

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ADRIANA JOURIS

**Avaliação da qualidade de conferências
científicas através da visibilidade na Web**

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em Ciência
da Computação

Prof. Dr. José Palazzo Moreira de Oliveira
Orientador

Porto Alegre, março de 2011.

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Jouris, Adriana

Avaliação da qualidade de conferências científicas através da visibilidade na web / Adriana Jouris – Porto Alegre: Programa de Pós-Graduação em Computação, 2011.

76 f.:il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação. Porto Alegre, BR – RS, 2011. Orientador: José Palazzo Moreira de Oliveira.

1. Visibilidade na Web. 2. Avaliação de Conferência. 3. Classificação. 4. Webometria. I. Oliveira, José Palazzo Moreira de. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenador do PPGC: Prof. Álvaro Freitas Moreira

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Ao Prof. Dr. José Palazzo Moreira de Oliveira, por ter aceitado ser meu orientador, por estar sempre disponível para esclarecer minhas dúvidas e pelos seus valiosos ensinamentos.

Ao Instituto de Informática, pela infraestrutura disponibilizada, aos funcionários e colegas do Instituto, que sempre estiveram prontos para esclarecer nossas dúvidas e ajudar no que fosse preciso, e aos professores do PPGC.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	6
LISTA DE FIGURAS.....	7
LISTA DE TABELAS	8
LISTA DE GRÁFICOS	9
RESUMO.....	10
ABSTRACT	11
1 INTRODUÇÃO	12
1.1 Motivações.....	13
1.2 Objetivos.....	13
1.3 Contribuições	13
1.4 Organização do Texto	14
2 AVALIAÇÃO DE CONFERÊNCIA	125
2.1 Indicadores Bibliométricos	16
2.1.1 Deficiências dos indicadores Bibliométricos	20
2.2 Visibilidade na Web	24
2.2.1 Webometria	26
2.2.1.1 Objetivos.....	27
2.2.1.2 Trabalhos relacionados	28
2.2.1.3 Funcionamento da metodologia.....	29
2.2.1.4 Considerações	35
3 MODELO PARA AVALIAÇÃO DE CONFERÊNCIA CIENTÍFICA.....	36
3.1 Identificar os Indicadores	37
3.2 Coletar dados da conferências.....	38
3.2.1 Ferramentas de coleta de dados	38
3.2.1.1 Yahoo	38
3.2.1.2 Harzing's Publish or Perish	39
3.2.2 Coleta dos dados.....	40
3.3 Correlacionar os dados coletados com o valor do ranking da QUALIS	44
3.3.1 Análise dos dados	44
3.3.1.1 Coeficiente de correlação	44
3.3.1.2 Gráfico de Dispersão	46
3.4 Definir um modelo-padrão dos indicadores.....	50
3.5 Avaliar a qualidade das conferências a partir deste modelo.....	52
4 EXPERIMENTO	53
4.1 Resultados	54
5 CONCLUSÃO.....	58
5.1 Trabalhos Futuros	59
REFERÊNCIAS	60
APÊNDICE A AMOSTRA DE DADOS DAS CONFERÊNCIAS	67

APÊNDICE B CONFERÊNCIAS COM OS VALORES PADRONIZADOS.....	70
APÊNDICE C AMOSTRA DE DADOS DAS CONFERÊNCIAS AVALIADAS.	75

LISTA DE ABREVIATURAS E SIGLAS

FI	Fator de Impacto
ISI	<i>Institute for Scientific Information</i>
SI	Sistemas de Informação
GA	Google Acadêmico
CC	Ciência da Computação
WIF	<i>Web Impact Factor</i>

LISTA DE FIGURAS

Figura 2.1: Esquema da webometria	30
Figura 3.1: Etapas do processo de avaliação de qualidade.....	37
Figura 3.2: Interface do Yahoo Site Explorer.....	39
Figura 3.3: Interface de entrada de dados do Publish or Perish	40
Figura 3.4: Interface de consulta do Publish or Perish.....	41
Figura 3.5: Tabela de classificação do coeficiente de Pearson.....	45
Figura 3.6: Esquema da webometria adaptada para conferência.....	50

LISTA DE TABELAS

Tabela 3.1: Conferências analisadas.....	42
Tabela 3.2: Conferências com indicador QUALIS	43
Tabela 3.3: Grau de correlação de cada indicador.....	45
Tabela 3.4: Tabela dos pesos e padrões atribuídos aos indicadores GA.....	50
Tabela 3.5: Tabela dos pesos e padrões atribuídos aos indicadores h-index.....	51
Tabela 3.6: Tabela dos pesos e padrões atribuídos aos indicadores pagerank	51
Tabela 3.7: Tabela dos pesos e padrões atribuídos aos indicadores visibilidade	51
Tabela 3.8: Tabela dos pesos atribuídos aos estratos	52
Tabela 4.1: Classificação de conferências	53
Tabela 4.2: Comparação entre os resultados	56

LISTA DE GRÁFICOS

Gráfico 3.1: Correlação entre inlinks e valor ranking QUALIS.....	46
Gráfico 3.2: Correlação entre artigos e valor ranking QUALIS.....	47
Gráfico 3.3: Correlação entre h-index e valor ranking QUALIS	47
Gráfico 3.4: Correlação entre citações e valor ranking QUALIS.....	48
Gráfico 3.5: Correlação entre citações/artigo e valoring rank QUALIS	48
Gráfico 3.6: Correlação entre PageRank e valor ranking QUALIS	49
Gráfico 3.7: Correlação entre autores/artigo e valor ranking QUALIS	49
Gráfico 3.8: Distribuição dos estratos considerando o conjunto de Conferências	52
Gráfico 4.1: Correlação entre modelo BD e QUALIS	55
Gráfico 4.2: Correlação entre modelo Eng. e QUALIS	55
Gráfico 4.3: Correlação entre modelo SI e QUALIS	56

RESUMO

Com o aumento constante dos eventos científicos realizados surge a necessidade de se ter uma ferramenta para auxiliar na avaliação desses eventos de maneira rápida e precisa a fim de evidenciar a sua qualidade. A tarefa de avaliação atualmente em uso tem alta probabilidade de falhas e geralmente envolve muito trabalho humano afetando o custo. Assim, este trabalho apresenta um estudo das principais métricas para avaliação de conferências científicas utilizadas. Essas métricas foram separadas em dois grupos. O primeiro, estuda as características internas, os chamados indicadores bibliométricos, e o segundo grupo analisa a perspectiva externa, ou seja, os indicadores de visibilidade na web. A partir deste estudo, apresenta-se um modelo para avaliação de Conferências Científicas, na área de Ciência da Computação, com o principal objetivo de gerar um indicador de qualidade, analisando múltiplos critérios, tais como a quantidade de citações dos artigos daquela conferência e a quantidade de links que a apontam e utilizando as tecnologias web disponíveis para obtenção dos dados.

Este modelo tem o objetivo de contribuir para o processo de avaliação de conferências de uma maneira mais transparente. O resultado do *ranking* indicado pelo modelo é gerado a partir de dados obtidos das conferências científicas. Nos experimentos, foi possível observar o novo *ranking* dessas conferências baseados na visibilidade na web como uma nova maneira para medir a qualidade das conferências, podendo demonstrar que este método é adequado.

Palavras-Chave: Visibilidade na web, avaliação de conferência, classificação, webometria.

Assessment of scientific conferences quality by means of visibility on the web

ABSTRACT

With the steady increase of the number of scientific conferences it is necessary the existence of a tool to assist the quality evaluation of these events in a quick and accurate manner. The manual assessment task presently in use has a high probability of failures and the huge human work needed affects its cost. For that reason this work presents a study and the application of the main metrics for evaluating scientific conferences. These metrics were separated into two groups: the first studies the internal features, the bibliometric indicators, and the second group analyzes the external perspective or the visibility indicators in the web. From this study, we developed a model for the evaluation of Scientific Conferences in Computer Science with the main objective of generating a quality indicator, analyzing multiple criteria such as the number of citations of articles that conference and the number of links that point and using the web technologies available for data collection.

This model is intended to contribute to the evaluation process of conferences in a more clear and explicit way. The result of the ranking is generated from the obtained data of the scientific conferences. In these experiments, it was possible to verify the new ranking of these conferences based on the visibility in the web as a useful way for measuring the quality of the conference and thus demonstrating the suitability of this method.

Keywords: Visibility on the web, conference assessment, classification, webometrics.

1. INTRODUÇÃO

À medida que o número de eventos científicos cresce, o pesquisador se depara com a questão de como identificar bons eventos para publicar seus trabalhos, já que publicar em eventos com qualidade assegurada é um dos critérios mais importantes na área de Ciência da Computação para obter mais notas altas em avaliação de pesquisadores, grupos de pesquisa e até mesmo instituições. Podemos citar também a importância de divulgar os bons trabalhos como forma de acelerar o processo de descobertas científicas e tecnológicas, criando um ambiente colaborativo de discussão e troca de informações que podem influenciar as pesquisas em curso.

O desafio que surge então é como usar as tecnologias disponíveis, principalmente a partir da internet, para automatizar o processo de avaliação de eventos científicos, coletando dados (artigos, citações e assim por diante) em bases disponíveis na web, e comparando o nível de qualidade desses eventos com padrões pré-convencionados. E, através de um sistema-padrão que obtenha informações atualizadas instantaneamente, formular um modelo que permita identificar e ranquear os eventos de acordo com sua qualidade. Porém, o desenvolvimento dessas ferramentas tem um custo relativamente alto e grande probabilidade de falhas.

Dentre os inúmeros critérios testados e defendidos pelos estudiosos do assunto para inferir a qualidade de uma conferência, podemos citar: membros do comitê de programa, fator de impacto, tradição do evento, tópicos do evento, entre outros. Neste contexto, este trabalho tem por objetivo desenvolver um modelo para avaliação da qualidade de Conferências Científicas na área da Ciência da Computação baseando em características de visibilidade na web, visando indicar qual é o nível da qualidade científica de determinada conferência através da combinação de múltiplos critérios, tais como a quantidade de citações dos artigos daquela conferência e a quantidade de links que a apontam. Este modelo utilizará como referência o documento QUALIS da CAPES (CAPES, 2008) e análises estatísticas com objetivo de encontrar o coeficiente de correlação entre as variáveis que caracterizem a qualidade de uma conferência.

A idéia do trabalho é que, a partir deste modelo, possa ser definido um padrão de indicadores para compor um sistema de avaliação que permita inferir a qualidade de uma conferência e forneça ao usuário um *ranking* para auxiliar na escolha da melhor. Esta avaliação é realizada em função da sua qualidade intrínseca e por sua visibilidade externa.

1.1 Motivação

Uma das maiores preocupações para os pesquisadores em geral é encontrar canais de comunicação eficientes para divulgar o resultado do seu trabalho. Quando o canal a ser avaliado para a submissão é uma revista ou uma publicação regular, o trabalho de avaliar a qualidade é relativamente facilitado em função de se poder criar um esquema-padrão para isso. Porém, quando se trata da submissão de artigos em eventos como conferências ou congressos, esta tarefa se torna complexa, pois nos deparamos com questionamentos em relação à organização e classificação das incontáveis conferências que acontecem todos os anos em todo o mundo: como atribuir maior ou menor qualidade aos eventos, se não há um padrão seguido pelos organizadores?

A principal motivação para o presente trabalho vem da possibilidade de acrescentar uma solução que, mesmo não resolvendo a questão definitivamente, pelo menos apresente uma alternativa eficaz e mais avançada que as atuais para avaliar a qualidade de publicações e conferências.

1.2 Objetivos

O objetivo deste trabalho é desenvolver um modelo para avaliar a qualidade das Conferências Científicas, levando em consideração os indicadores mais utilizados atualmente nos trabalhos científicos e propondo novos indicadores.

Como objetivos específicos, é possível citar:

- Identificar indicadores já propostos nos trabalhos científicos com esta finalidade.
- Especificar um modelo para avaliação de Conferências Científicas: o foco é modelar as informações da conferência com base na perspectiva externa, que poderá indicar a qualidade para fins de recomendação.
- Realizar análises estatísticas objetivando encontrar o coeficiente de correlação entre as variáveis determinadas no modelo e as variáveis do documento da QUALIS visando caracterizar o grau de importância para avaliação da qualidade de uma conferência.
- Determinar um modelo-padrão dos indicadores visando classificar a qualidade da Conferência Científica.
- Identificar se é possível realizar esse tipo de avaliação da qualidade baseando-se neste modelo.

1.3 Contribuições

A principal contribuição deste trabalho é apresentar um modelo de avaliação da qualidade científica, baseado em visibilidade na *web*, o qual servirá como um novo indicador para conferências de CC, podendo realizar uma análise do *website* da conferência em nível mundial e resultar num *ranking* de qualidade.

Outra contribuição é apresentar um modelo baseado nas características *web*, que capturam aspectos importantes que não são considerados pelas métricas existentes (para conferências); apresentar e detalhar um conjunto de características que possam ser utilizadas como indicadores de qualidade para Conferências Científicas.

1.4 Organização do texto

O restante do documento está estruturado conforme descrito a seguir:

- No capítulo 2 são apresentadas algumas formas de avaliação de conferências, abordando este tema, de acordo com os trabalhos relacionados, mostrando o que foi desenvolvido por outros pesquisadores.
- No capítulo 3 apresenta-se o modelo elaborado para avaliação de conferências científicas, bem como, as ferramentas utilizadas para coleta dos dados do modelo e os dados coletados e as correlações entre os indicadores de qualidade;
- No capítulo 4 descreve alguns experimentos realizados e os resultados obtidos;
- As conclusões e trabalhos futuros são relatados no capítulo 5.

2. AVALIAÇÃO DE CONFERÊNCIA

É consenso que a avaliação da produção científica dos pesquisadores revela o estágio de amadurecimento das políticas de ensino e pesquisa de uma determinada instituição e até mesmo de um país. A comparação da evolução quantitativa e qualitativa desta produção ao longo do tempo fornece subsídios sobre o progresso do ente avaliado e serve para diagnosticar e propor eventuais melhorias no processo.

Dentre as opções que se colocam aos pesquisadores para divulgarem seus trabalhos, a inserção de artigos em revistas especializadas e a submissão de artigos em conferências de reconhecida qualidade são duas das formas que mais visibilidade trazem ao trabalho publicado. Assim, é importante que haja uma padronização, com a definição de elementos-chave para avaliar e identificar a qualidade de um evento científico.

Os *rankings* que mostram a avaliação da qualidade das revistas são de especial interesse para acadêmicos e para as instituições, pois estas costumam atribuir uma importância significativa às publicações em revistas de maior prestígio para avaliar seus pesquisadores, influenciando também nas promoções e no reconhecimento do seu trabalho pelos pares (KATERATTANAKUL; HAN; HONG, 2003; MILLER; RAINER, 2005; MYLONOPOULOS; THEOHARAKIS, 2001).

Na literatura, encontramos diferentes métodos que permitem medir a produtividade científica de pesquisadores, revistas, conferências, universidades e instituições. Entre eles, encontramos métodos quantitativos e qualitativos, ou mesmo uma combinação de ambos. Os métodos podem ser subdivididos em bibliometria¹, cienciometria², informetria³ e, atualmente, também em webometria⁴ e cibermetria⁵ (VANTI, 2007).

¹Bibliometria é “o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada” Tague-Sutcliffe (1992, p.1) apud MACIAS-CHAPULA (1998, P.134).

²Cienciometria é “o estudo dos aspectos quantitativos da ciência enquanto uma disciplina ou atividade econômica” Tague-Sutcliffe (1992, p.1) apud MACIAS-CHAPULA (1998, P.134).

³Informetria é “o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social, e não apenas aos cientistas” Tague-Sutcliffe (1992, p.1) apud MACIAS-CHAPULA (1998, P.134).

⁴Webometria é “o estudo dos aspectos quantitativos da construção e da utilização dos recursos de informação, estruturas e tecnologias da *web*, utilizando enfoques bibliométricos e informétricos” Björneborn (2004, p.12).

⁵Cibermetria é “o estudo dos aspectos quantitativos da construção e utilização de recursos de informação, estruturas e tecnologias em toda a Internet, com base em abordagens bibliométricas e informétricas” Björneborn (2004, p.13).

Cada um destes métodos se destina a medir, sob enfoques distintos, aspectos específicos de uma área de conhecimento. Para a realização deste estudo, foi dada ênfase à webometria.

Para realizar a avaliação dos eventos, os indicadores bibliométricos já foram motivo de muita pesquisa e continuam sendo ainda uma das métricas de classificação muito utilizada por pesquisadores e instituições de pesquisa. A necessidade de determinar a qualidade de conferências científicas e revistas tem motivado vários trabalhos na literatura (CAPES, 2008; EGGHE, 2006; HIRSCH, 2005; THOMPSON, 2009).

Com esta finalidade, recentemente, surgiram também os indicadores de visibilidade na *web* que estão ganhando importância na pesquisa para avaliação da qualidade de instituições de ensino e até mesmo de conferências e grupos de pesquisa. Entre todos, o maior destaque atual é da webometria, que vem ganhando cada vez mais interesse dos pesquisadores para avaliar o conhecimento e a produção científica das universidades do mundo todo como, por exemplo, o *ranking* das universidades desenvolvido por um grupo de pesquisa do laboratório Consejo Superior de Investigaciones Científicas (CSIC) da Espanha (RWWU, 2009).

A seguir, apresentamos alguns trabalhos realizados com os indicadores bibliométricos e os indicadores de visibilidade na *web*.

2.1 Indicadores Bibliométricos

Uma das métricas criadas para avaliar eventos científicos foi a dos indicadores bibliométricos, que, segundo Tague-Sutcliffe (1992, p.1) apud MACIAS-CHAPULA (1998, P.134), são:

O estudo dos aspectos quantitativos da produção, da disseminação e o uso da informação registrada. A bibliometria desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão.

Portanto, a bibliometria serve para representar os estudos que tentam quantificar os processos de comunicação científica e produtividade, fornecendo subsídios na formulação da política científica e tecnológica de determinada área de conhecimento. A bibliometria examina o número de citações que foram feitas em um determinado trabalho e os relacionamentos entre eles.

Os indicadores bibliométricos também podem ser utilizados para avaliar a qualidade e a produtividade das pesquisas científicas, por meio da medição do número de publicações e citações. Identificar o índice de qualidade de uma determinada conferência ou *journals* é de fundamental importância, pois pode auxiliar os pesquisadores a determinar onde publicar seus trabalhos.

Diversos estudiosos tem-se utilizado de indicadores bibliométricos para elaborar um *ranking* de qualidade de *journals* e conferências científicas. Os indicadores mais populares são o Fator de Impacto (FI), o *h-index*, o *g-index* e o programa QUALIS da CAPES. O FI foi elaborado por Eugene Garfield e desenvolvido pelo *Institute for Scientific Information* - ISI e avalia a qualidade de uma revista científica, ou de um evento científico, ou mesmo de um pesquisador ou Instituição Científica (AMIN; MABE, 2000). O FI de um periódico é calculado com base em um período de dois anos, a partir do número de vezes que os artigos publicados nesse período foram citados como referências no ano em questão, dividido pelo número de artigos científicos publicados nesses dois anos.

O *h-index* foi proposto por Hirsch (2005), que o define como sendo o número de “h” trabalhos com número igual ou superior a “h” citações. Assim, dizer que um pesquisador tem *h-index* = 5 significa dizer que este pesquisador publicou 5 artigos que receberam 5 ou mais citações. O autor argumenta que o *h-index* é preferível a outros critérios utilizados, tais como número total de artigos, número total de citações e citações por artigo.

A vantagem do *h-index* é combinar uma avaliação tanto da quantidade (número de artigos) como da qualidade (impacto ou citações destes artigos) (SCHUBERT; BRAUN; GLÄNZEL, 2006). Um acadêmico não pode ter um *h-index* elevado sem publicar um número substancial de artigos. Estes artigos precisam ser citados por outros pesquisadores para serem incluídos no *h-index*. Finalmente, talvez a mais forte indicação de que o *h-index* está se tornando uma medida geral para realizações, é a Thomson ISI atualmente o ter incluído como uma nova característica no “relatório de citação” da *Web of Science*.

O *h-index* ignora o número de citações para cada um dos artigos além do que é necessário para alcançar um determinado *h-index*. No entanto, um pesquisador ou *journal* com um *h-index* de 5 teoricamente poderia ter um total de 25 citações (5 para cada artigo), mas poderia também ter mais de 3.000 citações (4 artigos com 750 citações cada um e 1 artigo com 5 citações). Assim, a fim de dar mais peso aos artigos mais citados surge o *g-index*. O *g-index* foi proposto por Egghe (2006), que assim o definiu: dado um conjunto de artigos classificados em ordem decrescente do número de citações que eles receberam, o *g-index* é o maior número que os artigos superiores *g* receberam (em conjunto) pelo menor g^2 citações. Embora o *g-index* ainda não tenha atraído muita atenção, parece ser um complemento muito útil para o *h-index* (HARZING; WAL, 2008)

Já o QUALIS, foi desenvolvido pela CAPES para a classificação de veículos de divulgação da produção científica, tecnológica, artística, etc., dos programas de pós-graduação, tendo sido implantado em 1998 e utilizado para a composição de indicadores fundamentais para a avaliação do referido nível de ensino (CAPES, 2008). O documento da área de Ciência da Computação (CC) apresenta os critérios definidos pelo Comitê de Computação para a elaboração do QUALIS da área. Assim, foram considerados, como veículos, *journals*, periódicos, conferências nacionais, conferências internacionais, etc.. O QUALIS abrange o conjunto dos veículos de publicação relacionados pela área, classificando-os em conceito A, B ou C, ou então como "Não Classificado" - no caso de ser publicação pertinente à área, mas que não preencha os requisitos de classificação - ou "Inadequado" (D). A cada ano/período, o conjunto dos veículos contemplados será ajustado à medida que outros forem mencionados nos relatórios encaminhados à CAPES, ou os parâmetros de qualidade dos eventos sofram alterações. Ainda, os critérios para a classificação dos veículos, ou a própria classificação de um veículo em particular, poderão ser revistos para contemplar o dinamismo, a evolução e particularidades da área de CC, sempre segundo as diretrizes e normas da CAPES.

Neste trabalho optou-se por seguir esta classificação da QUALIS da CAPES, embora tenha sido alterada recentemente.

Com base no documento QUALIS da CAPES, Souto et al. (2007) propõe um modelo (semi) automático para classificação de Conferência Científica na área de CC, baseada em ontologias e inferência de regras. A ontologia desenvolvida recebeu o nome de OntoQualis. A classificação das conferências foi realizada em duas etapas: na primeira, foram realizadas as coletas de dados e a instanciação destes, e na segunda, a classificação das conferências. Os resultados desta avaliação demonstraram um nível satisfatório em comparação com a classificação QUALIS.

Hakan e Hussain (2008) demonstraram uma nova abordagem baseada em inferência de regras, criando um modelo *fuzzy*, e considerou diferentes fatores para acessar a qualidade global. Os fatores principais que afetam o *ranking* das conferências incluem a qualidade dos membros do comitê de programa, a razão da aceitação dos artigos, o impacto destes artigos, localizações convenientes e muitos mais. O modelo, na verdade, aprimora o *ranking*, adicionando mais fatores que afetam esta qualidade. Assim, melhoram o sistema de recuperação de informação baseada no raciocínio humano.

Um índice não dimensional chamado *Fuzzy Index* tem sido proposto e utilizado para embaralhar os corpos de pesquisa previamente classificados.

Neste modelo, foram adicionados diferentes parâmetros, que devem ser considerados na classificação de diferentes corpos de pesquisa. Em hipótese, os autores concluem que o modelo incorporou com sucesso o efeito dos parâmetros e, quanto mais dava importância (peso) a qualquer desses parâmetros, mais obtinha o efeito nos resultados finais. Não obstante, o trabalho é limitado a um determinado domínio e a alguns parâmetros. O modelo é bastante flexível e pode-se acrescentar o efeito de tantos parâmetros quantos quisermos. A classificação final pode mudar, adicionando ou excluindo parâmetros de input. O estudo demonstrou com sucesso a aplicação do *Fuzzy Index* em *ranking journals*, conferências e autores no campo da computação gráfica, em especial, e no da CC em geral.

A análise das citações é uma parte integrante da bibliometria, utilizada para avaliar artigos científicos e, desta forma, contribuir para uma melhor compreensão da atividade de pesquisa dos autores e das instituições (HIRSCH, 2005). A avaliação da produtividade dos cientistas tem-se baseado no número de trabalhos por eles publicados. Esta medida tornou-se popular para a avaliação do desempenho científico das instituições de pesquisa, pesquisadores, conferências e journals.

Baseando-se em análise de citações para determinar a qualidade de *journals* Katerattanakul et al. (2003) propuseram, no seu estudo, uma combinação de sete indicadores. Estes indicadores foram combinados a partir dos dados coletados: citações por artigo, taxa de não citados, 20 + taxa de citações, média anual de citações por artigo, relação citado-citando, autocitações e impacto atual do artigo. Com base na pontuação média de cada journal nos sete índices de citações, o autor conclui que os cinco principais *journals* de computação assemelham-se aos resultados relatados em vários estudos anteriores (MILLER; RAINER, 2005; MYLONOPOULOS; THEOHARAKIS, 2001).

O Lister et al. (2008) compara o *ranking* das conferências e revistas da área de CC do CORE¹ (*Computing Research and Education*) com a frequência de citação da ACE

¹CORE <http://core.edu.au/>

(*Australasian Computing Education Conference*) em 2005, 2006 e 2007. O autor considerou que as taxas de citação são uma medida de opinião favorável (“estima”) e, por isso, deveria haver uma relação positiva entre taxas de citação e *rankings*. Os *rankings* da conferência do CORE parecem refletir nas citações da ACE, mas apresentam ainda algumas inconsistências em seus resultados, já que o *site* CORE não fornece detalhes sobre o processo de classificação das conferências: somente descreve a estrutura ampla e menciona que acadêmicos fizeram uma avaliação preliminar através de um painel, que foi liberado para consulta e *feedback*. Com base neste painel, foi determinada a classificação final das conferências, o que torna um pouco obscura a forma utilizada para a avaliação das conferências por não especificar detalhadamente os critérios utilizados para a realização desse *ranking* no *site* da CORE.

O Harzing e Wal (2008) propuseram, em seu estudo, o uso de uma fonte alternativa de dados: o Google acadêmico (GA). Devido à sua mais ampla gama de fonte de dados, esse sistema disponibiliza uma cobertura mais abrangente que beneficia as publicações acadêmicas em fontes que não são (bem) abrangidas pelo ISI, o que permite uma democratização da análise de citações, uma vez que fornece dados a cada acesso às citações de todos os pesquisadores, independentemente dos meios financeiros da sua instituição. Com base nos dados do GA, foram utilizadas três alternativas (*h-index*, *g-index* e o número de citações por artigo) em contraposição à tradicional fonte de dados bibliométricos ISI, que fornece o *Journal Impact Factors* (JIF) anual para avaliar o impacto do *journal*.

Analisando os resultados, estas métricas apresentaram fortes correlações com o JIF tradicional. Elas conduzem os pesquisadores e instituições a considerarem as medidas de impacto baseadas no GA um índice mais abrangente e possivelmente mais exato.

Um método alternativo é o esquema automático de classificação de inovação baseado em termos que podem avaliar as publicações pelos tópicos que utilizam. Proposto por Liu et al. (2009), este esquema de *ranking* pode ser usado para avaliar a qualidade da inovação das conferências, *journals* e estudiosos com base em suas contribuições inovadoras. Uma ideia inovadora é apresentada principalmente pela introdução de um novo termo domínio. A qualidade de uma ideia é aferida observando-se a variação do tempo da popularidade de seu termo domínio associado. Uma maneira possível para julgar a qualidade de inovação das conferências e dos estudiosos automaticamente ocorre através da avaliação dos seus produtos de pesquisa. Para o desenvolvimento da pesquisa, o citado autor utilizou os critérios clássicos baseados em citações, nas quais uma publicação citada por muitos pode obter uma nota mais elevada por influenciar os demais. O autor, porém, desenvolveu um novo fator nos critérios de classificação, que chamou de Fator de Inovação (ou INF). Quando um novo termo proposto por uma publicação se torna popular mais tarde, então esta publicação recebe uma graduação INF elevada, pois constituiu não somente uma ideia nova, mas também uma nova linha de pesquisa do domínio.

A contagem INF das conferências e estudiosos na área de banco de dados, que são classificados a partir do algoritmo *Irak* (*Term-based Innovation Ranking*) proposto por Liu et al. (2009), é realizada utilizando-se a extração de termos de nomes de sessão em procedimentos. Esse é também um método híbrido, pois são aplicadas regras de associações, combinando análises linguísticas e métodos estatísticos. O *ranking* obtido foi comparado com o resultado do prêmio de inovação “SIGMOD” em um determinado período e os estudiosos que receberam o prêmio também ocupavam uma excelente

posição no *ranking*. Com isso, o método de *ranking* baseado em termos pode ser uma alternativa ao esquema de *ranking* clássico baseado em citação.

Embora tenham sido feitos avanços no modo de avaliar as conferências, com o desenvolvimento de novas metodologias, a maior dificuldade ainda reside no esforço necessário para coletar e organizar os dados, o que, muitas vezes, é feito de forma manual. O avanço das tecnologias *web* tem facilitado este trabalho, permitindo automatizar alguns processos de coleta, recuperação e análise das informações que possibilitam inferir a qualidade.

Depois de descrever alguns conceitos de bibliometria e trabalhos realizados utilizando análise bibliométrica como um indicador de qualidade para avaliações científicas de pesquisadores, conferências, *journals* ou até mesmo de universidades em nível mundial, relataremos algumas deficiências desta metodologia abordadas por diversos pesquisadores.

2.1.1 Deficiências dos indicadores bibliométricos

Depois de uma visão geral da bibliometria e da utilização de seus indicadores, vamos abordar alguns problemas técnicos e metodológicos na utilização de dados de citação e publicação no contexto avaliativo.

O *ranking* das universidades tem atraído muito a atenção dos decisores políticos, do mundo científico e dos meios de comunicação públicos, particularmente no *ranking* publicado desde 2003 pela Universidade de Jiao Tong, de Shanghai (SJTU, 2009). Aqui, como em outros casos, os elementos bibliométricos, tais como número de publicações e citações, desempenham um papel fundamental.

Algumas vezes, os cientistas executam seu papel como membros invisíveis dentro da universidade ao referir em seus próprios trabalhos anteriores os estudos de outros cientistas. Os motivos para citar ou não uma referência a um artigo específico podem variar consideravelmente (BROOKS, 1986; MACROBERTS; MACROBERTS, 1988; VINKLER, 1998 apud VAN RAAN, 2005). Então, sem dúvida, o processo de citação é um problema complexo, e certamente não fornece um “ideal” para monitorar o desempenho científico (MACROBERTS; MACROBERTS, 1996). Há, no entanto, evidências suficientes de que essas referências não são motivos tão diferentes ou determinados aleatoriamente de tal forma que o fenômeno da citação perderia o seu papel como uma medida confiável do impacto (VAN RAAN, 1998). A aplicação da análise de citações para toda produção de um grupo de pesquisadores durante um longo período pode se tornar um forte indicador de desempenho científico.

Mas, segundo Van Raan (2005), o núcleo do problema não está necessariamente na produção dos dados e sim, muitas vezes, nas pessoas responsáveis pela avaliação do desempenho da pesquisa científica. Por exemplo, os cientistas, quando no papel de chefes das instituições ou os administradores da ciência, quando no âmbito do governo, além de outros responsáveis políticos que demonstram uma atitude incentivadora das análises bibliométricas “*quick and dirty*”, considerando que a melhor qualidade está disponível.

O *ranking* das instituições de pesquisa utilizando os métodos bibliométricos é um instrumento inadequado para a avaliação de desempenho da pesquisa, mesmo no nível das grandes instituições (VAN RAAN, 2005). O problema, entretanto, não é a classificação como tal, e sim os indicadores utilizados para a classificação, que são, muitas vezes, insuficientes por não evoluírem. Essa situação faz parte do problema mais

amplo da insuficiência na aplicação desenvolvida com os indicadores bibliométricos utilizados por pessoas que não têm competências claras e experiência na área dos estudos quantitativos da ciência.

Os indicadores bibliométricos têm certo potencial, mas não evoluíram de maneira a incorporar os avanços tecnológicos. Além disso, só são plenamente confiáveis sob duas condições: o sistema técnico e a metodologia em que esses indicadores são baseados. Estes devem ser suficientemente avançados e sofisticados, de maneira que as armadilhas e as fontes de erro em análises de citação sejam conhecidas a ponto de todas as correções relevantes poderem ser feitas.

Antes de abordar alguns problemas relacionados com a utilização de *ranking* na avaliação da qualidade da pesquisa científica, as condições necessárias para uma aplicação avançada bem-sucedida de indicadores bibliométricos como ferramenta de apoio para *peer-review* são discutidas.

Com a análise de citação como elemento central, os indicadores “bibliométricos” medem o impacto da pesquisa publicada em uma perspectiva internacional. Por que devemos aplicar a análise bibliométrica na realização da pesquisa? *Peer-review*, sem dúvida, é e deve continuar a ser um dos processos importante de julgamento de qualidade. Mas *peer-review* é relacionado com decisões que podem ter graves deficiências e desvantagens (MOXHAM; ANDERSON, 1992; HORROBIN, 1990).

A dependência dos resultados na escolha dos membros da comissão é um dos principais problemas. Essa dependência pode resultar em conflito de interesses, desconhecimento da qualidade, ou um viés negativo contra as pessoas mais jovens ou recém-chegadas ao campo.

Embora *peer-review* seja amplamente utilizado para avaliar a qualidade dos resultados da pesquisa, o processo de revisão é caro, demorado e trabalhoso. Shanks et al. (2007), argumenta que medidas como índices de citação não são adequados para sistemas de informação.

O processo técnico mais central sobre o qual a análise de citação baseia-se inteiramente é a adequação entre as citando-publicações e as publicações-citadas. O significado deste processo é o seguinte: quando em uma publicação (a qual é “citando-publicação”) a referência é dada para outra publicação (a qual é “citada-publicação”), a referência tem de ser identificada nos índices de citação. Este é de fato o mais importante procedimento técnico que deve ser executado pelo produtor de tais índices.

Porém, ainda ocorre uma quantidade considerável de erros no processo de correspondência líder na citando-citando e, conseqüentemente, uma “perda” de citações de publicações específicas. Em média, o número de referências não correspondidas é de aproximadamente 7% das citações correspondentes (MOED, 2002). Frequentemente ocorrem problemas de não-correspondência relatada a respeito das publicações escritas por grandes grupos de autores, devido a variações e erros em nomes dos autores, em particular os de países que não falam inglês. Também ocorrem erros em *journal* nos números de volumes e nos números de página inicial, devido a discrepâncias de *journals* com duplo sistema de numeração-volume ou de volumes combinados, ou devido à aplicação de sistemas de numeração para diferentes artigos.

Assim, estas citações não correspondidas são altamente distribuídas de forma desigual em situações específicas, o que pode causar um aumento da porcentagem de citações perdidas de até 30%. Portanto, se os índices de citações são utilizados para fins de avaliação, todos esses possíveis erros devem ser corrigidos, tanto quanto possível.

Outro problema diz respeito à atribuição de publicações às organizações específicas, como institutos, departamentos universitários, e até mesmo, em um nível de agregação elevado, à organização principal, por exemplo, as universidades. Não há como simplesmente digitalizar os índices de citações, a fim de relacionar todas as publicações de uma universidade, por exemplo, pois variações no nome da mesma universidade e citação de grupos ou instituições de uma organização mencionada em vez da universidade em que a pesquisa de fato ocorre, inviabilizam a tarefa. Essa questão ocorre também com as conferências científicas. Foi possível registrar que a ausência de padronização do nome das conferências apresenta-se como uma dificuldade para os estudiosos que trabalham com análise de citações. Mais uma vez, grandes esforços na limpeza e reconstrução dos índices de citação são necessários para resolver este problema.

Porém a análise de visibilidade na web por inlinks externos poderia evitar este problema, por obter os dados de várias fontes. É pouco provável que haja uma unificação nos nomes: por exemplo, uma conferência X não é sempre descrita como 'X', podendo aparecer também como 'Xy'. Embora isso, os links não mudam sua descrição: a conferência *Entity-Relationship* (ER), no ano de 2009, tinha a URL 'www.inf.ufrgs.br/ER2009' e isto não mudará.

Primeiramente, todos os problemas técnicos têm de ser resolvidos antes de a discussão da metodologia aplicada fazer algum sentido. Então, depois de terem sido feitas todas as correções, a limpeza e a unificação, os problemas metodológicos relativos ao design e ao cálculo dos indicadores serão verificados como pertinentes ou não e isso deve ser resolvido também.

Uma questão a ser considerada é o idioma. Trabalhos recentes (GRUPP; HINZE; SCHMOCH, 2001; VAN LEEUWEN; VAN RAAN, 2001) mostram que o maior cuidado deve ser tomado na interpretação dos dados bibliométricos em uma avaliação comparativa dos sistemas nacionais de pesquisa. O valor medido dos indicadores de impacto das atividades de pesquisa, no nível de uma instituição e até mesmo de um país, depende fortemente de saber se inclui ou exclui publicações em *journals* escritos em outros idiomas além do inglês. Isto ocorre devido ao simples fato de, especialmente, *journals* em francês, em japonês e em alemão não abrangerem o idioma inglês, no qual os artigos têm um impacto consideravelmente menor. Portanto, se os pesquisadores usam esses *journals* de língua diferente do inglês, estas publicações são certamente incluídas formalmente em índices de citação. Mas geralmente o impacto de tais publicações é relativamente baixo. Assim, no cálculo do indicador de impacto elas são computadas, mas contribuem muito pouco.

Portanto, a falta de publicações em inglês "atenua" consideravelmente a medida de impacto de uma universidade. O uso do idioma alemão nos *journals* cobertos pelo índice de citação pode reduzir cerca de 25% na medida de impacto (VAN LEEUWEN; VAN RAAN, 2001). O mesmo efeito observa-se nos *journals* de língua francesa e de língua japonesa. Estes resultados ilustram mais uma vez os indicadores que devem ser interpretados no contexto de suas limitações inerentes, como, neste caso, os efeitos da língua, mesmo em nível macro, de países inteiros.

Há ainda o problema do viés norte-americano em dados de citação. Na maioria das vezes, os *rankings* baseados em indicadores bibliométricos por meio de índices de citação são dominados pelas universidades e instituições de pesquisa norte-americanas. Embora, indiscutivelmente, estas instituições sejam de classe mundial, elas têm o efeito de dominante no tráfego de citações.

As análises bibliométricas aplicadas não são sofisticadas o suficiente. Assim, o verdadeiro problema não é o uso de indicadores bibliométricos, como tais, mas sim a aplicação de medidas bibliométricas menos desenvolvidas.

Como *peer-review* é uma das formas de avaliação da pesquisa, os indicadores bibliométricos devem agir como uma ferramenta de apoio para fazer *peer-review* mais objetivos e transparentes (RINIA et al., 1998). Profissionais da bibliometria podem atuar aqui como consultores, a fim de agregar valor ao processo de *peer-review* e evitar o uso fraudulento que causa danos às universidades, institutos e estudiosos. Um exemplo recente das melhores práticas é a aplicação de indicadores bibliométricos combinada com *peer-review* na avaliação de todos os grupos de pesquisa universitários em química e engenharia química na Holanda (VSNU, 2002).

As transformações que vêm ocorrendo em todas as áreas do saber são influenciadas pela presença forte das novas tecnologias de informação. A expansão da ciência trouxe a necessidade de avaliação e acompanhamento do desenvolvimento e dos avanços alcançados pelas diversas áreas do conhecimento. Existem inúmeras formas de medir voltadas para avaliar a ciência e os fluxos de informação, dentre as quais podemos citar: a cienciometria, a bibliometria e a webometria. Cada uma dessas técnicas quantitativas busca enfoques diferentes para medir a difusão do conhecimento científico.

Uma metodologia que vem ganhando muita importância em nível mundial é a webometria. Esta metodologia está sendo usada no *ranking* das universidades através da visibilidade na web, do GA, da quantidade de 'arquivos ricos' e do tamanho das páginas, resultando em novos indicadores de qualidade. Este processo poderá ser utilizado para determinar a qualidade das conferências científicas da área de CC e fornecer um *ranking* de impacto mais avançado, pois um dos problemas relatados anteriormente aborda a utilização de indicadores bibliométricos desatualizados, entre outros fatores e algumas deficiências em utilizar *peer-review*.

De modo geral, a ideia não é a substituição de análise bibliométrica, ou *peer-review* pela webometria, pois os aspectos subjetivos não são meramente negativos. Em qualquer avaliação, deve haver espaço para as ideias intuitivas de cientistas. No entanto, para uma melhoria substancial da tomada de decisões sobre questões de atividades científicas, os indicadores webométricos podem ser usados em paralelo com outros métodos de avaliação. A utilização de vários métodos em conjunto poderia resultar num indicador de qualidade mais confiável.

Tem-se a sabedoria convencional de que a informação que os estudiosos relacionados disponibilizam na web tem sido dramaticamente incrementada (BOKOS; POULOS; KORFIATIS, 2007). Ao mesmo tempo, a web torna-se o principal meio de comunicar e distribuir informação. Os resultados das pesquisas levam-nos à compreensão, melhoria e exploração adequada da web para a pesquisa científica. Métodos, técnicas e ferramentas originadas no método bibliométrico para aplicação nas avaliações dos estudos científicos têm afetado a maioria das avaliações (BAR-ILAN; PERITZ, 2002).

Com o constante aumento no número de conferências, as métricas existentes utilizando bibliometria estão se tornando inadequadas para medir a qualidade das conferências de CC. Pode-se dizer que até inaplicáveis para as conferências emergentes, já que as estatísticas de citações históricas não estão prontamente disponíveis. Estas estatísticas de citação levam muito tempo para acumular. Por exemplo, o banco de dados de grandes conferências e *journals* que, entre 1994 e 2003, tiveram mais citações remontam a cinco ou mais anos (THOR; RHAM, 2005). Assim, exigem o conhecimento

disciplinar: a menos que seja um acontecimento conhecido no domínio familiar de um investigador, este pode não estar ciente das suas estatísticas de citação.

As técnicas existentes para medir a reputação e qualidade das conferências são determinadas pela bibliometria, principalmente as métricas baseadas em citações como o FI. Ou seja, a qualidade de um elemento é diretamente proporcional às suas características bibliométricas. Por exemplo, se o número de citações das publicações de uma conferência passou de certo limiar, a conferência é considerada de boa qualidade. No entanto, Zhuang et al. (2007) afirmou que essas técnicas não são suficientes para medir a qualidade estabelecida, pois as estatísticas de citação levam tempo para acumular. Conforme o das conferências em CC, salientando que, para conferências emergentes ou jovens, as estatísticas de citações históricas não estão facilmente disponíveis, tornando métricas baseadas em citação inaplicáveis. O mesmo vale para conferências bem-estabelecidas no estudo relatado acima, mesmo o banco de dados de grandes conferências e *journals*, entre 1994 e 2003, mostra que muitas das citações remontam há cinco ou mais anos.

Zhuang et al. (2007), em seu estudo, assumiu a hipótese de que a qualidade de uma conferência está intimamente correlacionada com os membros do seu PC. Esta hipótese, juntamente com o descrito acima, inspirou a realização do estudo sobre problema de uma nova perspectiva para avaliar a qualidade de conferências através da análise das características dos membros do seu PC.

Atualmente, há diversos métodos disponíveis para avaliação da qualidade de uma conferência. Pode-se utilizar um conjunto de indicadores, tais como h-index, inlinks, Pagerank, etc. para a realização de *rankings* das conferências e até mesmo das universidades e instituições. É evidente que, devido à vasta gama de informações disponibilizada na web em diferentes bases de dados, e com os sistemas e as ferramentas de busca cada vez mais aprimorados e atualizados, devemos usufruir dessas tecnologias disponíveis para gerar novos indicadores. Estes indicadores podem gerar novos índices de qualidade. Esta nova maneira de medir a qualidade a partir da web fornece informações ao usuário com maior precisão e fácil acesso, e com custo mais acessível do que, por exemplo, o FI da ISI. Também poderá evitar alguns problemas como: de interesse, que pode ocorrer por meio de *peer-review*, e da bibliometria, por meio da utilização de indicadores desatualizados.

2.2 Visibilidade na Web

Outra maneira de medir a qualidade de uma conferência ocorre por meio da visibilidade na *web*. Esta é uma das percepções externas da qualidade de um trabalho e pode ser medida através de uma análise dos *links* que apontam para o *site* da conferência. Para isso, foram desenvolvidas várias metodologias e ferramentas, entre as quais podemos citar como bastante popular o *PageRank*, desenvolvido pelos fundadores do Google, Larry Page e Sergey Brin na Universidade de Stanford. O *PageRank* é usado em conjunto com uma série de diferentes fatores, a fim de encontrar respostas mais relevantes para uma determinada busca. Infelizmente, nem a forma como esses fatores estão computados, nem como eles são combinados com o *PageRank* são de domínio público (BIANCHINI; GORI; SCARSELLI, 2005).

O *PageRank* funciona como um sistema que atribui notas para páginas na *web*. A classificação das páginas baseia-se na vasta estrutura de *links* como um indicador do valor de uma página individual. Essencialmente, o *PageRank* interpreta um *link* da página A para a página B como um voto da página A para a página B. Mas olha além do

volume de votos, ou *links*, que uma página recebe, analisa também a página que dá o voto. Os votos dados por páginas "importantes" pesam mais e ajudam a tornar outras páginas "importantes." *Sites* importantes, de alta qualidade recebem uma nota de avaliação maior, que é gravada a cada busca feita. Esta noção de importância da página é independente do seu conteúdo e lembra a noção de citações na literatura científica.

Um *PageRank* elevado significa que muitos usuários na *web* estão interessados na página e mais usuários vão provavelmente visitá-la se comparada com páginas que possuem um *PageRank* baixo. Considerando a efetividade dos resultados da pesquisa do Google e sua adoção por muitas ferramentas de busca *web*, o *PageRank* parece capturar bem a importância ou a qualidade das páginas *web*. De acordo com as pesquisas recentes, a maioria dos usuários está satisfeita com os resultados apresentados por esta ferramenta de busca (CHO; ADAMS; ROY, 2005).

Em outro estudo, os autores apresentaram um método que estabelecia uma média de classificação ao longo dos estudos para determinar a sua importância. Utilizou este método com 9 estudos publicados entre 1991 e 2003 para produzir uma classificação composta por 50 revistas mais importantes ao longo desses estudos. Os 2 primeiros estudos utilizam uma análise de citações para classificar as revistas e os 7 restantes empregam percepções dos que responderam ao questionário para classificar as revistas. Segundo os autores, o método de análise de citações é digno de interesse uma vez que pretende ser mais objetivo do que o uso das percepções dos consultados Miller e Rainer (2005). Neste estudo, os autores utilizam duas metodologias para avaliação das revistas, a bibliometria para realização da análise de citação e a análise de redes sociais através de questionários *on-line* para que pesquisadores e participantes respondam.

Esta segunda metodologia utiliza-se de outra forma para medir a visibilidade na *web*, por meio das chamadas 'redes sociais' que têm crescido significativamente nos últimos anos, em função do avanço de processamento computacional, da popularização da Internet e do aumento do leque de assuntos e áreas que a utilizam. Uma rede social é basicamente uma estrutura constituída de 'nós', representando entidades de diferentes grupos, ligados a diferentes tipos de relações, com fins que vão desde o estabelecimento de amizades até a montagem de estruturas compartilhadas de conhecimentos no âmbito institucional. Enxergar e entender as relações sociais entre os indivíduos ou outras entidades é conhecido como Análise de Redes Sociais (WASSERMAN; FAUST, 1994).

Outro estudo utilizando questionários *on-line* foi realizado por Mylonopoulos e Theoharakis (2001), que pedia uma classificação de até 10 revistas que os participantes considerassem como as melhores em termos de contribuição para a área de SI. Esta pesquisa demonstrou que usuários de diferentes países determinam diferentes prioridades para o gerenciamento SI e que a perceptível qualidade das revistas é um conceito complexo, pois a avaliação pode variar consideravelmente, dependendo das medidas adotadas e dos diversos enfoques adotados pelos pesquisadores em diferentes partes do mundo. Com esta pesquisa, os autores mostraram uma perspectiva global da interpretação de cada participante em relação à qualidade e ao impacto das revistas.

Ainda no quesito visibilidade na *web*, Zhuang et al. (2007) propõe que seja utilizado um conjunto de novas heurísticas para descobrir automaticamente conferências com prestígio (e também trabalhos de baixa qualidade) por meio da mineração das características dos membros do Comitê de Programa (CP). Combinadas com um sistema de classificação, conseguem uma precisão maior. Os resultados podem ser aplicados diretamente em bibliotecas digitais já existentes para complementar e reforçar a bibliometria. Para identificar a reputação e a qualidade de uma conferência, os citados

autores desenvolveram cinco heurísticas: número de membros do CP da conferência (hipoteticamente, quanto mais membros o comitê tiver mais rigoroso será o processo de avaliação dos artigos), número médio de publicação por membro do comitê (coleta a média de publicação na ACM), número médio de co-autores (assume que pesquisadores renomados tendem a publicar mais artigos em co-autoria com seus colegas), grau de proximidade do CP (utiliza Análise de Redes Sociais para quantificar a localização do indivíduo em uma comunidade. O pressuposto é que uma conferência de alta qualidade tem um programa composto por um grupo de pesquisadores renomados, que estão localizados em destaque na rede social daquela comunidade), grau de intermediação do CP (mostra a influência do indivíduo no fluxo de informações na rede social). O resultado da avaliação com aplicação destas heurísticas mostra que foram eficazes na detecção de conferências de baixa qualidade e na recomendação e classificação, inclusive de novas conferências com boa qualidade.

Assim, existem diversos estudos utilizando redes sociais para gerar novos indicadores que podem ser uma nova alternativa no *ranking* de *journal*. Bollen et al. (2005), em seu estudo de *journals*, analisou as redes sociais a partir da citação e do *download* de dados, determinando o impacto dos *rankings* de *journal* dessas redes por meio de um conjunto de métricas da centralidade das redes sociais. O impacto resultante dos *rankings* dos *journals* foi comparado com o FI da ISI. Os resultados indicaram que, embora as métricas de rede social e o *ranking* da ISI se desviem moderadamente para redes *journals* baseadas em citação, eles diferem consideravelmente nas redes de *journal* derivadas dos dados de *download*. Portanto, este estudo conseguiu capturar os resultados de impacto dos *journals* que não são capturados pelo ISI. Além disso, estes resultados levantam questões com a relação da validade do ISI como única avaliação de impacto de *journal*, e sugerem a possibilidade de elaboração de indicadores de impacto com base na utilização das informações em geral.

Já Martins et al. (2009), inspirado em Zhuang et al. (2007), propõem duas abordagens diferentes para a classificação automática de Conferências Científicas de CC. Na primeira, utilizam um conjunto de novas métricas, baseadas em citações bibliográficas. Na segunda, utiliza alguns dos critérios que um especialista considera para dar a sua opinião sobre a qualidade das conferências, entre os quais podemos citar tradição, taxa de aceitação, reputação dos membros do comitê de programa, entre outros. E a partir delas tenta automatizar a avaliação, por meio de técnicas de aprendizagem de máquina, classificando as conferências como A, B ou C (alta, média ou baixa). Com isso, concluem afirmando que os mais eficazes são aqueles com base em citações seguidas por informações sobre a tradição do evento.

2.2.1 Webometria

O termo webometria foi criado por Almind e Ingwersen em 1997. É uma das áreas de estudo que vem assumindo crescente importância nas análises quantitativas na Internet, despertando o interesse dos pesquisadores no assunto. Thelwall et al. (2003, p. 81) assim definem a webometria:

A webometria trata dos aspectos quantitativos tanto da construção quanto do uso da *web*, compreendendo quatro áreas principais de pesquisa: análise de conteúdo das páginas *web*; análise da estrutura dos *weblinks*; análise do uso da *web* (exploração dos programas que registram os comportamentos de pesquisa e busca na *web*) e análise de tecnologias na *web* (incluindo o desempenho das ferramentas de busca).

Björneborn (2004, p.12) também definiu a webometria como o “estudo dos aspectos quantitativos da construção e da utilização dos recursos de informação, estruturas e

tecnologias da *web*, utilizando enfoques bibliométricos e informétricos”. Eles consideram webometria como uma parte da bibliometria.

Portanto, esta definição abrange os aspectos quantitativos da construção e da utilização da *web*, abrangendo as quatro principais áreas de pesquisa da webometria: análise do conteúdo da página *web*; análise de estrutura de *links* da *web*; análise de uso da *web* e análise da tecnologia *web* como, por exemplo, as ferramentas de busca. Todas estas áreas de pesquisa incluem estudos longitudinais de mudanças na dinâmica *web*, por exemplo, conteúdo das páginas, estruturas de ligação, padrões de usabilidade e novas tecnologias disponíveis.

Outra metodologia que está sendo utilizada atualmente é a cibermetria, muitas vezes usada como sinônimo da webometria. Estas duas metodologias são mais adotadas em ciência da informação, uma terminologia diferenciada para distinguir entre os estudos sobre a *web* e os estudos de todas as aplicações da Internet. A cibermetria é definida por Björneborn (2004, p.13) como: “o estudo dos aspectos quantitativos da construção e utilização de recursos de informação, estruturas e tecnologias em toda a Internet, com base em abordagens bibliométricas e informétricas”.

Portanto, cibermetria engloba estudos estatísticos dos grupos de discussão, listas de discussão e outras formas de comunicações na Internet, além de abranger toda a comunicação mediada por computador com aplicações da Internet. Também abrange medidas quantitativas da tecnologia de *backbone*, topologia e tráfego (MOLYNEUX; WILLIAMS, 1999 apud BJÖRNEBORN, 2004). Já a webometria se restringe ao estudo dos padrões de produção de informação, busca, recuperação, divulgação e utilização na WWW, que é o componente da Internet o qual utiliza textos, imagens e transferência de arquivos, para fornecer informações acessadas através dos bilhões de páginas *web* de todo o mundo.

Uma razão para usar a webometria é mostrar a ligação com a bibliometria e a informetria destacando uma perspectiva da ciência da informação. A webometria é uma forma de reconhecer a importância da rede como meio de informação e comunicação para a ciência, os estudiosos, as instituições e outros nos quais os estudos quantitativos estão sendo utilizados (VANTI, 2002).

2.2.1.1 Objetivos

A webometria tem como objetivos avaliar o acesso aos *sites* da *web*, avaliando o número de páginas por *website*, o número de *links* por página, o tamanho da página, o número de citações e o número de *links* recuperados por uma ferramenta de busca. A análise destes dados permite verificar a produção de pesquisadores e instituições, ajudando a aprimorar os resultados das ferramentas de busca. Por exemplo, o *link* é considerado um relevante indicador da importância de um determinado *site* na *web*, já que ressalta a sua posição no resultado retornado pelas ferramentas.

Outros objetivos são avaliar o acesso a determinados *websites*, detectar a presença de instituições, conferências, *journals* e pesquisadores na rede, melhorar a eficiência das ferramentas de busca na recuperação da informação (VANTI, 2007), assim podendo demonstrar para a comunidade acadêmica e política qual a importância da publicação na *web*, não só para a divulgação do conhecimento acadêmico, mas para a medição de atividades científicas, seu desempenho e impacto.

2.2.1.2 Trabalhos relacionados

Embora webometria ainda seja um campo novo, diversos artigos relatam o reconhecimento crescente da importância da metodologia e a existência de pesquisas. Smith (1999) foi pioneiro na aplicação da análise de citação de *journals* eletrônicos australianos, selecionando 22 periódicos aleatoriamente, por meio da ferramenta de busca AltaVista para computar os *links* de cada *website*. Nesse procedimento não encontrou correlação entre seus resultados e o FI calculado pelo ISI. Concluiu então que *links* de *journals* eletrônicos são fundamentalmente diferentes das citações. Ford e Harter (2000) analisaram um conjunto heterogêneo de 39 *journals* e também não encontraram correlação significativa entre as medidas de *links* e o FI da ISI. Uma correlação significativa entre os *inlinks* e o FI foi encontrada por Vaughan e Hysen (2002), que analisaram um conjunto de *journals* de algumas disciplinas específicas as quais são indexadas pelo ISI.

Os *journals* pesquisados eram tradicionais, com *websites* associados. O fato de haver correlação desta vez possivelmente reflete um avanço na *web*.

O impacto dos *journals on-line*, medido pela contagem total de *inlinks*, é, portanto, associado ao impacto *off-line*, medido pela contagem média de citações por artigo. Estas correlações são importantes por sugerir que os *links* podem ser criados por motivos semelhantes aos de citações e, assim, podem ter potencial de uso em cálculos semelhantes aos que utilizam indicadores bibliométricos.

Vaughan e Thelwall (2003) incorporaram dois fatores: o conteúdo do *website* de um *journal* e a sua idade. E confirmaram que a contagem de *inlinks* dos *websites* se correlaciona com o FI. Quanto mais antigo o *journal* era, mais *inlinks* atraíam; da mesma forma, os *journals* com mais conteúdo *on-line*. Foram utilizados todos os artigos publicados em 1997 em 46 *journals* da ciência da informação, que produziram predominantemente correlações significativas, sugerindo que os impactos de citação *on-line* e *off-line* são de alguma forma fenômenos semelhantes.

Uma das medidas utilizadas hoje para avaliação de universidades em nível mundial é o "*Webometrics Ranking of World Universities*", que foi oficialmente lançado em 2004 e é atualizado a cada 6 meses (janeiro e julho). A avaliação é feita em mais de 18 mil instituições de ensino superior por todo o mundo e no final apresenta uma lista do *ranking* das 8.000 (mil) melhores instituições, sendo disponibilizada no *site* oficial¹. A metodologia deste *ranking* considera as análises quantitativas de conteúdos disponibilizados na *web*, especialmente aqueles relacionados a processos de geração e comunicação do conhecimento científico acadêmico, avaliando as atividades científicas, o desempenho e o impacto (RWWU, 2009). A webometria ainda leva em consideração o volume de informação publicada, o impacto e a visibilidade de tais conteúdos na *web*. Os indicadores utilizados são baseados na *web* e correlacionados com os indicadores bibliométricos e cienciométricos.

A webometria também pode ser utilizada para classificar as melhores universidades e instituições em nível mundial. Os indicadores do *ranking* são fornecidos de forma a mostrar o compromisso das instituições com as publicações *web*. Se o desempenho de uma instituição está abaixo da sua posição esperada de acordo com a sua excelência

¹<http://www.webometrics.info>

acadêmica, as autoridades dessa instituição deverão reconsiderar a sua política web, promovendo aumentos na qualidade da produtividade científica e na quantidade das suas publicações eletrônicas.

A quantidade de pesquisa científica produzida pelos acadêmicos é a principal razão para atrair os *links*. As universidades com melhores pesquisadores ou com renome internacional tendem a atrair mais *links*. A hipótese é de que os pesquisadores que produzem mais conteúdo na *web* têm maior capacidade para atrair *links* (HARRIES; THELWALL, 2004).

A Universidade Federal do Rio Grande do Sul (UFRGS) foi avaliada entre as 300 melhores instituições de Ensino Superior do mundo. No último *ranking* semestral organizado pelo Ministério da Educação da Espanha, com 18 mil participantes, a universidade gaúcha conquistou a 243ª posição. No Brasil, a Universidade de São Paulo ocupa a melhor colocação (53ª). O índice leva em conta itens como a qualidade das pesquisas acadêmicas, o prestígio em âmbito internacional, o comprometimento dos professores, a relação da instituição com a comunidade e a qualidade do diálogo entre a universidade e setores da economia. O *ranking* é liderado em sua última edição pelas universidades americanas, que ocuparam as primeiras posições. A universidade de Harvard ocupou a primeira posição no *ranking*, seguida pelo Instituto de Tecnologia de Massachusetts e pela universidade de Stanford.

2.2.1.3 Funcionamento da metodologia

A webometria pode ser utilizada em toda a *web*, incluindo agrupamentos de *site* (*clusters*), domínios, ferramentas de busca, páginas *web*, URLs e outros *links*. Uma forma adequada de *ranking* é combinar um conjunto de indicadores que medem estes diferentes aspectos. Almind & Ingwersen propuseram o primeiro indicador *web* conhecido como *Web Impact Factor* (WIF), baseado na análise do *link*, e que combina o número de *links* externos (*inlinks*) com o número de páginas do *site*, numa relação de 1:1, entre visibilidade e tamanho. Esse índice é usado para o *ranking*, mas acrescentando dois novos indicadores para o componente tamanho: número de documentos, medido a partir do número de arquivos ricos em um domínio *web* e do número de publicações coletadas pelo GA. Os quatro indicadores são descritos a seguir:

- Tamanho (*Size*). Número de páginas recuperadas de quatro ferramentas de busca: Google, Yahoo, Live Search e Exalead. Para cada ferramenta, os resultados são normalizados para log-1 considerando o valor mais elevado. Então, para cada domínio, os resultados máximo e mínimo são excluídos, e para cada instituição é atribuído um *ranking* de acordo com a soma combinada.
- Visibilidade (*Visibility*). O número total de *links* externos únicos recebidos (*inlinks*) por um *site* só pode ser obtido de ferramentas confiáveis como: Yahoo Explorer, Live Search e Exalead. Para cada ferramenta, os resultados são normalizados para log-1 considerando o valor mais elevado, e então combinados para gerar o *ranking*.
- Arquivos ricos (*Rich Files*). Depois de avaliar a sua relevância para as atividades acadêmicas e as publicações, e considerando o volume dos diferentes formatos de arquivo, foram selecionados os seguintes: Adobe Acrobat (.Pdf), PostScript (.Ps), Microsoft Word (.Doc) e Microsoft PowerPoint (.ppt). Estes dados foram extraídos com a utilização do Google e a fusão dos resultados para cada tipo de arquivo após normalização de *log* é feita como descrito anteriormente.

- Acadêmico (*Scholar*). O GA fornece o número de artigos e citações para cada domínio acadêmico. Estes resultados representam artigos, relatórios e outros itens acadêmicos.

A melhor maneira de construir o *ranking* é combinando um conjunto de indicadores que medem esses diferentes aspectos. Os quatro *rankings* foram combinados de acordo com o esquema da figura 2.1, na qual cada indicador tem um peso diferente.

WEBOMETRICS RANK	
VISIBILITY (external inlinks) 50%	SIZE (web pages) 20%
	RICH FILES 15%
	SCHOLAR 15%

Figura 2.1: Esquema da webometria (RWWU, 2009).

Os dados dos quatro indicadores foram obtidos a partir dos resultados quantitativos fornecidos pelas principais ferramentas de pesquisa. Na realidade, os valores obtidos representam a visibilidade que o *site* em questão tem na Internet.

Os *websites* dos ministérios Iranianos são, de fato, uma entrada no ambiente virtual, com informações valiosas sobre os funcionários e membros da sociedade. Danesh et al. (2008) analisaram a visibilidade, o WIF e a taxa de colaboração dos *websites* dos ministérios do governo iraniano. Foram utilizados métodos webométricos no processo e análise de link. Os *inlinks*, *Self-links* e *co-links* dos *sites* em estudo foram totalizados e, em seguida, agrupados (*clusters*) e também foi aplicado um escalonamento múltiplo dimensional. Todos os 20 *sites* pertencentes aos ministérios do governo iraniano, no período de um mês, foram computados por meio da ferramenta de busca Yahoo. Esta pesquisa mostrou uma colaboração entre os *websites* dos ministérios e destacou também a importância da qualidade do conteúdo do site como um dos principais fatores para atrair *links*. Os *sites* que obtiveram maior número de *inlinks* são os que têm mais visibilidade e podem ser vistos como um indicador da importância global e da importância do *website* na rede. Outro ponto importante a ser abordado é a necessidade de realização de planos de melhoria da qualidade e do conteúdo dos *websites*.

Para verificar a visibilidade na *web* dos *websites* da biblioteca nacional - *National Library Websites*, Osareh et al. (2006) utilizou também os métodos webométricos. Desta forma, foram computados os *inlinks* e *co-links* do *website* da biblioteca, analisando a sua visibilidade, a sua importância na *web*, em nível nacional e internacional, e também a taxa de colaboração entre os *websites*. Portanto, usando métodos webométricos, pode ser vista a visibilidade que um *website* tem no mundo. Além disso, é importante mencionar a realização de estudos que tenham sido derivados de modelos de pesquisa de citações impressas (PLOTT; TUNGER, 2005). Devemos prestar atenção a algumas considerações. Enquanto em *journals* científicos, índices de citações são as ferramentas mais importantes para os estudos de citação, no ambiente *web*, as ferramentas de pesquisa desempenham parcialmente este papel (KOUSHA, 2002).

As ferramentas de busca são fáceis de usar e suficientemente poderosas para uma extração automática de dados na *web*. Por longo período, o uso de ferramentas de busca

tem sido motivo de diversas discussões por causa das suas limitações como: a instabilidade nos seus resultados (BAR-ILAN, 1998; ROUSSEAU, 1997), a cobertura incompleta da *web* (GILES; LAWRENCE, 1999) e a falta de confiabilidade nas bases de dados (MAT-HASSAN et al., 2004). Assim, estudos recentes têm mostrado que as ferramentas de busca atuais melhoraram a sua consistência e confiabilidade (BAR-ILAN, 2005). Nas análises webométricas, as ferramentas de busca são muito úteis, pois desempenham um papel fundamental no processo de obtenção dos dados, permitindo uma cobertura maior e mais abrangente nas bases de dados.

Um dos resultados da pesquisa acima mostrou que os *websites* forneciam várias questões as quais evidenciavam mais qualidade na *web* por meio da visibilidade nos *inlinks* ao comparar com as citações de trabalhos anteriores.

Para exibir e visualizar as universidades mais importantes do mundo, Aguillo e Ortega (2009) mostram as características topológicas e descrevem as relações entre as universidades na *web*. As primeiras 1.000 instituições de ensino superior foram selecionadas a partir do *Web Ranking of World Universities*. Esta ordem do *ranking* das universidades foi estipulada em função de quatro características principais da *web* e do seu domínio institucional, a saber: o volume de conteúdo, que é medido pelo número de páginas de acesso livre; a sua visibilidade na *web* por meio do número de *inlinks* recebidos; o número de arquivos ricos, que é utilizado como um indicador, pois estes são uma forma para divulgar os dados técnicos, científicos e os resultados; e o número total de documentos indexados no GA, que é um indicador das publicações científicas na Internet. Cada domínio *web* é classificado pela agregação linear destes indicadores. A construção do *ranking* webométrico (WR) se dá pelo Esquema utilizado na figura 2.1. E os relacionamentos de *link* dos *websites* foram computados pela ferramenta de busca Yahoo! Search. Os dados obtidos constituem-se em gráficos de rede e mapas geográficos. Utilizaram-se técnicas de análise de redes sociais usadas para analisar e descrever as propriedades estruturais de toda a rede e seus nós.

Os resultados evidenciam, segundo eles que as redes universitárias *world-class* constituem-se de sub-redes nacionais que se fundem em um núcleo central, às quais as principais universidades de cada país adicionam *links* internacionais. Os Estados Unidos dominam a rede mundial e, na Europa, as sub-redes inglesas e alemãs se destacam.

Para verificar a visibilidade das universidades iranianas de ciência médica na *web* Keshtkar et al. (2009) aplicaram o método webométrico por meio do número de páginas da *web*, *inlinks* e também o fator de impacto *web*. Utilizou a ferramenta de busca AltaVista para obter os dados. As universidades foram classificadas usando os indicadores webométricos. Os resultados revelaram que as universidades iranianas de ciência médica não tiveram muito impacto na *web* e não são muito conhecidas internacionalmente. No entanto, a principal razão baseia-se nas barreiras linguísticas: *sites* escritos em línguas nativas têm tendência de atrair menor público.

Segundo eles, algumas universidades iranianas também sofrem de problemas técnicos em seus *websites*. Para torná-los mais atrativos, os formuladores de políticas de tais universidades e também os gestores dos *websites* devem desenvolver planos para melhorar a qualidade e a visibilidade na *web*, tornando os *sites* ativos e mais ricos em conteúdo e informações.

Para determinar o *ranking* das universidades australianas, Boell et al. (2008) usaram o método webométrico para gerar o fator de impacto *web* que dependia do tamanho (*size*) e do pessoal.

As ferramentas de busca podem ser empregadas para coleta de dados de forma eficiente e automatizada por meio das URLs das universidades. Os citados autores compararam os WIFs relacionados ao pessoal com o fator de impacto relacionado ao tamanho. O resultado mostra que os *rankings* baseados em WIFs relacionados ao pessoal se correlacionam melhor com um *ranking* estabelecido pelo Instituto de Melbourne do que os WIFs dependentes do tamanho.

Portanto, o estudo baseado em métodos webométricos está investigando o aspecto do *ranking* das instituições acadêmicas. Há também a necessidade de compreender melhor o que os WIFs estão medindo, uma vez que eles parecem introduzir um novo aspecto em *rankings* universitários que diferem dos *rankings* clássicos, principalmente com base na qualidade da pesquisa. WIFs produzem resultados levemente diferentes, fornecendo informações melhores e mais amplas, que podem refletir mais nas universidades na medida em que elas são "linkadas" pelo público em geral. Este fator pode influenciar na hora de um estudante ou pesquisador escolher uma boa universidade. O estudo deles mostra que, se avaliados adequadamente, os indicadores baseados na web podem constituir uma parte de um critério múltiplo baseado no *ranking* das instituições de ensino superior.

Jowkar e Beglou (2008), em seu estudo, relacionaram o *ranking* web com a contagem dos links baseando-se em dois *rankings* conhecidos internacionalmente: o *Academic Ranking of World Universities*, apresentado pela universidade de *Shanghai*, e o *Times Higher Education Supplement*.

A pesquisa deles analisou os aspectos webométricos, verificando a relação entre *rankings* da universidade de *Shanghai* e a taxa de seus links, e encontrou correlação significativa entre os links e os dois *rankings*. A maior correlação foi entre os links externos. Desde que os estudos webométricos foram realizados em revistas eletrônicas e *sites* de universidades e desde que a documentação foi criada e usada em ambientes científicos, esses estudos podem ser usados como evidência para a compreensão das relações científicas (KOUSHA, 2003).

Boudourides et al. (1999), exploraram e documentaram o estudo de viabilidade da bibliometria na *web*. O objetivo foi testar a medida em que ele analisa cienciometria como, por exemplo, análise co-citação, no nível da *web*, que pode produzir resultados coerentes e interpretáveis. A análise webométrica é referida nos estudos quantitativos de comunicação na rede realizados por intermédio de *links web*. Uma exploração em webometria foi feita para a autoorganização da sociedade da informação europeia e outra para a constituição de instituições da *Triple Helix*, ou seja, as universidades, os governos, e as indústrias na Europa. Isto foi realizado por meio da análise do número de ligações recebidas por estes *websites* e da verificação da sua visibilidade na *web*. Para realizar análise de correspondência múltipla, foi utilizado o algoritmo de agrupamento por *clusters* que pode ser observado nas correlações significativas entre os dados.

Assim, diz Boudourides et al. (1999):

A ideia de considerar a *web*, paralelamente à bibliometria e à cienciometria como uma rede de citações, é bastante recente. Embora diferentes métodos de pesquisas de distribuição de dados sejam amplamente utilizados em um contexto de recuperação de informações, muito pouco é ainda conhecido no âmbito das pesquisas semânticas e do processamento de conteúdo na condição de uma tentativa realizada através da webometria.

Nas universidades nigerianas, a análise de *links* implicaria examinar o padrão das relações com cada um dos outros *websites*, bem como das universidades e instituições

não-universitárias, dentro e fora da Nigéria. As características desta ligação nos dizem muito sobre as atividades das universidades.

Agarin e Nwagwu (2008), em seu estudo com indicadores webométricos utilizaram a ferramenta de busca AltaVista, dados sobre *links* foram coletados aleatoriamente a partir de 1000 páginas *web* selecionadas de 30 universidades da Nigéria para estudar o padrão e a frequência dos *outlinks* e *inlinks*. Os *websites* tem um total de 44.567 ligações, representando uma média de 45 ligações por página. Destes, 81,2% eram *inlinks* de *websites* das universidades nigerianas para outros *sites*, enquanto 18,8% eram *outlinks* de outros *websites* para os da universidade nigeriana. As universidades têm maior percentagem de *links* por páginas *web* de destino do que os outros *websites*, que retornaram menos de 50% das páginas *web* como destino. As páginas *web* das universidades nigerianas não fazem links entre si. Segundo os citados autores, os pesquisadores nigerianos têm a tendência de utilizar os métodos tradicionais limitados para publicar os seus trabalhos científicos, em vez de usar os métodos eletrônicos amplamente acessíveis. Além disso, os *sites* parecem ter mais *links* com os *websites* não acadêmicos do que com os acadêmicos. O resultado mostra que existe um baixo nível geral de utilização da *web* para compartilhar e divulgar as informações produzidas por universidades nigerianas.

Com a finalidade de testar a viabilidade dos indicadores webométricos para descrição e classificação das atividades da universidade, conforme mostrado em seus *sites*, Aguillo et al. (2006) compilaram e analisaram um grande conjunto de 9.330 instituições em todo o mundo. Usando características avançadas das ferramentas de busca, o tamanho, a visibilidade e os números de arquivos ricos foram obtidos para cada um dos domínios institucionais das universidades. Encontrou-se correlação estatística entre a classificação da *web* construída em uma combinação de dados webométricos e outros *rankings* universitários, com base em indicadores bibliométricos.

Os resultados deles mostram que as medidas cybermétricas poderiam ser úteis para refletir a contribuição das instituições orientadas tecnologicamente, aumentando a visibilidade dos países em desenvolvimento, e melhorando os *rankings* baseados nos dados do ISI.

Uma comparação entre o *ranking web* e os indicadores bibliométricos foi realizada para determinar uma possível correlação entre eles, em uma análise comparativa entre o *ranking web* e os outros três *rankings* universitários baseados em indicadores bibliométricos tradicionais e outras medidas de prestígio e impacto. Essential Science Indicators¹, Ranking of World Universities (2004), elaborado pela Universidade de Shanghai², e o World University Ranking (2004), publicada pelo Times Higher Education Supplement³.

Foram selecionadas as 100 principais universidades do *ranking* da Universidade de Shanghai. E a correlação entre o *ranking* da *web* e as outras três classificações acadêmicas mostrou uma associação positiva entre eles. O idioma e a produtividade científica são critérios utilizados para *links* entre os *websites* de destino e outras instituições, ou até mesmo organizações não universitárias.

¹ <http://www.esi-topics.com/>

² <http://www.arwu.org/>

³ <http://www.timeshighereducation.co.uk/>

Algumas das deficiências das medidas bibliométricas tradicionais observadas em diferentes estudos podem ser supridas pelos indicadores cybermétricos. A webometria poderia ser útil para avaliar as universidades de Terceiro Mundo, cujos recursos financeiros e acesso às principais publicações científicas são menores. Na web, tanto as informações acadêmicas (em língua local) quanto as científicas (principalmente em inglês) são publicadas.

Por outro lado, as universidades tecnológicas são geralmente subestimadas pelos indicadores bibliométricos, devido ao baixo número de artigos publicados em revistas científicas neste domínio. Assim, estes resultados sugerem que os indicadores cybermétricos são úteis para avaliar as universidades tecnológicas enquanto que os bibliométricos tradicionais têm limitações.

A fim de aproveitarem melhor sua presença na web, alguns pesquisadores ressaltam a importância para as instituições e universidades de terem seus *websites* atualizados, com seus recursos funcionando corretamente, atividades e desempenho, fornecendo aos visitantes uma visão geral da instituição (RWWU, 2009).

- **URL de nomeação** - cada instituição deve escolher um único domínio institucional, que pode ser usado por todos os sites da instituição.
- **Conteúdo: Criar** - uma grande presença na web é possível apenas com o esforço de um grande grupo de autores.
- **Conteúdo: Converter** – tornar, se possível, os conteúdos todos disponíveis na internet, incluindo relatórios de atividades passadas.
- **Interligação** - se o seu conteúdo não for conhecido (design ruim, informação limitada, ou linguagem minoritária), o tamanho for escasso ou tiver baixa qualidade, o site provavelmente receberá poucos links de outros *sites*.
- **Linguagem, especialmente inglês** - Versões da língua, principalmente em inglês, são obrigatórias.
- **Arquivos ricos e mídia** – a largura da banda está crescendo exponencialmente, por isso é um bom investimento para arquivar todos os meios materiais produzidos em repositórios da *web*.
- **Desenhos amistosos da ferramenta de pesquisa** - usar diretórios ou páginas estáticas e evitar menus de navegação pesados baseados em Flash, Java ou JavaScript, que podem bloquear o acesso robô.
- **Popularidade e estatísticas** - número de visitas é importante, mas também é importante monitorar a sua origem, distribuição e as causas por que elas acessam os *sites* das universidades.
- **Arquivamento e persistência** – deve ser obrigatório manter uma cópia do material antigo ou desatualizado no site.
- **Normas para o enriquecimento de sites** - a utilização de títulos significativos e meta-tags descritivos podem aumentar a visibilidade das páginas.

A Internet nos fornece um canal impressionante para conectar-nos e contribuímos para o mundo do aprendizado. Em geral, só temos a ganhar compartilhando nossos conhecimentos e pesquisas com o mundo e, ao fazê-lo de forma interessante, o *ranking* WR poderá melhorar também.

Estudar *websites* das universidades para visualizar a sua presença no mundo virtual da *web* e, em geral, avaliar o seu desempenho, é o mesmo que avaliar o desempenho acadêmico e de pesquisa das universidades. Vários anos atrás, o mundo das universidades tinha sido classificado por suas características bibliométricas, contudo este método apresenta limitações. Uma das medidas importantes que está sendo muito utilizada no *ranking* é a webometria. Esta medida fornece o impacto dos *websites* das universidades na rede por meio de análise de *links*.

Analisando os estudos de pesquisas relatadas, podemos considerar a importância dos *links* para formação do *ranking* das universidades e a sua validação como uma fonte de novas informações. Com ligações entre os *websites* de universidades, por exemplo, uma correlação positiva entre a contagem de *links* e a atividade de pesquisa que fornece alguma evidência de que a criação de *links* não é completamente aleatória pode ser útil para estudar o comportamento acadêmico.

A evolução do *ranking* webométrico das universidades é o resultado da necessidade de medir a adoção da *web* pelas universidades para suas pesquisas acadêmicas, ensino e aprendizagem. As universidades com *ranking* superior, presumivelmente, são aquelas que integraram a *web* em sua cultura de pesquisas, ensino e aprendizagem. Elas tendem a ter mais recursos e informações em seus *websites* e, com isso, têm maior probabilidade de manter mais *links* para outros *sites* e vice-versa. Assim, aumenta seu impacto e visibilidade na *web*, gerando percepção positiva sobre elas e as partes interessadas.

2.2.1.4 Considerações

Analisando os estudos anteriores, pode-se considerar que existem limitações nos métodos de avaliação da qualidade de Conferência Científica, conforme descrito anteriormente, como problemas de interesse que podem ocorrer por meio de *peer-review* e da utilização de indicadores desatualizados, no caso da bibliometria.

Após o estudo sobre os trabalhos relacionados, percebeu-se a necessidade de criar um modelo para a avaliação de conferências que pudessem fazer uso de indicadores mais atualizados e disponíveis na *web* e evidenciar a qualidade delas. A ideia é, a partir dos métodos webométricos, tentar aplicar os indicadores que possam ser utilizados para avaliar Conferências Científicas e outras características externas de visibilidade na *web*, com um conjunto de indicadores de qualidade obtidos através das ferramentas de buscas e *softwares* disponíveis na *web* para desenvolver o modelo. Partindo deste conjunto, pôde-se definir onde cada indicador deveria ser obtido, ou seja, identificar qual fonte de dados serviria para cada indicador.

Para obter estes dados, foi necessário identificar ferramentas que suprissem estas necessidades e que retornassem o maior número de dados indexados por elas. Depois de computar os dados, foi preciso verificar se cada indicador podia comprovar a existência de alguma correlação com o *ranking* do QUALIS. Após este procedimento, foram definidos padrões para os indicadores do modelo. Tais modelos poderão ser verificados no capítulo 3.

3. MODELO PARA AVALIAÇÃO DE CONFERÊNCIA CIENTÍFICA

Têm sido realizadas muitas pesquisas com a webometria, especialmente os impactos de *sites* entre si e o fator de impacto na *web*. No entanto, há poucas pesquisas tratando sobre os *sites* das conferências. Este estudo visa suprir esta lacuna analisando os *sites* das Conferências Científicas da área de CC de acordo com os indicadores webométricos. Há várias análises que utilizam a webometria para obter uma avaliação de visibilidade dos seus *websites*, como relatado anteriormente, mas nenhuma aborda a utilização desse método para conferências.

O trabalho é inspirado em (ZHUANG et al., 2007). No entanto, o modelo que em desenvolvimento é mais complexo, uma vez que não apenas diferenciará entre conferências prestigiadas e não prestigiadas, mas buscará distinguir conferências de alta, média e baixa qualidade (A, B ou C). Além disso, o método não se baseia nos aspectos do comitê de programa, e sim explora um número mais elevado de características baseadas na visibilidade na web como critérios para avaliar a qualidade das conferências.

A partir das metodologias estudadas, buscou-se estruturar um modelo que levasse em consideração a visibilidade na *web* e considerasse também alguns indicadores bibliométricos, a fim de determinar de maneira segura e rápida (automatizando todos os processos possíveis, com a utilização das ferramentas atualmente disponíveis) o conceito de qualidade atribuído a uma determinada conferência, através da análise do seu *link* (perspectiva externa).

A partir da conclusão obtida no estudo anterior, que apresenta deficiências nos modelos tradicionais, motivou a criação de um modelo utilizando indicadores de visibilidade do método webométrico e outros indicadores para ter um conjunto mais completo e tentar mostrar um resultado mais preciso por meio das características *web*.

O objetivo deste trabalho é chamar a atenção para as muitas e novas possibilidades que a análise webométrica oferece atualmente para os que desejam explorar as bases de dados como um arquivo e como um instrumento de análise. Acredita-se que seja preciso aprender a explorar bases de dados *on-line* não somente para ter acesso a conteúdo e informações, mas também para traçar as tendências e o desenvolvimento da sociedade, das conferências científicas e das áreas de pesquisa.

Portanto, a proposta deste trabalho é definir um modelo para avaliação de Conferência Científica. Tal modelo é baseado em indicadores da webometria e outros indicadores e tem como objetivo avaliar as conferências por meio dos dados disponíveis na *web*. Para isso, foram especificados padrões aos indicadores de acordo com os dados obtidos com esta finalidade.

Este trabalho visa demonstrar:

- a visibilidade e o impacto da presença das conferências na *web*;
- uma possibilidade de avaliar *sites* de Conferências Científicas, relativamente a abrangência, conteúdo e atualidade, inferindo a qualidade por meio dos indicadores definidos no modelo.

Para a definição deste modelo, foram elaboradas as seguintes etapas que constituem de um processo, conforme apresentado na figura 3.1.

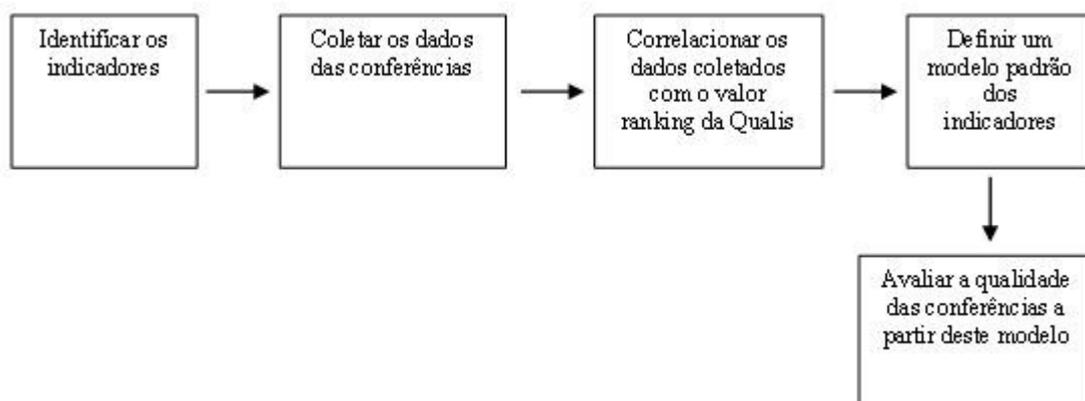


Figura 3.1: Etapas do processo de avaliação de qualidade

Todo o processo de desenvolvimento para a elaboração do modelo foi realizado em 5 etapas, detalhadas a seguir.

3.1 Identificar os indicadores

Nesta etapa do processo foram identificados os indicadores atualizados que podiam ser fornecidos por meio da *web* e no final determinar a qualidade de uma dada conferência.

Para a caracterização deste modelo de avaliação de conferências foram considerados os indicadores que compõem a webometria: o **Google Acadêmico**, que fornece a quantidade de artigos, a média de autores por artigo, a média de citação por artigo, o *h-index* e citações de cada conferência; e a **Visibilidade**, que considera todos os *links* externos (*inlinks*). Outro indicador considerado na construção do modelo foi o **PageRank**.

Os critérios considerados para a composição do modelo de avaliação de conferência de acordo com a webometria e estudos realizados foram:

- **Visibilidade (*visibility*):** são os *links* externos (*inlinks*), isto é, todos os *links* que apontam para a *url* da conferência por edição. Este termo é aqui considerado igual a "citação" em recursos impressos.
- **PageRank:** o *pagerank* é obtido a partir da *url* da conferência para cada edição.
- **H-index:** o *h-index* de uma conferência é obtido a partir da ferramenta *Publish or Perish*.

- Citações (Google Acadêmico): são todas as informações obtidas pela ferramenta *Publish or Perish* que utiliza a base de dados do GA. Neste *ranking* foram considerados 4 indicadores que são:
 - Total de Citações: total de citações que uma conferência recebeu durante todo o período de existência.
 - Média de Citações por artigo: todas as citações de uma conferência dividida pelo total de artigos.
 - Média de autores por artigo: média de autores por artigo de todos os artigos retornados de uma conferência.
 - Total de Artigos: total de artigos de uma dada conferência de todo o período de existência (limitados a 1000).

3.2 Coletar dados das conferências

A partir da definição dos indicadores, foram identificadas as fontes de dados que tinham estas informações e as ferramentas de busca e *software* que pudessem obter maior número de dados das conferências e assim realizar a coleta desses dados.

3.2.1 Ferramentas de Coleta dos Dados

Os estudos webométricos requerem alguns mecanismos que permitam efetuar a busca, a extração, a quantificação, a representação e a visualização das informações disponíveis na *web*. As ferramentas de busca, os programas mapeadores e os programas para representação e visualização de redes são as ferramentas utilizadas para este fim.

Uma das formas de obter dados para os estudos webométricos é por meio das ferramentas de busca e o do *software Publish or Perish*, que consegue indexar grande parte das páginas disponíveis na *web*.

Para a coleta de dados, o primeiro passo foi determinar as ferramentas a serem utilizadas para obter os dados necessários conforme o modelo determinado no capítulo anterior, e o segundo, explorar os comandos e as combinações possíveis a serem empregados na pesquisa.

Para a obtenção dos dados neste trabalho, foram utilizadas as ferramentas descritas a seguir.

3.2.1.1 Yahoo

As ferramentas de busca são utilizadas na webometria para encontrar informações a partir da análise de grandes quantidades de dados. Existem várias ferramentas, porém algumas delas oferecem resultados mais precisos para efeito de análises na webometria.

Para determinar qual ferramenta de busca seria necessária para coletar os dados neste estudo, foi levado em consideração o percentual de páginas *web* que podiam ser indexadas. Uma das ferramentas encontradas para obtenção dos dados, que tende a suprir tal necessidade, é o Site Explorer do Yahoo¹.

¹ <http://br.yahoo.com/>

O Site Explorer do Yahoo permite a exploração de todas as páginas indexadas pelo Yahoo! Cadê?² na *web*. É uma ferramenta que permite acessar a informação sobre a performance e a presença *on-line* de um *site*. Essa ferramenta permite descobrir páginas que têm *links* para aquele *site* (*inlinks*), um dos indicadores utilizados neste trabalho com grande possibilidade de acurácia.

Efetuar uma pesquisa no Site Explorer é fácil como fazer uma busca, mas, ao invés de colocar um termo de busca, coloca-se a URL, por exemplo, a URL www.lirmm.fr/caise08/ da conferência CAiSE - *Conference on Advanced Information Systems Engineering* realizada em 2008. Como retorno, o Site Explorer fornece informações dos *links* do *site* inserido e a quantidade de *links* externos que apontam para aquele *site*, os *inlinks*. Fornece ainda a opção de mostrar todos os *links* externos (*inlinks*) exceto os do domínio da URL, conforme demonstrado na figura 3.2.

The screenshot shows the Yahoo Site Explorer interface. At the top, the URL <http://www.lirmm.fr/caise08/> is entered in the search bar, and the 'Explorar URL' button is visible. Below the search bar, the 'Resultados' section shows 'Páginas (31)' and 'Inlinks (202)'. The 'Exibir Inlinks' dropdown is set to 'Exceto esse domínio'. The results list includes:

- Alessandro F. Garcia - Principal
text/html <http://www.comp.lancs.ac.uk/~garciaa/index.htm> - 18k - cache
- Related Conferences
text/html <http://portal.cin.ufpe.br/ler/Lists/Related%20Conferences/AllItems.aspx>
- Homepage of Wil van der Aalst
text/html <http://is.tn.tue.nl/staff/wvdaalst/> - 24k - cache
- CAISE'07
text/html <http://caise07.idi.ntnu.no/> - 3k - cache
- Borbala Online Conference Services
text/html <http://www.borbala.com/> - 40k - cache
- Roel Wieringa
text/html <http://wwwhome.cs.utwente.nl/~roelw/> - 5k - cache

Figura 3.2: Interface do Yahoo Site Explorer

3.2.1.2 Harzing's Publish or Perish

Um mecanismo muito conhecido e utilizado para avaliar a qualidade externa é o *Publish or Perish*, um programa de *software* simples e gratuito, desenvolvido por Anne-Wil Harzing. Este *software* recupera e analisa citações acadêmicas de autores, *journals* e conferências. Ele usa o GA para obter os dados das citações (HARZING, 2008). Ao realizar uma pesquisa nesse programa, o usuário recupera uma lista das referências pertinentes à expressão de busca e gera estatísticas de produtividade e de impacto de autores, *journals* e conferências. A partir dos dados obtidos, é calculada uma série de métricas, entre elas o *h-index* da conferência, do *journal* e do autor (HIRSCH, 2005), a

² <http://cade.search.yahoo.com/>

média de citações por artigo, tanto de pesquisador quanto de conferência e *journal*, o total de citações de um determinado pesquisador ou de uma determinada conferência, etc.. Cabe salientar que os resultados retornados pelo GA são baseados na análise de citações.

Para realizar uma busca no *software Publish or Perish*, basta informar um termo, por exemplo, a sigla da conferência CaiSE. Como resultado, são fornecidos vários indicadores, dentre os quais foram considerados neste trabalho os seguintes: número total de artigos; número total de citações; número médio de citações por artigo; número médio de autores por artigo e o *h-index*, conforme ilustrado na figura 3.3.

The screenshot shows the 'Journal impact analysis' window of Harzing's Publish or Perish. The 'Query' section has 'CAISE' entered in the 'Journal title' field. The 'Results' section displays the following summary statistics:

Papers:	591	Cites/paper:	3.41	h-index:	20	AWCR:	184.67
Citations:	2015	Cites/author:	982.14	g-index:	28	AW-index:	13.59
Years:	62	Papers/author:	333.15	hc-index:	17	AWCRpA:	84.85
Cites/year:	32.50	Authors/paper:	2.26	hI-index:	8.33		
				hI,norm:	11		

Below the summary is a table of individual articles with columns for Cites, Per year, Rank, Authors, Title, and Year. The first few rows are:

Cites	Per year	Rank	Authors	Title	Year
107	5.63	308	G WiederholcP, S Ja...	Dealing with granularity of time in te...	1991
58	9.67	311	K Bařna, B Benatall...	Model-driven web service development	2004
39	0.00	160	M Adams, AHM ter ...	Facilitating flexibility and dynamic ex...	
37	7.40	316	BF van Dongen, W...	Verification of EPCs: Using reduction...	2005
32	6.40	56	A Moreira, J Araújo...	A concern-oriented requirements en...	2005
30	2.31	317	K Pohl, R Domges, ...	Towards Method-Driven Trace Capt...	1997
29	0.00	318	K Smolander, VP Ta...	How to Combine Tools and Methods i...	90
29	4.83	101	MC Jaeger, S Tang	Ranked matching for service descrip...	2004
29	0.00	231	M Blaschka	FIESTA: A Framework for Schema E...	
28	5.60	161	B Benatallah, F Cas...	On temporal abstractions of web ser...	2005
28	0.00	294	MM Fonkam, WA Gray	An approach to eliciting the semantic...	

Figura 3.3: Interface de entrada de dados do *Publish or Perish*

3.2.2 Coleta dos dados

Para auxiliar na elaboração do modelo e confirmação dos parâmetros a serem aplicados, realizou-se uma pesquisa na qual foram selecionadas 44 Conferências Científicas da área de Ciência da Computação, com ênfase em Banco de Dados, consideradas de nível internacional. Nesta seleção, foram considerados os seguintes indicadores: o *h-index* global (acumulado de todas as edições) da conferência, a quantidade de artigos daquela conferência, o total de citações de todos os artigos, o número médio de citações por artigo e a média de autores por artigo. Estes dados foram extraídos através da ferramenta *Publish or Perish*. Para a obtenção dos *links* externos (*inlinks*), foi restringido o intervalo, considerando-se as conferências que ocorreram entre 2004 e 2008. Para obter estes dados, foi utilizado o Yahoo Site Explorer através da URL de cada edição da conferência e no final foi efetuada a soma total das edições ocorridas no período. Há alguns casos especiais em que a conferência é realizada de forma bi-anual e outros dois casos de conferências criadas recentemente, como a

“ADMA - *International Conference on Advanced Data Mining and Applications*”, que teve sua primeira edição no ano de 2005, e a “CIDM - *IEEE Symposium on Computational Intelligence and Data Mining*”, que se iniciou em 2007.

Para a pesquisa realizada na ferramenta *Publish or Perish* com a finalidade de obter dados, foi informada a sigla, o nome ou parte do nome de uma dada conferência e foram considerados os que obtiveram o número mais alto de ocorrências em artigos. Em cada pesquisa efetuada, ocorreu uma limpeza nos dados. Esta limpeza permite ajudar na precisão de tais dados evitando informações erradas, como, por exemplo, com a sigla VLDB, que é uma conferência e também um *journal*. Esta ferramenta permite visualizar as informações obtidas possibilitando desmarcar as que não pertencem à conferência, conforme figura 3.4, podendo-se também restringir algumas palavras durante a consulta.

The screenshot shows the 'Harzing's Publish or Perish' software interface. The main window displays a 'Multi-query center' with a table of queries. The first query is 'VLDB, conference, NOT journal: eng' with 997 papers, 21065 citations, and a query date of 17/07/2009. An 'Edit Query' dialog box is open, showing the current query and options to modify it. The 'None of the words' field is circled in red. The 'Results' section shows a summary of the query results: 997 papers, 21065 citations, 34 years, and 619.56 citations per year. A table below shows the top results, with the first three rows circled in red.

Name	Papers	Cites	Cites	Auth	h	g	Query Date
VLDB, conference, NOT journal: eng	997	21065	619.56	2.47	73	117	17/07/2009

Cites	Per year	Year	Author	Title	Year	Journal
9	0.31	397	W Kent	Data model theory meets a practical ...	1981	... the seventh international ...
9	0.31	399	K Karlsson	Reduced cover-trees and their appli...	1981	... the seventh international ...

Figura 3.4: Interface de consulta do *Publish or Perish*.

O *PageRank* foi obtido através do *site* da conferência pela ferramenta de busca Google. Em alguns casos, o *site* não está mais ativo, então foi dado como resultado nulo. Foi considerado o *ranking* de cada conferência, levando-se em conta o mesmo período utilizado para o cálculo dos *links* externos, e foi calculada a média de todas as edições, inclusive quando o resultado foi zero, de acordo com a quantidade de edições que a página mantivesse no ar.

Todos os dados coletados na tabela 3.1 foram extraídos até a data de 17/07/2009. Como estes dados podem sofrer modificações, está sendo especificada a sua data de coleta.

Tabela 3.1: Conferências analisadas.

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações / artigo</i>	<i>Autores / artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
KDD	24	737	5168	7,01	2,42	360	5,40
SIGMOD	140	1000	70865	70,87	2,48	952	4,80
VLDB	73	997	21065	21,13	2,47	1011	5,60
ICDE	100	996	40209	40,37	2,82	1202	5,75
PODS	11	321	763	2,38	2,16	952	4,80
ER	18	312	1102	3,53	2,10	383	5,00
JCDL	14	256	730	2,85	2,70	1020	5,80
ICDM	33	769	6276	8,16	2,82	1369	4,60
CIKM	17	374	1526	4,08	2,25	1111	5,25
SIGIR	9	300	583	1,94	2,12	1197	5,20
PKDD	13	286	967	3,38	2,47	1461	5,50
EDBT	29	434	4658	10,73	2,21	343	3,00
SDM	15	121	1003	8,29	2,80	841	5,00
ICDT	21	183	1669	9,12	2,10	158	5,00
CIDM	4	103	74	0,72	2,85	47	
DaWaK	11	91	388	4,26	2,43	47	4,40
ODBASE	5	33	105	3,18	2,55	169	4,60
DEXA	33	1000	7176	7,18	2,56	47	4,40
ADBIS	9	185	477	2,58	2,22	244	4,00
WIDM	5	41	125	3,05	2,24	218	4,40
IDEAS	19	331	1653	4,99	2,65	88	2,25
ECDL	10	101	442	4,38	2,32	1242	6,50
SSDBM	31	354	4130	11,67	2,93	214	5,00
SPIRE	18	164	1176	7,17	2,36	268	4,75
DOLAP	3	31	41	1,32	2,35	101	4,25
MIR	4	24	59	2,46	2,79	193	4,60
DBPL	16	143	1092	7,64	2,15	87	5,00
DASFAA	20	262	1511	5,77	2,58	405	4,60
WebDB	20	141	1837	13,03	2,62	213	5,00
ECIR	8	88	559	6,35	2,35	609	5,50

TREC	44	999	9931	9,94	2,94	133	4,60
RIDE	33	283	4422	15,63	2,84	56	4,67
XSym	3	24	41	1,71	2,71	112	4,25
ADMA	1	23	5	0,22	2,17	248	5,00
FQAS	6	74	159	2,15	2,46	54	5,00
NLDB	7	37	280	7,57	2,70	70	4,67
GIR	8	24	195	8,13	2,42	82	2,60
DIWeb	3	14	35	2,50	2,36	18	5,00
TAKMA	1	1	1	1,00	5,00		
IKE	5	37	100	2,70	2,54	27	3,67
DCC	51	1000	12677	12,68	2,28		
CIDR	22	90	2485	27,61	3,41	63	5,00
ASIS&T	6	40	118	2,95	1,80	462	5,60
ISMIR	19	147	1445	9,83	2,46	632	5,00

Para a definição do indicador final da conferência, utilizaram-se métodos estatísticos, como o coeficiente de correlação, para verificar se os resultados obtidos em cada consulta traduziam a real indicação de qualidade daquele evento, e se os parâmetros adotados eram condizentes e estavam suficientemente calibrados para atingir o resultado pretendido. O modelo de comparação que assegurou ser condizente o indicador final auferido foi o conceito da QUALIS (já descrito no item 2.1 “Indicadores Bibliométricos”), ao qual se acrescentou o valor *ranking* utilizado para a atribuição do conceito. Quanto mais próximo estivesse o resultado, maior seria a precisão do modelo sugerido.

Na tabela 3.2, são mostradas as conferências com seus respectivos valores de *ranking*, de acordo com o conceito QUALIS.

Tabela 3.2: Conferências com indicador QUALIS.

<i>Conferência</i>	<i>Valor ranking</i>	<i>Conceito QUALIS</i>	<i>Conferência</i>	<i>Valor ranking</i>	<i>Conceito QUALIS</i>
KDD	0,943	A	SSDBM	0,581	B
SIGMOD	0,927	A	SPIRE	0,576	B
VLDB	0,899	A	DOLAP	0,568	B
ICDE	0,898	A	MIR	0,562	B
PODS	0,863	A	DBPL	0,535	B
ER	0,856	A	DASFAA	0,515	B
JCDL	0,856	A	WebDB	0,508	B

ICDM	0,854	A	ECIR	0,495	B
CIKM	0,819	A	TREC	0,486	B
SIGIR	0,811	A	RIDE	0,459	B
PKDD	0,748	A	XSym	0,444	B
EDBT	0,711	A	ADMA	0,433	B
SDM	0,698	A	FQAS	0,420	C
ICDT	0,667	A	NLDB	0,414	C
CIDM	0,659	A	GIR	0,411	C
DaWaK	0,620	B	DIWeb	0,375	C
ODBASE	0,611	B	TAKMA	0,361	C
DEXA	0,609	B	IKE	0,352	C
ADBIS	0,604	B	DCC	0,333	C
WIDM	0,600	B	CIDR	0,333	C
IDEAS	0,581	B	ASIS&T	0,323	C
ECDL	0,581	B	ISMIR	0,321	C

3.3 Correlacionar os dados coletados com o valor do ranking da QUALIS

Nesta etapa, após ter computado os dados das conferências, foi correlacionado cada indicador com o valor do ranking da QUALIS para verificar se tal indicador poderia ser utilizado como critério de qualidade de uma dada conferência.

3.3.1 Análise dos dados

Para realizar a análise dos dados coletados e verificar se condizem como indicadores de qualidade, foi utilizado o coeficiente de correlação entre o *ranking* da QUALIS e os indicadores definidos no modelo. Para melhor visualizar esta correlação, utilizaram-se os gráficos de dispersão, conforme descrito a seguir.

3.3.1.1 Coeficiente de correlação

O coeficiente de correlação procura demonstrar a existência ou não de relações entre variáveis de estudo. Esta informação é utilizada para tomadas de decisão, sendo importante tanto a informação de que existe uma relação linear entre as variáveis como o conhecimento do grau de intensidade dessa relação (MCGRANE; SMAILES, 2002).

Se considerarmos duas variáveis x e y não independentes, pode-se dizer que, se elas estiverem estreitamente relacionadas, o conhecimento dos valores de uma delas nos diz muito sobre o valor da outra, ao contrário do que ocorre se o relacionamento entre as duas for fraco. Assim, para medir o relacionamento linear entre as variáveis, utiliza-se o coeficiente de correlação. Os valores que a correlação pode assumir estão entre -1

(inversamente proporcional) e 1 (diretamente proporcional). O valor -1 indica uma perfeita correlação linear negativa. O valor 1 indica uma perfeita correlação linear positiva. Quando o valor for zero, isso indica que não existe correlação linear entre as variáveis.

Para definição da correlação entre variáveis quantitativas, o método utilizado foi o Coeficiente de Correlação Linear de Pearson, que é um valor numérico medidor da intensidade da associação linear existente entre duas variáveis, obtido a partir de uma série de observações, e que é calculado a partir da equação 3.1:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (3.1)$$

Uma das propostas para interpretação do coeficiente de correlação Pearson é a metodologia desenvolvida por SHIMAKURA (2006), que estabelece um sistema de classificação para o grau de correlação, conforme figura 3.5.

Valor de ρ (+ ou -)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Figura 3.5: Tabela de classificação do coeficiente de Pearson (SHIMAKURA, 2006).

A tabela 3.3 mostra o grau de correlação calculado para cada indicador. Este cálculo foi realizado com cada indicador e valor do *ranking* da Qualis. Por exemplo, os *inlinks* das Conferências Científicas correlacionaram-se com o valor de *ranking* do Qualis, gerando como resultado o grau de correlação de $r=0,639$. Os *inlinks*, artigos, *h-index* e as citações foram as variáveis que apresentaram maior grau de correlação (Tabela 3.3), o que, segundo a figura 3.5, indica uma correlação moderada. Já os autores por artigo apresentaram uma correlação bem fraca com o QUALIS.

Tabela 3.3 – Grau de Correlação de cada Indicador

<i>Indicador</i>	<i>Grau de Correlação</i>
<i>inlinks</i>	0,639
artigos	0,487
<i>h-index</i>	0,440

citações	0,411
citações/artigo	0,289
PageRank	0,243
autores/artigo	-0,191

3.3.1.2 Gráfico de Dispersão

Um dispositivo bastante útil para se verificar associação entre duas variáveis quantitativas é o gráfico de dispersão que analisa o comportamento de um conjunto medido através do coeficiente de correlação.

O cálculo do coeficiente da correlação Pearson foi realizado utilizando as variáveis X e Y. No gráfico 3.1, o X representa a quantidade total de *inlinks* das Conferências Científicas, e o Y representa o valor *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,639$.

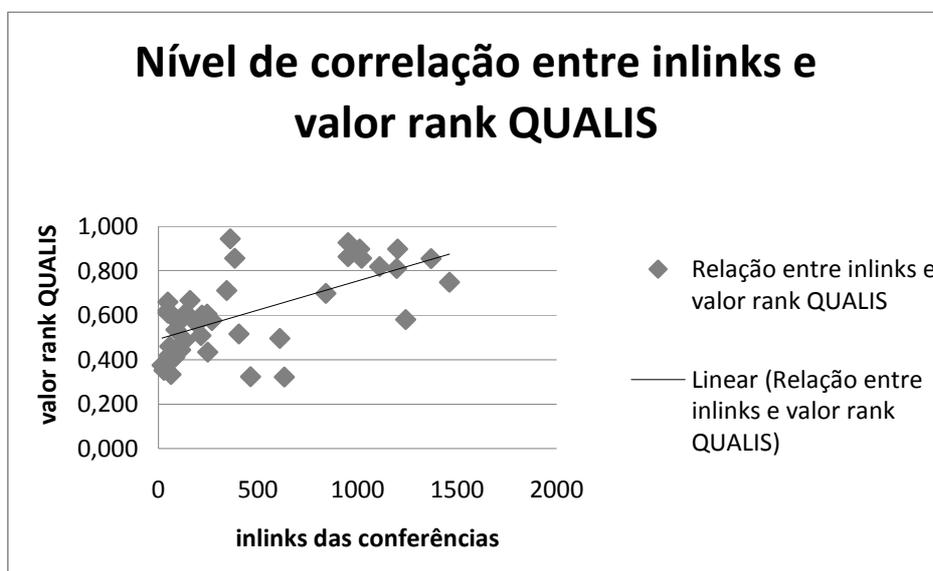


Gráfico 3.1: Correlação entre *inlinks* e valor do *ranking* QUALIS

Por intermédio do tratamento e da análise das amostras coletadas no período examinado por este trabalho, foi possível observar a existência de uma relação entre os *inlinks* e o valor do *ranking* da QUALIS, indicando uma correlação moderada e diretamente proporcional, ou seja, à medida que ocorreu um aumento nos *inlinks*, observou-se um aumento proporcional do valor *ranking* da QUALIS.

No gráfico 3.2, a variável X representa a quantidade total de artigos das Conferências Científicas e Y representa o valor do *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,487$.

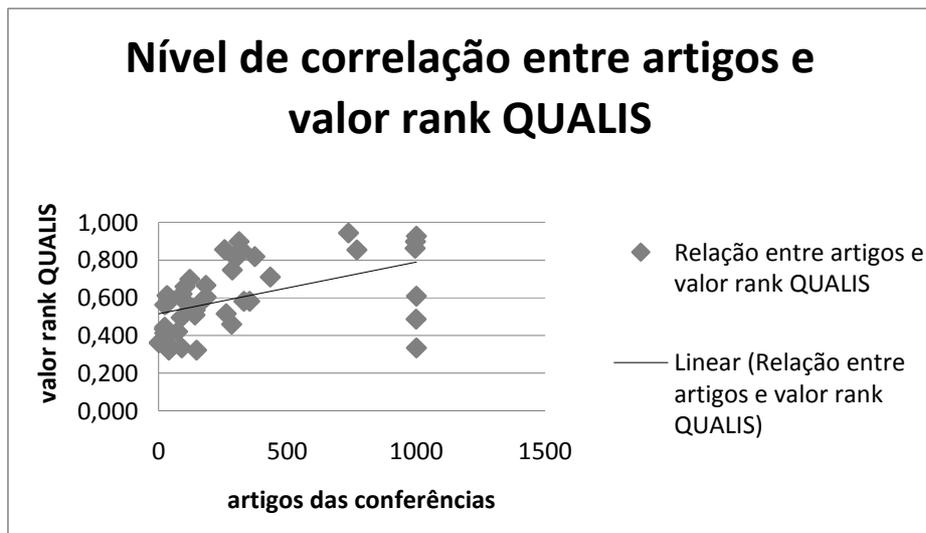


Gráfico 3.2: Correlação entre artigos e valor do *ranking* QUALIS

No cálculo da correlação entre *h-index* e o valor do *ranking* da QUALIS, o X representa o *h-index* das Conferências Científicas e o Y representa o valor do *ranking* da QUALIS. Neste cálculo, obtivemos como grau de correlação o valor de $r=0,440$, conforme gráfico 3.3.

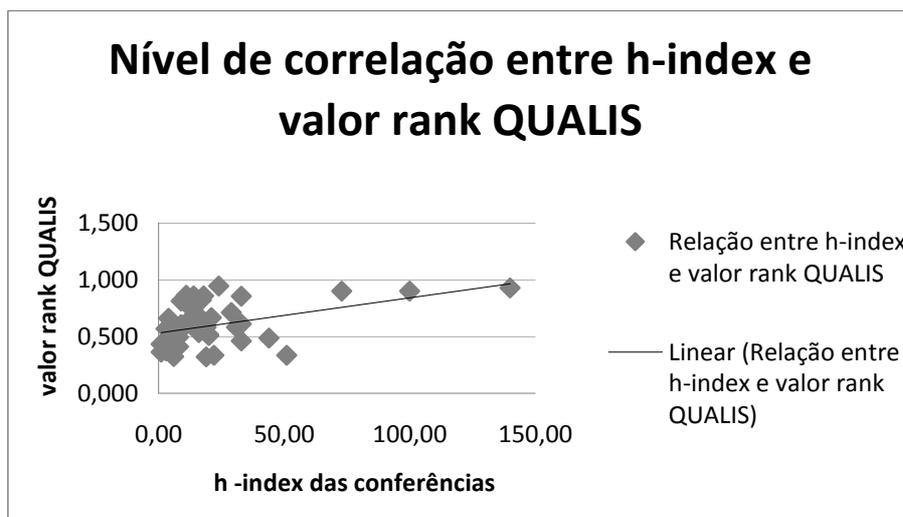


Gráfico 3.3: Correlação entre h-index e valor do rank QUALIS

No gráfico 3.4, a variável X representa o número total de citações das Conferências Científicas e Y representa o valor do *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,411$.

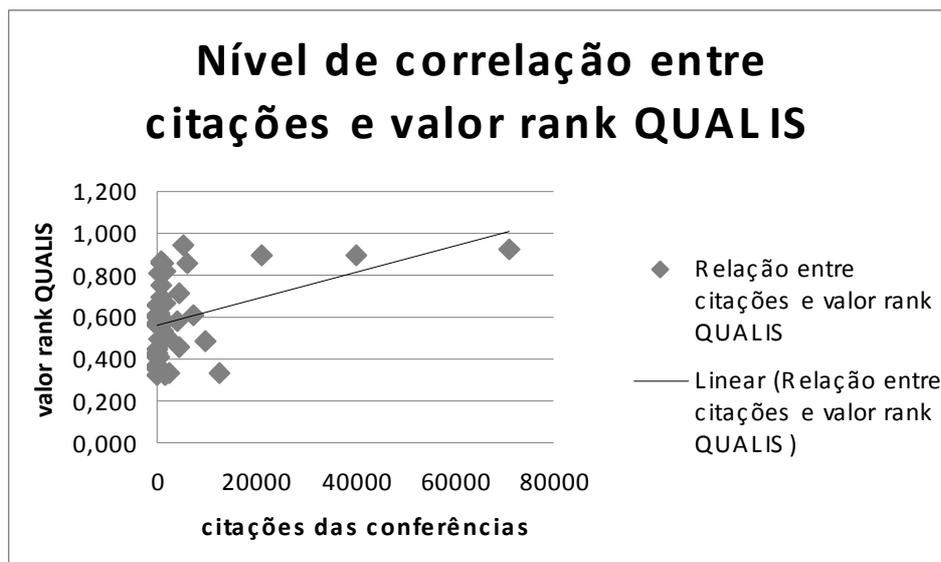


Gráfico 3.4: Correlação entre citações e valor do *ranking* QUALIS

No gráfico 3.5, a variável X representa o total das citações por artigo das Conferências Científicas e Y representa o valor do *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,289$.

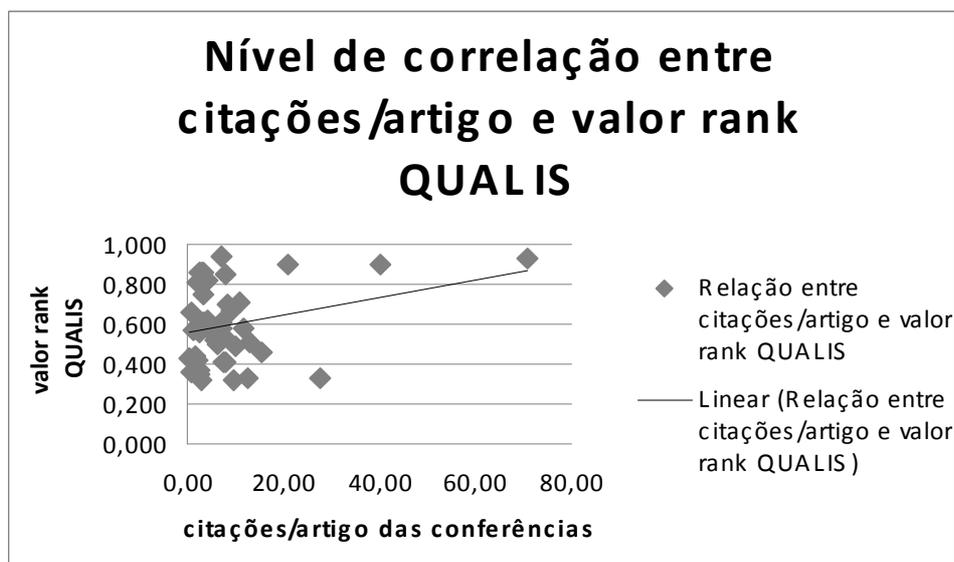


Gráfico 3.5: Correlação entre citações/artigo e valor do *ranking* QUALIS

Por intermédio do tratamento e da análise das amostras coletadas no período examinado por este trabalho, foi possível observar a existência de uma relação entre as citações por artigo e o valor do *ranking* da QUALIS, indicando uma correlação fraca.

No gráfico 3.6, a variável X representa o total do *PageRank* das Conferências Científicas e Y representa o valor do *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,243$.

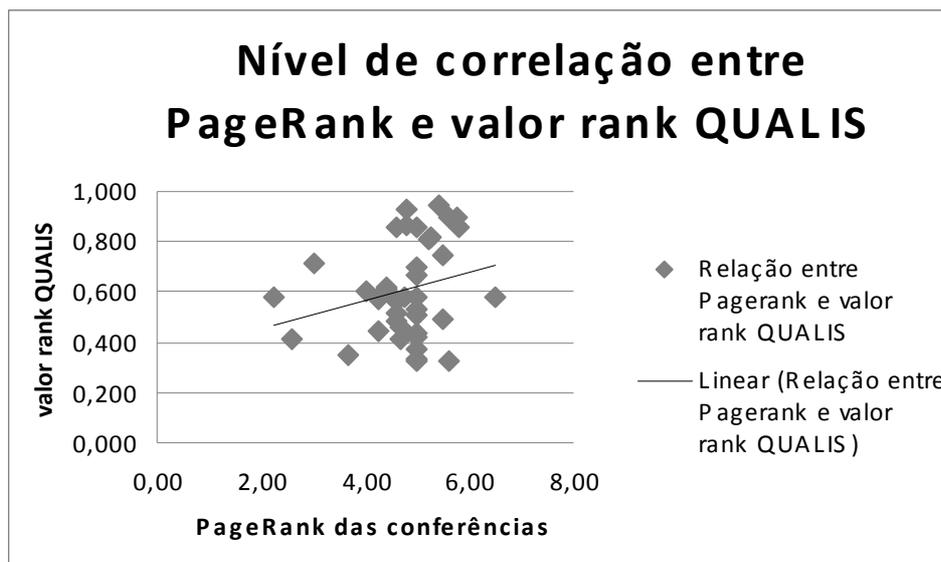


Gráfico 3.6: Correlação entre PageRank e valor do rank QUALIS

No gráfico 3.7, a variável X representa o total de autores por artigo das Conferências Científicas e Y representa o valor do *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=-0,191$.

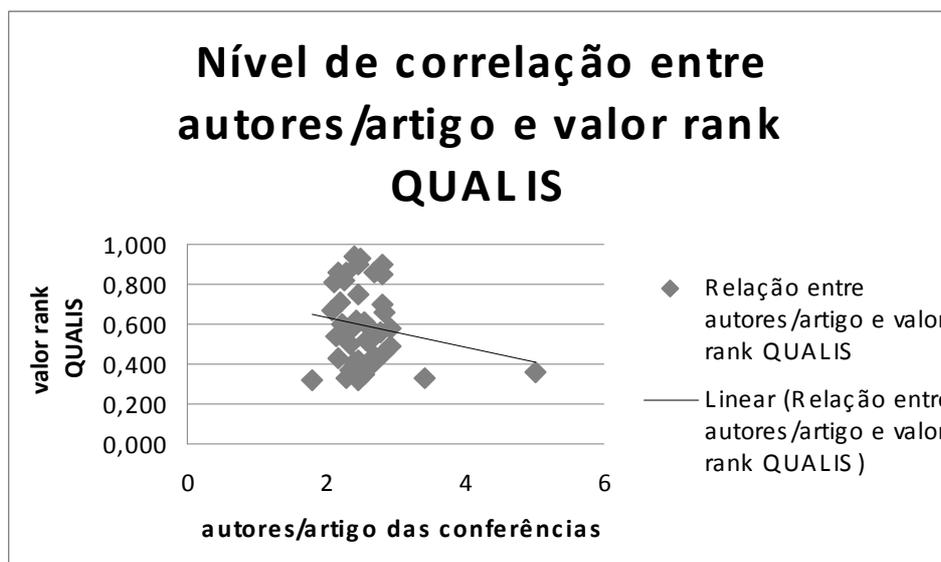


Gráfico 3.7: Correlação entre autores/artigo e valor do *ranking* QUALIS

Observou-se uma correlação linear moderada entre as variáveis - artigos, *h-index*, *inlinks* e citações – e o valor do *ranking* da QUALIS. As correlações para as variáveis citação/artigo e *PageRank* apresentaram valores de $r=0,289$ e $r=0,243$ respectivamente, indicando correlações diretamente proporcionais (fraca positiva). Para a variável autores/artigo, o valor expresso pelo coeficiente de Pearson foi de $r=-0,191$, indicando uma correlação bem fraca e inversamente proporcional.

3.4 Definir um modelo-padrão dos indicadores

Para construir um esquema de avaliação de conferências, foram correlacionados os indicadores definidos no modelo e o valor do *ranking* da QUALIS, resultando no grau de correlação conforme tabela 3.3, então foi preciso definir o conjunto final de indicadores que fazem parte do modelo e determinou-se um esquema de *ranking* para o modelo a partir dos dados coletados sobre 44 Conferências Científicas da área de CC. Foram então definidos padrões e pesos aos indicadores, bem como os pesos atribuídos aos estratos.

A melhor maneira de construir o *ranking* é combinando um conjunto de indicadores que medem esses diferentes aspectos. Os quatro *rankings* foram combinados de acordo com o esquema da figura 3.6, na qual cada indicador tem um peso diferente. Para a definição dos pesos do esquema, foi considerado de acordo com a tabela 3.3, atribuindo-se maior peso para os indicadores que obtiveram maior grau de correlação.

Ranking Conference	
Visibilidade (inlinks)	PageRank 10%
50%	
H-index	Google Acadêmico
20%	20%

Figura 3.6: Esquema da webometria adaptada para conferência

Os dados dos quatro indicadores foram obtidos a partir dos resultados quantitativos fornecidos pelas principais ferramentas de pesquisa. Na realidade, os valores obtidos representam a visibilidade que o *site* em questão tem na *web*.

Tendo definido o esquema para avaliação da qualidade e realizado a coleta dos dados, foram atribuídos padrões para cada indicador. Considerando um conjunto de 44 conferências científicas da área de Ciência da Computação do grupo 2 da QUALIS, de que fazem parte “Bancos de Dados, Recuperação de Informação, Bibliotecas Digitais e Mineração de Dados”.

Neste processo para definir cada padrão dos indicadores foi levado em conta os dados obtidos ordenando em ordem decrescente e definidas três categorias para cada indicador. Os padrões podem ser visualizados nas tabelas abaixo.

Tabela 3.4: Tabela dos pesos e padrões atribuídos aos indicadores GA

<i>Artigos</i>	<i>Autor/Artigo</i>	<i>Citação</i>	<i>Citação/Artigo</i>	<i>Peso %</i>	<i>Peso</i>
1-100	2,3	500	0,1-3	35%	1,66
300	2,6	2000	3,01-8	70%	3,32
+301	+2,61	+2001	+8,01	100%	5

A partir dos dados referentes às Conferências Científicas coletados e ordenados em tabelas de maneira decrescente, foram definidas três categorias: a primeira categoria

representa 35% do peso que equivale ao valor de 1,66 (por exemplo, uma conferência com 98 artigos receberia um valor de 1,66); na segunda categoria, com peso de 70% do valor (por exemplo, uma conferência com citação entre 501 a 2000 receberia um valor de 3,32) e na terceira é o peso máximo de acordo com os padrões definidos.

No final, realiza-se o somatório dos quatro indicadores que compõem o GA de acordo com cada peso definido totalizando os 20% do esquema da figura 3.6.

Tabela 3.5: Tabela dos pesos e padrões atribuídos aos indicadores h-index

<i>h-index</i>	<i>Peso %</i>	<i>Peso</i>
1-10	35%	6,66
11-20	70%	16,32
+≥21	100%	20

Na tabela 3.5, utiliza-se as três categorias citadas, mas com os valores diferenciados de acordo com o esquema da figura 3.6. O *h-index* equivale 20%.

Tabela 3.6: Tabela dos pesos e padrões atribuídos aos indicadores Pagerank

<i>PageRank</i>	<i>Peso %</i>	<i>Peso</i>
1-2	35%	3,33
3-4	70%	6,66
+≥5	100%	10

Na Tabela 3.6, utiliza-se as três categorias citadas, mas com os valores diferenciados de acordo com o esquema da figura 3.6. O *pagerank* equivale a 10%.

Tabela 3.7: Tabela dos pesos e padrões atribuídos aos indicadores visibilidade

<i>Visibilidade</i>	<i>Peso %</i>	<i>Peso</i>
1-150	35%	16,66
151-400	70%	33,20
+≥401	100%	50

Para a visibilidade, a partir dos *inlinks* obtidos de cada conferência, utilizaram-se as categorias citadas gerando os pesos e padrões. A tabela 3.7 equivale a 50% do esquema da figura 3.6 definido para classificar a qualidade das Conferências Científicas.

Indicadores que compõem o esquema da figura 3.6 são: visibilidade (V); *pagerank* (PG); Google Acadêmico (GA) e o *h-index* (H).

No final é calculado o somatório de todos os valores dos indicadores e gerado um índice de qualidade.

$$\text{Índice} = ((V) + (PG) + (GA) + (H))$$

Este índice é analisado e classificado de acordo com os valores da tabela 3.8, e atribuído um estrato a este índice como (A, B, C ou D).

Os pesos atribuídos para cada um dos estratos são indicados na tabela 3.8.

Tabela 3.8: Tabela dos pesos atribuídos aos estratos

<i>A</i>	<i>B</i>	<i>C</i>	
100	70	50	
100-71	70-51	50-1	

3.5 Avaliar a qualidade das conferências a partir deste modelo

Nesta etapa do processo, com os dados obtidos, foi auferida a qualidade das conferências (como A, B, C ou D), de acordo com os padrões definidos na tabela 3.8.

A aplicação dos critérios propostos acima gera uma distribuição nos estratos conforme gráfico 3.8. Esta distribuição mostra que 25% das conferências são classificadas como B, respectivamente, 27% foram classificadas como C e 48% as conferências de maior prestígio.

Como o conjunto de 44 conferências científicas da área de Ciência da Computação do grupo 2 da QUALIS, de que fazem parte “Bancos de Dados, Recuperação de Informação, Bibliotecas Digitais e Mineração de Dados”, sendo que nenhuma satisfaz os requisitos definidos, D=0 para este conjunto.

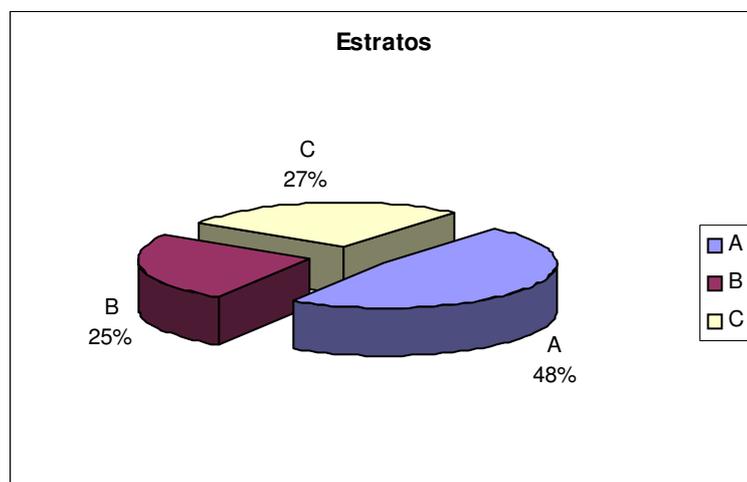


Gráfico 3.8: Distribuição dos estratos, considerando o conjunto de conferências.

4. EXPERIMENTOS

Neste capítulo apresentaremos três conjuntos de experimentos realizados para avaliar a efetividade do modelo para classificação de Conferências Científicas. O objetivo é mostrar o funcionamento do modelo definido no capítulo 3.

No primeiro experimento, todos os critérios apresentados foram utilizados para criar um *ranking* das 44 conferências. Foram utilizados os dados coletadas no capítulo 4 e aplicado o modelo definido no capítulo 3. No segundo experimento, utilizamos novamente todos os indicadores apresentados para classificar essas conferências e utilizamos uma amostra de dados de 14 conferências científicas da área de Sistemas de Informação. E no último experimento foi utilizado um conjunto de 62 conferências da área de Engenharia de Software e Métodos Formais. O segundo e o terceiro experimentos foram utilizados para confirmar a validação e efetividade do modelo que pode ser visualizado os resultados do *ranking* no Apêndice C. Nos experimentos, foi possível observar o novo *ranking* dessas conferências baseados na visibilidade na web como uma nova maneira para medir a qualidade das conferências, podendo mostrar que este método é adequado.

Na tabela 4.1, podemos verificar a classificação das conferências com o seu valor de *ranking* e seu estrato de acordo com os padrões determinados anteriormente.

Tabela 4.1 – Classificação de conferências BD

<i>Conferência</i>	<i>Valor ranking</i>	<i>Conceito</i>	<i>Conferência</i>	<i>Valor ranking</i>	<i>Conceito</i>
ICDE	1,000	A	CIDR	0,633	B
VLDB	0,983	A	TREC	0,633	B
ICDM	0,967	A	RIDE	0,616	B
SIGMOD	0,950	A	DEXA	0,600	B
SDM	0,930	A	ADMA	0,565	B
ISMIR	0,913	A	MIR	0,565	B
JCDL	0,896	A	ADBIS	0,548	B
CIKM	0,896	A	WIDM	0,548	B
PKDD	0,896	A	DBPL	0,546	B

DASFAA	0,863	A	IDEAS	0,529	B
PODS	0,846	A	DaWaK	0,496	C
SSDBM	0,832	A	FQAS	0,416	C
KDD	0,798	A	DIWeb	0,416	C
ECIR	0,783	A	NLDB	0,416	C
ECDL	0,783	A	XSym	0,400	C
SIGIR	0,766	A	ODBASE	0,399	C
EDBT	0,765	A	DOLAP	0,383	C
ICDT	0,765	A	IKE	0,383	C
WebDB	0,762	A	GIR	0,383	C
ER	0,762	A	DCC	0,367	C
ASIS&T	0,733	A	CIDM	0,350	C
SPIRE	0,695	B	TAKMA	0,166	C

O objetivo do primeiro experimento consiste em validar o método. Buscou-se identificar variações nas qualificações com base no modelo.

4.1 Resultados

Nestes experimentos foi possível avaliar os resultados do *ranking* das conferências obtidos com a aplicação do modelo.

A partir do novo valor de *ranking* da conferência, foram utilizados métodos estatísticos, como o coeficiente de correlação, para verificar se os resultados obtidos no *ranking* de cada consulta traduziam a real indicação de qualidade daquele evento, e se os parâmetros adotados eram condizentes e estavam suficientemente calibrados para atingir o resultado pretendido. O modelo de comparação que assegurou ser condizente ao novo *ranking* auferido foi o conceito da QUALIS (já descrito no item 2.1 “Indicadores Bibliométricos”), ao qual se acrescentou o valor *ranking* utilizado para a atribuição do conceito. Quanto mais próximo estivesse o resultado, maior seria a precisão do modelo sugerido.

Para demonstrar e verificar se o resultado do novo *ranking* condiz foi utilizado o coeficiente de correlação entre o valor de *ranking* da QUALIS e o valor de *ranking* obtido por meio da aplicação dos padrões do modelo. Para melhor visualizar esta correlação, utilizaram-se os gráficos de dispersão, conforme descrito a seguir.

No gráfico 4.1, a variável X representa o novo valor de *ranking* das Conferências Científicas de Banco de Dados e Y representa o valor de *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,637$.

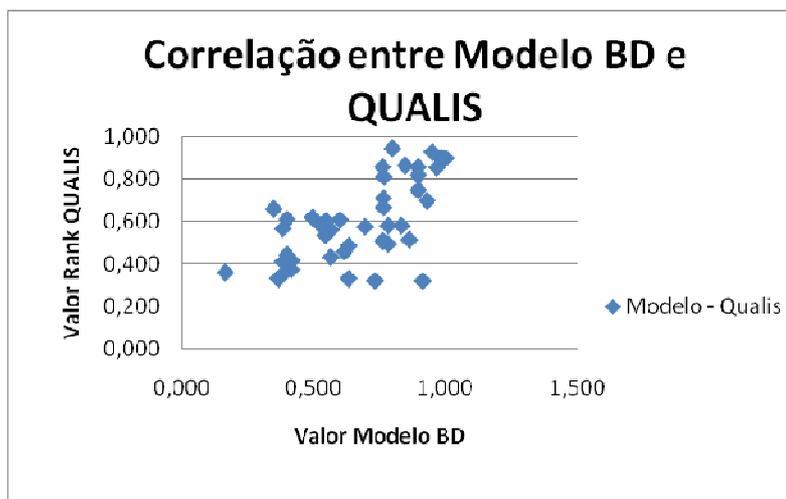


Gráfico 4.1: Correlação entre modelo BD e QUALIS

No gráfico 4.2, a variável X representa o novo valor de *ranking* das Conferências Científicas de Sistema de Informação e Y representa o valor de *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,688$.

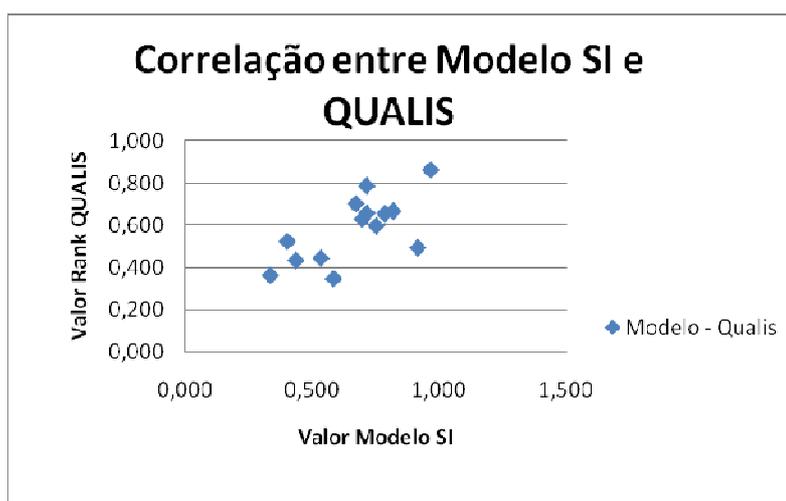


Gráfico 4.2: Correlação entre modelo SI e QUALIS

Nos gráficos 4.1 e 4.2 é possível visualizar os resultados da comparação entre os *rankings* confirmando que este modelo pode ser utilizado para avaliações de conferências, pois foi possível evidenciar a existência de uma correlação entre o novo valor de *ranking* e o valor de *ranking* da QUALIS, indicando uma correlação moderada e diretamente proporcional, ou seja, à medida que ocorreu um aumento nos resultados do novo valor de *ranking*, tanto no gráfico 4.1 quanto no 4.2 observou-se um aumento proporcional do valor *ranking* da QUALIS.

No gráfico 4.3, a variável X representa o novo valor de *ranking* das Conferências Científicas de Engenharia de Software e Y representa o valor de *ranking* da QUALIS. Como resultado, obtivemos o grau de correlação de $r=0,968$.

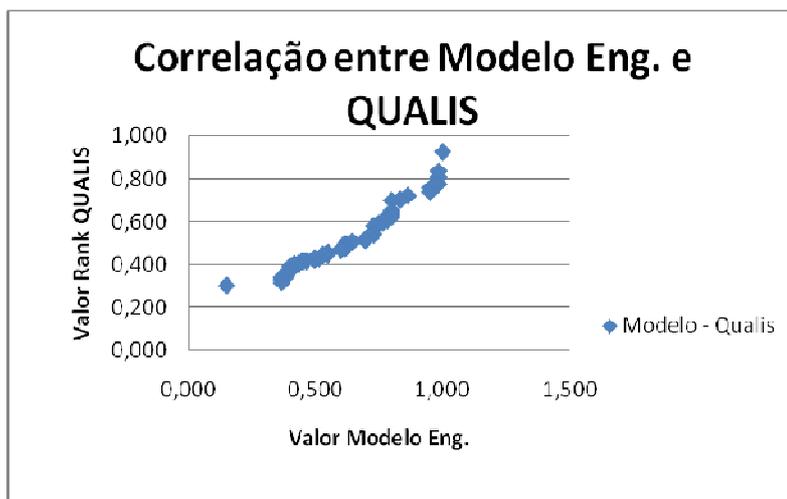


Gráfico 4.3: Correlação entre modelo Eng. e QUALIS

Neste gráfico diremos que as variáveis estão positivamente correlacionadas com o valor de r muito próximo de 1 apresentando uma correlação linear muito forte, conforme figura 3.5.

Na tabela 4.2, podemos verificar a classificação das conferências com o seu novo valor de *ranking* e seu estrato de acordo com os padrões determinados anteriormente e também o valor de *ranking* e o valor de estratos da QUALIS.

Tabela 4.2 – Comparação entre os resultados

<i>Conferência</i>	<i>Valor ranking QUALIS</i>	<i>Conceito QUALIS</i>	<i>Valor ranking</i>	<i>Conceito</i>
CAiSE	0,861	A	0,963	A
AMCIS	0,494	B	0,913	A
SCC	0,667	A	0,816	A
BPM	0,655	A	0,783	A
ECIS	0,598	B	0,750	A
ICEIS	0,657	A	0,712	A
CoopIS	0,786	A	0,711	A
IRI	0,630	A	0,695	B
ICIS	0,701	A	0,669	B
EMMSAD	0,348	C	0,582	B
IBIMA	0,444	B	0,532	B
ICEIMT	0,435	B	0,433	C
ICSOC	0,524	B	0,399	C
IAIM	0,364	C	0,333	C

A tabela 4.2 mostra os resultados da classificação gerada pelo modelo de classificação e os resultados da QUALIS que classificou apenas Conferências Científicas de Sistema de Informação. Considerando 14 conferências, sete foram classificadas como sendo de nível "A", somente duas foram consideradas como "A" no modelo e sendo de nível "B" pelos resultados da QUALIS. Quatro foram classificadas como de nível "B" tendo um diferencial maior com os resultados da QUALIS. Três conferências foram classificadas como de nível "C" cumprindo os critérios.

Analisando os resultados, podemos visualizar que sete foram classificadas de acordo com os resultados do *ranking* das conferências da QUALIS-CAPES e que os resultados das outras sete foram interpretados com um pequeno diferencial de um nível para mais ou para menos, sendo que nenhuma conferência mudou dois níveis. Uma hipótese para esta diferença pode ser a utilização dos dados de comparação um pouco desatualizados.

5. CONCLUSÃO

Neste trabalho foram estudadas várias métricas de classificação baseadas na qualidade intrínseca, seja de publicações, de autores ou de conferências. Como relatado anteriormente, o objetivo é fornecer subsídios aos pesquisadores indicando-lhes os materiais de maior qualidade, evitando assim que se perca muito tempo pesquisando documentos que não serão úteis.

Inúmeras métricas foram desenvolvidas com essa finalidade, sem que se possa dizer que uma delas seja definitiva ou exponencialmente mais importante do que outra. Todas têm o seu valor, dependendo do contexto em que são aplicadas. O avanço da tecnologia traz novas possibilidades de ferramentas, mas também novos desafios pela ampliação do horizonte de pesquisa. Seja priorizando os aspectos de visibilidade na *web*, tão em voga atualmente ou então observando as diretrizes da bibliometria, o desafio é conciliar as melhores características de cada métrica, adaptando-as às peculiaridades da Ciência da Computação, para poder apresentar o melhor resultado no menor espaço de tempo.

Neste trabalho preocupou-se com o desenvolvimento de um modelo. Este modelo tem o objetivo de contribuir para o processo de avaliação de conferências de uma maneira mais transparente. O resultado do *ranking* indicado pelo modelo é gerado a partir de dados obtidos na *web*.

A partir do modelo proposto, foi iniciada a coleta manual dos dados. Para validação dos indicadores externos, foi feita a correlação estatística dos resultados entre o valor *Ranking* da QUALIS e os indicadores escolhidos, verificando se há correlação ou não dos mesmos.

A principal contribuição deste trabalho foi apresentar um modelo de avaliação da qualidade científica, baseado em visibilidade na *web*, o qual servirá como um novo indicador para conferências de CC, podendo realizar uma análise do *website* da conferência em nível mundial e resultar num *ranking* de qualidade.

Outra contribuição foi apresenta um modelo baseado nas características *web*, que capturam aspectos importantes que não são considerados pelas métricas existentes (para conferências); apresentar e detalhar um conjunto de características que podem ser utilizadas como indicadores de qualidade para Conferências Científicas.

Nos experimentos, foi possível observar o novo *ranking* dessas conferências baseados na visibilidade na *web* como uma nova maneira para medir a qualidade das conferências, podendo comprovar que este método é adequado.

5.1 Trabalhos Futuros

Como trabalho futuro pretende-se por em prática este modelo, desenvolvendo um protótipo com interface *web*, para que seja possível a implementação de testes. Com este ambiente em produção, será possível validar os resultados. Com a possibilidade de uma consulta on-line e simplificada para avaliar a qualidade de conferências, espera-se que os pesquisadores se apoiem nesta ferramenta para a sua tomada de decisão quanto ao submeter ou não seus trabalhos em determinado evento científico, recompensando desta forma a realização deste trabalho.

REFERÊNCIAS

- AGARIN, O.; NWAGWU, W. E. Nigerian University Websites: A Webometric Analysis. *Webology*, v.5, n.4, 2008. Disponível em: <http://www.webology.ir/2008/v5n4/a62.html>. Acesso em: 05 mar. 2009.
- AGUILLO, I. F.; GRANADINO, B.; ORTEGA, J. L.; PRIETO, J. A. Scientific research activity and communication measured with cybermetrics indicators: Research Articles. *J. Am. Soc. Inf. Sci. Technol.*, New York, v.57, n.10, p. 1296-1302, 2006.
- AGUILLO, I. F.; ORTEGA, J. L. Mapping world-class universities on the web. *Inf. Process. Manage.*, Tarrytown, v.45, n.2, p. 272-279, 2009.
- AMIN, M.; MABE, M. Impact factors: use and abuse. *Perspectives in publishing*, Amsterdam, n.1, p. 1-6, out. 2000.
- BAR-ILAN, J. On the overlap, the precision and estimated recall of search engines, a case study of the query “Erdos”. *Scientometrics*, v.42, n.2, p. 207–228, 1998.
- BAR-ILAN, J.; PERITZ, B. Informetric theories and methods for exploring the internet: an analytical survey of recent research literature. *Library Trends*, v.50, n.3, p. 371-392, 2002.
- BAR-ILAN, J. 2005. Expectations versus reality – Web search engines at the beginning of 2005. In: Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics, 2005, p. 87-96.
- BIANCHINI, M.; GORI, M.; SCARSELLI, F. Inside PageRank. *ACM Trans. Internet Technol.*, New York, v.5, n.1, p. 92-128, feb. 2005.
- BJÖRNEBORN, L. **Small-world link structures across an academic Web space: a library and information science approach.** 2004, 399 p. Dissertation (PHD), Department of Informations Studies, Royal School of Library and Information Science, Copenhagen, Denmark, 2004.

BOELL, S. K.; WILSON, C. S.; COLE, F. T. H. A Webometric Analysis of Australian Universities using Staff and Size dependent Web Impact Factors (WIF). In: INTERNATIONAL CONFERENCE ON WEBOMETRICS, INFORMETRICS AND SCIENTOMETRICS & COLLNET MEETING, 4., 9., 2008, Berlin, Alemanha. **Proceedings...** [S.l.]: Gesellschaft für Wissenschaftsforschung, 2008.

BOKOS, G.; POULOS, M.; KORFIATIS, N. 2007. Towards the Construction of a Global Bibliometric Indicator. In: ISSI'07: Proceeding of the 11th biennial International Conference on Scientometrics and Informetrics, 2007, Madrid, Spain. Center for Scientific Information and Documentation (CINDOC), 2007, p. 920-921.

BOLLEN, J.; SMITH, J.; LUCE, R.; SOMPEL, V. de H. Toward alternative metrics of journal impact: a comparison of download and citation data. *Inf. Process. Manage.*, Tarrytown, v.41, n.6, p. 1419-1440, 2005.

BOUDOURIDES, M. A.; SIGRIST, B.; ALEVIZOS, P. D. Webometrics and the Self-Organization of the European Information Society. 1999. Disponível em: <http://hyperion.math.upatras.gr/webometrics/>, Acesso em: nov. 2009.

CAPES, Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Disponível em: http://qualis.capes.gov.br/arquivos/avaliacao/webqualis/criterios2004_2006/Criterios_Qualis_2005_02.pdf. Acesso em: out. 2008.

CHO, J.; ADAMS, R. E.; ROY, S. 2005. Page quality: in search of an unbiased web ranking. In: SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international Conference on Management of Data, 2005, New York, NY, USA. Anais...ACM, 2005. p. 551-562.

DANESH, F.; SOHEILI, F.; SHAFIEI, A. 2008. Hyperlink analysis of Iranian ministries websites. In: WIS'08: Proceedings of the Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, 2008, Berlin, Alemanha. Gesellschaft für Wissenschaftsforschung, 2008.

EGGHE, L. Theory and practice of the g-index. *Scientometrics*, v.69, n.1, p.131-152, abr. 2006.

FORD, C.; HARTE, S. Web-based analysis of E-journal impact: Approaches, problems, and issues. *Journal of the American Society for Information Science*, v.51, n.13, p. 1159-76, 2000.

GILES, C. L.; LAWRENCE, S. Accessibility of information on the web. *Nature*, v. 400, p. 107-109, 1999.

GRUPP, H.; HINZE, S.; SCHMOCH, U. International alignment and scientific regard as macro-indicators for international comparisons of publications. *Scientometrics*, v.51, p.359-380, 2001.

HAKAN G.; HUSSAIN, S. 2008. Ranking Journals, Conferences and Authors in Computer Graphics: A Fuzzy Reasoning. In IADIS '08: Proceedings International Conference on Informatics, 2008, Amsterdam, Holanda. p.75-80.

HARRIES; THELWALL. Do better scholars' web publications have significantly higher online impact?. *Journal of the American Society for Information Science and Technology*, New York, v.55, n.i2. 149-159, 2004.

HARZING, Research in International and Cross-cultural Management. Disponível em: <http://www.harzing.com/>, Acesso em: out. 2008.

HARZING, AW .K.; WAL, R. V. D. Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*. v.8, n.1, p. 61-73, jan. 2008.

HIRSCH, J.E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science*, USA, v.102, n.46, p. 16569-16572, nov. 2005.

HORROBIN, D. F. The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association*, v.263, p. 1438-1441, 1990.

JOWKAR, A.; BEGLOU, R. R. 2008. Study of the Relationship Between World University Rankings and Count of Links to Them: Based on Shanghai University Ranking and Times Higher Education. In: WIS'08: Proceedings of the Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, 2008, Berlin, Alemanha. Gesellschaft für Wissenschaftsforschung, 2008.

KATERATTANAKUL, P.; HAN, B.; HONG, S. Objective quality ranking of computing journals. *Communications of the ACM* . New York, v.46, n.10, p.111-114, out. 2003.

KESHTKAR, A.A.; AMINPOUR, F.; KABIRI, P.; OTROJ, Z. Webometric analysis of Iranian universities of medical sciences. *Scientometrics*, Budapest, v.80, n.1, p.253-64. Jul. 2009.

KOUSHA, K. A comparative study on Iranian newspapers websites via web impact factor. *Informology*, v.1, n.2, p.87-114, 2002.

KOUSHA, K. A comparison of Iranian newspaper websites with the application of web impact factor (WIF). *Ettela'Shenasi*, v.1, n.2, p. 87-114, 2003.

LISTER, R.; BOX, I. 2008. A citation analysis of the ACE2005--2007 proceedings, with reference to the June 2007 CORE conference and journal rankings. In: ACE'08: Proceedings of the Tenth Conference on Australasian Computing Education, 2008, Wollongong, NSW, Australia. Anais...ACM, 2008. P.93-102.

LIU, H.; HE, J.; DU, X.; ZHOU, X.; LI, Z. 2009. IRank: A Term-Based Innovation Ranking System for Conferences and Scholars. In: APWeb/WAIM '09: Proceedings of the Joint International Conferences on Advances in Data and Web Management, 2009, Suzhou, China. Anais...ACM, 2009. P.126-137.

MACIAS-CHAPULA, C. A. O papel da informática e da ciênciometria e sua perspectiva nacional e internacional. *Ciência da Informação*. Brasília, v.27, n.2, p.134-140, maio/ago. 1998.

MACROBERTS, B.R.; MACROBERTS, M.H. Problems of citation analysis. *Scientometrics*, v.36, n.3, p. 435-444, 1996.

MARTINS, W. S.; GONCALVES, M. A.; LAENDER, A. H. F.; PAPPAS, G. L. 2009. Learning to Assess the Quality of Scientific Conferences: A Case Study in Computer Science. In: JCDL'09: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, 2009, Austin, Texas, USA. Anais...ACM, 2009. P. 193-202.

MAT-HASSAN, M.; BAR-ILAN, J.; LEVENE, M. Dynamics of search engine rankings: a case study. In: Proceedings of the 3rd International Workshop on Web Dynamics, 2004, New York. Disponível em: http://www.dcs.bbk.ac.uk/webDyn3/webdyn3_proceedings.pdf .

MCGRANE, A.; SMAILES, J. *Estatística aplicada à administração com Excel*. São Paulo: Atlas, 2002.

MILLER, M. D.; RAINER, Jr. R. K. Examining differences across journal rankings. *Communications of the ACM*. New York, v.48, n.2, p. 91-94, 2005.

MOED, H. F. The impact factors debate: the ISI's uses and limits. *Nature*, v.415, p. 731-732, 2002.

MOXHAM, H.; ANDERSON, J. Peer review. A view from the inside. *Science and Technology Policy*, v.5, n.1, p. 7-15, 1992.

MYLONOPOULOS, N. A.; THEOHARAKIS, V. Global Perceptions of IS Journals: Where is the Best IS Research Published?. *Communications of the ACM*, New York, v.44, n.9, p. 29, set. 2001.

OSAREH, F.; MAKTABIFARD, L.; ZEINOLABEDINI, M. H. 2006. Collaboration analysis of World National Library websites via webometric methods. In: Proceedings on International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, 2006, Nancy, França. SRDI-INIST-CNRS, 2006. p. 215-224.

PLOTT, C.; TUNGER, D. Bibliometric analysis as part of a trend recognition system in science. *Iranian Journal of Information Science and Technology*. v.3, n.2, p. 1-17, jul./dez., 2005.

RINIA et al. Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy*, v.27, p.95-107, 1998.

ROUSSEAU, R. Situations: An exploratory study. **Cybermetrics**, v.1, n.1, 1997. Disponível em: <<http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>>. Acesso em: mar. 2010.

RWWU, Ranking Web of World Universities. Disponível em: <<http://www.webometrics.info>>. Acesso em: mar. 2009.

SCHUBERT, A.; BRAUN, T.; GLÄNZEL, W. A hirsh-type index for journals. *Scientometrics*, v.69, n.1, p. 169-173, 2006.

SHANKS, G.; FISHER, J.; LAMP, J. W. A Ranking List for Information Systems Journals. *Australasian Journal of Information Systems*, v.14, n.2, p. 5-18, 2007.

SHIMAKURA, S. E. Interpretação do coeficiente de correlação. Ago. 2006. Disponível em:
<http://leg.ufpr.br/~silvia/CE003/node74.html>, Acesso em: abril 2009.

SJTU, SHANGHAI JIAO TONG UNIVERSITY (2003), Academic Ranking of World Universities. Disponível em: <http://www.arwu.org/aboutARWU.jsp>, Acesso em: jan. 2010.

SMITH, A.G. A tale of two Web spaces: Comparing sites using web impact factors. *Journal of Documentation*, v.55, n.5, p. 577-592, 1999.

SOUTO, M. A. M.; WARPECHOWSKI, M.; PALAZZO J. M. O. 2007. An Ontological Approach for the Quality Assessment of Computer Science Conferences. In: QoIS '07 - International Workshop on Quality of Information Systems, 2007, Auckland. *Advances in Conceptual Modeling - Foundations and Applications*. Berlin: Springer-Verlag, 2007. v. 1. p. 202-212.

THELWALL, M.; VAUGAN, L. Scholarly use of the web: What are the key inducers of links to journal web sites?. *Journal of the American Society for Information Science and Technology*, New York, v.54, n.1, p. 29-38, 2003.

THELWALL, M.; VAUGHAN, L.; BJÖRNEBORN, L. (forthcoming). Webometrics. *Annual Review of Information Science and Technology*, 39, 2003.

Thomson Reuters Science. Disponível em: <http://science.thomsonreuters.com/>, Acesso em: 25 ago. 2009.

THOR, A.; RHAM, E. Citation analysis of database publications. *SIGMOD Record*, v.34, n. 4, 2005.

VAN RAAN, A. F. J. In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, v.43, p. 129-139, 1998.

VAN RAAN, A. F. J.; VAN LEEUWEN, TH. N. Identifying the Fields for Mapping RTD Excellence in the Life Sciences . A First Approach. *Report to the European Commission*, Brussels, 2001.

VAN RAAN, A .F. J. Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, Leiden, v.62, n.1, p. 133-143, 2005.

VANTI, Nadia. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão de conhecimento. *Ciência da Informação*, Brasília, v. 31, n.2, p. 152-162, maio/ago. 2002.

VANTI, Nadia. **Links Hipertextuais na Comunicação Científica:** análise webométrica dos sítios acadêmicos latino-americanos em Ciências Sociais 2007, 292 p. Tese (Doutorado em Comunicação e Informação), Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2007.

VSNU, Chemistry and Chemical Engineering. Assessment of Research Quality. Netherlands. Utrecht: VSNU, p.172, out. 2002. Disponível em: <http://media.leidenuniv.nl/legacy/assessment-chemistry-1996-2000.pdf>, Acesso em: fev. 2010.

VAUGHAN, L.; HYSEN, K. Relationship between links to journal Web sites and impact factors. *Aslib Proceedings*, v.54, n.6, p. 356-361, 2002.

ZHUANG, Z.; ELMACIOGLU, E.; LEE, D.; GILES, C. L. 2007. Measuring conference quality by mining program committee characteristics. In: JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, 2007, New York, NY, USA. Anais...ACM, 2007. p. 225-234.

WASSERMAN S.; FAUST K. *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press, 1994.

APÊNDICE A – AMOSTRA DE DADOS DAS CONFERÊNCIAS

Tabela A.1 – Conferências Científicas de Sistema de Informação

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações/Artigo</i>	<i>Autores/Artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
CAiSE	20	591	2015	3,41	2,26	631	5,40
CoopIS	21	133	1585	11,92	2,51	188	4,60
ICIS	11	218	603	2,77	2,46	396	4,00
SCC	8	413	540	1,31	2,96	540	5,20
ICEIS	11	922	1199	1,30	2,61	195	4,00
BPM	8	94	383	4,07	2,70	557	5,40
IRI	12	233	690	2,96	2,90	341	4,80
ECIS	10	95	369	3,88	2,27	602	5,25
ICSOC	6	144	169	1,17	2,54	127	4,20
AMCIS	19	432	2520	5,83	2,26	490	5,00
IBIMA	3	9	20	2,22	1,78	205	4,00
ICEIMT	2	6	16	2,67	3,00	17	5,00
IAIM	2	2	4	2,00	1,00	4	2,33
EMMSAD	6	21	88	4,19	2,14	373	5,33

Tabela A.2 – Conferências Científicas de Engenharia de Software

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações/Artigo</i>	<i>Autores/Artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
ICSE	50	1000	10648	10,65	2,52	1662	5,20
FSE	40	250	5138	20,55	2,14	597	5,40
ICSM	40	443	7422	16,75	2,64	621	5,00
ISRE	21	148	2602	17,58	2,17	361	5,00
SEFM	16	64	886	13,84	2,39	314	5,60

RE	35	407	4802	11,80	2,72	361	5,00
ICSR	20	100	1797	17,97	2,28	210	4,50
ASE	30	273	4120	15,09	2,41	203	5,00
ISSRE	21	141	2590	18,37	2,20	245	4,50
ESEC/FSE	40	250	5138	20,55	2,14	713	5,40
SEKE	14	414	1471	3,55	2,35	255	4,50
ESEM	10	71	329	4,63	3,13	13	3,50
CBSE	30	361	3677	10,19	2,35	66	3,67
METRICS	26	191	2419	12,66	2,49	56	3,50
AOSD	21	215	2039	9,48	2,57	144	4,75
FM	42	1000	10687	10,69	2,29	355	5,67
FME	38	424	6145	14,49	2,17	355	5,67
TACAS	65	663	18868	28,46	2,65	371	5,25
WCRE	36	384	5227	13,61	2,44	326	4,40
ISSTA	13	112	1288	11,50	1,88	286	5,00
CAV	106	1000	43378	43,38	2,55	469	5,75
ICCBSS	18	188	1192	6,34	2,51	10	0,00
UML	49	254	7623	30,01	2,40	283	4,50
FASE	38	423	6010	14,21	2,58	424	5,00
AMAST	34	452	4799	10,62	2,04	162	5,00
TAPSOFT	44	409	7652	18,71	1,83	1033	6,00
WICSA	21	170	1471	8,65	2,71	157	4,50
EASE	10	64	326	5,09	2,42	62	5,00
FOSSACS	36	235	6104	25,97	2,05	320	5,25
CSMR	30	333	3388	10,17	2,60	210	4,25
ICSSEA	6	64	154	2,41	2,50	53	3,00
ICSEA	7	201	436	2,17	2,71	138	5,00
SPLC	15	129	979	7,59	2,44	249	4,33
ECSA	4	24	40	1,67	2,08	104	4,50
QSIC	18	284	1231	4,33	2,95	189	4,75
XP	17	426	1255	2,95	2,38	417	3,75
FMOODS	28	217	2829	13,04	2,29	132	5,00
AGILE	7	20	113	5,65	2,30	233	4,25
SPICE	4	27	40	1,48	1,96	137	4,00

PROFES	17	296	1323	4,47	2,57	109	4,00
Ada-Europe	19	568	2115	3,72	2,30	100	3,50
ICFEM	24	404	2878	7,12	2,42	247	5,00
ISWM-Mensura	11	70	347	4,96	2,43	39	3,00
IFM	11	66	480	7,27	2,09	141	5,00
SQM	11	266	626	2,35	2,07	36	3,00
CSEE&T	13	213	845	3,97	2,32	70	4,00
IDEAS	5	24	111	4,63	2,83	98	4,00
ECMDA	21	159	1307	8,22	2,95	27	4,50
VLFM	3	6	93	15,50	2,33	266	5,33
AOSE	35	245	6311	25,76	2,74	88	3,50
EuroSPI	7	67	182	2,72	2,04	130	4,50
PROSim	6	47	128	2,72	2,32	84	4,50
CONQUEST	5	26	155	5,96	1,81	0	0,00
ESELAW	3	25	30	1,20	3,00	69	4,00
FATES/RV	18	92	1092	11,87	2,64	242	5,33
FMICS	7	77	156	2,03	2,73	106	4,80
SC	17	133	1026	7,71	2,57	315	5,00
EA	9	38	285	7,50	2,89	67	4,00
SELMAS	10	27	370	13,70	3,52	75	3,67
IWPSE	13	101	744	7,37	2,66	125	4,50
SAVCBS	11	51	396	7,76	2,51	34	4,20
SEW	15	192	1218	6,34	2,61	312	0,00

APÊNDICE B - CONFERÊNCIAS COM VALORES PADRONIZADOS

Tabela B.1 – Conferências de Banco de Dados com valores de acordo com o modelo

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações/Artigo</i>	<i>Autores/Artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
ICDE	20,00	5,00	5,00	5,00	5,00	50,00	10,00
VLDB	20,00	5,00	5,00	5,00	3,32	50,00	10,00
ICDM	20,00	5,00	5,00	5,00	5,00	50,00	6,66
SIGMOD	20,00	5,00	5,00	5,00	3,32	50,00	6,66
SDM	16,32	3,32	3,32	5,00	5,00	50,00	10,00
ISMIR	16,32	3,32	3,32	5,00	3,32	50,00	10,00
JCDL	16,32	3,32	3,32	1,66	5,00	50,00	10,00
CIKM	16,32	5,00	3,32	3,32	1,66	50,00	10,00
PKDD	16,32	3,32	3,32	3,32	3,32	50,00	10,00
DASFAA	16,32	3,32	3,32	3,32	3,32	50,00	6,66
PODS	16,32	5,00	3,32	1,66	1,66	50,00	6,66
SSDBM	20,00	5,00	5,00	5,00	5,00	33,20	10,00
KDD	20,00	5,00	5,00	3,32	3,32	33,20	10,00
ECIR	6,66	1,66	3,32	3,32	3,32	50,00	10,00
ECDL	6,66	3,32	1,66	3,32	3,32	50,00	10,00
SIGIR	6,66	3,32	3,32	1,66	1,66	50,00	10,00
EDBT	20,00	5,00	5,00	5,00	1,66	33,20	6,66
ICDT	20,00	3,32	3,32	5,00	1,66	33,20	10,00
WebDB	16,32	3,32	3,32	5,00	5,00	33,20	10,00
ER	16,32	5,00	3,32	5,00	3,32	33,20	10,00
ASIS&T	6,66	1,66	1,66	1,66	1,66	50,00	10,00

SPIRE	16,32	3,32	3,32	3,32	3,32	33,20	6,66
CIDR	20,00	1,66	5,00	5,00	5,00	16,66	10,00
TREC	20,00	5,00	5,00	5,00	5,00	16,66	6,66
RIDE	20,00	3,32	5,00	5,00	5,00	16,66	6,66
DEXA	20,00	5,00	5,00	3,32	3,32	16,66	6,66
ADMA	6,66	1,66	1,66	1,66	1,66	33,20	10,00
MIR	6,66	1,66	1,66	1,66	5,00	33,20	6,66
ADBIS	6,66	3,32	1,66	1,66	1,66	33,20	6,66
WIDM	6,66	1,66	1,66	3,32	1,66	33,20	6,66
DBPL	16,32	3,32	3,32	3,32	1,66	16,66	10,00
IDEAS	16,32	5,00	3,32	3,32	5,00	16,66	3,32
DaWaK	16,32	1,66	1,66	3,32	3,32	16,66	6,66
FQAS	6,66	1,66	1,66	1,66	3,32	16,66	10,00
DIWeb	6,66	1,66	1,66	1,66	3,32	16,66	10,00
NLDB	6,66	1,66	1,66	3,32	5,00	16,66	6,66
XSym	6,66	1,66	1,66	1,66	5,00	16,66	6,66
ODBASE	6,66	1,66	1,66	3,32	3,32	16,66	6,66
DOLAP	6,66	1,66	1,66	1,66	3,32	16,66	6,66
IKE	6,66	1,66	1,66	1,66	3,32	16,66	6,66
GIR	6,66	1,66	1,66	5,00	3,32	16,66	3,32
DCC	20,00	5,00	5,00	5,00	1,66	0,00	0,00
CIDM	6,66	3,32	1,66	1,66	5,00	16,66	0,00
TAKMA	6,66	1,66	1,66	1,66	5,00	0,00	0,00

Tabela B.2 – Conferências de Sistema de Informação com valores de acordo com o modelo

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações/Artigo</i>	<i>Autores/Artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
CAiSE	16,32	5,00	5,00	5,00	5,00	50,00	10,00
AMCIS	16,32	5,00	5,00	3,32	1,66	50,00	10,00
SCC	6,66	5,00	3,32	1,66	5,00	50,00	10,00
BPM	6,66	1,66	1,66	3,32	5,00	50,00	10,00
ECIS	6,66	1,66	1,66	3,32	1,66	50,00	10,00
ICEIS	16,32	5,00	3,32	1,66	5,00	33,20	6,66

CoopIS	16,32	3,32	3,32	5,00	3,32	33,20	6,66
IRI	16,32	3,32	3,32	1,66	5,00	33,20	6,66
ICIS	16,32	3,32	3,32	1,66	2,46	33,20	6,66
EMMSAD	6,66	1,66	1,66	3,32	1,66	33,20	10,00
IBIMA	6,66	1,66	1,66	1,66	1,66	33,20	6,66
ICEIMT	6,66	1,66	1,66	1,66	5,00	16,66	10,00
ICSOC	6,66	3,32	1,66	1,66	3,32	16,66	6,66
IAIM	6,66	1,66	1,66	1,66	1,66	16,66	3,33

Tabela B.3 – Conferências de Engenharia de Software com valores de acordo com o modelo

<i>Conferência</i>	<i>H-index</i>	<i>Artigos</i>	<i>Citações</i>	<i>Citações/Artigo</i>	<i>Autores/Artigo</i>	<i>Inlinks</i>	<i>PageRank</i>
TAPSOFT	20,00	5,00	5,00	5,00	1,66	50,00	10,00
CAV	20,00	5,00	5,00	5,00	3,32	50,00	10,00
FM	20,00	5,00	5,00	5,00	1,66	33,20	10,00
FME	20,00	5,00	5,00	5,00	1,66	33,20	10,00
SEFM	16,32	1,66	3,32	5,00	3,32	33,20	10,00
FSE	20,00	3,32	5,00	5,00	1,66	50,00	10,00
ESEC/FSE	20,00	3,32	5,00	5,00	1,66	50,00	10,00
FATES/RV	16,32	1,66	3,32	5,00	5,00	33,20	10,00
VLFM	6,66	1,66	1,66	5,00	3,32	33,20	10,00
TACAS	20,00	5,00	5,00	5,00	5,00	33,20	10,00
FOSSACS	20,00	3,32	5,00	5,00	1,66	33,20	10,00
ICSE	20,00	5,00	5,00	5,00	3,32	50,00	10,00
ICSM	20,00	5,00	5,00	5,00	5,00	50,00	10,00
FASE	20,00	5,00	5,00	5,00	3,32	50,00	10,00
RE	20,00	5,00	5,00	5,00	5,00	33,20	10,00
AMAST	20,00	5,00	5,00	5,00	1,66	33,20	10,00
ASE	20,00	3,32	5,00	5,00	3,32	33,20	10,00
ICFEM	20,00	5,00	5,00	3,32	3,32	33,20	10,00
ISRE	20,00	3,32	5,00	5,00	1,66	33,20	10,00
ISSTA	16,32	3,32	3,32	5,00	1,66	33,20	10,00
SC	16,32	3,32	3,32	3,32	3,32	33,20	10,00
FMOODS	20,00	3,32	5,00	5,00	1,66	16,66	10,00

IFM	16,32	1,66	1,66	3,32	1,66	16,66	10,00
ICSEA	6,66	3,32	1,66	1,66	5,00	16,66	10,00
EASE	6,66	1,66	1,66	3,32	3,32	16,66	10,00
FMICS	6,66	1,66	1,66	1,66	5,00	16,66	6,66
QSIC	16,32	3,32	3,32	3,32	5,00	33,20	6,66
AOSD	20,00	3,32	5,00	5,00	3,32	16,66	6,66
UML	20,00	3,32	5,00	5,00	3,32	33,20	6,66
WICSA	20,00	3,32	3,32	5,00	5,00	33,20	6,66
ISSRE	20,00	3,32	5,00	5,00	1,66	33,20	6,66
SEKE	16,32	5,00	3,32	3,32	3,32	33,20	6,66
ICSR	16,32	3,32	3,32	5,00	1,66	33,20	6,66
ECMDA	20,00	3,32	3,32	5,00	5,00	16,66	6,66
IWPSE	16,32	3,32	3,32	3,32	5,00	16,66	6,66
PROSim	6,66	1,66	1,66	1,66	3,32	16,66	6,66
ECSA	6,66	1,66	1,66	1,66	1,66	16,66	6,66
EuroSPI	6,66	1,66	1,66	1,66	1,66	16,66	6,66
WCRE	20,00	5,00	5,00	5,00	3,32	33,20	6,66
SPLC	16,32	3,32	3,32	3,32	3,32	33,20	6,66
CSMR	20,00	5,00	5,00	5,00	3,32	33,20	6,66
AGILE	6,66	1,66	1,66	3,32	1,66	33,20	6,66
SAVCBS	16,32	1,66	1,66	3,32	3,32	16,66	6,66
PROFES	16,32	3,32	3,32	3,32	3,32	16,66	6,66
CSEE&T	16,32	3,32	3,32	3,32	3,32	16,66	6,66
IDEAS	6,66	1,66	1,66	3,32	5,00	16,66	6,66
EA	6,66	1,66	1,66	3,32	5,00	16,66	6,66
ESELAW	6,66	1,66	1,66	1,66	5,00	16,66	6,66
SPICE	6,66	1,66	1,66	1,66	1,66	16,66	6,66
XP	16,32	5,00	3,32	1,66	3,32	50,00	6,66
CBSE	20,00	5,00	5,00	5,00	3,32	16,66	6,66
SELMAS	6,66	1,66	1,66	5,00	5,00	16,66	6,66
AOSE	20,00	3,32	5,00	5,00	5,00	16,66	6,66
METRICS	20,00	3,32	5,00	5,00	3,32	16,66	6,66
Ada-Europe	16,32	5,00	5,00	3,32	1,66	16,66	6,66
ESEM	6,66	1,66	1,66	3,32	5,00	16,66	6,66

ISWM-Mensura	16,32	1,66	1,66	3,32	3,32	16,66	6,66
SQM	16,32	3,32	3,32	1,66	1,66	16,66	6,66
ICSSEA	6,66	1,66	1,66	1,66	3,32	16,66	6,66
SEW	16,32	3,32	3,32	3,32	5,00	33,20	0,00
ICCBSS	16,32	3,32	3,32	3,32	3,32	16,66	0,00
CONQUEST	6,66	1,66	1,66	3,32	1,66	0,00	0,00

APÊNDICE C - AMOSTRA DE DADOS DAS CONFERÊNCIAS AVALIADAS

Tabela C.1 – Classificação de Conferências Científicas de Sistema de Informação

<i>Conferência</i>	<i>Valor ranking</i>	<i>Conceito</i>
CAiSE	0,963	A
AMCIS	0,913	A
SCC	0,816	A
BPM	0,783	A
ECIS	0,750	A
ICEIS	0,712	A
CoopIS	0,711	A
IRI	0,695	B
ICIS	0,669	B
EMMSAD	0,582	B
IBIMA	0,532	B
ICEIMT	0,433	C
ICSOC	0,399	C
IAIM	0,333	C

Tabela C.2 – Classificação de Conferências Científicas de Engenharia de Software

Conferência	Valor ranking	Conceito	Conferência	Valor ranking	Conceito
ICSM	1,000	A	FMOODS	0,616	B
CAV	0,983	A	CBSE	0,616	B

ICSE	0,983	A	AOSE	0,616	B
FASE	0,983	A	VLFM	0,615	B
TAPSOFT	0,967	A	AOSD	0,600	B
FSE	0,950	A	ECMDA	0,600	B
ESEC/FSE	0,950	A	METRICS	0,600	B
XP	0,863	A	AGILE	0,548	B
TACAS	0,832	A	Ada-Europe	0,546	B
RE	0,832	A	IWPSE	0,546	B
FM	0,799	A	PROFES	0,529	B
FME	0,799	A	CSEE&T	0,529	B
AMAST	0,799	A	IFM	0,513	B
ASE	0,798	A	SAVCBS	0,496	C
ICFEM	0,798	A	ISWM-Mensura	0,496	C
FOSSACS	0,782	A	SQM	0,496	C
ISRE	0,782	A	ICCBSS	0,463	C
WCRE	0,782	A	ICSEA	0,450	C
CSMR	0,782	A	SELMAS	0,433	C
UML	0,765	A	EASE	0,433	C
WICSA	0,765	A	IDEAS	0,416	C
ISSRE	0,748	A	EA	0,416	C
FATES/RV	0,745	A	ESEM	0,416	C
SEFM	0,728	A	FMICS	0,400	C
ISSTA	0,728	A	ESELAW	0,400	C
SC	0,728	A	PROSim	0,383	C
QSIC	0,711	A	ICSSEA	0,383	C
SEKE	0,711	A	ECSA	0,366	C
ICSR	0,695	B	EuroSPI	0,366	C
SPLC	0,695	B	SPICE	0,366	C
SEW	0,645	B	CONQUEST	0,150	C