Contagem Erros; (C) Conveniência; (A):Aleatória; (*) Obra Referência

Tabela 1 – Estudos sobre a qualidade da Wikipédia

A metodologia utilizada está na base das críticas a que alguns dos estudos referidos foram sujeitos, especialmente quando as suas conclusões apontavam numa direcção diferente daquela que seria expectável, pelo que a pertinência e o contributo deste trabalho para a clarificação do tema em estudo ficam assim indissociáveis da credibilidade que a metodologia a adoptar trazer às suas conclusões.

[Vanderstoep e Johnston 2009] descrevem e discutem as vantagens e condicionantes dos vários aspectos que caracterizam os métodos de investigação comumente utilizados. Partindo destes princípios gerais, vamos de seguida elencar os aspectos particulares da metodologia a utilizar neste caso, que devem ser tidos em conta de modo a não limitar o âmbito e a aceitação das conclusões a que esta conduzir:

- ☐ Comparar apenas itens comparáveis: como termos de comparação para aferir a qualidade relativa da Wikipédia não deverão ser utilizadas obras de referência, mas sim enciclopédias generalistas, de modo a garantir que estamos a comparar produtos que pertencem ao mesmo segmento e que são alternativas mútuas entre si.
- ☐ Caracterizar toda a população: o âmbito do estudo deve ser, se possível, global e não cingir-se apenas a uma área do conhecimento, pois só assim poderemos concluir sobre a qualidade da Wikipédia como um todo.
- ☐ Isenção: os avaliadores não devem saber a proveniência dos artigos que avaliam (prova cega), de modo a garantir que não se deixam influenciar por ideias preconcebidas que possam ter.
- ☐ Representatividade da amostra: a amostra deve ser representativa da população que pretendemos estudar, recolhida aleatoriamente e com uma dimensão que garanta a margem de erro que se pretende, de modo a que não se possa afirmar que esta não é representativa da população ou que os elementos escolhidos para análise favorecem um dos pratos da balança. Com isto não queremos dizer que a utilização de amostras de conveniência ou outras não seja válida, mas tão-somente que, no ponto de vista de uma abordagem positivista do problema, esta metodologia conduz a resultados mais fidedignos.
- ☐ Utilização de uma escala ajustada ao que queremos medir: a qualidade de cada elemento da amostra (artigo) deve ser aferida por um valor dentro de uma escala (nota), resultante de uma apreciação global e comparada com a qualidade do seu par proveniente da enciclopédia generalista a utilizar na sua comparação. Embora a contagem de erros seja precisa para aferir um ditado, per se, parece-nos curta para avaliar uma redacção.
- ☐ Utilizar instrumentos de medida fiáveis: as avaliações devem ser efectuadas por peritos na área do artigo em causa, cuja credibilidade não possa ser posta em causa.

Os comentários efectuados às metodologias utilizadas nos estudos atrás elencados não pretendem indiciar que estas não são válidas, mas sim alertar para as limitações às conclusões e abrangência que a sua aplicação acarreta. Iremos de seguida debruçar-nos sobre as escolhas que foram feitas em termos metodológicos para cumprir estes requisitos.

Dimensão da Amostra

A amostra a utilizar deverá ser estatisticamente representativa dos artigos que compõem a Wikipédia, definindo-se à partida uma margem de erro e um intervalo de confiança que conduzam a um compromisso aceitável entre a precisão do estudo e a dimensão da amostra a tratar.

Tendo em conta os mais de três milhões de artigos que compõem a Wikipédia, [Ruane 2005] aconselha uma amostra de cerca de 750⁵ artigos, enquanto [Gaur e Gaur 2009] consideram que uma amostra excelente deverá conter 500 artigos, embora uma amostra de 200-300 seja considerada adequada. Ambos os autores usam a *rule of thumb* para chegar a estes valores.

⁵ 0,025% por cada milhão de elementos do universo a amostrar.

[Adams, Raeside, e Khan 2007], [Given 2008], [Urdan 2005] e [Vanderstoep e Johnston 2009] propõem a utilização da seguinte fórmula:

$$n_0 = Z^2 \frac{p(1-p)}{d^2}$$

n_0 = Dimensão da amostra, Z = Valor standard da normal,

α = Nível de significância, p = Proporção da população e d = margem de erro.

Na Figura 2, podemos ver a variação da dimensão da amostra com a margem de erro para os níveis de significância $\alpha=95\%$ e $\alpha=99\%$, a que correspondem os valores de Z^2 de 1,96 e 2,58 respectivamente, e para os valores de p de 1/3 para três alternativas⁶ de resposta no que diz respeito à classificação dos artigos e de 1/5 para cinco. Caso a dimensão da amostra obtida fosse superior a 5% do universo, poderíamos diminuir a dimensão desta. No caso em análise, tal não se verifica.

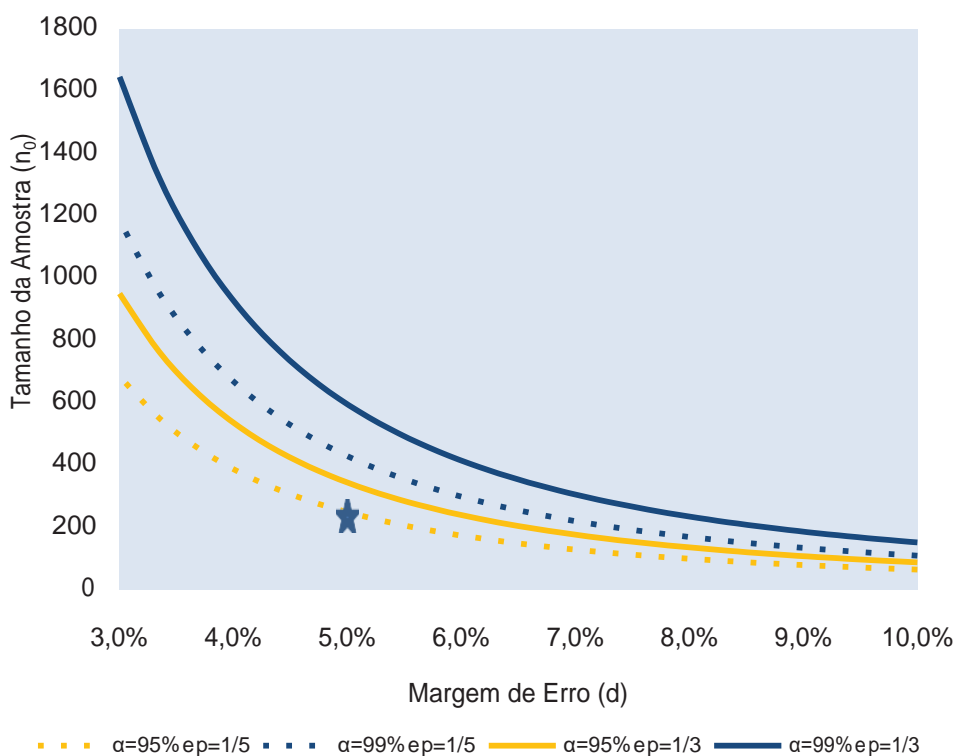


Figura 2 – Estudos sobre a qualidade da Wikipédia

Tendo em conta o comumente aceitável neste tipo de estudos [Gaur e Gaur, 2009], em termos de nível de significância e margem de erro, as dificuldades em escolher entre muitas alternativas de classificação e as implicações que a dimensão da amostra tem na exequibilidade deste trabalho, optámos por utilizar um nível de significância de 95%, uma margem de erro de 5%, e cinco alternativas de classificação. Nestas condições ($\alpha=95\%$, $Z=1,96$, $p=1/5$ e $d=5\%$), a dimensão da amostra a estudar (n_0) será de 245 pares de artigos – 490 avaliações.

O processo de extracção aleatória artigos das versões *online* de ambas as enciclopédias realizou-se em Janeiro de 2010, tendo sido recolhidos os títulos, a categoria⁷, o *URL* e o conteúdo das entradas, tendo sido deixadas de fora as secções *Related Articles* e *External Web sites* na Britânica e os seus correspondentes na Wikipédia, *Articles Related*, *See Also* e *External Links*.

⁶ Vanderstoep, S. & Johnston, D. (2009), p.31.

⁷ A categoria foi apenas recolhida na Britânica, visto ser esta a única a disponibilizar este tipo de informação.

Devido à forma como funciona o hipertexto, corríamos o risco de, a partir de um artigo e explorando todas as suas ligações e as ligações das ligações, obter todo o conteúdo da Enciclopédia partindo apenas de um artigo.

Como a Wikipédia contém vinte e cinco vezes mais entradas do que a Britânica, teria sido mais fácil extrair aleatoriamente o artigo da Britânica, procurando seguidamente a entrada correspondente na Wikipédia. Contudo, dado que a Britânica não permite a selecção aleatória de artigos, tivemos de usar o procedimento oposto, que conduziu à extracção de 6.382 entradas da Wikipédia, até serem encontradas 245 entradas⁸ que estivessem simultaneamente presentes nas duas enciclopédias.

Critérios avaliação:

Platão definia conhecimento como uma crença verdadeira com uma razão (*true belief with an account*); esta definição foi adaptada e hoje é comumente aceite definir conhecimento como uma crença verdadeira justificada (*justified true belief*). [Floridi 2004] defende que a informação deve ser definida como verdadeira – o que levanta a questão do que chamar à informação não verdadeira e mesmo a atribuir à palavra verdade um significado imutável no tempo e independente do observador. Quando alguém procura informação, espera que a informação que obtém seja correcta; no entanto, a veracidade desta não está contida no seu significado literal [Fetzer 2004].

Apesar de o conceito de informação variar consoante o autor e o contexto, parece razoável seguir a definição de [Dunn 2008] e assumir, no âmbito deste estudo, que a informação só se transforma em conhecimento quando satisfaz os três testes de Platão (*believed, justified, true*). A quantidade da informação que é, também, conhecimento, será então o nosso instrumento de medida para avaliar a colaboração em massa, enquanto metodologia credível na produção de conhecimento.

Um dos factores críticos no sucesso deste estudo é a escolha do método de aferir e quantificar a qualidade dos artigos, sendo necessário encontrar critérios suficientemente abrangentes que sejam aplicáveis a todo o tipo de conteúdos. [Wang e Strong 1996] propõem uma classificação dos atributos a que os utilizadores dão mais importância no que diz respeito à qualidade dos dados, agrupando estes em quatro grupos: Intrínsecos, Contextuais, Representacionais e Acessibilidade. Os atributos englobados nos grupos *Representacional* e *Acessibilidade* estão directamente dependentes do sistema que suporta a informação – o que, no caso em estudo, é similar em ambos os casos. O grupo *Contextual*, como o próprio nome indica, refere-se a características que variam com o consumidor da informação e o contexto em que esta será utilizada, que também não é relevante para este estudo. Assim iremos cingir-nos apenas às propriedades intrínsecas da informação.

Se considerarmos apenas as características intrínsecas constatamos que, para os consumidores de informação, a credibilidade é o factor mais importante, seguido da exactidão e da objectividade [Wang e Strong 1996]. Com base nesta informação, revela-se essencial que, no estudo a realizar, as entradas das enciclopédias omitam a sua proveniência, de modo a evitar que os peritos não utilizem a credibilidade da fonte nas suas avaliações.

Escolha Avaliadores

⁸Na altura da amostragem, os dados oficiais indicavam que a Wikipédia continha 25 vezes mais artigos do que a Britânica; se dividirmos 6.382 por 245, obtemos aproximadamente uma relação de 24:1 (erro de 4%)

O estudo de [Wang e Strong 1996] dá-nos pistas importantes na escolha do perfil dos peritos a convidar para avaliar os artigos. Quem se dedica ao ensino está habituado a avaliar conteúdos produzidos pelos alunos – pondo de lado a credibilidade da fonte – e concentrando-se apenas na exactidão e objectividade dos conteúdos, o que torna este grupo profissional ideal para realizar uma avaliação deste tipo.

Como será pedido a cada perito que avalie sempre pares de artigos, não se põe o problema de obtermos resultados não comparáveis, pois o contexto, grau de exigência e critérios são invariáveis dentro de cada par analisado.

Escala Avaliação

Os peritos avaliarão os artigos, utilizando uma escala de Likert de cinco níveis (Mau, Medíocre, Suficiente, Bom e Muito Bom), aos quais faremos corresponder os valores {1,2,3,4,5}. Como o número de alternativas de resposta condiciona a dimensão da amostra a estudar, como veremos mais à frente, e um maior número de opções iria complicar ainda mais a vida aos avaliadores, optámos por esta alternativa, que nos parece um bom compromisso [Likert 1932].

De modo a normalizar eventuais diferenças nos graus de exigência dos avaliadores, iremos transformar a escala original, na diferença entre as avaliações de cada par de artigos. Nesta nova escala {-4,-3,-2,-1,0,1,2,3,4} os valores positivos indicam uma qualidade superior da Wikipédia em relação à Britânica e os negativos o oposto, sendo o valor absoluto a ordem de grandeza da diferença de qualidade, podendo ser interpretada qualitativamente de acordo com a Tabela 2.

Qualidade Informação Wikipédia quando comparada com a Britânica		
Diferença entre as notas	4	Incomparavelmente Melhor
	3	Muitíssimo Melhor
	2	Muito Melhor
	1	Melhor
	0	Equivalente
	-1	Pior
	-2	Muito Pior
	-3	Muitíssimo Pior
	-4	Incomparavelmente Pior

Tabela 2: Significado das Diferenças

Número de Avaliadores por artigo

A escolha de apenas utilizar um avaliador por cada par de artigos prende-se com a elevada dimensão da amostra (245 pares de artigos – 490 avaliações). A utilização de três ou cinco avaliadores por cada par iria fazer disparar o número de avaliadores envolvidos tornando o estudo inexecutável.

Para obviar este problema foi utilizada uma metodologia que passou pela obtenção de uma amostra não aleatória, de pequena dimensão, susceptível de ser avaliada por múltiplos avaliadores de modo a concluir sobre a variabilidade das respostas destes e, partindo daí, usar essa incerteza nas avaliações com apenas um avaliador. Embora não seja possível provar que a variação das respostas se iria manter constante em todos os ramos da ciência, esta é a aproximação possível, tendo em conta o número de avaliadores com credibilidade e boa vontade suficientes que foi possível encontrar para colaborar neste estudo.

Por razões de conveniência, e presumida uma maior facilidade em recrutar avaliadores na área de Gestão, foi este o tema escolhido para a selecção dos artigos, tendo sido recrutado um painel de 12 avaliadores – Professores na área de Gestão. Para a escolha dos temas dos artigos foi utilizado um suplemento da revista *Executive Digest*, em que são elencados 50 conceitos de gestão [Cardoso 1997], para obter um leque de assuntos alargado e não muito específico, de modo a serem susceptíveis de ter sido incluídos em enciclopédias generalistas.

O processo de constituição da amostra é similar ao do estudo principal, em que são apenas seleccionados os temas que constam em ambas as enciclopédias, com a diferença de que o processo não termina quando for encontrado um número de pares predeterminado à partida, mas sim quando forem analisados os 50 temas elencados na revista.

Os resultados da procura dos 50 temas de Gestão encontraram 44 artigos (88%) na Wikipédia e apenas 4 artigos na Britânica⁹ (8%). Os quatro pares de artigos versam sobre os temas *Just in Time* (JIT), *Activity Based Costing* (ABC), *Outsourcing* (OUT) e *Total Quality Management* (TQM). Foram submetidos à avaliação dos doze Professores de Gestão – 96 avaliações.

Os resultados, em termos de diferença de nota, para cada tema encontram-se sumariados na Figura 3.

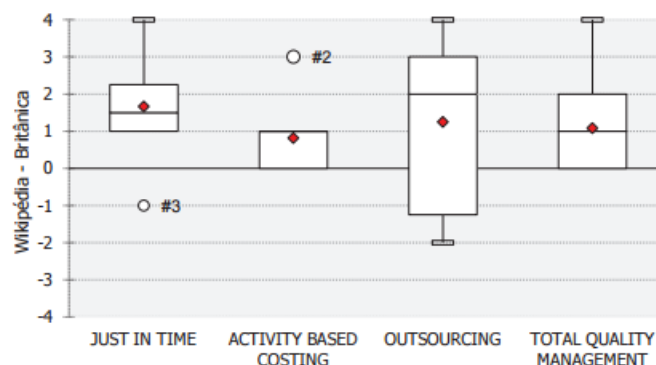


Figura 3 – Boxplots da variabilidade das respostas por tema

Em termos médios, os docentes avaliaram com uma nota de $3,6 \pm 0,3^{10}$ os artigos provenientes da Wikipédia e com $2,4 \pm 0,2^{10}$ os seus homónimos da Britânica. Em termos globais, a diferença¹¹ média entre as notas atribuídas pelos múltiplos avaliadores, foi de $+1,2 \pm 0,4$.

A variabilidade¹² das diferenças entre as respostas dos avaliadores dentro do mesmo par de artigos, tiveram um desvio médio absoluto da média de 1,1 e uma mediana do desvio absoluto da mediana de 0,9. Estas medidas de dispersão – a razão de ser deste estudo complementar – irão posteriormente ser utilizadas para simular a variação de resultados que seria expectável obter caso tivéssemos 12 avaliadores por cada par de artigos no estudo global.

Embora não seja possível provar que a variabilidade encontrada seja constante em todos os ramos da ciência, e seja similar à encontrada para a Gestão, esta é a aproximação possível, tendo em conta o número de avaliadores com credibilidade e boa vontade suficientes que foi possível encontrar para colaborar neste estudo.

⁹ Estes quatro temas foram também encontrados na Wikipédia.

¹⁰ Incerteza estatística = $\frac{\delta}{\sqrt{n}}$

¹¹ Escala [-4, 4]

¹² Média Desvio Absoluto Média = $\frac{1}{n} \sum_{i=1}^n |X_i - \mu|$

Mediana Desvio Absoluto Mediana = Mediana $|x_i - M|$

Isenção dos Avaliadores

Embora nada nos fizesse duvidar à partida da isenção dos avaliadores escolhidos foi elaborado junto da população em que iriam ser recrutados os avaliadores um pequeno inquérito de modo a caracterizar os hábitos de consulta de informação e a confiança que deposita na informação obtida, através de instrumentos similares à Wikipédia – Fontes Alternativas - Vox Populi – e à Britânica – Fontes Clássicas -Vox Expertorum.

O inquérito (ver Figura 4) foi deliberadamente construído com um número mínimo de questões e sem perguntas que fizessem a caracterização dos inquiridos, de modo a tentar maximizar a taxa de resposta, que é habitualmente muito baixa nestes casos.

Caracterização Hábitos de Consulta Informação

1. Quando necessita obter informação/aprofundar um determinado tema, onde procura primeiro informação?

Motor de Busca / Wikipedia ✓

Enciclopédia / Dicionário / Obra de Referência

2. Confia na Informação obtida através de um Motor de Busca / Wikipedia?

1 2 3 4 5

Desconfio completamente Confio plenamente

3. Confia na Informação obtida através de uma Enciclopédia / Dicionário / Obra de Referência?

1 2 3 4 5

Desconfio completamente Confio plenamente

Figura 4 – Inquérito

Apesar da baixa taxa de resposta (9,6%) foram obtidas 63 respostas, que permitiram concluir que 75% dos docentes utiliza numa primeira abordagem na procura de informação fontes alternativas, embora apenas 14% deposite nestas uma confiança igual ao superior à que deposita nas fontes clássicas. Na Figura 5 podemos ver os resultados do inquérito.

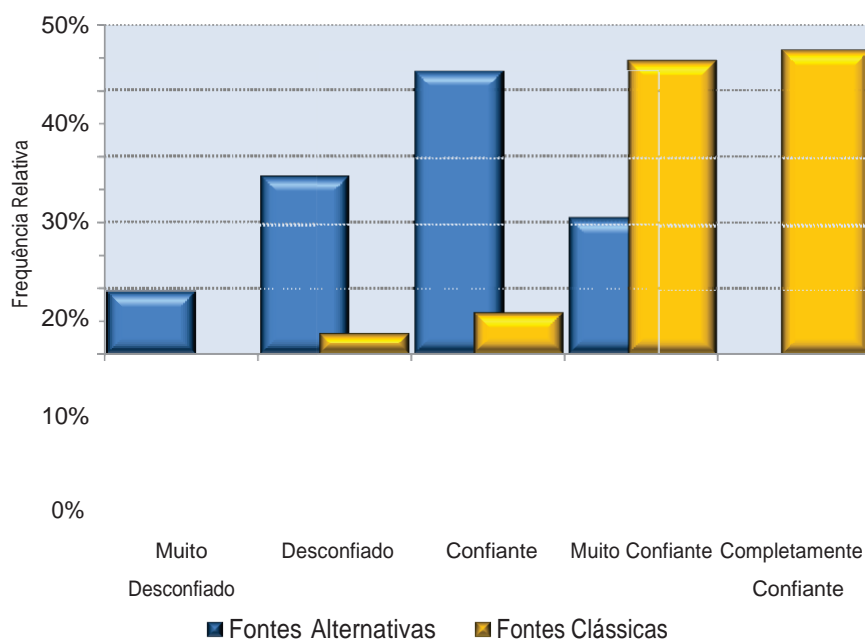


Figura 5 – Estudo sobre a confiança nas fontes de informação

Os resultados obtidos mostram claramente, a desconfiança por parte do universo dos avaliadores, na qualidade da informação disponibilizada pela Wikipédia.

4. Apresentação de resultados

A amostra inicial de 245 pares de artigos foi dividida em quatro grandes grupos: Artes & Entretenimento, Ciência & Tecnologia, História & Sociedade e Viagens & Geografia, sendo posteriormente os artigos subdivididos consoante a área do conhecimento a que pertenciam de modo a serem distribuídos pelos avaliadores.

Foram encontrados avaliadores para todas as áreas do conhecimento constantes das três primeiras categorias, tendo sido avaliados todos os pares de artigos. Em relação ao grupo Viagens & Geografia, os 57 pares de artigos desta categoria são compostos por um artigo de História, um de Pintura, um de Arquitectura e 54 de Geografia.

Se analisarmos os valores médios das avaliações divididas pelas categorias anteriormente apresentadas, temos (Figura 6):

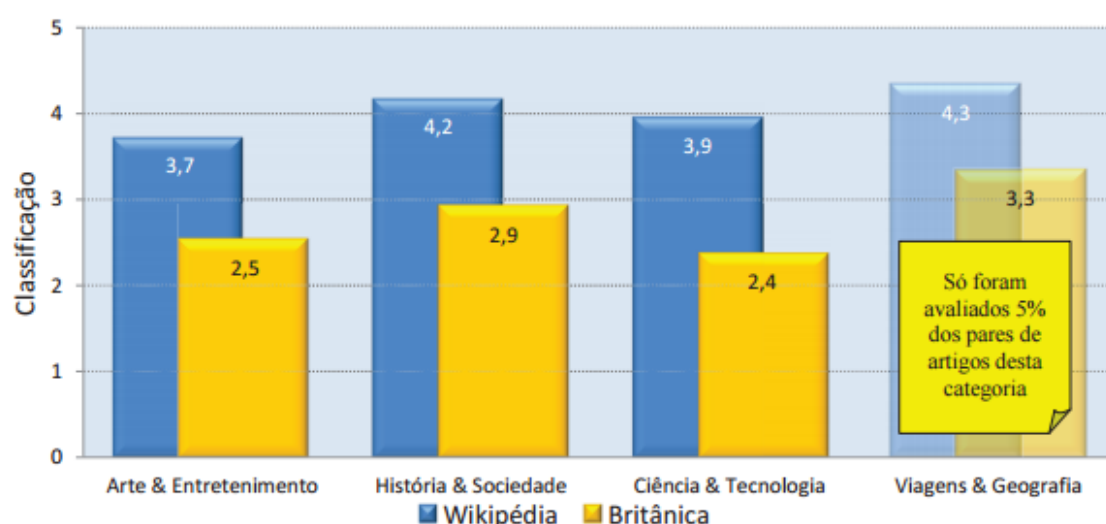


Figura 6 - Média das Avaliações Wikipédia e Britânica Divididas por Categorias

Apesar de todos os esforços desenvolvidos, foi impossível encontrar avaliadores na área da Geografia dispostos a colaborar com o estudo. Os artigos em causa são, na sua totalidade, compostos por descrições de cidades, regiões, países, montanhas e rios; e como não foram avaliados, limitam as conclusões deste estudo a todas as áreas do conhecimento à excepção da Geografia.

Na Figura 7, a área das esferas é proporcional ao número de avaliações que atribuíram o mesmo binómio de classificações (W,B)¹³. Das 25 combinações possíveis, apenas os binómios (1,1), (1,3) e (1,5) não foram utilizados pelos avaliadores.

¹³ A título de exemplo, o binómio (4,2) corresponde à atribuição de uma nota de 4 valores para a Wikipédia (abcissas) e de 2 valores à Britânica (ordenadas) e ocorreu 27 vezes.

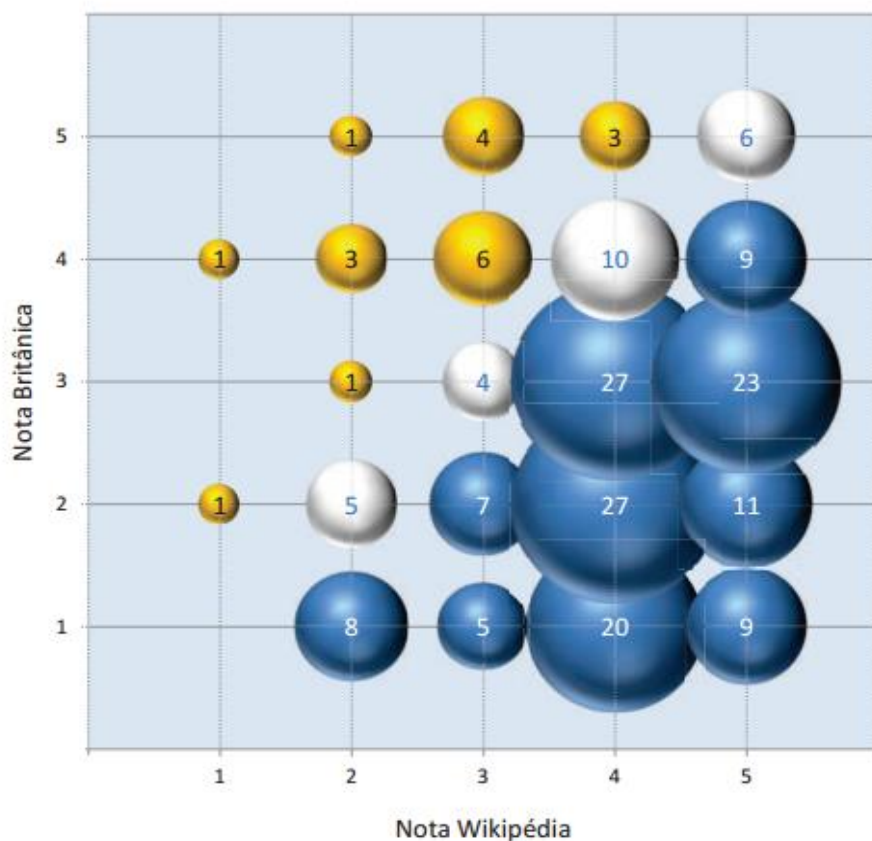


Figura 7 - Avaliações agrupadas por binómio de notas iguais

Iremos de seguida testar se as diferenças encontradas são estatisticamente significantes. Para tal, vamos começar por testar a normalidade das notas atribuídas à Wikipédia e à Britânica, testando as hipóteses:

H_0 : a variável é normalmente distribuída;

H_1 : a variável não é normalmente distribuída.

Foi efectuado o teste de normalidade, com os seguintes resultados (Tabela 3):

	Wikipédia		Britânica	
	Shapiro-Wilk	Probabilidade	Shapiro-Wilk	Probabilidade
Notas	0,834	0,000	0,902	0,000

Tabela 3: Teste Shapiro-Wilk – Avaliações Wikipédia e Britânica

A probabilidade associada ao teste de normalidade é inferior a 0,05 em todas as variáveis; logo, rejeita-se a hipótese nula nos dois casos, concluindo-se que nenhuma das variáveis é normalmente distribuída.

Iremos, pois, utilizar um teste não paramétrico para verificar se existem diferenças estatisticamente significativas entre as variáveis. Aplicou-se o teste de *Wilcoxon* para amostras emparelhadas, usando o seguinte teste de hipóteses:

H_0 : $\mu_1 = \mu_2$ (as médias são iguais);

H_1 : $\mu_1 \neq \mu_2$ (as médias são diferentes).

Foram obtidos os seguintes resultados (Tabela 4):

	Média			Probabilidade (bicaudal)
	Wikipédia	Britânica	Z	
Notas	4,0	2,6	-9,235	0,000

Tabela 4: Teste de Wilcoxon – Avaliações Wikipédia e Britânica

Como o valor da probabilidade é inferior a 0,05, rejeita-se a hipótese nula e conclui-se que existe diferença significativa entre as médias, ou seja, as diferenças encontradas são estatisticamente significativas.

Em termos globais – não considerando os artigos de Geografia - constata-se que em 90% dos pares de artigos avaliados a Wikipédia teve uma classificação igual ou superior à da Britânica (Figura 8).

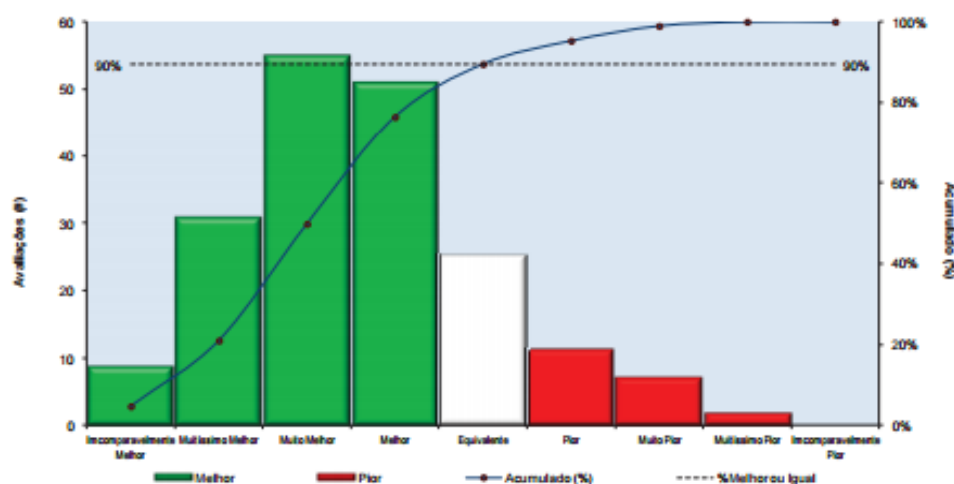


Figura 8: Qualidade informação da Wikipédia quando comparada com a Britânica

Se considerarmos a média (+1,4) e a mediana (+1,0) das diferenças entre as avaliações da Wikipédia e da Britânica e as combinarmos com as medidas de variabilidade obtidas nas avaliações dos artigos de gestão poderemos ter uma aproximação do resultado que obteríamos se tivéssemos utilizado, também aqui, 12 avaliadores¹⁴ por artigo. Tendo em conta que a introdução de múltiplos avaliadores traduziu-se num desvio médio absoluto da média de 1,1 e numa mediana do desvio absoluto da mediana de 0,9; podemos inferir que a média das diferenças entre as avaliações situa-se no intervalo [+0,3 ; +2,5] e a mediana em [+0,1 ; +2,9].

Seja qual for a medida de variabilidade escolhida, o resultado do intervalo é sempre positivo, ou seja, constata-se que, em média e em mediana, as classificações atribuídas à Wikipédia são sempre superiores às da Britânica, mesmo que cada par de artigos tivesse sido avaliado por múltiplos avaliadores.

¹⁴ Teria sido necessário efectuar 5.880 avaliações.

5. Conclusões

Este estudo é uma tentativa de ajudar a clarificar a discussão em torno do rigor e qualidade científica dos conteúdos da Wikipédia e contribuir para uma melhor percepção da qualidade de uma ferramenta utilizada por milhões diariamente e, conseqüentemente, concluir sobre a viabilidade do modelo de funcionamento que lhe está subjacente – a colaboração em massa.

Assim, e ao contrário da opinião preconcebida da maioria, mostrou-se que a Wikipédia tem uma qualidade média superior à sua concorrente directa, a Britânica, o que a transforma num caso de sucesso e permite concluir que a colaboração em massa é um mecanismo que permite criar e gerir conteúdos de qualidade, pelo menos dentro do contexto e das condicionantes do caso estudado.

A característica aberta da tecnologia, que permite que todos possam alterar os conteúdos da Wikipédia e que é o cerne do seu sucesso, é simultaneamente uma ameaça à sua credibilidade, a par da possibilidade de amadores criarem conteúdos sobre temas que não dominam. Como contraponto a estas legítimas críticas, a Wikipédia dispõe, essencialmente, da colaboração em massa na defesa contra ataques de vandalismo e na vontade de manter a qualidade do projecto por parte de quem colabora.

A percepção da qualidade de um produto baseia-se na confiança que os consumidores têm neste: a confiança não pode ser imposta ou comprada, tem de ser ganha ao longo do tempo. Apesar do crescimento constante do número de utilizadores e de artigos disponíveis, o projecto – e a filosofia que o suporta – ainda não conquistou a mesma confiança que o poria, pelo menos, ao nível de outras enciclopédias generalistas que há muito ganharam o reconhecimento e a confiança dos consumidores.

O facto de a utilização da Wikipédia ser gratuita e a da Britânica ser paga poderá justificar, em parte, a enorme diferença entre o número de utilizadores e mesmo a percepção sobre a qualidade destas, que ficou patente nas respostas ao inquérito sobre os hábitos de consulta de informação.

A opinião preconcebida da qualidade dos conteúdos da Wikipédia e os resultados da avaliação desta numa prova cega, dentro de um mesmo grupo, contribui para reforçar a isenção do painel de avaliadores e a conseqüente relevância dos resultados obtidos, ou seja, eventuais faltas de isenção por parte de alguns avaliadores tenderiam a favorecer a Britânica.

Se na Wikipédia excluirmos os poucos artigos que originam controvérsia e guerras de alterações, exacerbadas por paixões políticas, religiosas ou outras, a esmagadora maioria dos artigos é construída por utilizadores que dominam e se interessam pelo tema em causa e que tentam criar conteúdos de qualidade. Não nos parece plausível que alguém sem conhecimentos específicos sobre um determinado tema tenha motivação para criar e editar conteúdos em conjunto com outros que dominam o assunto em causa.

Uma explicação para o sucesso da Wikipédia poderá passar pelo altruísmo de indivíduos que se reúnem à volta de temas sobre os quais partilham a mesma paixão, sendo expectável possuírem

uma dose assinalável de conhecimento sobre estes. Por outras palavras, no caso estudado, a colaboração em massa parece auto-organizar-se, resultando dessa organização uma auto-avaliação e uma autocorreção efectuada pelos pares.

Cabe aos gestores e líderes destas comunidades sociais, assentes na colaboração em massa, não o papel de moldar a comunidade e o modo como esta se organiza, mas sim tornar o projecto atractivo, de modo a agregar o maior número possível de contribuidores, condição necessária – embora não suficiente – ao sucesso do projecto.

6. Referências

- Adams, J., Raeside, R. & Khan, H., *Research Methods for Graduate Business and Social Science Students*, Sage Publications Inc., New Delhi, India, 2007.
- Anderson, C., *The Long Tail: Why the future of business is selling less of more*, Hyperion Books, New York, USA, 2006.
- Benkler, Y., *The Wealth of Networks How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven, USA, 2006.
- Bragues, G., *Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's entries on seven great minds*, University of Guelph-Humber, <http://ssrn.com/abstract=978177> (12 de Março de 2011), 2007.
- Broughton, B., *Wikipedia: The Missing Manual*, O'Reilly Media, First Edition, Sebastopol, California, USA, 2008.
- Cardoso, J., "Manual de Gestão : Guia dos Conceitos de A a Z - Executive Digest", *Abril/Controljornal*, (1997).
- Christensen, C., *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston, USA, 1997.
- Clauson, K., Polen, H., Boulos, M., & Dzenowagis, J., "Scope, Completeness, and Accuracy of Drug Information in Wikipedia", *The Annals of Pharmacotherapy*, Vol. 42, N. 12, (2008), 1814-1821.
- Cohen, M., *The Cambridge Companion to Ancient Greek Law*, Cambridge University Press, Cambridge, UK, 2005.
- Devgan, L., Powe, D., Blakey, B., & Makary, M., "Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference", *Journal of the American College of Surgeons*, Vol. 205, n° 3, (2007), 76-77.
- Dunn, J., *Information in Computer Science in Philosophy of Information*, Vol.8, Edited by Adriaans, P., & Benthem, J., Elsevier, Amsterdam, The Netherlands, 2008.
- Fetzer, J., "Information: Does it Have to be True?", *Minds and Machines Journal*, Vol. 14, N° 2, Springer Netherlands, (2004), 223-229.
- Floridi, L., "Outline of a Theory of Strongly Semantic Information", *Minds and machines Journal*, Vol. 14, N° 2, Springer Netherlands, (2004), 197- 221.
- Gaur, A., & Gaur, S., *Statistical Methods for Practice and Research: A Guide to Data Analysis using SPSS*, Response Books, Business books from Sage (Second edition), New Delhi, India, 2009.
- Ghazawneh, A., "Managing Mass Collaboration: Toward a Process Framework", Master Thesis, School of Economics and Management, Lund University, Sweden, (2008).

- Giles, J., "Special Report: Internet Encyclopaedias Go Head to Head", *Nature*, Vol. 438, 14 December 2005, Nature Publishing Group, (2005), 900-901.
- Given, L., *The Sage Encyclopedia of Qualitative Research Methods*, Sage Publications Ltd., Thousand Oaks, California, USA, 2008.
- Gobillot, E., *Leadershift: Reinventing leadership for the age of mass collaboration*, Kogan Page Limited, Philadelphia, USA, 2009.
- Hobbes, T., *Leviathan*, Basil Blackwell, Oxford, England, 1957 [1651].
- Howe, J., *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*, Three Rivers Press, New York, USA, 2009.
- Lévy, P., *Collective Intelligence: Mankind's Emerging World in Cyberspace*, Perseus, Cambridge, USA, 1997.
- Lih, A., *The Wikipedia Revolution – How a Bunch of Nobodies Created the World's Greatest Encyclopedia*, Hyperion eBook, New York, USA, 2009.
- Likert, R., "A Technique for the Measurement of Attitudes", *Archives of Psychology*, 140, (1932), 1-55.
- Pender, M., Kaye, L., Lisa, K., Christopher, D., & Satyamurthy, A., "Putting Wikipedia to the Test: A Case Study", *Special Libraries Association Annual Conference Seattle*, Washington, University of Queensland and Department of Neurology, Royal Brisbane and Women's Hospital, (2008).
- Rector, L. H., "Comparison of Wikipedia and Other Encyclopedias for Accuracy, Breadth, and Depth in Historical Articles", *Reference Services Review Journal*, Vol. 36, nº1, (2008), 7-22.
- Rosenzweig, R., "Can History be Open Source? Wikipedia and the Future of the Past", *The Journal of American History*, Vol. 93, nº1, (2006), 117-146.
- Rousseau, J., *The Social Contract or Principles of Political Right*, Constitution Society, <http://www.constitution.org/jjr/socon.htm>, (12 de Agosto de 2012), 2012 [1762].
- Ruane, J., *Essentials of Research Methods: A Guide to Social Science Research*, Blackwell Publishing Ltd, Boston, USA, 2005.
- Shirky, C., *Here Comes Everybody: The Power of Organization Without Organizations*, The Penguin Press, New York, USA, 2008.
- Shirky, C., *Cognitive Surplus: How Technology Makes Consumers into Collaborators*, The Penguin Press, New York, USA, 2010.
- Sokal, A., "A Physicist Experiments with Cultural Studies" *Lingua Franca*, Vol. 6, nº4 (May-June), pp.62-64. Reprinted in *The Sokal hoax: The sham that shook the Academy*, edited by the editors of *Lingua Franca*, Lincoln: University of Nebraska, (1996), 49-53.

Sunstein, C., *Infotopia: How Many Minds Produce Knowledge*, Oxford University Press, New York, USA, 2006.

Surowiecki, J., *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Anchor Books, New York, USA, 2004.

Urdan, T., *Statistics in Plain English*, Lawrence Erlbaum Associates Publishers, New Jersey, USA, 2005.

Vanderstoep, S., & Johnston, D., *Research Methods For Everyday Life: Blending Qualitative and Quantitative Approaches*, Published by Jossey-Bass, San Francisco, USA, 2009.

Wang, R., & Strong, D., "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems*, Spring, Vol. 12, nº4, (1996), 5-34.