



Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Modelo Automático de Qualidade para Sítios Web

José Américo Alves Sustelo Rio
(licenciado)

Dissertação apresentada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para a obtenção do Grau de Mestre em Engenharia Informática

Orientador científico: Professor Doutor Fernando Brito e Abreu

Julho de 2010

Nº do aluno: 20942

Nome: José Américo Alves Sustelo Rio

Título da dissertação:

Modelo Automático de Qualidade para Sítios Web

Palavras-Chave:

- Sítios web
- Indicadores de qualidade
- ISO9126
- Avaliação
- Recolha automática
- Domínio da aplicação

Keywords:

- Websites
- Quality indicators
- ISO9126
- Assessment
- Automatic collection
- Application domain

Agradecimentos

Em primeiro lugar gostaria de agradecer ao Professor Doutor Fernando Brito e Abreu, pela sua orientação e supervisão, e pelo seu encorajamento e amizade ao longo do decurso do mestrado, tanto da dissertação como da elaboração do programa de suporte. Foi incansável e insubstituível.

Gostaria de agradecer aos outros membros do QUASAR, especialmente à Raquel Porciúncula, Jorge Freitas e Miguel Goulão, pela troca de experiências e apoio.

Também quero agradecer ao meu colega José Reis, pelo incentivo e amizade ao longo do trabalho. Ao Ricardo Martins, da Itcode, pela paciência, crítica e capacidade e ajuda nas tarefas profissionais, quando não estava presente.

Por último, queria agradecer à minha mulher, Paula Nunes, pela sua paciência inesgotável, incentivo e troca de ideias acerca de Qualidade.

[Esta página foi intencionalmente deixada em branco]

Resumo

A avaliação da qualidade de presenças na web deve levar em consideração várias vertentes como o design, estrutura dos sítios e sua navegação, conteúdos, funcionalidades e desempenho. Embora se conheçam várias iniciativas de avaliação, estas são maioritariamente de cariz qualitativo, geralmente suportadas na opinião de peritos, o que torna as avaliações caras, morosas e, acima de tudo, subjectivas.

O objectivo desta dissertação é o de propor uma avaliação quantitativa e automatizada da qualidade de presenças na web. Para tal será consolidado um modelo de qualidade adequado e proposto um metamodelo representativo dos conceitos do domínio do problema. Este último deve ser instanciado para casos concretos, através da recolha automática com um *crawler* para captura da estrutura e analisadores léxico-sintácticos para análise de conteúdos. Serão usados indicadores quantitativos que permitam uma avaliação em valor absoluto (caracterizando os pontos fortes e fracos de um sítio) e relativo (permitindo a construção automática de classificações ordenadas de sítios). Será ainda efectuada análise estatística aos indicadores provenientes do modelo de qualidade para sítios web.

Espera-se que a utilização da aproximação proposta possa ser utilizada quer pelas equipas de desenvolvimento de sítios na web, por exemplo no estabelecimento e avaliação de conformidade com padrões mínimos, quer no apoio à decisão da subcontratação desse desenvolvimento, permitindo, por exemplo, a comparação de portfolios de sub-contratantes potenciais.

[Esta página foi intencionalmente deixada em branco]

Abstract

Assessing the quality of web presences should take into consideration various aspects such as design, structure and navigation of sites, content, functionality and performance. Although several initiatives of assessment are known, these are mostly qualitative in nature, generally supported by the view of experts, which makes assessments expensive, time consuming and, above all, subjective.

The aim of this thesis is to propose a quantitative and automated quality web presence assessment. We will consolidate a quality model and propose a metamodel representing the concepts of the problem domain. The latter must be instantiated for specific cases, through the collection with an automated crawler to capture the structure and lexical-syntactic analyzer for analysis of contents. Indicators will be used to enable a quantitative assessment in absolute value (characterizing the strengths and weaknesses of a site) and on (allowing the automatic construction of classifications ranked sites). It will also be carried out statistical analysis of the indicators from the quality model for websites.

It is hoped that using the proposed approach can be used either by the development teams of websites, for example in the establishment and assessment of compliance with minimum standards, both in supporting the decision of outsourcing this development, allowing, for example, comparing portfolios of potential sub-contractors.

[Esta página foi intencionalmente deixada em branco]

Índice

1	Introdução	2
1.1	Web Engineering	2
1.2	Motivação	3
1.3	Hipóteses e Problemas da investigação	5
1.3.1	Hipóteses:	5
1.3.2	Questões de investigação.....	5
1.3.3	Fronteiras da investigação:	6
1.4	Contribuições esperadas	6
1.5	Metodologia a seguir no decurso do trabalho	7
1.6	Organização da dissertação.....	8
2	Trabalho relacionado	10
2.1	WebQEM (Luis Olsina et al, 98-2008).....	10
2.1.1	Qualidade de aplicações web.....	11
2.1.2	O processo de avaliação WebQEM.....	11
2.1.3	Automatização do processo usando WebQEM_Tool.....	13
2.1.4	Implementação ou prova de conceito	13
2.1.5	Trabalho futuro do autor.....	14
2.1.6	Comentários a este trabalho.....	14
2.2	2QCV3Q (Mitch et al.)	16
2.2.1	O modelo 2QCV3Q.....	16

2.2.2	Implementação do modelo.....	18
2.2.3	Capacidade de automatização.....	19
2.2.4	Implementação ou prova de conceito	19
2.2.5	Comentários a este trabalho.....	19
2.3	MILE (Blas, et al. 2002I).....	21
2.3.1	Modelo MILE.....	21
2.3.2	Implementação do modelo.....	22
2.3.3	Capacidade de automatização.....	22
2.3.4	Implementação ou prova de conceito	22
2.3.5	Comentários a este trabalho.....	22
2.4	Minerva (Minerva Working Group, 2005).....	24
2.4.1	Modelo Minerva	24
2.4.2	Implementação do modelo.....	25
2.4.3	Capacidade de automatização.....	26
2.4.4	Implementação ou prova de conceito	26
2.4.5	Comentários a este trabalho.....	26
2.5	Enoteam (Polillo, 2005).....	28
2.5.1	Modelo Enoteam	28
2.5.2	Implementação do modelo.....	28
2.5.3	Capacidade de automatização.....	30
2.5.4	Implementação ou prova de conceito	30
2.5.5	Comentários a este trabalho.....	30
2.6	Modelo de qualidade de sítios web 5D (Signore, 2005).....	31
2.6.1	O modelo	31
2.6.2	Capacidade de automatização.....	33
2.6.3	Implementação ou prova de conceito	33
2.6.4	Comentários a este trabalho.....	33

2.7	Web Quality Model (Calero, 2005)	34
2.7.1	Web quality model	34
2.7.2	Implementação do modelo.....	34
2.7.3	Capacidade de automatização.....	35
2.7.4	Implementação ou prova de conceito	35
2.7.5	Comentários a este trabalho.....	35
2.8	Web Q-Model (Cimino e Micali, 2008)	36
2.8.1	Modelo de qualidade Web Q-Model	36
2.8.2	Implementação do modelo.....	37
2.8.3	Capacidade de automatização.....	37
2.8.4	Implementação ou prova de conceito	37
2.8.5	Comentários a este trabalho.....	38
2.9	Tabela comparativa	39
2.10	Análise dos resultados	39
3	Modelo de qualidade para presenças na web.....	42
3.1	Modelo de qualidade.....	43
3.2	Lista de características	45
3.3	Proposta de ontologia para avaliação.....	47
3.4	Descrição de indicadores quantitativos.....	49
3.5	Hipóteses a ensaiar.....	52
4	Procedimento Experimental	54
4.1	Amostra.....	54
4.2	Planeamento da experiencia.....	56
4.3	Recolha de dados	56
4.4	Descrição do ambiente de pesquisa e do Programa de recolha de dados	60
4.5	Avaliação dos dados, ou recolha de características indirectas.....	60
4.6	Estatísticas do processamento.....	62

5	Análise de dados	64
5.1	Análise de componentes principais.....	64
5.2	Análise da variância.....	66
5.3	Modelo de qualidade.....	73
5.3.1	Exemplo de limites para indicadores.....	75
5.4	Relação das características de qualidade com o PageRank do Google	77
5.4.1	Funcionamento do page rank.....	77
5.4.2	Recolha do PageRank	78
5.4.3	Análise estatística	79
5.4.4	Regressão Ordinal	79
5.4.5	Aplicação da regressão ordinal.....	81
5.4.6	Redução de variáveis	83
5.4.7	Regressão ordinal com conjunto reduzido de variáveis	85
5.4.8	Regressão ordinal dependente do domínio.....	87
5.4.9	Modelo de predição de probabilidade de ser um determinado PageRank	91
6	Conclusões e trabalho futuro	96
6.1	Conclusões	96
6.2	Trabalho Futuro	97
	Bibliografia.....	98
	Anexo 1 – Caracterização da amostra	101
	Anexo 2 Introdução à Engenharia de Software experimental	105
	Tipos de variáveis.....	105
	Aderência à distribuição	106
	Testes de Hipóteses	107
	Métodos de análise de dados – testes estatísticos.....	108
	Anexo 3 Folha de Excel usado para predição - fórmulas	113

Índice de Figuras

Figura 2-1 Avaliação dos trabalhos de Olsina et Al.....	15
Figura 2-2 Avaliação do trabalho de Mitch.....	20
Figura 2-3 Avaliação do trabalho de Blas	23
Figura 2-4 Avaliação do trabalho efectuado no projecto Minerva.....	27
Figura 2-5 Avaliação do trabalho de Polilo (Enoteam).....	30
Figura 2-6 Avaliação do trabalho de Signore	33
Figura 2-7 Avaliação do trabalho de Calero.....	35
Figura 2-8 Avaliação do trabalho de Cirino	38
Figura 3-1 Modelo de qualidade interna e externa da ISO 9126.....	44
Figura 3-2 Modelo para qualidade em uso da ISO9126.....	44
Figura 3-3 Modelo de qualidade proposto para sítios web.....	45
Figura 3-4 Ontologia para recolha de dados para modelo de qualidade proposto	48
Figura 4-1 Esquema da base de dados.....	58
Figura 4-2 Valores por omissão que escolhemos	59
Figura 5-1 Numero de variáveis x variabilidade explicada.....	65
Figura 5-2 Médias x erros de validação.....	68
Figura 5-3 Médias vs tamanho de página.....	68
Figura 5-4 Médias vs numero de ligações internos	69
Figura 5-5 Exemplo de indicador não dependente do domínio - maiúsculas.....	75
Figura 5-6 Indicador tamanho de página - dependente do domínio (menor é melhor)..	76

[Esta página foi intencionalmente deixada em branco]

Índice de Tabelas

Tabela 2-1 – Dimensões do modelo 2QCV3Q.....	16
Tabela 2-2 – Atributos do modelo 2QCV3Q.....	18
Tabela 2-3 - Cruzamento entre princípios de qualidade e fases de implementação.....	26
Tabela 2-4 Características e sub-carecterísticas do modelo de qualidade Enoteam.....	29
Tabela 2-5 tabela comparativa dos trabalhos	39
Tabela 3-1 Indicadores de qualidade passíveis de recolha automática	46
Tabela 3-2 Tabela de definição dos indicadores	50
Tabela 3-3 Definição do indicador <code>efficiency_css_size</code>	50
Tabela 3-4 Definição do indicador <code>links_num_extern_links</code>	51
Tabela 4-1 Áreas e Países da amostra	54
Tabela 4-2 lista de sítios da amostra.....	55
Tabela 4-3 Estatísticas da base de dados	62
Tabela 4-4 estatísticas das funções que extraem os indicadores	62
Tabela 5-1 Variância total explicada.....	65
Tabela 5-2 Resultado da análise Anova para as 7 variáveis normais.....	67
Tabela 5-3 Teste Kruskal-Wallis.....	70
Tabela 5-4 Métricas de funcionalidade vs Domínio da Aplicação.....	71
Tabela 5-5 Métricas de fiabilidade vs Domínio da Aplicação	71
Tabela 5-6 Métricas de Usabilidade vs Domínio da Aplicação	72
Tabela 5-7 Métricas de Eficiência vs Domínio da Aplicação	72
Tabela 5-8 Métricas de Manutenção vs Domínio da Aplicação.....	72
Tabela 5-9 Métricas de Portabilidade vs Domínio da Aplicação	72
Tabela 5-10 Parte do Modelo de qualidade não dependente do domínio.....	73
Tabela 5-11 Parte do Modelo de qualidade dependente do domínio	74
Tabela 5-12 Page rank da amostra.....	79

Tabela 5-13 Funções de ligação para regressão ordinal	80
Tabela 5-14 valores de correlação para a regressão ordinal, com a função de link= logit	81
Tabela 5-15 valores de correlação para a regressão ordinal, com a função de ligação= complementary log-log.....	81
Tabela 5-16 Estimação de parâmetros para modelo de regressão nominal com todas as variáveis.....	82
Tabela 5-17 Análise de Co-variância – Spearman	83
Tabela 5-18 Lista de variáveis candidatas para modelo reduzido	85
Tabela 5-19 Coeficientes para o conjunto reduzido de variáveis, com a função de ligação= complementary log-log	85
Tabela 5-20 valores de correlação para a regressão ordinal com número de variáveis reduzido, com a função de ligação= complementary log-log.....	86
Tabela 5-21 Fit para modelo com conjunto de variáveis reduzido.....	86
Tabela 5-22 Crosstab do rank x rank previsto para o modelo com o conjunto reduzido de variáveis	86
Tabela 5-23 Parâmetros do modelo para domínio Jornais	87
Tabela 5-24 Fit do modelo para domínio Jornais	87
Tabela 5-25 Crosstab do rank x rank previsto para o modelo específico do domínio Jornais.....	88
Tabela 5-26 Parâmetros do modelo para domínio Bancos	88
Tabela 5-27 Fit do modelo para domínio Bancos	89
Tabela 5-28 Crosstab do rank x rank previsto para o modelo específico do domínio Bancos	89
Tabela 5-29 Parâmetros do modelo para domínio Aviação	90
Tabela 5-30 Fit do modelo para domínio Aviação	90
Tabela 5-31 Excel para predição do page rank, modelo reduzido independente do domínio de aplicação, exemplo para usatoday.com.....	92
Tabela 5-32 Excel para predição do page rank, modelo reduzido dependente do domínio de aplicação, exemplo para latimes.com.	93
Tabela A-0-1 Lista de jornais	102
Tabela A-0-2 Lista de Bancos	103
Tabela A-0-3 lista de sites de Companhias de Aviação	104

Capítulo 1

Introdução

1	Introdução.....	2
1.1	Web Engineering	2
1.2	Motivação	3
1.3	Hipóteses e Problemas da investigação	5
1.3.1	Hipóteses:	5
1.3.2	Questões secundárias	5
1.3.3	Fronteiras da investigação:	6
1.4	Contribuições esperadas	6
1.5	Metodologia a seguir no decurso do trabalho	7
1.6	Organização da dissertação.....	8

1 Introdução

1.1 Web Engineering

A *Web Engineering*, ou Engenharia Web, é uma disciplina emergente, que não cabe no âmbito da Engenharia de Software devido principalmente à natureza multidisciplinar da web e outros factores tais como as características das aplicações, processo diferente de desenvolvimento (por exemplo, tempos de desenvolvimento mais curtos), utilizadores diversos, diferente evolução e ciclo de vida, disponibilidade (sempre *online*), estrutura subjacente (arquitectura e rede), aspectos legais, técnicos e éticos, etc. A Engenharia Web engloba o design, desenvolvimento, evolução e avaliação de qualidade das aplicações web [1-3].

Podemos dizer que a Engenharia Web é uma disciplina multidisciplinar que contém parte das disciplinas da Engenharia de Software. Contudo, nenhuma delas contém a outra totalmente, nem a sua intersecção dá como resultado um conjunto vazio. Assim, além das disciplinas da Engenharia de Software, a Engenharia Web contém outras disciplinas devido à sua natureza multidisciplinar, como por exemplo o design (de *layouts* para as páginas do sítio) [4]. Um dos pontos que diferem também é o facto da Engenharia de Web dever ser encarada mais como um serviço do que um produto, devido à sua constante evolução como software, design e conteúdo [5]. Embora parte do conhecimento e ferramentas usadas na Engenharia de Software possam ser usadas na Engenharia da Web, está a ser estudada a possibilidade de um WEBBOK, para agregar o conhecimento, da mesma forma que o SWEBOK agrega o conhecimento da Engenharia de Software [6].

O desenvolvimento de aplicações web tem sido feito de um modo directo, onde a ausência do uso de ferramentas de Engenharia direccionadas para a web se traduz nas seguintes percentagens [7]:

- 84% dos projectos não atinge as necessidades de negócio
- 53% dos projectos não providencia a funcionalidade requerida
- 79% dos projectos tem atrasos
- 63% dos projectos excede o orçamento

Assim, à medida que a web aumenta de complexidade, a necessidade de ferramentas, metodologias e de melhores práticas direccionadas para a web também aumenta, sendo necessário usar metodologias, técnicas, avaliação de qualidade, processos de qualidade e melhores ferramentas aplicadas à Engenharia Web [8].

Podemos definir Engenharia Web como “o uso de princípios científicos, de engenharia e de gestão e aproximações sistemáticas com o intuito de desenvolver, implantar e manter sistemas e aplicações de alta qualidade baseadas na web” [9].

1.2 Motivação

Em 20 anos, a web evoluiu bastante, até ao ponto de as grandes companhias de software estarem a migrar o seu software tradicional para a web. Algumas delas já lá estão, como o Google Apps[10] ou o Microsoft Office, na versão 2010. Outras companhias com a Adobe desejam estar em força na web brevemente (ou na nuvem, como se pode chamar a uma aplicação que está completamente online).

Esta tendência em crescimento motiva a necessidade de modelos de qualidade mais adequados para suportar a avaliação de sítios web em geral e aplicações web em particular.

Podemos dizer que a avaliação da qualidade web é importante segundo vários pontos de vista:

- É importante para a empresa de software que desenvolve, para ajudar a definir e implementar os processos de qualidade;

- É importante para as empresas que contratam as que desenvolvem, para ajudar a escolher entre empresas fornecedoras de Web Engineering /Web Design, baseado nos seus portfolios.
- Uma organização terceira (um observatório independente, empresa de consultoria, organismos de certificação, ou regulador de mercado governamental) pode também produzir classificações e identificar boas práticas, assim como assinalar implementações pobres [11].
- Pode ser ainda importante para os utilizadores em geral das mesmas aplicações.

Quando se tenta exprimir a qualidade de sítios web, existem certos aspectos a considerar: a qualidade real ou percebida é ligada fortemente ao *web design*, e este é, em grande medida, subjectivo e muito difícil de medir, sem o auxílio de peritos. Além disso, a qualidade das partes de conteúdo (informação textual, imagens , animações ou filmes) é também difícil de medir sem intervenção humana.

Assim, quando pensamos em medir a qualidade web, encontramos-nos numa encruzilhada. Podemos construir um modelo mais completo e compreensivo, que tenha uma maior cobertura dos aspectos de qualidade Web, mas que requeira uma intervenção de especialistas humanos para colheita de indicadores e métricas, resultando numa avaliação mais subjectiva e cara. Além disso, sítios web que suportam tarefas comerciais, evoluem rapidamente [12], por isso a opção de termos a avaliação de peritos pode nem ser exequível. Considere-se, por exemplo, a mudança diária de um jornal na web, não só dos conteúdos, mas do tipo de conteúdos.

Por outro lado, podemos construir um modelo mais pequeno, onde tentamos que as variáveis sejam independentes, e ainda que se acabe com a referida subjectividade introduzida pelos peritos. Para que um modelo seja objectivo, em vez de subjectivo, a solução que nos parece mais indicada é que os indicadores ou métricas sejam recolhidos automaticamente. Outra solução seria termos um número grande de peritos, que fariam a subjectividade tender para zero, por cancelamento.

Tomando estes factores em conta, a avaliação baseada em peritos dá-nos uma aproximação mais completa, mas nem tem uma boa relação custo/benefício, nem nos fornece a base para uma avaliação justa, objectiva, e que possa ser feita de modo regular, como é necessário num panorama actual de sítios e aplicações web em constante mutação. Assim, decidimos explorar uma aproximação completamente

automática para a avaliação de qualidade de sítios web. Terá uma cobertura menor (nem todos os indicadores são passíveis de automação), mas terá uma relação custo/benefício boa, dará resultados rápidos e será objectiva.

Devido à nossa experiência neste campo, não acreditamos na possibilidade de estabelecer um modelo universal. Diferentes domínios de aplicação poderão requerer modelos adaptados. Contudo, não é claro quais são as fronteiras dessa adaptação. Precisamos também de provas estatísticas sólidas que mostrem quais os indicadores identificados que são dependentes do domínio da aplicação e quais os que não são.

1.3 Hipóteses e Problemas da investigação

Seguindo o método científico, vamos definir as hipóteses a investigar, assim como alguns problemas que surgem com estas.

1.3.1 Hipóteses:

As hipóteses que se pretendem resolver no decurso desta investigação são as seguintes:

- É possível definir um modelo automático de presenças na web, contendo apenas indicadores que se possam recolher automaticamente?
- É possível elaborar um programa que recolha dados e indicadores de uma amostra automaticamente, e avaliar esta amostra segundo o modelo de qualidade para sítios web?

1.3.2 Questões de investigação

Com as hipóteses, surgem algumas questões secundárias, de operacionalidade, que esperamos também responder no decurso desta dissertação:

- O que é a qualidade web?
- Como se classifica um sítio com respeito à qualidade?
- Como definir um modelo de qualidade web com características/indicadores adequados para recolha automática?
- Que pesos devem possuir essas características/indicadores?
- Como se validam os resultados?

1.3.3 Fronteiras da investigação:

O que cai dentro da investigação em curso é a definição de um modelo ontológico para a qualidade web, com grupos de características (ou somente características) e atributos (ou somente indicadores). É também necessário estudar a operação de *web crawlers* e recolha automática de dados, neste caso páginas web, e como mapear as características ou extrair indicadores automaticamente.

O que sai fora do âmbito de investigação é a definição de um modelo de qualidade envolvendo a avaliação não automática, tipicamente executado por peritos. Sai também fora do âmbito da investigação a avaliação do código do servidor, visto que não temos acesso a este.

1.4 Contribuições esperadas

As contribuições esperadas para o avanço do conhecimento nesta área serão a respostas às questões de investigação atrás identificadas.

No decurso da investigação, esperamos ainda contribuir com o modelo de qualidade automático e a definição de como se recolhem as suas características automaticamente. Esperamos ainda contribuir com código usado na investigação, que irá ser distribuído para outros investigadores na área, no site do grupo QUASAR.

1.5 Metodologia a seguir no decurso do trabalho

Esta investigação tem um duplo objectivo: fazer um apanhado/resumo de ontologias de características/indicadores de qualidade para sítios web e destas, tentar ver quais as passíveis de análise quantitativa, ou seja, de quantificação automática, e fazer a respectiva prova de conceito, a implementação, e sobre os dados resultantes faremos investigação com análise estatística para provar ou refutar as hipóteses.

A metodologia usada será a seguinte:

Rever os trabalhos feitos na última década sobre a matéria; Classificar estes mesmos trabalhos segundo critérios: Modelo de qualidade, Ontologia, Lista de características, Indicadores, Âmbito (geral ou só para uma área), Passível de automatização, Se testa casos Reais e Novidades propostas.

Destes trabalhos fazer uma lista de características e sub-características (ou indicadores); Renomear e eliminar duplicados das mesmas características e indicadores; Extrair os indicadores passíveis de automatização. Adicionar indicadores que não estão contemplados. Agrupar estes indicadores/características de forma a fazer um **modelo automático de qualidade da web**.

Em seguida temos o procedimento experimental, que é o seguinte: 1) Construção de um modelo em base de dados que vai servir de suporte ao nosso modelo de qualidade; 2) Construção ou aproveitamento de um *crawler* que vai guardar os dados (componentes) obtidos em base de dados

3) Fase de Recolha: Recolha automática de sítios a partir de uma amostra que consiste em 3 domínios de sítios, com mais de 30 sítios em cada domínio. Extrair na recolha alguns indicadores directos que só se podem extrair na fase de recolha (como tempo de download).

4) Fase de análise de indicadores: extracção dos restantes indicadores da amostra guardada anteriormente.

5) Análise de Resultados, ou Análise estatística – feita com o SPSS, e tenta responder a: Tentativa de reduzir o número de indicadores; Análise de quais são independentes e dependentes do domínio da aplicação; Tentativa de construção de

modelo de qualidade; Verificação de modelo de qualidade com base no conjunto dos indicadores de qualidade de sítios web

1.6 Organização da dissertação

Esta dissertação está organizada como se segue. O capítulo 1, contém a introdução da dissertação, incluindo hipóteses, problemas, metodologia e motivação.

O capítulo 2 é a revisão do trabalho relacionado, como ponto de partida para o nosso modelo. Para cada artigo, contém comentários à possibilidade de automatização, bem como às contribuições e implementações. Contém uma tabela comparativa e ainda alguns contributos para o nosso trabalho.

O capítulo 3 é onde esboçamos o modelo de qualidade e a lista de características a recolher, descrevemos alguns indicadores quantitativos e definimos as hipóteses a ensaiar na parte experimental.

No capítulo 4 mostramos como fizemos o nosso procedimento experimental. Começamos por descrever a amostra, depois o planeamento da experiência, a recolha de dados, a descrição do ambiente, e por fim, a fase de análise de dados.

O capítulo 5 é o capítulo da análise dos dados. Neste capítulo apresentamos a análise estatística realizada com o auxílio do SPSS, para obter respostas às nossas hipóteses de investigação.

O último capítulo contém as conclusões deste trabalho e identifica algumas vias para o trabalho futuro.

Capítulo 2

Trabalho Relacionado

2	Trabalho relacionado	10
2.1	WebQEM	10
2.2	2QCV3Q-Mitch et al	16
2.3	MILE	21
2.4	Minerva	24
2.5	Polillo	28
2.6	SIGNORE	31
2.7	Calero	34
2.8	Web Q-Model	36
2.9	Tabela comparativa	39
2.10	Análise dos resultados	39

Este capítulo é a revisão do trabalho relacionado, como ponto de partida para o nosso modelo. Para cada artigo, contém comentários à possibilidade de automatização, bem como às contribuições e implementações. Contém uma tabela comparativa e ainda alguns contributos para o nosso trabalho.

2 Trabalho relacionado

2.1 WebQEM (Luis Olsina et al, 98-2008)

Luis Olsina começou a investigação nesta área nos finais de década de 90 [13-16]. Baseado nas definições de qualidade da engenharia do software da ISO 9126, definiu o seu modelo e ontologia de características, que foi sucessivamente refinado e testado para vários tipos de sítios (domínios), até que atingiu a forma mais madura no capítulo de livro *Web quality*.

Segundo este investigador, os requisitos de qualidade para uma aplicação Web podem variar dependendo do tipo de entidade, dos pontos de vista do utilizador ou *developer*, ou ainda do contexto de utilização. Assim, o modelo de qualidade e logo a importância relativa das características, deve ter em conta 3 pontos de vista, do visitante, do *developer* e do gestor. O utilizador pode ainda ser dividido em utilizadores gerais e especialistas.

Na norma ISO 9126 o modelo de qualidade é dividido em: Funcionalidade, Fiabilidade, Usabilidade, Eficiência, Manutenção e Portabilidade. Na revisão mais recente da norma ISO 9126-1 as características passaram a ser internas (divididas de forma semelhante), externas (divididas de forma semelhante), e de qualidade em uso (semelhantes à norma 9241-11): Eficácia, Produtividade, Segurança e Satisfação.

Neste conjunto de trabalhos de Olsina et al. o foco é nas características externas e de qualidade em uso, devido à dificuldade de medição das características internas.

2.1.1 Qualidade de aplicações web

O autor refere Powell [17] dizendo que as aplicações *web* “envolvem uma mistura entre design gráfico e desenvolvimento de software, entre marketing e computação, entre comunicações internas e relações externas, e entre arte e tecnologia”.

Com efeito, englobam programação, arquitectura de informação, gestão de conteúdos, navegação, apresentação e estética (ou design), audiências múltiplas de utilizadores, aspectos legais e éticos, segurança e performance da rede e até sistemas e *browsers* heterogéneos. A Qualidade de aplicações Web tem os seus próprios atributos, distintos da Engenharia de Software, devido à sua natureza multidimensional e multidisciplinar, por exemplo: i) as aplicações web vão continuar a ser de conteúdos e orientadas para os documentos; ii) as aplicações web são interactivas, centradas no utilizador, baseadas em hipermédia, onde a interface tem um papel principal; iii) a web requer uma maior ligação entre arte e ciência que a encontrada em aplicações de software; iv) tem que se ter em conta aspectos como internacionalização e acessibilidade para pessoas com deficiência física; v) devem conter premissas básicas como pesquisa e navegação; vi) segurança é um ponto central em aplicações transaccionais; vii) estão sempre a evoluir; viii) o meio onde correm é menos previsível; ix) contem aspectos como privacidade de conteúdos e direitos legais dos elementos usados na sua concepção.

Contudo, apesar de serem um artefacto distinto das aplicações de software tradicionais que correm directamente sobre o sistema operativo, também envolvem código e executáveis, persistência, requisitos, arquitectura, design e especificações de teste.

Assim, o autor considera que o modelo da norma também é aplicável em grande parte às aplicações Web. Mas considera também que devem ser feitas algumas modificações e adicionadas algumas características, como por exemplo, o conteúdo.

2.1.2 O processo de avaliação WebQEM

O *Web Quality Evaluation Method* (WebQEM) é um método centrado num modelo de categorias de inspecção. É um modelo quantitativo por oposição a qualitativo, ou seja, objectivo em vez de subjectivo.

As fases usadas inicialmente na construção do modelo eram os seguintes: 1- Seleccionar um sítio ou conjunto de sítios competitivos para avaliar ou comparar; 2- Especificar objectivos e ponto de vista do utilizador; 3- Definir as características de

qualidade do sítio web e atributos; 4-Definir critérios aos atributos e aplicar medidas de atributos; 5-Agregar preferências elementares de modo a atingir as preferências globais de qualidade do sítio web, 6-Analisar, avaliar e comparar resultados parciais e globais.

Estas fases foram refinados mais tarde em: 1-Definição e especificação de requisitos de qualidade; 2-Medição e avaliação elementar (etapas: design e implementação), 3-Avaliação global (etapas: design e implementação); 4-Conclusão e recomendações.

Estas fases são definidas como:

1-Definição e especificação de requisitos de qualidade - Nesta fase, avaliadores clarificam os objectivos da avaliação e o ponto de vista do utilizador. Seleccionam um modelo de qualidade, por exemplo, as características do modelo ISO, adicionando os atributos modificados para o domínio web. Depois identificam a importância relativa desses componentes para a audiência web esperada (visitante, *developer* e gestor) e a cobertura requerida.

2-Medição e avaliação elementar (etapas: design e implementação) - Nesta fase, definem-se duas grandes etapas: desenhar a avaliação elementar e sua execução ou implementação. A primeira divide-se ainda em dois processos: a)definição de métricas e b)definição de indicadores elementares. Cada atributo pode ser quantificado por uma ou mais métricas. Para o processo de definição de métricas pode-se seleccionar apenas uma métrica para cada atributo da árvore de requisitos de qualidade, dado um projecto de medição específico. A métrica contém a definição da medição seleccionada do método de cálculo e escala. As métricas podem ser directas ou mapeadas numa função ou equação. O valor dos indicadores pode ter uma escala contínua ou discreta (por exemplo, não satisfaz, marginal e satisfaz).

3-Avaliação global (etapas: design e implementação) - A fase da avaliação global tem duas etapas, concepção e execução (design e implementação). Na etapa da concepção são seleccionados os critérios de agregação e um modelo de pontuação. Existem dois tipos de modelos: aditivos lineares e baseados em pontuações multi-critério não lineares, mas os dois usam pesos para considerar a importância relativa dos indicadores. Assim, depois da escolha do modelo, podemos calcular os indicadores parciais e globais na etapa de execução. Podemos dizer que os indicadores qualidade global e “qualidade em uso” representam o grau de satisfação de conformidade com os requisitos, do ponto vista do utilizador.

4-Conclusão e recomendações - A conclusão do processo de avaliação inclui a documentação dos produtos web, a especificação dos requisitos, métricas e indicadores da qualidade, modelos elementares e globais e critérios de decisão bem como a recolha de medidas e valores dos indicadores elementares, parciais e globais. Pode-se então analisar e compreender os pontos fortes e fracos do produto avaliado relativamente às necessidades de informação e sugerir recomendações justificadas.

2.1.3 Automatização do processo usando WebQEM_Tool

Foi desenvolvida uma ferramenta baseada em web (webQEM_Tool) para servir de suporte à administração dos projectos de avaliação, que permite editar os requisitos não funcionais e calcular indicadores baseados em dois modelos de agregação. Em seguida, editando indicadores elementares automaticamente ou manualmente, a ferramenta agrega os elementos e calcula o indicador de qualidade global para cada aplicação.

A ferramenta mostra os resultados de avaliação usando páginas web com informação em texto, tabelas e gráficos, gerando páginas dinamicamente obtidas de tabelas da camada de dados (persistência).

2.1.4 Implementação ou prova de conceito

Foram estudados três grupos de sítios: museus, sítios académicos e lojas. Para os dois primeiros, a árvore de requisitos da qualidade é como a seguinte, com algumas variações: Usabilidade, Funcionalidade, Fiabilidade e Eficiência.

A usabilidade divide-se em compreensão global do sítio (indicadores como presença de mapa do sítio, índice global, tabela de conteúdos, qualidade dos *labels*, *Tours* -só nos museus), feedback e *help* (*help*, último *update*, contactos, FAQ, *Survey Questionary*), e da interface (coesão, permanência da apresentação e estabilidade dos controlos principais, preferência estética e uniformidade de estilo), e miscelâneas como suporte a língua estrangeira e *downloads*.

A funcionalidade engloba indicadores como pesquisa, navegação e funções específicas de domínio como relevância, carrinho de compras e qualidade do catálogo de produtos.

A fiabilidade do sítio engloba indicadores de não deficiência, como erros de *ligações*, e erros devido aos *browsers* diferentes.

A eficiência engloba indicadores como a acessibilidade de informação (suporte para versão de texto, legibilidade sem imagens) e desempenho (tamanho da página). Para as lojas (ou *Shopping Cart*) a árvore é a seguinte: Usabilidade, Funcionalidade, Conteúdo e Fiabilidade.

Na usabilidade temos métricas que estão relacionadas com a compreensão, operação e atracção visual do sítio.

Na funcionalidade temos métricas relacionadas com a função específica da loja e funcionamento do *shopping cart*.

No conteúdo temos indicadores de informação, acessibilidade do conteúdo (conforme descrito atrás na eficiência).

Na fiabilidade temos indicadores de não deficiência ou maturidade (*ligações*) e outras deficiências (dependentes de *browsers*).

No modelo de agregação foi usado o LSP (*Logic Scoring of Preference*) – usa operadores baseados na potenciação com pesos - em vez da adição linear.

Foram analisadas as lojas Amazon e Cuspide, tendo em conta um ponto de vista de visitante geral. O resultado é uma tabela sumário das características calculadas e respectivos resultados. Aqui podemos observar algumas deficiências de imediato.

Numa análise final, podemos englobar as melhores características de cada aplicação no desenvolvimento de uma futura aplicação no mesmo domínio (loja on-line).

2.1.5 Trabalho futuro do autor

O grupo de que o autor faz parte está a implementar uma Framework de avaliação e medida mais robusta chamada INCAMI (*Information Need, Concept model, Attribute, Metric, and Indicator*) que é fundamentada na especificação ontológica de métricas e indicadores.

2.1.6 Comentários a este trabalho

Nos artigos iniciais o autor avança o conceito de *web engineering* e diferenças que existem entre a engenharia de software tradicional, fundamentando-se na natureza das aplicações web que é uma mistura de conteúdos de informação, funcionalidades e serviços.

Fundamentando-se na ISO 9126-1, define qualidade como qualidade interna, externa e qualidade em uso e o modelo de qualidade com os seis grupos de

características definidos na norma. Adiciona que qualidade em uso é a medida da excelência da qualidade de um produto. Propõe também que seja adicionado o conteúdo como uma característica extra.

Define um método de avaliação de qualidade baseado num modelo ou conjunto de características que agregadas dão um indicador global da qualidade. Este conjunto de características pode mudar conforme o domínio. As características têm pesos associados e a agregação pode ser adição ou uma função ou operador. Em vez de utilizadores normais propões a utilização de especialistas para minimizar o tempo. Este processo é passível de automatização como demonstrado pela ferramenta WebQEM_Tool.

Os pontos mais fracos deste método são a complexidade na agregação de indicadores, e as características são divididas em 7 grupos mas só são usados 4. Este modelo permite a mudança da lista de características e até grupos de características e atributos para cada domínio, ou seja, o modelo é um meta-modelo que pode ser instanciado, que pode ser tanto uma fraqueza como um ponto forte. No caso de o domínio já ter sido estudado, considera-se um ponto forte, no caso de um domínio novo o modelo não pode ser aplicado.

O gráfico seguinte representa a nossa avaliação deste(s) trabalhos(s), onde as contribuições estão numa escala de 0 a 4.

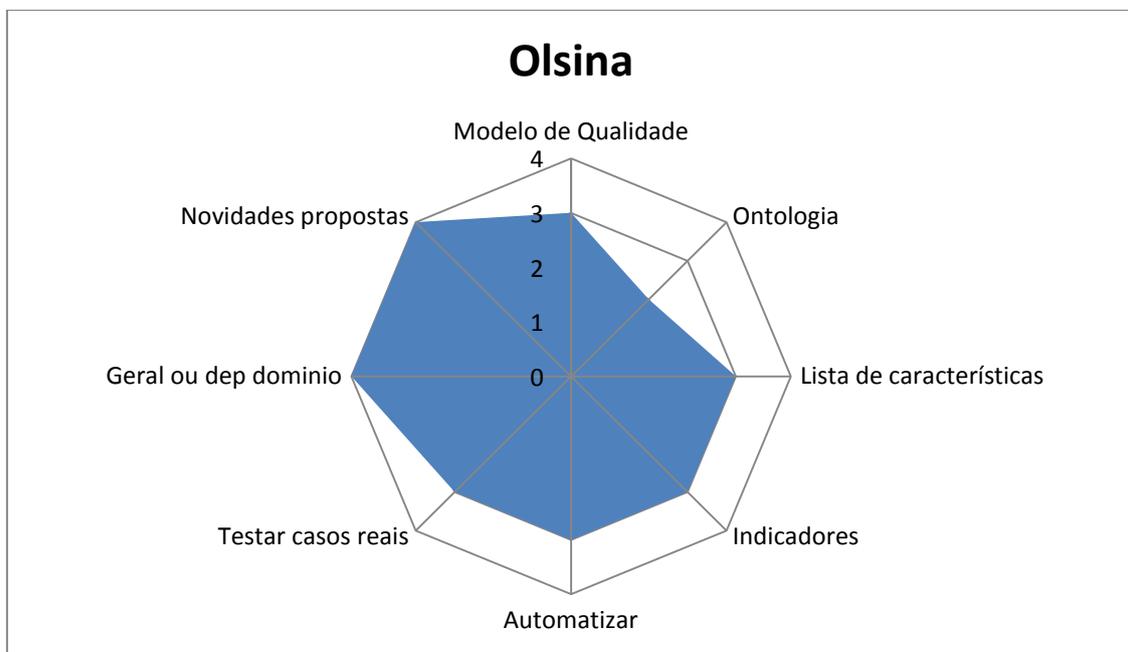


Figura 2-1 Avaliação dos trabalhos de Olsina et Al

2.2 2QCV3Q (Mitch et al.)

O trabalho de Mitch et al. também usa a definição de qualidade da ISO que define qualidade como a “totalidade das características de uma entidade que é suportada pela sua capacidade de satisfazer requisitos implicados ou pedidos” [18]. Dois requisitos para avaliar sítios web saem desta definição: 1) avaliação geral das características dos sítios web; 2) avaliação de quão bem o sítio satisfaz as necessidades específicas.

2.2.1 O modelo 2QCV3Q

Esse modelo serve para os donos dos sítios e *developers* avaliarem a qualidade do sítio e incorporar as conclusões na concepção do sítio.

Assumindo que um sítio web é um sistema hipermédia principalmente preocupado com a comunicação, os autores usaram retórica clássica para criar o modelo, usando os sete *loci* ou *argumenta* de Cícero, muito à semelhança dos jornalistas que usam as questões: quem, o quê, porquê, onde, quando e onde.

Este modelo permite uma aproximação *multistakeholder* que considera os pontos de vista de todos os envolvidos (o dono do sítio ou gestor, os utilizadores e os envolvidos na concepção e implementação).

As sete dimensões do modelo são definidas na tabela 2-1.

Quis? (Quem?)	Identidade
Quid? (O quê?)	Conteúdo
Cur? (Porquê?)	Serviços
Ubi? (Onde?)	Localização
Quando?	Gestão
Quomodo? (Como?)	Usabilidade
Quibus Auxilius? (Com que meios?)	Viabilidade

Tabela 2-1 – Dimensões do modelo 2QCV3Q

A identidade é afectada pelo design gráfico e personalização. Para avaliar o conteúdo, avalia-se a qualidade da informação em relação ao domínio do sítio. Os serviços são o conjunto de funcionalidades que os utilizadores esperam encontrar no sítio. Estes serviços devem ainda oferecer segurança. A dimensão localização é acerca da facilidade de aceder ao sítio (por exemplo, uma URL fácil) e do modo de interagir

com este. A gestão envolve fazer *updates* de informação e assegurar a funcionalidade do sítio (por exemplo, não existirem ligações inexistentes). A usabilidade reúne características como o uso do sítio com variados *browsers* preferencialmente sem *plugins* (e de outras plataformas, por exemplo móveis) e a acessibilidade de portadores de deficiência física. A viabilidade é uma nova dimensão que não tinha sido proposta no primeiro artigo e reúne aspectos como recursos humanos, recursos de hardware e software – é o investimento no sítio.

Ciceronian Loci	Atributos
Quis (Quem?)	Identificação
Identificação	<ul style="list-style-type: none"> ▪ Marca (organização ou companhia); carisma ▪ Imagem Caracterização <ul style="list-style-type: none"> ▪ Design ▪ Personalização
Quid (o quê?)	Cobertura
Conteúdo	<ul style="list-style-type: none"> ▪ Domínio refere os objectivos do dono e dos utilizadores ▪ Valor da informação e ligações Exactidão <ul style="list-style-type: none"> ▪ Qualidade da informação ▪ Fonte(s), autor(es)
Cur (Porquê?)	Funcionalidades
Serviços	<ul style="list-style-type: none"> ▪ Adequação aos objectivos do dono ▪ Adequação aos objectivos dos utilizadores Controlo <ul style="list-style-type: none"> ▪ Exactidão ▪ Segurança, ética, e privacidade
Ubi (Onde?)	Acessibilidade
Localização	<ul style="list-style-type: none"> ▪ URL intuitivo ▪ Recuperação Interactividade <ul style="list-style-type: none"> ▪ Informação de contactos ▪ Construção da comunidade
Quando	Actualidade
Gestão	<ul style="list-style-type: none"> ▪ Updates e revisões ▪ Datas Manutenção <ul style="list-style-type: none"> ▪ Check-up ▪ Ferramentas
Quomodo (Como?)	Acessibilidade
Usabilidade	<ul style="list-style-type: none"> ▪ Requisitos de hardware e de software ▪ Pessoas com deficiências Navegabilidade

	<ul style="list-style-type: none"> ▪ Estrutura, orientação ▪ Tempos de download
	<p>Inteligibilidade</p> <ul style="list-style-type: none"> ▪ Línguas ▪ Nível de terminologia
Quibus Auxiliis (Com que meios?)	<p>Recursos</p> <ul style="list-style-type: none"> ▪ Recursos humanos e financeiros ▪ Tempo
Viabilidade	<p>Tecnologia de informação e comunicação</p> <ul style="list-style-type: none"> ▪ Hardware (computador, redes) ▪ Software (Implementação, integração)

Tabela 2-2 – Atributos do modelo 2QCV3Q

2.2.2 Implementação do modelo

As fases da implementação deste modelo são as seguintes: 1-Uma fase de configuração inicial, que inclui a análise e especificação dos requisitos de avaliação; 2-Uma fase de concepção, que define o plano de avaliação e técnicas; 3-Uma fase de realização que aplica técnicas de questionários e modalidades de medidas especificadas nos planos de avaliação.

Esta fase de configuração inicial deve considerar os seguintes elementos: o propósito da avaliação, o tipo e domínio do sítio, a fase de desenvolvimento do sítio, os objectivos do dono do sítio, os utilizadores e os seus perfis. Baseada na informação recolhida na fase inicial, passamos à fase seguinte.

Na fase de concepção identificam-se as modalidades de avaliação apropriadas para os atributos do modelo referentes aos requisitos de qualidade definidos anteriormente. Nesta fase devemos determinar modalidades de questionários que podem variar dependendo das técnicas e ferramentas utilizadas e número de utilizadores. Em alguns casos, onde os atributos são qualitativos, têm de intervir especialistas. Para os atributos quantitativos pode ser adoptada uma aproximação baseada no uso de identificadores por *checklist*.

Na fase de realização aplicam-se técnicas de questionários e modalidades de medida especificadas no plano de avaliação. Normalmente baseiam a avaliação em várias visitas ao sítio, mas pode acontecer que precisem de ficheiros que incluem programação, só disponíveis para o *webmaster*. Usando métodos apropriados, comparam os resultados com o perfil de qualidade definido na primeira fase. A

comparação pode ser enriquecida usando gráficos, por exemplo, tipo radar. Este processo pode ser interativo.

2.2.3 Capacidade de automatização

A avaliação pode ser parcialmente automatizada para alguns atributos usando ferramentas de software, por exemplo, *site watchers* e avaliadores, medidores de tamanho de página, etc.

2.2.4 Implementação ou prova de conceito

Este modelo foi desenvolvido durante cerca de três anos e aplicado em domínios como turismo, educação, empresariais e sítios de serviço ao cliente.

2.2.5 Comentários a este trabalho

Este modelo é baseado na retórica clássica sendo as características um pouco diferentes das definidas pela ISO. Contudo, se os sítios são geralmente da informação, faz sentido usar uma lógica semelhante à dos jornalistas.

Esta aproximação é independente do domínio, apesar de funcionar como meta-modelo e instanciada para vários domínios. Pode ainda ter vários níveis de detalhe. Segundo o autor, a estrutura simples facilita a compreensão e aplicação.

As fraquezas deste método são: não é automático ou só algumas características é que são. Os sítios alvos são os sítios de informação.

O gráfico seguinte representa a nossa avaliação deste trabalho, onde as contribuições estão numa escala de 0 a 4.

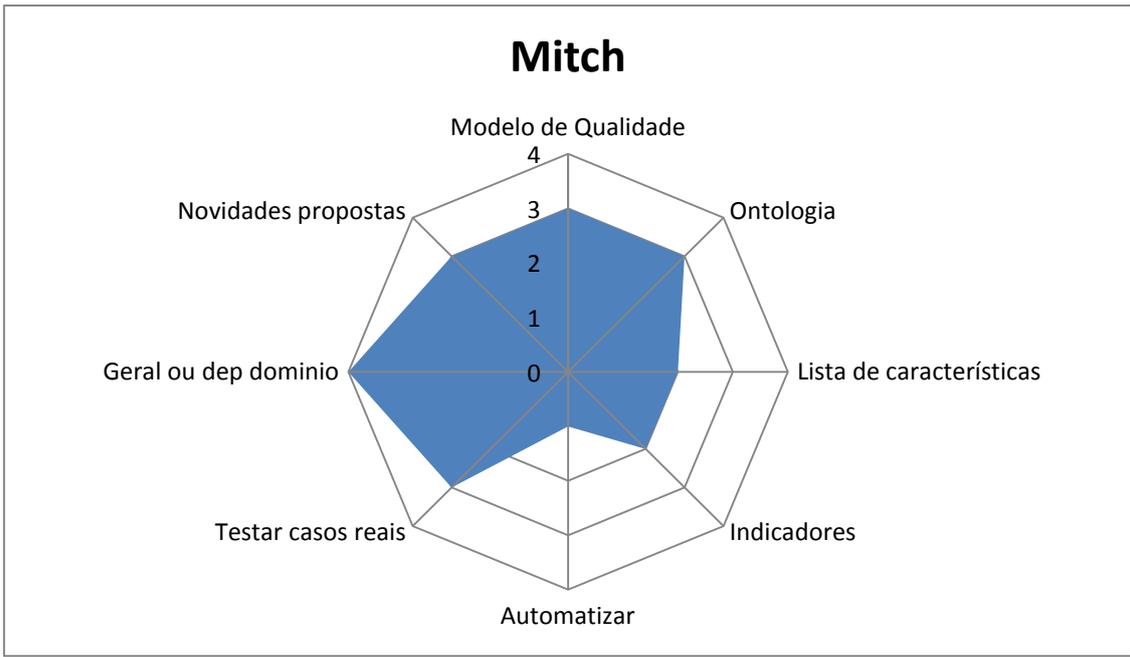


Figura 2-2 Avaliação do trabalho de Mitch

2.3 MILE (Blas, et al. 2002I).

O método MILE (método de avaliação Milano-Lugano) combina métodos de inspeção e testes empíricos [19]. Este método é completamente dependente de analistas humanos, sejam especialistas ou utilizadores finais para atribuir as pontuações. O método contém características que são passíveis de atribuição de pesos.

O método MILE+ é uma evolução deste método, que introduz uma distinção entre as análises dependentes e independentes das aplicações. Também introduz uma actividade específica, a inspeção técnica.

2.3.1 Modelo MILE

O método MILE é baseado numa combinação de testes de inspeção (avaliação de especialistas que exploram a aplicação) e testes empíricos (utilizadores finais que usam a aplicação, com a supervisão de especialistas de usabilidade).

São introduzidos dois conceitos: **Tarefas abstractas (AT)** – Lista de acções genéricas que podem ser aplicadas a um conjunto grande de aplicações; **Tarefas concretas (CT)** – Lista de acções específicas que são definidas para uma determinada aplicação. O método MILE tenta assim separar os diferentes níveis de análise: tecnologia, navegação, conteúdo, comunicação, gráfico, etc.

Os níveis do modelo MILE são os seguintes: **Conteúdo** – este nível analisa a quantidade do conteúdo (em termos de eficácia e qualidade da comunicação); **Serviços** – todas as funcionalidades que um sitio web oferece aos seus utilizadores; **Navegação** – 1) os modos diferentes pelos quais o utilizador pode atingir uma parte específica da informação; 2) as ligações utilizadas para passar de uma parte específica de informação para outra; **Características cognitivas da interface** – o utilizador compreende e lembra-se do conteúdo e estrutura da aplicação. Está relacionada com escolhas feitas no design; **Gráfico e estético** – design gráfico (cores, tipos de fontes, imagens, etc.) e *layout* (distribuição espacial dos elementos gráficos na página). Este método foi usado com sucesso no domínio de sítios de museus.

O MILE+ é uma evolução do anterior. As actividades de avaliação que compõem este modelo são: i) avaliação técnica – problemas que não estão relacionados com partes específicas da aplicação, ou seja, podem ser aplicados todos os sítios; ii) avaliação de experiência do utilizador final; iii) avaliação de utilizador baseado em cenários. Estas duas últimas actividades estão ligadas com a natureza da aplicação,

sendo por isso específicas para o domínio da mesma. Esta evolução do método consegue separar a análise dependente e independente da aplicação mas é bastante virada para a usabilidade e quase completamente dependente de utilizadores que testam.

2.3.2 Implementação do modelo

O método requer os seguintes passos: i) selecção da parte relevante da aplicação; ii) selecção das tarefas abstractas que são relevantes para os cenários de utilizadores destino; iii) execução das tarefas abstractas, dando pontuações a cada atributo e para cada cenário de utilizadores; iv) atribuir pesos aos atributos e tarefas escolhidas (uma determinada tarefa pode ser mais relevante do que outra); v) produção de medidas de avaliação quantitativas (aplicar pesos às pontuações).

2.3.3 Capacidade de automatização

O método pode ser automatizado por ferramentas externas.

2.3.4 Implementação ou prova de conceito

O método MILE original foi usado com sucesso no domínio de sítios de museus. A evolução do método foi usada também no mesmo domínio, os museus, mas com diferentes grupos de utilizadores e avaliadores.

2.3.5 Comentários a este trabalho

Os pontos fortes deste trabalho ou novidades são a introdução de actividades técnicas (independentes da aplicação) e específicas (dependentes da aplicação). Embora este método só tenha sido testado no domínio de sítios de museus, pode ser aplicado a outro domínio. Contudo, este método não vai de encontro ao nosso objectivo: o método não pode ser automatizado; é bastante dependente dos utilizadores e avaliadores.

Em seguida temos o gráfico representativo das contribuições propostas por este método:

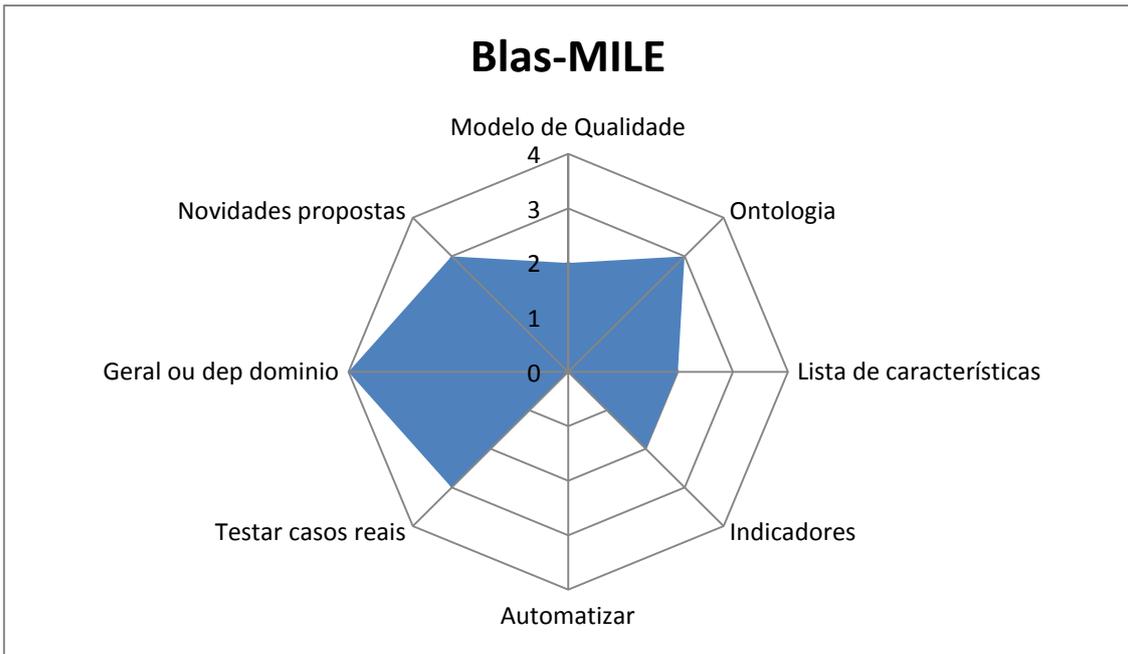


Figura 2-3 Avaliação do trabalho de Blas

2.4 Minerva (Minerva Working Group, 2005)

O projecto Minerva representa uma iniciativa importante no campo da qualidade de sítios web [20]. Este projecto tenta constituir um modelo de qualidade baseado em 10 princípios (ou áreas) de qualidade.

2.4.1 Modelo Minerva

O modelo de qualidade do projecto Minerva consiste em 10 princípios de qualidade:

1. **Transparência** – um sítio web de qualidade deve ser transparente, definir univocamente a sua identidade, o seu propósito ou missão, e organização responsável pela sua gestão.
2. **Efectividade** – um sítio web de qualidade deve ser formado por conteúdo que crie um sítio web efectivo para os utilizadores. Este conteúdo deve ser seleccionado de forma relevante, exacto e com informação de suporte. Esta área tem ainda a apresentação e navegação.
3. **Manutenção** – um sítio web de qualidade deve implementar uma política de qualidade de serviço de forma a assegurar que o sítio tem manutenção e é actualizado a um nível apropriado. Deve ainda ter manutenção técnica, não obsolescência de conteúdos, manutenção de conteúdos e estes devem estar actualizados.
4. **Acessibilidade** – um sítio web de qualidade deve ser acessível a todos os utilizadores, independentemente da tecnologia usada ou limitações físicas dos utilizadores, incluindo navegação, conteúdo e elementos interactivos. Subáreas: acessibilidade técnica e tecnologia de acessibilidade.
5. **Centrado no utilizador** – um sítio web de qualidade deve ser centrado no utilizador levando em conta as necessidades destes e assegurando relevância e facilidade de uso respondendo à avaliação e *feedback*. Deve existir consulta, envolvimento e contribuição dos utilizadores.
6. **Resposta** - um sítio web de qualidade deve assegurar resposta, permitindo aos utilizadores contactar o sítio e receber a resposta apropriada. Quando adequado devem ser encorajadas perguntas, partilha de informação, e discussões com e entre os utilizadores.

7. Multilíngue - um sítio web de qualidade deve ter em conta a importância da língua providenciando pelo menos um nível de acesso mínimo em mais do que uma linguagem.
8. Inter-operação - um sítio web de qualidade deve procurar ser inter-operável entre redes culturais para habilitar os utilizadores a localizarem facilmente o conteúdo e os serviços que precisam. Áreas deste princípio: *metadata*, tecnologias usadas no sítio, exposição de informação a motores de pesquisa e facilidade de descoberta por utilizadores e ferramentas, e ainda suporte a pesquisa distribuída ou remota.
9. Gestão - um sítio web de qualidade deve ser gerido de modo a que respeite os requisitos legais como privacidade e outros, e deve conter de forma clara os termos e condições em que o sítio web e os conteúdos podem ser usados. Subáreas: direitos de propriedade do conteúdo, direitos de propriedade do sítio, protecção do dono do sítio contra litigação, privacidade do utilizador, reutilização do conteúdo.
10. Preservação - um sítio web de qualidade deve adoptar estratégias e normas de forma a assegurar que o sítio e os seus conteúdos podem ser preservados no longo prazo. Estas estratégias contêm escolhas de modelo de dados, formato de ficheiros, tecnologias de apresentação do sítio e média. Procedimentos de *backup* e *restore* devem existir assim como planos de recuperação de desastres.

Cada princípio tem uma lista de verificação.

2.4.2 Implementação do modelo

O modelo pode ser implementado nas diversas fases de ciclo de vida de sítio web. Estas fases podem ser identificadas como: 1-Planeamento do sítio web; 2-Design do sítio web; 3-Seleção do conteúdo; 4-Processo de digitalização; 5-Arquivo e preservação do material digital principal; 6-Introdução de *metadata*; 7-Implementação do sítio web; 8-Publicação on-line; 9-Manutenção e resposta a *feedback* de utilizadores.

Em seguida mostra-se uma grelha com o cruzamento das fases e princípios avaliados de 1 a 3.

	Planeamento	Design	Seleção do conteúdo	Digitalização	Arquivo	Metadada	Implementação	Publicação	Manutenção
Transparência	2	3	1	1	1	1	3	3	2
Efectividade	2	3	3	1	1	2	3	3	3
Manutenção	2	1	2	2	1	1	1	3	3
Acessibilidade	3	3	1	2	1	1	3	1	1
Centrado no utilizador	2	3	1	1	1	1	3	1	2
Responsivo	2	2	3	1	1	1	2	3	3
Multilingue	3	3	2	2	1	1	3	2	1
Inter-operação	3	3	1	3	2	3	3	2	2
Gestão	1	1	3	1	2	1	1	1	1
Preservação	1	1	2	3	3	3	1	1	2

Tabela 2-3 - Cruzamento entre princípios de qualidade e fases de implementação

O autor aponta ainda as fases mais críticas da qualidade: Planeamento do sítio web; Design do sítio web; Implementação do sítio web; Publicação on-line.

2.4.3 Capacidade de automatização

Estes princípios de qualidade podem ser automatizados em parte com recurso a ferramentas externas.

2.4.4 Implementação ou prova de conceito

Cada um destes princípios tem uma lista de verificação e existem guias para a sua implementação em concreto. Foram feitos estudos em sítios culturais ou de museus.

2.4.5 Comentários a este trabalho

Este modelo de qualidade aproxima-se mais de um modelo de qualidade padrão para um outro domínio. Consiste em 10 princípios de qualidade que na implementação contêm lista de verificação, algumas das quais passíveis de automatização. É no entanto, dirigido apenas a sítios culturais e de museus.

Segundo a nossa classificação, podemos rever este trabalho com o gráfico seguinte:

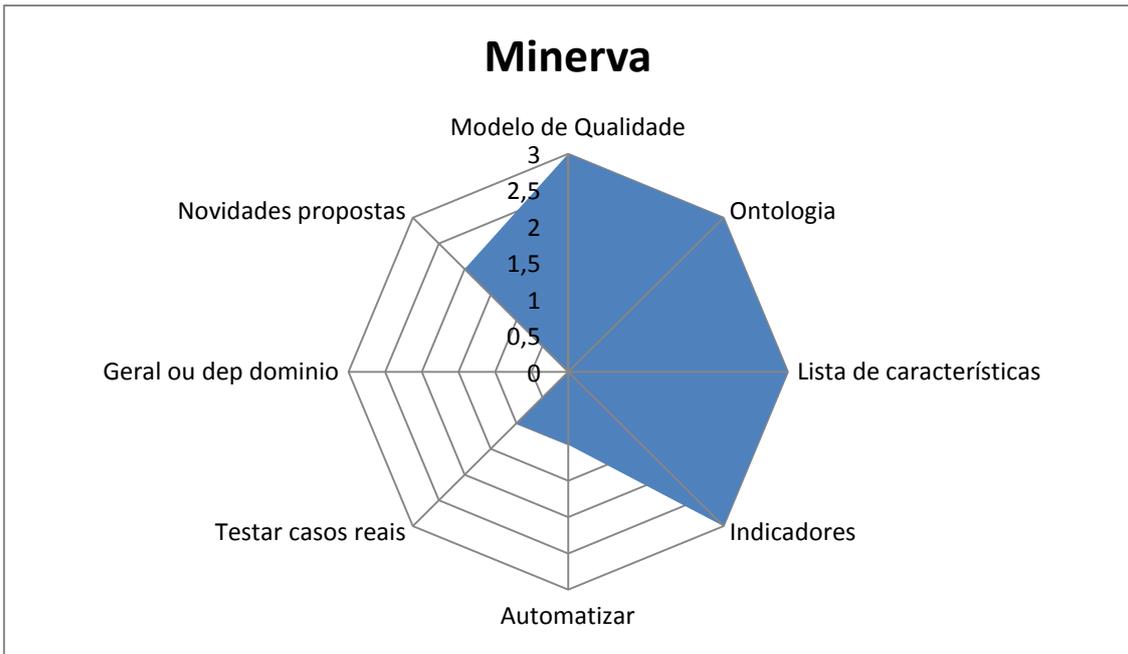


Figura 2-4 Avaliação do trabalho efectuado no projecto Minerva

2.5 Enoteam (Polillo, 2005)

Segundo o autor [21], para fazer a avaliação de um sítio web, é necessário estar definido um modelo de qualidade. O modelo de qualidade apresentado a seguir contempla características externas e de qualidade em uso, deixando de fora as internas.

2.5.1 Modelo Enoteam

Este modelo divide-se em 8 macro-características: Objectivos (estratégia e requisitos, definidos por consultores); Arquitectura ou web design (Web designers); Comunicação (visual designer); Funcionalidade (integradores de sistemas); Gestão de conteúdos (editores de conteúdos); Operação do sítio (Web master); Gestão do servidor; Ligação internet.

O autor sugere que algumas destas características se podem juntar num macro-modelo, obtendo assim as seguintes características: Arquitectura (estrutura geral do sítio e navegação, coerência de layout); Comunicação (comunicação da marca, coerência com a imagem da organização, graficamente atractivo); Funcionalidade (mede as funcionalidades que o sítio apresenta, e se são adequadas e apresentadas correctamente); Conteúdos (qualidade do conteúdos, compreensão); Gestão (operação do sítio, funcionamento correcto do servidor, qualidade da ligação internet); Acessibilidade (capacidade para todos os utilizadores acederem rapidamente ao sítio sem problemas); Usabilidade (mede o grau de dificuldade ou facilidade com que se usa o sítio).

2.5.2 Implementação do modelo

O modelo é implementado usando um sistema de votação para cada característica, e pode ser representado num diagrama tipo radar com uma escala de 0 a 4, onde 0 significa “muito mau” e 4 significa “muito bom”. Mais tarde este modelo foi modificado para usar pesos em cada sub-característica, em vez da soma ou media simples.

Assim, é sugerido usar o seguinte esquema de perguntas:

Característica	Sub-característica	Pergunta
Arquitectura	Estrutura	A estrutura informativa do sítio é adequada?
	Mapa do sítio	Existe um mapa do sítio que representa com clareza a sua estrutura?
	Navegação	A navegação do sítio é adequada?
Comunicação	Homepage	A página inicial comunica claramente o objectivo do sítio?
	Imagem de marca	O sítio é coerente com a imagem de marca?
	Design	O design do sítio é adequado?
Funcionalidade	Adequação	As funcionalidades do sítio são adequadas?
	Correcção	As funcionalidades do sítio são apresentadas correctamente?
Conteúdo	Categorização/Rótulos	A informação é classificada de modo adequado?
	Estilo	O estilo do texto é adequado à web.
	Informação	A informação é adequada, pertinente, de confiança e actualizada?
	Localização	O sítio é localizado correctamente?
Gestão	Disponibilidade	O sítio está sempre activo e disponível?
	Monitorização	O uso do sítio é monitorizado adequadamente?
	Actualizações	O sítio é constantemente actualizado e melhorado?
	Relação com utilizadores	Existe mecanismo de <i>feedback</i> ?
Acessibilidade	Tempo de acesso	Os tempos de acesso são adequados?
	Disponibilidade	O sítio é fácil de encontrar?
	Independência do browser	O sítio é acessível com todos os browsers?
	Acessibilidade para deficientes	O sítio é usável por utilizadores portadores de deficiência?
Usabilidade	Eficácia	O utilizador obtém os resultados esperados de modo preciso e completo?
	Eficiência	O esforço dispendido pelo utilizador para obter os resultados é aceitável?
	Satisfação do utente	O sítio é confortável e bem aceite pelo utilizador?

Tabela 2-4 Características e sub-características do modelo de qualidade Enoteam.

2.5.3 Capacidade de automatização

O modelo como proposto não é passível de automatização, por duas razões: o sistema de votos, que requer peritos ou utilizadores, e algumas das características, que são subjectivas.

2.5.4 Implementação ou prova de conceito

Não existe nos artigos; o modelo proposto é teórico.

2.5.5 Comentários a este trabalho

Os pontos fortes desta aproximação são a sua simplicidade. Para o nosso objectivo, as suas fraquezas são a impossibilidade de automatização e algumas características serem subjectivas e outras ainda só serem possíveis de medir durante o tempo de operação.

Gráfico de avaliação do modelo:

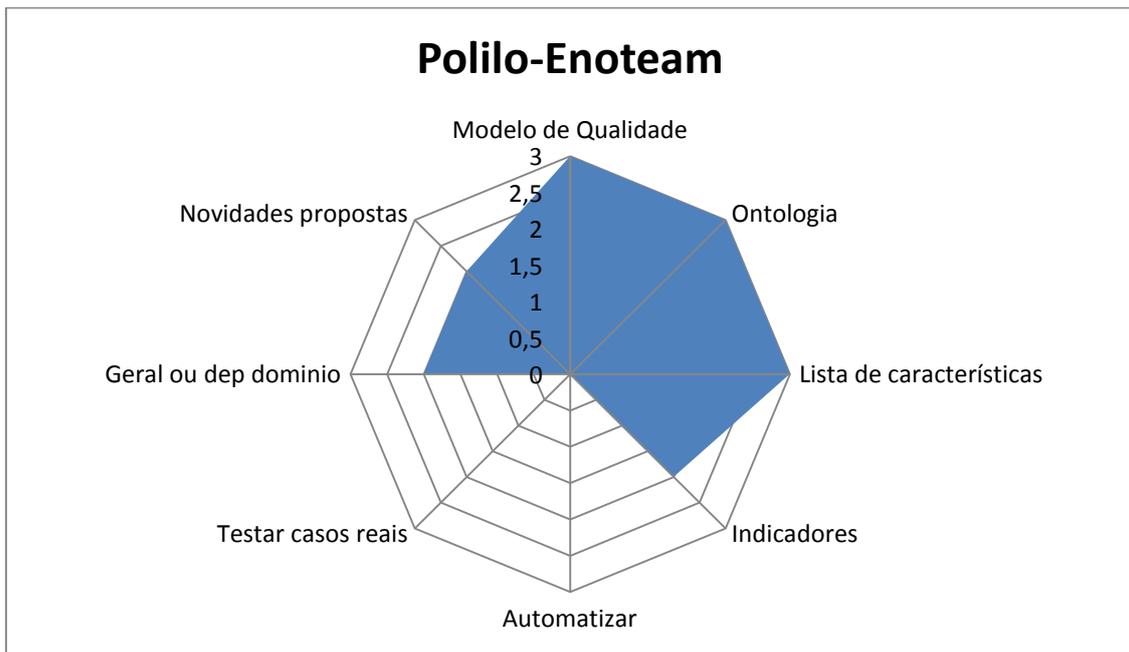


Figura 2-5 Avaliação do trabalho de Polilo (Enoteam)

2.6 Modelo de qualidade de sítios web 5D (Signore, 2005)

O modelo para qualidade de sítios web de Signore, embora não tenha sido o primeiro por ordem cronológica, é o que mais se assemelha ao trabalho que vamos efectuar [22].

Este trabalho tenta definir um modelo de qualidade (modelo conceptual do sítio/página) e um conjunto de características de sítios web passíveis de serem medidas automaticamente. É também feita uma investigação de como algumas características internas afectam características externas. Contudo, algumas das características definidas ainda não são passíveis de automatização completa, sendo que precisam de análise humana.

O autor aponta algumas limitações à avaliação de qualidade que devem ser tomadas em consideração: Critérios gerais – devem ser tidos em conta critérios que tem a ver com o domínio do sítio, ou posicionamento da página ou parte da página – podemos resolver este problema dando um peso aos critérios. Os critérios devem ser quantitativos em vez de qualitativos. Os critérios devem ser ortogonais (e não aparecer repetidos e sobrepostos). Muitos critérios de avaliação têm a ver com a acessibilidade ou usabilidade. Deve existir granularidade (muitas vezes não há distinção entre um sítio (conjunto de páginas com uma base comum) e uma página. Perspectivas diferentes de qualidade: Utilizador – interessado na qualidade externa, em aspectos como usabilidade e funcionalidade; *Developer* – interessado em aspectos técnicos, como compatibilidade, evolução, manutenção, portabilidade, etc.

2.6.1 O modelo

A maior parte da informação é obtida através de análise do código, incluindo as folhas de estilo. Assim, neste trabalho, começa-se por definir um modelo de sítio e uma base de dados de qualidade, onde são guardados os resultados obtidos por algumas ferramentas já existentes (por exemplo *crawlers*).

2.6.1.1 Modelo de sítio

O modelo de sítio, na aproximação do artigo, consiste no seguinte: Um sítio é uma colecção de páginas. Uma página é uma colecção de componentes de página. Um

componente de página pode incluir outros componentes de página. Uma página tem uma folha de estilo. Uma folha de estilo pode importar outra folha de estilo.

A página tem certas propriedades, como título e metadados, e podem ser testadas para código válido.

Os componentes de página têm propriedades, como: Tipo (div, célula de tabela); Propósito (cabeçalho, corpo, índice/menu, rodapé, navegação, etc); Número de ligações (internas de pagina, internas de sítio e externas).

2.6.1.2 Modelo de qualidade

O modelo considerado para avaliação é um modelo 5D, que compreende: Correção / Validação; Apresentação; Conteúdo; Navegação; Interação. É ignorada aqui a dimensão do desempenho, que nós não vamos ignorar no nosso modelo.

Em seguida é apresentado um conjunto de características a serem analisadas, de acordo com este modelo:

Correção / Validação -A dimensão Correção / Validação pode ser avaliada recorrendo ao DTD ou outro cabeçalho da página com indicação da formatação usada.

Apresentação - A apresentação divide-se nas características: *layout* da página, Apresentação do texto, Apresentação multimédia, Apresentação das ligações e Formulários.

Conteúdo - O conteúdo tem a ver com o tamanho dos parágrafos e títulos, se tem sumários ou leads, autor, última actualização, etc. Existem algumas características que não mensuráveis automaticamente, mas que caem nesta categoria, como a correção da informação, uso de linguagem, informação mais detalhada para utilizadores avançados e menos detalhada para utilizadores ocasionais, etc.

Navegação - A navegação tem a ver com as tipologias das ligações de menu e de submenus, e logo a navegação no sítio e posição no mesmo.

Interação - O modo principal de interação é com formulários. Aqui analisamos a transparência (o utilizador é informado das consequências de submeter um formulário; recuperação (*undo*); ajuda.

2.6.2 Capacidade de automatização

Em 2005 estavam a começar a implementar uma ferramenta automatizada, que seria calibrada com teste com utilizadores reais. A calibração tem a ver com a atribuição de pesos às características definidas anteriormente.

2.6.3 Implementação ou prova de conceito

Não encontramos na literatura prova de conceito.

2.6.4 Comentários a este trabalho

Os pontos fortes deste trabalho são efectivamente a tentativa de definição de um modelo de qualidade passível de automatização. Este modelo e sua definição de algumas características, e sua medição, foram bastante importantes na definição do nosso modelo. Como fraqueza só podemos apontar o facto de alguma das características definidas serem difíceis de medir automaticamente, e serem subjectivas.

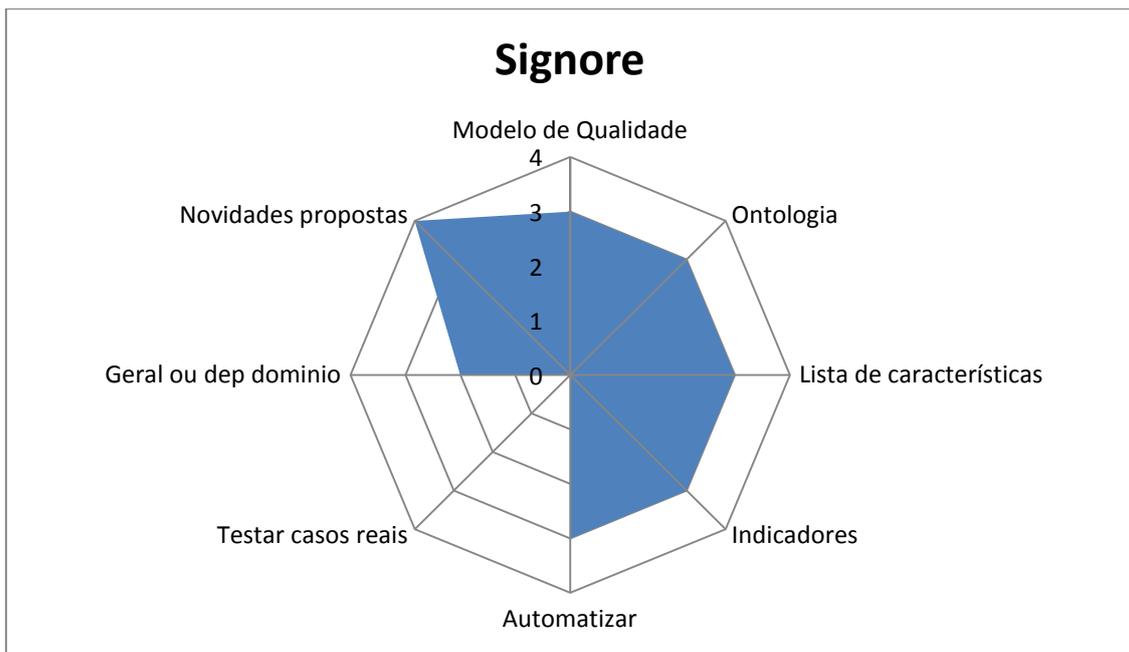


Figura 2-6 Avaliação do trabalho de Signore

2.7 Web Quality Model (Calero, 2005)

O trabalho de Calero incide na classificação de métricas propostas na literatura, visando a sua correcta identificação e definição [23]. Foram avaliadas e classificadas 325 métricas, provenientes de 60 artigos.

2.7.1 Web quality model

O modelo chamado Web Quality Model, ou WQM, consiste num cubo com 3 dimensões, que são:

Características Web: Navegação – funcionalidades para aceder à informação e navegar no sítio; Conteúdo - (dados, ou seja, texto, figura, imagens, vídeos, e ainda programas, aplicações, e outros. A fronteira não está bem definida); Apresentação – A apresentação está relacionada com o modo como o conteúdo e navegação são apresentados ao utilizador.

Características de qualidade: Funcionalidade – conjunto de atributos que descrevem um conjunto de funcionalidades e suas propriedades; Fiabilidade – conjunto de atributos que medem a capacidade do produto manter o nível de desempenho nas condições estabelecidas por um dado período de tempo; Usabilidade – conjunto de atributos que medem a capacidade do produto ser compreendido, a nível de funcionamento e operação, por um conjunto de utilizadores definido ou implícito; Eficiência – conjunto de atributos que são suportados na relação entre o nível de performance e os recursos utilizados; Portabilidade - conjunto de atributos que englobam a capacidade do software ser transformado de um ambiente para outro; Capacidade de manutenção – Conjunto de atributos que medem o esforço que é preciso para fazer modificações pedidas.

Processos do ciclo de vida: Desenvolvimento; Operação; Manutenção; Esforço; Reutilização.

A dimensão das características de qualidade está dividida nas características definidas na ISO9126. A dimensão de características web, inclui os aspectos chamados clássicos da web, ou seja, a navegação, conteúdo e navegação.

2.7.2 Implementação do modelo

Este trabalho consiste essencialmente em classificação de características.

2.7.3 Capacidade de automatização

Algumas das características definidas são automatizáveis.

2.7.4 Implementação ou prova de conceito

O trabalho incide na definição de características e sua classificação. Não temos conhecimento de implementação.

2.7.5 Comentários a este trabalho

O trabalho consiste na classificação exaustiva de métricas encontradas na literatura, segundo um cubo e seus eixos, divididos em subcategorias.

Pontos Fortes – Tenta fazer um reaproveitamento e ordenação das características de qualidade existentes na literatura, com a finalidade de por ordem na classificação das mesmas.

As suas maiores fraquezas são que não tem instruções para a sua implementação. O modelo também não se encontra bem definido, uma vez que para a quantidade de características estudadas, há algumas que se sobrepõem.

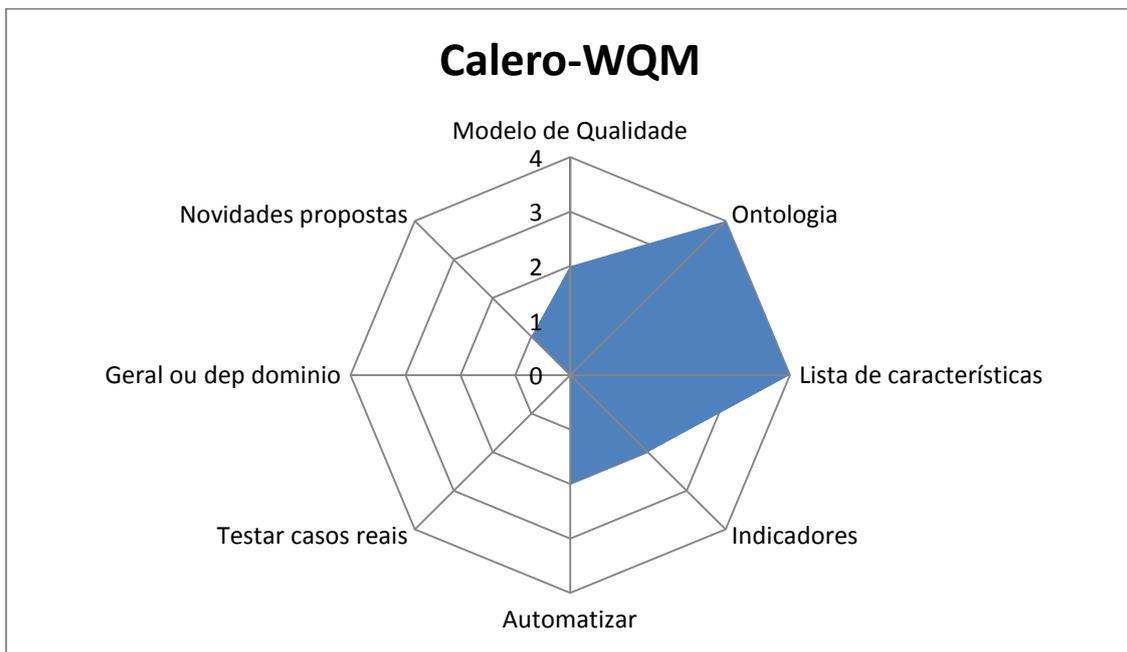


Figura 2-7 Avaliação do trabalho de Calero

2.8 Web Q-Model (Cimino e Micali, 2008)

Cimino e Micali propuseram o Web q-model, no onde partiram das propostas existente e requisitos gerais para os modelos de qualidade [24]. Foi feita uma colecção e eliminaram duplicados, tentando extrapolar um modelo geral para todos os domínios.

2.8.1 Modelo de qualidade Web Q-Model

O Web q-model é um modelo de qualidade Web que tem em conta 3 níveis diferentes de qualidade e, devido a isso, pode ser realizado em tempos de execução diferentes, conforme o que se quer medir. As características estão situadas na qualidade em uso.

A metodologia usada para a construção deste método foi a seguinte: No primeiro passo reuniram-se todos os atributos de todos os modelos; no segundo passo, separaram-se os atributos em grupos, segundo o seu significado semântico. Neste passo foram eliminadas as repetições e os grupos forma optimizados. O resultado foi um grupo de 6 dimensões: Comunicação e interface; Conteúdo; Navegação; Gestão e acessibilidade; Interactividade; Acessibilidade para pessoas com deficiência. As dimensões são definidas da seguinte forma:

Comunicação e interface - A comunicação e interface têm a ver com o grafismo das páginas e sua distribuição (rótulos, títulos, ícons, qualidade das imagens, disposição dos elementos da página, identidade visual, cores, fonte, identificação de ligações, parágrafos, elementos multimédia, comportamento da interface).

Conteúdo - O conteúdo agrega as características pondo de lado a estrutura do sítio: qualidade de informação, actualidade de conteúdos, cobertura do domínio do sítio ou amplitude, precisão do conteúdo, selecção e relevância do conteúdo, compreensão e facilidade de leitura do conteúdo, oferecer pelo menos um serviço básico noutra língua e conteúdo multimédia ligado com o texto da página onde está inserido.

Navegação - As características inseridas na navegação são: estrutura do sítio ou navegação fácil; ajudar o utilizador a saber onde está – *bread crumbs*; usar ferramentas de navegação adequadas, fáceis e intuitivas; suportar navegação para trás; navegação com o mínimo de *clicks*.

Gestão e acessibilidade - O grupo gestão e acessibilidade tem as características: requisitos de software e hardware - verificar browsers e plataformas, evitar *plugins* e extensões proprietárias; controlar a presença do sítio nos principais motores de pesquisa; verificar correcção do código; otimizar o tempo das páginas (*download* + *parsing* + *rendering*) e *streaming*; manutenção - monitorizar o sítio, operação ininterrupta, acessibilidade técnica; assegurar funcionalidade adequada e correcção; usar políticas de segurança e garantir uso correcto de informação pessoal; gestão de erros - verificar reacções do sistema, mensagens de erro e apagar ou reparar ligações erradas.

Interactividade - O grupo interactividade tem as características: assegurar a interacção adequada em funcionalidade; considerar o envolvimento do utilizador e contribuição (comentários, *guest book*, etc); help e informação de contacto; assegurar respostas aos utilizadores rápidas, pertinentes e educadas; satisfação do utilizador fazer o sítio aceitável, agradável e fácil de usar; usar a melhor posição e largura para os elementos.

Acessibilidade para portadores de deficiência - A acessibilidade para pessoas portadores de deficiência é a conformidade às *guidelines* W3C [25].

Cada característica tem várias sub-características que são numeradas e classificadas segundo 3 níveis, Básico (Q), Normal (QQ), ou Emocionante (QQQ), e distribuídas num quadro nas respectivas colunas.

2.8.2 Implementação do modelo

A implementação do modelo consiste na avaliação das características como uma *checklist*, que depois são somadas para atingir o nível Q, QQ, ou QQQ.

2.8.3 Capacidade de automatização

Muitas das características propostas requerem avaliação por pessoas, limitando assim a sua automatização.

2.8.4 Implementação ou prova de conceito

Não existe implementação ou prova de conceito.

2.8.5 Comentários a este trabalho

Este trabalho permite uma abordagem simples para avaliação de sítios. Contudo a natureza subjectiva de algumas características não permite a avaliação automática. Introduz 3 níveis de qualidade, talvez indo beber inspiração aos 3 níveis de acessibilidade (A, AA, AAA) [25].

A nossa classificação a este modelo é a seguinte:

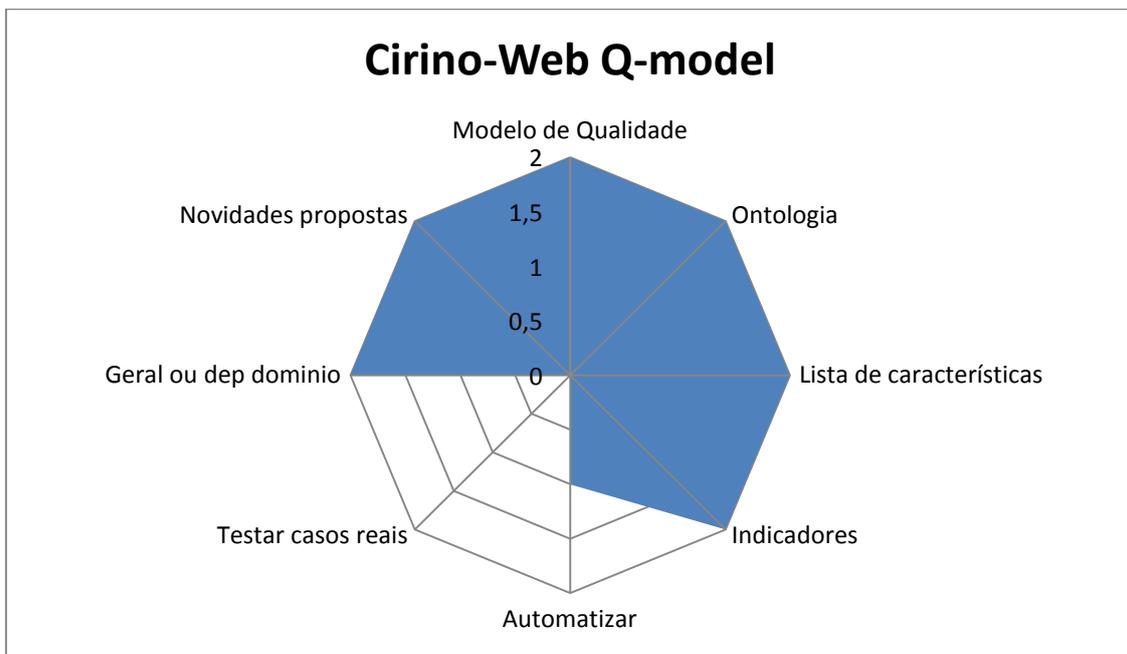


Figura 2-8 Avaliação do trabalho de Cirino

2.9 Tabela comparativa

Apresenta-se na tabela 2-5 a tabela comparativa destes trabalhos revistos.

	Olsina - WebQem	Mitch - 2QCV3Q	Blas- MILE	Min erva	Polilo- Enoteam	Sign ore	Calero- WQM	Cirino-Web Q-model	Mé dia
Modelo de Qualidade	3	3	2	3	3	3	2	2	2,6
Ontologia	2	3	3	3	3	3	4	2	2,9
Lista de características	3	2	2	3	3	3	4	2	2,8
Indicadores	3	2	2	3	2	3	2	2	2,4
Automatizar	3	1	0	1	0	3	2	1	1,4
Testar casos reais	3	3	3	1	0	0	0	0	1,3
Geral ou dep domínio	4	4	4	0	2	2	0	2	2,3
Novidades propostas	4	3	3	2	2	4	1	2	2,6
Total	25	21	19	16	15	21	15	13	

Tabela 2-5 tabela comparativa dos trabalhos

As classificações da tabela são entre 0 a 4. O modelo de qualidade, e a ontologia têm o valor mais alto quanto mais definido estiver. A lista de características e os indicadores também seguem este sistema. A capacidade de automatização tem o valor mais alto quando é possível automatizar. Se os preponentes testam casos reais tem maior pontuação mas são penalizados se não o descrevem bem. Se é um modelo dependente do domínio, tem pontuação maior (4) sendo que não dependente do domínio recebe 2. Quanto mais novidades propostas, mais pontuação recebem.

2.10 Análise dos resultados

Estes trabalhos são, na sua maior parte, virados para uma análise compreensiva por peritos, que dão lugar a uma avaliação demorada e subjectiva, excepto o trabalho[22], que é um trabalho académico sobre a possibilidade de recolha automática.

Devido a estes factores, conseguimos utilizar algumas das suas contribuições e outras não. O que conseguimos utilizar foi o conjunto da lista de características /indicadores definidos nesta literatura, embora tenhamos adicionado algumas e rejeitado

mais ainda, devido à sua incapacidade de automação. Conseguimos também utilizar as mesmas macrocaracterísticas de alguns deles, que por sua vez são baseados na ISO9126. Tiramos algumas ideias da ontologia para recolha, especialmente do trabalho de Signore [22].

Como vemos pela coluna das médias, as grandes necessidades nestas áreas, são um modelo de qualidade automatizado, e a análise de casos reais. São estes os maiores contributos desta dissertação.

Capítulo 3

Modelo de qualidade para presenças na web

3	Modelo de qualidade para presenças na web.....	42
3.1	Modelo de qualidade.....	43
3.2	Lista de características	45
3.3	Proposta de ontologia para avaliação.....	47
3.4	Descrição de indicadores quantitativos.....	49
3.5	Hipóteses a ensaiar.....	52

Neste capítulo é onde esboçamos o modelo de qualidade e a lista de características a recolher, descrevemos alguns indicadores quantitativos e definimos as hipóteses a ensaiar na parte experimental.

3 Modelo de qualidade para presenças na web

Vários modelos de qualidade para presenças na Web são propostos na literatura, como se pode ler no capítulo do trabalho relacionado, cada um com o seu conjunto de características e sub-características (ou indicadores ou métricas). Contudo, a maior parte das características e os próprios modelos não são apropriados para recolha e classificação automática (uma excepção, embora não completa, é o trabalho de Signori). Os modelos propostos prevêm a classificação de uma grande parte das características e seus indicadores, por humanos, sejam utilizadores comuns ou peritos.

Como se pode depreender, este tipo de modelos não são reproduzíveis, pois é impossível ter sempre o mesmo painel de analisadores. E mesmo se fosse possível, estes não iriam analisar uma presença na web sempre da mesma forma, pois a subjectividade do factor humano iria influenciar os resultados.

Assim, precisamos de um modelo que contenha características/indicadores que sejam quantitativos, de modo que a sua recolha e avaliação seja objectiva e não subjectiva. Ou seja, se os indicadores forem quantitativos, podemos obter resultados objectivos e reproduzíveis.

O caminho que nos parece o mais lógico para a obtenção destes indicadores, será começar por fazer uma análise e colecção dos indicadores propostos pela literatura estudada, adicionando os que porventura falem para uma análise mais abrangente. Em seguida, dividiremos estes indicadores em categorias ou características, que irão formar o modelo de qualidade.

3.1 Modelo de qualidade

Podemos encontrar na literatura várias interpretações do que deve ser um modelo de qualidade. Alguns investigadores usam as chamadas características web, que são a navegação, conteúdo e apresentação. Outros investigadores usam as divisões da ISO/IEC 9126 [26] como base para o seu trabalho. Há investigadores que tentam combinar estas categorias, formando um cubo de categorias [23]. Outros investigadores ainda, agrupam os indicadores em características ou categorias diferentes destas últimas, e que têm a ver com o modelo proposto por eles.

O modelo de qualidade que vamos construir, vai ser baseado na ISO9126, tomando as contribuições dos artigos atrás descritos. Porque a ISO9126? A Engenharia de Software tem vários pontos em comum com a Engenharia Web, mas também muitos pontos diferentes. Contudo, existe já uma certa familiaridade na divisão em categorias da ISO9126, o que certamente beneficiará este modelo. Adicionalmente, ao revermos a lista de categorias passíveis de automação, vimos que quase todas se enquadravam nesta definição, com poucas excepções, de que falaremos à frente.

Assim, tomamos esta norma como ponto de partida, acrescentando ou retirando pontos conforme se adequem, enriquecendo o modelo com as contribuições dos autores estudados, e por fim acrescentando as nossas próprias contribuições. Como o objectivo desta dissertação é construir um modelo automatizado, haverá características que fariam sentido mas que terão de ser retiradas devido à sua impossibilidade de automatização, ou porque não temos acesso ao código fonte.

Com efeito, como o objectivo é construir um modelo completamente automatizado de indicadores de qualidade de sítios web, tanto na recolha como na sua quantificação, só temos acesso a elementos estáticos e aos artefactos que resultam dos programas que são executados no servidor, e que são compostos maioritariamente por html, folhas de estilo, *javascript* e componentes multimédia. A análise da programação no servidor em C#, PHP, PERL, JAVA, ou outra linguagem *server-side* estão fora das nossas possibilidades de análise automática. Também se encontram fora das nossas possibilidades de análise o tipo de servidor usado, o SGBD usado, e o hardware.

Como nota, podemos separar o lado cliente do lado servidor, devido à *www* possuir claramente uma arquitectura cliente-servidor. O cliente é o browser e interpreta,

executa ou faz *rendering* de html, media, *scripts* e formatação. Os *browsers* também são diferentes entre eles ao fazer estas operações, mas esta avaliação está fora do nosso escopo. No lado do servidor, estão localizados todos os componentes e operações que decorrem antes da resposta ser devolvida ao *browser*, isto é, o web server, o SGBD e base de dados, e os *scripts* ou programas de servidor que processam ou executam o output em html. A parte do servidor é a parte a que não temos acesso.

A ISO9126 define um modelo de qualidade que contém qualidade externa, qualidade interna e qualidade em uso. O seu modelo para a qualidade interna e externa é o mostrado na figura 3-1.

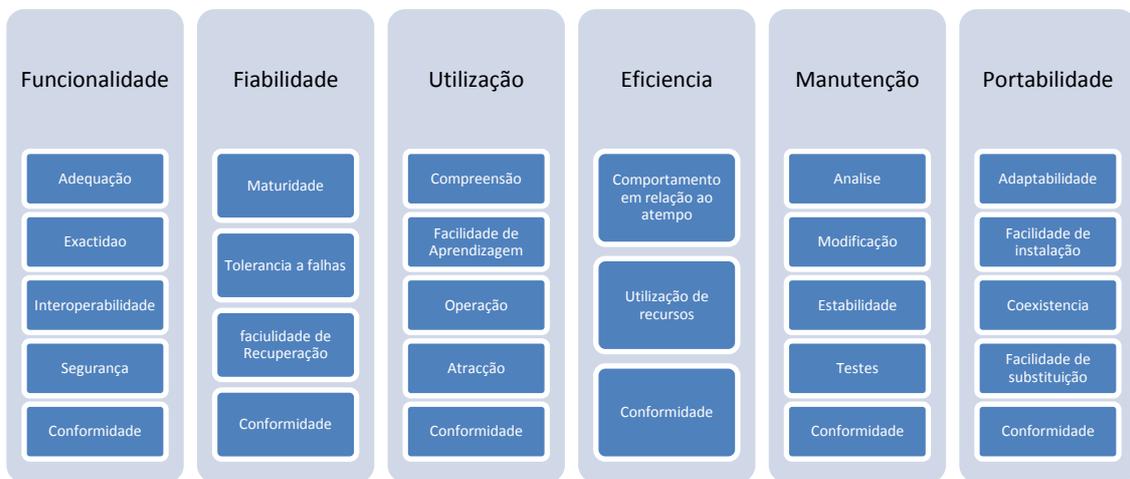


Figura 3-1 Modelo de qualidade interna e externa da ISO 9126

O modelo para a qualidade em uso é o mostrado na figura 3-2.



Figura 3-2 Modelo para qualidade em uso da ISO9126

Como vamos usar um processo automático, vamos considerar somente a qualidade externa, pois será a única vertente a que teremos acesso com uma ferramenta de recolha automática. Assim, o nosso modelo de qualidade englobará as citadas características de qualidade da ISO9126.

3.2 Lista de características

O nosso primeiro passo foi a definição de um modelo automático de classificação para a qualidade dos sítios, onde cada atributo de qualidade é descrito em termos de indicadores (métricas) passíveis de recolha automática. Tentámos não reinventar a roda, mas em vez disso olhar para as longas listas de métricas de qualidade web presentes na literatura, e seleccionar as passíveis de recolha automática. Após este passo, ainda considerámos alguns indicadores adicionais omitidos na literatura. O modelo de qualidade web resultante foi organizado como as características de topo da ISO9126, como se segue:

Eficiência (E) - inclui aspectos relacionados com o tamanho e tempo de carregamento.

Funcionalidade (F) - inclui navegação, formulários, identidade e outros aspectos relacionados com a funcionalidade oferecida pelo sítio.

Manutenção (M) - inclui aspectos relacionados com os itens a manter (*scripts*, folhas de estilo, tables)

Portabilidade (P) - inclui aspectos relacionados com o layout das páginas, o uso de normas html, etc

Fiabilidade (R) – inclui aspectos relacionados com a validação e o estado das ligações

Usabilidade (U) – inclui aspectos relacionados com a acessibilidade, multimédia e conteúdos de texto

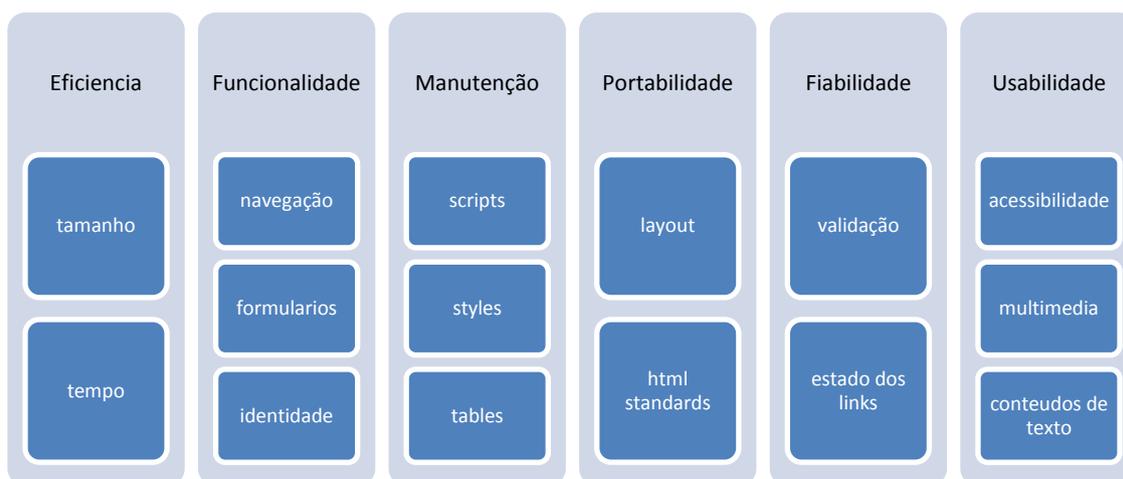


Figura 3-3 Modelo de qualidade proposto para sítios web

Identificámos mais de 60 indicadores de qualidade, que podemos recolher automaticamente, e que estão descritos na tabela 3-1.

	Indicador de qualidade de sítios web	Significado
E	efficiency_css_size	tamanho do css por pagina
E	efficiency_homepage_load_time	tempo de carregamento da homepage
E	efficiency_img_size [22]	Tamanho da imagem
E	efficiency_javascript_size [22]	Tamanho <i>scripts</i> por pagina
E	efficiency_page_load_time [18]	load time da pagina
E	efficiency_page_size	Tamanho pagina
F	forms_form_info_request [14],[20]	Presença de formulário de contactos
F	forms_labels	Numero de <i>tags</i> de label
F	identity_author [18]	Presença de indentificação de autor
F	identity_logo [18]	Presença do logo
F	identity_sitename_title [20]	Presença do nome do sitio no titulo
F	navigation_bar [22]	Presença de barra de navegação
F	navigation_breadcrumbs [22]	Presença de breadcrumbs (indicador path)
F	navigation_quality_of_links [18]	Presença de titulo da pagina no ligação
M	manutencao_num_scripts	Número de <i>scripts</i> por página
M	manutencao_num_styles	Número de css por página
M	manutencao_num_tables [23]	Número de tabelas por página
P	pagelayout_device_specific [22]	Presença de css especificos para dispositivos
P	pagelayout_html_standards	Uso de notação html na formatação
P	pagelayout_num_divs [22]	Número de divs
P	pagelayout_num_frames [23]	Número de frames
P	pagelayout_num_tables [22]	Número de tables
P	pagelayout_num_tables_inside_tables	Presença de tabelas dentro de tabelas
R	links_average_num_words [23]	Media do numero de palavras nos ligações
R	links_links_title [14]	ligações com atributo title
R	links_num_broken_links[14], [22]	Número de ligações com erros
R	links_num_extern_broken_links [22]	Número de ligações com erros para outros sítios
R	links_num_extern_links [22]	Número de ligações para outros sítios
R	links_num_image_links [23]	Número de ligações com imagens
R	links_num_intern_broken_links [22]	Número de ligações internas partidas
R	links_num_intern_links [22]	Numero de ligações internas
R	links_num_links [14], [22]	Número de ligações
R	links_num_non_implemented_links[14]	Número de ligações não implementadas
R	links_page_without_links [23]	Pagina sem ligações no sitio
R	validation_errors [22]	Erros de html por página
R	validation_warnings	Avisos de html por página
U	accessibility_img_alt [22]	Presença do atributo alt nas imagens
U	accessibility_img_title [23]	Presença do atributo title nas imagens
U	accessibility_validate_access [22],[18],[21]	Problemas de acessibilidade por página
U	multimedia_num_img [22]	Número de imagens por pagina
U	text_font_size_average_em	Media do tamanho da fonte em "em" (<i>percentagem</i>) no css
U	text_font_size_average_px	Media do tamanho da fonte em "pixels" no css
U	text_font_size_max_em	Máximo do tamanho da fonte em "em" (<i>percentagem</i>) no css
U	text_font_size_max_px	Máximo do tamanho da fonte em "pixel" (<i>percentagem</i>) no css
U	text_font_size_min_em	Mínimo do tamanho da fonte em "em" (<i>percentagem</i>) no css
U	text_font_size_min_px	Mínimo do tamanho da fonte em "pixel" (<i>percentagem</i>) no css
U	text_heading_len [22]	Média do numero de caracteres dos headings
U	text_heading_reverse_order [22]	Número de headings em ordem reversa
U	text_italic_text	Número de textos em italic com mais de 20 caracteres
U	text_num_diferent_colors	Número de cores diferentes no css
U	text_num_diferent_fonts [22]	Números fonts diferentes no css
U	text_num_sentences_in_paragraph [22]	Número frases por paragrafo
U	text_num_subheading_heading [22]	Número de subheadings por heading
U	text_num_syllables_in_word [22]	Número de sílabas por palavra
U	text_num_words_in_sentence [22]	Número de palavras por frase
U	text_num_words_meta_description	Número de palavras na metatag description
U	text_num_words_meta_keywords	Número de palavras na metatag keywords
U	text_paragraph_max_size [22]	Tamanho máximo do parágrafo
U	text_paragraph_size [22]	Tamanho do parágrafo
U	text_subheading_len [22]	Tamanho de subheading
U	text_total_newlines [22]	Número total de newlines
U	text_total_sentences [22]	Total de frases
U	text_total_syllables [22]	Total de sílabas
U	text_total_words [22]	Total de palavras
U	text_uppercase_text	Número de frases em maiúsculas

Tabela 3-1 Indicadores de qualidade passíveis de recolha automática

3.3 Proposta de ontologia para avaliação

A nossa proposta de ontologia para avaliar sítios web automaticamente está relacionada com o nível de decomposição que um sítio web pode ter, mas não indo tão longe que torne ambígua a sua definição, pelo menos por meios automáticos [22]. Assim, ao fazer uma ontologia que permita medir automaticamente, deve-se decompor só até ao nível em que tenhamos a certeza que o componente é exactamente aquilo que está definido. O que acabamos de dizer está relacionado com a simplicidade da definição do html, no presente, o html 4.01 ou o xhtml 1.0 [27, 28]. Com efeito, é difícil de, automaticamente, confirmar sem erros se um determinado componente é um menu ou se determinado ligação faz parte da navegação. Com a entrada do html 5, que vai substituir simultaneamente o html 4 e o xhtml (tornado obsoleto), e que define algumas etiquetas adicionais, este problema fica resolvido [29].

Deste modo, podemos dizer que um sítio contém páginas, e uma delas é a página de entrada (*homepage*). As páginas incluem *scripts* e folhas de estilo sendo que estes podem ser repetidos e incluídos por várias páginas. As páginas incluem elementos multimédia, tais como imagens e outros. Na presente versão de html, não devemos ir além destas divisões. Numa versão futura, o html 5, podemos dividir as páginas em mais componentes, usando por exemplo as novas `<menu>`, `<nav>`, `<video>`, `<audio>`, etc.

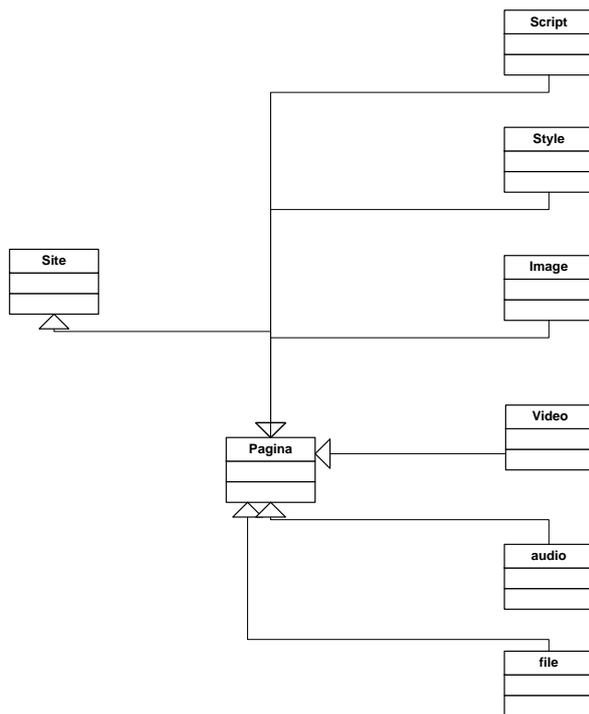


Figura 3-4 Ontologia para recolha de dados para modelo de qualidade proposto

Torna-se necessário dividir os elementos multimédia em classes separadas, pois estes têm atributos diferentes. Por exemplo, os vídeos e imagens têm largura e altura, o vídeo e o áudio tem duração, o vídeo e o áudio tem *codecs* de áudio, o vídeo tem *codecs* de vídeo, etc. Quanto aos *scripts* e folhas de estilo, estes são incluídos numa página, mas muitas vezes repetem-se bastante, sendo os mesmos ao longo do sítio, ou pelo menos parte deles. Assim, também faria sentido uma ligação com o sítio, contudo ficam ligados com a página pois é a sua ligação real.

Esta ontologia vai permitir a separação dos elementos do sítio web a serem analisados, para a lista de características a recolher. Dado o volume enorme de dados que teremos de tratar, optamos por usar uma base de dados relacional.

3.4 Descrição de indicadores quantitativos

Neste ponto vamos descrever nosso modelo de qualidade com mais algum detalhe, nomeadamente os indicadores quantitativos e seu modo de cálculo. Como vimos os indicadores estão devidos em 6 macro-características, sendo que alguns indicadores são booleanos, outros números, e alguns indicadores a sua granularidade é o sítio, outros a página.

Todas as características são calculadas por página e depois é feito a média excepto as indicadas a seguir, em que é feita a soma: *form_info_request*, *text_total_words*, *text_total_syllables*, *text_total_sentences*, *text_total_newlines* e a contagem no indicador *links_page_without_links*

As características seguintes são calculadas por sitio: *text_num_diferent_fonts*, *text_font_size* (*avegare*, *Max*, *min*, *px*, *em*), *text_num_diferent_colors*.

Páginas diferentes podem ter número de *scripts* ou *css* diferente por isso estes são por página. Os tamanhos representados na tabela 4.2 são em bytes, e o tempo é representado em segundos.

Característica	Indicador de qualidade de sítios web	Significado	escala
Eficiência	efficiency_css_size	tamanho do css por pagina	number
Eficiência	efficiency_homepage_load_time	tempo de carregamento da homepage	number
Eficiência	efficiency_img_size	Tamanho da imagem	number
Eficiência	efficiency_javascript_size	Tamanho <i>scripts</i> por pagina	number
Eficiência	efficiency_page_load_time	load time da pagina	number
Eficiência	efficiency_page_size	Tamanho pagina	number
Funcionalidade	forms_form_info_request	Presença de formulário de contactos	0-1
Funcionalidade	forms_labels	Numero de <i>tags</i> de label	number
Funcionalidade	identity_author	Presença de indentificação de autor	0-1
Funcionalidade	identity_logo	Presença do logo	0-1
Funcionalidade	identity_sitename_title	Presença do nome do sitio no titulo	0-1
Funcionalidade	navigation_bar	Presença de barra de navegação	0-1
Funcionalidade	navigation_breadcrumbs	Presença de breadcrumbs (indicador path)	0-1
Funcionalidade	navigation_quality_of_links	Presença de titulo da pagina no ligação	number
Manutenção	manutencao_num_scripts	Número de <i>scripts</i> por página	number
Manutenção	manutencao_num_styles	Número de <i>css</i> por página	number
Manutenção	manutencao_num_tables	Número de tabelas por página	number
Portabilidade	pagelayout_device_specific	Presença de <i>css</i> especificos para dispositivos	number
Portabilidade	pagelayout_html_standards	Uso de notação html na formatação	0-1
Portabilidade	pagelayout_num_divs	Número de <i>divs</i>	number
Portabilidade	pagelayout_num_frames	Número de frames	number
Portabilidade	pagelayout_num_tables	Número de tables	number
Portabilidade	pagelayout_num_tables_inside_tables	Presença de tabelas dentro de tabelas	number
Fiabilidade	links_average_num_words	Media do numero de palavras nos ligações	number
Fiabilidade	links_links_title	ligações com atributo title	number
Fiabilidade	links_num_broken_links	Número de ligações com erros	number
Fiabilidade	links_num_extern_broken_links	Número de ligações com erros para outros sítios	number
Fiabilidade	links_num_extern_links	Número de ligações para outros sítios	number
Fiabilidade	links_num_image_links	Número de ligações com imagens	number
Fiabilidade	links_num_intern_broken_links	Numero de ligações internos partidos	number
Fiabilidade	links_num_intern_links	Numero de ligações internos	number
Fiabilidade	links_num_links	Número de ligações	number
Fiabilidade	links_num_non_implemented_links	Número de ligações não implementadas	number
Fiabilidade	links_page_without_links	Pagina sem ligações no sítio	number
Fiabilidade	validation_errors	Erros de html por pagina	number

Característica	Indicador de qualidade de sítios web	Significado	escala
Fiabilidade	validation_warnings	Avisos de html por pagina	number
Usabilidade	accessibility_img_alt	Presença do atributo alt nas imagens	0-1
Usabilidade	accessibility_img_title	Presença do atributo title nas imagens	0-1
Usabilidade	accessibility_validate_access	Problemas de acessibilidade por página	number
Usabilidade	multimedia_num_img	Número de imagens por pagina	number
Usabilidade	text_font_size_average_em	Media do tamanho da fonte em "em" (percentagem) no css	number
Usabilidade	text_font_size_average_px	Media do tamanho da fonte em "pixels" no css	number
Usabilidade	text_font_size_max_em	Máximo do tamanho da fonte em "em" (percentagem) no css	number
Usabilidade	text_font_size_max_px	Máximo do tamanho da fonte em "pixel" (percentagem) no css	number
Usabilidade	text_font_size_min_em	Mínimo do tamanho da fonte em "em" (percentagem) no css	number
Usabilidade	text_font_size_min_px	Mínimo do tamanho da fonte em "pixel" (percentagem) no css	number
Usabilidade	text_heading_len [20]	Média do numero de caracteres dos headings	number
Usabilidade	text_heading_reverse_order	Número de headings em ordem reversa	number
Usabilidade	text_italic_text	Número de textos em italic com mais de 20 caracteres	number
Usabilidade	text_num_diferent_colors	Número de cores diferentes no css	number
Usabilidade	text_num_diferent_fonts	Números fonts diferentes no css	number
Usabilidade	text_num_sentences_in_paragraph	Número frases por paragrafo	number
Usabilidade	text_num_subheading_heading	Número de subheadings por heading	number
Usabilidade	text_num_syllables_in_word	Número de sílabas por palavra	number
Usabilidade	text_num_words_in_sentence	Número de palavras por frase	number
Usabilidade	text_num_words_meta_description	Número de palavras na metatag description	number
Usabilidade	text_num_words_meta_keywords	Número de palavras na metatag keywords	number
Usabilidade	text_paragraph_max_size	Tamanho máximo do parágrafo	number
Usabilidade	text_paragraph_size	Tamanho do parágrafo	number
Usabilidade	text_subheading_len	Tamanho de subheading	number
Usabilidade	text_total_newlines	Número total de newlines	number
Usabilidade	text_total_sentences	Total de frases	number
Usabilidade	text_total_syllables	Total de sílabas	number
Usabilidade	text_total_words	Total de palavras	number
Usabilidade	text_uppercase_text	Número de frases em maiúsculas	number

Tabela 3-2 Tabela de definição dos indicadores

Como exemplo mostramos o algoritmo para cálculo de dois indicadores.

Nome:	Significado:	Tipo de dados:
efficiency_css_size	Número de css por página	float
<p>Algoritmo:</p> <p>Contadores a zero (tamanho total, numero css)</p> <p>Para todas as páginas do sítio</p> <p style="padding-left: 40px;">Retirar código da Base Dados</p> <p style="padding-left: 40px;">Para todos os ficheiros css</p> <p style="padding-left: 80px;">Contador = contador + 1</p> <p style="padding-left: 80px;">Tamanho total = tamanho total + tamanho css</p> <p style="padding-left: 40px;">Guardar na tabela página características</p> <p>Média do sítio= tamanho total / contador</p> <p>Guardar na tabela sítio características</p>		
<p>Recomendações: menor é melhor, pois aumenta a rapidez.</p>		

Tabela 3-3 Definição do indicador efficiency_css_size

Nome: links_num_extern_links	Significado: Numero ligações externas por página	Tipo de dados: float
<p>Algoritmo:</p> <p>Contadores a zero (numero)</p> <p>Para todas as páginas do sítio</p> <p style="padding-left: 40px;">Retirar código da Base Dados</p> <p style="padding-left: 40px;">Encontrar tag '<a>'</p> <p style="padding-left: 80px;">Se atributo href conter 'http://'</p> <p style="padding-left: 120px;">Ligação= conteúdo de href</p> <p style="padding-left: 120px;">Se ligação não contiver o <hostname do sítio></p> <p style="padding-left: 160px;">Contador = contador + 1</p> <p style="padding-left: 40px;">Guardar na tabela página características</p> <p>Média do sítio= tamanho total / contador</p> <p>Guardar na tabela sítio características</p>		
Recomendações: menor é melhor, pois aumenta a rapidez.		

Tabela 3-4 Definição do indicador links_num_extern_links

3.5 Hipóteses a ensaiar

Este trabalho divide-se em duas partes, uma onde se tenta criar um modelo de qualidade, com um conjunto de indicadores/características, agrupadas de acordo com a ISO9126, e é desenvolvida uma ferramenta que permite extrair estes indicadores automaticamente, guarda-los numa base de dados e exporta-los para Excel. Na segunda parte do trabalho tentamos fazer algum trabalho experimental, nomeadamente a nível da relação dos indicadores com a sua área de aplicação, e com o PageRank do Google. A razão de compararmos com o PageRank está descrita mais à frente. Assim, as nossas hipóteses são as seguintes:

- É possível fazer uma recolha automática de indicadores, e com isso classificar automaticamente um sítio web quanto à sua qualidade? Este ponto divide-se em duas partes, uma recolha e análise de indicadores e a outra a comprovação que o conjunto de indicadores constitui de facto uma classificação de um sítio web quanto à qualidade.
- Os descritores das características de qualidade são sensíveis ao domínio da aplicação web? Analisaremos os indicadores retirados e concluiremos sobre a sua dependência ou não do domínio da aplicação (Jornais, Bancos, Aviação).
- É possível construir um modelo de previsão do PageRank usado no Google com base nos nossos indicadores? Tentaremos relacionar o PageRank com os nossos indicadores.

No capítulo seguinte tentaremos provar ou não cada uma destas hipóteses, recorrendo às ferramentas da Eng. de software Experimental, nomeadamente a análise estatística.

Capítulo 4

Procedimento Experimental

4	Procedimento Experimental	54
4.1	Amostra.....	54
4.2	Planeamento da experiencia.....	56
4.3	Recolha de dados	56
4.4	Descrição do ambiente de pesquisa e do Programa de recolha de dados	60
4.5	Avaliação dos dados, ou recolha de características indirectas.....	60
4.6	Estatísticas do processamento.....	62

No capítulo 4 mostramos como fizemos o nosso procedimento experimental. Começamos por descrever a amostra, depois o planeamento da experiencia, a recolha de dados, a descrição do ambiente, e por fim, a fase de analise de dados.

4 Procedimento Experimental

4.1 Amostra

Escolhemos 3 domínios de aplicação que tem uma lógica de negócio bastante bem definida, e estão espalhados pelo mundo inteiro: *Jornais, Bancos e Aviação*. Para cada um destes domínios seleccionamos mais de 30 sítios, divididos por áreas geograficamente distantes e representativas do universo. Contudo, em alguns sítios, tivemos de seleccionar apenas a parte inglesa do sítio, pois algumas características só podiam ser recolhidas nesta língua, ou pelos menos, com um alfabeto ocidental. O tamanho da amostra tem 111 sítios, embora no inicio tivesse cerca de 150, no decurso de 2009/2010 alguns deixaram de existir (especialmente jornais) e outros juntaram-se (aviação). Também eliminamos alguns valores que não faziam sentido.

Area	Países
América - Norte e GB	UK, USA, Canada
América - Sul	Brazil, Argentina, Venezuela
Europa Central	França, Alemanha, Belgica
Europa - Sul	Portugal, Espanha, Itália
Europa - Norte	Finlândia, Suécia, Dinamarca, Noruega
Ásia	India, Singapura, Hong Kong, Coreia do Sul, Japão
África - Sul	África do Sul, Nigéria, Angola
África - Norte	Egipto, Marrocos, Tunísia, ...

Tabela 4-1 Áreas e Países da amostra

Os jornais foram escolhidos pela sua tiragem e número de páginas vistas. Os bancos pelo número de clientes e capitalização. As agências de viagens pelo número de passageiros e capitalização. Alguns destes dados, como os dos jornais, encontram-se online em sítios especializados, outros para países menos desenvolvidos, tivemos de ir por meios menos credíveis, quanto à sua tiragem. A informação da capitalização é mais fácil de encontrar, e serviu-nos de base para escolhermos os sítios de aviação e bancos.

Jornais		Bancos		Companhias de Aviação	
País	Nome	País	Nome	País	Nome
USA	USA Today	USA	Bank of America	USA	Southwest Airlines
USA	The Wall Street Journal	USA	JPMorgan Chase	USA	American Airlines
USA	The New York Times	USA	Citigroup	USA	Delta Air Lines
USA	Los Angeles Times	USA	Wells Fargo	USA	United Airlines
UK	The Sun	UK	HSBC Holdings	Canada	Air Canada
UK	Daily Mail	UK	RBS	Irlanda	Ryanair
UK	Daily Mirror	UK	Standard Chartered	UK	easyJet
UK	The Daily Telegraph	UK	Barclays	UK	British Airways
UK		UK		UK	BMI
Brasil	O Estado de São Paulo	Brasil	Itau	Brasil	TAM Airlines
Brasil	Jornal do Brasil	Brasil	BB	Brasil	Gol Transportes Aéreos
Brasil	O Globo	Brasil	Bradesco	Brasil	Varig
Argentina	La Nación	Argentina	Banco de Galicia y Buenos Aires	Argentina	Aerolíneas Argentinas
Venezuela	El Nacional	Venezuela	Mercantil Banco Universal	Colombia	Avianca
Venezuela	El Universal	Venezuela	Banco Exterior	Chile	LAN Airlines
Venezuela	Diario 2001	Venezuela	Banco Federal		
Alemanha	Bild	Alemanha	Commerzbank	França	Air France
Alemanha	Süddeutsche Zeitung	Alemanha	Postbank	Holanda	KLM
Alemanha	Die Zeit	Alemanha	Deutsche Bank	Alemanha	Lufthansa
França	Ouest-France	França	BNP Paribas	Alemanha	Air Berlin
França	La Tribune	França	Société Générale	Suíça	Swiss International Air Lines
França	Le Parisien	França	Credite Agricole	Belgica	Brussels Airlines
Bélgica	Gazet van Antwerpen	Bélgica	KBC Groep	Espanha	Iberia Airlines
Bélgica	Grenz-Echo	Bélgica	Dexia	Itália	Alitalia
Bélgica	Le Soir en ligne	Bélgica	Fortis	Espanha	Air Europa
Holanda	De Telegraaf	Portugal	Banco BPI	Portugal	TAP Portugal
Austria	Kronen Zeitung	Portugal	BCEP	Itália	Air One
Portugal	Diario de Noticias	Portugal	BES	Espanha	Vueling Airlines
Portugal	Jornal de Noticias	Espanha	Banco Popular	Dln./Su/Noa	SAS Group
Portugal	O Público	Espanha	BBVA	Noruega	Norwegian Air Shuttle
Espanha	El País	Espanha	Santader	Finlândia	Finnair
Espanha	El Mundo	Italia	Cassa di Risparmio di Firenze	China	China Southern Airlines
Espanha	ABC	Noruega	Sparebanken NOR	China	China Eastern Airlines
Italia	Corriere della Sera	Noruega	Norges Bank	China	Air China
Italia	La Repubblica	Dinamarca	Amagerbanken	Japão	Japan Airlines
Noruega	Aftenposten	Dinamarca	Jyske Bank Private Banking	Japão	All Nippon Airways
Noruega	Dagbladet	Suécia	Swedbank	Singapura	Singapore Airlines
Dinamarca	Árhus Stiftstidende	China	ICBC	África do Sul	South African Airways
Dinamarca	Børsen	China	China Construction Bank of China	Angola	TAAG Angola Airlines
Suécia	Expressen	China	Bank of China	Kenia	Kenya Airways
China	People's Daily	Japan	Mitsubishi UFJ Fnl, MUFG (Japan)	Etiopia	Ethiopian Airlines
China	Guangzhou Daily	Angola	Banco Fomento de Angola (BFA)	Egipto	EgyptAir
China	Global Times	Nigeria	Zenith Bank	Marrocos	Royal Air Maroc
Japan	Yomiuri Shimbun	Nigeria	UBA United Bank for Africa	Tunisia	Tunisair
Japan	Asahi Shimbun	Nigeria	Ecobank	Marrocos	Atlas Blue
Japan	Mainichi Shimbun	África do Sul	The Standard Bank of South Africa		
India	Anandabazar Patrika	África do Sul	ABSA		
India	The Times of India	Egipto	National Bank Of Egypt		
India	Malayala Manorama	Egipto	Bank of Alexandria		
Angola	Jornal de Angola	Marrocos	Bank Al-Maghrib		
Nigeria	Daily Champion	Marrocos	Attijariwafa Bank		
Nigeria	Nigerian Tribune	Marrocos	Banque Centrale Populaire		
Nigeria	This Day				
África do Sul	The citizen				
África do Sul	Daily News				
Egipto	Al-Ahram				
Egipto	Al Gomhuriya				
Tunisia	La Presse				
Tunisia	Le quotidien				

Tabela 4-2 lista de sítios da amostra

Se o sítio não usasse alfabeto ocidental, encontrávamos o ponto de entrada para a versão em inglês do sítio. Este passo é extremamente importante para os sítios asiáticos e do

norte de África. No Anexo A reproduzem-se os nomes, endereços, países e as referências usadas.

4.2 Planeamento da experiência

Após termos a amostra, começamos a planear a experiência. O primeiro passo necessário é a recolha recursiva dos elementos /componentes do sítio, como páginas, *scripts*, folhas de estilo, componentes multimédia, etc. Para este fim vamos usar um *crawler*, ou aproveitando um que já existe ou implementando um. Neste ponto a escolha ainda não tinha sido feita.

A recolha dos indicadores irá ser feita em duas fases, uma para os indicadores directos, que pode ser feita na fase de recolha dos dados, e outra fase para os indicadores indirectos através de análise das páginas e outros componentes. Chamamos à primeira fase a *recolha de dados* e a segunda fase a *classificação dos dados*. A granularidade de alguns indicadores é por sítio (por exemplo presença de formulário de informação), enquanto a granularidade de outros é por página (por exemplo, o numero de imagens). Para estes últimos, aplica-se uma operação de agregação, por exemplo, a média para a maior parte dos casos.

Em seguida, usando um pacote de software de estatística, analisaremos o conjunto de indicadores à luz das nossas hipóteses. No nosso caso escolhemos o SPSS devido à experiência com este software e às suas capacidades de análise.

4.3 Recolha de dados

Neste ponto foi efectuada uma pesquisa pelos *crawlers* existentes, e que fossem código aberto (open source) e livre (que pudéssemos modificar). Encontramos *crawlers* em versões *desktop* (software executável para sistemas operativos, como Windows e Unix) e web. Como estamos a analisar outputs de sítios web, faz todo o sentido que escolhamos um *crawler* que corra via *web server*, até porque podemos disponibilizar mais tarde esta ferramenta online.

A linguagem em que a nossa pesquisa incidiu, pesando o que se disse anteriormente, foi o PHP, devido à sua facilidade de programação, e integração com base de dados. Mas uma das maiores vantagens é que conseguimos ter o mesmo código

praticamente em linha de comando ou se for executado num web server. Efectivamente, só o que temos de mudar entre estes dois ambientes é o código de quebra de linha.

Foram estudados estes dois crawlers para o seu possível aproveitamento:

- PHPCrawl 0.70 (<http://sourceforge.net/projects/phpcrawl>) é uma classe escrita em php para fazer o *crawling* ou *spidering* de sitios web. Suporta filtros, limitadores, cookie e outras funcionalidades. Fazendo a redefinição de um método especial na classe, os utilizadores podem escolher o que fazer aos dados que o crawler encontra.
- **PHP Crawler 0.80** (<http://sourceforge.net/projects/php-crawler/>) é um *script* simples de pesquisa de sítios web, para sítios pequenos a médios. Foi escrito em php e usa mysql para armazenamento de dados.

Após várias experiências com estes dois crawlers e tentativas de fazer um crawler específico para o nosso propósito, implementado também em php, chegamos à conclusão que o primeiro crawler era aproveitável, embora tivesse de ser fortemente modificado, no que toca à persistência de dados, e de recolha de indicadores directos, para por exemplo, obter o tempo de carregamento da página. Modificamos o crawler para usar persistência, ou seja, guardar os dados recolhidos, numa base de dados. O segundo crawler já tinha suporte a base de dados, sendo no entanto mais limitado.

Para o layer de persistência usamos uma base de dados Mysql. Esta base de dados contém o nosso sample, e vai conter também os componentes recolhidos (as páginas, *scripts*, folhas de estilo, elementos multimédia, etc) e os indicadores recolhidos e a sua agregação por sítio. Usando uma base de dados local permite uma maior rapidez no pós-processamento, ou seja, a extracção das características /indicadores a que chamamos classificação.

O modelo de persistência é mostrado na figura 5-1.

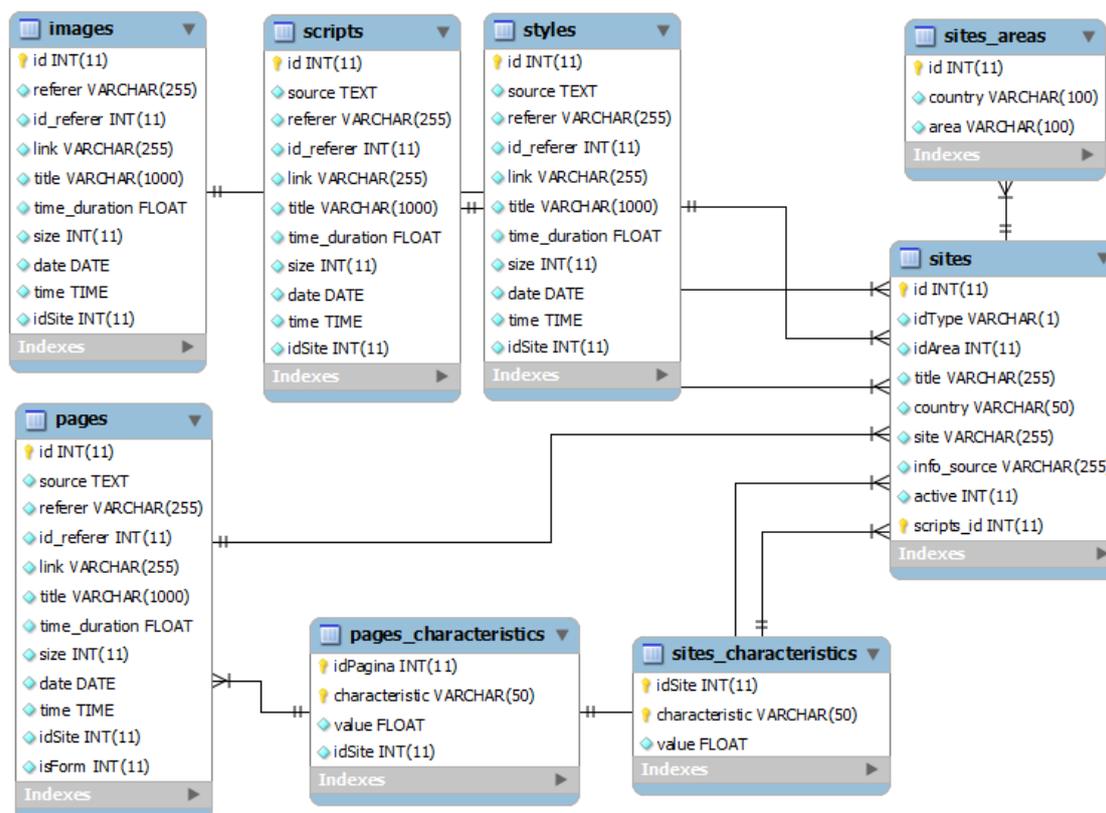


Figura 4-1 Esquema da base de dados.

Podemos dividir esta base de dados em 3 áreas: a primeira, tem as áreas geográficas da amostra e países, bem como os URLs de entrada, no caso de alfabeto ocidental, e de entrada em inglês, no caso de não ter alfabeto ocidental. Esta área é composta pelas tabelas *sites_areas* e *sites*. A segunda área é composta pelas tabelas de recolha de dados, ou sejam, as *pages*, *styles*, *scripts* e *images*. Estas tabelas também contêm alguns indicadores directos. A terceira área é onde se guardam as características obtidas por análise dos componentes. Existem duas tabelas, uma para as características/indicadores obtidos para as páginas e outra para as carecterísticas/indicadores cuja granularidade é o sítio. Temos ainda uma outra tabela, não representada na figura, que agrega todos os indicadores, por sítio, e as áreas, num formato desnormalizado, para exportação em Excel. Ao obtermos este formato em Excel, podemos importar no SPSS.

Optamos por fazer tabelas para as características dos sítios e características das páginas, em que se pudesse facilmente adicionar características. Assim, ficam com poucas colunas mas muitas linhas (registos) o que esperamos que não seja um problema

para o processamento. Com efeito o Mysql com tabelas com muitos registos requer alguma optimização.

Depois de adicionarmos persistência ao *crawler*, redefinindo um método, para guardar os dados em base de dados, modificamos o código do *crawler* para tirar o tempo de *download* e outros indicadores directos, tais como presença de formulários, título da página, etc. Em seguida fizemos um *script* que é chamado inicialmente e que corre o *crawler* num ciclo para todos os sites, para cada elemento (sitio) da nossa amostra. Além disso, define alguns valores por omissão para a classe do *crawler*, tais como ficheiros a seguir e a ignorar, se usa cookies e limites de tráfego. No nosso caso usamos os valores de ficheiros a receber para texto, html, js e css, para ignorar filmes e áudio (precisamos de receber as imagens para as medir) e o limite de tráfego para cada sítio definimos como 5 Megabytes:

```
$crawler->addReceiveContentType("/text/html/js/css");  
$crawler->addNonFollowMatch("(swf|flv|avi|mov|mp3)$/ i");  
$crawler->setCookieHandling(true);  
$crawler->setTrafficLimit(5000 * 1024);
```

Figura 4-2 Valores por omissão que escolhemos

Este ciclo para todos os sites também põe uma *flag* de fim de recolha no registo do sítio, quando acaba a obtenção de dados. Esta *flag* serve para uma possível falha do programa a meio, o que permite recomeçar de onde paramos.

Neste indicadores directos que recolhemos nesta fase, começamos também por recolher todas as classes e ids de elementos html, únicos, e sua contagem. Contudo, após algumas experiencias, concluímos que não era possível categorizar os sub-elementos de uma página com base unicamente nos seus nomes de classes de css e ids, com algumas excepções que utilizámos, como por exemplos para os menus e navegação. Também recolhemos as extensões únicas de ficheiros, mas não chegamos a usar. Este fase do programa durou menos de dois dias.

4.4 Descrição do ambiente de pesquisa e do Programa de recolha de dados

Para correr o programa podemos usar o browser ou linha de comando, em Linux ou Windows, ou ainda em OSX (todavia este último é desaconselhado pois o Mysql corre muito lentamente).

Se usarmos o browser, Precisamos do Apache, PHP, e Mysql, e uma ferramenta gráfica para gerir o Mysql se possível. Em Linux, em muitas distribuições é somente necessário instalar o PHP com mod_php e o Mysql, outras já o trazem. O Apache vem em quase todas as distribuições. Em ambiente Windows, podemos usar um pacote que providencia o Apache, Php, Mysql e phpmyadmin (gestão web do mysql) tal como o Wampserver ou o Xamp.

Depois de configurarmos a base de dados e a popularmos com o nosso sample, corremos o programa com o url no browser : http://<host>/<dir_programa>/index.php. Ou podemos omitir o index.php se tivermos o *server* configurado para tal.

Se optarmos por correr o programa em linha de comando, precisamos do mesmo, menos o apache (será usado o PHP cli - *command line interface*). Temos de adicionar o interpretador de PHP à *path* dos executáveis. Para executar o programa corremos o script run.bat.

No nosso caso usamos uma máquina Windows mas podíamos perfeitamente ter usado uma máquina Linux, e correremos o programa na consola.

4.5 Avaliação dos dados, ou recolha de características indirectas

A avaliação dos dados, como chamamos à segunda fase de recolha de indicadores indirectos, consiste em fazer *data mining* dos componentes guardados na nossa base de dados local, e destes extrair os outros indicadores, que não são directos. Para isso fizemos um programa, também em PHP, e que em consola é executado com a *script* evaluate.bat, que chama o ficheiro evaluate.php e este por sua vez analisa código de cada sítio, para as 60+ características. Esta parte do programa foi desenvolvida completamente de raiz, pois era muito específica e não existe muito trabalho nesta área, a não ser para a validação, erros e warnings. É fornecido um ficheiro comprimido com o código completo num CD anexo a esta dissertação e em apêndice pode-se encontrar uma descrição mais pormenorizada.

Nesta fase de análise de dados tivemos o problema de, ao limitarmos o número de componentes a um número fixo em bytes, por vezes os componentes das páginas que eram outros ficheiros, por exemplo os *scripts*, folhas de estilo e imagens, não se encontravam na base de dados. Assim, tivemos de providenciar este programa de um mecanismo que tirasse as URLs em falta, fossem elas código ou componentes multimédia. Aqui chegou-se a pensar que seria melhor construir um crawler de raiz, uma vez que grande parte do trabalho iria ficar duplicado, e já tínhamos feitos testes em código, mas por uma questão de tempo, ficou com esta correcção para ficheiros em falta, nesta segunda fase.

A segunda fase usa as mesmas tabelas de base de dados da anterior, ou seja as características por páginas são guardadas na tabela de características de páginas e depois é feita uma operação de agregação, normalmente média ou soma, estas são guardadas na tabela de características dos sítios. As características de sítios são guardadas directamente na tabela de sítios.

Por fim, na última fase do programa, exportamos as características para a tabela desnormalizada, onde ficamos com um registo por sítio. Além, das características, também incluímos o tipo de sítio e as áreas e os países.

Como tínhamos 111 casos na nossa amostra, mas limitamos a recolha na primeira fase a 5 mega bytes por sítio, obtivemos cerca de 700K registos na tabela de características de páginas e pouco mais de 9K registo na tabela de características de sítios. Estes valores foram depois passados para a tabela de exportação com cada característica numa coluna e cada linha um sítio, a fim de exportarmos para Excel, e por fim importarmos no SPSS.

Esta fase do programa demorou mais de uma semana, dividida por dois computadores, e esta lentidão deve-se ao acesso à base de dados. O Mysql com tantos registos precisa de mais memória ram alocada para ele e da optimização de alguns parâmetros, para funcionar com maior rapidez. Contudo, como estávamos a meio do programa, ao parar perderíamos mais tempo ainda.

No final da recolha de indicadores, ainda tivemos de apagar alguns sítios, pois estes não tinham valores. Devido ao ano instável para a economia, alguns sítios tinham deixado de existir, outros tinham mudado de endereço, e outros ainda não davam resultados apropriados.

4.6 Estatísticas do processamento

A base de dados utilizada ficou com os seguintes valores, depois das duas fases.

Tabela	Registos	Tamanho
imagens	162,103	24.0 MB
paginas	16,749	432.0 MB
paginas_caracteristicas	689,661	63.0 MB
paginas_links	0	1.0 KB
rel_class_page	96,51	9.9 MB
rel_id_page	248,533	25.4 MB
scripts	7,441	34.8 MB
sites	153	23.2 KB
sites_area	8	2.4 KB
sites_caracteristicas	9,164	823.4 KB
sites_caracteristicas_export	152	43.8 KB
styles	5,309	30.6 MB
unique_class	2,12	70.2 KB
unique_ext	0	1.0 KB
unique_id	18,251	709.3 KB
15 tabela(s) - Soma:	1,256,154	621.4 MB

Tabela 4-3 Estatísticas da base de dados

Como exemplo damos também algumas estatísticas do programa desenvolvido.

caracteristicas	descrição	linhas
acessibilidade	funções de acessibilidade	127
eficiencia	funções de eficiencia	365
forms	funções de formulários	116
identity	funções de texto	142
links	funções de ligações	415
manutencao	funções de manutenção	121
multimedia	funções de multimédia	99
navegacao	funções de navegação	190
page_layout	funções de layout	230
text	funções de texto num. 1	925
text2	funções de texto num. 2	65
utils	utilitários, somas copias, etc	368

Tabela 4-4 estatísticas das funções que extraem os indicadores

O tempo de recolha, fase1, foi de dois dias e da fase 2, análise de indicadores, de mais de uma semana repartido por dois computadores (embora o segundo a 50%).

Capítulo 5

Análise de dados

5	Análise de dados.....	64
5.1	Análise de componentes principais.....	64
5.2	Análise da variância.....	66
5.3	Modelo de qualidade.....	73
5.3.1	Exemplo de limites para indicadores.....	75
5.4	Relação das características de qualidade com o PageRank do Google	77
5.4.1	Funcionamento do page rank.....	77
5.4.2	Recolha do PageRank.....	78
5.4.3	Análise estatística	79
5.4.4	Regressão Ordinal	79
5.4.5	Aplicação da regressão ordinal.....	81
5.4.6	Redução de variáveis	83
5.4.7	Regressão ordinal com conjunto reduzido de variáveis	85
5.4.8	Regressão ordinal dependente do domínio.....	87
5.4.9	Modelo de predição de probabilidade de ser um determinado PageRank	91

O capítulo 5 é o capítulo da análise dos dados. Neste capítulo fazemos apresentamos a análise estatística realizada com o auxílio do SPSS, para obter respostas às nossas hipóteses de investigação.

5 Análise de dados

Neste capítulo pretende-se ensaiar um conjunto de hipóteses, que serão validadas (ou refutadas) estatisticamente com base numa amostra significativa de casos reais. Este capítulo será desenvolvido com utilização do pacote SPSS de análise estatística.

5.1 Análise de componentes principais

Depois de importar os dados para o Spss, começamos com alguma análise de dados exploratória usando Análise de Componentes Principais. Este método não paramétrico permite-nos extrair informação relevante de grandes quantidades de dados, e que possam ser confusos. O ACP providencia-nos um caminho para a tentativa de redução de um conjunto de dados complexo, e com muitas variáveis, para um conjunto de dados com uma dimensão menor e pode revelar a dinâmica escondida e simplificada que está por detrás dos dados.

Como temos um número consideravelmente grande de métricas, queremos analisar se a sua variabilidade pode ser explicada por um número mais pequeno de variáveis, chamados componentes em ACP. Frequentemente, poucas variáveis podem explicar uma grande percentagem dessa variabilidade, especialmente quando a matriz de correlação evidencia a presença de multicolinearidade.

Na tabela 6-1 podemos ver estes valores e o seu cumulativo.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9,414	14,483	14,483	9,414	14,483	14,483
2	5,596	8,609	23,092	5,596	8,609	23,092
3	4,348	6,690	29,781	4,348	6,690	29,781
4	3,729	5,737	35,518	3,729	5,737	35,518
5	2,868	4,412	39,930	2,868	4,412	39,930
6	2,445	3,762	43,692	2,445	3,762	43,692
7	2,270	3,493	47,185	2,270	3,493	47,185
8	2,082	3,204	50,388	2,082	3,204	50,388
9	1,909	2,937	53,326	1,909	2,937	53,326
10	1,785	2,746	56,072	1,785	2,746	56,072
11	1,681	2,587	58,659	1,681	2,587	58,659
12	1,654	2,544	61,203	1,654	2,544	61,203
13	1,533	2,358	63,561	1,533	2,358	63,561
14	1,469	2,260	65,821	1,469	2,260	65,821
15	1,428	2,196	68,018	1,428	2,196	68,018
16	1,332	2,049	70,066	1,332	2,049	70,066
17	1,291	1,986	72,052	1,291	1,986	72,052
18	1,196	1,840	73,892	1,196	1,840	73,892
19	1,149	1,768	75,660	1,149	1,768	75,660
20	1,108	1,705	77,365	1,108	1,705	77,365
21	,992	1,526	78,891			

Tabela 5-1 Variância total explicada

Note-se o corte no 1, na antepenúltima coluna.

Scree Plot

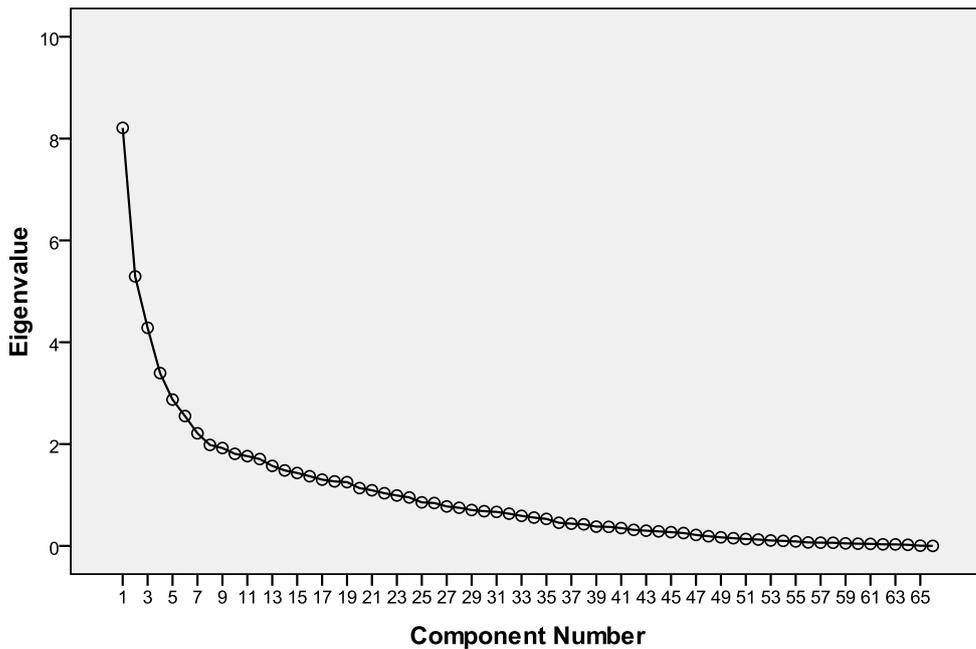


Figura 5-1 Numero de variáveis x variabilidade explicada

Os resultados mostram que só com mais de 20 das variáveis (cerca de 24) conseguimos explicar 80% da variabilidade providenciada pelo conjunto de variáveis total. Assim, a redução dos nossos dados complexos para um modelo mais reduzido através de PCA não foi efectiva, pois ainda tínhamos um número muito grande de variáveis. Por outro lado, isto quer dizer que as variáveis são independentes.

5.2 Análise da variância

Um dos nossos objectivos principais é analisar se podemos construir um modelo de qualidade universal para sítios web em vez de termos um modelo personalizado, ou seja, que é modificado para cada domínio de aplicação. Para fazer esta investigação, adoptamos uma aproximação *bottom-up*. Como temos um número considerável de variáveis descritivas de qualidade de sítios web, fizemos uma análise de variância para determinar se o domínio da aplicação (variável exploratória definida em escala nominal, com 3 categorias Jornais, Bancos e Aviação) tem um impacto em cada uma das variáveis descritivas da qualidade de sítios web.

Em estatística, a análise da variância (abreviada ANOVA) é uma técnica usada para comparar médias de duas ou mais amostras, ou grupos (de casos relacionados com a mesma categoria da variável a explorar). O teste ANOVA testa a hipótese nula de que amostras são tiradas da mesma população. Este teste produz a estatística F, a relação da variância calculada entre as médias e a variância das amostras. Se as médias de grupo são tiradas da mesma população, a variância entre as médias do grupo deve ser mais baixa que a variância das amostras, de acordo com o teorema do limite central. No nosso caso isso significa que a variável de qualidade descritiva correspondente pode ser considerada como descritiva num modelo de qualidade de sítios web universal. Uma maior relação implica contudo que as amostras foram tiradas de populações diferentes. No nosso caso isso implica que a variável de qualidade descritiva correspondente pode ser considerada para inclusão num modelo de qualidade de sítios web personalizado por domínio de aplicação.

Como existem testes ANOVA paramétricos, realizamos um teste de aderência Kolmogorof-Smirnov, com uma significância $\alpha=5\%$ para testar o quanto a distribuição

de cada indicador de sítios web pode ser aproximada à distribuição normal. O teste de significância K-S é mostrado na tabela 6-4 até à tabela 6-9. Como pode ser visto, apenas 7 indicadores de qualidade nos sítios web tem evidência estatística de distribuição normal. Para estes usamos o teste paramétrico One-Way Anova.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
accessibility_img_alt	Between Groups	,045	2	,022	,289	,750
	Within Groups	8,629	111	,078		
	Total	8,674	113			
accessibility_validate_access	Between Groups	254913,298	2	127456,649	4,768	,010
	Within Groups	2,967E+06	111	26730,279		
	Total	3,222E+06	113			
efficiency_page_size	Between Groups	2,376E+10	2	1,188E+10	12,606	,000
	Within Groups	1,046E+11	111	9,426E+08		
	Total	1,284E+11	113			
links_num_intern_links	Between Groups	58175,893	2	29087,946	14,798	,000
	Within Groups	218187,118	111	1965,650		
	Total	276363,010	113			
text_num_sentences_in_paragraph	Between Groups	5,071	2	2,535	,848	,431
	Within Groups	331,812	111	2,989		
	Total	336,883	113			
text_num_syllables_in_word	Between Groups	1,637	2	,819	1,385	,255
	Within Groups	65,588	111	,591		
	Total	67,226	113			
text_num_words_in_sentence	Between Groups	38,462	2	19,231	1,160	,317
	Within Groups	1840,028	111	16,577		
	Total	1878,490	113			

Tabela 5-2 Resultado da análise Anova para as 7 variáveis normais

Para o teste One-way Anova, comparamos os valores da estatística-F com o valor de $F(n-1; p-n)$ ou seja $F(2;108)$, que é 3,08, como tirado de tabelas. Quando o valor F é maior que este valor, pode-se rejeitar a hipótese nula que as amostras vêm da mesma

população, ou seja, podemos concluir que o domínio de aplicação afecta os valores médios. No nosso caso as 3 variáveis que obedecem à distribuição normal sofrem este efeito.

Podemos ainda ver os gráficos de comparação de médias em variáveis que aderem à distribuição normal.

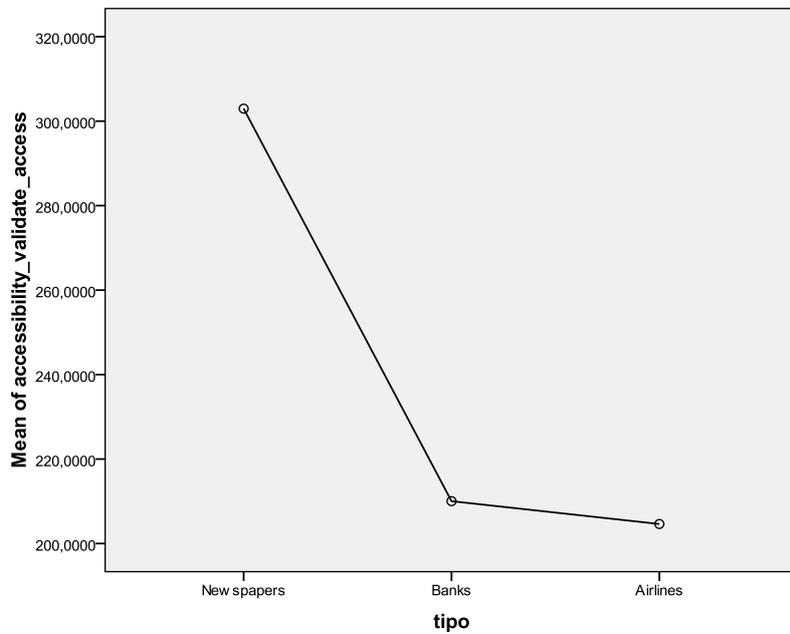


Figura 5-2 Médias x erros de validação

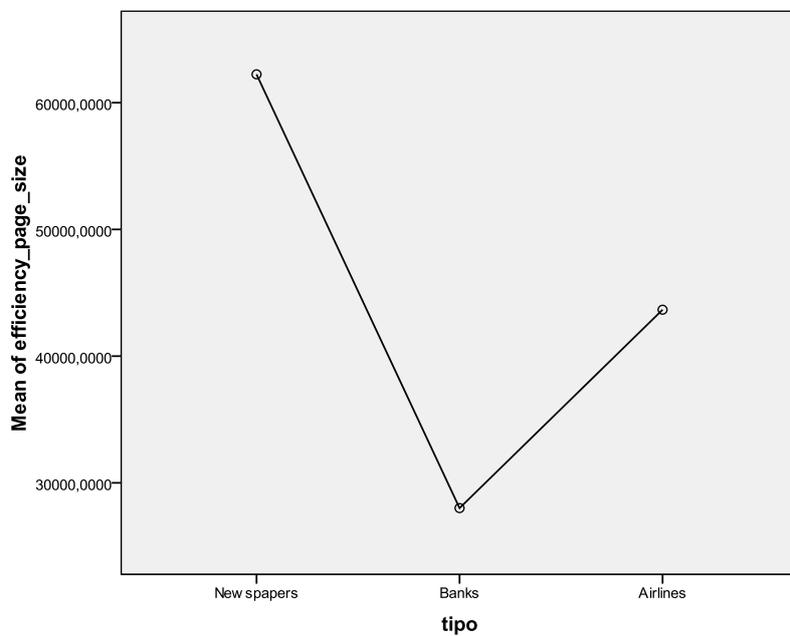


Figura 5-3 Médias vs tamanho de página

Neste gráfico vemos que o tamanho media das páginas é menor nos bancos, seguido das companhias de aviação e os jornais têm as páginas maiores.

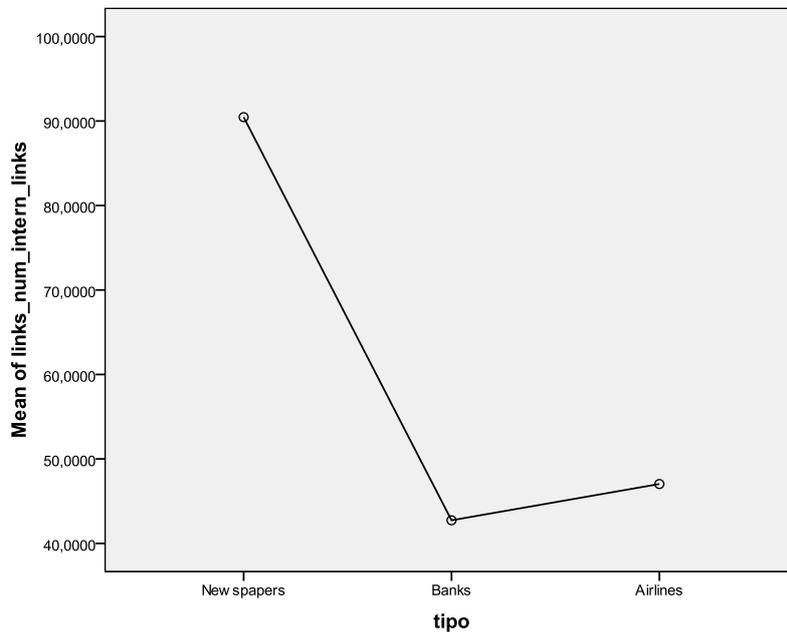


Figura 5-4 Médias vs numero de ligações internos

Neste ultimo gráfico, vemos que os números de ligações internas são muito maiores nos jornais que nos outros, o que é perfeitamente normal, dado serem sítios com informação muito interligada.

Para os outros indicadores usamos o teste equivalente não-paramétrico, o teste Kruskal-Wallis (ou H-test), cuja estatística correspondente é mostrada na tabela 5-4 até à tabela 5-9. Comparamos com o valor 5,991, também tirado de tabelas H.

Test Statistics ^{**}										
	accessibility_img_title	validation_errors	validation_warnings	efficiency_css_size	efficiency_html_page_load_time	efficiency_img_size	efficiency_javascript_size	efficiency_page_load_time	forms_form_pedido_informacao	forms_labels
Chi-Square	10,038	3,404	3,056	32,219	1,922	28,062	25,938	,310	8,445	10,155
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	,007	,182	,217	,000	,382	,000	,000	,856	,015	,006
	identity_author	identity_logo	identity_site_name_title	links_average_num_words	links_links_title	links_num_broken_links	links_num_external_broken_links	links_num_external_links	links_num_image_links	links_num_internal_broken_links
Chi-Square	6,845	2,501	4,390	6,372	8,878	4,169	,556	43,605	25,266	4,109
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	,033	,286	,111	,041	,012	,124	,757	,000	,000	,128
	links_num_links	links_num_noimplemented_links	links_page_without_links	manutencao_num_scripts	manutencao_num_styles	manutencao_num_tables	multimedia_num_img	navigation_bar	navigation_breadcrumbs	navigation_quality_of_links
Chi-Square	33,625	5,071	11,707	9,843	9,792	9,147	11,228	,336	1,393	6,799
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	,000	,079	,003	,007	,007	,010	,004	,845	,498	,033
	pagelayout_device_specific	pagelayout_html_standards	pagelayout_num_divs	pagelayout_num_frames	pagelayout_num_tables	pagelayout_num_tables_in_side_tables	text_font_size_average_em	text_font_size_average_px	text_font_size_max_em	text_font_size_max_px
Chi-Square	1,463	8,663	28,661	4,228	8,099	6,842	1,424	4,276	1,099	4,620
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	,481	,013	,000	,121	,017	,033	,491	,118	,577	,099
	text_font_size_min_em	text_font_size_min_px	text_heading_level	text_heading_reverse_order	text_italic_text	text_num_diferent_colors	text_num_diferent_fonts	text_num_subheading_heading	text_num_words_meta_description	text_num_words_meta_keywords
Chi-Square	1,557	4,212	6,137	22,718	7,092	54,909	31,765	12,721	15,901	6,196
df	2	2	2	2	2	2	2	2	2	2
Asymp. Sig.	,459	,122	,046	,000	,029	,000	,000	,002	,000	,045
	text_paragraph_max_size	text_paragraph_size	text_subheading_len	text_total_newlines	text_total_sentences	text_total_syllables	text_total_words	text_uppercase_text		
Chi-Square	6,644	,641	15,883	,205	7,765	5,047	5,636	2,900		
df	2	2	2	2	2	2	2	2		
Asymp. Sig.	,036	,726	,000	,903	,021	,080	,060	,235		

Tabela 5-3 Teste Kruskal-Wallis

Nas mesmas tabelas, a última coluna (chamada “Conclusão”) mostra o rotulo “EFEITO” nos casos em que o teste ANOVA revelou a detecção de uma diferença estatística significativa na variância entre os grupos. Como tal, os indicadores de qualidade de sítios web marcados com EFEITO devem ser considerados como possíveis candidatos para um modelo de qualidade de sítios web específico do domínio, enquanto os restantes são candidatos a um modelo de qualidade independente do domínio. Note-se que para todas as características da ISO9126 podemos encontrar indicadores dependentes e independentes do domínio.

As seguintes tabelas representam os resultados da aderência `distribuição normal, e análise dos testes de variância, para cada indicador de qualidade de sítios web, agrupados pelas características da ISO9126.

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
forms_form_pedido_informacao	,000		8,445	EFEITO
forms_labels	,000		10,155	EFEITO
identity_author	,000		6,845	EFEITO
identity_logo	,002		2,501	
identity_sitename_title	,000		4,390	
navigation_bar	,000		,336	
navigation_breadcrumbs	,000		1,393	
navigation_quality_of_links	,000		6,799	EFEITO

Tabela 5-4 Métricas de funcionalidade vs Domínio da Aplicação

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
validation_errors	,000		3,404	
validation_warnings	,000		3,056	
links_average_num_words	,000		6,372	EFEITO
links_links_title	,000		8,878	EFEITO
links_num_broken_links	,001		4,169	
links_num_extern_broken_links	,000		,556	
links_num_extern_links	,000		43,605	EFEITO
links_num_image_links	,000		25,266	EFEITO
links_num_intern_broken_links	,001		4,109	
links_num_intern_links	,063	14,798		EFEITO
links_num_links	,028		33,625	EFEITO
links_num_non_implemented_links	,000		5,071	
links_page_without_links	,000		11,707	EFEITO

Tabela 5-5 Métricas de fiabilidade vs Domínio da Aplicação

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
accessibility_img_alt	,309	,289		
accessibility_img_title	,000		10,038	EFEITO
accessibility_validate_access	,056	4,768		EFEITO
multimedia_num_img	,003		11,228	EFEITO
text_font_size_average_em	,000		1,424	
text_font_size_average_px	,000		4,276	
text_font_size_max_em	,000		1,099	
text_font_size_max_px	,000		4,620	
text_font_size_min_em	,000		1,557	
text_font_size_min_px	,000		4,212	
text_heading_len	,000		6,137	EFEITO
text_heading_reverse_order	,000		22,718	EFEITO

text_italic_text	,000		7,092	EFEITO
text_num_diferent_colors	,000		54,909	EFEITO
text_num_diferent_fonts	,000		31,765	EFEITO
text_num_sentences_in_paragraph	,083	,848		
text_num_subheading_heading	,000		12,721	EFEITO
text_num_syllables_in_word	,994	1,385		
text_num_words_in_sentence	,173	1,160		
text_num_words_meta_description	,000		15,901	EFEITO
text_num_words_meta_keywords	,000		6,196	EFEITO
text_paragraph_max_size	,001		6,644	EFEITO
text_paragraph_size	,014		,641	
text_subheading_len	,000		15,883	EFEITO
text_total_newlines	,000		,205	
text_total_sentences	,004		7,765	EFEITO
text_total_syllables	,001		5,047	
text_total_words	,000		5,636	
text_uppercase_text	,000		2,900	

Tabela 5-6 Métricas de Usabilidade vs Domínio da Aplicação

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
efficiency_css_size	,000		32,219	EFEITO
efficiency_homepage_load_time	,017		1,922	
efficiency_img_size	,000		28,062	EFEITO
efficiency_javascript_size	,000		25,938	EFEITO
efficiency_page_load_time	,001		,310	
efficiency_page_size	,087	12,606		EFEITO

Tabela 5-7 Métricas de Eficiência vs Domínio da Aplicação

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
manutencao_num_scripts	,019		9,843	EFEITO
manutencao_num_styles	,006		9,792	EFEITO
manutencao_num_tables	,000		9,147	EFEITO

Tabela 5-8 Métricas de Manutenção vs Domínio da Aplicação

Indicador de Qualidade de Sítios Web	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusão
pagelayout_device_specific	,000		1,463	
pagelayout_html_standards	,000		8,663	EFEITO
pagelayout_num_divs	,002		28,661	EFEITO
pagelayout_num_frames	,000		4,228	
pagelayout_num_tables	,000		8,099	EFEITO
pagelayout_num_tables_inside_tables	,000		6,842	EFEITO

Tabela 5-9 Métricas de Portabilidade vs Domínio da Aplicação

5.3 Modelo de qualidade

Podemos assim construir um modelo que qualidade para sítios web que tem duas componentes, ou seja, um modelo de qualidade com uma base independente do domínio e outra dependente.

Categoria	Indicador de Qualidade de Sítios Web
Funcionalidade	identity_logo
	identity_sitename_title
	navigation_bar
	navigation_breadcrumbs
Fiabilidade	validation_errors
	validation_warnings
	links_num_broken_links
	links_num_extern_broken_links
	links_num_intern_broken_links
Usabilidade	accessibility_img_alt
	text_font_size_average_em
	text_font_size_average_px
	text_font_size_max_em
	text_font_size_max_px
	text_font_size_min_em
	text_font_size_min_px
	text_num_sentences_in_paragraph
	text_num_syllables_in_word
	text_num_words_in_sentence
	text_paragraph_size
	text_total_newlines
	text_total_syllables
	text_total_words
text_uppercase_text	
Eficiência	efficiency_homepage_load_time
	efficiency_page_load_time
Portabilidade	pagelayout_device_specific
	pagelayout_num_frames

Tabela 5-10 Parte do Modelo de qualidade não dependente do domínio

Categoria	Indicador de Qualidade de Sítios Web
Funcionalidade	forms_form_pedido_informacao
	forms_labels
	identity_author
	navigation_quality_of_links
Fiabilidade	links_average_num_words
	links_links_title
	links_num_extern_links
	links_num_image_links
	links_num_intern_links
	links_num_links
	links_page_without_links
Usabilidade	accessibility_img_title
	accessibility_validate_access
	multimedia_num_img
	text_heading_len
	text_heading_reverse_order
	text_italic_text
	text_num_diferent_colors
	text_num_diferent_fonts
	text_num_subheading_heading
	text_num_words_meta_description
	text_num_words_meta_keywords
	text_paragraph_max_size
	text_subheading_len
	text_total_sentences
Eficiência	efficiency_css_size
	efficiency_img_size
	efficiency_javascript_size
	efficiency_page_size
Manutenção	manutencao_num_scripts
	manutencao_num_styles
	manutencao_num_tables
Portabilidade	pagelayout_html_standards
	pagelayout_num_divs
	pagelayout_num_tables
	pagelayout_num_tables_inside_tables

Tabela 5-11 Parte do Modelo de qualidade dependente do domínio

Cada indicador deve ter um intervalo de valores no qual deve estar contido, para o caso dos numéricos. Para os booleanos, deve-se definir se é bom para a qualidade o verdadeiro ou falso. No caso dos indicadores numéricos, pode-se identificar limites das seguintes maneiras:

- Comparar com a avaliação de peritos que teriam de avaliar todos os sítios da amostra e ainda classificar os sítios como bons ou maus a nível de qualidade e numa escala, divididos pelas macro-características (exequível mas bastante moroso).

- Fazer um ranking dos sítios quanto às macro-características e depois através de combinações, inferir a sua qualidade (não exequível pois eram muitas combinações).
- Definir intervalos (ou máximos ou mínimos) baseados na informação estatística que dispomos, e considerando que se for um *outlier*, existirá um problema com esse indicador.

Devido à dificuldade de implementar os dois primeiros, vamos usar o terceiro método. Podemos pensar num modelo de qualidade, de acordo com as médias de cada indicador numérico e baseando-nos no facto que para as variáveis numéricas não devemos deixar o valor ser um *outlier*. Os outliers são definidos como sendo:

$$\text{Outlier Superior} = \text{Quartil 3} + 1,5 \text{ DIQ}$$

$$\text{Outlier Inferior} = \text{Quartil 1} - 1,5 \text{ DIQ}$$

Onde o Quartil 1 é o percentil 25, o quartil 3 é o percentil 75, e o DIQ- distancia inter quartil é a diferença entre o percentil 75 e o percentil 25.

Contudo, todos os indicadores terão de ser analisados, pois enquanto que para uns será definido um máximo, para outros será definido um mínimo. Os indicadores dependentes do domínio terão limites conforme o domínio a ser avaliado.

Para os indicadores booleanos, será atribuído o valor aceitável ou não aceitável.

5.3.1 Exemplo de limites para indicadores

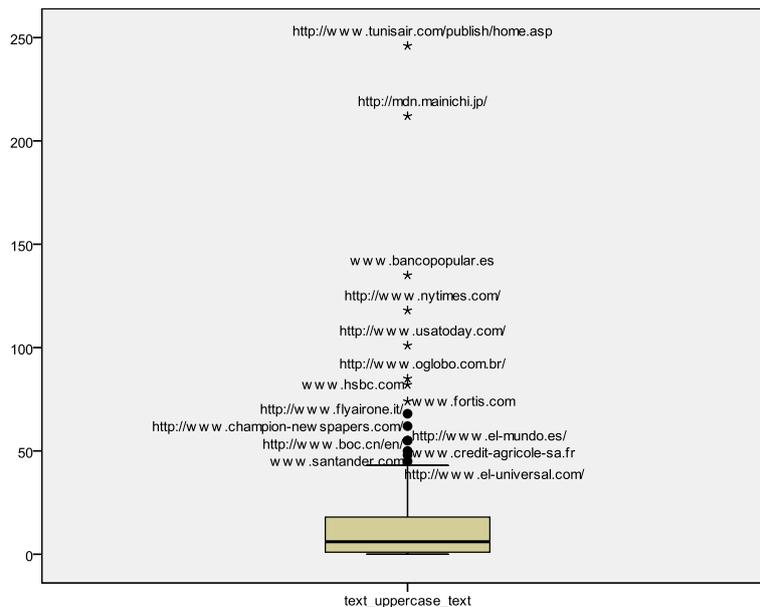


Figura 5-5 Exemplo de indicador não dependente do domínio - maiúsculas

Este indicador mede os blocos de texto em maiúsculas. Assim, não é recomendável ter muito texto em maiúsculas pois é de leitura difícil e logo este indicador deve ser majorado.

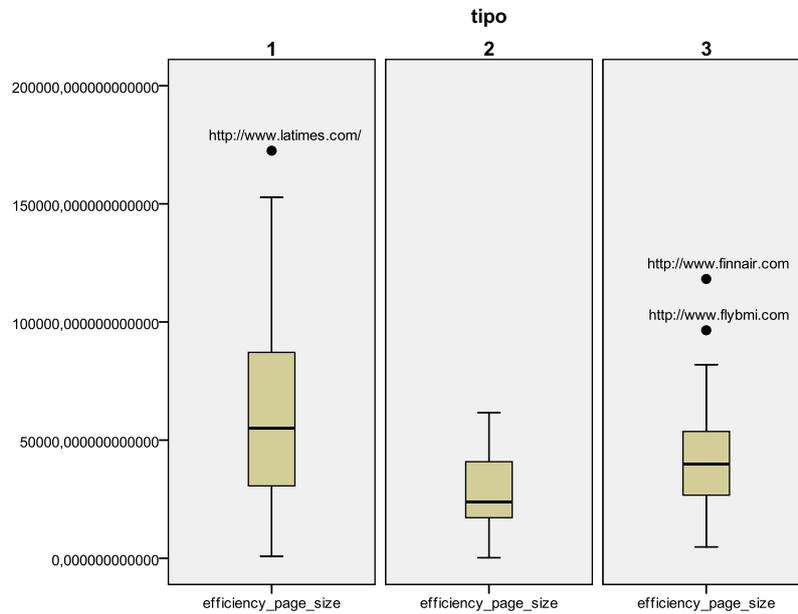


Figura 5-6 Indicador tamanho de página - dependente do domínio (menor é melhor)

Neste caso observamos alguns casos que saem da amostra. Estes têm páginas muito grandes comparado com os outros sítios do seu domínio, e os utilizadores vão sentir essa diferença, pois traduz-se em maior tempo para espera para as páginas carregarem.

5.4 Relação das características de qualidade com o PageRank do Google

O coração do Google, o sistema de pesquisa mais usado na internet, é o PageRank, um sistema de classificação e ordenação de páginas desenvolvido pelos fundadores do Google na Universidade de Stanford. O PageRank é um conjunto de algoritmos de análise de ligações que atribui pesos numéricos a cada elemento de um conjunto de documentos hiperligados, como a Web, com a intenção de medir a sua importância relativa no conjunto de documentos. Este peso é usado no motor de busca Google para mostrar os seus resultados e o valor de um elemento E, chamado PageRank de E, e escrito como $PR(E)$. O “Page” de PageRank vem do nome de um dos seus criadores Larry Page.

5.4.1 Funcionamento do page rank

Como podemos ver na página do Google, o seu funcionamento é o seguinte:

‘O PageRank assenta na natureza excepcionalmente democrática da Web, ao utilizar a sua vasta estrutura de ligações como uma medida do valor de uma página individual. Essencialmente, o Google interpreta uma ligação da página A para a página B, como um voto da página A em direcção à página B. No entanto, o Google analisa mais do que o simples volume dos votos ou as ligações recebidos por uma página; analisa, também, a página que lança o voto. Os votos dados pelas páginas que são, por si próprias, "importantes", têm maior peso e ajudam a tornar outras páginas "importantes.”

Os sítios importantes e de alta qualidade recebem um PageRank mais alto, que o Google recorda, cada vez que efectua uma pesquisa. Naturalmente que uma página importante não tem qualquer significado se não corresponder à sua pesquisa, pelo que o Google combina o PageRank com sofisticadas técnicas de correspondência de texto, com o objectivo de encontrar páginas que sejam, ao mesmo tempo, importantes e relevantes para a sua pesquisa. O Google vai muito além do número de vezes que um termo aparece numa página, examinando todos os aspectos do conteúdo dessa página (bem como o conteúdo das páginas que fornecem ligações para essa página), para determinar o melhor resultado possível para a sua pesquisa’.

Assim, embora seja um algoritmo de ligações, admitimos que mede de algum modo a qualidade de um sítio. Através de processos estatísticos, veremos se o podemos relacionar com a nossa lista de indicadores/características.

5.4.2 Recolha do PageRank

Para a recolha do PageRank, fizemos um programa em PHP que extrai o PageRank, da base de dados do Google. Este programa utiliza um conjunto de funções em PHP que envia a assinatura da toolbar do Google que é instalado nos *browsers*, para que consiga retirar este valor. Os valores recolhidos para a nossa amostra são mostrados na tabela 5-12.

idSite	Site	P Rank	idSite	Site	P Rank
1	http://www.usatoday.com/	8	72	http://www.bancoexterior.com/	4
2	http://europe.wsj.com/home-page	6	73	http://www.bancofederal.com/	5
3	http://www.nytimes.com/	9	74	www.commerzbank.com	6
4	http://www.latimes.com/	8	75	www.postbank.de	7
5	http://www.thesun.co.uk/sol/homepage/	7	76	www.db.com	7
6	http://www.dailymail.co.uk/home/index.html	7	77	www.bnpparibas.com	7
7	http://www.mirror.co.uk/	8	78	www.ir.socgen.com	5
8	http://www.telegraph.co.uk/	8	79	www.credit-agricole-sa.fr	6
10	http://www.jb.com.br/	7	80	www.kbc.com	7
11	http://www.oglobo.com.br/	8	81	www.dexia.com	6
12	http://www.lanacion.com.ar/	8	82	www.fortis.com	7
13	http://www.el-nacional.com	7	84	www.millenniumbcp.pt	7
14	http://www.el-universal.com/	7	85	www.bes.pt	7
15	http://www.2001.com.ve/	6	86	www.bancopopular.es	6
16	www.bild.de	7	87	www.bbva.es	7
17	www.sueddeutsche.de	7	88	www.santander.com	9
18	http://www.zeit.de	8	91	www.norges-bank.no/english/	6
19	http://www.ouest-france.fr/	7	92	www.amagerbanken.dk/	5
20	http://www.latribune.fr/accueil/a-la-une.html	7	94	www.swedbank.com/	6
21	http://www.leparisien.fr	7	95	http://www.icbc.com.cn/icbc/sy/default.htm	6
22	www.gva.be	7	97	http://www.boc.cn/en/	7
25	http://www.telegraaf.nl/	7	98	http://www.mufg.jp/english/	6
26	http://www.krone.at/krone/S1/kmprog/index.html	7	100	http://www.zenithbank.com/index.cfm	7
27	www.dn.pt	7	101	http://www.ubagroup.com/	6
28	www.jnoticias.pt	7	102	http://www.ecobank.com/english/others/home.aspx	7
29	http://www.publico.pt/	7	110	http://www.southwest.com/	7
30	www.elpais.es	8	111	http://www.aa.com/	8
31	http://www.el-mundo.es/	9	113	http://www.united.com/	7
32	http://www.abc.es/	8	114	www.aircanada.com	7
33	http://www.corriere.it/	8	116	http://www.easyjet.com/	7
34	http://www.repubblica.it/	8	117	http://www.britishairways.com	8
35	http://www.aftenposten.no/	8	118	http://www.flybmi.com	6
36	http://www.dagbladet.no/	7	120	http://www.voegol.com.br/Paginas/home.aspx	6
37	http://www.stiften-aarhus.dk/	3	122	http://www.aerolineas.com.ar/home.asp	6
38	http://www.borsen.dk/	7	123	http://www.avianca.com	0
39	http://www.expressen.se/	8	124	http://www.lan.com/cgi-bin/country_selector.cgi	7
40	http://www.people.com.cn/	9	125	http://www.airfrance.fr/	7
42	http://en.huanqiu.com/	0	126	http://www.klm.com/travel/gb_en/index_default.html	6
44	http://www.asahi.com/english/	7	127	http://www.lufthansa.de/	0
45	http://mdn.mainichi.jp/	7	128	http://www.airberlin.com/	7
47	http://timesofindia.indiatimes.com/default1.cms	5	129	http://www.swiss.com	8
49	http://www.jornaldeangola.com/	6	132	www.alitalia.com	6
50	http://www.champion-newspapers.com/	0	134	http://www.flytap.com/	7
51	http://www.tribune.com.ng/	6	135	http://www.flyairone.it/	7
52	http://www.thisdayonline.com/	7	136	http://www.vueling.com/	7
53	http://www.citizen.co.za/	5	138	http://www.norwegian.no	6
54	http://www.dailynews.co.za/	6	139	http://www.finnair.com	7
57	http://www.lapresse.tn	7	140	http://www.flychinasouthern.com/	6
60	www.jpmorgan Chase.com	7	143	http://www.jal.co.jp/en/	5
61	www.citigroup.com	7	144	http://www.ana.co.jp/	7
63	www.hsbc.com	7	145	http://www.singaporeair.com/saa/index.jsp	0
64	www.rbs.com	7	146	http://ww4.flysaa.com/	8

65	www.standardchartered.com	7	148	http://www.kenya-airways.com/home/	7
66	www.barclays.com	7	149	http://www.ethiopianairlines.com/	7
67	http://www.itau.com.br/	6	150	http://www.egyptair.com/English/Pages/splashpage.aspx	7
69	http://www.bradesco.com.br/	7	151	http://www.royalairmaroc.com/	7
70	http://www.e-galicia.com/	6	152	http://www.tunisair.com/publish/home.asp	6

Tabela 5-12 Page rank da amostra

5.4.3 Análise estatística

Para realizar a análise estatística, o primeiro passo foi ver se conseguimos relacionar de algum modo as variáveis que temos com os valores do PageRank.

Existem vários tipos de regressões, e várias são disponibilizadas pelo programa SPSS, que é o programa que temos vindo a utilizar para a análise estatística. Contudo, sabemos que a nossa variável dependente é ordinal, ou seja, temos valores para a variável dependente que estão ordenados, mas não sabemos se os valores são equidistantes. Por outras palavras, não sabemos se a distância entre o 1 e o 2, é a mesma distância que entre o 5 e o 6, por exemplo. Devido a este facto, a regressão apropriada é a regressão ordinal.

5.4.4 Regressão Ordinal

A regressão ordinal permite-nos construir modelos, gerar previsões e avaliar a importância das variáveis nelas envolvidas, quando a variável dependente é de natureza ordinal. Neste caso, os modelos lineares não funcionam muito bem, pois estes modelos assumem que a variável dependente é medida numa escala de intervalo e as simplificações assumidas não são satisfeitas. Com uma variável ordinal, o importante é a ordem de categorias. Se se colapsar duas categorias adjacentes numa categoria maior, a alteração no modelo é mínima e os modelos antes da junção e depois devem ser bastante semelhantes. Na regressão linear, não se podem fundir categorias.

A forma básica do modelo linear generalizado é mostrada na equação seguinte:

$$\text{link}(\gamma_{ij}) = \theta_j - [\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}] \quad \text{- Equação 1}$$

Onde:

- link() É a função de ligação
- γ_{ij} É a probabilidade acumulada da categoria j para o caso i
- θ_j É o corte (interceção) para a categoria j (chamado threshold em inglês)
- p É o número de coeficiente da regressão
- $x_{i1} \dots x_{ip}$ São os valores dos predictores para o caso i
- $\beta_1 \dots \beta_p$ São os coeficientes da regressão

Em lugar de prever as probabilidades cumulativas, o modelo prevê uma função desses valores, que é a função de ligação. Esta função deve ser escolhida baseada no problema a resolver.

O modelo pode ser dividido em 3 partes:

- Componente de Localização: o que se encontra dentro do parêntesis recto, e que inclui os coeficientes e variáveis de predição.
- Componente de escala: a escala é uma modificação opcional ao modelo que tem em consideração as diferenças em variabilidade nas variáveis predictoras. Este componente aparece em denominador do modelo, e o modelo toma a forma

$$link(\gamma_{ij}) = \frac{\theta_j - [\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ij}]}{e^{\tau_1 z_{i1} + \tau_m z_{im}}} \text{ - Equação 2}$$

onde $z_{i1} \dots z_{im}$ são os predictores dos componentes de escala, e $\tau_1 \dots \tau_m$ são os coeficientes dos componentes de escala.

- Função de ligação - é uma transformação das probabilidades que nos deixa estimar o modelo. Estão disponíveis no SPSS 5 funções de ligação:

Função	Forma	Aplicação Típica
Logit	$\log(x / (1-x))$	Categorias distribuídas uniformemente
Complementary log-log	$\log(-\log(1-x))$	Categorias mais altas mais prováveis
Negative log-log	$-\log(-\log(x))$	Categorias mais altas mais prováveis
Probit	$F^{-1}(x)$	Variável latente tem distribuição normal
Cauchit (inverse Cauchy)	$\tan(\pi(x-0.5))$	Variável latente tem muitos valores extremos

Tabela 5-13 Funções de ligação para regressão ordinal

5.4.5 Aplicação da regressão ordinal

Vamos fazer uma regressão ordinal com os valores pré-definidos do SPSS.

Neste caso temos valores de correlação diferentes de 0, mas que ainda não são os melhores, quando usamos a função de link logit.

Pseudo R-Square	
Cox and Snell	,745
Nagelkerke	,793
McFadden	,486

Link function: Logit.

Tabela 5-14 valores de correlação para a regressão ordinal, com a função de link= logit

Consultando a literatura do SPSS, consideramos as várias funções de ligação e vemos que a função de ligação que mais se adequa à nossa lista de casos é o Complementary – log-log, pois temos a maior parte dos valores de PageRank na metade superior, ou seja, entre 6 e 9. Para esta regressão, com esta função de ligação, temos a correlação:

Pseudo R-Square	
Cox and Snell	,940
Nagelkerke	1,000
McFadden	1,000

Link function:

Complementary Log-log.

Tabela 5-15 valores de correlação para a regressão ordinal, com a função de ligação= complementary log-log

Ou seja, com todas as variáveis, os valores de correlação que temos são bastantes grandes. Assim, concluímos que é possível ter um modelo que consiga prever o pageRank mas que tem muitas variáveis. O problema é que este modelo é um modelo de probabilidades, e as variáveis que são ordinais (discretas) vão ter tantos casos quanto valores possível, gerando assim um modelo muito grande.

Threshold	Estimate	Location	Estimate
[rank = 4]	-2,405	pagelayout_num_frames	17,332
[rank = 5]	-1,718	pagelayout_num_tables	,020
[rank = 6]	-,422	pagelayout_num_tables_inside_tables	-,014
[rank = 7]	1,666	text_font_size_average_em	4,093
[rank = 8]	2,807	text_font_size_average_px	,002
Location	Estimate	text_font_size_max_em	-1,468
accessibility_img_alt	-,011	text_font_size_max_px	-,021
accessibility_img_title	-,342	text_font_size_min_em	-1,908
accessibility_validate_access	,002	text_font_size_min_px	,032
validation_errors	-,045	text_heading_len	,001
validation_warnings	,003	text_heading_reverse_order	,004
efficiency_css_size	,000	text_italic_text	,000
efficiency_homepage_load_time	,283	text_num_diferent_colors	,022
efficiency_img_size	,000	text_num_diferent_fonts	-,055
efficiency_javascript_size	,000	links_num_intern_links	-,100
efficiency_page_load_time	-,302	links_num_links	,101
efficiency_page_size	,000	text_num_sentences_in_paragraph	-,131
forms_form_pedido_informacao	,051	text_num_subheading_heading	-,011
forms_labels	1,150	text_num_syllables_in_word	,210
identity_author	,634	text_num_words_in_sentence	-,083
identity_logo	-,305	text_num_words_meta_description	-,012
identity_sitename_title	-,734	text_num_words_meta_keywords	,003
links_average_num_words	,076	text_paragraph_max_size	,000
links_links_title	-,004	text_paragraph_size	,007
links_num_broken_links	,205	text_subheading_len	-,005
links_num_extern_broken_links	-12,160	text_total_newlines	,000
links_num_extern_links	-,105	text_total_sentences	,000
links_num_image_links	,027	text_total_syllables	,000
links_num_intern_broken_links	0	text_total_words	,000
links_num_non_implemented_links	-,047	text_uppercase_text	,004
links_page_without_links	-,015	[area_=1]	,787
manutencao_num_scripts	,043	[area_=2]	-1,394
manutencao_num_styles	-,100	[area_=3]	,243
manutencao_num_tables	,015	[area_=4]	-,652
multimedia_num_img	-,037	[area_=5]	-2,224
navigation_bar	,095	[area_=6]	-1,106
navigation_breadcrumbs	-,283	[area_=7]	,402
navigation_quality_of_links	,325	[area_=8]	0
pagelayout_device_specific	,261	[tipo=1]	,943
pagelayout_html_standards	-,518	[tipo=2]	-,191
pagelayout_num_divs	-,005	[tipo=3]	0

Tabela 5-16 Estimação de parâmetros para modelo de regressão nominal com todas as variáveis

Como se vê, o caminho a seguir é o da redução de variáveis.

Fizemos uma simplificação neste modelo que foi o colapso das categorias que tem o PR ≤ 4 , na mesma categoria. Como vimos anteriormente, esta junção não influencia o modelo

5.4.6 Redução de variáveis

Para a redução de variáveis fizemos uma análise de co-variância, mas incluindo a variável Rank. Como algumas das variáveis não têm distribuição normal, tivemos de usar o coeficiente de Spearman.

Assim, a lista que obtivemos foi a seguinte:

Métrica	Correlação
links_num_extern_links	,395**
links_num_links	,331**
links_num_image_links	,310**
text_heading_reverse_order	,301**
links_links_title	,293**
links_num_intern_links	,289**
text_num_diferent_colors	,283**
text_num_subheading_heading	,276**
pagelayout_device_specific	,229*
pagelayout_num_tables	-,227*
text_num_words_meta_description	,224*
text_num_words_meta_keywords	,224*
accessibility_img_title	,223*
text_subheading_len	,221*
text_font_size_max_em	,215*
forms_labels	,212*
pagelayout_html_standards	,212*
text_font_size_average_em	,209*
text_heading_len	,207*
manutencao_num_tables	-,203*
text_font_size_min_em	,203*
efficiency_css_size	,197*
efficiency_page_load_time	-,195*
efficiency_page_size	,195*
identity_author	,195*
pagelayout_num_divs	,195*
text_num_diferent_fonts	,195*
text_font_size_max_px	,192*
pagelayout_num_tables_inside_tables	-,189*

Tabela 5-17 Análise de Co-variância – Spearman

Da observação da tabela, nota que podemos ter bastantes caso correlacionados entre si.

Começamos pelos maiores suspeitos, os que começam por *link*.

Para estes, fizemos a correlação de todos os links, tiramos os que tinham:

1. maior co-variância entre eles
2. em caso de empate, o que tinha a correlação maior com o PR

Assim, para as variáveis, ou métricas, de *link*, ficaram:

links_num_extern_links
links_num_image_links
links_links_title

Para as métricas do grupo *text* fizemos igual, o que resultou nas métricas:

text_heading_reverse_order
text_num_diferent_colors
text_font_size_max_em
text_font_size_max_px

Para as métricas de *pagelayout* ficamos com a seguinte lista, após análise de co-variância de Spearman

pagelayout_device_specific
pagelayout_num_tables
pagelayout_num_divs

No grupo das variáveis da eficiência sai o *pagesize* e ficam as outras:

efficiency_css_size
efficiency_page_size

Ficamos com o seguinte conjunto reduzido de variáveis, o qual vamos fazer a regressão nominal novamente.

links_num_extern_links
links_num_image_links
text_heading_reverse_order
links_links_title
text_num_diferent_colors
pagelayout_device_specific
pagelayout_num_tables
accessibility_img_title
text_font_size_max_em
forms_labels
efficiency_css_size
efficiency_page_size
identity_author
pagelayout_num_divs
text_font_size_max_px

Tabela 5-18 Lista de variáveis candidatas para modelo reduzido

5.4.7 Regressão ordinal com conjunto reduzido de variáveis

Fazendo a regressão ordinal novamente com o SPSS, e com este conjunto de variáveis, obtemos os seguintes coeficientes ou variáveis predictoras.

		Estimate
Threshold	[rank = 4]	-2,243
	[rank = 5]	-1,573
	[rank = 6]	-,312
	[rank = 7]	1,759
	[rank = 8]	3,028
Location	links_num_extern_links	-,001
	links_num_image_links	,023
	text_heading_reverse_order	,002
	links_links_title	-,004
	text_num_diferent_colors	,017
	pagelayout_device_specific	-,009
	pagelayout_num_tables	-,021
	accessibility_img_title	-,204
	text_font_size_max_em	,108
	forms_labels	,579
	efficiency_css_size	,000
	efficiency_page_size	,000
	identity_author	,736
	pagelayout_num_divs	-,007
	text_font_size_max_px	,002
	[tipo=1]	,350
	[tipo=2]	-,324
	[tipo=3]	0
	[area_=1]	-,010
	[area_=2]	-,304
[area_=3]	,842	
[area_=4]	-,144	
[area_=5]	-,662	
[area_=6]	-,450	
[area_=7]	1,031	
[area_=8]	0	

Tabela 5-19 Coeficientes para o conjunto reduzido de variáveis, com a função de ligação= complementary log-log

Vemos que perdemos um pouco da capacidade de predição, como se pode observar pelas tabelas seguintes. O *pseudo R-Square* dá-nos a significância prática do modelo.

Cox and Snell	,905
Nagelkerke	,963
McFadden	,839

Link function:

Complementary Log-log.

Tabela 5-20 valores de correlação para a regressão ordinal com numero de varaiveis reduzido, com a função de ligação= complementary log-log

Observamos também os parâmetros de aderência, ou *fit*, que são bastante bons.

	Chi-Square	df	Sig.
Pearson	484,715	536	,945
Deviance	259,178	536	1,000

Link function: Complementary Log-log.

Tabela 5-21 Fit para modelo com conjunto de varaiveis reduzido

Para ver quão bom seria o nosso modelo, efectuamos uma tabela cruzada que relaciona o PR real (na coluna da esquerda) com o PR previsto (na linha de cima). O modelo será melhor se obtivermos um maior número de cruzamentos onde o número destes dois é igual.

rank * Predicted Response Category Crosstabulation

		Predicted Response Category				Total
		6	7	8	9	
rank	4	2	5	0	0	7
	5	2	4	0	0	6
	6	4	18	1	0	23
	7	3	52	1	0	56
	8	0	8	8	1	17
	9	0	1	3	0	4
Total		11	88	13	1	113

Tabela 5-22 Crosstab do rank x rank previsto para o modelo com o conjunto reduzido de variáveis

Analizamos a tabela e vemos que temos uma boa capacidade de predição, pois temos um grande número de casos que se pode prever. Contudo, podemos melhorar.

5.4.8 Regressão ordinal dependente do domínio

Segundo o nosso estudo, parte das variáveis ou métricas são dependentes do domínio da aplicação. Assim, o modelo será mais preciso se o dividirmos em 3 modelos, um para cada domínio de aplicação.

5.4.8.1 Domínio Jornais

Voltamos a fazer a regressão ordinal, mas somente para o domínio dos jornais, é só considerando esta área de amostra. Conseguimos reduzir ainda mais as variáveis sem perder o fit. Seguem-se os resultados obtidos.

Parameter Estimates		Estimate
Threshold	[rank = 4]	-2,3
	[rank = 5]	-1,73
	[rank = 6]	-0,829
	[rank = 7]	1,477
	[rank = 8]	3,315
Location	links_num_extern_links	0,006
	links_num_image_links	-0,008
	text_heading_reverse_order	0,002
	links_links_title	-0,006
	text_num_diferent_colors	0,02
	pagelayout_device_specific	0,022
	pagelayout_num_tables	-0,013
	accessibility_img_title	-0,172
	text_font_size_max_em	-0,139
	forms_labels	1,151
	efficiency_css_size	-2,76E-06
	[area_=1]	-0,32
	[area_=2]	-2,293
	[area_=3]	1,286
	[area_=4]	0,327
	[area_=5]	-0,663
	[area_=6]	-0,001
[area_=7]	0,898	
[area_=8]	0 ^a	

Tabela 5-23 Parâmetros do modelo para domínio Jornais

Pseudo R-Square	
Cox and Snell	,942
Nagelkerke	1,000
McFadden	1,000

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	179,964	217	,968
Deviance	97,022	217	1,000

Link function: Complementary Log-log.

Tabela 5-24 Fit do modelo para domínio Jornais

Como podemos observar, o *fit* aumenta bastante, mesmo diminuindo as variáveis. Fazendo uma tabela cruzada, ou *crosstabs*, para este domínio obtemos a tabela seguinte.

rank * Rank Previsto ModParc Jornais Crosstabulation

Count		Rank Previsto ModParc Jornais					Total
		4	6	7	8	9	
rank	4	0	1	2	0	0	3
	5	0	1	1	0	0	2
	6	1	1	2	1	0	5
	7	0	0	19	3	0	22
	8	0	0	3	8	2	13
	9	0	0	0	3	0	3
Total		1	3	27	15	2	48

Tabela 5-25 Crosstab do rank x rank previsto para o modelo específico do domínio Jornais

A predição maior dá-se quando se conseguem maximizar os pares (rank previsto, rank) com o mesmo valor. Tomámos como exemplo o valor previsto 7. Conseguimos prever 19 casos com 7, sobrando 4 abaixo e 3 acima.

5.4.8.2 Domínio Bancos

Neste caso, não conseguimos tirar nenhuma variável ao modelo reduzido, sem perder poder de predição. É uma amostra mais heterogénea.

Parameter Estimates

		Estimate			Estimate
Threshold	[rank = 4]	-3,301	Location	forms_labels	-0,377
	[rank = 5]	-1,727		efficiency_css_size	-2,05E-06
	[rank = 6]	0,31		efficiency_page_size	1,18E-05
	[rank = 7]	3,439		identity_author	-1,891
Location	links_num_extern_links	0,047	pagelayout_num_divs	0,052	
	links_num_image_links	-0,18	text_font_size_max_px	0,025	
	text_heading_reverse_order	-0,027	[area_=2]	3,227	
	links_links_title	0,019	[area_=3]	0,852	
	text_num_diferent_colors	-0,034	[area_=4]	0,308	
	pagelayout_device_specific	-0,01	[area_=5]	0,739	
	pagelayout_num_tables	0,007	[area_=6]	-2,826	
	accessibility_img_title	-0,436	[area_=7]	0,767	
	text_font_size_max_em	0,45	[area_=8]	0 ^a	

Tabela 5-26 Parâmetros do modelo para domínio Bancos

Pseudo R-Square	
Cox and Snell	,894
Nagelkerke	1,000
McFadden	1,000

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	100,514	107	,658
Deviance	50,497	107	1,000

Link function: Complementary Log-log.

Tabela 5-27 Fit do modelo para domínio Bancos

Podemos ainda observar a predição para este modelo.

rank * Rank Previsto ModParc Bancos Crosstabulation

Count

		Rank Previsto ModParc Bancos			Total
		5	6	7	
rank	4	0	1	0	1
	5	1	1	1	3
	6	0	7	3	10
	7	0	2	16	18
	9	0	0	1	1
Total		1	11	21	33

Tabela 5-28 Crosstab do rank x rank previsto para o modelo específico do domínio Bancos

Neste caso também temos uma predição bastante boa. Tomemos como exemplo novamente o valor 7 do rank previsto. Aqui podemos prever 7 casos, acertando ao lado em 2 casos abaixo e dois casos acima.

5.4.8.3 Domínio Aviação

Para este domínio temos uma amostra mais homogênea e foi novamente possível tirar algumas variáveis do fim, partindo do modelo reduzido.

Parameter Estimates		Estimate
Threshold	[rank = 4]	-4,077
	[rank = 5]	-3,685
	[rank = 6]	-1,64
	[rank = 7]	1,947
Location	links_num_extern_links	-0,131
	links_num_image_links	0,086
	text_heading_reverse_order	0,015
	links_links_title	-0,027
	text_num_diferent_colors	0,097
	pagelayout_device_specific	0,058
	pagelayout_num_tables	0,022
	accessibility_img_title	-5,908
	text_font_size_max_em	2,078
	forms_labels	1,231
	efficiency_css_size	-9,68E-06
	[area_ =1]	-2,979
	[area_ =2]	0,362
	[area_ =3]	-0,294
	[area_ =4]	-4,79
	[area_ =5]	-4,023
[area_ =6]	-4,489	
[area_ =7]	-1,94	
[area_ =8]	0 ^a	

Tabela 5-29 Parâmetros do modelo para domínio Aviação

Pseudo R-Square	
Cox and Snell	,923
Nagelkerke	1,000
McFadden	1,000

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	47,139	106	1,000
Deviance	35,444	106	1,000

Link function: Complementary Log-log.

Tabela 5-30 Fit do modelo para domínio Aviação

Como vemos, conseguimos construir modelos específicos para o domínio, como menos variáveis, e melhor *fit*.

5.4.9 Modelo de predição de probabilidade de ser um determinado PageRank

Vamos agora mostrar como se podem usar as variáveis dos modelos para prever o PageRank.

Revertendo a função *complementary log log*, conseguimos fazer um modelo de predição no Excel. Este modelo funciona do seguinte modo:

Na coluna *estimates*, preenchemos os predictores. Na coluna *current site*, pomos indicadores recolhidos. A coluna seguinte multiplica as duas anteriores. A soma desta coluna aparece na célula a amarelo.

Na coluna azul, pomos os *Thresholds*. A coluna rosa seguinte calcula para cada célula a diferença entre o *Threshold* acima e o somatório na célula amarela, conforme a equação 1.

Na linha da probabilidade acumulada, tiramos o inverso da função de ligação.

Como a função de ligação é:

$$y = \log(-\log(1-x))$$

Equação 3 – Complementary log-log

A sua inversa é:

$$x = 1 - e^{-e^y}$$

Equação 4 – Inversa da função de ligação Complementary log-log

Onde x é a probabilidade acumulada e y a função de ligação.

Podemos escrever a função completa como:

$$\log(-\log(1-x)) = \theta_j - [\beta_1 x_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ij}]$$

Equação 5 – Complementary log-log e equação generalizada

Na linha da probabilidade absoluta, é fazemos as diferenças e na célula “*result*” seleccionamos a maior probabilidade e fazemos um **hlookup** (pesquisa) da coluna de baixo, obtendo o PR esperado.

Mostramos em seguida as tabelas excel para predição do PR.

	P4	P5	P6	P7	P8	P9
Accumulated	1%	2%	6%	39%	83%	100%
Absolute prob.	1%	1%	4%	33%	44%	17%
	4	5	6	7	8	9

Thresholds ->	-1,91	-1,24	0,01	2,037	3,31
-------------------------	-------	-------	------	-------	------

2,74	-4,650	-3,980	-2,730	-,703	,570
------	--------	--------	--------	-------	------

	ESTIMATES	CURRENT SITE	
links_num_extern_links	-,001	35,9895	-0,0198
links_num_image_links	,023	12,5158	0,2893
text_heading_reverse_order	,002	11,6632	0,018
links_links_title	-,004	6,68421	-0,0268
text_num_diferent_colors	,017	70	1,1885
pagelayout_device_specific	-,009	0,19048	-0,0018
pagelayout_num_tables	-,021	9,17647	-0,191
accessibility_img_title	-,204	0	0
text_font_size_max_em	,108	0	0
forms_labels	,579	1	0,5789
efficiency_css_size	,000	46217,9	-0,1483
efficiency_page_size	,000	53923,7	0,2743
identity_author	,736	0	0
pagelayout_num_divs	-,007	75,0392	-0,4964
text_font_size_max_px	,002	44	0,083
[tipo=1]	,350	1	0,3502
[tipo=2]	-,324		0
[tipo=3]	0		0
[area_=1]	-,010		0
[area_=2]	-,304		0
[area_=3]	,842	1	0,842
[area_=4]	-,144		0
[area_=5]	-,662		0
[area_=6]	-,450		0
[area_=7]	1,031		0
[area_=8]	0		0

	RESULT
Per	44%
PR	8

Tabela 5-31 Excel para predição do page rank, modelo reduzido independente do domínio de aplicação, exemplo para usatodays.com.

	P4	P5	P6	P7	P8	P9	
Accumulated		1%	2%	5%	34%	77%	100%
Absolute prob.		1%	1%	4%	28%	43%	23%
	4	5	6	7	8	9	

Thresholds ->						
	-1,91	-1,24	0,01	2,037	3,31	

2,927646	-4,838	-4,168	-2,918	-,891	,382
----------	--------	--------	--------	-------	------

	ESTIMATES	CURRENT SITE	
links_num_extern_links	0,006	31,8667	0,1912
links_num_image_links	-0,008	5,36667	-0,04293
text_heading_reverse_order	0,002	7,63333	0,015267
links_links_title	-0,006	6,03333	-0,0362
text_num_diferent_colors	0,02	39	0,78
pagelayout_device_specific	0,022	21	0,462
pagelayout_num_tables	-0,013	4,06667	-0,05287
accessibility_img_title	-0,172	0,01481	-0,00255
text_font_size_max_em	-0,139	1,5	-0,2085
forms_labels	1,151	1	1,151
efficiency_css_size	-2,76E-06	223067	-0,61477
[area_=1]	-0,32		0
[area_=2]	-2,293		0
[area_=3]	1,286	1	1,286
[area_=4]	0,327		0
[area_=5]	-0,663		0
[area_=6]	-0,001		0
[area_=7]	0,898		0
[area_=8]	0		0

	RESULT
Per	43%
PR	8

Tabela 5-32 Excel para predição do page rank, modelo reduzido dependente do domínio de aplicação, exemplo para latimes.com.

5.5 Verificação do Modelo

Fizemos nova análise, que implica nova recolha de dados, e análise de características.

Analizamos para os sites New York Daily News (<http://www.nydailynews.com/>), conseguindo prever com sucesso o seu PageRank que é de 8. Para os sites US BANK (<http://www.usbank.com/>) e Lloyds TSB (www.lloydstsb.com) com um rank previsto de 7, que é o real. Analisamos ainda para US Airways (www.usairways.com) – PR de 7 e previsto de 7, e ainda para aviação, vimos o Jet2 www.jet2.com cujo PR é de 7 e conseguimos prever 7. Contudo esta amostra é pequena e não tão dispersa geograficamente como desejaríamos.

[Esta página foi intencionalmente deixada em branco]

Capítulo 6

Conclusões e trabalho futuro

6	Conclusões e trabalho futuro	96
6.1	Conclusões	96
6.2	Trabalho Futuro	97

O último capítulo contém as conclusões deste trabalho e identifica algumas vias para o trabalho futuro.

6 Conclusões e trabalho futuro

6.1 Conclusões

Nesta dissertação propomos a base para construir um modelo de qualidade de sítios web completamente automatizado. Conseguimos com sucesso recolher dados automaticamente de uma amostra composta de mais de 100 sítios médios e grandes. Destes dados conseguimos retirar automaticamente indicadores de qualidade, de acordo com o nosso modelo que tem mais de 60 indicadores, agrupados segundo o standard ISO 9126.

Como estamos preocupados com que indicadores deverão entrar no nosso modelo, do número elevado que recolhemos automaticamente, realizamos uma Análise de Componentes Principais. Contudo os seus resultados não deixaram reduzir consideravelmente a complexidade do nosso conjunto de dados para uma dimensão mais baixa.

A segunda hipótese desta tese, era a generalidade do modelo. Queríamos avaliar se o domínio da aplicação influenciava os valores dos nossos indicadores. Para este fim, o nosso *sample* estava dividido em 3 domínios distintos de aplicação: Jornais, Bancos e Aviação, de onde recolhemos os indicadores definidos no modelo de qualidade para sítios Web. Em seguida, fizemos uma análise de dados ANOVA extensiva. Nas tabelas apresentadas podemos observar que cerca de metade dos nossos indicadores são dependentes do domínio, isto é, há evidência estatística que o seu valor é influenciado pelo domínio da aplicação. Por exemplo observamos que a contudo da informação textual, varia bastante entre domínios. Isto deve-se em grande parte ao domínio dos jornais que põe grande ênfase na apresentação de texto.

A terceira hipótese desta tese, era se conseguíamos relacionar o PageRank do Google com os indicadores do modelo. O PageRank do Google, embora tenha um ênfase muito grande nos *links* de rede, é de alguma forma uma medida de qualidade dos

sites. Fizemos uma análise logística ordinal e conseguimos fazer um modelo que relaciona, mas que no entanto era muito grande (continha muitas variáveis). Através de análise de correlação, conseguimos diminuir o número de variáveis para menos de uma dezena, onde o modelo manteve a mesma correlação. Este modelo é instanciado para os 3 diferentes domínios de aplicação, devido às conclusões que tiramos: metade dos indicadores são dependentes do domínio de aplicação

Dividimos o nosso modelo de qualidade para sítios web, considerando as duas componentes, uma universal, e outra dependente do domínio.

6.2 Trabalho Futuro

Com base neste estudo preliminar, pretendemos aperfeiçoar o nosso modelo de qualidade de site, considerando as duas componentes: a "universal" (inclui indicadores independentes do domínio), e a outra que deve ser personalizadas para um domínio específico (indicadores dependentes do domínio).

Para aumentar a validade externa de nosso modelo que pretendemos estender a recolha de dados para outros domínios de aplicação. Para melhorar a validade interna pretendemos recolher uma amostra maior.

Por último, mas não menos importante, queremos usar nossos indicadores de qualidade, como variáveis explicativas de um índice de classificação. O nosso próximo passo aqui será obter *ranks* produzidos por especialistas humanos em diversos domínios de aplicação. Iremos usar regressão logística ordinal para calibrar (obter os coeficientes) do nosso modelo de qualidade para cada um desses domínios.

A ISO9126 será em breve substituída pela ISO25010, que actualmente é um *Full Comited Draft* [30]. Planeamos rever o conjunto de métricas de acordo com as características propostas por esta nova norma. A norma supostamente introduz duas novas características: segurança e compatibilidade. Enquanto podemos medir a compatibilidade e esta é de algum modo relacionada com a portabilidade, a segurança será mais problemática para medir automaticamente.

[Esta página foi intencionalmente deixada em branco]

Bibliografia

- [1] C. Calero, *et al.*, Eds., *Handbook of Research on Web Information Systems Quality*. Information Science Reference (IGI Global), 2008, p.^pp. Pages.
- [2] G. Rossi, *et al.*, Eds., *Web Engineering: Modelling and Implementing Web Applications* (Human-Computer Interaction Series. London: Springer, 2008, p.^pp. Pages.
- [3] G. Kappel, *et al.*, Eds., *Web Engineering: The Discipline of Systematic Development of Web Applications*. John Wiley & Sons, 2006, p.^pp. Pages.
- [4] A. Navarro, "A SWEBOK-based Viewpoint of the Web Engineering Discipline," *j-jucs*, vol. 15, pp. 3169--3200, 2009.
- [5] W. Werner, *et al.*, "Web Engineering for Intranets: Rethinking Software Engineering," 2002.
- [6] A. Abran, *et al.*, Eds., *Guide to the Software Engineering Body of Knowledge (SWEBOK)*. IEEE Computer Society, 2004, p.^pp. Pages.
- [7] A. Ginige, "Web engineering: managing the complexity of web systems development," 2002, pp. 721--729.
- [8] E. Mendes and N. Mosley, Eds., *Web Engineering*. Springer, 2005, p.^pp. Pages.
- [9] A. Ginige and S. Murugesan, "Guest Editors Introduction: The Essence of Web Engineering— Managing the Diversity and Complexity of Web Application Development," *IEEE MultiMedia*, vol. 8, pp. 22--25, 2001.
- [10] Google. (2010, 1st April). *Google Apps*. Available: www.google.com/apps/
- [11] S. M. Dias, *et al.*, "Portuguese Local e-Government," in *Fifth International Conference on Enterprise Information Systems (ICEIS)*, Angers, 2003.
- [12] M. Y. Ivory and R. Megraw, "Evolution of web site design patterns," *ACM Transactions on Information Systems (TOIS)*, vol. 23, pp. 463-497, 2005.
- [13] L. Olsina, *et al.*, "Web Quality," in *Web Engineering*, E. Mendes and N. Mosley, Eds., ed: Springer 2005, pp. 109-142.
- [14] L. Olsina and G. Rossi, "Measuring Web Application Quality with WebQEM," *IEEE MultiMedia*, vol. 9, pp. 20-29, 2002.
- [15] L. Olsina and G. Rossi, "Toward Web-site Quantitative Evaluation: Defining Quality Characteristics And Attributes," presented at the WebNet World Conference on the WWW and Internet (WEBNET), Honolulu, Hawaii, 1999.
- [16] L. Olsina, *et al.*, "Specifying Quality Requirements for the Web 2.0 Applications," presented at the 7th International workshop on Web-oriented software technology, New York, 2008.
- [17] T. A. Powell, *et al.*, *Web site engineering: beyond Web page design*: Prentice-Hall, Inc., 1998.

- [18] L. Mich, *et al.*, "Evaluating and Designing the Quality of Web Sites," *IEEE MultiMedia*, vol. 10, pp. 34-43, 2003.
- [19] D. B. N., *et al.*, "Evaluating The Features Of Museum Websites: (The Bologna Report)," in *Museums and the Web, 2002*, Boston, 2002, p. 11.
- [20] Working_Group_5, *Quality Principles for Cultural Websites: a Handbook*: Minerva Project, 2003.
- [21] R. Polillo, "Un modello di qualità per i siti web," *Mondo Digitale* vol. June, pp. 32-44, 2005.
- [22] O. Signore, "A Comprehensive Model for Web Sites Quality," presented at the Seventh IEEE International Symposium on Web Site Evolution (WSE'2005), Budapest, Hungary, 2005.
- [23] C. Calero, *et al.*, "Classifying web metrics using the web quality model," *Online Information Review*, vol. 29, pp. 227-248, 2005.
- [24] S. Cimino and F. Micali, "Web Q-Model: a new approach to the quality," presented at the The 26th Annual CHI Conference on Human Factors in Computing Systems (CHI 2008), Florence, Italy, 2008.
- [25] I. Jacobs, *et al.*, "Web Content Accessibility Guidelines 1.0," techreport1999.
- [26] *ISO/IEC 9126: Information Technology - Software Product Evaluation - Software Quality Characteristics and Metrics*, 2001.
- [27] I. Jacobs, *et al.*, "HTML 4.01 Specification," techreport1999.
- [28] S. Pemberton, "XHTML® 1.0: The Extensible HyperText Markup Language - A Reformulation of HTML 4 in XML 1.0," techreport2000.
- [29] D. Hyatt and I. Hickson, "HTML 5," techreport2009.
- [30] *Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality models - FCD*, ISO, 2010.

Anexo 1 – Caracterização da amostra

As tabelas apresentam a amostra total que analisamos, incluindo as ligações e fontes.

País	Nome	Sítio
USA	USA Today	http://www.usatoday.com/
USA	The Wall Street Journal	http://europe.wsj.com/home-page
USA	The New York Times	http://www.nytimes.com/
USA	Los Angeles Times	http://www.latimes.com/
UK	The Sun	http://www.thesun.co.uk/sol/homepage/
UK	Daily Mail	http://www.dailymail.co.uk/home/index.html
UK	Daily Mirror	http://www.mirror.co.uk/
UK	The Daily Telegraph	http://www.telegraph.co.uk/
<i>Fontes</i> http://en.wikipedia.org/wiki/List_of_newspapers_in_the_world_by_circulation		
Brasil	O Estado de São Paulo	http://www.estado.com.br/
Brasil	Jornal do Brasil	http://www.jb.com.br/
Brasil	O Globo	http://www.oglobo.com.br/
Argentina	La Nación	http://www.lanacion.com.ar/
Venezuela	El Nacional	http://www.el-nacional.com
Venezuela	El Universal	http://www.el-universal.com/
Venezuela	Diario 2001	http://www.2001.com.ve/
<i>Fontes</i> http://www.worldpress.org/newspapers/AMERICAS/Brazil.cfm http://www.worldpress.org/newspapers/AMERICAS/Argentina.cfm http://www.worldpress.org/newspapers/AMERICAS/Venezuela.cfm		
Alemanha	Bild	www.bild.de
Alemanha	Süddeutsche Zeitung	www.sueddeutsche.de
Alemanha	Die Zeit	http://www.zeit.de
França	Ouest-France	http://www.ouest-france.fr/
França	La Tribune	http://www.latribune.fr/accueil/a-la-une.html
França	Le Parisien	http://www.leparisien.fr/actualites-informations-direct-videos-parisien
Bélgica	Gazet van Antwerpen	www.gva.be
Bélgica	Grenz-Echo	www.grenzecho.be
Bélgica	Le Soir en ligne	www.lesoir.com
Holanda	De Telegraaf	http://www.telegraaf.nl/
Austria	Kronen Zeitung	http://www.krone.at/krone/S1/kmprog/index.html
<i>Fontes</i> http://en.wikipedia.org/wiki/List_of_newspapers_in_the_world_by_circulation http://en.wikipedia.org/wiki/List_of_German_newspapers http://en.wikipedia.org/wiki/List_of_French_newspapers		

País	Nome	Sítio
		http://dir.yahoo.com/News_and_Media/Newspapers/By_Region/Countries/Belgium/
Portugal	Diário de Notícias	www.dn.pt
Portugal	Jornal de Notícias	www.jnoticias.pt
Portugal	O Público	http://www.publico.pt/
Espanha	El País	www.elpais.es
Espanha	El Mundo	http://www.el-mundo.es/
Espanha	ABC	http://www.abc.es/
Italia	Corriere della Sera	http://www.corriere.it/
Italia	La Repubblica	http://www.repubblica.it/
Fontes	http://dir.yahoo.com/News_and_Media/Newspapers/By_Region/Countries/Portugal/ http://worldpress.org/newspapers/EUROPE/Portugal.cfm http://worldpress.org/newspapers/EUROPE/Spain.cfm http://en.wikipedia.org/wiki/List_of_newspapers_in_the_world_by_circulation	
Noruega	Aftenposten	http://www.aftenposten.no/
Noruega	Dagbladet	http://www.dagbladet.no/
Dinamarca	Århus Stiftstidende	http://www.stiften-aarhus.dk/
Dinamarca	Børsen	http://www.borsen.dk/
Suécia	Expressen	http://www.expressen.se/
Fontes	http://worldpress.org/newspapers/EUROPE/Norway.cfm http://worldpress.org/newspapers/EUROPE/Denmark.cfm http://worldpress.org/newspapers/EUROPE/Sweden.cfm	
China	People's Daily	http://www.people.com.cn/
China	Guangzhou Daily	http://gzdaily.dayoo.com/html/2009-05/11/node_1.htm
China	Global Times	http://en.huanqiu.com/www/english/china/ChinaNews/index.html
Japan	Yomiuri Shimbun	http://www.yomiuri.co.jp/
Japan	Asahi Shimbun	http://www.asahi.com/english/
Japan	Mainichi Shimbun	http://mdn.mainichi.jp/
India	Anandabazar Patrika	http://www.anandabazar.com/
India	The Times of India	http://timesofindia.indiatimes.com/default1.cms
India	Malayala Manorama	http://www.manoramaonline.com
Fontes	http://en.wikipedia.org/wiki/List_of_newspapers_in_the_world_by_circulation	
Angola	Jornal de Angola	http://www.jornaldeangola.com/
Nigeria	Daily Champion	http://www.champion-newspapers.com/
Nigeria	Nigerian Tribune	http://www.tribune.com.ng/
Nigeria	This Day	http://www.thisdayonline.com/
África do Sul	The citizen	http://www.citizen.co.za/
África do Sul	Daily News	http://www.dailynews.co.za/
Fontes	http://worldpress.org/newspapers/AFRICA/Angola.cfm http://worldpress.org/newspapers/AFRICA/Nigeria.cfm http://worldpress.org/newspapers/AFRICA/South_Africa.cfm	
Egipto	Al-Ahram	http://www.ahram.org.eg/
Egipto	Al Gomhuriya	www.algomhuria.net.eg/
Tunisia	La Presse	http://www.lapresse.tn
Tunisia	Le quotidien	www.tunisie.com/lequotidien
Fontes	http://en.wikipedia.org/wiki/List_of_newspapers_in_the_world_by_circulation http://worldpress.org/newspapers/AFRICA/Tunisia.cfm	

Tabela A-0-1 Lista de jornais

País	Nome	Sítio
USA	Bank of America	www.bankofamerica.com
USA	JPMorgan Chase	www.jpmorganchase.com
USA	Citigroup	www.citigroup.com
USA	Wells Fargo	www.wellsfargo.com
UK	HSBC Holdings	www.hsbc.com
UK	RBS	www.rbs.com
UK	Standard Chartered	www.standardchartered.com
UK	Barclays	www.barclays.com
Fontes	http://skorcareer.com.my/blog/10-worlds-largest-banks-list-in-2008/2008/06/18/zib/bloomberg	

País	Nome	Sítio
Brasil	Itau	http://www.itau.com.br/
Brasil	BB	http://www.bb.com.br/portalbb/home/geral/index.bb
Brasil	Bradesco	http://www.bradesco.com.br/
Argentina	Banco de Galicia y Buenos Aires	http://www.e-galicia.com/
Venezuela	Mercantil Banco Universal	http://www.bancomercantil.com/
Venezuela	Banco Exterior	http://www.bancoexterior.com/
Venezuela	Banco Federal	http://www.bancofederal.com/
Fontes	http://www.bcb.gov.br/fis/top50/port/ArquivoZip.asp http://www.quazell.com/bank/both_arg.htm http://www.quazell.com/bank/both_ven.htm	
Alemanha	Commerzbank	www.commerzbank.com
Alemanha	Postbank	www.postbank.de
Alemanha	Deutsche Bank	www.db.com
França	BNP Paribas	www.bnpparibas.com
França	Société Générale	www.ir.socgen.com
França	Credite Agricole	www.credit-agricole-sa.fr
Bélgica	KBC Groep	www.kbc.com
Bélgica	Dexia	www.dexia.com
Bélgica	Fortis	www.fortis.com
Fontes	zib/bloomberg	
Portugal	Banco BPI	www.bpi.bancobpi.pt
Portugal	BCP	www.millenniumbcp.pt
Portugal	BES	www.bes.pt
Espanha	Banco Popular	www.bancopopular.es
Espanha	BBVA	www.bbva.es
Espanha	Santader	www.santander.com
Italia	Cassa di Risparmio di Firenze	www.carifirenze.it/
Fontes	zib/bloomberg http://www.google.com/Top/Business/Financial_Services/Banking_Services/Banks_and_Institutions/Regional/Europe/Italy/	
Noruega	Sparebanken NOR	www.nor.no/
Noruega	Norges Bank	www.norges-bank.no/english/
Dinamarca	Amagerbanken	www.amagerbanken.dk/
Dinamarca	Jyske Bank Private Banking	www.jbpb.com
Suécia	Swedbank	www.swedbank.com/
Fontes	http://www.google.com/Top/Business/Financial_Services/Banking_Services/Banks_and_Institutions/Regional/Europe/Norway/ http://www.google.com/Top/Business/Financial_Services/Banking_Services/Banks_and_Institutions/Regional/Europe/Denmark/ http://www.google.com/Top/Business/Financial_Services/Banking_Services/Banks_and_Institutions/Regional/Europe/Sweden/	
China	Industrial & Commercial Bank of China, ICBC	http://www.icbc.com.cn/icbc/sy/default.htm
China	China Construction Bank of China	http://www.ccb.com/portal/en/home/index.html
China	Bank of China	http://www.boc.cn/en/
Japan	Mitsubishi UFJ Financial, MUFG (Japan)	http://www.mufg.jp/english/
Fontes	http://skorcareer.com.my/blog/10-worlds-largest-banks-list-in-2008/2008/06/18/	
Angola	Banco Fomento de Angola (BFA)	http://www.bfa.ao/
Nigeria	Zenith Bank	http://www.zenithbank.com/index.cfm
Nigeria	UBA United Bank for Africa	http://www.ubagroup.com/
Nigeria	Ecobank	http://www.ecobank.com/english/others/home.aspx
África do Sul	The Standard Bank of South Africa	http://www.standardbank.co.za/
África do Sul	ABSA	http://www.absa.co.za/absacoza/
Fontes	http://uk.reuters.com/article/rbssFinancialServicesAndRealEstateNews/idUKL1427529820080714 http://www.zenithbank.com/bestglobalbank.cfm http://www.google.com/Top/Regional/Africa/Nigeria/Business_and_Economy/Financial_Services/Banks/ http://www.google.com/Top/Regional/Africa/South_Africa/Business_and_Economy/Financial_Services/Banks/	
Egipto	National Bank Of Egypt	www.nbe.com.eg/
Egipto	Bank of Alexandria	www.alexbank.com
Marrocos	Bank Al-Maghrib	http://www.bkam.ma/
Marrocos	Attijariwafa Bank	http://www.attijariwafabank.com/Pages/default.aspx
Marrocos	Banque Centrale Populaire	http://www.gbp.ma/
Fontes	http://www.google.com/Top/Regional/Africa/Egypt/Business_and_Economy/Financial_Services/Banks/ http://internationalbusiness.wikia.com/wiki/Morocco%27s_Largest_Banks	

Tabela A-0-2 Lista de Bancos

País	Nome	Sítio
USA	Southwest Airlines	http://www.southwest.com/
USA	American Airlines	http://www.aa.com/
USA	Delta Air Lines	http://www.delta.com/
USA	United Airlines	http://www.united.com/
Canada	Air Canada	www.aircanada.com
Irlanda	Ryanair	http://www.ryanair.com/
UK	easyJet	http://www.easyjet.com/
UK	British Airways	http://www.britishairways.com
UK	BMI	http://www.flybmi.com
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_North_America http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Europe
Brasil	TAM Airlines	http://www.tam.com.br/
Brasil	Gol Transportes Aéreos	http://www.voegol.com.br/Paginas/home.aspx
Brasil	Varig	http://portal.varig.com.br/br/varig/varig/index.html
Argentina	Aerolíneas Argentinas	http://www.aerolineas.com.ar/home.asp
Colômbia	Avianca	http://www.avianca.com/
Chile	LAN Airlines	http://www.lan.com/cgi-bin/country_selector.cgi
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_South_America
França	Air France	http://www.airfrance.fr/
Holanda	KLM	http://www.klm.com/travel/gb_en/index_default.html
Alemanha	Lufthansa	http://www.lufthansa.de/
Alemanha	Air Berlin	http://www.airberlin.com/
Suíça	Swiss International Air Lines	http://www.swiss.com
Belgica	Brussels Airlines	brusselsairlines.com
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Europe
Espanha	Iberia Airlines	www.iberia.com
Itália	Alitalia	www.alitalia.com
Espanha	Air Europa	http://www.aireuropa.com
Portugal	TAP Portugal	http://www.flytap.com/
Itália	Air One	http://www.flyairone.it/
Espanha	Vueling Airlines	http://www.vueling.com/
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Europe
Dinamarca, Suécia, Noruega	SAS Group	http://www.sasgroup.net/
Noruega	Norwegian Air Shuttle	http://www.norwegian.no
Finlândia	Finnair	http://www.finnair.com
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Europe
China	China Southern Airlines	http://www.flychinasouthern.com/
China	China Eastern Airlines	http://www.ce-air.com/
China	Air China	http://www.airchina.com.cn/index.shtml
Japão	Japan Airlines	http://www.jal.co.jp/en/
Japão	All Nippon Airways	http://www.ana.co.jp/
Singapura	Singapore Airlines	http://www.singaporeair.com/saa/index.jsp
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Asia
África do Sul	South African Airways	http://www4.flysaa.com/
Angola	TAAG Angola Airlines	http://www.taag.com.br/
Kenia	Kenya Airways	http://www.kenya-airways.com/home/
Etiopia	Ethiopian Airlines	http://www.ethiopianairlines.com/
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Africa http://en.wikipedia.org/wiki/List_of_airlines_of_Angola
Egipto	EgyptAir	http://www.egyptair.com/English/Pages/splashpage.aspx
Marrocos	Royal Air Maroc	http://www.royalairmaroc.com/
Tunisia	Tunisair	http://www.tunisair.com/publish/home.asp
Marrocos	Atlas Blue	http://www.atlas-blue.com/
	Fontes	http://en.wikipedia.org/wiki/List_of_largest_airlines_in_Africa

Tabela A-0-3Lista de sítios de Companhias de Aviação

Anexo 2 Introdução à Engenharia de Software experimental

A Estatística permite retirar uma informação valiosa dos nossos dados, usando técnicas próprias e conseqüentemente melhorando o processo de tomada de decisão.

Tipos de variáveis

As variáveis podem ser classificadas em independentes e dependentes. As variáveis independentes, também designadas factores ou variáveis explicativas, são as que podem ser manipuladas na investigação experimental e consoante o número de variáveis, comportam análises unifactoriais (com apenas uma variável independente) ou multifactoriais (com várias variáveis independentes). As variáveis dependentes são aquelas cujo efeito das independentes pretendemos medir.

Existem quatro tipos de escalas: nominal, ordinal, intervalo e razão. Para as escalas nominal e ordinal (valores discretos) são utilizados testes não paramétricos e para as escalas métricas (valores contínuos), ou seja, em intervalo e de razão podem também ser utilizados os testes paramétricos, se as variáveis tiverem uma distribuição normal.

Quando uma variável independente está definida numa escala de valores discretos, cada um desses valores é habitualmente designado de categoria.

		Valores discretos		Valores Contínuos	
Tipo de variável		Nominal	Ordinal	Intervalo	Razão
Definições	Existe uma ordem significativa?	Não	Sim	Sim	Sim
	Iguais distâncias entre os valores da escala têm significados iguais?	Não	Não	Sim	Sim
	O cálculo da razão faz sentido?	Não	Não	Não	Sim
Operações autorizadas	Média			X	X
	Mediana		X	X	X
	Moda	X	X	X	X
	Amplitude		X	X	X
	Variância			X	X
	Desvio padrão			X	X

Tabela B-0-1 Resumo das definições e do tipo de operações autorizadas por tipo variável

Aderência à distribuição

A distribuição normal é importante porque em muitos casos, representa de forma aproximada a relação entre a “magnitude” e a “significância” de relações entre duas variáveis, dependendo do tamanho da amostra.

O tipo de distribuição condiciona os testes que podem ser aplicados, por isso queremos saber se uma variável segue (adere) uma dada distribuição estatística. Muitas vezes estamos interessados na forma como a distribuição pode ser aproximada pela distribuição normal, para tal podemos utilizar várias abordagens como a estatística descritiva, gráficos e testes de aderência à distribuição. Os testes de aderência à distribuição mais utilizados são:

Kolmogorov-Smirnov – Teste baseado na diferença máxima entre a distribuição cumulativa da amostra e a distribuição cumulativa hipotética. Testa $H_0: X \sim N(\mu; \sigma)$ e $H_1: \neg X \sim N(\mu; \sigma)$, se a estatística Z é significativa, então a hipótese de que a respectiva distribuição é normal (H_0) deve ser rejeitada.

Teste de Lilliefors - Este teste é basicamente uma correção para o teste de Kolmogorov-Smirnov, aplicável quando os parâmetros da distribuição normal hipotética são estimados a partir da amostra de dados, se a estatística Z é significativa, então a hipótese de que a respectiva distribuição é normal deve ser rejeitada (o mesmo que para o teste de KS).

Teste de Shapiro-Wilks' W - É o teste preferido da normalidade devido às suas propriedades mais poderosas em relação a uma ampla gama de métodos alternativos, se a estatística W é significativa (ou seja, $p \leq \alpha$), então a hipótese de que a respectiva distribuição é normal deve ser rejeitada.

O valor-p, significância estatística de um resultado, representa a probabilidade de erro que está envolvido em aceitar o resultado observado como válido, isto é, como representativo da população. Um valor-p de 5% (ou seja, 1/20) indica que há uma probabilidade de 5% que a relação entre as variáveis encontradas na nossa amostra é uma casualidade. Para os testes de aderência, o valor-p é a probabilidade de que a diferença observada entre a distribuição cumulativa da amostra e a distribuição cumulativa hipotética ocorrerem por casualidade. Os valores mais utilizados são 5% (1/20) para a linha de fronteira estatisticamente significativa, 1% (1/100) para estatisticamente significativo e 0.5% (1/200) ou 0.1% (1/1000) para altamente significativo.

Testes de Hipóteses

Os testes de hipóteses consideram duas hipóteses, a nula H_0 , na qual definimos o critério que pretendemos, e a hipótese alternativa H_1 com o critério contrário à H_0 .

A hipótese nula pode ser verdadeira ou falsa, assim como ser aceite ou rejeitada. Existem dois tipos de erros associados à decisão, um erro do tipo I que ocorre quando a hipótese nula é rejeitada, apesar de ser verdadeira, sendo que o α é usado para representar a probabilidade de ocorrência de um erro do tipo I e um erro do tipo II que ocorre quando a hipótese nula não é rejeitada, apesar de ser falsa, sendo o β é usado para representar a probabilidade de ocorrência de um erro do tipo II.

A potência de um teste de hipóteses é a probabilidade $(1 - \beta)$ de rejeitar a hipótese nula quando é falsa, a qual deve ser elevada, devendo ser no mínimo 0.80.

Métodos de análise de dados – testes estatísticos

Existem diversos métodos de análise, mais concretamente:

- Teste de Proporção
- Testes de Inferência para dados discretos e contínuos
 - Testes paramétricos
 - Testes não-paramétricos
- A análise de regressão
 - Regressão linear simples
 - Modelo de regressão não-linear (por exemplo, análise de regressão logística)
- Análise multivariada de dados
 - Análise factorial
 - Análise de clusters
 - Análise discriminante

Testes de Inferência para dados discretos e contínuos

Testes paramétricos

Os testes paramétricos garantem uma validação mais forte, a sua potência $(1-\beta)$ é maior do que para os testes não-paramétricos. Podem ser aplicados para análise entre grupos ou dentro de grupos.

t-Student (uma amostra) – Este método é aplicado se a média de uma variável quantitativa difere de um valor do teste de hipótese (o valor de teste é uma constante especificada). Este método pressupõe que os dados da variável resultante são normalmente distribuídos, porém, é bastante robusto para desvios da normalidade. Testa-se a $H_0: \mu = k$ (a média da variável não difere significativamente do valor K) e $H_1: \mu \neq k$ (a média da variável difere significativamente do valor k). Quando n , graus de liberdade, é inferior ou igual a trinta, rejeitamos H_0 , para um nível de significância α , se $|T_{\text{calc}}| > t_{1-\alpha/2} (n-1)$. Por outro lado, quando o n é superior a trinta a distribuição t-student transforma-se em normal $N(0;1)$; se $p \leq \alpha$, rejeitamos H_0 e aceitamos H_1 , sendo as proporções diferentes, mas se $p > \alpha$, aceitamos H_0 pois as proporções são similares.

t-Student (duas amostras independentes) – Este teste permite inferir a igualdade de médias nas populações de duas amostras (grupos). Os participantes devem ser divididos aleatoriamente em dois grupos, de modo que qualquer diferença na resposta é devido aos tratamentos e não a outros factores. Este teste usa estatísticas diferentes, dependendo do resultado ter variâncias homogêneas ou não-homogênea nos

dois grupos. Esta homogeneidade pode ser avaliada com o teste de Levene. Definimos as hipóteses $H_0: \mu_A = \mu_B$ (as médias das variáveis em cada grupo, definidos por cada tratamento, são os mesmos) e $H_1: H_0: \mu_A \neq \mu_B$ (as médias das variáveis em cada grupo não são iguais). Quando $n \leq 30$ (graus de liberdade), rejeitamos H_0 , para um nível de significância α , se $|T_{\text{calc}}| > t_{1-\alpha/2 (n-1)}$. Por outro lado, quando $n > 30$, a distribuição t-student transforma-se em normal $N(0;1)$; se $p \leq \alpha$, rejeitamos H_0 e aceitamos H_1 , sendo as médias diferentes, mas se $p > \alpha$, aceitamos H_0 pois as médias são similares.

ANOVA One-Way (Análise de Variância Unifactorial) – Este procedimento é usado para testar se as médias (determinadas por um factor variável) entre vários grupos são iguais. Por isso, permite testar se existe uma variância na variável resultante que é devida ao factor. Esta é uma extensão do teste-t com duas amostras. A validade dos resultados da ANOVA pressupõe que os dados cumprem as seguintes condições: seleccionam-se amostras aleatórias das populações, de forma a garantir a independência e aleatoriedade dos erros e as populações devem ter distribuição normal e variâncias iguais (homogéneas). As hipóteses que são testadas são H_0 (as médias de cada grupo são iguais), isto é, $\forall i, j: \mu_i = \mu_j (i \neq j)$ e H_1 (as médias de cada grupo não são iguais), ou seja, $\exists i, j: \mu_i \neq \mu_j (i \neq j)$. Assim, rejeitamos H_0 para um determinado nível de significância α , se $F_{\text{calc}} > F_{1-\alpha (k-1, n-k)}$. Sendo K o nº de grupos e n o nº total de dados no estudo (utiliza-se a tabela de distribuição F). Para determinar quais os grupos, cujas médias diferem podem ser utilizados dois tipos de testes, contrastes a priori e post hoc.

ANOVA Factorial (Análise de Variância Multifactorial) - Este procedimento é usado para testar se um determinado conjunto de factores tem um efeito significativo sobre uma determinada variável, permitindo determinar o efeito de cada factor e avaliar a interacção entre os factores (moderação). Este é um caso particular de uma metodologia de análise de regressão multivariada, denominada "General Linear Model" (GLM). Se considerarmos 3 factores, F1, F2 e F3, os efeitos principais são os gerados por cada factor isolado, tal como na ANOVA Unifactorial, representados por F1, F2 e F3, o efeitos da interacção são os gerados pelo acção combinada de todas as combinações de factores, isto é, F1*F2, F1*F3, F2*F3, F1*F2*F3, e o efeito global GML é representado por $I + F1 + F2 + F3 + F1*F2 + F1*F3 + F2*F3 + F1*F2*F3$, sendo I um termo de intercepção, semelhante ao utilizado na regressão linear. Quanto aos testes de hipóteses, são testadas hipóteses para cada factor e para cada interacção. Para cada factor, H_0 (as médias esperadas de cada grupo são iguais), isto é, $\mu_1 = \dots =$

μ_k e H_1 (pelo menos uma média é diferente), $\exists i, j: \mu_i \neq \mu (i \neq j)$. Assim, rejeitamos H_0 para um determinado nível de significância, α , se $F_{\text{calc}} > F_{1-\alpha (k-1, n-k)}$. Para cada interação, H_0 (não há interação entre os factores), isto é, $\forall i, j: \gamma_{i,j} = 0 (i \neq j)$ e H_1 (existe interação pelo menos entre 2 factores), $\exists i, j: \gamma_{i,j} \neq 0 (i \neq j)$. Assim, rejeitamos H_0 para um determinado nível de significância, α , se $F_{\text{calc}} > F_{1-\alpha (k-1, n-k)}$. Para aumentar o poder do teste, as populações devem ser normais e com variâncias homogêneas, no entanto, o ANOVA multifactorial é robusto a desvios da normalidade, embora os dados devem ser simétricas e quanto à variância, existem alternativas para a utilização deste método quando a homogeneidade não é assumida. Quando existem diferenças entre grupos, a estatística F global permite testar que pelo menos um grupo tenha a média diferente dos outros grupos e se no efeito global mostrou significância, usamos os testes post hoc para avaliar as diferenças entre as médias específicas, alguns destes testes post hoc são aplicados quando as variâncias iguais são assumidas e outros quando não são.

Testes não-paramétricos (entre os grupos)

Teste Binomial (teste de proporções) - Este teste permite comparar as proporções de ocorrência de um dos dois possíveis valores de uma variável dicotómica sobre o número total de casos. A probabilidade de ocorrência de cada um dos valores possíveis da variável dicotómica segue uma distribuição Binomial. Se a amostra for suficientemente grande ($n > 30$), a Binomial aproxima de uma distribuição Normal, neste caso podemos usar a estatística do teste Z: $Z \sim N(0, 1)$. A decisão baseia-se no intervalo de rejeição bilateral, que depende do valor de α , $]-\infty, -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}, +\infty[$, então se Z pertence ao intervalo de rejeição, rejeitamos H_0 e aceitamos H_1 (as proporções são diferentes), mas se Z está fora do intervalo de rejeição, aceitamos H_0 e rejeitamos H_1 (proporções são semelhantes). Definimos as hipóteses, $H_0: p_x = p_0, p_y = 1 - p_0$, isto é, não há evidência estatística de que as proporções esperadas na população são as que estão sendo testados, e $H_1: p_x \neq p_0, p_y \neq 1 - p_0$, isto é, as proporções esperadas na população são significativamente diferentes das testadas. Na tomada de decisão do teste, se $p \leq \alpha$ (estatística Z significativa), rejeitamos H_0 e aceitamos H_1 , se $p > \alpha$ (estatística Z não significativa), aceitamos H_0 e rejeitamos H_1 .

Qui-Quadrado (teste de proporções) - Este teste compara as frequências observadas e esperadas em cada categoria, para testar que todas as categorias que contenham a mesma proporção dos valores e também pode testar se cada categoria contém uma proporção especificada de valores. O teste Qui-Quadrado pode ser usado para testar se duas ou mais amostras independentes (grupos) diferem relativamente a um determinado factor. O Qui-Quadrado opera numa tabela de contingência, em que as linhas e colunas representam as categorias das duas variáveis e cada célula contém o número de observações para um dado par de valores (factor, variável resultante). As condições prévias do Qui-Quadrado são: a amostra deve ser suficientemente grande ($n > 20$), todos os valores de contingência devem ser > 1 e pelo menos 80% dos valores de contingência devem ser > 5 . Definimos as hipóteses, H_0 : as proporções esperadas dos grupos (da população) são semelhantes (os grupos não diferem significativamente no tamanho e o efeito do factor é insignificante) e H_1 : as proporções dos grupos são diferentes (os grupos diferem significativamente no tamanho e o efeito do factor não é desprezável). Na tomada de decisão do teste, se $p \leq \alpha$ (estatística Z significativa), rejeitamos H_0 e aceitamos H_1 (as proporções são diferentes), mas se $p > \alpha$ (estatística Z não significativa), aceitamos H_0 e rejeitamos H_1 (as proporções são semelhantes).

Teste Mann-Whitney (teste-U) - Este teste é o equivalente não-paramétrico para o teste-t, em vez de comparar a média das duas amostras, compara sua tendência central para detectar diferenças. Este pode ser usado para testar se duas amostras diferem relativamente a um determinado factor. As duas amostras testadas devem ser similares na forma. Definimos as hipóteses H_0 : as duas populações de onde as amostras para os dois grupos foram tomadas, têm tendência central similar (os grupos não são afectados pelo factor variável) e H_1 : as duas amostras não têm tendência central semelhante (os grupos são afectados pelo factor variável), para testar as hipóteses é utilizada a estatística U.

Teste Kruskal-Wallis (H-test) - É uma extensão do teste-U, sendo o equivalente não-paramétrico para a análise de variância unifactorial. Avalia se várias amostras independentes têm um parâmetro comum de localização, em que cada amostra é um grupo de indivíduos que corresponde à aplicação de um determinado tratamento (a nível do factor variável). Assim como no teste-U, as duas amostras testadas devem ser similares na forma. Definem-se as hipóteses H_0 : a distribuição das populações de onde foi extraído cada grupo têm o mesmo parâmetro de localização (os grupos não diferem

significativamente e o efeito do factor é insignificante) e H_1 : pelo menos uma das distribuições tem um parâmetro de localização que é menor ou maior do que os outros (pelo menos grupo difere significativamente e o efeito do factor não é desprezável). A distribuição estatística calculada do teste-H é aproximada à do Qui-quadrado, a partir de uma tabela do Qui-quadrado com um dado df (graus de liberdade) e de um nível de significância α fixado (probabilidade de um erro do tipo I), obtemos um valor crítico do Qui-Quadrado para ser comparado com a estatística H calculada. Na tomada de decisão do teste, rejeitamos H_0 , para um dado nível de significância α , se $H_{\text{calc}} > H_{(1-\alpha, df)}$.

ANOVA Factorial não-paramétrico - Este procedimento é usado para testar se um determinado conjunto de factores tem um efeito significativo sobre uma determinada variável, permitindo determinar o efeito de cada factor e avaliar a interacção entre os factores (moderação). Este procedimento é semelhante ao ANOVA Factorial (paramétrico), mas a estatística-H é calculada com base nas categorias dos casos em cada grupo, enquanto que na versão paramétrica é utilizada a estatística-F, que é calculada sobre os valores da variável resultante. Se considerarmos 3 factores, F_1 , F_2 e F_3 , os efeitos principais são os gerados por cada factor isolado, representados por F_1 , F_2 e F_3 , o efeitos da interacção são os gerados pelo acção combinada de todas as combinações de factores, isto é, $F_1 * F_2$, $F_1 * F_3$, $F_2 * F_3$, $F_1 * F_2 * F_3$, e o efeito global GML é representado por $I + F_1 + F_2 + F_3 + F_1 * F_2 + F_1 * F_3 + F_2 * F_3 + F_1 * F_2 * F_3$, sendo I um termo de intercepção, semelhante ao utilizado na regressão linear. Para a avaliação dos efeitos principais, definem-se as hipóteses H_0 : a distribuição das populações de onde foi extraído cada grupo têm o mesmo parâmetro de localização (os grupos não diferem significativamente e o efeito do factor é insignificante) e H_1 : pelo menos uma das distribuições tem um parâmetro de localização que é menor ou maior do que os outros (pelo menos um grupo difere significativamente e o efeito do factor não é desprezável). Na tomada de decisão rejeitamos H_0 , para um dado nível de significância α , se: $H_{\text{calc}} > H_{(1-\alpha, df)}$. Obtem-se o valor crítico $H_{(1-\alpha, df)}$ da tabela Qui-quadrado apresentada no teste de Kruskal-Wallis. Para a avaliação dos efeitos da interacção (um por cada interacção), definimos as hipóteses H_0 (não há nenhuma interacção entre os factores), isto é, $\forall i, j: \gamma_{i,j} = 0 (i \neq j)$ e H_1 (há interacção entre pelo menos dois factores), isto é, $\exists i, j: \gamma_{i,j} \neq 0 (i \neq j)$. Na tomada de decisão, rejeitamos H_0 , para um dado nível de significância α , se $H_{\text{calc}} > H_{(1-\alpha, df)}$.

Anexo 3 Folha de Excel usado para predição - fórmulas

Mostramos neste anexo o Excel usado para predição, modelo específico para domínio Jornais.

	P4	P5	P6	P7	P8	P9
Accumulated	=1-EXP(-EXP(C8))	=1-EXP(-EXP(D8))	=1-EXP(-EXP(E8))	=1-EXP(-EXP(F8))	=1-EXP(-EXP(G8))	1
Absolute prob.	=C2	=+D2-C2	=+E2-D2	=+F2-E2	=+G2-F2	=+H2-G2
	4	5	6	7	8	9

Thresholds ->	-1,91	-1,24	0,01	2,037	3,31
-------------------------	-------	-------	------	-------	------

=+SUM(D11:D29)	=C6-\$B\$8	=D6-\$B\$8	=E6-\$B\$8	=F6-\$B\$8	=G6-\$B\$8
----------------	------------	------------	------------	------------	------------

	ESTIMATES	CURRENT SITE	
links_num_extern_links	0,006	31,8667	=+C11*B11
links_num_image_links	-0,008	5,36667	=+C12*B12
text_heading_reverse_order	0,002	7,63333	=+C13*B13
links_links_title	-0,006	6,03333	=+C14*B14
text_num_diferent_colors	0,02	39	=+C15*B15
pagelayout_device_specific	0,022	21	=+C16*B16
pagelayout_num_tables	-0,013	4,06667	=+C17*B17
accessibility_img_title	-0,172	0,0148148	=+C18*B18
text_font_size_max_em	-0,139	1,5	=+C19*B19
forms_labels	1,151	1	=+C20*B20
efficiency_css_size	-0,000002756	223067	=+C21*B21
[area_=1]	-0,32		=+C22*B22
[area_=2]	-2,293		=+C23*B23
[area_=3]	1,286	1	=+C24*B24
[area_=4]	0,327		=+C25*B25
[area_=5]	-0,663		=+C26*B26
[area_=6]	-0,001		=+C27*B27
[area_=7]	0,898		=+C28*B28
[area_=8]	0		=+C29*B29

Per

PR

RESULT
=+LARGE(C3:H3;1)
=HLOOKUP(G13;C3:H4;2;0)